

INSIGHTS INTO HOW THE STRUCTURAL FEATURES OF DNA ADDUCTS  
DICTATE LOCAL HELICAL CONFORMATION AND REPAIRABILITY: A  
COMPUTATIONAL STUDY

Ryan Kung  
Bachelor of Science, University of Lethbridge, 2015

A thesis submitted  
in partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Department of Chemistry and Biochemistry  
University of Lethbridge  
LETHBRIDGE, ALBERTA, CANADA

© Ryan Kung, 2024

INSIGHTS INTO HOW THE STRUCTURAL FEATURES OF DNA ADDUCTS  
DICTATE LOCAL HELICAL CONFORMATION AND REPAIRABILITY: A  
COMPUTATIONAL STUDY

RYAN KUNG

Date of Defence: December 10, 2021

|                                  |           |       |
|----------------------------------|-----------|-------|
| Dr. Stacey Wetmore<br>Supervisor | Professor | Ph.D. |
|----------------------------------|-----------|-------|

|                                                         |                     |       |
|---------------------------------------------------------|---------------------|-------|
| Dr. Peter Dibble<br>Thesis Examination Committee Member | Associate Professor | Ph.D. |
|---------------------------------------------------------|---------------------|-------|

|                                                           |           |       |
|-----------------------------------------------------------|-----------|-------|
| Dr. Michael Gerken<br>Thesis Examination Committee Member | Professor | Ph.D. |
|-----------------------------------------------------------|-----------|-------|

|                                                             |                     |       |
|-------------------------------------------------------------|---------------------|-------|
| Dr. Jean-Denys Hamel<br>Thesis Examination Committee Member | Assistant Professor | Ph.D. |
|-------------------------------------------------------------|---------------------|-------|

|                                                                                                     |           |       |
|-----------------------------------------------------------------------------------------------------|-----------|-------|
| Dr. Irina Paci<br>External Examiner<br>University of Victoria<br>Victoria, British Columbia, Canada | Professor | Ph.D. |
|-----------------------------------------------------------------------------------------------------|-----------|-------|

|                                                         |           |       |
|---------------------------------------------------------|-----------|-------|
| Dr. Marc Roussel<br>Chair, Thesis Examination Committee | Professor | Ph.D. |
|---------------------------------------------------------|-----------|-------|



## **ABSTRACT**

DNA damage occurs regularly and results in various biological consequences, including cancer. One specific type of DNA damage is the formation of adducts, which arise from various sources, including cigarette smoke and pesticides. The local helical structure of adducted DNA is dictated by the chemical composition of the lesion. Furthermore, various known conformations of adducted DNA have been shown to differentially impact DNA repair and replication. Although this underscores how the lesion dictates biological outcomes, structure–function relationships have yet to be fully explored. Thus, this thesis uses computational modelling to examine the complex interplay between the adduct chemical composition (e.g., bulky moiety shape, adduct linker type, and lesion number) and the resulting damaged DNA structure and lesion mutagenicity. Overall, novel trends in the helical conformation and lesion repairability as a function of the chemical structure of the DNA adduct are uncovered, which have implications in the severity of the long-term biological consequences.

## ACKNOWLEDGMENTS

First of all, I would like to thank my supervisor, Dr. Stacey Wetmore, for all of her encouragement and support. I can honestly say that I would not have been able to achieve this without her and I am deeply grateful to have such a wonderful mentor. I would also like to thank my committee members, Dr. Peter Dibble and Dr. Michael Gerken, for all the support and advice they have provided through my graduate journey. Next, I would like to thank Dr. Jean-Denys Hamel for agreeing to be my external internal examiner and Dr. Marc Roussel for chairing my defense. Finally, I would like to thank my external examiner, Dr. Irina Paci, for her valuable time and insight.

Thank-you to the Natural Sciences and Engineering Research Council (NSERC), Alberta Innovates – Technology Futures, and the University of Lethbridge for student scholarships. I would also like to thank the University of Lethbridge, the University of Calgary, Westgrid and Compute Canada for providing valuable computational resources.

I consider myself blessed to have worked alongside many incredible people throughout my years at the University of Lethbridge. Included among them are the high school and undergraduate students that I have had the pleasure of supervising during my time as a graduate student: Chasen Jackle, Simmone D'souza, Emma Arnell, Cassidy Griffith, Janelle Bykowski, Trinity Deak, Michaela Reitsma, Chalalai Phawaphootanon and Rian Opperman. Furthermore, I am very grateful for all the past and current members of the Wetmore lab: Dr. Purshotam Sharma, Dr. Katie Wilson, Dr. Stefan Lenz, Dr. Mohamed Aboelnga, Dr. Preethi Seelam, Dr. Shahin Sowlati Hashjin, Dr. James McFarlane, Briana Boychuk, Rajwinder Kaur, Dylan Nikkel, Makay Murray, Prakash

Amruth Raj, Nathania Takyi, Cynthia Fonderson, Rebecca Jeong, Bimaldeep Singh, Lindey Felske, Priya Bhutani and Hanzala Hussain. I will treasure the many great times we have shared for the rest of my life.

Finally, I want to thank all my family and friends. I consider myself extremely blessed to have so many amazing people in my life. Lastly, I need to thank my wonderful wife, Katri, whose love and support mean more to me than I could ever describe. Thank-you for adventuring on this journey with me.

## **Table of Contents**

|                         |       |
|-------------------------|-------|
| Abstract .....          | iii   |
| Acknowledgments .....   | iv    |
| Table of Contents ..... | vi    |
| List of Schemes.....    | xviii |
| List of Tables .....    | xviii |
| List of Figures .....   | xix   |
| Abbreviations .....     | xxx   |

## **Chapter 1: Introduction**

|                                      |    |
|--------------------------------------|----|
| 1.1 General Overview .....           | 1  |
| 1.2 DNA Structure .....              | 1  |
| 1.3 DNA Damage .....                 | 4  |
| 1.3.1 DNA Nucleobase Adducts .....   | 5  |
| 1.4 Replication of DNA Adducts ..... | 9  |
| 1.5 Repair of DNA Adducts .....      | 11 |

|                                            |    |
|--------------------------------------------|----|
| 1.6 Computer Modelling of DNA Adducts..... | 13 |
| 1.7 Thesis Overview.....                   | 13 |
| 1.8 References .....                       | 16 |

## **Chapter 2: Effect of the Size and Shape of Nitrogen-Containing Aromatics on the Conformational Preferences of DNA Containing Damaged Guanine**

|                                                                                                                                                          |    |
|----------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 2.1 Introduction .....                                                                                                                                   | 23 |
| 2.2 Computational Methods .....                                                                                                                          | 28 |
| 2.2.1 DFT Calculations.....                                                                                                                              | 28 |
| 2.2.2 MD Simulations.....                                                                                                                                | 29 |
| 2.3 Results and Discussion.....                                                                                                                          | 33 |
| 2.3.1 DFT Calculations Reveal that the Shape of the NCA Moiety has a Greater Influence on the Intrinsic Adduct Structure than the Size.....              | 33 |
| 2.3.1.1 Structure at the Guanine–NCA Linkage .....                                                                                                       | 33 |
| 2.3.1.2 Structure at the Glycosidic Linkage.....                                                                                                         | 34 |
| 2.3.2 MD Simulations Illustrate how the DNA Structural Characteristics and Conformational Preferences Depend on the Size and Shape of the NCA Moiety ... | 35 |
| 2.3.2.1 General Structural Features of DNA Conformational Themes .....                                                                                   | 35 |

|                                                                                                                 |    |
|-----------------------------------------------------------------------------------------------------------------|----|
| 2.3.2.2 The Shape of the NCA Moiety Affects the Lesion Structural Characteristics in DNA .....                  | 37 |
| 2.3.2.3 The Size and Shape of the NCA Moiety Affect Lesion-Site Noncovalent Interactions.....                   | 38 |
| 2.3.2.4 The Shape of the NCA Moiety Affects its Solvent Accessibility.....                                      | 39 |
| 2.3.2.5 The Size and Shape of the NCA Moiety Affect the Relative Stability of Different DNA Conformations ..... | 42 |
| 2.3.3 Implications to the repair propensity and mutagenicity of N-linked C8-dG adducts .....                    | 45 |
| 2.4 Conclusion.....                                                                                             | 48 |
| 2.5 References .....                                                                                            | 49 |

### **Chapter 3: Structure of an Unusual Tetracyclic Deoxyguanosine Adduct: Implications for Frameshift Mutagenicity by Ortho–Cyano Nitroanilines**

|                                   |    |
|-----------------------------------|----|
| 3.1 Introduction .....            | 56 |
| 3.2 Experimental Procedures.....  | 59 |
| 3.2.1 Materials and Methods ..... | 59 |
| 3.2.2 Synthesis.....              | 60 |
| 3.3 Computational Details.....    | 62 |

|                                                          |    |
|----------------------------------------------------------|----|
| 3.3.1 DFT Calculations.....                              | 62 |
| 3.3.2 MD Simulations.....                                | 63 |
| 3.4 Results and Discussion.....                          | 66 |
| 3.4.1 Spectral Properties and Hydrolytic Stability ..... | 66 |
| 3.4.2 Conformation.....                                  | 71 |
| 3.4.3 MD Simulations of Adducted <i>NarI</i> .....       | 75 |
| 3.5 Conclusions .....                                    | 79 |
| 3.6 References .....                                     | 80 |

## **Chapter 4: Effect of Guanine Adduct Size, Shape and Linker Type on the Conformation of Damaged DNA: A Molecular Dynamics Study**

|                                                                                                                                                              |    |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 4.1 Introduction .....                                                                                                                                       | 85 |
| 4.2 Materials and Methods .....                                                                                                                              | 91 |
| 4.2.1 DFT Calculations.....                                                                                                                                  | 91 |
| 4.2.2 MD Simulations.....                                                                                                                                    | 92 |
| 4.3 Results .....                                                                                                                                            | 95 |
| 4.3.1 DFT Calculations Suggest that the Shape, but not Size, of the Bulky Moiety<br>Impacts the Inherent Structural Preferences of C-linked dG Adducts ..... | 95 |

|                                                                                                                                                                                              |     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 4.3.1.1 Structure at the Aryl–Nucleobase Linker .....                                                                                                                                        | 95  |
| 4.3.1.2 Preferred Glycosidic Orientation of C-linked Nucleosides .....                                                                                                                       | 96  |
| 4.3.2 MD Simulations Reveal that Bulky Moiety Shape and Size Impact the C-linked Adducted DNA Conformational Preference .....                                                                | 96  |
| 4.3.2.1 Overall Helical Features are Independent of C-linked Adduct.....                                                                                                                     | 96  |
| 4.3.2.2 C-linked Adduct Orientation is Dependent on Lesion Shape.....                                                                                                                        | 98  |
| 4.3.2.3 Nonbonded Lesion-Site Interactions are Dependent on the Damaged DNA Conformation as well as the Bulky Group Size and Shape, Particularly for the <i>syn</i> Adduct Orientation ..... | 98  |
| 4.3.2.4 Lesion Size Affects the Solvent Exposure of the Aryl Group.....                                                                                                                      | 100 |
| 4.3.2.5 The Preferred Conformation of C-linked Adducted DNA is Impacted by Lesion Shape and Size .....                                                                                       | 102 |
| 4.3.3 DFT Calculations Show that the Aryl-G Linker Orientation Preferences are Dependent on Lesion Shape and not Size for O-linked dG Adducts .....                                          | 104 |
| 4.3.3.1 Structure at the Aryl–O–Nucleobase Linker .....                                                                                                                                      | 104 |
| 4.3.3.2 Preferred Glycosidic Orientation of O-linked Nucleosides.....                                                                                                                        | 105 |
| 4.3.4 MD Simulations Show that the O-linked Adducted DNA Structural Preferences are Dictated by the Aryl Moiety Shape and Size.....                                                          | 106 |
| 4.3.4.1 Overall Helical Features are Independent of O-linked Adduct .....                                                                                                                    | 106 |
| 4.3.4.2 O-linked Adduct Orientation is Affected by Aryl Moiety Shape.....                                                                                                                    | 106 |
| 4.3.4.3 The Strength of Nonbonded Lesion-Site Interactions are Dependent on the Bulky Group Size and Shape for O-linked Adducts, Especially in the <i>syn</i> Orientation .....              | 107 |



|                                                                                                                                                                            |     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 4.3.4.4 The Lesion Size Rather Than Shape Primarily Determines Solvent Exposure of the Aryl Group for O-linked Adducts .....                                               | 109 |
| 4.3.4.5 The Shape and Size of the O-linked Adduct Impacts the Helical Conformational Preferences .....                                                                     | 110 |
| 4.4 Discussion .....                                                                                                                                                       | 113 |
| 4.4.1 The Linker Type has a Subtle Effect on the Preferred Damaged DNA Conformation Due to the Unique Inherent Structural Features of C, O and N-linked C8-dG Adducts..... | 113 |
| 4.4.2 The Shape of the Aryl Group Affects the Conformations Intrinsically Accessible to the C8-dG Adducts and therefore the Local Helical Structure of Damaged DNA .....   | 116 |
| 4.4.3 The Bulky Group Size has an Even Larger Effect on Damaged DNA Structure than Shape and Linker Type.....                                                              | 117 |
| 4.5 Conclusion.....                                                                                                                                                        | 120 |
| 4.6 References .....                                                                                                                                                       | 121 |

## **Chapter 5: Conformational Preference and Fluorescence Response of a C-Linked C8-Biphenyl-Guanine Lesion in the *NarI* Mutational Hotspot: Evidence for Enhanced *Syn* Adduct Formation**

|                        |     |
|------------------------|-----|
| 5.1 Introduction ..... | 128 |
|------------------------|-----|

|                                                                                                                |     |
|----------------------------------------------------------------------------------------------------------------|-----|
| 5.2 Materials and Methods .....                                                                                | 132 |
| 5.2.1 Materials .....                                                                                          | 132 |
| 5.2.2 Methods .....                                                                                            | 133 |
| 5.3 Results .....                                                                                              | 135 |
| 5.3.1 Optical Properties of FBP-dG in <i>NarI</i> .....                                                        | 135 |
| 5.3.2 Conformational Analysis of FBP dG in <i>NarI</i> .....                                                   | 138 |
| 5.4 Discussion .....                                                                                           | 149 |
| 5.4.1 Conformation of FBP-dG in <i>NarI</i> : Comparison to other C-linked and N-linked<br>C8-dG Adducts ..... | 149 |
| 5.4.2 Fluorescence of FBP-dG in Fully Paired versus SMI <i>NarI</i> Duplex .....                               | 151 |
| 5.5 Conclusions .....                                                                                          | 153 |
| 5.6 References .....                                                                                           | 154 |

## **Chapter 6: Impact of DNA Adduct Size, Number and Relative Position on the Toxicity of Aromatic Amines: A Molecular Dynamics Case Study of <sup>AN</sup>dG and <sup>AP</sup>dG Containing DNA**

|                        |     |
|------------------------|-----|
| 6.1 Introduction ..... | 159 |
|------------------------|-----|

|                                                                                                                                                                                                                                                                      |     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 6.2 Computational Details.....                                                                                                                                                                                                                                       | 167 |
| 6.3 Results .....                                                                                                                                                                                                                                                    | 170 |
| 6.3.1 General Structural Features of Monoadducted DNA .....                                                                                                                                                                                                          | 170 |
| 6.3.1.1 B Conformation of Monoadducted DNA .....                                                                                                                                                                                                                     | 170 |
| 6.3.1.2 S Conformation of Monoadducted DNA.....                                                                                                                                                                                                                      | 171 |
| 6.3.2 Conformational Preference of Monoadducted DNA.....                                                                                                                                                                                                             | 172 |
| 6.3.3 General Structural Features of Diadducted DNA .....                                                                                                                                                                                                            | 177 |
| 6.3.3.1 B Conformation of Diadducted DNA.....                                                                                                                                                                                                                        | 177 |
| 6.3.3.2 S Conformation of Diadducted DNA .....                                                                                                                                                                                                                       | 178 |
| 6.3.4 Conformational Preference of Diadducted DNA .....                                                                                                                                                                                                              | 179 |
| 6.3.4.1 <sup>AN</sup> dG Diadducted DNA.....                                                                                                                                                                                                                         | 180 |
| 6.3.4.2 <sup>AP</sup> dG Diadducted DNA .....                                                                                                                                                                                                                        | 182 |
| 6.4 Discussion .....                                                                                                                                                                                                                                                 | 186 |
| 6.4.1 Despite Other N-linked C8-dG Adducts Exhibiting Sequence Dependence<br>Conformational Heterogeneity, <sup>AN</sup> dG and <sup>AP</sup> dG Induce Monoadducted DNA<br>Conformational Homogeneity, which is Exclusively Controlled by Aryl Moiety<br>Size ..... | 186 |
| 6.4.2 The Preferred Damaged DNA Conformation Can Change Upon Addition of a<br>Second Adduct, with Neighboring Lesions Stabilizing the B Conformation and<br>Next Nearest Damaged Sites Stabilizing the S Conformation Regardless of Adduct<br>Size .....             | 188 |

|                                                                                                                                                               |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 6.4.3 Implications of Single and Cluster Lesions on the Mutagenicity and Repair of <sup>AN</sup> dG and <sup>AP</sup> dG-containing DNA .....                 | 189 |
| 6.4.3.1 Toxicity of <sup>AN</sup> dG and <sup>AP</sup> dG in Monoadducted DNA is Most Significantly Dictated by Lesion Size Rather than Sequence Context..... | 190 |
| 6.4.3.2 Adduct Toxicity can Increase or Decrease Upon a Second Damaging Event Depending on Lesion Size and Relative Position.....                             | 191 |
| 6.5 Conclusions .....                                                                                                                                         | 194 |
| 6.6 References .....                                                                                                                                          | 195 |

## **Chapter 7: Effects of a Second Local DNA Damage Event on the Mutagenicity of the Human Carcinogen 4-Aminobiphenyl: A Molecular Dynamics Study of Damaged DNA Structure**

|                                                                                 |     |
|---------------------------------------------------------------------------------|-----|
| 7.1 Introduction .....                                                          | 204 |
| 7.2 Computational Details.....                                                  | 211 |
| 7.3 Results .....                                                               | 215 |
| 7.3.1 General Structural Features of <sup>ABP</sup> dG Monoadducted DNA .....   | 215 |
| 7.3.1.1 Monoadducted DNA Containing an <i>anti</i> Lesion .....                 | 216 |
| 7.3.1.2 Monoadducted DNA Containing a <i>syn</i> Lesion .....                   | 216 |
| 7.3.2 <i>anti/syn</i> Preferences of <sup>ABP</sup> dG in Monoadducted DNA..... | 218 |

|                                                                                                                                                                                                |     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 7.3.3 General Structural Features of <sup>ABP</sup> dG Diadducted DNA.....                                                                                                                     | 222 |
| 7.3.3.1 Impact of a Second Lesion on DNA Containing an <i>anti</i> Adduct .....                                                                                                                | 223 |
| 7.3.3.2 Impact of a Second Lesion on DNA Containing a <i>syn</i> Adduct.....                                                                                                                   | 223 |
| 7.3.4 <i>anti/syn</i> Preferences of <sup>ABP</sup> dG in Diadducted DNA .....                                                                                                                 | 225 |
| 7.4 Discussion .....                                                                                                                                                                           | 230 |
| 7.4.1 Sequence Context Impacts the Preferred Conformation of <sup>ABP</sup> dG in<br>Monoadducted DNA .....                                                                                    | 230 |
| 7.4.2 Inherent Biphenyl Flexibility Results in Unique Conformational Preferences<br>.....                                                                                                      | 232 |
| 7.4.3 Addition of a Secondary <sup>ABP</sup> dG Adduct Differentially Impacts the Helical<br>Structure, Conformational Preferences and Carcinogenicity Depending on Lesion<br>Separation ..... | 233 |
| 7.5 Conclusions .....                                                                                                                                                                          | 235 |
| 7.6 References .....                                                                                                                                                                           | 236 |

## **Chapter 8: Understanding the Nucleotide Excision Repair of the 2-Acetylaminofluorene-C8-dG Adduct: Impact of Number of Adducted Sites and Helical Conformation**

|                        |     |
|------------------------|-----|
| 8.1 Introduction ..... | 244 |
|------------------------|-----|

|                                                                                                                                                                                                                                                                             |     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 8.2 Computational Details.....                                                                                                                                                                                                                                              | 251 |
| 8.2.1 General Simulation Protocol .....                                                                                                                                                                                                                                     | 251 |
| 8.2.2 Model Generation for Isolated DNA Duplexes.....                                                                                                                                                                                                                       | 253 |
| 8.2.3 Model Generation for Rad4–DNA Complexes .....                                                                                                                                                                                                                         | 254 |
| 8.2.4 Umbrella Sampling of the Rad4–DNA Binding Pathway.....                                                                                                                                                                                                                | 255 |
| 8.2.5 Unrestrained MD Simulations of Rad4–DNA Complexes .....                                                                                                                                                                                                               | 256 |
| 8.3 Results and Discussion.....                                                                                                                                                                                                                                             | 257 |
| 8.3.1 The Conformation of <sup>AAF</sup> dG Monoadducted DNA Impacts the Predicted<br>Barriers and Stabilities of Rad4 Binding, Indicating Differing Repair Propensities<br>Recognition Process .....                                                                       | 257 |
| 8.3.2 The Binding of Rad4 to Diadducted DNA is Distinct from the Monoadducted<br>Systems Being More Thermodynamically and Less Kinetically Preferred .....                                                                                                                  | 262 |
| 8.3.3 Rad4 Binding to Undamaged DNA is Energetically Feasible During the<br>Coordination of the BHDs and the Base Flipping Residues but Overall is<br>Thermodynamically and Kinetically Unfavourable, Which Provides Further<br>Evidence of a Kinetic Gating Mechanism..... | 265 |
| 8.3.4 Unlike Other Lesions, Neither Helical Distortions nor BHD2 Interactions in<br>the Prebound Rad4–DNA Complex are Sufficient Indicators of NER Efficiency for<br><sup>AAF</sup> dG Adducted DNA Mechanism.....                                                          | 268 |
| 8.3.5 Helical Bending at the Damaged Site in the Productively Bound Rad4–DNA<br>Complex is Correlated with Calculated Barriers and Complex Stabilities .....                                                                                                                | 273 |

|                                                                     |            |
|---------------------------------------------------------------------|------------|
| 8.4 Conclusions .....                                               | 276        |
| 8.5 References .....                                                | 279        |
| <br><b>Chapter 9: Conclusions</b>                                   |            |
| 9.1 Thesis Overview.....                                            | 284        |
| 9.2 Future Work .....                                               | 290        |
| 9.3 Final Remarks .....                                             | 297        |
| 9.4 References .....                                                | 298        |
| <br><b>Appendix A: Supplemental Information for Chapter 2.....</b>  | <b>304</b> |
| <br><b>Appendix B: Supplemental Information for Chapter 3 .....</b> | <b>397</b> |
| <br><b>Appendix C: Supplemental Information for Chapter 4.....</b>  | <b>443</b> |
| <br><b>Appendix D: Supplemental Information for Chapter 5.....</b>  | <b>481</b> |
| <br><b>Appendix E: Supplemental Information for Chapter 6 .....</b> | <b>517</b> |

**Appendix F: Supplemental Information for Chapter 7 .....565**

**Appendix G: Supplemental Information for Chapter 8.....615**



## List of Schemes

|                                                                                                                                                   |    |
|---------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Scheme 3.1: Synthesis of closed, cyclic adduct (1) and open, para adduct (2) from 2-aminobenzonitrile and 4-aminobenzonitrile, respectively ..... | 58 |
|---------------------------------------------------------------------------------------------------------------------------------------------------|----|

## List of Tables

|                                                                                                                                               |     |
|-----------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Table 3.1: Diagnostic $^{13}\text{C}$ -NMR Chemical Shifts in DMSO .....                                                                      | 68  |
| Table 3.2: Summary of Rate Data for Acid-Catalyzed Hydrolysis of <b>1</b> and <b>2</b> .....                                                  | 71  |
| Table 5.1: Thermal Melting Parameters of FBP-dG modified <i>NarI</i> .....                                                                    | 136 |
| Table 5.2: MM-PBSA relative energies (kJ/mol) for different conformations of FPh-dG- and FBP-dG-adducted fully paired <i>NarI</i> duplex..... | 138 |
| Table 5.3: Properties of C-linked C8-aryl-dG adducts in fully paired and SMI <i>NarI</i> 12-mer duplexes .....                                | 152 |
| Table 6.1: DNA duplexes modelled in Chapter 6 .....                                                                                           | 166 |
| Table 7.1: DNA duplexes modelled in Chapter 7 .....                                                                                           | 210 |
| Table 8.1: Relative free energies (kJ/mol) for Rad4–DNA binding with respect to the corresponding initially bound structure .....             | 259 |

## List of Figures

|                                                                                                                                                                                                                                                                                 |    |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figure 1.1: Structures of A) the four canonical nucleotides, including atom numbering and the $\chi$ dihedral angle and B) a double-stranded B-DNA duplex, highlighting the major and minor grooves.....                                                                        | 3  |
| Figure 1.2: Structures of the WC A:T (left) and G:C (right) base pairs .....                                                                                                                                                                                                    | 3  |
| Figure 1.3: DNA structures extracted from duplexes, highlighting the twist and rise helical parameters .....                                                                                                                                                                    | 4  |
| Figure 1.4: Structures of sample DNA damage products, with the damaged portion shown in red .....                                                                                                                                                                               | 5  |
| Figure 1.5: Structures of sample DNA adducts, highlighting the diversity in chemical structures. ....                                                                                                                                                                           | 6  |
| Figure 1.6: Example structures of C8-dG adducted DNA, in the major groove (B, left), minor groove wedge (W, center) and intercalated stacked (S, right) conformations .....                                                                                                     | 7  |
| Figure 1.7: Structures of a generic C8-dG adduct (R = bulky group) in the <i>anti</i> (left) and <i>syn</i> (right) orientations, which expose the WC (blue) and Hoogsteen (red) face of the damaged G to the opposing strand, respectively. ....                               | 8  |
| Figure 1.8: Crystal structures of C8-dG adducted DNA duplexes containing a deletion frameshift mutation around <sup>AF</sup> dG (left, PDB ID: 1AX6) or a C to A point mutation opposite <sup>AP</sup> dG (right, PDB ID: 1AXU; black arrow points to the mispaired base). .... | 11 |
| Figure 2.1: Structures and chemical numbering of N-linked DNA adducts at the C <sup>8</sup> -site of 2'-deoxyguanosine. ....                                                                                                                                                    | 24 |

|                                                                                                                                                                                                                                                                                                         |    |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figure 2.2: MD representative structures of the major groove (B, left), wedge (W, middle) and base-displaced stacked (S, right) conformations of DNA containing N-linked C <sup>8</sup> -dG adducts .....                                                                                               | 25 |
| Figure 2.3: Effect of ring-fusion on the $\phi$ torsion angle, illustrated using the approximately planar <sup>3</sup> APHN dG (A, C) and nonplanar <sup>1</sup> ANP dG (B, D) nucleobase (A, B) and nucleoside (C, D) adducts. ....                                                                    | 34 |
| Figure 2.4: Discrete hydrogen-bonding interactions observed during MD simulations on the (A) B, (B) W and (C) S conformations of DNA containing N-linked C <sup>8</sup> -dG adducts. <sup>2</sup> ANP dG adducted DNA is shown as an example .....                                                      | 36 |
| Figure 2.5: Effect of size and shape of the C <sup>8</sup> -moiety on the lesion-site interaction energy (LSIE) within the B, W and S conformations of damaged DNA containing an N-linked dG adducts .....                                                                                              | 40 |
| Figure 2.6: Dependence of the solvent accessible surface area (SASA, Å <sup>2</sup> ) on the shape and size of the hydrophobic NCA for the B (red), W (blue) and S (green) conformations of damaged DNA containing N-linked dG adducts .....                                                            | 41 |
| Figure 2.7: Dependence of the solvent accessible surface area (SASA, Å <sup>2</sup> ) and lesion-site interaction energy (LSIE, kJ mol <sup>-1</sup> ) on the shape and size of the hydrophobic C <sup>8</sup> -moiety for the (A) B and (B) S conformations of DNA damaged by N-linked dG adducts..... | 43 |
| Figure 3.1: IR spectral overlay traces of <b>1</b> and <b>2</b> depicting the CN stretch for adduct <b>2</b> ....                                                                                                                                                                                       | 66 |
| Figure 3.2: <sup>1</sup> H NMR spectra of <b>1</b> and <b>2</b> displaying aromatic peaks and C8-NH for <b>2</b> .....                                                                                                                                                                                  | 67 |

|                                                                                                                                                                                                                                                                                                                                                   |    |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figure 3.3: (A) UV spectral overlay traces of <b>1</b> (black trace) and <b>2</b> (red trace). (B) Excitation and emission spectra of <b>1</b> in water. (C) Picture highlighting the blue fluorescence of <b>1</b> (right) and lack of fluorescence for adduct <b>2</b> (left).....                                                              | 68 |
| Figure 3.4: Changes in absorbance spectra of (A) adduct <b>2</b> and (B) adduct <b>1</b> in 0.1 M HCl at 37 °C recorded in ~ 1 min intervals (dashed red traces, initial trace in black) .....                                                                                                                                                    | 71 |
| Figure 3.5: Selective 1D ROESY spectrum of <b>1</b> in DMSO-d <sub>6</sub> at ambient temperature highlighting the cross-peak from the sugar H1' to the aromatic H12 proton for a <i>syn</i> conformational preference .....                                                                                                                      | 72 |
| Figure 3.6: B3LYP-D3(BJ)/6-31G(d,p) <i>anti</i> (left) and <i>syn</i> (right) minima for the free nucleosides of A) <b>1</b> with no constraints, B) <b>2</b> with no constraints, C) <b>1</b> with the 5'-OH constrained, and D) <b>2</b> with the 5'-OH unconstrained.....                                                                      | 73 |
| Figure 3.7: PMF for the rotation about the glycosidic bond for the free nucleoside adduct <b>1</b> (blue) and <b>2</b> (green) in water.....                                                                                                                                                                                                      | 75 |
| Figure 3.8: Representative SMI structures. A) Distorted SMI containing <b>1</b> in the <i>anti</i> conformation, B) planar, stabilized SMI containing <b>1</b> in the <i>syn</i> conformation, C) distorted SMI containing <b>2</b> in the <i>anti</i> conformation, and D) distorted SMI containing <b>2</b> in the <i>syn</i> conformation..... | 77 |
| Figure 3.9: Representative structures of the FP duplexes containing adduct <b>1</b> in the A) B conformation or B) S conformation or containing adduct <b>2</b> in the C) B conformation or D) W conformation .....                                                                                                                               | 78 |

|                                                                                                                                                                                                                                                                                                  |     |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 4.1: General conformational themes for C8-dG adducted DNA as shown using MD representative structures of <sup>Py</sup> dG adducted DNA in the major groove (B, left), wedge (W, middle), and base-displaced stacked (S, right) conformations .....                                        | 88  |
| Figure 4.2: Structures, chemical numbering and nomenclature for model N, O and C-linked DNA adducts at the C8-site of 2'-deoxyguanosine. Key dihedral angles for the C-linked adducts defined for representative lesions .....                                                                   | 89  |
| Figure 4.3: Hydrogen-bonding interactions at the lesion site in the (A) B, (B) W and (C) S conformations of adducted DNA, using <sup>2NP</sup> dG adducted DNA as examples for the C-linked systems (top) and <sup>2ONP</sup> dG adducted DNA as examples for the O-linked systems (bottom)..... | 97  |
| Figure 4.4: Effect of C8-moiety A) size and shape in the B conformation, B) size and shape in the W conformation, C) shape in the S conformation, and D) size in the S conformation on the lesion-site interaction energy (LSIE) for damaged DNA containing a C-linked dG adduct .....           | 99  |
| Figure 4.5: Dependence of the solvent accessible surface area (SASA, Å <sup>2</sup> ) on the shape and size of the hydrophobic NCA for the B (red), W (blue) and S (green) conformations of damaged DNA containing a C-linked dG adduct .....                                                    | 10  |
| Figure 4.6: Dependence of the solvent accessible surface area (SASA, Å <sup>2</sup> ) and lesion-site interaction energy (LSIE, kJ/mol) on the shape and size of the hydrophobic C8-moiety for the (A) B and (B) S conformations of DNA containing a C-linked dG adduct.....                     | 103 |
| Figure 4.7: Effect of C8-moiety A) size and shape in the B conformation, B) size and shape in the W conformation, C) shape in the S conformation, and D) size in the S                                                                                                                           |     |

|                                                                                                                  |     |
|------------------------------------------------------------------------------------------------------------------|-----|
| conformation on the lesion-site interaction energy (LSIE) for damaged DNA containing an O-linked dG adduct ..... | 108 |
|------------------------------------------------------------------------------------------------------------------|-----|

|                                                                                                                                                                                                                                               |     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 4.8: Dependence of the solvent accessible surface area (SASA, Å <sup>2</sup> ) on the shape and size of the hydrophobic NCA for the B (red), W (blue) and S (green) conformations of damaged DNA containing an O-linked dG adduct..... | 110 |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|

|                                                                                                                                                                                                                                                                               |     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 4.9: Dependence of the solvent accessible surface area (SASA, Å <sup>2</sup> ) and lesion-site interaction energy (LSIE, kJ/mol) on the shape and size of the hydrophobic C8-moiety for the (A) B and (B) S conformations of DNA containing an O-linked dG adduct..... | 111 |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|

|                                                                                                                                                                                                                                  |     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 4.10: Effect of aryl moiety size on the SASA (A and B) and LSIE (C and D) for the B (A and C) and S (B and D) conformations, averaged over C8-dG adducts with a C-linker (Green), O-linker (Red) or N-linker (Blue) ..... | 118 |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|

|                                                                                                                                                                                                                                                                                                                                                  |     |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 5.1: A) N-linked AF-dG, AAF-dG and ABP-dG adducts, B) C-linked FBP-dG adduct and C) sequences of the fully paired and slippage mutagenic intermediate (SMI) duplexes, where N:M=C:G, G:C, A:T or T:A in the fully paired duplex, and N=C, G, A or T in the SMI duplex; N:M=C:G represents the fully paired 12-mer <i>NarI</i> sequence... | 129 |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|

|                                                                                                                                                                                                                                                                                          |     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 5.2: Excitation and emission spectra of FBP-dG-modified 12-mer <i>NarI</i> oligonucleotides in the single-strand, fully paired duplex and SMI duplex. All spectra of oligonucleotides (6 mM) were recorded in 50 mM sodium phosphate buffer, pH 7, with 0.1 M NaCl at 10 °C ..... | 137 |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|

|                                                                                                                                                                                                                                                   |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 5.3: FBP-dG adducted <i>NarI</i> (N=C) in the A) major-groove, B) wedge conformation, C) intercalated conformation, and D) base-displaced intercalated conformation Only the three base pairs about the lesion are shown for clarity ..... | 139 |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|

|                                                                                                                                                                                                                                                   |            |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| Figure 5.4: FPh-dG adducted <i>NarI</i> (N=C) in the A) major-groove, B) wedge conformation, C) intercalated conformation, and D) base-displaced intercalated conformation Only the three base pairs about the lesion are shown for clarity ..... | 140        |
| Figure 5.5: Frequency of the $\xi$ dihedral angle (Figure D.2, Appendix D) for FBP-dG-adducted fully paired <i>NarI</i> duplex in the major-groove (B), stacked (S), intercalated (I), and wedge (W) conformations .....                          | 142        |
| Figure 5.6: FBP-dG-adducted <i>NarI</i> paired against a –2 deletion strand with the adduct in the A) <i>anti</i> or B) <i>syn</i> orientation. Only the lesion site including the 2 base pairs flanking the bulge are shown for clarity .....    | 143        |
| Figure 5.7: Overlay of the –2 deletion site for FPh-dG (blue) and FBP-dG (red) adducted helices with the adduct in A) <i>anti</i> or B) <i>syn</i> orientation .....                                                                              | <b>144</b> |
| Figure 5.8: Frequency of the $\xi$ dihedral angle (Figure D.2, Appendix D) for FBP-dG-adducted SMI DNA in the <i>anti</i> and <i>syn</i> conformations .....                                                                                      | 145        |
| Figure 5.9: Variable temperature proton decoupled $^{19}\text{F}$ NMR of A) FPh-dG- and B) FBP-dG-adducted fully paired <i>NarI</i> duplex; single-strand (ss), major-groove (B), wedge (W), and intercalated (I) .....                           | 145        |
| Figure 5.10: The isotopic effect for the proton decoupled $^{19}\text{F}$ NMR of the FBP-dG-adducted fully paired <i>NarI</i> duplex in 100% $\text{D}_2\text{O}$ and 90% $\text{H}_2\text{O}$ 10% $\text{D}_2\text{O}$ .....                     | 148        |
| Figure 5.11: The isotopic effect for the proton decoupled $^{19}\text{F}$ NMR of the FPh-dG-adducted fully paired <i>NarI</i> duplex in 100% $\text{D}_2\text{O}$ and 90% $\text{H}_2\text{O}$ 10% $\text{D}_2\text{O}$ .....                     | 149        |

|                                                                                                                                                                                                                                                                                                                                                                                  |     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 6.1: Chemical structures, as well as atom numbering and the $\chi$ dihedral angles, for C8-dG DNA adducts arising from AA compounds, including the <sup>AN</sup> dG and <sup>AP</sup> dG lesions (red) investigated in the present study .....                                                                                                                            | 160 |
| Figure 6.2: Example structures of A) <sup>AN</sup> dG adducted DNA in the B conformation and B) <sup>AP</sup> dG adducted DNA in the S conformation, highlighting the lesion and flanking base pairs. C) <i>NarI</i> oligomer, highlighting the sequence context of the damage sites considered in the present work (G <sub>1</sub> , G <sub>2</sub> , and G <sub>3</sub> )..... | 161 |
| Figure 6.3: Structures of the <sup>AN</sup> dG and <sup>AP</sup> dG adducts considered in the present work (top), as well as the different monoadducted (B and S) and diadducted (BB, BS, SB and SS) helical conformations modelled for each sequence combination .....                                                                                                          | 165 |
| Figure 6.4: A) LSIE <sup>avg</sup> and B) SASA <sup>avg</sup> for DNA containing a single <sup>AN</sup> dG or <sup>AP</sup> dG lesion at G <sub>1</sub> (red), G <sub>2</sub> (blue) or G <sub>3</sub> (green).....                                                                                                                                                              | 173 |
| Figure 6.5: MD representative structures of the preferred <sup>AN</sup> dG and <sup>AP</sup> dG monoadducted DNA conformations in different sequence contexts .....                                                                                                                                                                                                              | 175 |
| Figure 6.6: $\Delta(\text{LSIE}^{\text{avg}})$ versus $\Delta(\text{SASA}^{\text{avg}})$ for <sup>AN</sup> dG damaged DNA in the B (top) or S (bottom) conformation upon addition of a second adduct with the first lesion at G <sub>1</sub> (red), G <sub>2</sub> (blue) or G <sub>3</sub> (green).....                                                                         | 181 |
| Figure 6.7: $\Delta(\text{LSIE}^{\text{avg}})$ versus $\Delta(\text{SASA}^{\text{avg}})$ for <sup>AP</sup> dG damaged DNA in the B (top) or S (bottom) conformation upon addition of a second adduct with the first lesion at G <sub>1</sub> (red), G <sub>2</sub> (blue) or G <sub>3</sub> (green).....                                                                         | 183 |
| Figure 6.8: MD representative structures of the preferred <sup>AN</sup> dG and <sup>AP</sup> dG diadducted DNA conformations in different sequence .....                                                                                                                                                                                                                         | 185 |



|                                                                                                                                                                                                                                                                                                                                  |     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 7.1: Chemical structures of A) the <sup>ABP</sup> dG nucleotide investigated in the current study, including atom numbering and identification of key dihedral angles and B) various AA-derived N-linked C8-dG adducts .....                                                                                              | 205 |
| Figure 7.2: MD representative structures of <sup>ABP</sup> dG containing DNA in the B, S and W conformations .....                                                                                                                                                                                                               | 207 |
| Figure 7.3: MD representative structures of <sup>ABP</sup> dG containing DNA in the B, S, W and I conformations .....                                                                                                                                                                                                            | 213 |
| Figure 7.4: LSIE <sup>avg</sup> and SASA <sup>avg</sup> for DNA containing a single <sup>AN</sup> dG or <sup>AP</sup> dG averaged over all <i>NarI</i> three sites and <sup>ABP</sup> dG lesion at G <sub>1</sub> (red), G <sub>2</sub> (blue) or G <sub>3</sub> (green) in the <i>anti</i> or <i>syn</i> orientations .....     | 219 |
| Figure 7.5: MD representative structures of the preferred conformations of <sup>ABP</sup> dG monoadducted DNA for different sequence contexts. Only the base pairs in the 5'-CG <sub>1</sub> G <sub>2</sub> CG <sub>3</sub> C sequence context are shown in each case .....                                                      | 221 |
| Figure 7.6: Δ(LSIE <sup>avg</sup> ) versus Δ(SASA <sup>avg</sup> ) for <sup>ABP</sup> dG DNA upon addition of a second adduct with the first lesion in the <i>anti</i> (top) or <i>syn</i> (bottom) conformation for neighboring adducts (the G <sub>1</sub> G <sub>2</sub> sequence context) .....                              | 226 |
| Figure 7.7: Δ(LSIE <sup>avg</sup> ) versus Δ(SASA <sup>avg</sup> ) for <sup>ABP</sup> dG DNA upon addition of a second adduct with the first lesion in the <i>anti</i> (top) or <i>syn</i> (bottom) conformation for next-nearest neighboring adducts (the G <sub>2</sub> G <sub>3</sub> sequence context) .....                 | 227 |
| Figure 7.8: Δ(LSIE <sup>avg</sup> ) versus Δ(SASA <sup>avg</sup> ) for <sup>ABP</sup> dG DNA upon addition of a second adduct with the first lesion in the <i>anti</i> (top) or <i>syn</i> (bottom) conformation for adducts separated by two intervening nucleotides (the G <sub>1</sub> G <sub>3</sub> sequence context) ..... | 228 |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 7.9: MD representative structures of the preferred conformations of <sup>ABP</sup> dG diadducted DNA for different sequence contexts. Only the base pairs in the 5'-CG <sub>1</sub> G <sub>2</sub> CG <sub>3</sub> C sequence context are shown in each case .....                                                                                                                                                                                                                                                                                                                                     | 230 |
| Figure 8.1: A) Crystal structure of a representative Rad4–DNA complex (PDB ID: 6CFI). The chemical structure of B) CPD, C) <i>cis</i> -B[a]P-dG and D) the <sup>AAF</sup> dG adduct considered in the present work.....                                                                                                                                                                                                                                                                                                                                                                                       | 245 |
| Figure 8.2: Structures of <sup>AAF</sup> dG-containing DNA in the A) B conformation, B) S conformation, and C) W conformation .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 248 |
| Figure 8.3: MD representative structures of the four (damaged and undamaged) duplexes considered in the present work complexed with Rad4 .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                | 253 |
| Figure 8.4: Rad4– DNA complex during the binding pathway of Rad4 to <sup>AAF</sup> dG adducted DNA in the S conformation, starting from the first capture state, highlighting the different stages modelled using umbrella sampling including the A) initial association of the BHDs to the damaged DNA site (Stage 1), B) coordination of the base flipping residues to the 5' interstrand neighbor of the adducted nucleotide (Stage 2), C) flipping of the base opposing the lesion and its 3' neighbor (Stage 3), and D) insertion of the BHD3 β-hairpin loop into the interhelical space (Stage 4) ..... | 255 |
| Figure 8.5: Free energy profile for the binding of Rad4 to monoadducted DNA in the B conformation (blue), monoadducted DNA in the S conformation (orange), diadducted DNA in the BB conformation (yellow) or undamaged DNA (grey) over four umbrella sampling stages .....                                                                                                                                                                                                                                                                                                                                    | 258 |

|                                                                                                                                                                                                                                                                                                                               |     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 8.6: Average (and standard error) for the DNA structural properties and Rad4–DNA interactions (AlphaSpace volume) of the prebound state from quintuplicate 1 $\mu$ s MD simulations after the initial coordination (Stage 1), and the NER repair efficiencies of the adducted duplexes .....                           | 269 |
| Figure 8.7: Average (and standard error) for the DNA structural properties and Rad4–DNA interactions (AlphaSpace volume) of the bound state from quintuplicate 1 $\mu$ s MD simulations after the insertion of the BHD3 into the interhelical space (Stage 4), and the NER repair efficiencies of the adducted duplexes ..... | 273 |
| Figure 9.1: Structures of the N6-dA (left) and N2-dG (right) adducts derived aristolochic acids... ..                                                                                                                                                                                                                         | 291 |

## List of Abbreviations

|                      |                         |
|----------------------|-------------------------|
| <sup>1</sup> ANP-dG  | 1-aminonaphthyl-dG      |
| <sup>1</sup> APHN-dG | 1-aminophenanthrenyl-dG |
| 1NP                  | 1-nitropyrene           |
| <sup>1</sup> NP-dG   | 1-naphthyl-dG           |
| <sup>1</sup> ONP-dG  | 1-naphthol-dG           |
| <sup>1</sup> OPHN-dG | 1-phenanthrenol-dG      |
| <sup>1</sup> PHN-dG  | 1-phenanthrenyl-dG      |
| <sup>2</sup> ANP-dG  | 2-aminonaphthyl-dG      |
| 2NP                  | 2-naphthylamine         |
| <sup>2</sup> NP-dG   | 2-naphthyl-dG           |
| <sup>2</sup> ONP-dG  | 2-naphthol-dG           |
| <sup>3</sup> ABP-dG  | 3-aminobiphenyl-dG      |
| <sup>3</sup> APHN-dG | 3-aminophenanthrenyl-dG |
| <sup>3</sup> BP-dG   | 3-biphenyl-dG           |
| 3NBA                 | 3-nitrobenzanthrone     |
| <sup>3</sup> OPHN-dG | 3-phenanthrenol-dG      |
| <sup>3</sup> PHN-dG  | 3-phenanthrenyl-dG      |
| <sup>4</sup> ABP-dG  | 4-aminobiphenyl-dG      |
| 8OG                  | 8-oxoguanine            |
| A                    | adenine                 |
| AA                   | aromatic amines         |
| AAF                  | N2-acetylaminofluorene  |

|                   |                                             |
|-------------------|---------------------------------------------|
| <sup>AAF</sup> dG | 2-acetylaminofluorenyl-guanosine            |
| IARC              | International Agency for Research on Cancer |
| ABP               | aminobiphenyl                               |
| Adducts           | addition products                           |
| AF                | 2-aminofluorene                             |
| <sup>AF</sup> dG  | 2-aminofluorenyl-dG                         |
| AN                | aniline                                     |
| <sup>AN</sup> dG  | aniline-dG                                  |
| AP                | 1-aminopyrene                               |
| <sup>AP</sup> dG  | 1-aminopyrenyl-dG                           |
| AS                | AlphaSpace                                  |
| B                 | B-DNA type                                  |
| BHD               | β-hairpin domain                            |
| BHD1              | β-hairpin domain 1                          |
| BHD2              | β-hairpin domain 2                          |
| BHD3              | β-hairpin domain 3                          |
| BSE               | block standard error                        |
| C                 | cytosine                                    |
| CD                | circular dichroism                          |
| CNCNNA            | 2,6-dicyano-4-nitroaniline                  |
| CNNA              | 2-cyano-4-nitroaniline                      |
| COM               | center of mass                              |
| CPD               | pyrimidine dimer                            |
| D-A               | donor-acceptor                              |

|                     |                                                    |
|---------------------|----------------------------------------------------|
| DFT                 | density functional theory                          |
| dG                  | 2'-deoxyguanosine                                  |
| dR                  | 2'-deoxyribose                                     |
| <sup>FAAF</sup> dG  | fluorine-labeled <sup>AAF</sup> dG                 |
| FBP-dG              | C8-fluoro-biphenyl-dG                              |
| FP                  | fully paired                                       |
| FPh-dG              | C8-fluoro-phenyl-dG                                |
| G                   | guanine                                            |
| GAFF                | general Amber force field                          |
| GG-NER              | global genome NER                                  |
| IARC                | International Agency for Research on Cancer        |
| IQ                  | 2-amino-3-methylimidazo[4,5-f]quinolone            |
| LIE                 | linear interaction energy                          |
| LSIE                | lesion site interaction energy                     |
| LSIE <sup>avg</sup> | average LSIE                                       |
| MD                  | molecular dynamics                                 |
| MF                  | mutation frequency                                 |
| MM/PBSA             | molecular mechanics Poisson Boltzmann surface area |
| NA                  | nitroaniline                                       |
| NCA                 | nitrogen containing aromatic                       |
| NER                 | nucleotide excision repair                         |
| NMR                 | nuclear magnetic resonance                         |
| NP                  | 1-nitropyrene                                      |

|                               |                                                      |
|-------------------------------|------------------------------------------------------|
| OBP <sub>dG</sub>             | 3-biphenol-dG                                        |
| O <sup>Ph</sup> <sub>dG</sub> | O-linked-phenolic-dG                                 |
| O <sup>Py</sup> <sub>dG</sub> | 1-pyrenol-dG                                         |
| OTA                           | ochratoxin A                                         |
| PAH                           | polycyclic aromatic hydrocarbon                      |
| PCP <sub>dG</sub>             | pentachlorophenyl-dG                                 |
| PDB                           | protein data bank                                    |
| PES                           | potential energy surface                             |
| Ph <sub>dG</sub>              | phenyl-dG                                            |
| Ph-dG                         | C8-phenyl-dG                                         |
| PhIP                          | 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine      |
| PMF                           | potential of mean force                              |
| Py <sub>dG</sub>              | 1-pyrenyl-dG                                         |
| Py-dG                         | C8-pyrene-dG                                         |
| RMSD                          | root-mean-square deviation                           |
| RT                            | residence time                                       |
| S                             | base-displaced stacked                               |
| SASA                          | solvent accessible surface area                      |
| SASA <sup>avg</sup>           | average SASA                                         |
| SMI                           | slippage mutagenic intermediate                      |
| SSBs                          | single-strand breaks                                 |
| T                             | thymine                                              |
| TFIIH                         | 10-subunit general transcription factor II H complex |
| TGD                           | transglutaminase domain                              |

|                                    |                                                                                      |
|------------------------------------|--------------------------------------------------------------------------------------|
| $T_m$                              | thermal melting temperature.                                                         |
| $T_o\text{dG}$                     | tolyl-dG                                                                             |
| W                                  | wedge                                                                                |
| WC                                 | Watson-Crick                                                                         |
| XP                                 | xeroderma pigmentosum                                                                |
| XPC                                | XPC-RAD23B-CETN2 complex                                                             |
| $\Delta(\text{LSIE}^{\text{avg}})$ | $\text{LSIE}^{\text{avg}}$ of the same adduct site in the absence of a second lesion |
| $\Delta(\text{SASA}^{\text{avg}})$ | $\text{SASA}^{\text{avg}}$ of the same adduct site in the absence of a second lesion |
| $\theta^{\text{avg}}$              | average $\theta$                                                                     |
| $\phi^{\text{avg}}$                | average $\phi$                                                                       |
| $\chi^{\text{avg}}$                | average $\chi$                                                                       |
| $\xi^{\text{avg}}$                 | average $\xi$                                                                        |



# Chapter 1: Introduction<sup>a</sup>

## 1.1 General Overview

Deoxyribonucleic acid (DNA) plays a critical role in cells, containing genetic information that is passed from one generation to the next. Unfortunately, the fidelity of the genetic code is frequently compromised as DNA is regularly damaged by various sources, including UV light, industrial chemicals and pollutants.<sup>1, 2</sup> One specific type of damage is the chemical formation of bulky nucleobase addition products (adducts).<sup>2</sup> These DNA adducts are known to change the local DNA structure, impacting the interactions between DNA and various DNA-binding biomolecules.<sup>3-5</sup> As a result, DNA damage modulates numerous biological processes including replication, transcription and DNA folding/packing. By disrupting these key processes, DNA damage has the potential to cause significant biological harm, including aging and cancer in humans.<sup>6</sup> Fortunately, numerous DNA repair pathways have evolved to minimize the consequences of DNA damage.<sup>7</sup> Through the use of computational methods, this thesis considers the factors that dictate adducted DNA structure and its interaction with critical cellular machinery, including DNA repair enzymes. The following sections will discuss the structure of DNA, the origin and structure of damaged DNA, bulky nucleobase adducts, the replication and repair of adducted DNA, and previous uses of computational chemistry to study DNA damage.

## 1.2 DNA Structure

DNA is a biopolymer consisting of nucleotide subunits. Each DNA nucleotide contains a phosphate group, a 2'-deoxyribose sugar and one of the four canonical nucleobases, namely adenine (A), guanine (G), thymine (T) and cytosine (C, Figure 1.1A). The

---

<sup>a</sup>The *Journal of American Chemical Society* Reference Style is used in this Chapter.

nucleobases can be classified as the two-ringed purines (A and G) and the one-ringed pyrimidines (T and C). Within cells, DNA naturally exists as a double-stranded duplex (Figure 1.1B) containing two antiparallel and complementary strands in which the nucleobases selectively hydrogen bond. Specifically, the bases form Watson-Crick (WC) hydrogen bonds, with A:T base pairs containing two strong hydrogen bonds and the G:C pairs forming three (Figure 1.2). In order for the WC faces of the nucleobases to interact with each other, the nucleotides adopt an *anti* glycosidic orientation ( $\chi$  ( $\angle(\text{O4}'\text{--C1}'\text{--N9--C4})$  for the purines and  $\angle(\text{O4}'\text{--C1}'\text{--N1--C2})$  for the pyrimidines) =  $180^\circ \pm 90^\circ$ ). B-DNA has two grooves in the backbone of the helix, with the major groove being the larger gap and the minor groove being the smaller gap (Figure 1.1B). Under cellular conditions, DNA duplexes adopt the B-form that is characterized by a  $36^\circ$  twist (Figure 1.3), a rise between the base pairs of 3.3 Å, and a C2'-endo sugar pucker.

Maintaining the structure of DNA is critical to preserve interactions with other biomolecules and promote various cellular processes. For example, deviations from the canonical structure could lead to improper folding of DNA around histone proteins and hinder the formation of chromosomes.<sup>8</sup> Modifications to the structure of DNA can also impact transcription, leading to mutations during protein synthesis.<sup>9</sup> Additionally, structural disruptions to the double helix can impede DNA replication, potentially resulting in severe cellular and biological consequences, including cancer and other severe illnesses.<sup>10</sup> However, as discussed in the next section, DNA can be damaged, which can cause significant changes to the helical structure.

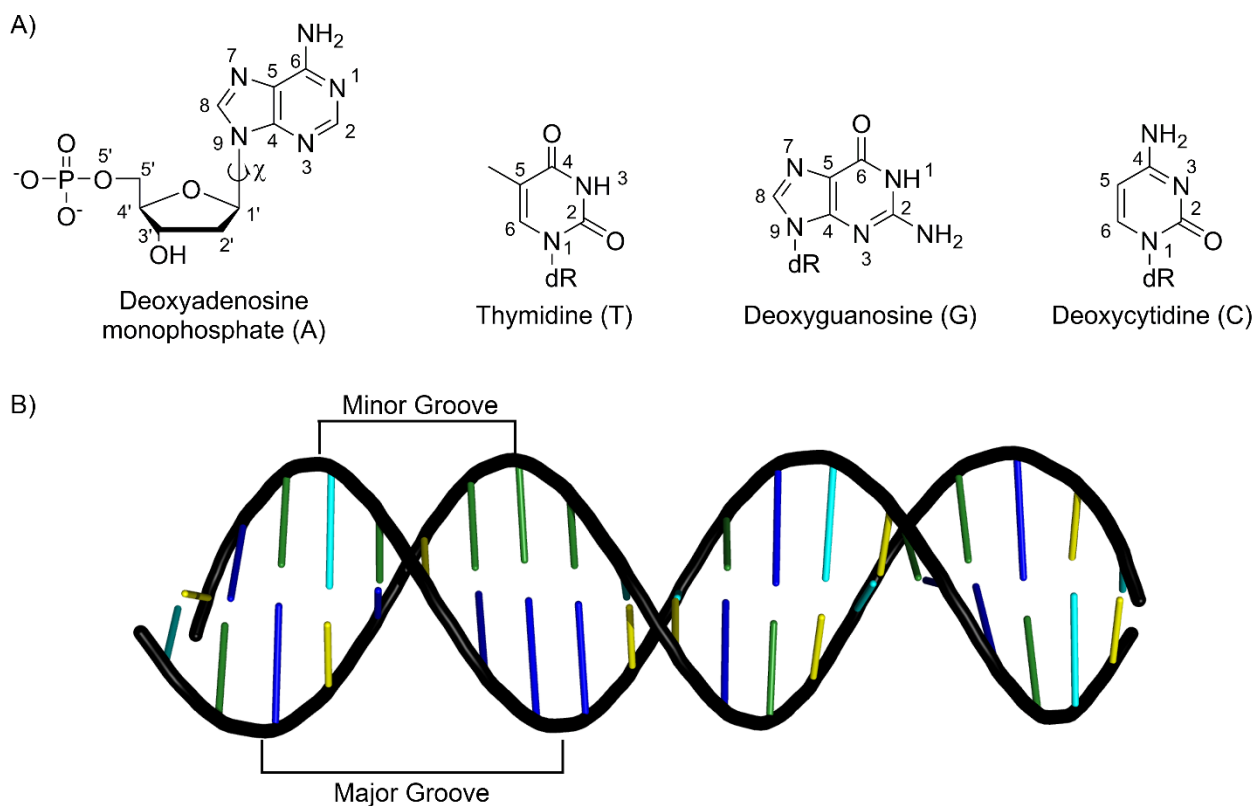


Figure 1.1: Structures of A) the four canonical nucleotides, including atom numbering and the  $\chi$  dihedral angle (defined as  $\angle(\text{O4}'\text{--C1}'\text{--N9--C4})$  for purines and  $\angle(\text{O4}'\text{--C1}'\text{--N1--C2})$  for pyrimidines) and B) a double-stranded B-DNA duplex, highlighting the major and minor grooves.

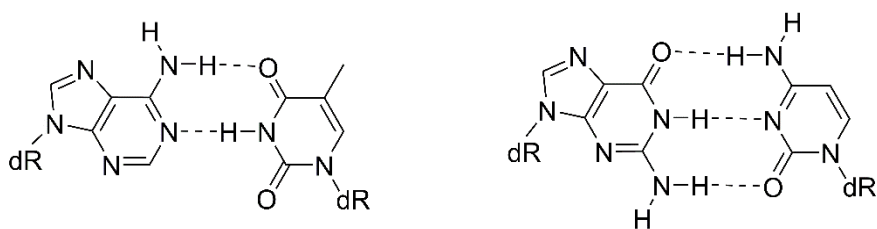


Figure 1.2: Structures of the WC A:T (left) and G:C (right) base pairs.

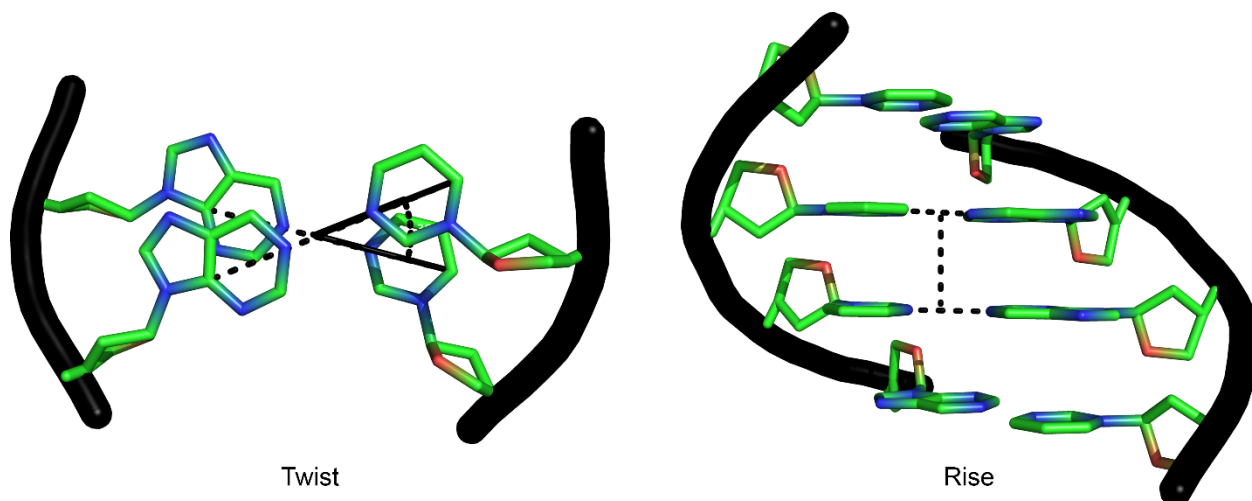


Figure 1.3: DNA structures extracted from duplexes, highlighting the twist and rise helical parameters.

### 1.3 DNA Damage

DNA damage is a common occurrence in cells, with estimates that damage arises up to 60,000 times per day for a human cell.<sup>11</sup> DNA damage can occur from numerous endogenous and exogenous sources, including pesticides, hormones, UV radiation, industrial agents, environmental pollutants and anticancer drugs.<sup>2</sup> The nature of the resulting damage is chemically variable, leading to products such as nucleobase oxidation (e.g., G to 8-oxoguanine),<sup>12</sup> alkylation (e.g., A to ethenoadenine),<sup>13</sup> and deamination (e.g., C to uracil),<sup>14</sup> DNA single and double strand breaks,<sup>15</sup> DNA crosslinks (e.g., cyclobutane pyrimidine dimer),<sup>16</sup> and bulky nucleobase or phosphate adducts (e.g., G to pentachlorophenol-dG, Figure 1.4).<sup>2, 17</sup> Since the structure of the damage widely varies, the cellular and biological outcomes range considerably. For example, DNA damage can have no effect on an individual's health or lead to cell death, cancer, or neurodegenerative diseases.<sup>2, 4, 6</sup> To help mitigate the negative effects of DNA damage, cells have evolved DNA repair pathways and alternative replication processes, which increase organism

survivability.<sup>7</sup> This thesis focuses on the helical structural impact and repair of DNA nucleobase adducts, which are further detailed in the next subsection.

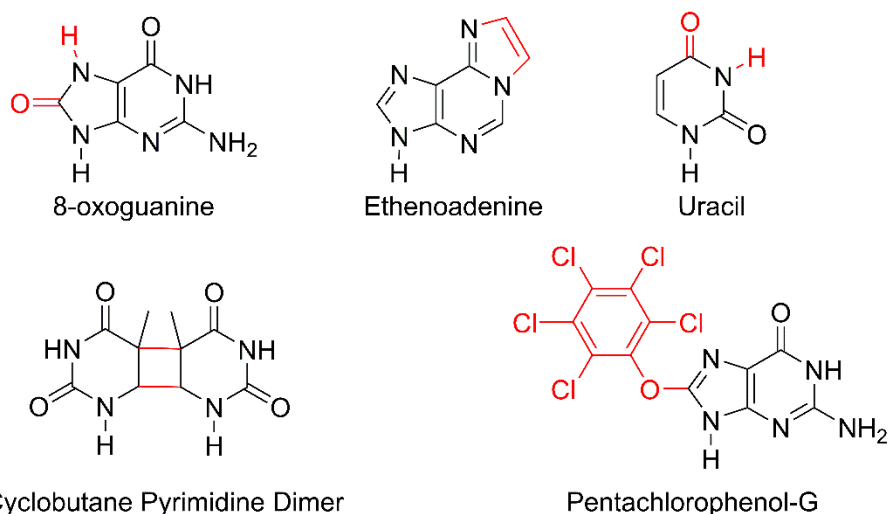


Figure 1.4: Structures of sample DNA damage products, with the damaged portion shown in red.

### 1.3.1 DNA Nucleobase Adducts

Bulky nucleobase adducts are primarily formed upon covalent binding of chemically reactive species to a DNA nitrogenous nucleobase.<sup>2, 18</sup> The origins of the corresponding damaging agents vary, with some being naturally found in cells (e.g., hormones) and others arising from the environment (e.g., cigarette smoke).<sup>2, 18</sup> For example, one class of known DNA damaging agents that are highly toxic to mammals is aromatic amines (AAs), which are chemically activated in cells to form highly unstable nitrenium complexes that in turn bind to DNA.<sup>19, 20</sup> Depending on the chemical agent, the structure of the resulting bulky group can vary, including its size, shape, charge, presence/absence of heteroatoms, and type of linker connecting the moiety to the damaged base (Figure 1.5). Further adding to the complexity of the DNA adduct chemical space,

there are numerous sites that can be damaged on each of the four canonical nucleobases, including C8-dA, N7-dA, N6-dA, C8-dG, O2-dT, O4-dT, O2-dC, N7-dG, O6-dG and N2-dG (Figure 1.1A).<sup>2</sup> Although adducts can arise at different nucleotide sites, the formation of adducts at the C8 site of G is among the most prevalent.<sup>2, 18, 21</sup> C8-dG adducts are known to form from various chemical groups including AAs (tobacco smoke, charred red meat and deisel exhaust), polyaromatic hydrocarbons (PAHs, automobile exhaust), aryl hydrazines (edible mushrooms) and phenoxyis (pesticides).<sup>2, 18</sup> This thesis focuses on a structurally diverse range of C8-dG adducts.

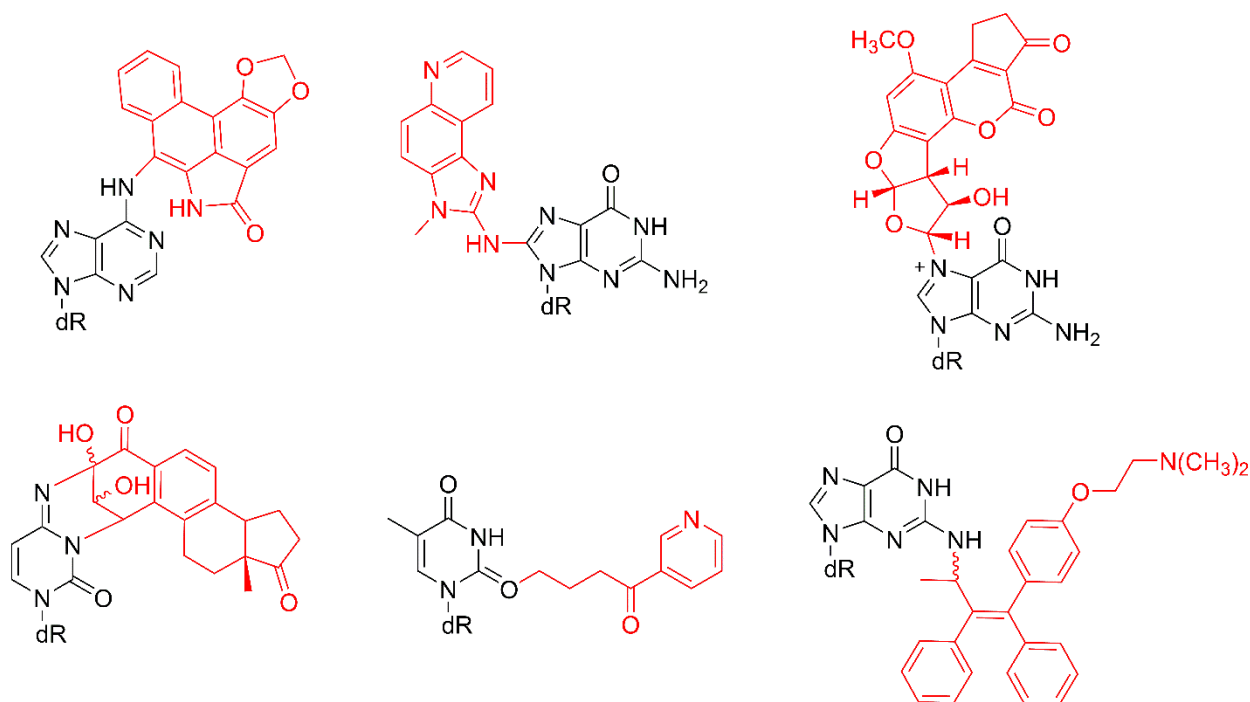


Figure 1.5: Structures of sample DNA adducts, highlighting the diversity in chemical structures. The bulky group is shown in red.

As with adducts in general, the chemical composition and mutagenic outcomes of C8-dG adducts can vary significantly. For example, aniline, which forms <sup>AN</sup>dG adducts, is considered nonmutagenic,<sup>22</sup> whereas both pentachlorophenol and 4-aminobiphenyl (ABP),

the precursors to the <sup>PCP</sup>dG and <sup>ABP</sup>dG adducts, are considered carcinogenic to humans (Group 1) by the International Agency for Research on Cancer (IARC).<sup>23, 24</sup> Thus, the chemical makeup of the adduct dictates the downstream biological consequences, likely because of changes in the structure of the duplex and its interactions with DNA repair and replication enzymes.

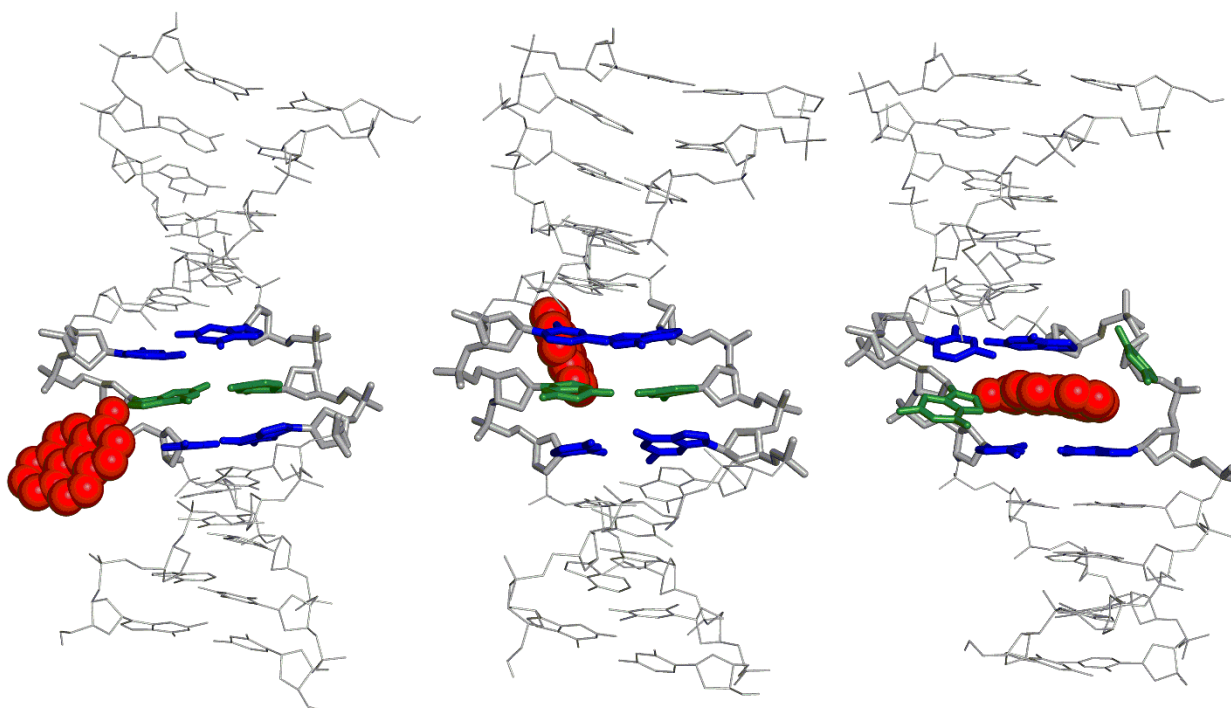


Figure 1.6: Example structures of C8-dG adducted DNA, in the major groove (B, left), minor groove wedge (W, center) and intercalated stacked (S, right) conformations. The damaged and pairing bases are shown in green, flanking bases in blue and bulky group in red.

Despite the range in the chemical composition of known C8-dG adducts, three broad structural categories have been identified for C8-dG adducted DNA (Figure 1.6).<sup>2, 18, 25</sup> Specifically, the bulky group can be positioned in the larger major groove, the smaller minor groove, or the interhelical space. The major groove (B) conformation of C8-dG

adducted DNA has the bulky group in the major groove, with the damaged nucleotide adopting an *anti* orientation, which allows the canonical WC face of the damaged base to hydrogen bond with the pairing strand (Figure 1.7). In contrast, for the wedge (W) conformation, the Hoogsteen hydrogen-bonding face is exposed to the pairing strand when the bulky moiety is positioned in the minor groove of the helix due to the *syn* orientation of the adduct (Figure 1.7). Finally, the *syn* orientation is also adopted at the lesion site when the damaged DNA is in an intercalated stacked (S) conformation, which positions the bulky group in the interhelical space and prevents hydrogen bonding between the adducted G and pairing C. Notably, even within each of these conformations, the exact structure of adducted DNA can vary in the degree of helical distortion, the hydrogen bonds formed at the lesion site and the bulky group orientation with respect to the damaged base.

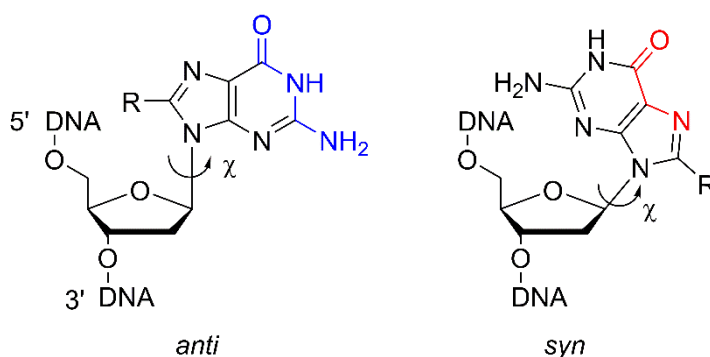


Figure 1.7: Structures of a generic C8-dG adduct (R = bulky group) in the *anti* (left) and *syn* (right) orientations, which expose the WC (blue) and Hoogsteen (red) face of the damaged G to the opposing strand, respectively.

A significant volume of research has been completed to understand the structure of adducted DNA, particularly through the use of computer modelling (primarily molecular dynamics (MD) simulations)<sup>2, 26-29</sup> and advanced nuclear magnetic resonance (NMR) techniques (such as <sup>19</sup>F-NMR using fluorine-labeled adducts).<sup>30-34</sup> As expected, the



chemical makeup of the bulky moiety has been shown to have a large impact on the conformational heterogeneity of the damaged DNA.<sup>2, 18, 25</sup> Multiple factors are known to affect the exact conformational structure of adducted DNA, including the bulky group size, charge, and presence/absence of heteroatoms, and sequence context.<sup>2, 18, 25-27, 30, 32-35</sup> Consequently, DNA adducts have a range of mutagenicities, with some precursors being carcinogenic to humans, while others are classified as nonmutagenic.<sup>23, 24</sup> Therefore, a significant portion of this thesis focuses on the how the C8-dG adduct chemical structure impacts the conformational heterogeneity of adducted DNA and the resulting cellular consequences. Since the long-term mutagenicity of an adduct is dependent on the ability of the cell to repair the damage and to perform error-free bypass of the lesion should it be replicated before being repaired, general concepts related to the replication and repair of DNA adducts are discussed in the following sections.

#### **1.4 Replication of DNA Adducts**

Replication of DNA occurs during cell division and passes genetic information from one cell generation to the next. For canonical DNA, this is carried out by highly accurate replicative polymerases, with error rates of  $< 10^{-9}$  per base pair.<sup>10</sup> In order to maintain DNA, replicative polymerases rely on the interactions between the DNA template strand and the incoming 2'-deoxynucleotide triphosphate (dNTP).<sup>4, 10</sup> These standard polymerases have evolved to form tightly bound complexes at the active site when the nucleotide is correctly paired, affording small misincorporation rates. However, the active site of a standard replicative polymerase is often significantly disrupted when bound to adducted DNA. As a result, replication stalls and the translesion synthesis (TLS) response is initiated. This process includes the recruitment of specialized TLS polymerases that have more flexible

active sites, which allows bypass of noncanonical DNA, but at a much greater error rate ( $10^{-3}$  to  $10^{-4}$  per base pair) than standard polymerases, even for undamaged regions.<sup>4, 10</sup> By modifying the interactions between the template strand and the incoming dNTP to accommodate the damage (e.g., by changing the *anti/syn* preference of the adducted nucleotide or repositioning the damaged base at the replication fork), mutations can occur during TLS. Depending on the adducted DNA and TLS polymerase, certain types of mutations occur with greater frequencies, with point mutations (i.e., the insertion of a noncomplementary nucleotide across from the damage site) and frameshift mutations (i.e., the insertion or deletion of bases into the growing strand) being the most common (Figure 1.8).<sup>2, 36, 37</sup> Notably, the formation of deletion mutations is believed to be correlated with the ability of adducted DNA to stabilize a Streisinger slippage mutagenic intermediate (SMI) complex in which the pairing strand is two bases shorter than the adducted strand, resulting in two unpaired bases at the lesion site.<sup>38</sup> Notably, this thesis investigates the structures of several SMI complexes to help rationalize differences in lesion toxicity. Additionally, the glycosylic orientation of the adduct can impact the replication outcomes, with the lesions that induce the canonical *anti* orientation less likely to cause replicative errors than the promutagenic *syn* conformation.<sup>22</sup> Thus, this thesis examines the *anti:syn* ratio of damaged nucleotides as an indicator for mutational frequency. Overall, prior to replication, damaged DNA would ideally be repaired through one of the pathways discussed in the following section.

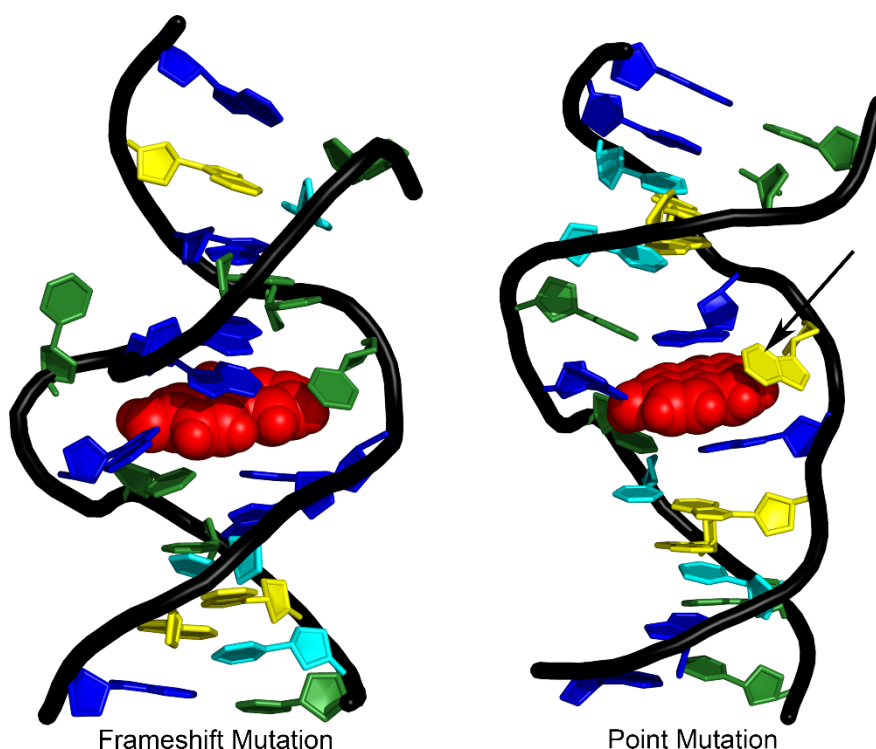


Figure 1.8: Crystal structures of C8-dG adducted DNA duplexes containing a deletion frameshift mutation around <sup>AF</sup>dG (left, PDB ID: 1AX6) or a C to A point mutation opposite <sup>AP</sup>dG (right, PDB ID: 1AXU; black arrow points to the mispaired base). G nucleosides are in blue, C nucleosides in green, A nucleosides in yellow and T nucleosides in cyan. The bulky group is shown in red.

### 1.5 Repair of DNA Adducts

Cells have evolved multiple repair pathways to revert damaged DNA to its canonical form. Each of these repair pathways specializes in the treatment of different lesions. For example, DNA alkylation can be repaired by the direct repair pathway,<sup>39</sup> oxidized nucleobases can be repaired through base excision repair (BER),<sup>40</sup> and bulky DNA adducts are removed via the complex nucleotide excision repair (NER) pathway.<sup>41, 42</sup> NER is especially interesting due to the wide variety of lesions recognized and excised by this process.<sup>41</sup> In addition to bulky DNA adducts, mismatched DNA nucleotides, DNA damage induced by UV light and DNA crosslinks are all recognized and cleaved by NER.<sup>41</sup> NER

is subdivided into two pathways, namely global genome NER (GG-NER) and transcription coupled NER (TC-NER).<sup>41, 42</sup> The difference between these pathways lies in the process of recognizing the damaged site.<sup>43</sup> Despite the functional diversity of GG-NER and TC-NER, both pathways share the same process for damage incision, repair and ligation.<sup>43</sup> As the names suggest, TC-NER is focused on repairing actively-expressed genes, whereas GG-NER is responsible for repairing both transcribed and nontranscribed regions of DNA. The GG-NER recognition process is thought to proceed through a “kinetic gating” mechanism in which the diffusion of the repair enzymes along DNA is faster than the binding of unmodified DNA.<sup>44</sup> Consequently, noncanonical DNA is more prone to complexation, allowing for the localized binding of repair proteins at damaged sites. There are several factors that are believed to impact the efficiency of protein binding and damage recognition, including changes in the local stacking and hydrogen-bonding interactions, increases in the helical dynamics and increases in the helical distortion at the lesion site.<sup>44-54</sup> Notably, the presence of DNA adducts is known to cause all of these structural changes.

After NER damage recognition, the damaged strand is cleaved by an endonuclease, which removes a short section (24–32 nucleotides) of single-stranded DNA containing the lesion.<sup>55</sup> Subsequently, the missing nucleotides are replaced by DNA polymerases. Finally, DNA ligases seal the nicks in the sugar-phosphate backbone on both sides of the newly inserted strand. Overall, the rate-limiting step in the NER pathway is the identification of the lesions. Therefore, the repairability of damaged DNA is largely dependent on the local structure at the lesion site. Part of this thesis investigates how the structure of adducted DNA impacts the NER recognition process and overall repairability of the adducted DNA, using the modelling techniques briefly described in the following section.

## 1.6 Computer Modelling of DNA Adducts

Computer modelling is a useful tool to provide valuable atomic level information about the structure of chemical systems. Computational techniques have become a common approach for studying biomolecules, including nucleic acids and proteins.<sup>56-58</sup> Previously, both quantum mechanical (QM) and molecular mechanical (MM) methods have been used to study adducted DNA.<sup>2, 26, 28, 29, 33, 35, 59-61</sup> Specifically, QM calculations have been used to understand the inherent structure of damaged DNA by modelling isolated nucleobases, nucleosides and nucleotides.<sup>2, 35, 59</sup> This work has shown the preferential orientations of the bulky group with respect to the nucleobase and the intrinsic *anti/syn* glycosidic preferences. MM methods have primarily been used to study the structure and stability of short oligomeric chains (both single stranded and double stranded) containing adducts, including time-dependent molecular dynamics (MD) techniques.<sup>2, 26, 28, 29, 33</sup> MD work has primarily been used to characterize the conformational space of adducted DNA duplexes, as well as to understand the interactions between adducted DNA and replicative<sup>60, 61</sup> or repair proteins.<sup>48, 49</sup> Such work has been essential for advancing our understanding of the impact of adduct structure on the mutagenic outcomes of DNA damage. My work will use these established computational methods to investigate the structure of a range of C8-dG adducted DNA strands, including in the context of NER machinery, and increase our knowledge of the factors that determine adducted DNA structure and mutagenicity.

## 1.7 Thesis Overview

This thesis uses computational chemistry to understand the structure and repair of adducted DNA. To start, Chapter 2 uses a multiscale approach to investigate the impact of bulky group size and shape on a systematic set of seven N-linked C8-dG adducts that

contain from one to four aryl rings. These adducts were chosen to identify the origin of the mutagenicity differences between the one-ringed <sup>AN</sup>dG adduct and the four-ringed <sup>AP</sup>dG adduct. Specifically, damaged nucleobase and nucleoside models were considered using QM methodology (density functional theory (DFT)) to understand the inherent structural preferences of isolated adducts. Furthermore, adducted DNA helical models were studied using all-atom MD simulations to probe the conformational preferences and helical features induced by the different adducts. This work highlights the effects of both lesion size and shape on the structure of damaged DNA. Next, Chapter 3 studies the mutagenic effect upon ring cyclization by examining two N-linked C8-dG cyano-aniline adducts using MD simulations. Notably, these azo dye-derived lesions are substituted versions of the single-ringed <sup>AN</sup>dG adduct studied in Chapter 2. This chapter models the conformational preferences of fully paired duplexes, as well as the SMI complexes that cause frameshift mutations. By studying the *ortho* and *para* cyano-substituted C8-dG N-linked aniline adducts, this work highlights the impact of functional group position on the structure of adducted DNA.

By investigating a set of 14 systematic lesions corresponding to the C-linked and O-linked versions of the seven adducts considered in Chapter 2, Chapter 4 further explores the impact of the lesion size and shape as well as newly considers the effect of the linkage type on the preferred adducted DNA structure. Following the approach taken in Chapter 2, nucleobase and nucleoside models were considered, along with fully paired DNA duplexes. Although this work shows that the linkage type and bulky group shape impact the adduct orientation within the duplex, the bulky group size is highlighted as the primary factor that determines the conformational preferences. Chapter 5 uses MD simulations and <sup>19</sup>F-NMR

to further probe the impact of linkage type and lesion size by examining the F-labelled phenyl and biphenyl C-linked C8-dG adducts. Although these lesions are equivalent to the N-linked biphenyl adducts studied in Chapter 2 and experimentally investigated by other groups,<sup>19, 32</sup> the C-linked analogues were studied as potential fluorescent probes of DNA damage. This work highlights the significant effect of adding a second aromatic ring through (4-) *para* substitution of the aryl group of C-linked adducts on the structures of fully paired and SMI duplexes.

The vast majority of work on DNA adducts, including the first portion of this thesis, has studied DNA containing a single local lesion (monoadducted DNA). However, evidence suggests that more than one adduct may be formed within short sequences of DNA (e.g., diadducted DNA) under certain circumstances.<sup>37, 62-64</sup> Chapters 6 and 7 add the number of lesions to the list of factors that affect the preferred helical conformation by considering the structural and associated biological consequences of three N-linked C8-dG adducts within a short DNA sequence. Specifically, Chapter 6 considers monoadducted and diadducted DNA duplexes containing the one-ringed and four-ringed <sup>AN</sup>dG and <sup>AP</sup>dG adducts considered in Chapter 2 due to their differing mutagenicities,<sup>24, 65</sup> while Chapter 7 studies the model <sup>ABP</sup>dG adduct that is a known indicator of bladder cancer and displays unique conformational heterogeneity in select monoadducted DNA sequences.<sup>25, 30, 32</sup> These chapters represent the first in-depth structural studies of diadducted DNA and reveal important lesion separation effects. Notably, it is uncovered that a second damaging event may increase or decrease lesion mutagenicity as well as the severity of the resulting biological consequences depending on the adduct sizes and separation. Lastly, Chapter 8 uses comprehensive advanced MD sampling methods to probe the GG-NER damage

recognition process for the N-linked C8-dG<sup>AA</sup>FdG adduct, which was chosen due to the availability of repair data for both monoadducted and diadducted duplexes.<sup>66</sup> Beyond providing the first structural look at the repair of<sup>AA</sup>FdG adducted DNA, this study is the first to compare the binding pathways for monoadducted DNA in the major groove and intercalated conformations, and the first to investigate the binding pathway of diadducted DNA. Overall, this chapter supports the previously proposed “kinetic gating” mechanism<sup>44</sup> and expands our understanding of the repair of bulky DNA adducts. Combined, my work unveils novel trends in how the adduct chemical composition dictates damaged DNA structure and repairability as well as the resulting toxicity of C8-dG nucleobase adducts.

## 1.8 References

1. Ignatov, A. V.; Bondarenko, K. A.; Makarova, A. V., Non-Bulky Lesions in Human DNA: The Ways of Formation, Repair, and Replication. *Acta Naturae* **2017**, 9 (3), 12-26.
2. Wilson, K. A.; Kung, R. W.; Wetmore, S. D., Chapter Seven - Toxicology of DNA Adducts Formed Upon Human Exposure to Carcinogens: Insights Gained from Molecular Modeling. In *Advances in Molecular Toxicology*, Fishbein, J. C.; Heilman, J. M., Eds. Elsevier: 2016; Vol. 10, pp 293-360.
3. Dimitri, A.; Burns, J. A.; Broyde, S.; Scicchitano, D. A., Transcription Elongation Past O6-Methylguanine by Human Rna Polymerase Ii and Bacteriophage T7 Rna Polymerase. *Nucleic Acids Res.* **2008**, 36 (20), 6459-6471.
4. Broyde, S.; Wang, L.; Rechkoblit, O.; Geacintov, N. E.; Patel, D. J., Lesion Processing: High-Fidelity Versus Lesion-Bypass DNA Polymerases. *Trends in Biochemical Sciences* **2008**, 33 (5), 209-219.
5. Cai, Y.; Wang, L.; Ding, S.; Schwaid, A.; Geacintov, N. E.; Broyde, S., A Bulky DNA Lesion Derived from a Highly Potent Polycyclic Aromatic Tumorigen Stabilizes Nucleosome Core Particle Structure. *Biochemistry* **2010**, 49 (46), 9943-9945.
6. Hoeijmakers, J. H. J., DNA Damage, Aging, and Cancer. *New England Journal of Medicine* **2009**, 361 (15), 1475-1485.
7. Basu, A.; Broyde, S.; Iwai, S.; Kisker, C., DNA Damage, Mutagenesis, and DNA Repair. *J. Nucleic Acids* **2010**, 182894, 1 pp.



8. Mann, D. B.; Springer, D. L.; Smerdon, M. J., DNA Damage Can Alter the Stability of Nucleosomes: Effects Are Dependent on Damage Type. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, 94 (6), 2215-2220.
9. Xu, L.; Wang, W.; Chong, J.; Shin, J. H.; Xu, J.; Wang, D., Rna Polymerase Ii Transcriptional Fidelity Control and Its Functional Interplay with DNA Modifications. *Critical reviews in biochemistry and molecular biology* **2015**, 50 (6), 503-519.
10. Sale, J. E.; Lehmann, A. R.; Woodgate, R., Y-Family DNA Polymerases and Their Role in Tolerance of Cellular DNA Damage. *Nature Reviews Molecular Cell Biology* **2012**, 13 (3), 141-152.
11. Carol Bernstein; Anil R. Prasad; Nfonsam, V.; Bernstein, H., DNA Damage, DNA Repair and Cancer. In *New Research Directions in DNA Repair*, Chen, C., Ed. IntechOpen: 2013.
12. Rechko, O.; Malinina, L.; Cheng, Y.; Geacintov, N. E.; Broyde, S.; Patel, D. J., Impact of Conformational Heterogeneity of Oxog Lesions and Their Pairing Partners on Bypass Fidelity by Y Family Polymerases. *Structure (Cambridge, MA, U. S.)* **2009**, 17 (5), 725-736.
13. Kielhorn, J.; Melber, C.; Wahnschaffe, U.; Aitio, A.; Mangelsdorf, I., Vinyl Chloride: Still a Cause for Concern. *Environ. Health Perspect.* **2000**, 108 (7), 579-588.
14. Yonekura, S. I.; Nakamura, N.; Yonei, S.; Zhang-Akiyama, Q. M., Generation, Biological Consequences and Repair Mechanisms of Cytosine Deamination in DNA. *Journal of Radiation Research* **2009**, 50 (1), 19-26.
15. Jakobsen, K. P.; Andersen, A. H.; Bjergbaek, L., Abortive Activity of Topoisomerase I: A Challenge for Genome Integrity? *Curr. Genet.* **2019**, 65 (5), 1141-1144.
16. Rebillard, A.; Lagadic-Gossman, D.; Dimanche-Boitrel, M. T., Cisplatin Cytotoxicity: DNA and Plasma Membrane Targets. *Curr. Med. Chem.* **2008**, 15 (26), 2656-2663.
17. Ma, B.; Zarth, A. T.; Carlson, E. S.; Villalta, P. W.; Upadhyaya, P.; Stepanov, I.; Hecht, S. S., Identification of More Than 100 Structurally Unique DNA-Phosphate Adducts Formed During Rat Lung Carcinogenesis by the Tobacco-Specific Nitrosamine 4-(Methylnitrosamino)-1-(3-Pyridyl)-1-Butanone. *Carcinogenesis* **2018**, 39 (2), 232-241.
18. Millen, A. L.; Sharma, P.; Wetmore, S. D., C8-Linked Bulky Guanosine DNA Adducts: Experimental and Computational Insights into Adduct Conformational Preferences and Resulting Mutagenicity. *Future Med. Chem.* **2012**, 4 (15), 1981-2007.
19. Wang, S.; Hanna, D.; Sugamori, K. S.; Grant, D. M., Primary Aromatic Amines and Cancer: Novel Mechanistic Insights Using 4-Aminobiphenyl as a Model Carcinogen. *Pharmacol. Ther.* **2019**, 200, 179-189.

20. Humphreys, W. G.; Kadlubar, F. F.; Guengerich, F. P., Mechanism of C8 Alkylation of Guanine Residues by Activated Arylamines: Evidence for Initial Adduct Formation at the N7 Position. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89* (17), 8278-8282.
21. Hwa Yun, B.; Guo, J.; Bellamri, M.; Turesky, R. J., DNA Adducts: Formation, Biological Effects, and New Biospecimens for Mass Spectrometric Measurements in Humans. *Mass Spectrom. Rev.* **2020**, *39* (1-2), 55-82.
22. Shapiro, R.; Ellis, S.; Hingerty, B. E.; Broyde, S., Effect of Ring Size on Conformations of Aromatic Amine-DNA Adducts: The Aniline-C8 Guanine Adduct Resides in the B-DNA Major Groove. *Chem. Res. Toxicol.* **1998**, *11* (4), 335-341.
23. Baan, R.; Grosse, Y.; Straif, K.; Secretan, B.; El Ghissassi, F.; Bouvard, V.; Benbrahim-Tallaa, L.; Guha, N.; Freeman, C.; Galichet, L.; Coglianò, V., A Review of Human Carcinogens; Part F: Chemical Agents and Related Occupations. *Lancet Oncol.* **2009**, *10* (12), 1143-1144.
24. The International Agency for Research on Cancer, *Overall Evaluations of Carcinogenicity. An Updating of Iarc Monographs Volumes 1 to 42*. IARC: 1987.
25. Cho, B. P., Dynamic Conformational Heterogeneities of Carcinogen-DNA Adducts and Their Mutagenic Relevance. *Journal of Environmental Science and Health Part C Environmental Carcinogenesis & Ecotoxicology Reviews* **2004**, *22* (2), 57-90.
26. Broyde, S.; Hingerty, B. E.; Shapiro, R.; Norman, D., Unusual Hydrogen Bonding Patterns in 2-Aminofluorene (Af) and 2-Acetylaminofluorene (Aaf) Modified DNA. *Environ. Sci. Res.* **1990**, *40* (Nitroarenes), 113-23.
27. Cai, Y.; Kropachev, K.; Xu, R.; Tang, Y.; Kolbanovskii, M.; Kolbanovskii, A.; Amin, S.; Patel, D. J.; Broyde, S.; Geacintov, N. E., Distant Neighbor Base Sequence Context Effects in Human Nucleotide Excision Repair of a Benzo[a]Pyrene-Derived DNA Lesion. *J. Mol. Biol.* **2010**, *399* (3), 397-409.
28. Yang, S.; Diem, M.; Liu, J. D. H.; Wesseling, S.; Vervoort, J.; Oostenbrink, C.; Rietjens, I., Cellular Levels and Molecular Dynamics Simulations of Estragole DNA Adducts Point at Inefficient Repair Resulting from Limited Distortion of the Double-Stranded DNA Helix. *Arch. Toxicol.* **2020**, *94* (4), 1349-1365.
29. Yang, S.; Liu, J. D. H.; Diem, M.; Wesseling, S.; Vervoort, J.; Oostenbrink, C.; Rietjens, I., Molecular Dynamics and in Vitro Quantification of Safrole DNA Adducts Reveal DNA Adduct Persistence Due to Limited DNA Distortion Resulting in Inefficient Repair. *Chem. Res. Toxicol.* **2020**, *33* (9), 2298-2309.
30. Cho, B. P.; Beland, F. A.; Marques, M. M., NMR Structural Studies of a 15-mer DNA Sequence from a ras Protooncogene, Modified at the First Base of Codon 61 with the Carcinogen 4-Aminobiphenyl. *Biochemistry* **1992**, *31* (40), 9587-602.

31. Jain, N.; Li, Y.; Zhang, L.; Meneni, S. R.; Cho, B. P., Probing the Sequence Effects on NarI-Induced -2 Frameshift Mutagenesis by Dynamic <sup>19</sup>F NMR, Uv, and Cd Spectroscopy. *Biochemistry* **2007**, *46* (46), 13310-13321.
32. Jain, V.; Vaidyanathan, V. G.; Patnaik, S.; Gopal, S.; Cho, B. P., Conformational Insights into the Lesion and Sequence Effects for Arylamine-Induced Translesion DNA Synthesis: <sup>19</sup>F NMR, Surface Plasmon Resonance, and Primer Kinetic Studies. *Biochemistry* **2014**, *53* (24), 4059-4071.
33. Norman, D.; Abuaf, P.; Hingerty, B. E.; Live, D.; Grunberger, D.; Broyde, S.; Patel, D. J., NMR and Computational Characterization of the N-(Deoxyguanosin-8-yl)Aminofluorene Adduct [(Af)G] Opposite Adenosine in DNA: (Af)G[Syn]-a[Anti] Pair Formation and Its Ph Dependence. *Biochemistry* **1989**, *28* (18), 7462-7476.
34. Zhang, N.; Ding, S.; Kolbanovskiy, A.; Shastry, A.; Kuzmin, V. A.; Bolton, J. L.; Patel, D. J.; Broyde, S.; Geacintov, N. E., NMR and Computational Studies of Stereoisomeric Equine Estrogen-Derived DNA Cytidine Adducts in Oligonucleotide Duplexes: Opposite Orientations of Diastereomeric Forms. *Biochemistry* **2009**, *48* (30), 7098-7109.
35. Wilson, K. A.; Wetmore, S. D., Complex Conformational Heterogeneity of the Highly Flexible O6-Benzyl-Guanine DNA Adduct. *Chem. Res. Toxicol.* **2014**, *27* (7), 1310-1325.
36. Solomon, M. S.; Morgenthaler, P.-M. L.; Turesky, R. J.; Essigmann, J. M., Mutational and DNA Binding Specificity of the Carcinogen 2-Amino-3,8-Dimethylimidazo[4,5-F]Quinoxaline. *J. Biol. Chem.* **1996**, *271* (31), 18368-18374.
37. Fuchs, R. P. P., DNA Binding Spectrum of the Carcinogen N-Acetoxy-N-2-Acetylaminofluorene Significantly Differs from the Mutation Spectrum. *J. Mol. Biol.* **1984**, *177* (1), 173-180.
38. Xu, L.; Cho, B. P., Conformational Insights into the Mechanism of Acetylaminofluorene-dG-Induced Frameshift Mutations in the NarI Mutational Hotspot. *Chem. Res. Toxicol.* **2016**, *29* (2), 213-226.
39. Sedgwick, B.; Bates, P. A.; Paik, J.; Jacobs, S. C.; Lindahl, T., Repair of Alkylated DNA: Recent Advances. *DNA Repair* **2007**, *6* (4), 429-442.
40. Jia, L.; Shafirovich, V.; Geacintov, N. E.; Broyde, S., Lesion Specificity in the Base Excision Repair Enzyme Hneil1: Modeling and Dynamics Studies. *Biochemistry* **2007**, *46* (18), 5305-5314.
41. Friedberg, E. C., Nucleotide Excision Repair: Biology. In *Encyclopedia of Biological Chemistry (Second Edition)*, Lennarz, W. J.; Lane, M. D., Eds. Academic Press: Waltham, 2013; pp 337-340.

42. Kong, M.; Liu, L.; Chen, X.; Driscoll, K. I.; Mao, P.; Bohm, S.; Kad, N. M.; Watkins, S. C.; Bernstein, K. A.; Wyrick, J. J.; Min, J. H.; Van Houten, B., Single-Molecule Imaging Reveals That Rad4 Employs a Dynamic DNA Damage Recognition Process. *Mol. Cell* **2016**, *64* (2), 376-387.
43. Fuss, J. O.; Cooper, P. K., DNA Repair: Dynamic Defenders against Cancer and Aging. *PLoS Biol.* **2006**, *4* (6), e203-e203.
44. Velmurugu, Y.; Chen, X.; Slogoff Sevilla, P.; Min, J.-H.; Ansari, A., Twist-Open Mechanism of DNA Damage Recognition by the Rad4/XPC Nucleotide Excision Repair Complex. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (16), E2296.
45. Min, J.-H.; Pavletich, N. P., Recognition of DNA Damage by the Rad4 Nucleotide Excision Repair Protein. *Nature* **2007**, *449* (7162), 570-575.
46. Cai, Y.; Patel, D. J.; Geacintov, N. E.; Broyde, S., Differential Nucleotide Excision Repair Susceptibility of Bulky DNA Adducts in Different Sequence Contexts: Hierarchies of Recognition Signals. *J. Mol. Biol.* **2009**, *385* (1), 30-44.
47. Hess, M. T.; Schwitter, U.; Petretta, M.; Giese, B.; Naegeli, H., Bipartite Substrate Discrimination by Human Nucleotide Excision Repair. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (13), 6664.
48. Mu, H.; Zhang, Y.; Geacintov, N. E.; Broyde, S., Lesion Sensing During Initial Binding by Yeast XPC/Rad4: Toward Predicting Resistance to Nucleotide Excision Repair. *Chem. Res. Toxicol.* **2018**, *31* (11), 1260-1268.
49. Mu, H.; Geacintov, N. E.; Min, J.-H.; Zhang, Y.; Broyde, S., Nucleotide Excision Repair Lesion-Recognition Protein Rad4 Captures a Pre-Flipped Partner Base in a Benzo[a]Pyrene-Derived DNA Lesion: How Structure Impacts the Binding Pathway. *Chem. Res. Toxicol.* **2017**, *30* (6), 1344-1354.
50. Reeves, D. A.; Mu, H.; Kropachev, K.; Cai, Y.; Ding, S.; Kolbanovskiy, A.; Kolbanovskiy, M.; Chen, Y.; Krzeminski, J.; Amin, S.; Patel, D. J.; Broyde, S.; Geacintov, N. E., Resistance of Bulky DNA Lesions to Nucleotide Excision Repair Can Result from Extensive Aromatic Lesion-Base Stacking Interactions. *Nucleic Acids Res.* **2011**, *39* (20), 8752-8764.
51. Yang, W., Poor Base Stacking at DNA Lesions May Initiate Recognition by Many Repair Proteins. *DNA Repair* **2006**, *5* (6), 654-666.
52. Kropachev, K.; Ding, S.; Terzidis, M. A.; Masi, A.; Liu, Z.; Cai, Y.; Kolbanovskiy, M.; Chatgililoglu, C.; Broyde, S.; Geacintov, N. E.; Shafirovich, V., Structural Basis for the Recognition of Diastereomeric 5',8-Cyclo-2'-Deoxypurine Lesions by the Human Nucleotide Excision Repair System. *Nucleic Acids Res.* **2014**, *42* (8), 5020-5032.

53. Maillard, O.; Camenisch, U.; Clement, F. C.; Blagoev, K. B.; Naegeli, H., DNA Repair Triggered by Sensors of Helical Dynamics. *Trends Biochem. Sci.* **2007**, 32 (11), 494-499.
54. Geacintov, N. E.; Broyde, S.; Buterin, T.; Naegeli, H.; Wu, M.; Yan, S.; Patel, D. J., Thermodynamic and Structural Factors in the Removal of Bulky DNA Adducts by the Nucleotide Excision Repair Machinery. *Biopolymers* **2002**, 65 (3), 202-210.
55. Aziz Sancar; Laura A. Lindsey-Boltz; Keziban Ünsal-Kaçmaz; Linn, S., Molecular Mechanisms of Mammalian DNA Repair and the DNA Damage Checkpoints. *Annual Review of Biochemistry* **2004**, 73 (1), 39-85.
56. De Jong, H., Modeling and Simulation of Genetic Regulatory Systems: A Literature Review. *J. Comput. Biol.* **2002**, 9 (1), 67-103.
57. Mackerell, A. D., Empirical Force Fields for Biological Macromolecules: Overview and Issues. *J. Comput. Chem.* **2004**, 25 (13), 1584-1604.
58. Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J., Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nature Reviews Drug Discovery* **2004**, 3 (11), 935-949.
59. Millen, A. L.; Kamenz, B. L.; Leavens, F. M. V.; Manderville, R. A.; Wetmore, S. D., Conformational Flexibility of C8-Phenoxyguanine Adducts in Deoxydinucleoside Monophosphates. *The Journal of Physical Chemistry B* **2011**, 115 (44), 12993-13002.
60. Zhang, L.; Shapiro, R.; Broyde, S., Molecular Dynamics of a Food Carcinogen–DNA Adduct in a Replicative DNA Polymerase Suggest Hindered Nucleotide Incorporation and Extension. *Chem. Res. Toxicol.* **2005**, 18 (9), 1347-1363.
61. Zhang, L.; Rechkoblit, O.; Wang, L.; Patel, D. J.; Shapiro, R.; Broyde, S., Mutagenic Nucleotide Incorporation and Hindered Translocation by a Food Carcinogen C8-dG Adduct in *Sulfolobus Solfataricus* P2 DNA Polymerase Iv (Dpo4): Modeling and Dynamics Studies. *Nucleic Acids Res.* **2006**, 34 (11), 3326-3337.
62. Jain, V.; Hilton, B.; Lin, B.; Jain, A.; MacKerell, A. D.; Zou, Y.; Cho, B. P., Structural and Thermodynamic Insight into *Escherichia Coli* UvrABC-Mediated Incision of Cluster Diacetylaminofluorene Adducts on the NarI Sequence. *Chem. Res. Toxicol.* **2013**, 26 (8), 1251-1262.
63. Koffel-Schwartz, N.; Verdier, J.-M.; Bichara, M.; Freund, A.-M.; Daune, M. P.; Fuchs, R. P. P., Carcinogen-Induced Mutation Spectrum in Wild-Type, UvrA and UvrC Strains of *Escherichia Coli*: Strain Specificity and Mutation-Prone Sequences. *J. Mol. Biol.* **1984**, 177 (1), 33-51.
64. Roy, A.; Fuchs, R. P. P., Mutational Spectrum Induced in *Saccharomyces Cerevisiae* by the Carcinogen N-2-Acetylaminofluorene. *Mol. Gen. Genet.* **1994**, 245 (1), 69-77.

65. The International Agency for Research on Cancer, *Diesel and Gasoline Engine Exhausts and Some Nitroarenes*. IARC: 2015.
66. Hilton, B.; Gopal, S.; Xu, L.; Mazumder, S.; Musich, P. R.; Cho, B. P.; Zou, Y., Dissociation Dynamics of XPC-Rad23b from Damaged DNA Is a Determining Factor of NER Efficiency. *PLoS One* **2016**, *11* (6), e0157784.

## Chapter 2: Effect of the Size and Shape of Nitrogen-Containing Aromatics on the Conformational Preferences of DNA Containing Damaged Guanine<sup>a,b,c</sup>

### 2.1 Introduction

Nitrogen containing aromatics (NCAs), such as aromatic amines, heterocyclic aromatic amines and nitroaromatic compounds, are well known to impact the environment and human health.<sup>1</sup> A number of potentially carcinogenic NCAs are proven chemical mutagens, and originate from a variety of sources such as food products,<sup>2</sup> cigarette smoke<sup>3</sup> and automobile waste.<sup>4</sup> In living cells, NCAs undergo bioactivation to yield reactive nitrenium ion intermediates that covalently bind to specific sites on the DNA nucleobases.<sup>5-</sup>  
<sup>6</sup> Of the four canonical nucleobases (adenine (A), cytosine (C), guanine (G) and thymine (T)), the electron-rich G is the most common site for nitrenium ion attachment.<sup>7</sup> An abundant damage product arises from the direct attachment of the activated nitrogen atom of the nitrenium ion to the C<sup>8</sup>-site of 2'-deoxyguanosine (dG).<sup>8-9</sup> The resulting N-linked C<sup>8</sup>-dG adducts (addition products) contain a C–N–C linkage between the nucleobase and the aromatic moiety (Figure 2.1).<sup>10</sup>

N-linked C<sup>8</sup>-dG adducts have been shown to originate from a variety of NCAs that exhibit a range in carcinogenic potential including, but not limited to, 2-aminofluorene (AF),<sup>11</sup> 2-acetylaminofluorene (AAF),<sup>12</sup> aniline (AN),<sup>13</sup> 4-aminobiphenyl (ABP),<sup>14</sup> 2-amino-3-methylimidazo[4,5-f]quinolone (IQ),<sup>15</sup> 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP),<sup>15</sup> 1-nitropyrene (1NP),<sup>16</sup> 2-naphthylamine (2NP)<sup>17</sup> and 3-

---

<sup>a</sup>The *Journal of American Chemical Society* Reference Style is used in this Chapter.

<sup>b</sup>Reprinted with permission from Kung, R. W.; Sharma, P.; Wetmore, S. D., Effect of Size and Shape of Nitrogen-Containing Aromatics on Conformational Preferences of DNA Containing Damaged Guanine. *J Chem Inf Model* **2018**, 58 (7), 1415-1425.

<sup>c</sup>R.W.K. performed DFT calculations and MD simulations, along with the corresponding analysis. Writing was done by R.W.K., P.S. and S.D.W.

nitrobenzanthrone (3NBA).<sup>18</sup> For example, <sup>AN</sup>dG (Figure 2.1A) is derived from AN,<sup>19</sup> which is used to manufacture dyes<sup>13</sup> and is considered to not be classifiable as a human carcinogen (group 3) by the International Agency for Research on Cancer (IARC). Alternatively, 2NAP, the precursor of a ring-opened C<sup>8</sup>-dG lesion,<sup>17</sup> was also previously used to manufacture azo dyes, but is a group 1 (human) carcinogen. On the other hand, 1NP is an abundant aromatic compound in the environment that has been associated with <sup>1AP</sup>dG (Figure 2.1A) formation,<sup>16</sup> and has been classified as a group 2B (possible human) carcinogen.

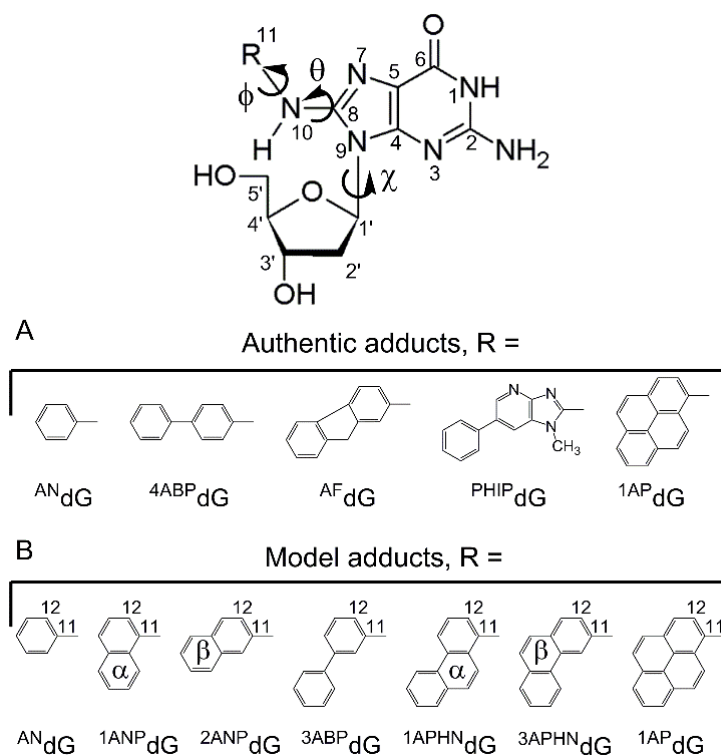


Figure 2.1: Structures and chemical numbering of N-linked DNA adducts at the C<sup>8</sup>-site of 2'-deoxyguanosine. (A) Biologically-relevant (authentic) adducts and (B) model adducts considered in the present work.



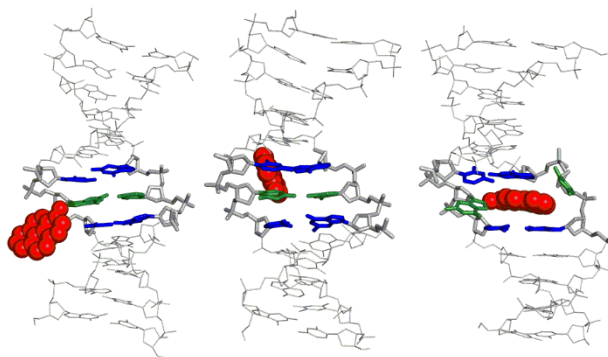


Figure 2.2: MD representative structures of the major groove (B, left), wedge (W, middle) and base-displaced stacked (S, right) conformations of DNA containing N-linked C<sup>8</sup>-dG adducts. <sup>1</sup>APdG adducted DNA is shown as an example.

Despite substantial progress in understanding the impact of NCAs on human health, significant gaps exist in our knowledge of their mechanism of toxicity in living cells. In attempts fill this void, the structures of the associated C<sup>8</sup>-dG adducted DNA have been under investigation for many years. The vast majority of these studies have focused on adducts originating from AAF and AF (<sup>AA</sup>FdG and <sup>A</sup>FdG, respectively), with NMR spectroscopy and computational chemistry being particularly pivotal to these investigations.<sup>10, 20-33</sup> Collectively, these studies have led to the identification of three general damaged DNA conformational themes. Specifically, the B-DNA type (B) conformation (Figure 2.2, left) locates the NCA-substituent in the DNA major groove and the damaged dG adopts the *anti* glycosidic orientation (described by the dihedral angle  $\chi = 180^\circ \pm 90^\circ$ , Figure 2.1). Alternatively, the damaged dG adopts a *syn* glycosidic orientation ( $\chi = 0^\circ \pm 90^\circ$ ) in the wedge (W) conformation (Figure 2.2, middle), which locates the NCA moiety in the DNA minor groove. Although the *syn* orientation is also adopted by the damaged dG in the base-displaced stacked (S) conformation (Figure 2.2, right), this conformation intercalates the aromatic moiety between the flanking base pairs, thereby

displacing the damaged guanine moiety and the opposing base toward the DNA major groove.

Since N-linked C<sup>8</sup>-dG adducts originate from NCAs that differ considerably in terms of their chemical structure, there is remarkable variation in the size and shape of the C<sup>8</sup>-moieties in the associated adducts (Figure 2.1A). Furthermore, the structure of the adducted DNA and the mutagenic potential associated with these adducts fluctuate significantly. For example, the <sup>4</sup>ABP-dG adduct is associated with both substitution and frameshift mutations, and induces a dominant B conformation<sup>34</sup> and a secondary S conformation<sup>31</sup> of damaged DNA. However, <sup>PhIP</sup>-dG mainly causes substitution mutations,<sup>35</sup> and induces a mixture of the B and S conformations.<sup>36</sup> On the other hand, <sup>AF</sup>-dG mainly causes base substitution mutations,<sup>37-38</sup> and induces a mixture of B and S damaged DNA conformations,<sup>36</sup> whereas <sup>AAF</sup>-dG mainly promotes frameshift mutations,<sup>39-40</sup> and a small amount of substitution mutations,<sup>40</sup> but induces a mixture of the B, W and S DNA conformations.<sup>21</sup> Finally, <sup>1</sup>AP-dG predominantly induces frameshift mutagenesis in bacteria,<sup>41</sup> and results in the S conformation of damaged DNA.<sup>42</sup>

Despite previous efforts towards elucidating the relationship between adduct structure and mutagenicity for individual N-linked C<sup>8</sup>-dG lesions,<sup>22, 24-25, 42-43</sup> systematic attempts to understand the effects of the size and shape of the NCA moiety on the structural characteristics of damaged (adducted) DNA are scarce. In fact, to the best of our knowledge, only one such study used molecular mechanics energy minimizations to study three related adducts (<sup>AN</sup>-dG, <sup>AF</sup>-dG and <sup>1</sup>AP-dG, Figure 2.1A) in two 9-mer DNA sequences (5'-TGAXGCCG and 5'-GAATXGTG, X = modified nucleotide).<sup>44</sup> This work revealed that <sup>AN</sup>-dG distorts the DNA structure less than the larger lesions, which rationalized its

observed weaker mutagenicity. However, these lesions do not reflect systematic size expansion since the NCA moiety of <sup>AF</sup>dG contains a methylene bridge between the two aromatic rings, a chemical characteristic absent in the other adducts. Furthermore, although the previous study provides some insights into the effects of ring size, the effects of the shape of the NCA moiety on the conformational features of damaged DNA are not known. A systematic understanding of the effects of both ring size and shape is critical for developing meaningful hypotheses regarding the relationship between the structures adopted by damaged DNA and the mutagenic potential of the associated lesions.

To fill this critical void in the damaged DNA literature, the present work employs density functional theory and classical explicit-solvent molecular dynamics (MD) simulations are used to systematically study the influence of adduct composition on the structural and energetic characteristics of three plausible (i.e., B, S and W) conformations of N-linked C<sup>8</sup>-dG adducted DNA. We computationally design a range of model adducts (Figure 2.1B) to analyze the effects of incremental N-linked ring expansion at the C<sup>8</sup>-position of dG, and carefully assess the conformational features of the resulting modified DNA. Specifically, the smallest lesion (<sup>AN</sup>dG) contains a single phenyl ring, which is subsequently extended through addition of a second 6-membered aromatic ring through  $\alpha$  (2,3-) or  $\beta$  (3,4-) fusion to yield the uniquely shaped <sup>1ANP</sup>dG or <sup>2ANP</sup>dG adducts, or through (3-) substitution to form the <sup>3ABP</sup>dG adduct (Figure 2.1B). To further extend the aromatic system, a naphthalene ring was added to <sup>AN</sup>dG through  $\alpha$  or  $\beta$  fusion to generate the differently shaped <sup>1APHN</sup>dG or <sup>3APHN</sup>dG (Figure 2.1B). Finally, <sup>1AP</sup>dG includes all four aromatic ring locations adopted in the smaller adducts (Figure 2.1B). Although our primary criteria for the design of these model adducts was to systematically study the effects of the

size and shape of the aromatic amine moiety on the structural properties of damaged DNA, the adducts investigated include DNA lesions that have been experimentally verified to exhibit varied mutagenic and carcinogenic outcomes, namely <sup>AN</sup>dG<sup>13</sup> and <sup>1AP</sup>dG.<sup>41</sup> Our computational analysis reveals that the effects of incremental ring expansion on the conformational properties of damaged DNA depend on the size and shape of the NCA moiety. In addition to providing valuable structural data for authentic DNA lesions, our work complements a growing body of literature that has used model DNA adducts to probe the structure of damaged DNA<sup>27, 45-46</sup> and/or polymerase function,<sup>45, 47-48</sup> and will inspire future experimental studies that further develop the relationship between adduct structure and mutagenicity.

## 2.2 Computational Methods

### 2.2.1 DFT Calculations

To analyze the intrinsic conformational preferences of the adducts, the potential energy surface was mapped as a function of the  $\theta$  ( $\angle(\text{N9-C8-N10-C11})$ ) and  $\phi$  ( $\angle(\text{C8-N10-C11-C12})$ ) dihedral angles using nucleobase models, and as a function of  $\chi$  ( $\angle(\text{O4'-C1'-N9-C4})$ ) and  $\theta$  using nucleoside models for representative (<sup>AN</sup>dG, <sup>1ANP</sup>dG, <sup>3APHN</sup>G and <sup>1AP</sup>dG) adducts (Figure 2.1). Specifically, surfaces were generated using B3LYP/6-31G(d) optimizations by altering and constraining each dihedral angle from 0° to 350° in 10° increments. The nucleoside model was built from the lowest energy conformation obtained for the nucleobase model by replacing H9 with 2'-deoxyribose in the B-DNA conformation obtained in a previous conformational analysis<sup>49</sup> (i.e., C2'-endo pucker,  $\beta$  ( $\angle(\text{H-O5'-O5'-C5'-C4'})$ ) dihedral angle fixed to 180°, and the dihedral angle at the 3' terminus ( $\angle(\text{H-O3'-C3'-H3'})$ ) locked to 60°). Minima identified from these scans were subsequently fully

optimized using B3LYP and B3LYP-D3(BJ) with the 6-31G(d) basis set, which led to similar structural parameters. Relative energies were obtained with B3LYP-D3(BJ)/6-311+G(2df,p), and include scaled (0.9813) zero-point vibrational energy (ZPVE) corrections. All DFT calculations were performed using Gaussian 09 (revision D.01).<sup>50</sup>

## 2.2.2 MD Simulations

MD simulations were carried out by reiteratively incorporating each adduct (<sup>AN</sup>dG, <sup>1ANP</sup>dG, <sup>2ANP</sup>dG, <sup>3ABP</sup>dG, <sup>1APHN</sup>dG, <sup>3APHN</sup>dG or <sup>1AP</sup>dG, Figure 2.1B) at the X position in the 5'-CTCGGCXCCATC-3' oligonucleotide and paired opposite C in the complementary strand. This oligomer contains the *NarI* sequence (underlined) that is a well-known hotspot for deletion mutations induced by NCA carcinogens.<sup>23, 51</sup> Starting structures were generated from representative MD structures of damaged DNA in the B, W and S conformations containing the previously studied structurally-related, O-linked C8-phenoxy-dG adduct.<sup>52</sup> To delineate the conformational features of the damaged DNA, two rotamers with respect to  $\theta$  (i.e.,  $\sim 0^\circ$  and  $\sim 180^\circ$ ) must be considered for <sup>AN</sup>dG ( $\phi \sim 0$ ), while four rotamers with respect to  $\theta$  and  $\phi$  (i.e.,  $\theta \sim 0^\circ$  and  $\phi \sim 0^\circ$ ,  $\theta \sim 0^\circ$  and  $\phi \sim 180^\circ$ ,  $\theta \sim 180^\circ$  and  $\phi \sim 0^\circ$ , and  $\theta \sim 180^\circ$  and  $\phi \sim 180^\circ$ ) must be considered for each of the 6 remaining adducts, for each of the B, W and S conformations (Figure 2.2). The natural nucleotides were treated with the AMBER14SB force field,<sup>53</sup> while the General Amber Force Field (GAFF)<sup>54</sup> parameters were used to supplement the AMBER14SB force field for the adducts using antechamber.<sup>55</sup> The R.E.D. program (version III.4)<sup>56</sup> was used to calculate the partial charges of the adducted nucleotides using the B3LYP/6-31G(d) optimized structures of the *anti* and *syn* nucleoside. The assigned atom types and charges are provided in Tables A.14–A.20 in Appendix A.

Each DNA system was neutralized with 22 sodium ions using the tleap module of AMBER 14.<sup>57</sup> The damaged DNA was then solvated using an octahedral TIP3P<sup>58</sup> water box, with a minimum distance of 8 Å between the DNA and the edge of the box. The DNA hydrogen atoms were constrained using the SHAKE algorithm. A cut off distance of 8 Å was used for nonbonded interactions, and the particle-mesh Ewald method along with periodic boundary conditions was used to describe Coulombic interactions. While restraining DNA atoms with a 500 kcal mol<sup>-1</sup> Å<sup>-2</sup> force constant, each system was minimized using 500 steps of the steepest descent (SD) minimization algorithm, followed by 500 steps of the conjugate gradient (CG) minimization algorithm. The DNA was subsequently energy minimized without constraints using 1000 steps of SD minimization and 1500 steps of CG minimization. Subsequently, each system was heated to 300 K for 20 ps under conditions of constant volume using the Langevin thermostat ( $\gamma = 1$ ), while imposing a weak (10 kcal mol<sup>-1</sup> Å<sup>-2</sup>) constraint on DNA. Finally, production simulations were performed on each DNA system at constant temperature (300 K) and pressure (1 atm) using the Langevin and Berendsen algorithms, respectively. A 2 fs time step was used for the production runs, with the trajectory sampled every 2.5 ps. During the production phase, a 25 kcal mol<sup>-1</sup> Å<sup>-2</sup> harmonic distance restraint, with an average heavy atom distance of 3.0 Å, was used on the Watson-Crick hydrogen bonds in the terminal base pairs to prevent unravelling.

Due to the large number of resulting DNA systems that must be considered (78 in total), preliminary analysis was done on DNA containing the smallest <sup>AN</sup>dG or largest <sup>1AP</sup>dG adduct to identify a suitable MD sampling protocol. Specifically, eighteen 20 ns trial simulations were generated for <sup>AN</sup>dG and <sup>1AP</sup>dG adducted DNA (i.e., two and four simulations for each of <sup>AN</sup>dG and <sup>1AP</sup>dG adducted DNA in each of the B, W and S

conformations, respectively). Additional 20 ns trial simulations were subsequently initiated from representative structures of all unique damaged DNA conformations isolated from the preliminary calculations. The 20 ns production simulations were extended to 120 ns for the 18 stable conformations identified from the iterative set of trial simulations. The most stable B, W and S conformations according to molecular mechanics Poisson Boltzmann surface area (MM/PBSA) calculations identified from the 20 and 120 ns production simulations are highly similar (Table A.1, Appendix A). Furthermore, two additional independent 20 ns replicas were obtained for the lowest energy B and S conformations of <sup>AN</sup>dG and <sup>1AP</sup>dG adducted DNA using a different set of initial velocities for each trajectory. These replicas were also subsequently extended to 120 ns. All replicas yield the same structural and energetic conclusions regardless of the simulation length (Table A.2, Appendix A), with standard deviations between replicas for the average structural features and lesion-site interaction energies (LSIE) energies reported in the main text being less than 14.3° and 6.2 kJ mol<sup>-1</sup>, respectively. Therefore, the conformational themes of DNA containing the remaining adducts were investigated using the same rigorous pre-production phase conformational sampling approach, followed by 20 ns production simulations. All resulting DNA conformations were ranked with respect to their conformational free energies deduced from MM/PBSA calculations<sup>59</sup> and normal mode analysis for the enthalpy and entropy components, respectively, using a total of 400 frames of the water and sodium free trajectories. We acknowledge the shortcomings associated with MM/PBSA calculations; however, it was necessary to reduce the large number of conformations generated in our rigorous pre-production sampling step in some fashion, and it was anticipated that inaccuracies in the methodology would be relatively consistent across systems due to the overall similarities in the structures considered. This is supported by consistent MM/PBSA

conformational free energies across replicas (standard deviation in the average across replicas < 7 kJ mol, Table A.2, Appendix A). In total, our detailed approach involved over 4.5  $\mu$ s of production simulation time (excluding trial simulations). All MD simulations were conducted using the PMEMD module of AMBER 14.<sup>57, 60</sup>

The lowest energy DNA structure corresponding to each conformational theme for each adduct was further analyzed. Structural analysis was carried out using the ptraj and cpptraj modules of AMBER14. Hydrogen bonding occupancies were calculated using a cut-off of 3.4 Å for the donor–acceptor distance and 120° for the donor–hydrogen–acceptor angle. The desired number of representative structures from the simulations, as determined based on the dihedral plots for each system, were obtained by clustering the MD simulations with respect to atoms constituting the  $\chi$ ,  $\theta$  and  $\phi$  dihedral angles using the hieragglo algorithm. LSIE were calculated as the sum of the interactions between the adduct and the opposing base, the interactions between the bases of the flanking base pairs, and the interaction of the adducted base pair and the flanking base pairs using linear-interaction energy (LIE) of cpptraj. Therefore, the LSIE is a sum of the hydrogen bonding of three G:C base pairs and stacking energy of 2 base steps. DNA helical parameters were examined using the Curves+/Canal program,<sup>61</sup> modelling the lesion as G.



## 2.3 Results and Discussion

### 2.3.1 DFT Calculations Reveal that the Shape of the NCA Moiety has a Greater Influence on the Intrinsic Adduct Structure than the Size

#### 2.3.1.1 Structure at the Guanine–NCA Linkage

Contour plots of the DFT nucleobase energy as a function of the conformation at the (C–N–C) linkage connecting the NCA moiety and G (defined by the  $\theta$  and  $\phi$  dihedral angles, Figure 2.1) reveal intrinsic structural differences in the N-linked adducts with differently shaped NCA moieties (Figure A.1, Appendix A). Specifically, despite differing in the size of the NCA moiety, <sup>AN</sup>dG and <sup>3APHN</sup>G adopt stable conformations at both  $\phi \sim 0^\circ$  and  $180^\circ$  (Figure A.1A and B, Appendix A), with minima exhibiting less than  $20^\circ$  deviations from planarity (Figures 2.3A, and A.2–A.3, Appendix A). In contrast, <sup>1ANP</sup>dG and <sup>1AP</sup>dG adopt similar stable conformations at  $\phi \sim 0^\circ$ , while conformations with  $\phi \sim 180^\circ$  exhibit steric repulsion between hydrogens on the  $\alpha$ -fused ring and the nucleobase moiety (Figure A.1C and D, Appendix A), which induces significant ( $65 - 75^\circ$ ) twist within the NCA moiety in the corresponding set of stable structures (Figures 2.3B and A.4–A.5, Appendix A). This distortion will prevail in any adduct in the series with an  $\alpha$ -fused ring, and may limit the stable conformations available to the adducts within DNA duplexes. Nevertheless, irrespective of the type of ring fusion, stable conformations were obtained at two distinct  $\theta$  orientations for all adducts corresponding to  $\sim 0^\circ$  and  $\sim 180^\circ$  (Figures A.1–A.5, Appendix A). Overall, these results emphasize that the shape of the NCA moiety significantly affects the structure at the nucleobase–carcinogen linkage.

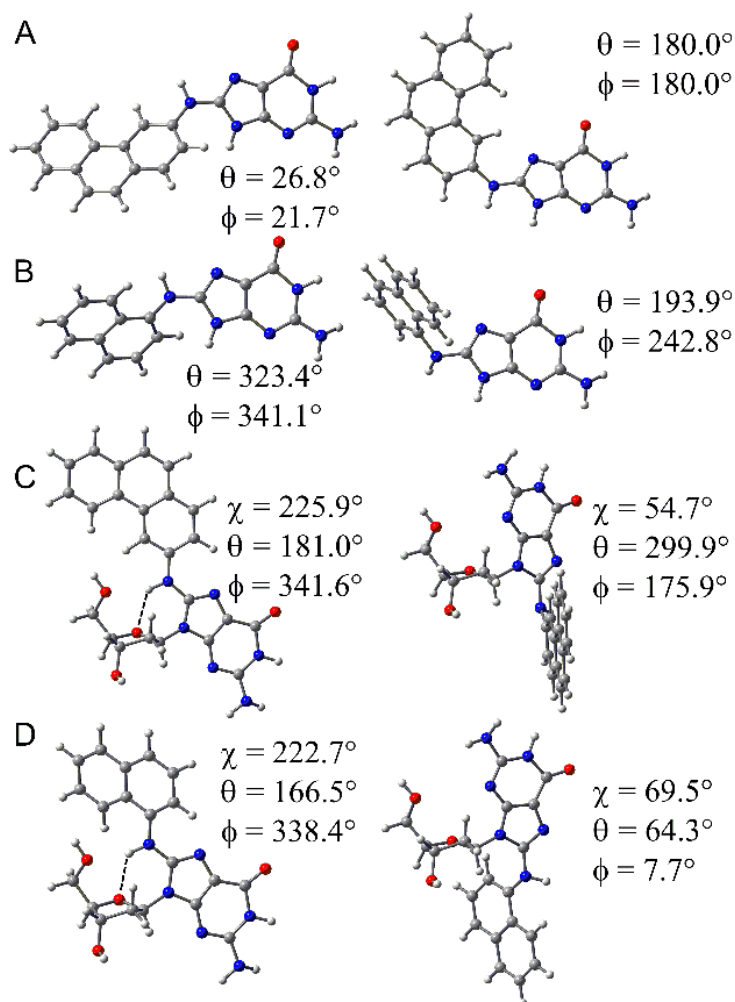


Figure 2.3: Effect of ring-fusion on the  $\phi$  torsion angle, illustrated using the approximately planar  ${}^3\text{APHN}$  dG (A, C) and nonplanar  ${}^1\text{ANP}$  dG (B, D) nucleobase (A, B) and nucleoside (C, D) adducts.

### 2.3.1.2 Structure at the Glycosidic Linkage

Contour plots of the DFT nucleoside energy as a function of the conformation at the guanine–deoxyribose glycosidic (N–C) linkage (defined by the  $\chi$  dihedral angle, Figure 2.1) and the guanine–NCA (C–N) linkage (defined by  $\theta$ ) reveal that all adducts intrinsically prefer the *anti* glycosidic orientation irrespective of the size of the NCA moiety or type of ring fusion ( $\chi = 226 - 228^\circ$ , Figures A.6 – A.10, Appendix A). However, a low (40 kJ mol $^{-1}$ ) *anti/syn* rotational barrier affords reasonable flexibility at the glycosidic linkage (Figure

A.6, Appendix A), which suggests both the *anti* and *syn* conformations may be feasible when the adduct is incorporated into DNA. More importantly, unlike the nucleobase–NCA linkage, the intrinsic structural features at the glycosidic linkage are not significantly affected by the size or shape of the C<sup>8</sup>-moiety (Figure 2.3C and D, and A.6–A.10, Appendix A) since the N-linker spatially separates the aromatic moiety from the sugar.

## **2.3.2 MD Simulations Illustrate how the DNA Structural Characteristics and Conformational Preferences Depend on the Size and Shape of the NCA Moiety**

### *2.3.2.1 General Structural Features of DNA Conformational Themes*

MD simulations reveal characteristic differences in the structural features of the B, W and S conformations that occur irrespective of the chemical composition of the adduct (Figures A.11 – A.17, Appendix A). Specifically, in the B conformation, three persistent Watson-Crick (WC) hydrogen bonds are formed between the lesion and the opposing cytosine (C), and an additional interaction occurs between the linker N–H and O4' of the damaged nucleotide (Figure 2.4A and Table A.6, Appendix A), which is intrinsic to the damaged nucleosides (Figure 2.3C and D). Due to the extrahelical (major groove) location of the C<sup>8</sup>-moiety, this conformation minimally disrupts the overall B-DNA geometry of the helix (Table A.7, Appendix A), and maintains average lesion-site noncovalent interactions (LSIE =  $-475 \pm 2$  kJ mol<sup>-1</sup>) that are equivalent to the corresponding natural DNA oligomer ( $\sim -472$  kJ mol<sup>-1</sup>, Table A.8, Appendix A). In contrast, the W conformation involves Hoogsteen hydrogen bonding between the lesion and the opposing C, and an interaction between the linker N–H group and the O4' of the nucleotide 3' with respect to the lesion (Figure 2.4B and Table A.9, Appendix A). With the exception of an increase in roll (by up

to 6° for the pseudostep composed of the base pairs flanking the lesion site), the W conformation does not significantly affect the lesion-site structural parameters, despite the *syn* adduct conformation (Table A.7, Appendix A). However, the lesion-site noncovalent interactions in this conformation (up to  $-372 \text{ kJ mol}^{-1}$ ) are significantly weaker than the B conformation, as well as the control (undamaged) helix (Table A.8, Appendix A), mainly due to weaker hydrogen bonding within the lesion base pair. In contrast, the S conformation is characterized by intercalation of the C<sup>8</sup>-moiety and the loss of lesion interstrand hydrogen bonding due to the displacement of the opposing C, which interacts with the phosphate oxygen of the nearest nucleotide present on its 3'-side (Figure 2.4C and Table A.10, Appendix A). Due to the absence of hydrogen bonding between the lesion and the opposing base, the lesion-site interaction energy (up to  $-358 \text{ kJ mol}^{-1}$ ) is weaker for this conformation compared to the other conformations and the control helix (Table A.8, Appendix A). Furthermore, the intercalation of the NCA moiety increases the roll (up to 13°) and duplex untwisting (up to 21°) at the lesion site compared to natural DNA (Table A.7, Appendix A). Despite lesion-site structural differences, the B, S and W conformations associated with all adducts maintain WC hydrogen bonding within the base pairs flanking the lesion (Figures A.11 – A.17, and Tables A.6, A.9 and A.10, Appendix A).

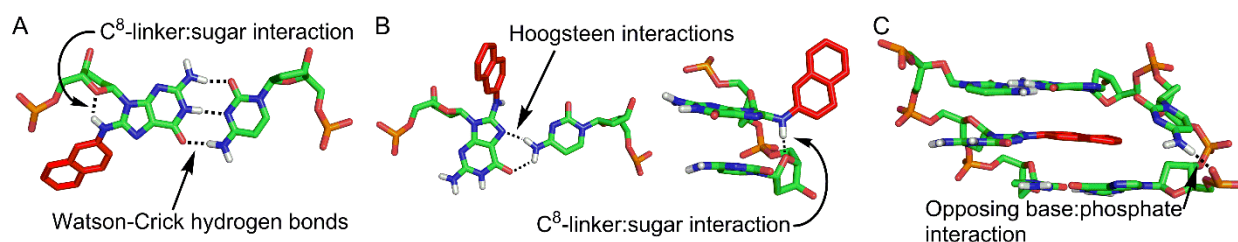


Figure 2.4: Discrete hydrogen-bonding interactions observed during MD simulations on the (A) B, (B) W and (C) S conformations of DNA containing N-linked C<sup>8</sup>-dG adducts. <sup>2ANP</sup>dG adducted DNA is shown as an example.

### 2.3.2.2 The shape of the NCA moiety affects the lesion structural characteristics in DNA

Despite overall similarities in the global structural features of adducted DNA in the three conformational themes, MD simulations reveal a marked dependence of several adduct structural characteristics on the shape of the NCA moiety. First, although all adducts adopt consistent  $\chi$  orientations in the B (average  $\chi$  ( $\chi^{\text{av}}$ ) of  $216 - 231^\circ$ ) and W ( $\chi^{\text{av}} = 64 - 71^\circ$ ) conformations (Table A.11, Appendix A) that align with the DFT predicted lowest energy *anti* and *syn* structures for the isolated nucleosides (Figures A.7 – A.10, Appendix A), the orientation of the damaged G about the glycosidic bond in the S conformation depends on the shape of the C<sup>8</sup>-moiety. Specifically, the three adducts that expand upon single-ringed <sup>AN</sup>dG through single ring substitution (<sup>3ABP</sup>dG) or naphthalene ring fusion (<sup>1APHN</sup>dG or <sup>3APHN</sup>dG) adopt a noticeably different glycosidic orientation ( $\chi^{\text{av}} = 119^\circ - 132^\circ$ ) compared to other adducts ( $\chi^{\text{av}} = 38^\circ - 48^\circ$ , Table A.11, Appendix A), which are more similar to the inherent nucleoside *syn* conformation. This difference occurs to most efficiently stack the aromatic C<sup>8</sup>-systems within the helix (Figure A.18, Appendix A). Second, although all adducts are nearly planar about the guanine–NCA linkage (average  $\theta$  ( $\theta^{\text{av}}$ ) =  $180 \pm 26^\circ$ ) in the S conformation to maximize the stacking interactions of the C<sup>8</sup>-moiety within DNA, and the guanine and NCA rings are almost orthogonal ( $\theta^{\text{av}} = 62 - 76^\circ$ ) in the W conformation to accommodate the C<sup>8</sup>-moiety in the narrow DNA minor groove, the  $\theta$ -orientation in the B conformation depends on the shape of the C<sup>8</sup>-moiety (Table A.11, Appendix A). Specifically, adducts with  $\alpha$ -ring fusion are more planar ( $\theta^{\text{av}} = 165 - 175^\circ$ ) than the other adducts ( $\theta^{\text{av}} = 125 - 140^\circ$ ), which permits (albeit weak) hydrogen bonding between the  $\alpha$ -fused ring and the DNA backbone (Figure A.19, Appendix A). Third, all adducts in the B conformation adopt a consistent  $\phi$ -orientation within the NCA moiety

(average  $\phi$  ( $\phi^{av}$ ) of  $3^\circ - 18^\circ$ ); however, in the W and S conformers, adducts without an  $\alpha$ -fused ring adopt  $\phi \sim 180^\circ$  (except the W conformation of  ${}^3\text{APHN dG}$ ; Table A.11, Appendix A), while adducts with an  $\alpha$ -fused ring adopt  $\phi \sim 0^\circ$  to minimize steric repulsion with the backbone (Figure A.20, Appendix A). Overall, these results suggest that despite little size dependence, the N-linked lesions possess specific shape-dependent conformational signatures in duplex DNA.

### 2.3.2.3 The Size and Shape of the NCA Moiety Affect Lesion-Site Noncovalent Interactions

Although the structural distortion (nonplanarity) in the adduct about the guanine–NCA linkage in the B conformation depends on the shape of the NCA moiety, the lesion-site interactions are similar for all adducts (within  $2.3 \text{ kJ mol}^{-1}$ ) due to the major groove location of the bulky moiety (Figure 2.5A). Nevertheless, the size and shape of the  $\text{C}^8$ -moiety significantly affects the magnitude of lesion-site noncovalent interactions in the W conformation (Figure 2.5B). Specifically, despite a small (up to  $3 \text{ kJ mol}^{-1}$ ) effect of inclusion of a second ring through  $\alpha$  or  $\beta$ -ring fusion to the single phenyl moiety of  ${}^{\text{AN}}\text{dG}$  (i.e.,  ${}^1\text{ANP dG}$  and  ${}^2\text{ANP dG}$ ) on the lesion-site stabilization, a larger ( $\sim 10 \text{ kJ mol}^{-1}$ ) enhancement occurs upon phenyl ring substitution (i.e.,  ${}^3\text{ABP dG}$ ). On the other hand, although the addition of a naphthalene ring to  ${}^{\text{AN}}\text{dG}$  through  $\alpha$ -fusion (i.e.,  ${}^1\text{APHN dG}$ ) significantly decreases the lesion-site stabilization (by  $\sim 14 \text{ kJ mol}^{-1}$ ), naphthalene addition by  $\beta$ -fusion leads to similar stabilization (i.e.,  ${}^{\text{AN}}\text{dG}$  and  ${}^3\text{APHN dG}$  lesion-site interactions differ by only  $\sim 2 \text{ kJ mol}^{-1}$ ). Nevertheless, the largest  ${}^1\text{AP dG}$  adduct leads to slightly greater stabilization compared to the smallest  ${}^{\text{AN}}\text{dG}$  lesion (by  $\sim 6 \text{ kJ mol}^{-1}$ ). Thus, although NCA moieties of all sizes and shapes are well accommodated in the major groove in the B conformation, there is a complicated dependence of the lesion-site stability on both the size

and shape of the C<sup>8</sup>-moiety for the W conformation due to steric constraints within the narrow minor groove.

In comparison to the B and W conformations, the S conformation exhibits larger and more systematic size- and shape-dependent effects on the lesion-site stabilization. Specifically, in terms of shape (Figure 2.5C), although the position of the second phenyl ring in the C<sup>8</sup>-system has a relatively small ( $\sim 6 \text{ kJ mol}^{-1}$ ) effect on the lesion-site interactions, stronger shape effects occur upon addition of naphthalene to the first phenyl ring. Indeed,  $\alpha$ -fusion of naphthalene to <sup>AN</sup>dG (<sup>1APHN</sup>dG) on average provides  $\sim 21 \text{ kJ mol}^{-1}$  greater stabilization compared to  $\beta$ -fusion (<sup>3APHN</sup>dG, Figure 2.5C and Table A.8, Appendix A). In terms of size effects (Figure 2.5D), ring addition through initial  $\alpha$ -fusion exhibits a consistent increase (up to  $23 \text{ kJ mol}^{-1}$ ) in lesion-site stabilization upon addition of the first and second rings, as well as a subsequent increase for addition of the third ring ( $\sim 10 \text{ kJ mol}^{-1}$ ). However, the effects of ring extension through  $\beta$ -fusion are less systematic, with addition of the third ring destabilizing the system compared to the second ring (Figure 2.5D). Overall, due to differences in stacking between the lesion and the flanking base pairs, ring expansion through  $\alpha$ -fusion contributes more significantly and systematically to lesion-site stabilization in the S conformation compared to  $\beta$ -fusion.

#### 2.3.2.4 The Shape of the NCA Moiety Affects its Solvent Accessibility

In addition to variations in lesion-site noncovalent interactions, the extent of energetic destabilization caused by solvation (hydration) of the hydrophobic NCA moiety will depend on the size and shape of the C<sup>8</sup>-moiety, and the conformation of damaged DNA. Specifically, the relative extent of hydrophobic destabilization can be assessed in terms of

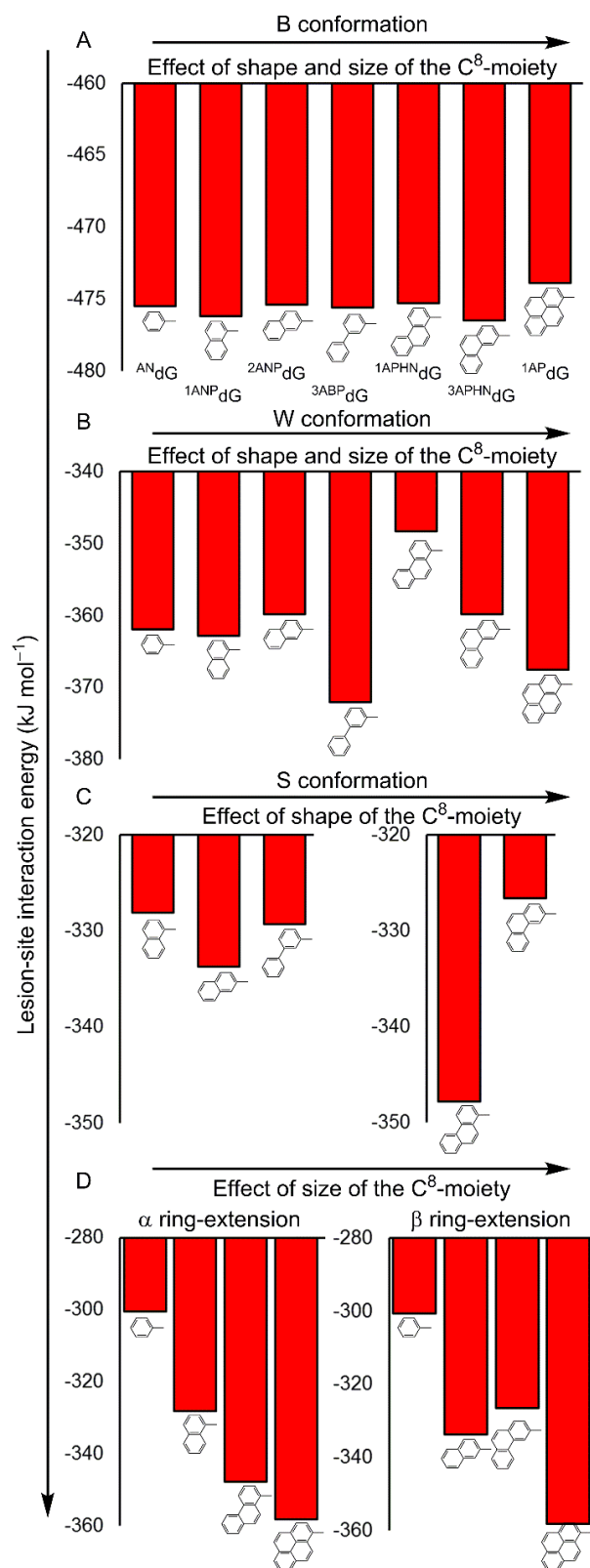


Figure 2.5: Effect of size and shape of the C<sup>8</sup>-moiety on the lesion-site interaction energy (LSIE) within the B, W and S conformations of damaged DNA containing an N-linked dG adducts



the solvent accessible surface area (SASA) of the NCA moieties (Figure 2.6 and Table A.12, Appendix A). In the B or W conformations, which solvent expose the NCA-moiety in the DNA (major or minor) groove, expansion from a single phenyl ring (<sup>AN</sup>dG) by ring fusion (<sup>1ANP</sup>dG and <sup>2ANP</sup>dG) slightly increases the solvent accessibility (up to  $\sim 31 \text{ \AA}^2$ ), while a much greater increase (at least  $67 \text{ \AA}^2$ ) occurs upon ring substitution (<sup>3ABP</sup>dG). This increase is consistently maintained for the largest adducts (<sup>1APHN</sup>dG, <sup>3APHN</sup>dG and <sup>1AP</sup>dG), which all contain the same ring substitution position. Therefore, the larger lesions may induce greater DNA destabilization compared to smaller adducts. Regardless of the size or shape, the NCA moieties within the S conformation are significantly less solvent exposed compared to the B and W conformations (Figure 2.6). These features may have implications for the relative stability of the three damaged DNA conformations (*vide infra*).

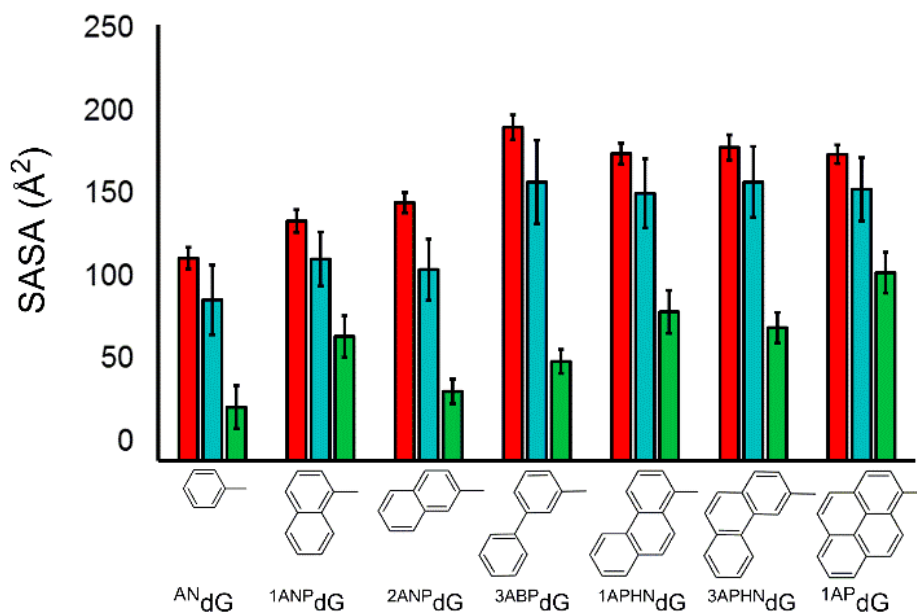


Figure 2.6: Dependence of the solvent accessible surface area (SASA,  $\text{\AA}^2$ ) on the shape and size of the hydrophobic NCA for the B (red), W (blue) and S (green) conformations of damaged DNA containing N-linked dG adducts.

### *2.3.2.5 The Size and Shape of the NCA Moiety Affect the Relative Stability of Different DNA Conformations*

Analysis of the combined effect of (stabilizing) noncovalent interactions at the lesion site and (destabilizing) hydrophobic interactions of the NCA moiety provides clues into the conformational preferences of damaged DNA possessing differently sized and shaped NCA moieties. Our analysis reveals that since the W conformation has much weaker lesion-site interactions than the B conformation (Table A.8, Appendix A) and maintains substantial hydrophobic destabilization of the C<sup>8</sup>-moiety (Table A.12, Appendix A), the relative contribution of the W conformation for all studied adducts is likely marginal, regardless of the NCA moiety size or shape. Indeed, the proposed inaccessibility of the W conformation for the studied adducts is supported by the fact that this conformation has been observed experimentally only for DNA containing the <sup>AA</sup>FdG lesion that contains a substituent on the N-linker atom,<sup>21</sup> and for DNA containing <sup>AF</sup>dG paired against mismatches.<sup>26, 29, 62</sup>

Although the lesion-site interactions in the B conformation are independent of the C<sup>8</sup>-moiety size or shape, the small size of the NCA moiety in the single-ringed <sup>AN</sup>dG lesion results in significantly less solvent accessibility compared to any other adduct (Figure 2.7A), and hence less pronounced destabilization of the B conformation occurs due to hydrophobic interactions. Furthermore, <sup>AN</sup>dG leads to significantly weaker lesion-site interactions in the S conformation compared to any other lesion (Figure 2.7B). When coupled with the significantly greater lesion-site stabilization provided by <sup>AN</sup>dG to the B conformation compared to the S conformation, these features suggest that DNA containing the <sup>AN</sup>dG adduct will exclusively prefer the B conformation. This conclusion aligns with the previously suggested B conformational preference for <sup>AN</sup>dG adducted DNA of a

different sequence,<sup>44</sup> which suggests that the B conformation dominates irrespective of sequence context.

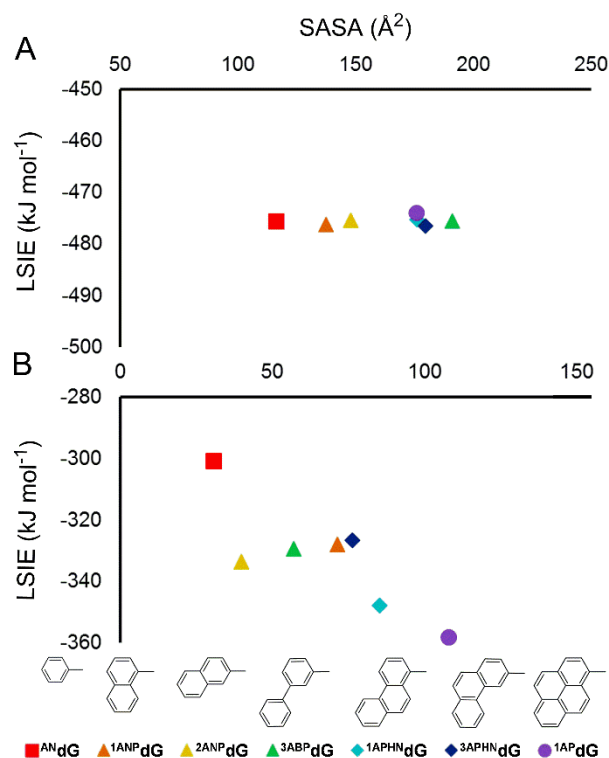


Figure 2.7: Dependence of the solvent accessible surface area (SASA, Å²) and lesion-site interaction energy (LSIE, kJ mol⁻¹) on the shape and size of the hydrophobic C<sup>8</sup>-moiety for the (A) B and (B) S conformations of DNA damaged by N-linked dG adducts.

In contrast to the smallest <sup>AN</sup>dG adduct, although the B conformation of DNA adducted with <sup>1APHN</sup>dG or the largest <sup>1AP</sup>dG lesion possess a significant lesion-site interaction energy, the associated bulky moieties are among the most solvent exposed, and thereby most significantly destabilize the B conformation (Figure 2.7A). Furthermore, despite substantial solvent exposure of the bulky moieties in the S conformation, the overall solvent accessibility is less for the S than the B conformation. In addition, the large bulky moieties of <sup>1APHN</sup>dG and <sup>1AP</sup>dG lead to substantial lesion-site stabilization in the S conformation through enhanced stacking with the flanking base pairs (Figure 2.7B), being the largest amongst all model adducts. This suggests that only the S conformation is

accessible for adducted DNA containing  $^{1\text{APHN}}\text{dG}$  and  $^{1\text{AP}}\text{dG}$ . This proposal is supported at least in part by previous NMR studies, which revealed that  $^{1\text{AP}}\text{dG}$  adopts an S conformation in solution in a sequence context similar to the one considered in the present study.<sup>42</sup> Regardless, comparison of  $^{\text{AN}}\text{dG}$ ,  $^{1\text{APHN}}\text{dG}$  and  $^{1\text{AP}}\text{dG}$  illustrates the effect of NCA size on the preferred DNA conformation.

The remaining 4 adducts ( $^{1\text{ANP}}\text{dG}$ ,  $^{2\text{ANP}}\text{dG}$ ,  $^{3\text{ABP}}\text{dG}$  and  $^{3\text{APHN}}\text{dG}$ ) display a similar balance between the stability provided to the lesion site and the solvent accessibility of the bulky moiety in the S conformation (Figure 2.7B). Specifically, in all cases, the solvent accessibility of the C<sup>8</sup>-moiety is lower than for the 1APHN and 1AP moieties, while the lesion-site interactions are greater than for the single-ringed AN. This suggests that the S conformation may also be accessible to DNA containing these polycyclic adducts, regardless of the size or shape of the NCA moiety. Nevertheless, subtle size and shape effects will likely dictate the B:S ratio for these intermediate sized adducts. Specifically, the bulky moieties containing two fused rings ( $^{1\text{ANP}}\text{dG}$  and  $^{2\text{ANP}}\text{dG}$ ) are significantly less solvent exposed in the B conformation (Figure 2.7A), suggesting a greater preference for the B conformation for these adducts compared to  $^{3\text{ABP}}\text{dG}$  and  $^{3\text{APHN}}\text{dG}$ , and further highlighting the influence of the C<sup>8</sup>-moiety shape (i.e.,  $^{1\text{ANP}}\text{dG}$  versus  $^{3\text{ABP}}\text{dG}$ ) and size (i.e.,  $^{2\text{ANP}}\text{dG}$  versus  $^{3\text{APHN}}\text{dG}$ ). Additionally, for  $^{1\text{ANP}}\text{dG}$  and  $^{2\text{ANP}}\text{dG}$ , the greater solvent accessibility in the B conformer, the lower solvent accessibility in the S conformer, and the larger average lesion-site interaction energies in the S conformer suggests  $^{2\text{ANP}}\text{dG}$  may lead to the S conformation more frequently than  $^{1\text{ANP}}\text{dG}$ , further illustrating the effect of NCA shape. Thus, overall, the B:S ratio likely decreases as  $^{1\text{ANP}}\text{dG} > ^{2\text{ANP}}\text{dG} \gg ^{3\text{ABP}}\text{dG} \sim ^{3\text{APHN}}\text{dG}$  for these intermediate sized adducts. In addition, the relative solvent accessibility

of these adducts in the B conformer of damaged DNA suggests the smaller  $^{1\text{ANP}}\text{dG}$  and  $^{2\text{ANP}}\text{dG}$  lesions will predominantly adopt the B conformation, while  $^{3\text{ABP}}\text{dG}$  and  $^{3\text{APHN}}\text{dG}$  will predominantly adopt the S conformation.

In summary, this comparison clearly emphasizes that both the extent (size) and direction (shape) of ring expansion affect the damaged DNA conformations associated with N-linked adducts. Specifically, the conformation adopted is determined by a balance in the stabilization provided by lesion-site interactions and destabilization due to hydrophobic effects arising from solvent exposure of the NCA moiety. Thus, the smallest N-linked adduct with a single-ringed NCA moiety ( $^{\text{AN}}\text{dG}$ ), as well as the two adducts containing NCA moieties with two fused rings ( $^{1\text{ANP}}\text{dG}$  and  $^{2\text{ANP}}\text{dG}$ ), may preferentially induce a B conformation, with  $^{\text{AN}}\text{dG}$  exhibiting the greatest propensity to induce the B conformation. On the other hand, DNA adducted with a biphenyl-type double-ring system ( $^{3\text{ABP}}\text{dG}$ ) or larger (3 or 4-) ringed systems ( $^{3\text{APHN}}\text{dG}$ ,  $^{1\text{APHN}}\text{dG}$  and  $^{1\text{AP}}\text{dG}$ ) may predominantly adopt the S conformation, with  $^{1\text{APHN}}\text{dG}$  and  $^{1\text{AP}}\text{dG}$  exhibiting the greatest propensity to induce the S conformation.

### **2.3.3 Implications to the repair propensity and mutagenicity of N-linked C<sup>8</sup>-dG adducts**

The persistence and mutagenicity of bulky DNA adducts depend on their propensity to be recognized and repaired in cells. Bulky lesions are typically repaired by nucleotide excision repair (NER), which removes 24–32 nucleotides from the damage-containing strand.<sup>63</sup> It is widely believed that lesion sensing by recognition factors involved in the initial NER steps is influenced by changes in the thermodynamic stabilization at the site of the DNA lesion, which can, in turn, be influenced by a number of factors, including impaired Watson-Crick hydrogen bonding,<sup>64-65</sup> decreased stacking,<sup>66-67</sup> increased helical

distortions,<sup>68-69</sup> and increased helical<sup>70-71</sup> and backbone dynamics,<sup>68</sup> compared to natural (undamaged) DNA.<sup>72</sup> In this context, a crystal structure of Rad4, the yeast ortholog of the eukaryotic NER recognition factor (XPC-Rad23) bound to damaged DNA revealed that lesion recognition is accompanied by insertion of a  $\beta$ -hairpin into the DNA major groove at the damage site, along with simultaneous flipping of the lesion and opposing base out of the helix.<sup>73</sup> Further studies have proposed that lesion-site structural and thermodynamic perturbations facilitate repair by increasing the residence time of the recognition factor, decreasing the opening time associated with opposing base flipping, and influencing the rate of DNA untwisting and nucleotide flipping.<sup>73-76</sup> On the basis of these experimental observations, recent computational studies have further proposed that (i) lesion recognition involves a correlated motion involving scanning of the locally destabilized region by the recognition factor,  $\beta$ -hairpin insertion and lesion eviction,<sup>77</sup> (ii) lesion recognition and flipping of the opposing base by XPC-Rad23B is enhanced by decreased lesion-site stacking,<sup>78</sup> and (iii) a pre-flipped opposing base can foster binding of the recognition factor to damaged DNA.<sup>79</sup>

Due to a similar lesion-site structure, including intact Watson-Crick hydrogen bonding and stacking stabilization comparable to undamaged DNA, our results suggest that the <sup>AN</sup>dG, <sup>1ANP</sup>dG and <sup>2ANP</sup>dG adducts that preferentially induce the B conformation will have similar repair propensities. Since the  $\beta$ -hairpin of the recognition factor is inserted at the lesion site through the DNA major groove, it is possible that the favored major groove positioning of the NCA moieties of these adducts may obstruct the insertion of the  $\beta$  hairpin, and hence repair. However, even if left unrepaired, the associated damaged DNA may behave similar to undamaged DNA since the B conformation will allow the Watson-

Crick edge of the damaged guanine to pair with complementary cytosine during replication. Therefore, these lesions will likely be weakly mutagenic. This proposal aligns with the fact that the smallest (<sup>AN</sup>dG) lesion has been previously reported to induce the B conformation and evidence that AN is a weak mutagen.<sup>13</sup>

In contrast, due to greater lesion site changes, including disrupted Watson-Crick pairing, increased helical distortions (i.e., lesion-site untwisting, Table A.7, Appendix A) and the flipped-out location of the opposing base, the <sup>1APHN</sup>dG, <sup>1AP</sup>dG, <sup>3APHN</sup>dG and <sup>3ABP</sup>dG adducts that preferentially induce the S conformation are more likely be recognized compared to the lesions that prefer the B conformation. However, the lesion site in the S conformation of <sup>1APHN</sup>dG and <sup>1AP</sup>dG adducted DNA is substantially (at least 18.5 kJ mol<sup>-1</sup>) more stable compared to <sup>3APHN</sup>dG and <sup>3ABP</sup>dG adducted DNA, mainly due to enhanced stacking interactions. This greater lesion-site stability may provide resistance to NER recognition and repair for <sup>1APHN</sup>dG and <sup>1AP</sup>dG compared to <sup>3ABP</sup>dG and <sup>3APHN</sup>dG. Comparison of <sup>1APHN</sup>dG and <sup>1AP</sup>dG to <sup>3APHN</sup>dG highlights that an  $\alpha$ -fused ring is important for imparting NER resistance to larger polycyclic lesions that induce the S conformation of adducted DNA.

Overall, the predicted differences in the relative repair propensities of <sup>1AN</sup>dG and <sup>1AP</sup>dG illustrate the effect of the size of the NCA moiety on NER recognition, whereas the difference in the repair propensities of <sup>1APHN</sup>dG and <sup>3APHN</sup>dG demonstrates the influence of the NCA moiety shape on imparting NER resistance. Furthermore, the anticipated relative repair propensities suggests that the studied lesions will have different survival rates, and thereby differential mutagenic outcomes in cells. This proposal correlates with previous experimental replication outcomes for a series of size-expanded N<sup>2</sup>-linked dG adducts (see,

for example, <sup>80-82</sup>). However, to fully appreciate the mutagenic outcomes associated with C<sup>8</sup>-dG lesions, and further probe the structure-activity relationship related to lesion size and shape, the polymerase processing of adducted DNA must be investigated.

## 2.4 Conclusion

In summary, combined quantum mechanical and molecular dynamics calculations are used to examine seven NCA containing C<sup>8</sup>-dG adducts, and thereby assess the effect of adduct size and shape on DNA structure. DFT calculations demonstrate the significant effect of the NCA moiety shape on the geometry at the nucleobase–carcinogen linkage. MD simulations substantiate that the adducts can be divided based on the B:S ratio of damaged DNA conformations, with the smallest <sup>AN</sup>dG adduct displaying the greatest propensity for the B conformation, the largest <sup>AP</sup>dG adduct having the greatest propensity for the S conformation, and the remainder of the adducts falling between the two extremes, with the ratio dictated by the size and shape of the bulky moiety. Taken together, the present contribution highlights an emerging theme of the effect of the bulky moiety shape on the damaged DNA conformational and associated repair outcomes,<sup>45</sup> which complements previously established effects in the literature such as size,<sup>44, 83</sup> linker type,<sup>84</sup> ionization state<sup>84</sup> and exocyclic substituents<sup>20, 27, 84</sup> of the C<sup>8</sup>-linked aromatic moiety. Furthermore, the incremental structural effects our work uncovers for the systematic series of N-linked adducts suggests that these model systems can serve as useful biochemical tools to further probe the cellular processing of DNA adducted by potential carcinogens that are ubiquitous in our environment. Importantly, this work provides the necessary groundwork for future computational studies of these model adducts interacting with polymerases and repair enzymes to fully characterize their mutagenic effects.



## 2.5 References

1. Kovacic, P.; Somanathan, R., Nitroaromatic compounds: Environmental toxicity, carcinogenicity, mutagenicity, therapy and mechanism. *Journal of Applied Toxicology* **2014**, 34 (8), 810-824.
2. Gibis, M., Heterocyclic aromatic amines in cooked meat products: causes, formation, occurrence, and risk assessment. *Comprehensive Reviews in Food Science and Food Safety* **2016**, 15 (2), 269-302.
3. Stabbert, R., et al., Analysis of aromatic amines in cigarette smoke. *Rapid Communications in Mass Spectrometry* **2003**, 17 (18), 2125-2132.
4. Wei, E.; Shu, H. P., Nitroaromatic carcinogens in diesel soot: a review of laboratory findings. *American journal of public health* **1983**, 73 (9), 1085-1088.
5. Scribner, J. D., et al., Mechanisms of action of carcinogenic aromatic amines: an investigation using mutagenesis in bacteria. *Chemico-biological interactions* **1979**, 26 (1), 11-25.
6. Hatch, F. T., et al., Extended quantitative structure–activity relationships for 80 aromatic and heterocyclic amines: structural, electronic, and hydrophobic factors affecting mutagenic potency. *Environmental and molecular mutagenesis* **2001**, 38 (4), 268-291.
7. Beland, F. A., Carcinogenesis and Mutagenesis. 1990; Vol. 94, pp 267-325.
8. Shamovsky, I., et al., Theoretical studies of chemical reactivity of metabolically activated forms of aromatic amines toward DNA. *Chemical Research in Toxicology* **2012**, 25 (10), 2236-2252.
9. Humphreys, W. G., et al., Mechanism of C8 alkylation of guanine residues by activated arylamines: evidence for initial adduct formation at the N7 position. *Proceedings of the National Academy of Sciences* **1992**, 89 (17), 8278-8282.
10. Millen, A. L., et al., C8-linked bulky guanosine DNA adducts: experimental and computational insights into adduct conformational preferences and resulting mutagenicity. *Future medicinal chemistry* **2012**, 4 (15), 1981-2007.
11. Krauss, R. S., et al., The formation of 2-aminofluorene-dna adducts in vivo: Evidence for peroxidase-mediated activation. *Journal of Biochemical and Molecular Toxicology* **1989**, 4 (2), 111-117.
12. Howard, P., et al., The binding of N-hydroxy-2-acetylaminofluorene to DNA and repair of the adducts in primary rat hepatocyte cultures. *Carcinogenesis* **1981**, 2 (2), 97-102.

13. Jacobson, M. D., et al., Synthesis and conformation of a dinucleoside monophosphate modified by aniline. *Chemical research in toxicology* **1988**, 1 (3), 152-159.
14. Beland, F. A., et al., Arylamine-DNA adducts in vitro and in vivo: their role in bacterial mutagenesis and urinary bladder carcinogenesis. *Environmental health perspectives* **1983**, 49, 125-134.
15. Schut, H. A.; Snyderwine, E. G., DNA adducts of heterocyclic amine food mutagens: implications for mutagenesis and carcinogenesis. *Carcinogenesis* **1999**, 20 (3), 353-368.
16. Howard, P. C., et al., Formation of DNA adducts in vitro and in *Salmonella typhimurium* upon metabolic reduction of the environmental mutagen 1-nitropyrene. *Cancer research* **1983**, 43 (5), 2052-2058.
17. Beland, F. A.; Kadlubar, F. F., Formation and persistence of arylamine DNA adducts in vivo. *Environmental Health Perspectives* **1985**, 62, 19.
18. Arlt, V. M., et al., Identification of three major DNA adducts formed by the carcinogenic air pollutant 3-nitrobenzanthrone in rat lung at the C8 and N2 position of guanine and at the N6 position of adenine. *International journal of cancer* **2006**, 118 (9), 2139-2146.
19. Jacobson, M. D., et al., Synthesis and conformation of a dinucleoside monophosphate modified by aniline. *Chemical Research in Toxicology* **1988**, 1 (3), 152-9.
20. Mu, H., et al., Nucleotide excision repair of 2-acetylaminofluorene- and 2-aminofluorene-(C8)-guanine adducts: molecular dynamics simulations elucidate how lesion structure and base sequence context impact repair efficiencies. *Nucleic Acids Research* **2012**, 40 (19), 9675-9690.
21. Patnaik, S.; Cho, B. P., Structures of 2-acetylaminofluorene modified DNA revisited: Insight into conformational heterogeneity. *Chemical Research in Toxicology* **2010**, 23 (11), 1650-1652.
22. O'Handley, S. F., et al., Structural characterization of an N-acetyl-2-aminofluorene (AAF) modified DNA oligomer by NMR, energy minimization, and molecular dynamics. *Biochemistry* **1993**, 32 (10), 2481-2497.
23. Burnouf, D., et al., Single adduct mutagenesis: strong effect of the position of a single acetylaminofluorene adduct within a mutation hot spot. *Proceedings of the National Academy of Sciences of the United States of America* **1989**, 86 (11), 4147-4151.
24. Mao, B., et al., Solution structure of the aminofluorene [AF]-external conformer of the anti-[AF]-C8-dG adduct opposite dC in a DNA duplex. *Biochemistry* **1998**, 37 (1), 95-106.

25. Mao, B., et al., Solution structure of the aminofluorene [AF]-intercalated conformer of the syn-[AF]-C8-dG adduct opposite dC in a DNA duplex. *Biochemistry* **1998**, 37 (1), 81-94.
26. Norman, D., et al., NMR and computational characterization of the N-(deoxyguanosin-8-yl)aminofluorene adduct [(AF)G] opposite adenosine in DNA: (AF)G[syn].cntdot.A[anti] pair formation and its pH dependence. *Biochemistry* **1989**, 28 (18), 7462-7476.
27. Jain, V., et al., Conformational and thermodynamic properties modulate the nucleotide excision repair of 2-aminofluorene and 2-acetylaminofluorene dG adducts in the NarI sequence. *Nucleic Acids Research* **2012**, 40 (9), 3939-3951.
28. Xu, L.; Cho, B. P., Conformational insights into the mechanism of acetylaminofluorene-dG-induced frameshift mutations in the NarI mutational hotspot. *Chemical Research in Toxicology* **2016**, 29 (2), 213-226.
29. Jain, N., et al., Influence of flanking sequence context on the conformational flexibility of aminofluorene-modified dG adduct in dA mismatch DNA duplexes. *Nucleic Acids Research* **2009**, 37 (5), 1628-1637.
30. Broyde, S., et al., DNA adduct structure–function relationships: comparing solution with polymerase structures. *Chemical Research in Toxicology* **2008**, 21 (1), 45-52.
31. Cho, B., Structure–function characteristics of aromatic amine-DNA adducts. In *The Chemical Biology of DNA Damage*, Wiley-VCH Verlag GmbH & Co. KGaA: 2010; pp 217-238.
32. Lukin, M.; de los Santos, C., NMR structures of damaged DNA. *Chemical Reviews* **2006**, 106 (2), 607-686.
33. Patel, D. J., et al., Nuclear magnetic resonance solution structures of covalent aromatic amine– DNA adducts and their mutagenic relevance. *Chemical research in toxicology* **1998**, 11 (5), 391-407.
34. Cho, B. P., et al., NMR structural studies of a 15-mer DNA sequence from a ras protooncogene, modified at the first base of codon 61 with the carcinogen 4-aminobiphenyl. *Biochemistry* **1992**, 31 (40), 9587-9602.
35. Shibutani, S., et al., Mutagenesis of the N-(deoxyguanosin-8-yl)-2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine DNA adduct in mammalian cells: sequence context effects. *Journal of Biological Chemistry* **1999**, 274 (39), 27433-27438.
36. Brown, K., et al., Solution structure of the 2-amino-1- methyl-6-phenylimidazo[4,5-b]pyridine C8-deoxyguanosine adduct in duplex DNA. *Proceedings of the National Academy of Sciences* **2001**, 98 (15), 8507-8512.

37. Bichara, M.; Fuchs, R. P., DNA binding and mutation spectra of the carcinogen N-2-aminofluorene in *Escherichia coli*: A correlation between the conformation of the premutagenic lesion and the mutation specificity. *Journal of molecular biology* **1985**, *183* (3), 341-351.
38. Gupta, P. K., et al., Comparison of mutagenesis induced in single-and double-stranded M13 viral DNA by treatment with N-hydroxy-2-aminofluorene. *Carcinogenesis* **1988**, *9* (8), 1337-1345.
39. Tan, X., et al., Mutagenic events in *Escherichia coli* and mammalian cells generated in response to acetylaminofluorene-derived DNA adducts positioned in the *NarI* restriction enzyme site. *Biochemistry* **2002**, *41* (48), 14255-14262.
40. Shibutani, S.; Grollman, A. P., Molecular mechanisms of mutagenesis by aromatic amines and amides. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **1997**, *376* (1), 71-78.
41. Stanton, C. A., et al., The mutagenicity and DNA base sequence changes induced by 1-nitroso- and 1-nitropyrene in the *cl* gene of lambda prophage. *Carcinogenesis* **1988**, *9* (7), 1153-1158.
42. Mao, B., et al., Solution conformation of the N-(deoxyguanosin-8-yl)-1-aminopyrene ([AP] dG) adduct opposite dC in a DNA duplex. *Biochemistry* **1996**, *35* (39), 12659-12670.
43. Mu, H., et al., Role of structural and energetic factors in regulating repair of a bulky DNA lesion with different opposite partner bases. *Biochemistry* **2013**, *52* (33), 5517-5521.
44. Shapiro, R., et al., Effect of ring size on conformations of aromatic amine– DNA adducts: the aniline– C8 guanine adduct resides in the B-DNA major groove. *Chemical research in toxicology* **1998**, *11* (4), 335-341.
45. Sproviero, M., et al., Structural and biochemical impact of C8-aryl-guanine adducts within the *NarI* recognition DNA sequence: influence of aryl ring size on targeted and semi-targeted mutagenicity. *Nucleic Acids Research* **2014**, *42* (21), 13405-13421.
46. Jain, V., et al., Unusual sequence effects on nucleotide excision repair of arylamine lesions: DNA bending/distortion as a primary recognition factor. *Nucleic Acids Research* **2013**, *41* (2), 869-880.
47. Wyss, L. A., et al., Specific incorporation of an artificial nucleotide opposite a mutagenic DNA adduct by a DNA polymerase. *Journal of the American Chemical Society* **2015**, *137* (1), 30-33.
48. Wyss, L. A., et al., The use of an artificial nucleotide for polymerase-based recognition of carcinogenic O(6)-alkylguanine DNA adducts. *Nucleic Acids Research* **2016**, *44* (14), 6564-6573.

49. Millen, A. L., et al., Computational and experimental evidence for the structural preference of phenolic C-8 purine adducts. *The Journal of Physical Chemistry A* **2008**, *112* (16), 3742-53.
50. Frisch, M. J., et al., Gaussian 09, Revision D.01. Wallingford CT, 2009.
51. Fuchs, R. P., et al., Hot spots of frameshift mutations induced by the ultimate carcinogen N-acetoxy-N-2-acetylaminofluorene. *Nature* **1981**, *294* (5842), 657-659.
52. Kuska, M. S., et al., Structural Influence of C8-Phenoxy-Guanine in the NarI Recognition DNA Sequence. *Chemical Research in Toxicology* **2013**, *26* (9), 1397-1408.
53. Maier, J. A., et al., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **2015**, *11* (8), 3696-3713.
54. Wang, J., et al., Development and Testing of a General Amber Force Field. *Journal of Computational Chemistry* **2004**, *25* (9), 1157-1174.
55. Wang, J., et al., Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of molecular graphics & modelling* **2006**, *25* (2), 247-60.
56. Dupradeau, F. Y., et al., The R.E.D. tools: advances in RESP and ESP charge derivation and force field library building. *Physical chemistry chemical physics : PCCP* **2010**, *12* (28), 7821-39.
57. Case, D. A., et al., *Amber 14*. 2014.
58. Jorgensen, W. L., et al., Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, *79* (2), 926-935.
59. Miller, B. R., et al., MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *Journal of Chemical Theory and Computation* **2012**, *8* (9), 3314-3321.
60. Salomon-Ferrer, R., et al., Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J Chem Theory Comput* **2013**, *9* (9), 3878-88.
61. Lavery, R., et al., Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Research* **2009**, *37* (17), 5917-5929.
62. Abuaf, P., et al., Solution conformation of the N-(deoxyguanosin-8-yl)aminofluorene adduct opposite deoxyinosine and deoxyguanosine in DNA by NMR and computational characterization. *Chemical Research in Toxicology* **1995**, *8* (3), 369-378.

63. Rubbi, C. P.; Milner, J., Analysis of nucleotide excision repair by detection of single-stranded DNA transients. *Carcinogenesis* **2001**, 22 (11), 1789-1796.
64. Hess, M. T., et al., Bipartite substrate discrimination by human nucleotide excision repair. *Proceedings of the National Academy of Sciences* **1997**, 94 (13), 6664-6669.
65. Cai, Y., et al., Differential nucleotide excision repair susceptibility of bulky DNA adducts in different sequence contexts: hierarchies of recognition signals. *Journal of Molecular Biology* **2009**, 385 (1), 30-44.
66. Reeves, D. A., et al., Resistance of bulky DNA lesions to nucleotide excision repair can result from extensive aromatic lesion–base stacking interactions. *Nucleic Acids Research* **2011**, 39 (20), 8752-8764.
67. Yang, W., Poor base stacking at DNA lesions may initiate recognition by many repair proteins. *DNA Repair* **2006**, 5 (6), 654-666.
68. Kropachev, K., et al., Structural basis for the recognition of diastereomeric 5',8-cyclo-2'-deoxypurine lesions by the human nucleotide excision repair system. *Nucleic Acids Research* **2014**, 42 (8), 5020-5032.
69. Geacintov, N. E., et al., Thermodynamic and structural factors in the removal of bulky DNA adducts by the nucleotide excision repair machinery. *Biopolymers* **2002**, 65 (3), 202-210.
70. Maillard, O., et al., DNA repair triggered by sensors of helical dynamics. *Trends in Biochemical Sciences* **2007**, 32 (11), 494-499.
71. Isaacs, R. J.; Spielmann, H. P., A model for initial DNA lesion recognition by NER and MMR based on local conformational flexibility. *DNA Repair* **2004**, 3 (5), 455-464.
72. Cai, Y., et al., Nucleotide excision repair efficiencies of bulky carcinogen–DNA adducts are governed by a balance between stabilizing and destabilizing interactions. *Biochemistry* **2012**, 51 (7), 1486-1499.
73. Min, J.-H.; Pavletich, N. P., Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature* **2007**, 449 (7162), 570-575.
74. Velmurugu, Y., et al., Twist-open mechanism of DNA damage recognition by the Rad4/XPC nucleotide excision repair complex. *Proceedings of the National Academy of Sciences* **2016**, 113 (16), E2296-E2305.
75. Kong, M., et al., Single-molecule imaging reveals that rad4 employs a dynamic DNA damage recognition process. *Molecular Cell* **2016**, 64 (2), 376-387.

76. Chen, X., et al., Kinetic gating mechanism of DNA damage recognition by Rad4/XPC. **2015**, 6, 5849.
77. Mu, H., et al., Recognition of damaged DNA for nucleotide excision repair: A correlated motion mechanism with a mismatched cis-syn thymine dimer lesion. *Biochemistry* **2015**, 54 (34), 5263-5267.
78. Cai, Y., et al., Free energy profiles of base flipping in intercalative polycyclic aromatic hydrocarbon-damaged DNA duplexes: energetic and structural relationships to nucleotide excision repair susceptibility. *Chemical Research in Toxicology* **2013**, 26 (7), 1115-1125.
79. Mu, H., et al., Nucleotide excision repair lesion-recognition protein Rad4 captures a pre-flipped partner base in a benzo[a]pyrene-derived DNA lesion: how structure impacts the binding pathway. *Chemical Research in Toxicology* **2017**, 30 (6), 1344-1354.
80. Choi, J.-Y.; Guengerich, F. P., Adduct size limits efficient and error-free bypass across bulky N2-guanine DNA lesions by human DNA polymerase  $\eta$ . *Journal of Molecular Biology* **2005**, 352 (1), 72-90.
81. Zhang, H.; Guengerich, F. P., Effect of N2-guanine modifications on early steps in catalysis of polymerization by *Sulfolobus solfataricus* P2 DNA polymerase Dpo4 T239W. *Journal of Molecular Biology* **2010**, 395 (5), 1007-1018.
82. Zhang, H., et al., Versatility of Y-family *Sulfolobus solfataricus* DNA polymerase Dpo4 in translesion synthesis past bulky N2-alkylguanine adducts. *Journal of Biological Chemistry* **2009**, 284 (6), 3563-3576.
83. Sproviero, M., et al., Enhancing bulge stabilization through linear extension of C8-aryl-guanine adducts to promote polymerase blockage or strand realignment to produce a C:C mismatch. *Chemical Research in Toxicology* **2015**, 28 (8), 1647-1658.
84. Sharma, P., et al., Influence of the linkage type and functional groups in the carcinogenic moiety on the conformational preferences of damaged DNA: structural and energetic characterization of carbon- and oxygen-linked C8-phenolic-guanine adducts. *Chemical Research in Toxicology* **2015**, 28 (4), 782-796.

# Chapter 3: Structure of an Unusual Tetracyclic Deoxyguanosine Adduct: Implications for Frameshift Mutagenicity by *Ortho*-Cyano Nitroanilines<sup>a,b,c</sup>

## 3.1 Introduction

Polycyclic nitroaromatics and arylamines exhibit potent mutagenicity,<sup>1,2</sup> while monocyclic analogs are usually much less potent and are regarded as borderline carcinogens.<sup>3-5</sup> Both undergo bioactivation to afford nitrenium ion electrophiles that react at the C8-site of guanine in DNA to generate covalently *N*-linked C8-2'-deoxyguanosine (C8-dG) lesions.<sup>1-5</sup> However, the polycyclic *N*-linked C8-dG adducts can adopt the promutagenic *syn* conformation, which permits favorable stacking interactions between the planar polycyclic aromatic moiety and the flanking base-pairs, generating an intercalated base-displaced stacked (S) conformer in a fully paired (FP) duplex.<sup>6-8</sup> They also induce a higher frequency of GC deletion mutations in bacterial mutagenesis within the *Escherichia coli* *NarI* endonuclease sequence (5'-CGGCGC-3') containing a GpC repeat.<sup>9-11</sup> Their ability to stabilize a slipped mutagenic intermediate (SMI) that contains the C8-dG adduct and the 5'-flanking C within a two-nucleotide bulge in the template strand provides a mechanism for -2 frameshift mutagenesis.<sup>12-14</sup> In contrast, single-ringed *N*-linked C8-dG adducts typically adopt the *anti* conformation for efficient base pairing with the complementary C to produce the major-groove B-type conformer with minimal impact on

---

<sup>a</sup>The *Chemical Research and Toxicology* Reference Style is used in this Chapter.

<sup>b</sup>Reprinted with permission from Manning, T.W.; Al-Abdul-Wahid, S., Manderville, R.A., Josephy, P.D., Kung, R.W., and Wetmore S.D. (2019) Structure of an Unusual Tetracyclic Deoxyguanosine Adduct: Implications for Frameshift Mutagenicity of *ortho*-Cyano Nitroanilines. *Chem. Res. Toxicol.* 33, 584-593.

<sup>c</sup>R.W.K. performed DFT calculations and MD simulations, while experimental work was done by T.W.M., S.A.A.W., P.D.J. and R.A.M. Writing was done by R.W.K., S.D.W., T.W.M., S.A.A.W., P.D.J. and R.A.M.

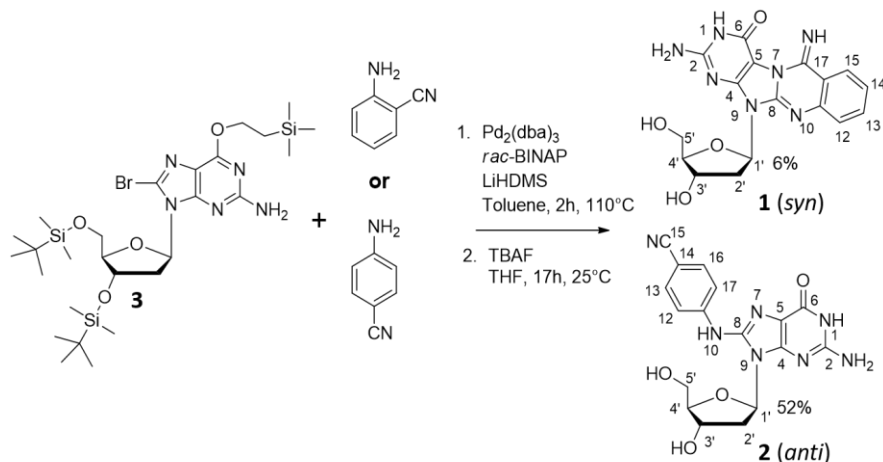


the B-form duplex.<sup>15</sup> Consequently, monocyclic nitroaromatics and arylamines are only weakly mutagenic and generally lack an ability to induce –2 frameshift mutagenesis.<sup>3-5</sup>

Nitroaniline (NA) derivatives containing a cyano substituent are components of azo dyes and have recently been found to display potent mutagenicity in the Ames test using the frameshift-sensitive tester strain TA98.<sup>16,17</sup> The potency of 2-cyano-4-nitroaniline (CNNA) in TA98 is ~ 2,500-fold higher than 4NA, while 2,6-dicyano-4-nitroaniline (CNCNNA) is ~ 49,000-fold higher than 4NA and is a more potent mutagen in TA98 than the polycyclic derivatives nitropyrene and 2-nitrofluorene.<sup>17</sup> Possible mechanisms for the observed “cyano effect” include the impact of the cyano substituent on the structure of the *N*-linked C8-dG lesion.<sup>17</sup> For both CNNA and CNCNNA, metabolic activation to the nitrenium ion would generate an *N*-linked C8-dG adduct bearing an *ortho*-CN substituent. According to Baldwin’s rules of ring closures,<sup>18</sup> it would be favorable for the *N*-linked C8-dG adduct to undergo a cyclization process involving N7 of dG and the *ortho*-CN group of the C8-attached aryl moiety to generate a quinazolinimine ring<sup>19,20</sup> as part of a fused tetracyclic C8,N7-dG adduct structure.<sup>17</sup> Such an unusual polycyclic C8-dG adduct derived from a monocyclic NA precursor may exhibit potent frameshift mutagenicity.

To test the above hypothesis, two model isomeric *N*-linked C8-aniline-dG adducts derived from 2- and 4-aminobenzonitrile were chemically synthesized using Pd-catalyzed Buchwald-Hartwig cross-coupling (Scheme 3.1).<sup>4,21</sup> Our results clearly demonstrate formation of the closed C8,N7-dG cyclic adduct **1** from attachment of 2-aminobenzonitrile, while attachment of the 4-isomer affords the open analog **2**, since it is not feasible for the *para*-CN to undergo cyclization with N7 of the dG moiety. The cyclic analog **1** displays impressively bright blue fluorescence in water, while the open analog **2** is devoid of fluorescent properties. IR and <sup>13</sup>C NMR of adduct **1** confirm formation of the

quinazolinimine ring with conversion of the *ortho*-CN substituent into the exocyclic imine with a downfield NH resonance that suggests intramolecular H-bonding with the O6 carbonyl of the dG moiety. The cyclic adduct **1** is also more resistant to acid-catalyzed deglycosylation than the open isomer **2**. The NMR spectral features of the two C8-aniline-dG adducts also demonstrate a conformational change with the open derivative **2** favoring the *anti* conformation, while the closed polycyclic derivative **1** adopts the promutagenic *syn* conformation. To test the propensity of the adducts to induce frameshift mutagenesis, they were both modeled within the reiterated G3 site of the *NarI* sequence (5'-CTCTCGGCXCCATC, X = C8-aniline-dG), which is a hotspot for -2 frameshift mutations mediated by polycyclic *N*-linked C8-dG adducts.<sup>9-11</sup> MD simulations predict that only the fused tetracyclic C8,N7-dG adduct **1** is able to stabilize the promutagenic S conformer in the FP duplex, while the FP duplex containing the open adduct **2** is more likely to maintain the canonical B-form. Adduct **1** is also better able to stabilize a planar *syn* conformation within a SMI duplex. Overall, our findings provide an important rationale for the observed “cyano effect” in mutagenicity assays performed on nitrobenzonitriles.



Scheme 3.1: Synthesis of closed, cyclic adduct (**1**) and open, para adduct (**2**) from 2-aminobenzonitrile and 4-aminobenzonitrile, respectively.

## 3.2 Experimental Procedures

### 3.2.1 Materials and Methods

All reagents purchased from commercial sources were used without further purification unless otherwise noted. Infrared spectra were recorded using FT-IR-ATR Bruker Alpha equipped with diamond crystal and the data was processed using OPUS 7.5 spectroscopy software. All UV-vis absorbances ( $A$ ) were measured on a Cary 300-Bio UV-Vis spectrophotometer equipped with a  $6 \times 6$  Multicell Block Peltier, stirrer and temperature controller with Probe Series II. Measurements were carried out using 10 mm light path quartz glass cells with a baseline correction. Stock solutions of the adducts were prepared at 5 mM using spectroscopic DMSO, prior to dilution with other solvents for spectrophotometric measurements wherein the concentration of DMSO was assumed to be negligible. Fluorescence measurements were recorded using an Edinburgh Instrument Spectrofluorometer FS5. Both excitation and emission spectra were measured in quartz cells (108.002F-QS) with a path length of 10 mm at 25 °C (298K) and both excitation and emission slit widths were kept constant at 3 nm. The quantum yield was calculated using the comparative method with quinine bisulfite in 0.5 M  $H_2SO_4$  ( $\Phi_f = 0.54$ ) serving as fluorescent standard. Mass spectra were acquired on either a quadrupole ion trap or on a Q-TOF high-resolution (HRMS) instrument, both using an electrospray ionization source. Acid-catalyzed kinetics of hydrolysis for adducts **1** and **2** were obtained spectrophotometrically, using procedures previously reported for O-linked<sup>22</sup> and C-linked C8-dG adducts.<sup>23</sup> Reactions were initiated by addition of 10  $\mu$ L of DMSO stock solutions of **1** and **2** to 1990  $\mu$ L of either 0.1 M HCl or 0.2 M HCl/KCl buffer, pH 2, for a final adduct concentration of 25  $\mu$ M. Spectral measurements were taken at various time intervals at 37.2°C (for 0.1 M HCl) and 48.4 °C (for 0.2 M HCl/KCl buffer) until no further change

was observed in the absorbance spectrum indicating that deglycosylation was complete. Formation of the deglycosylated product was then verified using mass spectrometry. NMR spectra were recorded on Bruker Avance DPX 300, 400, or 600 MHz spectrometers in DMSO-d<sub>6</sub> at 25 °C. One-dimensional (1D) <sup>1</sup>H experiments were typically performed with 16 scans and a relaxation delay of 10 seconds between scans, for a total acquisition time of 3.3 minutes, while one-dimensional <sup>13</sup>C experiments were typically performed with 256 scans and a relaxation delay of 5 seconds between scans, for a total acquisition time of 26 minutes. <sup>1</sup>H-<sup>13</sup>C HSQC experiments were collected with 384 increments spanning 165 ppm in the indirect dimension, a relaxation delay of 1.5 seconds between scans, and 2 scans per increment, for a total acquisition time of ~ 20 minutes. <sup>1</sup>H-<sup>13</sup>C HMBC experiments were collected with 256 increments spanning 220 ppm in the indirect dimension, a relaxation delay of 1.5 seconds between scans, and 4 scans per increment, for a total acquisition time of 30 minutes. The <sup>1</sup>H-<sup>13</sup>C HMBC experiments were optimized for a long-range coupling constant of 8 Hz, while a one-bond filter was used to attenuate one-bond J-couplings between 130-180 Hz. ROESY NMR spectra were collected on a Bruker AVANCE III 600 MHz spectrometer equipped with a 5 mm “TXI” probe. The sample temperature was regulated at 298 ± 1 K. 1D selective ROESY spectra were acquired using the *selrogp* sequence provided by Bruker using a mixing time of 200 ms

### 3.2.2 Synthesis

8-Bromo-3',5'-O-bis(*tert*-butyldimethylsilyl)-O<sup>6</sup>-(trimethylsilylethyl)-2'-deoxy-guanosine (**3**, Scheme 3.1) was prepared as previously reported.<sup>22</sup> Compound **3** (613 mg, 0.91 mmol), 2-aminobenzonitrile or 4-aminobenzonitrile (215 mg, 1.82 mmol), Pd<sub>2</sub>dba<sub>3</sub> (167 mg, 0.182 mmol), and *rac*-2,2'-bis(diphenylphosphino)-1,1'-binaphthyl (BINAP, 340

mg, 0.546 mmol) were placed in a 50 mL oven-dried round bottom flask fitted with a condenser. The flask was evacuated and purged with argon several times. Dry, degassed toluene (5 mL) was added to the flask and the reaction vessel was heated to 80°C. After the mixture was stirred for 20 minutes, lithium bis(trimethylsilyl)amide (1.0 M in THF, 1.82 mL, 1.82 mmol) was added and the vessel was heated to 110°C. The reaction was stirred for 2 hours and then cooled to room temperature. The crude mixture was diluted with dichloromethane and filtered through Celite. The filtrate was dried over sodium sulphate, filtered and concentrated *in vacuo* and subjected to silica flash chromatography, eluting in dichloromethane with 10% methanol (adduct **1**) and hexane with 20% ethyl acetate (adduct **2**). Crude fractions were collected, concentrated and redissolved in anhydrous THF (5 ml). Tetrabutylammomium fluoride (880 uL, 0.88 mmol) was added to the reaction and the mixture was stirred for 17 hours at room temperature under an argon atmosphere. The mixture was then concentrated and diluted with 30 mL of methanol before being placed in the freezer at -20°C. After 3 hours, a precipitate had formed. The suspension was centrifuged, and the pellet was washed twice with dichloromethane and thrice with methanol. The pellet was dried *in vacuo* to yield 20 mg of compound **1** as a light green powder (6% yield) and 185 mg of compound **2** as a white powder (52% yield, see Appendix B Figures B.1-B.4 and B.6-B.10 for NMR spectra).

**Adduct 1.** <sup>1</sup>H NMR (300 MHz, DMSO-d<sub>6</sub>): δ = 15.51 (bs, 1H, NH), 10.12 (bs, 1H, N1H), 8.38 (d, *J* = 8.34 Hz, 1H, H15) 7.84, (t, *J* = 7.51 Hz, 1H, H13), 7.62 (d, *J* = 8.19 Hz, 1H, H12), 7.48 (t, *J* = 7.16 Hz, 1H, H14), 6.65 (m, 1H, H1'), 6.15 (bs, 2H, NH<sub>2</sub>), 5.40 (m, 1H, 5'OH), 5.27 (d, *J* = 3.79 Hz, 1H, 3'OH), 4.53 (m, 1H, H3'), 3.93 (m, 1H, H4'), 3.76, 3.60 (m, 2H, H5',5''), 3.22 (m, 1H, H2'), 2.08 (m, 1H, H2''). <sup>13</sup>C NMR (150 MHz, DMSO-

d<sub>6</sub>): signals missing for C2 and C6 due to line broadening,  $\delta$  = 152.2 (C4), 151.3 (C13), 146.2 (C11), 141.8 (C8), 135.0 (C13), 126.4 (C12), 125.2 (C15), 124.7 (C14), 99.6 (C5), 88.6 (C4'), 83.2 (C1'), 72.1 (C3'), 63.0 (C5'), 36.6 (C2'). HRMS (ESI/Q-TOF) calcd for C<sub>17</sub>H<sub>18</sub>N<sub>7</sub>O<sub>4</sub><sup>+</sup> [M+H<sup>+</sup>] 384.1420, found 384.1415.

**Adduct 2:** <sup>1</sup>H NMR (600 MHz, DMSO-d<sub>6</sub>):  $\delta$  = 10.60 (s, 1H, N1H), 9.17 (s, 1H, N10H), 7.90 (d,  $J$  = 8.92 Hz, 2H, H12,17), 7.72 (d,  $J$  = 8.99 Hz, 2H, H13,16), 6.42 (bs, 2H, NH<sub>2</sub>), 6.33 (dd  $J$  = 9.51, 5.81 Hz, 1H, H1'), 6.02 (bs, 1H, 5'OH), 5.35 (d,  $J$  = 3.66 Hz, 1H, 3'OH), 4.42 (m, 1H, H3'), 3.94 (m, 1H, H4'), 3.76 (m, 2H, H5',5''), 2.50 (m, 1H, H2'), 2.05 (m, 1H, H2''). <sup>13</sup>C NMR (100 MHz, DMSO-d<sub>6</sub>):  $\delta$  = 156.2, 153.6, 150.0 (C4), 145.3 (C11), 142.2 (C8), 133.6 (C13,16), 120.1 (C15), 117.6 (C12,17), 112.7 (C5), 102.2 (C14), 87.7 (C4'), 83.4 (C1'), 71.7 (C3'), 61.8 (C5'), 39.2 (C2'). HRMS (ESI/Q-TOF) calcd for C<sub>17</sub>H<sub>18</sub>N<sub>7</sub>O<sub>4</sub><sup>+</sup> [M+H<sup>+</sup>] 384.1420, found 384.1417.

### 3.3 Computational Details

#### 3.3.1 DFT Calculations

Nucleobase models of the adducts derived from 2-aminobenzonitrile and 4-aminobenzonitrile, **1** and **2**, respectively, were built using GaussView 6.0.<sup>24</sup> The structures were optimized with B3LYP-D3(BJ)/6-31G(d,p) and scaled (0.955) zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311++G(2df,2p) energies were obtained. Subsequently, nucleoside models were built by replacing the N9 hydrogen in the global minima for the nucleobase models with 2'-deoxyribose. A C2'-endo puckered sugar was used based on a previously published conformational analysis.<sup>25</sup> Scans were performed for rotation about the  $\chi$  glycosidic dihedral angle ( $\angle(\text{O4}'\text{--C1}'\text{--N9--C4})$ ) from  $\chi = 0^\circ$  to  $350^\circ$  in

10° increments using constrained B3LYP-D3(BJ)/6-31G(d,p) optimizations. Nucleoside scans were repeated with the 5' and 3'-hydroxy groups constrained such that  $\angle(\text{H}-\text{O}5'-\text{C}5'-\text{C}4') = 180^\circ$  and  $\angle(\text{H}-\text{O}3'-\text{C}3'-\text{H}3') = 60^\circ$ , which corresponds to the hydroxy groups interacting with solvent rather than intramolecular hydrogen bonding with the nucleobase. Nucleoside minima and transition states (TS) identified from the scans were fully optimized using B3LYP-D3(BJ)/6-31G(d,p) and relative energies determined with B3LYP-D3(BJ)/6-311++G(2df,2p), which include scaled B3LYP-D3(BJ)/6-31G(d,p) zero-point vibrational energy corrections. Single-point calculations were also performed in water using the IEF-PCM implicit model. All DFT calculations were performed using Gaussian 16.<sup>26</sup>

### 3.3.2 MD Simulations

To gain greater insight into the conformational space of the free nucleosides in water, solvated models of the **1** and **2** nucleosides were built based on the DFT optimized structures. Furthermore, initial structures of FP duplex and SMI complex models containing adducts **1** and **2** were generated from those for similar C8-dG adducts<sup>27</sup> using GaussView 6.0. The adducts were placed at the G3 (X) position in the *NarI* containing sequence 5'-CTCTCGGCXCCATC. The complementary sequence was 3'-GAGAGCCGCGGTAG for the FP duplex model and 3'-GAGAGCCGCGGTAG for the SMI complex. Within the FP systems, the damaged DNA was modelled in the B, W and S conformations that have been reported for other C8-dG aromatic-amine derived lesions.<sup>28,29</sup> Within the SMI complex, simulations were performed with the damaged nucleotide in the *anti* and *syn* orientations. Thus, in total, 10 modified DNA systems were considered, as well as two free nucleoside models. Natural DNA was described using the parmbsc1 force field within Amber18, while

the description of the modified bases was supplemented using GAFF parameters.<sup>30,31</sup> The partial charges for the adducts were calculated at the HF/6-31G(d) level within the RESP charge fitting scheme using the R.E.D. (Version III.4) software package.<sup>32</sup> Information about the resulting atom types and partial charges can be found in the Supporting Information.

Each system was prepared by neutralizing the solute with sodium cations and solvating the solute in a TIP3P water box using the tleap module in Amber 18, such that there was a minimum distance of 10 Å between the solute and the edge of the water box in each direction. Each system was minimized over four stages that consisted of 2500 steps of steepest decent proceeded by 2500 steps of conjugate gradient. In the first three stages, a 100 kcal/(mol Å<sup>2</sup>) restraint was added to the solute atoms, then the heavy atoms and then the solute atoms, respectively. The heating process was similarly divided into six 20 ps stages during which the temperature was raised from 10 to 310 K, each step increasing the temperature by 50 K. During heating, the system was treated as constant volume, with the solute restrained by 25 kcal/(mol Å<sup>2</sup>). After heating, the systems were equilibrated over five 20 ps stages during which the restraint force on the solute was decreased from 20 to 15, 10, 5 and finally 1.5 kcal/(mol Å<sup>2</sup>).

For the free nucleosides, umbrella sampling was performed to assess the rotational barrier about the glycosidic bond ( $\chi$  dihedral angle). Specifically, simulations were completed on sampling windows generated every 8° between 0° and 360°. After the equilibration steps described above, each system was further equilibrated for 6 ns with a 30 kcal/(mol rad<sup>2</sup>) restraint placed on the atoms defining the  $\chi$  torsion angle (O4', C1', N9, and C4). Each window was then simulated for 15 ns with the same restraint on the  $\chi$  torsion



angle and data was collected every 0.1 ps, which resulted in adequate sampling between the windows (Figure B.24, Appendix B). The potential of mean force (PMF) was generated for rotation about the glycosidic bond using the weighted histogram analysis method (WHAM) and a tolerance of 0.01 kcal/mol.<sup>33</sup>

For each of the 10 unique modified DNA systems, five MD production replicas were generated using the same initial structure and different initial velocities. Each replica was run for 1  $\mu$ s, resulting in a total of 50  $\mu$ s MD production simulation time. Each replica was analyzed every 100 ps using the cpptraj module within Amber 18. The backbone root mean square deviation (RMSD) was calculated with respect to the initial structure of the MD production simulation (Table B.2, Appendix B). Key dihedral angles of the lesion were monitored, including  $\chi$ , as well as  $\theta$  ( $\angle(\text{N9-C8-N10-C11})$ ) and  $\phi$  ( $\angle(\text{C8-N10-C11-C12})$ ) for **2**, which determine the position of the bulky group with respect to the base (Table B.3, Appendix B). Additionally, the H-bonding occupancies for the lesion site and flanking base pairs were determined using a heavy-atom distance cut-off of 3.4 Å and angle cut-off of 120°. For each system, the trajectories were clustered based on the atoms in the adduct and flanking base pairs using the hieragglo algorithm to obtain representative structures. The interplanar angle was calculated between the ring atoms of the neighboring base pair 5' (i.e., G7 and its pairing base) and 3' (i.e., C10 and its pairing base) with respect to the two unpaired bases (i.e., the adducted G and its 5' neighboring C) for the SMI structures in the *syn* orientation using an in-house script. The interplanar angle was also determined between each flanking base pair and the aromatic ring atoms of the lesion for the FP conformations containing the *syn* adduct conformations.

### 3.4 Results and Discussion

#### 3.4.1 Spectral Properties and Hydrolytic Stability

Comparative analysis of the spectra of **1** and **2** provided diagnostic information to support the structure of the cyclic adduct **1** (Figures 3.1-3.3). Firstly, nitrile groups produce a prominent infrared (IR) stretch at  $\sim 2250\text{ cm}^{-1}$ .<sup>34</sup> This signature band was present in the IR spectrum of **2** (red trace, Figure 3.1), but was absent for **1** (black trace, Figure 3.1), which was consistent with the proposed cyclic structure.  $^1\text{H-NMR}$  evidence also supports formation of the cyclic “closed” state of **1**. Typically, *N*-linked dG adducts such as **2** produce a singlet between  $\delta$  7.5 - 9.5 ppm in  $^1\text{H-NMR}$  spectra, due to the proton on the C8-linked  $sp^3$ -hybridized nitrogen (C8-NH, red trace for **2**, Figure 3.2).<sup>4,5</sup> This singlet was absent from the  $^1\text{H-NMR}$  of **1** (black trace, Figure 3.2) and instead a broad singlet at  $\sim \delta$  15.5 ppm was observed, which was assigned to the exocyclic imino proton produced from cyclization (Figure B.1, Appendix B). Its downfield shift was ascribed to its participation in an intramolecular H-bond with O6 of the dG moiety.

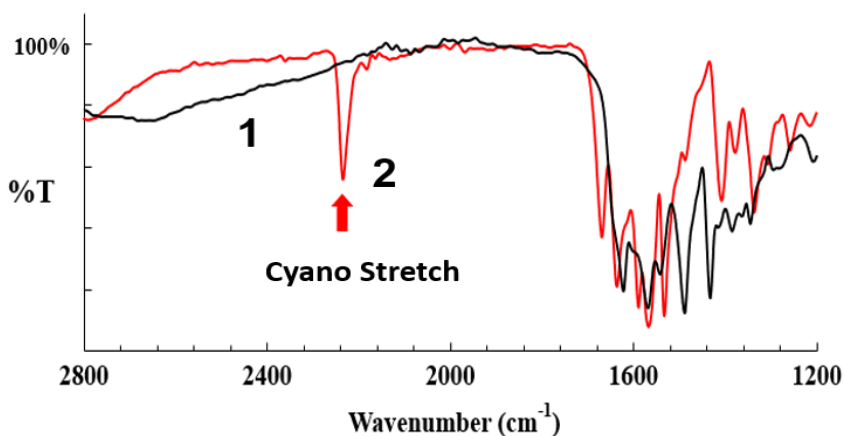


Figure 3.1: IR spectral overlay traces of **1** and **2** depicting the CN stretch for adduct **2**.

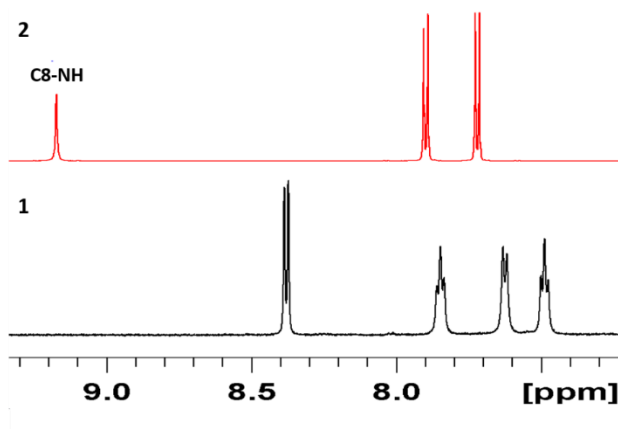


Figure 3.2:  $^1\text{H}$  NMR spectra of **1** and **2** displaying aromatic peaks and C8-NH for **2**.

The optical character of adducts **1** and **2** also provided additional reinforcement of adduct structure. The *para* adduct **2** produced a single broad absorption peak at  $\lambda_{\text{max}} = 312$  nm (red trace, Figure 3.3A), while the cyclic adduct **1** generated two absorption peaks with a sharp absorbance at  $\lambda_{\text{max}} = 290$  nm and a broad red-shifted absorbance at  $\lambda_{\text{max}} = 369$  nm (black trace, Figure 3.3A). This difference in absorbance was also accompanied by a significant difference in fluorescent character. Adduct **2** did not produce any appreciable fluorescence in water or  $\text{CH}_3\text{CN}$ , while the cyclic adduct **1** displayed pronounced blue emission with  $\lambda_{\text{em}} = 457$  nm following excitation at 369, or 290 nm, with a quantum yield ( $\Phi_f$ ) of 0.68 in water (Figures 3.3B/C). The fluorescence of **1** was not influenced by solvent polarity, showing similar emission intensity in water,  $\text{CH}_3\text{CN}$  or  $\text{CH}_2\text{Cl}_2$ . Cyclization would produce a rigid, planar structure devoid of free rotation, which would explain the solvent-independent fluorescence of **1**.

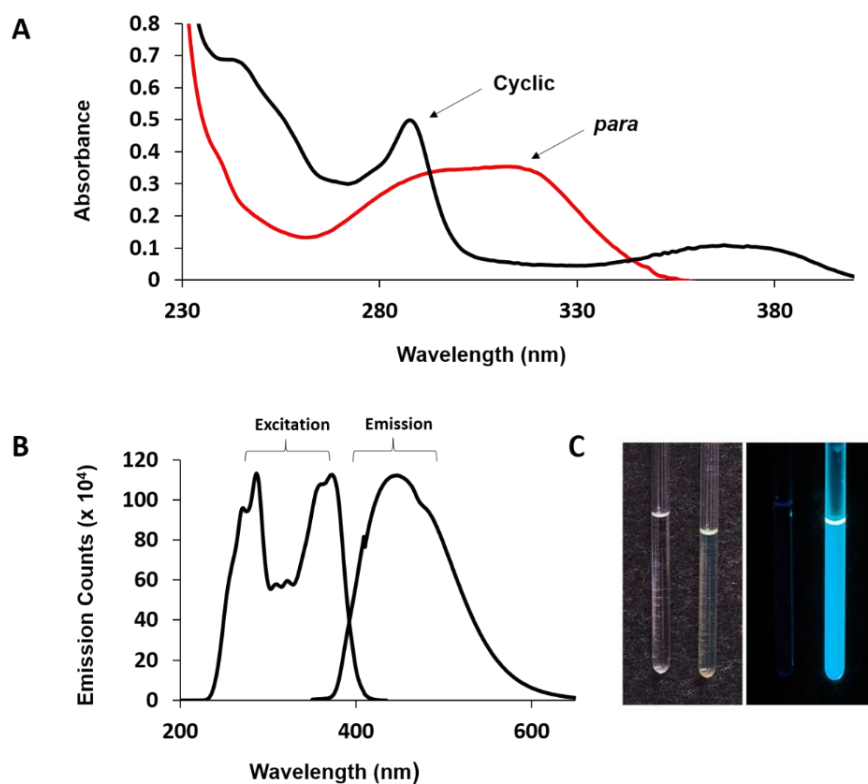


Figure 3.3: (A) UV spectral overlay traces of **1** (black trace) and **2** (red trace). (B) Excitation and emission spectra of **1** in water. (C) Picture highlighting the blue fluorescence of **1** (right) and lack of fluorescence for adduct **2** (left).

Table 3.1: Diagnostic <sup>13</sup>C-NMR Chemical Shifts in DMSO.

| $\delta$ (ppm) <sup>a</sup>  | C5    | C8    | Imino | CN    | C2'  |
|------------------------------|-------|-------|-------|-------|------|
| <b>dG</b>                    | 116.8 | 136.0 | —     | —     | 39.5 |
| <b>8-oxo-dG</b> <sup>b</sup> | 98.5  | 151.6 | —     | —     | 35.6 |
| <b>1</b>                     | 99.7  | 141.8 | 151.4 | —     | 36.6 |
| <b>2</b>                     | 112.8 | 142.3 | —     | 120.1 | 39.2 |

<sup>a</sup>Chemical shifts are given in ppm measured at 150 MHz at ambient temperature in DMSO-d<sub>6</sub>. <sup>b</sup>Chemical shifts for 8-oxo-dG taken from Cho.<sup>36</sup>

Diagnostic <sup>13</sup>C-NMR chemical shifts of **1** and **2** (Table 3.1) were obtained through <sup>1</sup>H-<sup>13</sup>C correlations (Figures B.3, B.4, B.8, B.9, Appendix B) and confirmed covalent bond

formation between the *ortho*-CN substituent and N7 of the dG nucleobase to afford **1**. The chemical shifts of C5 reveal a change in N7-hybridization. The unmodified nucleobase dG contains an sp<sup>2</sup>-hybridized N7 and the adjacent C5 atom resonates at 116.8 ppm.<sup>35</sup> *N*-Linked adducts such as **2** do not induce a change in N7-hybridization and the C5 chemical shift of **2** (112.8 ppm) was similar to dG. In contrast, the oxidized 8-oxo-dG lesion adopts the keto-tautomer, which contains an sp<sup>3</sup>-hybridized N7. The C5 atom of this oxidized lesion resonates upfield at 98.5 ppm.<sup>36</sup> Adduct cyclization to produce **1** would also induce an sp<sup>3</sup>-hybridized N7, and its C5 chemical shift (99.7 ppm) was similar to the corresponding chemical shift within 8-oxo-dG. The downfield shift of C8 in adducts **1** (141.8), **2** (142.3) and 8-oxo-dG (151.6) relative to unmodified dG (136.0) was representative of adduction at C8. Cyclization was also evidenced by the diagnostic <sup>13</sup>C-NMR peaks associated with the nitrile and imino groups. Aromatic nitriles possess a <sup>13</sup>C chemical shift ~120 ppm,<sup>37</sup> which was observed for adduct **2** (120.1 ppm). In adduct **1** the cyano group is converted into an exocyclic imine,<sup>19</sup> which resonates downfield at 151.4 ppm.

Rates of acid-catalyzed hydrolysis of adducts **1** and **2** were determined using a spectrophotometric procedure, as previously carried out for *C*-linked,<sup>23</sup> *O*-linked<sup>22</sup> and other *N*-linked C8-dG adducts.<sup>38</sup> For **2**, the deglycosylated nucleobase displays greater absorbance at ~ 320 nm in 0.1 M HCl at 37 °C (Figure 3.4A) than the starting nucleoside, while for **1** the greatest change in absorbance occurred at ~ 250 nm, with the deglycosylated nucleobase exhibiting a decrease in absorbance (insert, Figure 3.4B). That the observed spectral changes were due to acid-catalyzed hydrolysis of the starting nucleosides was confirmed by ESI<sup>+</sup>-MS (Figures B.12-B.15, Appendix B). The changes in absorbance as a function of time were utilized to determine apparent first-order rate constants ( $k_{\text{obs}}$  (min<sup>-1</sup>))

and half-lives ( $t_{1/2}$  (min)) for **1** and **2** in 0.1 M HCl at 37.2 °C and in 0.2 M HCl/KCl pH 2.0 buffer at 48.4 °C, in order to draw direct comparisons to the corresponding rates for unmodified dG (Table 3.2).<sup>39</sup> In acidic media, the first step of dG hydrolysis involves protonation at N7 (N7H<sup>+</sup>-dG,  $pK_a \sim 2.34$ <sup>40</sup>) followed by rate-limiting unimolecular cleavage of the glycosidic bond. In general, C8-dG adducts are more reactive than dG, with C-linked and O-linked adducts being ~ 4- to 45-fold more reactive than dG under acidic conditions,<sup>22,23</sup> while N-linked adducts are only 2- to 5-fold more reactive.<sup>38</sup> Relief of steric strain, stabilization of the developing negative charge at N9 upon removal of the sugar moiety, and the influence of the C8-substituent on the proton affinity of N7, influence the relative rates of depurination for C8-dG adducts compared to dG.<sup>22,23,38</sup> In the present case, **1** and **2** were ~ 5-fold more reactive than dG in 0.1 M HCl at 37.2 °C, as expected for N-linked C8-dG adducts.<sup>38</sup> However, at pH 2 (48.4 °C), the cyclic adduct **1** was more stable to hydrolysis than dG, while the open analog **2** was still more reactive (~ 2-fold, Table 3.2). Thus, the open adduct **2** is 2- to 5-fold more reactive than dG to hydrolysis in acidic media, as reported previously for other N-linked C8-dG adducts,<sup>38</sup> while the cyclic adduct **1** is more resistant to hydrolysis than both dG and **1** at pH 2.0. Clearly, the involvement of N7 in formation of the quinazolinimine ring of **1** changes the anticipated site of protonation to the imine N-atoms of the attached C8-aryl moiety, which likely plays a key role in the hydrolytic stability of **1** in pH 2 buffer compared to dG and **2**.

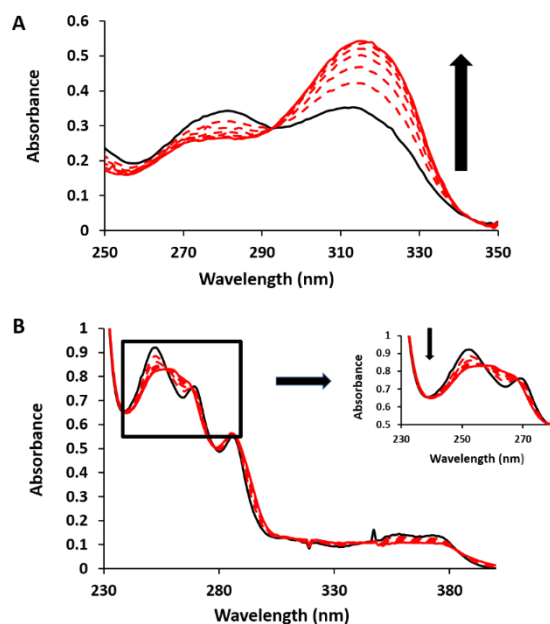


Figure 3.4: Changes in absorbance spectra of (A) adduct **2** and (B) adduct **1** in 0.1 M HCl at 37 °C recorded in ~ 1 min intervals (dashed red traces, initial trace in black).

Table 3.2: Summary of Rate Data for Acid-Catalyzed Hydrolysis of **1** and **2**.

| Adduct          | $k_{\text{obs}}$ (min <sup>-1</sup> ), $t_{1/2}$ (min) <sup>a</sup> | $k/k(\text{dG})$ | $k_{\text{obs}}$ (min <sup>-1</sup> ), $t_{1/2}$ (min) <sup>b</sup> | $k/k(\text{dG})$ |
|-----------------|---------------------------------------------------------------------|------------------|---------------------------------------------------------------------|------------------|
| <b>1</b>        | 0.172 ± 0.005, 4.03                                                 | 4.4              | 0.029 ± 0.001, 23.9                                                 | 0.83             |
| <b>2</b>        | 0.214 ± 0.005, 3.24                                                 | 5.5              | 0.062 ± 0.003, 11.2                                                 | 1.8              |
| dG <sup>c</sup> | 0.0391, 17.7                                                        | 1                | 0.035, 19.7                                                         | 1                |

<sup>a</sup>Determined in 0.1 M HCl at 37.2 °C. <sup>b</sup>Determined in 0.2 M HCl/KCl pH 2.0 buffer at 48.4 °C. <sup>c</sup>Data for dG taken from ref. 39.

### 3.4.2 Conformation

The mutagenic potential of a C8-dG adduct is associated with its ability to adopt the promutagenic *syn* conformation.<sup>2,6</sup> Discrimination of the *syn* and *anti* conformations through <sup>13</sup>C- and <sup>1</sup>H-NMR chemical shifts have been reported in the literature.<sup>35,36</sup> A switch from *anti* to *syn* is correlated with a downfield shift of H<sub>2</sub>' from ~ 2.5 to 3.2 ppm and an

upfield shift of C2' from  $\sim 39.0$  to  $36.0$  ppm. These signature shifts are displayed in the  $^1\text{H}$ -NMR and  $^{13}\text{C}$ -NMR spectra (Figures B.1-B.4, B.6-B.9, Appendix B) of **1** and **2**, and indicate that **1** preferentially adopts the *syn* conformation with H<sub>2'</sub> and C2' resonating at 3.19–3.26 ppm and 36.6 ppm (Table 3.1), respectively. Conversely, **2** adopts an *anti* conformation with H<sub>2'</sub> and C2' resonating at 2.47–2.52 ppm and 39.2 ppm.

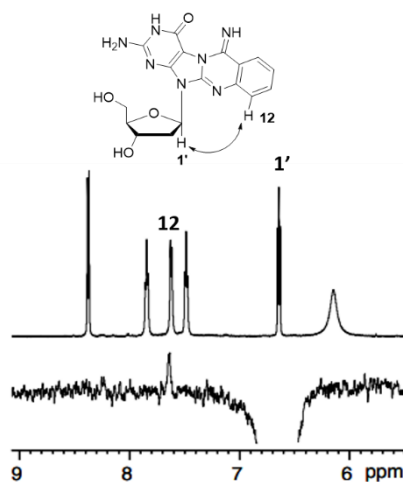


Figure 3.5: Selective 1D ROESY spectrum of **1** in DMSO- $\text{d}_6$  at ambient temperature highlighting the cross-peak from the sugar H1' to the aromatic H12 proton for a *syn* conformational preference.

Further evidence of adduct conformation was obtained by collecting 1D selective ROESY spectra. The cyclic adduct **1** displays a diagnostic ROESY cross-peak from the sugar H1' to the aromatic H12 (Figure 3.5), which is consistent with the chemical shift data for the *syn* preference. In contrast, the open adduct **2** exhibits a ROESY cross-peak from the sugar H5' to the aromatic H12 proton (Figure B.10, Appendix B), which is consistent with an *anti* preference.



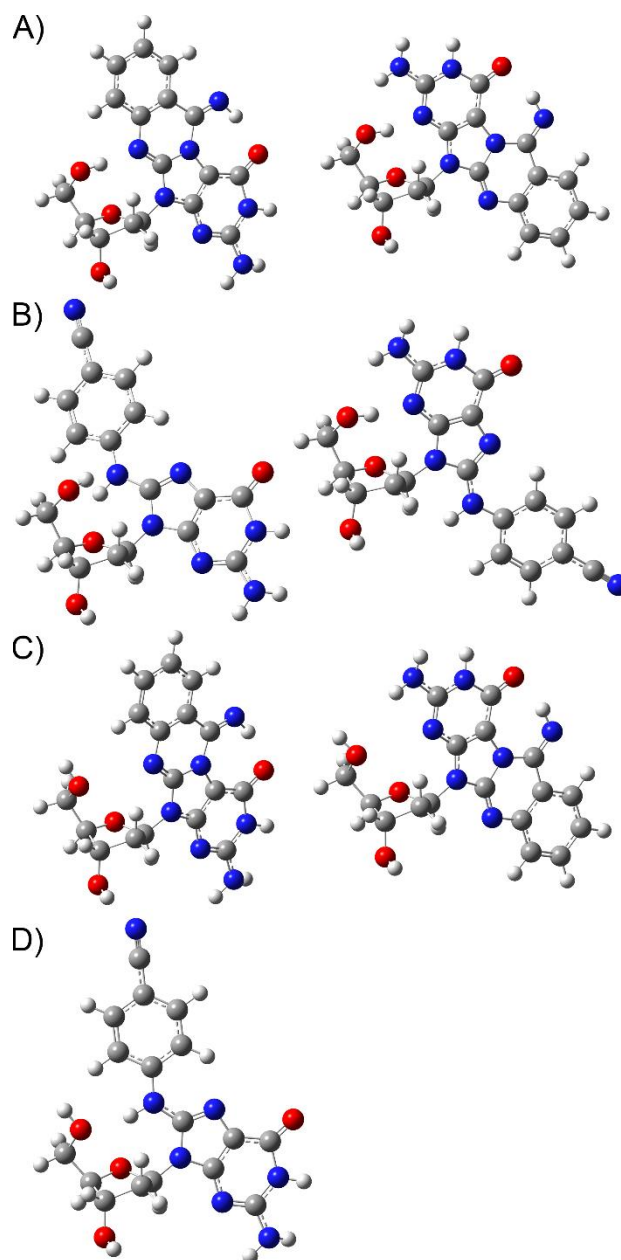


Figure 3.6: B3LYP-D3(BJ)/6-31G(d,p) *anti* (left) and *syn* (right) minima for the free nucleosides of A) **1** with no constraints, B) **2** with no constraints, C) **1** with the 5'-OH constrained, and D) **2** with the 5'-OH unconstrained.

Additional insight into adduct structure and conformation was obtained through DFT calculations. The two nucleosides were built from minima obtained at the nucleobase level. Notably, within the nucleobase model for **1**, the energetically favored orientation contains a N–H $\cdots$ O6 intramolecular hydrogen bond (Figure B.16, Appendix B), which is

consistently maintained within all DFT calculations and MD simulations (occupancy > 90 %). For **2**, we observed that all three minima obtained were within 3 kJ/mol, with the most stable structure positioning the bulky group and nucleobase in the same plane (Figure B.17, Appendix B). In contrast to the NMR data, gas-phase DFT calculations that scan about the  $\chi$  dihedral angle in the isolated nucleosides suggest that **2** is more stable in the *syn* orientation ( $\chi = 57.0^\circ$ ) by 13.2 kJ/mol, while the *anti* orientation of **1** ( $241.2^\circ$ ) is slightly stabilized (by 1.0 kJ/mol, Figures 3.6, B.18-B.21, Appendix B). However, these models likely overestimate the stability provided by a O5'-H $\cdots$ N sugar-nucleobase hydrogen bond between the terminal 5'-hydroxy group of the sugar and the nucleobase adduct. Indeed, when the 5'-hydroxy group is constrained away from the nucleobase to mimic stronger sugar-solvent interactions, the *syn* orientation ( $253.9^\circ$ ) is slightly more stable than *anti* ( $56.6^\circ$ ) for **1** (by 0.8 kJ/mol) and completely destabilized for **2** (no *syn* minimum isolated, and *anti* occurs with  $\chi = 225.3^\circ$ ; see Figures B.19, B.21-B.23, Appendix B). Implicit water further stabilizes *syn* **1** (from 0.8 kJ/mol to 2.1 kJ/mol, Table B.1, Appendix B). To further assess the impact of solvent-nucleoside interactions on the preferential stabilization of the *syn* orientation of the lesions, umbrella sampling was used to characterize the conformational space of the isolated nucleoside adducts in a water box (Figure B.24, Appendix B). In explicit solvent, both nucleosides **1** and **2** have *anti* and *syn* minima at  $\chi \sim 220^\circ$  and  $\sim 50^\circ$ , respectively (Figure 3.7), and adopt a range of sugar puckerings (Table B.2, Appendix B). For **1**, the *syn* orientation is more stable than the *anti* conformation by 1.5 kJ/mol, whereas the *anti* orientation is more stable for **2** by 0.6 kJ/mol. This trend also agrees with the NMR data, although the small relative energies and low rotational energy barriers (< 15 kJ/mol) suggest conformational flexibility.

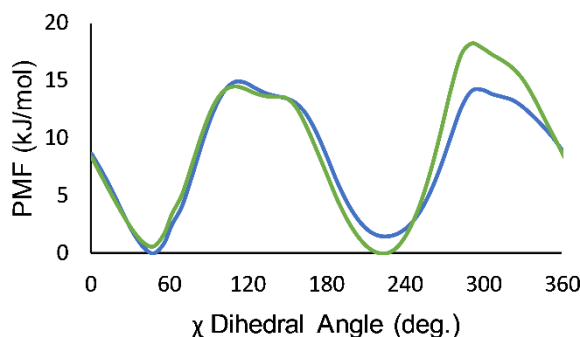


Figure 3.7: PMF for the rotation about the glycosidic bond ( $\chi = \angle(\text{O4}'\text{--C1}'\text{--N9--C4})$ ) for the free nucleoside adduct **1** (blue) and **2** (green) in water.

### 3.4.3 MD Simulations of Adducted *NarI*

The NMR predicted and computationally supported *syn* preference of the adduct **1** nucleoside has important implications for mutagenicity as the lesion may stabilize the SMI complex to promote frameshift mutations. To investigate this possibility, *anti* and *syn* adducts **1** and **2** were modelled in the reiterated G3 (X) site of the *NarI* sequence (5'-CTCTCG1G2CXCCATC, X = **1** or **2**) within SMI and FP duplexes. When each *anti* lesion is modelled in the SMI complexes, the structural differences between the duplexes are minimal, although both lesions adopt a range of sugar puckers (Tables B.3-B.5, Appendix B). Specifically, both lesions position the bulky moiety in the major groove and the Watson-Crick (WC) face of the adducted G is directed towards the opposing strand (Figures B.25 and B.26, Appendix B). In fact, the adduct forms WC hydrogen bonds with the 5'-flanking C in the opposing strand, causing the identity of the unpaired base to shift to the neighboring or next-nearest neighboring 5'-G (Tables B.6 and B.7, Appendix B). Notably, the sequence will therefore have a significant impact on the location of the unpaired G, as the damaged G3 base can pair with the C originally opposite G2 in the sequence considered

in the present work, which in turn can pair with the C across from G1 (leaving G1 unpaired). Therefore, either G1, G2, or the adducted G3 can be unpaired as there are only two Cs in the pairing strand. However, periodically during the simulations, a single C also simultaneously interacts with two Gs (see, for example, Figure B.27, Appendix B). Regardless of the specific unpaired G, either the unpaired *anti* lesion or the other unpaired G is positioned to stack in the helix, which can lead to narrowing of the DNA helix. The resulting helical narrowing disrupts multiple base pairs 5' to the lesion (Figure B.28, Appendix B). Overall, the distorted structures for the SMI containing *anti* **1** and **2** suggest that the *anti* adduct orientations destabilize the complex.

In contrast to the *anti* lesion orientation, MD simulations indicate that the SMI duplexes containing the *syn* adducts adopt two conformations that can be classified based on the interplanar angle between the base pairs flanking the unpaired bases as bent ( $> 40^\circ$ ) or planar ( $< 40^\circ$ ) (Figures 3.8, B.29-B.31, Appendix B). Stacking between the adduct and surrounding nucleobases is a significant stabilization factor for both conformations due the lack of hydrogen bonding between the lesion and surrounding bases (Table B.8, Appendix B). Nevertheless, the base pairs flanking the lesion site are maintained (Table B.9, Appendix B). Although both lesions adopt both conformations, adduct **1** is better able to stabilize the planar conformation than the bent, with the average lifetime of each state before switching conformations (i.e., the residence time (RT)) being larger for the planar conformation (596 ps compared to 468 ps; Tables B.10 and B.11, and Figure B.32, Appendix B). In contrast, adduct **2** favors the distorted conformation (RT = 406 ps) over the planar conformation (RT = 306 ps; Tables B.10, B.11 and Figure B.33, Appendix B). In addition to differences in the identity of the most stable conformation, **1** has overall

larger residence times compared to **2**, implying that **2** is less able to stabilize the SMI. The greater ability of the closed adduct **1** to stabilize the SMI in the *syn* orientation likely arises due to more stable  $\pi$ – $\pi$  stacking interactions between the lesion bulky moiety and flanking base pairs in the preferred planar conformation.

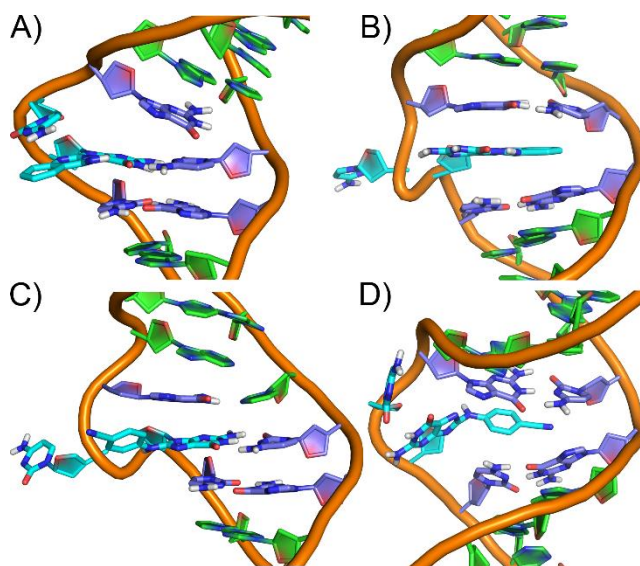


Figure 3.8: Representative SMI structures. A) Distorted SMI containing **1** in the *anti* conformation, B) planar, stabilized SMI containing **1** in the *syn* conformation, C) distorted SMI containing **2** in the *anti* conformation, and D) distorted SMI containing **2** in the *syn* conformation.

As seen for the SMI complexes, the structures of the FP helices are similar when either adduct **1** or **2** adopts the *anti* conformation, which places the bulky moiety in the major groove, maintains the structural features of canonical B-DNA, and allows the lesion site sugar to adopt a range of conformations (Table B.5, Appendix B). Indeed, both lesions maintain WC base pairing with the opposing C, with minimal helical distortions (Figures 3.9, B.34, B.35, and Tables B.12 and B.13, Appendix B). However, when the lesions are in the *syn* orientation, the FP duplexes adopt either a base-displaced stacked (S) or wedge (W) conformation (Figures 3.9, B.36-B.39, Appendix B). Specifically, the S conformation

displaces the adducted G and pairing C bases towards the major groove such that no hydrogen bonds exist between these bases. Instead, the S conformation is stabilized by stacking between the adduct and the flanking base pairs. In the W conformation, although the adduct is also in the *syn* orientation, the bulky group is located in the minor groove and the Hoogsteen face of the adduct is available to hydrogen bond with the opposing C. The dominant conformation of the damaged FP duplexes is dictated by the ability of the adduct to stack with the flanking base pairs.

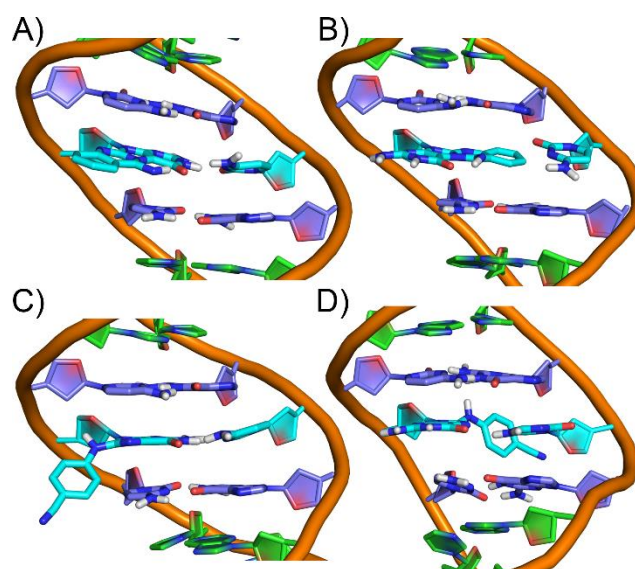


Figure 3.9: Representative structures of the FP duplexes containing adduct **1** in the A) B conformation or B) S conformation or containing adduct **2** in the C) B conformation or D) W conformation.

Regardless of whether simulations on the FP duplex containing adduct **2** are initiated from an S or W conformation, a similar structure is obtained that positions the bulky group in the minor groove and contains one or two hydrogen bond(s) between the damaged G and the opposing C (> 40% occupancy, Tables B.14-B.17, Appendix B). Since the bulky moiety is considerably nonplanar with respect to the flanking base pairs (average

interplanar angle =  $55.5^{\circ} \pm 6.8^{\circ}$ , Figures B.40, B.41 and Table B.18, Appendix B), no significant stabilization is obtained from stacking interactions involving the bulky moiety of the *syn* adduct. When coupled with the undistorted B-conformation of damaged FP duplexes containing the *anti* lesion, it can be concluded that adduct **2** is more likely to maintain the canonical B-form duplex. In contrast, *syn* **1** is nicely positioned between the flanking base pairs to form stabilizing stacking interactions in the FP duplexes in simulations initiated from the S conformation, with an average interplanar angle between the lesion and flanking base pairs of  $12.5^{\circ} \pm 0.7^{\circ}$  (Table B.18, Appendix B). Moreover, for the majority of the FP duplex simulations initiated from a W conformation, the bulky group repositions to intercalate within the helix and stack with the flanking base pairs. In instances where the bulky moiety remains in the minor groove, only one hydrogen bond occurs between the lesion and the opposing C with low occupancy (on average 17%; Table B.17, Appendix B). This suggests that **1** likely primarily adopts the S conformation when in the *syn* orientation. Nevertheless, the adduct also stabilizes in the B conformation of damaged DNA, suggesting that there may be conformational heterogeneity for FP duplexes containing this lesion.

### 3.5 Conclusions

Our work highlights the involvement of an *ortho*-cyano substituent in formation of an unusual polycyclic *N*-linked C8,N7-dG adduct. The resulting planar fluorescent adduct **1** adopts the promutagenic *syn* conformation. In contrast, adduct **2** derived from *para*-cyano-aniline maintains the *anti* conformation and is devoid of fluorescent properties. MD simulations demonstrate that while **1** and **2** both lead to distorted SMI helices regardless of the *anti* or *syn* glycosidic orientation, only *syn* **1** stabilizes the SMI as evidenced by a

persistent non-distorted planar duplex structure that benefits from stacking interactions between the lesion and surrounding nucleobases. Furthermore, although both *anti* **1** and **2** are well accommodated and maintain interstrand WC pairing in the B conformation of the FP duplexes, *syn* **2** is unable to stabilize the S conformation of damaged FP duplexes, while *syn* **1** preferentially stabilizes an intercalated lesion conformation. Thus, the MD simulations suggest that only **1** is able to stabilize the promutagenic S conformation in FP and SMI duplexes, likely due to a greater potential of the closed polycyclic ring to stack with the flanking base pairs. Thus, the “cyano effect” noted for the potent mutagenicity of cyano-nitroanilines in the Ames test utilizing frame-shift sensitive strains may stem from the ability of the monocyclic nitroaniline precursor to produce a polycyclic adduct that features the structural attributes of *N*-linked C8-dG adducts derived from polycyclic nitroaromatic and arylamine precursors that are potent frame-shift mutagens.

### 3.6 References

- (1) Josephy, P. D., and Mannervik, B. (2006) *Molecular Toxicology*, 2nd ed., pp 517-518, Oxford University Press, Inc., New York.
- (2) Cho, B. (2010) Structure-function characteristics of aromatic amine-DNA adducts, in *The Chemical Biology of DNA Damage* (Geacintov, N. E. and Broyde, S., Eds) pp. 217-238, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.
- (3) Skipper, P. L., Kim, M. Y., Sun, H.-L. P., Wogan, G. N., and Tannenbaum, S. R. (2010) Monocyclic aromatic amines as potential human carcinogens: old is new again. *Carcinogenesis* **31**, 50-58.
- (4) Böge, N., Gräsl, S., and Meier C. (2006) Synthesis and properties of oligonucleotides containing C8-deoxyguanosine arylamine adducts of borderline carcinogens. *J. Org. Chem.* **71**, 9728-9738.
- (5) Naiman, K., Dračinský, M., Hodek, P., Martínková, M., Schmeiser, H. H., Frei, E., and Stiborová, M. (2012) Formation, persistence, and identification of DNA adducts



formed by the carcinogenic environmental pollutant *o*-anisidine in rats. *Toxicol. Sci.* 127, 348-359.

(6) Patel, D. J., Mao, B., Gu, Z., Hingerty, B. E., Gorin, A., Basu, A. K., and Broyde, S. (1998) Nuclear magnetic resonance solution structures of covalent aromatic amine–DNA adducts and their mutagenic relevance. *Chem. Res. Toxicol.* 11, 391-407.

(7) Mao, B., Hingerty, B. E., Broyde, S., and Patel, D. J. (1998) Solution structure of the aminofluorene [AF]-intercalated conformer of the *syn*-[AF]-C8-dG adduct opposite dC in a DNA duplex. *Biochemistry* 37, 81-94.

(8) Wang, F., DeMuro, N. E., Elmquist, C. E., Stover, J. S., Rizzo, C. J., and Stone, M. P. (2006) Base-displaced intercalated structure of the food mutagen 2-amino-3-methylimidazo[4,5-*f*]quinoline in the recognition sequence of the *NarI* restriction enzyme, a hotspot for –2 bp deletions. *J. Am. Chem. Soc.* 128, 10085-10095.

(9) Fuchs, R. P. P., Schwartz, N., and Daune, M. P. (1981) Hot spots of frameshift mutations induced by the ultimate carcinogen N-acetoxy-N-2-acetylaminofluorene. *Nature* 294, 657-659.

(10) Burnouf, D., Koehl, P., and Fuchs, R. (1989) Single adduct mutagenesis: strong effect of the position of a single acetylaminofluorene adduct within a mutation hot spot. *Proc. Natl. Acad. Sci.* 86, 4147-4151.

(11) Broschard, T. H., Koffel-Schwartz, N., and Fuchs, R. P. (1999) Sequence-dependent modulation of frameshift mutagenesis at *NarI*-derived mutation hot spots. *J. Mol. Biol.* 288, 191-199.

(12) Milhe, C., Fuchs, R. P. P., and Lefevre, J. F. (1996) NMR data show that the carcinogen N-2-acetylaminofluorene stabilises an intermediate of –2 frameshift mutagenesis in a region of high mutation frequency. *Eur. J. Biochem.* 235, 120-127.

(13) Xu, L., and Cho, B. P. (2016) Conformational insights into the mechanism of acetylaminofluorene-dG-induced frameshift mutations in the *NarI* mutational hotspot. *Chem. Res. Toxicol.* 29, 213-226.

(14) Hoffmann, G. R., and Fuchs, R. P. P. (1997) Mechanisms of frameshift mutations: insight from aromatic amines. *Chem. Res. Toxicol.* 10, 347-359.

(15) Shapiro, R., Ellis, S., Hingerty, B. E., and Broyde, S. (1998) Effect of ring size on conformations of aromatic amine–DNA adducts: The aniline–C8 guanine adduct resides in the B-DNA major groove. *Chem. Res. Toxicol.* 11, 335-341.

(16) Josephy, P. D., Zahid, M., Dhanoa, J., de Souza, G. B. D., Groom, H., and Lambie, M. (2016) Potent mutagenicity in the Ames test of 2-cyano-4-nitroaniline and 2,6-dicyano-4-nitroaniline, components of disperse dyes. *Environ. Mol. Mutagen.* 57, 10-16.

- (17) Josephy, P. D., Dhanoa, J., Elzawy, G., Heney, K., Petrie, L., and Senis C. (2018) Structure-activity investigation of the potentiating effect of cyano substitution on nitroaniline mutagenicity in the Ames test. *Environ. Mol. Mutagen.* 59, 114-122.
- (18) Baldwin, J. E. (1976) Rules for ring closure. *J. Chem. Soc., Chem. Commun.* 734-736.
- (19) Quintero, C., Valderrama, M., Becerra, A., Daniluc, C. G., and Rojas, R. S. (2015) Synthesis of 4(3*H*)quinazolinimines by reaction of (*E*)-*N*-(aryl)-acetimidoyl or -benzimidoyl chloride with amines. *Org. Biomol. Chem.* 13, 6183-6193.
- (20) Ramanathan, M., Liu, Y.-H., Peng, S.-M., and Liu, S.-T. (2017) Synthesis of substituted quinazolin-4(3*H*)-imines from aryldiazonium salts, nitriles and 2-cyanoanilines via a metal-free tandem approach. *Org. Lett.* 19, 5840-5843.
- (21) Elmquist, C. E., Stover, J. S., Wang, Z., and Rizzo, C. J. (2004) Site-specific synthesis and properties of oligonucleotides containing C8-deoxyguanosine adducts of the dietary mutagen IQ. *J. Am. Chem. Soc.* 126, 11189-11201.
- (22) Kuska, M. S., Yazdi, M. M., Witham, A. A., Dahlmann, H. A., Sturla, S. J., Wetmore, S. D., and Manderville, R. A. (2013) Influence of chlorine substitution on the hydrolytic stability of biaryl ether nucleoside adducts produced by phenolic toxins. *J. Org. Chem.* 78, 7176-7185.
- (23) Schlitt, K. M., Sun, K. M., Paugh, R. J., Millen, A. L., Navarro-Whyte, L., Wetmore, S. D., and Manderville, R. A. (2009) Concerning the hydrolytic stability of 8-aryl-2'-deoxyguanosine nucleoside adducts: implications for abasic site formation at physiological pH. *J. Org. Chem.* 74, 5793-5802.
- (24) Dennington, R., Keith, T. A., and Millam, J. M. (2016) *Gaussview 6.0*, Semichem Inc. Shawnee Mission KS.
- (25) Millen, A. L., McLaughlin, C. K., Sun, K. M., Manderville, R. A., and Wetmore, S. D. (2008) Computational and experimental evidence for the structural preference of phenolic C-8 purine adducts. *J. Phys. Chem. A* 112, 3742-3753.
- (26) Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Petersson, G. A., Nakatsuji, H., Li, X., Caricato, M., Marenich, A. V., Bloino, J., Janesko, B. G., Gomperts, R., Mennucci, B., Hratchian, H. P., Ortiz, J. V., Izmaylov, A. F., Sonnenberg, J. L., Williams; Ding, F., Lipparini, F., Egidi, F., Goings, J., Peng, B., Petrone, A., Henderson, T., Ranasinghe, D., Zakrzewski, V. G., Gao, J., Rega, N., Zheng, G., Liang, W., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Throssell, K., Montgomery Jr., J. A., Peralta, J. E., Ogliaro, F., Bearpark, M. J., Heyd, J. J., Brothers, E. N., Kudin, K. N., Staroverov, V. N., Keith, T. A., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A. P., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Millam, J. M., Klene, M., Adamo, C., Cammi, R., Ochterski, J. W., Martin, R.

L., Morokuma, K., Farkas, O., Foresman, J. B., and Fox, D. J. (2016) *Gaussian 16 Rev. C.01*, Wallingford, CT.

(27) Berger, F. D., Sturla, S. J., Kung, R. W., Montana, T., Wetmore, S. D., and Manderville, R. A. (2018) Conformational preference and fluorescence response of a C-linked C8-biphenyl-guanine lesion in the *NarI* mutational hotspot: evidence for enhanced *syn* adduct formation. *Chem Res Toxicol.* 31, 37-47.

(28) Wilson, K. A., Kung, R. W., and Wetmore, S. D. (2016) Chapter Seven - Toxicology of DNA adducts formed upon human exposure to carcinogens: insights gained from molecular modeling. In *Advances in Molecular Toxicology*, Fishbein, J. C.; Heilman, J. M., Eds. Elsevier: Vol. 10, pp 293-360.

(29) Millen, A. L., Sharma, P., and Wetmore, S. D. (2012) C8-linked bulky guanosine DNA adducts: Experimental and computational insights into adduct conformational preferences and resulting mutagenicity. *Future Med. Chem.* 4, 1981-2007.

(30) Ivani, I., Dans, P. D., Noy, A., Pérez, A., Faustino, I., Hospital, A., Walther, J., Andrio, P., Goñi, R., Balaceanu, A., Portella, G., Battistini, F., Gelpí, J. L., González, C., Vendruscolo, M., Laughton, C. A., Harris, S. A., Case, D. A., and Orozco, M. (2015) Parmbsc1: a refined force field for DNA simulations. *Nature Methods* 13, 55.

(31) Case, D. A., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, T.E., Cruzeiro, V. W. D., Darden, T. A., Duke, R. E., Ghoreishi, D., Gilson, M. K., Gohlke, H., Goetz, A. W., Greene, D., Harris, R., Homeyer, N., Izadi, S., Kovalenko, A., Kurtzman, T., Lee, T. S., LeGrand, S., Li, P., Lin, C., Liu, J., Luchko, T., Luo, R., Mermelstein, D. J., Merz, K. M., Miao, Y., Monard, G., Nguyen, C., Nguyen, H., Omelyan, I., Onufriev, A., Pan, F., Qi, R., Roe, D. R., Roitberg, A., Sagui, C., Schott-Verdugo, S., Shen, J., Simmerling, C. L., Smith, J., Salomon-Ferrer, R., Swails, J., Walker, R. C., Wang, J., Wei, H., Wolf, R. M., Wu, X., Xiao, L., York, D. M., Kollman, P. A. (2018) *Amber 2018*, University of California: San Francisco.

(32) Dupradeau, F.-Y., Pigache, A., Zaffran, T., Savineau, C., Lelong, R., Grivel, N., Lelong, D., Rosanski, W., and Cieplak, P. (2010) The R.E.D. tools: advances in resp and esp charge derivation and force field library building. *Phys. Chem. Chem. Phys.* 12, 7821-7839.

(33) Grossfield, A., *WHAM: an implementation of the weighted histogram analysis method*, [http://membrane.urmc.rochester.edu/wordpress/?page\\_id=126](http://membrane.urmc.rochester.edu/wordpress/?page_id=126), version 2.0.9.1.

(34) Kitson, R. E., and Griffith, N. E. (1952) Infrared absorption band due to nitrile stretching vibration. *Anal. Chem.* 24, 334-337.

(35) Rankin, K. M., Sproviero, M., Rankin, K., Sharma, P., Wetmore, S. D., and Manderville, R. A. (2012) C8-Heteroaryl-2'-deoxyguanosine adducts as conformational fluorescent probes in the *NarI* recognition sequence. *J. Org. Chem.* 77, 10498-10508.

- (36) Cho, B. P. (1993) Structure of oxidatively damaged nucleic acid adducts: pH dependence of the  $^{13}\text{C}$  NMR spectra of 8-oxoguanosine and 8-oxoadenosine. *Mag. Reson. Chem.* 31, 1048-1053.
- (37) Bailey, W. F., and Cioffi, E. A. (1987) Carbon-13 NMR chemical shifts of representative nitriles and nitro compounds. *Magn. Reson. Chem.* 25, 181-183.
- (38) Novak, M., Ruenz, M., Kazerani, S., Toth, K., Nguyen, T.-M., and Heinrich, B. (2002) Kinetics of hydrolysis of 8-(arylamino)-2'-deoxyguanosines. *J. Org. Chem.* 67, 2303-2308.
- (39) Hevesi, L., Wolfson-Davidson, E., Nagy, J. B., Nagy, O. B., and Bruylants, A. (1972) Contribution to the mechanism of the acid-catalyzed hydrolysis of purine nucleosides. *J. Am. Chem. Soc.* 94, 4715-4720.
- (40) Da Costa, C. P., and Sigel, H. (2003) Acid-base and metal ion binding properties of guanylyl(3'→5')guanosine ( $\text{GpG}^-$ ) and 2'-deoxyguanosine[d(GpG) $^-$ ] in aqueous solution. *Inorg. Chem.* 42, 3475-3482.

## Chapter 4: Effect of Guanine Adduct Size, Shape and Linker Type on the Conformation of Damaged DNA: A Molecular Dynamics Study<sup>a,b,c</sup>

### 4.1 Introduction

DNA damaging agents are ubiquitous in our environment and include sources such as cigarette smoke, vehicle exhaust, pesticides and industrial agents.<sup>1</sup> In cells, key molecular components of these agents undergo bioactivation to form highly reactive species that attack various DNA positions to form adducts (addition products). Although cells have several DNA repair mechanisms,<sup>2, 3</sup> some forms of damage persist for long periods of time, leading to severe biological consequences, including several diseases.<sup>4</sup> Indeed, the formation of bulky nucleobase adducts is highly correlated with many cancers, including breast, liver and lung cancers.<sup>1, 5-8</sup> Previous work has hinted at a complex relationship between the chemical structure of the damaged nucleotide, the conformational preference of the adducted duplex, and the overall lesion mutagenicity or toxicity.<sup>9-13</sup> Numerous factors have been suggested to influence the conformational preference of damaged DNA, including the lesion size, shape, and aromaticity, length of chain connecting the damage moiety to the nucleobase, presence and location of heteroatoms, linker type and nucleobase binding site.<sup>1, 10, 14-19</sup> Nevertheless, the complex interconnectivity between the chemical composition of the lesion and DNA conformational outcomes is still poorly understood.

---

<sup>a</sup>The *Journal of American Chemical Society* Reference Style is used in this Chapter.

<sup>b</sup>Adapted from Kung, R.W.; Deak, T.K.; Bykowski, J.F.; Wetmore, S.D., Effect of Guanine Adduct Size, Shape and Linker Type on the Conformation of Damaged DNA: A Molecular Dynamics Study, [*in preparation*].

<sup>c</sup>R.W.K. and S.D.W. designed the project. R.W.K. generated adduct forcefield parameters and created initial starting structures. J.F.B. and R.W.K. performed DFT calculations. J.F.B. and T.K.D. performed molecular dynamics simulations and analyzed the trajectories. R.W.K. interpreted the data and wrote the first draft of the manuscript.

Although DNA adducts have been discovered for each of the four canonical nucleobases (adenine (A), cytosine (C), guanine (G) and thymine (T)), G adducts are the most prevalent in cells.<sup>1</sup> Bulky moieties can add to different guanine sites (N1, N2, O6, N7 and C8), with N2 and C8 being the most susceptible positions for attack.<sup>16</sup> To date, many C8-dG lesions have been identified that differ in how the carcinogen is linked to the C8 site. For example, 2-acetylaminofluorenyl-guanosine (<sup>AAF</sup>dG), 2-aminofluorenyl-dG (<sup>AF</sup>dG), 4-aminobiphenyl-dG (<sup>4ABP</sup>dG), 1-aminopyrenyl-dG (<sup>AP</sup>dG) and analine-dG (<sup>AN</sup>dG) have N-linkages.<sup>1, 10, 14, 16</sup> On the other hand, the pentachlorophenyl-dG (<sup>PCP</sup>dG) and phenolic-dG (<sup>O<sup>Ph</sup></sup>dG) adducts have O-linkages,<sup>1, 16, 20</sup> while phenyl-dG (<sup>Ph</sup>dG), ochratoxin A-dG (OTA-dG) and tolyl-dG (<sup>T<sup>o</sup></sup>dG) have direct C-linkages between the bulk of the damage and the adducted G.<sup>1, 13, 16, 21</sup>

Despite sharing a common G binding site, a range of biological outcomes are associated with the precursors to C8-dG lesions, even for those with the same linkage type.<sup>1, 16</sup> For example, PCP, the precursor to <sup>PCP</sup>dG, was used as a pesticide and disinfectant in the 20<sup>th</sup> century, but has been heavily regulated following the International Agency for Research on Cancer (IARC) labelling PCP as carcinogenic to humans (group 1).<sup>22</sup> In contrast, phenol, the precursor to O-linked <sup>O<sup>Ph</sup></sup>dG, is not labeled as classifiable as carcinogenic to humans (group 3).<sup>23</sup> Alternatively, 4ABP, which is found in tobacco and dyes and leads to N-linked <sup>ABP</sup>dG, has been classified as a group 1 carcinogen due to strong connections to the formation of bladder cancer.<sup>24</sup> Unfortunately, other compounds still lack sufficient evidence to be definitively labelled as carcinogenic or non-carcinogenic. For example, 1-nitropyrene, a precursor to N-linked <sup>AP</sup>dG, has been classified as probably carcinogenic to humans (group 2A).<sup>25</sup> Therefore, despite many compounds sharing a

common nucleobase binding position following metabolic activation, the associated C8-dG adducts have diverse biological consequences.

The toxicity of a lesion is at least partially linked to the ability of DNA repair enzymes to revert the associated damaged DNA to its undamaged state and the capacity of replication enzymes to duplicate DNA without errors. Repair of bulky DNA lesions is accomplished through the nucleotide excision repair (NER) pathway,<sup>26</sup> whereas the replication of bulky DNA adducts is generally accomplished by the translesion synthesis (TLS) response.<sup>27</sup> Several helical structural factors have been identified to impact lesion recognition and overall NER efficiency, including disruption of canonical Watson–Crick (WC) hydrogen bonding, changes in the base pair stacking interactions, and helical backbone distortions.<sup>28-36</sup> Additionally, changes in the local structure around the damage site, including rotation about the glycosidic bond, are believed to promote mutations during replication.<sup>1, 19, 37</sup> Collectively, these works illustrate that the impact of adduct composition on DNA repair and replication is dependent on the conformation of the adducted DNA.

Previous works in the field have primarily used nuclear magnetic resonance (NMR) spectroscopy and computational modelling to determine the conformational preferences of adducted DNA.<sup>1, 10, 38-45</sup> Although NMR methods, specifically <sup>19</sup>F NMR using F-labelled adducts, have been effectively utilized to determine the conformational ratios of damaged helices,<sup>13, 17, 18, 46-48</sup> the atomic information provided by these techniques is limited. Therefore, computational methods have been used to obtain complementary structural details of adducted DNA. Previous literature on several common N-linked C8-dG lesions (<sup>AF</sup>dG, <sup>AAF</sup>dG and <sup>ABP</sup>dG) has outlined three general conformational themes of the associated adducted duplexes (Figure 4.1).<sup>14, 15, 17, 18, 46-50</sup> In the major groove (B)

conformation, the bulky group is located in the DNA major groove such that the damaged nucleoside adopts an *anti* orientation ( $\chi = 180^\circ \pm 90^\circ$ ) and WC hydrogen bonding occurs between the adduct and the pairing dC. The wedge (W) conformation places the bulky moiety in the DNA minor groove as the damaged nucleoside adopts the *syn* orientation ( $\chi = 0^\circ \pm 90^\circ$ ), which exposes the Hoogsteen binding face of the adduct to the opposing strand. The adduct also adopts the *syn* orientation in the base-displaced stacked (S) conformation, but positions the bulky group in the interstrand space to  $\pi$ - $\pi$  stack with the flanking base pairs. Consequently, both the damaged and opposing bases are displaced into the major groove in the S conformer. Multiple studies have highlighted that the adducted duplexes often adopt more than a single conformation with the exact conformational heterogeneity depending on the system.<sup>1, 16-18, 42, 43, 49</sup> Although these conformations have been observed for several C8-dG lesions,<sup>1, 10, 14, 16-19, 42, 43, 49, 51</sup> the impact of the size and shape of the bulky group, and the type of aryl-base linker on the preferred helical structure is still unclear.

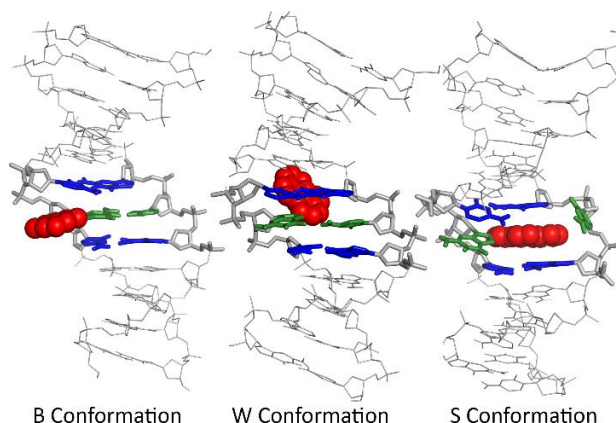


Figure 4.1: General conformational themes for C8-dG adducted DNA as shown using MD representative structures of <sup>Py</sup>dG adducted DNA in the major groove (B, left), wedge (W, middle), and base-displaced stacked (S, right) conformations.

To probe the effect of bulky moiety composition on the helical conformational heterogeneity of damaged DNA, prior computational work studied a systematic series of seven N-linked aromatic amine C8-dG adducts, with bulky groups incrementally increasing



in size from one to four aromatic rings (Figure 4.2).<sup>14</sup> The two-ringed lesions were chosen based on the addition of an aromatic ring to <sup>AN</sup>dG through  $\alpha$  (2,3-) fusion (<sup>1ANP</sup>dG),  $\beta$  (3,4-) fusion (<sup>2ANP</sup>dG) or (3-) substitution (<sup>3ABP</sup>dG, Figure 4.2). Correspondingly, the three-ringed moieties were generated by the addition of a naphthyl group by  $\alpha$  fusion (<sup>1APHN</sup>dG) or  $\beta$  fusion (<sup>3APHN</sup>dG). All six of these adducts are composed of subcomponents of the bulky group associated with the final four-ringed <sup>AP</sup>dG lesion. For each lesion, density functional theory (DFT) calculations and MD simulations were used to analyze the impact of the varying bulky moieties on the intrinsic structural preferences of the adducted site and the overall duplex. These data demonstrated the impact of both lesion size and shape on the preferred helical conformation. Specifically, <sup>AP</sup>dG and <sup>1APHN</sup>dG show the greatest propensity for the S conformation, while the smaller one and two ringed systems (<sup>AN</sup>dG, <sup>1ANP</sup>dG and <sup>2ANP</sup>dG) preferentially induced a B-type conformation. Finally, the <sup>3ABP</sup>dG and <sup>3APHN</sup>dG adducts were predicted to lead to a conformational heterogeneous mixture of the

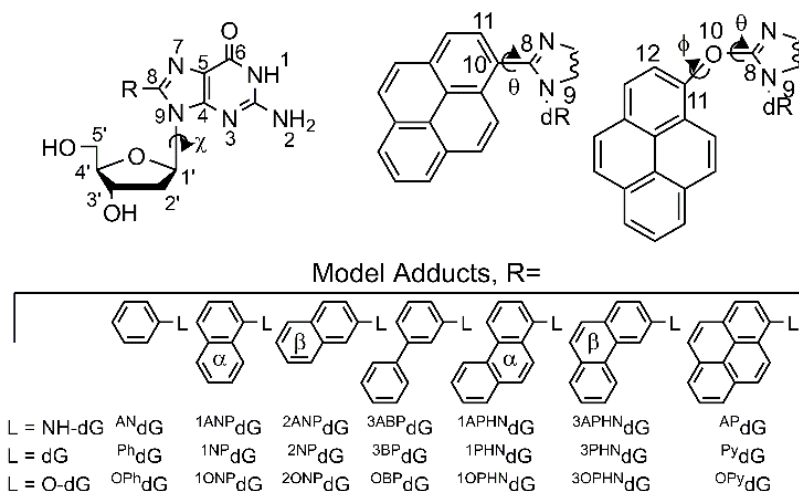


Figure 4.2: Structures, chemical numbering and nomenclature for model N, O and C-linked DNA adducts at the C8-site of 2'-deoxyguanosine. Key dihedral angles for the C-linked ( $\chi$  ( $\angle$  (O4'-C1'-N9-C4)) and  $\theta$  ( $\angle$  (N9-C8-C10-C11))) and O-linked ( $\chi$  ( $\angle$  (O4'-C1'-N9-C4)),  $\theta$  ( $\angle$  (N9-C8-O10-C11)) and  $\phi$  ( $\angle$  (C8-O10-C11-C12))) adducts defined for representative lesions.

B and S conformers. Although this study provided valuable insight into the effects of both adduct size and shape on the adopted helical conformation, more information is needed to validate whether the noted size and shape effects are consistent across different types of linkages due to the variety of nucleobase–carcinogen linkers observed in biological systems.<sup>1, 16</sup>

Since computer modelling has proven to be an invaluable tool to draw connections between the bulky group chemical composition and the structure of adducted DNA containing C8-dG lesions,<sup>1, 14, 16, 20, 33, 46, 52</sup> the present study uses a computational approach to consider 14 unique C-linked and O-linked C8-dG adducts (Figure 4.2). Specifically, DFT was used to study isolated nucleobase and nucleoside models of eight representative adducts that vary in adduct size, shape and linkage type, which uncovers the intrinsically preferred nucleobase-bulky moiety orientations. Subsequently, specific helical features and general conformational preferences of a 12-mer DNA oligomer containing each adduct at the G<sub>3</sub> site of the *NarI* sequence (5'–G<sub>1</sub>G<sub>2</sub>CG<sub>3</sub>CC) were revealed using MD simulations. Overall, the present data highlights key differences in the preferred placement of the bulky moiety with respect to the nucleobase upon varying the lesion shape and type of aryl–base linker. Additionally, the greater influence of adduct size than shape on the conformation heterogeneity of the adducted duplex previously reported for N-linked adducts is confirmed for the C-linked and O-linked analogues. Overall, this work provides the first systematic investigation into the impact of the nucleobase–carcinogen linkage type (C vs O vs N) on the behavior of the adduct as a free nucleobase, nucleoside and within a DNA helix. Such information can be applied to better predict the toxicity of novel bulky DNA adducts and

potentially allow for the design of better anticancer drugs that more effectively damage the DNA of cancer cells.<sup>2, 53</sup>

## 4.2 Materials and Methods

### 4.2.1 DFT Calculations

To understand the inherent structures of O and C-linked damaged G, the potential energy surface (PES) was initially mapped for the nucleobase and nucleoside models as previously described for other C8-dG adducts.<sup>14, 20, 52, 54, 55</sup> Initial structures were built by modifying those previously published for the corresponding N-linked C8-dG adducts using Gaussview 5.0.<sup>56</sup> The nucleoside models adopt the biologically-relevant C2'-endo puckering, and the 5' and 3' terminal hydroxy groups were constrained to the approximate values expected for B-DNA (namely  $\angle(\text{H-O5'}-\text{C5'}-\text{C4'}) = 180^\circ$  and  $\angle(\text{H-O3'}-\text{C3'}-\text{H3'}) = 60^\circ$ ). For the O-linked nucleobase adducts, the PES was considered as a function of both  $\theta$  ( $\angle(\text{N9-C8-O10-C11})$ ) and  $\phi$  ( $\angle(\text{C8-O10-C11-C12})$ ), while the PES of each C-linked nucleobase adduct was considered as a function of  $\theta$  ( $\angle(\text{N9-C8-C10-C11})$ ). The impact of the relative orientation of the nucleobase and sugar was also considered in nucleoside models by searching the PES as a function of  $\theta$  and  $\chi$  ( $\angle(\text{O4'}-\text{C1'}-\text{N9-C4})$ ). In each case, the dihedral angles dictating the orientation of the bulky group with respect to the nucleobase and the nucleobase with respect to the sugar were constrained in  $10^\circ$  steps from  $0-350^\circ$  and the rest of the system optimized with B3LYP/6-31G(d). Each minimum identified from the PES scans was subsequently fully optimized using B3LYP-D3(BJ)/6-31G(d). Reported energies were obtained at the B3LYP-D3(BJ)/6-311+G(2df,p) level of theory and include scaled (0.9813) zero-point vibrational corrections. These calculations were performed for the phenyl, 1-naphthyl, 3-phenanthrenyl and 1-pyrenyl C and O-linked

adducts (Figure 4.2), which provides a representative set of adducts to study the effects of the bulky moiety size and shape on the preferred damaged nucleobase and nucleoside orientations. All DFT calculations were done using Gaussian 09 (revision D.01).<sup>57</sup>

#### 4.2.2 MD Simulations

A total of 14 C-linked and O-linked C8-dG adducts were considered (Figure 4.2), namely the C-linked <sup>Ph</sup>dG, 1-naphthyl-dG (<sup>1NP</sup>dG), 2-naphthyl-dG (<sup>2NP</sup>dG), 3-biphenyl-dG (<sup>3BP</sup>dG), 1-phenanthrenyl-dG (<sup>1PHN</sup>dG), 3-phenanthrenyl-dG (<sup>3PHN</sup>dG), and 1-pyrenyl-dG (<sup>Py</sup>dG) adducts, and the O-linked <sup>O<sup>Ph</sup></sup>dG, 1-naphthol-dG (<sup>1ONP</sup>dG), 2-naphthol-dG (<sup>2ONP</sup>dG), 3-biphenol-dG (<sup>OBP</sup>dG), 1-phenanthrenol-dG (<sup>1OPHN</sup>dG), 3-phenanthrenol-dG (<sup>3OPHN</sup>dG), and 1-pyrenol-dG (<sup>OPy</sup>dG) adducts. Each adduct was incorporated at the X position opposite C in the 5'-CTCGGCXCCATC-3' DNA sequence, which contains the *NarI* mutational hotspot (underlined).<sup>58</sup> Each adducted duplex was modelled in the B, W and S helical conformations, with the starting structures generated from representative structures of the corresponding N-linked adducted DNA in the relevant conformation.<sup>14</sup> For a given adduct and duplex conformation, multiple rotamers with respect to  $\theta$  and  $\phi$  (when applicable) were considered, with the exception of <sup>Ph</sup>dG, which was considered with respect to  $\theta$  due to symmetry. Specifically, for each damaged DNA conformation, the remaining 6 C-linked adducts and O-linked <sup>O<sup>Ph</sup></sup>dG were modelled with two possible  $\theta$  values (i.e.,  $\sim 0^\circ$  and  $\sim 180^\circ$ ), while four rotamers (i.e.,  $\theta \sim 0^\circ$  and  $\phi \sim 0^\circ$ ,  $\theta \sim 0^\circ$  and  $\phi \sim 180^\circ$ ,  $\theta \sim 180^\circ$  and  $\phi \sim 0^\circ$ , and  $\theta \sim 180^\circ$  and  $\phi \sim 180^\circ$ ) were modelled for the remaining six O-linked adducts.

The AMBER14SB force field was used to describe the canonical DNA bases,<sup>59</sup> while the general amber force field (GAFF) was used to supplement the AMBER parameters for the damaged nucleotides.<sup>60</sup> Charges for non-standard residues were

developed using the R.E.D. program (version III.4)<sup>61</sup> to ensure compatibility with the AMBER force field. B3LYP-D3(BJ)/6-31G(d) optimized nucleotides in the *anti* and *syn* orientations were used as the R.E.D. input structures. Using the tleap module of the AMBER 14 program,<sup>62</sup> each damaged DNA duplex and the corresponding canonical helix were solvated in an octahedral TIP3P water box such that there was a minimum of 8 Å between the DNA and the edge of the box in all directions. Twenty-two sodium ions were added to neutralize the solute. The systems were minimized over 4 stages. First, the solvent was optimized using 500 steps of steepest decent (SD) minimization, with a 500 kcal mol<sup>-1</sup> Å<sup>-2</sup> force restraint on DNA atoms. This was followed by 500 steps of the conjugate gradient (CG) algorithm with the same restraint on DNA. The third stage involved 1000 steps of unconstrained SD energy minimization. Finally, the entire system was minimized without constraints using 1500 steps of the CG procedure. The systems were then heated to 300 K under constant volume for 20 ps. The Langevin thermostat ( $\gamma = 1$ ) was implemented during heating along with a 10 kcal mol<sup>-1</sup> Å<sup>-2</sup> restraint on the DNA. Subsequently, MD production simulations were performed on each system for 100 ns with a constant pressure of 1 atm (implemented using the Langevin algorithm) and a constant temperature of 300 K (maintained by the Berendsen procedure). Three 120 ns replicas on select N-linked adducts suggested that structural and energetic conclusions were consistent with the 100 ns timescale.<sup>14</sup> The heavy atoms involved in each WC hydrogen bond in the terminal base pairs were restrained to 3.0 Å using a 25 kcal mol<sup>-1</sup> Å<sup>-2</sup> harmonic restraint to prevent unraveling during the production simulation. An 8 Å cutoff was used for nonbonded interactions and supplemented with the particle-mesh Ewald method for Coulombic interactions. During all MD production simulations, SHAKE was applied to the hydrogens and a time step of 2 fs was implemented. The periodic boundary condition was

used during every stage of the MD procedure as implemented using the PMEMD module of AMBER 14.<sup>62</sup> Each trajectory was saved and analyzed every 2.5 ps.

As previously done for the analogous N-linked adducts,<sup>14</sup> the multiple damaged DNA conformations for each (*anti* and *syn*) glycosidic orientation of the damaged nucleotide were ranked with respect to the free energies as determined using MM/PBSA and normal mode analysis calculations. The MM/PBSA calculations were performed on 1000 frames of solvent and sodium free trajectories. We acknowledge that errors have been observed with using MM/PBSA energies, but we anticipate that errors within the MM/PBSA protocol are minimized when comparing the highly similar structures in the present work. Due to the large number of adducts and conformations considered in the present work, our analysis focused on the one adduct rotamer per lesion glycosidic conformation that yields the most stable adducted DNA helix.

Structural analysis was performed using the CPPTRAJ module in AMBER.<sup>63</sup> Specifically, the hydrogen-bonding occupancies were calculated using a 3.4 Å heavy-atom cutoff distance and a 120° donor-hydrogen-acceptor angle. Key dihedral angles within the damaged nucleotide ( $\chi$ ,  $\theta$ , and  $\phi$ ) were also determine as well as the DNA backbone RMSD and solvent accessible surface area (SASA) of the bulky group. Representative structures were generated for each trajectory by clustering with respect to the  $\chi$ ,  $\theta$ , and  $\phi$  dihedral angles using the heiragglo algorithm. Lesion-site interaction energies (LSIEs) were calculated using the linear-interaction energy (LIE) command in CPPTRAJ. Specifically, the average hydrogen-bond strengths were determined for the adducted base pair and the flanking base pairs, while the average stacking interactions were calculated between the

adduct or pairing C and the four flanking nucleobases. The reported LSIE is the sum of these average hydrogen-bonding and stacking energies.

## 4.3 Results

### 4.3.1 DFT Calculations Suggest that the Shape, but not Size, of the Bulky Moiety Impacts the Inherent Structural Preferences of C-linked dG Adducts

#### 4.3.1.1 Structure at the Aryl–Nucleobase Linker

The intrinsic structural preferences for the  $\theta$  dihedral angle describing the rotation about the bond connecting the bulky group and damaged G show significant bulky group shape effects for C-linked nucleobase adducts (Figure C.1, Appendix C). Notably, the PESs are almost identical for one-ringed <sup>Ph</sup>G and  $\beta$ -fused three-ringed <sup>3PHN</sup>G, showing a planar preference between the aryl group and G (minima at  $\theta \sim 0^\circ$  and  $\sim 180^\circ$ , Figures C.2, C.3, Appendix C). In contrast, the  $\alpha$ -fused ring containing two-ringed <sup>1NP</sup>G and four-ringed <sup>Py</sup>G adducts have four minima that are twisted by 25–40° from planar in either direction (Figures C.4, C.5, Appendix C), which is caused by steric repulsion between the  $\alpha$ -fused ring and N7 ( $\theta \approx 0^\circ$ ) or N9 ( $\theta \approx 180^\circ$ ) of G. The larger degree of twist ( $\sim 36^\circ$ ) near  $\theta = 180^\circ$  compared to  $\theta = 0^\circ$  (twist of  $\sim 25^\circ$ ) is likely due to greater steric repulsion between the  $\alpha$ -fused ring and N9–H than N7. This also suggests that the degree of twisting may be larger in larger (duplex) models due to the increased bulk around N9. Overall, this highlights that key differences in the orientation about the C-linkage in C8-dG adducts prevail due to the shape, but not size, of the aryl group.

#### 4.3.1.2 Preferred Glycosidic Orientation of C-linked Nucleosides

The DFT energy as a function of the guanine–bulky moiety linkage (defined by  $\theta$ ) and nucleobase–sugar linkage (defined by  $\chi$ ) for the four C-linked nucleoside adducts reveals a clear adduct shape dependence. Consistent with the nucleobase model, the  $\alpha$ -fused ring containing adducts ( $^{1\text{NP}}\text{dG}$  and  $^{\text{Py}}\text{dG}$ ) have a high-energy ridge at  $\theta \approx 180^\circ$  (green, Figures C.6, Appendix C) due to steric clashes between the  $\alpha$ -fused ring and the sugar moiety. In contrast, the  $^{\text{Ph}}\text{dG}$  and  $^{3\text{PHN}}\text{dG}$  adducts have additional minima ( $\theta = 180^\circ \pm 40^\circ$ ), which include the global minimum. Despite this difference, all adducts can adopt stable *anti* ( $\chi = 230^\circ - 250^\circ$ ) and *syn* ( $\chi = 80^\circ - 90^\circ$ ) glycosidic orientations, with *anti* being thermodynamically preferred by  $< 12$  kJ/mol in all cases and a small ( $\sim 20$  kJ/mol) barrier separating the *anti* and *syn* energetic valleys (Figures C.7-C.10, Appendix C). Overall, although the shape of the adduct impacts the preferred orientation about the nucleobase–bulky moiety linker, neither the shape nor the size significantly affect the adduct glycosidic conformational flexibility.

### 4.3.2 MD Simulations Reveal that Bulky Moiety Shape and Size Impact the C-linked Adducted DNA Conformational Preference

#### 4.3.2.1 Overall Helical Features are Independent of C-linked Adduct

MD simulations reveal that damaged DNA containing each of the 7 C-linked adducts in the B, W and S conformations have consistent structural features (Figures C.11-C.18, Appendix C). Regardless of the specific adduct or conformation, the hydrogen bonding at the terminal base pairs and the base pairs flanking the adduct site maintain WC hydrogen bonding for the majority of the trajectories, with the average hydrogen-bonding



occupancy of the three G:C hydrogen-bonds being  $> 90\%$  for all systems (Table C.1, Appendix C). For the B conformation, three strong WC hydrogen bonds occur between the adducted G and the pairing C (occupancy  $\geq 98\%$ , Table C.2, Appendix C). The W conformation has considerably weaker hydrogen bonding than the B conformer or natural base pairs at the adduct site, with two periodic hydrogen bonds between the *syn* Hoogsteen face of the adduct and the pairing C (occupancies  $> 75\%$ , Figure 4.3, Table C.2, Appendix C). In the S conformation of damaged DNA, although no hydrogen bonding occurs between the damaged G and the pairing C since the two bases are extrahelical, most systems have a transient hydrogen bond (average occupancy  $\approx 60\%$ ) between N4–H of the displaced C and the 5'-phosphate group (Table C.2, Appendix C). Instead, the bulky group of the adduct stacks within the DNA duplex. Overall, the structural features of DNA containing C-linked adducts is generally independent of the aryl moiety shape and size.

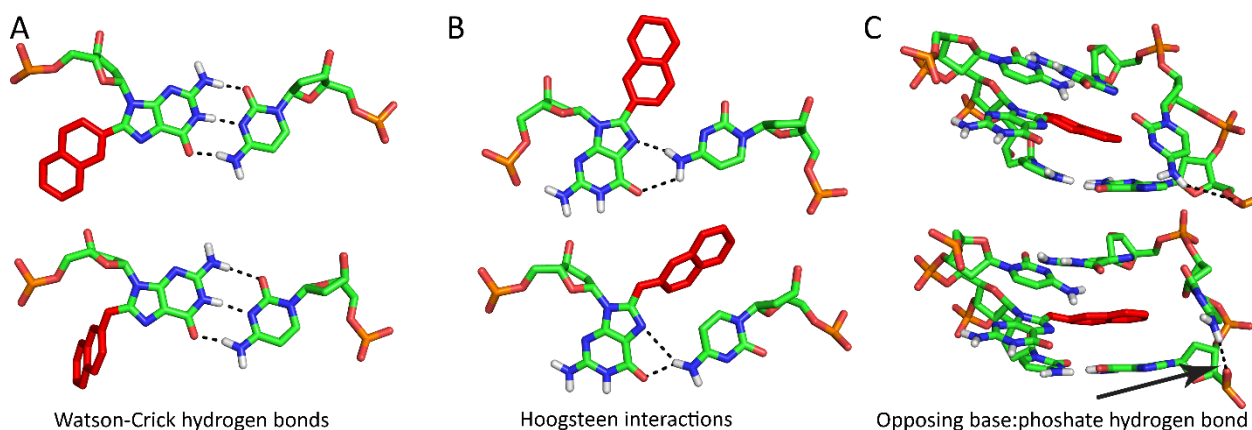


Figure 4.3: Hydrogen-bonding interactions at the lesion site in the (A) B, (B) W and (C) S conformations of adducted DNA, using  $^{2\text{NP}}$ dG adducted DNA as examples for the C-linked systems (top) and  $^{2\text{ONP}}$ dG adducted DNA as examples for the O-linked systems (bottom).

#### 4.3.2.2 C-linked Adduct Orientation is Dependent on Lesion Shape

Within each of the three adducted DNA conformations, the average  $\chi$  ( $\chi^{\text{avg}}$ ) and  $\theta$  ( $\theta^{\text{avg}}$ ) are comparable for all seven C-linked adducts (Table C.3, Appendix C). Specifically, the lesion-site glycosidic orientation is consistent for damaged DNA in the B ( $\chi^{\text{avg}} = 219^\circ - 244^\circ$ ), S ( $37^\circ - 53^\circ$ ) and W ( $40^\circ - 72^\circ$ ) helical conformations. In contrast,  $\theta^{\text{avg}}$  in the B ( $3^\circ - 13^\circ$ ), S ( $343^\circ - 358^\circ$ ) and W ( $180^\circ - 204^\circ$ ) duplex conformations shows a notably aryl shape effect on the linker orientation. As seen for the DFT models, bulky moieties lacking an  $\alpha$ -ring ( $^{\text{Ph}}\text{dG}$ ,  $^{2\text{NP}}\text{dG}$ ,  $^{3\text{BP}}\text{dG}$  and  $^{3\text{PHN}}\text{dG}$ ) are consistently more planar ( $\theta^{\text{avg}} \approx 5^\circ$ ) than those containing an  $\alpha$ -ring ( $\theta^{\text{avg}} = 10^\circ - 25^\circ$ ,  $^{1\text{NP}}\text{dG}$ ,  $^{1\text{PHN}}\text{dG}$  and  $^{\text{Py}}\text{dG}$ ). Thus, overall, the  $\chi$  dihedral angle is independent of the C-linked lesions size and shape, whereas the shape of the bulky group dictates the twist in aryl-G linkage.

#### 4.3.2.3 Nonbonded Lesion-Site Interactions are Dependent on the Damaged DNA Conformation as well as the Bulky Group Size and Shape, Particularly for the syn Adduct Orientation

The stability of the lesion site is modulated by both hydrogen bonding between the adducted nucleobase and the pairing C, and their stacking interactions with the flanking base pairs. Within the B conformation, the maximum difference in the LSIEs between any two C-linked adducts systems being  $\geq 7$  kJ/mol, with an average LSIE ( $\text{LSIE}^{\text{avg}}$ ) of  $\sim -479$  kJ/mol (Figure 4.4, Table C.4, Appendix C). The LSIE is significantly less in the W (LSIE =  $-335$  to  $-378$  kJ/mol) than the B conformation. Interestingly, although the LSIE fluctuates with the adduct for the W conformer, with  $^{1\text{NP}}\text{dG}$  being the least stable and  $^{1\text{PHN}}\text{dG}$  being the most stable (Figure 4.4, Table C.4, Appendix C), there is no clear lesion shape or size effects.

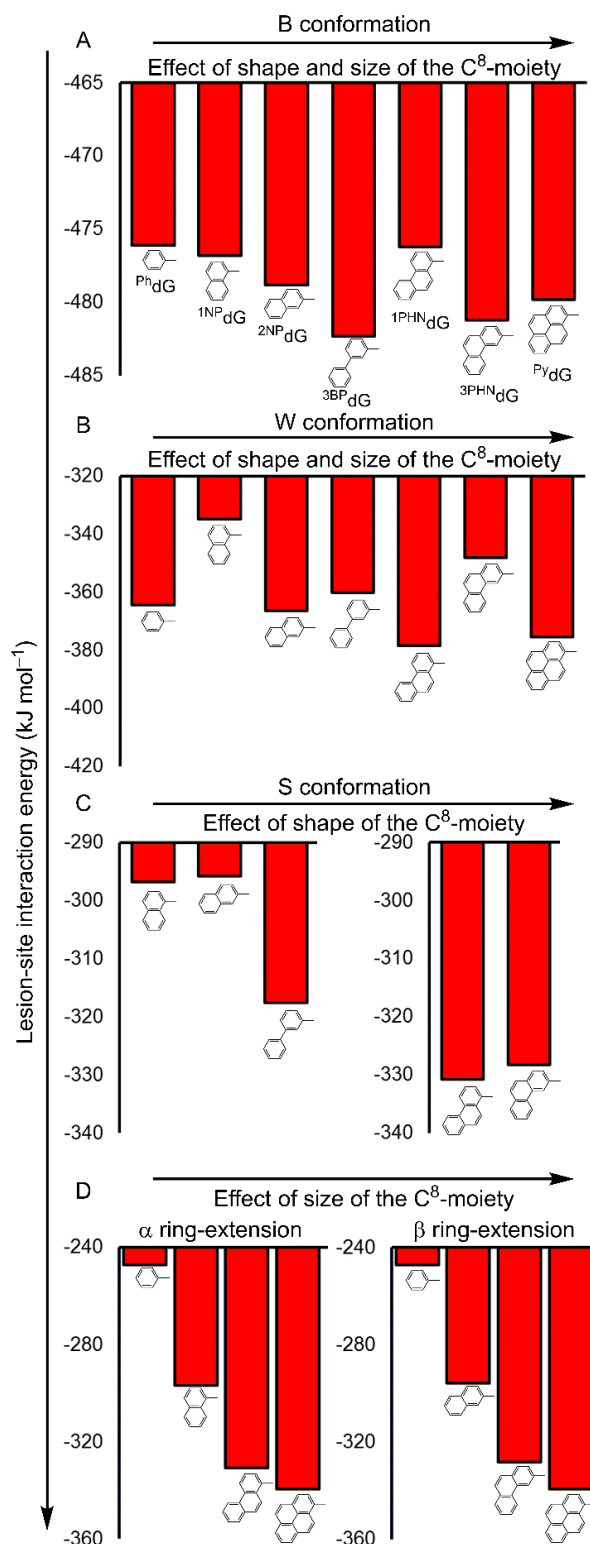


Figure 4.4: Effect of C8-moiety A) size and shape in the B conformation, B) size and shape in the W conformation, C) shape in the S conformation, and D) size in the S conformation on the lesion-site interaction energy (LSIE) for damaged DNA containing a C-linked dG adduct.

In the S conformation, the <sup>Ph</sup>dG adduct results in the smallest LSIE (−247.1 kJ/mol, Table C.4, Appendix C). Adding a second ring increases the LSIEs from <sup>Ph</sup>dG adducted DNA, with fused rings resulting in a smaller LSIE increase ( $\sim -49$  kJ/mol) than ring substitution (−70.4 kJ/mol, Figure 4.4C). The three-ringed lesions have greater lesion-site stability than the two-ringed lesions, with the LSIEs increasing by  $\sim -80$  kJ/mol for <sup>1PHN</sup>dG and <sup>3PHN</sup>dG. Finally, the largest four-ringed <sup>Py</sup>dG adduct leads to the greatest increase in the LSIE of <sup>Ph</sup>dG in the S damaged DNA conformation (−92.3 kJ/mol). Overall, although the LSIEs of C-linked adducted DNA have no clear adduct dependence for the B and W conformations, the LSIE for the S conformation of damaged DNA has a minor aryl moiety shape impact (ring-substitution  $\gg$   $\alpha$ -ring fusion  $>$   $\beta$ -ring fusion). The lesion size significantly influences the strength of the lesion-site interactions, with the LSIE increasing with bulky group size (Figure 4.4D).

#### 4.3.2.4 Lesion Size Affects the Solvent Exposure of the Aryl Group

Due to the hydrophobic nature of the aryl groups, increased solvent exposure of the bulky moiety of the adducts considered in the present work is anticipated to destabilize the local helical structure. For C-linked adducted DNA in all three conformations, the smallest phenyl-derived <sup>Ph</sup>dG systems have the lowest average SASA (SASA<sup>avg</sup>), although the B (98 Å<sup>2</sup>) and W (78 Å<sup>2</sup>) conformations are significantly more solvent exposed than the S conformation (42 Å<sup>2</sup>, Figure 4.5, Table C.5, Appendix C). Interestingly, despite DNA in the B conformation having similar SASA<sup>avg</sup> values after adding an  $\alpha$ - or  $\beta$ -fused ring to the single-ringed system ( $\sim 131$  Å<sup>2</sup>),  $\alpha$  fusion results in greater solvent exposure for adducted DNA in the S (<sup>1NP</sup>dG = 56 Å<sup>2</sup> and <sup>2NP</sup>dG = 45 Å<sup>2</sup>) and W (<sup>1NP</sup>dG = 137 Å<sup>2</sup> and <sup>2NP</sup>dG = 107 Å<sup>2</sup>) conformations. An even larger SASA value occurs upon ring substitution (<sup>3BP</sup>dG) for

both the B ( $178 \text{ \AA}^2$ ) and W ( $153 \text{ \AA}^2$ ) conformations, resulting in the most solvent exposed adduct for both conformations. Notably, the solvent exposure for the largest  $^{1\text{PHN}}\text{dG}$ ,  $^{3\text{PHN}}\text{dG}$  and  $^{\text{Py}}\text{dG}$  adducts are similar ( $\sim 168 \text{ \AA}^2$ ) when the DNA adopts the B conformation, whereas the phenanthrenyl adducts are more solvent exposed ( $\sim 147 \text{ \AA}^2$ ) than  $^{\text{Py}}\text{dG}$  ( $136.6 \text{ \AA}^2$ ) for W adducted. Interestingly,  $^{3\text{BP}}\text{dG}$  has a SASA value between the two naphthyl lesions for the S conformation of damaged DNA. Adducted DNA in the S conformation shows an increase in solvent exposure from the smaller moieties as the number of aryl rings in the bulky group increases to three ( $\sim 67 \text{ \AA}^2$ ) or four ( $92 \text{ \AA}^2$ ). Remarkably, the most exposed  $^{\text{Py}}\text{dG}$  adduct in the S conformation of DNA is still more shielded than the smallest  $^{\text{Ph}}\text{dG}$  lesion in the B conformation, highlighting the large degree of solvent shielding in the intercalated orientation. Overall, there is no consistent lesion shape effect on the  $\text{SASA}^{\text{avg}}$  although the aryl group size does impact the solvent exposure, with the incremental increase in solvent exposure with the number of rings in the bulky moiety being particularly evident for the S conformation.

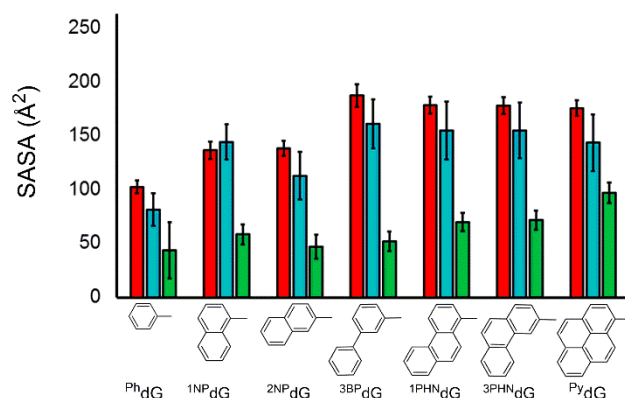


Figure 4.5: Dependence of the solvent accessible surface area (SASA,  $\text{\AA}^2$ ) on the shape and size of the hydrophobic NCA for the B (red), W (blue) and S (green) conformations of damaged DNA containing a C-linked dG adduct.

#### 4.3.2.5 The Preferred Conformation of C-linked Adducted DNA is Impacted by Lesion Shape and Size

By considering both the favorable nonbonded interaction energies (LSIE) and the unfavorable bulky group solvent exposure (SASA), information about the conformational preferences of adducted DNA can be inferred. For the seven C-linked adducts considered, the W damaged DNA conformation has weaker nonbonded interactions than the B conformation and high degree of solvent exposure compared to the S conformation, suggesting this conformation lacks strong lesion-site interactions to offset significant bulky moiety solvation (Figures 4.4, 4.5). Indeed, the W conformation is only known to be adopted when a C is paired across from <sup>AAF</sup>dG<sup>17</sup> or <sup>AF</sup>dG is mispaired.<sup>64-66</sup> Since the W conformation is unlikely to be adopted for DNA containing the C-linked lesions, we focus our discussion on the B and S conformations.

Although there are minor changes in the LSIEs of DNA containing different C-linked adducts in the B conformation (less than 7 kJ/mol), the SASA increases with lesion size according to <sup>Ph</sup>dG  $\ll$  <sup>1NP</sup>dG  $\approx$  <sup>2NP</sup>dG  $\ll$  <sup>1PHN</sup>dG  $\approx$  <sup>3PHN</sup>dG  $\approx$  <sup>Py</sup>dG  $<$  <sup>3BP</sup>dG (Figure 4.6). Thus, the stability of the B conformation will decrease for the larger (ring-substituted) adducts. However, the stability of the S conformation must be considered to understand the conformational heterogeneity of adducted DNA. Although <sup>Ph</sup>dG adducted DNA in the S conformation is only minimally destabilized by the hydrophobic interactions of the aryl moiety, the weaker nonbonded interactions at the lesion site compared to the B conformation ( $< -250$  kJ/mol) suggest that <sup>Ph</sup>dG adducted DNA is exclusively stabilized in the B conformation. In contrast to the smallest lesion, the largest <sup>Py</sup>dG adduct is associated

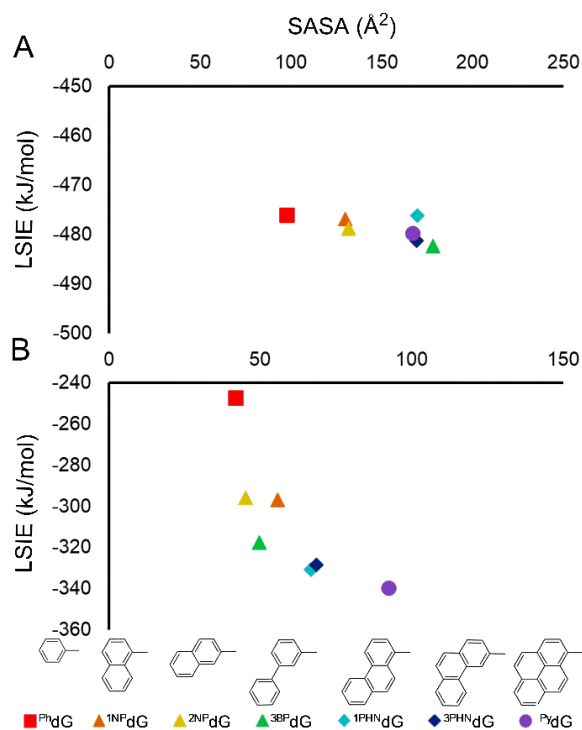


Figure 4.6: Dependence of the solvent accessible surface area (SASA,  $\text{\AA}^2$ ) and lesion-site interaction energy (LSIE, kJ/mol) on the shape and size of the hydrophobic C8-moiety for the (A) B and (B) S conformations of DNA containing a C-linked dG adduct.

with the greatest lesion-site interactions in the S conformation of adducted DNA and the greatest degree of solvent shielding compared to the B conformer, suggesting that the S conformation is adopted for the vast majority of DNA duplexes containing  $\text{PydG}$ . Although  $1\text{PHNdG}$  and  $3\text{PHNdG}$  have smaller SASAs than  $\text{PydG}$ , the nonbonded interactions are weaker when the adducted duplexes are in the S conformation, suggesting that the two C-linked phenanthrenyl derived adducts likely have similar propensities to preferentially induce the S over the B conformation. Despite the lower SASA,  $3\text{BPdG}$  exhibits more than a 20 kJ/mol decrease in the lesion-site interactions compared to  $\text{PydG}$  in the S conformer, suggesting a lower propensity for  $3\text{BPdG}$  adducted DNA to adopt an S conformation. Therefore, we propose that the  $3\text{BPdG}$  adduct will result in an equal or greater level of the B conformation

compared to S. Lastly, the  $^{1\text{NP}}\text{dG}$  and  $^{2\text{NP}}\text{dG}$  adducts lack strong nonbonded interactions in the S conformation, being  $> 20$  kJ/mol less than those for  $^{3\text{BP}}\text{dG}$ . When coupled with the significantly smaller destabilization of the  $^{1\text{NP}}\text{dG}$  and  $^{2\text{NP}}\text{dG}$  damaged DNA B conformer due to decreased solvation of the bulky group compared to  $^{3\text{BP}}\text{dG}$ , the two naphthyl-derived lesions likely preferentially induce the B conformation, with the  $\beta$ -fused  $^{2\text{NP}}\text{dG}$  likely promoting a greater occupancy of the S conformation due to less hydrophobic destabilization than  $\alpha$ -fused  $^{1\text{NP}}\text{dG}$ .

In summary, the one-ringed and two-fused-ringed adducts likely primarily adopt the B conformation, with  $^{\text{Ph}}\text{dG}$ -containing DNA exclusively adopting B and the  $\beta$ -fused ring lesion inducing a greater S preference. In contrast, DNA containing the two-ringed substituted  $^{3\text{BP}}\text{dG}$  adduct will likely prefer the B conformation with some portion of S. Conversely, the largest four-ringed  $^{\text{Py}}\text{dG}$  adduct likely predominately induces the S conformation, while three-ringed  $^{1\text{PHN}}\text{dG}$  and  $^{3\text{PHN}}\text{dG}$  mainly cause the S conformation, with a small percentage of the B orientation. In other words, the B:S conformational ratio is predicted to decrease as  $^{\text{Ph}}\text{dG} \gg ^{1\text{NP}}\text{dG} > ^{2\text{NP}}\text{dG} \gg ^{3\text{BP}}\text{dG} \gg ^{1\text{PHN}}\text{dG} \sim ^{3\text{PHN}}\text{dG} > ^{\text{Py}}\text{dG}$ .

### **4.3.3 DFT Calculations Show that the Aryl-G Linker Orientation Preferences are Dependent on Lesion Shape and not Size for O-linked dG Adducts**

#### *4.3.3.1 Structure at the Aryl–O–Nucleobase Linker*

DFT energy as a function of the  $\theta$  and  $\phi$  dihedral angles for the  $^{\text{OPh}}\text{G}$ ,  $^{1\text{ONP}}\text{G}$ ,  $^{3\text{OPHN}}\text{G}$  and  $^{\text{OPy}}\text{G}$  nucleobase adducts shows common features despite notable lesion shape effects (Figure C.18, Appendix C). For all four O-linked nucleobase adducts considered,  $\theta = 0^\circ \pm 90^\circ$  is destabilized compared to  $\theta = 180^\circ \pm 90^\circ$ , due to repulsion between the linking O and



N7 of G, and the aryl group and N9–H. As a result, the global minima for these lesions maintains planarity between the aryl group and the damaged G with  $\theta \approx 180^\circ$  (Figures C.19–22, Appendix C). Despite this commonality, the O-linked lesions show a clear bulky moiety shape effect. Specifically, lesions that lack an  $\alpha$ -fused ring ( $^{\text{OPh}}\text{G}$  and  $^{3\text{OPHN}}\text{G}$ ) have two minima at  $\theta \approx 180^\circ$  and  $\phi \approx 0^\circ$  or  $180^\circ$  (Figures C.19, C.20, Appendix C). In contrast, the  $\alpha$ -fused ring containing adducts ( $^{1\text{ONP}}\text{G}$  and  $^{\text{OPy}}\text{dG}$ ) have a stable minimum at  $\theta \approx 180^\circ$  and  $\phi \approx 0^\circ$  (Figures C.21–C.22, Appendix C) and two higher energy minima at  $\phi = 180^\circ \pm 30^\circ$  because of steric clashes between C–H of the  $\alpha$ -fused ring and N9–H. Therefore, despite all four O-linked adducts being structurally similar for  $\phi = 0^\circ \pm 90^\circ$ , the energy landscape when  $\phi = 180^\circ \pm 90^\circ$  is dependent on the lesion shape.

#### 4.3.3.2 Preferred Glycosidic Orientation of O-linked Nucleosides

The energy as a function of the sugar–nucleobase linkage ( $\chi$ ) and the O–C8 linker ( $\theta$ ) were also calculated for the four O-linked adducts that vary in both bulky group size and shape (Figure C.23, Appendix C). As with the O-linked nucleobase model, all adducts prefer  $\theta = 180^\circ \pm 90^\circ$  over  $\theta = 0^\circ \pm 90^\circ$ . Despite the low *anti/syn* conversion barrier ( $\leq 20$  kJ/mol), all four damaged nucleosides have an *anti* global minimum ( $\chi = 233 - 257^\circ$ ) and *syn* minima ( $\chi = 67 - 115^\circ$ ) that are only slightly less stable ( $< 11$  kJ/mol higher in energy, Figures C.24 – C.27, Appendix C). Overall, neither the bulky group size nor shape affect the glycosidic conformational preferences of the O-linked adducts, with both the *anti* and *syn* orientations, being thermodynamically and kinetically accessible.

#### 4.3.4 MD Simulations Show that the O-linked Adducted DNA Structural Preferences are Dictated by the Aryl Moiety Shape and Size

##### 4.3.4.1 Overall Helical Features are Independent of O-linked Adduct

For O-linked adducted DNA, the WC hydrogen bonding in the terminal and flanking base pairs is maintained regardless of duplex the conformation (occupancy > 90%, Table C.1, Appendix C). Additionally, damaged duplexes in the B conformation maintain WC hydrogen bonding at the adduct site ( $\geq 98\%$  occupancy, Figures C.28-34, Table C.2, Appendix C). Comparatively, the W conformation has sporadic hydrogen bonding between the adducted G and pairing C (occupancy > 90% but < 125%, Figure 4.3, Table C.2, Appendix C). <sup>OPy</sup>dG containing DNA in the W conformation is unique due to the presence of another hydrogen bond between N4–H of the pairing C and O6 of the 3' flanking G (Figure C.35, Table C.2, Appendix C). In the S conformation, as with the N-linked adducts,<sup>14</sup> the previously noted hydrogen bond between the displaced C and the 3' neighboring phosphate group is present (up to  $\sim 74\%$  occupancy, Table C.2, Appendix C), except for the <sup>1OPHN</sup>dG system for which the interaction was not present.

##### 4.3.4.2 O-linked Adduct Orientation is Affected by Aryl Moiety Shape

The glycosidic bond orientation is independent of the O-linked adduct, with  $\chi^{\text{avg}}$  being highly conserved in the B ( $230^\circ - 240^\circ$ ), S ( $45^\circ - 55^\circ$ ) and W ( $40^\circ - 50^\circ$ ) conformations (Table C.3, Appendix C). Within the B conformation, the O-linked adducts without an  $\alpha$ -fused ring (<sup>OPh</sup>dG, <sup>2ONP</sup>dG, <sup>OBP</sup>dG and <sup>3OPHN</sup>dG) are less planar ( $\theta^{\text{avg}}$  of  $142.4^\circ$  to  $147.6^\circ$ ) than other (<sup>1ONP</sup>dG, <sup>1OPHN</sup>dG and <sup>OPy</sup>dG) adducts ( $\theta^{\text{avg}}$  of  $160.2^\circ$  to  $162.6^\circ$ , Table C.3, Appendix C). However, in the W conformation, the presence of an  $\alpha$ -fused ring has

the opposite effect, with O-linked adducts lacking the  $\alpha$ -ring being less planar ( $\theta^{\text{avg}}$  of  $196.0^\circ$  to  $198.9^\circ$ ) than those with an  $\alpha$ -fused ring ( $\theta^{\text{avg}}$  of  $176.6^\circ$  to  $187.3^\circ$ ). Within the S conformation, there is no clear trend in lesion planarity with respect to bulky group size or shape ( $\theta^{\text{avg}}$  between  $173.7^\circ$  and  $207.1^\circ$ ). Notably, the unfavourable steric interactions for  $\theta = 0^\circ \pm 90^\circ$  seen for the DFT models (Figures C.18, C.23, Appendix C) are consistent with the  $\theta$  values seen in the helical models. Unlike  $\theta$ , the average  $\phi$  ( $\phi^{\text{avg}}$ ) in O-linked adducted DNA is only consistent for the B conformer ( $\phi^{\text{avg}} = 81^\circ \pm 10^\circ$ ), varying in the S ( $7.8^\circ$  to  $325.5^\circ$ ) and W ( $321.7^\circ$  to  $47.7^\circ$ ) conformations with no straightforward lesion size or shape effect.

#### *4.3.4.3 The Strength of Nonbonded Lesion-Site Interactions are Dependent on the Bulky Group Size and Shape for O-linked Adducts, Especially in the syn Orientation*

DNA in the B conformation containing O-linked adducts have comparable nonbonded interactions, with a  $\text{LSIE}^{\text{avg}}$  of  $-490 \pm 4$  kJ/mol (Figure 4.7, Table C.6, Appendix C). Within the W conformation, lesions containing a substituted ring ( $^{\text{OBP}}\text{dG}$ ,  $^{1\text{OPHN}}\text{dG}$ ,  $^{3\text{OPHN}}\text{dG}$  and  $^{\text{OPy}}\text{dG}$ ) have stronger adduct-site interactions ( $-393$  to  $-400$  kJ/mol) than those lacking ring substitution ( $-380$  to  $-387$  kJ/mol, Figure 4.7, Table C.6, Appendix C). Beyond this aryl-moiety shape effect, in the B and W conformations the O-linked adducts have no clear lesion shape trends.

For the O-linked adducts in the S DNA conformation, the one-ringed  $^{\text{OPh}}\text{dG}$  lesion results in the smallest LSIE ( $-265.4$  kJ/mol, Figure 4.7, Table C.6, Appendix C). Notably, adding a second ring strengthens the nonbonded interactions, with the addition of an  $\alpha$ -fused ring ( $^{1\text{ONP}}\text{dG}$ ) being less stabilizing ( $-28.6$  kJ/mol) than a  $\beta$ -fused ring ( $^{2\text{ONP}}\text{dG} =$

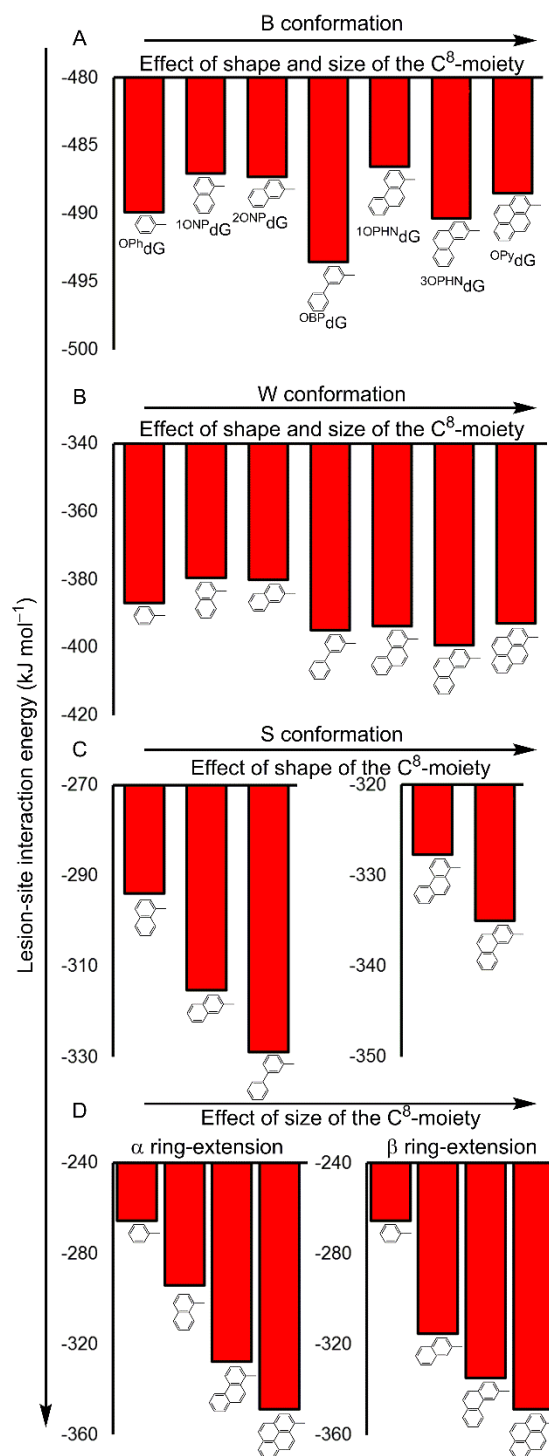


Figure 4.7: Effect of C8-moiety A) size and shape in the B conformation, B) size and shape in the W conformation, C) shape in the S conformation, and D) size in the S conformation on the lesion-site interaction energy (LSIE) for damaged DNA containing an O-linked dG adduct.

–49.8 kJ/mol) or ring substitution ( $^{OBP}dG = -63.5$  kJ/mol). The three ringed phenanthrenyl systems further increase the stability (by –62.3 kJ/mol and –69.6 kJ/mol for  $^{1OPHN}dG$  and  $^{3OPHN}dG$ , respectively). Lastly, the four-ringed  $^{OPy}dG$  adduct leads to the greatest increase in LSIE (–83.3 kJ/mol) out of the seven O-linked systems considered. Overall, there is a clear dependence of the LSIE for the S conformer on the bulky group size (Figure 4.7D) and shape (Figure 4.7C).

#### *4.3.4.4 The Lesion Size Rather Than Shape Primarily Determines Solvent Exposure of the Aryl Group for O-linked Adducts*

Since the O-linked and C-linked adducts both have the same hydrophobic aryl substituents, a larger SASA will similarly result in duplex destabilization for the O-linked adducts as discussed for the C-linked lesions. The single-ringed  $^{OPh}dG$  lesion has the smallest SASA among the O-linked adducts in the B (118.5 Å<sup>2</sup>), S (29.4 Å<sup>2</sup>) and W (64.4 Å<sup>2</sup>) conformations (Figure 4.8, Table C.5, Appendix C). For the B and W conformations, the  $^{1ONP}dG$  and  $^{2ONP}dG$  adducts are 25 to 32 Å<sup>2</sup> more solvent exposed than  $^{OPh}dG$ . Interestingly, the largest  $^{OPy}dG$  adduct is the least solvent exposed of the four remaining adducts in the B (179.3 Å<sup>2</sup>) and W (120.3 Å<sup>2</sup>) orientations, whereas  $^{OBP}dG$  is the most solvent exposed (187.8 Å<sup>2</sup> for the B conformer and 139.9 Å<sup>2</sup> for the W). Notably, the O-linked phenanthrenyl adducts have similar SASA values when comparing the  $\alpha$ -fused adducts to the  $\beta$ -fused version in the B (~ 183 Å<sup>2</sup>) and W (~ 125 Å<sup>2</sup>) conformations.

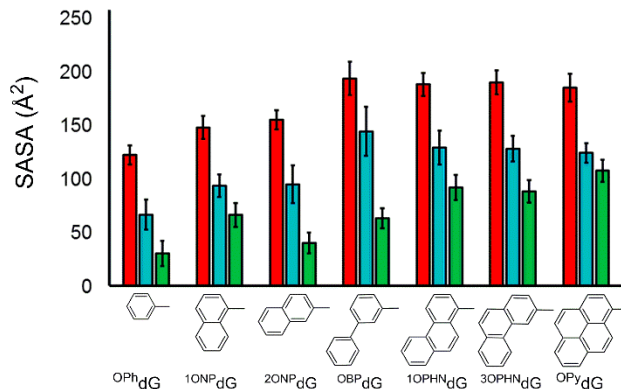


Figure 4.8: Dependence of the solvent accessible surface area (SASA, Å<sup>2</sup>) on the shape and size of the hydrophobic NCA for the B (red), W (blue) and S (green) conformations of damaged DNA containing an O-linked dG adduct.

O-linked adducted DNA in the S conformation shows a clear size pattern on the adduct, with the degree of solvent exposure being proportional to the number of aromatic rings in the lesion (Figure 4.8, Table C.5, Appendix C). Specifically, one-ringed <sup>OPh</sup>dG is the most solvent shielded, the four-ringed <sup>OPy</sup>dG adduct is the most solvent exposed (104.2 Å<sup>2</sup>), and the three-ringed <sup>1OPHN</sup>dG (89.2 Å<sup>2</sup>) and <sup>3OPHN</sup>dG (85.6 Å<sup>2</sup>) adducts are the next most exposed adducts in the S conformation. Out of the two-ringed adducts, <sup>2ONP</sup>dG is the most solvent shielded (38.9 Å<sup>2</sup>), while <sup>1ONP</sup>dG and <sup>OBP</sup>dG have similar solvent exposure (~63 Å<sup>2</sup>). Notably, the least shielded adduct in the S conformation (<sup>OPy</sup>dG) is less solvent exposed than the one-ringed <sup>OPh</sup>dG adduct in the B conformation. Overall, the SASA of the O-linked adducts is primarily dictated by lesion size with the shape having minimal impact.

#### 4.3.4.5 The Shape and Size of the O-linked Adduct Impacts the Helical Conformational Preferences

Analysis of both the stabilizing nonbonded interaction at the lesion site and destabilizing hydrophobic effect of the hydrocarbon bulky moiety can be used to suggest

the conformational preferences of O-linked adducted DNA (Figure 4.9). For the O-linked adducts, adducted DNA in the W conformation lacks the large degree of solvent shielding of the S conformation (Figure 4.8) and the strong lesion-site interactions of the B conformation (Figure 4.7), suggesting that the W conformation will not be a dominant conformation of O-linked adducted DNA as seen for most C8-dG lesions.<sup>1, 10, 14, 19, 51</sup> Despite this, due to the O-linked adducts having reasonable lesion-site interaction strengths ( $\sim -390$  kJ/mol) and aryl-group solvent exposures (less than in the B conformation but greater than S) compared to the C-linked and N-linked versions of the adducts,<sup>14</sup> it is plausible that these adducts may stabilize W DNA conformation for a greater occupancy than the other linker types when paired with a non-canonical nucleotide.

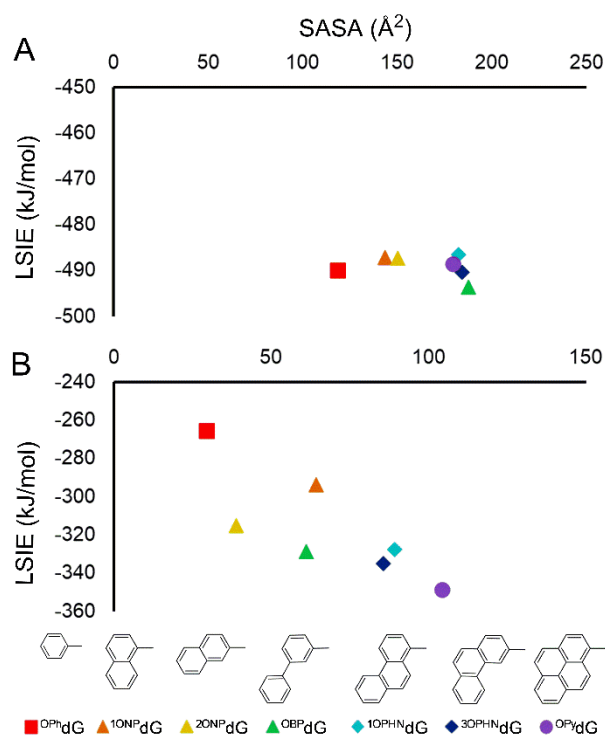


Figure 4.9: Dependence of the solvent accessible surface area (SASA, Å²) and lesion-site interaction energy (LSIE, kJ/mol) on the shape and size of the hydrophobic C8-moiety for the (A) B and (B) S conformations of DNA containing an O-linked dG adduct.

Due to consistent LSIEs, the stability of the B DNA conformer of O-linked adducted DNA can be subdivided into three groups based on the adduct SASAs, which is directly related to the size of the bulky group. Specifically, in the B conformation, the one ringed <sup>OPh</sup>dG is the least destabilized by bulky group hydrophobic interactions, which is followed by the naphthyl derived adducts, while the remaining four adducts have the highest solvent exposure levels. In the S conformation, both <sup>OPh</sup>dG and <sup>1ONP</sup>dG result in the weakest interactions at the lesion site, suggesting that <sup>OPh</sup>dG and <sup>1ONP</sup>dG will exclusively adopt the B conformation. In contrast, the largest <sup>OPy</sup>dG adduct has the strongest nonbonded adduct-site interactions in the S conformation and is significantly more solvent shielded than in the B conformation, indicating that the four-ringed adduct will extensively stabilize the S conformation. The <sup>3OPHN</sup>dG adduct also has large LSIE in the S conformation, being only ~ 15 kJ/mol less than that of <sup>OPy</sup>dG, and the bulky group is significantly solvent shielded in the S conformer compared to in the major groove orientation. Thus, <sup>3OPHN</sup>dG adducted DNA likely adopts the S conformation to a similar extent as DNA containing the largest <sup>OPy</sup>dG. In contrast, <sup>1OPHN</sup>dG adducted DNA will likely have a smaller S preference than <sup>3OPHN</sup>dG because of a smaller LSIE compared to <sup>1OPHN</sup>dG, although <sup>1OPHN</sup>dG will still predominately induce the S conformation. Finally, <sup>2ONP</sup>dG and <sup>OBP</sup>dG likely induce conformational heterogeneity. DNA containing the  $\beta$ -fused ring naphthyl adduct probably adopts the B conformation in most cases due to the <sup>2ONP</sup>dG adduct being better shielded from solvent and having stronger nonbonded interactions in the S conformation. Despite the <sup>OBP</sup>dG containing DNA having high LSIE values in the B conformation, the biphenyl moiety has a large degree of solvent shielding in the S conformation, and a large LSIE (–329 kJ/mol), suggesting that damaged DNA containing <sup>OBP</sup>dG will likely induce the S conformation of adducted DNA more than the B. Thus, the B:S ratio of O-linked adducted



DNA likely decreases with helices containing  ${}^{\text{OPh}}\text{dG} \approx {}^{\text{1ONP}}\text{dG} \gg {}^{\text{2ONP}}\text{dG} \gg {}^{\text{OBP}}\text{dG} \gg {}^{\text{1OPHN}}\text{dG} > {}^{\text{3OPHN}}\text{dG} \approx {}^{\text{OPy}}\text{dG}$ .

## 4.4 Discussion

### 4.4.1 The Linker Type has a Subtle Effect on the Preferred Damaged DNA Conformation Due to the Unique Inherent Structural Features of C, O and N-linked C8-dG Adducts

Comparison of the inherent structural properties of the systematic set of C and O-linked adducts considered in the present work coupled with those previously uncovered for the analogous N-linked variants<sup>14</sup> can provide important insights into the dependence of the damaged DNA conformations on the nature of the chemical linker between the dG nucleobase and the bulky aryl moiety. The O and N-linked adducts are structurally related due to the presence of a heteroatom in the nucleobase–bulky group linker. As a result, the aryl moieties in the O and N-linked nucleobase and nucleoside adducts possess similar conformational flexibility with respect to the nucleobase. Nevertheless, the amine linker affords unique noncovalent interactions compared to an oxygen atom, which stabilizes different conformations of the aryl group relative to the nucleobase. Indeed, the amine linker forms a N–H•••O5' hydrogen bond at the nucleoside level,<sup>14</sup> while the corresponding orientations of the O-linked adducts are unstable due to intramolecular repulsion involving the linker oxygen atom (Figures C.18-C.27, Appendix C). The C-linked adducts are distinct from the O and N-linked variants due to the direct covalent attachment of the aryl group to C8 of G, which reduces the number of energetically accessible nucleobase and nucleoside conformations. Regardless, all N, O and C-linked nucleoside adducts are preferentially stabilized in the *anti* glycosidic conformation and have thermodynamically and kinetically

accessible *syn* orientations, suggesting that multiple conformations of damaged DNA may be possible for all linker types.

Although the global structural features of the most prevalent B and S conformations of damaged DNA are preserved for all adducts regardless of the linker type, the differences in the inherent structural features of the C, O and N-linked nucleobase and nucleoside adducts are maintained in the DNA helix. As a result, despite the inherent flexibility of the adducts, the bulky aryl group is differentially positioned with respect to the damaged dG depending on the nucleobase–aryl moiety linker type. This in turn results in subtle differences in the stabilizing noncovalent interactions at the damage site and destabilizing hydrophobic interactions of the aryl group with the identity of the linker (Figure 4.10, Tables C.7, C.8, Appendix C). For example, a direct C–C linkage of the bulky moiety to dG reduces the overall adduct size for the C-linked variants regardless of the aryl group or (B/S) damaged DNA conformation, which results in the bulky moiety being less solvent exposed (Figure 4.10A,B, Table C.7, Appendix C). While the magnitude of the lesion-site interactions in the B helical conformation are similar for C and N-linked adducts (Figure 4.10C, Table C.8, Appendix C), the reduced flexibility about the nucleobase–aryl group C-linkage leads to a stark reduction in the bulky group stacking potential in the S conformation (Figure 4.10D, Table C.8, Appendix C), suggesting that the C-linked adducts are less likely to induce the S conformation of damaged DNA than the N-linked counterparts. Notably, work on other C8-dG lesions has shown the reverse effect for the analogous (4-) substituted N-linked <sup>FABP</sup>dG and C-linked <sup>FBP</sup>dG adducted helices, which exhibit a greater B conformational preference for the N-linked (67%) than C-linked (20%) lesions. Despite similar linker lengths and flexibility as well as solvent accessibility of the

aryl group, the unique orientation of the bulky moiety results in slightly greater lesion-site interactions for the O-linked adducts compared to the N-linked variants in the B conformation, but significantly reduced lesion-site nonbonded interactions in the S conformation of damaged DNA, especially for the smaller one and two-ringed adducts. Therefore, the S conformation of O-linked damaged DNA is likely less favorable than for N-linked damaged DNA. Although the lesion-site interactions are consistently stronger for the O-linked adducts than the C-linked analogues in both the B and S conformations, C-linked adducts built based on ring substitution (i.e., biphenyl, phenanthrenyl and pyrenyl systems) are more solvent shielded in the S conformation than the O-linked counterparts, which suggests a marginally greater S preference for larger O-linked adducts compared to the C-linked versions. Thus, the differential intramolecular interactions in the nucleoside units arising from changes in adduct flexibility and/or hydrogen-bonding properties of the linkage result in a linker-dependent nucleobase–aryl group relative orientation in DNA duplexes, which in turn impacts the relative populations of the B and S conformations of damaged DNA. Furthermore, the O-linked adducts have stronger LSIEs and lower SASAs in the W conformation compared to either the C-linked or N-linked versions,<sup>14</sup> suggesting that in mispaired sequences the occupancy of the W conformation will be highest for DNA containing O-linked adducts. Overall, this suggests that the highest occupancy of the S conformation will be for DNA containing N-linked adducts, followed by O-linked adducts, with C-linked lesions having the greatest propensity for the B conformer.

#### 4.4.2 The Shape of the Aryl Group Affects the Conformations Intrinsically Accessible to the C8-dG Adducts and therefore the Local Helical Structure of Damaged DNA

The model adducts considered in the present study afford the unique opportunity to systematically investigate the impact of lesion shape on the adducted DNA structure (Figure 4.2). Specifically,  $\alpha$  and  $\beta$  ring fusion was considered for adducts with two or three fused rings, while the biphenyl lesions allow comparison of ring fusion and ring substitution. In isolated nucleobase and nucleoside adducts, the presence of an  $\alpha$ -fused ring heavily impacts the structural landscape due to steric interactions with the damaged nucleobase and/or sugar moiety (Figures C.1, C.6, C.18, C.23, Appendix C).<sup>14</sup> This translates to restricted conformations about the nucleobase–aryl moiety linker (i.e.,  $\theta = 180^\circ$  is inaccessible for C-linked adducts and  $\phi = 180^\circ$  is inaccessible for O-linked and N-linked analogues),<sup>14</sup> which is preserved in DNA duplexes. Therefore, although the solvent accessible surface area is generally highly similar regardless of ring fusion, the lesion shape affects the magnitude of the lesion-site interactions in both the damaged DNA B and S conformations. However, the impact of the directionality of ring expansion is dependent on linker type. Specifically, for the C-linked adducts, the stabilizing lesion-site interactions in the B conformation of adducted DNA are slightly less for systems with an  $\alpha$ -fused ring than a  $\beta$ -fused ring, while the noncovalent interactions are highly similar in the S conformer, suggesting that damaged DNA containing a  $\beta$ -fused system (<sup>2NP</sup>dG and <sup>3PHN</sup>dG) may have a slightly greater B:S ratio. In contrast, O-linked adducts with  $\beta$ -fused rings significantly increase the lesion-site interactions in the S conformation and decrease aryl group solvation (especially for the naphthyl moiety), which decreases the B:S ratio for the corresponding damaged DNA. The shape dependence of the damaged DNA conformation is much more

complicated for N-linked adducts,<sup>14</sup> with the lesion size also playing a role. Specifically, although there is essentially no shape impact on the lesion-site interactions in the B conformation of damaged DNA,  $\beta$  fusion of two rings or  $\alpha$  fusion of three rings increases the lesion-site interactions in the S conformation. In contrast to aryl group expansion through ring fusion, ring substitution of the single ring systems to yield the biphenyl adducts results in similar solvent exposure to the largest lesions for B damaged DNA and solvent exposure more similar to that of the two ring fused systems for S damaged DNA for all linkage types (Figure 4.10, Table C.7, Appendix C). Furthermore, for the C and O-linked lesions, although the biphenyl substituent results in only slightly greater lesion-site interactions than ring fusion in the B damaged DNA conformer, the interactions are significantly greater in the S conformation, suggesting an increase in the presence of the S conformer upon adduct expansion through ring substitution compared to ring fusion. Although the S conformer is likely also more prevalent for N-linked biphenyl lesions than naphthyl aryl groups, the behavior is likely more similar to  $\alpha$ -ring fusion due to increased solvent exposure of the aryl group in the B conformer of damaged DNA and decreased stabilizing lesion-site interactions in the S conformer for <sup>3</sup>ABP dG compared to <sup>1</sup>ANP dG and <sup>2</sup>ANP dG. Overall, the shape of the bulky group can significantly impact the lesion site interactions, especially in the S conformer of damaged DNA, which in turn affects the conformational heterogeneity of adducted DNA.

#### **4.4.3 The Bulky Group Size has an Even Larger Effect on Damaged DNA Structure than Shape and Linker Type**

The model lesions investigated in the present work systematically increase the size

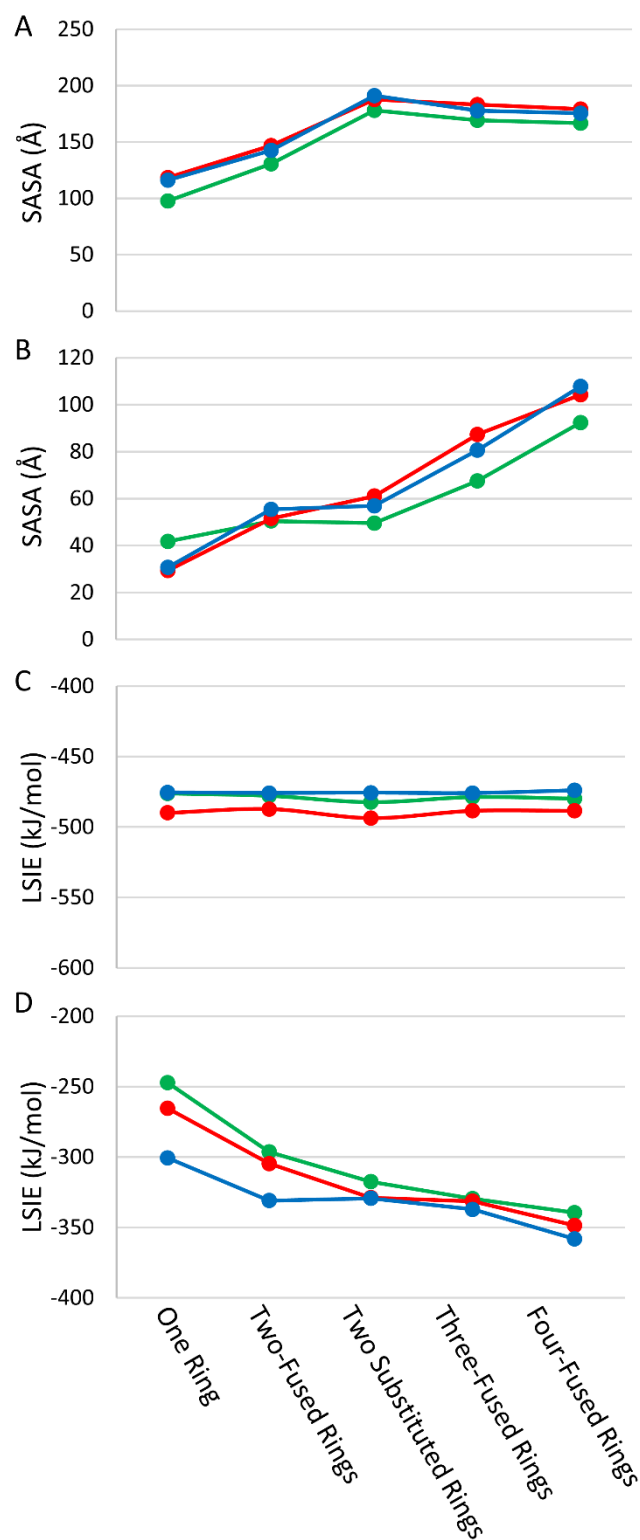


Figure 4.10: Effect of aryl moiety size on the SASA (A and B) and LSIE (C and D) for the B (A and C) and S (B and D) conformations, averaged over C8-dG adducts with a C-linker (Green), O-linker (Red) or N-linker (Blue).

of the aryl moiety from a single phenyl ring to the four-ringed pyrenyl group. With the exception of the biphenyl moiety, the larger adducts are generated by extending the smallest system through  $\alpha$  (2,3-) or  $\beta$  (3,4-) fusion of two, three, or four rings. No major structural deviations occur to the damaged DNA upon increasing the bulky group size. In the B conformation of damaged DNA, the stabilizing lesion-site interactions are negligibly impacted by the aryl ring size, while there is an increase in the destabilizing hydrophobic interactions as a single ring bulky moiety is expanded to two or three fused rings, with the addition of a fourth ring not further increasing the solvent exposure. Although the adduct solvent exposure does not rise as sharply with bulky moiety expansion for the S conformation, the stabilizing lesion-site interactions significantly increase with an increase in the aryl group size (Figure 4.10, Tables C.7, C.8, Appendix C). However, the increase in the lesion-site stability due to ring expansion is dependent on the linker type, decreasing according to C (–92 kJ/mol) > O (–83 kJ/mol) > N (–58 kJ/mol). Nevertheless, regardless of linker type, damaged DNA containing a single-ring adduct likely exclusively adopts the B conformation, while the S conformer will be preferred for DNA containing the largest four-ringed adduct. Furthermore, the B:S ratio will decrease as the phenyl adducts are increased in size through two- or three-ring fusion. This ratio change is primarily due to an increase in the magnitude of the lesion-site interaction energies for the S conformer as the phenyl group is expanded to naphthyl (average increase across all systems of ~ 40 kJ/mol) to phenanthryl (average increase of approximately 22 kJ/mol) and finally pyrenyl (average increase of 16 kJ/mol, Table C.8, Appendix C). Notably, by systematically increasing the number of aryl rings in the bulky group, our work complements and correlates with previous work on N-linked and C-linked adducts that show that larger bulky groups decrease the B:S conformational ratio for C8-dG adducts.<sup>13, 14, 19</sup> In summary, despite the

dependence of the adducted DNA conformation on the lesion shape and linker type, the size of the aryl moiety clearly drives the B:S ratio, which is smallest for the largest lesions.

#### 4.5 Conclusion

In summary, the present work used a combination of quantum mechanics and molecular dynamics simulation to investigate 14 unique C-linked and O-linked C8-dG adducts. When coupled with previous conclusions on the 7 analogous N-linked adducts,<sup>14</sup> this comprehensive systematic series of model lesions provides a thorough understanding of the effect of the bulky moiety shape and size, and the nucleobase–aryl moiety linker type on the structure and conformational preference of adducted DNA. The C-linked lesions are inherently less flexible due to the absence of a linker atom between the aryl group and G, whereas the O-linked adducts have fewer stable minima compared to the N-linked systems due to changes in the hydrogen-bonding properties of the linker. These differences are preserved in the DNA helices and impact the stabilizing lesion site interactions and destabilizing bulky group solvent exposure, which in turn shifts in the helical conformational heterogeneity such that N-linked and O-linked adducts have a greater preference for the S and W conformers, respectively. We theorize that the greater W stability could partly arise from the lack of hydrogen bonding donor on the linking atom, as the N-linked <sup>AA</sup>FdG adduct that is known to adopt the W conformer<sup>10, 18</sup> has an acetyl group on the linking amine. Despite the shape of the bulky group, specifically the presence of an  $\alpha$ - or  $\beta$ -fused ring, significantly impacting the dihedral angle preferences that dictate the position of the aryl group with respect to the damaged nucleobase, the conformational stabilities are generally independent of bulky moiety shape. Comparatively, the number of aryl rings in the bulky group generally does not affect the helical structure, but does



significantly decrease the B:S conformational ratio as the aryl ring size increases. Thus, this work adds to the current literature on DNA adducts by underscoring the impact of lesion size and shape on helical conformation and adds the aryl–nucleobase linker type to the list of adduct features that affect adducted DNA structure, which in turn have significant biological consequences including changing the repair and replication outcomes.

Overall, this comprehensive incremental series of model adducts can be used in future works to better understand the interactions between adducted DNA and various cellular processes. Specifically, upcoming studies can selectively choose adducts that cover a wide range of B:S ratios to study a variety of repair and replication outcomes. By controlling the structure of adducted DNA, valuable insight into the relationship between adduct structure and biological outcomes, including specific mechanisms and potential outcomes of DNA repair and replication, will be obtained.

#### 4.6 References

1. Wilson, K. A.; Kung, R. W.; Wetmore, S. D., Chapter Seven - Toxicology of DNA Adducts Formed Upon Human Exposure to Carcinogens: Insights Gained From Molecular Modeling. In *Advances in Molecular Toxicology*, Fishbein, J. C.; Heilman, J. M., Eds. Elsevier: 2016; Vol. 10, pp 293-360.
2. Helleday, T.; Petermann, E.; Lundin, C.; Hodgson, B.; Sharma, R. A., DNA repair pathways as targets for cancer therapy. *Nat. Rev. Cancer* **2008**, 8 (3), 193-204.
3. Aziz Sancar; Laura A. Lindsey-Boltz; Keziban Ünsal-Kaçmaz; Linn, S., Molecular Mechanisms of Mammalian DNA Repair and the DNA Damage Checkpoints. *Annual Review of Biochemistry* **2004**, 73 (1), 39-85.
4. Cancer, I. A. f. R. o., IARC monographs on the evaluation of carcinogenic risks to humans. *IARC Monogr. Eval. Carcinog. Risks Hum.* **2008**, 97, 3-471.
5. Munia, A.; Giese, R. W.; Polvani, S.; Galli, A.; Cellai, F.; Peluso, M. E. M., Chapter Six - Bulky DNA Adducts, Tobacco Smoking, Genetic Susceptibility, and Lung Cancer Risk. In *Advances in Clinical Chemistry*, Makowski, G. S., Ed. Elsevier: 2017; Vol. 81, pp 231-277.

6. Wakabayashi, K.; Nagao, M.; Esumi, H.; Sugimura, T., FOOD-DERIVED MUTAGENS AND CARCINOGENS. *Cancer Research* **1992**, 52 (7), S2092-S2098.
7. Cavalieri, E. L.; Rogan, E. G., Depurinating estrogen-DNA adducts, generators of cancer initiation: their minimization leads to cancer prevention. *Clin. Transl. Med.* **2016**, 5, 15.
8. Feigelson, H. S.; Henderson, B. E., Estrogens and breast cancer. In *Endocrinology of Breast Cancer*, Springer: 1999; pp 55-67.
9. Kung, R. W.; Sharma, P.; Wetmore, S. D., Effect of Size and Shape of Nitrogen-Containing Aromatics on Conformational Preferences of DNA Containing Damaged Guanine. *J Chem Inf Model* **2018**, 58 (7), 1415-1425.
10. Jain, V.; Vaidyanathan, V. G.; Patnaik, S.; Gopal, S.; Cho, B. P., Conformational Insights into the Lesion and Sequence Effects for Arylamine-Induced Translesion DNA Synthesis: 19F NMR, Surface Plasmon Resonance, and Primer Kinetic Studies. *Biochemistry* **2014**, 53 (24), 4059-4071.
11. Millen, A. L.; McLaughlin, C. K.; Sun, K. M.; Manderville, R. A.; Wetmore, S. D., Computational and Experimental Evidence for the Structural Preference of Phenolic C-8 Purine Adducts. *The Journal of Physical Chemistry A* **2008**, 112 (16), 3742-3753.
12. Mu, H.; Zhang, Y.; Geacintov, N. E.; Broyde, S., Lesion Sensing during Initial Binding by Yeast XPC/Rad4: Toward Predicting Resistance to Nucleotide Excision Repair. *Chem Res Toxicol* **2018**, 31 (11), 1260-1268.
13. Berger, F. D.; Sturla, S. J.; Kung, R. W.; Montana, T.; Wetmore, S. D.; Manderville, R. A., Conformational Preference and Fluorescence Response of a C-Linked C8-Biphenyl-Guanine Lesion in the NarI Mutational Hotspot: Evidence for Enhanced Syn Adduct Formation. *Chem. Res. Toxicol.* **2018**, 31 (1), 37-47.
14. Kung, R. W.; Sharma, P.; Wetmore, S. D., Effect of Size and Shape of Nitrogen-Containing Aromatics on Conformational Preferences of DNA Containing Damaged Guanine. *J. Chem. Inf. Model.* **2018**, 58 (7), 1415-1425.
15. Sproviero, M.; Verwey, A. M.; Rankin, K. M.; Witham, A. A.; Soldatov, D. V.; Manderville, R. A.; Fekry, M. I.; Sturla, S. J.; Sharma, P.; Wetmore, S. D., Structural and Biochemical Impact of C8-Aryl-guanine Adducts within the NarI Recognition DNA Sequence: Influence of Aryl Ring Size on Targeted and Semi-Targeted Mutagenicity. *Nucleic Acids Res.* **2014**, 42 (21), 13405-13421.
16. Millen, A. L.; Sharma, P.; Wetmore, S. D., C8-Linked Bulky Guanosine DNA Adducts: Experimental and Computational Insights into Adduct Conformational Preferences and Resulting Mutagenicity. *Future Med. Chem.* **2012**, 4 (15), 1981-2007.

17. Patnaik, S.; Cho, B. P., Structures of 2-Acetylaminofluorene Modified DNA Revisited: Insight into Conformational Heterogeneity. *Chem. Res. Toxicol.* **2010**, *23* (11), 1650-1652.
18. Cho, B. P.; Zhou, L., Probing the Conformational Heterogeneity of the Acetylaminofluorene-Modified 2'-Deoxyguanosine and DNA by 19F NMR Spectroscopy. *Biochemistry* **1999**, *38* (23), 7572-7583.
19. Shapiro, R.; Ellis, S.; Hingerty, B. E.; Broyde, S., Effect of Ring Size on Conformations of Aromatic Amine-DNA Adducts: The Aniline-C8 Guanine Adduct Resides in the B-DNA Major Groove. *Chem. Res. Toxicol.* **1998**, *11* (4), 335-341.
20. Millen, A. L.; McLaughlin, C. K.; Sun, K. M.; Manderville, R. A.; Wetmore, S. D., Computational and experimental evidence for the structural preference of phenolic C-8 purine adducts. *The Journal of Physical Chemistry A* **2008**, *112* (16), 3742-53.
21. Mantle, P. G.; Faucet-Marquis, V.; Manderville, R. A.; Squillaci, B.; Pfohl-Leszkowicz, A., Structures of covalent adducts between DNA and ochratoxin A: A new factor in debate about genotoxicity and human risk assessment. *Chem. Res. Toxicol.* **2009**, *23* (1), 89-98.
22. Guyton, K. Z.; Loomis, D.; Grosse, Y.; El Ghissassi, F.; Bouvard, V.; Benbrahim-Tallaa, L.; Guha, N.; Mattock, H.; Straif, K., Carcinogenicity of pentachlorophenol and some related compounds. *The Lancet Oncology* **2016**, *17* (12), 1637-1638.
23. Humans, I. W. G. o. t. E. o. C. R. t.; Cancer, I. A. f. R. o.; Organization, W. H., *Re-evaluation of Some Organic Chemicals, Hydrazine and Hydrogen Peroxide*. IARC: 1999.
24. Cancer, I. A. f. R. o., *Some aromatic amines, organic dyes, and related exposures*. IARC Press, International Agency for Research on Cancer: 2010; Vol. 99.
25. Cancer, I. A. f. R. o., Diesel and gasoline engine exhausts and some nitroarenes. *IARC monographs on the evaluation of carcinogenic risks to humans* **1988**, *105*.
26. Sancar, A.; Reardon, J. T., Nucleotide Excision Repair in E. Coli and Man. In *Advances in Protein Chemistry*, Academic Press: 2004; Vol. 69, pp 43-71.
27. Sale, J. E.; Lehmann, A. R.; Woodgate, R., Y-family DNA Polymerases and their Role in Tolerance of Cellular DNA Damage. *Nature Reviews Molecular Cell Biology* **2012**, *13* (3), 141-152.
28. Min, J.-H.; Pavletich, N. P., Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature* **2007**, *449* (7162), 570-575.

29. Cai, Y.; Patel, D. J.; Geacintov, N. E.; Broyde, S., Differential nucleotide excision repair susceptibility of bulky DNA adducts in different sequence contexts: Hierarchies of recognition signals. *J. Mol. Biol.* **2009**, *385* (1), 30-44.
30. Hess, M. T.; Schwitter, U.; Petretta, M.; Giese, B.; Naegeli, H., Bipartite substrate discrimination by human nucleotide excision repair. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (13), 6664.
31. Mu, H.; Zhang, Y.; Geacintov, N. E.; Broyde, S., Lesion Sensing during Initial Binding by Yeast XPC/Rad4: Toward Predicting Resistance to Nucleotide Excision Repair. *Chem. Res. Toxicol.* **2018**, *31* (11), 1260-1268.
32. Mu, H.; Geacintov, N. E.; Min, J.-H.; Zhang, Y.; Broyde, S., Nucleotide Excision Repair Lesion-Recognition Protein Rad4 Captures a Pre-Flipped Partner Base in a Benzo[a]pyrene-Derived DNA Lesion: How Structure Impacts the Binding Pathway. *Chem. Res. Toxicol.* **2017**, *30* (6), 1344-1354.
33. Reeves, D. A.; Mu, H.; Kropachev, K.; Cai, Y.; Ding, S.; Kolbanovskiy, A.; Kolbanovskiy, M.; Chen, Y.; Krzeminski, J.; Amin, S.; Patel, D. J.; Broyde, S.; Geacintov, N. E., Resistance of bulky DNA lesions to nucleotide excision repair can result from extensive aromatic lesion-base stacking interactions. *Nucleic Acids Res.* **2011**, *39* (20), 8752-8764.
34. Yang, W., Poor base stacking at DNA lesions may initiate recognition by many repair proteins. *DNA Repair* **2006**, *5* (6), 654-666.
35. Kropachev, K.; Ding, S.; Terzidis, M. A.; Masi, A.; Liu, Z.; Cai, Y.; Kolbanovskiy, M.; Chatgililoglu, C.; Broyde, S.; Geacintov, N. E.; Shafirovich, V., Structural basis for the recognition of diastereomeric 5',8-cyclo-2'-deoxypurine lesions by the human nucleotide excision repair system. *Nucleic Acids Res.* **2014**, *42* (8), 5020-5032.
36. Geacintov, N. E.; Broyde, S.; Buterin, T.; Naegeli, H.; Wu, M.; Yan, S.; Patel, D. J., Thermodynamic and structural factors in the removal of bulky DNA adducts by the nucleotide excision repair machinery. *Biopolymers* **2002**, *65* (3), 202-210.
37. Xu, L.; Cho, B. P., Conformational Insights into the Mechanism of Acetylaminofluorene-dG-Induced Frameshift Mutations in the NarI Mutational Hotspot. *Chem. Res. Toxicol.* **2016**, *29* (2), 213-226.
38. Velmurugu, Y.; Chen, X.; Slogoff Sevilla, P.; Min, J. H.; Ansari, A., Twist-open mechanism of DNA damage recognition by the Rad4/XPC nucleotide excision repair complex. *Proc Natl Acad Sci U S A* **2016**, *113* (16), E2296-305.
39. Velmurugu, Y.; Chen, X.; Slogoff Sevilla, P.; Min, J.-H.; Ansari, A., Twist-open mechanism of DNA damage recognition by the Rad4/XPC nucleotide excision repair complex. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (16), E2296.

40. Broyde, S.; Hingerty, B. E.; Shapiro, R.; Norman, D., Unusual hydrogen bonding patterns in 2-aminofluorene (AF) and 2-acetylaminofluorene (AAF) modified DNA. *Environ. Sci. Res.* **1990**, *40* (Nitroarenes), 113-23.
41. Cai, Y.; Kropachev, K.; Xu, R.; Tang, Y.; Kolbanovskii, M.; Kolbanovskii, A.; Amin, S.; Patel, D. J.; Broyde, S.; Geacintov, N. E., Distant neighbor base sequence context effects in human nucleotide excision repair of a benzo[a]pyrene-derived DNA lesion. *J. Mol. Biol.* **2010**, *399* (3), 397-409.
42. Yang, S.; Diem, M.; Liu, J. D. H.; Wesseling, S.; Vervoort, J.; Oostenbrink, C.; Rietjens, I., Cellular levels and molecular dynamics simulations of estragole DNA adducts point at inefficient repair resulting from limited distortion of the double-stranded DNA helix. *Arch. Toxicol.* **2020**, *94* (4), 1349-1365.
43. Yang, S.; Liu, J. D. H.; Diem, M.; Wesseling, S.; Vervoort, J.; Oostenbrink, C.; Rietjens, I., Molecular Dynamics and In Vitro Quantification of Safrole DNA Adducts Reveal DNA Adduct Persistence Due to Limited DNA Distortion Resulting in Inefficient Repair. *Chem. Res. Toxicol.* **2020**, *33* (9), 2298-2309.
44. Cho, B. P.; Beland, F. A.; Marques, M. M., NMR structural studies of a 15-mer DNA sequence from a ras protooncogene, modified at the first base of codon 61 with the carcinogen 4-aminobiphenyl. *Biochemistry* **1992**, *31* (40), 9587-602.
45. Jain, N.; Li, Y.; Zhang, L.; Meneni, S. R.; Cho, B. P., Probing the Sequence Effects on NarI-Induced -2 Frameshift Mutagenesis by Dynamic <sup>19</sup>F NMR, UV, and CD Spectroscopy. *Biochemistry* **2007**, *46* (46), 13310-13321.
46. Norman, D.; Abuaf, P.; Hingerty, B. E.; Live, D.; Grunberger, D.; Broyde, S.; Patel, D. J., NMR and Computational Characterization of the N-(Deoxyguanosin-8-yl)aminofluorene Adduct [(AF)G] Opposite Adenosine in DNA: (AF)G[syn]·A[anti] Pair Formation and its pH Dependence. *Biochemistry* **1989**, *28* (18), 7462-7476.
47. Zhang, N.; Ding, S.; Kolbanovskiy, A.; Shastry, A.; Kuzmin, V. A.; Bolton, J. L.; Patel, D. J.; Broyde, S.; Geacintov, N. E., NMR and Computational Studies of Stereoisomeric Equine Estrogen-Derived DNA Cytidine Adducts in Oligonucleotide Duplexes: Opposite Orientations of Diastereomeric Forms. *Biochemistry* **2009**, *48* (30), 7098-7109.
48. Patel, D. J.; Mao, B.; Gu, Z.; Hingerty, B. E.; Gorin, A.; Basu, A. K.; Broyde, S., Nuclear Magnetic Resonance Solution Structures of Covalent Aromatic Amine–DNA Adducts and Their Mutagenic Relevance. *Chem. Res. Toxicol.* **1998**, *11* (5), 391-407.
49. Cai, A.; Wilson, K. A.; Patnaik, S.; Wetmore, S. D.; Cho, B. P., DNA base sequence effects on bulky lesion-induced conformational heterogeneity during DNA replication. *Nucleic Acids Res.* **2018**, *46* (12), 6356-6370.

50. Vaidyanathan, V. G.; Cho, B. P., Sequence effects on translesion synthesis of an aminofluorene-DNA adduct: Conformational, thermodynamic, and primer extension kinetic studies. *Biochemistry* **2012**, *51* (9), 1983-1995.
51. Cho, B. P.; Beland, F. A.; Marques, M. M., NMR structural studies of a 15-mer DNA duplex from a ras protooncogene modified with the carcinogen 2-aminofluorene: conformational heterogeneity. *Biochemistry* **1994**, *33* (6), 1373-84.
52. Mu, H.; Kropachev, K.; Wang, L.; Zhang, L.; Kolbanovskiy, A.; Kolbanovskiy, M.; Geacintov, N. E.; Broyde, S., Nucleotide excision repair of 2-acetylaminofluorene- and 2-aminofluorene-(C8)-guanine adducts: molecular dynamics simulations elucidate how lesion structure and base sequence context impact repair efficiencies. *Nucleic Acids Res.* **2012**, *40* (19), 9675-9690.
53. Mao, B.; Vyas, R. R.; Hingerty, B. E.; Broyde, S.; Basu, A. K.; Patel, D. J., Solution Conformation of the N-(Deoxyguanosin-8-yl)-1-aminopyrene ([AP]dG) Adduct Opposite dC in a DNA Duplex. *Biochemistry* **1996**, *35* (39), 12659-12670.
54. Millen, A. L.; Kamenz, B. L.; Leavens, F. M. V.; Manderville, R. A.; Wetmore, S. D., Conformational Flexibility of C8-Phenoxyguanine Adducts in Deoxydinucleoside Monophosphates. *The Journal of Physical Chemistry B* **2011**, *115* (44), 12993-13002.
55. Denington, R.; Keith, T.; Millam, J. M., Gaussview. **2009**.
56. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J., Gaussian 09, Revision B.01. Wallingford CT, 2009.
57. Mao, B.; Gorin, A.; Gu, Z.; Hingerty, B. E.; Broyde, S.; Patel, D. J., Solution structure of the aminofluorene-intercalated conformer of the syn [AF]-C8-dG adduct opposite a -2 deletion site in the NarI hot spot sequence context. *Biochemistry* **1997**, *36* (47), 14479-14490.
58. Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham, T. E.; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.;

Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A., {*Amber 14*}. 2014.

59. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, 25 (9), 1157-1174.

60. Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P., The R.E.D. Tools: Advances in RESP and ESP Charge Derivation and Force Field Library Building. *Phys. Chem. Chem. Phys.* **2010**, 12 (28), 7821-7839.

61. Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham, T. E.; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A., *Amber 14*. 2014.

62. Roe, D. R.; Cheatham, T. E., PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, 9 (7), 3084-3095.

63. Jain, N.; Meneni, S.; Jain, V.; Cho, B. P., Influence of flanking sequence context on the conformational flexibility of aminofluorene-modified dG adduct in dA mismatch DNA duplexes. *Nucleic Acids Research* **2009**, 37 (5), 1628-1637.

64. Abuaf, P.; Hingerty, B. E.; Broyde, S.; Grunberger, D., Solution Conformation of the N-(Deoxyguanosin-8-yl)aminofluorene Adduct opposite Deoxyinosine and Deoxyguanosine in DNA by NMR and Computational Characterization. *Chem. Res. Toxicol.* **1995**, 8 (3), 369-378.

65. Norman, D.; Abuaf, P.; Hingerty, B. E.; Live, D.; Grunberger, D.; Broyde, S.; Patel, D. J., NMR and computational characterization of the N-(deoxyguanosin-8-yl)aminofluorene adduct [(AF)G] opposite adenosine in DNA: (AF)G[syn].cntdot.A[anti] pair formation and its pH dependence. *Biochemistry* **1989**, 28 (18), 7462-7476.

## Chapter 5: Conformational Preference and Fluorescence Response of a C-Linked C8-Biphenyl-Guanine Lesion in the *NarI* Mutational Hotspot: Evidence for Enhanced *Syn* Adduct Formation<sup>a,b,c</sup>

### 5.1 Introduction

Endogenous and exogenous electrophiles can damage DNA to produce bulky covalent DNA adducts (addition products). If adducts evade DNA repair, they may initiate mutagenesis and ultimately carcinogenesis.<sup>1,2</sup> Bulky adducts at the C8-site of 2'-deoxyguanosine (C8-dG) are commonly produced by chemical carcinogens.<sup>3</sup> Examples include aromatic amines, such as 2-aminofluorene (AF), N-acetylaminofluorene (AAF) and 4-aminobiphenyl (ABP), which undergo metabolic activation to afford N-linked C8-dG adducts (Figure 5.1A).<sup>4,5</sup> Phenolic toxins, can produce O-linked C8-dG adducts<sup>6-9</sup> or C-linked C8-dG adducts<sup>10-13</sup> upon oxidative metabolism. C-linked varieties are also produced by the phenolic food toxin ochratoxin A (OTA)<sup>11-15</sup>, the polycyclic aromatic hydrocarbon (PAH) benzo[a]pyrene<sup>16</sup> or carcinogenic aryl hydrazines.<sup>17,18</sup>

---

<sup>a</sup>The *Chemical Research and Toxicology* Reference Style is used in this Chapter.

<sup>b</sup>Reprinted with permission from Berger, F. D., Sturla, S. J., Kung, R. W., Montana, T., Wetmore, S. D., and Manderville, R. A. (2018) Conformational preference and fluorescence response of a C-linked C8-biphenyl-guanine lesion in the *NarI* mutational hotspot: evidence for enhanced *syn* adduct formation. *Chem. Res. Toxicol.* 31, 37-47.

<sup>c</sup>R.W.K. performed DFT calculations and MD simulations. R.W.K. and T.M. performed and analysed the NMR work, while the remainder of the experimental work, including synthesis was done by F.D.B., and R.A.M. Writing was done by R.W.K., S.D.W., S.J.S., F.D.B. and R.A.M.



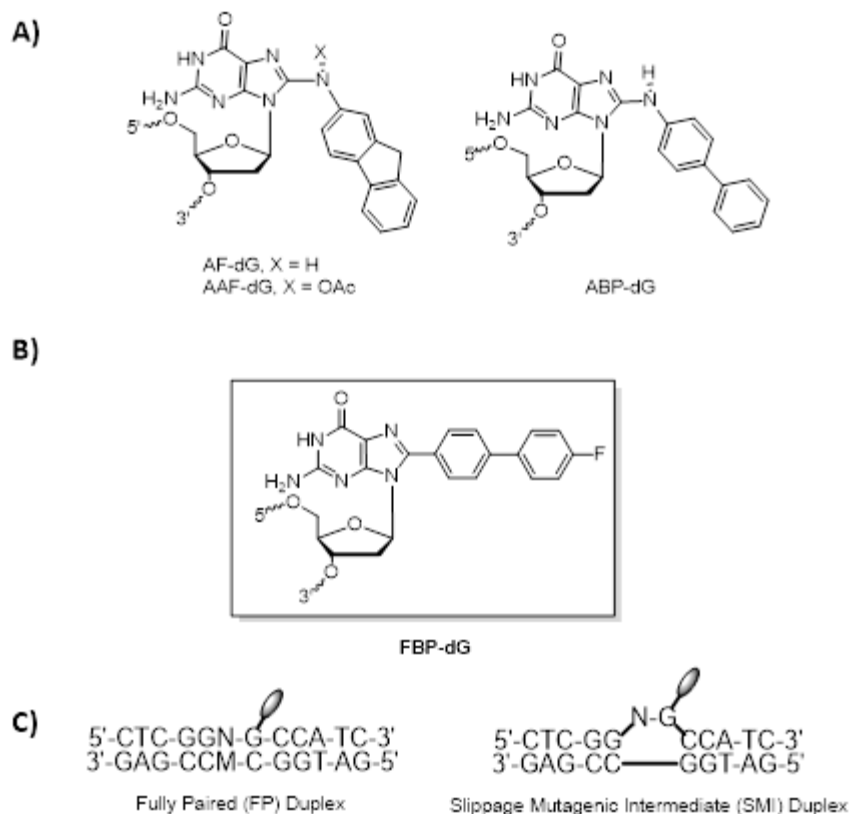


Figure 5.1: A) N-linked AF-dG, AAF-dG and ABP-dG adducts, B) C-linked FBP-dG adduct and C) sequences of the fully paired and slippage mutagenic intermediate (SMI) duplexes, where N:M=C:G, G:C, A:T or T:A in the fully paired duplex, and N=C, G, A or T in the SMI duplex; N:M=C:G represents the fully paired 12-mer *NarI* sequence.

Attachment of the bulky aryl moiety at the C8-site of dG can shift the glycosidic angle from the preferred *anti*-conformation for unmodified dG to the *syn*-conformation for C8-aryl-dG, which disrupts Watson-Crick (W-C) H-bonding with the opposite cytosine within the DNA duplex.<sup>4,5</sup> For duplexes containing a C8-aryl-dG, three unique conformations have been described:<sup>19</sup> the major-groove B-type conformer, in which the modified-dG adopts an *anti* conformation, and the aryl moiety resides in the major groove;<sup>20</sup> the intercalated base-displaced stacked (S) conformer,<sup>21,22</sup> in which the glycosidic linkage of the modified-dG adopts the *syn* orientation and the aryl moiety is intrahelical;

and the minor-groove wedge (W) conformer,<sup>23-25</sup> in which the adduct also adopts a *syn* glycosidic linkage, but the aryl moiety resides in the minor groove. The relative occupancy of each conformer depends on different helical twisting and  $\pi$ -stacking interactions, induced by the different linkage (C vs. N vs. O),<sup>26,27</sup> the size and planarity of the aryl moiety,<sup>28-30</sup> and the flanking sequence.<sup>25,31,32</sup> ABP-dG and AF-dG, examples for N-linked C8-dG adducts, exhibit S/B-heterogeneity, with AF-dG displaying a higher proportion of the population in the S-state due to increased planarity of the aryl moiety compared to the twisted biphenyl of ABP-dG.<sup>33,34</sup> AAF-dG, with an acetyl group attached to the C8-N atom, exhibits an even stronger *syn* preference than AF-dG leading to both S and W conformers.<sup>34</sup> In general, AF-dG and AAF-dG are more mutagenic than ABP-dG due to increased aryl planarity and stronger tendency to adopt the promutagenic *syn* conformation.<sup>19,33-35</sup>

The *Escherichia coli* *NarI* endonuclease sequence (5'-CG<sub>1</sub>G<sub>2</sub>CG<sub>3</sub>C-3') has commonly been used to study the mutagenic potential of N-linked adducts. Adducts and the resulting conformational distortion of the helix influence the mutation frequency and characteristic in the *NarI* sequence. Insertion of AAF-dG at the G<sub>3</sub>-site is well known to induce a higher frequency of GC deletion mutations in bacterial mutagenesis via a two-base slippage mechanism.<sup>36,37</sup> AAF-dG strongly stabilizes the so-called Streisinger slippage mutagenic intermediate (SMI, Figure 5.1C), which provides a rationale for its tendency to induce -2 deletion mutations.<sup>38-40</sup>

C-linked aryl rings lack the flexible tether, which results in different impacts on the DNA duplex conformation in the *NarI* sequence context. Smaller aryl moieties, such as the phenyl-dG (Ph-dG), favor the B-type conformation and lack the ability to stabilize the SMI duplex.<sup>29</sup> However, adducts with linearly-extended bulkier aryl moieties, such as the

pyrene-dG (Py-dG), distort the duplex more and can stabilize the SMI duplex.<sup>30</sup> These studies were assisted by fluorescent properties of C-linked aryl-dG adducts, which permitted insight into the conformational change induced by the adducts. Py-dG in duplex DNA displayed dramatic changes in excitation wavelength compared to the single-strand form, as it was feasible for the pyrene moiety to engage in strong stacking interactions inside the duplex.<sup>30</sup> This suggested the potential design of a new probe to serve as a fluorescent reporter for predicting base flipping and SMI formation on the basis of changes in emission intensity and/or wavelength. To serve this purpose, aryl-dG adducts should be able to stabilize the SMI duplex, and exhibit emission that is sensitive to solvent polarity. Unfortunately, Py-dG is not suitable for this purpose, as the adduct lacks strong fluorescence solvatochromism and failed to exhibit fluorescence that was sensitive to the adduct conformation.<sup>30</sup>

Donor-acceptor biphenyls are aromatic chromophores that exhibit strong fluorescence solvatochromism.<sup>41,42</sup> Attachment of electron donating and electron withdrawing groups at each end of the biphenyl system affords push-pull dyes with high sensitivity toward the polarity of the environment. This knowledge prompted the synthesis of a C-linked dG, containing a 4-fluoro-biphenyl moiety (FBP-dG, Figure 5.1B), in which dG could serve as the donor, while the fluoro-phenyl ring would act as the acceptor. Furthermore, the biphenyl system extends the phenyl ring in a linear fashion for predicted SMI duplex stabilization<sup>30</sup> and the fluorine atom permits the use of <sup>19</sup>F NMR spectroscopy to define adduct conformation within the duplex.<sup>31,34</sup> This probe allows direct comparison with the properties of the corresponding N-linked ABP-dG adduct.<sup>20,34</sup> FBP-dG was incorporated into the *NarI* sequence at the G<sub>3</sub>-site containing various bases (N) 5' to the

lesion site of the adduct (Figure 5.1). Additionally, the fluorophenyl-dG (FPh-dG, Figure 5.1) was synthesized for  $^{19}\text{F}$  NMR studies to complement the data on Ph-dG and permit full comparison to FBP-dG. By a combination of experimental and computational approaches, we determined the conformational preference of FBP-dG within the *NarI* duplex, and furthermore determined the ability of FBP-dG to stabilize the SMI duplex and distinguish the SMI duplex from the fully paired duplex using fluorescence spectroscopy. These data demonstrate the propensity of C-linked C8-F-biphenyl-dG to adopt an intercalated (I) structure and exhibit B/I-heterogeneity within the *NarI* duplex and furthermore suggest FBP-dG as a chemical probe for monitoring SMI formation by fluorescence spectroscopy.

## 5.2 Materials and Methods

### 5.2.1 Materials

Boronic acids (4fluorobiphenylboronic acid and 4-fluorophenyl boronic acid),  $\text{Pd}(\text{OAc})_2$ , 3,3',3''-phosphinidynetris-(benzenesulfonic acid) trisodium salt (TPPTS), *N,N*-dimethylformamide diethyl acetal, 4,4'-dimethoxytrityl chloride, 2-cyanoethyl *N,N*-diisopropyl-chlorophosphoramidite, and other commercial products used for the synthesis of FBP-dG and FPh-dG phosphoramidites were used as received. All unmodified phosphoramidites (bz-dA-CE, ac-dC-CE, dmf-dG-CE and dT-CE), activator (0.25 M 5-(ethylthio)-1H-tetrazole in  $\text{CH}_3\text{CN}$ ), oxidizing agent (0.02 M  $\text{I}_2$  in THF/pyridine/ $\text{H}_2\text{O}$ , 70/20/10, v/v/v), deblock (3% dichloroacetic acid in dichloromethane), cap A (THF/2,6-lutidine/acetic anhydride), cap B (methylimidazole in THF), and 1000 Å controlled pore glass (CPG) solid supports were purchased from Glen Research (Sterling, VA). Unmodified oligonucleotides were purchased from Sigma Genosys (Oakville, ON).

### 5.2.2 Methods

Suzuki cross-coupling reactions of boronic acids with 8-Br-dG to afford FBP-dG and FPh-dG were performed as described previously for other C-linked C8-dG adducts.<sup>29,30</sup> NMR spectra were recorded in CDCl<sub>3</sub> or DMSO-d<sub>6</sub> on 300 and 600 MHz Bruker spectrometers, chemical shift were referenced to TMS (0 ppm) or the respective residual solvent peak. Full synthetic details and NMR Spectra for modified phosphoramidites, NMR spectra and ESI-MS analysis of modified *NarI* oligonucleotides are available in Appendix D. All adducted *NarI* oligonucleotide substrates were prepared on a 1 μmol scale using a BioAutomation MerMade 12 automatic DNA synthesizer using standard or modified β-cyanoethylphosphoramidite chemistry.

The DNA concentration was determined by UV absorption, the extinction coefficient was calculated using the IDT calculator at <http://eu.idtdna.com/analyzer/applications/oligoanalyzer> assuming the same absorbance properties for dG and FBP-dG and FPh-dG ( $\epsilon(5'\text{-CTCGGCGCCATC})=102100$  L/(mole·cm)). Thermal stability of *NarI* oligonucleotides were measured by variable temperature UV analysis using a Cary 300-Bio UV-Vis spectrophotometer equipped with a 6x6 multicell Peltier block-heating unit and Hellma 114-QS 10 mm light path cells. *T<sub>m</sub>* values were determined as previously outlined.<sup>29,30</sup> Oligonucleotide samples were prepared in 50 mM phosphate buffer, pH 7, with 100 mM NaCl, using equivalent amounts (6.0 μM) of the unmodified or FPh-dG and FBP-dG modified *NarI* oligonucleotide and its complementary strand. Circular dichroism (CD) spectra were recorded on a Jasco J-815 CD spectropolarimeter equipped with a 1x6 Multicell block thermal controller and a water

circulator unit. Spectra were collected at 10 °C between 200 and 400 nm, with a bandwidth of 1 nm and scanning speed of 100 nm/min, as previously described.<sup>29,30</sup> All fluorescence spectra were recorded on a Cary Eclipse Fluorescence spectrophotometer equipped with a 1x4 Multicell block Peltier stirrer and temperature controller. Both excitation and emission spectra were recorded for the FPh-dG and FBP-dG modified *NarI* oligonucleotide hybridized to its complementary strand at 10 °C.

For dynamic <sup>19</sup>F NMR experiments, the single-stranded DNA (ssDNA) samples containing FPh-dG and FBP-dG were dissolved in a 100% D<sub>2</sub>O solution, and filtered. A stoichiometric amount of the complementary strand was added to form double-stranded DNA (dsDNA). <sup>19</sup>F NMR experiments were performed for both modified duplexes at temperatures ranging from 10-70 °C on a 16.4 T Bruker Avance III HD NMR spectrometer. Each experiment was carried out using proton decoupling, with 10k scans and a sweep width of 250 ppm. The spectra were analyzed and figures prepared using the MestReNova software.

Molecular dynamics (MD) simulations were conducted on the fully paired and SMI *NarI* duplexes (N=C) bearing FPh-dG and FBP-dG at the G<sub>3</sub> position. The modified fully paired duplexes were examined in major-groove, wedge, intercalated and base-displaced conformations, while FPh-dG and FBP-dG were considered in the *anti* and *syn* orientations in the SMI. Each system was placed in an octahedral TIP3P water box with charge equivalencing sodium ions. A total of 100 ns of production MD simulations were performed for each system using the PMEMD module of AMBER14.<sup>43</sup> The amber14SB force field was used,<sup>43</sup> which was supplemented by GAFF parameters<sup>44</sup> as required for the

adducts. Key dihedral angles at the lesion sight, backbone RMSD, and representative structures were obtained using the cptraj module of AMBERTOOLS.<sup>45</sup> Free energies were calculated using the MMPBSA method.<sup>46</sup> Full details of the simulations are provided in the Appendix D.

## 5.3 Results

### 5.3.1 Optical Properties of FBP-dG in *NarI*

FBP-dG and FPh-dG containing *NarI* 12-mer were synthesized using standard solid phase synthesis (Scheme D.1, Appendix D). Solvatochromic properties of the FBP-dG nucleoside were determined, as well as the conformational influence of FBP-dG in the 12-mer *NarI* sequence, using UV-VIS spectroscopy, fluorescence and CD. *NarI* bearing FPh-dG was used for <sup>19</sup>F NMR studies.

On the basis of UV derived thermal melting parameters of the various 12-mer *NarI* duplexes, FBP-dG appeared to strongly destabilized all four fully paired duplexes compared to their unmodified duplex, whereas FBP-dG stabilized the SMI duplex (Table 5.1). When the 5'-flanking base was a G or C, the degree of the destabilization was slightly less ( $\Delta T_m = -10.0$  °C for G,  $-12.5$  °C for C) than observed with an A or T 5'-flanking base ( $\Delta T_m = -14.5$  °C for A,  $-14.0$  °C for T, Table 5.1). In the 12-mer *NarI* duplex (N:M=C:G), FBP-dG was less destabilizing than Ph-dG ( $\Delta T_m = -15.7$  °C) in the same sequence context.<sup>29</sup> In the SMI duplexes containing FBP-dG, the modified duplexes were more stable when the 5'-flanking base was a pyrimidine ( $\Delta T_m = +7.0$  °C for T,  $+6.0$  °C for C) versus a purine ( $\Delta T_m = +4.5$  °C for A,  $1.5$  °C for G).

Table 5.1: Thermal Melting Parameters of FBP-dG modified *NarI*.

5'-CTC-GGN-X-CCA-TC-3'

5'-CTC-GG-NX-CCA-TC-3'

3'-GAG-CCM-C-GGT-AG-5'

3'-GAG-CC----GGT-AG-5'

| Fully Paired Duplex | $T_m$ (°C) <sup>a</sup> | $\Delta T_m$ (°C) <sup>b</sup> | SMI Duplex    | $T_m$ (°C) | $\Delta T_m$ (°C) |
|---------------------|-------------------------|--------------------------------|---------------|------------|-------------------|
| N:M=C:G, X=dG       | 63.5                    | /                              | N=C, X=dG     | 39.5       | /                 |
| N:M=C:G, X=FBP-dG   | 51.0                    | -12.5                          | N=C, X=FBP-dG | 45.5       | + 6.0             |
| N:M=G:C, X=dG       | 62.0                    | /                              | N=G, X=dG     | 41.0       | /                 |
| N:M=G:C, X=FBP-dG   | 52.0                    | -10.0                          | N=G, X=FBP-dG | 42.5       | +1.5              |
| N:M=A:T, X=dG       | 61.0                    | /                              | N=A, X=dG     | 38.0       | /                 |
| N:M=A:T, X=FBP-dG   | 46.5                    | -14.5                          | N=A, X=FBP-dG | 42.5       | +4.5              |
| N:M=T:A, X=dG       | 61.0                    | /                              | N=T, X=dG     | 38.0       | /                 |
| N:M=T:A, X=FBP-dG   | 47.0                    | -14.0                          | N=T, X=FBP-dG | 45.0       | +7.0              |

<sup>a</sup> $T_m$  values of duplexes (6  $\mu$ M) measured in 50 mM sodium phosphate buffer, pH 7, with 0.1 M NaCl, a heating rate of 1 °C/min and are reproducible within 3%. <sup>b</sup> $\Delta T_m = T_m$  (modified duplex) –  $T_m$  (unmodified duplex).

The quantum yield and fluorescence properties of the FBP-dG nucleoside were initially determined in H<sub>2</sub>O, CH<sub>3</sub>CN and CHCl<sub>3</sub> for comparison to the solvatochromic properties of other C-linked nucleosides (Table D.1, Appendix D).<sup>29</sup> FBP-dG in the free nucleoside form is strongly emissive in H<sub>2</sub>O ( $\lambda_{\text{max}}$ =286 nm,  $\lambda_{\text{em}}$ =424 nm,  $\Phi_f$ =0.52) and is weakly sensitive to solvent polarity. These characteristics were also noted for the simple Ph-dG nucleoside ( $\lambda_{\text{max}}$ =277 nm,  $\lambda_{\text{em}}$ =395 nm,  $\Phi_f$ =0.44) that lacks donor-acceptor (D-A) character.<sup>29</sup> Biphenyl systems that exhibit D-A character typically display quenched emission in water, and light-up in non-polar solvents. A red-shift in emission wavelength ( $\lambda_{\text{em}}$ ), as observed for FBP-dG is a typical behavior for biphenyl systems (Table D.1, Appendix D).<sup>41</sup>



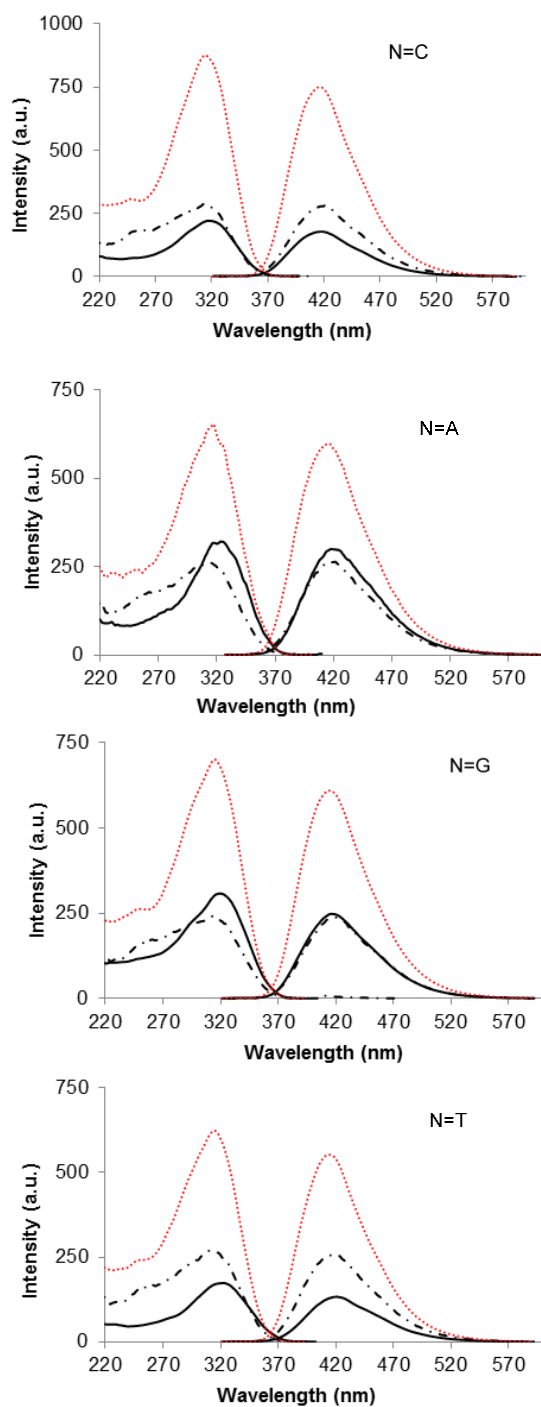


Figure 5.2: Excitation and emission spectra of FBP-dG-modified 12-mer *NarI* oligonucleotides in the single-strand (black dashed-dotted traces), fully paired duplex (solid black traced) and SMI duplex (dotted red traces). All spectra of oligonucleotides (6 mM) were recorded in 50 mM sodium phosphate buffer, pH 7, with 0.1 M NaCl at 10 °C.

Inserting FBP-dG into the 12-mer *NarI* sequence caused a 29 nm red-shift in excitation wavelength and little change in emission wavelength (Figure 5.2, Table D.2, Appendix D), suggesting formation of a more planar biphenyl adduct structure. Emission from FBP-dG hybridized to the fully paired complementary strand was slightly quenched, when the 5'-flanking base was a pyrimidine (N=C or T), whereas virtually no change in emission intensity was observed when the 5'-flanking base was a purine (N=G or A) (Figure 5.2). The  $\lambda_{\text{ex}}$  in the fully paired duplexes a red-shifted 4-12 nm relative to the ssDNA was observed, suggesting increased  $\pi$ -stacking interactions (Table D.2, Appendix D). While  $\lambda_{\text{em}}$  was blue-shifted, up to -4 nm, suggesting placement of FBP-dG in a less polar environment (Table D.2, Appendix D). Comparing the emission intensity of the fully paired duplex to the SMI duplex, a 4.2-fold increase for the sequences containing a 5'-flanking pyrimidine, and a 2.0-fold increase containing 5'-flanking purine was observed (Figure 5.2). This fluorescence response suggested a decrease in  $\pi$ -stacking interactions and less exposure of the biphenyl to the polar aqueous solvent.

Table 5.2: MM-PBSA relative energies (kJ/mol) for different conformations of FPh-dG- and FBP-dG-adducted fully paired *NarI* duplex.

| Aryl group | Major-groove | Wedge | Stacked | Intercalated |
|------------|--------------|-------|---------|--------------|
| FPh        | 0.0          | 24.8  | 65.0    | 60.8         |
| FBP        | 0.0          | 43.4  | 36.4    | 2.3          |

### 5.3.2 Conformational Analysis of FBP-dG in *NarI*

In order to determine the conformational preference of FBP-dG within the fully paired and SMI *NarI* duplex (N=C), MD simulations were carried out on both duplex

structures and comparison was drawn to the corresponding structures determined for the smaller FPh-dG. Previous calculations indicate that the unsubstituted Ph-dG strongly favors the B-type conformation in the fully paired *NarI* duplex structure.<sup>29</sup> In the SMI duplex, Ph-dG adopts the *anti* conformation (favored by 15.1 kJ/mol relative to the *syn* conformation). The fully paired *NarI* duplexes containing FBP-dG or FPh-dG were also analyzed using variable temperature <sup>19</sup>F NMR spectroscopy, to highlight the influence of the biphenyl on the duplex conformation.

MD simulations isolated four unique stable conformations for duplexes containing FPh-dG or FBP-dG (Figures 5.3, 5.4, Table 5.2 and Table D.4, Appendix D). In all

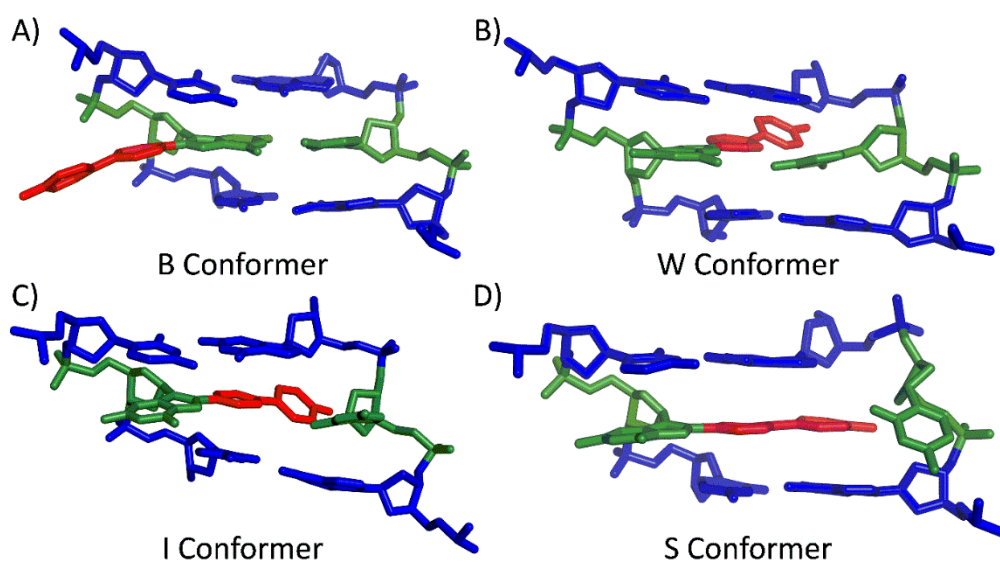


Figure 5.3: FBP-dG adducted *NarI* (N=C) in the A) major-groove conformation in which the biphenyl moiety is in the major groove, while the damaged G is in the *anti* conformation and maintains Watson-Crick hydrogen bonding with the opposing C, B) wedge conformation in which the bulky moiety is in the minor groove, while the adducted G is *syn* and forms two Hoogsteen interactions with the pairing C, C) intercalated conformation in which the damaged G is *syn*, and the biphenyl moiety and opposing C are intercalated within the helix, and D) base-displaced intercalated conformation in which the adducted G is *syn* and the bulky moiety displaces the opposing C into the extrahelical major groove location. Only the three base pairs about the lesion are shown for clarity, with the damaged G and opposing C in green, the bulky moiety in red, and flanking base pairs in blue.

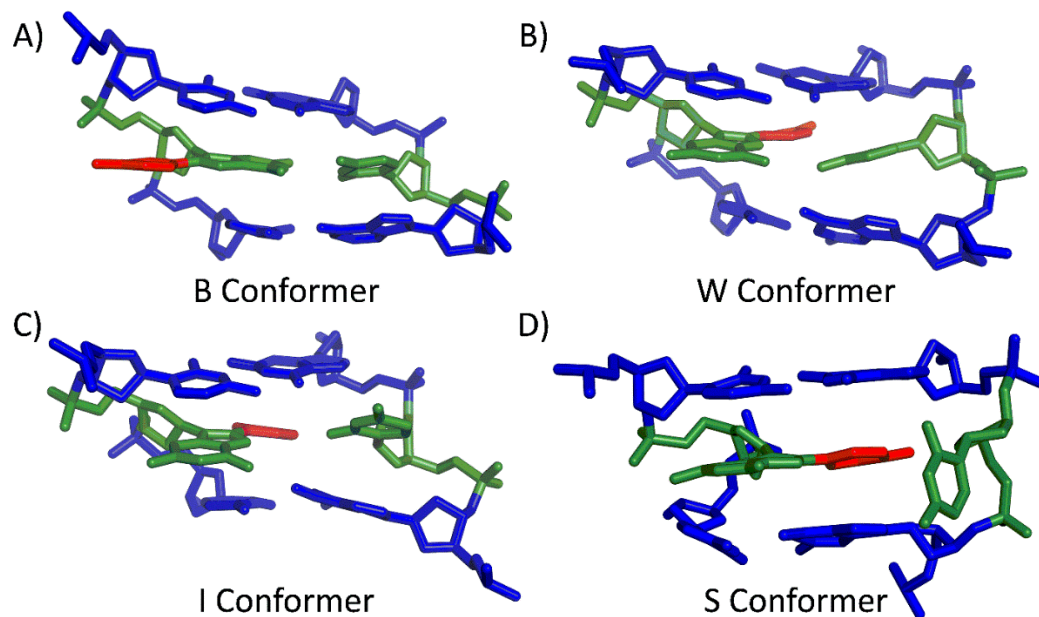


Figure 5.4: FPh-dG adducted *NarI* (N=C) in the A) major-groove conformation in which the phenyl moiety is in the major groove, while the damaged G is in the *anti* conformation and maintains Watson-Crick hydrogen bonding with the opposing C, B) wedge conformation in which the bulky moiety is in the minor groove, while the adducted G is *syn* and forms two Hoogsteen interactions with the pairing C, C) intercalated conformation in which the damaged G is *syn*, and the phenyl moiety and opposing C are intercalated within the helix, and D) base-displaced intercalated conformation in which the adducted G is *syn* and the bulky moiety displaces the opposing C into the extrahelical major groove location. Only the three base pairs about the lesion are shown for clarity, with the damaged G and opposing C in green, the bulky moiety in red, and flanking base pairs in blue.

conformations, the Watson-Crick hydrogen bonding in the base pairs flanking the adduct remains intact, indicating that any structural deviations occur solely at the modified dG. In the major-groove or B-type conformation, FBP-dG adopts the *anti* glycosidic orientation. The bulky moiety is located in the major groove and adopts a planar arrangement about the guanine–aryl bond ( $\theta$ , Figures 5.3A and 5.4A, Table D.4, Appendix D), which minimizes helical distortion and maintains three stable Watson-Crick hydrogen bonds at the lesion site

(> 95% occupancy). The fluorine atom exclusively interacts with solvent. In all other adducted DNA conformations, the lesion adopts the *syn* conformation.

The wedge conformer places the biphenyl in the minor groove (Figure 5.3B), and the dG forms two hydrogen bonds with the opposing dC ( $\approx 35$  and 75% occupancy). Similar to the major-groove conformation, the fluorine atom is exclusively exposed to solvent in this conformer. Since the biphenyl is positioned outside the helix in both the major-groove and wedge conformations, free rotation occurs about the  $\xi$  inter-ring dihedral angle within the bulky moiety, resulting in various relative orientations of the two rings throughout the MD simulations (i.e., standard deviation > 100°; Table D.4, Appendix D).

In the base-displaced stacked conformer, the biphenyl is located within the helix, and the opposing base is fully displaced into the major groove. Unlike FPh-dG-containing DNA, the fluorine atom has no strong interactions (e.g., hydrogen bonds or stacking contacts) with other parts of the DNA helix. MD simulations predict another closely related conformation (denoted the intercalated (I) conformer) in which the biphenyl stacks with the flanking bases, and partially displaces the opposing dC such that the interactions with FBP-dG are disrupted (Figures 5.3C and 5.4C). Nevertheless, the opposing dC remains within the helix and stacks with the 5'-interstrand base. Another notable difference is the local environment of the fluorine atom. Specifically, since the aryl moiety intercalates within the helix, the fluorine atom of FPh-dG partially stacks with the 3'-dC, and is more shielded from solvent than the fluorine atom in the major-groove and wedge conformations. In contrast, due to the larger bulky moiety, the fluorine atom of FBP-dG-containing DNA has no significant interactions with the DNA helix. The adducts in both intercalated conformations have reduced conformational flexibility within the aryl moiety with respect

to the  $\xi$  dihedral angle compared to the B and W counterparts (i.e., standard deviation  $< 30^\circ$  for S and I, but  $> 100^\circ$  for B and W; Table D.4, and Figure 5.5, Appendix D). Furthermore, the aryl moiety is on average less twisted in the stacked and intercalated conformations (average twist of  $19^\circ$  and  $26^\circ$ , respectively) compared to in the B and W conformers ( $33^\circ$  and  $31^\circ$ , respectively).

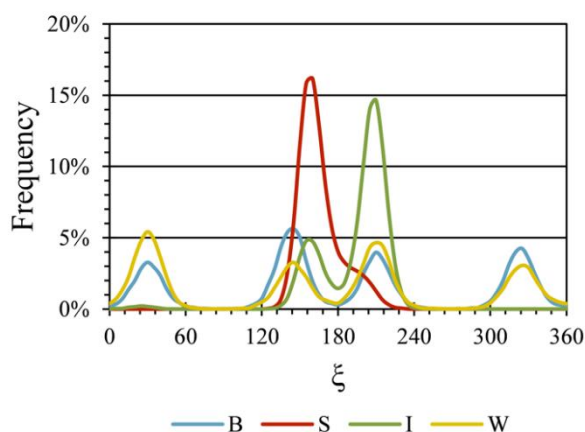


Figure 5.5: Frequency of the  $\xi$  dihedral angle (Figure D.2, Appendix D) for FBP-dG-adducted fully paired *NarI* duplex in the major-groove (B), stacked (S), intercalated (I), and wedge (W) conformations.

The major-groove orientation is the most stable conformation for both adducts (Table 5.2). Nevertheless, despite the structural similarities of the FBP- and FPh-containing duplexes, the relative energies for the next stable conformer vary significantly. Specifically, for the smaller FPh-dG lesion, the next most stable DNA conformation is the wedge conformer, which is 24.8 kJ/mol higher in energy. Furthermore, the small size of the phenyl moiety leads to energetically inaccessible intercalated conformations ( $> 60$  kJ/mol relative to the major-groove conformer). In contrast, the second most stable DNA conformation associated with FBP-dG is the intercalated conformer, which is only 2.3 kJ/mol higher in energy. Similarly, the base-displaced stacked conformation is slightly more stable than the

wedge orientation, although both are  $> 35$  kJ/mol above the most stable major-groove orientation.

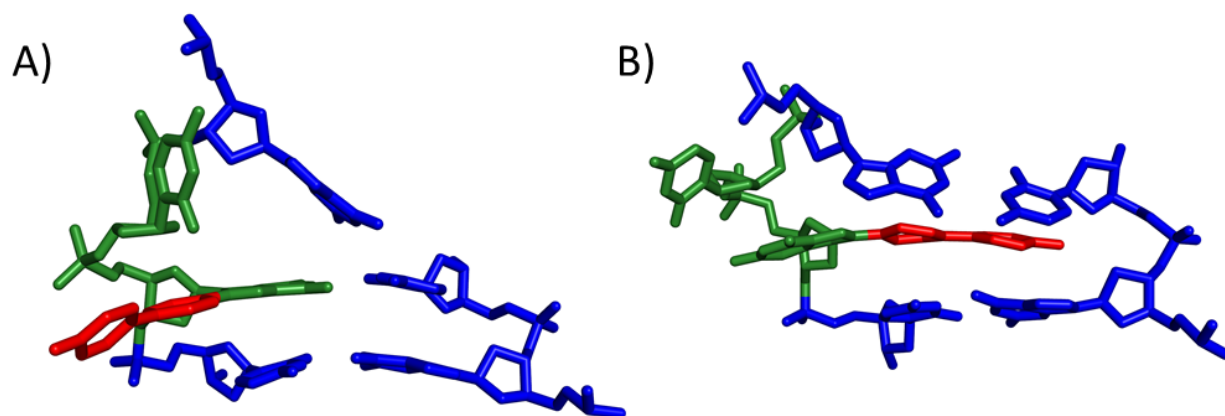


Figure 5.6: FBP-dG-adducted *NarI* paired against a  $-2$  deletion strand with the adduct in the A) *anti* or B) *syn* orientation. Only the lesion site including the 2 base pairs flanking the bulge are shown for clarity, with the damaged G and unpaired C in green, the bulky moiety in red, and flanking base pairs in blue.

For the SMI duplexes, two stable conformations were isolated corresponding to the two glycosylic bond orientations of FBP-dG. The *anti* orientation about the glycosidic bond places the aryl moiety within the major groove of the helix (Figure 5.6A) and is stabilized by Watson-Crick hydrogen bonding between the FBP-dG and the opposing 5'-dC. Due to the bulge formation, the 5' base pair is weakened ( $< 50\%$  occupancy of Watson-Crick bonds) and the unpaired 5'-dC falls into the major groove. The position of the unpaired dC is highly flexible, alternating between a stacked orientation with the aryl moiety ( $\approx 40\%$  of simulation time) and intrastrand hydrogen bonding via N4-H to a phosphate oxygen of the 5' nucleotide ( $\approx 60\%$  of simulation time). This predicted structure for FBP-dG is considerably similar to that previously reported for the *anti* Ph-dG in the same sequence context (Figure 5.7A). When FBP-dG adopts the *syn* glycosidic orientation, dG resides in the major groove and the biphenyl is intercalated such that it stacks with the base pairs

flanking the bulge region (Figure 5.6B). As discussed for the fully paired structures, the biphenyl  $\pi$ - $\pi$  interactions in the *syn* orientation limit the rotational freedom about the  $\xi$  dihedral angle, resulting in a more planar biphenyl system in the *syn* orientation (average twist of 25°) compared to the *anti* orientation (average twist of 31°, Figure 5.8). Although the *syn* orientation of FBP-dG is similar to that previously reported for Ph-dG<sup>29</sup> (RMSD = 1.25 Å), the biphenyl is in a position that maximizes stacking interactions with all four nucleobases in the flanking base pairs (Figure 5.7B). As a result, the energy difference between the *anti* and *syn* conformations are small, with *anti* being only 2.8 kJ/mol more stable. This structural and energetic data correlates with the observed increased stability of the SMI duplexes containing FBP-dG compared to Ph-dG. Specifically, although both *anti* structures similarly lack helical interactions for the aryl group within the SMI duplex, FBP-dG *syn* structure exhibits increased stacking interactions between the aryl moiety and the flanking base pairs compared to Ph-dG, which increases the accessibility of this conformation and overall helical stability.

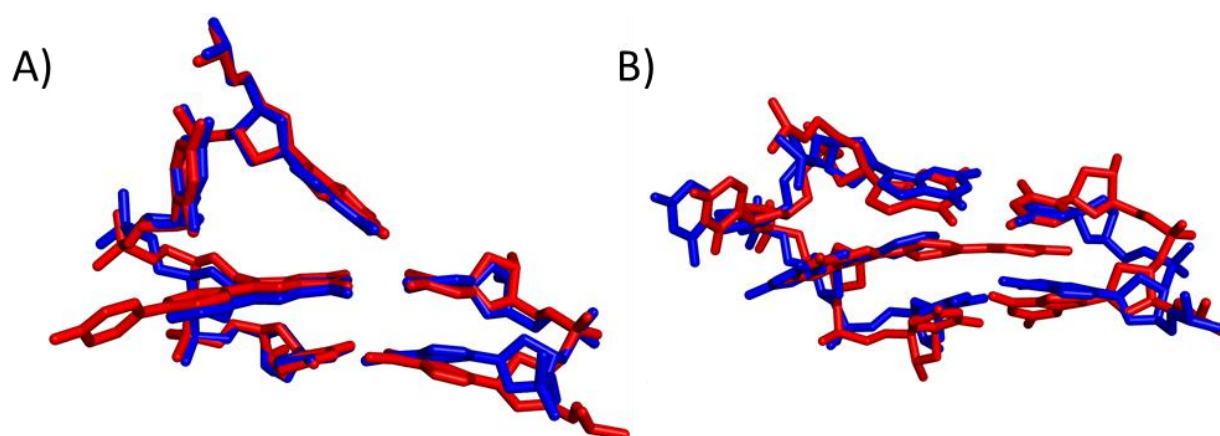


Figure 5.7: Overlay of the -2 deletion site for FPh-dG (blue) and FBP-dG (red) adducted helices with the adduct in A) *anti* or B) *syn* orientation. Only the lesion site including the 2 base pairs flanking the bulge are shown for clarity.



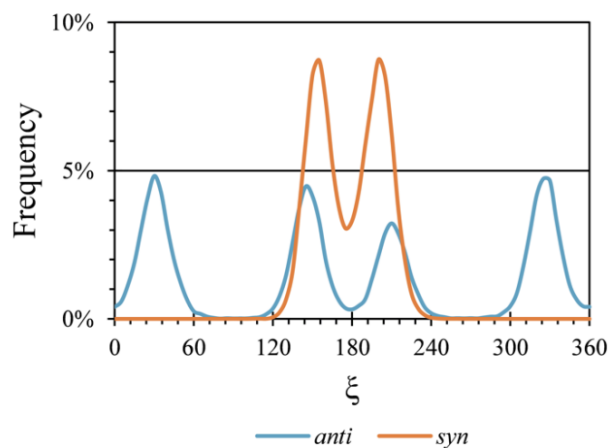


Figure 5.8: Frequency of the  $\xi$  dihedral angle (Figure D.2, Appendix D) for FBP-dG-adducted SMI DNA in the *anti* and *syn* conformations.

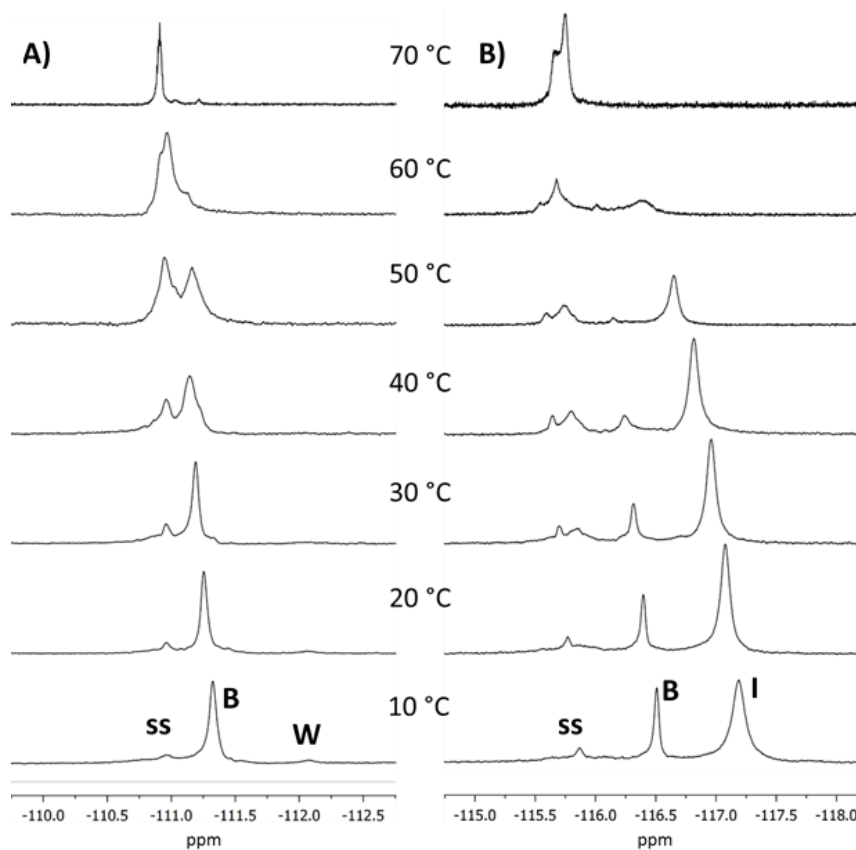


Figure 5.9: Variable temperature proton decoupled  $^{19}\text{F}$  NMR of A) FPh-dG- and B) FBP-dG-adducted fully paired *NarI* duplex; single-strand (ss), major-groove (B), wedge (W), and intercalated (I).

$^{19}\text{F}$  NMR data was collected at variable temperatures for FPh-dG and FBP-dG-containing DNA duplexes (Figure 5.9). At 10 °C, the  $^{19}\text{F}$  NMR spectrum of FPh-dG adducted DNA (Figure 5.9A) contains one major peak at -111.26 ppm (85% of the total adduct signal) and two small peaks at -110.96 ppm (10%) and -112.07 ppm (5%). The most downfield peak at -110.96 ppm corresponds to ssDNA since the variable temperature data reveals that this peak gains intensity with increased temperature, while the other peaks decrease in intensity, which correlates with an increased concentration of ssDNA upon denaturing of dsDNA (Figure 5.9). Since the MD simulations suggest that the major-groove conformation is highly favored for the FPh-dG, the major NMR signal at -111.26 ppm (85%) was ascribed to the most stable B-form. The minor upfield peak at -112.07 ppm (5%) was attributed to the formation of the W-form. These assignments were further supported by the persistence of the peak at -111.26 ppm at higher temperatures, correlating with the greater stability of the B-form over the W-form dsDNA conformation (Table 5.2). These assignments suggest that the B-form makes up 94% of the total duplex for the FPh-dG adducted *NarI*, with the W-form making up the remaining 6%, which is consistent with the relative energies of these duplexes determined by the MD simulations (Table 5.2).

The  $^{19}\text{F}$  NMR spectra of FBP-dG-containing DNA at 10 °C contains three peaks corresponding to ssDNA and dsDNA (Figure 5.9B). The small downfield peak at -115.77 ppm (4%) corresponds to ssDNA, while the two major peaks at -116.40 ppm (19%) and -117.08 ppm (77%) correspond to dsDNA. The single stranded peak was assigned based on the variable temperature data, which indicates that the peak at -115.7 ppm increases with sample heating and corresponds to the formation of ssDNA due to annealing. Based on MD simulations, FBP-dG-containing DNA was predicted to be stable in two

conformations, the major-groove and intercalated conformers (Table 5.2). The calculated energy difference between the major-groove and intercalated conformer of FBP-dG-containing DNA (2.3 kJ/mol) is considerably smaller than the difference between the two most stable FPh-dG adducted DNA conformations (24.8 kJ/mol), making unambiguous assignment of the peaks at -116.40 and -117.08 ppm difficult. For several N-linked C8-dG adducts, previously characterized by  $^{19}\text{F}$  NMR, the fluorine atom in the S-type intercalated conformer resonates up-field relative to the B-type conformer.<sup>31,34</sup> Furthermore, at 50 °C the peak at -116.40 ppm disappears, while the peak at -117.08 ppm maintains half of the original area. Previous work on N-linked C8-dG adducts has shown, that S-type intercalated structures with the aromatic moiety in the duplex are known to be more stable.<sup>31,34</sup> On the basis of these features, the downfield peak at -116.40 ppm (19%), that is closer to the ssDNA peak was ascribed to the B-type conformer, while the up-field peak at -117.08 ppm (77%) was assigned to the intercalated conformer. In this conformer, the fluorine atom is more shielded because the biphenyl is stacked within the DNA duplex. These assignments suggest that the intercalated conformer represents 80% of the FBP-dG adducted *NarI* duplex, while the B-type conformer represents 20%.

To further clarify which peaks represent the intercalated (S and I) and groove (B and W) conformations, 1D  $^{19}\text{F}$  NMR spectra were obtained for the adducted helices in 90%  $\text{H}_2\text{O}$  and 10%  $\text{D}_2\text{O}$  at 20 °C. By comparing the H–D isotopic effect, the degree of fluorine solvation can be determined. Specifically, the peaks with the greatest change in chemical shift will correspond to the conformation in which the fluorine label is most solvated. Previous work on fluorine-labeled adducts has demonstrated that a downfield chemical shift of  $\approx 0.2$  ppm when the deuterium content is decreased from 100% to 10% is indicative

of a solvent exposed conformation.<sup>49</sup> Therefore, a chemical shift difference of  $\approx 0.08$  ppm corresponds to a conformation with the fluorine atom shielded from solvent. For the FBP-dG adducted DNA, the chemical shifts for the major (-117.08 ppm) and minor (-116.40 ppm) peaks move 0.04 ppm and 0.17 ppm downfield to -116.23 ppm and -117.04 ppm, respectively (Figure 5.10). This confirms that the major peak is solvent shielded and likely corresponds to the I conformation, while the major peak is solvent exposed and corresponds to the B conformation. For FPh-dG-adducted DNA, the major peak shifts by 0.14 ppm (from -111.27 ppm to -111.13 ppm, Figure 5.11), while the minor peak shifts by 0.10 ppm (from -112.07 to -111.97 ppm). Although these changes are smaller than previously reported for other solvent exposed conformations, the bulky group of FPh-dG is the smallest studied using this technique to date and does not extend the fluorine atom into the solvent sphere to the same extent as larger lesions. Therefore, the observed H-D isotopic effect supports our previous assignment of both peaks to DNA structures that are solvent exposed.

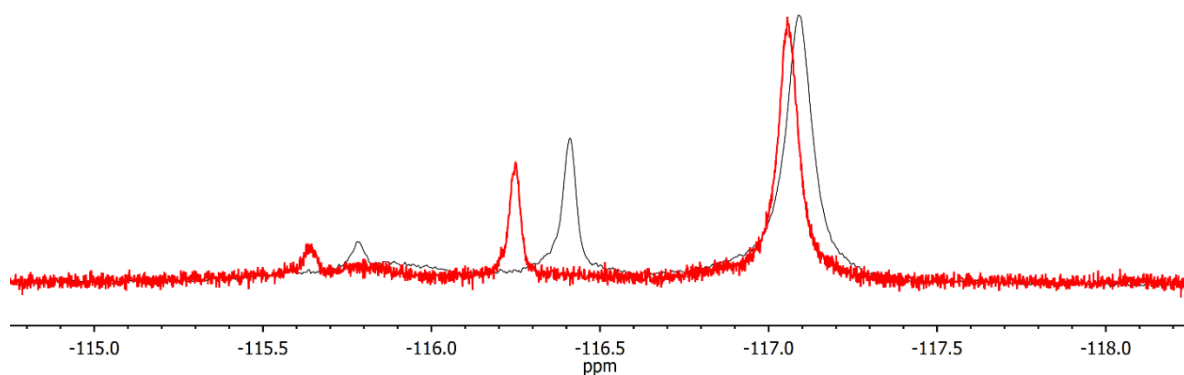


Figure 5.10: The isotopic effect for the proton decoupled  $^{19}\text{F}$  NMR of the FBP-dG-adducted fully paired *NarI* duplex in 100%  $\text{D}_2\text{O}$  (black) and 90%  $\text{H}_2\text{O}$  10%  $\text{D}_2\text{O}$  (red).

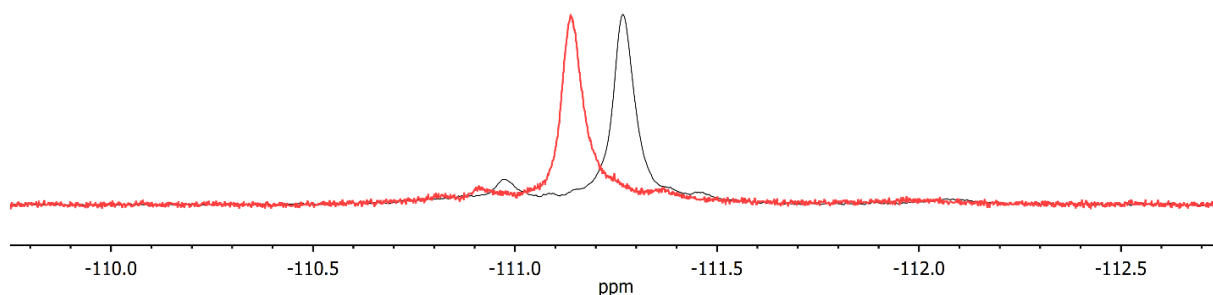


Figure 5.11: The isotopic effect for the proton decoupled  $^{19}\text{F}$  NMR of the FPh-dG-adducted fully paired *NarI* duplex in 100%  $\text{D}_2\text{O}$  (black) and 90%  $\text{H}_2\text{O}$  10%  $\text{D}_2\text{O}$  (red).

## 5.4 Discussion

### 5.4.1 Conformation of FBP-dG in *NarI*: Comparison to other C-linked and N-linked C8-dG Adducts

On the basis of MD simulations, the *NarI* 12-mer DNA duplex with FBP-dG at the  $\text{G}_3$  position was predicted to display conformational heterogeneity between anti B-type and syn intercalated (I) structures. This model was consistent with  $^{19}\text{F}$  NMR data, which indicated a 20:80 B:I ratio (Figure 5.9B). For the smaller FPh-dG,  $^{19}\text{F}$  NMR analysis suggested a 94:6 B:W ratio (Figure 5.9C), consistent with the MD prediction of a strong anti-preference, as well as previous MD predictions for Ph-dG adducted DNA.<sup>29</sup>

The conformational preferences of FBP-dG are comparable to other N-linked adducts.<sup>19</sup> On the basis of  $^{19}\text{F}$  NMR data FABP-dG and FAF-dG in a  $\text{TG}^*\text{A}$  sequence context ( $\text{G}^*$  represents the N-linked C8-dG adduct) exhibited B:S ratios of 100:0 for FABP-dG and 55:45 for FAF-dG.<sup>20,49</sup> In the 12-mer *NarI* sequence context (i.e.,  $\text{CG}^*\text{CC}$ ), FAF-dG exhibits a 25:75 B:S ratio.<sup>31</sup> There is no  $^{19}\text{F}$  NMR study of the N-linked FABP-dG in the *NarI* sequence; however, FAF-dG and FABP-dG have been studied in  $\text{CG}^*\text{CA}$  and

CG\*CT sequence contexts, and FABP-dG was predicted to exhibit 95:5 and 67:33 B:S ratios, while the B:S ratios for FAF-dG were 36:64 and 20:80.<sup>34</sup> The B:S ratio for FAF-dG in the CG\*CT sequence (20:80)<sup>34</sup> is almost identical to the corresponding ratio in the authentic *NarI* sequence (25:75),<sup>31</sup> suggesting that the B:S ratio of 67:33 for FABP-dG in the CG\*CT sequence<sup>34</sup> is representative of its expected ratio in the fully paired *NarI* duplex. The B:I ratio of 20:80 of FBP-dG suggests a stronger syn preference, despite the calculated equilibrium between the *syn* and *anti* conformations. In fact, the B:I ratio determined for FBP-dG is much more similar to the B:S ratios determined for AF-dG. The conformational characteristics suggests greater mutagenicity for the C-linked FBP-dG compared to the N-linked ABP-dG due to its greater propensity to generate the promutagenic syn conformation.<sup>19,33-35</sup>

FBP-dG appears to stabilize the SMI duplex, especially containing a pyrimidine 5' to the adduct (Table 5.1). The degree of SMI duplex stabilization by FBP-dG was essentially identical to that for the C-linked Py-dG (Table D.3, Appendix D),<sup>30</sup> which was predicted to favor a *syn* conformation by 12.6 kJ/mol (Table D.3, Appendix D). In the favored *syn* conformation, the pyrene can stack with its flanking bases.<sup>30</sup> The MD calculations suggested a conformational heterogeneity for FBP-dG between *anti* and *syn* in the SMI duplex. The predicted structure of the *anti* conformation is similar to that previously reported for the *anti* Ph-dG (Figure 5.7A),<sup>29</sup> which destabilizes the SMI duplex ( $\Delta T_m$  vs unmodified= -2.9 °C, Table D.3, Appendix D). On the other hand, in the *syn* conformation, increased stacking interactions between the biphenyl and its flanking bases are predicted compared to the smaller phenyl (Figure 5.7B). Thus, adoption of the *syn* conformation provides a rationale for the stability of the SMI duplex containing FBP-dG.

The conformational stability of the SMI is thought to be a critical determinant for the -2 frameshift mutation frequency in NarI sequences and has been shown to correlate with the ability of N-linked C8-dG adducts to form stable S-conformers in the fully paired duplex.<sup>31</sup> Like FBP-dG, N-linked C8-dG adducts stabilize the NarI SMI duplex. For example AAF-dG is strongly stabilizing compared to the unmodified oligonucleotide ( $\Delta T_m = +15\text{ }^\circ\text{C}$ ),<sup>38,39</sup> whereas AF-dG is only mildly stabilizing ( $\Delta T_m$  values of  $+3.3\text{ }^\circ\text{C}$ <sup>31</sup> and  $+6.0\text{ }^\circ\text{C}$ <sup>24</sup>). AF-dG produces 10-fold fewer -2 deletion mutations compared to AAF-dG,<sup>50</sup> but AF-dG is expected to induce a higher frequency of deletion mutations than ABP-dG due to its greater tendency to adopt the S-type duplex structure.<sup>34</sup> Based on these comparisons, the deletion mutation frequency anticipated to arise from the C-linked FBP-dG adduct may be very similar to that from AF-dG, since the adducts exhibit similar anti/syn ratios in the fully paired duplex, and have almost identical capacities to stabilize the SMI duplex, as evidenced by UV thermal melting experiments.

#### 5.4.2 Fluorescence of FBP-dG in Fully Paired versus SMI NarI Duplex

Based on fluorescence spectroscopy, the biphenyl nucleoside lacks D-A character.<sup>29</sup> However, when placed within the NarI oligonucleotide, the  $\lambda_{ex}$  shifts to the red by 29 nm in the single-strand and up to 37 nm in the fully paired duplex. These changes suggest increased biphenyl planarity within the helical environment, which would be expected to increase the D-A character of FBP-dG. In the fully paired duplex with a pyrimidine 5' to the FBP-dG, the emission was quenched compared to the single strand (Figure 5.2, Table 5.3). The MD simulations coupled with the <sup>19</sup>F NMR data suggests that FBP-dG can adopt an intercalated syn conformation that favors  $\pi$ -stacking interactions between the biphenyl and the flanking bases. A wide variety of fluorescent nucleobase probes display quenched

emission in duplex structures and this is typically ascribed to  $\pi$ -stacking interactions, which can lead to radiationless decay.<sup>51</sup> In the SMI duplex the emission intensity is turned on, because the biphenyl and the 5'-flanking base are unpartnered. (Figure 5.2). The turn-on emission response of FBP-dG in the SMI duplex may be useful for monitoring SMI formation in the NarI sequence upon synthesis with DNA polymerases.

Table 5.3: Properties of C-linked C8-aryl-dG adducts in fully paired and SMI NarI 12-mer duplexes.

| Aryl group | Duplex       | $E_{\text{syn}}/E_{\text{anti}}^a$ | $\Delta T_m^b$ | $I_{\text{rel}}^c$ |
|------------|--------------|------------------------------------|----------------|--------------------|
| FBP        | fully paired | 2.3/0.0                            | -12.5          | 0.64               |
|            | SMI          | 2.8/0.0                            | +6.0           | 4.2                |

<sup>a</sup> Calculated relative energies (kJ/mol) of *syn* versus *anti* adduct conformation. <sup>b</sup>  $\Delta T_m$  (°C) of adducted duplex vs. unmodified control. <sup>c</sup> Relative emission intensity of ss vs fully paired or SMI vs fully paired duplexes.

To our knowledge, there is no literature reporting an O-linked biphenyl-dG adduct for comparison the corresponding C- and N-linked derivatives. However, the structural influence of the O-linked phenol-dG adduct has previously been published.<sup>8</sup> <sup>19</sup>F NMR studies and MD calculations indicate that O-linked phenyl-dG favors a B-form structure.<sup>8</sup> This is also suggested for the N-linked aniline-dG adduct<sup>52</sup> and the C-linked phenyl-dG.<sup>29</sup> Compared to the duplex containing the O-linked derivative, the B-form duplex containing the C-linked phenyl-dG adduct is less stable ( $\Delta T_m = -15.7^{29}$  vs  $-10.8$  °C<sup>8</sup>) due to the stronger *syn* preference of the C-linked lesion that lacks the flexible tether separating the phenyl ring from the nucleobase. This increased *syn* preference for the C-linked lesion also provides a rationale for the greater propensity of the FBP-dG lesion to produce the I-type



structure compared to the N-linked ABP-dG adduct that favors the B-type structure. In fact, the C-linked FBP-dG behaves more like N-linked AF-dG in the NarI oligonucleotide than the corresponding ABP-dG. The methylene linkage between the two aromatic rings in AF-dG makes it planar for efficient  $\pi$ -stacking interactions, which provides a rationale for its much greater propensity to produce the stacked S-type conformation and stabilize the SMI duplex.<sup>31,34</sup> For the C-linked FBP-dG adduct, its increased syn preference must stem from the difference in the length and flexibility of the linker between biphenyl and dG, and this may also influence how the biphenyl is situated in the helix.

## 5.5 Conclusions

We have inserted a C-linked C8-biphenyl-dG bearing a fluorine atom (FBP-dG) into the G<sub>3</sub>-site of the 12-mer *NarI* sequence, and characterized its emission response and impact on fully paired and slippage mutagenic intermediate (SMI) duplex stability. In the fully paired *NarI* duplex, MD simulations combined with <sup>19</sup>F NMR demonstrate conformational heterogeneity between a major-groove B-type *anti*-adduct conformation (B) and an intercalated *syn*-adduct conformation (I), with a predicted B:I ratio of 20:80. This ratio contrasts the expected *anti:syn* ratio of 67:33 for the corresponding N-linked aminobiphenyl-dG (ABP-dG), but is similar to the corresponding ratio (25:75 in *NarI*) for the N-linked 2-aminofluorene-dG (AF-dG), which adopts a planar adduct conformation due to the methylene linkage between the biphenyl rings. MD simulations for the C-linked FBP-dG adduct predict increased *syn* preference due to the lack of the flexible linker separating the dG component from the biphenyl moiety, which provides a rationale for its AF-dG-like properties within the fully paired and SMI *NarI* duplexes. In the SMI duplex, FBP-dG and the 5'-flanking base are unpartnered within a 2-base bulge and a turn-on

fluorescence response is observed. This response suggests the potential of such modifications in DNA to enable emission monitoring the formation of SMI structures, which are mediated by polymerase processing of chemically damaged DNA, and thereby understanding lesion-induced mutagenicity.

## 5.6 References

- (1) Beland, F. A., and Kadlubar, F. F. (1990) Chemical carcinogenesis and mutagenesis, in *Handbook of Experimental Pharmacology* (Cooper, C. S.; Grover, P. L., Eds.) Vol. 94/I, pp 267-325, Springer-Verlag, Heidelberg.
- (2) Luch, A. (2005) Nature and nurture – lessons from chemical carcinogenesis. *Nat. Rev. Cancer* 5, 113-125.
- (3) Millen, A. L., Sharma, P., and Wetmore, S. D. (2012) C8-linked bulky guanosine DNA adducts: Experimental and computational insights into adduct conformational preferences and resulting mutagenicity. *Future Med. Chem.* 4, 1981-2007.
- (4) Patel, D. J., Mao, B., Gu, Z., Hingerty, B. E., Gorin, A., Basu, A. K., and Broyde, S. (1998) Nuclear magnetic resonance solution structures of covalent aromatic amine-DNA adducts and their mutagenic relevance. *Chem. Res. Toxicol.* 11, 391-407.
- (5) Cho, B. P. (2004) Dynamic conformational heterogeneities of carcinogen-DNA adducts and their mutagenic relevance. *J. Environ. Sci. Health.* 22, 57-90
- (6) Dai, J., Wright, M. W., and Manderville, R. A. (2003) An oxygen-bonded C8-deoxy-guanosine nucleoside adduct of pentachlorophenol by peroxidase activation: evidence for ambident C8 reactivity by phenoxyl radicals. *Chem. Res. Toxicol.* 16, 817-821.
- (7) Dai, J., Sloat, A. L., Wright, M. W., and Manderville, R. A. (2005) Role of phenoxyl radicals in DNA adduction by chlorophenol xenobiotics following peroxidase activation. *Chem. Res. Toxicol.* 18, 771-779.
- (8) Kuska, M. S., Witham, A. A., Sproviero, M., Manderville, R. A., Majdi Yazdi, M., Sharma, P., and Wetmore, S. D. (2013) Structural influence of C8-phenoxy-guanine in the *NarI* recognition DNA sequence. *Chem. Res. Toxicol.* 26, 1397-1408.
- (9) Witham, A. A., Verwey, A. M. R., Sproviero, M., Manderville, R. A., Sharma, P., and Wetmore, S. D. (2015) Chlorine functionalization of a model phenolic C8-guanine adduct increases conformational rigidity and blocks extension by a Y-family DNA polymerase. *Chem. Res. Toxicol.* 28, 1346-1356.

- (10) Manderville, R. A. (2005) Ambident reactivity of phenoxyl radicals in DNA adduction. *Can. J. Chem.* 83, 1261-1267.
- (11) Dai, J., Wright, M. W., and Manderville, R. A. (2003) Ochratoxin A forms a carbon-bonded C8-deoxyguanosine nucleoside adduct: Implication for C8-reactivity by a phenolic radical. *J. Am. Chem. Soc.* 125, 3716-3717.
- (12) Faucet, V., Pfohl-Leszkowicz, A., Dai, J., Castegnaro, M., and Manderville, R. A. (2004) Evidence for covalent DNA adduction by ochratoxin A following chronic exposure to rat and subacute exposure to pig. *Chem. Res. Toxicol.* 17, 1289-1296.
- (13) Mantle, P. G., Faucet-Marquis, V., Manderville, R. A., Squillaci, B., and Pfohl-Leszkowicz, A. (2010) Structures of covalent adducts between DNA and ochratoxin A: A new factor in debate about genotoxicity and human risk assessment. *Chem. Res. Toxicol.* 23, 89-98.
- (14) Sharma, P., Manderville, R. A., and Wetmore, S. D. (2014) Structural and energetic characterization of the major DNA adduct formed from the food mutagen ochratoxin A in the NarI hotspot sequence: influence of adduct ionization on the conformational preferences and implications for the NER propensity. *Nucleic Acids Res.* 42, 11831-11845.
- (15) Manderville, R. A., and Wetmore, S. D. (2017) Mutagenicity of ochratoxin A: role for a carbon-linked C8-deoxyguanosine adduct? *J. Agric. Food Chem.* DOI: 10.1021/acs.jafc.6b03897.
- (16) Rogan, E. G., Cavalieri, E. L., Tibbels, S. R., Cremonesi, P., Warner, C. D., Nagel, D. L., Tomer, K. B., Cerney, R. L., and Gross, M. L. (1988) Synthesis and identification of benzo[a]pyrene-guanine nucleoside adducts formed by electrochemical oxidation and by horseradish peroxidase catalyzed reaction of benzo[a]pyrene with DNA. *J. Am. Chem. Soc.* 110, 4023-4029.
- (17) Hiramoto, K., Kaku, M., Sueyoshi, A., Fujise, M., and Kikugawa, K. (1995) DNA base and deoxyribose modification by the carbon-centered radical generated from 4-(hydroxymethyl) benzenediazonium salt, a carcinogen in mushroom. *Chem. Res. Toxicol.* 8, 356-362.
- (18) Gannett, P. M., Powell, J. H., Rao, R., Shi, X., Lawson, T., Kolar, C., and Toth, B. (1999) C8-Arylguanine and C8-aryl原因ine formation in calf thymus DNA from arenediazonium ions. *Chem. Res. Toxicol.* 12, 297-304.
- (19) Cho, B. (2010) Structure-function characteristics of aromatic amine-DNA adducts, in *The Chemical Biology of DNA Damage* (Geacintov, N. E. and Broyde, S., Eds) pp. 217-238, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.
- (20) Cho, B. P., Beland, F. A., and Marques, M. M. (1992) NMR structural studies of a 15-mer DNA sequence from ras protooncogene modified at the first base of codon 61 with the carcinogen 4-aminobiphenyl. *Biochemistry* 31, 9587-9602.

- (21) Mao, B., Hingerty, B. E., Broyde, S., and Patel, D. J. (1998) Solution structure of the aminofluorene [AF]-intercalated conformer of the *syn*-[AF]-C8-dG adduct opposite dC in a DNA duplex. *Biochemistry* 37, 81-94.
- (22) Wang, F., DeMuro, N. E., Elmquist, C. E., Stover, J. S., Rizzo, C. J., and Stone, M. P. (2006) Base-displaced intercalated structure of the food mutagen 2-amino-3-methylimidazo[4,5-*f*]quinoline in the recognition sequence of the *NarI* restriction enzyme, a hotspot for -2 bp deletions. *J. Am. Chem. Soc.* 128, 10085-10095.
- (23) Shapiro, R., Hingerty, B. E., and Broyde, S. (1989) Minor-groove binding models for acetylaminofluorene modified DNA. *J. Biomol. Struct. Dyn.* 7, 493-513.
- (24) Elmquist, C. E., Wang, F., Stover, J. S., Stone, M. P., and Rizzo, C. J. (2007) Conformational differences of the C8-deoxyguanosine adduct of 2-amino-3-methylimidazo[4,5-*f*]quinoline (IQ) within the *NarI* recognition sequence. *Chem. Res. Toxicol.* 20, 445-454.
- (25) Wang, F., Elmquist, C. E., Stover, J. S., Rizzo, C. J., and Stone, M. P. (2007) DNA sequence modulates the conformation of the food mutagen 2-amino-3-methylimidazo[4,5-*f*]quinoline in the recognition sequence of the *NarI* restriction enzyme. *Biochemistry* 46, 8498-8516.
- (26) Sharma, P., Majdi Yazdi, M., Merriman, A., Manderville, R. A., and Wetmore, S. D. (2015) Influence of the linkage type and functional groups in the carcinogenic moiety on the conformational preferences of damaged DNA: structural and energetic characterization of carbon- and oxygen-linked C8-phenolic-guanine adducts. *Chem. Res. Toxicol.* 28, 782-796.
- (27) Manderville, R. A., and Wetmore, S. D. (2016) Understanding the mutagenicity of O-linked and C-linked guanine DNA Adducts: A combined experimental and computational approach. *Chem. Res. Toxicol.* 30, 177-188.
- (28) Josephy, P. D., and Mannervik, B. (2006) *Molecular Toxicology*, 2nd ed., pp 517-518, Oxford University Press, Inc., New York.
- (29) Sproviero, M., Verwey, A. M. R., Rankin, K. M., Witham, A. A., Soldatov, D. V., Manderville, R. A., Fekry, M. I., Sturla, S. J., Sharma, P., and Wetmore, S. D. (2014) Structural and biochemical impact of C8-aryl-guanine adducts within the *NarI* recognition DNA sequence: influence of aryl ring size on targeted and semi-targeted mutagenicity. *Nucleic Acids Res.* 42, 13405-13421.
- (30) Sproviero, M., Verwey, A. M. R., Witham, A. A., Manderville, R. A., Sharma, P., and Wetmore, S. D. (2015) Enhancing bulge stabilization through linear extension of C8-aryl-guanine adducts to promote polymerase blockage or strand realignment to produce a C:C mismatch. *Chem. Res. Toxicol.* 28, 1647-1658.

- (31) Jain, N., Li, Y., Zhang, L., Meneni, S. R., and Cho, B. P. (2007) Probing the sequence effects on NarI-induced -2 frameshift mutagenesis by dynamic  $^{19}\text{F}$  NMR, UV, and CD spectroscopy. *Biochemistry* 46, 13310-13321.
- (32) Jain, V., Hilton, B., Lin, B., Patnaik, S., Liang, F., Darian, E., Zou, Y., Mackerell, A. D., Jr., and Cho, B. P. (2013) Unusual sequence effects on nucleotide excision repair of arylamine lesions: DNA bending/distortion as a primary recognition factor. *Nucleic Acids Res.* 41, 869-880.
- (33) Liang, F., and Cho, B. P. (2010) Enthalpy-entropy contribution to carcinogen-induced DNA conformational heterogeneity. *Biochemistry* 49, 259-266.
- (34) Jain, V., Vaidyanathan, V. G., Patnaik, S., Gopal, S., and Cho, B. P. (2014) Conformational insights into the lesion and sequence effects for arylamine-induced translesion DNA synthesis:  $^{19}\text{F}$  NMR, surface plasmon resonance, and primer kinetic studies. *Biochemistry* 53, 4059-4071.
- (35) Verghis, S. B. M., Essigmann, J. M., Kadlubar, F. F., Morningstar, M. L., and Lasko, D. D. (1997) Specificity of mutagenesis by 4-aminobiphenyl: mutations at G residues in bacteriophage M13 DNA and G→C transversions at a unique dG8-ABP lesion in single-stranded DNA. *Carcinogenesis* 18, 2403-2414.
- (36) Fuchs, R. P. P., Schwartz, N., and Daune, M. P. (1981) Hot spots of frameshift mutations induced by the ultimate carcinogen N-acetoxy-N-2-acetylaminofluorene. *Nature* 294, 657-659.
- (37) Burnouf, D., Koehl, P., and Fuchs, R. (1989) Single adduct mutagenesis: strong effect of the position of a single acetylaminofluorene adduct within a mutation hot spot. *Proc. Natl. Acad. Sci.* 86, 4147-4151.
- (38) Zhou, Y., and Romano, L. J. (1983) Solid-phase synthesis of oligonucleotides containing site-specific N-(2'-deoxyguanosin-8-yl)-2-(acetylaminofluorene) adducts using 9-fluorenylmethoxycarbonyl as the base-protecting group. *Biochemistry* 32, 14043-14052.
- (39) Milhe, C., Fuchs, R. P. P., and Lefevre, J. F. (1996) NMR data show that the carcinogen N-2-acetylaminofluorene stabilises an intermediate of -2 frameshift mutagenesis in a region of high mutation frequency. *Eur. J. Biochem.* 235, 120-127.
- (40) Xu, L., and Cho, B. P. (2016) Conformational Insights into the Mechanism of Acetylaminofluorene-dG-Induced Frameshift Mutations in the NarI Mutational Hotspot. *Chem. Res. Toxicol.* 29, 213-226.
- (41) Maus, M., Rettig, W., Bonafoux, D., and Lapouyade, R. (1999) Photoinduced intramolecular charge transfer in a series of differently twisted donor-acceptor biphenyls as revealed by fluorescence. *J. Phys. Chem. A.* 103, 3388-3401.

- (42) Fukuda, R., and Ehara, M. (2013) Electronic excited states and electronic spectra of biphenyl: a study using many-body wavefunction methods and density functional theories. *Phys. Chem. Chem. Phys.*, *15*, 17426-17434.
- (43) Case, D. A., Babin, V., Berryman, J. T., Betz, R. M., Cai, Q., Cerutti, D. S., Cheatham, III, T. E., Darden, T. A., Duke, R. E., Gohlke, H., Goetz, A. W., Gusarov, S., Homeyer, N., Janowski, P., Kaus, J., Kolossváry, I., Kovalenko, A., Lee, T. S., LeGrand, S., Luchko, T., Luo, R., Madej, B., Merz, K. M., Paesani, F., Roe, D. R., Roitberg, A., Sagui, C., Salomon-Ferrer, R., Seabra, G., Simmerling, C. L., Smith, W., Swails, J., Walker, R. C., Wang, J., Wolf, R. M., Wu X., and Kollman P. A. (2014), AMBER 14, University of California, San Francisco.
- (44) Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and testing of a general AMBER force field. *J. Computat. Chem.* *25*, 1157-1174.
- (45) Case, D. A., Cheatham, III, T. E., Darden, T., Gohlke, H., Luo, R., Merz, Jr., K. M., Onufriev, A., Simmerling, C., Wang B., and Woods., R. (2005) The Amber biomolecular simulation programs. *J. Computat. Chem.* *26*, 1668-1688.
- (46) Miller III, B. R., McGee Jr, T. D., Swails, J. M., Homeyer, N., Gohlke, H., and Roitberg, A. E. (2012) MMPBSA.py: an efficient program for end-state free energy calculations. *J. Chem. Theory Comput.* *8*, 3314-3321.
- (47) Kypr, J., Kejnovská, I., Renčiuk, D., and Vorlíčková, M. (2009) Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.* *37*, 1713–1725.
- (48) Sandineni, A., Lin, B., MacKerell, Jr., A. A., and Cho, B. P. (2013) Structure and thermodynamic insights on acetylaminofluorene-modified deletion DNA duplexes as models for frameshift mutagenesis. *Chem. Res. Toxicol.* *26*, 937-91.
- (49) Zhou, L., Rajabzadeh, M., Traficante, D. D., and Cho, B. P. (1997) Conformational heterogeneity of aryl-amine-modified DNA: <sup>19</sup>F NMR evidence. *J. Am. Chem. Soc.* *119*, 5384-5389.
- (50) Shibutani, S. (2004) Requirements for frame-shift deletion during translesion synthesis. *Environ. Mutagen Res.* *26*, 135-141.
- (51) Rankin, K. M., Sproviero, M., Rankin, K., Sharma, P., Wetmore, S. D., and Manderville, R. A. (2012) C8-Heteroaryl-2'-deoxyguanosine adducts as conformational fluorescent probes in the *NarI* recognition sequence. *J. Org. Chem.* *77*, 10498-10508.
- (52) Shapiro, R., Ellis, S., Hingerty, B. E., and Broyde, S. (1998) Effect of ring size on conformations of aromatic amine–DNA adducts: The aniline–C8 Guanine adduct resides in the B-DNA major groove. *Chem. Res. Toxicol.* *11*, 335-341.

## Chapter 6: Impact of DNA Adduct Size, Number and Relative Position on the Toxicity of Aromatic Amines: A Molecular Dynamics Case Study of <sup>AN</sup>dG and <sup>AP</sup>dG Containing DNA<sup>a,b,c</sup>

### 6.1 Introduction

Humans are commonly exposed to aromatic amines (AAs), a well-known class of DNA damaging agents.<sup>1-3</sup> For example, 4-aminobiphenyl (ABP) was historically used as a rubber antioxidant and dye intermediate, and was more recently identified in tobacco smoke,<sup>4</sup> while aminofluorene (AF) and N2-acetylamino fluorene (AAF) were designed as pesticides.<sup>5</sup> Alternatively, 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP) and 2-amino-3-methylimidazo[4,5-f]quinolone (IQ) are found in cooked meats (e.g., beef, chicken, fish and pork).<sup>2, 6</sup> 1-aminopyrene (AP) is a metabolite of 1-nitropyrene (NP), which is the most abundant nitroarene in diesel and other fossil fuel emissions,<sup>7</sup> and has been found in human urine after exposure to diesel exhaust.<sup>8</sup> The smallest AA, single-ringed aniline (AN), is predominately used to manufacture urethane polymers.<sup>9, 10</sup>

Upon incorporation into the body, AAs can be bioactivated into highly reactive species that bind to the DNA nucleobases to form DNA addition products (adducts), primarily N-linked lesions at the C8 site of guanine (Figure 6.1).<sup>11</sup> Previous work has

---

<sup>a</sup>The *Journal of American Chemical Society* Reference Style is used in this Chapter.

<sup>b</sup>Reprinted with permission from Kung, R. W.; Deak, T. K.; Griffith-Salik, C. A.; Takyi, N. A.; Wetmore, S. D., Impact of DNA Adduct Size, Number, and Relative Position on the Toxicity of Aromatic Amines: A Molecular Dynamics Case Study of <sup>AN</sup>dG- and <sup>AP</sup>dG-Containing DNA Duplexes. *J. Chem. Inf. Model.* **2021**, 61 (5), 2313-2327.

<sup>c</sup>R.W.K., T.K.D., C.A.G.S. and N.A.T. performed molecular dynamics simulations and initial analysis using scripts written by R.W.K. R.W.K. interpreted the data and wrote the first draft of the manuscript.

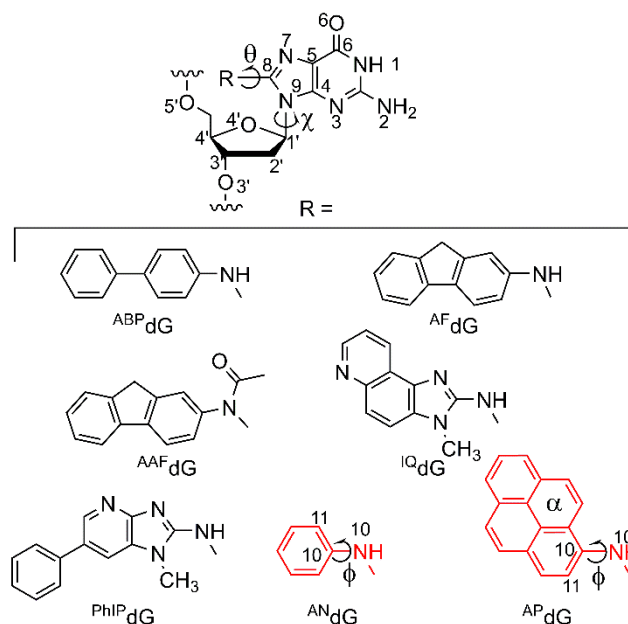


Figure 6.1: Chemical structures, as well as atom numbering and the  $\chi$  ( $\angle(\text{O4}'\text{--C1}'\text{--N9--C4})$ ),  $\theta$  ( $\angle(\text{N9--C8--N10--C10})$ ), and  $\phi$  ( $\angle(\text{C8--N10--C10--C11})$ ) dihedral angles, for C8-dG DNA adducts arising from AA compounds, including the <sup>AN</sup>dG and <sup>AP</sup>dG lesions (red) investigated in the present study.

uncovered three main conformational themes for damaged DNA containing an N-linked C8-dG AA adduct (Figure 6.2A,B, E.1, Appendix E).<sup>12-22</sup> In the first conformation (denoted the major groove or B conformation), the *anti* orientation about the glycosidic bond at the adduct site ( $\chi = 180^\circ \pm 90^\circ$ ) positions the bulky moiety in the DNA major groove and permits canonical Watson-Crick (WC) hydrogen bonding with the opposing strand. In the second (intercalated or stacked, S) conformation, the adduct adopts the *syn* glycosidic orientation ( $\chi = 0^\circ \pm 90^\circ$ ), which places the bulky group in the interstrand space, disrupts base pairing with the complementary strand, and shifts the damaged G toward the major groove. In the final (wedge or W) conformation (Figure E.1, Appendix E), the adduct also adopts the *syn* orientation, but the aryl moiety is located in the DNA minor groove, which aligns the Hoogsteen face of the lesion to hydrogen bond with the opposing base. Many



AA-derived N-linked lesions have the potential to induce multiple damaged DNA conformations, with the exact ratio being heavily dependent on several factors including the adduct size and shape, the presence of heteroatoms, and the sequence context.<sup>13, 14, 23, 24</sup>

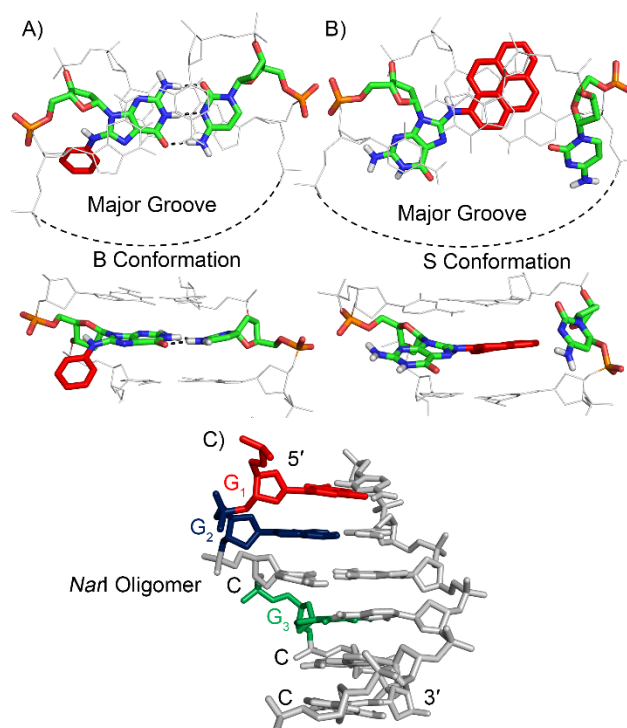


Figure 6.2: Example structures of A) <sup>AN</sup>dG adducted DNA in the B conformation and B) <sup>AP</sup>dG adducted DNA in the S conformation, highlighting the lesion and flanking base pairs. C) *NarI* oligomer, highlighting the sequence context of the damage sites considered in the present work (G<sub>1</sub>, G<sub>2</sub>, and G<sub>3</sub>).

An abundance of experimental and computational work on AA-adducted DNA has revealed correlations between the conformation of the damaged helix and the mutagenicity of AAs. For example, DNA containing <sup>ABP</sup>dG adopts both the B and S conformations<sup>14, 24, 25</sup> and the <sup>ABP</sup>dG lesion causes a mixture of point and frameshift mutations,<sup>26</sup> which correlates with ABP being classified as carcinogenic to humans (Group 1) by the International Agency for Research on Cancer (IARC).<sup>27</sup> Similarly, duplexes containing

<sup>AF</sup>dG or <sup>AAF</sup>dG possess B/S/W conformational heterogeneity,<sup>5, 23, 24</sup> resulting in point mutations, while <sup>AAF</sup>dG also induces base deletions. Although AF and AAF have not been formally classified by the IARC, 2-nitrofluorene leads to the same C8-dG adducts and has been labelled a possible human carcinogen (Group 2B).<sup>7</sup> Four-ringed <sup>AP</sup>dG, one of the largest AA-derived lesions, induces the S conformation of damaged DNA<sup>13, 20, 21</sup> and results in both point and deletion mutations,<sup>28, 29</sup> with the NP precursor to <sup>AP</sup>dG considered to be probably carcinogenic to humans (Group 2A).<sup>7</sup> Notably, the mutagenic adducts discussed to this point can all adopt the *syn* glycosidic conformation to some extent depending on the adduct structure and sequence context. In contrast, the nonmutagenic, single-ringed <sup>AN</sup>dG adduct exclusively leads to the *anti* lesion orientation associated with the B conformation of damaged DNA<sup>13, 30</sup> and the IARC deems AN to be not classifiable as carcinogenic to humans (Group 3).<sup>10</sup>

In direct connection to mutagenicity, the structure of damaged DNA also provides vital clues about the repair propensity of AA-derived adducts. For example, studies on <sup>AAF</sup>dG and <sup>AF</sup>dG in different sequence contexts revealed a positive correlation between the preference for a distorted S conformation of damaged DNA and the resulting repair efficiency.<sup>22</sup> Other work on DNA containing the fluorine-labeled <sup>FAF</sup>dG, <sup>FAAF</sup>dG and <sup>FABP</sup>dG adducts uncovered an increased repairability of damaged duplexes that exhibit greater distortions from canonical DNA structure.<sup>31</sup> Additionally, the repair resistance of <sup>PhP</sup>dG in deletion duplexes compared to the fully paired counterpart can be rationalized based on stronger lesion site stacking interactions that stabilize the deletion structures.<sup>32</sup> Overall, these select examples highlight that uncovering the preferred

conformation of damaged DNA is a critical step toward understanding the severity of the toxicity associated with AA exposure.

Despite the abundance of literature on the structure of AA-adducted DNA, the vast majority of studies have focused on duplexes containing a single adduct (monoadduct). However, although DNA modification spectra indicate that many G sites can be damaged upon exposure to aromatic amines, specific GC rich regions (e.g., the *NarI* sequence) result in a significantly greater number of mutations.<sup>33</sup> In fact, there is a nonlinear increase in the mutational frequency with an increase in the number of damaged G sites.<sup>33-35</sup> For example, an approximate 3-fold increase in the number of AAF-dG adducts in the *URA3* gene results in an approximate 10-fold increase in the mutational frequency.<sup>35</sup> This greater mutagenic rate in G-rich sequences could at least partially arise due to the presence and synergistic effects of multiple adducts in close proximity. Indeed, multiple damaged sites have been shown to exist within 24 base pairs for other types of DNA damage,<sup>29, 36-45</sup> including 8-oxoguanine (8OG), abasic sites, 5-methylcytosine, and single-strand breaks (SSBs).<sup>36-41</sup> Furthermore, there is evidence that the presence of one of these lesions can promote the formation of a second lesion within the same local area.<sup>45, 46</sup> Perhaps more importantly, the repairability of a given lesion has been shown to decrease and the mutagenic potential has been shown to increase when a second damaged site is in close proximity.<sup>29, 37, 38, 40</sup> Although cluster lesions related to damage such as 8OG, abasic sites, 5MeC and SSBs may more readily arise in cells, cluster lesions associated with bulky DNA adducts may also be possible upon exposure to genotoxins, especially in damage-prone GC-rich regions.<sup>46</sup> Indeed, the formation of cluster adducts (diadducts) for aromatic amines could help account for previously unexplainable complex mutational profiles.<sup>35, 47</sup> Furthermore, although there

is large decrease in AA adduct levels with time through lesion repair or cell death, some adducts persist weeks or even months after the initial exposure.<sup>48, 49</sup> Thus, there is a chance that cluster adducts could form, particularly at damage hotspots, upon chronic exposure to even low levels of damaging agents.

Although the above discussion suggests that cluster lesions arising from aromatic amine exposure should be investigated, to the best of our knowledge, only one study to date has considered the impact of cluster adducts on the structure and repair of damaged DNA.<sup>50</sup> Specifically, NMR, CD, UV spectroscopy, and nucleotide excision assays, as well as complementary short-timescale single molecular dynamics (MD) simulations, revealed that fluorine-labeled <sup>FAAF</sup>dG at two sites in the AA deletion hotspot *NarI* sequence (5'-G<sub>1</sub>G<sub>2</sub>CG<sub>3</sub>CC, Figure 6.2C) results in a complex heterogeneous mixture of damaged DNA conformations and repair efficiencies, which vary with the number of adducts and base pairs between the lesions.<sup>50</sup> Despite this insightful information for <sup>FAAF</sup>dG, relatively little is known about lesion mutagenicity and repair efficiency for cluster adducts compared to monoadducts.

Due to the insights computational chemistry has provided about the structure, mutagenicity and repair susceptibility of DNA containing a single AA-dG lesion,<sup>13, 14, 20-22, 25, 30, 51, 52</sup> and the lack of corresponding information for cluster adducts, this study uses detailed MD simulations to investigate the structure and function of <sup>AN</sup>dG and <sup>AP</sup>dG monoadducted and diadducted DNA (Figure 6.3). These two lesions were selected as representative adducts that span the size (single versus four-ringed aryl groups), the induced adducted DNA structure (predominately B versus S),<sup>13, 20, 21, 30</sup> and mutagenicity

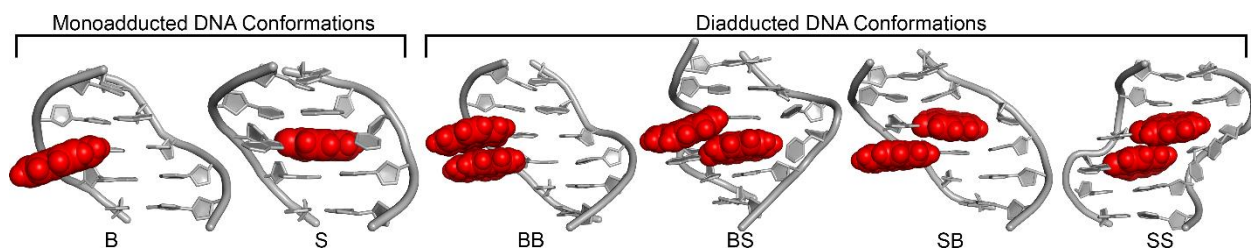


Figure 6.3: Structures of the  $^{AN}dG$  and  $^{AP}dG$  adducts considered in the present work (top), as well as the different monoadducted (B and S) and diadducted (BB, BS, SB and SS) helical conformations modelled for each sequence combination (structures for the  $^{AP}dG$  lesion in the  $G_1$  or  $G_1G_2$  sequence context shown as representative examples). The bulky groups are shown in red.

(nonmutagenic versus probably carcinogenic precursors)<sup>7, 10</sup> of AA-derived N-linked lesions. The adducts are investigated in the context of the *NarI* sequence<sup>53, 54</sup> to complement existing studies on  $^{FAAF}dG$  diadducts,<sup>50</sup> and expand upon the  $G_3$  sequence context previously considered for  $^{AN}dG$  and  $^{AP}dG$  monoadducted DNA.<sup>13</sup> Initially, 12 monoadducted duplexes are considered with  $^{AN}dG$  or  $^{AP}dG$  in (*anti* and *syn*) glycosidic orientations that correspond to the B and S DNA conformations at each of three sites ( $G_1$ ,  $G_2$  and  $G_3$ , Table 1), which reveals clues regarding sequence dependent mutagenicity and provide a point of reference for understanding the impact of the second lesion. Subsequently, 24 diadducted DNA oligomers are considered with two  $^{AN}dG$  or  $^{AP}dG$  adducts in all four *anti/syn* (B/S) conformational combinations in three sequence contexts ( $G_1G_2$ ,  $G_1G_3$  and  $G_2G_3$ , Table 1), which represent different distances between the lesions. This work fills a critical void in the literature by providing the first in-depth structural analysis of diadducted DNA that systematically compares the influences of lesion size, number, and sequence context (spatial separation). Through this investigation, we uncover a differential impact of a second lesion on the helical conformational preferences, and

Table 6.1: DNA duplexes modelled in the current study.<sup>a</sup>

| System       | Adduct           | Label                                                   | Adduct Position and Conformation |                |                |
|--------------|------------------|---------------------------------------------------------|----------------------------------|----------------|----------------|
|              |                  |                                                         | G <sub>1</sub>                   | G <sub>2</sub> | G <sub>3</sub> |
| Undamaged    | None             | Natural                                                 | –                                | –              | –              |
| Monoadducted | <sup>AN</sup> dG | G <sub>1</sub> <sup>B</sup>                             | B                                | –              | –              |
|              |                  | G <sub>1</sub> <sup>S</sup>                             | S                                | –              | –              |
|              |                  | G <sub>2</sub> <sup>B</sup>                             | –                                | B              | –              |
|              |                  | G <sub>2</sub> <sup>S</sup>                             | –                                | S              | –              |
|              |                  | G <sub>3</sub> <sup>B</sup>                             | –                                | –              | B              |
|              |                  | G <sub>3</sub> <sup>S</sup>                             | –                                | –              | S              |
|              | <sup>AP</sup> dG | G <sub>1</sub> <sup>B</sup>                             | B                                | –              | –              |
|              |                  | G <sub>1</sub> <sup>S</sup>                             | S                                | –              | –              |
|              |                  | G <sub>2</sub> <sup>B</sup>                             | –                                | B              | –              |
|              |                  | G <sub>2</sub> <sup>S</sup>                             | –                                | S              | –              |
|              |                  | G <sub>3</sub> <sup>B</sup>                             | –                                | –              | B              |
|              |                  | G <sub>3</sub> <sup>S</sup>                             | –                                | –              | S              |
| Diadducted   | <sup>AN</sup> dG | G <sub>1</sub> <sup>B</sup> G <sub>2</sub> <sup>B</sup> | B                                | B              | –              |
|              |                  | G <sub>1</sub> <sup>B</sup> G <sub>2</sub> <sup>S</sup> | B                                | S              | –              |
|              |                  | G <sub>1</sub> <sup>S</sup> G <sub>2</sub> <sup>B</sup> | S                                | B              | –              |
|              |                  | G <sub>1</sub> <sup>S</sup> G <sub>2</sub> <sup>S</sup> | S                                | S              | –              |
|              |                  | G <sub>1</sub> <sup>B</sup> G <sub>3</sub> <sup>B</sup> | B                                | –              | B              |
|              |                  | G <sub>1</sub> <sup>B</sup> G <sub>3</sub> <sup>S</sup> | B                                | –              | S              |
|              |                  | G <sub>1</sub> <sup>S</sup> G <sub>3</sub> <sup>B</sup> | S                                | –              | B              |
|              |                  | G <sub>1</sub> <sup>S</sup> G <sub>3</sub> <sup>S</sup> | S                                | –              | S              |
|              |                  | G <sub>2</sub> <sup>B</sup> G <sub>3</sub> <sup>B</sup> | –                                | B              | B              |
|              |                  | G <sub>2</sub> <sup>B</sup> G <sub>3</sub> <sup>S</sup> | –                                | B              | S              |
|              |                  | G <sub>2</sub> <sup>S</sup> G <sub>3</sub> <sup>B</sup> | –                                | S              | B              |
|              |                  | G <sub>2</sub> <sup>S</sup> G <sub>3</sub> <sup>S</sup> | –                                | S              | S              |
|              | <sup>AP</sup> dG | G <sub>1</sub> <sup>B</sup> G <sub>2</sub> <sup>B</sup> | B                                | B              | –              |
|              |                  | G <sub>1</sub> <sup>B</sup> G <sub>2</sub> <sup>S</sup> | B                                | S              | –              |
|              |                  | G <sub>1</sub> <sup>S</sup> G <sub>2</sub> <sup>B</sup> | S                                | B              | –              |
|              |                  | G <sub>1</sub> <sup>S</sup> G <sub>2</sub> <sup>S</sup> | S                                | S              | –              |
|              |                  | G <sub>1</sub> <sup>B</sup> G <sub>3</sub> <sup>B</sup> | B                                | –              | B              |
|              |                  | G <sub>1</sub> <sup>B</sup> G <sub>3</sub> <sup>S</sup> | B                                | –              | S              |
|              |                  | G <sub>1</sub> <sup>S</sup> G <sub>3</sub> <sup>B</sup> | S                                | –              | B              |
|              |                  | G <sub>1</sub> <sup>S</sup> G <sub>3</sub> <sup>S</sup> | S                                | –              | S              |
|              |                  | G <sub>2</sub> <sup>B</sup> G <sub>3</sub> <sup>B</sup> | –                                | B              | B              |
|              |                  | G <sub>2</sub> <sup>B</sup> G <sub>3</sub> <sup>S</sup> | –                                | B              | S              |
|              |                  | G <sub>2</sub> <sup>S</sup> G <sub>3</sub> <sup>B</sup> | –                                | S              | B              |
|              |                  | G <sub>2</sub> <sup>S</sup> G <sub>3</sub> <sup>S</sup> | –                                | S              | S              |

<sup>a</sup>Each system was simulated for 500 ns in triplicate, resulting in a total of 1.5 μs production time for each duplex.

therefore the replication and repair outcomes, depending on the lesion size and spatial separation. As a result, our work expands the critical roles of adduct size and sequence context known for monoadducted DNA structure to diadducted DNA, and further uncovers the biological implications of cluster lesions in the toxicity of AAs.

## 6.2 Computational Details

The <sup>AN</sup>dG and <sup>AP</sup>dG lesions (Figure 6.1) were modelled in a 16-mer DNA strand containing the *NarI* sequence (5'–CTCTCG<sub>1</sub>G<sub>2</sub>CG<sub>3</sub>CCATCAC–3', underlined), which was paired with the fully complementary strand (Figure 6.2C). Specifically, each adduct was initially incorporated at a single G<sub>1</sub>, G<sub>2</sub> or G<sub>3</sub> site, and subsequently simultaneously incorporated at the G<sub>1</sub> and G<sub>2</sub> (G<sub>1</sub>G<sub>2</sub>), G<sub>1</sub> and G<sub>3</sub> (G<sub>1</sub>G<sub>3</sub>), or G<sub>2</sub> and G<sub>3</sub> (G<sub>2</sub>G<sub>3</sub>) sites. All possible combinations of (*anti/syn*) lesion glycosidic orientations were considered. Specifically, monoadducted DNA was considered in the major groove B and intercalated S conformations (Figure 6.3), which were previously reported to be the stable conformations of <sup>AN</sup>dG and <sup>AP</sup>dG monoadducted DNA.<sup>13, 21, 30</sup> For diadducted DNA, four conformational themes were considered (Figure 6.3), namely the BB, BS, SB and SS conformations, with the first symbol corresponding to the 5' lesion conformation. All DNA structures were built based on those previously reported for DNA containing <sup>AN</sup>dG or <sup>AP</sup>dG at G<sub>3</sub>.<sup>13</sup> Undamaged DNA of the same sequence was also considered as a control. Thus, a total of 37 unique systems were studied.

Natural DNA nucleotides were modelled using the parmbsc1 force field in Amber18.<sup>55</sup> Although the bsc1 modification can lead to overstabilization of B-DNA,<sup>56</sup> this force field is accepted to generally provide a good description of DNA.<sup>57, 58</sup> Parameters for the damaged nucleotides are available in the SI and were taken from previously published

work.<sup>13</sup> Specifically, the parameter set for the damaged bases was supplemented with GAFF parameters and the Amber compatible partial charges were generated using the RESP charge fitting scheme in the R.E.D. (Version III.4) software package.<sup>59</sup>

Using the tleap module of Amber18, each helix was solvated in a TIP3P water box such that there was a minimum of 10 Å of water between the solute and the edge of the box in each direction. Each system was neutralized with sodium cations. Based on previous work,<sup>60</sup> it is expected that the key structural features of the damaged DNA around the lesion site will be relatively independent of the monovalent ion concentration under typical salt conditions. Subsequently, minimization, heating and equilibration steps were completed following the comprehensive protocol previously used to study other N-linked adducts.<sup>61</sup> Specifically, minimization was conducted over four stages that consisted of 2500 steps of steepest decent followed by 2500 steps of conjugate gradient minimization. The first stage was completed with a restraint on the solute atoms, while the second stage was performed with a restraint on the heavy atoms. A restraint was again applied to the solute in the third minimization stage, while the whole system was minimized without restraints in the final step. The restraints in the first three minimization stages were applied using a force constant of 100 kcal/(mol Å<sup>2</sup>). The system was then heated by increasing the temperature to 310 K in six 20 ps stages at constant volume, with a 25 kcal/(mol Å<sup>2</sup>) restraint applied to the solute. Equilibration was then performed for 100 ps, with the restraint on the solute weakened every 20 ps from 20 to 15, 10, 5 and 1.5 kcal/(mol Å<sup>2</sup>). All simulations used a 10 Å nonbonded cutoff.

To obtain optimal starting structures for the final MD production trajectories, the equilibrated systems were initially simulated for 100 ns. A representative structure for each



system was then obtained based on the configuration of the lesion site(s) using the cluster command within cpptraj.<sup>62</sup> Specifically, the hieragglo algorithm was used on key dihedral angles at the lesion site(s) (including  $\chi$  ( $\angle(\text{O4}'\text{--C1}'\text{--N9--C4})$ ),  $\theta$  ( $\angle(\text{N9--C8--N10--C10})$ ), and  $\phi$  ( $\angle(\text{C8--N10--C10--C11})$ , Figure 6.1). The resulting representative structures were then minimized, heated and equilibrated using the same protocol discussed above prior to the final production trajectories. During the production phase, a 25 kcal/(mol Å<sup>2</sup>) restraint was applied to the hydrogen bonds in the terminal base pairs to prevent helical fraying.<sup>63</sup> 500 ns MD production simulations were carried out in triplicate using different initial velocities and a 2 fs timestep.<sup>65-67</sup> All MD simulations were conducted using the Amber18 suite.<sup>68</sup> Overall, 111 MD production simulations were performed, generating 27  $\mu\text{s}$  of data for each lesion (55.5  $\mu\text{s}$  total, including undamaged DNA).

Each MD replicate was independently analyzed every 50 ps. Backbone root-mean-square deviations (RMSDs) were calculated with respect to the first structure of the production phase using the cpptraj tool in Amber18 and signify that all systems are stable (Table E.1, Appendix E). The cpptraj tool in Amber18 was also used to calculate key dihedral angles at the lesion site(s) (including  $\chi$ ,  $\theta$  and  $\phi$ , Figure 6.1) and the solvent accessible surface area (SASA) of the bulky group. The lesion site interaction energy (LSIE) was calculated as the sum of all noncovalent interactions between the adduct and the opposing base, the adduct and the four flanking nucleobases, and the opposing base and the four flanking nucleobases using the linear interaction energy (LIE) command in cpptraj. The hydrogen-bonding occupancies were calculated using a heavy atom distance cutoff of 3.4 Å and a heavy atom–hydrogen–heavy atom angle cutoff of 120°. The base pairing at the terminal ends was maintained in all trajectories (Table E.2, Appendix E), while the

lesion site WC base pairing was maintained when damaged DNA adopts the B conformation (Tables E.3, E.4, Appendix E). Hydrogen bonding in the base pairs flanking the lesion site was also preserved unless stated otherwise in the main text (Tables E.5-E.8, Appendix E). Since the three replicas for each system sampled similar structures, final representative structures were obtained by clustering over the total 1.5  $\mu$ s dataset for each system based on atoms in the nucleotide base pairs between G<sub>1</sub> and G<sub>3</sub> (i.e., the 5'–G<sub>1</sub>G<sub>2</sub>CG<sub>3</sub>–3' duplex, Figures E.2-E.7, Appendix E). The pseudobase step helical parameters were determined by treating the *NarI* sequence as one unit and calculating the pseudostep parameters between the C5:G28 and A12:T21 base pairs (Figure E.8, Appendix E) using the Curves+/Canal program.<sup>69</sup> Reported averages were calculated over the three replicates, with standard deviations calculated between the three averages. The block standard error (BSE) of the dihedral angles, SASA and LSIE values were also calculated to account for autocorrelation using the block averaging method, with the values extrapolated from all frames of the three replicas fitted to a monomolecular growth model.

## 6.3 Results

### 6.3.1 General Structural Features of Monoadducted DNA

#### 6.3.1.1 B Conformation of Monoadducted DNA

As per undamaged DNA, regardless of the adduct (<sup>AN</sup>dG or <sup>AP</sup>dG) or damage location (G<sub>1</sub>, G<sub>2</sub> or G<sub>3</sub>), the lesion adopts an *anti* orientation about the glycosidic bond in the B conformation of damaged DNA ( $\chi^{\text{avg}} \sim 200 - 230^\circ$ , Figure E.9, Table E.9, Appendix E). This lesion orientation exposes the aryl moiety to solvent in the major groove, directed toward the 3' end of the damaged strand (Figures 6.2A, E.2, Appendix E). While the bulky moiety for the smaller <sup>AN</sup>dG lesion is slightly nonplanar with respect to the damaged G

( $\theta^{\text{avg}} = 128 - 142^\circ$ ), the larger  $^{\text{AP}}\text{dG}$  adduct is more planar ( $\theta^{\text{avg}} = 157 - 177^\circ$ , Figure E.10, Table E.10, Appendix E), regardless of sequence context. The reduced twist in  $^{\text{AP}}\text{dG}$  has been previously noted at  $\text{G}_3$ ,<sup>13</sup> which was attributed to weak hydrogen bonding between the  $\alpha$ -fused ring (Figure 6.1) and the DNA backbone. Nevertheless,  $\phi^{\text{avg}}$  adopts a narrow range regardless of the lesion or position ( $1 - 17^\circ$ , Figure E.11, Table E.11, Appendix E). The resulting lesion orientation allows for hydrogen bonding between the linker N–H and O4' of the damaged nucleotide in all systems (most prevalent for  $^{\text{AN}}\text{dG}$  at  $\text{G}_2$  and  $\text{G}_3$ ), as well as between the linker N–H and O3' of the 5'-dC with respect to the lesion for  $^{\text{AN}}\text{dG}$  at  $\text{G}_1$  (Figure E.12, Table E.12, Appendix E). Since the C opposing the lesion also adopts the *anti* orientation, the B conformation preserves the three Watson-Crick hydrogen bonds of undamaged DNA (occupancy > 97%, Table E.3, Appendix E) and stacking with the flanking nucleotides at the lesion site (Table E.13, Appendix E). The helical parameters of undamaged DNA are not significantly impacted by the presence of a single *anti*  $^{\text{AN}}\text{dG}$  or  $^{\text{AP}}\text{dG}$  lesion regardless of the sequence context (Figure E.2, Appendix E), with the largest deviation being < 2 Å for the rise, shift and slide, and < 10° for the roll, tilt, twist and axial bend (Table E.14, Appendix E).

### 6.3.1.2 S Conformation of Monoadducted DNA

In the S conformation of damaged DNA, the single lesion adopts the *syn* glycosidic orientation ( $\chi^{\text{avg}} \sim 40 - 60^\circ$ , Figure E.9, Table E.9, Appendix E). As a result, the aryl moiety is intercalated into the helical core, which displaces the opposing C into the major groove in the 3' direction of the damaged strand (Figure 6.2B, E.3, Appendix E). Nevertheless, the hydrogen bonding in the flanking base pairs is maintained (occupancy > 92%, Table E.5, Appendix E). While the larger AP moiety adopts a nearly planar orientation with respect to

dG ( $\theta^{\text{avg}} = 186 - 194^\circ$ , Figure E.10, Appendix E), the AN moiety is perpendicular to the plane of the nucleobase ( $\theta^{\text{avg}} \sim 90^\circ$  or  $\sim 270^\circ$ , Figures E.13, E.14, Table E.10, Appendix E). Regardless,  $\phi^{\text{avg}}$  is consistent across adducts and sequences ( $\phi^{\text{avg}} = 0 \pm 7^\circ$ , Figure E.11, Table E.11, Appendix E). The greater nonplanarity of <sup>AN</sup>dG in the S damaged DNA conformation arises due to shifting of the bulky group toward the minor groove (Figures E.3, E.14), resulting in a W-like structure. Indeed, (albeit limited) hydrogen bonding occurs between the Hoogsteen face of <sup>AN</sup>dG and the amino group of the opposing C, as well as the linking N–H of the adduct and the flanking bases (Table E.12, Appendix E). In contrast, hydrogen bonding with the pairing C is absent for <sup>AP</sup>dG. Instead, the entire damaged G stacks with the surrounding nucleobases (Figure E.3, Appendix E). The S conformations of monoadducted DNA are modestly distorted from canonical DNA, with the largest deviations in helical parameters occurring in the slide (up to 3.7 Å change), roll (up to 21.9° change), tilt (up to 22.9° change) and twist (up to 17.0° change, Figure E.3, Table E.14, Appendix E). The <sup>AP</sup>dG lesion affects more helical parameters at each site than <sup>AN</sup>dG (Table E.15, Appendix E), which may at least in part arise due to <sup>AN</sup>dG-containing DNA frequently adopting a W-like conformation.

### 6.3.2 Conformational Preference of Monoadducted DNA

To identify the preferred conformation(s) of damaged DNA containing a single <sup>AN</sup>dG or <sup>AP</sup>dG lesion at different locations ( $G_1$ ,  $G_2$  or  $G_3$ ), differences in the structural features of the corresponding damaged helices in each potential conformation must be carefully compared. Our structural analysis reveals that the presence of a single lesion results in minimal helical distortion and therefore global helical structural changes occurring outside the local lesion site will not significantly influence the relative conformer stability for these

systems. To better understand the impact of the structural changes at the lesion site for a given damaged DNA system, we use an analogous approach to that successfully applied to DNA containing N-linked C8-dG adducts of varying bulky moiety size.<sup>13</sup> Specifically, the impact of the lesion site conformation on both stabilizing (noncovalent interactions at the lesion site) and destabilizing (SASA of the aryl group) factors are considered. We note that, although discrete solute–solvent interactions at the lesion site may also change with the damaged DNA conformation, the largest difference between damaged DNA conformations will be the apolar interactions arising from the major groove versus intercalated placement of the hydrophobic bulky moiety.

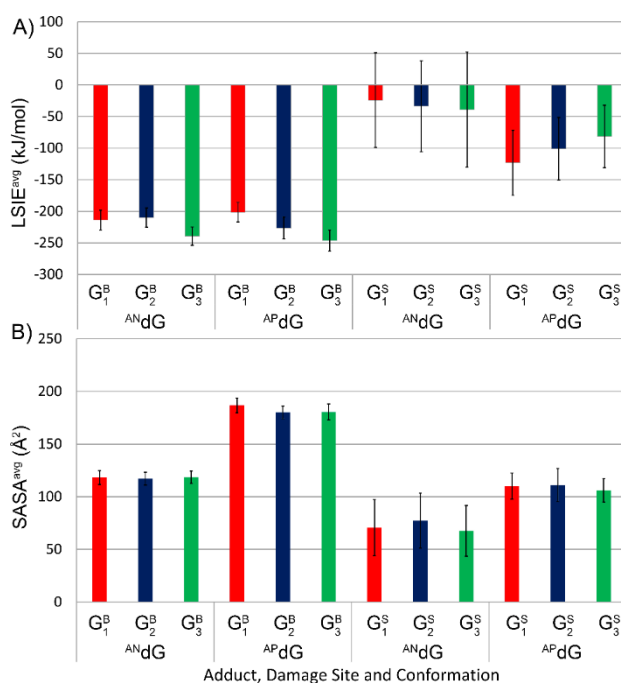


Figure 6.4: A) LSIE<sup>avg</sup> and B) SASA<sup>avg</sup> for DNA containing a single <sup>AN</sup>dG or <sup>AP</sup>dG lesion at G<sub>1</sub> (red), G<sub>2</sub> (blue) or G<sub>3</sub> (green), including the dataset standard deviation.

The LSIE is estimated as the sum of the hydrogen-bonding and stacking interactions within the lesion base pair, and between the damaged and flanking base pairs. In the B

conformation, the  $LSIE^{avg}$  ranges from  $-202$  to  $-246$  kJ/mol regardless of the size of the aryl moiety (Figures 6.4A, E.15, Table E.13, Appendix E). For DNA containing  $^{AN}dG$ ,  $LSIE^{avg}$  is  $\sim 25 - 30$  kJ/mol larger when the lesion is at  $G_3$  compared to  $G_1$  or  $G_2$ , while  $LSIE^{avg}$  decreases by  $\sim 20$  kJ/mol with  $^{AP}dG$  position as  $G_3 > G_2 > G_1$ . Thus, there is a modest sequence context dependence for this stabilizing factor, which results in the B conformer being most stable at  $G_3$ . In contrast, the  $SASA^{avg}$  is independent of sequence, being larger for  $^{AP}dG$  ( $180 - 187 \text{ \AA}^2$ ) than  $^{AN}dG$  ( $117 - 119 \text{ \AA}^2$ , Figures 6.4B, E.16, Table E.16, Appendix E). This suggests that the B conformation will be less stable for the larger  $^{AP}dG$  lesion than  $^{AN}dG$  due to increased hydrophobic interactions of the bulkier aryl moiety.

In the S conformation of damaged DNA,  $^{AN}dG$  results in very weak  $LSIE^{avg}$  ( $-24$  to  $-39$  kJ/mol, Figures 6.4A, E.15, Table E.13, Appendix E), being relatively sequence independent. The significant decrease in the  $LSIE^{avg}$  for the S compared to the B conformation of  $^{AN}dG$  monoadducted DNA is accompanied by a modest decrease in the  $SASA^{avg}$  (by  $40 - 52 \text{ \AA}^2$ , Figures 6.4B, E.16, Table E.16, Appendix E). However, the standard deviations in  $LSIE^{avg}$  and  $SASA^{avg}$  are large for the S conformers ( $> 70$  kJ/mol and  $> 24 \text{ \AA}^2$ ) due to the aryl moiety shifting from an intercalated (S) to a minor groove (W) position (Figure E.3, Appendix E). This points to the inability of *syn*  $^{AN}dG$  to form strong stacking interactions in the S conformation of damaged DNA. As a result, as previously reported at  $G_3$ ,<sup>13</sup> DNA containing  $^{AN}dG$  can be concluded to be exclusively stabilize in the B conformation in all sequence contexts (Figure 6.5).

The  $LSIE^{avg}$  is significantly more stabilizing for  $^{AP}dG$  than  $^{AN}dG$  adducted DNA in the S conformation (by up to  $\sim 100$  kJ/mol, Figure 6.4A, Table E.13, Appendix E). Furthermore, the overall solvent exposure of the  $^{AP}dG$  aryl group is considerably less for

the S conformer than that for the B conformer (by 69 – 77 Å<sup>2</sup>, Figure 6.4B, Table E.16, Appendix E). Thus, the lesion-site stability provided to the S conformation of damaged DNA by the enhanced stacking potential of the larger AP moiety, coupled with the greater destabilization of the B conformation due to significant solvent exposure of the hydrophobic bulky moiety, will lead to a preference for the S conformer of <sup>AP</sup>dG monoadducted DNA for all sequence contexts (Figure 6.5) as previously reported at G<sub>3</sub>.<sup>13</sup>

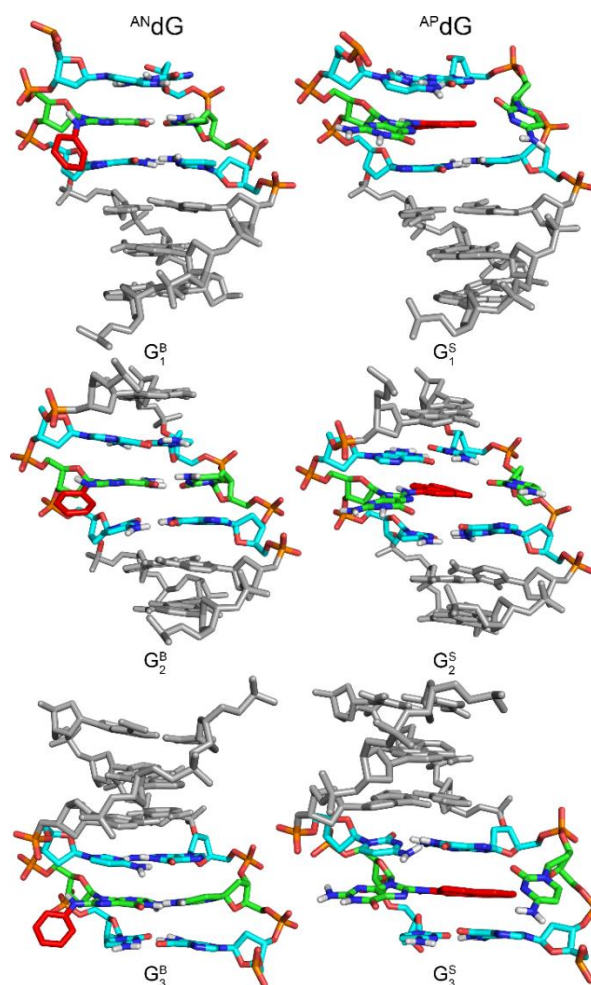


Figure 6.5: MD representative structures of the preferred <sup>AN</sup>dG and <sup>AP</sup>dG monoadducted DNA conformations in different sequence contexts. Only the central base pairs from the 5' flanking base pair of G<sub>1</sub> to the 3' flanking base pair of G<sub>3</sub> in the *NarI* sequence context are shown in each case, with the adducted bases and pairing Cs shown in green, the bulky moieties in red, and the flanking bases in cyan.

Interestingly,  $\text{LSIE}^{\text{avg}}$  for the S conformation exhibits a notable sequence dependence, decreasing by  $\sim 30$  kJ/mol as  $G_1 > G_2 > G_3$  (Figure 6.4A, Table E.13, Appendix E), which is opposite to the trend discussed for the B conformer. Notably, the greater stability at  $G_1$  arises from more favourable interactions between the adduct and pairing C (Table E.13, Appendix E), likely caused by changes in the range of motion in the displaced C with the identity of the flanking bases. Therefore, there is a modest sequence effect on the relative conformer stability of  $^{\text{AP}}\text{dG}$  adducted DNA, with the stability of the S conformation being greatest in the  $G_1$  sequence context and least stable at  $G_3$ .

It is interesting to consider whether the presence of one adduct increases the likelihood of a second lesion forming within close proximity. Our data indicates that a single lesion does not significantly change the structure of the surrounding DNA, suggesting that the major feature controlling the impact of one adduct on the formation of a second lesion is whether the bulky moiety blocks the major groove nucleobase binding site. As  $^{\text{AP}}\text{dG}$  induces the S conformation of monoadducted DNA, we propose that this lesion will not block access to the major groove side of nearby guanines. In contrast,  $^{\text{AN}}\text{dG}$  monoadducted DNA adopts the B conformation with the bulky moiety in the major groove directed toward the 3' end of the damaged strand. Therefore, the formation of an  $^{\text{AN}}\text{dG}$  lesion at  $G_1$  may block the formation of a second adduct at  $G_2$ . Nevertheless, the dynamical and structural impacts of  $^{\text{AN}}\text{dG}$  at all other sites are minimal. Therefore, the formation of cluster lesions is likely dictated by other factors such as the persistent concentration of the carcinogen.



### 6.3.3 General Structural Features of Diadducted DNA

In order to understand the impact of a second lesion on damaged DNA, the following sections systematically compare the structure and stability of duplexes containing cluster lesions to those containing a single lesion. Due to the large number of cluster lesions considered, the impact of adding a second lesion will be separately assessed for each combination of adduct, damaged site and *anti/syn* glycosidic orientation (B/S damaged DNA conformation). Within the context of the *NarI* sequence (5'-G<sub>1</sub>G<sub>2</sub>CG<sub>3</sub>CC, Figure 6.2C), all diadducts are formed within the same strand, with the G<sub>1</sub>G<sub>2</sub> sequence context representing neighboring lesions, G<sub>2</sub>G<sub>3</sub> being lesions that are one nucleotide removed, and G<sub>1</sub>G<sub>3</sub> involving two intervening nucleotides. The nomenclature used to differentiate between the diadducted systems lists the damage sites in numerical order and indicates the damaged DNA conformation as a superscript. For example, the G<sub>1</sub><sup>B</sup>G<sub>2</sub><sup>S</sup> strand contains an *anti* lesion at G<sub>1</sub> associated with the damaged DNA B conformer and a *syn* lesion at G<sub>2</sub> associated with the damaged DNA S conformer. When the impact of a second lesion is being discussed, the site under consideration will be indicated using bold font. For example, **G<sub>1</sub>**G<sub>2</sub><sup>S</sup> indicates the impact of a lesion at G<sub>2</sub> on the G<sub>1</sub> site is under investigation, while G<sub>1</sub><sup>B</sup>**G<sub>2</sub>**<sup>S</sup> indicates that the change at G<sub>2</sub> due to a lesion at G<sub>1</sub> is being discussed.

#### 6.3.3.1 B Conformation of Diadducted DNA

When <sup>AN</sup>dG or <sup>AP</sup>dG-containing DNA adopts the B conformation, the addition of a second lesion has little impact on the structure near the lesion site. Specifically, key lesion dihedral angles generally change by < 10° ( $\chi^{\text{avg}}$ ,  $\theta^{\text{avg}}$  and  $\phi^{\text{avg}}$ , Figures E.9–E.11, Tables E.17–E.19) and the bulky moiety remains solvent exposed in the major groove (Figure E.16, Table E.20, Appendix E). Hydrogen bonds are also maintained, including WC

hydrogen bonding in the lesion and flanking pairs (Table E.4, E.6–E.8, Appendix E), and hydrogen bonding between the linker N–H and O4' within the damaged nucleotide or O3' of the 5' nucleotide for <sup>AN</sup>dG at G<sub>1</sub> (Figure E.12, Table E.21, Appendix E). The overall helical structure is also generally preserved in diadducted DNA (Figure E.4, Appendix E). However, the degree of distortion increases upon addition of a second lesion, with the extent of deformation primarily depending on the conformation of the second lesion (Tables E.15, E.22, Appendix E). Specifically, for both adducts, the addition of a *syn* lesion (S conformer) results in greater deviations in helical parameters (by up to 6.7 Å and 79.1°) than an *anti* lesion (B conformer, 3.6 Å and 18.6°). Nevertheless, the helical structural deviations are generally only slightly larger for <sup>AP</sup>dG (up to 6.7 Å and 79.1°) than <sup>AN</sup>dG (5.9 Å and 37.6°, Tables E.15, E.22, Appendix E). Although the helical distortions can change with sequence context, the effect is not consistent across lesion sizes and conformations.

### 6.3.3.2 S Conformation of Diadducted DNA

When <sup>AN</sup>dG-containing DNA in the S conformation is subjected to a second damaging event, significant structural changes can occur depending on the lesion size, conformation and relative location. Generally, key adduct dihedral angles are not significantly impacted upon the addition of a second damage site ( $\chi^{\text{avg}}$ ,  $\theta^{\text{avg}}$  and  $\phi^{\text{avg}}$ , Figures E.9–E.11, Tables E.17–E.19, Appendix E). One notable deviation from monoadducted DNA is a > 80° change in the  $\chi$  dihedral angle of <sup>AP</sup>dG in the G<sup>B</sup><sub>1</sub>G<sup>S</sup><sub>2</sub>, G<sup>S</sup><sub>1</sub>G<sup>S</sup><sub>2</sub>, G<sup>S</sup><sub>1</sub>G<sup>S</sup><sub>2</sub>, G<sup>S</sup><sub>2</sub>G<sup>S</sup><sub>3</sub> and G<sup>S</sup><sub>2</sub>G<sup>S</sup><sub>3</sub> systems (Figure E.9, Table E.17, Appendix E), which occurs due to repositioning of ribose to permit hydrogen bonding between the linking N–H and O4', while maintaining the adduct stacking seen in monoadducted DNA (Figures E.17, E.18, Tables E.23, E.24,

Appendix E). Although a second lesion affects  $\theta^{avg}$  of  $^{AN}dG$  due to changes in the bimodal distribution of  $\theta$  ( $\sim 90^\circ$  or  $270^\circ$ ), the aryl ring remains perpendicular to the damaged nucleobase in the minor groove for  $^{AN}dG$  diadducted DNA (Figures E.19, E.20, Appendix E). As for monoadducted DNA, diadducted DNA maintains WC hydrogen bonding in the base pairs flanking the damaged sites ( $> 90\%$  occupancy, Tables E.6–E.8, Appendix E), with the exception of  $^{AP}dG$  at  $G_2^S G_3^B$  and  $G_2^S G_3^S$  (occupancies as low as 42%). The intrahelical and extrahelical positions of the opposing C are maintained for the  $^{AN}dG$  and  $^{AP}dG$  lesions, respectively (Table E.23, Appendix E). Interestingly, both  $^{AP}dG$  adducts generally stack with the corresponding 3' and 5' flanking base pairs in the S conformation of  $G_1^S G_2^S$  and  $G_2^S G_3^S$  damaged DNA, while the  $G_1^S G_2^S$  neighboring adducts exhibit minimal mutual overlap due to the helical twist (Figure E.7, E.18, Appendix E). The helical distortion of diadducted DNA with at least one *syn* lesion (S conformer) is typically greater than that for the corresponding monoadducted helix (Figure E.21, Tables E.15, E.22, Appendix E). However, the extent of distortion depends on the conformation associated with the second lesion, with the B conformation resulting in less distortion (average deviation  $< 6.7$  Å and  $79.1^\circ$ ) than S ( $< 11.5$  Å and  $255.2^\circ$ ). Furthermore, the distortion is dependent on lesion size (i.e.,  $^{AN}dG$  results in smaller changes than  $^{AP}dG$ ) and sequence (i.e., the distortion is the least for two intervening bases ( $G_1 G_3$ ) and the greatest for next-nearest diadducts).

#### 6.3.4 Conformational Preference of Diadducted DNA

The conformational preference of monoadducted DNA highlights the important interplay between the LSIE and SASA of the bulky moiety. To assess the impact of a second adduct on the conformational preference of damaged DNA, the change in  $LSIE^{avg}$  and  $SASA^{avg}$  upon addition of the second lesion was calculated (Figures 6.6, 6.7, Tables

E.20, E.24–E.26, Appendix E). Specifically,  $\Delta(\text{LSIE}^{\text{avg}})$  ( $\Delta(\text{SASA}^{\text{avg}})$ ) denotes the  $\text{LSIE}^{\text{avg}}$  ( $\text{SASA}^{\text{avg}}$ ) of a given adduct site in the presence of a second lesion (diadducted DNA) minus the  $\text{LSIE}^{\text{avg}}$  ( $\text{SASA}^{\text{avg}}$ ) of the same adduct site in the absence of a second lesion (monoadducted DNA). Thus, negative  $\Delta(\text{LSIE}^{\text{avg}})$  and  $\Delta(\text{SASA}^{\text{avg}})$  values reflect increased stability of a lesion conformation upon the addition of the second adduct. Due to the differential conformational preference of  $^{\text{AN}}\text{dG}$  and  $^{\text{AP}}\text{dG}$  monoadducted DNA, the effect of each lesion type will be considered separately below.

#### 6.3.4.1 $^{\text{AN}}\text{dG}$ Diadducted DNA

The impact of a second  $^{\text{AN}}\text{dG}$  lesion on the stability of the *anti* adduct orientation associated with the damaged DNA B conformation is minimal (Figure 6.6, Table E.26, Appendix E). Specially,  $\Delta(\text{LSIE}^{\text{avg}})$  falls between  $-10$  and  $8$  kJ/mol, while select cluster lesions increase shielding of the bulky moiety by up to  $\sim 22$  Å<sup>2</sup>. The largest effects occur for the nearest neighbor G<sub>1</sub>G<sub>2</sub> sequence context, which is most stabilized when both lesions adopt the *anti* conformation ( $\text{G}_1^{\text{P}}\text{G}_2^{\text{B}}$  and  $\text{G}_1^{\text{B}}\text{G}_2^{\text{B}}$ ), and modestly stabilized when the second adduct is in the *syn* orientation ( $\text{G}_1^{\text{S}}\text{G}_2^{\text{B}}$  and  $\text{G}_1^{\text{B}}\text{G}_2^{\text{S}}$ ). Interestingly,  $\text{G}_1^{\text{S}}\text{G}_2^{\text{B}}$  is more solvent shielded than  $\text{G}_1^{\text{P}}\text{G}_2^{\text{S}}$  due to the helical twist and displacement of the damaged *syn* G<sub>1</sub> toward the major groove. All other combinations of lesion conformations and positions result in minimal changes to the stability of the  $^{\text{AN}}\text{dG}$  damaged DNA B conformation (Figure 6.6).

The impact of a second adduct on the conformational stability is more diverse when  $^{\text{AN}}\text{dG}$  adopts the *syn* glycosidic conformation (S conformer of damaged DNA, Figure 6.6, Table E.26, Appendix E). Indeed,  $\Delta(\text{LSIE}^{\text{avg}})$  ranges from  $-123.1$  kJ/mol (stabilizing) to  $60.6$  kJ/mol (destabilizing) and  $\Delta(\text{SASA}^{\text{avg}})$  extends from  $-16.1$  to  $26.0$  Å<sup>2</sup>. The damaged

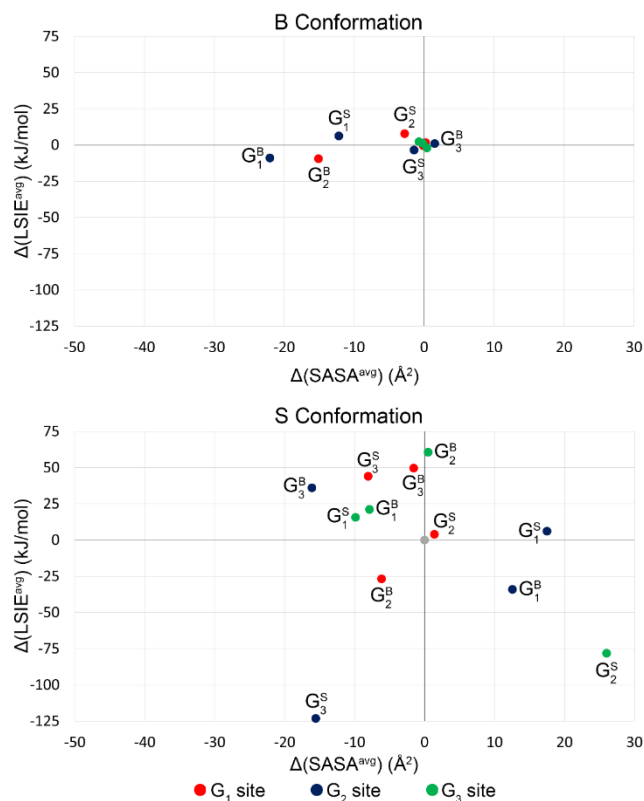


Figure 6.6:  $\Delta(\text{LSIE}^{\text{avg}})$  versus  $\Delta(\text{SASA}^{\text{avg}})$  for  $\text{ANdG}$  damaged DNA in the B (top) or S (bottom) conformation upon addition of a second adduct with the first lesion at  $\text{G}_1$  (red),  $\text{G}_2$  (blue) or  $\text{G}_3$  (green). Labels represent the site and conformation of the second lesion. Negative values indicate increased stability.

DNA S conformation is most stabilized in the next-nearest neighbor sequence, with both lesions in the *syn* orientation (i.e.,  $\text{G}_2^{\text{S}}\text{G}_3^{\text{S}}$  and  $\text{G}_2^{\text{S}}\text{G}_1^{\text{S}}$ ). Indeed, the  $\text{G}_2$  lesion site interaction increases by  $-123.1$  kJ/mol and the adduct is  $15.6 \text{ \AA}^2$  less solvent exposed, while the  $\text{G}_3$  lesion interactions increase by  $-78.2$  kJ/mol, although the adduct is solvent deshielded by  $26.0 \text{ \AA}^2$ . The damaged DNA S conformation is also stabilized in the nearest neighbor context when the second lesion adopts the B conformation (i.e.,  $\text{G}_1^{\text{S}}\text{G}_2^{\text{B}}$  and  $\text{G}_1^{\text{B}}\text{G}_2^{\text{S}}$ ), albeit to a smaller extent ( $\Delta(\text{LSIE}^{\text{avg}})$  up to  $\sim 27$  kJ/mol and  $\Delta(\text{SASA}^{\text{avg}})$  up to  $\sim 6 \text{ \AA}^2$ ). In all other cases, the addition of a second lesion has a minimal or detrimental effect on the stability of the S conformer (Figure 6.6).

Based on the relative impact of the second lesion on the stability of the B and S DNA conformations at <sup>AN</sup>dG sites, the preferred conformation of <sup>AN</sup>dG diadducted DNA at lesion sites separated by two base pairs (G<sub>1</sub>G<sub>3</sub>) can be concluded to be the same as that for the corresponding monoadducted DNA regardless of the second lesion (*anti/syn*) glycosidic orientation (Figure 6.8). This suggests that G<sub>1</sub>G<sub>3</sub> <sup>AN</sup>dG diadducted DNA will maintain an overall B conformation at both lesion positions, which reflects the spatial separation between the damaged sites. For neighboring adducts (i.e., the G<sub>1</sub>G<sub>2</sub> sequence context), the B conformation is further stabilized at both locations, while only modest changes occur to the stability of the S conformation, suggesting that the *anti* lesion conformation will continue to be preferred at both damaged sites (G<sub>1</sub><sup>B</sup>G<sub>2</sub><sup>B</sup>). Most interestingly, for the next-nearest neighbor sequence context (G<sub>2</sub>G<sub>3</sub>), a second *syn* adduct enhances the stability of the S duplex (i.e., G<sub>2</sub><sup>S</sup>G<sub>3</sub><sup>S</sup>), while the stability of the B conformation is unaffected. In contrast to monoadducted DNA, this suggests that the <sup>AN</sup>dG lesion may be uniquely stabilized in the *syn* orientation in diadducted DNA when damaged sites are one nucleotide removed.

#### 6.3.4.2 <sup>AP</sup>dG Diadducted DNA

As discussed for <sup>AN</sup>dG, there is minimal impact of a second lesion on the LSIE for *anti* <sup>AP</sup>dG in the B conformer of damaged DNA ( $\Delta(\text{LSIE}^{\text{avg}}) = 0 \pm 6$  kJ/mol, Figure 6.7, Table E.26, Appendix E). The only exception is the G<sub>1</sub><sup>B</sup>G<sub>2</sub><sup>S</sup> combination ( $\Delta(\text{LSIE}^{\text{avg}}) = -25.8$  kJ/mol). However,  $\Delta(\text{SASA}^{\text{avg}})$  spans a larger range for <sup>AP</sup>dG than <sup>AN</sup>dG, being up to  $-42.3$  Å<sup>2</sup>. As for <sup>AN</sup>dG, the largest stabilizing effects occur for the nearest neighbor G<sub>1</sub>G<sub>2</sub> sequence context, which decrease according to G<sub>1</sub><sup>B</sup>G<sub>2</sub><sup>S</sup> > G<sub>1</sub><sup>B</sup>G<sub>2</sub><sup>B</sup> > G<sub>1</sub><sup>B</sup>G<sub>2</sub><sup>B</sup> > G<sub>1</sub><sup>S</sup>G<sub>2</sub><sup>B</sup>. Furthermore, the stability of the B conformation at each damaged site is not significantly impacted when

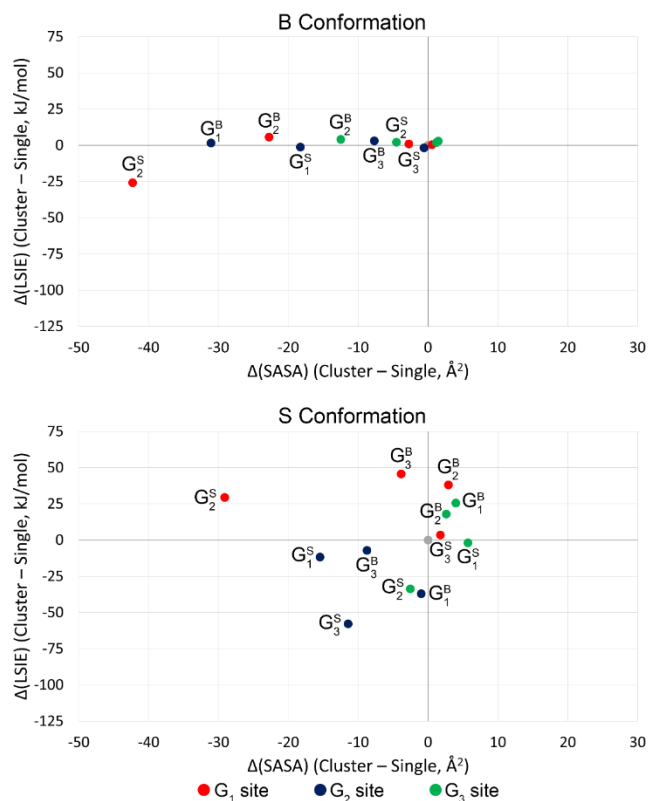


Figure 6.7:  $\Delta(\text{LSIE}^{\text{avg}})$  versus  $\Delta(\text{SASA}^{\text{avg}})$  for  $\text{APdG}$  damaged DNA in the B (top) or S (bottom) conformation upon addition of a second adduct with the first lesion at  $\text{G}_1$  (red),  $\text{G}_2$  (blue) or  $\text{G}_3$  (green). Labels represent the site and conformation of the second lesion. Negative values indicate increased stability.

the lesions are spatially separated (i.e., the  $\text{G}_1\text{G}_3$  sequence). However, unlike  $\text{ANdG}$ , two *anti*  $\text{APdG}$  lesions at the next-nearest neighbor  $\text{G}_2\text{G}_3$  position experience (albeit minor) increased solvent shielding due to the large size of the aryl ring ( $7 - 13 \text{ \AA}^2$  for  $\text{G}_2^{\text{B}}\text{G}_3^{\text{B}}$  and  $\text{G}_2^{\text{B}}\text{G}_3^{\text{S}}$ ).

When  $\text{APdG}$  adopts the *syn* orientation associated with the S conformer of damaged DNA, there is a small to moderate change in the LSIE and SASA upon addition of a second damage site. As discussed for  $\text{ANdG}$ , the most stabilizing effects occur for the next-nearest neighbor sequence with both lesions in the *syn* orientation (i.e.,  $\text{G}_2^{\text{S}}\text{G}_3^{\text{S}}$  and  $\text{G}_2^{\text{S}}\text{G}_3^{\text{B}}$ ), which

results in  $\Delta(\text{LSIE}^{\text{avg}})$  up to  $-57.8$  kJ/mol and  $\Delta(\text{SASA}^{\text{avg}})$  up to  $-11.5$  Å<sup>2</sup>. However, *syn* <sup>AP</sup>dG at G<sub>2</sub> is stabilized upon addition of a second lesion at any position (G<sub>1</sub> or G<sub>3</sub>) in any (*anti/syn*) glycosidic orientation (B/S damaged DNA conformer). In contrast, the impact of a second lesion on the stability of *syn* <sup>AP</sup>dG at G<sub>3</sub> is minimal (except for G<sub>2</sub><sup>B</sup> as discussed), while *syn* <sup>AP</sup>dG at G<sub>1</sub> is destabilized by the addition of another lesion, regardless of the second adduct location or conformation.

Based on the relative impact of a second damaging event on the conformational stability of <sup>AP</sup>dG-containing DNA, the overall conformational preference of diadducted DNA will significantly depend on the spatial separation between the damaged sites. There is no notable effect on the lesion site stability upon addition of a second lesion two nucleotides removed (i.e., G<sub>1</sub>G<sub>3</sub>), while the stability of the *syn* lesion orientation (S DNA conformer) is enhanced in the next-nearest neighbor (G<sub>2</sub>G<sub>3</sub>) sequence context. This suggests that the S conformation of damaged DNA will persist in these sequence contexts regardless of the number of adducts formed (Figure 6.8). In contrast, for the G<sub>1</sub>G<sub>2</sub> combination, the *anti* lesion orientation associated with the damaged DNA B conformer is stabilized by the presence of a second lesion, while there is minimal stabilization or even destabilization of the *syn* orientation associated with the S conformation. Therefore, the B conformation of <sup>AP</sup>dG damaged DNA is likely stabilized in the nearest neighbor G<sub>1</sub>G<sub>2</sub> sequence context, particularly for the G<sub>1</sub><sup>B</sup>G<sub>2</sub><sup>B</sup> combination. When coupled with the fact that the S conformer is the most stable at G<sub>1</sub> in <sup>AP</sup>dG monoadducted DNA, our work on diadducted DNA highlights the critical interplay between the two adducts and the stark impact of lesion spatial separation on the preferred damaged DNA conformation.



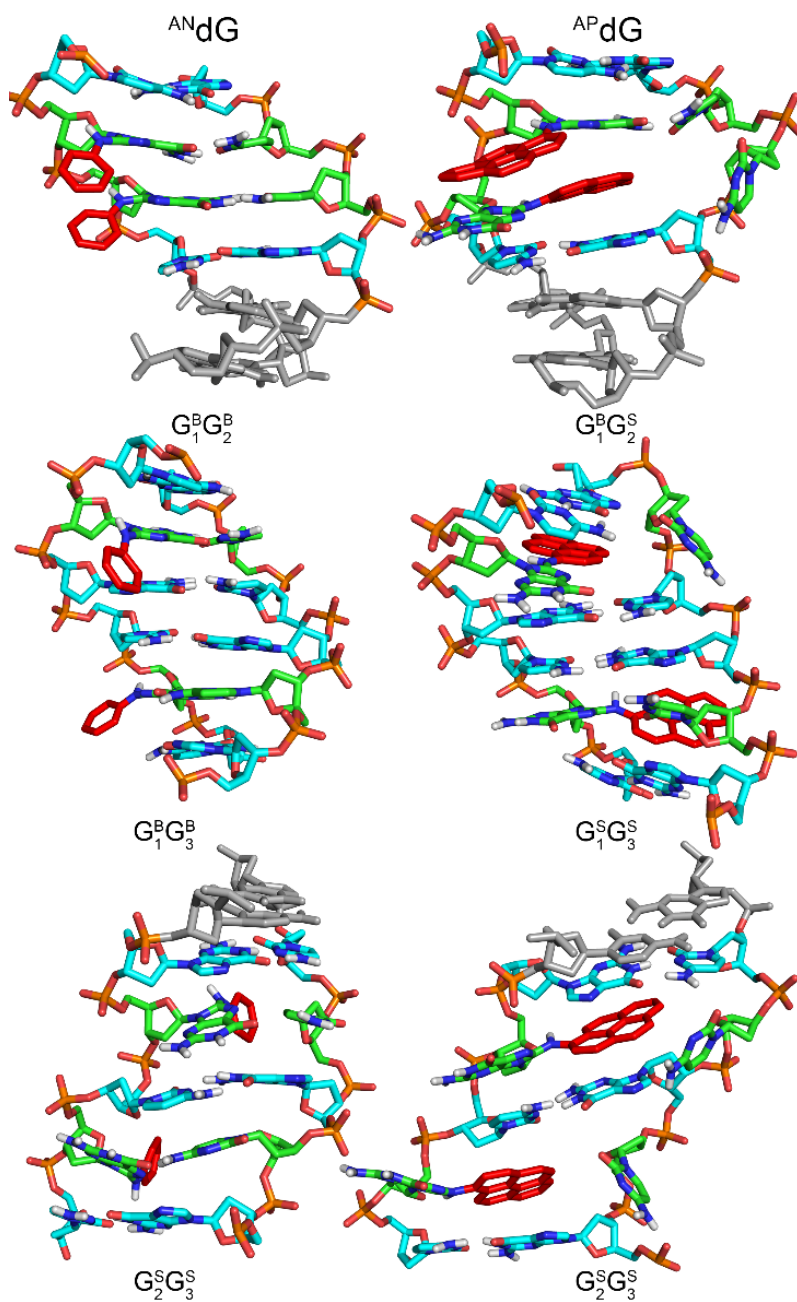


Figure 6.8: MD representative structures of the preferred <sup>AN</sup>dG and <sup>AP</sup>dG diadducted DNA conformations in different sequence contexts. Only the central base pairs from the 5' flanking base pair of G<sub>1</sub> to the 3' flanking base pair of G<sub>3</sub> in the *NarI* sequence context are shown in each case, with the adducted bases and pairing Cs shown in green, the bulky moieties in red, and the flanking bases in cyan.

## 6.4 Discussion

### 6.4.1 Despite Other N-linked C8-dG Adducts Exhibiting Sequence Dependence Conformational Heterogeneity, <sup>AN</sup>dG and <sup>AP</sup>dG Induce Monoadducted DNA Conformational Homogeneity, which is Exclusively Controlled by Aryl Moiety Size

As for other AA-derived adducts,<sup>13, 20, 24</sup> the major groove (B) conformation can be isolated for DNA containing a single <sup>AN</sup>dG or <sup>AP</sup>dG lesion at G<sub>1</sub>, G<sub>2</sub>, or G<sub>3</sub> in the *NarI* containing duplex. The B conformation of monoadducted DNA preserves the general structural features of undamaged DNA regardless of the lesion size or sequence context, with both the <sup>AN</sup>dG and <sup>AP</sup>dG lesions adopting the *anti* conformation that exposes the aryl moiety to solvent, permits WC hydrogen bonding with the opposing C, and results in little helical distortion (Figures 6.2A, E.2, Appendix E). However, for the *syn* lesion orientation, the larger <sup>AP</sup>dG adduct adopts the S conformation of damaged DNA in which the nearly planar intercalated lesion nicely stacks with the flanking base pairs (Figures 6.2B, E.3, Appendix E). In contrast, the smaller single-ringed AN moiety is unable to sufficiently stabilize the helix in an intrahelical position. Instead, <sup>AN</sup>dG shifts toward the minor groove (W-like conformation) and yields a twisted arrangement about the nucleobase–aryl moiety N-linker (Figure E.3, E.14, Appendix E). Although the larger AP aryl moiety more significantly distorts helical parameters from canonical values, especially for the S conformer, the overall structural impact is minor and sequence independent (Tables E.14, E.15, Appendix E). Due to greater lesion site interactions of the B conformer and comparable solvent exposure of the aryl moiety in both helical conformations (Figure 6.4), <sup>AN</sup>dG is predicted to preferentially stabilize the B conformation of damaged DNA (Figure 6.5). In contrast, the decreased solvent accessibility and sufficient stacking interactions of

the larger aryl moiety upon intercalation into the duplex are predicted to preferentially stabilize the S conformation of <sup>AP</sup>dG monoadducted DNA (Figure 6.4).

Overall, our finding that <sup>AN</sup>dG induces the B conformation, while <sup>AP</sup>dG induces the S conformation of damaged DNA highlights the clear influence of aryl ring size on the preferred damaged DNA conformation. This conclusion holds regardless of the sequence context, although the stability of the S conformation of <sup>AP</sup>dG monoadducted DNA decreases slightly as  $G_1 > G_2 > G_3$ , suggesting a modest sequence effect for large aryl groups. The computed conformational homogeneity for <sup>AN</sup>dG and <sup>AP</sup>dG monoadducted DNA aligns with previous suggestions from our group for the lesions at  $G_3$ ,<sup>13</sup> as well as proposals based on NMR data for damaged DNA containing <sup>AP</sup>dG in a similar sequence context as  $G_3$ <sup>21</sup> and molecular modelling for <sup>AN</sup>dG and <sup>AP</sup>dG in different sequence contexts.<sup>20, 30</sup> Thus, the present study generalizes previous conclusions regarding to the preferred conformations of <sup>AN</sup>dG and <sup>AP</sup>dG monoadducted DNA regardless of sequence context. Interestingly, the lack of sequence dependent conformational homogeneity exhibited by <sup>AN</sup>dG and <sup>AP</sup>dG adducted DNA contrasts the variation in the conformational preference with the identity of the flanking base pairs reported for other N-linked C8-dG lesions,<sup>14, 18, 24, 31</sup> as well as adducts with other linker types.<sup>52, 70-72</sup> This suggests that in the extreme limits of the smallest and largest N-linked dG AA-derived lesions, adduct size outweighs sequence context as the primary determinant of damaged DNA conformation.

#### 6.4.2 The Preferred Damaged DNA Conformation Can Change Upon Addition of a Second Adduct, with Neighboring Lesions Stabilizing the B Conformation and Next Nearest Damaged Sites Stabilizing the S Conformation Regardless of Adduct Size

With the exception of a single study on <sup>FAAF</sup>dG containing DNA,<sup>50</sup> little is known about the structure of DNA containing more than one adduct. The present work reveals that a second <sup>AN</sup>dG or <sup>AP</sup>dG adduct generally does not impact the structure of a given lesion site regardless of the adduct glycosidic (*anti/syn*) orientation, with the bulky group helical location, and lesion site hydrogen-bonding and intrahelical stacking interactions being largely consistent between corresponding monoadducted and diadducted DNA duplexes. However, diadducted DNA helices are generally more distorted than monoadducted DNA duplexes (Figures 6.5, 6.8, E.2–E.7, E.21, Appendix E), especially when both lesions adopt the *syn* orientation (S conformer) or have large aryl moieties. Although the sequence context also impacts the degree of distortion, no clear sequence trend occurs across lesion sizes or conformations. Most importantly, a second lesion can influence the preferred lesion site conformation, an effect that is heavily dependent on the distance between the lesion sites. Specifically, the preferred conformational homogeneity is preserved when the lesions in diadducted DNA are separated by two intervening base pairs (the G<sub>1</sub>G<sub>3</sub> sequence context), regardless of the lesion size. However, when two lesions occur at neighboring positions (G<sub>1</sub>G<sub>2</sub>), there is a significant increase in the stability of the (*anti*) B conformation. This effect is so pronounced that the larger <sup>AP</sup>dG adduct can stabilize the B conformation of damaged DNA at the G<sub>1</sub> site when a second lesion is located at G<sub>2</sub> despite the instability of the B conformer for monoadducted DNA. In contrast, when two lesions are next-nearest neighbours (G<sub>2</sub>G<sub>3</sub>), a sharp increase in the stability of the S conformation occurs,

particularly when both lesions adopt the *syn* orientation ( $G_2G_3$ ). In the case of AN exposed DNA, this enhanced stability means that cluster lesions can induce the S conformation that does not occur for monoadducted DNA. These conformational changes may result in differential toxicity for diadducted DNA compared to monoadducted DNA, as discussed in the next section. Regardless of whether these effects result in a complete change in the adopted conformation of damaged DNA or lead to conformational heterogeneity, the ability of a second lesion to promote a conformational shift within damaged DNA depending on adduct separation complements the previously reported conformational dependence on the sequence context for DNA containing a single AA or other N-linked C8-dG adduct<sup>23, 24, 31, 50</sup> and extends this concept to distance between neighboring adducts.

#### **6.4.3 Implications of Single and Cluster Lesions on the Mutagenicity and Repair of <sup>AN</sup>dG and <sup>AP</sup>dG-containing DNA**

The structural differences identified between <sup>AN</sup>dG and <sup>AP</sup>dG monoadducted and diadducted DNA will likely have implications for the biological outcomes associated with the replication and repair of diadducted DNA. Indeed, as discussed in the Introduction, previous literature has highlighted that the structure of damaged DNA induced by AA-derived N-linked dG adducts provides critical clues about the biological outcomes following replication by error-prone translesion synthesis (TLS) polymerases. Specifically, the stabilization of the *syn* lesion glycosidic conformation (S/W DNA conformations) is associated with mutagenic point and frameshift (deletion) mutations,<sup>23, 24, 26, 29</sup> while nonmutagenic lesions primarily adopt *anti* orientations (B DNA conformation).<sup>13, 30</sup> In addition to TLS outcomes, an important determinant of mutagenicity is lesion persistence as determined by the ability of the cell to repair the damage site. The removal of bulky

adducts from the genome is accomplished by the nucleotide excision repair (NER) pathway, which cleaves 24 – 32 nucleotides from damaged DNA strands.<sup>73</sup> The first and rate-limiting NER step in the human global genome NER (GG-NER) pathway, which is responsible for the repair of lesions across the entire genome, is the recognition of the damaged site by the XPC-Rad23B-Cen2 complex. A significant amount of work has been done to understand the GG-NER recognition step,<sup>52, 74, 75</sup> with the majority of studies focusing on yeast orthologs.<sup>76-79</sup> Overall, GG-NER efficiency has been determined to depend on several structural factors, including decreased nonbonded lesion site interactions (hydrogen bonding<sup>52, 77, 80</sup> and stacking<sup>32, 74, 76, 79</sup>), increased helical dynamics,<sup>75, 76, 78, 81</sup> and increased helical distortion<sup>81, 82</sup> compared to undamaged DNA, which impact the ability of the repair proteins to be localized in the damaged region of DNA<sup>83</sup> and form a productive bound complex.

#### *6.4.3.1 Toxicity of <sup>AN</sup>dG and <sup>AP</sup>dG in Monoadducted DNA is Most Significantly Dictated by Lesion Size Rather than Sequence Context*

In the three sequence contexts considered in the present work, <sup>AN</sup>dG monoadducted DNA is structurally similar to unmodified DNA, maintaining WC hydrogen bonding, significant stacking interactions and minimal helical distortions at the lesion site (Figure 6.5). Thus, we propose that <sup>AN</sup>dG will likely be repair resistant. Nevertheless, the small bulky group size and the predicted *anti* orientation of the damaged nucleotide suggest that replication past the damage will be error free. In contrast, <sup>AP</sup>dG exclusively stabilizes the S helical conformation in monoadducted DNA, which disrupts lesion site hydrogen bonding and induces (albeit modest) helical distortions. As a result, we hypothesize that <sup>AP</sup>dG is more likely to be recognized than <sup>AN</sup>dG during GG-NER. However, if <sup>AP</sup>dG escapes repair,

the *syn* <sup>AP</sup>dG nucleotide is expected to cause polymerase stalling and/or frameshift (deletion) mutations upon replication analogous to other AA-derived N-linked dG lesions that stabilize similar S conformations of damaged DNA.<sup>21, 22, 24, 27</sup> Since the conformations adopted by <sup>AN</sup>dG and <sup>AP</sup>dG monoadducted DNA are sequence independent, we propose that the lesion sequence context is less consequential for the cellular outcomes than adduct size. The contrasting conformational homogeneity and the hypothesized associated differential biological outcomes of the <sup>AN</sup>dG and <sup>AP</sup>dG adducts in the context of monoadducted DNA directly correlate with experimental mutagenesis data,<sup>26, 27</sup> as well as the IARC classifications of AN as nonmutagenic<sup>9</sup> and the NP precursor to <sup>AP</sup>dG as probably carcinogenic.<sup>7</sup> Additionally, our calculations further support literature emphasizing the impact of lesion size on the repairability and mutagenicity of adducted DNA.<sup>11, 28, 64, 65</sup> However, unlike other N-linked adducts for which the sequence context affects the DNA structure and biological outcomes,<sup>12, 22, 29, 42</sup> our calculations predict that <sup>AN</sup>dG and <sup>AP</sup>dG lesions induce the same conformation of damaged DNA and therefore biological outcomes irrespective of the identity of the flanking bases.

#### *6.4.3.2 Adduct Toxicity can Increase or Decrease Upon a Second Damaging Event Depending on Lesion Size and Relative Position*

In terms of cluster lesions, the recognition of one adduct will most often result in the removal of both damaged sites due to the excision of multiple nucleotides during GG-NER. Furthermore, we predict the structures of diadducted DNA to be either similar to or more distorted than the corresponding conformations of monoadducted duplexes (i.e., when comparing the same lesion at the same location in the same conformation). These factors support a new hypothesis that the repair propensities of cluster lesions will be the same or greater than those of the corresponding monoadducts in instances where the lesion

site conformation is maintained upon a second damaging event. However, our data suggests that the preferred lesion site conformations of diadducted DNA, and therefore the repair propensity and mutagenic outcomes, are dependent on lesion separation in addition to the lesion size effects noted for <sup>AN</sup>dG and <sup>AP</sup>dG monoadducted DNA.

For two intervening base pairs (the G<sub>1</sub>G<sub>3</sub> sequence context), diadducted DNA adopts the same helical conformations at both positions as the corresponding monoadducted DNA duplexes, supporting similar replication outcomes upon the second damaging event. In contrast, when two lesions are positioned with one intervening nucleotide (G<sub>2</sub>G<sub>3</sub>), the S (*syn*) conformation is preferentially stabilized at both lesion sites for both the <sup>AN</sup>dG and <sup>AP</sup>dG adducts. This is the same conformational preference for <sup>AP</sup>dG monoadducted DNA and therefore we proposed that <sup>AP</sup>dG cluster lesions will lead to similar mutagenic outcomes as a single lesion in this sequence context. However, the enhanced S conformational stability will result in a B to S conformational change for <sup>AN</sup>dG adducted DNA upon a second damaging event, which is hypothesized to increase the repair propensity for <sup>AN</sup>dG cluster lesions that are one nucleotide removed. Furthermore, if the lesions escape repair, we propose that the G<sub>2</sub>G<sub>3</sub> <sup>AN</sup>dG cluster lesions will likely be more mutagenic than the monoadducts due to the *syn* orientation preference. Since <sup>AN</sup>dG monoadducted DNA is well accepted to be nonmutagenic,<sup>13, 30</sup> our data suggests that there may be a profound increase in the mutagenicity of <sup>AN</sup>dG upon addition of a second lesion. Furthermore, the size independence of the increased stabilization of the S conformation for lesions one nucleotide removed suggests that other adducts preferentially stabilized in the *anti* orientation in monoadducted duplexes could become more mutagenetic upon the addition of a second lesion. This proposal aligns with reports that the presence of a second



damaging event enhances mutagenicity for other types of cluster lesions (e.g., SSB and 8OG).<sup>29, 37</sup>

In stark contrast to lesions one nucleotide removed, neighboring adducts ( $G_1G_2$ ) preferentially stabilize the B conformation for both  $^{AN}dG$  ( $G^B_1G^B_2$ ) and  $^{AP}dG$  ( $G^B_1G^B_2$ ) diadducted DNA. Although this leads to similar predictions regarding the repairability and mutagenicity of the smaller  $^{AN}dG$  lesion,  $^{AP}dG$  diadducted DNA is the least distorted, and therefore likely the most repair resistant, in this context. This proposed decrease in repairability is unique among all other combinations of the  $^{AN}dG$  and  $^{AP}dG$  lesions considered in the present work. Nevertheless, this hypothesis correlates with the greater repair resistance previously reported for 8OG in the presence of other (U or abasic) damaged sites.<sup>37, 40</sup> However, since neighboring  $^{AP}dG$  adducts are predicted to uniquely stabilize the nonmutagenic B conformation of damaged DNA, we propose that replication errors will occur less frequently for neighboring  $^{AP}dG$  diadducts than other spatial separations or monoadducted DNA. The increased stability of the B conformation for neighboring lesions regardless of adduct size indicates that other types of damage that induce the S conformation of monoadducted DNA may also be less mutagenic upon the formation of a second adduct. Thus, unlike  $^{AN}dG$  and other types of damage (e.g., SSB<sup>37, 38</sup> and 8OG<sup>37, 40</sup>), we hypothesize that  $^{AP}dG$  cluster lesions may generally be less mutagenic than  $^{AP}dG$  monoadducts regardless of lesion separation due to a higher repair propensity (shorter cellular lifetimes) and/or uniquely stabilized lesion site conformations.

Overall, our data highlight the complexity of the biological impact of cluster lesions, which has direct implications for the toxicity of the associated AA. Indeed, as previously reported for  $^{FAAF}dG$ ,<sup>50</sup> the repair susceptibility and mutagenic outcome can differ significantly depending on the lesion number and relative position within the duplex.

Nevertheless, <sup>AN</sup>dG, <sup>AP</sup>dG and <sup>FAAF</sup>dG<sup>50</sup> diadducts are all most repair prone when the lesions are one nucleotide removed. Perhaps most importantly, our calculations emphasize for the first time that lesion size likely plays a significant role in the impact of cluster lesions on repairability and mutagenicity. Specifically, depending on the lesion separation, the mutagenicity of the smaller <sup>AN</sup>dG adduct is proposed to be enhanced in the cluster lesion context similar to other (nonbulky) lesion types (i.e., SSB<sup>37, 38</sup> and 8OG<sup>37, 40</sup>), while the mutagenicity of the larger <sup>AP</sup>dG adduct is hypothesized to be less severe upon a second damaging event.

## 6.5 Conclusions

In summary, the present work represents the first in-depth structural analysis of diadducted DNA. Specifically, one or two AA-derived single-ringed <sup>AN</sup>dG and four-ringed <sup>AP</sup>dG adducts were placed in the *anti* and/or *syn* glycosidic orientation(s) at all combinations of three sites (G<sub>1</sub>, G<sub>2</sub> and/or G<sub>3</sub>) in oligomers containing the *NarI* recognition sequence. Unlike other N-linked C8-dG adducts,<sup>1, 14, 18, 19, 24</sup> the preferred conformation of <sup>AN</sup>dG and <sup>AP</sup>dG monoadducted DNA is exclusively determined by the adduct size, with the larger <sup>AP</sup>dG lesion inducing the mutagenic S conformation and the smaller <sup>AN</sup>dG adduct resulting in the nonmutagenic B conformation regardless of sequence context. Despite the conformational homogeneity of monoadducted DNA, the addition of a second lesion can change the preferred conformation depending on the adduct relative positions (number of intervening nucleotides). Surprisingly, following a second damaging event, sequences with neighboring damaged bases increase the stability of the *anti* lesion orientation, systems with adducts one nucleotide removed preferentially stabilize the *syn* conformation, and lesions separated by two or more base pairs adopt helical structures similar to

monoadducted DNA in the same sequence context. Notably, these general conclusions mean that <sup>AP</sup>dG is uniquely stabilized in the nonmutagenic *anti* (B) conformation in the nearest neighbor sequence context (G<sub>1</sub>G<sub>2</sub>), while <sup>AN</sup>dG is stabilized in the mutagenic *syn* (S) orientation in the next-nearest neighbor sequence (G<sub>2</sub>G<sub>3</sub>). These conformational changes suggest that <sup>AN</sup>dG could become more mutagenetic and <sup>AP</sup>dG less mutagenic upon cluster adduct formation. Therefore, the lesion size, number, and spatial separation synergistically control the damaged DNA conformation, and therefore the lesion repairability and mutagenicity. This finding complements vast literature on the structural features of AA-derived monoadducted DNA and provides new information about the impact of the number of lesions and mediating base pairs on the preferred conformation of diadducted DNA. Importantly, our calculations add the number of adducts and their spatial separation to the growing list of factors (including adduct size, shape, presence/absence of heteroatoms, linkage type and the sequence context) that determine the helical structure and therefore the biological outcomes, of damaged DNA, and emphasize the need for future research to further understand the repair and replication of diadducted DNA for a range of lesions.

## 6.6 References

1. Wilson, K. A.; Kung, R. W.; Wetmore, S. D., Chapter Seven - Toxicology of DNA Adducts Formed Upon Human Exposure to Carcinogens: Insights Gained from Molecular Modeling. In *Advances in Molecular Toxicology*, Fishbein, J. C.; Heilman, J. M., Eds. Elsevier: 2016; Vol. 10, pp 293-360.
2. International Agency for Research on Cancer, *Some Aromatic Amines, Organic Dyes, and Related Exposures*. IARC Press, International Agency for Research on Cancer: 2010; Vol. 99.
3. Wang, S.; Hanna, D.; Sugamori, K. S.; Grant, D. M., Primary Aromatic Amines and Cancer: Novel Mechanistic Insights Using 4-Aminobiphenyl as a Model Carcinogen. *Pharmacol. Ther.* **2019**, 200, 179-189.

4. International Agency for Research on Cancer, *Review of Human Carcinogens*. International Agency for Research on Cancer: 2012.
5. Sittig's Handbook of Toxic and Hazardous Chemicals and Carcinogens, 4th Edition R. P. Pohanish. Noyes/William Andrew Publishing: Norwich, New York. 2002. 2608 Pp. \$495. Isbn: 0-8155-1459-X. *Organic Process Research & Development* **2002**, 6 (5), 740-740.
6. Gooderham, N. J.; Zhu, H.; Lauber, S.; Boyce, A.; Creton, S., Molecular and Genetic Toxicology of 2-Amino-1-Methyl-6-Phenylimidazo[4,5-B]Pyridine (PhIP). *Mutat. Res.* **2002**, 506–507, 91-99.
7. The International Agency for Research on Cancer, *Diesel and Gasoline Engine Exhausts and Some Nitroarenes*. IARC: 2015.
8. Laumbach, R.; Tong, J.; Zhang, L.; Ohman-Strickland, P.; Stern, A.; Fiedler, N.; Kipen, H.; Kelly-McNeil, K.; Liroy, P.; Zhang, J., Quantification of 1-Aminopyrene in Human Urine after a Controlled Exposure to Diesel Exhaust. *J. Environ. Monit.* **2009**, 11 (1), 153-159.
9. Kahl, T.; Schröder, K.-W.; Lawrence, F. R.; Marshall, W. J.; Höke, H.; Jäckh, R., Aniline. *Ullmann's Encyclopedia of Industrial Chemistry* **2011**.
10. The International Agency for Research on Cancer, *Overall Evaluations of Carcinogenicity. An Updating of Iarc Monographs Volumes 1 to 42*. IARC: 1987.
11. Millen, A. L.; Sharma, P.; Wetmore, S. D., C8-Linked Bulky Guanosine DNA Adducts: Experimental and Computational Insights into Adduct Conformational Preferences and Resulting Mutagenicity. *Future Med. Chem.* **2012**, 4 (15), 1981-2007.
12. Patel, D. J.; Mao, B.; Gu, Z.; Hingerty, B. E.; Gorin, A.; Basu, A. K.; Broyde, S., Nuclear Magnetic Resonance Solution Structures of Covalent Aromatic Amine–DNA Adducts and Their Mutagenic Relevance. *Chem. Res. Toxicol.* **1998**, 11 (5), 391-407.
13. Kung, R. W.; Sharma, P.; Wetmore, S. D., Effect of Size and Shape of Nitrogen-Containing Aromatics on Conformational Preferences of DNA Containing Damaged Guanine. *J. Chem. Inf. Model.* **2018**, 58 (7), 1415-1425.
14. Cai, A.; Wilson, K. A.; Patnaik, S.; Wetmore, S. D.; Cho, B. P., DNA Base Sequence Effects on Bulky Lesion-Induced Conformational Heterogeneity During DNA Replication. *Nucleic Acids Res.* **2018**, 46 (12), 6356-6370.
15. Sproviero, M.; Verwey, A. M.; Rankin, K. M.; Witham, A. A.; Soldatov, D. V.; Manderville, R. A.; Fekry, M. I.; Sturla, S. J.; Sharma, P.; Wetmore, S. D., Structural and Biochemical Impact of C8-Aryl-Guanine Adducts within the NarI Recognition DNA Sequence: Influence of Aryl Ring Size on Targeted and Semi-Targeted Mutagenicity. *Nucleic Acids Res.* **2014**, 42 (21), 13405-13421.

16. Vaidyanathan, V. G.; Cho, B. P., Sequence Effects on Translesion Synthesis of an Aminofluorene-DNA Adduct: Conformational, Thermodynamic, and Primer Extension Kinetic Studies. *Biochemistry* **2012**, *51* (9), 1983-1995.
17. Cho, B. P.; Beland, F. A.; Marques, M. M., NMR Structural Studies of a 15-mer DNA Duplex from a ras Protooncogene Modified with the Carcinogen 2-Aminofluorene: Conformational Heterogeneity. *Biochemistry* **1994**, *33* (6), 1373-84.
18. Patnaik, S.; Cho, B. P., Structures of 2-Acetylaminofluorene Modified DNA Revisited: Insight into Conformational Heterogeneity. *Chem. Res. Toxicol.* **2010**, *23* (11), 1650-1652.
19. Cho, B. P.; Zhou, L., Probing the Conformational Heterogeneity of the Acetylaminofluorene-Modified 2'-Deoxyguanosine and DNA by 19F NMR Spectroscopy. *Biochemistry* **1999**, *38* (23), 7572-7583.
20. Nolan, S. J.; Vyas, R. R.; Hingerty, B. E.; Ellis, S.; Broyde, S.; Shapiro, R.; Basu, A. K., Solution Properties and Computational Analysis of an Oligodeoxynucleotide Containing N-(Deoxyguanosin-8-yl)-1-Aminopyrene. *Carcinogenesis* **1996**, *17* (1), 133-44.
21. Mao, B.; Vyas, R. R.; Hingerty, B. E.; Broyde, S.; Basu, A. K.; Patel, D. J., Solution Conformation of the N-(Deoxyguanosin-8-yl)-1-Aminopyrene ([AP]dG) Adduct Opposite dC in a DNA Duplex. *Biochemistry* **1996**, *35* (39), 12659-12670.
22. Mu, H.; Kropachev, K.; Wang, L.; Zhang, L.; Kolbanovskiy, A.; Kolbanovskiy, M.; Geacintov, N. E.; Broyde, S., Nucleotide Excision Repair of 2-Acetylaminofluorene- and 2-Aminofluorene-(C8)-Guanine Adducts: Molecular Dynamics Simulations Elucidate How Lesion Structure and Base Sequence Context Impact Repair Efficiencies. *Nucleic Acids Res.* **2012**, *40* (19), 9675-9690.
23. Jain, V.; Hilton, B.; Patnaik, S.; Zou, Y.; Chiarelli, M. P.; Cho, B. P., Conformational and Thermodynamic Properties Modulate the Nucleotide Excision Repair of 2-Aminofluorene and 2-Acetylaminofluorene dG Adducts in the NarI Sequence. *Nucleic Acids Res.* **2012**, *40* (9), 3939-51.
24. Jain, V.; Vaidyanathan, V. G.; Patnaik, S.; Gopal, S.; Cho, B. P., Conformational Insights into the Lesion and Sequence Effects for Arylamine-Induced Translesion DNA Synthesis: 19F NMR, Surface Plasmon Resonance, and Primer Kinetic Studies. *Biochemistry* **2014**, *53* (24), 4059-4071.
25. Broyde, S.; Hingerty, B. E., Conformations of DNA Modified by the Carcinogen 4-Aminobiphenyl. *Annals of the New York Academy of Sciences* **1986**, *463* (Colloq. Biol. Sci., 2nd, 1984), 86-9.
26. Cohen, S. M.; Boobis, A. R.; Meek, M. E.; Preston, R. J.; McGregor, D. B., 4-Aminobiphenyl and DNA Reactivity: Case Study within the Context of the 2006 IPCS

Human Relevance Framework for Analysis of a Cancer Mode of Action for Humans. *Crit. Rev. Toxicol.* **2006**, 36 (10), 803-19.

27. Baan, R.; Grosse, Y.; Straif, K.; Secretan, B.; El Ghissassi, F.; Bouvard, V.; Benbrahim-Tallaa, L.; Guha, N.; Freeman, C.; Galichet, L.; Coglian, V., A Review of Human Carcinogens; Part F: Chemical Agents and Related Occupations. *Lancet Oncol.* **2009**, 10 (12), 1143-1144.
28. Malia, S. A.; Basu, A. K., Mutagenic Specificity of Reductively Activated 1-Nitropyrene in Escherichia Coli. *Biochemistry* **1995**, 34 (1), 96-104.
29. Watt, D. L.; Utzat, C. D.; Hilario, P.; Basu, A. K., Mutagenicity of the 1-Nitropyrene-DNA Adduct N-(Deoxyguanosin-8-yl)-1-Aminopyrene in Mammalian Cells. *Chem. Res. Toxicol.* **2007**, 20 (11), 1658-1664.
30. Shapiro, R.; Ellis, S.; Hingerty, B. E.; Broyde, S., Effect of Ring Size on Conformations of Aromatic Amine-DNA Adducts: The Aniline-C8 Guanine Adduct Resides in the B-DNA Major Groove. *Chem. Res. Toxicol.* **1998**, 11 (4), 335-341.
31. Jain, V.; Hilton, B.; Lin, B.; Patnaik, S.; Liang, F.; Darian, E.; Zou, Y.; MacKerell, A. D., Jr.; Cho, B. P., Unusual Sequence Effects on Nucleotide Excision Repair of Arylamine Lesions: DNA Bending/Distortion as a Primary Recognition Factor. *Nucleic Acids Res.* **2013**, 41 (2), 869-880.
32. Reeves, D. A.; Mu, H.; Kropachev, K.; Cai, Y.; Ding, S.; Kolbanovskiy, A.; Kolbanovskiy, M.; Chen, Y.; Krzeminski, J.; Amin, S.; Patel, D. J.; Broyde, S.; Geacintov, N. E., Resistance of Bulky DNA Lesions to Nucleotide Excision Repair Can Result from Extensive Aromatic Lesion-Base Stacking Interactions. *Nucleic Acids Res.* **2011**, 39 (20), 8752-8764.
33. Fuchs, R. P. P., DNA Binding Spectrum of the Carcinogen N-Acetoxy-N-2-Acetylaminofluorene Significantly Differs from the Mutation Spectrum. *J. Mol. Biol.* **1984**, 177 (1), 173-180.
34. Koffel-Schwartz, N.; Verdier, J.-M.; Bichara, M.; Freund, A.-M.; Daune, M. P.; Fuchs, R. P. P., Carcinogen-Induced Mutation Spectrum in Wild-Type, UvrA and UvrC Strains of Escherichia Coli: Strain Specificity and Mutation-Prone Sequences. *J. Mol. Biol.* **1984**, 177 (1), 33-51.
35. Roy, A.; Fuchs, R. P. P., Mutational Spectrum Induced in Saccharomyces Cerevisiae by the Carcinogen N-2-Acetylaminofluorene. *Mol. Gen. Genet.* **1994**, 245 (1), 69-77.
36. Budworth, H.; Dianova, II; Podust, V. N.; Dianov, G. L., Repair of Clustered DNA Lesions. Sequence-Specific Inhibition of Long-Patch Base Excision Repair by 8-Oxoguanine. *J. Biol. Chem.* **2002**, 277 (24), 21300-5.

37. Cunniffe, S.; O'Neill, P.; Greenberg, M. M.; Lomax, M. E., Reduced Repair Capacity of a DNA Clustered Damage Site Comprised of 8-Oxo-7,8-Dihydro-2'-Deoxyguanosine and 2-Deoxyribonolactone Results in an Increased Mutagenic Potential of These Lesions. *Mutat. Res.* **2014**, 762, 32-9.
38. Sakofsky, C. J.; Saini, N.; Klimczak, L. J.; Chan, K.; Malc, E. P.; Mieczkowski, P. A.; Burkholder, A. B.; Fargo, D.; Gordenin, D. A., Repair of Multiple Simultaneous Double-Strand Breaks Causes Bursts of Genome-Wide Clustered Hypermutation. *PLoS Biol.* **2019**, 17 (9), e3000464.
39. Karwowski, B. T., The Influence of (5'r)- and (5's)-5',8-Cyclo-2'-Deoxyadenosine on UDG and hAPE1 Activity. Tandem Lesions Are the Base Excision Repair System's Nightmare. *Cells* **2019**, 8 (11).
40. Shikazono, N.; Akamatsu, K., Mutagenic Potential of 8-Oxo-7,8-Dihydroguanine (8-oxoG) Is Influenced by Nearby Clustered Lesions. *Mutat. Res.* **2018**, 810, 6-12.
41. Georgakilas, A. G.; O'Neill, P.; Stewart, R. D., Induction and Repair of Clustered DNA Lesions: What Do We Know So Far? *Radiat. Res.* **2013**, 180 (1), 100-9.
42. Zhang, N.; Lin, C.; Huang, X.; Kolbanovskiy, A.; Hingerty, B. E.; Amin, S.; Broyde, S.; Geacintov, N. E.; Patel, D. J., Methylation of Cytosine at C5 in a CpG Sequence Context Causes a Conformational Switch of a Benzo[a]Pyrene Diol Epoxide-N 2-Guanine Adduct in DNA from a Minor Groove Alignment to Intercalation with Base Displacement. *J. Mol. Biol.* **2005**, 346 (4), 951-965.
43. Feng, Z.; Hu, W.; Rom, W. N.; Beland, F. A.; Tang, M.-s., N-Hydroxy-4-Aminobiphenyl-DNA Binding in Human p53 Gene: Sequence Preference and the Effect of C5 Cytosine Methylation. *Biochemistry* **2002**, 41 (20), 6414-6421.
44. Tang, M. S.; Zheng, J. B.; Denissenko, M. F.; Pfeifer, G. P.; Zheng, Y., Use of UvrABC Nuclease to Quantify Benzo[a]Pyrene Diol Epoxide-DNA Adduct Formation at Methylated Versus Unmethylated CpG Sites in the p53 Gene. *Carcinogenesis* **1999**, 20 (6), 1085-1089.
45. Denissenko, M. F.; Chen, J. X.; Tang, M.-s.; Pfeifer, G. P., Cytosine Methylation Determines Hot Spots of DNA Damage in the Human P53 Gene. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, 94 (8), 3893.
46. Ross, M. K.; Said, B.; Shank, R. C., DNA-Damaging Effects of Genotoxins in Mixture: Modulation of Covalent Binding to DNA. *Toxicol. Sci.* **2000**, 53 (2), 224-236.
47. Hoffmann, G. R.; Fuchs, R. P. P., DNA Sequence Analysis of Mutations Induced by N-2-Acetylamin-7-Iodofluorene in Plasmid Pbr322 in Escherichia Coli. *J. Mol. Biol.* **1990**, 213 (2), 239-246.

48. Cui, X. S.; Eriksson, L. C.; Möller, L., Formation and Persistence of DNA Adducts During and after a Long-Term Administration of 2-Nitrofluorene. *Mutat. Res.* **1999**, 442 (1), 9-18.
49. Beland, F. A.; Kadlubar, F. F., Formation and Persistence of Arylamine DNA Adducts in Vivo. *Environ. Health Perspect.* **1985**, 62, 19-30.
50. Jain, V.; Hilton, B.; Lin, B.; Jain, A.; MacKerell, A. D.; Zou, Y.; Cho, B. P., Structural and Thermodynamic Insight into Escherichia Coli UvrABC-Mediated Incision of Cluster Diacetylaminofluorene Adducts on the NarI Sequence. *Chem. Res. Toxicol.* **2013**, 26 (8), 1251-1262.
51. Broyde, S.; Wang, L.; Zhang, L.; Rechkoblit, O.; Geacintov, N. E.; Patel, D. J., DNA Adduct Structure-Function Relationships: Comparing Solution with Polymerase Structures. *Chem. Res. Toxicol.* **2008**, 21 (1), 45-52.
52. Cai, Y.; Patel, D. J.; Geacintov, N. E.; Broyde, S., Differential Nucleotide Excision Repair Susceptibility of Bulky DNA Adducts in Different Sequence Contexts: Hierarchies of Recognition Signals. *J. Mol. Biol.* **2009**, 385 (1), 30-44.
53. Burnouf, D.; Koehl, P.; Fuchs, R. P., Single Adduct Mutagenesis: Strong Effect of the Position of a Single Acetylaminofluorene Adduct within a Mutation Hot Spot. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, 86 (11), 4147-4151.
54. Fuchs, R. P. P.; Schwartz, N.; Daune, M. P., Hot Spots of Frameshift Mutations Induced by the Ultimate Carcinogen N- Acetoxy-N-2-Acetylaminofluorene. *Nature* **1981**, 294 (5842), 657-659.
55. Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; Portella, G.; Battistini, F.; Gelpí, J. L.; González, C.; Vendruscolo, M.; Laughton, C. A.; Harris, S. A.; Case, D. A.; Orozco, M., Parmbsc1: A Refined Force Field for DNA Simulations. *Nat. Methods* **2015**, 13, 55.
56. Zgarbová, M.; Jurečka, P.; Šponer, J.; Otyepka, M., A- to B-DNA Transition in Amber Force Fields and Its Coupling to Sugar Pucker. *J. Chem. Theory Comput.* **2018**, 14 (1), 319-328.
57. Galindo-Murillo, R.; Robertson, J. C.; Zgarbová, M.; Šponer, J.; Otyepka, M.; Jurečka, P.; Cheatham, T. E., Assessing the Current State of Amber Force Field Modifications for DNA. *J. Chem. Theory Comput.* **2016**, 12 (8), 4114-4127.
58. Salsbury, A. M.; Lemkul, J. A., Recent Developments in Empirical Atomistic Force Fields for Nucleic Acids and Applications to Studies of Folding and Dynamics. *Curr. Opin. Struct. Biol.* **2020**, 67, 9-17.
59. Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P., The R.E.D. Tools: Advances in Resp and Esp



Charge Derivation and Force Field Library Building. *Phys. Chem. Chem. Phys.* **2010**, *12* (28), 7821-7839.

60. Galindo-Murillo, R.; Cheatham, T. E., 3rd, Lessons Learned in Atomistic Simulation of Double-Stranded DNA: Solvation and Salt Concerns [Article V1.0]. *Living J Comput Mol Sci* **2019**, *1* (2).

61. Manning, T. W.; Al-Abdul-Wahid, M. S.; Manderville, R. A.; Josephy, P. D.; Kung, R. W.; Wetmore, S. D., Structure of an Unusual Tetracyclic Deoxyguanosine Adduct: Implications for Frameshift Mutagenicity of Ortho-Cyano Nitroanilines. *Chem. Res. Toxicol.* **2020**, *33* (2), 584-593.

62. Roe, D. R.; Cheatham, T. E., Ptraaj and Cpptraaj: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084-3095.

63. Minhas, V.; Sun, T.; Mirzoev, A.; Korolev, N.; Lyubartsev, A. P.; Nordenskiöld, L., Modeling DNA Flexibility: Comparison of Force Fields from Atomistic to Multiscale Levels. *J. Phys. Chem. B* **2020**, *124* (1), 38-49.

64. Zgarbova, M.; Otyepka, M.; Sponer, J.; Lankas, F.; Jurecka, P., Base Pair Fraying in Molecular Dynamics Simulations of DNA and Rna. *J. Chem. Theory Comput.* **2014**, *10* (8), 3177-3189.

65. Le Grand, S.; Götz, A. W.; Walker, R. C., Spfp: Speed without Compromise—a Mixed Precision Model for Gpu Accelerated Molecular Dynamics Simulations. *Comput. Phys. Commun.* **2013**, *184* (2), 374-380.

66. Gotz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C., Routine Microsecond Molecular Dynamics Simulations with Amber on Gpus. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8* (5), 1542-1555.

67. Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C., Routine Microsecond Molecular Dynamics Simulations with Amber on Gpus. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9* (9), 3878-3888.

68. Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. *Amber 2018*, University of California: San Francisco, 2018.

69. Lavery, R.; Moakher, M.; Maddocks, J. H.; Petkeviciute, D.; Zakrzewska, K., Conformational Analysis of Nucleic Acids Revisited: Curves+. *Nucleic Acids Res.* **2009**, *37* (17), 5917-5929.
70. Kathuria, P.; Sharma, P.; Manderville, R. A.; Wetmore, S. D., Molecular Dynamics Simulations of Mismatched DNA Duplexes Associated with the Major C8-Linked 2'-Deoxyguanosine Adduct of the Food Mutagen Ochratoxin A: Influence of Opposing Base, Adduct Ionization State, and Sequence on the Structure of Damaged DNA. *Chem. Res. Toxicol.* **2018**, *31* (8), 712-720.
71. Cai, Y.; Kropachev, K.; Xu, R.; Tang, Y.; Kolbanovskii, M.; Kolbanovskii, A.; Amin, S.; Patel, D. J.; Broyde, S.; Geacintov, N. E., Distant Neighbor Base Sequence Context Effects in Human Nucleotide Excision Repair of a Benzo[a]Pyrene-Derived DNA Lesion. *J. Mol. Biol.* **2010**, *399* (3), 397-409.
72. Cai, Y.; Patel, D. J.; Geacintov, N. E.; Broyde, S., Dynamics of a Benzo[a]Pyrene-Derived Guanine DNA Lesion in TGT and CGC Sequence Contexts: Enhanced Mobility in TGT Explains Conformational Heterogeneity, Flexible Bending, and Greater Susceptibility to Nucleotide Excision Repair. *J. Mol. Biol.* **2007**, *374* (2), 292-305.
73. Rubbi, C. P.; Milner, J., Analysis of Nucleotide Excision Repair by Detection of Single-Stranded DNA Transients. *Carcinogenesis* **2001**, *22* (11), 1789-1796.
74. Yang, W., Poor Base Stacking at DNA Lesions May Initiate Recognition by Many Repair Proteins. *DNA Repair* **2006**, *5* (6), 654-666.
75. Maillard, O.; Camenisch, U.; Clement, F. C.; Blagoev, K. B.; Naegeli, H., DNA Repair Triggered by Sensors of Helical Dynamics. *Trends Biochem. Sci.* **2007**, *32* (11), 494-499.
76. Mu, H.; Zhang, Y.; Geacintov, N. E.; Broyde, S., Lesion Sensing During Initial Binding by Yeast XPC/Rad4: Toward Predicting Resistance to Nucleotide Excision Repair. *Chem. Res. Toxicol.* **2018**, *31* (11), 1260-1268.
77. Min, J.-H.; Pavletich, N. P., Recognition of DNA Damage by the Rad4 Nucleotide Excision Repair Protein. *Nature* **2007**, *449* (7162), 570-575.
78. Velmurugu, Y.; Chen, X.; Slogoff Sevilla, P.; Min, J.-H.; Ansari, A., Twist-Open Mechanism of DNA Damage Recognition by the Rad4/XPC Nucleotide Excision Repair Complex. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (16), E2296.
79. Mu, H.; Geacintov, N. E.; Min, J.-H.; Zhang, Y.; Broyde, S., Nucleotide Excision Repair Lesion-Recognition Protein Rad4 Captures a Pre-Flipped Partner Base in a Benzo[a]Pyrene-Derived DNA Lesion: How Structure Impacts the Binding Pathway. *Chem. Res. Toxicol.* **2017**, *30* (6), 1344-1354.

80. Hess, M. T.; Schwitter, U.; Petretta, M.; Giese, B.; Naegeli, H., Bipartite Substrate Discrimination by Human Nucleotide Excision Repair. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, 94 (13), 6664.
81. Kropachev, K.; Ding, S.; Terzidis, M. A.; Masi, A.; Liu, Z.; Cai, Y.; Kolbanovskiy, M.; Chatgililoglu, C.; Broyde, S.; Geacintov, N. E.; Shafirovich, V., Structural Basis for the Recognition of Diastereomeric 5',8-Cyclo-2'-Deoxypurine Lesions by the Human Nucleotide Excision Repair System. *Nucleic Acids Res.* **2014**, 42 (8), 5020-5032.
82. Geacintov, N. E.; Broyde, S.; Buterin, T.; Naegeli, H.; Wu, M.; Yan, S.; Patel, D. J., Thermodynamic and Structural Factors in the Removal of Bulky DNA Adducts by the Nucleotide Excision Repair Machinery. *Biopolymers* **2002**, 65 (3), 202-210.
83. Kong, M.; Liu, L.; Chen, X.; Driscoll, K. I.; Mao, P.; Bohm, S.; Kad, N. M.; Watkins, S. C.; Bernstein, K. A.; Wyrick, J. J.; Min, J. H.; Van Houten, B., Single-Molecule Imaging Reveals That Rad4 Employs a Dynamic DNA Damage Recognition Process. *Mol. Cell* **2016**, 64 (2), 376-387.

## Chapter 7: Effects of a Second Local DNA Damage Event on the Mutagenicity of the Human Carcinogen 4-Aminobiphenyl: A Molecular Dynamics Study of Damaged DNA Structure<sup>a,b,c</sup>

### 7.1 Introduction

4-aminobiphenyl (ABP) is a potent bladder mutagen that has been classified as carcinogenic to humans (Group 1) by the International Agency for Research on Cancer (IARC).<sup>1,2</sup> Human exposure to ABP has historically occurred from rubber antioxidants and dye manufacturing.<sup>3</sup> Today, humans are more commonly exposed to ABP through benzidine and benzidine-based dyes, and cigarette smoke.<sup>3</sup> ABP is an aromatic amine (AA), which is a class of chemical agents known to be mutagenic and carcinogenic to humans.<sup>1-3</sup> In fact, more than one eighth of known or suspected human carcinogens are an AA or a related biological precursor.<sup>4</sup>

In cells, ABP undergoes bioactivation by cytochrome P450 and N-acetyltransferase to produce reactive intermediates that can form adducts (addition products) with all four DNA nucleobases. However, the C8-dG adduct (denoted <sup>ABP</sup>dG) is the major stable product (Figure 7.1A).<sup>5</sup> This C8-dG lesion is similar to those known to be formed by other AAs including aminofluorene (AF), N2-acetylaminofluorene (AAF), 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP), 2-amino-3-methylimidazo[4,5-f]quinolone (IQ), aniline (AN), and aminopyrene (AP, Figure 7.1B). Interestingly, <sup>AF</sup>dG and its acetylated

---

<sup>a</sup>The *Journal of American Chemical Society* Reference Style is used in this Chapter.

<sup>b</sup>Adapted from Kung, R.W.; Takyi, N.A.; Wetmore, S.D., Effects of a Second Local DNA Damage Event on the Mutagenicity of the Human Carcinogen 4-Aminobiphenyl: A Molecular Dynamics Study of Damaged DNA Structure, [*in preparation*].

<sup>c</sup>R.W.K. and N.A.T. performed molecular dynamics simulations and initial analysis using scripts written by R.W.K. R.W.K. interpreted the data and wrote the first draft of the manuscript.

counterpart <sup>AA</sup>FdG are structurally similar to <sup>ABP</sup>dG, differing by the presence of a methylene bridge between the aromatic rings in the bulky aryl group.<sup>6</sup> Nevertheless, AAs exhibit varying degrees of mutagenicity. For example, the nitro precursors to AF and AAF have been classified by the IARC as being possibly carcinogenic to humans (Group 2B), while AP has been classified as being probably carcinogenic to humans (Group 2A).<sup>7</sup> Alternatively, AN is considered to be not classifiable as carcinogenic to humans (Group 3).<sup>8</sup>

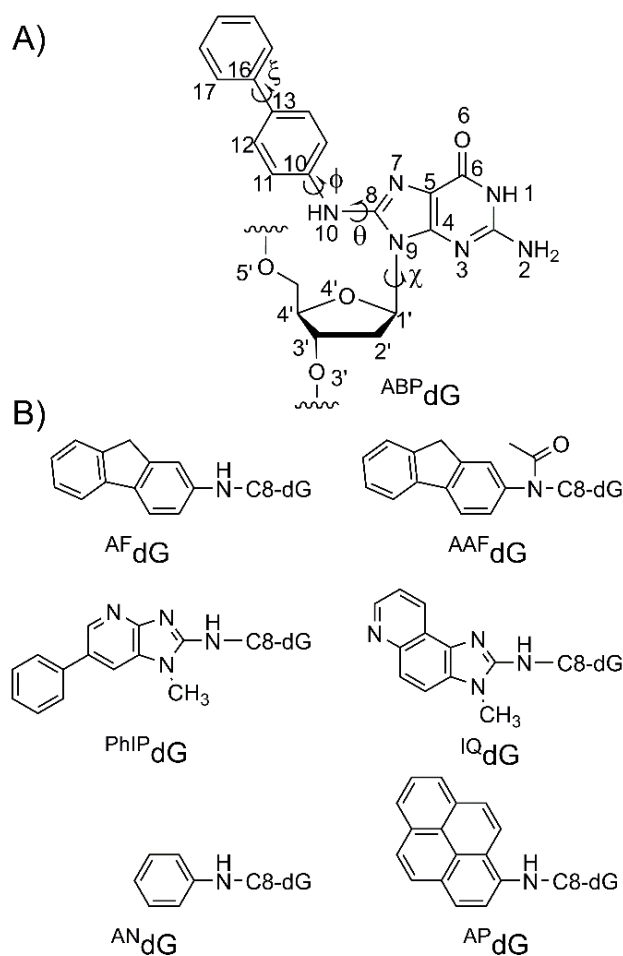


Figure 7.1: Chemical structures of A) the <sup>ABP</sup>dG nucleotide investigated in the current study, including atom numbering and identification of key dihedral angles ( $\chi$  ( $\angle(\text{O4}'\text{-C1}'\text{-N9}\text{-C4})$ ),  $\theta$  ( $\angle(\text{N9}\text{-C8}\text{-N10}\text{-C10})$ ),  $\phi$  ( $\angle(\text{C8}\text{-N10}\text{-C10}\text{-C11})$ ), and  $\xi$  ( $\angle(\text{C12}\text{-C13}\text{-C16}\text{-C17})$ ) and B) various AA-derived N-linked C8-dG adducts.

The varied mutagenic potential of AA-derived adducts is at least in part dictated by the structure of the corresponding damaged DNA. Previous NMR and computational studies have elucidated three main conformational themes of DNA containing AA-derived C8-dG adducts.<sup>9-19</sup> First, in the major groove (B) conformation, the adducted G nucleotide adopts an *anti* glycosidic orientation ( $\chi$  ( $\angle(\text{O4}'\text{--C1}'\text{--N9--C4})$ ) =  $180^\circ \pm 90^\circ$ ), which positions the bulky aryl moiety in the major groove of the DNA duplex (Figure 7.2). This conformation maintains canonical Watson-Crick (WC) hydrogen bonding between the damaged G and pairing C. In the other two conformational themes, the damaged base adopts a *syn* glycosidic orientation ( $\chi = 0^\circ \pm 90^\circ$ , Figure 7.2). In the base-displaced intercalated (stacked, S) conformation, the bulky group is situated in the interhelical space, which displaces the damaged and pairing bases into the major groove. In the wedge (W) conformation, the bulky group is located in the DNA minor groove and the Hoogsteen face of the damaged G interacts with the opposing base. The structures of damaged DNA containing <sup>ABP</sup>dG as well as the structurally similar <sup>AF</sup>dG and <sup>AAF</sup>dG lesions have been studied using fluorine-labelled nucleotide analogues (namely <sup>FABP</sup>dG, <sup>FAF</sup>dG and <sup>FAAF</sup>dG), which reveal a conformational heterogeneity (B/S/W) that is heavily dependent on the adduct and sequence context.<sup>6, 11, 15, 16, 20-25</sup> For example, <sup>ABP</sup>dG exclusively induces the B conformation of damaged DNA in three sequences (i.e., 5'-TTXAC,<sup>22</sup> 5'-TCXCA<sup>6</sup> and 5'-CTGXTC<sup>11</sup>, where X = <sup>ABP</sup>dG), while B:S conformational heterogeneity is seen in other sequence contexts (i.e., B:S ratio of 40:60 for 5'-TCXCT and 67:33 for 5'-CTXGTC).<sup>11</sup> Interestingly, the addition of a methylene bridge to <sup>ABP</sup>dG to form <sup>AF</sup>dG results in greater *syn* preference due to increased stacking from the more rigid bulky moiety.<sup>6, 26</sup> Despite some AAs inducing multiple conformations, all DNA sequences studied to date that contain a single <sup>AN</sup>dG or <sup>AP</sup>dG lesion exclusively adopt the B or S conformation, respectively.<sup>10, 17,</sup>

<sup>18, 27</sup> Based on these studies, several factors that determine the structure of damaged DNA have emerged, including the adduct size and shape, the presence of heteroatoms in the bulky moiety, and the sequence context.<sup>6, 10, 11, 28</sup>

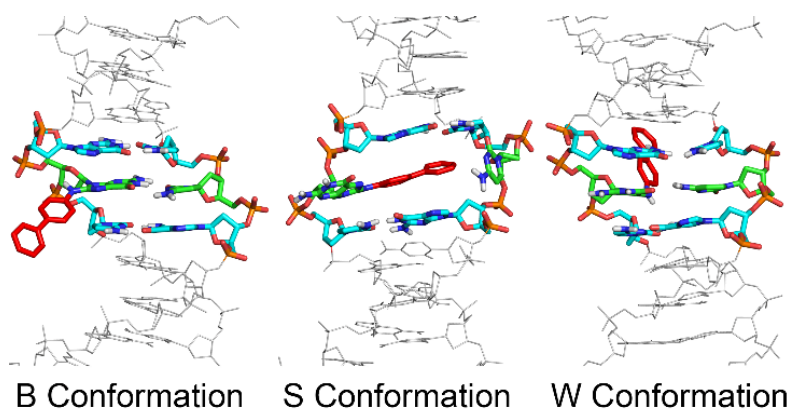


Figure 7.2: MD representative structures of <sup>ABP</sup>dG containing DNA in the B, S and W conformations as viewed from the major groove. The damaged base pair is in green, flanking base pairs in cyan, and bulky aryl group in red.

Although much work has been done to understand the structure and mutagenicity of AA-derived monoadducted DNA,<sup>6, 11, 15, 16, 20-25</sup> including the specific impact of ABP damage,<sup>29-34</sup> our knowledge of the structural preferences of DNA containing more than one adduct in close proximity is limited. However, there is a nonlinear increase in the mutation frequency (MF, number of post-replication mutations divided by the number of adducted bases) with an increase in the number of AA adducted G sites.<sup>35-37</sup> For example, increasing the number of <sup>AAF</sup>dG adducts in a section of DNA by a factor of three results in a 10-fold increase in the MF.<sup>37</sup> It is plausible that the greater MF arises from the synergistic effect of multiple adducts within a short DNA sequence. Indeed, *in vitro* work on AAF has shown that an adduct can form at different G sites within the genome, with GC rich regions being the most prone to mutations.<sup>35</sup> Furthermore, the presence of two AA-derived adducts could

help explain previously-reported obscure complex mutational profiles (e.g., combinations of frameshift and transition mutations).<sup>37, 38</sup> In fact, cluster lesions (multiple damage sites within 24 base pairs) have been observed for other types of DNA damage (i.e., 8-oxoguanine, uracil and single-strand breaks).<sup>39-49</sup> Additionally, increases in mutagenicity and decreases in repairability have been reported for such lesions upon formation of multiple damaged centers, highlighting differences in biological outcomes when several damage sites occur within close proximity.<sup>40, 41, 43, 46</sup> Notably, some AA-derived adducts are repair resistant and therefore persist in cells weeks or months after initial AA exposure.<sup>50, 51</sup> Thus, chronic exposure to ABP arising from, for example, the long term use of cigarettes and/or select dyes could lead to diadducted DNA, especially in damage hotspots (i.e., GC rich regions).

Despite the potential role of diadducts in AA-derived mutagenicity, only two studies have investigated the structure of diadducted DNA to the best of our knowledge.<sup>25, 52</sup> The first investigation used short-timescale (100 ns) molecular dynamics (MD) simulations in conjunction with various spectroscopic analysis and nucleotide excision repair (NER) assays to study the impact of <sup>FAAF</sup>dG at different combinations of two G sites in the AA-deletion *NarI* hotspot sequence (5'-G<sub>1</sub>G<sub>2</sub>CG<sub>3</sub>CC).<sup>25</sup> This work highlights the complex heterogeneous conformations of <sup>FAAF</sup>dG diadducted DNA and the corresponding repair efficiencies, which vary with the number of adducts and the proximity of the damaged bases. The second investigation used triplicate (500 ns) MD simulations to examine the structure of monoadducted and diadducted DNA containing the AA-derived single-ringed <sup>AN</sup>dG or four-ringed <sup>AP</sup>dG adduct in the *NarI* sequence context.<sup>52</sup> This work revealed that the presence of a second local lesion can stabilize helical conformations not



observed for monoadducted DNA depending on the lesion proximity, which likely results in altered replication and repair outcomes and thereby increases or decreases lesion toxicity depending on the bulky moiety size. Although <sup>ABP</sup>dG has been widely studied as an important model adduct because of its significant biological relevance,<sup>6, 11, 22</sup> <sup>ABP</sup>dG has only been considered in the context of monoadducted DNA to date.

Due to the useful insights into the structure of AA-derived damaged DNA provided by computational chemistry,<sup>10, 19, 26, 27, 53</sup> the present work uses MD simulations to investigate <sup>ABP</sup>dG diadducted DNA and thereby expand our currently limited understanding of the impact of cluster adducts on DNA structure and AA-derived biological outcomes. Specifically, 0.5  $\mu$ s MD simulations in quintuplicate are used to characterize the structure of 18 DNA helices containing one or two <sup>ABP</sup>dG adducts in different conformations and sequence contexts (Table 7.1). Monoadducted DNA containing <sup>ABP</sup>dG in the *anti* or *syn* glycosidic conformation at the G<sub>1</sub>, G<sub>2</sub>, or G<sub>3</sub> site in the *NarI* sequence is initially considered to yield insight into the conformational preferences of this lesion in sequence contexts that complement those previously studied,<sup>6, 11</sup> as well as to provide a frame of reference for understanding the impact of a second damage site. Subsequently, diadducted DNA is investigated that contains two <sup>ABP</sup>dG adducts in all combinations of (*anti/syn*) glycosidic orientations in three sequence contexts (i.e., G<sub>1</sub>G<sub>2</sub>, G<sub>1</sub>G<sub>3</sub> and G<sub>2</sub>G<sub>3</sub>). Through comparisons to previous literature on chemically similar adducts<sup>6, 23, 26, 54</sup> and <sup>ABP</sup>dG in different sequence contexts,<sup>6, 11, 22</sup> this work furthers our understanding of how the lesion chemical structure and sequence context impact the structure of monoadducted DNA. Furthermore, our work uncovers unique trends in the conformational preferences of diadducted DNA with respect to lesion separation for an adduct known to induce conformational

heterogeneity, which support conclusions previously reported from computational studies on AA-derived lesions known to induce a single conformation of damaged DNA (<sup>AN</sup>dG and <sup>AP</sup>dG).<sup>52</sup> Overall, this study helps elucidate factors that determine the mutagenic potential of ABP and provides greater insight into the complex relationship between the adduct chemical composition, number and separation, and the structure of the resulting damaged DNA, which has implications for the biological outcomes associated with AA exposure.

Table 7.1: DNA duplexes modelled in the current study.<sup>a</sup>

| System                 | Adduct                 | Label                                                         | Adduct Position and Conformation |                |                |
|------------------------|------------------------|---------------------------------------------------------------|----------------------------------|----------------|----------------|
|                        |                        |                                                               | G <sub>1</sub>                   | G <sub>2</sub> | G <sub>3</sub> |
| Undamaged Monoadducted | None <sup>ABP</sup> dG | Natural                                                       | –                                | –              | –              |
|                        |                        | G <sub>1</sub> <sup>anti</sup>                                | <i>anti</i>                      | –              | –              |
|                        |                        | G <sub>1</sub> <sup>syn</sup>                                 | <i>syn</i>                       | –              | –              |
|                        |                        | G <sub>2</sub> <sup>anti</sup>                                | –                                | <i>anti</i>    | –              |
|                        |                        | G <sub>2</sub> <sup>syn</sup>                                 | –                                | <i>syn</i>     | –              |
|                        |                        | G <sub>3</sub> <sup>anti</sup>                                | –                                | –              | <i>anti</i>    |
| Diadducted             | <sup>ABP</sup> dG      | G <sub>3</sub> <sup>syn</sup>                                 | –                                | –              | <i>syn</i>     |
|                        |                        | G <sub>1</sub> <sup>anti</sup> G <sub>2</sub> <sup>anti</sup> | <i>anti</i>                      | <i>anti</i>    | –              |
|                        |                        | G <sub>1</sub> <sup>anti</sup> G <sub>2</sub> <sup>syn</sup>  | <i>anti</i>                      | <i>syn</i>     | –              |
|                        |                        | G <sub>1</sub> <sup>syn</sup> G <sub>2</sub> <sup>anti</sup>  | <i>syn</i>                       | <i>anti</i>    | –              |
|                        |                        | G <sub>1</sub> <sup>syn</sup> G <sub>2</sub> <sup>syn</sup>   | <i>syn</i>                       | <i>syn</i>     | –              |
|                        |                        | G <sub>1</sub> <sup>anti</sup> G <sub>3</sub> <sup>anti</sup> | <i>anti</i>                      | –              | <i>anti</i>    |
|                        |                        | G <sub>1</sub> <sup>anti</sup> G <sub>3</sub> <sup>syn</sup>  | <i>anti</i>                      | –              | <i>syn</i>     |
|                        |                        | G <sub>1</sub> <sup>syn</sup> G <sub>3</sub> <sup>anti</sup>  | <i>syn</i>                       | –              | <i>anti</i>    |
|                        |                        | G <sub>1</sub> <sup>syn</sup> G <sub>3</sub> <sup>syn</sup>   | <i>syn</i>                       | –              | <i>syn</i>     |
|                        |                        | G <sub>2</sub> <sup>anti</sup> G <sub>3</sub> <sup>anti</sup> | –                                | <i>anti</i>    | <i>anti</i>    |
|                        |                        | G <sub>2</sub> <sup>anti</sup> G <sub>3</sub> <sup>syn</sup>  | –                                | <i>anti</i>    | <i>syn</i>     |
|                        |                        | G <sub>2</sub> <sup>syn</sup> G <sub>3</sub> <sup>anti</sup>  | –                                | <i>syn</i>     | <i>anti</i>    |
|                        |                        | G <sub>2</sub> <sup>syn</sup> G <sub>3</sub> <sup>syn</sup>   | –                                | <i>syn</i>     | <i>syn</i>     |

<sup>a</sup>0.5  $\mu$ s MD simulations in quintuplicate were performed, resulting in a total of 2.5  $\mu$ s production data for each duplex.

## 7.2 Computational Details

Initial models of damaged DNA were built from previously published structures of 16-mer helices containing one or two <sup>AN</sup>dG adduct(s) in the 5'–CTCTCG<sub>1</sub>G<sub>2</sub>CG<sub>3</sub>CCATCAC–3' sequence (*NarI* recognition sequence underlined).<sup>52</sup> Specifically, unique damaged duplexes were created from <sup>AN</sup>dG representative structures with the lesion(s) at the same site(s) in the *anti* or *syn* orientation by modifying the AN aryl group to ABP using GaussView 6.<sup>55</sup> Monoadducted DNA sequences were considered with (*anti* or *syn*) <sup>ABP</sup>dG at the G<sub>1</sub>, G<sub>2</sub>, or G<sub>3</sub> site. Diadducted DNA sequences contain <sup>ABP</sup>dG lesions at G<sub>1</sub> and G<sub>2</sub> (G<sub>1</sub>G<sub>2</sub>), G<sub>1</sub> and G<sub>3</sub> (G<sub>1</sub>G<sub>3</sub>), or G<sub>2</sub> and G<sub>3</sub> (G<sub>2</sub>G<sub>3</sub>). For each diadducted sequence, the adducts were considered in all four combinations of the *anti* and *syn* conformations (i.e., *anti/anti*, *anti/syn*, *syn/anti*, and *syn/syn*). The corresponding unmodified helix was also investigated as a control, which was generated using the nucleic acid builder tool in Amber18.<sup>56</sup> Thus, in total, 19 duplexes were considered (Table 7.1).

Due to the success of previous modeling and to afford accurate comparisons to published work on other AA-derived damage, the computational protocol previously used to study <sup>AN</sup>dG and <sup>AP</sup>dG cluster lesions was adopted for <sup>ABP</sup>dG.<sup>52</sup> Specifically, the parcb3c1 force field was used to describe the natural nucleotides.<sup>57</sup> The partial charges for the <sup>ABP</sup>dG nucleotide were generated using the RESP charge fitting scheme in the R.E.D. (Version III.4) software package.<sup>58</sup> Atom types were assigned using Antechamber of Amber18<sup>56</sup> and manually verified. Each duplex was solvated in a TIP3P water box using the tleap module of Amber18, such that there was a minimum of 10 Å between the solute and the edge of the water box in each direction. The systems were then neutralized with Na<sup>+</sup>.

Each system underwent an extensive minimization, heating and equilibration protocol.<sup>52, 59</sup> Specifically, the minimization was performed over four stages, with each stage consisting of 2500 steps using the steepest decent algorithm, proceeded by 2500 steps of the conjugate gradient algorithm. First, the solvent was minimized with a 100 kcal/(mol Å<sup>2</sup>) restraint on the solute. Second, the hydrogen atom positions were optimized with a 100 kcal/(mol Å<sup>2</sup>) restraint on all heavy atoms. Next, the solvent was minimized using the same constraint on the solute. Finally, minimization of the entire system was completed without restraints. Heating was then done to 310 K at constant volume over six stages, with the temperature increased from 10 K by 50 K in each stage. Each heating stage was 20 ps, and used a 1 fs timestep and a 25 kcal/(mol Å<sup>2</sup>) restraint on the solute. Equilibration was performed for 100 ps using a 2 fs timestep, with the restraint on the solute decreasing every 20 ps from 20 to 15, 10, 5, and 1.5 kcal/(mol Å<sup>2</sup>).

500 ns MD production simulations were performed in quintuplicate using different initial velocities in the PMEMD module of Amber18.<sup>60-62</sup> During the equilibration and production phases, the heavy atom distances in the WC hydrogen bonds of the terminal base pairs were restrained to less than 3 Å using a force constant of 25 kcal/(mol Å<sup>2</sup>), which prevents helical unwinding known to occur during MD simulations.<sup>63</sup> SHAKE was used for all equilibration and production simulations along with the periodic boundary condition. Furthermore, a nonbonded cutoff of 10 Å was implemented. Analysis was performed on each of the five replicas for each system, with standard deviations reported over all frames from the five replicas unless otherwise specified.

For simulations initiated with a starting *syn* lesion orientation, the trajectories were clustered by classifying each frame based on whether damaged DNA adopted the S, W or

intercalated (I) conformation (Figure 7.3 and F.1, Appendix F). The conformation was categorized as W when the distance between N4 of the pairing C and one of the Hoogsteen hydrogen-bond acceptors of the adducted G (O6 or N7) was less than 3.5 Å (Figure F.1A, Appendix F). The damaged DNA was determined to be in the S conformation when the distances between the center of mass (COM) of the ring atoms in the pairing C and the COM of the ring atoms in all four flanking nucleotides were  $> 6$  Å (Figure F.1B, Appendix F), a cutoff determined through visual inspection. All remaining structures were classified as the I conformation, which positions the bulky group in the interhelical space of DNA without fully displacing or forming hydrogen bonds with the opposing C, and has been previously reported for the C-linked para-biphenyl C8-dG adduct.<sup>54</sup> All frames of a given conformation (across all replicas) were used to obtain a representative structure using the hieragglo clustering algorithm in cpptraj based on all atoms in the duplex region spanning 5'-G<sub>1</sub>G<sub>2</sub>CG<sub>3</sub> (Figures F.2-F.13, Appendix F).

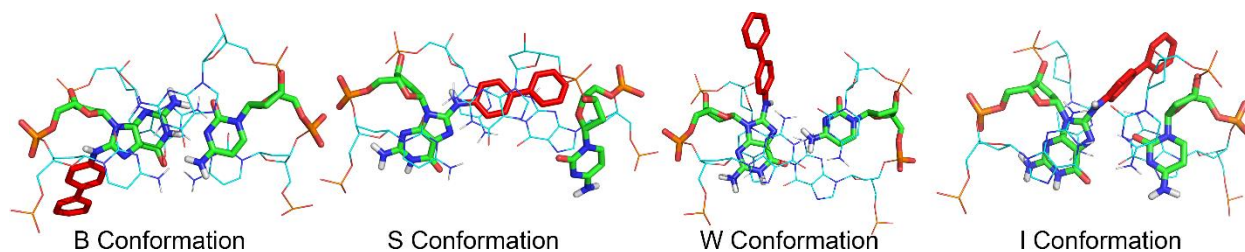


Figure 7.3: MD representative structures of <sup>ABP</sup>dG containing DNA in the B, S, W and I conformations, with the major groove on the bottom side of the figure and the minor groove on the top. The damaged base pair is in green (tube), flanking base pairs in cyan (wireframe), and bulky aryl group in red.

The cpptraj module of Amber was used to perform trajectory analysis unless otherwise specified. The backbone root-mean-square deviations (RMSDs) were recorded

with respect to the first frame of the production simulation to ensure system stability (Table F.1, Appendix F). Hydrogen-bond occupancies for the lesion site, flanking base pairs, and terminal pairs were calculated according to a heavy atom distance  $< 3.2 \text{ \AA}$  and a heavy atom–hydrogen–heavy atom angle  $> 120^\circ$  (Table F.2-F.6, Appendix F). Key dihedral angles that dictate the lesion site conformation were examined, including  $\chi$  ( $\angle(\text{O4}'\text{--C1}'\text{--N9--C4})$ ),  $\theta$  ( $\angle(\text{N9--C8--N10--C10})$ ),  $\phi$  ( $\angle(\text{C8--N10--C10--C11})$ ) and  $\xi$  ( $\angle(\text{C12--C13--C16--C17})$ , Figure 7.1). The  $\chi$  dihedral angle was also considered for the C opposing the adduct ( $\chi^{\text{opp}}$ ,  $\angle(\text{O4}'\text{--C1}'\text{--N1--C2})$ ). The C1'–C1' interstrand distance between the damaged G and pairing C was calculated and the solvent accessible surface area (SASA) of the bulky moiety was evaluated. For trajectories involving the *syn* orientation of the lesion, the structural features were independently analyzed for each (S, W, and I) damaged DNA conformation.

Helical parameters were calculated using the Curves+/Canal program. To simplify the evaluation of the base step parameters near the lesion site and permit direct comparison between systems, pseudostep parameters were used, which were defined between the base pairs 5' and 3' with respect to the *NarI* sequence (Figure F.14, Appendix F). To estimate the nonbonded interactions at the lesion sites, the lesion site interaction energy (LSIE) was evaluated as the sum of the hydrogen-bonding and stacking interactions calculated using the LIE command in cpptraj. Specifically, the hydrogen-bonding portion includes the interaction energy between the adduct and the opposing C base, while the stacking portion includes the interactions between the adduct or opposing C and the four flanking nucleobases. Interactions between the aryl moiety of the adduct and the sugar–phosphate

groups of the surrounding nucleotides were also included in the LSIE. For DNA containing a *syn* lesion, the LSIE was calculated over the unclustered trajectories.

## 7.3 Results

### 7.3.1 General Structural Features of <sup>ABP</sup>dG Monoadducted DNA

Although <sup>ABP</sup>dG damaged DNA solely adopts the B conformation when the adduct assumes an *anti* glycosidic orientation, the damaged duplex interconverts between three conformations upon incorporation of a *syn* lesion (W, S and I, Figure 7.3, Table F.7, Appendix F). The B, S and W conformations are well established in the literature regarding AA-derived damaged DNA as discussed in the Introduction (Figure 7.2),<sup>9-19</sup> while the I conformation has been previously reported for the C-linked para-biphenyl C8-dG adduct.<sup>54</sup> In the I conformation, the opposing C is stacked in the helix (unlike the S conformer), yet does not hydrogen bond with the *syn* adduct (unlike the W conformer). Rotation about the glycosidic bond can occur for the C opposing the *syn* adduct due to dynamic (intra and extrahelical) repositioning, although the N4-H hydrogen-bonding potential is not impacted. Due to the large number of duplexes considered across all damaged positions and conformations, a nomenclature is invoked that signifies the <sup>ABP</sup>dG position (G<sub>1</sub>, G<sub>2</sub> or G<sub>3</sub>) using a subscript and the adduct orientation (*anti* or *syn*) or DNA conformation (B, I, S or W) using a superscript. For example, G<sub>2</sub><sup>S</sup> (G<sub>2</sub><sup>B</sup>) denotes an adduct at the G<sub>2</sub> position with a S (B) helical conformation, whereas G<sub>2</sub><sup>syn</sup> indicates an adduct at the same position in the *syn* orientation (i.e., the combined analysis of the S, I and W duplex conformations).

### 7.3.1.1 Monoadducted DNA Containing an *anti* Lesion

Damaged DNA containing a single <sup>ABP</sup>dG lesion in the *anti* orientation adopts the B conformation previously seen for AA-derived lesions,<sup>10, 27, 53</sup> and therefore maintains the structural features of undamaged DNA regardless of sequence context (Figure 7.3, Tables F.4, F.6, F.8, F.11-F.12, Appendix F). Specifically, WC hydrogen bonding remains intact at the lesion site and there is minimal helical distortion. However, the glycosylic orientation of the damaged nucleoside slightly deviates from that of undamaged DNA, with average  $\chi$  ( $\chi^{\text{avg}}$ ) values ranging between 225° – 230° for the adduct compared to 249° – 257° for natural G (Table F.9, Appendix F). The  $\theta$  ( $\theta^{\text{avg}}$  = 129.1–142.5°) and  $\phi$  ( $\phi^{\text{avg}}$  = 0° – 6°, Table F.9, Appendix F) dihedral angles that at least in part dictate the relative nucleobase–bulky moiety orientation are consistent across the trajectories, likely due to hydrogen bonding between the N–H linker (N10–H) and O4' of the damaged nucleotide or O3' of the 5' nucleotide (Table F.10, Appendix F). Although the average  $\xi$  ( $\xi^{\text{avg}}$ ) dihedral angle of approximately 0° (Table F.9, Appendix F) suggests a planar relative orientation of the two phenyl rings in the bulky moiety, this arises due to the presence of four nearly equally distributed slightly twisted conformations ( $\xi$  = 30, 150, 210 and 330°, Figure F.15, Appendix F), as reflected in the average interplanar angle between the two ABP rings (~33°, Table F.8, Appendix F). Overall, the helical parameters for damaged DNA the B conformation do not deviate from those of canonical undamaged DNA.

### 7.3.1.2 Monoadducted DNA Containing a *syn* Lesion

As discussed previously, damaged DNA containing a *syn* adduct undergoes dynamical interconversion between three distinct conformations (S, W and I, Figure 7.3, Table F.7, Appendix F). The I conformer is consistently the most populated (55-98%),



followed by S (2-31%) and W (0-15%), with the exact conformational distribution dependent on the sequence context. Notably, the G<sub>2</sub> and G<sub>3</sub> sites (flanked by G and C) stabilize a S conformation significantly more often ( $30 \pm 1\%$ ) than the G<sub>1</sub> site (2%, flanked by two Cs). Regardless, the base pairs flanking the lesion site remain intact (occupancies > 82%, Table F.4, Appendix F). The structural features at the lesion site are generally consistent across sequence context for the most populated I and S conformations. For example,  $\chi^{\text{avg}} = 45^\circ - 62^\circ$ ,  $\phi^{\text{avg}} = 357^\circ - 4^\circ$  (Table F.9, Appendix F) and average interplanar angle between rings of the ABP moiety =  $25^\circ - 31^\circ$ , with the exception of only G<sub>2</sub> for which  $\chi^{\text{avg}}$  ( $164.4^\circ$ ) differs due to a change in the sugar orientation to permit hydrogen bonding between the linking N–H and O4' in the damaged nucleoside (84.2% occupancy, Figure F.16, Table F.10, Appendix F) as previously reported for <sup>AP</sup>dG in the same sequence context.<sup>52</sup> Nevertheless,  $\theta^{\text{avg}}$  changes with helical conformation ( $\theta^{\text{avg}} \approx 250^\circ$  for I and  $190^\circ$  for S). The interstrand (C1'–C1') distance is greater than canonical DNA ( $10.7 - 10.8 \text{ \AA}$ ) for the S conformation ( $12.6 - 15.3 \text{ \AA}$ , Table F.8, Appendix F), but smaller for the I conformer ( $8.3 - 8.7 \text{ \AA}$ ), which permits the ABP ring farthest from the damaged G to stack with the minor groove face of the sugar in the opposing nucleotide (Figure 7.3).

Although  $\chi$  and  $\phi$  in the W conformation align with the I and S conformers,  $\theta$  differs according to sequence context, which orients the lesion in the 3' direction of the damaged strand for G<sub>1</sub><sup>W</sup> ( $226.6^\circ$ ) and G<sub>3</sub><sup>W</sup> ( $258.9^\circ$ ) and the 5' direction for G<sub>2</sub><sup>W</sup> ( $79.5^\circ$ ). This reflects differences in neighboring hydrogen-bond acceptors as the adduct linker N–H hydrogen bond with the interstrand flanking G in G<sub>1</sub><sup>W</sup> and G<sub>3</sub><sup>W</sup> (occupancy > 84%), but hydrogen bonds with O4' of the 3' sugar in G<sub>2</sub><sup>W</sup> due to the presence of a 5' interstrand C (occupancy of 42%, Table F.10, Appendix F). This change in lesion orientation and local hydrogen bonding

impacts the interstrand distance. Specifically,  $G^W$  has a larger C1'–C1' distance (11.3 Å) compared to the more similar  $G^Y$  (7.6 Å) and  $G^Z$  (9.0 Å) damaged helices (Table F.8, Appendix F). Thus, unlike the I and S conformations, the W conformation shows an interesting sequence-dependent lesion site conformation. Nevertheless, due to the low occupancy of the W conformation (< 16%), the *syn* orientation results in minimal interactions between the adduct and pairing base. Regardless of the damaged DNA conformation, the helix is moderately distorted, with the slide, tilt and twist generally changing by less than 5 Å and 15°, although larger increases of up to 5.9 Å and 41° occur for damage at G<sub>1</sub> (Tables F.11-12, Appendix F).

### 7.3.2 *anti/syn* Preferences of <sup>ABP</sup>dG in Monoadducted DNA

Due to the dynamical interconversion between several helical conformations for damaged DNA containing *syn* <sup>ABP</sup>dG, the conformational features of damaged DNA are best assessed based on the preferred (*anti/syn*) lesion glycosidic orientation. <sup>ABP</sup>dG monoadducted DNA has consistent global structural features regardless of the lesion glycosidic orientation (Tables F.4, F.8, F.9, F.11, F.12, Appendix F), suggesting that structural changes away from the adducted site will not significantly influence the relative conformer stability. As a result, to better understand the impact of the lesion site for a given monoadducted DNA sequence, an approach analogous to that successfully used to understand the preferred conformations of DNA containing N-linked C8-dG adducts of varying bulky moiety size is implemented.<sup>10, 52</sup> Specifically, the preferred lesion conformation is assessed based on the stabilizing noncovalent interactions at the lesion site (LSIE<sup>avg</sup>) and the destabilizing solvation of the hydrophobic bulky group (SASA<sup>avg</sup>). While these factors are believed to be the largest and therefore the most important differences between damaged DNA

conformations, we acknowledge that other factors such as discrete solute–solvent interactions may also change with damaged DNA conformation.

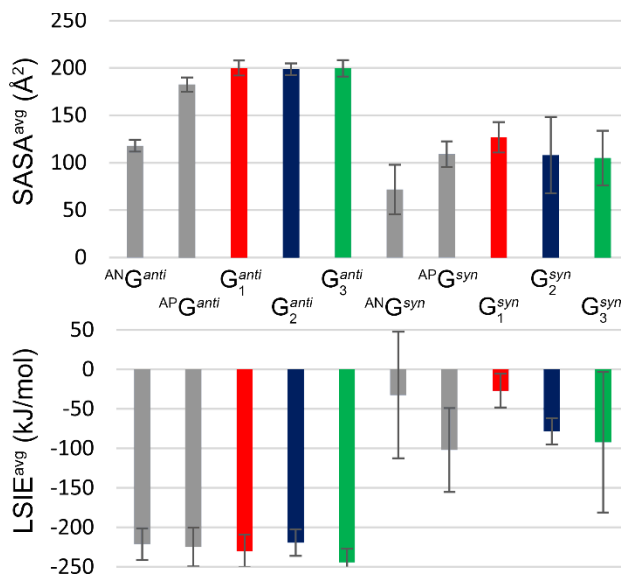


Figure 7.4: LSIE<sup>avg</sup> and SASA<sup>avg</sup> for DNA containing a single <sup>AN</sup>dG or <sup>AP</sup>dG averaged over all *NarI* three sites (data from <sup>52</sup>) and <sup>ABP</sup>dG lesion at G<sub>1</sub> (red), G<sub>2</sub> (blue) or G<sub>3</sub> (green) in the *anti* or *syn* orientations, including the dataset standard deviation.

Although the average SASA of the aryl moiety in *anti* <sup>ABP</sup>dG in the B conformer of damaged DNA is highly similar across all sequence contexts ( $199 \pm 1 \text{ Å}^2$ , Figure 7.4, Table F.13, Appendix F), there are minor variations in the LSIE<sup>avg</sup> due to changes in the stacking potential with the identity of the neighboring bases (Figure 7.4, Table F.14, Appendix F). Specifically, the LSIE<sup>avg</sup> is most stabilizing at the G<sub>3</sub> position ( $-244.3 \text{ kJ/mol}$ ), being  $\sim 15 \text{ kJ/mol}$  greater than that at G<sub>1</sub> ( $-229.7 \text{ kJ/mol}$ ) and  $\sim 25 \text{ kJ/mol}$  greater than that at G<sub>2</sub> ( $-219.3 \text{ kJ/mol}$ ). Thus, there is a modest dependence of the lesion site stability on the sequence context for the B conformer of damaged DNA, which decreases as  $G_3 > G_1 > G_2$ . The impact of sequence context on the noncovalent lesion site interactions is even greater

for the *syn* lesion orientation, with the  $LSIE^{avg}$  decreasing as  $G_3$  (−92.2 kJ/mol) >  $G_2$  (−78.2 kJ/mol) >  $G_1$  (−26.9 kJ/mol, Figure 7.4, Table F.14, Appendix F). While bulky group interactions with the DNA backbone are a major stabilizing contribution to the  $LSIE^{avg}$  at all three sites (> 40 kJ/mol), a large portion of the sequence effect can be attributed to unfavourable electrostatic interactions between hydrogen-bond acceptors of the adduct (O6 and N7) and the pairing C (O2 and N3, Figures F.5-F.13, Table F.14, Appendix F), with differences in the interaction strengths occurring due to changes in the relative orientation of the adduct and pairing C the identity of the flanking nucleobases. Further enhancing the sequence effects, the bulky group of *syn*  $^{ABP}dG$  is most solvent exposed at  $G_1$  (126.8 Å<sup>2</sup>, Figure 7.4, Table F.13, Appendix F), followed by  $G_2$  (108.1 Å<sup>2</sup>) and  $G_3$  (104.9 Å<sup>2</sup>). The greater solvent shielding of the biphenyl moiety at  $G_2$  and  $G_3$  can be attributed to the increased intercalation associated with the greater presence of the S conformation at these sites (30 ± 1%) compared to  $G_1$  (2%, Table F.7, Appendix F).

Overall, comparison with the properties predicted for other AA-derived N-linked C8-dG lesions ( $^{AN}dG$  and  $^{AP}dG$ )<sup>10, 52</sup> suggests that  $^{ABP}dG$  will adopt both the *anti* and *syn* glycosidic orientations in monoadducted DNA. Previous computational work has rationalized the *anti* conformational homogeneity for  $^{AN}dG$  regardless of sequence context based on the much larger  $LSIE^{avg}$  for the B compared to the S conformer of damaged DNA coupled with a small decrease in the bulky moiety  $SASA^{avg}$  for *anti* to *syn* conversion (Figure 7.4). In contrast, the S conformer of  $^{AP}dG$  containing DNA dominates at all G positions in the *NarI* sequence due to a significant decrease in the aryl group  $SASA^{avg}$  upon *anti* to *syn* conversion and the strong stacking of the lesion in the S conformer (Figure 7.4). Although  $^{ABP}dG$  results in a similar bulky moiety  $SASA^{avg}$  as  $^{AP}dG$  regardless of the adduct

glycosidic orientation and similar  $\text{LISE}^{\text{avg}}$  for the *anti* lesion conformation, differences arise in the  $\text{LSIE}^{\text{avg}}$  for the *syn* adducts, which opens the door for  $^{\text{ABP}}\text{dG}$  lesion site conformational heterogeneity. However, the exact *anti:syn* ratio depends on the sequence context. Specifically, the  $\text{LSIE}^{\text{avg}}$  for *syn*  $^{\text{AP}}\text{dG}$  at  $\text{G}_1$  is more similar to  $^{\text{AN}}\text{dG}$ , suggesting a stronger *anti* preference. Nevertheless, the greater solvent shielding upon *anti* to *syn* conversion for  $^{\text{ABP}}\text{dG}$  compared to  $^{\text{AN}}\text{dG}$  rationalizes the presence of some *syn* orientation for  $^{\text{ABP}}\text{dG}$ . In contrast, the  $\text{LSIE}^{\text{avg}}$  is much larger for  $^{\text{ABP}}\text{dG}$  at the  $\text{G}_2$  and  $\text{G}_3$  sites than  $\text{G}_1$ , which is coupled to a larger decrease in the  $\text{SASA}^{\text{avg}}$  upon *anti* to *syn* conformational change at  $\text{G}_2$  and  $\text{G}_3$  than  $\text{G}_1$ , suggesting dominance of the *syn* conformation (Figure 7.5)

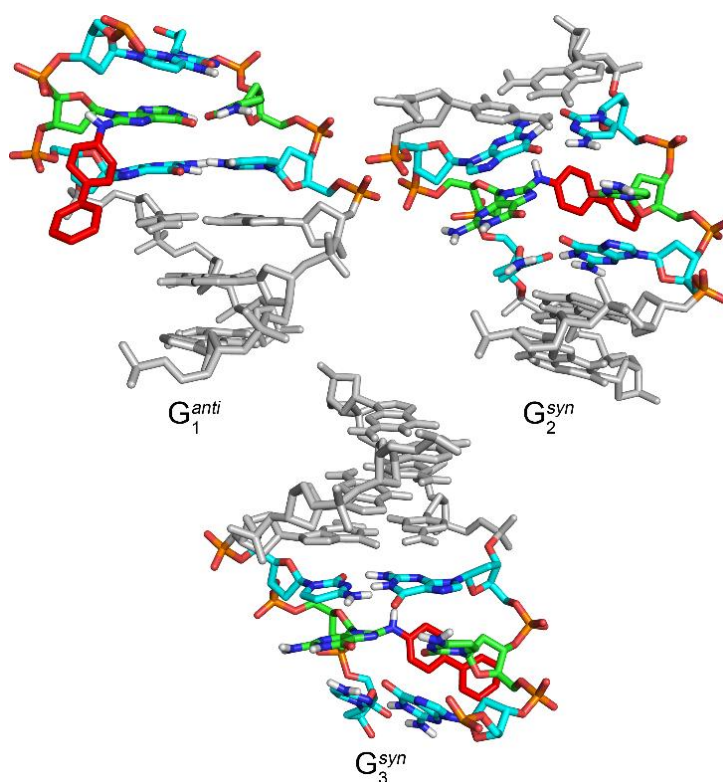


Figure 7.5: MD representative structures of the preferred conformations of  $^{\text{ABP}}\text{dG}$  monoadducted DNA for different sequence contexts. Only the base pairs in the 5'-CG<sub>1</sub>G<sub>2</sub>CG<sub>3</sub>C sequence context are shown in each case, with the adducted base and pairing C shown in green, the bulky moiety in red, and the flanking bases in cyan.

as predicted for <sup>AP</sup>dG. Interestingly, the similar conformational heterogeneity predicted at G<sub>2</sub> and G<sub>3</sub> correlates with both positions having a 3'-C with respect to the damaged site. The stabilization of <sup>ABP</sup>dG in both glycosidic conformations seen for <sup>AN</sup>dG and <sup>AP</sup>dG indicates a “middle-of-the-road” behaviour for the biphenyl-containing adduct compared to the one-ringed and four-ringed lesions.

### 7.3.3 General Structural Features of <sup>ABP</sup>dG Diadducted DNA

To understand the impact of a second damaging event on DNA structure, two adducts were modelled within the *NarI* sequence for all combinations of *anti/syn* glycosidic orientations as neighboring nucleotides (G<sub>1</sub>G<sub>2</sub>), next-nearest neighbors (G<sub>2</sub>G<sub>3</sub>) or nucleotides separated by two intervening nucleotides (G<sub>1</sub>G<sub>3</sub>). In the diadducted DNA nomenclature, damage sites are listed numerically, with the position denoted as a subscript and the conformation as a superscript (e.g., G<sub>1</sub><sup>*anti*</sup>G<sub>2</sub><sup>*syn*</sup> represents the duplex with an *anti* adduct at G<sub>1</sub> and a *syn* lesion at G<sub>2</sub>). When discussing the impact of a second damaging event, the site under consideration is indicated using bold font (e.g., **G**<sub>1</sub><sup>*anti*</sup>G<sub>2</sub><sup>*syn*</sup> represents studying the impact of the *syn* lesion at G<sub>2</sub> on the *anti* adduct at G<sub>1</sub>). For cluster lesions, Δ signifies the effect of a second lesion on a given parameter (i.e., the average value of a dihedral angle, LSIE or SASA for diadducted DNA minus that for the corresponding monoadducted system). For example, Δχ (−1.6°) for **G**<sub>1</sub><sup>*P*</sup>G<sub>2</sub><sup>*syn*</sup> is the lesion site χ in G<sub>1</sub><sup>*P*</sup> monoadducted DNA (226.3°) subtracted from the corresponding χ in diadducted DNA with a second G<sub>2</sub><sup>*syn*</sup> adduct (224.7°).

### 7.3.3.1 Impact of a Second Lesion on DNA Containing an *anti* Adduct

The addition of a second lesion generally has a minimal impact on the local structure of DNA containing *anti*<sup>ABP</sup>dG (Tables F.5, F.6, F.15-F.22, Appendix F). The first exception is the **G<sub>1</sub><sup>anti</sup>G<sub>2</sub><sup>syn</sup>** system where the addition of a neighboring *syn* lesion decreases the WC base pairing of the adducted G<sub>1</sub> site (occupancy 77.2%, Table F.6, Appendix F) due to displacement of the pairing C toward the minor groove to permit stacking with the aryl group of G<sub>2</sub><sup>syn</sup> (Figure F.17, Appendix F). The second exception is a shift in  $\theta$  for the neighboring **G<sub>1</sub><sup>anti</sup>G<sub>2</sub><sup>syn</sup>** and **G<sub>1</sub><sup>syn</sup>G<sub>2</sub><sup>anti</sup>** systems ( $\Delta\theta = 28 \pm 1^\circ$ , Table F.18, Appendix F) due to favourable stacking interactions between the major groove bulky group and the neighboring displaced *syn* G. Overall, helices containing two *anti* adducts are generally undistorted, while those containing one *syn* lesion are moderately distorted (Tables F.11-F.12, Appendix F). In fact, the helical distortion for diadducted DNA containing one *anti* and one *syn* lesion is comparable to monoadducted DNA containing a *syn* adduct. This suggests the level of helical distortion in diadducted DNA is more significantly dictated by the *anti/syn* orientation of the second lesion than the sequence context or lesion separation.

### 7.3.3.2 Impact of a Second Lesion on DNA Containing a *syn* Adduct

Although all three helical conformations associated with a *syn* lesion (S, I and W) occur for most diadducted duplexes, the conformational heterogeneity of monoadducted DNA changes depending on the location and orientation of the second lesion (Table F.7, Appendix F). Although the W conformation is minimally populated, the addition of a second <sup>ABP</sup>dG decreases the prevalence of the I conformer and increases the occupancy of the S conformer. Exceptions arise for a next-nearest neighbor *anti* adduct, which almost

exclusively results in the I conformation for  $G_2^{syn}G_3^{anti}$  diadducted DNA and leads to a nearly equal distribution of all three conformations for  $G_2^{anti}G_3^{syn}$ .

Irrespective of the specific helical conformation associated with  $syn^{ABP}dG$ , the addition of a second lesion generally does not disrupt the WC hydrogen bonding of the base pairs flanking a *syn* adduct (occupancy > 85%, Table F.5, Appendix F). The only exception is neighboring ( $G_1^{syn}G_2^{syn}$ ) *syn* lesions for which disruption of the 3' flanking base pair increases stacking interactions with the displaced  $G_2$  (Figure F.18, Appendix F). Moreover, the key dihedral angles at the lesion site ( $\chi$ ,  $\theta$ ,  $\phi$  and  $\xi$ ) are generally independent of the presence of a second adduct and therefore the bulky moiety is consistently positioned at corresponding positions in diadducted and monoadducted DNA (Tables F.18-F.20, F.24-F.26, Appendix F). A notable deviation in  $\chi^{avg}$  is seen when an intra-nucleotide hydrogen bond forms that is coupled with backbone distortions in diadducted DNA ( $\chi^{avg}$  values > 90°; Tables F.24-F.26, Figure F.16, Appendix F).

As with monoadducted DNA, a *syn* lesion results in greater structural distortion of diadducted DNA compared to the *anti* orientation. For example, the range in the interstrand C1'–C1' distance in diadducted DNA containing a *syn* lesion (7.7 Å to 15.3 Å) is significantly larger than an *anti* lesion (10.7 Å to 10.8 Å, Table F.15). Nevertheless, the helical distortion at the damage site is generally comparable to monoadducted DNA with a *syn* adduct, with most parameters changing by less than 5 Å or 35° from undamaged DNA, although some deviate by up to 6 Å or 53° (Tables F.11-F.12, Appendix F). Overall, the dominant structural features of monoadducted containing a  $syn^{ABP}dG$  adduct are maintained upon inclusion of a second lesion, with slight adjustments occurring at the lesion site for some systems.



### 7.3.4 *anti/syn* Preferences of <sup>ABP</sup>dG in Diadducted DNA

To gain an understanding of the impact of a second lesion on the preferred <sup>ABP</sup>dG conformations in DNA duplexes, the changes in LSIE<sup>avg</sup> ( $\Delta(\text{LSIE}^{\text{avg}})$ ) and SASA<sup>avg</sup> ( $\Delta(\text{SASA}^{\text{avg}})$ ) upon addition of a second lesion were evaluated, with a negative  $\Delta(\text{LSIE}^{\text{avg}})$  or  $\Delta(\text{SASA}^{\text{avg}})$  reflecting increased stability of a given lesion site upon addition of the second adduct. Most notably, regardless of lesion separation or sequence context, the conformational heterogeneity of <sup>ABP</sup>dG monoadducted DNA likely persists in diadducted DNA. Nevertheless, the exact conformational mixture is dependent on the lesion separation, with certain distances between adducts preferentially stabilizing unique glycosidic orientations compared to monoadducted DNA.

When a lesion is present at G<sub>1</sub> in either glycosidic orientation, the addition of a second adduct at a neighboring (G<sub>2</sub>) position increases the stabilizing lesion site interactions (by at least –10 kJ/mol) and the bulky moiety solvent shielding (by at least –25 Å<sup>2</sup>) regardless of the second lesion orientation, with the effect being slightly larger for *anti* G<sub>2</sub> (Figures 7.6, F.19, Table F.27, Appendix F). Although the impact of a G<sub>2</sub> lesion is significant for G<sub>1</sub><sup>syn</sup>, the change is likely not enough to overcome the inherent instability of the *syn* adduct orientation at G<sub>1</sub> (Figure 7.4), suggesting that the *anti* conformation will remain preferred at G<sub>1</sub> in G<sub>1</sub>G<sub>2</sub> diadducted DNA. Indeed, while the addition of an adduct with either glycosidic conformation at G<sub>1</sub> stabilizes the G<sub>2</sub> lesion site, an *anti* G<sub>1</sub> lesion results in significantly greater lesion site interactions and solvent shielding than a *syn* lesion (differences in the  $\Delta(\text{LSIE}^{\text{avg}})$  and  $\Delta(\text{SASA}^{\text{avg}})$  of up to 17.9 kJ/mol and 26.3 Å<sup>2</sup>, respectively), regardless of whether <sup>ABP</sup>dG is *anti* or *syn* at G<sub>2</sub> (Figures 7.6, F.19 Table

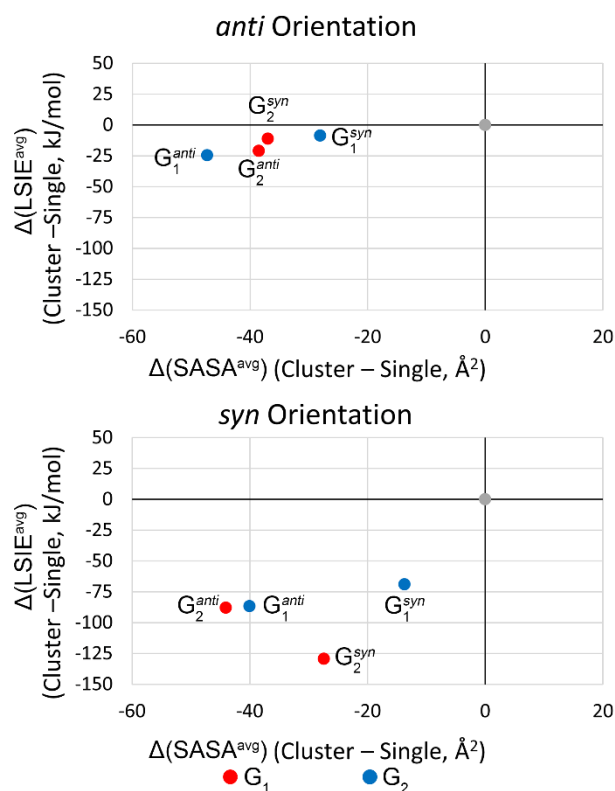


Figure 7.6:  $\Delta(\text{LSIE}^{\text{avg}})$  versus  $\Delta(\text{SASA}^{\text{avg}})$  for <sup>ABP</sup>dG DNA upon addition of a second adduct with the first lesion in the *anti* (top) or *syn* (bottom) conformation for neighboring adducts (the G<sub>1</sub>G<sub>2</sub> sequence context), original G<sub>1</sub> site in red and G<sub>2</sub> site in blue, with labels representing the site and conformation of the second lesion. Negative values indicate increased stability.

F.27). Therefore, when coupled with the significant solvent shielding of the lesions and large lesion site noncovalent interactions in G<sub>1</sub><sup>*anti*</sup>G<sub>2</sub><sup>*anti*</sup> diadducted DNA, this suggests that G<sub>1</sub><sup>*anti*</sup>G<sub>2</sub><sup>*anti*</sup> is likely the dominant duplex conformation, followed by G<sub>1</sub><sup>*anti*</sup>G<sub>2</sub><sup>*syn*</sup>. Overall, although the enhanced LSIE and decreased SASA for neighboring lesions regardless of their conformation suggests that G<sub>1</sub>G<sub>2</sub> <sup>ABP</sup>dG diadducted DNA will maintain the conformational heterogeneity of the corresponding monoadducted DNA duplexes, the presence of neighboring lesions tends to increase the prevalence of the *anti* orientation at G<sub>2</sub> compared to monoadducted DNA, while not impact the G<sub>1</sub><sup>*anti*</sup> preference.

Next-nearest neighbor adducts ( $G_2G_3$ ) only slightly enhance the stability of an *anti* lesion site, with changes in the bulky moiety solvent accessible surface area and lesion site interactions being less than  $7 \text{ \AA}^2$  and  $15 \text{ kJ/mol}$  compared to monoadducted DNA, respectively (Figures 7.7, F.19 (top), Table F.27, Appendix F). In contrast, the stability of the *syn* orientation can be significantly enhanced (Figures 7.7, F.19, Table F.27, Appendix F), especially when the second damage site is also *syn* ( $\Delta(\text{SASA}^{\text{avg}})$  up to  $-50 \text{ \AA}^2$  and  $\Delta(\text{LSIE}^{\text{avg}})$  up to  $-123 \text{ kJ/mol}$ ). When coupled with the inherent *syn* preferences for both the  $G_2$  and  $G_3$  sites (Figure 7.4), this implies that next-nearest adducted nucleotides will both adopt a *syn* orientation for the vast majority of the time. Interestingly, although next-

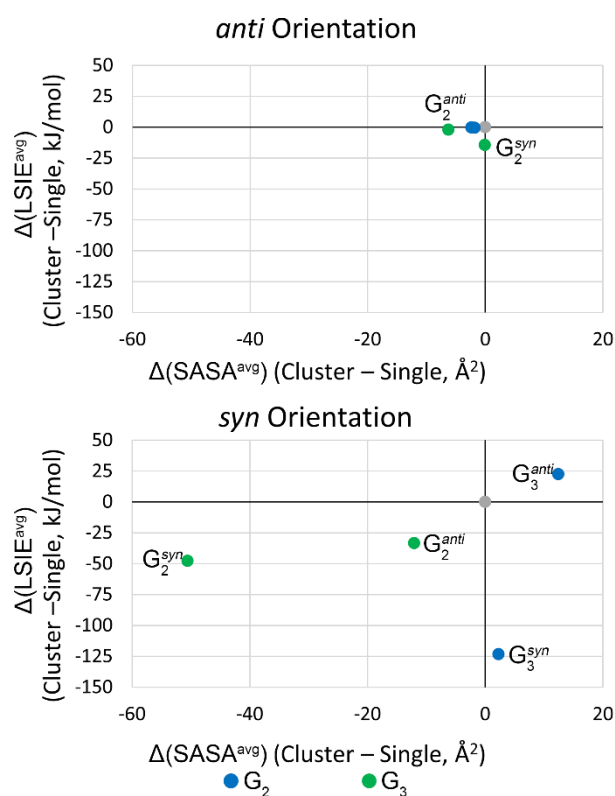


Figure 7.7:  $\Delta(\text{LSIE}^{\text{avg}})$  versus  $\Delta(\text{SASA}^{\text{avg}})$  for  $^{\text{ABP}}\text{dG}$  DNA upon addition of a second adduct with the first lesion in the *anti* (top) or *syn* (bottom) conformation for next-nearest neighboring adducts (the  $G_2G_3$  sequence context), original  $G_2$  site in blue and  $G_3$  site in green, with labels representing the site and conformation of the second lesion. Negative values indicate increased stability.

nearest lesions will significantly decrease the probability of an *anti* adduct orientation, there is a stark directional impact on the stability of diadducted DNA containing an *anti* and *syn* adduct, where  $G_2^{syn}G_3^{anti}$  is destabilized and  $G_2^{anti}G_3^{syn}$  is stabilized compared to monoadducted DNA, suggesting that  $G_2^{anti}G_3^{syn}$  will have a higher occupancy than  $G_2^{syn}G_3^{anti}$ .

For adducts separated by two intervening nucleotides ( $G_1G_3$ ), the lesion site interactions and solvent accessibility of adducts in the *anti* orientation do not change regardless of the lesion orientation at the second site (Figures 7.8, F.19 (top), Table F.27, Appendix F). Comparatively, the stability of a *syn* lesion site increases with the addition of a second lesion two nucleotides removed. At  $G_3$ , the LSIE is significantly greater when the

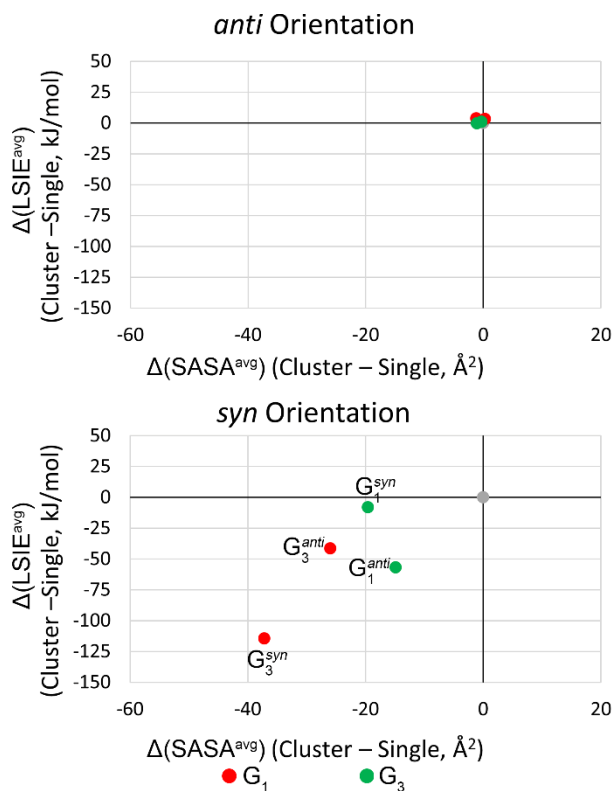


Figure 7.8:  $\Delta(LSIE^{avg})$  versus  $\Delta(SASA^{avg})$  for  $^{ABP}dG$  DNA upon addition of a second adduct with the first lesion in the *anti* (top) or *syn* (bottom) conformation for adducts separated by two intervening nucleotides (the  $G_1G_3$  sequence context), original  $G_1$  site in red and  $G_3$  site in green, with labels representing the site and conformation of the second lesion. Negative values indicate increased stability.

G<sub>1</sub> lesion adopts the *anti* over the *syn* orientation (> 45 kJ/mol difference, Figures 7.8, F.19, Table F.27, Appendix F). Due to the further stabilization of the inherently preferred *syn* orientation at G<sub>3</sub> (Figure 7.4) upon the addition of a second lesion, G<sub>3</sub><sup>*syn*</sup> likely dominates over G<sub>3</sub><sup>*anti*</sup> for G<sub>1</sub>G<sub>3</sub> diadducted DNA. Although the *syn* orientation is also stabilized at G<sub>1</sub> by the addition of a lesion at G<sub>3</sub> compared to monoadducted DNA, G<sub>1</sub><sup>*syn*</sup> is inherently unstable and the G<sub>3</sub> site has larger lesion site interactions when the G<sub>1</sub> adduct is *anti*. This suggests that *anti* will be the most prevalent conformation at G<sub>1</sub>. Overall, although G<sub>1</sub>G<sub>3</sub> diadducted DNA likely maintains conformational heterogeneity, the preferred G<sub>1</sub><sup>*anti*</sup> and G<sub>3</sub><sup>*syn*</sup> orientations of monoadducted DNA likely persist, suggesting that addition of a second lesions separated by two (or more) intervening nucleotides from the first damage site does not further impact the local conformation of damaged DNA.

In summary, the preferred conformation of each lesion site in <sup>ABP</sup>dG diadducted DNA depends on the spatial separation between the adducts, with the G<sub>1</sub><sup>*anti*</sup>G<sub>2</sub><sup>*anti*</sup>, G<sub>1</sub><sup>*anti*</sup>G<sub>2</sub><sup>*syn*</sup>, G<sub>1</sub><sup>*anti*</sup>G<sub>3</sub><sup>*syn*</sup> and G<sub>2</sub><sup>*syn*</sup>G<sub>3</sub><sup>*syn*</sup> conformations being the most prevalent (Figure 7.9). For all sequence contexts considered in the present work, a 5' *syn* adduct and a 3' *anti* adduct is the least desirable combination, highlighting an interesting directionality in the conformational preference of diadducted DNA. Although conformational heterogeneity likely prevails for <sup>ABP</sup>dG in diadducted DNA regardless of sequence context, the *anti* lesion orientation is uniquely preferentially stabilized at both lesion sites in G<sub>1</sub>G<sub>2</sub> diadducted DNA, while the *syn* conformation is uniquely preferentially stabilized at both sites in G<sub>2</sub>G<sub>3</sub> diadducted DNA. Thus, although the preferred conformations of monoadducted DNA are maintained when adducts are at least two nucleotides removed (G<sub>2</sub>G<sub>3</sub> or G<sub>1</sub>G<sub>3</sub>), neighboring adducts(G<sub>1</sub>G<sub>2</sub>) increase and next-nearest adducts decrease the prevalence of the *anti* adduct

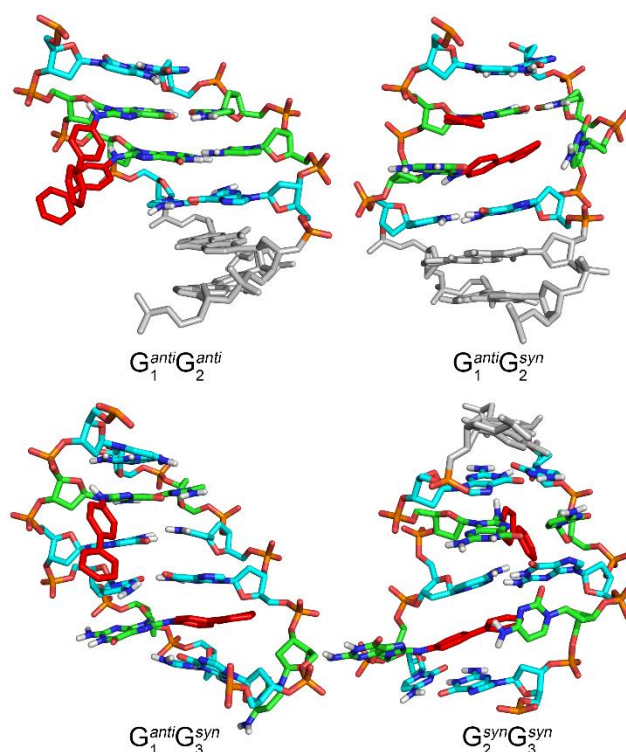


Figure 7.9: MD representative structures of the preferred conformations of <sup>ABP</sup>dG diadducted DNA for different sequence contexts. Only the base pairs in the 5'-CG<sub>1</sub>G<sub>2</sub>CG<sub>3</sub>C sequence context are shown in each case, with the adducted bases and pairing Cs shown in green, the bulky moieties in red, and the flanking bases in cyan.

orientation at both sites beyond the levels expected in monoadducted DNA. The <sup>ABP</sup>dG lesion thereby highlights a unique shift in the *anti:syn* conformational heterogeneity depending on lesion separation.

## 7.4 Discussion

### 7.4.1 Sequence Context Impacts the Preferred Conformation of <sup>ABP</sup>dG in Monoadducted DNA

Although the conformation of DNA containing <sup>ABP</sup>dG in the *NarI* sequence has not been previously reported, <sup>19</sup>F NMR and the <sup>FABP</sup>dG fluorine-labeled lesion have been used

to understand the structure of <sup>ABP</sup>dG monoadducted DNA in a variety of sequence contexts.<sup>6, 11, 22</sup> When ABP-adducted G adopts the *anti* glycosidic orientation at any G site in the *NarI* recognition sequence, the present work predicts the damaged DNA maintains a B conformation with the bulky group in the major groove and canonical WC hydrogen bonding at the lesion site, as previously seen for C8-dG adducts.<sup>9-19, 52, 59</sup> When the lesion acquires the *syn* orientation, the damaged DNA dynamically interconverts between three conformations, (I, S and W), with I being the most populated for <sup>ABP</sup>dG monoadducted DNA followed by S and then W. Most importantly, we predict that <sup>ABP</sup>dG can be stabilized in both the *anti* or *syn* glycosidic orientation in monoadducted DNA regardless of sequence context. However, the G<sub>1</sub> position (X in 5'-TCXGC) has the greatest preference for the *anti* orientation among the *NarI* sites. In fact, <sup>ABP</sup>dG monoadducted DNA is predicted to predominately adopt the B conformation when the lesion is in the 5'-TCXGC (X = G<sub>1</sub>) sequence context, which is similar to the conformational homogeneity reported for <sup>FABP</sup>dG in other sequence contexts (5'-TTXAC,<sup>22</sup> 5'-TCXCA<sup>6</sup> and 5'-CTXGTC<sup>11</sup>, X = lesion). Notably, 5'-TCXCA results in the same 5'-nearest and next-nearest neighboring bases as the G<sub>1</sub> sequence context, but differs in the identity of the 3' bases, while 5'-CTXGTC contains the same 3'-flanking G as G<sub>1</sub> and 5'-TTXAC is a completely unique sequence. Although the *anti* lesion orientation can likely also be adopted at G<sub>2</sub> and G<sub>3</sub>, the adducted nucleotide is predicted to be preferentially stabilized in the *syn* orientation at these sites. This suggests that damaged DNA with the 5'-CGXCG (X = G<sub>2</sub>) and 5'-GCXCC (X = G<sub>3</sub>) sequence contexts may have similar conformational heterogeneity as <sup>FABP</sup>dG in the 5'-TCXCT sequence (40:60 *anti:syn* ratio).<sup>6</sup> Interestingly, all three of these sequences contain a 3'-C with respect to the *syn* lesion.<sup>6, 11, 22</sup> Previous work has shown that the next nearest 3' base can also impact the conformational heterogeneity of <sup>FABP</sup>dG monoadducted DNA,

with the 60% *syn* lesion in 5'-TCXCT being reduced to 0% in the 5'-TCXCA sequence context.<sup>6</sup> Thus, our new data combined with previous experimental analysis of <sup>FABP</sup>dG underline the sensitivity of the local helical structure to the identity of the surrounding base pairs. Overall, when coupled with previous studies,<sup>6, 11, 22</sup> our work highlights the complex effect of sequence context on the conformational heterogeneity of monoadducted DNA, with evidence emerging that a 3' flanking C likely promotes a *syn* orientation more than other nucleotides.

#### 7.4.2 Inherent Biphenyl Flexibility Results in Unique Conformational Preferences

<sup>ABP</sup>dG is structurally similar to other C8-dG adducts previously studied in the literature, namely the N-linked <sup>AF</sup>dG<sup>6, 13, 19, 22, 23, 26, 27, 64</sup> and the C-linked biphenyl lesion (<sup>BP</sup>dG).<sup>54</sup> In fact, the <sup>AF</sup>dG and <sup>ABP</sup>dG adducts are often compared in the literature since they only differ by a methylene bridge between the aromatic rings in the <sup>AF</sup>dG bulky moiety, which induces planarity and prevents the free rotation between the <sup>ABP</sup>dG rings.<sup>6, 26</sup> The increased rigidity of <sup>AF</sup>dG is commonly believed to enhance the stability of the *syn* lesion orientation through greater stacking interactions with the flanking bases compared to <sup>ABP</sup>dG.<sup>6, 26</sup> For example, the B:S ratios of <sup>FABP</sup>dG in the 5'-TCXCT and 5'-TCXCA are 40:60 and 100:0, respectively, while those for <sup>FAF</sup>dG are 10:90 and 34:66.<sup>21</sup> The proposed enhanced *syn* stability for <sup>FAF</sup>dG is supported by our prediction that <sup>ABP</sup>dG at G<sub>1</sub> will predominately adopt the *anti* orientation and the reported greater *syn* <sup>FAF</sup>dG preference in the same sequence context (B:S ratio of 42:58).<sup>23</sup> However, our data suggest that <sup>ABP</sup>dG may similarly preferentially stabilize the *syn* orientation at G<sub>2</sub> and G<sub>3</sub>, whereas different B:S ratios have been reported for <sup>FAF</sup>dG monoadducted DNA (69:31 for G<sub>2</sub> and 35:65 for G<sub>3</sub>), with *anti* <sup>FAF</sup>dG predicted to dominate at G<sub>2</sub>.<sup>23</sup> The ability of flexible <sup>ABP</sup>dG to stabilize



the *syn* orientation to the same degree as or more than rigid <sup>FAF</sup>dG could arise because of the preferred I conformation of <sup>ABP</sup>dG adducted DNA compared to the strict S conformer of <sup>AF</sup>dG adducted DNA.<sup>19, 64</sup> Thus, although the planar fluorene moiety of <sup>AF</sup>dG has greater potential to stack with the flanking base pairs, the flexible biphenyl moiety of <sup>ABP</sup>dG can strongly stack with the sugar–phosphate backbone in an intrahelical position. Indeed, the ring farthest from the damaged G twists to form  $\pi$ –interactions with the minor groove (H1'H2'H4') face of the opposing deoxyribose, with similar sugar– $\pi$  interactions previously shown to be stabilizing.<sup>65, 66</sup> Furthermore, unlike the extrahelical position of the opposing C in the S conformation, the opposing C stacks with the flanking bases in the I conformation. Therefore, in contrary to the proposed impact of the flexible aryl group of <sup>ABP</sup>dG, newly identified local helical conformation in the present work can result in a greater *syn* preference for <sup>ABP</sup>dG at G<sub>2</sub> and G<sub>3</sub> than expected based on reported B:S ratios for <sup>FAF</sup>dG adducted DNA.<sup>6, 26</sup> Interestingly, the I conformer has previously been reported for <sup>BP</sup>dG,<sup>54</sup> the flexible C-linked analogue of <sup>ABP</sup>dG in which the aryl group is directly bound to C8, further supporting that it is the bulky group flexibility that leads to unique behaviour of <sup>ABP</sup>dG. Thus, our work complements and further clarifies published literature that highlights the sensitivity of the damaged DNA helical conformation to the lesion chemical composition.<sup>6, 10, 53, 54</sup>

#### **7.4.3 Addition of a Secondary <sup>ABP</sup>dG Adduct Differentially Impacts the Helical Structure, Conformational Preferences and Carcinogenicity Depending on Lesion Separation**

Although the local structure of monoadducted DNA in the B conformation is not greatly perturbed by the addition of a second lesion, the presence of a second damage site

impacts the ratio of damaged (I, S and W) DNA conformations when the lesion adopts the *syn* glycosidic orientation. Specifically, although monoadducted DNA primarily adopts the I conformation, the S conformer is the most populated for diadducted DNA (Table F.7, Appendix F). As a result, cluster lesions can influence the *anti:syn* ratio, and likely biological consequences, of <sup>ABP</sup>dG in damaged duplexes. Indeed, lesions at neighboring G<sub>1</sub>G<sub>2</sub> sites are predicted to prefer the *anti* orientation at both positions (G<sub>1</sub><sup>*anti*</sup>G<sub>2</sub><sup>*anti*</sup>). Although the *syn* conformation at G<sub>2</sub> is likely also prevalent (G<sub>1</sub><sup>*anti*</sup>G<sub>2</sub><sup>*syn*</sup>), cluster <sup>ABP</sup>dG lesions clearly induce *anti* conformational homogeneity at the G<sub>1</sub> site. This likely increases the carcinogenicity associated with <sup>ABP</sup>dG since the *anti* adduct orientation is anticipated to result in greater repair resistance than the *syn* orientation due to the more canonical structure of the DNA in the B conformation.<sup>52, 67</sup> This prediction correlates with experimental work on <sup>FAAF</sup>dG diadducts,<sup>25</sup> which revealed that G<sub>1</sub>G<sub>2</sub> cluster lesions have the lowest repair efficiency. In contrast, when two adducts are one nucleotide removed (G<sub>2</sub>G<sub>3</sub>), the <sup>ABP</sup>dG *syn* orientation is significantly stabilized at both sites, suggesting that the *anti* orientation will be even less prevalent than seen for the corresponding monoadducted DNA helices. The strong preference for the distorted G<sub>2</sub><sup>*syn*</sup>G<sub>3</sub><sup>*syn*</sup> structure likely increases the reparability of the lesion, which provides a structural rationalization for the experimentally observed greatest repair efficiency of <sup>FAAF</sup>dG in the same sequence context.<sup>25</sup> Consequently, due to the greater likelihood of adduct repair for G<sub>2</sub>G<sub>3</sub> cluster lesions, this sequence decreases the expected carcinogenicity of <sup>ABP</sup>dG. Finally, each adduct at sites two-nucleotide removed (G<sub>1</sub>G<sub>3</sub>) has the same conformational preference as the corresponding monoadducted DNA, suggesting multiple lesions separated by at least two nucleotides have little impact on the local helical environment or the repair propensities of either site. Overall, the *anti:syn* adduct separation effects seen for <sup>ABP</sup>dG generally align with the trends previously reported

for <sup>AN</sup>dG and <sup>AP</sup>dG diadducted DNA,<sup>52</sup> suggesting that the conformational preference of N-linked C8-dG diadducted DNA is more dependent on the lesion separation than the chemical composition of the aryl moiety. In summary, the addition of a second lesion can maintain the conformational preference and carcinogenicity of monoadducted DNA or change the *anti:syn* conformational ratio and mutagenic severity of the adduct depending on the lesion separation.

## 7.5 Conclusions

The present study uses MD simulations to investigate the structure of 18 DNA helices containing one or two <sup>ABP</sup>dG adducts in different (*anti/syn*) glycosidic conformations and sequence contexts (Table 7.1). Incorporation of a single <sup>ABP</sup>dG lesion in the *anti* orientation does not significantly disrupt the DNA duplex, with the bulky moiety situated in the major groove and lesion site WC pairing maintained. When a single <sup>ABP</sup>dG adduct adopts the *syn* orientation, a dynamic mixture of three damaged DNA conformations prevails, with I being more populated than S or W. Although the *anti* and *syn* orientations likely exist in a mixture regardless of sequence context, the dominant conformation of <sup>ABP</sup>dG monoadducted DNA shows a distinct sequence dependence, with the lesion preferentially adopting the *anti* orientation at G<sub>1</sub> and the *syn* conformer at G<sub>2</sub> and G<sub>3</sub>. Although the local helical structure associated with an *anti* lesion is not impacted by the presence of a second lesion in either glycosidic orientation, *syn* <sup>ABP</sup>dG preferentially induces the S over the I conformation in diadducted DNA. More importantly, the lesion site *anti:syn* ratio can be significantly impacted by a second damaged site, with the magnitude and direction of the equilibrium shift being dependent on the lesion separation. Specifically, neighboring G<sub>1</sub>G<sub>2</sub> lesions are predicted to preferentially stabilize the *anti*

orientation at both sites, which may reduce DNA repair recognition and thereby increase carcinogenicity.<sup>67</sup> In contrast, the next-nearest neighbor sequence (G<sub>2</sub>G<sub>3</sub>) is proposed to result in significant stabilization of *syn* <sup>ABP</sup>dG, which may increase adduct repairability and decrease negative biological consequences. Finally, ABP-damaged DNA containing two lesions separated by two intervening nucleotides (G<sub>1</sub>G<sub>3</sub>) is structurally similar at each lesion site to the corresponding monoadducted DNA duplexes, suggesting that the *anti:syn* ratio does not change upon a second damage event at least two or more nucleotides removed, resulting in mutagenic outcomes as monoadducted DNA. Interestingly, the lesion separation dependence observed for <sup>ABP</sup>dG diadducted DNA directly aligns with that previously reported for <sup>AN</sup>dG and <sup>AP</sup>dG diadducted DNA,<sup>52</sup> suggesting that these trends may hold true for other AA-derived N-linked C8-dG adducts regardless of bulky moiety size or flexibility. Overall, the present work advances our understanding of how several factors, including number of adducts, distance between adducts, bulky moiety size and shape, and sequence context, determine the structure of damaged DNA and therefore the anticipated biological outcomes.

## 7.6 References

1. International Agency for Research on Cancer, *Some Aromatic Amines, Organic Dyes, and Related Exposures*. IARC Press, International Agency for Research on Cancer: 2010; Vol. 99.
2. International Agency for Research on Cancer, Working Group on the Evaluation of Carcinogenic Risks to Humans. Meeting, *A Review of Human Carcinogens*. International Agency for Research on Cancer: 2012.
3. Baan, R.; Grosse, Y.; Straif, K.; Secretan, B.; El Ghissassi, F.; Bouvard, V.; Benbrahim-Tallaa, L.; Guha, N.; Freeman, C.; Galichet, L.; Coglianò, V., A Review of Human Carcinogens; Part F: Chemical Agents and Related Occupations. *Lancet Oncol.* **2009**, *10* (12), 1143-1144.

4. National Toxicology Program 12th Report on Carcinogens. *Rep. Carcinog.* **2011**, 12, iii-499.
5. Li, F.; Wang, H.; Wang, Y.; Feng, S.; Hu, B.; Zhang, X.; Wang, J.; Li, W.; Cheng, M., Computational Investigation Reveals Picrasidine C as Selective Pparalpha Lead: Binding Pattern, Selectivity Mechanism and Adme/Tox Profile. *J Biomol Struct Dyn* **2019**, 1-18.
6. Jain, V.; Vaidyanathan, V. G.; Patnaik, S.; Gopal, S.; Cho, B. P., Conformational Insights into the Lesion and Sequence Effects for Arylamine-Induced Translesion DNA Synthesis: 19F NMR, Surface Plasmon Resonance, and Primer Kinetic Studies. *Biochemistry* **2014**, 53 (24), 4059-4071.
7. The International Agency for Research on Cancer, *Diesel and Gasoline Engine Exhausts and Some Nitroarenes*. IARC: 2015.
8. The International Agency for Research on Cancer, *Overall Evaluations of Carcinogenicity. An Updating of Iarc Monographs Volumes 1 to 42*. IARC: 1987.
9. Patel, D. J.; Mao, B.; Gu, Z.; Hingerty, B. E.; Gorin, A.; Basu, A. K.; Broyde, S., Nuclear Magnetic Resonance Solution Structures of Covalent Aromatic Amine–DNA Adducts and Their Mutagenic Relevance. *Chem. Res. Toxicol.* **1998**, 11 (5), 391-407.
10. Kung, R. W.; Sharma, P.; Wetmore, S. D., Effect of Size and Shape of Nitrogen-Containing Aromatics on Conformational Preferences of DNA Containing Damaged Guanine. *J. Chem. Inf. Model.* **2018**, 58 (7), 1415-1425.
11. Cai, A.; Wilson, K. A.; Patnaik, S.; Wetmore, S. D.; Cho, B. P., DNA Base Sequence Effects on Bulky Lesion-Induced Conformational Heterogeneity During DNA Replication. *Nucleic Acids Res.* **2018**, 46 (12), 6356-6370.
12. Sproviero, M.; Verwey, A. M.; Rankin, K. M.; Witham, A. A.; Soldatov, D. V.; Manderville, R. A.; Fekry, M. I.; Sturla, S. J.; Sharma, P.; Wetmore, S. D., Structural and Biochemical Impact of C8-Aryl-Guanine Adducts within the NarI Recognition DNA Sequence: Influence of Aryl Ring Size on Targeted and Semi-Targeted Mutagenicity. *Nucleic Acids Res.* **2014**, 42 (21), 13405-13421.
13. Vaidyanathan, V. G.; Cho, B. P., Sequence Effects on Translesion Synthesis of an Aminofluorene-DNA Adduct: Conformational, Thermodynamic, and Primer Extension Kinetic Studies. *Biochemistry* **2012**, 51 (9), 1983-1995.
14. Cho, B. P.; Beland, F. A.; Marques, M. M., NMR Structural Studies of a 15-mer DNA Duplex from a ras Protooncogene Modified with the Carcinogen 2-Aminofluorene: Conformational Heterogeneity. *Biochemistry* **1994**, 33 (6), 1373-84.
15. Patnaik, S.; Cho, B. P., Structures of 2-Acetylaminofluorene Modified DNA Revisited: Insight into Conformational Heterogeneity. *Chem. Res. Toxicol.* **2010**, 23 (11), 1650-1652.

16. Cho, B. P.; Zhou, L., Probing the Conformational Heterogeneity of the Acetylaminofluorene-Modified 2'-Deoxyguanosine and DNA by 19F NMR Spectroscopy. *Biochemistry* **1999**, 38 (23), 7572-7583.
17. Nolan, S. J.; Vyas, R. R.; Hingerty, B. E.; Ellis, S.; Broyde, S.; Shapiro, R.; Basu, A. K., Solution Properties and Computational Analysis of an Oligodeoxynucleotide Containing N-(Deoxyguanosin-8-yl)-1-Aminopyrene. *Carcinogenesis* **1996**, 17 (1), 133-44.
18. Mao, B.; Vyas, R. R.; Hingerty, B. E.; Broyde, S.; Basu, A. K.; Patel, D. J., Solution Conformation of the N-(Deoxyguanosin-8-yl)-1-Aminopyrene ([AP]dG) Adduct Opposite dC in a DNA Duplex. *Biochemistry* **1996**, 35 (39), 12659-12670.
19. Mu, H.; Kropachev, K.; Wang, L.; Zhang, L.; Kolbanovskiy, A.; Kolbanovskiy, M.; Geacintov, N. E.; Broyde, S., Nucleotide Excision Repair of 2-Acetylaminofluorene- and 2-Aminofluorene-(C8)-Guanine Adducts: Molecular Dynamics Simulations Elucidate How Lesion Structure and Base Sequence Context Impact Repair Efficiencies. *Nucleic Acids Res.* **2012**, 40 (19), 9675-9690.
20. Xu, L.; Vaidyanathan, V. G.; Cho, B. P., Real-Time Surface Plasmon Resonance Study of Biomolecular Interactions between Polymerase and Bulky Mutagenic DNA Lesions. *Chem. Res. Toxicol.* **2014**, 27 (10), 1796-1807.
21. Jain, V.; Hilton, B.; Lin, B.; Patnaik, S.; Liang, F.; Darian, E.; Zou, Y.; MacKerell, A. D., Jr.; Cho, B. P., Unusual Sequence Effects on Nucleotide Excision Repair of Arylamine Lesions: DNA Bending/Distortion as a Primary Recognition Factor. *Nucleic Acids Res.* **2013**, 41 (2), 869-880.
22. Zhou, L.; Rajabzadeh, M.; Traficante, D. D.; Cho, B. P., Conformational Heterogeneity of Arylamine-Modified DNA: 19F NMR Evidence. *J. Am. Chem. Soc.* **1997**, 119 (23), 5384-5389.
23. Jain, V.; Hilton, B.; Patnaik, S.; Zou, Y.; Chiarelli, M. P.; Cho, B. P., Conformational and Thermodynamic Properties Modulate the Nucleotide Excision Repair of 2-Aminofluorene and 2-Acetylaminofluorene dG Adducts in the NarI Sequence. *Nucleic Acids Res.* **2012**, 40 (9), 3939-3951.
24. Xu, L.; Cho, B. P., Conformational Insights into the Mechanism of Acetylaminofluorene-dG-Induced Frameshift Mutations in the NarI Mutational Hotspot. *Chem. Res. Toxicol.* **2016**, 29 (2), 213-226.
25. Jain, V.; Hilton, B.; Lin, B.; Jain, A.; MacKerell, A. D.; Zou, Y.; Cho, B. P., Structural and Thermodynamic Insight into Escherichia Coli UvrABC-Mediated Incision of Cluster Diacetylaminofluorene Adducts on the NarI Sequence. *Chem. Res. Toxicol.* **2013**, 26 (8), 1251-1262.

26. Millen, A. L.; Sharma, P.; Wetmore, S. D., C8-Linked Bulky Guanosine DNA Adducts: Experimental and Computational Insights into Adduct Conformational Preferences and Resulting Mutagenicity. *Future Med. Chem.* **2012**, *4* (15), 1981-2007.
27. Shapiro, R.; Ellis, S.; Hingerty, B. E.; Broyde, S., Effect of Ring Size on Conformations of Aromatic Amine-DNA Adducts: The Aniline-C8 Guanine Adduct Resides in the B-DNA Major Groove. *Chem. Res. Toxicol.* **1998**, *11* (4), 335-341.
28. Jain, V.; Hilton, B.; Patnaik, S.; Zou, Y.; Chiarelli, M. P.; Cho, B. P., Conformational and Thermodynamic Properties Modulate the Nucleotide Excision Repair of 2-Aminofluorene and 2-Acetylaminofluorene dG Adducts in the NarI Sequence. *Nucleic Acids Res.* **2012**, *40* (9), 3939-51.
29. Wang, S.; Hanna, D.; Sugamori, K. S.; Grant, D. M., Primary Aromatic Amines and Cancer: Novel Mechanistic Insights Using 4-Aminobiphenyl as a Model Carcinogen. *Pharmacol. Ther.* **2019**, *200*, 179-189.
30. Cho, B. P.; Beland, F. A.; Marques, M. M., One-Dimensional Multiple Quantum Filtration 1h NMR Spectra of a 15-mer DNA Duplex Modified by the Carcinogen 4-Aminobiphenyl. *Magn. Reson. Chem.* **1993**, *31* (11), 1008-10.
31. Cohen, S. M.; Boobis, A. R.; Meek, M. E.; Preston, R. J.; McGregor, D. B., 4-Aminobiphenyl and DNA Reactivity: Case Study within the Context of the 2006 IPCS Human Relevance Framework for Analysis of a Cancer Mode of Action for Humans. *Crit. Rev. Toxicol.* **2006**, *36* (10), 803-19.
32. Shapiro, R.; Ellis, S.; Hingerty, B. E.; Broyde, S., Major and Minor Groove Conformations of DNA Trimers Modified on Guanine or Adenine by 4-Aminobiphenyl: Adenine Adducts Favor the Minor Groove. *Chem. Res. Toxicol.* **1995**, *8* (1), 117-127.
33. Nauwelaers, G.; Bellamri, M.; Fessard, V.; Turesky, R. J.; Langouet, S., DNA Adducts of the Tobacco Carcinogens 2-Amino-9h-Pyrido[2,3-B]Indole and 4-Aminobiphenyl Are Formed at Environmental Exposure Levels and Persist in Human Hepatocytes. *Chem. Res. Toxicol.* **2013**, *26* (9), 1367-77.
34. Broyde, S.; Hingerty, B. E.; Srinivasan, A. R., Influence of the Carcinogen 4-Aminobiphenyl on DNA Conformation. *Carcinogenesis (London)* **1985**, *6* (5), 719-725.
35. Fuchs, R. P. P., DNA Binding Spectrum of the Carcinogen N-Acetoxy-N-2-Acetylaminofluorene Significantly Differs from the Mutation Spectrum. *J. Mol. Biol.* **1984**, *177* (1), 173-180.
36. Koffel-Schwartz, N.; Verdier, J.-M.; Bichara, M.; Freund, A.-M.; Daune, M. P.; Fuchs, R. P. P., Carcinogen-Induced Mutation Spectrum in Wild-Type, Uvra and Umuc Strains of Escherichia Coli: Strain Specificity and Mutation-Prone Sequences. *J. Mol. Biol.* **1984**, *177* (1), 33-51.

37. Roy, A.; Fuchs, R. P. P., Mutational Spectrum Induced in *Saccharomyces Cerevisiae* by the Carcinogen N-2-Acetylaminofluorene. *Mol. Gen. Genet.* **1994**, *245* (1), 69-77.
38. Hoffmann, G. R.; Fuchs, R. P. P., DNA Sequence Analysis of Mutations Induced by N-2-Acetylaminofluorene in Plasmid Pbr322 in *Escherichia Coli*. *J. Mol. Biol.* **1990**, *213* (2), 239-246.
39. Budworth, H.; Dianova, II; Podust, V. N.; Dianov, G. L., Repair of Clustered DNA Lesions. Sequence-Specific Inhibition of Long-Patch Base Excision Repair by 8-Oxoguanine. *J. Biol. Chem.* **2002**, *277* (24), 21300-5.
40. Cunniffe, S.; O'Neill, P.; Greenberg, M. M.; Lomax, M. E., Reduced Repair Capacity of a DNA Clustered Damage Site Comprised of 8-Oxo-7,8-Dihydro-2'-Deoxyguanosine and 2-Deoxyribonolactone Results in an Increased Mutagenic Potential of These Lesions. *Mutat. Res.* **2014**, *762*, 32-9.
41. Sakofsky, C. J.; Saini, N.; Klimczak, L. J.; Chan, K.; Malc, E. P.; Mieczkowski, P. A.; Burkholder, A. B.; Fargo, D.; Gordenin, D. A., Repair of Multiple Simultaneous Double-Strand Breaks Causes Bursts of Genome-Wide Clustered Hypermutation. *PLoS Biol.* **2019**, *17* (9), e3000464.
42. Karwowski, B. T., The Influence of (5'r)- and (5's)-5',8-Cyclo-2'-Deoxyadenosine on UDG and hAPE1 Activity. Tandem Lesions Are the Base Excision Repair System's Nightmare. *Cells* **2019**, *8* (11).
43. Shikazono, N.; Akamatsu, K., Mutagenic Potential of 8-Oxo-7,8-Dihydroguanine (8-oxoG) Is Influenced by Nearby Clustered Lesions. *Mutat. Res.* **2018**, *810*, 6-12.
44. Georgakilas, A. G.; O'Neill, P.; Stewart, R. D., Induction and Repair of Clustered DNA Lesions: What Do We Know So Far? *Radiat. Res.* **2013**, *180* (1), 100-9.
45. Zhang, N.; Lin, C.; Huang, X.; Kolbanovskiy, A.; Hingerty, B. E.; Amin, S.; Broyde, S.; Geacintov, N. E.; Patel, D. J., Methylation of Cytosine at C5 in a CpG Sequence Context Causes a Conformational Switch of a Benzo[a]Pyrene Diol Epoxide-N 2-Guanine Adduct in DNA from a Minor Groove Alignment to Intercalation with Base Displacement. *J. Mol. Biol.* **2005**, *346* (4), 951-965.
46. Watt, D. L.; Utzat, C. D.; Hilario, P.; Basu, A. K., Mutagenicity of the 1-Nitropyrene-DNA Adduct N-(Deoxyguanosin-8-yl)-1-Aminopyrene in Mammalian Cells. *Chem. Res. Toxicol.* **2007**, *20* (11), 1658-1664.
47. Feng, Z.; Hu, W.; Rom, W. N.; Beland, F. A.; Tang, M.-s., N-Hydroxy-4-Aminobiphenyl-DNA Binding in Human p53 Gene: Sequence Preference and the Effect of C5 Cytosine Methylation. *Biochemistry* **2002**, *41* (20), 6414-6421.
48. Tang, M. S.; Zheng, J. B.; Denissenko, M. F.; Pfeifer, G. P.; Zheng, Y., Use of UvrABC Nuclease to Quantify Benzo[a]Pyrene Diol Epoxide-DNA Adduct Formation at



Methylated Versus Unmethylated CpG Sites in the p53 Gene. *Carcinogenesis* **1999**, 20 (6), 1085-1089.

49. Denissenko, M. F.; Chen, J. X.; Tang, M.-s.; Pfeifer, G. P., Cytosine Methylation Determines Hot Spots of DNA Damage in the Human P53 Gene. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, 94 (8), 3893.

50. Cui, X. S.; Eriksson, L. C.; Möller, L., Formation and Persistence of DNA Adducts During and after a Long-Term Administration of 2-Nitrofluorene. *Mutat. Res.* **1999**, 442 (1), 9-18.

51. Beland, F. A.; Kadlubar, F. F., Formation and Persistence of Arylamine DNA Adducts in Vivo. *Environ. Health Perspect.* **1985**, 62, 19-30.

52. Kung, R. W.; Deak, T. K.; Griffith-Salik, C. A.; Takyi, N. A.; Wetmore, S. D., Impact of DNA Adduct Size, Number, and Relative Position on the Toxicity of Aromatic Amines: A Molecular Dynamics Case Study of (an)dG- and (AP)dG-Containing DNA Duplexes. *J. Chem. Inf. Model.* **2021**, 61 (5), 2313-2327.

53. Wilson, K. A.; Kung, R. W.; Wetmore, S. D., Chapter Seven - Toxicology of DNA Adducts Formed Upon Human Exposure to Carcinogens: Insights Gained from Molecular Modeling. In *Advances in Molecular Toxicology*, Fishbein, J. C.; Heilman, J. M., Eds. Elsevier: 2016; Vol. 10, pp 293-360.

54. Berger, F. D.; Sturla, S. J.; Kung, R. W.; Montana, T.; Wetmore, S. D.; Manderville, R. A., Conformational Preference and Fluorescence Response of a C-Linked C8-Biphenyl-Guanine Lesion in the NarI Mutational Hotspot: Evidence for Enhanced Syn Adduct Formation. *Chem. Res. Toxicol.* **2018**, 31 (1), 37-47.

55. Dennington, R.; Keith, T. A.; Millam, J. M. S. *Gaussview, Version 6.0.16*, Semichem Inc.: Shawnee Mission KS, 2016.

56. Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. *Amber 2018*, University of California: San Francisco, 2018.

57. Ivani, I.; Dans, P. D.; Noy, A.; Perez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goni, R.; Balaceanu, A.; Portella, G.; Battistini, F.; Gelpi, J. L.; Gonzalez, C.; Vendruscolo, M.; Laughton, C. A.; Harris, S. A.; Case, D. A.; Orozco, M., Parmbsc1: A Refined Force Field for DNA Simulations. *Nat. Methods* **2016**, 13 (1), 55-8.

58. Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P., The R.E.D. Tools: Advances in Resp and Esp Charge Derivation and Force Field Library Building. *Phys. Chem. Chem. Phys.* **2010**, *12* (28), 7821-7839.
59. Manning, T. W.; Al-Abdul-Wahid, M. S.; Manderville, R. A.; Josephy, P. D.; Kung, R. W.; Wetmore, S. D., Structure of an Unusual Tetracyclic Deoxyguanosine Adduct: Implications for Frameshift Mutagenicity of Ortho-Cyano Nitroanilines. *Chem. Res. Toxicol.* **2020**, *33* (2), 584-593.
60. Le Grand, S.; Götz, A. W.; Walker, R. C., Spfp: Speed without Compromise—a Mixed Precision Model for Gpu Accelerated Molecular Dynamics Simulations. *Comput. Phys. Commun.* **2013**, *184* (2), 374-380.
61. Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C., Routine Microsecond Molecular Dynamics Simulations with Amber on Gpus. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9* (9), 3878-3888.
62. Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C., Routine Microsecond Molecular Dynamics Simulations with Amber on Gpus. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8* (5), 1542-1555.
63. Zgarbova, M.; Otyepka, M.; Sponer, J.; Lankas, F.; Jurecka, P., Base Pair Fraying in Molecular Dynamics Simulations of DNA and Rna. *J. Chem. Theory Comput.* **2014**, *10* (8), 3177-3189.
64. Jain, N.; Meneni, S.; Jain, V.; Cho, B. P., Influence of Flanking Sequence Context on the Conformational Flexibility of Aminofluorene-Modified dG Adduct in dA Mismatch DNA Duplexes. *Nucleic Acids Res.* **2009**, *37* (5), 1628-1637.
65. Wilson, K. A.; Kung, R. W.; D'Souza, S.; Wetmore, S. D., Anatomy of Noncovalent Interactions between the Nucleobases or Ribose and  $\Pi$ -Containing Amino Acids in Rna-Protein Complexes. *Nucleic Acids Res.* **2021**, *49* (4), 2213-2225.
66. Wilson, K. A.; Wells, R. A.; Abendong, M. N.; Anderson, C. B.; Kung, R. W.; Wetmore, S. D., Landscape of  $\Pi$ - $\Pi$  and Sugar- $\Pi$  Contacts in DNA-Protein Interactions. *J. Biomol Struct Dyn* **2016**, *34* (1), 184-200.
67. Lin, J.; Kadlubar, F. F.; Spitz, M. R.; Zhao, H.; Wu, X., A Modified Host Cell Reactivation Assay to Measure DNA Repair Capacity for Removing 4-Aminobiphenyl Adducts: A Pilot Study of Bladder Cancer. *Cancer Epidemiol Biomarkers Prev* **2005**, *14* (7), 1832-6.
68. Rubbi, C. P.; Milner, J., Analysis of Nucleotide Excision Repair by Detection of Single-Stranded DNA Transients. *Carcinogenesis* **2001**, *22* (11), 1789-1796.

69. Cai, Y.; Patel, D. J.; Geacintov, N. E.; Broyde, S., Differential Nucleotide Excision Repair Susceptibility of Bulky DNA Adducts in Different Sequence Contexts: Hierarchies of Recognition Signals. *J. Mol. Biol.* **2009**, *385* (1), 30-44.
70. Yang, W., Poor Base Stacking at DNA Lesions May Initiate Recognition by Many Repair Proteins. *DNA Repair* **2006**, *5* (6), 654-666.
71. Maillard, O.; Camenisch, U.; Clement, F. C.; Blagoev, K. B.; Naegeli, H., DNA Repair Triggered by Sensors of Helical Dynamics. *Trends Biochem. Sci.* **2007**, *32* (11), 494-499.
72. Mu, H.; Zhang, Y.; Geacintov, N. E.; Broyde, S., Lesion Sensing During Initial Binding by Yeast XPC/Rad4: Toward Predicting Resistance to Nucleotide Excision Repair. *Chem. Res. Toxicol.* **2018**, *31* (11), 1260-1268.
73. Min, J.-H.; Pavletich, N. P., Recognition of DNA Damage by the Rad4 Nucleotide Excision Repair Protein. *Nature* **2007**, *449* (7162), 570-575.
74. Velmurugu, Y.; Chen, X.; Slogoff Sevilla, P.; Min, J.-H.; Ansari, A., Twist-Open Mechanism of DNA Damage Recognition by the Rad4/XPC Nucleotide Excision Repair Complex. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (16), E2296.
75. Mu, H.; Geacintov, N. E.; Min, J.-H.; Zhang, Y.; Broyde, S., Nucleotide Excision Repair Lesion-Recognition Protein Rad4 Captures a Pre-Flipped Partner Base in a Benzo[a]Pyrene-Derived DNA Lesion: How Structure Impacts the Binding Pathway. *Chem. Res. Toxicol.* **2017**, *30* (6), 1344-1354.
76. Reeves, D. A.; Mu, H.; Kropachev, K.; Cai, Y.; Ding, S.; Kolbanovskiy, A.; Kolbanovskiy, M.; Chen, Y.; Krzeminski, J.; Amin, S.; Patel, D. J.; Broyde, S.; Geacintov, N. E., Resistance of Bulky DNA Lesions to Nucleotide Excision Repair Can Result from Extensive Aromatic Lesion-Base Stacking Interactions. *Nucleic Acids Res.* **2011**, *39* (20), 8752-8764.
77. Hess, M. T.; Schwitter, U.; Petretta, M.; Giese, B.; Naegeli, H., Bipartite Substrate Discrimination by Human Nucleotide Excision Repair. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (13), 6664.
78. Kropachev, K.; Ding, S.; Terzidis, M. A.; Masi, A.; Liu, Z.; Cai, Y.; Kolbanovskiy, M.; Chatgililoglu, C.; Broyde, S.; Geacintov, N. E.; Shafirovich, V., Structural Basis for the Recognition of Diastereomeric 5',8-Cyclo-2'-Deoxypurine Lesions by the Human Nucleotide Excision Repair System. *Nucleic Acids Res.* **2014**, *42* (8), 5020-5032.
79. Geacintov, N. E.; Broyde, S.; Buterin, T.; Naegeli, H.; Wu, M.; Yan, S.; Patel, D. J., Thermodynamic and Structural Factors in the Removal of Bulky DNA Adducts by the Nucleotide Excision Repair Machinery. *Biopolymers* **2002**, *65* (3), 202-210.

## **Chapter 8: Understanding the Nucleotide Excision Repair of the 2-Acetylaminofluorene-C8-dG Adduct: Impact of Number of Adducted Sites and Helical Conformation<sup>a,b,c</sup>**

### **8.1 Introduction**

DNA damage is ubiquitous in the genome. If persistent in cells, DNA damage can result in severe biological consequences, including neurodegenerative diseases and cancers.<sup>1, 2</sup> Fortunately, several repair pathways have evolved to revert damaged DNA to the canonical form.<sup>2</sup> One such pathway is nucleotide excision repair (NER), which is an evolutionary conserved pathway that is responsible for the repair of structurally diverse forms of noncanonical DNA.<sup>1-5</sup> Among other types of damage, NER excises mismatched DNA, intrastrand crosslinks (e.g., nucleobase dimers) from UV light and bulky DNA adducts,<sup>2-4</sup> including those considered in the present thesis. In order to remove these structurally varied lesions, NER involves a unique damage recognition process. Specifically, in eukaryotic cells, initial damage recognition is performed by the XPC-RAD23B-CETN2 complex (XPC).<sup>1, 2, 6</sup> Notably, mutations to XPC in humans can result in xeroderma pigmentosum (XP), a genetic disorder that greatly increases an individual's susceptibility to cancer, highlighting the importance of XPC to DNA damage response and overall organism survival.<sup>7</sup>

Although XPC plays a critical role in the long-term survivability of cells, the NER pathway consists of many proteins with varied functions. Upon binding to damaged DNA,

---

<sup>a</sup>The *Journal of American Chemical Society* Reference Style is used in this Chapter.

<sup>b</sup>Adapted from Kung, R.W.; Wetmore, S.D., Understanding the Nucleotide Excision Repair of the 2-Acetylaminofluorene-C8-dG Adduct: Impact of Number of Adducted Sites and Helical Conformation, [*in preparation*].

<sup>c</sup>R.W.K. performed all calculations, interpreted the data and wrote the first draft of the manuscript.

XPC recruits the 10-subunit general transcription factor II H complex (TFIIH), which unwinds DNA and confirms the presence of a lesion.<sup>8,9</sup> Subsequently, the damaged DNA strand is cleaved by the XPF-ERCC1 and XPG endonucleases before DNA polymerases use the remaining undamaged strand as a template to replace the cleaved strand.<sup>8</sup> Lastly, DNA ligases are recruited to seal the breaks in the backbone, restoring the DNA to its canonical form. A significant amount of work has been performed to understand the NER lesion recognition process, with several papers examining XPC and Rad4, the yeast ortholog of XPC.<sup>3, 6, 8-17</sup> Due to a lack of experimental structures of the XPC–DNA complex, most structural work geared toward understanding the NER binding and recognition process has been done on the Rad4 system.<sup>3, 8, 10-12, 17-19</sup>

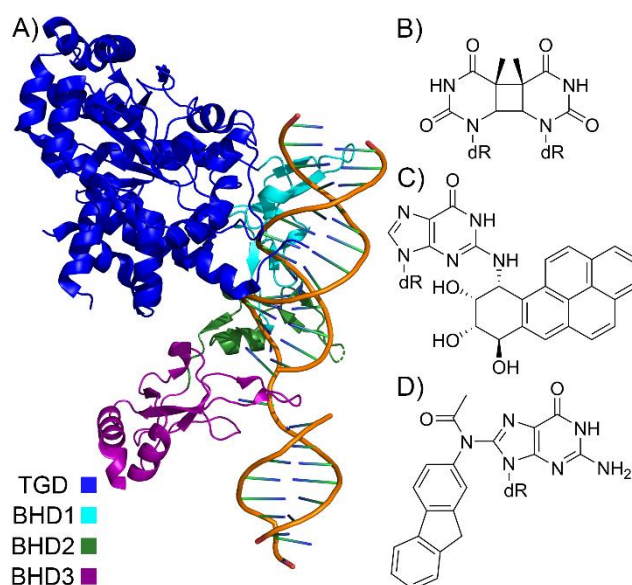


Figure 8.1: A) Crystal structure of a representative Rad4–DNA complex (PDB ID: 6CFI), with TGD in blue, BHD1 in cyan, BHD2 in green and BHD3 in purple. The chemical structure of B) CPD, C) *cis*-B[a]P-dG and D) the <sup>AAF</sup>dG adduct considered in the present work.

The transglutaminase domain (TGD) and β-hairpin domain 1 (BHD1) of Rad4 are responsible for the nonspecific binding of DNA (Figure 8.1A), allowing for the one-

dimensional diffusion of the complex along the duplex<sup>9, 10</sup> Interestingly, the structures of these domains are generally consistent between apo Rad4 and the protein bound to various DNA systems (e.g. undamaged DNA, *cis*-B[a]P-dG, 6-4 photoproduct, and cyclobutane pyrimidine dimer (CPD)).<sup>10, 11, 14</sup> However, it is the  $\beta$ -hairpin domain 2 (BHD2) and  $\beta$ -hairpin domain 3 (BHD3) that are tasked with probing for damage in the minor groove and major groove of the helix, respectively.<sup>9, 10, 16</sup> The lesion verification process of Rad4 involves binding of damaged DNA, which leads to the formation of a cavity in the duplex due to expulsion of the nucleotide across from the damage and its 3' neighbor out of the helix.<sup>10, 14</sup> Specifically, after the initial association of BHD2 and BHD3 to the helix, the lesion pairing nucleotide and its 3' neighbor are flipped from the interhelical DNA space into the protein binding pocket. The insertion of the two nucleobases of the undamaged strand into the Rad4 binding pocket, which allows for binding without direct contact with the lesion, is considered the primary reason for the broad specificity of NER.<sup>6, 9</sup> Lastly, the  $\beta$ -hairpin loop of BHD3 is inserted into the interstrand space, leading to a productively bound state.<sup>10, 14</sup>

A T-jump study has shown that the DNA opening step associated with helical unwinding and base flipping are the rate-limiting step for Rad4 binding to DNA.<sup>6</sup> Recent reports have suggested that Rad4/XPC binding to damaged DNA functions via a 'kinetic gating' mechanism.<sup>3, 6, 10, 11</sup> This theory proposes that both undamaged and damaged DNA can be opened by Rad4/XPC, but the high barrier for protein binding to canonical DNA results in the recognition protein complex diffusing prior to binding. This proposal is supported by observations that undamaged DNA can be bound by Rad4/XPC, but not cleaved, when the protein is tethered to the DNA.<sup>6, 18</sup>

Recent experimental work has strived to understand the human NER efficiencies of DNA adducts derived from 2-acetylaminofluorene (AAF).<sup>20-22</sup> Specifically, AAF is of particular interest as it is a potent carcinogen that was originally designed for use as a pesticide, but was never commercially implemented due its mutagenic potential.<sup>23</sup> Despite limited human exposure, AAF-derived lesions have been extensively studied as model adducts.<sup>20, 22, 24-41</sup> A significant body of work on DNA containing C8-dG AAF (<sup>AAF</sup>dG, Figure 8.1D), including <sup>19</sup>F NMR studies on fluorine-labeled <sup>AAF</sup>dG (<sup>FAAF</sup>dG), has revealed that <sup>AAF</sup>dG containing DNA adopts three conformations, namely B, S and W (Figure 8.2), in the mutational hotspot of the *NarI* recognition sequence (5'-G<sub>1</sub>G<sub>2</sub>CG<sub>3</sub>CC).<sup>22, 29, 32, 33, 38-41</sup> Specifically, the B conformation maintains canonical Watson-Crick (WC) base pairing at the lesion site, with the damaged nucleotide in the *anti* orientation and the bulky group positioned in the major groove of the duplex. Thus, for the B conformation, the minor groove is relatively unaffected by the presence of a lesion site. The remaining two helical conformations involve the rotation of the damaged nucleotide to a *syn* orientation. The stacked (S) conformation displaces the damaged G and pairing base into the major groove, while positioning the aromatic group in the interstrand space. Lastly, in the W conformation, the aryl moiety is located in the minor groove of the duplex, with the damaged G forming Hoogsteen interactions with the pairing base. Interestingly, the exact conformational ratio for damaged DNA and the human repair efficiency are dependent on the specific sequence context.<sup>21, 22</sup> Other work has shown that DNA containing two <sup>FAAF</sup>dG adducts in close proximity (diadducts) has an even more complex conformational heterogeneity and varying human repair efficiency.<sup>20, 21</sup> Thus, previous work has shown that the damaged DNA structure and NER susceptibility are dependent on both the number

of lesions and the local sequence context, although the atomic level structural details for these trends remain unclear.

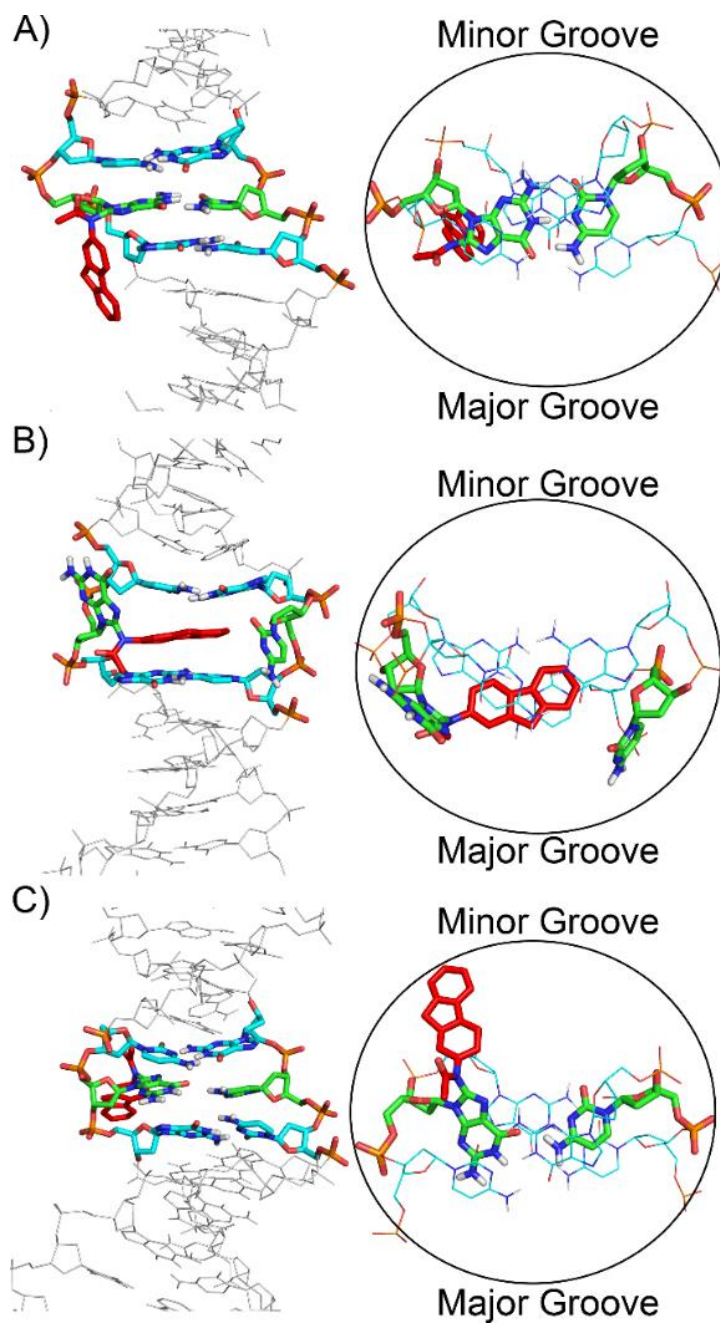


Figure 8.2: Structures of <sup>AAF</sup>dG-containing DNA in the A) B conformation, B) S conformation, and C) W conformation. The adducted base pair is shown in green, the bulky moiety in red, and the flanking base pairs in cyan.



Despite an abundance of work done to understand the structure of isolated damaged DNA,<sup>24, 28, 33, 40, 42</sup> relatively little is known about the structure of DNA bound to Rad4. Computational methods, primarily molecular dynamics (MD) simulations, have been utilized to provide valuable insight into the structure of Rad4–DNA complexes and the NER recognition process.<sup>10-12, 16-19</sup> Umbrella sampling techniques have been used to predict the binding pathway of Rad4 to DNA containing a *cis-syn* cyclobutane pyrimidine dimer (CPD, Figure 8.1B), 10R-(+)-*cis-anti*-benzo[a]pyrene-N2-dG (*cis*-B[a]P-dG, Figure 8.1C), or mismatched DNA. These works have collectively revealed a correlated mechanism where DNA base flipping occurs congruently with the association of the  $\beta$ -hairpin domains, instead of a proposed conformation capture mechanism in which base flipping for both the lesion pairing nucleotide and its 3' neighbor occurring prior to Rad4 diffusing to the location.<sup>10, 11, 16, 17</sup> These studies have highlighted a range of barriers for Rad4 recognition, with the barrier being the smallest for CPD-containing DNA (17.6 kJ/mol)<sup>10</sup> and *cis*-B[a]P-dG adducted DNA (21.3 kJ/mol),<sup>11</sup> whereas the Rad4 activation barrier for mismatched DNA ranged from 21.5 kJ/mol (CCC/CCC) to 64.9 kJ/mol (TAT/TAT) depending on the mismatched sequence.<sup>16, 17</sup> Additionally, long-time scale ( $\geq 1 \mu\text{s}$ ) MD simulations have been employed to rationalize the experimentally observed NER efficiencies of several substrates.<sup>12, 18, 19</sup> Specifically, Rad4–DNA complexes containing GGG and CCC mismatches, pyrimidine–pyrimidone 6-4 photoproduct (6-4PP), N2-dG polycyclic aromatic hydrocarbon adducts (PAHs, B[a]P and DB[a,l]P derived adducts) or C8-dG aromatic amine PhIP adducts were simulated by docking available NMR/MD-derived structures of the damaged duplexes to the apo protein.<sup>43</sup> This work has identified several structural factors of the initially bound Rad4–DNA complex that are predictors for efficient

binding, including DNA helix untwisting and bending, and the interactions between BHD2 and the DNA minor groove.<sup>12, 18, 19</sup>

Despite the insight provided by the previous computational studies on the binding of Rad4/XPC, only adducted DNA in a single conformation has been considered. However, adducted DNA commonly results in conformational heterogeneity,<sup>24, 28, 35, 40, 44, 45</sup> especially for <sup>AA</sup>FdG.<sup>20, 25, 33</sup> Furthermore, all adducts previously considered position the bulky group in the DNA minor groove (namely (+) *trans*-B[a]P-dG:dC and (–) *trans*-B[a]P-dG:dC) or in the interhelical space ((+) *cis*-B[a]P-dG:dC, (+) *cis*-B[a]P-dG:dT, (+) *cis*-B[a]P-dG:dA, (+) *cis*-B[a]P-dG:del, PhIP-C8-dG:dC, PhIP-C8-dG:del, 14*R*-DB[a,l]P-dA:dT and 14*R*-DB[a,l]P-dG:dC).<sup>11, 24, 35, 40</sup> Notably, the minor groove of the duplex is deviated from canonical DNA for all these systems considered. Therefore, to the best of our knowledge, the structure of the Rad4–DNA complex has not yet been considered to date for DNA containing a major groove lesion or DNA containing more than one local damage site. It is unclear how a bulky group in the major groove of the helix or multiple damage sites may impact BHD3 insertion or change interactions of BHD2 with the unmodified DNA minor groove.

To shed additional light on the NER recognition process, the present work employs numerous MD simulations to rationalize the differing Rad4 recognition capability for <sup>AA</sup>FdG. The model <sup>AA</sup>FdG lesion was chosen due to <sup>AA</sup>FdG damaged DNA exhibiting conformational heterogeneity,<sup>20, 25, 33</sup> including the B conformation that places the bulky moiety in the major groove, and the availability of experimental information for both mono and diadducted DNA.<sup>21</sup> The G<sub>1</sub> <sup>AA</sup>FdG monoadducted DNA and G<sub>2</sub>G<sub>3</sub> <sup>AA</sup>FdG diadducted DNA are considered as these represent the two repair efficiency extremes (incision rate

differing by a factor of 10), with DNA containing a single lesion at G<sub>1</sub> being the most repair prone and the G<sub>2</sub>G<sub>3</sub> diadducted helices being the most repair resistant of the <sup>AAF</sup>dG systems considered.<sup>21</sup> G<sub>1</sub> adducted DNA is considered in both the dominant B (46%) and S (34%) conformations,<sup>22</sup> while G<sub>2</sub>G<sub>3</sub> diadducted DNA was considered in the BB conformation based on the suggested stability of the conformation.<sup>20</sup> Consequently, this work will provide information into the influence of helical conformation at the damage site (B or S) and the impact from multiple DNA damage sites (monoadducted or diadducted DNA) on the Rad4 binding process, and general NER efficiencies. Furthermore, the binding of undamaged DNA was considered as a control, which will afford information about the impact of the damage and permit consideration of the proposed kinetic gating mechanism. Moreover, classical MD simulations on the monoadducted, diadducted and undamaged Rad4–DNA complexes in the initially bound state provide insight into the structural features that may be indicative of more efficient Rad4 damage recognition and cellular repairability. Finally, classical MD simulations on the productively bound complexes containing the three adducted helices provide further information about the structures after BHD3 insertion. Thus, this work provides the foundation for future studies to fully understand how the structural features of the damaged duplex impact the reparability of the lesion, and its overall toxicity.

## **8.2 Computational Details**

### **8.2.1 General Simulation Protocol**

The present work models <sup>AAF</sup>dG monoadducted and diadducted DNA as an isolated duplex and in complex with Rad4, with the model building described in detail in the following sections. An undamaged DNA system with the corresponding sequence as

monoadducted DNA was also considered to allow for more direct comparisons of Rad4 binding to damaged and undamaged DNA.

All MD simulations were performed with the AMBER18 suite of programs.<sup>46</sup> The parameters for <sup>AAF</sup>dG were generated to be compatible with the Amber force field using an analogous methodology to that implemented for other adducts.<sup>44, 45, 47</sup> Specifically, the R.E.D. program (version III.4)<sup>48</sup> was used to generate partial charges from B3LYP/6-31G(d) optimized structures of the *anti* and *syn* <sup>AAF</sup>dG nucleosides. Antechamber<sup>49</sup> was used to assign atom types. The remainder of the system was treated with the ff14SB forcefield<sup>50</sup> for the protein or OL15<sup>51</sup> for canonical DNA. Each system was neutralized with sodium cations and solvated in a TIP3P water cubic box, such that there was at least 10 Å between the edge of the water box and the solute, using the tleap program. Next, the systems were minimized, heated, and equilibrated using a multistep approach that was previously successfully used for adducted DNA helices.<sup>44, 52</sup> Briefly, the systems were minimized over four stages consisting of 2500 steps of steepest decent and 2500 steps of conjugate gradient minimization. A 100 kcal/(mol Å<sup>2</sup>) restraint was added to the solute for the first and third stages, with all heavy atoms restrained in the second stage. The final stage minimized the entire system without restraints. The systems were then gradually heated to 310 K over six 20 ps stages at constant volume, with a 25 kcal/(mol Å<sup>2</sup>) restraint applied to the solute. Equilibration was performed in five 20 ps stages, with the solute restraint decreased from 20 to 15, 10, 5 and 1.5 kcal/(mol Å<sup>2</sup>). All production simulations were run with a 2 fs timestep and a 10 Å nonbonded interaction cutoff.

### 8.2.2 Model Generation for Isolated DNA Duplexes

To generate the Rad4–DNA complexes, local helical structures of the damaged DNA must first be obtained. Three 16-mer helices based on the 5′–CTCTCG<sub>1</sub>G<sub>2</sub>CG<sub>3</sub>CCATCAC–3′ sequence used in <sup>FAAF</sup>dG NMR experiments,<sup>20, 22</sup> which contain the <sup>AAF</sup>dG *NarI* mutational hotspot sequence (underlined), were built from representative structures of damaged duplexes containing <sup>AP</sup>dG C8-dG adducts in the same sequence<sup>52</sup> using GaussView 6.0.16.<sup>53</sup> Duplexes were considered in the B and S conformations with <sup>AAF</sup>dG at G<sub>1</sub> and the BB conformation with <sup>AAF</sup>dG adducts simultaneously at G<sub>2</sub> and G<sub>3</sub>. The starting structure for undamaged B-DNA was built using the NAB program. Each system was then treated with the general MD simulation protocol

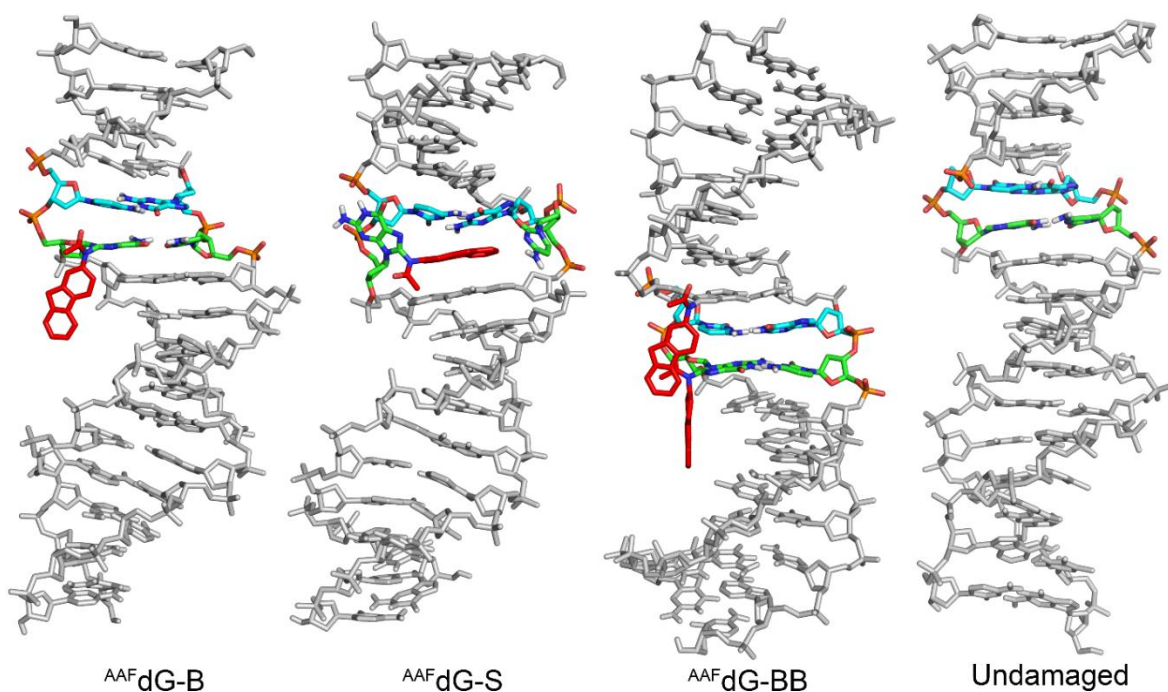


Figure 8.3: MD representative structures of the four (damaged and undamaged) duplexes considered in the present work complexed with Rad4, with the probed base pair in green, the flanking base pair that is broken during binding in cyan, and the bulky group(s) in red.

described above and simulated for 100 ns. Subsequently, the cluster command in cpptraj was used to obtain representative structures based on the lesion site(s) (i.e., the heavy atoms of the damaged base pair(s), Figure 8.3). The representative structures were then used to build protein–DNA complexes as described in the following section.

### 8.2.3 Model Generation for Rad4–DNA Complexes

The initial model for Rad4 was built from the crystal structure of the apo enzyme (PDB ID: 2QSF),<sup>14</sup> with the unresolved sections of the protein added via homology modelling using the SWISS-MODEL webserver.<sup>54</sup> The DNA duplex was docked to the apo Rad4 to maintain the nonspecific interactions with TGD and BHD1 based on the X-ray structure of Rad4 bound to DNA containing CPD (PDB ID: 2QSG).<sup>14</sup> Base pairs were added to both terminal ends of the DNA to match the sequence studied in the experimental work by the Cho research group (University of Rhode Island, unpublished), which resulted in a 27-mer (5'–ATTCTCTCG<sub>1</sub>G<sub>2</sub>CG<sub>3</sub>CCATCACA<sup>CT</sup>CACAC–3' for undamaged and monoadducted DNA and 5'–CTCTCG<sub>1</sub>G<sub>2</sub>CG<sub>3</sub>CCATCACCACA<sup>CT</sup>CACAC–3' for diadducted DNA). The sequence was slightly different for the modeled monoadducted and diadducted DNA systems compared to the experimental sequence (5'–ATTCTCTCG<sub>1</sub>G<sub>2</sub>CG<sub>3</sub>CCATCACA<sup>CT</sup>CA–3') to maintain consistent positioning of the probed (adducted) site. Specifically, the recognized damage site (i.e., the BHD3 insertion site) for each of the four systems (Figure 8.3) was at the same location as the productively bound CPD-containing DNA–Rad4 complex in the crystal structure (PDB ID: 2QSG),<sup>14</sup> with respect to the nonspecific DNA TGD and BHD1 binding domains. The diadducted system was bound such that BHD3 insertion occurs at the G<sub>3</sub> lesion site. Protonation states of the amino acids in the protein were determined using PROPKA<sup>55</sup> and visually

confirmed. Each of the four Rad4 systems were solvated, minimized, heated, and equilibrated as described for the general simulation protocol. The complexes were subsequently further equilibrated for 100 ns without restraints to obtain starting structures for the umbrella sampling calculations.

### 8.2.4 Umbrella Sampling of the Rad4–DNA Binding Pathway

The formation of the productive bound complex upon Rad4 binding to <sup>AAF</sup>dG monoadducted and diadducted DNA as well as undamaged DNA was modelled over four stages based on a previous umbrella sampling study on Rad4 binding to CPD containing DNA (Figure 8.4).<sup>10</sup> Specifically, the binding process was divided into: 1) the initial association of BHD2 and BHD3 to the binding site, 2) the coordination of the base flipping residue (MET498) to the 5' interstrand neighbor of the recognized (probed) nucleotide, 3) the flipping of the base opposing the lesion and its 3' neighbor, and 4) the insertion of the

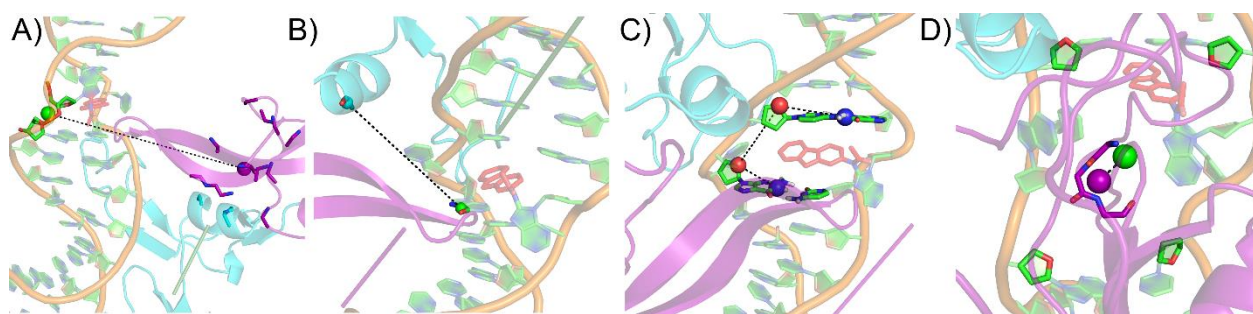


Figure 8.4: Rad4–DNA complex during the binding pathway of Rad4 to <sup>AAF</sup>dG adducted DNA in the S conformation, starting from the first capture state, highlighting the different stages modelled using umbrella sampling including the A) initial association of the BHDs to the damaged DNA site (Stage 1), B) coordination of the base flipping residues to the 5' interstrand neighbor of the adducted nucleotide (Stage 2), C) flipping of the base opposing the lesion and its 3' neighbor (Stage 3), and D) insertion of the BHD3  $\beta$ -hairpin loop into the interhelical space (Stage 4). The structures are color-coded as in Figure 8.1 and the AAF bulky moiety is shown in red. The atoms involved in the umbrella sampling CV definitions are shown as sticks, with the center of masses for the CV regions being shown as spheres.

$\beta$ -hairpin of BHD3 into the interhelical space (Figure 8.4, Tables G.1-G.4, Appendix G). Each binding stage consisted of a varying number of restrained MD (rMD) simulations (40+ per stage, Table G.5, Appendix G), with each simulation run for > 150 ns. Each window was restrained using AMBER harmonic restraints according to the values in Table G.5 (Appendix G). The harmonic anchors were set to a one window interval distance above and below the constrained collective variable (CV) for that stage. Prior to sampling in each window, the structure was equilibrated for 1 ns, with each initial structure sequentially built from a nearby window to minimize large changes in the structures or CVs. CVs were collected every 0.1 ps. Histograms of the CV sampled from each simulation were examined to ensure adequate overlap between, and sufficient sampling within, the windows (Figures G.1-G.4, Appendix G). The weighted histogram analysis method (WHAM)<sup>56</sup> was used to generate the potential of mean force (PMF, Figures G.5-G.8, Appendix G). Notably, Rad4 binding to undamaged DNA consisted of the same initial general stages as binding to the adducted helices, although our simulations revealed that a second base flipping stage was required to prepare for BHD3 insertion (Table G.4, Appendix G). For all base flipping stages, the reported energetic barrier corresponds to the highest energy point (CV) along the sampled path.

### **8.2.5 Unrestrained MD Simulations of Rad4–DNA Complexes**

Unrestrained MD simulations were run on the initially coordinated and productively bound Rad4–DNA complexes to better understand the corresponding structural features. Specifically, five 1  $\mu$ s MD trajectories were generated from the end of Stage 1 (initial association) and Stage 4 (BHD3 insertion) for Rad4 binding to each damaged DNA system. The initially coordinate (Stage1) system was also considered for unmodified DNA. Frames



were saved every 50 ps for analysis. The DNA helical twist was determined using the Curves+/Canal software.<sup>57</sup> In order to measure and easily compare the changes in helical twisting at the Rad4–DNA binding site, a single helical twist was calculated over the 6-mer segment of the duplex that includes the three base pairs 5' with respect to the recognized (probed) site and the two base pairs in the 3' direction (i.e., the two base pairs on either side of the two flipped nucleotides, Figure G.9, Appendix G). The interactions between BHD2 and the duplex were determined every 5 ns using AlphaSpace3.<sup>58</sup>

## **8.3 Results and Discussion**

### **8.3.1 The Conformation of <sup>AAF</sup>dG Monoadducted DNA Impacts the Predicted Barriers and Stabilities of Rad4 Binding, Indicating Differing Repair Propensities**

The binding of Rad4 to <sup>AAF</sup>dG monoadducted DNA in the B and S conformations was examined using advanced umbrella sampling methods over four stages. The MD representative structure of <sup>AAF</sup>dG in the B conformation (Figure 8.3) is similar to those previously reported for DNA containing <sup>AAF</sup>dG<sup>22, 41</sup> and other N-linked C8-dG lesions,<sup>24, 33, 47</sup> with WC base pairing at the lesion site remaining intact, the bulky group being oriented in the 3' direction and the minor groove face matching that of canonical DNA. Monoadducted DNA in the S conformation orients the bulky group in the interhelical space to stack with the flanking bases, which displaces the adducted G into the major groove such that N1–H and N2–H hydrogen bond with the phosphate backbone of the 5' neighboring nucleotide, as previously seen for <sup>AAF</sup>dG containing DNA.<sup>22, 29, 33</sup> Consequently, damaged DNA in the S conformation lacks hydrogen bonding between the adducted G and pairing C.

Models of the DNA–protein complexes were generated by docking MD representative structures from simulations on the damaged duplexes to apo Rad4 (PDB ID: 2QSF). The final coordination distances of the BHDs are similar for the monoadducted DNA complexes, with DNA containing the S conformation ( $\sim 18$  Å) only having slightly longer BHDs distances than the B conformation ( $\sim 17$  Å, Figures G.5-G.6, Appendix G). During the initial probing, the energy changes are minimal with no high energy regions along the pathway. Consequently, as this process is near barrierless, the energies for the binding process were scaled for each system, such that the energy after the initial probing of the BHDs (Stage 1) was set to 0 kJ/mol, allowing for direct comparisons between the different systems for the subsequent binding steps.

The coordination of the base flipping residues (Stage 2) does not significantly impact the stability of the complex ( $0 \pm 5$  kJ/mol) from the initially bound state formed following coordination of the BHDs (Stage 1, Figure 8.5, Table 8.1). Interestingly, the

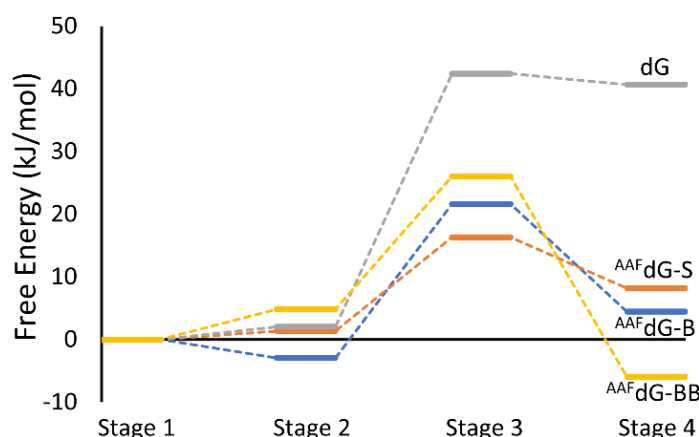


Figure 8.5: Free energy profile for the binding of Rad4 to monoadducted DNA in the B conformation (blue), monoadducted DNA in the S conformation (orange), diadducted DNA in the BB conformation (yellow) or undamaged DNA (grey) over four umbrella sampling stages (the four stages of Rad4 binding are described in Figure 8.4 and Tables G.1-G.5, Appendix G).

Table 8.1: Relative free energies (kJ/mol) for Rad4–DNA binding with respect to the corresponding initially bound structure.<sup>a</sup>

| Stage | <sup>AAF</sup> dG-B | <sup>AAF</sup> dG-S | <sup>AAF</sup> dG-BB | Undamaged |
|-------|---------------------|---------------------|----------------------|-----------|
| 1     | 0.0                 | 0.0                 | 0.0                  | 0.0       |
| 2     | −2.9                | 1.4                 | 4.8                  | 2.1       |
| 3     | 21.6                | 16.3                | 26.0                 | 42.4      |
| 4     | 4.4                 | 8.2                 | −6.0                 | 40.7      |

<sup>a</sup>The relative energies are reported with respect to the structure after Stage 1. The four stages of Rad4 binding are described in Figure 8.4 and Tables G.1-G.5 (Appendix G).

coordination of Rad4 to DNA containing a single <sup>AAF</sup>dG lesion at G<sub>1</sub> in the B conformation results in minimal stabilization (−2.9 kJ/mol), whereas coordination to the duplex in the S conformation results in nominal destabilization (1.4 kJ/mol). Despite the relatively similar energies after Stage 2, there are differences in the interactions between the BHDs and the C paired with the adducted G. At this stage in the binding process, the key residues that have been previously seen to stabilize the flipped nucleobases, namely the PHE “flipping path” from the BHD3 domain and Arg494 from the BHD2 domain,<sup>10, 16</sup> are positioning by the bases to facilitate the formation of a helical cavity. For the complex with DNA in the B conformation, the only direct interaction with the pairing C nucleotide is the ARG494 with DNA in the S conformation, the C that would be paired with the adduct is in extrahelical position such that it is hydrogen bonding to the 5′ neighboring phosphate group. Consequently, the Arg494 is still hydrogen bonding to the 5′ phosphate group as seen with the B conformation complex, but the displaced C also appears to be in a better position to be stabilized by the PHE “flipping path”, composed of Phe597, Phe599, and Phe556 (Figure G.10, Appendix G). Regardless of these differences, there are only minor energy barriers for the coordination, with the monoadducted DNA systems having a similar barrier of ~ 8 kJ/mol (Figures G.5-G.6, Appendix G), regardless of the conformation. Overall, there are minimal differences in the energetics associated with the coordination of the

BHDs (Stage 1) or the base flipping residues (Stage 2), with the process likely to readily occur for both monoadducted systems.

The base flipping stage (Stage 3) has the largest energetic barrier along the Rad4 recognition process regardless of the system considered (Figure 8.5, Table 8.1), and thereby corresponds to the transition state of the Rad4 recognition process. Interestingly, the global minimum for the base pair opening process for all four Rad4–DNA complexes occurs at similar coordinates along the flipping pathway, which corresponds to stable WC base pairing for the flipped nucleotides (Figure G.5-G.8, Appendix G). As the nucleobase is displaced from the helix during the base opening step, the stability decreases until a plateau is reached. The formation of the helical cavity for the BHD3 insertion is energetically accessible for DNA containing <sup>AA</sup>FdG in the S conformation, with a barrier of only 16.3 kJ/mol. This low barrier can likely be attributed to the lack of WC hydrogen bonding at the lesion site due to the inherent displacement of the adduct and pairing C in the S conformation. Indeed, the base flipping barrier is higher for the B conformation of <sup>AA</sup>FdG monoadducted DNA (21.6 kJ/mol). Potentially, it is the altered orientation of the PHE “flipping path and Arg494 could partly account for the lower barrier for the base flipping step of Rad4 binding to adducted DNA in the S conformation. Notably, for both monoadducted duplexes the flipped nucleobases are stabilized by key interactions with the PHE “flipping path”, as well as Arg494, which forms strong hydrogen bonds with the negatively charged phosphate backbone of the flipped bases (Figure G.11, Appendix G). In general, the barriers for base flipping of <sup>AA</sup>FdG damaged DNA in the B conformer are relatively low considering the stability of the G:C base pairs that are broken during cavity formation.

Overall, the barriers for Rad4 binding of <sup>AAF</sup>dG monoadducted DNA are comparable to those previously reported for DNA containing a CPD (17.6 kJ/mol)<sup>10</sup> or *cis*-B[a]P-dG (21.3 kJ/mol)<sup>11</sup> lesion. Interestingly, the barrier for Rad4 binding to B[a]P-N2-dG adducted DNA in an intercalated conformation (i.e., with the pairing C displaced into the major groove, the bulky group in the interstrand space and the damaged G in the minor groove), is larger than the <sup>AAF</sup>dG containing DNA in the stacked intercalated (S) conformation, being within 1 kJ/mol of <sup>AAF</sup>dG in the B conformation.<sup>11</sup> This suggests that the unique interactions within the helical conformations caused by the adduct chemical structure could lead to repair efficiency differences despite the DNA structure similarities. Notably, since <sup>AAF</sup>dG, CPD and *cis*-B[a]P-dG are known to be repaired by NER,<sup>10, 11, 21</sup> the similar barriers of <sup>AAF</sup>dG monoadducted DNA to the other lesions supports the validity of our approach.

The stability of the productive bound complex following insertion of the BHD3 domain (Stage 4) is highly system dependent (Figure 8.5, Table 8.1). Both monoadducted duplexes form stable insertion complexes (stability increased by > 8 kJ/mol) when the leading edge of BHD3 (SER603, THE604 and VAL605) is a similar distance from the center of the helical cavity (~ 7 Å, Figures G.5-G.6, Appendix G). During the insertion of BHD3, monoadducted DNA in the B conformation exhibits a moderate increase in stability (–17.2 kJ/mol), which results in a productive bound complex that is only 4.4 kJ/mol higher in energy than the complex formed after initial probing (Stage 1). Interestingly, for Rad4 binding to <sup>AAF</sup>dG adducted DNA in the S conformation, the complex is only 8.1 kJ/mol more stable than the transition state upon insertion of the BHD3, resulting in a final energy of 8.2 kJ/mol. Thus, although the binding of DNA in the B conformation has a larger base

opening barrier than DNA in the S conformation, the stabilization gained following BHD3 insertion is also slightly larger when the lesion adopts an *anti* rather than *syn* orientation. Overall, the B conformation results in a higher barrier for base flipping (by 5.3 kJ/mol) than the S conformation during Rad4 association and binding but forms a more stable bound state (by 3.8 kJ/mol). Thus, the damage recognition by Rad4 of monoadducted DNA in the B conformation is thermodynamically favoured compared to the S conformation which is kinetically preferred. Overall, these differences in the energetics for the binding of Rad4 to <sup>AAF</sup>dG monoadducted DNA in the B and S conformation indicate that the exact conformation of the adducted DNA impacts the recognition propensity, and overall repairability of the lesion.

### **8.3.2 The Binding of Rad4 to Diadducted DNA is Distinct from the Monoadducted Systems Being More Thermodynamically and Less Kinetically Preferred**

In order to understand the impact of a second local lesion site on Rad4 binding and lesion recognition, diadducted DNA in the BB conformation was modeled over the same four stages as monoadducted DNA. As with monoadducted DNA in the B conformation, the structure of diadducted DNA in the BB conformation (Figure 8.3) aligns with previously studied N-linked C8-dG lesions,<sup>24, 33, 47</sup> with the WC base pairing maintained at both lesion sites. Consequently, as with monoadducted DNA in the B and S conformations, the initial coordination of the BHDs (Stage 1) to diadducted DNA does not result in significant energy changes (Figure G.7, Appendix G). Thus, the further binding stages were scaled for the diadducted system as was done with monoadducted DNA complexes.

The coordination of the base flipping residues (Stage 2) is slightly less stable for diadducted DNA (4.8 kJ/mol) compared to monoadducted DNA ( $\leq 1.4$  kJ/mol). However,

monoadducted DNA in the B conformation and the diadducted DNA in the BB conformation have similar free energy plots during coordination, with two minima separated by a relatively low barrier and the stability decreasing for coordination distances  $> 20 \text{ \AA}$  or  $< 9 \text{ \AA}$  (Figures G.5, G.7, Appendix G). Despite the similarity, the barrier for coordination is larger for diadducted DNA in the BB conformation ( $\sim 13 \text{ kJ/mol}$ , Figure G.8, Appendix G) compared to the monoadducted DNA systems ( $\sim 8 \text{ kJ/mol}$ ). Notably, the higher energy barrier and less stable coordinated complex of Rad4 to diadducted DNA compared to monoadducted DNA aligns the experimental data the diadducted DNA is more repair resistant than the monoadducted DNA systems.<sup>21</sup>

The global minimum during the formation of the helical cavity for diadducted DNA is similar to the monoadducted systems, matching the canonical hydrogen bonding of the GC base pair 5' to the recognized adduct (Figure G.7, Appendix G). As with monoadducted DNA, the base pair opening process is the rate limiting step for the binding of diadducted DNA by Rad4 and thereby corresponds to the transition state of the Rad4 recognition process (Figures 8.1, G.7, Appendix G). Despite the formation of the helical cavity for diadducted DNA by Rad4 having a larger barrier for diadducted DNA in the BB conformation ( $26.0 \text{ kJ/mol}$ ) than either monoadducted system ( $\leq 21.6 \text{ kJ/mol}$ ), the energy required is still significantly lower than the energy associated with G:C base pairs. Interestingly, the higher energetic cost of base opening for the BB conformation of DNA containing two local lesions may be due to steric interactions in the major groove between the adduct at the bound site ( $G_3$ ) and the bulky group of the nearby lesion ( $G_2$ , Figure 8.3). Nevertheless, the binding of the  $G_3$  site was considered in the present study, whereas the recognition of the  $G_2$  site may have a lower barrier due to the 3' orientation of the bulky

group in the B conformation. Experimental evidence shows that the neighboring G<sub>1</sub>G<sub>2</sub> diadducted system is more efficiently repaired than other <sup>AAF</sup>dG diadducted sequence contexts (e.g. the G<sub>2</sub>G<sub>3</sub> sequence modeled), which may arise due to the BDH3 insertion occurring at the diadducted site, such that the bases pairing both lesions are flipped into the protein binding pocket (similar to CPD).<sup>10</sup> Furthermore, if the G<sub>1</sub> site and the 5' flanking base pair are flipped during Rad4 binding for the G<sub>1</sub>G<sub>2</sub> or G<sub>1</sub>G<sub>3</sub> sequences, the presence of the 3' adduct would likely have minimal impact due to the 3' directionality of the bulky group in the major groove. Overall, comparison to the barriers for damaged DNA suggests that the presence of a bulky moiety has a significant impact on Rad4 binding, regardless of the helical location of the bulky group (i.e., the major groove or interhelical space) or number of lesions (i.e., monoadducted or diadducted DNA).

As the differing conformation of monoadducted DNA had significant changes to the energetics of the insertion of the BHD3 domain (Stage 4) into the helical cavity, diadducted DNA also differs in the stability provided (Figure 8.5, Table 8.1). Notably, the leading edge of Rad4's BHD3 is similarly distanced with respect to the center of the helical cavity for diadducted DNA compared to the monoadducted systems (~ 7 Å, Figures G.5-G.7, Appendix G). Compared to monoadducted DNA, the BB diadducted DNA helix gains the most stability during BHD3 insertion (total of -32.0 kJ/mol). Consequently, the G<sub>2</sub>G<sub>3</sub> diadducted system has the most stable final bound state (-6.0 kJ/mol), being the only damaged system considered with a more stable bound than prebound state. This result is particularly interesting as it suggests that the larger the base flipping barrier (Stage 3) for <sup>AAF</sup>dG containing DNA, the greater the stability gained during the final insertion step (Stage 4) and the more stable the productive bound complex. Notably, the reasonable energetics



associated with the binding of Rad4 to <sup>AAF</sup>dG diadducted and monoadducted DNA further agrees with Rad4 implementing a correlated motion mechanism as previously preposed.<sup>10, 11, 16, 17</sup> Interestingly, the binding of BB diadducted DNA by Rad4 is more thermodynamically favourable than monoadducted DNA but less kinetically favourable. Considering the known higher repair resistance of the <sup>AAF</sup>dG diadducted helix, this provides further evidence that the rate recognition for lesions is kinetically and not thermodynamically determined, aligning with the proposed kinetic gating mechanism.<sup>21</sup>

### **8.3.3 Rad4 Binding to Undamaged DNA is Energetically Feasible During the Coordination of the BHDs and the Base Flipping Residues but Overall is Thermodynamically and Kinetically Unfavourable, Which Provides Further Evidence of a Kinetic Gating Mechanism**

In order to gain a better understanding of the NER recognition process, the binding of Rad4 to undamaged DNA was modelled. The isolated undamaged DNA duplex maintains the helical features expected for canonical B-DNA. As with the adducted system, the initial coordination of the BHDs to unmodified DNA (Stage 1) does not have any significant barriers (Figure G.8, Appendix G), with the resulting initially probed complex scaled for simpler comparisons with the sequential binding stages.

The energy associated with coordination of the base flipping residues (Stage 2) for undamaged DNA (2.1 kJ/mol, Figure G.8, Appendix G) is comparable to the damaged DNA systems, being larger than the monoadducted DNA complexes but less than the diadducted DNA system. The barrier along the coordination pathway for undamaged DNA (~ 7 kJ/mol) is also more similar to monoadducted DNA (~ 8 kJ/mol) than diadducted DNA

(~ 13 kJ/mol). Interestingly, this suggests that unmodified DNA could readily have initial association to Rad4 prior to the formation of a bound complex.

The undamaged system is unique in that two US stages were required to form the helical cavity (Stage 3 and Stage 3b, Figure G.8, Table G.4, Appendix G), whereas the damaged DNA systems only required a single base flipping step. Furthermore, undamaged DNA results in a notably higher barrier for the base flipping (42.4 kJ/mol) than any of the <sup>AAF</sup>dG adducted DNA systems. Nevertheless, the cost of Rad4 binding to unmodified DNA is not excessively large, especially compared to the stability of the WC hydrogen bonds that are broken upon base flipping.<sup>59</sup> This aligns with the experimentally-observed Rad4–DNA bound complexes when the protein is confined to a short segment of undamaged DNA (e.g., when the TGD of Rad4 is covalently tethered to the DNA).<sup>6, 18</sup> Despite these differences, the Stage 3 PMF of undamaged DNA has a similar minimum location and overall shape as the adducted DNA.

Unlike <sup>AAF</sup>dG adducted DNA, natural DNA results in a final bound product in which the BHD3 is inserted further into the duplex (~ 4 Å, Figure G.8, Appendix G). Consequently, the Rad4 binding to an unmodified duplex is only slightly stabilized upon BHD3 insertion (–1.8 kJ/mol, Figure 8.1, Table 8.1), which results in the final complex being 40.7 kJ/mol higher in energy than the initial coordination state (Stage 1). The higher barrier associated with Stage 3 for unmodified DNA coupled with the less stable productively bound complex aligns with the experimental evidence that Rad4 generally does not form productively bound complexes with canonical DNA, being both kinetically and thermodynamically unfavourable.<sup>3</sup>

Notably, although this barrier for Rad4 binding to undamaged DNA (42.4 kJ/mol) is higher than the <sup>AAF</sup>dG adducted DNA systems studied here (<26.0 kJ/mol), as well as the previously investigated CPD-containing DNA (17.6 kJ/mol)<sup>10</sup>, *cis*-B[a]P-dG adducted DNA (21.3 kJ/mol)<sup>11</sup> and CCC/CCC mismatched DNA (21.5 kJ/mol){Panigrahi, 2020 #17} systems, it is lower than some mismatched sequences previously studied (i.e., that of TAT/TAT at 64.9 kJ/mol and TTT/TTT at 51.0).{Panigrahi, 2020 #17} This is interesting as the binding for our undamaged DNA is more energetically accessible than the medium and low specificity TTT/TTT and TAT/TAT mismatches, respectively, suggesting that this barrier is not too high for Rad4 to bind in select situations. Despite this, the bound state of undamaged DNA is not as stable as the damaged DNA systems, which may partially account for the lack of binding as the energy of Rad4-DNA dissociation at the insertion site is minimal.

Overall, the energies along the binding pathway of Rad4 to undamaged, monoadducted and diadducted DNA correlate with the known NER efficiencies of <sup>AAF</sup>dG adducted DNA. Specifically, unmodified DNA is predicted to have the highest energetic barriers for Rad4 binding and Rad4 has been shown to not bind or cleave undamaged DNA.<sup>3</sup> <sup>AAF</sup>dG adducted DNA containing one lesion results in a lower barrier regardless of lesion site conformation than diadducted DNA, which correlates with the reported greater NER efficiency of the G<sub>1</sub> site compared to G<sub>2</sub>G<sub>3</sub>.<sup>21</sup>

Additionally, the umbrella sampling data supports the previously proposed kinetic gating mechanism.<sup>3, 6, 10, 11</sup> Specifically, it has been postulated that undamaged DNA does not bind to Rad4 due to the unfavorable kinetics, which suggests that that Rad4 binding to undamaged DNA will be slower than the Rad4 diffusion along the DNA duplex. The

fact that the trend in the barriers rather than the stability of the bound complexes matches the reported NER efficiencies<sup>21</sup> supports a kinetically, not thermodynamically, driven binding. The reasonable predicted energy barriers for Rad4 binding to <sup>AAF</sup>dG adducted DNA (16 – 26 kJ/mol) that are significantly smaller than that of undamaged DNA provide further evidence of a kinetic gating mechanism. The exact energy requirement for the binding of NER recognition proteins is likely dependent on the lesion site sequence context, especially the 5' pair flanking the lesion that is broken during BHD insertion, with the strength of stacking and hydrogen bonding at the insertion site likely contributing to the binding barrier height. This suggests that A:T rich sequences may have lower barriers than the currently studied G:C sequence, due to the weaker hydrogen bonding and stacking interactions, and therefore the impact of the identity of the 5' base with respect to the damage that is also displaced during binding on the binding efficiency would be interesting to consider in future work. Indeed, sequence effects have been previously reported, with undamaged DNA in the CCC/GGG sequence being easier to open than CGC/GCG sequence.<sup>18</sup>

#### **8.3.4 Unlike Other Lesions, Neither Helical Distortions nor BHD2 Interactions in the Prebound Rad4–DNA Complex are Sufficient Indicators of NER Efficiency for <sup>AAF</sup>dG Adducted DNA**

Although modelling the Rad4 damage recognition binding mechanism can provide valuable insight into the NER efficiency for DNA systems, this is a complex modeling problem requiring extensive computational resources. Alternatively, previous work computational work has suggested that select features of the prebound (probing) Rad4–DNA complex (i.e., helical untwisting at binding site, extent of BHD2 interactions in the

minor groove and helical bending) can be indicators of relative repair efficiency without utilizing advanced sampling methods.<sup>8, 12</sup> Thus, to compare the structural features of the Rad4 probing complex, 1  $\mu$ s MD simulations were performed in quintuplicate on the Rad4–DNA complex formed following the initial DNA probing by the BHDs (Stage 1 of US, Figure 8.4) for each of the three damaged DNA and the undamaged DNA complexes.

For each system, the degree of helical untwisting in the bound DNA was determined since a large twist in DNA after the initial binding has been previously suggested to be an indicator of enhanced NER efficiency.<sup>11, 12</sup> The helical untwisting was determined as the difference in the twist of the Rad4-bound DNA and the unbound canonical DNA (in the same sequence), as measured over the 6-mer duplex segment around the lesion site, which includes the two flanking base pairs on either side of the two bound bases that will be flipped (Figure G.11, Appendix G). For monoadducted DNA, the S conformation ( $28.2 \pm$

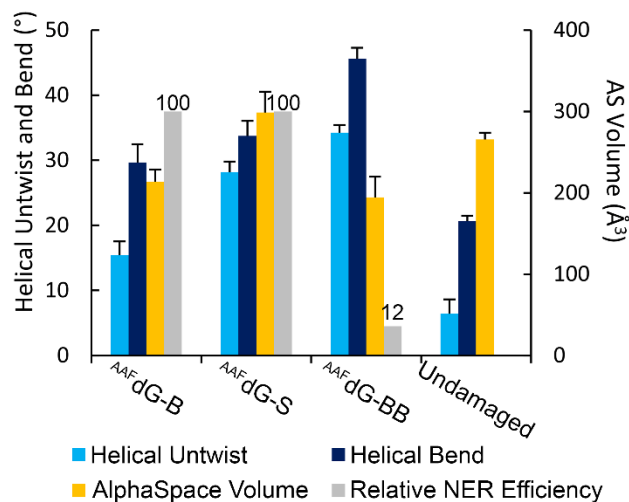


Figure 8.6: Average (and standard error) for the DNA structural properties and Rad4–DNA interactions (AlphaSpace volume) of the prebound state from quintuplicate 1  $\mu$ s MD simulations after the initial coordination (Stage 1), and the NER repair efficiencies of the adducted duplexes. Specifically, the helical untwist, helical bending and the BHD2-minor groove interaction (AlphaSpace) volume are reported along with the relative NER efficiency for the adducted site(s) from reference (from <sup>21</sup>).

1.6°) undergoes more unwinding than the B conformation ( $15.4 \pm 2.1^\circ$ ), which matches the lower energetic barrier for the binding of the S than B conformation (Figure 8.6, Table G.6, Appendix G). This finding aligns with the increased helical distortion of the S conformer in the absence of Rad4 noted previously for N-linked adducts (e.g., <sup>AN</sup>dG, <sup>ANP</sup>dG, <sup>ABP</sup>dG, and <sup>AP</sup>dG).<sup>52</sup> This also supports previous computational work on the initially bound Rad4–DNA complexes that determined that other bulky adducted DNA systems have helical untwisting up to 60° depending on the damage, with all systems that are excised having untwisting > 10°.<sup>11</sup> Despite the previously reported range in helical untwisting, only the 14R-DB[*a,l*]P-dG containing DNA is untwisted more than 40°, with most damaged helices (e.g., *cis*-B[*a*]P-dG:dC, *cis*-B[*a*]P-dG:dT, *trans*-B[*a*]P-dG:dC and PhIP-C8-dG:dC) being untwisted by approximately  $20^\circ \pm 10^\circ$ . Consequently, <sup>AAF</sup>dG damaged DNA has similar degrees of helical untwisting when bound to Rad4 as other bulky DNA adducts, including another N-linked C8-dG lesion (PhIP-C8-dG).<sup>11</sup>

Comparatively, the unmodified DNA has the smallest change in the helical twist ( $6.4 \pm 2.2^\circ$ ) compared to isolated canonical DNA (Figure 8.6, Table G.6, Appendix G), which aligns with the fact that Rad4 binding to undamaged DNA has the highest barrier among the systems considered. The low amount of untwisting for undamaged DNA (6.4°) also aligns with the untwist values for adducted systems that are not repaired by the NER pathway (i.e., 14R-DB[*a,l*]P-dA:dT, (+) *cis*-B[*a*]P-dG:del and (+) *cis*-B[*a*]P-dG:dA).<sup>11</sup> The pre-bound complexes for diadducted DNA results in the greatest untwisting ( $34.3 \pm 1.2^\circ$ ) despite diadducted DNA having the lowest human NER efficiency and largest Rad4 binding barrier among the adducted systems. This highlights some of the complexity in understanding the NER efficiencies of adducted DNA that although helical untwisting may

indicate the repair propensity for some systems, the method does work for diadducted DNA complex.

Interestingly, the degree of helical untwisting upon initial association of Rad4 correlates with the amount of helical bending at the BHD2 and BHD3 probed site (Figure 8.6, Table G.6, Appendix G), supporting the correlation between these helical untwist and bend previously seen for 6-4PP.<sup>19</sup> Indeed, undamaged DNA is the least bent ( $20.6 \pm 0.8^\circ$ ) followed by monoadducted DNA, with <sup>AAF</sup>dG adducted DNA in the B conformation being less bent ( $29.7 \pm 2.8^\circ$ ) than in the S conformation ( $33.8 \pm 2.3^\circ$ ). Finally, diadducted DNA is the more bent around the damaged sites ( $45.6 \pm 1.7^\circ$ ). Thus, although the degree of helical untwisting and helical bending for the prebound (Stage 1) Rad4–DNA complexes correlates with the reported NER binding and efficiency for monoadducted and unmodified DNA, the diadducted model showcases that these indicators are not universal and cannot be applied to all systems.

In addition to helical untwisting, the degree of interaction between the BHD2 domain and the DNA minor groove after the initial association of Rad4 to damaged DNA has also been suggested to be positively correlated to the repair efficiency.<sup>12</sup> As done previously, the interaction pocket (alpha space) volume was used to reflect the surface area and curvature of the minor groove region that is occupied by the BHD2 in order to quantify the extent of the BHD2–DNA interactions.<sup>12</sup> Interestingly, the level of interaction between BHD2 with the DNA minor groove is similar for the B conformation of monoadducted DNA and the BB conformation of diadducted DNA (Figure 8.6, Table G.6, Appendix G). These complexes have the smallest average volumes ( $214 \text{ \AA}^3$  and  $195 \text{ \AA}^3$ , respectively). In comparison, the natural DNA complex shows a greater interaction volume ( $266 \text{ \AA}^3$ ), while

the S conformer of monoadducted DNA has the largest interaction space (299 Å<sup>3</sup>). Notably, the largest BHD2–minor groove interaction volume exists for monoadducted DNA in the S conformation, which is the only system considered in the present work that modifies canonical DNA in the minor groove. Nevertheless, despite the S conformation having the largest interaction and the greatest repair propensity, the extent of BHD2 interactions is not a suitable indicator of NER efficiency for all helical systems.

Overall, despite the larger AS volume, the small helical distortions (helical bend and twist) for undamaged DNA could potentially increase the barrier for Rad4 binding, resulting in the lack of Rad4 binding to canonical DNA. Nevertheless, the BHD2–DNA interaction pocket volume may not be the best metric for the NER efficiency of major groove lesions as the unmodified system has a larger AS volume than monoadducted DNA in the B conformation. Thus, neither helical distortions nor the BHD2 interaction volume for the initially bound complexes appears to be a clear indicator for the overall experimental human NER efficiency<sup>12</sup> or computationally predicted Rad4 recognition propensity. Notably, previous work<sup>8, 12</sup> that indicated both factors may play a role in predicting NER efficiencies examined damaged DNA containing the bulky group in the minor groove or interhelical space, both of which modify the minor groove at the lesion site. Therefore, BHD2 interactions in the minor groove of the duplex may require a change in the minor groove interfaces to reflect the lesion repairability in human cells. Consequently, the changes in the helical structure upon Rad4 binding may be a more universal indicator for NER efficiency for monoadducted DNA, although this trend cannot be applied to all adducted systems as it does not hold for diadducted systems. In summary, although the helical distortions and BHD2 interactions may provide some insight into the potential



repairability of damaged DNA, there is no clear single factor that universally correlates with the NER efficiency. Therefore, modelling of the base-flipping mechanism or potentially a final Rad4–DNA bound state, may be required to accurately predict NER efficiency for damaged DNA complexes.

### 8.3.5 Helical Bending at the Damaged Site in the Productively Bound Rad4–DNA Complex is Correlated with Calculated Barriers and Complex Stabilities

MD simulations on the bound DNA complexes generated upon BHD3 insertion (Stage 4) reveal that the helical untwisting upon Rad4 binding (ranging between 28° and 44°, Figure 8.7, Table G.6, Appendix G) to monoadducted DNA depends on the helical conformation, with the B conformation being the least untwisted damaged duplex for the initially bound state ( $15.4 \pm 2.1^\circ$ , Figure 8.7, Table G.6, Appendix G), but the most

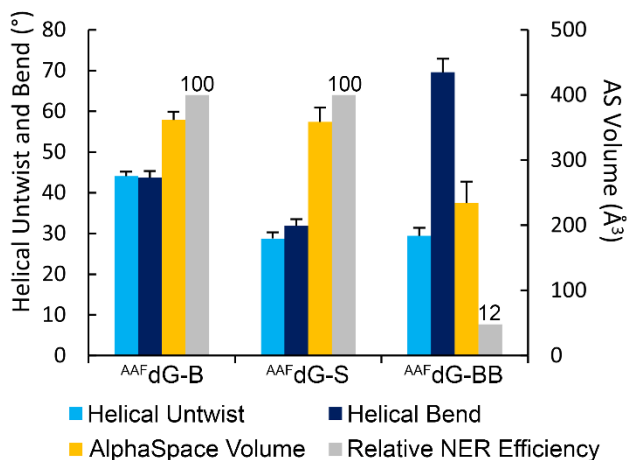


Figure 8.7: Average and standard error for the DNA structural properties and Rad4–DNA interactions (Alphaspace volume) of the bound state from quintuplicate 1  $\mu$ s MD simulations after the insertion of the BHD3 into the interhelical space (Stage 4), and the NER repair efficiencies of the adducted duplexes. Specifically, the helical untwist, helical bending and the BHD2-minor groove interaction (AlphaSpace) volume are reported along with the relative NER efficiency for the adducted site(s) (from <sup>21</sup>).

untwisted duplex following BHD3 insertion ( $44.0 \pm 1.1^\circ$ ). Monoadducted DNA in the S conformation has similar helical twist values after initial Rad4 association ( $28.2 \pm 1.6^\circ$ ) and in the productive bound state ( $28.7 \pm 1.5^\circ$ ), suggesting that the BHD3 insertion did not significantly change the helical untwisting for <sup>AAF</sup>dG-containing intercalated DNA. Despite diadducted DNA having the greatest helical untwisting after coordination of the BHDs ( $34.3 \pm 1.2^\circ$ ), the helical untwist is similar for the productively bound complex involving diadducted DNA ( $29.4 \pm 2.0^\circ$ ). Notably, the degree of helical untwisting for the productively bound complexes of monoadducted DNA in the S conformation and diadducted DNA are within  $2^\circ$  (Figure 8.7, Table G.6, Appendix G). Overall, despite some changes to the helical twist occurring during Rad4 binding, there does not appear to be a clear trend between the helical untwisting of the bound Rad4–DNA complex and the stability of the productive bound complex or the overall binding barrier (Figure 8.5).

In contrast to the bending of the prebound Rad4–DNA complexes, the helical bend for the bound complexes follows the same trend as the binding barrier. Indeed, diadducted DNA has the largest barrier (26.0 kJ/mol, Figure 8.5) and the greatest helical bend ( $69.6 \pm 3.3^\circ$ , Figure 8.7, Table G.6, Appendix G). Comparatively, monoadducted DNA in the S conformation has the smallest Rad4 binding barrier (16.3 kJ/mol) and is the least bent damaged system in the productively bound complexes ( $31.9 \pm 1.7^\circ$ ). Lastly, monoadducted DNA in the B conformation has an intermediate barrier (21.6 kJ/mol) and helical bend ( $43.7 \pm 1.6^\circ$ ). Thus, the larger binding barrier could be due at least in part to the energetic cost associated with the greater helical bending. Furthermore, a larger helical bend at the lesion site could also result in a greater stability gain from BHD3 insertion and the corresponding stability of the opened Rad4–DNA system (Figures 8.5, 8.7). Thus, the

helical bending of the bound site appears to be correlated with both the recognition barrier height, the final bound complex stability as well as the overall NER efficiency.<sup>21</sup>

As expected, the interacting volume is significantly larger ( $\geq 40 \text{ \AA}^3$ ) for the three adducted DNA systems in the productive bound complex than for the same systems after the initial coordination of the BHDs (Figure 8.7, Table G.6, Appendix G). The BHD2 alpha space volume for both the B ( $362 \text{ \AA}^3$ ) and S ( $359 \text{ \AA}^3$ ) conformations of monoadducted DNA are remarkably similar considering the larger ( $85 \text{ \AA}^3$ ) difference in the interaction pocket volume for the prebound complex. Although there are some differences in the helical distortions for the two <sup>AAF</sup>dG monoadducted DNA systems, the similar AS volumes indicate the initial conformation of the duplex may have minimal effect on the interactions between the bound Rad4 and the damaged DNA. Interestingly, as discussed for the initially bound complex, the diadducted DNA–Rad4 complex has a significantly lower alpha space volume ( $234 \text{ \AA}^3$ ) than either of the monoadducted DNA–Rad4 complexes, which may account for the experimentally observed poor NER efficiency.<sup>21</sup>

Although modelling Rad4–DNA bound structures is more computationally challenging than the prebound state, the monoadducted and diadducted complexes provide insight into the binding process. Overall, despite the degree of helical untwisting in the bound Rad4–DNA complexes having no clear correlation with the binding barrier strength or bound complex stability, there is positive correlation between the change in the helical bending around the lesion site upon BHD3 insertion and the stability of the resulting bound complexes. Moreover, the extent of helical bending is related to the energetic cost for Rad4 binding, with the more bend systems having a larger barrier and lower experimental NER efficiency. Interestingly, the previous work on *cis*-B[a]P-dG adducted DNA supports the

bending being a useful indicator of Rad4 binding barriers.<sup>11</sup> Specifically, *cis*-B[a]P-dG adducted DNA in a productive binding state has a bend angle of *cis*-B[a]P-dG (45°) and a barrier of 21.3 kJ/mol),<sup>11</sup> which is remarkably similar to the helical bending and Rad4 barrier values of Rad4 binding to <sup>AA</sup>FdG monoadducted DNA in the B conformation of 44° and 21.6 kJ/mol, respectively. Lastly, the minor groove interaction volume for the ‘opened’ bound complexes do not show consistent trends with the predicted barriers, with the monoadducted DNA systems having similar levels of interaction and the diadducted DNA system being notably smaller. Thus, unlike the helical twist and BHD2-minor groove interaction volume, it appears that the helical bending is correlated with the barrier of the bound Rad4–DNA systems.

## 8.4 Conclusions

By examining the Rad4 binding process to monoadducted DNA in the B and S conformations, diadducted DNA and undamaged DNA, this work has highlighted that there are differences in the energetics with different systems being kinetically and thermodynamically favoured (Figure 8.5). Notably, the thermodynamically preferred Rad4-diadducted DNA complex is experimentally shown to be the most repair resistant<sup>21</sup> providing evidence that the NER recognition is determined kinetically, with the stability of the rate-limiting base-flipping step indicating the lesion repairability. Interestingly, the S conformation has the smallest barrier (16.3 kJ/mol) followed by the B conformation (21.6 kJ/mol), suggesting that DNA in the S conformation will be more efficiently recognized and repaired than in the B conformation. The lower barrier of the S conformation could be partly accounted for by the different interactions with the PHE “flipping path” and Arg494 (Figure G.10, Appendix G), which have been previously shown to be key residues in

stabilizing the flipped nucleobases.<sup>10, 16</sup> Compared to the monoadducted duplexes, the diadducted DNA system in the BB conformation (26.0 kJ/mol) is less readily bound by Rad4. Finally, despite the binding of canonical DNA by Rad4 being energetically feasible (barrier of 42.4 kJ/mol), it is neither kinetically nor thermodynamically favoured, which supports the experimental evidence that Rad4 does not form productively bound complexes with undamaged DNA.<sup>3</sup> In summary, the predicted Rad4–DNA binding barriers align with the experimentally determined NER efficiencies,<sup>3, 21</sup> which could account for the 10-fold difference in repair efficiency for the better repaired G<sub>1</sub> monoadducted helix compared to G<sub>2</sub>G<sub>3</sub> cluster lesion system. Overall, the predicted energy required for DNA opening by Rad4 is dependent on the conformation of the <sup>AA</sup>FdG damaged DNA (B versus S) and the number of local lesions (monoadducted DNA versus diadducted DNA).

The high barrier Rad4 binding to undamaged DNA supports the proposed kinetic gating mechanism, in which damage detection by Rad4/XPC at a given DNA site is determined based on a competition between the residence time of the protein at that site and the time required for Rad4/XPC to form a productive bound complex.<sup>3, 6, 10, 11</sup> as Rad4 diffusion along the DNA should preferentially occur instead of DNA binding. Notably, this barrier is still significantly lower than the stability of canonical WC bonding of a single base pair,<sup>59</sup> showcasing the impressive ability of Rad4 to make the base opening feasible when the protein is artificially confined to the local DNA.<sup>6</sup> In fact, the experimentally-reported repair resistance of diadducted DNA compared to monoadducted DNA<sup>22</sup> also indicates that the recognition process is kinetically rather than thermodynamically driven, as it is the trend in productive complex formation barriers that align with the NER repairability and not the stability of the bound complex. Overall, the Rad4 binding barrier

size increasing from monoadducted DNA to diadducted DNA and then to unmodified DNA, matching the repairability trends of these systems,<sup>3, 21, 22</sup> further emphasizes the link between lesion recognition kinetics and overall NER efficiency.

NMR work on <sup>FAAF</sup>dG indicates that the diadducted systems have complex conformational heterogeneities, with multiple conformations being potentially stabilized at each damage site.<sup>20</sup> Therefore, since the monoadducted system indicates that changing the helical conformation results in differing energetic barriers, more in-depth investigations should be done to understand the conformational effects on the NER recognition for diadducted DNA beyond the BB considered in this work (i.e., BS, SB or SS). Other possible avenues for future work could investigate the impact of different sequence contexts on the recognition of DNA lesions, as the surrounding bases at the lesion site are known to significantly impact the conformational heterogeneity.<sup>20, 22, 24, 28, 40, 41</sup> Beyond the changes in the preferred conformations, future work may also investigate the consequences of changing the identify of the second base pair that is disrupted during the DNA opening process (i.e., the pair 5' to the adducted base).

In addition to modelling the Rad4 recognition process, long (1  $\mu$ s) simulations on the prebound (Stage 1) and bound (Stage4) Rad4–DNA complexes were simulated to better understand the impact of number of lesions (undamaged, monoadducted or diadducted) and helical conformation on Rad4 binding and NER efficiency. Notably, previous work on Rad4–adducted DNA prebound complexes, with the DNA duplexes in minor groove or intercalated conformations, suggested that the BHD2 interaction pocket (AlphaSpace) volume in the minor groove and the helical untwist and bending at the lesion sites could be used as a metric for determining NER efficiency.<sup>12</sup> Although the extent of helical

distortions (untwisting and bending) was positively correlated with the determined binding barriers for undamaged and monoadducted DNA, the trend did not extend to the diadducted complex. Furthermore, despite the <sup>AAF</sup>dG adducted DNA in the S conformation being the most kinetically favoured to form a productive bound state and having largest BHD2 interaction (AlphaSpace) volume, the volume of remaining three systems do not align with the calculated Rad4 binding barriers. This suggests that neither the helical distortions nor the extent of BHD2 interactions universal indicators for NER efficiency and care must be taken when predicting the overall NER efficiency using these factors.

As with the prebound state, after Rad4 binding the BHD2 minor groove interaction pocket volume does not reflect the calculated binding barriers, with the monoadducted systems having similar values and the diadducted system being significantly lower. Unlike for the prebound system, the helical untwisting for the bound complex does not reflect the repair susceptibility with the monoadducted DNA in the S conformation and diadducted DNA in the BB conformation having similar twists. Interestingly, the extent of helical bend appears to be related to both the stability of the bound complex (positive correlation) and the calculated barrier (negative correlation). Thus, despite the challenges involved in the modelling of bound Rad4-DNA complexes, determining the lesion site helical bending may be a potential indicator of NER efficiency. Overall, this work highlights that the number of lesions and the exact helical conformation at the damage site have a complex relationship with the Rad4 binding process, and general NER efficiencies.

## 8.5 References

1. Marteiijn, J. A.; Lans, H.; Vermeulen, W.; Hoeijmakers, J. H. J., Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat. Rev. Mol. Cell Biol.* **2014**, *15* (7), 465-481.

2. Helleday, T.; Petermann, E.; Lundin, C.; Hodgson, B.; Sharma, R. A., DNA repair pathways as targets for cancer therapy. *Nat Rev Cancer* **2008**, *8* (3), 193-204.
3. Chen, X.; Velmurugu, Y.; Zheng, G.; Park, B.; Shim, Y.; Kim, Y.; Liu, L.; Van Houten, B.; He, C.; Ansari, A.; Min, J. H., Kinetic gating mechanism of DNA damage recognition by Rad4/XPC. *Nat Commun* **2015**, *6*, 5849.
4. Kumar, N.; Raja, S.; Van Houten, B., The involvement of nucleotide excision repair proteins in the removal of oxidative DNA damage. *Nucleic Acids Research* **2020**, *48* (20), 11227-11243.
5. Kelley, M. R.; Fishel, M. L., *DNA Repair in Cancer Therapy: Molecular Targets and Clinical Applications*. Elsevier Science: 2016.
6. Velmurugu, Y.; Chen, X.; Slogoff Sevilla, P.; Min, J.-H.; Ansari, A., Twist-open mechanism of DNA damage recognition by the Rad4/XPC nucleotide excision repair complex. *Proceedings of the National Academy of Sciences* **2016**, *113* (16), E2296.
7. Bradford, P. T.; Goldstein, A. M.; Tamura, D.; Khan, S. G.; Ueda, T.; Boyle, J.; Oh, K. S.; Imoto, K.; Inui, H.; Moriwaki, S. I.; Emmert, S.; Pike, K. M.; Raziuddin, A.; Plona, T. M.; DiGiovanna, J. J.; Tucker, M. A.; Kraemer, K. H., Cancer and neurologic degeneration in xeroderma pigmentosum: long term follow-up characterises the role of DNA repair. *J. Med. Genet.* **2011**, *48* (3), 168-176.
8. Paul, D.; Mu, H.; Tavakoli, A.; Dai, Q.; Chen, X.; Chakraborty, S.; He, C.; Ansari, A.; Broyde, S.; Min, J. H., Tethering-facilitated DNA 'opening' and complementary roles of beta-hairpin motifs in the Rad4/XPC DNA damage sensor protein. *Nucleic Acids Res* **2020**, *48* (21), 12348-12364.
9. Mu, H.; Geacintov, N. E.; Broyde, S.; Yeo, J. E.; Scharer, O. D., Molecular basis for damage recognition and verification by XPC-RAD23B and TFIIH in nucleotide excision repair. *DNA Repair (Amst)* **2018**, *71*, 33-42.
10. Mu, H.; Geacintov, N. E.; Zhang, Y.; Broyde, S., Recognition of Damaged DNA for Nucleotide Excision Repair: A Correlated Motion Mechanism with a Mismatched cis-syn Thymine Dimer Lesion. *Biochemistry* **2015**, *54* (34), 5263-5267.
11. Mu, H.; Geacintov, N. E.; Min, J.-H.; Zhang, Y.; Broyde, S., Nucleotide Excision Repair Lesion-Recognition Protein Rad4 Captures a Pre-Flipped Partner Base in a Benzo[a]pyrene-Derived DNA Lesion: How Structure Impacts the Binding Pathway. *Chemical Research in Toxicology* **2017**, *30* (6), 1344-1354.
12. Mu, H.; Zhang, Y.; Geacintov, N. E.; Broyde, S., Lesion Sensing during Initial Binding by Yeast XPC/Rad4: Toward Predicting Resistance to Nucleotide Excision Repair. *Chemical Research in Toxicology* **2018**, *31* (11), 1260-1268.
13. Kong, M.; Liu, L.; Chen, X.; Driscoll, K. I.; Mao, P.; Bohm, S.; Kad, N. M.; Watkins, S. C.; Bernstein, K. A.; Wyrick, J. J.; Min, J. H.; Van Houten, B., Single-Molecule Imaging Reveals that Rad4 Employs a Dynamic DNA Damage Recognition Process. *Molecular Cell* **2016**, *64* (2), 376-387.
14. Min, J.-H.; Pavletich, N. P., Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature* **2007**, *449* (7162), 570-575.
15. Lee, Y.-C.; Cai, Y.; Mu, H.; Broyde, S.; Amin, S.; Chen, X.; Min, J.-H.; Geacintov, N. E., The relationships between XPC binding to conformationally diverse DNA adducts and their excision by the human NER system: Is there a correlation? *DNA Repair* **2014**, *19*, 55-63.
16. Pitta, K.; Krishnan, M., Molecular Mechanism, Dynamics, and Energetics of Protein-Mediated Dinucleotide Flipping in a Mismatched DNA: A Computational Study of the RAD4-DNA Complex. *J Chem Inf Model* **2018**, *58* (3), 647-660.



17. Panigrahi, A.; Vemuri, H.; Aggarwal, M.; Pitta, K.; Krishnan, M., Sequence specificity, energetics and mechanism of mismatch recognition by DNA damage sensing protein Rad4/XPC. *Nucleic Acids Res* **2020**, *48* (5), 2246-2257.
18. Paul, D.; Mu, H.; Dai, Q.; Tavakoli, A.; He, C.; Broyde, S.; Min, J.-H., Impact of DNA sequences in the DNA duplex opening by the Rad4/XPC nucleotide excision repair complex. *bioRxiv* **2020**, 2020.01.16.909549.
19. Paul, D.; Mu, H.; Zhao, H.; Ouerfelli, O.; Jeffrey, P. D.; Broyde, S.; Min, J. H., Structure and mechanism of pyrimidine-pyrimidone (6-4) photoproduct recognition by the Rad4/XPC nucleotide excision repair complex. *Nucleic Acids Res* **2019**, *47* (12), 6015-6028.
20. Jain, V.; Hilton, B.; Lin, B.; Jain, A.; Mackerell, A. D.; Zou, Y.; Cho, B. P., Structural and Thermodynamic Insight into Escherichia coli UvrABC-Mediated Incision of Cluster Diacetylaminofluorene Adducts on the NarI Sequence. *Chemical Research in Toxicology* **2013**, *26* (8), 1251-1262.
21. Hilton, B.; Gopal, S.; Xu, L.; Mazumder, S.; Musich, P. R.; Cho, B. P.; Zou, Y., Dissociation Dynamics of XPC-RAD23B from Damaged DNA Is a Determining Factor of NER Efficiency. *PLoS One* **2016**, *11* (6), e0157784.
22. Jain, V.; Hilton, B.; Patnaik, S.; Zou, Y.; Chiarelli, M. P.; Cho, B. P., Conformational and thermodynamic properties modulate the nucleotide excision repair of 2-aminofluorene and 2-acetylaminofluorene dG the NarI sequence. *Nucleic Acids Res.* **2012**, *40* (9), 3939-3951.
23. Sittig's Handbook of Toxic and Hazardous Chemicals and Carcinogens, 4th Edition R. P. Pohanish. Noyes/William Andrew Publishing: Norwich, New York. 2002. 2608 pp. \$495. ISBN: 0-8155-1459-X. *Organic Process Research & Development* **2002**, *6* (5), 740-740.
24. Wilson, K. A.; Kung, R. W.; Wetmore, S. D., Chapter Seven - Toxicology of DNA Adducts Formed Upon Human Exposure to Carcinogens: Insights Gained From Molecular Modeling. In *Advances in Molecular Toxicology*, Fishbein, J. C.; Heilman, J. M., Eds. Elsevier: 2016; Vol. 10, pp 293-360.
25. Cho, B. P.; Zhou, L., Probing the Conformational Heterogeneity of the Acetylaminofluorene-Modified 2'-Deoxyguanosine and DNA by 19F NMR Spectroscopy. *Biochemistry* **1999**, *38* (23), 7572-7583.
26. Donny-Clark, K.; Shapiro, R.; Broyde, S., Accommodation of an N-(Deoxyguanosin-8-yl)-2-acetylaminofluorene Adduct in the Active Site of Human DNA Polymerase  $\alpha$ : Hoogsteen or Watson-Crick Base Pairing? *Biochemistry* **2009**, *48* (1), 7-18.
27. Hingerty, B.; Broyde, S., Conformation of the deoxydinucleoside monophosphate dCpdG modified at carbon 8 of guanine with 2-(acetylamino)fluorene. *Biochemistry* **1982**, *21* (13), 3243-52.
28. Jain, V.; Vaidyanathan, V. G.; Patnaik, S.; Gopal, S.; Cho, B. P., Conformational Insights into the Lesion and Sequence Effects for Arylamine-Induced Translesion DNA Synthesis: 19F NMR, Surface Plasmon Resonance, and Primer Kinetic Studies. *Biochemistry* **2014**, *53* (24), 4059-4071.
29. O'Handley, S. F.; Sanford, D. G.; Xu, R.; Lester, C. C.; Hingerty, B. E.; Broyde, S.; Krugh, T. R., Structural characterization of an N-acetyl-2-aminofluorene (AAF) modified DNA oligomer by NMR, energy minimization, and molecular dynamics. *Biochemistry* **1993**, *32* (10), 2481-97.
30. Grad, R.; Shapiro, R.; Hingerty, B. E.; Broyde, S., A Molecular Mechanics and Dynamics Study of the Minor Adduct between DNA and the Carcinogen 2-(Acetylamino)fluorene (dG-N2-AAF). *Chem. Res. Toxicol.* **1997**, *10* (10), 1123-1132.
31. Broyde, S.; Hingerty, B., An internal fluorene model for iodo-N-2-acetylaminofluorene modified DNA. *Chem.-Biol. Interact.* **1983**, *47* (1), 69-78.

32. Broyde, S.; Hingerty, B. E., DNA backbone conformation in AAF modified dCpdG: variable conformational modes of achieving base displacement. *Chem.-Biol. Interact.* **1982**, *40* (1), 113-19.
33. Patnaik, S.; Cho, B. P., Structures of 2-Acetylaminofluorene Modified DNA Revisited: Insight into Conformational Heterogeneity. *Chemical Research in Toxicology* **2010**, *23* (11), 1650-1652.
34. Shapiro, R.; Sidawi, D.; Miao, Y.-S.; Hingerty, B. E.; Schmidt, K. E.; Moskowitz, J.; Broyde, S., Conformation of Amine-Modified DNA: 2-Aminofluorene- and 2-(Acetylamino)fluorene-Modified Deoxydinucleoside Monophosphates with all Possible Nearest Neighbors. A Comparison of Search and Optimization Methods. *Chemical Research in Toxicology* **1994**, *7* (2), 239-253.
35. Xu, L.; Cho, B. P., Conformational Insights into the Mechanism of Acetylaminofluorene-dG-Induced Frameshift Mutations in the NarI Mutational Hotspot. *Chemical Research in Toxicology* **2016**, *29* (2), 213-226.
36. Zaliznyak, T.; Bonala, R.; Johnson, F.; de los Santos, C., Structure and stability of duplex DNA containing the 3-(deoxyguanosin-N 2-yl)-2-acetylaminofluorene (dG (N 2)-AAF) lesion: a bulky adduct that persists in cellular DNA. *Chemical research in toxicology* **2006**, *19* (6), 745-752.
37. Broyde, S.; Hingerty, B. E.; Shapiro, R.; Norman, D., Unusual hydrogen bonding patterns in 2-aminofluorene (AF) and 2-acetylaminofluorene (AAF) modified DNA. *Environ. Sci. Res.* **1990**, *40* (Nitroarenes), 113-23.
38. Hingerty, B. E.; Broyde, S., Base displacement in AAF and AF modified dCpdG: syn and anti Guanine. *International Journal of Quantum Chemistry* **1982**, *22* (S9), 125-136.
39. Broyde, S.; Hingerty, B., Base displacement in AAF-modified Z-DNA. *Carcinogenesis (London)* **1985**, *6* (1), 151-4.
40. Cho, B. P., Dynamic Conformational Heterogeneities of Carcinogen-DNA Adducts and their Mutagenic Relevance. *Journal of Environmental Science and Health Part C Environmental Carcinogenesis & Ecotoxicology Reviews* **2004**, *22* (2), 57-90.
41. Jain, V.; Hilton, B.; Lin, B.; Patnaik, S.; Liang, F.; Darian, E.; Zou, Y.; MacKerell, A. D., Jr.; Cho, B. P., Unusual Sequence Effects on Nucleotide Excision Repair of Arylamine Lesions: DNA Bending/Distortion as a Primary Recognition Factor. *Nucleic Acids Research* **2013**, *41* (2), 869-880.
42. Millen, A. L.; Sharma, P.; Wetmore, S. D., C8-Linked Bulky Guanosine DNA Adducts: Experimental and Computational Insights into Adduct Conformational Preferences and Resulting Mutagenicity. *Future Medicinal Chemistry* **2012**, *4* (15), 1981-2007.
43. Geacintov, N. E.; Broyde, S., Repair-Resistant DNA Lesions. *Chem Res Toxicol* **2017**, *30* (8), 1517-1548.
44. Manning, T. W.; Al-Abdul-Wahid, M. S.; Manderville, R. A.; Josephy, P. D.; Kung, R. W.; Wetmore, S. D., Structure of an Unusual Tetracyclic Deoxyguanosine Adduct: Implications for Frameshift Mutagenicity of ortho-Cyano Nitroanilines. *Chemical Research in Toxicology* **2020**, *33* (2), 584-593.
45. Berger, F. D.; Sturla, S. J.; Kung, R. W.; Montana, T.; Wetmore, S. D.; Manderville, R. A., Conformational Preference and Fluorescence Response of a C-Linked C8-Biphenyl-Guanine Lesion in the NarI Mutational Hotspot: Evidence for Enhanced Syn Adduct Formation. *Chem Res Toxicol* **2018**, *31* (1), 37-47.
46. Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.;

- Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. *AMBER 2018*, University of California: San Francisco, 2018.
47. Kung, R. W.; Sharma, P.; Wetmore, S. D., Effect of Size and Shape of Nitrogen-Containing Aromatics on Conformational Preferences of DNA Containing Damaged Guanine. *J. Chem. Inf. Model.* **2018**, *58* (7), 1415-1425.
48. Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P., The R.E.D. Tools: Advances in RESP and ESP Charge Derivation and Force Field Library Building. *Phys. Chem. Chem. Phys.* **2010**, *12* (28), 7821-7839.
49. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A., Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *Journal of Molecular Graphics and Modelling* **2006**, *25* (2), 247-260.
50. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **2015**, *11* (8), 3696-3713.
51. Galindo-Murillo, R.; Robertson, J. C.; Zgarbová, M.; Šponer, J.; Otyepka, M.; Jurečka, P.; Cheatham, T. E., Assessing the Current State of Amber Force Field Modifications for DNA. *Journal of Chemical Theory and Computation* **2016**, *12* (8), 4114-4127.
52. Kung, R. W.; Deak, T. K.; Griffith-Salik, C. A.; Takyi, N. A.; Wetmore, S. D., Impact of DNA Adduct Size, Number, and Relative Position on the Toxicity of Aromatic Amines: A Molecular Dynamics Case Study of (AN)dG- and (AP)dG-Containing DNA Duplexes. *J Chem Inf Model* **2021**, *61* (5), 2313-2327.
53. Dennington, R.; Keith, T. A.; Millam, J. M. S. *GaussView, Version 6.0.16*, Semichem Inc.: Shawnee Mission KS, 2016.
54. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T., SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* **2018**, *46* (W1), W296-W303.
55. Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H., PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *Journal of Chemical Theory and Computation* **2011**, *7* (2), 525-537.
56. Grossfield, A. *WHAM: an implementation of the weighted histogram analysis method*, version 2.0.9.1.; [http://membrane.urmc.rochester.edu/wordpress/?page\\_id=126](http://membrane.urmc.rochester.edu/wordpress/?page_id=126).
57. Lavery, R.; Moakher, M.; Maddocks, J. H.; Petkeviciute, D.; Zakrzewska, K., Conformational Analysis of Nucleic Acids Revisited: Curves+. *Nucleic Acids Research* **2009**, *37* (17), 5917-5929.
58. Rooklin, D.; Wang, C.; Katigbak, J.; Arora, P. S.; Zhang, Y., AlphaSpace: Fragment-Centric Topographical Mapping To Target Protein-Protein Interaction Interfaces. *J Chem Inf Model* **2015**, *55* (8), 1585-99.
59. Poater, J.; Swart, M.; Bickelhaupt, F. M.; Fonseca Guerra, C., B-DNA structure and stability: the role of hydrogen bonding, pi-pi stacking interactions, twist-angle, and solvation. *Org Biomol Chem* **2014**, *12* (26), 4691-700.

## Chapter 9: Conclusions<sup>a</sup>

### 9.1 Thesis Overview

How the chemical composition of DNA nucleobase adducts determines the conformational heterogeneity of the damaged helix, and the resulting toxicology, has been previously studied in literature.<sup>1-13</sup> Notably, in conjunction with previous studies,<sup>2, 8, 14-18</sup> this thesis demonstrates the value of computational techniques, including MD simulations and DFT calculations, in understanding the structure of damaged DNA, and how DNA repair proteins interact with adducted DNA. By modelling over 25 distinct lesions in more than 130 unique damaged helices, the work presented in this thesis strengthens past observations<sup>19</sup> on how the size of the bulky group impacts the conformational heterogeneity of the DNA, while unveiling novel trends in how the bulky group shape, nucleobase–aryl moiety linker type, location of functional groups on the aryl moiety, number of lesions and lesion separation of diadducted DNA determines the structure of adducted DNA. Additionally, this thesis highlights the impact of helical conformation and number of lesions on the NER recognition process of the model <sup>AA</sup>FdG bulky DNA adduct. Overall, this thesis contributes to the understanding on how the chemical composition of the DNA adduct effects the structure and biological consequences of damaged DNA.

Chapter 2 used a multiscale modelling approach to examine a systematic set of seven N-linked C8-dG DNA adducts that vary in bulky moiety shape and size from one (<sup>A</sup>NdG) to four (<sup>A</sup>PdG) aryl rings. Specifically, DFT nucleobase and nucleoside models were used to probe the nucleobase–carcinogen and nucleobase–sugar linkages. Notably, this showed that the shape of the bulky group (i.e., the presence/absence of an  $\alpha$ -fused ring) has

---

<sup>a</sup>The *Journal of American Chemical Society* Reference Style is used in this Chapter.

a significant effect on the preferred conformation of the nucleobase–carcinogen linker. For all seven lesions, MD simulations of the fully paired adducted duplexes in the major groove (B), intercalated stacked (S) and wedge (W) conformations were generated. This work highlighted that the W conformation is generally unfavourable, and that the exact B:S (*anti:syn*) conformational ratio is dependent on the specific adduct, with the smallest <sup>AN</sup>dG adduct having the greatest propensity for the B conformation, the largest <sup>AP</sup>dG adduct having the greatest propensity for the S conformation, and the remaining five lesions falling between the two extremes. Overall, in addition to strengthening previously suggested trends in how the bulky group size impacts conformational heterogeneity,<sup>19</sup> this work was the first to consider the impact of bulky group shape to the growing list of factors that determine the conformation of adducted DNA. Beyond providing insights to better predict the potential toxicity of novel bulky DNA adducts, this understanding of the factors that determine mutagenicity can be used to design more efficient anticancer drugs with more cellular resistance than the current generation treatments possess.<sup>38, 39</sup>

In order to understand the extreme mutagenic differences experimentally observed for two N-linked (*para*- or *ortho*-substituted) cyano-aniline adducts arising from components found in azo dyes,<sup>20</sup> DFT calculations and MD simulations were implemented. The *ortho*-substituted cyano-aniline adduct is unique as it was suggested to undergo a cyclization reaction to form a fused tetracyclic C8,N7-dG adduct.<sup>20</sup> Thus, the cyclized *ortho*-lesion was shown to have greater stacking potential in both the fully paired and deletion (SMI) duplexes than the noncyclic *para*-cyano aniline adduct. Consequently, the cyclic variant is predicted to stabilize a promutagenic *syn* glycosylic orientation in nucleosides, fully paired duplexes and deletion duplexes, whereas the linear *para*-isomer

stabilizes the nonmutagenic *anti* orientation. In summary, this work highlights how the addition of a functional group to the bulky moiety can significantly impact the preferred helical structure and the resulting mutagenicity of C8-dG adducts through a novel cyclization reaction.<sup>20</sup>

Chapter 4 further probes the factors that determine the helical conformation of adducted DNA by considering 14 unique DNA adducts. Specifically, the computational methodology developed in Chapter 2 was applied to C-linked and O-linked derivatives of the seven model N-linked C8-dG adducts previously studied, with the goal of understanding how the nucleobase–aryl moiety linker type effects the conformational preference of the adducted DNA. Furthermore, this selection of lesions allows us to further consider the impact of bulky aryl moiety shape and size on the conformational heterogeneity. Within the DFT nucleobase and nucleoside models, the linker type as well as the previously noted aryl moiety shape was shown to impact the preferred orientation of the aryl group with respect to the nucleobase, with the C-linked lesions being less flexible and the O-linked versions stabilizing fewer minima than the N-linked variants. Notably, structural differences are also seen for the helical models, which have a change in the conformational heterogeneity depending on the linker type, such that the N-linked and O-linked lesions have a greater *syn* preference than the C-linked adducts. Overall, out of the B, S and W conformations considered in the study, only the B and S conformations were suggested to be stabilized for any given lesion, with the W conformation generally being unfavoured. Despite this, a greater W preference for the O-linked adducts can arise due to a lack of bonding donor on the linking atom. This aligns with previous work on the N-linked <sup>AAF</sup>dG adduct, which has an acetyl group on the linking atom and often stabilizes

the W conformation.<sup>6, 21</sup> When coupled with Chapter 2, the results from this chapter highlight that the aryl moiety shape has minimal impact on the adopted helical conformation of adducted DNA, and thus the resulting biological outcomes, compared to the aryl moiety size and type of nucleobase–aryl group linker. Overall, this work used an incremental series of 21 model adducts to perform the first systematic investigation of linker type to date. This information could be applied in future studies to probe the interactions between adducted DNA and DNA-interacting proteins by choosing adducts, or rationally design new lesions, that induce a desired conformational heterogeneity in DNA.

Chapter 5 examines two F-labelled C-linked C8-dG adducts, namely <sup>FPh</sup>dG and <sup>FBP</sup>dG, using both <sup>19</sup>F-NMR and MD simulations. These lesions highlight the effect of aryl group size on the preferred DNA helical conformations by revealing a large shift from an *anti* (B) orientation (94% for <sup>FPh</sup>dG) to a *syn* (I) orientation (80% for <sup>FBP</sup>dG) upon the substitution of a phenyl ring in the *para* position. In fact, for both the fully paired and SMI deletion complexes, the biphenyl <sup>FBP</sup>dG adduct has a significantly greater *syn* preference than the single-ringed <sup>FPh</sup>dG. Through comparisons to previous literature, this work also further highlights how the linkage type can impact the conformational heterogeneity. Specifically, the N-linked biphenyl equivalent (<sup>ABP</sup>dG) has an *anti:syn* ratio of 67:33<sup>6</sup> compared to the determined 20:80 ratio for <sup>FBP</sup>dG in a similar sequence context. Consequently, this study exemplifies the significance of both aryl moiety size and nucleobase–aryl linkage type on determining the structural preferences and biological outcomes of damaged DNA.

Despite the amount of work done previously,<sup>2, 4, 6, 9, 10, 12, 13, 16, 18, 19, 21, 22</sup> and in this thesis, to understand the helical structures of adducted DNA, the vast majority of this work

has exclusively studied DNA containing a single lesion (monoadducted DNA). Hence, Chapters 6 and 7 attempt to broaden our understanding of the impact of formation of DNA adducts by providing the first in-depth structural investigations into the helical conformational preferences of short sections of DNA containing two lesions (cluster adducts or diadducted DNA). Specifically, MD simulations were used to probe the structures of <sup>AN</sup>dG (Chapter 6), <sup>AP</sup>dG (Chapter 6) or <sup>ABP</sup>dG (Chapter 7) monoadducted and diadducted DNA containing the damaged *NarI* recognition sequence (5'–G<sub>1</sub>G<sub>2</sub>CG<sub>3</sub>CC). These lesions were chosen as representative adducts that span the range of lesion size (one, two or four aryl rings), mutagenicity (nonmutagenic, probably carcinogenic and carcinogenic precursors) and induced adducted duplex conformations (predominantly B, heterogenetic mixture of B and S, and predominantly S). In these chapters, monoadducted DNA containing a <sup>AN</sup>dG, <sup>AP</sup>dG or <sup>ABP</sup>dG adduct in the *anti* or *syn* orientations at the G<sub>1</sub>, G<sub>2</sub> or G<sub>3</sub> site of the *NarI* sequence were modelled along with diadducted DNA containing two <sup>AN</sup>dG, <sup>AP</sup>dG or <sup>ABP</sup>dG adducts at the G<sub>1</sub> and G<sub>2</sub>, G<sub>1</sub> and G<sub>3</sub>, or G<sub>2</sub> and G<sub>3</sub> sites in either the *anti/anti*, *anti/syn*, *syn/anti* or *syn/syn* orientations. Monoadducted DNA possessed conformational homogeneity for the single-ringed <sup>AN</sup>dG (the nonmutagenic B conformation) and the four-ringed <sup>AP</sup>dG (the mutagenic S conformation). Comparatively, the <sup>ABP</sup>dG monoadducted systems were determined to have conformational heterogeneity, with the exact *anti:syn* ratio depending on the sequence context. For all adducts, upon addition of a second damage site, the preferred helical conformations shift, with the one-nucleotide removed (G<sub>2</sub>G<sub>3</sub>) diadducts stabilizing the *syn* adduct orientations, whereas the neighboring (G<sub>1</sub>G<sub>2</sub>) sequence increases the stability of the *anti* orientation. Thus, the <sup>AN</sup>dG lesion could become more mutagenic in diadducted DNA due to the stabilization of the promutagenic *syn* orientation (G<sub>2</sub><sup>*syn*</sup>G<sub>3</sub><sup>*syn*</sup>), whereas <sup>AP</sup>dG may be less mutagenic due to *anti*



becoming stabilized in diadducted DNA ( $G_1^{anti}G_2^{syn}$ ). In contrast, when the lesions are separated by two base pairs, the preferred conformations at both damage sites align with the corresponding monoadducted helices, suggesting that separation by even greater distances will not alter the monoadducted DNA conformations in the same sequence context. Consequently, this work uncovers the importance of lesion separation on resulting toxicity, with the separation distance increasing or decreasing the mutagenicity depending on the exact distance between the damage sites. Overall, these chapters add the number of lesions to the list of factors that are known to impact conformational preferences. Importantly, this work emphasizes the need for future studies in order to understand how diadducted DNA impacts the formation of cancer and other diseases.

Chapter 8 examines the GG-NER recognition pathway for  $^{AAF}dG$  adducted DNA using MD simulations in conjunction with advanced umbrella sampling. The initial binding of the Rad4 recognition protein was modelled over four US stages for DNA duplexes containing the model  $^{AAF}dG$  adduct in the B (monoadducted), S (monoadducted) or BB (diadducted) conformations, as well as undamaged DNA. The calculated energetic barrier for the rate determining step of the binding process matches those reported for DNA containing a different adduct (*cis*-B[a]P-dG)<sup>11</sup> or a photodimer (CPD),<sup>23</sup> and was found to be smallest for monoadducted DNA in the S conformation, followed by monoadducted DNA in the B conformation, and diadducted DNA in the BB conformation, and was the highest for canonical DNA. Notably, the barriers align with the known NER efficiency, with monoadducted DNA being more repair prone than diadducted DNA, and undamaged DNA generally not being bound by Rad4. Thus, this work provides unique insight into the “kinetic gating” mechanism<sup>24</sup> that is critical for understanding the complex NER pathway.

Additionally, unrestrained MD simulations of the prebound and bound complexes suggested that the helical bending of the Rad4–DNA bound systems is the best indicator of NER efficiency compared to the helical twist or the pocket volume of protein–DNA minor groove interactions, which were previously suggested factors for other DNA lesions.<sup>8</sup> Importantly, this is the first work to model the recognition process for a major groove lesions, and the first to study the mechanism for diadducted DNA, highlighting that major groove lesions and diadducted DNA have different interactions with the recognition protein and, therefore, require different indicators for NER susceptibility than was previously suggested.<sup>8</sup> Overall, this thesis successfully uses computational methods to fill voids in the literature by revealing novel trends in how the chemical composition of the adducted DNA impact the resulting helical conformations and the lesion repairability, as well as the potential mutagenic consequences.

## 9.2 Future Work

The multiscale DFT and MD computational approaches developed and implemented in this thesis could be applied to other bulky DNA adducts to elucidate the most probable helical conformations and thereby gain insight into their potential mutagenic effects. As highlighted by Chapters 2 and 4 of this thesis, other model adducts can be systematically studied to better understand the impact of other lesion chemical features, such as the presence and location of heteroatoms on the overall helical structure and biological outcomes. Notably, Chapter 3 highlights the vast mutagenic differences from varying the location (*ortho* or *para*) of an aryl-substituted functional group. Furthermore, the unsubstituted phenol agent (i.e., precursor to <sup>O</sup>PhdG), is not classifiable as carcinogenic to humans (group 3), whereas the chlorine-substituted pentachlorophenyl agent, (i.e.,

precursor to <sup>PCP</sup>dG), is carcinogenic to humans.<sup>25</sup> Therefore, a potential project would be examining the systematic addition of chlorine atoms to <sup>OPh</sup>dG to determine how the number and location of chlorine atoms on the aryl moiety influence the helical conformation and resulting mutagenicity. Another structural factor that has been found to influence the mutagenic potential of DNA damage is the carcinogen–nucleobase binding site. This thesis focuses on DNA adducts formed at C8-dG, although some damaging agents (e.g., AAF and AP)<sup>18, 26</sup> can also form adducts at different locations with potentially different levels of toxicity. For example, aristolochic acids can form adducts at both the N2-dG and N6-dA sites (Figure 9.1), with the A lesion being more mutagenic in cells than the G derivative.<sup>27</sup> Unfortunately, limited studies have systematically explored the effect on adduct structure (e.g., aryl group shape and size) on helical conformation and biological outcome for different carcinogen–nucleobase binding sites. Thus, future studies could expand our knowledge by systematically comparing lesions as done in Chapter 2 for DNA adducts formed at other sites, including N2-dG, N6-dA, N4-dC and O2-dT. Another helical feature that is well known to impact the conformational heterogeneity of adducted DNA is the

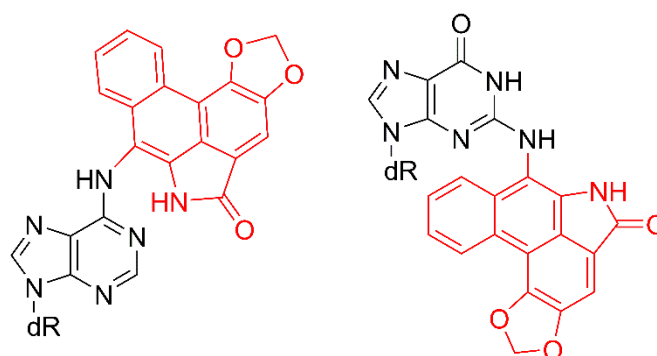


Figure 9.1: Structures of the N6-dA (left) and N2-dG (right) adducts derived from aristolochic acids. The bulky group is shown in red.

sequence context of the damage site, with not only the flanking base pairs changing the conformational preferences, but also the next nearest base pairs.<sup>6</sup> Perhaps unsurprisingly, certain sequences are more mutation prone than others.<sup>28, 29</sup> Consequently, in-depth investigations of the sequence context effects on the conformational heterogeneity could provide valuable information into the mutation susceptibility of different genes.

In addition to systematic studies to probe the effect of specific adduct features, other work could apply the computational methodology developed on identified DNA adducts in which the structure of the damaged duplex is not known. For example, lucidin, a mutagen found in the madder root that has been used in red dyes, forms adducts at the N2-dG and N6-dA sites, although the helical conformations of these lesions are currently not known.<sup>30</sup> Another compound used in dyes is *o*-toluidine, which is carcinogenic to humans (group 1).<sup>31</sup> Despite the toxicity of this agent, there are no studies investigating the helical structure of the N2-dG, N7-dG and C8-dG adducts formed by the *o*-toluidine precursor in cells.<sup>31</sup> Moreover, new potentially mutagenic lesions are being regularly identified, such as the recently detected N-6-HPB-dAdo lesion, which is formed from *N'*-nitrosonornicotine (NNN) in tobacco.<sup>32</sup> Notably, such structural studies are useful in the classification of compounds as carcinogenic, allowing for stricter government regulations to limit human exposure.

Beyond examining isolated DNA models, computational methods have been used to understand the interactions between adducted DNA and replication,<sup>33-35</sup> repair<sup>8, 11</sup> or DNA folding proteins (i.e., histones).<sup>36, 37</sup> Importantly, understanding how proteins replicate and repair damaged DNA could allow for the development of new anticancer drugs with increased potency, as current therapeutics suffer from drug resistance.<sup>38, 39</sup>

Although the interactions of adducted DNA with bypass polymerases have been studied for several DNA lesions,<sup>33-35</sup> the factors that determine what TLS polymerases are responsible for the bypass and extension at any given adducted site are not well known and could be probed using computational methods, such as the MD techniques used in this thesis. Furthermore, insight into the conformational preferences of the adducted DNA duplexes modeled in this thesis would provide a valuable information on the relevant lesion orientations for future polymerase models. Interestingly, although RNA polymerases are also known to stall at adducted DNA sites,<sup>40-42</sup> there are relatively few studies, such as Perlow *et al.*<sup>43</sup>, that examine how the structure of the adducted DNA impacts transcription outcomes, with most work focusing on DNA polymerases due to the longer term consequences of mutations. Thus, future work could be geared towards understanding how the chemical structure of the adduct, and the resulting damaged duplex, can impact protein synthesis.

Beyond impacting the DNA replication and transcription, the structure of adducted DNA also determines the reparability of the lesion. Unfortunately, the NER recognition of monoadducted DNA in the GG-NER pathway is still poorly understood. Despite Chapter 8 examining the Rad4 binding pathway of a major groove lesion for the first time, adducted DNA in the W conformation has yet to be considered. Despite W being a minor conformation for <sup>AAF</sup>dG adducted DNA in the sequence contexts that have been studied,<sup>44, 45</sup> uncovering the recognition pathway for damaged DNA in the W conformation is necessary to fully understand the repair of <sup>AAF</sup>dG adducted DNA in a given sequence context due to the induced conformational heterogeneity. Notably, since the <sup>AAF</sup>dG lesion is known to have varied NER efficiency depending on the lesion site sequence context (i.e.,

the G<sub>1</sub>, G<sub>2</sub> or G<sub>3</sub> site of *NarI*) and the resulting differing conformational heterogeneities,<sup>7</sup> it would be of interest to model the remaining sequences and helical conformations to gain a better understanding of the structural features that dictate the repairability of a given site. Specifically, by systematically comparing the *NarI* monoadducted <sup>AAF</sup>dG systems, the extent that the repairability of the damage can be attributed to the structural features unique to the specific sequence context (e.g., helical twist and the strength of stacking interactions at lesion site) rather than changes in the conformational preferences (i.e., B/S/W ratio) can be ascertained. Beyond the changes in the preferred conformations, future work may also investigate the consequences of changing the identity of the second base pair that is disrupted during the DNA opening process. In addition to monoadducted DNA, NMR work on <sup>FAAF</sup>dG indicates that the diadducted systems have complex conformational heterogeneities, with multiple conformations being potentially stabilized at each damage site.<sup>30</sup> Therefore, since the monoadducted system indicates that changing the helical conformation results in differing energetic barriers, more in-depth investigations should be done to understand the conformational effects on the NER recognition for diadducted DNA. Further evidence to support the conformational and sequence dependence on the repair efficiency gained from a comprehensive study of <sup>AAF</sup>dG recognition could be confirmed by considering the Rad4 binding pathway of similar bulky C8-dG lesions, such as <sup>AF</sup>dG or <sup>ABP</sup>dG.

In addition to impacting the repair and replication of DNA adducts, the adduct structure and resulting helical conformation likely influence the properties of nucleosomes. Although the presence of a bulky DNA adduct can impact nucleosome folding of damaged DNA around the histone core,<sup>36, 37</sup> relatively little is known about the impact of the helical

conformation and position of the adduct on the overall DNA folding within the nucleosome. Importantly, the presence of adducts can affect histone post-translational modifications,<sup>36, 37</sup> potentially modifying cellular functions (e.g., transcription and repair). Thus, determining adducted DNA–histone interactions could reveal unique insights into the associated biological outcomes. Notably, future work could use a systematic approach to consider how adduct structure, helical conformation and the position of the damage site with respect to the histone core particle impact the ability of DNA to tightly fold into chromatin, thus changing the cellular interactions with the DNA.

In addition to studying isolated damage events with a duplex, further work to understand cluster DNA adducts is required. At the simplest level, studies using the methodologies developed in Chapters 6 and 7 could be implemented to understand how DNA containing other adducts (e.g., <sup>AF</sup>dG) differ upon addition of a second damage site, thus strengthening the lesion separation effects uncovered in this work. Additionally, to date, only diadducted DNA containing two of the same adduct have been studied, whereas it is plausible that two different nucleobase adducts that arise from the same source could form in close proximity. For example, it would be of interest to consider how the presence of an N2-dG and a C8-dG AAF-derived adducts<sup>22</sup> in close proximity impact the conformational preferences and resulting mutagenicity compared to the independent monoadducted DNA. Furthermore, more work should be done to understand how the presence of a second damage site impacts the TLS replication process. This is of specific interest as diadducted DNA may have the bulky group in different positions with respect to the helix (i.e., the major groove and interhelical space for the BS conformation of <sup>AP</sup>dG diadducted DNA at the G<sub>1</sub>G<sub>2</sub> sites in Chapter 6); however, TLS polymerases often have an

open pocket on one side of DNA,<sup>46</sup> allowing for a bulky group to only be accommodated in a specific way at the active site. Lastly, for cluster lesions, it would be interesting to consider how the presence of a nearby common non-adduct damage site (e.g., 8OG or abasic sites) that are commonly present in DNA may impact the helical structure of the adducted DNA.

Finally, similar to DNA adducts being used as potential probes for specific DNA environments (e.g., SMI complexes in Chapter 5), DNA adducts could be incorporated into nonconventional nucleic acid systems. For example, DNA origami, which is a form of nanotechnology that involves using DNA molecules that self-assemble into specifically shaped nanoparticles, has become increasingly popular, having potential applications in various fields including cancer diagnosis and drug delivery.<sup>47, 48</sup> Beyond the use of the canonical DNA nucleotides, modifications can be added to increase functionality of the new DNA shapes.<sup>49</sup> By selectively incorporating DNA adducts into DNA origami, structural features including the local helical twist, bend and nonbonded interactions can be modified to the designers' specifications by adjusting the chemical structure of the bulky moiety. Another increasingly popular, nonconventional use of DNA is the synthetic design of DNA aptamers.<sup>50</sup> DNA aptamers are short oligomeric sections that are constructed to bind to a specific target with high affinity and selectivity. DNA modifications, including the addition of bulky DNA adducts, have been used to increase the specificity of DNA aptamers.<sup>51-55</sup> Although adducts can be incorporated into aptamers to finely tune binding preferences, the inclusion of various systematic lesions, like the 21 model adducts studied in Chapters 2 and 4, can be done during the initial design process to better sample different combinations of modifications within the aptamer. Overall, although this thesis provides



novel insight into the structural factors that determine the conformations of adducted DNA, the resulting interactions with repair proteins, and the lesion mutagenicity, further work is needed to fully understand the complex interplay of factors that determine helical structure and biological outcomes of DNA adducts and their potential uses in synthetic nucleic acids systems.

### **9.3 Final Remarks**

In conclusion, this thesis provides unique insight into the structural factors that determine the conformation of adducted DNA and thus the resulting biological consequences of this form of damage. This work reveals the significance of bulky group shape and size, the presence and position of functional groups on the aryl moiety, and the type of linkage connecting the aryl carcinogen to the nucleobase on the resulting helical structure. Furthermore, this thesis presents the first in-depth structural studies on diadducted DNA, revealing a previously unknown lesion separation effect for the mutagenicity of damaged duplexes. Moreover, the Rad4–DNA binding process of monoadducted and diadducted is probed, confirming the dependence of the NER efficiency on the helical conformation and number of lesions, providing novel insight into the “kinetic gating” recognition mechanism,<sup>24</sup> and showcasing a novel structural feature (i.e., the extent of helical bending of the protein-bound DNA) that can be used as an indicator of overall NER recognition efficiency. The future research avenues outlined in this chapter would even further advance our understanding of how the chemical composition of the adduct allows for the determination of the helical structure and biological outcomes. Research that highlights the mutagenic pathway of DNA adducts could be used to establish government regulations for compounds in which the human carcinogenicity is not currently fully

documented. Additionally, understanding how cells tolerate damaged DNA could allow for the development of novel therapeutics that have greater efficacy due to a decrease in anticancer damage repair and replication bypass. Overall, this thesis underscores the complex interplay between adduct chemical structure, the induced DNA conformational heterogeneity, and the resulting DNA replication and repair outcomes.

## 9.4 References

1. Ignatov, A. V.; Bondarenko, K. A.; Makarova, A. V., Non-bulky Lesions in Human DNA: The Ways of Formation, Repair, and Replication. *Acta Naturae* **2017**, 9 (3), 12-26.
2. Wilson, K. A.; Kung, R. W.; Wetmore, S. D., Chapter Seven - Toxicology of DNA Adducts Formed Upon Human Exposure to Carcinogens: Insights Gained From Molecular Modeling. In *Advances in Molecular Toxicology*, Fishbein, J. C.; Heilman, J. M., Eds. Elsevier: 2016; Vol. 10, pp 293-360.
3. Hoeijmakers, J. H. J., DNA Damage, Aging, and Cancer. *New England Journal of Medicine* **2009**, 361 (15), 1475-1485.
4. Millen, A. L.; Sharma, P.; Wetmore, S. D., C8-Linked Bulky Guanosine DNA Adducts: Experimental and Computational Insights into Adduct Conformational Preferences and Resulting Mutagenicity. *Future Med. Chem.* **2012**, 4 (15), 1981-2007.
5. Broyde, S.; Wang, L.; Rechkoblit, O.; Geacintov, N. E.; Patel, D. J., Lesion Processing: High-Fidelity Versus Lesion-Bypass DNA Polymerases. *Trends in Biochemical Sciences* **2008**, 33 (5), 209-219.
6. Jain, V.; Vaidyanathan, V. G.; Patnaik, S.; Gopal, S.; Cho, B. P., Conformational Insights into the Lesion and Sequence Effects for Arylamine-Induced Translesion DNA Synthesis: 19F NMR, Surface Plasmon Resonance, and Primer Kinetic Studies. *Biochemistry* **2014**, 53 (24), 4059-4071.
7. Hilton, B.; Gopal, S.; Xu, L.; Mazumder, S.; Musich, P. R.; Cho, B. P.; Zou, Y., Dissociation Dynamics of XPC-RAD23B from Damaged DNA Is a Determining Factor of NER Efficiency. *PLoS One* **2016**, 11 (6), e0157784.
8. Mu, H.; Zhang, Y.; Geacintov, N. E.; Broyde, S., Lesion Sensing during Initial Binding by Yeast XPC/Rad4: Toward Predicting Resistance to Nucleotide Excision Repair. *Chem. Res. Toxicol.* **2018**, 31 (11), 1260-1268.
9. Yang, S.; Liu, J. D. H.; Diem, M.; Wesseling, S.; Vervoort, J.; Oostenbrink, C.; Rietjens, I., Molecular Dynamics and In Vitro Quantification of Safrole DNA Adducts

Reveal DNA Adduct Persistence Due to Limited DNA Distortion Resulting in Inefficient Repair. *Chem. Res. Toxicol.* **2020**, *33* (9), 2298-2309.

10. Yang, S.; Diem, M.; Liu, J. D. H.; Wesseling, S.; Vervoort, J.; Oostenbrink, C.; Rietjens, I., Cellular levels and molecular dynamics simulations of estragole DNA adducts point at inefficient repair resulting from limited distortion of the double-stranded DNA helix. *Arch. Toxicol.* **2020**, *94* (4), 1349-1365.

11. Mu, H.; Geacintov, N. E.; Min, J.-H.; Zhang, Y.; Broyde, S., Nucleotide Excision Repair Lesion-Recognition Protein Rad4 Captures a Pre-Flipped Partner Base in a Benzo[a]pyrene-Derived DNA Lesion: How Structure Impacts the Binding Pathway. *Chem. Res. Toxicol.* **2017**, *30* (6), 1344-1354.

12. Wilson, K. A.; Wetmore, S. D., Complex conformational heterogeneity of the highly flexible O6-benzyl-guanine DNA adduct. *Chem. Res. Toxicol.* **2014**, *27* (7), 1310-1325.

13. Zhang, L.; Shapiro, R.; Broyde, S., Molecular Dynamics of a Food Carcinogen–DNA Adduct in a Replicative DNA Polymerase Suggest Hindered Nucleotide Incorporation and Extension. *Chem. Res. Toxicol.* **2005**, *18* (9), 1347-1363.

14. Mu, H.; Geacintov, N. E.; Broyde, S.; Yeo, J. E.; Scharer, O. D., Molecular basis for damage recognition and verification by XPC-RAD23B and TFIIH in nucleotide excision repair. *DNA Repair (Amst)* **2018**, *71*, 33-42.

15. Cai, Y.; Patel, D. J.; Geacintov, N. E.; Broyde, S., Dynamics of a Benzo[a]pyrene-Derived Guanine DNA Lesion in TGT and CGC Sequence Contexts: Enhanced Mobility in TGT Explains Conformational Heterogeneity, Flexible Bending, and Greater Susceptibility to Nucleotide Excision Repair. *J. Mol. Biol.* **2007**, *374* (2), 292-305.

16. Cai, Y.; Ding, S.; Geacintov, N. E.; Broyde, S., Intercalative Conformations of the 14R (+)- and 14S (-)-trans-anti-DB[a,l]P-N6-dA Adducts: Molecular Modeling and MD Simulations. *Chem. Res. Toxicol.* **2011**, *24* (4), 522-531.

17. Shapiro, R.; Sidawi, D.; Miao, Y.-S.; Hingerty, B. E.; Schmidt, K. E.; Moskowitz, J.; Broyde, S., Conformation of Amine-Modified DNA: 2-Aminofluorene- and 2-(Acetylamino)fluorene-Modified Deoxydinucleoside Monophosphates with all Possible Nearest Neighbors. A Comparison of Search and Optimization Methods. *Chem. Res. Toxicol.* **1994**, *7* (2), 239-253.

18. Grad, R.; Shapiro, R.; Hingerty, B. E.; Broyde, S., A Molecular Mechanics and Dynamics Study of the Minor Adduct between DNA and the Carcinogen 2-(Acetylamino)fluorene (dG-N2-AAF). *Chem. Res. Toxicol.* **1997**, *10* (10), 1123-1132.

19. Shapiro, R.; Ellis, S.; Hingerty, B. E.; Broyde, S., Effect of Ring Size on Conformations of Aromatic Amine-DNA Adducts: The Aniline-C8 Guanine Adduct Resides in the B-DNA Major Groove. *Chem. Res. Toxicol.* **1998**, *11* (4), 335-341.

20. Josephy, P. D.; Dhanoa, J.; Elzawy, G.; Heney, K.; Petrie, L.; Senis, C., Structure–activity investigation of the potentiating effect of cyano substitution on nitroaniline mutagenicity in the ames test. *Environmental and Molecular Mutagenesis* **2018**, 59 (2), 114-122.
  21. Cho, B. P.; Zhou, L., Probing the Conformational Heterogeneity of the Acetylaminofluorene-Modified 2'-Deoxyguanosine and DNA by <sup>19</sup>F NMR Spectroscopy. *Biochemistry* **1999**, 38 (23), 7572-7583.
  22. Isgro, T. A.; Mathew, N.; Topper, R. Q., Structural characterization of N-acetyl-2-aminofluorene (AAF) adducts to guanine and deoxyguanosine via a molecular mechanics, semi-empirical, and density functional theory cascade. *Journal of Molecular Structure: THEOCHEM* **2004**, 710 (1), 31-43.
  23. Mu, H.; Geacintov, N. E.; Zhang, Y.; Broyde, S., Recognition of Damaged DNA for Nucleotide Excision Repair: A Correlated Motion Mechanism with a Mismatched cis-syn Thymine Dimer Lesion. *Biochemistry* **2015**, 54 (34), 5263-5267.
  24. Velmurugu, Y.; Chen, X.; Slogoff Sevilla, P.; Min, J.-H.; Ansari, A., Twist-open mechanism of DNA damage recognition by the Rad4/XPC nucleotide excision repair complex. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, 113 (16), E2296.
  25. Humans, I. W. G. o. t. E. o. C. R. t., IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. In *Pentachlorophenol and Some Related Compounds*, International Agency for Research on Cancer
- © International Agency for Research on Cancer, 2019. For more information contact [publications@iarc.fr](mailto:publications@iarc.fr): Lyon (FR), 2019.
26. Vyas, R.; Efthimiopoulos, G.; Tokarsky, E. J.; Malik, C. K.; Basu, A. K.; Suo, Z., Mechanistic Basis for the Bypass of a Bulky DNA Adduct Catalyzed by a Y-Family DNA Polymerase. *J. Am. Chem. Soc.* **2015**, 137 (37), 12131-42.
  27. Kathuria, P.; Sharma, P.; Wetmore, S. D., Adenine versus guanine DNA adducts of aristolochic acids: role of the carcinogen-purine linkage in the differential global genomic repair propensity. *Nucleic Acids Res.* **2015**, 43 (15), 7388-7397.
  28. Roy, A.; Fuchs, R. P. P., Mutational spectrum induced in *Saccharomyces cerevisiae* by the carcinogen N-2-acetylaminofluorene. *Mol. Gen. Genet.* **1994**, 245 (1), 69-77.
  29. Fuchs, R. P. P., DNA binding spectrum of the carcinogen N-acetoxy-N-2-acetylaminofluorene significantly differs from the mutation spectrum. *J. Mol. Biol.* **1984**, 177 (1), 173-180.
  30. Ishii, Y.; Inoue, K.; Takasu, S.; Jin, M.; Matsushita, K.; Kuroda, K.; Fukuhara, K.; Nishikawa, A.; Umemura, T., Determination of lucidin-specific DNA adducts by

liquid chromatography with tandem mass spectrometry in the livers and kidneys of rats given lucidin-3-O-primeveroside. *Chem. Res. Toxicol.* **2012**, 25 (5), 1112-8.

31. Tajima, Y.; Toyoda, T.; Hirayama, Y.; Matsushita, K.; Yamada, T.; Ogawa, K.; Watanabe, K.; Takamura-Enya, T.; Totsuka, Y.; Wakabayashi, K.; Miyoshi, N., Novel o-Toluidine Metabolite in Rat Urine Associated with Urinary Bladder Carcinogenesis. *Chem. Res. Toxicol.* **2020**, 33 (7), 1907-1914.

32. Li, Y. P.; Hecht, S. S., Identification of an N'-Nitrosornicotine-Specific Deoxyadenosine Adduct in Rat Liver and Lung DNA. *Chem. Res. Toxicol.* **2021**, 34 (4), 992-1003.

33. Wilson, K. A.; Wetmore, S. D., Molecular Insights into the Translesion Synthesis of Benzyl-Guanine from Molecular Dynamics Simulations: Structural Evidence of Mutagenic and Nonmutagenic Replication. *Biochemistry* **2017**, 56 (13), 1841-1853.

34. Wilson, K. A.; Garden, J. L.; Wetmore, N. T.; Felske, L. R.; Wetmore, S. D., DFT and MD Studies of Formaldehyde-Derived DNA Adducts: Molecular-Level Insights into the Differential Mismatching Potentials of the Adenine, Cytosine, and Guanine Lesions. *The Journal of Physical Chemistry A* **2019**, 123 (29), 6229-6240.

35. Chandani, S.; Loechler, E. L., Molecular Modeling Benzo [a] pyrene N 2-dG Adducts in the Two Overlapping Active Sites of the Y-Family DNA Polymerase Dpo4. *Journal of Molecular Graphics and Modelling* **2007**, 25 (5), 658-670.

36. Fu, I.; Cai, Y.; Zhang, Y.; Geacintov, N. E.; Broyde, S., Entrapment of a Histone Tail by a DNA Lesion in a Nucleosome Suggests the Lesion Impacts Epigenetic Marking: A Molecular Dynamics Study. *Biochemistry* **2016**, 55 (2), 239-242.

37. Cai, Y.; Wang, L.; Ding, S.; Schwaid, A.; Geacintov, N. E.; Broyde, S., A Bulky DNA Lesion Derived from a Highly Potent Polycyclic Aromatic Tumorigen Stabilizes Nucleosome Core Particle Structure. *Biochemistry* **2010**, 49 (46), 9943-9945.

38. Helleday, T.; Petermann, E.; Lundin, C.; Hodgson, B.; Sharma, R. A., DNA repair pathways as targets for cancer therapy. *Nat. Rev. Cancer* **2008**, 8 (3), 193-204.

39. Alt, A.; Lammens, K.; Chiocchini, C.; Lammens, A.; Pieck, J. C.; Kuch, D.; Hopfner, K. P.; Carell, T., Bypass of DNA lesions generated during anticancer treatment with cisplatin by DNA polymerase  $\eta$ . *Science* **2007**, 318 (5852), 967-70.

40. Donahue, B. A.; Fuchs, R. P. P.; Reines, D.; Hanawalt, P. C., Effects of aminofluorene and acetylaminofluorene DNA adducts on transcriptional elongation by RNA polymerase II. *J. Biol. Chem.* **1996**, 271 (18), 10588-10594.

41. Choi, D. J.; Roth, R. B.; Liu, T. M.; Geacintov, N. E.; Scicchitano, D. A., Incorrect base insertion and prematurely terminated transcripts during T7 RNA polymerase transcription elongation past benzo a pyrenediol epoxide-modified DNA. *J. Mol. Biol.* **1996**, 264 (2), 213-219.

42. Cline, S. D.; Riggins, J. N.; Tornaletti, S.; Marnett, L. J.; Hanawalt, P. C., Malondialdehyde adducts in DNA arrest transcription by T7 RNA polymerase and mammalian RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101* (19), 7275-7280.
43. Perlow, R. A.; Kolbanovskii, A.; Hingerty, B. E.; Geacintov, N. E.; Broyde, S.; Scicchitano, D. A., DNA adducts from a tumorigenic metabolite of benzo a pyrene block human RNA polymerase II elongation in a sequence- and stereochemistry-dependent manner. *J. Mol. Biol.* **2002**, *321* (1), 29-47.
44. Jain, V.; Hilton, B.; Patnaik, S.; Zou, Y.; Chiarelli, M. P.; Cho, B. P., Conformational and thermodynamic properties modulate the nucleotide excision repair of 2-aminofluorene and 2-acetylaminofluorene dG adducts in the NarI sequence. *Nucleic Acids Res.* **2012**, *40* (9), 3939-3951.
45. Jain, V.; Hilton, B.; Lin, B.; Patnaik, S.; Liang, F.; Darian, E.; Zou, Y.; MacKerell, A. D., Jr.; Cho, B. P., Unusual Sequence Effects on Nucleotide Excision Repair of Arylamine Lesions: DNA Bending/Distortion as a Primary Recognition Factor. *Nucleic Acids Res.* **2013**, *41* (2), 869-880.
46. Sale, J. E.; Lehmann, A. R.; Woodgate, R., Y-family DNA Polymerases and their Role in Tolerance of Cellular DNA Damage. *Nature Reviews Molecular Cell Biology* **2012**, *13* (3), 141-152.
47. Seeman, N. C.; Sleiman, H. F., DNA nanotechnology. *Nature Reviews Materials* **2018**, *3* (1).
48. Hong, F.; Zhang, F.; Liu, Y.; Yan, H., DNA Origami: Scaffolds for Creating Higher Order Structures. *Chem. Rev.* **2017**, *117* (20), 12584-12640.
49. Tapio, K.; Bald, I., The potential of DNA origami to build multifunctional materials. *Multifunctional Materials* **2020**, *3* (3), 032001.
50. Van Riesen, A. J.; Fadock, K. L.; Deore, P. S.; Desoky, A.; Manderville, R. A.; Sowlati-Hashjin, S.; Wetmore, S. D., Manipulation of a DNA aptamer-protein binding site through arylation of internal guanine residues. *Org. Biomol. Chem.* **2018**, *16* (20), 3831-3840.
51. Hermann, T.; Patel, D. J., Biochemistry - Adaptive recognition by nucleic acid aptamers. *Science* **2000**, *287* (5454), 820-825.
52. Dolot, R.; Lam, C. H.; Sierant, M.; Zhao, Q.; Liu, F. W.; Nawrot, B.; Egli, M.; Yang, X., Crystal structures of thrombin in complex with chemically modified thrombin DNA aptamers reveal the origins of enhanced affinity. *Nucleic Acids Res.* **2018**, *46* (9), 4819-4830.

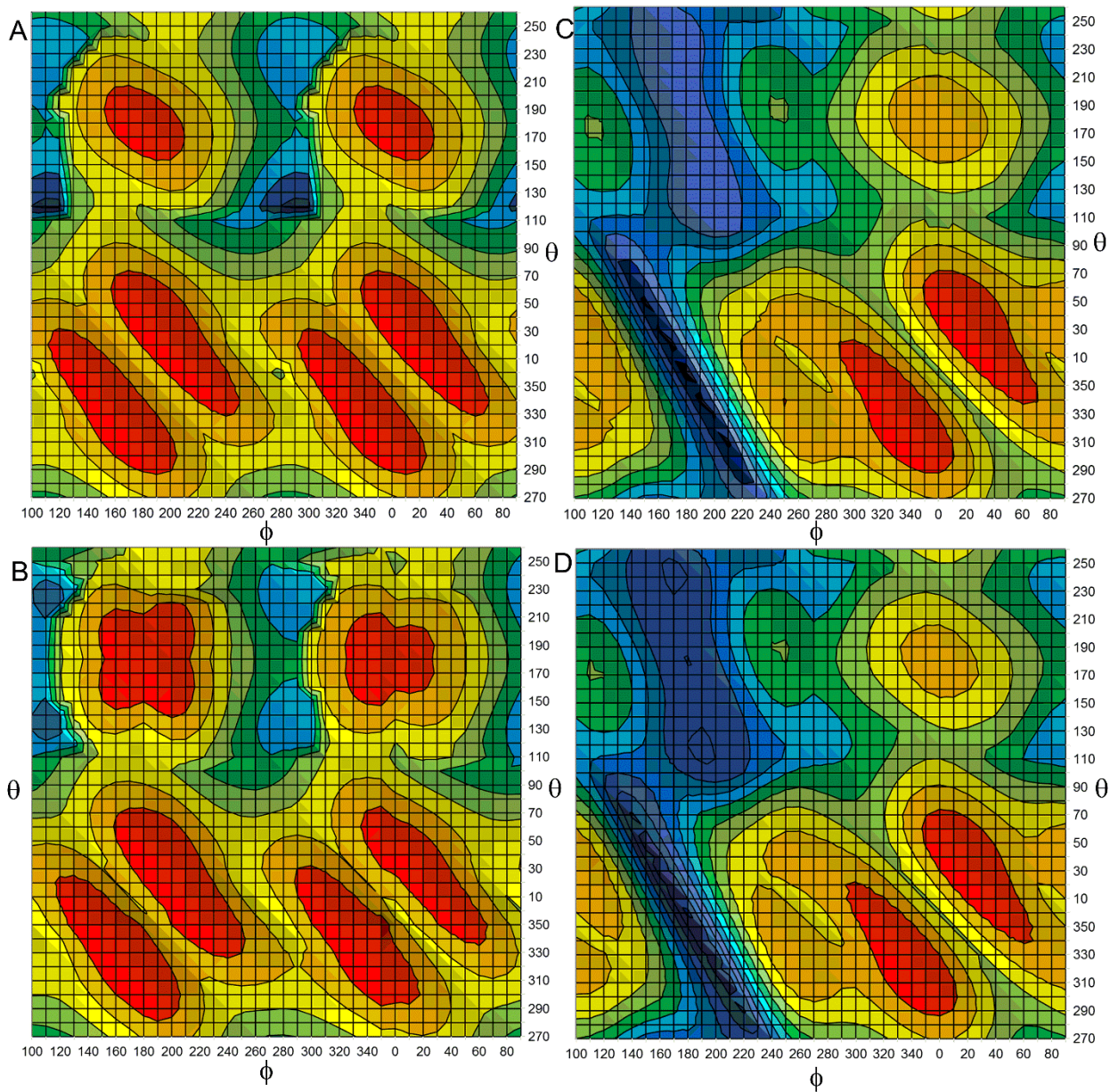
53. Valsangkar, V.; Vangaveti, S.; Lee, G. W.; Fahssi, W. M.; Awan, W. S.; Huang, Y. C.; Chen, A. L. A.; Sheng, J., Structural and Binding Effects of Chemical Modifications on Thrombin Binding Aptamer (TBA). *Molecules* **2021**, 26 (15).
54. Smirnov, I.; Kolganova, N.; Troisi, R.; Sica, F.; Timofeev, E., Expanding the recognition interface of the thrombin-binding aptamer HD1 through modification of residues T3 and T12. *Molecular Therapy-Nucleic Acids* **2021**, 23, 863-871.
55. Bao, H. L.; Ishizuka, T.; Yamashita, A.; Furukoji, E.; Asada, Y.; Xu, Y., Improving Thermodynamic Stability and Anticoagulant Activity of a Thrombin Binding Aptamer by Incorporation of 8-trifluoromethyl-2'-deoxyguanosine. *J. Med. Chem.* **2021**, 64 (1), 711-718.

## **Appendix A**

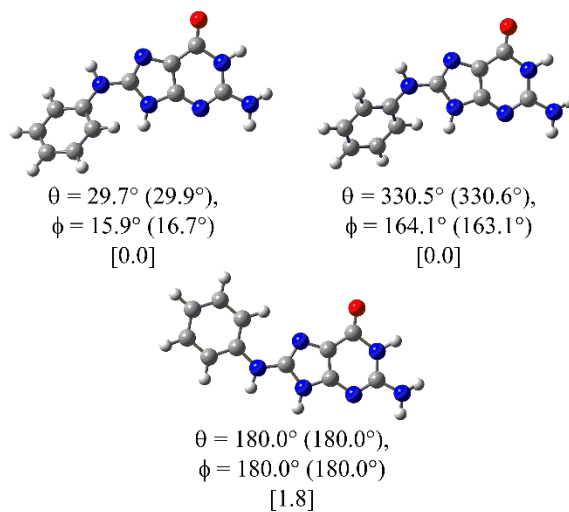
### **Supplementary Information for Chapter 2: Effect of Size and Shape of Nitrogen-Containing Aromatics on Conformational Preferences of DNA Containing Damaged Guanine**

Contains Figures A.1–A.20 and Tables A.1–A.58

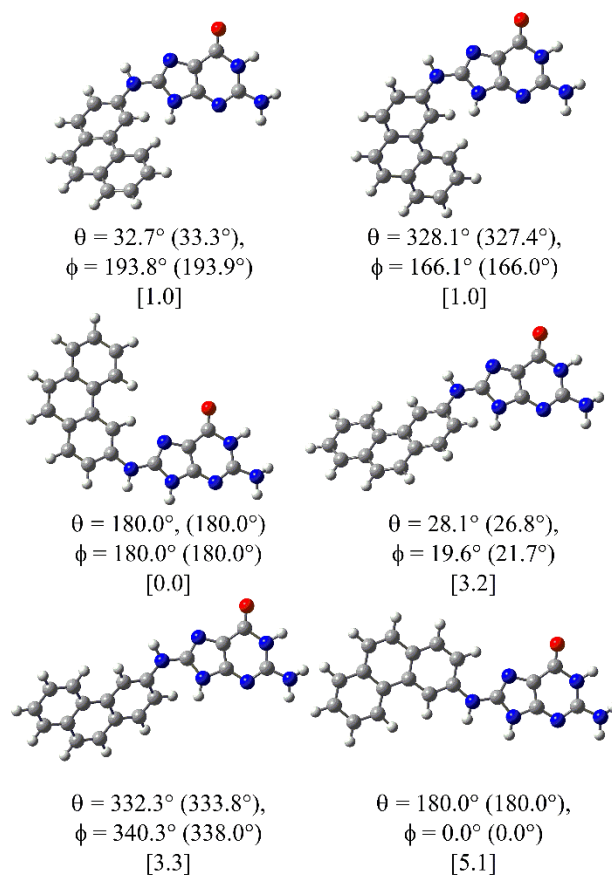




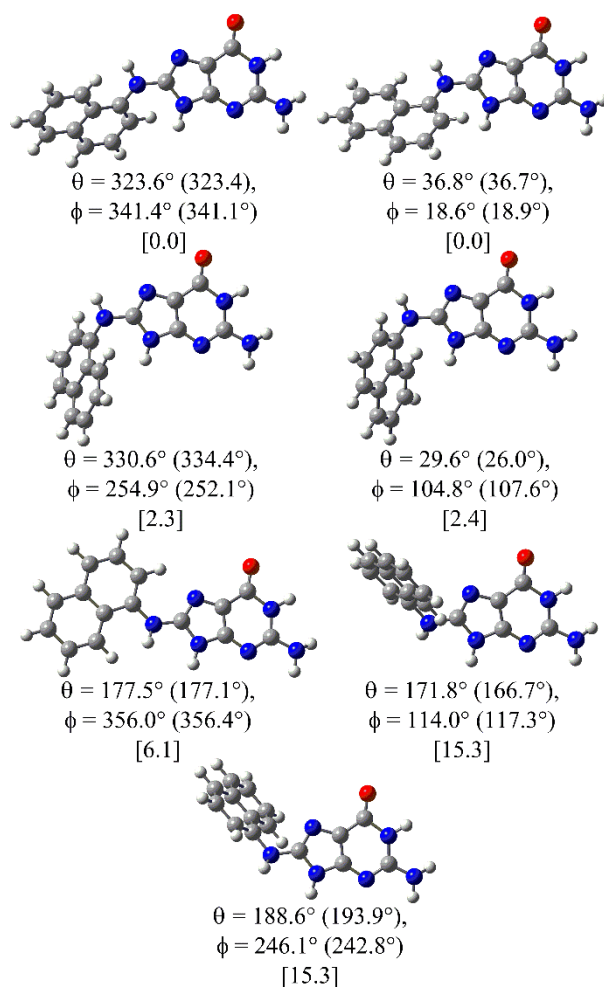
**Figure A.1.** Contour plots of the B3LYP/6-31G(d) potential energy for the (A)  $^{\text{AN}}\text{G}$ , (B)  $^{\text{3APHN}}\text{G}$ , (C)  $^{\text{1ANP}}\text{G}$  and (D)  $^{\text{1AP}}\text{G}$  nucleobase adducts with respect to  $\theta$  and  $\phi$  (deg). The relative energy is represented by color; red represents the lowest energy regions, and each change in color in the surrounding regions represents an increase in the energy by 5 kJ mol $^{-1}$ .



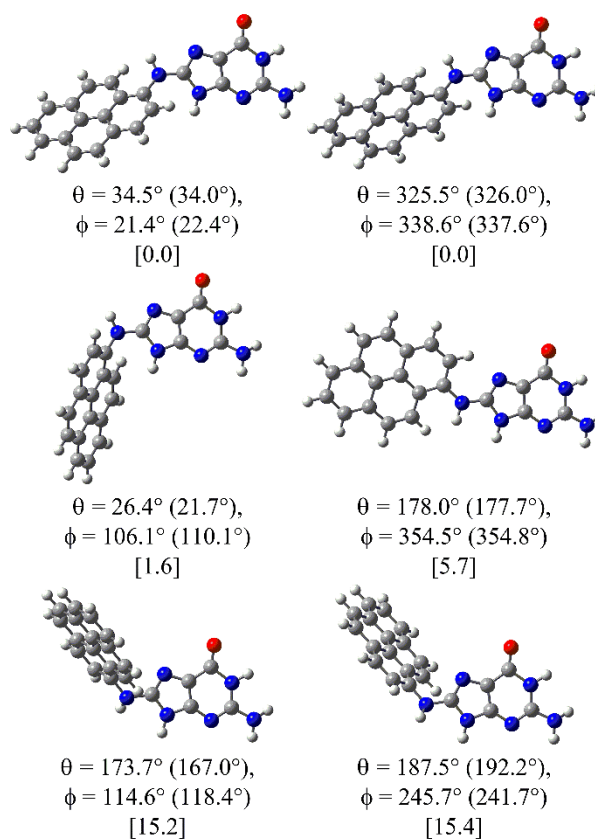
**Figure A.2.** Select B3LYP/6-31G(d) and B3LYP-D3(BJ)/6-31G(d) (in parenthesis) dihedral angles ( $\theta$  and  $\phi$ , deg) for the fully optimized minima identified from the contour plots for the <sup>AN</sup>G nucleobase adduct. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311+G(2df,p) relative energies are provided in parentheses ( $\text{kJ mol}^{-1}$ ).



**Figure A.3.** Select B3LYP/6-31G(d) and B3LYP-D3(BJ)/6-31G(d) (in parenthesis) dihedral angles ( $\theta$  and  $\phi$ , deg) for the fully optimized minima identified from the contour plots for the <sup>3</sup>APHNG nucleobase adduct. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311+G(2df,p) relative energies are provided in parentheses (kJ mol<sup>-1</sup>).

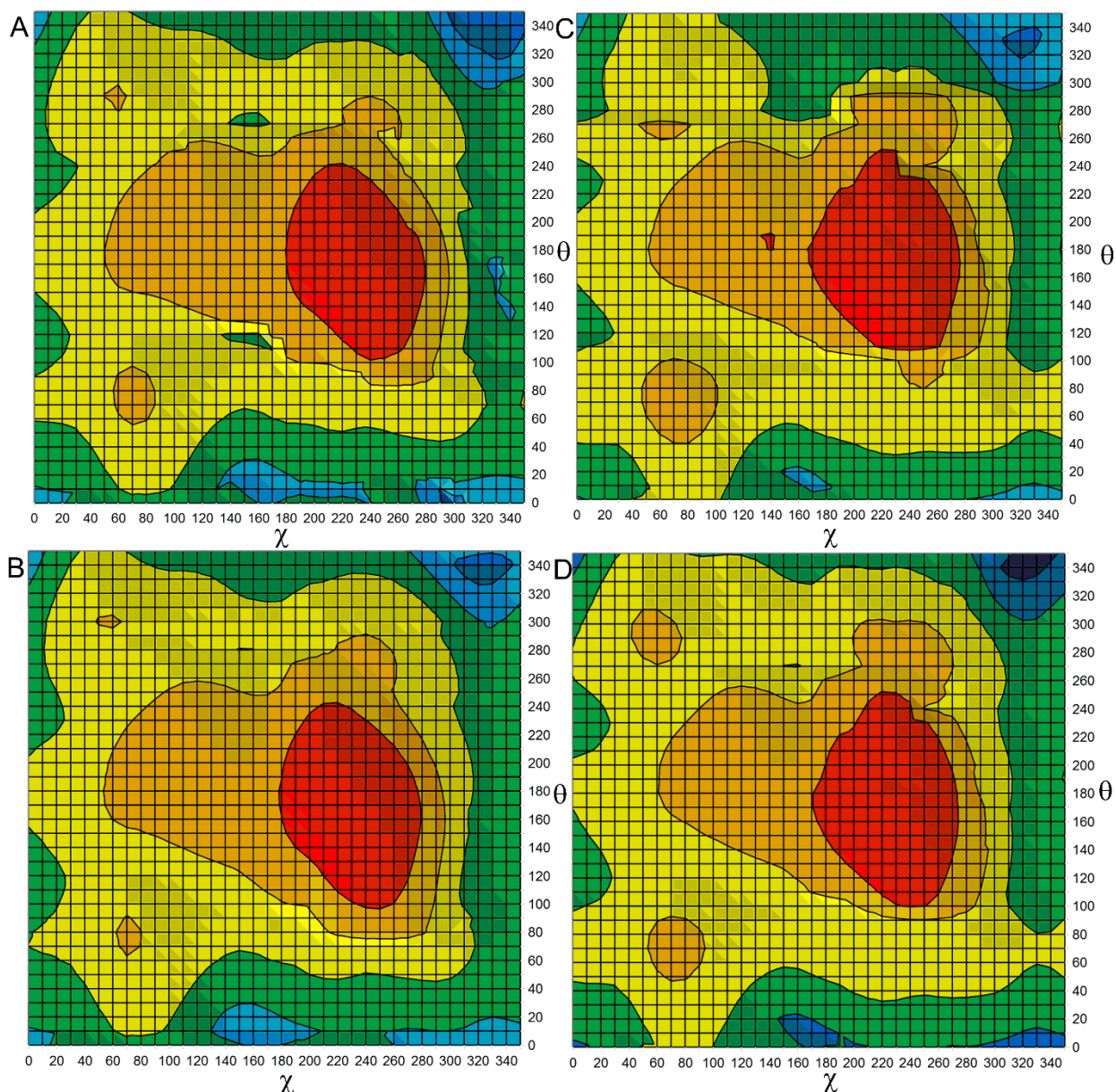


**Figure A.4.** Select B3LYP/6-31G(d) and B3LYP-D3(BJ)/6-31G(d) (in parenthesis) dihedral angles ( $\theta$  and  $\phi$ , deg) for the fully optimized minima identified from the contour plots for the  $^1\text{ANP}$ G nucleobase adduct. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311+G(2df,p) relative energies are provided in parentheses ( $\text{kJ mol}^{-1}$ ).

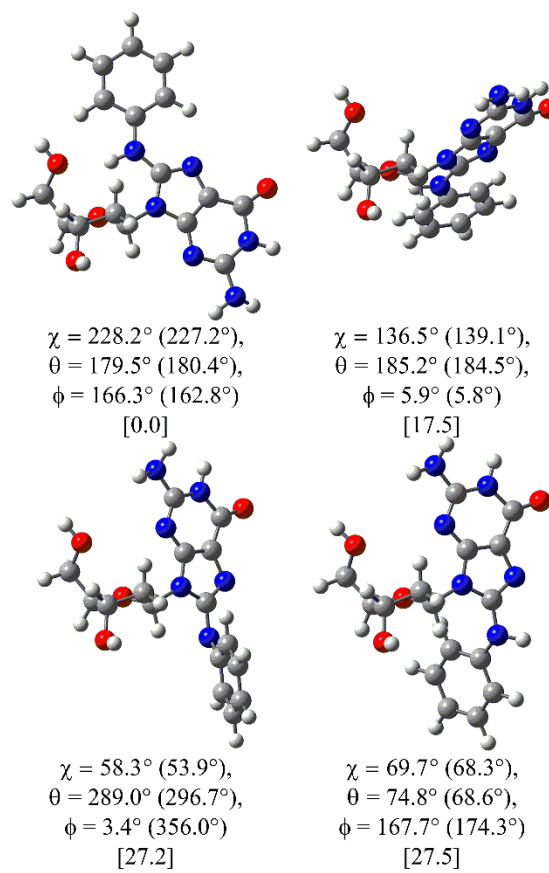


**Figure A.5.** Select B3LYP/6-31G(d) and B3LYP-D3(BJ)/6-31G(d) (in parenthesis) dihedral angles ( $\theta$  and  $\phi$ , deg) for the fully optimized minima identified from the contour plots for the  $^1\text{APG}$  nucleobase adduct. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311+G(2df,p) relative energies are provided in parentheses ( $\text{kJ mol}^{-1}$ ).

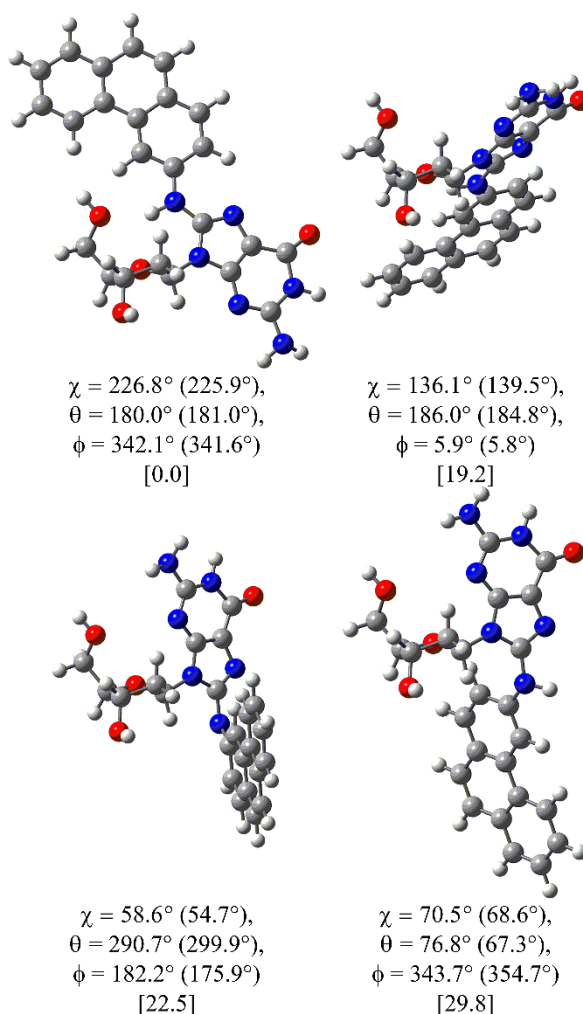




**Figure A.6.** Contour plots of the B3LYP/6-31G(d) potential energy for the (A)  $^{\text{AN}}\text{dG}$ , (B)  $^{\text{3APHN}}\text{dG}$ , (C)  $^{\text{1ANP}}\text{dG}$  and (D)  $^{\text{1AP}}\text{dG}$  nucleoside adducts with respect to  $\theta$  and  $\chi$  (deg). The relative energy is represented by color; red represents the lowest energy regions, and each change in color in the surrounding regions represents an increase in the energy by 20 kJ mol $^{-1}$ .

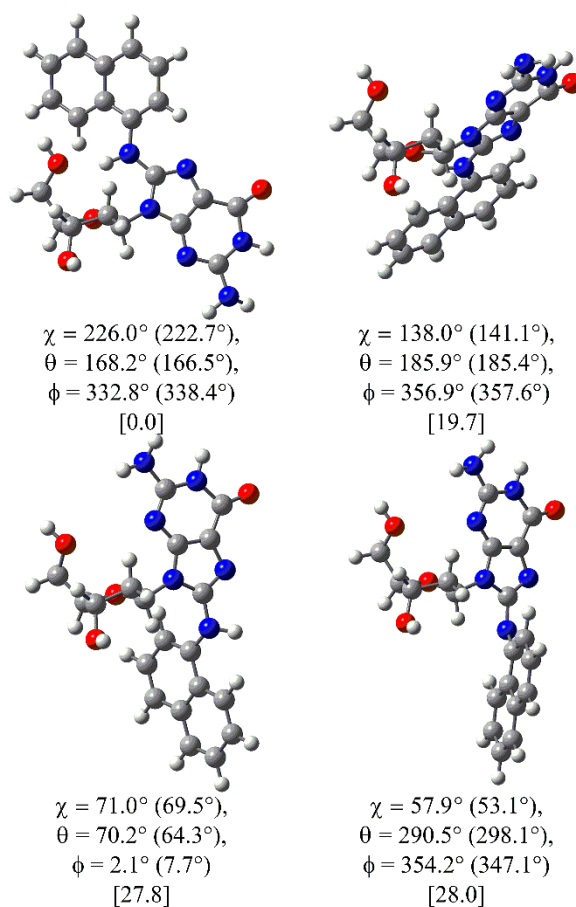


**Figure A.7.** Select B3LYP/6-31G(d) and B3LYP-D3(BJ)/6-31G(d) (in parenthesis) dihedral angles ( $\chi$ ,  $\theta$  and  $\phi$ , deg) for the fully optimized minima identified from the contour plots for the <sup>AN</sup>dG nucleoside adduct. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311+G(2df,p) relative energies are provided in parentheses (kJ mol<sup>-1</sup>).

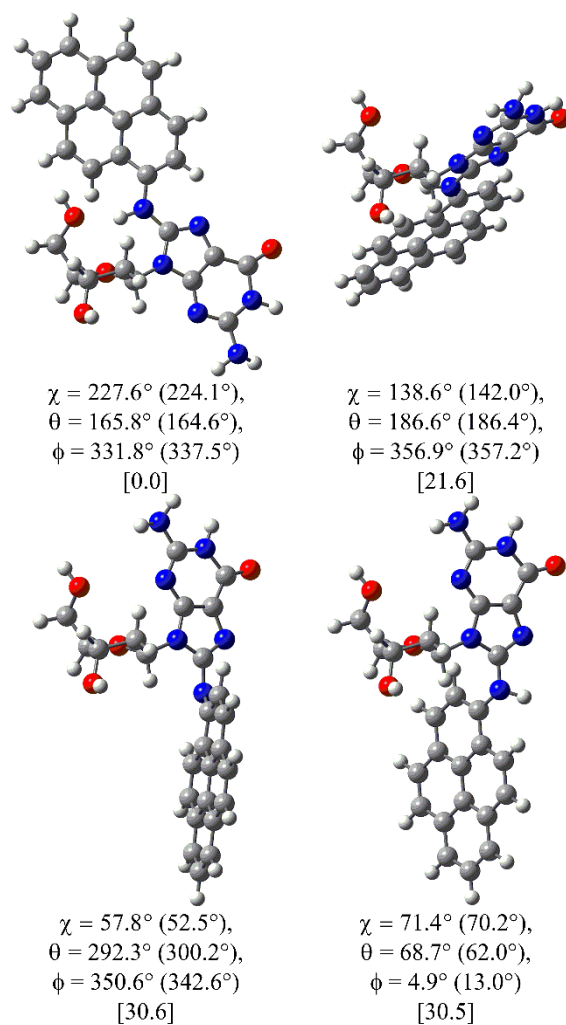


**Figure S8.** Select B3LYP/6-31G(d) and B3LYP-D3(BJ)/6-31G(d) (in parenthesis) dihedral angles ( $\chi$ ,  $\theta$  and  $\phi$ , deg) for the fully optimized minima identified from the contour plots for the <sup>3</sup>APHN dG nucleoside adduct. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311+G(2df,p) relative energies are provided in parentheses (kJ mol<sup>-1</sup>).

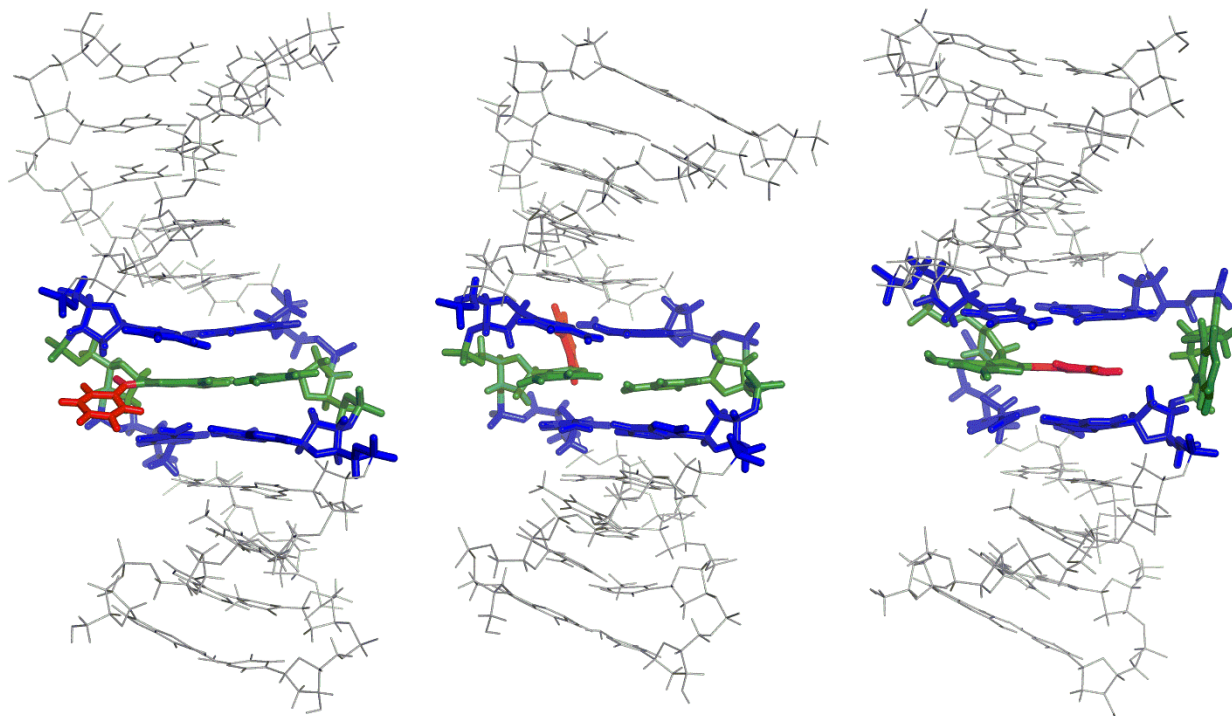




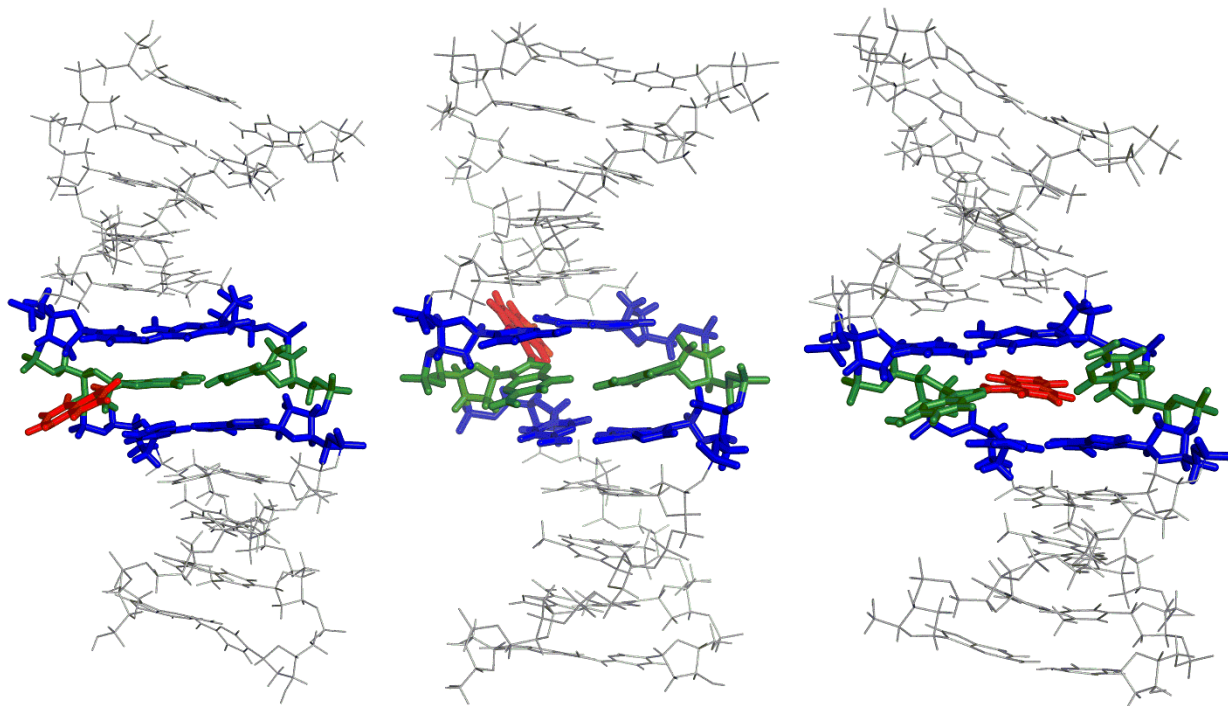
**Figure A.9.** Select B3LYP/6-31G(d) and B3LYP-D3(BJ)/6-31G(d) (in parenthesis) dihedral angles ( $\chi$ ,  $\theta$  and  $\phi$ , deg) for the fully optimized minima identified from the contour plots for the  $^{1\text{ANP}}$ dG nucleoside adduct. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311+G(2df,p) relative energies are provided in parentheses ( $\text{kJ mol}^{-1}$ ).



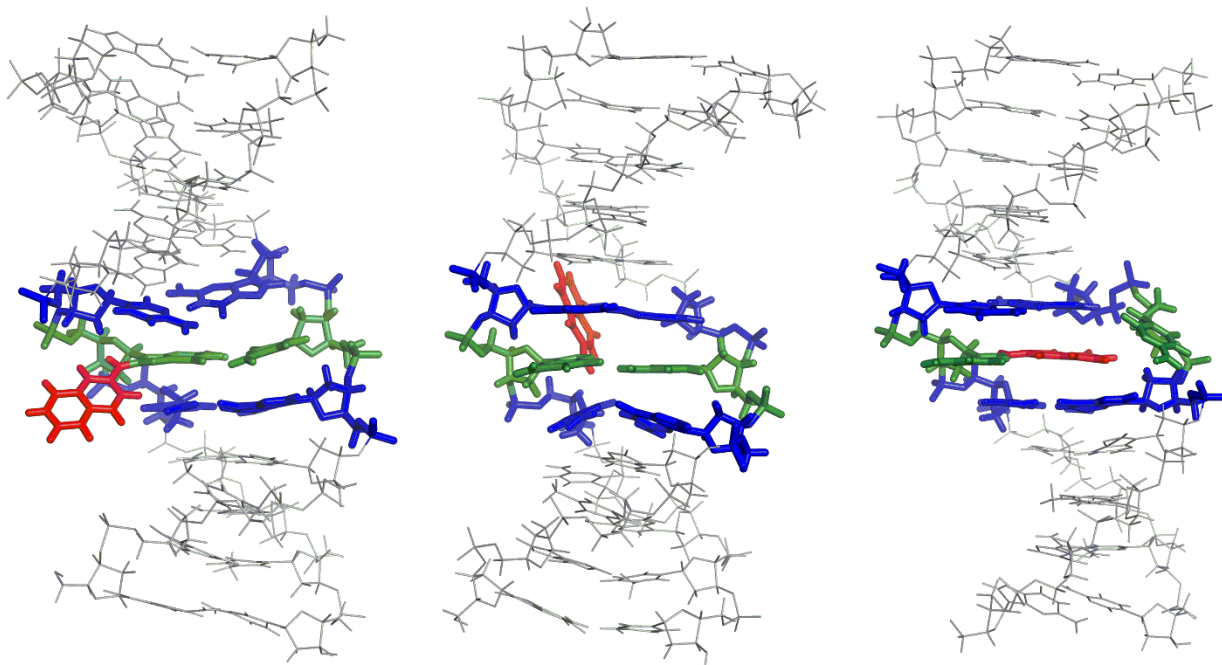
**Figure A.10.** Select B3LYP/6-31G(d) and B3LYP-D3(BJ)/6-31G(d) (in parenthesis) dihedral angles ( $\chi$ ,  $\theta$  and  $\phi$ , deg) for the fully optimized minima identified from the contour plots for the  $^1\text{APdG}$  nucleoside adduct. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311+G(2df,p) relative energies are provided in parentheses ( $\text{kJ mol}^{-1}$ ).



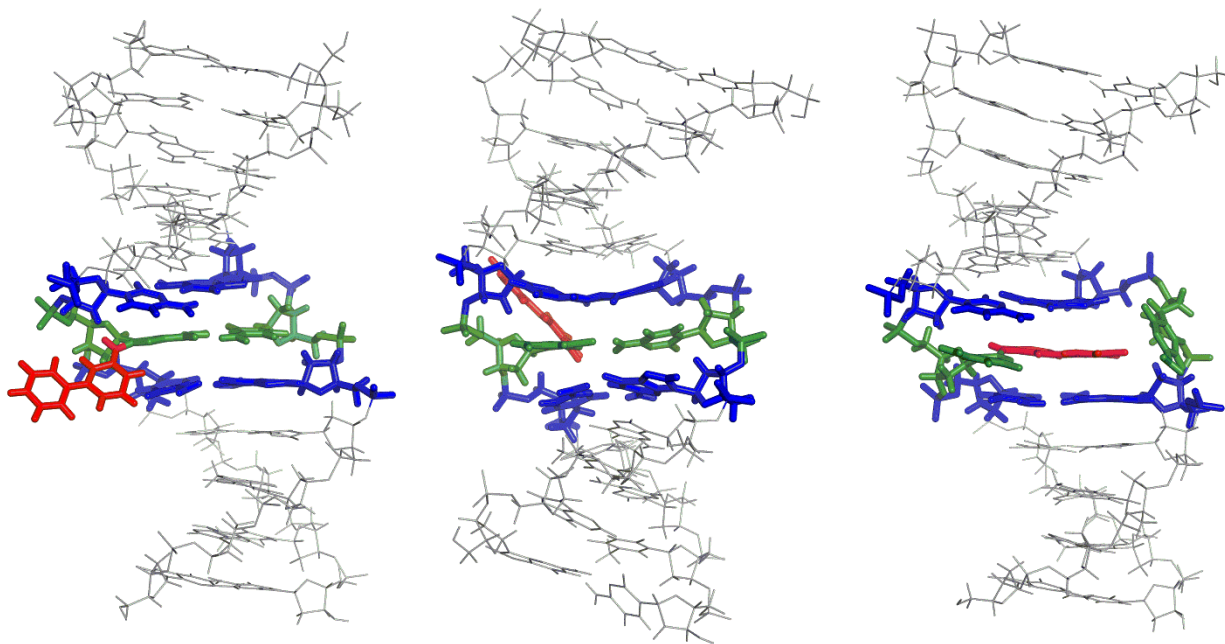
**Figure A.11.** Representative structures obtained from MD simulations of the B (left), W (middle) and S (right) conformations of <sup>AN</sup>dG adducted DNA. The C<sup>8</sup>-moiety is shown in red, the damaged dG and the opposing nucleotide in green, the base pairs flanking the lesion in blue, and rest of the DNA helix in grey.



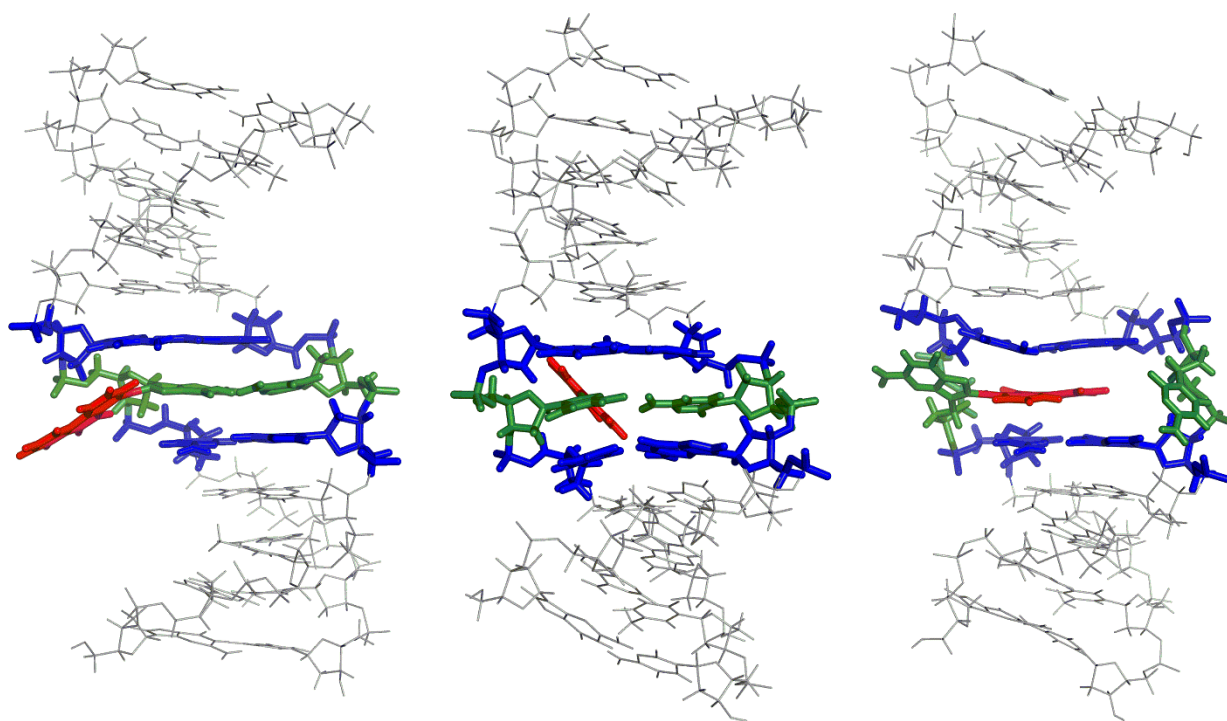
**Figure A.12.** Representative structures obtained from MD simulations of the B (left), W (middle) and S (right) conformations of  $^{1\text{ANP}}$ dG adducted DNA. The C<sup>8</sup>-moiety is shown in red, the damaged dG and the opposing nucleotide in green, the base pairs flanking the lesion in blue, and rest of the DNA helix in grey.



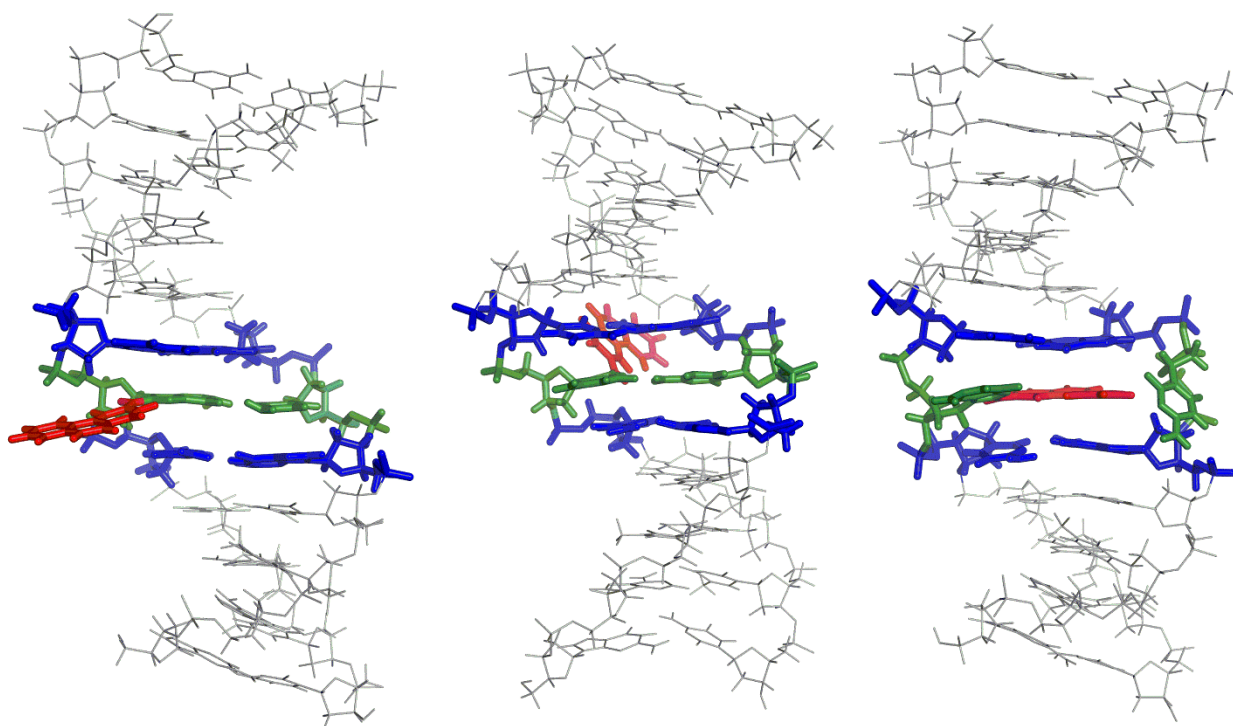
**Figure A.13.** Representative structures obtained from MD simulations of the B (left), W (middle) and S (right) conformations of  $^{2\text{ANP}}\text{dG}$  adducted DNA. The  $\text{C}^8$ -moiety is shown in red, the damaged dG and the opposing nucleotide in green, the base pairs flanking the lesion in blue, and rest of the DNA helix in grey.



**Figure A.14.** Representative structures obtained from MD simulations of the B (left), W (middle) and S (right) conformations of <sup>3</sup>ABP-dG adducted DNA. The C<sup>8</sup>-moiety is shown in red, the damaged dG and the opposing nucleotide in green, the base pairs flanking the lesion in blue, and rest of the DNA helix in grey.

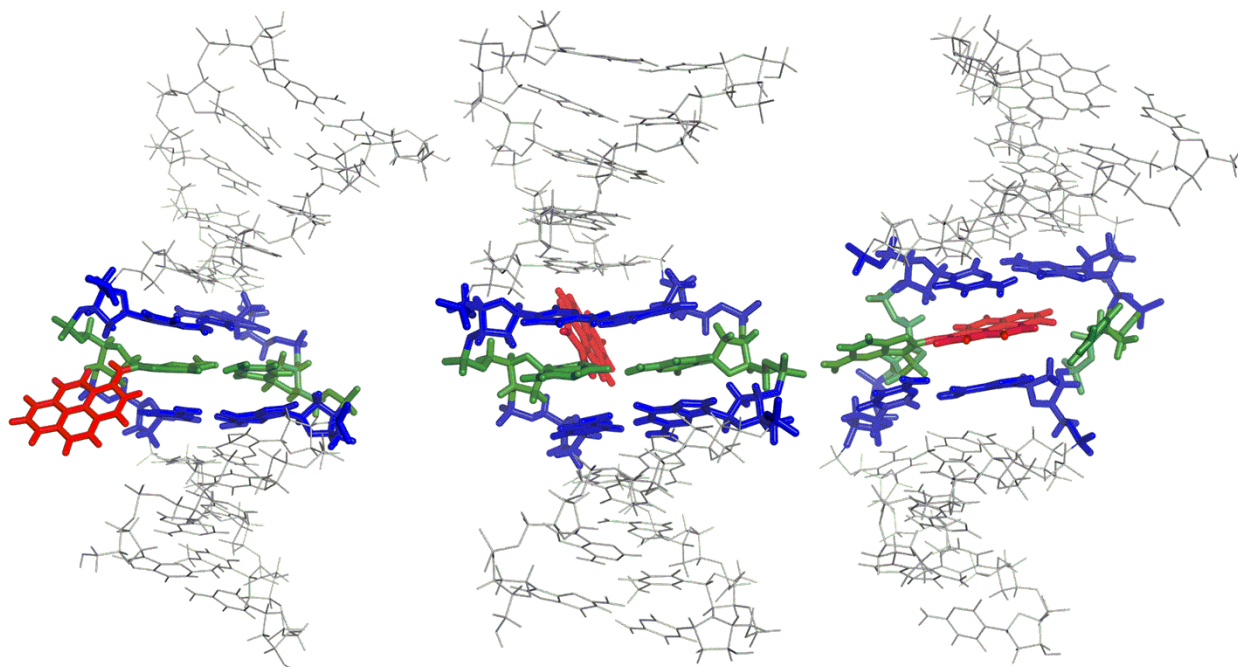


**Figure A.15.** Representative structures obtained from MD simulations of the B (left), W (middle) and S (right) conformations of  $^{1\text{APHN}}$ dG adducted DNA. The C<sup>8</sup>-moiety is shown in red, the damaged dG and the opposing nucleotide in green, the base pairs flanking the lesion in blue, and rest of the DNA helix in grey.

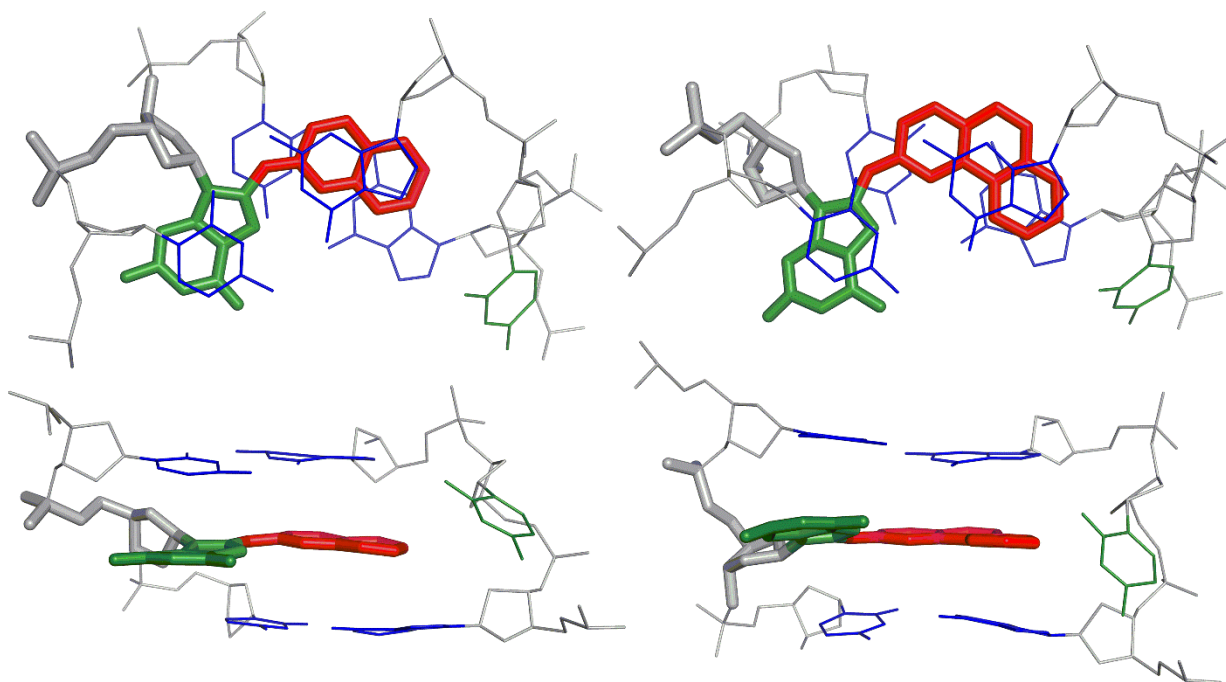


**Figure A.16.** Representative structures obtained from MD simulations of the B (left), W (middle) and S (right) conformations of <sup>3</sup>APHN-dG adducted DNA. The C<sup>8</sup>-moiety is shown in red, the damaged dG and the opposing nucleotide in green, the base pairs flanking the lesion in blue, and rest of the DNA helix in grey.

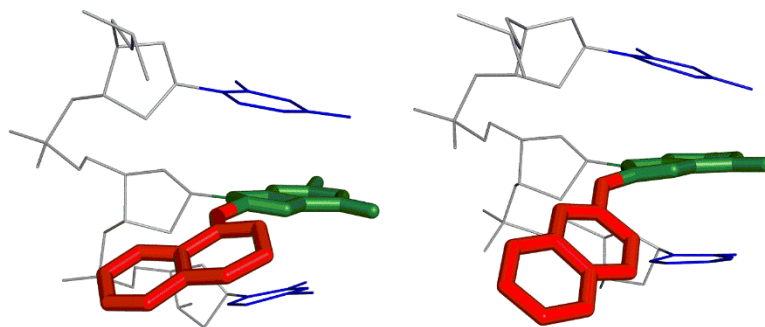




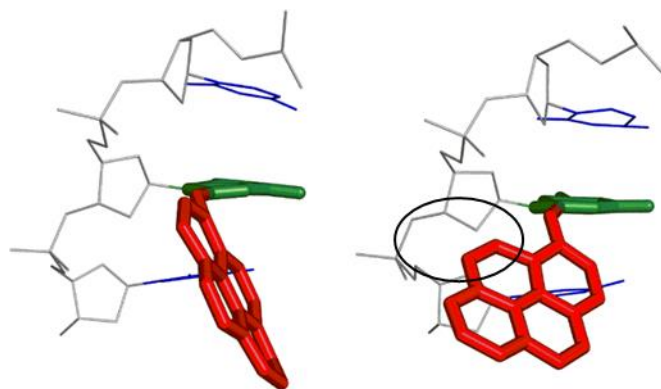
**Figure A.17.** Representative structures obtained from MD simulations of the B (left), W (middle) and S (right) conformations of <sup>1</sup>APdG adducted DNA. The C<sup>8</sup>-moiety is shown in red, the damaged dG and the opposing nucleotide in green, the base pairs flanking the lesion in blue, and rest of the DNA helix in grey.



**Figure A.18.** Lesion-site trimer of base pairs from  $^{2\text{ANP}}\text{dG}$  (left,  $\chi = 45.2^\circ$ ) and  $^{3\text{APHN}}\text{dG}$  (right,  $\chi = 132.3^\circ$ ) adducted DNA in the S conformer, highlighting how the changing  $\chi$  allows different NCA moieties to optimize stacking interactions with the flanking base pairs. The C<sup>8</sup>-moiety is shown in red, the damaged G and the opposing base in green, the base pairs flanking the lesion in blue, and the DNA backbone in grey.



**Figure A.19.** Structures of the  $^1\text{ANP-dG}$  (left,  $\theta = 165.3^\circ$ ) and  $^2\text{ANP-dG}$  (right,  $\theta = 139.9^\circ$ ) adducts in the B conformer of damaged DNA, highlighting the change in  $\theta$  with the presence (left) or absence (right) of an  $\alpha$ -fused ring due to the presence of a weak hydrogen bond between the bulky moiety and phosphate. The  $\text{C}^8$ -moiety is shown in red, the damaged G in green, the bases flanking the lesion in blue, and the DNA backbone in grey.



**Figure A.20.** Structures of the  $^1\text{APdG}$  adduct in the B conformer of damaged DNA with  $\phi \sim 0^\circ$  (left) or  $180^\circ$  (right), highlighting minimal steric repulsion between the bulky group and the backbone for  $\phi \sim 0^\circ$ . The  $\text{C}^8$ -moiety is shown in red, the damaged G moiety in green, the bases flanking the lesion in blue, and the DNA backbone in grey.

**Table A.1.** Comparison of the average structural and energetic parameters (standard deviations in parentheses) for damaged DNA containing select adducts deduced from 20 ns and 120 ns MD simulations.

| Conformation | Parameter                                 | <sup>AN</sup> dG |               | <sup>1AP</sup> dG |               |
|--------------|-------------------------------------------|------------------|---------------|-------------------|---------------|
|              |                                           | 20 ns            | 120 ns        | 20 ns             | 120 ns        |
| B            | RMSD (Å)                                  | 4.0 (0.7)        | 4.2 (0.7)     | 2.6 (0.5)         | 4.2 (0.7)     |
|              | $\chi^{\text{av}}$ (deg)                  | 227.8 (12.7)     | 226.6 (13.2)  | 218.5 (12.2)      | 215.2 (17.4)  |
|              | $\theta^{\text{av}}$ (deg)                | 139.9 (20.6)     | 140.8 (22.1)  | 166.4 (20.6)      | 167.2 (25.5)  |
|              | $\phi^{\text{av}}$ (deg)                  | 2.9 (12.2)       | 3.2 (12.1)    | 17.6 (11.3)       | 15.0 (25.9)   |
|              | LSIE (kJ mol <sup>-1</sup> ) <sup>a</sup> | -475.5 (19.2)    | -474.8 (17.1) | -473.9 (17.9)     | -473.3 (18.3) |
| W            | RMSD (Å)                                  | 3.8 (0.8)        | 3.9 (0.7)     | 4.1 (0.7)         | 4.0 (0.7)     |
|              | $\chi^{\text{av}}$ (deg)                  | 67.7 (13.2)      | 69.3 (13.4)   | 64.0 (12.5)       | 65.8 (12.9)   |
|              | $\theta^{\text{av}}$ (deg)                | 76.1 (18.1)      | 74.0 (17.8)   | 68.2 (12.9)       | 70.0 (13.3)   |
|              | $\phi^{\text{av}}$ (deg)                  | 1.2 (9.9)        | 1.7 (10.0)    | 1.1 (10.2)        | 359.9 (10.2)  |
|              | LSIE (kJ mol <sup>-1</sup> ) <sup>a</sup> | -361.9 (24.9)    | -349.2 (39.8) | -367.5 (19.1)     | -356.7 (34.5) |
| S            | RMSD (Å)                                  | 2.5 (0.4)        | 3.3 (1.0)     | 5.0 (0.7)         | 4.7 (0.7)     |
|              | $\chi^{\text{av}}$ (deg)                  | 42.1 (13.2)      | 54.7 (23.4)   | 48.4 (33.6)       | 45.3 (19.0)   |
|              | $\theta^{\text{av}}$ (deg)                | 167.0 (15.9)     | 147.9 (37.3)  | 205.6 (11.0)      | 175.9 (18.0)  |
|              | $\phi^{\text{av}}$ (deg)                  | 357.6 (10.9)     | 358.7 (10.8)  | 353.2 (11.5)      | 352.8 (10.9)  |
|              | LSIE (kJ mol <sup>-1</sup> ) <sup>a</sup> | -300.6 (17.1)    | -300.3 (17.6) | -358.2 (21.5)     | -355.6 (25.0) |

<sup>a</sup>Lesion-site interaction energies (LSIE) were calculated as the sum of the van der Waals and electrostatic interactions between the bases constituting the 5' and 3' base pairs flanking the lesion, the lesions and the opposing base, and the lesion and the 5' and 3' flanking base pairs.

**Table A.2.** Comparison of the average structural and energetic parameters (standard deviations in parentheses) for damaged DNA containing select adducts deduced from replicate 20 ns and 120 ns MD simulations.<sup>a</sup>

| System      | Parameter                                 | <sup>AN</sup> dG |                 |                 |                                | <sup>1AP</sup> dG |                 |                 |                                |
|-------------|-------------------------------------------|------------------|-----------------|-----------------|--------------------------------|-------------------|-----------------|-----------------|--------------------------------|
|             |                                           | Replicate 1      | Replicate 2     | Replicate 3     | Replicate Average <sup>b</sup> | Replicate 1       | Replicate 2     | Replicate 3     | Replicate Average <sup>b</sup> |
| B<br>20 ns  | RMSD (Å)                                  | 4.0 (0.7)        | 4.0 (0.7)       | 4.2 (0.8)       | 4.1 (0.1)                      | 2.6 (0.5)         | 4.3 (0.7)       | 4.2 (0.7)       | 3.7 (1.0)                      |
|             | $\chi^{av}$ (deg)                         | 227.8 (12.7)     | 226.7 (12.9)    | 226.6 (12.9)    | 227.0 (0.5)                    | 218.5 (12.2)      | 218.9 (13.3)    | 222.4 (13.8)    | 219.9 (1.8)                    |
|             | $\theta^{av}$ (deg)                       | 139.9 (20.6)     | 139.8 (21.6)    | 140.5 (22.2)    | 140.1 (0.3)                    | 166.4 (20.6)      | 167.0 (20.9)    | 161.3 (25.0)    | 164.9 (2.6)                    |
|             | $\phi^{av}$ (deg)                         | 2.9 (12.2)       | 3.1 (12.1)      | 3.3 (12.1)      | 3.1 (0.2)                      | 17.6 (11.3)       | 16.8 (11.6)     | 15.5 (11.5)     | 16.6 (0.9)                     |
|             | LSIE (kJ mol <sup>-1</sup> ) <sup>c</sup> | -475.5 (19.2)    | -474.6 (17.0)   | -473.8 (17.2)   | -474.6 (0.7)                   | -473.9 (17.9)     | -475.3 (16.6)   | -477.3 (16.7)   | -475.5 (1.4)                   |
|             | G (kJ mol <sup>-1</sup> ) <sup>d</sup>    | -24010.0 (81.5)  | -24009.8 (85.5) | -24013.8 (85.4) | -24011.2 (1.8)                 | -23774.0 (79.8)   | -23757.3 (90.4) | -23771.7 (91.3) | -23767.7 (7.4)                 |
| B<br>120 ns | RMSD (Å)                                  | 4.2 (0.7)        | 4.0 (0.7)       | 4.0 (0.7)       | 4.1 (0.1)                      | 4.2 (0.7)         | 4.3 (0.7)       | 4.2 (0.7)       | 4.2 (0.1)                      |
|             | $\chi^{av}$ (deg)                         | 226.6 (13.2)     | 227.2 (12.7)    | 227.2 (12.7)    | 227.0 (0.3)                    | 215.2 (17.4)      | 220.1 (13.0)    | 216.1 (16.5)    | 217.1 (2.1)                    |
|             | $\theta^{av}$ (deg)                       | 140.8 (22.1)     | 139.5 (21.5)    | 139.3 (21.5)    | 139.9 (0.7)                    | 167.2 (25.5)      | 163.8 (22.2)    | 168.3 (24.3)    | 166.6 (2.1)                    |
|             | $\phi^{av}$ (deg)                         | 3.2 (12.1)       | 3.0 (12.0)      | 3.0 (12.0)      | 3.1 (0.1)                      | 15.0 (25.9)       | 16.2 (12.0)     | 14.9 (12.4)     | 15.4 (0.6)                     |
|             | LSIE (kJ mol <sup>-1</sup> ) <sup>c</sup> | -474.8 (17.1)    | -474.4 (17.4)   | -474.9 (16.9)   | -474.7 (0.2)                   | -473.3 (18.3)     | -475.2 (17.2)   | -474.8 (17.4)   | -474.4 (0.8)                   |
|             | G (kJ mol <sup>-1</sup> ) <sup>d</sup>    | -24012.6 (86.7)  | -24004.0 (79.5) | -24013.1 (89.8) | -24009.9 (4.2)                 | -23762.5 (84.1)   | -23764.5 (87.2) | -23762.6 (86.7) | -23763.2 (0.9)                 |
| S<br>20 ns  | RMSD (Å)                                  | 2.5 (0.4)        | 2.6 (0.7)       | 2.7 (0.5)       | 2.6 (0.1)                      | 5.0 (0.7)         | 4.4 (0.8)       | 4.5 (0.7)       | 4.6 (0.3)                      |
|             | $\chi^{av}$ (deg)                         | 42.1 (13.2)      | 43.7 (12.6)     | 47.6 (13.1)     | 44.5 (2.3)                     | 48.4 (33.6)       | 51.7 (30.6)     | 42.8 (13.0)     | 47.6 (3.7)                     |
|             | $\theta^{av}$ (deg)                       | 167.0 (15.9)     | 164.9 (15.5)    | 161.0 (15.0)    | 164.3 (2.5)                    | 205.6 (11.0)      | 179.4 (18.1)    | 172.3 (14.5)    | 185.8 (14.3)                   |
|             | $\phi^{av}$ (deg)                         | 357.6 (10.9)     | 357.9 (10.7)    | 358.8 (10.8)    | 358.1 (0.5)                    | 353.2 (11.5)      | 353.7 (11.1)    | 352.2 (10.8)    | 353.0 (0.6)                    |
|             | LSIE (kJ mol <sup>-1</sup> ) <sup>c</sup> | -300.6 (17.1)    | -300.9 (16.2)   | -306.2 (15.6)   | -302.6 (2.6)                   | -358.2 (21.5)     | -356.8 (17.7)   | -344.4 (35.9)   | -353.1 (6.2)                   |
|             | G (kJ mol <sup>-1</sup> ) <sup>d</sup>    | -23925.1 (85.0)  | -23921.1 (85.6) | -23915.7 (91.6) | -23920.6 (3.8)                 | -23738.9 (88.2)   | -23724.9 (83.9) | -23723.5 (84.2) | -23729.1 (6.9)                 |
| S<br>120 ns | RMSD (Å)                                  | 3.3 (1.0)        | 2.6 (0.6)       | 3.0 (1.0)       | 3.0 (0.3)                      | 4.7 (0.7)         | 4.7 (0.7)       | 4.6 (0.7)       | 4.7 (0.1)                      |
|             | $\chi^{av}$ (deg)                         | 54.7 (23.4)      | 43.7 (13.1)     | 50.5 (17.9)     | 49.6 (4.5)                     | 45.3 (19.0)       | 52.8 (33.3)     | 42.7 (18.4)     | 46.9 (4.3)                     |
|             | $\theta^{av}$ (deg)                       | 147.9 (37.3)     | 163.9 (15.7)    | 160.5 (15.6)    | 157.4 (6.9)                    | 175.9 (18.0)      | 181.7 (20.0)    | 181.7 (20.4)    | 179.8 (2.7)                    |
|             | $\phi^{av}$ (deg)                         | 358.7 (10.8)     | 358.2 (10.8)    | 359.4 (10.8)    | 358.8 (0.5)                    | 352.8 (10.9)      | 354.5 (11.7)    | 352.8 (10.9)    | 353.4 (0.8)                    |
|             | LSIE (kJ mol <sup>-1</sup> ) <sup>c</sup> | -300.3 (17.6)    | -300.8 (16.1)   | -306.8 (15.8)   | -302.6 (3.0)                   | -355.6 (25.0)     | -351.6 (27.3)   | -358.9 (22.4)   | -355.4 (3.0)                   |
|             | G (kJ mol <sup>-1</sup> ) <sup>d</sup>    | -23923.3 (100.8) | -23922.7 (90.2) | -23925.5 (82.0) | -23923.8 (1.2)                 | -23718.0 (84.4)   | -23719.4 (90.2) | -23725.6 (88.5) | -23721.0 (3.3)                 |

<sup>a</sup>Replicas were obtained from the same starting structure with different initial velocities. <sup>b</sup>Replicate averages calculated as the sum and standard deviation between the three replicate values reported. <sup>c</sup>Lesion-site interaction energies (LSIE) were calculated as the sum of the van der Waals and electrostatic interactions between the bases constituting the 5' and 3' base pairs flanking the lesion, the lesions and the opposing base, and the lesion and the 5' and 3' flanking base pairs. <sup>d</sup>G energies calculated using MM/PBSA enthalpies and normal mode entropies.

**Table A.3.** MM/PBSA free energy rankings ( $G^{\text{rel}}$ , kJ mol<sup>-1</sup>), and average  $\theta$  and  $\phi$  dihedral angles (deg) deduced from MD simulations on various damaged DNA conformations containing the <sup>AN</sup>G, <sup>1ANP</sup>dG or <sup>2ANP</sup>dG adducts.

| Adduct             | Conformation | Occupancy <sup>a</sup><br>(%) | $\theta^{\text{av}}$ (deg) | $\phi^{\text{av}}$ (deg) | $G^{\text{rel}}$ (kJ mol <sup>-1</sup> ) |
|--------------------|--------------|-------------------------------|----------------------------|--------------------------|------------------------------------------|
| <sup>AN</sup> dG   | B            | 100                           | 139.9                      | 2.9                      | 0.0                                      |
|                    |              | 100                           | 134.3                      | 184.0                    | 6.7                                      |
|                    | W            | 100                           | 76.1                       | 1.2                      | 47.0                                     |
|                    |              | 100                           | 75.4                       | 1.4                      | 62.8                                     |
|                    | S            | 100                           | 167.0                      | 357.6                    | 85.0                                     |
|                    |              | 100                           | 167.7                      | 178.6                    | 96.8                                     |
| <sup>1ANP</sup> dG | B            | 100                           | 165.3                      | 17.1                     | 0.0                                      |
|                    |              | 100                           | 131.1                      | 165.5                    | 34.2                                     |
|                    |              | 100                           | 157.9                      | 219.9                    | 28.7                                     |
|                    | W            | 100                           | 70.7                       | 359.6                    | 39.0                                     |
|                    |              | 100                           | 88.5                       | 181.0                    | 66.1                                     |
|                    |              | 58                            | 225.4                      | 356.8                    | 69.1                                     |
|                    |              | 47                            | 263.9                      | 156.5                    | 104.1                                    |
|                    |              | 53                            | 218.8                      | 190.4                    | 88.1                                     |
|                    | S            | 100                           | 189.2                      | 355.6                    | 52.0                                     |
|                    |              | 100                           | 303.5                      | 4.2                      | 100.8                                    |
|                    |              | 100                           | 166.6                      | 357.6                    | 77.4                                     |
|                    |              | 100                           | 294.5                      | 7.9                      | 127.5                                    |
|                    |              | 100                           | 151.8                      | 166.8                    | 80.4                                     |
| <sup>2ANP</sup> dG | B            | 100                           | 139.9                      | 5.4                      | 0.0                                      |
|                    |              | 47                            | 130.7                      | 183.8                    | 11.2                                     |
|                    |              | 53                            | 170.1                      | 188.0                    | 17.2                                     |
|                    |              | 29                            | 121.5                      | 185.3                    | 6.9                                      |
|                    |              | 35                            | 148.1                      | 183.9                    | 2.3                                      |
|                    | W            | 100                           | 76.2                       | 180.5                    | 40.5                                     |
|                    |              | 100                           | 70.7                       | 0.9                      | 42.9                                     |
|                    |              | 44                            | 166.1                      | 358.1                    | 83.0                                     |
|                    |              | 100                           | 228.5                      | 177.5                    | 66.4                                     |
|                    | S            | 100                           | 177.1                      | 175.2                    | 62.3                                     |
|                    |              | 100                           | 168.1                      | 354.1                    | 92.6                                     |
|                    |              | 30                            | 294.6                      | 4.0                      | 101.0                                    |
|                    |              | 70                            | 298.8                      | 7.4                      | 79.7                                     |
|                    |              | 100                           | 232.8                      | 179.9                    | 81.9                                     |

<sup>a</sup>Occupancies of hydrogen-bonding interactions were determined using a cutoff of 3.4 Å for the donor–acceptor distance and 120° for the donor–hydrogen–acceptor angle.

**Table A.4.** MM/PBSA free energy rankings ( $G^{\text{rel}}$ , kJ mol<sup>-1</sup>), and average  $\theta$  and  $\phi$  dihedral angles (deg) deduced from MD simulations on various damaged DNA conformations containing the <sup>3</sup>ABP dG, <sup>1</sup>APHN dG and <sup>3</sup>APHN dG adducts.

| Adduct               | Conformation | Occupancy <sup>a</sup><br>(%) | $\theta^{\text{av}}$ (deg) | $\phi^{\text{av}}$ (deg) | $G^{\text{rel}}$ (kJ mol <sup>-1</sup> ) |
|----------------------|--------------|-------------------------------|----------------------------|--------------------------|------------------------------------------|
| <sup>3</sup> ABP dG  | B            | 100                           | 124.6                      | 5.3                      | 0.0                                      |
|                      |              | 100                           | 163.7                      | 189.8                    | 18.7                                     |
|                      |              | 41                            | 171.3                      | 12.2                     | 7.3                                      |
|                      | W            | 100                           | 62.4                       | 183.8                    | 54.7                                     |
|                      |              | 100                           | 72.1                       | 0.8                      | 78.2                                     |
|                      |              | 100                           | 225.6                      | 177.5                    | 84.4                                     |
|                      | S            | 100                           | 196.5                      | 175.9                    | 37.6                                     |
|                      |              | 100                           | 69.3                       | 348.0                    | 54.8                                     |
|                      |              | 100                           | 166.0                      | 356.0                    | 79.1                                     |
|                      |              | 100                           | 73.7                       | 169.7                    | 73.4                                     |
| <sup>1</sup> APHN dG | B            | 100                           | 174.8                      | 17.8                     | 0.0                                      |
|                      |              | 27                            | 126.8                      | 10.8                     | 17.3                                     |
|                      |              | 100                           | 154.3                      | 224.1                    | 25.5                                     |
|                      |              | 37                            | 141.3                      | 160.5                    | 38.4                                     |
|                      | W            | 100                           | 71.7                       | 359.0                    | 53.7                                     |
|                      |              | 100                           | 73.2                       | 188.9                    | 82.4                                     |
|                      |              | 100                           | 232.4                      | 178.1                    | 110.4                                    |
|                      |              | 100                           | 232.1                      | 356.9                    | 74.3                                     |
|                      | S            | 100                           | 206.6                      | 5.9                      | 35.9                                     |
|                      |              | 100                           | 300.4                      | 1.2                      | 108.7                                    |
|                      |              | 84                            | 166.7                      | 354.1                    | 84.2                                     |
|                      |              | 100                           | 152.1                      | 164.9                    | 113.8                                    |
| <sup>3</sup> APHN dG | B            | 100                           | 135.4                      | 8.6                      | 0.0                                      |
|                      |              | 30                            | 163.4                      | 10.2                     | 14.8                                     |
|                      |              | 70                            | 118.9                      | 8.3                      | 17.0                                     |
|                      |              | 55                            | 174.0                      | 189.8                    | 6.4                                      |
|                      |              | 45                            | 131.5                      | 186.9                    | 11.0                                     |
|                      |              | 100                           | 148.2                      | 188.0                    | 10.4                                     |
|                      | W            | 100                           | 72.5                       | 0.1                      | 49.3                                     |
|                      |              | 100                           | 58.7                       | 186.1                    | 51.7                                     |
|                      |              | 100                           | 229.4                      | 171.8                    | 63.4                                     |
|                      |              | 30                            | 230.0                      | 355.9                    | 55.9                                     |
|                      | S            | 100                           | 191.9                      | 177.5                    | 50.9                                     |
|                      |              | 100                           | 166.6                      | 354.0                    | 67.6                                     |
|                      |              | 100                           | 298.2                      | 11.8                     | 89.0                                     |
|                      |              | 65                            | 77.0                       | 175.1                    | 172.6                                    |

<sup>a</sup>Occupancies of hydrogen-bonding interactions were determined using a cutoff of 3.4 Å for the donor–acceptor distance and 120° for the donor–hydrogen–acceptor angle.



**Table A.5.** MM/PBSA free energy rankings ( $G^{\text{rel}}$ , kJ mol<sup>-1</sup>), and average  $\theta$  and  $\phi$  dihedral angles (deg) deduced from MD simulations on various damaged DNA conformations containing the <sup>1</sup>APdG adduct.

| Adduct            | Conformation | Occupancy <sup>a</sup><br>(%) | $\theta^{\text{av}}$ (deg) | $\phi^{\text{av}}$ (deg) | $G^{\text{rel}}$ (kJ mol <sup>-1</sup> ) |
|-------------------|--------------|-------------------------------|----------------------------|--------------------------|------------------------------------------|
| <sup>1</sup> APdG | B            | 100                           | 166.4                      | 17.6                     | 0.0                                      |
|                   |              | 100                           | 135.1                      | 167.1                    | 35.2                                     |
|                   |              | 65                            | 179.3                      | 12.7                     | 14.5                                     |
|                   |              | 35                            | 127.1                      | 4.8                      | 33.2                                     |
|                   | W            | 100                           | 68.2                       | 1.1                      | 37.5                                     |
|                   |              | 100                           | 82.0                       | 182.6                    | 53.9                                     |
|                   |              | 46                            | 259.1                      | 150.0                    | 92.0                                     |
|                   |              | 54                            | 215.3                      | 193.4                    | 112.6                                    |
|                   | S            | 100                           | 205.6                      | 353.2                    | 35.1                                     |
|                   |              | 100                           | 225.0                      | 198.4                    | 63.2                                     |
|                   |              | 100                           | 177.7                      | 354.3                    | 43.8                                     |
|                   |              | 100                           | 156.2                      | 165.8                    | 89.6                                     |

<sup>a</sup>Occupancies of hydrogen-bonding interactions were determined using a cutoff of 3.4 Å for the donor–acceptor distance and 120° for the donor–hydrogen–acceptor angle.

**Table A.6.** Occupancies (%) of the hydrogen bonds within the trimer (5'–CXC–3') containing the lesion pair (X:C) and the flanking base pairs in the B conformation of adducted DNA.<sup>a</sup>

| Base pair           | Hydrogen bond     | <sup>AN</sup> dG | <sup>1ANP</sup> dG | <sup>2ANP</sup> dG | <sup>3ABP</sup> dG | <sup>1APHN</sup> dG | <sup>3APHN</sup> dG | <sup>1AP</sup> dG |
|---------------------|-------------------|------------------|--------------------|--------------------|--------------------|---------------------|---------------------|-------------------|
| 5'–C:G <sup>b</sup> | N1(G)–H•••N3(C)   | 99.9             | 99.9               | 100.0              | 100.0              | 99.9                | 99.9                | 100.0             |
|                     | N2(G)–H•••O2(C)   | 99.8             | 99.8               | 99.8               | 99.9               | 99.9                | 99.8                | 99.9              |
|                     | N4(C)–H•••O6(G)   | 98.0             | 98.6               | 98.8               | 98.5               | 98.7                | 98.6                | 98.7              |
| X:C                 | N2(X)–H•••O2(C)   | 99.9             | 99.9               | 100.0              | 99.9               | 99.9                | 100.0               | 99.9              |
|                     | N1(X)–H•••N3(C)   | 99.9             | 99.9               | 99.9               | 99.9               | 100.0               | 99.9                | 99.4              |
|                     | N4(C)–H•••O6(X)   | 98.7             | 98.4               | 97.7               | 98.3               | 98.7                | 98.3                | 97.6              |
|                     | N10(X)–H•••O4'(X) | 81.0             | 57.1               | 74.4               | 66.0               | 64.8                | 65.5                | 63.1              |
| 3'–C:G <sup>b</sup> | N1(G)–H•••N3(C)   | 100.0            | 100.0              | 99.9               | 99.7               | 100.0               | 100.0               | 100.0             |
|                     | N2(G)–H•••O2(C)   | 99.8             | 99.9               | 99.7               | 99.6               | 99.9                | 99.8                | 99.8              |
|                     | N4(C)–H•••O6(G)   | 99.2             | 98.9               | 99.1               | 98.8               | 98.8                | 99.2                | 98.5              |

<sup>a</sup>Hydrogen-bonding interactions were determined using a cutoff of 3.4 Å for the donor–acceptor distance and 120° for the donor–hydrogen–acceptor angle. <sup>b</sup>5'–C:G and 3'–C:G represent the flanking base pairs present 5' and 3' with respect to the lesion, respectively.

**Table A.7.** Average and standard deviations (in parentheses) of pseudostep parameters at the lesion site within different conformations of N-linked adducted DNA.<sup>a</sup>

| Nucleoside          | Conformation | Shift (Å)  | Slide (Å)  | Rise (Å)  | Tilt (deg) | Roll (deg) | Twist (deg) |
|---------------------|--------------|------------|------------|-----------|------------|------------|-------------|
| dG                  | B            | 0.4 (0.8)  | −1.0 (0.9) | 6.5 (0.4) | 2.2 (5.0)  | 9.9 (7.5)  | 63.1 (6.3)  |
| <sup>AN</sup> dG    |              | 0.1 (0.8)  | −1.8 (0.9) | 6.4 (0.4) | 4.2 (4.8)  | 9.5 (7.4)  | 59.5 (5.5)  |
| <sup>1ANP</sup> dG  |              | −0.1 (0.8) | −2.4 (0.8) | 6.7 (0.4) | 2.8 (4.9)  | 8.4 (6.9)  | 59.4 (4.8)  |
| <sup>2ANP</sup> dG  |              | 0.2 (0.8)  | −1.9 (0.9) | 6.5 (0.4) | 3.9 (4.8)  | 8.9 (7.3)  | 60.7 (5.0)  |
| <sup>3ABP</sup> dG  |              | 0.3 (0.8)  | −1.6 (1.0) | 6.4 (0.4) | 4.8 (5.0)  | 8.0 (7.5)  | 60.0 (6.5)  |
| <sup>1APHN</sup> dG |              | −0.1 (0.8) | −2.7 (0.8) | 6.8 (0.5) | 2.8 (5.0)  | 7.2 (7.3)  | 59.1 (4.9)  |
| <sup>3APHN</sup> dG |              | 0.0 (0.9)  | −2.3 (0.8) | 6.6 (0.4) | 3.4 (4.9)  | 9.6 (7.3)  | 59.9 (4.9)  |
| <sup>1AP</sup> dG   |              | −0.1 (0.8) | −2.6 (0.9) | 6.8 (0.5) | 3.1 (4.9)  | 8.0 (7.1)  | 60.0 (4.9)  |
| <sup>AN</sup> dG    | W            | −0.1 (0.8) | −1.1 (0.8) | 6.2 (0.4) | 2.5 (5.2)  | 14.1 (7.2) | 59.7 (5.1)  |
| <sup>1ANP</sup> dG  |              | −0.2 (0.7) | −1.1 (0.8) | 6.2 (0.4) | 1.9 (5.0)  | 13.7 (6.4) | 58.9 (4.7)  |
| <sup>2ANP</sup> dG  |              | −0.3 (0.7) | −1.4 (0.7) | 6.2 (0.4) | 3.1 (5.0)  | 14.0 (6.4) | 59.7 (5.0)  |
| <sup>3ABP</sup> dG  |              | 0.1 (0.8)  | −1.3 (0.8) | 6.2 (0.4) | 3.0 (5.1)  | 14.7 (6.5) | 58.0 (4.7)  |
| <sup>1APHN</sup> dG |              | 0.0 (0.7)  | −0.9 (0.9) | 6.3 (0.4) | 3.0 (5.5)  | 12.0 (6.8) | 60.5 (5.5)  |
| <sup>3APHN</sup> dG |              | 0.0 (0.7)  | −1.2 (0.8) | 6.3 (0.4) | 3.0 (5.4)  | 11.7 (6.5) | 60.0 (5.4)  |
| <sup>1AP</sup> dG   |              | −0.2 (0.7) | −1.3 (0.8) | 6.2 (0.4) | 2.3 (5.1)  | 14.5 (6.9) | 59.1 (5.4)  |
| <sup>AN</sup> dG    | S            | −0.4 (0.9) | 0.9 (1.0)  | 6.9 (0.5) | 3.1 (5.8)  | 15.8 (6.7) | 42.3 (6.9)  |
| <sup>1ANP</sup> dG  |              | 0.1 (1.0)  | −3.4 (1.6) | 6.8 (0.6) | 0.9 (6.0)  | 16.7 (8.0) | 56.6 (8.0)  |
| <sup>2ANP</sup> dG  |              | −1.0 (0.9) | −2.4 (1.2) | 6.9 (0.5) | 2.2 (5.0)  | 10.4 (9.1) | 49.5 (6.0)  |
| <sup>3ABP</sup> dG  |              | −0.8 (0.9) | −1.4 (1.5) | 6.9 (0.5) | 1.6 (4.7)  | −1.0 (8.0) | 43.8 (5.9)  |
| <sup>1APHN</sup> dG |              | −0.3 (0.8) | −1.5 (1.2) | 6.6 (0.4) | 2.2 (5.2)  | −2.9 (9.5) | 41.1 (5.0)  |
| <sup>3APHN</sup> dG |              | −0.7 (0.7) | 0.3 (3.2)  | 6.9 (0.5) | 1.9 (4.8)  | −0.8 (7.1) | 44.5 (5.7)  |
| <sup>1AP</sup> dG   |              | −0.2 (0.7) | −2.9 (3.7) | 7.1 (0.7) | −1.3 (5.9) | 20.5 (7.9) | 50.7 (11.2) |

<sup>a</sup>Pseudostep parameters at the lesion site were calculated using a pseudostep consisting of the base pairs 5' and 3' with respect to the lesion.

**Table S8.** Average and standard deviations (in parentheses) of lesion-site van der Waals and electrostatic interactions (kJ mol<sup>-1</sup>) within different conformations of N-linked adducted DNA.<sup>a</sup>

| Nucleoside          | Conformation | Lesion Stacking energy with 5' bases | Lesion Stacking energy with 3' bases | Base pair energy at the 5' base pair | Base pair energy at the 3' base pair | Base pair energy at the lesion site | Lesion-site Interaction energy <sup>b</sup> |
|---------------------|--------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|-------------------------------------|---------------------------------------------|
| dG                  | B            | -61.0 (9.7)                          | -56.6 (10.5)                         | -118.2 (7.4)                         | -118.0 (7.4)                         | -118.4 (7.3)                        | -472.2 (17.7)                               |
| <sup>AN</sup> dG    |              | -65.5 (10.3)                         | -63.5 (9.8)                          | -117.7 (7.9)                         | -118.5 (7.2)                         | -110.3 (7.1)                        | -475.5 (19.2)                               |
| <sup>1ANP</sup> dG  |              | -65.6 (10.0)                         | -63.4 (10.1)                         | -118.2 (7.4)                         | -118.5 (7.3)                         | -110.5 (7.1)                        | -476.2 (19.0)                               |
| <sup>2ANP</sup> dG  |              | -66.7 (10.2)                         | -62.6 (10.0)                         | -118.6 (7.3)                         | -118.1 (7.5)                         | -109.4 (7.6)                        | -475.4 (19.2)                               |
| <sup>3ABP</sup> dG  |              | -64.7 (10.3)                         | -64.5 (9.7)                          | -118.1 (7.6)                         | -118.3 (8.0)                         | -110.0 (7.2)                        | -475.6 (19.3)                               |
| <sup>1APHN</sup> dG |              | -63.8 (10.5)                         | -64.4 (10.0)                         | -118.7 (7.2)                         | -118.4 (7.3)                         | -110.0 (6.9)                        | -475.3 (19.0)                               |
| <sup>3APHN</sup> dG |              | -66.5 (10.0)                         | -63.8 (9.9)                          | -118.1 (7.5)                         | -118.4 (7.3)                         | -109.7 (7.3)                        | -476.5 (18.9)                               |
| <sup>1AP</sup> dG   |              | -65.5 (10.4)                         | -62.5 (10.2)                         | -118.6 (7.2)                         | -117.8 (7.7)                         | -109.5 (7.9)                        | -473.9 (17.9)                               |
| <sup>AN</sup> dG    | W            | -47.4 (8.3)                          | -44.9 (10.7)                         | -117.9 (7.3)                         | -116.0 (8.2)                         | -35.8 (17.8)                        | -361.9 (24.9)                               |
| <sup>1ANP</sup> dG  |              | -51.3 (8.2)                          | -45.0 (10.2)                         | -118.0 (7.1)                         | -114.8 (8.8)                         | -33.6 (15.8)                        | -362.8 (23.4)                               |
| <sup>2ANP</sup> dG  |              | -48.6 (7.8)                          | -40.9 (9.7)                          | -118.2 (7.2)                         | -115.5 (8.4)                         | -36.6 (15.4)                        | -359.8 (22.7)                               |
| <sup>3ABP</sup> dG  |              | -52.7 (7.8)                          | -44.3 (9.9)                          | -117.3 (7.6)                         | -115.8 (8.1)                         | -41.8 (9.4)                         | -372.0 (19.2)                               |
| <sup>1APHN</sup> dG |              | -49.2 (9.3)                          | -43.0 (12.2)                         | -117.9 (7.4)                         | -113.9 (9.6)                         | -24.3 (24.9)                        | -348.3 (31.6)                               |
| <sup>3APHN</sup> dG |              | -48.2 (8.4)                          | -45.0 (13.1)                         | -118.0 (7.5)                         | -114.9 (8.7)                         | -33.7 (19.0)                        | -359.8 (27.0)                               |
| <sup>1AP</sup> dG   |              | -53.4 (7.6)                          | -43.3 (9.3)                          | -118.1 (7.2)                         | -115.1 (9.0)                         | -37.6 (9.4)                         | -367.5 (19.1)                               |
| <sup>AN</sup> dG    |              | -44.2 (11.6)                         | -20.3 (11.3)                         | -117.7 (13.4)                        | -118.5 (14.9)                        | N/A                                 | -300.6 (17.1)                               |
| <sup>1ANP</sup> dG  | S            | -51.9 (11.0)                         | -45.9 (16.2)                         | -119.1 (13.4)                        | -111.2 (19.0)                        | N/A                                 | -328.1 (19.9)                               |
| <sup>2ANP</sup> dG  |              | -52.5 (9.0)                          | -46.2 (11.3)                         | -117.8 (13.3)                        | -117.2 (13.7)                        | N/A                                 | -333.7 (16.6)                               |
| <sup>3ABP</sup> dG  |              | -44.2 (12.2)                         | -49.0 (10.8)                         | -118.1 (13.2)                        | -118.0 (13.1)                        | N/A                                 | -329.3 (17.7)                               |
| <sup>1APHN</sup> dG |              | -64.6 (13.0)                         | -51.3 (17.5)                         | -116.9 (13.5)                        | -115.0 (17.3)                        | N/A                                 | -347.8 (19.4)                               |
| <sup>3APHN</sup> dG |              | -53.3 (21.8)                         | -41.0 (8.9)                          | -114.9 (14.4)                        | -117.4 (13.7)                        | N/A                                 | -326.6 (20.5)                               |
| <sup>1AP</sup> dG   |              | -63.8 (14.3)                         | -61.6 (12.5)                         | -118.5 (13.3)                        | -114.3 (15.3)                        | N/A                                 | -358.2 (21.5)                               |

<sup>a</sup>Energies calculated with LIE of cpptraj. <sup>b</sup>Calculated as the sum of all other columns.

**Table A.9.** Occupancies (%) of the hydrogen bonds within the trimer (5'–CXC–3') containing the lesion pair (X:C) and the flanking base pairs in the W conformation of adducted DNA.<sup>a</sup>

| Base pair           | Hydrogen Bond        | <sup>AN</sup> dG | <sup>1ANP</sup> dG | <sup>2ANP</sup> dG | <sup>3ABP</sup> dG | <sup>1APHN</sup> dG | <sup>3APHN</sup> dG | <sup>1AP</sup> dG |
|---------------------|----------------------|------------------|--------------------|--------------------|--------------------|---------------------|---------------------|-------------------|
| 5'–C:G <sup>b</sup> | N1(G)–H•••N3(C)      | 99.9             | 99.9               | 99.9               | 99.9               | 99.9                | 99.9                | 99.9              |
|                     | N2(G)–H•••O2(C)      | 99.7             | 99.7               | 99.8               | 99.5               | 99.8                | 99.6                | 99.7              |
|                     | N4(C)–H•••O6(G)      | 98.7             | 98.8               | 98.7               | 98.7               | 98.7                | 98.8                | 98.9              |
| X:C                 | N4(C)–H•••O6(X)      | 78.9             | 79.7               | 76.3               | 80.1               | 93.0                | 73.0                | 75.0              |
|                     | N4(C)–H•••N7(X)      | 42.7             | 45.4               | 56.1               | 59.8               | 38.5                | 38.5                | 61.0              |
| X:3'–C <sup>c</sup> | N10(X)–H•••O4'(3'–C) | 67.4             | 48.4               | 68.0               | 30.6               | 49.7                | 41.0                | 37.3              |
| 3'–C:G <sup>b</sup> | N1(G)–H•••N3(C)      | 99.9             | 99.9               | 99.9               | 99.9               | 99.8                | 99.9                | 99.8              |
|                     | N2(G)–H•••O2(8)      | 99.3             | 99.2               | 99.3               | 99.4               | 99.4                | 99.4                | 99.0              |
|                     | N4(C)–H•••O6(G)      | 98.1             | 97.2               | 98.1               | 98.0               | 95.6                | 97.7                | 97.8              |

<sup>a</sup>Hydrogen-bonding interactions were determined using a cutoff of 3.4 Å for the donor–acceptor distance and 120° for the donor–hydrogen–acceptor angle. <sup>b</sup>5'–C:G and 3'–C:G represent the flanking base pairs present 5' and 3' with respect to the lesion, respectively. <sup>c</sup>X:3'–C represents the interaction between the lesion (X) and the intrastrand C present 3' with respect to the lesion.

**Table A.10.** Occupancies (%) of the hydrogen bonds within the trimer (5'–CXC–3') containing the lesion (X), the opposing extrahelical base C (ex-C), and the flanking base pairs in the S conformation of the adducted DNA.<sup>a</sup>

| Interaction            | Hydrogen Bond                     | <sup>AN</sup> dG | <sup>1ANP</sup> dG | <sup>2ANP</sup> dG | <sup>3ABP</sup> dG | <sup>1APHN</sup> dG | <sup>3APHN</sup> dG | <sup>1AP</sup> dG |
|------------------------|-----------------------------------|------------------|--------------------|--------------------|--------------------|---------------------|---------------------|-------------------|
| 5'–C:G <sup>b</sup>    | N1(G)–H•••N3(C)                   | 99.9             | 100.0              | 100.0              | 100.0              | 100.0               | 99.9                | 100.0             |
|                        | N2(G)–H•••O2(C)                   | 99.8             | 99.9               | 99.7               | 99.6               | 99.7                | 99.4                | 99.9              |
|                        | N4(C)–H•••O6(G)                   | 97.9             | 99.0               | 98.3               | 99.0               | 97.7                | 97.7                | 98.7              |
| exC: 5'–G <sup>c</sup> | N4(exC)–H•••OP(3'–G) <sup>b</sup> | 68.2             | 10.8               | 65.5               | 79.7               | 51.9                | 67.2                | 19.8              |
| 3'–C:G <sup>b</sup>    | N2(G)–H•••O2(C)                   | 99.7             | 99.9               | 99.9               | 100.0              | 99.9                | 96.3                | 99.9              |
|                        | N1(G)–H•••N3(C)                   | 99.2             | 93.5               | 99.8               | 99.9               | 96.4                | 67.3                | 99.7              |
|                        | N4(C)–H•••O6(17)                  | 96.5             | 87.0               | 97.3               | 98.6               | 92.9                | 66.3                | 91.5              |

<sup>a</sup>Hydrogen-bonding interactions were determined using a cutoff of 3.4 Å for the donor–acceptor distance and 120° for the donor–hydrogen–acceptor angle. <sup>b</sup>5'–C:G and 3'–C:G represent the flanking base pairs present 5'– and 3'– with respect to the lesion, respectively. <sup>c</sup>exC:5'–G represents the interaction between the extrahelical cytosine (exC) opposing the lesion, and a phosphate oxygen (OP) on the 5'–G.

**Table A.11.** Average values and standard deviation (in parentheses) for the backbone RMSD<sup>a</sup> (Å) and key dihedral angles (deg) for different conformations of adducted DNA.

| Adduct              | Conformation | RMSD      | $\chi$       | $\theta$     | $\phi$       |
|---------------------|--------------|-----------|--------------|--------------|--------------|
| <sup>AN</sup> dG    | B            | 4.0 (0.7) | 227.8 (12.7) | 139.9 (20.6) | 2.9 (12.2)   |
|                     | W            | 3.8 (0.8) | 67.7 (13.2)  | 76.1 (18.1)  | 1.2 (9.9)    |
|                     | S            | 2.5 (0.4) | 42.1 (13.2)  | 167.0 (15.9) | 357.6 (10.9) |
| <sup>1ANP</sup> dG  | B            | 4.0 (0.7) | 221.7 (13.1) | 165.3 (21.2) | 17.1 (11.0)  |
|                     | W            | 2.1 (0.3) | 66.2 (12.6)  | 70.7 (13.5)  | 359.6 (9.1)  |
|                     | S            | 2.3 (0.4) | 38.2 (15.8)  | 189.2 (18.3) | 355.6 (9.8)  |
| <sup>2ANP</sup> dG  | B            | 4.0 (0.7) | 226.8 (12.7) | 139.9 (22.3) | 5.4 (11.4)   |
|                     | W            | 1.9 (0.4) | 64.7 (12.8)  | 76.2 (14.2)  | 180.5 (9.8)  |
|                     | S            | 2.4 (0.5) | 45.2 (21.7)  | 177.1 (14.1) | 175.2 (10.5) |
| <sup>3ABP</sup> dG  | B            | 2.2 (0.5) | 230.8 (12.2) | 124.6 (13.9) | 5.3 (12.8)   |
|                     | W            | 2.2 (0.5) | 69.1 (12.6)  | 62.4 (14.7)  | 183.8 (9.7)  |
|                     | S            | 2.9 (0.5) | 119.4 (18.9) | 196.5 (13.8) | 175.9 (10.2) |
| <sup>1APHN</sup> dG | B            | 3.5 (0.6) | 216.5 (11.0) | 174.8 (12.2) | 17.8 (10.8)  |
|                     | W            | 2.3 (0.4) | 67.9 (12.9)  | 71.7 (13.9)  | 359.0 (10.4) |
|                     | S            | 3.3 (0.5) | 120.4 (23.1) | 206.6 (15.0) | 5.9 (11.9)   |
| <sup>3APHN</sup> dG | B            | 2.1 (0.4) | 226.9 (13.1) | 135.4 (25.9) | 8.6 (12.9)   |
|                     | W            | 2.4 (0.4) | 71.2 (13.4)  | 72.5 (15.0)  | 0.1 (10.4)   |
|                     | S            | 2.7 (0.5) | 132.3 (24.4) | 191.9 (18.1) | 177.5 (11.1) |
| <sup>1AP</sup> dG   | B            | 2.6 (0.5) | 218.5 (12.2) | 166.4 (20.6) | 17.6 (11.3)  |
|                     | W            | 4.1 (0.7) | 64.0 (12.5)  | 68.2 (12.9)  | 1.1 (10.2)   |
|                     | S            | 5.0 (0.7) | 48.4 (33.6)  | 205.6 (11.0) | 353.2 (11.5) |

<sup>a</sup>RMSD is calculated with respect to the first frame of the production run.

**Table A.12.** Average and standard deviations (in parentheses) of the solvent accessible surface area (SASA, Å<sup>2</sup>) of the NCA moiety within different conformations of N-linked adducted DNA.

| Nucleoside          | Conformation | SASA         |
|---------------------|--------------|--------------|
| <sup>AN</sup> dG    | B            | 116.1 (6.3)  |
| <sup>1ANP</sup> dG  |              | 137.3 (6.5)  |
| <sup>2ANP</sup> dG  |              | 147.8 (5.8)  |
| <sup>3ABP</sup> dG  |              | 191.1 (7.1)  |
| <sup>1APHN</sup> dG |              | 176.0 (6.1)  |
| <sup>3APHN</sup> dG |              | 179.5 (7.3)  |
| <sup>1AP</sup> dG   |              | 175.6 (5.3)  |
| <sup>AN</sup> dG    | W            | 92.2 (20.0)  |
| <sup>1ANP</sup> dG  |              | 115.5 (15.5) |
| <sup>2ANP</sup> dG  |              | 109.5 (17.6) |
| <sup>3ABP</sup> dG  |              | 159.7 (24.0) |
| <sup>1APHN</sup> dG |              | 153.1 (19.7) |
| <sup>3APHN</sup> dG |              | 159.7 (20.3) |
| <sup>1AP</sup> dG   |              | 155.5 (18.2) |
| <sup>AN</sup> dG    | S            | 30.7 (12.3)  |
| <sup>1ANP</sup> dG  |              | 71.2 (11.9)  |
| <sup>2ANP</sup> dG  |              | 39.8 (7.0)   |
| <sup>3ABP</sup> dG  |              | 56.9 (6.8)   |
| <sup>1APHN</sup> dG |              | 85.3 (12.3)  |
| <sup>3APHN</sup> dG |              | 76.2 (8.8)   |
| <sup>1AP</sup> dG   |              | 107.8 (11.7) |

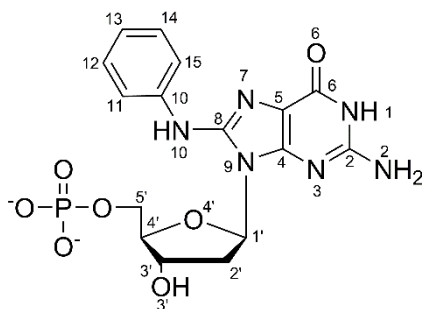


**Table A.13.** Occupancies (%) for the Watson-Crick hydrogen bonds within the base pairs present at the 5'– and 3'– terminals with respect to the lesion respectively, in three conformations of the adducted DNA.<sup>a</sup>

| Conformation | Base pair <sup>b</sup> | Hydrogen bond   | ANdG  | 1ANP dG | 2ANP dG | 3ABP dG | 1APHN dG | 3APHN dG | 1AP dG |
|--------------|------------------------|-----------------|-------|---------|---------|---------|----------|----------|--------|
| B            | 5'–G:C                 | N1(G)–H•••N3(C) | 100.0 | 100.0   | 100.0   | 100.0   | 100.0    | 100.0    | 100.0  |
|              |                        | N2(G)–H•••O2(C) | 99.9  | 99.9    | 99.9    | 99.9    | 99.9     | 99.9     | 99.9   |
|              |                        | N4(C)–H•••O6(G) | 98.9  | 98.8    | 99.0    | 99.2    | 99.1     | 98.9     | 99.1   |
|              | 3'–G:C                 | N1(G)–H•••N3(C) | 100.0 | 100.0   | 100.0   | 100.0   | 100.0    | 100.0    | 100.0  |
|              |                        | N2(G)–H•••O2(C) | 99.6  | 99.8    | 99.6    | 99.8    | 99.7     | 99.7     | 99.6   |
|              |                        | N4(C)–H•••O6(G) | 98.8  | 98.8    | 98.8    | 98.8    | 98.8     | 98.6     | 98.8   |
| W            | 5'–G:C                 | N1(G)–H•••N3(C) | 100.0 | 100.0   | 100.0   | 100.0   | 100.0    | 100.0    | 100.0  |
|              |                        | N2(G)–H•••O2(C) | 100.0 | 99.9    | 99.9    | 100.0   | 99.9     | 99.8     | 99.9   |
|              |                        | N4(C)–H•••O6(G) | 98.8  | 98.6    | 98.9    | 99.1    | 98.5     | 98.4     | 98.9   |
|              | 3'–G:C                 | N1(G)–H•••N3(C) | 100.0 | 100.0   | 100.0   | 100.0   | 100.0    | 100.0    | 100.0  |
|              |                        | N2(G)–H•••O2(C) | 99.7  | 99.8    | 99.6    | 99.8    | 99.7     | 99.7     | 99.7   |
|              |                        | N4(C)–H•••O6(G) | 98.8  | 98.8    | 98.6    | 98.9    | 98.5     | 98.5     | 98.8   |
| S            | 5'–G:C                 | N1(G)–H•••N3(C) | 100.0 | 100.0   | 100.0   | 100.0   | 100.0    | 100.0    | 100.0  |
|              |                        | N2(G)–H•••O2(C) | 99.9  | 99.8    | 99.9    | 99.9    | 99.9     | 99.9     | 99.9   |
|              |                        | N4(C)–H•••O6(G) | 99.1  | 98.8    | 98.8    | 98.7    | 98.8     | 98.8     | 99.0   |
|              | 3'–G:C                 | N1(G)–H•••N3(C) | 100.0 | 100.0   | 100.0   | 100.0   | 100.0    | 100.0    | 100.0  |
|              |                        | N2(G)–H•••O2(C) | 99.7  | 99.7    | 99.7    | 99.8    | 99.6     | 99.7     | 99.7   |
|              |                        | N4(C)–H•••O6(G) | 98.6  | 98.7    | 98.8    | 98.9    | 97.9     | 98.5     | 98.8   |

<sup>a</sup>Hydrogen-bonding interactions were determined using a cutoff of 3.4 Å for the donor–acceptor distance and 120° for the donor–hydrogen–acceptor angle. <sup>b</sup>5'–C:G and 3'–C:G present the terminal base pairs on the 5' and 3' terminal of the lesion containing strand, respectively.

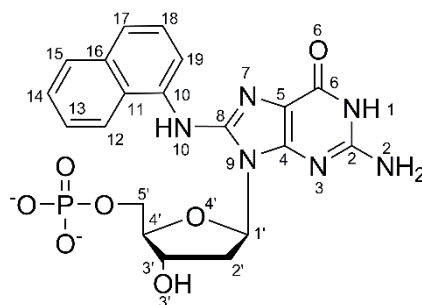
**Table A.14.** MOL2 file containing partial charges and atom types of the <sup>AN</sup>dG adduct used for MD simulations (atom numbers are provided in the associated figure).



|    |      |        |         |         |    |   |     |         |
|----|------|--------|---------|---------|----|---|-----|---------|
| 1  | P    | 4.7050 | 0.0520  | -1.1680 | P  | 1 | LIG | 1.2090  |
| 2  | O5'  | 4.1730 | 1.0760  | 0.0000  | OS | 1 | LIG | -0.4900 |
| 3  | OP1  | 4.6090 | 0.7320  | -2.4680 | O2 | 1 | LIG | -0.7864 |
| 4  | OP2  | 5.9220 | -0.5810 | -0.6380 | O2 | 1 | LIG | -0.7953 |
| 5  | O3   | 0.0000 | 0.0000  | 0.0000  | OS | 1 | LIG | -0.5474 |
| 6  | C5'  | 3.1010 | 1.9490  | -0.3030 | CI | 1 | LIG | -0.0102 |
| 7  | H5'  | 2.8330 | 2.4680  | 0.6250  | H1 | 1 | LIG | 0.0752  |
| 8  | H5'' | 3.3910 | 2.7140  | -1.0420 | H1 | 1 | LIG | 0.0752  |
| 9  | C4'  | 1.8690 | 1.2180  | -0.8380 | CT | 1 | LIG | 0.1830  |
| 10 | H4'  | 1.0270 | 1.9170  | -0.9140 | H1 | 1 | LIG | 0.0978  |
| 11 | O4'  | 2.1700 | 0.7140  | -2.1590 | OS | 1 | LIG | -0.4082 |
| 12 | C1'  | 1.9220 | -0.6850 | -2.2280 | CT | 1 | LIG | 0.1596  |
| 13 | H1'  | 0.9300 | -0.8780 | -2.6450 | H2 | 1 | LIG | 0.0902  |
| 14 | C3'  | 1.4280 | 0.0000  | 0.0000  | CT | 1 | LIG | 0.1484  |
| 15 | H3'  | 1.8240 | 0.0530  | 1.0220  | H1 | 1 | LIG | 0.0720  |
| 16 | C2'  | 2.0070 | -1.1870 | -0.7800 | CT | 1 | LIG | -0.0617 |
| 17 | H2'  | 3.0480 | -1.3520 | -0.4920 | HC | 1 | LIG | 0.0466  |
| 18 | H2'' | 1.4440 | -2.1140 | -0.6290 | HC | 1 | LIG | 0.0466  |
| 19 | N9   | 2.8790 | -1.2770 | -3.1530 | N* | 1 | LIG | -0.0563 |
| 20 | C8   | 2.6100 | -2.0090 | -4.3140 | CR | 1 | LIG | 0.4678  |
| 21 | C4   | 4.2540 | -1.2720 | -3.0400 | CB | 1 | LIG | 0.1484  |
| 22 | N7   | 3.6880 | -2.4790 | -4.8900 | NB | 1 | LIG | -0.5254 |
| 23 | C5   | 4.7300 | -2.0310 | -4.1010 | CB | 1 | LIG | 0.0762  |
| 24 | N4   | 4.9730 | -0.6260 | -2.0850 | NC | 1 | LIG | -0.5421 |
| 25 | C6   | 6.1500 | -2.2180 | -4.2500 | C  | 1 | LIG | 0.5106  |
| 26 | C2   | 6.2750 | -0.7780 | -2.1980 | CA | 1 | LIG | 0.6729  |
| 27 | O6   | 6.7970 | -2.8430 | -5.0700 | O  | 1 | LIG | -0.5407 |
| 28 | N1   | 6.8450 | -1.5130 | -3.1980 | NA | 1 | LIG | -0.4810 |
| 29 | H1   | 7.8500 | -1.6540 | -3.2180 | H  | 1 | LIG | 0.3501  |

|    |     |         |         |         |    |   |     |         |
|----|-----|---------|---------|---------|----|---|-----|---------|
| 30 | N2  | 7.0960  | -0.2220 | -1.2480 | N2 | 1 | LIG | -0.8752 |
| 31 | H21 | 8.0220  | 0.0670  | -1.5350 | H  | 1 | LIG | 0.3884  |
| 32 | H22 | 6.6240  | 0.4340  | -0.6340 | H  | 1 | LIG | 0.3884  |
| 33 | N10 | 1.3100  | -2.3020 | -4.7270 | N2 | 1 | LIG | -0.5894 |
| 34 | H10 | 1.3080  | -3.1770 | -5.2390 | H  | 1 | LIG | 0.3511  |
| 35 | C10 | 0.4170  | -1.3450 | -5.2830 | CA | 1 | LIG | 0.2156  |
| 36 | C11 | 0.6440  | 0.0360  | -5.2250 | CA | 1 | LIG | -0.0547 |
| 37 | H11 | 1.5430  | 0.4310  | -4.7640 | HA | 1 | LIG | 0.1212  |
| 38 | C15 | -0.7540 | -1.8230 | -5.8900 | CA | 1 | LIG | -0.2053 |
| 39 | H15 | -0.9360 | -2.8950 | -5.9310 | HA | 1 | LIG | 0.1458  |
| 40 | C12 | -0.2970 | 0.9160  | -5.7630 | CA | 1 | LIG | -0.2343 |
| 41 | H12 | -0.1060 | 1.9850  | -5.7080 | HA | 1 | LIG | 0.1637  |
| 42 | C14 | -1.6780 | -0.9360 | -6.4340 | CA | 1 | LIG | -0.1747 |
| 43 | H14 | -2.5760 | -1.3250 | -6.9060 | HA | 1 | LIG | 0.1548  |
| 44 | C13 | -1.4580 | 0.4430  | -6.3720 | CA | 1 | LIG | -0.1124 |
| 45 | H13 | -2.1820 | 1.1350  | -6.7910 | HA | 1 | LIG | 0.1321  |

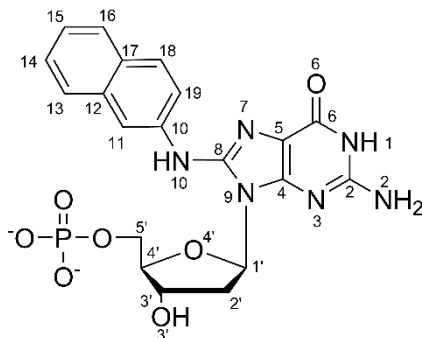
**Table A.15.** MOL2 file containing partial charges and atom types of the <sup>1</sup>ANP dG adduct used for MD simulations (atom numbers are provided in the associated figure).



|    |      |        |         |         |    |   |     |         |
|----|------|--------|---------|---------|----|---|-----|---------|
| 1  | P    | 4.9460 | 0.3260  | -1.3040 | P  | 1 | LIG | 1.1661  |
| 2  | O5'  | 4.1310 | 1.0120  | 0.0000  | OS | 1 | LIG | -0.4975 |
| 3  | OP1  | 5.0700 | 1.3460  | -2.3940 | O2 | 1 | LIG | -0.7680 |
| 4  | OP2  | 6.0760 | -0.4560 | -0.7070 | O2 | 1 | LIG | -0.7680 |
| 5  | O3'  | 0.0000 | 0.0000  | 0.0000  | OS | 1 | LIG | -0.5419 |
| 6  | C5'  | 3.0960 | 1.8960  | -0.3140 | CI | 1 | LIG | -0.0178 |
| 7  | H5'  | 3.4180 | 2.6460  | -1.0320 | H1 | 1 | LIG | 0.0805  |
| 8  | H5'' | 2.8360 | 2.4120  | 0.6020  | H1 | 1 | LIG | 0.0805  |
| 9  | C4'  | 1.8580 | 1.1950  | -0.8480 | CT | 1 | LIG | 0.1759  |
| 10 | H4'  | 1.0450 | 1.9050  | -0.9320 | H1 | 1 | LIG | 0.1016  |
| 11 | O4'  | 2.1310 | 0.6740  | -2.1470 | OS | 1 | LIG | -0.4096 |
| 12 | C1'  | 1.8680 | -0.6970 | -2.2080 | CT | 1 | LIG | 0.1529  |

|    |      |         |         |         |    |   |     |         |
|----|------|---------|---------|---------|----|---|-----|---------|
| 13 | H1'  | 0.8770  | -0.8680 | -2.5990 | H2 | 1 | LIG | 0.0818  |
| 14 | C3'  | 1.4070  | 0.0000  | 0.0000  | CT | 1 | LIG | 0.1305  |
| 15 | H3'  | 1.7900  | 0.0630  | 1.0110  | H1 | 1 | LIG | 0.0783  |
| 16 | C2'  | 1.9740  | -1.1870 | -0.7680 | CT | 1 | LIG | -0.0617 |
| 17 | H2'  | 3.0120  | -1.3450 | -0.5030 | HC | 1 | LIG | 0.0463  |
| 18 | H2'' | 1.4240  | -2.1060 | -0.6070 | HC | 1 | LIG | 0.0463  |
| 19 | N    | 2.7840  | -1.3310 | -3.1400 | N* | 1 | LIG | -0.0146 |
| 20 | C    | 4.1010  | -1.0480 | -3.4690 | CR | 1 | LIG | 0.2646  |
| 21 | C6   | 2.4920  | -2.4870 | -3.8070 | CB | 1 | LIG | 0.1347  |
| 22 | N1   | 4.6200  | -1.9060 | -4.2590 | NB | 1 | LIG | -0.4291 |
| 23 | C7   | 3.6230  | -2.8360 | -4.4770 | CB | 1 | LIG | 0.0774  |
| 24 | N2   | 1.2990  | -3.1310 | -3.7570 | NC | 1 | LIG | -0.5059 |
| 25 | C8   | 3.6080  | -4.0190 | -5.2880 | C  | 1 | LIG | 0.4773  |
| 26 | C9   | 1.2730  | -4.1920 | -4.4890 | CA | 1 | LIG | 0.6026  |
| 27 | O5   | 4.4510  | -4.5170 | -5.9710 | O  | 1 | LIG | -0.5400 |
| 28 | N3   | 2.3330  | -4.6330 | -5.2100 | NA | 1 | LIG | -0.4049 |
| 29 | H    | 2.2290  | -5.4250 | -5.8090 | H  | 1 | LIG | 0.3318  |
| 30 | N4   | 0.1230  | -4.9160 | -4.5980 | N2 | 1 | LIG | -0.8791 |
| 31 | H6   | 0.2020  | -5.8970 | -4.7470 | H  | 1 | LIG | 0.3940  |
| 32 | H7   | -0.5810 | -4.6540 | -3.9430 | H  | 1 | LIG | 0.3940  |
| 33 | N5   | 4.7320  | 0.0820  | -2.9520 | N2 | 1 | LIG | -0.1783 |
| 34 | H8   | 4.1240  | 0.8710  | -3.0190 | H  | 1 | LIG | 0.2217  |
| 35 | C10  | 6.0550  | 0.3790  | -3.4120 | CM | 1 | LIG | -0.0616 |
| 36 | C11  | 7.1720  | -0.2710 | -2.8070 | CM | 1 | LIG | 0.0743  |
| 37 | C12  | 6.2440  | 1.3280  | -4.3640 | CM | 1 | LIG | -0.0968 |
| 38 | H9   | 5.3930  | 1.8050  | -4.8190 | HA | 1 | LIG | 0.1341  |
| 39 | C13  | 8.4670  | 0.0840  | -3.2370 | CA | 1 | LIG | 0.1635  |
| 40 | C14  | 7.0360  | -1.2460 | -1.7850 | CA | 1 | LIG | -0.2174 |
| 41 | H10  | 6.0550  | -1.5200 | -1.4480 | HA | 1 | LIG | 0.1272  |
| 42 | C15  | 7.5470  | 1.6890  | -4.7810 | CA | 1 | LIG | -0.1760 |
| 43 | H11  | 7.6660  | 2.4430  | -5.5370 | HA | 1 | LIG | 0.1573  |
| 44 | C16  | 9.5900  | -0.5530 | -2.6450 | CA | 1 | LIG | -0.2093 |
| 45 | H12  | 10.5740 | -0.2810 | -2.9810 | HA | 1 | LIG | 0.1434  |
| 46 | C17  | 8.6280  | 1.0790  | -4.2380 | CA | 1 | LIG | -0.2327 |
| 47 | H13  | 9.6220  | 1.3420  | -4.5540 | HA | 1 | LIG | 0.1526  |
| 48 | C18  | 8.1300  | -1.8340 | -1.2340 | CA | 1 | LIG | -0.1129 |
| 49 | H14  | 8.0110  | -2.5710 | -0.4600 | HA | 1 | LIG | 0.1354  |
| 50 | C19  | 9.4290  | -1.4860 | -1.6720 | CA | 1 | LIG | -0.1435 |
| 51 | H15  | 10.2850 | -1.9630 | -1.2300 | HA | 1 | LIG | 0.1397  |

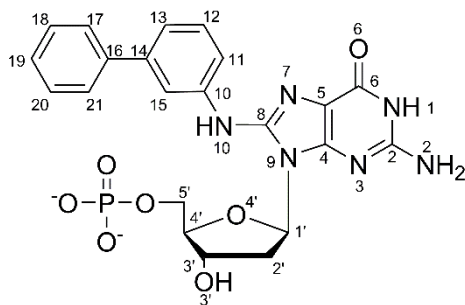
**Table A.16.** MOL2 file containing partial charges and atom types of the <sup>2</sup>ANP dG adduct used for MD simulations (atom numbers are provided in the associated figure).



|    |      |        |         |         |    |   |     |         |
|----|------|--------|---------|---------|----|---|-----|---------|
| 1  | P    | 4.9500 | 0.3390  | -1.3140 | P  | 1 | LIG | 1.1662  |
| 2  | O5'  | 4.1340 | 1.0060  | 0.0000  | OS | 1 | LIG | -0.4947 |
| 3  | OP1  | 5.0740 | 1.3760  | -2.3880 | O2 | 1 | LIG | -0.7681 |
| 4  | OP2  | 6.0790 | -0.4520 | -0.7290 | O2 | 1 | LIG | -0.7681 |
| 5  | O3   | 0.0000 | 0.0000  | 0.0000  | OS | 1 | LIG | -0.5427 |
| 6  | C5'  | 3.1000 | 1.8950  | -0.3000 | CI | 1 | LIG | -0.0213 |
| 7  | H5'  | 3.4210 | 2.6520  | -1.0120 | H1 | 1 | LIG | 0.0819  |
| 8  | H5'' | 2.8450 | 2.4020  | 0.6220  | H1 | 1 | LIG | 0.0819  |
| 9  | C4'  | 1.8570 | 1.2030  | -0.8360 | CT | 1 | LIG | 0.1693  |
| 10 | H4'  | 1.0460 | 1.9160  | -0.9090 | H1 | 1 | LIG | 0.1034  |
| 11 | O4'  | 2.1200 | 0.6960  | -2.1430 | OS | 1 | LIG | -0.4066 |
| 12 | C1'  | 1.8670 | -0.6760 | -2.2160 | CT | 1 | LIG | 0.1659  |
| 13 | H1'  | 0.8770 | -0.8500 | -2.6110 | H2 | 1 | LIG | 0.0788  |
| 14 | C3'  | 1.4070 | 0.0000  | 0.0000  | CT | 1 | LIG | 0.1359  |
| 15 | H3'  | 1.7900 | 0.0530  | 1.0120  | H1 | 1 | LIG | 0.0773  |
| 16 | C2'  | 1.9750 | -1.1790 | -0.7800 | CT | 1 | LIG | -0.0594 |
| 17 | H2'  | 3.0120 | -1.3380 | -0.5180 | HC | 1 | LIG | 0.0441  |
| 18 | H2'' | 1.4250 | -2.1000 | -0.6270 | HC | 1 | LIG | 0.0441  |
| 19 | N9   | 2.7890 | -1.3010 | -3.1500 | N* | 1 | LIG | -0.0377 |
| 20 | C8   | 4.1030 | -1.0080 | -3.4810 | CR | 1 | LIG | 0.4496  |
| 21 | C4   | 2.5170 | -2.4780 | -3.7870 | CB | 1 | LIG | 0.1103  |
| 22 | N7   | 4.6370 | -1.8800 | -4.2460 | NB | 1 | LIG | -0.5057 |
| 23 | C5   | 3.6540 | -2.8290 | -4.4440 | CB | 1 | LIG | 0.0865  |
| 24 | N3   | 1.3330 | -3.1380 | -3.7240 | NC | 1 | LIG | -0.5039 |
| 25 | C6   | 3.6620 | -4.0360 | -5.2180 | C  | 1 | LIG | 0.5121  |
| 26 | C2   | 1.3250 | -4.2180 | -4.4270 | CA | 1 | LIG | 0.6173  |
| 27 | O6   | 4.5190 | -4.5440 | -5.8760 | O  | 1 | LIG | -0.5469 |
| 28 | N1   | 2.3950 | -4.6650 | -5.1310 | NA | 1 | LIG | -0.4417 |

|    |     |         |         |         |    |   |     |         |
|----|-----|---------|---------|---------|----|---|-----|---------|
| 29 | H1  | 2.3050  | -5.4750 | -5.7070 | H  | 1 | LIG | 0.3416  |
| 30 | N2  | 0.1860  | -4.9610 | -4.5210 | N2 | 1 | LIG | -0.8699 |
| 31 | H21 | 0.2770  | -5.9440 | -4.6480 | H  | 1 | LIG | 0.3898  |
| 32 | H22 | -0.5260 | -4.6910 | -3.8780 | H  | 1 | LIG | 0.3898  |
| 33 | N10 | 4.7040  | 0.1470  | -2.9940 | N2 | 1 | LIG | -0.4461 |
| 34 | H10 | 4.0660  | 0.9130  | -3.0200 | H  | 1 | LIG | 0.2949  |
| 35 | C10 | 6.0190  | 0.5050  | -3.3890 | CA | 1 | LIG | 0.0740  |
| 36 | C11 | 6.2680  | 1.7590  | -3.8480 | CA | 1 | LIG | -0.1637 |
| 37 | H11 | 5.4580  | 2.4550  | -3.9930 | HA | 1 | LIG | 0.1344  |
| 38 | C19 | 7.0920  | -0.4100 | -3.2090 | CA | 1 | LIG | -0.1758 |
| 39 | H19 | 6.8840  | -1.4000 | -2.8560 | HA | 1 | LIG | 0.1831  |
| 40 | C12 | 7.5910  | 2.1810  | -4.1490 | CA | 1 | LIG | 0.0512  |
| 41 | C18 | 8.3590  | -0.0320 | -3.5020 | CA | 1 | LIG | -0.2300 |
| 42 | H18 | 9.1660  | -0.7300 | -3.3690 | HA | 1 | LIG | 0.1580  |
| 43 | C17 | 8.6530  | 1.2740  | -3.9790 | CA | 1 | LIG | 0.1844  |
| 44 | C13 | 7.8740  | 3.4880  | -4.6240 | CA | 1 | LIG | -0.2158 |
| 45 | H13 | 7.0630  | 4.1820  | -4.7560 | HA | 1 | LIG | 0.1417  |
| 46 | C16 | 9.9740  | 1.6910  | -4.2850 | CA | 1 | LIG | -0.2256 |
| 47 | H16 | 10.7820 | 0.9920  | -4.1540 | HA | 1 | LIG | 0.1487  |
| 48 | C14 | 9.1500  | 3.8610  | -4.9100 | CA | 1 | LIG | -0.1211 |
| 49 | H14 | 9.3540  | 4.8530  | -5.2700 | HA | 1 | LIG | 0.1365  |
| 50 | C15 | 10.2180 | 2.9490  | -4.7380 | CA | 1 | LIG | -0.1481 |
| 51 | H15 | 11.2220 | 3.2570  | -4.9690 | HA | 1 | LIG | 0.1401  |

**Table A.17.** MOL2 file containing partial charges and atom types of the <sup>3</sup>ABP dG adduct used for MD simulations (atom numbers are provided in the associated figure).

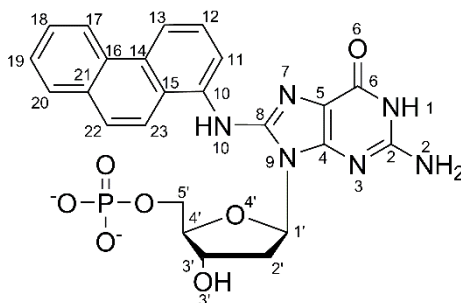


|   |     |        |         |         |    |   |     |         |
|---|-----|--------|---------|---------|----|---|-----|---------|
| 1 | P   | 4.9540 | 0.3210  | -1.2910 | P  | 1 | LIG | 1.1662  |
| 2 | O5' | 4.1290 | 1.0200  | 0.0000  | OS | 1 | LIG | -0.4955 |
| 3 | OP1 | 5.0810 | 1.3290  | -2.3930 | O2 | 1 | LIG | -0.7681 |
| 4 | OP2 | 6.0840 | -0.4490 | -0.6780 | O2 | 1 | LIG | -0.7681 |
| 5 | O3' | 0.0000 | 0.0000  | 0.0000  | OS | 1 | LIG | -0.5421 |
| 6 | C5' | 3.0930 | 1.8970  | -0.3300 | CI | 1 | LIG | -0.0196 |

|    |      |         |         |         |    |   |     |         |
|----|------|---------|---------|---------|----|---|-----|---------|
| 7  | H5'1 | 3.4150  | 2.6390  | -1.0570 | H1 | 1 | LIG | 0.0822  |
| 8  | H5'2 | 2.8290  | 2.4240  | 0.5790  | H1 | 1 | LIG | 0.0822  |
| 9  | C4'  | 1.8570  | 1.1880  | -0.8590 | CT | 1 | LIG | 0.1652  |
| 10 | H4'  | 1.0430  | 1.8960  | -0.9510 | H1 | 1 | LIG | 0.1045  |
| 11 | O4'  | 2.1310  | 0.6550  | -2.1530 | OS | 1 | LIG | -0.4045 |
| 12 | C1'  | 1.8680  | -0.7170 | -2.2030 | CT | 1 | LIG | 0.1618  |
| 13 | H1'  | 0.8770  | -0.8900 | -2.5920 | H2 | 1 | LIG | 0.0837  |
| 14 | C3'  | 1.4070  | 0.0000  | 0.0000  | CT | 1 | LIG | 0.1348  |
| 15 | H3'  | 1.7900  | 0.0720  | 1.0110  | H1 | 1 | LIG | 0.0780  |
| 16 | C2'  | 1.9750  | -1.1940 | -0.7580 | CT | 1 | LIG | -0.0575 |
| 17 | H2'1 | 3.0130  | -1.3480 | -0.4930 | HC | 1 | LIG | 0.0444  |
| 18 | H2'2 | 1.4250  | -2.1110 | -0.5880 | HC | 1 | LIG | 0.0444  |
| 19 | N9   | 2.7840  | -1.3660 | -3.1260 | N* | 1 | LIG | -0.0535 |
| 20 | C8   | 4.1050  | -1.0990 | -3.4580 | CR | 1 | LIG | 0.4020  |
| 21 | C4   | 2.5000  | -2.5600 | -3.7280 | CB | 1 | LIG | 0.1306  |
| 22 | N7   | 4.6320  | -2.0050 | -4.1890 | NB | 1 | LIG | -0.4841 |
| 23 | C5   | 3.6370  | -2.9460 | -4.3650 | CB | 1 | LIG | 0.0745  |
| 24 | N3   | 1.3090  | -3.2040 | -3.6490 | NC | 1 | LIG | -0.4981 |
| 25 | C6   | 3.6350  | -4.1770 | -5.0980 | C  | 1 | LIG | 0.5012  |
| 26 | C2   | 1.2920  | -4.3060 | -4.3170 | CA | 1 | LIG | 0.6022  |
| 27 | O6   | 4.4910  | -4.7170 | -5.7330 | O  | 1 | LIG | -0.5467 |
| 28 | N1   | 2.3600  | -4.7880 | -5.0000 | NA | 1 | LIG | -0.4250 |
| 29 | H1   | 2.2640  | -5.6160 | -5.5490 | H  | 1 | LIG | 0.3383  |
| 30 | N2   | 0.1440  | -5.0380 | -4.3930 | N2 | 1 | LIG | -0.8680 |
| 31 | H21  | 0.2250  | -6.0270 | -4.4870 | H  | 1 | LIG | 0.3898  |
| 32 | H22  | -0.5680 | -4.7390 | -3.7640 | H  | 1 | LIG | 0.3898  |
| 33 | N10  | 4.7040  | 0.0670  | -3.0080 | N2 | 1 | LIG | -0.3808 |
| 34 | H10  | 4.0510  | 0.8200  | -3.0030 | H  | 1 | LIG | 0.2905  |
| 35 | C10  | 6.0120  | 0.4610  | -3.3830 | CA | 1 | LIG | 0.0469  |
| 36 | C15  | 6.2360  | 1.8120  | -3.6080 | CA | 1 | LIG | -0.1137 |
| 37 | H15  | 5.4080  | 2.4990  | -3.5780 | HA | 1 | LIG | 0.1068  |
| 38 | C11  | 7.0730  | -0.4330 | -3.4560 | CA | 1 | LIG | -0.0925 |
| 39 | H11  | 6.9110  | -1.4790 | -3.2990 | HA | 1 | LIG | 0.1433  |
| 40 | C14  | 7.5060  | 2.2950  | -3.9050 | cp | 1 | LIG | -0.0167 |
| 41 | C12  | 8.3330  | 0.0450  | -3.7620 | CA | 1 | LIG | -0.2240 |
| 42 | H12  | 9.1540  | -0.6480 | -3.8190 | HA | 1 | LIG | 0.1649  |
| 43 | C16  | 7.7230  | 3.7520  | -4.1390 | cp | 1 | LIG | 0.0101  |
| 44 | C13  | 8.5600  | 1.3930  | -3.9840 | CA | 1 | LIG | -0.0936 |
| 45 | H13  | 9.5530  | 1.7430  | -4.1980 | HA | 1 | LIG | 0.1157  |
| 46 | C17  | 7.1880  | 4.7030  | -3.2750 | CA | 1 | LIG | -0.0792 |
| 47 | H17  | 6.6280  | 4.3820  | -2.4150 | HA | 1 | LIG | 0.1073  |

|    |     |        |        |         |    |   |     |         |
|----|-----|--------|--------|---------|----|---|-----|---------|
| 48 | C21 | 8.4700 | 4.1930 | -5.2290 | CA | 1 | LIG | -0.0822 |
| 49 | H21 | 8.8760 | 3.4750 | -5.9180 | HA | 1 | LIG | 0.1165  |
| 50 | C18 | 7.3910 | 6.0550 | -3.4940 | CA | 1 | LIG | -0.1877 |
| 51 | H18 | 6.9760 | 6.7740 | -2.8100 | HA | 1 | LIG | 0.1447  |
| 52 | C20 | 8.6730 | 5.5450 | -5.4490 | CA | 1 | LIG | -0.1870 |
| 53 | H20 | 9.2470 | 5.8650 | -6.3000 | HA | 1 | LIG | 0.1448  |
| 54 | C19 | 8.1350 | 6.4820 | -4.5820 | CA | 1 | LIG | -0.1130 |
| 55 | H19 | 8.2930 | 7.5320 | -4.7520 | HA | 1 | LIG | 0.1337  |

**Table A.18.** MOL2 file containing partial charges and atom types of the <sup>1</sup>APHN dG adduct used for MD simulations (atom numbers are provided in the associated figure).

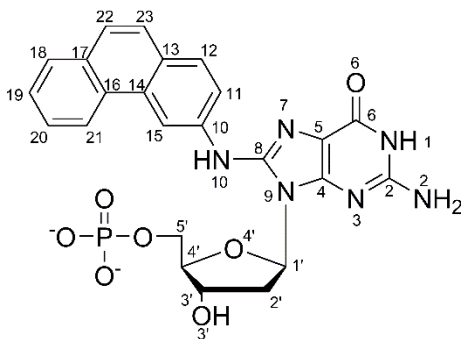


|    |      |        |         |         |    |   |     |         |
|----|------|--------|---------|---------|----|---|-----|---------|
| 1  | P    | 4.9390 | 0.3380  | -1.3210 | P  | 1 | LIG | 1.1662  |
| 2  | O5'  | 4.1360 | 1.0060  | 0.0000  | OS | 1 | LIG | -0.4960 |
| 3  | OP1  | 5.0550 | 1.3730  | -2.3970 | O2 | 1 | LIG | -0.7680 |
| 4  | OP2  | 6.0730 | -0.4550 | -0.7460 | O2 | 1 | LIG | -0.7680 |
| 5  | O3   | 0.0000 | 0.0000  | 0.0000  | OS | 1 | LIG | -0.5443 |
| 6  | C5'  | 3.1000 | 1.8960  | -0.2910 | CI | 1 | LIG | -0.0174 |
| 7  | H5'  | 3.4190 | 2.6590  | -0.9980 | H1 | 1 | LIG | 0.0811  |
| 8  | H5'' | 2.8460 | 2.3970  | 0.6360  | H1 | 1 | LIG | 0.0811  |
| 9  | C4'  | 1.8580 | 1.2080  | -0.8300 | CT | 1 | LIG | 0.1705  |
| 10 | H4'  | 1.0470 | 1.9210  | -0.8980 | H1 | 1 | LIG | 0.1016  |
| 11 | O4'  | 2.1200 | 0.7070  | -2.1400 | OS | 1 | LIG | -0.4090 |
| 12 | C1'  | 1.8680 | -0.6660 | -2.2190 | CT | 1 | LIG | 0.1650  |
| 13 | H1'  | 0.8790 | -0.8380 | -2.6160 | H2 | 1 | LIG | 0.0790  |
| 14 | C3'  | 1.4070 | 0.0000  | 0.0000  | CT | 1 | LIG | 0.1414  |
| 15 | H3'  | 1.7900 | 0.0480  | 1.0120  | H1 | 1 | LIG | 0.0750  |
| 16 | C2'  | 1.9750 | -1.1760 | -0.7860 | CT | 1 | LIG | -0.0599 |
| 17 | H2'  | 3.0120 | -1.3360 | -0.5240 | HC | 1 | LIG | 0.0433  |
| 18 | H2'' | 1.4250 | -2.0970 | -0.6370 | HC | 1 | LIG | 0.0433  |
| 19 | N9   | 2.7920 | -1.2820 | -3.1550 | N* | 1 | LIG | -0.0483 |
| 20 | C8   | 4.1100 | -0.9870 | -3.4700 | CR | 1 | LIG | 0.3864  |
| 21 | C4   | 2.5270 | -2.4550 | -3.8040 | CB | 1 | LIG | 0.1609  |
| 22 | N7   | 4.6520 | -1.8510 | -4.2380 | NB | 1 | LIG | -0.4514 |
| 23 | C5   | 3.6710 | -2.7980 | -4.4530 | CB | 1 | LIG | 0.0424  |



|    |     |         |         |         |    |   |     |         |
|----|-----|---------|---------|---------|----|---|-----|---------|
| 24 | N3  | 1.3430  | -3.1170 | -3.7540 | NC | 1 | LIG | -0.5186 |
| 25 | C6  | 3.6850  | -3.9970 | -5.2400 | C  | 1 | LIG | 0.5168  |
| 26 | C2  | 1.3430  | -4.1910 | -4.4660 | CA | 1 | LIG | 0.6210  |
| 27 | O6  | 4.5450  | -4.4980 | -5.8990 | O  | 1 | LIG | -0.5494 |
| 28 | N1  | 2.4180  | -4.6300 | -5.1660 | NA | 1 | LIG | -0.4362 |
| 29 | H1  | 2.3330  | -5.4340 | -5.7500 | H  | 1 | LIG | 0.3377  |
| 30 | N2  | 0.2050  | -4.9360 | -4.5750 | N2 | 1 | LIG | -0.8693 |
| 31 | H21 | 0.3020  | -5.9190 | -4.7020 | H  | 1 | LIG | 0.3884  |
| 32 | H22 | -0.5100 | -4.6730 | -3.9330 | H  | 1 | LIG | 0.3884  |
| 33 | N10 | 4.7180  | 0.1580  | -2.9650 | N2 | 1 | LIG | -0.4445 |
| 34 | H10 | 4.0870  | 0.9300  | -2.9790 | H  | 1 | LIG | 0.3021  |
| 35 | C10 | 6.0390  | 0.4770  | -3.4140 | CA | 1 | LIG | 0.0699  |
| 36 | C15 | 6.2480  | 1.3350  | -4.5210 | CA | 1 | LIG | 0.0883  |
| 37 | C11 | 7.0970  | -0.0340 | -2.7190 | CA | 1 | LIG | -0.1367 |
| 38 | H11 | 6.9150  | -0.6850 | -1.8850 | HA | 1 | LIG | 0.1436  |
| 39 | C14 | 7.5620  | 1.6900  | -4.8780 | CA | 1 | LIG | 0.0228  |
| 40 | C23 | 5.1440  | 1.8530  | -5.2910 | CA | 1 | LIG | -0.1798 |
| 41 | H23 | 4.1430  | 1.5500  | -5.0490 | HA | 1 | LIG | 0.1390  |
| 42 | C12 | 8.4040  | 0.2890  | -3.0970 | CA | 1 | LIG | -0.1725 |
| 43 | H12 | 9.2330  | -0.1240 | -2.5510 | HA | 1 | LIG | 0.1602  |
| 44 | C16 | 7.7770  | 2.6060  | -5.9960 | CA | 1 | LIG | 0.0658  |
| 45 | C13 | 8.6300  | 1.1380  | -4.1430 | CA | 1 | LIG | -0.2018 |
| 46 | H13 | 9.6430  | 1.3750  | -4.4010 | HA | 1 | LIG | 0.1547  |
| 47 | C22 | 5.3460  | 2.6840  | -6.3220 | CA | 1 | LIG | -0.2403 |
| 48 | H22 | 4.5110  | 3.0530  | -6.8920 | HA | 1 | LIG | 0.1586  |
| 49 | C21 | 6.6690  | 3.0960  | -6.7040 | CA | 1 | LIG | 0.1173  |
| 50 | C17 | 9.0590  | 3.0330  | -6.4040 | CA | 1 | LIG | -0.1932 |
| 51 | H17 | 9.9330  | 2.6860  | -5.8890 | HA | 1 | LIG | 0.1406  |
| 52 | C20 | 6.8570  | 3.9810  | -7.7840 | CA | 1 | LIG | -0.1752 |
| 53 | H20 | 5.9930  | 4.3420  | -8.3140 | HA | 1 | LIG | 0.1388  |
| 54 | C18 | 9.2230  | 3.8930  | -7.4540 | CA | 1 | LIG | -0.1292 |
| 55 | H18 | 10.2120 | 4.2010  | -7.7420 | HA | 1 | LIG | 0.1390  |
| 56 | C29 | 8.1100  | 4.3760  | -8.1570 | CA | 1 | LIG | -0.1704 |
| 57 | H19 | 8.2450  | 5.0520  | -8.9830 | HA | 1 | LIG | 0.1492  |

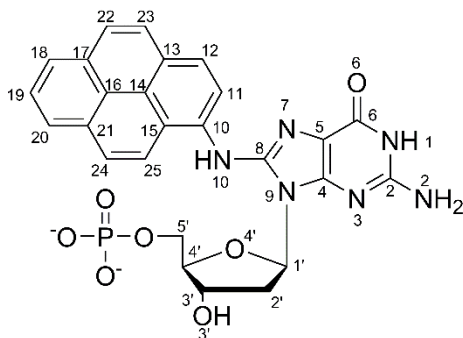
**Table A.19.** MOL2 file containing partial charges and atom types of the <sup>3</sup>APHN dG adduct used for MD simulations (atom numbers are provided in the associated figure).



|    |      |        |         |         |    |   |     |         |
|----|------|--------|---------|---------|----|---|-----|---------|
| 1  | P    | 4.7030 | 0.0550  | -1.1760 | P  | 1 | LIG | 1.2091  |
| 2  | O5'  | 4.1740 | 1.0720  | 0.0000  | OS | 1 | LIG | -0.4894 |
| 3  | OP1  | 4.6050 | 0.7450  | -2.4710 | O2 | 1 | LIG | -0.7862 |
| 4  | OP2  | 5.9200 | -0.5830 | -0.6540 | O2 | 1 | LIG | -0.7955 |
| 5  | O3   | 0.0000 | 0.0000  | 0.0000  | OS | 1 | LIG | -0.5485 |
| 6  | C5'  | 3.1030 | 1.9480  | -0.2950 | CI | 1 | LIG | -0.0049 |
| 7  | H5'  | 2.8380 | 2.4620  | 0.6370  | H1 | 1 | LIG | 0.0737  |
| 8  | H5'' | 3.3920 | 2.7170  | -1.0290 | H1 | 1 | LIG | 0.0737  |
| 9  | C4'  | 1.8690 | 1.2220  | -0.8310 | CT | 1 | LIG | 0.1806  |
| 10 | H4'  | 1.0280 | 1.9230  | -0.9010 | H1 | 1 | LIG | 0.0973  |
| 11 | O4'  | 2.1640 | 0.7260  | -2.1570 | OS | 1 | LIG | -0.4092 |
| 12 | C1'  | 1.9180 | -0.6730 | -2.2330 | CT | 1 | LIG | 0.1631  |
| 13 | H1'  | 0.9240 | -0.8650 | -2.6470 | H2 | 1 | LIG | 0.0901  |
| 14 | C3'  | 1.4280 | 0.0000  | 0.0000  | CT | 1 | LIG | 0.1539  |
| 15 | H3'  | 1.8240 | 0.0460  | 1.0220  | H1 | 1 | LIG | 0.0696  |
| 16 | C2'  | 2.0060 | -1.1830 | -0.7870 | CT | 1 | LIG | -0.0625 |
| 17 | H2'  | 3.0470 | -1.3500 | -0.5020 | HC | 1 | LIG | 0.0462  |
| 18 | H2'' | 1.4430 | -2.1100 | -0.6400 | HC | 1 | LIG | 0.0462  |
| 19 | N9   | 2.8730 | -1.2580 | -3.1640 | N* | 1 | LIG | -0.0458 |
| 20 | C8   | 2.6000 | -1.9810 | -4.3310 | CR | 1 | LIG | 0.4616  |
| 21 | C4   | 4.2480 | -1.2530 | -3.0550 | CB | 1 | LIG | 0.1258  |
| 22 | N7   | 3.6770 | -2.4460 | -4.9130 | NB | 1 | LIG | -0.5189 |
| 23 | C5   | 4.7210 | -2.0040 | -4.1240 | CB | 1 | LIG | 0.0919  |
| 24 | N3   | 4.9690 | -0.6150 | -2.0960 | NC | 1 | LIG | -0.5254 |
| 25 | C6   | 6.1410 | -2.1900 | -4.2770 | C  | 1 | LIG | 0.5023  |
| 26 | C2   | 6.2710 | -0.7660 | -2.2140 | CA | 1 | LIG | 0.6442  |
| 27 | O6   | 6.7850 | -2.8100 | -5.1030 | O  | 1 | LIG | -0.5410 |
| 28 | N1   | 6.8390 | -1.4930 | -3.2220 | NA | 1 | LIG | -0.4656 |
| 29 | H1   | 7.8430 | -1.6340 | -3.2440 | H  | 1 | LIG | 0.3496  |
| 30 | N2   | 7.0940 | -0.2180 | -1.2620 | N2 | 1 | LIG | -0.8684 |
| 31 | H21  | 8.0210 | 0.0700  | -1.5460 | H  | 1 | LIG | 0.3883  |

|    |      |         |         |         |    |   |     |         |
|----|------|---------|---------|---------|----|---|-----|---------|
| 32 | H22  | 6.6230  | 0.4330  | -0.6410 | H  | 1 | LIG | 0.3883  |
| 33 | N10  | 1.3000  | -2.2720 | -4.7430 | N2 | 1 | LIG | -0.5844 |
| 34 | H10  | 1.2910  | -3.1520 | -5.2460 | H  | 1 | LIG | 0.3515  |
| 35 | C10  | 0.4040  | -1.3200 | -5.2950 | CA | 1 | LIG | 0.2030  |
| 36 | C11  | 0.6360  | 0.0710  | -5.2180 | CA | 1 | LIG | -0.0909 |
| 37 | H11  | 1.5320  | 0.4600  | -4.7470 | HA | 1 | LIG | 0.1418  |
| 38 | C15  | -0.7530 | -1.7900 | -5.9070 | CA | 1 | LIG | -0.2595 |
| 39 | H15  | -0.9090 | -2.8640 | -5.9430 | HA | 1 | LIG | 0.1563  |
| 40 | C12  | -0.2930 | 0.9410  | -5.7500 | CA | 1 | LIG | -0.2651 |
| 41 | H12  | -0.1140 | 2.0120  | -5.6900 | HA | 1 | LIG | 0.1669  |
| 42 | C14  | -1.7080 | -0.9190 | -6.4670 | CA | 1 | LIG | 0.0047  |
| 43 | C13  | -1.4710 | 0.4860  | -6.3810 | CA | 1 | LIG | 0.1187  |
| 44 | C16  | -2.9230 | -1.3900 | -7.1230 | CA | 1 | LIG | 0.0738  |
| 45 | C22  | -2.4200 | 1.4080  | -6.9330 | CA | 1 | LIG | -0.1827 |
| 46 | H22' | -2.2110 | 2.4720  | -6.8510 | HA | 1 | LIG | 0.1524  |
| 47 | C17  | -3.8380 | -0.4320 | -7.6570 | CA | 1 | LIG | 0.1448  |
| 48 | C21  | -3.2440 | -2.7600 | -7.2590 | CA | 1 | LIG | -0.2055 |
| 49 | H21' | -2.5720 | -3.5150 | -6.8670 | HA | 1 | LIG | 0.1475  |
| 50 | C23  | -3.5530 | 0.9700  | -7.5430 | CA | 1 | LIG | -0.2670 |
| 51 | H23  | -4.2670 | 1.6770  | -7.9590 | HA | 1 | LIG | 0.1612  |
| 52 | C18  | -5.0190 | -0.8780 | -8.2950 | CA | 1 | LIG | -0.2005 |
| 53 | H18  | -5.7040 | -0.1350 | -8.6960 | HA | 1 | LIG | 0.1440  |
| 54 | C20  | -4.4060 | -3.1720 | -7.8880 | CA | 1 | LIG | -0.1450 |
| 55 | H20  | -4.6230 | -4.2330 | -7.9770 | HA | 1 | LIG | 0.1421  |
| 56 | C19  | -5.3040 | -2.2240 | -8.4120 | CA | 1 | LIG | -0.1486 |
| 57 | H19  | -6.2150 | -2.5510 | -8.9050 | HA | 1 | LIG | 0.1462  |

**Table A.20.** MOL2 file containing partial charges and atom types of the <sup>1</sup>APdG adduct used for MD simulations (atom numbers are provided in the associated figure).



|   |     |        |        |         |    |   |     |         |
|---|-----|--------|--------|---------|----|---|-----|---------|
| 1 | P   | 4.9760 | 0.3860 | -1.3260 | P  | 1 | LIG | 1.1662  |
| 2 | O5' | 4.1390 | 0.9990 | 0.0000  | OS | 1 | LIG | -0.4956 |
| 3 | OP1 | 5.1120 | 1.4640 | -2.3580 | O2 | 1 | LIG | -0.7681 |

|    |      |         |         |         |    |   |     |         |
|----|------|---------|---------|---------|----|---|-----|---------|
| 4  | OP2  | 6.1010  | -0.4220 | -0.7550 | O2 | 1 | LIG | -0.7681 |
| 5  | O3   | 0.0000  | 0.0000  | 0.0000  | OS | 1 | LIG | -0.5424 |
| 6  | C5'  | 3.1040  | 1.8950  | -0.2820 | CI | 1 | LIG | -0.0196 |
| 7  | H5'  | 3.4230  | 2.6620  | -0.9830 | H1 | 1 | LIG | 0.0821  |
| 8  | H5'' | 2.8550  | 2.3890  | 0.6500  | H1 | 1 | LIG | 0.0821  |
| 9  | C4'  | 1.8570  | 1.2140  | -0.8200 | CT | 1 | LIG | 0.1633  |
| 10 | H4'  | 1.0480  | 1.9310  | -0.8780 | H1 | 1 | LIG | 0.1054  |
| 11 | O4'  | 2.1080  | 0.7240  | -2.1370 | OS | 1 | LIG | -0.4032 |
| 12 | C1'  | 1.8650  | -0.6500 | -2.2260 | CT | 1 | LIG | 0.1513  |
| 13 | H1'  | 0.8780  | -0.8250 | -2.6240 | H2 | 1 | LIG | 0.0873  |
| 14 | C3'  | 1.4070  | 0.0000  | 0.0000  | CT | 1 | LIG | 0.1347  |
| 15 | H3'  | 1.7910  | 0.0400  | 1.0120  | H1 | 1 | LIG | 0.0784  |
| 16 | C2'  | 1.9750  | -1.1690 | -0.7950 | CT | 1 | LIG | -0.0572 |
| 17 | H2'  | 3.0130  | -1.3300 | -0.5360 | HC | 1 | LIG | 0.0443  |
| 18 | H2'' | 1.4270  | -2.0920 | -0.6530 | HC | 1 | LIG | 0.0443  |
| 19 | N9   | 2.7970  | -1.2560 | -3.1630 | N* | 1 | LIG | -0.0519 |
| 20 | C8   | 4.0990  | -0.9300 | -3.5110 | CR | 1 | LIG | 0.3777  |
| 21 | C4   | 2.5500  | -2.4460 | -3.7900 | CB | 1 | LIG | 0.1445  |
| 22 | N7   | 4.6490  | -1.7930 | -4.2750 | NB | 1 | LIG | -0.4566 |
| 23 | C5   | 3.6890  | -2.7690 | -4.4570 | CB | 1 | LIG | 0.0570  |
| 24 | N3   | 1.3850  | -3.1370 | -3.7080 | NC | 1 | LIG | -0.4990 |
| 25 | C6   | 3.7200  | -3.9790 | -5.2240 | C  | 1 | LIG | 0.4977  |
| 26 | C2   | 1.3980  | -4.2210 | -4.4050 | CA | 1 | LIG | 0.5921  |
| 27 | O6   | 4.5820  | -4.4690 | -5.8910 | O  | 1 | LIG | -0.5456 |
| 28 | N1   | 2.4720  | -4.6420 | -5.1190 | NA | 1 | LIG | -0.4138 |
| 29 | H1   | 2.3970  | -5.4580 | -5.6890 | H  | 1 | LIG | 0.3351  |
| 30 | N2   | 0.2780  | -4.9950 | -4.4810 | N2 | 1 | LIG | -0.8656 |
| 31 | H21  | 0.3970  | -5.9770 | -4.5990 | H  | 1 | LIG | 0.3895  |
| 32 | H22  | -0.4320 | -4.7410 | -3.8300 | H  | 1 | LIG | 0.3895  |
| 33 | N10  | 4.6780  | 0.2390  | -3.0360 | N2 | 1 | LIG | -0.3628 |
| 34 | H10  | 4.0050  | 0.9680  | -2.9720 | H  | 1 | LIG | 0.2783  |
| 35 | C10  | 5.9550  | 0.6490  | -3.5100 | CA | 1 | LIG | 0.0814  |
| 36 | C15  | 6.1060  | 1.8360  | -4.2310 | CA | 1 | LIG | 0.0633  |
| 37 | C11  | 7.0640  | -0.1160 | -3.1790 | CA | 1 | LIG | -0.2616 |
| 38 | H11  | 6.9210  | -1.0300 | -2.6350 | HA | 1 | LIG | 0.1751  |
| 39 | C14  | 7.4130  | 2.2570  | -4.5810 | CA | 1 | LIG | 0.0181  |
| 40 | C23  | 4.9910  | 2.6640  | -4.6450 | CA | 1 | LIG | -0.1706 |
| 41 | H23  | 3.9870  | 2.3450  | -4.4390 | HA | 1 | LIG | 0.1296  |
| 42 | C12  | 8.3300  | 0.2810  | -3.5480 | CA | 1 | LIG | -0.1240 |
| 43 | H12  | 9.1750  | -0.3300 | -3.2860 | HA | 1 | LIG | 0.1527  |
| 44 | C16  | 7.6070  | 3.4880  | -5.2930 | CA | 1 | LIG | 0.0859  |

|    |      |         |        |         |    |   |     |         |
|----|------|---------|--------|---------|----|---|-----|---------|
| 45 | C13  | 8.5310  | 1.4720 | -4.2390 | CA | 1 | LIG | 0.0232  |
| 46 | C22  | 5.1770  | 3.8110 | -5.3090 | CA | 1 | LIG | -0.2254 |
| 47 | H22' | 4.3310  | 4.4020 | -5.6120 | HA | 1 | LIG | 0.1573  |
| 48 | C21  | 6.4980  | 4.2810 | -5.6550 | CA | 1 | LIG | 0.0462  |
| 49 | C17  | 8.9100  | 3.9110 | -5.6360 | CA | 1 | LIG | 0.0488  |
| 50 | C24  | 9.8500  | 1.9270 | -4.6100 | CA | 1 | LIG | -0.1744 |
| 51 | H24  | 10.6920 | 1.3130 | -4.3430 | HA | 1 | LIG | 0.1419  |
| 52 | C20  | 6.7030  | 5.4740 | -6.3400 | CA | 1 | LIG | -0.1411 |
| 53 | H20  | 5.8560  | 6.0770 | -6.6170 | HA | 1 | LIG | 0.1383  |
| 54 | C18  | 9.0760  | 5.1110 | -6.3220 | CA | 1 | LIG | -0.1473 |
| 55 | H18  | 10.0680 | 5.4350 | -6.5840 | HA | 1 | LIG | 0.1409  |
| 56 | C25  | 10.0320 | 3.0790 | -5.2670 | CA | 1 | LIG | -0.1962 |
| 57 | H25  | 11.0200 | 3.4070 | -5.5350 | HA | 1 | LIG | 0.1453  |
| 58 | C19  | 7.9830  | 5.8840 | -6.6680 | CA | 1 | LIG | -0.2217 |
| 59 | H19  | 8.1290  | 6.8080 | -7.1990 | HA | 1 | LIG | 0.1629  |

**Table A.21.** Cartesian coordinates of the nucleobase minimum for the <sup>AN</sup>G adduct with  $\theta = 29.9^\circ$  and  $\phi = 16.7^\circ$ .

| Calculated energy (in Hartrees) = -829.3022069 |             |             |             |
|------------------------------------------------|-------------|-------------|-------------|
| Atom                                           | X           | Y           | Z           |
| N                                              | -0.06926100 | 0.46196200  | -0.34668600 |
| C                                              | 0.36255100  | -0.83952200 | -0.13182800 |
| N                                              | -0.62197900 | -1.66773600 | 0.12555900  |
| C                                              | -1.76302700 | -0.88013200 | 0.06605300  |
| C                                              | -3.14133000 | -1.22591300 | 0.27780200  |
| O                                              | -3.67376800 | -2.28962300 | 0.54653900  |
| N                                              | -3.95774100 | -0.04763900 | 0.13461900  |
| H                                              | -4.93548300 | -0.21028200 | 0.34652800  |
| C                                              | -3.52219600 | 1.21669300  | -0.15888600 |
| N                                              | -4.47557500 | 2.21503400  | -0.19219900 |
| H                                              | -5.37876500 | 1.96528100  | -0.57618400 |
| H                                              | -4.10773300 | 3.07863000  | -0.57202800 |
| N                                              | -2.25953000 | 1.51823900  | -0.35395900 |
| C                                              | -1.44222000 | 0.44483200  | -0.21787800 |
| H                                              | 0.49182900  | 1.25207700  | -0.62913500 |
| N                                              | 1.67717600  | -1.22237200 | -0.27218300 |
| H                                              | 1.79065200  | -2.22480500 | -0.33485400 |
| C                                              | 2.82627500  | -0.44676400 | -0.06812300 |

|   |            |             |             |
|---|------------|-------------|-------------|
| C | 4.04225100 | -0.90686200 | -0.59942700 |
| C | 2.80909700 | 0.75946800  | 0.65048000  |
| C | 5.21007000 | -0.17334200 | -0.42118000 |
| C | 3.98409000 | 1.49507500  | 0.80382400  |
| H | 1.89486700 | 1.09888900  | 1.12470500  |
| C | 5.18943700 | 1.03977400  | 0.27166900  |
| H | 3.95361800 | 2.42493600  | 1.36501200  |
| H | 4.05801900 | -1.83803800 | -1.15985900 |
| H | 6.14109900 | -0.54780600 | -0.83727900 |
| H | 6.10043800 | 1.61583900  | 0.39968000  |

**Table A.22.** Cartesian coordinates of the nucleobase minimum for the <sup>AN</sup>G adduct with  $\theta = 180.0^\circ$  and  $\phi = 0.0^\circ$ .

| Calculated energy (in Hartrees) = -829.3011175 |             |             |             |
|------------------------------------------------|-------------|-------------|-------------|
| Atom                                           | X           | Y           | Z           |
| N                                              | -0.71971200 | -1.62217300 | -0.00300100 |
| C                                              | 0.34768300  | -0.72606200 | 0.00014100  |
| N                                              | -0.04848800 | 0.52486400  | 0.00689100  |
| C                                              | -1.43564900 | 0.45607300  | 0.00749200  |
| C                                              | -2.41328200 | 1.50716100  | 0.00050600  |
| O                                              | -2.29293000 | 2.72075700  | -0.00934100 |
| N                                              | -3.73435600 | 0.93089300  | -0.00467900 |
| H                                              | -4.47160500 | 1.62174600  | -0.08528700 |
| C                                              | -4.04281500 | -0.40239300 | -0.00206000 |
| N                                              | -5.38073900 | -0.73663600 | -0.06938500 |
| H                                              | -6.02269400 | -0.13814300 | 0.43574200  |
| H                                              | -5.53578000 | -1.72206600 | 0.10426700  |
| N                                              | -3.13657200 | -1.35285600 | 0.00977600  |
| C                                              | -1.87219800 | -0.86355200 | -0.00026400 |
| H                                              | -0.69216600 | -2.63138300 | -0.01400800 |
| N                                              | 1.63298000  | -1.20806800 | -0.00257200 |
| H                                              | 1.73139600  | -2.21225600 | -0.00625100 |

|   |            |             |             |
|---|------------|-------------|-------------|
| C | 2.83877100 | -0.48739000 | -0.00063700 |
| C | 4.03030100 | -1.23011600 | -0.00360500 |
| C | 2.89475400 | 0.91416400  | 0.00401800  |
| C | 5.26172700 | -0.58395500 | -0.00187500 |
| C | 4.13920600 | 1.54342300  | 0.00566500  |
| H | 1.97546400 | 1.48378600  | 0.00615900  |
| C | 5.32598100 | 0.81060800  | 0.00282400  |
| H | 4.17293700 | 2.62943100  | 0.00916300  |
| H | 3.98743900 | -2.31795600 | -0.00735900 |
| H | 6.17304700 | -1.17551000 | -0.00427200 |
| H | 6.28673000 | 1.31636700  | 0.00412800  |

**Table A.23.** Cartesian coordinates of the nucleobase minimum for the <sup>AN</sup>G adduct with  $\theta = 330.6^\circ$  and  $\phi = 163.1^\circ$ .

| Calculated energy (in Hartrees) = -829.3021875 |             |             |             |
|------------------------------------------------|-------------|-------------|-------------|
| Atom                                           | X           | Y           | Z           |
| N                                              | 0.06882000  | 0.46199600  | -0.33584200 |
| C                                              | -0.36201300 | -0.83862600 | -0.11443300 |
| N                                              | 0.62295600  | -1.66482700 | 0.14819700  |
| C                                              | 1.76340000  | -0.87670400 | 0.08514400  |
| C                                              | 3.14441300  | -1.22562400 | 0.27292300  |
| O                                              | 3.68027600  | -2.29385500 | 0.51566900  |
| N                                              | 3.96079700  | -0.04898500 | 0.11733800  |
| H                                              | 4.95462100  | -0.24310600 | 0.16175500  |
| C                                              | 3.52277300  | 1.21696400  | -0.16560000 |
| N                                              | 4.48804200  | 2.18985700  | -0.33512400 |
| H                                              | 5.29755000  | 2.13326300  | 0.27071300  |
| H                                              | 4.08715700  | 3.11884300  | -0.37642400 |
| N                                              | 2.25680800  | 1.52327800  | -0.32862700 |
| C                                              | 1.44219500  | 0.44576300  | -0.20945800 |
| H                                              | -0.49169300 | 1.24843600  | -0.62957200 |
| N                                              | -1.67618500 | -1.22376900 | -0.25156700 |

|   |             |             |             |
|---|-------------|-------------|-------------|
| H | -1.78837700 | -2.22640000 | -0.31278200 |
| C | -2.82675200 | -0.44772000 | -0.06130200 |
| C | -2.81422500 | 0.76651800  | 0.64383100  |
| C | -4.04019700 | -0.91551400 | -0.59196100 |
| C | -3.99091900 | 1.50197900  | 0.78451700  |
| C | -5.20964600 | -0.18176200 | -0.42647700 |
| H | -4.05253700 | -1.85298700 | -1.14188600 |
| C | -5.19352500 | 1.03908600  | 0.25288500  |
| H | -6.13849200 | -0.56221400 | -0.84199900 |
| H | -1.90239000 | 1.11259800  | 1.11779500  |
| H | -3.96388300 | 2.43819400  | 1.33524100  |
| H | -6.10584600 | 1.61520300  | 0.37076300  |

**Table A.24.** Cartesian coordinates of the nucleobase minimum for the  $^1\text{ANP}^{\text{G}}$  adduct with  $\theta = 36.7^\circ$  and  $\phi = 18.9^\circ$ .

| Calculated energy (in Hartrees) = -983.0063609 |            |             |             |
|------------------------------------------------|------------|-------------|-------------|
| Atom                                           | X          | Y           | Z           |
| N                                              | 1.09192600 | 0.52696100  | 0.40380500  |
| C                                              | 0.55296900 | -0.60271600 | -0.19375000 |
| N                                              | 1.46763100 | -1.43568300 | -0.62933700 |
| C                                              | 2.67290500 | -0.83352500 | -0.29511100 |
| C                                              | 4.02267800 | -1.26433400 | -0.53165400 |
| O                                              | 4.46766600 | -2.26564300 | -1.06692300 |
| N                                              | 4.93726100 | -0.27501800 | -0.02096200 |
| H                                              | 5.90790200 | -0.48530300 | -0.22340800 |
| C                                              | 4.60605800 | 0.90045900  | 0.59772200  |
| N                                              | 5.64682500 | 1.73327200  | 0.95776200  |
| H                                              | 6.48952200 | 1.28431200  | 1.29512600  |
| H                                              | 5.33695300 | 2.49337500  | 1.55071900  |
| N                                              | 3.36769200 | 1.28069400  | 0.81074300  |
| C                                              | 2.46186400 | 0.38912300  | 0.33829600  |
| H                                              | 0.58681700 | 1.28238200  | 0.84345900  |



|   |             |             |             |
|---|-------------|-------------|-------------|
| N | -0.80283900 | -0.85214200 | -0.21753800 |
| H | -1.00778800 | -1.78862200 | -0.54221800 |
| C | -1.81593300 | 0.11472400  | -0.36474900 |
| C | -3.14501200 | -0.24382800 | 0.03753400  |
| C | -1.56522100 | 1.37561300  | -0.87817700 |
| C | -4.19727300 | 0.71088200  | -0.14311000 |
| C | -3.45607600 | -1.50305900 | 0.61215200  |
| C | -2.60748400 | 2.32015200  | -1.00852400 |
| H | -0.56929100 | 1.62966200  | -1.22437900 |
| C | -5.51987600 | 0.34535900  | 0.21948700  |
| C | -3.89764700 | 1.99786200  | -0.66282000 |
| C | -4.75083900 | -1.82347700 | 0.95671400  |
| H | -2.65823900 | -2.21079500 | 0.81064200  |
| H | -2.37836200 | 3.30080900  | -1.41542500 |
| C | -5.79565800 | -0.89443600 | 0.74901900  |
| H | -4.96888800 | -2.79046600 | 1.40042800  |
| H | -4.70275700 | 2.71741800  | -0.78137500 |
| H | -6.31558300 | 1.07101300  | 0.07193000  |
| H | -6.81381400 | -1.15802500 | 1.02039700  |

**Table A.25.** Cartesian coordinates of the nucleobase minimum for the  $^{1\text{ANP}}\text{G}$  adduct with  $\theta = 323.4^\circ$  and  $\phi = 341.1^\circ$ .

| Calculated energy (in Hartrees) = -983.006339 |             |             |             |
|-----------------------------------------------|-------------|-------------|-------------|
| Atom                                          | X           | Y           | Z           |
| N                                             | -1.09272700 | 0.53158000  | 0.39535500  |
| C                                             | -0.55347200 | -0.59686200 | -0.20423200 |
| N                                             | -1.46786400 | -1.42835700 | -0.64328100 |
| C                                             | -2.67332200 | -0.82666100 | -0.30897500 |
| C                                             | -4.02294800 | -1.26876600 | -0.52475700 |
| O                                             | -4.46741300 | -2.28358400 | -1.03441200 |
| N                                             | -4.93791300 | -0.28334700 | -0.00734800 |
| H                                             | -5.90818600 | -0.57162600 | -0.06030200 |
| C                                             | -4.60707300 | 0.89876200  | 0.59885400  |
| N                                             | -5.64776300 | 1.66938500  | 1.07777200  |
| H                                             | -6.49300600 | 1.68554900  | 0.52032400  |
| H                                             | -5.33926200 | 2.59176300  | 1.35954200  |
| N                                             | -3.36894400 | 1.29391600  | 0.78387600  |

|   |             |             |             |
|---|-------------|-------------|-------------|
| C | -2.46269100 | 0.39272400  | 0.33071800  |
| H | -0.58798500 | 1.28227100  | 0.84349100  |
| N | 0.80218300  | -0.84642700 | -0.22848800 |
| H | 1.00748600  | -1.78261000 | -0.55354500 |
| C | 1.81700500  | 0.11943300  | -0.36653000 |
| C | 3.14485600  | -0.24424700 | 0.03551700  |
| C | 1.56944100  | 1.38457900  | -0.87111500 |
| C | 4.19900700  | 0.71000500  | -0.13628400 |
| C | 3.45283000  | -1.50805000 | 0.60169200  |
| C | 2.61353100  | 2.32821300  | -0.99273000 |
| H | 0.57450200  | 1.64281400  | -1.21700200 |
| C | 5.52033600  | 0.33971800  | 0.22608500  |
| C | 3.90253600  | 2.00126000  | -0.64714100 |
| C | 4.74643300  | -1.83306900 | 0.94631300  |
| H | 2.65361700  | -2.21591800 | 0.79409800  |
| H | 2.38683000  | 3.31228600  | -1.39273100 |
| C | 5.79313200  | -0.90430200 | 0.74718900  |
| H | 4.96201800  | -2.80350800 | 1.38362900  |
| H | 4.70906100  | 2.72031300  | -0.75890800 |
| H | 6.31747400  | 1.06514500  | 0.08528100  |
| H | 6.81034200  | -1.17147200 | 1.01860500  |

**Table A.26.** Cartesian coordinates of the nucleobase minimum for the  $^{1\text{ANP}}\text{G}$  adduct with  $\theta = 334.4^\circ$  and  $\phi = 252.1^\circ$ .

| Calculated energy (in Hartrees) = -983.0056003 |            |             |             |
|------------------------------------------------|------------|-------------|-------------|
| Atom                                           | X          | Y           | Z           |
| N                                              | 0.56164000 | 0.10570700  | 0.08398800  |
| C                                              | 0.40044400 | -1.13466500 | -0.50854600 |
| N                                              | 1.53787900 | -1.73256800 | -0.76899100 |
| C                                              | 2.50632800 | -0.83615000 | -0.33490400 |
| C                                              | 3.93904100 | -0.92844100 | -0.36786900 |
| O                                              | 4.68257700 | -1.80249600 | -0.78203800 |
| N                                              | 4.50869200 | 0.26763700  | 0.19834600  |
| H                                              | 5.51965100 | 0.29881600  | 0.13323900  |
| C                                              | 3.82179500 | 1.34062700  | 0.69991300  |
| N                                              | 4.57301900 | 2.41471700  | 1.13885700  |
| H                                              | 5.43763500 | 2.19205800  | 1.61724700  |

|   |             |             |             |
|---|-------------|-------------|-------------|
| H | 4.00766200  | 3.08898200  | 1.64031000  |
| N | 2.51184600  | 1.40825900  | 0.73132000  |
| C | 1.92195400  | 0.31088700  | 0.19481700  |
| H | -0.16282800 | 0.76045000  | 0.34149100  |
| N | -0.84376200 | -1.67130300 | -0.81357300 |
| H | -0.75040100 | -2.66313800 | -1.00069200 |
| C | -2.02305100 | -1.29957200 | -0.11303400 |
| C | -2.60850600 | -0.01649200 | -0.35246600 |
| C | -2.62022500 | -2.18177100 | 0.76529700  |
| C | -3.77053900 | 0.35582500  | 0.39998200  |
| C | -2.09930500 | 0.89075500  | -1.32092200 |
| C | -3.79459000 | -1.82417900 | 1.46364000  |
| H | -2.16324300 | -3.15271000 | 0.93593100  |
| C | -4.33180600 | 1.64420500  | 0.19672700  |
| C | -4.34551400 | -0.57420800 | 1.30447400  |
| C | -2.67550400 | 2.12916100  | -1.49652200 |
| H | -1.26184400 | 0.58317100  | -1.93717600 |
| H | -4.24219600 | -2.53759200 | 2.14911200  |
| C | -3.79359600 | 2.51650400  | -0.72083800 |
| H | -2.27775600 | 2.80825400  | -2.24492000 |
| H | -5.23360600 | -0.28489000 | 1.85952000  |
| H | -5.20516300 | 1.92632300  | 0.77903200  |
| H | -4.23528700 | 3.49818100  | -0.86492200 |

**Table A.27.** Cartesian coordinates of the nucleobase minimum for the  $^1\text{ANP}^{\text{G}}$  adduct with  $\theta = 26.0^\circ$  and  $\phi = 107.6^\circ$ .

| Calculated energy (in Hartrees) = -983.0055845 |             |             |             |
|------------------------------------------------|-------------|-------------|-------------|
| Atom                                           | X           | Y           | Z           |
| N                                              | -0.56203000 | 0.11250900  | 0.07006900  |
| C                                              | -0.40148700 | -1.12705100 | -0.52430900 |
| N                                              | -1.53918400 | -1.72345700 | -0.78664300 |
| C                                              | -2.50723600 | -0.82682700 | -0.35204600 |
| C                                              | -3.93959300 | -0.93013500 | -0.36022300 |
| O                                              | -4.68295900 | -1.81720100 | -0.74604700 |
| N                                              | -4.50854500 | 0.26216700  | 0.21455200  |
| H                                              | -5.51683000 | 0.21945100  | 0.30831700  |
| C                                              | -3.82133200 | 1.34175900  | 0.70132500  |
| N                                              | -4.57040900 | 2.35754900  | 1.26496600  |

|   |             |             |             |
|---|-------------|-------------|-------------|
| H | -5.44822000 | 2.57438000  | 0.80841900  |
| H | -4.00926100 | 3.17985300  | 1.45100500  |
| N | -2.51185500 | 1.42399800  | 0.70050300  |
| C | -1.92217700 | 0.31735500  | 0.18305100  |
| H | 0.16300100  | 0.76478000  | 0.33223700  |
| N | 0.84277800  | -1.66366700 | -0.83007800 |
| H | 0.74772100  | -2.65469500 | -1.02086800 |
| C | 2.01956900  | -1.29892000 | -0.12104300 |
| C | 2.61064400  | -0.01704800 | -0.35218700 |
| C | 2.60806800  | -2.18619500 | 0.75792500  |
| C | 3.76930200  | 0.34869000  | 0.40866200  |
| C | 2.11021800  | 0.89551000  | -1.32019700 |
| C | 3.77924400  | -1.83502300 | 1.46490300  |
| H | 2.14654700  | -3.15599700 | 0.92263200  |
| C | 4.33607600  | 1.63590500  | 0.21355200  |
| C | 4.33545400  | -0.58645100 | 1.31344500  |
| C | 2.69170700  | 2.13255200  | -1.48785800 |
| H | 1.27542300  | 0.59284700  | -1.94256200 |
| H | 4.21990200  | -2.55220900 | 2.15091700  |
| C | 3.80641500  | 2.51334700  | -0.70409600 |
| H | 2.30094700  | 2.81568100  | -2.23631300 |
| H | 5.22089900  | -0.30204900 | 1.87521000  |
| H | 5.20669400  | 1.91303100  | 0.80230900  |
| H | 4.25232200  | 3.49401900  | -0.84193600 |

**Table A.28.** Cartesian coordinates of the nucleobase minimum for the  $^{1\text{ANP}}\text{G}$  adduct with  $\theta = 177.1^\circ$  and  $\phi = 356.4^\circ$ .

| Calculated energy (in Hartrees) = -983.0036838 |             |             |             |
|------------------------------------------------|-------------|-------------|-------------|
| Atom                                           | X           | Y           | Z           |
| N                                              | -1.47743200 | -1.43280100 | -0.13440700 |
| C                                              | -0.56851000 | -0.37866200 | -0.05444600 |
| N                                              | -1.16273900 | 0.78611900  | 0.04854900  |
| C                                              | -2.52029300 | 0.49329600  | 0.04768500  |
| C                                              | -3.65489100 | 1.36946000  | 0.12239800  |
| O                                              | -3.73235700 | 2.58355000  | 0.20660000  |
| N                                              | -4.86521200 | 0.58833500  | 0.07437400  |
| H                                              | -5.70581300 | 1.15415700  | 0.04964200  |

|   |             |             |             |
|---|-------------|-------------|-------------|
| C | -4.95336100 | -0.77365100 | -0.02640600 |
| N | -6.21988600 | -1.31487400 | -0.11482900 |
| H | -6.94344400 | -0.86850200 | 0.43519400  |
| H | -6.21141400 | -2.32286700 | -0.01967500 |
| N | -3.90515400 | -1.56245800 | -0.09243900 |
| C | -2.73687900 | -0.87528800 | -0.06583800 |
| H | -1.28432400 | -2.42358300 | -0.15939300 |
| N | 0.77835500  | -0.65737300 | -0.06535800 |
| H | 1.01457600  | -1.58629600 | -0.37796500 |
| C | 1.84932400  | 0.25444300  | -0.04994400 |
| C | 3.17578600  | -0.29525800 | 0.01637000  |
| C | 1.66166200  | 1.62482100  | -0.09290200 |
| C | 4.28913600  | 0.60637500  | -0.03551900 |
| C | 3.44481200  | -1.68433400 | 0.14029700  |
| C | 2.77605700  | 2.49143400  | -0.10792800 |
| H | 0.65574900  | 2.02029300  | -0.12136700 |
| C | 5.60495000  | 0.07465400  | -0.00972900 |
| C | 4.06072100  | 2.00520700  | -0.09772400 |
| C | 4.73493600  | -2.16707400 | 0.16959400  |
| H | 2.63067000  | -2.39373800 | 0.25986200  |
| H | 2.59773800  | 3.56224400  | -0.14536400 |
| C | 5.82997200  | -1.27925500 | 0.08203500  |
| H | 4.91006700  | -3.23408600 | 0.27260800  |
| H | 4.91463300  | 2.67618400  | -0.12807000 |
| H | 6.43996100  | 0.76897500  | -0.05866800 |
| H | 6.84439500  | -1.66683400 | 0.10151400  |

**Table A.29.** Cartesian coordinates of the nucleobase minimum for the  $^1\text{ANP}$ G adduct with  $\theta = 166.7^\circ$  and  $\phi = 117.3^\circ$ .

| Calculated energy (in Hartrees) = -983.0002316 |             |             |             |
|------------------------------------------------|-------------|-------------|-------------|
| Atom                                           | X           | Y           | Z           |
| N                                              | -1.48124800 | -1.39739000 | -1.15125000 |
| C                                              | -0.34412700 | -1.17495100 | -0.38209200 |
| N                                              | -0.57007500 | -0.38824000 | 0.63835900  |
| C                                              | -1.91387900 | -0.05776200 | 0.53554100  |
| C                                              | -2.72672400 | 0.78878300  | 1.36497400  |
| O                                              | -2.45934300 | 1.44643200  | 2.35542400  |

|   |             |             |             |
|---|-------------|-------------|-------------|
| N | -4.07400500 | 0.80849300  | 0.85215400  |
| H | -4.68138000 | 1.44953100  | 1.34969100  |
| C | -4.53846000 | 0.13990000  | -0.24829300 |
| N | -5.85575500 | 0.35352500  | -0.60041100 |
| H | -6.51624900 | 0.46510800  | 0.15886900  |
| H | -6.17052900 | -0.29492500 | -1.31148000 |
| N | -3.78682000 | -0.63707700 | -0.99331400 |
| C | -2.50050800 | -0.68112800 | -0.56450300 |
| H | -1.53229600 | -1.88097600 | -2.03615500 |
| N | 0.86973000  | -1.73728000 | -0.77492200 |
| H | 0.82901100  | -2.74824900 | -0.85793100 |
| C | 2.07944300  | -1.29232200 | -0.17092800 |
| C | 2.53105800  | 0.04600900  | -0.39091700 |
| C | 2.84407900  | -2.17535900 | 0.56122300  |
| C | 3.76485000  | 0.45148100  | 0.21222000  |
| C | 1.82421700  | 0.97793000  | -1.19328300 |
| C | 4.07694500  | -1.77449100 | 1.12264600  |
| H | 2.48144600  | -3.18727100 | 0.72522300  |
| C | 4.21660600  | 1.78350100  | 0.01871400  |
| C | 4.52149800  | -0.48408000 | 0.96649400  |
| C | 2.29626000  | 2.25856800  | -1.36810000 |
| H | 0.90650800  | 0.66941100  | -1.67898000 |
| H | 4.65759600  | -2.49142100 | 1.69561600  |
| C | 3.49889700  | 2.67093600  | -0.74833400 |
| H | 1.74202100  | 2.95887600  | -1.98636000 |
| H | 5.45963400  | -0.16225300 | 1.41037900  |
| H | 5.14894000  | 2.08704600  | 0.48806900  |
| H | 3.85630700  | 3.68735900  | -0.88683300 |

**Table A.30.** Cartesian coordinates of the nucleobase minimum for the  $^1\text{ANP}^{\text{G}}$  adduct with  $\theta = 193.9^\circ$  and  $\phi = 242.8^\circ$ .

| Calculated energy (in Hartrees) = -983.0002151 |            |             |             |
|------------------------------------------------|------------|-------------|-------------|
| Atom                                           | X          | Y           | Z           |
| N                                              | 1.47685000 | -1.37099600 | -1.17772700 |
| C                                              | 0.34258600 | -1.16798900 | -0.39910500 |
| N                                              | 0.57049200 | -0.39991700 | 0.63500000  |
| C                                              | 1.91243500 | -0.06244000 | 0.53231000  |

|   |             |             |             |
|---|-------------|-------------|-------------|
| C | 2.73225800  | 0.74965700  | 1.38889100  |
| O | 2.47485800  | 1.36081200  | 2.41124800  |
| N | 4.07820700  | 0.77972400  | 0.87334400  |
| H | 4.72528700  | 1.27705100  | 1.47444500  |
| C | 4.53598700  | 0.14825300  | -0.25155500 |
| N | 5.88190500  | 0.26823000  | -0.53199000 |
| H | 6.31157100  | 1.16076100  | -0.32230100 |
| H | 6.11208400  | -0.07917600 | -1.45473200 |
| N | 3.77573900  | -0.58958600 | -1.02714700 |
| C | 2.49692400  | -0.66489700 | -0.58061500 |
| H | 1.52713400  | -1.84371600 | -2.06850600 |
| N | -0.87153900 | -1.72676200 | -0.79736700 |
| H | -0.83054000 | -2.73708000 | -0.88872000 |
| C | -2.07984700 | -1.28805500 | -0.18600900 |
| C | -2.53159900 | 0.05248900  | -0.39190700 |
| C | -2.84289500 | -2.17771100 | 0.53967400  |
| C | -3.76358700 | 0.45273700  | 0.21827400  |
| C | -1.82643100 | 0.99165100  | -1.18726700 |
| C | -4.07407600 | -1.78165900 | 1.10822000  |
| H | -2.48014500 | -3.19118500 | 0.69327700  |
| C | -4.21522300 | 1.78672300  | 0.03849900  |
| C | -4.51859400 | -0.48966100 | 0.96566000  |
| C | -2.29840500 | 2.27407300  | -1.34889800 |
| H | -0.91031700 | 0.68692500  | -1.67843800 |
| H | -4.65337500 | -2.50375800 | 1.67603000  |
| C | -3.49922500 | 2.68108000  | -0.72211100 |
| H | -1.74576100 | 2.97990700  | -1.96234700 |
| H | -5.45533200 | -0.17170200 | 1.41523500  |
| H | -5.14618600 | 2.08627600  | 0.51311000  |
| H | -3.85670400 | 3.69885300  | -0.85018500 |

**Table A.31.** Cartesian coordinates of the nucleobase minimum for the <sup>3</sup>APHN<sup>+</sup>G adduct with  $\theta = 33.3^\circ$  and  $\phi = 193.9^\circ$ .

| Calculated energy (in Hartrees) = -1136.7154178 |             |             |             |
|-------------------------------------------------|-------------|-------------|-------------|
| Atom                                            | X           | Y           | Z           |
| N                                               | -1.66986700 | -0.33928200 | -0.61953400 |
| C                                               | -1.67468000 | -1.54206300 | 0.07343000  |

|   |             |             |             |
|---|-------------|-------------|-------------|
| N | -2.82014300 | -1.78546800 | 0.66455300  |
| C | -3.61796500 | -0.69813600 | 0.33755400  |
| C | -4.97142400 | -0.39414200 | 0.71207700  |
| O | -5.77663400 | -1.00687700 | 1.39244700  |
| N | -5.35096800 | 0.87508700  | 0.14576900  |
| H | -6.27248300 | 1.18275400  | 0.43482100  |
| C | -4.57095400 | 1.68769900  | -0.63253500 |
| N | -5.11835300 | 2.89293100  | -1.02398100 |
| H | -6.10623100 | 2.89103600  | -1.24652000 |
| H | -4.55356900 | 3.35763600  | -1.72428600 |
| N | -3.33610900 | 1.39611200  | -0.96998600 |
| C | -2.92161000 | 0.21372700  | -0.45193100 |
| N | -0.61273500 | -2.41935300 | 0.04929200  |
| H | -0.85809500 | -3.33846300 | 0.39220400  |
| C | 0.74954100  | -2.10436900 | -0.04124800 |
| C | 1.24040100  | -0.81877700 | 0.15591300  |
| C | 1.64810400  | -3.15317100 | -0.35044100 |
| C | 2.61437900  | -0.52610700 | 0.02061300  |
| C | 2.99384500  | -2.89263000 | -0.45897000 |
| H | 1.26214800  | -4.15652400 | -0.50836400 |
| C | 3.14737200  | 0.81583000  | 0.20249000  |
| C | 3.51185700  | -1.58735100 | -0.29042500 |
| H | 3.68283700  | -3.70061300 | -0.69087600 |
| C | 2.33283700  | 1.93184500  | 0.50076200  |
| C | 4.55217100  | 1.02707200  | 0.07119900  |
| C | 4.91317600  | -1.32680700 | -0.41873700 |
| C | 2.87194200  | 3.19342400  | 0.66911900  |
| C | 5.07986100  | 2.32642700  | 0.24892500  |
| C | 5.41377800  | -0.07437900 | -0.24302500 |
| H | 5.57179300  | -2.15799600 | -0.65742900 |
| C | 4.25892400  | 3.39565700  | 0.54417000  |
| H | 2.21997300  | 4.03129100  | 0.89854700  |
| H | 6.15281400  | 2.46849100  | 0.14689900  |
| H | 6.47941900  | 0.11552200  | -0.33924100 |
| H | 4.67874800  | 4.38841400  | 0.67792700  |
| H | 1.25984800  | 1.80922400  | 0.60000300  |
| H | 0.55618600  | -0.04102200 | 0.46687600  |
| H | -0.93241500 | 0.02258700  | -1.20656800 |



**Table A.32.** Cartesian coordinates of the nucleobase minimum for the  ${}^3\text{APHN}^{\text{G}}$  adduct with  $\theta = 327.4^\circ$  and  $\phi = 166.0^\circ$ .

| Calculated energy (in Hartrees) = -1136.7153814 |             |             |             |
|-------------------------------------------------|-------------|-------------|-------------|
| Atom                                            | X           | Y           | Z           |
| N                                               | 1.66917700  | -0.32454400 | -0.60170400 |
| C                                               | 1.67705200  | -1.53262100 | 0.08194200  |
| N                                               | 2.82486500  | -1.78045600 | 0.66669600  |
| C                                               | 3.62141000  | -0.69089300 | 0.34438200  |
| C                                               | 4.98518200  | -0.40253200 | 0.69303500  |
| O                                               | 5.80323600  | -1.03553700 | 1.33864200  |
| N                                               | 5.36504800  | 0.86744900  | 0.12884900  |
| H                                               | 6.34222200  | 1.09360200  | 0.27496000  |
| C                                               | 4.57619100  | 1.69434400  | -0.62525000 |
| N                                               | 5.16481100  | 2.84020100  | -1.12102600 |
| H                                               | 5.83495000  | 3.30284400  | -0.51920800 |
| H                                               | 4.48537400  | 3.47426300  | -1.52302000 |
| N                                               | 3.32874000  | 1.42176100  | -0.93048900 |
| C                                               | 2.92307500  | 0.22535000  | -0.43809900 |
| N                                               | 0.61526400  | -2.40989000 | 0.05779000  |
| H                                               | 0.86315900  | -3.33212400 | 0.39014100  |
| C                                               | -0.74713300 | -2.09928200 | -0.03634000 |
| C                                               | -1.24303000 | -0.81523600 | 0.15908900  |
| C                                               | -1.64163900 | -3.15140200 | -0.34654300 |
| C                                               | -2.61779400 | -0.52762700 | 0.02102600  |
| C                                               | -2.98800000 | -2.89572100 | -0.45777200 |
| H                                               | -1.25201500 | -4.15355900 | -0.50308000 |
| C                                               | -3.15606600 | 0.81251700  | 0.20071100  |
| C                                               | -3.51098800 | -1.59220000 | -0.29090000 |
| H                                               | -3.67363400 | -3.70630800 | -0.69050200 |
| C                                               | -2.34614300 | 1.93187600  | 0.49899700  |
| C                                               | -4.56141900 | 1.01856800  | 0.06694900  |
| C                                               | -4.91298100 | -1.33686100 | -0.42174900 |
| C                                               | -2.89008400 | 3.19166800  | 0.66515300  |
| C                                               | -5.09408700 | 2.31620200  | 0.24251500  |
| C                                               | -5.41856100 | -0.08618800 | -0.24774800 |
| H                                               | -5.56810700 | -2.17062800 | -0.66104300 |
| C                                               | -4.27758900 | 3.38876500  | 0.53790100  |
| H                                               | -2.24159600 | 4.03229800  | 0.89449800  |
| H                                               | -6.16736900 | 2.45426400  | 0.13853500  |
| H                                               | -6.48471000 | 0.09975500  | -0.34593700 |
| H                                               | -4.70127200 | 4.38012900  | 0.66982300  |

|   |             |             |             |
|---|-------------|-------------|-------------|
| H | -1.27282900 | 1.81308600  | 0.59963400  |
| H | -0.56224500 | -0.03490600 | 0.47103200  |
| H | 0.93246300  | 0.03757300  | -1.18956400 |

**Table A.33.** Cartesian coordinates of the nucleobase minimum for the  ${}^3\text{APHN}^{\text{N}}\text{G}$  adduct with  $\theta = 180.0^\circ$  and  $\phi = 180.0^\circ$ .

| Calculated energy (in Hartrees) = -1136.7153462 |             |             |             |
|-------------------------------------------------|-------------|-------------|-------------|
| Atom                                            | X           | Y           | Z           |
| N                                               | -2.99392500 | -1.87130400 | 0.00181000  |
| C                                               | -1.67760900 | -1.41392200 | 0.00052800  |
| N                                               | -1.60401600 | -0.10303500 | 0.00260600  |
| C                                               | -2.92415600 | 0.32590900  | 0.00468300  |
| C                                               | -3.46167900 | 1.65644400  | -0.00613000 |
| O                                               | -2.91354800 | 2.74606600  | -0.02101100 |
| N                                               | -4.90089800 | 1.58993900  | -0.00821100 |
| H                                               | -5.34394200 | 2.49772700  | -0.09197600 |
| C                                               | -5.66432600 | 0.45377300  | 0.00051600  |
| N                                               | -7.03311300 | 0.61818600  | -0.06407700 |
| H                                               | -7.41921200 | 1.40989800  | 0.43535500  |
| H                                               | -7.53006200 | -0.24595900 | 0.11331600  |
| N                                               | -5.15667600 | -0.75769200 | 0.01570500  |
| C                                               | -3.80129600 | -0.75206300 | 0.00261700  |
| N                                               | -0.64391300 | -2.31707700 | -0.00112900 |
| H                                               | -0.90135500 | -3.29251100 | -0.00433000 |
| C                                               | 0.73596400  | -2.05502200 | -0.00178500 |
| C                                               | 1.26115000  | -0.77002800 | -0.00013300 |
| C                                               | 1.60255200  | -3.17307500 | -0.00383500 |
| C                                               | 2.65541800  | -0.56138100 | -0.00005900 |
| C                                               | 2.96528100  | -2.98386400 | -0.00382400 |
| H                                               | 1.18903200  | -4.17967900 | -0.00534700 |
| C                                               | 3.22991400  | 0.77609000  | 0.00187000  |
| C                                               | 3.52789700  | -1.68738400 | -0.00176800 |
| H                                               | 3.63019900  | -3.84374500 | -0.00535800 |
| C                                               | 2.43830500  | 1.94736400  | 0.00306400  |
| C                                               | 4.64875500  | 0.92041100  | 0.00257000  |
| C                                               | 4.94644200  | -1.49429600 | -0.00127700 |
| C                                               | 3.01752700  | 3.20279900  | 0.00508700  |
| C                                               | 5.21581500  | 2.21511400  | 0.00473200  |
| C                                               | 5.48441100  | -0.24496500 | 0.00094400  |
| H                                               | 5.58680200  | -2.37305900 | -0.00264700 |

|   |             |             |             |
|---|-------------|-------------|-------------|
| C | 4.41791700  | 3.34147600  | 0.00602200  |
| H | 2.38443100  | 4.08534600  | 0.00580900  |
| H | 6.29924600  | 2.30751100  | 0.00525900  |
| H | 6.56247300  | -0.10599200 | 0.00139500  |
| H | 4.86722900  | 4.33060600  | 0.00760400  |
| H | 1.35659000  | 1.87233800  | 0.00215700  |
| H | 0.57220000  | 0.05996600  | 0.00108400  |
| H | -3.32714000 | -2.82439000 | -0.00409300 |

**Table A.34.** Cartesian coordinates of the nucleobase minimum for the <sup>3</sup>APHN<sup>G</sup> adduct with  $\theta = 26.8^\circ$  and  $\phi = 21.7^\circ$ .

| Calculated energy (in Hartrees) = -1136.7143743 |             |             |             |
|-------------------------------------------------|-------------|-------------|-------------|
| Atom                                            | X           | Y           | Z           |
| N                                               | 2.30200400  | 0.39726100  | 0.32317500  |
| C                                               | 1.85732000  | -0.77311900 | -0.27630300 |
| N                                               | 2.83834500  | -1.56359600 | -0.64273500 |
| C                                               | 3.98893600  | -0.89098000 | -0.25622500 |
| C                                               | 5.37088500  | -1.25407300 | -0.40679000 |
| O                                               | 5.89681900  | -2.24083200 | -0.89338700 |
| N                                               | 6.20207500  | -0.20812700 | 0.13224500  |
| H                                               | 7.19233800  | -0.37036400 | -0.01058100 |
| C                                               | 5.77569300  | 0.95914600  | 0.70663900  |
| N                                               | 6.74975700  | 1.85208100  | 1.10702700  |
| H                                               | 7.59333400  | 1.45451400  | 1.50182800  |
| H                                               | 6.36766100  | 2.60640900  | 1.66415000  |
| N                                               | 4.50906800  | 1.27687500  | 0.84147300  |
| C                                               | 3.67959800  | 0.33012500  | 0.33735500  |
| N                                               | 0.52403500  | -1.09424300 | -0.38838500 |
| H                                               | 0.36353300  | -2.07292300 | -0.58216200 |
| C                                               | -0.56947400 | -0.22166200 | -0.44369400 |
| C                                               | -1.83247400 | -0.70866900 | -0.12323400 |
| C                                               | -0.42384000 | 1.13043200  | -0.83537500 |
| C                                               | -2.97805300 | 0.10803200  | -0.17162100 |
| C                                               | -1.53026900 | 1.95191700  | -0.86344400 |
| H                                               | 0.54241800  | 1.50552200  | -1.15327900 |
| C                                               | -4.31022400 | -0.38328200 | 0.15251300  |
| C                                               | -2.81877300 | 1.47816300  | -0.53568400 |
| H                                               | -1.41936400 | 2.98939700  | -1.16823600 |
| C                                               | -4.56680700 | -1.72471100 | 0.51386100  |
| C                                               | -5.41228400 | 0.52086900  | 0.10499200  |

|   |             |             |             |
|---|-------------|-------------|-------------|
| C | -3.95629900 | 2.34677000  | -0.57312400 |
| C | -5.84398600 | -2.15825100 | 0.81788300  |
| C | -6.70846400 | 0.05481000  | 0.42102300  |
| C | -5.19907700 | 1.89008900  | -0.26451500 |
| H | -3.80188600 | 3.38491100  | -0.85594200 |
| C | -6.92719200 | -1.26183700 | 0.77304900  |
| H | -6.00973400 | -3.19621900 | 1.09150000  |
| H | -7.53652100 | 0.75798700  | 0.38036000  |
| H | -6.05776800 | 2.55542100  | -0.29449800 |
| H | -7.92912600 | -1.60572500 | 1.01296900  |
| H | -3.75230800 | -2.43941100 | 0.55419300  |
| H | -1.91331500 | -1.74545700 | 0.18470600  |
| H | 1.73790400  | 1.11820700  | 0.74864600  |

**Table A.35.** Cartesian coordinates of the nucleobase minimum for the <sup>3</sup>APHN<sup>G</sup> adduct with  $\theta = 333.8^\circ$  and  $\phi = 338.0^\circ$ .

| Calculated energy (in Hartrees) = -1136.7143537 |             |             |             |
|-------------------------------------------------|-------------|-------------|-------------|
| Atom                                            | X           | Y           | Z           |
| N                                               | -2.30268200 | 0.40375800  | 0.30971900  |
| C                                               | -1.85784200 | -0.76630700 | -0.29017000 |
| N                                               | -2.83864300 | -1.55712100 | -0.65661100 |
| C                                               | -3.98932600 | -0.88525200 | -0.26920600 |
| C                                               | -5.37039500 | -1.26087500 | -0.39540000 |
| O                                               | -5.89483700 | -2.26212300 | -0.85313700 |
| N                                               | -6.20185000 | -0.21849200 | 0.15003300  |
| H                                               | -7.18691600 | -0.45701200 | 0.16030900  |
| C                                               | -5.77623100 | 0.95677000  | 0.70868600  |
| N                                               | -6.74631600 | 1.78964700  | 1.22994600  |
| H                                               | -7.62107000 | 1.83870500  | 0.72213300  |
| H                                               | -6.37530400 | 2.70027500  | 1.47175600  |
| N                                               | -4.51081700 | 1.29037800  | 0.81203200  |
| C                                               | -3.68028400 | 0.33423600  | 0.32763700  |
| N                                               | -0.52462200 | -1.08638300 | -0.40553500 |
| H                                               | -0.36355000 | -2.06556500 | -0.59614300 |
| C                                               | 0.56989200  | -0.21506800 | -0.45187000 |
| C                                               | 1.83161400  | -0.70568600 | -0.13164600 |
| C                                               | 0.42706400  | 1.13969400  | -0.83542700 |
| C                                               | 2.97834200  | 0.10977800  | -0.17230400 |
| C                                               | 1.53464300  | 1.95982600  | -0.85589300 |
| H                                               | -0.53799800 | 1.51803200  | -1.15304100 |

|   |             |             |             |
|---|-------------|-------------|-------------|
| C | 4.30913600  | -0.38534500 | 0.15175200  |
| C | 2.82177700  | 1.48236200  | -0.52823200 |
| H | 1.42583500  | 2.99938400  | -1.15433500 |
| C | 4.56305900  | -1.72924700 | 0.50565900  |
| C | 5.41255800  | 0.51752100  | 0.11200700  |
| C | 3.96055600  | 2.34960500  | -0.55786900 |
| C | 5.83898900  | -2.16640400 | 0.80976000  |
| C | 6.70741200  | 0.04774400  | 0.42791200  |
| C | 5.20205200  | 1.88928400  | -0.24953000 |
| H | 3.80817400  | 3.38971600  | -0.83450400 |
| C | 6.92354900  | -1.27129800 | 0.77249600  |
| H | 6.00266000  | -3.20618700 | 1.07765700  |
| H | 7.53652900  | 0.74998900  | 0.39327300  |
| H | 6.06172700  | 2.55358500  | -0.27354500 |
| H | 7.92448200  | -1.61802200 | 1.01249600  |
| H | 3.74748700  | -2.44301500 | 0.54020500  |
| H | 1.91050000  | -1.74448700 | 0.16994200  |
| H | -1.73875000 | 1.12104500  | 0.74162500  |

**Table A.36.** Cartesian coordinates of the nucleobase minimum for the  ${}^3\text{APHN}^{\text{G}}$  adduct with  $\theta = 180.0^\circ$  and  $\phi = 0.0^\circ$ .

| Calculated energy (in Hartrees) = -1136.7133475 |             |             |             |
|-------------------------------------------------|-------------|-------------|-------------|
| Atom                                            | X           | Y           | Z           |
| N                                               | -2.81301400 | -1.57816200 | 0.00049300  |
| C                                               | -1.85929600 | -0.56250000 | 0.00204700  |
| N                                               | -2.40021300 | 0.63267000  | 0.00603100  |
| C                                               | -3.76948800 | 0.40101900  | 0.00652000  |
| C                                               | -4.86427900 | 1.32962100  | -0.00314900 |
| O                                               | -4.88703400 | 2.54880100  | -0.01570900 |
| N                                               | -6.10794100 | 0.60162100  | -0.00747100 |
| H                                               | -6.92164200 | 1.20034500  | -0.09001600 |
| C                                               | -6.25686600 | -0.75891600 | -0.00209000 |
| N                                               | -7.54543300 | -1.24866900 | -0.06917000 |
| H                                               | -8.25533600 | -0.72818700 | 0.43125500  |
| H                                               | -7.58461300 | -2.24530100 | 0.10512700  |
| N                                               | -5.24468200 | -1.59588300 | 0.01207100  |
| C                                               | -4.04713700 | -0.96088500 | 0.00145100  |
| N                                               | -0.52545500 | -0.89115000 | 0.00080300  |
| H                                               | -0.31185300 | -1.87750400 | -0.00300300 |
| C                                               | 0.58965100  | -0.03995500 | 0.00175600  |

|   |             |             |             |
|---|-------------|-------------|-------------|
| C | 1.84883500  | -0.63236800 | 0.00003200  |
| C | 0.47097900  | 1.36893900  | 0.00445500  |
| C | 3.02908800  | 0.13316900  | 0.00098800  |
| C | 1.62025100  | 2.12976900  | 0.00550600  |
| H | -0.51128100 | 1.82111800  | 0.00551200  |
| C | 4.35831500  | -0.46231300 | -0.00083700 |
| C | 2.90912600  | 1.55366200  | 0.00391200  |
| H | 1.53678300  | 3.21357900  | 0.00748900  |
| C | 4.57653300  | -1.85812400 | -0.00393700 |
| C | 5.50008900  | 0.39286300  | 0.00048000  |
| C | 4.08517500  | 2.36998900  | 0.00510300  |
| C | 5.85232700  | -2.39126400 | -0.00564400 |
| C | 6.79435700  | -0.17512400 | -0.00128200 |
| C | 5.32712400  | 1.81633900  | 0.00349900  |
| H | 3.96028700  | 3.44982600  | 0.00728300  |
| C | 6.97488700  | -1.54358600 | -0.00429500 |
| H | 5.98627100  | -3.46930500 | -0.00808100 |
| H | 7.65261900  | 0.49227900  | -0.00026800 |
| H | 6.21478200  | 2.44335400  | 0.00439100  |
| H | 7.97621800  | -1.96454700 | -0.00566400 |
| H | 3.73071900  | -2.53709500 | -0.00512300 |
| H | 1.90932500  | -1.71679100 | -0.00208700 |
| H | -2.66686600 | -2.57718300 | -0.00755300 |

**Table A.37.** Cartesian coordinates of the nucleobase minimum for the  $^{1\text{AP}}\text{G}$  adduct with  $\theta = 34.0^\circ$  and  $\phi = 22.4^\circ$ .

| Calculated energy (in Hartrees) = -1212.9766572 |            |             |             |
|-------------------------------------------------|------------|-------------|-------------|
| Atom                                            | X          | Y           | Z           |
| N                                               | 2.43931300 | 0.38308200  | 0.43537700  |
| C                                               | 1.99297000 | -0.68480100 | -0.32959900 |
| N                                               | 2.97255300 | -1.39845800 | -0.83146100 |
| C                                               | 4.12532800 | -0.77989900 | -0.36778000 |
| C                                               | 5.50689000 | -1.09866800 | -0.59915400 |
| O                                               | 6.03151600 | -1.99217800 | -1.24217200 |
| N                                               | 6.33975500 | -0.13759800 | 0.07787700  |
| H                                               | 7.32818500 | -0.26351000 | -0.10789200 |
| C                                               | 5.91477300 | 0.92326600  | 0.83166500  |
| N                                               | 6.88858600 | 1.75690900  | 1.34436900  |
| H                                               | 7.74458000 | 1.31481200  | 1.65633600  |
| H                                               | 6.51151400 | 2.41247500  | 2.01773400  |

|   |             |             |             |
|---|-------------|-------------|-------------|
| N | 4.64840600  | 1.20127300  | 1.03816600  |
| C | 3.81690500  | 0.33245500  | 0.41196900  |
| H | 1.87521700  | 1.04113900  | 0.95304500  |
| N | 0.65764800  | -1.00066900 | -0.45202900 |
| H | 0.51515800  | -1.90023100 | -0.89240300 |
| C | -0.41153800 | -0.08682100 | -0.46960800 |
| C | -1.70804200 | -0.54995200 | -0.13605500 |
| C | -0.21735600 | 1.25638100  | -0.82093900 |
| C | -2.80325900 | 0.36291100  | -0.19867300 |
| C | -1.96440400 | -1.90304100 | 0.26740000  |
| C | -1.28216000 | 2.14816600  | -0.84277000 |
| H | 0.76996200  | 1.59234500  | -1.11861100 |
| C | -4.12376000 | -0.08781100 | 0.09911300  |
| C | -2.58708200 | 1.73046600  | -0.54770200 |
| C | -3.22652800 | -2.33160600 | 0.54807400  |
| H | -1.12957500 | -2.58699900 | 0.37811700  |
| H | -1.10767100 | 3.18268600  | -1.12628100 |
| C | -4.35270200 | -1.44843600 | 0.46390600  |
| C | -5.22301700 | 0.82248800  | 0.03605800  |
| C | -3.70868600 | 2.62298100  | -0.59591800 |
| H | -3.39820800 | -3.35887500 | 0.85853100  |
| C | -5.66106100 | -1.87209000 | 0.74615000  |
| C | -6.51290500 | 0.35314600  | 0.32675700  |
| C | -4.96924400 | 2.19003400  | -0.32101100 |
| H | -3.52800600 | 3.66066400  | -0.86449700 |
| C | -6.72694000 | -0.97900200 | 0.67560300  |
| H | -5.83258400 | -2.90881800 | 1.02334800  |
| H | -7.34923900 | 1.04551000  | 0.27816000  |
| H | -5.81096400 | 2.87596800  | -0.36607700 |
| H | -7.73341500 | -1.32250800 | 0.89685000  |

**Table A.38.** Cartesian coordinates of the nucleobase minimum for the  $^{1\text{AP}}\text{G}$  adduct with  $\theta = 326.0^\circ$  and  $\phi = 337.6^\circ$ .

| Calculated energy (in Hartrees) = -1212.9766571 |             |             |             |
|-------------------------------------------------|-------------|-------------|-------------|
| Atom                                            | X           | Y           | Z           |
| N                                               | -2.43931800 | 0.38308600  | 0.43537700  |
| C                                               | -1.99297100 | -0.68479800 | -0.32959600 |
| N                                               | -2.97255000 | -1.39846100 | -0.83145500 |
| C                                               | -4.12532900 | -0.77989900 | -0.36778400 |
| C                                               | -5.50688500 | -1.09868200 | -0.59915100 |

|   |             |             |             |
|---|-------------|-------------|-------------|
| O | -6.03151700 | -1.99220000 | -1.24215100 |
| N | -6.33975700 | -0.13759800 | 0.07787300  |
| H | -7.32818800 | -0.26353400 | -0.10787900 |
| C | -5.91478100 | 0.92326700  | 0.83165100  |
| N | -6.88859000 | 1.75692100  | 1.34435000  |
| H | -7.74458500 | 1.31482300  | 1.65631900  |
| H | -6.51151200 | 2.41246800  | 2.01773000  |
| N | -4.64841200 | 1.20128000  | 1.03815200  |
| C | -3.81691000 | 0.33245900  | 0.41196300  |
| N | -0.65764800 | -1.00066400 | -0.45202200 |
| H | -0.51515800 | -1.90022100 | -0.89240700 |
| C | 0.41154000  | -0.08681700 | -0.46959900 |
| C | 1.70804000  | -0.54995000 | -0.13605000 |
| C | 0.21735700  | 1.25638700  | -0.82092800 |
| C | 2.80326000  | 0.36291200  | -0.19866800 |
| C | 1.96440400  | -1.90304000 | 0.26740500  |
| C | 1.28216200  | 2.14817100  | -0.84275900 |
| H | -0.76996000 | 1.59235100  | -1.11860200 |
| C | 4.12376200  | -0.08781100 | 0.09911300  |
| C | 2.58708500  | 1.73046700  | -0.54769400 |
| C | 3.22652700  | -2.33160600 | 0.54807500  |
| H | 1.12957500  | -2.58699800 | 0.37812700  |
| H | 1.10767600  | 3.18269200  | -1.12626700 |
| C | 4.35270500  | -1.44844100 | 0.46390500  |
| C | 5.22301600  | 0.82248400  | 0.03605500  |
| C | 3.70868800  | 2.62298200  | -0.59591200 |
| H | 3.39820200  | -3.35887600 | 0.85853300  |
| C | 5.66106100  | -1.87209200 | 0.74614000  |
| C | 6.51291100  | 0.35314200  | 0.32674700  |
| C | 4.96924700  | 2.19003100  | -0.32100900 |
| H | 3.52800800  | 3.66066500  | -0.86448700 |
| C | 6.72694500  | -0.97900200 | 0.67558900  |
| H | 5.83259500  | -2.90881900 | 1.02333900  |
| H | 7.34923800  | 1.04551300  | 0.27814500  |
| H | 5.81096900  | 2.87596300  | -0.36607700 |
| H | 7.73341700  | -1.32251900 | 0.89683200  |
| H | -1.87522400 | 1.04116400  | 0.95301900  |



**Table A.39.** Cartesian coordinates of the nucleobase minimum for the  $^1\text{APG}$  adduct with  $\theta = 21.7^\circ$  and  $\phi = 110.1^\circ$ .

| Calculated energy (in Hartrees) = -1212.9762440 |             |             |             |
|-------------------------------------------------|-------------|-------------|-------------|
| Atom                                            | X           | Y           | Z           |
| N                                               | -1.71781700 | -0.08005700 | 0.05700800  |
| C                                               | -1.84349900 | -1.26341700 | -0.64941300 |
| N                                               | -3.09028200 | -1.58768800 | -0.89392200 |
| C                                               | -3.83181200 | -0.56033500 | -0.32446400 |
| C                                               | -5.25344600 | -0.36878000 | -0.25490800 |
| O                                               | -6.18039800 | -1.04469700 | -0.66992500 |
| N                                               | -5.53756100 | 0.85312200  | 0.45347900  |
| H                                               | -6.52789900 | 1.00719900  | 0.60430500  |
| C                                               | -4.62131500 | 1.71797100  | 0.98944500  |
| N                                               | -5.11844000 | 2.80432300  | 1.68441300  |
| H                                               | -5.95046000 | 3.23885200  | 1.30374600  |
| H                                               | -4.39135600 | 3.47293300  | 1.90800400  |
| N                                               | -3.32450800 | 1.53236200  | 0.91548400  |
| C                                               | -2.99996000 | 0.38450100  | 0.26992700  |
| H                                               | -0.86368200 | 0.37994400  | 0.33709400  |
| N                                               | -0.75359000 | -2.01379400 | -1.06812100 |
| H                                               | -1.05084600 | -2.95154300 | -1.31148700 |
| C                                               | 0.52612000  | -1.91585300 | -0.46554500 |
| C                                               | 1.31670100  | -0.76041000 | -0.65891100 |
| C                                               | 1.01443700  | -2.97963700 | 0.30063100  |
| C                                               | 2.57527900  | -0.66916300 | 0.01146500  |
| C                                               | 0.91553900  | 0.31941100  | -1.51931200 |
| C                                               | 2.25799600  | -2.90814200 | 0.91483900  |
| H                                               | 0.39244800  | -3.86086500 | 0.43206300  |
| C                                               | 3.37161700  | 0.50606900  | -0.12700000 |
| C                                               | 3.04909100  | -1.75453100 | 0.80792100  |
| C                                               | 1.67971100  | 1.43919000  | -1.64813500 |
| H                                               | -0.00301400 | 0.21534900  | -2.08544200 |
| H                                               | 2.61427800  | -3.74308300 | 1.51193400  |
| C                                               | 2.92290200  | 1.58484500  | -0.94660600 |
| C                                               | 4.62430500  | 0.60613500  | 0.55139600  |
| C                                               | 4.31545600  | -1.62905900 | 1.47082800  |
| H                                               | 1.36343400  | 2.24178500  | -2.30921800 |
| C                                               | 3.71794100  | 2.73548100  | -1.06005400 |
| C                                               | 5.38440100  | 1.77709100  | 0.40879100  |
| C                                               | 5.06562800  | -0.49961100 | 1.35385000  |
| H                                               | 4.66294900  | -2.46188200 | 2.07655800  |

|   |            |             |             |
|---|------------|-------------|-------------|
| C | 4.93352500 | 2.82858600  | -0.38620100 |
| H | 3.37475500 | 3.55522100  | -1.68557400 |
| H | 6.33671500 | 1.85367800  | 0.92678500  |
| H | 6.02065900 | -0.41630100 | 1.86574300  |
| H | 5.53687600 | 3.72635200  | -0.48475700 |

**Table A.40.** Cartesian coordinates of the nucleobase minimum for the  $^1\text{APG}$  adduct with  $\theta = 177.7^\circ$  and  $\phi = 354.8^\circ$ .

| Calculated energy (in Hartrees) = -1212.9740911 |             |             |             |
|-------------------------------------------------|-------------|-------------|-------------|
| Atom                                            | X           | Y           | Z           |
| N                                               | -2.94633700 | -1.48193600 | -0.16389400 |
| C                                               | -2.00722500 | -0.45673100 | -0.06327100 |
| N                                               | -2.56703700 | 0.72232000  | 0.06665700  |
| C                                               | -3.93231100 | 0.46963200  | 0.05977400  |
| C                                               | -5.04079000 | 1.37683300  | 0.15574600  |
| O                                               | -5.08197500 | 2.59046700  | 0.26641900  |
| N                                               | -6.27335000 | 0.63273400  | 0.09284000  |
| H                                               | -7.09707300 | 1.22331900  | 0.08219400  |
| C                                               | -6.40151200 | -0.72357300 | -0.03818500 |
| N                                               | -7.68289000 | -1.22549200 | -0.13712700 |
| H                                               | -8.39441700 | -0.76887400 | 0.42002500  |
| H                                               | -7.70491800 | -2.23508400 | -0.06404000 |
| N                                               | -5.37669000 | -1.54109100 | -0.12294000 |
| C                                               | -4.18904700 | -0.88921600 | -0.08286300 |
| N                                               | -0.66910500 | -0.77302400 | -0.09056600 |
| H                                               | -0.46231500 | -1.72474600 | -0.34853000 |
| C                                               | 0.43516200  | 0.09168000  | -0.04901300 |
| C                                               | 1.73222700  | -0.48476800 | 0.01623100  |
| C                                               | 0.28018100  | 1.48466800  | -0.07734200 |
| C                                               | 2.86895100  | 0.37988100  | -0.01316400 |
| C                                               | 1.96660800  | -1.89815700 | 0.11722000  |
| C                                               | 1.39524100  | 2.31082600  | -0.07823600 |
| H                                               | -0.71810400 | 1.89958300  | -0.10145400 |
| C                                               | 4.18796000  | -0.16652900 | 0.01174500  |
| C                                               | 2.69799600  | 1.79565000  | -0.06321300 |
| C                                               | 3.22491400  | -2.41928500 | 0.14309300  |
| H                                               | 1.13070300  | -2.58482600 | 0.21952500  |
| H                                               | 1.25350000  | 3.38801400  | -0.10605000 |
| C                                               | 4.38258100  | -1.57859100 | 0.07811400  |
| C                                               | 5.32486300  | 0.69925000  | -0.02682900 |

|   |             |             |             |
|---|-------------|-------------|-------------|
| C | 3.85661600  | 2.63911400  | -0.09475000 |
| H | 3.36938200  | -3.49311000 | 0.22891600  |
| C | 5.68735600  | -2.09630300 | 0.09626400  |
| C | 6.61032600  | 0.13626100  | -0.00812200 |
| C | 5.11366400  | 2.11808300  | -0.08122600 |
| H | 3.70543600  | 3.71494800  | -0.13257800 |
| C | 6.78754700  | -1.24449300 | 0.05150400  |
| H | 5.82841000  | -3.17277900 | 0.14768200  |
| H | 7.47359500  | 0.79595200  | -0.03889300 |
| H | 5.98362900  | 2.76884800  | -0.10887100 |
| H | 7.79176900  | -1.65863300 | 0.06550000  |
| H | -2.78378900 | -2.47631700 | -0.23043200 |

**Table A.41.** Cartesian coordinates of the nucleobase minimum for the  ${}^1\text{APG}$  adduct with  $\theta = 167.0^\circ$  and  $\phi = 118.4^\circ$ .

| Calculated energy (in Hartrees) = -1212.9706277 |             |             |             |
|-------------------------------------------------|-------------|-------------|-------------|
| Atom                                            | X           | Y           | Z           |
| N                                               | 2.97329200  | 1.29324600  | -1.22716200 |
| C                                               | 1.77482200  | 1.34356800  | -0.52333100 |
| N                                               | 1.78976600  | 0.60258300  | 0.55459500  |
| C                                               | 3.04864700  | 0.01856300  | 0.56102900  |
| C                                               | 3.63742900  | -0.90567400 | 1.49089000  |
| O                                               | 3.19571900  | -1.42920500 | 2.49864600  |
| N                                               | 4.98109400  | -1.21584800 | 1.06990000  |
| H                                               | 5.42562000  | -1.92385000 | 1.64307400  |
| C                                               | 5.62381300  | -0.72662400 | -0.03518700 |
| N                                               | 6.89152200  | -1.21024500 | -0.28717600 |
| H                                               | 7.47665500  | -1.39301400 | 0.51866900  |
| H                                               | 7.36343700  | -0.68494700 | -1.01278600 |
| N                                               | 5.07698400  | 0.12414400  | -0.87254000 |
| C                                               | 3.80248000  | 0.44042500  | -0.53261300 |
| N                                               | 0.71440300  | 2.09286900  | -1.02917900 |
| H                                               | 0.95671600  | 3.06392000  | -1.19733800 |
| C                                               | -0.59084500 | 1.94122000  | -0.48638700 |
| C                                               | -1.27952700 | 0.71457700  | -0.60906200 |
| C                                               | -1.20924900 | 3.03991800  | 0.11315400  |
| C                                               | -2.59464800 | 0.61479100  | -0.06339100 |
| C                                               | -0.72206700 | -0.42946600 | -1.27244700 |
| C                                               | -2.50120700 | 2.95049800  | 0.61660200  |
| H                                               | -0.65753400 | 3.97273900  | 0.20229200  |

|   |             |             |             |
|---|-------------|-------------|-------------|
| C | -3.31014300 | -0.61686800 | -0.15017700 |
| C | -3.21125700 | 1.74446700  | 0.55613600  |
| C | -1.40799200 | -1.60131800 | -1.35581100 |
| H | 0.26008400  | -0.34657000 | -1.72064700 |
| H | -2.96147400 | 3.81721600  | 1.08309500  |
| C | -2.71752000 | -1.74606800 | -0.78933500 |
| C | -4.62451500 | -0.72422400 | 0.39815600  |
| C | -4.53481800 | 1.60831400  | 1.09384400  |
| H | -0.96859500 | -2.45649000 | -1.86269700 |
| C | -3.43581200 | -2.94980800 | -0.85823900 |
| C | -5.30415800 | -1.94820900 | 0.30642600  |
| C | -5.20917000 | 0.42891300  | 1.02317000  |
| H | -4.98842500 | 2.47596200  | 1.56625900  |
| C | -4.71368500 | -3.04760200 | -0.31342100 |
| H | -2.98063500 | -3.80886000 | -1.34414100 |
| H | -6.30349500 | -2.02907200 | 0.72621000  |
| H | -6.20950400 | 0.33869300  | 1.43856200  |
| H | -5.25525600 | -3.98741000 | -0.37364600 |
| H | 3.16376200  | 1.69172500  | -2.13516600 |

**Table A.42.** Cartesian coordinates of the nucleobase minimum for the  $^1\text{APG}$  adduct with  $\theta = 192.2^\circ$  and  $\phi = 241.7^\circ$ .

| Calculated energy (in Hartrees) = -1212.9705809 |             |             |             |
|-------------------------------------------------|-------------|-------------|-------------|
| Atom                                            | X           | Y           | Z           |
| N                                               | -2.96132200 | 1.26047100  | -1.25586500 |
| C                                               | -1.77189400 | 1.33877500  | -0.53951000 |
| N                                               | -1.79339500 | 0.62361300  | 0.55560700  |
| C                                               | -3.04713900 | 0.02915900  | 0.56166200  |
| C                                               | -3.64890800 | -0.85577300 | 1.52105300  |
| O                                               | -3.22653800 | -1.32503800 | 2.56320700  |
| N                                               | -4.98720800 | -1.18217500 | 1.09566500  |
| H                                               | -5.49385500 | -1.74628900 | 1.76835700  |
| C                                               | -5.61468700 | -0.73944700 | -0.03745100 |
| N                                               | -6.92319400 | -1.13594500 | -0.22319900 |
| H                                               | -7.16080900 | -2.07395000 | 0.07479400  |
| H                                               | -7.26255300 | -0.91261100 | -1.15057400 |
| N                                               | -5.05257700 | 0.06769400  | -0.90736900 |
| C                                               | -3.79270200 | 0.42037400  | -0.54918800 |
| N                                               | -0.71167600 | 2.08562200  | -1.05115500 |
| H                                               | -0.95446800 | 3.05624700  | -1.22196600 |

|   |             |             |             |
|---|-------------|-------------|-------------|
| C | 0.59198600  | 1.93748700  | -0.50360100 |
| C | 1.28004500  | 0.70963100  | -0.61753700 |
| C | 1.20960200  | 3.03877100  | 0.09183600  |
| C | 2.59343900  | 0.61137400  | -0.06773100 |
| C | 0.72322700  | -0.43736800 | -1.27629000 |
| C | 2.50000900  | 2.95066200  | 0.59962500  |
| H | 0.65848300  | 3.97248900  | 0.17467800  |
| C | 3.30804800  | -0.62134700 | -0.14618800 |
| C | 3.20921600  | 1.74374100  | 0.54769000  |
| C | 1.40832300  | -1.61028000 | -1.35191600 |
| H | -0.25747600 | -0.35541000 | -1.72803600 |
| H | 2.95958300  | 3.81936000  | 1.06310100  |
| C | 2.71626200  | -1.75330300 | -0.78131000 |
| C | 4.62079900  | -0.72700700 | 0.40637000  |
| C | 4.53103200  | 1.60909800  | 1.09005300  |
| H | 0.96981100  | -2.46759600 | -1.85602200 |
| C | 3.43386300  | -2.95789300 | -0.84239800 |
| C | 5.29971700  | -1.95195800 | 0.32252900  |
| C | 5.20457700  | 0.42877900  | 1.02735100  |
| H | 4.98399400  | 2.47869700  | 1.55947500  |
| C | 4.71017200  | -3.05392500 | -0.29364200 |
| H | 2.97946500  | -3.81897700 | -1.32547100 |
| H | 6.29781200  | -2.03152500 | 0.74548800  |
| H | 6.20357100  | 0.33978900  | 1.44620300  |
| H | 5.25128300  | -3.99436700 | -0.34788300 |
| H | -3.14686700 | 1.64218800  | -2.17203200 |

**Table A.43.** Cartesian coordinates of nucleoside minimum for the <sup>AN</sup>dG adduct with  $\chi = 227.2^\circ$ ,  $\theta = 180.4^\circ$  and  $\phi = 162.8^\circ$ .

| Calculated energy (in Hartrees) = -1250.4854482 |             |             |             |
|-------------------------------------------------|-------------|-------------|-------------|
| Atom                                            | X           | Y           | Z           |
| N                                               | -0.84202700 | 0.32347400  | -0.11559100 |
| C                                               | 0.06232400  | -0.74988600 | -0.04674900 |
| N                                               | -0.54490100 | -1.91429500 | -0.00085800 |
| C                                               | -1.89847600 | -1.61754400 | -0.04169500 |
| C                                               | -3.03925300 | -2.48762500 | 0.00076800  |
| O                                               | -3.12725200 | -3.70111500 | 0.09206600  |
| N                                               | -4.24544200 | -1.70028900 | -0.07107500 |
| H                                               | -5.08997600 | -2.25185500 | 0.02762600  |
| C                                               | -4.32431200 | -0.33736800 | -0.15689400 |

|   |             |             |             |
|---|-------------|-------------|-------------|
| N | -5.58941000 | 0.22118700  | -0.15202200 |
| H | -6.30295800 | -0.28725100 | -0.66060300 |
| H | -5.56504000 | 1.20516500  | -0.39106700 |
| N | -3.26988700 | 0.44311500  | -0.18902300 |
| C | -2.10414600 | -0.24778300 | -0.12015000 |
| N | 1.41072300  | -0.50879400 | -0.04820600 |
| H | 1.69743000  | 0.43503300  | -0.28702200 |
| C | 2.43617000  | -1.46536900 | 0.00756800  |
| C | 3.72625700  | -1.05714600 | -0.37224100 |
| C | 2.23712600  | -2.78599900 | 0.44057700  |
| C | 4.79797000  | -1.94097400 | -0.30363800 |
| C | 3.32178700  | -3.65948100 | 0.49934100  |
| H | 1.24123600  | -3.11556000 | 0.70159400  |
| C | 4.60560100  | -3.25156000 | 0.13632400  |
| H | 3.15201200  | -4.67921100 | 0.83439900  |
| H | 3.87288300  | -0.04604200 | -0.73864300 |
| H | 5.78655100  | -1.60425700 | -0.60507600 |
| H | 5.44044700  | -3.94415100 | 0.18651600  |
| C | -0.60317000 | 1.74155700  | -0.06187500 |
| O | 0.36525400  | 2.09741000  | -1.05335800 |
| C | -0.01632200 | 2.27075200  | 1.24976500  |
| C | 1.01657100  | 3.31326500  | -0.65402300 |
| C | 0.57389700  | 3.60745100  | 0.79822700  |
| H | 0.77293100  | 1.60245700  | 1.60367400  |
| H | -0.77514100 | 2.37650100  | 2.03049600  |
| C | 2.51381900  | 3.13646400  | -0.81880100 |
| H | 0.67677500  | 4.13767200  | -1.29256500 |
| O | -0.41405900 | 4.63366600  | 0.71372600  |
| H | 1.41845800  | 3.92337700  | 1.42275600  |
| O | 2.93424900  | 2.14680000  | 0.11905800  |
| H | 3.01048600  | 4.09952500  | -0.63122500 |
| H | 2.73077400  | 2.82502500  | -1.84961300 |
| H | -0.72588900 | 4.81727800  | 1.61356000  |
| H | 3.88960000  | 2.01090400  | 0.03808000  |
| H | -1.56456400 | 2.21798500  | -0.27641000 |

**Table A.44.** Cartesian coordinates of nucleoside minimum for the <sup>AN</sup>dG adduct with  $\chi = 139.1^\circ$ ,  $\theta = 184.5^\circ$  and  $\phi = 5.8^\circ$ .

| Calculated energy (in Hartrees) = -1250.4780571 |             |             |             |
|-------------------------------------------------|-------------|-------------|-------------|
| Atom                                            | X           | Y           | Z           |
| N                                               | 0.49152500  | 0.08143800  | -0.27512200 |
| C                                               | -0.81650300 | 0.56225300  | -0.08625600 |
| N                                               | -0.85858000 | 1.85218400  | 0.15099000  |
| C                                               | 0.46315300  | 2.26195400  | 0.09826700  |
| C                                               | 1.01543400  | 3.57396400  | 0.28455500  |
| O                                               | 0.48209900  | 4.64270700  | 0.53443300  |
| N                                               | 2.44614400  | 3.51979000  | 0.14042300  |
| H                                               | 2.90594600  | 4.40143200  | 0.33644600  |
| C                                               | 3.18686500  | 2.40755900  | -0.14680700 |
| N                                               | 4.56051200  | 2.56653200  | -0.19904200 |
| H                                               | 4.89430300  | 3.42481300  | -0.62149100 |
| H                                               | 5.01891100  | 1.74349600  | -0.57042300 |
| N                                               | 2.66530900  | 1.21712800  | -0.32044600 |
| C                                               | 1.31507600  | 1.19473400  | -0.16436700 |
| N                                               | -1.86321400 | -0.31725600 | -0.18094400 |
| H                                               | -1.60097700 | -1.29111200 | -0.07799300 |
| C                                               | -3.23528600 | -0.03458200 | -0.11519000 |
| C                                               | -4.11219500 | -1.13240800 | -0.13400600 |
| C                                               | -3.75914800 | 1.26565600  | -0.05196100 |
| C                                               | -5.48800900 | -0.93619500 | -0.09175700 |
| C                                               | -5.14187300 | 1.44260700  | -0.01359100 |
| H                                               | -3.08368800 | 2.10953000  | -0.01893000 |
| C                                               | -6.01554100 | 0.35545200  | -0.03234500 |
| H                                               | -5.53642700 | 2.45393900  | 0.03700200  |
| H                                               | -3.70615400 | -2.14049000 | -0.18761100 |
| H                                               | -6.14944600 | -1.79832700 | -0.10679700 |
| H                                               | -7.08992900 | 0.50981500  | 0.00084200  |
| C                                               | 0.80996500  | -1.25631100 | -0.73809400 |
| O                                               | 0.14133600  | -2.19234900 | 0.11743400  |
| C                                               | 2.28020200  | -1.65767400 | -0.67476800 |
| C                                               | 0.91877600  | -3.39693900 | 0.25065200  |
| C                                               | 2.17173200  | -3.18299300 | -0.61092700 |
| H                                               | 2.72784700  | -1.27902000 | 0.24309200  |
| H                                               | 2.85838300  | -1.29651200 | -1.52780800 |
| C                                               | 1.20032300  | -3.65677900 | 1.72095400  |

|   |            |             |             |
|---|------------|-------------|-------------|
| H | 0.34073800 | -4.23438200 | -0.15603200 |
| O | 1.89190700 | -3.76299000 | -1.88540600 |
| H | 3.06120100 | -3.63545900 | -0.15580600 |
| O | 2.14535300 | -2.70379800 | 2.18584000  |
| H | 1.58264700 | -4.68553000 | 1.82182900  |
| H | 0.25152100 | -3.58926500 | 2.27285400  |
| H | 2.66224800 | -3.60771000 | -2.45362900 |
| H | 2.31781000 | -2.87765900 | 3.12247300  |
| H | 0.40982000 | -1.37660800 | -1.75359600 |

**Table A.45.** Cartesian coordinates of nucleoside minimum for the <sup>AN</sup>dG adduct with  $\chi = 68.3^\circ$ ,  $\theta = 68.6^\circ$  and  $\phi = 174.3^\circ$ .

| Calculated energy (in Hartrees) = -1250.4741386 |             |             |             |
|-------------------------------------------------|-------------|-------------|-------------|
| Atom                                            | X           | Y           | Z           |
| N                                               | 0.22199600  | -0.47054500 | 0.47999300  |
| C                                               | -0.37944500 | -1.72717200 | 0.60990700  |
| N                                               | 0.46435300  | -2.71758600 | 0.46032000  |
| C                                               | 1.68165800  | -2.10922100 | 0.20356300  |
| C                                               | 2.96661800  | -2.69288000 | -0.07459100 |
| O                                               | 3.32578000  | -3.85684400 | -0.12862900 |
| N                                               | 3.92368200  | -1.64862600 | -0.33135000 |
| H                                               | 4.83179100  | -1.99420600 | -0.61997200 |
| C                                               | 3.67761700  | -0.30193500 | -0.30465900 |
| N                                               | 4.72306300  | 0.54034800  | -0.63187100 |
| H                                               | 5.63743300  | 0.28373200  | -0.27931500 |
| H                                               | 4.48051700  | 1.50709500  | -0.44867400 |
| N                                               | 2.50354200  | 0.21264500  | -0.03790300 |
| C                                               | 1.54951300  | -0.72348600 | 0.18795500  |
| N                                               | -1.70966800 | -1.87660800 | 0.98581200  |
| H                                               | -1.84018200 | -2.77323600 | 1.44007700  |
| C                                               | -2.80773000 | -1.49787100 | 0.17135300  |
| C                                               | -4.09676600 | -1.81548400 | 0.62258300  |
| C                                               | -2.65216500 | -0.80813500 | -1.03674600 |
| C                                               | -5.21173900 | -1.45173800 | -0.12600300 |
| C                                               | -3.77880800 | -0.43287900 | -1.76788900 |
| H                                               | -1.66529900 | -0.55327700 | -1.40271600 |
| C                                               | -5.06181500 | -0.75227400 | -1.32544700 |
| H                                               | -3.64252700 | 0.10686900  | -2.70087500 |
| H                                               | -4.21619900 | -2.34385000 | 1.56534600  |
| H                                               | -6.20309200 | -1.71117000 | 0.23518600  |



|   |             |             |             |
|---|-------------|-------------|-------------|
| H | -5.93282600 | -0.46125100 | -1.90460500 |
| C | -0.40954600 | 0.81811700  | 0.64295900  |
| O | -0.44024700 | 1.48126200  | -0.61737200 |
| C | 0.30198800  | 1.80344200  | 1.57215600  |
| C | -0.51542800 | 2.89718200  | -0.39970100 |
| C | -0.31514300 | 3.12670100  | 1.11410200  |
| H | 1.37262700  | 1.79521200  | 1.36235700  |
| H | 0.12288000  | 1.58944500  | 2.63044000  |
| C | 0.49537300  | 3.60704200  | -1.28447900 |
| H | -1.52042200 | 3.24613100  | -0.66922800 |
| O | -1.61579100 | 3.33812500  | 1.66647000  |
| H | 0.34197500  | 3.98139700  | 1.31141300  |
| O | 1.80086600  | 3.48191800  | -0.73008800 |
| H | 0.20000900  | 4.66633100  | -1.35700500 |
| H | 0.44056900  | 3.16614700  | -2.29004300 |
| H | -1.51133900 | 3.45468100  | 2.62355500  |
| H | 2.41158000  | 3.94556200  | -1.32127400 |
| H | -1.42621700 | 0.62484800  | 0.99689500  |

**Table A.46.** Cartesian coordinates of nucleoside minimum for the <sup>AN</sup>dG adduct with  $\chi = 53.9^\circ$ ,  $\theta = 296.7^\circ$  and  $\phi = 356.0^\circ$ .

| Calculated energy (in Hartrees) = -1250.4739861 |             |             |             |
|-------------------------------------------------|-------------|-------------|-------------|
| Atom                                            | X           | Y           | Z           |
| N                                               | 0.08879300  | -0.45457900 | -0.27720100 |
| C                                               | -0.53535800 | -1.61779200 | -0.73620900 |
| N                                               | 0.25338500  | -2.66256300 | -0.74211600 |
| C                                               | 1.47225900  | -2.18188700 | -0.28730500 |
| C                                               | 2.72764400  | -2.85999700 | -0.10595700 |
| O                                               | 3.05586400  | -4.01481200 | -0.31991300 |
| N                                               | 3.70563600  | -1.93353300 | 0.40348800  |
| H                                               | 4.63780900  | -2.32586500 | 0.47031700  |
| C                                               | 3.49534900  | -0.61015200 | 0.68633800  |
| N                                               | 4.58275700  | 0.13077400  | 1.11119600  |
| H                                               | 5.21095800  | -0.32446200 | 1.76290200  |
| H                                               | 4.28895600  | 1.05151200  | 1.41677300  |
| N                                               | 2.34232100  | -0.01230900 | 0.52591000  |
| C                                               | 1.38658000  | -0.82659900 | 0.01857200  |
| N                                               | -1.85092400 | -1.60022200 | -1.18039900 |
| H                                               | -2.03038100 | -2.35391800 | -1.83288800 |
| C                                               | -2.94241900 | -1.27533500 | -0.34053800 |

|   |             |             |             |
|---|-------------|-------------|-------------|
| C | -2.78184300 | -1.00278300 | 1.02486400  |
| C | -4.22229500 | -1.20439700 | -0.90829500 |
| C | -3.88657500 | -0.64739200 | 1.79807700  |
| C | -5.32029600 | -0.86776200 | -0.12245000 |
| H | -4.34417400 | -1.40414600 | -1.96958000 |
| C | -5.16054100 | -0.57811900 | 1.23464500  |
| H | -6.30574900 | -0.82127900 | -0.57741400 |
| H | -1.80159700 | -1.08715900 | 1.48076800  |
| H | -3.74727100 | -0.44317700 | 2.85635900  |
| H | -6.01763200 | -0.30693600 | 1.84333000  |
| C | -0.41469000 | 0.89866900  | -0.38555700 |
| O | 0.39334300  | 1.61328700  | -1.31316500 |
| C | -0.36269100 | 1.73104900  | 0.89639700  |
| C | 0.36051600  | 3.01184000  | -0.99161700 |
| C | -0.43418800 | 3.14747900  | 0.32312200  |
| H | 0.59666700  | 1.57316800  | 1.39331800  |
| H | -1.18666600 | 1.50172700  | 1.57825400  |
| C | 1.77255700  | 3.57640400  | -0.94258200 |
| H | -0.19045300 | 3.53764000  | -1.78155100 |
| O | -1.76669200 | 3.50860900  | -0.04630100 |
| H | 0.00734800  | 3.89164900  | 0.99545400  |
| O | 2.38143700  | 3.29525600  | 0.31506300  |
| H | 1.70900900  | 4.66432200  | -1.10534100 |
| H | 2.34256700  | 3.13643800  | -1.77268100 |
| H | -2.29155600 | 3.56290400  | 0.76734400  |
| H | 3.27213800  | 3.67430600  | 0.29194800  |
| H | -1.43941800 | 0.82222600  | -0.75797500 |

**Table A.47.** Cartesian coordinates of nucleoside minimum for the  $^{1\text{ANP}}$ dG adduct with  $\chi = 222.7^\circ$ ,  $\theta = 166.5^\circ$  and  $\phi = 338.4^\circ$ .

| Calculated energy (in Hartrees) = -1404.1903584 |             |             |             |
|-------------------------------------------------|-------------|-------------|-------------|
| Atom                                            | X           | Y           | Z           |
| N                                               | -1.52525200 | 0.29828500  | -0.18423500 |
| C                                               | -0.56545200 | -0.72509700 | -0.11377400 |
| N                                               | -1.10394200 | -1.92326100 | -0.11804500 |
| C                                               | -2.47027700 | -1.70194200 | -0.19094100 |
| C                                               | -3.56098700 | -2.63554500 | -0.19937900 |
| O                                               | -3.58240900 | -3.85377000 | -0.13779900 |
| N                                               | -4.80774000 | -1.91579900 | -0.28435000 |
| H                                               | -5.62157500 | -2.51658000 | -0.22118700 |

|   |             |             |             |
|---|-------------|-------------|-------------|
| C | -4.96217700 | -0.55751300 | -0.33471000 |
| N | -6.25686300 | -0.07182900 | -0.34633400 |
| H | -6.92851700 | -0.60558700 | -0.88510000 |
| H | -6.28383000 | 0.91829000  | -0.55816200 |
| N | -3.95370000 | 0.28190300  | -0.31836300 |
| C | -2.75198600 | -0.34343900 | -0.23889000 |
| N | 0.76929000  | -0.41806100 | -0.10139800 |
| H | 1.00934100  | 0.56472000  | -0.17877900 |
| C | 1.82108600  | -1.28624300 | 0.22622700  |
| C | 3.13519300  | -0.92924400 | -0.22774000 |
| C | 1.64253900  | -2.42951500 | 0.98756200  |
| C | 4.25646300  | -1.69003900 | 0.23871200  |
| C | 3.37243400  | 0.12998000  | -1.14531900 |
| C | 2.75494300  | -3.19756500 | 1.38819000  |
| H | 0.64193400  | -2.72675000 | 1.27098800  |
| C | 5.56335200  | -1.29162300 | -0.15264200 |
| C | 4.03815800  | -2.82765400 | 1.05570800  |
| C | 4.65295400  | 0.48135600  | -1.51559500 |
| H | 2.52681800  | 0.63463700  | -1.59913100 |
| H | 2.58364400  | -4.08398400 | 1.99245400  |
| C | 5.76438400  | -0.22256500 | -0.99429300 |
| H | 4.80881700  | 1.28082100  | -2.23503000 |
| C | -1.37757800 | 1.72463300  | -0.02382800 |
| O | -0.35859800 | 2.20242100  | -0.90430600 |
| C | -0.93529400 | 2.18827400  | 1.36767100  |
| C | 0.23114500  | 3.39380700  | -0.35048700 |
| C | -0.36595700 | 3.57456700  | 1.06136400  |
| H | -0.15071500 | 1.53046500  | 1.75017900  |
| H | -1.76562700 | 2.21091200  | 2.07943400  |
| C | 1.73860700  | 3.22636500  | -0.37299700 |
| H | -0.05225800 | 4.25674700  | -0.96420700 |
| O | -1.38845000 | 4.56246000  | 0.93865900  |
| H | 0.39656300  | 3.88218600  | 1.78735600  |
| O | 2.06283400  | 2.18621100  | 0.54219400  |
| H | 2.21195900  | 4.17595300  | -0.08289200 |
| H | 2.06156300  | 2.97542700  | -1.39304100 |
| H | -1.79705700 | 4.67343600  | 1.81121600  |
| H | 3.02369200  | 2.05324600  | 0.55187600  |
| H | -2.34694100 | 2.15800300  | -0.28740400 |
| H | 6.77113900  | 0.06404800  | -1.28478700 |
| H | 6.40908200  | -1.86242500 | 0.22271900  |

|   |            |             |            |
|---|------------|-------------|------------|
| H | 4.89309000 | -3.40678300 | 1.39355000 |
|---|------------|-------------|------------|

**Table A.48.** Cartesian coordinates of nucleoside minimum for the  $^{1\text{ANP}}\text{dG}$  adduct with  $\chi = 141.1^\circ$ ,  $\theta = 185.4^\circ$  and  $\phi = 357.6^\circ$ .

| Calculated energy (in Hartrees) = -1404.1821590 |             |             |             |
|-------------------------------------------------|-------------|-------------|-------------|
| Atom                                            | X           | Y           | Z           |
| N                                               | -1.20610400 | -0.03236400 | -0.23142700 |
| C                                               | -0.15090400 | -0.94087100 | -0.02580500 |
| N                                               | -0.56431600 | -2.17113600 | 0.16406900  |
| C                                               | -1.94322200 | -2.09877800 | 0.05692700  |
| C                                               | -2.91965500 | -3.14397600 | 0.18061800  |
| O                                               | -2.79628400 | -4.33646100 | 0.40742200  |
| N                                               | -4.23765100 | -2.59616900 | -0.00193500 |
| H                                               | -4.98005000 | -3.26958600 | 0.14827300  |
| C                                               | -4.53879300 | -1.28998800 | -0.26934100 |
| N                                               | -5.87957100 | -0.96432700 | -0.36730200 |
| H                                               | -6.47279200 | -1.64549600 | -0.82601700 |
| H                                               | -6.01343800 | -0.02519800 | -0.72145600 |
| N                                               | -3.63353300 | -0.34812900 | -0.38365600 |
| C                                               | -2.36500400 | -0.79638300 | -0.18877700 |
| N                                               | 1.13306500  | -0.46205200 | -0.06107900 |
| H                                               | 1.19246000  | 0.53086000  | 0.12446700  |
| C                                               | 2.32985100  | -1.18613400 | 0.01223300  |
| C                                               | 3.54965400  | -0.43398200 | -0.11512400 |
| C                                               | 2.37156600  | -2.55918000 | 0.18651700  |
| C                                               | 4.79607700  | -1.13357800 | -0.00987500 |
| C                                               | 3.58158200  | 0.96721300  | -0.34890100 |
| C                                               | 3.61348400  | -3.22619600 | 0.25989500  |
| H                                               | 1.44419600  | -3.10741300 | 0.28043600  |
| C                                               | 6.00612500  | -0.39751400 | -0.10616400 |
| C                                               | 4.80117900  | -2.54015000 | 0.17800000  |
| C                                               | 4.77520600  | 1.64924900  | -0.44239100 |
| H                                               | 2.65923100  | 1.52324300  | -0.48674800 |
| H                                               | 3.61475000  | -4.30362700 | 0.39960500  |
| C                                               | 6.00287900  | 0.96270000  | -0.31147200 |
| H                                               | 4.77142200  | 2.71981200  | -0.62723200 |
| C                                               | -1.02971800 | 1.33876500  | -0.67789100 |
| O                                               | -0.08032600 | 1.97186400  | 0.19043600  |
| C                                               | -2.26872100 | 2.22604800  | -0.60624900 |
| C                                               | -0.39712200 | 3.36685700  | 0.36034400  |

|   |             |             |             |
|---|-------------|-------------|-------------|
| C | -1.64033200 | 3.61781400  | -0.50473900 |
| H | -2.82948400 | 2.00460600  | 0.30106300  |
| H | -2.92752300 | 2.10758800  | -1.46878400 |
| C | -0.57971900 | 3.66825800  | 1.83792800  |
| H | 0.43719800  | 3.96292800  | -0.02530000 |
| O | -1.16680600 | 4.09348300  | -1.76489100 |
| H | -2.32234100 | 4.33913700  | -0.03829600 |
| O | -1.79591000 | 3.08457600  | 2.28183600  |
| H | -0.58676600 | 4.76273500  | 1.96705700  |
| H | 0.28625600  | 3.26561300  | 2.38319800  |
| H | -1.93822700 | 4.22710200  | -2.33711500 |
| H | -1.90281500 | 3.28215900  | 3.22353900  |
| H | -0.60551000 | 1.32348300  | -1.69007600 |
| H | 6.93940500  | 1.50852400  | -0.38356700 |
| H | 6.94495500  | -0.93865900 | -0.01737800 |
| H | 5.75399200  | -3.05730400 | 0.25083000  |

**Table A.49.** Cartesian coordinates of nucleoside minimum for the  $^{1\text{ANP}}$ dG adduct with  $\chi = 69.5^\circ$ ,  $\theta = 64.3^\circ$  and  $\phi = 7.7^\circ$ .

| Calculated energy (in Hartrees) = -1404.1788452 |             |             |             |
|-------------------------------------------------|-------------|-------------|-------------|
| Atom                                            | X           | Y           | Z           |
| N                                               | -0.85583900 | -0.45924900 | -0.37159800 |
| C                                               | -0.08465500 | -1.62208200 | -0.27248800 |
| N                                               | -0.80225400 | -2.70299600 | -0.09521500 |
| C                                               | -2.11217500 | -2.25320400 | -0.05523600 |
| C                                               | -3.33272500 | -2.98439000 | 0.15274400  |
| O                                               | -3.53848100 | -4.17551300 | 0.31441900  |
| N                                               | -4.44407700 | -2.06981900 | 0.17084700  |
| H                                               | -5.32714800 | -2.51081100 | 0.40114900  |
| C                                               | -4.37670900 | -0.71250000 | 0.00534700  |
| N                                               | -5.55774000 | -0.00272300 | 0.11255800  |
| H                                               | -6.37888400 | -0.42792700 | -0.30117800 |
| H                                               | -5.42811800 | 0.96418200  | -0.16136200 |
| N                                               | -3.25680500 | -0.06403100 | -0.19457600 |
| C                                               | -2.16606300 | -0.86947800 | -0.19773100 |
| N                                               | 1.29011400  | -1.61710200 | -0.47117500 |
| H                                               | 1.60148600  | -2.55551900 | -0.69992900 |
| C                                               | 2.18584800  | -0.91348100 | 0.38069300  |
| C                                               | 3.55472600  | -0.82977600 | -0.02915900 |
| C                                               | 1.77143400  | -0.31324200 | 1.55227200  |

|   |             |             |             |
|---|-------------|-------------|-------------|
| C | 4.48339400  | -0.14768400 | 0.82175300  |
| C | 4.02482300  | -1.38905300 | -1.24560800 |
| C | 2.69087500  | 0.39422500  | 2.35779700  |
| H | 0.73349900  | -0.37449300 | 1.85576600  |
| C | 5.84798000  | -0.09240600 | 0.43560100  |
| C | 4.01968500  | 0.46572500  | 2.01517500  |
| C | 5.35620800  | -1.31323000 | -1.59117200 |
| H | 3.31775500  | -1.85635300 | -1.92291400 |
| H | 2.33345300  | 0.86438200  | 3.26920800  |
| C | 6.27982300  | -0.66703500 | -0.73790600 |
| H | 5.69760400  | -1.74346100 | -2.52835300 |
| C | -0.37106300 | 0.88696400  | -0.56829200 |
| O | -0.53494900 | 1.61857500  | 0.64280600  |
| C | -1.10166400 | 1.72356800  | -1.61979700 |
| C | -0.60552500 | 3.01821100  | 0.33752100  |
| C | -0.68601400 | 3.13528900  | -1.19979900 |
| H | -2.17897500 | 1.59784200  | -1.50371700 |
| H | -0.79963300 | 1.47349700  | -2.64163800 |
| C | -1.77480000 | 3.64407100  | 1.07823100  |
| H | 0.32056300  | 3.50359400  | 0.67097500  |
| O | 0.62870600  | 3.46999900  | -1.64853700 |
| H | -1.41465900 | 3.89219500  | -1.51219500 |
| O | -2.99475300 | 3.30736200  | 0.42684200  |
| H | -1.62476800 | 4.73587000  | 1.09378900  |
| H | -1.75840900 | 3.28099100  | 2.11587900  |
| H | 0.60379000  | 3.51827500  | -2.61680500 |
| H | -3.71429800 | 3.72175100  | 0.92474900  |
| H | 0.68983100  | 0.80348700  | -0.81955800 |
| H | 7.32825800  | -0.61583500 | -1.01755300 |
| H | 6.54980400  | 0.41966600  | 1.08906900  |
| H | 4.73044100  | 0.99378200  | 2.64481700  |

**Table A.50.** Cartesian coordinates of nucleoside minimum for the  $^{1\text{ANP}}$ dG adduct with  $\chi = 53.1^\circ$ ,  $\theta = 298.1^\circ$  and  $\phi = 347.1^\circ$ .

| Calculated energy (in Hartrees) = -1404.1786742 |            |             |             |
|-------------------------------------------------|------------|-------------|-------------|
| Atom                                            | X          | Y           | Z           |
| N                                               | 0.78307900 | -0.43781700 | -0.17868600 |
| C                                               | 0.06177800 | -1.59867100 | -0.46249700 |
| N                                               | 0.80274300 | -2.67749300 | -0.49315500 |
| C                                               | 2.08837200 | -2.22345400 | -0.23572500 |
| C                                               | 3.33007400 | -2.94464800 | -0.16086500 |

|   |             |             |             |
|---|-------------|-------------|-------------|
| O | 3.58595000  | -4.12545000 | -0.32755500 |
| N | 4.39814600  | -2.03190100 | 0.15726000  |
| H | 5.31554500  | -2.46253400 | 0.14188900  |
| C | 4.27437500  | -0.68481100 | 0.36998900  |
| N | 5.43178900  | 0.03189500  | 0.61263700  |
| H | 6.11817200  | -0.41092600 | 1.21212600  |
| H | 5.20935000  | 0.98021300  | 0.89332300  |
| N | 3.13409600  | -0.04580100 | 0.30543200  |
| C | 2.09299400  | -0.84751300 | -0.02119500 |
| N | -1.30121700 | -1.54766200 | -0.72523300 |
| H | -1.60186900 | -2.37574300 | -1.22808100 |
| C | -2.22545400 | -1.08191500 | 0.24716900  |
| C | -3.55180700 | -0.76828800 | -0.18843200 |
| C | -1.86897100 | -0.90604900 | 1.57063600  |
| C | -4.50132100 | -0.30850000 | 0.78036600  |
| C | -3.95549400 | -0.88590300 | -1.54333600 |
| C | -2.80502100 | -0.41399200 | 2.50757100  |
| H | -0.86766000 | -1.16734800 | 1.89551200  |
| C | -5.82614400 | -0.02742100 | 0.35557500  |
| C | -4.09632200 | -0.13128100 | 2.12961100  |
| C | -5.24880200 | -0.60028500 | -1.92120900 |
| H | -3.22409700 | -1.17789100 | -2.28929700 |
| H | -2.49736700 | -0.28863900 | 3.54193700  |
| C | -6.19667500 | -0.17658300 | -0.96127100 |
| H | -5.54069500 | -0.68996400 | -2.96350300 |
| C | 0.30522700  | 0.92609700  | -0.26410600 |
| O | 1.03122600  | 1.60633100  | -1.28050600 |
| C | 0.50625500  | 1.77518500  | 0.99147100  |
| C | 1.06238400  | 3.00989700  | -0.98003400 |
| C | 0.40291400  | 3.18299400  | 0.40350400  |
| H | 1.50979900  | 1.60432100  | 1.38707500  |
| H | -0.24538000 | 1.57155500  | 1.75952700  |
| C | 2.48465200  | 3.54226700  | -1.07629300 |
| H | 0.44931300  | 3.53754200  | -1.72143600 |
| O | -0.95370600 | 3.56055300  | 0.15967100  |
| H | 0.91989300  | 3.92996700  | 1.01641300  |
| O | 3.20635100  | 3.26668300  | 0.12176100  |
| H | 2.42880400  | 4.62882300  | -1.25038700 |
| H | 2.96259700  | 3.07709400  | -1.94960700 |
| H | -1.39552700 | 3.63661900  | 1.01958700  |
| H | 4.09850500  | 3.62483200  | 0.00606700  |

|   |             |            |             |
|---|-------------|------------|-------------|
| H | -0.75333900 | 0.87077000 | -0.53059400 |
| H | -7.21489700 | 0.04266300 | -1.26961300 |
| H | -6.54611900 | 0.31304400 | 1.09545600  |
| H | -4.82211300 | 0.22934000 | 2.85326400  |

**Table A.51.** Cartesian coordinates of nucleoside minimum for the  ${}^3\text{APHN}$  dG adduct with  $\chi = 225.9^\circ$ ,  $\theta = 181.0^\circ$  and  $\phi = 341.6^\circ$ .

| Calculated energy (in Hartrees) = -1557.8987699 |             |             |             |
|-------------------------------------------------|-------------|-------------|-------------|
| Atom                                            | X           | Y           | Z           |
| N                                               | -2.51965200 | 0.23390500  | -0.10745900 |
| C                                               | -1.44284600 | -0.66040000 | 0.00516500  |
| N                                               | -1.82743600 | -1.91620500 | -0.01403300 |
| C                                               | -3.20715500 | -1.86747400 | -0.14304300 |
| C                                               | -4.17059400 | -2.93011400 | -0.19730300 |
| O                                               | -4.04122400 | -4.14162000 | -0.13589400 |
| N                                               | -5.49262100 | -2.37150300 | -0.33798600 |
| H                                               | -6.22700600 | -3.06943700 | -0.30975400 |
| C                                               | -5.81294900 | -1.04317600 | -0.40262000 |
| N                                               | -7.15554000 | -0.72256200 | -0.47452600 |
| H                                               | -7.73300600 | -1.33738000 | -1.03543600 |
| H                                               | -7.29752400 | 0.25630500  | -0.69236100 |
| N                                               | -4.91842400 | -0.08429500 | -0.34643600 |
| C                                               | -3.65352200 | -0.55519700 | -0.21007300 |
| N                                               | -0.16555500 | -0.17486000 | 0.10153700  |
| H                                               | -0.05066000 | 0.81530000  | -0.08981900 |
| C                                               | 1.01827700  | -0.91462800 | 0.20051900  |
| C                                               | 2.21247200  | -0.25269100 | -0.07231700 |
| C                                               | 1.04803200  | -2.27521000 | 0.58951500  |
| C                                               | 3.46141900  | -0.88327000 | 0.05627100  |
| C                                               | 2.26603400  | -2.90799400 | 0.71431500  |
| H                                               | 0.11846500  | -2.79890900 | 0.76405800  |
| C                                               | 4.71907200  | -0.20036800 | -0.21217200 |
| C                                               | 3.49004700  | -2.24736500 | 0.46745400  |
| H                                               | 2.29240700  | -3.95248800 | 1.01477000  |
| C                                               | 4.78837500  | 1.14333800  | -0.64331300 |
| C                                               | 5.94445500  | -0.90892000 | -0.03513000 |
| C                                               | 4.74425100  | -2.91922800 | 0.62170900  |
| C                                               | 5.99867500  | 1.76796600  | -0.88154000 |
| C                                               | 7.17007000  | -0.24930900 | -0.28270200 |
| C                                               | 5.92165800  | -2.27930200 | 0.38660200  |



|   |             |             |             |
|---|-------------|-------------|-------------|
| H | 4.73396900  | -3.95974200 | 0.93660800  |
| C | 7.20450700  | 1.06697800  | -0.69714800 |
| H | 6.01761400  | 2.80172600  | -1.21544000 |
| H | 8.09427100  | -0.80434500 | -0.14154500 |
| H | 6.86955800  | -2.79622500 | 0.51052800  |
| H | 8.15494300  | 1.55858100  | -0.88392300 |
| C | -2.54309300 | 1.67100700  | -0.01576700 |
| O | -1.56742100 | 2.21816900  | -0.90834500 |
| C | -2.17885100 | 2.25872200  | 1.35045500  |
| C | -1.18194300 | 3.51981900  | -0.43781400 |
| C | -1.79646000 | 3.69065400  | 0.97091000  |
| H | -1.31828400 | 1.72932300  | 1.76830200  |
| H | -3.01058600 | 2.20899600  | 2.05909900  |
| C | 0.33133500  | 3.62128500  | -0.46794100 |
| H | -1.60819900 | 4.28665300  | -1.09575200 |
| O | -2.93844500 | 4.53029500  | 0.80806100  |
| H | -1.07934300 | 4.13104500  | 1.67438400  |
| O | 0.84263300  | 2.69857800  | 0.49234500  |
| H | 0.62417000  | 4.65297900  | -0.22505200 |
| H | 0.69095500  | 3.38413700  | -1.47851400 |
| H | -3.35672200 | 4.63148100  | 1.67726800  |
| H | 1.81055900  | 2.73863100  | 0.49826600  |
| H | -3.55068100 | 1.97647400  | -0.31393400 |
| H | 2.15482500  | 0.77845400  | -0.39512500 |
| H | 3.87292000  | 1.70298000  | -0.80785800 |

**Table A.52.** Cartesian coordinates of nucleoside minimum for the  ${}^3\text{APHN}$  dG adduct with  $\chi = 139.5^\circ$ ,  $\theta = 184.8^\circ$  and  $\phi = 5.8^\circ$ .

| Calculated energy (in Hartrees) = -1557.8907377 |             |             |             |
|-------------------------------------------------|-------------|-------------|-------------|
| Atom                                            | X           | Y           | Z           |
| N                                               | -2.21786700 | -0.00674600 | -0.28093600 |
| C                                               | -1.09251800 | -0.82604200 | -0.08420000 |
| N                                               | -1.40518000 | -2.07881000 | 0.14797100  |
| C                                               | -2.78774200 | -2.11349000 | 0.08266300  |
| C                                               | -3.67752700 | -3.22641500 | 0.25843200  |
| O                                               | -3.45639200 | -4.40029300 | 0.50706000  |
| N                                               | -5.03799700 | -2.78450600 | 0.10386800  |
| H                                               | -5.72210500 | -3.50842500 | 0.29137600  |
| C                                               | -5.44561500 | -1.51183400 | -0.18376500 |
| N                                               | -6.80956800 | -1.29084400 | -0.24699800 |

|   |             |             |             |
|---|-------------|-------------|-------------|
| H | -7.36274900 | -2.02626700 | -0.67063500 |
| H | -7.02524800 | -0.37384300 | -0.61834500 |
| N | -4.61830600 | -0.50759400 | -0.34786600 |
| C | -3.31447900 | -0.85393700 | -0.18145500 |
| N | 0.15574600  | -0.26369700 | -0.16491100 |
| H | 0.16575300  | 0.74534000  | -0.06464500 |
| C | 1.39900700  | -0.90301900 | -0.09515500 |
| C | 2.53159800  | -0.09234300 | -0.09903400 |
| C | 1.54121000  | -2.30936700 | -0.04030100 |
| C | 3.83024100  | -0.62955400 | -0.04972200 |
| C | 2.80871800  | -2.84850200 | 0.00710500  |
| H | 0.65816600  | -2.93306300 | -0.01920800 |
| C | 5.02930800  | 0.19760900  | -0.05310300 |
| C | 3.97122700  | -2.04741000 | 0.00532000  |
| H | 2.92376200  | -3.92862300 | 0.05294900  |
| C | 4.98956900  | 1.60869100  | -0.11027300 |
| C | 6.30758000  | -0.43357400 | 0.00242800  |
| C | 5.27575700  | -2.63405600 | 0.05985200  |
| C | 6.14656400  | 2.36601600  | -0.11180700 |
| C | 7.47623000  | 0.36147200  | 0.00016600  |
| C | 6.39655200  | -1.86370700 | 0.05928400  |
| H | 5.34938000  | -3.71793700 | 0.10216700  |
| C | 7.40451200  | 1.73888500  | -0.05580600 |
| H | 6.08177600  | 3.44958100  | -0.15708500 |
| H | 8.44152700  | -0.13711300 | 0.04331200  |
| H | 7.38314800  | -2.31774100 | 0.10109300  |
| H | 8.31221200  | 2.33572700  | -0.05713500 |
| C | -2.15792200 | 1.37166700  | -0.73314100 |
| O | -1.25649900 | 2.08044100  | 0.12651700  |
| C | -3.46304600 | 2.15830600  | -0.66035600 |
| C | -1.68156400 | 3.44655900  | 0.28950300  |
| C | -2.94318100 | 3.59541000  | -0.57231600 |
| H | -4.00082300 | 1.90083400  | 0.25119800  |
| H | -4.11449500 | 1.98228900  | -1.51881300 |
| C | -1.88406900 | 3.74192200  | 1.76606000  |
| H | -0.89823700 | 4.10529800  | -0.10179900 |
| O | -2.51136000 | 4.09563900  | -1.83802900 |
| H | -3.67654700 | 4.26649000  | -0.10892700 |
| O | -3.05098300 | 3.06939700  | 2.21654300  |
| H | -1.97491000 | 4.83335200  | 1.88937000  |
| H | -0.98884000 | 3.41001300  | 2.31177600  |

|   |             |            |             |
|---|-------------|------------|-------------|
| H | -3.29240900 | 4.16570700 | -2.40858800 |
| H | -3.17077900 | 3.26395300 | 3.15732800  |
| H | -1.74176700 | 1.38582900 | -1.74908500 |
| H | 2.39079800  | 0.98301700 | -0.14756600 |
| H | 4.03399600  | 2.11975700 | -0.15564500 |

**Table A.53.** Cartesian coordinates of nucleoside minimum for the  $^3\text{APHN}$ dG adduct with  $\chi = 54.7^\circ$ ,  $\theta = 299.9^\circ$  and  $\phi = 175.9^\circ$ .

| Calculated energy (in Hartrees) = -1557.8895159 |             |             |             |
|-------------------------------------------------|-------------|-------------|-------------|
| Atom                                            | X           | Y           | Z           |
| N                                               | 1.40590300  | -0.63621600 | -0.56097700 |
| C                                               | 1.09476300  | -1.96163500 | -0.87874500 |
| N                                               | 1.92416100  | -2.82953000 | -0.35680700 |
| C                                               | 2.85112700  | -2.05847700 | 0.32815100  |
| C                                               | 4.01383500  | -2.45288800 | 1.07745000  |
| O                                               | 4.49885200  | -3.55143000 | 1.28938400  |
| N                                               | 4.65712600  | -1.28327000 | 1.61819300  |
| H                                               | 5.53465600  | -1.48368300 | 2.08417700  |
| C                                               | 4.24144200  | 0.01237500  | 1.46353300  |
| N                                               | 5.03249900  | 1.00524000  | 2.01054900  |
| H                                               | 5.42643600  | 0.81667300  | 2.92473400  |
| H                                               | 4.56792300  | 1.90422800  | 1.95218600  |
| N                                               | 3.17330400  | 0.35156800  | 0.78747400  |
| C                                               | 2.54009500  | -0.70554900 | 0.22644200  |
| N                                               | 0.03297500  | -2.27990400 | -1.71577700 |
| H                                               | 0.16445900  | -3.18226500 | -2.15663500 |
| C                                               | -1.30634800 | -1.93530600 | -1.43945900 |
| C                                               | -1.67414200 | -1.27332900 | -0.27667800 |
| C                                               | -2.28714900 | -2.23830600 | -2.41104100 |
| C                                               | -2.99824000 | -0.83963400 | -0.06405600 |
| C                                               | -3.59325300 | -1.85696100 | -2.20544800 |
| H                                               | -1.99521000 | -2.75453400 | -3.32136800 |
| C                                               | -3.39253100 | -0.09224200 | 1.12056700  |
| C                                               | -3.98129800 | -1.13893900 | -1.05042600 |
| H                                               | -4.34767700 | -2.09010800 | -2.95231900 |
| C                                               | -2.48412100 | 0.25129200  | 2.14833200  |
| C                                               | -4.74936300 | 0.32710300  | 1.25593400  |
| C                                               | -5.33339500 | -0.70707800 | -0.86607300 |
| C                                               | -2.88727500 | 0.97756500  | 3.25372300  |
| C                                               | -5.13709600 | 1.06710500  | 2.39638700  |

|   |             |             |             |
|---|-------------|-------------|-------------|
| C | -5.70286500 | -0.00224700 | 0.23731000  |
| H | -6.06171700 | -0.95305200 | -1.63467800 |
| C | -4.22599800 | 1.39150300  | 3.38132900  |
| H | -2.16633100 | 1.22495800  | 4.02756600  |
| H | -6.17520300 | 1.37787700  | 2.48429000  |
| H | -6.73066100 | 0.32588100  | 0.36745900  |
| H | -4.53974300 | 1.96052900  | 4.25170700  |
| C | 0.84841800  | 0.56038500  | -1.15950700 |
| O | 1.85813200  | 1.19896900  | -1.93180900 |
| C | 0.37319900  | 1.63403500  | -0.17884600 |
| C | 1.57699100  | 2.60556100  | -2.00402700 |
| C | 0.36319600  | 2.86302200  | -1.08849000 |
| H | 1.11528500  | 1.75634800  | 0.61262000  |
| H | -0.60814600 | 1.41722600  | 0.25136200  |
| C | 2.81952600  | 3.41083100  | -1.65493100 |
| H | 1.28909800  | 2.85031700  | -3.03417300 |
| O | -0.78684600 | 2.88535800  | -1.93637000 |
| H | 0.46101200  | 3.79786500  | -0.52524100 |
| O | 2.97960900  | 3.49958600  | -0.24131500 |
| H | 2.70719300  | 4.41526400  | -2.09381400 |
| H | 3.68366600  | 2.92325000  | -2.12741000 |
| H | -1.56514900 | 3.00610900  | -1.37014000 |
| H | 3.78010900  | 4.01904400  | -0.07717700 |
| H | 0.02901800  | 0.23696600  | -1.80528800 |
| H | -0.91761200 | -1.09117700 | 0.47421600  |
| H | -1.44826400 | -0.06191700 | 2.08163800  |

**Table A.54.** Cartesian coordinates of nucleoside minimum for the  ${}^3\text{APHN}$ dG adduct with  $\chi = 68.6^\circ$ ,  $\theta = 67.3^\circ$  and  $\phi = 354.7^\circ$ .

| Calculated energy (in Hartrees) = -1557.8864595 |            |             |             |
|-------------------------------------------------|------------|-------------|-------------|
| Atom                                            | X          | Y           | Z           |
| N                                               | 1.84983800 | -0.50148900 | 0.44164300  |
| C                                               | 1.03585000 | -1.64004000 | 0.41641100  |
| N                                               | 1.70072600 | -2.74037800 | 0.16691600  |
| C                                               | 3.01271200 | -2.33077500 | -0.00035700 |
| C                                               | 4.18549700 | -3.09759000 | -0.32583400 |
| O                                               | 4.33824300 | -4.29335100 | -0.50825400 |
| N                                               | 5.31723600 | -2.21662400 | -0.44757700 |
| H                                               | 6.16086900 | -2.68280700 | -0.76117700 |
| C                                               | 5.30776100 | -0.85886300 | -0.27056000 |

|   |             |             |             |
|---|-------------|-------------|-------------|
| N | 6.49415000  | -0.18414900 | -0.48441300 |
| H | 7.33824700  | -0.63719900 | -0.15528100 |
| H | 6.42016900  | 0.78410000  | -0.19496800 |
| N | 4.23231200  | -0.17777200 | 0.03680300  |
| C | 3.12311900  | -0.95025400 | 0.14085500  |
| N | -0.31075800 | -1.60025300 | 0.75782200  |
| H | -0.60716900 | -2.50218100 | 1.11190500  |
| C | -1.30887200 | -0.93079400 | 0.00927900  |
| C | -2.62918700 | -1.05324300 | 0.42377900  |
| C | -0.99781400 | -0.13883300 | -1.11656300 |
| C | -3.68107700 | -0.40580900 | -0.25185600 |
| C | -2.01172800 | 0.52086400  | -1.77704600 |
| H | 0.02583500  | -0.03451100 | -1.45273200 |
| C | -5.07381100 | -0.52594700 | 0.15633600  |
| C | -3.36039000 | 0.40981700  | -1.37676400 |
| H | -1.77322900 | 1.13693200  | -2.64018400 |
| C | -5.48793600 | -1.31808300 | 1.25046300  |
| C | -6.07375200 | 0.18214500  | -0.57404700 |
| C | -4.39978400 | 1.09941100  | -2.08001500 |
| C | -6.81900100 | -1.40959800 | 1.61324600  |
| C | -7.42777600 | 0.07392500  | -0.18376700 |
| C | -5.69964400 | 0.99297300  | -1.69623800 |
| H | -4.12265400 | 1.71354800  | -2.93316100 |
| C | -7.80104800 | -0.70692500 | 0.89156700  |
| H | -7.10576300 | -2.02814000 | 2.45894000  |
| H | -8.17610000 | 0.62087700  | -0.75203600 |
| H | -6.48271000 | 1.51976400  | -2.23518900 |
| H | -8.84595900 | -0.78067500 | 1.17905900  |
| C | 1.44307000  | 0.85431400  | 0.72750200  |
| O | 1.55530800  | 1.62961100  | -0.46236700 |
| C | 2.28677100  | 1.61282800  | 1.75330900  |
| C | 1.70607100  | 3.01149400  | -0.10707600 |
| C | 1.90687300  | 3.05819200  | 1.42332800  |
| H | 3.34541900  | 1.44666600  | 1.54920300  |
| H | 2.05142600  | 1.33123700  | 2.78438200  |
| C | 2.83952200  | 3.62555800  | -0.91071800 |
| H | 0.77741100  | 3.54455200  | -0.34743200 |
| O | 0.64661900  | 3.42981800  | 1.98440100  |
| H | 2.69142700  | 3.76850100  | 1.70813000  |
| O | 4.09239000  | 3.22475800  | -0.36635800 |
| H | 2.72803700  | 4.72139600  | -0.87530700 |

|   |             |             |             |
|---|-------------|-------------|-------------|
| H | 2.73296500  | 3.30362000  | -1.95650900 |
| H | 0.74574100  | 3.43563600  | 2.94913200  |
| H | 4.78604100  | 3.63457600  | -0.90332400 |
| H | 0.40011700  | 0.80609800  | 1.05382600  |
| H | -2.83405400 | -1.65943700 | 1.29993300  |
| H | -4.75362300 | -1.87322000 | 1.82365900  |

**Table A.55.** Cartesian coordinates of nucleoside minimum for the  $^1\text{APdG}$  adduct with  $\chi = 224.1^\circ$ ,  $\theta = 164.6^\circ$  and  $\phi = 337.5^\circ$ .

| Calculated energy (in Hartrees) = -1634.1616705 |             |             |             |
|-------------------------------------------------|-------------|-------------|-------------|
| Atom                                            | X           | Y           | Z           |
| N                                               | -2.65120200 | 0.23691900  | -0.17530600 |
| C                                               | -1.56371900 | -0.65010800 | -0.20737200 |
| N                                               | -1.93707300 | -1.90458100 | -0.31620600 |
| C                                               | -3.32193800 | -1.86353300 | -0.35220100 |
| C                                               | -4.27900900 | -2.93094600 | -0.43326000 |
| O                                               | -4.13664900 | -4.14145100 | -0.48002600 |
| N                                               | -5.61163800 | -2.38025200 | -0.44154700 |
| H                                               | -6.33695500 | -3.08799800 | -0.42507900 |
| C                                               | -5.94609100 | -1.05560800 | -0.37019000 |
| N                                               | -7.29290200 | -0.74830500 | -0.32653100 |
| H                                               | -7.90180200 | -1.32180200 | -0.89783700 |
| H                                               | -7.45669000 | 0.24338200  | -0.45107400 |
| N                                               | -5.05778100 | -0.09285100 | -0.28946100 |
| C                                               | -3.78261300 | -0.55572500 | -0.27632000 |
| N                                               | -0.28084700 | -0.16876100 | -0.17164700 |
| H                                               | -0.17388200 | 0.83987800  | -0.20711000 |
| C                                               | 0.87815400  | -0.90327900 | 0.10360600  |
| C                                               | 2.12388900  | -0.36664300 | -0.31085900 |
| C                                               | 0.84151500  | -2.11094700 | 0.81806900  |
| C                                               | 3.32877700  | -0.99029900 | 0.13297800  |
| C                                               | 2.23157500  | 0.77427300  | -1.17697300 |
| C                                               | 2.01806100  | -2.73774600 | 1.19812800  |
| H                                               | -0.11815100 | -2.54452100 | 1.06518000  |
| C                                               | 4.59477500  | -0.41762300 | -0.19557100 |
| C                                               | 3.27620100  | -2.18924800 | 0.90281400  |
| C                                               | 3.44072000  | 1.31990100  | -1.49363300 |
| H                                               | 1.33002000  | 1.17552700  | -1.62653900 |
| H                                               | 1.96664300  | -3.66535000 | 1.76242300  |
| C                                               | 4.66582100  | 0.76481500  | -0.99288300 |

|   |             |             |             |
|---|-------------|-------------|-------------|
| C | 5.79960600  | -1.02800000 | 0.27175800  |
| C | 4.50107700  | -2.78761900 | 1.34353000  |
| H | 3.49738900  | 2.16825700  | -2.17161300 |
| C | 5.92010100  | 1.31885400  | -1.29101000 |
| C | 7.03147300  | -0.43523700 | -0.04939400 |
| C | 5.70865200  | -2.23038400 | 1.05028400  |
| H | 4.44383400  | -3.70247500 | 1.92802500  |
| C | 7.08826800  | 0.72435400  | -0.81873100 |
| H | 5.97055400  | 2.21632800  | -1.90231800 |
| H | 7.94746200  | -0.89801700 | 0.30866600  |
| H | 6.62967200  | -2.69095900 | 1.39777700  |
| H | 8.05198100  | 1.16522000  | -1.05802700 |
| C | -2.68480800 | 1.65391100  | 0.09578900  |
| O | -1.76938200 | 2.32572000  | -0.77150100 |
| C | -2.25401200 | 2.06537100  | 1.50701200  |
| C | -1.32022300 | 3.54123800  | -0.14200200 |
| C | -1.88130500 | 3.53402700  | 1.29622400  |
| H | -1.37672700 | 1.49073600  | 1.81423100  |
| H | -3.05307400 | 1.92418000  | 2.24055400  |
| C | 0.19411400  | 3.57895700  | -0.22023800 |
| H | -1.73814100 | 4.40145800  | -0.67757300 |
| O | -3.02734300 | 4.38411300  | 1.28161100  |
| H | -1.13777200 | 3.88414500  | 2.02254000  |
| O | 0.68726000  | 2.53027400  | 0.60528100  |
| H | 0.54945800  | 4.56063000  | 0.12589500  |
| H | 0.50773100  | 3.44527800  | -1.26512600 |
| H | -3.41302500 | 4.37386500  | 2.17140600  |
| H | 1.65727000  | 2.52716700  | 0.57702700  |
| H | -3.71114800 | 1.97411100  | -0.10679800 |

**Table A.56.** Cartesian coordinates of nucleoside minimum for the  $^{1\text{AP}}\text{dG}$  adduct with  $\chi = 142.0^\circ$ ,  $\theta = 186.4^\circ$  and  $\phi = 357.2^\circ$ .

| Calculated energy (in Hartrees) = -1634.1528866 |            |             |             |
|-------------------------------------------------|------------|-------------|-------------|
| Atom                                            | X          | Y           | Z           |
| N                                               | 2.18671300 | 0.09706400  | -0.54759700 |
| C                                               | 1.09459700 | -0.76538100 | -0.40654200 |
| N                                               | 1.43798900 | -2.01922500 | -0.23856500 |
| C                                               | 2.82038100 | -2.00259400 | -0.26984200 |
| C                                               | 3.73181800 | -3.10129300 | -0.12775400 |
| O                                               | 3.53061700 | -4.28090800 | 0.07337400  |

|   |             |             |             |
|---|-------------|-------------|-------------|
| N | 5.08038500  | -2.62854400 | -0.24673700 |
| H | 5.77876800  | -3.34765000 | -0.09714300 |
| C | 5.45647400  | -1.33236000 | -0.47117800 |
| N | 6.81886000  | -1.07535000 | -0.51089200 |
| H | 7.38666100  | -1.78928000 | -0.95314400 |
| H | 7.00474500  | -0.15353000 | -0.88871600 |
| N | 4.61357300  | -0.33947700 | -0.59374200 |
| C | 3.31289100  | -0.71199800 | -0.46440000 |
| N | -0.17856700 | -0.24268500 | -0.45949200 |
| H | -0.23483000 | 0.73094400  | -0.69998900 |
| C | -1.39772300 | -0.85790800 | -0.16089300 |
| C | -2.58858800 | -0.09617000 | -0.31805000 |
| C | -1.45994500 | -2.18076600 | 0.30131900  |
| C | -3.83583200 | -0.69074600 | 0.04601900  |
| C | -2.61077900 | 1.24774300  | -0.82513200 |
| C | -2.68114000 | -2.74568100 | 0.63752700  |
| H | -0.54262000 | -2.74763400 | 0.38343600  |
| C | -5.04943100 | 0.05310100  | -0.07244500 |
| C | -3.88229400 | -2.03130400 | 0.53276900  |
| C | -3.76855100 | 1.95621500  | -0.94123300 |
| H | -1.69602600 | 1.72680300  | -1.16540700 |
| H | -2.70663900 | -3.77116200 | 0.99682600  |
| C | -5.02957500 | 1.39301400  | -0.56279200 |
| C | -6.29585200 | -0.54091300 | 0.29912100  |
| C | -5.14647400 | -2.60140000 | 0.89374400  |
| H | -3.75168800 | 2.96813600  | -1.33767900 |
| C | -6.23266900 | 2.10847400  | -0.67002000 |
| C | -7.47323900 | 0.21337700  | 0.17615700  |
| C | -6.30238800 | -1.89119100 | 0.78534900  |
| H | -5.15993600 | -3.62463700 | 1.26068800  |
| C | -7.44026600 | 1.52136900  | -0.30243100 |
| H | -6.21000800 | 3.12816300  | -1.04570300 |
| H | -8.41958600 | -0.23980900 | 0.45977800  |
| H | -7.25428200 | -2.33514300 | 1.06435000  |
| H | -8.36398700 | 2.08641600  | -0.39029000 |
| C | 2.05548300  | 1.56508100  | -0.63919200 |
| O | 3.28729600  | 2.17567300  | -0.88107500 |
| C | 1.48076800  | 2.17645700  | 0.66536500  |
| C | 3.67603300  | 3.01152200  | 0.22316200  |
| C | 2.36085800  | 3.41050700  | 0.89845700  |
| H | 1.59621400  | 1.46452500  | 1.48274300  |



|   |            |            |             |
|---|------------|------------|-------------|
| H | 0.42471100 | 2.45028600 | 0.57636000  |
| C | 4.66798600 | 2.31555200 | 1.14313100  |
| H | 4.15629600 | 3.89170200 | -0.21378400 |
| O | 1.87763000 | 4.56035600 | 0.20875900  |
| H | 2.49809000 | 3.61534200 | 1.96757600  |
| O | 3.97652900 | 1.35254000 | 1.93013500  |
| H | 5.13940900 | 3.08075900 | 1.78115800  |
| H | 5.43428100 | 1.84033100 | 0.52100400  |
| H | 1.02611100 | 4.80311700 | 0.60405200  |
| H | 4.61678900 | 0.91678300 | 2.51065700  |
| H | 1.41302700 | 1.76552600 | -1.50842300 |

**Table A.57.** Cartesian coordinates of nucleoside minimum for the  $^{1\text{AP}}\text{dG}$  adduct with  $\chi = 70.2^\circ$ ,  $\theta = 62.0^\circ$  and  $\phi = 13.0^\circ$ .

| Calculated energy (in Hartrees) = -1634.1492840 |             |             |             |
|-------------------------------------------------|-------------|-------------|-------------|
| Atom                                            | X           | Y           | Z           |
| N                                               | -1.98114000 | -0.51606200 | -0.37525800 |
| C                                               | -1.30317100 | -1.72742200 | -0.20092400 |
| N                                               | -2.09799700 | -2.72673700 | 0.08996800  |
| C                                               | -3.36504000 | -2.16712700 | 0.13072700  |
| C                                               | -4.63190900 | -2.77448800 | 0.43717900  |
| O                                               | -4.92655100 | -3.92715000 | 0.70508800  |
| N                                               | -5.66458100 | -1.77212400 | 0.41188400  |
| H                                               | -6.57102200 | -2.11795000 | 0.70537000  |
| C                                               | -5.49425800 | -0.44296700 | 0.13186800  |
| N                                               | -6.60879100 | 0.36962000  | 0.21678800  |
| H                                               | -7.47739300 | -0.01708300 | -0.13276700 |
| H                                               | -6.41209200 | 1.29721500  | -0.14044500 |
| N                                               | -4.33398800 | 0.09128100  | -0.15671100 |
| C                                               | -3.31273700 | -0.79993400 | -0.12671700 |
| N                                               | 0.05879800  | -1.84917300 | -0.43762800 |
| H                                               | 0.29765300  | -2.82294900 | -0.59252600 |
| C                                               | 1.03426200  | -1.11809400 | 0.29425300  |
| C                                               | 2.35744500  | -1.08652700 | -0.20238200 |
| C                                               | 0.70563300  | -0.43545200 | 1.46907200  |
| C                                               | 3.35222500  | -0.37873900 | 0.53598300  |
| C                                               | 2.73836000  | -1.73151500 | -1.42653100 |
| C                                               | 1.67069100  | 0.28091100  | 2.16516300  |
| H                                               | -0.31390800 | -0.46056200 | 1.83483900  |
| C                                               | 4.70032300  | -0.35368100 | 0.07105100  |

|   |             |             |             |
|---|-------------|-------------|-------------|
| C | 3.00345000  | 0.31636100  | 1.73291600  |
| C | 4.02694600  | -1.70805100 | -1.86733300 |
| H | 1.97233900  | -2.22093600 | -2.01844200 |
| H | 1.39296200  | 0.80603800  | 3.07501700  |
| C | 5.05661200  | -1.03052600 | -1.13409200 |
| C | 5.69892300  | 0.35416500  | 0.80766100  |
| C | 4.02656900  | 1.02255000  | 2.44876200  |
| H | 4.29518400  | -2.19704800 | -2.80032800 |
| C | 6.39038300  | -0.99573600 | -1.57066700 |
| C | 7.01890500  | 0.36052100  | 0.33192000  |
| C | 5.31489800  | 1.03870300  | 2.01037400  |
| H | 3.74711300  | 1.54653200  | 3.35929800  |
| C | 7.35823700  | -0.30839800 | -0.84263500 |
| H | 6.65963900  | -1.51203200 | -2.48843200 |
| H | 7.77891600  | 0.89850300  | 0.89261800  |
| H | 6.08052200  | 1.57503600  | 2.56486000  |
| H | 8.38603900  | -0.29153400 | -1.19409300 |
| C | -1.39611500 | 0.76206600  | -0.70679900 |
| O | -1.44078000 | 1.60384100  | 0.44089300  |
| C | -2.09982700 | 1.56932600  | -1.79850300 |
| C | -1.39988800 | 2.97314000  | 0.01537800  |
| C | -1.53691200 | 2.96556800  | -1.52254400 |
| H | -3.17699600 | 1.55610800  | -1.62761500 |
| H | -1.87191500 | 1.20510100  | -2.80521400 |
| C | -2.47663300 | 3.76440300  | 0.73791000  |
| H | -0.42088600 | 3.39764100  | 0.27175500  |
| O | -0.21777100 | 3.13080500  | -2.04571700 |
| H | -2.20479200 | 3.75972400  | -1.87544200 |
| O | -3.74832300 | 3.48382200  | 0.16332100  |
| H | -2.23166500 | 4.83531500  | 0.64927900  |
| H | -2.44676800 | 3.49277700  | 1.80288300  |
| H | -0.27986600 | 3.09773600  | -3.01292100 |
| H | -4.40716900 | 4.00308400  | 0.64667700  |
| H | -0.35884300 | 0.56677300  | -0.99288400 |

**Table A.58.** Cartesian coordinates of nucleoside minimum for the <sup>1</sup>APdG adduct with  $\chi = 52.5^\circ$ ,  $\theta = 300.2^\circ$  and  $\phi = 342.6^\circ$ .

| Calculated energy (in Hartrees) = -1634.1490684 |            |             |             |
|-------------------------------------------------|------------|-------------|-------------|
| Atom                                            | X          | Y           | Z           |
| N                                               | 1.88711900 | -0.46267400 | -0.25264100 |

|   |             |             |             |
|---|-------------|-------------|-------------|
| C | 1.27820000  | -1.68404400 | -0.54615100 |
| N | 2.10342500  | -2.69996600 | -0.50793400 |
| C | 3.33199200  | -2.13741700 | -0.19186400 |
| C | 4.62080000  | -2.75349500 | -0.02692800 |
| O | 4.97965300  | -3.91287200 | -0.14809200 |
| N | 5.59161900  | -1.75067400 | 0.32971900  |
| H | 6.53970700  | -2.10572300 | 0.37895700  |
| C | 5.34799500  | -0.41400900 | 0.50090500  |
| N | 6.42747100  | 0.39956800  | 0.79281900  |
| H | 7.10962100  | 0.02653300  | 1.44254600  |
| H | 6.11243600  | 1.33166000  | 1.03694000  |
| N | 4.16632000  | 0.12870400  | 0.35374300  |
| C | 3.21421100  | -0.76162300 | -0.01170500 |
| N | -0.06955200 | -1.74826300 | -0.87296200 |
| H | -0.28666700 | -2.61399500 | -1.35482300 |
| C | -1.06598700 | -1.29976200 | 0.03587600  |
| C | -2.35764600 | -1.01180400 | -0.45929300 |
| C | -0.77753700 | -1.12128500 | 1.39362900  |
| C | -3.36082700 | -0.56624200 | 0.45285800  |
| C | -2.69455200 | -1.13769000 | -1.84848700 |
| C | -1.74788200 | -0.65615200 | 2.27274200  |
| H | 0.21355700  | -1.36655100 | 1.76016000  |
| C | -4.67889100 | -0.29326100 | -0.01844400 |
| C | -3.05077100 | -0.37968100 | 1.83398700  |
| C | -3.95505700 | -0.87950300 | -2.29479100 |
| H | -1.91502000 | -1.41869600 | -2.54826200 |
| H | -1.50464500 | -0.53319900 | 3.32496500  |
| C | -4.99514700 | -0.46049500 | -1.40018700 |
| C | -5.68650800 | 0.15293800  | 0.89032800  |
| C | -4.08142200 | 0.07523400  | 2.72216800  |
| H | -4.19101500 | -0.97308600 | -3.35158900 |
| C | -6.30062000 | -0.19171200 | -1.84063100 |
| C | -6.97723200 | 0.40908400  | 0.40345700  |
| C | -5.34160200 | 0.32668100  | 2.27356700  |
| H | -3.83256100 | 0.20970100  | 3.77184500  |
| C | -7.27847100 | 0.23614200  | -0.94609000 |
| H | -6.53976500 | -0.31940100 | -2.89306400 |
| H | -7.74435700 | 0.74772700  | 1.09490800  |
| H | -6.11405400 | 0.66608600  | 2.95867100  |
| H | -8.28386500 | 0.43941900  | -1.30365700 |
| C | 1.31002300  | 0.85762700  | -0.39732300 |

|   |             |            |             |
|---|-------------|------------|-------------|
| O | 2.04765500  | 1.57721800 | -1.37708200 |
| C | 1.35609400  | 1.73910100 | 0.85138800  |
| C | 1.95593300  | 2.98326400 | -1.10047100 |
| C | 1.18552300  | 3.12496900 | 0.22780200  |
| H | 2.33940200  | 1.65445400 | 1.31916500  |
| H | 0.57020300  | 1.48857400 | 1.56952900  |
| C | 3.33952100  | 3.61776100 | -1.10294300 |
| H | 1.36292400  | 3.45403900 | -1.89444700 |
| O | -0.17491200 | 3.38859000 | -0.12109900 |
| H | 1.59626100  | 3.92031800 | 0.85984100  |
| O | 3.98477000  | 3.42359700 | 0.15355700  |
| H | 3.21905500  | 4.69308800 | -1.31050300 |
| H | 3.91600000  | 3.17002100 | -1.92423000 |
| H | -0.68444500 | 3.44301600 | 0.70228300  |
| H | 4.85473000  | 3.84473100 | 0.09507600  |
| H | 0.27951700  | 0.71439500 | -0.73218400 |

## **Appendix B**

**Supplementary Information for Chapter 3: Structure of an Unusual Tetracyclic Deoxyguanosine Adduct: Implications for Frameshift Mutagenicity by *Ortho*-Cyano Nitroanilines.**

Contains Figures B.1–B.41 and Tables B.1–B.20

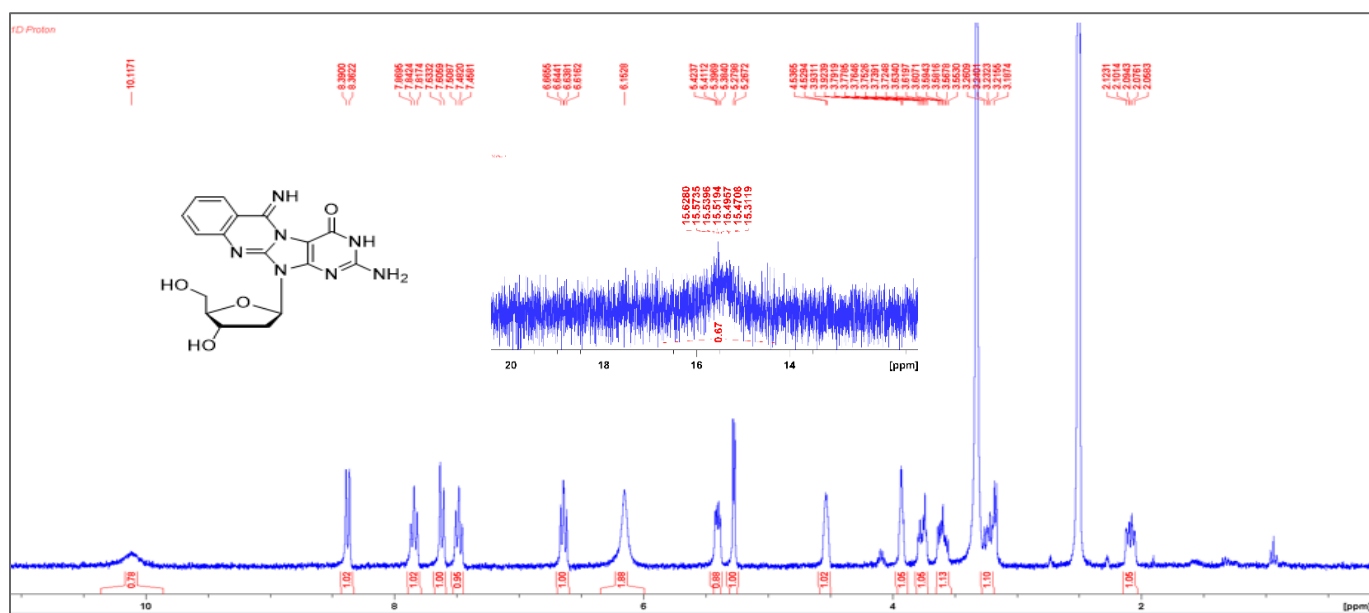


Figure B.1. <sup>1</sup>H NMR spectrum of compound 1

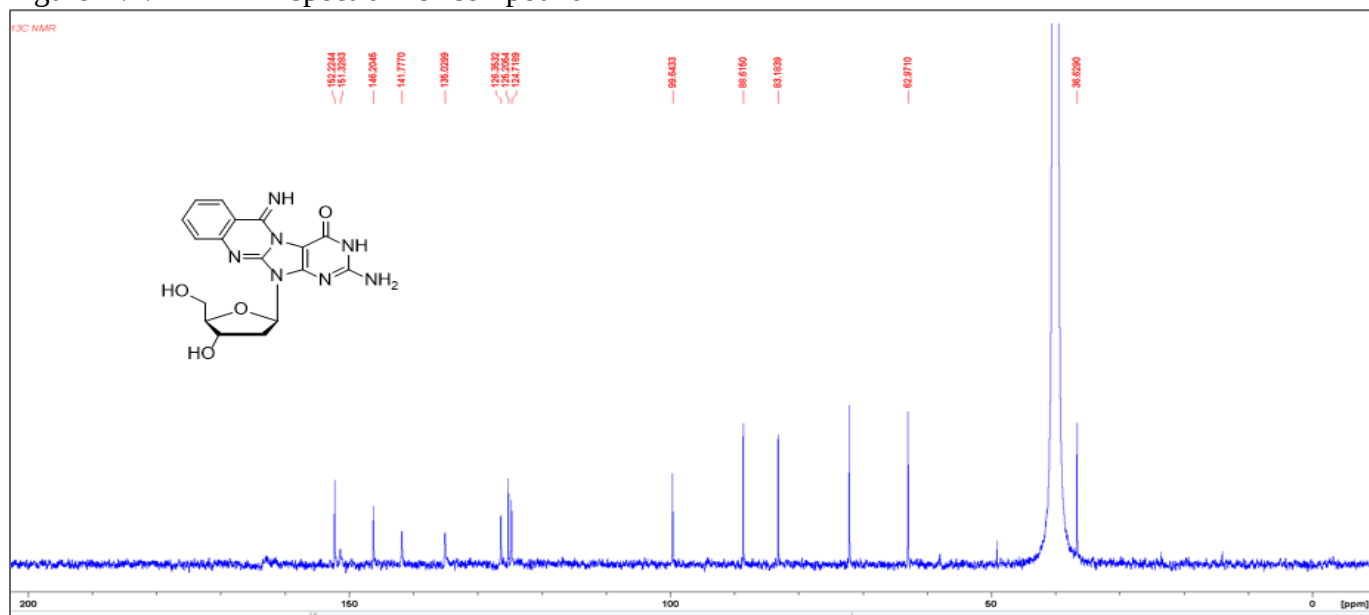


Figure B.2. <sup>13</sup>C NMR spectrum of compound 1

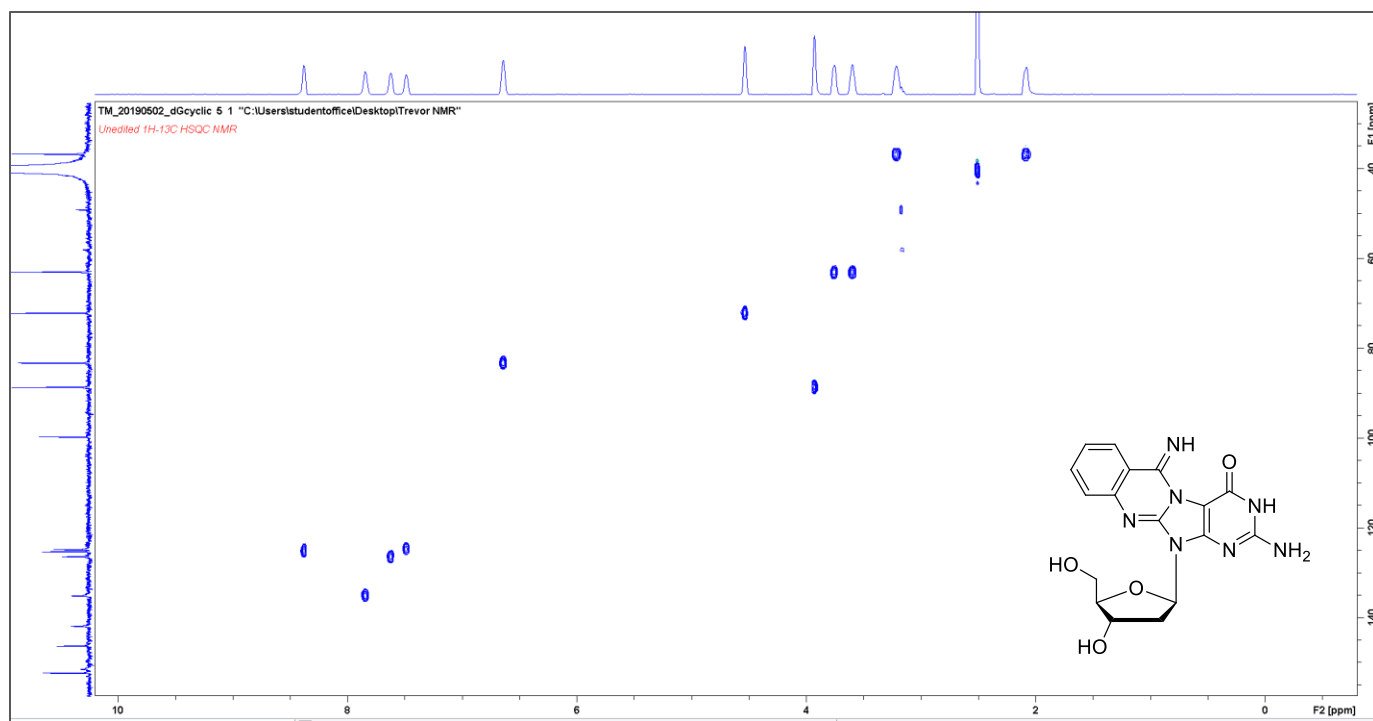


Figure B.3.  $^{13}\text{C}$  - $^1\text{H}$  HSQC NMR spectrum of compound **1**

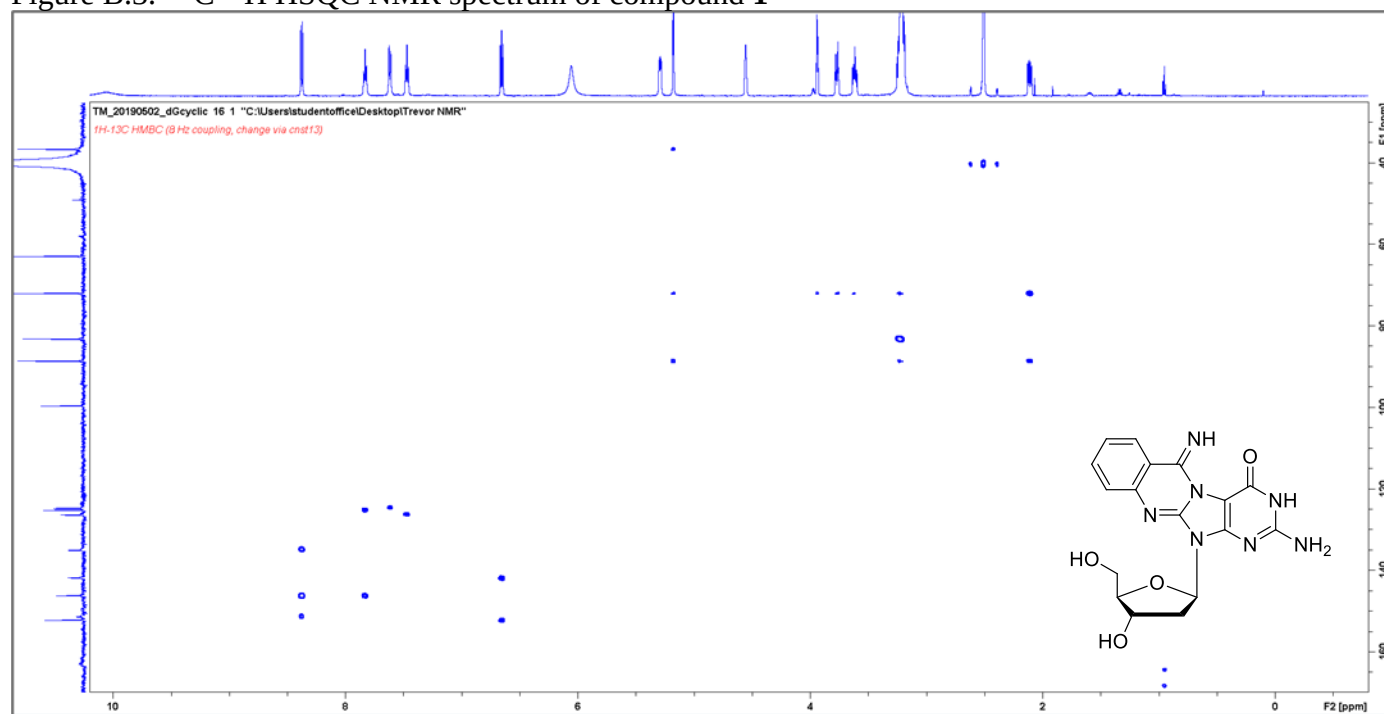
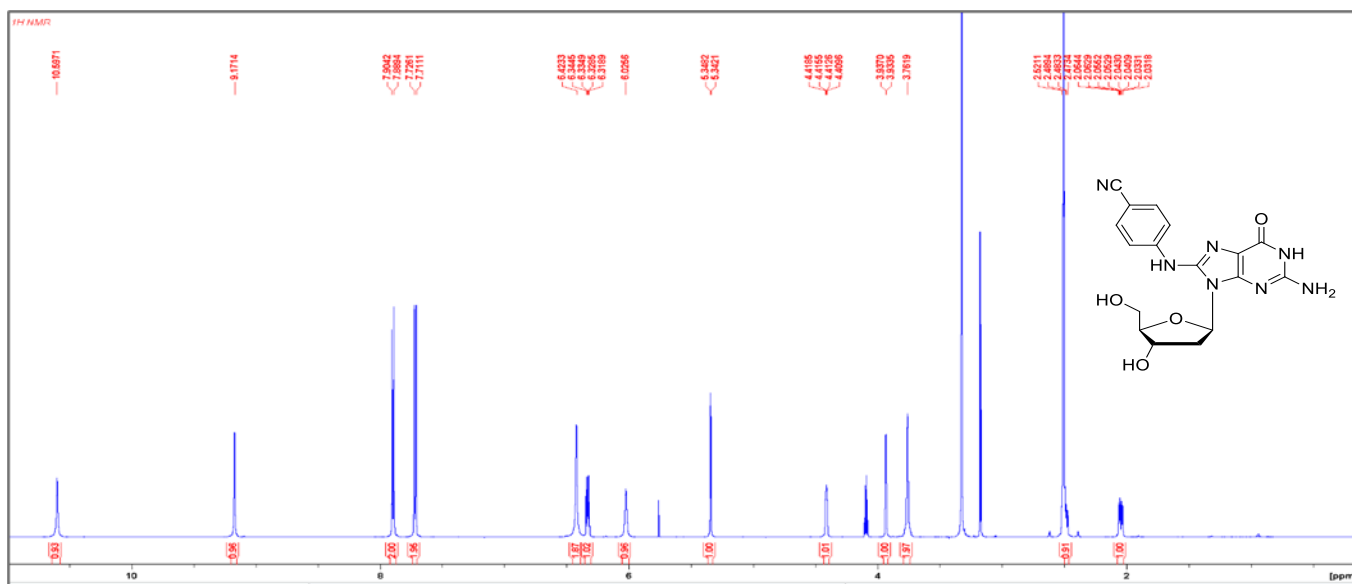
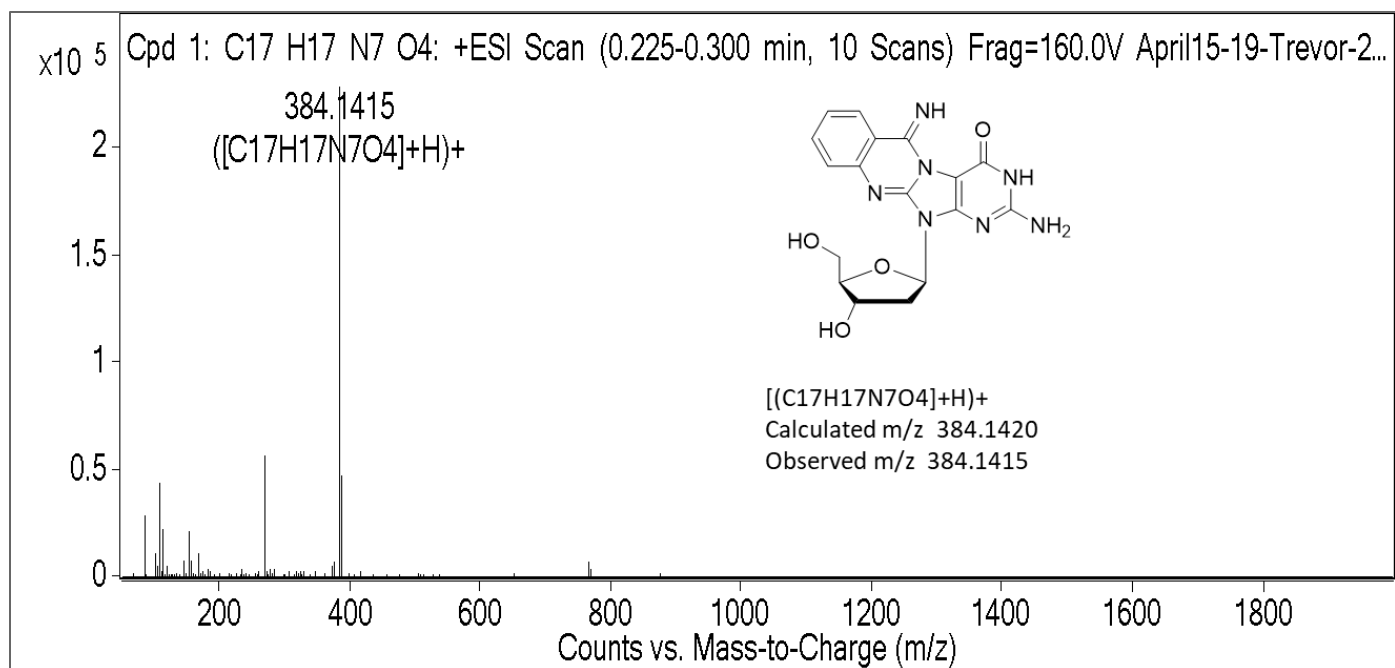


Figure B.4.  $^{13}\text{C}$  - $^1\text{H}$  HMBC NMR spectrum of compound **1**





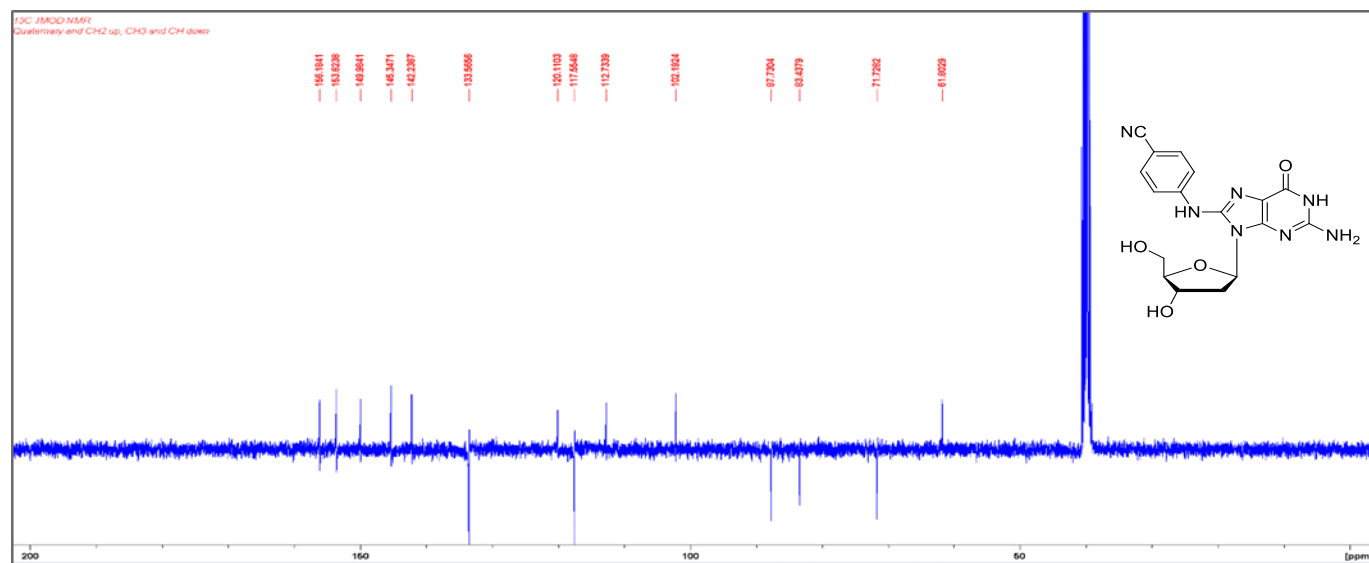


Figure B.7. <sup>13</sup>C J-mod spectrum of compound 2

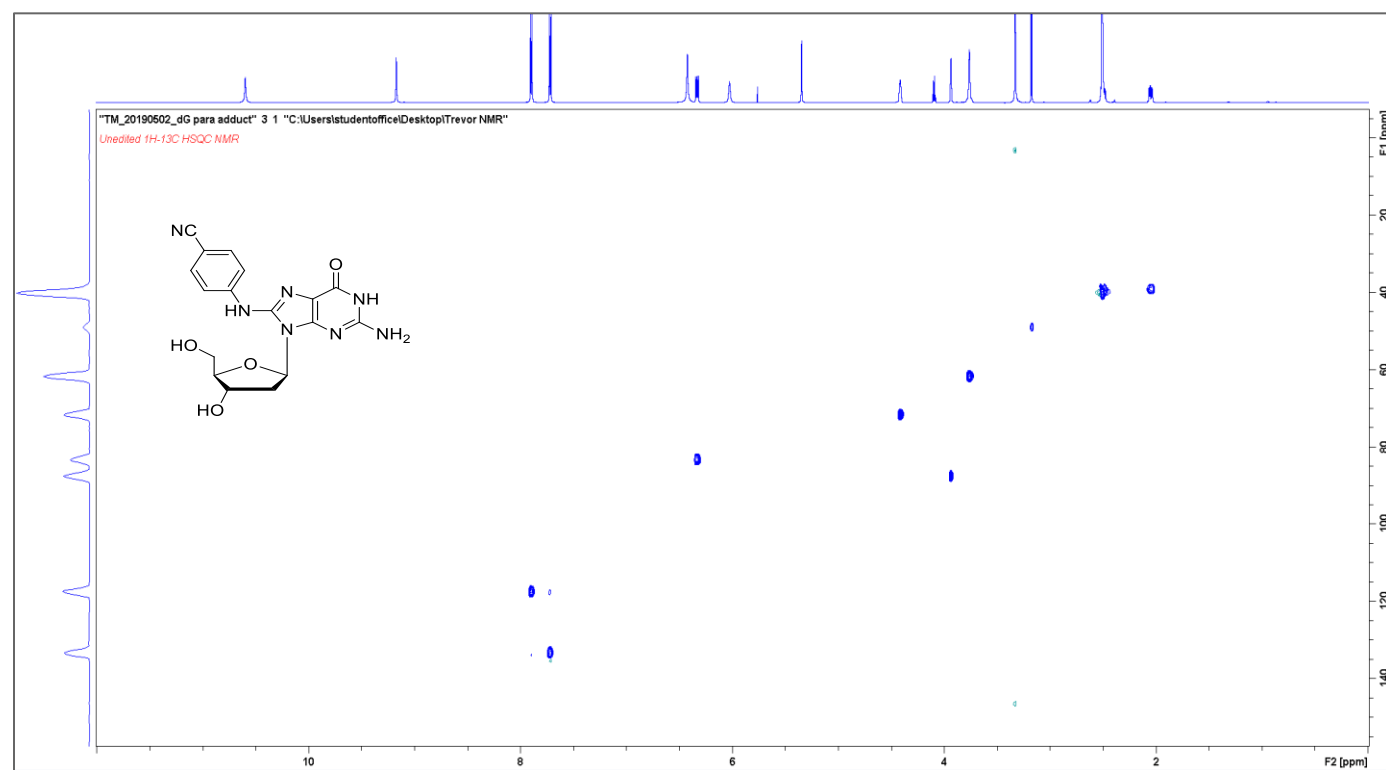


Figure B.8. <sup>13</sup>C - <sup>1</sup>H HSQC NMR spectrum of compound 2

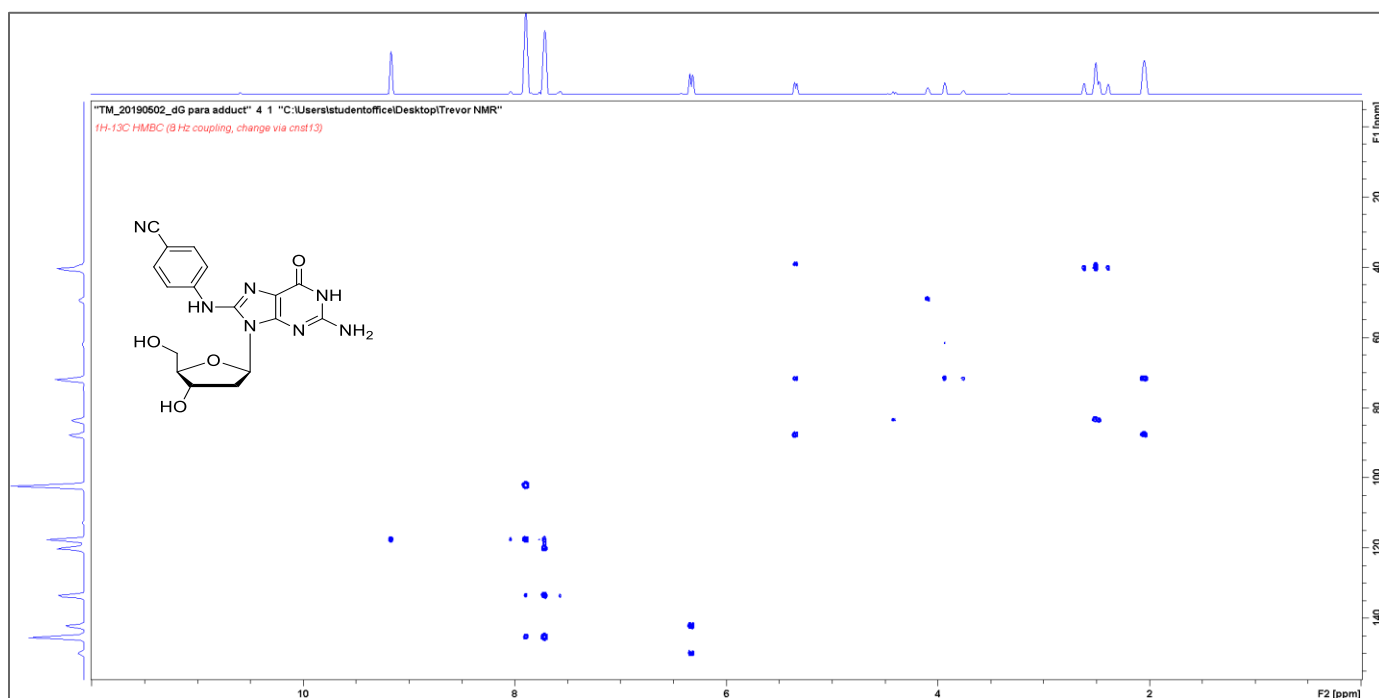


Figure B.9.  $^{13}\text{C}$  -  $^1\text{H}$  HMBC NMR spectrum of compound 2

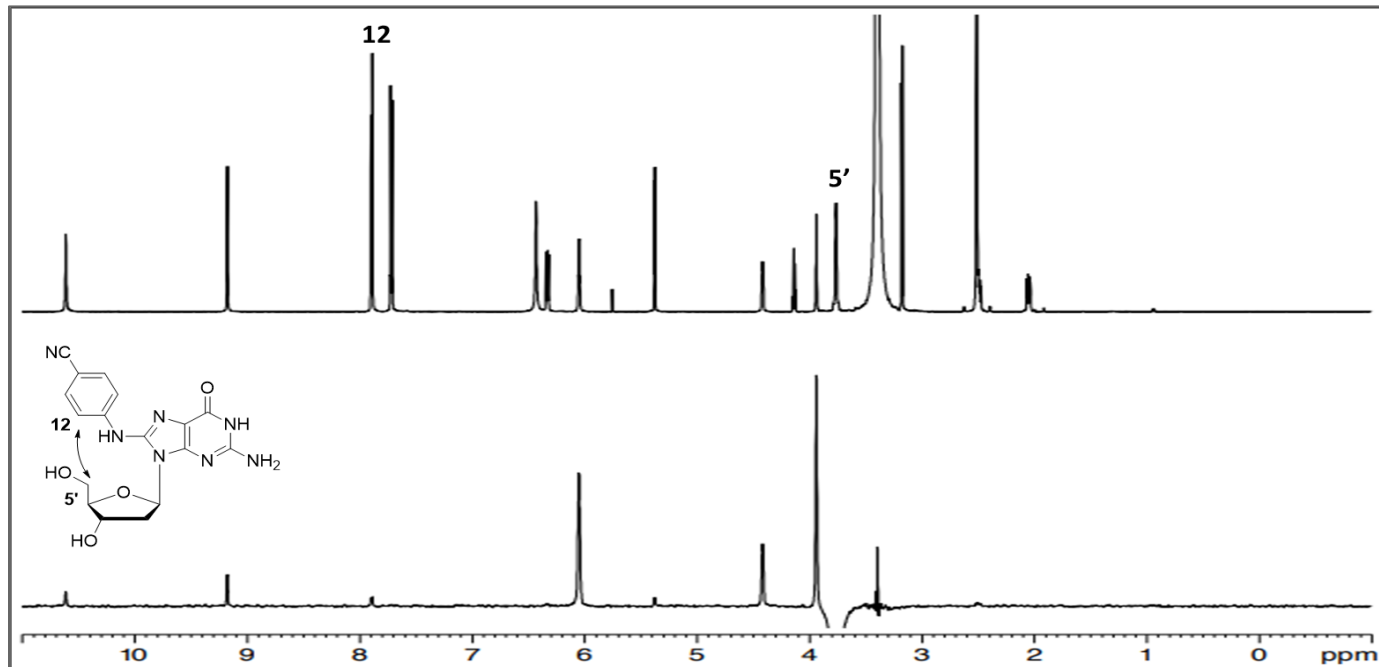


Figure B.10. ROESY spectrum of compound 2

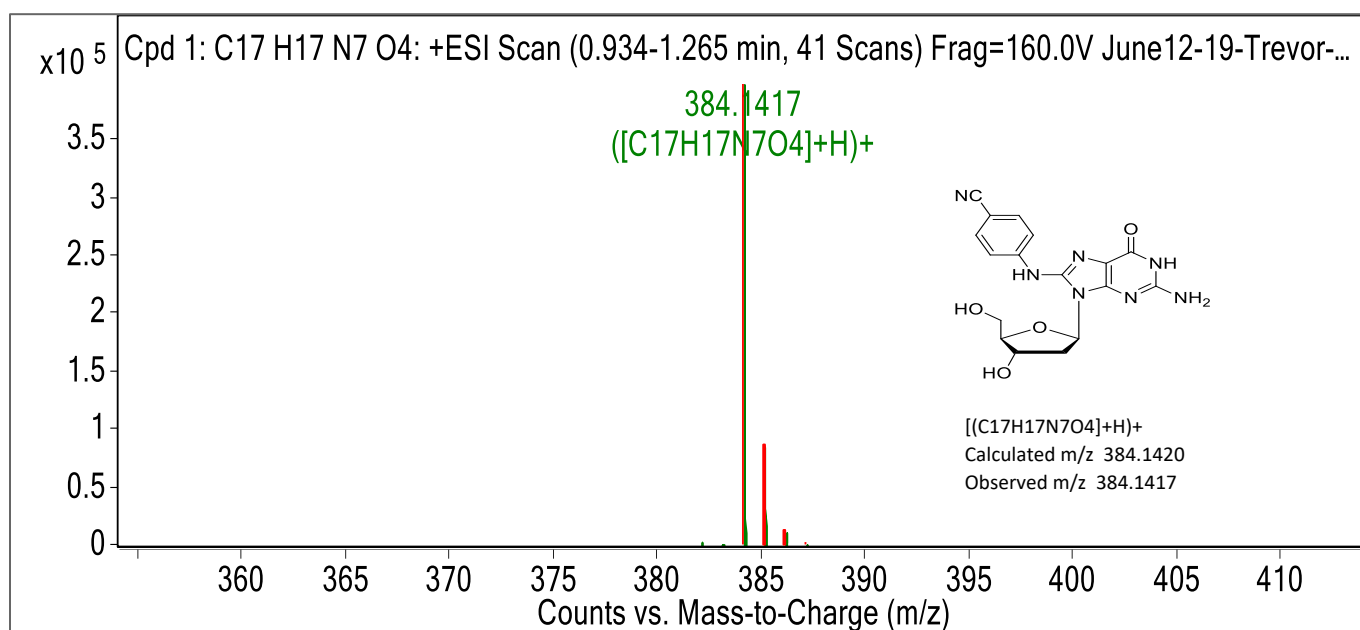


Figure B.11. HRMS spectrum of compound 2

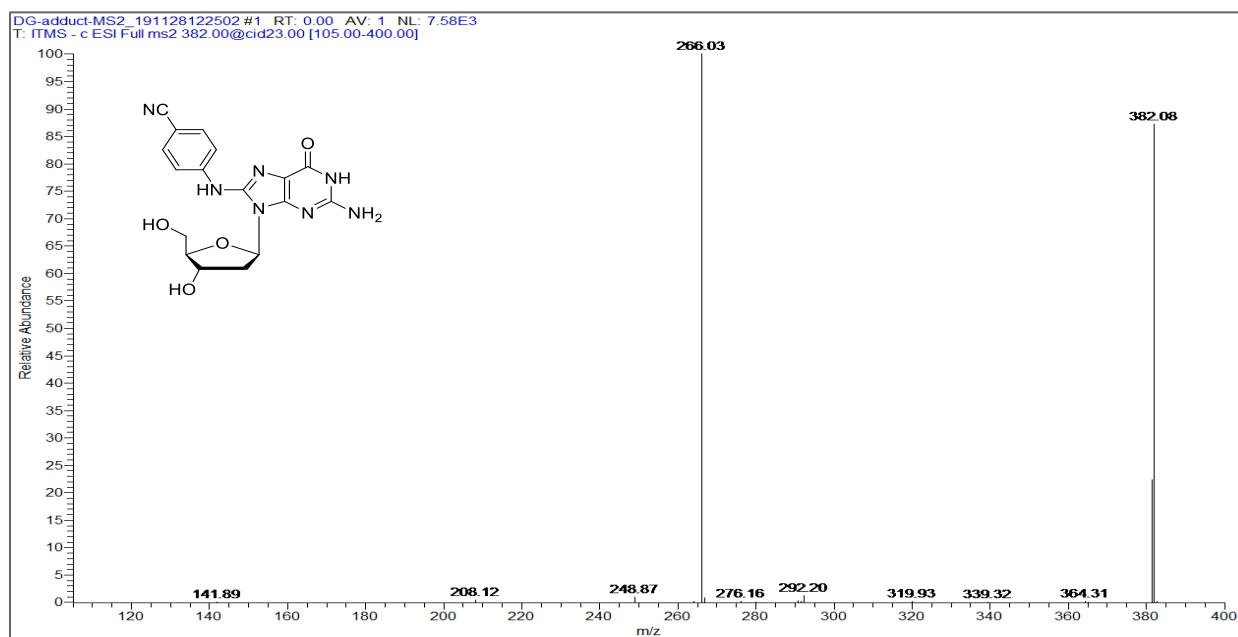


Figure B.12. Mass spectrum of compound 2 in DMSO stock solution

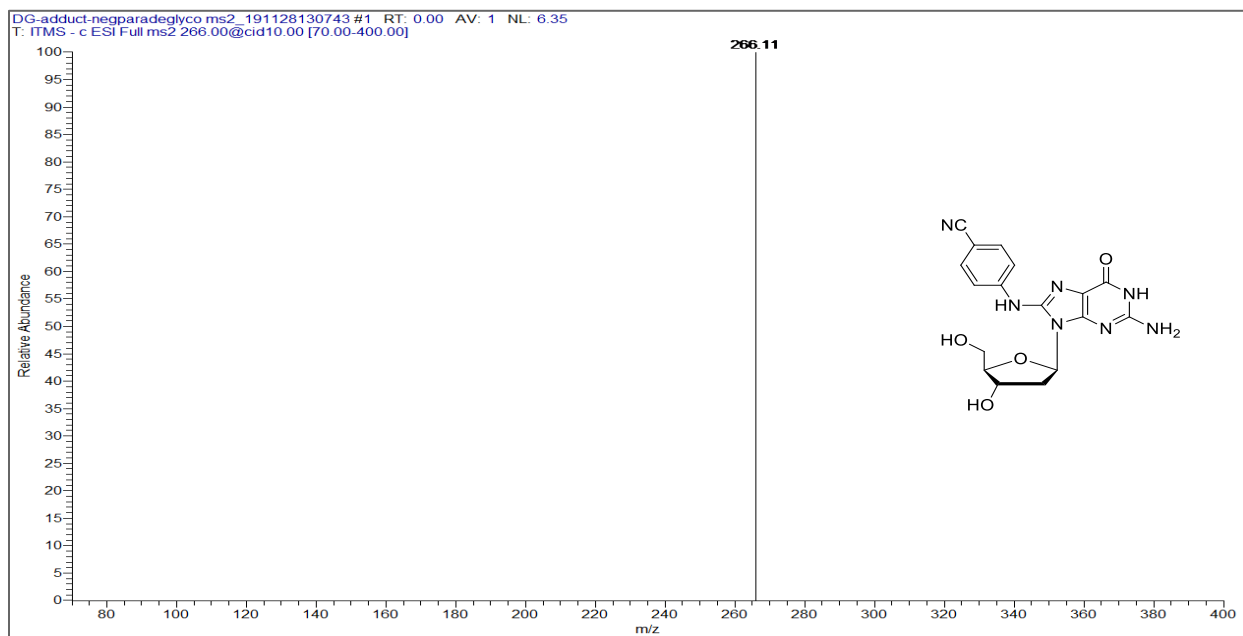


Figure B.13. Mass spectrum of compound 2 following exposure to 0.1M HCl at 37.2°C

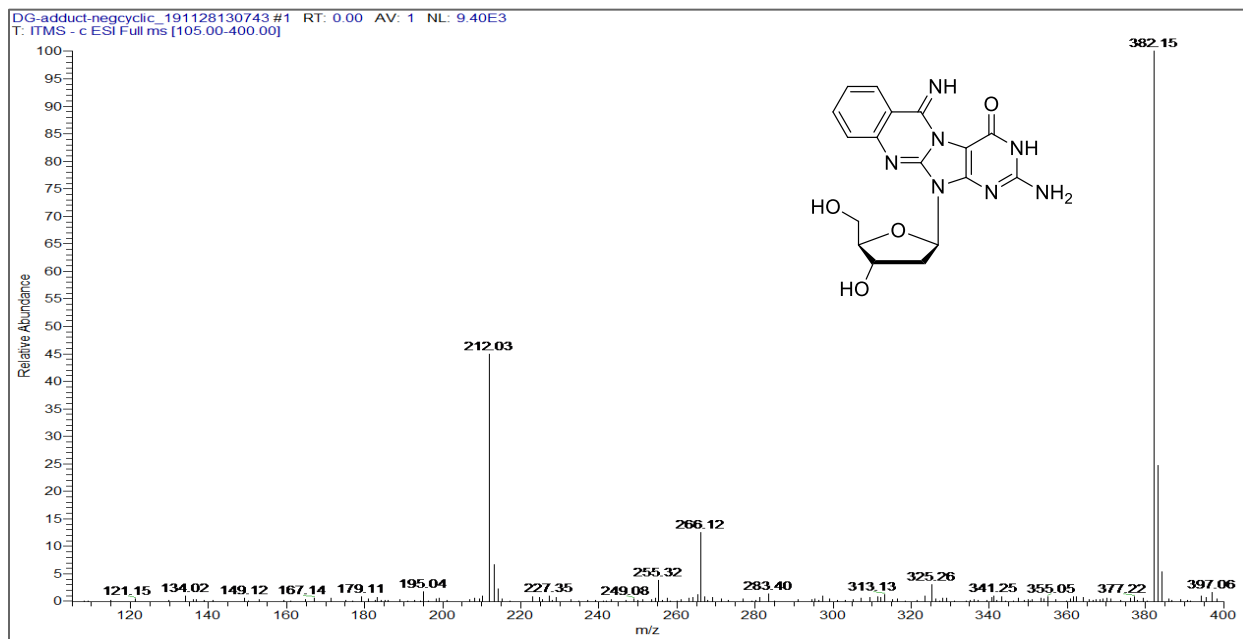


Figure B.14. Mass spectrum of compound 1 in DMSO stock solution

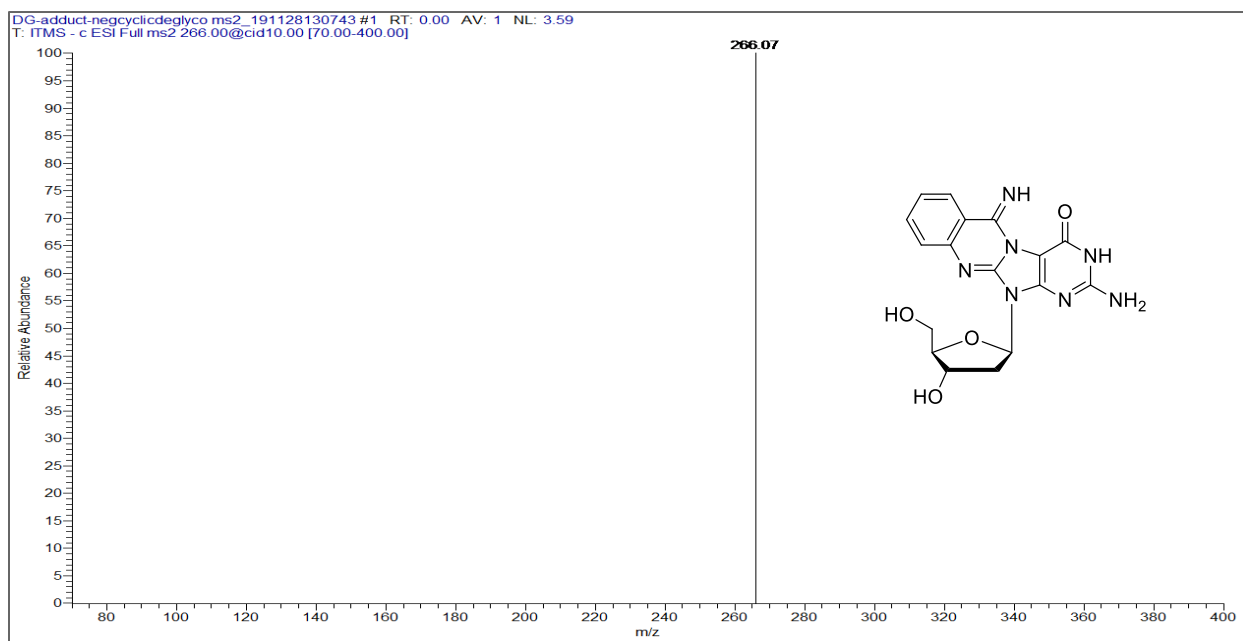


Figure B.15. Mass spectrum of compound **1** following exposure to 0.1M HCl at 37.2°C

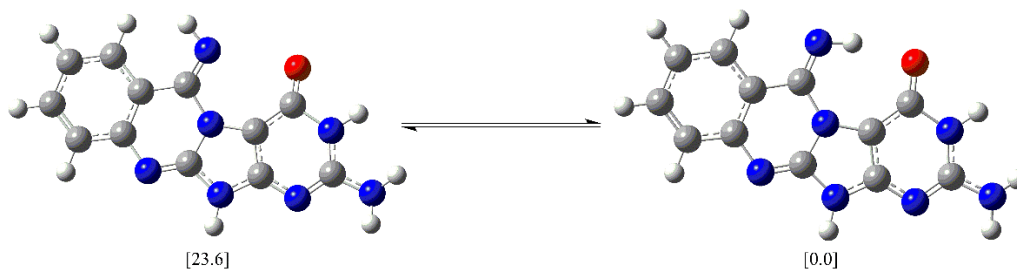


Figure B.16: Structures and relative stabilities of the **1** nucleobase minima as calculated at the B3LYP-D3(BJ)/6-31G(d,p) level of theory. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311++G(2df,2p) relative energies are provided in parentheses (kJ/mol).

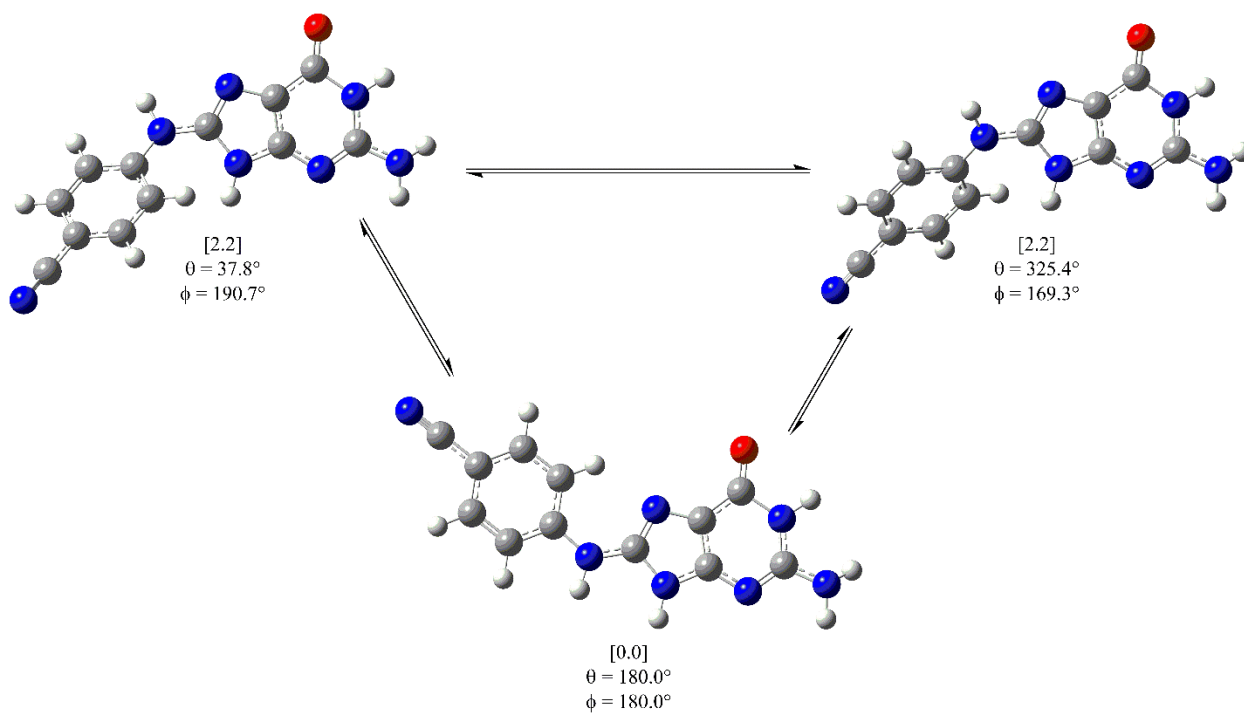


Figure B.17: Select dihedrals ( $\theta$  and  $\phi$ , deg) for the 2 nucleobase minima as calculated at the B3LYP-D3(BJ)/6-31G(d,p) level of theory. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311++G(2df,2p) relative energies are provided in parentheses (kJ/mol).

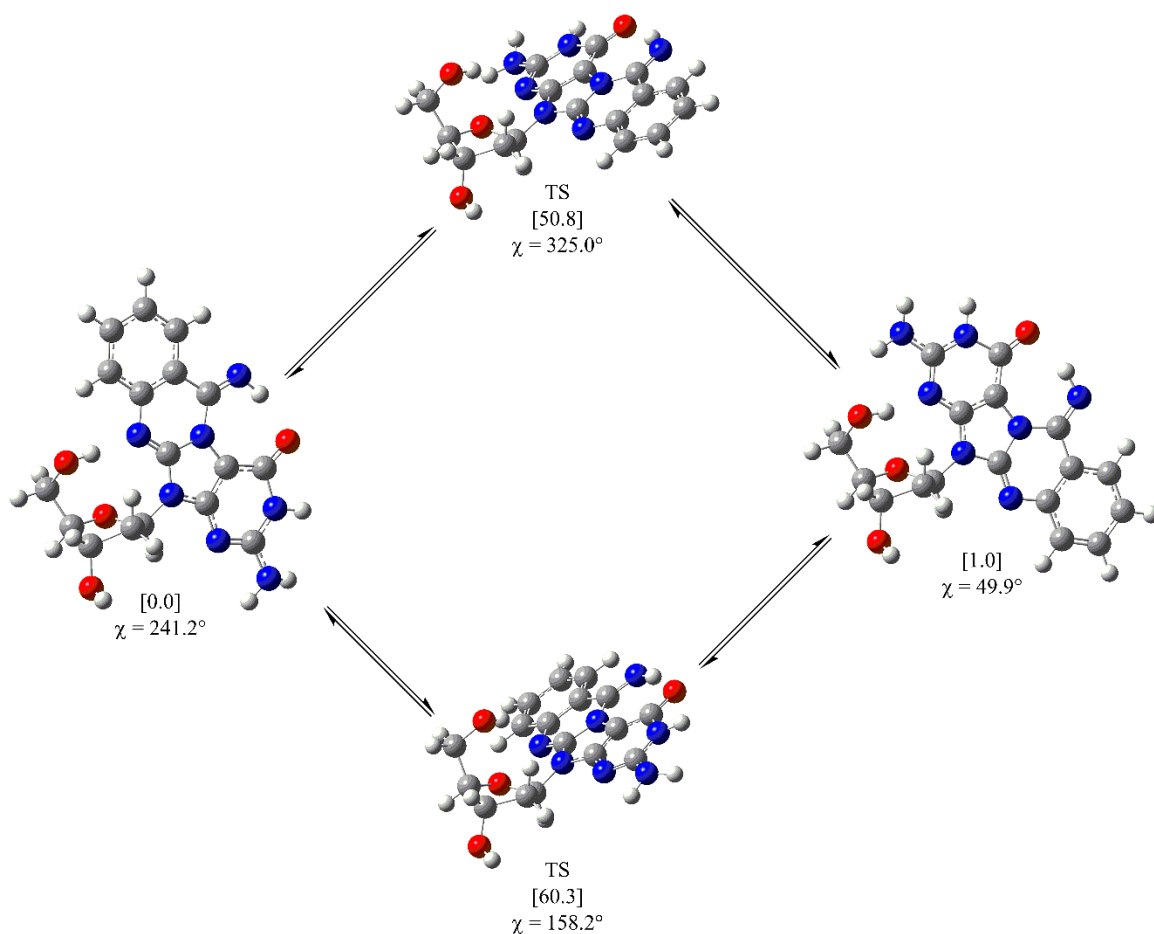


Figure B.18: Select dihedral angles ( $\chi$ , deg) for the **1** nucleoside minima and transition states (TS) as calculated at the B3LYP-D3(BJ)/6-31G(d,p) level of theory. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311++G(2df,2p) relative energies are provided in parentheses (kJ/mol).

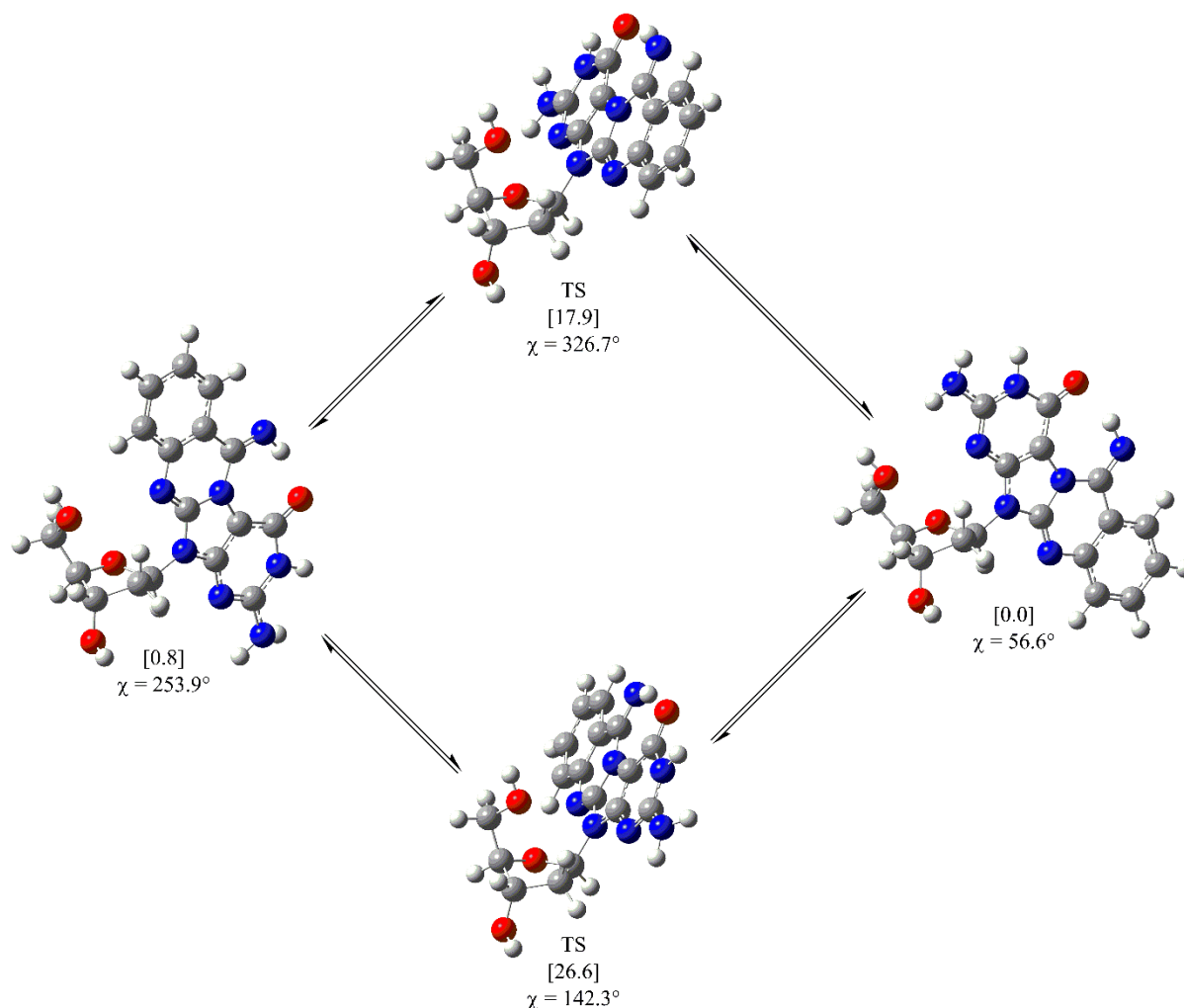


Figure B.19: Select dihedral angles ( $\chi$ , deg) for the **1** nucleoside minima and transition states (TS) as calculated at the B3LYP-D3(BJ)/6-31G(d,p) level of theory with the 5'-hydroxyl group fixed away from the adduct. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311++G(2df,2p) relative energies are provided in parentheses (kJ/mol).



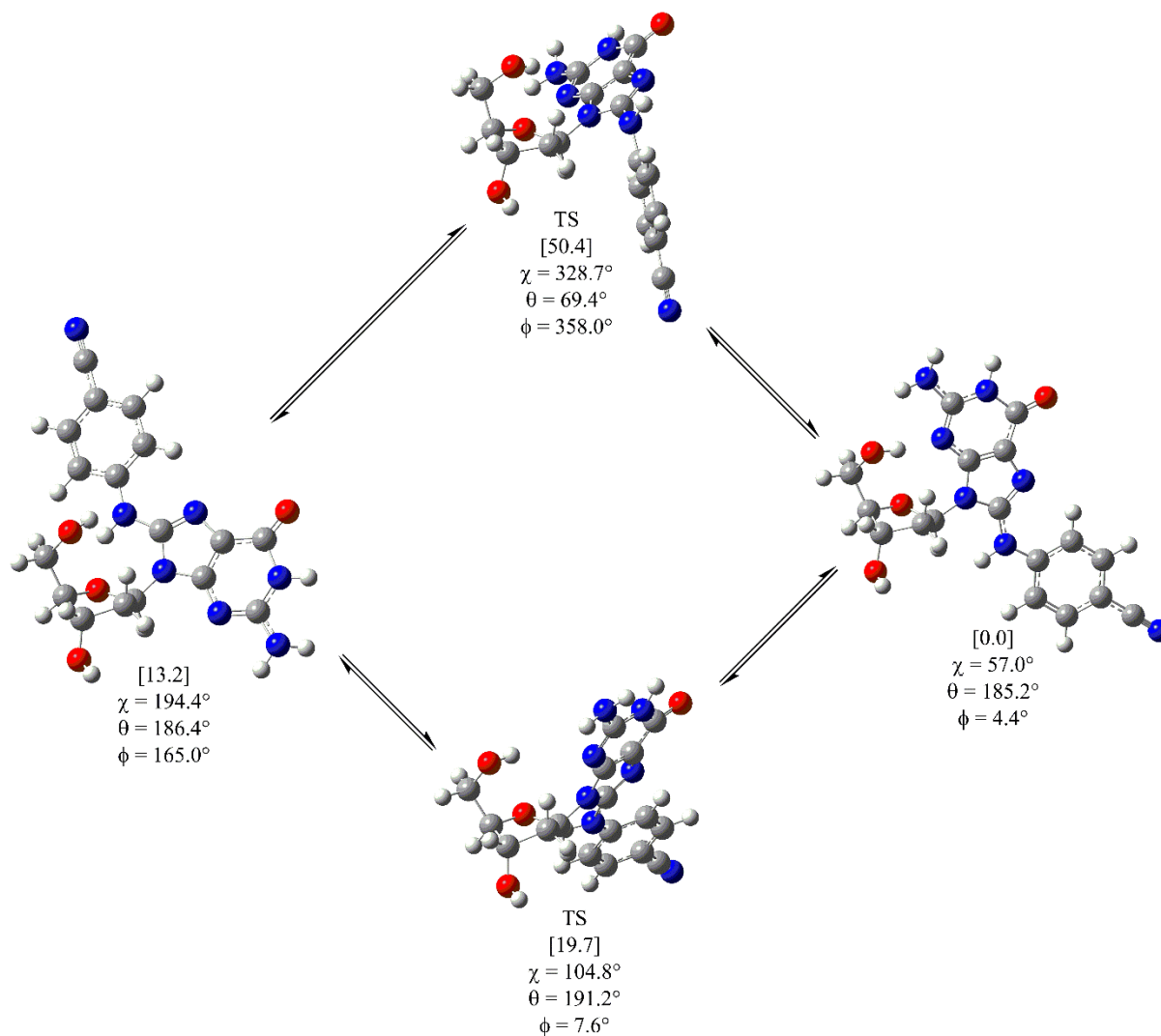


Figure B.20: Select dihedral angles ( $\chi$ ,  $\theta$  and  $\phi$ , deg) for the 2 nucleoside minima and transition states (TS) as calculated at the B3LYP-D3(BJ)/6-31G(d,p) level of theory. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311++G(2df,2p) relative energies are provided in parentheses (kJ/mol).

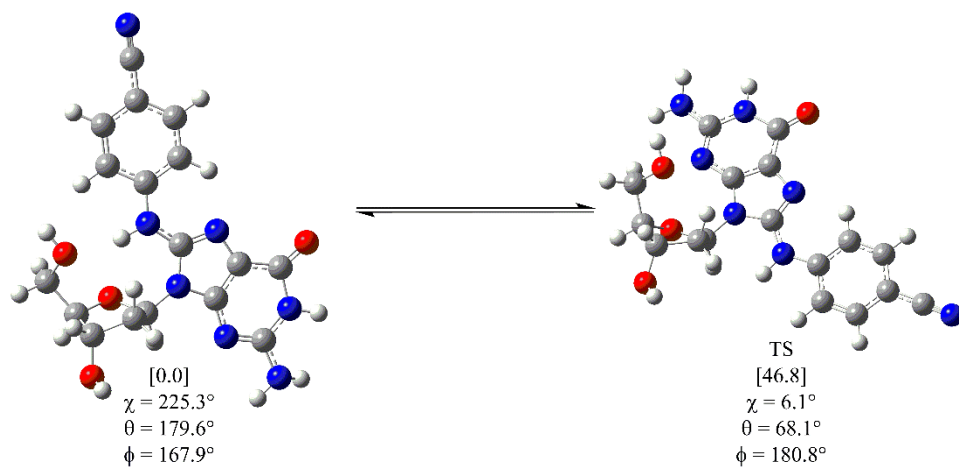


Figure B.21: Select dihedral angles ( $\chi$ ,  $\theta$  and  $\phi$ , deg) for the 2 nucleoside minima and transition states (TS) as calculated at the B3LYP-D3(BJ)/6-31G(d,p) level of theory with the 5'-hydroxyl group fixed away from the adduct. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311++G(2df,2p) relative energies are provided in parentheses (kJ/mol).

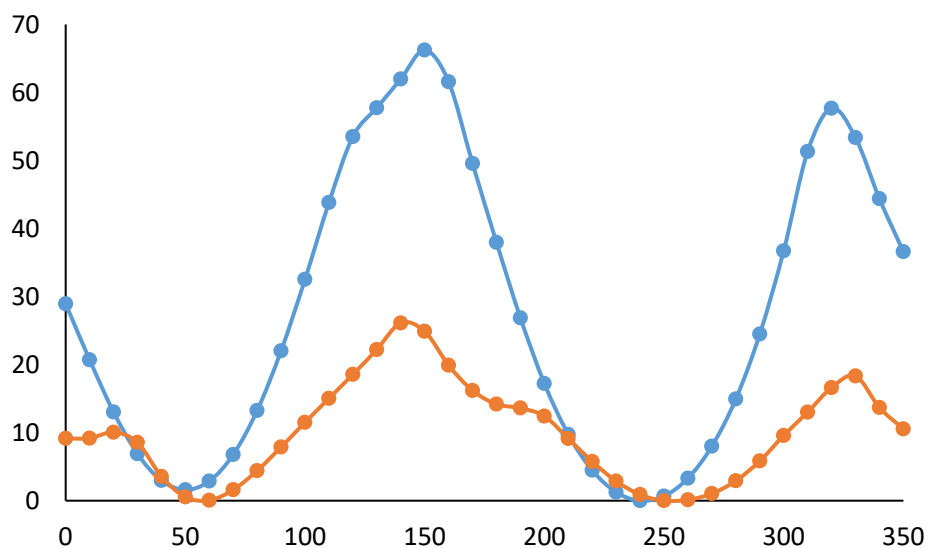


Figure B.22: Potential energy surface for the rotation about the  $\chi$  dihedral angle (deg) in the **1** nucleoside with the 5'-hydroxyl group constrained (orange) or unconstrained (blue). Calculated every  $10^\circ$  at the B3LYP-D3(BJ)/6-31G(d,p) level of theory and report in kJ/mol.

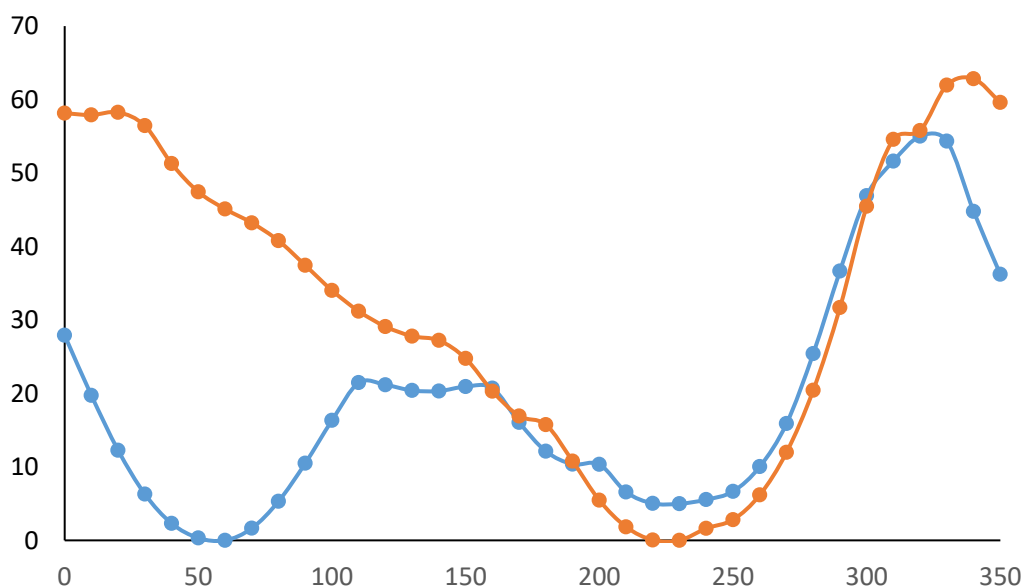


Figure B.23: Potential energy surface for the rotation about the  $\chi$  dihedral angle (deg) in the **2** nucleoside with the 5'-hydroxyl group constrained (orange) or unconstrained (blue). Calculated every  $10^\circ$  at the B3LYP-D3(BJ)/6-31G(d,p) level of theory and report in kJ/mol.

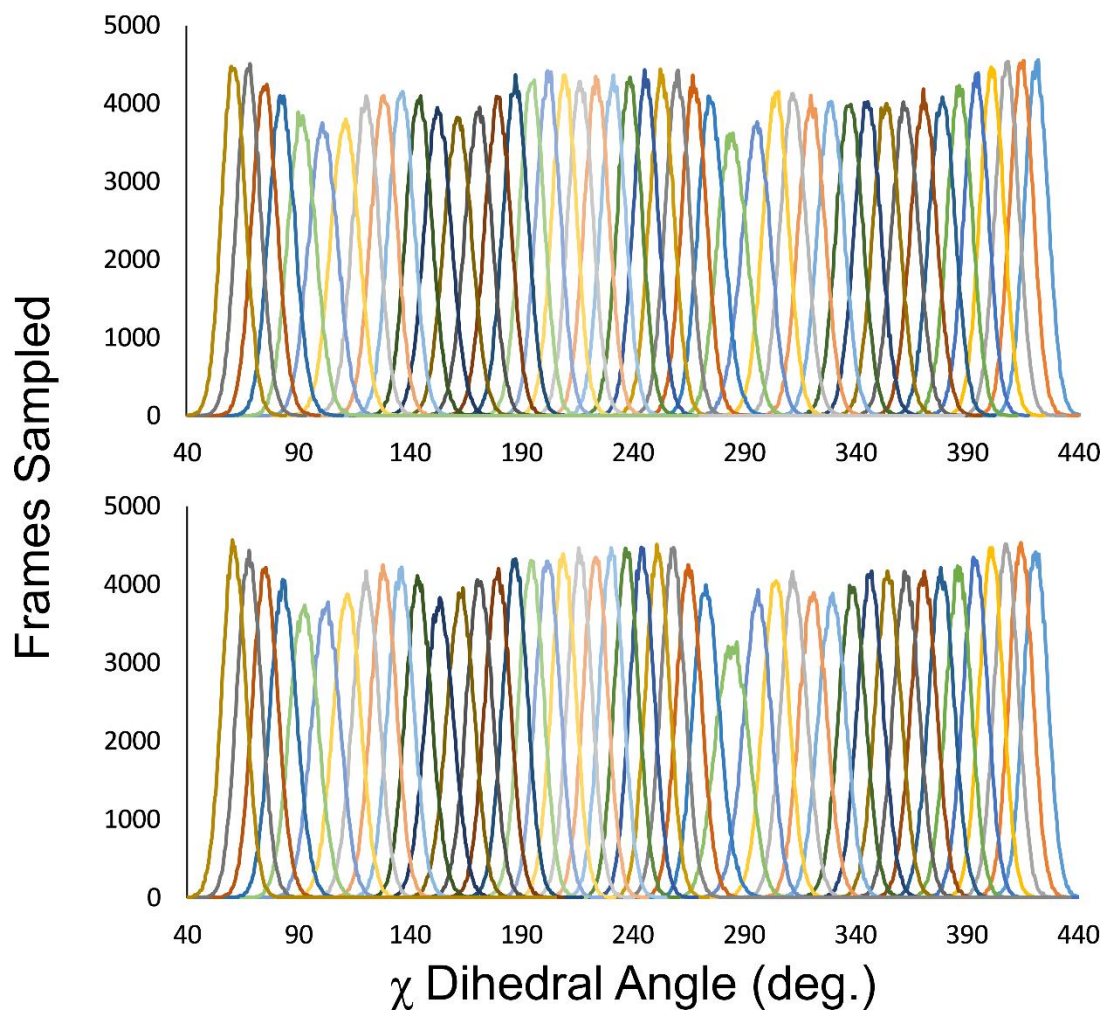


Figure B.24: Histograms of the  $\chi$  dihedral angle sampled during umbrella sampling for the **1** (top) and **2** (bottom) free nucleosides with a 30 kcal/(mol rad<sup>2</sup>) restraint.

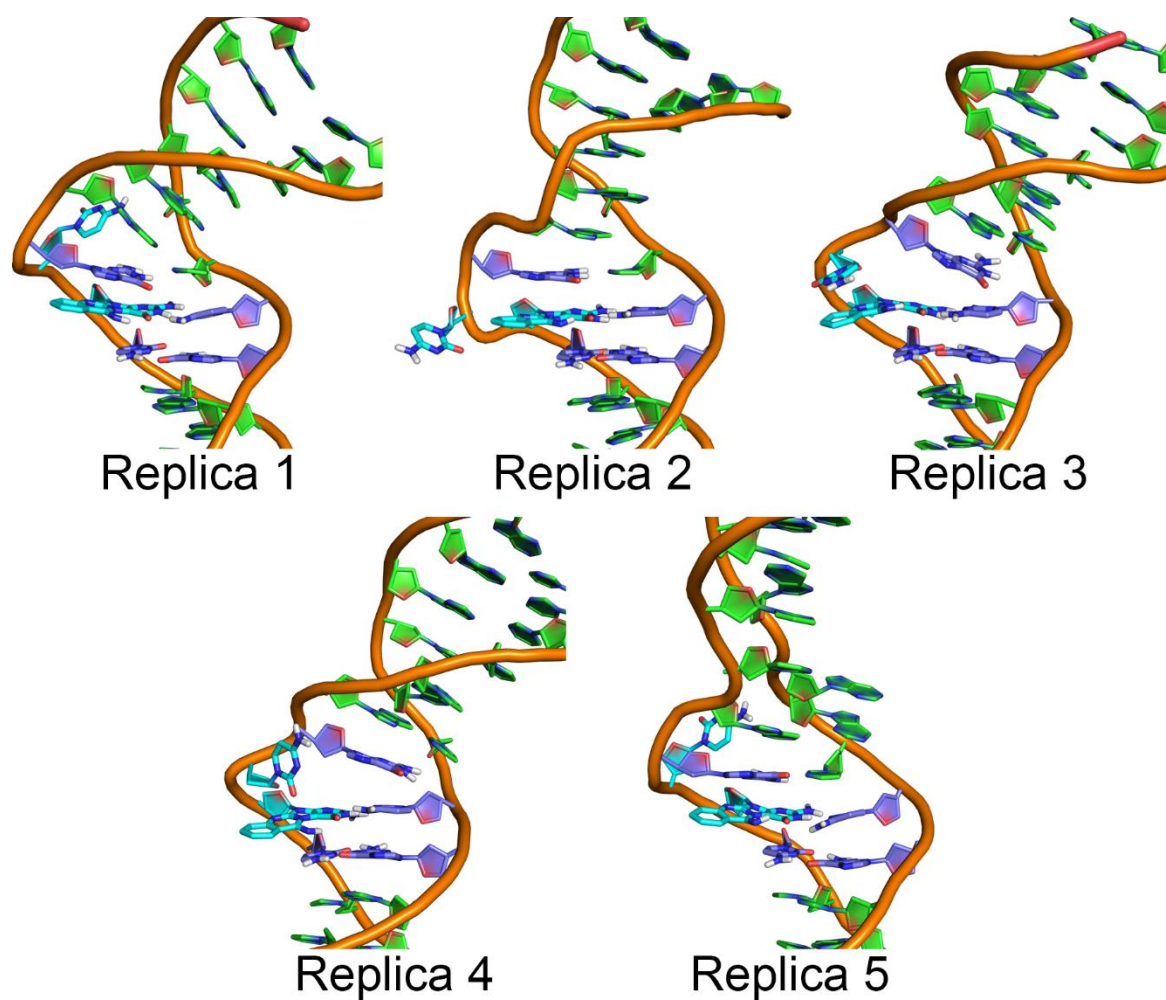


Figure B.25: Representative structures obtained from MD simulations for the five 1  $\mu$ s replicas of the distorted **1** adducted SMI duplex in the *anti* orientation. The adducted base and unpaired C are shown in cyan, and the flanking bases are shown in purple.

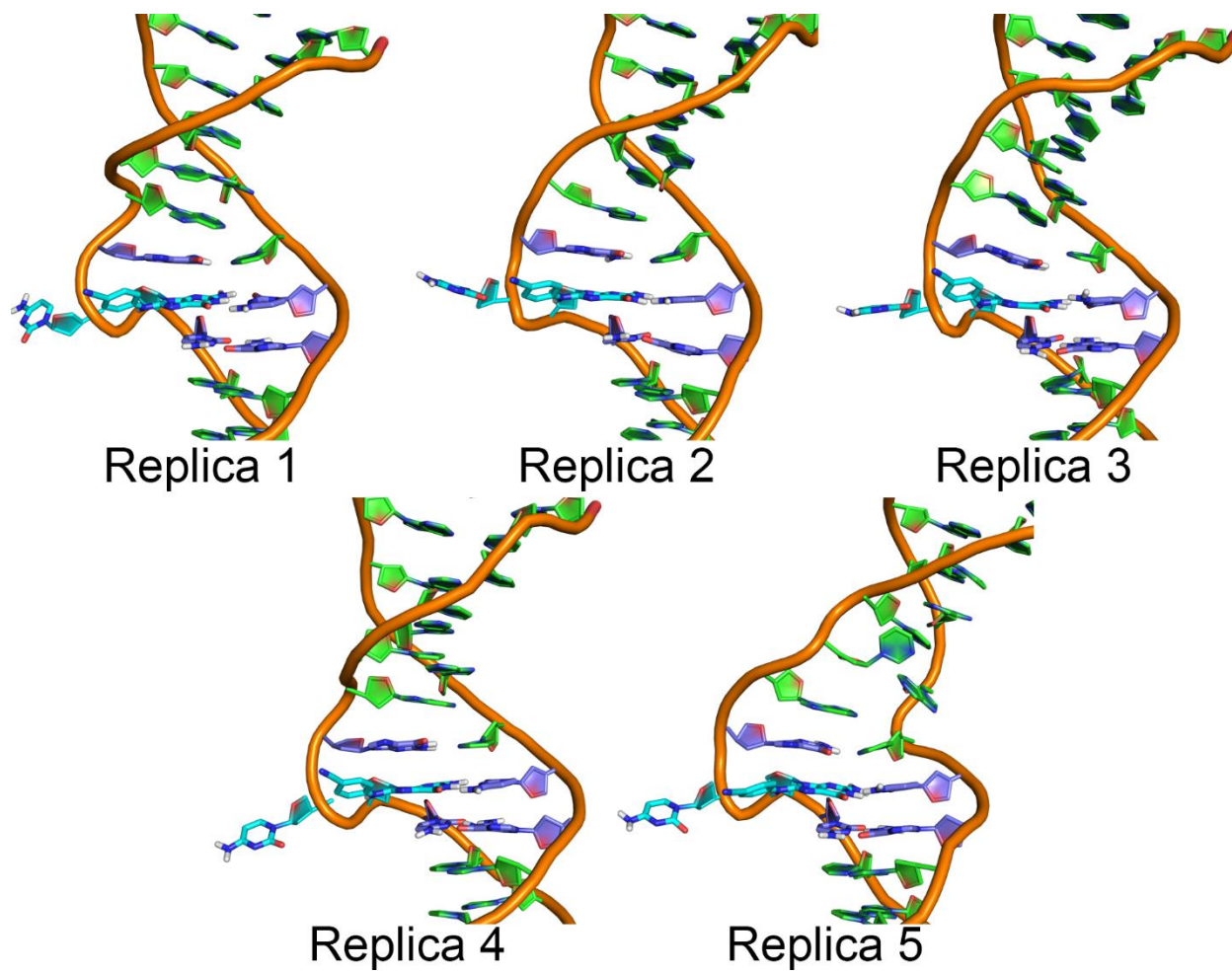


Figure B.26: Representative structures obtained from MD simulations for the five 1  $\mu$ s replicas of the distorted 2 adducted SMI duplex in the *anti* orientation. The adducted base and unpaired C are shown in cyan, and the flanking bases are shown in purple.



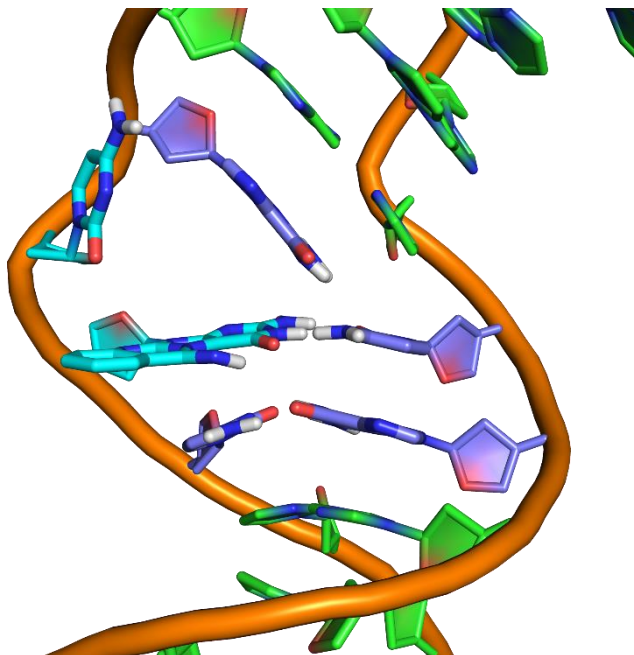


Figure B.27: Example structure of a SMI helix with the adducted G3 and G2 bases hydrogen bonding to a single C. The adducted base and unpaired C are shown in cyan, and the flanking bases are shown in purple.

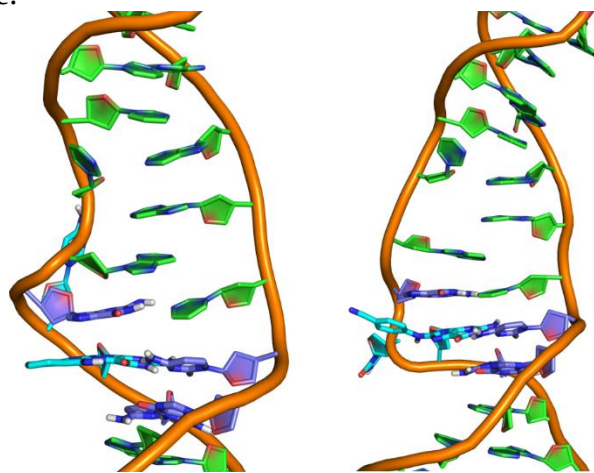


Figure B.28: Example structures of the distorted SMI helices highlighting observed helical distortions 5' to the lesion site with the **1** (left) and **2** (right) adducts in the *anti* orientation. The adducted base and unpaired C are shown in cyan, and the flanking bases are shown in purple.

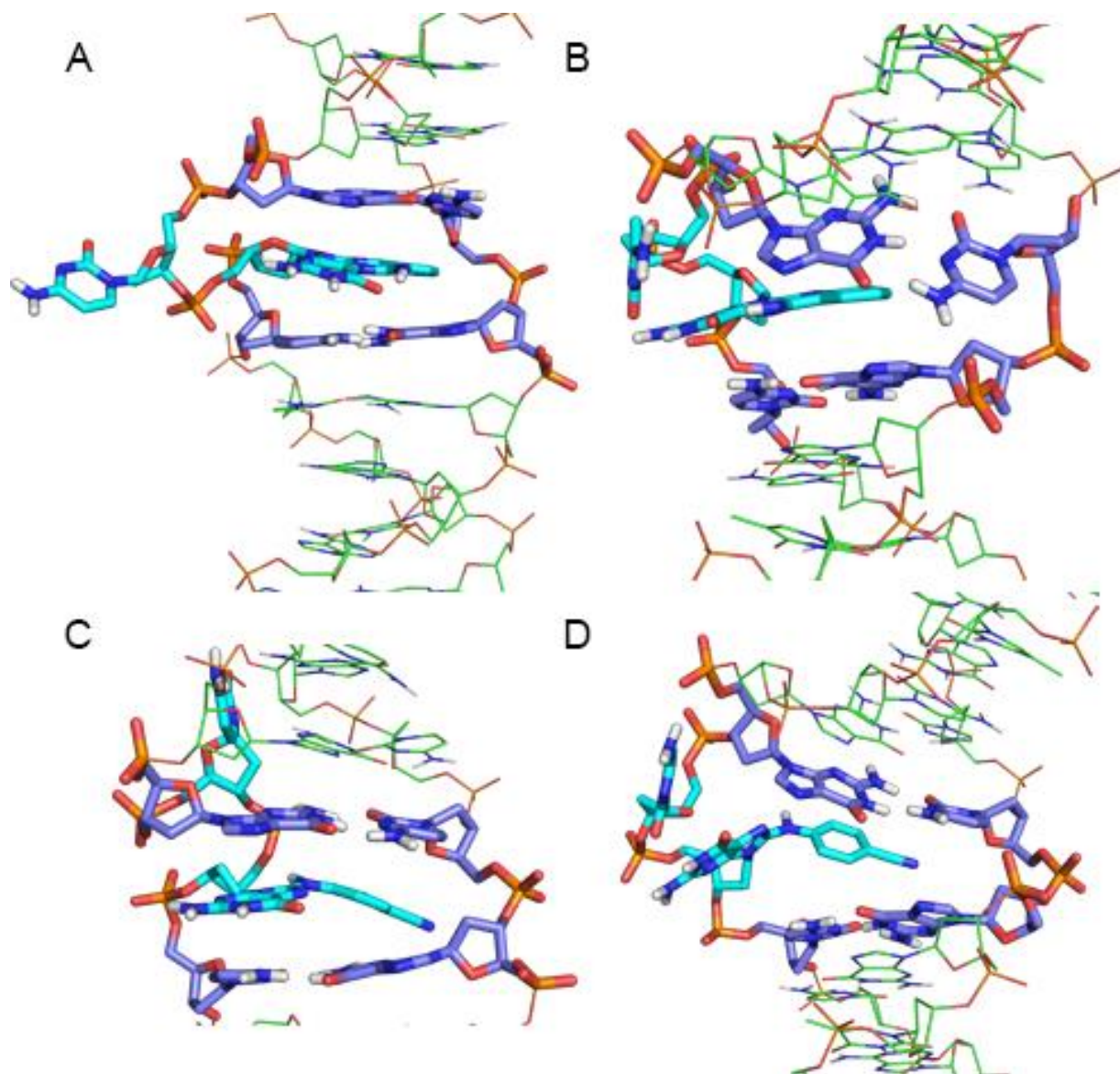


Figure B.29: SMI complex representative structures, highlighting the ability of both lesions to adopt planar and bent structures, for the (A) planar 1 adduct, (B) distorted bent 1 adduct, (C) planar 2 adduct and (D) distorted bent 2 adduct in the *syn* orientation. The unpaired nucleotides and flanking base-pairs are coloured cyan.



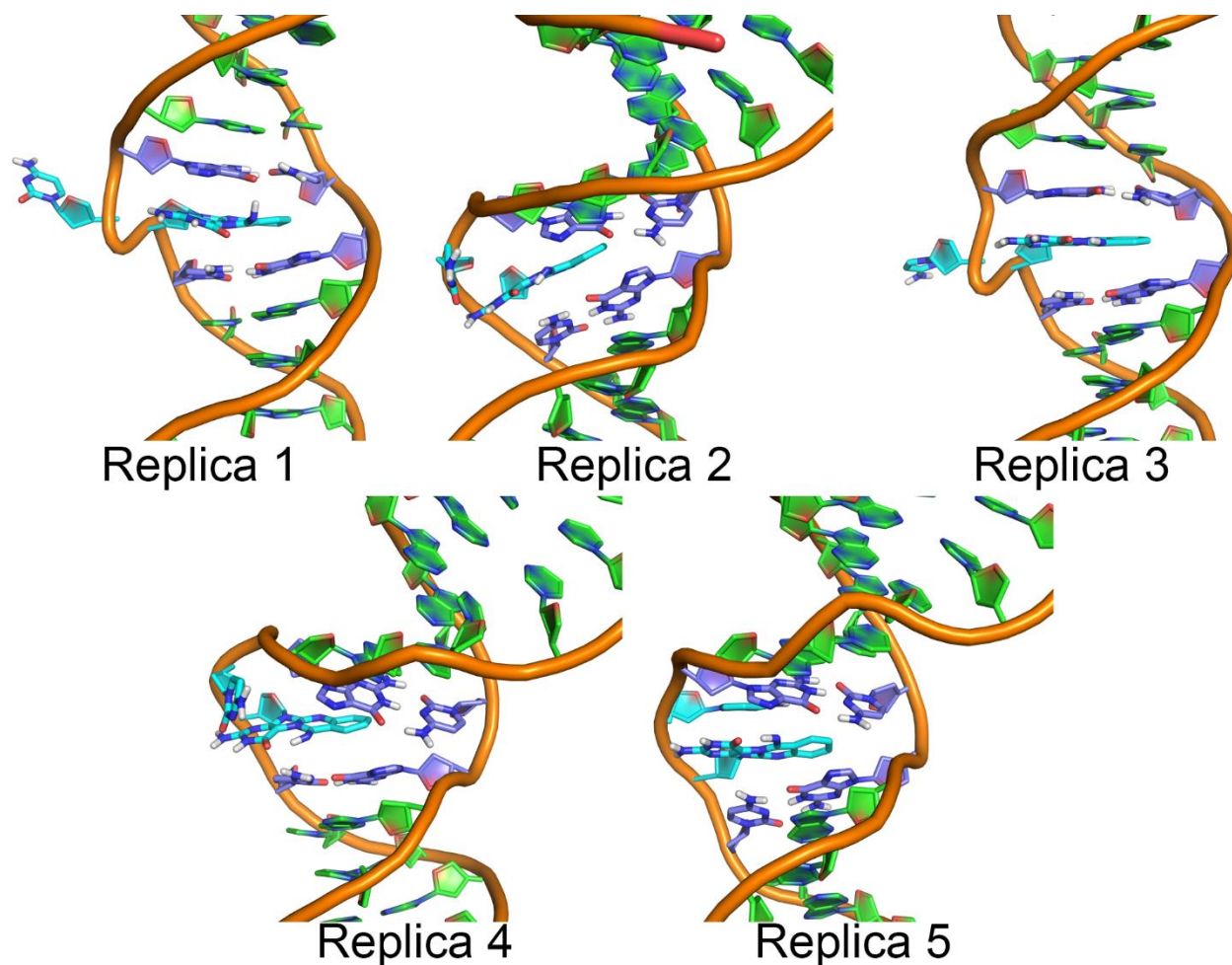


Figure B.30: Representative structures obtained from MD simulations for the five 1  $\mu$ s replicas of **1** adducted SMI duplex in the *syn* orientation. The adducted base and unpaired C are shown in cyan, and the flanking bases are shown in purple.

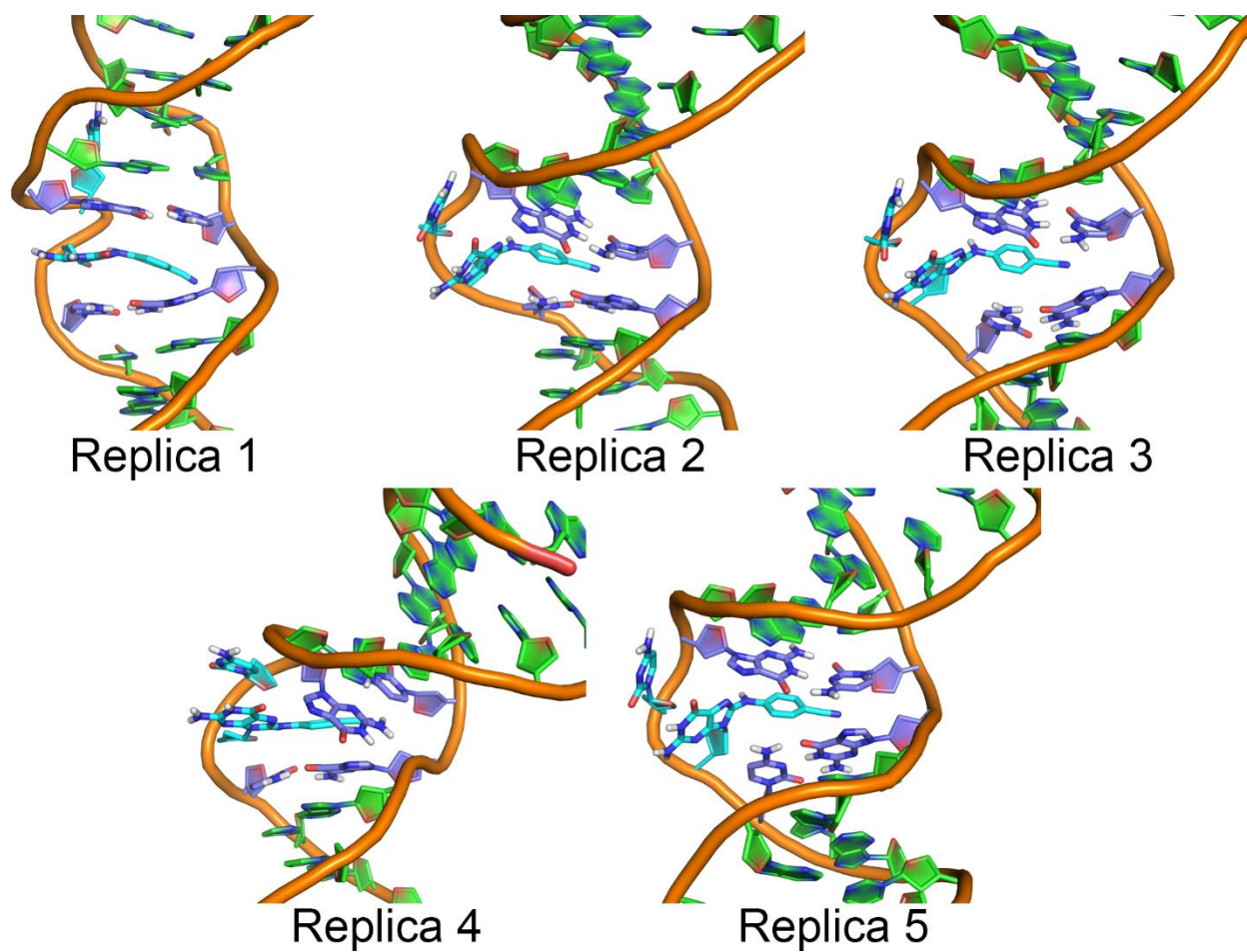


Figure B.31: Representative structures obtained from MD simulations for the five 1  $\mu$ s replicas of 2 adducted SMI duplex in the *syn* orientation. The adducted base and unpaired C are shown in cyan, and the flanking bases are shown in purple.

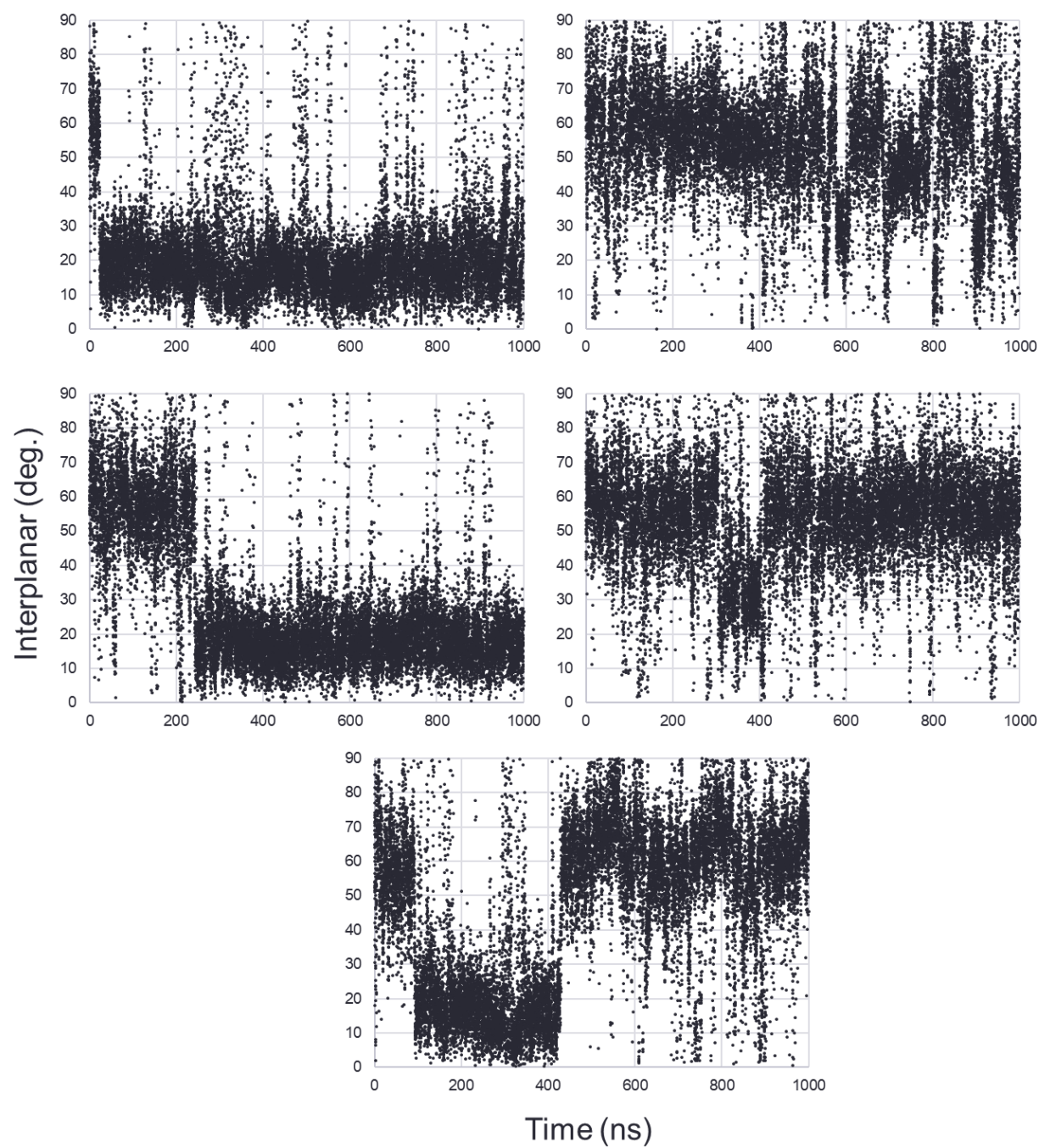


Figure B.32: Interplanar angle between the base pairs flanking the lesion site in the SMI containing *syn* adduct **1** over the course of the five MD replicate simulations.



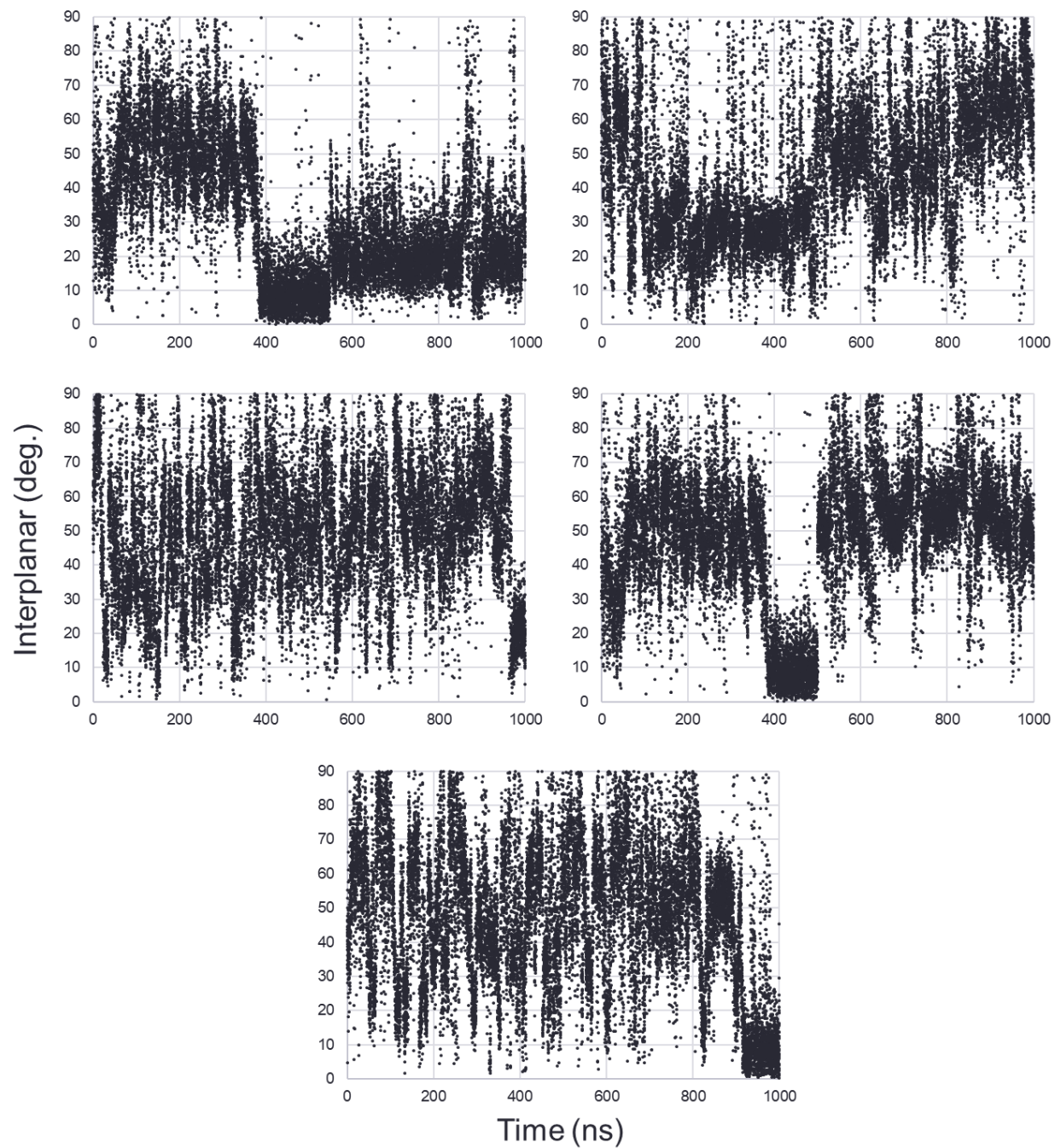


Figure B.33: Interplanar angle between the base-pairs flanking the lesion site in the SMI containing *syn* adduct **2** over the course of the five MD replicate simulations.

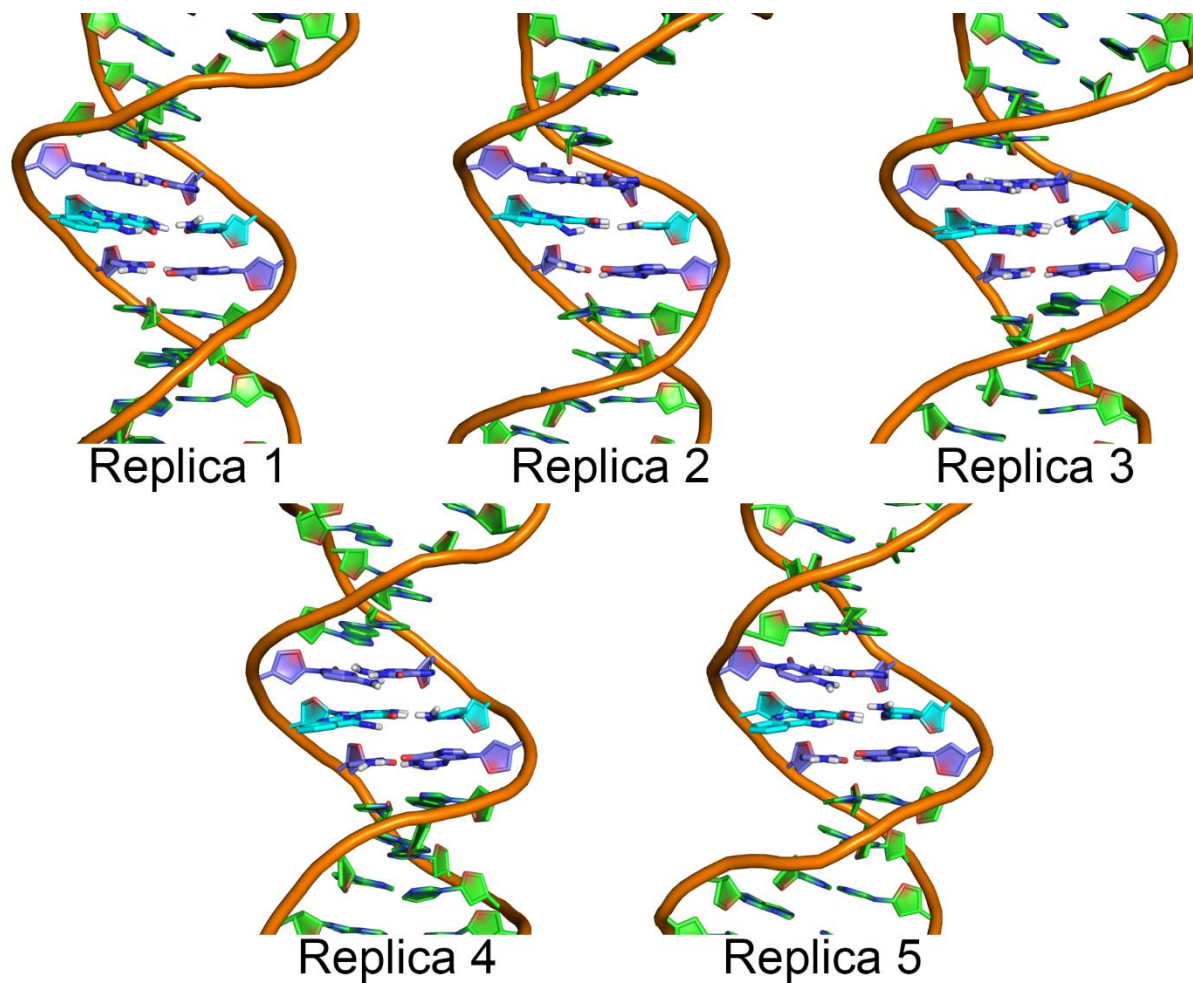


Figure B.34: Representative structures obtained from MD simulations for the five 1  $\mu$ s replicas of **1** adducted FP duplex in the B conformation. The adducted base and pairing C are shown in cyan and the flanking bases are shown in purple.

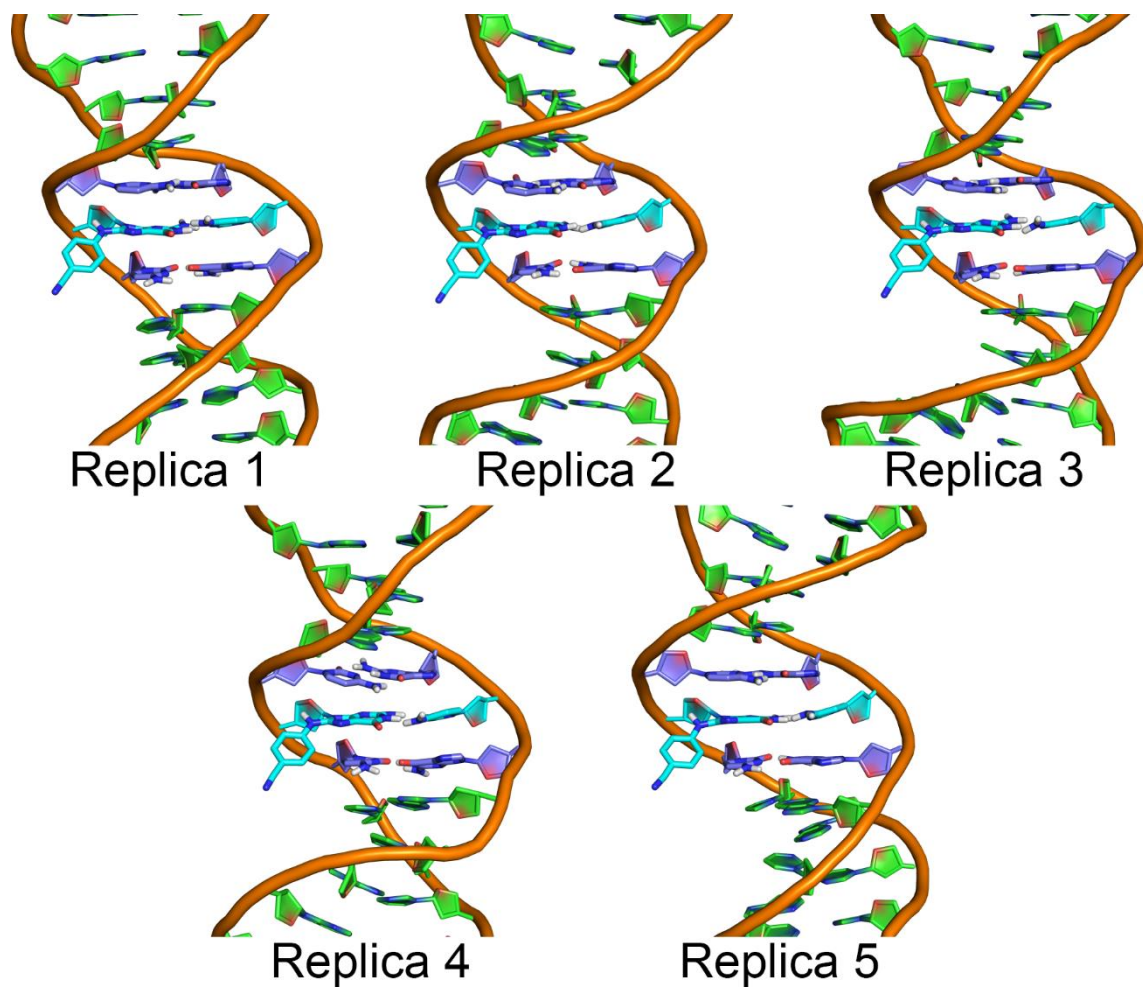


Figure B.35: Representative structures obtained from MD simulations for the five 1  $\mu$ s replicas of 2 adducted FP duplex in the B conformation. The adducted base and pairing C are shown in cyan and the flanking bases are shown in purple.

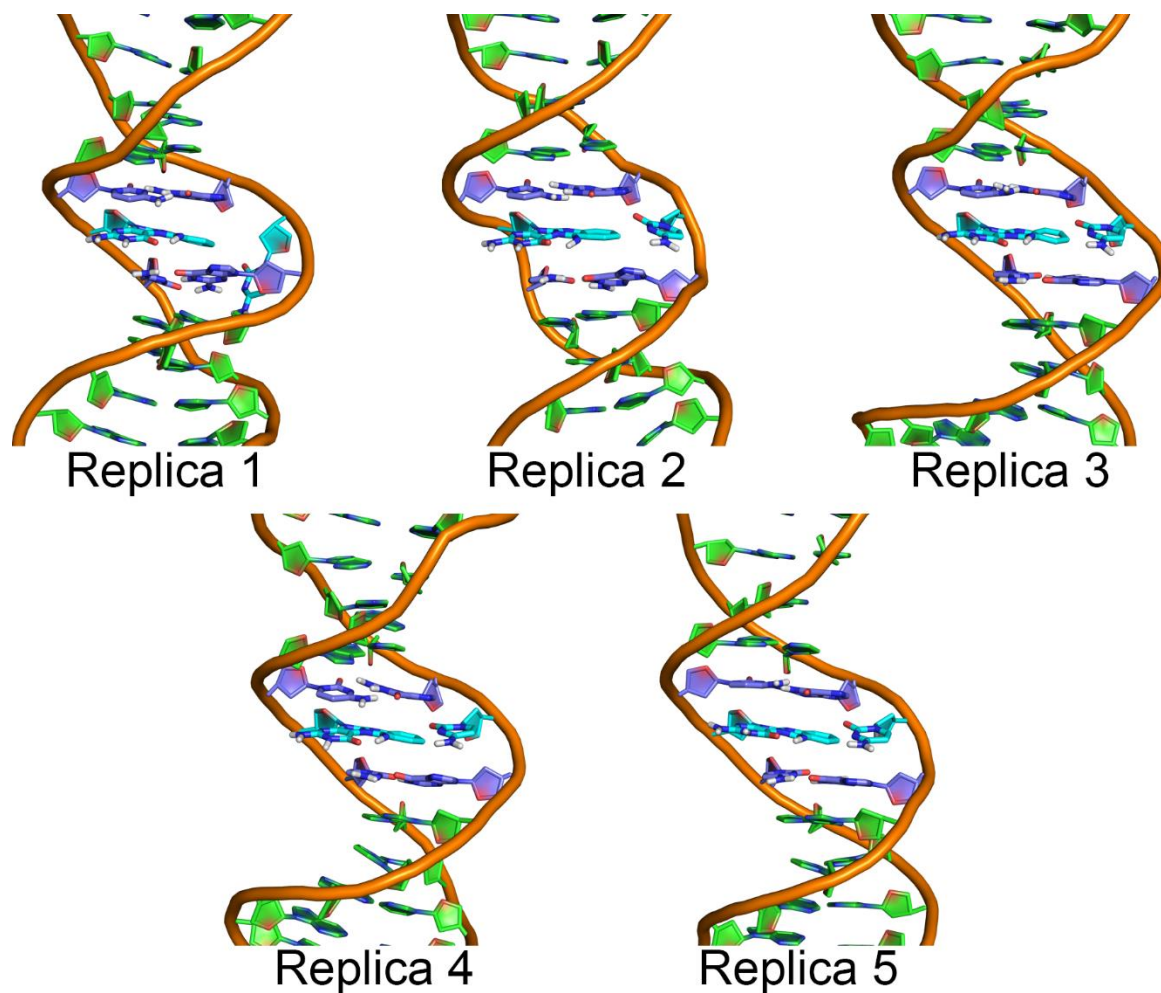


Figure B.36: Representative structures obtained from MD simulations for the five 1  $\mu$ s replicas of **1** adducted FP duplex in the S conformation. The adducted base and pairing C are shown in cyan and the flanking bases are shown in purple.



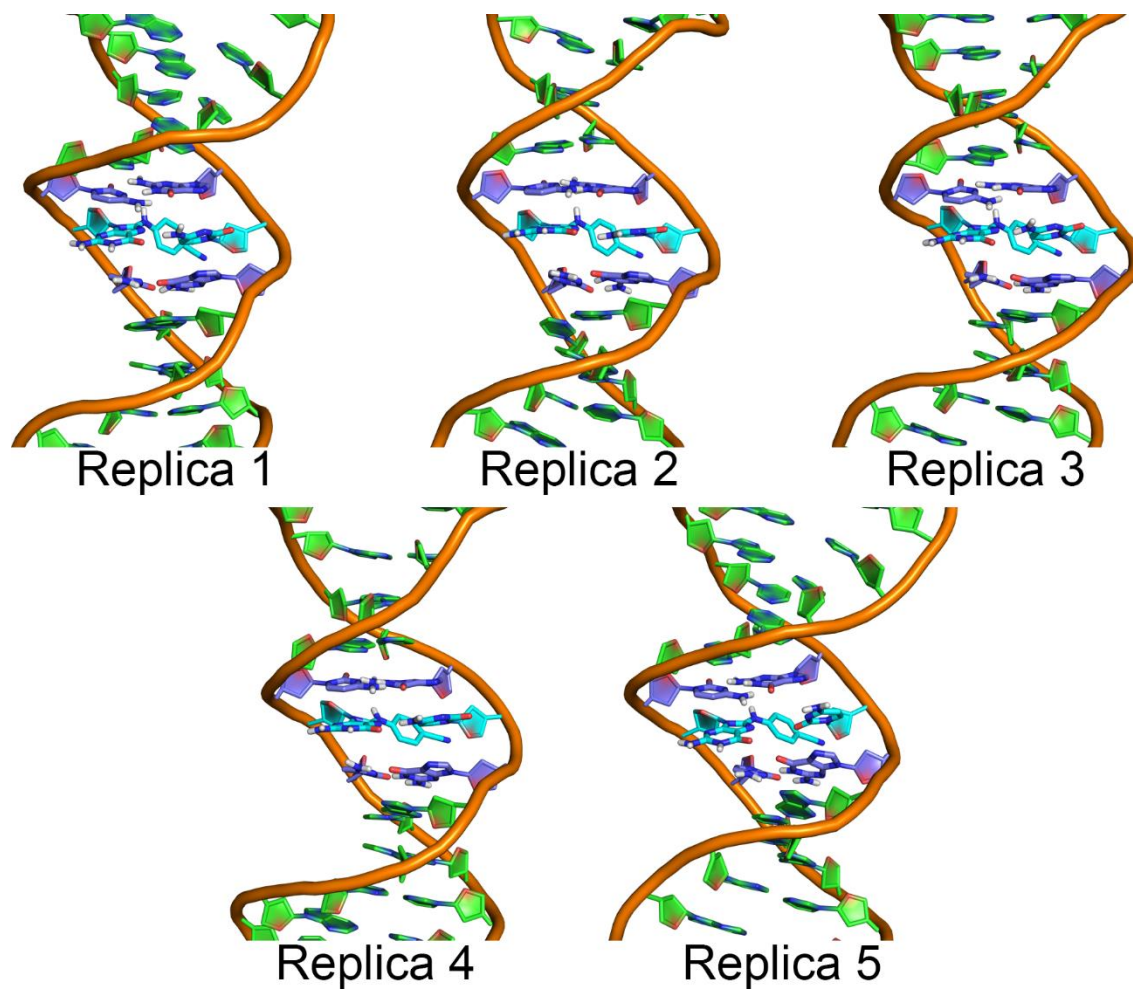


Figure B.37: Representative structures obtained from MD simulations for the five 1  $\mu$ s replicas of **2** adducted FP duplex in the S conformation. The adducted base and pairing C are shown in cyan and the flanking bases are shown in purple.



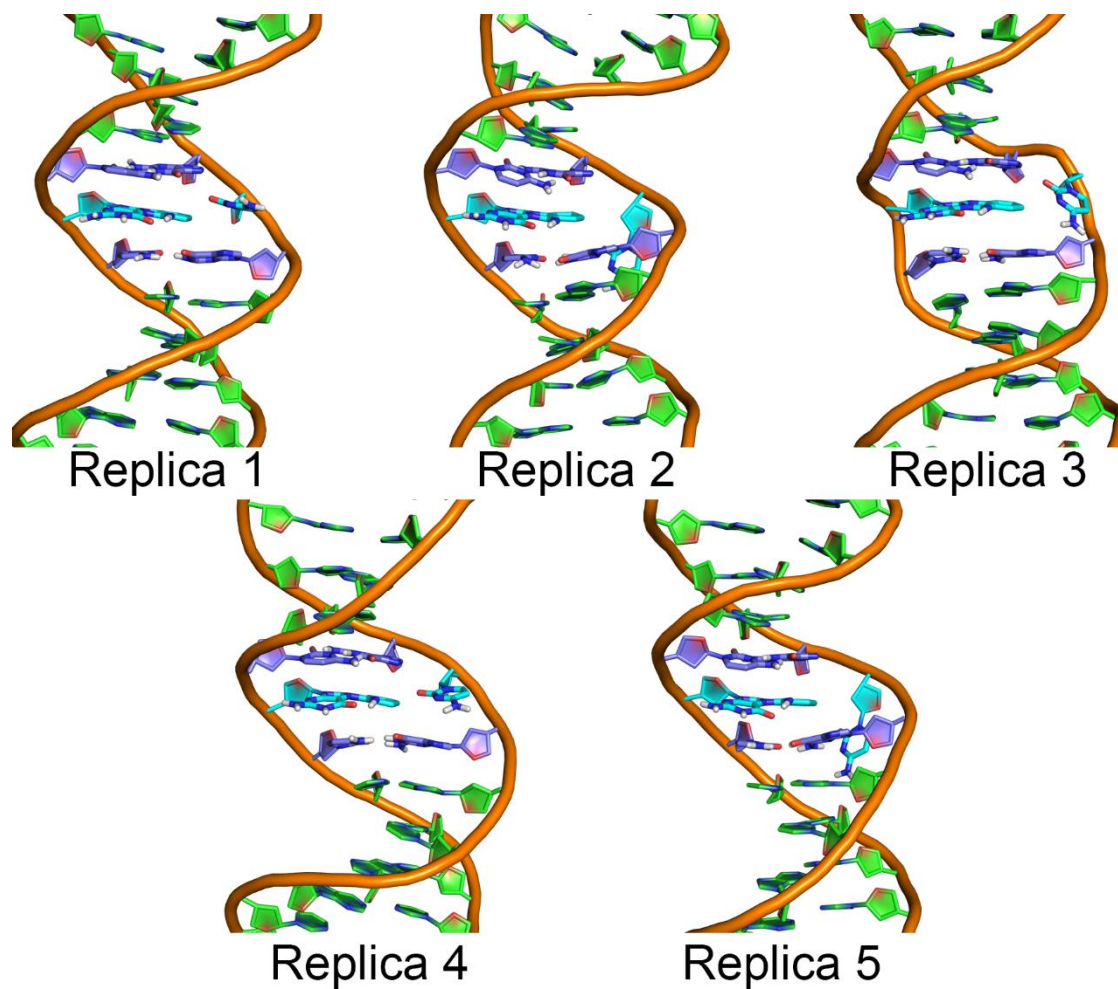


Figure B.38: Representative structures obtained from MD simulations for the five 1  $\mu$ s replicas of **1** adducted FP duplex in the W conformation. The adducted base and pairing C are shown in cyan and the flanking bases are shown in purple.

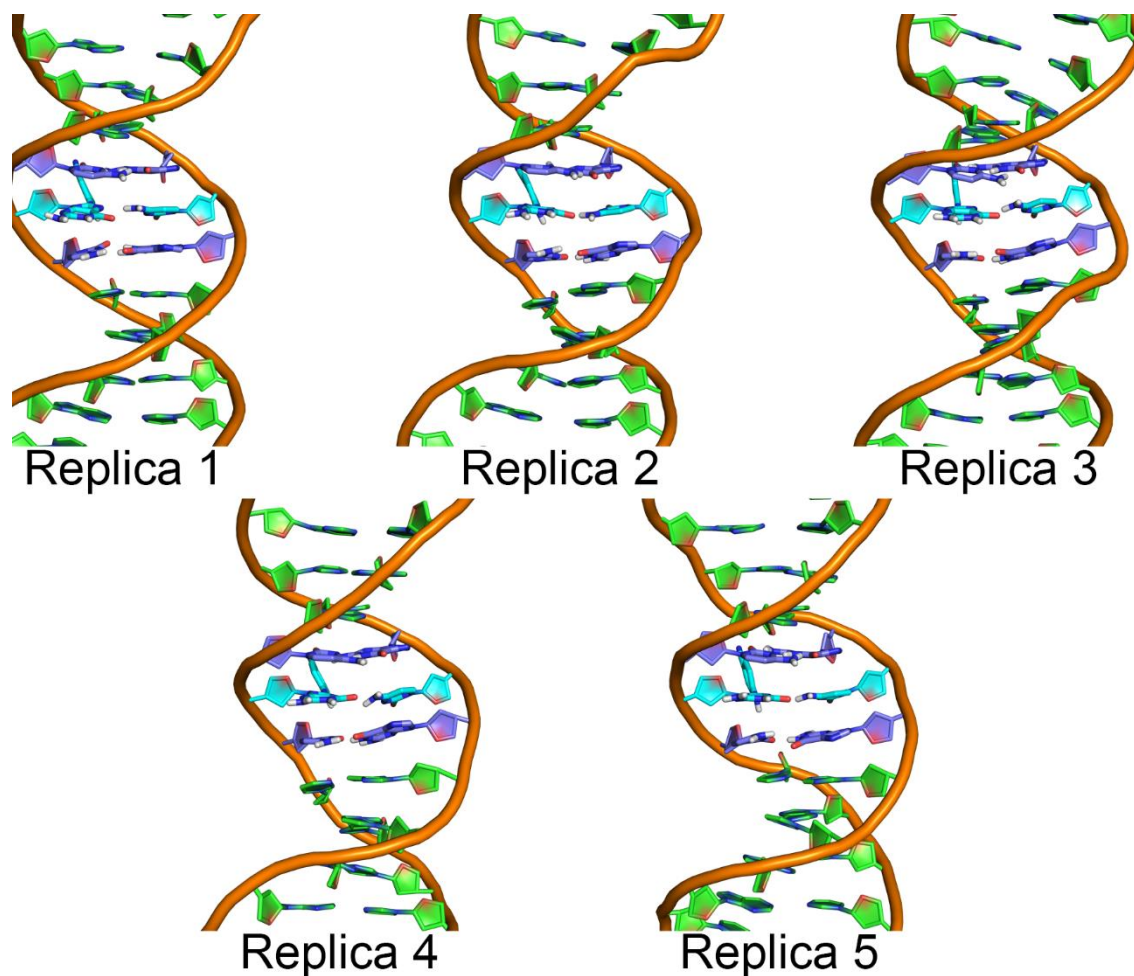


Figure B.39: Representative structures obtained from MD simulations for the five 1  $\mu$ s replicas of 2 adducted FP duplex in the W conformation. The adducted base and pairing C are shown in cyan and the flanking bases are shown in purple.

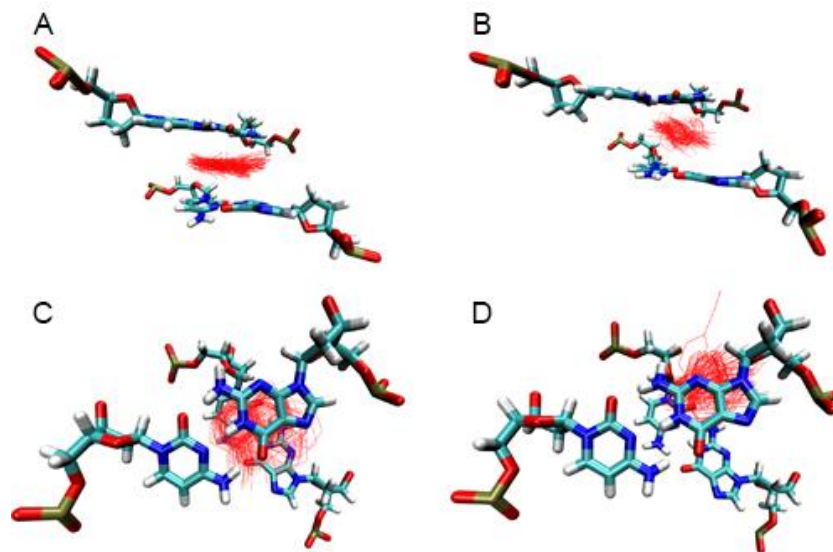


Figure B.40: Positions of the bulky moieties (red) with respect to the flanking base-pairs in the FP duplexes in the S conformation for **1** (A and C) and **2** (B and D). Figures were generated by overlaying 1000 frames across the trajectories based on the atoms of the flanking base pairs.

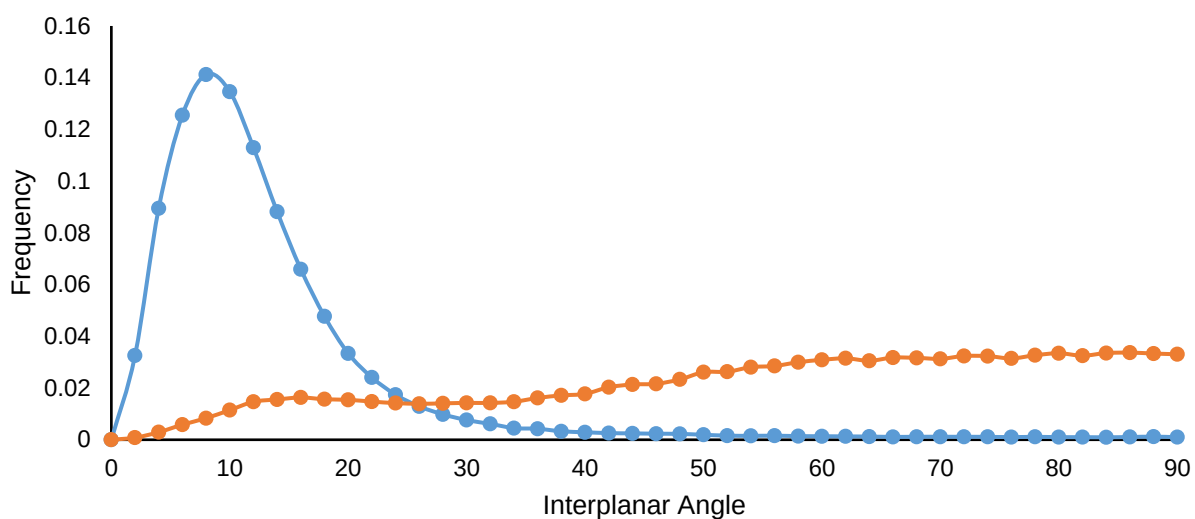


Figure B.41: Histogram of the interplanar angles (deg) between the aromatic portion of the bulky moiety of *syn* **1** (blue) and **2** (orange) and the 5'- or 3'-base pairs from the five MD replicate simulations of the FP duplex.

Table B.1. Comparisons between gas-phase and implicit solvent B3LYP-D3(BJ)/6-311++G(2df,2p) relative energies (kJ/mol) for the *anti* and *syn* conformations of the **1** and **2** nucleosides.

| Adduct   | 5' OH  | Structure   | Gas Phase | Implicit Water |
|----------|--------|-------------|-----------|----------------|
| <b>1</b> | Frozen | <i>anti</i> | 0.8       | 2.9            |
|          |        | <i>syn</i>  | 0.0       | 0.0            |
|          |        | TB.1        | 17.9      | 17.5           |
|          |        | TB.2        | 26.6      | 20.3           |
|          | Free   | <i>anti</i> | 0.0       | 0.0            |
|          |        | <i>syn</i>  | 1.0       | 3.2            |
|          |        | TB.1        | 50.8      | 48.1           |
|          |        | TB.2        | 60.3      | 50.0           |
| <b>2</b> | Frozen | <i>anti</i> | 0.0       | 0.0            |
|          |        | TS          | 46.8      | 38.9           |
|          | Free   | <i>anti</i> | 13.2      | 15.2           |
|          |        | <i>syn</i>  | 0.0       | 0.0            |
|          |        | TB.1        | 50.4      | 42.6           |
|          |        | TB.2        | 19.7      | 19.5           |

Table B.2: Average and standard deviations (parentheses) of the sugar pucker occupancies over all windows during umbrella sampling on the adduct nucleosides.

| Adduct   | C3'-endo  | C4'-exo    | O4'-endo   | C1'-exo     | C2'-endo   |
|----------|-----------|------------|------------|-------------|------------|
| <b>1</b> | 4.8 (3.4) | 10.5 (8.7) | 11.6 (3.2) | 32.0 (11.5) | 26.1 (6.2) |
| <b>2</b> | 3.6 (3.2) | 9.6 (8.4)  | 13.0 (4.4) | 35.3 (11.2) | 25.5 (5.6) |
|          | C3'-exo   | C4'-endo   | O4'-exo    | C1'-endo    | C2'-exo    |
| <b>1</b> | 8.5 (7.1) | 3.3 (4.9)  | 0.7 (0.7)  | 2.3 (3.1)   | 1.5 (1.7)  |
| <b>2</b> | 7.6 (7.6) | 3.0 (4.5)  | 0.6 (0.7)  | 2.3 (3.2)   | 1.3 (1.7)  |

Table B.3: Average and standard deviations (parentheses) of the DNA backbone RMSDs (Å) across 1  $\mu$ s MD simulations.

| Adduct   | Strand Length | Conformation | Replica 1 | Replica 2 | Replica 3 | Replica 4 | Replica 5 |
|----------|---------------|--------------|-----------|-----------|-----------|-----------|-----------|
| <b>1</b> | SMI           | <i>anti</i>  | 5.8 (1.0) | 5.1 (0.9) | 2.6 (0.6) | 4.6 (1.7) | 5.4 (1.3) |
|          |               | <i>syn</i>   | 3.4 (0.6) | 2.9 (0.9) | 4.6 (1.5) | 3.1 (1.1) | 3.4 (1.1) |
|          | FP            | B            | 3.0 (0.6) | 2.8 (0.6) | 3.1 (0.6) | 2.6 (0.6) | 2.7 (0.6) |
|          |               | S            | 2.5 (0.5) | 3.1 (0.7) | 3.2 (0.7) | 3.5 (0.6) | 2.5 (0.6) |
|          |               | W            | 2.5 (0.7) | 3.3 (0.7) | 3.3 (0.7) | 2.6 (0.6) | 2.6 (0.6) |
| <b>2</b> | SMI           | <i>anti</i>  | 5.8 (0.8) | 6.2 (1.5) | 3.6 (0.8) | 3.9 (0.6) | 6.1 (0.8) |
|          |               | <i>syn</i>   | 3.6 (0.7) | 3.2 (0.7) | 3.2 (0.7) | 3.9 (0.9) | 3.9 (1.0) |
|          | FP            | B            | 2.3 (0.5) | 2.9 (0.6) | 2.6 (0.5) | 2.9 (0.7) | 2.9 (0.6) |

|  |  |   |           |           |           |           |           |
|--|--|---|-----------|-----------|-----------|-----------|-----------|
|  |  | S | 3.4 (0.8) | 2.5 (0.7) | 3.3 (0.8) | 3.8 (1.0) | 2.8 (0.6) |
|  |  | W | 3.2 (0.7) | 2.5 (0.5) | 2.6 (0.7) | 2.9 (0.7) | 3.4 (0.7) |

Table B.4: Average and standard deviations (parentheses) of the key dihedral angles (deg) for the positioning of the adduct with respect to the DNA backbone across 1  $\mu$ s MD simulations.

| Adduct   | Strand Length | Conformation | Dihedral | Replica 1    | Replica 2    | Replica 3    | Replica 4    | Replica 5    | Average over Replicas |
|----------|---------------|--------------|----------|--------------|--------------|--------------|--------------|--------------|-----------------------|
| <b>1</b> | SMI           | <i>anti</i>  | $\chi$   | 231.7 (47.9) | 218.2 (22.8) | 240.9 (22.1) | 242.2 (18.8) | 220.3 (20.8) | 230.6 (10.0)          |
|          |               | <i>syn</i>   | $\chi$   | 53.2 (70.5)  | 61.2 (24.7)  | 24.0 (24.5)  | 57.4 (12.0)  | 50.1 (17.1)  | 49.2 (13.1)           |
|          | FP            | B            | $\chi$   | 222.9 (12.3) | 222.0 (13.0) | 221.5 (12.2) | 221.8 (12.1) | 221.9 (12.4) | 222.0 (0.5)           |
|          |               | S            | $\chi$   | 41.1 (11.2)  | 42.6 (11.8)  | 41.2 (11.7)  | 40.7 (11.4)  | 41.3 (11.5)  | 41.4 (0.7)            |
|          |               | W            | $\chi$   | 42.0 (11.6)  | 41.4 (11.7)  | 43.7 (11.6)  | 42.1 (12.0)  | 41.3 (11.3)  | 42.1 (0.9)            |
| <b>2</b> | SMI           | <i>anti</i>  | $\chi$   | 203.3 (12.8) | 204.1 (15.4) | 195.7 (15.1) | 202.2 (11.9) | 200.3 (12.5) | 201.1 (3.0)           |
|          |               |              | $\theta$ | 200.7 (21.2) | 200.3 (31.1) | 196.5 (30.9) | 202.3 (19.4) | 199.6 (22.6) | 199.9 (1.9)           |
|          |               |              | $\phi$   | 180.0 (10.3) | 180.5 (10.6) | 180.0 (10.5) | 179.4 (10.5) | 180.4 (10.4) | 180.1 (0.4)           |
|          |               | <i>syn</i>   | $\chi$   | 177.0 (39.7) | 77.3 (35.1)  | 84.7 (39.8)  | 125.5 (46.4) | 91.8 (52.1)  | 111.2 (36.8)          |
|          |               |              | $\theta$ | 207.9 (24.7) | 188.4 (80.3) | 236.8 (39.2) | 200.2 (43.9) | 229.5 (41.1) | 212.6 (18.1)          |
|          |               |              | $\phi$   | 356.5 (13.2) | 359.4 (11.1) | 358.0 (11.6) | 3.9 (12.8)   | 357.4 (11.4) | 359.0 (2.6)           |
|          | FP            | B            | $\chi$   | 230.3 (13.4) | 230.8 (13.0) | 231.5 (12.9) | 230.0 (14.3) | 230.7 (13.1) | 230.7 (0.5)           |
|          |               |              | $\theta$ | 137.3 (24.9) | 136.9 (24.3) | 136.6 (24.2) | 137.3 (25.8) | 136.9 (24.9) | 137.0 (0.3)           |
|          |               |              | $\phi$   | 2.9 (12.0)   | 2.9 (12.0)   | 2.7 (11.9)   | 2.7 (11.8)   | 3.0 (11.9)   | 2.8 (0.1)             |
|          |               | S            | $\chi$   | 47.0 (15.3)  | 46.9 (15.4)  | 48.2 (15.4)  | 49.3 (15.1)  | 43.8 (14.8)  | 47.1 (1.8)            |
|          |               |              | $\theta$ | 235.7 (34.5) | 227.5 (39.4) | 232.9 (34.1) | 250.8 (35.5) | 217.1 (28.7) | 232.8 (11.0)          |
|          |               |              | $\phi$   | 183.0 (10.8) | 181.6 (10.9) | 181.3 (10.9) | 182.3 (10.3) | 178.5 (10.4) | 181.4 (1.5)           |
|          |               | W            | $\chi$   | 54.2 (17.4)  | 60.3 (15.8)  | 59.0 (15.9)  | 57.5 (17.2)  | 57.7 (16.0)  | 57.7 (2.0)            |
|          |               |              | $\theta$ | 123.9 (66.3) | 87.4 (54.7)  | 107.4 (61.1) | 121.1 (66.2) | 103.3 (65.0) | 108.6 (13.2)          |
|          |               |              | $\phi$   | 359.1 (11.2) | 1.5 (10.1)   | 0.4 (10.7)   | 359.7 (11.0) | 0.6 (10.7)   | 0.2 (0.8)             |

Table B.5: Average and standard deviations (parentheses) of the sugar pucker occupancies over all replicas for each SMI and FP containing adducts **1** and **2**.

| Adduct   | System            | <b>C3'-endo</b> | <b>C4'-exo</b>  | <b>O4'-endo</b> | <b>C1'-exo</b>  | <b>C2'-endo</b> |
|----------|-------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| <b>1</b> | SMI – <i>anti</i> | 1.1 (1.1)       | 3.9 (1.4)       | 21.9 (11.3)     | 28.4 (4.3)      | 33.8 (11.1)     |
|          | SMI – <i>syn</i>  | 7.5 (4.2)       | 29.7 (15.0)     | 12.7 (9.0)      | 17.6 (9.5)      | 14.8 (8.1)      |
|          | FP – B            | 0.0 (0.0)       | 4.5 (0.2)       | 69.0 (0.5)      | 24.7 (0.4)      | 1.3 (0.5)       |
|          | FP – S            | 0.0 (0.0)       | 4.3 (0.6)       | 62.4 (4.7)      | 33.0 (5.3)      | 0.3 (0.0)       |
|          | FP – W            | 0.0 (0.0)       | 4.0 (0.9)       | 59.0 (7.6)      | 35.3 (9.0)      | 1.4 (0.6)       |
| <b>2</b> | SMI – <i>anti</i> | 0.6 (0.2)       | 1.7 (0.5)       | 9.0 (2.7)       | 33.6 (2.4)      | 51.8 (4.5)      |
|          | SMI – <i>syn</i>  | 8.1 (4.1)       | 14.5 (8.6)      | 4.3 (1.9)       | 34.3 (4.4)      | 34.5 (11.3)     |
|          | FP – B            | 0.1 (0.0)       | 10.6 (0.2)      | 65.2 (1.1)      | 23.1 (0.3)      | 1.1 (0.7)       |
|          | FP – S            | 0.2 (0.0)       | 6.9 (0.9)       | 44.3 (7.9)      | 44.6 (8.0)      | 3.6 (1.2)       |
|          | FP – W            | 0.0 (0.0)       | 1.5 (2.1)       | 9.1 (6.4)       | 70.2 (5.6)      | 18.3 (2.9)      |
|          |                   | <b>C3'-exo</b>  | <b>C4'-endo</b> | <b>O4'-exo</b>  | <b>C1'-endo</b> | <b>C2'-exo</b>  |
| <b>1</b> | SMI – <i>anti</i> | 10.5 (4.8)      | 0.2 (0.2)       | 0.0 (0.0)       | 0.0 (0.0)       | 0.0 (0.0)       |
|          | SMI – <i>syn</i>  | 13.8 (15.5)     | 2.8 (3.9)       | 0.0 (0.0)       | 0.2 (0.2)       | 0.8 (0.3)       |
|          | FP – B            | 0.5 (0.3)       | 0.0 (0.0)       | 0.0 (0.0)       | 0.0 (0.0)       | 0.0 (0.0)       |
|          | FP – S            | 0.0 (0.0)       | 0.0 (0.0)       | 0.0 (0.0)       | 0.0 (0.0)       | 0.0 (0.0)       |
|          | FP – W            | 0.3 (0.2)       | 0.0 (0.0)       | 0.0 (0.0)       | 0.0 (0.0)       | 0.0 (0.0)       |
| <b>2</b> | SMI – <i>anti</i> | 2.8 (1.2)       | 0.2 (0.4)       | 0.0 (0.1)       | 0.1 (0.0)       | 0.3 (0.2)       |
|          | SMI – <i>syn</i>  | 3.0 (1.5)       | 0.1 (0.0)       | 0.1 (0.0)       | 0.1 (0.1)       | 1.1 (0.6)       |
|          | FP – B            | 0.0 (0.0)       | 0.0 (0.0)       | 0.0 (0.0)       | 0.0 (0.0)       | 0.0 (0.0)       |
|          | FP – S            | 0.3 (0.3)       | 0.0 (0.0)       | 0.0 (0.0)       | 0.0 (0.0)       | 0.0 (0.0)       |
|          | FP – W            | 0.9 (0.3)       | 0.0 (0.0)       | 0.0 (0.0)       | 0.0 (0.0)       | 0.0 (0.0)       |

Table B.6: Occupancies (%) of the hydrogen bonds of the adduct and unpaired C when the damaged G is in the *anti* conformation for the SMI models.<sup>a</sup>

| Donor       | Acceptor  | Adduct   | Replica 1 | Replica 2 | Replica 3 | Replica 4 | Replica 5 | Average | Standard Deviation |
|-------------|-----------|----------|-----------|-----------|-----------|-----------|-----------|---------|--------------------|
| dC24:N4H    | Lesion:O6 | <b>1</b> | 77.3%     | 94.1%     | 95.8%     | 97.1%     | 97.6%     | 92.4%   | 7.7%               |
|             |           | <b>2</b> | 97.3%     | 93.9%     | 59.9%     | 98.5%     | 98.1%     | 89.5%   | 14.9%              |
| Lesion:N1H  | dC24:N3   | <b>1</b> | 78.1%     | 95.1%     | 96.8%     | 98.4%     | 98.4%     | 93.3%   | 7.7%               |
|             |           | <b>2</b> | 98.2%     | 95.7%     | 60.5%     | 99.4%     | 99.6%     | 90.7%   | 15.2%              |
| Lesion:N2H  | dC24:O2   | <b>1</b> | 78.1%     | 95.4%     | 97.8%     | 98.5%     | 99.4%     | 93.8%   | 8.0%               |
|             |           | <b>2</b> | 99.2%     | 98.5%     | 60.6%     | 99.5%     | 99.6%     | 91.5%   | 15.4%              |
| dC8:N4H     | dA27:OP1  | <b>1</b> | 0.8%      | 0.0%      | 0.0%      | 0.0%      | 27.8%     | 5.7%    | 11.1%              |
| dC8:N4H     | dG6:N3    | <b>1</b> | 39.5%     | 0.0%      | 0.1%      | 0.0%      | 40.0%     | 15.9%   | 19.5%              |
| dC8:N4H     | dG7:OP2   | <b>1</b> | 0.0%      | 0.0%      | 46.1%     | 32.1%     | 40.0%     | 23.7%   | 19.8%              |
| Lesion:N17H | Lesion:O6 | <b>1</b> | 99.0%     | 98.8%     | 98.4%     | 96.4%     | 99.2%     | 98.4%   | 1.0%               |
| Lesion:N1H  | dC24:O2   | <b>1</b> | 23.7%     | 21.2%     | 27.8%     | 28.3%     | 32.0%     | 26.6%   | 3.8%               |
| Lesion:N10H | dC10:O4'  | <b>2</b> | 35.0%     | 33.8%     | 40.0%     | 37.7%     | 42.1%     | 37.7%   | 3.1%               |
| Lesion:N10H | dC10:O5'  | <b>2</b> | 48.5%     | 38.5%     | 37.7%     | 48.0%     | 35.8%     | 41.7%   | 5.4%               |
| Lesion:N2H  | dC10:O4'  | <b>2</b> | 32.8%     | 25.6%     | 19.6%     | 32.8%     | 27.8%     | 27.7%   | 4.9%               |
| Lesion:N1H  | dC24:O4'  | <b>2</b> | 0.0%      | 0.0%      | 27.4%     | 0.0%      | 0.0%      | 5.5%    | 11.0%              |
| Lesion:N2H  | dC24:O4'  | <b>2</b> | 0.0%      | 0.0%      | 26.5%     | 0.0%      | 0.0%      | 5.3%    | 10.6%              |

<sup>a</sup>Hydrogen-bonding interactions were determined using a cut-off of 3.4 Å for the donor–acceptor distance and 120° for the donor–hydrogen–acceptor angle.



Table B.7: Occupancies (%) of the hydrogen bonds of the flanking bases when the lesion is started from the *anti* conformation for the SMI models. <sup>a</sup>

| Donor    | Acceptor | Adduct   | Replica 1 | Replica 2 | Replica 3 | Replica 4 | Replica 5 | Average | Standard Deviation |
|----------|----------|----------|-----------|-----------|-----------|-----------|-----------|---------|--------------------|
| dG23:N1H | dC10:N3  | <b>1</b> | 99.9%     | 99.9%     | 99.8%     | 99.9%     | 99.9%     | 99.9%   | 0.0%               |
|          |          | <b>2</b> | 100.0%    | 99.9%     | 100.0%    | 99.9%     | 100.0%    | 99.9%   | 0.0%               |
| dG23:N2H | dC10:O2  | <b>1</b> | 99.8%     | 99.8%     | 99.8%     | 99.8%     | 99.8%     | 99.8%   | 0.0%               |
|          |          | <b>2</b> | 99.7%     | 99.9%     | 99.8%     | 99.7%     | 99.8%     | 99.8%   | 0.1%               |
| dC10:N4H | dG23:O6  | <b>1</b> | 98.7%     | 98.7%     | 98.5%     | 98.7%     | 98.6%     | 98.6%   | 0.1%               |
|          |          | <b>2</b> | 98.9%     | 98.8%     | 98.9%     | 98.9%     | 98.7%     | 98.8%   | 0.1%               |
| dG7:N1H  | dC25:N3  | <b>1</b> | 68.4%     | 84.8%     | 21.4%     | 4.9%      | 79.0%     | 51.7%   | 32.3%              |
|          |          | <b>2</b> | 92.7%     | 74.0%     | 56.8%     | 95.8%     | 89.2%     | 81.7%   | 14.5%              |
| dG7:N2H  | dC25:O2  | <b>1</b> | 73.6%     | 86.7%     | 22.1%     | 4.9%      | 81.2%     | 53.7%   | 33.5%              |
|          |          | <b>2</b> | 93.2%     | 79.5%     | 58.2%     | 96.6%     | 94.2%     | 84.3%   | 14.4%              |
| dC25:N4H | dG7:O6   | <b>1</b> | 65.5%     | 82.9%     | 20.4%     | 4.7%      | 76.9%     | 50.1%   | 31.6%              |
|          |          | <b>2</b> | 90.2%     | 64.4%     | 54.8%     | 92.8%     | 84.1%     | 77.2%   | 15.0%              |
| dG7:N1H  | dC24:N3  | <b>2</b> | 2.2%      | 5.9%      | 39.6%     | 0.9%      | 0.7%      | 9.9%    | 15.0%              |
| dG7:N2H  | dC24:O2  | <b>2</b> | 2.1%      | 5.6%      | 39.5%     | 0.8%      | 0.7%      | 9.7%    | 15.0%              |
| dC24:N4H | dG7:O6   | <b>2</b> | 1.8%      | 4.8%      | 38.4%     | 0.8%      | 0.8%      | 9.3%    | 14.6%              |

<sup>a</sup>Hydrogen-bonding interactions were determined using a cut-off of 3.4 Å for the donor–acceptor distance and 120° for the donor–hydrogen–acceptor angle.

Table B.8: Occupancies (%) of the hydrogen bonds of the adduct and unpaired C when the damaged G is in the *syn* conformation for the SMI models. <sup>a</sup>

| Donor       | Acceptor   | Adduct   | Replica 1 | Replica 2 | Replica 3 | Replica 4 | Replica 5 | Average | Standard Deviation |
|-------------|------------|----------|-----------|-----------|-----------|-----------|-----------|---------|--------------------|
| dC8:N4H     | dG7:OP2    | <b>1</b> | 1.1%      | 31.3%     | 14.7%     | 52.2%     | 5.5%      | 21.0%   | 18.7%              |
|             |            | <b>2</b> | 28.2%     | 64.1%     | 73.2%     | 82.3%     | 65.0%     | 62.5%   | 18.4%              |
| dC8:N4H     | dG7:N3     | <b>1</b> | 0.0%      | 1.0%      | 0.0%      | 0.0%      | 30.0%     | 6.2%    | 11.9%              |
| Lesion:N1H  | dC8:O2     | <b>1</b> | 0.5%      | 16.2%     | 7.2%      | 27.4%     | 3.0%      | 10.8%   | 9.8%               |
| Lesion:N17H | Lesion:O6  | <b>1</b> | 93.1%     | 98.2%     | 92.2%     | 98.9%     | 97.9%     | 96.0%   | 2.8%               |
| dC8:N4H     | dG:6@N3    | <b>2</b> | 27.9%     | 0.0%      | 0.0%      | 0.0%      | 0.9%      | 5.8%    | 11.1%              |
| dC8:N4H     | dG:6@O4'   | <b>2</b> | 28.5%     | 0.0%      | 0.0%      | 0.0%      | 2.9%      | 6.3%    | 11.2%              |
| Lesion:N10H | dC10:O4'   | <b>2</b> | 56.2%     | 13.0%     | 20.2%     | 59.6%     | 25.6%     | 34.9%   | 19.2%              |
| Lesion:N10H | dC10:O5'   | <b>2</b> | 28.2%     | 0.1%      | 0.0%      | 0.0%      | 1.4%      | 5.9%    | 11.1%              |
| dG7:N2H     | Lesion:OP1 | <b>2</b> | 60.8%     | 0.0%      | 0.0%      | 0.0%      | 8.5%      | 13.9%   | 23.7%              |

<sup>a</sup>Hydrogen-bonding interactions were determined using a cut-off of 3.4 Å for the donor–acceptor distance and 120° for the donor–hydrogen–acceptor angle.

Table B.9: Occupancies (%) of the hydrogen bonds of the flanking bases when the lesion is started from the *syn* conformation for the SMI models. <sup>a</sup>

| Donor    | Acceptor | Adduct | Replica 1 | Replica 2 | Replica 3 | Replica 4 | Replica 5 | Average | Standard Deviation |
|----------|----------|--------|-----------|-----------|-----------|-----------|-----------|---------|--------------------|
| dG7:N1H  | dC24:N3  | 1      | 99.8%     | 99.3%     | 99.8%     | 99.6%     | 99.8%     | 99.7%   | 0.2%               |
|          |          | 2      | 99.3%     | 99.0%     | 99.1%     | 41.1%     | 99.7%     | 87.6%   | 23.3%              |
| dG7:N2H  | dC24:O2  | 1      | 99.4%     | 99.0%     | 99.7%     | 98.2%     | 99.4%     | 99.2%   | 0.5%               |
|          |          | 2      | 99.5%     | 98.0%     | 98.7%     | 41.1%     | 99.4%     | 87.3%   | 23.1%              |
| dC24:N4H | dG7:O6   | 1      | 97.7%     | 97.4%     | 97.8%     | 98.4%     | 98.1%     | 97.9%   | 0.3%               |
|          |          | 2      | 95.0%     | 96.4%     | 95.7%     | 39.8%     | 96.6%     | 84.7%   | 22.4%              |
| dG23:N1H | dC10:N3  | 1      | 99.5%     | 99.9%     | 100.0%    | 99.9%     | 99.3%     | 99.7%   | 0.3%               |
|          |          | 2      | 98.5%     | 94.9%     | 97.7%     | 99.2%     | 99.1%     | 97.9%   | 1.6%               |
| dG23:N2H | dC10:O2  | 1      | 99.8%     | 99.7%     | 99.9%     | 99.8%     | 99.8%     | 99.8%   | 0.1%               |
|          |          | 2      | 98.2%     | 97.6%     | 98.7%     | 99.1%     | 99.2%     | 98.6%   | 0.6%               |
| dC10:N4H | dG23:O6  | 1      | 98.4%     | 98.2%     | 99.1%     | 98.9%     | 98.1%     | 98.5%   | 0.4%               |
|          |          | 2      | 96.5%     | 90.4%     | 94.2%     | 97.6%     | 96.1%     | 95.0%   | 2.5%               |
| dC24:NH4 | dG23:OP2 | 1      | 2.1%      | 61.3%     | 18.3%     | 63.8%     | 43.9%     | 37.9%   | 24.1%              |
|          |          | 2      | 16.0%     | 26.7%     | 31.7%     | 6.7%      | 29.4%     | 22.1%   | 9.4%               |
| dG7:N1H  | dG23:OP2 | 2      | 0.0%      | 0.0%      | 0.0%      | 55.5%     | 0.0%      | 11.1%   | 22.2%              |
| dG7:N2H  | dG23:OP2 | 2      | 0.0%      | 0.0%      | 0.0%      | 55.7%     | 0.0%      | 11.1%   | 22.3%              |
| dC24:NH4 | dG7:N3   | 2      | 0.0%      | 0.0%      | 0.0%      | 38.6%     | 0.0%      | 7.7%    | 15.5%              |
| dG7:N2H  | dC24:N3  | 2      | 26.0%     | 17.6%     | 19.1%     | 9.3%      | 20.9%     | 18.6%   | 5.4%               |

<sup>a</sup>Hydrogen-bonding interactions were determined using a cut-off of 3.4 Å for the donor–acceptor distance and 120° for the donor–hydrogen–acceptor angle.

Table B.10: Average residence times (ps, amount of time the helix adopts one conformation before switching conformations) for the planar and bent conformations of the *syn* SMI structures.<sup>a</sup>

| Adduct | Conformation | Replica 1 | Replica 2 | Replica 3 | Replica 4 | Replica 5 | Average   |
|--------|--------------|-----------|-----------|-----------|-----------|-----------|-----------|
| 1      | Planar       | 1167      | 137       | 1136      | 138       | 402       | 596 ± 464 |
|        | Bent         | 99        | 556       | 363       | 639       | 683       | 468 ± 215 |
| 2      | Planar       | 587       | 291       | 207       | 200       | 242       | 306 ± 144 |
|        | Bent         | 275       | 348       | 434       | 548       | 425       | 406 ± 91  |

<sup>a</sup>Structures were classified as planar or bent according to whether the interplanar angle between the base pairs flanking the lesion site is less than or greater than 40°, respectively.

Table B.11: Histogram of residence times for the SMI adducts switching between the planar and bent conformations for all five replicas.<sup>a</sup>

| Residence time (ns) | 1    | 2     |
|---------------------|------|-------|
| < 0.5               | 8785 | 12609 |
| 0.5 to 1.0          | 574  | 715   |
| 1.0 to 1.5          | 243  | 294   |
| 1.5 to 2.0          | 133  | 160   |
| 2.0 to 2.5          | 87   | 115   |
| 2.5 to 3.0          | 75   | 68    |
| 3.0 to 3.5          | 52   | 54    |
| 3.5 to 4.0          | 45   | 39    |
| 4.0 to 4.5          | 21   | 27    |
| 4.5 to 5.0          | 34   | 16    |
| > 5.0               | 164  | 140   |

<sup>a</sup>Each simulation was sampled every 50 ps and classified as planar or bent according to whether the interplanar angle between the base pairs flanking the lesion site is less than or greater than 40°, respectively.

Table B.12: Occupancies (%) of the hydrogen bonds at the lesion site containing the lesion pair in the B conformation for the FP models. <sup>a</sup>

| Donor       | Acceptor   | Adduct   | Replica 1 | Replica 2 | Replica 3 | Replica 4 | Replica 5 | Average | Standard Deviation |
|-------------|------------|----------|-----------|-----------|-----------|-----------|-----------|---------|--------------------|
| Lesion:N1H  | dC24:N3    | <b>1</b> | 99.9%     | 99.9%     | 99.9%     | 99.9%     | 99.8%     | 99.9%   | 0.0%               |
|             |            | <b>2</b> | 99.8%     | 99.8%     | 99.8%     | 99.8%     | 99.6%     | 99.8%   | 0.1%               |
| Lesion:N2H  | dC24:O2    | <b>1</b> | 99.9%     | 99.9%     | 99.9%     | 99.9%     | 99.9%     | 99.9%   | 0.0%               |
|             |            | <b>2</b> | 100.0%    | 99.9%     | 99.9%     | 99.9%     | 99.9%     | 99.9%   | 0.0%               |
| dC24:N4H    | Lesion:O6  | <b>1</b> | 99.3%     | 99.4%     | 99.2%     | 99.4%     | 99.3%     | 99.3%   | 0.1%               |
|             |            | <b>2</b> | 98.3%     | 98.0%     | 98.1%     | 98.1%     | 98.1%     | 98.1%   | 0.1%               |
| Lesion:N17H | Lesion:O6  | <b>1</b> | 98.9%     | 98.9%     | 98.8%     | 98.7%     | 98.8%     | 98.8%   | 0.1%               |
| Lesion:N10H | Lesion:O4' | <b>2</b> | 61.8%     | 61.6%     | 61.4%     | 60.4%     | 60.8%     | 61.2%   | 0.6%               |

<sup>a</sup>Hydrogen-bonding interactions were determined using a cut-off of 3.4 Å for the donor–acceptor distance and 120° for the donor–hydrogen–acceptor angle.

Table B.13: Occupancies (%) of the hydrogen bonds of the flanking bases in the B Conformation for the FP models. <sup>a</sup>

| Donor    | Acceptor | Adduct   | Replica 1 | Replica 2 | Replica 3 | Replica 4 | Replica 5 | Average | Standard Deviation |
|----------|----------|----------|-----------|-----------|-----------|-----------|-----------|---------|--------------------|
| dG25:N1H | dC8:N3   | <b>1</b> | 99.9%     | 99.9%     | 99.9%     | 99.9%     | 99.9%     | 99.9%   | 0.0%               |
|          |          | <b>2</b> | 100.0%    | 100.0%    | 100.0%    | 100.0%    | 100.0%    | 100.0%  | 0.0%               |
| dG25:N2H | dC8:O2   | <b>1</b> | 99.9%     | 99.8%     | 99.9%     | 99.9%     | 99.9%     | 99.9%   | 0.0%               |
|          |          | <b>2</b> | 99.8%     | 99.8%     | 99.8%     | 99.7%     | 99.8%     | 99.8%   | 0.1%               |
| dC8:N4H  | dG25:O6  | <b>1</b> | 98.6%     | 98.3%     | 98.6%     | 98.5%     | 98.6%     | 98.5%   | 0.1%               |
|          |          | <b>2</b> | 98.8%     | 99.1%     | 99.0%     | 98.9%     | 99.0%     | 98.9%   | 0.1%               |
| dG23:N1H | dC10:N3  | <b>1</b> | 99.9%     | 99.9%     | 99.9%     | 99.8%     | 99.9%     | 99.9%   | 0.0%               |
|          |          | <b>2</b> | 99.9%     | 100.0%    | 100.0%    | 99.9%     | 99.9%     | 100.0%  | 0.0%               |
| dG23:N2H | dC10:O2  | <b>1</b> | 99.5%     | 99.7%     | 99.7%     | 99.8%     | 99.7%     | 99.7%   | 0.1%               |
|          |          | <b>2</b> | 99.5%     | 99.2%     | 99.4%     | 99.4%     | 99.4%     | 99.4%   | 0.1%               |
| dC10:N4H | dG23:O6  | <b>1</b> | 98.3%     | 98.5%     | 98.3%     | 98.3%     | 98.5%     | 98.4%   | 0.1%               |
|          |          | <b>2</b> | 99.1%     | 99.2%     | 98.9%     | 99.0%     | 99.0%     | 99.0%   | 0.1%               |

<sup>a</sup>Hydrogen-bonding interactions were determined using a cut-off of 3.4 Å for the donor–acceptor distance and 120° for the donor–hydrogen–acceptor angle.

Table B.14: Occupancies (%) of the hydrogen bonds at the lesion site containing the lesion pair in the S conformation for the FP models. <sup>a</sup>

| Donor       | Acceptor  | Adduct   | Replica 1 | Replica 2 | Replica 3 | Replica 4 | Replica 5 | Average | Standard Deviation |
|-------------|-----------|----------|-----------|-----------|-----------|-----------|-----------|---------|--------------------|
| dC24:NH4    | dG23:OP2  | <b>1</b> | 5.8%      | 31.2%     | 16.1%     | 10.4%     | 14.8%     | 15.7%   | 8.5%               |
| Lesion:N17H | Lesion:O6 | <b>1</b> | 96.2%     | 95.7%     | 95.0%     | 95.9%     | 94.6%     | 95.5%   | 0.6%               |
| Lesion:N10H | dG25:N2   | <b>2</b> | 44.7%     | 38.5%     | 40.7%     | 38.2%     | 26.8%     | 37.8%   | 5.9%               |
| dC24:NH4    | Lesion:O6 | <b>2</b> | 56.7%     | 43.9%     | 45.2%     | 42.8%     | 16.6%     | 41.0%   | 13.2%              |

<sup>a</sup>Hydrogen-bonding interactions were determined using a cut-off of 3.4 Å for the donor–acceptor distance and 120° for the donor–hydrogen–acceptor angle.

Table B.15: Occupancies (%) of the hydrogen bonds of the flanking bases when the helix containing a lesion is started from the S conformation for the FP models. <sup>a</sup>

| Donor    | Acceptor | Adduct   | Replica 1 | Replica 2 | Replica 3 | Replica 4 | Replica 5 | Average | Standard Deviation |
|----------|----------|----------|-----------|-----------|-----------|-----------|-----------|---------|--------------------|
| dG25:N1H | dC8:N3   | <b>1</b> | 100.0%    | 100.0%    | 100.0%    | 99.9%     | 100.0%    | 100.0%  | 0.0%               |
|          |          | <b>2</b> | 100.0%    | 99.9%     | 100.0%    | 100.0%    | 100.0%    | 100.0%  | 0.0%               |
| dG25:N2H | dC8:O2   | <b>1</b> | 99.9%     | 99.9%     | 99.9%     | 99.9%     | 99.9%     | 99.9%   | 0.0%               |
|          |          | <b>2</b> | 100.0%    | 100.0%    | 99.9%     | 100.0%    | 100.0%    | 99.9%   | 0.0%               |
| dC8:N4H  | dG25:O6  | <b>1</b> | 98.8%     | 98.9%     | 98.9%     | 99.1%     | 98.9%     | 98.9%   | 0.1%               |
|          |          | <b>2</b> | 99.2%     | 99.0%     | 99.2%     | 98.7%     | 99.2%     | 99.1%   | 0.2%               |
| dG23:N1H | dC10:N3  | <b>1</b> | 99.7%     | 99.6%     | 99.7%     | 99.7%     | 99.7%     | 99.7%   | 0.0%               |
|          |          | <b>2</b> | 93.7%     | 94.3%     | 74.8%     | 69.9%     | 98.3%     | 86.2%   | 11.5%              |
| dG23:N2H | dC10:O2  | <b>1</b> | 99.9%     | 99.8%     | 99.9%     | 99.9%     | 99.8%     | 99.9%   | 0.0%               |
|          |          | <b>2</b> | 94.1%     | 97.2%     | 75.4%     | 70.4%     | 99.7%     | 87.4%   | 12.0%              |
| dC10:N4H | dG23:O6  | <b>1</b> | 97.3%     | 97.4%     | 97.1%     | 97.4%     | 96.5%     | 97.1%   | 0.3%               |
|          |          | <b>2</b> | 92.0%     | 91.9%     | 73.2%     | 67.7%     | 92.2%     | 83.4%   | 10.7%              |

<sup>a</sup>Hydrogen-bonding interactions were determined using a cut-off of 3.4 Å for the donor–acceptor distance and 120° for the donor–hydrogen–acceptor angle.

Table B.16: Occupancies (%) of the hydrogen bonds at the lesion site containing the lesion pair in the W conformation for the FP models. <sup>a</sup>

| Donor       | Acceptor  | Adduct   | Replica 1 | Replica 2 | Replica 3 | Replica 4 | Replica 5 | Average | Standard Deviation |
|-------------|-----------|----------|-----------|-----------|-----------|-----------|-----------|---------|--------------------|
| Lesion:N17H | Lesion:O6 | <b>1</b> | 94.4%     | 97.1%     | 95.1%     | 94.8%     | 98.3%     | 95.9%   | 1.5%               |
| dC24:NH4    | dG23:OP2  | <b>1</b> | 21.6%     | 2.2%      | 38.0%     | 21.2%     | 0.0%      | 16.6%   | 14.0%              |
| dC24:N4H    | Lesion:O6 | <b>2</b> | 38.6%     | 63.2%     | 51.1%     | 38.4%     | 47.0%     | 47.7%   | 9.2%               |
| dC24:N4H    | Lesion:N7 | <b>2</b> | 37.8%     | 60.9%     | 49.0%     | 37.4%     | 48.3%     | 46.7%   | 8.7%               |
| Lesion:N10H | dC10:O4'  | <b>2</b> | 24.2%     | 37.3%     | 41.7%     | 45.1%     | 43.5%     | 38.4%   | 7.5%               |
| dC24:N4H    | dC10:O2   | <b>2</b> | 10.0%     | 8.3%      | 18.2%     | 32.5%     | 22.4%     | 18.3%   | 8.8%               |

<sup>a</sup>Hydrogen-bonding interactions were determined using a cut-off of 3.4 Å for the donor–acceptor distance and 120° for the donor–hydrogen–acceptor angle.

Table B.17: Occupancies (%) of the hydrogen bonds of the flanking bases when the helix containing a lesion is started from the W conformation for the FP models. <sup>a</sup>

| Donor    | Acceptor | Adduct   | Replica 1 | Replica 2 | Replica 3 | Replica 4 | Replica 5 | Average | Standard Deviation |
|----------|----------|----------|-----------|-----------|-----------|-----------|-----------|---------|--------------------|
| dG25:N1H | dC8:N3   | <b>1</b> | 100.0%    | 99.9%     | 100.0%    | 100.0%    | 99.9%     | 100.0%  | 0.0%               |
|          |          | <b>2</b> | 99.9%     | 99.9%     | 99.9%     | 99.9%     | 99.9%     | 99.9%   | 0.0%               |
| dG25:N2H | dC8:O2   | <b>1</b> | 99.9%     | 99.9%     | 99.8%     | 99.9%     | 99.9%     | 99.9%   | 0.0%               |
|          |          | <b>2</b> | 99.7%     | 99.6%     | 99.7%     | 99.5%     | 99.8%     | 99.6%   | 0.1%               |
| dC8:N4H  | dG25:O6  | <b>1</b> | 99.1%     | 98.8%     | 98.6%     | 99.2%     | 98.4%     | 98.8%   | 0.3%               |
|          |          | <b>2</b> | 98.7%     | 99.0%     | 98.5%     | 98.4%     | 98.6%     | 98.6%   | 0.2%               |
| dG23:N1H | dC10:N3  | <b>1</b> | 99.5%     | 99.8%     | 99.8%     | 99.6%     | 99.6%     | 99.7%   | 0.1%               |
|          |          | <b>2</b> | 99.0%     | 99.6%     | 99.3%     | 99.5%     | 99.5%     | 99.4%   | 0.2%               |
| dG23:N2H | dC10:O2  | <b>1</b> | 99.8%     | 99.8%     | 99.9%     | 99.8%     | 99.7%     | 99.8%   | 0.1%               |
|          |          | <b>2</b> | 98.7%     | 97.4%     | 97.1%     | 97.8%     | 97.9%     | 97.8%   | 0.5%               |
| dC10:N4H | dG23:O6  | <b>1</b> | 96.4%     | 98.2%     | 98.1%     | 97.0%     | 98.7%     | 97.7%   | 0.8%               |
|          |          | <b>2</b> | 97.2%     | 98.2%     | 98.4%     | 98.3%     | 98.2%     | 98.1%   | 0.5%               |

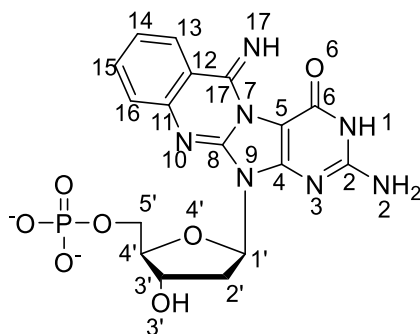
<sup>a</sup>Hydrogen-bonding interactions were determined using a cut-off of 3.4 Å for the donor–acceptor distance and 120° for the donor–hydrogen–acceptor angle.

Table B.18: Average and standard deviations (in parenthesis) of the interplanar angle (deg) between the aromatic portion of the bulky moiety and the 5' or 3' base-pairs for the FP duplexes with the adduct in the *syn* orientation.<sup>a</sup>

| Direction    | Adduct | Replica 1      | Replica 2      | Replica 3      | Replica 4      | Replica 5      | Average    |
|--------------|--------|----------------|----------------|----------------|----------------|----------------|------------|
| 5' base-pair | 1      | 11.0<br>(12.2) | 9.6 (11.2)     | 10.0<br>(10.4) | 10.1<br>(10.8) | 11.4<br>(13.0) | 10.4 (0.7) |
|              | 2      | 56.7<br>(20.7) | 50.9<br>(23.5) | 53.9<br>(21.0) | 60.9<br>(19.1) | 37.8<br>(23.8) | 52.0 (7.8) |
| 3' base-pair | 1      | 16.1<br>(13.1) | 13.6 (9.3)     | 14.0<br>(11.3) | 14.7<br>(11.5) | 15.0<br>(12.1) | 14.7 (0.9) |
|              | 2      | 55.2<br>(20.5) | 59.7<br>(23.6) | 63.9<br>(18.9) | 66.9<br>(17.5) | 49.1<br>(27.3) | 58.9 (6.3) |
| Both         | 1      | 13.6<br>(12.9) | 11.6<br>(10.5) | 12.0<br>(11.1) | 12.4<br>(11.4) | 13.2<br>(12.7) | 12.5 (0.7) |
|              | 2      | 55.9<br>(20.6) | 55.3<br>(24.0) | 58.9<br>(20.6) | 63.9<br>(18.5) | 43.4<br>(26.2) | 55.5 (6.8) |

<sup>a</sup>Interplanar angle calculated between the ring atoms of either the 5' or 3' flanking bases and the aromatic rings of the bulky group using an in-house script.

Table B.19. MOL2 file containing partial charges and atom types of the **1** adduct used for MD simulations (atom numbers are provided in the associated figure).

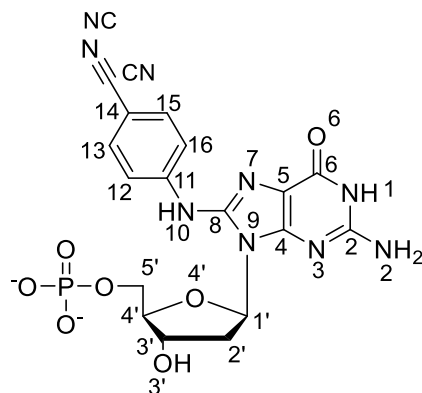


|    |      |       |       |        |    |   |     |         |
|----|------|-------|-------|--------|----|---|-----|---------|
| 1  | P    | 5.238 | 2.263 | 0.407  | P  | 1 | LIG | 1.2091  |
| 2  | O5'  | 3.926 | 1.364 | 0.000  | OS | 1 | LIG | -0.4826 |
| 3  | OP1  | 4.795 | 3.635 | 0.696  | O2 | 1 | LIG | -0.7854 |
| 4  | OP2  | 6.043 | 1.432 | 1.314  | O2 | 1 | LIG | -0.7966 |
| 5  | O3'  | 0.000 | 0.000 | 0.000  | OS | 1 | LIG | -0.5525 |
| 6  | C5'  | 2.896 | 1.974 | -0.733 | CI | 1 | LIG | -0.0043 |
| 7  | H5'  | 2.408 | 2.742 | -0.135 | H1 | 1 | LIG | 0.0792  |
| 8  | H5'' | 3.282 | 2.438 | -1.635 | H1 | 1 | LIG | 0.0792  |
| 9  | C4'  | 1.855 | 0.949 | -1.131 | CT | 1 | LIG | 0.1315  |
| 10 | H4'  | 0.980 | 1.480 | -1.489 | H1 | 1 | LIG | 0.1007  |
| 11 | O4'  | 2.356 | 0.135 | -2.168 | OS | 1 | LIG | -0.3713 |

|    |      |        |        |        |    |   |     |         |
|----|------|--------|--------|--------|----|---|-----|---------|
| 12 | C1'  | 1.961  | -1.189 | -1.952 | CT | 1 | LIG | 0.1256  |
| 13 | H1'  | 0.969  | -1.356 | -2.338 | H2 | 1 | LIG | 0.0846  |
| 14 | C3'  | 1.408  | 0.000  | 0.000  | CT | 1 | LIG | 0.1727  |
| 15 | H3'  | 1.784  | 0.318  | 0.964  | H1 | 1 | LIG | 0.0582  |
| 16 | C2'  | 1.987  | -1.344 | -0.436 | CT | 1 | LIG | -0.0313 |
| 17 | H2'  | 3.005  | -1.441 | -0.088 | HC | 1 | LIG | 0.0330  |
| 18 | H2'' | 1.400  | -2.188 | -0.092 | HC | 1 | LIG | 0.0330  |
| 19 | N9   | 2.814  | -2.077 | -2.709 | N* | 1 | LIG | -0.0620 |
| 20 | C8   | 2.351  | -2.810 | -3.788 | CB | 1 | LIG | 0.2835  |
| 21 | N7   | 3.431  | -3.415 | -4.375 | N* | 1 | LIG | -0.0363 |
| 22 | C5   | 4.589  | -3.029 | -3.648 | CB | 1 | LIG | -0.2219 |
| 23 | C6   | 5.978  | -3.308 | -3.781 | C  | 1 | LIG | 0.5360  |
| 24 | O6   | 6.588  | -3.978 | -4.572 | O  | 1 | LIG | -0.5808 |
| 25 | N1   | 6.711  | -2.644 | -2.778 | NA | 1 | LIG | -0.4165 |
| 26 | H1   | 7.699  | -2.753 | -2.866 | H  | 1 | LIG | 0.3444  |
| 27 | C2   | 6.191  | -1.852 | -1.816 | CA | 1 | LIG | 0.6281  |
| 28 | N2   | 7.057  | -1.259 | -0.959 | N2 | 1 | LIG | -0.8567 |
| 29 | H21  | 7.929  | -1.697 | -0.764 | H  | 1 | LIG | 0.3949  |
| 30 | H22  | 6.621  | -0.790 | -0.196 | H  | 1 | LIG | 0.3949  |
| 31 | N3   | 4.927  | -1.617 | -1.704 | NC | 1 | LIG | -0.5009 |
| 32 | C4   | 4.171  | -2.205 | -2.644 | CB | 1 | LIG | 0.3147  |
| 33 | N10  | 1.132  | -2.890 | -4.123 | NC | 1 | LIG | -0.4245 |
| 34 | C17  | 3.279  | -4.257 | -5.500 | CM | 1 | LIG | 0.4418  |
| 35 | N17  | 4.198  | -4.859 | -6.100 | N2 | 1 | LIG | -0.8649 |
| 36 | H17  | 5.115  | -4.706 | -5.728 | H  | 1 | LIG | 0.4607  |
| 37 | C11  | 0.864  | -3.684 | -5.229 | CA | 1 | LIG | 0.1305  |
| 38 | C12  | -0.464 | -3.803 | -5.653 | CA | 1 | LIG | -0.1441 |
| 39 | H12  | -1.221 | -3.271 | -5.108 | HA | 1 | LIG | 0.1073  |
| 40 | C16  | 1.865  | -4.363 | -5.921 | CA | 1 | LIG | -0.0028 |
| 41 | C13  | -0.775 | -4.581 | -6.741 | CA | 1 | LIG | -0.1299 |
| 42 | H13  | -1.799 | -4.665 | -7.059 | HA | 1 | LIG | 0.1477  |
| 43 | C14  | 0.227  | -5.262 | -7.437 | CA | 1 | LIG | -0.2012 |
| 44 | H14  | -0.024 | -5.870 | -8.288 | HA | 1 | LIG | 0.1464  |
| 45 | C15  | 1.533  | -5.151 | -7.027 | CA | 1 | LIG | -0.1279 |
| 46 | H15  | 2.323  | -5.662 | -7.541 | HA | 1 | LIG | 0.1564  |



Table B.20. MOL2 file containing partial charges and atom types of the **2** adduct used for MD simulations (atom numbers are provided in the associated figure).



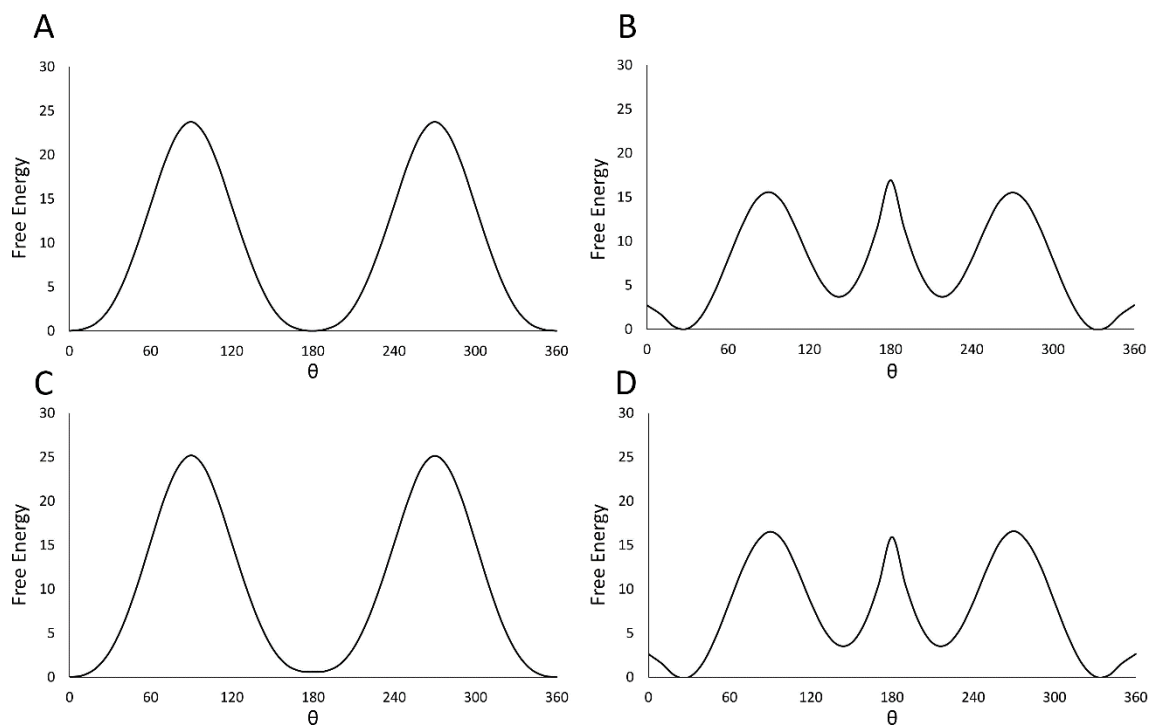
|    |      |       |        |        |    |   |     |         |
|----|------|-------|--------|--------|----|---|-----|---------|
| 1  | P    | 4.929 | 0.561  | -1.290 | P  | 1 | LIG | 1.2091  |
| 2  | O5'  | 4.126 | 1.182  | 0.000  | OS | 1 | LIG | -0.4826 |
| 3  | OP1  | 4.917 | 1.555  | -2.373 | O2 | 1 | LIG | -0.7854 |
| 4  | OP2  | 6.138 | -0.082 | -0.753 | O2 | 1 | LIG | -0.7966 |
| 5  | O3'  | 0.000 | 0.000  | 0.000  | OS | 1 | LIG | -0.5525 |
| 6  | C5'  | 3.014 | 1.997  | -0.235 | CI | 1 | LIG | -0.0043 |
| 7  | H5'  | 2.720 | 2.416  | 0.719  | H1 | 1 | LIG | 0.0792  |
| 8  | H5'' | 3.265 | 2.819  | -0.899 | H1 | 1 | LIG | 0.0792  |
| 9  | C4'  | 1.855 | 1.216  | -0.825 | CT | 1 | LIG | 0.1315  |
| 10 | H4'  | 1.005 | 1.870  | -0.975 | H1 | 1 | LIG | 0.1007  |
| 11 | O4'  | 2.278 | 0.694  | -2.080 | OS | 1 | LIG | -0.3713 |
| 12 | C1'  | 1.896 | -0.657 | -2.200 | CT | 1 | LIG | 0.2759  |
| 13 | H1'  | 0.874 | -0.697 | -2.557 | H2 | 1 | LIG | 0.0289  |
| 14 | C3'  | 1.407 | 0.000  | 0.000  | CT | 1 | LIG | 0.1727  |
| 15 | H3'  | 1.787 | 0.044  | 1.014  | H1 | 1 | LIG | 0.0582  |
| 16 | C2'  | 1.983 | -1.182 | -0.775 | CT | 1 | LIG | -0.0313 |
| 17 | H2'  | 3.013 | -1.341 | -0.497 | HC | 1 | LIG | 0.0330  |
| 18 | H2'' | 1.434 | -2.101 | -0.623 | HC | 1 | LIG | 0.0330  |
| 19 | N9   | 2.741 | -1.278 | -3.195 | N* | 1 | LIG | -0.0566 |
| 20 | C8   | 2.801 | -0.867 | -4.521 | CR | 1 | LIG | 0.2579  |
| 21 | N7   | 3.488 | -1.645 | -5.266 | NB | 1 | LIG | -0.4147 |
| 22 | C5   | 3.912 | -2.651 | -4.421 | CB | 1 | LIG | 0.0338  |
| 23 | C6   | 4.721 | -3.800 | -4.704 | C  | 1 | LIG | 0.4880  |
| 24 | O6   | 5.220 | -4.171 | -5.723 | O  | 1 | LIG | -0.5423 |
| 25 | N1   | 4.910 | -4.556 | -3.524 | NA | 1 | LIG | -0.3889 |
| 26 | H1   | 5.522 | -5.337 | -3.633 | H  | 1 | LIG | 0.3280  |
| 27 | C2   | 4.405 | -4.259 | -2.302 | CA | 1 | LIG | 0.5609  |

|    |     |       |        |         |    |   |     |         |
|----|-----|-------|--------|---------|----|---|-----|---------|
| 28 | N2  | 4.744 | -5.102 | -1.287  | N2 | 1 | LIG | -0.8603 |
| 29 | H21 | 4.900 | -6.061 | -1.505  | H  | 1 | LIG | 0.3910  |
| 30 | H22 | 4.236 | -4.947 | -0.444  | H  | 1 | LIG | 0.3910  |
| 31 | N3  | 3.672 | -3.226 | -2.061  | NC | 1 | LIG | -0.4977 |
| 32 | C4  | 3.469 | -2.445 | -3.152  | CB | 1 | LIG | 0.1794  |
| 33 | N10 | 2.135 | 0.277  | -4.881  | N2 | 1 | LIG | -0.2892 |
| 34 | H10 | 1.983 | 0.919  | -4.135  | H  | 1 | LIG | 0.2402  |
| 35 | C11 | 1.944 | 0.790  | -6.162  | CA | 1 | LIG | 0.1738  |
| 36 | C12 | 1.354 | 2.054  | -6.249  | CA | 1 | LIG | -0.2411 |
| 37 | H12 | 1.090 | 2.582  | -5.350  | HA | 1 | LIG | 0.1622  |
| 38 | C16 | 2.278 | 0.112  | -7.333  | CA | 1 | LIG | -0.0856 |
| 39 | H16 | 2.749 | -0.847 | -7.285  | HA | 1 | LIG | 0.1125  |
| 40 | C13 | 1.100 | 2.629  | -7.472  | CA | 1 | LIG | -0.1466 |
| 41 | H13 | 0.645 | 3.601  | -7.524  | HA | 1 | LIG | 0.1693  |
| 42 | C15 | 2.016 | 0.697  | -8.557  | CA | 1 | LIG | -0.1947 |
| 43 | H15 | 2.278 | 0.171  | -9.456  | HA | 1 | LIG | 0.1870  |
| 44 | C14 | 1.429 | 1.953  | -8.643  | CA | 1 | LIG | 0.0003  |
| 45 | CN  | 1.163 | 2.552  | -9.927  | C1 | 1 | LIG | 0.3527  |
| 46 | NC  | 0.950 | 3.027  | -10.938 | N1 | 1 | LIG | -0.4879 |

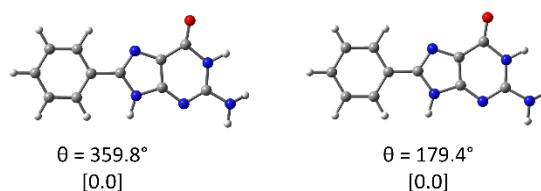
## **Appendix C**

### **Supplementary Information for Chapter 4: Effect of Guanine Adduct Size, Shape and Linker Type on the Conformation of Damaged DNA: A Molecular Dynamics Study**

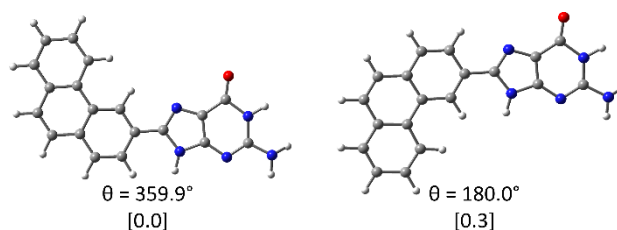
Contains Figures C.1–C.35 and Tables C.1–C.8



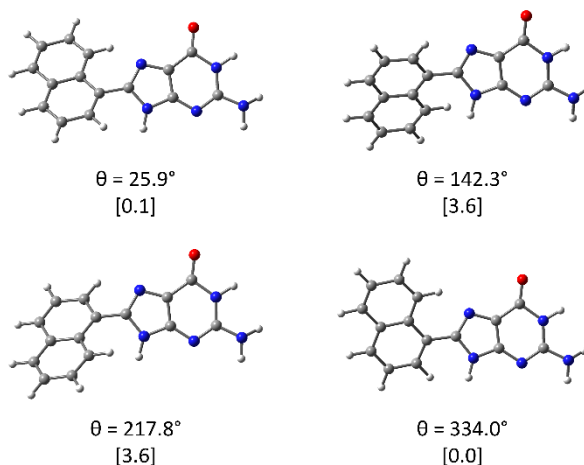
**Figure C.1.** Potential energy surface for the rotation about the  $\theta$  dihedral (deg) in the C-linked (A)  $^{Ph}G$ , (B)  $^{1NPG}$ , (C)  $^{3PHNG}$  and (D)  $^{PyG}$  adducts, calculated at the B3LYP/6-31G(d) level of theory. Energies reported in kJ/mol.



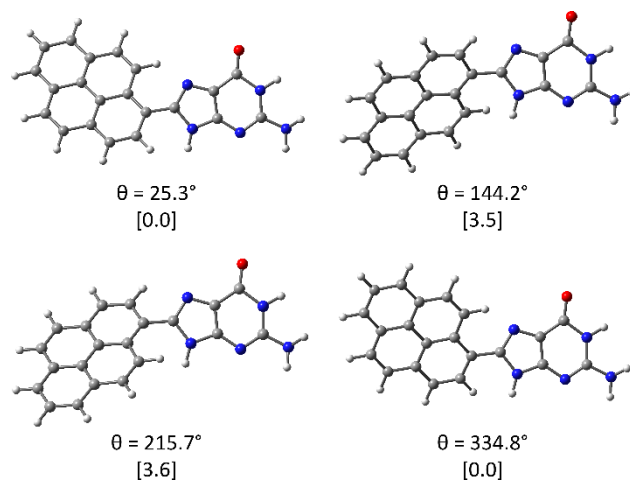
**Figure C.2.** Select B3LYP/6-31G(d) dihedral angle ( $\theta$ ) for the fully optimized minima identified from the 1D for the <sup>Ph</sup>G nucleobase adduct. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311+G(2df,p) relative energies are provided in parentheses (kJ/mol).



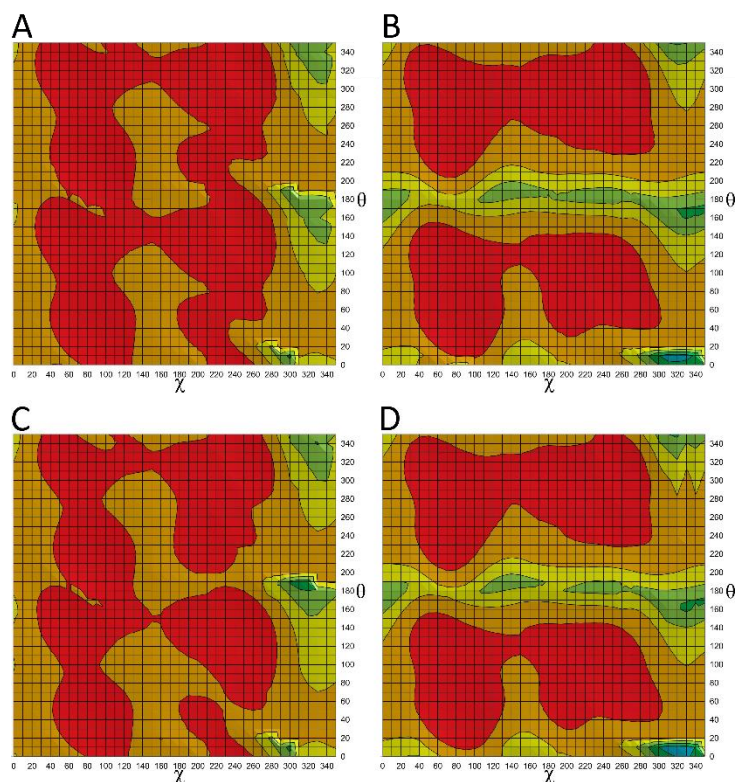
**Figure C.3.** Select B3LYP/6-31G(d) dihedral angle ( $\theta$ ) for the fully optimized minima identified from the 1D for the <sup>3PHN</sup>G nucleobase adduct. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311+G(2df,p) relative energies are provided in parentheses (kJ/mol).



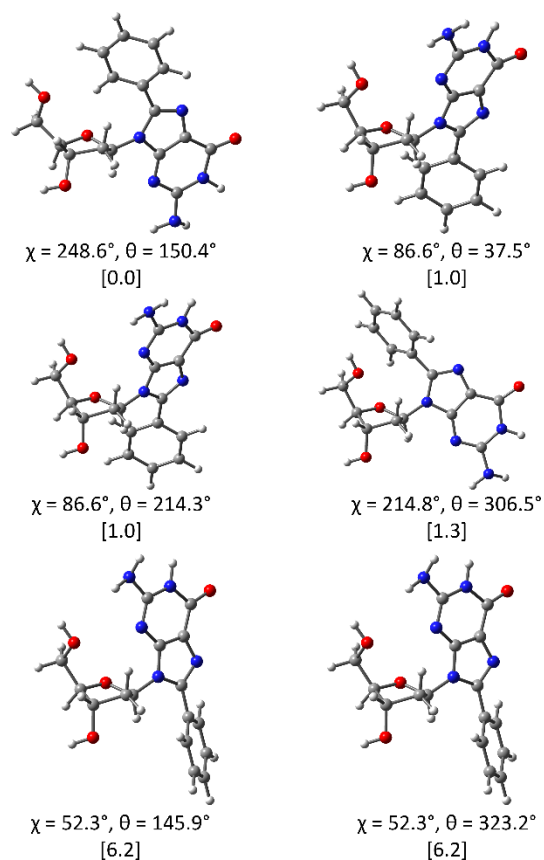
**Figure C.4.** Select B3LYP/6-31G(d) dihedral angle ( $\theta$ ) for the fully optimized minima identified from the 1D for the <sup>1NP</sup>G nucleobase adduct. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311+G(2df,p) relative energies are provided in parentheses (kJ/mol).



**Figure C.5.** Select B3LYP/6-31G(d) dihedral angle ( $\theta$ ) for the fully optimized minima identified from the 1D for the <sup>Py</sup>G nucleobase adduct. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311+G(2df,p) relative energies are provided in parentheses (kJ/mol).

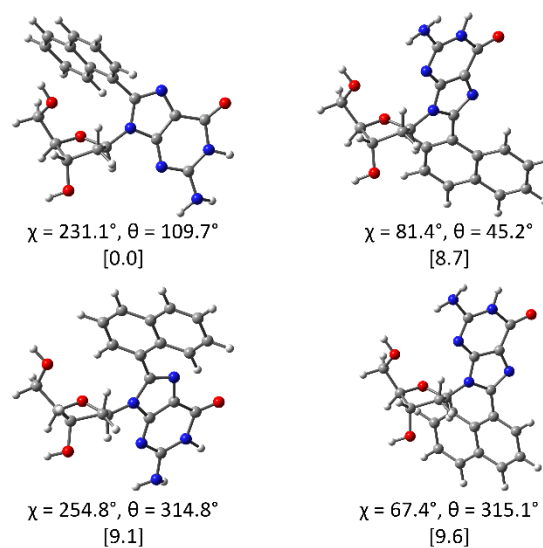


**Figure C.6.** Contour plots of the B3LYP/6-31G(d) potential energy for the (A) <sup>Ph</sup>dG, (B) <sup>1NP</sup>dG, (C) <sup>3PHN</sup>dG and (D) <sup>Py</sup>dG nucleoside adducts with respect to  $\theta$  and  $\chi$  (deg). The relative energy is represented by color; red represents the lowest energy regions, and each change in color in the surrounding regions represents an increase in the energy by 20 kJ/mol.

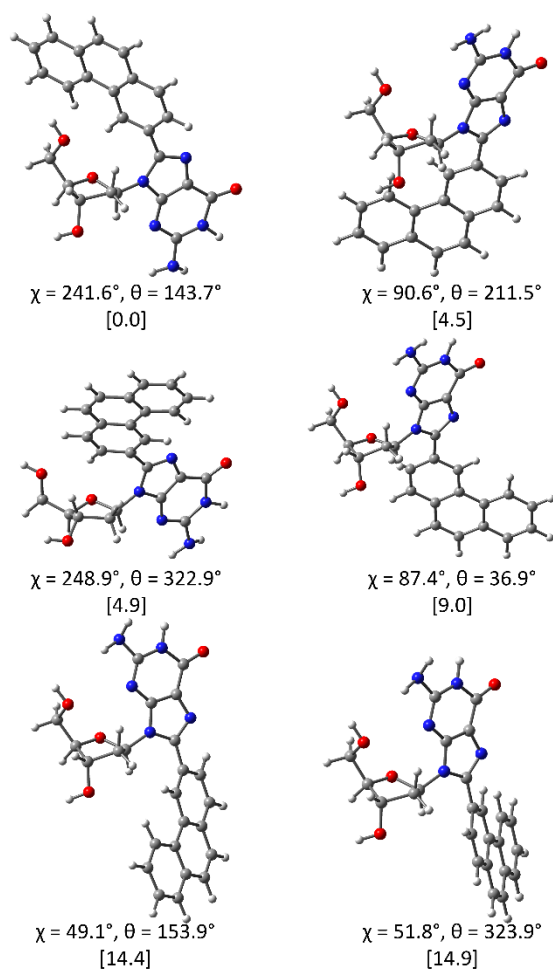


**Figure C.7.** Select B3LYP/6-31G(d) dihedral angles ( $\chi$  and  $\theta$ , deg) for the fully optimized minima identified from the contour plots for the <sup>Ph</sup>dG nucleoside adduct. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311+G(2df,p) relative energies are provided in parentheses (kJ/mol).

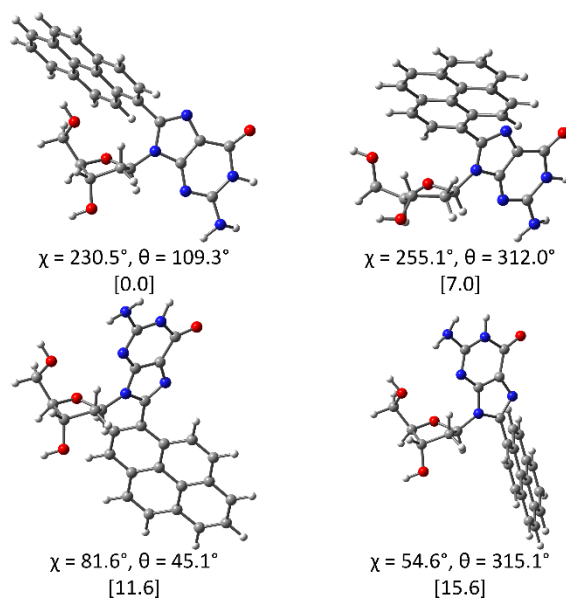




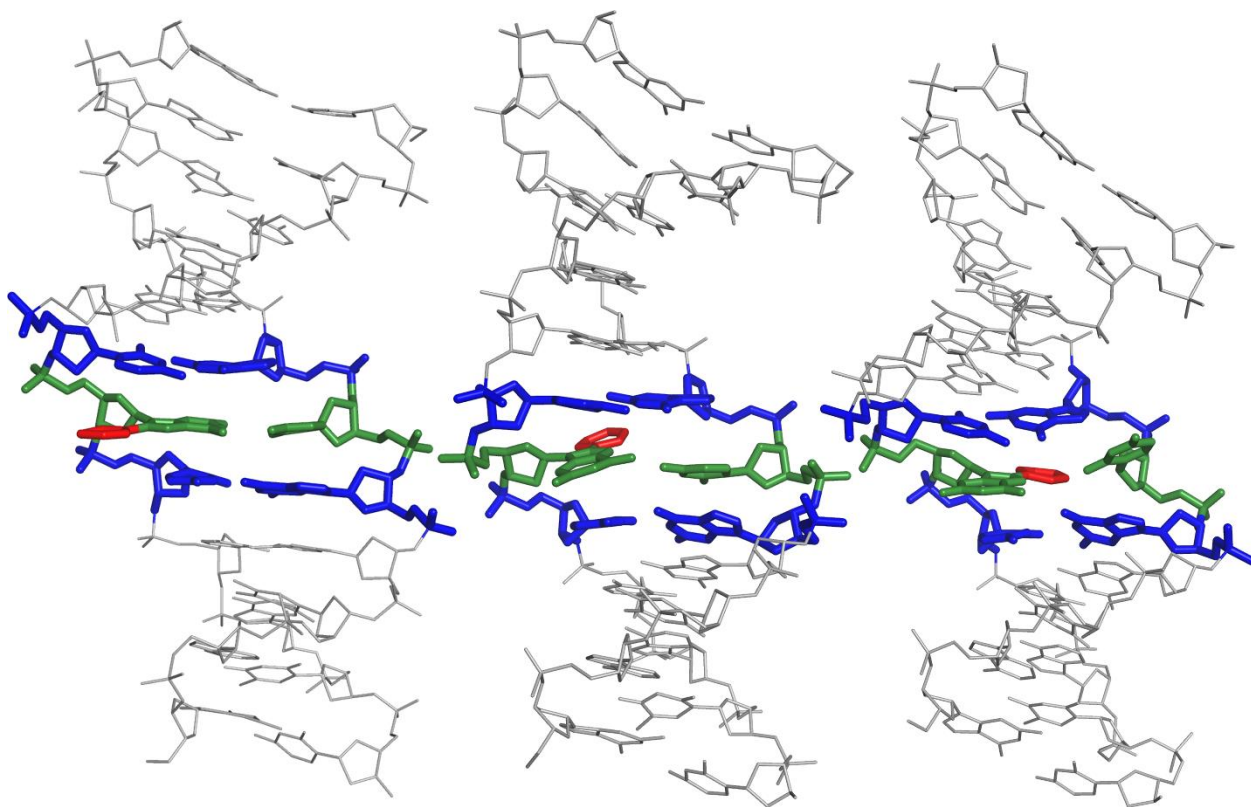
**Figure C.8.** Select B3LYP/6-31G(d) dihedral angles ( $\chi$  and  $\theta$ , deg) for the fully optimized minima identified from the contour plots for the <sup>1</sup>NPdG nucleoside adduct. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311+G(2df,p) relative energies are provided in parentheses (kJ/mol).



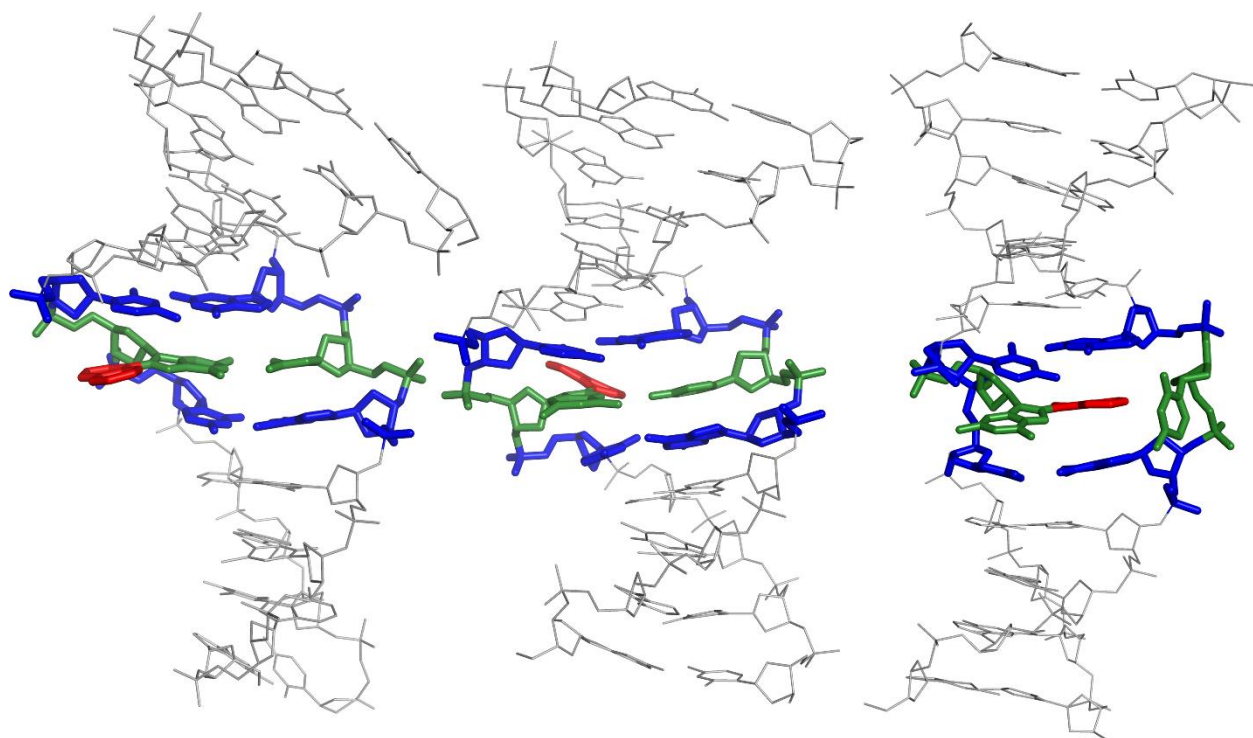
**Figure C.9.** Select B3LYP/6-31G(d) dihedral angles ( $\chi$  and  $\theta$ , deg) for the fully optimized minima identified from the contour plots for the  $^3\text{PHN}$ -dG nucleoside adduct. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311+G(2df,p) relative energies are provided in parentheses (kJ/mol).



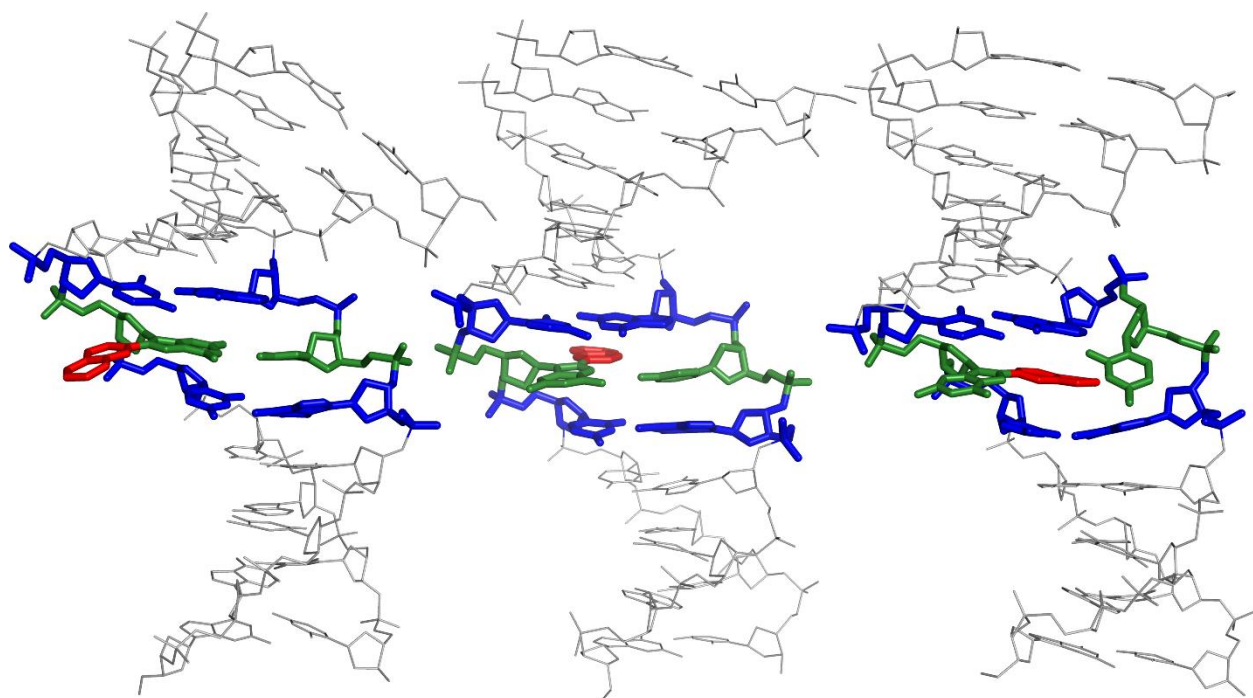
**Figure C.10.** Select B3LYP/6-31G(d) dihedral angles ( $\chi$  and  $\theta$ , deg) for the fully optimized minima identified from the contour plots for the <sup>Py</sup>dG nucleoside adduct. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311+G(2df,p) relative energies are provided in parentheses (kJ/mol).



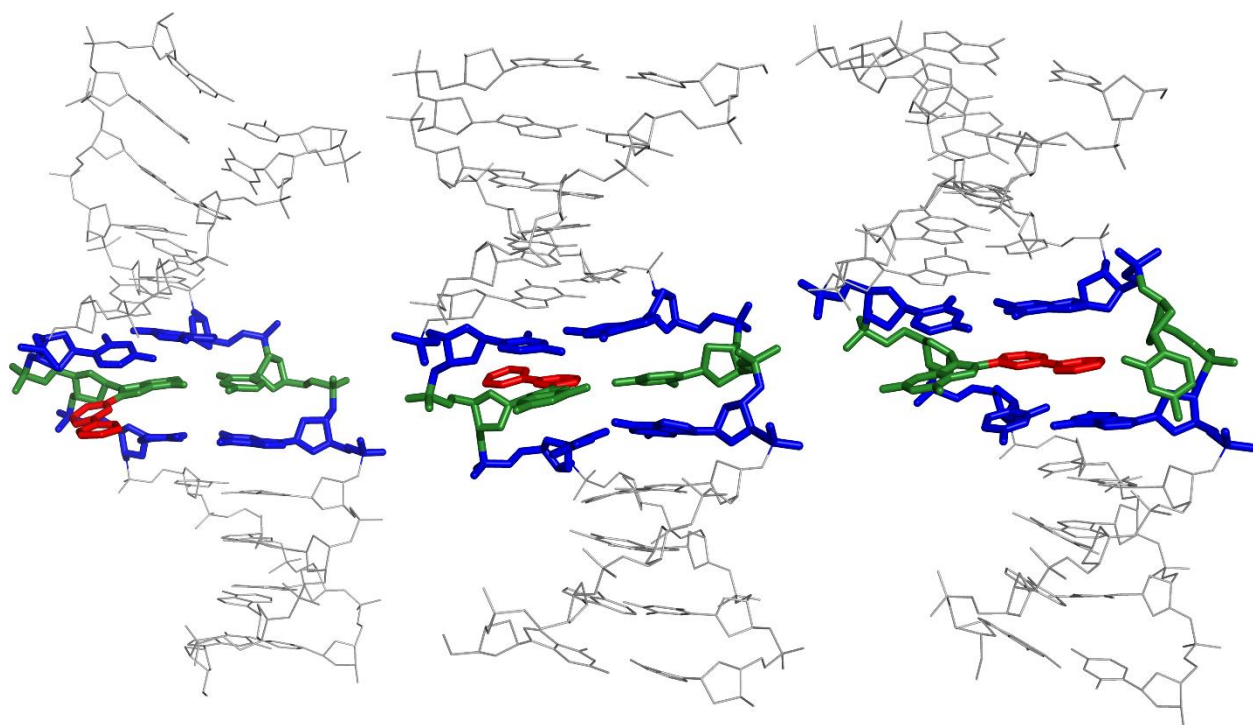
**Figure C.11.** Representative structures obtained from MD simulations of the B (left), W (middle) and S (right) conformations of <sup>Ph</sup>dG adducted DNA. The bulky group is shown in red, the damaged dG and the opposing nucleotide in green, the base pairs flanking the lesion in blue, and rest of the DNA helix in grey.



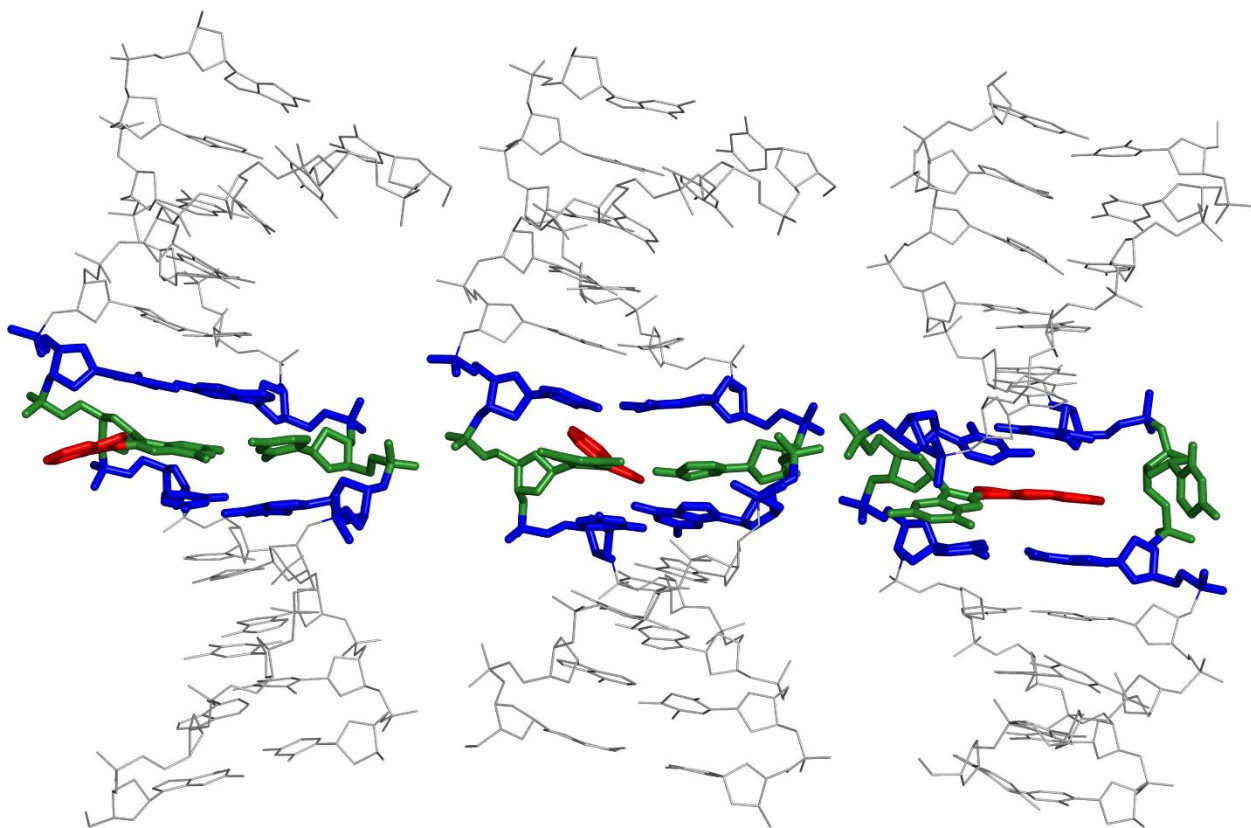
**Figure C.12.** Representative structures obtained from MD simulations of the B (left), W (middle) and S (right) conformations of  $^{1\text{NP}}$ dG adducted DNA. The bulky group is shown in red, the damaged dG and the opposing nucleotide in green, the base pairs flanking the lesion in blue, and rest of the DNA helix in grey.



**Figure C.13.** Representative structures obtained from MD simulations of the B (left), W (middle) and S (right) conformations of  $2^{\text{NP}}$ dG adducted DNA. The bulky group is shown in red, the damaged dG and the opposing nucleotide in green, the base pairs flanking the lesion in blue, and rest of the DNA helix in grey.

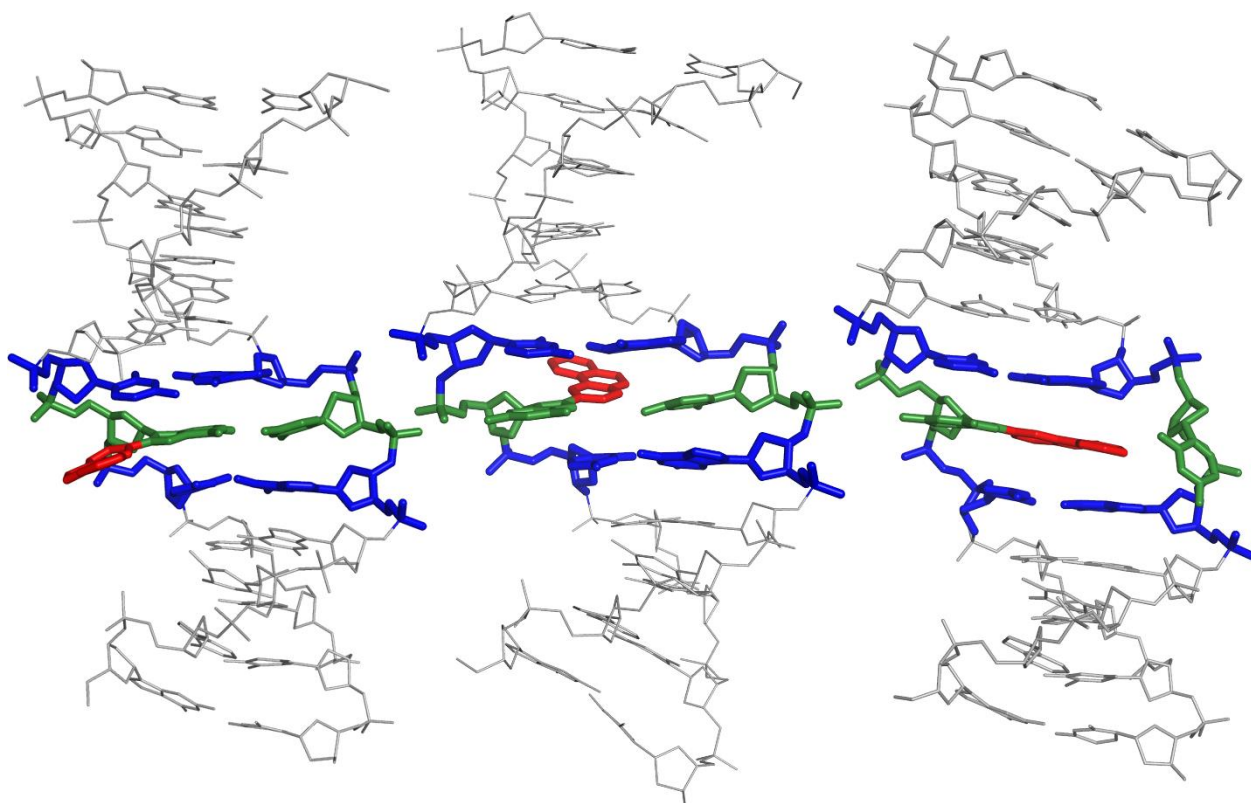


**Figure C.14.** Representative structures obtained from MD simulations of the B (left), W (middle) and S (right) conformations of  $^{3BP}dG$  adducted DNA. The bulky group is shown in red, the damaged dG and the opposing nucleotide in green, the base pairs flanking the lesion in blue, and rest of the DNA helix in grey.

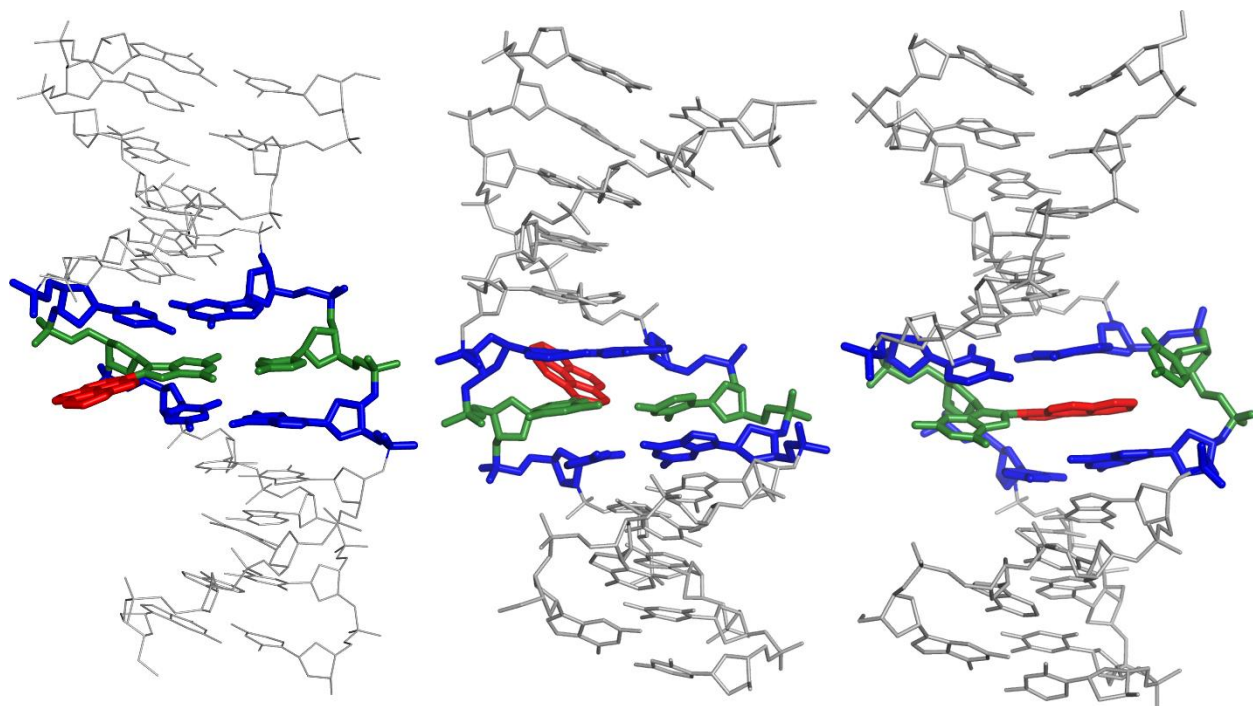


**Figure C.15.** Representative structures obtained from MD simulations of the B (left), W (middle) and S (right) conformations of  $1^{\text{PHN}}$ dG adducted DNA. The bulky group is shown in red, the damaged dG and the opposing nucleotide in green, the base pairs flanking the lesion in blue, and rest of the DNA helix in grey.

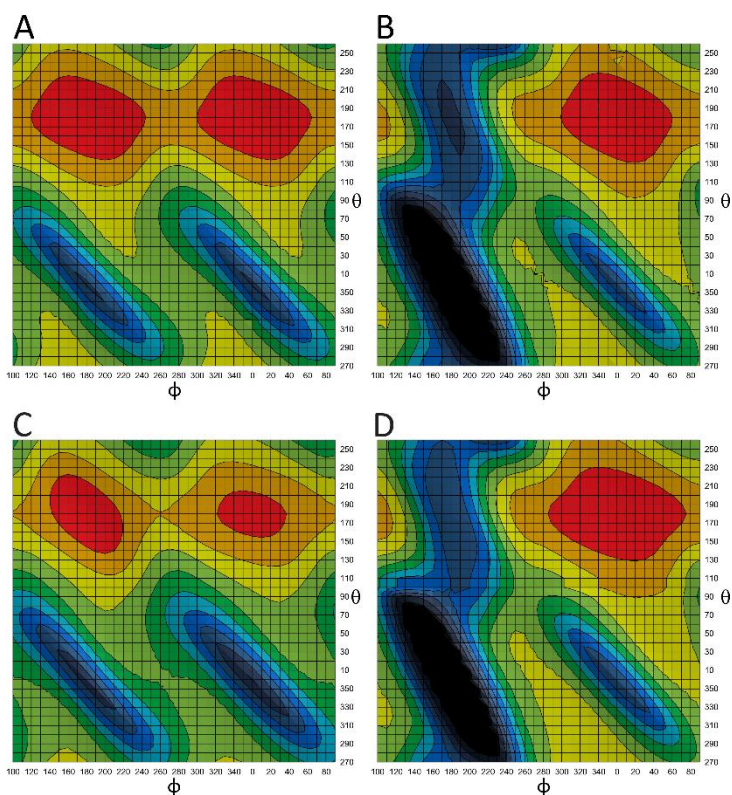




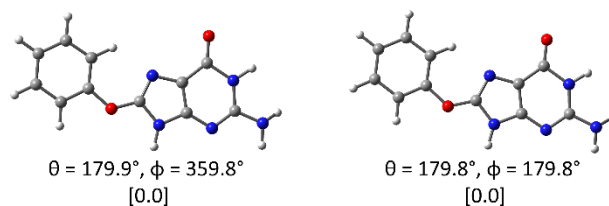
**Figure C.16.** Representative structures obtained from MD simulations of the B (left), W (middle) and S (right) conformations of  $^{3\text{PHN}}$ dG adducted DNA. The bulky group is shown in red, the damaged dG and the opposing nucleotide in green, the base pairs flanking the lesion in blue, and rest of the DNA helix in grey.



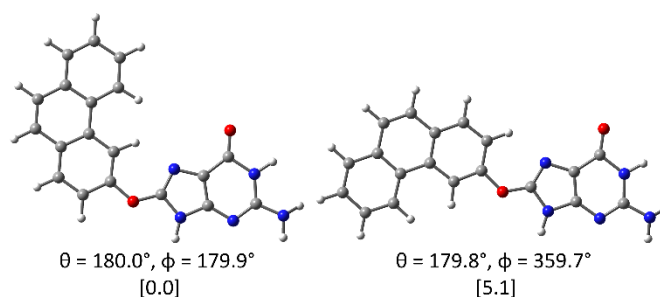
**Figure C.17.** Representative structures obtained from MD simulations of the B (left), W (middle) and S (right) conformations of <sup>P</sup>vdG adducted DNA. The bulky group is shown in red, the damaged dG and the opposing nucleotide in green, the base pairs flanking the lesion in blue, and rest of the DNA helix in grey.



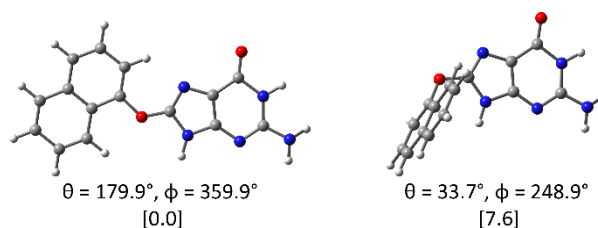
**Figure C.18.** Contour plots of the B3LYP/6-31G(d) potential energy for the (A)  $^{OPh}G$ , (B)  $^{1ONP}G$ , (C)  $^{3OPHN}G$  and (D)  $^{OPy}G$  nucleoside adducts with respect to  $\theta$  and  $\phi$  (deg). The relative energy is represented by color; red represents the lowest energy regions, and each change in color in the surrounding regions represents an increase in the energy by 5 kJ/mol.



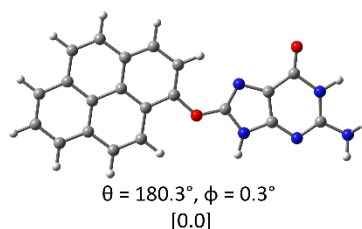
**Figure C.19.** Select B3LYP/6-31G(d) dihedral angles ( $\theta$  and  $\phi$ , deg) for the fully optimized minima identified from the contour plots for the  $^{\text{OPh}}\text{G}$  nucleobase adduct. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311+G(2df,p) relative energies are provided in parentheses (kJ/mol).



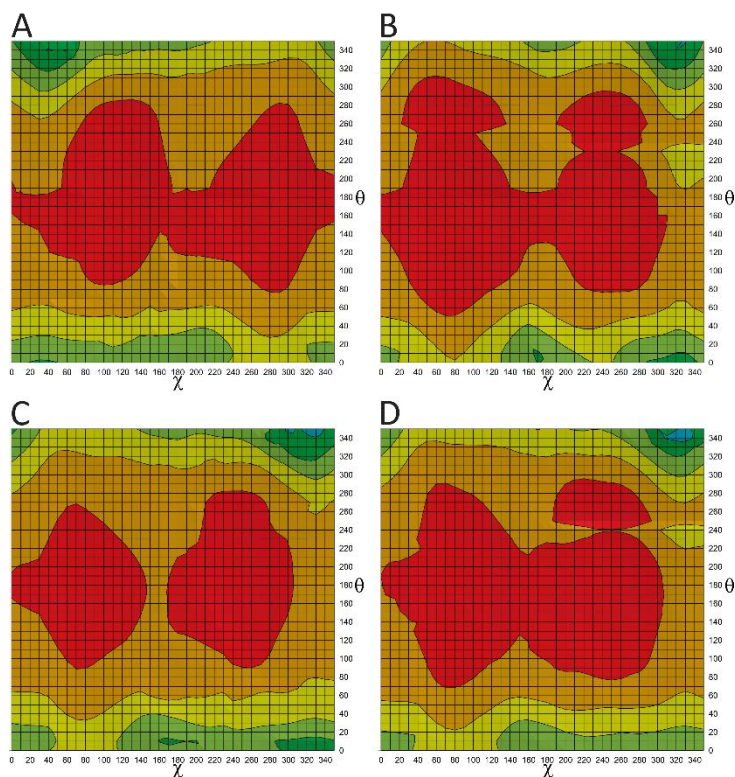
**Figure C.20** Select B3LYP/6-31G(d) dihedral angles ( $\theta$  and  $\phi$ , deg) for the fully optimized minima identified from the contour plots for the  $^{\text{3OPHNu}}\text{G}$  nucleobase adduct. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311+G(2df,p) relative energies are provided in parentheses (kJ/mol).



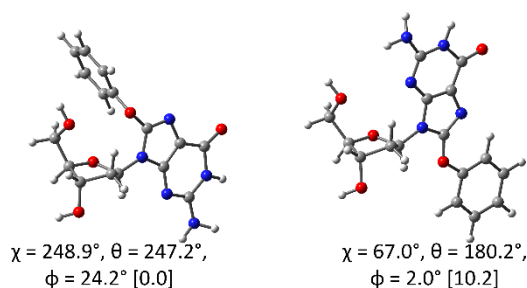
**Figure C.21** Select B3LYP/6-31G(d) dihedral angles ( $\theta$  and  $\phi$ , deg) for the fully optimized minima identified from the contour plots for the  $^{1\text{ONPG}}$  nucleobase adduct. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311+G(2df,p) relative energies are provided in parentheses (kJ/mol).



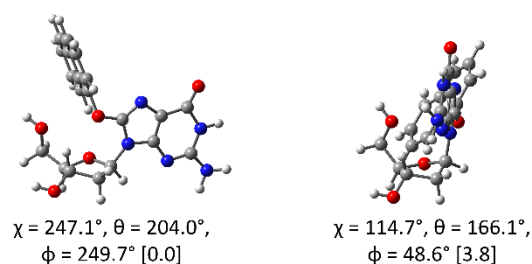
**Figure C.22.** Select B3LYP/6-31G(d) dihedral angles ( $\theta$  and  $\phi$ , deg) for the fully optimized minima identified from the contour plots for the  $^{\text{OPyG}}$  nucleobase adduct. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311+G(2df,p) relative energies are provided in parentheses (kJ/mol).



**Figure C.23.** Contour plots of the B3LYP/6-31G(d) potential energy for the (A)  $^{\text{OPhdG}}$ , (B)  $^{\text{1ONPdG}}$ , (C)  $^{\text{3OPHNdG}}$  and (D)  $^{\text{OPy dG}}$  nucleoside adducts with respect to  $\theta$  and  $\chi$  (deg). The relative energy is represented by color; red represents the lowest energy regions, and each change in color in the surrounding regions represents an increase in the energy by 20 kJ.

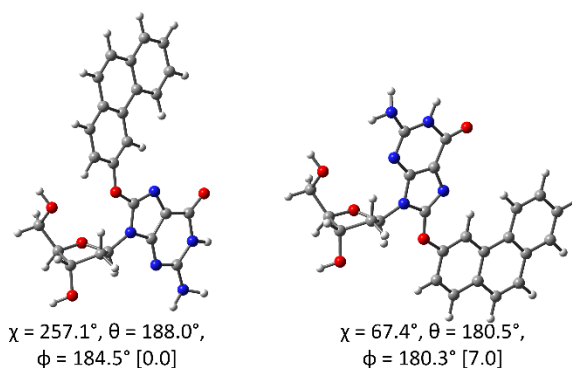


**Figure C.24.** Select B3LYP/6-31G(d) dihedral angles ( $\chi$ ,  $\theta$  and  $\phi$ , deg) for the fully optimized minima identified from the contour plots for the  $^{\text{OP}^{\text{h}}}\text{dG}$  nucleoside adduct. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311+G(2df,p) relative energies are provided in parentheses (kJ/mol).

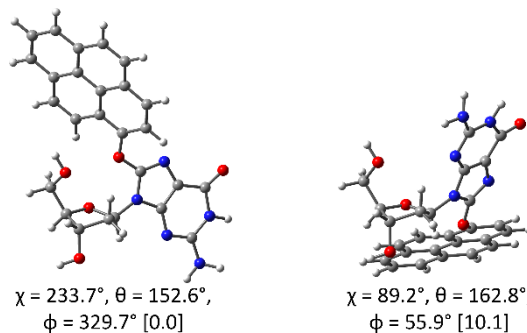


**Figure C.25.** Select B3LYP/6-31G(d) dihedral angles ( $\chi$ ,  $\theta$  and  $\phi$ , deg) for the fully optimized minima identified from the contour plots for the  $^{\text{1ONP}^{\text{d}}}\text{dG}$  nucleoside adduct. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311+G(2df,p) relative energies are provided in parentheses (kJ/mol).



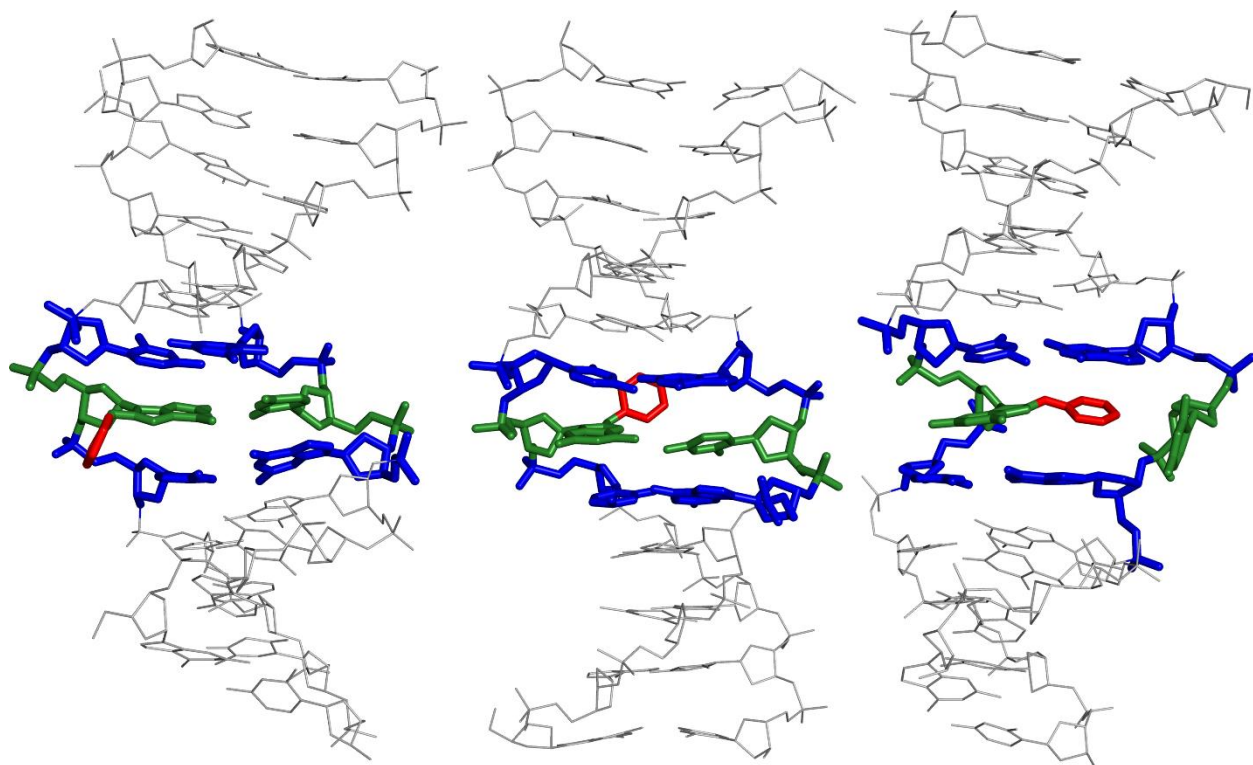


**Figure C.26.** Select B3LYP/6-31G(d) dihedral angles ( $\chi$ ,  $\theta$  and  $\phi$ , deg) for the fully optimized minima identified from the contour plots for the <sup>3OPHN</sup>dG nucleoside adduct. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311+G(2df,p) relative energies are provided in parentheses (kJ/mol).

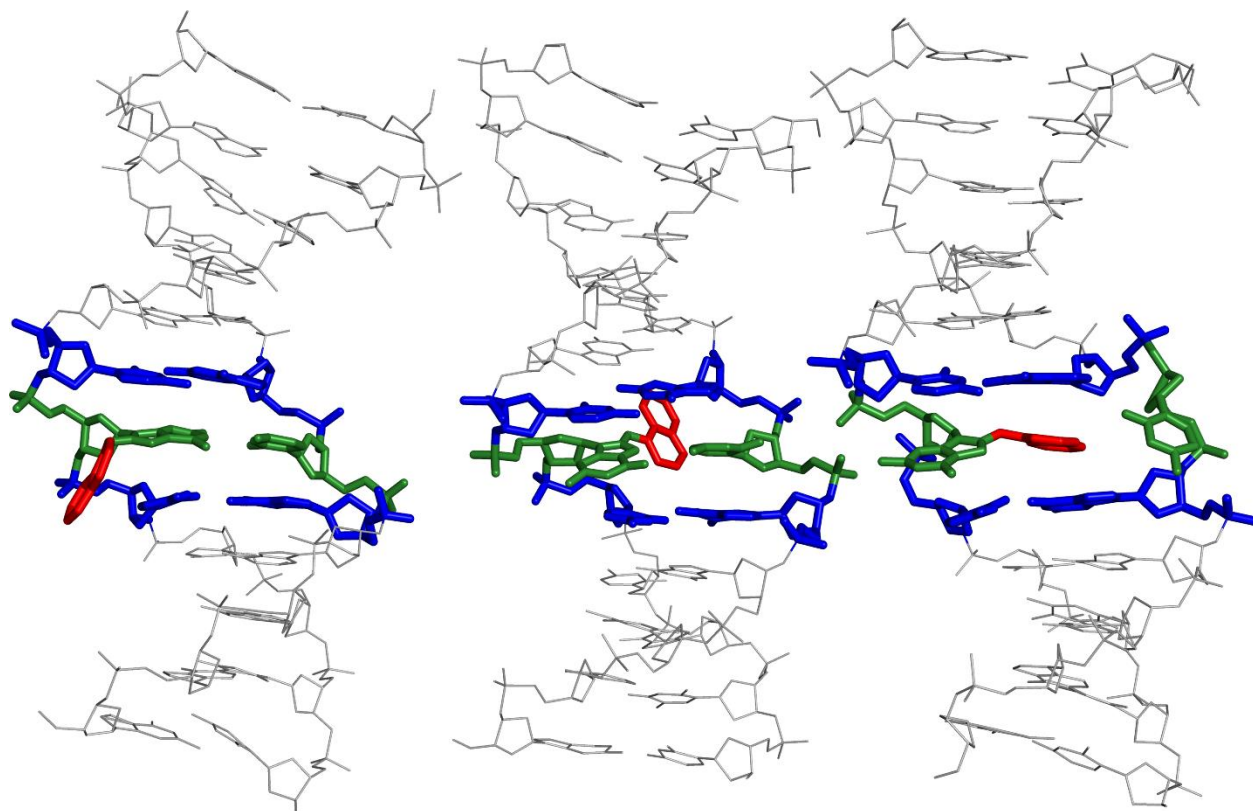


**Figure C.27.** Select B3LYP/6-31G(d) dihedral angles ( $\chi$ ,  $\theta$  and  $\phi$ , deg) for the fully optimized minima identified from the contour plots for the <sup>OPy</sup>dG nucleoside adduct. Zero-point vibrational energy corrected B3LYP-D3(BJ)/6-311+G(2df,p) relative energies are provided in parentheses (kJ/mol).

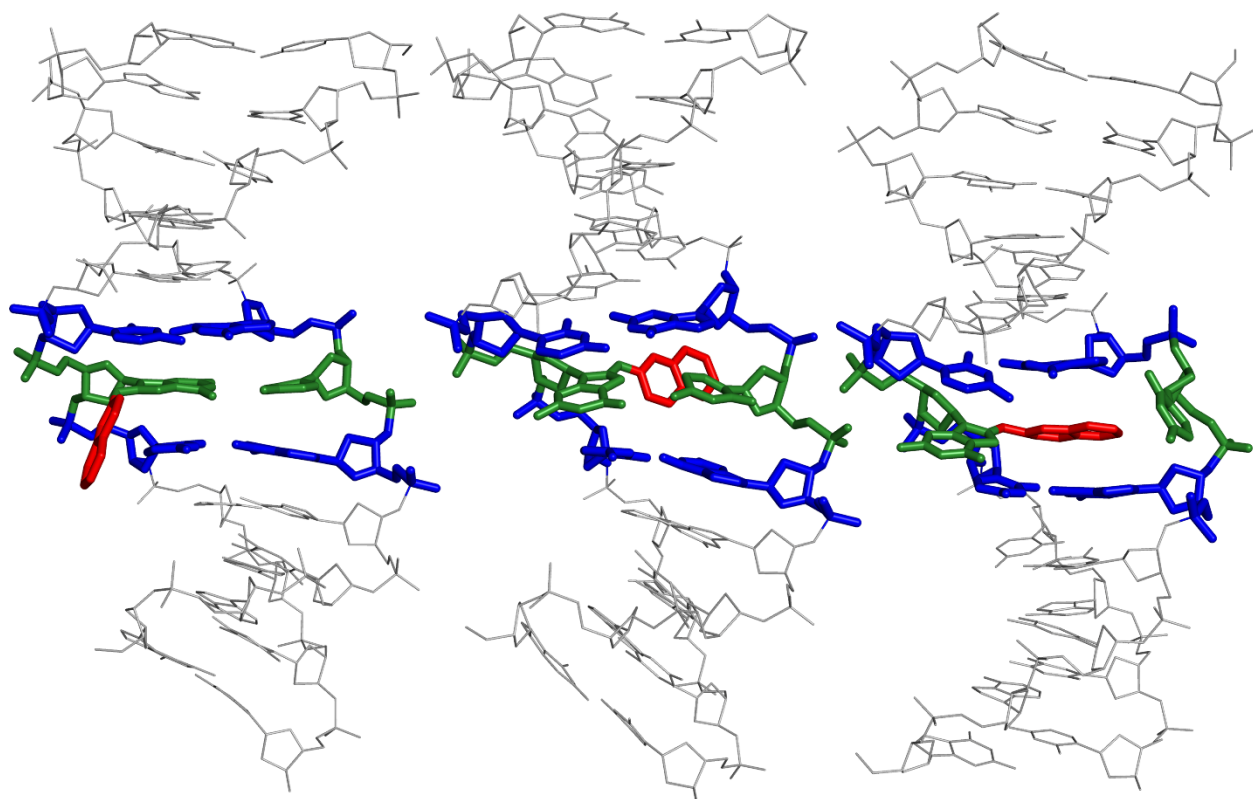




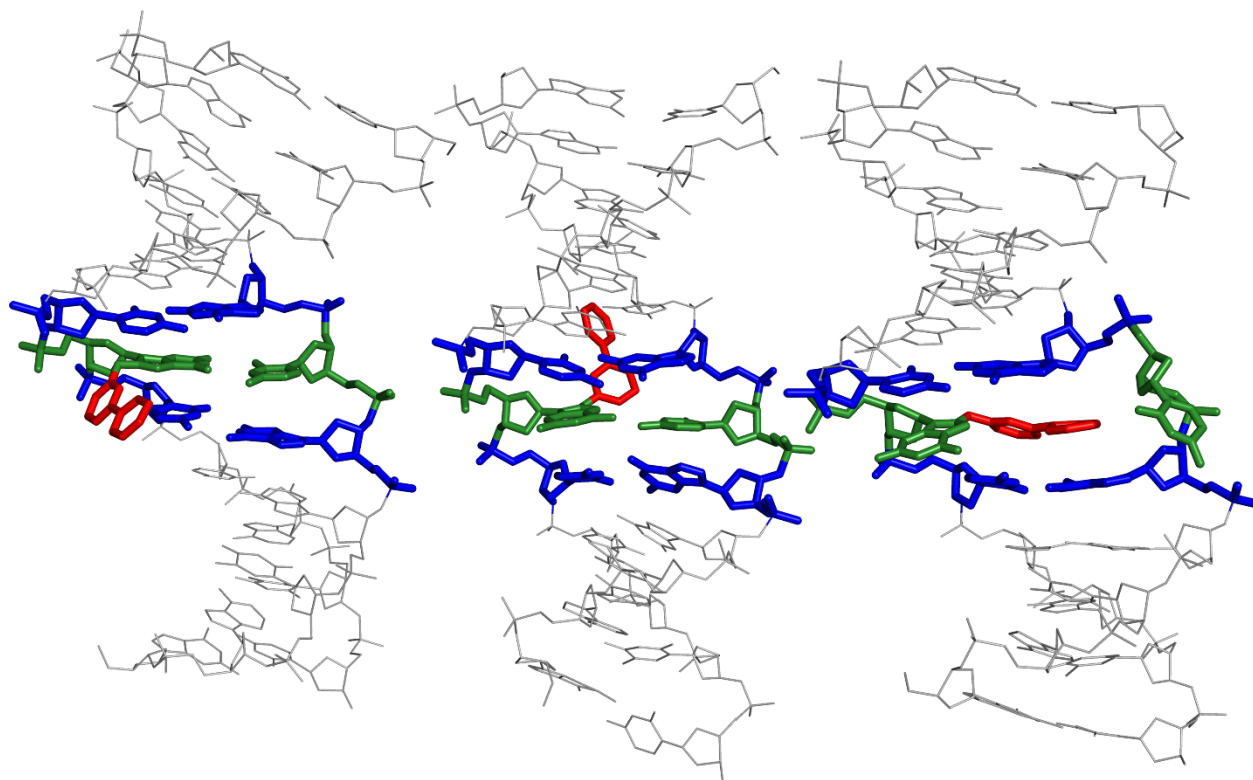
**Figure C.28.** Representative structures obtained from MD simulations of the B (left), W (middle) and S (right) conformations of  $\text{OP}^{\text{h}}\text{dG}$  adducted DNA. The bulky group is shown in red, the damaged dG and the opposing nucleotide in green, the base pairs flanking the lesion in blue, and rest of the DNA helix in grey.



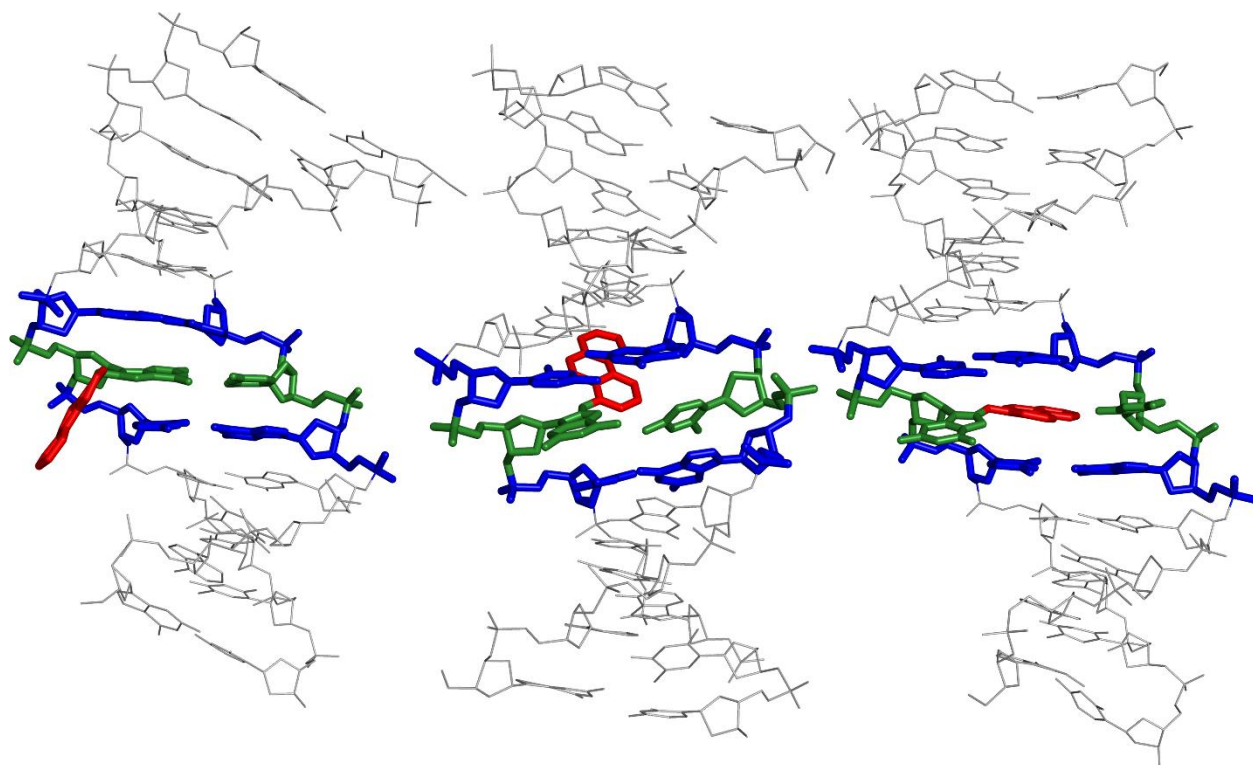
**Figure C.29.** Representative structures obtained from MD simulations of the B (left), W (middle) and S (right) conformations of  $^{1\text{ONP}}\text{dG}$  adducted DNA. The bulky group is shown in red, the damaged dG and the opposing nucleotide in green, the base pairs flanking the lesion in blue, and rest of the DNA helix in grey.



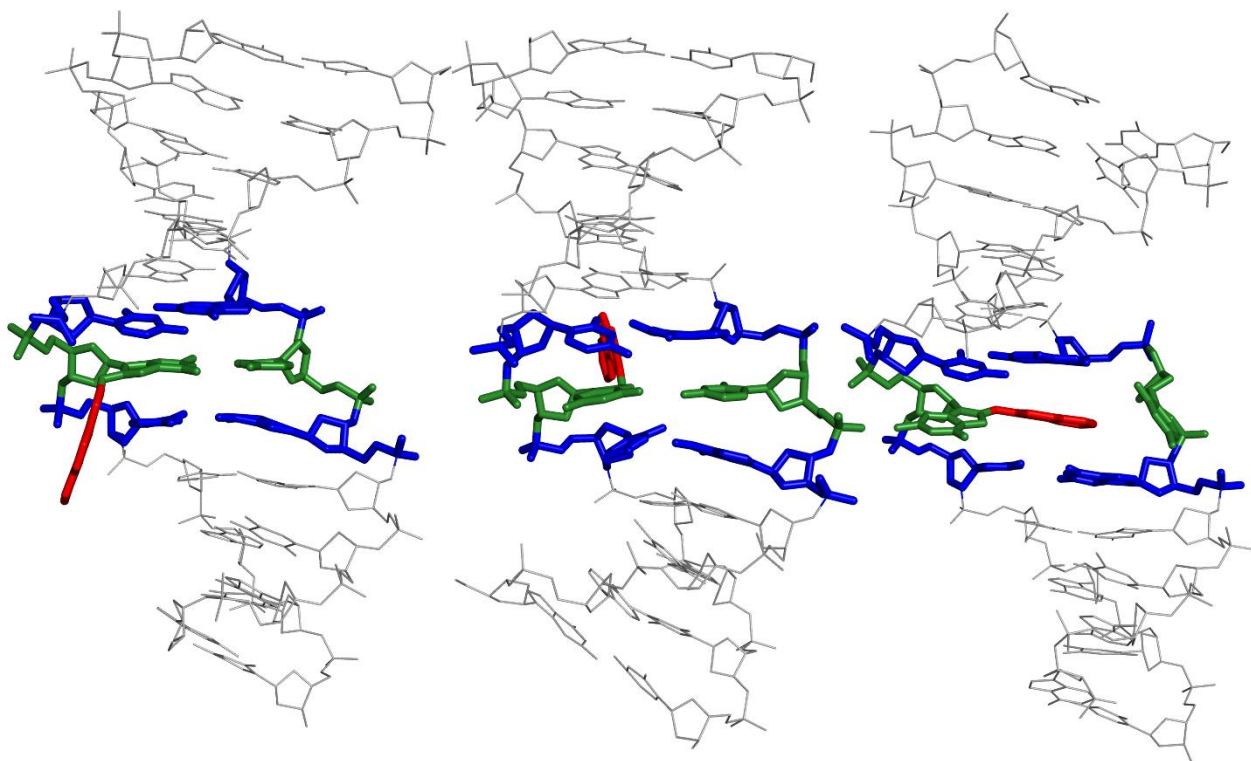
**Figure C.30.** Representative structures obtained from MD simulations of the B (left), W (middle) and S (right) conformations of  $^{2\text{ONP}}$ dG adducted DNA. The bulky group is shown in red, the damaged dG and the opposing nucleotide in green, the base pairs flanking the lesion in blue, and rest of the DNA helix in grey.



**Figure C.31.** Representative structures obtained from MD simulations of the B (left), W (middle) and S (right) conformations of <sup>OBp</sup>dG adducted DNA. The bulky group is shown in red, the damaged dG and the opposing nucleotide in green, the base pairs flanking the lesion in blue, and rest of the DNA helix in grey.

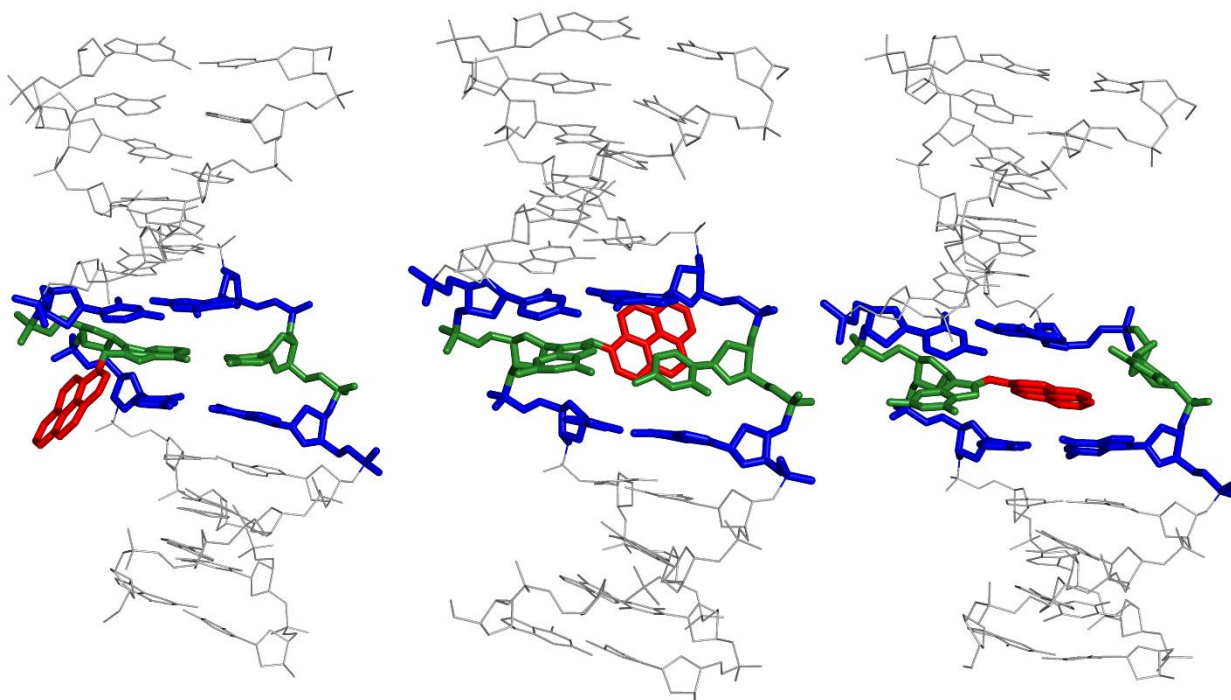


**Figure C.32.** Representative structures obtained from MD simulations of the B (left), W (middle) and S (right) conformations of  $^{10\text{PHN}}$ dG adducted DNA. The bulky group is shown in red, the damaged dG and the opposing nucleotide in green, the base pairs flanking the lesion in blue, and rest of the DNA helix in grey.

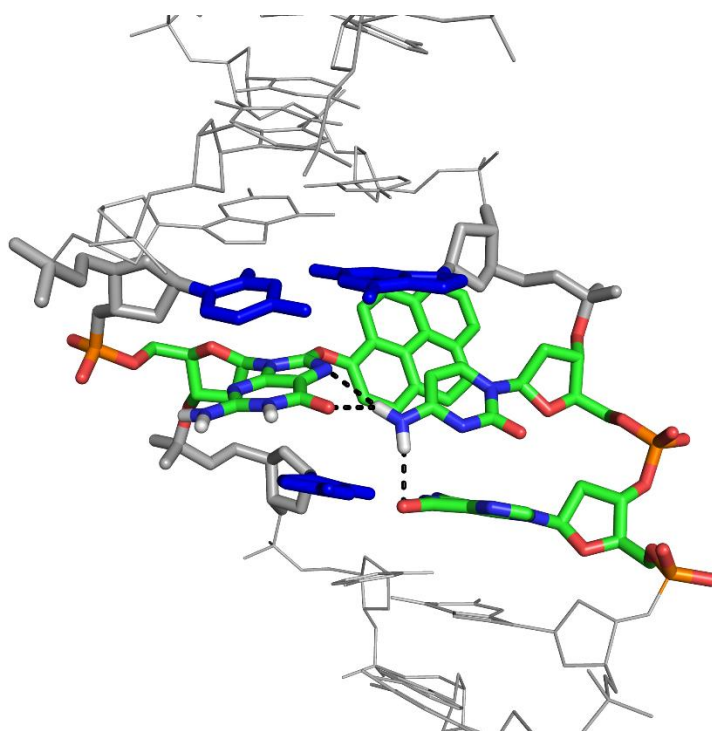


**Figure C.33.** Representative structures obtained from MD simulations of the B (left), W (middle) and S (right) conformations of <sup>3</sup>OPHNdG adducted DNA. The bulky group is shown in red, the damaged dG and the opposing nucleotide in green, the base pairs flanking the lesion in blue, and rest of the DNA helix in grey.





**Figure C.34.** Representative structures obtained from MD simulations of the B (left), W (middle) and S (right) conformations of <sup>OPydg</sup> adducted DNA. The bulky group is shown in red, the damaged dG and the opposing nucleotide in green, the base pairs flanking the lesion in blue, and rest of the DNA helix in grey.



**Figure C.35.** Structure of the unique hydrogen bonding pattern for the lesion site of <sup>OPy</sup>dG in the W conformation.



**Table C.1.** Average occupancies (%) for the Watson-Crick hydrogen bonds within the base pairs present at the 5'– and 3'– terminals and flanking the damage site with respect to the lesion respectively, in three conformations of the adducted DNA.<sup>a</sup>

| Conformation | Adduct                       | Terminal Base Pairs |              | Flanking Base Pairs |             |
|--------------|------------------------------|---------------------|--------------|---------------------|-------------|
|              |                              | 5'-dC1:dG24         | 3'-dC12:dG13 | 5'-dC6:dG19         | 3'-dG8:dC17 |
| B            | <sup>Ph</sup> dG             | 99.6                | 99.4         | 99.4                | 99.5        |
|              | <sup>1NP</sup> dG            | 99.7                | 99.5         | 99.6                | 99.5        |
|              | <sup>2NP</sup> dG            | 99.5                | 99.5         | 99.1                | 99.5        |
|              | <sup>3BP</sup> dG            | 99.6                | 99.5         | 99.1                | 99.5        |
|              | <sup>1PHN</sup> dG           | 99.6                | 99.5         | 99.5                | 99.4        |
|              | <sup>3PHN</sup> dG           | 99.6                | 99.4         | 99.1                | 99.5        |
|              | <sup>Py</sup> dG             | 99.6                | 99.5         | 99.9                | 99.6        |
|              | <sup>O<sup>Ph</sup></sup> dG | 99.6                | 99.5         | 99.5                | 99.4        |
|              | <sup>1ONP</sup> dG           | 99.6                | 99.5         | 99.5                | 99.3        |
|              | <sup>2ONP</sup> dG           | 99.6                | 99.4         | 99.5                | 99.3        |
|              | <sup>OBP</sup> dG            | 99.6                | 99.4         | 99.5                | 99.3        |
|              | <sup>1OPHN</sup> dG          | 99.6                | 99.4         | 99.4                | 99.3        |
|              | <sup>3OPHN</sup> dG          | 99.6                | 99.6         | 99.4                | 99.3        |
|              | <sup>OPy</sup> dG            | 99.5                | 99.5         | 99.4                | 99.3        |
| W            | <sup>Ph</sup> dG             | 99.6                | 99.5         | 99.6                | 97.5        |
|              | <sup>1NP</sup> dG            | 99.6                | 99.4         | 99.5                | 98.6        |
|              | <sup>2NP</sup> dG            | 99.6                | 99.4         | 99.6                | 98.3        |
|              | <sup>3BP</sup> dG            | 99.6                | 99.6         | 99.6                | 98.8        |
|              | <sup>1PHN</sup> dG           | 99.6                | 99.4         | 99.5                | 98.6        |
|              | <sup>3PHN</sup> dG           | 99.6                | 99.4         | 99.5                | 98.4        |
|              | <sup>Py</sup> dG             | 99.6                | 99.5         | 99.5                | 98.3        |
|              | <sup>O<sup>Ph</sup></sup> dG | 99.7                | 99.5         | 99.6                | 98.9        |
|              | <sup>1ONP</sup> dG           | 99.7                | 99.5         | 99.3                | 99.3        |
|              | <sup>2ONP</sup> dG           | 99.6                | 99.6         | 99.6                | 99.0        |
|              | <sup>OBP</sup> dG            | 99.5                | 99.5         | 99.5                | 98.2        |
|              | <sup>1OPHN</sup> dG          | 99.6                | 99.5         | 99.4                | 96.7        |
|              | <sup>3OPHN</sup> dG          | 99.6                | 99.5         | 99.5                | 98.8        |
|              | <sup>OPy</sup> dG            | 99.6                | 99.5         | 99.3                | 98.5        |
| S            | <sup>Ph</sup> dG             | 99.6                | 99.2         | 99.7                | 97.2        |
|              | <sup>1NP</sup> dG            | 99.6                | 99.5         | 99.1                | 99.3        |
|              | <sup>2NP</sup> dG            | 99.5                | 99.5         | 99.6                | 99.0        |
|              | <sup>3BP</sup> dG            | 99.6                | 99.4         | 99.6                | 99.2        |
|              | <sup>1PHN</sup> dG           | 99.6                | 99.5         | 98.9                | 99.1        |
|              | <sup>3PHN</sup> dG           | 99.5                | 99.5         | 99.6                | 99.5        |
|              | <sup>Py</sup> dG             | 99.6                | 99.5         | 99.1                | 98.2        |
|              | <sup>O<sup>Ph</sup></sup> dG | 99.6                | 99.5         | 99.3                | 99.3        |
|              | <sup>1ONP</sup> dG           | 99.6                | 99.5         | 99.1                | 91.4        |
|              | <sup>2ONP</sup> dG           | 99.5                | 99.5         | 99.4                | 97.5        |
|              | <sup>OBP</sup> dG            | 99.5                | 99.5         | 99.5                | 96.2        |
|              | <sup>1OPHN</sup> dG          | 99.2                | 99.5         | 99.5                | 97.8        |
|              | <sup>3OPHN</sup> dG          | 99.6                | 99.4         | 99.2                | 97.0        |
|              | <sup>OPy</sup> dG            | 99.4                | 99.5         | 99.3                | 98.6        |

<sup>a</sup>Hydrogen-bonding interactions were determined using a cutoff of 3.4 Å for the donor–acceptor distance and 120° for the donor–hydrogen–acceptor angle.

**Table C.2.** Occupancies (%) for the Watson-Crick hydrogen bonds involving the damaged guanine and pairing dC with respect to the lesion respectively, in three conformations of the adducted DNA.<sup>a</sup>

| Conformation | Adduct                       | dG7(N1H) –<br>dC18(N3) | dG7(N2H) –<br>dC18(O2) | dC18(N4H) –<br>dG7(O6) | dC18(N4H) –<br>dG7(N7) | dC18(N4H) –<br>dG17(OP2) | dC18(N4H) –<br>dG17(O6) |
|--------------|------------------------------|------------------------|------------------------|------------------------|------------------------|--------------------------|-------------------------|
| B            | <sup>Ph</sup> dG             | 100.0                  | 99.9                   | 98.6                   | -                      | -                        | -                       |
|              | <sup>1NP</sup> dG            | 99.9                   | 99.9                   | 98.8                   | -                      | -                        | -                       |
|              | <sup>2NP</sup> dG            | 99.8                   | 99.9                   | 98.6                   | -                      | -                        | -                       |
|              | <sup>3BP</sup> dG            | 99.9                   | 99.9                   | 98.8                   | -                      | -                        | -                       |
|              | <sup>1PHN</sup> dG           | 99.9                   | 99.8                   | 98.7                   | -                      | -                        | -                       |
|              | <sup>3PHN</sup> dG           | 99.9                   | 99.9                   | 98.7                   | -                      | -                        | -                       |
|              | <sup>Py</sup> dG             | 99.8                   | 99.9                   | 98.7                   | -                      | -                        | -                       |
|              | <sup>O<sup>Ph</sup></sup> dG | 99.9                   | 99.8                   | 98.3                   | -                      | -                        | -                       |
|              | <sup>1ONP</sup> dG           | 100.0                  | 99.8                   | 98.4                   | -                      | -                        | -                       |
|              | <sup>2ONP</sup> dG           | 99.8                   | 99.7                   | 98.0                   | -                      | -                        | -                       |
|              | <sup>OBP</sup> dG            | 99.8                   | 99.7                   | 98.0                   | -                      | -                        | -                       |
|              | <sup>1OPHN</sup> dG          | 99.9                   | 99.8                   | 98.1                   | -                      | -                        | -                       |
|              | <sup>3OPHN</sup> dG          | 99.9                   | 99.8                   | 98.2                   | -                      | -                        | -                       |
|              | <sup>OPy</sup> dG            | 100.0                  | 99.8                   | 98.2                   | -                      | -                        | -                       |
| W            | <sup>Ph</sup> dG             | -                      | -                      | 52.9                   | 29.7                   | -                        | -                       |
|              | <sup>1NP</sup> dG            | -                      | -                      | 91.8                   | 29.7                   | -                        | -                       |
|              | <sup>2NP</sup> dG            | -                      | -                      | 43.9                   | 63.9                   | -                        | -                       |
|              | <sup>3BP</sup> dG            | -                      | -                      | 54.9                   | 68.5                   | -                        | -                       |
|              | <sup>1PHN</sup> dG           | -                      | -                      | 62.1                   | 70.8                   | -                        | -                       |
|              | <sup>3PHN</sup> dG           | -                      | -                      | 30.8                   | 44.4                   | -                        | -                       |
|              | <sup>Py</sup> dG             | -                      | -                      | 46.5                   | 69.9                   | -                        | -                       |
|              | <sup>O<sup>Ph</sup></sup> dG | -                      | -                      | 81.5                   | 23.4                   | -                        | -                       |
|              | <sup>1ONP</sup> dG           | -                      | -                      | 40.3                   | 50.0                   | -                        | -                       |
|              | <sup>2ONP</sup> dG           | -                      | -                      | 69.1                   | 38.3                   | -                        | -                       |
|              | <sup>OBP</sup> dG            | -                      | -                      | 76.5                   | 45.3                   | -                        | -                       |
|              | <sup>1OPHN</sup> dG          | -                      | -                      | 65.7                   | 51.4                   | -                        | -                       |
|              | <sup>3OPHN</sup> dG          | -                      | -                      | 73.7                   | 43.1                   | -                        | -                       |
|              | <sup>OPy</sup> dG            | -                      | -                      | 39.2                   | 55.4                   | -                        | 39.7                    |
| S            | <sup>Ph</sup> dG             | -                      | -                      | -                      | -                      | 0.0                      | -                       |
|              | <sup>1NP</sup> dG            | -                      | -                      | -                      | -                      | 77.0                     | -                       |
|              | <sup>2NP</sup> dG            | -                      | -                      | -                      | -                      | 48.9                     | -                       |
|              | <sup>3BP</sup> dG            | -                      | -                      | -                      | -                      | 73.6                     | -                       |
|              | <sup>1PHN</sup> dG           | -                      | -                      | -                      | -                      | 78.9                     | -                       |
|              | <sup>3PHN</sup> dG           | -                      | -                      | -                      | -                      | 71.0                     | -                       |
|              | <sup>Py</sup> dG             | -                      | -                      | -                      | -                      | 72.7                     | -                       |
|              | <sup>O<sup>Ph</sup></sup> dG | -                      | -                      | -                      | -                      | 42.5                     | -                       |
|              | <sup>1ONP</sup> dG           | -                      | -                      | -                      | -                      | 72.0                     | -                       |
|              | <sup>2ONP</sup> dG           | -                      | -                      | -                      | -                      | 58.3                     | -                       |
|              | <sup>OBP</sup> dG            | -                      | -                      | -                      | -                      | 74.0                     | -                       |
|              | <sup>1OPHN</sup> dG          | -                      | -                      | -                      | -                      | 0.0                      | -                       |
|              | <sup>3OPHN</sup> dG          | -                      | -                      | -                      | -                      | 35.8                     | -                       |
|              | <sup>OPy</sup> dG            | -                      | -                      | -                      | -                      | 35.1                     | -                       |

<sup>a</sup>Hydrogen-bonding interactions were determined using a cutoff of 3.4 Å for the donor–acceptor distance and 120° for the donor–hydrogen–acceptor angle.

**Table C.3.** Average values and standard deviation (in parentheses) for the backbone RMSD<sup>a</sup> (Å) and key dihedral angles (deg) for different conformations of adducted DNA.

| Adduct              | Conformation | RMSD      | $\chi$       | $\theta$     | $\phi$       |
|---------------------|--------------|-----------|--------------|--------------|--------------|
| <sup>Ph</sup> dG    | B            | 2.3 (0.5) | 226.7 (9.2)  | 3.5 (6.7)    | N/A          |
|                     | W            | 2.8 (0.5) | 53.2 (20.1)  | 180.7 (9.9)  | N/A          |
|                     | S            | 2.9 (0.5) | 37.8 (15.2)  | 357.9 (8.1)  | N/A          |
| <sup>1NP</sup> dG   | B            | 2.2 (0.5) | 219.6 (12.9) | 11.3 (9.7)   | N/A          |
|                     | W            | 2.5 (0.5) | 71.5 (8.5)   | 204.0 (6.1)  | N/A          |
|                     | S            | 2.4 (0.3) | 40.8 (7.4)   | 344.3 (6.4)  | N/A          |
| <sup>2NP</sup> dG   | B            | 2.2 (0.5) | 227.3 (9.7)  | 3.5 (7.4)    | N/A          |
|                     | W            | 2.9 (0.5) | 46.7 (20.0)  | 180.5 (8.9)  | N/A          |
|                     | S            | 2.4 (0.4) | 36.6 (9.8)   | 354.2 (8.1)  | N/A          |
| <sup>3BP</sup> dG   | B            | 2.4 (0.5) | 226.2 (9.1)  | 5.2 (7.2)    | N/A          |
|                     | W            | 2.7 (0.4) | 58.6 (22.6)  | 183.3 (10.0) | N/A          |
|                     | S            | 2.3 (0.5) | 36.1 (10.2)  | 354.0 (6.8)  | N/A          |
| <sup>1PHN</sup> dG  | B            | 2.6 (0.6) | 232.9 (18.8) | 12.6 (9.4)   | N/A          |
|                     | W            | 2.7 (0.5) | 41.7 (8.9)   | 202.5 (6.3)  | N/A          |
|                     | S            | 2.0 (0.4) | 51.6 (9.9)   | 343.8 (6.2)  | N/A          |
| <sup>3PHN</sup> dG  | B            | 2.4 (0.5) | 243.3 (13.9) | 4.8 (7.2)    | N/A          |
|                     | W            | 2.5 (0.4) | 44.7 (13.2)  | 185.4 (9.5)  | N/A          |
|                     | S            | 2.2 (0.5) | 46.5 (10.9)  | 354.2 (8.2)  | N/A          |
| <sup>Py</sup> dG    | B            | 2.6 (0.6) | 232.2 (13.7) | 13.0 (7.5)   | N/A          |
|                     | W            | 2.8 (0.5) | 40.2 (9.1)   | 202.7 (6.2)  | N/A          |
|                     | S            | 2.0 (0.4) | 52.4 (9.6)   | 344.4 (6.1)  | N/A          |
| <sup>OPh</sup> dG   | B            | 2.5 (0.6) | 236.0 (20.8) | 144.6 (19.1) | 90.4 (34.6)  |
|                     | W            | 2.6 (0.5) | 49.9 (13.0)  | 196.0 (43.9) | 4.7 (48.8)   |
|                     | S            | 3.9 (1.4) | 45.2 (12.0)  | 173.7 (21.2) | 131.9 (56.0) |
| <sup>1ONP</sup> dG  | B            | 2.0 (0.4) | 230.8 (23.8) | 160.2 (16.0) | 71.3 (18.6)  |
|                     | W            | 2.5 (0.5) | 51.4 (12.5)  | 176.6 (48.5) | 356.2 (55.7) |
|                     | S            | 2.5 (0.4) | 48.0 (10.1)  | 204.5 (29.3) | 7.8 (64.1)   |
| <sup>2ONP</sup> dG  | B            | 2.0 (0.4) | 231.6 (17.5) | 142.4 (19.4) | 77.6 (89.1)  |
|                     | W            | 2.8 (0.6) | 51.2 (13.4)  | 198.9 (40.0) | 32.1 (64.5)  |
|                     | S            | 2.7 (0.7) | 40.6 (12.1)  | 187.3 (17.5) | 153.1 (18.2) |
| <sup>OBP</sup> dG   | B            | 2.2 (0.5) | 233.8 (17.1) | 147.6 (16.9) | 90.7 (71.1)  |
|                     | W            | 2.3 (0.5) | 54.6 (12.9)  | 163.8 (44.1) | 9.8 (65.8)   |
|                     | S            | 2.4 (0.5) | 44.7 (12.0)  | 207.1 (18.0) | 135.1 (18.1) |
| <sup>1OPHN</sup> dG | B            | 2.9 (0.6) | 239.6 (27.2) | 162.6 (15.0) | 74.7 (18.0)  |
|                     | W            | 2.0 (0.4) | 51.7 (12.8)  | 187.3 (33.1) | 33.7 (43.5)  |
|                     | S            | 2.6 (0.5) | 40.8 (12.2)  | 188.2 (20.1) | 150.7 (18.3) |
| <sup>3OPHN</sup> dG | B            | 2.4 (0.5) | 234.3 (17.4) | 147.1 (15.3) | 87.9 (22.5)  |
|                     | W            | 2.3 (0.4) | 52.1 (13.0)  | 196.3 (35.4) | 47.7 (37.6)  |
|                     | S            | 2.1 (0.4) | 44.7 (11.8)  | 188.9 (21.5) | 325.5 (17.0) |
| <sup>OPy</sup> dG   | B            | 2.7 (0.6) | 232.4 (23.0) | 161.6 (16.1) | 74.8 (18.8)  |
|                     | W            | 3.3 (0.6) | 46.5 (10.2)  | 186.5 (15.1) | 321.7 (14.6) |
|                     | S            | 4.6 (0.6) | 45.5 (10.8)  | 190.6 (18.8) | 322.9 (16.9) |

<sup>a</sup>RMSD is calculated with respect to the first frame of the production run.

**Table C.4.** Average and standard deviations (in parentheses) of lesion-site van der Waals and electrostatic interactions (kJ mol<sup>-1</sup>) within different conformations of C-linked adducted DNA.<sup>a</sup>

| Nucleoside         | Conformation | Lesion Stacking energy with 5' bases | Lesion Stacking energy with 3' bases | Base pair energy at the 5' base pair | Base pair energy at the 3' base pair | Base pair energy at the lesion site | Lesion-site Interaction energy <sup>b</sup> |
|--------------------|--------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|-------------------------------------|---------------------------------------------|
| <sup>Ph</sup> dG   | B            | -63.6 (10.4)                         | -63.6 (9.7)                          | -117.8 (7.6)                         | -118.4 (7.5)                         | -112.8 (6.9)                        | -476.1 (19.9)                               |
| <sup>1NP</sup> dG  |              | -79.8 (12.5)                         | -46.6 (9.5)                          | -119.0 (6.9)                         | -118.2 (7.5)                         | -113.2 (7.0)                        | -476.8 (17.3)                               |
| <sup>2NP</sup> dG  |              | -83.3 (11.8)                         | -46.1 (9.6)                          | -117.4 (8.2)                         | -118.4 (7.3)                         | -113.5 (7.0)                        | -478.8 (17.0)                               |
| <sup>3BP</sup> dG  |              | -84.8 (11.4)                         | -47.3 (9.6)                          | -117.4 (7.8)                         | -118.4 (7.4)                         | -114.4 (6.9)                        | -482.3 (16.5)                               |
| <sup>1PHN</sup> dG |              | -62.5 (9.9)                          | -63.3 (10.7)                         | -118.6 (7.1)                         | -118.0 (7.6)                         | -113.7 (6.9)                        | -476.2 (17.7)                               |
| <sup>3PHN</sup> dG |              | -67.1 (9.9)                          | -64.8 (9.9)                          | -117.5 (7.7)                         | -118.4 (7.4)                         | -113.5 (7.1)                        | -481.2 (16.7)                               |
| <sup>Pv</sup> dG   |              | -63.7 (9.9)                          | -64.9 (9.9)                          | -118.8 (7.0)                         | -118.3 (7.3)                         | -114.1 (7.0)                        | -479.8 (17.3)                               |
| <sup>Ph</sup> dG   | W            | -50.6 (9.7)                          | -38.1 (10.4)                         | -118.8 (7.1)                         | -114.3 (12.2)                        | -42.6 (7.4)                         | -364.4 (21.3)                               |
| <sup>1NP</sup> dG  |              | -45.8 (11.5)                         | -46.9 (10.1)                         | -118.3 (7.3)                         | -113.4 (9.6)                         | -10.4 (30.9)                        | -334.9 (44.2)                               |
| <sup>2NP</sup> dG  |              | -53.8 (10.8)                         | -37.9 (10.1)                         | -118.7 (7.1)                         | -115.3 (10.3)                        | -40.4 (10.0)                        | -366.1 (18.4)                               |
| <sup>3BP</sup> dG  |              | -52.1 (10.8)                         | -38.4 (10.3)                         | -118.6 (7.2)                         | -115.8 (8.6)                         | -41.8 (8.6)                         | -366.6 (18.6)                               |
| <sup>1PHN</sup> dG |              | -55.3 (10.3)                         | -44.4 (9.9)                          | -118.2 (7.3)                         | -116.2 (8.2)                         | -44.4 (8.1)                         | -378.4 (16.3)                               |
| <sup>3PHN</sup> dG |              | -51.2 (12.2)                         | -40.5 (10.5)                         | -118.5 (7.3)                         | -114.4 (9.5)                         | -28.0 (22.9)                        | -352.6 (32.0)                               |
| <sup>Pv</sup> dG   |              | -59.3 (10.3)                         | -40.5 (9.9)                          | -118.0 (7.1)                         | -115.7 (8.4)                         | -42.0 (8.2)                         | -375.5 (16.5)                               |
| <sup>Ph</sup> dG   | S            | -25.0 (17.4)                         | 10.2 (18.2)                          | -118.8 (6.8)                         | -113.5 (9.9)                         | N/A                                 | -247.1 (35.5)                               |
| <sup>1NP</sup> dG  |              | -52.9 (8.7)                          | -10.0 (12.3)                         | -115.1 (7.7)                         | -118.7 (7.8)                         | N/A                                 | -296.8 (17.1)                               |
| <sup>2NP</sup> dG  |              | -44.1 (13.8)                         | -15.9 (15.9)                         | -118.3 (7.1)                         | -117.5 (8.0)                         | N/A                                 | -295.8 (26.5)                               |
| <sup>3BP</sup> dG  |              | -43.0 (9.4)                          | -38.0 (11.9)                         | -118.7 (7.1)                         | -117.8 (7.8)                         | N/A                                 | -317.5 (18.6)                               |
| <sup>1PHN</sup> dG |              | -43.9 (9.7)                          | -53.9 (8.8)                          | -116.2 (7.6)                         | -116.8 (8.2)                         | N/A                                 | -330.8 (17.8)                               |
| <sup>3PHN</sup> dG |              | -38.1 (9.1)                          | -53.2 (8.3)                          | -118.5 (7.0)                         | -118.4 (7.2)                         | N/A                                 | -328.2 (16.3)                               |
| <sup>Pv</sup> dG   |              | -48.0 (10.2)                         | -58.2 (9.6)                          | -116.6 (7.4)                         | -116.6 (10.4)                        | N/A                                 | -339.4 (20.5)                               |

<sup>a</sup>Energies calculated with LIE of cpptraj. <sup>b</sup>Calculated as the sum of all other columns.

**Table C.5.** Average and standard deviations (in parentheses) of the solvent accessible surface area (SASA, Å<sup>2</sup>) of the bulky moiety within different conformations of adducted DNA.

| Conformation | Adduct             | SASA         | Adduct                       | SASA         |
|--------------|--------------------|--------------|------------------------------|--------------|
| B            | <sup>Ph</sup> dG   | 97.7 (5.8)   | <sup>O<sup>Ph</sup></sup> dG | 118.5 (8.7)  |
|              | <sup>1NP</sup> dG  | 129.9 (7.5)  | <sup>1ONP</sup> dG           | 143.5 (10.4) |
|              | <sup>2NP</sup> dG  | 131.7 (6.6)  | <sup>2ONP</sup> dG           | 150.3 (8.6)  |
|              | <sup>3BP</sup> dG  | 178.1 (9.9)  | <sup>OBP</sup> dG            | 187.8 (15.2) |
|              | <sup>1PHN</sup> dG | 169.6 (7.4)  | <sup>1OPHN</sup> dG          | 182.4 (10.3) |
|              | <sup>3PHN</sup> dG | 169.2 (7.5)  | <sup>3OPHN</sup> dG          | 184.3 (10.7) |
|              | <sup>Py</sup> dG   | 166.9 (6.9)  | <sup>OPy</sup> dG            | 179.3 (12.5) |
| W            | <sup>Ph</sup> dG   | 77.7 (14.4)  | <sup>O<sup>Ph</sup></sup> dG | 64.4 (13.5)  |
|              | <sup>1NP</sup> dG  | 137.3 (15.6) | <sup>1ONP</sup> dG           | 90.7 (10.1)  |
|              | <sup>2NP</sup> dG  | 107.4 (20.8) | <sup>2ONP</sup> dG           | 91.9 (17.1)  |
|              | <sup>3BP</sup> dG  | 153.0 (21.6) | <sup>OBP</sup> dG            | 139.9 (22.1) |
|              | <sup>1PHN</sup> dG | 147.3 (25.3) | <sup>1OPHN</sup> dG          | 125.2 (15.4) |
|              | <sup>3PHN</sup> dG | 147.4 (28.7) | <sup>3OPHN</sup> dG          | 124.1 (11.6) |
|              | <sup>Py</sup> dG   | 136.6 (24.9) | <sup>OPy</sup> dG            | 120.3 (8.9)  |
| S            | <sup>Ph</sup> dG   | 41.8 (24.7)  | <sup>O<sup>Ph</sup></sup> dG | 29.4 (11.6)  |
|              | <sup>1NP</sup> dG  | 55.8 (8.9)   | <sup>1ONP</sup> dG           | 64.2 (10.7)  |
|              | <sup>2NP</sup> dG  | 45.1 (10.7)  | <sup>2ONP</sup> dG           | 38.9 (9.5)   |
|              | <sup>3BP</sup> dG  | 49.6 (8.5)   | <sup>OBP</sup> dG            | 61.1 (9.1)   |
|              | <sup>1PHN</sup> dG | 66.7 (8.2)   | <sup>1OPHN</sup> dG          | 89.2 (11.4)  |
|              | <sup>3PHN</sup> dG | 68.4 (8.3)   | <sup>3OPHN</sup> dG          | 85.6 (10.3)  |
|              | <sup>Py</sup> dG   | 92.4 (8.9)   | <sup>OPy</sup> dG            | 104.2 (10.0) |

**Table C.6.** Average and standard deviations (in parentheses) of lesion-site van der Waals and electrostatic interactions (kJ mol<sup>-1</sup>) within different conformations of O-linked adducted DNA.<sup>a</sup>

| Nucleoside           | Conformation | Lesion Stacking energy with 5' bases | Lesion Stacking energy with 3' bases | Base pair energy at the 5' base pair | Base pair energy at the 3' base pair | Base pair energy at the lesion site | Lesion-site Interaction energy <sup>b</sup> |
|----------------------|--------------|--------------------------------------|--------------------------------------|--------------------------------------|--------------------------------------|-------------------------------------|---------------------------------------------|
| O <sup>Ph</sup> dG   | B            | -70.9 (12.7)                         | -73.8 (13.2)                         | -118.3 (7.4)                         | -118.1 (7.5)                         | -108.8 (7.4)                        | -489.9 (18.9)                               |
| 1ONP <sup>d</sup> G  |              | -68.5 (13.1)                         | -74.2 (12.1)                         | -118.6 (7.2)                         | -117.7 (7.9)                         | -108.1 (7.4)                        | -487.1 (19.2)                               |
| 2ONP <sup>d</sup> G  |              | -71.0 (13.1)                         | -71.9 (12.4)                         | -118.4 (7.4)                         | -117.6 (7.9)                         | -108.4 (8.1)                        | -487.3 (19.8)                               |
| OBP <sup>d</sup> G   |              | -73.1 (12.5)                         | -74.6 (11.9)                         | -118.1 (7.7)                         | -118.3 (7.5)                         | -109.4 (7.1)                        | -493.6 (17.6)                               |
| 1OPHN <sup>d</sup> G |              | -67.6 (12.7)                         | -75.2 (12.4)                         | -118.4 (7.4)                         | -117.5 (7.9)                         | -107.8 (7.6)                        | -486.6 (19.5)                               |
| 3OPHN <sup>d</sup> G |              | -70.9 (12.8)                         | -74.1 (12.1)                         | -118.4 (7.4)                         | -118.0 (7.6)                         | -108.9 (7.3)                        | -490.3 (18.2)                               |
| OPy <sup>d</sup> G   |              | -68.0 (13.1)                         | -76.9 (13.3)                         | -118.4 (7.4)                         | -117.5 (7.8)                         | -107.7 (7.5)                        | -488.5 (20.1)                               |
| O <sup>Ph</sup> dG   | W            | -55.1 (13.1)                         | -44.9 (18.4)                         | -118.9 (7.1)                         | -115.9 (9.5)                         | -52.2 (9.8)                         | -387.0 (21.0)                               |
| 1ONP <sup>d</sup> G  |              | -47.0 (14.1)                         | -49.3 (11.9)                         | -117.3 (7.7)                         | -116.6 (8.1)                         | -49.3 (7.0)                         | -379.5 (21.5)                               |
| 2ONP <sup>d</sup> G  |              | -45.1 (15.2)                         | -52.0 (16.9)                         | -118.6 (7.2)                         | -116.2 (8.9)                         | -48.2 (24.3)                        | -380.2 (32.9)                               |
| OBP <sup>d</sup> G   |              | -48.8 (11.0)                         | -68.5 (13.1)                         | -118.5 (7.2)                         | -112.3 (11.8)                        | -46.8 (17.0)                        | -394.9 (29.6)                               |
| 1OPHN <sup>d</sup> G |              | -47.9 (14.3)                         | -62.1 (17.3)                         | -118.0 (7.6)                         | -113.3 (12.8)                        | -52.4 (10.4)                        | -393.7 (27.8)                               |
| 3OPHN <sup>d</sup> G |              | -47.1 (12.7)                         | -64.1 (14.3)                         | -118.4 (7.3)                         | -116.3 (8.7)                         | -53.4 (9.2)                         | -399.3 (21.1)                               |
| OPy <sup>d</sup> G   |              | -41.3 (14.0)                         | -57.3 (13.0)                         | -118.6 (6.9)                         | -116.8 (10.2)                        | -59.5 (8.3)                         | -393.6 (22.5)                               |
| O <sup>Ph</sup> dG   | S            | -32.5 (15.7)                         | 2.5 (15.4)                           | -117.1 (7.7)                         | -118.3 (7.8)                         | N/A                                 | -265.4 (24.4)                               |
| 1ONP <sup>d</sup> G  |              | -34.6 (16.5)                         | -26.0 (23.2)                         | -118.0 (7.2)                         | -115.3 (9.7)                         | N/A                                 | -293.9 (27.2)                               |
| 2ONP <sup>d</sup> G  |              | -36.5 (13.1)                         | -45.3 (15.7)                         | -118.0 (7.3)                         | -115.4 (11.3)                        | N/A                                 | -315.2 (25.9)                               |
| OBP <sup>d</sup> G   |              | -31.7 (10.4)                         | -63.4 (16.4)                         | -118.7 (7.2)                         | -115.1 (14.8)                        | N/A                                 | -328.9 (24.6)                               |
| 1OPHN <sup>d</sup> G |              | -40.8 (15.8)                         | -53.6 (16.3)                         | -118.0 (7.0)                         | -115.3 (9.6)                         | N/A                                 | -327.7 (28.0)                               |
| 3OPHN <sup>d</sup> G |              | -46.7 (15.7)                         | -55.6 (15.9)                         | -117.4 (7.3)                         | -115.2 (11.8)                        | N/A                                 | -335.0 (29.9)                               |
| OPy <sup>d</sup> G   |              | -49.9 (15.5)                         | -64.4 (17.0)                         | -117.6 (7.2)                         | -116.9 (9.1)                         | N/A                                 | -348.7 (30.8)                               |

<sup>a</sup>Energies calculated with LIE of cpptraj. <sup>b</sup>Calculated as the sum of all other columns.

**Table C.7.** Dependence of the average SASA ( $\text{\AA}^2$ ) on the size of the aromatic bulky group for C-linked, O-linked and N-linked adducts.

| Conformation | Aryl Group                         | C-linked | O-linked | N-linked <sup>a</sup> | Average <sup>b</sup> |
|--------------|------------------------------------|----------|----------|-----------------------|----------------------|
| B            | One Ring <sup>c</sup>              | 97.7     | 118.5    | 116.1                 | 110.8                |
|              | Two-Fused Rings <sup>d</sup>       | 130.8    | 146.9    | 142.6                 | 140.1                |
|              | Two Substituted Rings <sup>e</sup> | 178.1    | 187.8    | 191.1                 | 185.7                |
|              | Three-Fused Rings <sup>f</sup>     | 169.4    | 183.4    | 177.8                 | 176.8                |
|              | Four-Fused Rings <sup>g</sup>      | 166.9    | 179.3    | 175.6                 | 173.9                |
| S            | One Ring <sup>c</sup>              | 41.8     | 29.4     | 30.7                  | 34.0                 |
|              | Two-Fused Rings <sup>d</sup>       | 50.5     | 51.6     | 55.5                  | 52.5                 |
|              | Two Substituted Rings <sup>e</sup> | 49.6     | 61.1     | 56.9                  | 55.9                 |
|              | Three-Fused Rings <sup>f</sup>     | 67.6     | 87.4     | 80.7                  | 78.6                 |
|              | Four-Fused Rings <sup>g</sup>      | 92.4     | 104.2    | 107.8                 | 101.5                |

<sup>a</sup>Data from DOI: 10.1021/acs.jcim.8b00238. <sup>b</sup>Average calculated over all three linkage types. <sup>c</sup>Averaged over the phenyl-derived adducts. <sup>d</sup>Averaged over the naphthyl-derived adducts. <sup>e</sup>Averaged over the biphenyl-derived adducts. <sup>f</sup>Averaged over the phenanthrenyl-derived adducts. <sup>g</sup>Averaged over the pyrenyl-derived adducts.

**Table C.8.** Dependence of the average LSIE (kJ/mol) on the size of the aromatic bulky group for C-linked, O-linked and N-linked<sup>a</sup> (adducts).

| Conformation | Aryl Group                         | C-linked | O-linked | N-linked | Average <sup>b</sup> |
|--------------|------------------------------------|----------|----------|----------|----------------------|
| B            | One Ring <sup>c</sup>              | −476.1   | −489.9   | −475.5   | −480.5               |
|              | Two-Fused Rings <sup>d</sup>       | −477.8   | −487.2   | −475.8   | −480.3               |
|              | Two Substituted Rings <sup>e</sup> | −482.3   | −493.6   | −475.6   | −483.8               |
|              | Three-Fused Rings <sup>f</sup>     | −478.7   | −488.4   | −475.9   | −481.0               |
|              | Four-Fused Rings <sup>g</sup>      | −479.8   | −488.5   | −473.9   | −480.7               |
| S            | One Ring <sup>c</sup>              | −247.1   | −265.4   | −300.6   | −271.0               |
|              | Two-Fused Rings <sup>d</sup>       | −296.3   | −304.6   | −330.9   | −310.6               |
|              | Two Substituted Rings <sup>e</sup> | −317.5   | −328.9   | −329.3   | −325.2               |
|              | Three-Fused Rings <sup>f</sup>     | −329.5   | −331.3   | −337.2   | −332.7               |
|              | Four-Fused Rings <sup>g</sup>      | −339.4   | −348.7   | −358.2   | −348.8               |

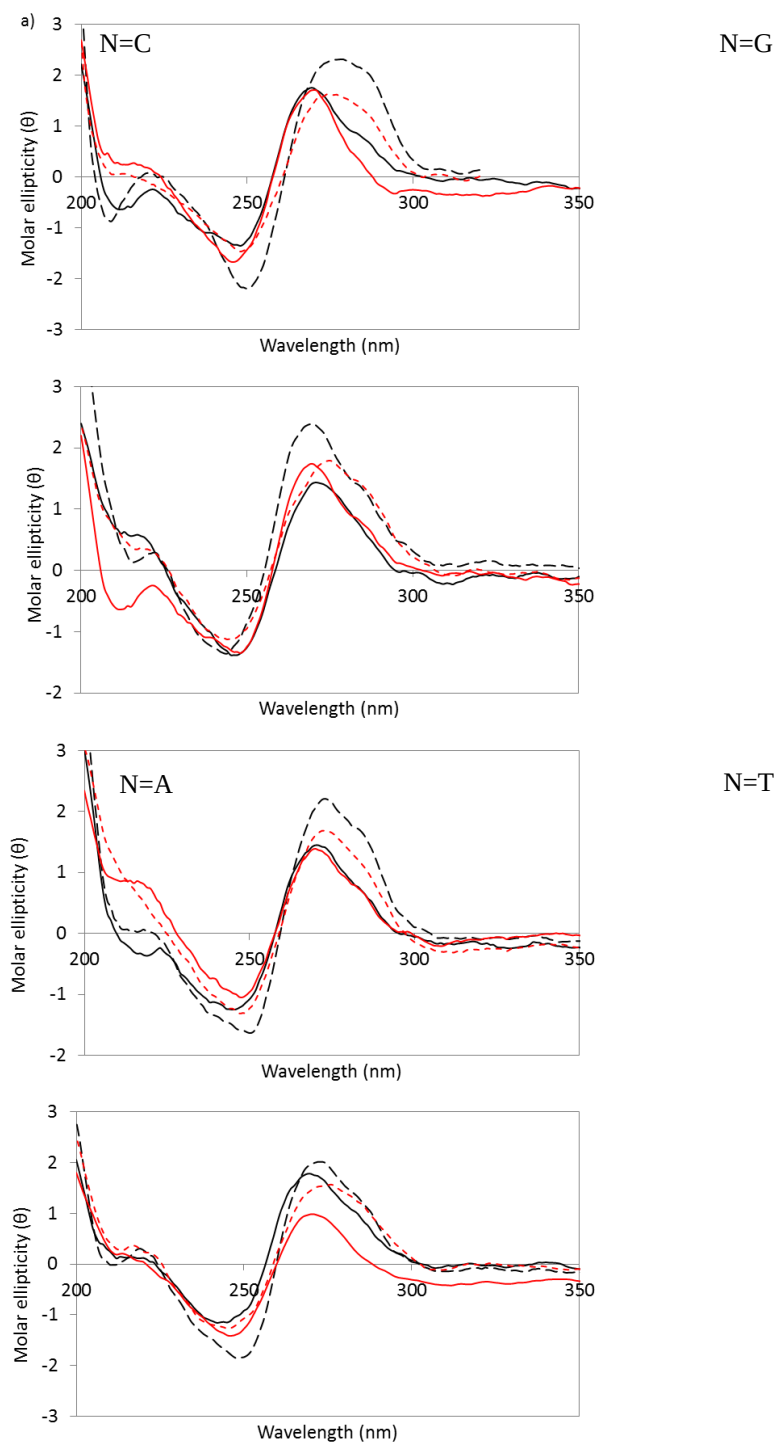
<sup>a</sup>Data from DOI: 10.1021/acs.jcim.8b00238. <sup>b</sup>Average calculated over all three linkage types. <sup>c</sup>Averaged over the phenyl-derived adducts. <sup>d</sup>Averaged over the naphthyl-derived adducts. <sup>e</sup>Averaged over the biphenyl-derived adducts. <sup>f</sup>Averaged over the phenanthrenyl-derived adducts. <sup>g</sup>Averaged over the pyrenyl-derived adducts.



## Appendix D

### **Supplementary Information for Chapter 4: Conformational Preference and Fluorescence Response of a C-Linked C8-Biphenyl-Guanine Lesion in the *NarI* Mutational Hotspot: Evidence for Enhanced *syn*-Adduct Formation**

Contains information on the synthesis of the modified DNA, the molecular dynamics protocol, Scheme D.1, Figures D.1–D.30 and Tables D.1–D.7



**Figure D.1.** CD spectral overlays of *NarI* duplexes, illustrating a general B-form pattern with a major positive band at 275 nm, a negative band at 240 nm and crossover at 260 nm. *NarI*(G):*NarI*(FP) (red lines), *NarI*(G):*NarI*(SMI) (dashed red lines), *NarI*(<sup>FBP</sup>G):*NarI*(FP) (black lines), *NarI*(<sup>FBP</sup>G):*NarI*(SMI) (dashed black lines) with N=A; C; G or T.

**Table D.1.** Solvatochromic Properties of FBP-dG.

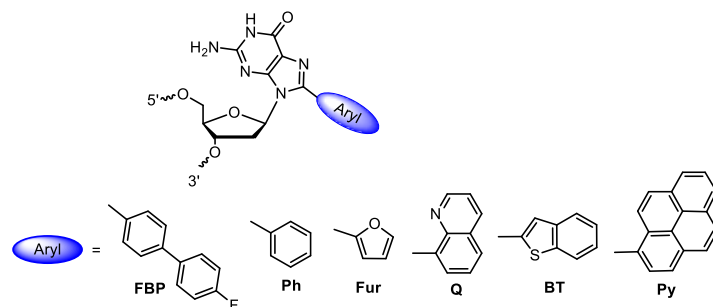
| Solvent            | $\lambda_{\text{ex}}$ (nm) | $\lambda_{\text{em}}$ (nm) | $\lambda_{\text{fl}}$ |
|--------------------|----------------------------|----------------------------|-----------------------|
| H <sub>2</sub> O   | 286                        | 424                        | 0.52                  |
| CHCl <sub>3</sub>  | 303                        | 410                        | 0.48                  |
| CH <sub>3</sub> CN | 305                        | 422                        | 0.60                  |

**Table D.2.** Photophysical parameters for the modified *NarI* (12mer).

| N | M              | $\lambda_{\text{ex}}$<br>(nm) <sup>a</sup> | $\Delta\lambda_{\text{ex}}$<br>(nm) <sup>b</sup> | $\lambda_{\text{em}}$<br>(nm) | $\Delta\lambda_{\text{em}}$<br>(nm) | $\Delta\nu$<br>(cm <sup>-1</sup> ) <sup>c</sup> |
|---|----------------|--------------------------------------------|--------------------------------------------------|-------------------------------|-------------------------------------|-------------------------------------------------|
| C | / <sup>f</sup> | 314                                        | /                                                | 421                           | /                                   | 8089                                            |
|   | G              | 318                                        | 4                                                | 417                           | -4                                  | 7466                                            |
|   | -2             | 314                                        | 0                                                | 417                           | -4                                  | 7866                                            |
| G | /              | 313                                        | /                                                | 420                           | /                                   | 8118                                            |
|   | C              | 319                                        | 6                                                | 416                           | -4                                  | 7305                                            |
|   | -2             | 315                                        | 2                                                | 414                           | -6                                  | 7576                                            |
| A | /              | 315                                        | /                                                | 422                           | /                                   | 8034                                            |
|   | T              | 323                                        | 8                                                | 420                           | -2                                  | 7141                                            |
|   | -2             | 315                                        | 0                                                | 413                           | -9                                  | 7533                                            |
| T | /              | 311                                        | /                                                | 420                           | /                                   | 8323                                            |
|   | A              | 323                                        | 12                                               | 420                           | 0                                   | 7141                                            |
|   | -2             | 315                                        | 4                                                | 415                           | -5                                  | 7640                                            |

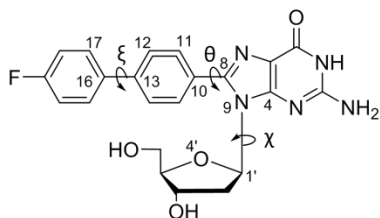
<sup>a</sup>All spectra for single strand and duplexes (6.0  $\mu\text{M}$ ) were recorded in 50 mM sodium phosphate buffer, pH 7, with 0.1 M NaCl. <sup>b</sup>Change in excitation or emission maximum for duplex versus single strand. <sup>c</sup>Stokes' shift ( $\Delta\nu$ ) is calculated as  $(1/\lambda_{\text{ex}} - 1/\lambda_{\text{em}})$  <sup>d</sup> $I_{\text{rel}} = I_{\text{single-strand}}/I_{\text{FP}}$ . <sup>e</sup> $I_{\text{rel}} = I_{\text{SMI}}/I_{\text{FP}}$ . <sup>f</sup>/ indicates optical properties of the modified base in the single strand.

**Table D.3. Properties of C-linked C8-aryl-dG adducts in fully paired and SMI *Nar*I 12-mer duplexes.**



| Aryl group       | Duplex       | fully paired<br>$E_{\text{syn}}/E_{\text{anti}}^a$ | fully paired<br>$\Delta T_m^b$ | fully paired<br>$I_{\text{rel}}^c$ |
|------------------|--------------|----------------------------------------------------|--------------------------------|------------------------------------|
| FBP              | fully paired | 2.3/0.0                                            | -12.5                          | 0.64                               |
|                  | SMI          | 2.8/0.0                                            | +6.0                           | 4.2                                |
| Ph <sup>d</sup>  | fully paired | 26.4/0.0                                           | -15.7                          | 0.38                               |
|                  | SMI          | 15.1/0.0                                           | -2.9                           | 0.80                               |
| Fur <sup>d</sup> | fully paired | 14.2/0.0                                           | -9.1                           | 0.45                               |
|                  | SMI          | 14.2/0.0                                           | -2.6                           | 2.97                               |
| Q <sup>d</sup>   | fully paired | 25.1/0.0                                           | -18.6                          | 0.41                               |
|                  | SMI          | 0.0/23.8                                           | 0.0                            | 4.92                               |
| BT <sup>e</sup>  | fully paired | 18.4/0.0                                           | -15.8                          | 1.45                               |
|                  | SMI          | 10.0/0.0                                           | +3.8                           | 1.05                               |
| Py <sup>e</sup>  | fully paired | 7.9/0.0                                            | -20.1                          | 0.85                               |
|                  | SMI          | 0.0/12.6                                           | +6.2                           | 0.61                               |

<sup>a</sup> Calculated relative energies (kJ/mol) of *syn* versus *anti* adduct conformation. <sup>b</sup>  $\Delta T_m$  (°C) of adducted duplex vs. unmodified control. <sup>c</sup> Relative emission intensity of ss vs fully paired or SMI vs fully paired duplexes. <sup>d</sup> Values taken from Ref. 29. <sup>e</sup> Values taken from Ref. 30.

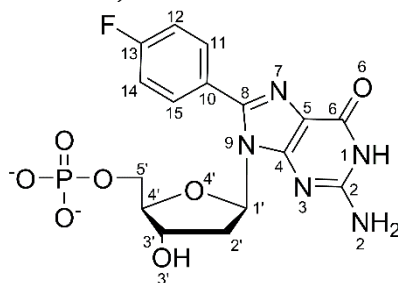


**Figure D.2.** Chemical numbering and definition of key dihedral angles in the <sup>FBP</sup>dG adduct, with  $\chi = \angle(\text{O4}'\text{C1}'\text{N9C4})$ ,  $\theta = \angle(\text{N9C8C10C11})$ , and  $\xi = \angle(\text{C12C13C16C17})$ .

**Table D.4:** Average values of key dihedral angles (deg.) and backbone RMSD for clustered simulations on <sup>FPh</sup>dG or <sup>FBP</sup>dG-adducted DNA.

| Adduct            | Conformation            | $\chi$ (°)    | $\theta$ (°)  | $\xi$ (°)     | RMSD (Å)    |
|-------------------|-------------------------|---------------|---------------|---------------|-------------|
| <sup>FPh</sup> dG | Major-Groove            | −133.4 (9.3)  | −176.7 (6.8)  | N/A           | 2.35 (0.55) |
|                   | Wedge                   | 38.9 (10.0)   | 175.9 (8.4)   | N/A           | 2.39 (0.43) |
|                   | Intercalated            | 38.9 (13.5)   | −5.0 (8.5)    | N/A           | 3.95 (0.67) |
|                   | Base-Displaced          | 35.8 (8.5)    | −5.6 (7.3)    | N/A           | 3.90 (0.57) |
| <sup>FBP</sup> dG | Major-Groove            | −132.5 (9.3)  | −177.0 (6.9)  | 162.9 (100.8) | 2.32 (0.52) |
|                   | Wedge                   | 61.3 (22.7)   | −176.7 (10.0) | 42.7 (107.8)  | 2.79 (0.54) |
|                   | Intercalated            | 35.6 (10.7)   | −4.1 (7.4)    | −162.0 (28.9) | 3.06 (0.53) |
|                   | Base-Displaced          | 40.3 (10.0)   | −7.7 (7.5)    | 167.6 (16.8)  | 3.00 (0.51) |
|                   | −2 strand - <i>anti</i> | −131.6 (11.1) | 4.6 (8.2)     | 18.9 (101.9)  | 3.22 (0.55) |
|                   | −2 strand - <i>syn</i>  | 40.2 (10.4)   | 176.0 (6.7)   | −179.5 (25.3) | 3.69 (0.65) |

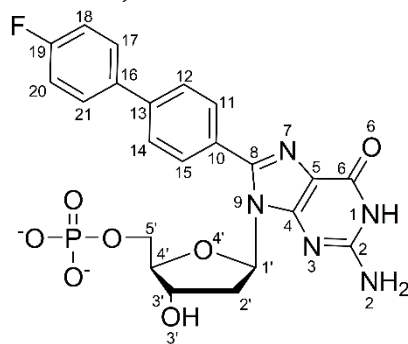
**Table D.5:** MOL2 file of the F-Phenyl-dG adduct generated using the RED program (atom numbering provided in the figure below).



|    |      |       |        |        |    |   |     |         |
|----|------|-------|--------|--------|----|---|-----|---------|
| 1  | P    | 4.682 | 0.033  | -1.207 | P  | 1 | LIG | 1.2093  |
| 2  | O5'  | 4.186 | 1.03   | 0      | OS | 1 | LIG | -0.4938 |
| 3  | OP1  | 4.581 | 0.757  | -2.483 | O2 | 1 | LIG | -0.7864 |
| 4  | OP2  | 5.894 | -0.639 | -0.716 | O2 | 1 | LIG | -0.7950 |
| 5  | O3'  | 0     | 0      | 0      | O  | 1 | LIG | -0.5494 |
| 6  | C5'  | 3.127 | 1.932  | -0.26  | CI | 1 | LIG | -0.0222 |
| 7  | H5'  | 2.887 | 2.429  | 0.688  | H1 | 1 | LIG | 0.0806  |
| 8  | H5'' | 3.42  | 2.712  | -0.98  | H1 | 1 | LIG | 0.0806  |
| 9  | C4'  | 1.869 | 1.246  | -0.794 | CT | 1 | LIG | 0.1858  |
| 10 | H4'  | 1.04  | 1.964  | -0.816 | H1 | 1 | LIG | 0.0992  |
| 11 | O4'  | 2.118 | 0.795  | -2.146 | OS | 1 | LIG | -0.4181 |
| 12 | C1'  | 1.887 | -0.604 | -2.254 | CT | 1 | LIG | 0.1229  |
| 13 | H1'  | 0.883 | -0.795 | -2.639 | H2 | 1 | LIG | 0.0912  |
| 14 | C3'  | 1.429 | 0      | 0      | CT | 1 | LIG | 0.1455  |
| 15 | H3'  | 1.825 | 0.015  | 1.023  | H1 | 1 | LIG | 0.0749  |
| 16 | C2'  | 2.004 | -1.158 | -0.825 | CT | 1 | LIG | -0.0637 |
| 17 | H2'  | 3.047 | -1.33  | -0.555 | HC | 1 | LIG | 0.0455  |
| 18 | H2'' | 1.444 | -2.09  | -0.699 | HC | 1 | LIG | 0.0455  |
| 19 | N9   | 2.815 | -1.158 | -3.232 | N* | 1 | LIG | -0.0240 |
| 20 | C8   | 2.515 | -1.806 | -4.443 | CK | 1 | LIG | 0.2328  |
| 21 | C4   | 4.182 | -1.246 | -3.094 | CB | 1 | LIG | 0.1719  |
| 22 | N7   | 3.587 | -2.314 | -5.013 | NB | 1 | LIG | -0.5106 |
| 23 | C5   | 4.634 | -1.972 | -4.19  | CB | 1 | LIG | 0.1374  |
| 24 | N3   | 4.92  | -0.696 | -2.094 | NC | 1 | LIG | -0.4924 |
| 25 | C6   | 6.045 | -2.234 | -4.327 | C  | 1 | LIG | 0.4724  |
| 26 | C2   | 6.214 | -0.915 | -2.198 | CA | 1 | LIG | 0.5847  |
| 27 | O6   | 6.669 | -2.851 | -5.169 | O  | 1 | LIG | -0.5308 |
| 28 | N1   | 6.76  | -1.631 | -3.227 | NA | 1 | LIG | -0.4411 |
| 29 | H1   | 7.755 | -1.825 | -3.24  | H  | 1 | LIG | 0.3482  |
| 30 | N2   | 7.049 | -0.449 | -1.215 | N2 | 1 | LIG | -0.8464 |
| 31 | H21  | 7.996 | -0.206 | -1.473 | H  | 1 | LIG | 0.3871  |

|    |     |        |        |        |    |   |     |         |
|----|-----|--------|--------|--------|----|---|-----|---------|
| 32 | H22 | 6.607  | 0.204  | -0.577 | H  | 1 | LIG | 0.3871  |
| 33 | C10 | 1.172  | -1.882 | -5.041 | CA | 1 | LIG | 0.0197  |
| 34 | C11 | 0.287  | -0.789 | -5.048 | CA | 1 | LIG | -0.1202 |
| 35 | H11 | 0.581  | 0.153  | -4.595 | HA | 1 | LIG | 0.1545  |
| 36 | C15 | 0.796  | -3.065 | -5.7   | CA | 1 | LIG | -0.1290 |
| 37 | H15 | 1.492  | -3.897 | -5.719 | HA | 1 | LIG | 0.1692  |
| 38 | C12 | -0.957 | -0.885 | -5.669 | CA | 1 | LIG | -0.2080 |
| 39 | H12 | -1.648 | -0.049 | -5.686 | HA | 1 | LIG | 0.1696  |
| 40 | C14 | -0.441 | -3.169 | -6.328 | CA | 1 | LIG | -0.2164 |
| 41 | H14 | -0.746 | -4.077 | -6.837 | HA | 1 | LIG | 0.1699  |
| 42 | C13 | -1.303 | -2.077 | -6.295 | CA | 1 | LIG | 0.2515  |
| 43 | F   | -2.506 | -2.175 | -6.896 | F  | 1 | LIG | -0.1896 |

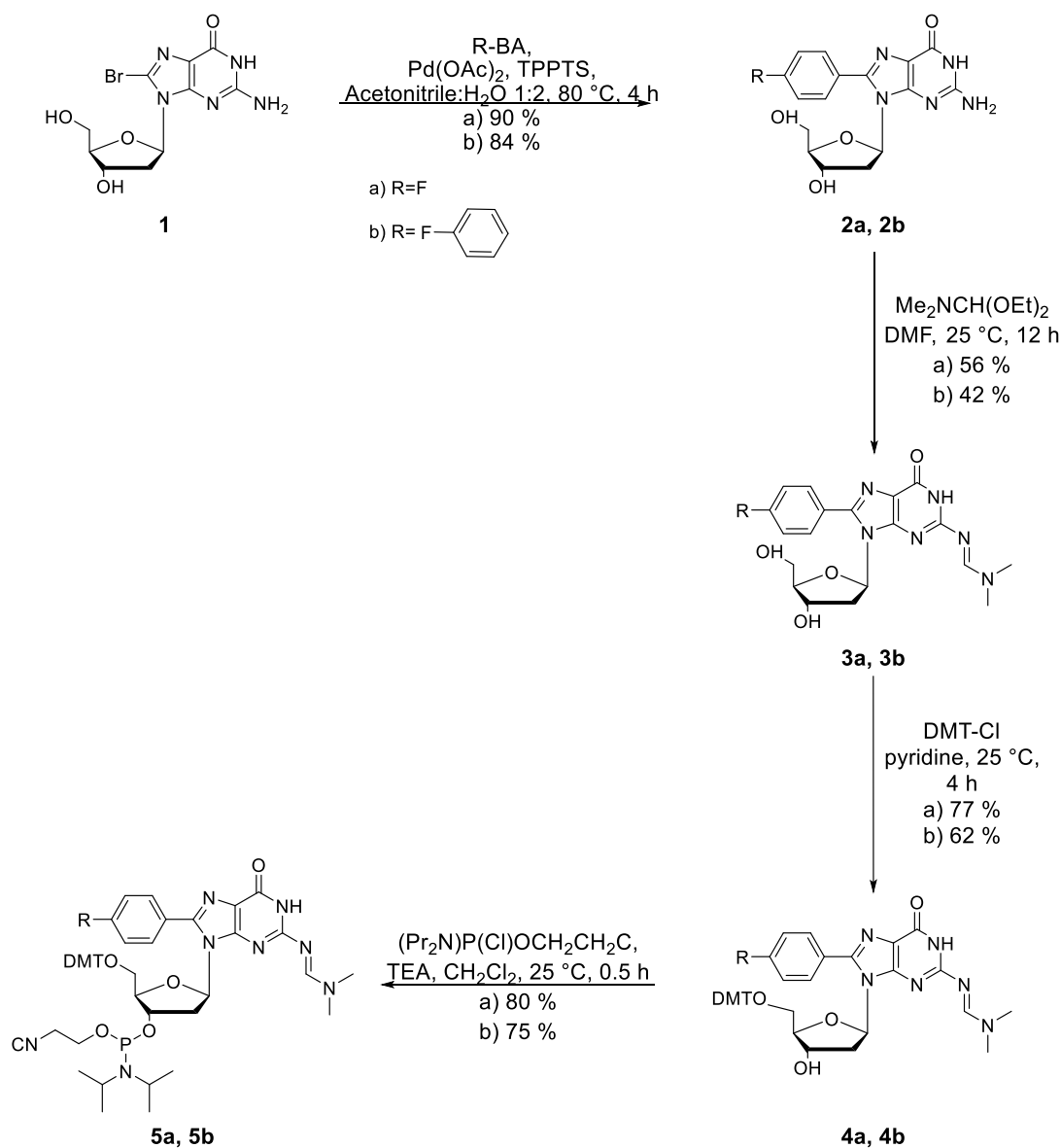
**Table D.6:** MOL2 file of the FBP-dG adduct generated using the RED program (atom numbering provided in the figure below).



|    |      |       |        |        |    |   |     |         |
|----|------|-------|--------|--------|----|---|-----|---------|
| 1  | P    | 4.701 | 0.017  | -1.112 | P  | 1 | LIG | 1.2091  |
| 2  | O5   | 4.162 | 1.098  | 0      | OS | 1 | LIG | -0.4910 |
| 3  | OP1  | 4.608 | 0.63   | -2.446 | O2 | 1 | LIG | -0.7864 |
| 4  | OP2  | 5.918 | -0.584 | -0.547 | O2 | 1 | LIG | -0.7952 |
| 5  | O3'  | 0     | 0      | 0      | OS | 1 | LIG | -0.5473 |
| 6  | C5'  | 3.088 | 1.952  | -0.351 | CI | 1 | LIG | -0.0175 |
| 7  | H5'  | 2.801 | 2.504  | 0.552  | H1 | 1 | LIG | 0.0816  |
| 8  | H5'' | 3.384 | 2.69   | -1.114 | H1 | 1 | LIG | 0.0816  |
| 9  | C4'  | 1.872 | 1.189  | -0.877 | CT | 1 | LIG | 0.1523  |
| 10 | H4'  | 1.025 | 1.877  | -0.991 | H1 | 1 | LIG | 0.1052  |
| 11 | O4'  | 2.196 | 0.641  | -2.175 | OS | 1 | LIG | -0.3937 |
| 12 | C1'  | 1.937 | -0.755 | -2.206 | CT | 1 | LIG | 0.1104  |
| 13 | H1'  | 0.946 | -0.94  | -2.629 | H2 | 1 | LIG | 0.1002  |
| 14 | C3'  | 1.429 | 0      | 0      | CT | 1 | LIG | 0.1513  |
| 15 | H3'  | 1.824 | 0.085  | 1.02   | H1 | 1 | LIG | 0.0733  |
| 16 | C2'  | 2.005 | -1.213 | -0.741 | CT | 1 | LIG | -0.0655 |



|    |      |        |        |        |    |   |     |         |
|----|------|--------|--------|--------|----|---|-----|---------|
| 17 | H2'  | 3.044  | -1.38  | -0.446 | HC | 1 | LIG | 0.0462  |
| 18 | H2'' | 1.436  | -2.131 | -0.558 | HC | 1 | LIG | 0.0462  |
| 19 | N9   | 2.898  | -1.392 | -3.104 | N* | 1 | LIG | -0.0064 |
| 20 | C8   | 2.64   | -2.089 | -4.3   | CK | 1 | LIG | 0.2410  |
| 21 | C4   | 4.258  | -1.167 | -3.096 | CB | 1 | LIG | 0.1563  |
| 22 | N7   | 3.735  | -2.301 | -4.998 | NB | 1 | LIG | -0.5275 |
| 23 | C5   | 4.75   | -1.734 | -4.267 | CB | 1 | LIG | 0.1547  |
| 24 | N3   | 4.955  | -0.501 | -2.139 | NC | 1 | LIG | -0.5181 |
| 25 | C6   | 6.162  | -1.658 | -4.549 | C  | 1 | LIG | 0.4914  |
| 26 | C2   | 6.246  | -0.405 | -2.378 | CA | 1 | LIG | 0.6308  |
| 27 | O6   | 6.817  | -2.079 | -5.484 | O  | 1 | LIG | -0.5370 |
| 28 | N1   | 6.832  | -0.942 | -3.489 | NA | 1 | LIG | -0.4811 |
| 29 | H1   | 7.838  | -0.893 | -3.61  | H  | 1 | LIG | 0.3568  |
| 30 | N2   | 7.046  | 0.219  | -1.455 | N2 | 1 | LIG | -0.8602 |
| 31 | H21  | 7.89   | 0.667  | -1.786 | H  | 1 | LIG | 0.3885  |
| 32 | H22  | 6.53   | 0.754  | -0.762 | H  | 1 | LIG | 0.3885  |
| 33 | C10  | 1.316  | -2.567 | -4.726 | CA | 1 | LIG | 0.0296  |
| 34 | C11  | 0.362  | -3.091 | -3.839 | CA | 1 | LIG | -0.0793 |
| 35 | H11  | 0.583  | -3.163 | -2.779 | HA | 1 | LIG | 0.1102  |
| 36 | C15  | 1.015  | -2.56  | -6.098 | CA | 1 | LIG | -0.1559 |
| 37 | H15  | 1.76   | -2.188 | -6.793 | HA | 1 | LIG | 0.1604  |
| 38 | C12  | -0.863 | -3.56  | -4.307 | CA | 1 | LIG | -0.1924 |
| 39 | H12  | -1.592 | -3.937 | -3.595 | HA | 1 | LIG | 0.1403  |
| 40 | C14  | -0.206 | -3.035 | -6.56  | CA | 1 | LIG | -0.1317 |
| 41 | H14  | -0.404 | -3.038 | -7.628 | HA | 1 | LIG | 0.1382  |
| 42 | C13  | -1.177 | -3.538 | -5.675 | cp | 1 | LIG | 0.0223  |
| 43 | C16  | -2.485 | -4.033 | -6.171 | cp | 1 | LIG | 0.0372  |
| 44 | C17  | -3.103 | -5.156 | -5.593 | CA | 1 | LIG | -0.1293 |
| 45 | H17  | -2.602 | -5.688 | -4.789 | HA | 1 | LIG | 0.1506  |
| 46 | C21  | -3.144 | -3.392 | -7.234 | CA | 1 | LIG | -0.1522 |
| 47 | H21' | -2.7   | -2.508 | -7.683 | HA | 1 | LIG | 0.1472  |
| 48 | C18  | -4.332 | -5.625 | -6.052 | CA | 1 | LIG | -0.2504 |
| 49 | H18  | -4.807 | -6.497 | -5.616 | HA | 1 | LIG | 0.1744  |
| 50 | C20  | -4.372 | -3.848 | -7.707 | CA | 1 | LIG | -0.2233 |
| 51 | H20  | -4.888 | -3.35  | -8.521 | HA | 1 | LIG | 0.1690  |
| 52 | C19  | -4.949 | -4.96  | -7.106 | CA | 1 | LIG | 0.3046  |
| 53 | F    | -6.139 | -5.408 | -7.557 | F  | 1 | LIG | -0.2078 |



**Scheme D.1:** Synthesis of FPh-dG phosphoramidite and FBP-dG phosphoramidite.

**Synthesis of Modified *NarI*.** FPh-dG and FBP-dG were synthesized by Suzuki-Miyaura cross-coupling between 8-Br-dG and 4-fluorophenylboronic acid or 4-(4-fluorophenyl)phenylboronic acid and converted into the phosphoramidite (Scheme D.1) by adaption of the method described previously for other C-linked C8-aryl-dG adducts.<sup>1-2</sup> Solid phase DNA synthesis was then used to incorporate the corresponding phosphoramidite into the 12-mer *NarI* sequence (5'-CTCGGNXCCATC, N=C,

X=FBP-dG, FPh-dG) and additionally in oligonucleotides with varying 5'-flanking base (N = A, G or T). The various adducted 12-mer sequences were annealed to a complementary 12-mer sequence or a truncated 10mer sequence 5'-GATGGCCGAG modeling the SMI.

**8-(aryl)-2'-deoxyguanosine (2a, 2b).** The title compound was synthesized by adaption of a previously described method.<sup>1</sup> 8-Br-dG (1) (0.3 g, 0.9 mmol), Pd(OAc)<sub>2</sub> (0.005 g, 0.026 mmol), TPPTS (0.03 mg, 0.05 mmol), Na<sub>2</sub>CO<sub>3</sub> (0.3 g, 1.7 mmol) and **a) F-phenyl-boronic acid** (0.14 g, 1.2 mmol), **b) 4-(4-Fluorophenyl)phenylboronic acid** (0.23 g, 1.0 mmol) were dissolved in 2:1 H<sub>2</sub>O: acetonitrile allowed to react for 4 h under reflux at 80 °C. The reaction was diluted with ca. 20 mL of water and the pH adjusted to 6–7 with 1 M aq. HCl. The mixture was allowed to cool to 0 °C for 16 h hours and the precipitate was recovered by vacuum filtration as solid,

**2a)** 0.25 g (80 %), <sup>1</sup>H NMR (300 MHz, DMSO-d<sub>6</sub>) δ 1.97-2.04 (m, 1H), 3.07-3.17 (m, 1H), 3.48-3.54 (m, 1H), 3.60-3.66 (m, 1H), 3.76-3.80 (m, 1H), 4.30-4.32 (t, 1H), 5.15-5.21 (t, 1H), 5.99-6.04 (t, 1H), 6.59 (s, 1H), 7.33-7.39 (t, 2H), 7.65-7.70 (m, 2H). <sup>13</sup>C NMR (101 MHz, DMSO-d<sub>6</sub>) δ 36.62, 62.12, 71.21, 84.67, 87.92, 115.53, 115.76, 117.12, 131.34, 131.42, 145.57, 152.05, 154.53, 161.36, 163.82. HRMS calcd for C<sub>16</sub>H<sub>17</sub>FN<sub>5</sub>O<sub>4</sub> [M+H<sup>+</sup>]: 362.1265, found: 362.1234. **b)** 0.34 g (90 %). <sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>) δ 2.04-2.08 (m, 1H), 3.23-3.27 (m, 1H), 3.59 (s, 1H), 3.70-3.72 (d, 1H), 3.86 (s, 1H), 4.39 (s, 1H), 5.20 (s, 1H), 5.87 (2, 1H), 6.14-6.17 (t, 1H), 6.74-6.75 (bt, 2H), 7.32-7.36 (t, 2H), 7.73-7.74 (d, 2H), 7.83-7.85 (d, 4H). <sup>13</sup>C NMR (101 MHz, DMSO-d<sub>6</sub>) δ 37.23, 62.68, 71.78, 85.44, 88.42, 115.93, 116.13, 117.86, 126.93, 129.011, 129.73, 135.95, 139.74, 145.55, 152.55, 156.94, 161.03, 162.54, 163.50. HRMS calcd for C<sub>22</sub>H<sub>21</sub>FN<sub>5</sub>O<sub>4</sub> [M+H<sup>+</sup>]: 438.1578, found: 438.1560.

**N<sup>2</sup>-(Dimethylformamidyl)-8-(aryl)-2'-deoxyguanosine (3a, 3b).** The title compound was synthesized by adaption of a previously described method.<sup>2</sup> **2a**) (0.55 g, 1.322 mmol), **2b**) (0.34 g, 0.78 mmol) and dimethyl-formamide-diethyl-acetal (a) 1.1 mL, 6.08 mmol, b) 0.53 mL, 3.11 mmol) were dissolved in DMF in a round bottomed flask and allowed to react for 16 h at room temperature under argon atmosphere. The resulting solution was then evaporated to dryness, washed with MeOH and used without further purification to yield product. **a**) 0.32 g (51 %) <sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>) δ 2.065-2.11 (m, 1H), 3.04 (s, 3H), 3.15 (s, 3H), 3.18-3.25 (m, 1H), 3.38 (s, 1H), 3.54-3.57 (m, 1H), 3.65-3.68 (d, 2H), 3.81-3.82 (d, 2H), 4.92 (s, 1H), 5.25 (s, 1H), 6.06-6.09 (t, 1H), 7.38-7.42 (t, 2H), 7.70-7.73 (q, 2H), 8.51 (s, 1H). <sup>13</sup>C NMR (400 MHz, DMSO-d<sub>6</sub>) δ 34.84, 37.17, 41.09, 48.84, 62.17, 71.21, 84.98, 87.95, 115.93, 116.15, 131.70, 131.78, 147.42, 150.92, 157.13, 157.78, 158.38, 161.80, 162.26, 164.26. HRMS calcd for C<sub>19</sub>H<sub>22</sub>FN<sub>6</sub>O<sub>4</sub> [M+H<sup>+</sup>]: 417.1687, found: 417.1664. **b**) 0.16 g (42 %). <sup>1</sup>H NMR (300 MHz, DMSO-d<sub>6</sub>) δ 2.01-2.08 (m, 1H), 3.0 (s, 3H), 3.11 (s, 3H), 3.17-3.23 (m, 1H), 3.51-3.57 (m, 1H), 3.61-3.66 (m, 1H), 3.76-3.81 (m, 1H), 4.40-4.41 (d, 1H), 4.82-4.86 (t, 1H), 5.17-5.18 (d, 1H), 6.10-6.15 (t, 1H), 7.25-7.31 (m, 2H), 7.69-7.71 (d, 2H), 7.75-7.81 (m, 4H), 8.47 (s, 1H). <sup>13</sup>C NMR (101 MHz, DMSO-d<sub>6</sub>) δ 34.84, 37.17, 41.09, 48.84, 62.17, 71.21, 84.98, 87.95, 115.93, 116.15, 120.31, 126.89, 126.92, 131.69, 131.78, 147.42, 150.92, 157.13, 157.78, 158.38, 161.80, 164.26. HRMS calcd for C<sub>25</sub>H<sub>26</sub>FN<sub>6</sub>O<sub>4</sub> [M+H<sup>+</sup>]: 493.2000, found: 493.1969.

**5'-O-(4,4'-Dimethoxytrityl)-N<sup>2</sup>-(Dimethylformamidyl)-8-(aryl)-2'-deoxyguanosine (4a, 4b).** The title compound was synthesized by adaption of a previously described method.<sup>2</sup> In a round bottomed flask N<sup>2</sup>-(Dimethylformamidyl)-8-(aryl)-2'-deoxyguanosine **3a**) 0.23 g, 0.47 mmol, **3b**) 0.3 g, 0.6 mmol) was dissolved in anhydrous pyridine (10 mL) and the resulting solution was cooled to 0 °C. Subsequently, DMT-Cl (a) 0.190 g, 0.56

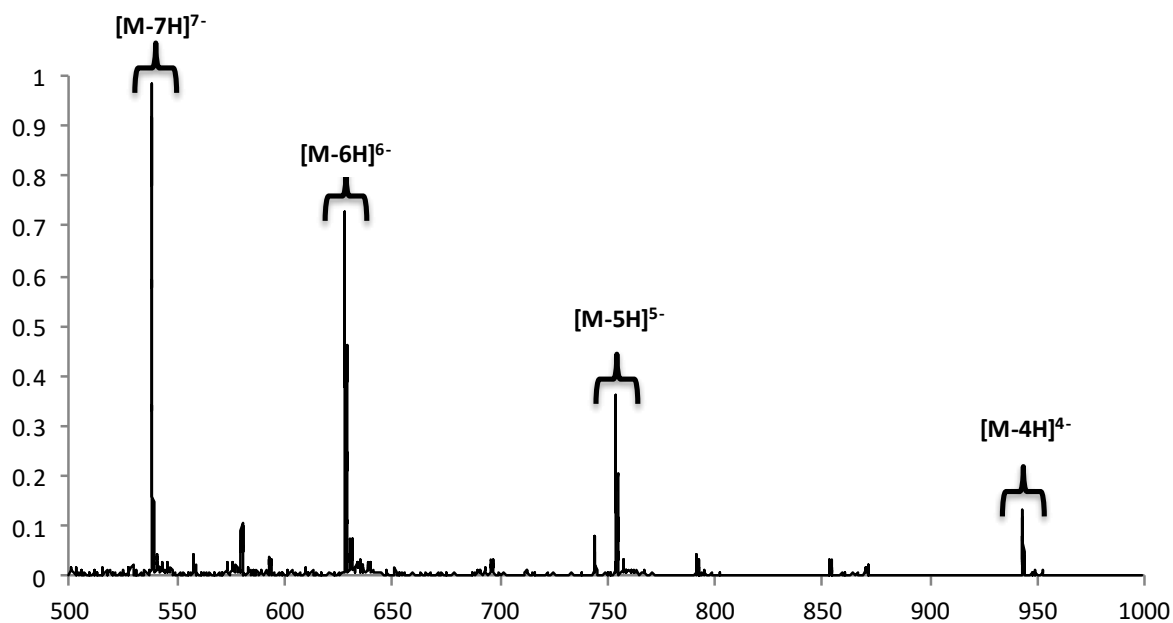
mmol, b) 0.25 g, 0.73 mmol) in pyridine (10 mL) was added dropwise. The mixture was allowed to react at room temperature under argon atmosphere. The reaction was monitored by TLC and after all starting material has been converted into the product (4 h), the resulting solution was diluted with CH<sub>2</sub>Cl<sub>2</sub> and washed with water (2x10 mL). TEA (a few drops) was added, and the mixture was evaporated to dryness. The crude mixture was purified on with silica column chromatography and eluted with MeOH:CH<sub>2</sub>Cl<sub>2</sub>:TEA (5:19:5) to yield product. **a)** 0.240 g (70 %) <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>) δ 2.22-2.31 (m, 1H), 3.06 (s, 3H), 3.11 (s, 3H), 3.20-3.25 (m, 1H), 3.31-3.33 (m, 1H), 3.45 (s, 1H), 3.77-3.78 (d, 6H), 4.00 (s, 1H), 4.70 (s, 1H), 5.38-5.40 (d, 1H), 6.19-6.23 dd, 1H), 6.79-6.85 (q, 4H), 7.24 (s, 2H), 7.25 (s, 5H), 7.37-7.48 (m, 4H), 7.86-7.91 (dd, 2H), 8.39 (s, 1H), 11.52 (bs, 1H). <sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 35.08, 37.72, 41.29, 55.23, 64.24, 72.74, 83.74, 84.90, 86.33, 113.09, 115.48, 115.70, 120.45, 126.32, 126.35, 126.90, 127.85, 128.06, 129.96, 129.99, 131.71, 131.79, 135.78, 135.81, 144.72, 148.80, 151.03, 155.78, 157.79, 157.81, 158.50, 162.32, 164.80. HRMS calcd for C<sub>40</sub>H<sub>40</sub>FN<sub>6</sub>O<sub>6</sub> [M+H<sup>+</sup>]: 719.2993, found: 719.2936 **b)** 0.3 g (62 %) <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 2.19-2.26 (m, 1H), 2.81 (s, 3H), 2.93 (s, 3H), 3.19-3.25 (m, 1H), 3.34-3.38 (m, 1H), 3.49-3.53 (m, 1H), 3.69 (s, 6H), 3.99-4.04 (q, 1H), 4.80-4.85 (q, 1H), 6.21-6.24 (q, 1H), 6.68-6.72 (dd, 4H), 7.08-7.18 (m 5H), 7.23-7.26 (dd, 4H), 7.35-7.37 (d, 2H), 7.50-7.53 (m, 4H), 7.82-7.84 (d, 2H), 8.31 (s, 1H), 9.14-9.58 (bs, 1H). <sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 35.04, 37.84, 41.29, 55.20, 64.38, 72.60, 83.94, 85.11, 86.26, 113.05, 115.68, 115.89, 120.60, 128.12, 128.69, 128.77, 130.02, 130.11, 135.87, 135.93, 136.29, 136.32, 141.21, 144.79, 149.77, 151.16, 155.74, 157.74, 158.43, 161.47, 163.93. HRMS calcd for C<sub>46</sub>H<sub>44</sub>FN<sub>6</sub>O<sub>6</sub> [M+H<sup>+</sup>]: 795.3306, found: 795.3245.

**3'-O-[(2-Cyanoethoxy)(diisopropylamino)phosphine]-5'-O-(4,4'-Dimethoxytrityl)-N<sup>2</sup>-(Dimethylformamidyl)-8-(aryl)-2'-deoxyguanosine (5).** The title compound was

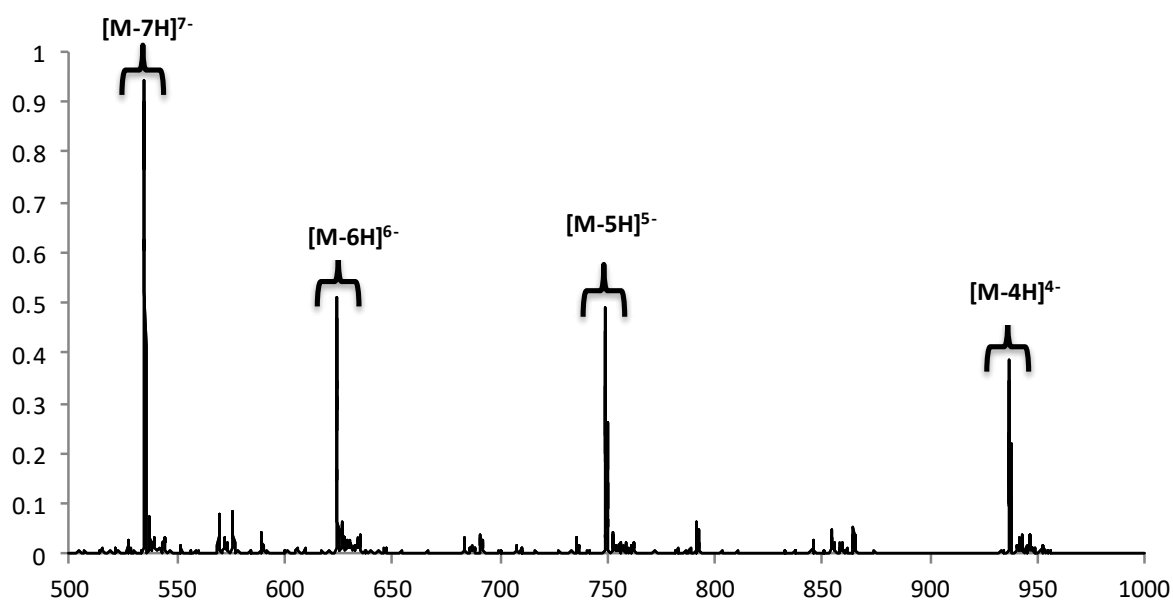
synthesized by adaption of a previously described method.<sup>2</sup> 5'-O-(4,4'-Dimethoxytrityl)-N<sup>2</sup>-(Dimethylformamidyl)-8-(aryl)-2'-deoxyguanosine **4a**) 0.23 g, 0.3 mmol, **4b**) 0.25 g, 0.3 mmol) was dissolved in anhydrous CH<sub>2</sub>Cl<sub>2</sub> (10 mL) and anhydrous TEA (a) 0.2 mL, 1.3 mmol b) 0.2 mL, 1.4 mmol), subsequently 2-cyanoethyl-N,N-diisopropylchlorophosphoramidite (0.08 mL, 0.38 mmol) was added. The reaction was monitored by TLC and after all starting material has been converted into the product (30 min) the mixture was added dropwise to ice cold hexane and washed with hexane to yield product a) 0.13 g (45 %)<sup>44</sup> **a**) 0.97-1.02 (m, 6H), 1.06-1.10 (m, 6H), 2.25-2.41 (m, 2H), 2.69-2.71 (m, 3H), 2.89 (s, 3H), 3.02-3.03 (d, 3H), 3.43-3.49 (m, 5H), 3.68-3.69 (m, 7H), 4.11 (s, 2H), 4.9-5.14 (d, 1H), 6.1 (s, 1H), 6.65-6.70 (m, 4H), 7.07-7.15 (m, 9H), 7.22-7.35 (2H), 7.83 (2H), 8.3-8.4 (d, 1H), 10.1 (s, 1H) <sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 8.9, 21.43, 24.5, 24.5, 35.20, 35.23, 37.14, 41.19, 41.26, 41.31, 43.23, 43.34, 43.35, 55.28, 55.24, 62.95, 63.55, 73.14, 73.36, 74.30, 74.49, 8.78, 84.86, 84.93, 112.98, 113.02, 115.49, 115.54, 115.70, 115.76, 117.09, 117.63, 117.73, 120.69, 120.70, 126.30, 126.33, 128.07, 128.13, 129.93, 130.00, 130.06, 131.70, 131.73, 131.79, 135.92, 144.76, 144.86, 148.82, 148.85, 155.92, 155.99, 157.83, 157.88, 158.25, 162.31, 162.33, 164.82. HRMS calcd for C<sub>49</sub>H<sub>57</sub>FN<sub>8</sub>O<sub>7</sub>P [M+H<sup>+</sup>]: 919.4072, found: 919.4013.

**b**) 0.18 g (75 %). <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>) δ 0.78-0.97 (m, 12H), 2.22-2.35 (m, 3H), 2.89 (s, 3H), 3.02 (s, 3H), 3.23-3.78 (m, 12H), 4.05-4.19 (m, 1H), 4.95-5.14 (m, 1H), 6.18-6.25 (m, 1H), 6.65-6.72 (m, 4H), 7.09-7.19 (m, 6H), 7.24-7.29 (m, 4H), 7.36-7.39 (m, 2H), 7.53-7.62 (d, 4H), 7.89-7.98 (m, 2H), 8.31-8.39 (d, 1H), 8.68-8.69 (bs, 1H). <sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 19.63, 20.26, 21.59, 22.7, 24.23, 24.34, 35.20, 36.79, 40.89, 41.94, 43.00, 43.13, 52.76, 55.10, 57.58, 58.29, 60.10, 63.09, 72.80, 74.16, 84.00, 85.84, 112.82, 115.56, 115.78, 116.49, 116.711, 117.52, 117.66. 120.49, 126.61, 126.85, 127.61, 127.89, 128.54,

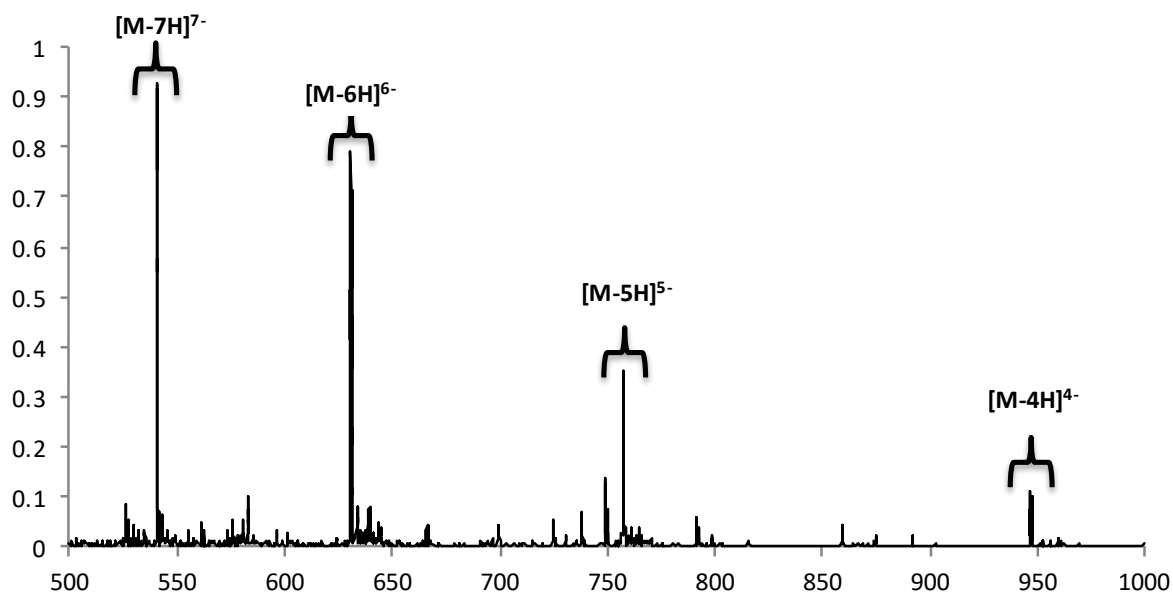
128.62, 129.81, 135.63, 136.02, 141.11, 144.46, 144.71, 149.27, 150.70, 150.99, 155.71, 155.74, 155.90, 157.66, 157.79, 158.19, 161.29, 163.74.  $^{31}\text{P}$  NMR (300 MHz,  $\text{CDCl}_3$ )  $\delta$  149.34, 148.99 (73.53 %), 14.29 (27.47 %). HRMS calcd for  $\text{C}_{55}\text{H}_{61}\text{FN}_8\text{O}_7\text{P}$   $[\text{M}+\text{H}]^+$ : 995.4385, found: 995.4304.



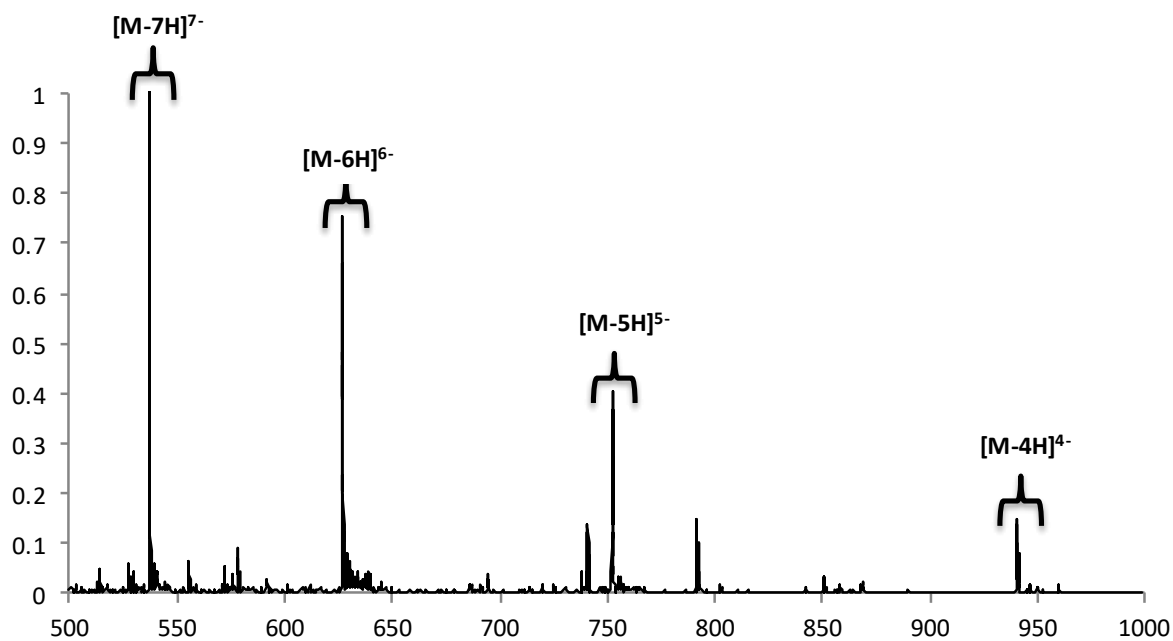
**Figure D.3:** ESI Spectrum of NarI (N=A, X=FBPG).



**Figure D.4:** ESI<sup>-</sup> Spectrum of *NarI* (N=C, X=<sup>FBP</sup>G).

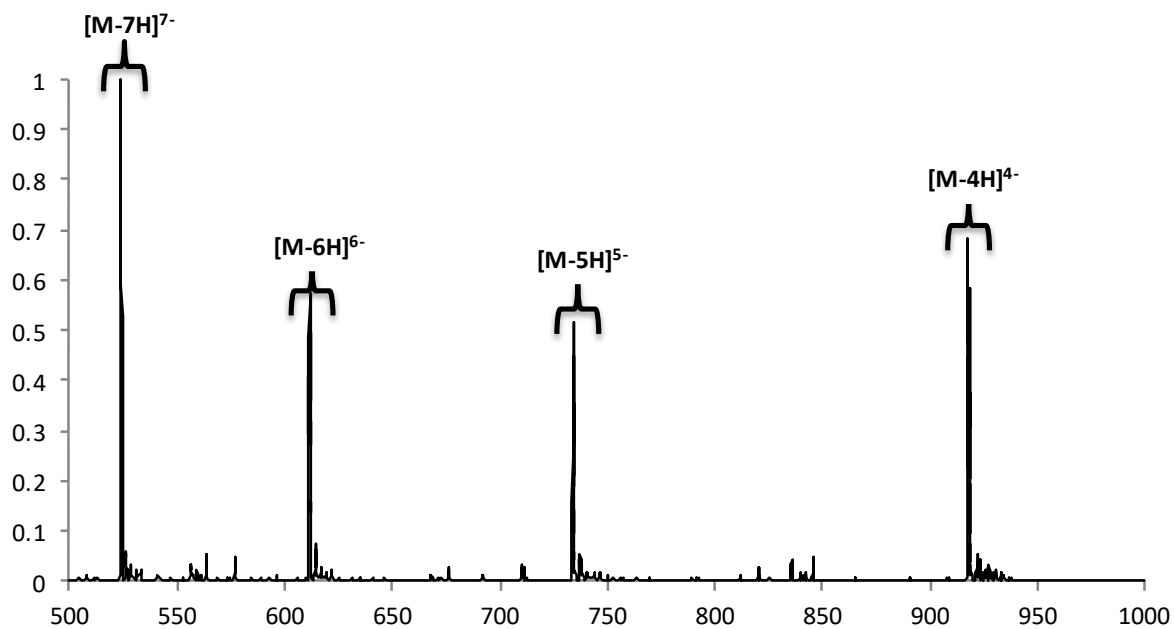


**Figure D.5:** ESI<sup>-</sup> Spectrum of *NarI* (N=G, X=<sup>FBP</sup>G).



**Figure D.6:** ESI<sup>-</sup> Spectrum of *NarI* (N=T, X=<sup>FBP</sup>G).

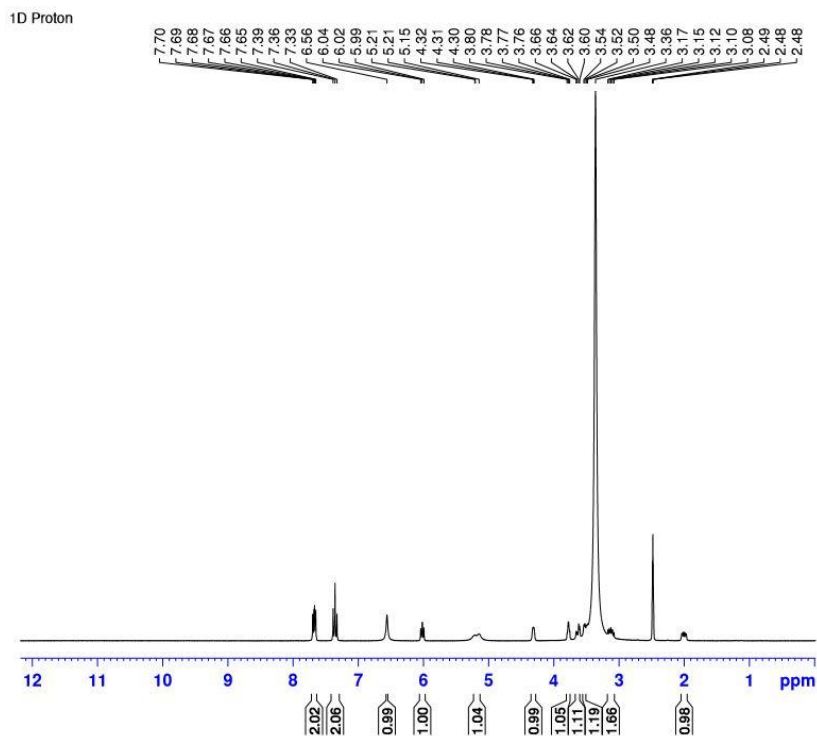




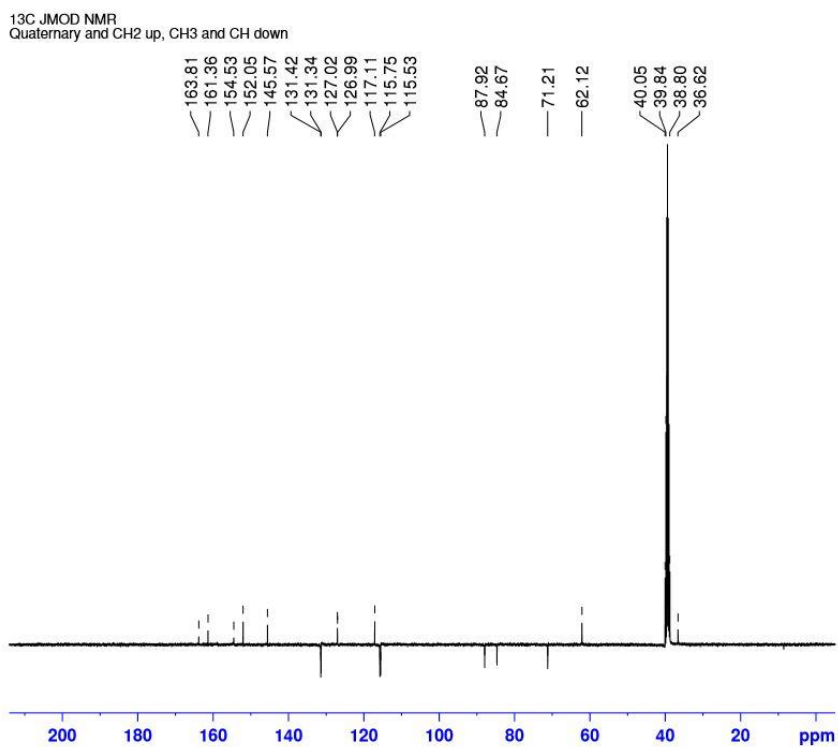
**Figure D.7:** ESI<sup>-</sup> Spectrum of *NarI* (N=C, X=<sup>FPhG</sup>).

**Table D.7.** ESI-MS Analysis of C8-Aryl-dG Modified *NarI* (N,X) oligonucleotides.:

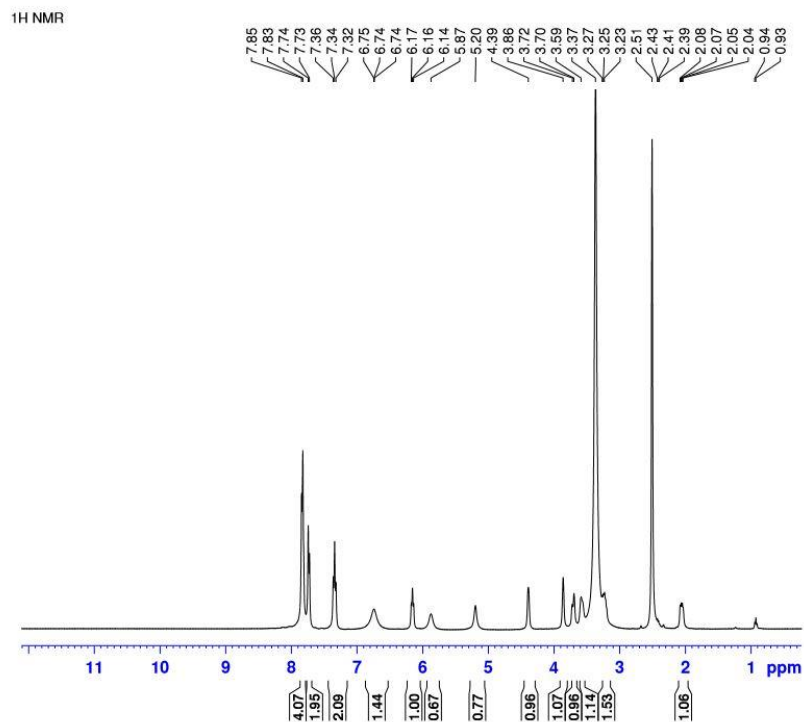
| Oligonucleotide                        | Formula                                                                            | Calc Mass | Exptl m/z            |        | Exptl Mass |
|----------------------------------------|------------------------------------------------------------------------------------|-----------|----------------------|--------|------------|
| <i>NarI</i> (N=A, X= <sup>FBP</sup> G) | C <sub>127</sub> H <sub>154</sub> FN <sub>44</sub> O <sub>70</sub> P <sub>11</sub> | 3774.64   | [M-7H] <sup>7-</sup> | 538.28 | 3774.96    |
|                                        |                                                                                    |           | [M-6H] <sup>6-</sup> | 628.16 | 3774.96    |
|                                        |                                                                                    |           | [M-5H] <sup>5-</sup> | 753.97 | 3774.96    |
|                                        |                                                                                    |           | [M-4H] <sup>4-</sup> | 942.69 | 3774.76    |
| <i>NarI</i> (N=C, X= <sup>FBP</sup> G) | C <sub>126</sub> H <sub>154</sub> FN <sub>42</sub> O <sub>71</sub> P <sub>11</sub> | 3750.68   | [M-7H] <sup>7-</sup> | 534.88 | 3751.16    |
|                                        |                                                                                    |           | [M-6H] <sup>6-</sup> | 624.16 | 3750.96    |
|                                        |                                                                                    |           | [M-5H] <sup>5-</sup> | 749.22 | 3751.11    |
|                                        |                                                                                    |           | [M-4H] <sup>4-</sup> | 936.75 | 3751.00    |
| <i>NarI</i> (N=G, X= <sup>FBP</sup> G) | C <sub>127</sub> H <sub>154</sub> FN <sub>44</sub> O <sub>71</sub> P <sub>11</sub> | 3790.96   | [M-7H] <sup>7-</sup> | 540.56 | 3790.92    |
|                                        |                                                                                    |           | [M-6H] <sup>6-</sup> | 630.89 | 3791.34    |
|                                        |                                                                                    |           | [M-5H] <sup>5-</sup> | 757.19 | 3789.95    |
|                                        |                                                                                    |           | [M-4H] <sup>4-</sup> | 946.47 | 3789.88    |
| <i>NarI</i> (N=T, X= <sup>FBP</sup> G) | C <sub>127</sub> H <sub>155</sub> FN <sub>41</sub> O <sub>72</sub> P <sub>11</sub> | 3767.56   | [M-7H] <sup>7-</sup> | 537.00 | 3766.00    |
|                                        |                                                                                    |           | [M-6H] <sup>6-</sup> | 626.66 | 3765.96    |
|                                        |                                                                                    |           | [M-5H] <sup>5-</sup> | 752.19 | 3765.95    |
|                                        |                                                                                    |           | [M-4H] <sup>4-</sup> | 940.44 | 3765.76    |
| <i>NarI</i> (N=C, X= <sup>FPh</sup> G) | C <sub>120</sub> H <sub>150</sub> FN <sub>42</sub> O <sub>71</sub> P <sub>11</sub> | 3674.65   | [M-7H] <sup>7-</sup> | 524.00 | 3675.00    |
|                                        |                                                                                    |           | [M-6H] <sup>6-</sup> | 611.5  | 3675.00    |
|                                        |                                                                                    |           | [M-5H] <sup>5-</sup> | 734.00 | 3574.00    |
|                                        |                                                                                    |           | [M-4H] <sup>4-</sup> | 917.24 | 3674.77    |



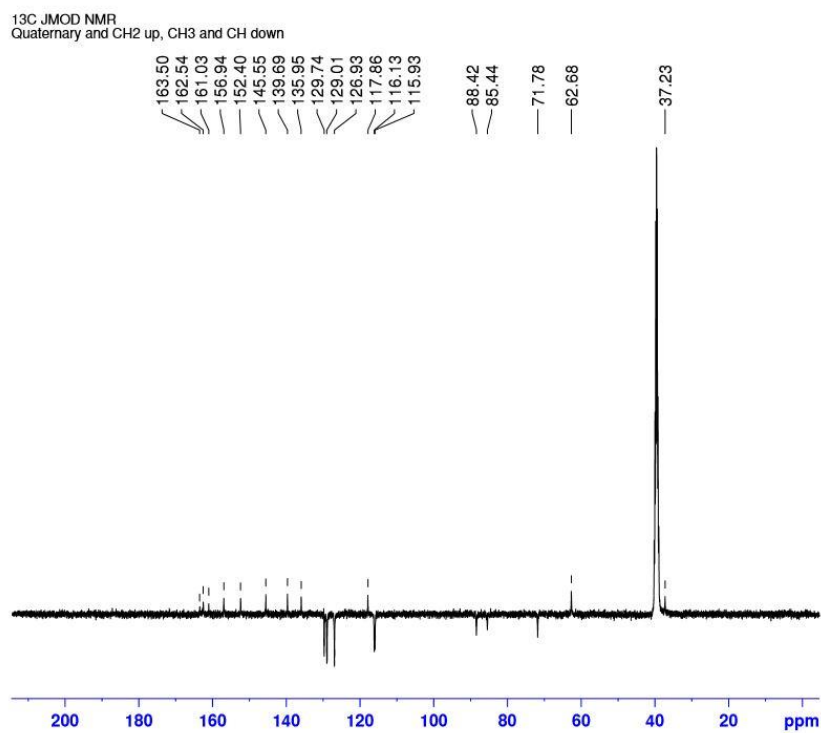
**Figure D.8:**  $^1\text{H}$  NMR of 2a in  $\text{DMSO-d}_6$ .



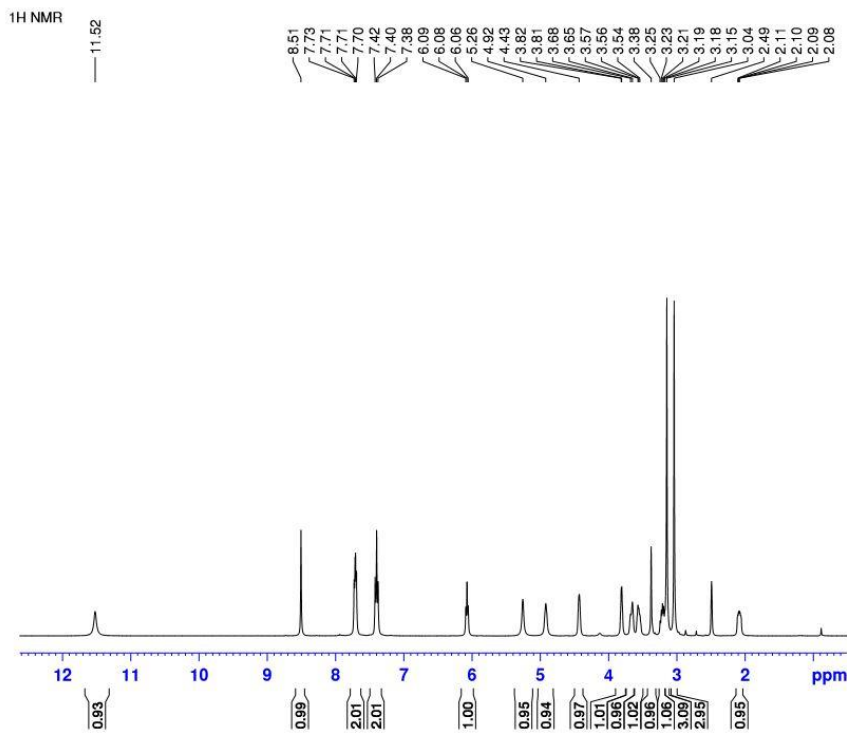
**Figure D.9:**  $^{13}\text{C}$  NMR (DEPT-135) of 2a in  $\text{DMSO-d}_6$ .



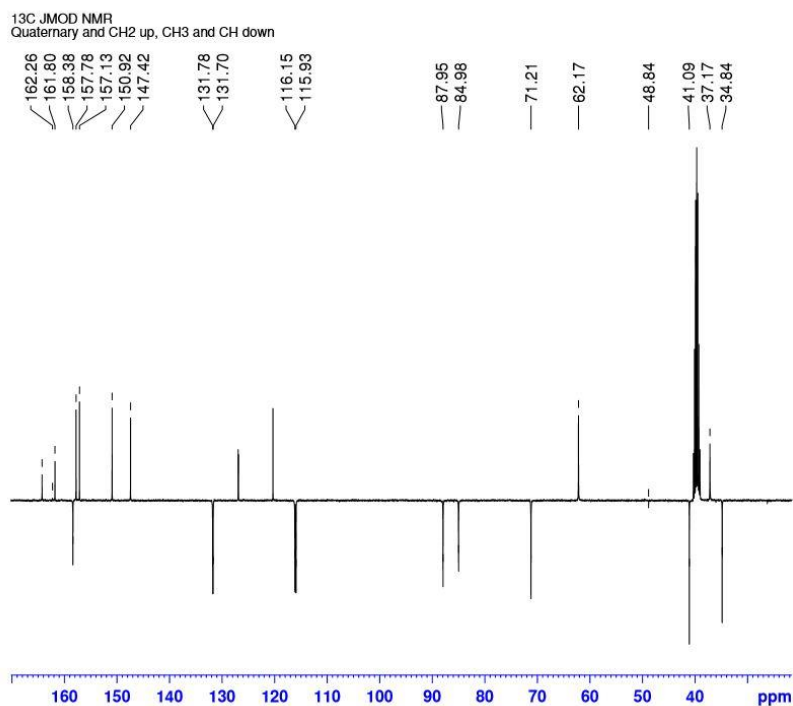
**Figure D.10:** <sup>1</sup>H NMR of 2b in DMSO-d<sub>6</sub>.



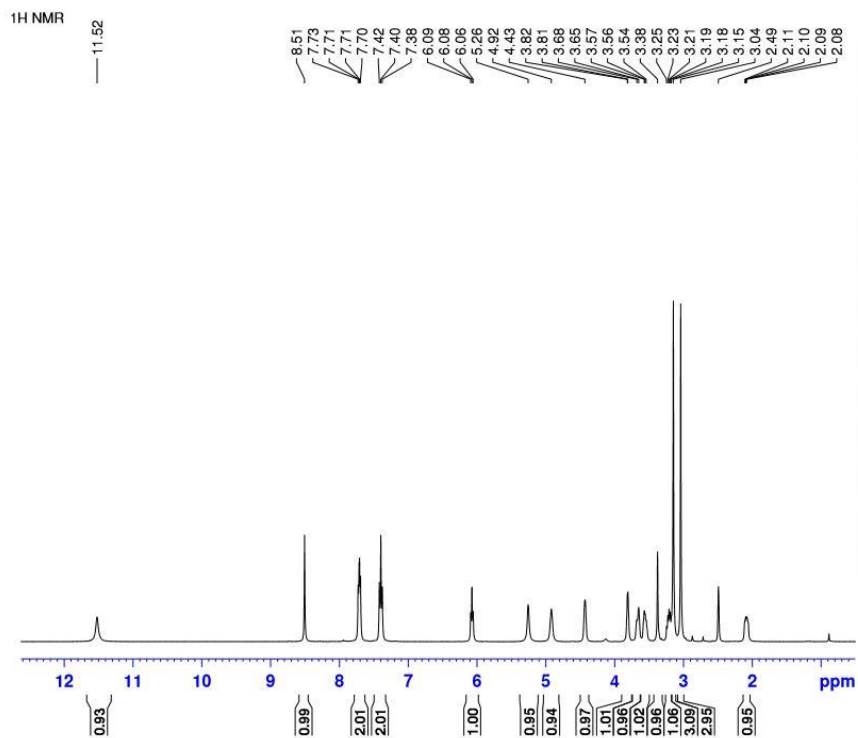
**Figure D.11:** <sup>13</sup>C NMR (DEPT-135) of 2b in DMSO-d<sub>6</sub>.



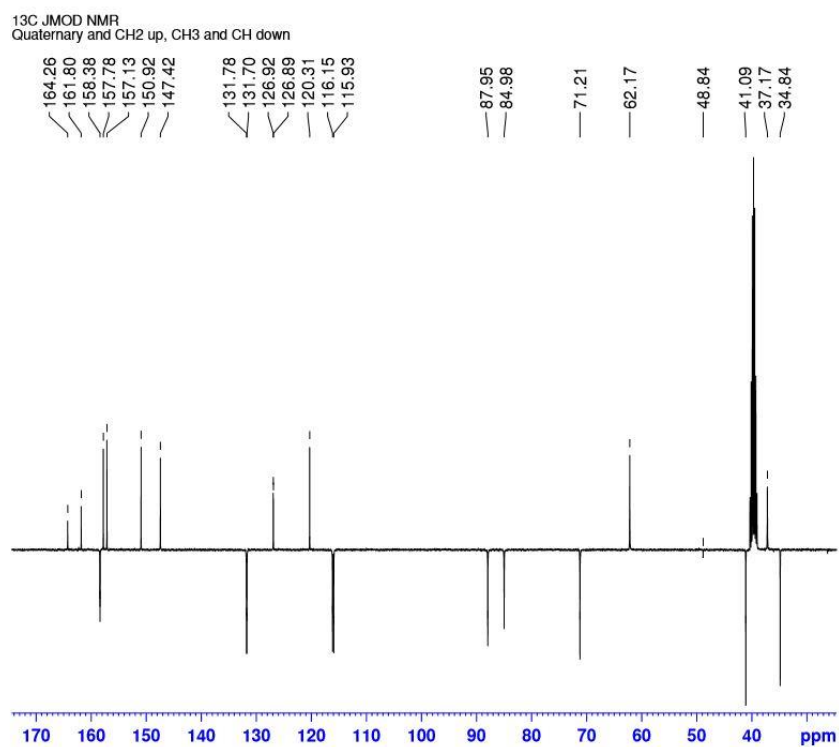
**Figure D.12:** <sup>1</sup>H NMR of 3a in DMSO-d<sub>6</sub>.



**Figure D.13:** <sup>13</sup>C NMR (DEPT-135) of 3a in DMSO-d<sub>6</sub>.



**Figure D.14:** <sup>1</sup>H NMR of 3b in DMSO-d<sub>6</sub>.



**Figure D.15:** <sup>13</sup>C NMR (DEPT-135) of 3b in DMSO-d<sub>6</sub>.

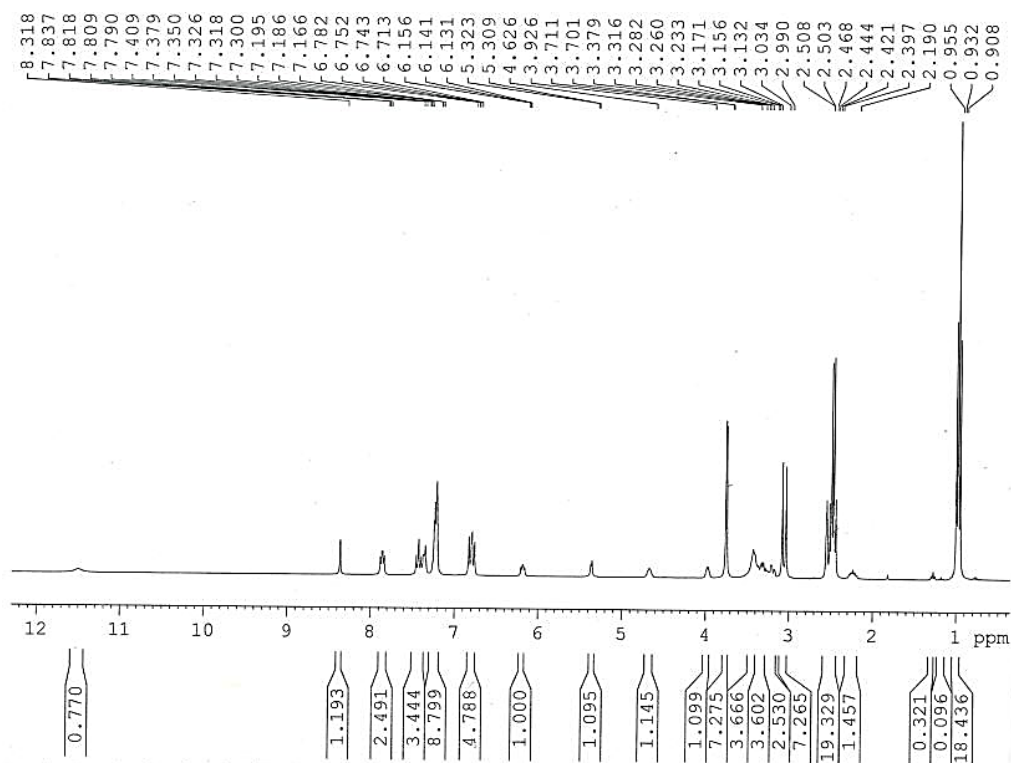


Figure D.16:  $^1\text{H}$  NMR of 4a in  $\text{CDCl}_3$ .

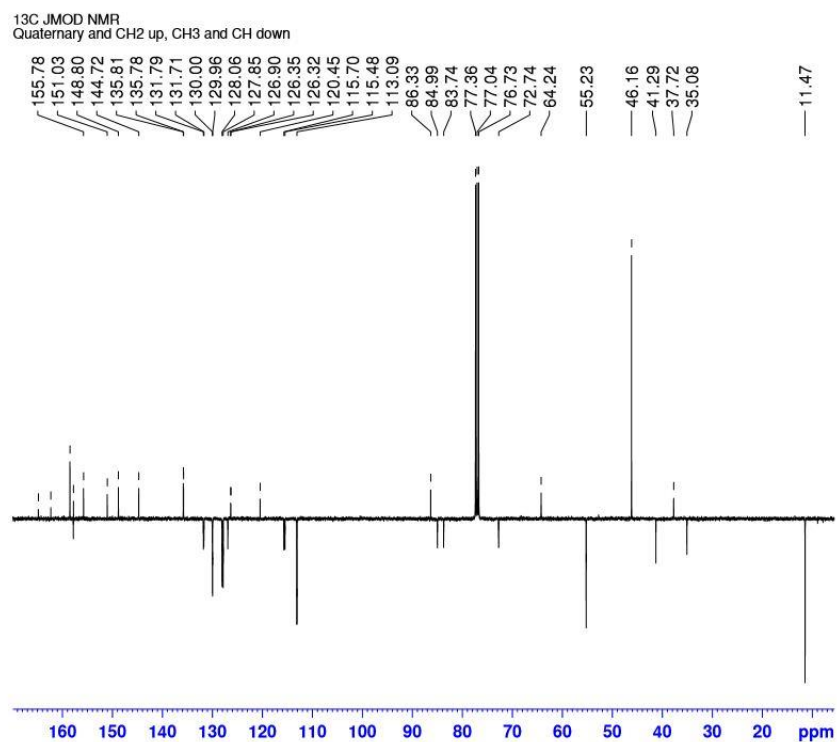


Figure D.17:  $^{13}\text{C}$  NMR (DEPT-135) of 4a in  $\text{CDCl}_3$ .

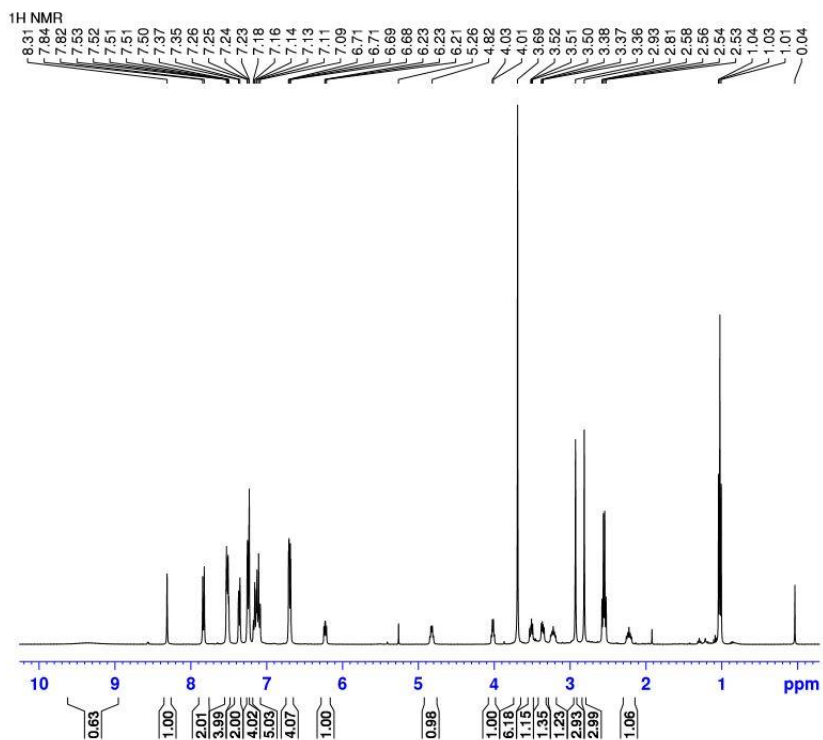


Figure D.18: <sup>1</sup>H NMR of 4b in CDCl<sub>3</sub>.

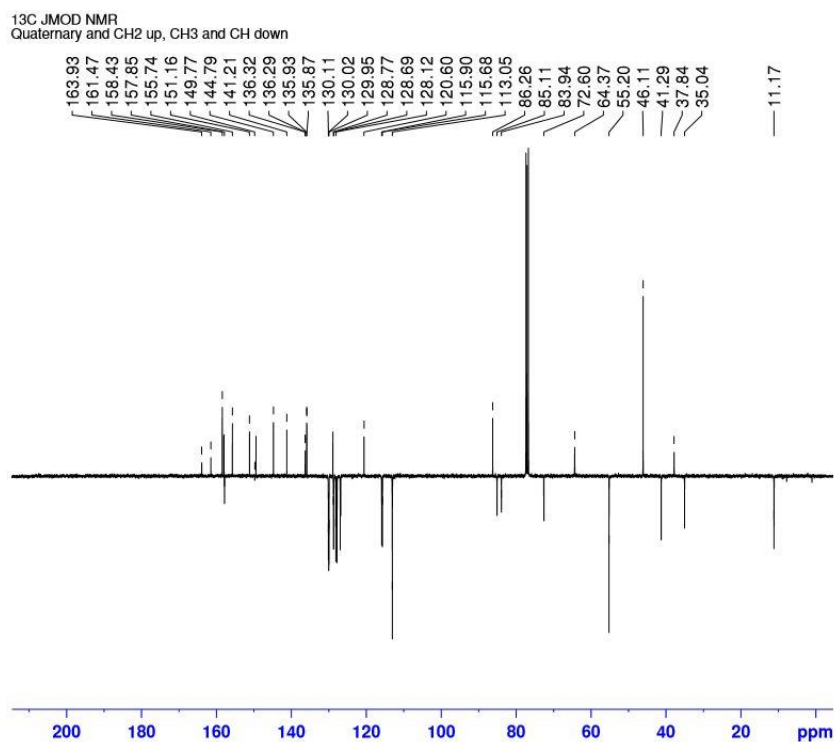
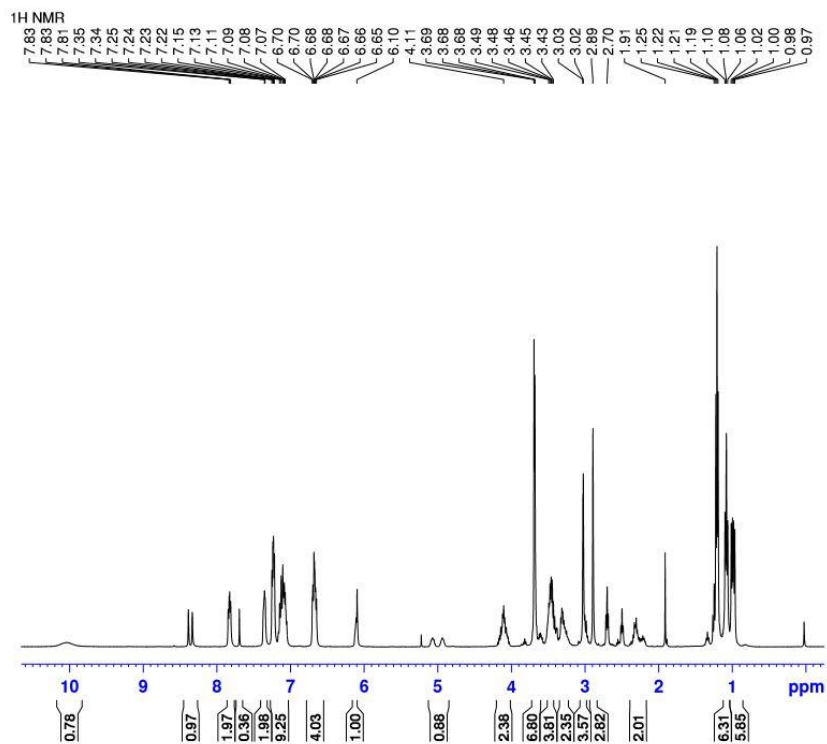
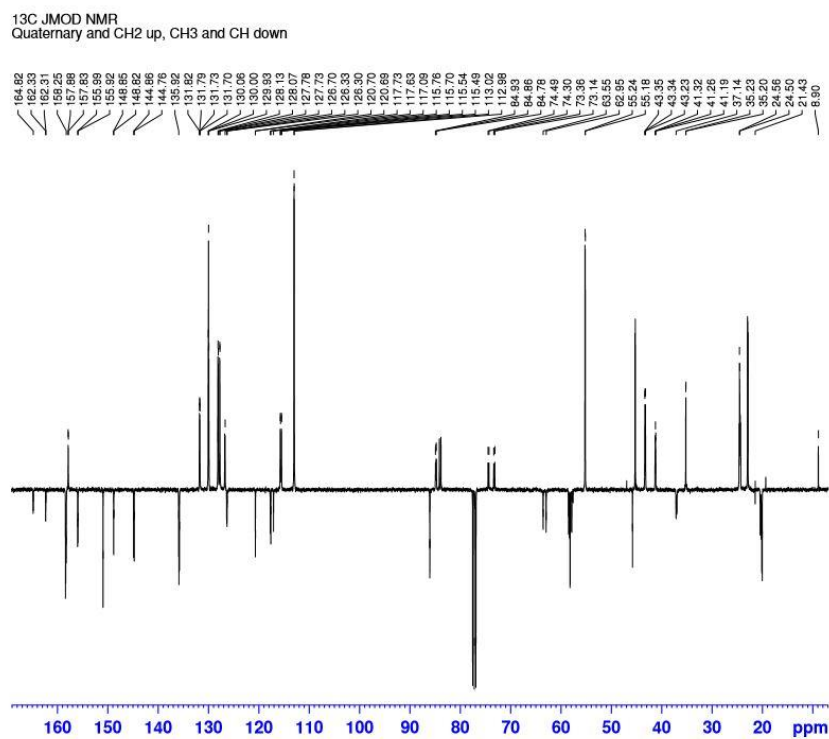


Figure D.19: <sup>13</sup>C NMR (DEPT-135) of 4b in CDCl<sub>3</sub>.





**Figure D.20:** <sup>1</sup>H NMR of 5a in CDCl<sub>3</sub>.



**Figure D.21:** <sup>13</sup>C NMR (DEPT-135) of 5a in CDCl<sub>3</sub>.

<sup>31</sup>P NMR

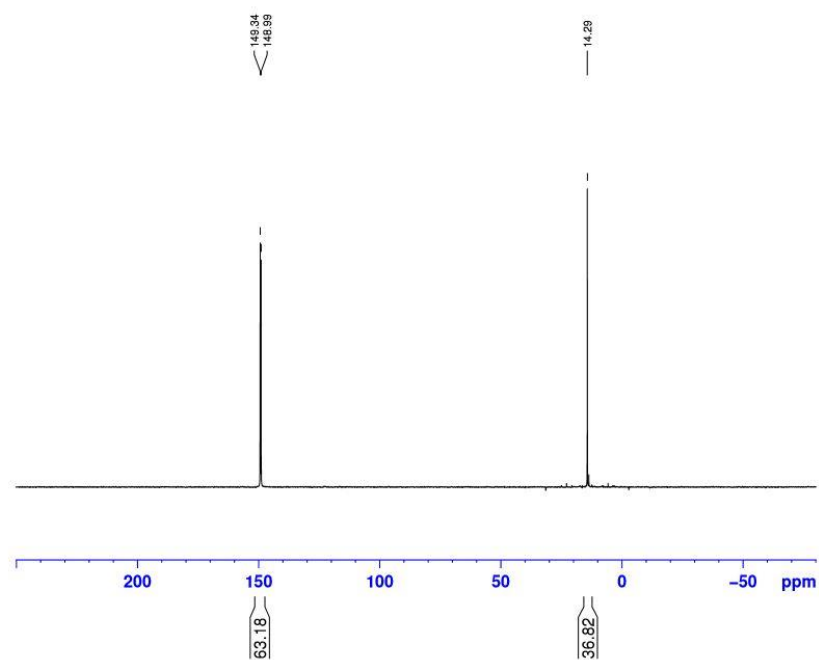


Figure D.22: <sup>31</sup>P NMR of 5a in CDCl<sub>3</sub>.

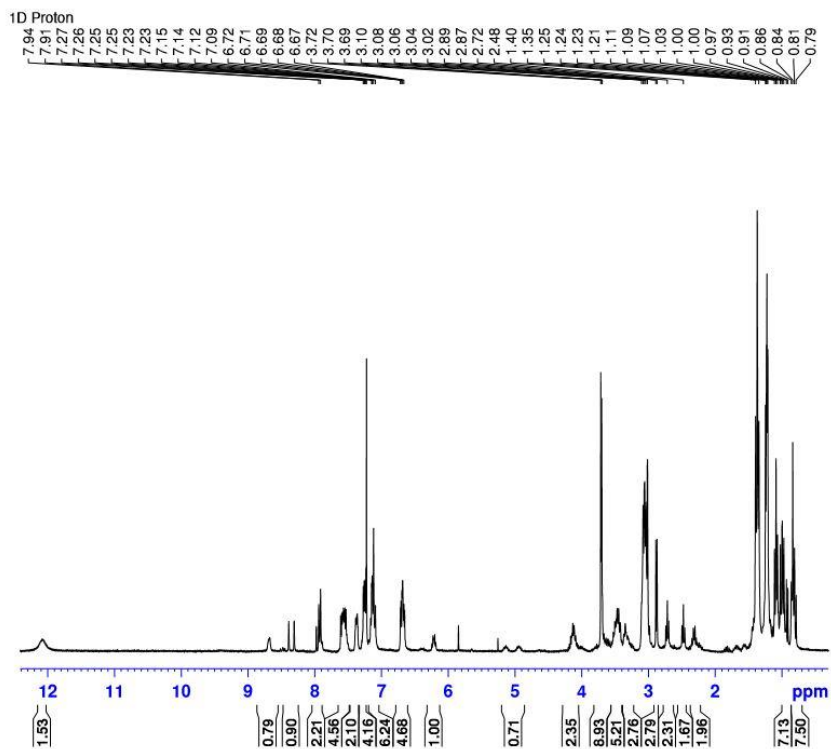
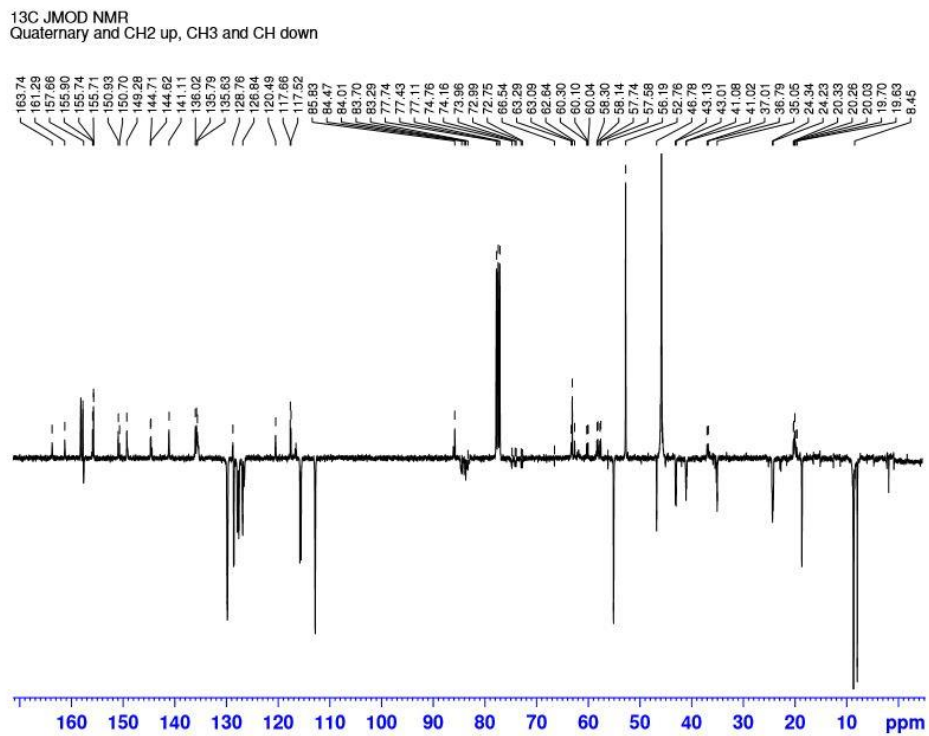
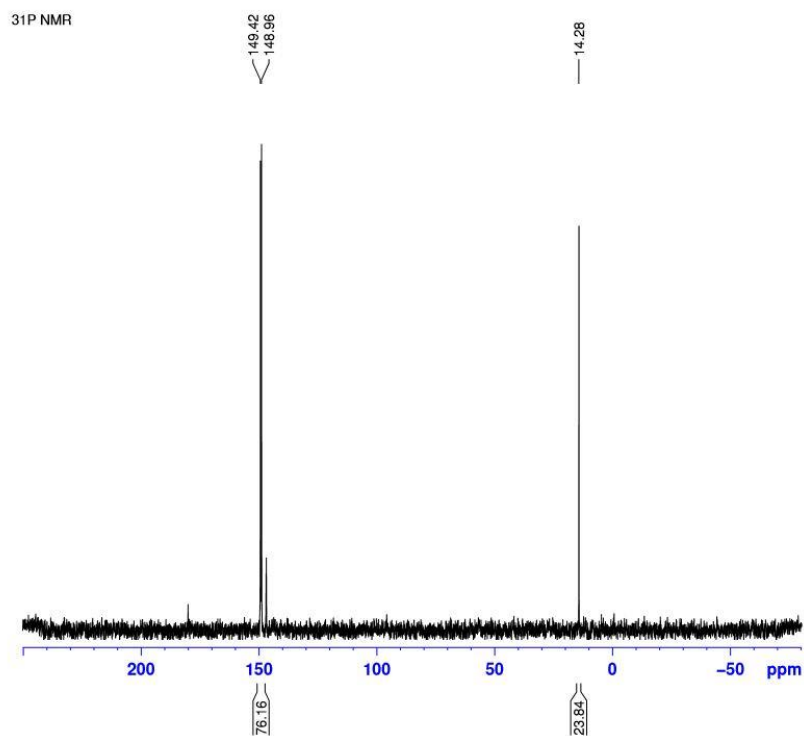


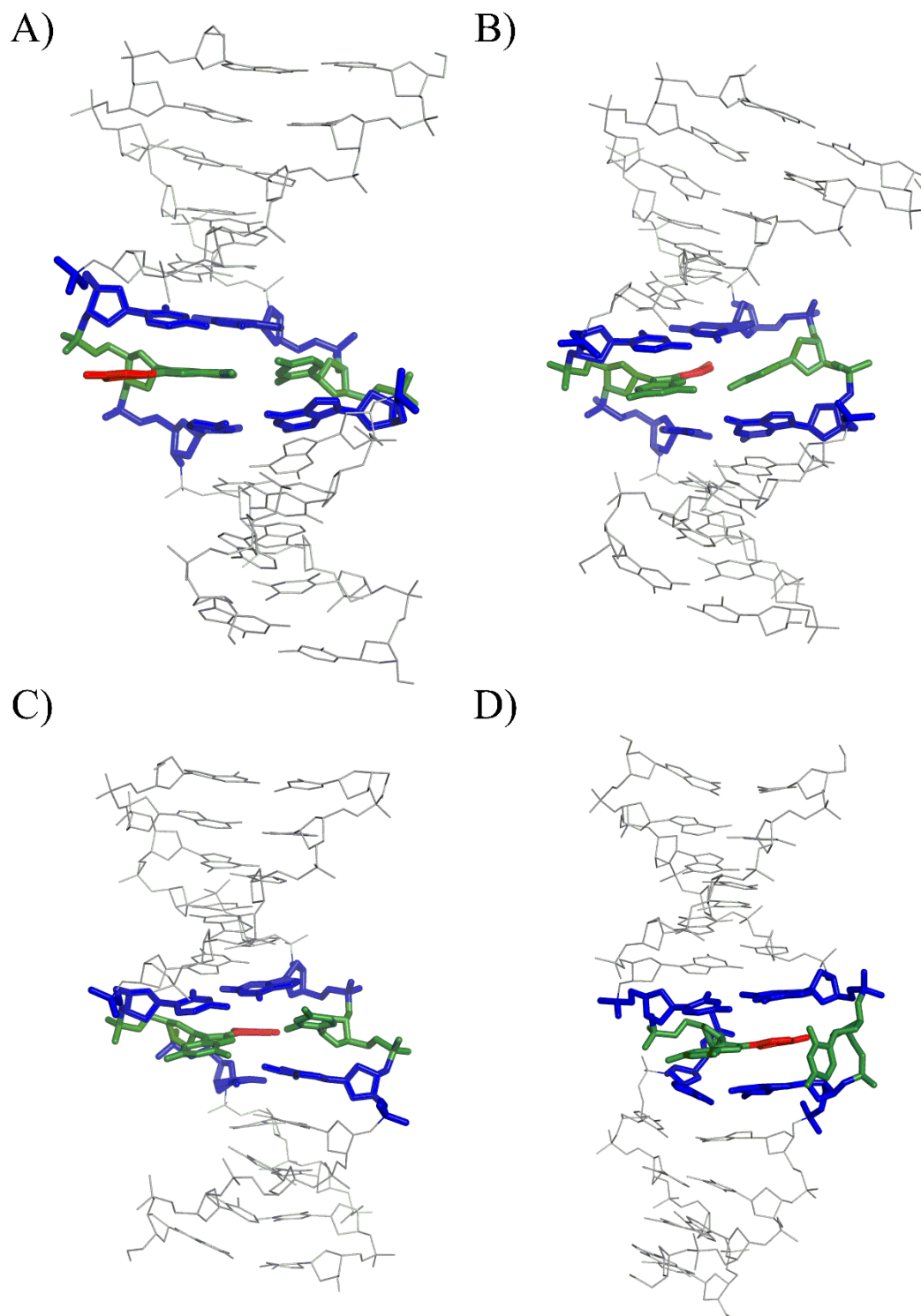
Figure D.23: <sup>1</sup>H NMR of 5b in CDCl<sub>3</sub>.



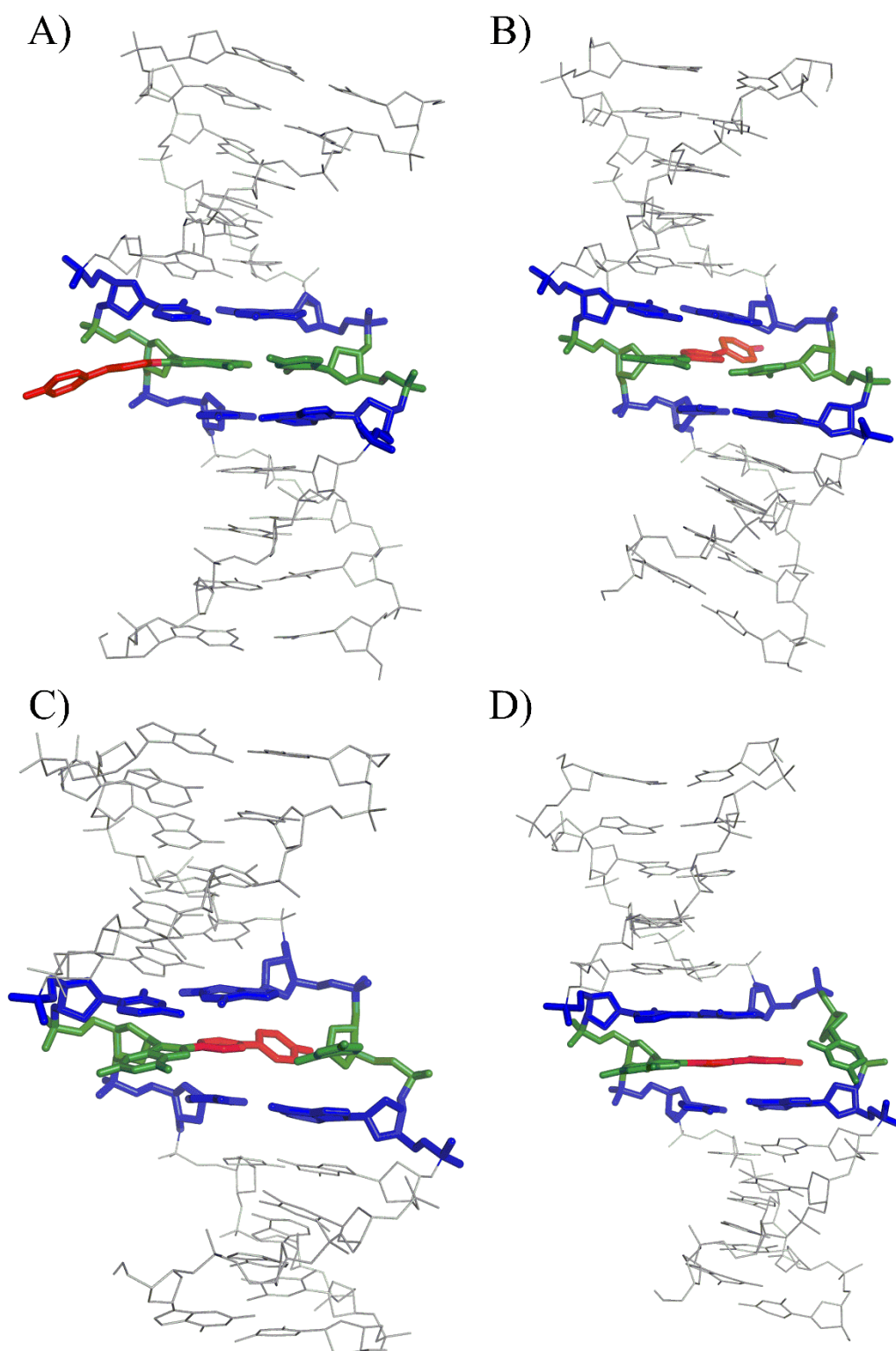
**Figure D.24:** <sup>13</sup>C NMR (DEPT-135) of 5b in CDCl<sub>3</sub>.



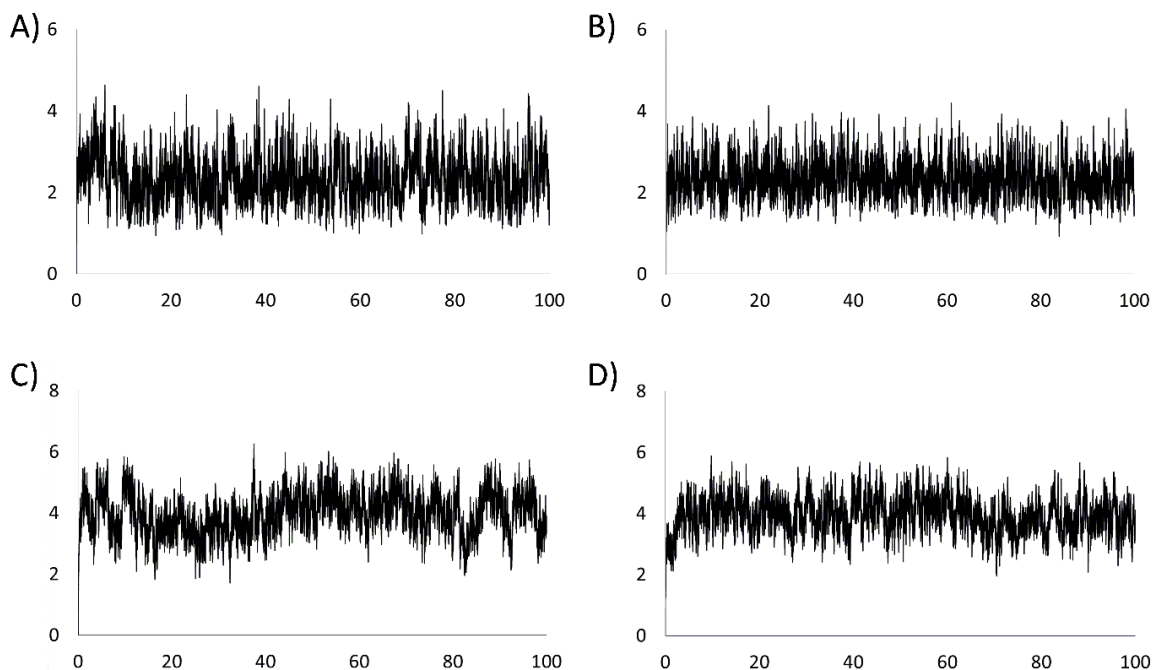
**Figure D.25:** <sup>31</sup>P NMR of 5b in CDCl<sub>3</sub>.



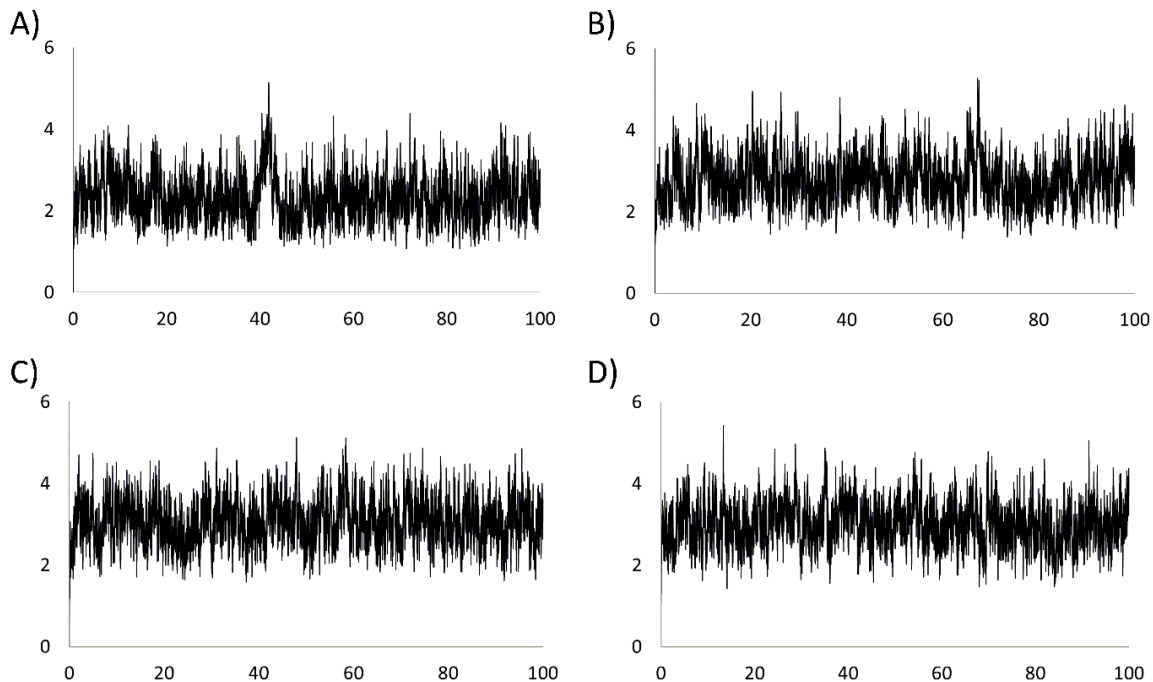
**Figure D.26:** The A) major-groove, B) wedge, C) intercalated, and D) base-displaced intercalated conformations of FPh-dG adducted *NarI* (N = C). The damaged G and opposing C in green, the bulky moiety in red, and flanking base pairs in blue.



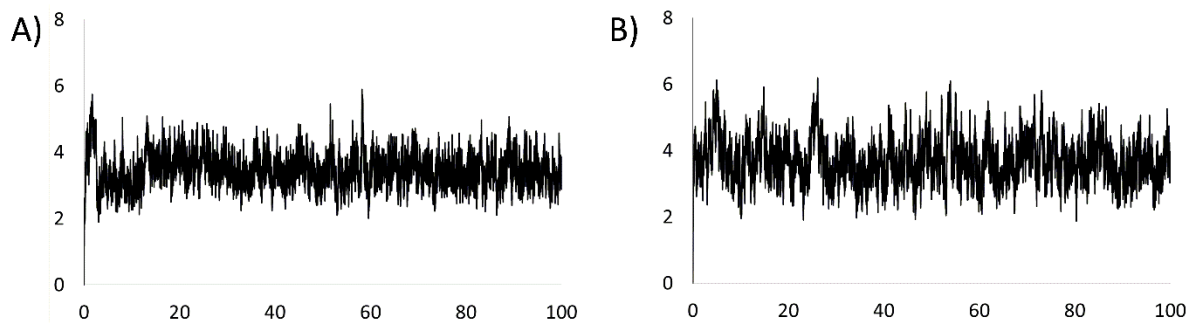
**Figure D.27:** The A) major-groove, B) wedge, C) intercalated, and D) base-displaced intercalated conformations of FBP-dG adducted *NarI* (N = C). The damaged G and opposing C in green, the bulky moiety in red, and flanking base pairs in blue.



**Figure D.28:** The helical backbone root mean square deviation (RMSD, Å) over the 100 ns production trajectories of the A) major-groove, B) wedge, C) intercalated, and D) base-displaced intercalated conformations of FPh-dG adducted fully paired *NarI* ( $N = C$ ).



**Figure D.29:** The helical backbone root mean square deviation (RMSD, Å) over the 100 ns production trajectories of the A) major-groove, B) wedge, C) intercalated, and D) base-displaced intercalated conformations of FBP-dG adducted fully paired *NarI* ( $N = C$ ).



**Figure D.30:** The helical backbone root mean square deviation (RMSD, Å) over the 100 ns production trajectories of the A) *anti*, and B) *syn* conformations of FBP-dG adducted SMI *NarI* (N = C).

**Molecular Dynamics Simulations.** Both the F-phenyl-dG (<sup>FPh</sup>dG) and F-biphenyl-dG (<sup>FBP</sup>dG) adducts were studied in the 5′-CTCGGCXCCATC-3′ 12-mer containing the *NarI* sequence (underlined), with X being the fluorine-labeled adduct. The <sup>FBP</sup>dG adduct was also studied in a two-base bulge helix using the same primary 12-mer strand paired with the 5′-GAGCCGGTAG-3′ 10-mer strand as done previously for other DNA adducts.<sup>1</sup> The natural nucleotides were treated with the amber14SB force field, while GAFF parameters were used to supplement the amber14SB force field for the adducts. Antechamber of the AMBER toolset was used to assign atom types, while R.E.D. version III.4<sup>3</sup> was used to calculate the partial charges of the damaged residues based on B3LYP/6-31G(d) optimized structures of the *anti* and *syn* conformers. Specifically, the R.E.D. program uses the initial structures to calculate the molecular electrostatic potential (MEP) surfaces at the HF/6-31G\* level of theory, and then fits the atomic charges to the MEP surfaces according to a restrained electrostatic potential. Importantly, the charges generated using the R.E.D. program are fully compatible with the Amber force field.<sup>3</sup> Although there are not suitable experimental quantities to test our force field against for these systems since the only experimental data available is presented in this work, the agreement between NMR and MD data provides confidence that our approach is reasonable. Molecular dynamics (MD) simulations were run on the adducted DNA duplexes using the PMEMD module of AMBER 14.

For the fully paired helices, initial structures were generated that correspond to the B-type (major-groove), W-type (wedge) and S-type (base-displaced intercalated) conformations discussed in the literature for N-linked C8-dG adducts. Specifically, the B-type conformer places the bulky moiety within the major groove of the helix, while the W-type conformer places the bulky moiety in the minor groove. In contrast, the S



conformation intercalates the bulky moiety within the helix and displaces the opposing base into the major groove. These conformers were generated by modifying representative structures of previously studied C8-aryl-dG adducts<sup>4</sup> using GaussView 5.0. For the B and W conformations, the  $\chi$  dihedral angle ( $\chi = \angle(\text{O4}'\text{C1}'\text{N9C4})$ ) for each conformation initially considered was selected based on the value in the modified representative structure, while maintaining interactions between the damaged G and opposing C. For the S conformer, the  $\chi$  dihedral was selected to keep the bulky group planar with respect to the flanking bases since the adduct has limited flexibility due to direct attachment of the aromatic group to C8 of dG. From these starting conformations, multiple MD simulations were performed for shorter timescales (20 ns) with bulky moiety orientations that differ in the relative orientation with respect to the damaged nucleobase ( $\theta = \angle(\text{N9C8C10C11})$ , Figure D.2). Specifically,  $\theta$  was initially set to  $0^\circ$ ,  $30^\circ$  or  $60^\circ$  in order to enhance sampling of possible bulky moiety orientations, while the dihedral angle between the two rings of the biphenyl system ( $\xi = \angle(\text{C12C3C16C17})$ , Figure D.2) was set equal to  $0^\circ$ , which was justified based on the sampling of many values over the production runs (Figure 5). It is worth noting that because of the lack of flexibility in the C–C linker, the bulky group cannot be directed in the 5' or 3' direction within a groove as reported in the literature for N-linked C8-dG adducts. In cases where these preliminary simulations led to new conformations, the corresponding representative structures were then used as initial coordinates for further simulations in an iterative process. Free energies estimated from the molecular mechanics-Poisson Boltzmann surface area (MM-PBSA) approach were then used to determine the most favorable structures. Through this cycle, the intercalated conformation (I) was isolated where the bulky moiety and opposing C are both intercalated within the DNA duplex. The

lowest energy structures obtained from the iterative phase sampling simulations were then used for the 100 ns production MD simulations described below. Similarly, preliminary conformations for the bulged structures were built based on the X-ray structure of the AF-dG adduct opposite a -2 deletion site (PDB ID: 1AX6), with GaussView 5.0 used to modify the lesion, generate *anti* lesion orientations and consider several input orientations with differing values of key dihedral angles.

Using the TLEAP module of AMBER, the DNA systems were solvated within a 8 Å octahedral TIP3P water box, such that there was a minimum of 8 Å between the DNA and the edge of the water box in any direction. Furthermore, the systems were neutralized with 22 or 20 sodium counter-ions for the full and deletion helices respectively. A cutoff distance of 8 Å was used for Leonard-Jones non-bonded interactions, with the particle-mesh Ewald method used to describe Coulombic interactions. The periodic boundary condition was used for all simulation steps. Minimization was performed using 500 steps of the steepest descent (SD) algorithm followed by 500 steps of the conjugate gradient (CG) minimization algorithm, while restraining the DNA with a 500 kcal mol<sup>-1</sup> Å<sup>-2</sup> force constant. The DNA was subsequently optimized without constraints using 1000 steps of SD minimization and 1500 steps of CG minimization. Following minimization, each system was heated to 300 K over 20 ps under conditions of constant volume using the Langevin thermostat algorithm, while imposing a weak constraint on DNA (10 kcal mol<sup>-1</sup> Å<sup>-2</sup>). Finally, 100 ns production simulations were performed at constant temperature (300 K) and pressure (1 atm) using the Langevin and Berendsen algorithms, respectively, and the hydrogen atoms were constrained using SHAKE. A 2 fs time step was used for the production runs, with the trajectory sampled every 1250 steps. During the production phase, 25 kcal mol<sup>-1</sup> Å<sup>-2</sup> distance restraints were used on the Watson-Crick hydrogen bonds in the

terminal base pairs to prevent unravelling that is a common artifact of MD simulations on DNA.<sup>5</sup>

Structural analysis was carried out on each 100 ns trajectory using the PTRAJ and CPPTRAJ modules of AMBER 14. The stability of the systems was confirmed using the backbone root-mean-square deviation (RMSD) compared to the first frame of the production run (Table S4, Figures D.28 and D.29). Representative structures were obtained by clustering the MD simulations with respect to the relative orientation of the damaged nucleobase and 2'-deoxyribose ( $\chi$ ), and the bulky moiety and damaged nucleobase ( $\theta$ ). The reported structures occur over the entire simulation trajectory, with the exception of the <sup>FPh</sup>dG adducted DNA wedge and base-displaced intercalated conformations. The wedge structure interconverts between the reported conformer (46.6% of the simulation time) and a slightly higher energy wedge conformer (32.1 kJ/mol) that deviates in  $\chi$  by approximately 40°, while the base-displaced intercalated conformer dominates the trajectory (87.4% of the simulation time) and interconverts with a wedge conformation. A total of 400 frames of the water and sodium free trajectories corresponding to the clustered simulations were sampled to calculate the enthalpy according to the MMPBSA method, and the entropy was estimated using the normal mode analysis method in AMBER.

## References

1. Sproviero, M.; Verwey, A. M.; Rankin, K. M.; Witham, A. A.; Soldatov, D. V.; Manderville, R. A.; Fekry, M. I.; Sturla, S. J.; Sharma, P.; Wetmore, S. D., Structural and Biochemical Impact of C8-Aryl-guanine Adducts within the NarI Recognition DNA Sequence: Influence of Aryl Ring Size on Targeted and Semi-Targeted Mutagenicity. *Nucleic Acids Res.* **2014**, 42 (21), 13405-13421.
2. Sproviero, M.; Verwey, A. M.; Witham, A. A.; Manderville, R. A.; Sharma, P.; Wetmore, S. D., Enhancing Bulge Stabilization through Linear Extension of C8-Aryl-Guanine Adducts to Promote Polymerase Blockage or Strand Realignment to Produce a C:C Mismatch. *Chem. Res. Toxicol.* **2015**, 28 (8), 1647-1658.

3. Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P., The R.E.D. Tools: Advances in RESP and ESP Charge Derivation and Force Field Library Building. *Phys. Chem. Chem. Phys.* **2010**, *12* (28), 7821-7839.
4. Sharma, P.; Majdi Yazdi, M.; Merriman, A.; Manderville, R. A.; Wetmore, S. D., Influence of the Linkage Type and Functional Groups in the Carcinogenic Moiety on the Conformational Preferences of Damaged DNA: Structural and Energetic Characterization of Carbon- and Oxygen-Linked C8-Phenolic-Guanine Adducts. *Chem. Res. Toxicol.* **2015**, *28* (4), 782-796.
5. Zgarbova, M.; Otyepka, M.; Sponer, J.; Lankas, F.; Jurecka, P., Base Pair Fraying in Molecular Dynamics Simulations of DNA and RNA. *J. Chem. Theor. Comp.* **2014**, *10* (8), 3177-3189.

## Appendix E

**Supplementary Information for Chapter 6: Impact of DNA Adduct Size, Number, and Relative Position on the Toxicity of Aromatic Amines: A Molecular Dynamics Case Study of <sup>AN</sup>dG- and <sup>AP</sup>dG-Containing DNA Duplexes**

Contains Figures E.1–E.21 and Tables E.1–E.28

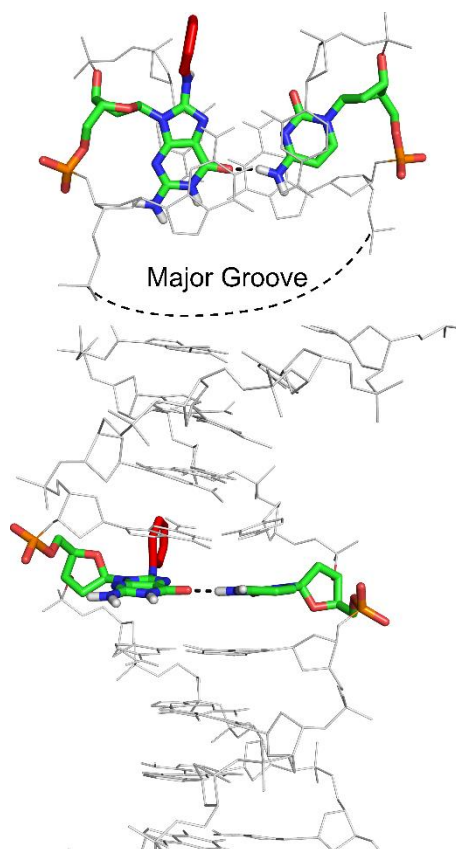


Figure E.1: Structure of <sup>AN</sup>dG adducted DNA in the W conformation, with the major groove highlighted and the bulky aryl group (red) in the minor groove.

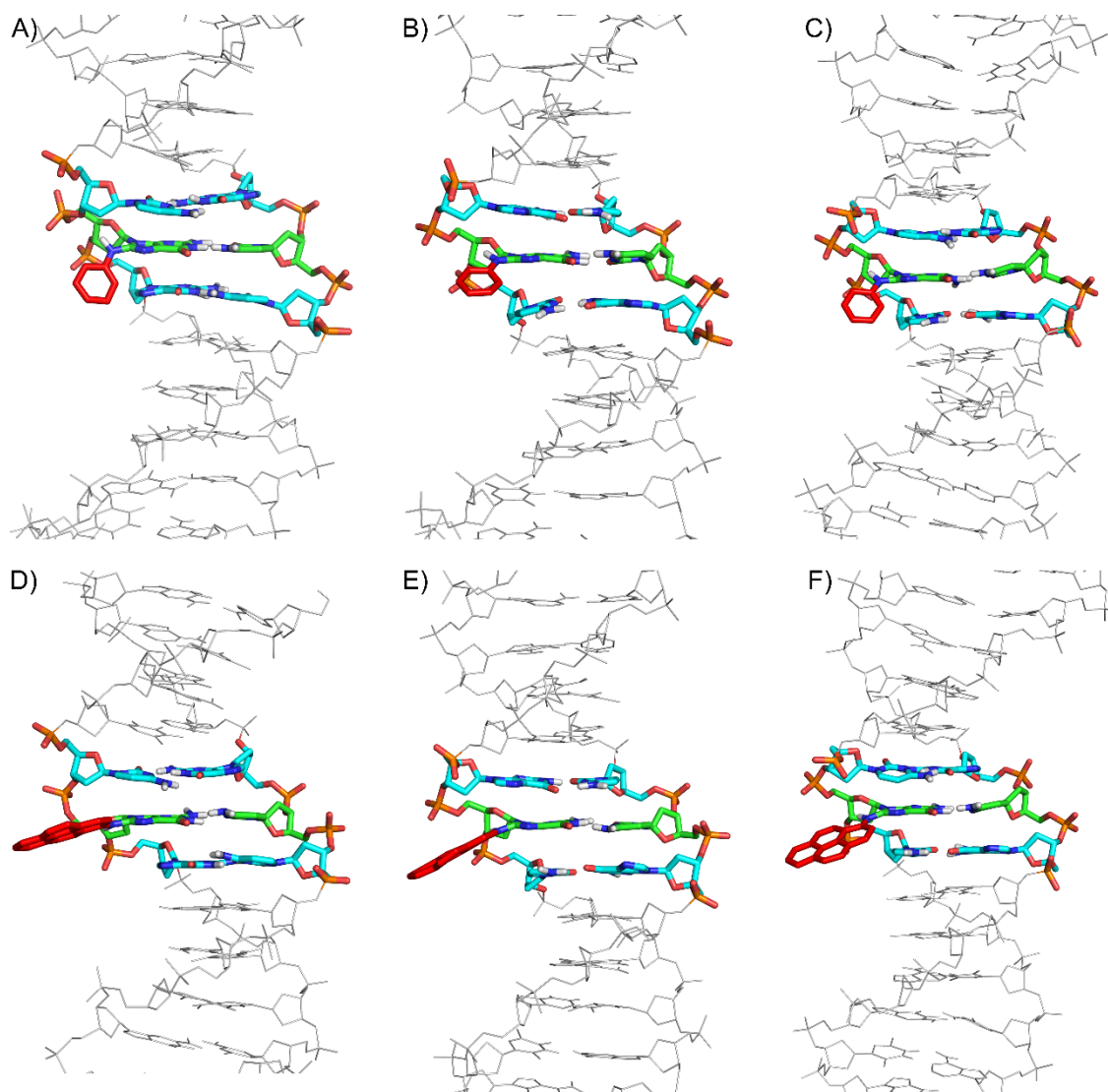


Figure E.2: MD representative structures across the three 500 ns replicas for <sup>AN</sup>dG (top) or <sup>AP</sup>dG (bottom) monoadducted DNA at G<sub>1</sub> (left, A, D), G<sub>2</sub> (middle B, E) or G<sub>3</sub> (right, C, F) in the B conformation. The adducted base pair is shown in green, the bulky moiety in red, and the flanking base pairs in cyan.

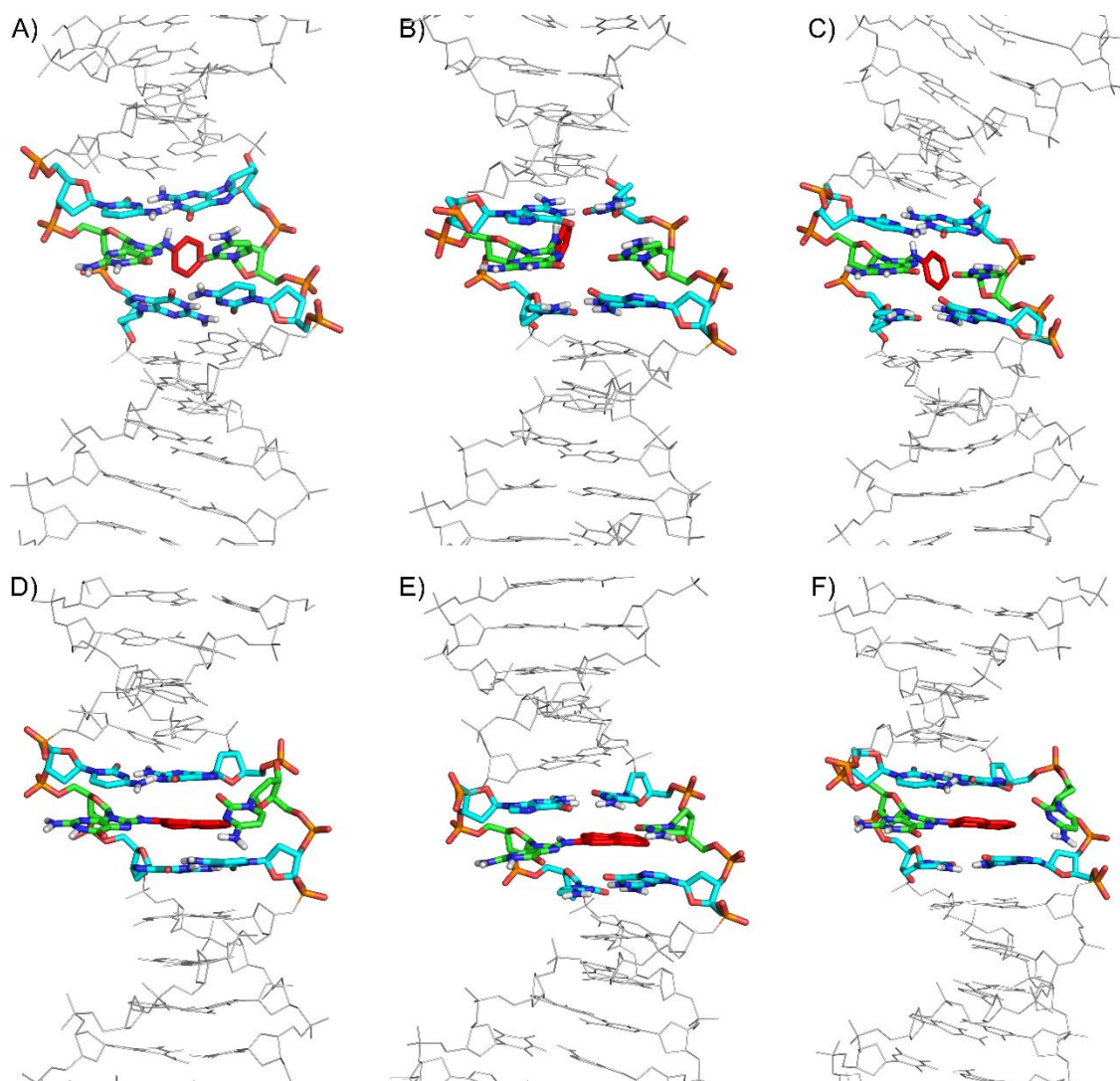


Figure E.3: MD representative structures across the three 500 ns replicas for <sup>AN</sup>dG (top) or <sup>AP</sup>dG (bottom) monoadducted DNA at G<sub>1</sub> (left, A, D), G<sub>2</sub> (middle B, E) or G<sub>3</sub> (right, C, F) in the S conformation. The adducted base pair is shown in green, the bulky moiety in red, and the flanking base pairs in cyan.



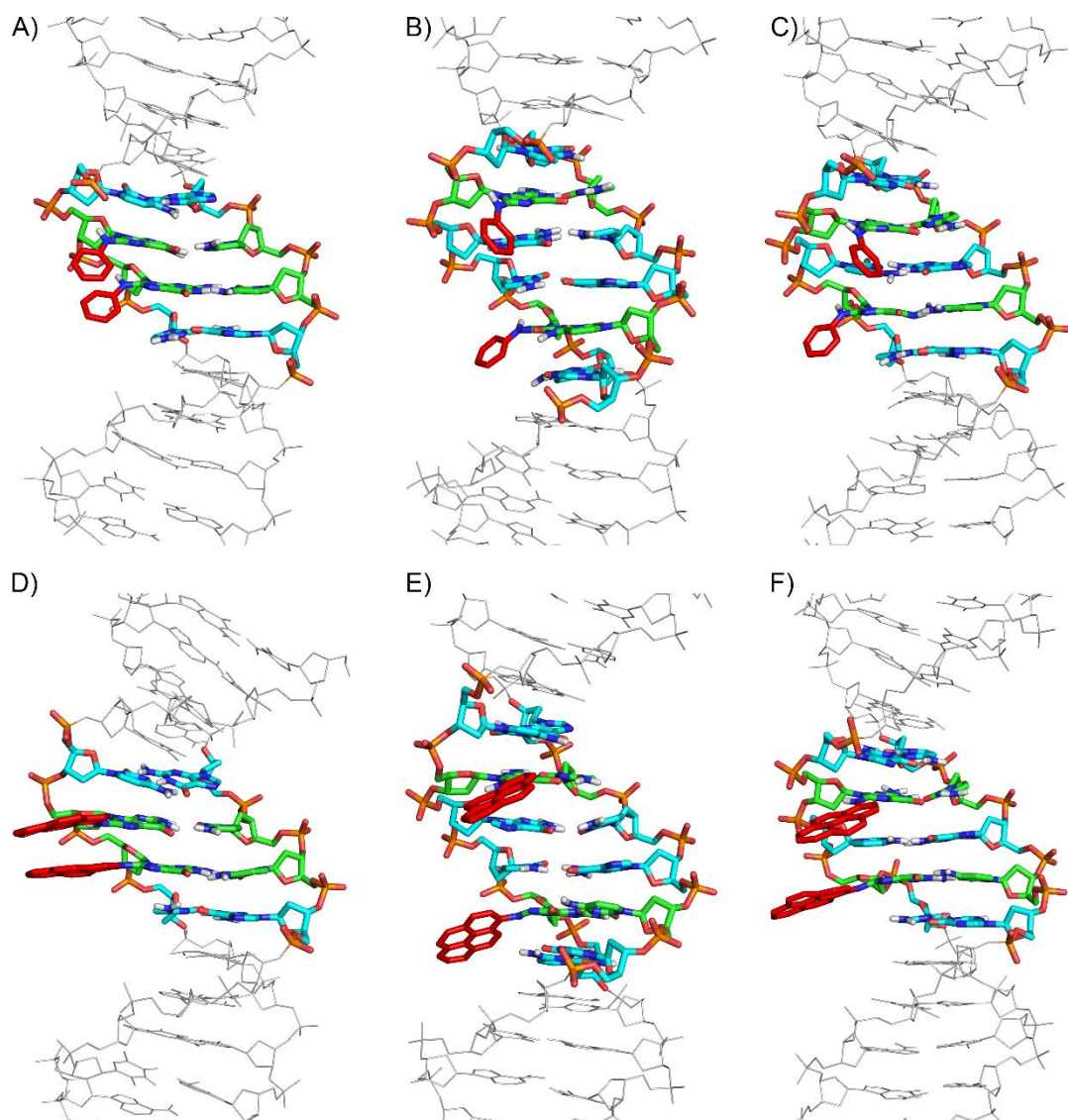


Figure E.4: MD representative structures across the three 500 ns replicas for <sup>AN</sup>dG (top) or <sup>AP</sup>dG (bottom) diadducted DNA in the  $G_1^B G_2^B$  (left, A, D),  $G_1^B G_3^B$  (middle, B, E) or  $G_2^B G_3^B$  (right, C, F) conformations. The adducted base pair is shown in green, the bulky moiety in red, and the flanking base pairs in cyan.

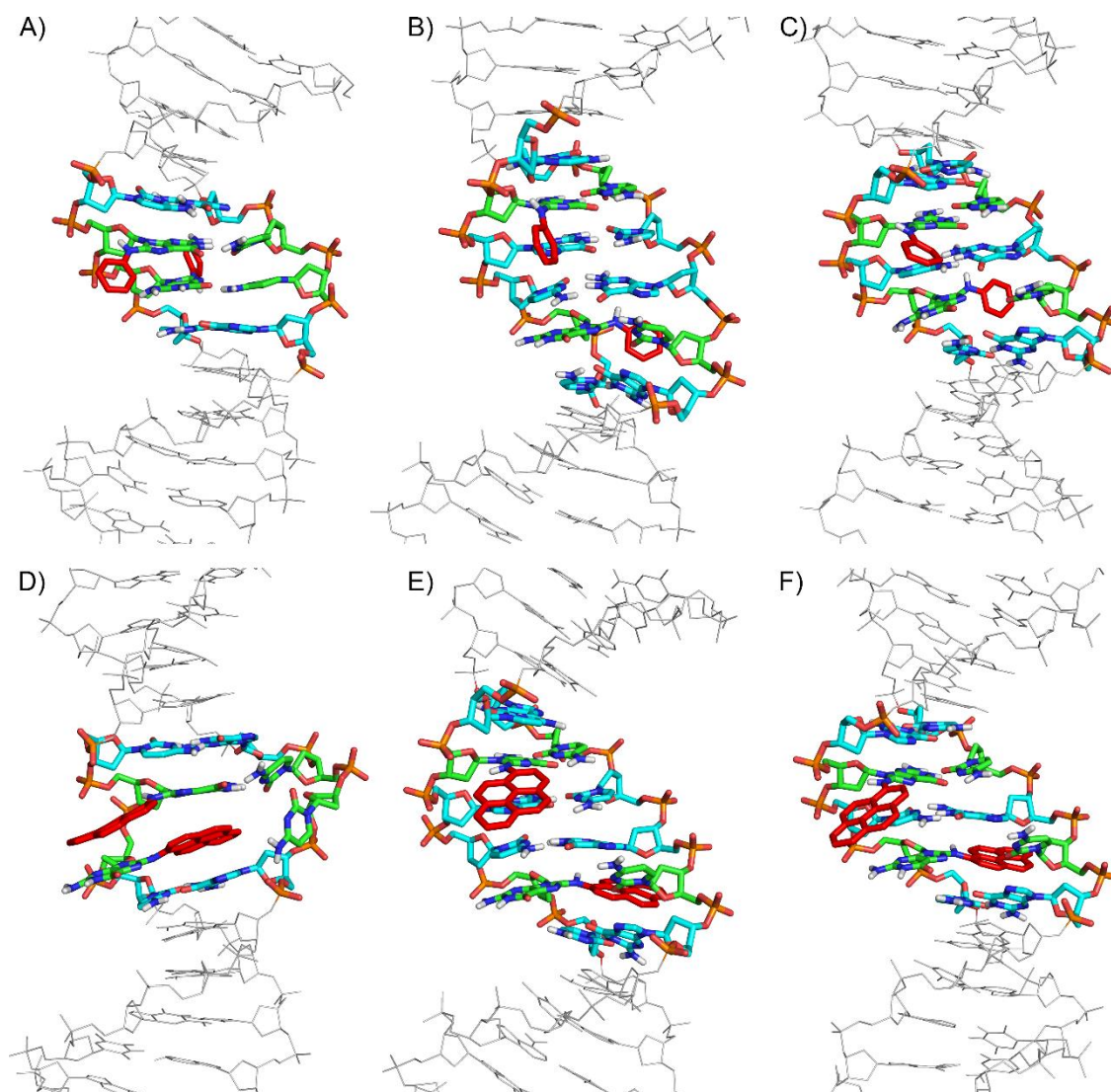


Figure E.5: MD representative structures across the three 500 ns replicas for <sup>AN</sup>dG (top) or <sup>AP</sup>dG (bottom) diadducted DNA in the  $G_1^P G_2^S$  (left, A, D),  $G_1^P G_3^S$  (middle, B, E) or  $G_2^P G_3^S$  (right, C, F) conformations. The adducted base pair is shown in green, the bulky moiety in red, and the flanking base pairs in cyan.

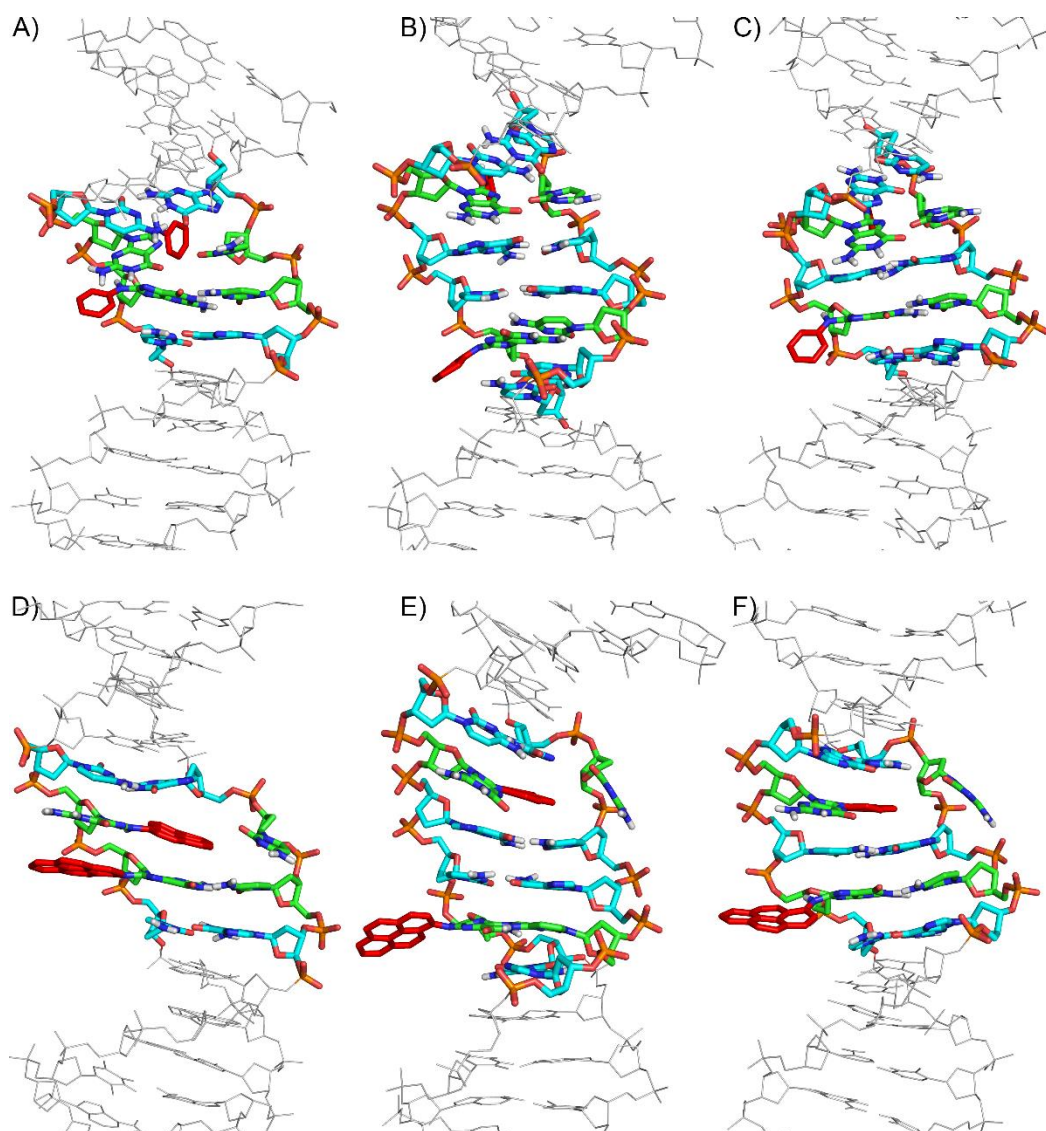


Figure E.6: MD representative structures across the three 500 ns replicas for <sup>AN</sup>dG (top) or <sup>AP</sup>dG (bottom) diadducted DNA in the  $G_1^S G_2^B$  (left, A, D),  $G_1^S G_3^B$  (middle, B, E) or  $G_2^S G_3^B$  (right, C, F) conformations. The adducted base pair is shown in green, the bulky moiety in red, and the flanking base pairs in cyan.



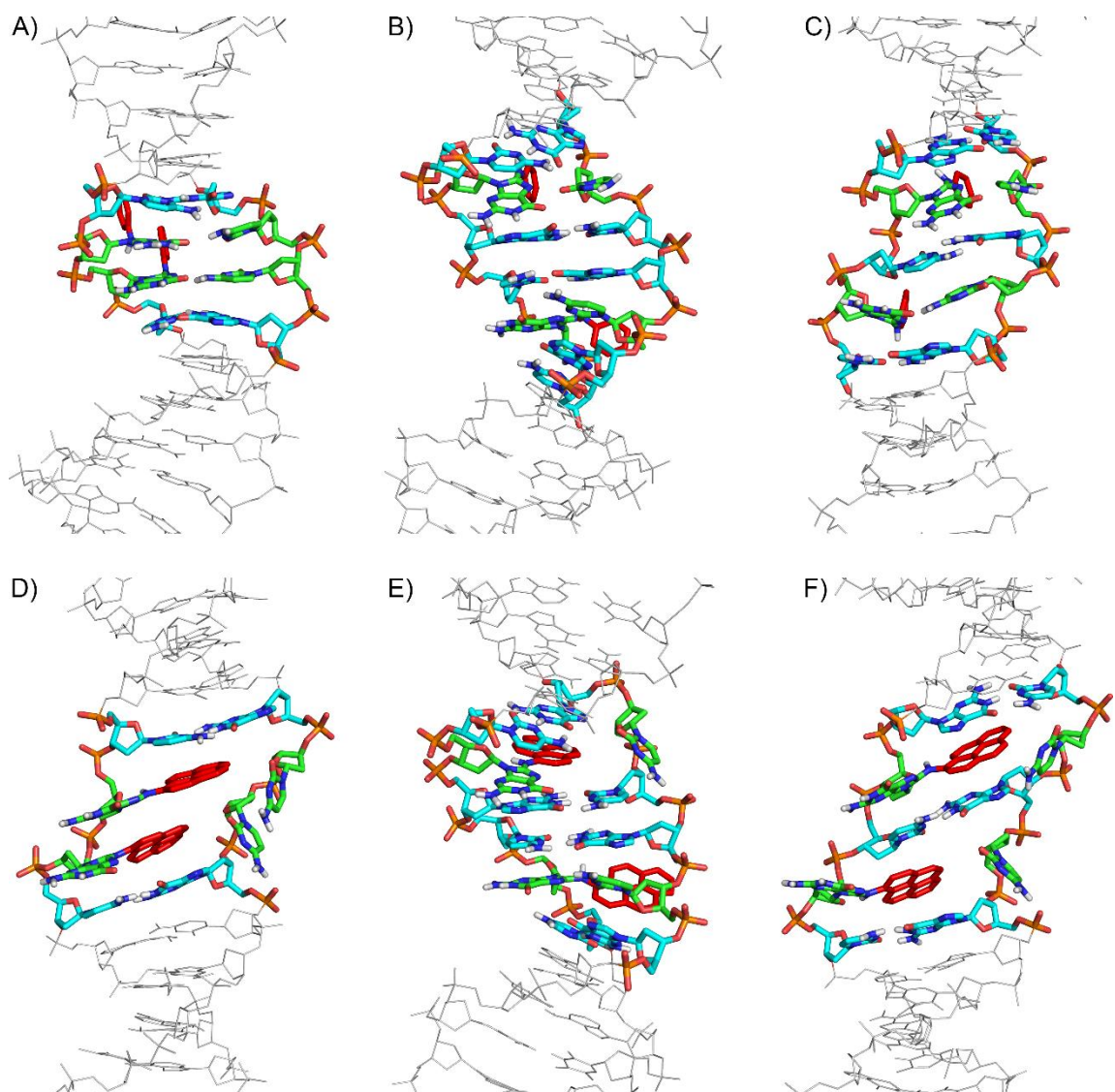


Figure E.7: MD representative structures across the three 500 ns replicas for <sup>AN</sup>dG (top) or <sup>AP</sup>dG (bottom) diadducted DNA in the  $G_1^S G_2^S$  (left, A, D),  $G_1^S G_3^S$  (middle, B, E) or  $G_2^S G_3^S$  (right, C, F) conformations. The adducted base pair is shown in green, the bulky moiety in red, and the flanking base pairs in cyan.

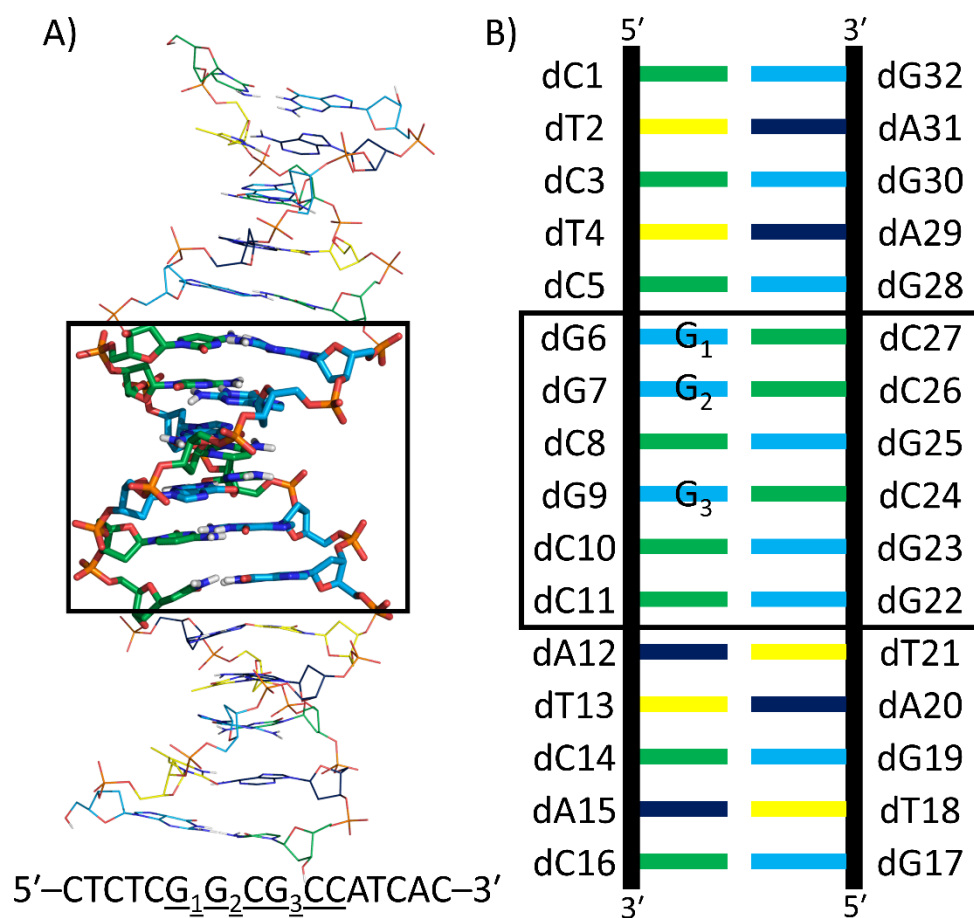


Figure E.8: A) Structure and B) nucleotide numbering of the (undamaged) *NarI* containing sequence investigated in the present work. Helical parameters were calculated using a pseudostep between the base pairs on either side of the box.

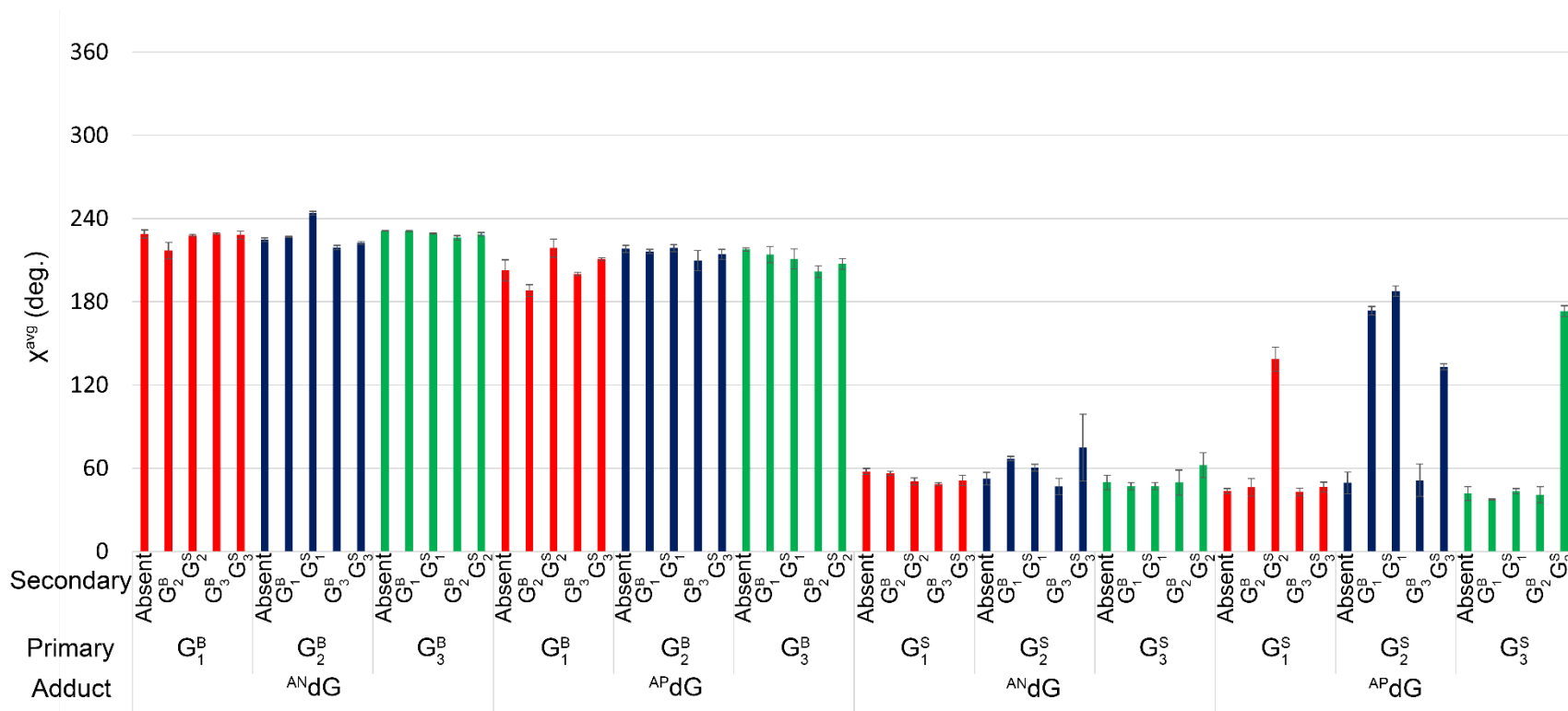


Figure E.9:  $\chi^{\text{avg}}$  (deg.) of the adduct in monoadducted (Secondary = Absent) and diadducted DNA, with standard deviations calculated between replicas. The bars are color coded based on the lesion under investigation, with the G<sub>1</sub> site in red, G<sub>2</sub> site in green and G<sub>3</sub> site in blue (see main text Figure 1 for  $\chi$  definition).

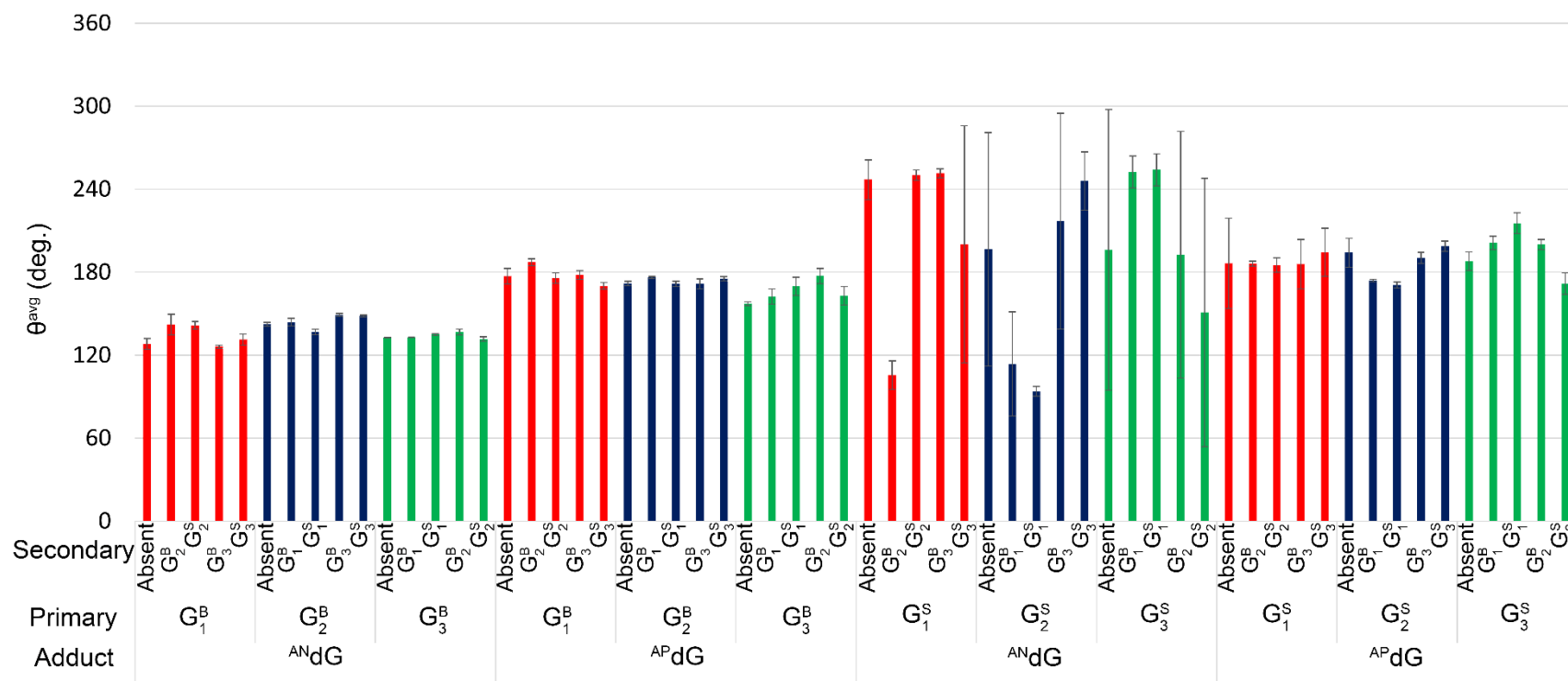


Figure E.10:  $\theta^{\text{avg}}$  (deg.) of the adduct in monoadducted (Secondary = Absent) and diadducted DNA, with standard deviations calculated between replicas. The bars are color coded based on the lesion under investigation, with the  $G_1$  site in red,  $G_2$  site in green and  $G_3$  site in blue (see main text Figure 1 for  $\theta$  definition).

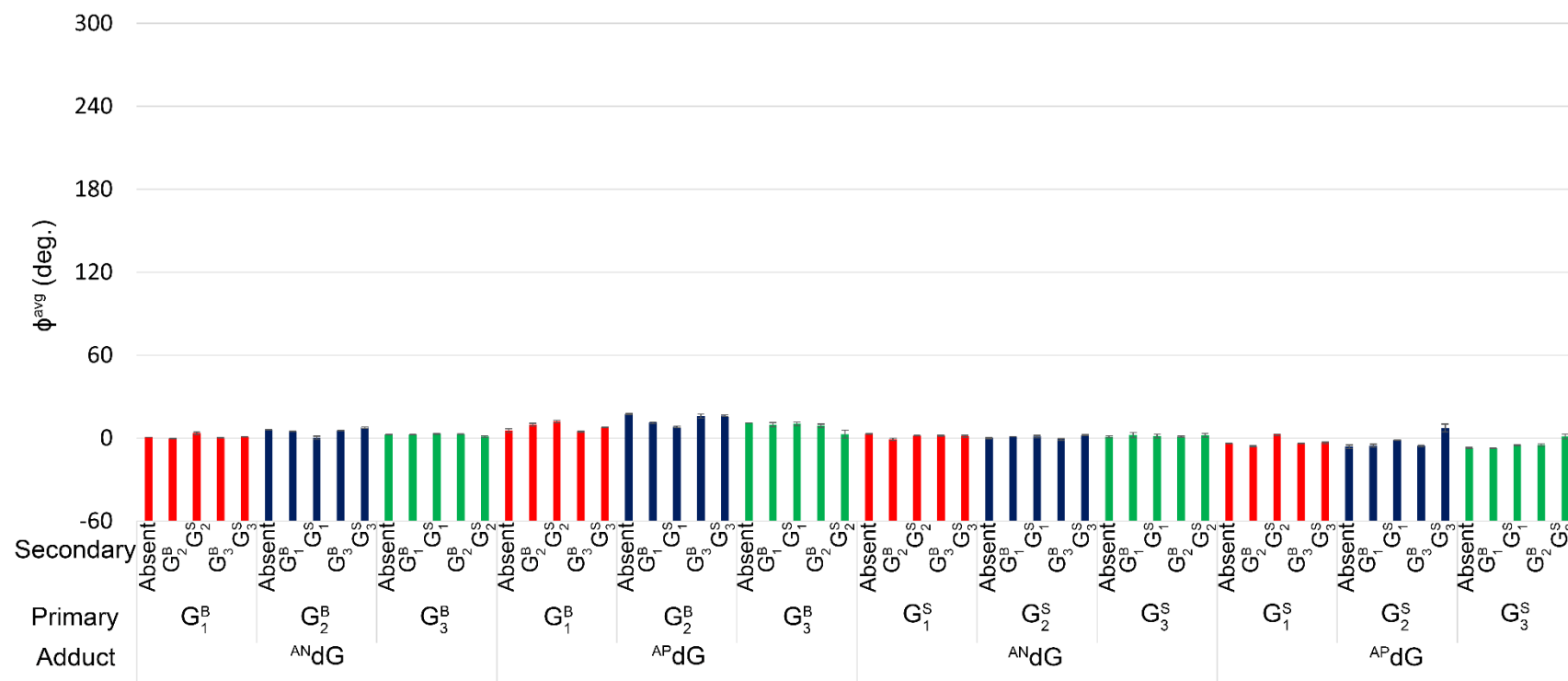


Figure E.11:  $\phi^{\text{avg}}$  (deg.) of the adduct in monoadducted (Secondary = Absent) and diadducted DNA, with standard deviations calculated between replicas. The bars are color coded based on the lesion under investigation, with the G<sub>1</sub> site in red, G<sub>2</sub> site in green and G<sub>3</sub> site in blue (see main text Figure 1 for  $\phi$  definition).



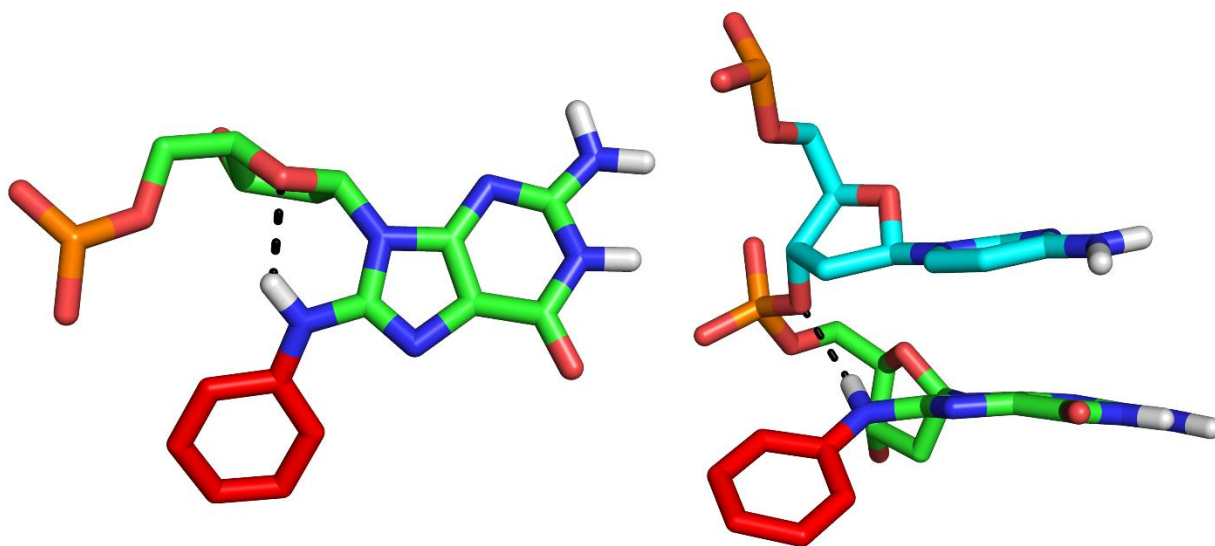


Figure E.12: MD representative structure of  $G_1^B \text{ANdG}$  monoadducted DNA, highlighting the linking N–H hydrogen bonding with O4' of the same nucleotide seen for both  $\text{ANdG}$  and  $\text{APdG}$  at all positions (left) or O3' of the 5' flanking nucleotide seen for  $\text{ANdG}$  and  $G_1$  (right). The adducted base is shown in green, the bulky moiety in red, and the flanking base in cyan.

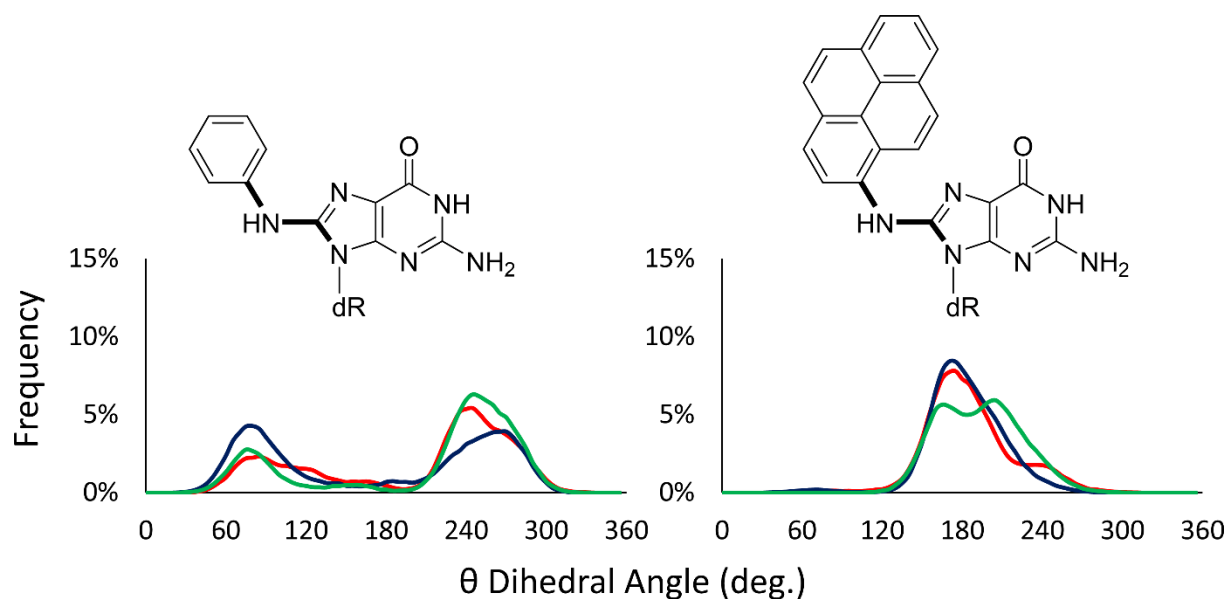


Figure E.13: Histogram of the  $\theta$  dihedral angle (bolded) over all MD replicas for  $^{AN}dG$ - (left) and  $^{AP}dG$ -containing DNA duplexes (right) at  $G_1$  (red),  $G_2$  (blue) or  $G_3$  (green) in monoadducted DNA in the S conformation.

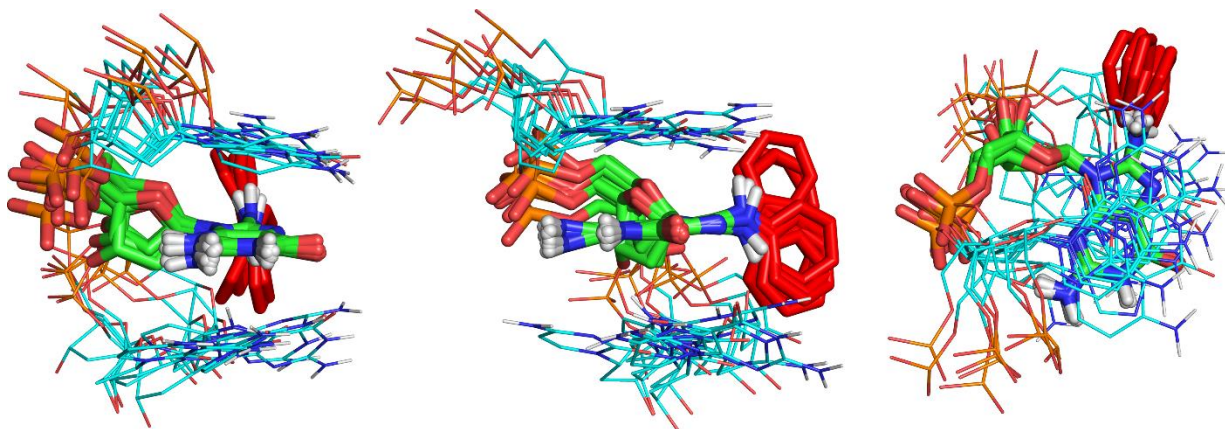


Figure E.14: Different views of an overlay of the lesion site from representative structures obtained from each MD simulation of  $^{AN}dG$  monoadducted DNA in the S conformation, highlighting the nonplanarity of the bulky group with respect to the nucleobase. The adducted base is shown in green, the bulky moiety in red, and the flanking bases in cyan.

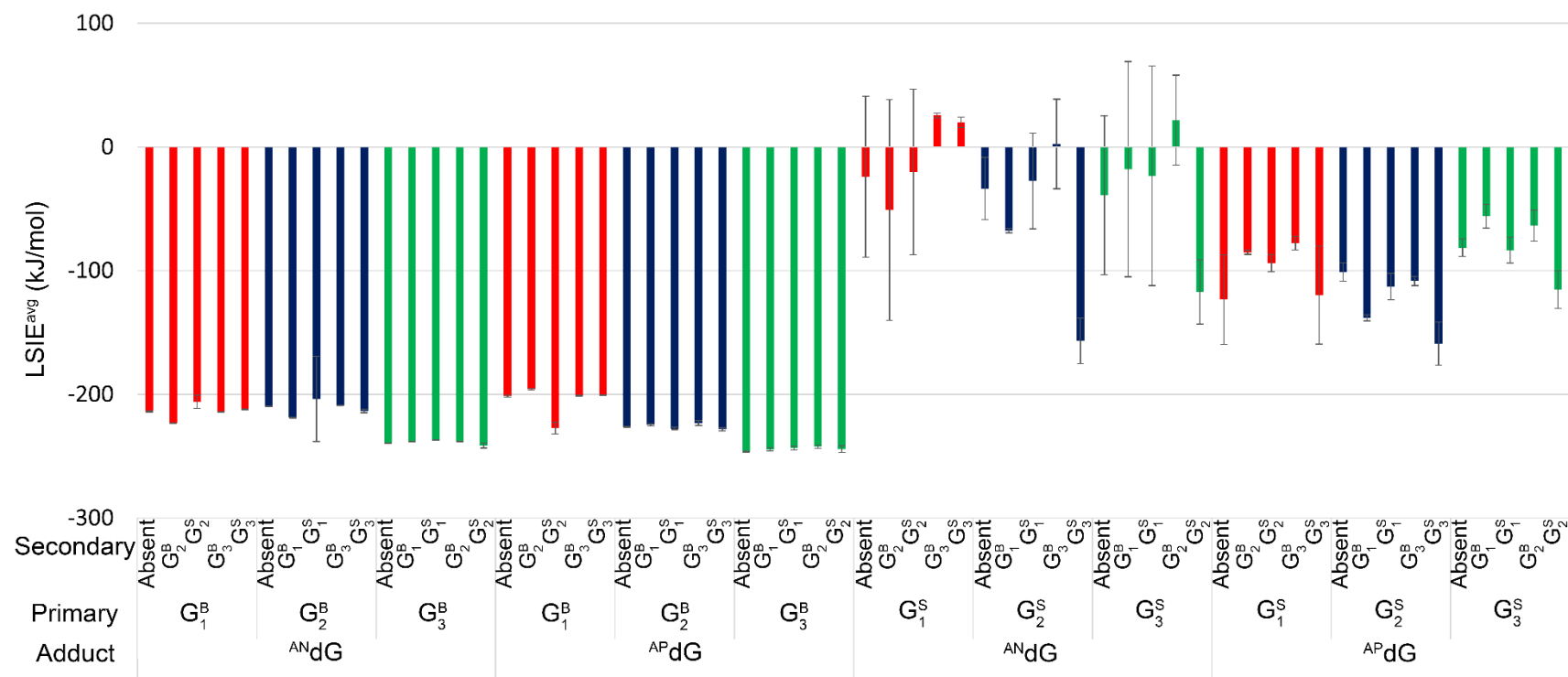


Figure E.15: LSIE<sup>avg</sup> (kJ/mol) of the adduct in monoadducted (Secondary = Absent) and diadducted DNA duplexes, with standard deviations calculated between replicas. The bars are color coded based on the lesion under investigation, with the G<sub>1</sub> site in red, G<sub>2</sub> site in green and G<sub>3</sub> site in blue.

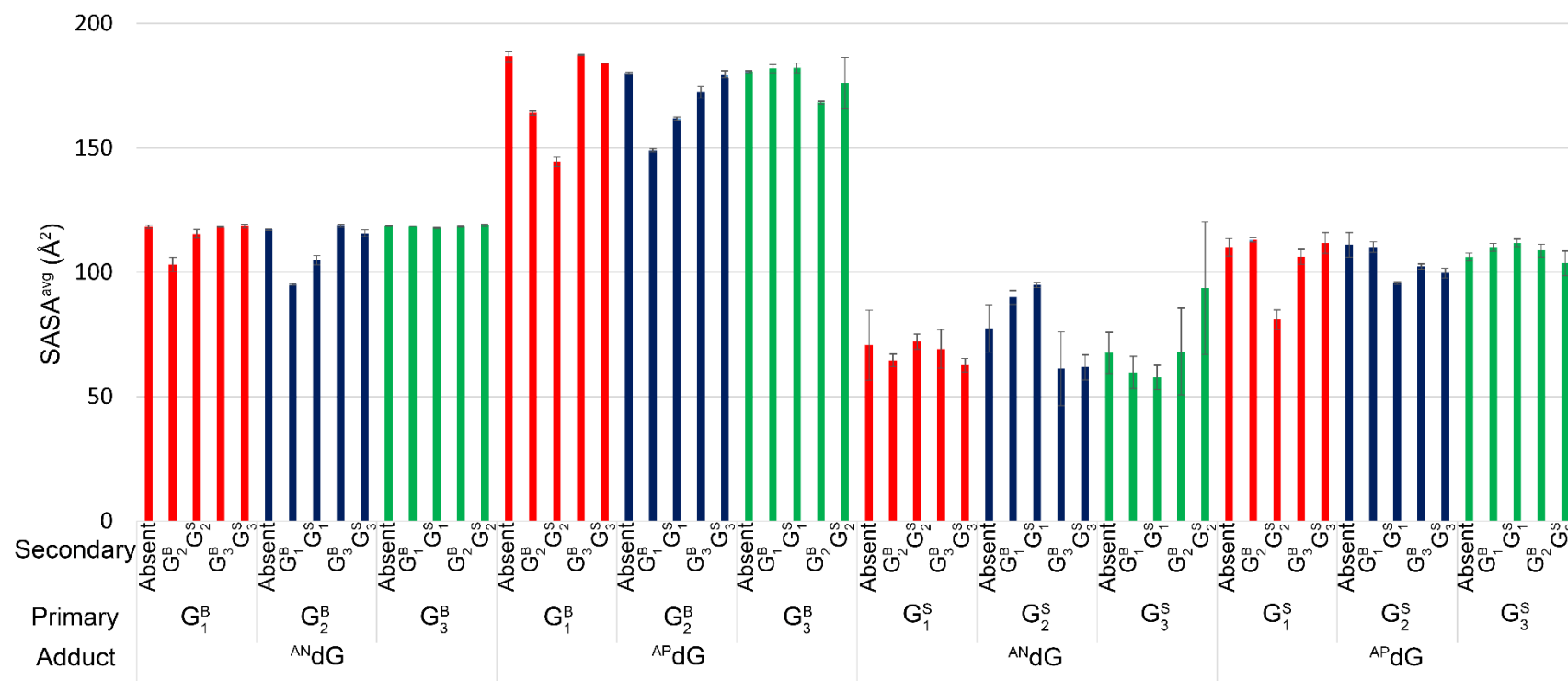


Figure E.16: SASA<sup>avg</sup> (Å<sup>2</sup>) of the adduct in monoadducted (Secondary = Absent) and diadducted DNA duplexes, with standard deviations calculated between replicas. The bars are color coded based on the lesion under investigation, with the G<sub>1</sub> site in red, G<sub>2</sub> site in green and G<sub>3</sub> site in blue.

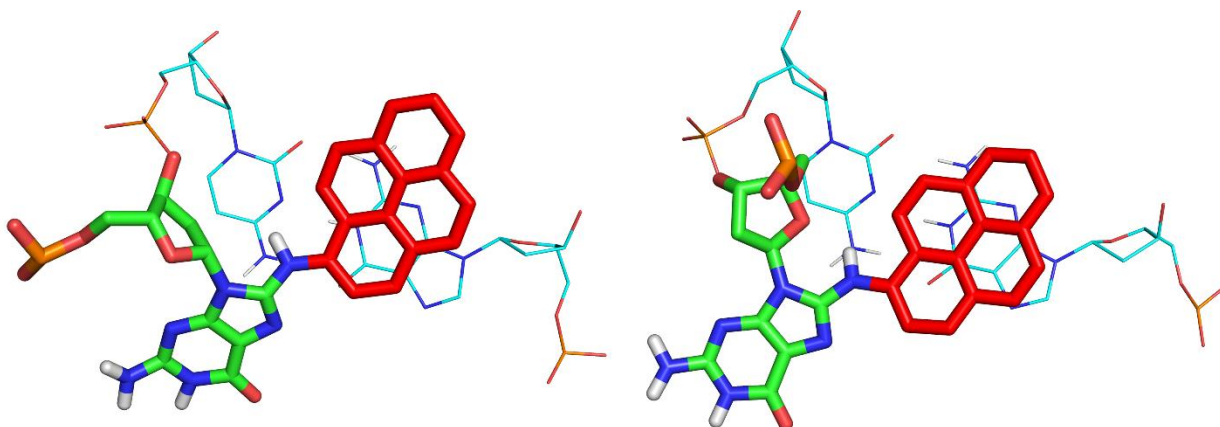


Figure E.17: MD representative structures of <sup>AP</sup>dG monoadducted DNA at G<sub>3</sub> (left) and diadducted DNA at G<sub>2</sub>G<sub>3</sub> (right), highlighting the shift in sugar orientation in the S conformation at G<sub>3</sub> upon addition of the second lesion. The adducted base is shown in green, the bulky moiety in red, and the flanking bases in cyan.

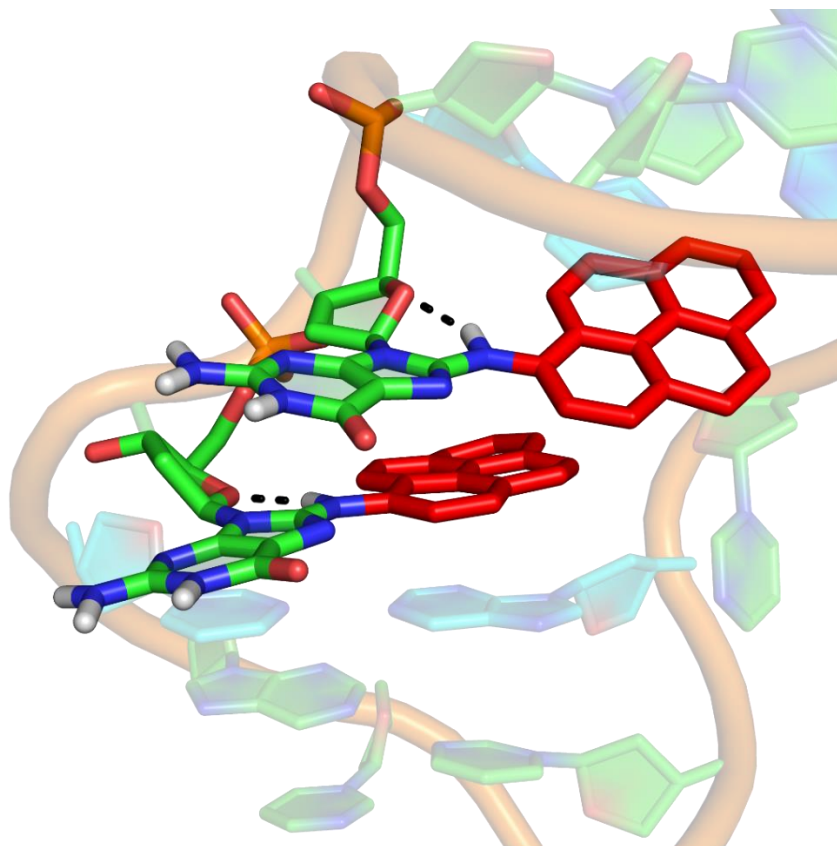


Figure E.18: MD representative structure of <sup>AP</sup>dG diadducted DNA in the G<sub>1</sub><sup>S</sup>G<sub>2</sub><sup>S</sup> conformation, highlighting the distorted sugar at G<sub>2</sub> and G<sub>3</sub>, the intramolecular hydrogen bonds between the linker N–H and O4', and the lack of lesion overlay due to helical twist. The adducted bases are shown in green and the bulky moiety in red.

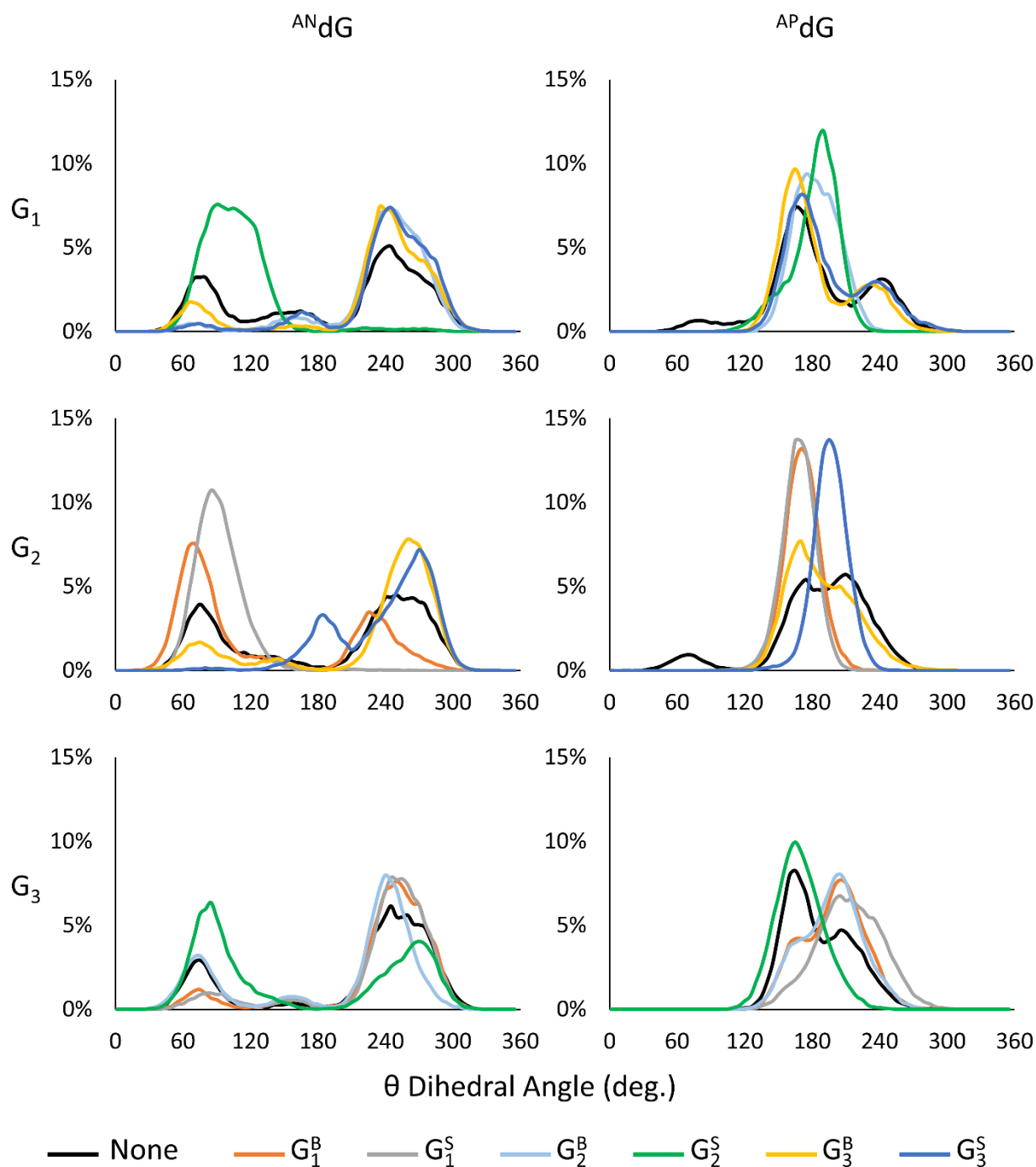


Figure E.19: Histograms of the  $\theta$  dihedral angle over all MD replicas for  $^{AN}dG$  (left) and  $^{AP}dG$  (right) adducted DNA duplexes in the S conformation at  $G_1$  (top),  $G_2$  (middle) or  $G_3$  (bottom). The color indicates the presence and location of a second lesion, with 'None' corresponding to monoadducted DNA.

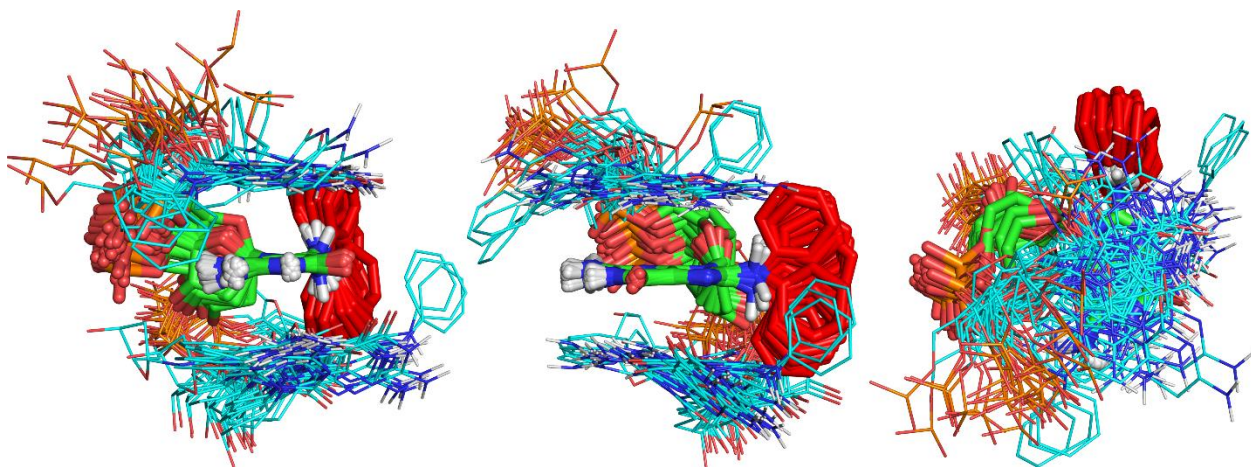


Figure E.20: Different views of an overlay of the lesion site from representative structures obtained from each MD simulation of <sup>AN</sup>dG diadducted DNA in the S conformation, highlighting the nonplanarity of the bulky group with respect to the nucleobase. The adducted base is shown in green, the bulky moiety in red, and the flanking bases in cyan.

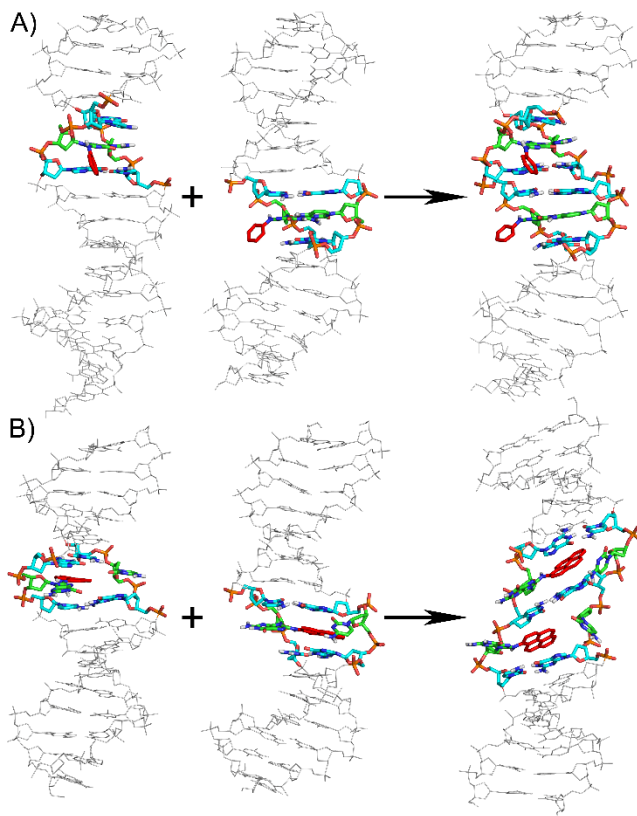


Figure E.21: Structures of A) the  $G_1^P$  (left),  $G_3^P$  (middle) and  $G_1^P G_3^P$  (right) <sup>AN</sup>dG-containing helices and B) the  $G_2^S$  (left),  $G_3^S$  (middle) and  $G_2^S G_3^S$  (right) <sup>AP</sup>dG-containing helices, highlighting the range in helical distortion observed for monoadducted and diadducted DNA duplexes.



Table E.1: Average (standard deviation in parentheses) backbone RMSD (Å) across MD simulations of undamaged and damaged DNA.<sup>a</sup>

| System       | Primary Lesion              | Secondary Lesion            | ANdG      |           |           | APdG      |           |           |
|--------------|-----------------------------|-----------------------------|-----------|-----------|-----------|-----------|-----------|-----------|
|              |                             |                             | Replica 1 | Replica 2 | Replica 3 | Replica 1 | Replica 2 | Replica 3 |
| Undamaged    | absent                      | absent                      | 2.9 (0.7) | 2.9 (0.7) | 3.0 (0.7) | 2.9 (0.7) | 2.9 (0.7) | 3.0 (0.7) |
| Monoadducted | G <sub>1</sub> <sup>B</sup> | absent                      | 4.4 (0.9) | 4.1 (0.8) | 4.3 (0.8) | 4.4 (0.8) | 4.2 (0.8) | 4.2 (0.8) |
|              | G <sub>1</sub> <sup>S</sup> | absent                      | 4.4 (0.7) | 4.4 (0.7) | 5.7 (1.0) | 5.3 (0.8) | 5.6 (1.0) | 5.3 (0.7) |
|              | G <sub>2</sub> <sup>B</sup> | absent                      | 4.2 (0.8) | 4.1 (0.8) | 4.2 (0.8) | 4.5 (0.8) | 4.2 (0.8) | 4.6 (0.7) |
|              | G <sub>2</sub> <sup>S</sup> | absent                      | 4.5 (0.8) | 4.9 (0.8) | 4.3 (0.7) | 6.4 (0.8) | 5.8 (0.6) | 5.6 (0.6) |
|              | G <sub>3</sub> <sup>B</sup> | absent                      | 3.1 (0.6) | 3.5 (0.8) | 3.2 (0.6) | 3.2 (0.6) | 3.4 (0.7) | 3.2 (0.7) |
|              | G <sub>3</sub> <sup>S</sup> | absent                      | 3.5 (0.5) | 3.6 (0.7) | 4.0 (0.8) | 4.5 (0.8) | 4.6 (0.6) | 4.5 (1.1) |
| Diadducted   | G <sub>1</sub> <sup>B</sup> | G <sub>2</sub> <sup>B</sup> | 2.6 (0.6) | 2.7 (0.6) | 2.7 (0.6) | 3.2 (0.7) | 3.3 (0.7) | 3.7 (0.7) |
|              | G <sub>1</sub> <sup>B</sup> | G <sub>2</sub> <sup>S</sup> | 3.5 (0.7) | 4.0 (0.5) | 4.1 (0.5) | 3.6 (0.7) | 3.9 (0.7) | 3.4 (0.7) |
|              | G <sub>1</sub> <sup>S</sup> | G <sub>2</sub> <sup>B</sup> | 2.9 (0.5) | 3.0 (0.6) | 3.2 (0.6) | 3.8 (0.7) | 3.7 (0.7) | 4.0 (0.6) |
|              | G <sub>1</sub> <sup>S</sup> | G <sub>2</sub> <sup>S</sup> | 6.5 (0.7) | 7.0 (0.7) | 6.7 (0.7) | 4.9 (1.0) | 4.9 (1.1) | 4.7 (0.9) |
|              | G <sub>1</sub> <sup>B</sup> | G <sub>3</sub> <sup>B</sup> | 3.8 (0.8) | 3.8 (0.8) | 3.4 (0.7) | 3.4 (0.7) | 3.4 (0.7) | 3.7 (0.7) |
|              | G <sub>1</sub> <sup>B</sup> | G <sub>3</sub> <sup>S</sup> | 4.1 (0.7) | 3.9 (0.8) | 3.8 (0.7) | 3.7 (0.7) | 3.3 (0.6) | 3.2 (0.6) |
|              | G <sub>1</sub> <sup>S</sup> | G <sub>3</sub> <sup>B</sup> | 3.7 (0.7) | 3.9 (0.7) | 4.0 (0.7) | 3.7 (0.8) | 3.6 (0.8) | 3.4 (0.7) |
|              | G <sub>1</sub> <sup>S</sup> | G <sub>3</sub> <sup>S</sup> | 4.2 (0.8) | 4.2 (0.7) | 4.1 (0.7) | 5.3 (0.8) | 4.9 (0.8) | 5.2 (0.7) |
|              | G <sub>2</sub> <sup>B</sup> | G <sub>3</sub> <sup>B</sup> | 3.5 (0.6) | 3.1 (0.6) | 3.2 (0.6) | 3.8 (0.8) | 3.7 (0.8) | 3.9 (0.8) |
|              | G <sub>2</sub> <sup>B</sup> | G <sub>3</sub> <sup>S</sup> | 3.7 (0.6) | 3.8 (0.6) | 3.5 (0.6) | 7.0 (0.7) | 7.0 (0.6) | 6.9 (1.5) |
|              | G <sub>2</sub> <sup>S</sup> | G <sub>3</sub> <sup>B</sup> | 3.3 (0.6) | 4.6 (1.3) | 3.4 (0.7) | 5.8 (1.0) | 5.0 (0.7) | 5.1 (0.7) |
|              | G <sub>2</sub> <sup>S</sup> | G <sub>3</sub> <sup>S</sup> | 7.4 (1.2) | 5.7 (2.3) | 3.3 (0.7) | 4.9 (1.0) | 5.0 (0.9) | 5.3 (1.1) |

<sup>a</sup>RMSDs calculated with respect to the first frame of the corresponding 500 ns MD production phase.



Table E.2: Occupancies (%) of the Watson-Crick hydrogen bonds at the terminal ends of DNA helices.<sup>a</sup>

| System       | Adduct | Site(s)                                                 | 5'-terminal end |           |           | 3'-terminal end |           |           |
|--------------|--------|---------------------------------------------------------|-----------------|-----------|-----------|-----------------|-----------|-----------|
|              |        |                                                         | Replica 1       | Replica 2 | Replica 3 | Replica 1       | Replica 2 | Replica 3 |
| Undamaged    | absent | N/A                                                     | 99.4            | 99.5      | 99.4      | 99.7            | 99.8      | 99.7      |
| Monoadducted | ANdG   | G <sub>1</sub> <sup>B</sup>                             | 99.1            | 99.2      | 99.9      | 99.8            | 99.8      | 99.9      |
|              |        | G <sub>2</sub> <sup>B</sup>                             | 99.4            | 99.3      | 99.8      | 99.7            | 99.7      | 99.8      |
|              |        | G <sub>3</sub> <sup>B</sup>                             | 99.0            | 99.7      | 99.8      | 99.7            | 99.8      | 99.8      |
|              | APdG   | G <sub>1</sub> <sup>B</sup>                             | 99.2            | 99.3      | 99.8      | 99.1            | 99.8      | 99.8      |
|              |        | G <sub>2</sub> <sup>B</sup>                             | 99.1            | 99.2      | 99.8      | 99.7            | 99.7      | 99.8      |
|              |        | G <sub>3</sub> <sup>B</sup>                             | 99.1            | 99.2      | 99.7      | 99.7            | 99.7      | 99.7      |
|              | ANdG   | G <sub>1</sub> <sup>S</sup>                             | 99.1            | 99.3      | 99.7      | 99.7            | 99.8      | 99.7      |
|              |        | G <sub>2</sub> <sup>S</sup>                             | 99.6            | 99.2      | 99.7      | 99.7            | 99.8      | 99.7      |
|              |        | G <sub>3</sub> <sup>S</sup>                             | 99.7            | 99.2      | 99.1      | 99.8            | 99.8      | 99.1      |
|              | APdG   | G <sub>1</sub> <sup>S</sup>                             | 99.3            | 99.1      | 99.7      | 99.7            | 99.8      | 99.7      |
|              |        | G <sub>2</sub> <sup>S</sup>                             | 99.2            | 99.0      | 99.7      | 99.8            | 99.8      | 99.7      |
|              |        | G <sub>3</sub> <sup>S</sup>                             | 99.2            | 99.2      | 99.8      | 99.7            | 99.7      | 99.8      |
| Diadducted   | ANdG   | G <sub>1</sub> <sup>B</sup> G <sub>2</sub> <sup>B</sup> | 99.2            | 99.1      | 99.7      | 99.8            | 99.9      | 99.7      |
|              |        | G <sub>1</sub> <sup>B</sup> G <sub>3</sub> <sup>B</sup> | 99.1            | 99.1      | 99.7      | 99.8            | 99.8      | 99.7      |
|              |        | G <sub>2</sub> <sup>B</sup> G <sub>3</sub> <sup>B</sup> | 99.3            | 99.3      | 99.8      | 99.7            | 99.8      | 99.8      |
|              | APdG   | G <sub>1</sub> <sup>B</sup> G <sub>2</sub> <sup>B</sup> | 99.3            | 99.1      | 99.8      | 99.8            | 99.8      | 99.8      |
|              |        | G <sub>1</sub> <sup>B</sup> G <sub>3</sub> <sup>B</sup> | 99.3            | 99.2      | 99.7      | 99.7            | 99.7      | 99.7      |
|              |        | G <sub>2</sub> <sup>B</sup> G <sub>3</sub> <sup>B</sup> | 99.2            | 99.2      | 99.9      | 99.7            | 99.8      | 99.9      |
|              | ANdG   | G <sub>1</sub> <sup>B</sup> G <sub>2</sub> <sup>S</sup> | 99.1            | 99.3      | 99.8      | 99.9            | 99.7      | 99.8      |
|              |        | G <sub>1</sub> <sup>B</sup> G <sub>3</sub> <sup>S</sup> | 99.1            | 99.2      | 99.7      | 99.8            | 99.7      | 99.7      |
|              |        | G <sub>2</sub> <sup>B</sup> G <sub>3</sub> <sup>S</sup> | 99.3            | 99.4      | 99.8      | 99.8            | 99.8      | 99.8      |
|              | APdG   | G <sub>1</sub> <sup>B</sup> G <sub>2</sub> <sup>S</sup> | 99.1            | 99.1      | 99.8      | 99.8            | 99.8      | 99.8      |
|              |        | G <sub>1</sub> <sup>B</sup> G <sub>3</sub> <sup>S</sup> | 99.2            | 99.1      | 99.7      | 99.7            | 99.7      | 99.7      |
|              |        | G <sub>2</sub> <sup>B</sup> G <sub>3</sub> <sup>S</sup> | 99.2            | 99.4      | 99.7      | 99.8            | 99.9      | 99.7      |
|              | ANdG   | G <sub>1</sub> <sup>S</sup> G <sub>2</sub> <sup>B</sup> | 99.3            | 99.2      | 99.8      | 99.7            | 99.8      | 99.8      |
|              |        | G <sub>1</sub> <sup>S</sup> G <sub>3</sub> <sup>B</sup> | 99.2            | 99.1      | 99.7      | 99.7            | 99.7      | 99.7      |
|              |        | G <sub>2</sub> <sup>S</sup> G <sub>3</sub> <sup>B</sup> | 99.4            | 99.2      | 99.7      | 99.8            | 99.8      | 99.7      |
|              | APdG   | G <sub>1</sub> <sup>S</sup> G <sub>2</sub> <sup>B</sup> | 99.1            | 99.4      | 99.7      | 99.8            | 99.8      | 99.7      |
|              |        | G <sub>1</sub> <sup>S</sup> G <sub>3</sub> <sup>B</sup> | 99.1            | 99.0      | 99.7      | 99.7            | 99.7      | 99.7      |
|              |        | G <sub>2</sub> <sup>S</sup> G <sub>3</sub> <sup>B</sup> | 99.4            | 99.2      | 99.8      | 99.9            | 99.7      | 99.8      |
|              | ANdG   | G <sub>1</sub> <sup>S</sup> G <sub>2</sub> <sup>S</sup> | 99.2            | 99.1      | 99.8      | 99.8            | 99.8      | 99.8      |
|              |        | G <sub>1</sub> <sup>S</sup> G <sub>3</sub> <sup>S</sup> | 99.1            | 99.1      | 99.8      | 99.7            | 99.7      | 99.8      |
|              |        | G <sub>2</sub> <sup>S</sup> G <sub>3</sub> <sup>S</sup> | 99.2            | 99.6      | 99.8      | 99.8            | 99.7      | 99.8      |
|              | APdG   | G <sub>1</sub> <sup>S</sup> G <sub>2</sub> <sup>S</sup> | 99.4            | 99.3      | 99.7      | 99.8            | 99.7      | 99.7      |
|              |        | G <sub>1</sub> <sup>S</sup> G <sub>3</sub> <sup>S</sup> | 99.0            | 99.3      | 99.8      | 99.7            | 99.7      | 99.8      |
|              |        | G <sub>2</sub> <sup>S</sup> G <sub>3</sub> <sup>S</sup> | 99.2            | 99.2      | 99.8      | 99.7            | 99.8      | 99.8      |

<sup>a</sup>Occupancies determined as the percent of the simulation that at least three hydrogen bonds exist according to a heavy atom distance < 3.2 Å and a heavy atom–hydrogen–heavy atom angle > 120°.

Table E.3: Occupancies (%) of the Watson-Crick hydrogen bonds at the lesion site for undamaged and monoadducted DNA duplexes in the B conformation.<sup>a</sup>

| Hydrogen Bond <sup>b</sup> | Adduct <sup>c</sup> | Site                        | Replica 1 | Replica 2 | Replica 3 | Average | Standard Deviation <sup>d</sup> |
|----------------------------|---------------------|-----------------------------|-----------|-----------|-----------|---------|---------------------------------|
| (X)N2–H<br>•••O2(C)        | absent              | G <sub>1</sub>              | 99.9      | 99.9      | 99.9      | 99.9    | 0.0                             |
|                            |                     | G <sub>2</sub>              | 99.8      | 99.8      | 99.7      | 99.8    | 0.0                             |
|                            |                     | G <sub>3</sub>              | 99.9      | 100.0     | 99.9      | 99.9    | 0.0                             |
|                            | ANdG                | G <sub>1</sub> <sup>B</sup> | 99.8      | 99.9      | 99.9      | 99.9    | 0.0                             |
|                            |                     | G <sub>2</sub> <sup>B</sup> | 99.9      | 99.8      | 99.9      | 99.9    | 0.0                             |
|                            |                     | G <sub>3</sub> <sup>B</sup> | 99.9      | 100.0     | 100.0     | 100.0   | 0.0                             |
|                            | APdG                | G <sub>1</sub> <sup>B</sup> | 99.8      | 99.8      | 99.9      | 99.8    | 0.1                             |
|                            |                     | G <sub>2</sub> <sup>B</sup> | 99.8      | 99.7      | 99.9      | 99.8    | 0.1                             |
|                            |                     | G <sub>3</sub> <sup>B</sup> | 99.8      | 99.7      | 99.8      | 99.8    | 0.0                             |
| (X)N1–H<br>•••N3(C)        | absent              | G <sub>1</sub>              | 100.0     | 99.9      | 100.0     | 100.0   | 0.0                             |
|                            |                     | G <sub>2</sub>              | 100.0     | 99.9      | 99.9      | 99.9    | 0.0                             |
|                            |                     | G <sub>3</sub>              | 100.0     | 100.0     | 100.0     | 100.0   | 0.0                             |
|                            | ANdG                | G <sub>1</sub> <sup>B</sup> | 98.6      | 99.3      | 99.3      | 99.1    | 0.3                             |
|                            |                     | G <sub>2</sub> <sup>B</sup> | 99.9      | 99.7      | 99.8      | 99.8    | 0.1                             |
|                            |                     | G <sub>3</sub> <sup>B</sup> | 99.7      | 99.8      | 99.8      | 99.8    | 0.0                             |
|                            | APdG                | G <sub>1</sub> <sup>B</sup> | 99.9      | 99.6      | 99.9      | 99.8    | 0.1                             |
|                            |                     | G <sub>2</sub> <sup>B</sup> | 99.9      | 100.0     | 100.0     | 100.0   | 0.0                             |
|                            |                     | G <sub>3</sub> <sup>B</sup> | 99.9      | 99.7      | 99.9      | 99.8    | 0.1                             |
| (C)N4–H<br>•••O6(X)        | absent              | G <sub>1</sub>              | 98.9      | 98.7      | 99.1      | 98.9    | 0.1                             |
|                            |                     | G <sub>2</sub>              | 98.8      | 98.8      | 98.8      | 98.8    | 0.0                             |
|                            |                     | G <sub>3</sub>              | 99.3      | 99.1      | 99.1      | 99.2    | 0.1                             |
|                            | ANdG                | G <sub>1</sub> <sup>B</sup> | 97.0      | 97.6      | 97.7      | 97.4    | 0.3                             |
|                            |                     | G <sub>2</sub> <sup>B</sup> | 98.8      | 98.5      | 98.5      | 98.6    | 0.2                             |
|                            |                     | G <sub>3</sub> <sup>B</sup> | 98.3      | 98.2      | 98.4      | 98.3    | 0.1                             |
|                            | APdG                | G <sub>1</sub> <sup>B</sup> | 97.6      | 97.5      | 97.7      | 97.6    | 0.1                             |
|                            |                     | G <sub>2</sub> <sup>B</sup> | 99.3      | 99.2      | 99.3      | 99.2    | 0.1                             |
|                            |                     | G <sub>3</sub> <sup>B</sup> | 98.5      | 98.3      | 98.5      | 98.5    | 0.1                             |

<sup>a</sup>Occupancies determined as the percent of the simulation that a hydrogen bonds exist according to a heavy atom distance < 3.2 Å and a heavy atom–hydrogen–heavy atom angle > 120°. <sup>b</sup>X represents the damaged nucleotide for adducted DNA or the undamaged nucleotide for natural DNA. <sup>c</sup>Absent indicates undamaged DNA. <sup>d</sup>Standard deviation calculated over the occupancies from the three replicates.

Table E.4: Occupancies (%) of the Watson-Crick hydrogen bonds at the lesion sites for diadducted DNA duplexes in the B conformation.<sup>a</sup>

| Hydrogen Bond <sup>b</sup> | Adduct           | Primary Lesion              | Secondary Lesion  |                   |                   |                   |                   |                   |
|----------------------------|------------------|-----------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
|                            |                  |                             | G <sub>1</sub> -B | G <sub>1</sub> -S | G <sub>2</sub> -B | G <sub>2</sub> -S | G <sub>3</sub> -B | G <sub>3</sub> -S |
| (X)N2–H•••O2(C)            | <sup>AN</sup> dG | G <sub>1</sub> <sup>B</sup> | N/A               | N/A               | 99.9              | 95.3              | 99.9              | 99.7              |
|                            |                  | G <sub>2</sub> <sup>B</sup> | 99.8              | 99.7              | N/A               | N/A               | 99.8              | 99.8              |
|                            |                  | G <sub>3</sub> <sup>B</sup> | 99.9              | 99.9              | 99.9              | 99.8              | N/A               | N/A               |
|                            | <sup>AP</sup> dG | G <sub>1</sub> <sup>B</sup> | N/A               | N/A               | 99.8              | 97.8              | 99.9              | 99.9              |
|                            |                  | G <sub>2</sub> <sup>B</sup> | 99.8              | 99.3              | N/A               | N/A               | 99.8              | 99.8              |
|                            |                  | G <sub>3</sub> <sup>B</sup> | 99.9              | 99.8              | 99.8              | 99.4              | N/A               | N/A               |
| (X)N1–H•••N3(C)            | <sup>AN</sup> dG | G <sub>1</sub> <sup>B</sup> | N/A               | N/A               | 99.5              | 95.5              | 99.1              | 98.4              |
|                            |                  | G <sub>2</sub> <sup>B</sup> | 99.9              | 99.5              | N/A               | N/A               | 99.8              | 99.9              |
|                            |                  | G <sub>3</sub> <sup>B</sup> | 99.7              | 99.2              | 99.8              | 99.6              | N/A               | N/A               |
|                            | <sup>AP</sup> dG | G <sub>1</sub> <sup>B</sup> | N/A               | N/A               | 99.1              | 94.9              | 99.7              | 99.6              |
|                            |                  | G <sub>2</sub> <sup>B</sup> | 99.9              | 99.5              | N/A               | N/A               | 99.9              | 100.0             |
|                            |                  | G <sub>3</sub> <sup>B</sup> | 99.8              | 99.9              | 99.9              | 99.9              | N/A               | N/A               |
| (C)N4–H•••O6(X)            | <sup>AN</sup> dG | G <sub>1</sub> <sup>B</sup> | N/A               | N/A               | 97.4              | 82.6              | 97.2              | 96.3              |
|                            |                  | G <sub>2</sub> <sup>B</sup> | 98.9              | 97.6              | N/A               | N/A               | 98.6              | 98.8              |
|                            |                  | G <sub>3</sub> <sup>B</sup> | 98.1              | 97.5              | 98.1              | 98.1              | N/A               | N/A               |
|                            | <sup>AP</sup> dG | G <sub>1</sub> <sup>B</sup> | N/A               | N/A               | 95.6              | 90.8              | 97.1              | 97.6              |
|                            |                  | G <sub>2</sub> <sup>B</sup> | 99.5              | 98.7              | N/A               | N/A               | 99.0              | 99.3              |
|                            |                  | G <sub>3</sub> <sup>B</sup> | 98.4              | 98.5              | 98.5              | 98.6              | N/A               | N/A               |

<sup>a</sup>Occupancies determined as the percent of the simulation that a hydrogen bonds exist according to a heavy atom distance < 3.2 Å and a heavy atom–hydrogen–heavy atom angle > 120°. <sup>b</sup>X represents the damaged nucleotide.

Table E.5: Occupancies (%) of the Watson-Crick hydrogen bonds flanking the nucleotides at the lesion site in monoadducted DNA duplexes.<sup>a,b</sup>

| Adduct           | Site                        | Replica 1 | Replica 2 | Replica 3 | Replica 1 | Replica 2 | Replica 3 |
|------------------|-----------------------------|-----------|-----------|-----------|-----------|-----------|-----------|
|                  |                             | dC5:dG28  |           |           | dG7:dC26  |           |           |
| <sup>AN</sup> dG | G <sub>1</sub> <sup>B</sup> | 99.1      | 98.9      | 99.1      | 98.8      | 98.7      | 98.7      |
|                  | G <sub>1</sub> <sup>S</sup> | 99.0      | 98.9      | 97.4      | 99.2      | 98.9      | 98.1      |
| <sup>AP</sup> dG | G <sub>1</sub> <sup>B</sup> | 98.7      | 98.7      | 98.8      | 98.9      | 98.8      | 98.8      |
|                  | G <sub>1</sub> <sup>S</sup> | 98.1      | 98.4      | 98.8      | 99.1      | 98.9      | 99.1      |
|                  |                             | dG6:dC27  |           |           | dC8:dG25  |           |           |
| <sup>AN</sup> dG | G <sub>2</sub> <sup>B</sup> | 99.1      | 98.9      | 99.0      | 99.4      | 99.1      | 99.3      |
|                  | G <sub>2</sub> <sup>S</sup> | 98.5      | 96.0      | 98.2      | 97.8      | 98.4      | 97.9      |
| <sup>AP</sup> dG | G <sub>2</sub> <sup>B</sup> | 98.9      | 98.8      | 99.1      | 99.3      | 99.2      | 99.3      |
|                  | G <sub>2</sub> <sup>S</sup> | 97.8      | 98.5      | 98.5      | 95.0      | 98.1      | 92.9      |
|                  |                             | dC8:dG25  |           |           | dC10:dG23 |           |           |
| <sup>AN</sup> dG | G <sub>3</sub> <sup>B</sup> | 99.2      | 99.1      | 99.2      | 98.9      | 98.9      | 98.8      |
|                  | G <sub>3</sub> <sup>S</sup> | 99.4      | 98.7      | 99.2      | 97.0      | 97.5      | 92.3      |
| <sup>AP</sup> dG | G <sub>3</sub> <sup>B</sup> | 99.1      | 98.7      | 98.9      | 98.8      | 98.7      | 99.0      |
|                  | G <sub>3</sub> <sup>S</sup> | 98.8      | 98.2      | 98.5      | 93.0      | 92.2      | 96.1      |

<sup>a</sup>Nucleotide numbering is illustrated in Figure E.8. <sup>b</sup>Occupancies determined as the percent of the simulation that at least three hydrogen bonds exist according to a heavy atom distance < 3.2 Å and a heavy atom–hydrogen–heavy atom angle > 120°.

Table E.6: Occupancies (%) of the Watson-Crick hydrogen bonds flanking the nucleotides at the lesion sites in G<sub>1</sub>G<sub>2</sub> diadducted DNA duplexes.<sup>a,b</sup>

| Adduct           | Sites                                                   | Replica 1 | Replica 2 | Replica 3 | Replica 1 | Replica 2 | Replica 3 |
|------------------|---------------------------------------------------------|-----------|-----------|-----------|-----------|-----------|-----------|
|                  |                                                         | dC5:dG28  |           |           | dC8:dG25  |           |           |
| <sup>AN</sup> dG | G <sub>1</sub> <sup>B</sup> G <sub>2</sub> <sup>B</sup> | 98.9      | 99.0      | 99.0      | 99.4      | 99.4      | 99.2      |
|                  | G <sub>1</sub> <sup>B</sup> G <sub>2</sub> <sup>S</sup> | 99.4      | 99.3      | 99.4      | 96.7      | 97.3      | 97.6      |
|                  | G <sub>1</sub> <sup>S</sup> G <sub>2</sub> <sup>B</sup> | 98.9      | 99.1      | 99.0      | 99.2      | 99.5      | 99.3      |
|                  | G <sub>1</sub> <sup>S</sup> G <sub>2</sub> <sup>S</sup> | 97.7      | 98.4      | 98.6      | 98.3      | 97.6      | 97.7      |
| <sup>AP</sup> dG | G <sub>1</sub> <sup>B</sup> G <sub>2</sub> <sup>B</sup> | 98.0      | 98.5      | 98.5      | 99.0      | 99.3      | 99.2      |
|                  | G <sub>1</sub> <sup>B</sup> G <sub>2</sub> <sup>S</sup> | 97.8      | 98.6      | 98.6      | 99.3      | 99.3      | 99.2      |
|                  | G <sub>1</sub> <sup>S</sup> G <sub>2</sub> <sup>B</sup> | 98.2      | 97.7      | 98.5      | 99.3      | 99.0      | 99.2      |
|                  | G <sub>1</sub> <sup>S</sup> G <sub>2</sub> <sup>S</sup> | 98.4      | 98.4      | 98.2      | 99.3      | 99.4      | 99.2      |

<sup>a</sup>Nucleotide numbering is illustrated in Figure E.8. <sup>b</sup>Occupancies determined as the percent of the simulation that at least three hydrogen bonds exist according to a heavy atom distance < 3.2 Å and a heavy atom–hydrogen–heavy atom angle > 120°.

Table E.7: Occupancies (%) of the Watson-Crick hydrogen bonds flanking the nucleotides at the lesion sites in G<sub>1</sub>G<sub>3</sub> diadducted DNA duplexes.<sup>a,b</sup>

| Adduct           | Sites                                                   | Replica 1 | Replica 2 | Replica 3 | Replica 1 | Replica 2 | Replica 3 |
|------------------|---------------------------------------------------------|-----------|-----------|-----------|-----------|-----------|-----------|
|                  |                                                         | dC5:dG28  |           |           | dG7:dC26  |           |           |
| <sup>AN</sup> dG | G <sub>1</sub> <sup>B</sup> G <sub>3</sub> <sup>B</sup> | 99.0      | 99.2      | 99.1      | 98.8      | 99.0      | 98.6      |
|                  | G <sub>1</sub> <sup>B</sup> G <sub>3</sub> <sup>S</sup> | 99.0      | 99.1      | 99.1      | 99.0      | 99.0      | 98.8      |
|                  | G <sub>1</sub> <sup>S</sup> G <sub>3</sub> <sup>B</sup> | 99.0      | 98.9      | 98.7      | 98.7      | 99.1      | 99.2      |
|                  | G <sub>1</sub> <sup>S</sup> G <sub>3</sub> <sup>S</sup> | 98.6      | 98.8      | 99.1      | 99.1      | 99.5      | 99.3      |
| <sup>AP</sup> dG | G <sub>1</sub> <sup>B</sup> G <sub>3</sub> <sup>B</sup> | 98.7      | 98.8      | 98.7      | 99.0      | 99.2      | 98.9      |
|                  | G <sub>1</sub> <sup>B</sup> G <sub>3</sub> <sup>S</sup> | 98.8      | 98.6      | 98.8      | 99.0      | 99.0      | 98.9      |
|                  | G <sub>1</sub> <sup>S</sup> G <sub>3</sub> <sup>B</sup> | 98.5      | 98.5      | 98.1      | 99.2      | 99.1      | 99.1      |
|                  | G <sub>1</sub> <sup>S</sup> G <sub>3</sub> <sup>S</sup> | 98.5      | 97.2      | 98.4      | 99.1      | 99.3      | 99.1      |
|                  |                                                         | dC8:dG25  |           |           | dC10:dG23 |           |           |
| <sup>AN</sup> dG | G <sub>1</sub> <sup>B</sup> G <sub>3</sub> <sup>B</sup> | 99.0      | 99.2      | 99.0      | 98.7      | 98.9      | 98.7      |
|                  | G <sub>1</sub> <sup>B</sup> G <sub>3</sub> <sup>S</sup> | 99.2      | 99.1      | 99.2      | 93.5      | 94.8      | 98.9      |
|                  | G <sub>1</sub> <sup>S</sup> G <sub>3</sub> <sup>B</sup> | 99.0      | 99.2      | 99.0      | 98.7      | 98.7      | 98.9      |
|                  | G <sub>1</sub> <sup>S</sup> G <sub>3</sub> <sup>S</sup> | 99.0      | 99.3      | 99.1      | 95.6      | 98.4      | 94.2      |
| <sup>AP</sup> dG | G <sub>1</sub> <sup>B</sup> G <sub>3</sub> <sup>B</sup> | 98.7      | 98.9      | 98.7      | 98.7      | 98.9      | 98.9      |
|                  | G <sub>1</sub> <sup>B</sup> G <sub>3</sub> <sup>S</sup> | 98.8      | 99.1      | 99.1      | 93.7      | 93.0      | 92.6      |
|                  | G <sub>1</sub> <sup>S</sup> G <sub>3</sub> <sup>B</sup> | 98.7      | 98.7      | 98.3      | 98.8      | 98.7      | 98.9      |
|                  | G <sub>1</sub> <sup>S</sup> G <sub>3</sub> <sup>S</sup> | 99.1      | 99.0      | 99.1      | 95.4      | 94.5      | 93.7      |

<sup>a</sup>Nucleotide numbering is illustrated in Figure E.8. <sup>b</sup>Occupancies determined as the percent of the simulation that at least three hydrogen bonds exist according to a heavy atom distance < 3.2 Å and a heavy atom–hydrogen–heavy atom angle > 120°.

Table E.8: Occupancies (%) of the Watson-Crick hydrogen bonds flanking the nucleotides at the lesion sites in G<sub>2</sub>G<sub>3</sub> diadducted DNA duplexes.<sup>a,b</sup>

| Adduct | Sites                                                  | Replica 1 | Replica 2 | Replica 3 |
|--------|--------------------------------------------------------|-----------|-----------|-----------|
|        |                                                        | dG6:dC27  |           |           |
| ANdG   | G <sub>2</sub> G <sub>3</sub> <sup>B<sup>B</sup></sup> | 98.9      | 98.8      | 99.1      |
|        | G <sub>2</sub> G <sub>3</sub> <sup>B<sup>S</sup></sup> | 99.0      | 99.1      | 99.0      |
|        | G <sub>2</sub> G <sub>3</sub> <sup>S<sup>B</sup></sup> | 99.0      | 97.7      | 98.7      |
|        | G <sub>2</sub> G <sub>3</sub> <sup>S<sup>S</sup></sup> | 98.8      | 98.9      | 98.9      |
| APdG   | G <sub>2</sub> G <sub>3</sub> <sup>B<sup>B</sup></sup> | 98.8      | 98.8      | 98.7      |
|        | G <sub>2</sub> G <sub>3</sub> <sup>B<sup>S</sup></sup> | 98.9      | 98.9      | 99.1      |
|        | G <sub>2</sub> G <sub>3</sub> <sup>S<sup>B</sup></sup> | 97.2      | 97.8      | 97.9      |
|        | G <sub>2</sub> G <sub>3</sub> <sup>S<sup>S</sup></sup> | 99.4      | 99.3      | 99.3      |
|        |                                                        | dC8:dG25  |           |           |
| ANdG   | G <sub>2</sub> G <sub>3</sub> <sup>B<sup>B</sup></sup> | 99.1      | 99.2      | 99.2      |
|        | G <sub>2</sub> G <sub>3</sub> <sup>B<sup>S</sup></sup> | 99.1      | 98.2      | 98.9      |
|        | G <sub>2</sub> G <sub>3</sub> <sup>S<sup>B</sup></sup> | 97.6      | 97.8      | 97.8      |
|        | G <sub>2</sub> G <sub>3</sub> <sup>S<sup>S</sup></sup> | 94.0      | 98.9      | 98.6      |
| APdG   | G <sub>2</sub> G <sub>3</sub> <sup>B<sup>B</sup></sup> | 99.0      | 98.9      | 99.0      |
|        | G <sub>2</sub> G <sub>3</sub> <sup>B<sup>S</sup></sup> | 99.2      | 99.0      | 98.9      |
|        | G <sub>2</sub> G <sub>3</sub> <sup>S<sup>B</sup></sup> | 46.2      | 94.2      | 92.1      |
|        | G <sub>2</sub> G <sub>3</sub> <sup>S<sup>S</sup></sup> | 77.1      | 75.4      | 41.8      |
|        |                                                        | dC10:dG23 |           |           |
| ANdG   | G <sub>2</sub> G <sub>3</sub> <sup>B<sup>B</sup></sup> | 98.7      | 98.9      | 98.9      |
|        | G <sub>2</sub> G <sub>3</sub> <sup>B<sup>S</sup></sup> | 93.5      | 95.6      | 94.7      |
|        | G <sub>2</sub> G <sub>3</sub> <sup>S<sup>B</sup></sup> | 99.0      | 98.9      | 99.0      |
|        | G <sub>2</sub> G <sub>3</sub> <sup>S<sup>S</sup></sup> | 98.7      | 97.6      | 99.3      |
| APdG   | G <sub>2</sub> G <sub>3</sub> <sup>B<sup>B</sup></sup> | 98.9      | 98.6      | 98.7      |
|        | G <sub>2</sub> G <sub>3</sub> <sup>B<sup>S</sup></sup> | 93.8      | 94.7      | 95.9      |
|        | G <sub>2</sub> G <sub>3</sub> <sup>S<sup>B</sup></sup> | 99.0      | 99.0      | 99.0      |
|        | G <sub>2</sub> G <sub>3</sub> <sup>S<sup>S</sup></sup> | 99.1      | 99.0      | 98.7      |

<sup>a</sup>Nucleotide numbering is illustrated in Figure E.8. <sup>b</sup>Occupancies determined as the percent of the simulation that at least three hydrogen bonds exist according to a heavy atom distance < 3.2 Å and a heavy atom–hydrogen–heavy atom angle > 120°.

Table E.9: Average (standard deviation in parentheses)  $\chi$  dihedral angle (deg.) for each replica and across all replicas ( $\chi^{\text{avg}}$ ) for undamaged and monoadducted DNA.<sup>a</sup>

| Adduct              | Site                        | Replica 1    | Replica 2    | Replica 3    | $\chi^{\text{avg}}$ | BSE <sup>c</sup> |
|---------------------|-----------------------------|--------------|--------------|--------------|---------------------|------------------|
| absent <sup>b</sup> | G <sub>1</sub>              | 256.2 (19.3) | 252.9 (18.9) | 252.5 (19.4) | 253.9 (2.0)         | 1.2              |
|                     | G <sub>2</sub>              | 244.1 (16.7) | 241.1 (19.7) | 244.3 (15.2) | 243.2 (1.8)         | 0.9              |
|                     | G <sub>3</sub>              | 254.1 (16.2) | 254.1 (15.8) | 254.5 (15.3) | 254.2 (0.2)         | 0.4              |
| ANdG                | G <sub>1</sub> <sup>B</sup> | 231.0 (12.8) | 225.5 (17.7) | 229.9 (14.2) | 228.8 (2.9)         | 0.7              |
|                     | G <sub>2</sub> <sup>B</sup> | 225.0 (13.1) | 223.4 (14.6) | 225.9 (12.2) | 224.8 (1.3)         | 0.6              |
|                     | G <sub>3</sub> <sup>B</sup> | 230.7 (12.5) | 231.0 (12.2) | 231.2 (12.3) | 231.0 (0.3)         | 0.1              |
| APdG                | G <sub>1</sub> <sup>B</sup> | 196.0 (16.3) | 210.9 (21.6) | 201.2 (19.7) | 202.7 (7.5)         | 13.3             |
|                     | G <sub>2</sub> <sup>B</sup> | 215.9 (16.6) | 217.6 (13.8) | 221.0 (12.3) | 218.1 (2.6)         | 0.8              |
|                     | G <sub>3</sub> <sup>B</sup> | 216.5 (21.8) | 217.9 (21.9) | 218.7 (21.1) | 217.7 (1.1)         | 3.1              |
| ANdG                | G <sub>1</sub> <sup>S</sup> | 55.3 (12.8)  | 59.5 (16.3)  | 58.2 (14.4)  | 57.6 (2.2)          | 0.6              |
|                     | G <sub>2</sub> <sup>S</sup> | 48.4 (16.3)  | 57.4 (16.8)  | 51.7 (17.6)  | 52.5 (4.6)          | 0.9              |
|                     | G <sub>3</sub> <sup>S</sup> | 50.3 (21.9)  | 54.7 (18.8)  | 44.5 (14.9)  | 49.8 (5.2)          | 1.2              |
| APdG                | G <sub>1</sub> <sup>S</sup> | 41.6 (16.9)  | 45.2 (28.1)  | 44.1 (20.2)  | 43.6 (1.9)          | 0.9              |
|                     | G <sub>2</sub> <sup>S</sup> | 55.9 (38.7)  | 51.9 (46.4)  | 40.8 (19.9)  | 49.6 (7.8)          | 3.0              |
|                     | G <sub>3</sub> <sup>S</sup> | 38.2 (16.6)  | 39.4 (14.9)  | 47.4 (54.2)  | 41.7 (5.0)          | 4.8              |

<sup>a</sup> $\chi = \angle(\text{O4}'\text{--C1}'\text{--N9--C4})$ . See Figure 1 in the main text. <sup>b</sup>Undamaged DNA. <sup>c</sup>Block standard error (BSE) extrapolated from all frames of the three replicas fitted to a monomolecular growth model.

Table E.10: Average (standard deviation in parentheses)  $\theta$  dihedral angle (deg.) for each replica and across all replicas ( $\theta^{\text{avg}}$ ) for monoadducted DNA.<sup>a</sup>

| Adduct | Site                        | Replica 1    | Replica 2    | Replica 3    | $\theta^{\text{avg}}$ | BSE <sup>b</sup> |
|--------|-----------------------------|--------------|--------------|--------------|-----------------------|------------------|
| ANdG   | G <sub>1</sub> <sup>B</sup> | 124.9 (22.7) | 132.6 (31.6) | 126.8 (25.2) | 128.1 (4.0)           | 1.7              |
|        | G <sub>2</sub> <sup>B</sup> | 141.8 (23.5) | 143.9 (25.1) | 141.2 (22.9) | 142.3 (1.4)           | 2.1              |
|        | G <sub>3</sub> <sup>B</sup> | 132.7 (19.6) | 132.6 (18.6) | 132.5 (19.1) | 132.6 (0.1)           | 0.2              |
| APdG   | G <sub>1</sub> <sup>B</sup> | 182.8 (41.3) | 171.6 (38.7) | 177.0 (41.3) | 177.1 (5.6)           | 18.9             |
|        | G <sub>2</sub> <sup>B</sup> | 173.1 (22.0) | 172.3 (19.3) | 169.9 (20.7) | 171.8 (1.7)           | 11.5             |
|        | G <sub>3</sub> <sup>B</sup> | 158.1 (33.4) | 157.8 (33.5) | 155.6 (33.4) | 157.2 (1.4)           | 11.5             |
| ANdG   | G <sub>1</sub> <sup>S</sup> | 263.0 (40.2) | 241.9 (26.0) | 235.6 (58.0) | 246.8 (14.4)          | 11.4             |
|        | G <sub>2</sub> <sup>S</sup> | 244.9 (77.7) | 99.1 (85.7)  | 245.4 (69.7) | 196.5 (84.4)          | 10.5             |
|        | G <sub>3</sub> <sup>S</sup> | 262.9 (27.0) | 79.3 (86.9)  | 246.2 (22.4) | 196.1 (101.6)         | 8.5              |
| APdG   | G <sub>2</sub> <sup>S</sup> | 171.9 (21.2) | 163.3 (49.1) | 223.6 (33.8) | 186.3 (32.6)          | 16.6             |
|        | G <sub>2</sub> <sup>S</sup> | 182.3 (58.1) | 197.3 (29.9) | 202.5 (28.7) | 194.1 (10.5)          | 13.0             |
|        | G <sub>3</sub> <sup>S</sup> | 195.3 (30.6) | 181.8 (28.0) | 186.6 (28.1) | 187.9 (6.8)           | 18.2             |

<sup>a</sup> $\theta = \angle(\text{N9--C8--N10--C10})$ . See Figure 6.1 in the main text. <sup>b</sup>Block standard error (BSE) extrapolated from all frames of the three replicas fitted to a monomolecular growth model.



Table E.11: Average (standard deviation in parentheses)  $\phi$  dihedral angle (deg.) for each replica and across all replicas ( $\phi^{\text{avg}}$ ) for monoadducted DNA.<sup>a</sup>

| Adduct           | Site                        | Replica 1   | Replica 2   | Replica 3   | $\phi^{\text{avg}}$ | BSE <sup>b</sup> |
|------------------|-----------------------------|-------------|-------------|-------------|---------------------|------------------|
| <sup>AN</sup> dG | G <sub>1</sub> <sup>B</sup> | 0.6 (11.8)  | 0.4 (11.9)  | 0.6 (11.8)  | 0.5 (0.1)           | 0.2              |
|                  | G <sub>2</sub> <sup>B</sup> | 6.0 (12.3)  | 6.0 (12.4)  | 6.3 (12.3)  | 6.1 (0.1)           | 0.2              |
|                  | G <sub>3</sub> <sup>B</sup> | 2.7 (11.9)  | 2.5 (11.8)  | 2.8 (11.9)  | 2.7 (0.1)           | 0.1              |
| <sup>AP</sup> dG | G <sub>1</sub> <sup>B</sup> | 4.3 (12.3)  | 7.3 (13.2)  | 5.0 (12.6)  | 5.5 (1.3)           | 0.4              |
|                  | G <sub>2</sub> <sup>B</sup> | 16.7 (12.1) | 17.8 (11.7) | 17.5 (11.6) | 17.3 (0.4)          | 0.2              |
|                  | G <sub>3</sub> <sup>B</sup> | 10.8 (13.0) | 11.0 (12.9) | 10.8 (13.0) | 10.9 (0.1)          | 0.5              |
| <sup>AN</sup> dG | G <sub>1</sub> <sup>S</sup> | 3.3 (11.1)  | 2.7 (10.9)  | 2.8 (10.8)  | 2.9 (0.2)           | 0.2              |
|                  | G <sub>2</sub> <sup>S</sup> | -0.5 (10.6) | 0.4 (10.7)  | -0.1 (10.9) | -0.1 (0.3)          | 0.1              |
|                  | G <sub>3</sub> <sup>S</sup> | 1.4 (10.7)  | 0.0 (9.9)   | 1.5 (9.4)   | 1.0 (0.7)           | 15.7             |
| <sup>AP</sup> dG | G <sub>1</sub> <sup>S</sup> | -3.8 (11.2) | -4.0 (11.3) | -3.3 (10.4) | -3.7 (0.3)          | 0.1              |
|                  | G <sub>2</sub> <sup>S</sup> | -4.8 (12.3) | -6.1 (12.1) | -7.8 (11.6) | -6.3 (1.2)          | 0.3              |
|                  | G <sub>3</sub> <sup>S</sup> | -7.3 (11.1) | -7.5 (11.0) | -6.6 (11.7) | -7.2 (0.4)          | 0.3              |

<sup>a</sup> $\phi = \angle(\text{C8-N10-C10-C11})$ . See Figure 1 in the main text. <sup>b</sup>Block standard error (BSE) extrapolated from all frames of the three replicas fitted to a monomolecular growth model.

Table E.12: Occupancies (%) of the non-Watson-Crick hydrogen bonds at the lesion site of monoadducted DNA.<sup>a,b</sup>

| Adduct            | Site                        | Hydrogen Bond <sup>c</sup> | Replica 1 | Replica 2 | Replica 3 | Average <sup>d</sup> | Standard Deviation <sup>e</sup> |
|-------------------|-----------------------------|----------------------------|-----------|-----------|-----------|----------------------|---------------------------------|
| AN <sub>d</sub> G | G <sub>1</sub> <sup>B</sup> | (X6)N10–H•••O3'(C5)        | 33.0      | 25.8      | 29.9      | 29.6                 | 2.9                             |
|                   | G <sub>1</sub> <sup>B</sup> | (X6)N10–H•••O4'(X6)        | 48.5      | 52.2      | 50.3      | 50.3                 | 1.5                             |
|                   | G <sub>2</sub> <sup>B</sup> | (X7)N10–H•••O4'(X7)        | 80.2      | 79.9      | 79.5      | 79.9                 | 0.3                             |
|                   | G <sub>3</sub> <sup>B</sup> | (X9)N10–H•••O4'(X9)        | 75.1      | 75.2      | 74.6      | 75.0                 | 0.3                             |
| AP <sub>d</sub> G | G <sub>1</sub> <sup>B</sup> | (X6)N10–H•••O4'(X6)        | 53.5      | 50.1      | 53.7      | 52.4                 | 1.6                             |
|                   | G <sub>2</sub> <sup>B</sup> | (X7)N10–H•••O4'(X7)        | 48.5      | 48.8      | 46.3      | 47.9                 | 1.1                             |
|                   | G <sub>3</sub> <sup>B</sup> | (X9)N10–H•••O4'(X9)        | 60.7      | 60.7      | 59.9      | 60.4                 | 0.4                             |
| AN <sub>d</sub> G | G <sub>1</sub> <sup>S</sup> | (C27)N4–H•••O6(X6)         | 63.5      | 0.0       | 0.0       | 21.2                 | 29.9                            |
|                   | G <sub>1</sub> <sup>S</sup> | (X6)N10–H•••N1(G28)        | 19.4      | 21.3      | 0.0       | 13.5                 | 9.6                             |
|                   | G <sub>1</sub> <sup>S</sup> | (X6)N10–H•••N2(G28)        | 45.9      | 50.7      | 0.0       | 32.2                 | 22.9                            |
|                   | G <sub>1</sub> <sup>S</sup> | (X6)N10–H•••O4'(G7)        | 6.1       | 0.0       | 46.9      | 17.7                 | 20.8                            |
|                   | G <sub>1</sub> <sup>S</sup> | (C27)N4–H•••OP2(C26)       | 5.7       | 8.0       | 10.1      | 7.9                  | 1.8                             |
|                   | G <sub>2</sub> <sup>S</sup> | (C26)N4–H•••O6(X7)         | 18.8      | 57.5      | 28.6      | 35.0                 | 16.4                            |
|                   | G <sub>2</sub> <sup>S</sup> | (C26)N4–H•••N7(X7)         | 7.9       | 22.9      | 15.5      | 15.4                 | 6.1                             |
|                   | G <sub>2</sub> <sup>S</sup> | (X7)N10–H•••O2(C27)        | 17.6      | 5.7       | 12.6      | 11.9                 | 13.5                            |
|                   | G <sub>2</sub> <sup>S</sup> | (X7)N10–H•••O4'(C8)        | 17.3      | 34.1      | 10.4      | 20.6                 | 7.1                             |
|                   | G <sub>2</sub> <sup>S</sup> | (X7)N10–H•••N2(G6)         | 28.2      | 13.0      | 28.1      | 23.1                 | 16.1                            |
|                   | G <sub>3</sub> <sup>S</sup> | (C24)N4–H•••O6(X9)         | 0.0       | 37.1      | 0.0       | 12.4                 | 17.5                            |
|                   | G <sub>3</sub> <sup>S</sup> | (X9)N10–H•••O4'(C10)       | 0.0       | 33.8      | 0.0       | 11.3                 | 15.9                            |
|                   | G <sub>3</sub> <sup>S</sup> | (X9)N10–H•••N2(G25)        | 37.1      | 12.7      | 50.8      | 33.5                 | 15.7                            |
|                   | G <sub>3</sub> <sup>S</sup> | (X9)N10–H•••N3(G25)        | 37.6      | 21.9      | 21.0      | 26.8                 | 7.6                             |
| AP <sub>d</sub> G | G <sub>1</sub> <sup>S</sup> | (C27)N4–H•••OP2(C26)       | 60.7      | 0.0       | 33.0      | 31.3                 | 24.8                            |
|                   | G <sub>1</sub> <sup>S</sup> | (C26)N4–H•••O6(X6)         | 0.0       | 0.0       | 29.7      | 9.9                  | 14.0                            |
|                   | G <sub>1</sub> <sup>S</sup> | (X6)N10–H•••N2(G28)        | 0.0       | 6.1       | 36.4      | 14.1                 | 15.9                            |
|                   | G <sub>2</sub> <sup>S</sup> | (X7)N10–H•••N2(G6)         | 7.3       | 8.6       | 10.1      | 8.7                  | 1.2                             |
|                   | G <sub>2</sub> <sup>S</sup> | (X7)N10–H•••O4'(X7)        | 11.5      | 16.0      | 0.0       | 9.2                  | 6.8                             |
|                   | G <sub>2</sub> <sup>S</sup> | (C26)N4–H•••OP2(G25)       | 19.8      | 34.1      | 23.3      | 25.7                 | 6.1                             |
|                   | G <sub>3</sub> <sup>S</sup> | (C24)N4–H•••OP2(G23)       | 30.5      | 43.6      | 36.1      | 36.7                 | 5.4                             |

<sup>a</sup>Nucleotide numbering is illustrated in Figure E.8. <sup>b</sup>Occupancies determined as the percent of the simulation that a hydrogen bonds exist according to a heavy atom distance < 3.2 Å and a heavy atom–hydrogen–heavy atom angle > 120°. <sup>c</sup>X represents the damaged nucleotide. <sup>d</sup>Average occupancy calculated over the three replicates. <sup>e</sup>Standard deviation calculated over the three replicates.

Table E.13: Average (standard deviation in parentheses) lesion site interaction energy (kJ/mol) for undamaged and monoadducted DNA duplexes.

| Adduct              | Site                        | LSIE <sup>avg, b</sup> | Dataset Standard Deviation <sup>c</sup> | LSIE BSE <sup>d</sup> | Average Stacking <sup>e</sup> | Dataset Standard Deviation <sup>c</sup> | Average Hydrogen Bonding <sup>f</sup> | Dataset Standard Deviation <sup>c</sup> |
|---------------------|-----------------------------|------------------------|-----------------------------------------|-----------------------|-------------------------------|-----------------------------------------|---------------------------------------|-----------------------------------------|
| absent <sup>a</sup> | G <sub>1</sub>              | −215.9 (0.7)           | 13.7                                    | 0.1                   | −97.3 (0.6)                   | 12.1                                    | −118.6 (0.0)                          | 7.4                                     |
|                     | G <sub>2</sub>              | −208.4 (0.1)           | 14.8                                    | 0.1                   | −90.7 (0.3)                   | 12.3                                    | −117.7 (0.1)                          | 7.6                                     |
|                     | G <sub>3</sub>              | −235.7 (0.5)           | 14.4                                    | 0.0                   | −117.2 (0.4)                  | 12.3                                    | −118.6 (0.1)                          | 7.3                                     |
| <sup>AN</sup> dG    | G <sub>1</sub> <sup>B</sup> | −213.9 (0.4)           | 15.8                                    | 0.2                   | −104.5 (0.5)                  | 12.9                                    | −109.4 (0.2)                          | 8.7                                     |
|                     | G <sub>2</sub> <sup>B</sup> | −209.9 (0.2)           | 15.2                                    | 0.0                   | −99.8 (0.2)                   | 12.9                                    | −110.2 (0.1)                          | 7.3                                     |
|                     | G <sub>3</sub> <sup>B</sup> | −239.4 (0.2)           | 14.6                                    | 0.1                   | −129.6 (0.1)                  | 12.7                                    | −109.8 (0.1)                          | 7.5                                     |
| <sup>AP</sup> dG    | G <sub>1</sub> <sup>B</sup> | −201.5 (0.7)           | 15.5                                    | 2.6                   | −96.7 (0.4)                   | 13.9                                    | −104.7 (0.4)                          | 8.0                                     |
|                     | G <sub>2</sub> <sup>B</sup> | −226.3 (0.3)           | 17.1                                    | 0.2                   | −119.2 (0.4)                  | 14.8                                    | −107.1 (0.3)                          | 6.9                                     |
|                     | G <sub>3</sub> <sup>B</sup> | −246.4 (0.4)           | 16.8                                    | 0.9                   | −140.5 (0.3)                  | 14.5                                    | −105.9 (0.1)                          | 7.7                                     |
| <sup>AN</sup> dG    | G <sub>1</sub> <sup>S</sup> | −24.1 (65.1)           | 74.9                                    | 1.7                   | −40.0 (35.8)                  | 44.4                                    | 15.9 (29.3)                           | 33.5                                    |
|                     | G <sub>2</sub> <sup>S</sup> | −33.7 (25.1)           | 72.0                                    | 1.5                   | −48.7 (15.0)                  | 46.2                                    | 15.1 (10.1)                           | 30.1                                    |
|                     | G <sub>3</sub> <sup>S</sup> | −39.1 (64.2)           | 91.3                                    | 2.1                   | −47.9 (39.7)                  | 57.5                                    | 8.9 (24.6)                            | 36.5                                    |
| <sup>AP</sup> dG    | G <sub>1</sub> <sup>S</sup> | −123.3 (36.2)          | 51.6                                    | 1.1                   | −103.4 (21.6)                 | 33.7                                    | −19.9 (15.1)                          | 23.8                                    |
|                     | G <sub>2</sub> <sup>S</sup> | −101.2 (7.4)           | 49.6                                    | 0.8                   | −106.8 (4.4)                  | 37.6                                    | 5.6 (3.6)                             | 16.1                                    |
|                     | G <sub>3</sub> <sup>S</sup> | −81.5 (7.1)            | 49.4                                    | 1.3                   | −86.7 (4.9)                   | 37.8                                    | 5.2 (2.3)                             | 13.9                                    |

<sup>a</sup>Undamaged DNA. <sup>b</sup>Average of the sum of stacking and hydrogen-bonding interactions for the adducted base, bulky moiety and opposing C, with standard deviation calculated between the replicas. <sup>c</sup>Standard deviation calculated over the 1.5  $\mu$ s dataset from the three replicas. <sup>d</sup>Block standard error (BSE) extrapolated from all frames of the three replicas fitted to a monomolecular growth model. <sup>e</sup>Average of the stacking between the adduct (base and bulky moiety) plus the opposing C and the flanking nucleobases, with standard deviation calculated between the replicas. <sup>f</sup>Average of the hydrogen-bonding interaction between the adduct and the pairing C, with standard deviation calculated between the replicas.

Table E.14: Average (standard deviation in parentheses) pseudostep parameters<sup>a</sup> for undamaged and monoadducted DNA.

| Adduct              | Site                        | Rise (Å)   | Shift (Å)  | Slide (Å) | Roll (deg.)  | Tilt (deg.)  | Twist (deg.) | Axial Bend (deg.) |
|---------------------|-----------------------------|------------|------------|-----------|--------------|--------------|--------------|-------------------|
| absent <sup>b</sup> | N/A                         | 22.7 (0.0) | 0.5 (0.1)  | 3.0 (0.1) | −9.7 (0.5)   | 0.2 (0.7)    | −117.0 (0.6) | 39.4 (0.4)        |
| <sup>AN</sup> dG    | G <sub>1</sub> <sup>B</sup> | 22.6 (0.0) | 1.5 (0.1)  | 3.2 (0.1) | −8.5 (0.4)   | −5.3 (0.8)   | −119.9 (0.6) | 39.1 (0.3)        |
|                     | G <sub>2</sub> <sup>B</sup> | 22.5 (0.0) | 1.1 (0.2)  | 4.3 (0.0) | −12.7 (0.1)  | −5.5 (0.4)   | −121.8 (0.1) | 39.3 (0.4)        |
|                     | G <sub>3</sub> <sup>B</sup> | 22.4 (0.3) | 0.0 (0.1)  | 3.5 (0.3) | −16.3 (2.0)  | −0.5 (1.8)   | −117.4 (2.5) | 39.0 (0.8)        |
| <sup>AP</sup> dG    | G <sub>1</sub> <sup>B</sup> | 22.4 (0.0) | 2.3 (0.1)  | 4.2 (0.1) | −7.3 (1.4)   | −6.2 (1.0)   | −126.6 (2.0) | 46.5 (0.1)        |
|                     | G <sub>2</sub> <sup>B</sup> | 22.6 (0.2) | 0.9 (0.2)  | 4.1 (0.6) | −13.3 (1.5)  | −4.2 (2.2)   | −120.8 (3.4) | 37.6 (2.2)        |
|                     | G <sub>3</sub> <sup>B</sup> | 22.7 (0.0) | −0.5 (0.0) | 3.4 (0.1) | −14.7 (0.6)  | −2.2 (0.1)   | −121.6 (0.3) | 39.5 (0.2)        |
| <sup>AN</sup> dG    | G <sub>1</sub> <sup>S</sup> | 21.5 (1.2) | 2.4 (1.7)  | 6.0 (3.1) | −28.1 (14.1) | −22.7 (11.1) | −124.8 (5.1) | 49.0 (5.3)        |
|                     | G <sub>2</sub> <sup>S</sup> | 22.5 (0.1) | 0.7 (0.1)  | 3.7 (0.5) | −28.5 (5.1)  | −3.4 (2.3)   | −117.3 (0.7) | 45.1 (0.4)        |
|                     | G <sub>3</sub> <sup>S</sup> | 22.6 (1.4) | −0.8 (1.4) | 4.5 (3.0) | −20.6 (7.3)  | 14.9 (11.6)  | −119.8 (6.8) | 44.4 (7.3)        |
| <sup>AP</sup> dG    | G <sub>2</sub> <sup>S</sup> | 22.4 (0.2) | 2.9 (0.5)  | 4.8 (0.3) | −24.5 (5.6)  | −13.8 (3.2)  | −129.9 (4.5) | 44.1 (2.5)        |
|                     | G <sub>2</sub> <sup>S</sup> | 21.8 (0.4) | 1.5 (0.3)  | 6.1 (1.1) | −31.6 (3.9)  | −5.7 (1.2)   | −131.6 (2.5) | 46.8 (0.8)        |
|                     | G <sub>3</sub> <sup>S</sup> | 21.9 (0.4) | −2.5 (0.8) | 6.7 (1.0) | −21.6 (2.3)  | 14.5 (1.4)   | −134.0 (1.0) | 46.7 (1.2)        |

<sup>a</sup>The pseudostep is defined in Figure E.8. <sup>b</sup>Undamaged DNA.

Table E.15: Relative deviations in the pseudostep parameters from undamaged DNA, with the number of Xs representing the degree of distortion.<sup>a, b</sup>

| Adduct | Site(s)                                                 | Rise | Shift | Slide     | Roll      | Tilt | Twist     | Axial Bend |
|--------|---------------------------------------------------------|------|-------|-----------|-----------|------|-----------|------------|
| ANdG   | G <sub>1</sub> <sup>B</sup>                             | -    | -     | -         | -         | -    | -         | -          |
|        | G <sub>2</sub> <sup>B</sup>                             | -    | -     | -         | -         | -    | -         | -          |
|        | G <sub>3</sub> <sup>B</sup>                             | -    | -     | -         | -         | -    | -         | -          |
| APdG   | G <sub>1</sub> <sup>B</sup>                             | -    | -     | -         | -         | -    | -         | -          |
|        | G <sub>2</sub> <sup>B</sup>                             | -    | -     | -         | -         | -    | -         | -          |
|        | G <sub>3</sub> <sup>B</sup>                             | -    | -     | -         | -         | -    | -         | -          |
| ANdG   | G <sub>1</sub> <sup>S</sup>                             | -    | -     | X         | X         | XX   | -         | -          |
|        | G <sub>2</sub> <sup>S</sup>                             | -    | -     | -         | X         | -    | -         | -          |
|        | G <sub>3</sub> <sup>S</sup>                             | -    | -     | -         | X         | X    | -         | -          |
| APdG   | G <sub>1</sub> <sup>S</sup>                             | -    | X     | -         | X         | X    | X         | -          |
|        | G <sub>2</sub> <sup>S</sup>                             | -    | -     | X         | XX        | -    | X         | -          |
|        | G <sub>3</sub> <sup>S</sup>                             | -    | X     | X         | X         | X    | X         | -          |
| ANdG   | G <sub>1</sub> <sup>B</sup> G <sub>2</sub> <sup>B</sup> | -    | X     | X         | -         | -    | X         | -          |
|        | G <sub>1</sub> <sup>B</sup> G <sub>3</sub> <sup>B</sup> | -    | -     | -         | -         | -    | -         | -          |
|        | G <sub>2</sub> <sup>B</sup> G <sub>3</sub> <sup>B</sup> | -    | -     | X         | -         | -    | -         | -          |
| APdG   | G <sub>1</sub> <sup>B</sup> G <sub>2</sub> <sup>B</sup> | -    | X     | X         | -         | X    | X         | -          |
|        | G <sub>1</sub> <sup>B</sup> G <sub>3</sub> <sup>B</sup> | -    | -     | X         | -         | -    | X         | -          |
|        | G <sub>2</sub> <sup>B</sup> G <sub>3</sub> <sup>B</sup> | -    | -     | X         | X         | X    | X         | -          |
| ANdG   | G <sub>1</sub> <sup>B</sup> G <sub>2</sub> <sup>S</sup> | -    | -     | -         | -         | -    | -         | -          |
|        | G <sub>1</sub> <sup>B</sup> G <sub>3</sub> <sup>S</sup> | -    | -     | XX        | X         | X    | X         | X          |
|        | G <sub>2</sub> <sup>B</sup> G <sub>3</sub> <sup>S</sup> | -    | -     | XX        | X         | X    | X         | -          |
| APdG   | G <sub>1</sub> <sup>B</sup> G <sub>2</sub> <sup>S</sup> | -    | -     | X         | -         | -    | XXX       | -          |
|        | G <sub>1</sub> <sup>B</sup> G <sub>3</sub> <sup>S</sup> | X    | -     | XXX       | X         | X    | XX        | -          |
|        | G <sub>2</sub> <sup>B</sup> G <sub>3</sub> <sup>S</sup> | -    | X     | XX        | X         | -    | XX        | -          |
| ANdG   | G <sub>1</sub> <sup>S</sup> G <sub>2</sub> <sup>B</sup> | X    | X     | XX        | XX        | XXX  | X         | X          |
|        | G <sub>1</sub> <sup>S</sup> G <sub>3</sub> <sup>B</sup> | X    | X     | XX        | XXX       | XX   | -         | X          |
|        | G <sub>2</sub> <sup>S</sup> G <sub>3</sub> <sup>B</sup> | -    | -     | X         | XXX       | -    | -         | X          |
| APdG   | G <sub>1</sub> <sup>S</sup> G <sub>2</sub> <sup>B</sup> | -    | XX    | XX        | X         | X    | XXX       | -          |
|        | G <sub>1</sub> <sup>S</sup> G <sub>3</sub> <sup>B</sup> | -    | -     | X         | XX        | X    | XX        | -          |
|        | G <sub>2</sub> <sup>S</sup> G <sub>3</sub> <sup>B</sup> | -    | -     | -         | X         | -    | XXXX<br>X | -          |
| ANdG   | G <sub>1</sub> <sup>S</sup> G <sub>2</sub> <sup>S</sup> | -    | -     | -         | -         | -    | -         | -          |
|        | G <sub>1</sub> <sup>S</sup> G <sub>3</sub> <sup>S</sup> | XX   | -     | XXXX      | XXXX      | -    | X         | XX         |
|        | G <sub>2</sub> <sup>S</sup> G <sub>3</sub> <sup>S</sup> | -    | X     | XXX       | -         | -    | XXXX<br>X | -          |
| APdG   | G <sub>1</sub> <sup>S</sup> G <sub>2</sub> <sup>S</sup> | -    | XXX   | XXX       | XXXX      | X    | XXXX<br>X | -          |
|        | G <sub>1</sub> <sup>S</sup> G <sub>3</sub> <sup>S</sup> | X    | -     | XXX       | XX        | -    | XX        | X          |
|        | G <sub>2</sub> <sup>S</sup> G <sub>3</sub> <sup>S</sup> | -    | XX    | XXXX<br>X | XXXX<br>X | X    | XXXX<br>X | -          |

<sup>a</sup>The pseudostep is defined in Figure E.8. <sup>b</sup>Each X represents a 2 Å or 10° change from the value in undamaged DNA, with five Xs indicating changes > 12 Å or 60°.

Table E.16: Average (standard deviation in parentheses) solvent accessible surface area of the hydrophobic aryl bulky group (SASA, Å<sup>2</sup>) for monoadducted DNA.

| Adduct | Site                        | Replica 1    | Replica 2    | Replica 3    | SASA <sup>avg, a</sup> | Dataset Standard Deviation <sup>b</sup> | BSE <sup>c</sup> |
|--------|-----------------------------|--------------|--------------|--------------|------------------------|-----------------------------------------|------------------|
| ANdG   | G <sub>1</sub> <sup>B</sup> | 117.7 (6.1)  | 119.0 (6.9)  | 118.0 (6.5)  | 118.2 (0.7)            | 6.5                                     | 0.2              |
|        | G <sub>2</sub> <sup>B</sup> | 117.1 (6.0)  | 117.4 (6.1)  | 116.8 (5.9)  | 117.1 (0.3)            | 6.0                                     | 0.2              |
|        | G <sub>3</sub> <sup>B</sup> | 118.4 (6.0)  | 118.6 (5.9)  | 118.5 (5.8)  | 118.5 (0.1)            | 5.9                                     | 0.1              |
| APdG   | G <sub>1</sub> <sup>B</sup> | 188.7 (6.0)  | 184.3 (7.3)  | 187.1 (6.8)  | 186.7 (2.2)            | 6.9                                     | 11.2             |
|        | G <sub>2</sub> <sup>B</sup> | 180.5 (6.5)  | 179.8 (6.1)  | 179.9 (5.7)  | 180.0 (0.4)            | 6.1                                     | 16.5             |
|        | G <sub>3</sub> <sup>B</sup> | 181.0 (7.7)  | 180.6 (7.6)  | 180.3 (7.7)  | 180.6 (0.3)            | 7.6                                     | 18.8             |
| ANdG   | G <sub>1</sub> <sup>S</sup> | 66.3 (22.1)  | 59.4 (14.2)  | 86.5 (32.6)  | 70.7 (14.1)            | 26.8                                    | 2.4              |
|        | G <sub>2</sub> <sup>S</sup> | 70.2 (23.4)  | 88.2 (25.9)  | 73.7 (25.5)  | 77.4 (9.6)             | 26.2                                    | 2.1              |
|        | G <sub>3</sub> <sup>S</sup> | 69.2 (29.1)  | 74.9 (22.2)  | 58.7 (16.8)  | 67.6 (8.2)             | 24.2                                    | 2.6              |
| APdG   | G <sub>2</sub> <sup>S</sup> | 106.7 (9.9)  | 109.6 (14.8) | 113.7 (10.9) | 110.0 (3.5)            | 12.4                                    | 0.6              |
|        | G <sub>2</sub> <sup>S</sup> | 116.7 (21.1) | 108.0 (11.3) | 108.5 (11.1) | 111.1 (4.9)            | 15.8                                    | 1.1              |
|        | G <sub>3</sub> <sup>S</sup> | 107.0 (11.4) | 107.0 (10.1) | 104.3 (11.8) | 106.1 (1.6)            | 11.2                                    | 0.7              |

<sup>a</sup>Average SASA over the three replicates with the standard deviation (parentheses) calculated between the replicas. <sup>b</sup>Standard deviation calculated over the 1.5 μs dataset from the three replicas. <sup>c</sup>Block standard error (BSE) extrapolated from all frames of the three replicas fitted to a monomolecular growth model.

Table E.17: Average (standard deviation in parentheses)  $\chi$  dihedral angle (deg.) for each replica and across all replicas ( $\chi^{\text{avg}}$ ) for diadducted DNA.<sup>a</sup>

| Adduct | Primary Lesion | Secondary Lesion | Replica 1    | Replica 2    | Replica 3    | $\chi^{\text{avg}}$ | BSE <sup>b</sup> | $\Delta\chi^c$ |
|--------|----------------|------------------|--------------|--------------|--------------|---------------------|------------------|----------------|
| ANdG   | G <sup>P</sup> | G <sup>B</sup>   | 216.4 (22.5) | 223.1 (20.7) | 211.2 (23.6) | 216.9 (6.0)         | 1.6              | -11.9          |
|        |                | G <sup>S</sup>   | 227.1 (11.2) | 228.6 (10.9) | 228.0 (11.6) | 227.9 (0.8)         | 0.3              | -0.9           |
|        |                | G <sup>B</sup>   | 228.3 (15.7) | 229.4 (14.6) | 229.5 (14.2) | 229.1 (0.7)         | 0.7              | 0.3            |
|        |                | G <sup>S</sup>   | 228.3 (14.9) | 225.1 (17.0) | 230.9 (13.3) | 228.1 (2.9)         | 0.7              | -0.7           |
|        | G <sup>B</sup> | G <sup>P</sup>   | 226.8 (12.3) | 227.2 (12.1) | 226.4 (12.7) | 226.8 (0.4)         | 0.3              | 2.0            |
|        |                | G <sup>S</sup>   | 244.1 (22.6) | 242.5 (27.7) | 245.1 (30.0) | 243.9 (1.3)         | 2.3              | 19.1           |
|        |                | G <sup>B</sup>   | 220.7 (15.8) | 217.7 (17.8) | 218.8 (17.2) | 219.1 (1.5)         | 1.1              | -5.7           |
|        |                | G <sup>S</sup>   | 222.6 (11.0) | 221.3 (14.7) | 223.2 (10.8) | 222.4 (1.0)         | 0.6              | -2.4           |
|        | G <sup>S</sup> | G <sup>P</sup>   | 231.0 (12.6) | 231.5 (12.3) | 230.6 (12.6) | 231.0 (0.5)         | 0.2              | 0.0            |
|        |                | G <sup>S</sup>   | 228.9 (14.4) | 229.5 (13.1) | 229.1 (13.3) | 229.2 (0.3)         | 0.4              | -1.8           |
|        |                | G <sup>B</sup>   | 224.4 (16.7) | 227.2 (14.4) | 227.0 (14.5) | 226.2 (1.6)         | 0.8              | -4.8           |
|        |                | G <sup>S</sup>   | 229.9 (14.8) | 227.4 (15.8) | 228.6 (15.5) | 228.6 (1.3)         | 0.7              | -2.4           |
| APdG   | G <sup>P</sup> | G <sup>B</sup>   | 191.3 (21.7) | 189.8 (20.6) | 183.4 (15.8) | 188.2 (4.2)         | 12.6             | -14.5          |
|        |                | G <sup>S</sup>   | 222.8 (19.8) | 222.2 (21.6) | 211.1 (24.9) | 218.7 (6.6)         | 1.8              | 16.0           |
|        |                | G <sup>B</sup>   | 201.0 (20.3) | 198.6 (19.0) | 200.3 (19.6) | 200.0 (1.2)         | 2.0              | -2.8           |
|        |                | G <sup>S</sup>   | 211.7 (20.0) | 209.7 (20.6) | 210.6 (20.5) | 210.7 (1.0)         | 2.0              | 8.0            |
|        | G <sup>B</sup> | G <sup>P</sup>   | 214.4 (18.3) | 216.8 (14.4) | 217.4 (13.2) | 216.2 (1.6)         | 3.4              | -1.9           |
|        |                | G <sup>S</sup>   | 221.5 (18.1) | 218.2 (20.3) | 216.3 (19.5) | 218.7 (2.6)         | 2.4              | 0.5            |
|        |                | G <sup>B</sup>   | 201.4 (20.3) | 215.1 (14.8) | 212.4 (16.7) | 209.6 (7.3)         | 1.4              | -8.5           |
|        |                | G <sup>S</sup>   | 216.3 (15.0) | 210.2 (19.1) | 216.3 (15.5) | 214.3 (3.5)         | 1.2              | -3.9           |
|        | G <sup>S</sup> | G <sup>P</sup>   | 220.8 (20.1) | 209.4 (22.7) | 212.1 (22.0) | 214.1 (6.0)         | 3.2              | -3.6           |
|        |                | G <sup>S</sup>   | 218.4 (18.4) | 210.5 (22.1) | 203.8 (21.1) | 210.9 (7.3)         | 3.2              | -6.8           |
|        |                | G <sup>B</sup>   | 204.5 (25.4) | 203.9 (24.0) | 196.8 (22.4) | 201.8 (4.3)         | 3.8              | -15.9          |
|        |                | G <sup>S</sup>   | 203.5 (19.6) | 211.3 (20.8) | 207.2 (21.0) | 207.3 (3.9)         | 1.9              | -10.4          |
| ANdG   | G <sup>S</sup> | G <sup>B</sup>   | 54.5 (14.3)  | 56.9 (13.4)  | 57.3 (13.6)  | 56.2 (1.5)          | 0.5              | -1.4           |
|        |                | G <sup>S</sup>   | 53.4 (16.8)  | 48.5 (14.2)  | 49.9 (15.5)  | 50.6 (2.5)          | 0.3              | -7.1           |
|        |                | G <sup>B</sup>   | 49.4 (14.6)  | 48.8 (13.1)  | 47.3 (14.1)  | 48.5 (1.1)          | 0.6              | -9.1           |
|        |                | G <sup>S</sup>   | 50.1 (14.7)  | 48.0 (13.9)  | 55.3 (18.2)  | 51.1 (3.7)          | 0.3              | -6.5           |
|        | G <sup>S</sup> | G <sup>P</sup>   | 65.0 (16.8)  | 68.0 (15.7)  | 67.7 (15.1)  | 66.9 (1.6)          | 0.4              | 14.4           |
|        |                | G <sup>S</sup>   | 57.6 (12.8)  | 61.9 (12.8)  | 62.0 (12.6)  | 60.5 (2.5)          | 0.5              | 8.0            |
|        |                | G <sup>B</sup>   | 43.3 (14.0)  | 53.6 (17.6)  | 43.6 (14.0)  | 46.8 (5.8)          | 0.8              | -5.7           |
|        |                | G <sup>S</sup>   | 53.5 (23.7)  | 70.0 (36.1)  | 101.0 (34.4) | 74.8 (24.1)         | 3.3              | 22.3           |
|        | G <sup>S</sup> | G <sup>P</sup>   | 49.0 (17.2)  | 44.0 (14.8)  | 47.7 (14.3)  | 46.9 (2.6)          | 0.6              | -2.9           |
|        |                | G <sup>S</sup>   | 44.6 (15.3)  | 49.9 (15.1)  | 46.8 (16.0)  | 47.1 (2.6)          | 0.6              | -2.7           |
|        |                | G <sup>B</sup>   | 44.5 (15.1)  | 60.1 (17.3)  | 44.7 (15.2)  | 49.8 (9.0)          | 1.0              | -0.1           |
|        |                | G <sup>S</sup>   | 52.4 (15.9)  | 64.7 (15.7)  | 69.9 (14.1)  | 62.3 (8.9)          | 1.0              | 12.5           |
| APdG   | G <sup>S</sup> | G <sup>B</sup>   | 41.1 (13.3)  | 44.4 (24.6)  | 53.3 (35.9)  | 46.2 (6.3)          | 2.2              | 2.6            |
|        |                | G <sup>S</sup>   | 146.9 (26.0) | 129.5 (37.9) | 139.5 (21.5) | 138.6 (8.8)         | 5.1              | 95.0           |
|        |                | G <sup>B</sup>   | 45.8 (19.9)  | 42.1 (19.2)  | 40.6 (17.9)  | 42.9 (2.7)          | 0.5              | -0.8           |
|        |                | G <sup>S</sup>   | 45.1 (24.3)  | 50.5 (41.2)  | 43.6 (18.9)  | 46.4 (3.6)          | 1.7              | 2.8            |
|        | G <sup>S</sup> | G <sup>P</sup>   | 174.1 (20.6) | 176.3 (18.6) | 170.1 (24.9) | 173.5 (3.1)         | 10.3             | 123.9          |
|        |                | G <sup>S</sup>   | 190.8 (16.0) | 183.7 (17.9) | 188.7 (16.4) | 187.7 (3.7)         | 14.8             | 138.2          |
|        |                | G <sup>B</sup>   | 64.7 (35.1)  | 43.0 (25.7)  | 46.2 (32.0)  | 51.3 (11.7)         | 2.6              | 1.7            |
|        |                | G <sup>S</sup>   | 131.4 (32.4) | 135.5 (15.6) | 132.2 (17.3) | 133.1 (2.2)         | 5.2              | 83.5           |
|        | G <sup>S</sup> | G <sup>P</sup>   | 37.8 (16.4)  | 37.4 (15.5)  | 37.0 (15.5)  | 37.4 (0.4)          | 0.3              | -4.3           |
|        |                | G <sup>S</sup>   | 43.6 (27.2)  | 45.2 (41.5)  | 41.5 (21.9)  | 43.4 (1.9)          | 2.7              | 1.8            |
|        |                | G <sup>B</sup>   | 39.1 (29.5)  | 36.1 (18.0)  | 47.4 (45.1)  | 40.9 (5.9)          | 2.8              | -0.8           |
|        |                | G <sup>S</sup>   | 170.3 (18.9) | 171.7 (18.7) | 177.7 (21.0) | 173.2 (4.0)         | 15.6             | 131.6          |

<sup>a</sup> $\chi = \angle(\text{O4}'-\text{C1}'-\text{N9}-\text{C4})$ . See Figure 1 in the main text. <sup>b</sup>Block standard error (BSE) extrapolated from all frames of the three replicas fitted to a monomolecular growth model. <sup>c</sup>Calculated as the difference in  $\chi^{\text{avg}}$  for the lesion in diadducted DNA minus monoadducted DNA.

Table E.18: Average (standard deviation in parentheses)  $\theta$  dihedral angle (deg.) for each replica and across all replicas ( $\theta^{\text{avg}}$ ) for diadducted DNA.<sup>a</sup>

| Adduct | Primary Lesion              | Secondary Lesion            | Replica 1    | Replica 2    | Replica 3    | $\theta^{\text{avg}}$ | BSE <sup>b</sup> | $\Delta\theta^c$ |
|--------|-----------------------------|-----------------------------|--------------|--------------|--------------|-----------------------|------------------|------------------|
| ANdG   | G <sup>P</sup>              | G <sup>B</sup> <sub>2</sub> | 141.5 (37.3) | 134.9 (35.0) | 149.9 (39.6) | 142.1 (7.5)           | 7.2              | 14.0             |
|        |                             | G <sup>S</sup> <sub>2</sub> | 144.8 (21.3) | 139.8 (19.3) | 139.9 (20.2) | 141.5 (2.9)           | 1.2              | 13.4             |
|        |                             | G <sup>B</sup> <sub>3</sub> | 127.4 (27.9) | 125.2 (25.8) | 125.5 (25.0) | 126.0 (1.2)           | 1.5              | -2.1             |
|        |                             | G <sup>S</sup> <sub>3</sub> | 130.4 (26.7) | 135.5 (29.7) | 127.6 (23.9) | 131.2 (4.0)           | 1.0              | 3.0              |
|        | G <sup>B</sup> <sub>2</sub> | G <sup>P</sup>              | 143.8 (24.1) | 141.2 (23.0) | 146.6 (25.2) | 143.9 (2.7)           | 0.9              | 1.6              |
|        |                             | G <sup>S</sup> <sub>1</sub> | 134.7 (21.7) | 136.4 (18.3) | 138.7 (24.1) | 136.6 (2.0)           | 1.5              | -5.7             |
|        |                             | G <sup>B</sup> <sub>3</sub> | 147.8 (27.7) | 150.0 (29.6) | 149.3 (29.2) | 149.0 (1.1)           | 6.0              | 6.7              |
|        |                             | G <sup>S</sup> <sub>3</sub> | 148.9 (22.0) | 148.3 (26.8) | 147.7 (22.2) | 148.3 (0.6)           | 3.6              | 6.0              |
|        | G <sup>B</sup> <sub>3</sub> | G <sup>P</sup>              | 132.8 (19.2) | 132.8 (18.6) | 133.1 (19.3) | 132.9 (0.2)           | 0.2              | 0.3              |
|        |                             | G <sup>S</sup> <sub>1</sub> | 135.6 (21.7) | 134.4 (20.6) | 134.6 (20.8) | 134.9 (0.7)           | 0.5              | 2.3              |
|        |                             | G <sup>B</sup> <sub>2</sub> | 139.2 (26.4) | 135.4 (22.9) | 135.4 (23.8) | 136.7 (2.2)           | 1.1              | 4.1              |
|        |                             | G <sup>S</sup> <sub>2</sub> | 130.0 (22.3) | 133.2 (27.0) | 131.6 (23.4) | 131.6 (1.6)           | 2.5              | -1.0             |
| APdG   | G <sup>P</sup>              | G <sup>B</sup> <sub>2</sub> | 185.3 (19.9) | 186.0 (18.7) | 190.1 (16.5) | 187.1 (2.6)           | 13.6             | 10.0             |
|        |                             | G <sup>S</sup> <sub>2</sub> | 174.2 (18.6) | 172.7 (18.7) | 180.0 (20.4) | 175.7 (3.9)           | 16.5             | -1.5             |
|        |                             | G <sup>B</sup> <sub>3</sub> | 178.3 (41.3) | 180.8 (40.2) | 174.7 (43.0) | 178.0 (3.1)           | 18.4             | 0.8              |
|        |                             | G <sup>S</sup> <sub>3</sub> | 167.6 (39.8) | 172.5 (38.2) | 170.0 (38.4) | 170.1 (2.4)           | 15.9             | -7.1             |
|        | G <sup>B</sup> <sub>2</sub> | G <sup>P</sup>              | 176.0 (16.9) | 175.3 (16.1) | 177.0 (15.3) | 176.1 (0.9)           | 14.7             | 4.3              |
|        |                             | G <sup>S</sup> <sub>1</sub> | 169.6 (21.6) | 171.3 (21.9) | 173.5 (25.0) | 171.5 (2.0)           | 13.9             | -0.3             |
|        |                             | G <sup>B</sup> <sub>3</sub> | 175.6 (29.1) | 168.6 (24.4) | 170.3 (23.6) | 171.5 (3.7)           | 16.9             | -0.3             |
|        |                             | G <sup>S</sup> <sub>3</sub> | 174.2 (20.9) | 177.1 (23.0) | 174.6 (20.9) | 175.3 (1.6)           | 15.3             | 3.5              |
|        | G <sup>B</sup> <sub>3</sub> | G <sup>P</sup>              | 156.4 (32.4) | 167.2 (36.5) | 163.6 (36.0) | 162.4 (5.5)           | 14.6             | 5.2              |
|        |                             | G <sup>S</sup> <sub>1</sub> | 163.5 (31.1) | 169.0 (34.9) | 176.7 (33.5) | 169.7 (6.6)           | 16.7             | 12.5             |
|        |                             | G <sup>B</sup> <sub>2</sub> | 172.3 (35.4) | 176.2 (31.1) | 183.2 (29.7) | 177.3 (5.5)           | 17.9             | 20.1             |
|        |                             | G <sup>S</sup> <sub>2</sub> | 155.4 (38.7) | 165.1 (37.6) | 168.1 (39.3) | 162.9 (6.7)           | 13.4             | 5.7              |
| ANdG   | G <sup>S</sup> <sub>1</sub> | G <sup>B</sup> <sub>2</sub> | 117.3 (20.7) | 99.6 (21.5)  | 99.8 (46.5)  | 105.6 (10.2)          | 6.7              | -141.3           |
|        |                             | G <sup>S</sup> <sub>2</sub> | 246.5 (81.7) | 253.9 (29.0) | 250.2 (60.0) | 250.2 (3.7)           | 4.6              | 3.3              |
|        |                             | G <sup>B</sup> <sub>3</sub> | 251.0 (58.1) | 254.8 (30.1) | 248.4 (43.8) | 251.4 (3.2)           | 8.0              | 4.6              |
|        |                             | G <sup>S</sup> <sub>3</sub> | 249.7 (60.9) | 249.4 (34.4) | 101.0 (51.1) | 200.0 (85.8)          | 8.7              | -46.8            |
|        | G <sup>S</sup> <sub>2</sub> | G <sup>P</sup>              | 157.0 (83.6) | 89.8 (68.2)  | 93.9 (72.9)  | 113.6 (37.7)          | 10.7             | -82.9            |
|        |                             | G <sup>S</sup> <sub>1</sub> | 97.9 (25.3)  | 91.7 (18.8)  | 91.8 (20.5)  | 93.8 (3.6)            | 1.1              | -102.7           |
|        |                             | G <sup>B</sup> <sub>3</sub> | 261.5 (22.3) | 126.7 (88.3) | 262.3 (22.2) | 216.8 (78.0)          | 8.2              | 20.4             |
|        |                             | G <sup>S</sup> <sub>3</sub> | 261.7 (39.9) | 254.0 (41.9) | 222.0 (38.9) | 245.9 (21.1)          | 5.0              | 49.4             |
|        | G <sup>S</sup> <sub>3</sub> | G <sup>P</sup>              | 245.0 (74.8) | 246.2 (28.2) | 265.7 (39.1) | 252.3 (11.6)          | 7.5              | 56.2             |
|        |                             | G <sup>S</sup> <sub>1</sub> | 248.2 (29.7) | 267.4 (67.1) | 246.4 (58.9) | 254.0 (11.6)          | 8.0              | 57.9             |
|        |                             | G <sup>B</sup> <sub>2</sub> | 246.4 (22.5) | 89.6 (72.3)  | 241.8 (40.2) | 192.6 (89.2)          | 10.3             | -3.5             |
|        |                             | G <sup>S</sup> <sub>2</sub> | 262.9 (52.0) | 91.6 (57.9)  | 98.1 (80.9)  | 150.8 (97.1)          | 11.1             | -45.3            |
| APdG   | G <sup>S</sup> <sub>1</sub> | G <sup>B</sup> <sub>2</sub> | 184.9 (18.8) | 185.6 (19.5) | 188.1 (20.1) | 186.2 (1.7)           | 16.1             | -0.1             |
|        |                             | G <sup>S</sup> <sub>2</sub> | 179.1 (24.2) | 187.4 (19.0) | 188.8 (18.7) | 185.1 (5.2)           | 14.4             | -1.2             |
|        |                             | G <sup>B</sup> <sub>3</sub> | 206.2 (38.2) | 178.3 (28.7) | 172.8 (23.2) | 185.8 (17.9)          | 18.2             | -0.5             |
|        |                             | G <sup>S</sup> <sub>3</sub> | 213.4 (36.2) | 179.6 (26.2) | 189.9 (35.9) | 194.3 (17.3)          | 17.3             | 8.0              |
|        | G <sup>S</sup> <sub>2</sub> | G <sup>P</sup>              | 174.0 (14.4) | 174.8 (15.2) | 172.9 (16.4) | 173.9 (0.9)           | 1.5              | -20.1            |
|        |                             | G <sup>S</sup> <sub>1</sub> | 168.3 (15.8) | 172.5 (13.9) | 171.2 (14.5) | 170.7 (2.1)           | 12.2             | -23.3            |
|        |                             | G <sup>B</sup> <sub>3</sub> | 194.9 (24.1) | 189.0 (32.5) | 187.0 (29.5) | 190.3 (4.1)           | 17.0             | -3.7             |
|        |                             | G <sup>S</sup> <sub>3</sub> | 195.2 (16.9) | 202.7 (14.4) | 198.0 (14.0) | 198.6 (3.7)           | 5.7              | 4.6              |
|        | G <sup>S</sup> <sub>3</sub> | G <sup>P</sup>              | 195.7 (28.6) | 205.3 (27.2) | 202.4 (26.2) | 201.1 (5.0)           | 11.8             | 13.2             |
|        |                             | G <sup>S</sup> <sub>1</sub> | 220.3 (28.9) | 206.6 (28.8) | 219.2 (27.8) | 215.4 (7.6)           | 6.5              | 27.5             |
|        |                             | G <sup>B</sup> <sub>2</sub> | 204.2 (25.1) | 197.5 (27.9) | 198.0 (27.6) | 199.9 (3.7)           | 12.2             | 12.0             |
|        |                             | G <sup>S</sup> <sub>2</sub> | 167.6 (19.3) | 180.5 (23.3) | 166.9 (18.2) | 171.7 (7.7)           | 13.9             | -16.2            |

<sup>a</sup> $\theta = \angle(\text{N9}-\text{C8}-\text{N10}-\text{C10})$ . See Figure 1 in the main text. <sup>b</sup>Block standard error (BSE) extrapolated from all frames of the three replicas fitted to a monomolecular growth model.

<sup>c</sup>Calculated as the difference in  $\theta^{\text{avg}}$  for the lesion in diadducted DNA minus monoadducted DNA.



Table E.19: Average (standard deviation in parentheses)  $\phi$  dihedral angle (deg.) for each replica and across all replicas ( $\phi^{\text{avg}}$ ) for diadducted DNA.<sup>a</sup>

| Adduct | Primary Lesion | Secondary Lesion | Replica 1    | Replica 2    | Replica 3    | $\phi^{\text{avg}}$ | BSE <sup>b</sup> | $\Delta\phi^c$ |
|--------|----------------|------------------|--------------|--------------|--------------|---------------------|------------------|----------------|
| ANdG   | G <sup>P</sup> | G <sup>B</sup>   | 359.5 (12.8) | 359.9 (13.1) | 359.7 (12.3) | 359.7 (0.2)         | 0.4              | −0.8           |
|        |                | G <sup>S</sup>   | 4.9 (12.3)   | 3.1 (11.8)   | 2.6 (11.9)   | 3.5 (1.2)           | 0.3              | 3.0            |
|        |                | G <sup>B</sup>   | 0.6 (11.7)   | 0.4 (11.9)   | 359.9 (11.9) | 0.3 (0.3)           | 0.8              | −0.2           |
|        |                | G <sup>S</sup>   | 1.1 (12.0)   | 1.0 (12.2)   | 0.8 (12.0)   | 1.0 (0.1)           | 0.1              | 0.5            |
|        | G <sup>B</sup> | G <sup>P</sup>   | 4.8 (11.6)   | 4.9 (11.9)   | 5.2 (11.6)   | 5.0 (0.2)           | 0.2              | −1.1           |
|        |                | G <sup>S</sup>   | 358.7 (11.7) | 0.8 (11.5)   | 1.4 (11.7)   | 0.3 (1.2)           | 0.2              | −5.8           |
|        |                | G <sup>B</sup>   | 6.0 (12.4)   | 5.3 (12.4)   | 5.5 (12.4)   | 5.6 (0.4)           | 0.2              | −0.5           |
|        |                | G <sup>S</sup>   | 8.0 (12.2)   | 6.8 (12.5)   | 7.9 (12.3)   | 7.5 (0.6)           | 0.1              | 1.5            |
|        | G <sup>S</sup> | G <sup>P</sup>   | 2.7 (12.0)   | 2.3 (12.0)   | 2.6 (12.1)   | 2.5 (0.2)           | 0.2              | −0.1           |
|        |                | G <sup>S</sup>   | 2.6 (12.2)   | 3.3 (12.1)   | 2.8 (12.2)   | 2.9 (0.3)           | 0.1              | 0.3            |
|        |                | G <sup>B</sup>   | 2.6 (12.2)   | 3.0 (12.0)   | 2.8 (12.1)   | 2.8 (0.2)           | 0.2              | 0.2            |
|        |                | G <sup>S</sup>   | 1.0 (12.0)   | 1.8 (12.2)   | 0.4 (12.2)   | 1.1 (0.7)           | 1.6              | −1.6           |
| APdG   | G <sup>P</sup> | G <sup>B</sup>   | 10.5 (11.6)  | 10.3 (11.4)  | 8.4 (10.9)   | 9.7 (1.1)           | 0.4              | 4.2            |
|        |                | G <sup>S</sup>   | 12.0 (11.6)  | 13.1 (11.4)  | 11.0 (11.6)  | 12.0 (1.1)          | 0.3              | 6.5            |
|        |                | G <sup>B</sup>   | 4.5 (12.4)   | 5.4 (12.6)   | 4.3 (12.7)   | 4.8 (0.6)           | 0.3              | −0.7           |
|        |                | G <sup>S</sup>   | 7.4 (13.8)   | 7.9 (13.4)   | 8.2 (13.5)   | 7.8 (0.4)           | 0.6              | 2.3            |
|        | G <sup>B</sup> | G <sup>P</sup>   | 11.5 (11.7)  | 11.5 (11.6)  | 10.2 (11.3)  | 11.1 (0.7)          | 0.5              | −6.2           |
|        |                | G <sup>S</sup>   | 8.6 (12.3)   | 8.3 (12.6)   | 7.3 (12.4)   | 8.0 (0.7)           | 0.4              | −9.3           |
|        |                | G <sup>B</sup>   | 13.9 (12.9)  | 17.3 (12.3)  | 16.6 (12.1)  | 15.9 (1.8)          | 0.5              | −1.4           |
|        |                | G <sup>S</sup>   | 16.7 (12.1)  | 15.3 (12.5)  | 16.8 (12.0)  | 16.3 (0.8)          | 0.4              | −1.1           |
|        | G <sup>S</sup> | G <sup>P</sup>   | 11.8 (12.9)  | 8.7 (13.1)   | 8.6 (13.1)   | 9.7 (1.8)           | 0.5              | −1.2           |
|        |                | G <sup>S</sup>   | 12.2 (13.3)  | 9.6 (13.2)   | 9.2 (13.2)   | 10.3 (1.6)          | 0.5              | −0.6           |
|        |                | G <sup>B</sup>   | 8.3 (12.3)   | 10.6 (12.6)  | 7.8 (12.3)   | 8.9 (1.5)           | 0.4              | −2.0           |
|        |                | G <sup>S</sup>   | 358.9 (13.2) | 5.6 (13.6)   | 3.9 (13.7)   | 2.8 (2.8)           | 0.6              | −8.1           |
| ANdG   | G <sup>S</sup> | G <sup>B</sup>   | 358.0 (10.7) | 359.6 (10.1) | 359.6 (10.1) | 359.1 (0.9)         | 0.3              | −3.9           |
|        |                | G <sup>S</sup>   | 2.1 (10.5)   | 1.6 (10.6)   | 1.5 (10.6)   | 1.7 (0.3)           | 0.2              | −1.2           |
|        |                | G <sup>B</sup>   | 1.5 (10.7)   | 2.2 (10.5)   | 1.4 (10.6)   | 1.7 (0.4)           | 0.1              | −1.2           |
|        |                | G <sup>S</sup>   | 1.8 (10.5)   | 1.8 (10.7)   | 0.1 (10.6)   | 1.2 (1.0)           | 0.8              | −1.7           |
|        | G <sup>S</sup> | G <sup>P</sup>   | 0.9 (10.8)   | 1.2 (10.8)   | 1.1 (10.7)   | 1.0 (0.2)           | 0.2              | 1.1            |
|        |                | G <sup>S</sup>   | 359.9 (10.7) | 1.7 (10.7)   | 1.6 (10.7)   | 1.1 (0.8)           | 0.2              | 1.1            |
|        |                | G <sup>B</sup>   | 358.6 (10.2) | 359.8 (10.8) | 358.5 (10.1) | 359.0 (0.8)         | 0.2              | −1.0           |
|        |                | G <sup>S</sup>   | 2.4 (11.3)   | 2.4 (11.0)   | 1.5 (10.9)   | 2.1 (0.5)           | 0.2              | 2.2            |
|        | G <sup>S</sup> | G <sup>P</sup>   | 180.9 (9.6)  | 180.9 (9.5)  | 184.9 (10.4) | 182.2 (2.3)         | 13.0             | 1.2            |
|        |                | G <sup>S</sup>   | 180.5 (9.5)  | 183.4 (10.6) | 180.7 (9.5)  | 181.5 (1.6)         | 14.8             | 0.5            |
|        |                | G <sup>B</sup>   | 181.6 (9.3)  | 180.5 (9.9)  | 181.0 (9.4)  | 181.0 (0.6)         | 14.5             | 0.1            |
|        |                | G <sup>S</sup>   | 183.9 (10.8) | 181.2 (10.5) | 180.9 (10.6) | 182.0 (1.6)         | 15.7             | 1.0            |
| APdG   | G <sup>S</sup> | G <sup>B</sup>   | 353.3 (11.0) | 353.9 (11.1) | 354.5 (11.2) | 353.9 (0.6)         | 0.2              | −2.4           |
|        |                | G <sup>S</sup>   | 1.7 (11.7)   | 2.3 (11.5)   | 2.9 (11.3)   | 2.3 (0.6)           | 0.3              | 6.0            |
|        |                | G <sup>B</sup>   | 355.9 (11.0) | 356.3 (11.2) | 356.1 (11.2) | 356.1 (0.2)         | 0.1              | −0.2           |
|        |                | G <sup>S</sup>   | 357.3 (11.0) | 356.0 (11.8) | 356.6 (11.2) | 356.6 (0.6)         | 0.2              | 0.3            |
|        | G <sup>S</sup> | G <sup>P</sup>   | 354.3 (11.8) | 353.2 (11.6) | 355.8 (12.3) | 354.4 (1.3)         | 0.4              | 0.7            |
|        |                | G <sup>S</sup>   | 358.1 (11.1) | 358.6 (11.4) | 358.9 (11.2) | 358.5 (0.4)         | 0.4              | 4.8            |
|        |                | G <sup>B</sup>   | 355.1 (12.4) | 353.6 (11.6) | 354.2 (12.1) | 354.3 (0.7)         | 0.5              | 0.6            |
|        |                | G <sup>S</sup>   | 9.0 (12.7)   | 3.2 (12.1)   | 9.7 (11.9)   | 7.3 (3.5)           | 0.7              | 13.6           |
|        | G <sup>S</sup> | G <sup>P</sup>   | 352.6 (11.1) | 352.7 (11.2) | 352.5 (11.1) | 352.6 (0.1)         | 0.1              | −0.3           |
|        |                | G <sup>S</sup>   | 355.5 (11.7) | 354.4 (11.7) | 354.7 (11.2) | 354.9 (0.6)         | 0.4              | 2.0            |
|        |                | G <sup>B</sup>   | 354.5 (11.1) | 354.1 (11.0) | 356.2 (11.9) | 354.9 (1.1)         | 0.7              | 2.1            |
|        |                | G <sup>S</sup>   | 2.1 (13.2)   | 2.4 (14.1)   | 358.6 (13.4) | 1.1 (1.7)           | 0.6              | 8.2            |

<sup>a</sup> $\phi = \angle(\text{C8}–\text{N10}–\text{C10}–\text{C11})$ . See Figure 1 in the main text. <sup>b</sup>Block standard error (BSE) extrapolated from all frames of the three replicas fitted to a monomolecular growth model.

<sup>c</sup>Calculated as the difference in  $\phi^{\text{avg}}$  for the lesion in diadducted DNA minus monoadducted DNA.

Table E.20: Average (standard deviation in parentheses) solvent accessible surface area of the hydrophobic aryl bulky group (SASA, Å<sup>2</sup>) for diadducted DNA.

| Primary Lesion              | Secondary Lesion <sup>a</sup> | <sup>AN</sup> dG      |                                         |                  | <sup>AP</sup> dG      |                                         |                  |
|-----------------------------|-------------------------------|-----------------------|-----------------------------------------|------------------|-----------------------|-----------------------------------------|------------------|
|                             |                               | SASA <sup>avg b</sup> | Dataset Standard Deviation <sup>c</sup> | BSE <sup>d</sup> | SASA <sup>avg b</sup> | Dataset Standard Deviation <sup>c</sup> | BSE <sup>d</sup> |
| G <sub>1</sub> <sup>B</sup> | absent                        | 118.2 (0.7)           | 6.5                                     | 0.2              | 186.7 (2.2)           | 6.9                                     | 11.2             |
|                             | G <sub>2</sub> <sup>B</sup>   | 103.1 (2.9)           | 11.9                                    | 0.7              | 164.0 (0.8)           | 6.1                                     | 0.3              |
|                             | G <sub>2</sub> <sup>S</sup>   | 115.4 (1.7)           | 6.7                                     | 0.3              | 144.4 (1.7)           | 12.2                                    | 0.7              |
|                             | G <sub>3</sub> <sup>B</sup>   | 118.1 (0.1)           | 6.4                                     | 0.7              | 187.3 (0.2)           | 6.8                                     | 7.7              |
|                             | G <sub>3</sub> <sup>S</sup>   | 118.4 (0.7)           | 6.8                                     | 0.8              | 184.0 (0.0)           | 7.8                                     | 18.7             |
| G <sub>2</sub> <sup>B</sup> | absent                        | 117.1 (0.3)           | 6.0                                     | 0.2              | 180.0 (0.4)           | 6.1                                     | 16.5             |
|                             | G <sub>1</sub> <sup>B</sup>   | 95.0 (0.3)            | 10.7                                    | 0.2              | 149.0 (0.6)           | 7.5                                     | 0.4              |
|                             | G <sub>1</sub> <sup>S</sup>   | 104.9 (1.9)           | 12.0                                    | 0.8              | 161.8 (0.7)           | 8.0                                     | 0.5              |
|                             | G <sub>3</sub> <sup>B</sup>   | 118.6 (0.6)           | 6.6                                     | 0.3              | 172.4 (2.4)           | 15.9                                    | 16.8             |
|                             | G <sub>3</sub> <sup>S</sup>   | 115.6 (1.3)           | 6.4                                     | 0.2              | 179.5 (1.4)           | 7.9                                     | 16.2             |
| G <sub>3</sub> <sup>B</sup> | absent                        | 118.5 (0.1)           | 5.9                                     | 0.1              | 180.6 (0.3)           | 7.6                                     | 18.8             |
|                             | G <sub>1</sub> <sup>B</sup>   | 118.4 (0.1)           | 6.0                                     | 0.1              | 181.8 (1.7)           | 7.9                                     | 19.2             |
|                             | G <sub>1</sub> <sup>S</sup>   | 117.7 (0.2)           | 6.5                                     | 0.1              | 182.1 (1.9)           | 8.0                                     | 19.0             |
|                             | G <sub>2</sub> <sup>B</sup>   | 118.3 (0.2)           | 6.7                                     | 0.2              | 168.1 (0.6)           | 18.4                                    | 5.7              |
|                             | G <sub>2</sub> <sup>S</sup>   | 118.9 (0.5)           | 7.2                                     | 1.6              | 176.1 (10.2)          | 18.2                                    | 18.0             |
| G <sub>1</sub> <sup>S</sup> | absent                        | 70.7 (14.1)           | 26.8                                    | 2.4              | 110.0 (3.5)           | 12.4                                    | 0.6              |
|                             | G <sub>2</sub> <sup>B</sup>   | 64.6 (2.5)            | 22.5                                    | 1.5              | 112.9 (0.9)           | 11.8                                    | 0.4              |
|                             | G <sub>2</sub> <sup>S</sup>   | 72.1 (3.0)            | 20.6                                    | 0.8              | 80.9 (3.9)            | 12.2                                    | 0.7              |
|                             | G <sub>3</sub> <sup>B</sup>   | 69.2 (7.7)            | 22.8                                    | 1.8              | 106.2 (3.0)           | 12.3                                    | 0.7              |
|                             | G <sub>3</sub> <sup>S</sup>   | 62.6 (2.7)            | 17.1                                    | 1.5              | 111.8 (4.2)           | 12.6                                    | 0.7              |
| G <sub>2</sub> <sup>S</sup> | absent                        | 77.4 (9.6)            | 26.2                                    | 2.1              | 111.1 (4.9)           | 15.8                                    | 1.1              |
|                             | G <sub>1</sub> <sup>B</sup>   | 89.9 (2.8)            | 21.9                                    | 0.7              | 110.1 (2.1)           | 13.4                                    | 0.4              |
|                             | G <sub>1</sub> <sup>S</sup>   | 94.8 (1.0)            | 17.1                                    | 1.1              | 95.6 (0.5)            | 10.9                                    | 0.9              |
|                             | G <sub>3</sub> <sup>B</sup>   | 61.2 (14.8)           | 23.5                                    | 2.1              | 102.3 (1.0)           | 11.7                                    | 0.6              |
|                             | G <sub>3</sub> <sup>S</sup>   | 61.8 (5.0)            | 20.4                                    | 1.8              | 99.6 (2.0)            | 9.4                                     | 0.3              |
| G <sub>3</sub> <sup>S</sup> | absent                        | 67.6 (8.2)            | 24.2                                    | 2.6              | 106.1 (1.6)           | 11.2                                    | 0.7              |
|                             | G <sub>1</sub> <sup>B</sup>   | 59.7 (6.5)            | 19.6                                    | 1.6              | 110.1 (1.6)           | 10.1                                    | 0.4              |
|                             | G <sub>1</sub> <sup>S</sup>   | 57.7 (4.9)            | 18.9                                    | 1.8              | 111.8 (1.5)           | 13.2                                    | 0.9              |
|                             | G <sub>2</sub> <sup>B</sup>   | 68.1 (17.4)           | 26.8                                    | 2.7              | 108.7 (2.5)           | 11.7                                    | 0.9              |
|                             | G <sub>2</sub> <sup>S</sup>   | 93.6 (26.7)           | 30.2                                    | 3.8              | 103.6 (5.0)           | 13.5                                    | 0.8              |

<sup>a</sup>Absent indicates monoadducted DNA. <sup>b</sup>Average SASA over the three replicates, with the standard deviation (parentheses) calculated between the replicates. <sup>c</sup>Standard deviation calculated over the 1.5 μs dataset from the three replicas. <sup>d</sup>Block standard error (BSE) extrapolated from all frames of the three replicas fitted to a monomolecular growth model.

Table E.21: Occupancies (%) for the non-Watson-Crick hydrogen bonds at the diadducted DNA lesion sites in the B conformation.<sup>a,b</sup>

| Adduct | Primary Lesion              | Hydrogen Bond <sup>c</sup> | Secondary Lesion |                   |                   |                   |                   |                   |                   |
|--------|-----------------------------|----------------------------|------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
|        |                             |                            | None             | G <sub>1</sub> -B | G <sub>1</sub> -S | G <sub>2</sub> -B | G <sub>2</sub> -S | G <sub>3</sub> -B | G <sub>3</sub> -S |
| ANdG   | G <sub>1</sub> <sup>B</sup> | (X6)N10–H•••O3'(C5)        | 29.6             | N/A               | N/A               | 23.1              | 0.0               | 35.5              | 23.1              |
|        | G <sub>1</sub> <sup>B</sup> | (X6)N10–H•••O4'(X6)        | 50.3             | N/A               | N/A               | 50.0              | 80.7              | 46.8              | 56.5              |
|        | G <sub>2</sub> <sup>B</sup> | (X7)N10–H•••O4'(X7)        | 79.9             | 79.1              | 69.3              | N/A               | N/A               | 77.1              | 82.5              |
|        | G <sub>3</sub> <sup>B</sup> | (X9)N10–H•••O3'(C8)        | 0.0              | 0.0               | 0.0               | 0.0               | 42.3              | N/A               | N/A               |
|        | G <sub>3</sub> <sup>B</sup> | (X9)N10–H•••O4'(X9)        | 75.0             | 75.1              | 74.0              | 73.6              | 57.2              | N/A               | N/A               |
| APdG   | G <sub>1</sub> <sup>B</sup> | (X6)N10–H•••O4'(X6)        | 52.4             | N/A               | N/A               | 81.2              | 65.8              | 53.4              | 50.4              |
|        | G <sub>2</sub> <sup>B</sup> | (X7)N10–H•••O4'(X7)        | 47.9             | 69.1              | 74.9              | N/A               | N/A               | 58.9              | 51.5              |
|        | G <sub>3</sub> <sup>B</sup> | (X9)N10–H•••O4'(X9)        | 60.4             | 57.3              | 59.5              | 69.3              | 56.7              | N/A               | N/A               |

<sup>a</sup>Nucleotide numbering is illustrated in Figure E.8. <sup>b</sup>Occupancies determined as the percent of the simulation that a hydrogen bonds exist according to a heavy atom distance < 3.2 Å and a heavy atom–hydrogen–heavy atom angle > 120°. <sup>c</sup>X represents the damaged nucleotide

Table E.22: Average (standard deviation in parentheses) pseudostep parameters<sup>a</sup> within undamaged and diadducted DNA.

| Adduct              | Sites                                                   | Rise (Å)   | Shift (Å)  | Slide (Å)  | Roll (deg.)  | Tilt (deg.) | Twist (deg.)  | Axial Bend (deg.) |
|---------------------|---------------------------------------------------------|------------|------------|------------|--------------|-------------|---------------|-------------------|
| absent <sup>b</sup> | N/A                                                     | 22.7 (0.0) | 0.5 (0.1)  | 3.0 (0.1)  | −9.7 (0.5)   | 0.2 (0.7)   | −117.0 (0.6)  | 39.4 (0.4)        |
| AN <sub>d</sub> G   | G <sub>1</sub> <sup>B</sup> G <sub>2</sub> <sup>B</sup> | 22.3 (0.1) | 2.7 (0.2)  | 5.1 (0.2)  | −8.7 (0.3)   | −9.3 (1.2)  | −128.0 (1.3)  | 39.4 (0.0)        |
|                     | G <sub>1</sub> <sup>B</sup> G <sub>3</sub> <sup>B</sup> | 22.4 (0.0) | 1.1 (0.1)  | 4.1 (0.1)  | −13.5 (0.4)  | −6.1 (0.9)  | −122.5 (0.3)  | 38.8 (0.2)        |
|                     | G <sub>2</sub> <sup>B</sup> G <sub>3</sub> <sup>B</sup> | 22.2 (0.0) | 0.6 (0.1)  | 5.4 (0.1)  | −18.2 (0.4)  | −7.3 (0.8)  | −125.6 (0.7)  | 39.9 (0.2)        |
| AP <sub>d</sub> G   | G <sub>1</sub> <sup>B</sup> G <sub>2</sub> <sup>B</sup> | 22.0 (0.1) | 3.7 (0.1)  | 6.6 (0.1)  | −11.2 (0.9)  | −11.9 (0.6) | −135.6 (0.6)  | 40.0 (0.0)        |
|                     | G <sub>1</sub> <sup>B</sup> G <sub>3</sub> <sup>B</sup> | 22.3 (0.1) | 1.6 (0.0)  | 5.3 (0.2)  | −10.8 (1.3)  | −8.8 (0.5)  | −132.8 (0.8)  | 39.5 (0.2)        |
|                     | G <sub>2</sub> <sup>B</sup> G <sub>3</sub> <sup>B</sup> | 22.2 (0.1) | −0.2 (0.1) | 6.1 (0.2)  | −19.7 (1.9)  | −10.2 (0.8) | −131.5 (0.7)  | 41.2 (0.7)        |
| AN <sub>d</sub> G   | G <sub>1</sub> <sup>B</sup> G <sub>2</sub> <sup>S</sup> | 22.2 (0.0) | 2.0 (0.3)  | 4.3 (0.2)  | −14.7 (3.0)  | −9.2 (1.8)  | −119.7 (1.7)  | 41.0 (0.8)        |
|                     | G <sub>1</sub> <sup>B</sup> G <sub>3</sub> <sup>S</sup> | 21.3 (0.5) | −0.8 (0.4) | 8.3 (0.7)  | −25.9 (1.9)  | 19.0 (0.7)  | −132.9 (2.0)  | 50.0 (0.7)        |
|                     | G <sub>2</sub> <sup>B</sup> G <sub>3</sub> <sup>S</sup> | 20.9 (0.6) | −1.3 (0.8) | 8.6 (1.6)  | −28.0 (6.3)  | 16.8 (10.0) | −132.1 (2.6)  | 49.2 (4.2)        |
| AP <sub>d</sub> G   | G <sub>1</sub> <sup>B</sup> G <sub>2</sub> <sup>S</sup> | 24.3 (0.1) | 0.1 (0.1)  | −0.8 (0.1) | −14.2 (1.3)  | −9.6 (0.8)  | −151.2 (1.1)  | 38.5 (0.1)        |
|                     | G <sub>1</sub> <sup>B</sup> G <sub>3</sub> <sup>S</sup> | 20.7 (0.3) | −1.3 (0.2) | 9.7 (0.4)  | −23.2 (1.9)  | 11.2 (1.7)  | −140.9 (0.9)  | 44.7 (0.9)        |
|                     | G <sub>2</sub> <sup>B</sup> G <sub>3</sub> <sup>S</sup> | 21.0 (0.5) | −2.4 (0.2) | 8.9 (0.9)  | −23.8 (5.9)  | 9.0 (4.0)   | −142.4 (6.0)  | 45.7 (1.3)        |
| AN <sub>d</sub> G   | G <sub>1</sub> <sup>S</sup> G <sub>2</sub> <sup>B</sup> | 20.6 (1.1) | 3.3 (0.9)  | 8.7 (1.0)  | −35.9 (4.3)  | −32.5 (1.6) | −127.6 (2.2)  | 54.7 (1.9)        |
|                     | G <sub>1</sub> <sup>S</sup> G <sub>3</sub> <sup>B</sup> | 20.1 (0.2) | 2.8 (0.3)  | 8.9 (0.3)  | −43.0 (2.8)  | −28.4 (2.7) | −126.9 (0.2)  | 55.0 (2.4)        |
|                     | G <sub>1</sub> <sup>S</sup> G <sub>3</sub> <sup>S</sup> | 21.6 (0.8) | 0.8 (0.6)  | 6.5 (2.0)  | −47.3 (12.2) | −7.0 (1.1)  | −125.7 (3.6)  | 55.3 (5.2)        |
| AP <sub>d</sub> G   | G <sub>1</sub> <sup>S</sup> G <sub>2</sub> <sup>B</sup> | 22.0 (0.1) | 4.7 (0.3)  | 7.0 (0.1)  | −21.7 (1.5)  | −10.7 (1.0) | −148.4 (1.3)  | 41.8 (0.3)        |
|                     | G <sub>1</sub> <sup>S</sup> G <sub>3</sub> <sup>B</sup> | 21.5 (0.2) | 2.3 (0.1)  | 6.6 (0.4)  | −30.5 (3.4)  | −13.8 (0.9) | −139.6 (1.7)  | 47.7 (2.5)        |
|                     | G <sub>1</sub> <sup>S</sup> G <sub>3</sub> <sup>S</sup> | 21.5 (0.3) | 1.0 (0.2)  | 3.5 (6.1)  | −27.2 (13.4) | −6.8 (5.5)  | −37.9 (179.4) | 45.4 (10.9)       |
| AN <sub>d</sub> G   | G <sub>1</sub> <sup>S</sup> G <sub>2</sub> <sup>S</sup> | 22.6 (0.2) | 0.5 (1.1)  | 1.7 (1.7)  | −17.0 (4.5)  | −1.3 (8.0)  | −116.7 (0.8)  | 41.7 (0.1)        |
|                     | G <sub>1</sub> <sup>S</sup> G <sub>3</sub> <sup>S</sup> | 18.3 (0.6) | 0.8 (0.5)  | 12.6 (0.4) | −56.9 (1.7)  | −1.7 (4.2)  | −135.8 (2.7)  | 61.5 (0.2)        |
|                     | G <sub>2</sub> <sup>S</sup> G <sub>3</sub> <sup>S</sup> | 22.3 (1.9) | −2.3 (1.3) | 10.1 (0.3) | −12.8 (40.0) | −2.6 (14.9) | −47.5 (175.1) | 46.4 (14.2)       |
| AP <sub>d</sub> G   | G <sub>1</sub> <sup>S</sup> G <sub>2</sub> <sup>S</sup> | 22.9 (0.9) | 7.5 (0.5)  | 10.6 (0.6) | 32.3 (3.3)   | 19.4 (4.0)  | 138.2 (3.8)   | 29.5 (2.6)        |
|                     | G <sub>1</sub> <sup>S</sup> G <sub>3</sub> <sup>S</sup> | 20.3 (0.0) | −0.1 (0.2) | 9.3 (0.7)  | −38.8 (3.6)  | 5.9 (0.8)   | −144.6 (6.2)  | 50.4 (2.1)        |
|                     | G <sub>2</sub> <sup>S</sup> G <sub>3</sub> <sup>S</sup> | 22.4 (1.1) | −5.2 (1.7) | 14.5 (0.6) | 50.3 (7.1)   | −17.0 (5.7) | 116.0 (8.8)   | 41.0 (4.2)        |

<sup>a</sup>The pseudostep is defined in Figure E.8. <sup>b</sup>Undamaged DNA.

Table E.23: Occupancies (%) for the non-Watson-Crick hydrogen bonds at the diadducted DNA damage sites in the S conformation.<sup>a,b</sup>

| Adduct            | Primary Lesion              | Hydrogen Bond <sup>c</sup> | Secondary Lesion    |                             |                             |                             |                             |                             |                             |
|-------------------|-----------------------------|----------------------------|---------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
|                   |                             |                            | Absent <sup>d</sup> | G <sub>1</sub> <sup>B</sup> | G <sub>1</sub> <sup>S</sup> | G <sub>2</sub> <sup>B</sup> | G <sub>2</sub> <sup>S</sup> | G <sub>3</sub> <sup>B</sup> | G <sub>3</sub> <sup>S</sup> |
| AN <sub>d</sub> G | G <sub>1</sub> <sup>S</sup> | (C27)N4–H•••O6(X6)         | 21.2                | N/A                         | N/A                         | 8.0                         | 50.3                        | 0.0                         | 0.0                         |
|                   |                             | (X6)N10–H•••N1(G28)        | 13.5                | N/A                         | N/A                         | 13.7                        | 0.0                         | 18.1                        | 22.9                        |
|                   |                             | (X6)N10–H•••N2(G28)        | 32.2                | N/A                         | N/A                         | 49.2                        | 0.0                         | 47.7                        | 48.4                        |
|                   |                             | (X6)N10–H•••N3(G28)        | 0.0                 | N/A                         | N/A                         | 10.8                        | 0.0                         | 0.0                         | 10.3                        |
|                   |                             | (X6)N10–H•••O4'(X7)        | 17.7                | N/A                         | N/A                         | 0.0                         | 93.7                        | 0.0                         | 0.0                         |
|                   |                             | (C27)N4–H•••OP2(C26)       | 7.9                 | N/A                         | N/A                         | 0.0                         | 0.0                         | 0.0                         | 6.3                         |
|                   | G <sub>2</sub> <sup>S</sup> | (C26)N4–H•••OP2(G25)       | 0.0                 | 6.8                         | 0.0                         | N/A                         | N/A                         | 0.0                         | 15.8                        |
|                   |                             | (C26)N4–H•••O6(X7)         | 35.0                | 76.0                        | 88.5                        | N/A                         | N/A                         | 9.7                         | 15.2                        |
|                   |                             | (C26)N4–H•••N7(X7)         | 15.4                | 0.0                         | 0.0                         | N/A                         | N/A                         | 0.0                         | 0.0                         |
|                   |                             | (X6)N10–H•••O4'(X7)        | 17.7                | 0.0                         | 93.7                        | N/A                         | N/A                         | 0.0                         | 0.0                         |
|                   |                             | (X7)N10–H•••O4'(C8)        | 20.6                | 37.8                        | 80.8                        | N/A                         | N/A                         | 10.6                        | 0.0                         |
|                   |                             | (X7)N10–H•••O2(C27)        | 11.9                | 0.0                         | 0.0                         | N/A                         | N/A                         | 14.2                        | 0.0                         |
|                   |                             | (X7)N10–H•••N2(G6)         | 23.1                | 0.0                         | 0.0                         | N/A                         | N/A                         | 25.4                        | 36.0                        |
|                   |                             | (X7)N10–H•••O4'(X7)        | 0.0                 | 0.0                         | 0.0                         | N/A                         | N/A                         | 0.0                         | 17.5                        |
|                   |                             | (X7)N1–H•••O6(G25)         | 0.0                 | 8.8                         | 12.4                        | N/A                         | N/A                         | 0.0                         | 0.0                         |
|                   | G <sub>3</sub> <sup>S</sup> | (C24)N4–H•••O6(X9)         | 12.4                | 9.8                         | 9.8                         | 16.6                        | 53.1                        | N/A                         | N/A                         |
|                   |                             | (C24)N4–H•••N7(X9)         | 0.0                 | 0.0                         | 0.0                         | 0.0                         | 21.8                        | N/A                         | N/A                         |
|                   |                             | (X9)N10–H•••O2(C10)        | 0.0                 | 0.0                         | 0.0                         | 0.0                         | 15.4                        | N/A                         | N/A                         |
|                   |                             | (X9)N10–H•••O4'(C10)       | 11.3                | 0.0                         | 0.0                         | 15.2                        | 16.8                        | N/A                         | N/A                         |
|                   |                             | (X9)N10–H•••N1(G25)        | 0.0                 | 9.3                         | 9.1                         | 0.0                         | 0.0                         | N/A                         | N/A                         |
|                   |                             | (X9)N10–H•••N2(G25)        | 33.5                | 44.1                        | 46.6                        | 36.9                        | 11.9                        | N/A                         | N/A                         |
|                   |                             | (X9)N10–H•••N3(G25)        | 26.8                | 23.8                        | 26.7                        | 12.6                        | 0.0                         | N/A                         | N/A                         |
| AP <sub>d</sub> G | G <sub>1</sub> <sup>S</sup> | (C27)N4–H•••OP2(C26)       | 31.3                | N/A                         | N/A                         | 6.7                         | 0.0                         | 49.4                        | 28.3                        |
|                   |                             | (X6)N10–H•••N2(G28)        | 14.1                | N/A                         | N/A                         | 0.0                         | 0.0                         | 10.6                        | 13.5                        |
|                   |                             | (C26)N4–H•••O6(X6)         | 9.9                 | N/A                         | N/A                         | 0.0                         | 0.0                         | 0.0                         | 0.0                         |
|                   |                             | (X6)N10–H•••O4'(X6)        | 0.0                 | N/A                         | N/A                         | 0.0                         | 90.6                        | 0.0                         | 0.0                         |
|                   | G <sub>2</sub> <sup>S</sup> | (X7)N10–H•••O4'(X7)        | 0.0                 | 97.5                        | 98.2                        | N/A                         | N/A                         | 11.4                        | 95.7                        |
|                   |                             | (C26)N4–H•••OP2(G25)       | 25.7                | 62.1                        | 83.0                        | N/A                         | N/A                         | 40.1                        | 0.0                         |
|                   | G <sub>3</sub> <sup>S</sup> | (C24)N4–H•••OP2(G23)       | 36.7                | 19.4                        | 8.7                         | 20.0                        | 53.5                        | N/A                         | N/A                         |
|                   |                             | (X9)N10–H•••N2(G25)        | 0.0                 | 10.4                        | 20.2                        | 9.6                         | 0.0                         | N/A                         | N/A                         |
|                   |                             | (X9)N10–H•••N3(G25)        | 0.0                 | 0.0                         | 10.2                        | 0.0                         | 0.0                         | N/A                         | N/A                         |
|                   |                             | (X9)N10–H•••O4'(X9)        | 0.0                 | 0.0                         | 0.0                         | 0.0                         | 93.6                        | N/A                         | N/A                         |

<sup>a</sup>Nucleotide numbering is illustrated in Figure E.8. <sup>b</sup>Occupancies determined as the percent of the simulation that a hydrogen bonds exist according to a heavy atom distance < 3.2 Å and a heavy atom–hydrogen–heavy atom angle > 120°. <sup>c</sup>X represents the damaged nucleotide. <sup>d</sup>Absent indicates monoadducted DNA.

Table E.24: Average (standard deviation in parentheses) lesion site interaction energies (kJ/mol) for monoadducted and diadducted DNA duplexes in the B conformation.

| Adduct | Primary Lesion | Secondary Lesion <sup>a</sup> | LSIE <sup>avg b</sup> | Dataset Standard Deviation <sup>c</sup> | LSIE BSE <sup>d</sup> | Average Stacking <sup>e</sup> | Dataset Standard Deviation <sup>c</sup> | Average Hydrogen Bonding <sup>f</sup> | Dataset Standard Deviation <sup>c</sup> |
|--------|----------------|-------------------------------|-----------------------|-----------------------------------------|-----------------------|-------------------------------|-----------------------------------------|---------------------------------------|-----------------------------------------|
| ANdG   | G <sup>B</sup> | absent                        | −213.9 (0.4)          | 15.8                                    | 0.2                   | −104.5 (0.5)                  | 12.9                                    | −109.4 (0.2)                          | 8.7                                     |
|        |                | G <sup>B</sup>                | −223.4 (0.2)          | 15.4                                    | 0.2                   | −113.6 (0.2)                  | 13.2                                    | −109.8 (0.1)                          | 8.0                                     |
|        |                | G <sup>S</sup>                | −206.0 (5.4)          | 29.0                                    | 0.4                   | −108.4 (1.2)                  | 19.6                                    | −97.6 (4.2)                           | 16.2                                    |
|        |                | G <sup>B</sup>                | −214.4 (0.4)          | 15.4                                    | 0.2                   | −104.9 (0.2)                  | 12.6                                    | −109.5 (0.1)                          | 8.6                                     |
|        |                | G <sup>S</sup>                | −212.2 (0.5)          | 17.5                                    | 0.2                   | −103.6 (0.5)                  | 13.3                                    | −108.7 (0.2)                          | 10.3                                    |
|        | G <sup>B</sup> | absent                        | −209.9 (0.2)          | 15.2                                    | 0.0                   | −99.8 (0.2)                   | 12.9                                    | −110.2 (0.1)                          | 7.3                                     |
|        |                | G <sup>B</sup>                | −218.9 (0.4)          | 15.9                                    | 0.3                   | −108.3 (0.5)                  | 14.0                                    | −110.6 (0.1)                          | 7.0                                     |
|        |                | G <sup>S</sup>                | −203.7 (34.4)         | 40.9                                    | 0.9                   | −94.8 (33.5)                  | 38.9                                    | −108.9 (1.0)                          | 8.5                                     |
|        |                | G <sup>B</sup>                | −208.9 (0.1)          | 15.8                                    | 0.1                   | −99.0 (0.0)                   | 13.2                                    | −109.9 (0.1)                          | 7.5                                     |
|        |                | G <sup>S</sup>                | −213.4 (1.5)          | 15.3                                    | 0.4                   | −103.0 (1.2)                  | 13.0                                    | −110.4 (0.3)                          | 7.1                                     |
|        | G <sup>B</sup> | absent                        | −239.4 (0.2)          | 14.6                                    | 0.1                   | −129.6 (0.1)                  | 12.7                                    | −109.8 (0.1)                          | 7.5                                     |
|        |                | G <sup>B</sup>                | −238.2 (0.2)          | 15.1                                    | 0.1                   | −128.6 (0.2)                  | 12.9                                    | −109.6 (0.0)                          | 7.8                                     |
|        |                | G <sup>S</sup>                | −237.1 (0.3)          | 16.2                                    | 0.2                   | −127.8 (0.1)                  | 13.2                                    | −109.3 (0.1)                          | 8.4                                     |
|        |                | G <sup>B</sup>                | −238.1 (0.3)          | 14.9                                    | 0.2                   | −128.5 (0.3)                  | 13.1                                    | −109.6 (0.0)                          | 7.6                                     |
|        |                | G <sup>S</sup>                | −241.4 (1.9)          | 15.3                                    | 0.2                   | −131.4 (1.7)                  | 12.9                                    | −110.0 (0.2)                          | 7.9                                     |
| APdG   | G <sup>B</sup> | absent                        | −201.5 (0.7)          | 15.5                                    | 2.6                   | −96.7 (0.4)                   | 13.9                                    | −104.7 (0.4)                          | 8.0                                     |
|        |                | G <sup>B</sup>                | −195.8 (0.7)          | 18.7                                    | 0.1                   | −93.8 (0.7)                   | 16.9                                    | −102.0 (0.1)                          | 9.0                                     |
|        |                | G <sup>S</sup>                | −227.3 (4.7)          | 33.5                                    | 0.6                   | −127.7 (3.6)                  | 26.1                                    | −99.6 (2.9)                           | 15.8                                    |
|        |                | G <sup>B</sup>                | −201.1 (0.4)          | 15.6                                    | 0.7                   | −96.8 (0.4)                   | 13.9                                    | −104.3 (0.1)                          | 8.2                                     |
|        |                | G <sup>S</sup>                | −200.6 (0.3)          | 16.3                                    | 0.5                   | −95.3 (0.2)                   | 14.2                                    | −105.2 (0.1)                          | 8.3                                     |
|        | G <sup>B</sup> | absent                        | −226.3 (0.3)          | 17.1                                    | 0.2                   | −119.2 (0.4)                  | 14.8                                    | −107.1 (0.3)                          | 6.9                                     |
|        |                | G <sup>B</sup>                | −224.8 (0.9)          | 19.7                                    | 0.1                   | −117.3 (0.8)                  | 18.1                                    | −107.5 (0.1)                          | 6.6                                     |
|        |                | G <sup>S</sup>                | −227.6 (1.0)          | 22.1                                    | 0.2                   | −121.4 (0.7)                  | 20.3                                    | −106.1 (0.4)                          | 8.4                                     |
|        |                | G <sup>B</sup>                | −223.3 (2.0)          | 17.5                                    | 0.1                   | −117.7 (1.4)                  | 15.0                                    | −105.6 (0.5)                          | 7.1                                     |
|        |                | G <sup>S</sup>                | −228.1 (1.4)          | 18.7                                    | 0.2                   | −120.9 (1.2)                  | 16.0                                    | −107.3 (0.3)                          | 6.9                                     |
|        | G <sup>B</sup> | absent                        | −246.4 (0.4)          | 16.8                                    | 0.9                   | −140.5 (0.3)                  | 14.5                                    | −105.9 (0.1)                          | 7.7                                     |
|        |                | G <sup>B</sup>                | −244.5 (1.3)          | 16.8                                    | 1.0                   | −138.8 (1.1)                  | 14.7                                    | −105.7 (0.2)                          | 7.7                                     |
|        |                | G <sup>S</sup>                | −243.5 (1.5)          | 16.7                                    | 1.3                   | −137.9 (1.2)                  | 14.6                                    | −105.6 (0.3)                          | 7.6                                     |
|        |                | G <sup>B</sup>                | −242.3 (1.3)          | 17.9                                    | 0.1                   | −137.3 (1.0)                  | 15.8                                    | −105.1 (0.3)                          | 7.6                                     |
|        |                | G <sup>S</sup>                | −244.3 (2.7)          | 18.2                                    | 0.2                   | −139.1 (3.0)                  | 16.3                                    | −105.1 (0.4)                          | 7.7                                     |

<sup>a</sup>Absent indicates monoadducted DNA. <sup>b</sup>Average of the sum of stacking and hydrogen-bonding interactions for the adducted base, bulky moiety and opposing C, with standard deviation calculated between the replicas. <sup>c</sup>Standard deviation calculated over the 1.5  $\mu$ s dataset from the three replicas. <sup>d</sup>Block standard error (BSE) extrapolated from all frames of the three replicas fitted to a monomolecular growth model. <sup>e</sup>Average of the stacking between the adduct (base and bulky moiety) plus the opposing C and the flanking nucleobases, with standard deviation calculated between the replicas. <sup>f</sup>Average of the hydrogen-bonding interaction between the adduct and the pairing C, with standard deviation calculated between the replicas.

Table E.25: Average (standard deviation in parentheses) lesion site interaction energy (kJ/mol) for monoadducted and diadducted DNA duplexes in the S conformation.

| Adduct | Primary Lesion              | Secondary Lesion <sup>a</sup> | LSIE <sup>avg, b</sup> | Dataset Standard Deviation <sup>c</sup> | LSIE BSE <sup>d</sup> | Average Stacking <sup>e</sup> | Dataset Standard Deviation <sup>c</sup> | Average Hydrogen Bonding <sup>f</sup> | Dataset Standard Deviation <sup>c</sup> |
|--------|-----------------------------|-------------------------------|------------------------|-----------------------------------------|-----------------------|-------------------------------|-----------------------------------------|---------------------------------------|-----------------------------------------|
| ANdG   | G <sub>1</sub> <sup>S</sup> | absent                        | −24.1 (65.1)           | 74.9                                    | 1.7                   | −40.0 (35.8)                  | 44.4                                    | 15.9 (29.3)                           | 33.5                                    |
|        |                             | G <sub>2</sub> <sup>B</sup>   | −50.9 (89.2)           | 96.0                                    | 2.4                   | −58.0 (48.4)                  | 55.6                                    | 7.1 (40.8)                            | 42.7                                    |
|        |                             | G <sub>2</sub> <sup>S</sup>   | −20.2 (66.9)           | 74.5                                    | 1.8                   | −25.9 (45.1)                  | 51.4                                    | 5.7 (21.8)                            | 25.5                                    |
|        |                             | G <sub>3</sub> <sup>B</sup>   | 25.5 (1.8)             | 27.4                                    | 0.5                   | −13.6 (1.2)                   | 24.4                                    | 39.1 (0.9)                            | 8.3                                     |
|        |                             | G <sub>3</sub> <sup>S</sup>   | 19.9 (4.1)             | 36.4                                    | 0.7                   | −17.2 (4.0)                   | 29.2                                    | 37.1 (0.4)                            | 11.2                                    |
|        | G <sub>2</sub> <sup>S</sup> | absent                        | −33.7 (25.1)           | 72.0                                    | 1.5                   | −48.7 (15.0)                  | 46.2                                    | 15.1 (10.1)                           | 30.1                                    |
|        |                             | G <sub>1</sub> <sup>B</sup>   | −67.7 (1.8)            | 37.7                                    | 0.5                   | −78.0 (1.5)                   | 28.4                                    | 10.3 (0.2)                            | 17.2                                    |
|        |                             | G <sub>1</sub> <sup>S</sup>   | −27.6 (38.7)           | 55.8                                    | 1.0                   | −39.8 (37.3)                  | 46.4                                    | 12.3 (1.5)                            | 18.0                                    |
|        |                             | G <sub>3</sub> <sup>B</sup>   | 2.4 (36.2)             | 57.2                                    | 1.2                   | −25.2 (22.0)                  | 37.7                                    | 27.6 (14.2)                           | 23.1                                    |
|        |                             | G <sub>3</sub> <sup>S</sup>   | −156.8 (18.3)          | 48.2                                    | 1.3                   | −117.3 (9.1)                  | 29.5                                    | −39.5 (9.3)                           | 21.8                                    |
|        | G <sub>3</sub> <sup>S</sup> | absent                        | −39.1 (64.2)           | 91.3                                    | 2.1                   | −47.9 (39.7)                  | 57.5                                    | 8.9 (24.6)                            | 36.5                                    |
|        |                             | G <sub>1</sub> <sup>B</sup>   | −18.0 (87.0)           | 92.8                                    | 2.3                   | −28.2 (44.2)                  | 50.4                                    | 10.3 (42.8)                           | 44.8                                    |
|        |                             | G <sub>1</sub> <sup>S</sup>   | −23.4 (88.8)           | 95.0                                    | 2.4                   | −31.9 (45.6)                  | 51.5                                    | 8.5 (43.2)                            | 45.8                                    |
|        |                             | G <sub>2</sub> <sup>B</sup>   | 21.5 (36.4)            | 61.4                                    | 1.3                   | −10.5 (22.9)                  | 40.5                                    | 32.1 (13.5)                           | 24.4                                    |
|        |                             | G <sub>2</sub> <sup>S</sup>   | −117.3 (26.0)          | 59.8                                    | 1.3                   | −83.4 (11.7)                  | 35.9                                    | −33.9 (15.5)                          | 28.1                                    |
| APdG   | G <sub>1</sub> <sup>S</sup> | absent                        | −123.3 (36.2)          | 51.6                                    | 1.1                   | −103.4 (21.6)                 | 33.7                                    | −19.9 (15.1)                          | 23.8                                    |
|        |                             | G <sub>2</sub> <sup>B</sup>   | −85.2 (1.6)            | 25.6                                    | 0.2                   | −82.4 (1.4)                   | 24.2                                    | −2.8 (0.3)                            | 3.5                                     |
|        |                             | G <sub>2</sub> <sup>S</sup>   | −93.9 (7.0)            | 32.1                                    | 0.5                   | −81.8 (9.1)                   | 31.2                                    | −12.1 (2.7)                           | 9.0                                     |
|        |                             | G <sub>3</sub> <sup>B</sup>   | −77.6 (5.5)            | 46.0                                    | 1.0                   | −78.4 (3.7)                   | 32.3                                    | 0.8 (1.8)                             | 16.0                                    |
|        |                             | G <sub>3</sub> <sup>S</sup>   | −119.8 (39.7)          | 56.9                                    | 1.2                   | −103.7 (25.1)                 | 39.5                                    | −16.1 (14.6)                          | 22.5                                    |
|        | G <sub>2</sub> <sup>S</sup> | absent                        | −101.2 (7.4)           | 49.6                                    | 0.8                   | −106.8 (4.4)                  | 37.6                                    | 5.6 (3.6)                             | 16.1                                    |
|        |                             | G <sub>1</sub> <sup>B</sup>   | −138.2 (2.4)           | 22.7                                    | 0.2                   | −135.2 (3.1)                  | 23.4                                    | −3.0 (0.8)                            | 4.7                                     |
|        |                             | G <sub>1</sub> <sup>S</sup>   | −113.0 (10.6)          | 40.0                                    | 1.1                   | −105.8 (10.6)                 | 38.6                                    | −7.1 (0.4)                            | 4.9                                     |
|        |                             | G <sub>3</sub> <sup>B</sup>   | −108.4 (3.8)           | 41.0                                    | 0.7                   | −110.3 (3.4)                  | 32.9                                    | 1.9 (2.2)                             | 11.6                                    |
|        |                             | G <sub>3</sub> <sup>S</sup>   | −159.0 (17.4)          | 34.7                                    | 0.8                   | −144.5 (16.5)                 | 33.6                                    | −14.6 (1.4)                           | 5.5                                     |
|        | G <sub>3</sub> <sup>S</sup> | absent                        | −81.5 (7.1)            | 49.4                                    | 1.3                   | −86.7 (4.9)                   | 37.8                                    | 5.2 (2.3)                             | 13.9                                    |
|        |                             | G <sub>1</sub> <sup>B</sup>   | −56.0 (9.7)            | 49.3                                    | 0.7                   | −67.9 (7.0)                   | 37.3                                    | 12.0 (2.8)                            | 13.9                                    |
|        |                             | G <sub>1</sub> <sup>S</sup>   | −83.5 (10.4)           | 69.8                                    | 1.7                   | −79.6 (4.2)                   | 43.7                                    | −3.8 (6.2)                            | 30.8                                    |
|        |                             | G <sub>2</sub> <sup>B</sup>   | −63.6 (12.5)           | 50.3                                    | 0.9                   | −72.9 (9.2)                   | 38.0                                    | 9.3 (3.3)                             | 14.4                                    |
|        |                             | G <sub>2</sub> <sup>S</sup>   | −115.3 (15.2)          | 37.4                                    | 1.1                   | −112.2 (15.9)                 | 38.1                                    | −3.1 (0.8)                            | 6.0                                     |

<sup>a</sup>Absent indicates monoadducted DNA. <sup>b</sup>Average of the sum of stacking and hydrogen-bonding interactions for the adducted base, bulky moiety and opposing C, with standard deviation calculated between the replicas. <sup>c</sup>Standard deviation calculated over the 1.5  $\mu$ s dataset from the three replicas. <sup>d</sup>Block standard error (BSE) extrapolated from all frames of the three replicas fitted to a monomolecular growth model. <sup>e</sup>Average of the stacking between the adduct (base and bulky moiety) plus the opposing C and the flanking nucleobases, with standard deviation calculated between the replicas. <sup>f</sup>Average of the hydrogen-bonding interaction between the adduct and the pairing C, with standard deviation calculated between the replicas.

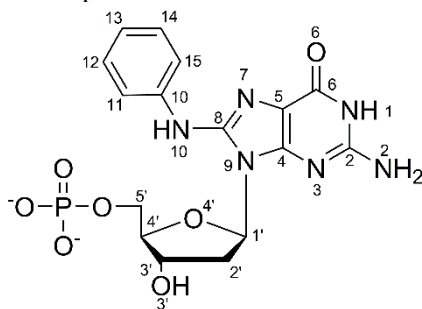
Table E.26: Average  $\Delta(\text{SASA}^{\text{avg}})$  and  $\Delta(\text{LSIE}^{\text{avg}})$  upon the addition of a second lesion.

| Primary Lesion              | Secondary Lesion <sup>a</sup> | <sup>AN</sup> dG                                                   |                                                           | <sup>AP</sup> dG                                                   |                                                           |
|-----------------------------|-------------------------------|--------------------------------------------------------------------|-----------------------------------------------------------|--------------------------------------------------------------------|-----------------------------------------------------------|
|                             |                               | $\Delta(\text{SASA}^{\text{avg}})^{\text{b}}$<br>(Å <sup>2</sup> ) | $\Delta(\text{LSIE}^{\text{avg}})^{\text{c}}$<br>(kJ/mol) | $\Delta(\text{SASA}^{\text{avg}})^{\text{b}}$<br>(Å <sup>2</sup> ) | $\Delta(\text{LSIE}^{\text{avg}})^{\text{c}}$<br>(kJ/mol) |
| G <sub>1</sub> <sup>B</sup> | absent                        | 0.0                                                                | 0.0                                                       | 0.0                                                                | 0.0                                                       |
|                             | G <sub>2</sub> <sup>B</sup>   | −15.1                                                              | −9.5                                                      | −22.8                                                              | 5.7                                                       |
|                             | G <sub>2</sub> <sup>S</sup>   | −2.8                                                               | 7.8                                                       | −42.3                                                              | −25.8                                                     |
|                             | G <sub>3</sub> <sup>B</sup>   | −0.1                                                               | −0.5                                                      | 0.6                                                                | 0.4                                                       |
|                             | G <sub>3</sub> <sup>S</sup>   | 0.2                                                                | 1.7                                                       | −2.7                                                               | 0.9                                                       |
| G <sub>2</sub> <sup>B</sup> | absent                        | 0.0                                                                | 0.0                                                       | 0.0                                                                | 0.0                                                       |
|                             | G <sub>1</sub> <sup>B</sup>   | −22.1                                                              | −8.9                                                      | −31.1                                                              | 1.6                                                       |
|                             | G <sub>1</sub> <sup>S</sup>   | −12.2                                                              | 6.2                                                       | −18.3                                                              | −1.2                                                      |
|                             | G <sub>3</sub> <sup>B</sup>   | 1.5                                                                | 1.0                                                       | −7.7                                                               | 3.1                                                       |
|                             | G <sub>3</sub> <sup>S</sup>   | −1.5                                                               | −3.5                                                      | −0.6                                                               | −1.8                                                      |
| G <sub>3</sub> <sup>B</sup> | absent                        | 0.0                                                                | 0.0                                                       | 0.0                                                                | 0.0                                                       |
|                             | G <sub>1</sub> <sup>B</sup>   | −0.1                                                               | 1.2                                                       | 1.2                                                                | 1.9                                                       |
|                             | G <sub>1</sub> <sup>S</sup>   | −0.7                                                               | 2.3                                                       | 1.5                                                                | 2.9                                                       |
|                             | G <sub>2</sub> <sup>B</sup>   | −0.2                                                               | 1.3                                                       | −12.5                                                              | 4.1                                                       |
|                             | G <sub>2</sub> <sup>S</sup>   | 0.4                                                                | −2.0                                                      | −4.5                                                               | 2.1                                                       |
| G <sub>1</sub> <sup>S</sup> | absent                        | 0.0                                                                | 0.0                                                       | 0.0                                                                | 0.0                                                       |
|                             | G <sub>2</sub> <sup>B</sup>   | −6.2                                                               | −26.8                                                     | 2.9                                                                | 38.1                                                      |
|                             | G <sub>2</sub> <sup>S</sup>   | 1.4                                                                | 4.0                                                       | −29.1                                                              | 29.4                                                      |
|                             | G <sub>3</sub> <sup>B</sup>   | −1.6                                                               | 49.6                                                      | −3.9                                                               | 45.7                                                      |
|                             | G <sub>3</sub> <sup>S</sup>   | −8.1                                                               | 44.0                                                      | 1.8                                                                | 3.5                                                       |
| G <sub>2</sub> <sup>S</sup> | absent                        | 0.0                                                                | 0.0                                                       | 0.0                                                                | 0.0                                                       |
|                             | G <sub>1</sub> <sup>B</sup>   | 12.6                                                               | −34.0                                                     | −1.0                                                               | −37.1                                                     |
|                             | G <sub>1</sub> <sup>S</sup>   | 17.5                                                               | 6.1                                                       | −15.5                                                              | −11.8                                                     |
|                             | G <sub>3</sub> <sup>B</sup>   | −16.1                                                              | 36.1                                                      | −8.7                                                               | −7.2                                                      |
|                             | G <sub>3</sub> <sup>S</sup>   | −15.6                                                              | −123.1                                                    | −11.5                                                              | −57.8                                                     |
| G <sub>3</sub> <sup>S</sup> | absent                        | 0.0                                                                | 0.0                                                       | 0.0                                                                | 0.0                                                       |
|                             | G <sub>1</sub> <sup>B</sup>   | −7.9                                                               | 21.1                                                      | 4.0                                                                | 25.5                                                      |
|                             | G <sub>1</sub> <sup>S</sup>   | −9.9                                                               | 15.7                                                      | 5.7                                                                | −1.9                                                      |
|                             | G <sub>2</sub> <sup>B</sup>   | 0.5                                                                | 60.6                                                      | 2.6                                                                | 17.9                                                      |
|                             | G <sub>2</sub> <sup>S</sup>   | 26.0                                                               | −78.2                                                     | −2.5                                                               | −33.7                                                     |

<sup>a</sup>Absent indicates monoadducted DNA. <sup>b</sup> $\Delta(\text{SASA}^{\text{avg}})$  denotes the  $\text{SASA}^{\text{avg}}$  of a given adduct site in the presence of a second lesion (diadducted DNA) minus the  $\text{SASA}^{\text{avg}}$  of the same adduct site in the absence of a second lesion (monoadducted DNA). <sup>c</sup> $\Delta(\text{LSIE}^{\text{avg}})$  denotes the  $\text{LSIE}^{\text{avg}}$  of a given adduct site in the presence of a second lesion (diadducted DNA) minus the  $\text{LSIE}^{\text{avg}}$  of the same adduct site in the absence of a second lesion (monoadducted DNA).



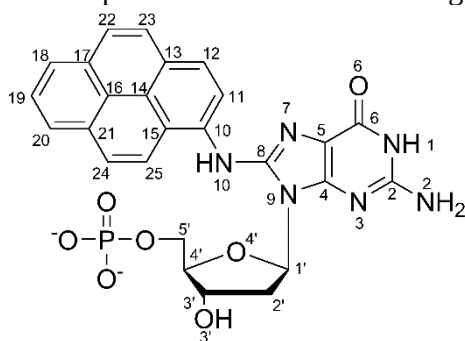
Table E.27: MOL2 file containing partial charges and atom types of the <sup>AN</sup>dG adduct used for MD simulations (atom numbers are provided in the associated figure).



|    |      |        |         |         |    |   |     |         |
|----|------|--------|---------|---------|----|---|-----|---------|
| 1  | P    | 4.7050 | 0.0520  | -1.1680 | P  | 1 | LIG | 1.2090  |
| 2  | O5'  | 4.1730 | 1.0760  | 0.0000  | OS | 1 | LIG | -0.4900 |
| 3  | OP1  | 4.6090 | 0.7320  | -2.4680 | O2 | 1 | LIG | -0.7864 |
| 4  | OP2  | 5.9220 | -0.5810 | -0.6380 | O2 | 1 | LIG | -0.7953 |
| 5  | O3   | 0.0000 | 0.0000  | 0.0000  | OS | 1 | LIG | -0.5474 |
| 6  | C5'  | 3.1010 | 1.9490  | -0.3030 | CI | 1 | LIG | -0.0102 |
| 7  | H5'  | 2.8330 | 2.4680  | 0.6250  | H1 | 1 | LIG | 0.0752  |
| 8  | H5'' | 3.3910 | 2.7140  | -1.0420 | H1 | 1 | LIG | 0.0752  |
| 9  | C4'  | 1.8690 | 1.2180  | -0.8380 | CT | 1 | LIG | 0.1830  |
| 10 | H4'  | 1.0270 | 1.9170  | -0.9140 | H1 | 1 | LIG | 0.0978  |
| 11 | O4'  | 2.1700 | 0.7140  | -2.1590 | OS | 1 | LIG | -0.4082 |
| 12 | C1'  | 1.9220 | -0.6850 | -2.2280 | CT | 1 | LIG | 0.1596  |
| 13 | H1'  | 0.9300 | -0.8780 | -2.6450 | H2 | 1 | LIG | 0.0902  |
| 14 | C3'  | 1.4280 | 0.0000  | 0.0000  | CT | 1 | LIG | 0.1484  |
| 15 | H3'  | 1.8240 | 0.0530  | 1.0220  | H1 | 1 | LIG | 0.0720  |
| 16 | C2'  | 2.0070 | -1.1870 | -0.7800 | CT | 1 | LIG | -0.0617 |
| 17 | H2'  | 3.0480 | -1.3520 | -0.4920 | HC | 1 | LIG | 0.0466  |
| 18 | H2'' | 1.4440 | -2.1140 | -0.6290 | HC | 1 | LIG | 0.0466  |
| 19 | N9   | 2.8790 | -1.2770 | -3.1530 | N* | 1 | LIG | -0.0563 |
| 20 | C8   | 2.6100 | -2.0090 | -4.3140 | CR | 1 | LIG | 0.4678  |
| 21 | C4   | 4.2540 | -1.2720 | -3.0400 | CB | 1 | LIG | 0.1484  |
| 22 | N7   | 3.6880 | -2.4790 | -4.8900 | NB | 1 | LIG | -0.5254 |
| 23 | C5   | 4.7300 | -2.0310 | -4.1010 | CB | 1 | LIG | 0.0762  |
| 24 | N4   | 4.9730 | -0.6260 | -2.0850 | NC | 1 | LIG | -0.5421 |
| 25 | C6   | 6.1500 | -2.2180 | -4.2500 | C  | 1 | LIG | 0.5106  |
| 26 | C2   | 6.2750 | -0.7780 | -2.1980 | CA | 1 | LIG | 0.6729  |
| 27 | O6   | 6.7970 | -2.8430 | -5.0700 | O  | 1 | LIG | -0.5407 |
| 28 | N1   | 6.8450 | -1.5130 | -3.1980 | NA | 1 | LIG | -0.4810 |
| 29 | H1   | 7.8500 | -1.6540 | -3.2180 | H  | 1 | LIG | 0.3501  |
| 30 | N2   | 7.0960 | -0.2220 | -1.2480 | N2 | 1 | LIG | -0.8752 |
| 31 | H21  | 8.0220 | 0.0670  | -1.5350 | H  | 1 | LIG | 0.3884  |

|    |     |         |         |         |    |   |     |         |
|----|-----|---------|---------|---------|----|---|-----|---------|
| 32 | H22 | 6.6240  | 0.4340  | -0.6340 | H  | 1 | LIG | 0.3884  |
| 33 | N10 | 1.3100  | -2.3020 | -4.7270 | N2 | 1 | LIG | -0.5894 |
| 34 | H10 | 1.3080  | -3.1770 | -5.2390 | H  | 1 | LIG | 0.3511  |
| 35 | C10 | 0.4170  | -1.3450 | -5.2830 | CA | 1 | LIG | 0.2156  |
| 36 | C11 | 0.6440  | 0.0360  | -5.2250 | CA | 1 | LIG | -0.0547 |
| 37 | H11 | 1.5430  | 0.4310  | -4.7640 | HA | 1 | LIG | 0.1212  |
| 38 | C15 | -0.7540 | -1.8230 | -5.8900 | CA | 1 | LIG | -0.2053 |
| 39 | H15 | -0.9360 | -2.8950 | -5.9310 | HA | 1 | LIG | 0.1458  |
| 40 | C12 | -0.2970 | 0.9160  | -5.7630 | CA | 1 | LIG | -0.2343 |
| 41 | H12 | -0.1060 | 1.9850  | -5.7080 | HA | 1 | LIG | 0.1637  |
| 42 | C14 | -1.6780 | -0.9360 | -6.4340 | CA | 1 | LIG | -0.1747 |
| 43 | H14 | -2.5760 | -1.3250 | -6.9060 | HA | 1 | LIG | 0.1548  |
| 44 | C13 | -1.4580 | 0.4430  | -6.3720 | CA | 1 | LIG | -0.1124 |
| 45 | H13 | -2.1820 | 1.1350  | -6.7910 | HA | 1 | LIG | 0.1321  |

Table E.28: MOL2 file containing partial charges and atom types of the <sup>AP</sup>dG adduct used for MD simulations (atom numbers are provided in the associated figure).



|    |      |        |         |         |    |   |     |         |
|----|------|--------|---------|---------|----|---|-----|---------|
| 1  | P    | 4.9760 | 0.3860  | -1.3260 | P  | 1 | LIG | 1.1662  |
| 2  | O5'  | 4.1390 | 0.9990  | 0.0000  | OS | 1 | LIG | -0.4956 |
| 3  | OP1  | 5.1120 | 1.4640  | -2.3580 | O2 | 1 | LIG | -0.7681 |
| 4  | OP2  | 6.1010 | -0.4220 | -0.7550 | O2 | 1 | LIG | -0.7681 |
| 5  | O3   | 0.0000 | 0.0000  | 0.0000  | OS | 1 | LIG | -0.5424 |
| 6  | C5'  | 3.1040 | 1.8950  | -0.2820 | CI | 1 | LIG | -0.0196 |
| 7  | H5'  | 3.4230 | 2.6620  | -0.9830 | H1 | 1 | LIG | 0.0821  |
| 8  | H5'' | 2.8550 | 2.3890  | 0.6500  | H1 | 1 | LIG | 0.0821  |
| 9  | C4'  | 1.8570 | 1.2140  | -0.8200 | CT | 1 | LIG | 0.1633  |
| 10 | H4'  | 1.0480 | 1.9310  | -0.8780 | H1 | 1 | LIG | 0.1054  |
| 11 | O4'  | 2.1080 | 0.7240  | -2.1370 | OS | 1 | LIG | -0.4032 |
| 12 | C1'  | 1.8650 | -0.6500 | -2.2260 | CT | 1 | LIG | 0.1513  |
| 13 | H1'  | 0.8780 | -0.8250 | -2.6240 | H2 | 1 | LIG | 0.0873  |
| 14 | C3'  | 1.4070 | 0.0000  | 0.0000  | CT | 1 | LIG | 0.1347  |
| 15 | H3'  | 1.7910 | 0.0400  | 1.0120  | H1 | 1 | LIG | 0.0784  |
| 16 | C2'  | 1.9750 | -1.1690 | -0.7950 | CT | 1 | LIG | -0.0572 |

|    |      |         |         |         |    |   |     |         |
|----|------|---------|---------|---------|----|---|-----|---------|
| 17 | H2'  | 3.0130  | -1.3300 | -0.5360 | HC | 1 | LIG | 0.0443  |
| 18 | H2'' | 1.4270  | -2.0920 | -0.6530 | HC | 1 | LIG | 0.0443  |
| 19 | N9   | 2.7970  | -1.2560 | -3.1630 | N* | 1 | LIG | -0.0519 |
| 20 | C8   | 4.0990  | -0.9300 | -3.5110 | CR | 1 | LIG | 0.3777  |
| 21 | C4   | 2.5500  | -2.4460 | -3.7900 | CB | 1 | LIG | 0.1445  |
| 22 | N7   | 4.6490  | -1.7930 | -4.2750 | NB | 1 | LIG | -0.4566 |
| 23 | C5   | 3.6890  | -2.7690 | -4.4570 | CB | 1 | LIG | 0.0570  |
| 24 | N3   | 1.3850  | -3.1370 | -3.7080 | NC | 1 | LIG | -0.4990 |
| 25 | C6   | 3.7200  | -3.9790 | -5.2240 | C  | 1 | LIG | 0.4977  |
| 26 | C2   | 1.3980  | -4.2210 | -4.4050 | CA | 1 | LIG | 0.5921  |
| 27 | O6   | 4.5820  | -4.4690 | -5.8910 | O  | 1 | LIG | -0.5456 |
| 28 | N1   | 2.4720  | -4.6420 | -5.1190 | NA | 1 | LIG | -0.4138 |
| 29 | H1   | 2.3970  | -5.4580 | -5.6890 | H  | 1 | LIG | 0.3351  |
| 30 | N2   | 0.2780  | -4.9950 | -4.4810 | N2 | 1 | LIG | -0.8656 |
| 31 | H21  | 0.3970  | -5.9770 | -4.5990 | H  | 1 | LIG | 0.3895  |
| 32 | H22  | -0.4320 | -4.7410 | -3.8300 | H  | 1 | LIG | 0.3895  |
| 33 | N10  | 4.6780  | 0.2390  | -3.0360 | N2 | 1 | LIG | -0.3628 |
| 34 | H10  | 4.0050  | 0.9680  | -2.9720 | H  | 1 | LIG | 0.2783  |
| 35 | C10  | 5.9550  | 0.6490  | -3.5100 | CA | 1 | LIG | 0.0814  |
| 36 | C15  | 6.1060  | 1.8360  | -4.2310 | CA | 1 | LIG | 0.0633  |
| 37 | C11  | 7.0640  | -0.1160 | -3.1790 | CA | 1 | LIG | -0.2616 |
| 38 | H11  | 6.9210  | -1.0300 | -2.6350 | HA | 1 | LIG | 0.1751  |
| 39 | C14  | 7.4130  | 2.2570  | -4.5810 | CA | 1 | LIG | 0.0181  |
| 40 | C23  | 4.9910  | 2.6640  | -4.6450 | CA | 1 | LIG | -0.1706 |
| 41 | H23  | 3.9870  | 2.3450  | -4.4390 | HA | 1 | LIG | 0.1296  |
| 42 | C12  | 8.3300  | 0.2810  | -3.5480 | CA | 1 | LIG | -0.1240 |
| 43 | H12  | 9.1750  | -0.3300 | -3.2860 | HA | 1 | LIG | 0.1527  |
| 44 | C16  | 7.6070  | 3.4880  | -5.2930 | CA | 1 | LIG | 0.0859  |
| 45 | C13  | 8.5310  | 1.4720  | -4.2390 | CA | 1 | LIG | 0.0232  |
| 46 | C22  | 5.1770  | 3.8110  | -5.3090 | CA | 1 | LIG | -0.2254 |
| 47 | H22' | 4.3310  | 4.4020  | -5.6120 | HA | 1 | LIG | 0.1573  |
| 48 | C21  | 6.4980  | 4.2810  | -5.6550 | CA | 1 | LIG | 0.0462  |
| 49 | C17  | 8.9100  | 3.9110  | -5.6360 | CA | 1 | LIG | 0.0488  |
| 50 | C24  | 9.8500  | 1.9270  | -4.6100 | CA | 1 | LIG | -0.1744 |
| 51 | H24  | 10.6920 | 1.3130  | -4.3430 | HA | 1 | LIG | 0.1419  |
| 52 | C20  | 6.7030  | 5.4740  | -6.3400 | CA | 1 | LIG | -0.1411 |
| 53 | H20  | 5.8560  | 6.0770  | -6.6170 | HA | 1 | LIG | 0.1383  |
| 54 | C18  | 9.0760  | 5.1110  | -6.3220 | CA | 1 | LIG | -0.1473 |
| 55 | H18  | 10.0680 | 5.4350  | -6.5840 | HA | 1 | LIG | 0.1409  |
| 56 | C25  | 10.0320 | 3.0790  | -5.2670 | CA | 1 | LIG | -0.1962 |
| 57 | H25  | 11.0200 | 3.4070  | -5.5350 | HA | 1 | LIG | 0.1453  |

|    |     |        |        |         |    |   |     |         |
|----|-----|--------|--------|---------|----|---|-----|---------|
| 58 | C19 | 7.9830 | 5.8840 | -6.6680 | CA | 1 | LIG | -0.2217 |
| 59 | H19 | 8.1290 | 6.8080 | -7.1990 | HA | 1 | LIG | 0.1629  |

## **Appendix F**

**Supplementary Information for Chapter 7: Effects of a Second Local DNA Damage Event on the Mutagenicity of the Human Carcinogen 4-Aminobiphenyl: A Molecular Dynamics Study of Damaged DNA Structure**

Contains Figures F.1–F.19 and Tables F.1–F.28

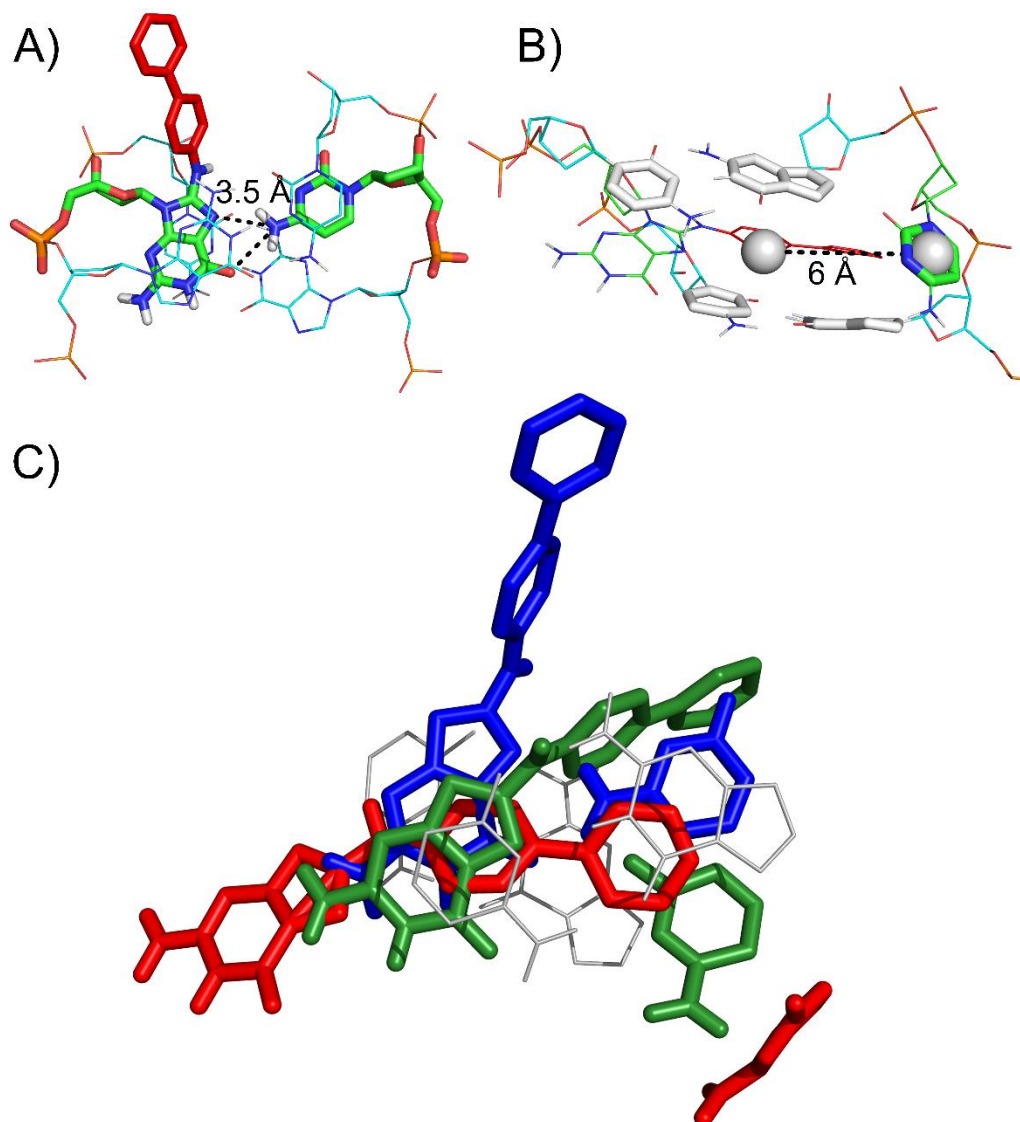


Figure F.1: Structural cluster criteria for of <sup>ABP</sup>dG adducted DNA containing the adduct in the *syn* orientation. A) The maximum distance criteria used to assign a structure to the conformation. B) The minimum distance criteria used to assign a structure to the S conformation. C) The relative locations of the adducted base, bulky group and pairing C with respect to the flanking nucleobases (grey) for DNA in the S (red), W (blue) and I (green) conformations.

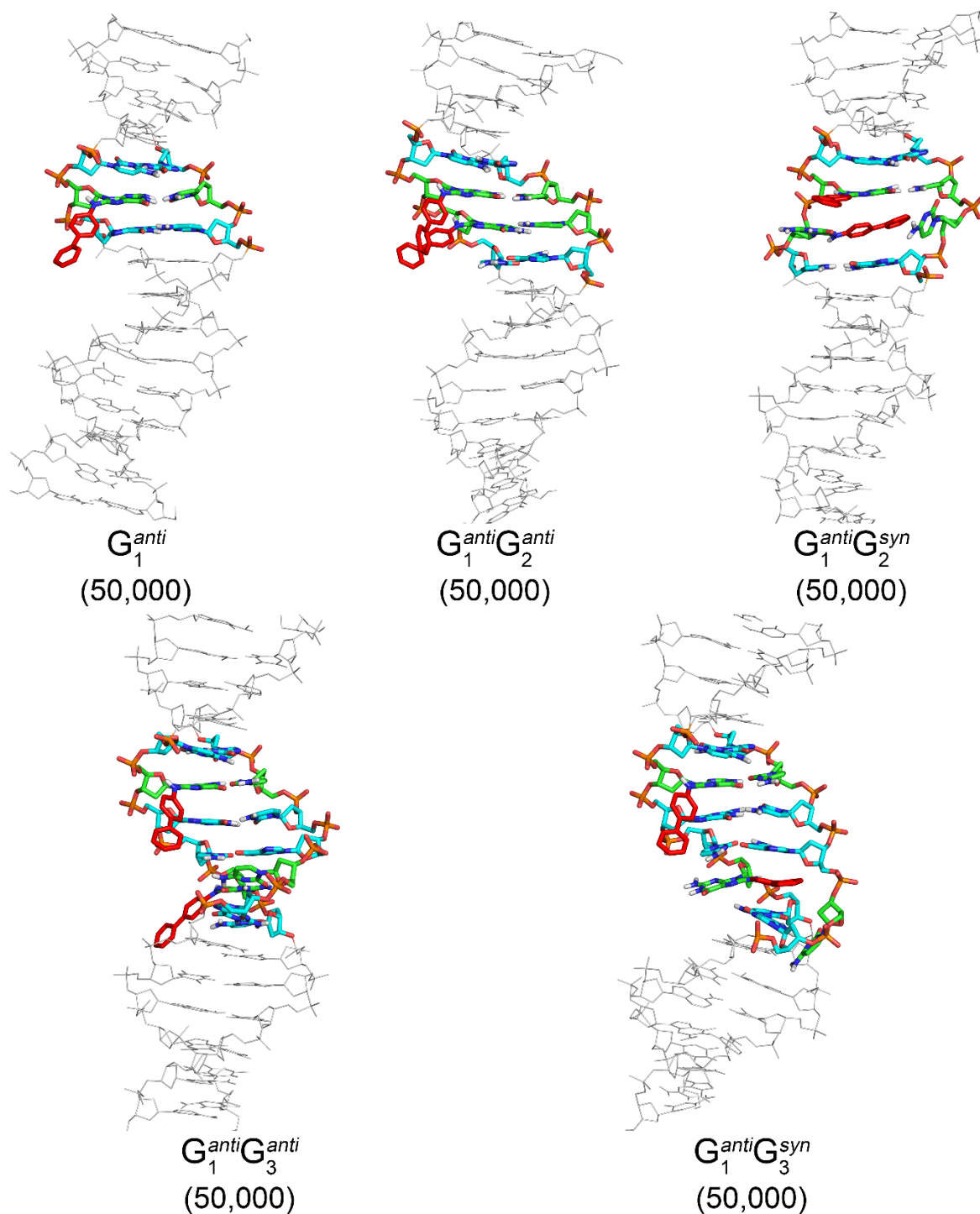


Figure F.2: MD representative structures for damaged DNA containing *anti*<sup>ABP</sup>dG at G<sub>1</sub> calculated across five 500 ns replicas by clustering with respect to the helical conformation. The values in parentheses indicates the number of frames populated for a given cluster among the 50,000 analyzed. The adducted base pair is shown in green, the bulky moiety in red, and the flanking base pairs in cyan.

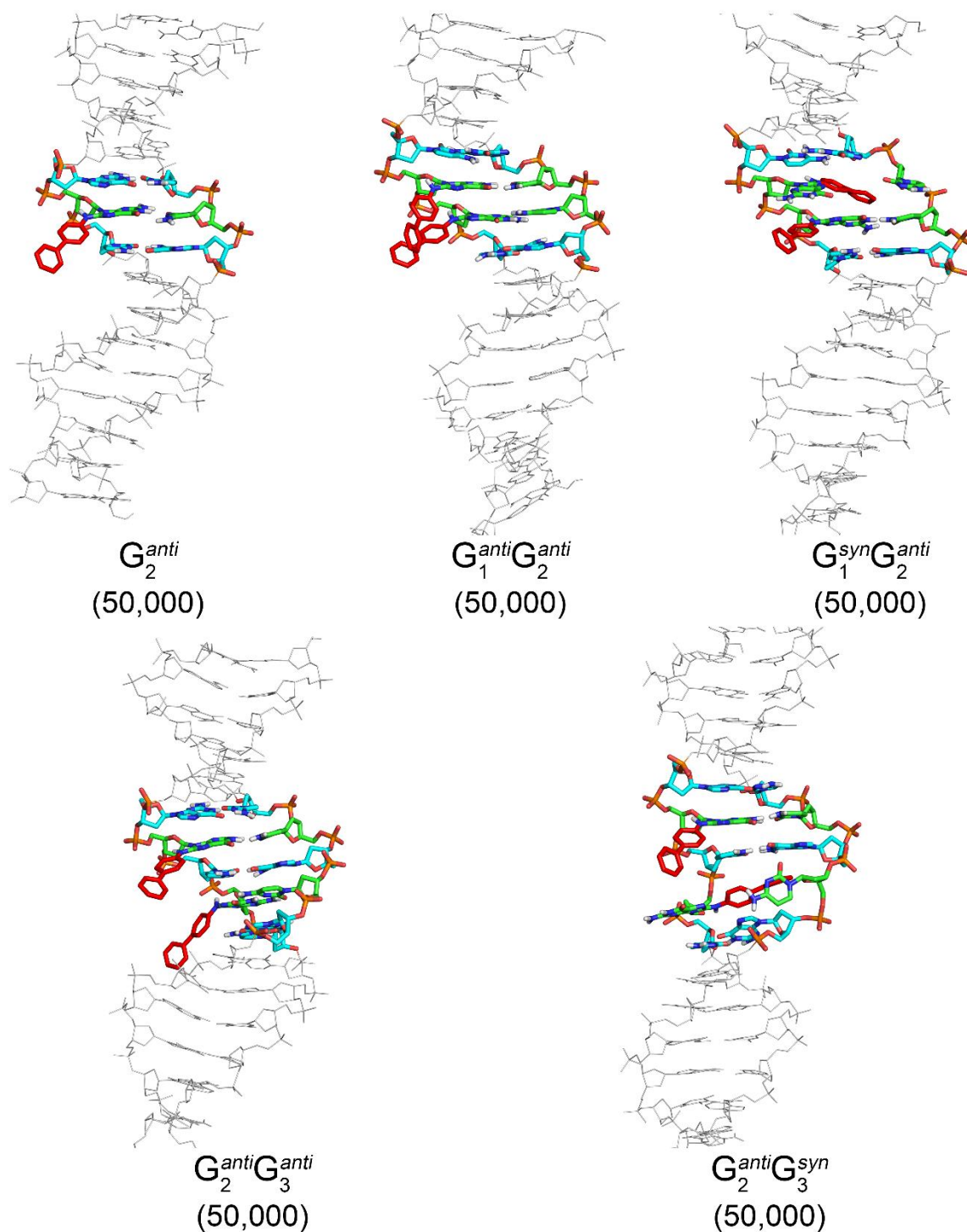


Figure F.3: MD representative structures for damaged DNA containing *anti*<sup>ABP</sup>dG at G<sub>2</sub> calculated across five 500 ns replicas by clustering with respect to the helical conformation. The values in parentheses indicates the number of frames populated for a given cluster among the 50,000 analyzed. The adducted base pair is shown in green, the bulky moiety in red, and the flanking base pairs in cyan.



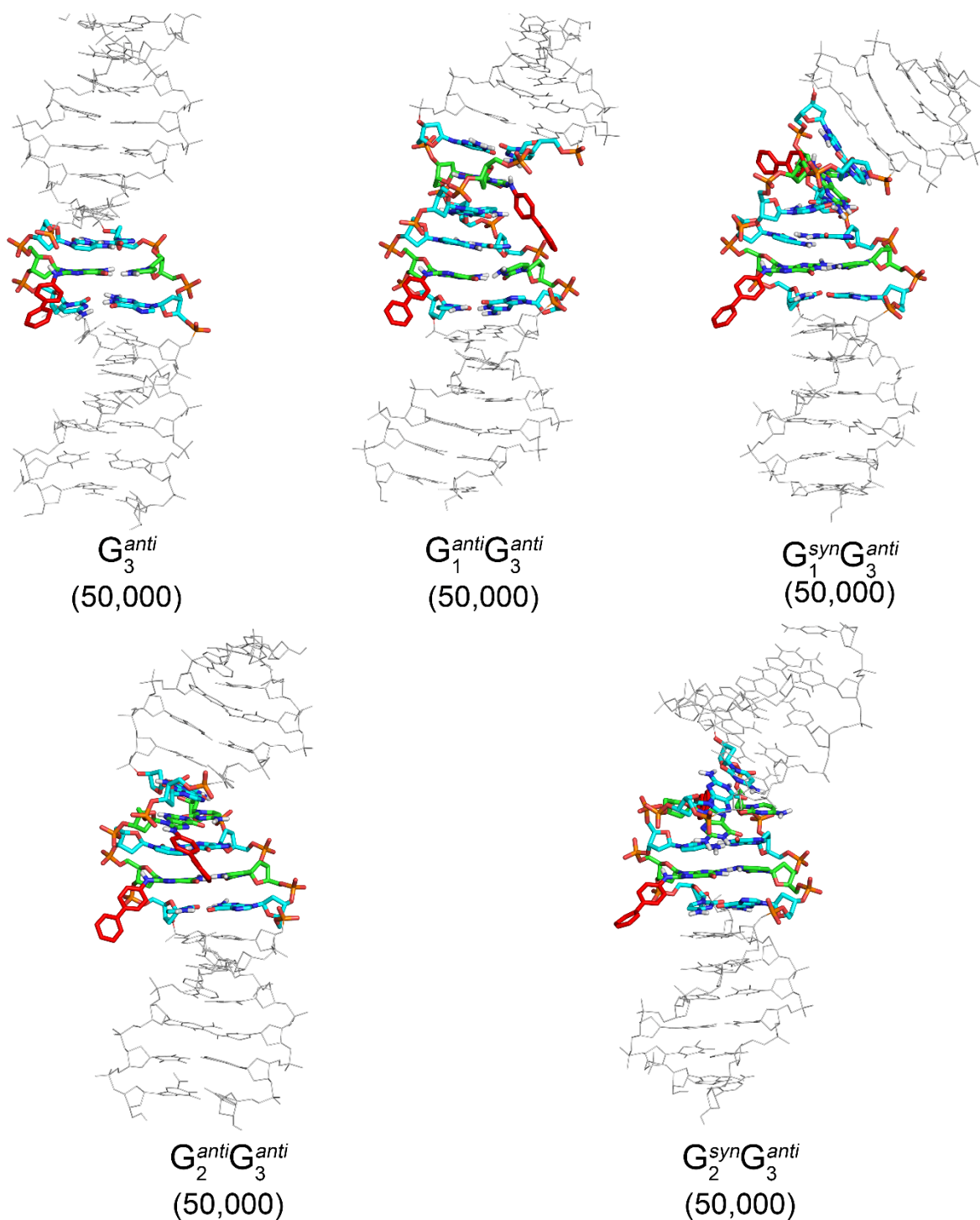


Figure F.4: MD representative structures for damaged DNA containing *anti*<sup>ABP</sup>dG at G<sub>3</sub> calculated across five 500 ns replicas by clustering with respect to the helical conformation. The values in parentheses indicates the number of frames populated for a given cluster among the 50,000 analyzed. The adducted base pair is shown in green, the bulky moiety in red, and the flanking base pairs in cyan.

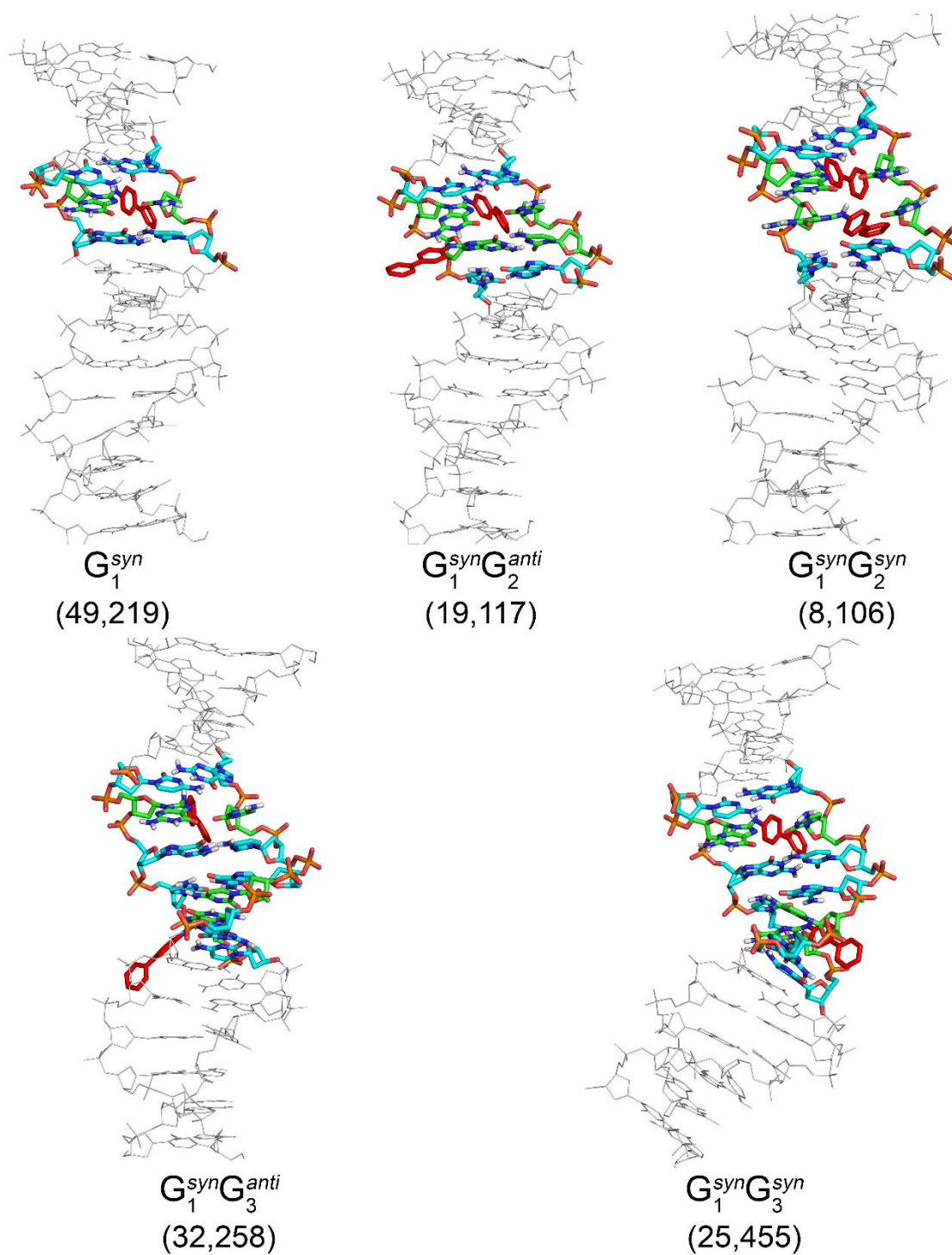


Figure F.5: MD representative structures for damaged DNA containing  $\text{syn}^{\text{ABP}}\text{dG}$  at  $G_1$  calculated across five 500 ns replicas by clustering with respect to the I helical conformation. The values in parentheses indicates the number of frames populated for a given cluster among the 50,000 analyzed. The adducted base pair is shown in green, the bulky moiety in red, and the flanking base pairs in cyan.

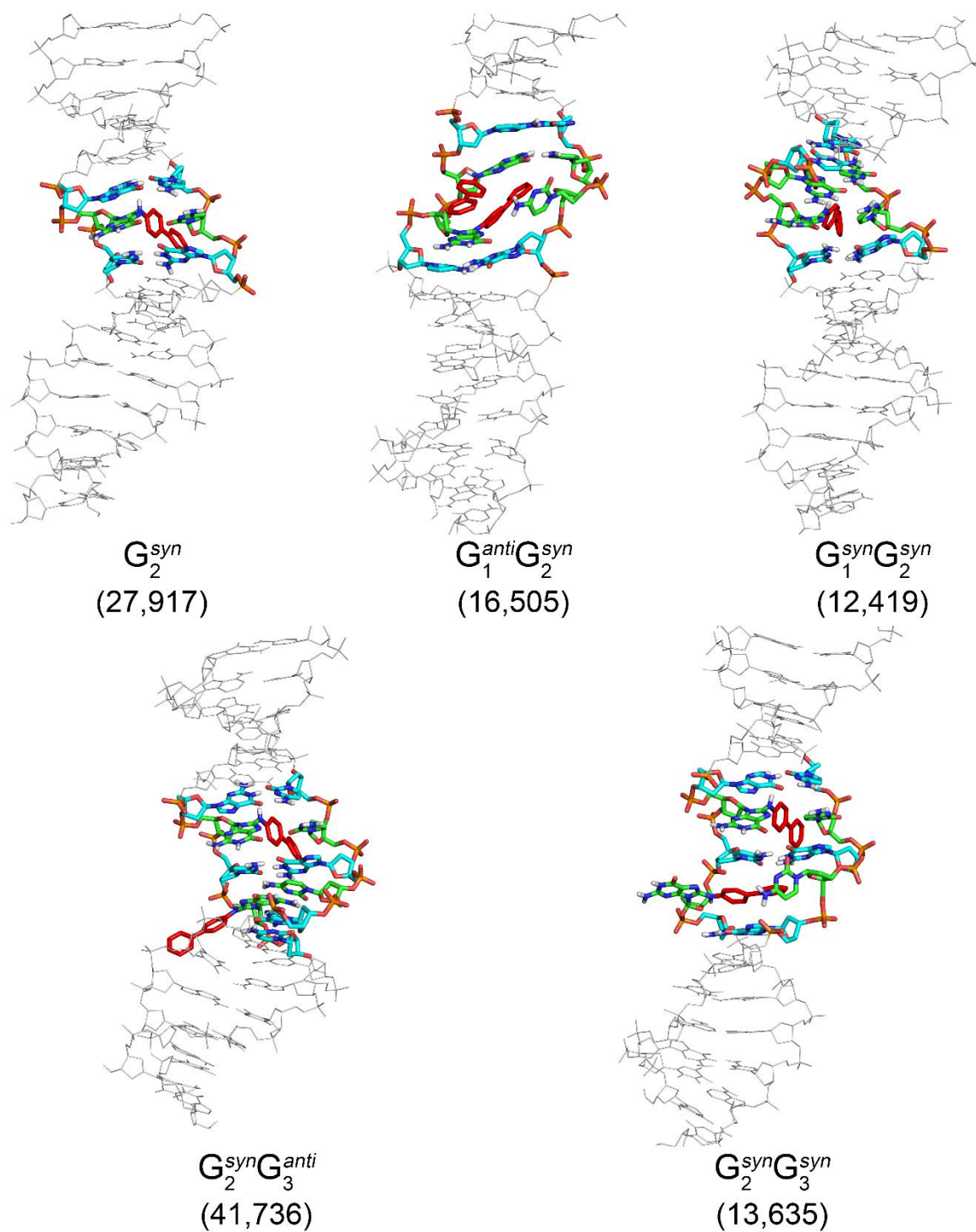


Figure F.6: MD representative structures for damaged DNA containing  $syn^{ABP}dG$  at  $G_2$  calculated across five 500 ns replicas by clustering with respect to the I helical conformation. The values in parentheses indicates the number of frames populated for a given cluster among the 50,000 analyzed. The adducted base pair is shown in green, the bulky moiety in red, and the flanking base pairs in cyan.

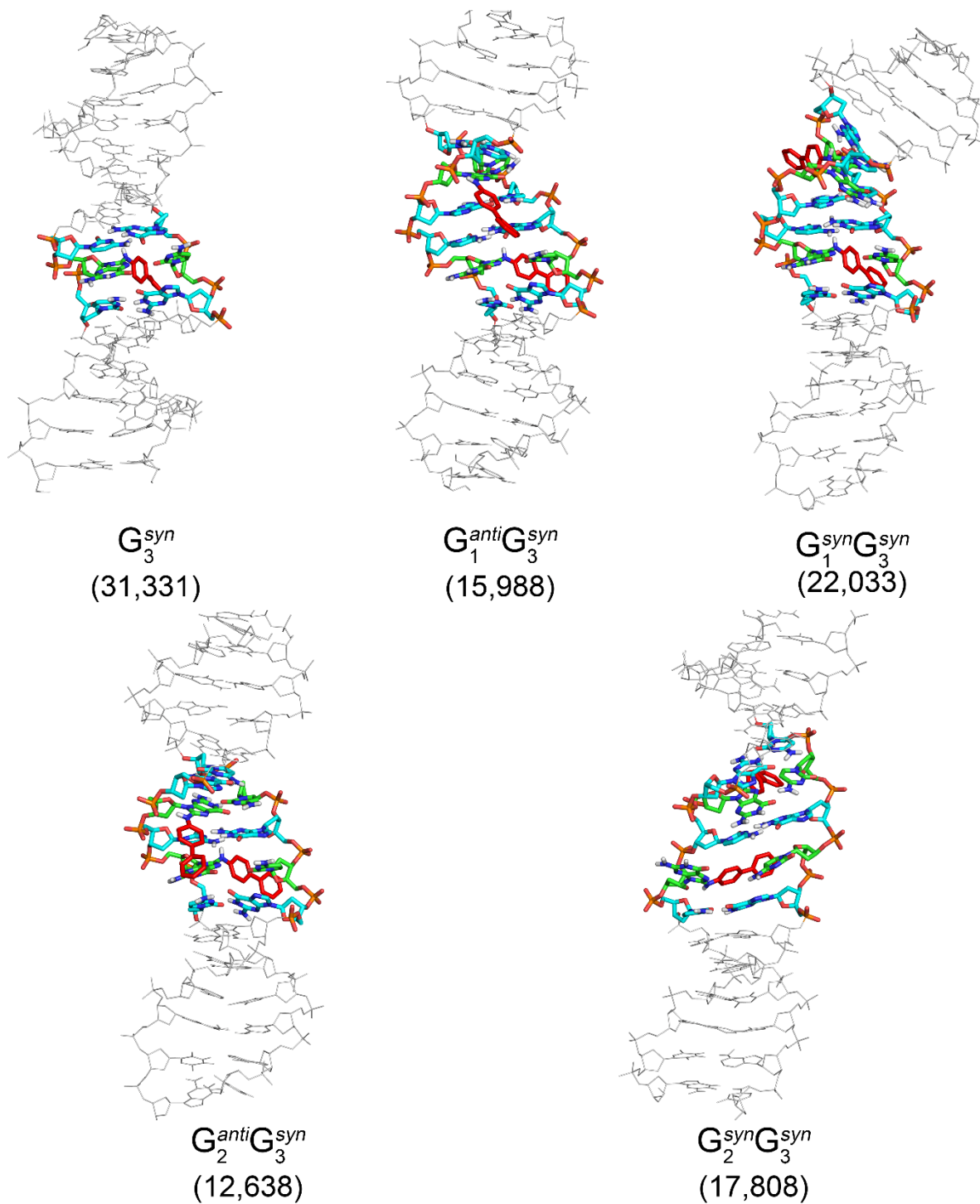


Figure F.7: MD representative structures for damaged DNA containing  $syn^{ABP}dG$  at  $G_3$  calculated across five 500 ns replicas by clustering with respect to the I helical conformation. The values in parentheses indicates the number of frames populated for a given cluster among the 50,000 analyzed. The adducted base pair is shown in green, the bulky moiety in red, and the flanking base pairs in cyan.

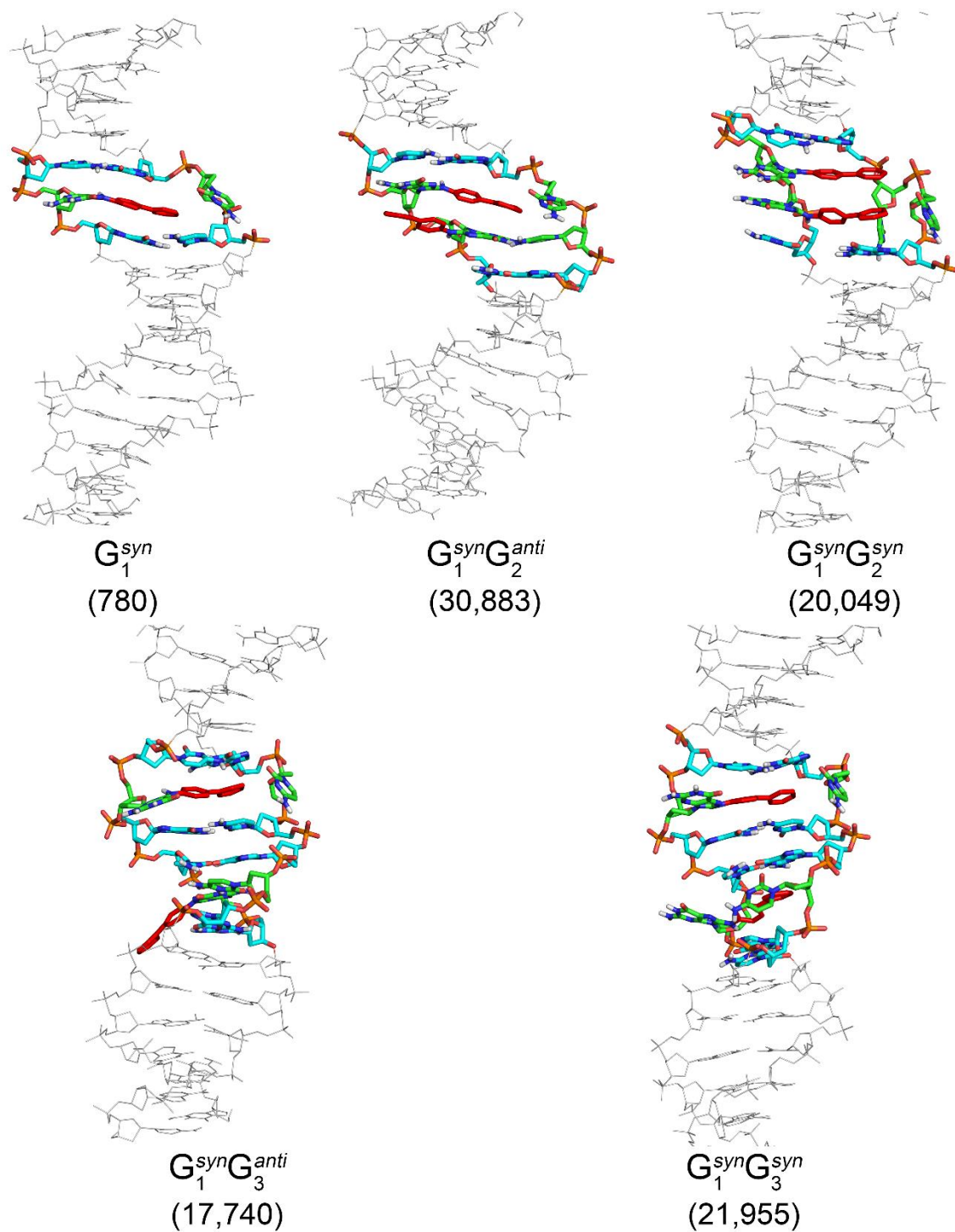


Figure F.8: MD representative structures for damaged DNA containing  $\text{syn}^{\text{ABP}}\text{dG}$  at  $G_1$  calculated across five 500 ns replicas by clustering with respect to the S helical conformation. The values in parentheses indicates the number of frames populated for a given cluster among the 50,000 analyzed. The adducted base pair is shown in green, the bulky moiety in red, and the flanking base pairs in cyan.



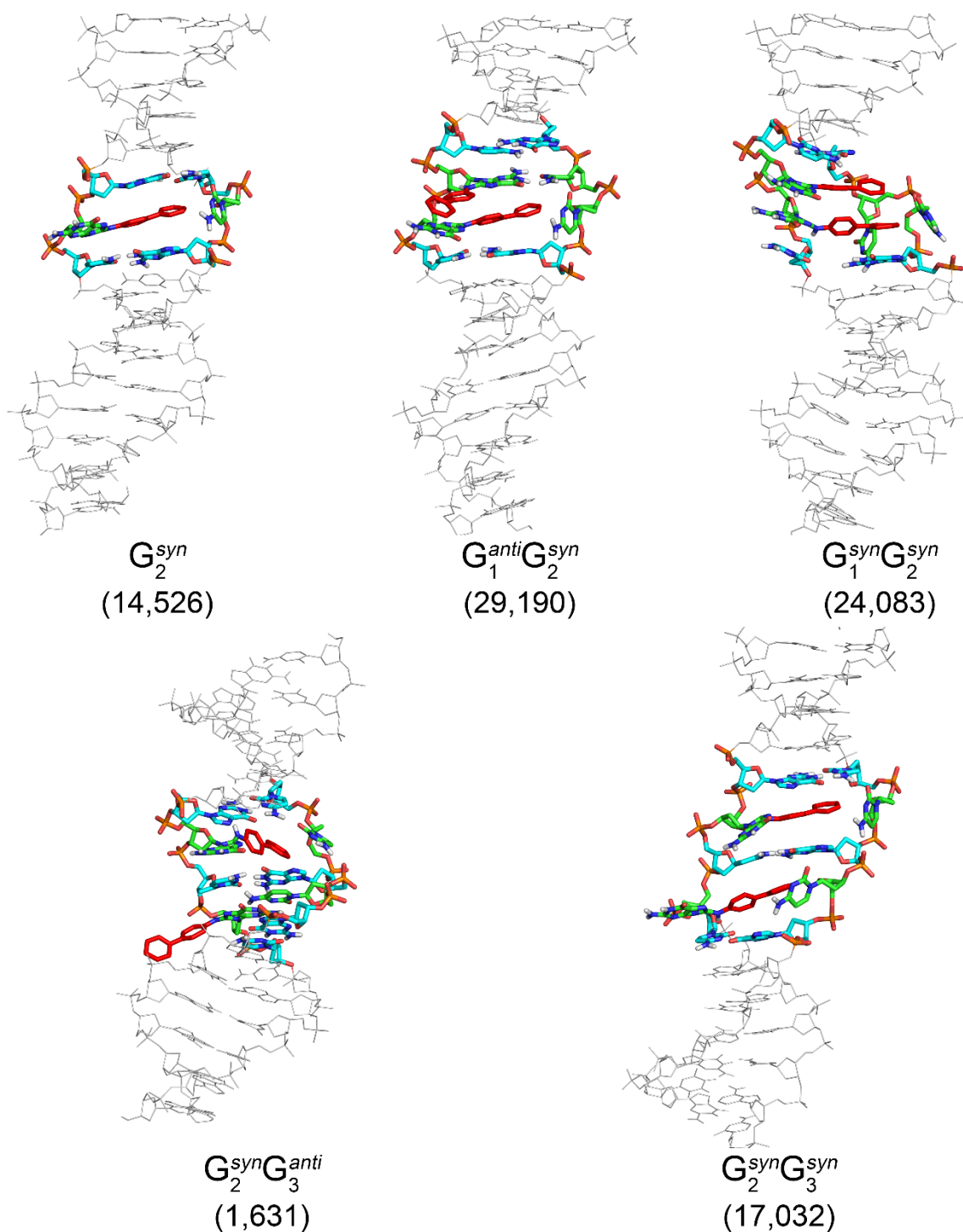


Figure F.9: MD representative structures for damaged DNA containing  $\text{syn}^{\text{ABP}}\text{dG}$  at  $\text{G}_2$  calculated across five 500 ns replicas by clustering with respect to the S helical conformation. The values in parentheses indicates the number of frames populated for a given cluster among the 50,000 analyzed. The adducted base pair is shown in green, the bulky moiety in red, and the flanking base pairs in cyan.

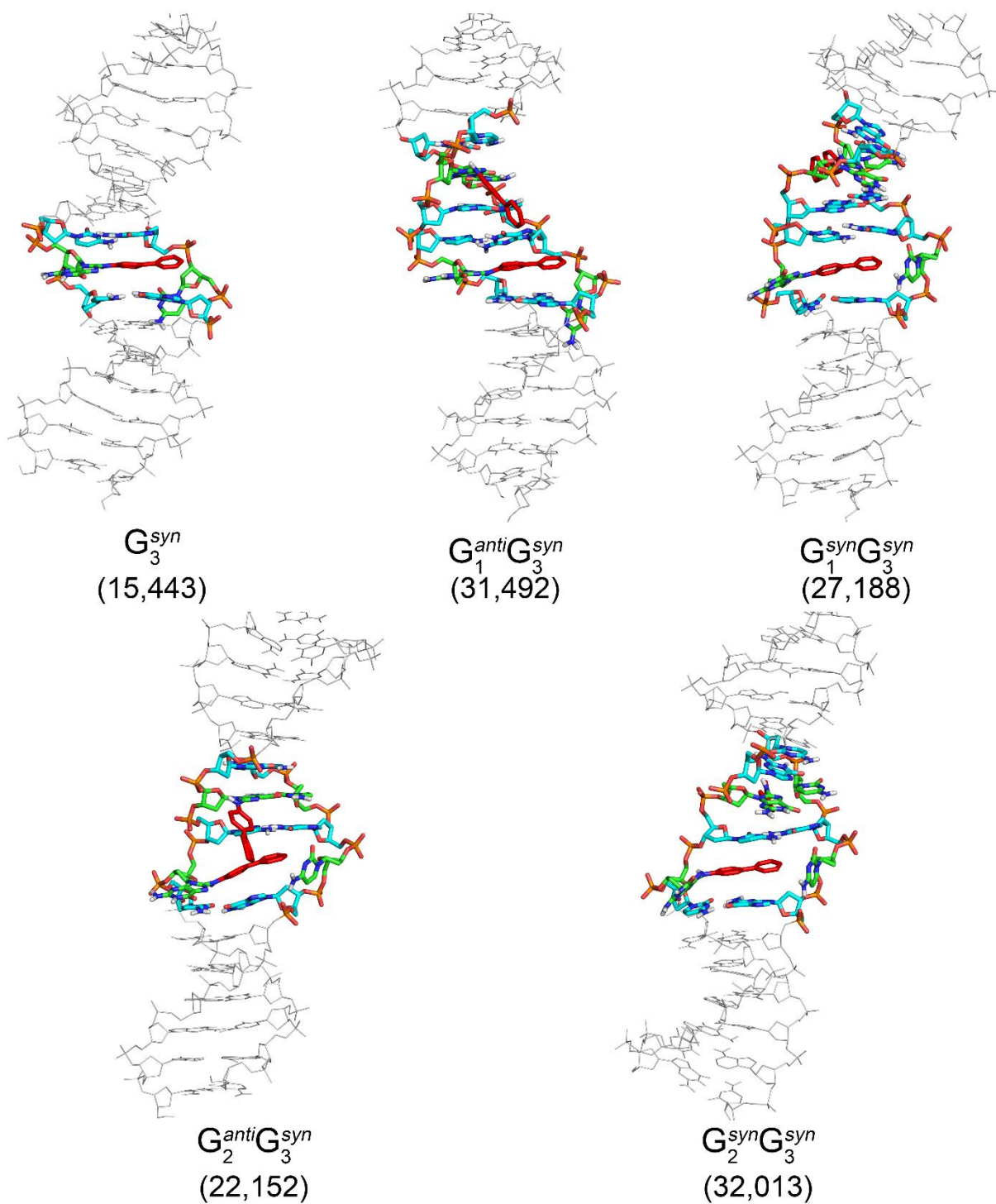


Figure F.10: MD representative structures for damaged DNA containing  $\text{syn}^{\text{ABP}}\text{dG}$  at  $\text{G}_3$  calculated across five 500 ns replicas by clustering with respect to the S helical conformation. The values in parentheses indicates the number of frames populated for a given cluster among the 50,000 analyzed. The adducted base pair is shown in green, the bulky moiety in red, and the flanking base pairs in cyan.

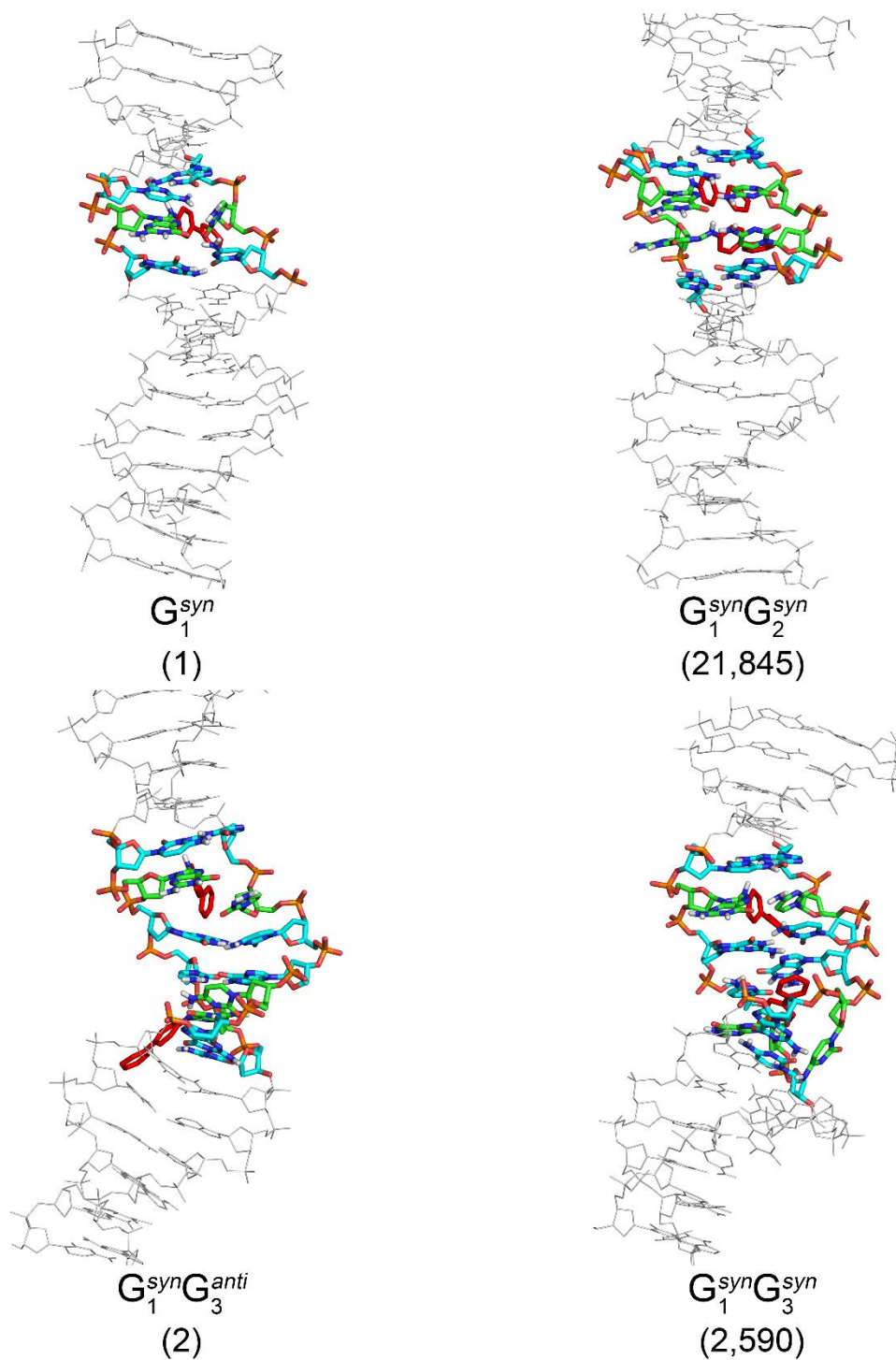


Figure F.11: MD representative structures for damaged DNA containing  $syn^{ABP}dG$  at  $G_1$  calculated across five 500 ns replicas by clustering with respect to the W helical conformation. The values in parentheses indicates the number of frames populated for a given cluster among the 50,000 analyzed. The adducted base pair is shown in green, the bulky moiety in red, and the flanking base pairs in cyan.



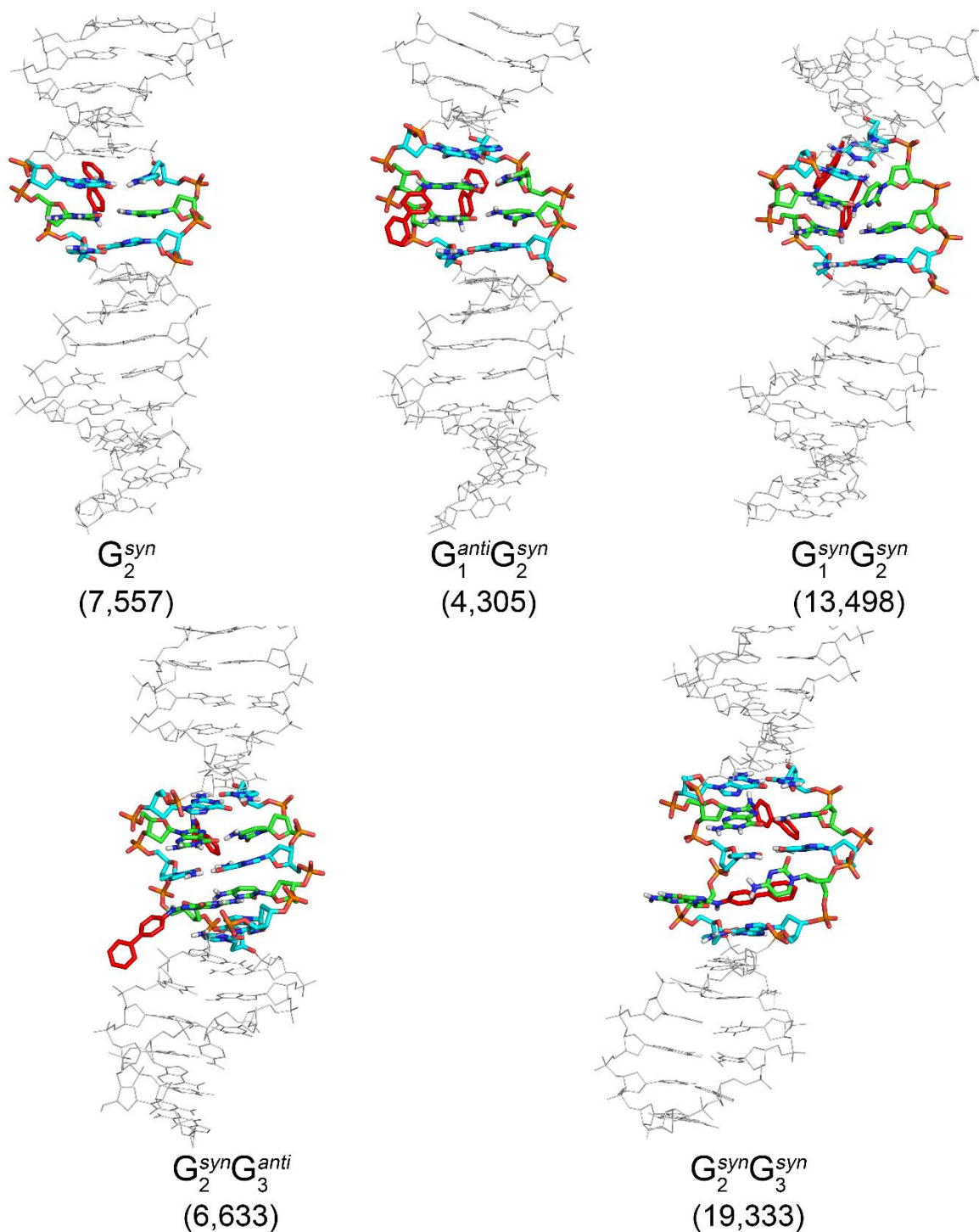


Figure F.12: MD representative structures for damaged DNA containing  $\text{syn}^{\text{ABP}}\text{dG}$  at  $\text{G}_2$  calculated across five 500 ns replicas by clustering with respect to the W helical conformation. The values in parentheses indicates the number of frames populated for a given cluster among the 50,000 analyzed. The adducted base pair is shown in green, the bulky moiety in red, and the flanking base pairs in cyan.

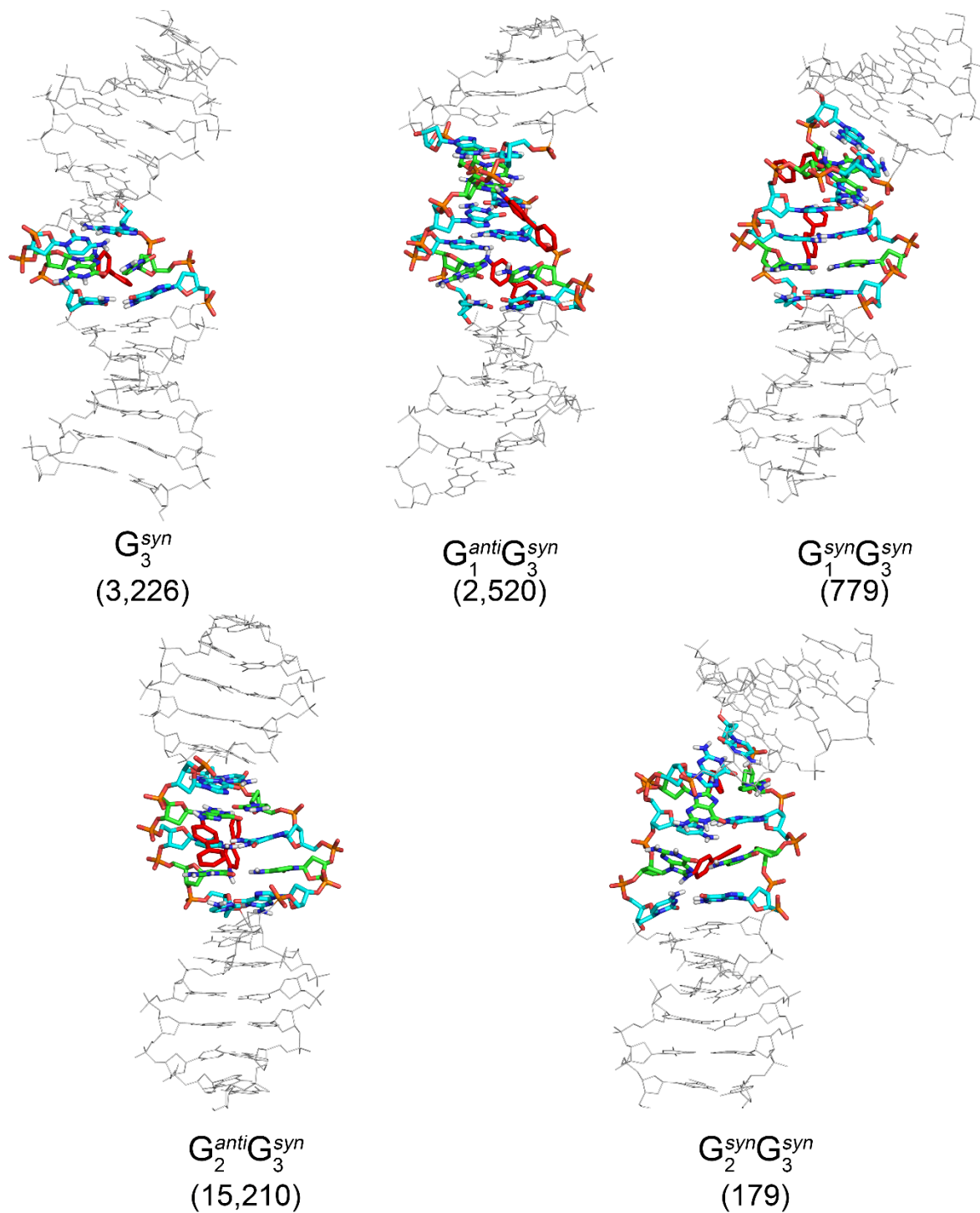


Figure F.13: MD representative structures for damaged DNA containing  $\text{syn}^{\text{ABP}}\text{dG}$  at  $\text{G}_3$  calculated across five 500 ns replicas by clustering with respect to the W helical conformation. The values in parentheses indicates the number of frames populated for a given cluster among the 50,000 analyzed. The adducted base pair is shown in green, the bulky moiety in red, and the flanking base pairs in cyan.

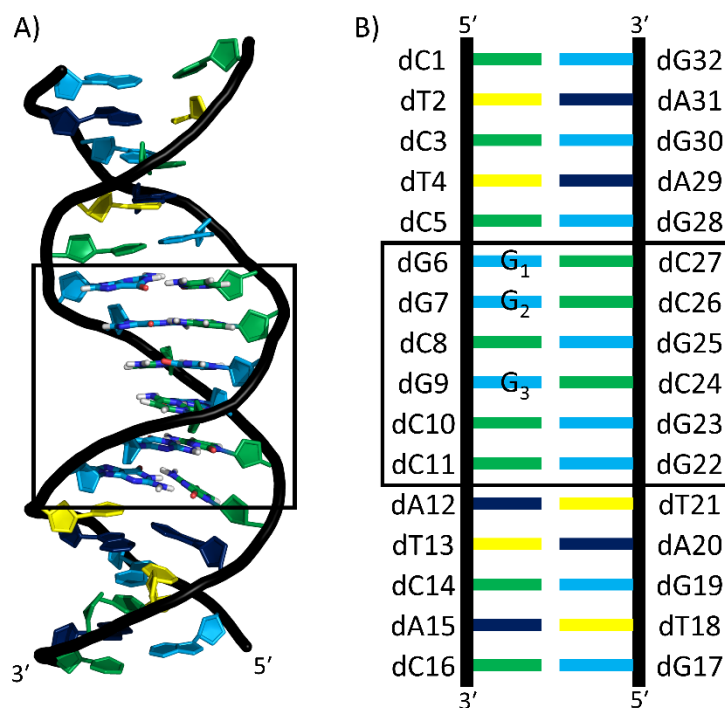


Figure F.14: A) Structure and B) nucleotide numbering of the (undamaged) *NarI* containing sequence investigated in the present work. Helical parameters were calculated with Curves+/Canal using a pseudostep between the base pairs on either side of the box.

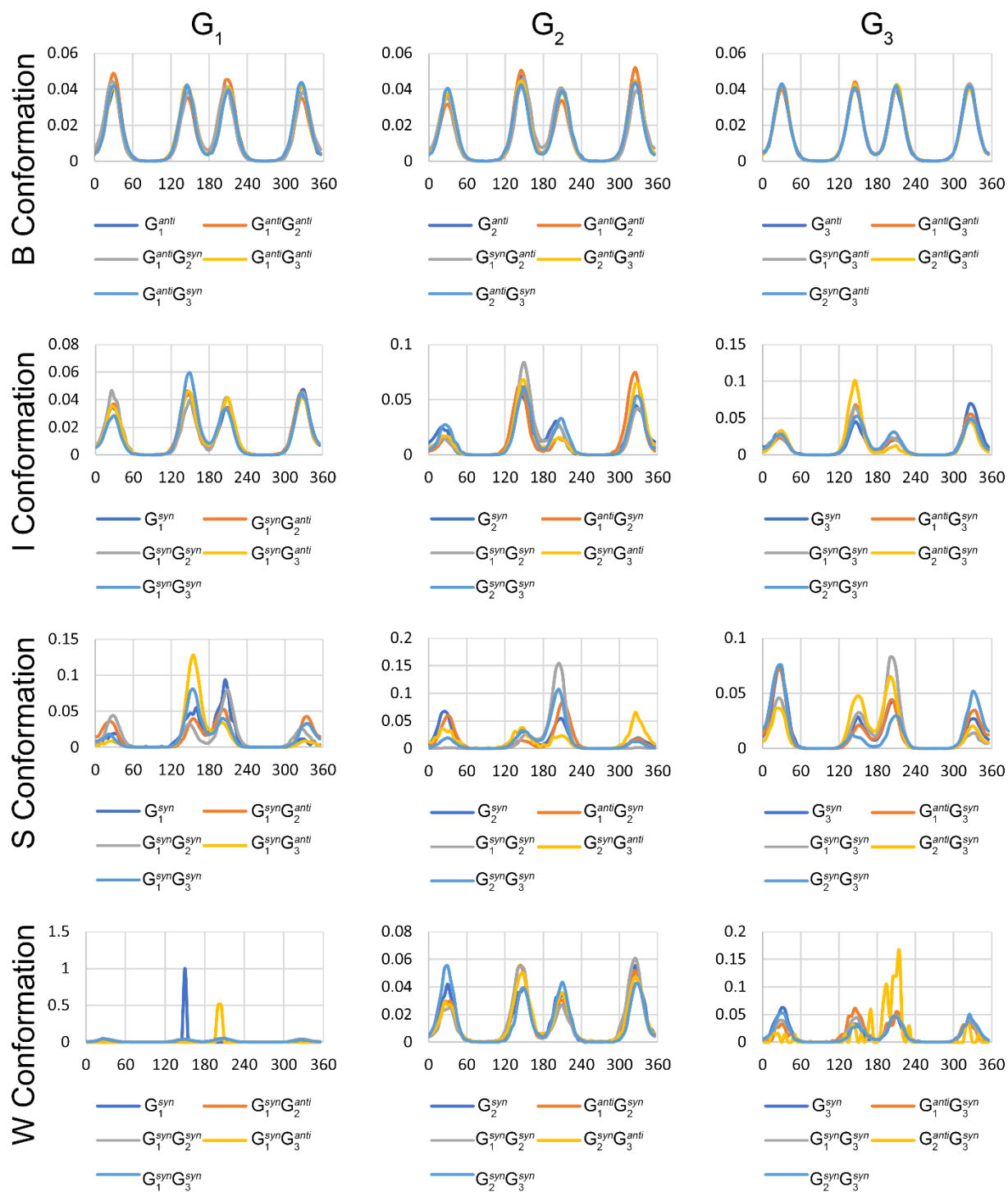


Figure F.15: Histograms of the  $\xi$  dihedral angle over all MD replicas for  $^{ABP}$ dG adducts at  $G_1$  (left),  $G_2$  (middle) and  $G_3$  (right) for each clustered conformation. The color indicates the damaged DNA system.

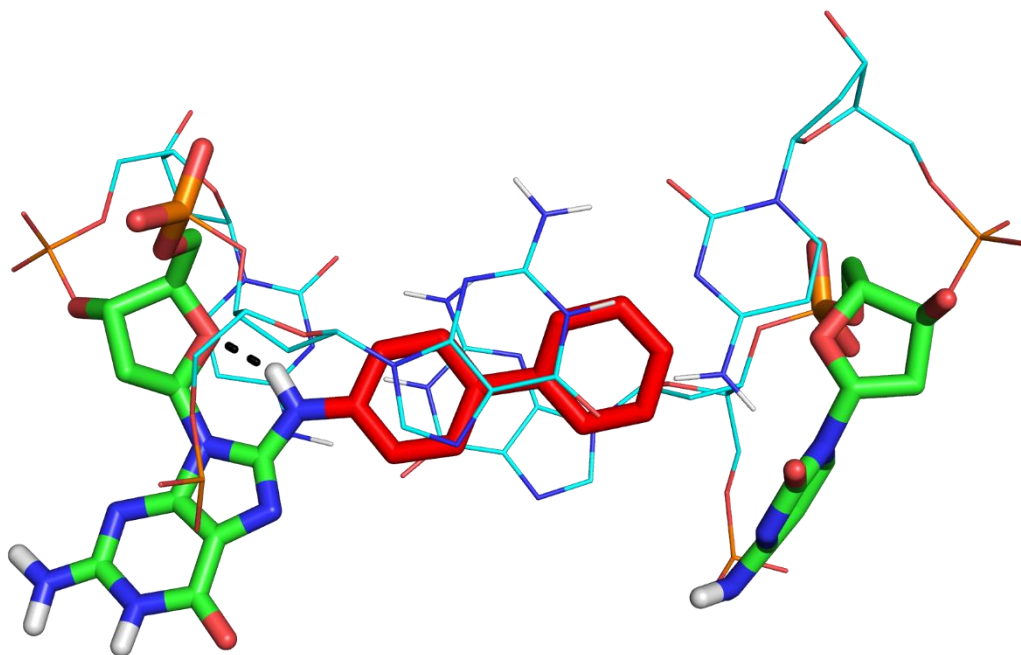


Figure F.16: Representative structure of  $G_2^S$  monoadducted DNA, highlighting the hydrogen bond between linker N–H and O4' achieved through a sugar rotation that also shifts  $\chi$ .

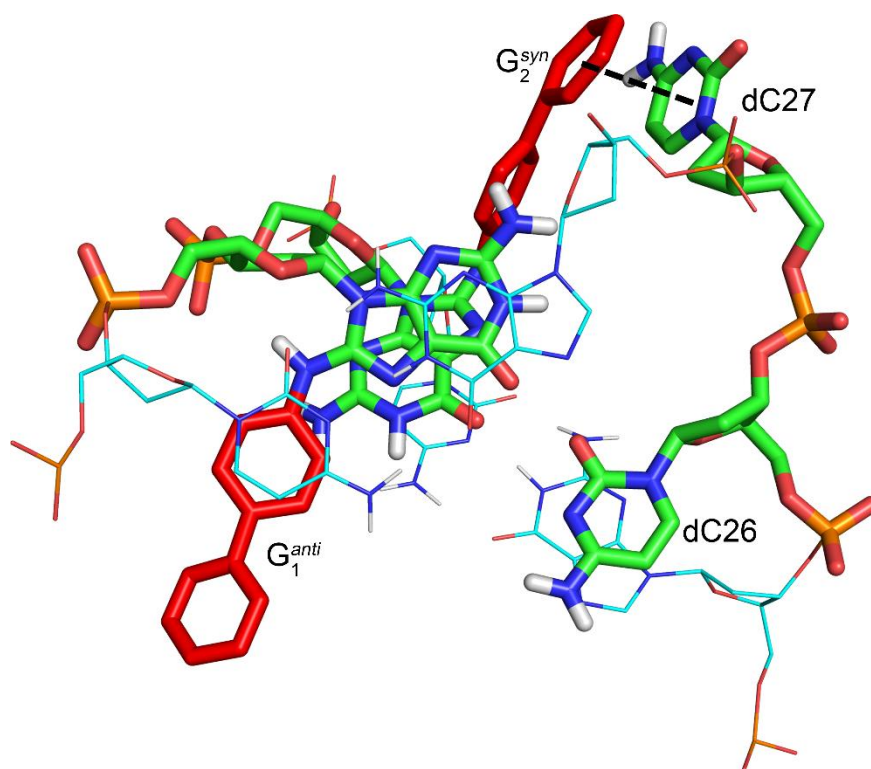


Figure F.17: Structure of  $G_1^{anti}G_2^{syn}$  diadducted DNA, highlighting disruption of the WC hydrogen bonds at  $G_1$  due to displacement of the pairing C into the minor groove to stack with the bulky group of the flanking  $G_2^{syn}$  adduct.

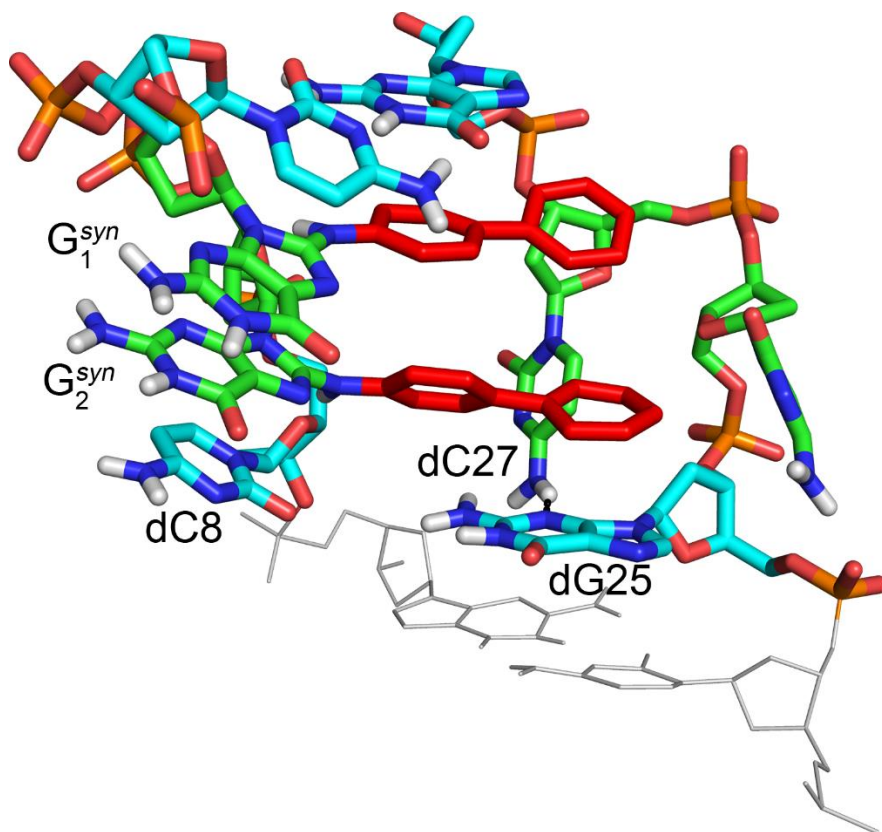


Figure F.18: Representative structure of  $G_1^{syn}G_2^{syn}$  diadducted DNA, highlighting the disruption in the WC hydrogen bonding in the 3' flanking base pair to increase stacking with  $G_2$ .

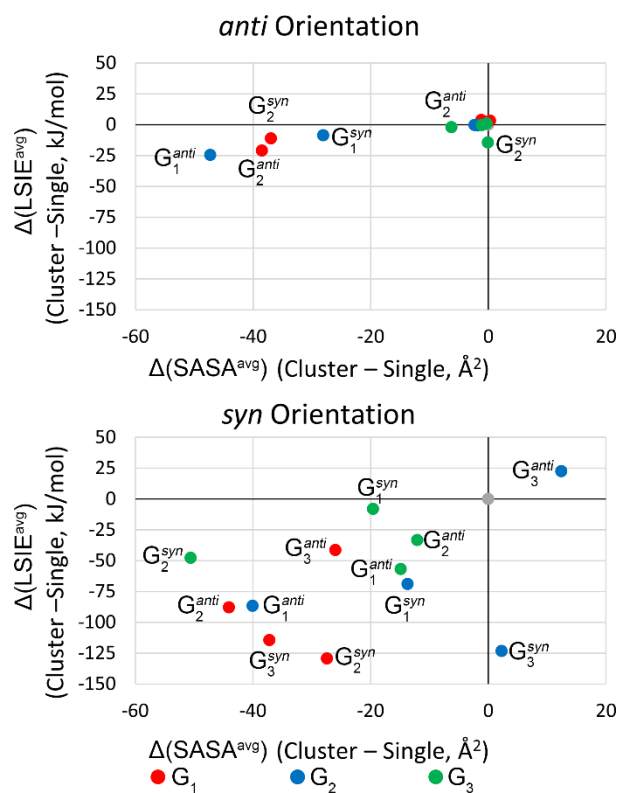


Figure F.19:  $\Delta(\text{LSIE}^{\text{avg}})$  versus  $\Delta(\text{SASA}^{\text{avg}})$  for <sup>ABP</sup>dG containing DNA upon addition of a second adduct with the first lesion in the *anti* (top) or *syn* (bottom) conformation at G<sub>1</sub> (red), G<sub>2</sub> (blue) or G<sub>3</sub> (green). Labels represent the site and conformation of the second lesion. Negative values indicate increased stability.



Table F.1: Average (standard deviation in parentheses) backbone RMSD (Å) across MD simulations of undamaged and damaged DNA.<sup>a</sup>

| System              | Primary Lesion                 | Secondary Lesion               | Replica 1 | Replica 2 | Replica 3 | Replica 4 | Replica 5 |
|---------------------|--------------------------------|--------------------------------|-----------|-----------|-----------|-----------|-----------|
| <b>Undamaged</b>    | N/A                            | N/A                            | 2.8 (0.6) | 2.6 (0.6) | 2.8 (0.6) | 2.8 (0.6) | 3.0 (0.6) |
| <b>Monoadducted</b> | G <sup>anti</sup> <sub>1</sub> | absent                         | 3.8 (0.8) | 3.8 (0.8) | 4.5 (0.7) | 4.5 (0.7) | 4.8 (0.6) |
|                     | G <sup>n</sup> <sub>1</sub>    | absent                         | 5.0 (0.7) | 5.3 (0.7) | 5.2 (0.7) | 5.1 (0.8) | 5.2 (0.7) |
|                     | G <sup>anti</sup> <sub>2</sub> | absent                         | 4.5 (0.9) | 4.8 (0.7) | 4.4 (0.8) | 4.0 (0.8) | 4.4 (0.8) |
|                     | G <sup>n</sup> <sub>2</sub>    | absent                         | 3.3 (0.9) | 6.1 (0.6) | 6.1 (0.9) | 5.5 (0.6) | 5.7 (0.7) |
|                     | G <sup>anti</sup> <sub>3</sub> | absent                         | 3.2 (0.6) | 3.0 (0.7) | 3.1 (0.6) | 3.2 (0.6) | 3.4 (0.7) |
|                     | G <sup>n</sup> <sub>3</sub>    | absent                         | 4.4 (0.5) | 4.4 (0.7) | 4.8 (0.8) | 5.6 (0.9) | 4.6 (0.6) |
| <b>Diadducted</b>   | G <sup>anti</sup> <sub>1</sub> | G <sup>anti</sup> <sub>2</sub> | 2.9 (0.7) | 2.9 (0.7) | 2.7 (0.6) | 3.0 (0.7) | 2.7 (0.6) |
|                     | G <sup>n</sup> <sub>1</sub>    | G <sup>anti</sup> <sub>2</sub> | 3.4 (0.8) | 4.1 (0.8) | 3.2 (0.5) | 4.1 (1.1) | 3.8 (1.0) |
|                     | G <sup>anti</sup> <sub>1</sub> | G <sup>n</sup> <sub>2</sub>    | 5.5 (1.2) | 3.3 (0.6) | 3.9 (0.8) | 4.1 (0.5) | 5.0 (1.5) |
|                     | G <sup>n</sup> <sub>1</sub>    | G <sup>n</sup> <sub>2</sub>    | 6.5 (0.9) | 5.2 (0.9) | 5.8 (0.9) | 6.3 (0.8) | 6.0 (1.2) |
|                     | G <sup>anti</sup> <sub>1</sub> | G <sup>anti</sup> <sub>3</sub> | 3.2 (0.8) | 3.3 (0.8) | 3.5 (0.7) | 3.6 (0.8) | 3.4 (0.7) |
|                     | G <sup>n</sup> <sub>1</sub>    | G <sup>anti</sup> <sub>3</sub> | 3.9 (0.9) | 3.8 (0.6) | 3.6 (0.6) | 3.8 (0.6) | 3.8 (0.7) |
|                     | G <sup>anti</sup> <sub>1</sub> | G <sup>n</sup> <sub>3</sub>    | 4.0 (0.9) | 3.3 (0.6) | 3.6 (0.7) | 3.5 (0.8) | 3.9 (0.7) |
|                     | G <sup>n</sup> <sub>1</sub>    | G <sup>n</sup> <sub>3</sub>    | 3.8 (0.7) | 4.3 (0.8) | 4.3 (0.9) | 4.6 (1.0) | 3.8 (0.8) |
|                     | G <sup>anti</sup> <sub>2</sub> | G <sup>anti</sup> <sub>3</sub> | 3.3 (0.6) | 3.1 (0.6) | 3.6 (0.6) | 3.1 (0.6) | 3.3 (0.6) |
|                     | G <sup>n</sup> <sub>2</sub>    | G <sup>anti</sup> <sub>3</sub> | 3.3 (0.7) | 3.0 (0.6) | 3.0 (0.6) | 3.0 (0.6) | 4.3 (1.0) |
|                     | G <sup>anti</sup> <sub>2</sub> | G <sup>n</sup> <sub>3</sub>    | 3.8 (0.6) | 4.1 (0.9) | 4.1 (0.7) | 4.1 (1.0) | 3.7 (0.6) |
|                     | G <sup>n</sup> <sub>2</sub>    | G <sup>n</sup> <sub>3</sub>    | 5.7 (1.1) | 5.4 (1.0) | 5.7 (1.0) | 4.2 (1.0) | 5.4 (0.8) |

<sup>a</sup>RMSDs calculated with respect to the first frame of the corresponding 500 ns MD production phase.

Table F.2: Occupancies (%) of the Watson-Crick hydrogen bonds at the 5' end of damaged DNA helix.<sup>a</sup>

| System              | Primary Lesion                 | Secondary Lesion               | Replica 1 | Replica 2 | Replica 3 | Replica 4 | Replica 5 |
|---------------------|--------------------------------|--------------------------------|-----------|-----------|-----------|-----------|-----------|
| <b>Undamaged</b>    | N/A                            | N/A                            | 93.3      | 90.4      | 94.5      | 92.7      | 94.6      |
| <b>Monoadducted</b> | G <sub>1</sub> <sup>anti</sup> | absent                         | 99.2      | 99.2      | 99.6      | 99.6      | 99.8      |
|                     | G <sub>1</sub> <sup>syn</sup>  | absent                         | 99.0      | 99.1      | 99.3      | 94.2      | 91.4      |
|                     | G <sub>2</sub> <sup>anti</sup> | absent                         | 99.2      | 99.7      | 99.3      | 96.0      | 93.9      |
|                     | G <sub>2</sub> <sup>syn</sup>  | absent                         | 99.3      | 99.2      | 99.1      | 99.2      | 93.7      |
|                     | G <sub>3</sub> <sup>anti</sup> | absent                         | 99.2      | 99.3      | 99.0      | 94.8      | 91.4      |
|                     | G <sub>3</sub> <sup>syn</sup>  | absent                         | 99.1      | 99.1      | 99.2      | 93.4      | 87.0      |
| <b>Diadducted</b>   | G <sub>1</sub> <sup>anti</sup> | G <sub>2</sub> <sup>anti</sup> | 99.3      | 99.1      | 99.3      | 91.7      | 90.5      |
|                     | G <sub>1</sub> <sup>syn</sup>  | G <sub>2</sub> <sup>anti</sup> | 99.5      | 99.3      | 99.3      | 93.3      | 96.8      |
|                     | G <sub>1</sub> <sup>anti</sup> | G <sub>2</sub> <sup>syn</sup>  | 99.3      | 99.2      | 99.6      | 90.4      | 92.5      |
|                     | G <sub>1</sub> <sup>syn</sup>  | G <sub>2</sub> <sup>syn</sup>  | 99.3      | 99.0      | 99.3      | 93.8      | 91.2      |
|                     | G <sub>1</sub> <sup>anti</sup> | G <sub>3</sub> <sup>anti</sup> | 99.4      | 99.3      | 99.3      | 94.7      | 90.1      |
|                     | G <sub>1</sub> <sup>syn</sup>  | G <sub>3</sub> <sup>anti</sup> | 99.0      | 99.2      | 99.3      | 84.1      | 91.9      |
|                     | G <sub>1</sub> <sup>anti</sup> | G <sub>3</sub> <sup>syn</sup>  | 99.6      | 99.1      | 99.1      | 91.7      | 88.3      |
|                     | G <sub>1</sub> <sup>syn</sup>  | G <sub>3</sub> <sup>syn</sup>  | 99.2      | 99.2      | 99.1      | 93.6      | 99.4      |
|                     | G <sub>2</sub> <sup>anti</sup> | G <sub>3</sub> <sup>anti</sup> | 99.3      | 99.2      | 99.2      | 91.2      | 99.2      |
|                     | G <sub>2</sub> <sup>syn</sup>  | G <sub>3</sub> <sup>anti</sup> | 99.0      | 99.3      | 99.4      | 94.7      | 93.6      |
|                     | G <sub>2</sub> <sup>anti</sup> | G <sub>3</sub> <sup>syn</sup>  | 99.1      | 99.2      | 99.1      | 94.6      | 91.1      |
|                     | G <sub>2</sub> <sup>syn</sup>  | G <sub>3</sub> <sup>syn</sup>  | 99.2      | 99.3      | 99.1      | 94.7      | 87.9      |

<sup>a</sup>Occupancies determined as the percentage of the simulation that at least three hydrogen bonds exist according to a heavy atom distance < 3.2 Å and a heavy atom–hydrogen–heavy atom angle > 120°.

Table F.3: Occupancies (%) of the Watson-Crick hydrogen bonds at the 3' end of damaged DNA helix.<sup>a</sup>

| System              | Primary Lesion                 | Secondary Lesion               | Replica 1 | Replica 2 | Replica 3 | Replica 4 | Replica 5 |
|---------------------|--------------------------------|--------------------------------|-----------|-----------|-----------|-----------|-----------|
| <b>Undamaged</b>    | N/A                            | N/A                            | 96.9      | 97.5      | 96.0      | 97.1      | 97.0      |
| <b>Monoadducted</b> | G <sub>1</sub> <sup>anti</sup> | absent                         | 99.0      | 99.1      | 99.5      | 99.6      | 99.7      |
|                     | G <sub>3</sub> <sup>syn</sup>  | absent                         | 99.8      | 99.8      | 99.8      | 96.8      | 95.8      |
|                     | G <sub>2</sub> <sup>anti</sup> | absent                         | 99.6      | 99.4      | 99.7      | 97.4      | 97.2      |
|                     | G <sub>2</sub> <sup>syn</sup>  | absent                         | 99.8      | 99.7      | 99.8      | 99.9      | 95.8      |
|                     | G <sub>3</sub> <sup>anti</sup> | absent                         | 99.8      | 99.7      | 99.8      | 97.1      | 94.9      |
|                     | G <sub>3</sub> <sup>syn</sup>  | absent                         | 99.7      | 99.7      | 99.8      | 96.6      | 97.0      |
| <b>Diadducted</b>   | G <sub>1</sub> <sup>anti</sup> | G <sub>2</sub> <sup>anti</sup> | 99.8      | 99.8      | 99.8      | 97.4      | 97.1      |
|                     | G <sub>3</sub> <sup>syn</sup>  | G <sub>2</sub> <sup>anti</sup> | 99.8      | 99.8      | 99.8      | 96.6      | 98.3      |
|                     | G <sub>1</sub> <sup>anti</sup> | G <sub>2</sub> <sup>syn</sup>  | 99.7      | 99.8      | 99.8      | 97.2      | 97.3      |
|                     | G <sub>1</sub> <sup>syn</sup>  | G <sub>2</sub> <sup>syn</sup>  | 99.8      | 99.8      | 99.7      | 97.0      | 96.9      |
|                     | G <sub>1</sub> <sup>anti</sup> | G <sub>3</sub> <sup>anti</sup> | 99.8      | 99.6      | 99.8      | 95.8      | 97.1      |
|                     | G <sub>3</sub> <sup>syn</sup>  | G <sub>3</sub> <sup>anti</sup> | 99.8      | 99.8      | 99.7      | 94.2      | 97.1      |
|                     | G <sub>1</sub> <sup>anti</sup> | G <sub>3</sub> <sup>syn</sup>  | 99.8      | 99.8      | 99.8      | 97.1      | 97.1      |
|                     | G <sub>1</sub> <sup>syn</sup>  | G <sub>3</sub> <sup>syn</sup>  | 99.8      | 99.7      | 99.8      | 97.4      | 99.8      |
|                     | G <sub>2</sub> <sup>anti</sup> | G <sub>3</sub> <sup>anti</sup> | 99.8      | 99.7      | 99.8      | 97.2      | 99.7      |
|                     | G <sub>2</sub> <sup>syn</sup>  | G <sub>3</sub> <sup>anti</sup> | 99.8      | 99.8      | 99.7      | 97.9      | 96.8      |
|                     | G <sub>2</sub> <sup>anti</sup> | G <sub>3</sub> <sup>syn</sup>  | 99.7      | 99.8      | 99.9      | 97.2      | 97.3      |
|                     | G <sub>2</sub> <sup>syn</sup>  | G <sub>3</sub> <sup>syn</sup>  | 99.8      | 99.8      | 99.8      | 96.8      | 97.4      |

<sup>a</sup>Occupancies determined as the percentage of the simulation that at least three hydrogen bonds exist according to a heavy atom distance < 3.2 Å and a heavy atom–hydrogen–heavy atom angle > 120°.

Table F.4: Occupancies (%) of the Watson-Crick hydrogen bonds flanking the lesion site in monoadducted DNA.<sup>a,b</sup>

| Site        | Base Pair | Replica 1 | Replica 2 | Replica 3 | Replica 4 | Replica 5 |
|-------------|-----------|-----------|-----------|-----------|-----------|-----------|
| $G_1^{nti}$ | dC5:dG28  | 99.0      | 99.0      | 99.1      | 99.1      | 99.0      |
|             | dG7:dC26  | 98.9      | 98.9      | 98.7      | 98.9      | 98.8      |
| $G_1^n$     | dC5:dG28  | 98.9      | 98.9      | 99.0      | 98.9      | 98.9      |
|             | dG7:dC26  | 99.0      | 99.0      | 98.9      | 98.9      | 99.1      |
| $G_2^{nti}$ | dG6:dC27  | 98.8      | 99.2      | 98.9      | 99.1      | 99.1      |
|             | dC8:dG25  | 99.3      | 99.5      | 99.3      | 99.3      | 99.2      |
| $G_2^n$     | dG6:dC27  | 97.5      | 98.5      | 99.1      | 99.1      | 99.0      |
|             | dC8:dG25  | 98.3      | 97.9      | 98.3      | 97.9      | 95.3      |
| $G_3^{nti}$ | dC8:dG25  | 98.9      | 99.0      | 98.9      | 99.0      | 99.1      |
|             | dC10:dG23 | 98.7      | 98.7      | 98.8      | 98.9      | 98.9      |
| $G_3^n$     | dC8:dG25  | 99.0      | 98.5      | 99.1      | 99.1      | 99.2      |
|             | dC10:dG23 | 88.7      | 82.3      | 91.5      | 89.5      | 83.8      |

<sup>a</sup>Nucleotide numbering is illustrated in Figure F.14. <sup>b</sup>Occupancies determined as the percentage of the simulation that at least three hydrogen bonds exist according to a heavy atom distance < 3.2 Å and a heavy atom–hydrogen–heavy atom angle > 120°.

Table F.5: Occupancies (%) of the Watson-Crick hydrogen bonds flanking the lesion sites in diadducted DNA.<sup>a,b</sup>

| Sites                  | Base Pair | Replica 1 | Replica 2 | Replica 3 | Replica 4 | Replica 5 |
|------------------------|-----------|-----------|-----------|-----------|-----------|-----------|
| $G_1^{anti}G_2^{anti}$ | dC5:dG28  | 99.1      | 99.0      | 98.9      | 98.9      | 99.0      |
|                        | dC8:dG25  | 99.4      | 99.3      | 99.2      | 99.2      | 99.4      |
| $G_1^{anti}G_2^n$      | dC5:dG28  | 98.2      | 98.7      | 98.6      | 99.0      | 98.5      |
|                        | dC8:dG25  | 97.7      | 98.4      | 97.6      | 98.0      | 98.7      |
| $G_1^nG_2^{anti}$      | dC5:dG28  | 99.0      | 99.1      | 97.6      | 98.5      | 98.9      |
|                        | dC8:dG25  | 99.3      | 99.1      | 99.2      | 98.6      | 99.3      |
| $G_1^nG_2^n$           | dC5:dG28  | 98.6      | 97.1      | 98.8      | 98.1      | 98.3      |
|                        | dC8:dG25  | 97.8      | 0.1       | 85.7      | 98.8      | 30.9      |
| $G_1^{anti}G_3^{anti}$ | dC5:dG28  | 98.9      | 99.1      | 99.1      | 99.0      | 99.1      |
|                        | dG7:dC26  | 98.9      | 99.0      | 98.8      | 99.0      | 98.7      |
|                        | dC8:dG25  | 99.0      | 98.8      | 99.0      | 99.2      | 99.0      |
|                        | dC10:dG23 | 98.8      | 98.7      | 98.8      | 98.8      | 98.7      |
| $G_1^{anti}G_3^n$      | dC5:dG28  | 98.9      | 98.9      | 99.1      | 99.1      | 99.0      |
|                        | dG7:dC26  | 98.7      | 97.3      | 98.3      | 98.2      | 98.9      |
|                        | dC8:dG25  | 98.8      | 97.9      | 98.8      | 98.1      | 99.2      |
|                        | dC10:dG23 | 81.3      | 87.9      | 76.2      | 83.3      | 95.0      |
| $G_1^nG_3^{anti}$      | dC5:dG28  | 99.0      | 98.5      | 98.1      | 98.3      | 98.1      |
|                        | dG7:dC26  | 98.8      | 98.9      | 98.7      | 98.8      | 98.9      |
|                        | dC8:dG25  | 98.9      | 99.2      | 99.0      | 98.8      | 99.2      |
|                        | dC10:dG23 | 98.8      | 98.7      | 98.9      | 98.6      | 98.6      |
| $G_1^nG_3^n$           | dC5:dG28  | 98.7      | 98.5      | 98.3      | 98.1      | 98.7      |
|                        | dG7:dC26  | 99.1      | 98.8      | 98.7      | 97.9      | 98.3      |
|                        | dC8:dG25  | 98.6      | 99.0      | 98.8      | 98.8      | 98.3      |
|                        | dC10:dG23 | 98.2      | 94.1      | 92.8      | 77.0      | 90.5      |
| $G_2^{anti}G_3^{anti}$ | dG6:dC27  | 99.0      | 98.8      | 98.9      | 99.0      | 99.0      |
|                        | dC8:dG25  | 99.1      | 98.9      | 99.2      | 99.2      | 99.1      |
|                        | dC10:dG23 | 98.8      | 98.8      | 98.8      | 98.7      | 98.7      |
| $G_2^{anti}G_3^n$      | dG6:dC27  | 99.0      | 98.9      | 99.3      | 99.2      | 99.1      |
|                        | dC8:dG25  | 98.2      | 99.0      | 98.3      | 98.9      | 99.0      |
|                        | dC10:dG23 | 92.7      | 99.1      | 96.3      | 98.5      | 95.3      |
| $G_2^nG_3^{anti}$      | dG6:dC27  | 98.9      | 99.0      | 99.1      | 98.9      | 98.6      |
|                        | dC8:dG25  | 98.2      | 98.0      | 98.1      | 98.3      | 94.4      |
|                        | dC10:dG23 | 99.0      | 99.0      | 98.8      | 99.0      | 99.2      |
| $G_2^nG_3^n$           | dG6:dC27  | 98.9      | 99.0      | 98.9      | 99.0      | 99.1      |
|                        | dC8:dG25  | 97.8      | 98.2      | 53.7      | 98.3      | 86.7      |
|                        | dC10:dG23 | 98.5      | 96.4      | 98.0      | 98.8      | 98.6      |

<sup>a</sup>Nucleotide numbering is illustrated in Figure F.14. <sup>b</sup>Occupancies determined as the percentage of the simulation that at least three hydrogen bonds exist according to a heavy atom distance < 3.2 Å and a heavy atom–hydrogen–heavy atom angle > 120°.

Table F.6: Occupancies (%) of the Watson-Crick hydrogen bonds at G<sub>1</sub>, G<sub>2</sub> or G<sub>3</sub> for undamaged, monoadducted and diadducted DNA.<sup>a</sup>

| System       | Primary Lesion                 | Secondary Lesion               | Replica 1 | Replica 2 | Replica 3 | Replica 4 | Replica 5 | Average | Standard Deviation <sup>b</sup> |
|--------------|--------------------------------|--------------------------------|-----------|-----------|-----------|-----------|-----------|---------|---------------------------------|
| Undamaged    | G <sub>1</sub>                 | N/A                            | 99.1      | 99.2      | 99.1      | 99.0      | 99.1      | 99.1    | 0.1                             |
|              | G <sub>2</sub>                 | N/A                            | 98.7      | 98.6      | 98.7      | 98.8      | 98.8      | 98.7    | 0.1                             |
|              | G <sub>3</sub>                 | N/A                            | 99.3      | 99.2      | 99.3      | 99.3      | 99.2      | 99.2    | 0.0                             |
| Monoadducted | G <sub>1</sub> <sup>anti</sup> | absent                         | 98.8      | 98.6      | 98.1      | 98.5      | 98.9      | 98.6    | 0.3                             |
|              | G <sub>2</sub> <sup>anti</sup> | absent                         | 99.6      | 99.6      | 99.6      | 99.6      | 99.8      | 99.7    | 0.1                             |
|              | G <sub>3</sub> <sup>anti</sup> | absent                         | 98.9      | 99.3      | 99.4      | 99.3      | 99.4      | 99.3    | 0.2                             |
| Diadducted   | G <sub>1</sub> <sup>anti</sup> | G <sub>2</sub> <sup>anti</sup> | 98.6      | 98.4      | 98.9      | 99.1      | 99.1      | 98.8    | 0.3                             |
|              | G <sub>1</sub> <sup>anti</sup> | G <sub>2</sub> <sup>syn</sup>  | 82.3      | 99.1      | 78.6      | 28.1      | 98.1      | 77.2    | 25.9                            |
|              | G <sub>1</sub> <sup>anti</sup> | G <sub>3</sub> <sup>anti</sup> | 98.3      | 98.4      | 98.4      | 98.2      | 97.7      | 98.2    | 0.3                             |
|              | G <sub>1</sub> <sup>anti</sup> | G <sub>3</sub> <sup>syn</sup>  | 98.4      | 98.6      | 98.7      | 98.9      | 98.8      | 98.7    | 0.2                             |
|              | G <sub>2</sub> <sup>anti</sup> | G <sub>1</sub> <sup>anti</sup> | 99.8      | 99.7      | 99.8      | 99.7      | 99.7      | 99.7    | 0.0                             |
|              | G <sub>2</sub> <sup>anti</sup> | G <sub>1</sub> <sup>syn</sup>  | 99.2      | 98.9      | 99.5      | 98.3      | 98.5      | 98.9    | 0.4                             |
|              | G <sub>2</sub> <sup>anti</sup> | G <sub>3</sub> <sup>anti</sup> | 99.6      | 99.4      | 99.5      | 99.7      | 99.5      | 99.6    | 0.1                             |
|              | G <sub>2</sub> <sup>anti</sup> | G <sub>3</sub> <sup>syn</sup>  | 99.5      | 99.6      | 99.6      | 99.4      | 99.6      | 99.5    | 0.1                             |
|              | G <sub>3</sub> <sup>anti</sup> | G <sub>1</sub> <sup>anti</sup> | 99.2      | 99.5      | 99.3      | 99.3      | 99.5      | 99.4    | 0.1                             |
|              | G <sub>3</sub> <sup>anti</sup> | G <sub>1</sub> <sup>syn</sup>  | 99.5      | 99.3      | 99.3      | 99.4      | 99.6      | 99.4    | 0.1                             |
|              | G <sub>3</sub> <sup>anti</sup> | G <sub>2</sub> <sup>anti</sup> | 99.4      | 99.3      | 99.4      | 99.3      | 99.5      | 99.4    | 0.1                             |
|              | G <sub>3</sub> <sup>anti</sup> | G <sub>2</sub> <sup>syn</sup>  | 99.0      | 99.1      | 99.2      | 99.1      | 99.0      | 99.1    | 0.1                             |

<sup>a</sup>Occupancies determined as the percentage of the simulation that at least three hydrogen bonds exist according to a heavy atom distance < 3.2 Å and a heavy atom–hydrogen–heavy atom angle > 120°. <sup>b</sup>Standard deviation calculated between the replicas.

Table F.7: Occupancies (%) and number of frames<sup>a</sup> (in parentheses) of the S, I, and W conformations of <sup>ABP</sup>dG monoadducted and diadducted DNA.

| Primary Lesion                   | Secondary Lesion <sup>b</sup>  | Intercalated   | Stacking       | Wedge          |
|----------------------------------|--------------------------------|----------------|----------------|----------------|
| <b>G<sub>1</sub><sup>n</sup></b> | absent                         | 98.44% (49219) | 1.56% (780)    | 0.00% (1)      |
|                                  | G <sub>2</sub> <sup>anti</sup> | 38.23% (19117) | 61.77% (30883) | 0.00% (0)      |
|                                  | G <sub>2</sub> <sup>n</sup>    | 16.21% (8106)  | 40.10% (20049) | 43.69% (21845) |
|                                  | G <sub>3</sub> <sup>anti</sup> | 64.52% (32258) | 35.48% (17740) | 0.00% (2)      |
|                                  | G <sub>3</sub> <sup>n</sup>    | 50.91% (25455) | 43.91% (21955) | 5.18% (2590)   |
| <b>G<sub>2</sub><sup>n</sup></b> | absent                         | 55.83% (27917) | 29.05% (14526) | 15.11% (7557)  |
|                                  | G <sub>2</sub> <sup>anti</sup> | 33.01% (16505) | 58.38% (29190) | 8.61% (4305)   |
|                                  | G <sub>1</sub> <sup>n</sup>    | 24.84% (12419) | 48.17% (24083) | 27.00% (13498) |
|                                  | G <sub>3</sub> <sup>anti</sup> | 83.47% (41736) | 3.26% (1631)   | 13.27% (6633)  |
|                                  | G <sub>3</sub> <sup>n</sup>    | 27.27% (13635) | 34.06% (17032) | 38.67% (19333) |
| <b>G<sub>3</sub><sup>n</sup></b> | absent                         | 62.66% (31331) | 30.89% (15443) | 6.45% (3226)   |
|                                  | G <sub>2</sub> <sup>anti</sup> | 31.98% (15988) | 62.98% (31492) | 5.04% (2520)   |
|                                  | G <sub>1</sub> <sup>n</sup>    | 44.07% (22033) | 54.38% (27188) | 1.56% (779)    |
|                                  | G <sub>2</sub> <sup>anti</sup> | 25.28% (12638) | 44.30% (22152) | 30.42% (15210) |
|                                  | G <sub>2</sub> <sup>n</sup>    | 35.62% (17808) | 64.03% (32013) | 0.36% (179)    |

<sup>a</sup>A total of 10,000 frames were analyzed from each replica, resulting in a total of 50,000 frames for each system.

<sup>b</sup>Absent indicates monoadducted DNA.

Table F.8: Average (standard deviation in parentheses) structural features of monoadducted DNA, including the C1'–C1' distance at the lesion site, the  $\chi^{\text{opp}}$  ( $\chi$  for the C opposing the adduct) dihedral angle and the angle between the two ring planes of the biphenyl moiety.

| System       | Conformation | Site                        | C1'–C1' (Å)            | $\chi^{\text{opp}}$ (deg.) | Ring Interplanar Angle (deg.) <sup>b</sup> |
|--------------|--------------|-----------------------------|------------------------|----------------------------|--------------------------------------------|
| undamaged    | N/A          | G <sub>1</sub>              | 10.7 (0.2)             | 252.1 (19.3)               | N/A                                        |
|              |              | G <sub>2</sub>              | 10.8 (0.2)             | 247.7 (16.1)               | N/A                                        |
|              |              | G <sub>3</sub>              | 10.7 (0.2)             | 253.2 (16.9)               | N/A                                        |
| monoadducted | B            | G <sub>1</sub> <sup>B</sup> | 10.7 (0.2)             | 256.4 (19.2)               | 33.7 (11.5)                                |
|              |              | G <sub>2</sub> <sup>B</sup> | 10.8 (0.2)             | 250.0 (15.6)               | 33.4 (11.4)                                |
|              |              | G <sub>3</sub> <sup>B</sup> | 10.7 (0.2)             | 259.1 (15.1)               | 33.5 (11.5)                                |
|              | I            | G <sub>1</sub> <sup>I</sup> | 8.3 (0.5)              | 239.5 (10.8)               | 30.6 (11.5)                                |
|              |              | G <sub>2</sub> <sup>I</sup> | 8.4 (0.8)              | 236.4 (11.5)               | 27.9 (11.0)                                |
|              |              | G <sub>3</sub> <sup>I</sup> | 8.7 (1.0)              | 242.3 (54.8)               | 29.6 (11.3)                                |
|              | S            | G <sub>1</sub> <sup>S</sup> | 12.9 (2.0)             | 247.3 (17.4)               | 27.6 (11.3)                                |
|              |              | G <sub>2</sub> <sup>S</sup> | 15.3 (1.4)             | 247.5 (20.7)               | 25.5 (9.7)                                 |
|              |              | G <sub>3</sub> <sup>S</sup> | 12.6 (2.1)             | 245.3 (33.8)               | 28.4 (10.9)                                |
|              | W            | G <sub>1</sub> <sup>W</sup> | 7.6 (N/A) <sup>a</sup> | 281.6 (N/A) <sup>a</sup>   | 34.1 (0.0)                                 |
|              |              | G <sub>2</sub> <sup>W</sup> | 11.4 (0.4)             | 249.9 (14.1)               | 33.9 (11.5)                                |
|              |              | G <sub>3</sub> <sup>W</sup> | 9.0 (0.6)              | 63.1 (11.9)                | 31.8 (10.6)                                |

<sup>a</sup>Conformation only has one frame for this system. <sup>b</sup>Calculated using vectors in cpptraj.



Table F.9: Average (standard deviation in parentheses)  $\chi^{\text{avg}}$ ,  $\theta^{\text{avg}}$ ,  $\phi^{\text{avg}}$ , and  $\xi^{\text{avg}}$  dihedral angles (deg.) for each conformation across all replicas for undamaged and monoadducted DNA.<sup>a,b</sup>

| System       | Conformation      | Site                                                          | $\chi^{\text{avg}}$     | $\theta^{\text{avg}}$    | $\phi^{\text{avg}}$     | $\xi^{\text{avg}}$       |
|--------------|-------------------|---------------------------------------------------------------|-------------------------|--------------------------|-------------------------|--------------------------|
| undamaged    | N/A               | G <sub>1</sub>                                                | 256.9<br>(14.9)         | N/A                      | N/A                     | N/A                      |
|              |                   | G <sub>2</sub>                                                | 249.4<br>(15.5)         | N/A                      | N/A                     | N/A                      |
|              |                   | G <sub>3</sub>                                                | 253.5<br>(15.3)         | N/A                      | N/A                     | N/A                      |
| monoadducted | B ( <i>anti</i> ) | G <sub>1</sub> <sup>B</sup> (G <sub>1</sub> <sup>anti</sup> ) | 226.3<br>(16.7)         | 129.1 (30.9)             | 0.0 (11.8)              | −4.4 (18.9)              |
|              |                   | G <sub>2</sub> <sup>B</sup> (G <sub>2</sub> <sup>anti</sup> ) | 225.2<br>(12.7)         | 142.5 (29.8)             | 5.7 (12.2)              | −13.7 (20.8)             |
|              |                   | G <sub>3</sub> <sup>B</sup> (G <sub>3</sub> <sup>anti</sup> ) | 228.8<br>(13.9)         | 133.3 (23.3)             | 2.3 (11.9)              | 3.5 (22.7)               |
|              | <i>syn</i>        | G <sub>1</sub> <sup>n</sup>                                   | 49.6 (13.6)             | 253.0 (29.8)             | 3.1 (10.6)              | 134.4 (94.9)             |
|              |                   | G <sub>2</sub> <sup>n</sup>                                   | 66.6 (53.4)             | 229.6 (68.3)             | 0.3 (10.5)              | 186.4 (82.7)             |
|              |                   | G <sub>3</sub> <sup>n</sup>                                   | 46.1 (15.3)             | 237.0 (34.2)             | 2.1 (11.1)              | 176.6 (105.7)            |
|              | I                 | G <sub>1</sub> <sup>I</sup>                                   | 49.7 (13.3)             | 254.4 (25.1)             | 3.3 (10.6)              | 29.5 (95.1)              |
|              |                   | G <sub>2</sub> <sup>I</sup>                                   | 46.8 (15.4)             | 256.9 (31.1)             | 0.5 (10.0)              | −15.5 (89.7)             |
|              |                   | G <sub>3</sub> <sup>I</sup>                                   | 48.2 (18.7)             | 248.5 (43.5)             | 3.1 (10.7)              | −17.5 (94.5)             |
|              | S                 | G <sub>1</sub> <sup>S</sup>                                   | 50.6 (22.5)             | 197.4 (35.5)             | −2.6 (11.6)             | 3.0 (70.9)               |
|              |                   | G <sub>2</sub> <sup>S</sup>                                   | 164.4<br>(39.9)         | 180.2 (23.6)             | −1.4 (11.6)             | 18.8 (28.8)              |
|              |                   | G <sub>3</sub> <sup>S</sup>                                   | 58.1 (44.5)             | 196.7 (40.6)             | −3.0 (11.9)             | 3.7 (77.3)               |
|              | W                 | G <sub>1</sub> <sup>W</sup>                                   | 61.8 (N/A) <sup>c</sup> | 226.6 (N/A) <sup>c</sup> | 19.2 (N/A) <sup>c</sup> | −25.9 (N/A) <sup>c</sup> |
|              |                   | G <sub>2</sub> <sup>W</sup>                                   | 61.4 (14.8)             | 79.5 (19.5)              | 2.4 (10.1)              | −54.0 (95.2)             |
|              |                   | G <sub>3</sub> <sup>W</sup>                                   | 48.9 (13.8)             | 258.9 (22.3)             | 7.2 (11.1)              | −19.8 (92.8)             |

<sup>a</sup> $\chi = \angle(\text{O4}'\text{--C1}'\text{--N9--C4})$ ,  $\theta = \angle(\text{N9--C8--N10--C10})$ ,  $\phi = \angle(\text{C8--N10--C10--C11})$  and  $\xi = \angle(\text{C12--C13--C16--C17})$ .

<sup>b</sup>Standard deviation calculated over all frames corresponding to the specified conformation. <sup>c</sup>Conformation only has one frame for this system.

Table F.10: Occupancies (%) of the non-Watson-Crick hydrogen bonds at the lesion site of monoadducted DNA.<sup>a,b</sup>

| Site                             | Hydrogen Bond           | Occupancy | Site                             | Hydrogen Bond            | Occupancy |
|----------------------------------|-------------------------|-----------|----------------------------------|--------------------------|-----------|
| <b>G<sub>1</sub><sup>B</sup></b> | (G6)N10–<br>H•••O3'(C5) | 34.2%     | <b>G<sub>1</sub><sup>I</sup></b> | (G6)N10–<br>H•••N1(G28)  | 19.8%     |
|                                  | (G6)N10–<br>H•••O4'(G6) | 41.9%     |                                  | (G6)N10–<br>H•••N2(G28)  | 57.4%     |
| <b>G<sub>2</sub><sup>B</sup></b> | (G7)N10–<br>H•••O4'(G7) | 79.1%     |                                  | (G6)N10–<br>H•••N3(G28)  | 19.2%     |
| <b>G<sub>3</sub><sup>B</sup></b> | (G9)N10–<br>H•••O4'(G9) | 66.0%     | <b>G<sub>2</sub><sup>I</sup></b> | (G7)N10–H•••N2(G6)       | 30.0%     |
| <b>G<sub>1</sub><sup>W</sup></b> | (G6)N10–<br>H•••N2(G28) | 100.0%    |                                  | (G7)N10–<br>H•••O2(C27)  | 19.5%     |
|                                  | (C26)N4–<br>H•••O6(G6)  | 100.0%    | <b>G<sub>3</sub><sup>I</sup></b> | (G9)N10–<br>H•••N2(G25)  | 36.6%     |
|                                  | (C27)N4–<br>H•••O6(G6)  | 100.0%    |                                  | (G9)N10–<br>H•••N3(G25)  | 27.4%     |
|                                  | (G28)N2–<br>H•••O4'(G6) | 100.0%    | <b>G<sub>1</sub><sup>S</sup></b> | (C27)N4–<br>H•••OP2(C26) | 36.0%     |
| <b>G<sub>2</sub><sup>W</sup></b> | (G7)N10–<br>H•••O4'(C8) | 42.2%     | <b>G<sub>2</sub><sup>S</sup></b> | (G7)N10–H•••O4'(G7)      | 84.2%     |
|                                  | (C26)N4–<br>H•••O6(G7)  | 75.6%     |                                  | (C26)N4–<br>H•••OP2(G25) | 62.7%     |
|                                  | (C26)N4–<br>H•••N7(G7)  | 44.2%     |                                  | (C27)N4–<br>H•••O2(C26)  | 46.6%     |
| <b>G<sub>3</sub><sup>W</sup></b> | (G9)N10–<br>H•••N2(G25) | 56.3%     | <b>G<sub>3</sub><sup>S</sup></b> | (G9)N10–H•••O4'(G9)      | 14.1%     |
|                                  | (G9)N10–<br>H•••N3(G25) | 27.9%     |                                  | (C23)N2–<br>H•••O2(C24)  | 14.4%     |
|                                  | (C24)N4–<br>H•••O6(G9)  | 67.2%     |                                  | (C23)N2–<br>H•••N3(C24)  | 14.0%     |
|                                  | (C24)N4–<br>H•••N7(G9)  | 41.4%     |                                  | (C24)N4–<br>H•••OP2(G23) | 10.9%     |
|                                  |                         |           |                                  |                          |           |

<sup>a</sup>Nucleotide numbering is illustrated in Figure F.14 <sup>b</sup>Occupancies determined as the percentage of the simulation that a hydrogen bonds exist according to a heavy atom distance < 3.2 Å and a heavy atom–hydrogen–heavy atom angle > 120°.

Table F.11: Average (standard deviation in parentheses) pseudostep parameters<sup>a</sup> for undamaged, monoadducted and diadducted DNA.

| System                                                          | Rise (Å)   | Shift (Å)  | Slide (Å) | Roll (deg.)  | Tilt (deg.)  | Twist (deg.)  |
|-----------------------------------------------------------------|------------|------------|-----------|--------------|--------------|---------------|
| <b>undamaged</b>                                                | 22.8 (0.8) | −0.9 (2.6) | 3.4 (2.6) | −8.8 (12.7)  | 1.7 (14.5)   | −115.0 (9.1)  |
| <b>G<sub>1</sub><sup>anti</sup></b>                             | 22.9 (0.8) | 0.7 (2.6)  | 2.9 (2.6) | −5.1 (11.9)  | 0.1 (13.2)   | −116.4 (10.1) |
| <b>G<sub>2</sub><sup>anti</sup></b>                             | 22.6 (1.0) | 0.8 (2.9)  | 4.3 (2.8) | −7.6 (12.7)  | −3.8 (13.9)  | −118.8 (9.5)  |
| <b>G<sub>3</sub><sup>anti</sup></b>                             | 22.5 (1.0) | −0.8 (2.6) | 4.7 (2.8) | −12.0 (13.0) | −0.6 (14.7)  | −118.6 (9.4)  |
| <b>G<sub>1</sub><sup>syn</sup></b>                              | 20.5 (1.8) | 4.9 (2.4)  | 8.4 (2.8) | −19.5 (13.9) | −39.1 (16.7) | −128.7 (10.0) |
| <b>G<sub>2</sub><sup>syn</sup></b>                              | 22.0 (2.0) | −0.4 (4.2) | 5.5 (4.5) | −29.5 (22.1) | −15.7 (20.4) | −125.2 (14.3) |
| <b>G<sub>3</sub><sup>syn</sup></b>                              | 20.7 (1.6) | −0.6 (2.2) | 9.4 (3.1) | −34.1 (15.0) | 6.6 (13.1)   | −127.8 (13.9) |
| <b>G<sub>1</sub><sup>anti</sup>G<sub>2</sub><sup>anti</sup></b> | 22.5 (1.1) | 2.8 (2.9)  | 4.7 (2.8) | −4.2 (13.9)  | −2.9 (15.0)  | −127.1 (10.5) |
| <b>G<sub>1</sub><sup>anti</sup>G<sub>3</sub><sup>anti</sup></b> | 22.4 (1.0) | 0.9 (2.8)  | 4.9 (2.8) | −8.9 (12.9)  | −1.6 (15.1)  | −122.7 (10.1) |
| <b>G<sub>2</sub><sup>anti</sup>G<sub>3</sub><sup>anti</sup></b> | 22.0 (1.3) | 1.2 (2.8)  | 6.6 (2.9) | −12.1 (14.6) | −7.2 (15.5)  | −126.6 (10.5) |
| <b>G<sub>1</sub><sup>anti</sup>G<sub>2</sub><sup>syn</sup></b>  | 24.5 (2.7) | 1.0 (5.7)  | 3.3 (4.3) | 0.5 (21.8)   | −8.0 (22.9)  | −164.1 (41.7) |
| <b>G<sub>1</sub><sup>anti</sup>G<sub>3</sub><sup>syn</sup></b>  | 21.8 (1.9) | 0.9 (3.0)  | 7.5 (3.9) | −22.9 (22.5) | 8.7 (18.6)   | −133.4 (14.1) |
| <b>G<sub>2</sub><sup>anti</sup>G<sub>3</sub><sup>syn</sup></b>  | 23.5 (2.8) | 1.4 (3.1)  | 2.8 (6.7) | −13.8 (22.7) | −1.9 (16.5)  | −145.2 (28.7) |
| <b>G<sub>1</sub><sup>syn</sup>G<sub>2</sub><sup>anti</sup></b>  | 22.4 (3.0) | 4.5 (5.0)  | 4.6 (5.1) | −11.8 (18.0) | −4.6 (38.6)  | −147.0 (23.4) |
| <b>G<sub>1</sub><sup>syn</sup>G<sub>3</sub><sup>anti</sup></b>  | 20.7 (2.3) | 2.8 (4.0)  | 7.7 (3.9) | −21.0 (15.0) | −27.2 (26.9) | −132.1 (13.1) |
| <b>G<sub>2</sub><sup>syn</sup>G<sub>3</sub><sup>anti</sup></b>  | 20.8 (1.8) | 1.1 (2.9)  | 8.6 (2.8) | −42.6 (16.4) | −27.3 (17.0) | −124.8 (10.0) |
| <b>G<sub>1</sub><sup>syn</sup>G<sub>2</sub><sup>syn</sup></b>   | 21.2 (2.8) | 3.4 (5.3)  | 7.0 (6.0) | −17.4 (24.9) | −29.3 (32.5) | −140.0 (31.7) |
| <b>G<sub>1</sub><sup>syn</sup>G<sub>3</sub><sup>syn</sup></b>   | 21.2 (4.1) | 2.5 (4.6)  | 6.9 (8.1) | −30.3 (37.2) | −14.3 (25.1) | −152.2 (30.6) |
| <b>G<sub>2</sub><sup>syn</sup>G<sub>3</sub><sup>syn</sup></b>   | 24.6 (2.7) | 0.5 (4.4)  | 2.9 (8.2) | −20.2 (44.2) | −6.3 (20.5)  | −167.4 (32.8) |

<sup>a</sup>The pseudostep is defined in Figure F.14.

Table F.12: Relative deviations in the pseudostep parameters from undamaged DNA, with the number of Xs representing the degree of distortion.<sup>a, b</sup>

| System                 | Rise (Å) | Shift (Å) | Slide (Å) | Roll (deg.) | Tilt (deg.) | Twist (deg.) |
|------------------------|----------|-----------|-----------|-------------|-------------|--------------|
| <b>undamaged</b>       | -        | -         | -         | -           | -           | -            |
| $G_1^{anti}$           | -        | -         | -         | -           | -           | -            |
| $G_2^{anti}$           | -        | -         | -         | -           | -           | -            |
| $G_3^{anti}$           | -        | -         | -         | -           | -           | -            |
| $G_1^{syn}$            | XX       | XX        | X         | XXXX        | X           | X            |
| $G_2^{syn}$            | -        | X         | -         | X           | XX          | X            |
| $G_3^{syn}$            | -        | XX        | X         | -           | XX          | X            |
| $G_1^{anti}G_2^{anti}$ | X        | -         | -         | -           | -           | X            |
| $G_1^{anti}G_3^{anti}$ | -        | -         | -         | -           | -           | -            |
| $G_2^{anti}G_3^{anti}$ | X        | X         | -         | -           | -           | X            |
| $G_1^{anti}G_2^{syn}$  | -        | -         | -         | -           | -           | XXXX         |
| $G_1^{anti}G_3^{syn}$  | -        | XX        | -         | -           | X           | X            |
| $G_2^{anti}G_3^{syn}$  | X        | -         | -         | -           | -           | XXX          |
| $G_1^{syn}G_2^{anti}$  | XX       | -         | -         | -           | -           | XXX          |
| $G_1^{syn}G_3^{anti}$  | X        | XX        | X         | XX          | X           | X            |
| $G_2^{syn}G_3^{anti}$  | X        | XX        | -         | XX          | XXX         | -            |
| $G_1^{syn}G_2^{syn}$   | XX       | X         | -         | XXX         | -           | XX           |
| $G_1^{syn}G_3^{syn}$   | X        | X         | -         | X           | XX          | XXX          |
| $G_2^{syn}G_3^{syn}$   | -        | -         | -         | -           | X           | XXXXX        |

<sup>a</sup>The pseudostep is defined in Figure F.14. <sup>b</sup>Each X represents a 2 Å or 10° change from the value for undamaged DNA, with five Xs indicating changes > 10 Å or 50°.

Table F.13: Average (standard deviation in parentheses) solvent accessible surface area of the hydrophobic aryl bulky group (SASA, Å<sup>2</sup>) for monoadducted and diadducted DNA.

| Primary Lesion                      | Secondary Lesion <sup>a</sup>  | SASA (Å <sup>2</sup> ) | Primary Lesion                     | Secondary Lesion <sup>a</sup>  | SASA (Å <sup>2</sup> ) |
|-------------------------------------|--------------------------------|------------------------|------------------------------------|--------------------------------|------------------------|
| <b>G<sub>1</sub><sup>anti</sup></b> | Absent                         | 199.9 (8.1)            | <b>G<sub>1</sub><sup>syn</sup></b> | Absent                         | 126.8 (16.0)           |
|                                     | G <sub>2</sub> <sup>anti</sup> | 161.4 (13.5)           |                                    | G <sub>2</sub> <sup>anti</sup> | 82.7 (44.4)            |
|                                     | G <sub>2</sub> <sup>syn</sup>  | 162.9 (27.3)           |                                    | G <sub>2</sub> <sup>syn</sup>  | 99.3 (38.4)            |
|                                     | G <sub>3</sub> <sup>anti</sup> | 200.2 (8.0)            |                                    | G <sub>3</sub> <sup>anti</sup> | 100.8 (45.0)           |
|                                     | G <sub>3</sub> <sup>syn</sup>  | 198.8 (10.2)           |                                    | G <sub>3</sub> <sup>syn</sup>  | 89.6 (42.1)            |
| <b>G<sub>2</sub><sup>anti</sup></b> | Absent                         | 198.9 (6.2)            | <b>G<sub>2</sub><sup>syn</sup></b> | Absent                         | 108.1 (40.4)           |
|                                     | G <sub>1</sub> <sup>anti</sup> | 151.6 (10.4)           |                                    | G <sub>1</sub> <sup>anti</sup> | 68.1 (44.1)            |
|                                     | G <sub>1</sub> <sup>syn</sup>  | 170.8 (21.2)           |                                    | G <sub>1</sub> <sup>syn</sup>  | 94.4 (36.5)            |
|                                     | G <sub>3</sub> <sup>anti</sup> | 197.0 (11.5)           |                                    | G <sub>3</sub> <sup>anti</sup> | 120.5 (16.4)           |
|                                     | G <sub>3</sub> <sup>syn</sup>  | 196.5 (10.2)           |                                    | G <sub>3</sub> <sup>syn</sup>  | 110.4 (35.8)           |
| <b>G<sub>3</sub><sup>anti</sup></b> | Absent                         | 199.7 (8.6)            | <b>G<sub>3</sub><sup>syn</sup></b> | Absent                         | 105.0 (28.7)           |
|                                     | G <sub>1</sub> <sup>anti</sup> | 199.4 (9.0)            |                                    | G <sub>1</sub> <sup>anti</sup> | 90.1 (29.8)            |
|                                     | G <sub>1</sub> <sup>syn</sup>  | 198.6 (10.1)           |                                    | G <sub>1</sub> <sup>syn</sup>  | 85.4 (44.3)            |
|                                     | G <sub>2</sub> <sup>anti</sup> | 193.4 (19.6)           |                                    | G <sub>2</sub> <sup>anti</sup> | 92.9 (56.0)            |
|                                     | G <sub>2</sub> <sup>syn</sup>  | 199.6 (10.3)           |                                    | G <sub>2</sub> <sup>syn</sup>  | 54.4 (22.8)            |

<sup>a</sup> Absent indicates monoadducted DNA.

Table F.14: Average (standard deviation in parentheses<sup>a</sup>) lesion site interaction energies (kJ/mol) for undamaged, monoadducted and diadducted DNA.

| Primary Lesion                      | Secondary Lesion <sup>b</sup>  | LSIE <sup>avg c</sup> | Average Stacking <sup>d</sup> | Average Hydrogen Bonding <sup>e</sup> | Average Bulky Group-Backbone Interactions <sup>f</sup> |
|-------------------------------------|--------------------------------|-----------------------|-------------------------------|---------------------------------------|--------------------------------------------------------|
| <b>G<sub>1</sub><sup>g</sup></b>    | absent                         | -219.9 (3.3)          | -101.2 (2.9)                  | -118.6 (1.8)                          | N/A                                                    |
| <b>G<sub>2</sub><sup>g</sup></b>    | absent                         | -205.6 (3.7)          | -88.3 (3.1)                   | -117.4 (1.9)                          | N/A                                                    |
| <b>G<sub>3</sub><sup>g</sup></b>    | absent                         | -236.3 (3.5)          | -117.8 (3.0)                  | -118.5 (1.7)                          | N/A                                                    |
| <b>G<sub>1</sub><sup>anti</sup></b> | absent                         | -229.7 (20.6)         | -102.9 (3.0)                  | -106.0 (2.1)                          | -20.9 (11.2)                                           |
|                                     | G <sub>2</sub> <sup>anti</sup> | -250.9 (19.9)         | -126.1 (3.3)                  | -106.0 (2.1)                          | -18.8 (10.6)                                           |
|                                     | G <sub>2</sub> <sup>syn</sup>  | -240.8 (17.0)         | -139.2 (6.8)                  | -75.1 (11.0)                          | -26.4 (14.3)                                           |
|                                     | G <sub>3</sub> <sup>anti</sup> | -226.4 (72.0)         | -101.4 (3.0)                  | -105.6 (2.4)                          | -19.5 (10.7)                                           |
|                                     | G <sub>3</sub> <sup>syn</sup>  | -225.9 (36.5)         | -100.7 (3.1)                  | -105.8 (2.1)                          | -19.4 (10.6)                                           |
| <b>G<sub>2</sub><sup>anti</sup></b> | absent                         | -219.3 (16.8)         | -99.5 (3.1)                   | -106.8 (1.7)                          | -12.9 (5.9)                                            |
|                                     | G <sub>1</sub> <sup>anti</sup> | -243.9 (16.2)         | -120.1 (3.3)                  | -107.1 (1.6)                          | -16.7 (6.2)                                            |
|                                     | G <sub>1</sub> <sup>syn</sup>  | -227.8 (76.7)         | -103.4 (7.3)                  | -105.2 (2.0)                          | -19.2 (7.4)                                            |
|                                     | G <sub>3</sub> <sup>anti</sup> | -219.8 (78.9)         | -99.9 (3.2)                   | -106.6 (1.7)                          | -13.3 (5.9)                                            |
|                                     | G <sub>3</sub> <sup>syn</sup>  | -219.6 (75.2)         | -99.1 (3.3)                   | -106.7 (1.8)                          | -13.8 (6.8)                                            |
| <b>G<sub>3</sub><sup>anti</sup></b> | absent                         | -244.3 (17.1)         | -125.7 (3.1)                  | -106.0 (2.0)                          | -12.6 (6.7)                                            |
|                                     | G <sub>1</sub> <sup>anti</sup> | -243.1 (17.2)         | -125.1 (3.1)                  | -106.0 (1.9)                          | -12.0 (6.6)                                            |
|                                     | G <sub>1</sub> <sup>syn</sup>  | -244.6 (56.6)         | -125.1 (3.1)                  | -106.0 (1.9)                          | -13.4 (8.0)                                            |
|                                     | G <sub>2</sub> <sup>anti</sup> | -246.2 (19.6)         | -125.4 (3.2)                  | -106.1 (1.8)                          | -14.7 (8.4)                                            |
|                                     | G <sub>2</sub> <sup>syn</sup>  | -258.7 (38.0)         | -127.9 (3.1)                  | -106.3 (1.8)                          | -24.5 (10.5)                                           |
| <b>G<sub>1</sub><sup>syn</sup></b>  | absent                         | -26.9 (21.2)          | -15.7 (5.4)                   | 34.0 (1.6)                            | -45.1 (12.1)                                           |
|                                     | G <sub>2</sub> <sup>anti</sup> | -114.6 (50.3)         | -81.7 (12.1)                  | 11.6 (4.3)                            | -44.5 (11.3)                                           |
|                                     | G <sub>2</sub> <sup>syn</sup>  | -156.1 (48.1)         | -90.1 (13.0)                  | -14.0 (6.5)                           | -51.9 (18.9)                                           |
|                                     | G <sub>3</sub> <sup>anti</sup> | -68.3 (72.0)          | -43.3 (9.6)                   | 21.7 (4.5)                            | -46.8 (11.4)                                           |
|                                     | G <sub>3</sub> <sup>syn</sup>  | -141.2 (79.7)         | -86.8 (12.0)                  | -6.4 (7.0)                            | -48.0 (9.5)                                            |
| <b>G<sub>2</sub><sup>syn</sup></b>  | absent                         | -78.2 (16.7)          | -50.2 (10.0)                  | 15.0 (5.8)                            | -43.0 (9.6)                                            |

|                               |              |                  |              |             |              |
|-------------------------------|--------------|------------------|--------------|-------------|--------------|
|                               | $G_1^{anti}$ | -164.9<br>(20.0) | -109.2 (7.9) | -8.1 (5.2)  | -47.6 (9.3)  |
|                               | $G_1^{syn}$  | -147.0<br>(61.2) | -84.5 (14.1) | -13.2 (6.1) | -49.2 (13.7) |
|                               | $G_3^{anti}$ | -55.6 (27.6)     | -33.5 (8.4)  | 20.7 (6.4)  | -42.8 (10.0) |
|                               | $G_3^{syn}$  | -201.5<br>(16.5) | -123.3 (5.9) | -35.1 (5.2) | -43.0 (10.7) |
| <b><math>G_3^{syn}</math></b> | absent       | -92.2 (89.1)     | -50.3 (13.4) | 7.3 (7.7)   | -48.6 (12.5) |
|                               | $G_1^{anti}$ | -149.1<br>(17.8) | -87.2 (11.8) | -3.1 (6.5)  | -58.8 (15.7) |
|                               | $G_1^{syn}$  | -100.2<br>(58.2) | -57.6 (12.0) | 6.3 (6.1)   | -48.9 (11.1) |
|                               | $G_2^{anti}$ | -125.6<br>(43.0) | -69.2 (8.4)  | -10.8 (6.1) | -45.5 (12.2) |
|                               | $G_2^{syn}$  | -139.9<br>(64.8) | -74.3 (6.9)  | -12.2 (2.6) | -53.4 (9.5)  |

<sup>a</sup>Standard deviation calculated over the five replicas. <sup>b</sup>Absent indicates monoadducted DNA.

<sup>c</sup>Average of the sum of stacking and hydrogen-bonding for the adducted base, bulky moiety and opposing C as well as bulky group-backbone interactions. <sup>d</sup>Average of the stacking interactions between the adduct (base and bulky moiety) plus the opposing C and the flanking nucleobases.

<sup>e</sup>Average of the hydrogen-bonding interactions between the adduct and the pairing C. <sup>f</sup>Average of the noncovalent interactions between the bulky moiety and the sugar–phosphate backbone of nearby nucleotides. <sup>g</sup>Undamaged helices.

Table F.15: Average (standard deviation in parentheses)<sup>a</sup> C1' – C1' interstrand distance (Å) for each conformation across all replicas for diadducted DNA.

| Primary Lesion | Secondary Lesion               | C1' – C1'  | $\Delta(\text{C1}' - \text{C1}')^b$ | Primary Lesion | Secondary Lesion               | C1' – C1'        | $\Delta(\text{C1}' - \text{C1}')^b$ |
|----------------|--------------------------------|------------|-------------------------------------|----------------|--------------------------------|------------------|-------------------------------------|
| G <sup>P</sup> | G <sub>2</sub> <sup>anti</sup> | 10.8 (0.2) | 0.0                                 | G <sup>S</sup> | G <sub>2</sub> <sup>anti</sup> | 13.9 (0.9)       | 1.0                                 |
| G <sup>P</sup> | G <sub>2</sub> <sup>syn</sup>  | 10.7 (0.8) | -0.1                                | G <sup>S</sup> | G <sub>2</sub> <sup>syn</sup>  | 12.7 (1.5)       | -0.2                                |
| G <sup>P</sup> | G <sub>3</sub> <sup>anti</sup> | 10.7 (0.2) | 0.0                                 | G <sup>S</sup> | G <sub>3</sub> <sup>anti</sup> | 14.7 (1.2)       | 1.8                                 |
| G <sup>P</sup> | G <sub>3</sub> <sup>syn</sup>  | 10.7 (0.2) | 0.0                                 | G <sup>S</sup> | G <sub>3</sub> <sup>syn</sup>  | 14.5 (1.8)       | 1.5                                 |
| G <sup>B</sup> | G <sub>1</sub> <sup>anti</sup> | 10.8 (0.2) | 0.0                                 | G <sup>S</sup> | G <sub>1</sub> <sup>anti</sup> | 14.6 (1.8)       | -0.7                                |
| G <sup>B</sup> | G <sub>1</sub> <sup>syn</sup>  | 10.7 (0.2) | -0.1                                | G <sup>S</sup> | G <sub>1</sub> <sup>syn</sup>  | 13.1 (1.9)       | -2.3                                |
| G <sup>B</sup> | G <sub>3</sub> <sup>anti</sup> | 10.8 (0.2) | 0.0                                 | G <sup>S</sup> | G <sub>3</sub> <sup>anti</sup> | 10.2 (1.8)       | -5.1                                |
| G <sup>B</sup> | G <sub>3</sub> <sup>syn</sup>  | 10.8 (0.2) | 0.0                                 | G <sup>S</sup> | G <sub>3</sub> <sup>syn</sup>  | 12.9 (2.9)       | -2.4                                |
| G <sup>B</sup> | G <sub>1</sub> <sup>anti</sup> | 10.7 (0.2) | 0.0                                 | G <sup>S</sup> | G <sub>1</sub> <sup>anti</sup> | 12.9 (1.8)       | 0.2                                 |
| G <sup>B</sup> | G <sub>1</sub> <sup>syn</sup>  | 10.7 (0.2) | 0.0                                 | G <sup>S</sup> | G <sub>1</sub> <sup>syn</sup>  | 14.0 (2.0)       | 1.4                                 |
| G <sup>B</sup> | G <sub>2</sub> <sup>anti</sup> | 10.7 (0.2) | 0.0                                 | G <sup>S</sup> | G <sub>2</sub> <sup>anti</sup> | 14.8 (1.7)       | 2.2                                 |
| G <sup>B</sup> | G <sub>2</sub> <sup>syn</sup>  | 10.8 (0.2) | 0.1                                 | G <sup>S</sup> | G <sub>2</sub> <sup>syn</sup>  | 15.3 (1.0)       | 2.7                                 |
| G <sup>I</sup> | G <sub>2</sub> <sup>anti</sup> | 9.4 (1.0)  | 1.1                                 | G <sup>W</sup> | G <sub>2</sub> <sup>anti</sup> | N/A <sup>c</sup> | N/A <sup>c</sup>                    |
| G <sup>I</sup> | G <sub>2</sub> <sup>syn</sup>  | 9.8 (1.5)  | 1.5                                 | G <sup>W</sup> | G <sub>2</sub> <sup>syn</sup>  | 10.6 (1.0)       | 2.9                                 |
| G <sup>I</sup> | G <sub>3</sub> <sup>anti</sup> | 8.5 (0.8)  | 0.3                                 | G <sup>W</sup> | G <sub>3</sub> <sup>anti</sup> | 7.7 (0.6)        | 0.0                                 |
| G <sup>I</sup> | G <sub>3</sub> <sup>syn</sup>  | 8.3 (0.9)  | 0.0                                 | G <sup>W</sup> | G <sub>3</sub> <sup>syn</sup>  | 9.3 (0.8)        | 1.7                                 |
| G <sup>L</sup> | G <sub>1</sub> <sup>anti</sup> | 11.5 (1.8) | 3.1                                 | G <sup>W</sup> | G <sub>1</sub> <sup>anti</sup> | 11.3 (0.8)       | -0.2                                |
| G <sup>L</sup> | G <sub>1</sub> <sup>syn</sup>  | 8.9 (1.1)  | 0.5                                 | G <sup>W</sup> | G <sub>1</sub> <sup>syn</sup>  | 10.5 (1.4)       | -0.9                                |
| G <sup>L</sup> | G <sub>3</sub> <sup>anti</sup> | 8.2 (0.6)  | -0.3                                | G <sup>W</sup> | G <sub>3</sub> <sup>anti</sup> | 10.4 (1.0)       | -1.1                                |
| G <sup>L</sup> | G <sub>3</sub> <sup>syn</sup>  | 8.5 (0.9)  | 0.1                                 | G <sup>W</sup> | G <sub>3</sub> <sup>syn</sup>  | 9.5 (0.6)        | -1.9                                |
| G <sup>L</sup> | G <sub>1</sub> <sup>anti</sup> | 8.6 (0.8)  | -0.1                                | G <sup>W</sup> | G <sub>1</sub> <sup>anti</sup> | 8.9 (0.6)        | -0.1                                |
| G <sup>L</sup> | G <sub>1</sub> <sup>syn</sup>  | 9.8 (2.1)  | 1.1                                 | G <sup>W</sup> | G <sub>1</sub> <sup>syn</sup>  | 11.3 (0.8)       | 2.3                                 |
| G <sup>L</sup> | G <sub>2</sub> <sup>anti</sup> | 10.8 (2.9) | 2.0                                 | G <sup>W</sup> | G <sub>2</sub> <sup>anti</sup> | 11.5 (0.5)       | 2.5                                 |
| G <sup>L</sup> | G <sub>2</sub> <sup>syn</sup>  | 12.8 (1.6) | 4.1                                 | G <sup>W</sup> | G <sub>2</sub> <sup>syn</sup>  | 11.4 (0.5)       | 2.4                                 |

<sup>a</sup>Standard deviation calculated over all frames corresponding to the specified conformation. <sup>b</sup>Calculated as the in C1' – C1' distance for the lesion in diadducted DNA minus the corresponding lesion in monoadducted DNA. <sup>c</sup>Conformation has no frames.



Table F.16: Average lesion site (standard deviation in parentheses)<sup>a</sup>  $\chi$  dihedral angle (deg.) for each conformation across all replicas ( $\chi^{\text{avg}}$ ) for diadducted DNA.<sup>b</sup>

| Primary Lesion              | Secondary Lesion               | $\chi^{\text{avg}}$ | $\Delta\chi^c$ | Primary Lesion | Secondary Lesion               | $\chi^{\text{avg}}$ | $\Delta\chi^c$   |
|-----------------------------|--------------------------------|---------------------|----------------|----------------|--------------------------------|---------------------|------------------|
| G <sup>P</sup>              | G <sub>2</sub> <sup>anti</sup> | 216.9 (22.8)        | -9.4           | G <sup>S</sup> | G <sub>2</sub> <sup>anti</sup> | 59.9 (35.5)         | 9.3              |
| G <sup>P</sup>              | G <sub>2</sub> <sup>syn</sup>  | 224.7 (26.6)        | -1.6           | G <sup>S</sup> | G <sub>2</sub> <sup>syn</sup>  | 55.0 (22.1)         | 4.4              |
| G <sup>P</sup>              | G <sub>3</sub> <sup>anti</sup> | 223.5 (17.6)        | -2.8           | G <sup>S</sup> | G <sub>3</sub> <sup>anti</sup> | 108.9 (52.7)        | 58.2             |
| G <sup>P</sup>              | G <sub>3</sub> <sup>syn</sup>  | 226.3 (16.5)        | 0.0            | G <sup>S</sup> | G <sub>3</sub> <sup>syn</sup>  | 118.0 (53.6)        | 67.4             |
| G <sup>B</sup>              | G <sub>1</sub> <sup>anti</sup> | 224.8 (12.7)        | -0.4           | G <sup>S</sup> | G <sub>1</sub> <sup>anti</sup> | 174.7 (50.0)        | 10.3             |
| G <sup>B</sup>              | G <sub>1</sub> <sup>syn</sup>  | 224.0 (43.4)        | -1.2           | G <sup>S</sup> | G <sub>1</sub> <sup>syn</sup>  | 63.8 (39.4)         | -100.7           |
| G <sup>B</sup>              | G <sub>3</sub> <sup>anti</sup> | 216.8 (17.3)        | -8.4           | G <sup>S</sup> | G <sub>3</sub> <sup>anti</sup> | 47.8 (15.2)         | -116.6           |
| G <sup>B</sup>              | G <sub>3</sub> <sup>syn</sup>  | 223.1 (15.0)        | -2.1           | G <sup>S</sup> | G <sub>3</sub> <sup>syn</sup>  | 131.3 (62.8)        | -33.1            |
| G <sup>B</sup>              | G <sub>1</sub> <sup>anti</sup> | 228.8 (13.8)        | 0.0            | G <sup>S</sup> | G <sub>1</sub> <sup>anti</sup> | 46.4 (15.3)         | -11.6            |
| G <sup>B</sup>              | G <sub>1</sub> <sup>syn</sup>  | 228.9 (13.4)        | 0.1            | G <sup>S</sup> | G <sub>1</sub> <sup>syn</sup>  | 118.3 (59.5)        | 60.3             |
| G <sup>B</sup>              | G <sub>2</sub> <sup>anti</sup> | 224.9 (16.1)        | -3.9           | G <sup>S</sup> | G <sub>2</sub> <sup>anti</sup> | 173.1 (57.7)        | 115.0            |
| G <sup>B</sup>              | G <sub>2</sub> <sup>syn</sup>  | 225.6 (16.6)        | -3.2           | G <sup>S</sup> | G <sub>2</sub> <sup>syn</sup>  | 164.8 (40.3)        | 106.7            |
| G <sup>I</sup>              | G <sub>2</sub> <sup>anti</sup> | 60.2 (15.1)         | 10.5           | G <sup>W</sup> | G <sub>2</sub> <sup>anti</sup> | N/A <sup>d</sup>    | N/A <sup>d</sup> |
| G <sup>I</sup>              | G <sub>2</sub> <sup>syn</sup>  | 56.2 (17.0)         | 6.5            | G <sup>W</sup> | G <sub>2</sub> <sup>syn</sup>  | 61.7 (17.1)         | -0.1             |
| G <sup>I</sup>              | G <sub>3</sub> <sup>anti</sup> | 50.6 (15.3)         | 0.9            | G <sup>W</sup> | G <sub>3</sub> <sup>anti</sup> | 36.9 (21.1)         | -24.9            |
| G <sup>I</sup>              | G <sub>3</sub> <sup>syn</sup>  | 50.0 (21.6)         | 0.3            | G <sup>W</sup> | G <sub>3</sub> <sup>syn</sup>  | 51.0 (40.1)         | -10.8            |
| G <sub>1</sub> <sup>I</sup> | G <sub>1</sub> <sup>anti</sup> | 163.5 (75.1)        | 1 16.7         | G <sup>W</sup> | G <sub>1</sub> <sup>anti</sup> | 77.4 (58.9)         | 16.0             |
| G <sub>1</sub> <sup>I</sup> | G <sub>1</sub> <sup>syn</sup>  | 57.9 (16.6)         | 11.1           | G <sup>W</sup> | G <sub>1</sub> <sup>syn</sup>  | 58.0 (14.7)         | -3.4             |
| G <sub>1</sub> <sup>I</sup> | G <sub>3</sub> <sup>anti</sup> | 46.1 (14.0)         | -0.7           | G <sup>W</sup> | G <sub>3</sub> <sup>anti</sup> | 54.3 (18.0)         | -7.1             |
| G <sub>1</sub> <sup>I</sup> | G <sub>3</sub> <sup>syn</sup>  | 50.4 (21.5)         | 3.6            | G <sup>W</sup> | G <sub>3</sub> <sup>syn</sup>  | 68.9 (34.2)         | 7.6              |
| G <sub>1</sub> <sup>I</sup> | G <sub>1</sub> <sup>anti</sup> | 46.6 (14.6)         | -1.6           | G <sup>W</sup> | G <sub>1</sub> <sup>anti</sup> | 48.6 (13.8)         | -0.3             |
| G <sub>1</sub> <sup>I</sup> | G <sub>1</sub> <sup>syn</sup>  | 64.9 (48.2)         | 16.7           | G <sup>W</sup> | G <sub>1</sub> <sup>syn</sup>  | 63.0 (24.1)         | 14.1             |
| G <sub>1</sub> <sup>I</sup> | G <sub>2</sub> <sup>anti</sup> | 91.9 (74.0)         | 43.7           | G <sup>W</sup> | G <sub>2</sub> <sup>anti</sup> | 68.6 (53.8)         | 19.7             |
| G <sub>1</sub> <sup>I</sup> | G <sub>2</sub> <sup>syn</sup>  | 139.1 (55.1)        | 90.9           | G <sup>W</sup> | G <sub>2</sub> <sup>syn</sup>  | 171.8 (40.3)        | 122.9            |

<sup>a</sup>Standard deviation calculated over all frames corresponding to the specified conformation. <sup>b</sup> $\chi = \angle(\text{O4}'-\text{C1}'-\text{N9}-\text{C4})$ .

<sup>c</sup>Calculated as the  $\chi^{\text{avg}}$  for the lesion in diadducted DNA minus the corresponding lesion in monoadducted DNA.

<sup>d</sup>Conformation has no frames.

Table F.17: Average lesion site (standard deviation in parentheses)<sup>a</sup>  $\chi^{\text{opp}}$  ( $\chi$  for the C opposing the adduct) dihedral angle (deg.) for each conformation across all replicas for diadducted DNA.<sup>b</sup>

| Primary Lesion | Secondary Lesion               | $\chi^{\text{opp}}$ | $\Delta\chi^{\text{opp c}}$ | Primary Lesion | Secondary Lesion               | $\chi^{\text{opp}}$ | $\Delta\chi^{\text{opp c}}$ |
|----------------|--------------------------------|---------------------|-----------------------------|----------------|--------------------------------|---------------------|-----------------------------|
| G <sup>P</sup> | G <sub>2</sub> <sup>anti</sup> | 233.0 (20.2)        | -23.4                       | G <sup>S</sup> | G <sub>2</sub> <sup>anti</sup> | 243.5 (19.5)        | -3.8                        |
| G <sup>P</sup> | G <sub>2</sub> <sup>syn</sup>  | 256.4 (19.1)        | 0.0                         | G <sup>S</sup> | G <sub>2</sub> <sup>syn</sup>  | 253.0 (29.9)        | 5.7                         |
| G <sup>P</sup> | G <sub>3</sub> <sup>anti</sup> | 253.9 (19.9)        | -2.5                        | G <sup>S</sup> | G <sub>3</sub> <sup>anti</sup> | 249.5 (21.0)        | 2.1                         |
| G <sup>P</sup> | G <sub>3</sub> <sup>syn</sup>  | 254.9 (19.0)        | -1.6                        | G <sup>S</sup> | G <sub>3</sub> <sup>syn</sup>  | 246.0 (24.6)        | -1.4                        |
| G <sup>B</sup> | G <sub>1</sub> <sup>anti</sup> | 250.9 (15.4)        | 0.8                         | G <sup>S</sup> | G <sub>1</sub> <sup>anti</sup> | 247.5 (39.9)        | 0.1                         |
| G <sup>B</sup> | G <sub>1</sub> <sup>syn</sup>  | 256.8 (18.3)        | 6.7                         | G <sup>S</sup> | G <sub>1</sub> <sup>syn</sup>  | 243.7 (34.4)        | -3.7                        |
| G <sup>B</sup> | G <sub>3</sub> <sup>anti</sup> | 250.6 (15.8)        | 0.6                         | G <sup>S</sup> | G <sub>3</sub> <sup>anti</sup> | 255.3 (18.0)        | 7.9                         |
| G <sup>B</sup> | G <sub>3</sub> <sup>syn</sup>  | 251.4 (15.9)        | 1.3                         | G <sup>S</sup> | G <sub>3</sub> <sup>syn</sup>  | 246.4 (26.5)        | -1.1                        |
| G <sup>B</sup> | G <sub>1</sub> <sup>anti</sup> | 258.4 (15.2)        | -0.8                        | G <sup>S</sup> | G <sub>1</sub> <sup>anti</sup> | 241.4 (34.3)        | -3.8                        |
| G <sup>B</sup> | G <sub>1</sub> <sup>syn</sup>  | 258.6 (15.0)        | -0.6                        | G <sup>S</sup> | G <sub>1</sub> <sup>syn</sup>  | 234.7 (25.4)        | -10.5                       |
| G <sup>B</sup> | G <sub>2</sub> <sup>anti</sup> | 256.5 (16.6)        | -2.6                        | G <sup>S</sup> | G <sub>2</sub> <sup>anti</sup> | 235.8 (24.3)        | -9.5                        |
| G <sup>B</sup> | G <sub>2</sub> <sup>syn</sup>  | 250.5 (17.9)        | -8.6                        | G <sup>S</sup> | G <sub>2</sub> <sup>syn</sup>  | 237.9 (24.5)        | -7.4                        |
| G <sup>I</sup> | G <sub>2</sub> <sup>anti</sup> | 234.1 (12.1)        | -5.4                        | G <sup>W</sup> | G <sub>2</sub> <sup>anti</sup> | N/A <sup>d</sup>    | N/A <sup>d</sup>            |
| G <sup>I</sup> | G <sub>2</sub> <sup>syn</sup>  | 258.8 (51.3)        | 19.3                        | G <sup>W</sup> | G <sub>2</sub> <sup>syn</sup>  | 316.7 (81.9)        | 35.1                        |
| G <sup>I</sup> | G <sub>3</sub> <sup>anti</sup> | 240.8 (11.3)        | 1.3                         | G <sup>W</sup> | G <sub>3</sub> <sup>anti</sup> | 224.3 (3.0)         | -57.3                       |
| G <sup>I</sup> | G <sub>3</sub> <sup>syn</sup>  | 147.2 (88.2)        | -92.4                       | G <sup>W</sup> | G <sub>3</sub> <sup>syn</sup>  | 67.2 (36.1)         | -214.3                      |
| G <sup>L</sup> | G <sub>1</sub> <sup>anti</sup> | 239.8 (67.8)        | 3.4                         | G <sup>W</sup> | G <sub>1</sub> <sup>anti</sup> | 248.0 (17.9)        | -1.9                        |
| G <sup>L</sup> | G <sub>1</sub> <sup>syn</sup>  | 69.9 (81.4)         | -166.5                      | G <sup>W</sup> | G <sub>1</sub> <sup>syn</sup>  | 251.2 (91.9)        | 1.3                         |
| G <sup>L</sup> | G <sub>3</sub> <sup>anti</sup> | 239.7 (11.9)        | 3.3                         | G <sup>W</sup> | G <sub>3</sub> <sup>anti</sup> | 256.4 (16.3)        | 6.5                         |
| G <sup>L</sup> | G <sub>3</sub> <sup>syn</sup>  | 62.2 (19.6)         | -174.2                      | G <sup>W</sup> | G <sub>3</sub> <sup>syn</sup>  | 61.8 (17.9)         | -188.2                      |
| G <sup>L</sup> | G <sub>1</sub> <sup>anti</sup> | 243.7 (69.3)        | 1.3                         | G <sup>W</sup> | G <sub>1</sub> <sup>anti</sup> | 62.5 (21.3)         | -0.6                        |
| G <sup>L</sup> | G <sub>1</sub> <sup>syn</sup>  | 232.0 (22.2)        | -10.4                       | G <sup>W</sup> | G <sub>1</sub> <sup>syn</sup>  | 249.3 (19.1)        | 186.2                       |
| G <sup>L</sup> | G <sub>2</sub> <sup>anti</sup> | 233.4 (22.5)        | -9.0                        | G <sup>W</sup> | G <sub>2</sub> <sup>anti</sup> | 247.6 (17.8)        | 184.5                       |
| G <sup>L</sup> | G <sub>2</sub> <sup>syn</sup>  | 202.0 (20.8)        | -40.3                       | G <sup>W</sup> | G <sub>2</sub> <sup>syn</sup>  | 197.9 (17.8)        | 134.8                       |

<sup>a</sup>Standard deviation calculated over all frames corresponding to the specified conformation. <sup>b</sup> $\chi^{\text{opp}} = \angle(\text{O4}'\text{--C1}'\text{--N1--C2})$ . <sup>c</sup>Calculated as the  $\chi^{\text{opp}}$  for the lesion in diadducted DNA minus the corresponding lesion in monoadducted DNA. <sup>d</sup>Conformation has no frames.

Table F.18: Average lesion site (standard deviation in parentheses)<sup>a</sup>  $\theta$  dihedral angle (deg.) for each conformation across all replicas ( $\theta^{avg}$ ) for diadducted DNA.<sup>b</sup>

| Primary Lesion              | Secondary Lesion               | $\theta^{avg}$  | $\Delta\theta^c$ | Primary Lesion              | Secondary Lesion               | $\theta^{avg}$   | $\Delta\theta^c$ |
|-----------------------------|--------------------------------|-----------------|------------------|-----------------------------|--------------------------------|------------------|------------------|
| G <sup>P</sup>              | G <sub>2</sub> <sup>anti</sup> | 138.6<br>(35.4) | 9.6              | G <sup>S</sup>              | G <sub>2</sub> <sup>anti</sup> | 196.8<br>(23.6)  | -0.6             |
| G <sup>P</sup>              | G <sub>2</sub> <sup>syn</sup>  | 158.0<br>(28.3) | 28.9             | G <sup>S</sup>              | G <sub>2</sub> <sup>syn</sup>  | 163.4<br>(40.3)  | -34.0            |
| G <sup>P</sup>              | G <sub>3</sub> <sup>anti</sup> | 133.4<br>(33.8) | 4.4              | G <sup>S</sup>              | G <sub>3</sub> <sup>anti</sup> | 189.4<br>(22.0)  | -8.0             |
| G <sup>P</sup>              | G <sub>3</sub> <sup>syn</sup>  | 131.9<br>(30.6) | 2.8              | G <sup>S</sup>              | G <sub>3</sub> <sup>syn</sup>  | 182.5<br>(28.9)  | -14.9            |
| G <sup>B</sup>              | G <sub>1</sub> <sup>anti</sup> | 149.2<br>(41.9) | 6.7              | G <sup>S</sup>              | G <sub>1</sub> <sup>anti</sup> | 157.5<br>(33.6)  | -22.7            |
| G <sup>B</sup>              | G <sub>1</sub> <sup>syn</sup>  | 169.9<br>(60.0) | 27.4             | G <sup>S</sup>              | G <sub>1</sub> <sup>syn</sup>  | 150.8<br>(43.9)  | -29.4            |
| G <sup>B</sup>              | G <sub>3</sub> <sup>anti</sup> | 149.2<br>(52.8) | 6.7              | G <sup>S</sup>              | G <sub>3</sub> <sup>anti</sup> | 234.4<br>(59.3)  | 54.2             |
| G <sup>B</sup>              | G <sub>3</sub> <sup>syn</sup>  | 143.9<br>(45.0) | 1.4              | G <sup>S</sup>              | G <sub>3</sub> <sup>syn</sup>  | 211.1<br>(53.1)  | 30.9             |
| G <sup>B</sup>              | G <sub>1</sub> <sup>anti</sup> | 133.3<br>(26.3) | 0.0              | G <sup>S</sup>              | G <sub>1</sub> <sup>anti</sup> | 183.2<br>(38.8)  | -13.5            |
| G <sup>B</sup>              | G <sub>1</sub> <sup>syn</sup>  | 132.9<br>(36.9) | -0.4             | G <sup>S</sup>              | G <sub>1</sub> <sup>syn</sup>  | 181.0<br>(30.1)  | -15.7            |
| G <sup>B</sup>              | G <sub>2</sub> <sup>anti</sup> | 139.7<br>(30.2) | 6.4              | G <sup>S</sup>              | G <sub>2</sub> <sup>anti</sup> | 173.3<br>(27.0)  | -23.4            |
| G <sup>B</sup>              | G <sub>2</sub> <sup>syn</sup>  | 133.7<br>(25.9) | 0.4              | G <sup>S</sup>              | G <sub>2</sub> <sup>syn</sup>  | 175.5<br>(25.8)  | -21.2            |
| G <sup>I</sup>              | G <sub>2</sub> <sup>anti</sup> | 228.6<br>(60.1) | -25.7            | G <sup>W</sup>              | G <sub>2</sub> <sup>anti</sup> | N/A <sup>d</sup> | N/A <sup>d</sup> |
| G <sup>I</sup>              | G <sub>2</sub> <sup>syn</sup>  | 211.4<br>(68.7) | -43.0            | G <sup>W</sup>              | G <sub>2</sub> <sup>syn</sup>  | 198.9<br>(66.0)  | -27.6            |
| G <sup>I</sup>              | G <sub>3</sub> <sup>anti</sup> | 251.2<br>(59.6) | -3.2             | G <sup>W</sup>              | G <sub>3</sub> <sup>anti</sup> | 265.3<br>(18.4)  | 38.8             |
| G <sup>I</sup>              | G <sub>3</sub> <sup>syn</sup>  | 254.0<br>(26.3) | -0.4             | G <sup>W</sup>              | G <sub>3</sub> <sup>syn</sup>  | 244.9<br>(53.1)  | 18.4             |
| G <sub>1</sub> <sup>I</sup> | G <sub>1</sub> <sup>anti</sup> | 127.8<br>(50.4) | -129.1           | G <sub>2</sub> <sup>W</sup> | G <sub>1</sub> <sup>anti</sup> | 120.2<br>(79.5)  | 40.7             |
| G <sub>1</sub> <sup>I</sup> | G <sub>1</sub> <sup>syn</sup>  | 282.8<br>(74.0) | 26.0             | G <sub>2</sub> <sup>W</sup> | G <sub>1</sub> <sup>syn</sup>  | 126.0<br>(87.8)  | 46.5             |
| G <sub>1</sub> <sup>I</sup> | G <sub>3</sub> <sup>anti</sup> | 263.4<br>(23.5) | 6.5              | G <sub>2</sub> <sup>W</sup> | G <sub>3</sub> <sup>anti</sup> | 254.7<br>(82.5)  | 175.2            |
| G <sub>1</sub> <sup>I</sup> | G <sub>3</sub> <sup>syn</sup>  | 266.0<br>(44.3) | 9.2              | G <sub>2</sub> <sup>W</sup> | G <sub>3</sub> <sup>syn</sup>  | 243.3<br>(37.4)  | 163.8            |
| G <sub>3</sub> <sup>I</sup> | G <sub>1</sub> <sup>anti</sup> | 251.4<br>(41.3) | 2.9              | G <sub>3</sub> <sup>W</sup> | G <sub>1</sub> <sup>anti</sup> | 259.3<br>(24.0)  | 0.4              |
| G <sub>3</sub> <sup>I</sup> | G <sub>1</sub> <sup>syn</sup>  | 222.7<br>(72.0) | -25.8            | G <sub>3</sub> <sup>W</sup> | G <sub>1</sub> <sup>syn</sup>  | 77.4 (59.6)      | -181.5           |
| G <sub>3</sub> <sup>I</sup> | G <sub>2</sub> <sup>anti</sup> | 208.2<br>(53.2) | -40.3            | G <sub>3</sub> <sup>W</sup> | G <sub>2</sub> <sup>anti</sup> | 91.2 (61.1)      | -167.7           |
| G <sub>3</sub> <sup>I</sup> | G <sub>2</sub> <sup>syn</sup>  | 160.8<br>(26.8) | -87.8            | G <sub>3</sub> <sup>W</sup> | G <sub>2</sub> <sup>syn</sup>  | 141.0<br>(29.9)  | -117.9           |

<sup>a</sup>Standard deviation calculated over all frames corresponding to the specified conformation. <sup>b</sup> $\theta = \angle(\text{N9}-\text{C8}-\text{N10}-\text{C10})$ .

<sup>c</sup>Calculated as the  $\theta^{avg}$  for the lesion in diadducted DNA minus the corresponding lesion in monoadducted DNA.

<sup>d</sup>Conformation has no frames. .

Table F.19: Average lesion site (standard deviation in parentheses)<sup>a</sup>  $\phi$  dihedral angle (deg.) for each conformation across all replicas ( $\phi^{avg}$ ) for diadducted DNA.<sup>b</sup>

| Primary Lesion              | Secondary Lesion               | $\phi^{avg}$ | $\Delta\phi^c$ | Primary Lesion              | Secondary Lesion               | $\phi^{avg}$     | $\Delta\phi^c$   |
|-----------------------------|--------------------------------|--------------|----------------|-----------------------------|--------------------------------|------------------|------------------|
| G <sup>P</sup>              | G <sub>2</sub> <sup>anti</sup> | -0.7 (12.2)  | -0.7           | G <sub>1</sub> <sup>S</sup> | G <sub>2</sub> <sup>anti</sup> | -0.8 (11.3)      | 1.9              |
| G <sup>P</sup>              | G <sub>2</sub> <sup>syn</sup>  | 3.4 (12.2)   | 3.4            | G <sub>1</sub> <sup>S</sup> | G <sub>2</sub> <sup>syn</sup>  | -2.1 (10.8)      | 0.5              |
| G <sup>P</sup>              | G <sub>3</sub> <sup>anti</sup> | 0.3 (11.8)   | 0.3            | G <sub>1</sub> <sup>S</sup> | G <sub>3</sub> <sup>anti</sup> | 1.9 (11.7)       | 4.5              |
| G <sup>P</sup>              | G <sub>3</sub> <sup>syn</sup>  | 0.4 (11.8)   | 0.4            | G <sub>1</sub> <sup>S</sup> | G <sub>3</sub> <sup>syn</sup>  | 0.0 (12.1)       | 2.6              |
| G <sub>2</sub> <sup>P</sup> | G <sub>1</sub> <sup>anti</sup> | 5.3 (11.5)   | -0.4           | G <sub>2</sub> <sup>S</sup> | G <sub>1</sub> <sup>anti</sup> | -1.6 (11.1)      | 0.0              |
| G <sub>2</sub> <sup>P</sup> | G <sub>1</sub> <sup>syn</sup>  | 3.4 (11.8)   | -2.4           | G <sub>2</sub> <sup>S</sup> | G <sub>1</sub> <sup>syn</sup>  | -2.7 (11.2)      | -1.1             |
| G <sub>2</sub> <sup>P</sup> | G <sub>3</sub> <sup>anti</sup> | 5.6 (12.3)   | -0.1           | G <sub>2</sub> <sup>S</sup> | G <sub>3</sub> <sup>anti</sup> | 1.1 (11.3)       | 2.7              |
| G <sub>2</sub> <sup>P</sup> | G <sub>3</sub> <sup>syn</sup>  | 5.4 (12.2)   | -0.3           | G <sub>2</sub> <sup>S</sup> | G <sub>3</sub> <sup>syn</sup>  | -0.4 (11.6)      | 1.2              |
| G <sub>3</sub> <sup>P</sup> | G <sub>1</sub> <sup>anti</sup> | 2.1 (11.8)   | -0.2           | G <sub>3</sub> <sup>S</sup> | G <sub>1</sub> <sup>anti</sup> | -6.1 (11.0)      | 0.0              |
| G <sub>3</sub> <sup>P</sup> | G <sub>1</sub> <sup>syn</sup>  | 2.4 (11.8)   | 0.1            | G <sub>3</sub> <sup>S</sup> | G <sub>1</sub> <sup>syn</sup>  | -1.8 (11.3)      | 4.3              |
| G <sub>3</sub> <sup>P</sup> | G <sub>2</sub> <sup>anti</sup> | 2.7 (11.9)   | 0.5            | G <sub>3</sub> <sup>S</sup> | G <sub>2</sub> <sup>anti</sup> | -1.5 (11.1)      | 4.6              |
| G <sub>3</sub> <sup>P</sup> | G <sub>2</sub> <sup>syn</sup>  | 0.3 (12.0)   | -2.0           | G <sub>3</sub> <sup>S</sup> | G <sub>2</sub> <sup>syn</sup>  | 3.0 (11.8)       | 9.1              |
| G <sub>1</sub> <sup>I</sup> | G <sub>2</sub> <sup>anti</sup> | 1.5 (12.1)   | -1.8           | G <sub>1</sub> <sup>W</sup> | G <sub>2</sub> <sup>anti</sup> | N/A <sup>d</sup> | N/A <sup>d</sup> |
| G <sub>1</sub> <sup>I</sup> | G <sub>2</sub> <sup>syn</sup>  | 0.6 (11.1)   | -2.6           | G <sub>1</sub> <sup>W</sup> | G <sub>2</sub> <sup>syn</sup>  | 2.7 (11.4)       | -16.5            |
| G <sub>1</sub> <sup>I</sup> | G <sub>3</sub> <sup>anti</sup> | 2.7 (10.7)   | -0.5           | G <sub>1</sub> <sup>W</sup> | G <sub>3</sub> <sup>anti</sup> | -1.7 (20.4)      | -21.0            |
| G <sub>1</sub> <sup>I</sup> | G <sub>3</sub> <sup>syn</sup>  | 3.3 (10.5)   | 0.1            | G <sub>1</sub> <sup>W</sup> | G <sub>3</sub> <sup>syn</sup>  | 5.4 (11.6)       | -13.8            |
| G <sub>2</sub> <sup>I</sup> | G <sub>1</sub> <sup>anti</sup> | -1.2 (10.6)  | 0.0            | G <sub>2</sub> <sup>W</sup> | G <sub>1</sub> <sup>anti</sup> | 0.5 (11.8)       | 0.0              |
| G <sub>2</sub> <sup>I</sup> | G <sub>1</sub> <sup>syn</sup>  | -2.9 (10.3)  | -1.7           | G <sub>2</sub> <sup>W</sup> | G <sub>1</sub> <sup>syn</sup>  | -0.1 (10.9)      | -0.6             |
| G <sub>2</sub> <sup>I</sup> | G <sub>3</sub> <sup>anti</sup> | -0.4 (9.9)   | 0.9            | G <sub>2</sub> <sup>W</sup> | G <sub>3</sub> <sup>anti</sup> | 2.1 (11.5)       | 1.6              |
| G <sub>2</sub> <sup>I</sup> | G <sub>3</sub> <sup>syn</sup>  | 1.8 (10.1)   | 3.1            | G <sub>2</sub> <sup>W</sup> | G <sub>3</sub> <sup>syn</sup>  | 6.7 (10.6)       | 6.2              |
| G <sub>3</sub> <sup>I</sup> | G <sub>1</sub> <sup>anti</sup> | 3.3 (10.1)   | 0.0            | G <sub>3</sub> <sup>W</sup> | G <sub>1</sub> <sup>anti</sup> | 8.1 (10.6)       | 0.0              |
| G <sub>3</sub> <sup>I</sup> | G <sub>1</sub> <sup>syn</sup>  | 3.2 (10.5)   | -0.1           | G <sub>3</sub> <sup>W</sup> | G <sub>1</sub> <sup>syn</sup>  | 0.3 (10.3)       | -7.7             |
| G <sub>3</sub> <sup>I</sup> | G <sub>2</sub> <sup>anti</sup> | 2.4 (10.6)   | -1.0           | G <sub>3</sub> <sup>W</sup> | G <sub>2</sub> <sup>anti</sup> | -0.7 (10.3)      | -8.7             |
| G <sub>3</sub> <sup>I</sup> | G <sub>2</sub> <sup>syn</sup>  | 3.4 (12.3)   | 0.0            | G <sub>3</sub> <sup>W</sup> | G <sub>2</sub> <sup>syn</sup>  | -1.1 (11.0)      | -9.2             |

<sup>a</sup>Standard deviation calculated over all frames corresponding to the specified conformation. <sup>b</sup> $\phi = \angle(C8-N10-C10-C11)$ .

<sup>c</sup>Calculated as the  $\phi^{avg}$  for the lesion in diadducted DNA minus the corresponding lesion in monoadducted DNA.

<sup>d</sup>Conformation has no frames.

Table F.20: Average lesion site (standard deviation in parentheses)<sup>a</sup>  $\xi$  dihedral angle (deg.) for each conformation across all replicas ( $\xi_{avg}$ ) for diadducted DNA.<sup>b</sup>

| Primary Lesion       | Secondary Lesion                    | $\xi_{avg}$  | $\Delta\xi^c$ | Primary Lesion       | Secondary Lesion                    | $\xi_{avg}$      | $\Delta\xi^c$    |
|----------------------|-------------------------------------|--------------|---------------|----------------------|-------------------------------------|------------------|------------------|
| <b>G<sup>P</sup></b> | <i>G<sub>2</sub><sup>anti</sup></i> | 24.4 (27.2)  | 28.8          | <b>G<sup>S</sup></b> | <i>G<sub>2</sub><sup>anti</sup></i> | 28.8 (91.3)      | 25.9             |
| <b>G<sup>P</sup></b> | <i>G<sub>2</sub><sup>syn</sup></i>  | 4.8 (12.1)   | 9.2           | <b>G<sup>S</sup></b> | <i>G<sub>2</sub><sup>syn</sup></i>  | 24.4 (90.1)      | 21.4             |
| <b>G<sup>P</sup></b> | <i>G<sub>3</sub><sup>anti</sup></i> | -7.5 (24.4)  | -3.1          | <b>G<sup>S</sup></b> | <i>G<sub>3</sub><sup>anti</sup></i> | -13.7 (54.6)     | -16.6            |
| <b>G<sup>P</sup></b> | <i>G<sub>3</sub><sup>syn</sup></i>  | 12.0 (12.9)  | 16.3          | <b>G<sup>S</sup></b> | <i>G<sub>3</sub><sup>syn</sup></i>  | -6.3 (85.8)      | -9.3             |
| <b>G<sup>B</sup></b> | <i>G<sub>1</sub><sup>anti</sup></i> | 0.8 (14.0)   | 14.5          | <b>G<sup>S</sup></b> | <i>G<sub>1</sub><sup>anti</sup></i> | -71.6 (91.3)     | -90.4            |
| <b>G<sup>B</sup></b> | <i>G<sub>1</sub><sup>syn</sup></i>  | 9.0 (8.2)    | 22.7          | <b>G<sup>S</sup></b> | <i>G<sub>1</sub><sup>syn</sup></i>  | 44.8 (90.5)      | 26.0             |
| <b>G<sup>B</sup></b> | <i>G<sub>3</sub><sup>anti</sup></i> | 9.3 (17.3)   | 23.0          | <b>G<sup>S</sup></b> | <i>G<sub>3</sub><sup>anti</sup></i> | -0.3 (92.6)      | -19.2            |
| <b>G<sup>B</sup></b> | <i>G<sub>3</sub><sup>syn</sup></i>  | -23.5 (34.3) | -9.8          | <b>G<sup>S</sup></b> | <i>G<sub>3</sub><sup>syn</sup></i>  | 13.5 (70.6)      | -5.3             |
| <b>G<sup>B</sup></b> | <i>G<sub>1</sub><sup>anti</sup></i> | 34.3 (33.2)  | 30.9          | <b>G<sup>S</sup></b> | <i>G<sub>1</sub><sup>anti</sup></i> | 29.8 (89.2)      | 26.1             |
| <b>G<sup>B</sup></b> | <i>G<sub>1</sub><sup>syn</sup></i>  | 22.5 (27.2)  | 19.1          | <b>G<sup>S</sup></b> | <i>G<sub>1</sub><sup>syn</sup></i>  | 13.4 (90.0)      | 9.6              |
| <b>G<sup>B</sup></b> | <i>G<sub>2</sub><sup>anti</sup></i> | 4.3 (27.3)   | 0.8           | <b>G<sup>S</sup></b> | <i>G<sub>2</sub><sup>anti</sup></i> | 6.6 (87.5)       | 2.9              |
| <b>G<sup>B</sup></b> | <i>G<sub>2</sub><sup>syn</sup></i>  | -12.3 (25.0) | -15.8         | <b>G<sup>S</sup></b> | <i>G<sub>2</sub><sup>syn</sup></i>  | -7.2 (85.4)      | -10.9            |
| <b>G<sup>I</sup></b> | <i>G<sub>2</sub><sup>anti</sup></i> | -57.9 (95.1) | -87.4         | <b>G<sup>W</sup></b> | <i>G<sub>2</sub><sup>anti</sup></i> | N/A <sup>d</sup> | N/A <sup>d</sup> |
| <b>G<sup>I</sup></b> | <i>G<sub>2</sub><sup>syn</sup></i>  | -12.5 (94.7) | -41.9         | <b>G<sup>W</sup></b> | <i>G<sub>2</sub><sup>syn</sup></i>  | -6.9 (95.3)      | 19.0             |
| <b>G<sup>I</sup></b> | <i>G<sub>3</sub><sup>anti</sup></i> | -5.8 (95.4)  | -35.2         | <b>G<sup>W</sup></b> | <i>G<sub>3</sub><sup>anti</sup></i> | 24.8 (2.8)       | 50.7             |
| <b>G<sup>I</sup></b> | <i>G<sub>3</sub><sup>syn</sup></i>  | -14.5 (93.5) | -43.9         | <b>G<sup>W</sup></b> | <i>G<sub>3</sub><sup>syn</sup></i>  | 8.9 (94.5)       | 34.7             |
| <b>G<sup>L</sup></b> | <i>G<sub>1</sub><sup>anti</sup></i> | -21.6 (93.0) | -6.1          | <b>G<sup>W</sup></b> | <i>G<sub>1</sub><sup>anti</sup></i> | -15.9 (95.0)     | 38.0             |
| <b>G<sup>L</sup></b> | <i>G<sub>1</sub><sup>syn</sup></i>  | -38.1 (93.1) | -22.7         | <b>G<sup>W</sup></b> | <i>G<sub>1</sub><sup>syn</sup></i>  | -65.4 (95.5)     | -11.4            |
| <b>G<sup>L</sup></b> | <i>G<sub>3</sub><sup>anti</sup></i> | -56.7 (93.1) | -41.3         | <b>G<sup>W</sup></b> | <i>G<sub>3</sub><sup>anti</sup></i> | 3.3 (95.7)       | 57.2             |
| <b>G<sup>L</sup></b> | <i>G<sub>3</sub><sup>syn</sup></i>  | 0.2 (94.1)   | 15.6          | <b>G<sup>W</sup></b> | <i>G<sub>3</sub><sup>syn</sup></i>  | 12.4 (94.6)      | 66.4             |
| <b>G<sup>L</sup></b> | <i>G<sub>1</sub><sup>anti</sup></i> | -15.4 (91.9) | 2.1           | <b>G<sup>W</sup></b> | <i>G<sub>1</sub><sup>anti</sup></i> | 21.4 (93.1)      | 41.2             |
| <b>G<sup>L</sup></b> | <i>G<sub>1</sub><sup>syn</sup></i>  | -28.8 (93.8) | -11.3         | <b>G<sup>W</sup></b> | <i>G<sub>1</sub><sup>syn</sup></i>  | -6.1 (92.5)      | 13.7             |
| <b>G<sup>L</sup></b> | <i>G<sub>2</sub><sup>anti</sup></i> | -45.2 (92.2) | -27.7         | <b>G<sup>W</sup></b> | <i>G<sub>2</sub><sup>anti</sup></i> | 9.6 (95.3)       | 29.4             |
| <b>G<sup>L</sup></b> | <i>G<sub>2</sub><sup>syn</sup></i>  | -62.8 (83.2) | -45.3         | <b>G<sup>W</sup></b> | <i>G<sub>2</sub><sup>syn</sup></i>  | -1.1 (11.0)      | -9.2             |

<sup>a</sup>Standard deviation calculated over all frames corresponding to the specified conformation. <sup>b</sup> $\xi = \angle(\text{C12-C13-C16-C17})$ . <sup>c</sup>Calculated as the  $\xi_{avg}$  for the lesion in diadducted DNA minus the corresponding lesion in monoadducted DNA.

<sup>d</sup>Conformation has no frames.

Table F.21: Average (standard deviation in parentheses) interplanar angle (deg.) between the two rings in the biphenyl moiety for each conformation and across all replicas for diadducted DNA.<sup>a</sup>

| Primary Lesion | Secondary Lesion  | Interplanar Angle | $\Delta$ Interplanar Angle <sup>b</sup> | Primary Lesion | Secondary Lesion  | Interplanar Angle | $\Delta$ Interplanar Angle <sup>b</sup> |
|----------------|-------------------|-------------------|-----------------------------------------|----------------|-------------------|-------------------|-----------------------------------------|
| G <sup>P</sup> | G <sup>anti</sup> | 32.4 (11.4)       | -1.3                                    | G <sup>S</sup> | G <sup>anti</sup> | 24.0 (10.1)       | -3.6                                    |
| G <sup>P</sup> | G <sup>2n</sup>   | 28.5 (11.1)       | -5.1                                    | G <sup>S</sup> | G <sup>2n</sup>   | 30.9 (10.3)       | 3.3                                     |
| G <sup>P</sup> | G <sup>3n</sup>   | 33.4 (11.4)       | -0.2                                    | G <sup>S</sup> | G <sup>3n</sup>   | 23.7 (9.7)        | -3.9                                    |
| G <sup>P</sup> | G <sup>1n</sup>   | 28.9 (10.7)       | -4.7                                    | G <sup>S</sup> | G <sup>1n</sup>   | 24.3 (10.2)       | -3.3                                    |
| G <sup>B</sup> | G <sup>anti</sup> | 32.4 (11.4)       | -1.0                                    | G <sup>S</sup> | G <sup>anti</sup> | 28.4 (10.5)       | 3.0                                     |
| G <sup>B</sup> | G <sup>2n</sup>   | 30.4 (11.5)       | -3.0                                    | G <sup>S</sup> | G <sup>2n</sup>   | 33.4 (10.9)       | 7.9                                     |
| G <sup>B</sup> | G <sup>3n</sup>   | 33.5 (11.5)       | 0.2                                     | G <sup>S</sup> | G <sup>3n</sup>   | 31.3 (12.7)       | 5.8                                     |
| G <sup>B</sup> | G <sup>1n</sup>   | 29.6 (11.0)       | -3.8                                    | G <sup>S</sup> | G <sup>1n</sup>   | 27.7 (11.2)       | 2.2                                     |
| G <sup>B</sup> | G <sup>anti</sup> | 33.4 (11.4)       | -0.1                                    | G <sup>S</sup> | G <sup>anti</sup> | 28.3 (10.5)       | 0.0                                     |
| G <sup>B</sup> | G <sup>2n</sup>   | 33.4 (11.4)       | -0.1                                    | G <sup>S</sup> | G <sup>2n</sup>   | 26.8 (10.4)       | -1.5                                    |
| G <sup>B</sup> | G <sup>3n</sup>   | 33.5 (11.5)       | 0.0                                     | G <sup>S</sup> | G <sup>3n</sup>   | 27.2 (10.2)       | -1.2                                    |
| G <sup>B</sup> | G <sup>1n</sup>   | 33.5 (11.4)       | 0.0                                     | G <sup>S</sup> | G <sup>1n</sup>   | 26.7 (10.2)       | -1.7                                    |
| G <sup>I</sup> | G <sup>anti</sup> | 31.3 (11.5)       | 0.7                                     | G <sup>W</sup> | G <sup>anti</sup> | N/A <sup>c</sup>  | N/A <sup>c</sup>                        |
| G <sup>I</sup> | G <sup>2n</sup>   | 31.1 (11.3)       | 0.5                                     | G <sup>W</sup> | G <sup>2n</sup>   | 32.6 (11.5)       | -1.6                                    |
| G <sup>I</sup> | G <sup>3n</sup>   | 30.8 (11.5)       | 0.2                                     | G <sup>W</sup> | G <sup>3n</sup>   | 29.1 (0.5)        | -5.0                                    |
| G <sup>I</sup> | G <sup>1n</sup>   | 29.9 (11.1)       | -0.7                                    | G <sup>W</sup> | G <sup>1n</sup>   | 31.0 (11.4)       | -3.1                                    |
| G <sup>I</sup> | G <sup>anti</sup> | 28.5 (12.2)       | 0.6                                     | G <sup>W</sup> | G <sup>anti</sup> | 32.1 (11.4)       | -1.9                                    |
| G <sup>I</sup> | G <sup>2n</sup>   | 34.6 (11.3)       | 6.7                                     | G <sup>W</sup> | G <sup>2n</sup>   | 33.9 (11.5)       | 0.0                                     |
| G <sup>I</sup> | G <sup>3n</sup>   | 29.7 (11.0)       | 1.8                                     | G <sup>W</sup> | G <sup>3n</sup>   | 32.7 (12.1)       | -1.3                                    |
| G <sup>I</sup> | G <sup>1n</sup>   | 30.5 (10.6)       | 2.6                                     | G <sup>W</sup> | G <sup>1n</sup>   | 31.4 (10.2)       | -2.5                                    |
| G <sup>I</sup> | G <sup>anti</sup> | 29.5 (11.2)       | -0.1                                    | G <sup>W</sup> | G <sup>anti</sup> | 32.2 (10.5)       | 0.4                                     |
| G <sup>I</sup> | G <sup>2n</sup>   | 30.3 (11.1)       | 0.6                                     | G <sup>W</sup> | G <sup>2n</sup>   | 33.4 (11.1)       | 1.6                                     |
| G <sup>I</sup> | G <sup>3n</sup>   | 30.1 (11.3)       | 0.5                                     | G <sup>W</sup> | G <sup>3n</sup>   | 32.8 (11.2)       | 1.0                                     |
| G <sup>I</sup> | G <sup>1n</sup>   | 31.8 (9.7)        | 2.1                                     | G <sup>W</sup> | G <sup>1n</sup>   | 29.5 (10.9)       | -2.3                                    |

<sup>a</sup>Standard deviation calculated over all frames corresponding to the specified conformation. <sup>b</sup>Calculated as the interplanar angle for the lesion in diadducted DNA minus monoadducted DNA. <sup>c</sup>Conformation has no frames.

Table F.22: Occupancies (%) of the non-Watson-Crick hydrogen bonds at the (*anti*) lesion site of diadducted DNA in the B conformation.<sup>a,b</sup>

| Primary Lesion              | Secondary Lesion               | Hydrogen Bond         |                        |                       |
|-----------------------------|--------------------------------|-----------------------|------------------------|-----------------------|
|                             |                                | (G6)N10–H ... O4'(G6) | (G6)N10–H ... O3'(C5)  |                       |
| G <sub>1</sub> <sup>B</sup> | Absent                         | 41.9%                 | 34.2%                  |                       |
| G <sub>1</sub> <sup>B</sup> | G <sub>2</sub> <sup>anti</sup> | 48.4%                 | 21.7%                  |                       |
| G <sub>1</sub> <sup>B</sup> | G <sub>2</sub> <sup>syn</sup>  | 60.2%                 | 0.0%                   |                       |
| G <sub>1</sub> <sup>B</sup> | G <sub>3</sub> <sup>anti</sup> | 45.4%                 | 28.1%                  |                       |
| G <sub>1</sub> <sup>B</sup> | G <sub>3</sub> <sup>syn</sup>  | 47.2%                 | 26.7%                  |                       |
|                             |                                | (G7)N10–H ... O4'(G7) | (C27)N4–H ... OP2(C26) | (G7)N10–H ... O4'(C8) |
| G <sub>2</sub> <sup>B</sup> | Absent                         | 74.1%                 | 0.0%                   | 0.0%                  |
| G <sub>2</sub> <sup>B</sup> | G <sub>1</sub> <sup>anti</sup> | 71.3%                 | 0.0%                   | 0.0%                  |
| G <sub>2</sub> <sup>B</sup> | G <sub>1</sub> <sup>syn</sup>  | 56.4%                 | 12.7%                  | 0.0%                  |
| G <sub>2</sub> <sup>B</sup> | G <sub>3</sub> <sup>anti</sup> | 70.3%                 | 0.0%                   | 9.1%                  |
| G <sub>2</sub> <sup>B</sup> | G <sub>3</sub> <sup>syn</sup>  | 71.7%                 | 0.0%                   | 0.0%                  |
|                             |                                | (G9)N10–H ... O4'(G9) | (G9)N10–H ... O3(C8)   |                       |
| G <sub>3</sub> <sup>B</sup> | Absent                         | 66.0%                 | 5.7%                   |                       |
| G <sub>3</sub> <sup>B</sup> | G <sub>1</sub> <sup>anti</sup> | 66.2%                 | 0.0%                   |                       |
| G <sub>3</sub> <sup>B</sup> | G <sub>1</sub> <sup>syn</sup>  | 64.4%                 | 10.3%                  |                       |
| G <sub>3</sub> <sup>B</sup> | G <sub>2</sub> <sup>anti</sup> | 64.5%                 | 8.6%                   |                       |
| G <sub>3</sub> <sup>B</sup> | G <sub>2</sub> <sup>syn</sup>  | 49.0%                 | 41.9%                  |                       |

<sup>a</sup>Nucleotide numbering is illustrated in Figure F.14. <sup>b</sup>Occupancies determined as the percentage of the simulation that a hydrogen bonds exist according to a heavy atom distance < 3.2 Å and a heavy atom–hydrogen–heavy atom angle > 120°.

Table F.23: Average (standard deviation in parentheses)<sup>a</sup> SASA of the hydrophobic bulky group (Å<sup>2</sup>) for each conformation across all replicas for diadducted DNA.

| Primary Lesion              | Secondary Lesion               | SASA <sup>avg</sup> | $\Delta(\text{SASA}^{\text{avg}})^{\text{b}}$ | Primary Lesion              | Secondary Lesion               | SASA <sup>avg</sup> | $\Delta(\text{SASA}^{\text{avg}})^{\text{b}}$ |
|-----------------------------|--------------------------------|---------------------|-----------------------------------------------|-----------------------------|--------------------------------|---------------------|-----------------------------------------------|
| G <sup>P</sup> <sub>1</sub> | G <sup>anti</sup> <sub>2</sub> | 161.4 (13.5)        | -38.5                                         | G <sup>S</sup> <sub>1</sub> | G <sup>anti</sup> <sub>2</sub> | 52.9 (13.7)         | -14.1                                         |
| G <sup>P</sup> <sub>1</sub> | G <sup>syn</sup> <sub>2</sub>  | 162.9 (27.3)        | -37.0                                         | G <sup>S</sup> <sub>1</sub> | G <sup>syn</sup> <sub>2</sub>  | 66.2 (25.2)         | -0.8                                          |
| G <sup>P</sup> <sub>1</sub> | G <sup>anti</sup> <sub>3</sub> | 200.2 (8.0)         | 0.3                                           | G <sup>S</sup> <sub>1</sub> | G <sup>anti</sup> <sub>3</sub> | 46.3 (17.8)         | -20.7                                         |
| G <sup>P</sup> <sub>1</sub> | G <sup>syn</sup> <sub>3</sub>  | 198.8 (10.2)        | -1.2                                          | G <sup>S</sup> <sub>1</sub> | G <sup>syn</sup> <sub>3</sub>  | 47.0 (16.6)         | -20.0                                         |
| G <sup>B</sup> <sub>2</sub> | G <sup>anti</sup> <sub>1</sub> | 151.6 (10.4)        | -47.3                                         | G <sup>S</sup> <sub>2</sub> | G <sup>anti</sup> <sub>1</sub> | 54.9 (30.7)         | 3.4                                           |
| G <sup>B</sup> <sub>2</sub> | G <sup>syn</sup> <sub>1</sub>  | 170.8 (21.2)        | -28.1                                         | G <sup>S</sup> <sub>2</sub> | G <sup>syn</sup> <sub>1</sub>  | 73.7 (34.7)         | 22.3                                          |
| G <sup>B</sup> <sub>2</sub> | G <sup>anti</sup> <sub>3</sub> | 197.0 (11.5)        | -1.8                                          | G <sup>S</sup> <sub>2</sub> | G <sup>anti</sup> <sub>3</sub> | 105.2 (21.3)        | 53.8                                          |
| G <sup>B</sup> <sub>2</sub> | G <sup>syn</sup> <sub>3</sub>  | 196.5 (10.2)        | -2.3                                          | G <sup>S</sup> <sub>2</sub> | G <sup>syn</sup> <sub>3</sub>  | 89.1 (51.5)         | 37.6                                          |
| G <sup>B</sup> <sub>3</sub> | G <sup>anti</sup> <sub>1</sub> | 199.4 (9.0)         | -0.3                                          | G <sup>S</sup> <sub>3</sub> | G <sup>anti</sup> <sub>1</sub> | 74.2 (23.0)         | 1.4                                           |
| G <sup>B</sup> <sub>3</sub> | G <sup>syn</sup> <sub>1</sub>  | 198.6 (10.1)        | -1.1                                          | G <sup>S</sup> <sub>3</sub> | G <sup>syn</sup> <sub>1</sub>  | 57.5 (22.7)         | -15.3                                         |
| G <sup>B</sup> <sub>3</sub> | G <sup>anti</sup> <sub>2</sub> | 193.4 (19.6)        | -6.3                                          | G <sup>S</sup> <sub>3</sub> | G <sup>anti</sup> <sub>2</sub> | 48.7 (21.6)         | -24.1                                         |
| G <sup>B</sup> <sub>3</sub> | G <sup>syn</sup> <sub>2</sub>  | 199.6 (10.3)        | 0.0                                           | G <sup>S</sup> <sub>3</sub> | G <sup>syn</sup> <sub>2</sub>  | 47.2 (16.4)         | -25.5                                         |
| G <sup>I</sup> <sub>1</sub> | G <sup>anti</sup> <sub>2</sub> | 131.6 (31.6)        | 4.0                                           | G <sup>W</sup> <sub>1</sub> | G <sup>anti</sup> <sub>2</sub> | N/A <sup>c</sup>    | N/A <sup>c</sup>                              |
| G <sup>I</sup> <sub>1</sub> | G <sup>syn</sup> <sub>2</sub>  | 102.4 (26.9)        | -25.3                                         | G <sup>W</sup> <sub>1</sub> | G <sup>syn</sup> <sub>2</sub>  | 124.5 (28.4)        | 2.5                                           |
| G <sup>I</sup> <sub>1</sub> | G <sup>anti</sup> <sub>3</sub> | 131.2 (21.4)        | 3.6                                           | G <sup>W</sup> <sub>1</sub> | G <sup>anti</sup> <sub>3</sub> | 102.2 (1.2)         | -19.8                                         |
| G <sup>I</sup> <sub>1</sub> | G <sup>syn</sup> <sub>3</sub>  | 123.5 (18.6)        | -4.1                                          | G <sup>W</sup> <sub>1</sub> | G <sup>syn</sup> <sub>3</sub>  | 121.1 (17.4)        | -0.9                                          |
| G <sup>I</sup> <sub>2</sub> | G <sup>anti</sup> <sub>1</sub> | 76.5 (47.0)         | -42.7                                         | G <sup>W</sup> <sub>2</sub> | G <sup>anti</sup> <sub>1</sub> | 139.2 (40.2)        | -22.5                                         |
| G <sup>I</sup> <sub>2</sub> | G <sup>syn</sup> <sub>1</sub>  | 96.2 (24.8)         | -23.0                                         | G <sup>W</sup> <sub>2</sub> | G <sup>syn</sup> <sub>1</sub>  | 125.3 (32.1)        | -36.4                                         |
| G <sup>I</sup> <sub>2</sub> | G <sup>anti</sup> <sub>3</sub> | 117.2 (10.9)        | -2.1                                          | G <sup>W</sup> <sub>2</sub> | G <sup>anti</sup> <sub>3</sub> | 144.7 (22.0)        | -17.0                                         |
| G <sup>I</sup> <sub>2</sub> | G <sup>syn</sup> <sub>3</sub>  | 116.2 (14.9)        | -3.0                                          | G <sup>W</sup> <sub>2</sub> | G <sup>syn</sup> <sub>3</sub>  | 125.5 (13.2)        | -36.2                                         |
| G <sup>I</sup> <sub>3</sub> | G <sup>anti</sup> <sub>1</sub> | 119.7 (18.9)        | -3.4                                          | G <sup>W</sup> <sub>3</sub> | G <sup>anti</sup> <sub>1</sub> | 115.0 (14.0)        | 0.7                                           |
| G <sup>I</sup> <sub>3</sub> | G <sup>syn</sup> <sub>1</sub>  | 117.0 (41.1)        | -6.1                                          | G <sup>W</sup> <sub>3</sub> | G <sup>syn</sup> <sub>1</sub>  | 153.9 (29.5)        | 39.6                                          |
| G <sup>I</sup> <sub>3</sub> | G <sup>anti</sup> <sub>2</sub> | 94.6 (42.0)         | -28.6                                         | G <sup>W</sup> <sub>3</sub> | G <sup>anti</sup> <sub>2</sub> | 156.0 (36.5)        | 41.7                                          |
| G <sup>I</sup> <sub>3</sub> | G <sup>syn</sup> <sub>2</sub>  | 73.1 (24.9)         | -50.1                                         | G <sup>W</sup> <sub>3</sub> | G <sup>syn</sup> <sub>2</sub>  | 101.2 (16.4)        | -13.1                                         |

<sup>a</sup>Standard deviation calculated over all frames corresponding to the specified conformation. <sup>b</sup>Calculated as the SASA<sup>avg</sup> for the lesion in diadducted DNA minus monoadducted DNA. <sup>c</sup>Conformation has no frames.



Table F.24: Occupancies (%) of the non-Watson-Crick hydrogen bonds at the (*syn*) lesion site of diadducted DNA in the I conformation.<sup>a,b</sup>

| Primary Lesion              | Secondary Lesion               | Hydrogen Bond             |                         |                         |                         |                         |
|-----------------------------|--------------------------------|---------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
|                             |                                | (G6)N10–H...<br>O4'(G/G7) | (G6)N10–H...<br>N1(G28) | (G6)N10–H...<br>N2(G28) | (G6)N10–H...<br>N3(G28) |                         |
| G <sub>1</sub> <sup>I</sup> | Absent                         | 0.0%                      | 19.8%                   | 57.4%                   | 19.2%                   |                         |
| G <sub>1</sub> <sup>I</sup> | G <sub>2</sub> <sup>anti</sup> | 5.5%                      | 6.7%                    | 36.6%                   | 7.0%                    |                         |
| G <sub>1</sub> <sup>I</sup> | G <sub>2</sub> <sup>syn</sup>  | 30.4%                     | 7.2%                    | 20.5%                   | 7.8%                    |                         |
| G <sub>1</sub> <sup>I</sup> | G <sub>3</sub> <sup>anti</sup> | 7.0%                      | 15.8%                   | 50.3%                   | 17.5%                   |                         |
| G <sub>1</sub> <sup>I</sup> | G <sub>3</sub> <sup>syn</sup>  | 0.0%                      | 15.9%                   | 56.9%                   | 25.6%                   |                         |
|                             |                                | (G7)N10–H...<br>N2(G/G6)  | (G7)N10–H...<br>O4'(G7) | (G7)N10–H...<br>O4'(C8) | (G7)N10–H...<br>O2(C27) | (C27)N4–H...<br>N4(C26) |
| G <sub>2</sub> <sup>I</sup> | Absent                         | 30.0%                     | 0.0%                    | 0.0%                    | 19.5%                   | 0.0%                    |
| G <sub>2</sub> <sup>I</sup> | G <sub>1</sub> <sup>anti</sup> | 0.0%                      | 36.0%                   | 23.3%                   | 0.0%                    | 0.0%                    |
| G <sub>2</sub> <sup>I</sup> | G <sub>1</sub> <sup>syn</sup>  | 0.0%                      | 0.0%                    | 10.6%                   | 0.0%                    | 24.7%                   |
| G <sub>2</sub> <sup>I</sup> | G <sub>3</sub> <sup>anti</sup> | 22.9%                     | 0.0%                    | 0.0%                    | 20.0%                   | 0.0%                    |
| G <sub>2</sub> <sup>I</sup> | G <sub>3</sub> <sup>syn</sup>  | 33.8%                     | 0.0%                    | 0.0%                    | 8.8%                    | 15.5%                   |
|                             |                                | (G9)N10–H...O4'(G9)       | (G9)N10–H...N2(G25)     | (G9)N10–H...N3(G25)     |                         |                         |
| G <sub>3</sub> <sup>I</sup> | Absent                         | 0.0%                      | 36.6%                   | 27.4%                   |                         |                         |
| G <sub>3</sub> <sup>I</sup> | G <sub>1</sub> <sup>anti</sup> | 0.0%                      | 40.9%                   | 34.0%                   |                         |                         |
| G <sub>3</sub> <sup>I</sup> | G <sub>1</sub> <sup>syn</sup>  | 15.5%                     | 26.3%                   | 20.8%                   |                         |                         |
| G <sub>3</sub> <sup>I</sup> | G <sub>2</sub> <sup>anti</sup> | 36.4%                     | 24.7%                   | 17.3%                   |                         |                         |
| G <sub>3</sub> <sup>I</sup> | G <sub>2</sub> <sup>syn</sup>  | 59.6%                     | 0.0%                    | 0.0%                    |                         |                         |

<sup>a</sup>Nucleotide numbering is illustrated in Figure F.14. <sup>b</sup>Occupancies determined as the percentage of the simulation that a hydrogen bonds exist according to a heavy atom distance < 3.2 Å and a heavy atom–hydrogen–heavy atom angle > 120°.

Table F.25: Occupancies (%) of the non-Watson-Crick hydrogen bonds at the (*syn*) lesion site of diadducted DNA in the S conformation.<sup>a,b</sup>

| Primary Lesion              | Secondary Lesion               | Hydrogen Bond                |                               |                              |                              |                            |                             |                               |                              |                               |
|-----------------------------|--------------------------------|------------------------------|-------------------------------|------------------------------|------------------------------|----------------------------|-----------------------------|-------------------------------|------------------------------|-------------------------------|
|                             |                                | (G6)N10–H...O4'(G6)          |                               | (G6)N10–H...O4'(G/G7)        |                              | (C27)N4–H...N3(G25)        |                             | (C27)N4–H...OP2(C26)          |                              |                               |
| G <sub>1</sub> <sup>S</sup> | Absent                         | 0.0%                         |                               | 0.0%                         |                              | 0.0%                       |                             | 36.0%                         |                              |                               |
| G <sub>1</sub> <sup>S</sup> | G <sub>2</sub> <sup>anti</sup> | 16.2%                        |                               | 0.0%                         |                              | 0.0%                       |                             | 22.0%                         |                              |                               |
| G <sub>1</sub> <sup>S</sup> | G <sub>2</sub> <sup>syn</sup>  | 0.0%                         |                               | 50.5%                        |                              | 30.5%                      |                             | 0.0%                          |                              |                               |
| G <sub>1</sub> <sup>S</sup> | G <sub>3</sub> <sup>anti</sup> | 55.3%                        |                               | 0.0%                         |                              | 0.0%                       |                             | 48.6%                         |                              |                               |
| G <sub>1</sub> <sup>S</sup> | G <sub>3</sub> <sup>syn</sup>  | 60.7%                        |                               | 0.0%                         |                              | 0.0%                       |                             | 48.2%                         |                              |                               |
|                             |                                | (G6)N10–<br>H ...<br>O4'(G7) | (G7)N10–<br>H ...<br>N2(G/G6) | (G7)N10–<br>H ...<br>O4'(G7) | (G7)N10–<br>H ...<br>O4'(C8) | (C8)N4–H<br>...<br>OP2(G7) | (C26)N4–<br>H ...<br>O2(C8) | (C26)N4–<br>H ...<br>OP2(G25) | (C27)N4–<br>H ...<br>O2(C26) | (C27)N4–<br>H ...<br>OP2(C26) |
| G <sub>2</sub> <sup>S</sup> | Absent                         | 0.0%                         | 0.0%                          | 84.2%                        | 0.0%                         | 0.0%                       | 0.0%                        | 62.7%                         | 46.6%                        | 0.0%                          |
| G <sub>2</sub> <sup>S</sup> | G <sub>1</sub> <sup>anti</sup> | 0.0%                         | 0.0%                          | 72.3%                        | 7.2%                         | 0.0%                       | 0.0%                        | 49.3%                         | 25.2%                        | 0.0%                          |
| G <sub>2</sub> <sup>S</sup> | G <sub>1</sub> <sup>syn</sup>  | 42.7%                        | 0.0%                          | 14.2%                        | 51.7%                        | 10.8%                      | 0.0%                        | 65.1%                         | 0.0%                         | 0.0%                          |
| G <sub>2</sub> <sup>S</sup> | G <sub>3</sub> <sup>anti</sup> | 0.0%                         | 20.4%                         | 0.0%                         | 7.5%                         | 6.5%                       | 28.2%                       | 24.4%                         | 15.5%                        | 29.9%                         |
| G <sub>2</sub> <sup>S</sup> | G <sub>3</sub> <sup>syn</sup>  | 0.0%                         | 8.5%                          | 56.2%                        | 0.0%                         | 10.0%                      | 0.0%                        | 63.9%                         | 44.0%                        | 0.0%                          |
|                             |                                | (G9)N10–H...O4'(G9)          |                               | (C23)N2–H...O2(C24)          |                              | (C23)N2–H...N3(C24)        |                             | (C24)N4–H...OP2(G23)          |                              |                               |
| G <sub>3</sub> <sup>S</sup> | Absent                         | 14.1%                        |                               | 14.4%                        |                              | 14.0%                      |                             | 10.9%                         |                              |                               |
| G <sub>3</sub> <sup>S</sup> | G <sub>1</sub> <sup>anti</sup> | 0.0%                         |                               | 21.2%                        |                              | 17.3%                      |                             | 0.0%                          |                              |                               |
| G <sub>3</sub> <sup>S</sup> | G <sub>1</sub> <sup>syn</sup>  | 50.2%                        |                               | 10.6%                        |                              | 16.0%                      |                             | 30.9%                         |                              |                               |
| G <sub>3</sub> <sup>S</sup> | G <sub>2</sub> <sup>anti</sup> | 74.0%                        |                               | 0.0%                         |                              | 5.7%                       |                             | 49.7%                         |                              |                               |
| G <sub>3</sub> <sup>S</sup> | G <sub>2</sub> <sup>syn</sup>  | 82.1%                        |                               | 0.0%                         |                              | 0.0%                       |                             | 53.4%                         |                              |                               |

<sup>a</sup>Nucleotide numbering is illustrated in Figure F.14. <sup>b</sup>Occupancies determined as the percentage of the simulation that a hydrogen bonds exist according to a heavy atom distance < 3.2 Å and a heavy atom–hydrogen–heavy atom angle > 120°.

Table F.26: Occupancies (%) of the non-Watson-Crick hydrogen bonds at the (*syn*) lesion site of diadducted DNA in the W conformation.<sup>a,b</sup>

| Primar<br>y<br>Lesion       |                                | Secondary<br>Lesion | Hydrogen Bond                 |                              |                             |                              |                             |                             |                              |                            |                            |
|-----------------------------|--------------------------------|---------------------|-------------------------------|------------------------------|-----------------------------|------------------------------|-----------------------------|-----------------------------|------------------------------|----------------------------|----------------------------|
|                             |                                |                     | (G6)N10–H<br>...<br>O4'(G/G7) | (G6)N10–H<br>...<br>N2(G28)  | (G6)N10–H<br>...<br>N3(G28) | (C26)N4–H<br>...<br>O6(G6)   | (C27)N4–H<br>...<br>O6(G6)  | (C27)N4–H<br>...<br>N7(G6)  | (G28)N2–H<br>...<br>O4'(G6)  |                            |                            |
| G <sub>1</sub> <sup>W</sup> | Absent                         |                     | 0.0%                          | 100.0%                       | 0.0%                        | 100.0%                       | 100.0%                      | 0.0%                        | 100.0%                       |                            |                            |
| G <sub>1</sub> <sup>W</sup> | G <sub>2</sub> <sup>anti</sup> |                     | N/A <sup>c</sup>              | N/A <sup>c</sup>             | N/A <sup>c</sup>            | N/A <sup>c</sup>             | N/A <sup>c</sup>            | N/A <sup>c</sup>            | N/A <sup>c</sup>             |                            |                            |
| G <sub>1</sub> <sup>W</sup> | G <sub>2</sub> <sup>syn</sup>  |                     | 31.4%                         | 27.5%                        | 0.0%                        | 11.0%                        | 53.3%                       | 45.3%                       | 0.0%                         |                            |                            |
| G <sub>1</sub> <sup>W</sup> | G <sub>3</sub> <sup>anti</sup> |                     | 0.0%                          | 50.0%                        | 50.0%                       | 0.0%                         | 0.0%                        | 0.0%                        | 0.0%                         |                            |                            |
| G <sub>1</sub> <sup>W</sup> | G <sub>3</sub> <sup>syn</sup>  |                     | 0.0%                          | 49.2%                        | 16.7%                       | 0.0%                         | 50.5%                       | 49.0%                       | 0.0%                         |                            |                            |
|                             |                                |                     | (G6)N10–H<br>...<br>O4'(G7)   | (G7)N10–H<br>...<br>N2(G/G6) | (G7)N10–H<br>...<br>O4'(C8) | (C26)N4–H<br>...<br>O6(G7)   | (C26)N4–H<br>...<br>N7(G7)  | (C26)N4–H<br>...<br>O2(C8)  | (C27)N4–H<br>...<br>OP2(C26) |                            |                            |
| G <sub>2</sub> <sup>W</sup> | Absent                         |                     | 0.0%                          | 0.0%                         | 42.2%                       | 75.6%                        | 44.2%                       | 0.0%                        | 0.0%                         |                            |                            |
| G <sub>2</sub> <sup>W</sup> | G <sub>1</sub> <sup>anti</sup> |                     | 0.0%                          | 5.9%                         | 23.1%                       | 70.4%                        | 26.8%                       | 7.1%                        | 9.7%                         |                            |                            |
| G <sub>2</sub> <sup>W</sup> | G <sub>1</sub> <sup>syn</sup>  |                     | 48.6%                         | 0.0%                         | 35.7%                       | 72.7%                        | 23.7%                       | 0.0%                        | 0.0%                         |                            |                            |
| G <sub>2</sub> <sup>W</sup> | G <sub>3</sub> <sup>anti</sup> |                     | 0.0%                          | 26.1%                        | 0.0%                        | 26.0%                        | 68.9%                       | 24.1%                       | 30.9%                        |                            |                            |
| G <sub>2</sub> <sup>W</sup> | G <sub>3</sub> <sup>syn</sup>  |                     | 0.0%                          | 40.8%                        | 0.0%                        | 42.7%                        | 58.8%                       | 0.0%                        | 0.0%                         |                            |                            |
|                             |                                |                     | (G9)N10–H<br>...<br>O4'(G9)   | (G9)N10–H<br>...<br>O2(C10)  | (G9)N10–H<br>...<br>N3(C10) | (G9)N10–H<br>...<br>O4'(C10) | (G9)N10–H<br>...<br>N2(G25) | (G9)N10–H<br>...<br>N3(G25) | (C23)N2–H<br>...<br>N3(C24)  | (C24)N4–H<br>...<br>O6(G9) | (C24)N4–H<br>...<br>N7(G9) |
| G <sub>3</sub> <sup>W</sup> | Absent                         |                     | 0.0%                          | 0.0%                         | 0.0%                        | 0.0%                         | 56.3%                       | 27.9%                       | 0.0%                         | 67.2%                      | 41.4%                      |
| G <sub>3</sub> <sup>W</sup> | G <sub>1</sub> <sup>anti</sup> |                     | 0.0%                          | 0.0%                         | 0.0%                        | 0.0%                         | 56.6%                       | 32.3%                       | 0.0%                         | 70.3%                      | 36.6%                      |
| G <sub>3</sub> <sup>W</sup> | G <sub>1</sub> <sup>syn</sup>  |                     | 0.0%                          | 35.9%                        | 0.0%                        | 0.0%                         | 5.1%                        | 0.0%                        | 72.5%                        | 48.5%                      | 6.4%                       |
| G <sub>3</sub> <sup>W</sup> | G <sub>2</sub> <sup>anti</sup> |                     | 11.2%                         | 0.0%                         | 0.0%                        | 39.8%                        | 0.0%                        | 0.0%                        | 0.0%                         | 77.7%                      | 45.3%                      |
| G <sub>3</sub> <sup>W</sup> | G <sub>2</sub> <sup>syn</sup>  |                     | 49.3%                         | 6.0%                         | 25.4%                       | 0.0%                         | 0.0%                        | 0.0%                        | 0.0%                         | 32.8%                      | 55.2%                      |

<sup>a</sup>Nucleotide numbering is illustrated in Figure F.14. <sup>b</sup>Occupancies determined as the percentage of the simulation that a hydrogen bonds exist according to a heavy atom distance < 3.2 Å and a heavy atom–hydrogen–heavy atom angle > 120°. <sup>c</sup>Conformation has no frames.

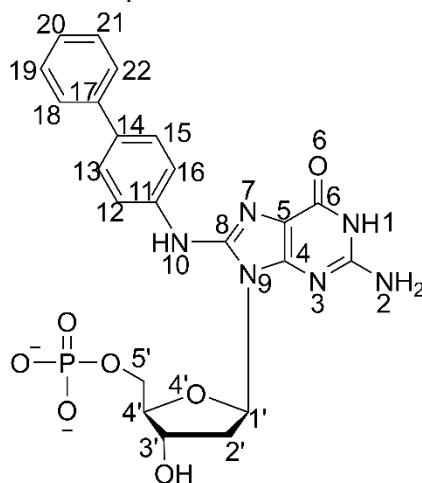
Table F.27:  $\Delta(\text{SASA}^{\text{avg}})$  and  $\Delta(\text{LSIE}^{\text{avg}})$  for  $^{\text{ABP}}\text{dG}$  diadducted DNA highlighting the impact upon addition of a second lesion.

| Primary Lesion                      | Secondary Lesion               | $\Delta(\text{SASA}^{\text{avg}})$<br>( $\text{\AA}^2$ ) <sup>a</sup> | $\Delta(\text{LSIE}^{\text{avg}})$<br>(kJ/mol) <sup>b</sup> |
|-------------------------------------|--------------------------------|-----------------------------------------------------------------------|-------------------------------------------------------------|
| <b>G<sub>1</sub><sup>qnti</sup></b> | absent                         | 0.0                                                                   | 0.0                                                         |
|                                     | G <sub>2</sub> <sup>qnti</sup> | −38.5                                                                 | −21.2                                                       |
|                                     | G <sub>2</sub> <sup>syn</sup>  | −37.0                                                                 | −11.0                                                       |
|                                     | G <sub>3</sub> <sup>qnti</sup> | 0.3                                                                   | 3.3                                                         |
|                                     | G <sub>3</sub> <sup>syn</sup>  | −1.2                                                                  | 3.9                                                         |
| <b>G<sub>2</sub><sup>qnti</sup></b> | absent                         | 0.0                                                                   | 0.0                                                         |
|                                     | G <sub>1</sub> <sup>qnti</sup> | −47.3                                                                 | −24.6                                                       |
|                                     | G <sub>1</sub> <sup>syn</sup>  | −28.1                                                                 | −8.6                                                        |
|                                     | G <sub>3</sub> <sup>qnti</sup> | −1.8                                                                  | −0.6                                                        |
|                                     | G <sub>3</sub> <sup>syn</sup>  | −2.3                                                                  | −0.3                                                        |
| <b>G<sub>3</sub><sup>qnti</sup></b> | absent                         | 0.0                                                                   | 0.0                                                         |
|                                     | G <sub>1</sub> <sup>qnti</sup> | −0.3                                                                  | 1.2                                                         |
|                                     | G <sub>1</sub> <sup>syn</sup>  | −1.1                                                                  | −0.3                                                        |
|                                     | G <sub>2</sub> <sup>qnti</sup> | −6.3                                                                  | −2.0                                                        |
|                                     | G <sub>2</sub> <sup>syn</sup>  | 0.0                                                                   | −14.4                                                       |
| <b>G<sub>1</sub><sup>syn</sup></b>  | absent                         | 0.0                                                                   | 0.0                                                         |
|                                     | G <sub>2</sub> <sup>qnti</sup> | −44.1                                                                 | −87.7                                                       |
|                                     | G <sub>2</sub> <sup>syn</sup>  | −27.4                                                                 | −129.2                                                      |
|                                     | G <sub>3</sub> <sup>qnti</sup> | −25.9                                                                 | −41.4                                                       |
|                                     | G <sub>3</sub> <sup>syn</sup>  | −37.2                                                                 | −114.4                                                      |
| <b>G<sub>2</sub><sup>syn</sup></b>  | absent                         | 0.0                                                                   | 0.0                                                         |
|                                     | G <sub>1</sub> <sup>qnti</sup> | −40.0                                                                 | −86.7                                                       |
|                                     | G <sub>1</sub> <sup>syn</sup>  | −13.7                                                                 | −68.8                                                       |
|                                     | G <sub>3</sub> <sup>qnti</sup> | 12.4                                                                  | 22.6                                                        |
|                                     | G <sub>3</sub> <sup>syn</sup>  | 2.3                                                                   | −123.3                                                      |
| <b>G<sub>3</sub><sup>syn</sup></b>  | absent                         | 0.0                                                                   | 0.0                                                         |
|                                     | G <sub>1</sub> <sup>qnti</sup> | −14.8                                                                 | −56.8                                                       |
|                                     | G <sub>1</sub> <sup>syn</sup>  | −19.6                                                                 | −8.0                                                        |
|                                     | G <sub>2</sub> <sup>qnti</sup> | −12.1                                                                 | −33.4                                                       |
|                                     | G <sub>2</sub> <sup>syn</sup>  | −50.6                                                                 | −47.6                                                       |

<sup>a</sup>Calculated as the SASA for the lesion in diadducted DNA minus the corresponding lesion in monoadducted DNA.

<sup>b</sup>Calculated as the difference in LSIE for the lesion in diadducted DNA minus the corresponding lesion in monoadducted DNA.

Table F.28: MOL2 file containing partial charges and atom types of the <sup>ABP</sup>dG adduct used for MD simulations (atom numbers are provided in the associated figure).



|    |      |       |        |        |    |   |     |         |
|----|------|-------|--------|--------|----|---|-----|---------|
| 1  | P    | 5.286 | 2.259  | 0.416  | P  | 1 | UNL | 1.2097  |
| 2  | O5'  | 3.956 | 1.39   | 0      | OS | 1 | UNL | -0.489  |
| 3  | OP1  | 4.85  | 3.558  | 0.947  | O2 | 1 | UNL | -0.7854 |
| 4  | OP2  | 6.186 | 1.327  | 1.111  | O2 | 1 | UNL | -0.7961 |
| 5  | O3'  | 0     | 0      | 0      | OS | 1 | UNL | -0.5564 |
| 6  | C5'  | 2.85  | 2.071  | -0.537 | CI | 1 | UNL | -0.0003 |
| 7  | H5'  | 3.153 | 2.703  | -1.366 | H1 | 1 | UNL | 0.0695  |
| 8  | H5'' | 2.378 | 2.694  | 0.218  | H1 | 1 | UNL | 0.0695  |
| 9  | C4'  | 1.851 | 1.054  | -1.038 | CT | 1 | UNL | 0.2036  |
| 10 | H4'  | 0.966 | 1.576  | -1.384 | H1 | 1 | UNL | 0.0894  |
| 11 | O4'  | 2.433 | 0.347  | -2.11  | OS | 1 | UNL | -0.4156 |
| 12 | C1'  | 2.056 | -1.002 | -2.047 | CT | 1 | UNL | 0.2371  |
| 13 | H1'  | 1.088 | -1.141 | -2.503 | H2 | 1 | UNL | 0.033   |
| 14 | C3'  | 1.407 | 0      | 0      | CT | 1 | UNL | 0.1661  |
| 15 | H3'  | 1.783 | 0.23   | 0.99   | H1 | 1 | UNL | 0.0605  |
| 16 | C2'  | 1.989 | -1.301 | -0.551 | CT | 1 | UNL | -0.043  |
| 17 | H2'  | 2.981 | -1.474 | -0.154 | HC | 1 | UNL | 0.0363  |
| 18 | H2'' | 1.371 | -2.163 | -0.331 | HC | 1 | UNL | 0.0363  |
| 19 | N9   | 2.973 | -1.791 | -2.824 | N* | 1 | UNL | 0.0089  |
| 20 | C8   | 4.359 | -1.885 | -2.756 | CR | 1 | UNL | 0.2068  |
| 21 | N7   | 4.841 | -2.715 | -3.604 | NB | 1 | UNL | -0.3943 |
| 22 | C5   | 3.745 | -3.213 | -4.283 | CB | 1 | UNL | 0.0202  |
| 23 | C6   | 3.665 | -4.177 | -5.34  | C  | 1 | UNL | 0.4962  |
| 24 | O6   | 4.514 | -4.811 | -5.894 | O  | 1 | UNL | -0.552  |
| 25 | N1   | 2.313 | -4.352 | -5.726 | NA | 1 | UNL | -0.3959 |
| 26 | H1   | 2.176 | -5.063 | -6.412 | H  | 1 | UNL | 0.3284  |
| 27 | C2   | 1.241 | -3.719 | -5.189 | CA | 1 | UNL | 0.5574  |
| 28 | N2   | 0.017 | -4.077 | -5.679 | N2 | 1 | UNL | -0.8548 |

|    |      |        |        |        |    |   |     |         |
|----|------|--------|--------|--------|----|---|-----|---------|
| 29 | H21  | -0.035 | -4.334 | -6.64  | H  | 1 | UNL | 0.3841  |
| 30 | H22  | -0.719 | -3.475 | -5.381 | H  | 1 | UNL | 0.3841  |
| 31 | N3   | 1.326  | -2.852 | -4.241 | NC | 1 | UNL | -0.4904 |
| 32 | C4   | 2.6    | -2.647 | -3.826 | CB | 1 | UNL | 0.1747  |
| 33 | N10  | 5.042  | -1.146 | -1.832 | N2 | 1 | UNL | -0.2954 |
| 34 | H10  | 4.564  | -0.353 | -1.463 | H  | 1 | UNL | 0.2901  |
| 35 | C11  | 6.409  | -1.227 | -1.515 | CA | 1 | UNL | 0.0946  |
| 36 | C12  | 7.353  | -1.908 | -2.275 | CA | 1 | UNL | -0.0501 |
| 37 | H12  | 7.06   | -2.419 | -3.166 | HA | 1 | UNL | 0.0915  |
| 38 | C16  | 6.828  | -0.572 | -0.358 | CA | 1 | UNL | -0.2184 |
| 39 | H16  | 6.101  | -0.077 | 0.258  | HA | 1 | UNL | 0.1271  |
| 40 | C13  | 8.679  | -1.92  | -1.871 | CA | 1 | UNL | -0.1843 |
| 41 | H13  | 9.392  | -2.438 | -2.486 | HA | 1 | UNL | 0.1484  |
| 42 | C15  | 8.154  | -0.585 | 0.022  | CA | 1 | UNL | -0.1187 |
| 43 | H15  | 8.44   | -0.088 | 0.932  | HA | 1 | UNL | 0.1295  |
| 44 | C14  | 9.114  | -1.262 | -0.727 | cp | 1 | UNL | -0.0043 |
| 45 | C17  | 10.545 | -1.284 | -0.313 | cp | 1 | UNL | 0.0284  |
| 46 | C18  | 11.181 | -0.125 | 0.128  | CA | 1 | UNL | -0.1012 |
| 47 | H18  | 10.637 | 0.802  | 0.148  | HA | 1 | UNL | 0.1132  |
| 48 | C22  | 11.284 | -2.464 | -0.356 | CA | 1 | UNL | -0.1032 |
| 49 | H22' | 10.81  | -3.375 | -0.674 | HA | 1 | UNL | 0.1212  |
| 50 | C19  | 12.51  | -0.145 | 0.515  | CA | 1 | UNL | -0.1867 |
| 51 | H19  | 12.982 | 0.763  | 0.847  | HA | 1 | UNL | 0.1441  |
| 52 | C21  | 12.615 | -2.484 | 0.027  | CA | 1 | UNL | -0.181  |
| 53 | H21' | 13.165 | -3.408 | -0.01  | HA | 1 | UNL | 0.1432  |
| 54 | C20  | 13.234 | -1.325 | 0.466  | CA | 1 | UNL | -0.1187 |
| 55 | H20  | 14.267 | -1.341 | 0.765  | HA | 1 | UNL | 0.1321  |

## **Appendix G**

**Supplementary Information for Chapter 8: Understanding the Nucleotide Excision Repair of the 2-Acetylaminofluorene-C8-dG Adduct: Impact of Number of Adducted Sites and Helical Conformation**

Contains Figures G.1–G.11 and Tables G.1–G.6

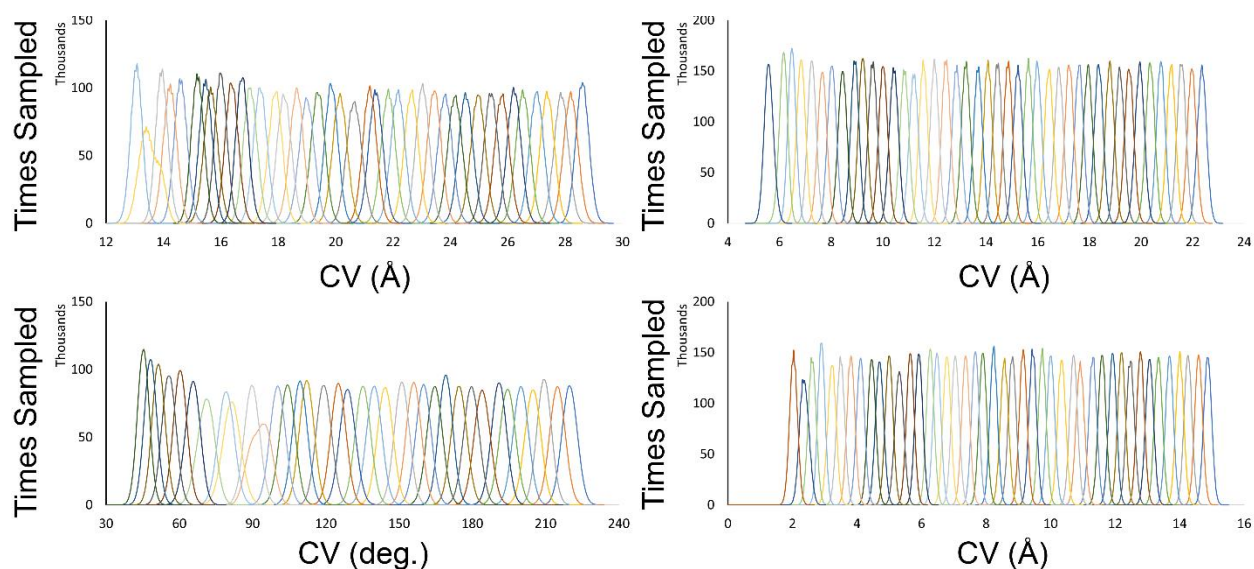


Figure G.1: Histograms of the overlap between the umbrella sampling windows during Stage 1 (top left), Stage 2 (top right), Stage 3 (bottom left) and Stage 4 (bottom right) for Rad4 binding to the B conformation of <sup>AAF</sup>dG monoadducted DNA. See Figure 4 for descriptions of the binding stages.

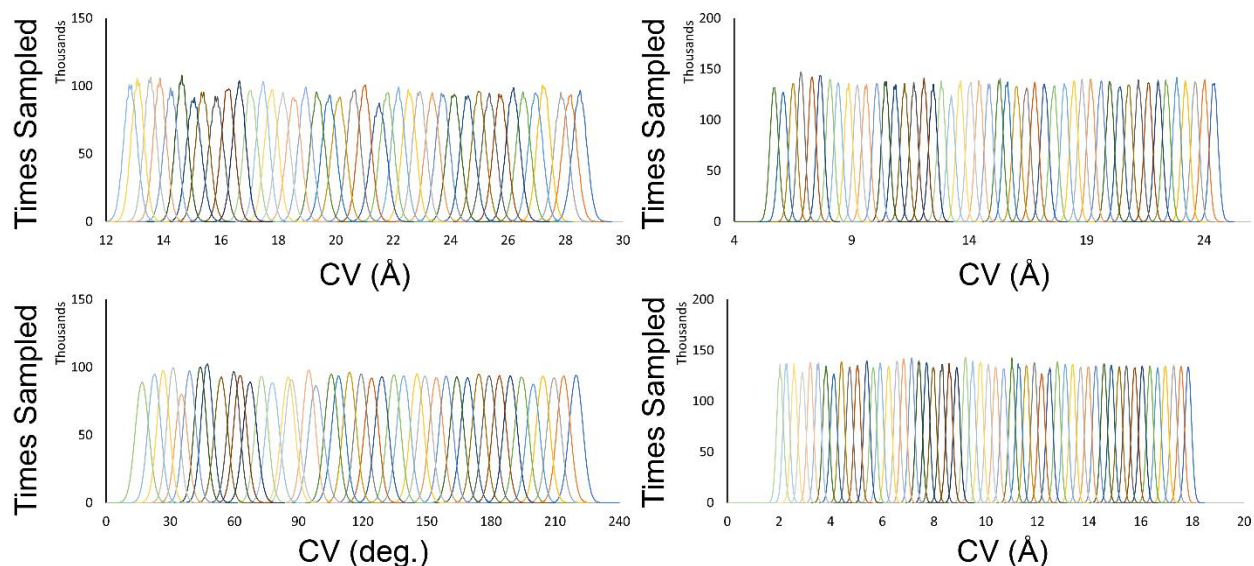


Figure G.2: Histograms of the overlap between the umbrella sampling windows during Stage 1 (top left), Stage 2 (top right), Stage 3 (bottom left) and Stage 4 (bottom right) for Rad4 binding to the S conformation of <sup>AAF</sup>dG monoadducted DNA. See Figure 4 for descriptions of the binding stages.



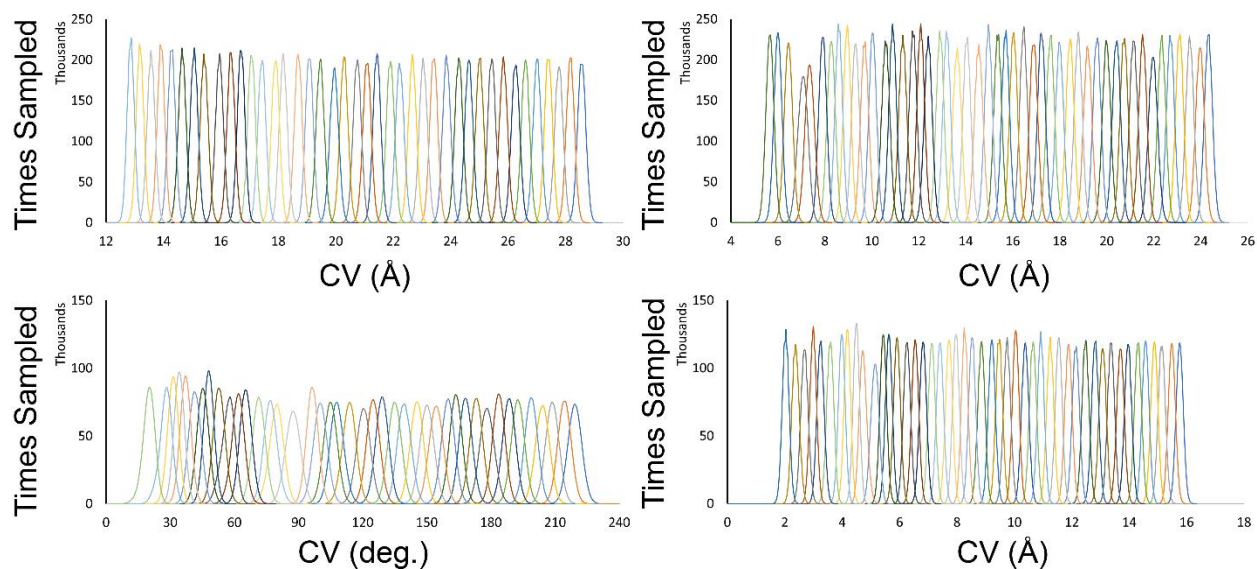


Figure G.3: Histograms of the overlap between the umbrella sampling windows during Stage 1 (top left), Stage 2 (top right), Stage 3 (bottom left) and Stage 4 (bottom right) for Rad4 binding to the BB conformation of <sup>AAF</sup>dG diadducted DNA. See Figure 4 for descriptions of the binding stages.

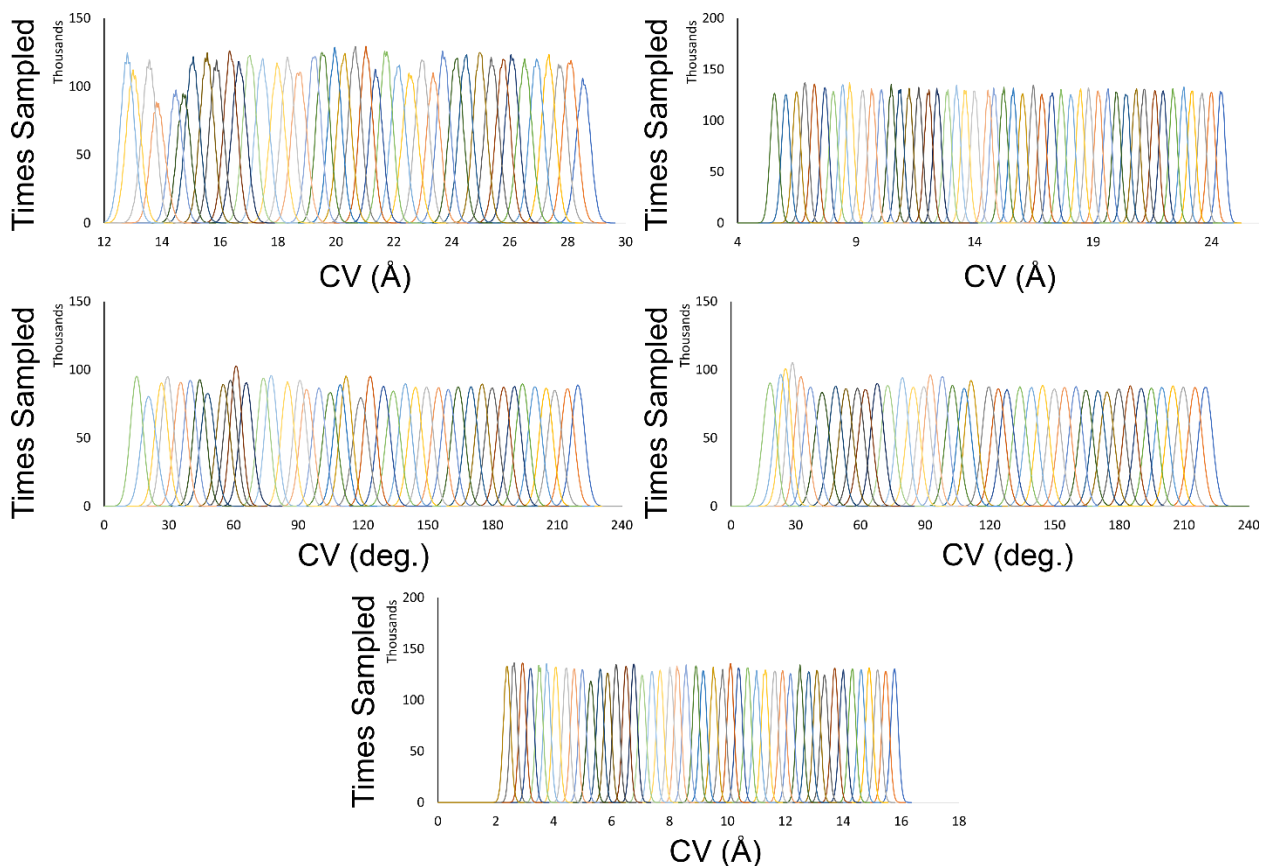


Figure G.4: Histograms of the overlap between the umbrella sampling windows during Stage 1 (top left), Stage 2 (top right), Stage 3 (middle left), Stage 3b (middle right) and Stage 4 (bottom) for Rad4 binding to undamaged DNA. See Figure 4 for descriptions of the binding stages.

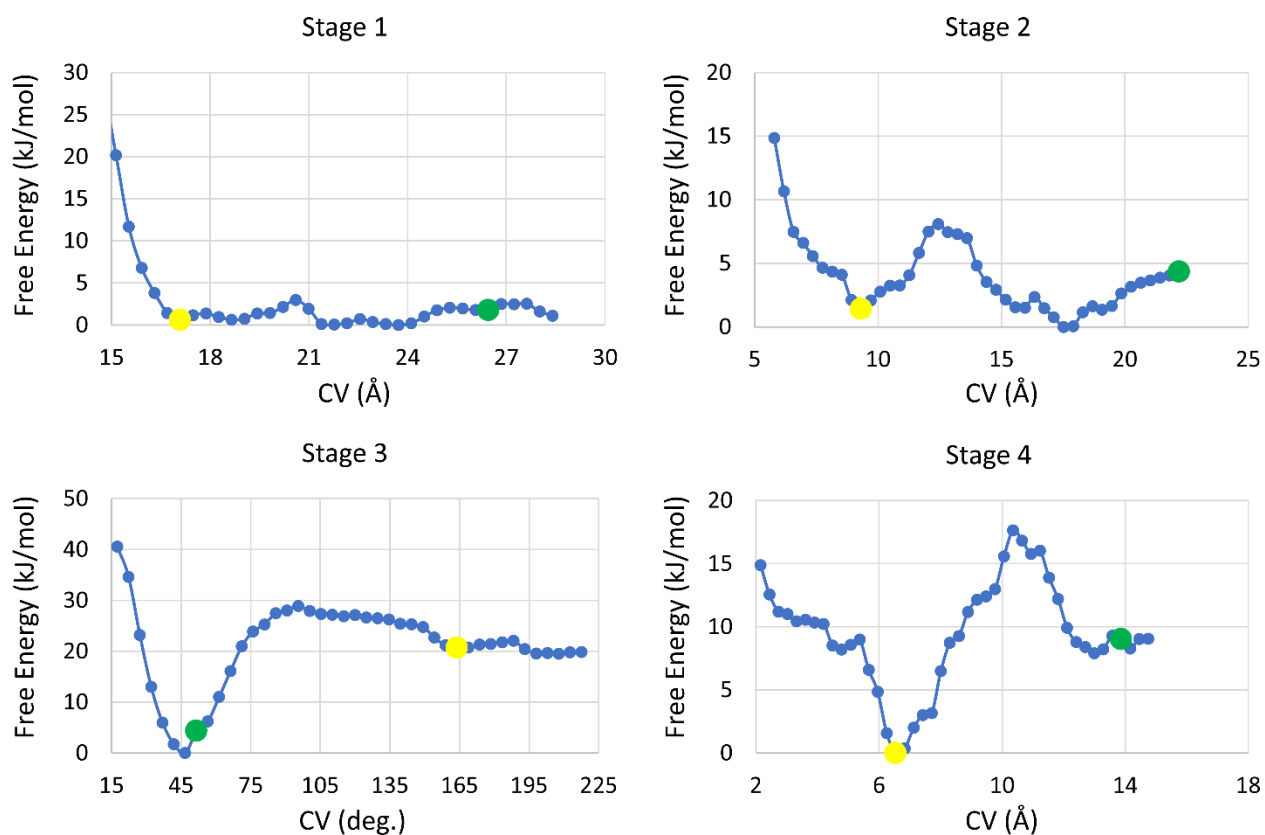


Figure G.5: Free energy plots for the different stages of Rad4 binding to the B conformation of <sup>AAF</sup>dG monoadducted DNA. See Figure 4 for descriptions of the binding stages. The green circle represents the starting CV, and the yellow circle represents the final data point.

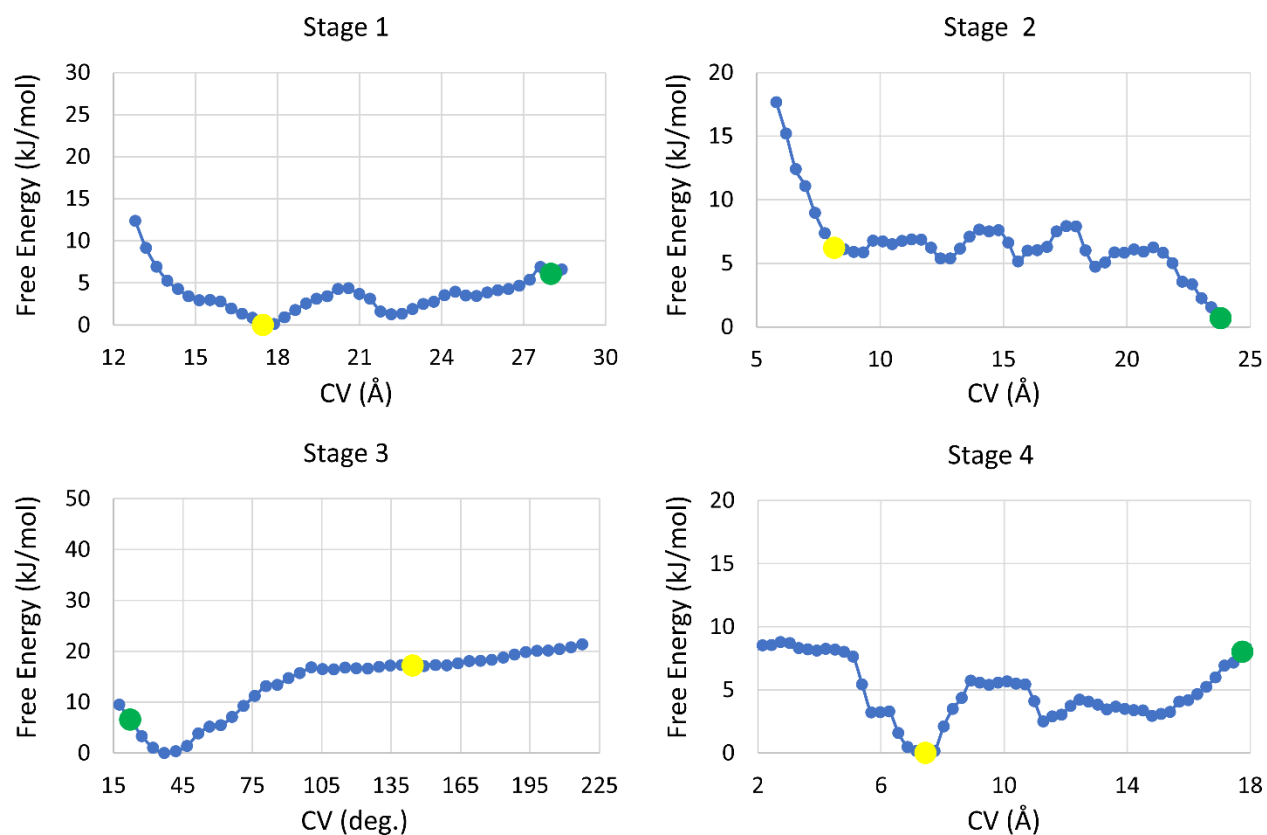


Figure G.6: Free energy plots for the different stages of Rad4 binding to the S conformation of <sup>AAF</sup>dG monoadducted DNA. See Figure 4 for descriptions of the binding stages. The green circle represents the starting CV, and the yellow circle represents the final data point.

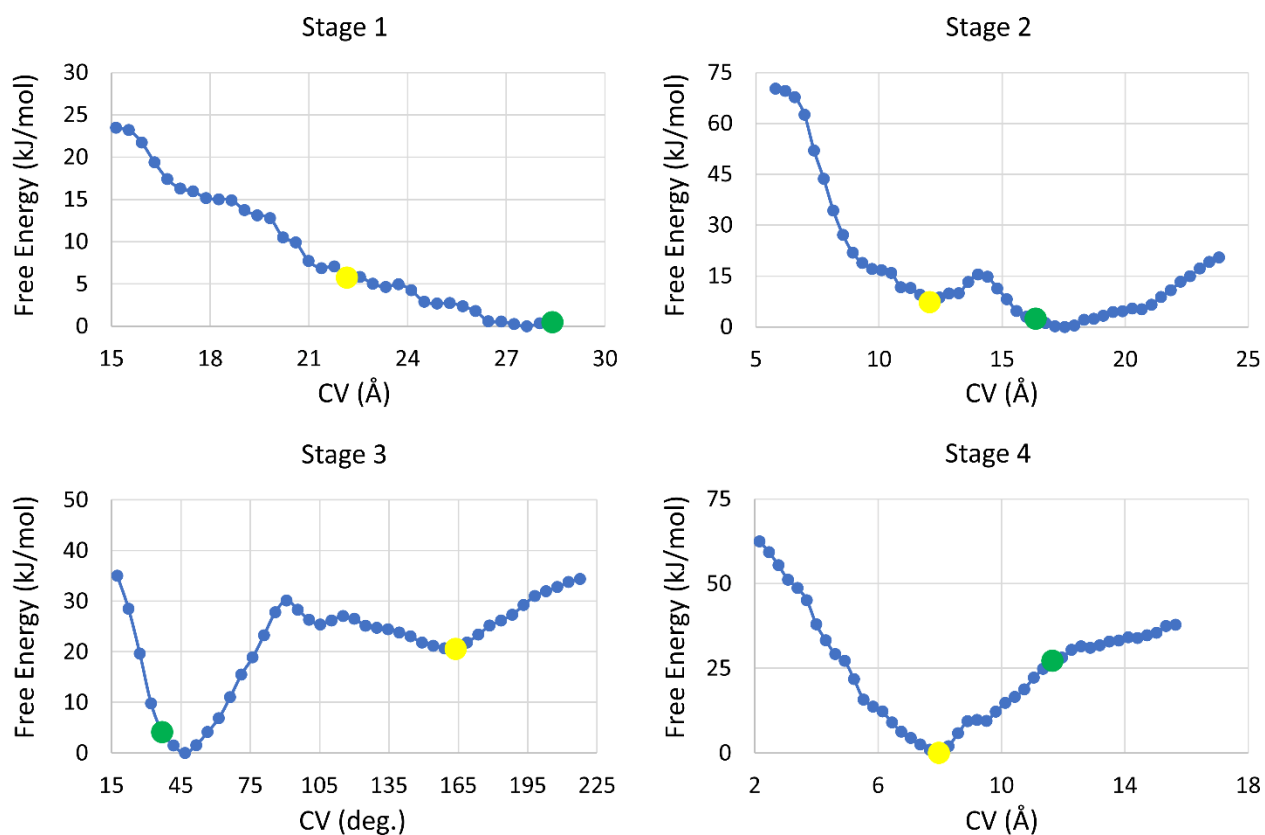


Figure G.7: Free energy plots for the different stages of Rad4 binding to the BB conformation of <sup>AAF</sup>dG diadducted DNA. See Figure 4 for descriptions of the binding stages. The green circle represents the starting CV, and the yellow circle represents the final data point.

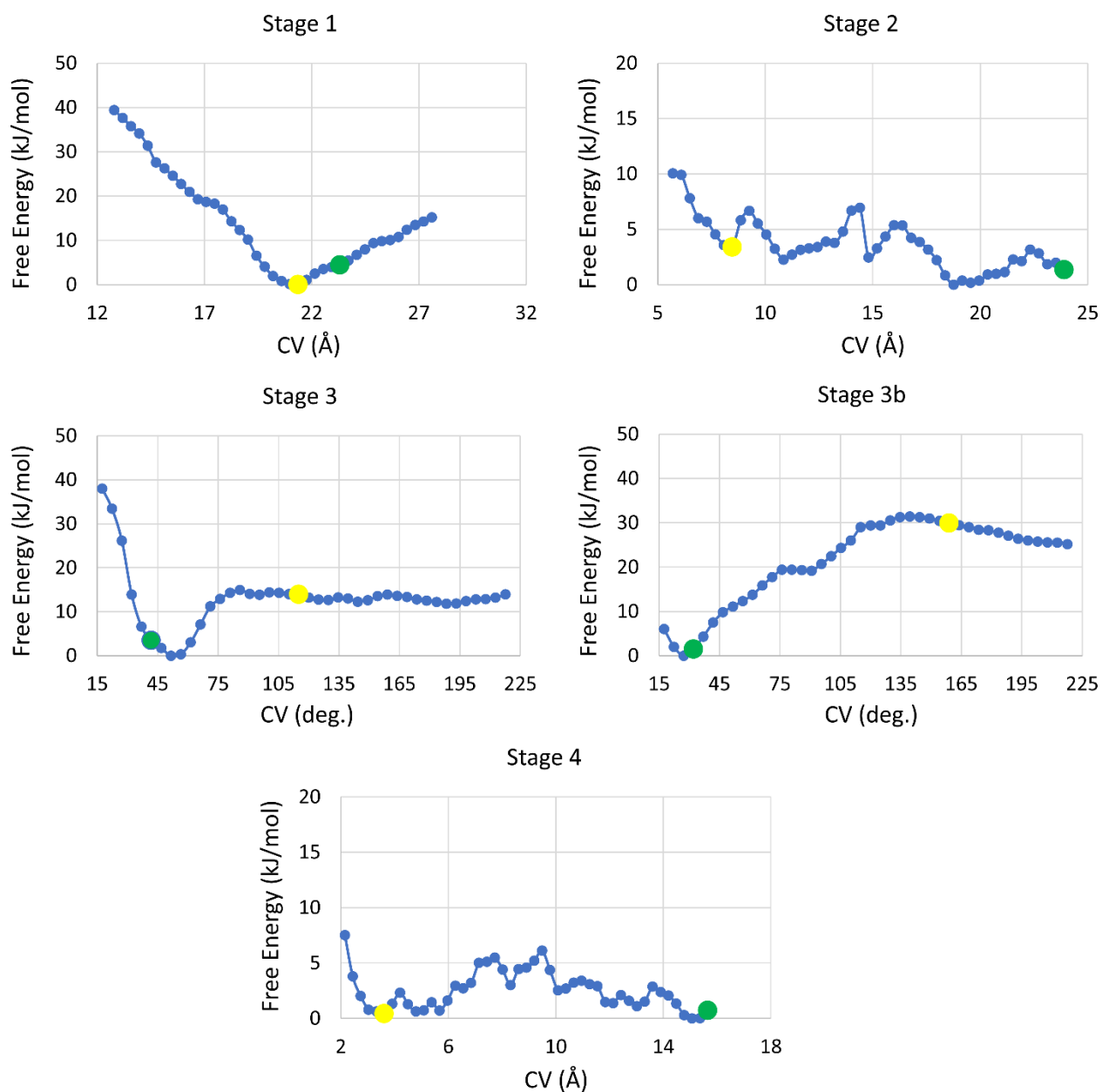


Figure G.8: Free energy plots for the different stages of Rad4 binding to undamaged DNA. See Figure 4 for descriptions of the binding stages. The green circle represents the starting CV, and the yellow circle represents the final data point.

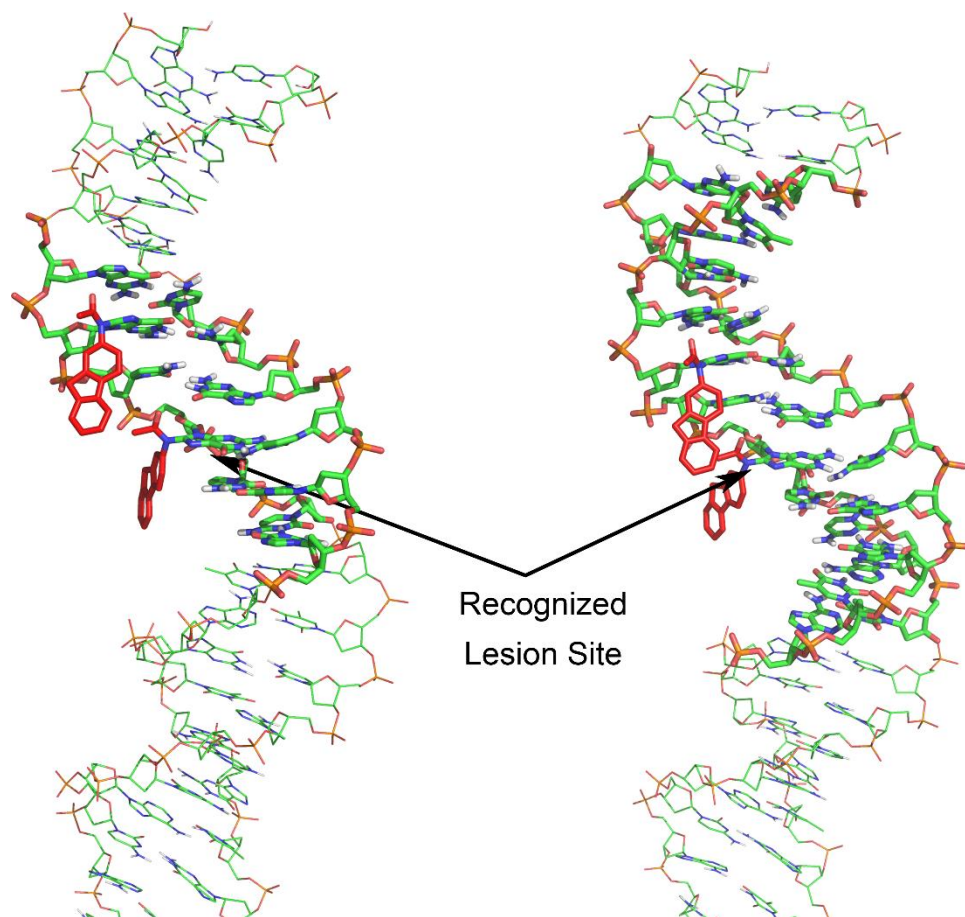


Figure G.9: MD representative structures of <sup>AAF</sup>dG diadducted DNA in the BB conformation. The DNA twist for all systems was calculated by defining a pseudostep between the base pairs on either side of the central damaged section (*NarI* sequence) shown in tubes (left). The DNA bending was calculated over the tube region on the right, with the AAF moiety shown in red. Beyond the first base pair on either side of the central fragment, the surrounding DNA shown as wireframe was not included in the calculations of these structural parameters.

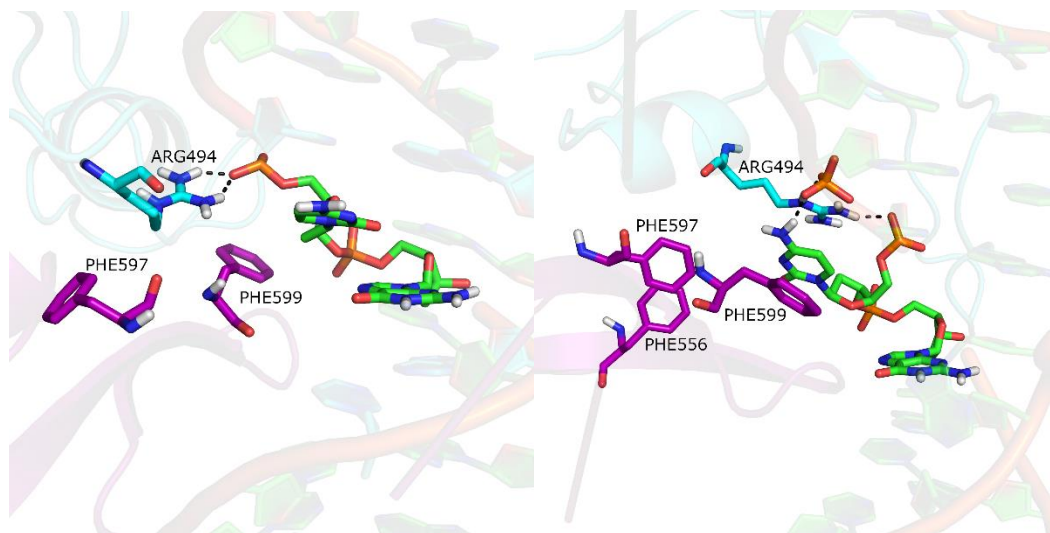


Figure G.10: MD structures of the Rad4-DNA complexes, with the duplex in the B conformation (left) or the S conformation (right), after the coordination of the base flipping residues (Stage 2), showing the differing interactions between the PHE “flipping path” residues of the BHD3 domain (purple) and Arg494 from the BHD2 domain (cyan), and the C pairing (green) across from the damage site.

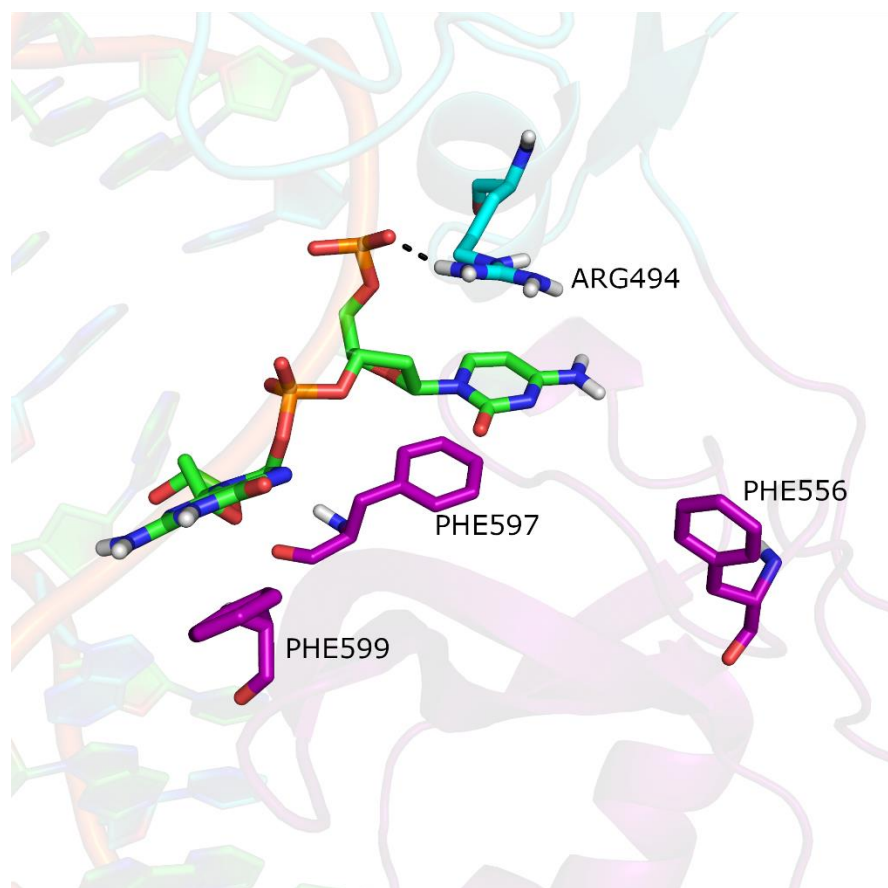


Figure G.11: MD structure of a Rad4 binding to <sup>AAF</sup>dG adducted DNA that was in the B conformation, after the formation of the helical cavity (Stage 3), highlighting the interactions of key stabilizing residues, namely the PHE “flipping path” from the BHD3 domain (purple) and Arg494 from the BHD2 domain (cyan), that stabilize the flipped nucleobases (green).



Table G.1: Collective variables used to sample the initial binding and recognition of the B conformation of <sup>AAF</sup>dG monoadducted DNA.

| US Step | Collective Variables                                                                                             |                                                        |                                                                                                                                                                                                                   |                                    |
|---------|------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------|
| Stage 1 | Distance Between COM1 and COM2                                                                                   |                                                        |                                                                                                                                                                                                                   |                                    |
|         | COM1 is the COM of sugar-phosphate backbone heavy atoms of the damaged nucleotide and its 3' flanking nucleotide |                                                        | COM2 is the COM of the main chain heavy atoms (excluding the O) of the binding pocket of Rad4 (ARG494, TYR497, MET498, ASN554, PHE556, ASN558, GLU560, PHE562, PRO592, VAL594, SER596, PHE597, PRO607 and LEU609) |                                    |
| Stage 2 | Distance Between COM1 and COM2                                                                                   |                                                        |                                                                                                                                                                                                                   |                                    |
|         | COM1 is the COM of the N7, C5, C6 and O6 atoms of the 5' interstrand neighbor to the bound adduct                |                                                        | COM2 is the COM of the MET498 main chain heavy atoms from the binding pocket of Rad4                                                                                                                              |                                    |
| Stage 3 | Dihedral Angle Formed from the Four Following Atoms                                                              |                                                        |                                                                                                                                                                                                                   |                                    |
|         | N1 of the 3' flanking G with respect to the adduct                                                               | O4' of the interstrand 3' C with respect to the adduct | O4' of the C paired with the adduct                                                                                                                                                                               | N4 of the C paired with the adduct |
| Stage 4 | Distance Between COM1 and COM2                                                                                   |                                                        |                                                                                                                                                                                                                   |                                    |
|         | COM1 is the COM of the four sugar rings of the base pairs flanking the two bound bases                           |                                                        | COM2 is the COM of the backbone of the leading edge of the BHD3 (SER603, THE604 and VAL605)                                                                                                                       |                                    |

Table G.2: Collective variables used to sample the initial binding and recognition of the S conformation of <sup>AAF</sup>dG monoadducted DNA. Bolded regions represent differences in the CVs from the <sup>AAF</sup>dG monoadducted DNA in the B conformation (Table G.1).

| US Step | Collective Variables                                                                                             |                                                        |                                                                                                                                                                                                                   |                                                              |
|---------|------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------|
| Stage 1 | Distance Between COM1 and COM2                                                                                   |                                                        |                                                                                                                                                                                                                   |                                                              |
|         | COM1 is the COM of sugar-phosphate backbone heavy atoms of the damaged nucleotide and its 3' flanking nucleotide |                                                        | COM2 is the COM of the main chain heavy atoms (excluding the O) of the binding pocket of Rad4 (ARG494, TYR497, MET498, ASN554, PHE556, ASN558, GLU560, PHE562, PRO592, VAL594, SER596, PHE597, PRO607 and LEU609) |                                                              |
| Stage 2 | Distance Between COM1 and COM2                                                                                   |                                                        |                                                                                                                                                                                                                   |                                                              |
|         | COM1 is the COM of the N7, C5, C6 and O6 atoms of the 5' interstrand neighbor to the bound adduct                |                                                        | COM2 is the COM of the MET498 main chain heavy atoms from the binding pocket of Rad4                                                                                                                              |                                                              |
| Stage 3 | Dihedral Angle Formed from the Four Following Atoms                                                              |                                                        |                                                                                                                                                                                                                   |                                                              |
|         | N1 of the 3' flanking G with respect to the adduct                                                               | O4' of the interstrand 3' C with respect to the adduct | <b>O4' of the 5' interstrand neighbor to the bound adduct</b>                                                                                                                                                     | <b>N1 of the 5' interstrand neighbor to the bound adduct</b> |
| Stage 4 | Distance Between COM1 and COM2                                                                                   |                                                        |                                                                                                                                                                                                                   |                                                              |
|         | COM1 is the COM of the four sugar rings of the base pairs flanking the two bound bases                           |                                                        | COM2 is the COM of the backbone of the leading edge of the BHD3 (SER603, THE604 and VAL605)                                                                                                                       |                                                              |

Table G.3: Collective variables used to sample the initial binding and recognition of <sup>AAF</sup>dG diadducted DNA.<sup>a</sup> Bolded regions represent differences in the CVs from the <sup>AAF</sup>dG monoadducted DNA in the B conformation (Table G.1).

| US Step | Collective Variables                                                                                             |                                                        |                                                                                                                                                                                                                   |                                                              |
|---------|------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------|
| Stage 1 | Distance Between COM1 and COM2                                                                                   |                                                        |                                                                                                                                                                                                                   |                                                              |
|         | COM1 is the COM of sugar-phosphate backbone heavy atoms of the damaged nucleotide and its 3' flanking nucleotide |                                                        | COM2 is the COM of the main chain heavy atoms (excluding the O) of the binding pocket of Rad4 (ARG494, TYR497, MET498, ASN554, PHE556, ASN558, GLU560, PHE562, PRO592, VAL594, SER596, PHE597, PRO607 and LEU609) |                                                              |
| Stage 2 | Distance Between COM1 and COM2                                                                                   |                                                        |                                                                                                                                                                                                                   |                                                              |
|         | COM1 is the COM of the N7, C5, C6 and O6 atoms of the 5' interstrand neighbor to the bound adduct                |                                                        | COM2 is the COM of the MET498 main chain heavy atoms from the binding pocket of Rad4                                                                                                                              |                                                              |
| Stage 3 | Dihedral Angle Formed from the Four Following Atoms                                                              |                                                        |                                                                                                                                                                                                                   |                                                              |
|         | N1 of the 3' flanking G with respect to the adduct                                                               | O4' of the interstrand 3' C with respect to the adduct | <b>O4' of the 5' interstrand neighbor to the bound adduct</b>                                                                                                                                                     | <b>N1 of the 5' interstrand neighbor to the bound adduct</b> |
| Stage 4 | Distance Between COM1 and COM2                                                                                   |                                                        |                                                                                                                                                                                                                   |                                                              |
|         | COM1 is the COM of the four sugar rings of the base pairs flanking the two bound bases                           |                                                        | COM2 is the COM of the backbone of the leading edge of the BHD3 (SER603, THE604 and VAL605)                                                                                                                       |                                                              |

<sup>a</sup>The diadducted DNA was considered in the BB conformation.

Table G.4: Collective variables used to sample the initial binding and recognition of undamaged DNA. Bolded regions represent differences in the CVs from the <sup>AAF</sup>dG monoadducted DNA in the B conformation (Table G.1).

| US Step  | Collective Variables                                                                                                         |                                                                                    |                                                                                                                                                                                                                   |                                                            |
|----------|------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------|
| Stage 1  | Distance Between COM1 and COM2                                                                                               |                                                                                    |                                                                                                                                                                                                                   |                                                            |
|          | COM1 is the COM of sugar-phosphate backbone heavy atoms of the recognized (probed) nucleotide and its 3' flanking nucleotide |                                                                                    | COM2 is the COM of the main chain heavy atoms (excluding the O) of the binding pocket of Rad4 (ARG494, TYR497, MET498, ASN554, PHE556, ASN558, GLU560, PHE562, PRO592, VAL594, SER596, PHE597, PRO607 and LEU609) |                                                            |
| Stage 2  | Distance Between COM1 and COM2                                                                                               |                                                                                    |                                                                                                                                                                                                                   |                                                            |
|          | COM1 is the COM of the N7, C5, C6 and O6 atoms of the 5' interstrand neighbor to the recognized (probed) nucleotide          |                                                                                    | COM2 is the COM of the MET498 main chain heavy atoms from the binding pocket of Rad4                                                                                                                              |                                                            |
| Stage 3  | Dihedral Angle Formed from the Four Following Atoms                                                                          |                                                                                    |                                                                                                                                                                                                                   |                                                            |
|          | N1 of the 3' flanking G with respect to the recognized (probed) nucleotide                                                   | O4' of the interstrand 3' C with respect to the recognized (probed) nucleotide     | O4' of the C paired with the recognized (probed) nucleotide                                                                                                                                                       | N4 of the C paired with the recognized (probed) nucleotide |
| Stage 3b | Dihedral Angle Formed from the Four Following Atoms                                                                          |                                                                                    |                                                                                                                                                                                                                   |                                                            |
|          | N3 of the interstrand 3' C with respect to the recognized (probed) nucleotide                                                | <b>O4' of the 3' flanking G with respect to the recognized (probed) nucleotide</b> | <b>O4' of the recognized (probed) nucleotide</b>                                                                                                                                                                  | <b>N1 of the recognized (probed) nucleotide</b>            |
| Stage 4  | Distance Between COM1 and COM2                                                                                               |                                                                                    |                                                                                                                                                                                                                   |                                                            |
|          | COM1 is the COM of the four sugar rings of the base pairs flanking the two bound bases                                       |                                                                                    | COM2 is the COM of the backbone of the leading edge of the BHD3 (SER603, THE604 and VAL605)                                                                                                                       |                                                            |

Table G.5: Umbrella sampling protocols for Rad4 binding to damaged and undamaged DNA.

| System    | Stage | Starting CV <sup>a</sup> | Ending CV <sup>b</sup> | Interval <sup>c</sup> | Number of Windows <sup>d</sup> | Restraint Force <sup>e</sup>     |
|-----------|-------|--------------------------|------------------------|-----------------------|--------------------------------|----------------------------------|
| AAFdG-B   | 1     | 12.6 Å                   | 28.6 Å                 | 0.4 Å                 | 41                             | 10 kcal/(mol Å <sup>2</sup> )    |
|           | 2     | 5.6 Å                    | 22.4 Å                 | 0.4 Å                 | 43                             | 20 kcal/(mol Å <sup>2</sup> )    |
|           | 3     | 15°                      | 220°                   | 5°                    | 42                             | 100 kcal/(mol rad <sup>2</sup> ) |
|           | 4     | 2.0 Å                    | 14.9 Å                 | 0.3 Å                 | 44                             | 40 kcal/(mol Å <sup>2</sup> )    |
| AAFdG-S   | 1     | 12.6 Å                   | 28.6 Å                 | 0.4 Å                 | 41                             | 10 kcal/(mol Å <sup>2</sup> )    |
|           | 2     | 5.6 Å                    | 24.4 Å                 | 0.4 Å                 | 48                             | 20 kcal/(mol Å <sup>2</sup> )    |
|           | 3     | 15°                      | 220°                   | 5°                    | 42                             | 100 kcal/(mol rad <sup>2</sup> ) |
|           | 4     | 2.0 Å                    | 17.9 Å                 | 0.3 Å                 | 54                             | 40 kcal/(mol Å <sup>2</sup> )    |
| AAFdG-BB  | 1     | 12.6 Å                   | 28.6 Å                 | 0.4 Å                 | 41                             | 10 kcal/(mol Å <sup>2</sup> )    |
|           | 2     | 5.6 Å                    | 24.4 Å                 | 0.4 Å                 | 48                             | 20 kcal/(mol Å <sup>2</sup> )    |
|           | 3     | 15°                      | 220°                   | 5°                    | 42                             | 100 kcal/(mol rad <sup>2</sup> ) |
|           | 4     | 2.0 Å                    | 15.8 Å                 | 0.3 Å                 | 47                             | 40 kcal/(mol Å <sup>2</sup> )    |
| Undamaged | 1     | 12.6 Å                   | 28.6 Å                 | 0.4 Å                 | 41                             | 10 kcal/(mol Å <sup>2</sup> )    |
|           | 2     | 5.6 Å                    | 22.4 Å                 | 0.4 Å                 | 43                             | 20 kcal/(mol Å <sup>2</sup> )    |
|           | 3     | 15°                      | 220°                   | 5°                    | 42                             | 100 kcal/(mol rad <sup>2</sup> ) |
|           | 3b    | 15°                      | 220°                   | 5°                    | 42                             | 100 kcal/(mol rad <sup>2</sup> ) |
|           | 4     | 2.0 Å                    | 15.8 Å                 | 0.3 Å                 | 47                             | 40 kcal/(mol Å <sup>2</sup> )    |

<sup>a</sup>Minimum CV value of the simulation windows considered. <sup>b</sup>Maximum CV value of the simulation windows considered. <sup>c</sup>Interval between each umbrella sampling window. <sup>d</sup>Number of umbrella sampling windows simulated for each stage. <sup>e</sup>Force of the restraint used to maintain the CV value in the desired range.

Table G.6: Helical untwist relative to unbound canonical DNA, helical bend around the bound site and the AS interaction volume<sup>a</sup> between BHD2 and the minor groove of the duplexes.<sup>b,c</sup>

| System    | Prebound State      |                  |                             | Bound State         |                  |                             |
|-----------|---------------------|------------------|-----------------------------|---------------------|------------------|-----------------------------|
|           | Helical Untwist (°) | Helical Bend (°) | AS Volume (Å <sup>3</sup> ) | Helical Untwist (°) | Helical Bend (°) | AS Volume (Å <sup>3</sup> ) |
| AAFdG-B   | 15.4 (2.1)          | 29.7 (2.8)       | 213.7 (15.0)                | 44.0 (1.1)          | 43.7 (1.6)       | 361.8 (12.6)                |
| AAFdG-S   | 28.2 (1.6)          | 33.8 (2.3)       | 298.9 (25.3)                | 28.7 (1.5)          | 31.9 (1.7)       | 358.9 (21.8)                |
| AAFdG-BB  | 34.3 (1.2)          | 45.6 (1.7)       | 194.5 (25.3)                | 29.4 (2.0)          | 69.6 (3.3)       | 234.1 (33.0)                |
| Undamaged | 6.4 (2.2)           | 20.6 (0.8)       | 265.8 (8.3)                 | — <sup>d</sup>      | — <sup>d</sup>   | — <sup>d</sup>              |

<sup>a</sup>The AlphaSpace (AS) volume was calculated using the AlphaSpace3 package. <sup>b</sup>Averages calculated over five 1 μs replicas. <sup>c</sup>Block Standard error (in parenthesis) calculated over the five replicates. <sup>d</sup>The bound state of the undamaged duplex by Rad4 was not modelled due to the lack of formation of a productive bound complex.