COMPUTATIONAL INSIGHTS INTO THE MOLECULAR BASIS FOR THE REPLICATION OF FLEXIBLE TOBACCO-DERIVED DNA LESIONS

Katie Wilson
Bachelor of Science, University of Lethbridge, 2013

A Thesis
Submitted to the School of Graduate Studies
of the University of Lethbridge
in Partial Fulfilment of the
Requirements for the Degree

DOCTOR OF PHILOSOPHY

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

COMPUTATIONAL INSIGHTS INTO THE MOLECULAR BASIS FOR THE REPLICATION OF FLEXIBLE TOBACCO-DERIVED DNA LESIONS

KATIE WILSON

Date of Defence: October 11, 2018

Dr. Stacey Wetmore Supervisor	Professor	Ph.D.
Dr. Ute Wieden-Kothe Thesis Examination Committee Member	Professor	Ph.D
Dr. Marc Roussel Thesis Examination Committee Member	Professor	Ph.D.
Dr. Roy Golsteyn Thesis Examination Committee Member	Associate Professor	Ph.D.
Dr. Rajeev Prabhakar External Examiner University of Miami Miami, Florida, USA	Professor	Ph.D.
Dr. Peter Dibble Chair, Thesis Examination Committee	Associate Professor	Ph.D.

Abstract

DNA is damaged by many agents in the environment and this affects many cellular processes, including DNA replication. The present thesis uses a multiscale computational modeling approach to study the intrinsic conformational and base-pairing preferences of flexible O6-G and O2-T alkyl DNA lesions, and the corresponding properties within DNA duplexes and polymerase active sites. Specifically, the replication of O6-benzylguanine (Bz-G) by a prototypical DNA polymerase, DNA polymerase IV, as well as the replication of the tobacco-derived carcinogens, O6-[4-oxo-4-(3-pyridyl)butyl]guanine (POB-G), O6-[4-hydroxy-4-(3-pyridyl)butyl]guanine (PHB-G), and O2-[4-oxo-4-(3-pyridyl)butyl] thymine (POB-T), by human DNA polymerases η and κ, were investigated. This work uncovers structural bases for the reported lesion mutagenicity and biological processing. Additionally, a consistent theoretical framework was used to provide insight into the previously unidentified general base for the nucleotidyl transfer reaction catalyzed by polymerase η. Overall, this thesis emphasizes the complex interplay between many factors that are required to replicate damaged DNA.

Acknowledgements

First of all, I would like to thank my supervisor, Dr. Stacey Wetmore. Your mentorship and encouragement over the years has been remarkable. I have learned so much while working in your lab. There is no truly sufficient way to say thank-you for all of the support you have given me. I would also like to thank my committee members, Dr. Marc Roussel and Dr. Ute Wieden-Kothe. I appreciate all of the advice and encouragement you have given me throughout my Ph.D. Thank-you to Dr. Roy Golsteyn for agreeing to participate in my thesis defence. Finally, thank-you to Dr. Rajeev Prabhakar from the University of Miami for your valuable time and interest in this thesis.

My sincere thanks also goes to Dr. Pedro Alexandrino Fernandes and Dr. Maria João Ramos, as well as the members of the Ramos group (especially Ana Rita Calixto, Rui Neves, João Coimbra, Eduardo Oliveira and Henrique Fernandes) for welcoming me into your group at the University of Porto. I had such a great experience in Porto and learned so much.

Additionally, I have had the wonderful opportunity to work with many other great students at the University of Lethbridge. Firstly, I would like to thank the high school and undergraduate students (Minette Abendong, Colin Anderson, Lindey Felske, Cynthia Fonderson, Josh Garden, Carl Holland, Devany Holland, Dylan Nikkel, Kariann Szemethy, Rachael Wells, and Natasha Wetmore) whose research has contributed either directly to this work or to other publications. To the past and present graduate students and post-doctoral fellows (Dr. Mohamed Aboelnga, Dr. Preethi Seelam, Dr. Purshotam Sharma,

Priya Bhutani, Briana Boychuk, Hanzala Hussain, Preetleen Kathuria, Rajwinder Kaur, Ryan Kung, Stefan Lenz, and Shahin Sowlati), thank you for your continued support and enlightening discussions. It has been a pleasure to work with all of you. A special thanks to Priya, Lindey, Rajwinder, Ryan, Stefan, Shahin, Hanzala, Briana, and Preethi; your friendship throughout my Ph.D. means so much to me.

Thanks to the Natural Sciences and Engineering Research Council (NSERC), Alberta Innovates – Technology Futures, and the University of Lethbridge for student scholarships. Additionally, computational resources from the New Upscale Cluster for Lethbridge to Enable Innovative Chemistry (NUCLEIC), and those provided by Westgrid and Compute/Calcul Canada, are greatly appreciated.

Finally, thank you to all my friends and family. I truly appreciate the love and support you have given me throughout my degree. To my parents, Cam and Jo-Anne, thank-you for always being there for me and encouraging me to follow my dreams. To my brother, Chris, I appreciate all of the support and friendship you have given me. To Seth, thank you for being by my side; your love, support, and understanding means the world to me.

Table of Contents

Abstract iii
Acknowledgements iv
Table of Contents vi
List of Tables xi
List of Figures xii
Abbreviations xviii

Chapter 1: Introduction

1.1. General Overview 1

1.2. DNA Structure 2

1.3. DNA Damage 4

1.4. Tobacco Induced DNA Damage 6

1.5. DNA Replication 9

1.6. Previous Computational Studies on DNA Damage and TLS 12

1.7. Scope of Thesis 14

1.8. Reference 17

Chapter 2: The Complex Conformational Heterogeneity of the Highly Flexibility O6-Benzyl-Guanine DNA Adduct

2.1. Introduction 28

2.2. Computational Details 34
2.2.1 Nucleoside Model 34
2.2.2 Nucleotide Model 35
2.2.3. DNA Model 35

2.3. Results 36
2.3.1. Nucleoside Model 36
2.3.2. Nucleotide Model 41

2.3.4. DNA Model 42
2.3.4.1. Structure of DNA Containing BzG Paired Opposite a Pyrimidine 44
2.3.4.2. Structure of DNA Containing BzG Paired Opposite a Purine 51
2.3.4.3. Energetics of DNA Helices Containing BzG 53

2.4. Discussion 56
2.4.1. Intrinsic Flexibility of the Bulky Moiety 56
2.4.2. Overview of Adducted DNA Conformations 57
2.4.3. Modified Nomenclature for Adducted DNA Conformations 58
2.4.4. Comparison of BzG Adducted DNA Conformation with Other Adducts 59
2.4.5. Distortion to the DNA Helix 61
2.4.6. Preferred Conformation about the Glycosidic Bond 62
2.4.7. Preferred Base Pairing 64
2.4.8. Biological Implications 65

2.5. References 69

Chapter 3: Conformational Flexibility of the Benzyl-guanine Adduct in a Bypass Polymerase Active Site Permits Replication: Insights from Molecular Dynamics Simulations

3.1. Introduction 80

3.2. Computational Details 85

3.3. Results 88
3.3.1. Replication of Canonical dG 88
3.3.1.1 Pre-insertion Complex 88
3.3.1.2. Insertion Complex 89
3.3.2. Replication of Bz-dG 93
3.3.2.1 Pre-insertion Complex 93
3.3.2.2. Insertion Complex 95

3.4. Discussion 97

3.5. Reference 103

Chapter 4: Molecular Insights into the Translesion Synthesis of Benzyl-Guanine from Molecular Dynamics Simulations: Structural Evidence for Mutagenic and Non-Mutagenic Replication

4.1. Introduction 110

4.2. Computational Details 115

4.3. Results 116
4.3.1. Criteria Used to Assess the Catalytic Nature of Dpo4 Ternary Complexes 116
4.3.2. The simulated orientation of dCTP opposite dG in the Dpo4 active site is conducive for replication 118
4.3.3. Due to stabilizing DNA–DNA and DNA–polymerase interactions, dCTP is favorably oriented for insertion opposite both anti and syn-Bz-dG 120
4.3.4. dTTP is favorably aligned for insertion opposite anti-Bz-dG, while the lack of DNA–DNA interactions likely precludes dTTP insertion opposite syn-Bz-dG 123
4.3.5. Distortions in the Dpo4 active site that alter DNA–polymerase and DNA–DNA interactions likely prevent insertion of a purine opposite Bz-dG 127
4.3.6. Deviations in key reaction parameters prevent –1 base deletion mutations upon Bz-dG replication by Dpo4 131

4.4. Discussion 133
4.4.1. MD simulations on the Dpo4 ternary insertion complexes for the replication of Bz-dG provide a structural explanation for the experimentally-observed insertion of dCTP and dTTP opposite Bz-dG 133
4.4.2. MD simulations of the –1 base deletion complexes provide the first structural evidence for the lack of deletion mutations upon Bz-dG replication 138

4.5. Conclusions 140

4.6. References 142

Chapter 5: Computational Insights into the Mutagenicity of Two Tobacco Derived Carcinogenic DNA Lesions

5.1. Introduction 149

5.2. Computational Details 154

5.3. Results and Discussion 156
5.3.1. POB-G and PHB-G have a high degree of inherent conformational flexibility 156
5.3.2. The Watson-Crick face of POB-G and PHB-G forms stable and minimally distorted base pairs with C, T, and A, while the Hoogsteen face of POB-G and PHB-G forms stable but distorted base pairs with G 158
5.3.3. Both lesions preferably extended the bulky moiety into the major groove of DNA, while PHB-G may also adopt an intercalated conformation 161
5.3.4. Both POB-G and PHB-G form stable interactions with T and A in the DNA duplex 165

5.3.5. Polymerase η insertion complexes indicate that dCTP and dTTP are best aligned for catalytic insertion opposite POB-G and PHB-G, but suggest that PHB-G may display a higher mutagenic potential 169

5.4. Conclusion 174

5.5. References 176

Chapter 6: Evidence for the replication of a syn DNA pyrimidine: A computational investigation of a highly mutagenic tobacco-derived thymine lesion

6.1. Introduction 180

6.2. Results and Discussion 184
6.2.1. POB-T is a highly flexible lesion 184
6.2.2. POB-T can adopt the canonical anti glycosidic orientation in the pol κ active site, while only syn-POB-T can be accommodated by pol η 186
6.2.3. Pol κ adopts a catalytically conducive conformation for the insertion of dGTP opposite POB-T 189
6.2.4. Unique syn glycosidic orientation explains experimentally observed dTTP and dATP insertion opposite syn-POB-T by pol η 193

6.3. Conclusion 197

6.4. Computational Details 198
6.4.1. DFT Calculations 198
6.4.2. MD Simulations 199

6.5. References 201

Chapter 7:

7.1. Introduction 206

7.2. Computational Details 210
7.2.1. Structural Preparation 210
7.2.2. Classical MD Simulations 211
7.2.2. Umbrella Sampling 211
7.2.2. QM/MM Calculations 212

7.3. Results and Discussion 214

7.4. Conclusion 223

7.5. References 225

Chapter 8: Conclusions

8.1. Thesis Overview 231

8.2. Future Work 236
8.2.1. DNA Lesions 236
8.2.2. DNA Polymerase Reaction Mechanism 239

8.3. Final Remarks 240

8.4. References 241

Appendix A: Supplementary Information for Chapter 2 248

Appendix B: Supplementary Information for Chapter 3 264

Appendix C: Supplementary Information for Chapter 4 271

Appendix D: Supplementary Information for Chapter 5 290

Appendix E: Supplementary Information for Chapter 6 333

Appendix F: Supplementary Information for Chapter 7 369

List of Tables

Table 2.1 Parameters of the BzG adduct incorporated into the G^3 position of the *Nar*I DNA sequence 43

Table 2.2 AMBER interaction energies between BzG in the G^3 position of the *Nar*I sequence and the surrounding nucleobases 45

Table 2.3 AMBER interaction energies between BzG in the G^3 position of the *Nar*I sequence and the surrounding nucleobases 46

Table 2.4 MM-PBSA total free energy for the *Nar*I sequence with the BzG adduct incorporated at the G^3 position and various opposing bases 55

Table 2.5 MM-PBSA free energy of helix binding for the *Nar*I sequence with the BzG adduct incorporated at the G^3 position and various opposing bases 56

Table 7.1 ONIOM reaction energies for the DNA pol η catalyzed nucleotidyl transfer reactions involving different general bases for the initial primer 3′-OH deprotonation step. 217

List of Figures

Figure 1.1 a) B-form DNA double helix and b) chemical structure of a single strand of DNA, with the DNA backbone (deoxyribose sugars and phosphate moieties) highlighted in black and the DNA nucleobases highlighted in blue 2

Figure 1.2 Structure and numbering of the canonical DNA nucleobases 2

Figure 1.3 Watson-Crick A:T and G:C base pairs, with the location of the DNA major and minor grooves highlighted 4

Figure 1.4 The *anti* (left) and *syn* (right) conformations of a DNA nucleoside. The χ dihedral angle is highlighted in blue and the hydrogen-bonding face (Watson-Crick face for *anti* or Hoogsteen face for *syn*) is highlighted in green 4

Figure 1.5 Examples of different classes of DNA damage 5

Figure 1.6 Methyl and formaldehyde lesions formed upon exposure to NNK 7

Figure 1.7 Pyridyloxobutyl (R= =O) and pyridylhydroxybutyl (R= –OH) lesions formed upon exposure to NNK 8

Figure 1.8 General mechanism for the nucleotidyl reaction mechanism catalyzed by DNA polymerases 10

Figure 1.9 Structure of a) a replicative DNA polymerase (bacteriophage RB69, PDB ID: 1IG9), and b) a TLS polymerase (human DNA polymerase η, PDB ID: 4ECS), highlighting the various domains (little finger= cyan, thumb=yellow, palm=red, fingers=green, exonuclease=blue, and amino-terminal=orange) 11

Figure 2.1 The BzG nucleoside in the (a) anti and (b) syn orientations, and definition of important dihedral angles including χ, θ, φ and ξ. 29

Figure 2.2 (a) PES as a function of χ and θ in BzG, and (b) corresponding β-constrained optimized minima and transition states. 37

Figure 2.3 (a) φ and (b) ξ dihedral angles for each point on the BzG χ versus θ PES. 38

Figure 2.4 (a) PES as a function of θ and φ in BzG, and (b) corresponding β-constraint optimized minima and transition states. 39

Figure 2.5 (a) PES as a function of θ and ξ in BzG, and (b) corresponding β-constraint optimized minima and transition states 39

Figure 2.6 Optimized BzG β-constrained nucleoside minima including important dihedral angles and relative energies 40

Figure 2.7 Optimized BzG nucleotide minima including important dihedral angles and relative energies 42

Figure 2.8 (a) Representative MD structure of BzG adducted DNA with the Bz moiety located in the major groove. (b-d) Interactions within *anti*–BzG base pairs in representative MD structures 47

Figure 2.9 Representative MD structures and interactions involving *syn*–BzG paired opposite various canonical nucleotides 49

Figure 2.10 Discrete interactions between the extrahelical (a) C or (b) T and the surrounding bases in representative structures from MD simulations with syn-BzG intercalated in the DNA helix and the pairing base extrahelical 51

Figure 2.11 BzG interactions in representative structures observed during MD simulations on adducted DNA helices with various pairing nucleotides 52

Figure 3.1 Structure of the Bz-dG nucleoside with the bulky moiety highlighted in red. Definitions of key dihedral angles and the sequence context considered in MD simulations are provided 81

Figure 3.2 Representative MD structures and corresponding average stacking interactions between dG* and the 3′–base pair in the (a) dG, (b) *anti*-Bz-dG, and (c) *syn*-Bz-dG pre-insertion complexes for replication by Dpo4 90

Figure 3.3 Representative MD structures of the (a) dG, (b) *anti*-Bz-dG, and (c) *syn*-Bz-dG insertion complexes for replication by Dpo4, and (d) the corresponding average stacking interactions between the dG*:dCTP and 3′–dC:dG pair 91

Figure 3.4 Representative MD structures and corresponding average hydrogen-bond strength between dG* and dCTP in the (a) dG, (b) *anti*-Bz-dG, and (c) *syn*-Bz-dG insertion complexes for replication by Dpo4 92

Figure 3.5 Occupancies of hydrogen bonds between the template strand or dCTP and primer strand, and the surrounding amino acid residues in the (a) dG, (b) anti-Bz-dG, and (c) syn-Bz-dG insertion complexes for replication by Dpo4..... 93

Figure 3.6 Overlay of representative MD structures of the (a) dG and anti-Bz-dG, and (b) dG and syn-Bz-dG pre-insertion complexes for replication by Dpo4..... 97

Figure 3.7 Overlay of structures and respective protein backbone RMSD from different TLS stages for a complete replication cycle for (a) dG or (b) *anti*-Bz-dG replication by Dpo4..... 102

Figure 4.1 Structures of the Bz-dG nucleoside, other O6-dG alkylation adducts and canonical dG, as well as the definitions of key rotatable dihedral angles in Bz-dG..... 111

Figure 4.2 Representative MD structure for the Dpo4 ternary complex of canonical dG replication depicting a) the orientation of dCTP with respect to the DNA, b) the dG:dCTP hydrogen-bonding arrangement and the percent occupancy of the hydrogen-bonding interactions, and c) the average reaction parameters..... 120

Figure 4.3 Representative MD structures depicting (a and b) the orientation of dCTP with respect to the DNA in the Dpo4 ternary complexes and (c and d) possible Bz-dG:dCTP hydrogen-bonding arrangements for (a and c) *anti*-Bz-dG and (b and d) *syn*-Bz-dG replication..... 121

Figure 4.4 Representative MD structures depicting (a and b) the orientation of dTTP with respect to the DNA in the Dpo4 ternary complexes and (c and d) possible Bz-dG:dTTP hydrogen-bonding arrangements for (a and c) *anti*-Bz-dG and (b and d) *syn*-Bz-dG replication..... 127

Figure 4.5 Representative MD structures depicting the orientation of the purine dNTP in the Dpo4 ternary complexes for a) *anti*-Bz-dG:*syn*-dGTP, b) *syn*-Bz-dG:*anti*-dGTP, c) *anti*-Bz-dG:*syn*-dATP, and d) *syn*-Bz-dG:*anti*-dATP..... 129

Figure 4.6 Representative MD structures depicting possible hydrogen-bonding arrangements of Bz-dG with purine dNTPs for a) *anti*-Bz-dG: *syn*-dGTP, b) *syn*-Bz-dG:*anti*-dGTP, c) *anti*-Bz-dG:*syn*-dATP, and d) *syn*-Bz-dG:*anti*-dATP replication..... 130

Figure 4.7 Representative MD structures of the Dpo4 –1 base deletion complexes for a) *anti*-Bz-dG, b) *syn*-Bz-dG in the major groove conformation, and c) *syn*-Bz-dG in the intercalated conformation..... 132

Figure 5.1 The G, POB-G and PHB-G nucleosides in the *anti* glycosidic conformation 151

Figure 5.2 a) Overlay of POB-G and PHB-G conformations identified using DFT. b) Most stable POB-G and PHB-G structure identified for each conformational category (relative energies provided) 157

Figure 5.3 DFT structures and binding energies for the most stable hydrogen-bonded pairs between the extended POB-G or PHB-G conformation and each canonical nucleobase 159

Figure 5.4 a) MD representative structures obtained for POB-G (left) and PHB-G (right) adducted DNA based on an initial hydrogen-bonded orientation of the lesion with the pairing C in the major groove. b) Overlay of lesion conformations adopted throughout the MD simulation, highlighting the deviation in bulky moiety orientation (red). c) Lesion hydrogen-bonding interactions within DNA..... 163

Figure 5.5 a) MD representative structure of POB-G adducted DNA with the lesion in the extended orientation opposite C. b) Overlay of lesion conformations adopted throughout the MD simulation, highlighting the deviation in bulky moiety orientation (red). c) Lesion hydrogen-bonding interactions within DNA 165

Figure 5.6 a) MD representative structure of PHB-G adducted DNA with the lesion in the extended orientation opposite C. b) Overlay of lesion conformations adopted throughout the MD simulation, highlighting the deviation in bulky moiety orientation (red). c) Lesion hydrogen-bonding interactions within DNA 165

Figure 5.7 MD representative structures of POB-G adducted DNA with the bulky moiety in the major groove opposite a) T, b) A, or c) G. The trimer containing the lesion pair (green), flanking pairs (blue), and bulky moiety (red) is shown on the left, while the lesion-site hydrogen bonding is shown on the right 167

Figure 5.8 MD representative structures of PHB-G adducted DNA with the bulky moiety in the major groove opposite a) T, b) A, or c) G. The trimer containing the lesion pair (green), flanking pairs (blue), and bulky moiety (red) is shown on the left, while the lesion-site hydrogen bonding is shown on the right 168

Figure 5.9 Hydrogen bonding between POB-G or PHB-G and an incoming dCTP, dTTP, or dATP in the polymerase η insertion complex 171

Figure 5.10 Overlay of the polymerase η insertion complex for POB-G and PHB-G replication for dCTP, dTTP, or dATP insertion..... 172

Figure 6.1 a) The T, anti-POB-T and syn-POB-T nucleosides, with WC face in blue and Hoogsteen face in green. Key dihedral angles for the lesions include α′, β′, γ′, δ′, ϵ′, ρ′, and χ. b) POB-T base pairings previously speculated in the literature 182

Figure 6.2 Overlay of DFT POB-T nucleobase conformations (bulky moiety highlighted in red, top) and most stable conformer from each category (bottom)..... 185

Figure 6.3 Overlay of MD representative structures for a) anti-POB-T bound in the pol κ active site (purple) and b) anti-POB-T or c) syn-POB-T bound in the pol η active site (blue) with crystal structures (grey) for dATP insertion opposite dT (PDB ID: 5W2C for pol κ and 4ECS for pol η) 187

Figure 6.4 DFT structures and binding energies for base pairs between the a) T or POB-T WC face or b) POB-T Hoogsteen (H) face and the canonical nucleobases..... 189

Figure 6.5 Structure and reaction coordinates for pol κ insertion of dATP opposite a) T or b) POB-T, and c) *syn*-dATP or d) dGTP insertion opposite POB-T..... 192

Figure 6.6 Structure and reaction coordinates for pol η insertion of dATP opposite a) T or b) POB-T, and c) syn-dATP, d) dCTP or e) dTTP insertion opposite POB-T..... 195

Figure 6.7 a) MD representative structure of adducted DNA containing POB-T:T. b) Overlay (based on damaged T heavy atoms) of adducted base pair throughout the MD simulation (bulky moiety in red). c) Lesion pair structure in DNA..... 197

Figure 7.1 The proposed general reaction mechanism for nucleotide addition by a DNA polymerase..... 207

Figure 7.2 Solvated QM/MM model for pol η, with the ONIOM high-level region highlighted in tube representation..... 213

Figure 7.3 Comparison of the initial crystal structure geometry (PDB ID: 4ECS, green) and a MD representative structure on the a) wild-type enzyme (light purple) or b) the Ser113Ala mutant enzyme (dark purple), showing residues that interact with the dNTP (left) or the 3′-primer terminus (right). 214

Figure 7.4 Reactant complexes obtained from umbrella sampling, with the energy and torsion angle, provided 216

Figure 7.5 ONIOM product complex and key reaction distances for the nucleotidyl transfer reaction catalyzed by DNA pol η involving a hydroxide ion acting as the general base 218

Figure 7.6 ONIOM stationary points and key reaction distances for the nucleotidyl transfer reaction catalyzed by DNA pol η involving Ser113 mediated proton transfer to Glu116 220

Figure 7.7 ONIOM stationary points and key reaction distances for the nucleotidyl transfer reaction catalyzed by the Ser113Ala pol η mutant involving water mediated proton transfer to Glu116 221

Figure 7.8 ONIOM stationary points and key reaction distances for the nucleotidyl transfer reaction catalyzed by DNA pol η involving Glu116 acting as the general base 223

Figure 8.1 POB and PHB lesions formed upon exposure to NNK for which structural and mutational information is not known 236

Figure 8.2 Estragole, methyleugenol, and safrole DNA lesions proposed for future structural investigations 237

List of Abbreviations

AAF	acetylaminofluorene
ABA	nitrobenzanthrone
ABP	aminobiphenyl
AF	2-aminofluorene
AFB	aflatoxin B_1
AP	aminopyrene
BP	benzopyrene
BPh	benzo[c]phenanthrene
Bz	benzyl
CM	carboxymethyl
DFT	density functional theory
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
dPer	deoxyperimidinone
Dpo4	DNA polymerase IV
Et-dG	O6-ethyl-guanine
EQ	estronequinone
ES	estragole
IARC	International Agency for Research on Cancer
IQ	2-amino-3-methylimidazo[4,5-*f*]quinoline
MD	molecular dynamics
Me-dG/MeG	O6-methyl-guanine
MIE	methyleugenol
MM	molecular mechanics
MMGBSA	molecular mechanics Generalized Born surface area
MMPBSA	molecular mechanics Poisson-Boltzmann surface area
Naph	naphthyl
NBA	nitrobenzanthrone
NER	nucleotide excision repair

NMR	nuclear magnetic resonance
NNAL	4- (methylnitrosamino)-1-(3-pyridyl)-1-butanol
NNK	4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone
ONIOM	our own N-layered integrated molecular orbital and molecular mechanics
OTA	ochratoxin A
PAH	polycyclic aromatic hydrocarbon
PDB	protein data bank
PC	product complex
PCM	polarizable continuum model
PES	potential energy surface
POB	pridyloxobutyl
pol	polymerase
Ph	phenyl
PHB	pyridylhydroxybutyl
PhIP	2-amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine
QM	quantum mechanics
RC	reactant complex
RESP	restrained electrostatic potential
rmsd	root-mean-square deviation
S	safrole
TLS	translesion synthesis
TS	transition state
UV	ultraviolet
WC	Watson-Crick
WHAM	weighted histogram analysis method
WHO	World Health Organization

Nucleobase and Deoxyribonucleotide Triphosphate Abbreviations

dA/A	adenine	dATP	deoxyadenosine triphosphate
dC/C	cytosine	dCTP	deoxycytidine triphosphate
dG/G	guanine	dGTP	deoxyguanosine triphosphate
dT/T	thymine	dTTP	deoxythymidine triphosphate

Single- and Triple-Letter Abbreviations for the Amino Acids

A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartate, aspartic acid
E	Glu	Glutamate, glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

1. Chapter 1: Introduction[a]

1.1. General Overview

Deoxyribonucleic acid (DNA) contains the genetic information of a cell. Unfortunately, DNA can be damaged by many agents in the environment such as ultraviolet (UV) radiation, pollution, and pesticides.[1-6] When DNA is damaged many cellular processes are affected, such as DNA replication and transcription.[7, 8] Specifically, DNA lesions can block standard DNA replication and lead to deleterious effects including cell death.[7, 8] Fortunately, our cells have developed an alternative replication mechanism known as translesion synthesis (TLS).[9] Nevertheless, TLS is an error prone process and commonly introduces mutations into DNA upon replication past a site of damage.[9] These mutations can cause permanent changes to the genetic code, which can lead to diseases such as cancer.[10] This thesis concentrates on understanding the process of replicating damaged DNA by investigating the structure of tobacco-derived DNA lesions and their related mutagenic consequences. Furthermore, the mechanism of action of a key human TLS polymerase is examined. The following sections will introduce the structure of DNA, forms of DNA damage, process of DNA replication, and contributions of computational chemistry to the field, as well as summarize how a multiscale modeling approach has been used throughout this thesis to provide a greater understanding of TLS.

[a]*Chemical Research in Toxicology* reference style used throughout chapter.

1.2. DNA Structure

DNA is composed of two anti-parallel strands interwound in a helical-like fashion (Figure 1.1a). Canonical DNA typically adopts a B-form helix, although other forms including the A- and Z-forms, are also possible.[11] B-form DNA is characterized by a right-handed helix, 36° twist, 10 base-pairs per turn, and a C2′-*endo* sugar pucker.[11] Two distinct grooves present in B-DNA, denoted the major and minor grooves, allow for interactions between DNA and proteins.

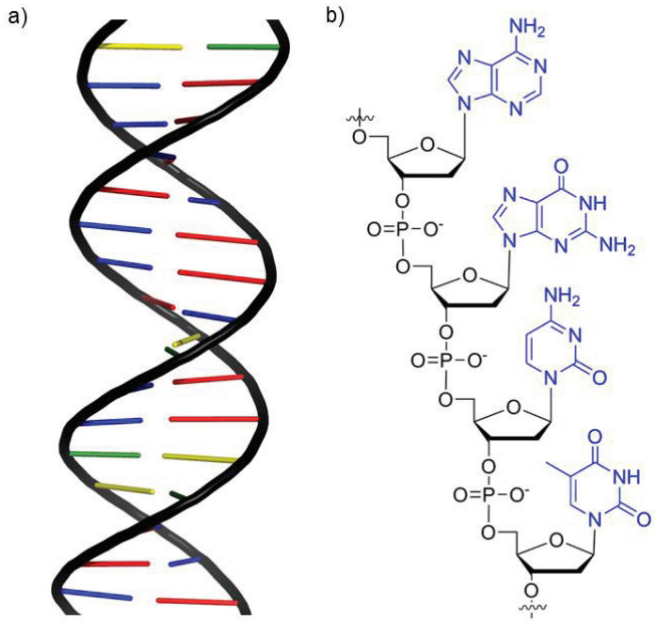

Figure 1.1. a) B-form DNA double helix and b) chemical structure of a single strand of DNA, with the DNA backbone (deoxyribose sugars and phosphate moieties) highlighted in black and the DNA nucleobases highlighted in blue.

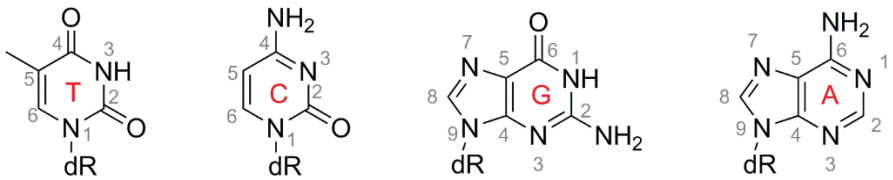

Figure 1.2. Structure and numbering of the canonical DNA nucleobases.

Each strand of DNA is composed of nucleotide subunits, which can be further decomposed into a nucleobase, deoxyribose sugar and phosphate moiety (Figure 1.1b). The genetic make-up of an organism is contained within the sequence of the DNA nucleobases. The canonical DNA nucleobases consist of the single-ringed pyrimidines, cytosine (C) and thymine (T), and the double-ring purines, guanine (G) and adenine (A, Figure 1.2). The chemical structure of the DNA nucleobases ensures selective Watson-Crick hydrogen bonding between single strands of DNA such that G pairs with C, and A pairs with T in the DNA double helix (Figure 1.3). These Watson-Crick base pairs require nucleosides to adopt the *anti* conformation about the glycosidic bond ($\chi = 180 \pm 90°$; $\chi = \angle(O4'C1'N9C4)$ for purines and $\chi = \angle(O4'C1'N1C2)$ for pyrimidines; Figure 1.4). When the base rotates about the glycosidic bond to adopt the *syn* conformation, the Hoogsteen face of the nucleobase is aligned to hydrogen bond with a pairing base. In DNA, this conformational change about the glycosidic bond commonly occurs upon DNA damage and frequently introduces base mispairs. Specifically, DNA damage can lead to nucleobase transitions (substitution of a purine for a purine or a pyrimidine for a pyrimidine), nucleobase transversions (substitution of a purine for a pyrimidine or a pyrimidine for a purine) or frameshift (insertion of one less or one more nucleobase into the primer strand than is present in the template strand) mutations.

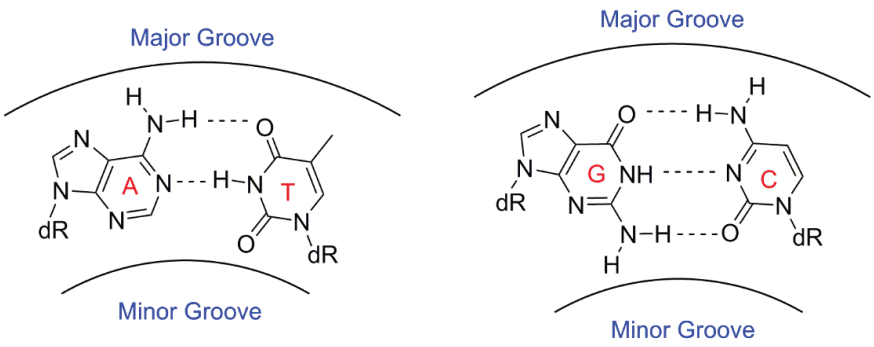

Figure 1.3. Watson-Crick A:T and G:C base pairs, with the location of the DNA major and minor grooves highlighted.

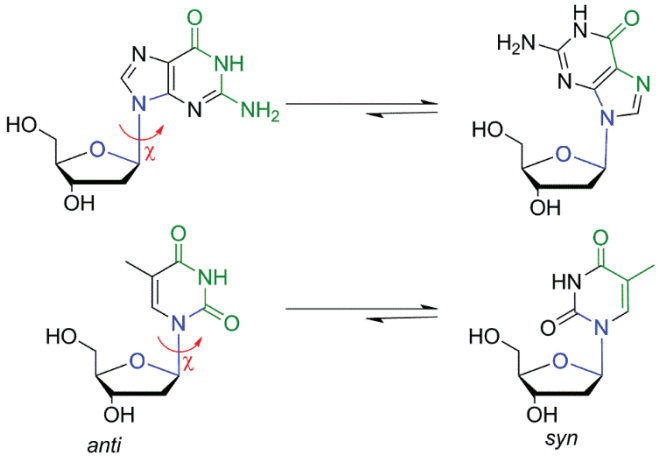

Figure 1.4. The *anti* (left) and *syn* (right) conformations of a DNA nucleoside. The χ dihedral angle is highlighted in blue and the hydrogen-bonding face (Watson-Crick face for *anti* or Hoogsteen face for *syn*) is highlighted in green.

1.3. DNA Damage

Unfortunately, DNA is damaged by both external (e.g., pollution, UV radiation, and agricultural contaminants) and internal (e.g., estrogen, reactive oxygen species, and *S*-adenosylmethionine) agents.[1-6] This damage can occur to any component of the DNA strand. For example, the DNA backbone can be cleaved (DNA strand breaks) or a nucleobase can be lost (formation of an abasic site).[12] Furthermore, chemical groups can be added to the DNA backbone or any nucleobase. Addition of a chemical group can be further divided based on the nature of the chemical reaction. For example, DNA

nucleobases can undergo oxidation (e.g., oxidation of G to generate 8-oxoguanine),[13] deamination (e.g., deamination of C to form uracil),[5] alkylation (e.g., methylation of A to yield 3-methyladenine),[14] cross-linking (e.g., covalently linking two adjacent intra or interstrand T nucleobases to yield a T-T dimer),[15] or addition of a larger chemical group to form a bulky DNA addition product (adduct, Figure 1.5).[16] Additionally, the DNA nucleobase damage site can be classified based whether the chemical modification occurs to an atom that is positioned in the major or minor groove when the lesion is in the canonical *anti* glycosidic conformation (Figures 1.3 and 1.5).

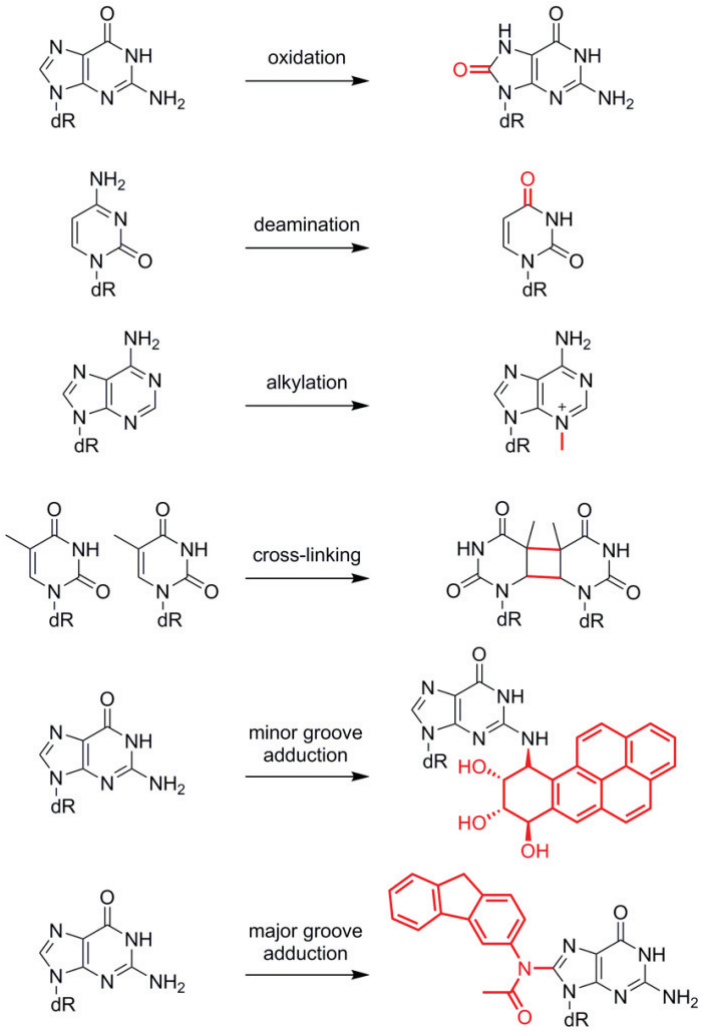

Figure 1.5. Examples of different classes of DNA damage.

1.4. Tobacco Induced DNA Damage

Among the vast array of DNA damaged products identified to date, this thesis focuses on lesions formed upon exposure to tobacco products. There is an estimated 1 billion people smoking tobacco worldwide at the beginning of the 21st century, and it is estimated that tobacco kills more than 7 million people per year.[17] Indeed, tobacco use is linked with many different cancers including lung, esophageal, laryngeal, oral, throat, kidney, bladder, pancreatic, stomach, and cervical.[17] Due to the known carcinogenic properties of conventional cigarettes, alternative tobacco products including hookah, electronic cigarettes, chewing tobacco, and the nicotine patch, are frequently advertised. Nevertheless, these alternative products still contain carcinogens,[17, 18] and the levels of the carcinogens in smokeless tobacco products are typically higher than conventional cigarettes.[19, 20] Additionally, marijuana is the third most popular recreational drug following alcohol and tobacco, and recent changes in legislation in Canada may lead to a further increase in the popularity of this drug.[21, 22] The most common method of marijuana use is smoking and although the exact chemical composition of marijuana smoke is not yet known, it is believed to contain many of the same toxic components as tobacco smoke including polyaromatic hydrocarbons (PAHs) and aromatic amines.[21, 23] Furthermore, an increasingly popular method of consuming marijuana is through smoking blunts and spliffs/mulled cigarettes, which combine tobacco and marijuana.[24, 25]

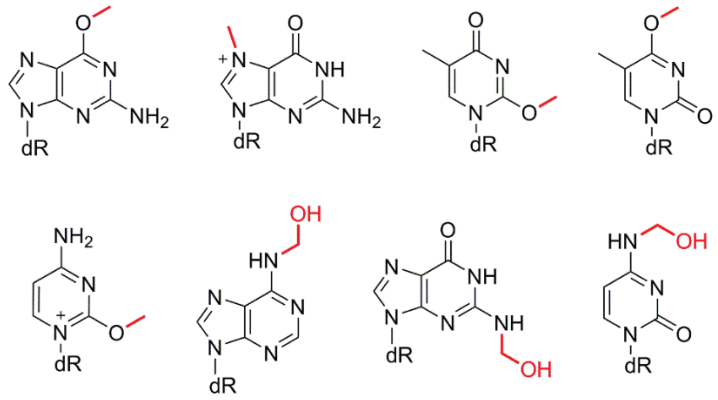

Figure 1.6. Methyl and formaldehyde lesions formed upon exposure to NNK.

Tobacco smoke contains thousands of chemicals, 70 of which are known carcinogens including tobacco specific nitrosamines (e.g., 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (also known as nicotine-derived nitrosaminoketone, NNK) and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol (NNAL)), aromatic amines (e.g., 1-aminonaphthalene, 2-aminonaphthalene, 3-aminobiphenyl, and 4-aminobiphenyl), and PAHs (e.g., benzo[*a*]pyrene, benzo[*b*]fluoranthene, benzo[*j*]fluoranthene, and benzo[*k*]fluoranthene).[26, 27] NNK is believed to be the main source of tobacco induced carcinogenesis.[28-30] Indeed, NNK has been shown to cause cancer in every species tested (i.e., mice, rats, hamsters, rabbits, pigs, monkeys, and humans) regardless of the route of administration.[31, 32] Upon metabolic activation, NNK forms a number of DNA lesions including methyl (N7-G, O6-G, N3-A, N1-A, N7-A, N3-C, N3-G, N3-T, O2-T, O4-T, and O4-C; Figure 1.6), formaldehyde (N6-A, N4-C, or N2-G; Figure 1.6), and pyridyloxobutyl (POB)/pyridylhydroxybutyl (PHB; O2-T, O4-T, N6-A, O6-G, N7-G, O2-C, N3-C, N4-C or phosphate; Figure 1.7) adducts.[30]

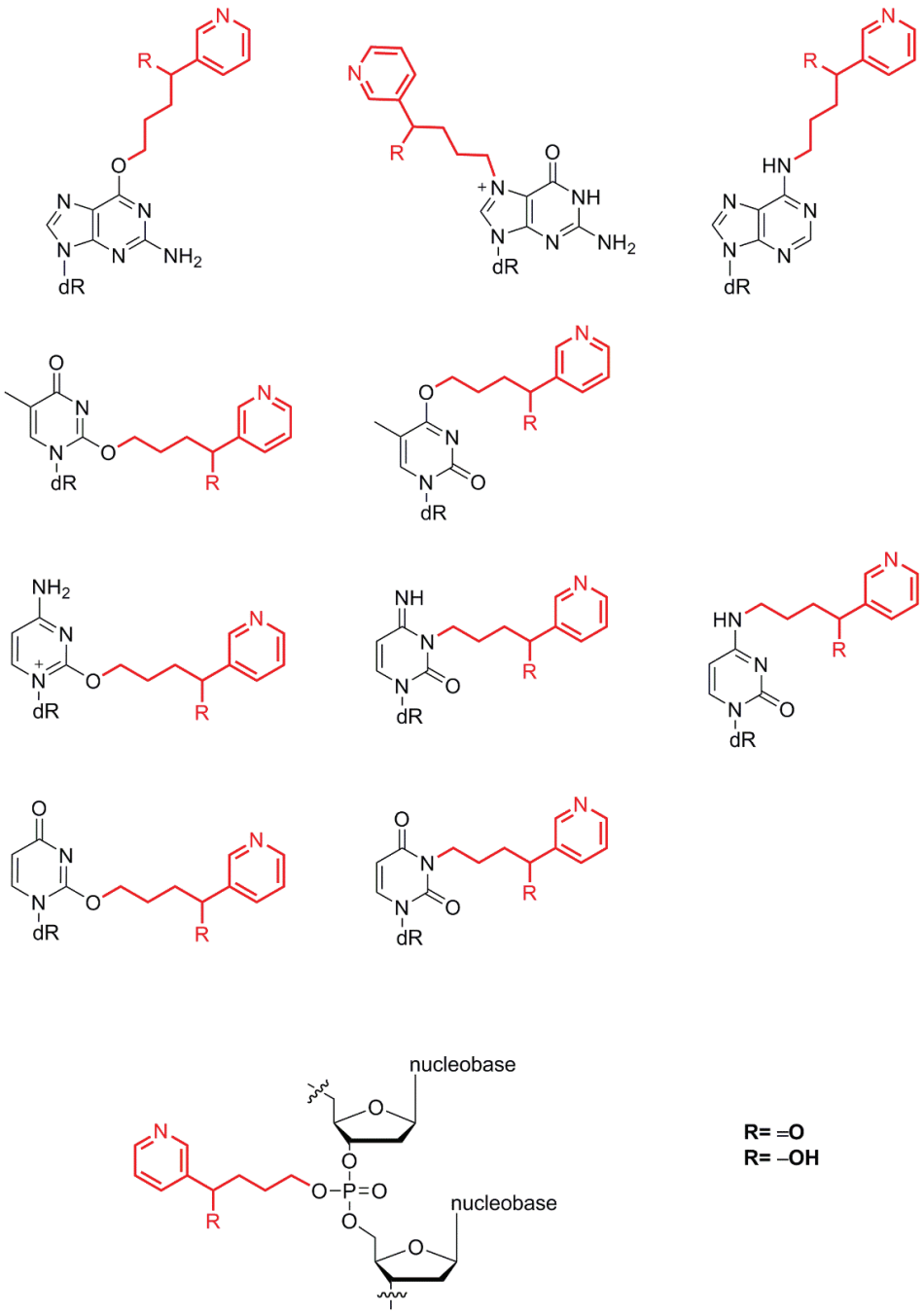

Figure 1.7. Pyridyloxobutyl (R= =O) and pyridylhydroxybutyl (R= –OH) lesions formed upon exposure to NNK.

Investigation of the mutagenicity of some of these lesions formed from NNK has been completed. Specifically, although the major methyl lesion, N7-Me-G, is not mutagenic,[33, 34] the second most prevalent tobacco-induced methyl lesion, O6-Me-G, lead to G → A

transition mutations.[35, 36] There is also evidence of mutagenic replication of some of the minor methyl lesions including O4-Me-T and O2-Me-T, which both lead to T → C and T → A mutations.[30] Additionally, although the N4-C and the N2-G formaldehyde lesions are non-mutagenic, the N6-A formaldehyde adduct has been reported to lead to A → T and A → C mutations.[89] Finally, mutagenicities of most of the POB and PHB lesions has not been studied to date. However, it is known that POB-G and POB-T lead to G → A[37-39] and T → A mutations,[40-46] respectively. Nevertheless, in order to gain a greater understanding of the mutagenicity of these lesions the pathways that process them must be understood.

1.5. DNA Replication

One such pathway that processes DNA is DNA replication. Although many enzymes are involved in the overall process, DNA polymerases are responsible for catalyzing the nucleotidyl transfer reaction that adds a nucleotide to the 3′ end of the primer strand.[47, 48] In the metal-ion assisted nucleotidyl transfer reaction, a general base deprotonates the primer 3′-OH, and the 3′-oxoanion subsequently attacks the dNTP α-phosphate moiety to form a phosphodiester bond and release pyrophosphate (Figure 1.8).[49-68] The nucleotide triphosphate is selected based on hydrogen-bonding interactions with the pairing template nucleobase and the steric constraints of the polymerase active site. Nevertheless, DNA damage sites can alter the base pairing preferences and sterics within the active site and therefore can lead to base-substitution mutations.

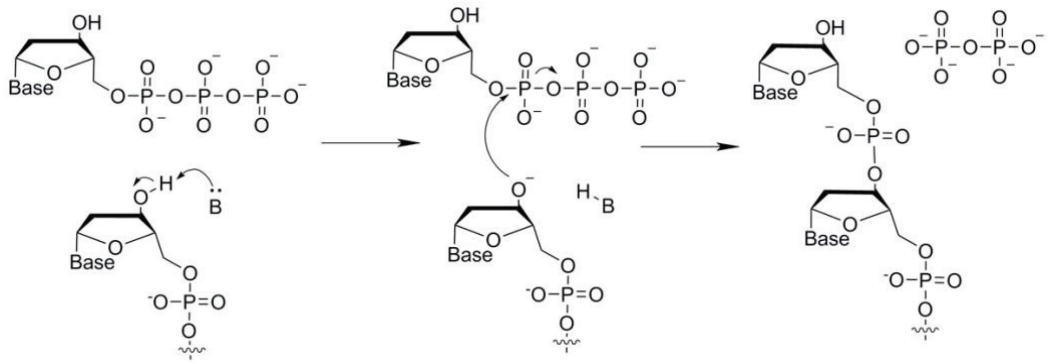

Figure 1.8. General mechanism for the nucleotidyl reaction mechanism catalyzed by DNA polymerases.

Besides leading to the incorrect incorporation of a nucleobase, DNA damage sites can block standard DNA replication due to the high substrate specificity of standard (replicative) polymerases. Stalled replication forks are highly unstable and can lead to deleterious consequences such as chromosomal rearrangements or cell death.[69, 70] Fortunately, an alternative pathway exists to replicate damaged DNA called translesion synthesis. Translesion synthesis (TLS) polymerases have more flexible active sites that permit replication of damaged DNA, which also leads to less accurate enzymes compared to replicative polymerases.[7] Indeed, the error rate of TLS polymerases is 10^{-3} to 10^{-4} compared to $< 10^{-9}$ for replicative polymerases.[7]

The differences between TLS and replicative polymerases are related to the structures of the enzymes.[71] In general, replicative polymerases contain a thumb domain to guide the newly formed dsDNA, a fingers domain to bind the incoming dNTP, a palm domain to catalyze the nucleotidyl transfer reaction, one or two exonuclease domains to proofread the newly synthesized DNA, and an amino-terminal domain to interact with other proteins (Figure 1.9).[71] In contrast, TLS polymerases have four domains, three of which (thumb,

fingers, and palm) are smaller in size, but have the same function, as in replicative polymerases. The little finger domain, or the polymerase associated domain, is unique to TLS polymerases and helps bind DNA (Figure 1.9).[71] Furthermore, the little finger domain has more sequence variability than the other domains and therefore is believed to play a role in determining substrate specificity.

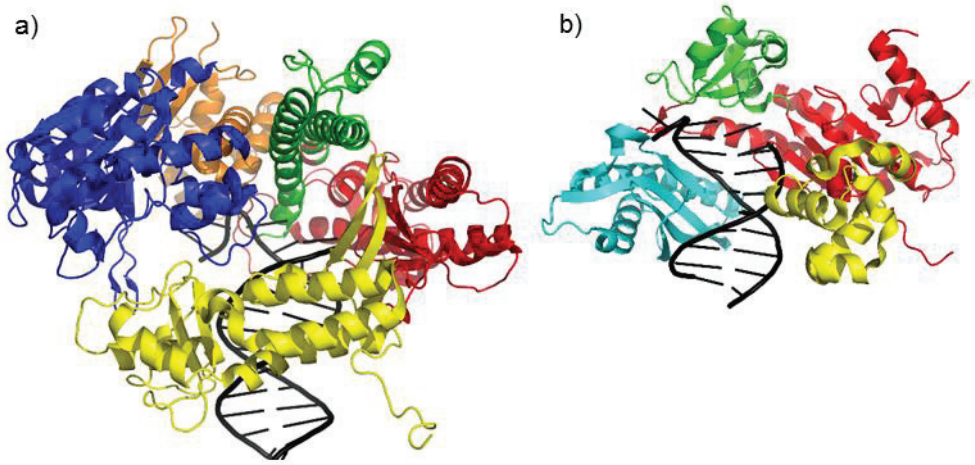

Figure 1.9. Structure of a) a replicative DNA polymerase (bacteriophage RB69, PDB ID: 1IG9), and b) a TLS polymerase (human DNA polymerase η, PDB ID: 4ECS), highlighting the various domains (little finger=cyan, thumb=yellow, palm=red, fingers=green, exonuclease=blue, and amino-terminal=orange).

Although all TLS polymerases are structurally related, the active site of eukaryotic TLS polymerases have different structural features. Therefore, each of the four eukaryotic TLS polymerases (Rev1, pol η, pol κ, and pol ι) are believed to be specific toward certain DNA lesions. Specifically, Rev1 inserts C opposite the template base regardless of the identity of the base (e.g., N2-G lesions or abasic sites).[72, 73] Furthermore, pol η is a key TLS polymerase due to the variety of lesions it has been shown to bypass, including pyrimidine dimers and major groove lesions,[74, 75] as well as the role of pol η in disease.[76, 77] Specifically, pol η is the only polymerase in which mutations to the enzyme have been directly linked

with cancer.[76, 77] The proposed role of pol κ in the cell is to replicate minor groove lesions, and act as an extender polymerase (i.e., pol κ extends DNA past a damage site that has been replicated by another polymerase).[9] Finally, pol ι has been proposed to use Hoogsteen hydrogen bonding to insert a nucleotide opposite major groove adducts, although cases in which pol ι replicates lesions using Watson-Crick base pairing have also been reported.[78] Besides the four discussed eukaryotic TLS polymerases, many other TLS polymerases exist as this class of enzymes is found in all domains of life.[71] Notably, the archaeal DNA polymerase IV (Dpo4) from *Sulfolobus solfataricus* was the first TLS polymerase to be crystallized,[79] and therefore a large portion of the key information about TLS has been gained by studying this system. Nevertheless, many questions regarding the function and substrate specificity of TLS polymerases still remain to be answered.

1.6. Previous Computational Studies on DNA Adducts and TLS

Computational chemistry has proven to be an effective tool to complement experimental studies. In particular, molecular modeling has provided atomic level details about DNA lesions and DNA–enzyme systems necessary to understand the structure of adducted DNA and the associated enzymatic processing.[16, 80] Previous work on DNA lesions has used quantum mechanical (QM) calculations to investigate the conformational preferences of damaged nucleobases about bulky moiety–nucleobase linkers and within the bulky moiety (see reference 16 and citations therein). Additionally, QM studies have investigated the conformational preference about the glycosidic bond for adducted nucleosides. Using the information obtained about the intrinsic conformational properties of the DNA adduct, the base-pairing preferences and damage site orientation in DNA duplexes have been modeled using molecular dynamics (MD) simulations.[16] These studies have been completed on a

wide range of DNA lesions, such as aromatic amines,[6] PAH,[81-88] aristolochic acid,[89-91] and phenolic lesions,[92-95] and have helped to begin to develop a structure–outcome relationship for DNA adducts. Additionally, information on the intrinsic lesion conformation and lesion orientation in DNA is critical for modeling DNA-enzymatic systems that process these damaged sites.

In order to gain insight into the effects of DNA adducts on replication, including the mutagenic consequences of lesion formation, MD simulations have been used (see reference 80 and citations therein). Specifically, MD simulations have investigated the local structure of the polymerase (i.e., Dpo4, pol ι, and pol κ) active site containing the well-studied aromatic amine and PAH adducts.[96-107] These works have considered the different possible conformations of adducts in the polymerase active sites and indicated that the preferred orientation of the lesion can be altered relative to that adopted in the DNA helix.[96-98] Additionally, the orientation and base pairing of the lesions in the active site predicted from molecular modeling have provided insight into experimentally observed replication outcomes. Specifically, non-mutagenic replication of AAF-G,[97, 98] as well as non-mutagenic and mutagenic replication of PhIP-G (i.e., G→A)[99] and S-B[c]Ph-A (i.e., A→C and A→T) has been observed.[100, 101] Conversely, S-B[a]P-A cannot be accommodated in the pol κ active site, which correlates with the experimentally-observed full blockage of the polymerase by this lesion.[102] Overall, although computational work on the replication of adducted DNA highlights that numerous factors play a role in determining whether mutagenic or non-mutagenic adduct replication will take place, further work is required to fully appreciate how the chemical composition of adducts dictate TLS outcomes.

To uncover details of the mechanisms used by enzymatic systems, hybrid QM/molecular mechanics (MM) approaches have been widely used.[108-110] In terms of DNA polymerases, previous studies have investigated critical aspects of the nucleotidyl transfer reaction catalyzed by DNA polymerases (i.e., DNA pol β, κ, λ, ι, and η, as well as T7 DNA polymerase and DNA polymerase IV) including the effects of the metal ions, roles of active site residues, and identity of the general base.[49-68] Although this work has given some key insights into the reaction mechanism, such as the role of the metal ions and the roles of active site residues in stabilizing the dNTP during catalysis, many different and contradictory conclusions have been made about the identity of the general base. Therefore, further work is required to fully understand the catalyzed reaction by systematically investigating the different possibilities for the general base in the nucleotidyl transfer reaction.

1.7. Scope of Thesis

This thesis uses computational chemistry to further understand the replication of highly flexible tobacco lesions by TLS polymerases. Specifically, methodology suitable for studying the structure and mutagenicity of highly flexible lesions is developed in Chapters 2 – 4 by investigating O6-benzyl-guanine (Bz-G). Bz-G is a synthetic lesion that serves as a simple, yet informative, model to gain an understanding of the replication of O6-G lesions. Chapter 2 begins with an in-depth investigation of the structure of the Bz-G nucleobase, nucleoside, and adducted DNA duplex. Chapter 3 investigates the replication of Bz-G by a prototypical DNA polymerase (Dpo4) during different stages of the TLS process. Finally, Chapter 4 uses MD simulations to investigate a number of structural and

energetic parameters that dictate the preferred dNTP insertion opposite the lesion. These chapters collectively demonstrate the importance of understanding the conformational flexibility of a DNA lesion with different models, as well as obtaining structural information at each stage during the TLS process, and provides the first structural explanation for the experimental mutagenesis spectrum for Bz-G.

Due to the success of the multiscale modeling approach developed in Chapters 2 – 4, the same method is used to investigate the replication of tobacco lesions by human DNA polymerases in Chapters 5 and 6. Specifically, Chapter 5 provides the first structural information on the replication of POB-G and PHB-G, by investigating pol η insertion complexes and post-replication DNA duplexes. This work highlights that a change in a single substituent in the bulky moiety of a DNA adduct can affect the lesion mutagenicity. Subsequently, Chapter 6 clarifies the previously reported replication outcomes for the minor groove POB-T lesion by explaining how two human polymerases (pol κ and η) process the lesion. Specifically, although pol κ typically bypasses minor groove adducts, *in vivo* studies indicate pol η replicates POB-T despite being known for processing major groove lesions. By uncovering novel structural features of POB-T replication, this work explains the unexpected experimental mutational spectrum and presents the first evidence for replication of a *syn* pyrimidine.

In Chapter 7, a greater understanding of the mechanism of action of pol η is obtained through a combined MD, umbrella sampling and QM/MM approach. This chapter clarifies the enzyme reaction mechanism by considering all possibilities for the general base by investigating previous proposals for pol η and other polymerases. Finally, Chapter 8

summarizes how this thesis has provided a greater understanding of the replication of highly flexible DNA lesions through a multiscale modeling approach. This chapter also outlines future research avenues that will build upon the current work to gain a more comprehensive understanding of the replication of damaged DNA.

1.8. Reference

(1) Helleday, T., Eshtad, S., and Nik-Zainal, S. (2014) Mechanisms underlying mutational signatures in human cancers. *Nat. Rev. Genet. 15*, 585-598.

(2) De Bont, R., and van Larebeke, N. (2004) Endogenous DNA damage in humans: A review of quantitative data. *Mutagenesis 19*, 169-185.

(3) Geacintov, N. E., Cosman, M., Hingerty, B. E., Amin, S., Broyde, S., and Patel, D. J. (1997) NMR solution structures of stereoisomeric polycyclic aromatic carcinogen-DNA adducts: Principles, patterns, and diversity. *Chem. Res. Toxicol. 10*, 111-146.

(4) Hingerty, B. E., and Broyde, S. (1990) Atomic resolution structures of DNA and DNA modified by carcinogens. *Int. J. High Perform. Comput. Appl. 4*, 11-21.

(5) Gates, K. S. (2009) An overview of chemical processes that damage cellular DNA: Spontaneous hydrolysis, alkylation, and reactions with radicals. *Chem. Res. Toxicol. 22*, 1747-1760.

(6) 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.

(7) Sale, J. E., Lehmann, A. R., and Woodgate, R. (2012) Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat. Rev. Mol. Cell Biol. 13*, 141-152.

(8) Xu, L., Wang, W., Chong, J., Shin, J. H., Xu, J., and Wang, D. (2015) RNA polymerase II transcriptional fidelity control and its functional interplay with DNA modifications. *Crit. Rev. Biochem. Mol. Biol. 50*, 503-519.

(9) Zhao, L., and Washington, M. (2017) Translesion synthesis: Insights into the selection and switching of DNA polymerases. *Genes 8*, 24.

(10) Zafar, M. K., and Eoff, R. L. (2017) Translesion DNA synthesis in cancer: Molecular mechanisms and therapeutic opportunities. *Chem. Res. Toxicol. 30*, 1942-1955.

(11) Voet, D., and Voet, J. G. (2011) *Biochemistry*. 4 ed., John Wiley & Sons.

(12) Dizdaroglu, M., Jaruga, P., Birincioglu, M., and Rodriguez, H. (2002) Free radical-induced damage to DNA: Mechanisms and measurement. *Free Radical Biol. Med. 32*, 1102-1115.

(13) Kryston, T. B., Georgiev, A. B., Pissis, P., and Georgakilas, A. G. (2011) Role of oxidative stress and DNA damage in human carcinogenesis. *Mutat. Res. 711*, 193-201.

(14) Fu, D., Calvo, J. A., and Samson, L. D. (2012) Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nat. Rev. Cancer 12*, 104-120.

(15) Cadet, J., Sage, E., and Douki, T. (2005) Ultraviolet radiation-mediated damage to cellular DNA. *Mutat. Res. Fund. Mol. Mech. Mut. 571*, 3-17.

(16) Wilson, K., Kung, R., and Wetmore, S. (2016) Toxicology of DNA adducts formed upon human exposure to carcinogens: Insights gained from molecular modeling. *Advances in Molecular Toxicology 10*, 293.

(17) Bilano, V., Gilmour, S., Moffiet, T., d'Espaignet, E. T., Stevens, G. A., Commar, A., Tuyl, F., Hudson, I., and Shibuya, K. (2015) Global trends and projections for tobacco use, 1990–2025: An analysis of smoking indicators from the WHO comprehensive information systems for tobacco control. *The Lancet 385*, 966-976.

(18) Harris, C. C. (2018) Tobacco smoking, e-cigarettes, and nicotine harm. *Proc. Natl. Acad. Sci. U. S. A.*

(19) Hecht, S. S., and Szabo, E. (2014) Fifty years of tobacco carcinogenesis research: From mechanisms to early detection and prevention of lung cancer. *Cancer Prev. Res. 7*, 1-8.

(20) Health, U. D. o., and Services, H. (2014) The health consequences of smoking—50 years of progress: A report of the surgeon general. *Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health 17*.

(21) Maertens, R. M., White, P. A., Rickert, W., Levasseur, G., Douglas, G. R., Bellier, P. V., McNamee, J. P., Thuppal, V., Walker, M., and Desjardins, S. (2009) The genotoxicity of mainstream and sidestream marijuana and tobacco smoke condensates. *Chem. Res. Toxicol. 22*, 1406-1414.

(22) Wei, B., Alwis, K. U., Li, Z., Wang, L., Valentin-Blasini, L., Sosnoff, C. S., Xia, Y., Conway, K. P., and Blount, B. C. (2016) Urinary concentrations of pah and voc metabolites in marijuana users. *Environ. Int. 88*, 1-8.

(23) Moir, D., Rickert, W. S., Levasseur, G., Larose, Y., Maertens, R., White, P., and Desjardins, S. (2007) A comparison of mainstream and sidestream marijuana and tobacco cigarette smoke produced under two machine smoking conditions. *Chem. Res. Toxicol. 21*, 494-502.

(24) Schauer, G. L., Rosenberry, Z. R., and Peters, E. N. (2017) Marijuana and tobacco co-administration in blunts, spliffs, and mulled cigarettes: A systematic literature review. *Addict. Behav. 64*, 200-211.

(25) Peters, E. N., Schauer, G. L., Rosenberry, Z. R., and Pickworth, W. B. (2016) Does marijuana "blunt" smoking contribute to nicotine exposure?: Preliminary product testing of nicotine content in wrappers of cigars commonly used for blunt smoking. *Drug Alcohol Depend. 168*, 119-122.

(26) Xue, J., Yang, S., and Seng, S. (2014) Mechanisms of cancer induction by tobacco-specific NNK and NNN. *Cancers 6*, 1138.

(27) Hecht, S. S. (1999) Tobacco smoke carcinogens and lung cancer. *J. Natl. Cancer Inst. 91*, 1194-1210.

(28) Hecht, S. S. (2003) Tobacco carcinogens, their biomarkers and tobacco-induced cancer. *Nat. Rev. Cancer 3*, 733.

(29) Lee, H.-W., Park, S.-H., Weng, M.-w., Wang, H.-T., Huang, W. C., Lepor, H., Wu, X.-R., Chen, L.-C., and Tang, M.-s. (2018) E-cigarette smoke damages DNA and reduces repair activity in mouse lung, heart, and bladder as well as in human lung and bladder cells. *Proc. Natl. Acad. Sci. U. S. A.*

(30) Peterson, L. A. (2016) Context matters: Contribution of specific DNA adducts to the genotoxic properties of the tobacco-specific nitrosamine NNK. *Chem. Res. Toxicol.*

(31) Hecht, S. S. (1998) Biochemistry, biology, and carcinogenicity of tobacco-specific N-nitrosamines. *Chem. Res. Toxicol. 11*, 559-603.

(32) Hecht, S. S. (2008) Progress and challenges in selected areas of tobacco carcinogenesis. *Chem. Res. Toxicol. 21*, 160-171.

(33) Fu, D., Calvo, J. A., and Samson, L. D. (2012) Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nat. Rev. Cancer 12*, 104-120.

(34) Larson, K., Sahm, J., Shenkar, R., and Strauss, B. (1985) Methylation-induced blocks to in vitro DNA replication. *Mutat. Res. 150*, 77-84.

(35) Pauly, G. T., and Moschel, R. C. (2001) Mutagenesis by O6-methyl-, O6-ethyl-, and O6-benzylguanine and O4-methylthymine in human cells: Effects of O6-alkylguanine-DNA alkyltransferase and mismatch repair. *Chem. Res. Toxicol. 14*, 894-900.

(36) Altshuler, K. B., Hodes, C. S., and Essigmann, J. M. (1996) Intrachromosomal probes for mutagenesis by alkylated DNA bases replicated in mammalian cells: A

comparison of the mutagenicities of O4-methylthymine and O6-methylguanine in cells with different DNA repair backgrounds. *Chem. Res. Toxicol. 9*, 980-987.

(37) Pauly, G. T., Peterson, L. A., and Moschel, R. C. (2002) Mutagenesis by O6-[4-oxo-4-(3-pyridyl)butyl]guanine in Escherichia coli and human cells. *Chem. Res. Toxicol. 15*, 165-169.

(38) Choi, J.-Y., Chowdhury, G., Zang, H., Angel, K. C., Vu, C. C., Peterson, L. A., and Guengerich, F. P. (2006) Translesion synthesis across O6-alkylguanine DNA adducts by recombinant human DNA polymerases. *J. Biol. Chem. 281*, 38244-38256.

(39) Gowda, A. S. P., and Spratt, T. E. (2016) DNA polymerase ν rapidly bypasses O6-methyl-dG but not O6-[4-(3-pyridyl)-4-oxobutyl-dG and O2-alkyl-dTs. *Chem. Res. Toxicol. 29*, 1894-1900.

(40) Gowda, A. S. P., Krishnegowda, G., Suo, Z., Amin, S., and Spratt, T. E. (2012) Low fidelity bypass of O2-(3-pyridyl)-4-oxobutylthymine, the most persistent bulky adduct produced by the tobacco specific nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone by model DNA polymerases. *Chem. Res. Toxicol. 25*, 1195-1202.

(41) Gowda, A. S. P., and Spratt, T. E. (2016) DNA polymerases η and ζ combine to bypass O(2)-[4-(3-pyridyl)-4-oxobutyl]thymine, a DNA adduct formed from tobacco carcinogens. *Chem. Res. Toxicol. 29*, 303-316.

(42) Hashimoto, K., Ohsawa, K.-i., and Kimura, M. (2004) Mutations induced by 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) in the lacz and cii genes of muta™ mouse. *Mutat. Res./Genet. Toxicol. Environ. Mutagen. 560*, 119-131.

(43) Sandercock, L. E., Hahn, J. N., Li, L., Luchman, H. A., Giesbrecht, J. L., Peterson, L. A., and Jirik, F. R. (2008) Mgmt deficiency alters the in vivo mutational spectrum of tissues exposed to the tobacco carcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK). *Carcinogenesis 29*, 866-874.

(44) Jasti, V. P., Spratt, T. E., and Basu, A. K. (2011) Tobacco-specific nitrosamine-derived O2-alkylthymidines are potent mutagenic lesions in SOS-induced Escherichia coli. *Chem. Res. Toxicol. 24*, 1833-1835.

(45) Weerasooriya, S., Jasti, V. P., Bose, A., Spratt, T. E., and Basu, A. K. (2015) Roles of translesion synthesis DNA polymerases in the potent mutagenicity of tobacco-specific nitrosamine-derived O(2)-alkylthymidines in human cells. *DNA repair 35*, 63-70.

(46) Du, H., Leng, J., Wang, P., Li, L., and Wang, Y. (2018) Impact of tobacco-specific nitrosamine-derived DNA adducts on the efficiency and fidelity of DNA replication in human cells. *J. Biol. Chem.*

(47) Patel, P. H., and Loeb, L. A. (2001) Getting a grip on how DNA polymerases function. *Nat. Struct. Mol. Biol. 8*, 656.

(48) Steitz, T. A. (1998) Structural biology: A mechanism for all polymerases. *Nature 391*, 231-232.

(49) Florián, J., Goodman, M. F., and Warshel, A. (2003) Computer simulation of the chemical catalysis of DNA polymerases: Discriminating between alternative nucleotide insertion mechanisms for T7 DNA polymerase. *J. Am. Chem. Soc. 125*, 8163-8177.

(50) Rittenhouse, R. C., Apostoluk, W. K., Miller, J. H., and Straatsma, T. (2003) Characterization of the active site of DNA polymerase β by molecular dynamics and quantum chemical calculation. *Proteins: Struct., Funct., Bioinf. 53*, 667-682.

(51) Lin, P., Pedersen, L. C., Batra, V. K., Beard, W. A., Wilson, S. H., and Pedersen, L. G. (2006) Energy analysis of chemistry for correct insertion by DNA polymerase β. *Proc. Natl. Acad. Sci. U. S. A. 103*, 13294-13299.

(52) Radhakrishnan, R., and Schlick, T. (2006) Correct and incorrect nucleotide incorporation pathways in DNA polymerase β. *Biochem. Biophys. Res. Commun. 350*, 521-529.

(53) Alberts, I. L., Wang, Y., and Schlick, T. (2007) DNA polymerase β catalysis: Are different mechanisms possible? *J. Am. Chem. Soc. 129*, 11100-11110.

(54) Batra, V. K., Perera, L., Lin, P., Shock, D. D., Beard, W. A., Pedersen, L. C., Pedersen, L. G., and Wilson, S. H. (2013) Amino acid substitution in the active site of DNA polymerase β explains the energy barrier of the nucleotidyl transfer reaction. *J. Am. Chem. Soc. 135*, 8078-8088.

(55) Cisneros, G. A., Perera, L., García-Díaz, M., Bebenek, K., Kunkel, T. A., and Pedersen, L. G. (2008) Catalytic mechanism of human DNA polymerase λ with Mg2+ and Mn2+ from ab initio quantum mechanical/molecular mechanical studies. *DNA Repair 7*, 1824-1834.

(56) Abashkin, Y. G., Erickson, J. W., and Burt, S. K. (2001) Quantum chemical investigation of enzymatic activity in DNA polymerase β. A mechanistic study. *J. Phys. Chem. B 105*, 287-292.

(57) Bojin, M. D., and Schlick, T. (2007) A quantum mechanical investigation of possible mechanisms for the nucleotidyl transfer reaction catalyzed by DNA polymerase β. *J. Phys. Chem. B 111*, 11244-11252.

(58) Lior-Hoffmann, L., Wang, L., Wang, S., Geacintov, N. E., Broyde, S., and Zhang, Y. (2012) Preferred wmsa catalytic mechanism of the nucleotidyl transfer reaction in human DNA polymerase κ elucidates error-free bypass of a bulky DNA lesion. *Nucleic Acids Res. 40*, 9193-9205.

(59) Prasad, B. R., Plotnikov, N. V., and Warshel, A. (2013) Addressing open questions about phosphate hydrolysis pathways by careful free energy mapping. *J. Phys. Chem. B 117*, 153-163.

(60) Lior-Hoffmann, L., Ding, S., Geacintov, N. E., Zhang, Y., and Broyde, S. (2014) Structural and dynamic characterization of polymerase κ's minor groove lesion processing reveals how adduct topology impacts fidelity. *Biochemistry 53*, 5683-5691.

(61) Wang, Y., and Schlick, T. (2008) Quantum mechanics/molecular mechanics investigation of the chemical reaction in Dpo4 reveals water-dependent pathways and requirements for active site reorganization. *J. Am. Chem. Soc. 130*, 13240-13250.

(62) Wang, L., Yu, X., Hu, P., Broyde, S., and Zhang, Y. (2007) A water-mediated and substrate-assisted catalytic mechanism for sulfolobus solfataricus DNA polymerase iv. *J. Am. Chem. Soc. 129*, 4731-4737.

(63) Li, Y., Bao, L., Zhang, R., Tang, X., Zhang, Q., and Wang, W. (2017) Insights into the error bypass of 1-nitropyrene DNA adduct by DNA polymerase ι: A QM/MM study. *Chem. Phys. Lett. 686*, 12-17.

(64) Hanwool, Y., and Arieh, W. (2017) Simulating the fidelity and the three mg mechanism of pol η and clarifying the validity of transition state theory in enzyme catalysis. *Proteins: Struct., Funct., Bioinf. 85*, 1446-1453.

(65) Stevens, D. R., and Hammes-Schiffer, S. (2018) Exploring the role of the third active site metal ion in DNA polymerase η with QM/MM free energy simulations. *J. Am. Chem. Soc. 140*, 8965-8969.

(66) Xiang, Y., and Warshel, A. (2008) Quantifying free energy profiles of proton transfer reactions in solution and proteins by using a diabatic fdft mapping. *J. Phys. Chem. B 112*, 1007-1015.

(67) A., M. R., Hanwool, Y., and Arieh, W. (2016) Exploring the mechanism of DNA polymerases by analyzing the effect of mutations of active site acidic groups in polymerase β. *Proteins: Struct., Funct., Bioinf. 84*, 1644-1657.

(68) Genna, V., Vidossich, P., Ippoliti, E., Carloni, P., and De Vivo, M. (2016) A self-activated mechanism for nucleic acid polymerization catalyzed by DNA/RNA polymerases. *J. Am. Chem. Soc. 138*, 14592-14598.

(69) Ghosal, G., and Chen, J. (2013) DNA damage tolerance: A double-edged sword guarding the genome. *Transl. Cancer Res. 2*, 107.

(70) Friedberg, E. C. (2005) Suffering in silence: The tolerance of DNA damage. *Nat. Rev. Mol. Cell Biol. 6*, 943.

(71) Eoff, R. L., Egli, M., and Guengerich, P. (2010) Impact of chemical adducts on translesion synthesis in replicative and bypass DNA polymerases: From structure to function. *The Chemical Biology of DNA Damage*, 299-330.

(72) Zhang, Y., Wu, X., Rechkoblit, O., Geacintov, N. E., Taylor, J.-S., and Wang, Z. (2002) Response of human rev1 to different DNA damage: Preferential dcmp insertion opposite the lesion. *Nucleic Acids Res. 30*, 1630-1638.

(73) Swan, M. K., Johnson, R. E., Prakash, L., Prakash, S., and Aggarwal, A. K. (2009) Structure of the human Rev1-DNA-dNTP ternary complex. *J. Mol. Biol. 390*, 699-709.

(74) McCulloch, S. D., Kokoska, R. J., Masutani, C., Iwai, S., Hanaoka, F., and Kunkel, T. A. (2004) Preferential cis-syn thymine dimer bypass by DNA polymerase eta occurs with biased fidelity. *Nature 428*, 97-100.

(75) Yang, W. (2014) An overview of y-family DNA polymerases and a case study of human DNA polymerase η. *Biochemistry 53*, 2793-2803.

(76) Lange, S. S., Takata, K.-i., and Wood, R. D. (2011) DNA polymerases and cancer. *Nat. Rev. Cancer 11*, 96-110.

(77) Rogozin, I. B., Goncearenco, A., Lada, A. G., De, S., Yurchenko, V., Nudelman, G., Panchenko, A. R., Cooper, D. N., and Pavlov, Y. I. (2018) DNA polymerase η mutational signatures are found in a variety of different types of cancer. *Cell Cycle 17*, 348-355.

(78) Guo, C., Kosarek-Stancel, J. N., Tang, T.-S., and Friedberg, E. (2009) Y-family DNA polymerases in mammalian cells. *Cell. Mol. Life Sci. 66*, 2363-2381.

(79) Masutani, C., Kusumoto, R., Yamada, A., Dohmae, N., Yokoi, M., Yuasa, M., Araki, M., Iwai, S., Takio, K., and Hanaoka, F. (1999) The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase eta. *Nature 399*, 700-704.

(80) Walker, A. R., and Cisneros, G. A. (2017) Computational simulations of DNA polymerases: Detailed insights on structure/function/mechanism from native proteins to cancer variants. *Chem. Res. Toxicol. 30*, 1922-1935.

(81) Yeh, H. J., Sayer, J. M., Liu, X., Altieri, A. S., Byrd, R. A., Lakshman, M. K., Yagi, H., Schurter, E. J., Gorenstein, D. G., and Jerina, D. M. (1995) NMR solution structure of a nonanucleotide duplex with a dG mismatch opposite a 10s adduct derived from trans addition of a deoxyadenosine N6-amino group to (+)-(7r, 8s, 9s, 10r)-7, 8-dihydroxy-9, 10-epoxy-7, 8, 9, 10-tetrahydrobenzo [a] pyrene: An unusual syn glycosidic torsion angle at the modified dA. *Biochemistry 34*, 13570-13581.

(82) Li, Z., Kim, H.-Y., Tamura, P. J., Harris, C. M., Harris, T. M., and Stone, M. P. (1999) Role of a polycyclic aromatic hydrocarbon bay region ring in modulating DNA adduct structure: The non-bay region (8s,9r,10s,11r)-N6-[11-(8,9,10,11-tetrahydro-8,9,10-trihydroxybenz[a]anthracenyl)]-2‘-deoxyadenosyl adduct in codon 61 of the human N-ras protooncogene. *Biochemistry 38*, 14820-14832.

(83) Li, Z., Mao, H., Kim, H.-Y., Tamura, P. J., Harris, C. M., Harris, T. M., and Stone, M. P. (1999) Intercalation of the (−)-(1r,2s,3r,4s)-N6-[1-benz[a]anthracenyl]-2‘-deoxyadenosyl adduct in an oligodeoxynucleotide containing the human N-ras codon 61 sequence. *Biochemistry 38*, 2969-2981.

(84) Volk, D. E., Thiviyanathan, V., Rice, J. S., Luxon, B. A., Shah, J. H., Yagi, H., Sayer, J. M., Yeh, H. J., Jerina, D. M., and Gorenstein, D. G. (2003) Solution structure of a cis-opened (10 r)-N 6-deoxyadenosine adduct of (9 s, 10 r)-9, 10-epoxy-7, 8, 9, 10-tetrahydrobenzo [a] pyrene in a DNA duplex. *Biochemistry 42*, 1410-1420.

(85) Mao, B., Gu, Z., Gorin, A., Chen, J., Hingerty, B. E., Amin, S., Broyde, S., Geacintov, N. E., and Patel, D. J. (1999) Solution structure of the (+)-cis-anti-benzo[a]pyrene-dA ([BP]dA) adduct opposite dT in a DNA duplex. *Biochemistry 38*, 10831-10842.

(86) Cai, Y., Ding, S., Geacintov, N. E., and Broyde, S. (2011) Intercalative conformations of the 14r (+)- and 14s (-)-trans-anti-db[a,l]p-N6-dA adducts: Molecular modeling and MD simulations. *Chem. Res. Toxicol. 24*, 522-531.

(87) Tan, J., Geacintov, N. E., and Broyde, S. (2000) Conformational determinants of structures in stereoisomeric cis-opened anti-benzo[a]pyrene diol epoxide adducts to adenine in DNA. *Chem. Res. Toxicol. 13*, 811-822.

(88) Cai, Y., Wang, L., Ding, S., Schwaid, A., Geacintov, N. E., and Broyde, S. (2010) A bulky DNA lesion derived from a highly potent polycyclic aromatic tumorigen stabilizes nucleosome core particle structure. *Biochemistry 49*, 9943-9945.

(89) Kathuria, P., Sharma, P., Abendong, M. N., and Wetmore, S. D. (2015) Conformational preferences of DNA following damage by aristolochic acids: Structural and energetic insights into the different mutagenic potential of the ALI and ALII-N-6-dA adducts. *Biochemistry 54*, 2414-2428.

(90) Lukin, M., Zaliznyak, T., Johnson, F., and de los Santos, C. (2012) Structure and stability of DNA containing an aristolactam II-dA lesion: Implications for the ner recognition of bulky adducts. *Nucleic Acids Res. 40*, 2759-2770.

(91) Kathuria, P., Sharma, P., and Wetmore, S. D. (2015) Adenine versus guanine DNA adducts of aristolochic acids: Role of the carcinogen-purine linkage in the differential global genomic repair propensity. *Nucleic Acids Res. 43*, 7388-7397.

(92) 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.

(93) 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 C-phenolic-guanine adducts. *Chem. Res. Toxicol. 6*, 6.

(94) Millen, A. L., Manderville, R. A., and Wetmore, S. D. (2010) Conformational flexibility of c8-phenoxyl-2'-deoxyguanosine nucleotide adducts. *J. Phys. Chem. B 114*, 4373-4382.

(95) Millen, A. L., Kamenz, B. L., Leavens, F. M., Manderville, R. A., and Wetmore, S. D. (2011) Conformational flexibility of C8-phenoxylguanine adducts in deoxydinucleoside monophosphates. *J. Phys. Chem. B 115*, 12993-13002.

(96) Zhang, L., Rechkoblit, O., Wang, L., Patel, D. J., Shapiro, R., and Broyde, S. (2006) 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. 34*, 3326-3337.

(97) Wang, L., and Broyde, S. (2006) A new anti conformation for N-(deoxyguanosin-8-yl)-2-acetylaminofluorene (AAF-dG) allows Watson–Crick pairing in the sulfolobus solfataricus p2 DNA polymerase iv (Dpo4). *Nucleic Acids Res. 34*, 785-795.

(98) Donny-Clark, K., Shapiro, R., and Broyde, S. (2008) Accommodation of an N-(deoxyguanosin-8-yl)-2-acetylaminofluorene adduct in the active site of human DNA polymerase ι: Hoogsteen or Watson−Crick base pairing? *Biochemistry 48*, 7-18.

(99) Gooderham, N. J., Zhu, H., Lauber, S., Boyce, A., and Creton, S. (2002) Molecular and genetic toxicology of 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP). *Mutat. Res. Fund. Mol. Mech. Mut. 506–507*, 91-99.

(100) Wang, L., Wu, M., Yan, S. F., Patel, D. J., Geacintov, N. E., and Broyde, S. (2005) Accommodation of a 1s-(−)-benzo[c]phenanthrenyl-N6-dA adduct in the y-family Dpo4 DNA polymerase active site: Structural insights through molecular dynamics simulations. *Chem. Res. Toxicol. 18*, 441-456.

(101) Donny-Clark, K., and Broyde, S. (2009) Influence of local sequence context on damaged base conformation in human DNA polymerase ι: Molecular dynamics studies of nucleotide incorporation opposite a benzo[a]pyrene-derived adenine lesion. *Nucleic Acids Res. 37*, 7095-7109.

(102) Jia, L., Geacintov, N. E., and Broyde, S. (2008) The N-clasp of human DNA polymerase κ promotes blockage or error-free bypass of adenine- or guanine-benzo[a]pyrenyl lesions. *Nucleic Acids Res. 36*, 6571-6584.

(103) Chandani, S., and Loechler, E. L. (2007) Molecular modeling benzo[a]pyrene N 2-dG adducts in the two overlapping active sites of the Y-family DNA polymerase Dpo4. *J. Mol. Graphics Modell. 25*, 658-670.

(104) Perlow-Poehnelt, R. A., Likhterov, I., Scicchitano, D. A., Geacintov, N. E., and Broyde, S. (2004) The spacious active site of a y-family DNA polymerase facilitates promiscuous nucleotide incorporation opposite a bulky carcinogen-DNA adduct: Elucidating the structure-function relationship through experimental and computational approaches. *J. Biol. Chem. 279*, 36951-36961.

(105) Xu, P., Oum, L., Geacintov, N. E., and Broyde, S. (2008) Nucleotide selectivity opposite a benzo[a]pyrene-derived N2-dG adduct in a y-family DNA polymerase: A 5′-slippage mechanism. *Biochemistry 47*, 2701-2709.

(106) Xu, P., Oum, L., Lee, Y.-C., Geacintov, N. E., and Broyde, S. (2009) Visualizing sequence-governed nucleotide selectivities and mutagenic consequences through a replicative cycle: Processing of a bulky carcinogen N2-dG lesion in a y-family DNA polymerase. *Biochemistry 48*, 4677-4690.

(107) Perlow-Poehnelt, R. A., Likhterov, I., Wang, L., Scicchitano, D. A., Geacintov, N. E., and Broyde, S. (2007) Increased flexibility enhances misincorporation: Temperature effects on nucleotide incorporation opposite a bulky carcinogen-DNA adduct by a y-family DNA polymerase. *J. Biol. Chem. 282*, 1397-1408.

(108) Senn, H. M., and Thiel, W. (2009) QM/MM methods for biomolecular systems. *Angew. Chem., Int. Ed. 48*, 1198-1229.

(109) Marino, T., Russo, N., and Prejano, M. (2018) QM cluster or QM/MM in computational enzymology: The test case of ligw-decarboxylase. *Front. Chem. 6*, 249.

(110) Hayashi, S., Uchida, Y., Hasegawa, T., Higashi, M., Kosugi, T., and Kamiya, M. (2017) QM/MM geometry optimization on extensive free-energy surfaces for examination of enzymatic reactions and design of novel functional properties of proteins. *Annu. Rev. Phys. Chem. 68*, 135-154.

Chapter 2: The Complex Conformational Heterogeneity of the Highly Flexibility O6-Benzyl-Guanine DNA Adduct[a,b,c]

2.1. Introduction

Cellular DNA can be damaged by many different exogenous and endogenous agents, which can cause mutations that lead to cancer, premature aging, and cardiovascular disease.[1-4] DNA modifications can occur to one of the four canonical nucleobases and include alkylation,[5, 6] oxidation,[7, 8] and deamination.[9, 10] In terms of alkylation, the addition of an alkyl or aryl group to the O6-position of G (Figure 2.1) leads to some of the most mutagenic and carcinogenic lesions.[11-13] Common forms of DNA methylation occur upon human exposure to tobacco specific nitrosamines, [6, 14-16] while DNA benzylation occurs upon exposure to chemicals (such as N-nitrobenzylmethylamine) used to study esophagus carcinogenesis.[17, 18] Due to the cytotoxic effects, therapeutic agents have been developed to exploit O6-G alkylation damage (for example, O6-benzylguanine is administered to patients to inactivate O6-alkylguanine-DNA alkyltransferase).[19, 20] However, these therapeutic agents are unfortunately carcinogenic to patients with long-term survival.[19] In order to understand the toxicity of these adducts and increase the effectiveness of related chemotherapeutic agents, an in-depth investigation of O6-G alkylation damage is required.

---

[a] Chemical Research in Toxicology reference style used throughout this chapter

[b] Reprinted with permission from Wilson, K.A. and Wetmore, S.D. (2014). Complex conformational heterogeneity of the highly flexible O6-benzyl-guanine DNA adduct. *Chem. Res. Toxicol.,* 27, 1310-1325.

[c] K.A.W. performed calculations, data analysis and wrote the first draft of the manuscript.

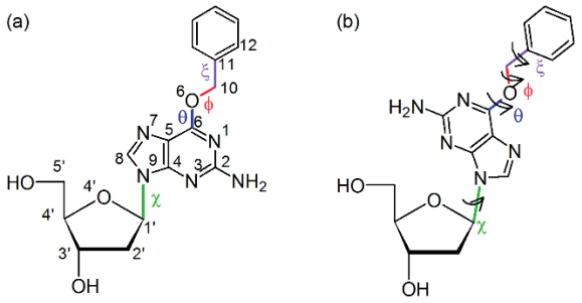

Figure 2.1. The BzG nucleoside in the (a) *anti* and (b) *syn* orientations, and definition of important dihedral angles including χ (∠(O4′C1′N9C4)) in green, θ (∠(C5C6O6C10)) in blue, φ (∠(C6O6C10C11)) in red and ξ (∠(O6C10C11C12)) in purple.

O6-G alkylation products include the smallest O6-methyl-G (MeG) and the larger O6-benzyl-G (BzG, Figure 2.1) lesions. In addition to being an important alkylation by-product, BzG is a member of the class of lesions known as bulky adducts (addition products), which are formed when chemicals composed of one or more aromatic rings are absorbed by our bodies and metabolized into intermediates that attack DNA.[21] The most widely studied, and therefore best understood, adducts are the N-linked G lesions in which the bulky moiety can modify the Watson-Crick (WC) hydrogen-bonding face (for example, when bonded at N2 and extended towards the pairing base) or can have no effect on the WC hydrogen-bonding face (for example, when bonded at C8). An abundance of literature has verified that DNA containing N-linked adducts can adopt three conformations,[22, 23] which either direct the bulky moiety into the major groove (B-conformation) or the minor groove (W-conformation), or intercalate the bulky moiety into the helix (S-conformation). The conformation adopted depends on the orientation of the bulky ring with respect to the nucleobase (θ) and the nucleobase (*anti/syn*) conformation with respect to the glycosidic bond (χ, Figure 2.1). Most importantly, the mutagenicity of N-linked G adducts has been directly correlated to the adducted helix conformation.[24-26] For example, frameshift

mutations commonly occur when intercalated structures are adopted,[26] while transition or transversion mutations occur when a G adduct is mispaired opposite G, A, or T due to changes in the Watson-Crick or Hoogsteen faces of G.[24, 25] Therefore, understanding the preferred conformation(s) of bulky adducts in DNA is the first step to revealing their mutagenic and cytotoxic consequences.

Information about the structural features of select O6-G alkylation products is available in the literature. Indeed, the structure of MeG has been well studied both computationally and experimentally.[19, 27-37] At the nucleoside level, computational work reveals that MeG prefers a conformation that directs the lone pairs of the linkage oxygen towards N7 of G and places the methyl group on the Watson-Crick binding face.[28] However, another computational study determined that the conformation with the lone pairs directed towards N1 becomes preferred when MeG is base paired, which permits Watson-Crick hydrogen bonding.[29] Experimental extension studies[27, 30, 34, 35] show that MeG can base pair with C in a wobble base pair, and binds to T in a pseudo-Watson-Crick geometry. When the stability of these free base pairs in solution are compared using computational chemistry, MeG:T is preferred over MeG:C by less than 5 kJ mol^{-1}.[29] However, the base pair stability differs in DNA, where a helix containing MeG:T has been measured to be less stable than the corresponding helix containing MeG:C.[34-37] Nevertheless, since the MeG:T binding arrangement more closely resembles the natural G:C base pair, T is frequently misinserted opposite MeG upon replication,[31-33] which leads to G → A transition mutations.

Less direct information about the structure and base-pairing preferences of the larger BzG alkylation adduct is available. Nevertheless, the replication of BzG has been studied with a variety of DNA polymerases.[32, 38-40] When replicated by Dpo4, an archaeal Y-family polymerase that copies damaged DNA, C is most commonly incorporated opposite BzG (70%), with T (15%) and A (15%) also (albeit less frequently) incorporated.[38] On the other hand, when BzG is replicated by pol η, which is the eukaryotic enzyme involved in the translesion synthesis of thymine dimers, C remains the most commonly inserted nucleobase (55%), while the frequency of T incorporation increases (31%) and the frequency of A incorporation decreases (11%).[32] Notably, stable base pairing with A or T may lead to transition or transversion mutations upon further replication. Furthermore, a deletion mutation occurs when BzG is replicated by pol η (although with a low (3%) frequency).[32] Interestingly, Dpo4 replicates BzG with a lower rate of mutation compared to other alkyl-G lesions,[29, 38] but the exact reason for this is unknown. To complement information about base incorporation opposite BzG, crystal structures of Dpo4 bound to DNA containing BzG:C or BzG:T have been published,[29, 38] which indicate these pairs are similar to the MeG variants. Specifically, BzG:C adopts a wobble base pair,[29, 38] while BzG:T adopts a pseudo-Watson-Crick base pair. [29, 38]

Despite the detailed research on the replication of BzG, literature on the N-linked G adducts suggests that the conformation of BzG adopted in cells must be unequivocally determined in order to elucidate the mutagenic consequences of this lesion. Notably, the BzG adduct is structurally distinct from other bulky adducts previously examined. First, the site of attachment of the bulky moiety in G adducts previously studied is typically either the C8

or N2 position, or in rare cases N7. Indeed, to the best of our knowledge, only one O6-G lesion has been previously discussed in the literature (namely O6-POB-G).[41] Second, the bulky moieties of most adducts are generally much larger, being composed of several fused rings (such as C8-AF-G,[42] C8-AAF-G,[42] C8-AP-G,[43] C8-B[α]P-6-G,[44] N7-AFB-G,[45] C8-PhIP-G,[46] C8-IQ-G,[47] and C8-3,4,-EQ-G[48]), although some smaller adducts have also been more recently investigated (such as C8-Ph-G,[49] and substituted analogues,[50] as well as the C8-PhO-G[51] variants). Third, the bulky moiety in many N-linked (such as C8-AF-G,[42] C8-AAF-G,[42] C8-AP-G,[43] and C8-IQ-G[52]) and C-linked (such as C8-Ph-G, C8-B[α]P-6-G,[44] and C8-3,4,-EQ-G[48]) adducts is generally rigid, with few flexible dihedral angles in the bulky group. Although some bulky moieties exhibit some intrinsic flexibility through, for example, the puckering in the bulky group (such as N2-B[α]P-G,[44] N2-[BPh]-G,[53] and N2-[BP]-G[53]) or ring substituent (such as C8-ABP-G,[54] C8-PhIP-G,[55] C8-(4-CH_2OH-Ph)-G,[56] C8-(4-COOH-Ph)-G,[56] and C8-OTA-G[57]), the ability to rotate about multiple dihedral angles in the linkage between G and the aromatic ring is rare. Indeed, the only exception in the literature to date is POB-G,[41] which like BzG has carbon spacers in the (–(CH_2)$_3$CO–) linkage. Finally, most other adducts involve a C–C or C–N bond linking the bulky moiety to the DNA base, with relatively few examples of O-linked adducts found in the literature (C8-PhO-G,[51] C8-OTA-G,[58] and O6-POB-G[41]). These differences may affect the structures adopted and therefore the mutagenicity of BzG compared to other adducts. For example, helices containing adducts with a large, rigid aromatic bulky moiety have been determined to more likely adopt the S-conformation in order to maximize nucleobase–nucleobase stacking interactions, while adducts with less rigid bulky moieties prefer major or minor groove orientations that permit rotation of the bulky group.[23, 51, 54, 59, 60] Furthermore, the site of attachment of the bulky moiety dictates changes in the hydrogen-bonding patterns

between the adduct and the canonical bases. Therefore, the conformation of BzG in DNA must be studied in detail.

The present work fills a void in the literature on BzG by unveiling the intrinsic conformation of the bulky moiety, the preferred orientation of the bulky group relative to the base, and the favoured conformation of the base about the glycosidic bond. Due to the atomic level details provided by computer simulations, computational chemistry is used to determine the preferred BzG conformation in models of varying size. First, the preferred orientation of the bulky group is determined relative to the G base, as well as the *anti*/*syn* preference, using a BzG nucleoside model and quantum mechanics. The effect of the DNA backbone on the intrinsic structure of the lesion is subsequently determined using a nucleotide model built from select low energy nucleoside conformations. Finally, the small model studies are used to direct MD simulations on adducted DNA helices, which will determine the effects of the pairing nucleotide and flanking base pairs on the BzG and adducted helical conformation, as well as the effect of BzG on the overall helix stability. For the first time, our calculations determine the most stable conformation and base pairing of the adduct, and thereby explain the previously reported mutagenic nature of BzG. Furthermore, our data expands the current literature on DNA adducts by considering a lesion containing a bulky moiety with a varied site of attachment, size, linkage type, and intrinsic conformational flexibility.

2.2. Computational Details

2.2.1. Nucleoside Model: In all nucleoside calculations, the C3′ hydroxyl of BzG was directed such that ∠(HC3′O3′H) = –60°, which has been shown to be the preferred conformation in related G adducts.[61] Furthermore, the C5′ hydroxyl was originally constrained (∠(C4′C5′O5′H) = 180°) to prevent unnatural interactions between the nucleobase and the C5′ hydroxyl (denoted as β-constraint) since this approach yields structures relevant to the DNA environment in previous studies of natural G, as well as other G adducts.[57, 62, 63] The preferred conformation of the BzG nucleoside was first considered by mapping the potential energy surface (PES) with B3LYP/6-31G(d) as a function of χ (∠(O4′C1′N9C4)), which dictates the (*anti*/*syn*) conformation of the base relative to the sugar moiety, and θ (∠(C5C6O6C10)), which dictates the orientation of the bulky group with respect to the base (Figure 2.1). Specifically, χ and θ were rotated and fixed in 10° increments from 0 to 360°, and the remainder of the molecule was relaxed at each point (except the additional β-constraint). Further information about the nucleoside conformation was then obtained by considering the dihedral angles φ (∠(C6O6C10C11)) and ξ (∠(O6C10C11C12)), which control the intrinsic conformation of the bulky moiety (Figure 2.1). Specifically, the PES was searched as a function of φ and θ or ξ and θ (with χ fixed to the minimum energy *anti* or *syn* value) using the same method as described for the χ and θ PES. Representative structures were optimized without β-constraints. All representative structures (with and without β-constraint) for the nucleoside model were performed using B3LYP-D3/6-311+G(2df,p)//B3LYP/6-31G(d), and the associated relative energies include scaled (0.9806) zero-point vibrational corrections.

2.2.2. Nucleotide Model: Previous computational work has determined PCM-B3LYP/6-31G(d) optimizations in water using a model that includes a Na^+ counterion yield the most biologically-relevant geometries for natural 5′-monophosphate 2′-deoxyguanosine.[62] Therefore, the same model was implemented in the present work, which was built by adding PO_3^- coupled with a Na^+ counterion to select β-constraint nucleoside minima. All nucleotide calculations implemented PCM-B3LYP-D3/6-311+G(2df,p)//PCM-B3LYP/6-31G(d), and relative energies include scaled (0.9806) zero-point vibrational corrections.

All nucleoside and nucleotide calculations were performed using Gaussian 09, Revisions C.01 or D.01.[64, 65]

2.2.3. DNA Model: MD simulations were performed on the *Nar*I (5′-CTCG[1]G[2]CG[3]CCATC) sequence with the BzG adduct in the G^3 position, adopting a variety of conformations and paired opposite different nucleotides. Specifically, *anti*–BzG was paired opposite *anti*–C, *anti*–T, *syn*–G and *syn*–A, while *syn*–BzG was paired opposite *anti*–C, *anti*–T and *anti*–G. The corresponding natural sequence (containing the *anti*–G:*anti*–C base pair at G^3) was also considered. The G^3 position was chosen since this is a known hotspot for mutagenesis.[66, 67] The starting χ, θ, φ and ξ dihedral angles (Figure 2.1) were chosen based on the quantum mechanical calculations, while ensuring BzG hydrogen bonds with the opposing nucleotide and avoids steric clashes with all neighbouring bases. Simulations were then conducted following the protocol used to investigate other adducted helices.[49, 51, 57, 68] Specifically, parameters for BzG were assigned according to the GAFF[69] and AMBER ff99SB[70] force fields using ANTECHAMBER 1.4 (Figure A.1, Appendix

A).[71] Partial charges for the BzG adduct were developed using RESP charge fitting from a HF/6-31G(d) calculation by the R.E.D.v.III.4 program (Table A.1, Appendix A).[72] Simulations on DNA helices neutralized with 22 Na^+ ions were carried out in an 8 Å TIP3P water box with the SANDER module of AMBER 12[73] using a 2 fs time step. Initial minimization of the solvent and ions was carried out with a 500 kcal/mol constraint on DNA. Subsequently, 1000 steps of unrestrained steepest descent and 1500 steps of unrestrained conjugate gradient minimization were performed on the entire system. The system was subsequently heated at constant volume from 0 to 300 K over 20 ps with a 10 kcal/mol restraint on DNA. The entire system was then simulated without restraints for 20 ns. All simulations have a stable RMSD (< 4.6 Å, Table 2.1). The MD simulations were analysed using the cpptraj module of AmberTools 13.[74] Simulations displaying a bi or trimodal distribution with respect to χ or φ were manually clustered. Furthermore, total and binding free energies were calculated for each simulation (or cluster) through the MM-PBSA[75] approximation using the Poisson-Boltzmann approach with snapshots taken every 50 ps.

2.3. Results

2.3.2. Nucleoside Model: To investigate the preferred conformation of the BzG nucleoside, the PES was first mapped as a function of the χ and θ dihedral angles, which control the relative orientations about the base–sugar and base–bulky moiety bonds, respectively (Figure 2.1). The PES (Figure 2.2a) reveals two *anti* minima (χ ≈ 240°, θ ≈ 0 and 180°) and two *syn* minima (χ ≈ 80°, θ ≈ 0 and 180°). In the relaxed structures (Figure 2.2b), the bulky moiety adopts a planar orientation in all minima, with θ ≈ 180° (O6 lone pairs directed

towards N7) being ~7 kJ mol^{-1} lower in energy than the corresponding θ ≈ 0° orientation. More importantly, the *syn* conformers are higher in energy than the *anti* conformers by approximately 12 kJ mol^{-1}. The rotational barriers about χ and θ are ~15 – 30 and ~25 kJ mol^{-1}, respectively. Interestingly, φ and ξ are dependent on θ, but independent of χ (Figure 2.3). Therefore, the PES was further mapped as a function of φ and θ or ξ and θ, while sequentially constraining the χ dihedral angle to the values in the *anti* and *syn* minima (χ = 80 or 240°). Since these PESs are very similar for the *anti* and *syn* conformers (Figures A.2 and A.3, Appendix A), only the lower energy *anti* surfaces will be discussed in more detail below.

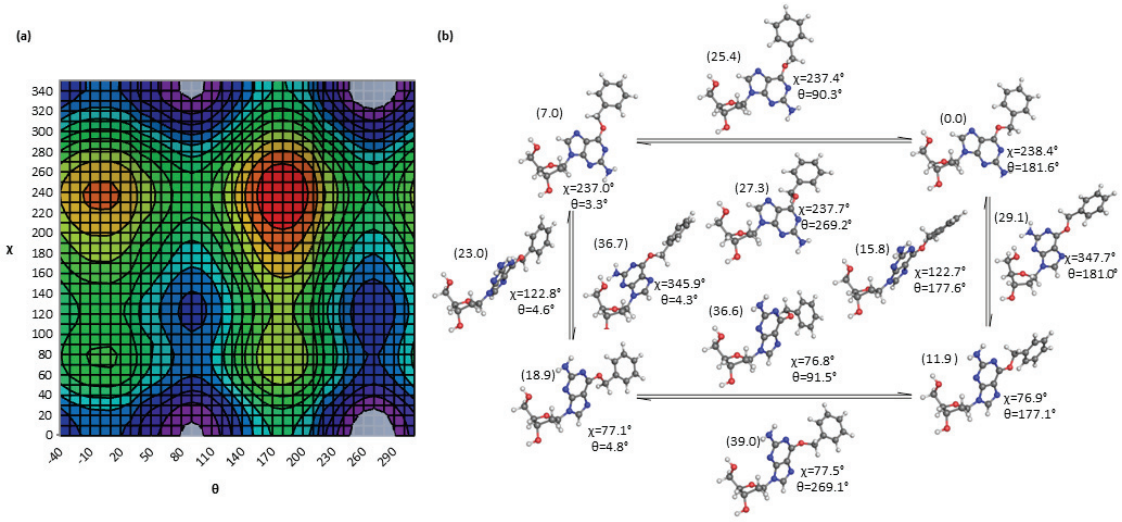

Figure 2.2. (a) PES (kJ mol^{-1}) as a function of χ and θ (degrees) in BzG (where red is the lowest energy region and each color change represents 5 kJ mol^{-1}), and (b) corresponding β-constrained optimized minima and transition states (χ and θ dihedral angles provided with energy (kJ mol^{-1}) in parentheses).

The PES with respect to φ and θ (Figure 2.4a) has 6 minima that adopt two θ values (0 and 180°) and each of three φ values (90, 180 and 270°). In all minima, the benzyl group is parallel (φ ≈ 180°) or perpendicular (φ ≈ 90 and 270°) to the guanine base, with an energy difference of less than 2 kJ mol^{-1} between the φ minima (Figure 2.4b). On the other hand,

the θ ≈ 180° minima are ~7 kJ mol^{-1} lower in energy than the corresponding θ ≈ 0° conformations. The barrier to rotation from either φ ≈ 90° to φ ≈ 180° or φ ≈ 180° to φ ≈ 270° is ~5 kJ mol^{-1}. However, due to steric repulsion between the benzyl ring and N1 of G when φ ≈ 0°, the rotational barrier between φ ≈ 90° and φ ≈ 270° is much larger, being 31.0 kJ mol^{-1} for the relaxed θ ≈ 180° transition state and ~65 kJ mol^{-1} for θ ≈ 0° (estimated from the PES since the corresponding fully optimized TS could not be characterized).

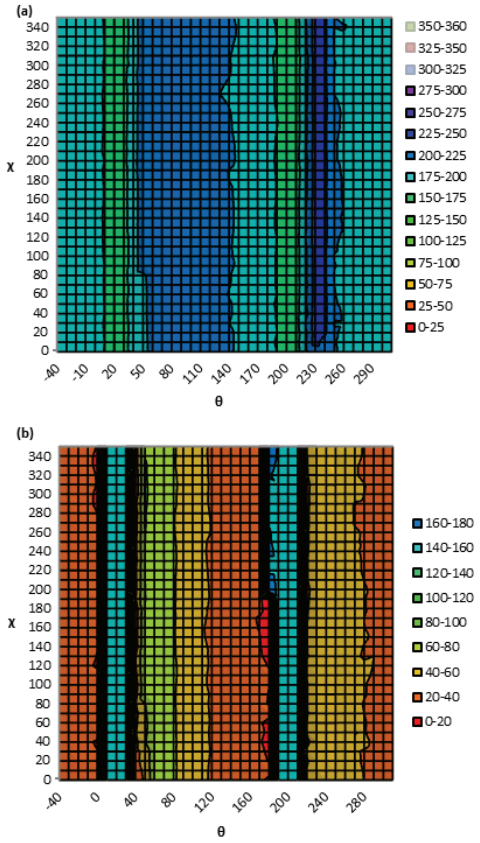

Figure 2.3. (a) φ and (b) ξ dihedral angles (degrees) for each point on the BzG χ versus θ (degrees) PES (Figure 2.2a).

The PES as a function of ξ and θ is very flat with respect to ξ (Figure 2.5a), with the average rotational barrier less than 3 kJ mol^{-1} for a given θ. Nevertheless, there are two minima at both θ ≈ 0 and 180° with ξ ≈ 20 and 160°. Therefore, the minima are not perfectly planar

or perpendicular, but rather are twisted by ~20° with respect to G (Figure 2.5b). Notably, rotation between these minima is essentially barrierless.

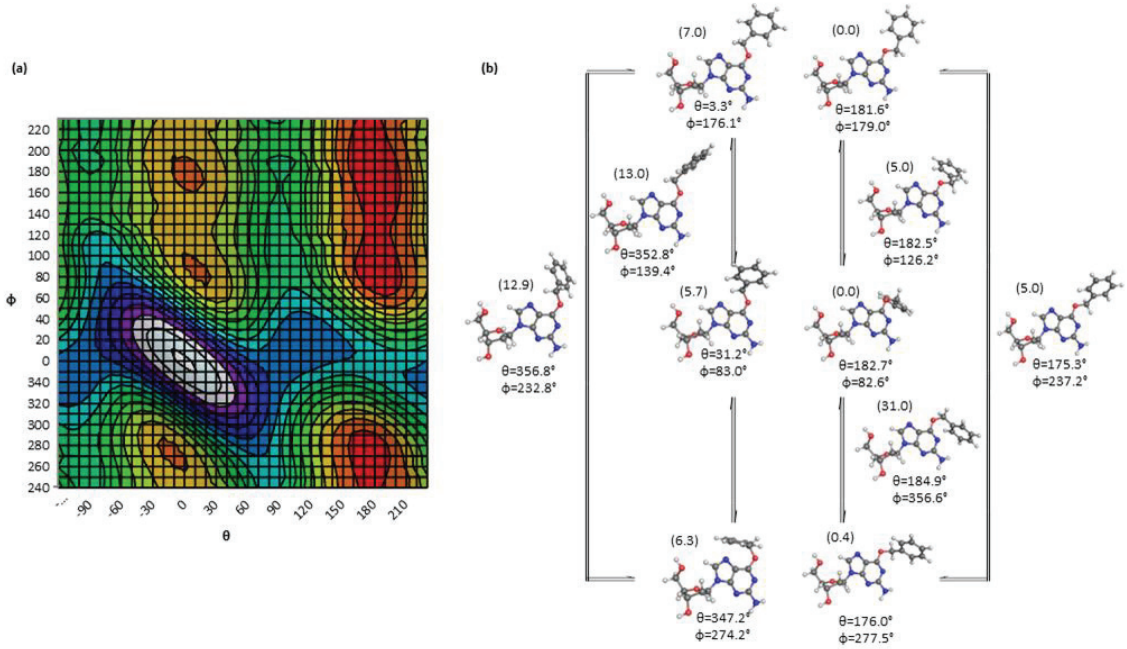

Figure 2.4. (a) PES (kJ mol^{-1}) as a function of θ and φ (degrees) in BzG (where red is the lowest energy region and each color change represents 5 kJ mol^{-1}), and (b) corresponding β-constraint optimized minima and transition states (θ and φ angles provided with energy (kJ mol^{-1}) in parentheses).

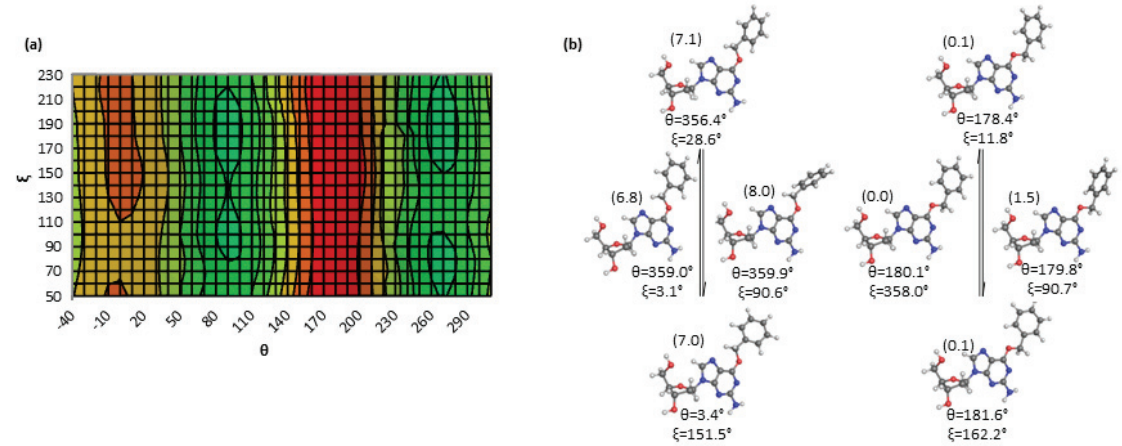

Figure 2.5. (a) PES (kJ mol^{-1}) as a function of θ and ξ (degrees) in BzG (where red is the lowest energy region and each color change represents 5 kJ mol^{-1}), and (b) corresponding β-constraint optimized minima and transition states (θ and ξ angles provided with energy (kJ mol^{-1}) in parentheses).

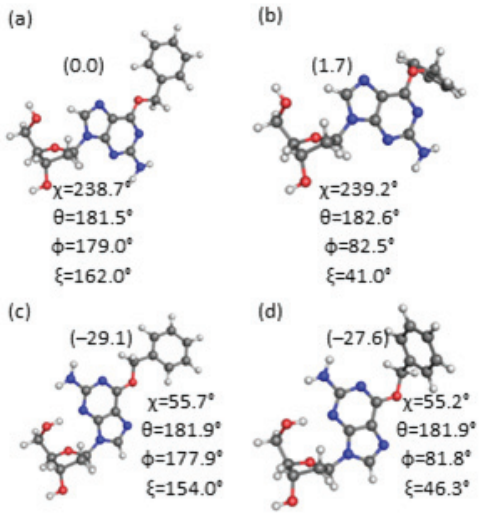

Figure 2.6. Optimized BzG β-constrained nucleoside minima including important dihedral angles (degrees) and relative energies given in parentheses (kJ mol^{-1}).

Based on the calculated PESs, select structures were chosen to more broadly compare the nucleoside minima (Figure 2.6). Specifically, structures with the lowest energy *anti* (χ = 240°) or *syn* (χ = 80°) arrangements and the lowest energy θ value (180°) were considered. Due to the similarity in energy, but differences in structure, both $\varphi \approx 90$ or 180° were considered. Finally, only the lowest energy ξ = 160° arrangement was chosen due to the small rotational barrier about ξ. When the resulting optimized (all but β-constraint removed) structures are compared (Figure 2.6), the conclusions previously highlighted are emphasized. Specifically, the two *anti* minima are ~11 kJ mol^{-1} more stable than the *syn* minima. Additionally, there is less than a 0.5 kJ mol^{-1} energy difference between structures with $\varphi \approx 90$ and 180°. Notably, two ξ values are adopted upon optimization (~40° (when $\varphi \approx 80°$) and ~180° (when $\varphi \approx 180°$)) despite the lack of defining intramolecular hydrogen bonds. Furthermore, when the β-constraint is removed and the minimum energy points re-optimized (Figure A.4, Appendix A), there is less than a 0.5° change in any of the key dihedral angles in the *anti* conformer. In contrast, an intramolecular C5′–H···N3 interaction

forms in the *syn* structures, which is coupled with an ~20° change in χ and an up to 6° change in ξ. This contact helps stabilizes the *syn* structure over the *anti* by nearly 30 kJ mol^{-1}, which underscores the importance of implementing the β-constraint when using a nucleoside model to gain structural information relevant to DNA.

2.3.2. Nucleotide Model: Although the *anti* conformation is preferred according to the β-constrained nucleoside model, the *anti*/*syn* energy difference is small (< 11 kJ mol^{-1}) and other factors (i.e., the 5′-phosphate) may alter the preferred conformation (i.e., due to steric constraints). Therefore, the effect of the 5′ phosphate was considered by adding a Na^+ neutralized phosphate to the β-constrained nucleoside minima as done in previous work on other DNA adducts.[57, 62] The optimized nucleotide structures (Figure 2.7) exhibit only small deviations in the χ, θ and φ dihedral angles compared to the nucleoside model, while larger changes occur in ξ, which is consistent with the previously discussed barrierless rotation about this dihedral angle. Nevertheless, the nucleotide model predicts an even smaller *anti/syn* energy difference (2 kJ mol^{-1}) than the nucleoside model (11 kJ mol^{-1}). Furthermore, the *syn* nucleotides can be lower in energy compared to the *anti* conformations depending on the φ adopted. However, the *syn* nucleotides are stabilized by an intramolecular hydrogen bond between the phosphate moiety and N2 of G that in order to maximize favorable stacking interactions with the flanking bases may not form in DNA, yet have been shown to be important for C8-AAF-G[76], C8-AF-G[76] and C8-IQ-G.[77] Most importantly, the *anti/syn* energy difference is small enough that the final BzG conformation will be determined by intrahelical interactions in the DNA environment.

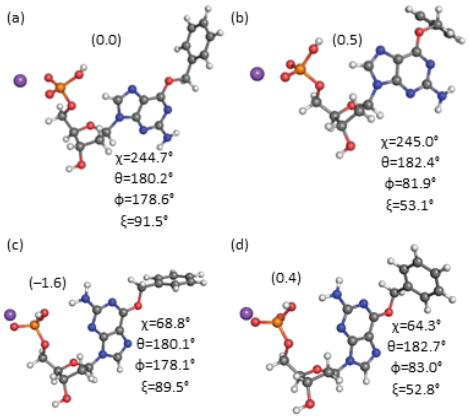

Figure 2.7. Optimized BzG nucleotide minima including important dihedral angles (degrees) and relative energies given in parentheses (kJ mol^{-1}).

2.3.4. DNA Model: To gain more information about the conformation of the adduct in DNA, BzG was placed in the *anti* and *syn* orientations at the G^3 position of the *Nar*I sequence opposite a variety of canonical nucleotides and MD simulations were conducted on each strand. DNA maintains a double helical structure without any kinks or bulges upon incorporation of the lesion (see Figures 2.8a, 2.9(top) and A.5, Appendix A). Indeed, absolute deviations are observed in the helical parameters for all base pairs (except the terminal base pairs since these are known to unwind throughout the course of MD simulations[78]) of up to 1.163 Å or 15.1° with respect to the natural DNA strand (Table A.2, Appendix A). Notably, standard deviations in helical parameters up to 1.138 Å or 14.2° occur in the simulation of the natural *Nar*I helix, which suggests the observed changes in the adducted helices are within the inherent dynamics of DNA. Furthermore, any changes in discrete interbase interactions are localized to the BzG base pair since the flanking hydrogen-bonded pairs retain the strength observed in the corresponding natural helix (Table 2.2). Additional details of the calculated structure will be discussed below as a function of the opposing base.

Table 2.1. Parameters from 20 ns MD simulations of the BzG adduct incorporated into the G^3 position of the *Nar*I DNA sequence.

Simulation	Cluster[a]	Occupancy	RMSD (Å)	BzG-χ[a] (°)	C2′ *endo*[b] (%)	dR Pucker[b]	BzG-θ[a] (°)	BzG-φ[a] (°)	BzG-ξ[a] (°)	C1′-C1′[c] (Å)
***anti*–G:*anti*–C**	all	1.00	3.392±0.742	251.3±18.7	45.1	C2′-*endo* (45.1%)	NA	NA	NA	10.7±0.2
***anti*–BzG:*anti*–C**	all	1.00	3.473±0.743	258.5±16.9	47.7	C2′-*endo* (47.7%)	73.4±75.4	180.0±14.3	195.3±94.0	11.3±0.4
***anti*–BzG:*anti*–T**	all	1.00	3.965±0.805	242.4±19.3	38.7	C2′-*endo* (38.7%)	152.0±67.8	181.4±15.6	172.9±94.4	10.6±0.2
***syn*–BzG:*anti*–C**	χ: 5-200°	0.73	4.329±0.852	53.9±14.4	5.8	C1′-*exo* (61.3%)	190.7±89.0	178.9±12.8	236.9±84.3	10.8±0.6
	χ: 200-365°	0.27	2.282±0.460	327.7±14.1	0.7	O4′-*endo* (70.9%)	158.3±67.5	182.2±10.9	165.8±103.6	10.9±0.2
***syn*–BzG:*anti*–T**	φ: 0-110°	0.03	2.556±0.499	42.4±10.8	0.0	O4′-*endo* (82.9%)	144.1±33.8	81.5±11.0	64.6±42.3	8.5±0.3
	φ: 110-220°	0.28	3.249±0.815	45.1±12.4	0.8	O4′-*endo* (53.5%)	336.7±73.7	174.3±16.3	199.6±101.5	8.7±0.8
	φ: 220-360°	0.70	2.491±0.595	42.6±14.3	0.3	O4′-*endo* (65.5%)	184.8±14.4	278.8±10.0	115.8±27.9	8.7±0.4
***syn*–BzG:*anti*–C**	extrahelical	1.00	4.002±1.629	56.3±12.2	10.6	C1′-*exo* (81.7%)	322.8±63.3	178.9±11.2	205.3±60.0	12.8±1.8
***syn*–BzG:*anti*–T**	extrahelical	1.00	5.298±1.192	60.9±12.4	13.1	C1′-*exo* (84.8%)	297.7±77.9	179.4±11.0	211.2±78.2	14.7±1.6
***anti*–BzG:*syn*–A**	all	1.00	3.803±0.690	252.4±17.8	45.3	C2′-*endo* (45.3%)	36.0±45.4	180.2±13.6	171.0±89.4	11.4±0.3
***anti*–BzG:*syn*–G**	φ: 0-120°	0.19	2.641±0.426	239.1±16.0	2.5	C3′-*endo* (48.5 %)	155.0±71.3	75.0±12.5	110.8±81.7	9.6±0.5
	φ: 120-240°	0.74	4.590±0.676	236.0±17.4	7.9	C3′-*endo* (48.2 %)	187.7±106.7	181.0±13.1	180.7±94.9	10.1±0.7
	φ: 240-360°	0.08	2.635±0.443	234.7±13.6	5.3	C3′-*endo* (58.3 %)	142.9±23.6	290.5±14.0	185.2±87.4	10.2±0.8
***syn*–BzG:*anti*–G**	φ: 0-125°	0.23	2.199±0.407	49.6±12.3	1.3	O4′-*endo* (49.8%)	172.0±19.7	79.1±10.9	65.6±20.7	10.6±0.6
	φ: 125-230°	0.23	4.204±0.861	46.9±13.4	0.0	O4′-*endo* (76.2%)	177.0±113.4	177.8±14.8	189.5±92.6	10.3±0.4
	φ: 230-360°	0.54	3.134±0.702	48.2±14.2	1.4	O4′-*endo* (51.7%)	191.6±20.2	276.0±10.0	283.7±21.0	10.8±0.4

[a]See Figure 2.1 for definitions of important dihedral angles. [b]Percentage of the simulation that BzG adopts the C2′–*endo* pucker and the most common sugar pucker adopted by the deoxyribose (dR) in BzG. [c]Distance between C1′ of BzG and C1′ of the pairing nucleotide.

2.3.4.1. Structure of DNA Containing BzG Paired Opposite a Pyrimidine: When BzG adopts the *anti* orientation opposite (*anti*) C or T, the benzyl moiety is located in the solvent exposed major groove ($\varphi \approx 180°$), χ deviates by less than 20° throughout the simulation and C2′-*endo* is the most prominent sugar pucker (Table 2.1). As anticipated from the nucleoside PESs (Figures 4 and 5), there is great variation in the ξ dihedral angle throughout the simulations due to the associated low rotational barriers. Despite the fact that the (quantum mechanically) calculated rotational barrier about θ is larger than about φ, there is little variation in φ, but significant deviation in θ throughout the simulation, likely due to steric constraints imposed on φ by the DNA helix. Although the C1′–C1′ distance indicates a slightly wider BzG:C pair (11.3 ± 0.4 Å) than the canonical G:C variant (10.7 ± 0.2 Å), the distance in BzG:T is close to that adopted by nucleobase pairs in natural helices (10.6 ± 0.2 Å, Table 2.1).

When *anti*–BzG is paired opposite *anti*–C, a wobble base pair is formed, which contains two stable hydrogen bonds (95% occupancy), and two transient interactions (< 20% occupancy, Figure 2.8b). The corresponding interaction energy is –42.1 ± 8.2 kJ mol^{-1}, which is ~76 kJ mol^{-1} less stable than the natural G:C Watson-Crick base pair (Table 2.2). However, the π–interactions between the *anti*–BzG:*anti*–C base pair and the flanking base pairs are stronger than the π–interactions involving the canonical G:C pair (by ~30 kJ mol^{-1}, Tables 2.3 and A.3, Appendix A).

Table 2.2. AMBER interaction energies (kJ mol^{-1}) for hydrogen-bonding interactions at the lesion site or in the flanking base pairs during 20 ns MD simulations with BzG at the G^3 position of the *Nar*I sequence.

Simulation	Cluster[a]	Occupancy	5′ Hbond G:C[b]	3′ Hbond G:C[c]	Hbond Bzg:nt[d]
***anti*–G:*anti*–C**	all	1.00	–117.6±7.7	–118.8±7.0	–118.4±7.2
***anti*–BzG:*anti*–C**	all	1.00	–118.7±7.1	–118.7±7.1	–42.1±8.2
***anti*–BzG:*anti*–T**	all	1.00	–119.3±6.7	–119.3±6.7	–20.4±4.8
***syn*–BzG:*anti*–C**	χ: 5-200°	0.73	–119.3±6.9	–119.3±6.9	–46.5±13.3
	χ: 200-365°	0.27	–118.0±7.3	–120.3±6.3	–45.5±9.7
***syn*–BzG:*anti*–T**	φ: 0-110°	0.03	–120.0±6.6	–116.8±7.4	–13.5±4.8
	φ: 110-220°	0.28	–118.8±7.1	–118.9±6.7	–16.3±5.0
	φ: 220-360°	0.70	–117.2±7.9	–117.9±7.2	–11.5±4.1
***syn*–BzG:*anti*–C**	extrahelical	1.00	–115.9±8.2	–118.6±7.0	–3.4±5.8
***syn*–BzG:*anti*–T**	extrahelical	1.00	–117.1±7.6	–117.8±7.6	–1.3±2.9
***anti*–BzG:*syn*–A**	all	1.00	–118.5±7.2	–118.5±7.2	–41.7±6.3
***anti*–BzG:*syn*–G**	φ: 0-120°	0.19	–116.0±7.8	–116.0±7.8	–10.6±7.6
	φ: 120-240°	0.74	–117.8±7.5	–117.8±7.5	–7.7±8.4
	φ: 240-360°	0.08	–116.4±8.1	–116.4±8.1	–12.3±6.9
***syn*–BzG:*anti*–G**	φ: 0-125°	0.23	–117.0±8.0	–118.5±7.0	–48.2±12.3
	φ: 125-230°	0.23	–116.1±8.6	–119.0±6.7	–36.2±9.5
	φ: 230-360°	0.54	–117.6±7.7	–119.1±6.9	–52.0±9.4

[a]See Figure 2.1 for definitions of χ and φ. [b]Interaction between residues 8 and 17. [c]Interaction between residues 6 and 19. [d]Interaction between residues 7 (BzG) and 18.

Table 2.3. AMBER interaction energies (kJ mol^{-1}) between BzG in the G^3 position of the *Nar*I sequence and the surrounding nucleobases from 20 ns MD simulations.

Simulation	Cluster[a]	Occupancy	BzG-intrastrand		BzG-interstrand		nt-intrastrand		nt-interstrand	
			5′[b]	3′[c]	5′[d]	3′[e]	5′[f]	3′[g]	5′[h]	3′[i]
***anti*–G:*anti*–C**	all	1.00	–14.8±12.0	–54.6±7.9	–21.6±6.5	23.6±6.3	–25.9±8.6	–53.3±8.0	12.6±3.0	31.5±5.6
***anti*–BzG:*anti*–C**	all	1.00	–22.4±6.8	–29.3±11.4	–20.4±7.0	–16.1±8.2	–31.2±5.7	–50.7±9.8	8.6±3.6	28.6±8.1
***anti*–BzG:*anti*–T**	all	1.00	–23.6±6.1	–35.8±7.5	–28.6±8.2	–10.6±5.4	–13.5±6.8	–27.0±5.2	0.2±2.2	10.7±5.9
***syn*–BzG:*anti*–C**	χ: 5-200°	0.73	–20.6±10.4	–23.3±8.7	–19.3±7.2	–19.8±8.2	–19.7±7.4	–18.2±17.5	–7.6±7.0	–2.4±15.5
	χ: 200-365°	0.27	–26.9±7.2	–13.8±6.7	–21.3±6.0	–23.5±9.5	–23.3±5.5	–16.1±7.7	–8.7±2.4	–9.5±7.7
***syn*–BzG:*anti*–T**	φ: 0-110°	0.03	–29.1±6.2	–28.6±6.7	–12.0±4.6	–4.2±4.9	–22.2±7.3	–24.0±4.5	–3.1±4.5	7.4±5.0
	φ: 110-220°	0.28	–37.6±5.5	–30.1±6.0	–16.0±7.2	–7.9±4.8	–12.5±8.8	–22.9±5.6	–1.2±3.9	9.5±5.4
	φ: 220-360°	0.70	–48.6±5.8	–21.6±4.7	–10.1±4.8	–7.9±6.0	–20.4±6.8	–25.5±5.0	–4.8±3.9	2.4±8.6
***syn*–BzG:*anti*–C**	extrahelical	1.00	–7.1±7.6	–27.2±7.5	–35.8±6.7	–20.4±6.5	–0.6±1.1	–19.9±22.1	–0.2±1.3	2.8±22.8
***syn*–BzG:*anti*–T**	extrahelical	1.00	–15.5±9.7	–29.2±8.0	–31.9±8.9	–21.1±7.7	3.4±8.4	0.6±1.7	0.0±1.0	1.6±1.7
***anti*–BzG:*syn*–A**	all	1.00	–22.3±6.7	–29.6±8.4	–18.2±5.8	–10.4±6.6	–13.1±5.3	–28.1±4.8	–2.5±2.6	3.6±3.6
***anti*–BzG:*syn*–G**	φ: 0-120°	0.19	–33.3±10.8	–27.4±10.0	–23.1±8.6	–5.8±8.9	–5.9±11.0	–10.1±8.5	–3.3±5.8	–7.2±7.0
	φ: 120-240°	0.74	–24.8±7.1	–24.3±7.2	–20.6±8.4	–1.0±9.6	–0.2±13.2	–8.6±11.3	–1.3±6.1	–5.3±8.6
	φ: 240-360°	0.08	–21.9±6.7	–34.6±8.9	–22.1±7.2	–11.7±7.0	–4.8±11.5	–13.3±11.1	0.2±5.6	–3.1±8.3
***syn*–BzG:*anti*–G**	φ: 0-125°	0.23	–27.7±6.4	–27.3±7.6	–10.5±5.3	–1.4±5.2	–9.0±8.4	–14.9±14.1	–20.0±9.8	–14.6±9.9
	φ: 125-230°	0.23	–35.6±6.8	–21.8±5.5	–9.2±6.7	–2.1±7.1	–6.4±10.6	–26.6±10.2	–14.0±8.0	–11.2±8.5
	φ: 230-360°	0.54	–45.8±7.1	–18.7±4.6	–11.8±5.1	–5.7±7.0	–5.8±6.7	–11.4±11.1	–17.9±6.4	–12.7±8.4

[a]See Figure 2.1 for definitions of χ and φ. [b]Interaction between residues 6 and 7 (BzG). [c]Interaction between residues 7 (BzG) and 8. [d]Interaction between residues 7 (BzG) and 19. [e]Interaction between residues 7 (BzG) and 17. [f]Interaction between residues 18 and 19. [g]Interaction between residues 18 and 17. [h]Interaction between residues 18 and 6. [i]Interaction between residues 18 and 8.

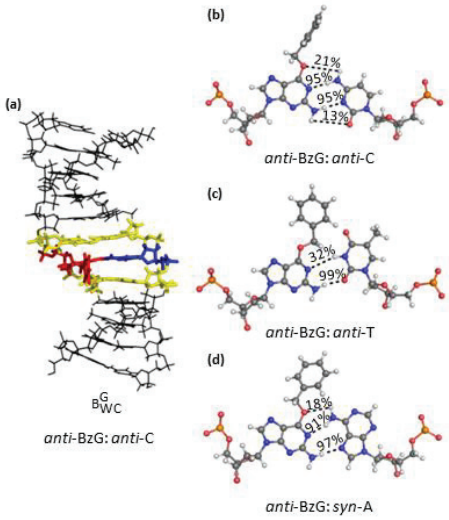

Figure 2.8. (a) Representative MD structure of BzG adducted DNA with the Bz moiety located in the major groove. (b-d) Interactions within *anti*–BzG base pairs in representative MD structures (hydrogen-bonding occupancies were defined based on a distance cut-off of < 3.4 Å and an angle cut-off of < 120°).

In DNA containing *anti*–BzG paired opposite *anti*–T, one hydrogen bond is formed between the amino group of BzG and O2 of the pairing T for 99% of the simulation, while another interaction is present between N1 of BzG and N3–H of T for only 32% of the simulation (Figure 2.8c). The resulting hydrogen bond is a pseudo-Watson-Crick base pair, which has a significantly weaker hydrogen-bonding energy (–20.4 ± 4.8 kJ mol^{-1}) compared to the corresponding G:C hydrogen-bonding interaction (–118.4±7.2 kJ mol^{-1}, Table 2.2). Nevertheless, the π–interactions between *anti*–BzG:*anti*–T and the neighbouring pairs (–128.2 kJ mol^{-1}, Tables 2.3 and A.3, Appendix A) are similar in strength to the C interactions (–102.5 kJ mol^{-1}, Tables 2.3 and A.3, Appendix A).

More than one BzG conformation is adopted when *syn*–BzG is paired with the (*anti*) pyrimidines (Figure 2.9). In the case of the C pair, the simulation can be clustered about two χ ranges (Table 2.1). In both structures, the benzyl moiety intercalates into the DNA

strand, but does not push the opposing C into an extrahelical position (Figures 2.9a and b). Nevertheless, C adopts a high *anti* orientation (χ = 270–300°) in both helices (χ = 281.1 ± 32.4°), which shifts the Watson-Crick hydrogen-bonding face away from the adduct. In the *syn*–BzG:C pairs, the benzyl moiety converts between an intercalated structure stacked with cytosine on the 3′ side with respect to the adduct (χ = 5 – 200°) and partial intercalation into the DNA helix from the major groove, which positions BzG in the same plane as cytosine (χ = 200 – 365°). Interestingly, 5′ intercalation with respect to the adduct reverts to 3′ intercalation upon simulation, possibly due to the orientation of the adduct relative to the helical twist, despite the inherent flexibly of BzG and the fact that the adduct base pair is flanked by G:C base pairs on both sides at G^3 in the *Nar*I sequence. The stacked structure is present for 73% of the simulation. These alternate binding arrangements are accompanied by different sugar puckers (C1′-*exo* or O4′-*endo*), yet maintain the C1′–C1′ distance of natural DNA (Table 2.1). In both conformations, one hydrogen bond is present in the *syn*–BzG:*anti*–C pair between the amino group of C and N7 of BzG for over 93% of the simulation, which yields one of the strongest (most negative) interactions between BzG and the opposing nucleotide among the pairs considered in the present work (–46 kJ mol^{-1}, Table 2.2, 2.3 and A.3, Appendix A). In addition, the sum of the 5′ and 3′ intra and interstrand π–interactions between BzG:C and the flanking pairs is very strong, being –83.0 kJ mol^{-1} (–85.5 kJ mol^{-1}) and –47.9 kJ mol^{-1} (–57.6 kJ mol^{-1}) for BzG and C interactions when χ = 5–200° (χ = 200-365°), respectively (the corresponding G and C interactions in the natural helix are –67.4 and –35.1 kJ mol^{-1}, respectively).

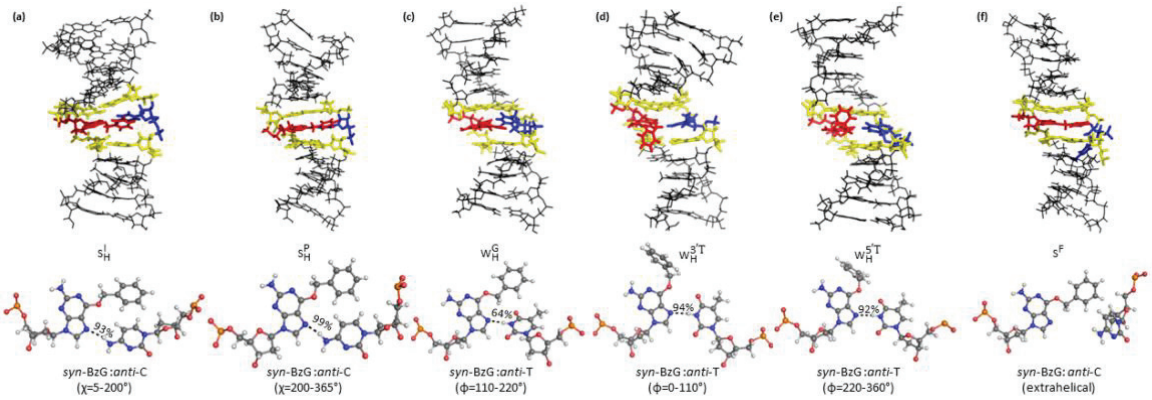

Figure 2.9. Representative MD structures (top) and interactions (bottom) involving *syn*–BzG observed during 20 ns MD simulations on the adducted *Nar*I DNA helices with various pairing nucleotides (hydrogen-bonding occupancies were defined based on a distance cut-off of < 3.4 Å and an angle cut-off of < 120°).

In DNA containing the *syn*–BzG:*anti*–T pair, three BzG orientations are intermittently adopted that primarily differ in φ (Table 2.1 and Figures 2.9c–e). When φ equals 175°, the bulky moiety is partially intercalated into the helix from the major groove and located beside (without displacing) the opposing T (Figure 2.9c). In this conformation, T adopts an *anti* orientation ($\chi = 225.1 \pm 16.3°$), which contrasts the high *anti* conformation adopted by C opposite *syn*–BzG. When φ equals 80 or 280°, the bulky group extends into the major groove, away from the opposing T, and adopts a T-shaped contact (involving the π–system of the benzyl ring and the edge of the intrastrand flanking nucleobase) with either the 3′ or 5′ flanking base with respect to BzG, respectively (Figures 2.9 d and e). Interestingly, there is much less variation in ξ for the T-shaped orientations compared to the free rotation observed in other conformations (Table 2.1). In all three BzG:T structures, one hydrogen bond forms between N3–H of T and N7 of BzG (Figures 2.9 c–e), which has a smaller occupancy when the benzyl moiety is intercalated (64%) than projected into the major groove (> 92%). The hydrogen-bonding interaction energy between BzG and T is slightly

greater for the intercalated structure (by up to 4.8 kJ mol^{-1}, Table 2.2). Nevertheless, the overall interaction energy is strongest for the 3′ T-shaped structure, but is still at least 72.9 kJ mol^{-1} smaller than that for the natural G:C dimer (Tables 2.2, 2.3 and A.3, Appendix A). In contrast, all intra and interstrand interactions with the BzG base pair are larger than for the natural G:C helix. Finally, the predominant sugar pucker is O4′-*endo* (with occupancy of 54 – 83%) and the C1′–C1′ distance is significantly shorter (by up to 2.2 Å) in the *syn*–BzG:*anti*–T dimers compared to the natural base pairs.

Since the most commonly observed intercalated structures of adducted DNA reported in the literature involve an extrahelical pairing pyrimidine,[43, 57, 68, 79, 80] simulations were initiated from a structure with BzG intercalated into the helix and the pairing (*anti*) C or T flipped out of the helix. Stable extrahelical conformations were present for the duration of the simulation (Figures 2.9f and A.5, Appendix A). In this conformation, the BzG adduct adopts the C1′-*exo* pucker and the BzG dihedrals behaves as discussed for other simulations (Table 2.1). Since the pairing base is extrahelical, hydrogen bonding does not occur between BzG and the opposing strand. Instead, BzG forms stacking interactions with the inter and intrastrand flanking bases that are ~30 kJ mol^{-1} stronger than the stacking in the natural helix (up to 7 kJ mol^{-1} stronger than the BzG stacking interactions when the pairing base is not extrahelical; Tables 2.3 and A.3, Appendix A). Additionally, since the pairing pyrimidine is outside the helix and typical DNA stacking interactions are not possible, alternative stabilizing interactions are formed. Specifically, the extrahelical C forms hydrogen bonds with the 5′-phosphate and O2 of the C +2 with respect to the BzG, while the extrahelical T forms a T-shaped interaction with the 3′ flanking base (Figure 2.10).

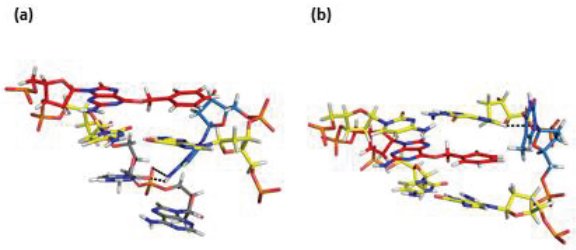

Figure 2.10. Discrete interactions between the extrahelical (a) C or (b) T and the surrounding bases in representative structures from MD simulations with *syn*-BzG intercalated in the DNA helix and the pairing base extrahelical.

2.3.4.2. Structure of DNA Containing BzG Paired Opposite a Purine: Only the *anti*–BzG:*syn*–A combination affords hydrogen bonding between A and BzG. In this pair (Figure 2.8d), the benzyl moiety is located in the major groove. The BzG adduct has a similar C2′-*endo* sugar pucker occupancy as G in the natural helix (~45%, Table 2.1), while the C1′–C1′ distance is slightly larger than the natural distance (by 0.6 Å). Two hydrogen bonds are present between BzG and A for the majority of the simulation (> 91%, Figure 2.8d), with a corresponding interaction energy of –41.7 ± 6.3 kJ mol^{-1} (Table 2.2). The π–interactions involving the *anti*–BzG:*syn*–A base pair are approximately –120.6 kJ mol^{-1} (Tables 2.3 and A.3, Appendix A), which is similar to the *anti*–BzG:*anti*–C pair (–102.5 kJ mol^{-1}, Tables 2.3 and A.3, Appendix A).

When BzG is paired opposite G in either the *anti* or *syn* orientation, three distinct BzG conformations occur with respect to φ (Figure 2.11). In the first (φ ≈ 180°), the benzyl moiety is solvent exposed and located in the major groove (Figure 2.11 and A.5, Appendix A). In this conformation, the π–interactions with the flanking bases are larger than those in the natural helix (by up to –27.3 kJ mol^{-1}, Tables 2.3 and A.3, Appendix A). In the other two BzG conformations, BzG adopts φ ≈ 80° or φ ≈ 280° and the benzyl moiety forms a T-

shaped interaction with the 3′ or 5′ flanking base with respect to BzG, respectively (Figure 2.11 and A.5, Appendix A). The T-shaped contact leads to stronger interactions between BzG and the flanking bases than observed in the natural helix (by up to –24.4 kJ mol^{-1}, Tables 2.3 and A.3, Appendix A).

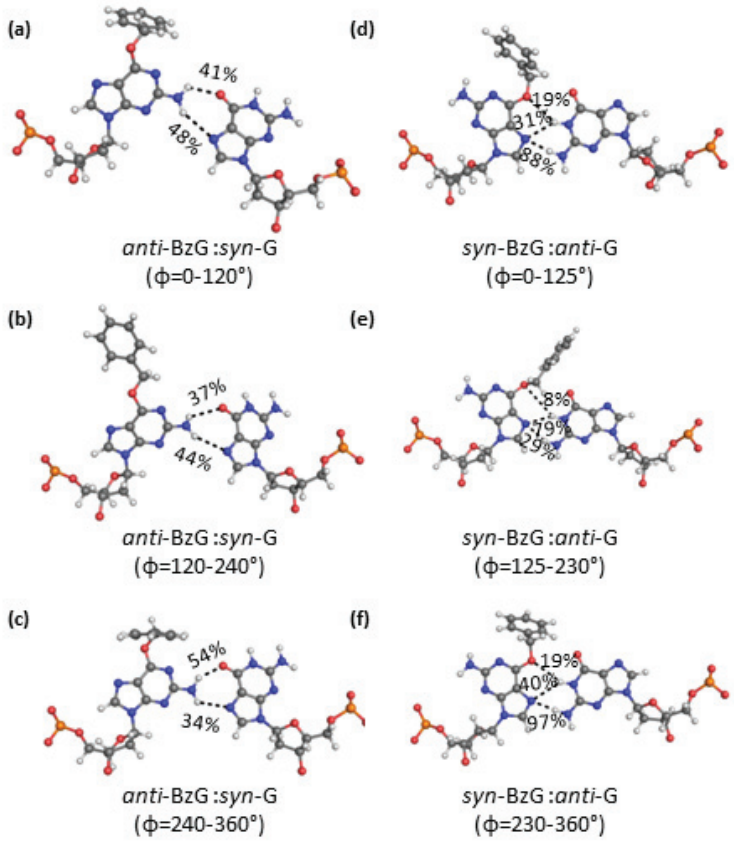

Figure 2.11. BzG interactions in representative structures observed during 20 ns MD simulations on adducted *Nar*I DNA helices with various pairing nucleotides (hydrogen-bonding occupancies were defined based on a distance cut-off of < 3.4 Å and an angle cut-off of < 120°).

In the BzG:G pairs, the C1′–C1′ distance is typically within 0.1 Å of the corresponding distance in the natural G:C pair, and BzG adopts C3′-*endo* or O4′-*endo* pucker in the *anti* or *syn* conformation, respectively (Table 2.1). The *anti*–BzG:*syn*–G pair contains two hydrogen bonds (Figures 2.10 a–c) with occupancies less than 55% regardless of the φ

conformation, which results in a small interaction energy (~–11 kJ mol^{-1}, Tables 2.2 and 2.3). In contrast, three hydrogen bonds form in the *syn*–BzG:*anti*–G pair in all φ configurations (Figures 2.10 d–f), but only one hydrogen bond has an occupancy greater than 50%, which always involves N7 of BzG. Nevertheless, the stabilities of the hydrogen-bonded *syn*–BzG:*anti*–G pairs (~–45 kJ mol^{-1}) are similar to those for other BzG pairs with two hydrogen bonds of high occupancy (Tables 2.2 and A.3, Appendix A). In both the *anti* and *syn* BzG conformations, the interactions between BzG and the inter or intrastrand flanking bases are greater (by up to 22.9 kJ mol^{-1}) than the corresponding natural G interaction energies (Table 2.3). In contrast, the interaction between the flanking bases and the pairing G is (up to 19.7 kJ mol^{-1}) smaller in *anti*–BzG:*syn*–G and (up to 23.4 kJ mol^{-1}) greater in *syn*–BzG:*anti*–G compared to the natural G:C (Tables 2.3 and A.3, Appendix A).

2.3.4.3. Energetics of DNA Helices Containing BzG: Due to the large number of helices and the resulting 16 structures considered in the present work, the total free energy (Tables 2.4 and A.4, Appendix A) is used to determine the preferred orientation of each BzG pairing combination prior to comparing the stabilities across strands with different compositions. The total free energy is the energy required for the two single DNA strands to bind, while neglecting the conformational and solvation changes in the monomer strands upon binding. When BzG is opposite C, four base pairing conformations are possible, namely *anti*–BzG:*anti*–C, *syn*–BzG:*anti*–C (intercalated stacked), *syn*–BzG:*anti*–C (intercalated) and *syn*–BzG:*anti*–C (extrahelical). Although the free energy is lower (helix is more stable) for the three *syn*–BzG:*anti*–C geometries than the *anti*–BzG:*anti*–C geometry, there is only a 6.2 kJ mol^{-1} difference between all orientations. Thus, all four geometries may exist in the cell and no structure can be discounted based on the free energy of helix formation. When

BzG is paired opposite T, five distinct base pairing conformations are adopted, the major groove conformation associated with *anti*–BzG, and the intercalated, extrahelical, 5′–T-shaped and 3′–T-shaped conformations associated with *syn*–BzG. However, the *anti*–BzG:*anti*–T major groove orientation and *syn*–BzG:*anti*–T (extrahelical) orientations are highly favored over the remaining *syn*–BzG orientations (by up to 91.7 kJ mol^{-1}; Tables 2.4 and A.4, Appendix A). Since the energy difference between the *anti*–BzG:*anti*–T major groove orientation and *syn*–BzG:*anti*–T (extrahelical) orientations is only 2.0 kJ mol^{-1}, both will be further considered. Finally, six distinct binding arrangements occur when BzG is paired opposite G, namely the major groove, 5′–T-shaped and 3′–T-shaped conformations associated with *anti*–BzG and the major groove, 5′–T-shaped and 3′–T-shaped conformations associated with *syn*–BzG. The *syn*–BzG:*anti*–G geometries are much more stable than the *anti*–BzG:*syn*–G (by up to 153.3 kJ mol^{-1}; Tables 2.4 and A.4, Appendix A). However, all *syn*–BzG structures are within the anticipated accuracy of our methodology and therefore must be further analyzed.

With the conformations for each base pairing combination limited by the total free energy, the preferred opposing base when BzG is located at the G^3 position in the *Nar*I sequence can be determined using the free energy of helix binding (relative ΔG, Tables 2.5 and A.5, Appendix A), which is the difference in free energy between the DNA helix and the two individual unbound strands. Although the free energies of helix binding are less accurate than the total free energies, the free energy of helix binding is the only method available to energetically compare different base pairing possibilities within a DNA helix over an entire MD simulation. The helix with the overall lowest free energy contains the *syn*–BzG:*anti*–C pair. When the helix stabilities for the preferred orientation(s) of the other base pairs are

compared, it is clear that the *syn*–BzG:*anti*–C pairing is the only combination that yields a more stable helix than the corresponding natural helix (Tables 2.5 and A.5, Appendix A). In fact, this pair is lower in energy than any other natural or damaged combination (by at least 6 kJ mol^{-1}). The next most stable binding conformation is the *syn*–BzG:*anti*–T arrangement (relative ΔG = 3.4 kJ mol^{-1}). The *syn*–BzG:*anti*–G arrangement is the next most stable, with the relative ΔG = 19.8 – 22.9 kJ mol^{-1} depending on the benzyl moiety conformation. Finally, helices with *anti*–BzG paired opposite *syn*–A, *anti*–T or *anti*–C all have approximately the same relative helix stability (~25 kJ mol^{-1} relative to the natural G:C pair, Tables 2.5 and A.5, Appendix A), which is not significantly larger than the *syn*–BzG:*anti*–G pair.

Table 2.4. MM-PBSA total free energy (kJ mol^{-1}) for the *Nar*I sequence with the BzG adduct incorporated at the G^3 position and various opposing bases from 20 ns MD simulations.

Simulation	Cluster[a]	Conformation	Occupancy	Relative G[b]
***anti*–BzG:*anti*–C**	all	B^{G}_{WC}	1.00	6.2
***syn*–BzG:*anti*–C**	χ: 5-200°	S^{I}_{H}	0.73	2.9
	χ: 200-365°	S^{P}_{H}	0.27	3.4
***syn*–BzG:*anti*–C**	extrahelical	S^{F}	1.00	0.0
***anti*–BzG:*anti*–T**	all	B^{G}_{WC}	1.00	0.0
***syn*–BzG:*anti*–T**	φ: 0-110°	$B^{3'T}_{H}$	0.26	91.7
	φ: 110-220°	S^{P}_{H}	0.11	46.0
	φ: 220-360°	$B^{5'T}_{H}$	0.70	26.6
***syn*–BzG:*anti*–T**	extrahelical	S^{F}	1.00	2.0
***anti*–BzG: *syn*–G**	φ: 0-120°	$B^{3'T}_{WC}$	0.19	79.4
	φ: 120-240°	B^{G}_{WC}	0.74	64.9
	φ: 240-360°	$B^{5'T}_{WC}$	0.08	153.3
***syn*–BzG:*anti*–G**	φ: 0-125°	$B^{3'T}_{H}$	0.23	21.7
	φ: 125-230°	B^{G}_{H}	0.23	11.7
	φ: 230-360°	$B^{5'T}_{H}$	0.54	0.0

[a]See Figure 2.1 for definitions of χ and φ. [b]Relative total free energy with respect to the lowest energy conformation of each pairing combination.

Table 2.5. MM-PBSA free energy of helix binding (kJ mol^{-1}) for the *Nar*I sequence with the BzG adduct incorporated at the G^3 position and various opposing bases from 20 ns MD simulations.

Simulation	Cluster[a]	Conformation	Occupancy	Relative ΔG [b]
***syn*–BzG:*anti*–C**	extrahelical	S^F	1.00	–27.6
***syn*–BzG:*anti*–C**	χ: 5-200°	S^I_H	0.73	–16.3
	χ: 200-365°	S^P_H	0.27	–6.8
***anti*–G:*anti*–C**	all	B^G_{WC}	1.00	0.0
***syn*–BzG:*anti*–T**	extrahelical	S^F	1.00	3.4
***syn*–BzG:*anti*–G**	φ: 0-125°	$B^{3'T}_H$	0.23	22.0
	φ: 125-230°	B^G_H	0.23	22.9
	φ: 230-360°	$B^{5'T}_H$	0.54	19.8
***anti*–BzG:*anti*–T**	all	B^G_{WC}	1.00	24.4
***anti*–BzG:*syn*–A**	all	B^G_{WC}	1.00	25.3
***anti*–BzG:*anti*–C**	all	B^G_{WC}	1.00	26.8

[a]See Figure 2.1 for definitions of χ and φ. [b]Relative free energy of helix binding with respect to the lowest energy pairing.

2.4. Discussion

2.4.1. Intrinsic Flexibility of the Bulky Moiety: In both the *anti* and *syn* conformations, the isolated BzG nucleoside and nucleotide preferentially adopt a conformation that places the O6 lone pairs towards N7 (θ ≈ 180°, Figures 1, 2 and 4–7), which orients the bulky moiety toward the Watson-Crick binding face of G and agrees with computational studies on the smaller MeG lesion.[28, 29] Nevertheless, rotation about the linkage bond is accessible, with an interconversion barrier of less than ~25 kJ mol^{-1} (Figures 2.2, 2.4 and 2.5). The orientation of the phenyl group is highly variable due to low rotational barriers about the φ (< 5 kJ mol^{-1}) and ξ (barrierless) dihedral angles (Figures 2.4 and 2.5). Together, this combination of linkage and bulky moiety dihedral angles leads to a parallel or perpendicular arrangement of the bulky group relative to the base. This relative ring

arrangement is maintained upon incorporating BzG into DNA. However, unlike reported for MeG,[81] significant rotation occurs about θ in the MD simulations. Nevertheless, 9 out of 16 structures adopt θ ≈ 180° on average, where alternate θ values occur in the remaining 7 structures to maximize hydrogen-bonding interactions (Figures 2.8b and d) or reduce steric repulsion (Figures 2.9d and 2.10a). The low rotational barrier about ξ in the nucleoside model leads to great flexibility about this dihedral in DNA. In contrast, there is little rotation about φ due to steric clashes despite the fact that the φ rotational barrier is ~20 kJ mol^{-1} lower than the θ rotational barrier according to the quantum mechanical models. The range of BzG conformations adopted in DNA is greater than previously reported for other adducts (for example 41, 49, 57, 82), which is likely due to a combination of the longer and more flexible (–CH_2–) linker, and the lack of discrete interactions between the bulky moiety and G. The implications of the BzG flexibility on the structures of the associated adducted DNA helices are discussed in the following sections.

2.4.2. Overview of Adducted DNA Conformations: When BzG is incorporated into DNA, many helical conformations direct the bulky moiety into the solvent exposed major groove (Figures 2.8, 2.10 and A.5, Appendix A), which is consistent with previous literature suggesting that the bulky moieties of more flexible adducts are generally found in a DNA groove.[23, 54, 59, 60] However, the BzG bulky moiety can also form T-shaped interactions with the 3′ or 5′ flanking base (Figures 2.9, 2.10 and A.5, Appendix A). Notably, the T-shaped interactions only form in helices containing a mismatch, and the contacts can be present for the majority of the MD simulation (up to 70% for a single T-shaped conformation). To the best of our knowledge, such T-shaped conformations have not been reported for any other DNA adducts. The novelty of these interactions for BzG

may not be surprising since the ability to form such contacts is dependent on the conformational flexibility of the adduct. In addition to the most widely adopted stacked orientations in the literature, which involve an extrahelical opposing pyrimidine,[43, 49, 57, 68, 83, 84] other intercalated structures are possible for BzG adducted DNA due to the small size of the benzyl ring (Figure 2.9). Specifically, the bulky moiety can intercalate beside the opposing base or between the opposing base and the 3′ flanking nucleobase with respect to the adduct. Such intercalated, yet not base displaced, structures have been previously reported for select (PAH) adducts.[85-90]

2.4.3. Modified Nomenclature for Adducted DNA Conformations: Nomenclature currently used to describe the conformations of adducted DNA is based on the structures adopted by DNA-containing a C8–major groove aromatic amine adduct. Although this classification system can naturally be extended to all major groove adducts, the terminology no longer describes the diversity of structures reported in the present study, or found when previous literature is carefully searched.[85-90] In the existing terminology, DNA structures are classified based on the helical location of the bulky moiety according to minor groove or wedged (W-conformation), major groove (B-conformation), or intercalated (S-conformation). The conformations adopted by BzG highlight the substantial structural variation within each of these classes. Therefore, we introduce a modified terminology that uses subscripts and superscripts to divide the three broad conformations according to distinctive structural features (Table A.6, Appendix A). For the W and B conformations, superscripts G or T are used to distinguish between conformations that place the bulky moiety in a (major/minor) DNA groove versus those that form T-shaped interactions with a flanking base. In addition, superscripts are also used to signify whether the T-shaped

interaction occurs with the 3′ or 5′ base (with respect to the adduct), while the subscript indicates whether the Hoogsteen (H) or Watson-Crick (WC) face of the adduct is available to hydrogen bond with the opposing base. Thus, for example, W_H^G indicates a wedged conformation with the bulky moiety in a DNA groove with the Hoogsteen face available for hydrogen bonding with the opposing base, while $B_{WC}^{3'T}$ represents a B-type conformation with the bulky moiety forming a T-shaped interaction with the 3′ base flanking the adduct, and the adduct forms hydrogen bonds with the opposing base using the Watson-Crick face. For the S conformation, the superscript P is used to indicate the bulky moiety is partially intercalated into the DNA helix from the DNA groove and lies beside the pairing nucleotide, while the superscripts I and F distinguish between full intercalation into the DNA helix without (I) and with (F) the opposing base flipped into an extrahelical position. This modified nomenclature clarifies important helical interactions that distinguish the conformation of the adduct in DNA.

2.4.4. Comparison of BzG Adducted DNA Conformation with Other Adducts: Using our new nomenclature for adducted DNA, *anti*–BzG paired opposite *anti*–C, *anti*–T or *syn*–A adopts the B_{WC}^G conformation, while *anti*–BzG paired opposite *syn*–G adopts the B_{WC}^G, $B_{WC}^{5'T}$ or $B_{WC}^{3'T}$ conformation. Among the B-conformations, the B_{WC}^G conformation identified for BzG has been most widely observed in the literature (for example, the C8-PhOH-G,[49] C8-OTA-G,[57] C8-AF-G,[22] O6-POB-G[41] and C8-AP-G[91] adducts adopt the B_{WC}^G conformation).

In the *syn* conformation, BzG adopts both B and S-conformations. Major groove conformations are adopted when *syn*–BzG is paired opposite *anti*–G (B_H^G, $B_H^{5'T}$ or $B_H^{3'T}$) or *anti*–T ($B_H^{5'T}$ or $B_H^{3'T}$). Conversely, previously studied adducts (i.e., PhOH-G,[49, 51] 8-Br-G,[92] Pyridyl-G[93] and Bth-G[68]) generally form the W_H^G conformation in the *syn* orientation, which stabilizes the G*:G mismatch. Despite the fact that unique T-shaped interactions can form with BzG and contribute up to ~10 kJ mol^{-1} to the intrastrand BzG non-covalent interactions, the sum of all π–interactions for the adducted pair in the $B_H^{5'T}$ or $B_H^{3'T}$ conformations is not necessarily stronger than those in B_H^G or S_H^P conformations (without the T-shaped contact), and highly depends on the relative alignment between the adducted and neighbouring pairs (Tables 2.2, 2.3 and A.3, Appendix A). Furthermore, the strength, and therefore likely the frequency, of these T-shaped conformations will be dependent on the flanking bases, where T-shaped interactions between the benzene (π–system) and a DNA nucleobase (edge) are strongest with A and weakest with T.[94]

When other adducts are in the *syn* orientation opposite C, the pairing base typically preferentially adopts an extrahelical position (for example, helices containing C8-BthG,[68] C8-AP-G,[43] N2-BP-G[80] C8-OTB-G[57] or AAII-A[79] adopt the S^F conformation). In stark contrast to MD studies on other (C8-BthG and C8-OTB-G)[57, 68] adducts, the barrier to rotate the pairing pyrimidine out of the helix is not surpassed during the 20 ns MD simulations on BzG adducted DNA in the present study. Instead, the S_H^P and S_H^I conformations are adopted, which are more stable than natural DNA. Nevertheless, the S^F conformation is stable during simulations initiated with an extrahelical pyrimidine and this conformation is in fact energetically preferred (Table 2.4). Nevertheless, the high stability of all *syn*-

BzG:*anti*-C pairing combinations suggests that the point at which the damage occurs in a cell may dictate the S conformation adopted. For example, if the double helix is damaged, the barrier to rotate the pairing C out of the helix will likely prohibit the S^F conformation and a conformation with C intrahelical (S^P_H or S^I_H) will form. Alternatively, if the DNA base is damaged when bound to a protein, steric constraints imposed by the protein will likely regulate the BzG adducted DNA conformation. Finally, if in the rare instance single-stranded DNA is damaged, then the S^F conformation may be adopted upon complexion with the complementary strand due to the anticipated intrinsic flexibility of BzG in single-stranded DNA.

Although the S^F conformation has been most widely cited in the literature, [43, 57, 68, 80, 84] the S^I_{WC} conformer has been observed for a variety of polyaromatic hydrocarbon (PAH) adducts, including the minor groove N2-G and major groove N6-A adducts.[85-90] Although BzG is a relatively small adduct with one ring in the bulky moiety, the PAH adducts that form the S^I_{WC} conformation are composed of four or five fused rings. Furthermore, BzG gains the flexibility required to both intercalate and hydrogen bond through the $–CH_2–$ linker, while the PAHs obtain the required flexibility through the puckering of a six-membered ring in the bulky group. Regardless, the larger PAHs lead to helical distortions. The global changes to the DNA helix formed upon incorporate of the BzG adduct will be discussed in the following section.

2.4.5. Distortion to the DNA Helix: Despite the large variation in the helical location of the bulky BzG moiety in DNA, no conformation of any nucleobase pairing combination

distorts the DNA helix through, for example, the introduction of a bulge or kink. Furthermore, any deviations in the helical parameters for BzG adducted DNA are consistent with the dynamics of natural helices.[95] Indeed, DNA is known to be flexible, with deviations in helical parameters (such as roll and twist) of up to 20° occurring with a change in the base pair sequence.[95] The general lack of distortion in the S^{I}_{H} structures is particularly noteworthy since there is little space between the base pairs in DNA (average calculated rise of 3.334 Å, Table A.2, Appendix A) to accommodate a bulky moiety. In fact, the average rise in DNA is comparable to the preferred vertical separation between stacked π–systems (~3.4 Å).[96] Indeed, many compounds that intercalate between stacked base pairs have been developed for pharmaceutical and scientific applications, and these compounds generally distort DNA[97-99] through unwinding (decrease in the twist angle), and/or an increase in the distance between neighboring base pairs (rise). The lack of helical distortion in BzG adducted DNA is consistent with previous findings that small adducts (with substituents composed of a single ring) generally do not significantly distort the helix, while larger DNA adducts (with bulkier substituents composed of multiple rings) cause notable helical distortion.[49, 68, 86, 87, 100] For example, the bulky (-)-trans-*anti*–DB[a,l]P-N6-dA adduct, which adopts the S^{I}_{WC} conformation opposite thymine, leads to an up to 38.6° deviation in the twist or 3.8 Å deviation in the rise in the adducted region,[87] while the smaller phenoxyl-G adducts does not cause helical distortion.[49]

2.4.6. Preferred Conformation about the Glycosidic Bond: Previous studies of C8-G adducts propose that a balance between unfavourable backbone–bulky moiety steric repulsion and favourable sugar–bulky moiety hydrogen bonding dictates the

conformational preference about the glycosidic bond,[57, 101, 102] which in turn influences the mutagenic and toxic consequences of adduct formation. However, the O6 bulky moiety in BzG does not interact with the DNA backbone. Therefore, the *anti* conformation is intrinsically preferred over the *syn* conformation in the nucleoside model as found for canonical G.[103] Nevertheless, the *anti/syn* energy difference is small (< 2 kJ mol^{-1} for the BzG nucleotide model). Therefore, the *anti/syn* preference of the adduct in DNA will be dictated by discrete interactions between BzG and the neighboring bases. Indeed, stabilizing interactions can cause a single structure to be adopted. For example, O^6-POB-G, which also contains a flexible linker, adopts only the *anti* orientation in DNA due to a strong hydrogen bond between the bulky moiety and pairing C.[41]

Within the DNA helix, the *anti* orientation of some previously studied G adducts have been stabilized through strong Watson-Crick hydrogen bonding with the opposing base.[21, 49, 68, 83, 104] In contrast, favourable stacking interactions with the flanking nucleobases have stabilized the *syn* conformation of previously studied adducts.[21, 49, 51, 57, 68, 83, 104] In the case of BzG, the preference for the *anti* or *syn* BzG orientation depends at least in part on discrete interactions with the bulky moiety. For example, the hydrogen-bond strength between an opposing *syn*–G and *anti*–BzG (–7 to –12 kJ mol^{-1}, Tables 2.2 and A.3, Appendix A) is significantly smaller than between an opposing *anti*–G and *syn*–BzG (–36 to –52 kJ mol^{-1}). Furthermore, the discrete stacking interactions for the adducted pair are larger for *syn*–BzG (–125 to –129 kJ mol^{-1}) than *anti*–BzG (–86 to –116 kJ mol^{-1}). The stronger discrete BzG interactions explain the greater calculated stability of the adducted helix with *syn*–BzG rather than *anti*–BzG opposing natural G.

Nevertheless, the *anti/syn* equilibrium of BzG in the DNA helix is not always simply dictated by discrete nucleobase interactions as discussed for G. For example, although the hydrogen-bonding (–42 kJ mol^{-1} for *anti*–BzG and –3 or –46 kJ mol^{-1} for *syn–BzG*) and stacking (–133 kJ mol^{-1} for *anti*–BzG and –110 to –143 kJ mol^{-1} for *syn*–BzG) interactions for the adducted BzG:C pair are either just as strong or stronger for the *anti*–BzG, the *syn* conformation is overall energetically preferred (Tables 2.2, 2.4, and A.3, Appendix A). Some reasons for the increased stabilization of the *syn* conformation include the increased solvation of the hydrophobic benzyl moiety in the B^{G}_{WC} conformation, which has been discussed in the literature for other adducts such as Bth-G,[68] PhOH-G,[49] PhO-G,[51] AAF-G[105] and IQ-G.[106] Additionally, for the S^{F} conformation, which has the smallest calculated hydrogen-bonding and reported stacking energies, contains additional discrete interactions with the extrahelical C that are not accounted for in our calculations of the discrete intrahelical interactions. Thus, the BzG adduct displays an *anti*/*syn* conformational heterogeneity that is highly dependent on the opposing base and the conformation of adducted DNA.

2.4.7. Preferred Base Pairing: According to MD simulations, the preferred base pairing of BzG is *syn*–BzG:*anti*–C > *syn*–BzG:*anti*–T > *syn*–BzG:*anti*–G > *anti*–BzG:*anti*–T ≈ *anti*–BzG:*syn*–A ≈ *anti*–BzG:*anti*–C (Table 2.4). In fact, *syn*–BzG:*anti*–C forms a more stable helix than the corresponding natural DNA due to particularly strong π–interactions between the BzG base pair and the flanking bases. Therefore, C may be incorporated opposite BzG at a higher rate than any other nucleobase. If the *syn*-BzG conformation is

adopted, pairing with T may also occur. Alternatively, if the *anti*-BzG conformation is adopted, helices of equal stability can be obtained with C, A or T incorporated opposite the lesion. Therefore, the nucleobase incorporated may be largely dependent on the polymerase used to replicate the DNA adduct as discussed in the next section.

2.4.8. Biological Implications: The alteration of a DNA nucleobase through the formation of a bulky adduct or an alkylation product can result in the incorrect replication or transcription of DNA, which can lead to mutations and, in the long term, cause cancer, premature aging or cardiovascular disease.[1-4] Large DNA lesions are typically removed by the nucleotide excision repair (NER) pathway, which relies on changes to DNA structure upon incorporation of a lesion to identify damaged sites.[107] Since BzG does not deform DNA, this lesion may not be readily repaired by NER, with the potential exception of the extrahelical (S^F) structures, but will likely be repaired by O6-alkylguanine-DNA alkyltransferase.[14] Therefore, BzG may lead to a high rate of mutation upon processing by biological machinery.

When BzG adducted DNA is replicated by Dpo4, C is inserted most frequently opposite BzG (70%), while A and T are also inserted (15% each).[38] The preferential incorporation of C is explained by the high stability of the strand with *syn*-BzG paired opposite this nucleobase (Table 2.4), which is in fact more stable in several conformations than natural DNA. This finding also explains the low rate of mutation when BzG is replicated by Dpo4.[29, 38] Nevertheless, a crystallography study of the replication of BzG adducted DNA by Dpo4 reveals only the *anti* orientation of the adduct bound in the enzymatic active site.[38]

Thus, the BzG conformation adopted may be different in a double strand versus when bound to a polymerase.

Experiments suggest that G is never incorporated opposite BzG by Dpo4.[29, 38] Although this initially appears to contradict our calculated relative helix stabilities (Table 2.4), modeling *syn*–BzG:*anti*–G in the Dpo4 active site reveals a less than 1 Å separation between N2 of BzG and Nη of R332 when the *syn*–BzG conformation with the bulky moiety in the minor groove is adopted (Figure A.6, Appendix A). Nevertheless, the *syn*–BzG:*anti*–C pair (in the S_H^P or S_H^I conformation) with the bulky moiety intercalated into the helix (rather than in the minor groove) is possible due to reduced steric repulsion with R332 in this *syn*-BzG orientation.

The relative BzG pairing with A and T is dependent on the polymerase considered, with equal (15%) incorporation by Dpo4[38] and preference for T by Pol η leads to a different proportion of pairing nucleotides (55% C, 31% T, 11% A).[32] This can be understood based on the possible adducted DNA conformations. Specifically, there is not space in the Dpo4 active site to accommodate the extrahelical T conformation opposite *syn*-BzG and the remaining *anti*-BzG conformations opposite T and A have similar energetics, which explains the experimental data. In contrast, all T and A BzG pairing conformations may be structurally feasible in the Pol η active site. Therefore, the close structural resemblance of the *anti*–BzG:*anti*–T rather than *anti*–BzG:*syn*–A base pair to G:C, coupled with the greater stability of BzG pairing with T, may be responsible for the observed preferred insertion of T over A by Pol η.

With the preferred BzG base pairings understood, the mutagenic behaviour can be considered. Although BzG will most likely pair with C, the *syn*–BzG:*anti*–C pair may be mutagenic due to the substantially different structure (intercalated with or without an extrahelical C) than natural G:C. Indeed, the replication of intercalated adducts (with or without and extrahelical pairing base) is known to cause frameshift mutations,[26] and therefore the adoption of intercalated conformations explains the frameshift mutations observed when BzG is replicated by pol η.[32] Additionally, the (high *anti*) orientation of the C opposing BzG in the S_H^P or S_H^I conformation may change the preferred C pairing upon replication. In the *anti*-BzG orientation, helices of equal stability are formed with BzG paired opposite A, T and C, and therefore there is the potential for mispairing upon replication. In particular, the incorporation of A opposite BzG can lead to G → T transversion mutations, while the incorporation of T opposite BzG can lead to G → A transition mutations, which agrees with the experimentally-observed pairing, and subsequent detection of transition and transversion mutations.[38]

In summary, the significant flexibility within the BzG moiety and about the glycosidic bond in the corresponding nucleoside and nucleotide is observed in all DNA models considered in the present work. This flexibility leads to many different DNA conformations, including unique T-shaped conformations and a variety of intercalated structures that have yet to be explicitly discussed for other DNA adducts. The conformation of BzG adopted in DNA is greatly dependent on a number of factors, including the pairing nucleotide, the discrete interactions in the helix and the solvation of the benzyl moiety. This in turn leads to a

complicated mutagenic and toxic profile that may invoke pairing with natural C, mispairs, or deletion mutations. Further studies with various DNA polymerases are required to fully explain the biological consequences of the BzG adduct.

2.5. References

(1) Hoeijmakers, J. H. J. (2009) DNA Damage, Aging, and Cancer. *New Engl. J Med. 361*, 1475-1485.

(2) Mattison, D., Shiromizu, K., and Nightingale, M. S. (1983) Oocyte Destruction by Polycyclic Aromatic Hydrocarbons. *Am. J. Ind. Med. 4*, 191.

(3) Ramos, K. S., and Moorthy, B. (2005) Bioactivation of Polycyclic Aromatic Hydrocarbon Carcinogens within the vascular Wall: Implications for Human Atherogenesis. *Drug Metab. Rev. 37*, 595-610.

(4) Jackson, S. P., and Bartek, J. (2009) The DNA-Damage Response in Human Biology and Disease. *Nature 461*, 1071-1078.

(5) Shrivastav, N., Li, D., and Essigmann, J. M. (2010) Chemical Biology of Mutagenesis and DNA repair: Cellular Responses to DNA Alkylation. *Carcinogenesis 31*, 59-70.

(6) Drabløs, F., Feyzi, E., Aas, P. A., Vaagbø, C. B., Kavli, B., Bratlie, M. S., Peña-Diaz, J., Otterlei, M., Slupphaug, G., and Krokan, H. E. (2004) Alkylation Damage in DNA and RNA—Repair Mechanisms and Medical Significance. *DNA Repair 3*, 1389-1407.

(7) Kanvah, S., Joseph, J., Schuster, G. B., Barnett, R. N., Cleveland, C. L., and Landman, U. (2009) Oxidation of DNA: Damage to Nucleobases. *Acc. Chem. Res. 43*, 280-287.

(8) Kuznetsova, A. A., Knorre, D. G., and Fedorova, O. S. (2009) Oxidation of DNA and its Components with Reactive Oxygen Species. *Russ. Chem. Rev. 78*, 659-678.

(9) Nabel, C. S., Lee, J. W., Wang, L. C., and Kohli, R. M. (2013) Nucleic Acid Determinants for Selective Deamination of DNA over RNA by Activation-Induced Deaminase. *Proc. Natl. Acad. Sci. U.S.A. 110*, 14225-14230.

(10) Nabel, C. S., Manning, S. A., and Kohli, R. M. (2011) The Curious Chemical Biology of Cytosine: Deamination, Methylation, and Oxidation as Modulators of Genomic Potential. *ACS Chem. Biol. 7*, 20-30.

(11) Loveless, A. (1969) Possible Relevance of O-6 Alkylation of Deoxyguanosine to the Mutagenicity and Carcinogenicity of Nitrosamines and Nitrosamides. *Nature 223*, 206-207.

(12) Newbold, R., Warren, W., Medcalf, A., and Amos, J. (1980) Mutagenicity of Carcinogenic Methylating Agents is Associated with a Specific DNA Modification. *Nature 283*, 596 - 599.

(13) Singer, B., Chavez, F., Goodman, M., Essigmann, J., and Dosanjh, M. (1989) Effect of 3' Flanking Neighbors on Kinetics of Pairing of dCTP or dTTP Opposite O6-methylguanine in a Defined Primed Oligonucleotide when Escherichia coli DNA Polymerase I is Used. *Proc. Natl. Acad. Sci. U.S.A. 86*, 8271-8274.

(14) Verbeek, B., Southgate, T. D., Gilham, D. E., and Margison, G. P. (2008) O6-Methylguanine-DNA Methyltransferase Inactivation and Chemotherapy. *Br. Med. Bull 85*, 17-33.

(15) Reh, B. D., DeBord, D. G., Butler, M. A., Reid, T. M., Mueller, C., and Fajen, J. M. (2000) O6-methylguanine DNA Adducts Associated with Occupational Nitrosamine Exposure. *Carcinogenesis 21*, 29-33.

(16) Taverna, P., and Sedgwick, B. (1996) Generation of an Endogenous DNA-methylating Agent by Nitrosation in Escherichia coli. *J. Bacteriol. 178*, 5105-5111.

(17) Autrup, H., and Stoner, G. D. (1982) Metabolism of N-nitrosamines by Cultured Human and Rat Esophagus. *Cancer Res. 42*, 1307-1311.

(18) Peterson, L. A. (1997) N-Nitrosobenzylmethylamine Is Activated to a DNA Benzylating Agent in Rats. *Chem. Res. Toxicol. 10*, 19-26.

(19) Margison, G. P., Santibáñez Koref, M. F., and Povey, A. C. (2002) Mechanisms of Carcinogenicity/Chemotherapy by O6-methylguanine. *Mutagenesis 17*, 483-487.

(20) Guarnieri, M., Biser-Rohrbaugh, A., Tyler, B., Gabikian, P., Bunton, T., Wu, Q., Weingart, J., and Carson, B. (2002) Toxicity of Intracranial and Intraperitoneal O6-benzyl guanine in Combination with BCNU Delivered Locally in a Mouse Model. *Cancer Chemoth. Pharm. 50*, 392-396.

(21) 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.

(22) Cho, B. P. (2004) Dynamic Conformational Heterogeneities of Carcinogen-DNA Adducts and Their Mutagenic Relevance. *J. Environ. Sci. Health, Part C 22*, 57-90.

(23) Cho, B. (2010) Structure–Function Characteristics of Aromatic Amine-DNA Adducts, In *The Chemical Biology of DNA Damage* pp 217-238, Wiley-VCH Verlag GmbH & Co. KGaA.

(24) Stover, J. S., Chowdhury, G., Zang, H., Guengerich, F. P., and Rizzo, C. J. (2006) Translesion Synthesis Past the C8- and N2-Deoxyguanosine Adducts of the Dietary Mutagen 2-Amino-3-methylimidazo[4,5-f]quinoline in the NarI Recognition Sequence by Prokaryotic DNA Polymerases. *Chem. Res. Toxicol. 19*, 1506-1517.

(25) Norman, D., Abuaf, P., Hingerty, B. E., Live, D., Grunberger, D., Broyde, S., and Patel, D. J. (1989) 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 28*, 7462-7476.

(26) Gu, Z., Gorin, A., Hingerty, B. E., Broyde, S., and Patel, D. J. (1999) Solution Structures of Aminofluorene [AF]-Stacked Conformers of the syn [AF]−C8-dG Adduct Positioned Opposite dC or dA at a Template-Primer Junction. *Biochemistry 38*, 10855-10870.

(27) Warren, J. J., Forsberg, L. J., and Beese, L. S. (2006) The Structural Basis for the Mutagenicity of O6-methyl-guanine Lesions. *Proc. Natl. Acad. Sci. U.S.A. 103*, 19701-19706.

(28) Zhachkina, A., Liu, M., Sun, X., Amegayibor, F. S., and Lee, J. K. (2009) Gas-Phase Thermochemical Properties of the Damaged Base O6-Methylguanine versus Adenine and Guanine. *J. Org. Chem. 74*, 7429-7440.

(29) Jena, N., and Bansal, M. (2011) Mutagenicity Associated with O6-methylguanine-DNA Damage and Mechanism of Nucleotide Flipping by AGT During Repair. *Physical biology 8*, 046007.

(30) Pence, M. G., Choi, J.-Y., Egli, M., and Guengerich, F. P. (2010) Structural Basis for Proficient Incorporation of dTTP Opposite O6-Methylguanine by Human DNA Polymerase ι. *J. Biol. Chem. 285*, 40666-40672.

(31) Snow, E. T., Foote, R. S., and Mitra, S. (1984) Base-Pairing Properties of O6-methylguanine in Template DNA During in vitro DNA Replication. *J. Biol. Chem. 259*, 8095-8100.

(32) Choi, J.-Y., Chowdhury, G., Zang, H., Angel, K. C., Vu, C. C., Peterson, L. A., and Guengerich, F. P. (2006) Translesion Synthesis across O6-Alkylguanine DNA Adducts by Recombinant Human DNA Polymerases. *J. Biol. Chem. 281*, 38244-38256.

(33) Haracska, L., Prakash, S., and Prakash, L. (2000) Replication Past O 6-Methylguanine by Yeast and Human DNA Polymerase η. *Mol. Cell Biol. 20*, 8001-8007.

(34) Patel, D. J., Shapiro, L., Kozlowski, S. A., Gaffney, B. L., and Jones, R. A. (1986) Structural Studies of the O6meG:T Interaction in the d(C-G-T-G-A-A-T-T-C-O6meG-C-G) Duplex. *Biochemistry 25*, 1036-1042.

(35) Patel, D. J., Shapiro, L., Kozlowski, S. A., Gaffney, B. L., and Jones, R. A. (1986) Structural Studies of the O6meG:C Interaction in the d(C-G-C-G-A-A-T-T-C-O6meG-C-G) Duplex. *Biochemistry 25*, 1027-1036.

(36) Gaffney, B. L., and Jones, R. A. (1989) Thermodynamic Comparison of the Base Pairs Formed by the Carcinogenic Lesion O6-methylguanine with Reference both to Watson-Crick Pairs and to Mismatched Pairs. *Biochemistry 28*, 5881-5889.

(37) Gaffney, B. L., Marky, L. A., and Jones, R. A. (1984) Synthesis and Characterization of a Set of Four Dodecadeoxyribonucleoside Undecaphosphates Containing O6-methylguanine Opposite Adenine, Cytosine, Guanine, and Thymine. *Biochemistry 23*, 5686-5691.

(38) Eoff, R. L., Angel, K. C., Egli, M., and Guengerich, F. P. (2007) Molecular Basis of Selectivity of Nucleoside Triphosphate Incorporation Opposite O6-Benzylguanine by Sulfolobus solfataricus DNA Polymerase Dpo4: Steady-State and Pre-Steady-State Kinetics and X-ray Crystallography of Correct and Incorrect Pairing. *J. Biol. Chem. 282*, 13573-13584.

(39) Gahlon, H. L., Schweizer, W. B., and Sturla, S. J. (2013) Tolerance of Base Pair Size and Shape in Postlesion DNA Synthesis. *J. Am. Chem. Soc. 135*, 6384-6387.

(40) Lim, S., Song, I., Guengerich, F. P., and Choi, J.-Y. (2012) Effects of N2-Alkylguanine, O6-Alkylguanine, and Abasic Lesions on DNA Binding and Bypass Synthesis by the Euryarchaeal B-Family DNA Polymerase Vent (exo–). *Chem. Res. Toxicol. 25*, 1699-1707.

(41) Peterson, L. A., Vu, C., Hingerty, B. E., Broyde, S., and Cosman, M. (2003) Solution Structure of an O6-[4-oxo-4-(3-Pyridyl)butyl]guanine Adduct in an 11mer DNA Duplex: Evidence for Formation of a Base Triplex. *Biochemistry 42*, 13134-13144.

(42) Shapiro, R., Sidawi, D., Miao, Y.-S., Hingerty, B. E., Schmidt, K. E., Moskowitz, J., and Broyde, S. (1994) 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. 7*, 239-253.

(43) Mao, B., Vyas, R. R., Hingerty, B. E., Broyde, S., Basu, A. K., and Patel, D. J. (1996) Solution Conformation of the N-(Deoxyguanosin-8-yl)-1-aminopyrene ([AP]dG) Adduct Opposite dC in a DNA Duplex. *Biochemistry 35*, 12659-12670.

(44) Penning, T. M. (2010) Polycyclic Aromatic Hydrocarbons: Multiple Metabolic Pathways and the DNA Lesions Formed. *The Chemical Biology of DNA Damage*, 131-155.

(45) Essigmann, J., Croy, R., Nadzan, A., Busby, W., Reinhold, V., Büchi, G., and Wogan, G. (1977) Structural Identification of the Major DNA Adduct Formed by Aflatoxin B1 in vitro. *Proc. Natl. Acad. Sci. U.S.A. 74*, 1870-1874.

(46) Brown, K., Hingerty, B. E., Guenther, E. A., Krishnan, V. V., Broyde, S., Turteltaub, K. W., and Cosman, M. (2001) Solution Structure of the 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine C8-deoxyguanosine Adduct in Duplex DNA. *Proc. Natl. Acad. Sci. U.S.A. 98*, 8507-8512.

(47) 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.

(48) Akanni, A., and Abul-Hajj, Y. J. (1999) Estrogen−Nucleic Acid Adducts: Dissection of the Reaction of 3,4-Estrone Quinone and Its Radical Anion and Radical Cation with Deoxynucleosides and DNA. *Chem. Res. Toxicol. 12*, 1247-1253.

(49) Omumi, A., Millen, A. L., Wetmore, S. D., and Manderville, R. A. (2011) Fluorescent Properties and Conformational Preferences of C-Linked Phenolic-DNA Adducts. *Chem. Res. Toxicol. 24*, 1694-1709.

(50) Läppchen, T., Pinas, V. A., Hartog, A. F., Koomen, G.-J., Schaffner-Barbero, C., Andreu, J. M., Trambaiolo, D., Löwe, J., Juhem, A., Popov, A. V., and den Blaauwen, T. (2008) Probing FtsZ and Tubulin with C8-Substituted GTP Analogs Reveals Differences in Their Nucleotide Binding Sites. *Chem. Biol. 15*, 189-199.

(51) 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.

(52) Wu, X., Shapiro, R., and Broyde, S. (1999) Conformational Analysis of the Major DNA Adduct Derived from the Food Mutagen 2-Amino-3-methylimidazo[4,5-f]quinoline. *Chem. Res. Toxicol. 12*, 895-905.

(53) Lin, C. H., Huang, X., Kolbanovskii, A., Hingerty, B. E., Amin, S., Broyde, S., Geacintov, N. E., and Patel, D. J. (2001) Molecular Topology of Polycyclic Aromatic Carcinogens Determines DNA Adduct Conformation: A Link to Tumorigenic Activity. *J. Mol. Biol. 306*, 1059-1080.

(54) Liang, F., and P. Cho, B. (2009) Enthalpy−Entropy Contribution to Carcinogen-Induced DNA Conformational Heterogeneity. *Biochemistry 49*, 259-266.

(55) Reeves, D. A., Mu, H., Kropachev, K., Cai, Y., Ding, S., Kolbanovskiy, A., Kolbanovskiy, M., Chen, Y., Krzeminski, J., and Amin, S. (2011) Resistance of Bulky DNA Lesions to Nucleotide Excision Repair Can Result from Extensive Aromatic Lesion–Base Stacking Interactions. *Nucleic Acids Res. 39*, 8752-8764.

(56) Kohda, K., Tsunomoto, H., Kasamatsu, T., Sawamura, F., Terashima, I., and Shibutani, S. (1997) Synthesis and Miscoding Specificity of Oligodeoxynucleotide Containing 8-Phenyl-2‘-deoxyguanosine. *Chem. Res. Toxicol. 10*, 1351-1358.

(57) Sharma, P., Manderville, R. A., and Wetmore, S. D. (2013) Modeling the Conformational Preference of the Carbon-Bonded Covalent Adduct Formed upon Exposure of 2′-Deoxyguanosine to Ochratoxin A. *Chem. Res. Toxicol. 26*, 803-816.

(58) Mantle, P. G., Faucet-Marquis, V., Manderville, R. A., Squillaci, B., and Pfohl-Leszkowicz, A. (2009) 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.

(59) Cho, B. P., and Zhou, L. (1999) Probing the Conformational Heterogeneity of the Acetylaminofluorene-Modified 2‘-Deoxyguanosine and DNA by 19F NMR Spectroscopy. *Biochemistry 38*, 7572-7583.

(60) 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.

(61) 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.

(62) Millen, A. L., Manderville, R. A., and Wetmore, S. D. (2010) Conformational Flexibility of C8-Phenoxyl-2′-deoxyguanosine Nucleotide Adducts. *J. Phys. Chem. B 114*, 4373-4382.

(63) Foloppe, N., Hartmann, B., Nilsson, L., and MacKerell, A. D. (2002) Intrinsic Conformational Energetics Associated with the Glycosyl Torsion in DNA: A Quantum Mechanical Study. *Biophys. J. 82*, 1554-1569.

(64) 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., Jr., J. A. M., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Keith, T., Kobayashi, R., Normand, J., Raghavchari, 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,

O., Foresman, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J. (2013) Gaussian 09, Gaussian, Inc., Wallingford CT.

(65) 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., Jr., J. A. M., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Keith, T., 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, O., Foresman, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J. (2010) Gaussian 09, Gaussian, Inc., Wallingford CT.

(66) Batra, V. K., Beard, W. A., Shock, D. D., Krahn, J. M., Pedersen, L. C., and Wilson, S. H. (2006) Magnesium-Induced Assembly of a Complete DNA Polymerase Catalytic Complex. *Structure 14*, 757-766.

(67) Burnouf, D., Koehl, P., and Fuchs, R. P. (1989) 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. 86*, 4147-4151.

(68) Manderville, R. A., Omumi, A., Rankin, K. M., Wilson, K. A., Millen, A. L., and Wetmore, S. D. (2012) Fluorescent C-Linked C8-Aryl-guanine Probe for Distinguishing syn from anti Structures in Duplex DNA. *Chem. Res. Toxicol. 25*, 1271-1282.

(69) 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. Comput. Chem. 25*, 1157–1174.

(70) Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins 65*, 712–725.

(71) Wang, J., Wang, W., Kollman, P. A., and Case, D. A. (2006) Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graphics Modell. 25*, 247–260.

(72) Cornell, W. D., Cieplak, P., Bayly, C. I., and Kollman, P. A. (1993) Application of RESP Charges to Calculate Conformational Energies, Hydrogen-Bond Energies, and Free-Energies of Solvation. *J. Am. Chem. Soc. 115*, 9620–9631.

(73) Case, D. A., Darden, T. A., III, T. E. C., Simmerling, C. L., Wang, J., R.E. Duke, R. L., Walker, R. C., Zhang, W., Merz, K. M., Roberts, B., Hayik, S., Roitberg, A., Seabra, G., Swails, J., Goetz, A. W., Kolossváry, I., Wong, K. F., Paesani, F., Vanicek, J., Wolf, R. M., Liu, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M.-J., Cui, G., Roe, D. R., Mathews, D. H., Seetin, M. G., Salomon-Ferrer, R., Sagui, C., Babin, V., Luchko, T., Gusarov, S., Kovalenko, A., and Kollman, P. A. (2012) AMBER 12, University of California, San Francisco.

(74) Case, D. A., Darden, T. A., Cheatham, T. E., III, Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Walker, R. C., Zhang, W., Merz, K. M., Roberts, B., Hayik, S., Roitberg, A., Seabra, G., Swails, J., Götz, A. W., Kolossvary, I., Wong, K. F., Paesani, F., Vanicek, J., Wolf, R. M., Liu, G., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, M., Hsieh, M.-J., Cui, G., Roe, D. R., Mathews, D. H., Seetin, M. G., Salomon-Ferrer, R., Sagui, C., Babin, V., Luchko, T., Gusarov, S., Kollman, P. A., and Kollman, P. A. (2013) AMBER 13, University of California, San Francisco.

(75) Miller, B. R., McGee, 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.

(76) Mu, H., Kropachev, K., Wang, L., Zhang, L., Kolbanovskiy, A., Kolbanovskiy, M., E. Geacintov, N., and Broyde, S. (2012) 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. 40*, 9675-9690.

(77) 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.

(78) Šponer, J., and Lankaš, F. (2006) *Computational Studies of RNA and DNA*. Vol. 2, Springer.

(79) Lukin, M., Zaliznyak, T., Johnson, F., and de los Santos, C. (2012) Structure and Stability of DNA Containing an Aristolactam II-dA Lesion: Implications for the NER Recognition of Bulky Adducts. *Nucleic Acids Res. 40*, 2759-2770.

(80) Cosman, M., Hingerty, B. E., Luneva, N., Amin, S., Geacintov, N. E., Broyde, S., and Patel, D. J. (1996) Solution Conformation of the (−)-cis-anti-Benzo[a]pyrenyl-dG Adduct Opposite dC in a DNA Duplex: Intercalation of the Covalently Attached BP Ring into the Helix with Base Displacement of the Modified Deoxyguanosine into the Major Groove. *Biochemistry 35*, 9850-9863.

(81) Parker, K., Cruzeiro-Hansson, L., and Goodfellow, J. M. (1993) Conformation and Flexibility of Oligonucleotides Containing O6-methylguanine: A Molecular Dynamics Study. *J. Chem. Soc., Faraday Trans. 89*, 2637-2650.

(82) Grad, R., Shapiro, R., Hingerty, B. E., and Broyde, S. (1997) A Molecular Mechanics and Dynamics Study of the Minor Adduct between DNA and the Carcinogen 2-(Acetylamino)fluorene (dG-N2-AAF). *Chem. Res. Toxicol. 10*, 1123-1132.

(83) Milhé, C., Dhalluin, C., Fuchs, R. P. P., and Lefèvre, F.-F. (1994) NMR Evidence of the Stabilisation by the Carcinogen N-2-acetylaminofluorene of a Frameshift Mutagenesis Intermediate. *Nucleic Acids Res. 22*, 4646-4652.

(84) Lukin, M., and de los Santos, C. (2006) NMR Structures of Damaged DNA. *Chem. Rev. 106*, 607-686.

(85) Cosman, M., Fiala, R., Hingerty, B. E., Laryea, A., Lee, H., Harvey, R. G., Amin, S., Geacintov, N. E., Broyde, S., and Patel, D. (1993) Solution Conformation of the (+)-trans-anti-[BPh]dA Adduct Opposite dT in a DNA Duplex: Intercalation of the Covalently Attached Benzo[c]phenanthrene to the 5'-side of the Adduct Site Without Disruption of the Modified Base Pair. *Biochemistry 32*, 12488-12497.

(86) Cai, Y., Geacintov, N. E., and Broyde, S. (2012) Nucleotide Excision Repair Efficiencies of Bulky Carcinogen–DNA Adducts Are Governed by a Balance between Stabilizing and Destabilizing Interactions. *Biochemistry 51*, 1486-1499.

(87) Cai, Y., Ding, S., Geacintov, N. E., and Broyde, S. (2011) Intercalative Conformations of the 14R (+)- and 14S (−)-trans-anti-DB[a,l]P-N6-dA Adducts: Molecular Modeling and MD Simulations. *Chem. Res. Toxicol. 24*, 522-531.

(88) Tang, Y., Liu, Z., Ding, S., Lin, C. H., Cai, Y., Rodriguez, F. A., Sayer, J. M., Jerina, D. M., Amin, S., Broyde, S., and Geacintov, N. E. (2012) Nuclear Magnetic Resonance Solution Structure of an N2-Guanine DNA Adduct Derived from the Potent Tumorigen Dibenzo[a,l]pyrene: Intercalation from the Minor Groove with Ruptured Watson–Crick Base Pairing. *Biochemistry 51*, 9751-9762.

(89) Wu, M., Yan, S., Patel, D. J., Geacintov, N. E., and Broyde, S. (2001) Cyclohexene Ring and Fjord Region Twist Inversion in Stereoisomeric DNA Adducts of Enantiomeric Benzo[c]phenanthrene Diol Epoxides. *Chem. Res. Toxicol. 14*, 1629-1642.

(90) Yan, S., Shapiro, R., Geacintov, N. E., and Broyde, S. (2001) Stereochemical, Structural, and Thermodynamic Origins of Stability Differences between Stereoisomeric Benzo[a]pyrene Diol Epoxide Deoxyadenosine Adducts in a DNA Mutational Hot Spot Sequence. *J. Am. Chem. Soc. 123*, 7054-7066.

(91) 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.

(92) Esposito, V., Randazzo, A., Piccialli, G., Petraccone, L., Giancola, C., and Mayol, L. (2004) Effects of an 8-bromodeoxyguanosine Incorporation on the Parallel Quadruplex Structure [d(TGGGT)]4. *Org. Biomol. Chem. 2*, 313-318.

(93) Dumas, A., and Luedtke, N. W. (2010) Cation-Mediated Energy Transfer in G-Quadruplexes Revealed by an Internal Fluorescent Probe. *J. Am. Chem. Soc. 132*, 18004-18007.

(94) Rutledge, L. R., Durst, H. F., and Wetmore, S. D. (2009) Evidence for Stabilization of DNA/RNA-Protein Complexes Arising from Nucleobase-Amino Acid Stacking and T-Shaped Interactions. *J. Chem. Theory Comput. 5*, 1400–1410.

(95) Pérez, A., Luque, F. J., and Orozco, M. (2011) Frontiers in Molecular Dynamics Simulations of DNA. *Acc. Chem. Res. 45*, 196-205.

(96) Hunter, C. A., and Lu, X.-J. (1997) DNA Base-Stacking Interactions: A Comparison of Theoretical Calculations with Oligonucleotide X-ray Crystal Structures. *J. Mol. Biol. 265*, 603-619.

(97) Snyder, R. D. (2007) Assessment of Atypical DNA Intercalating Agents in Biological and in silico Systems. *Mutat. Res.-Fund. Mol. M. 623*, 72-82.

(98) Hendry, L. B., Mahesh, V. B., Bransome Jr, E. D., and Ewing, D. E. (2007) Small Molecule Intercalation with Double Stranded DNA: Implications for Normal Gene Regulation and for Predicting the Biological Efficacy and Genotoxicity of Drugs and Other Chemicals. *Mutat. Res.-Fund. Mol. M. 623*, 53-71.

(99) Graves, D. E., and Velea, L. M. (2000) Intercalative Binding of Small Molecules to Nucleic Acids. *Curr. Org. Chem. 4*, 915-929.

(100) Kropachev, K., Kolbanovskiy, M., Liu, Z., Cai, Y., Zhang, L., Schwaid, A. G., Kolbanovskiy, A., Ding, S., Amin, S., Broyde, S., and Geacintov, N. E. (2013) Adenine–DNA Adducts Derived from the Highly Tumorigenic Dibenzo[a,l]pyrene Are Resistant to Nucleotide Excision Repair while Guanine Adducts Are Not. *Chem. Res. Toxicol. 26*, 783-793.

(101) Liang, F., and Cho, B. P. (2011) Conformational and Thermodynamic Impact of Bulky Aminofluorene Adduction on Simulated Translesion DNA Synthesis. *Chem. Res. Toxicol. 24*, 597-605.

(102) Cheng, X., Kelso, C., Hornak, V., de los Santos, C., Grollman, A. P., and Simmerling, C. (2005) Dynamic Behavior of DNA Base Pairs Containing 8-Oxoguanine. *J. Am. Chem. Soc. 127*, 13906-13918.

(103) Millen, A. L., Kamenz, B. L., Leavens, F. M. V., Manderville, R. A., and Wetmore, S. D. (2011) Conformational Flexibility of C8-Phenoxylguanine Adducts in Deoxydinucleoside Monophosphates. *J. Phys. Chem. B 115*, 12993-13002.

(104) Milhé, C., Fuchs, R. P. P., and Lefèvre, 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.

(105) Koehl, P., Valladier, P., Lefèvre, J.-F., and Fuchs, R. P. P. (1989) Strong Structural Effect of the Position of a Single Acetylaminofluorene Adduct Within a Mutation Hot Spot. *Nucleic Acids Res. 17*, 9531-9541.

(106) 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.

(107) de Laat, W. L., Jaspers, N. G. J., and Hoeijmakers, J. H. J. (1999) Molecular Mechanism of Nucleotide Excision Repair. *Genes Dev. 13*, 768-785.

Chapter 3: Conformational Flexibility of the Benzyl-guanine Adduct in a Bypass Polymerase Active Site Permits Replication: Insights from Molecular Dynamics Simulations[a,b,c]

3.1. Introduction

Many endogenous and exogenous substances, such as hormones, tobacco, processed meats, and by-products from the combustion of fossil fuels, damage our DNA through the formation of DNA addition products (adducts).[1-4] The DNA adducts identified to date involve different nucleobases (i.e., A, G, T, or C), sites of carcinogen attachment (e.g., N2, O6, or C8 of dG), carcinogen–nucleobase linker atoms (e.g., N, O, or C), and chemical composition of the carcinogen (e.g., the ring number, size, aromaticity, and substitution, as well as presence, length, conjugation, and substitution of the alkyl chain between the nucleobase and bulky moiety ring).[5, 6] To accompany this chemical diversity, DNA adducts exhibit a range of biological outcomes. For example, among the well-studied N-linked C8-dG adducts, the single-ringed N-linked C8-aniline-dG lesion lacks potent mutagenicity,[7] while N-linked C8-dG adducts with larger bulky moieties (e.g., aminofluorene (AF), 1-nitropyrene (AP), 3-nitrobenzanthrone (ABA), and 8-[(3-methyl-3H-imidazo[4,5-f]quinolin-2-yl)amino] (IQ)) exhibit potent, diverse mutagenic profiles, including base mispairs and deletion mutations.[8-10] As a result, these large lesions have been associated

---

[a] *Chemical Research in Toxicology* reference style used throughout this chapter

[b] Reprinted with permission from Wilson, K. A., and Wetmore, S. D. (2017). Conformational flexibility of the benzyl-guanine adduct in a bypass polymerase active site permits replication: Insights from molecular dynamics simulations. *Chem. Res. Toxicol.,* 30, 2013-2022.

[c] K.A.W. performed calculations, data analysis and wrote the first draft of the manuscript.

with various diseases, including bladder,[11] colorectal,[12] oral,[13] and pharyngeal cancer,[13] as well as inflammatory bowel disease.[14]

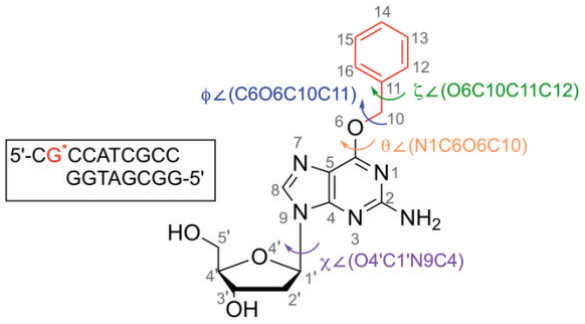

Figure 3.1. Structure of the Bz-dG nucleoside with the bulky moiety highlighted in red. Definitions of key dihedral angles and the sequence context considered in MD simulations are provided (with *anti* or *syn*-Bz-dG, or dG at the G* position).

The mutagenicity of DNA adducts arises from their effects on the process of DNA replication.[1-4] In general, the active sites of standard (replicative) DNA polymerases are unable to accommodate bulky nucleobase lesions and thus an alternative process known as translesion synthesis (TLS) is required to copy past DNA adducts.[15] TLS (bypass) polymerases are more flexible and have much less sterically constrained active sites than replicative polymerases, which allow these enzymes to replicate a wide variety of nucleobase lesions. As a consequence of this adaptability, TLS polymerases have been reported to have lower processivity, efficiency, and accuracy than replicative polymerases.[15] Nevertheless, since the first general information about TLS polymerases was only obtained ~ 15 years ago,[15, 16] there are currently many unknowns about this replication pathway.

Recent literature has shown that the TLS past a nucleobase damaged site can be divided into three general stages: 1) binding of the adduct in the bypass polymerase active site prior to dNTP binding (i.e., the pre-insertion complex), 2) insertion of a dNTP opposite the lesion (i.e., the insertion complex), and 3) replication of the 5′ base(s) with respect to the lesion (i.e., the extension complex(es)).[15, 17] Previous work has illustrated that bypass polymerases process damaged DNA through these steps less efficiently than natural DNA. Furthermore, research on adducted DNA replication has demonstrated that the mutagenic behavior and the inability to replicate damaged DNA may stem from problems at different stages in the TLS process. For example, the AF[18] and IQ[19] adducts can lead to insertion of the incorrect nucleotides, and the polycyclic aromatic hydrocarbon (PAH)[20] and acetylaminofluorene (AAF)[21] adducts can block dNTP insertion, while the 3-nitrobenzanthrone (3-NBA)[22] and AP[23] adducts can prevent extension. Therefore, gaining structural information along the entire TLS pathway is paramount for uncovering the function of this critical class of enzymes and understanding how DNA adducts are processed in cells.

The current study focuses on the TLS past an O6-dG adduct, namely O6-benzyl-guanine (Bz-dG, Figure 3.1). In general, O6-dG adducts are some of the most mutagenic and carcinogenic lesions currently identified in the literature to date.[24-26] Furthermore, O6-dG adducts are particularly intriguing due to their modified dG Watson-Crick hydrogen-bonding face (i.e., N1 is a hydrogen-bond acceptor rather than a hydrogen-bond donor). The Bz-dG adduct is of specific interest since it is a chemotherapeutic agent and related to adducts formed from known carcinogens, including tobacco.[27] Furthermore, Bz-dG has

been investigated experimentally as an informative, yet simple, model to gain an understanding of the replication of O6-dG lesions.[28-31]

Experimental kinetics studies have revealed that the replication of Bz-dG primarily results in successful dCTP insertion (i.e., ~ 70% of the replication products contain the Bz-dG:dC base pair).[28, 29, 32] However, dTTP[28, 29] and dATP[28, 29] misincorporation has also been reported, with the insertion frequency depending on the TLS polymerase considered (i.e., DNA polymerase IV (Dpo4) or REV1, as well as human polymerase η, ι, δ or κ), and the experimental approach implemented (i.e., steady state kinetics, pre-steady state kinetics, gel electrophoresis, or mass spectrometry). To complement these experimental studies, molecular dynamics (MD) simulations on DNA duplexes containing Bz-dG mismatches[33] and the corresponding DNA–Dpo4 insertion complexes[34] have provided atomic level structural explanations for the observed mutagenicity associated with Bz-dG replication. Furthermore, X-ray crystal structures of the Dpo4 extension complex have provided key structural data about the extension stage of the TLS process.[28] However, experimental steady state kinetic studies have also shown that a replication step prior to or during bond formation is significantly slower for Bz-dG than dG replication.[28] While it has been proposed that Bz-dG inhibits the nucleotidyl-transfer reaction,[28, 29] neither the pre-insertion complex nor the nucleotidyl-transfer reaction has been investigated for Bz-dG replication in order to rationalize this experimental observation. This lack of information arises at least in part due to the short-lived nature of the pre-insertion complex, as well as transition states along the reaction pathway, and thus challenges associated with crystallization of relevant DNA–polymerase complexes.

Since computational chemistry has been shown to be an important tool for gaining insight into the function of DNA polymerases (see, for example, 8, 35-48), and can readily provide key information about short-lived species, the present work uses computer calculations to gain molecular level details of the TLS stages prior to the bond formation reaction for the nonmutagenic insertion of dCTP opposite Bz-dG. Specifically, 100 ns production MD simulations were used to study the pre-insertion and insertion complexes for Dpo4 replication of Bz-dG at physiological salt concentrations. Since a previous computational study on Bz-dG adducted DNA suggests that the lesion may adopt either the *anti* or *syn* orientation in the double helix opposite dC,[33] both *anti* and *syn*-Bz-dG were bound within the Dpo4 active site to determine the impact of this potential conformational heterogeneity on the structure of the pre-insertion and insertion complexes. Furthermore, to provide a basis for evaluating the adducted DNA–polymerase complexes, the Dpo4 pre-insertion and insertion complexes associated with the replication of canonical dG were also modelled. Replication by Dpo4 is investigated since this is the most widely studied TLS polymerase (see, for example, 49-56) and an abundance of the experimental work on Bz-dG replication has been completed using Dpo4.[28] By combining our newly acquired structural data with that previously obtained from X-ray crystal structures of the Dpo4 extension complexes for Bz-dG replication,[28] indispensable structural details are now available at each critical stage of the Bz-dG TLS replication pathway. As a result, we demonstrate how a bypass polymerase can accommodate multiple orientations of a flexible DNA lesion at different stages during a single TLS replication cycle. Furthermore, although Bz-dG may also affect the nucleotidyl-transfer reaction, our structural insights suggest that a replication step prior to bond formation likely significantly affects the overall rate of lesion bypass, which

correlates with previous experimental observations.[28] This finding emphasizes the importance of studying each stage of the TLS process and therefore will likely encourage similar future studies for other DNA adducts.

3.2. Computational Details

All starting structures for molecular dynamics simulations were derived from a crystal structure of the Dpo4 ternary complex corresponding to dATP insertion opposite the 5′-dT with respect to a benzo[a]pyrene adenine adduct (PDB ID: 1S0M).[57] This crystal structure was chosen from other available Dpo4 ternary complexes due to the preserved substrate (i.e., the presence of a deoxyribose instead of dideoxyribose at the 3′-primer terminus and the presence of a canonical dNTP) and absence of mutations to the enzyme. However, Ca^{2+} ions were used to prevent the reaction and allow crystallization, which like any crystallization condition may affect the enzyme structure and/or may trap the system into a small part of the configurational space. Nevertheless, the coordination of the Ca^{2+} ions mimics that of the catalytically essential Mg^{2+} ions. Furthermore, the orientation of the 1S0M active site is conserved in 45 other crystal structures of Dpo4 ternary complexes obtained through a variety of crystallization conditions (see Figure B.1 and Appendix B for a list of the crystal structures considered). This comparison supports that we are not neglecting critical orientations of active site amino acids by focusing on one crystal structure.

The starting structures for the MD simulations were built from the 1S0M crystal structure by modifying the template DNA strand sequence to 5′–CG*CCATCGCC with dG or Bz-dG at the G* position. Pre-insertion complexes were generated for *anti* and *syn*-Bz-dG, and

anti-dG in the Dpo4 active site with no pairing dNTP, while insertion complexes include dCTP opposite the same nucleoside conformations. Two divalent ions (Ca405 and Ca408) were trapped during crystallization in a position that affords coordination to the phosphate backbone of the solvent exposed DNA, and therefore were removed since they do not interact with the polymerase and do not affect the active site conformation (Figure B.1, Appendix B). The two remaining crystallographic divalent ions (Ca403 and Ca404) possess active site positions and coordination consistent with previous research on TLS polymerases. Specifically, previous literature has established that the nucleotidyl-transfer reaction is dependent on two active-site divalent metal cations (typically Mg^{2+}), one of which assists dNTP binding (denoted the binding Mg^{2+} ion) through coordination to Asp7(Oδ2), Phe8(O), Asp105(Oδ2), dNTP(PαO1), dNTP(PβO1), and dNTP(PγO1), and the second of which assist catalysis (denoted the catalytic Mg^{2+} ion) through coordination to the primer terminus (O3′), Glu106(Oε1), Asp7(Oδ1), dNTP(PαO2), Asp105(Oδ1), and a water molecule (Figure B.2, Appendix B). Additionally, recent kinetic,[58-60] time-lapsed X-ray crystallography,[61-64] and computational studies[65-68] have concluded that the binding Mg^{2+} ion enters the active site with the dNTP, while the catalytic Mg^{2+} ion moves into the active site when the reaction coordinate is aligned. Therefore, Ca403 and Ca404 were removed for the pre-insertion complexes, and Ca403 and Ca404 were changed to catalytic Mg^{2+} ions for the insertion complexes in order to assess the structure of the complex directly prior to the nucleotidyl-transfer reaction.

The resulting complexes were prepared for simulation by adding hydrogen atoms to generate the natural protonation states of all DNA and protein residues. Furthermore, the systems were neutralized with Na^{+}, solvated with a periodic TIP3P water box such that the

sides are 10 Å from the solute, and buffered with 150 mM NaCl. All amino acids, undamaged nucleotides, ions (Na^+ and Cl^-), and water were modeled with the AMBER ff14SB force field. Bz-dG was described using parameters previously published by our group,[33] Mg^{2+} ions were described using the parameters from Allner *et al.*,[68] and dCTP parameters were adapted from the literature.[67] In the simulations, the SHAKE algorithm was used to constrain covalent bonds involving hydrogen and the particle mesh Ewald algorithm was used for long range electrostatic interactions, with a nonbonded cutoff of 8.0 Å.

Each DNA–Dpo4 complex was subjected to minimization, equilibration, and production steps using the pmemd module of Amber 14.[69] The first minimization phase involved 1000 steps of steepest decent minimization, followed by 3000 steps of conjugate gradient minimization, with a 500 kcal mol^{-1} $Å^{-2}$ force constraint on the protein and DNA. Next, the DNA was minimized using 1000 steps of steepest decent, followed by 3000 steps of conjugate gradient minimization, with a 500 kcal mol^{-1} $Å^{-2}$ force constraint on the protein. Finally, 1000 steps of steepest decent minimization, followed by 4000 steps of conjugate gradient minimization were performed on the entire system. Subsequently, the equilibration phase was completed using a Langevin thermostat ($\gamma = 1.0$) to heat the system from 0 to 300 K over 20 ps, with a 10 kcal mol^{-1} $Å^{-2}$ force constraint on the protein and DNA. To explore the active site conformational space, this process was repeated for several (up to 15) starting configurations. Since comparison of 46 crystal structures of Dpo4 ternary complexes suggests that the crystallization conditions do not significantly change the active site conformation, these starting configurations differed in the conformation and binding orientation of the template base (pre-insertion and insertion complexes) and pairing dCTP

(insertion complexes) in order to explore the template base alignment and lesion site hydrogen bonding. Furthermore, since previous work has indicated that Bz-dG is a highly flexible adduct as both an isolated nucleoside and within a DNA duplex, different conformations about key Bz-dG dihedral angles were considered. Based on each starting configuration, short (< 20 ns) production simulations were performed at 330 K. For each DNA–polymerase complex, a starting configuration was chosen for a final 100 ns unrestrained production simulation that best represents a catalytic complex based on previously determined criteria.[8, 35] The production simulation was performed at 330 K using a 2 fs time step.

Analysis of the simulations was performed with the cpptraj module of Amber 14.[69] Specifically, hydrogen-bonding occupancies are reported as the percentage of simulation time that the heavy atom distance is < 3.4 Å and the hydrogen-bonding angle is > 120°. Furthermore, Linear Interaction Energies were used to evaluate the strength of stacking (van der Waals component) and hydrogen-bonding (electrostatic component) interactions. Over the 100 ns production trajectories, all complexes showed a stable rmsd of the DNA and protein backbone atoms with respect to the first frame of the corresponding simulation (Table B.1, Appendix B).

3.3. Results

3.3.1. Replication of Canonical dG.

3.3.1.1. Pre-insertion Complex. In the pre-insertion complex (Figure 3.2a), dG adopts the *anti* orientation about the glycosidic bond (χ=236.1±22.0°) and the C2′-*endo* sugar pucker dominant in canonical DNA duplexes (Table B.1, Appendix B). The position of dG in the

Dpo4 active site is maintained through hydrogen bonds between the dG phosphate moiety, and Arg331(NωH) and Ser34(OγH) (96 and 20% occupancy, respectively; Figure B.3a, Appendix B). Although there are no hydrogen bonds between the dG nucleobase and Dpo4, and the interactions between dG and the 5′-intrastrand dC are insignificant (–0.5±0.5 kcal/mol), the position of the dG base is stabilized through stacking interactions with the 3′-intrastrand dC (–6.1±0.9 kcal/mol; Figure 3.2a and Table B.2, Appendix B). As expected, translocation of dG into the active site does not affect the canonical base pairing of 3′–dC:dG, which contains three hydrogen bonds (99–100% occupancy), a C1′–C1′ distance of 10.740±0.163 Å, and an interaction strength of –28.1±2.8 kcal/mol (Tables B.3, Appendix B). Furthermore, interactions between the phosphate backbone of 3′–dC:dG and Dpo4 help stabilize the pre-insertion complex. Specifically, the phosphate moiety of 3′–dC hydrogen bonds with Arg332(NωH), Arg332(NϵH), and Thr250(OγH) (> 99% occupancy), while the phosphate moiety of 3′–G hydrogen bonds with Lys152(NζH) (74% occupancy; Figure B.4a, Appendix B). Most importantly, the *anti* nucleoside conformation directs the dG Watson-Crick hydrogen-bond face towards a void in the Dpo4 active site (Figure 3.2a), which leaves space for dNTP binding opposite dG during the next stage of TLS.

3.3.1.2. Insertion Complex. In the dG:dCTP insertion complex (Figure 3.3a), both dG and dCTP adopt the *anti* glycosidic orientation (χ= 255.5±12.9 and 225.5±10.1°, respectively; Table B.1, Appendix B). However, while dG adopts the canonical B-DNA C2′-*endo* pucker, the dCTP adopts the C1′–*exo* sugar pucker. The incoming dCTP forms canonical Watson-Crick bonding with dG (with each hydrogen bond having > 99% occupancy), which results in a dG:dCTP pair with a C1′–C1′ distance of 10.773±0.138 Å, and an interaction strength of –30.7±2.7 kcal/mol (Figure 3.4). The position of dG is maintained

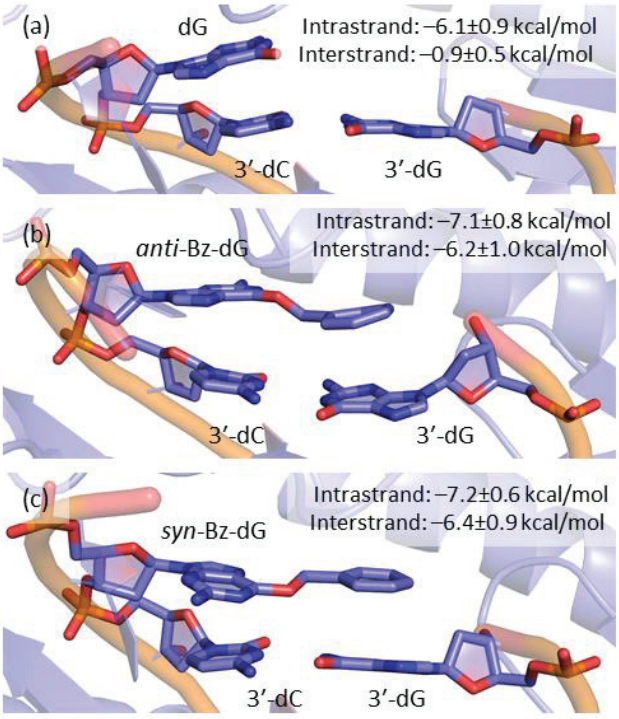

Figure 3.2. Representative MD structures and corresponding average (AMBER) stacking interactions between dG* and the 3′–base pair in the (a) dG, (b) *anti*-Bz-dG, and (c) *syn*-Bz-dG pre-insertion complexes for replication by Dpo4.

in the Dpo4 active site through hydrogen bonds between the dG phosphate moiety and Arg331(NωH) (69% occupancy; Figure 3.5, left). However, no interactions occur between Dpo4, and the nucleobases of dG or dCTP. Instead, the locations of the dG and dCTP bases are stabilized through interactions within DNA (Figure 3.3d). Specifically, significant stacking occurs between dG and the 3′–dC, and dCTP and the 3′–dG (~ –6 kcal/mol each; Table B.2, Appendix B). In contrast, minimal stacking occurs between dG, and the 5′–dC or 3′–dG, and dCTP and the 3′–dG (each weaker than –2 kcal/mol, Table B.2, Appendix B). The position of dCTP is further stabilized by many interactions with Dpo4 (Figure 3.5, right). Specifically, the γ-phosphate moiety hydrogen bonds with Arg51(NωH), Lys159(NζH), Tyr48(OH), and Tyr10(NH) (40–100% occupancy), while the β-phosphate

moiety hydrogen bonds with Thr45(OH) (67% occupancy) and Phe11(NH) (100% occupancy), and finally dNTP(O3′) hydrogen bonds with Tyr12(NH) (86% occupancy).

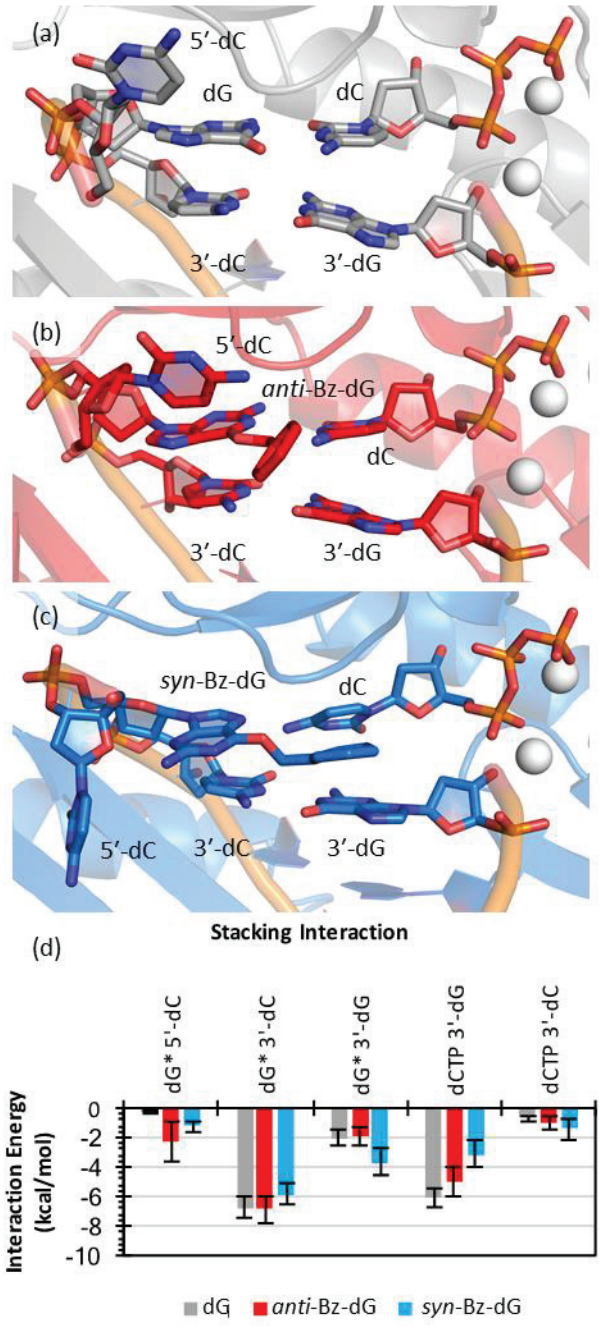

Figure 3.3. Representative MD structures of the (a) dG, (b) *anti*-Bz-dG, and (c) *syn*-Bz-dG insertion complexes for replication by Dpo4, and (d) the corresponding average (AMBER) stacking interactions between the dG*:dCTP and 3′–dC:dG pair.

As expected, the insertion of dCTP opposite dG does not affect the structure of the previously replicated 3′–dC:dG pair, which has a C1′–C1′ distance of 10.728±0.158 Å, three canonical Watson-Crick hydrogen bonds (> 99% occupancy), and an interaction energy of –28.9±2.6 kcal/mol (Table B.3, Appendix B). As discussed for the pre-insertion complexes, the location of the 3′–dC phosphate moiety is maintained through hydrogen bonds with Arg332(NωH), Arg332(NϵH), and Thr250(OγH) (> 95% occupancy), and the phosphate moiety of the 3′ interstand dG hydrogen bonds with Lys152(NζH) (53% occupancy; Figure 3.5). Although Na^+ ions in solution transiently coordinate with the solvent exposed phosphate moieties of the bound DNA throughout the simulation, no additional ions migrate into the active site.

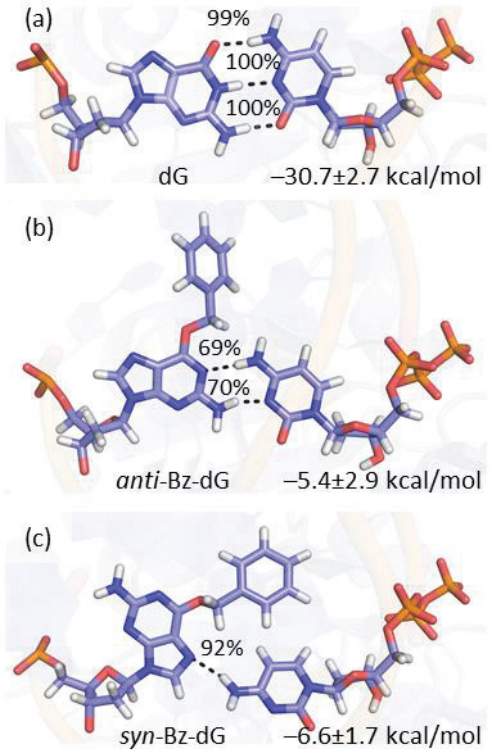

Figure 3.4. Representative MD structures and corresponding average (AMBER) hydrogen-bond strength between dG* and dCTP in the (a) dG, (b) *anti*-Bz-dG, and (c) *syn*-Bz-dG insertion complexes for replication by Dpo4.

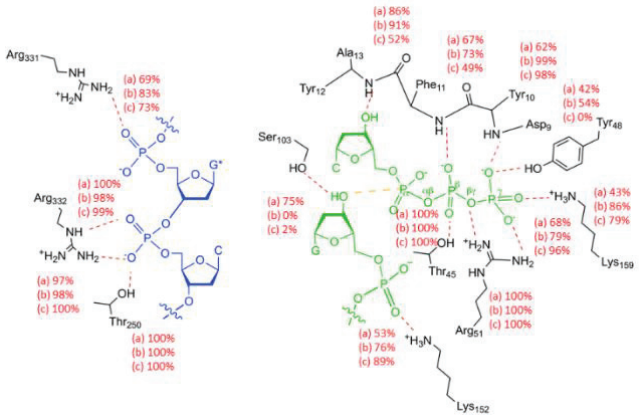

Figure 3.5. Occupancies of hydrogen bonds between the template strand (blue) or dCTP and primer strand (green), and the surrounding amino acid residues in the (a) dG, (b) *anti*-Bz-dG, and (c) *syn*-Bz-dG insertion complexes for replication by Dpo4.

Previous literature has identified several factors that must favorably align for the nucleotidyl-transfer reaction to occur. Specifically, the positions of the dNTP and terminal O3′ of the primer must be maintained through octahedral coordination of the two catalytic Mg^{2+} ions (Figure B.2, Appendix B).[42, 45, 70] Furthermore, the distance between the α-phosphate of the dNTP and the terminal O3′ of the primer strand must approach the van der Waals radii for O and P (3.5 Å), while X-ray and computational studies indicate that the in-line attack angle [∠O3′(primer 3′ end)–Pα(dNTP)–Oαβ(dNTP)] should range between 150 and 180°.[42, 45, 70] The dG:dCTP insertion complex studied in the present work contains the expected coordination of the Mg^{2+} ions (11 of the 12 coordination distances < 2.5 Å for > 90% of the simulation), average Pα–O3′ distance (3.529±0.158 Å), and in-line attack angle (168.8±5.8°; Table B.4, Appendix B).

3.3.2. Replication of Bz-dG.

3.3.2.1. Pre-insertion Complex. In the absence of an opposing dNTP, Bz-dG can readily adopt the *anti* (χ=246.4±15.1°) or *syn* (χ=42.0±15.7°) orientation about the glycosidic bond in the Dpo4 active site (Figure 3.2b), and the C2′-*endo* or C1′-*exo* sugar pucker,

respectively (Table B.1, Appendix B). Regardless of the glycosidic orientation, the lesion is bound in the active site through hydrogen bonds between the Bz-dG phosphate moiety, and Arg331(NωH) and Ser34(OγH) (> 83 and > 33% occupancy, respectively; Figure B.3b, Appendix B). However, no interactions occur between Dpo4 and the Bz-dG nucleobase or carcinogenic moiety. As a result, Bz-dG displays a high degree of conformational flexibility within the alkyl chain of the bulky moiety (θ=149.9±52.3 or 1.2±50.1°, for *anti or syn*-Bz-dG, respectively), as well as rotation of the phenyl ring (ξ=3.9±97.9 or 355.0±58.3°, for *anti or* syn-Bz-dG, respectively; Table B.1, Appendix B). However, there is little variation in φ (177.9±12.7 or 179.0±9.8°, for *anti or* syn-Bz-dG, respectively; Table B.1, Appendix B) such that the lesion adopts an extended conformation throughout the simulation (Figure 3.2b). The Bz-dG conformation within the Dpo4 active site consistently places the carcinogenetic moiety above the 3′-interstrand dG (Figures 3.2 and B.5, Appendix B). Nevertheless, the carcinogenic group of *anti*-Bz-dG is positioned towards N1 (θ≈180°), while that of *syn*-Bz-dG is positioned towards N7 (θ≈0°), which affords maximum overlap with the 3′–interstrand dG for each glycosidic orientation.

Regardless of the glycosidic orientation, this intercalated orientation leads to strong stacking interactions between the adduct, and the 3′–dC and 3′–dG (each ~ –7 kcal/mol), while a minimal interaction occurs between Bz-dG and the 5′–dC (~ –1 kcal/mol; Figure 3.2c and Table B.2, Appendix B). Despite interactions between the lesion and the previously replicated base pair, Bz-dG does not affect the canonical base pairing of 3′–dC:dG, which contains three hydrogen bonds (with > 97% occupancy), a C1′–C1′ distance of ~ 10.7 Å, and an interaction energy of ~ –28 kcal/mol (Tables B.3, Appendix B). Furthermore, the phosphate moiety of the 3′–dC hydrogen bonds with Arg332(NωH),

Arg332(NϵH) and Thr250(OγH) (> 92% occupancy), and the phosphate moiety of the 3′–dG hydrogen bonds with Lys152(NζH) (> 76% occupancy, Figure B.4, Appendix B).

3.3.2.2. Insertion Complex. When dCTP (χ=232.3±16.2° for *anti*-Bz-dG or 255.6±14.8° for *syn*-Bz-dG) is placed opposite *anti* (χ=267.2±13.7°) or *syn* (χ=29.3±14.8°) Bz-dG in the Dpo4 active site (Figures 3.3b and c), dCTP and Bz-dG adopt the C1′–*exo* or C2′–*endo* sugar pucker (Table B.1, Appendix B). While *anti*-Bz-dG forms a wobble base pair with the incoming dCTP that contains two hydrogen bonds, namely Bz-dG(N1)···dCTP(N4H) and Bz-dG(N2H)···dCTP(N3) (69 and 70% occupancy, respectively; Figure 3.4b), *syn*-Bz-dG forms a single dG(N7)···dCTP(N4H) hydrogen bond (92% occupancy; Figure 3.4c). The Bz-dG:dCTP hydrogen-bonding interaction is –5.4±2.9 or –6.6±1.7 kcal/mol for *anti* or *syn*-Bz-dG, respectively.

No interactions occur between the carcinogenic moiety, or the Bz-dG or dCTP nucleobases, and Dpo4. The lack of Dpo4 interactions with Bz-dG allows for significant rotation about key dihedral angles in the carcinogenic moiety. As a result, there is variation in the conformation of the alkyl chain (standard deviation in θ up to ~ 116°) and the position of the phenyl ring (standard deviation in ξ up to ~ 93°; Table B.1, Appendix B). However, Bz-dG remains in an extended conformation (φ ~ 180 with a standard deviation < 10°; Table B.1, Appendix B), with the bulky moiety solvent exposed in the major groove (Figure 3.3b and c).

The locations of the bases in the newly forming Bz-dG:dCTP pair are stabilized by stacking between the adduct and the 3′–dC (~ –6.8±0.9 or –5.9±0.7 kcal/mol for *anti* or *syn*-Bz-dG, respectively), and dCTP and the 3′–dG (–5.0±1.0 or –3.1±0.9 kcal/mol for *anti* or *syn*-Bz-dG, respectively). Additionally, regardless of the glycosidic orientation, weak interactions occur between Bz-dG, and the 5′–dC or 3′–dG, and dCTP and the 3′–dC (each < ~ –2 kcal/mol; Figure 3.3d and Table B.3, Appendix B).

Bz-dG is bound in the active site through a hydrogen bond between the phosphate moiety and Arg331(NωH) (> 73% occupancy; Figure 3.5, left). Furthermore, regardless of the Bz-dG glycosidic bond orientation, the position of the incoming dCTP is maintained through interactions between the dCTP backbone and Dpo4 (Figure 3.5, right). Specifically, the γ-phosphate moiety hydrogen bonds with Arg51(NωH), Lys159(NζH), and Tyr10(NH) (> 80% occupancy), while Tyr48(OH) hydrogen bonds with the γ-phosphate moiety (> 42% occupancy), and the β-phosphate moiety hydrogen bonds with Thr45(OH), and Phe11(NH) (> 49% or 100% occupancy, respectively). A hydrogen bond also exists between dNTP(O3′) and Tyr12(NH) (> 52% occupancy).

The 3′–dG:dC pair retains three canonical Watson-Crick hydrogen bonds (> 92% occupancy), an average C1′–C1′ distance of ~ 10.6–10.7 Å, and an average interaction energy of ~ –27 to – 29 kcal/mol (Table B.3, Appendix B). The location of the 3′–dC phosphate moiety is maintained through hydrogen bonds with Arg332 and Thr250 (> 95% occupancy), while the phosphate moiety of the 3′–interstand dG hydrogen bonds with Lys152 (> 76% occupancy, Figure B.4c, Appendix B). Furthermore, Na^+ ions in the

solution coordinate with the solvent exposed phosphate moieties of the DNA helix, but no additional ions migrate into the active site. The coordination of the active site Mg^{2+} ions is catalytic, with > 90% occupancy of a distance of < 2.5 Å for 11 of the 12 atoms coordinated to the Mg^{2+} ions (Table B.4, Appendix B). Additionally, the Pα–O3′ distance (~ 3.4 – 3.5 Å) and the in-line attack angle (~ 168–169°; Table B.4, Appendix B) are reaction ready.

3.4. Discussion

The present work complements previous MD structural information regarding the mutagenic bypass of Bz-dG by Dpo4[34] and X-ray structural data for the associated extension complex.[28] Indeed, when our newly acquired structural information is combined with previously published experimental X-ray structures of the dG and Bz-dG extension complexes,[28] molecular insights at each stage during a single TLS replication cycle are now available for dG and Bz-dG. As a result, our work sheds light on the experimentally observed nonmutagenic replication by Dpo4 for these bases, as well as the reduced dCTP binding affinity and insertion efficiency for Bz-dG.

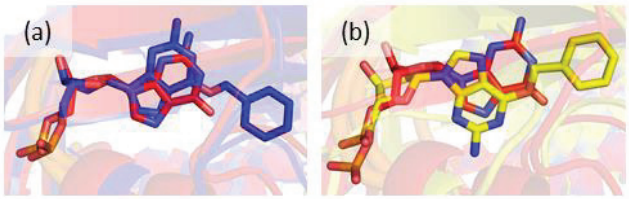

Figure 3.6. Overlay of representative MD structures of the (a) dG (red) and *anti*-Bz-dG (blue), and (b) dG (red) and *syn*-Bz-dG (yellow) pre-insertion complexes for replication by Dpo4.

During the first stage of TLS, the template nucleotide is positioned in the bypass polymerase active site prior to dNTP binding. Comparison of the Dpo4 pre-insertion

complexes for dG and (*anti* or *syn*) Bz-dG reveals that the template nucleotide is consistently held in the active site through contacts between the phosphate backbone and Dpo4. As a result, canonical and adducted dG adopt similar positions in the active site (Figure 3.6). The consistent orientation of canonical and adducted dG in Dpo4 pre-insertion complexes has previously been reported for the BP-dG adduct,[39] which provides further support that DNA lesions are recognized in the active site in the same manner as their natural counterparts. Although strong interactions between the template nucleotide and the 5′-base have been proposed to impede translocation of the TLS polymerase,[71] no such contacts are observed in the dG or Bz-dG pre-insertion complexes with Dpo4. Furthermore, although highly stabilizing (hydrogen-bonding or stacking) interactions between the bulky moiety and bypass polymerases have been proposed to stall replication for other adducts,[35, 72] no interactions between the bulky moiety and Dpo4 are observed for either Bz-dG glycosidic orientation. Thus, these global geometric features of the pre-insertion complexes are consistent with experimental data reporting bypass of dG and Bz-dG by Dpo4.[28]

The orientation of dG in the Dpo4 pre-insertion complex directs the Watson-Crick hydrogen-bonding face toward a void in the active site (Figures 3.2a), which is the positon the incoming dNTP must bind to yield the corresponding insertion complex (Figures 3.3a). Since the dG and Bz-dG nucleotides are positioned similarly in the Dpo4 pre-insertion complex, the Watson-Crick hydrogen-bonding face of *anti*-Bz-dG adopts the same orientation as that of dG (Figure 3.6a). In contrast, the change in the glycosidic orientation results in the Hoogsteen binding face of *syn*-Bz-dG being directed towards the dNTP

binding pocket. However, for both Bz-dG conformations, the bulky moiety is positioned over the 3′-flanking base pair in the dNTP binding site (Figures 3.2b and 2c). This orientation is stabilized by stacking interactions between the 3′-flanking pair and the template base, which are stronger for Bz-dG than canonical dG (by up to ~ 7 kcal/mol total, Table B.3, Appendix B). Since this conformation orients the bulky moiety in the position the incoming dNTP must bind regardless of the Bz-dG glycosidic orientation, the lesion must likely undergo a conformational change to permit dNTP binding opposite the lesion in a catalytically conducive orientation. This conformational change would at least in part contribute to the observed weaker binding of dCTP to the Bz-dG replication complex compared to the corresponding dG complex.[28] Interestingly, although the BP-dG lesion that leads to deletion mutations has been reported to be positioned similar to dG in the pre-insertion complex, the BP-dG bulky moiety is directed into the major groove pocket in the active site in the pre-insertion complex,[39] reflecting that the chemical structure of the lesion may affect early steps in the TLS process and thereby the replication outcome.

Despite the additional stability the stacking interactions between Bz-dG and the 3′–flanking base pair provide to the pre-insertion complex, the strength of these contacts are significantly less than other biologically-relevant interactions (i.e., hydrogen bonding in a canonical dG:dC Watson-Crick pair). Furthermore, although the bulky moiety maintains the position above the 3′–flanking pair in the pre-insertion complex, Bz-dG exhibits significant conformational flexibility in the Dpo4 active site, as previously reported for DNA nucleoside and duplex models.[33, 34] These factors suggest that Bz-dG could undergo a conformational change to afford dCTP binding. Although coordinating the

conformational change of Bz-dG with insertion of the correct dNTP is anticipated to be challenging, experimental studies have unanimously reported successful dCTP insertion during TLS by Dpo4.[28, 31]

To gain further insight into the TLS replication cycle for Bz-dG, Dpo4 insertion complexes were considered with dCTP inserted opposite dG or (*anti* or *syn*) Bz-dG. Successful dCTP binding is achieved through positioning the bulky moiety in the major groove of the growing DNA duplex. Nevertheless, as found in the present work for the Dpo4 pre-insertion complex and previous studies on DNA duplexes,[33] the bulky moiety exhibits flexibility in this position in the Dpo4 insertion complexes.

Although stable Watson-Crick hydrogen bonding occurs between a template dG and dCTP, less stable Bz-dG:dCTP hydrogen-bonding arises due to a reduction in the number, as well as occupancy, of the hydrogen bonds upon addition of the carcinogenic moiety to dG. Nevertheless, the wobble base pair for *anti*-Bz-dG and the single hydrogen bond for *syn*-Bz-dG are similar to the configurations of Bz-dG:dC base pairs reported for adducted DNA duplexes.[33] Since previous research has suggested that hydrogen bonding between the base being replicated and the incoming dNTP is an essential component of TLS,[31, 73, 74] the weaker interactions for Bz-dG may hinder dCTP insertion, and thereby contribute to the observed decreased dCTP binding affinity and dCTP insertion efficiency opposite Bz-dG compared to dG.[28]

Most importantly, within the Dpo4 insertion complexes, all DNA–protein interactions and DNA–DNA stacking interactions are similar, the catalytic Mg^{2+} ions are fully coordinated

in the active site, and the reaction coordinate is aligned for dCTP insertion regardless of whether the template base is canonical dG or damaged Bz-dG. Therefore, despite the bulky moiety being in the position of the incoming dNTP in the corresponding pre-insertion complex, dCTP can be inserted opposite dG and Bz-dG, and our results support the experimentally observed nonmutagenic bypass of dG and Bz-dG by Dpo4.[28] This contrasts results from a previous computational study of the AAF-dG lesion, which maintains Watson-Crick hydrogen bonding, but disrupts Mg^{2+} coordination and dNTP–protein interactions,[8] and therefore blocks replication bypass.[21]

Our newly characterized structural features of the Dpo4 pre-insertion and insertion complexes for Bz-dG replication illustrate the ability of the lesion to adopt, and Dpo4 to accommodate, different template nucleotide conformations during the various replication stages in order to facilitate bypass. This can be clearly seen when representative MD structures of the pre-insertion and insertion complexes are compared to X-ray crystal structures of dG and Bz-dG extension complexes (Figure 3.7). Specifically, for dG replication, the template base and the protein backbone maintain the same orientation across the three key TLS stages, namely pre-insertion, insertion and extension (Figure 3.7a). In contrast, the Bz-dG bulky moiety is positioned in the dNTP binding pocket in the pre-insertion complex, but maintains a similar major groove location in the insertion and extension complexes (Figure 3.7b). Interestingly, the orientation of Bz-dG in the extension complex is also consistent with the conformation reported in fully formed DNA duplexes,[33] which indicates that little conformational change will occur once dCTP has been inserted opposite the lesion.

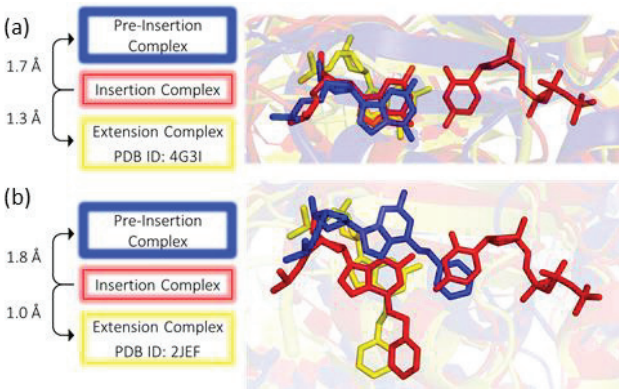

Figure 3.7. Overlay of structures and respective protein backbone RMSD from different TLS stages for a complete replication cycle, including a representative MD structure of the pre-insertion complex (blue), a representative MD structure of the dCTP insertion complex (red), and a crystal structure for the extension complex (yellow) for (a) dG or (b) *anti*-Bz-dG replication by Dpo4.

Overall, by characterizing structures along a single cycle of Bz-dG replication by Dpo4, the present work reveals a conformational change is required prior to or in conjunction with dNTP binding for Bz-dG replication that is not required for dG replication. This structural data is fully consistent with experimental kinetic evidence that a Bz-dG replication step prior to or during bond formation is significantly slower than for dG replication.[28] Thus, although future work must consider the reaction mechanism for the insertion of dCTP opposite Bz-dG to evaluate the effects of the adduct on the nucleotidyl-transfer reaction, the present study sheds light on the effects of Bz-dG on the early stages of TLS and emphasizes the importance of obtaining structural insight along the entire TLS process for other DNA lesions.

3.5. References

(1) Geacintov, N. E., Cosman, M., Hingerty, B. E., Amin, S., Broyde, S., and Patel, D. J. (1997) NMR solution structures of stereoisomeric polycyclic aromatic carcinogen-DNA adducts: Principles, patterns, and diversity. *Chem. Res. Toxicol. 10*, 111-146.

(2) Hingerty, B. E., and Broyde, S. (1990) Atomic resolution structures of DNA and DNA modified by carcinogens. *Int. J. High Perform. Comput. Appl. 4*, 11-21.

(3) Gates, K. S. (2009) An overview of chemical processes that damage cellular DNA: Spontaneous hydrolysis, alkylation, and reactions with radicals. *Chem. Res. Toxicol. 22*, 1747-1760.

(4) 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.

(5) Wilson, K., Kung, R., and Wetmore, S. (2016) Toxicology of DNA adducts formed upon human exposure to carcinogens: Insights gained from molecular modeling. *Advances in Molecular Toxicology 10*, 293.

(6) Stone, M. P., Huang, H., Brown, K. L., and Shanmugam, G. (2011) Chemistry and structural biology of DNA damage and biological consequences. *Chem. Biodivers. 8*, 1571-1615.

(7) 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.

(8) Wang, L., and Broyde, S. (2006) A new anti conformation for N-(deoxyguanosin-8-yl)-2-acetylaminofluorene (AAF-dG) allows watson–crick pairing in the sulfolobus solfataricus P2 DNA polymerase IV (Dpo4). *Nucleic Acids Res. 34*, 785-795.

(9) Gu, Z., Gorin, A., Krishnasamy, R., Hingerty, B. E., Basu, A. K., Broyde, S., and Patel, D. J. (1999) Solution structure of the N-(deoxyguanosin-8-yl)-1-aminopyrene ([AP]dG) adduct opposite dA in a DNA duplex. *Biochemistry 38*, 10843-10854.

(10) Mao, B., Vyas, R. R., Hingerty, B. E., Broyde, S., Basu, A. K., and Patel, D. J. (1996) Solution conformation of the N-(deoxyguanosin-8-yl)-1-aminopyrene ([AP]dG) adduct opposite dc in a DNA duplex. *Biochemistry 35*, 12659-12670.

(11) Besaratinia, A., and Tommasi, S. (2013) Genotoxicity of tobacco smoke-derived aromatic amines and bladder cancer: Current state of knowledge and future research directions. *The FASEB Journal 27*, 2090-2100.

(12) Rahman, U., Sahar, A., Khan, M. I., and Nadeem, M. (2014) Production of heterocyclic aromatic amines in meat: Chemistry, health risks and inhibition. A review. *LWT-Food Sci. Technol. 59*, 229-233.

(13) Pabiszczak, M., Banaszewski, J., Szmeja, Z., Szyfter, K., and Szyfter, W. (2000) Comparison of DNA adducts between oral, pharyngeal and larynx cancer. *Otolaryngol. Pol. 55*, 551-554.

(14) Li, L., Chan, R., Lu, L., Shen, J., Zhang, L., Wu, W., Wang, L., Hu, T., Li, M., and Cho, C. (2014) Cigarette smoking and gastrointestinal diseases: The causal relationship and underlying molecular mechanisms. *Int. J. Mol. Med. 34*, 372-380.

(15) Sale, J. E., Lehmann, A. R., and Woodgate, R. (2012) Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat. Rev. Mol. Cell Biol. 13*, 141-152.

(16) Masutani, C., Kusumoto, R., Yamada, A., Dohmae, N., Yokoi, M., Yuasa, M., Araki, M., Iwai, S., Takio, K., and Hanaoka, F. (1999) The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase η. *Nature 399*, 700-704.

(17) Liu, B., Xue, Q., Tang, Y., Cao, J., Guengerich, F. P., and Zhang, H. (2016) Mechanisms of mutagenesis: DNA replication in the presence of DNA damage. *Mutat. Res. Rev. Mutat. Res. 768*, 53-67.

(18) Tebbs, R. S., and Romano, L. J. (1994) Mutagenesis at a site-specifically modified NarI sequence by acetylated and deacetylated aminofluorene adducts. *Biochemistry 33*, 8998-9006.

(19) Bose, A., Millsap, A. D., DeLeon, A., Rizzo, C. J., and Basu, A. K. (2016) Translesion synthesis of the N2-2′-deoxyguanosine adduct of the dietary mutagen iq in human cells: Error-free replication by DNA polymerase κ and mutagenic bypass by DNA polymerases η, ζ, and Rev1. *Chem. Res. Toxicol. 29*, 1549-1559.

(20) Zang, H., Chowdhury, G., Angel, K. C., Harris, T. M., and Guengerich, F. P. (2006) Translesion synthesis across polycyclic aromatic hydrocarbon diol epoxide adducts of deoxyadenosine by sulfolobus solfataricus DNA polymerase Dpo4. *Chem. Res. Toxicol. 19*, 859-867.

(21) Vrtis, K. B., Markiewicz, R. P., Romano, L. J., and Rueda, D. (2013) Carcinogenic adducts induce distinct DNA polymerase binding orientations. *Nucleic Acids Res. 41*, 7843-7853.

(22) Gadkari, V. V., Tokarsky, E. J., Malik, C. K., Basu, A. K., and Suo, Z. (2014) Mechanistic investigation of the bypass of a bulky aromatic DNA adduct catalyzed by a Y-family DNA polymerase. *DNA Repair 21*, 65-77.

(23) Sherrer, S. M., Brown, J. A., Pack, L. R., Jasti, V. P., Fowler, J. D., Basu, A. K., and Suo, Z. (2009) Mechanistic studies of the bypass of a bulky single-base lesion catalyzed by a Y-family DNA polymerase. *J. Biol. Chem. 284*, 6379-6388.

(24) Loveless, A. (1969) Possible relevance of O-6 alkylation of deoxyguanosine to the mutagenicity and carcinogenicity of nitrosamines and nitrosamides. *Nature 223*, 206-207.

(25) Newbold, R., Warren, W., Medcalf, A., and Amos, J. (1980) Mutagenicity of carcinogenic methylating agents is associated with a specific DNA modification. *Nature 283*, 596 - 599.

(26) Singer, B., Chavez, F., Goodman, M., Essigmann, J., and Dosanjh, M. (1989) Effect of 3' flanking neighbors on kinetics of pairing of dCTP or dTTP opposite O6-methylguanine in a defined primed oligonucleotide when escherichia coli DNA polymerase I is used. *Proc. Natl. Acad. Sci. U.S.A. 86*, 8271-8274.

(27) Fu, D., Calvo, J. A., and Samson, L. D. (2012) Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nat Rev Cancer 12*, 104-120.

(28) Eoff, R. L., Angel, K. C., Egli, M., and Guengerich, F. P. (2007) Molecular basis of selectivity of nucleoside triphosphate incorporation opposite O6-benzylguanine by sulfolobus solfataricus DNA polymerase Dpo4: Steady-state and pre-steady-state kinetics and X-ray crystallography of correct and incorrect pairing. *J. Biol. Chem. 282*, 13573-13584.

(29) Choi, J.-Y., Chowdhury, G., Zang, H., Angel, K. C., Vu, C. C., Peterson, L. A., and Guengerich, F. P. (2006) Translesion synthesis across O6-alkylguanine DNA adducts by recombinant human DNA polymerases. *J. Biol. Chem. 281*, 38244-38256.

(30) Choi, J.-Y., and Guengerich, F. P. (2006) Kinetic evidence for inefficient and error-prone bypass across bulky N2-guanine DNA adducts by human DNA polymerase ι. *J. Biol. Chem. 281*, 12315-12324.

(31) Gahlon, H. L., Boby, M. L., and Sturla, S. J. (2014) O6-alkylguanine postlesion DNA synthesis is correct with the right complement of hydrogen bonding. *ACS Chem. Biol. 9*, 2807-2814.

(32) Choi, J.-Y., and Guengerich, F. P. (2008) Kinetic analysis of translesion synthesis opposite bulky N(2)- and O(6)-alkylguanine DNA adducts by human DNA polymerase rev1. *J.Biol. Chem. 283*, 23645-23655.

(33) Wilson, K. A., and Wetmore, S. D. (2014) Complex conformational heterogeneity of the highly flexible O6-benzyl-guanine DNA adduct. *Chem. Res. Toxicol. 27*, 1310-1325.

(34) Wilson, K. A., and Wetmore, S. D. (2017) Molecular insights into the translesion synthesis of benzyl-guanine from molecular dynamics simulations: Structural evidence of mutagenic and nonmutagenic replication. *Biochemistry 56*, 1841-1853.

(35) Zhang, L., Rechkoblit, O., Wang, L., Patel, D. J., Shapiro, R., and Broyde, S. (2006) 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. 34*, 3326-3337.

(36) Chandani, S., and Loechler, E. L. (2007) Molecular modeling benzo[a]pyrene N2-dG adducts in the two overlapping active sites of the Y-family DNA polymerase Dpo4. *J. Mol. Graph. Model. 25*, 658-670.

(37) Perlow-Poehnelt, R. A., Likhterov, I., Scicchitano, D. A., Geacintov, N. E., and Broyde, S. (2004) The spacious active site of a Y-Family DNA polymerase facilitates promiscuous nucleotide incorporation opposite a bulky carcinogen-DNA adduct: Elucidating the structure-function relationship through experimental and computational approaches. *J.Biol. Chem. 279*, 36951-36961.

(38) Xu, P., Oum, L., Geacintov, N. E., and Broyde, S. (2008) Nucleotide selectivity opposite a benzo[a]pyrene-derived N2-dG adduct in a Y-Family DNA polymerase: A 5′-slippage mechanism. *Biochemistry 47*, 2701-2709.

(39) Xu, P., Oum, L., Lee, Y.-C., Geacintov, N. E., and Broyde, S. (2009) Visualizing sequence-governed nucleotide selectivities and mutagenic consequences through a replicative cycle: Processing of a bulky carcinogen N2-dG lesion in a Y-Family DNA polymerase. *Biochemistry 48*, 4677-4690.

(40) Perlow-Poehnelt, R. A., Likhterov, I., Wang, L., Scicchitano, D. A., Geacintov, N. E., and Broyde, S. (2007) Increased flexibility enhances misincorporation: Temperature effects on nucleotide incorporation opposite A bulky carcinogen-DNA adduct by A y-family DNA polymerase. *J.Biol. Chem. 282*, 1397-1408.

(41) Wang, L., Wu, M., Yan, S. F., Patel, D. J., Geacintov, N. E., and Broyde, S. (2005) Accommodation of a 1s-(−)-benzo[c]phenanthrenyl-N6-dA adduct in the Y-Family Dpo4 DNA polymerase active site: Structural insights through molecular dynamics simulations. *Chem. Res. Toxicol. 18*, 441-456.

(42) Wang, L., Yu, X., Hu, P., Broyde, S., and Zhang, Y. (2007) A water-mediated and substrate-assisted catalytic mechanism for sulfolobus solfataricus DNA polymerase IV. *J. Am. Chem. Soc. 129*, 4731-4737.

(43) Prasad, B. R., Plotnikov, N. V., and Warshel, A. (2013) Addressing open questions about phosphate hydrolysis pathways by careful free energy mapping. *The Journal of Physical Chemistry B 117*, 153-163.

(44) Chu, X., Liu, F., Maxwell, B. A., Wang, Y., Suo, Z., Wang, H., Han, W., and Wang, J. (2014) Dynamic conformational change regulates the protein-DNA recognition: An investigation on binding of a Y-Family polymerase to its target DNA. *PLOS Comput. Biol. 10*, e1003804.

(45) Wang, Y., and Schlick, T. (2008) Quantum mechanics/molecular mechanics investigation of the chemical reaction in Dpo4 reveals water-dependent pathways and requirements for active site reorganization. *J. Am. Chem. Soc. 130*, 13240-13250.

(46) Lior-Hoffmann, L., Wang, L., Wang, S., Geacintov, N. E., Broyde, S., and Zhang, Y. (2012) Preferred wmsa catalytic mechanism of the nucleotidyl transfer reaction in human DNA polymerase κ elucidates error-free bypass of a bulky DNA lesion. *Nucleic Acids Res. 40*, 9193-9205.

(47) Lior-Hoffmann, L., Ding, S., Geacintov, N. E., Zhang, Y., and Broyde, S. (2014) Structural and dynamic characterization of polymerase κ's minor groove lesion processing reveals how adduct topology impacts fidelity. *Biochemistry 53*, 5683-5691.

(48) Jia, L., Geacintov, N. E., and Broyde, S. (2008) The N-clasp of human DNA polymerase κ promotes blockage or error-free bypass of adenine- or guanine-benzo[a]pyrenyl lesions. *Nucleic Acids Res. 36*, 6571-6584.

(49) Gahlon, H. L., Schweizer, W. B., and Sturla, S. J. (2013) Tolerance of base pair size and shape in postlesion DNA synthesis. *J. Am. Chem. Soc. 135*, 6384-6387.

(50) Sholder, G., and Loechler, E. L. (2015) A method to accurately quantitate intensities of 32P-DNA bands when multiple bands appear in a single lane of a gel is used to study dNTP insertion opposite a benzo[a]pyrene-dG adduct by sulfolobus DNA polymerases Dpo4 and dbh. *DNA Repair 25*, 97-103.

(51) 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.

(52) 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.

(53) 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: 19F NMR, surface plasmon resonance, and primer kinetic studies. *Biochemistry 53*, 4059-4071.

(54) Zhang, H., Eoff, R. L., Kozekov, I. D., Rizzo, C. J., Egli, M., and Guengerich, F. P. (2009) Versatility of Y-family sulfolobus solfataricus DNA polymerase Dpo4 in translesion synthesis past bulky N2-alkylguanine adducts. *J. Biol. Chem. 284*, 3563-3576.

(55) Walsh, J. M., Ippoliti, P. J., Ronayne, E. A., Rozners, E., and Beuning, P. J. (2013) Discrimination against major groove adducts by Y-family polymerases of the dinb subfamily. *DNA Repair 12*, 713-722.

(56) Kottur, J., Sharma, A., Gore, Kiran R., Narayanan, N., Samanta, B., Pradeepkumar, P. I., and Nair, Deepak T. (2015) Unique structural features in DNA polymerase IV enable efficient bypass of the N2 adduct induced by the nitrofurazone antibiotic. *Structure 23*, 56-67.

(57) Ling, H., Sayer, J. M., Plosky, B. S., Yagi, H., Boudsocq, F., Woodgate, R., Jerina, D. M., and Yang, W. (2004) Crystal structure of a benzo[a]pyrene diol epoxide adduct in a ternary complex with a DNA polymerase. *Proc. Natl. Acad. Sci. U.S.A. 101*, 2265-2269.

(58) Bakhtina, M., Lee, S., Wang, Y., Dunlap, C., Lamarche, B., and Tsai, M.-D. (2005) Use of viscogens, dNTPαS, and rhodium(III) as probes in stopped-flow experiments to obtain new evidence for the mechanism of catalysis by DNA polymerase β. *Biochemistry 44*, 5177-5187.

(59) Bakhtina, M., Roettger, M. P., Kumar, S., and Tsai, M.-D. (2007) A unified kinetic mechanism applicable to multiple DNA polymerases. *Biochemistry 46*, 5463-5472.

(60) Mendieta, J., Cases-González, C. E., Matamoros, T., Ramírez, G., and Menéndez-Arias, L. (2008) A mg2+-induced conformational switch rendering a competent DNA polymerase catalytic complex. *Proteins 71*, 565-574.

(61) Freudenthal, B. D., Beard, W. A., and Wilson, S. H. (2015) New structural snapshots provide molecular insights into the mechanism of high fidelity DNA synthesis. *DNA Repair 32*, 3-9.

(62) Batra, V. K., Beard, W. A., Shock, D. D., Krahn, J. M., Pedersen, L. C., and Wilson, S. H. (2006) Magnesium-induced assembly of a complete DNA polymerase catalytic complex. *Structure 14*, 757-766.

(63) Garcia-Diaz, M., Bebenek, K., Krahn, J. M., Pedersen, L. C., and Kunkel, T. A. (2007) Role of the catalytic metal during polymerization by DNA polymerase lambda. *DNA Repair 6*, 1333-1340.

(64) Yang, W., Weng, P. J., and Gao, Y. (2016) A new paradigm of DNA synthesis: Three-metal-ion catalysis. *Cell & Bioscience 6*, 51.

(65) Vidossich, P., and Magistrato, A. (2014) QM/MM molecular dynamics studies of metal binding proteins. *Biomolecules 4*, 616.

(66) Li, Y., Freudenthal, B. D., Beard, W. A., Wilson, S. H., and Schlick, T. (2014) Optimal and variant metal-ion routes in DNA polymerase β's conformational pathways. *J. Am. Chem. Soc. 136*, 3630-3639.

(67) Perlow, R. A., and Broyde, S. (2002) Toward understanding the mutagenicity of an environmental carcinogen: Structural insights into nucleotide incorporation preferences. *J. Mol. Biol. 322*, 291-309.

(68) Allnér, O., Nilsson, L., and Villa, A. (2012) Magnesium ion–water coordination and exchange in biomolecular simulations. *J. Chem. Theory Comput. 8*, 1493-1502.

(69) Case, D. A., Babin, V., Berryman, J. T., Betz, R. M., Cai, Q., Cerutti, D. S., T.E. Cheatham, I., 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.

(70) Brautigam, C. A., and Steitz, T. A. (1998) Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Curr. Opin. Struct. Biol. 8*, 54-63.

(71) Ling, H., Boudsocq, F., Woodgate, R., and Yang, W. (2001) Crystal structure of a Y-Family DNA polymerase in action: A mechanism for error-prone and lesion-bypass replication. *Cell 107*, 91-102.

(72) Sassa, A., Niimi, N., Fujimoto, H., Katafuchi, A., Grúz, P., Yasui, M., Gupta, R. C., Johnson, F., Ohta, T., and Nohmi, T. (2011) Phenylalanine 171 is a molecular brake for translesion synthesis across benzo[a]pyrene-guanine adducts by human DNA polymerase kappa. *Mutat. Res.-Gen. Tox. En. 718*, 10-17.

(73) Washington, M. T., Helquist, S. A., Kool, E. T., Prakash, L., and Prakash, S. (2003) Requirement of Watson-Crick hydrogen bonding for DNA synthesis by yeast DNA polymerase η. *Mol. Cell. Biol. 23*, 5107-5112.

(74) Wolfle, W. T., Washington, M. T., Kool, E. T., Spratt, T. E., Helquist, S. A., Prakash, L., and Prakash, S. (2005) Evidence for a Watson-Crick hydrogen bonding requirement in DNA synthesis by human DNA polymerase κ. *Mol. Cell. Biol. 25*, 7137-7143.

Chapter 4: Molecular Insights into the Translesion Synthesis of Benzyl-Guanine from Molecular Dynamics Simulations: Structural Evidence for Mutagenic and Non-Mutagenic Replication[a,b,c]

4.1. Introduction

DNA is damaged 60,000 times per cell per day by agents in our environment such as UV radiation, industrial chemicals, by-products from the combustion of fossil fuels, and agricultural contaminants.[1] One class of the resulting DNA lesions is alkylation adducts (addition products), which are formed when an alkyl or aryl group adds to a DNA nucleobase. Although alkylation adducts occur at low physiological levels, they are particularly harmful, especially at the O6 position of dG (Figure 4.1).[2, 3] Many O6-dG alkylation adducts have been reported in the literature, including O6-methyl-guanine (Me-dG), O6-ethyl-guanine (Et-dG), O6-benzyl-guanine (Bz-dG), O6-carboxymethyl-guanine (CM-dG), and O6-pyridyloxobutyl-guanine (POB-dG; Figure 4.1). These alkylation adducts can form upon exposure of guanine to many different compounds. For example, Me-dG and POB-dG can result from a tobacco carcinogen that has been linked to lung cancer.[4-8] Additionally, synthetic alkylation adducts, such as Bz-dG, have been developed as inhibitors for O6-alkylguanine-DNA alkyltransferase for cancer treatment, but are carcinogenic to patients with long-term survival.[9, 10]

---

[a] *Biochemistry* reference style used throughout this chapter

[b] Reprinted with permission from Wilson, K. A., and Wetmore, S.D. (2017). Molecular insights into the translesion synthesis of benzyl-guanine from molecular dynamics simulations: Structural evidence of mutagenic and nonmutagenic replication. *Biochemistry* 56, 1841-1853.

[c] K.A.W. performed calculations, data analysis and wrote the first draft of the manuscript.

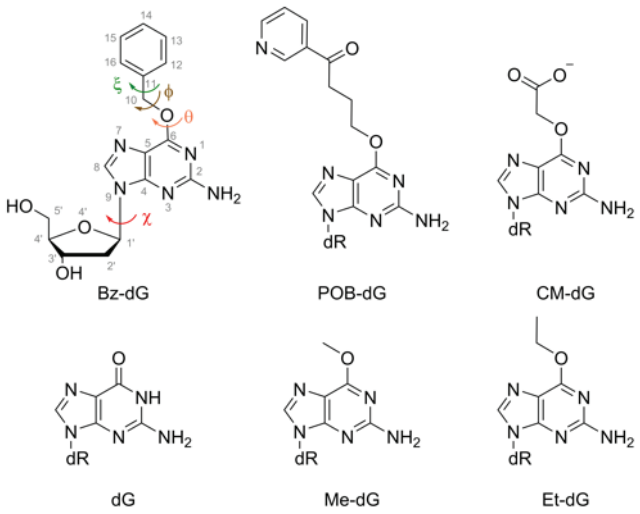

Figure 4.1. Structures of the Bz-dG nucleoside, other O6-dG alkylation adducts and canonical dG, as well as the definitions of key rotatable dihedral angles in Bz-dG [$\chi = \angle(O4'C1'N9C4)$, $\theta = \angle(N1C6O6C10)$, $\varphi = \angle(C6O6C10C11)$, and $\xi = \angle(O6C10C11C12)$].

The propensity of O6-dG alkylation adducts to cause cancer stems from their effects on DNA replication, specifically their tendency to form base mispairs. In general, such genetic mutations occur since the standard DNA replication pathway is impeded or completely blocked by bulky adducts.[11, 12] Therefore, an alternative replication pathway, namely translesion synthesis (TLS), typically processes these lesions. TLS (bypass) polymerases have more open and solvent accessible active sites than standard (replicative) polymerases.[11, 12] Although these features permit replication of bulky DNA lesions, TLS polymerases have lower selectivity, fidelity, efficiency, and processivity compared to replicative polymerases.[11, 12] Notably, the type of DNA damage has been shown to affect the mechanism and efficiency of TLS, and therefore understanding how TLS polymerases replicate different types of DNA damage will allow general conclusions to be made about this critical biological pathway. Unfortunately, while many studies have investigated the replication of N2-dG and C8-dG adducts (see, for example, 13-21), less

work has been completed on O6-dG adducts.[22-26] In particular, there is no structural information for the insertion of a dNTP opposite an O6-dG addition product.

Nevertheless, some key structural features are available that shed light on how O6-dG alkylation adducts are accommodated in a DNA duplex and the active site of a TLS polymerase. Specifically, molecular dynamics (MD) and nuclear magnetic resonance spectroscopy (NMR) data for damaged DNA duplexes suggest Me-dG adopts the *anti* ($\chi = \angle(O4'C1'N9C4) = 180\pm90°$) orientation opposite dC, with the methyl group located in the major groove.[27-31] While MD indicates that Bz-dG can assume the same conformation opposite a pyrimidine, Bz-dG can also adopt the *syn* ($\chi = 0\pm90°$) orientation with the bulky moiety intercalated into the helix with or without displacing an opposing dC.[32] Conversely, NMR and X-ray crystallographic data for the post-lesion synthesis complex (i.e., replication of the 5′-base with respect to the lesion) of Bz-dG or Me-dG adducted DNA bound to DNA polymerase IV (Dpo4) have only shown the adduct in the *anti* orientation opposite a pyrimidine.[23, 26, 33-37] Together, this data on the DNA duplexes and post-lesion synthesis complexes suggests that *anti*-O6-dG alkylation adducts form two hydrogen bonds with an opposing dC in a wobble base-pairing arrangement, and form two hydrogen bonds with an opposing dT in a pseudo Watson-Crick base-pairing arrangement (Figure C.1, Appendix C).[23, 27-32] These hydrogen-bonding patterns contrast the canonical Watson-Crick dG:dC base pair due to the modified hydrogen-bonding face of O6-dG adducts (i.e., N1 is a hydrogen-bond acceptor rather than a hydrogen-bond donor; Figure 4.1).

To complement MD, NMR, and X-ray crystallographic studies, kinetic studies have been completed for the replication of O6-dG alkylation adducts by a range of TLS polymerases.[22-25] The current work focuses on the TLS of Bz-dG, which is a common experimental model to investigate the replication of O6-dG alkylation adducts.[23-25, 38] The replication of Bz-dG by Dpo4 has been reported to result in dT, dA, and dG misincorporation frequencies (i.e., k_{cat}/K_m for incorrect dNTP insertion divided by k_{cat}/K_m for correct dCTP insertion) of 0.14, 0.21, and 0.059, respectively.[23] This correlates with information from mass spectroscopy, which suggests that nucleotide insertion opposite Bz-dG is ~70% dC, ~15% dT, and ~15% dA.[23] Nevertheless, another kinetic study has reported misincorporation frequencies of 0.024, 0.0091, and 0.04 for dT, dA, and dG, respectively.[20] Interestingly, unlike reported for the replication of other bulky adducts (e.g., C8-AP-dG and the C8 and N2-dG polycyclic aromatic hydrocarbons (PAH) adducts),[20, 39-42] deletion mutations are not formed upon replication of Bz-dG by Dpo4.[23] Additionally, both nascent Bz-dG:dC and Bz-dG:dT base pairs can be extended (extension frequency of 0.14 or 0.52 and 0.0034 or 0.04, respectively), [23, 25] while the Bz-dG:dA pair is extended at a very low frequency (0.00079).[25] The human TLS polymerases η and ι, as well as the human replicative polymerases T7$^-$ and δ, show little preference for dC or dT insertion opposite Bz-dG, with measured dT misincorporation frequencies of ~1.2. [23,26] In contrast, dA insertion is not observed opposite Bz-dG upon replication by polymerases η, ι, T7$^-$ and δ, with measured misincorporation frequencies of ~0.01.[23,26] Nevertheless, replication by human TLS polymerase κ leads to a dT misincorporation frequency of 0.55.[24] Overall, despite disagreement in the quantitative incorporation of each nucleobase opposite Bz-dG, these experimental kinetics studies generally agree that dC is most commonly inserted opposite Bz-dG and dT is also

inserted opposite Bz-dG. Furthermore, the data indicates that dA could possibly be inserted opposite Bz-dG. Unfortunately, the structural information on TLS of Bz-dG is currently limited to replication of the 5′-base with respect to the adduct.[23, 25] However, the mutagenicity of adduct replication stems from the insertion of a dNTP opposite the adduct, and structural information that explains the observed Bz-dG replication outcomes is currently unavailable.

The present work uses MD simulations to provide molecular level details of the replication of Bz-dG to understand the full range of possible TLS outcomes. Focus is placed on lesion bypass by Dpo4, which is considered to be a prototypical TLS polymerase and has been experimentally used to study the translesion synthesis of a range of DNA adducts (see, for example, 15-18, 21, 43-45). Specifically, MD simulations were completed on all possible ternary insertion complexes (i.e., insertion of each canonical dNTP opposite different (*anti/syn*) Bz-dG conformations) and ternary –1 base deletion complexes (i.e., insertion of the complementary dNTP opposite the 5′-base with respect to (*anti/syn*) Bz-dG, with no base opposite the lesion). Key structural and energetic parameters that dictate successful dNTP insertion were rigorously analyzed to reveal the most viable replication outcomes. Overall, our calculations provide the first structural insight into the experimentally-observed preferential insertion of dCTP and misincorporation of dTTP opposite Bz-dG, as well as explain the lack of deletion mutations for this adduct. Our findings provide a basis for understanding the replication of related O6-dG adducts.

4.2. Computational Details

MD simulations were performed on the Dpo4 ternary complex for the replication of Bz-dG based on a crystal structure of the Dpo4 ternary complex corresponding to dATP insertion opposite the 5′-dT with respect to a benzo[a]pyrene adenine adduct.[46] The nascent base pairs considered include *anti*-Bz-dG paired opposite *anti*-dCTP, *anti*-dTTP, *syn*-dGTP, or *syn*-dATP, and *syn*-Bz-dG paired opposite *anti*-dCTP, *anti*-dTTP, *anti*-dATP, or *anti*-dGTP. Additionally, a control simulation was performed for the ternary insertion complex with *anti*-dG paired opposite *anti*-dCTP. Furthermore, ternary –1 base deletion complexes were investigated in which no base is positioned opposite *anti* or *syn*-Bz-dG, and dGTP was paired opposite the 5′-dC with respect to the adduct. All natural amino acids, nucleotides, and the solvent were modeled with AMBER ff99SB parameters,[47] while the parameters for Bz-dG[32] and the dNTPs[48-50] were adapted from the literature. For each of the 11 complexes, a 20 ns AMBER MD simulation was carried out in explicit solvent (water and Na^+ counterions).[51-53] Each simulation was stable, with an overall backbone root-mean-square deviation (rmsd) of ~1.3–2.3 Å (Table C.1, Appendix C). To further confirm adequate sampling over the 20 ns, the trajectories for the insertion of *anti*-dCTP opposite *anti*-Bz-dG or *anti*-dG were extended to 100 ns. Extending the simulations led to only small deviations in the structures and do not change the overall conclusions (see Tables C.2–C.7, Appendix C, and Figures 4.2–4.4, C.2 and C.3 for a comparison of the data). Throughout the results and discussion, average interaction energies are reported with the dynamical information (standard deviations) provided in the Appendix C. Furthermore, the positions of the nucleobases are referenced with respect to the base being replicated (Bz-dG or dG). Additional details of the methods used in the present work are provided in Appendix C.

4.3. Results

4.3.1. Criteria Used to Assess the Catalytic Nature of Dpo4 Ternary Complexes:

Previous research has proposed a reaction mechanism for the replication of natural DNA by Dpo4,[54, 55] which involves dNTP selection and then addition to the growing DNA chain through O3′ attack on the dNTP α-phosphate (Figure C.4, Appendix C). Previous literature on the replication of natural and adducted DNA has identified many factors (discussed below) that must be favorably aligned in order for this reaction to occur. Although these structural and energetic parameters could be individually considered, they are clearly interdependent and adduct replication will only occur with high efficiency when all parameters are satisfactory for the reaction. Therefore, to determine the effects of Bz-dG on the reactant complex for lesion replication and the preferred dNTP insertion opposite the adduct, all of these previously identified factors required for successful replication of natural and adducted DNA were simultaneously investigated. While this detailed analysis of the reactant complexes provides key information about possible Bz-dG replication outcomes, we acknowledge that future work must investigate the structure and energetics of the transition state to fully understand the replication reaction.

Since previous computational studies show that Bz-dG can adopt many different conformations in duplex DNA,[32] we carefully monitored the conformations of Bz-dG, as well as the pairing base, in the polymerase active site. Furthermore, previous studies have reported that hydrogen bonding between the base being replicated and the incoming dNTP is an essential component of TLS.[25, 56, 57] In the case of a dG adduct, the ideal hydrogen-bonding interaction with the incoming dNTP mimics a natural dG:dC Watson-Crick base pair, with a C1′–C1′ distance of 10.8 Å, 3 hydrogen bonds, and a hydrogen-

bond strength of approximately –110 kJ mol^{-1}. Hydrogen bonding in the previously replicated base pairs should also be maintained upon lesion replication. Additionally, helical stacking between the nascent and 3′-base pairs, and the adduct and the 5′-base, should be consistent with a canonical DNA duplex. Indeed, strong interactions between the bulky moiety and the 5′-base have been proposed to impede translocation of the TLS polymerase.[58]

Beyond the important interactions within DNA (including the dNTP), several interactions between the adduct or dNTP and the polymerase must be maintained for successful replication. Specifically, computational and experimental studies have shown that the location of the adduct is maintained through hydrogen bonds between the adduct phosphate moiety and Dpo4 (Arg331 and Ser34; Figure C.5a, Appendix C).[39, 48, 55, 59-61] Nevertheless, interactions between the bulky moiety and the TLS polymerase have been proposed to stall replication.[48, 62] For example, a hydrogen-bonding interaction between the bulky moiety of PhIP-C8-dG and Arg322 (Dpo4),[48] or a stacking interaction between BP-N2-dG and Phe171 (polymerase κ)[62] has been hypothesized to stall replication. Additionally, the dNTP location is stabilized by hydrogen bonds between O3′ and Tyr12(NH), Oβ and Phe11(NH), Thr45(OγH) and Arg51(NζH), and Oγ and Tyr10(NH), Tyr48(OH) Arg51(NζH) and Lys159(NζH) (Figure C.5b, Appendix C).[39, 48, 55, 59-61] A conserved sugar–π interaction (dNTP–Tyr12) allows Dpo4 to be selective toward deoxyribose over ribose,[63, 64] and has been proposed to be crucial for the bypass of DNA adducts.[60] Additionally, the octahedral coordination of the two catalytic Mg^{2+} ions is essential for stabilizing the dNTP (nucleotide binding ion) and terminal O3′ of the primer (catalytic ion).

Finally, in order for the reaction to occur, the dNTP must be correctly aligned with respect to the 3′ terminus of the primer strand. Specifically, the distance between the terminal O3′ of the primer strand and the α-phosphate of the dNTP must approach the sum of the van der Waals radii for O and P (3.5 Å). Furthermore, previous X-ray crystallographic and computational data on natural DNA replication suggests that the in-line attack angle [∠O3′(primer 3′ end)–Pα(dNTP)–Oαβ(dNTP)] should fall between 150 and 180°.[54, 55, 65]

4.3.2. The simulated orientation of dCTP opposite dG in the Dpo4 active site is conducive for replication: When dCTP is paired opposite dG in the Dpo4 ternary insertion (control) complex (Figure 4.2), both dG and dCTP adopt the *anti* conformation (χ; Table C.2, Appendix C). The dG:dCTP and 3′-dC:dG base pairs contain three canonical Watson-Crick hydrogen bonds with 100% occupancy (i.e., the percent of the simulation time that the heavy atom distance is < 3.4 Å and the hydrogen-bonding angle is < 120°; Figure 4.2b and Table C.3, Appendix C), and an associated average interaction strength of –127.6 and –122.0 kJ mol^{-1}, respectively (Table C.4, Appendix C). The C1′–C1′ distance in the nascent dG:dCTP pair is on average ~10.8 Å (Table C.2, Appendix C), which is within the range expected for a catalytically-active complex. The stacking interaction between dG:dCTP and the 3′-dC:dG pair stabilizes the growing DNA duplex (–62.9 kJ mol^{-1}; Table C.4, Appendix C) . Conversely, there is little interaction between the 5′-dC and the nascent base pair (–1.0 kJ mol^{-1}; Table C.4, Appendix C). The location of dG is further maintained by hydrogen bonds between the phosphate moiety, and Ser34 and Arg331 for 36 and 98% of the simulation, respectively (Table C.5, Appendix C).

Additionally, many hydrogen-bonding interactions occur between the incoming dCTP and Dpo4. Specifically, dCTP forms multiple hydrogen bonds with Lys159 (69 and 37% occupancy) and Arg51 (84, 74, 57, and 51% occupancy; Table C.5 and Figure C.6, Appendix C), which stabilize dCTP binding to the DNA–polymerase complex by –191.4 and –127.7 kJ mol^{-1}, respectively (Table C.6, Appendix C). Furthermore, dCTP hydrogen bonds with Tyr10 (100% occupancy), Thr45 (100% occupancy), Tyr12 (79% occupancy), and Phe11 (67% occupancy), and forms a sugar–π interaction with Tyr12 (Figure C.7, Appendix C), with each residue contributing up to approximately –60 kJ mol^{-1} to dCTP binding (Table C.6, Appendix C). The Mg^{2+} ions are slightly repositioned in the active site compared to the crystallographic Ca^{2+} ions (rmsd of ~1.1 Å), which allows them to exhibit catalytically-active coordination, with 11 of the 12 coordination distances being < 2.5 Å for > 90% of the simulation (Table C.7, Appendix C). Indeed, the nucleotide binding and catalytic Mg^{2+} ions contribute a total of –937.1 and –435.7 kJ mol^{-1} to dCTP binding, respectively (Table C.6, Appendix C). Based on the dCTP location, the geometry of the complex is favorable for the reaction to occur, with an average Pα–O3′ distance of ~3.3 Å and an in-line attack angle of ~161° (Figure 4.2c). Overall, the simulated structure of the control ternary insertion complex is conducive for replication, which gives us confidence our methodology is predictive of experimental TLS outcomes and provides us with a frame of reference for evaluating the ternary insertion complexes corresponding to the replication of Bz-dG adducted DNA.

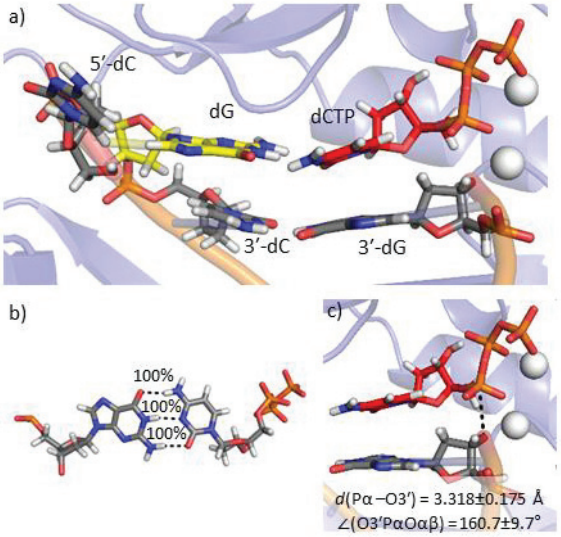

Figure 4.2. Representative MD structure for the Dpo4 ternary complex of canonical dG replication depicting a) the orientation of dCTP with respect to the DNA, b) the dG:dCTP hydrogen-bonding arrangement and the percent occupancy of the hydrogen-bonding interactions, and c) the average reaction parameters.

4.3.3. Due to stabilizing DNA–DNA and DNA–polymerase interactions, dCTP is favorably oriented for insertion opposite both *anti* and *syn*-Bz-dG: The bulky moiety of both *anti* and *syn*-Bz-dG adopts an extended conformation with respect to the damaged dG ($\varphi \approx 180°$; Table C.2, Appendix C). The bulky moiety of *anti*-Bz-dG extends into the major groove, while the bulky moiety of *syn*-Bz-dG intercalates between the 3′-base pair and the pairing *anti*-dCTP (Figure 4.3a and b). Furthermore, significant variation is observed in θ and ξ (standard deviations up to ~90°), and therefore the bulky moiety is directed toward N1 or N7 of the adducted dG and there is free rotation of the phenyl ring (Table C.2, Appendix C).

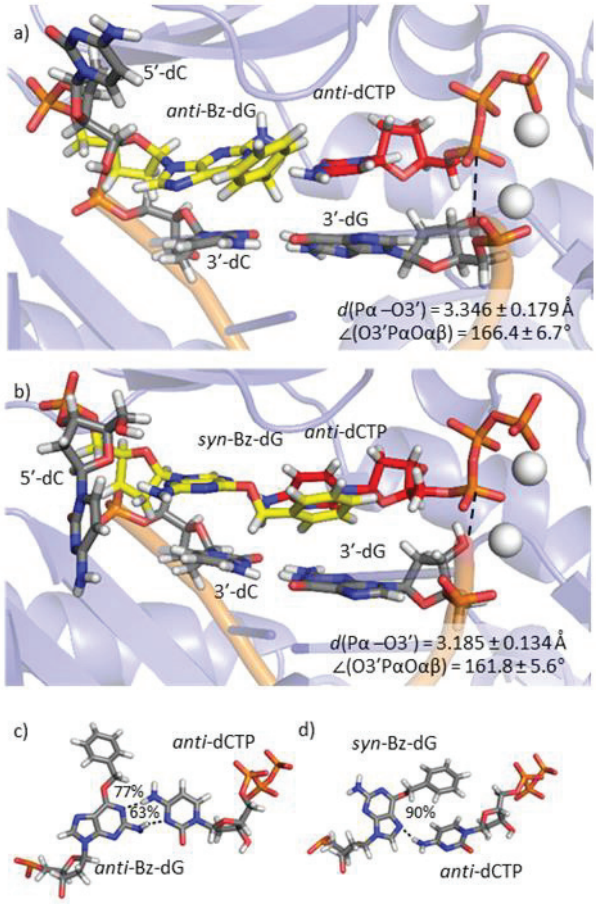

Figure 4.3. Representative MD structures depicting (a and b) the orientation of dCTP with respect to the DNA in the Dpo4 ternary complexes and (c and d) possible Bz-dG:dCTP hydrogen-bonding arrangements for (a and c) *anti*-Bz-dG and (b and d) *syn*-Bz-dG replication.

A wobble base pair is formed between *anti*-Bz-dG and *anti*-dCTP that contains two [Bz-dG(N1)···dCTP(N4H) and Bz-dG(N2H)···dCTP(N3)] hydrogen bonds for 63 and 77% of the simulation, respectively (Figure 4.3c). This hydrogen-bonding pattern is consistent with that observed in an MD study of Bz-dG adducted DNA,[32] and a crystal structure of the post-lesion synthesis complex for replication of Bz-dG containing DNA by Dpo4 (rmsd based on nucleobase heavy atoms of 0.383 Å; Figure C.8a, Appendix C).[23] Conversely, when *syn*-Bz-dG is paired opposite *anti*-dCTP, a single Bz-dG(N7)···dCTP(N4H) hydrogen bond is formed for 90% of the simulation (Figure 4.3d).

However, stacking interactions also occur between the *syn*-Bz-dG bulky moiety and the pairing dCTP (Figure 4.3b), which are not present during *anti*-Bz-dG or dG replication. The *syn*-Bz-dG hydrogen-bonding pattern is consistent with a structure reported in a MD study of Bz-dG adducted DNA,[32] and an analogous intercalated structure observed in an X-ray crystal structure of DNA containing the Bz-dG:dPer pair.[33] Both the *anti* and *syn*-Bz-dG:dCTP pairs are significantly weaker than the canonical Watson-Crick dG:dC pair due to the changes in the Watson-Crick binding face, with interaction strengths of –23.0 and –29.3 kJ mol^{-1}, respectively (Table C.4, Appendix C). Interestingly, regardless of the location of the bulky moiety, the strength of the stacking interaction between Bz-dG:dCTP and the 3′-dC:dG pair is comparable to that for the replication of canonical dG (approximately –60 kJ mol^{-1}; Table C.4, Appendix C) and the 3′-dG:dC pair retains > 90% occupancy of the three Watson-Crick hydrogen bonds (Table C.3, Appendix C). Furthermore, similar to the replication of natural dG, regardless of the Bz-dG glycosidic orientation, Bz-dG does not form strong interactions with the 5′-dC (approximately –2 kJ mol^{-1}; Table C.4, Appendix C).

Although no hydrogen bonds occur between the Bz-dG bulky moiety and Dpo4, hydrogen bonds exist between the phosphate moiety of *anti* or *syn*-Bz-dG, and Ser34 and Arg331 for > 60% of the simulation (Table C.5, Appendix C). Furthermore, hydrogen bonds are formed between dCTP, and Tyr10, Thr45, Phe11, Arg51, and Lys159 with similar geometric parameters as observed in the control simulation (Figure C.6, Appendix C), and the dR:Tyr12 sugar–π interaction is consistently maintained (Figure C.7, Appendix C). These interactions stabilize dCTP binding to the DNA–polymerase complex to a similar extent as for dG replication. Specifically, the largest contribution to

dCTP binding arises from Arg51 (approximately –190 kJ mol^{-1}) and Lys159 (approximately –150 kJ mol^{-1}), with significant contributions also coming from Tyr10, Tyr45, and Phe11 (up to approximately –60 kJ mol^{-1} each; Table C.6, Appendix C). In addition to the dNTP interactions with the DNA and protein, the coordination of the Mg^{2+} ions is catalytically favorable for both *anti* and *syn*-Bz-dG, with > 90% occupancy of a distance < 2.5 Å for 11 of the 12 coordinating atoms (Table C.7, Appendix C). As seen in the control simulation, the nucleotide binding and catalytic Mg^{2+} ions contribute approximately –950 or –440 kJ mol^{-1} to dCTP binding, respectively (Table C.6, Appendix C). Furthermore, the Pα–O3′ distance and in-line attack angle are reaction ready, being on average ~3.3 Å and ~166°, or ~3.2 Å and ~162° for *anti* or *syn*-Bz-dG replication, respectively (Figure 4.3). Taken together, our data indicates the Dpo4 active site conformation is preserved compared to the control simulation, and suggests the Dpo4 ternary insertion complex is conducive for dCTP insertion opposite either *anti* or *syn*-Bz-dG.

4.3.4. dTTP is favorably aligned for insertion opposite *anti*-Bz-dG, while the lack of DNA–DNA interactions likely precludes dTTP insertion opposite *syn*-Bz-dG: For both *anti* and *syn*-Bz-dG opposite *anti*-dTTP, an extended adduct conformation is adopted ($\varphi \approx 180°$; Table C.2, Appendix C), which places the bulky moiety in the major groove (Figure 4.4a and b). However, there is significant flexibility in the bulky moiety regardless of the orientation about the glycosidic bond, with standard deviations in θ and ξ being up to ~100° (Table C.2, Appendix C). Consequently, the phenyl ring is directed toward N1 or N7 of dG and exhibits free rotation.

In the *syn*-Bz-dG:*anti*-dTTP pair, there are no interbase hydrogen-bonding interactions, and this base pair is wider than a canonical pair by 1.1 Å (Figure 4.4). Due to this disruption, the stacking interaction between *syn*-Bz-dG:*anti*-dTTP and the 3′-dG:dC is on average ~10 kJ mol^{-1} less stable than observed in the control simulation (Table C.4, Appendix C). The distortion in the nascent base pair also decreases the occupancy of two Watson-Crick hydrogen bonds in the 3′-dG:dC pair (70 and 77%; Table C.3, Appendix C), and decreases the interaction energy by ~20 kJ mol^{-1} (Table C.4, Appendix C). Furthermore, *syn*-Bz-dG forms a strong interaction with the 5′-dC (approximately –25 kJ mol^{-1}; Table C.4, Appendix C), which may prevent translocation.[58] This lesion site misalignment that distorts key DNA–DNA interactions likely impedes dTTP insertion opposite *syn*-Bz-dG.

In contrast to *syn*-Bz-dG, viable interbase hydrogen bonding occurs for dTTP insertion opposite *anti*-Bz-dG in two distinct hydrogen-bonding patterns. Specifically, a pseudo Watson-Crick base pair is formed for the first 8.6 ns of the simulation, which contains two [Bz-dG(N1)···dTTP(N3H) and Bz-dG(N2H)···dTTP(O2)] hydrogen bonds (40 and 92% occupancy, respectively; Figure 4.4c). During the remaining 11.4 ns of the simulation, an alternative conformation exists due to repulsion between *anti*-Bz-dG and the pairing dTTP, which results in only a single Bz-dG(N2H)···dTTP(O4) hydrogen bond (88% occupancy; Figure C.9, Appendix C). Both hydrogen-bonding arrangements between *anti*-Bz-dG and *anti*-dTTP are weaker than the *anti*-Bz-dG:*anti*-dCTP interaction, with strengths of –16.1 and –8.3 kJ mol^{-1}, respectively (Table C.4, Appendix C). The pseudo Watson-Crick hydrogen bonding observed in our ternary insertion complex is similar to the crystal structure geometry of the *anti*-Bz-dG:*anti*-dT base pair in

the Dpo4 post-lesion synthesis complex (rmsd based on nucleobase heavy atoms is 0.523 Å; Figure C.8b, Appendix C).[23] Furthermore, our observed change in hydrogen-bonding orientation is consistent with significant disorder in the *anti*-Bz-dG:*anti*-dT base pair reported in this crystal structure, which was proposed to occur due to repulsion between the exocyclic carbonyl groups.[23] The presence of multiple hydrogen-bonding patterns is further supported by the significant disorder observed in the dT electron density for the *anti*-Me-dG:*anti*-dT pair in the Dpo4 post-lesion synthesis complex,[34] which led to the assignment of three distinct bonding orientations, including the pseudo Watson-Crick base pair and two orientations that reduce electrostatic repulsion between the carbonyl groups. The change in the hydrogen-bonding orientation observed in the present work affects the stacking interactions of the nascent adducted base pair. Specifically, while *anti*-Bz-dG does not interact with the 5′-dC during the first 8.6 ns of the simulation, *anti*-Bz-dG favorably stacks with the 5′-dC during the last 11.4 ns of the simulation (–17.1 kJ mol^{-1}; Table C.5, Appendix C). Nevertheless, changes to the *anti*-Bz-dG:*anti*-dT base pair geometry do not disrupt stacking with the 3′-flanking base pair (approximately –60 kJ mol^{-1} regardless of the hydrogen-bonding geometry; Table C.5, Appendix C) and the 3′-dG:dC pair retains three canonical Watson-Crick hydrogen bonds for > 90% of the simulation (Table C.3, Appendix C).

In general, there are no significant changes in the interactions between Bz-dG and the polymerase compared to the control simulation, regardless of the glycosidic orientation. Specifically, the phosphate moiety of Bz-dG hydrogen bonds with Ser34 and Arg331, which stabilizes the adduct location in the Dpo4 active site (Table C.5, Appendix C). No interactions form between the Bz-dG bulky moiety and the polymerase. The Tyr12:dR π-

interaction occurs throughout both simulations (Figure C.7, Appendix C). Additionally, hydrogen bonds between *anti*-dTTP and Phe11, Tyr10, Thr45, or Arg51 occur with similar frequencies as the control (Figure C.4 and Table C.5, Appendix C). As a result, Arg51 contributes approximately –185 kJ mol^{-1}, and Tyr10 and Thr45 each contribute up to approximately –60 kJ mol^{-1} to dTTP binding (Table C.6). Nevertheless, while the location and strength of the interaction between dTTP and Lys159 remain similar to the control for the replication of *anti*-Bz-dG, a change in the orientation of Lys159 relative to the dNTP for replication of *syn*-Bz-dG significantly increases the Lys159–dNTP interaction strength (by ~80 kJ mol^{-1}). This alteration in the Lys159 orientation will likely further impede catalysis since Lys159 has been shown to play a critical role in stabilizing the dNTP.

The maintenance of the majority of critical dTTP-polymerase interactions leads to reaction-ready parameters in the ternary insertion complex regardless of the Bz-dG glycosidic orientation or *anti*-Bz-dG hydrogen-bonding orientation. Indeed, 11 of the 12 interactions with the Mg^{2+} ions have a distance < 2.5 Å for > 90% of the simulation (Table C.7, Appendix C), the Pα–O3′ distance is ~3.4 Å, and the in-line attack angle is ~168° (Figure 4.4 and Table C.7, Appendix C). Overall, although dTTP is correctly positioned with respect to the growing DNA chain when paired opposite *anti* and *syn*-Bz-dG in the Dpo4 active site, dTTP will likely only be inserted opposite *anti*-Bz-dG due to the stark lack of DNA–DNA interactions and changes to the DNA–protein interactions in the *syn*-Bz-dG:*anti*-dTTP complex.

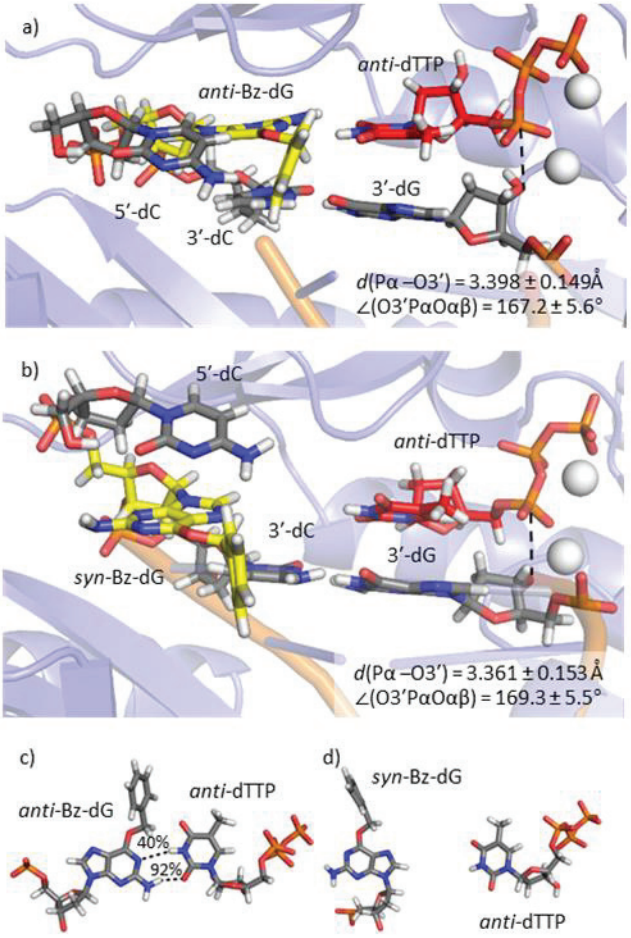

Figure 4.4. Representative MD structures depicting (a and b) the orientation of dTTP with respect to the DNA in the Dpo4 ternary complexes and (c and d) possible Bz-dG:dTTP hydrogen-bonding arrangements for (a and c) *anti*-Bz-dG and (b and d) *syn*-Bz-dG replication. The hydrogen-bonding arrangement for *anti*-Bz-dG:*anti*-dTTP occurs during the first 8.6 ns of the simulation. See Figure C.9 for the hydrogen-bonding pattern for *anti*-Bz-dG:*anti*-dTTP pair during the last 11.4 ns of the simulation.

4.3.5. Distortions in the Dpo4 active site that alter DNA–polymerase and DNA–DNA interactions likely prevent insertion of a purine opposite Bz-dG: When Bz-dG is in the *anti* orientation, the pairing purine dNTP adopts the *syn* orientation, while *syn*-Bz-dG pairs opposite an *anti*-purine dNTP in the Dpo4 active site (Figure 4.5). Similar to MD results for Bz-dG adducted DNA,[32] the Bz-dG bulky moiety is located in the major groove and adopts many orientations (standard deviations in φ, θ, and ξ of up to ~100°;

Table C.2, Appendix C). Specifically, the bulky moiety is either positioned towards N1 or N7 of dG, oriented perpendicular or parallel to dG, and there is free rotation of the phenyl moiety in the Dpo4 ternary insertion complexes.

Hydrogen bonding between Bz-dG and either canonical purine dNTP is not particularly favorable (Figure 4.6 and Table C.4, Appendix C). The strongest hydrogen bonding with a purine dNTP (–24.7 kJ mol^{-1}) occurs when *anti*-Bz-dG forms three [Bz-dG(N2H)···dGTP(O6), Bz-dG(N1)···dGTP(N1H), and Bz-dG(O6)···dGTP(N2H)] hydrogen bonds with *syn*-dGTP (7, 12 and 38% occupancy, respectively). Nevertheless, this base-pairing arrangement varies significantly throughout the simulation as evident by the low occupancies of the hydrogen bonds and high standard deviation for the interaction energy (±18.4 kJ mol^{-1}). In contrast, *anti*-Bz-dG forms two [Bz-dG(N2H)···dATP(N6H) and Bz-dG(N3)···dATP(N6H)] hydrogen bonds (83 and 71% occupancy, respectively) opposite *syn*-dATP (–13.4 kJ mol^{-1}), and *syn*-Bz-dG forms a single [Bz-dG(N7)···dATP(N6H)] hydrogen bond of low occupancy (22%) opposite *anti*-dATP (–6.3 kJ mol^{-1}). Furthermore, *syn*-Bz-dG does not form any hydrogen bonds with *anti*-dGTP. In direct correlation with these interactions, the C1′–C1′ distance between Bz-dG and dATP is smaller (by up to ~1.2 Å), while the C1′–C1′ distance between Bz-dG and dGTP is greater (by up to ~1.9 Å; Table C.2, Appendix C), than that of the canonical dG:dCTP pair in the Dpo4 active site. Nevertheless, the Watson-Crick hydrogen bonding in the 3′-flanking dC:dG is generally retained (Tables C.3 and C.4, Appendix C), while stacking interactions between the adducted base pair and 3′-dC:dG are generally weaker (by up to ~20 kJ mol^{-1}) than those for the dG:dCTP control (Table C.4, Appendix C).

Furthermore, Bz-dG generally forms strong interactions with the 5′-dC (up to approximately –25 kJ mol^{-1}) that are not present in the control.

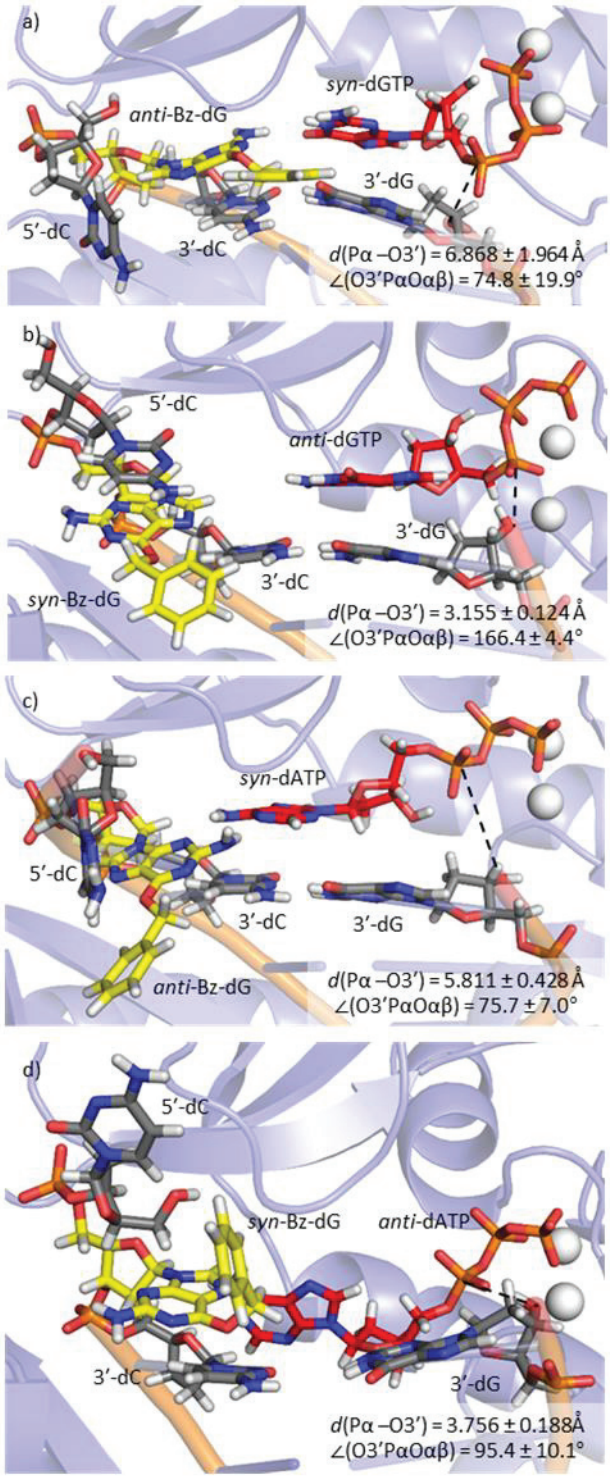

Figure 4.5. Representative MD structures depicting the orientation of the purine dNTP in the Dpo4 ternary complexes for a) *anti*-Bz-dG:*syn*-dGTP, b) *syn*-Bz-dG:*anti*-dGTP, c) *anti*-Bz-dG:*syn*-dATP, and d) *syn*-Bz-dG:*anti*-dATP.

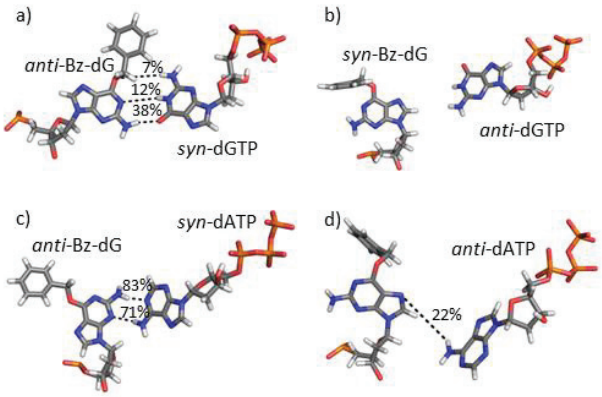

Figure 4.6. Representative MD structures depicting possible hydrogen-bonding arrangements of Bz-dG with purine dNTPs for a) *anti*-Bz-dG:*syn*-dGTP, b) *syn*-Bz-dG:*anti*-dGTP, c) *anti*-Bz-dG:*syn*-dATP, and d) *syn*-Bz-dG:*anti*-dATP replication.

Regardless of the Bz-dG glycosidic orientation and pairing purine dNTP, hydrogen bonds between the phosphate moiety of Bz-dG and Ser34 or Arg331 occur during the simulation (Table C.5, Appendix C). As observed for pyrimidine dNTP insertion opposite Bz-dG, no hydrogen bonds occur between the bulky moiety and polymerase for purine dNTP insertion opposite *syn*-Bz-dG. However, for purine dNTP insertion opposite *anti*-Bz-dG, hydrogen bonding occurs between Bz-dG(N2H) and the backbone oxygen of Gly58 for up to 34% of the simulation (Table C.5, Appendix C), which may impede translocation.[48, 62] In all cases for purine dNTP insertion opposite Bz-dG, the active site conformation is different than observed in the control, with key dNTP–polymerase interactions being disrupted. This change in the active site conformation significantly weakens DNA–polymerase interactions. For example, Arg51 contributes up to ~170 kJ mol^{-1} less than in the control (Figure C.6 and Table C.6, Appendix C). Furthermore, hydrogen bonds exist that are not observed in the control. For example, Tyr12(OH) hydrogen bonds with *syn*-dGTP(N7) paired opposite *anti*-Bz-dG, and Tyr12(OH) hydrogen bonds with *anti*-dATP(PαO) paired opposite *syn*-Bz-dG (Table C.5, Appendix

C). As a result, although the dNTP:Tyr12 sugar–π interaction does not occur for purine dNTP insertion opposite Bz-dG (Figure C.7, Appendix C), Tyr12 contributes approximately the same amount to dNTP binding as seen in the control.

Finally, even though *syn*-Bz-dG does not base pair with *anti*-dGTP, the reaction parameters are optimally aligned, with an ~3.2 Å reaction distance and an ~166° in-line attack angle (Figure 4.5), as well as full coordination of the catalytic Mg^{2+} ions (all 12 coordination interactions have > 90% occupancy of a distance < 2.5 Å; Table C.7, Appendix C). Nevertheless, the terminal 3′-OH in the template strand forms a direct hydrogen bond with dGTP(Pα) that is not present for canonical dG replication. Conversely, despite at least some hydrogen bonding in the lesion pair, the reaction distance and angle for *anti*-Bz-dG opposite *syn*-dGTP, or either glycosidic orientation of Bz-dG opposite dATP, are not conducive for the reaction (up to ~6.9 Å and ~75 or 95°, respectively; Figure 4.5). Furthermore, a maximum of 6 of the 12 atoms coordinating to Mg^{2+} have > 90% occupancy of a distance of < 2.5 Å for *anti*-Bz-dG opposite *syn*-dGTP or Bz-dG opposite dATP (Table C.7, Appendix C), which further suggests that these DNA–polymerase complexes are not catalytically competent. Taken together, our data suggests that the purines are unlikely to be inserted opposite either Bz-dG glycosidic orientation by Dpo4 due to the lack of hydrogen-bonding interactions with Bz-dG, changes in the active site conformation and/or the lack of reaction-ready parameters.

4.3.6. Deviations in key reaction parameters prevent –1 base deletion mutations upon Bz-dG replication by Dpo4: Consistent with the ternary insertion complexes, significant variation is observed in φ, θ and ξ in the deletion complex for both *anti* and

syn-Bz-dG, with standard deviations up to ~130° (Table C.8, Appendix C). As a result, the bulky moiety adopts many conformations in the polymerase active site, including being planar or perpendicular with respect to the adducted dG, positioned towards N1 or N7 of the adducted dG, and free rotation of the phenyl moiety. Furthermore, although the bulky moiety is always positioned in the major groove for *anti*-Bz-dG (Figure 4.7a), the bulky moiety shifts between the preferred major groove position and an intercalated position for *syn*-Bz-dG (Figure 4.7b and c). Regardless of the conformational flexibility of the unpaired Bz-dG, the 3′-dC:dG and 5′-dC:dGTP flanking pairs retain a C1′–C1′ distance of ~10.8 Å, > 95% occupancy of the three canonical dG:dC Watson-Crick hydrogen bonds (Table C.3, Appendix C) and have interaction energies of approximately –110 kJ mol^{-1} regardless of the Bz-dG glycosidic orientation (Table C.9, Appendix C). Additionally, the stacking interactions between Bz-dG and 3′-dG:dC are weaker than observed in the control ternary insertion complex for dG replication (by up to ~20 kJ mol^{-1}). Furthermore, the stacking interactions between Bz-dG and the 5′-dG:dCTP are even weaker (by ~15 kJ mol^{-1}) than the interaction between Bz-dG and the 3′-dG:dC in the ternary –1 base deletion complexes. Regardless of the Bz-dG glycosidic orientation, the Arg331 and Ser34 interactions with the phosphate moiety of the 5′-dC in the nascent pair are maintained. However, dGTP hydrogen-bonding interactions with the polymerase generally occur with < 16% occupancy (Table C.10, Appendix C), and the dGTP–Tyr12 sugar–π interaction is not aligned in either deletion complex (Figure C.10, Appendix C). Indeed, the dGTP–polymerase interactions generally contribute less than 10 kJ mol^{-1} to dGTP binding regardless of the orientation about the Bz-dG glycosidic bond (Table C.6, Appendix C). Additionally, neither the reaction distance (> 4.8 Å; Table C.8, Appendix C), the reaction angle (< 103.1°; Table C.8, Appendix C), nor the coordination

of the catalytic Mg^{2+} ions (< 3 coordinated atoms with > 90% occupancy of a distance of < 2.5 Å; Table C.7, Appendix C) is conducive for the reaction regardless of the Bz-dG glycosidic orientation. Therefore, our MD simulations suggest that a –1 base deletion mutation is unlikely to occur upon replication of Bz-dG by Dpo4.

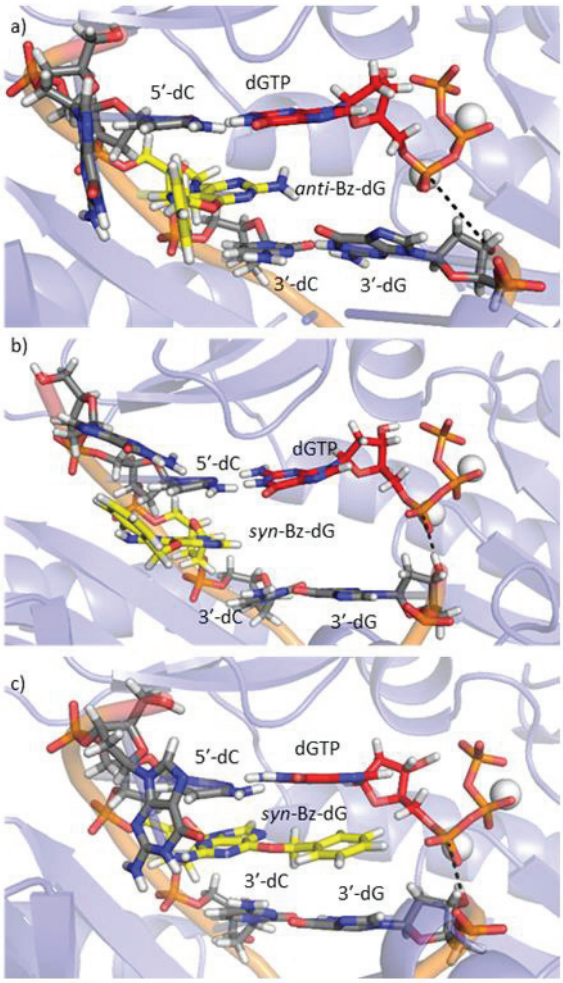

Figure 4.7. Representative MD structures of the Dpo4 –1 base deletion complexes for a) *anti*-Bz-dG, b) *syn*-Bz-dG in the major groove conformation, and c) *syn*-Bz-dG in the intercalated conformation.

4.4. Discussion

4.4.1. MD simulations on the Dpo4 ternary insertion complexes for the replication of Bz-dG provide a structural explanation for the experimentally-observed insertion of

dCTP and dTTP opposite Bz-dG: Numerous experimental studies have investigated the TLS of Bz-dG.[22-25] While these studies show significant variation in the frequency of base mispairs upon replication of Bz-dG, there is a general consensus that dC is the most commonly inserted nucleotide opposite Bz-dG, and that dT has the highest misincorporation frequency.[22-25] Additionally, both nascent Bz-dG:dC and Bz-dG:dT base pairs can be extended, and are in fact the only Bz-dG base pairs that have appreciable extension frequencies.[23, 25] Nevertheless, previous X-ray crystallographic work has only provided structural information for the Dpo4 replication of the 5′-base with respect to Bz-dG.[23, 25] However, lesion mutagenicity stems from the insertion of a dNTP opposite the adduct and therefore it is critical to obtain structural data for this step in order to understand the biological outcome of adduct replication. The present work provides the first atomic level details of the ternary insertion complexes for Bz-dG replication by Dpo4. Specifically, MD simulations were completed on all possible ternary insertion complexes with either *anti* or *syn*-Bz-dG conformations. To identify which complexes are conducive for replication, structural and energetic parameters that dictate successful dNTP insertion were thoroughly analyzed. Indeed, for the reaction to occur, dNTP binding in the active site must be stabilized by both dNTP–DNA and dNTP–polymerase interactions, the DNA must be stabilized by DNA–polymerase contacts, and the dNTP must be correctly aligned for insertion.

To validate our methodology and provide a frame of reference for evaluating the feasibility of adduct replication, the ternary insertion complex for the replication of canonical dG by Dpo4 was investigated. In this complex, the dG:dCTP base pair contains Watson-Crick hydrogen bonding, with an average interaction strength of approximately

–125 kJ mol^{-1}. Furthermore, stacking interactions occur between the nascent base pair and 3′-dC:dG, while there is a lack of interactions between the nascent base pair and 5′-dC. The location of the incoming dCTP is also stabilized through hydrogen bonds with surrounding polymerase residues, such as Lys159, Arg51, Tyr12 Phe11, Tyr10, and Thr45. Overall, these contacts lead to reaction parameters that are aligned for insertion, including an average reaction distance [O3′(primer 3′ end)–Pα(dNTP)] ~3.3 Å and an average in-line attack angle [∠(O3′(primer 3′ end)–Pα(dNTP)–Oαβ(dNTP))] of ~160.7°.

Conversely, when *anti*-dGTP or *anti*-dTTP is paired opposite *syn*-Bz-dG, no hydrogen bonding occurs between Bz-dG and the opposing dNTP, while hydrogen bonding between *anti*-dATP and *syn*-Bz-dG is very weak (i.e., an average interaction energy of approximately –6 kJ mol^{-1}). For these pairing combinations, the C1′–C1′ distance of the nascent base pair is up to 1.7 Å wider than the canonical dG:dCTP base pair (~10.8 Å). This weakens dNTP–DNA interactions (e.g., stacking between the nascent base pair and 3′-dC:dG) and disrupts key dNTP–polymerase interactions (e.g., Tyr12 and Lys159 are repositioned). Thus, although the reaction parameters are aligned for insertion in these cases, the disruption of key dNTP–DNA and dNTP–polymerase interactions likely prevent *anti*-dGTP, *anti*-dTTP or *anti*-dATP insertion opposite *syn*-Bz-dG.

For insertion of *syn*-dATP or *syn*-dGTP opposite *anti*-Bz-dG, weak and transient hydrogen-bonding interactions occurs in the nascent base pair in the Dpo4 active site (i.e., average interaction energies up to approximately –25 kJ mol^{-1} and standard deviations up to ~18 kJ mol^{-1}). Furthermore, the reaction parameters are significantly distorted, with a reaction distance > 5.8 Å and an in-line attack angle of ~75°. There are also significant

distortions to key dNTP–polymerase interactions (i.e., Arg51, Lys159, and Tyr12), which weaken dNTP binding. Consequently, although *syn*-dATP and *syn*-dGTP can form weak hydrogen-bonding interactions with *anti*-Bz-dG, this occurs at the expense of suitable DNA–polymerase interactions and reaction parameters, and therefore successful dNTP insertion is unlikely.

The final pairing options are *anti*-dCTP or *anti*-dTTP insertion opposite *anti*-Bz-dG, or *anti*-dCTP insertion opposite *syn*-Bz-dG. Hydrogen bonding between Bz-dG and the incoming dNTP is significantly weaker in these cases compared to dCTP insertion opposite canonical dG (by ~100 kJ mol^{-1}) as discussed for purine dNTP pairing with *anti*-Bz-dG, which arises due to a change in the Watson-Crick hydrogen-bonding face upon O6-dG alkylation (Figure 4.1). More importantly, these complexes have DNA–DNA and dNTP–DNA stacking interactions, and DNA–polymerase and dNTP–polymerase hydrogen-bonding interactions similar to Dpo4 insertion of dCTP opposite dG. Furthermore, these key interactions preserve reaction-ready parameters. This suggests that both *anti*-dCTP and *anti*-dTTP can be inserted opposite *anti*-Bz-dG, and *anti*-dCTP can also form a catalytically competent complex opposite *syn*-Bz-dG. Although this is the first time a base pair with *syn*-Bz-dG has be observed in a polymerase active site, the stable *syn*-Bz-dG orientation predicted when bound to Dpo4 is consistent with previous MD and X-ray studies on Bz-dG-adducted DNA.[32, 33]

Taken together, our structural data on Dpo4 ternary insertion complexes for insertion of each canonical dNTP opposite different Bz-dG conformations correlates with the experimentally-observed preference for dCTP insertion opposite Bz-dG, as well as the

greatest number of mispairs stemming from dTTP insertion.[22-25] Our findings are supported by crystal structures of Dpo4 post-lesion synthesis complexes that show viable reaction parameters for extension past only Bz-dG:dT and Bz-dG:dC base pairs.[23] Furthermore, our simulations show that the structural parameters for dCTP or dTTP binding opposite Bz-dG are within standard deviation of the binding of dCTP opposite dG.

While our simulations support the preferential insertion of dCTP and the misincorporation of dTTP opposite Bz-dG by Dpo4 observed in experimental studies,[22-25] insertion of dATP has also been reported in some experiments.[23] Nevertheless, dATP insertion opposite Bz-dG is not considered to be significant[24] and extension past a Bz-dG:dA base pair does not occur.[23, 24] Our work provides structural evidence that supports difficulties with dATP insertion opposite Bz-dG by Dpo4 due to altered DNA–DNA and DNA–polymerase interactions, but provides no further clarification on the discrepancies reported in these experimental studies. Experiments have also shown that Bz-dG strongly inhibits the enzymatic reaction compared to dG.[23] Our study reveals one factor that may contribute to the reduced catalytic efficiency of Bz-dG compared to dG replication, namely weak hydrogen bonding between the adducted dG and pairing dNTP. Indeed, previous studies have reported that hydrogen bonding between the base being replicated and the incoming dNTP is an essential component of TLS.[25, 56, 57] However, several factors likely play a role in the overall reduction in the catalytic efficiency beyond the structure of the reactant complex investigated in the current study, including potential changes in the nucleotidyl transfer reaction. Therefore, to complement our new explanations for the experimentally-observed preferred Bz-dG

replication outcomes, further computational studies should consider the structure of Bz-dG replication complexes at different stages of Dpo4 replication, including the pre-lesion and post-lesion synthesis complexes. Furthermore, the reaction mechanism should be investigated to unveil the effects of Bz-dG on the structures of the transition states and products, and reaction energetics associated with TLS by Dpo4.

4.4.2. MD simulations of the –1 base deletion complexes provide the first structural evidence for the lack of deletion mutations upon Bz-dG replication: Even though TLS of DNA adducts frequently results in deletion mutations,[20, 39-42, 66-68] Bz-dG replication rarely leads to base deletions.[23, 24] Due to the frequency of deletion mutations upon lesion bypass by Dpo4[67, 68] and a reported intercalated Bz-dG-adducted DNA conformation,[32, 33] which is a known precursor to deletion mutations,[69] the current work aimed to understand why Bz-dG replication by Dpo4 does not cause deletion mutations *in vivo*. Specifically, dGTP was positioned opposite the 5′-dC with respect to (*anti/syn*) Bz-dG, with no base opposite the lesion. MD simulations reveal that dNTP–polymerase interactions are severely distorted (e.g., Arg51, Lys159, Tyr10, Thr45, and Phe11), and neither the reaction distance (> 4.8 Å), the reaction angle (< 103.1°), nor the coordination of the catalytic Mg^{2+} ions (< 3 coordinated atoms) is conducive for dNTP insertion. Therefore, our data suggests that –1 base deletion mutations are unlikely to occur during Bz-dG replication by Dpo4. This is consistent with the lack of –1 base deletion mutations identified in experimental studies.[23, 24] Furthermore, we predict the reaction distance to be at least 1.3 Å greater for the Bz-dG deletion complexes than canonical dG replication, which supports the hypothesis based on experimental kinetic studies that deletion

mutations do not occur upon Bz-dG replication by Dpo4 since the terminal O3′ of the primer strand is too far from the dNTP.[23]

Previous literature on the replication of N2-dG and C8-dG adducts has proposed two mechanisms for the formation of deletion mutations during TLS: 1) the replication of a base (i.e., the adduct or flanking base) is skipped due to adduct intercalation, or 2) the adduct is looped out of the helix in order to translocate past the lesion.[65, 66] When coupled with experimental data for the replication of other adducts, our simulations provide insight into why Bz-dG does not form deletion mutations through either of these previously proposed mechanisms. First, skipping the replication of the adduct commonly occurs for lesions with bulky moieties composed of more than one aromatic ring (e.g., C8-BP-dG, N2-BP-dG, N2-Naph-dG, C8-ABA-dG, and C8-AP-dG),[20, 39-42, 66] which can fill the space of the missing pairing base and thus stabilize a –1 base deletion complex. Conversely, the single ring bulky moiety of the C8-Ph-dG adduct is insufficient to stabilize an intercalated structure and this lesion does not form deletion mutations.[16] Similar to C8-Ph-dG, our simulations suggest that the size and/or position of the Bz-dG bulky moiety in the polymerase active site precludes the lesion from filling the void of the missing base and stabilizing the intercalated precursor to a –1 base deletion mutation. Second, looping the adduct (e.g., N2-BP-dG) out of the helix commonly occurs to translocate past lesions that sterically clash with the polymerase during translocation and consequently leads to –1 base deletion mutations.[70, 71] However, our simulations suggest that the inherent flexibility of Bz-dG allows the adduct to adopt many positions in the Dpo4 active site. Therefore, it is unlikely that looping Bz-dG out of the helix is required for a TLS polymerase to translocate past the adduct. Since the bulky moiety in all O6-dG

adducts reported in the literature to date are small and highly flexible (Figure 4.1), these lesions are unlikely to cause –1 base deletion mutations. Nevertheless, the varying composition of the bulky moieties of O6-dG adducts, including the presence of heteroatoms, may affect discrete interactions of the adducts in the polymerase active site and therefore replication outcomes.

4.5. Conclusions

Computational studies of the TLS of DNA adducts are critical for obtaining structural information about the replication complexes. The present work provides currently missing atomic level details of the ternary insertion and –1 base deletion complexes for Bz-dG replication by Dpo4, and is the first study to outline these features for the replication of an O6-dG adduct. Due to the lack of critical hydrogen-bonding interactions, changes in the active site conformation and/or the lack of reaction-ready parameters, the purines are unlikely to be inserted opposite Bz-dG. Consistent with previous experimental kinetic studies that show dCTP and dTTP insertion opposite Bz-dG,[22-25] the calculated preserved active site conformation compared to natural dG replication suggests the Dpo4 ternary insertion complex is conducive for dCTP or dTTP insertion opposite Bz-dG. We also provide unique insight into the experimentally determined preference for dCTP insertion opposite Bz-dG due to the formation of catalytically competent reactant complexes for dCTP insertion opposite both *anti* and *syn*-Bz-dG, while the insertion of dTTP opposite *syn*-Bz-dG is unlikely due to the lack of DNA–DNA interactions. Finally, although the TLS of DNA adducts has been shown to commonly form deletion mutations,[20, 39-42, 66] these are not a predominant outcome for Bz-dG replication. The current work shows that deletion mutations will not occur upon Bz-dG replication due to

the small size and high degree of flexibility of the Bz-dG bulky moiety in the polymerase active site. Together, the structural insights acquired from our simulations provide a more complete understanding of Bz-dG replication. Furthermore, due to the similarity in previously observed orientations of O6-dG adducts in DNA duplexes, the calculated conformations of Bz-dG in the Dpo4 active site provide a basis for understanding the replication of other O6-dG adducts.

4.6. References

[1] Bernstein, C., and Bernstein, H. (2015) Epigenetic reduction of DNA repair in progression to gastrointestinal cancer, *World J. Gastrointest. Oncol. 7*, 30-46.

[2] Loveless, A. (1969) Possible relevance of o-6 alkylation of deoxyguanosine to the mutagenicity and carcinogenicity of nitrosamines and nitrosamides, *Nature 223*, 206-207.

[3] Singer, B., Chavez, F., Goodman, M., Essigmann, J., and Dosanjh, M. (1989) Effect of 3' flanking neighbors on kinetics of pairing of dCTP or dTTP opposite O6-methylguanine in a defined primed oligonucleotide when escherichia coli DNA polymerase I is used, *Proc. Natl. Acad. Sci. U.S.A. 86*, 8271-8274.

[4] Yuan, J.-M., Koh, W.-P., Murphy, S. E., Fan, Y., Wang, R., Carmella, S. G., Han, S., Wickham, K., Gao, Y.-T., Yu, M. C., and Hecht, S. S. (2009) Urinary levels of tobacco-specific nitrosamine metabolites in relation to lung cancer development in two prospective cohorts of cigarette smokers, *Cancer Research 69*, 2990-2995.

[5] Church, T. R., Anderson, K. E., Caporaso, N. E., Geisser, M. S., Le, C. T., Zhang, Y., Benoit, A. R., Carmella, S. G., and Hecht, S. S. (2009) A prospectively measured serum biomarker for a tobacco-specific carcinogen and lung cancer in smokers, *Cancer Epidemiol. Biomarkers Prev. 18*, 260-266.

[6] Alavanja, M., Baron, J., Brownson, R. C., Buffler, P. A., DeMarini, D. M., Djordjevic, M. V., Doll, R., Fontham, E. T., Gao, Y. T., and Gray, N. (2004) Tobacco smoke and involuntary smoking, *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans 83*, 1-1413.

[7] Hecht, S. S. (1998) Biochemistry, biology, and carcinogenicity of tobacco-specific N-nitrosamines, *Chem. Res. Toxicol. 11*, 559-603.

[8] Hecht, S. S. (2008) Progress and challenges in selected areas of tobacco carcinogenesis, *Chem. Res. Toxicol. 21*, 160-171.

[9] Margison, G. P., Santibáñez Koref, M. F., and Povey, A. C. (2002) Mechanisms of carcinogenicity/chemotherapy by O6-methylguanine, *Mutagenesis 17*, 483-487.

[10] Guarnieri, M., Biser-Rohrbaugh, A., Tyler, B., Gabikian, P., Bunton, T., Wu, Q., Weingart, J., and Carson, B. (2002) Toxicity of intracranial and intraperitoneal O6-benzyl guanine in combination with bcnu delivered locally in a mouse model, *Cancer Chemoth. Pharm. 50*, 392-396.

[11] Sale, J. E., Lehmann, A. R., and Woodgate, R. (2012) Y-family DNA polymerases and their role in tolerance of cellular DNA damage, *Nat. Rev. Mol. Cell Biol. 13*, 141-152.

[12] Ludmann, S., and Marx, A. (2016) Getting it right: How DNA polymerases select the right nucleotide, *Chima 70*, 203-206.

[13] Stover, J. S., Chowdhury, G., Zang, H., Guengerich, F. P., and Rizzo, C. J. (2006) Translesion synthesis past the C8- and N2-deoxyguanosine adducts of the dietary mutagen 2-amino-3-methylimidazo[4,5-f]quinoline in the NarI recognition sequence by prokaryotic DNA polymerases, *Chem. Res. Toxicol. 19*, 1506-1517.

[14] 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.

[15] Sholder, G., and Loechler, E. L. (2015) A method to accurately quantitate intensities of 32P-DNA bands when multiple bands appear in a single lane of a gel is used to study dNTP insertion opposite a benzo[a]pyrene-dG adduct by sulfolobus DNA polymerases Dpo4 and dbh, *DNA Repair 25*, 97-103.

[16] 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.*

[17] 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.

[18] 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: 19F NMR, surface plasmon resonance, and primer kinetic studies, *Biochemistry 53*, 4059-4071.

[19] Kirouac, K. N., and Ling, H. (2009) Structural basis of error-prone replication and stalling at a thymine base by human DNA polymerase ι, *EMBO J. 28*, 1644-1654.

[20] Gadkari, V. V., Tokarsky, E. J., Malik, C. K., Basu, A. K., and Suo, Z. (2014) Mechanistic investigation of the bypass of a bulky aromatic DNA adduct catalyzed by a Y-family DNA polymerase, *DNA Repair 21*, 65-77.

[21] Zhang, H., Eoff, R. L., Kozekov, I. D., Rizzo, C. J., Egli, M., and Guengerich, F. P. (2009) Versatility of Y-family sulfolobus solfataricus DNA polymerase Dpo4 in translesion synthesis past bulky N2-alkylguanine adducts, *J. Biol. Chem. 284*, 3563-3576.

[22] Lim, S., Song, I., Guengerich, F. P., and Choi, J.-Y. (2012) Effects of N2-alkylguanine, O6-alkylguanine, and abasic lesions on DNA binding and bypass synthesis by the euryarchaeal B-family DNA polymerase vent (exo–), *Chem. Res. Toxicol. 25*, 1699-1707.

[23] Eoff, R. L., Angel, K. C., Egli, M., and Guengerich, F. P. (2007) Molecular basis of selectivity of nucleoside triphosphate incorporation opposite O6-benzylguanine by sulfolobus solfataricus DNA polymerase Dpo4: Steady-state and pre-steady-state kinetics and X-ray crystallography of correct and incorrect pairing, *J. Biol. Chem. 282*, 13573-13584.

[24] Choi, J.-Y., Chowdhury, G., Zang, H., Angel, K. C., Vu, C. C., Peterson, L. A., and Guengerich, F. P. (2006) Translesion synthesis across O6-alkylguanine DNA adducts by recombinant human DNA polymerases, *J. Biol. Chem. 281*, 38244-38256.

[25] Gahlon, H. L., Boby, M. L., and Sturla, S. J. (2014) O6-alkylguanine postlesion DNA synthesis is correct with the right complement of hydrogen bonding, *ACS Chem. Biol. 9*, 2807-2814.

[26] Peterson, L. A., Vu, C., Hingerty, B. E., Broyde, S., and Cosman, M. (2003) Solution structure of an O6-[4-oxo-4-(3-pyridyl)butyl]guanine adduct in an 11mer DNA duplex: Evidence for formation of a base triplex, *Biochemistry 42*, 13134-13144.

[27] Jena, N., and Bansal, M. (2011) Mutagenicity associated with O6-methylguanine-DNA damage and mechanism of nucleotide flipping by AGT during repair, *Physical biology 8*, 046007.

[28] Patel, D. J., Shapiro, L., Kozlowski, S. A., Gaffney, B. L., and Jones, R. A. (1986) Structural studies of the o6meg:T interaction in the d(C-G-T-G-A-A-T-T-C-o6meg-C-G) duplex, *Biochemistry 25*, 1036-1042.

[29] Patel, D. J., Shapiro, L., Kozlowski, S. A., Gaffney, B. L., and Jones, R. A. (1986) Structural studies of the o6meg:C interaction in the d(C-G-C-G-A-A-T-T-C-o6meg-C-G) duplex, *Biochemistry 25*, 1027-1036.

[30] Gaffney, B. L., and Jones, R. A. (1989) Thermodynamic comparison of the base pairs formed by the carcinogenic lesion O6-methylguanine with reference both to watson-crick pairs and to mismatched pairs, *Biochemistry 28*, 5881-5889.

[31] Gaffney, B. L., Marky, L. A., and Jones, R. A. (1984) Synthesis and characterization of a set of four dodecadeoxyribonucleoside undecaphosphates containing O6-methylguanine opposite adenine, cytosine, guanine, and thymine, *Biochemistry 23*, 5686-5691.

[32] Wilson, K. A., and Wetmore, S. D. (2014) Complex conformational heterogeneity of the highly flexible O6-benzyl-guanine DNA adduct, *Chem. Res. Toxicol. 27*, 1310-1325.

[33] Kowal, E. A., Lad, R. R., Pallan, P. S., Dhummakupt, E., Wawrzak, Z., Egli, M., Sturla, S. J., and Stone, M. P. (2013) Recognition of O(6)-benzyl-2′-deoxyguanosine by a perimidinone-derived synthetic nucleoside: A DNA interstrand stacking interaction, *Nucleic Acids Res. 41*, 7566-7576.

[34] Eoff, R. L., Irimia, A., Egli, M., and Guengerich, F. P. (2007) Sulfolobus solfataricus DNA polymerase Dpo4 is partially inhibited by "wobble" pairing between O6-methylguanine and cytosine, but accurate bypass is preferred, *J.Biol. Chem. 282*, 1456-1467.

[35] Sriram, M., van der Marel, G. A., Roelen, H. L., van Boom, J. H., and Wang, A. H. (1992) Conformation of B-DNA containing O6-ethyl-G-C base pairs stabilized by minor groove binding drugs: Molecular structure of d(CGC[e6g]AATTCGCG complexed with hoechst 33258 or hoechst 33342, *The EMBO Journal 11*, 225-232.

[36] Sriram, M., Van der Marel, G. A., Roelen, H. L. P. F., Van Boom, J. H., and Wang, A. H. J. (1992) Structural consequences of a carcinogenic alkylation lesion on DNA: Effect of O6-ethylguanine [e6g] on the molecular structure of d(CGC[e6g]AATTCGCG)-netropsin complex, *Biochemistry 31*, 11823-11834.

[37] Warren, J. J., Forsberg, L. J., and Beese, L. S. (2006) The structural basis for the mutagenicity of O6-methyl-guanine lesions, *Proc. Natl. Acad. Sci. U.S.A. 103*, 19701-19706.

[38] Choi, J.-Y., and Guengerich, F. P. (2006) Kinetic evidence for inefficient and error-prone bypass across bulky N2-guanine DNA adducts by human DNA polymerase ι, *J. Biol. Chem. 281*, 12315-12324.

[39] Xu, P., Oum, L., Lee, Y.-C., Geacintov, N. E., and Broyde, S. (2009) Visualizing sequence-governed nucleotide selectivities and mutagenic consequences through a replicative cycle: Processing of a bulky carcinogen N2-dG lesion in a y-family DNA polymerase, *Biochemistry 48*, 4677-4690.

[40] Kirouac, K. N., Basu, A. K., and Ling, H. (2013) Structural mechanism of replication stalling on a bulky amino-polycyclic aromatic hydrocarbon DNA adduct by a y family DNA polymerase, *J. Mol. Biol. 425*, 4167-4176.

[41] Guengerich, F. P., Zhao, L., Pence, M. G., and Egli, M. (2015) Structure and function of the translesion DNA polymerases and interactions with damaged DNA, *Perspect. Sci. 4*, 24-31.

[42] Xu, P., Oum, L., Geacintov, N. E., and Broyde, S. (2008) Nucleotide selectivity opposite a benzo[a]pyrene-derived N2-dG adduct in a y-family DNA polymerase: A 5′-slippage mechanism, *Biochemistry 47*, 2701-2709.

[43] Gahlon, H. L., Schweizer, W. B., and Sturla, S. J. (2013) Tolerance of base pair size and shape in postlesion DNA synthesis, *J. Am. Chem. Soc. 135*, 6384-6387.

[44] Walsh, J. M., Ippoliti, P. J., Ronayne, E. A., Rozners, E., and Beuning, P. J. (2013) Discrimination against major groove adducts by Y-family polymerases of the dinb subfamily, *DNA Repair 12*, 713-722.

[45] Kottur, J., Sharma, A., Gore, Kiran R., Narayanan, N., Samanta, B., Pradeepkumar, P. I., and Nair, Deepak T. (2015) Unique structural features in DNA polymerase IV enable efficient bypass of the N2 adduct induced by the nitrofurazone antibiotic, *Structure 23*, 56-67.

[46] Ling, H., Sayer, J. M., Plosky, B. S., Yagi, H., Boudsocq, F., Woodgate, R., Jerina, D. M., and Yang, W. (2004) Crystal structure of a benzo[a]pyrene diol epoxide adduct in a ternary complex with a DNA polymerase, *Proc. Natl. Acad. Sci. U.S.A. 101*, 2265-2269.

[47] Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., and Shaw, D. E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field, *Proteins: Structure, Function, and Bioinformatics 78*, 1950-1958.

[48] Zhang, L., Rechkoblit, O., Wang, L., Patel, D. J., Shapiro, R., and Broyde, S. (2006) 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. 34*, 3326-3337.

[49] Perlow, R. A., and Broyde, S. (2002) Toward understanding the mutagenicity of an environmental carcinogen: Structural insights into nucleotide incorporation preferences, *J. Mol. Biol. 322*, 291-309.

[50] Zhang, L., Shapiro, R., and Broyde, S. (2005) Molecular dynamics of a food carcinogen−DNA adduct in a replicative DNA polymerase suggest hindered nucleotide incorporation and extension, *Chem. Res. Toxicol. 18*, 1347-1363.

[51] Case, D. A., Babin, V., Berryman, J. T., Betz, R. M., Cai, Q., Cerutti, D. S., T.E. Cheatham, I., 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.

[52] Case, D. A., Darden, T. A., Cheatham, T. E., III, Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Crowley, M., Walker, R. C., Zhang, W., Merz, K. M., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossvary, I., Wong, K. F., Paesani, F., Vanicek, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Yang, L., Tan, C., Mongan, J., Hornak, V., Cui, G., Mathews, D. H., Seetin, M. G., Sagui, C., Babin, V., and Kollman, P. A. (2008) AMBER tools, Version 1.0 ed., University of California, San Francisco.

[53] Case, D. A., Darden, T. A., III, T. E. C., Simmerling, C. L., Wang, J., R.E. Duke, R. L., Walker, R. C., Zhang, W., Merz, K. M., Roberts, B., Hayik, S., Roitberg, A., Seabra, G., Swails, J., Goetz, A. W., Kolossváry, I., Wong, K. F., Paesani, F., Vanicek, J., Wolf, R. M., Liu, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M.-J., Cui, G., Roe, D. R., Mathews, D. H., Seetin, M. G., Salomon-Ferrer, R., Sagui, C., Babin, V., Luchko, T., Gusarov, S., Kovalenko, A., and Kollman, P. A. (2012) AMBER 12, University of California, San Francisco.

[54] Wang, Y., and Schlick, T. (2008) Quantum mechanics/molecular mechanics investigation of the chemical reaction in Dpo4 reveals water-dependent pathways and requirements for active site reorganization, *J. Am. Chem. Soc. 130*, 13240-13250.

[55] Wang, L., Yu, X., Hu, P., Broyde, S., and Zhang, Y. (2007) A water-mediated and substrate-assisted catalytic mechanism for sulfolobus solfataricus DNA polymerase IV, *J. Am. Chem. Soc. 129*, 4731-4737.

[56] Washington, M. T., Helquist, S. A., Kool, E. T., Prakash, L., and Prakash, S. (2003) Requirement of watson-crick hydrogen bonding for DNA synthesis by yeast DNA polymerase η, *Mol. Cell. Biol. 23*, 5107-5112.

[57] Wolfle, W. T., Washington, M. T., Kool, E. T., Spratt, T. E., Helquist, S. A., Prakash, L., and Prakash, S. (2005) Evidence for a watson-crick hydrogen bonding requirement in DNA synthesis by human DNA polymerase κ, *Mol. Cell. Biol. 25*, 7137-7143.

[58] Ling, H., Boudsocq, F., Woodgate, R., and Yang, W. (2001) Crystal structure of a y-family DNA polymerase in action: A mechanism for error-prone and lesion-bypass replication, *Cell 107*, 91-102.

[59] Walsh, J. M., Parasuram, R., Rajput, P. R., Rozners, E., Ondrechen, M. J., and Beuning, P. J. (2012) Effects of non-catalytic, distal amino acid residues on activity of E. Coli dinb (DNA polymerase IV), *Environ. Mol. Mutagen. 53*, 766-776.

[60] Jarosz, D. F., Godoy, V. G., Delaney, J. C., Essigmann, J. M., and Walker, G. C. (2006) A single amino acid governs enhanced activity of dinb DNA polymerases on damaged templates, *Nature 439*, 225-228.

[61] Wang, L., and Broyde, S. (2006) A new anti conformation for N-(deoxyguanosin-8-yl)-2-acetylaminofluorene (AAF-dG) allows watson–crick pairing in the sulfolobus solfataricus P2 DNA polymerase IV (Dpo4), *Nucleic Acids Res. 34*, 785-795.

[62] Sassa, A., Niimi, N., Fujimoto, H., Katafuchi, A., Grúz, P., Yasui, M., Gupta, R. C., Johnson, F., Ohta, T., and Nohmi, T. (2011) Phenylalanine 171 is a molecular brake for translesion synthesis across benzo[a]pyrene-guanine adducts by human DNA polymerase kappa, *Mutat. Res.-Gen. Tox. En. 718*, 10-17.

[63] Wang, W., Wu, E. Y., Hellinga, H. W., and Beese, L. S. (2012) Structural factors that determine selectivity of a high fidelity DNA polymerase for deoxy-, dideoxy-, and ribonucleotides, *J.Biol. Chem. 287*, 28215-28226.

[64] Brown, J. A., and Suo, Z. (2011) Unlocking the sugar "steric gate" of DNA polymerases, *Biochemistry 50*, 1135-1142.

[65] Brautigam, C. A., and Steitz, T. A. (1998) Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes, *Curr. Opin. Struct. Biol. 8*, 54-63.

[66] Zhang, X., Lee, I., Zhou, X., and Berdis, A. J. (2006) Hydrophobicity, shape, and π-electron contributions during translesion DNA synthesis, *Journal of the American Chemical Society 128*, 143-149.

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

[68] McCulloch, S. D., and Kunkel, T. A. (2008) The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases, *Cell research 18*, 148-161.

[69] Mao, B., Cosman, M., Hingerty, B. E., Broyde, S., and Patel, D. J. (1995) Solution conformation of [AF] dG opposite a-1 deletion site in a DNA duplex: Intercalation of the covalently attached aminofluorene ring into the helix with base displacement of the C8-modified syn guanine into the major groove, *Biochemistry 34*, 6226-6238.

[70] Schneider, S., Schorr, S., and Carell, T. (2009) Crystal structure analysis of DNA lesion repair and tolerance mechanisms, *Curr. Opin. Struct. Biol. 19*, 87-95.

[71] Jha, V., Bian, C., Xing, G., and Ling, H. (2016) Structure and mechanism of error-free replication past the major benzo[a]pyrene adduct by human DNA polymerase κ, *Nucleic Acids Res.*

Chapter 5: Computational Insights into the Mutagenicity of Two Tobacco Derived Carcinogenic DNA Lesions[a,b,c]

5.1. Introduction

Nitrosamines are a large group of compounds that occur in the human diet, cosmetics, and tobacco, as well as flexible plastics such as balloons, condoms, and baby bottle nipples (1-3). Despite the high prevalence of nitrosamines in contemporary society, these compounds are very carcinogenic, with 90% of the 300 identified nitrosamines being known carcinogens (4). The nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) is formed during the curing of tobacco and has been identified in all tobacco products, including those used in conventional cigarettes, electronic cigarettes, and Hookahs (5-7). NNK is also the only tobacco component that has led to lung cancer in every species tested (i.e., mice, rats, hamsters, rabbits, pigs, monkeys, and humans) regardless of the route of administration (8,9). Therefore, the World Health Organization's International Agency for Research on Cancer has classified NNK as carcinogenic to humans (Group 1) (10). Nevertheless, changes in cigarette design since the 1950s and new routes of tobacco administration have increased human exposure to NNK, and consequently led to an

---

[a] *Nucleic Acids Research* reference style used throughout this chapter.

[b] Adapted from Wilson, K. A., Garden, J.L., Wetmore, N.T., and Wetmore, S.D. (2018). Computational insights into the mutagenicity of two tobacco derived carcinogenic DNA lesions. *Nucleic. Acids Res.* (DOI: 10.1093/nar/gky1071).

[c] J.L.G. performed calculations on the PHB-G nucleobase model and N.T.W. performed preliminary calculations on the PHB-G base pair model. K.A.W. performed calculations on the nucleobase, base pair, DNA and polymerase models, as well as data analysis on all models and wrote the first draft of the manuscript.

increase in the number of cases of adenocarcinoma (a form of lung cancer) (11,12). Indeed, lung cancer is the most common cancer caused by NNK and is also the leading cause of cancer death worldwide in part due to a survival rate of only 18% (13).

In cells, NNK can be reduced to 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol (NNAL), and cytochrome p450 enzymes convert both NNK and NNAL into species that react with DNA (5). The complex metabolism of NNK and NNAL gives rise to a variety of DNA lesions including methyl, formaldehyde, pyridyloxobutyl (POB), and pyridylhydroxybutyl (PHB) DNA adducts (5). The current study focuses on the larger POB-G and PHB-G lesions that result from attack at the O6 position of G (Figure 5.1). Both POB-G and PHB-G have a modified Watson-Crick hydrogen-bonding face compared to G (i.e., N1 becomes a hydrogen-bond acceptor; Figure 5.1). Therefore, this may cause these G lesions to no longer exhibit a preference for C, which could result in mutations upon replication. POB-G and PHB-G differ in the substitution of the chain connecting the adducted G to the bulky moiety ring (i.e., carbonyl versus hydroxyl group, respectively) and previous work on DNA adducts arising from known human carcinogens has shown that lesion mutagenicity is affected by small changes in chemical composition including linker substituents, ring substitution, and ionization state (14-25). Thus, POB-G and PHB-G may have different mutagenic profiles, and their consideration may afford a broader understanding of the effect of bulky moiety linker composition on lesion mutagenicity.

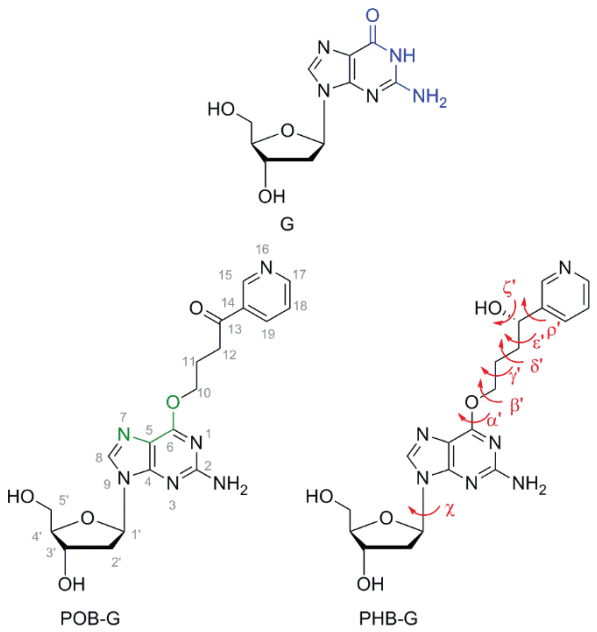

Figure 5.1. The G, POB-G and PHB-G nucleosides in the *anti* glycosidic conformation, with the Watson-Crick hydrogen-bonding face highlighted in blue and the Hoogsteen face in green. Key dihedral angles are defined for the lesions including χ [∠(O4′C1′N9C4)], α′ [∠(C5C6O6C10)], β′ [∠(C6O6C10C11)], γ′ [∠(O6C10C11C12)], δ′ [∠(C10C11C12C13)], ϵ′ [∠(C11C12C13C14)], and ρ′ [∠(C12C13C14C15)], as well as ζ′ for PHB-G [∠(C12C13O13H13)].

Detailed investigations have shown that POB-G is highly mutagenic (21,26,27). Specifically, POB-G has a 96% mutation rate in *E. coli* cells, exclusively leading to G → A transition mutations (26). The mutational spectrum in human HEK293 cells is slightly more complex, with the primary replication outcome still being G → A transition mutations (54%), but G → T mutations also occurring at a low frequency (2%) (26). *In vitro* kinetics studies support these reported mutational spectra, indicating that translesion synthesis (TLS) polymerases preferentially insert dCTP and dTTP opposite the lesion (21,27). Although the kinetic studies also report dGTP insertion, extension past the POB-G:G base pair does not occur (21,27). Furthermore, while polymerase η bypasses POB-G with a k_{cat}/K_m of 0.63 mM^{-1} s^{-1}, polymerase ι, κ, and ν bypass POB-G with a k_{cat}/K_m of < 0.005 mM^{-1} s^{-1}, which indicates that polymerase η likely plays a critical role in POB-G replication

(21,27). Nevertheless, POB-G replication is significantly slower compared to that of natural DNA, which has a k_{cat}/K_m of 160 mM^{-1} s^{-1} for the insertion of dCTP opposite G by polymerase η (21). While the mutagenicity of PHB-G is currently unknown, the NNAL parent compound that primarily forms PHB lesions has been reported to be tumorigenic in mice (28).

To complement mutagenicity data, structural studies have been performed on POB-G using a variety of models (29-31). Molecular mechanics (MM) minimization of the 5′–AT[POB-G] 2′-deoxytrinucleoside diphosphate revealed that the bulky moiety can be directed either 3′ or 5′ with respect to the lesion site, or extended in the same plane as the adducted G (29). Classical molecular dynamics (MD) simulations on POB-G adducted DNA (29) and NMR-based restrained MM minimizations (30) found that the bulky moiety is preferentially extended in the same plane as the adducted G. Nevertheless, while only one position of the POB-G bulky moiety was reported in the NMR structure, the inherent flexibility of the lesion is reflected in broad NMR signals (30). This correlates with computational work on the POB-G nucleobase (31) and 5′–AT[POB-G] 2′-deoxytrinucleoside diphosphate models (29), which indicate POB-G has a high degree of conformational flexibility. In terms of base pairing, POB-G has been shown both in computational and experimental work to form a wobble pair opposite C (29-31). Previous computational work from our group provided the first structural data for understanding the observed POB-G mutagenicity by characterizing nucleobase base dimers (31). Specifically, a stable, non-distorted pseudo-Watson-Crick POB-G:T pair, and marginally distorted pairs between the Watson-Crick face of POB-G and A or C were identified (31). Furthermore, a predicted stable, but highly-distorted, Hoogsteen POB-G:G pair was found (31), which correlates with the

experimentally-observed insertion of dGTP opposite POB-G, but the lack of subsequent extension (21,27). Overall, these previous studies have provided insight into the flexibility and base-pairing propensity of POB-G. Furthermore, these works highlight the critical role computational studies can play in unveiling how lesion structural properties relate to mutagenicity. Nevertheless, structural details of the adduct mispairs within the helical environment, as well as pairs within the polymerase active site, are key for understanding lesion mutagenicity, but remain unavailable. Additionally, no structural studies have clarified the similarities and/or differences in the flexibility and base-pairing preferences of PHB-G with respect to POB-G.

To shed further light on the mutagenicity of POB-G and provide the first clues regarding the mispairing potential of PHB-G, the present work uses quantum chemical calculations and molecular dynamics simulation methods to characterize the intrinsic conformational and hydrogen-bonding preferences of POB-G and PHB-G, as well as the corresponding properties in a DNA duplex and human TLS polymerase active site. Specifically, DFT nucleobase and nucleoside models are initially used to examine the energy penalty for rotation about key bonds in the bulky moiety and at the glycosidic linkage, which uncovers the accessible conformations of free POB-G and PHB-G. Subsequently, 28 hydrogen-bonded pairs for each lesion were characterized using DFT nucleobase models to understand possible interactions between different lesion orientations and the canonical DNA nucleobases. MD simulations on adducted DNA provide insight into the dynamic conformational landscape of POB-G and PHB-G, and the effects of the identity of the pairing base on the structure of post-replication duplexes. Finally, MD simulations reveal the POB-G and PHB-G conformational flexibility and base pairing within the confines of

the polymerase η active site, and thereby shed light on the relative propensity for the insertion of various dNTPs. Overall, our multiscaled computational approach provides structural insight into the mispairing potential of two key tobacco-derived DNA lesions and complements a growing body of literature focused on understanding how small changes in the chemical composition of adducts affect mutagenicity (14-25).

5.2. Computational Details

An AMBER conformational search about key rotatable bonds in the POB-G and PHB-G nucleobases (α′, β′, γ′, δ′, ϵ′, ρ′, and ζ′, Figure 5.1) as implemented in Hyperchem was used to determine the inherent conformational preference about the nucleobase–bulky moiety linker and within the bulky moiety. All orientations of POB-G and PHB-G isolated from the conformational search were subsequently optimized using density functional theory (DFT), specifically B3LYP-D3(BJ)/6-31G(d) and the relative energies were determined using B3LYP-D3/6-311+G(2df,2p). The structures thus obtained were grouped into conformational categories based on the orientation of the bulky moiety with respect to the adducted G. Subsequently, to assess the relative stability of the *anti* ($\chi \approx 220°$) and *syn* ($\chi \approx 60°$) conformations about the glycosidic bond in the POB-G and PHB-G nucleosides, 2′-deoxyribose was added to the most stable orientation from each conformational category of the nucleobase adducts. Finally, hydrogen-bonded nucleobase dimers between the Watson-Crick or Hoogsteen face of the most stable orientations of POB-G and PHB-G from each conformational category, and each of the four canonical DNA bases were investigated. These steps all involved B3LYP-D3(BJ)/6-31G(d) or M06-2X/6-31G(d) optimizations, and B3LYP-D3(BJ)/6-311+G(2df,2p) energy calculations using Gaussian 09 (revision D.01).

MD simulations (AMBER14SB) were performed on 5′–CTCGGCG*CCATC 12-mer DNA duplexes, with G*=POB-G or PHB-G in orientations that represent each conformational category and paired opposite C. Using the insight gained from these simulations about possible POB-G or PHB-G orientations in DNA, MD simulations were performed on duplexes containing the lesions mispaired against T, A, and G. Finally, MD simulations for dCTP, dTTP and dATP insertion opposite POB-G or PHB-G by polymerase η were initiated from a crystal structure of the insertion complex for dATP incorporation opposite T (PDB ID: 4ECS). DNA systems were neutralized and solvated such that 8 Å of TIP3P water surround the duplex, while polymerase systems were solvated such that 10 Å of TIP3P water surround the solute and the concentration of NaCl is ~0.150 M. All systems were minimized, heated and equilibrated. Subsequently, 20 ns pre-production simulations were performed to understand the inherent conformational dynamics at the lesion site. From these trial simulations, representative structures of unique conformations that cannot be easily converted into another conformation through standard MD sampling were chosen as starting points for further simulations. To enhance sampling of the accessible lesion conformations, this process was continued iteratively until no new structures were obtained. Based on these sampling simulations, the lesion orientation was used to select structures for final 300 (DNA) or 100 (polymerase) ns production simulations. Two additional 100 ns replicas were performed for each system (using different initial velocities) to ensure statistically relevant results were obtained. Due to negligible all-atom rmsds between each replica (generally < 2.5 Å; Table D.1, Appendix D), data from one replica will be discussed throughout the main text. The representative structure for each MD simulation shown in the figures was obtained by clustering the trajectory based on the rmsd of the damaged base

pair using the average linkage algorithm; however, all analysis was performed over the entire trajectory.

Full details of the computational protocol are provided in the Supplementary Data.

5.3. Results and Discussion

5.3.1. POB-G and PHB-G have a high degree of inherent conformational flexibility:

To understand the inherently preferred orientations of POB-G and PHB-G, the conformational landscape of both lesions is investigated using DFT nucleobase and nucleoside models. Consistent with the low rotational barriers previously reported for POB-G (31), both lesions are extremely flexible, with 302 and 710 unique conformations of POB-G and PHB-G identified, respectively (Figure 5.2a). The greater number of structures for PHB-G compared to POB-G arises due to multiple orientations of the hydroxy moiety. All structures were visually inspected and classified based on discrete interactions between the bulky moiety and adducted G as stacked, hydrogen bonded, T-shaped or extended (no interactions; Figure 5.2b). The majority of POB-G (62%) and PHB-G (52%) conformations do not contain an interaction between the bulky moiety and the adducted G (extended, Figure D.1, Appendix D). The relative distribution of the remaining conformations varies between the two adducts, although the separation between the categories is not as clear for PHB-G since some stacked conformations also involve hydrogen bonding between the bulky moiety hydroxy group and the adducted G. Most importantly, regardless of the lesion, substantial variation occurs within each conformational category (Figure 5.2a), which further highlights lesion flexibility.

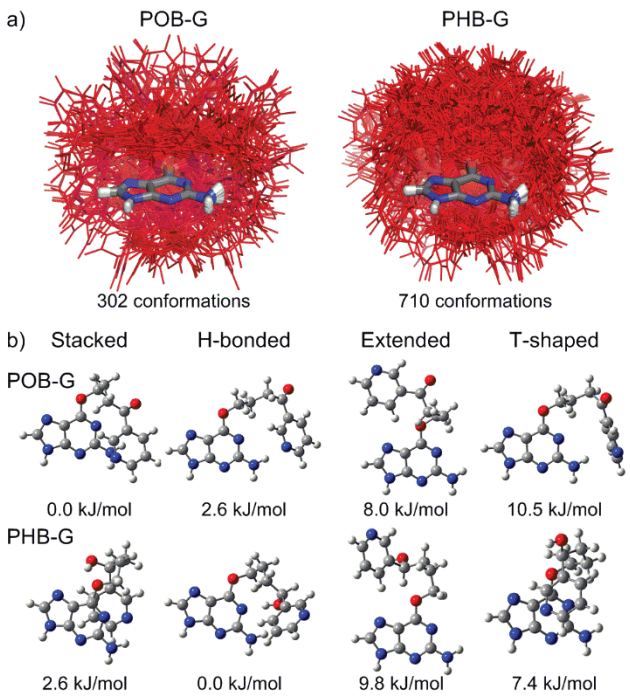

Figure 5.2. a) Overlay (based on G ring atoms) of POB-G (left) and PHB-G (right) conformations identified using DFT, with the bulky moiety highlighted in red. b) Most stable POB-G (top) and PHB-G (bottom) structure identified for each conformational category (relative energies provided). POB-G nucleobase geometries adapted from reference 31.

The most stable POB-G conformation is stacked, while the most stable hydrogen bonded, extended and T-shaped orientations are 2.6, 8.0 and 10.5 kJ/mol less favorable, respectively (Figure 5.2b, top). Conversely, the most stable PHB-G conformation is hydrogen bonded, while the most stable stacked, T-shaped and extended orientations are 2.6, 7.4 and 9.8 kJ/mol less favorable, respectively (Figure 5.2b, bottom). Nevertheless, the PHB-G conformations span a larger energy range than those for POB-G (by 15 kJ/mol; Figure D.2, Appendix D). Upon adding 2′-deoxyribose to the most stable POB-G and PHB-G nucleobase conformations, the bulky moiety orientations and the trends in relative stability are maintained (Figure D.3 and D.4, Appendix D). Additionally, the *anti* orientation is consistently ~ 15 kJ/mol more stable than the *syn* conformation, which indicates that both glycosidic orientations of the lesions may be energetically accessible. Overall, these

calculations on the damaged nucleobases and nucleosides reveal that both POB-G and PHB-G display a high degree of inherent conformational flexibility, adopting several conformational themes that may have unique biological implications.

5.3.2. The Watson-Crick face of POB-G and PHB-G forms stable and minimally distorted base pairs with C, T, and A, while the Hoogsteen face of POB-G and PHB-G forms stable but distorted base pairs with G: To understand the effects of the POB-G and PHB-G conformational heterogeneity on the lesion base-pairing potential, DFT is used to evaluate the geometry and strength of interactions between the Watson-Crick or Hoogsteen face of the most stable structure in each lesion conformational category and the canonical nucleobases. Additionally, since an NMR structure of POB-G adducted DNA predicted the bulky moiety to be projected straight into the major groove (30), the hydrogen-bonding potential of fully extended POB-G and PHB-G (i.e., all dihedral angles in the bulky moiety set to 180°) is also considered. Since neither the POB nor PHB bulky moiety orientation affects the trends in stability of the resulting base pairs (Table D.2 and Figures D.3-D.12, Appendix D), only the fully extended conformation will be discussed for simplicity. Nevertheless, the bulky moiety orientation may lead to additional interactions between the lesion and pairing base (e.g., hydrogen bonding between the bulky moiety and opposing base; Figures D.3-D.12, Appendix D), which may be important and will be discussed in more detail in the helical and polymerase environments.

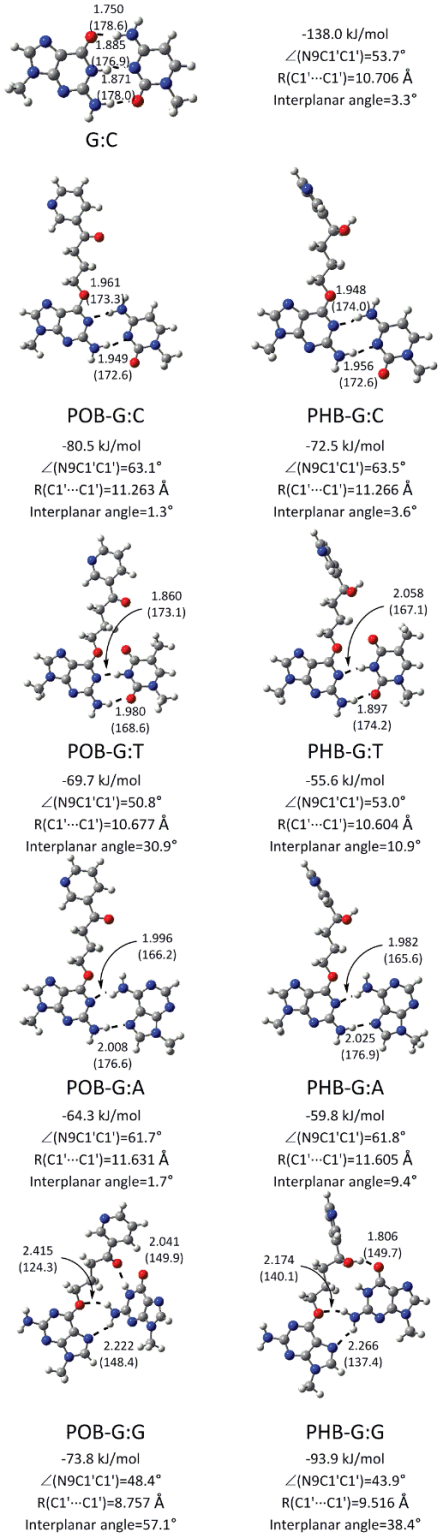

Figure 5.3. DFT structures (distances in Å and angles in deg.) and binding energies (kJ/mol) for the most stable hydrogen-bonded pairs between the extended POB-G (left) or PHB-G (right) conformation and each canonical nucleobase. Figures D.3-D.12 contain data for different lesion orientations. POB-G hydrogen-bonding dimer geometries adapted from reference 31.

Both POB-G and PHB-G form a wobble base pair with C using the Watson-Crick face of the lesion (Figure 5.3), which is consistent with previous studies on POB-G (29-31). This pair contains two hydrogen bonds and is weaker than the canonical G:C pair by ~ 58–65 kJ/mol. Both wobble pairs are wider than G:C by ~0.6 Å and the base pair opening is ~10° larger. Nevertheless, the base pairs are nearly planar (interplanar angle=1–4°). Regardless of the bulky moiety, adducted G forms a pseudo-Watson-Crick base pair with T (–69.7 or –55.6 kJ/mol for POB-G or PHB-G, respectively; Figure 5.3), which contains two hydrogen bonds. Although the G*:T mispairs slightly deviate from planarity (by 10–30°), both base pairs maintain the opening and width of natural DNA (50–54° and 10.6–10.7 Å, respectively), and therefore are overall less distorted than G*:C. Two hydrogen bonds are formed between the lesion Watson-Crick face and the Hoogsteen face of A (–64.3 or –59.8 kJ/mol for POB-G or PHB-G, respectively; Figure 5.3). The base pair opening and width of the G*:A pairs deviate from those of natural DNA (by ~5° and ~0.8 Å, respectively). The base pair between the Watson-Crick face of G* and the Hoogsteen face of G is weak (–38.1 or –29.8 kJ/mol, respectively; Table D.2, Appendix D) and significantly distorted (i.e., deviations up to ~2 Å and 55° in the base pair width and interplanar angles, respectively, Figures D.11 and D.12, Appendix D).

Several other DNA adducts use their Hoogsteen face to pair with the canonical nucleobases (18,32,33). In the case of POB-G and PHB-G, the strongest interactions with the lesion Hoogsteen face occur with G (–73.8 and –93.9 kJ/mol for POB-G and PHB-G, respectively, Figure 5.3). Nevertheless, these base pairs deviate from the natural G:C pair in the opening angle by up to ~25° and base pair width by up to ~2 Å, and the pairs are nonplanar by up

to ~57° (Figure 5.3). The hydrogen-bonding interactions between the Hoogsteen face of POB-G or PHB-G, and the Watson-Crick face of C, T or A are weak (–43 to –52 kJ/mol, Table D.2 and Figures D.3-D.10, Appendix D), and the pairs are structural distorted (i.e., deviations up to ~3 Å and 20° in the base pair width and opening angles, respectively, Figures D.3-D.8, Appendix D).

Overall, the Watson-Crick faces of POB-G and PHB-G form similar hydrogen-bonding interactions with the canonical nucleobases. In particular, stable pairs are characterized between the lesions and C, T, and A, with the T pairs being the least distorted. Conversely, the most stable base pairs with the Hoogsteen faces of the lesions occur with G. However, the high degree of distortion predicted for the isolated G*:G mispairs indicates that either the distorted base pairing geometry will not be adopted in the duplex environment or the pair will cause significant structural changes to the overall helical geometry.

5.3.3. Both lesions preferably extended the bulky moiety into the major groove of DNA, while PHB-G may also adopt an intercalated conformation: Since each lesion conformational category identified for the adducted nucleobases will result in a different bulky moiety position in DNA duplexes, MD simulations were initiated from each conformational theme for POB-G or PHB-G paired opposite C to better understand the accessible lesion orientations in DNA and the resulting impacts on DNA structure. Unsurprisingly, the bulky and intrinsically unstable T-shaped nucleobase orientation (Figure 5.2) does not fit within the confines of the duplex environment. Incorporation of the stacked POB-G or PHB-G conformation into DNA places the bulky moiety between the adducted base pair and the 5′ or 3′ flanking base pair. Regardless of the initial 3′ or 5′

stacked orientation, the bulky moiety consistently adopted an extended orientation within the first 1 ns of an unconstrained MD simulation (Figure D.13, Appendix D), which suggests the inherently stable stacked conformation of the damaged nucleobases (Figure 5.2) cannot be accommodated in the duplex. When the lesion hydrogen-bonded conformation is considered, the bulky moiety is situated in the location of the pairing base, forcing the opposing C into an extrahelical position. Although this conformation could be stabilized through stacking interactions between the bulky moiety and flanking base pairs, as well as hydrogen bonding between the bulky moiety and adducted G, POB-G acquires an extended conformation with the bulky moiety in the major groove and stacked with the 3′-G upon simulation, regardless of the initial starting point (Figure 5.4, left). This suggests that the POB-G hydrogen-bonded conformation is not stable in the duplex. Conversely, PHB-G remains intercalated throughout the simulation (Figure 5.4, right). In addition to bulky moiety stacking interactions with the flanking pairs and hydrogen-bonding interactions with the adducted G, the intercalated PHB-G position is stabilized through discrete hydrogen-bonding interactions between the bulky moiety hydroxy group and the 5′-G (N7, 46% or O6, 10%; Figure 5.4, right), which suggests the stability of this conformer is likely sequence dependent.

When the dynamics of the extended POB-G or PHB-G orientation is considered with the bulky moiety in the major groove, many lesion conformations occur in which the bulky moiety does not interact with the surrounding DNA (Figures 5.5 and 5.6). This directly correlates with the extended conformation being the inherently most common orientation for the isolated nucleobase (Figure 5.2). Nevertheless, in some of the sampled adducted DNA conformations, the bulky moiety forms a T-shaped interaction with the Hoogsteen

edge of the adducted G or a flanking base, or a hydrogen bond with the pairing C (N4H, ~20%), which further highlights the inherent lesion flexibility and the predicted similar stability of many nucleobase conformational themes. As predicted by DFT base pair models, both POB-G and PHB-G form a wobble pair with C that contains G*(N1)···C(N4H) and G*(N2H)···C(N3) hydrogen bonds (each occupied for > 87% of the simulation). The pairs are ~0.6 Å wider and the opening angle is ~10° larger than for canonical DNA (Figure 5.5c and 5.6c and Figures D.14 and D.15, Appendix D). The stability of POB-G:C (–61.8 kJ/mol) is less than PHB-G:C (–68.9 kJ/mol; Table D.2, Appendix D), as well as canonical G:C. Other global structural feature of the damaged DNA are consistent with those expected for natural DNA throughout the simulation, including the maintenance of the flanking base pairs and base step parameters (Figures D.14-D.15, Appendix D).

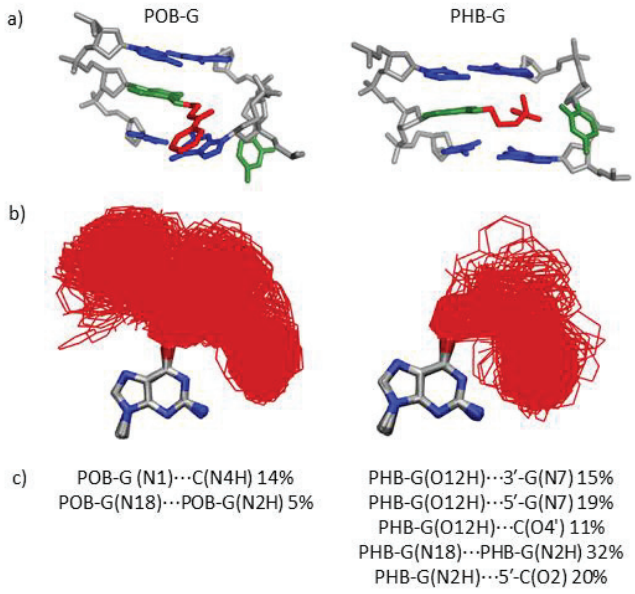

Figure 5.4. a) MD representative structures obtained for POB-G (left) and PHB-G (right) adducted DNA based on an initial hydrogen-bonded orientation of the lesion with the pairing C in the major groove. b) Overlay of lesion conformations adopted throughout the MD simulation, highlighting the deviation in bulky moiety orientation (red). c) Lesion hydrogen-bonding interactions within DNA.

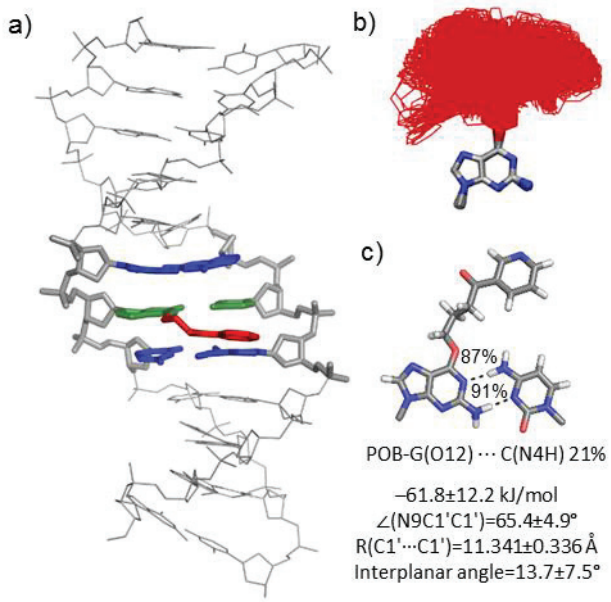

Figure 5.5. a) MD representative structure of POB-G adducted DNA with the lesion in the extended orientation opposite C. b) Overlay of lesion conformations adopted throughout the MD simulation, highlighting the deviation in bulky moiety orientation (red). c) Lesion hydrogen-bonding interactions within DNA.

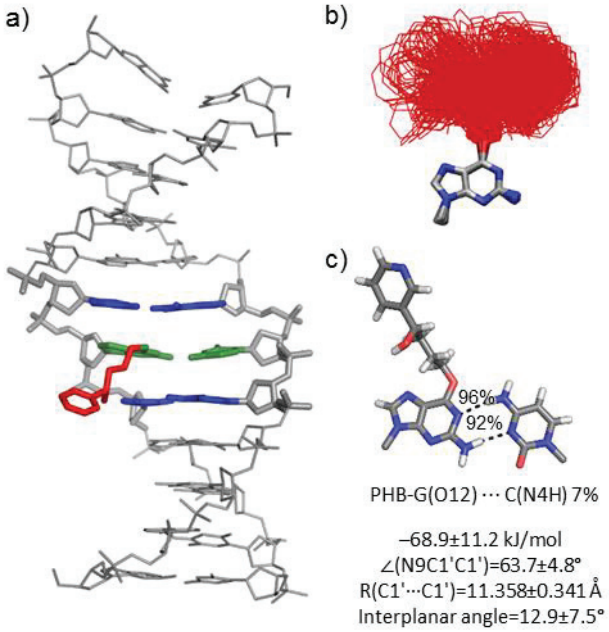

Figure 5.6. a) MD representative structure of PHB-G adducted DNA with the lesion in the extended orientation opposite C. b) Overlay of lesion conformations adopted throughout the MD simulation, highlighting the deviation in bulky moiety orientation (red). c) Lesion hydrogen-bonding interactions within DNA.

Overall, POB-G adducted DNA adopts one dominant conformational theme that positions the bulky moiety in the major groove and the damaged G hydrogen bonds with the opposing base, which correlates with the single structure of POB-G:C adducted DNA previously reported based on NMR (30). Nevertheless, the current work provides insight into the diversity of bulky moiety orientations, which is consistent with reported broad NMR peaks (30). In contrast, PHB-G adducted DNA can adopt two conformations: 1) the bulky moiety intercalates into the helix and the pairing base is displaced into the major groove, and 2) the bulky moiety is in the major groove and the adducted G hydrogen bonds with the opposing base. This difference in the conformational heterogeneity of POB-G and PHB-G damaged DNA highlights the potential influence of the bulky moiety composition on mutagenic outcomes. Indeed, it has been proposed that DNA lesions with the ability to stabilize intercalated conformations have a propensity to form deletion mutations (18,32,33). Therefore, our calculations predict that PHB-G may lead to deletion mutations, while the instability of the intercalated conformation for POB-G correlates with the absence of deletion mutations in the reported mutagenic profile for this lesion (21,27). Regardless, both POB-G and PHB-G can maintain hydrogen bonding with an opposing C such that there are only minor distortions to the DNA structure, which reflects the importance of this non-mutagenic pairing in the broader biological context and is consistent with the experimentally-observed non-mutagenic replication of POB-G (21,27).

5.3.4. Both POB-G and PHB-G form stable interactions with T and A in the DNA duplex: To probe the dynamical structure of post-replication duplexes that correspond to mutagenic events, the POB-G and PHB-G mispairs with T, A, and G were modeled in

DNA. When POB-G or PHB-G is paired opposite T, a pseudo-Watson-Crick pair is maintained that contains G*(N2H)···T(O2) and G*(N1)···T(N3H) hydrogen bonds (–41.7 or –55.5 kJ/mol for POB-G or PHB-G, respectively; Figure 5.7a and 5.8a and Table D.3, Appendix D). The G*(N1)···T(N3H) interaction is less persistent for POB-G (33%) than PHB-G (63%) due to the orientation of the bulky moiety at the site of attachment (i.e., C10–H is directed toward the POB-G Watson-Crick face to allow for T-shaped interaction between the pyridyl ring with the flanking base pairs, but C10–H is directed toward the PHB-G Hoogsteen face, with transient hydrogen bonding between the bulky moiety hydroxy group and flanking bases). Although the G*:T pairs are 15 to 20 kJ/mol weaker than the G*:C pairs within the helix, the G*:T mispairs cause little deviation in the canonical DNA structure, with a lesion base pair width of ~10.6 Å and opening of ~58°, as well as base step and base pair parameters consistent with natural DNA (Figures D.17-D.18, Appendix D).

When PHB-G is paired opposite A, the lesion forms persistent G*(N2H)···A(N7) and G*(N1)···A(N6H) hydrogen bonds (> 80%; Figures 5.7b and 5.8b and Table D.3, Appendix D). In contrast, POB-G forms transient G*(N2H)···A(N7) and G*(N1)···A(N6H) hydrogen bonds (< 50%; Figures 5.7b and 5.8b and Table D.3, Appendix D). These pairs are weaker than the G*:C pairs by ~12 kJ/mol. Furthermore, some deviations occur compared to canonical DNA for DNA containing POB-G:A or PHB-G:A, with the base pair width and opening angle increasing by ~1 Å and 10°, respectively (Figure 5.7b and 5.8b). Nevertheless, no significant deviations occur in the remaining helical parameters (Figure D.19-D.20, Appendix D).

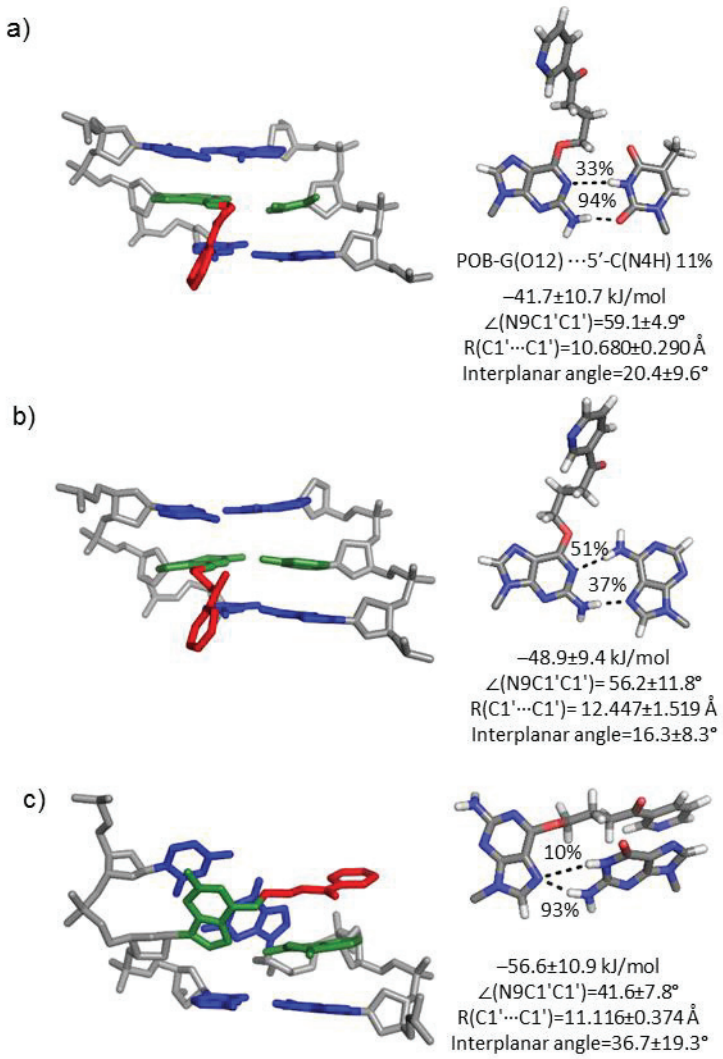

Figure 5.7. MD representative structures of POB-G adducted DNA with the bulky moiety in the major groove opposite a) T, b) A, or c) G. The trimer containing the lesion pair (green), flanking pairs (blue), and bulky moiety (red) is shown on the left, while the lesion-site hydrogen bonding is shown on the right.

When *syn*-POB-G or PHB-G is paired opposite G (Table D.3, Appendix D), a single hydrogen bond occurs between the adducted G(N7) and the pairing G (N2, for > 89% of the simulation; Figures 5.7c and 5.8c and Table D.6, Appendix D), which results in pairs that are up to 10 kJ/mol weaker than the G*:C pairs. Although there is little deviation from the canonical base pair width, the lesion mispair is not planar (interplanar angle ~35°). As

a result, the DNA helix is highly distorted with respect to all base step and base pair parameters surrounding the lesion site (Figures D.21-D.22, Appendix D). For example, the rise and twist between the adducted base pair and the 5′-flanking pair are 2 Å and ~90° smaller than for an undamaged helix, respectively.

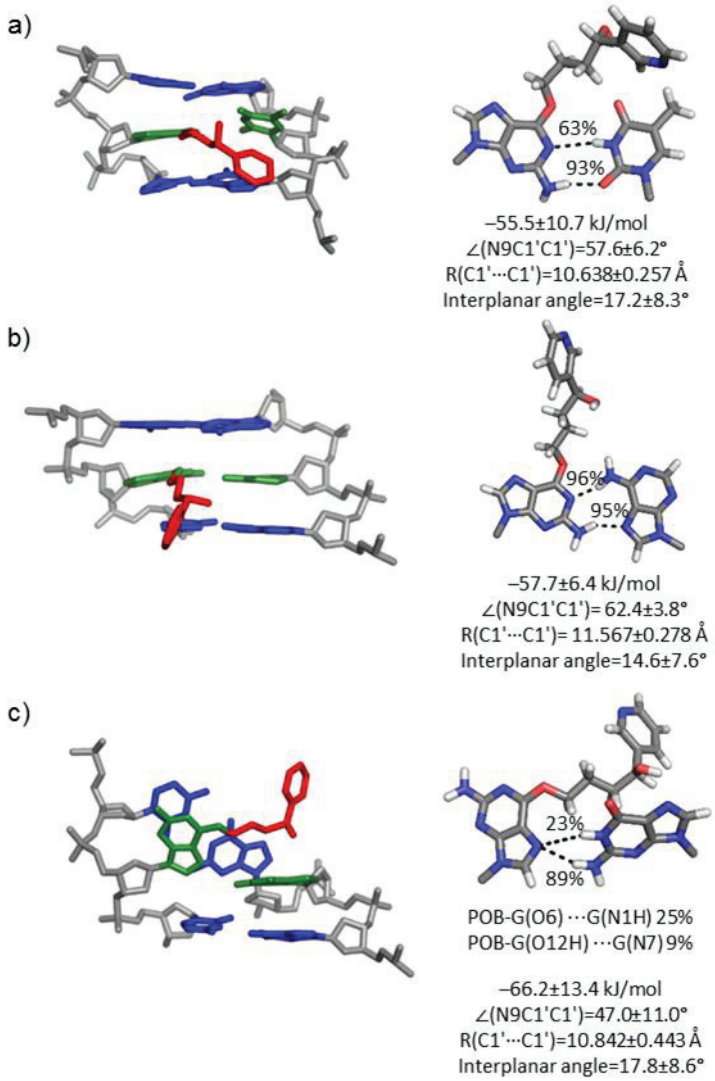

Figure 5.8. MD representative structures of PHB-G adducted DNA with the bulky moiety in the major groove opposite a) T, b) A, or c) G. The trimer containing the lesion pair (green), flanking pairs (blue), and bulky moiety (red) is shown on the left, while the lesion-site hydrogen bonding is shown on the right.

Overall, our simulations on post-replication DNA duplexes suggest that potential base substitution mutations formed upon replication of POB-G and PHB-G due to alterations in the Watson-Crick face of G will afford stable duplexes. Specifically, stable T and A mispairs can exist for both lesions, with the T mispairs causing the least distortion to the DNA duplex. However, the predicted highly distorted helices containing POB-G:G and PHB-G:G suggest that the G mispairs will destabilize the post-replication duplexes and may hinder lesion replication. This proposal is consistent with the observed formation of POB-G:G pairs, but the lack of extension past the lesion site by polymerase η, ι, κ, and ν *in vitro* (21,27).

5.3.5. Polymerase η insertion complexes indicate that dCTP and dTTP are best aligned for catalytic insertion opposite POB-G and PHB-G, but suggest that PHB-G may display a higher mutagenic potential: To understand the impact of constraints imposed by the polymerase η active site on the preferred lesion orientations and structures of lesion base pairs, the polymerase η insertion complexes were considered for dCTP, dTTP, or dATP incorporation opposite POB-G and PHB-G. The bulky moiety adopts many distinct extended conformations in the polymerase η major groove binding pocket (Figure D.23, Appendix D). However, in contrast to the lesions in unbound DNA, the bulky moiety is exclusively positioned away from the Watson-Crick face of the lesion in the polymerase η active site.

Regardless of the lesion conformation, all insertion complexes have key structural features that correctly align the dNTP for the phosphoryl transfer reaction. Specifically, Cys16, Phe17, Phe18, Tyr52, Arg55, Arg61, and Lys231 form hydrogen-bonding interactions with

the dNTP (Figure D.25, Appendix D) (34). Although minor differences exist in the interactions between the polymerase and each dNTP (Figure 5.10), all dNTPs maintain a network of dNTP-polymerase hydrogen bonds (Tables D.7-D.9, Appendix D). Additionally, all predicted structures for the polymerase η insertion complexes associated with POB-G and PHB-G replication maintain octahedral coordination of the binding and catalytic Mg^{2+} ions (Table D.10, Appendix D), which is known to be essential for catalysis (34). Nevertheless, the coordination of Oα2 to the binding Mg^{2+} ion adopts a wide spread of distances (Figure D.24, Appendix D). Finally, for a polymerase insertion complex to be catalytically active, the reaction coordinates should be aligned, which includes: 1) the distance between the 3′-primer end(O3′) and the dNTP(Pα) approaching the van der Waals radii for O and P (3.5 Å), and 2) the in-line attack angle [∠(3′-primer end(O3′)–dNTP(Pα)–dNTP(Oαβ))] approaching 180° (35-37). All POB-G and PHB-G replication complexes maintain this reaction coordinate orientation, with a distance < ~3.3 Å and angle > ~170° (Figure 5.9).

In addition to the contacts between polymerase η and the dNTP triphosphate tail, the position of the dNTP in the active site is stabilized by hydrogen-bonding interactions with the lesion in the template strand. Indeed, hydrogen bonding between the template base and incoming dNTP has been proposed to play an important role in dNTP selection by TLS polymerases (16,38,39). Consistent with DNA duplex models (Figure 5.5-5.8), POB-G and PHB-G form a wobble pair with dCTP that contains G*(N1)···dCTP(N4H) (> 40%) and G*(N2H)···dCTP(N3) (> 80%) interactions, and has an overall interaction strength of ~ –50 kJ/mol (Figure 5.9). The POB-G and PHB-G pairs with dCTP have opening angles

~10° wider than G:dCTP in the polymerase η active site, but have the same interplanar angle as G:dCTP. While the POB-G:dCTP pair has the same width as G:dCTP, the PHB-G:dCTP pair is ~0.7 Å wider.

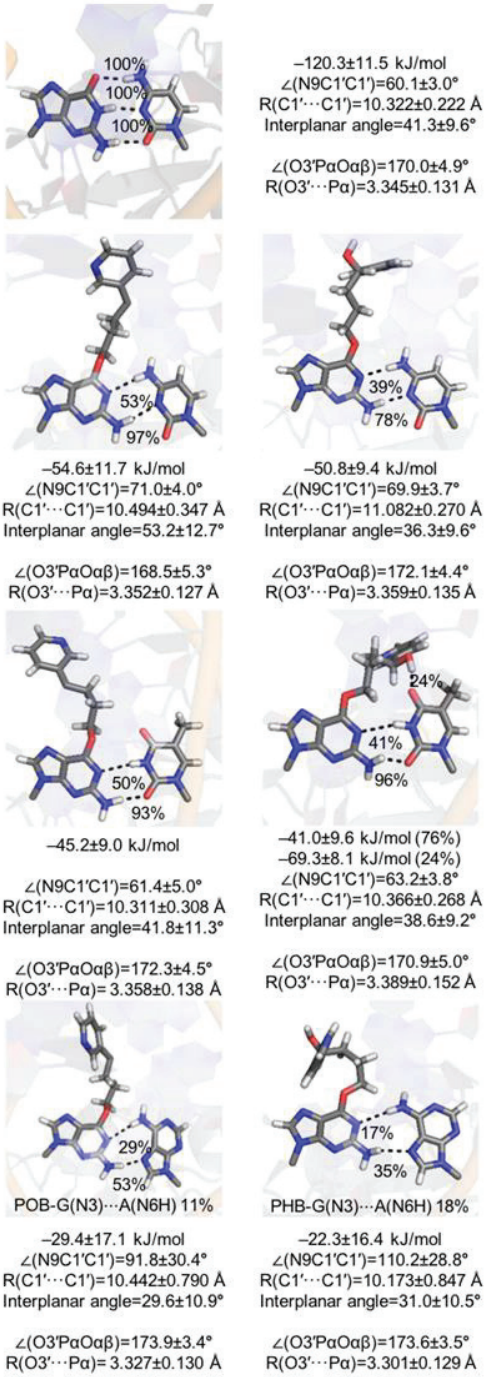

Figure 5.9. Hydrogen bonding between POB-G (left) or PHB-G (right) and an incoming dCTP, dTTP, or dATP in the polymerase η insertion complex.

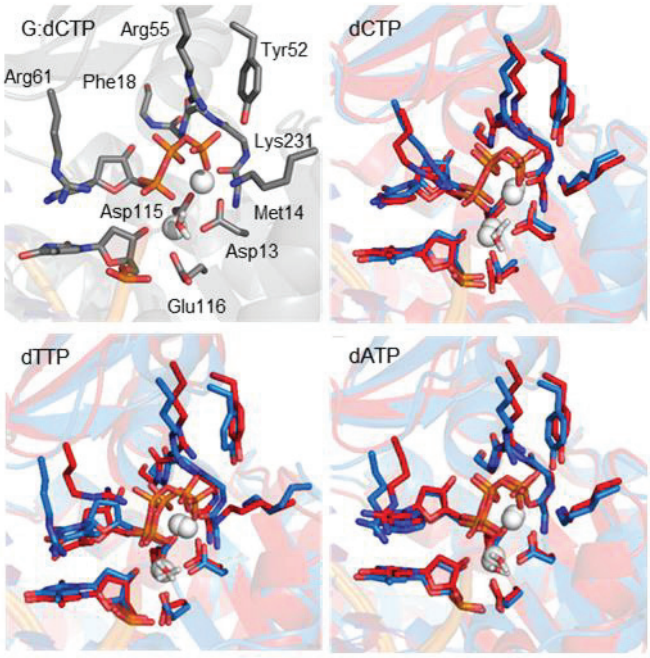

Figure 5.10. Overlay of the polymerase η insertion complex for POB-G (blue) and PHB-G (red) replication for dCTP, dTTP, or dATP insertion, highlighting the orientation of key active site residues around the dNTP.

The lesions also form two hydrogen bonds with dATP, namely the G*(N1)···dATP(N6H) (< 30%) and G*(N2H)···dATP(N7) (< 55%) hydrogen bonds. Nevertheless, dATP incorporation results in a weak base-pair strength (~ –25 kJ/mol; Figure 5.9). Furthermore, the dATP pairs contain wide opening angles (~90–110°), but possess a width and interplanar angle consistent with canonical DNA replication.

As seen for unbound DNA, a persistent pseudo-Watson-Crick pair occurs between dTTP and POB-G or PHB-G that contains an intermittent POB-G(N1)···dTTP(N3H) interaction (50 or 41%, respectively) and a well preserved POB-G(N2H)···dTTP(O2) hydrogen bond (> 93%; Figure 5.9). When only these two interactions are formed between the adducted

nucleobase and the incoming dTTP, both pairs have an interaction strength of ~ –45 kJ/mol. However, in contrast to unbound DNA and dTTP insertion opposite POB-G within the polymerase η active site, the PHB-G:dTTP pair can also contain a hydrogen bond between the bulky moiety hydroxy group and O4 of the pairing dTTP (24%). When this bulky moiety–dTTP contact is present, the interaction strength increases to ~ –70 kJ/mol. Irrespective of the involvement of the bulky moiety in dTTP binding, both the POB-G and PHB-G pairs maintain a structure consistent with canonical DNA.

Overall, regardless of the catalytically conducive alignment of the polymerase η active site, the lack of strong, persistent and undistorted hydrogen bonding between A and the lesion suggests that dATP is unlikely to be selectively incorporated opposite POB-G or PHB-G by polymerase η. On the other hand, dCTP insertion results in a more stable and a less distorted hydrogen-bonded pair with both lesions, as well as a correct active site configuration. Finally, persistent hydrogen bonding occurs between both lesions and an incoming dTTP, which leads to base pair structural parameters more consistent with natural DNA than dCTP. Therefore, our calculations predict that polymerase η replication of these two key tobacco induced DNA lesions will lead to preferential dNTP insertion according to dTTP > dCTP >> dATP. This is consistent with experimental studies on POB-G, which report a high frequency of G → A mutations and non-mutagenic replication (21,26,27), and highlights the predictive power of computer modeling for understanding the mutagenicity of DNA damage. In the absence of experimental data, the same computational approach reveals that the PHB-G:dTTP pair can be significantly more stable in the polymerase η active site than POB-G:dTTP, and therefore indicates that PHB-G may be more mutagenic. This effect is caused by a difference in the bulky moiety composition, and therefore this

finding complements previous literature that emphasizes the influence of adduct structure on the mutational outcomes of DNA lesions (14-25).

5.4. Conclusion

The current study uses a multiscale computational approach to provide structural data that rationalizes and predicts the mutagenicity of two tobacco-derived carcinogenic DNA lesions, namely POB-G and PHB-G. Quantum mechanical calculations reveal that both POB-G and PHB-G possess a high degree of inherent conformational flexibility, which may affect the biological processing of the lesions and emphasizes the importance of dynamical structural data for other flexible DNA lesions. Nevertheless, regardless of the conformation adopted, both lesions form stable and minimally distorted nucleobase pairs with C, T, and A, but highly distorted pairs with G. Indeed, molecular dynamics simulations predict that the G mispairs significantly alter the DNA duplex, suggesting that dGTP incorporation will not afford stable post-replication duplexes. This proposal is consistent with the previously reported formation of, but lack of extension past, POB-G:G pairs by TLS polymerases *in vitro* (21,27). In contrast, stable minimally distorted duplexes occur when POB-G or PHB-G are paired opposite C, T, and A. However, dATP weakly interacts with either lesion when bound within the polymerase, and therefore is unlikely to be selectively incorporated opposite POB-G or PHB-G, despite a catalytically conducive alignment of the polymerase η active site. This finding agrees with the observed low frequency of G → T mutations upon replication of POB-G *in vivo* (2%) (26). When dCTP and dTTP are placed opposite POB-G or PHB-G, the polymerase η active site is aligned for catalysis and the dNTP forms stable interactions with both lesions. However, the G*:dTTP pairs are less distorted, suggesting dTTP will be inserted more often than dCTP.

These findings are consistent with experimental studies that report an abundance of G → A mutations, as well as non-mutagenic replication, for POB-G (21,26,27), which validates our approach. Nevertheless, differences in the bulky moiety hydrogen-bonding patterns enhance the stability of the PHB-G:dTTP pair, suggesting that dTTP will be more often misinserted opposite PHB-G. Furthermore, the mutagenicity of PHB-G is likely increased by the stabilization of a conformation that intercalates the bulky moiety into the DNA helix, which has been associated with deletion mutations for other adducts. This contrasts the instability of an analogous conformer for POB-G, which coincides with the experimental mutagenicity spectrum. Nevertheless, future experimental work that probes the mutational profile of PHB-G is required to confirm the relative importance of deletion mutations for different tobacco-derived lesions. Overall, this study provides key structural insight into the reported mutagenicity of POB-G, uncovers the first clues about the mutagenicity of PHB-G, and adds to a growing body of literature that highlights the impact of small chemical changes to the bulky moiety on lesion mutagenicity.

5.5. References

1. Herrmann, S.S., Duedahl-Olesen, L., Christensen, T., Olesen, P.T. and Granby, K. (2015) Dietary exposure to volatile and non-volatile N-nitrosamines from processed meat products in denmark. *Food and Chemical Toxicology*, **80**, 137-143.

2. Al-Kaseem, M., Al-Assaf, Z. and Karabeet, F. (2014) Determination of seven volatile N-nitrosamines in fast food. *Pharmacology & Pharmacy*, **5**, 195.

3. Proksch, E. (2001) Review toxicological evaluation of nitrosamines in condoms. *International Journal of Hygiene and Environmental Health*, **204**, 103-110.

4. Lijinsky, W. (1986) The significance of N-nitroso compounds as environmental carcinogens. *Environmental Carcinogenesis Reviews*, **4**, 1-45.

5. Peterson, L.A. (2016) Context matters: Contribution of specific DNA adducts to the genotoxic properties of the tobacco-specific nitrosamine nnk. *Chem. Res. Toxicol.*

6. St.Helen, G., Benowitz, N.L., Dains, K.M., Havel, C., Peng, M. and Jacob, P. (2014) Nicotine and carcinogen exposure after water pipe smoking in hookah bars. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, **23**, 1055-1066.

7. Kim, H.-J. and Shin, H.-S. (2013) Determination of tobacco-specific nitrosamines in replacement liquids of electronic cigarettes by liquid chromatography–tandem mass spectrometry. *Journal of Chromatography A*, **1291**, 48-55.

8. Hecht, S.S. (1998) Biochemistry, biology, and carcinogenicity of tobacco-specific N-nitrosamines. *Chem. Res. Toxicol.*, **11**, 559-603.

9. Hecht, S.S. (2008) Progress and challenges in selected areas of tobacco carcinogenesis. *Chem. Res. Toxicol.*, **21**, 160-171.

10. Alavanja, M., Baron, J., Brownson, R.C., Buffler, P.A., DeMarini, D.M., Djordjevic, M.V., Doll, R., Fontham, E.T., Gao, Y.T. and Gray, N. (2004) Tobacco smoke and involuntary smoking. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*, **83**, 1-1413.

11. Hecht, S.S. (2014) It is time to regulate carcinogenic tobacco-specific nitrosamines in cigarette tobacco. *Cancer Prevention Research*, **7**, 639-647.

12. Czoli, C.D. and Hammond, D. (2018) Trends over time in tobacco-specific nitrosamines (tsnas) in whole tobacco and smoke emissions from cigarettes sold in canada. *Nicotine & Tobacco Research*, **20**, 649-653.

13. Siegel, R.L., Miller, K.D. and Jemal, A. (2015) Cancer statistics, 2015. *CA: a cancer journal for clinicians*, **65**, 5-29.

14. Patra, A., Nagy, L.D., Zhang, Q.Q., Su, Y., Muller, L., Guengerich, F.P. and Egli, M. (2014) Kinetics, structure, and mechanism of 8-oxo-7,8-dihydro-2 '-deoxyguanosine bypass by human DNA polymerase eta. *J. Biol. Chem.*, **289**, 16867-16882.

15. Gahlon, H.L., Schweizer, W.B. and Sturla, S.J. (2013) Tolerance of base pair size and shape in postlesion DNA synthesis. *J. Am. Chem. Soc.*, **135**, 6384-6387.

16. Gahlon, H.L., Boby, M.L. and Sturla, S.J. (2014) O6-alkylguanine postlesion DNA synthesis is correct with the right complement of hydrogen bonding. *ACS Chem. Biol.*, **9**, 2807-2814.

17. Kirouac, K.N., Basu, A.K. and Ling, H. (2013) Structural mechanism of replication stalling on a bulky amino-polycyclic aromatic hydrocarbon DNA adduct by a y family DNA polymerase. *J. Mol. Biol.*, **425**, 4167-4176.

18. Stover, J.S., Chowdhury, G., Zang, H., Guengerich, F.P. and Rizzo, C.J. (2006) Translesion synthesis past the C8- and N2-deoxyguanosine adducts of the dietary mutagen 2-amino-3-methylimidazo[4,5-f]quinoline in the NarI recognition sequence by prokaryotic DNA polymerases. *Chem. Res. Toxicol.*, **19**, 1506-1517.

19. Zhang, H., Eoff, R.L., Kozekov, I.D., Rizzo, C.J., Egli, M. and Guengerich, F.P. (2009) Versatility of Y-family sulfolobus solfataricus DNA polymerase Dpo4 in translesion synthesis past bulky N2-alkylguanine adducts. *J. Biol. Chem.*, **284**, 3563-3576.

20. Eoff, R.L., Angel, K.C., Egli, M. and Guengerich, F.P. (2007) Molecular basis of selectivity of nucleoside triphosphate incorporation opposite O6-benzylguanine by sulfolobus solfataricus DNA polymerase Dpo4: Steady-state and pre-steady-state kinetics and X-ray crystallography of correct and incorrect pairing. *J. Biol. Chem.*, **282**, 13573-13584.

21. Choi, J.-Y., Chowdhury, G., Zang, H., Angel, K.C., Vu, C.C., Peterson, L.A. and Guengerich, F.P. (2006) Translesion synthesis across O6-alkylguanine DNA adducts by recombinant human DNA polymerases. *J. Biol. Chem.*, **281**, 38244-38256.

22. Sherrer, S.M., Brown, J.A., Pack, L.R., Jasti, V.P., Fowler, J.D., Basu, A.K. and Suo, Z. (2009) Mechanistic studies of the bypass of a bulky single-base lesion catalyzed by a Y-family DNA polymerase. *J. Biol. Chem.*, **284**, 6379-6388.

23. Gadkari, V.V., Tokarsky, E.J., Malik, C.K., Basu, A.K. and Suo, Z. (2014) Mechanistic investigation of the bypass of a bulky aromatic DNA adduct catalyzed by a Y-family DNA polymerase. *DNA Repair*, **21**, 65-77.

24. Xu, P., Oum, L., Lee, Y.-C., Geacintov, N.E. and Broyde, S. (2009) Visualizing sequence-governed nucleotide selectivities and mutagenic consequences through a replicative cycle: Processing of a bulky carcinogen N2-dG lesion in a Y-Family DNA polymerase. *Biochemistry*, **48**, 4677-4690.

25. Wyss, L.A., Nilforoushan, A., Eichenseher, F., Suter, U., Blatter, N., Marx, A. and Sturla, S.J. (2015) Specific incorporation of an artificial nucleotide opposite a mutagenic DNA adduct by a DNA polymerase. *J. Am. Chem. Soc.*, **137**, 30-33.

26. Pauly, G.T., Peterson, L.A. and Moschel, R.C. (2002) Mutagenesis by O6-[4-oxo-4-(3-pyridyl)butyl]guanine in escherichia coli and human cells. *Chem. Res. Toxicol.*, **15**, 165-169.

27. Gowda, A.S.P. and Spratt, T.E. (2016) DNA polymerase ν rapidly bypasses O6-methyl-dG but not O6-[4-(3-pyridyl)-4-oxobutyl-dG and o2-alkyl-dts. *Chem. Res. Toxicol.*, **29**, 1894-1900.

28. Carmella, S., Ye, M., Upadhyaya, P. and S Hecht, S. (1999) *Stereochemistry of metabolites of a tobacco-specific lung carcinogen in smokers' urine*.

29. Park, S., Seetharaman, M., Ogdie, A., Ferguson, D. and Tretyakova, N. (2003) 3′-exonuclease resistance of DNA oligodeoxynucleotides containing O6-[4-oxo-4-(3-pyridyl) butyl] guanine. *Nucleic Acids Res.*, **31**, 1984-1994.

30. Peterson, L.A., Vu, C., Hingerty, B.E., Broyde, S. and Cosman, M. (2003) Solution structure of an O6-[4-oxo-4-(3-pyridyl)butyl]guanine adduct in an 11mer DNA duplex: Evidence for formation of a base triplex. *Biochemistry*, **42**, 13134-13144.

31. Wilson, K.A., Szemethy, K.G. and Wetmore, S.D. (2017) Conformational flexibility and base-pairing tendency of the tobacco carcinogen O6-[4-oxo-4-(3-pyridyl)butyl]guanine. *Biophysical Chemistry*, **228**, 25-37.

32. Norman, D., Abuaf, P., Hingerty, B.E., Live, D., Grunberger, D., Broyde, S. and Patel, D.J. (1989) 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*, **28**, 7462-7476.

33. Gu, Z., Gorin, A., Hingerty, B.E., Broyde, S. and Patel, D.J. (1999) Solution structures of aminofluorene [AF]-stacked conformers of the syn [AF]−C8-dG adduct positioned opposite dc or dA at a template-primer junction. *Biochemistry*, **38**, 10855-10870.

34. Nakamura, T., Zhao, Y., Yamagata, Y., Hua, Y.-J. and Yang, W. (2012) Watching DNA polymerase eta make a phosphodiester bond. *Nature*, **487**, 196-U177.

35. Brautigam, C.A. and Steitz, T.A. (1998) Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Curr. Opin. Struct. Biol.*, **8**, 54-63.

36. Wang, Y. and Schlick, T. (2008) Quantum mechanics/molecular mechanics investigation of the chemical reaction in Dpo4 reveals water-dependent pathways and requirements for active site reorganization. *J. Am. Chem. Soc.*, **130**, 13240-13250.

37. Wang, L., Yu, X., Hu, P., Broyde, S. and Zhang, Y. (2007) A water-mediated and substrate-assisted catalytic mechanism for sulfolobus solfataricus DNA polymerase IV. *J. Am. Chem. Soc.*, **129**, 4731-4737.

38. Washington, M.T., Helquist, S.A., Kool, E.T., Prakash, L. and Prakash, S. (2003) Requirement of Watson-Crick hydrogen bonding for DNA synthesis by yeast DNA polymerase η. *Mol. Cell. Biol.*, **23**, 5107-5112.

39. Wolfle, W.T., Washington, M.T., Kool, E.T., Spratt, T.E., Helquist, S.A., Prakash, L. and Prakash, S. (2005) Evidence for a Watson-Crick hydrogen bonding requirement in DNA synthesis by human DNA polymerase κ. *Mol. Cell. Biol.*, **25**, 7137-7143.

Chapter 6: Evidence for the replication of a *syn* DNA pyrimidine: A computational investigation of a highly mutagenic tobacco-derived thymine lesion[a,b,c]

6.1. Introduction

DNA encodes all information required for human life. However, this information is constantly threatened by many endogenous and exogenous damaging agents (1). Some particularly carcinogenic yet prevalent sources of DNA damage are nicotine products, such as conventional cigarettes and e-cigarette liquid (2, 3). The use of tobacco products is associated with several cancers (e.g., lung, head, and neck) and it is estimated to account for 30% of cancer deaths worldwide (4). One of the main causes of tobacco-induced cancers is believed to be 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), which is formed *in vivo* when nicotine is metabolized (5-7). In fact, NNK is the only tobacco component that has led to cancer in every species tested (i.e., mice, rats, hamsters, rabbits, pigs, monkeys, and humans) irrespective of the route of exposure (8, 9). Within the body, NNK is converted into species that react with DNA to form many different purine and pyrimidine

---

[a] *Proceedings of the National Academy of Science U.S.A.* reference style used throughout this chapter.

[b] Adapted from Wilson, K.A., Holland, C. and Wetmore, S.D. (2018). Evidence for the replication of a syn DNA pyrimidine: A computational investigation of a highly mutagenic tobacco-derived thymine lesion. *Nucleic. Acids Res.* (In preparation).

[c] C.H. performed the nucleobase model conformational search and preliminary calculations on the nucleoside and DNA models. K.A.W. performed calculations on the nucleobase, nucleoside, base pair, DNA and polymerase models, as well as data analysis on all models and wrote the first draft of the manuscript.

lesions such as methyl (O2-T, O4-T, or O6-G), formaldehyde (N6-A, N4-C, or N2-G) and pyridyloxobutyl (POB; O2-T, O4-T, N6-A, O6-G, N7-G, O2-C, N3-C, or N4-C) adducts (7).

DNA lesions caused by tobacco affect many cellular pathways (10, 11). For example, standard DNA replication is typically stalled by damaged nucleobases, and an alternative pathway known as translesion synthesis (TLS) is responsible for replicating past damaged sites. While TLS is critical for preventing cell death, it is error prone and can introduce mutations, which in turn can lead to cancer. Current literature on the replication of DNA adducts arising from known human carcinogens shows that TLS is affected by the chemical structure of the lesion (12-20). However, most studies on damaged DNA replication have considered purine lesions (see, for example, references (12-20)), while very little is known about the replication of damaged pyrimidines. In fact, only a few damaged pyrimidines have been identified in the literature to date (7, 21, 22).

A common pyrimidine adducts is O2-[4-oxo-4-(3-pyridyl)butyl]thymine (POB-T; Figure 6.1a), which is also the most persistent POB adduct formed from NNK, at least in part due to a repair resistance (23, 24). Additionally, the mutational signature (signature 4) associated with tobacco smoking in human cancers contains T→A mutations (25) that have not yet been explained, but could be associated with POB-T. As a result, researchers are striving to understand the structure and mutagenicity of this T lesion. POB-T is a minor groove lesion and has an altered Watson-Crick (WC) hydrogen-bonding face compared to canonical T (i.e., N3 becomes a hydrogen-bond acceptor; Figure 6.1a), and therefore may no longer preferentially pair with A. Indeed, *in vitro* studies indicate that DNA polymerase

η (pol η) preferentially inserts dATP and dTTP opposite POB-T, while DNA polymerase κ (pol κ) preferentially inserts dGTP over dATP opposite the lesion (26, 27). Interestingly, the unique outcomes for POB-T replication by pol κ and η indicate that these two human TLS polymerases process the lesion in different ways, although the exact pathways are currently unknown.

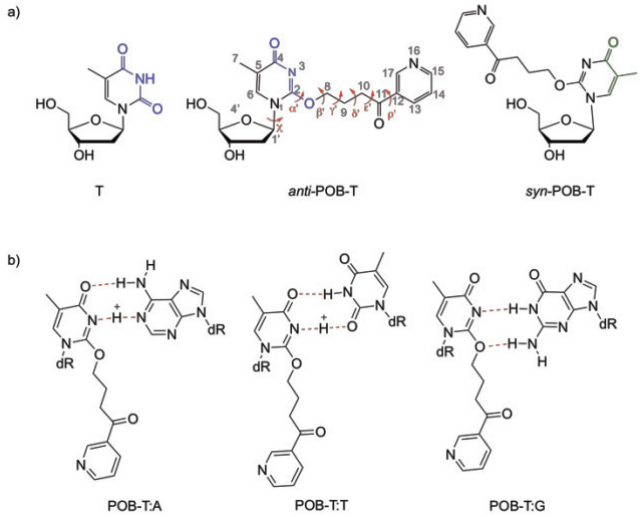

Figure 6.1. a) The T, *anti*-POB-T and *syn*-POB-T nucleosides, with WC face in blue and Hoogsteen face in green. Key dihedral angles for the lesions include α′ [∠(N1C2O2C8)], β′ [∠(C2O2C8C9)], γ′ [∠(O2C8C9C10)], δ′ [∠(C8C9C10C11)], ϵ′ [∠(C9C10C11C12)], ρ′ [∠(C10C11C12C13)] and χ [∠(C2N1C1′O4′)]. b) POB-T base pairings previously speculated in the literature (27, 28).

To clarify the polymerase responsible for POB-T bypass *in vivo*, single and double knockout strains of human cells (HEK293K) were considered for pol η, κ, ι, and ζ, and Rev1 (28, 29). Knockout of pol η and ζ has a significant effect on lesion bypass (bypass efficacy reduced from < 55% in wild-type cells to < 36% in the pol η knockout or < 22% in the double knockout), while knockout of pol κ, ι or Rev1 did not substantially change the efficiency of TLS past POB-T (28, 29). These results from the knockout strains are consistent with *in vitro* work indicating that pol η most efficiently bypasses POB-T (27).

Nevertheless, pol η replicates POB-T ~1000-fold less efficiently than natural DNA (26, 27). Furthermore, in wild-type cells, T→A mutations (i.e., dTTP incorporations) occur at a high frequency (up to ~50%) regardless of the organism (i.e., *E. coli*, mice, or humans) (28-32). These *in vivo* studies are consistent with POB-T replication by pol η, since *in vitro* studies indicate that pol η most commonly inserts dTTP opposite POB-T (26, 27). However, these conclusions contrast literature suggesting pol κ replicates minor groove lesions (33), while pol η replicates pyrimidine dimers and major groove adducts (34, 35). Indeed, pol κ has an open pocket on the minor groove side of the DNA duplex that can accommodate a bulky moiety, while pol η has an open pocket that aligns with the DNA major groove (36, 37). Nevertheless, pol κ has been proposed to require WC hydrogen bonding for rapid catalysis (28, 38), a condition that is not satisfied by POB-T.

There is currently no structural information available for POB-T replication. However, hydrogen-bonding patterns between the lesion and the canonical nucleobases have been proposed to rationalize the observed lesion mutagenicity. Specifically, the prevalence of dGTP insertion by pol κ was speculated to arise due to the formation of two hydrogen bonds between POB-T and dGTP (Figure 6.1b) (26, 27). On the other hand, the observed insertion of dATP and dTTP by pol η (26, 27) was explained based on interbase hydrogen bonding involving protonation of the lesion (i.e., at N3) or the pairing base (i.e., at N1 of A or O2 of T) (Figure 6.1b) (27, 28). However, the authors acknowledge that these proposed pairs do not explain why dGTP is not inserted by pol η (26, 27). Additionally, mass spectrometry data indicates that POB-T is not cationic (39), and there is no evidence to support the protonation of a canonical DNA nucleobase interacting with POB-T. It has therefore been suggested that protein–DNA interactions may play a large role in dictating dNTP insertion

opposite POB-T (27, 28). Thus, detailed structural information about the lesion and polymerase-DNA complex is required to elucidate factors that control POB-T replication by pol κ and η.

Previous work on other DNA lesions has shown that computational chemistry is an important tool for gaining molecular level insight into the function of DNA polymerases, including lesion accommodation within the active site and mutagenic outcomes (40-45). Therefore, the present work uses density functional theory (DFT) calculations and molecular dynamics (MD) simulations to provide the first structural information for POB-T, as well as the preferred base pairings and lesion replication. Specifically, DFT nucleobase models reveal the intrinsic conformational preference about the nucleobase–bulky moiety linker and within the bulky moiety of POB-T. Next, docking and MD simulations uncover how the lesion is accommodated in the pol κ and η active sites prior to dNTP binding. DFT models of isolated hydrogen-bonded pairs provide insight into possible interactions between POB-T and the canonical nucleobases. Finally, MD simulations on insertion complexes and select post-replication DNA duplexes shed light on how pol κ and η replicate POB-T and rationalize the varying propensities for dNTP insertion. Overall, our multiscale computational approach provides the structural insight necessary to clarify the apparent discrepancy in the DNA replication literature by highlighting differences in how two human TLS polymerases replicate POB-T and thereby expanding the biological role of pol η.

6.2. Results and Discussion

6.2.1. POB-T is a highly flexible lesion: To understand the inherent structural preference about the nucleobase–carcinogen linker and within the bulky moiety in the absence of steric constraints**,** the conformational landscape of the POB-T nucleobase was mapped using DFT. Calculations initiated from a linear orientation (i.e., α', β', γ', δ', ϵ', and $\rho' = 180°$, Figure 6.1a) indicate that the bulky moiety exhibits a ~41 kJ/mol preference to be positioned toward the WC face at the nucleobase–carcinogen linker (i.e., first carbon of the bulky moiety directed toward N3), with the corresponding rotational barrier being ~45 kJ/mol (Figure E.1-E.2, Appendix E). Nevertheless, POB-T is highly flexible about all other key bulky moiety dihedral angles (minima within ~5.5 kJ/mol and rotational barriers of < 30 kJ/mol; Figure E.1-E.2, Appendix E). Therefore, the interplay between various bulky moiety orientations needs to be considered.

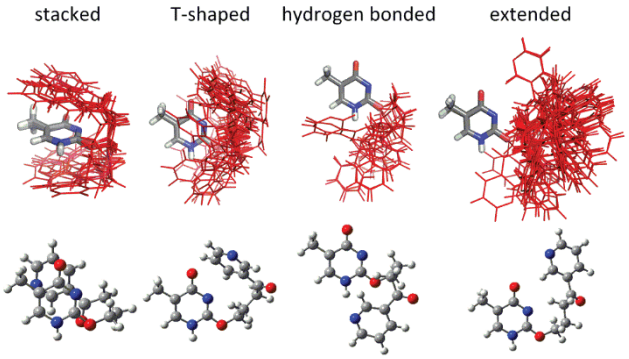

Figure 6.2. Overlay of DFT POB-T nucleobase conformations (based on damaged T heavy atoms, bulky moiety highlighted in red, top) and most stable conformer from each category (bottom).

Since full exploration of the POB-T conformational space would require characterization of a 7-dimensional energy surface, a usage-directed conformational search was performed to gain insight into the minimum energy regions on the surface. The resulting 164 unique adduct conformations (Figure E.3, Appendix E) were visually inspected and classified

based on the interactions between the bulky moiety and the adducted T as stacked, T-shaped, hydrogen bonded or extended (no interaction; Figure 6.2). Although there is substantial structural and energetic (up to ~50 kJ/mol) variation within each conformational category (Figure 6.2 and Figure E.4, Appendix E), the majority of accessible POB-T conformations are extended (49%), while there is an approximately equal number of stacked and T-shaped structures (~20% each), and fewer hydrogen bonded conformers (10%; Figure E.4a, Appendix E). However, within DNA none of the hydrogen bonded conformations can occur, and many of the T-shaped and extended conformations are excluded as well, due to the close proximity of the bulky moiety and the site of backbone attachment (Figure 6.2). Regardless, the lesion displays a consistent orientation at the nucleobase–carcinogen linker (i.e., first carbon of the bulky moiety is directed toward the WC face) and the bulky moiety is rarely linear in the 52 lowest energy structures (relative energy < 20 kJ/mol; Figure E.5, Appendix E).

6.2.2. POB-T can adopt the canonical *anti* glycosidic orientation in the pol κ active site, while only *syn*-POB-T can be accommodated by pol η: The orientation of POB-T in the pol κ and η active sites prior to dNTP insertion was investigated to gain insight into the impact of the enzyme and surrounding DNA on the lesion conformation. Initially, all POB-T conformers identified in the conformational search were manually docked into the pol κ active site such that the lesion adopts the canonical *anti* conformation about the glycosidic bond. Although most conformations of the isolated POB-T nucleobase do not fit in the pol κ active site, 37 of the 164 conformations are well accommodated (closest heavy atom distance between the lesion and polymerase or DNA > 2.0 Å; Figure E.6, Appendix E). All 37 conformers adopt an extended orientation, the most common

conformation of the isolated nucleobase. When MD simulations (100 ns in triplicate) were performed on the pol κ preinsertion complex, POB-T adopts orientations that span all 37 manually docked conformations (Figure E.7, Appendix E). These orientations position the bulky moiety in an open pocket near the little finger domain and away from the reactive center of pol κ. There is little change in the overall polymerase structure upon POB-T accommodation compared to a crystal structure of T replication (Figure 6.3a). This accommodation of POB-T correlates with evidence that pol κ replicates minor groove adducts, such as N2-G and N3-A lesions (36, 37).

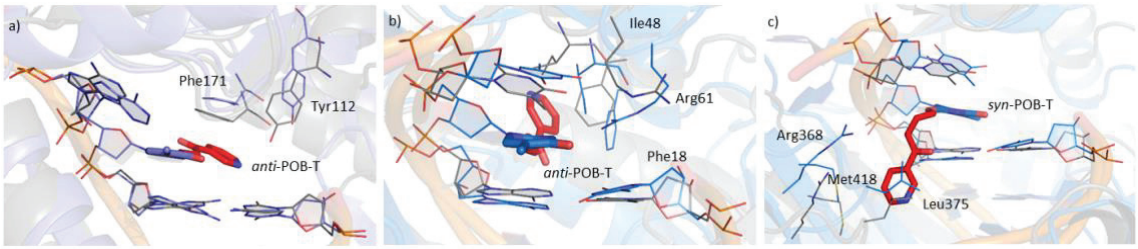

Figure 6.3. Overlay of MD representative structures for a) *anti*-POB-T bound in the pol κ active site (purple) and b) *anti*-POB-T or c) *syn*-POB-T bound in the pol η active site (blue) with crystal structures (grey) for dATP insertion opposite dT (PDB ID: 5W2C for pol κ and 4ECS for pol η).

Since pol η is generally believed to replicate major groove lesions (46), it is not clear how the enzyme exerts activity towards the minor groove POB-T adduct. Indeed, when all POB-T nucleobase orientations were manually docked into the pol η active site using the same protocol applied to pol κ, only 8 of the 164 conformations fit in the pol η active site (Figure E.8, Appendix E). However, these 8 conformations correspond to extended orientations that place the bulky moiety in close proximity to key residues in the catalytic palm domain (e.g., Phe18 and Arg61) and/or the binding position of the incoming dNTP. Furthermore, subsequent MD simulations reveal disruptions in the active site region that will compromise dNTP binding regardless of the initial bulky moiety orientation (Figure 6.3b

and E.9, Appendix E). This suggests that *anti*-POB-T is unlikely to be bypassed by pol η, which is consistent with the lack of replication of other minor groove lesions by pol η (36, 37), but contradicts *in vivo* and *in vitro* studies on POB-T (26-28). Nevertheless, pol η may accommodate POB-T in a different conformation.

A change in the glycosidic conformation for POB-T was considered, as the addition of various bulky moieties to purine nucleosides induces a *syn* glycosidic conformation (14, 47, 48). Although no such preference has been previously reported for a pyrimidine lesion, our DFT calculations suggest that the *syn*-POB-T conformation is only ~9 kJ/mol higher in energy than the *anti* conformation and the two conformations are connected by a modest ~35 kJ/mol barrier (Figure E.10, Appendix E). When *syn*-POB-T was manually docked into the pol η active site, significantly more POB-T conformations can be accommodated by pol η compared to *anti*-POB-T (30 versus 8; Figure E.8, Appendix E). These conformations extend the bulky moiety on the major groove side of the growing helix and into an open pocket near the pol η little finger domain. Subsequent MD simulations reveal only minor changes in the little finger domain to accommodate the lesion (Figure 6.3), which will likely have a much smaller impact on dNTP insertion than the disruption of the catalytic palm domain seen for *anti*-POB-T. While damaged purines have been shown to adopt a *syn* glycosidic orientation during replication (14, 47, 48), to the best of our knowledge this is the first evidence suggesting that a pyrimidine adopts the *syn* glycosidic conformation within a polymerase active site. Nevertheless, other nucleic processing enzymes, such as the TET family of enzymes (49), the NEIL1 DNA repair enzyme (50), and functional RNA (51), bind pyrimidines in the *syn* conformation. The unique POB-T glycosidic orientations accommodated by pol κ and η may explain the

observed differential replication outcomes, which will be probed in the following sections by analyzing the corresponding dNTP insertion complexes.

6.2.3. Pol κ adopts a catalytically conducive conformation for the insertion of dGTP opposite POB-T: Prior to investigating the POB-T base-pairing preferences in the pol κ active site, DFT calculations were used to explore potential hydrogen bonding between the POB-T WC face and each canonical nucleobase (Figure 6.4). POB-T was considered in the extended conformation, the only orientation accommodated by pol κ. In all dimers, a stacking interaction occurs between the POB-T bulky moiety and the pairing base, which complements interbase hydrogen bonding to stabilize the base pairs (Figure 6.4a). When POB-T is paired with A or G, two hydrogen bonds occur (Figure 6.4a). Nevertheless, POB-T pairing with G (~–94 kJ/mol) is significantly stronger than with A (~–70 kJ/mol). Furthermore, the POB-T:G hydrogen-bonding geometry is different from that proposed in the literature (27, 28) (Figure 6.1b) since our predicted orientation at the carcinogen–nucleobase linker directs the first carbon of the bulky moiety toward the lesion WC face and thus prevents O2 of the lesion from hydrogen bonding with the pairing G. Finally, POB-T mispairs with T or C contain a single hydrogen bond and have interaction energies of –70.6 or –80.7 kJ/mol, respectively (Figure 6.4a). All POB-T pairs are non-planar (interplanar angle ≈ 30°). Additionally, the base pairs exhibit distortion in the canonical base pair width (by up to 2.2 Å) and/or opening (by up to 30°), with the exception of the POB-T:*syn*-dATP pair, which maintains a width and opening consistent with canonical DNA.

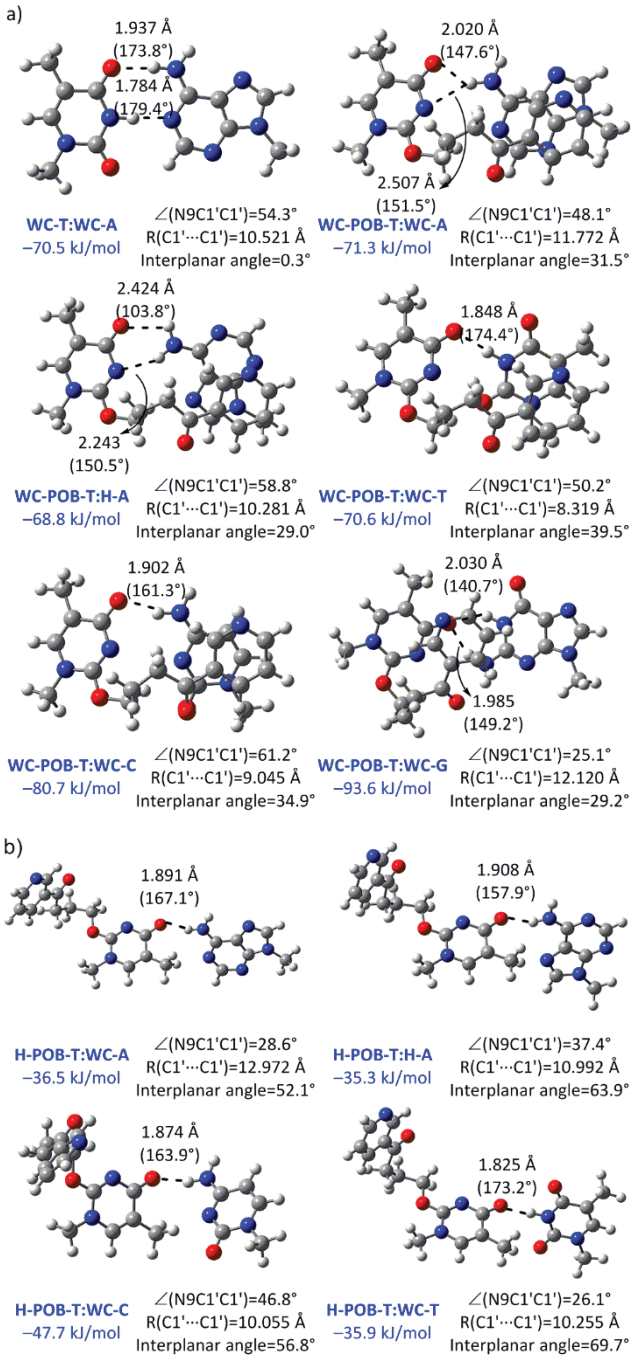

Figure 6.4. DFT structures and binding energies for base pairs between the a) T or POB-T WC face or b) POB-T Hoogsteen (H) face and the canonical nucleobases.

To gain a better understanding of POB-T replication by pol κ, MD simulations on the pol κ insertion complexes with dGTP or dATP paired opposite POB-T were performed. These pairs were considered since G and A form two hydrogen bonds with POB-T in the

nucleobase dimers and these pairs are observed experimentally (26, 27). Conversely, pairing C and T lead to only a single hydrogen bond with POB-T and does not correlate with an experimentally-observed mutagenic outcome and thus these pairs were not considered in the polymerase active site. Regardless of the opposing base, the bulky moiety adopts many distinct extended conformations in the insertion complexes. Nevertheless, the first carbon of the bulky moiety is preferentially directed toward the lesion WC face as observed for the isolated nucleobase (Figure E.11, Appendix E). The bound POB-T orientation consistently positions the bulky moiety in an open pocket of pol κ, where a hydrogen bond occurs between the bulky moiety N and Ser102 (Table E.1-E.3, Appendix E). Unlike the nucleobase dimers, the steric constraints of the active site prevent interactions between the bulky moiety and opposing base, which will destabilize the pairs.

To correctly align the dNTP for the phosphoryl transfer reaction, a number of active site features must be maintained, including octahedral coordination of the catalytic and binding Mg^{2+} ions, a reaction distance [d(3′-primer end(O3′)–dNTP(Pα))] of < 3.5 Å and a reaction angle [∠(3′-primer end(O3′)–dNTP(Pα)–dNTP(Oαβ))] of ~ 180° (52-54). These criteria are satisfied for dATP and dGTP insertion opposite POB-T by pol κ (Figure 6.5 and Table E.4, Appendix E). In addition, minimal differences occur in the interactions between the dNTP backbone, and the surrounding amino acids compared to dATP insertion opposite T (Figure E.12a and Tables E.1-E.3 and E.5, Appendix E).

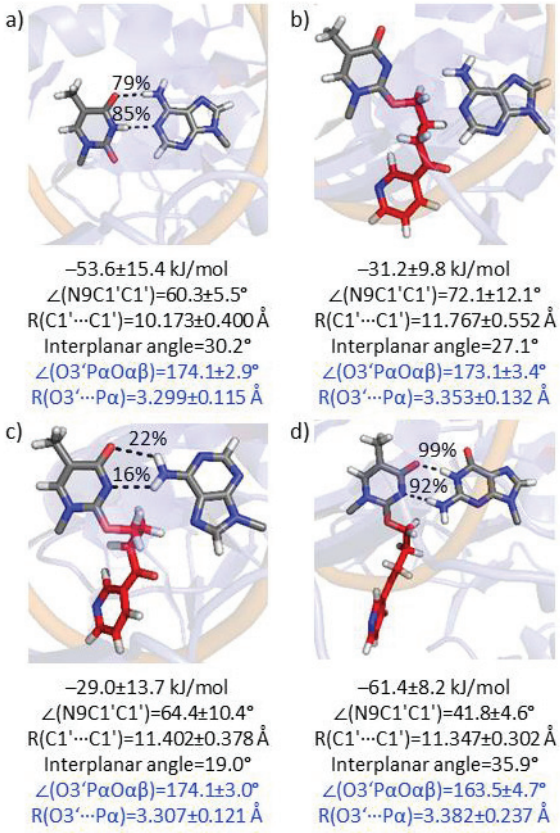

Figure 6.5. Structure and reaction coordinates for pol κ insertion of dATP opposite a) T or b) POB-T, and c) *syn*-dATP or d) dGTP insertion opposite POB-T.

Despite the correct alignment of the reaction coordinate, the largest differences in the dATP and dGTP pol κ insertion complexes occur in the nascent base pair. Specifically, POB-T does not hydrogen bond with an incoming *anti*-dATP. Instead, the bulky moiety chain or pyridyl ring form non-covalent C–H···π and π···π interactions with dATP (Figure E.11, Appendix E). This leads to an average POB-T:dATP interaction energy of –31.2 kJ/mol. In agreement with data for the DFT nucleobase dimers, changing the dATP glycosidic orientation results in a less distorted base pair in the pol κ active site, which maintains an opening and interplanar angle consistent with canonical DNA. Furthermore, POB-T(O4)···dATP(N6H) and POB-T(N3)···dATP(N6H) hydrogen bonds are transiently formed (< 22%), and the pair has an average interaction energy of –29.0 kJ/mol (Figure

6.5c). Nevertheless, despite the reaction coordinate being aligned for catalysis, weak interactions occur between POB-T and dATP, regardless of the dATP glyosidic orientation. Overall, this correlates with *in vitro* kinetic data suggesting that dATP is infrequently inserted opposite POB-T by pol κ (26, 27).

Although POB-T:dGTP is distorted compared to undamaged DNA (deviations in base pair width of ~1 Å and opening of ~20°), the POB-T(O4)···dGTP(N1H) and POB-T(N3)···dGTP(N2H) hydrogen bonds are persistent (> 90%; Figure 6.5d). This leads to an average interaction strength of –61.4 kJ/mol, which is ~12 or ~30 kJ/mol stronger than canonical T:dATP or POB-T:dATP, respectively (Figure 6.5d). When coupled with the catalytic orientation of the reaction coordinate, this correlates with experimental *in vitro* kinetic studies that indicate dGTP is inserted opposite POB-T at a higher rate than dATP (27). Although our calculations rationalize the *in vitro* data, *in vivo* studies indicate that knockout of pol κ does not influence POB-T replication and therefore it is unlikely that pol κ replicates POB-T *in vivo* (28). Since the predicted insertion complexes are aligned for catalysis, our data supports suggestions that the decreased reactivity of pol κ towards POB-T may stem from the lack of WC hydrogen bonding, which is proposed to be required for rapid catalysis (38).

6.2.4. Unique *syn* glycosidic orientation explains experimentally observed dTTP and dATP insertion opposite POB-T by pol η: MD simulations on preinsertion complexes predict that an extended orientation of POB-T is accommodated in the pol η active site in the *syn,* but not canonical *anti,* glycosidic orientation. However, it is unclear how the POB-T Hoogsteen face interacts with the canonical nucleobases. DFT calculations on pairs

between the POB-T Hoogsteen face and the canonical nucleobases reveal no interactions between the bulky moiety and any canonical pairing base. Although a stable Hoogsteen-POB-T:G base pair could not be isolated, a single hydrogen bond occurs between POB-T(O4) and A(N6H), C(N4H), or T(N3H) (Figure 6.4b). As a result, pairs involving the POB-T Hoogsteen face are all significantly less stable than canonical T:A (by 22 – 35 kJ/mol) and pairs involving the POB-T WC face (by 21 – 59 kJ/mol). Additionally, while all *syn*-POB-T base pairs retain the width of canonical DNA, the pairs are significantly nonplanar (by up to ~70°) and distorted with respect to the opening (by up to ~30°). The consistent weak and distorted interactions between the POB-T Hoogsteen face and the canonical nucleobases correlate with reports that pol η has a low dNTP selectivity during POB-T replication (26, 27).

To further understand the POB-T replication outcomes MD simulations on the pol η insertion complexes were performed. Specifically, dATP, dCTP, and dTTP were paired opposite the lesion in the pol η active site, since A, C, and T all form a hydrogen bond in the nucleobase pairs with POB-T. In all complexes, there is substantial variation in the bulky moiety conformation (Figure E.13, Appendix E). However, the lesion preferentially adopts an extended orientation that directs the bulky moiety away from the pairing base, toward the previously replicated DNA, and near residues in the little finger domain (i.e., Arg368, Met418, and Leu375). Additionally, as observed for pol κ, all pol η insertion complexes maintain octahedral coordination of the catalytic and binding Mg^{2+} ions (Table E.4, Appendix E), and the reaction distance and angle are < 3.5 Å and 170–175°, respectively (Figure 6.6). Furthermore, regardless of the pairing base, interactions between the dNTP backbone and surrounding amino acids are similar to those for the replication of

T (Figure E.12b and Tables E.5-E.8, Appendix E). Therefore, all complexes contain a catalytically conducive active site orientation.

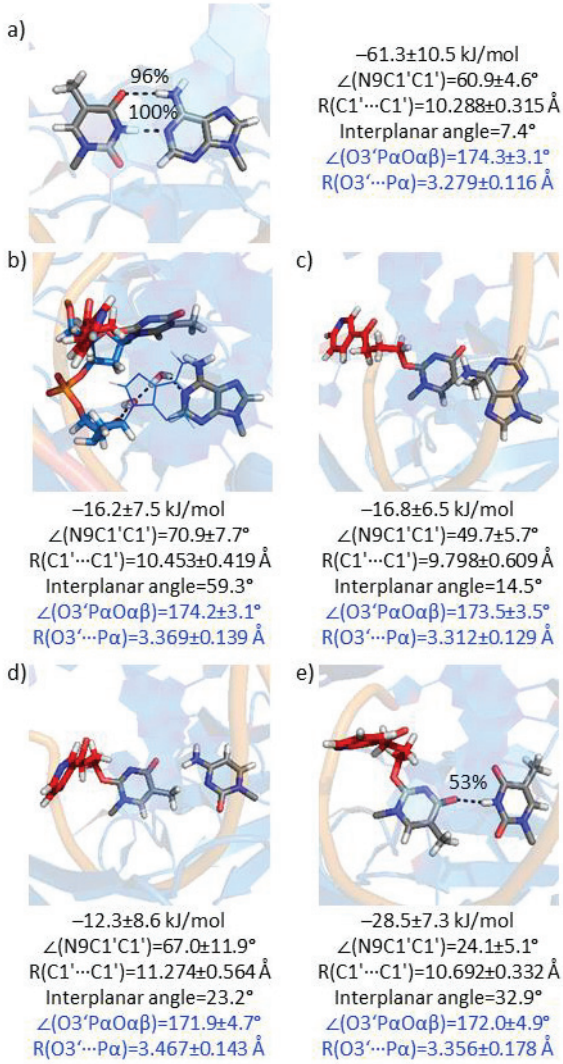

Figure 6.6. Structure and reaction coordinates for pol η insertion of dATP opposite a) T or b) POB-T, and c) *syn*-dATP, d) dCTP or e) dTTP insertion opposite POB-T.

Despite a consistent N–H···POB-T(O4) hydrogen bond in the lesion nucleobase pairs with A, C or T, differences occur in the interactions between the dNTP and *syn*-POB-T in the pol η active site. Indeed, steric constraints within the active site force *syn*-POB-T:*syn*-dATP and *syn*-POB-T:dCTP into geometries that lack interbase hydrogen bonding (Figure 6.6). This suggests that neither *syn*-dATP nor dCTP will be selectively incorporated opposite the

lesion by pol η. Interestingly, when dATP is placed opposite *syn*-POB-T, the lesion flips out of the helix into the major groove and toward the little finger domain of pol η. With *syn*-POB-T out of the helix, dATP is stabilized in the active site by a water-bridged hydrogen bond between dATP(N1) and the 3′-A(O4′) (Figure 6.6 and Figure E.13b, Appendix E). A crystal structure of dATP insertion opposite an abasic site by pol η shows a water-bridged hydrogen bond between the dATP nucleobase and the template backbone (55). This indicates that the experimentally-observed insertion of dATP opposite POB-T (26-28) likely occurs in a manner analogous to the replication of an abasic site.

In contrast to dCTP and dATP, dTTP paired opposite *syn*-POB-T by pol η leads to POB-T(O4)···dTTP(N3H) (53%) and POB-T(C7H)···dTTP(O2) interactions (~10%; Figure 6.6e), which contribute to an average interaction energy of –29 kJ/mol. Nevertheless, while the pair maintains a width consistent with natural DNA, it is not planar (interplanar angle=33°). However, since only dTTP directly interacts with *syn*-POB-T within the confines of the pol η active site, the structure of this pair was further probed using MD simulations on the corresponding post-replication DNA duplex. The helical environment leads to a more planar *syn*-POB-T:T pair (by ~15°) than the pol η active site, with the same POB-T(O4)···T(N3H) (92%) and POB-T(C7H)···dTTP(O2) (12%) hydrogen bonds, and stability (–32 kJ/mol; Figure 6.7). Since T is the only nucleobase that can directly interact with POB-T in the pol η active site, our simulations provide the first structural explanation for *in vitro* studies that indicate pol η preferentially inserts dTTP opposite POB-T (26, 27), as well as *in vivo* studies that report POB-T leads to T→A mutations (28). Our new proposal that pol η bypasses POB-T in the *syn* glycosidic conformation sheds light on the differential

POB-T replication outcomes for pol κ and η, and to the best of our knowledge provides the first evidence that a polymerase can replicate a damaged *syn* pyrimidine.

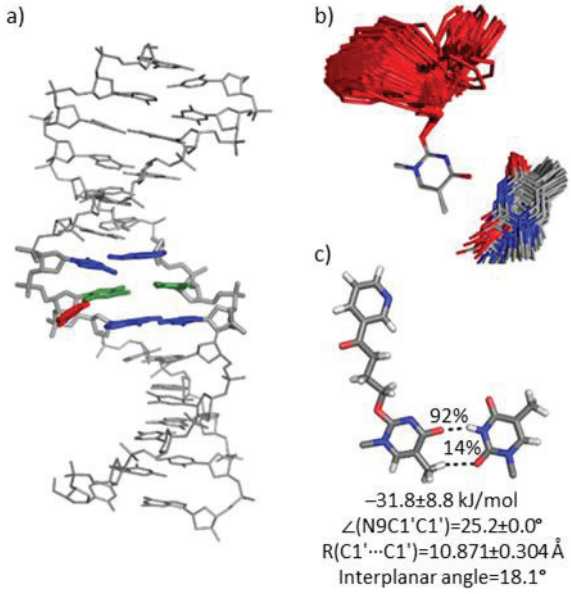

Figure 6.7. a) MD representative structure of adducted DNA containing POB-T:T. b) Overlay (based on damaged T heavy atoms) of adducted base pair throughout the MD simulation (bulky moiety in red). c) Lesion pair structure in DNA.

6.3. Conclusion

Our multiscale modeling provides the first structural information on the minor groove POB-T lesion, and thereby rationalizes the previously confusing mutational spectrum (26-28). Indeed, although pol κ typically replicates minor groove lesions, pol κ has been reported to insert dGTP opposite POB-T *in vitro,* but does not replicate the lesion *in vivo* (26-28). In contrast, *in vivo* work indicates that POB-T is replicated by pol η, which typically bypasses major groove lesions, and this replication leads to up to ~ 50% T→A mutations (26-28). Our results further clarify the role of these human polymerases by predicting that POB-T adopts different glycosidic orientations in the pol κ and η active

sites. Specifically, the *anti*-POB-T can fit in the pol κ active site and this POB-T orientation preferentially pairs with dGTP, which correlates with the experimentally observed replication outcome (26-28). Conversely, *anti*-POB-T cannot be accommodated by pol η and instead the *syn* conformation is adopted. In the *syn* conformation, the lesion preferentially pairs with dTTP, which correlates with previous reported *in vitro* and *in vivo* pol η data (26-28). We propose that the experimentally-observed non-mutagenic replication of POB-T by pol η occurs through stabilization of dATP opposite POB-T in a manner analogous to dATP insertion opposite an abasic site (55) rather than through conventional WC hydrogen bonding. All lesion conformations and associated hydrogen-bonding patterns between POB-T and the canonical nucleobases calculated in the present work differ from those previously hypothesized in the literature in the absence of accurate structural data, thus highlighting the need for this fundamental information. Overall, the current work provides the structural insights necessary to clarify an apparent discrepancy in the DNA replication literature by expanding the biological role of pol η. Furthermore, although other nucleic acid binding enzymes have been shown to bind pyrimidines in the *syn* glycosidic conformation (49-51), our results provide the first evidence for the replication of a *syn* pyrimidine. Most importantly, the current work provides structural support that indicates POB-T may lead to the T→A mutations observed in the mutational signature (signature 4) associated with tobacco smoking in human cancers (25).

6.4. Computational Details

6.4.1. DFT Calculations: Barriers for changing the α′, β′, γ′, δ′, ϵ′, and ρ′ dihedral angles in the bulky moiety of the POB-T nucleobase (Figure 6.1a) were determined using B3LYP-D3(BJ)/6-31G(d). To gain a more complete picture of possible POB-T nucleobase

orientations, an AMBER conformational search about α′, β′, γ′, δ′, ϵ′, and ρ′ as implemented in HyperChem was performed. All resulting lesion orientations were subsequently optimized using B3LYP-D3(BJ)/6-31G(d) and the relative energies determined using B3LYP-D3(BJ)/6-311++G(2df,2p). Hydrogen-bonded dimers between the WC or Hoogsteen face of POB-T and the canonical nucleobases were investigated using B3LYP-D3(BJ)/6-311++G(2df,2p)//M06-2X/6-31G(d). All calculations were performed using Gaussian 09 (revision D.01).

6.4.2. MD Simulations: The orientation of POB-T in the pol κ and η active sites was initially investigated by overlaying (adducted T heavy atoms) the nucleobase conformations identified in the conformational search onto the template T in a pol κ or η crystal structure (PDB IDs: 5W2C and 4ECS) corresponding to dATP insertion. Representative lesion orientations were then used to initiate MD simulations (AMBER ff14SB) on the pol κ and η preinsertion and insertion complexes with select dNTPs, as well as select 5′-CTGTGTC[POB-T]GCAGCG 14-mer DNA duplexes.

All complexes were solvated in a TIP3P octahedral water box such that the solute was at least 10.0 Å from the box edge. NaCl was added to DNA–polymerase systems to yield an ~ 0.150 M concentration, while DNA systems were neutralized with Na^+. Mg^{2+} ions were included in the insertion complexes. All systems were minimized, heated to 310 K, and equilibrated. 20 ns unconstrained pre-production simulations were performed to identify geometries that were distinctly different from the initial structures and cannot be easily converted during standard MD sampling. Structures with a hydrogen-bonding geometry and lesion orientation most conducive for the reaction were selected as initial structures for

the final 100 ns production simulations, which were run in triplicate. Since the replicas exhibit negligible all-atom rmsds (> 2.5 Å), data from one replica is discussed in the main text. While analysis was performed on the entire trajectory, a representative structure was chosen for figures based on the lesion orientation and hydrogen bonding. All MD simulations were performed using AMBER 14.

Full details of the computational protocol are provided in the Appendix E.

6.5. References

1. De Bont R & van Larebeke N (2004) Endogenous DNA damage in humans: A review of quantitative data. *Mutagenesis* 19(3):169-185.

2. Bilano V*, et al.* (2015) Global trends and projections for tobacco use, 1990–2025: An analysis of smoking indicators from the WHO comprehensive information systems for tobacco control. *The Lancet* 385(9972):966-976.

3. Harris CC (2018) Tobacco smoking, e-cigarettes, and nicotine harm. *Proc. Natl. Acad. Sci. U. S. A.* (DOI: 10.1073/pnas.1722636115).

4. Hang B (2010) Formation and repair of tobacco carcinogen-derived bulky DNA adducts. *Journal of Nucleic Acids* 2010:29.

5. Hecht SS (2003) Tobacco carcinogens, their biomarkers and tobacco-induced cancer. *Nat. Rev. Cancer* 3:733.

6. Lee H-W*, et al.* (2018) E-cigarette smoke damages DNA and reduces repair activity in mouse lung, heart, and bladder as well as in human lung and bladder cells. *Proc. Natl. Acad. Sci. U. S. A.*

7. Peterson LA (2016) Context matters: Contribution of specific DNA adducts to the genotoxic properties of the tobacco-specific nitrosamine NNK. *Chem. Res. Toxicol.*

8. Hecht SS (1998) Biochemistry, biology, and carcinogenicity of tobacco-specific N-nitrosamines. *Chem. Res. Toxicol.* 11(6):559-603.

9. Hecht SS (2008) Progress and challenges in selected areas of tobacco carcinogenesis. *Chem. Res. Toxicol.* 21(1):160-171.

10. Sale JE, Lehmann AR, & Woodgate R (2012) Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat. Rev. Mol. Cell Biol.* 13(3):141-152.

11. Xu L*, et al.* (2015) RNA polymerase II transcriptional fidelity control and its functional interplay with DNA modifications. *Crit. Rev. Biochem. Mol. Biol.* 50(6):503-519.

12. Patra A*, et al.* (2014) Kinetics, structure, and mechanism of 8-oxo-7,8-dihydro-2 '-deoxyguanosine bypass by human DNA polymerase eta. *J. Biol. Chem.* 289(24):16867-16882.

13. Kirouac KN, Basu AK, & Ling H (2013) Structural mechanism of replication stalling on a bulky amino-polycyclic aromatic hydrocarbon DNA adduct by a y family DNA polymerase. *J. Mol. Biol.* 425(22):4167-4176.

14. Stover JS, Chowdhury G, Zang H, Guengerich FP, & Rizzo CJ (2006) Translesion synthesis past the C8- and N2-deoxyguanosine adducts of the dietary mutagen 2-amino-3-methylimidazo[4,5-f]quinoline in the nari recognition sequence by prokaryotic DNA polymerases. *Chem. Res. Toxicol.* 19(11):1506-1517.

15. Zhang H*, et al.* (2009) Versatility of Y-family sulfolobus solfataricus DNA polymerase Dpo4 in translesion synthesis past bulky N2-alkylguanine adducts. *J. Biol. Chem.* 284(6):3563-3576.

16. Eoff RL, Angel KC, Egli M, & Guengerich FP (2007) Molecular basis of selectivity of nucleoside triphosphate incorporation opposite O6-benzylguanine by sulfolobus solfataricus DNA polymerase Dpo4: Steady-state and pre-steady-state kinetics and x-ray crystallography of correct and incorrect pairing. *J. Biol. Chem.* 282(18):13573-13584.

17. Choi J-Y*, et al.* (2006) Translesion synthesis across O6-alkylguanine DNA adducts by recombinant human DNA polymerases. *J. Biol. Chem.* 281(50):38244-38256.

18. Sherrer SM*, et al.* (2009) Mechanistic studies of the bypass of a bulky single-base lesion catalyzed by a Y-family DNA polymerase. *J. Biol. Chem.* 284(10):6379-6388.

19. Gadkari VV, Tokarsky EJ, Malik CK, Basu AK, & Suo Z (2014) Mechanistic investigation of the bypass of a bulky aromatic DNA adduct catalyzed by a Y-family DNA polymerase. *DNA Repair* 21:65-77.

20. Xu P, Oum L, Lee Y-C, Geacintov NE, & Broyde S (2009) Visualizing sequence-governed nucleotide selectivities and mutagenic consequences through a replicative cycle: Processing of a bulky carcinogen N2-dG lesion in a y-family DNA polymerase. *Biochemistry* 48(22):4677-4690.

21. Michel AK, Zarth AT, Upadhyaya P, & Hecht SS (2017) Identification of 4-(3-pyridyl)-4-oxobutyl-2′-deoxycytidine adducts formed in the reaction of DNA with 4-(acetoxymethylnitrosamino)-1-(3-pyridyl)-1-butanone: A chemically activated form of tobacco-specific carcinogens. *ACS Omega* 2(3):1180-1190.

22. Improta R, Santoro F, & Blancafort L (2016) Quantum mechanical studies on the photophysics and the photochemistry of nucleic acids and nucleobases. *Chem. Rev.* 116(6):3540-3593.

23. Lao Y, Yu N, Kassie F, Villalta PW, & Hecht SS (2007) Formation and accumulation of pyridyloxobutyl DNA adducts in f344 rats chronically treated with 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone and enantiomers of its metabolite, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol. *Chem. Res. Toxicol.* 20(2):235-245.

24. Upadhyaya P, Lindgren B, & Hecht SS (2009) Comparative levels of O6-methylguanine, pyridyloxobutyl-, and pyridylhydroxybutyl-DNA adducts in lung and liver of rats treated chronically with the tobacco-specific carcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK). *Drug Metabolism and Disposition*.

25. Alexandrov LB*, et al.* (2015) Clock-like mutational processes in human somatic cells. *Nat. Genet.* 47(12):1402-1407.

26. Gowda ASP, Krishnegowda G, Suo Z, Amin S, & Spratt TE (2012) Low fidelity bypass of O2-(3-pyridyl)-4-oxobutylthymine, the most persistent bulky adduct produced by the tobacco specific nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone by model DNA polymerases. *Chem. Res. Toxicol.* 25(6):1195-1202.

27. Gowda ASP & Spratt TE (2016) DNA polymerases η and ζ combine to bypass O(2)-[4-(3-pyridyl)-4-oxobutyl]thymine, a DNA adduct formed from tobacco carcinogens. *Chem. Res. Toxicol.* 29(3):303-316.

28. Weerasooriya S, Jasti VP, Bose A, Spratt TE, & Basu AK (2015) Roles of translesion synthesis DNA polymerases in the potent mutagenicity of tobacco-specific nitrosamine-derived O(2)-alkylthymidines in human cells. *DNA repair* 35:63-70.

29. Du H, Leng J, Wang P, Li L, & Wang Y (2018) Impact of tobacco-specific nitrosamine-derived DNA adducts on the efficiency and fidelity of DNA replication in human cells. *J. Biol. Chem.*

30. Hashimoto K, Ohsawa K-i, & Kimura M (2004) Mutations induced by 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) in the lacz and cii genes of muta™ mouse. *Mutat. Res./Genet. Toxicol. Environ. Mutagen.* 560(2):119-131.

31. Sandercock LE*, et al.* (2008) Mgmt deficiency alters the in vivo mutational spectrum of tissues exposed to the tobacco carcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK). *Carcinogenesis* 29(4):866-874.

32. Jasti VP, Spratt TE, & Basu AK (2011) Tobacco-specific nitrosamine-derived O2-alkylthymidines are potent mutagenic lesions in SOS-induced Escherichia coli. *Chem. Res. Toxicol.* 24(11):1833-1835.

33. Zhao L & Washington M (2017) Translesion synthesis: Insights into the selection and switching of DNA polymerases. *Genes* 8(1):24.

34. McCulloch SD*, et al.* (2004) Preferential cis-syn thymine dimer bypass by DNA polymerase eta occurs with biased fidelity. *Nature* 428(6978):97-100.

35. Yang W (2014) An overview of y-family DNA polymerases and a case study of human DNA polymerase η. *Biochemistry* 53(17):2793-2803.

36. Stefano M, Todor A, & J. SS (2017) Minor groove 3-deaza-adenosine analogues: Synthesis and bypass in translesion DNA synthesis. *Chem. Eur. J.* 23(5):1101-1109.

37. Choi J-Y, Angel KC, & Guengerich FP (2006) Translesion synthesis across bulky N2-alkyl guanine DNA adducts by human DNA polymerase κ. *J. Biol. Chem.* 281(30):21062-21072.

38. Wolfle WT*, et al.* (2005) Evidence for a Watson-Crick hydrogen bonding requirement in DNA synthesis by human DNA polymerase κ. *Mol. Cell. Biol.* 25(16):7137-7143.

39. Balbo S, Hecht SS, Upadhyaya P, & Villalta PW (2014) Application of a high-resolution mass-spectrometry-based DNA adductomics approach for identification of DNA adducts in complex mixtures. *Anal. Chem.* 86(3):1744-1752.

40. Wilson KA & Wetmore SD (2017) Molecular insights into the translesion synthesis of benzyl-guanine from molecular dynamics simulations: Structural evidence of mutagenic and nonmutagenic replication. *Biochemistry* 56(13):1841-1853.

41. Wilson KA & Wetmore SD (2017) Conformational flexibility of the benzyl-guanine adduct in a bypass polymerase active site permits replication: Insights from molecular dynamics simulations. *Chem. Res. Toxicol.* 30(11):2013-2022.

42. Lior-Hoffmann L, Ding S, Geacintov NE, Zhang Y, & Broyde S (2014) Structural and dynamic characterization of polymerase κ's minor groove lesion processing reveals how adduct topology impacts fidelity. *Biochemistry* 53(35):5683-5691.

43. Wang L & Broyde S (2006) A new anti conformation for N-(deoxyguanosin-8-yl)-2-acetylaminofluorene (AAF-dG) allows Watson-Crick pairing in the sulfolobus solfataricus p2 DNA polymerase IV (Dpo4). *Nucleic Acids Res.* 34(3):785-795.

44. Zhang L, Shapiro R, & Broyde S (2005) Molecular dynamics of a food carcinogen−DNA adduct in a replicative DNA polymerase suggest hindered nucleotide incorporation and extension. *Chem. Res. Toxicol.* 18(9):1347-1363.

45. Wang L*, et al.* (2005) Accommodation of a 1s-(−)-benzo[c]phenanthrenyl-N6-dA adduct in the y-family Dpo4 DNA polymerase active site: Structural insights through molecular dynamics simulations. *Chem. Res. Toxicol.* 18(3):441-456.

46. Vaisman A & Woodgate R (2017) Translesion DNA polymerases in eukaryotes: What makes them tick? *Crit. Rev. Biochem. Mol. Biol.* 52(3):274-303.

47. Norman D*, et al.* (1989) 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* 28(18):7462-7476.

48. Gu Z, Gorin A, Hingerty BE, Broyde S, & Patel DJ (1999) Solution structures of aminofluorene [AF]-stacked conformers of the syn [AF]−C8-dG adduct positioned opposite dC or dA at a template-primer junction. *Biochemistry* 38(33):10855-10870.

49. Hu L*, et al.* (2013) Crystal structure of tet2-DNA complex: Insight into tet-mediated 5mc oxidation. *Cell* 155(7):1545-1555.

50. Imamura K, Averill A, Wallace SS, & Doublié S (2012) Structural characterization of viral ortholog of human DNA glycosylase neil1 bound to thymine glycol or 5-hydroxyuracil-containing DNA. *J. Biol. Chem.* 287(6):4288-4298.

51. Sokoloski JE, Godfrey SA, Dombrowski SE, & Bevilacqua PC (2011) Prevalence of syn nucleobases in the active sites of functional rnas. *RNA* 17(10):1775-1787.

52. Brautigam CA & Steitz TA (1998) Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Curr. Opin. Struct. Biol.* 8(1):54-63.

53. Wang Y & Schlick T (2008) Quantum mechanics/molecular mechanics investigation of the chemical reaction in Dpo4 reveals water-dependent pathways and requirements for active site reorganization. *J. Am. Chem. Soc.* 130(40):13240-13250.

54. Wang L, Yu X, Hu P, Broyde S, & Zhang Y (2007) A water-mediated and substrate-assisted catalytic mechanism for sulfolobus solfataricus DNA polymerase IV. *J. Am. Chem. Soc.* 129(15):4731-4737.

55. Patra A*, et al.* (2015) Structural and kinetic analysis of nucleoside triphosphate incorporation opposite an abasic site by human translesion DNA polymerase η. *J. Biol. Chem.* 290(13):8028-8038.

Chapter 7: Exploring the Identity of the General Base for a DNA Polymerase Catalyzed Reaction Using QM/MM: The Case Study of Human TLS Polymerase η[a,b,c]

7.1. Introduction

DNA replication plays a large role in the accurate transmission and maintenance of genetic information, and therefore is fundamental to life. The process of DNA replication is carried out by enzymes known as DNA polymerases. Understanding the function of these enzymes is critical due to their health implications. Indeed, replication errors by polymerases are predicted to account for 66% of cancer driving mutations,[1] and changes to the function of DNA polymerases have been shown to unequivocally lead to cancer.[2-3] Furthermore, antibiotic and chemotherapy resistance has been linked to the action of DNA polymerases.[4-8] Seventeen human DNA polymerases have been identified to date, which vary in amino acid sequence, three-dimensional structure, and subunit composition.[9] Nevertheless, all polymerases are believed to share a metal ion (typically Mg^{2+}) mediated nucleotidyl transfer reaction.

---

[a] *Journal of the American Chemical Society* reference style used throughout this chapter.

[b] Adapted from Wilson, K.A., Fernandes, P.A., Ramos, M.J., and Wetmore, S.D. (2018). Exploring the Identity of the General Base for a DNA Polymerase Catalyzed Reaction Using QM/MM: The Case Study of Human TLS Polymerase η. *ACS Catalysis*. (submitted, cs-2018-04002m).

[c] K.A.W. performed calculations, data analysis and wrote the first draft of the manuscript. P.A.F., M.J.R. and S.D.W. played supervisory roles.

Many studies have provided key information about the reaction pathways catalyzed by a variety of polymerases.[10-11] The general proposed mechanism for the nucleotidyl transfer reaction consists of initial deprotonation of the primer 3′-OH by a general base to generate the activated 3′-oxyanion nucleophile (Figure 7.1). Subsequently, the 3′-oxyanion nucleophile attacks the α-phosphate of the incoming dNTP to form a phosphodiester bond and release pyrophosphate.[12] The roles of metal ions in the DNA polymerase mechanism have also been well studied.[12-15] Specifically, the catalytic Mg^{2+} ion lowers the pK_a of the primer 3′-OH, the binding Mg^{2+} ion stabilizes the negative charge on the leaving group pyrophosphate, and an additional Mg^{2+} ion has been proposed to potentially stabilize the transition state for pyrophosphate formation, assist in product release and/or prevent the reverse reaction. Despite the critical contributions of previous studies for understanding the overall mechanism, an important outstanding question is: what is the identity of the general base that initiates the reaction?

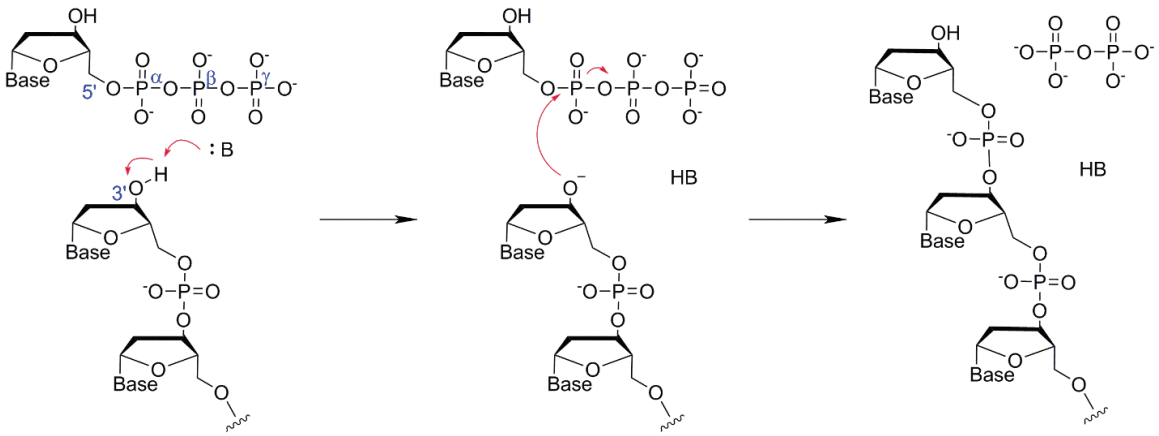

Figure 7.1. The proposed general reaction mechanism for nucleotide addition by a DNA polymerase.

Since identifying the general base with experimental techniques is challenging, several computational studies have characterized different mechanisms for the initial step in the polymerase catalyzed reaction.[13-14, 16-33] These works have each investigated a selection of the proposed mechanisms (typically one to three pathways) and have used a wide variety of methods ranging from quantum mechanical (QM) cluster approaches to empirical valence bond and quantum mechanics/molecular mechanics (QM/MM) calculations, with the size of the QM region typically being 52–80 atoms (with only two studies using a large QM region of 140 or 183 atoms). For polymerase (pol) β and λ, as well as T7 DNA polymerase, a conserved active site Asp/Glu residue has been proposed to play the role of the general base, either directly or assisted by water (Figure F.1a).[16-22] Alternatively, pol β has been proposed to invoke direct proton transfer to the dNTP α-phosphate (Figure F.1b).[23-24] Meanwhile, for many translesion synthesis (TLS) polymerases (i.e., Dpo4, as well as pol κ and ι), a water mediated substrate assisted (WMSA) mechanism has been suggested in which a series of water molecules transfer the primer 3′-OH proton to the dNTP α-phosphate group and then to the dNTP γ- or β-phosphate group through subsequent water-mediated proton transfer steps (Figure F.1c).[25-30] An active site water molecule that is hydrogen bonded to bulk water has also been considered as the general base for pol η (Figure F.1d),[13] while other studies on pol η and β have indicated that a hydroxide ion coordinated to the catalytic Mg^{2+} ion could play this role.[14, 31-32] Finally, a recent study proposed a self-activated mechanism (SAM) in which the leaving pyrophosphate deprotonates the primer 3′-OH (Figure F.1e) and an interaction between the 3′-oxyanion and the catalytic Mg^{2+} ion mediates subsequent movement of the polymerase along the DNA to align the primer for dNTP binding (translocation).[33] Overall, these previous studies on the reaction mechanisms catalyzed by DNA polymerases, highlight the discrepancy in

the literature surrounding the identity of the general base. Indeed, recent literature has emphasized the need for future studies to establish the identity of the general base.[13-14]

The current work aims to fill this gap in the literature by examining all previously proposed mechanisms using a consistent theoretical framework. Human pol η was chosen as the model for the current investigation due to the established biomedical importance of this enzyme. Specifically, pol η specializes in rescuing stalled replication forks at UV-induced pyrimidine dimers, as well as many other DNA lesions, by extending the primer past the site of damage.[34-35] Furthermore, pol η is the only DNA polymerase in which mutations to the enzyme or the lack of the polymerase has been explicitly linked with cancer.[2, 6] Indeed, lack of pol η causes the human syndrome xeroderma pigmentosum (XPV), which is characterized by elevated rates of skin malignancies.[36] Conversely, the presence of pol η can decrease cell sensitivity to anticancer agents such as cisplatin.[37-38] Due to its biomedical importance, a number of experimental studies have investigated the function of pol η, reporting observed k_{cat} values of 0.23 – 3.33 s^{-1},[15, 39-45] which correspond to an upper bound for the overall barrier for all reaction steps (including binding, conformational change(s), chemical step(s) and product release) of ~80 kJ/mol according to transition state theory. Additionally, these studies have proposed bulk water or Ser113 to be the general base that initiates the pol η catalyzed reaction.[15]

To explore the mechanism for the nucleotidyl transfer reaction catalyzed by human pol η, the current work uses molecular dynamics (MD) simulations, umbrella sampling, and ONIOM QM/MM calculations. Specifically, we considered all possibilities for the general base based on previous proposals from experimental and computational studies on pol η or

other DNA polymerases. Our results show that proton transfer to a hydroxide ion or to Glu116 through Ser113 leads to calculated barriers consistent with experimental kinetic data.[15, 39-45] However, a new pathway for the nucleotidyl transfer reaction is identified in this work that invokes direct proton abstraction by Glu116, which is the most thermodynamically favored pathway, and is fully consistent with all experimental kinetic and mutational data. In line with the common proposal that all DNA polymerases share a general mechanism,[9] Glu116 acting as the general base for the pol η catalyzed reaction is the first evidence that supports a TLS polymerase using a mechanism similar to that proposed for other DNA polymerases in which a conserved Asp/Glu initiates the reaction through proton abstraction.[16-22]

7.2. Computational Details

7.2.1. Structural Preparation: The X-ray crystal structure corresponding to pol η insertion of dATP opposite a template T in the presence of Ca^{2+} (PDB ID: 4ECS, 1.95 Å resolution) was used as a starting point for modeling. Missing residues (Thr155-Glu159), which are located in a loop region far away from the active site, were added manually using GaussView and the active site Ca^{2+} ions were converted to Mg^{2+} ions. All amino acids and nucleotides were assigned physiological protonation states (verified by PROPKA)[46] using the tLEAP module of AMBER 12.[47] dATP was modeled in the –4 charge state. The system was neutralized by Na^{+} ions and surrounded by a cubic box of TIP3P water such that the minimum distance between the box edge and solute is 10.0 Å.

7.2.2. Classical MD Simulations: The system was minimized in stages using AMBER 12. Specifically, water and counterions were minimized over 2500 steps of steepest descent and 2500 steps of conjugate gradient minimization with a 50 kcal/(mol·Å^2) constraint on the solute. Subsequently, the hydrogen atoms were minimized over 4000 steps of steepest descent and 4000 steps of conjugate gradient minimization, with a 50 kcal/(mol·Å^2) constraint on all heavy atoms. Next, 12500 steps of steepest descent and 12500 steps of conjugate gradient minimization were performed, with a 50 kcal/(mol·Å^2) constraint on the DNA and protein backbone. Finally, 30000 steps of unconstrained steepest descent and 30000 steps of unconstrained conjugate gradient minimization were performed. The minimized system was then heated to 310 K over 100 ps using the Langevin thermostat (γ=2.0), with a 10 kcal/(mol·Å^2) restraint on the solute. Finally, the system was equilibrated for 0.5 ns and simulated for 60 ns in an NPT ensemble at 310 K and 1 bar using the Langevin thermostat (γ=2.0) and Berendsen Barostat. For all calculations, a time step of 2 fs, the Particle Mech Ewald (PME) method was used to treat long-range electrostatic interactions, a non-bonded cutoff of 10.0 Å, the periodic boundary condition and SHAKE were implemented. All amino acids, nucleotides, Na^+ ions and the solvent were modeled with AMBER ff99SB parameters, while Mg^{2+} was modeled with the parameters from Allner et al,[48] and dATP parameters were adapted from the literature.[49]

7.2.3. Umbrella Sampling: Starting from the equilibrated pol η structure, umbrella sampling was used to investigate the orientation of the primer 3′-OH, which dictates the residue that is correctly aligned for the initial proton transfer step and is therefore critical to the overall reaction mechanism. Specifically, the ∠(H3′–C3′–O3′–HO3′) dihedral angle

was successively sampled in 5° intervals from 0 to 360°. Each window was minimized for 500 steps of steepest descent minimization followed by 1500 steps of conjugate gradient minimization, equilibrated for 50 ps and simulated for 2 ns, with a 30 kcal/(mol·rad^2) force constraint to ensure adequate sampling. The potential of mean force (PMF) was calculated for rotation about the dihedral angle using the weighted histogram analysis method (WHAM).[50]

7.2.4. QM/MM Calculations: The free energy profile for the nucleotidyl transfer reaction catalyzed by pol η was modeled using ONIOM QM/MM as implemented in Gaussian 09 (Rev. D.01).[51] The starting reactant geometries were taken from the minimized crystal structure or a structure from the umbrella sampling. To obtain a reactant, all Na^+ ions and water molecules beyond a 4 Å shell around the pol η–DNA complex were removed. The QM (high-level) region includes two to five water molecules, the dNTP (excluding the nucleobase), the 3′-primer terminus (C1′, H1′, C2′, H2′1, H2′2, C3′, H3′, O3′, O3′H, C4′, H4′, and O4′), the binding and catalytic Mg^{2+} ions, the amino acids coordinated to the Mg^{2+} ions (Asp13 (side-chain), Asp115 (side-chain), Glu116 (side-chain), and Met14 (backbone)), and the amino acids interacting with the triphosphate tail (Tyr52 (side-chain), Cys16 (backbone), Asp15 (backbone), Arg55 (side-chain) and Lys231 (side-chain; Figure 7.2). Depending on the reaction considered, the side-chains of Arg61 and/or Ser113, as well as the 5′-phosphate of the 3′-primer terminus, were also included in the high-level region. Therefore, the ONIOM model contained 127–140 atoms in the high-level region. All other polymerase, DNA and water atoms were modeled in the low-level (MM) region. Hydrogen-link atoms were used for covalent bonds that spanned the QM/MM boundary. All water molecules in the low-level region were frozen, while all other atoms were mobile during

the calculations. Preliminary linear scans along the reaction coordinate for each chemical step were performed using an increment of 0.10 Å, while finer scans with an increment of 0.01 Å were used in the transition state region. Where possible, the identified reactant (RC), intermediate (IC) and product (PC) complexes, as well as transition states (TS1 and TS2), were subsequently fully optimized. Stationary points were verified through vibrational frequency calculations, with the one imaginary frequency found for each transition state being visually inspected to verify the reaction coordinate. All calculations were performed using ONIOM(B3LYP-D3/6-311+G(2d,2p):AMBER ff99SB)// ONIOM(B3LYP/6-31G(d):AMBER ff99SB) and electronic embedding. Gibbs free energies and zero-point corrected energies were calculated by adding the thermal or scaled (0.9813) zero-point correction obtained from the frequency calculations to the ONIOM energies of each complex.

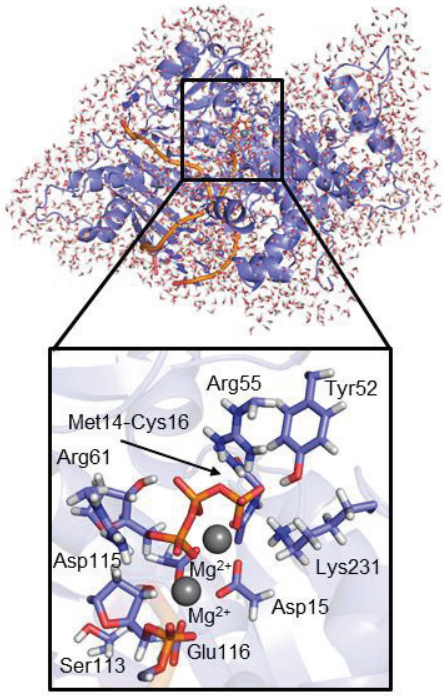

Figure 7.2. Solvated QM/MM model for pol η, with the ONIOM high-level region highlighted in tube representation.

7.3. Results and Discussion

The present work uses a consistent approach to systematically consider all possibilities for the general base in the nucleotidyl transfer reaction catalyzed by pol η (Figure 7.1) based on previous proposals from experimental and computational studies on this and other polymerases. Specifically, the pathways considered are primer 3′-OH deprotonation by 1) an active site water molecule that is hydrogen bonded to bulk water (denoted as bulk solvent),[13, 15] 2) the α-phosphate of dATP,[23-24] 3) water coordinated to the catalytic Mg^{2+} (i.e., the WMSA mechanism),[25-30] 4) a hydroxide ion coordinated to the catalytic Mg^{2+},[14, 31-32] 5) Glu116 facilitated by Ser113,[15] and 6) Glu116 directly (i.e., a conserved active site Asp/Glu).[16-22] Although each reaction mechanism differs in the general base, all pathways lead to an intermediate complex containing an activated 3′-oxyanion nucleophile, which subsequently attacks the α-phosphate of the incoming dATP to form a phosphodiester bond and generate pyrophosphate.

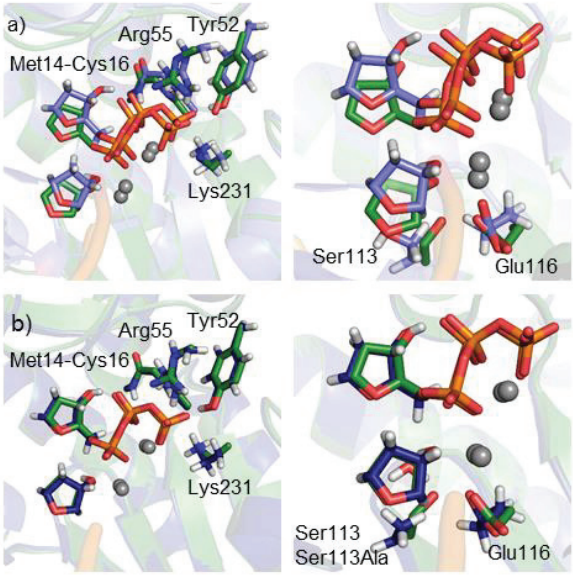

Figure 7.3. Comparison of the initial crystal structure geometry (PDB ID: 4ECS, green) and a MD representative structure from 60 ns simulation on the a) wild-type enzyme (light purple) or b) the Ser113Ala mutant enzyme (dark purple), showing residues that interact with the dNTP (left) or the 3′-primer terminus (right).

Bulk Solvent as the General Base: Several active site structural features are known to be required for a polymerase–DNA complex to be catalytically active, including the template:dATP hydrogen bonding, alignment of the reaction coordinate and coordination of the metal ions.[28-29, 52] During the classical MD simulation on the pol η complex corresponding to dATP insertion opposite T, Watson-Crick hydrogen bonding between the template T and the incoming dATP persists, and several hydrogen bonds occur between the dATP backbone and surrounding amino acids (i.e., Arg55, Lys231, Tyr52, Phe18, Phe17 and Cys16; Figure F.3 and 7.3). Additionally, both the catalytic and binding Mg^{2+} ions maintain octahedral coordination, and the distance (angle) between the primer 3′-O and the dNTP(Pα) is 3.3 Å (174°; Figure F.3). Throughout the classical MD simulation, the primer 3′-OH is directed toward and hydrogen bonds with bulk solvent (Figure F.3), which is also the global minimum energy orientation identified from umbrella sampling (Figure 7.4 and F.4). This is consistent with the general orientation and energetic trends previously reported for pol κ using a similar umbrella sampling methodology,[25] and with the conformations reported in a previous computational study on pol η.[14] Additionally, this orientation of the primer 3′-OH aligns with the previous suggestion that bulk solvent may act as a general base due to the presence of water in the vicinity of the primer 3′-OH according to time-lapsed X-ray crystallography.[15] Nevertheless, when proton transfer to bulk solvent is modeled with QM/MM, the mechanism proceeds to a high energy IC that lies ~155 kJ/mol above the corresponding RC (Table 7.1 and Figure 7.5). This is much greater than the estimated experimental barrier for pol η of ~80 kJ/mol,[15, 39-45] and therefore bulk water is unlikely the general base.

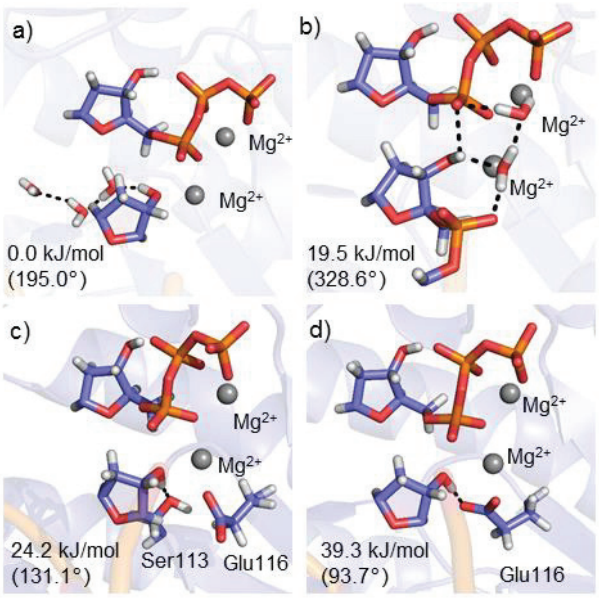

Figure 7.4. Reactant complexes obtained from umbrella sampling, with the energy and (∠(H3′–C3′–O3–HO3′) torsion angle, in parentheses) provided.

The α-phosphate of dNTP as the General Base: Based on umbrella sampling, a local minimum energy orientation of the primer 3′-OH contains a hydrogen bond between the hydroxy group and the α-phosphate (ΔG ~ 20 kJ/mol), and the barrier for rearrangement to this orientation from the global minimum energy conformation is ~25 kJ/mol (Figure 7.4 and S2). This orientation of the primer 3′-OH is also consistent with the general orientations and energetic trends previously reported in the literature.[14, 25] When the α-phosphate of dATP acts as the general base, the energy continuously increases during proton transfer, with the IC that contains the proton on the α-phosphate being ~113 kJ/mol higher in energy than the corresponding RC (Table 7.1 and Figure F.6). Since this energy difference is substantially higher than the estimated experimental barrier (~80 kJ/mol),[15, 39-45] it is unlikely that the dNTP α-phosphate will play the role of the general base.

Table 7.1. ONIOM(B3LYP-D3/6-311+G(2d,2p):AMBER ff99SB)// ONIOM(B3LYP/6-31G(d):AMBER ff99SB) reaction energies (kJ/mol) for the DNA pol η catalyzed nucleotidyl transfer reaction involving different general bases for the initial primer 3′-OH deprotonation step.

General Base	RC	RC*	TS1	IC	TS2	PC
Bulk Water	0.0	–	–	(157.9)[a]	–	–
dATP	0.0	19.5[b]	–	(127.0)[a]	–	–
Mg^{2+} Water (dATP)	0.0	19.5[b]	–	(92.0)[a]	–	–
Mg^{2+} Water (5'-phosphate)	0.0	19.5[b]	–	(90.9)[a]	–	–
Ser113 (Glu116)	0.0	24.2[b]	67.7 [67.5][c]	51.2 [56.8][c]	59.0 [58.4][c]	–21.9 [–23.6][c]
Water (Glu116)[d]	0.0	–	65.1 [64.0][c]	72.6 [73.4][c]	88.5 [84.7][c]	41.6 [41.8][c]
Glu116	0.0	39.3[b]	43.5 [38.1][c]	48.8 [44.5][c]	71.9 [64.7][c]	–54.5 [–57.0][c]

[a]Constrained optimization relative energy. [b]Free energy required to reorient the lowest energy reactant identified using umbrella sampling. [c]ΔG and [ΔE]. [d]Ser113Ala mutant pol η enzyme.

Proton Abstraction by Water Coordinated to the Catalytic Mg^{2+} (WMSA Mechanism): The local minimum energy orientation of the primer 3′-OH from umbrella sampling that contains a hydrogen bond between the primer 3′-OH and the α-phosphate is also aligned for the WMSA pathway proposed for other TLS polymerases.[25-30] Specifically, the primer 3′-OH is 1.520 Å away from a water molecule that is coordinated to the catalytic Mg^{2+} ion and this Mg^{2+} coordinated water hydrogen bonds with another water molecule that interacts with the dATP α-phosphate (Figure F.7). Nevertheless, when the Mg^{2+} coordinated water acts as the general base to pass the proton from the primer 3′-OH to the dATP α-phosphate through a water chain, the energy continuously increases with the corresponding IC being ~250 kJ/mol above the RC (Table 7.1). However, the Mg^{2+} coordinated water forms a second hydrogen bond with the 5′-phosphate group of the 3′-primer terminus, which was originally included in the QM/MM low-level region as done in previous computational studies on other polymerases.[25-30] To ensure the high energy is not a computational artifact, we reconsidered this mechanism using a model that includes the 5′-phosphate in the high-

level region (Figure F.8). In the resulting pathway, the terminal 3′-OH proton is shuttled through the Mg^{2+} coordinated water molecule to the 5′-phosphate of the primer 3′-terminal nucleotide in an uphill reaction, leading to an IC of ~80 kJ/mol (Figure F.7 and Table 7.1). Thus, as discussed for direct deprotonation by the α-phosphate or bulk solvent, the WMSA mechanism seems unlikely for pol η.

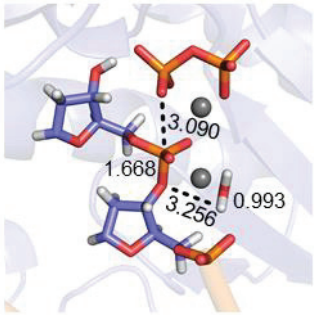

Figure 7.5. ONIOM product complex and key reaction distances for the nucleotidyl transfer reaction catalyzed by DNA pol η involving a hydroxide ion acting as the general base.

Hydroxide Ion as the General Base: While the Mg^{2+} coordinated water molecule is not able to assist the initial proton transfer, it has been proposed that this water molecule may be replaced by a hydroxide ion.[14, 31-32] Upon alignment of the primer 3′-OH to hydrogen bond with a Mg^{2+} coordinated hydroxide ion, barrierless conversion to the product complex occurs (Figure 7.5). This is consistent with the feasibility of this pathway reported in recent work on pol η.[14] Although our calculations suggest that a hydroxide ion bound to the catalytic Mg^{2+} ion could deprotonate the primer 3′-OH, hydroxide is present at a substantially lower concentration than water in the cell.[53] Furthermore, there will be an energetic cost to align the active site for the reaction, including the binding of a hydroxide ion to the catalytic Mg^{2+} ion. Although the free energy penalty for exchange of a water molecule for a hydroxide ion in a similar active site has been estimated to be

~11 kJ/mol,[54] hydroxide ions exhibit high reactivity toward DNA and proteins.[55-57] Therefore, a metal bound hydroxide may not be the ideal general base for this reaction.

Proton Abstraction by Glu116 Assisted by Ser113: Based on the proximity of Ser113 to the primer 3′-OH and Glu116 in a crystal structure of a pol η reactant mimic, Ser113 has been proposed to mediate an initial proton transfer from 3′-OH to Glu116.[15] Therefore, a reaction mechanism for proton abstraction by Glu116 assisted by Ser113 was investigated based on the minimized crystal structure geometry. In the RC, Ser113 simultaneously hydrogen bonds with Glu116 through the side-chain oxygen that is also coordinated to the catalytic Mg^{2+} ion (Figure 7.6). The first step of the reaction occurs with a free energy barrier of 43.5 kJ/mol, leading to an IC that is 51.2 kJ/mol higher in energy than the RC (Table 7.1). The free energy barrier for the second step is 7.8 kJ/mol, which leads to an exergonic reaction (PC = –21.9 kJ/mol). Nevertheless, in contrast to the crystal structure, Ser113 is consistently directed away from the active site, and no longer positioned between the primer 3′-OH and Glu116 throughout the classical MD simulation initiated from the same crystal structure geometry (Figure 7.3). Furthermore, although a complex with the 3′-OH hydrogen bonded to Ser113 can be identified that is ~24 kJ/mol higher in energy than the global minimum orientation, a complex in which Ser113 interacts with Glu116 does not occur during umbrella sampling. In summary, the overall calculated barrier (~75 kJ/mol including active site rearrangement) is consistent with the experimental barrier and indicates that Ser113 may be able to mediate the proton transfer step of the nucleotidyl transfer reaction. Nevertheless, the dynamic structural data (MD and umbrella sampling) for the RC* suggests Ser113 is not aligned to play a critical role in the reaction, which is consistent with mutational data indicating that the Ser113Ala mutant retains 95% of k_{cat}.[15]

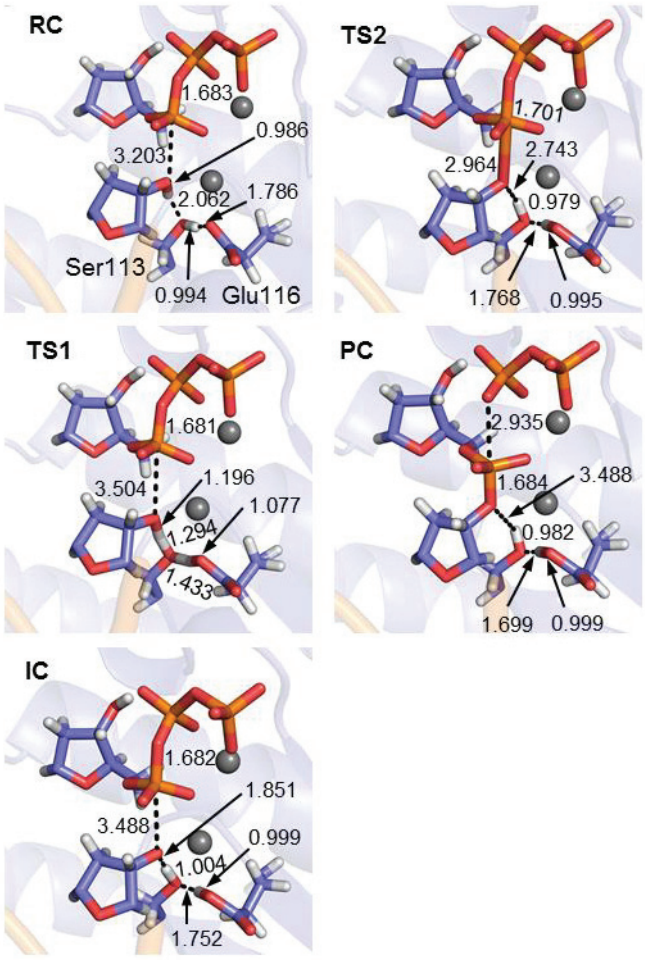

Figure 7.6. ONIOM stationary points and key reaction distances for the nucleotidyl transfer reaction catalyzed by DNA pol η involving Ser113 mediated proton transfer to Glu116.

To further investigate the role of Ser113 in the pol η mechanism, MD simulations and subsequent QM/MM calculations were performed on the Ser113Ala mutant. Classical MD simulations reveal that a water molecule bridges the interaction between the primer 3′-OH and Glu116 in the absence of Ser113 (Figure 7.3b). All other contacts in the wildtype complex are maintained in the mutant RC (Figure 7.3). When the corresponding reaction pathway is mapped, the first step occurs with a free energy of activation of 65.1 kJ/mol, resulting in an IC that is 72.6 kJ/mol higher in energy than the RC (Figure 7.7 and Table 7.1). Finally, the free energy barrier for the second step is 88.5 kJ/mol, which leads to an

overall endergonic (by 41.6 kJ/mol) reaction pathway for the Ser113Ala mutant. Experimentally the Ser113Ala mutant retains the same catalytic activity as the wildtype pol η complex and our calculated barrier is consistent with the experimental barrier (~80 kJ/mol),[15] which indicates that water may play the role of the general base, replacing Ser113 in the Ser113Ala mutant.

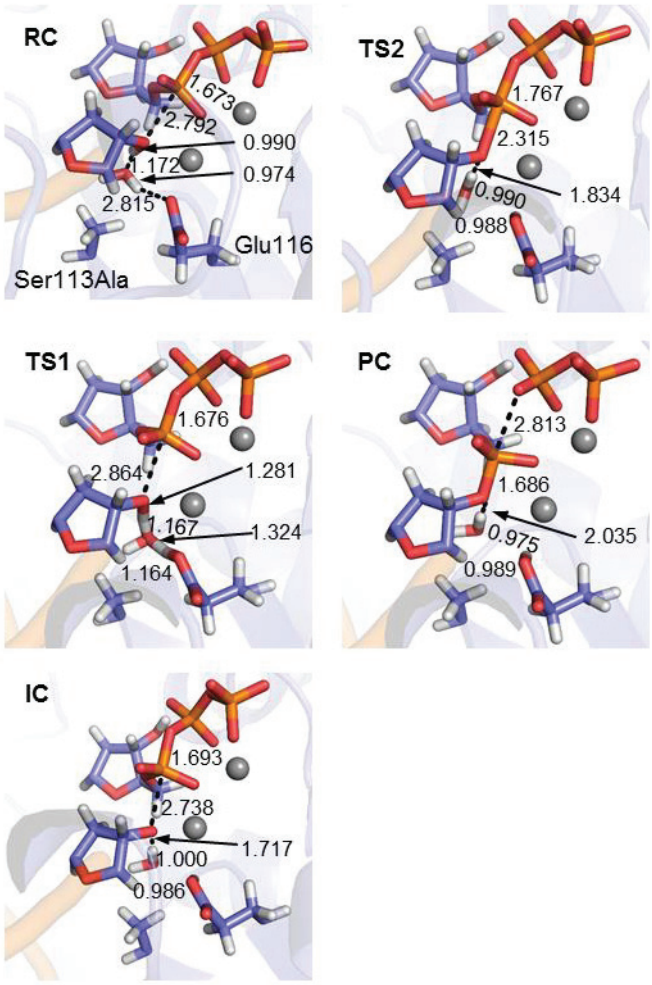

Figure 7.7. ONIOM stationary points and key reaction distances for the nucleotidyl transfer reaction catalyzed by the Ser113Ala pol η mutant involving water mediated proton transfer to Glu116.

Glu116 Directly Abstracting the Proton at the 3′-terminus: The final literature proposal for the general base in the nucleotidyl transfer reaction catalyzed by a DNA polymerase is a conserved active site Asp/Glu (Glu116 in pol η).[16-22] Based on umbrella sampling on pol η, the active site configuration that aligns Glu116 to act as the general base is 39.3 kJ/mol higher in energy than the global minimum (Figure 7.4). In this structure, one side-chain oxygen of Glu116 interacts with the catalytic Mg^{2+}, while the other oxygen hydrogen bonds with the primer 3′-OH. This reactant geometry more favorably aligns Glu116 to act as the general base than that modeled for other polymerases in which the same side-chain oxygen of the catalytic Asp/Glu simultaneously interacts with the catalytic Mg^{2+} and the primer 3′-OH.[16, 18, 21] Upon considering Glu116 of pol η as the general base, barrierless proton transfer leads to an IC that is only 9.5 kJ/mol higher in energy than the aligned RC* (Figure 7.8 and Table 7.1). Additionally, the barrier for phosphodiester bond formation and pyrophosphate generation is 23.1 kJ/mol and the reaction is exergonic (by 54.5 kJ/mol). Thus, this pathway leads to an overall barrier (~67 kJ/mol, including active site rearrangement) that is consistent with experimental data (barrier ~ 80 kJ/mol),[15, 39-45] which indicates that Glu116 may act as the general base.

Compared to the other reaction mechanisms characterized in the current work, the pathway in which Glu116 acts as the general base is the most thermodynamically favored, the only pathway consistent with the (MD and umbrella sampling) dynamical data, and consistent with all available experimental kinetic and mutational data for pol η.[15, 39-45] Our proposal contrasts conclusions from previous computational studies on other TLS polymerases that have not fully considered the dynamical structure of the active site.[25, 29] However, direct

involvement of Glu116 as the general base aligns the pol η mechanism with those for other DNA polymerases (pol β and λ).[16-22] Therefore, this work contributes to our understanding of the chemical details of TLS and points to a shared mechanism for different classes of polymerases.

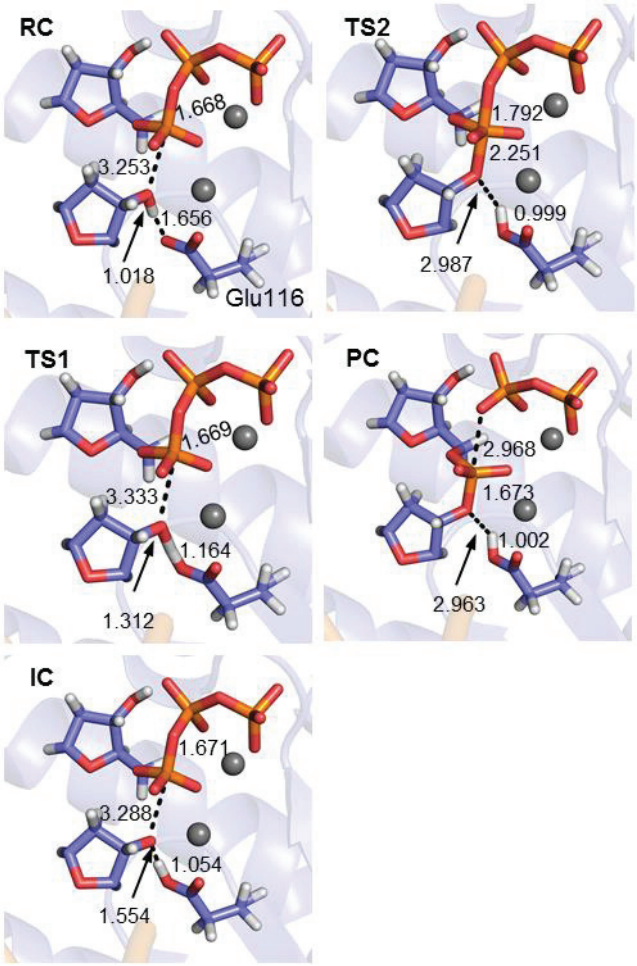

Figure 7.8. ONIOM stationary points and key reaction distances for the nucleotidyl transfer reaction catalyzed by DNA pol η involving Glu116 acting as the general base.

7.4. Conclusion

A consistent theoretical framework was used in the present work to investigate the range of previous proposals for the general base in the nucleotidyl transfer reaction catalyzed by

pol η. This comprehensive investigation of the pol η catalyzed reaction indicates that the dNTP α-phosphate moiety, bulk water, and the Mg^{2+} bound water molecule are not optimal choices for the general base due to the high predicted energy of the resulting intermediate complexes. In contrast, our calculations suggest that a pathway in which a hydroxide ion coordinated to the catalytic Mg^{2+} ion acts as a general base is consistent with available experimental kinetic data.[15, 39-45] However, the contribution of this pathway may be limited by the low availability in the pol η active site. Furthermore, it is energetically feasible for Ser113 to assist the first chemical step by mediating proton transfer to Glu116, and a water molecule could play the role of Ser113 in the Ser113Ala mutant. Nevertheless, classical MD simulations and umbrella sampling predict that Ser113 is preferentially directed away from the active site such that the side chain is no longer poised near the primer 3′-OH or Glu116, which indicates that Ser113 may not be the ideal general base. Finally, direct proton transfer from the primer 3′-OH to Glu116 leads to the most thermodynamically favorable pathway, and is the only pathway consistent with the dynamical data (i.e., classical MD and umbrella sampling) collected for pol η obtained in the present work, and available experimental kinetic and mutational data.[15, 39-45] Additionally, while all DNA polymerases are believed to share a general mechanism,[9-11] this is the first evidence that aligns the mechanism of a TLS polymerase with that proposed for other DNA polymerases (pol β and λ) in which a conserved Asp/Glu acts as the general base.[16-22] Taken together, our data contributes to answering a long-standing question in the DNA replication field regarding the identity of the general base for the nucleotidyl transfer reaction catalyzed by TLS polymerases. Future work should strive to determine whether our proposed reaction mechanism is the most favorable pathway for a broad range of other TLS polymerases.

7.5. References

1. Tomasetti, C.; Li, L.; Vogelstein, B., Stem Cell Divisions, Somatic Mutations, Cancer Etiology, and Cancer Prevention. *Science* **2017,** 355 (6331), 1330-1334.

2. Rogozin, I. B.; Goncearenco, A.; Lada, A. G.; De, S.; Yurchenko, V.; Nudelman, G.; Panchenko, A. R.; Cooper, D. N.; Pavlov, Y. I., DNA Polymerase H Mutational Signatures Are Found in a Variety of Different Types of Cancer. *Cell Cycle* **2018,** 17 (3), 348-355.

3. Loeb, L. A.; Monnat Jr, R. J., DNA Polymerases and Human Disease. *Nature Reviews Genetics* **2008,** 9 (8), 594.

4. Xie, K.; Doles, J.; Hemann, M. T.; Walker, G. C., Error-Prone Translesion Synthesis Mediates Acquired Chemoresistance. *Proc. Natl. Acad. Sci. U.S.A.* **2010,** 107 (48), 20792-20797.

5. Korzhnev, D. M.; Hadden, M. K., Targeting the Translesion Synthesis Pathway for the Development of Anti-Cancer Chemotherapeutics. *Journal of medicinal chemistry* **2016,** 59 (20), 9321-9336.

6. Lange, S. S.; Takata, K.; Wood, R. D., DNA Polymerases and Cancer. *Nature reviews cancer* **2011,** 11 (2), 96-110.

7. Warren, T. K.; Wells, J.; Panchal, R. G.; Stuthman, K. S.; Garza, N. L.; Van Tongeren, S. A.; Dong, L.; Retterer, C. J.; Eaton, B. P.; Pegoraro, G., Protection against Filovirus Diseases by a Novel Broad-Spectrum Nucleoside Analogue Bcx4430. *Nature* **2014,** 508 (7496), 402.

8. Bae, B.; Nayak, D.; Ray, A.; Mustaev, A.; Landick, R.; Darst, S. A., Cbr Antimicrobials Inhibit RNA Polymerase Via at Least Two Bridge-Helix Cap-Mediated Effects on Nucleotide Addition. *Proc. Natl. Acad. Sci. U.S.A.* **2015,** 112 (31), E4178-E4187.

9. Bebenek, K.; Kunkel, T. A. In *Advances in Protein Chemistry*, Elsevier: 2004; Vol. 69, pp 137-165.

10. Walker, A. R.; Cisneros, G. A., Computational Simulations of DNA Polymerases: Detailed Insights on Structure/Function/Mechanism from Native Proteins to Cancer Variants. *Chem. Res. Toxicol.* **2017,** 30 (11), 1922-1935.

11. Guengerich, F. P.; Zhao, L.; Pence, M. G.; Egli, M., Structure and Function of the Translesion DNA Polymerases and Interactions with Damaged DNA. *Perspect. Sci.* **2015,** 4 (0), 24-31.

12. Raper, A. T.; Reed, A. J.; Suo, Z., Kinetic Mechanism of DNA Polymerases: Contributions of Conformational Dynamics and a Third Divalent Metal Ion. *Chemical Reviews* **2018,** 118 (12), 6000-6025.

13. Hanwool, Y.; Arieh, W., Simulating the Fidelity and the Three Mg Mechanism of Pol H and Clarifying the Validity of Transition State Theory in Enzyme Catalysis. *Proteins* **2017,** 85 (8), 1446-1453.

14. Stevens, D. R.; Hammes-Schiffer, S., Exploring the Role of the Third Active Site Metal Ion in DNA Polymerase H with QM/MM Free Energy Simulations. *J. Am. Chem. Soc.* **2018,** 140 (28), 8965-8969.

15. Nakamura, T.; Zhao, Y.; Yamagata, Y.; Hua, Y.-J.; Yang, W., Watching DNA Polymerase Eta Make a Phosphodiester Bond. *Nature* **2012,** 487 (7406), 196-U77.

16. Florián, J.; Goodman, M. F.; Warshel, A., Computer Simulation of the Chemical Catalysis of DNA Polymerases: Discriminating between Alternative Nucleotide Insertion Mechanisms for T7 DNA Polymerase. *J. Am. Chem. Soc.* **2003,** 125 (27), 8163-8177.

17. Rittenhouse, R. C.; Apostoluk, W. K.; Miller, J. H.; Straatsma, T., Characterization of the Active Site of DNA Polymerase B by Molecular Dynamics and Quantum Chemical Calculation. *Proteins* **2003,** 53 (3), 667-682.

18. Lin, P.; Pedersen, L. C.; Batra, V. K.; Beard, W. A.; Wilson, S. H.; Pedersen, L. G., Energy Analysis of Chemistry for Correct Insertion by DNA Polymerase B. *Proc. Natl. Acad. Sci. U.S.A.* **2006,** 103 (36), 13294-13299.

19. Radhakrishnan, R.; Schlick, T., Correct and Incorrect Nucleotide Incorporation Pathways in DNA Polymerase B. *Biochemical and biophysical research communications* **2006,** 350 (3), 521-529.

20. Alberts, I. L.; Wang, Y.; Schlick, T., DNA Polymerase B Catalysis: Are Different Mechanisms Possible? *J. Am. Chem. Soc.* **2007,** 129 (36), 11100-11110.

21. Batra, V. K.; Perera, L.; Lin, P.; Shock, D. D.; Beard, W. A.; Pedersen, L. C.; Pedersen, L. G.; Wilson, S. H., Amino Acid Substitution in the Active Site of DNA Polymerase B Explains the Energy Barrier of the Nucleotidyl Transfer Reaction. *J. Am. Chem. Soc.* **2013,** 135 (21), 8078-8088.

22. Cisneros, G. A.; Perera, L.; García-Díaz, M.; Bebenek, K.; Kunkel, T. A.; Pedersen, L. G., Catalytic Mechanism of Human DNA Polymerase Λ with Mg2+ and Mn2+ from Ab Initio Quantum Mechanical/Molecular Mechanical Studies. *DNA Repair* **2008,** 7 (11), 1824-1834.

23. Abashkin, Y. G.; Erickson, J. W.; Burt, S. K., Quantum Chemical Investigation of Enzymatic Activity in DNA Polymerase B. A Mechanistic Study. *The Journal of Physical Chemistry B* **2001,** 105 (1), 287-292.

24. Bojin, M. D.; Schlick, T., A Quantum Mechanical Investigation of Possible Mechanisms for the Nucleotidyl Transfer Reaction Catalyzed by DNA Polymerase B. *The Journal of Physical Chemistry B* **2007,** 111 (38), 11244-11252.

25. Lior-Hoffmann, L.; Wang, L.; Wang, S.; Geacintov, N. E.; Broyde, S.; Zhang, Y., Preferred Wmsa Catalytic Mechanism of the Nucleotidyl Transfer Reaction in Human DNA Polymerase K Elucidates Error-Free Bypass of a Bulky DNA Lesion. *Nucleic Acids Res.* **2012,** 40 (18), 9193-9205.

26. Prasad, B. R.; Plotnikov, N. V.; Warshel, A., Addressing Open Questions About Phosphate Hydrolysis Pathways by Careful Free Energy Mapping. *The Journal of Physical Chemistry B* **2013,** 117 (1), 153-163.

27. Lior-Hoffmann, L.; Ding, S.; Geacintov, N. E.; Zhang, Y.; Broyde, S., Structural and Dynamic Characterization of Polymerase K's Minor Groove Lesion Processing Reveals How Adduct Topology Impacts Fidelity. *Biochemistry* **2014,** 53 (35), 5683-5691.

28. Wang, Y.; Schlick, T., Quantum Mechanics/Molecular Mechanics Investigation of the Chemical Reaction in Dpo4 Reveals Water-Dependent Pathways and Requirements for Active Site Reorganization. *J. Am. Chem. Soc.* **2008,** 130 (40), 13240-13250.

29. Wang, L.; Yu, X.; Hu, P.; Broyde, S.; Zhang, Y., A Water-Mediated and Substrate-Assisted Catalytic Mechanism for Sulfolobus Solfataricus DNA Polymerase IV. *J. Am. Chem. Soc.* **2007,** 129 (15), 4731-4737.

30. Li, Y.; Bao, L.; Zhang, R.; Tang, X.; Zhang, Q.; Wang, W., Insights into the Error Bypass of 1-Nitropyrene DNA Adduct by DNA Polymerase I: A QM/MM Study. *Chem. Phys. Lett.* **2017,** 686, 12-17.

31. Xiang, Y.; Warshel, A., Quantifying Free Energy Profiles of Proton Transfer Reactions in Solution and Proteins by Using a Diabatic Fdft Mapping. *The Journal of Physical Chemistry B* **2008,** 112 (3), 1007-1015.

32. A., M. R.; Hanwool, Y.; Arieh, W., Exploring the Mechanism of DNA Polymerases by Analyzing the Effect of Mutations of Active Site Acidic Groups in Polymerase B. *Proteins* **2016,** 84 (11), 1644-1657.

33. Genna, V.; Vidossich, P.; Ippoliti, E.; Carloni, P.; De Vivo, M., A Self-Activated Mechanism for Nucleic Acid Polymerization Catalyzed by DNA/RNA Polymerases. *J. Am. Chem. Soc.* **2016,** 138 (44), 14592-14598.

34. McCulloch, S. D.; Kokoska, R. J.; Masutani, C.; Iwai, S.; Hanaoka, F.; Kunkel, T. A., Preferential Cis-Syn Thymine Dimer Bypass by DNA Polymerase Eta Occurs with Biased Fidelity. *Nature* **2004,** 428 (6978), 97-100.

35. Yang, W., An Overview of Y-Family DNA Polymerases and a Case Study of Human DNA Polymerase H. *Biochemistry* **2014,** 53 (17), 2793-2803.

36. Masutani, C.; Kusumoto, R.; Yamada, A.; Dohmae, N.; Yokoi, M.; Yuasa, M.; Araki, M.; Iwai, S.; Takio, K.; Hanaoka, F., The XPV (Xeroderma Pigmentosum Variant) Gene Encodes Human DNA Polymerase H. *Nature* **1999,** 399 (6737), 700-704.

37. Zhao, Y.; Biertümpfel, C.; Gregory, M. T.; Hua, Y.-J.; Hanaoka, F.; Yang, W., Structural Basis of Human DNA Polymerase H-Mediated Chemoresistance to Cisplatin. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, 201202681.

38. Srivastava, A. K.; Han, C.; Zhao, R.; Cui, T.; Dai, Y.; Mao, C.; Zhao, W.; Zhang, X.; Yu, J.; Wang, Q.-E., Enhanced Expression of DNA Polymerase Eta Contributes to Cisplatin Resistance of Ovarian Cancer Stem Cells. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, 201421365.

39. Biertuempfel, C.; Zhao, Y.; Kondo, Y.; Ramon-Maiques, S.; Gregory, M.; Lee, J. Y.; Masutani, C.; Lehmann, A. R.; Hanaoka, F.; Yang, W., Structure and Mechanism of Human DNA Polymerase Eta. *Nature* **2010,** 465 (7301), 1044-U102.

40. Yang, W.; Weng, P. J.; Gao, Y., A New Paradigm of DNA Synthesis: Three-Metal-Ion Catalysis. *Cell & Bioscience* **2016,** 6 (1), 51.

41. Gao, Y.; Yang, W., Capture of a Third Mg2+ Is Essential for Catalyzing DNA Synthesis. *Science* **2016,** 352 (6291), 1334-1337.

42. Ummat, A.; Silverstein, T. D.; Jain, R.; Buku, A.; Johnson, R. E.; Prakash, L.; Prakash, S.; Aggarwal, A. K., Human DNA Polymerase H Is Pre-Aligned for dNTP Binding and Catalysis. *J. Mol. Biol.* **2012,** 415 (4), 627-634.

43. Silverstein, T. D.; Johnson, R. E.; Jain, R.; Prakash, L.; Prakash, S.; Aggarwal, A. K., Structural Basis for the Suppression of Skin Cancers by DNA Polymerase H. *Nature* **2010,** 465 (7301), 1039.

44. 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 H. *Science* **2007,** 318 (5852), 967-970.

45. Washington, M. T.; Johnson, R. E.; Prakash, L.; Prakash, S., The Mechanism of Nucleotide Incorporation by Human DNA Polymerase H Differs from That of the Yeast Enzyme. *Mol. Cell. Biol.* **2003,** 23 (22), 8316-8322.

46. Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H., Propka3: Consistent Treatment of Internal and Surface Residues in Empirical Pka Predictions. *J. Chem. Theory Comput.* **2011,** 7 (2), 525-537.

47. Case, D. A.; Darden, T. A.; III, T. E. C.; Simmerling, C. L.; Wang, J.; R.E. Duke, R. L.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 12*, University of California: San Francisco, 2012.

48. Allnér, O.; Nilsson, L.; Villa, A., Magnesium Ion–Water Coordination and Exchange in Biomolecular Simulations. *J. Chem. Theory Comput.* **2012,** 8 (4), 1493-1502.

49. 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.

50. Grossfield, A. *Wham: The Weighted Histogram Analysis Method*, version 2.0.9.

51. 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.; Jr., J. A. M.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavchari, 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, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision D.01; Gaussian, Inc.: Wallingford CT, 2013.

52. Brautigam, C. A.; Steitz, T. A., Structural and Functional Insights Provided by Crystal Structures of DNA Polymerases and Their Substrate Complexes. *Curr. Opin. Struct. Biol.* **1998,** 8 (1), 54-63.

53. Carvalho, A. T. P.; Fernandes, P. A.; Ramos, M. J., The Catalytic Mechanism of RNA Polymerase II. *J. Chem. Theory Comput.* **2011,** 7 (4), 1177-1188.

54. Ribeiro, A. J. M.; Ramos, M. J.; Fernandes, P. A., The Catalytic Mechanism of Hiv-1 Integrase for DNA 3′-End Processing Established by QM/MM Calculations. *J. Am. Chem. Soc.* **2012,** 134 (32), 13436-13447.

55. Whitaker, J. R.; Feeney, R. E.; Sternberg, M. M., Chemical and Physical Modification of Proteins by the Hydroxide Ion. *Critical Reviews in Food Science & Nutrition* **1983,** 19 (3), 173-212.

56. Smith, T. A.; Kirkpatrick, D. R.; Smith, S.; Smith, T. K.; Pearson, T.; Kailasam, A.; Herrmann, K. Z.; Schubert, J.; Agrawal, D. K., Radioprotective Agents to Prevent Cellular Damage Due to Ionizing Radiation. *Journal of Translational Medicine* **2017,** 15 (1), 232.

57. El-Kenawi, A.; Ruffell, B., Inflammation, Ros, and Mutagenesis. *Cancer Cell* **2017,** 32 (6), 727-729.

8. Chapter 8: Conclusions[a]

8.1 Thesis Overview

It has been previously determined that the conformations adopted by DNA lesions dictates their mutagenic consequences. Previous literature[1-5] and the work presented in this thesis show that computational studies are critical for obtaining important structural information. The current thesis first examines the conformational preferences of Bz-G, POB-G, PHB-G, and POB-T. For all systems, a systematic multiscale modeling approach was used that employs models of increasing size (from a nucleobase to a nucleoside to a nucleotide to a base pair dimer to an adducted DNA duplex to a DNA polymerase pre-insertion complex to a DNA polymerase insertion complex) and computational methods of varying accuracy (from QM to MD). Subsequently, the mechanism catalyzed by human DNA pol η was investigated using a combination of MD, umbrella sampling, and ONIOM QM/MM calculations. Overall, each model used in this thesis has provided critical insight into the replication of flexible DNA lesions by TLS polymerases.

In Chapter 2, the substantial conformational flexibility of Bz-G, including rotation of the bulky group with respect to the base and the internal conformation of the bulk moiety, is seen in the nucleoside and nucleotide models. This conformational flexibility suggests that the orientation adopted by Bz-G will be highly dependent on the local environment of the adduct. Upon incorporation of Bz-G into DNA, the conformational flexibility is maintained and leads to many damaged DNA conformations, including those not previously reported in the literature. The range of Bz-G conformations adopted in DNA arises due to a

[a] *Chemical Research in Toxicology* reference style used throughout this chapter.

combination of the flexible (–CH_2–) linker, the small adduct size, the discrete interactions in the helix, and the solvation of the benzyl moiety. These factors in turn lead to a complicated mutagenic profile that may involve pairing with natural C, mispairs with A or T, or deletion mutations, which is supported by previously reported experimental biochemical studies.[6-9] Nevertheless, my calculations suggest that pairing with C leads to the most stable helical structure, which is the first combined structural and energetic explanation for experimental results that report a higher rate of C incorporation than any other nucleobase upon the Bz-G replication.[6-9] Overall, Chapter 2 highlights the key information obtained from nucleoside, nucleotide, and DNA models on flexible DNA lesions that will assist in modeling the conformation and base pairing of DNA lesions in the polymerase active site.

To further understand the mutagenicity of Bz-G, the binding of the lesion in the Dpo4 active site was investigated (Chapter 3). These calculations show that Bz-G binds in the pre-insertion complex in the same manner as natural G due to lack of interactions between Dpo4 and the template base. As a result, the Bz-G bulky moiety is positioned over the 3′–flanking base pair and thereby occupies the dNTP binding pocket. Therefore, Bz-G must undergo a conformational change prior to or during dNTP binding in order for replication to proceed. This prediction is consistent with the experimentally observed decrease in catalytic rate prior to or during dNTP insertion opposite Bz-G compared to G.[7] This is the first atomic level detail about a pre-insertion complex for an O6-G lesion and highlights the importance of obtaining structural information at each stage during the TLS process for DNA lesions.

Subsequently, the Dpo4 insertion complexes for Bz-G replication (Chapter 4) were modeled. When dCTP or dTTP are paired opposite Bz-G in the Dpo4 active site, a catalytically conducive orientation is adopted. In contrast, due to the lack of critical hydrogen-bonding interactions, changes in the active site conformation and/or the lack of reaction-ready parameters, the purines are unlikely to be inserted opposite Bz-G. Additionally, although TLS of DNA adducts has been shown to commonly result in deletion mutations, this is not a predominant outcome for Bz-G replication, and I predict this is due to the small size and high degree of flexibility of the Bz-G bulky moiety in the polymerase active site. Overall, the observed structural parameters in the Dpo4 active site for Bz-G replication indicate that dCTP and dTTP will be inserted opposite the lesion, which correlates with the experimental mutational spectrum.[6-9] Furthermore, the observed differences between the DNA and polymerase models emphasize the effects of the enzyme active site on the predicted mutagenic outcomes. In summary, Chapters 2 – 4 provide significant chemical insights that explain the mutagenicity of the Bz-G adduct and highlight the predictive power of computer modeling for understanding the mutagenic outcomes associated with DNA damage.

Using the same multiscale modeling approach developed in Chapters 2 – 4, the mutagenicities of POB-G and PHB-G, which differ in a single substituent on the bulky moiety chain (i.e., carbonyl versus hydroxy), were investigated in Chapter 5. POB-G and PHB-G exhibit a high degree of conformational flexibility in nucleobase, nucleoside, DNA duplex, and polymerase models. In the DNA duplex and pol η active site, the bulky moiety is preferentially located in the DNA major groove. POB-G and PHB-G form stable pairs

opposite C and T, with the T pairs being the least distorted relative to canonical DNA. This is consistent with the previously observed predominant G → A mutations and the non-mutagenic replication of POB-G.[9-11] Furthermore, my calculations predict PHB-G to be more mutagenic than POB-G since the hydroxy group forms an additional hydrogen bond with the pairing dTTP that further stabilizes the PHB-G:dTTP pair in the pol η active site compare to the POB-G:dTTP pair. The mutagenicity of PHB-G is further increased by the stabilization of an intercalated DNA conformation that has been associated with deletion mutations for other adducts.[12-14] Since POB-G and PHB-G only differ in a single substituent on the bulky moiety linker chain, the differences in mutagenicity afford a broader understanding of how subtle chemical changes to the bulky moiety of a DNA adduct can affect the lesion mutagenicity.

In Chapter 6, the first atomic level details that clarify the replication outcomes for the most persistent tobacco lesion, namely POB-T, are provided. These calculations explain how human pol κ and η process this minor groove lesion. Although the experimental mutagenic profile for POB-T has been previously reported,[15-17] the findings are puzzling in terms of the human polymerase involved. Specifically, although pol κ typically replicates minor groove adducts, *in vivo* studies indicate pol η replicates POB-T despite being known for processing major groove adducts.[15-17] The multiscale modeling approach provides structural information on POB-T and thereby clarifies the unique replication outcomes for these two polymerases. Our calculations reveal that the *anti* glycosidic orientation of POB-T can fit in the pol κ active site, but only *syn*-POB-T is accommodated by pol η. These binding orientations rationalize the reported differential mutagenic spectra[15-17] based on the preferential stabilization of dGTP and dTTP opposite the lesion for pol κ and η,

respectively. By uncovering a unique approach for POB-T replication by pol η, Chapter 6 provides the insight necessary to clarify a discrepancy in the DNA replication literature by expanding the biological role of pol η and providing the first evidence for the replication of a damaged pyrimidine in the *syn* glycosidic orientation.

Finally, in Chapter 7, the pol η catalyzed nucleotidyl transfer reaction is investigated using MD, umbrella sampling, and ONIOM QM/MM calculations. Specifically, although there have been many studies that have investigated the reaction pathway for a variety of DNA polymerases, certain features of the mechanism remain unclear, including the identity of the general base for nucleophile formation.[18-37] Therefore, Chapter 7 investigates all possibilities for the general base based on previous proposals from experimental and computational studies on pol η and other DNA polymerases using a consistent theoretical framework. The characterized reaction pathways indicate that neither the dNTP α-phosphate nor water can play the role of the general base. Conversely, a hydroxide ion or a conserved active site Glu (Glu116) meditated by an active site Ser (Ser113) leads to calculated barriers consistent with experimental kinetic data,[38-45] and therefore may play the role of the general base. However, these pathways may be limited by the availability of hydroxide in the active site or the orientation of Ser113, respectively. Overall, the most thermodynamically favored pathway considered in this work invokes Glu116 as the general base for the nucleotidyl transfer reaction. This pathway is in agreement with the dynamical structural information obtained in Chapter 7 (i.e., MD and umbrella sampling), as well as with the available experimental kinetic and mutational data,[38-45] and aligns with the proposal made for the identity of the general base for other DNA polymerases.[18-24] Taken together, these results help to answer a long-standing question in the DNA replication

literature regarding the identity of the general base for the DNA polymerase reaction mechanism.

8.2. Future Work

8.2.1. DNA Lesions

The computational methodology developed and successfully applied in this thesis could be used to investigate other flexible DNA lesions, such as the other tobacco-induced POB and PHB adducts (i.e., PHB-O2-T,[46] POB-O4-T,[46] PHB-O4-T,[46] POB-N6-A,[47] PHB-N6-A,[47] POB-O2-C,[48] PHB-O2-C,[48] POB-N3-C,[48] PHB-N3-C,[48] POB-N4-C,[48] or PHB-N4-C;[48] Figure 8.1). While the current thesis studied the most abundant POB and PHB lesions (i.e., POB-O2-T and POB-O6-G), the recently discovered POB-N6-A and PHB-N6-A lesions are present at modest concentrations *in vivo,* and therefore the mutagenic consequences of these adducts is important to consider.[47] Additionally, understanding the biological implications of POB-O2-C is of particular interest since this major product of POB/PHB adduction of C is unique compared to other POB lesions due to its cationic nature and its frequent spontaneous deamination to form POB-O2-U.[48] Nevertheless, the experimental mutational spectra of these other POB and PHB lesions have not been reported to date. The ability of the new methodology developed in this thesis to predict and rationalize experimental mutational data makes this multiscale approach particularly suited to provide insight into the biochemical impact of these structurally-related DNA lesions.

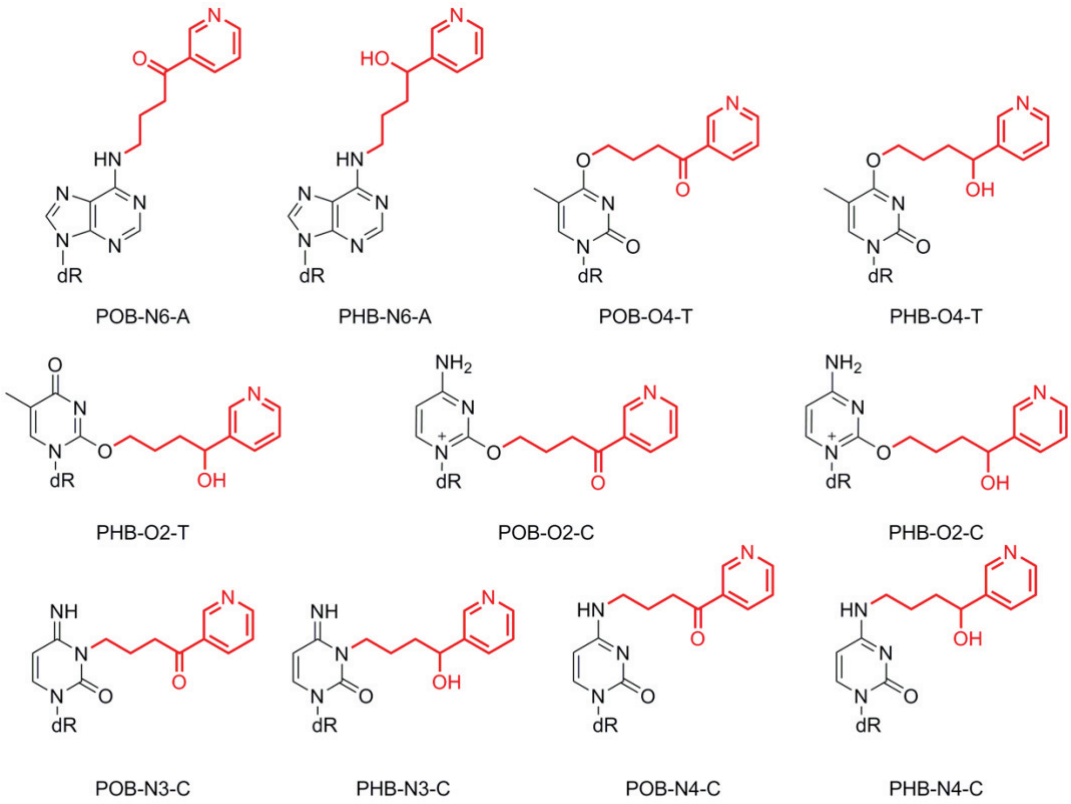

Figure 8.1. POB and PHB lesions formed upon exposure to NNK for which structural and mutational information is not known.

In addition to investigating the biological consequences of other tobacco lesions, the computational methodology developed in this thesis can be used to study the structure and associated mutagenic consequences of other flexible lesions. For example, several flexible DNA lesions (i.e., estragole-N6-A (ES-A),[49, 50] estragole-N2-G (ES-N2-G),[49, 50] estragole-C8-G (ES-C8-G),[49, 50] methyleugenol-N6-A (MIE-N6-A),[51, 52] methyleugenol-N2-G (MIE-N2-G),[51, 52] N2-3′-safrole-G (3′-S-N2-G),[53] N2-1′-safrole-G (1′-S-N2-G), [53] and N6-safrole-A (3′-S-N6-A);[53] Figure 8.2) have recently been identified. These lesions are formed upon exposure to estragole, methyleugenol, and safrole, which are used as additives in many food and cosmetic products, and found naturally in various herbs and spices, including mace, tarragon, basil, fennel, nutmeg, and anise.[49, 51-53] Nevertheless, no structural or mutagenic data is currently available for these estragole, methyleugenol, and

safrole lesions. Obtaining a structural understanding of these lesions is critical since this thesis and previous work on DNA adducts arising from known human carcinogens highlights that lesion mutagenicity can be affected by small changes in the chemical composition of the adduct, such as the damaged nucleobase, linker substituent, ring substitution, and ionization state.[7-9, 12, 54-62] Therefore, by investigating this series of flexible lesions, along with the POB and PHB adducts, a broader understanding of the connection between adduct chemical composition and biological outcomes will be obtained.

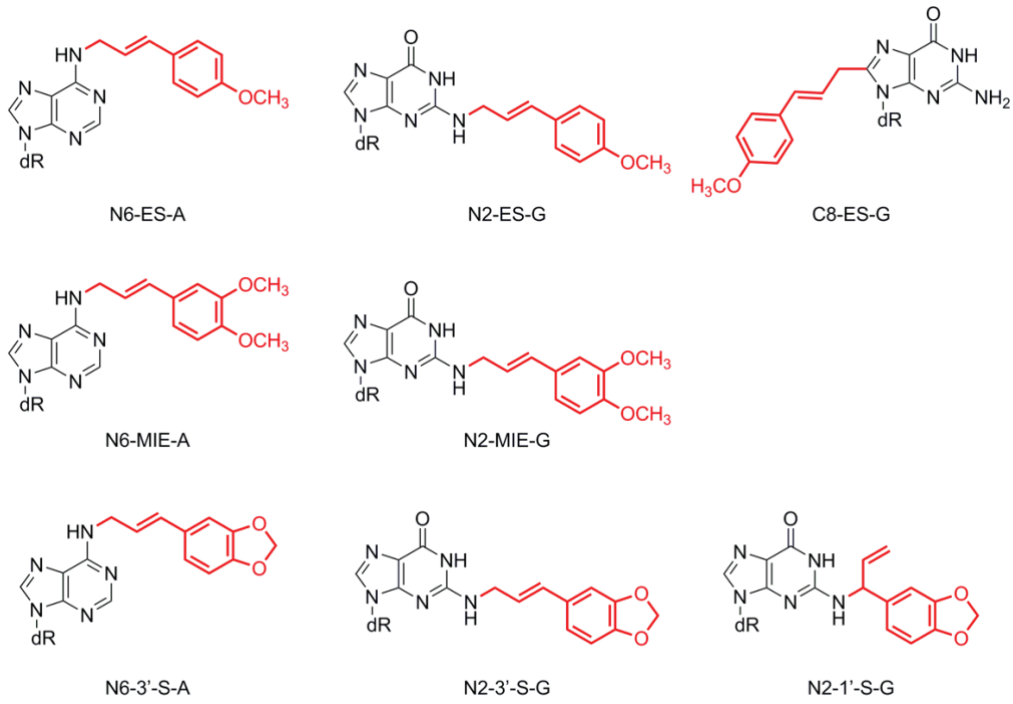

Figure 8.2. Estragole, methyleugenol, and safrole DNA lesions proposed for future structural investigations.

Besides gaining a greater understanding of the structure and mutagenicity of flexible DNA lesions, computational chemistry can be used to verify the expanded role of pol η in replicating O2-T lesions proposed in Chapter 6. Specifically, it would be of interest to investigate the bypass of other O2-alkyl-T lesions. Experimentally, a series of O2-alkyl-T

lesions has been synthesized (i.e., methyl, ethyl, n-propyl, iso-propyl, n-butyl, iso-butyl, and s-butyl) and the replication of these lesions has been investigated in human cells.[63] As is observed for POB-T,[15-17] the replication of straight chained O2-alkyl-T lesions is carried out by pol η.[63] Furthermore, the frequency of T → A transversion mutations for straight chained O2-alkyl-T lesions increases with chain length from 1% (methyl), 5% (ethyl) and ~13% (propyl and butyl).[63] The effect of chain length on the bypass of these lesions indicates that although the methyl and ethyl lesions may be accommodated in the canonical *anti* conformation in the pol η active site, the steric constraints of the active site likely cause the propyl and butyl lesions to adopt a *syn* orientation about the glycosidic bond, as seen for POB-T in Chapter 6. Furthermore, pol ι replicates the branched chain O2-alkyl-T lesions *in vivo* and leads to mostly non-mutagenic replication (< 4% T → A mutations).[63] The structural basis for the branched and straight chain alkyl-T lesions being replicated by different polymerases is not yet known. By using the same methodology applied in Chapter 6 to computationally investigate the replication of this series of O2-alkyl-T lesions, the structural basis for their mutagenicity can be identified and a greater understanding of key human DNA polymerases can be obtained.

8.2.2. DNA Polymerase Reaction Mechanism

Aside from understanding the mutagenic outcomes for the replication of DNA lesions, computer modeling can be used to obtain insight into the reaction mechanism catalyzed by DNA polymerases. Indeed, in Chapter 7, the mechanism of action of DNA pol η was investigated and my calculations provided substantial insight into the identity of the general base for the corresponding nucleotidyl transfer reaction. Due to the conserved active site configuration, future studies should investigate the proposed reaction mechanism for a

broad range of replicative and TLS polymerases to further elucidate whether a unified reaction mechanism for the general base exists for all DNA polymerases. By using a methodology consistent with that employed in Chapter 7, a quantitative comparison between different polymerases could be completed that is not possible based on previous literature (i.e., due to the variation in models and methods used). Additionally, although previous work has provided some insight into the role of metal ions in the reaction mechanism, future work that examines the role of the third metal in the reaction is required. Specifically, although a third metal ion has been proposed to stabilize the transition state for pyrophosphate formation, assist in product release and/or prevent the reverse reaction for pol β and η,[33, 34, 38, 64] it has also been proposed that an Arg residue could fulfill this role and therefore the impact of a third metal ion on the reaction barriers needs to be investigated. Furthermore, since polymerases are believed to share the same general reaction mechanism due to the conserved structure of the active site, the question still remains whether other DNA polymerases utilize a third metal ion during catalysis.

8.3. Final Remarks

In conclusion, this thesis provides the first comprehensive structural information on the replication of flexible O6-G and O2-T alkyl lesions. This work clarifies the mutational spectra for several lesions and the structural basis for their replication by pol η. Furthermore, the reaction mechanism catalyzed by pol η is now better characterized. The work proposed in this Chapter will help to further probe the mechanism of action of TLS polymerases, including the molecular basis for the mutagenic replication of flexible DNA lesions. Overall, this thesis emphasizes the complex interplay between many factors that are required to replicate damaged DNA.

8.4. References

(1) Wilson, K. A., and Wetmore, S. D. (2017) Molecular insights into the translesion synthesis of benzyl-guanine from molecular dynamics simulations: Structural evidence of mutagenic and nonmutagenic replication. *Biochemistry 56*, 1841-1853.

(2) Wilson, K. A., and Wetmore, S. D. (2017) Conformational flexibility of the benzyl-guanine adduct in a bypass polymerase active site permits replication: Insights from molecular dynamics simulations. *Chem. Res. Toxicol. 30*, 2013-2022.

(3) Wang, L., and Broyde, S. (2006) A new anti conformation for N-(deoxyguanosin-8-yl)-2-acetylaminofluorene (AAF-dG) allows Watson-Crick pairing in the sulfolobus solfataricus p2 DNA polymerase IV (Dpo4). *Nucleic Acids Res. 34*, 785-795.

(4) Zhang, L., Shapiro, R., and Broyde, S. (2005) Molecular dynamics of a food carcinogen−DNA adduct in a replicative DNA polymerase suggest hindered nucleotide incorporation and extension. *Chem. Res. Toxicol. 18*, 1347-1363.

(5) Wang, L., Wu, M., Yan, S. F., Patel, D. J., Geacintov, N. E., and Broyde, S. (2005) Accommodation of a 1s-(−)-benzo[c]phenanthrenyl-N6-dA adduct in the y-family Dpo4 DNA polymerase active site: Structural insights through molecular dynamics simulations. *Chem. Res. Toxicol. 18*, 441-456.

(6) Lim, S., Song, I., Guengerich, F. P., and Choi, J.-Y. (2012) Effects of N2-alkylguanine, O6-alkylguanine, and abasic lesions on DNA binding and bypass synthesis by the euryarchaeal b-family DNA polymerase vent (exo–). *Chem. Res. Toxicol. 25*, 1699-1707.

(7) Eoff, R. L., Angel, K. C., Egli, M., and Guengerich, F. P. (2007) Molecular basis of selectivity of nucleoside triphosphate incorporation opposite O6-benzylguanine by sulfolobus solfataricus DNA polymerase Dpo4: Steady-state and pre-steady-state kinetics and x-ray crystallography of correct and incorrect pairing. *J. Biol. Chem. 282*, 13573-13584.

(8) Gahlon, H. L., Boby, M. L., and Sturla, S. J. (2014) O6-alkylguanine postlesion DNA synthesis is correct with the right complement of hydrogen bonding. *ACS Chem. Biol. 9*, 2807-2814.

(9) Choi, J.-Y., Chowdhury, G., Zang, H., Angel, K. C., Vu, C. C., Peterson, L. A., and Guengerich, F. P. (2006) Translesion synthesis across O6-alkylguanine DNA adducts by recombinant human DNA polymerases. *J. Biol. Chem. 281*, 38244-38256.

(10) Pauly, G. T., Peterson, L. A., and Moschel, R. C. (2002) Mutagenesis by O6-[4-oxo-4-(3-pyridyl)butyl]guanine in Escherichia coli and human cells. *Chem. Res. Toxicol. 15*, 165-169.

(11) Gowda, A. S. P., and Spratt, T. E. (2016) DNA polymerase ν rapidly bypasses O6-methyl-dG but not O6-[4-(3-pyridyl)-4-oxobutyl-dG and O2-alkyl-dTs. *Chem. Res. Toxicol. 29*, 1894-1900.

(12) Stover, J. S., Chowdhury, G., Zang, H., Guengerich, F. P., and Rizzo, C. J. (2006) Translesion synthesis past the C8- and N2-deoxyguanosine adducts of the dietary mutagen 2-amino-3-methylimidazo[4,5-f]quinoline in the nari recognition sequence by prokaryotic DNA polymerases. *Chem. Res. Toxicol. 19*, 1506-1517.

(13) Norman, D., Abuaf, P., Hingerty, B. E., Live, D., Grunberger, D., Broyde, S., and Patel, D. J. (1989) 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 28*, 7462-7476.

(14) Gu, Z., Gorin, A., Hingerty, B. E., Broyde, S., and Patel, D. J. (1999) Solution structures of aminofluorene [AF]-stacked conformers of the syn [AF]−C8-dG adduct positioned opposite dC or dA at a template-primer junction. *Biochemistry 38*, 10855-10870.

(15) Gowda, A. S. P., Krishnegowda, G., Suo, Z., Amin, S., and Spratt, T. E. (2012) Low fidelity bypass of O2-(3-pyridyl)-4-oxobutylthymine, the most persistent bulky adduct produced by the tobacco specific nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone by model DNA polymerases. *Chem. Res. Toxicol. 25*, 1195-1202.

(16) Gowda, A. S. P., and Spratt, T. E. (2016) DNA polymerases η and ζ combine to bypass O(2)-[4-(3-pyridyl)-4-oxobutyl]thymine, a DNA adduct formed from tobacco carcinogens. *Chem. Res. Toxicol. 29*, 303-316.

(17) Weerasooriya, S., Jasti, V. P., Bose, A., Spratt, T. E., and Basu, A. K. (2015) Roles of translesion synthesis DNA polymerases in the potent mutagenicity of tobacco-specific nitrosamine-derived O(2)-alkylthymidines in human cells. *DNA repair 35*, 63-70.

(18) Florián, J., Goodman, M. F., and Warshel, A. (2003) Computer simulation of the chemical catalysis of DNA polymerases: Discriminating between alternative nucleotide insertion mechanisms for T7 DNA polymerase. *J. Am. Chem. Soc. 125*, 8163-8177.

(19) Rittenhouse, R. C., Apostoluk, W. K., Miller, J. H., and Straatsma, T. (2003) Characterization of the active site of DNA polymerase β by molecular dynamics and quantum chemical calculation. *Proteins: Struct., Funct., Bioinf. 53*, 667-682.

(20) Lin, P., Pedersen, L. C., Batra, V. K., Beard, W. A., Wilson, S. H., and Pedersen, L. G. (2006) Energy analysis of chemistry for correct insertion by DNA polymerase β. *Proc. Natl. Acad. Sci. U. S. A. 103*, 13294-13299.

(21) Radhakrishnan, R., and Schlick, T. (2006) Correct and incorrect nucleotide incorporation pathways in DNA polymerase β. *Biochem. Biophys. Res. Commun. 350*, 521-529.

(22) Alberts, I. L., Wang, Y., and Schlick, T. (2007) DNA polymerase β catalysis: Are different mechanisms possible? *J. Am. Chem. Soc. 129*, 11100-11110.

(23) Batra, V. K., Perera, L., Lin, P., Shock, D. D., Beard, W. A., Pedersen, L. C., Pedersen, L. G., and Wilson, S. H. (2013) Amino acid substitution in the active site of DNA polymerase β explains the energy barrier of the nucleotidyl transfer reaction. *J. Am. Chem. Soc. 135*, 8078-8088.

(24) Cisneros, G. A., Perera, L., García-Díaz, M., Bebenek, K., Kunkel, T. A., and Pedersen, L. G. (2008) Catalytic mechanism of human DNA polymerase λ with Mg2+ and Mn2+ from ab initio quantum mechanical/molecular mechanical studies. *DNA Repair 7*, 1824-1834.

(25) Abashkin, Y. G., Erickson, J. W., and Burt, S. K. (2001) Quantum chemical investigation of enzymatic activity in DNA polymerase β. A mechanistic study. *J. Phys. Chem. B 105*, 287-292.

(26) Bojin, M. D., and Schlick, T. (2007) A quantum mechanical investigation of possible mechanisms for the nucleotidyl transfer reaction catalyzed by DNA polymerase β. *J. Phys. Chem. B 111*, 11244-11252.

(27) Lior-Hoffmann, L., Wang, L., Wang, S., Geacintov, N. E., Broyde, S., and Zhang, Y. (2012) Preferred wmsa catalytic mechanism of the nucleotidyl transfer reaction in human DNA polymerase κ elucidates error-free bypass of a bulky DNA lesion. *Nucleic Acids Res. 40*, 9193-9205.

(28) Prasad, B. R., Plotnikov, N. V., and Warshel, A. (2013) Addressing open questions about phosphate hydrolysis pathways by careful free energy mapping. *J. Phys. Chem. B 117*, 153-163.

(29) Lior-Hoffmann, L., Ding, S., Geacintov, N. E., Zhang, Y., and Broyde, S. (2014) Structural and dynamic characterization of polymerase κ's minor groove lesion processing reveals how adduct topology impacts fidelity. *Biochemistry 53*, 5683-5691.

(30) Wang, Y., and Schlick, T. (2008) Quantum mechanics/molecular mechanics investigation of the chemical reaction in Dpo4 reveals water-dependent pathways and requirements for active site reorganization. *J. Am. Chem. Soc. 130*, 13240-13250.

(31) Wang, L., Yu, X., Hu, P., Broyde, S., and Zhang, Y. (2007) A water-mediated and substrate-assisted catalytic mechanism for sulfolobus solfataricus DNA polymerase IV. *J. Am. Chem. Soc. 129*, 4731-4737.

(32) Li, Y., Bao, L., Zhang, R., Tang, X., Zhang, Q., and Wang, W. (2017) Insights into the error bypass of 1-nitropyrene DNA adduct by DNA polymerase ι: A QM/MM study. *Chem. Phys. Lett. 686*, 12-17.

(33) Hanwool, Y., and Arieh, W. (2017) Simulating the fidelity and the three mg mechanism of pol η and clarifying the validity of transition state theory in enzyme catalysis. *Proteins: Struct., Funct., Bioinf. 85*, 1446-1453.

(34) Stevens, D. R., and Hammes-Schiffer, S. (2018) Exploring the role of the third active site metal ion in DNA polymerase η with QM/MM free energy simulations. *J. Am. Chem. Soc. 140*, 8965-8969.

(35) Xiang, Y., and Warshel, A. (2008) Quantifying free energy profiles of proton transfer reactions in solution and proteins by using a diabatic fdft mapping. *J. Phys. Chem. B 112*, 1007-1015.

(36) A., M. R., Hanwool, Y., and Arieh, W. (2016) Exploring the mechanism of DNA polymerases by analyzing the effect of mutations of active site acidic groups in polymerase β. *Proteins: Struct., Funct., Bioinf. 84*, 1644-1657.

(37) Genna, V., Vidossich, P., Ippoliti, E., Carloni, P., and De Vivo, M. (2016) A self-activated mechanism for nucleic acid polymerization catalyzed by DNA/RNA polymerases. *J. Am. Chem. Soc. 138*, 14592-14598.

(38) Nakamura, T., Zhao, Y., Yamagata, Y., Hua, Y.-J., and Yang, W. (2012) Watching DNA polymerase eta make a phosphodiester bond. *Nature 487*, 196-U177.

(39) Biertuempfel, C., Zhao, Y., Kondo, Y., Ramon-Maiques, S., Gregory, M., Lee, J. Y., Masutani, C., Lehmann, A. R., Hanaoka, F., and Yang, W. (2010) Structure and mechanism of human DNA polymerase eta. *Nature 465*, 1044-U1102.

(40) Yang, W., Weng, P. J., and Gao, Y. (2016) A new paradigm of DNA synthesis: Three-metal-ion catalysis. *Cell Biosci. 6*, 51.

(41) Gao, Y., and Yang, W. (2016) Capture of a third Mg2+ is essential for catalyzing DNA synthesis. *Science 352*, 1334-1337.

(42) Ummat, A., Silverstein, T. D., Jain, R., Buku, A., Johnson, R. E., Prakash, L., Prakash, S., and Aggarwal, A. K. (2012) Human DNA polymerase η is pre-aligned for dNTP binding and catalysis. *J. Mol. Biol. 415*, 627-634.

(43) Silverstein, T. D., Johnson, R. E., Jain, R., Prakash, L., Prakash, S., and Aggarwal, A. K. (2010) Structural basis for the suppression of skin cancers by DNA polymerase η. *Nature 465*, 1039.

(44) Alt, A., Lammens, K., Chiocchini, C., Lammens, A., Pieck, J. C., Kuch, D., Hopfner, K.-P., and Carell, T. (2007) Bypass of DNA lesions generated during anticancer treatment with cisplatin by DNA polymerase η. *Science 318*, 967-970.

(45) Washington, M. T., Johnson, R. E., Prakash, L., and Prakash, S. (2003) The mechanism of nucleotide incorporation by human DNA polymerase η differs from that of the yeast enzyme. *Mol. Cell. Biol. 23*, 8316-8322.

(46) Leng, J., and Wang, Y. (2017) Liquid chromatography-tandem mass spectrometry for the quantification of tobacco-specific nitrosamine-induced DNA adducts in mammalian cells. *Anal. Chem. 89*, 9124-9130.

(47) Carlson, E. S., Upadhyaya, P., Villalta, P. W., Ma, B., and Hecht, S. S. (2018) Analysis and identification of 2′-deoxyadenosine-derived adducts in lung and liver DNA of f-344 rats treated with the tobacco-specific carcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone and enantiomers of its metabolite 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol. *Chem. Res. Toxicol. 31*, 358-370.

(48) Michel, A. K., Zarth, A. T., Upadhyaya, P., and Hecht, S. S. (2017) Identification of 4-(3-pyridyl)-4-oxobutyl-2′-deoxycytidine adducts formed in the reaction of DNA with 4-(acetoxymethylnitrosamino)-1-(3-pyridyl)-1-butanone: A chemically activated form of tobacco-specific carcinogens. *ACS Omega 2*, 1180-1190.

(49) Ishii, Y., Suzuki, Y., Hibi, D., Jin, M., Fukuhara, K., Umemura, T., and Nishikawa, A. (2011) Detection and quantification of specific DNA adducts by liquid chromatography−tandem mass spectrometry in the livers of rats given estragole at the carcinogenic dose. *Chem. Res. Toxicol. 24*, 532-541.

(50) Paini, A., Punt, A., Scholz, G., Gremaud, E., Spenkelink, B., Alink, G., Schilter, B., van Bladeren, P. J., and Rietjens, I. M. C. M. (2012) In vivo validation of DNA adduct formation by estragole in rats predicted by physiologically based biodynamic modelling. *Mutagenesis 27*, 653-663.

(51) Tremmel, R., Herrmann, K., Engst, W., Meinl, W., Klein, K., Glatt, H., and Zanger, U. M. (2017) Methyleugenol DNA adducts in human liver are associated with sult1a1 copy number variations and expression levels. *Arch. Toxicol. 91*, 3329-3339.

(52) Herrmann, K., Engst, W., Appel, K. E., Monien, B. H., and Glatt, H. (2012) Identification of human and murine sulfotransferases able to activate hydroxylated metabolites of methyleugenol to mutagens in salmonella typhimurium and detection of associated DNA adducts using uplc–MS/MS methods. *Mutagenesis 27*, 453-462.

(53) Martati, E., Boonpawa, R., van den Berg, J. H. J., Paini, A., Spenkelink, A., Punt, A., Vervoort, J., van Bladeren, P. J., and Rietjens, I. M. C. M. (2014) Malabaricone C-containing mace extract inhibits safrole bioactivation and DNA adduct formation both in vitro and in vivo. *Food Chem. Toxicol. 66*, 373-384.

(54) Wilson, K., Kung, R., and Wetmore, S. (2016) Toxicology of DNA adducts formed upon human exposure to carcinogens: Insights gained from molecular modeling. *Advances in Molecular Toxicology 10*, 293.

(55) Patra, A., Nagy, L. D., Zhang, Q. Q., Su, Y., Muller, L., Guengerich, F. P., and Egli, M. (2014) Kinetics, structure, and mechanism of 8-oxo-7,8-dihydro-2 '-deoxyguanosine bypass by human DNA polymerase eta. *J. Biol. Chem. 289*, 16867-16882.

(56) Gahlon, H. L., Schweizer, W. B., and Sturla, S. J. (2013) Tolerance of base pair size and shape in postlesion DNA synthesis. *J. Am. Chem. Soc. 135*, 6384-6387.

(57) Kirouac, K. N., Basu, A. K., and Ling, H. (2013) Structural mechanism of replication stalling on a bulky amino-polycyclic aromatic hydrocarbon DNA adduct by a y family DNA polymerase. *J. Mol. Biol. 425*, 4167-4176.

(58) Zhang, H., Eoff, R. L., Kozekov, I. D., Rizzo, C. J., Egli, M., and Guengerich, F. P. (2009) Versatility of Y-family sulfolobus solfataricus DNA polymerase Dpo4 in translesion synthesis past bulky N2-alkylguanine adducts. *J. Biol. Chem. 284*, 3563-3576.

(59) Sherrer, S. M., Brown, J. A., Pack, L. R., Jasti, V. P., Fowler, J. D., Basu, A. K., and Suo, Z. (2009) Mechanistic studies of the bypass of a bulky single-base lesion catalyzed by a Y-family DNA polymerase. *J. Biol. Chem. 284*, 6379-6388.

(60) Gadkari, V. V., Tokarsky, E. J., Malik, C. K., Basu, A. K., and Suo, Z. (2014) Mechanistic investigation of the bypass of a bulky aromatic DNA adduct catalyzed by a Y-family DNA polymerase. *DNA Repair 21*, 65-77.

(61) Xu, P., Oum, L., Lee, Y.-C., Geacintov, N. E., and Broyde, S. (2009) Visualizing sequence-governed nucleotide selectivities and mutagenic consequences through a replicative cycle: Processing of a bulky carcinogen N2-dG lesion in a y-family DNA polymerase. *Biochemistry 48*, 4677-4690.

(62) Wyss, L. A., Nilforoushan, A., Eichenseher, F., Suter, U., Blatter, N., Marx, A., and Sturla, S. J. (2015) Specific incorporation of an artificial nucleotide opposite a mutagenic DNA adduct by a DNA polymerase. *J. Am. Chem. Soc. 137*, 30-33.

(63) Wu, J., Wang, P., Li, L., You, C., and Wang, Y. (2018) Cytotoxic and mutagenic properties of minor-groove O2-alkylthymidine lesions in human cells. *J. Biol. Chem. 293*, 8638-8644.

(64) Raper, A. T., Reed, A. J., and Suo, Z. (2018) Kinetic mechanism of DNA polymerases: Contributions of conformational dynamics and a third divalent metal ion. *Chem. Rev. 118*, 6000-6025.

Appendix A

Supplementary Information for Chapter 2: The Complex Conformational Heterogeneity of the Highly Flexibility O6-Benzyl-Guanine DNA Adduct

Contains Tables A.1–A.6 and Figures A.1–A.6

Table A.1. Partial charges for BzG obtained from HF/6 –31G(d) and RESP charge fitting.

Atom	Atom Type	Charge
P	P	1.1617
O	OS	–0.4974
O1	O2	–0.7662
O2	O2	–0.7662
O3	OS	–0.5408
C5'	CT	–0.059
H5'1	H1	0.0909
H5'2	H1	0.0909
C4'	CT	0.1406
H4'	H1	0.0743
O4'	OS	–0.4085
C1'	CT	0.1123
H1'	H2	0.151
C3'	CT	0.181
H3'	H1	0.071
C2'	CT	–0.0013
H2'1	HC	0.0386
H2'2	HC	0.0386
N9	N*	–0.0145
C8	CK	0.2263
H8	H5	0.1528
C4	CB	0.3054
N7	NB	–0.6079
C5	CB	0.0017
N3	NC	–0.6978
C6	CA	0.6199
C2	CQ	0.873
O6	OS	–0.2792
N1	NC	–0.7099
N2	N2	–0.9381
H21	H	0.3935
H22	H	0.3935
C10	CT	0.0224
H101	H1	0.0759
H102	H1	0.0759
C11	CA	0.0637
C12	CA	–0.1295
H12	HA	0.1307
C13	CA	–0.1226
H13	HA	0.1188
C14	CA	–0.1776
H14	HA	0.146
C15	CA	–0.1773
H15	HA	0.1465
C16	CA	–0.127
H16	HA	0.135

Table A.2. Helical parameters from 20 ns MD simulations on BzG opposite a variety of paring nucleotides.

Base Pair	Shear	Stretch	Stagger	Buckle	Propeller	Opening	HB	Shift	Slide	Rise	Tilt	Roll	Twist
G:C 1 –24	0.176	–0.089	0.218	–3.2	–15.2	–1.0	3	–0.067	–0.837	3.432	2.2	3.6	31.3
G:C 2 –23	–0.137	0.013	0.177	–7.7	–15.6	1.3	2	0.361	–0.076	3.362	1.3	1.4	37.1
G:C 3 –22	0.136	–0.047	0.052	–8.1	–5.7	0.1	3	0.071	–0.124	3.049	–0.4	8.1	22.9
G:C 4 –21	–0.094	–0.052	0.110	5.5	–3.8	–0.3	3	–0.380	–0.434	3.427	–1.4	4.1	34.7
G:C 5 –20	–0.099	–0.053	0.094	2.2	–10.2	–0.5	3	–0.046	–0.682	3.383	1.1	0.7	35.6
G:C 6 –19	0.113	–0.064	0.009	0.8	–9.6	–0.5	3	0.151	–0.260	3.323	0.0	8.7	29.4
G:C 8 –17	0.094	–0.051	0.130	1.2	–11.6	–0.8	3	–0.074	–0.664	3.344	–1.6	0.5	35.6
G:C 9 –16	0.116	–0.061	0.134	–4.1	–9.3	–0.1	3	0.530	–0.518	3.418	1.8	3.5	35.5
G:C 10 –15	0.081	0.018	–0.084	–2.9	–11.6	0.7	2	–0.608	–0.328	3.280	0.6	10.2	28.3
G:C 11 –14	–0.101	0.018	0.015	0.8	–12.5	1.5	2	0.120	–0.825	3.260	–0.4	0.3	30.2
G:C 12 –13	0.345	–0.041	–0.104	1.5	–4.6	0.8	3	0.431	–0.190	3.397	1.7	2.7	36.1
anti –BzG:*syn* –A 1 –24	0.569	–0.053	–0.251	4.7	–16.3	5.9	3	–1.021	0.592	3.505	–3.2	7.3	26.7
anti –BzG:*syn* –A 2 –23	–0.260	0.044	0.011	0.0	–15.8	–1.9	2	0.822	0.294	3.406	2.1	0.6	39.4
anti –BzG:*syn* –A 3 –22	0.109	–0.060	–0.042	–2.3	–8.5	–0.1	3	–0.153	–0.094	3.158	–2.7	9.0	25.1
anti –BzG:*syn* –A 4 –21	–0.117	–0.063	0.137	5.7	–6.4	–0.4	3	–0.230	–0.692	3.491	–0.6	4.8	34.4
anti –BzG:*syn* –A 5 –20	–0.124	–0.064	0.039	–1.5	–10.9	–0.4	3	0.294	–0.686	3.386	–0.1	1.6	34.8
anti –BzG:*syn* –A 6 –19	0.068	–0.042	0.100	–2.3	–4.7	–0.8	3	0.305	–1.121	6.609	2.2	1.0	61.0
anti –BzG:*syn* –A 8 –17	0.196	–0.050	–0.072	7.7	–13.4	0.4	3						
anti –BzG:*syn* –A 9 –16	0.173	–0.079	0.073	–2.9	–9.1	–0.3	3	–0.228	–1.295	3.575	–1.4	5.5	31.6
anti –BzG:*syn* –A 10 –15	0.067	0.025	–0.095	–6.4	–12.8	0.3	2	–0.196	–0.595	3.368	1.7	10.7	30.5
anti –BzG:*syn* –A 11 –14	–0.150	0.026	0.038	–2.6	–12.5	0.2	2	–0.172	–0.904	3.268	–0.3	–0.7	31.2
anti –BzG:*syn* –A 12 –13	0.310	–0.076	–0.116	1.1	–2.4	0.5	3	0.574	0.200	3.355	2.3	2.8	34.6
anti –BzG:*anti* –C 1 –24	0.175	–0.090	0.235	–4.5	–15.1	–1.1	3	–0.306	–0.964	3.427	2.1	2.2	28.2
anti –BzG:*anti* –C 2 –23	–0.157	0.024	0.151	–6.7	–12.0	0.5	2	0.277	0.020	3.282	1.5	1.9	34.2

anti –BzG:*anti* –C 3 –22	0.118	–0.054	–0.011	–3.2	–7.0	0.0	3	–0.065	–0.099	3.195	–0.9	8.1	28.7
anti –BzG:*anti* –C 4 –21	–0.128	–0.064	0.046	3.8	–5.8	–0.3	3	0.006	–0.991	3.525	0.2	4.8	31.5
anti –BzG:*anti* –C 5 –20	–0.124	–0.058	–0.046	–3.9	–10.6	–0.2	3	0.116	–0.723	3.411	–0.6	0.9	35.7
anti –BzG:*anti* –C 6 –19	0.083	–0.044	0.095	–5.1	–4.5	–0.6	3	–0.009	–0.222	6.631	3.5	–0.6	63.5
anti –BzG:*anti* –C 8 –17	0.160	–0.062	–0.052	4.4	–9.7	–0.6	3						
anti –BzG:*anti* –C 9 –16	0.148	–0.072	0.014	0.8	–9.5	–0.3	3	–0.142	–0.810	3.454	–0.8	4.0	30.1
anti –BzG:*anti* –C 10 –15	0.067	0.018	–0.092	–7.2	–13.1	0.8	2	–0.285	–0.445	3.471	0.7	9.7	33.3
anti –BzG:*anti* –C 11 –14	–0.158	0.037	0.136	–4.0	–12.0	–0.8	2	–0.214	–0.845	3.282	–1.3	0.0	30.0
anti –BzG:*anti* –C 12 –13	0.235	–0.097	–0.121	2.2	–1.5	1.5	3	0.799	0.675	3.297	4.3	3.1	33.7
anti –BzG:*anti* –T 1 –24	0.166	–0.087	0.240	–3.5	–15.4	–1.1	3	–0.126	–0.788	3.400	2.1	3.6	30.7
anti –BzG:*anti* –T 2 –23	–0.149	0.010	0.164	–6.4	–15.4	0.9	2	0.422	–0.042	3.370	1.4	0.8	37.6
anti –BzG:*anti* –T 3 –22	0.107	–0.044	0.044	–7.3	–6.0	0.0	3	0.066	–0.170	3.065	–0.6	7.9	23.8
anti –BzG:*anti* –T 4 –21	–0.109	–0.055	0.108	5.1	–4.2	–0.4	3	–0.385	–0.523	3.409	–1.3	3.6	34.3
anti –BzG:*anti* –T 5 –20	–0.119	–0.054	0.096	2.3	–8.6	–0.5	3	0.329	–0.604	3.355	0.7	0.2	34.3
anti –BzG:*anti* –T 6 –19	0.163	–0.061	–0.040	4.1	–3.8	–0.6	3	0.102	–2.204	6.920	1.0	5.9	60.5
anti –BzG:*anti* –T 8 –17	0.205	–0.060	0.034	–4.3	–13.3	0.8	3						
anti –BzG:*anti* –T 9 –16	0.166	–0.049	0.139	–10.3	–3.4	0.5	3	0.131	–0.666	3.492	0.4	4.7	35.0
anti –BzG:*anti* –T 10 –15	0.054	0.024	–0.032	0.7	–7.0	0.2	2	–0.120	–0.191	3.094	1.5	10.6	21.5
anti –BzG:*anti* –T 11 –14	–0.120	0.013	–0.007	4.5	–14.3	–0.2	2	–0.130	–0.633	3.284	–0.2	–3.0	30.0
anti –BzG:*anti* –T 12 –13	0.297	–0.083	–0.106	4.2	–3.4	0.3	3	0.609	0.997	3.490	3.9	4.2	37.3
anti –BzG:*syn* –G –1 1 –24	0.187	–0.098	0.282	–4.2	–17.8	–1.1	3	–0.007	–0.908	3.434	0.9	2.7	33.6
anti –BzG:*syn* –G –1 2 –23	–0.153	0.011	0.352	–8.2	–15.7	–1.8	2	0.016	0.066	3.292	2.2	0.5	32.2
anti –BzG:*syn* –G –1 3 –22	0.106	–0.046	–0.023	–3.8	–5.1	0.3	3	0.209	1.039	3.044	–1.5	9.1	13.8
anti –BzG:*syn* –G –1 4 –21	–0.130	–0.036	0.262	13.3	–5.6	0.4	3	–0.197	–0.695	3.594	1.0	5.4	33.9
anti –BzG:*syn* –G –1 5 –20	–0.143	–0.062	–0.023	2.6	–15.4	0.6	3	0.225	–0.350	3.367	2.3	0.2	38.7
anti –BzG:*syn* –G –1 6 –19	0.201	–0.083	–0.247	–0.8	–13.6	1.1	3	–0.238	–1.793	6.817	2.2	16.8	57.0

anti –BzG:*syn* –G –1 8 –17	0.116	–0.063	0.017	–8.9	–12.4	0.5	3						
anti –BzG:*syn* –G –1 9 –16	0.141	–0.034	0.185	–13.3	–3.4	0.2	3	0.501	–0.331	3.491	0.5	2.0	35.3
anti –BzG:*syn* –G –1 10 –15	0.067	0.033	0.085	0.9	–0.4	0.4	2	–0.294	–0.017	3.040	0.9	5.4	29.2
anti –BzG:*syn* –G –1 11 –14	–0.004	0.003	–0.146	17.3	4.3	2.9	2	0.281	–0.421	3.090	1.6	0.9	22.6
anti –BzG:*syn* –G –2 1 –24	0.197	–0.100	0.246	–2.9	–17.0	–1.0	3	0.043	–0.886	3.448	1.4	3.2	32.9
anti –BzG:*syn* –G –2 2 –23	–0.117	0.000	0.307	–8.1	–16.0	0.6	2	0.120	–0.066	3.309	1.8	1.0	33.4
anti –BzG:*syn* –G –2 3 –22	0.106	–0.046	0.014	–5.3	–5.3	–0.2	3	0.251	0.106	3.095	–1.0	8.0	21.3
anti –BzG:*syn* –G –2 4 –21	–0.134	–0.061	0.179	6.7	–6.7	–0.3	3	–0.142	–0.904	3.525	1.0	4.9	33.3
anti –BzG:*syn* –G –2 5 –20	–0.143	–0.066	–0.045	–1.4	–12.1	0.1	3	0.330	–0.591	3.328	0.8	1.7	35.5
anti –BzG:*syn* –G –2 6 –19	0.229	–0.065	–0.073	–1.0	–8.7	0.5	3	–0.476	–2.173	6.633	2.1	13.1	57.7
anti –BzG:*syn* –G –2 8 –17	0.132	–0.057	–0.077	–3.2	–13.7	1.0	3						
anti –BzG:*syn* –G –2 9 –16	0.146	–0.058	0.145	–8.3	–6.9	–0.1	3	0.277	–0.576	3.469	–0.2	5.5	33.8
anti –BzG:*syn* –G –2 10 –15	0.055	0.028	0.048	–2.0	–5.3	0.9	2	–0.196	–0.299	3.159	0.6	8.1	28.7
anti –BzG:*syn* –G –2 11 –14	–0.076	0.014	–0.027	8.7	–3.3	1.5	2	0.007	–0.582	3.169	0.6	0.4	26.5
anti –BzG:*syn* –G –2 12 –13	0.779	–0.528	0.642	14.6	–14.4	–0.6	1	0.994	1.976	3.573	10.3	1.5	27.6
anti –BzG:*syn* –G –3 1 –24	0.209	–0.104	0.293	–3.8	–18.0	–1.1	3	–0.054	–0.806	3.482	1.4	2.9	32.2
anti –BzG:*syn* –G –3 2 –23	–0.133	–0.007	0.360	–9.6	–15.2	1.1	2	0.582	0.338	3.342	3.0	1.4	35.9
anti –BzG:*syn* –G –3 3 –22	0.128	–0.047	0.041	–6.8	–0.9	–0.8	3	0.405	–0.361	3.061	–1.0	5.5	19.7
anti –BzG:*syn* –G –3 4 –21	–0.128	–0.063	0.227	6.3	–7.3	–0.4	3	–0.089	–1.089	3.630	2.4	2.9	36.4
anti –BzG:*syn* –G –3 5 –20	–0.198	–0.071	–0.094	–5.0	–13.1	0.3	3	0.156	–0.614	3.362	–0.8	–0.4	36.1
anti –BzG:*syn* –G –3 6 –19	0.047	–0.031	0.002	–4.7	–7.8	0.1	3	–0.651	–1.970	6.638	2.7	8.3	55.5
anti –BzG:*syn* –G –3 8 –17	0.075	–0.053	–0.127	–0.6	–11.0	0.1	3						
anti –BzG:*syn* –G –3 9 –16	0.114	–0.057	0.064	–7.3	–5.4	–0.1	3	0.121	–0.582	3.588	–0.9	2.2	34.6
anti –BzG:*syn* –G –3 10 –15	0.053	0.021	0.011	–0.4	–4.4	0.7	2	–0.144	0.017	3.209	0.8	6.0	32.6
anti –BzG:*syn* –G –3 11 –14	–0.027	0.012	–0.125	10.6	–2.7	1.6	2	0.090	–0.632	3.163	0.6	1.1	23.5
anti –BzG:*syn* –G –3 12 –13	3.416	–0.256	0.875	–8.2	34.1	24.4	0	1.978	2.536	3.982	–15.5	12.8	43.8

syn –BzG:*anti* –C 1 1 –24	0.168	–0.091	0.262	–4.4	–16.1	–1.0	3	–0.154	–0.695	3.428	1.9	3.4	31.4
syn –BzG:*anti* –C 1 2 –23	–0.150	0.007	0.251	–8.4	–14.7	1.9	2	0.799	0.268	3.294	3.1	1.9	35.2
syn –BzG:*anti* –C 1 3 –22	0.108	–0.045	–0.011	–4.7	–3.1	–0.8	3	–0.118	–0.290	3.165	–2.6	6.4	26.3
syn –BzG:*anti* –C 1 4 –21	–0.120	–0.068	0.168	2.6	–8.3	–0.6	3	0.024	–1.371	3.575	2.8	5.0	32.7
syn –BzG:*anti* –C 1 5 –20	–0.188	–0.070	–0.129	–7.6	–13.3	0.2	3	–0.244	–0.864	3.396	–0.9	1.4	34.9
syn –BzG:*anti* –C 1 6 –19	0.190	–0.064	0.096	–9.5	–7.6	–0.9	3	0.371	–0.671	7.041	–3.3	15.0	61.1
syn –BzG:*anti* –C 1 8 –17	0.094	–0.073	–0.346	1.3	–5.9	–0.1	3						
syn –BzG:*anti* –C 1 9 –16	0.190	–0.072	–0.108	0.2	–5.8	–0.1	3	–0.297	–1.100	3.436	–1.7	5.8	27.0
syn –BzG:*anti* –C 1 10 –15	0.069	0.021	–0.135	–8.1	–12.7	1.7	2	–0.356	–0.635	3.507	0.6	9.7	32.1
syn –BzG:*anti* –C 1 11 –14	–0.171	0.038	0.125	–5.0	–11.8	–1.2	2	–0.304	–0.882	3.284	–1.3	–0.3	30.5
syn –BzG:*anti* –C 1 12 –13	0.363	–0.053	–0.157	1.4	0.6	1.7	3	0.813	0.714	3.308	4.1	3.1	34.7
syn –BzG:*anti* –C 2 1 –24	0.207	–0.089	0.161	–5.2	–13.0	–0.7	3	–0.199	–0.543	3.435	1.7	2.8	31.7
syn –BzG:*anti* –C 2 2 –23	–0.117	0.012	0.156	–9.0	–14.3	2.6	2	0.649	0.101	3.254	1.3	3.4	33.0
syn –BzG:*anti* –C 2 3 –22	0.127	–0.050	–0.007	–2.5	–3.1	–1.0	3	0.965	–1.013	3.389	0.4	4.8	23.4
syn –BzG:*anti* –C 2 4 –21	–0.097	–0.074	0.036	–2.9	–12.2	–0.2	3	–0.228	–1.576	3.546	2.4	6.0	32.2
syn –BzG:*anti* –C 2 5 –20	–0.153	–0.052	–0.218	–12.9	–8.6	–0.4	3	–0.144	–0.635	3.145	–1.5	5.2	30.9
syn –BzG:*anti* –C 2 6 –19	0.193	–0.070	0.021	–2.1	–3.0	–2.2	3	–0.506	0.364	6.809	–5.8	2.6	62.7
syn –BzG:*anti* –C 2 8 –17	0.059	–0.026	–0.329	14.1	–8.1	–0.9	3						
syn –BzG:*anti* –C 2 9 –16	0.196	–0.074	–0.111	5.1	–11.9	0.6	3	–0.028	–1.284	3.592	–2.9	6.2	27.5
syn –BzG:*anti* –C 2 10 –15	0.070	0.021	–0.133	–9.5	–13.2	1.9	2	–0.217	–0.856	3.619	0.8	11.5	31.8
syn –BzG:*anti* –C 2 11 –14	–0.154	0.043	0.145	–6.5	–11.8	–0.8	2	–0.331	–0.931	3.290	–1.4	–1.0	31.2
syn –BzG:*anti* –C 2 12 –13	0.367	–0.068	–0.090	–0.3	0.2	1.2	3	0.662	0.457	3.313	2.9	3.1	33.0
syn –BzG:*anti* –G 1 1 –24	0.220	–0.109	0.199	0.9	–19.6	–0.9	3	–0.025	–0.929	3.498	1.3	3.2	33.4
syn –BzG:*anti* –G 1 2 –23	–0.186	0.016	0.271	–6.4	–16.4	–0.6	2	0.545	0.256	3.399	2.8	–1.5	39.9
syn –BzG:*anti* –G 1 3 –22	0.098	–0.045	0.045	–8.6	–6.5	0.0	3	–0.310	0.289	2.992	–2.1	8.4	21.0

syn –BzG:*anti* –G 1 4 –21	–0.170	–0.058	0.192	9.2	–4.0	–0.1	3	0.139	–0.785	3.538	1.4	4.1	31.9
syn –BzG:*anti* –G 1 5 –20	–0.128	–0.067	–0.013	1.8	–12.8	0.2	3	–0.077	–0.607	3.378	0.6	–0.3	38.0
syn –BzG:*anti* –G 1 6 –19	0.094	–0.055	–0.049	–0.9	–12.6	0.5	3	0.113	–1.836	6.681	2.3	8.7	59.7
syn –BzG:*anti* –G 1 8 –17	0.225	–0.070	–0.108	–4.3	–10.4	0.4	3						
syn –BzG:*anti* –G 1 9 –16	0.100	–0.050	0.096	–4.4	–6.1	–0.1	3	0.272	–0.641	3.384	–1.8	4.2	33.9
syn –BzG:*anti* –G 1 10 –15	0.051	0.014	–0.139	–4.6	–9.7	1.1	2	–0.091	–0.361	3.320	2.0	10.3	30.2
syn –BzG:*anti* –G 1 11 –14	–0.089	0.013	0.095	1.2	–8.2	1.3	2	0.008	–0.877	3.260	–1.5	1.2	28.3
syn –BzG:*anti* –G 1 12 –13	1.323	0.334	–0.324	1.2	13.6	20.5	2	1.102	1.069	3.481	0.9	10.3	33.6
syn –BzG:*anti* –G 2 1 –24	0.187	–0.098	0.193	–1.6	–15.5	–0.9	3	–0.063	–0.906	3.421	1.8	3.3	31.2
syn –BzG:*anti* –G 2 2 –23	–0.144	0.017	0.184	–5.2	–14.7	0.7	2	0.376	–0.108	3.342	1.8	0.5	36.8
syn –BzG:*anti* –G 2 3 –22	0.111	–0.050	–0.004	–5.3	–7.1	0.0	3	–0.084	–0.116	3.089	–1.8	8.2	25.4
syn –BzG:*anti* –G 2 4 –21	–0.146	–0.056	0.120	6.5	–3.3	–0.2	3	–0.033	–0.779	3.424	0.2	4.1	30.8
syn –BzG:*anti* –G 2 5 –20	–0.121	–0.059	0.003	4.0	–9.3	–0.1	3	0.125	–0.512	3.286	0.8	1.0	37.0
syn –BzG:*anti* –G 2 6 –19	0.166	–0.058	–0.110	6.3	–12.0	0.2	3	–0.618	–2.241	6.896	1.7	15.5	59.0
syn –BzG:*anti* –G 2 8 –17	0.231	–0.078	–0.085	–7.7	–10.0	0.0	3						
syn –BzG:*anti* –G 2 9 –16	0.083	–0.045	0.115	–8.6	–5.1	0.2	3	0.772	0.108	3.356	0.4	3.8	36.9
syn –BzG:*anti* –G 2 10 –15	0.069	0.033	–0.037	–3.5	–10.0	1.6	2	–0.296	–0.341	3.187	0.8	10.1	24.7
syn –BzG:*anti* –G 2 11 –14	–0.184	0.033	0.077	–0.3	–13.5	–1.0	2	–0.418	–0.714	3.278	–1.0	–1.2	31.9
syn –BzG:*anti* –G 2 12 –13	0.417	–0.028	–0.221	3.0	–1.3	3.1	3	0.842	0.799	3.361	4.2	3.8	35.7
syn –BzG:*anti* –G 3 1 –24	0.184	–0.096	0.257	–3.9	–17.3	–1.1	3	–0.126	–0.859	3.459	2.175	3.688	31.897
syn –BzG:*anti* –G 3 2 –23	–0.173	0.014	0.218	–9.6	–15.4	0.3	2	0.698	0.188	3.377	2.046	0.505	39.017
syn –BzG:*anti* –G 3 3 –22	0.106	–0.032	0.071	–9.9	–3.4	0.3	3	–0.114	–0.103	2.934	–1.013	8.591	18.780
syn –BzG:*anti* –G 3 4 –21	–0.133	–0.044	0.144	8.9	–1.2	–0.1	3	–0.194	–0.540	3.429	–0.088	3.600	33.927
syn –BzG:*anti* –G 3 5 –20	–0.110	–0.044	0.063	6.8	–9.2	0.1	3	–0.093	–0.747	3.344	–0.079	1.509	36.253
syn –BzG:*anti* –G 3 6 –19	0.044	–0.057	–0.080	5.9	–10.8	–1.0	3	0.253	–2.340	6.864	1.721	8.095	60.462
syn –BzG:*anti* –G 3 8 –17	0.160	–0.051	–0.075	–3.4	–7.9	0.1	3						

syn –BzG:*anti* –G 3 9 –16	0.118	–0.060	–0.004	–2.6	–5.5	–0.2	3	0.139	–0.490	3.353	0.246	4.158	31.935
syn –BzG:*anti* –G 3 10 –15	0.091	0.017	–0.126	–6.7	–13.4	0.8	2	–0.592	–0.267	3.413	0.225	8.852	31.127
syn –BzG:*anti* –G 3 11 –14	–0.181	0.046	0.131	–4.4	–13.1	–1.8	2	–0.208	–0.842	3.290	–1.480	–0.534	31.441
syn –BzG:*anti* –G 3 12 –13	0.311	–0.076	–0.192	3.7	0.0	0.9	3	0.788	0.863	3.268	4.939	3.452	33.387
syn –BzG:anti –T 1 1 –24	0.127	–0.073	0.222	–3.0	–13.1	–0.9	3	0.408	–0.776	3.275	2.0	1.1	31.9
syn –BzG:anti –T 1 2 –23	–0.123	0.011	0.184	–0.5	–11.0	1.7	2	–0.480	–0.550	3.164	1.4	2.9	28.9
syn –BzG:anti –T 1 3 –22	0.104	–0.048	–0.147	7.1	–12.1	0.8	3	–0.290	–0.274	3.321	–1.9	12.4	32.2
syn –BzG:anti –T 1 4 –21	–0.195	–0.090	0.065	6.0	–8.8	–0.7	3	0.385	–1.281	3.619	1.7	6.5	30.4
syn –BzG:anti –T 1 5 –20	–0.178	–0.074	–0.136	–7.3	–15.4	0.4	3	–0.006	–0.362	3.319	0.5	1.1	38.4
syn –BzG:anti –T 1 6 –19	0.194	–0.081	–0.034	–8.6	–10.2	1.0	3	0.309	–0.566	6.744	4.3	12.1	56.2
syn –BzG:anti –T 1 8 –17	0.139	–0.054	–0.044	–0.6	–2.3	–0.7	3						
syn –BzG:anti –T 1 9 –16	–0.023	–0.048	0.167	–5.4	–9.4	–0.4	3	0.308	0.555	3.548	0.3	–6.2	42.7
syn –BzG:anti –T 1 10 –15	0.122	0.015	0.051	–6.7	–9.3	–0.8	2	–1.062	–0.215	3.325	–2.0	8.9	29.5
syn –BzG:anti –T 1 11 –14	–0.134	0.015	0.210	–2.2	–7.1	–0.9	2	0.008	–0.717	3.291	–0.6	0.0	29.6
syn –BzG:anti –T 1 12 –13	0.287	–0.059	–0.346	14.1	2.7	–0.3	3	1.035	1.932	3.072	8.2	4.9	28.4
syn –BzG:anti –T 2 1 –24	0.142	–0.085	0.314	–5.1	–16.5	–1.0	3	–0.161	–0.659	3.330	2.4	4.2	29.7
syn –BzG:anti –T 2 2 –23	–0.167	0.009	0.158	–5.2	–14.1	0.1	2	0.666	0.210	3.358	2.3	–0.6	38.6
syn –BzG:anti –T 2 3 –22	0.099	–0.053	–0.001	–5.4	–7.0	–0.1	3	–0.549	0.198	3.150	–2.3	6.7	29.0
syn –BzG:anti –T 2 4 –21	–0.134	–0.067	0.112	3.5	–4.8	–0.8	3	0.419	–0.687	3.353	0.2	5.0	28.4
syn –BzG:anti –T 2 5 –20	–0.090	–0.046	0.118	4.1	–9.1	–0.1	3	0.005	–0.262	3.284	3.0	–1.0	36.5
syn –BzG:anti –T 2 6 –19	0.245	–0.079	–0.159	7.6	–11.2	0.7	3	0.790	–1.184	7.256	0.9	5.4	63.2
syn –BzG:anti –T 2 8 –17	0.142	–0.061	–0.251	1.2	–5.0	0.6	3						
syn –BzG:anti –T 2 9 –16	0.103	–0.062	–0.021	–2.9	–6.7	–0.4	3	0.044	–0.120	3.515	–0.6	1.1	36.6
syn –BzG:anti –T 2 10 –15	0.078	0.021	–0.090	–6.1	–10.9	1.6	2	–0.506	–0.240	3.399	–0.1	8.1	30.6
syn –BzG:anti –T 2 11 –14	–0.132	0.028	0.053	–0.4	–11.3	0.5	2	–0.113	–0.748	3.224	–0.9	1.1	28.7
syn –BzG:anti –T 2 12 –13	0.310	–0.076	–0.143	3.3	–3.1	0.6	3	0.605	0.340	3.347	2.9	2.6	35.9

syn –BzG:anti –T 3 1 –24	0.165	–0.037	0.179	–3.9	–16.5	–2.4	3	–0.003	–0.781	3.401	1.8	3.6	31.1
syn –BzG:anti –T 3 2 –23	–0.148	0.010	0.171	–6.7	–15.6	0.9	2	0.478	–0.030	3.362	1.9	0.7	37.2
syn –BzG:anti –T 3 3 –22	0.102	–0.046	0.027	–7.3	–6.8	0.0	3	–0.319	–0.028	3.089	–1.6	8.0	26.4
syn –BzG:anti –T 3 4 –21	–0.137	–0.061	0.100	4.6	–2.9	–0.6	3	0.365	–0.734	3.339	0.1	4.2	29.2
syn –BzG:anti –T 3 5 –20	–0.094	–0.037	0.088	6.4	–8.5	0.2	3	–0.297	–0.490	3.285	1.4	0.7	34.7
syn –BzG:anti –T 3 6 –19	0.159	–0.084	–0.165	7.8	–12.6	0.3	3	1.068	–1.820	7.024	1.9	12.3	65.4
syn –BzG:anti –T 3 8 –17	0.083	–0.038	–0.030	–3.2	–6.9	1.1	3						
syn –BzG:anti –T 3 9 –16	0.088	–0.055	0.065	–4.8	–5.1	–0.4	3	0.040	–0.267	3.439	0.1	1.9	34.6
syn –BzG:anti –T 3 10 –15	0.083	0.014	–0.079	–3.8	–11.1	1.3	2	–0.642	–0.197	3.309	0.1	9.5	28.0
syn –BzG:anti –T 3 11 –14	–0.140	0.024	0.088	–1.3	–12.1	0.1	2	–0.119	–0.771	3.304	–1.0	–0.2	30.7
syn –BzG:anti –T 3 12 –13	0.577	0.011	–0.166	1.2	2.3	5.5	3	0.924	0.767	3.392	3.2	4.7	36.4

Table A.3. AMBER (ff99SB) interaction energies (kJ mol $^{-1}$) between BzG in the G^3 position of the *Nar*I sequence and the surrounding nucleobases from 20 ns MD simulations.

Simulation	Cluster	Occupancy	BzG[a]	dNTP[b]	π–interactions[c]	Hbond[d]	Total Interaction[e]
***anti* –G:*anti* –C**	all	1.00	–67.4	–35.1	–102.5	–118.4	–220.9
***anti* –BzG:*anti* –C**	all	1.00	–88.2	–44.7	–132.9	–42.1	–175.0
***anti* –BzG:*syn* –A**	all	1.00	–80.5	–40.1	–120.6	–41.7	–162.3
***anti* –BzG:*anti* –T**	all	1.00	–98.6	–29.6	–128.2	–20.4	–148.6
***anti* –BzG:*syn* –G**	φ: 0 –120°	0.19	–89.6	–26.5	–116.1	–10.6	–126.7
	φ: 120 –240°	0.74	–70.7	–15.4	–86.1	–7.7	–93.8
	φ: 240 –360°	0.08	–90.3	–21.0	–111.3	–12.3	–123.6
***syn* –BzG:*anti* –C**	χ: 5 –200°	0.73	–83.0	–47.9	–130.9	–46.5	–177.4
	χ: 200 –365°	0.27	–85.5	–57.6	–143.1	–45.5	–188.6
***syn* –BzG:*anti* –C**	extrahelical	1.00	–90.5	–17.9	–108.4	–3.4	–112.8
***syn* –BzG:*anti* –G**	φ: 0 –125°	0.23	–66.9	–58.5	–125.4	–48.2	–173.6
	φ: 125 –230°	0.23	–68.7	–58.2	–126.9	–36.2	–163.1
	φ: 230 –360°	0.54	–82.0	–47.8	–129.8	–52.0	–181.8
***syn* –BzG:*anti* –T**	φ: 0 –110°	0.03	–73.9	–41.9	–115.8	–13.5	–129.3
	φ: 110 –220°	0.28	–91.6	–27.1	–118.7	–16.3	–135.0
	φ: 220 –360°	0.7	–88.2	–48.3	–136.5	–11.5	–148.0
***syn* –BzG:*anti* –T**	extrahelical	1.00	–97.7	5.6	–92.1	–1.3	–93.4

[a]π-interactions involving BzG. [b]π -interactions involving dNTP. [c]π -interactions involving BzG or dNTP. [d]Hydrogen-bonding energy between BzG and dNTP. [e]Sum of interactions a-d.

Table A.4. Total free energy (kJ mol^{-1}) of the *Nar*I sequence with the BzG adduct incorporated at the G^3 position and various opposing bases from 20 ns MD simulations.

Simulation	Cluster	Occupancy	TS	Bond	Angle	Dihedral	VdW	Elec	1 –4 VdW	1 –4 Elec	E PB	E cavity	E Complex	G	Relative G
***anti* –BzG: *anti* –C**	all	1.00	–210.8	860.7	1727.5	2138.9	–1707.3	15489	870.6	–14768.4	–26046.2	107	–21328.1	–21117.3	6.2
***syn* –BzG: *anti* –C**	χ: 5 –200°	0.73	–200.6	858.9	1752	2112.4	–1693.2	15350.3	868.3	–14791.1	–25885.4	106.5	–21321.2	–21120.6	2.9
	χ: 200 –365°	0.27	–215.9	851.9	1754.1	2123.9	–1707.1	15331.3	869	–14791.8	–25873.7	106.3	–21336.1	–21120.2	3.4
***syn* –BzG: *anti* –C**	extrahelical	1.00	–204.4	866.9	1741.0	2108.0	–1710.1	15561.4	866.4	–14771.3	–26095.2	104.9	–21327.9	–21123.5	0.0
***anti* –BzG: *anti* –T**	all	1.00	–217.4	853.6	1723.6	2157.8	–1735.5	15073.1	868.2	–13762.3	–26106.2	106.7	–20820.9	–20603.5	0.0
***syn* –BzG: *anti* –T**	ϕ: 0 –110°	0.26	–205	882.4	1789.2	2125.9	–1745.3	15498.4	854.1	–13813.1	–26414.3	105.9	–20716.8	–20511.7	91.7
	ϕ: 110 –220°	0.11	–203.6	857.3	1725.5	2155.7	–1724.3	15216.1	870.6	–13781.3	–26186.9	106.2	–20761.1	–20557.5	46
	ϕ: 220 –360°	0.70	–209.3	854	1740.9	2133.2	–1738	15161.1	863.8	–13784.4	–26122.5	105.6	–20786.3	–20576.9	26.6
***syn* –BzG: *anti* –T**	extrahelical	1.00	–188.1	857.9	1737.4	2116.5	–1693.2	15264.5	865.6	–13765.4	–26280.4	107.5	–20789.5	–20601.5	2.0
***anti* –BzG: *syn* –G**	ϕ: 0 –120°	0.19	–216	856.8	1754.1	2111.3	–1758.4	15750.7	865.7	–14523.9	–26291.5	105.4	–21129.7	–20913.6	79.4
	ϕ: 120 –240°	0.74	–207.1	856.8	1736.9	2125	–1723.6	15636.7	859.9	–14529.2	–26205.7	108	–21135.2	–20928.1	64.9
	ϕ: 240 –360°	0.08	–217.2	891.5	1739.1	2152.7	–1724.7	15423	866.2	–14540.4	–25971.4	107.2	–21056.9	–20839.7	153.3
***syn* –BzG: *anti* –G**	ϕ: 0 –125°	0.23	–208.7	878.4	1737	2131.8	–1745.4	15456.6	865.2	–14491.3	–26118	105.6	–21179.9	–20971.3	21.7
	ϕ: 125 –230°	0.23	–205.9	863.9	1752.8	2122.2	–1747.5	15598.9	866.5	–14541.3	–26208.8	106.2	–21187.2	–20981.3	11.7
	ϕ: 230 –360°	0.54	–205.4	871.6	1734	2108.3	–1738.4	15407	859	–14501.5	–26044.5	106.1	–21198.4	–20993	0.0

Table A.5. MM –PBSA free energy of helix binding (kJ mol $^{-1}$) for the *Nar*I sequence with the BzG adduct incorporated at the G^3 position and various opposing bases from 20 ns MD simulations.

Simulation	Cluster	Occupancy	TS	E VdW	E Elec	E PB	E cavity	E	ΔG	Relative ΔG
anti –G: *anti* –C	all	1.00	–206.1	–239.6	7989	–8151.4	–23.7	–425.9	–219.8	0.0
anti –BzG: *syn* –A	all	1.00	–212.5	–253.3	7874	–8002.8	–24.5	–406.9	–194.5	105.9
anti –BzG: *anti* –T	all	1.00	–217.4	–264.8	7943	–8066.1	–25.0	–412.8	–195.4	102.2
syn –BzG: *anti* –T	extrahelical	1.00	–45.0	–65.3	1907.3	–1932.2	–6.5	–96.7	–216.4	14.2
anti –BzG: *anti* –C	all	1.00	–210.8	–255.4	7946	–8069.5	–24.9	–403.8	–193.0	112.2
syn –BzG: *anti* –C	χ: 5 –200°	0.73	–200.6	–278.5	7762	–7892.5	–27.2	–436.7	–236.1	–68.3
	χ: 200 –365°	0.27	–215.9	–295.5	7731	–7849.6	–28.1	–442.5	–226.6	–28.5
syn –BzG: *anti* –C	extrahelical	1.00	–48.8	–77.8	1870.9	–1893.8	–7.3	–108.0	–247.4	–115.5
syn –BzG: *anti* –G	ϕ: 125 –230°	0.23	–208.7	–261.3	7940	–8060.3	–25.3	–406.5	–197.8	91.9
	ϕ: 230 –360°	0.23	–205.9	–260.4	7957	–8074.0	–25.3	–402.8	–196.9	95.9
	ϕ: 0 –110°	0.54	–205.4	–243.4	7858	–7995.7	–23.9	–405.3	–199.9	82.9

Table A.6. Overview of new nomenclature for adducted DNA conformations.

Conformation	Location of Bulky Group	Orientation of Adduct
$\mathbf{B}_{\mathbf{WC}}^{\mathbf{G}}$	Major groove	*anti*
$\mathbf{B}_{\mathbf{WC}}^{\mathbf{5'T}}$	Major groove T –shaped interaction with 5′ base	*anti*
$\mathbf{B}_{\mathbf{WC}}^{\mathbf{3'T}}$	Major groove T –shaped interaction with 3′ base	*anti*
$\mathbf{W}_{\mathbf{H}}^{\mathbf{G}}$	Minor groove	*syn*
$\mathbf{W}_{\mathbf{H}}^{\mathbf{5'T}}$	Minor groove T –shaped interaction with 5′ base	*syn*
$\mathbf{W}_{\mathbf{H}}^{\mathbf{3'T}}$	Minor groove T –shaped interaction with 3′ base	*syn*
$\mathbf{S}_{\mathbf{WC}}^{\mathbf{P}}$	Intercalated and beside pairing base	*anti*
$\mathbf{S}_{\mathbf{H}}^{\mathbf{P}}$	Intercalated and beside pairing base	*syn*
$\mathbf{S}_{\mathbf{WC}}^{\mathbf{I}}$	Intercalated and stacking with pairing base	*anti*
$\mathbf{S}_{\mathbf{H}}^{\mathbf{I}}$	Intercalated and stacking with pairing base	*syn*
$\mathbf{S}^{\mathbf{F}}$	Intercalated and opposing base extrahelical	*anti* or *syn*

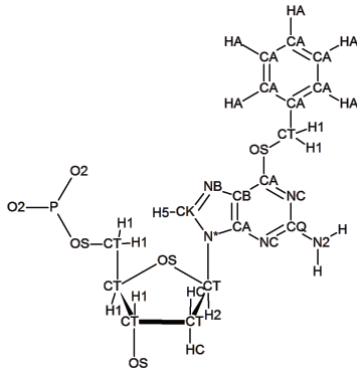

Figure A.1. AMBER atom types for the BzG adduct.

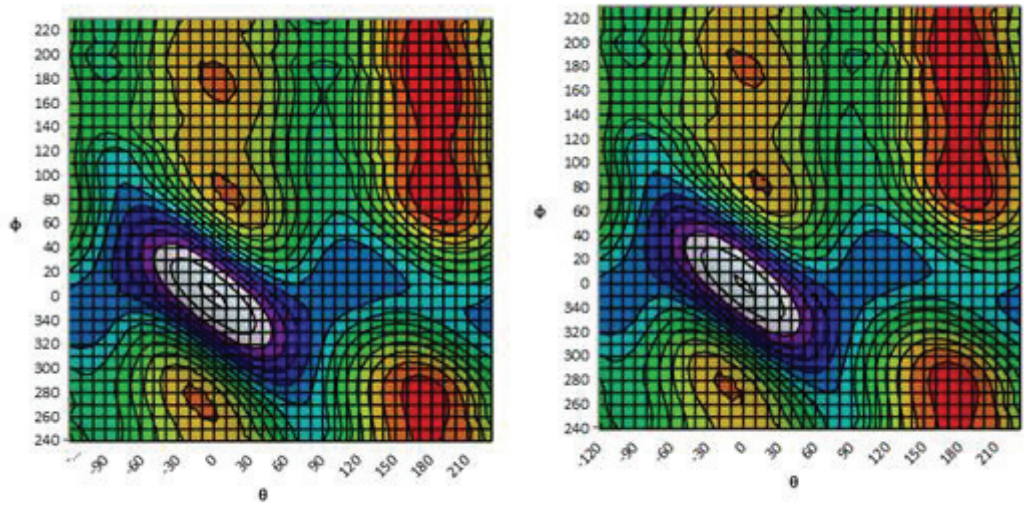

Figure A.2. PES (kJ mol^{-1}) as a function of θ and φ (degrees) in BzG (where red is the lowest energy region and each color change represents 5 kJ mol^{-1}) with χ constraint to 240° (left) or 80° (right)

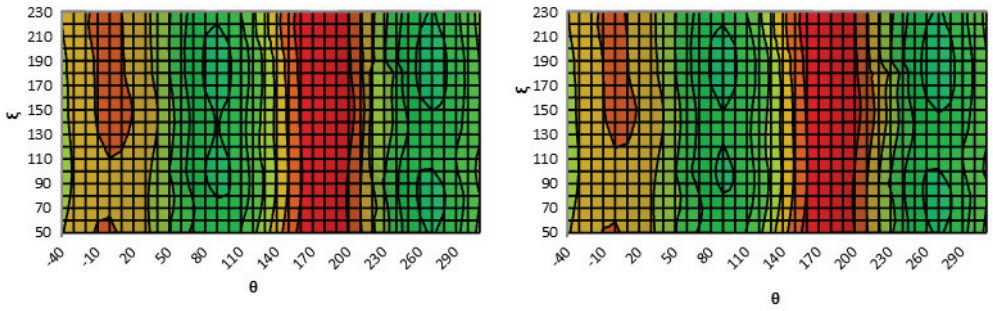

Figure A.3 PES (kJ mol $^{-1}$) as a function of θ and ξ (degrees) in BzG (where red is the lowest energy region and each color change represents 5 kJ mol $^{-1}$) with χ constraint to 240° (left) or 80° (right).

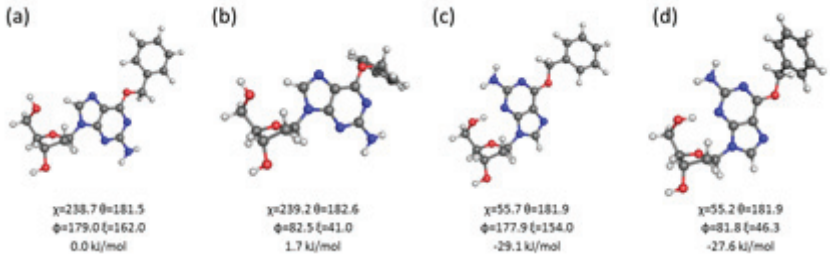

Figure A.4. Structures of the minima BzG non –constrained nucleosides and notable dihedrals (degrees).

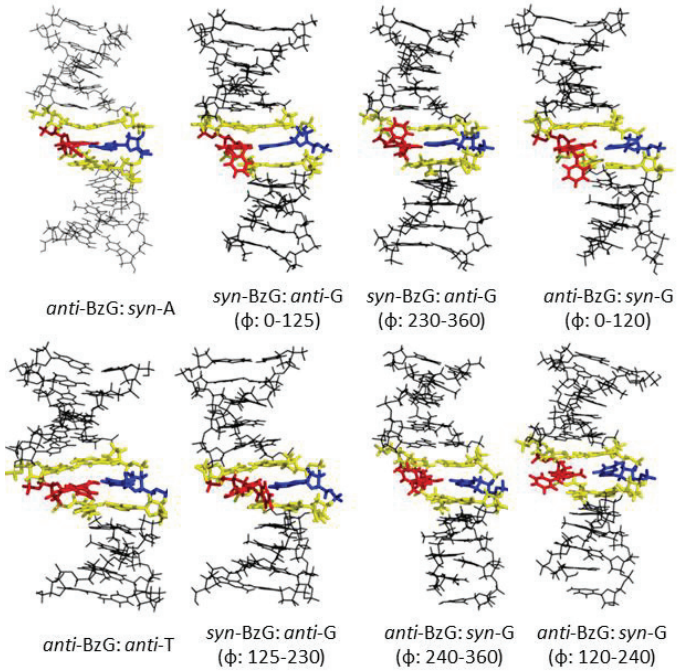

Figure A.5. Representative structure from 20 ns MD simulations on the *Nar*I DNA helixes with the BzG adduct (red) incorporated at the G^3 position. The pairing nucleotide is shown in blue and the flanking bases are shown in yellow.

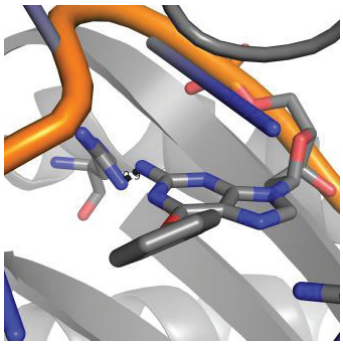

Figure A.6. Close contact between R332 and *syn*–BzG in Dpo4 with the BzG moiety located in the DNA minor groove (PDB ID: 1S0M).

Appendix B

Supplementary Information for Chapter 3: Conformational Flexibility of the Benzyl-Guanine Adduct in a Bypass Polymerase Active Site Permits Replication: Insights from Molecular Dynamics Simulations

Contains a list of Dpo4 crystal structure considered in present work, Tables B.1–B.4 and Figures B.1–B.5

Dpo4 Crystal Structures Considered in the Present Work

5EWD, 1S0M, 1S0O, 2AGQ, 2ASD, 2ATL, 2C22, 2J6S, 2J6T, 2R8I, 2UVR, 2UVU, 2UVV, 2UVW, 2V4Q, 2V4R, 2W8L, 2W9A, 2XCP, 3GII, 3GIJ, 3GIK, 3GIM, 3KHL, 3KHR, 3QZ7, 3QZ8, 3RAQ, 3RAX, 3RB3, 3RB4, 3RB6, 3RBD, 3RBE, 3T5H, 3T5J, 3T5K, 3T5L, 3V6H, 3V6J, 4JUZ, 4JV0, 4JV1, 4JV2, 4QW8, 4QWB

Table B.1. Average structural parameters with standard deviations during MD simulations on the preinsertion or dCTP insertion complex for the replication of dG or Bz-dG by Dpo4.

	dG	dG:dCTP	*anti*-Bz-dG	*syn*-Bz-dG	*anti*-Bz-dG:dCTP	*syn*-Bz-dG:dCTP
rmsd	1.641±0.353 Å	1.131±0.145 Å	1.578±0.236 Å	1.284±0.201 Å	1.178±0.122 Å	1.249±0.188 Å
dG* χ[a]	236.1±22.0°	255.5±12.9 °	246.4±15.1°	42.0±15.7°	267.2±13.7°	29.3±14.8°
dG* pucker	C2′-*endo*	C2′-*endo*	C2′-*endo*	C1′–*exo*	C2′-*endo*	C1′–*exo*
Bz-dG* θ[a]	NA[b]	NA[b]	149.9±52.3°	1.2±50.1°	106.5±116.1°	294.9±86.0°
Bz-dG* φ[a]	NA[b]	NA[b]	177.9±12.7°	179.0±9.8°	177.0±10.2°	178.8±10.4°
Bz-dG* ξ[a]	NA[b]	NA[b]	3.9±97.9°	355.0±58.3°	316.4±92.5°	50.0±87.3°
dCTP χ[a]	NA[b]	225.5±10.1°	NA[b]	NA[b]	232.3±16.2°	255.6±14.8°
dCTP pucker	NA[b]	C1′–*exo*	NA[b]	NA[b]	C1′–*exo*	C1′–*exo*
dG*:dCTP d(C1′–C1′)	NA[b]	10.773±0.138 Å	NA[b]	NA[b]	11.371±0.319 Å	11.082±0.263 Å

[a]dG* = dG or Bz-dG. See Figure 1 for the definitions of key nucleoside dihedral angles. [b]Not applicable.

Table B.2. Average and standard deviations for stacking interactions (kcal/mol) between the dG*:dCTP and 3′–dC:dG pairs, and dG* and 5′–dC (dG * = dG or Bz-dG) during MD simulations on the Dpo4 preinsertion and insertion complexes.

	dG	dG:dCTP[a]	*anti*-Bz-dG	*syn*-Bz-dG	*anti*-Bz-dG:dCTP[a]	*syn*-Bz-dG:dCTP[a]
dG*:5'–dC	–0.5±0.5	–0.3±0.1	–0.3±0.4	–1.2±0.4	–2.3±1.3	–1.3±0.4
dG*:3'–dC	–6.1±0.9	–6.7±0.7	–7.1±0.8	–7.2±0.6	–6.8±0.9	–5.9±0.7
dG*:3'–dG	–0.9±0.5	–2.0±0.5	–6.2±1.0	–6.4±0.9	–2.0±0.6	–3.7±0.9
dCTP:3'–dC	NA[b]	–6.0±0.6	NA[b]	NA[b]	–5.0±1.0	–3.1±0.9
dCTP:3'–dG	NA[b]	–0.7±0.2	NA[b]	NA[b]	–1.0±0.5	–1.5±0.7

[a]Strength of the (AMBER) stacking between dG or Bz-dG and the indicated flanking base. [b]Not applicable.

Table B.3. Occupancies (percentage of simulation time), average C1′–C1′ distance with standard deviation (Å), and average hydrogen-bond strength with standard deviation (kcal/mol) for the 3′–dC:dG pair with respect to dG or Bz-dG during MD simulations on the corresponding Dpo4 preinsertion and insertion complexes.

	dG(O6)···dC(HN4)[a]	dG(N1H)···dC(N3)[a]	dG(N2H)···dC(O2)[a]	d(C1′–C1′)	3′ Base Pair Hydrogen Bond[b]
dG	99%	100%	100%	10.740±0.163 Å	–28.1±2.8
dG:dCTP	99%	100%	100%	10.728±0.158 Å	–28.9±2.6
anti-Bz-dG	97%	98%	100%	10.619±0.186 Å	–27.6±3.0
syn-Bz-dG	98%	100%	100%	10.708±0.171 Å	–28.2±2.8
anti-Bz-dG:dCTP	92%	99%	100%	10.606±0.188 Å	–27.7±3.3
syn-Bz-dG:dCTP	100%	100%	100%	10.740±0.155 Å	–29.0±2.5

[a]Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of > 120°. [b]Strength of the (AMBER) hydrogen bond in the base pair 3′ with respect to dG or Bz-dG.

Table B.4. Average reaction parameters with standard deviations during MD simulations on the dCTP insertion complex for the replication of dG or Bz-dG by Dpo4.

	Reaction Distance[a]	Reaction Angle[b]	Coordination Catalytic Mg^{2+}[c]	Coordination Binding Mg^{2+}[c]
dG:dCTP	3.529±0.158 Å	168.8±5.8°	6/6	5/6 (PβO)
anti-Bz-dG:dCTP	3.477±0.152 Å	168.4±5.9°	6/6	5/6 (PβO)
syn-Bz-dG:dCTP	3.425±0.217 Å	168.9±5.7°	6/6	5/6 (PβO)

[a]Pα···O3′ distance. [b]∠(PαβPαO3′) angle. [c]Number of coordinated atoms within < 2.5 Å for < 90% of the simulation/Number of possible coordinated atoms, with atom not coordinated to Mg^{2+} throughout the simulation indicated in parentheses. See Figure S2 for coordination of Mg^{2+} ions.

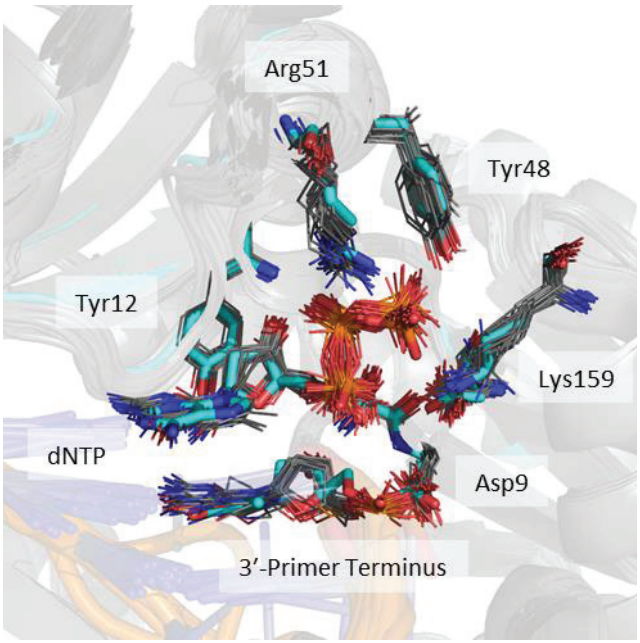

Figure B.1. Overlay of the active site region of 45 crystal structures of Dpo4 ternary complexes obtained under a variety of crystallization conditions. The active site from 1S0M, which was used as the starting point for the present work, is highlighted in cyan.

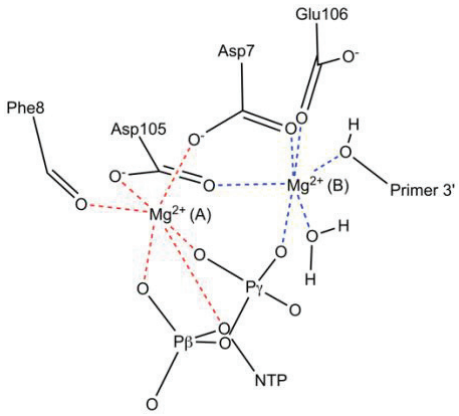

Figure B.2. Coordination of the dNTP binding [Mg^{2+} (A)] and catalytic [Mg^{2+} (B)] Mg^{2+} ions.

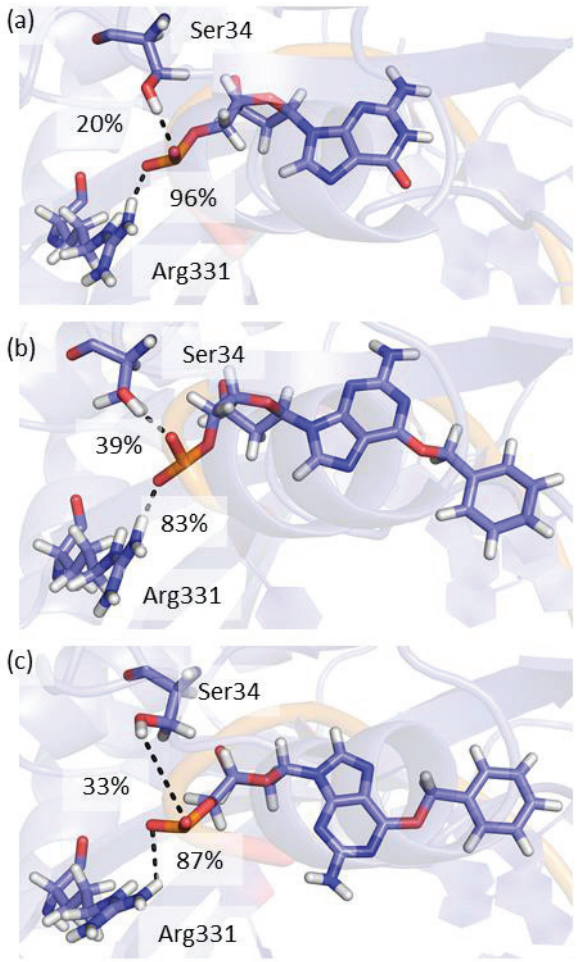

Figure B.3. Representative MD structures of the (a) dG, (b) *anti*-Bz-dG, and (c) *syn*-Bz-dG preinsertion complexes for replication by Dpo4 depicting the interactions between the template base and surrounding amino acids.

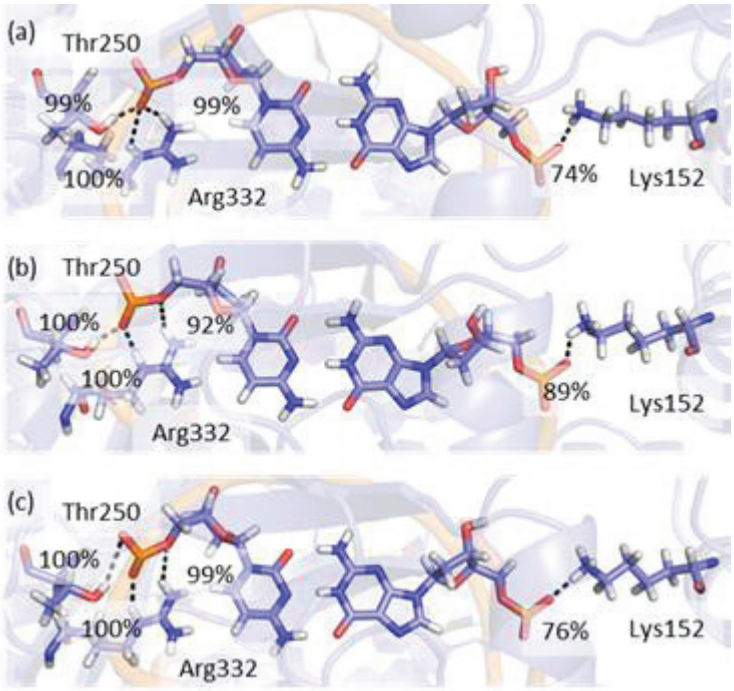

Figure B.4. Representative MD structures of the (a) dG, (b) *anti*-Bz-dG, and (c) *syn*-Bz-dG preinsertion complexes for replication by Dpo4 depicting the interactions between the 3′–dC:dG base pair and surrounding amino acids.

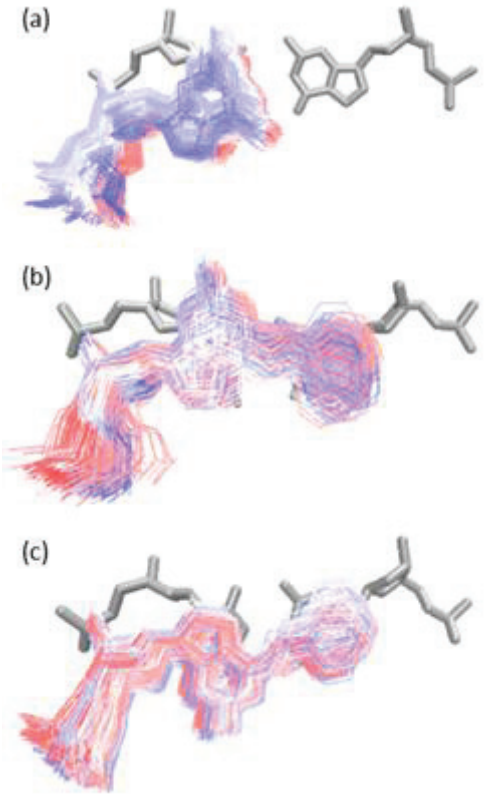

Figure B.5. Overlay of orientations adopted throughout the MD simulation with the simulation time indicated in color (red [0 ns] to white [50 ns] to blue [100 ns]) for the (a) dG, (b) *anti*-Bz-dG, and (c) *syn*-Bz-dG preinsertion complexes for replication by Dpo4. Only the 3′–dC:dG pair is shown for simplicity.

Appendix C

Supplementary Information for Chapter 4: Molecular Insights into the Translesion Synthesis of Benzyl-Guanine from Molecular Dynamics Simulations: Structural Evidence for Mutagenic and Nonmutagenic Replication

Contains full computational details, Figures C.1–C.10 and Tables C.1–C.10.

Full Computational Details

Starting Geometries: Initial structures for the MD simulations were obtained from chains B, E and F of a crystal structure of the Dpo4 ternary complex corresponding to the pairing of dATP opposite the 5′-dT with respect to the benzo[a]pyrene adenine adduct (PDB ID: 1S0M).[1] This crystal structure was chosen due to the presence of a dNTP (rather than a ddNTP), the near perfect coordination sphere of the catalytic divalent ions, and the lack of active site distortions. Two excess divalent ions (Ca405 and Ca408) were removed and the two remaining divalent ions (Ca403 and Ca404) were changed to the catalytic Mg^{2+} ions. We note that the position of the Mg^{2+} ions changes upon equilibration relative to the crystallographic Ca^{2+} ions in order to achieve the coordination required for catalysis (i.e., coordination of the terminal O3′ of the primer strand to the catalytic Mg^{2+} ion). Subsequently, the template DNA strand sequence was modified to 5′–CG*CCATCGCC for the ternary insertion complexes or 5′–GCG*CCATCCCC for the ternary –1 base deletion complexes, with Bz-dG positioned at G*. These sequences parallel that used in a previous MD study on Bz-dG adducted DNA.[2] To generate the ternary insertion complexes, Bz-dG was paired opposite each natural dNTP. In each starting complex, four key dihedral angles in the adducted nucleotide (Figure 4.1) were initially set to the lowest energy orientations previously determined using density functional theory (DFT) calculations on nucleoside models, and MD simulations on adducted DNA helices.[2] Specifically, the initial Bz-dG conformation contained $\theta \approx 0°$ (bulky moiety towards N7 of G), $\varphi \approx 180°$ (bulky moiety in a planar extended conformation), and $\xi \approx 0°$ (phenyl ring in the same plane as G and the methylene linker). Furthermore, χ was adjusted to the *anti* or *syn* orientation, such that the nascent base pairs considered include *anti*-Bz-dG paired

opposite *anti*-dCTP, *anti*-dTTP, *syn*-dGTP, or *syn*-dATP, and *syn*-Bz-dG paired opposite *anti*-dCTP, *anti*-dTTP, *anti*-dATP, or *anti*-dGTP. Additionally, a control simulation was performed for the ternary insertion complex with *anti*-dG at G^* paired opposite *anti*-dCTP. To generate the ternary –1 base deletion complexes, no base was positioned opposite *anti* or *syn*-Bz-dG, and dGTP was paired opposite the 5′-dC with respect to the adduct.

The resulting 11 DNA–polymerase complexes were prepared for minimization using the tleap module of Amber 11.[3] Specifically, hydrogen atoms were added to generate the natural protonation states of all DNA and protein residues. Furthermore, the systems were neutralized with Na^+ ions and solvated in a TIP3P octahedral water box such that the DNA–polymerase complex was at least 8.0 Å from the edge of the box. The resulting complexes each contain 341 amino acids, a dNTP, 18 (insertion) or 19 (deletion) nucleotides, 2 Mg^{2+} ions, 5 (insertion) or 6 (deletion) Na^+ ions, and ~900 water molecules. All natural amino acids, nucleotides, and the solvent were modeled with AMBER ff99SB parameters,[4] while the parameters for Bz-dG[2] and the dNTPs[5-7] were adapted from the literature.

Simulation Procedure: For all systems, the first minimization phase involved 1000 steps of steepest descent minimization, followed by 3000 steps of conjugate gradient minimization, with a 500 kcal mol^{-1} $Å^{-2}$ force constraint on the protein (including the Mg^{2+} ions) and DNA (including the dNTP). Next, the DNA was minimized using 1000 steps of steepest descent minimization, followed by 3000 steps of conjugate gradient minimization, with a 500 kcal mol^{-1} $Å^{-2}$ force constraint on the protein. Finally, 1000 steps of steepest descent minimization, followed by 4000 steps of conjugate gradient minimization, were performed on the entire unconstrained system. Subsequently, the equilibration phase was

completed using a Langevin thermostat ($\gamma = 1.0$) to heat the system from 0 to 300 K over 20 ps, with a 10 kcal mol^{-1} $Å^{-2}$ force constraint on the protein and DNA. Finally, a 20 ns unrestrained production MD simulation was performed on each of the 11 systems at 300 K and 1 bar. Each simulation was stable, with an overall backbone root-mean-square deviation (rmsd) of ~1.3–2.3 Å (Table C.1, Appendix C). To confirm adequate sampling over the 20 ns, the trajectories for the insertion of *anti*-dCTP opposite *anti*-Bz-dG or *anti*-dG were extended to 100 ns. Extending the simulations led to only small deviations in the structures and did not change the overall conclusions (see Tables C.2–C.7, Appendix C and Figures 4.2–4.4, C.2 and C.3 for a comparison of the data), which supports the use of 20 ns trajectories for the remaining systems. Throughout all minimization, equilibration and production calculations, the periodic boundary condition and a non-bonded cutoff of 10 Å were implemented. Additionally, SHAKE and a 0.002 ps time step were used in all equilibration and production steps. All minimization and equilibration calculations were performed using the sander module of Amber 12, while production simulations were performed using the pmemd module of Amber 12.[8] Analysis of the MD simulations was completed using AmberTools 14.[9] Specifically, the average linear interaction energies for discrete hydrogen-bonding (electrostatic component) and stacking (van der Waals component) interactions of the nascent base pair and 5′/3′-bases with respect to dG* were calculated across the simulation trajectory. Additionally, Molecular Mechanics/ Generalized Born Surface Area (MM/GBSA) pairwise energies were calculated for the discrete dNTP–polymerase interactions. We note that it was not feasible to calculate the entropy component of the pairwise energies due to the size of the system. Throughout the paper, average interaction energies are reported with the dynamical information (standard

deviations) provided in the Appendix C. Furthermore, the positions of the nucleobases are referenced with respect to dG*.

References

[1] Ling, H., Sayer, J. M., Plosky, B. S., Yagi, H., Boudsocq, F., Woodgate, R., Jerina, D. M., and Yang, W. (2004) Crystal structure of a benzo[a]pyrene diol epoxide adduct in a ternary complex with a DNA polymerase, *Proc. Natl. Acad. Sci. U.S.A. 101*, 2265-2269.

[2] Wilson, K. A., and Wetmore, S. D. (2014) Complex Conformational Heterogeneity of the Highly Flexible O6-Benzyl-guanine DNA Adduct, *Chem. Res. Toxicol. 27*, 1310-1325.

[3] Case, D. A., Darden, T. A., Cheatham, T. E., III, Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Crowley, M., Walker, R. C., Zhang, W., Merz, K. M., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossvary, I., Wong, K. F., Paesani, F., Vanicek, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Yang, L., Tan, C., Mongan, J., Hornak, V., Cui, G., Mathews, D. H., Seetin, M. G., Sagui, C., Babin, V., and Kollman, P. A. (2008) AMBER Tools, Version 1.0 ed., University of California, San Francisco.

[4] Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., and Shaw, D. E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field, *Proteins: Structure, Function, and Bioinformatics 78*, 1950-1958.

[5] Zhang, L., Rechkoblit, O., Wang, L., Patel, D. J., Shapiro, R., and Broyde, S. (2006) 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. 34*, 3326-3337.

[6] Perlow, R. A., and Broyde, S. (2002) Toward Understanding the Mutagenicity of an Environmental Carcinogen: Structural Insights into Nucleotide Incorporation Preferences, *J. Mol. Biol. 322*, 291-309.

[7] Zhang, L., Shapiro, R., and Broyde, S. (2005) Molecular Dynamics of a Food Carcinogen−DNA Adduct in a Replicative DNA Polymerase Suggest Hindered Nucleotide Incorporation and Extension, *Chem. Res. Toxicol. 18*, 1347-1363.

[8] Case, D. A., Darden, T. A., III, T. E. C., Simmerling, C. L., Wang, J., R.E. Duke, R. L., Walker, R. C., Zhang, W., Merz, K. M., Roberts, B., Hayik, S., Roitberg, A., Seabra, G., Swails, J., Goetz, A. W., Kolossváry, I., Wong, K. F., Paesani, F., Vanicek, J., Wolf, R. M., Liu, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M.-J., Cui, G., Roe, D. R., Mathews, D. H., Seetin, M. G., Salomon-Ferrer, R., Sagui, C., Babin, V., Luchko, T., Gusarov, S., Kovalenko, A., and Kollman, P. A. (2012) AMBER 12, University of California, San Francisco.

[9] Case, D., Babin, V., Berryman, J., Betz, R., Cai, Q., Cerutti, D., Cheatham Iii, T., Darden, T., Duke, R., and Gohlke, H. (2014) Amber 14.

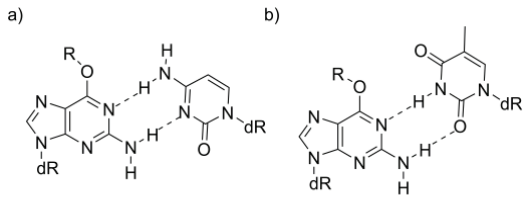

Figure C.1. Previously reported hydrogen-bonding arrangements between an O6-dG alkylation adduct (namely Me-dG (R=CH_3), Bz-dG (R=$C_2H_2C_6H_5$), or POB-dG (R=$C_4C_6OC_5NH_4$)) and (a) dC (wobble base pair) or (b) dT (pseudo Watson-Crick base pair).[23, 27-32]

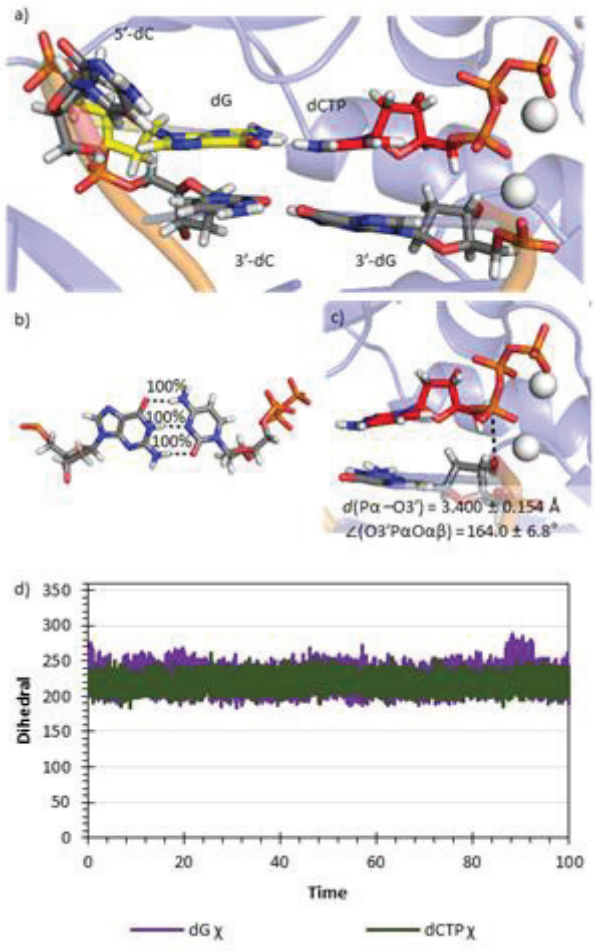

Figure C.2. Representative MD structure for the Dpo4 ternary complex of canonical dG replication from 100 ns MD simulations depicting a) the orientation of dCTP with respect to the DNA, b) the dG:dCTP hydrogen-bonding arrangement and the percent occupancy of the hydrogen-bonding interactions, c) the average reaction parameters, and d) the time evolution of key dG and dCTP dihedral angles.

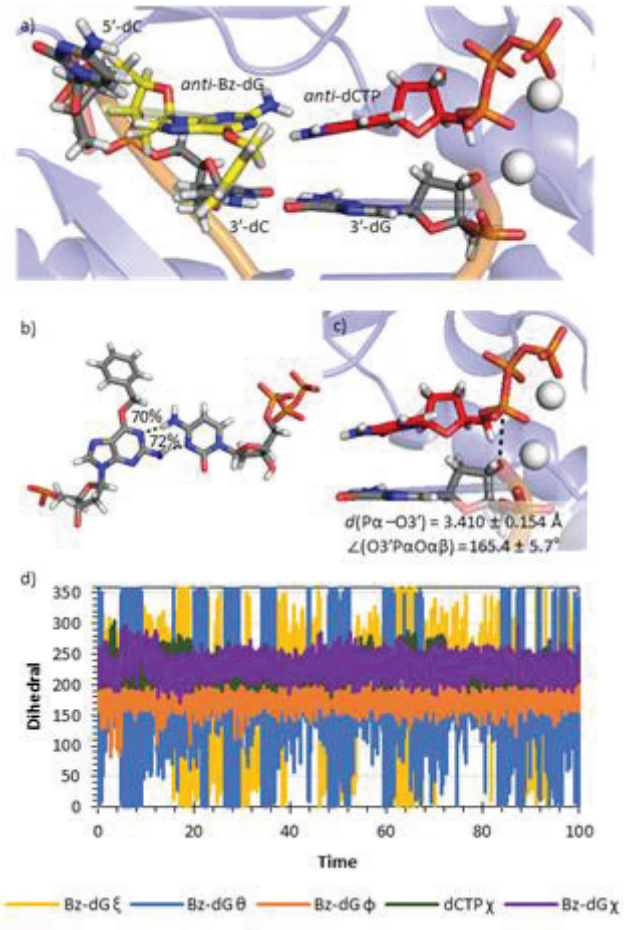

Figure C.3. Representative MD structure for the Dpo4 ternary complex of *anti*-Bz-dG replication from 100 ns MD simulations depicting a) the orientation of dCTP with respect to the DNA, b) the *anti*-Bz-dG:dCTP hydrogen-bonding arrangement and the percent occupancy of the hydrogen-bonding interactions, c) the average reaction parameters, and d) the time evolution of key Bz-dG and dCTP dihedral angles.

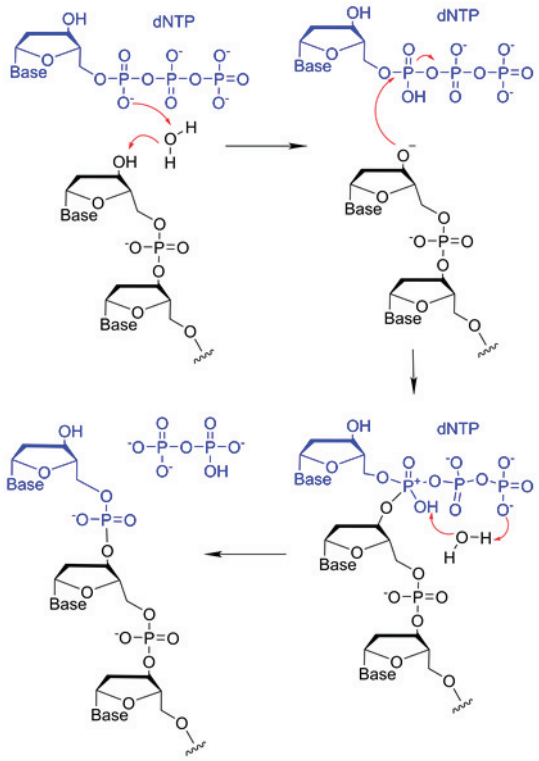

Figure C.4. Previously proposed reaction mechanism for the replication of natural DNA by Dpo4.[54, 55]

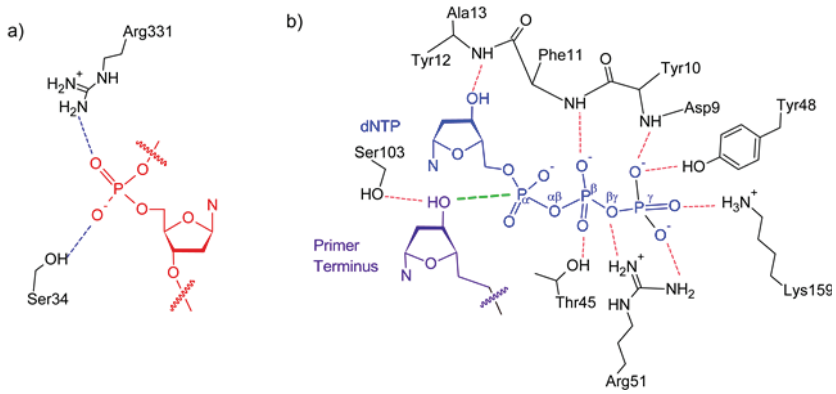

Figure C.5. Previously proposed hydrogen-bonding interactions between Dpo4 and a) the base being replicated or b) the dNTP.[39, 49, 55, 59-61]

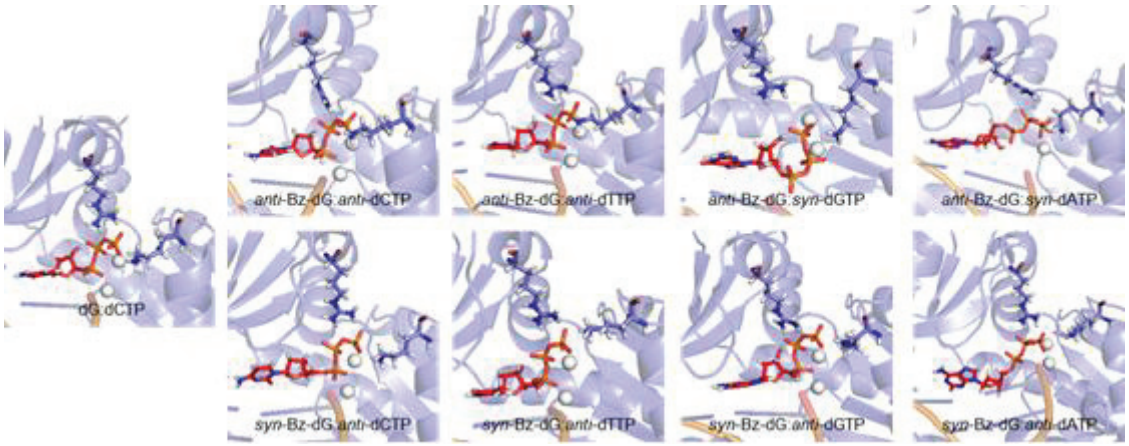

Figure C.6. Representative MD structures depicting the orientation of the dNTP with respect to Lys159 and Arg51 in the Dpo4 ternary complex for dG or Bz-dG replication.

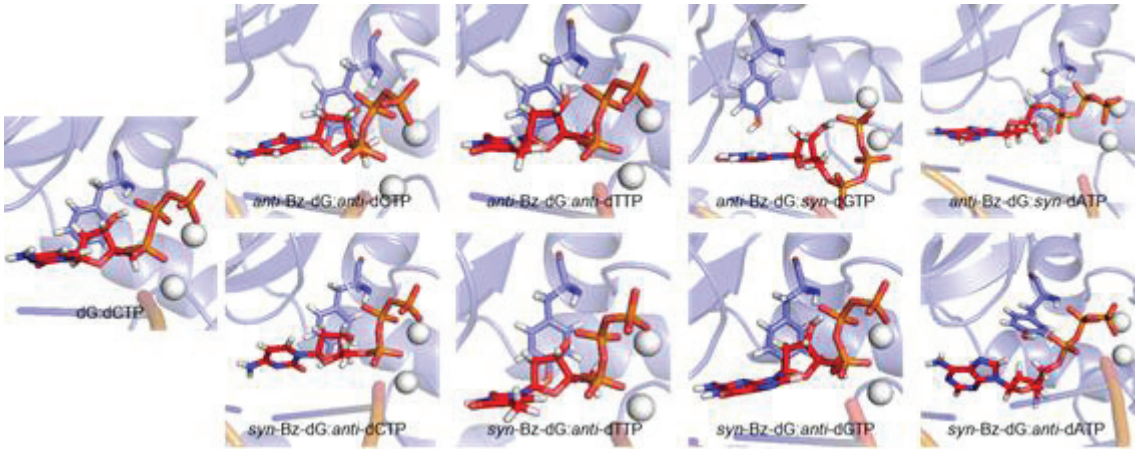

Figure C.7. Representative MD structures depicting the orientation of the dNTP with respect to Tyr12 in the Dpo4 ternary complex for dG or Bz-dG replication.

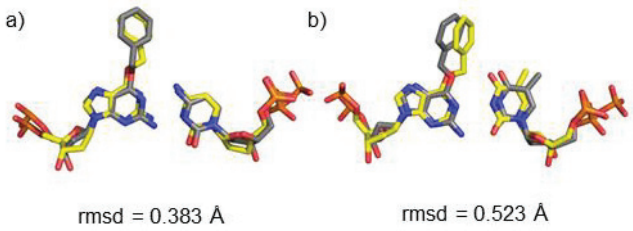

Figure C.8. Overlays of MD representative structures from the ternary insertion complexes (grey) and crystal structure of the Dpo4 post-lesion synthesis complex (yellow) for a) Bz-dG:dC (PDB ID: 2JEF) and b) Bz-dG:dT (PDB ID: 2JEI).

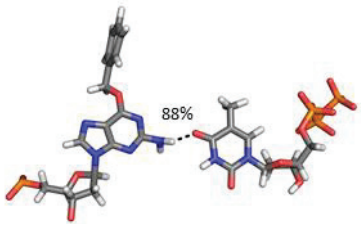

Figure C.9. Alternative hydrogen-bonding geometry for the *anti*-Bz-dG:*anti*-dTTP pair formed during the last 11.4 ns of the simulation on the Dpo4 ternary complex.

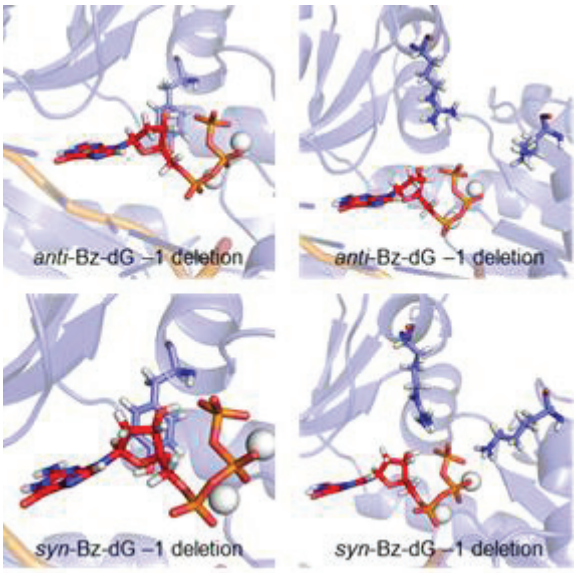

Figure C.10. Representative MD structures depicting the orientation of the dNTP with respect to Tyr12 (left) or Arg51 and Lys159 (right) in the Dpo4 ternary –1 deletion complex during *anti* (top) or *syn* (bottom) Bz-dG replication.

Table C.1. Backbone rmsds throughout the production MD simulation relative to the first frame for the ternary insertion and deletion complexes for Dpo4 replication of dG or Bz-dG.

Complex	Active Site Base Pair	rmsd
Ternary Insertion	*anti*-dG:*anti*-dCTP (20 ns) [a]	1.984±0.296 Å
	anti-dG:*anti*-dCTP (100 ns) [a]	1.224±0.190 Å
	anti-Bz-dG:*anti*-dCTP (20 ns) [a]	1.974±0.234 Å
	anti-Bz-dG:*anti*-dCTP (100 ns) [a]	1.102±0.139 Å
	syn-Bz-dG:*anti*-dCTP	1.945±0.252 Å
	anti-Bz-dG:*anti*-dTTP	1.295±0.181 Å
	syn-Bz-dG:*anti*-dTTP	1.653±0.212 Å
	anti-Bz-dG:*syn*-dGTP	1.968±0.291 Å
	syn-Bz-dG:*anti*-dGTP	1.690±0.203 Å
	anti-Bz-dG:*syn*-dATP	2.338±0.612 Å
	syn-Bz-dG:*anti*-dATP	2.298±0.435 Å
Deletion	*anti*-Bz-dG	1.911±0.273 Å
	syn-Bz-dG	1.381±0.254 Å

[a] Data from 20 and 100 ns simulations.

Table C.2. Structural parameters from MD simulations on the Dpo4 ternary (insertion) replication complexes of dG or Bz-dG.

Active Site Base Pair	dG* χ [a]	dG* Pucker [b]	dG* θ [a]	dG* φ [a]	dG* ξ [a]	dNTP χ	dNTP Pucker [b]	dG*:dNTP C1′–C1′ Distance [c]
anti-dG:*anti*-dCTP (20 ns) [d]	227.2± 14.1°	C1′-*exo* (40.5%)	NA [e]	NA [e]	NA [e]	216.6± 9.5°	C3′-*endo* (47.8%)	10.824± 0.147 Å
anti-dG:*anti*-dCTP (100 ns) [d]	224.4± 15.6°	C1′-*exo* (39.6%)	NA [e]	NA [e]	NA [e]	219.0± 10.2°	C3′-*endo* (36.5%)	10.809± 0.142 Å
anti-Bz-dG:*anti*-dCTP (20 ns) [d]	230.1± 16.9°	C1′-*exo* (43.4%)	168.9± 69.4°	178.4± 13.7°	185.1± 93.6°	226.1± 15.9°	C1′-*exo* (30.9%)	11.669± 0.419 Å
anti-Bz-dG:*anti*-dCTP (100 ns) [d]	233.1± 19.7°	C1′-*exo* (26.0%)	163.9± 83.5°	179.0± 13.7°	260.5± 56.4°	224.2± 14.9°	C1′-*exo* (27.0%)	11.490± 0.404 Å
syn-Bz-dG:*anti*-dCTP	48.2± 13.3°	C1′-*exo* (75.8%)	209.3± 84.0°	177.1± 9.9°	145.6± 90.7°	246.9± 17.6°	C1′-*exo* (50.0%)	11.094± 0.302 Å
anti-BzG:*anti*-TTP (pseudo Watson-Crick) [f]	248.7± 13.4°	C3′-*exo* (62.2%)	87.2± 90.5°	193.6± 15.1°	23.3± 94.6°	219.3± 10.9°	C3′-*endo* (48.2%)	10.815± 0.229 Å
anti-BzG:*anti*-TTP (alternative hydrogen bonding) [f]	254.4± 11.1°	C3′-*exo* (34.5%)	84.4± 120.7°	182.6± 13.9°	351.2± 90.8°	216.4± 10.7°	C3′-*endo* (48.3%)	11.571± 0.335 Å
syn-Bz-dG:*anti*-dTTP	47.1± 11.2°	C1′-*exo* (35.4%)	154.4± 120.0°	179.1± 17.3°	189.8± 102.8°	216.7± 10.0°	C3′-*endo* (39.4%)	11.950± 0.388 Å
anti-Bz-dG:*syn*-dGTP	244.8± 25.3°	C2′-*exo* (36.2%)	167.2± 93.4°	194.9± 41.9°	189.1± 94.3°	12.8± 21.5°	C1′-*exo* (40.7%)	11.541± 2.044 Å
syn-Bz-dG:*anti*-dGTP	40.0± 13.2°	C3′-*endo* (58.3%)	186.9± 58.2°	164.2± 47.4°	165.1± 98.8°	240.9± 10.9°	C3′-*endo* (70.4%)	12.460± 0.403 Å
anti-Bz-dG:*syn*-dATP	256.8± 12.7°	C2′-*exo* (56.3%)	145.0± 52.2°	183.9± 40.6°	174.1± 94.9°	321.7± 15.5°	C1′-*exo* (72.2%)	9.597± 0.318 Å
syn-Bz-dG:*anti*-dATP	53.1± 11.8°	C1′-*exo* (33.1%)	168.7± 73.2°	164.0± 65.6°	152.0± 96.2°	212.3± 25.0°	C2′-*endo* (45.3%)	10.004± 0.856 Å

[a] dG* dihedral angle (degrees). See Figure 1 for definitions of adduct dihedral angles. [b] Most common sugar pucker of dG* or the dNTP, and the percentage of the simulation that the sugar pucker is adopted. [c] Width of the dG* base pair. [d] Data from 20 and 100 ns simulations. [e] Not applicable. [f] See Figures 4 and S9 for a description of the different *anti*-BzG:*anti*-TTP hydrogen-bonding orientations.

Table C.3. Occupancies of the hydrogen bonds in the 3′ and 5′ flanking base pairs in MD simulations on the ternary insertion and deletion complexes for Dpo4 replication of dG or Bz-dG.

Complex	Active Site Base Pair	Location [a]	O6···HN4 [b]	N1H···N3 [b]	N2H···O2 [b]
Ternary Insertion	*anti*-dG:*anti*-dCTP (20 ns) [c]	3′	100%	100%	100%
	anti-dG:*anti*-dCTP (100 ns) [c]	3′	100%	100%	100%
	anti-Bz-dG:*anti*-dCTP (20 ns) [c]	3′	92%	98%	100%
	anti-Bz-dG:*anti*-dCTP (100 ns) [c]	3′	99%	99%	100%
	syn-Bz-dG:*anti*-dCTP	3′	98%	100%	100%
	anti-BzG:*anti*-TTP (pseudo Watson-Crick) [d]	3′	98%	100%	100%
	anti-BzG:*anti*-TTP (alternative hydrogen bonding) [d]	3′	90%	98%	100%
	syn-Bz-dG:*anti*-dTTP	3′	70%	77%	99%
	anti-Bz-dG:*syn*-dGTP	3′	90%	97%	100%
	syn-Bz-dG:*anti*-dGTP	3′	98%	100%	100%
	anti-Bz-dG:*syn*-dATP	3′	98%	100%	100%
	syn-Bz-dG:*anti*-dATP	3′	60%	59%	58%
Deletion	*anti*-Bz-dG	3′	99%	99%	95%
		5′	100%	100%	99%
	syn-Bz-dG	3′	96%	100%	99%
		5′	100%	99%	99%

[a] Location with respect to dG*. [b] Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of < 120°. [c] Data from 20 and 100 ns simulations. [d] See Figures 4 and S9 for a description of the different *anti*-BzG:*anti*-TTP hydrogen-bonding orientations.

Table C.4. Linear interaction energies (kJ mol^{-1}) for stacking and hydrogen-bonding interactions between dG* and the flanking base pairs in MD simulations on the ternary complexes for Dpo4 replication of dG or Bz-dG.

Active Site Base Pair	3' Hydrogen Bonding [a]	dG*:dNTP Interaction [b]	3′ Stacking [c]	5′ Stacking [c]
anti-dG:*anti*-dCTP (20 ns) [d]	–122.0±10.6	–127.6±11.3	–62.9±4.0	–1.0±1.2
anti-dG:*anti*-dCTP (100 ns) [d]	–114.9±31.7	–125.1±24.3	–61.8±8.2	–1.0±0.9
anti-Bz-dG:*anti*-dCTP (20 ns) [d]	–116.2±14.2	–23.0±11.7	–62.7±5.8	–1.3±2.4
anti-Bz-dG:*anti*-dCTP (100 ns) [d]	–119.3±12.7	–23.9±12.2	–61.9±6.1	–6.9±7.4
syn-BzG:*anti*-CTP	–117.2±17.2	–29.3±10.9	–60.5±5.8	–6.2±2.7
anti-BzG:*anti*-TTP (pseudo Watson-Crick) [e]	–120.6±7.4	–16.1±7.9	–69.7±4.6	–3.7±4.5
anti-BzG:*anti*-TTP (alternative hydrogen bonding) [e]	–113.9±13.8	–8.3±6.7	–64.4±4.9	–17.1±5.2
syn-BzG:*anti*-TTP	–102.7±24.6	0.8±2.9	–50.3±7.2	–25.2±5.3
syn-BzG:*anti*-GTP	–117.1±11.5	2.1±1.7	–47.3±4.6	–26.3±4.5
anti-BzG:*syn*-GTP	–115.3±14.9	–24.7±18.4	–61.6±7.8	–11.4±8.3
anti-BzG:*syn*-ATP	–120.7±11.4	–13.4±7.5	–45.7±5.2	–15.5±5.7
syn-BzG:*anti*-ATP	–76.3±48.5	–6.3±4.2	–40.6±7.5	–5.7±5.5

[a] Strength of the hydrogen bond in the base pair 3′ with respect to dG*. [b] Strength of the interaction between dG* and the pairing dNTP. [c] Strength of the stacking interaction between the dG base pair and the base pair 5′ or 3′ with respect to dG*. [d] Data from 20 and 100 ns simulations. [e] See Figures 4 and S9 for a description of the different *anti*-BzG:*anti*-TTP hydrogen-bonding orientations.

Table C.5. Occupancies of the hydrogen bonds between Dpo4 and the incoming dNTP or G* in MD simulations on the Dpo4 ternary complexes for dG or Bz-dG replication. [a,b]

Acceptor	Donor	*anti*-dG:*anti*-dCTP			*anti*-Bz-dG:*anti*-dCTP			*syn*-Bz-dG:*anti*-dCTP			*anti*-Bz-dG:*anti*-dTTP			*syn*-Bz-dG:*anti*-dTTP			*anti*-Bz-dG:*syn*-dGTP			*syn*-Bz-dG:*anti*-dGTP			*anti*-Bz-dG:*syn*-dATP			*syn*-Bz-dG:*anti*-dATP		
		%	Å	Deg.	%	Å	Deg.	%	Å	Deg.	%	Å	Deg.	%	Å	Deg.	%	Å	Deg.	%	Å	Deg.	%	Å	Deg.	%	Å	Deg.
dG*(O3′)	Ser34(Oγ H)	8%	2.8	152.6	NO [b]			12%	3.2	131.4	NO [b]			NO [b]			NO [b]			NO [b]			6%	2.9	149.9	54%	3.1	145.3
dG*(OP2)	Arg331(Nε H)	14%	3.3	135.5	9%	3.2	135.9	6%	3.0	151.8	35%	3.0	156.3	24%	3.2	139.3	28%	3.0	156.0	15%	3.2	138.1	88%	2.9	153.6	12%	3.2	140.9
dG*(OP1)	Arg331(Nη H)	98%	2.8	161.8	92%	2.8	161.1	61%	3.0	154.2	46%	2.9	153.6	100%	2.8	163.0	50%	3.0	152.0	100%	2.8	164.2	NO [b]			76%	2.9	157.2
dG*(OP2)	Arg331(Nη H)	NO [b]			NO [b]			NO [b]			65%	2.8	157.5	NO [b]			29%	2.9	158.1	NO [b]			88%	2.8	162.3	NO [b]		
dG*(OP)	Gly41(NH)	12%	3.0	138.0	34%	3.0	142.5	59%	3.0	143.5	NO [b]			NO [b]			42%	3.0	139.6	NO [b]			NO			63%	3.0	142.3
dG*(OP1)	Ser34(Oγ H)	36%	2.8	162.4	64%	2.8	160.8	NO [b]			30%	2.7	164.1	100%	2.7	164.9	NO [b]			100%	2.7	164.9	8%	2.7	162.9	34%	2.9	157.0
dG*(OP2)	Ser34(Oγ H)	NO [b]			22%	2.9	162.5	76%	2.8	164.6	NO [b]			NO [b]			60%	2.8	164.3	NO [b]			NO [b]			45%	2.9	159.3
dGTP(N7)	Tyr12(OH)	NO [b]			NO [b]			NO [b]			NO [b]			NO [b]			36%	3.0	152.5	NO [b]			NO [b]			NO [b]		
dNTP(O3′)	Tyr12(OH)	NO [b]			NO [b]			NO [b]			NO [b]			83%	3.1	162.8	34%	2.8	161.5	NO [b]			5%	3.1	136.2	NO [b]		
dNTP(O3′)	Tyr12(NH)	79%	3.1	162.9	86%	3.1	162.3	23%	3.2	160.5	81%	3.1	163.6	NO [b]			NO [b]			87%	3.1	165.5	NO [b]			NO [b]		
dNTP(O5′)	Tyr12(OH)	NO [b]			NO [b]			NO [b]			NO [b]			NO [b]			NO [b]			NO [b]			NO [b]			18%	3.2	144.2
dNTP(O5′)	dG 3'(O3')	NO [b]			NO [b]			NO [b]			NO [b]			NO [b]			NO [b]			98%	2.7	161.7	NO [b]			NO [b]		
dNTP(Oα1)	Tyr12(OH)	NO [b]			NO [b]			NO [b]			NO [b]			NO [b]			NO [b]			NO [b]			NO [b]			36%	3.0	158.5
dNTP(Oβ3)	Arg51(N□H)	74%	2.9	159.6	97%	2.9	161.0	82%	2.9	159.1	100%	2.9	161.3	100%	2.9	161.5	NO [b]			90%	3.0	141.4	NO [b]			NO [b]		
dNTP(Oβ)	Lys152(Nζ H)	NO [b]			NO [b]			94%	2.8	156.6	NO [b]			NO [b]			21%	2.9	139.1	NO [b]			NO [b]			NO [b]		
dNTP(Oβ)	Lys159(Nζ H)	NO [b]			NO [b]			NO [b]			NO [b]			NO [b]			34%	3.0	132.4	NO [b]			NO [b]			NO [b]		
dNTP(Oβ2)	Phe11(NH)	67%	3.2	158.9	54%	3.3	158.3	44%	3.2	157.9	58%	3.3	158	59%	3.3	156.8	NO [b]			99%	3.1	160.9	67%	3.1	156.6	70%	3.1	156.3
dNTP(Oβ1)	Thr45(Oγ H)	100%	2.7	162.4	100%	2.7	163.0	100%	2.7	163.5	100%	2.7	163.0	100%	2.7	163.8	NO [b]			10%	3.0	146.7	NO [b]			77%	2.7	165.8
dNTP(Oβ3)	Tyr10(NH)	NO [b]			NO [b]			35%	3.0	126.5	42%	3.1	125.9	46%	3.0	126.1	NO [b]			56%	2.9	129.4	13%	3.2	135.7	NO [b]		
dNTP(Oγ3)	Arg51(Nη H2)	57%	3.0	143.6	51%	3.1	137.0	55%	3.0	143.3	44%	3.2	135.9	NO [b]			NO [b]			NO [b]			NO [b]			30%	3.2	138.6
dNTP(Oγ2)	Arg51(Nη H2)	NO [b]			NO [b]			6%	3.3	134.2	NO [b]			23%	3.2	136.1	52%	2.8	156.4	13%	3.2	138.1	NO [b]			41%	3.2	157.4
dNTP(Oγ3)	Arg51(Nη H1)	84%	2.9	154.7	88%	2.8	159.9	88%	2.9	152.9	81%	2.8	161.1	64%	3.0	150.0	NO [b]			NO [b]			NO [b]			65%	2.9	164.2
dNTP(Oγ2)	Arg51(Nη H1)	51%	3.0	150.3	35%	3.0	148.3	62%	3.1	148.1	38%	3.0	153.3	69%	3.0	158.2	45%	2.9	148.0	97%	2.8	164.2	NO [b]			NO [b]		
dNTP(Oγ1)	Lys159(Nζ H)	37%	2.9	154.3	19%	2.9	152.6	NO [b]			NO [b]			77%	2.8	152.3	51%	2.8	156.0	58%	2.8	136.5	77%	2.8	158.1	NO [b]		
dNTP(Oγ2)	Lys159(Nζ H)	NO [b]			NO [b]			NO [b]			NO [b]			NO [b]			NO [b]			87%	2.9	157.7	NO [b]			NO [b]		
dNTP(Oγ3)	Lys159(Nζ H)	69%	2.8	154.1	75%	2.8	157.5	29%	3.1	138.7	93%	2.8	157.8	41%	3.0	149.6	47%	2.8	161.6	NO [b]			NO [b]			NO [b]		
dNTP(Oγ3)	Tyr10(NH)	100%	2.9	162.4	99%	2.9	167.1	100%	2.9	164.3	99%	2.9	167.6	100%	2.9	167.3	11%	3.1	139.1	49%	3.2	163	NO [b]			62%	3.0	136.4
dNTP(Oγ3)	Tyr48(OH)	24%	2.7	164.4	NO			40%	2.8	159.6	NO			84%	2.6	165.5	26%	2.7	165.8	98%	2.7	167.6	25%	2.7	163.0	NO [b]		
Gly58(O)	Bz-dG(N2H)	NO [b]			NO [b]			NO [b]			NO [b]			NO [b]			34%	3.0	141.7	NO [b]			24%	3.0	143.5	NO [b]		

[a] Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of > 120°. [b] Not observed.

Table C.6. MM/GBSA average pairwise energy contributions of individual residues to dNTP binding (kJ mol^{-1}) during MD simulations on the Dpo4 ternary complexes for dG or Bz-dG replication. [a]

Active Site Base Pair	Mg343	Mg342	Arg51	Lys159	Tyr10	Thr45	Phe11	Ala44	Gln14	Tyr12	Tyr48	Ile104	Lys152	Asp9	Lys78
anti-dG:*anti*-dCTP (20 ns) [b]	–937.1	–435.7	–191.4	–127.7	–58.2	–45.5	–28.8	–18.4	–15.6	–15.3	–14.7	–10.7	–10.0	–8.6	–1.9
anti-dG:*anti*-dCTP (100 ns) [b]	–930.4	–429.7	–178.3	–148.3	–57.7	–45.1	–27.0	–18.2	–15.8	–14.6	–6.3	–10.1	–10.3	–6.6	–1.8
anti-Bz-dG:*anti*-dCTP (20 ns) [b]	–918.7	–430.2	–185.7	–116.0	–57.8	–45.2	–27.1	–17.1	–15.0	–16.7	–15.5	–10.8	–12.1	–10.1	–6.2
anti-Bz-dG:*anti*-dCTP (100 ns) [b]	–907.4	–419.7	–187.3	–121.5	–58.5	–46.2	–27.4	–18.2	–15.0	–18.0	–17.4	–10.3	–13.1	–10.7	–3.3
syn-Bz-dG: *anti*-dCTP	–1056.7	–451.9	–193.6	–174.9	–57.9	–45.0	–24.5	–16.1	–16.0	–21.7	–20.8	–13.5	–9.1	–4.9	–18.3
anti-Bz-dG: *anti*-dTTP	–884.8	–390.0	–181.9	–114.8	–58.0	–44.6	–26.8	–18.5	–14.6	–16.1	–16.4	–9.5	–6.0	–11.3	–1.4
syn-Bz-dG: *anti*-dTTP	–991.4	–412.6	–188.0	–203.0	–59.8	–46.0	–27.9	–18.8	–18.8	–17.0	–43.3	–9.4	–8.9	13.4	–2.2
anti-Bz-dG:*syn*-dGTP	–661.9	–732.8	–58.9	–135.2	–13.7	–5.8	–6.5	–12.2	–3.8	–14.1	–0.9	–5.3	–30.8	–1.1	–6.6
syn-Bz-dG: *anti*-dGTP	–1154.7	–397.3	–119.1	–216.4	–48.5	–16.5	–43.3	–17.7	–14.9	–19.7	–46.7	–10.1	–6.4	13.9	–0.9
anti-Bz-dG:*syn*-dATP	–910.7	–437.6	–17.7	–80.1	–27.8	–10.4	–21.0	–15.4	–8.8	–8.8	–0.6	–4.0	–2.4	–0.5	–1.7
syn-Bz-dG: *anti*-dATP	–999.1	–546.4	–110.2	–13.1	–39.0	–40.7	–23.7	–13.9	–9.3	–30.1	1.1	–14.7	–8.3	–0.8	–11.5
anti-Bz-dG deletion	–896.2	–422.8	–18.9	–19.4	–9.9	–6.6	–4.8	–7.3	–1.9	0.2	0.1	–4.2	–6.8	–0.8	–2.5
syn-Bz-dG deletion	–970.6	–614.5	–26.9	–81.2	–11.6	–9.0	–3.6	–15.0	–2.1	–7.2	0.1	–7.8	–19.0	–7.9	–1.7

[a] Only residues that contributed more than 10 kJ mol^{-1} to dNTP binding are reported. [b] Data from 20 and 100 ns simulations.

Table C.7. Coordination of catalytic Mg^{2+} ions during MD simulations on the ternary complex for dG or Bz-dG replication. [a]

Active Site Base Pair	Catalytic Ion: Mg342						Nucleotide Binding Ion: Mg343					
	Primer (O3‘)	Glu106 (Oε1)	Asp7 (Oδ1)	dNTP (Oα2)	Asp105 (Oδ1)	Wat (O)	Asp7 (Oδ2)	Phe8 (O)	Asp105 (Oδ2)	dNTP (Oα2)	dNTP (Oβ2)	dNTP (Oγ1)
anti-dG:*anti*-dCTP (20 ns) [b]	94%	100%	100%	99%	100%	100%	100%	100%	100%	8%	100%	100%
anti-dG:*anti*-dCTP (100 ns) [b]	98%	100%	100%	100%	100%	100%	100%	100%	100%	5%	100%	100%
anti-Bz-dG:*anti*-dCTP (20 ns) [b]	97%	100%	100%	100%	100%	100%	100%	100%	100%	10%	100%	100%
anti-Bz-dG:*anti*-dCTP (100 ns) [b]	98%	100%	100%	100%	100%	100%	100%	100%	100%	4%	100%	100%
syn-Bz-dG:*anti*-dCTP	93%	100%	100%	92%	100%	100%	100%	100%	100%	62%	100%	100%
anti-BzG:*anti*-TTP (pseudo Watson-Crick) [d]	98%	100%	100%	99%	100%	100%	100%	100%	100%	15%	100%	100%
anti-BzG:*anti*-TTP (alternative hydrogen bonding) [d]	98%	100%	100%	100%	100%	100%	100%	100%	100%	0%	100%	100%
syn-Bz-dG:*anti*-dTTP	100%	100%	100%	97%	100%	100%	100%	100%	100%	33%	100%	100%
anti-Bz-dG:*syn*-dGTP	NO [c]	NO [c]	100%	NO [c]	100%	42%	100%	100%	100%	NO [c]	NO [c]	100%
syn-Bz-dG:*anti*-dGTP	99%	100%	100%	97%	100%	100%	100%	100%	100%	97%	100%	100%
anti-Bz-dG:*syn*-dATP	1%	100%	100%	NO [c]	100%	100%	100%	100%	100%	NO [c]	100%	NO [c]
syn-Bz-dG:*anti*-dATP	100%	100%	100%	NO [c]	100%	100%	NO [c]	100%	100%	NO [c]	100%	100%
anti-Bz-dG Deletion	NO [c]	NO [c]	76%	NO [c]	NO [c]	NO [c]	NO [c]	NO [c]	NO [c]	100%	100%	NO [c]
syn-Bz-dG Deletion	NO [c]	NO [c]	100%	NO [c]	NO [c]	NO [c]	NO [c]	NO [c]	NO [c]	100%	100%	NO [c]

[a] Percentage of the simulation that the distance between the Mg^{2+} ion and the specified atom is < 2.5 Å. [b] Data from 20 and 100 ns simulations. [c] Not observed. [d] See Figures 4 and S9 for a description of the different *anti*-BzG:*anti*-TTP hydrogen-bonding orientations.

Table C.8. Structural parameters from MD simulations on the Dpo4 –1 base deletion complex for Bz-dG replication.

G*	rmsd	G* Pucker[a]	G* χ[b]	G* θ[b]	G* φ[b]	G* ξ[b]	5′ Base Pair C1′–C1′ Distance [c]	3′ Base Pair C1′–C1′ Distance [c]	Reaction Distance [d]	Reaction Angle [e]
anti Bz-dG	1.911± 0.273 Å	C4′-*exo* (53.2%)	226.2± 8.8°	178.5± 132.2°	159.5± 43.7°	176.9± 88.6°	10.951±0 .222 Å	10.831± 0.178 Å	5.630± 0.492 Å	84.3± 15.8°
syn Bz-dG	1.381± 0.254 Å	O4′-*endo* (50.9%)	58.1± 19.2°	182.3± 94.3°	223.7± 49.1°	204.1± 95.6°	10.874±0 .195 Å	10.685± 0.195 Å	4.801± 1.019 Å	103.1± 16.2°

[a] Most common sugar pucker of Bz-dG and the percentage of the simulation that the sugar pucker is adopted. [b] Bz-dG dihedral angle (degrees). See Figure 1 for definitions of adduct dihedral angles. [c] Width of the base pair 5′ or 3′ with respect to Bz-dG. [d] Distance between O3′(primer 3′ end) and Pα(dNTP). [e] ∠(O3′(primer 3′ end)–Pα(dNTP)–Oαβ(dNTP)).

Table C.9. Strength (kJ mol^{-1}) of stacking and hydrogen-bonding interactions involving Bz-dG and the flanking base pairs in MD simulations on the Dpo4 –1 base deletion complex for Bz-dG replication.

G*	3′ Base Pair Hydrogen Bond [a]	5′ Base Pair Hydrogen Bond [a]	5′ Stacking [b]	3′ Stacking [b]
anti-Bz-dG	–110.0±8.3	–116.5±8.2	–15.4±8.2	–41.0±13.9
syn-Bz-dG	–111.2±6.9	–113.4±11.1	–14.5±9.2	–39.8±14.3

[a] Strength of the hydrogen bond in the base pair 5′ or 3′ with respect to Bz-dG. [b] Strength of the stacking interaction between Bz-dG and the 5′ or 3′ inter and intrastand bases with respect to Bz-dG

Table C.10. Occupancies for the hydrogen bonds between Dpo4 and the incoming dNTP or the dC being replicated in MD simulations on the Dpo4 deletion complexes for Bz-dG replication. [a]

Acceptor	Donor	*anti*-Bz-dG			*syn*-Bz-dG		
		%	Å	Deg.	%	Å	Deg.
dC354(OP)	Ser34(OγH)	79%	2.9	159.3	55%	2.9	162.8
dC354(OP)	Ser40(OγH)	NO [b]			7%	2.7	158.2
dC354(OP)	Arg331(NηH)	70%	3.0	153.4	86%	2.9	160.7
dC354(OP)	Arg36(NηH)	27%	2.9	156.5	NO [b]		
Ala44(O)	dGTP(O3')	NO [b]			9%	2.9	156.7
dGTP(O3′)	Thr45(OγH)	11%	2.8	160.5	8%	3.0	161.3
dGTP(Oγ)	Tyr10(NH)	10%	3.0	145.2	NO [b]		
dGTP(Oγ)	Arg51(NηH)	16%	2.9	151.2	16%	2.8	164.0
dGTP(Oγ)	Lys159(NζH)	NO [b]			74%	2.8	157.0
dGTP(Oγ)	Tyr10(NH)	NO [b]			16%	3.0	157.9

[a] Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of > 120°. [b] Not observed.

Appendix D

Supplementary Information for Chapter 5: Computational Insights into the Mutagenicity of Two Tobacco Derived Carcinogenic DNA Lesions

Contains full computational details, Tables D.1–D.11, and Figures D.1–D.25.

Full Computational Details

DFT Calculations

Nucleobase model: Due to the anticipated high degree of flexibility within POB-G and PHB-G, the inherent conformational preference about the nucleobase–carcinogen linker and within the bulky moiety was initially examined using a nucleobase model. Specifically, a usage directed conformational search about the α′, β′,γ′, δ′, ϵ′, ρ′, and ζ′ dihedral angles within the bulky moiety (Figure 5.1) was completed for each lesion as implemented in Hyperchem (1), with a maximum of 100,000 iterations or 1000 optimizations. Unique structures were defined based on an energy difference of > 0.2 kJ/mol and a heavy atom root-mean-square deviation (rmsd) of 0.25 Å. During the search, each adduct was modeled with AMBER99 charges and atom types (Table D.13, Appendix D) and structures within 40 kJ/mol of the most stable structure were saved. All orientations of POB-G and PHB-G isolated from the conformational search were subsequently optimized using B3LYP-D3(BJ)/6-31G(d) and the relative energies were determined using B3LYP-D3(BJ)/6-311+G(2df,2p). Each unique structure (based on an energy difference of > 0.2 kJ/mol and difference of > 0.2° in the key bulky moiety dihedral angles) was considered for further analysis, including the classification of the lesion orientation based on the position of the bulky moiety relative to the adducted G. Specifically, all structures were visually inspected and classified based on discrete interactions between the bulky moiety and adducted G as stacked, hydrogen bonded, T-shaped or extended (no direct interaction between the bulky moiety and adducted G).

Nucleoside model: To assess the relative stability of the *anti* ($\chi \approx 220°$) and *syn* ($\chi \approx 60°$) conformations about the glycosidic bond in the nucleoside adducts, 2′-deoxyribose was

added to the most stable orientation for each conformational category of POB-G and PHB-G identified using the nucleobase models. In the nucleoside adducts, the 2′-*endo* sugar pucker was used, and the O3′ hydroxy groups was oriented and the O5′ hydroxy groups was fixed to reflect the geometry of a non-terminal nucleoside (i.e., ∠(HC3′O3′H) ≈ −60° and ∠(C4′C5′O5′H) = 180°). Each nucleoside model was optimized in the *anti* and *syn* orientations using B3LYP-D3(BJ)/6-31G(d) and the relative energies were determined using B3LYP-D3(BJ)/6-311+G(2df,2p).

Hydrogen-Bonding Interactions: Hydrogen-bonded pairs between the Watson-Crick or Hoogsteen face of the most stable orientation of POB-G or PHB-G from each conformational category, and each of the four canonical DNA bases were optimized using M06-2X/6-31G(d). This method was used since previous work has shown that B3LYP-D3(BJ)/6-31G(d) optimizations of G mispairs give rise to a single, small imaginary frequency, while stable minima with nearly identical structures were identified when the same hydrogen-bonded complexes were optimized with M06-2X/6-31G(d) (2). Furthermore, no significant structural or energetic differences were observed between the most stable nucleobase and nucleoside conformations optimized with B3LYP-D3(BJ)/6-31G(d) and M06-2X/6-31G(d) (2). In the computational model, 2′-deoxyribose was replaced with a methyl group since the nucleoside model indicates that the sugar does not affect the conformation of the lesion. Furthermore, the methyl groups allow the base pair width (C1′–C1′ distance) and opening angle (∠(N9C1′C1′)) within the adducted base pairs to be assessed. The relative energies were determined using B3LYP-D3(BJ)/6-311+G(2df,2p).

All DFT calculations were performed using Gaussian 09 (revision D.01) (3).

Molecular Dynamics Simulations

Parameters: All natural amino acids, nucleotides, and the solvent were modeled with AMBER ff14SB parameters (4). Parameters for POB-G and PHB-G were assigned according to the GAFF (5) and AMBER ff14SB force fields using ANTECHAMBER 1.4 (6), and partial charges for the lesions were developed using RESP charge fitting from a HF/6-31G(d) calculation by the R.E.D.v.III.4 program (7,8) (Table D.11, Appendix D). The parameters for the dNTPs were adapted from the literature (9-11), Mg^{2+} was modeled with the parameters from Allner *et al.* (12), and Na^{+} and Cl^{-} were modeled with the monovalent ion parameters from Joung & Cheatham (13).

DNA model: MD simulations were performed on the 5′–CTCGGCG*CCATC 12-mer DNA helix. The initial choices of the key dihedral angles in the bulky moiety (α′, β′,γ′, δ′, ϵ′, ρ′, ζ′, and χ; Figure 5.1) and the lesion site base-pairing geometry were directed by the DFT calculations. Specifically, MD simulations were initially conducted with the lesion paired opposite C in a wobble base-pairing arrangement or with the lesion intercalated and the pairing C in the major groove. POB-G or PHB-G were positioned at G* in an orientation that represents each of the conformational categories identified in the nucleobase conformational search. Since the bulky moiety of the lesion resides in the major groove regardless of the orientation about the glycosidic bond, the conclusions about the preferred lesion orientation can be applied to POB-G and PHB-G regardless of the glycosidic orientation. Therefore, MD simulations were subsequently performed with the lesion

mispaired opposite T, A, or G, with the bulky moiety initiated in the preferred orientation identified from the simulations on DNA containing the lesions paired opposite C.

Initial DNA helices were prepared for simulation by neutralizing the system with 24 Na^+ ions, and solvating the DNA in a TIP3P octahedral water box such that the DNA was at least 8.0 Å from the box edge. The systems were minimized in a stepwise fashion. In the first minimization step, a 500 kcal mol^{-1} $Å^{-2}$ restraint was placed on the DNA, and 500 steps of steepest descent minimization and 500 steps of conjugate gradient minimization were performed. Subsequently, 1000 steps of unconstrained steepest descent minimization and 1500 steps of unconstrained conjugate gradient minimization were performed. The systems were then heated to 300 K over 20 ps using the Langevin thermostat (γ=1.0) with a 10 kcal mol^{-1} $Å^{-2}$ restraint on the DNA. Finally, each system was simulated for 100 ns at 300 K and 1 bar using the Langevin thermostat (γ=1.0), a Monte Carlo Barostat, and a time step of 2 fs. For all calculations, a non-bonded cutoff of 8.0 Å, the periodic boundary condition, and SHAKE were implemented. The simulations were run using the pmemd module of AMBER 14. For all duplexes, 20 ns pre-production simulations were performed to understand the inherent conformational flexibility of the lesion. From these trial simulations, representative structures for each unique lesion conformation were chosen as starting points for further simulations. The final 300 ns production MD simulation and two additional 100 ns MD simulations were run using different starting velocities to ensure the results are statistically significant. Due to negligible all-atom rmsds between the replicas (Table D.1, Appendix D), data from one replica will be discussed throughout the main text.

Polymerase η model: Initial structures for simulations on the polymerase η insertion complexes were obtained from a crystal structure of the reactant complex for the insertion of dATP opposite dT (PDB ID: 4ECS). This crystal structure was chosen due to the resolution, presence of the two ions in the active site, and lack of mutation to the dNTP or 3′ terminus. Missing residues (Thr155–Glu159) were added by hand using GaussView. Glycerol, Ca^{2+}, and pyrophosphate were removed from the system. In cases where multiple orientations of amino acids were present in the crystal structure, the orientation that was best aligned to interact with the surrounding residues was chosen. In cases where the amino acid was not positioned to interact with any surrounding residues, the higher occupied crystal structure orientation was chosen. To generate the insertion complexes, dCTP, dTTP, or dATP was paired opposite POB-dG or PHB-dG in the active site based on the base pair hydrogen-bonding patterns observed in the DNA helix. Adiditonally, the insertion complex for dCTP insertion opposite G was modeled as a control. The resulting 7 complexes were prepared for simulation using the tleap module of AMBER14. Specifically, hydrogen atoms were added to generate the native protonation states of all DNA and protein residues. Furthermore, NaCl was added to the water box to neutralize the system and to yield a final concentration of ~0.150 M (67 Na^+ ions and 56 Cl^- ions), and the system was solvated in a TIP3P octahedral water box such that the DNA–polymerase complex was at least 10.0 Å from the edge of the box.

The systems were minimized in a stepwise fashion. The first step minimized the water for 2500 steps of steepest descent and 2500 steps of conjugte gradient minimization, with a 50 kcal mol^{-1} $Å^{-2}$ on the rest of the system. Next, the hydrogen atoms were minimized for 4000 steps of steepest descent and 4000 steps of conjugant gradient minimization, with a

50 kcal mol^{-1} $Å^{-2}$ on the heavy atoms. Subsequently, the systems were minimized using a 50 kcal mol^{-1} $Å^{-2}$ restraint on the amino acid backbone for 12500 steps of steepest descent and 12500 steps of conjugate gradient minimization. Finally, the entire system was minimized without restraint for 5000 steps of steepest descent and 5000 steps of conjugant gradient minimization. The systems were then heated from 10 to 310 K in 6 steps, each increasing the temperature by 50 K over 10 ps using a 1 fs time step. Heating was performed with a 10 kcal mol^{-1} $Å^{-2}$ restraint on the solute using the Langevin thermostat (γ=1.0). The systems were then equilibrated over 5 steps, each of 20 ps, using a 2 fs time step, and a constraint of 20, 15, 10, 5, or 1 kcal mol^{-1} $Å^{-2}$ on the solute. Finally, a 100 ns unconstrained production simulation was performed on each system. MD simulations were run in triplicate using different initial velocities to ensure the results were statistically significantly. Due to negligible all-atom rmsds between each replica (Table D.1, Appendix D), data from one replica will be discussed throughout the main text.

Analysis: Analysis of the MD simulations was performed every 0.1 ns using the cpptraj module of Amber 14 (14). Additionally, 3DNA was used to understand the overall structure of the DNA helix (15). A representative structure for each MD simulation were obtained by clustering the simulations based on the rmsd of damaged base pair using the average linkage algorithm. Furthermore, the strength of the lesion hydrogen bonding with the pairing base in the DNA and polymerase models was calculated over 100 frames (i.e., every 3 ns for DNA models or 1 ns for polymerase models) using B3LYP-D3(BJ)/6-311+G(2df,2p) based on the geometry of the base pair in the representative structure, with hydrogen-capped nucleobases. When a hydrogen bond periodically exists between the

bulky moiety and opposing base, the binding strength was separated into two clusters based on the presence or absence of the additional interaction.

References

1. HyperChem. Professional 7.5 ed. Hypercube, Inc., Gainesville, Florida 32601, USA.

2. Wilson, K.A., Szemethy, K.G. and Wetmore, S.D. (2017) Conformational flexibility and base-pairing tendency of the tobacco carcinogen o6-[4-oxo-4-(3-pyridyl)butyl]guanine. *Biophysical Chemistry*, **228**, 25-37.

3. 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., Jr., J.A.M., Peralta, J.E., Ogliaro, F., Bearpark, M., Heyd, J.J., Brothers, E., Kudin, K.N., Staroverov, V.N., Keith, T., Kobayashi, R., Normand, J., Raghavchari, 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, O., Foresman, J.B., Ortiz, J.V., Cioslowski, J. and Fox, D.J. (2013). Revision D.01 ed. Gaussian, Inc., Wallingford CT.

4. Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E. and Simmerling, C. (2015) Ff14sb: Improving the accuracy of protein side chain and backbone parameters from ff99sb. *J Chem Theory Comput*, **11**, 3696-3713.

5. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W. and Kollman, P.A. (1995) A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.

6. Wang, J., Wang, W., Kollman, P.A. and Case, D.A. (2006) Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.*, **25**, 247–260.

7. 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.

8. Vanquelef, E., Simon, S., Marquant, G., Garcia, E., Klimerak, G., Delepine, J.C., Cieplak, P. and Dupradeau, F.-Y. (2011) R.E.D. Server: A web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res.*, **39**, W511-W517.

9. Zhang, L., Rechkoblit, O., Wang, L., Patel, D.J., Shapiro, R. and Broyde, S. (2006) 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.*, **34**, 3326-3337.

10. Perlow, R.A. and Broyde, S. (2002) Toward understanding the mutagenicity of an environmental carcinogen: Structural insights into nucleotide incorporation preferences. *J. Mol. Biol.*, **322**, 291-309.

11. Zhang, L., Shapiro, R. and Broyde, S. (2005) Molecular dynamics of a food carcinogen−DNA adduct in a replicative DNA polymerase suggest hindered nucleotide incorporation and extension. *Chem. Res. Toxicol.*, **18**, 1347-1363.

12. Allnér, O., Nilsson, L. and Villa, A. (2012) Magnesium ion–water coordination and exchange in biomolecular simulations. *J. Chem. Theory Comput.*, **8**, 1493-1502.

13. Joung, I.S. and Cheatham, T.E. (2008) Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B*, **112**, 9020-9041.

14. Case, D.A., Babin, V., Berryman, J.T., Betz, R.M., Cai, Q., Cerutti, D.S., T.E. Cheatham, I., 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). University of California, San Francisco.

15. Lu, X.-J. and Olson, W.K. (2008) 3dna: A versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protocols*, **3**, 1213-1227.

Table D.1. Heavy atom rmsd (Å) with respect to the representative structure for Trial 1.

	Trial 1[a]		Trial 2[b]		Trial 3[b]	
	Lesion Site	Entire system	Lesion Site	Entire system	Lesion Site	Entire system
DNA POB-G:C	0.875±0.178	1.887±0.399	0.879±0.184	1.889±0.391	0.875±0.191	1.890±0.426
DNA PHB-G:C	1.331±0.209	2.249±0.456	1.374±0.223	2.180±0.411	1.308±0.225	2.200±0.429
DNA POB-G:T	1.095±0.191	2.327±0.453	1.086±0.210	2.340±0.466	1.126±0.199	2.393±0.453
DNA PHB-G:T	1.581±0.249	2.468±0.550	1.157±0.250	2.501±0.571	1.569±0.239	2.878±0.611
DNA POB-G:A	1.442±0.727	2.175±0.508	1.859±0.344	0.766±0.164	0.791±0.196	1.928±0.372
DNA PHB-G:A	0.877±0.218	1.994±0.438	0.847±0.171	1.964±0.396	0.852±0.174	1.957±0.397
DNA POB-G:G	1.300±0.492	2.850±1.147	1.760±0.316	3.827±0.994	1.845±0.273	4.667±0.563
DNA PHB-G:G	2.244±0.208	5.772±0.805	2.231±0.197	6.052±0.697	1.620±0.229	2.621±0.871
Pol η POB-G:C	0.862±0.120	2.373±0.449	0.784±0.096	2.217±0.211	0.809±0.102	2.290±0.221
Pol η PHB-G:C	0.747±0.131	2.010±0.209	1.411±0.415	2.256±0.182	0.941±0.14	2.330±0.230
Pol η POB-G:T	1.059±0.182	2.090±0.268	0.987±0.230	2.107±0.159	1.120±0.180	2.356±0.121
Pol η PHB-G:T	1.202±0.264	2.167±0.267	1.646±0.100	2.472±0.250	1.704± 0.097	2.522±0.184
Pol η POB-G:A	1.157±0.384	2.065±0.330	2.385±0.341	1.044±0.256	0.947±0.117	2.228±0.166
Pol η PHB-G:A	1.468±0.402	2.077±0.237	2.200±0.116	2.331±0.165	2.197±0.228	2.283±0.130

[a]300 ns MD simulations on adducted DNA and 100 ns on polymerase η insertion complex. [b]100 ns MD simulations.

Table D.2. B3LYP-D3(BJ)/6-311+G(2df,2p)//M06-2X/6-31G(d) interaction energies (kJ/mol) for base pairs between the Watson-Crick or Hoogsteen hydrogen-bonding face of G, or various POB-G or PHB-G conformers (G*), and the canonical nucleobases.[a]

		Fully Extended		Stacked		Hydrogen Bonded		T-shaped		Extended	
	G	POB-G	PHB-G	POB-G	PHB-G	POB-G	PHB-G	POB-G	PHB-G	POB-G	PHB-G
Watson-Crick G*:C	–138.0	–80.5	–72.5	–79.9	–75.8	N/A[b]	N/A[b]	N/A[b]	N/A[b]	–64.4	–57.7
Watson-Crick G*:T	–80.4	–69.7	–55.6	–68.9	–75.4	N/A[b]	N/A[b]	N/A[b]	N/A[b]	–56.3	–52.8
Watson-Crick G*:A	–82.5	–64.3	–59.8	–64.7	–65.9	N/A[b]	N/A[b]	N/A[b]	N/A[b]	–50.6	–48.2
Watson-Crick G*:G	–69.3	–38.1	–29.8	–48.4	–34.6	N/A[b]	N/A[b]	N/A[b]	N/A[b]	–43.1	–36.4
Hoogsteen G*:C	–59.0	–54.6	–53.1	–59.3	–63.3	–59.3	–53.1	–58.4	–60.0	N/A[b]	N/A[b]
Hoogsteen G*:T	–42.4	–52.4	–50.6	–59.5	–69.9	–51.5	–46.9	–50.9	–71.1	N/A[b]	N/A[b]
Hoogsteen G*:A	–44.4	-42.7	–43.6	–43.2	–51.7	–43.3	–38.9	–41.4	–58.6	N/A[b]	N/A[b]
Hoogsteen G*:G	–69.3	–73.8	–93.9	–76.5	–71.9	–78.7	–68.6	–84.2	–78.0	N/A[b]	N/A[b]

[a]See Figures 5.3 and D.5–D.12. [b]Not applicable due to bulky moiety interactions with the respective hydrogen-bonding face.

Table D.3. Occupancies (%), average heavy atom distances (Å), and average angles (deg.) for hydrogen bonds in the adducted, and 3′ and 5′-flanking base pairs across the entire MD simulation trajectory for POB-G or PHB-G adducted DNA with the lesion paired opposite C.

	POB-G:C			PHB-G:C		
	Occupancy	Average Distance	Average Angle	Occupancy	Average Distance	Average Angle
5′-C(O2)···GN2H)	100%	2.882	163.7	100%	2.877	163.0
5′-C(N3)···G(N1)	100%	2.962	165.5	100%	2.963	164.8
5′-C(N4H)···G(O6)	99%	2.924	163.7	99%	2.934	162.2
G*(N1)···C(N4H)	86%	3.120	159.4	96%	3.020	163.4
G*(N2H)···C(N3)	91%	3.061	159.7	92%	3.035	160.8
G*(O12)···C(N4H)	29%	2.961	158.0	7%	2.977	159.0
3′-C(O2)···G(N2H)	100%	2.861	164.3	100%	2.882	163.1
3′-C(N3)···G(N1)	100%	2.952	164.3	100%	2.950	163.9
3′-C(N4H)···G(O6)	99%	2.937	164.3	99%	2.933	163.8

[a]Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of < 120°.

Table D.4. Occupancies (%), average heavy atom distances (Å), and average angles (deg.) for hydrogen bonds in the adducted, and 3′ and 5′-flanking base pairs across the entire MD simulation trajectory for POB-G or PHB-G adducted DNA with the lesion paired opposite T.

	POB-G:T			PHB-G:T		
	Occupancy	Average Distance	Average Angle	Occupancy	Average Distance	Average Angle
5′-C(O2)···G(N2H)	100%	2.878	163.9	100%	2.873	163.8
5′-C(N3)···G(N1)	100%	2.955	165.3	100%	2.955	165.1
5′-C(N4H)···G(O6)	99%	2.917	163.6	99%	2.919	162.9
G*(N1)···T(N3H)	32%	3.208	154.7	63%	3.136	156.8
G*(N2H)···T(O2)	94%	2.924	158.1	93%	2.892	158.1
G*(O12)···5′-C(N4H)	11%	2.988	154.4	6%	3.068	155.7
3′-C(O2)···G(N2H)	100%	2.861	163.7	100%	2.869	163.2
3′-C(N3)···G(N1)	100%	2.954	165.1	100%	2.954	164.9
3′-C(N4H)···G(O6)	98%	2.953	163.6	98%	2.946	163.4

[a]Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of < 120°.

Table D.5. Occupancies (%), average heavy atom distances (Å), and average angles (deg.) for hydrogen bonds in the adducted, and 3′ and 5′-flanking base pairs across the entire MD simulation trajectory for POB-G or PHB-G adducted DNA with the lesion paired opposite A.

	POB-G:A			PHB-G:A		
	Occupancy	Average Distance	Average Angle	Occupancy	Average Distance	Average Angle
5′-C(O2)···G(N2H)	100%	2.889	163.4	100%	2.885	163.4
5′-C(N3)···G(N1)	100%	2.964	165.0	100%	2.966	165.2
5′-C(N4H)···G(O6)	99%	2.928	162.7	99%	2.929	162.4
G*(N1)···A(N6H)	51%	3.143	159.8	96%	3.029	161.4
G*(N2H)···A(N7)	37%	3.040	160.2	95%	3.025	158.5
3′-C(O2)···G(N2H)	100%	2.876	163.6	100%	2.860	164.0
3′-C(N3)···G(N1)	100%	2.958	164.2	100%	2.957	164.7
3′-C(N4H)···G(O6)	97%	2.951	163.4	97%	2.962	163.6

[a]Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of < 120°.

Table D.6. Occupancies (%), average heavy atom distances (Å), and average angles (deg.) for hydrogen bonds in the adducted, and 3′ and 5′-flanking base pairs across the entire MD simulation trajectory for POB-G or PHB-G adducted DNA with the lesion paired opposite G.

	POB-G:G			PHB-G:G		
	Occupancy	Average Distance	Average Angle	Occupancy	Average Distance	Average Angle
5′-C(O2)···G(N2H)	100%	2.899	165.2	100%	2.883	164.2
5′-C(N3)···G(N1)	100%	2.965	165.7	100%	2.959	165.0
5′-C(N4H)···G(O6)	99%	2.923	164.5	99%	2.930	163.3
G*(N7)···G(N2H)	92%	2.997	154.9	90%	2.967	156.7
G*(N7)···G(N1H)	10%	3.208	141.7	NO[b]		
G*(O6)···G(N1H)	4%	3.191	151.9	48%	3.189	150.1
3′-C(O2)···G(N2H)	100%	2.854	163.9	100%	2.884	162.9
3′-C(N3)···G(N1)	100%	2.954	164.6	100%	2.960	165.3
3′-C(N4H)···G(O6)	96%	2.974	162.6	98%	2.940	162.8

[a]Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of < 120°. [b]Not Observed.

Table D.7. Occupancies (%), average heavy atom distances (Å), and average angles (deg.) for the hydrogen bonds between the template G, POB-G, PHB-G or opposing dCTP and the surrounding DNA:polymerase η complex across the entire MD simulation on the polymerase η insertion complex.

	G:dCTP			POB-G:dCTP			PHB-G:dCTP		
	Occupancy	Average Distance	Average Angle	Occupancy	Average Distance	Average Angle	Occupancy	Average Distance	Average Angle
3′-T(O4)···A(N6H)	95%	2.989	156.8	95%	2.99	157	82%	2.972	164.4
3′-T(N3H)···A(N1)	100%	2.937	163.3	100%	2.934	159.2	86%	2.946	164.1
dCTP(O2)···G(N2H)	100%	2.839	159.8	NO[b]			NO[b]		
dCTP(N3)···G(N1H)	100%	2.933	160.3	NO[b]			NO[b]		
dCTP(N4H)···G(O6)	100%	2.937	163.3	NO[b]			NO[b]		
dCTP(N3)···G*(N2H)	NO[b]			97%	2.999	158.1	77%	3.025	162.3
dCTP(N4H)···G*(N1)	NO[b]			53%	3.182	141.8	39%	3.117	159.7
dCTP(O3')···Phe18(NH)	95%	3.077	161.1	94%	3.058	163.1	83%	3.053	164.4
dCTP(Oα1)···Arg61(NH1H)	76%	2.83	157.5	90%	2.813	159.5	71%	2.802	161.3
dCTP(Oβ2)···Phe17(NH)	84%	3.193	157.6	83%	3.193	159.1	55%	3.227	164.1
dCTP(Oβ1)···Wat1(OH)	99%	2.7	162.6	50%	2.684	162.4	91%	2.677	165
Wat1(O)···Val50(NH)	100%	2.925	164	50%	2.956	163.8	100%	2.927	163.1
dCTP(Oβ1)···Wat2(OH)	67%	2.862	157.2	34%	2.907	154.8	52%	2.865	162
Wat2(O)···Arg61(NH1H)	69%	2.881	161.2	43%	2.892	159.9	82%	2.891	161.5
dCTP(Oβγ)···Arg55(NH2H)	96%	2.942	144.4	97%	2.911	149.6	56%	2.883	161.5
dCTP(Oγ3)···Tyr52(OH)	97%	2.586	165.5	90%	2.588	165.7	98%	2.572	167.5
dCTP(Oγ3)···Lys231(NZH)	61%	2.917	144.5	57%	2.923	144.1	42%	2.988	163
dCTP(Oγ1)···Lys231(NZH)	30%	3.074	147.1	22%	3.006	153.7	13%	2.895	160.4
dCTP(Oγ3)···Cys16(NH)	100%	2.92	166.8	98%	2.95	165.8	95%	2.941	166
dCTP(Oγ3)···Arg55(NH2H)	90%	2.939	149.8	78%	2.993	146.4	22%	2.94	158.2
dCTP(Oγ3)···Arg55(NH1H)	74%	3.041	143	68%	3.007	145.6	13%	2.94	156.9
dCTP(Oγ2)···Arg55(NH1H)	67%	3.077	152.3	67%	3.007	154.9	60%	2.912	164.3

[a]Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of < 120°. [b]Not Observed.

Table D.8. Occupancies (%), average heavy atom distances (Å), and average angles (deg.) for the hydrogen bonds between the lesion or opposing dTTP and the surrounding DNA:polymerase η complex across the entire MD simulation on the polymerase η insertion complex.

	POB-G:dTTP			PHB-G:dTTP		
	Occupancy	Average Distance	Average Angle	Occupancy	Average Distance	Average Angle
3′-T(O4)···A(N6H)	97%	2.997	156.8	96%	2.994	159.5
3′-T(N3H)···A(N1)	100%	2.942	161.3	100%	2.936	160.4
dTTP(O2)···G*(N2H)	93%	2.939	157.2	96%	2.933	154.1
dTTP(N3H)···G*(N1)	50%	3.187	141.5	41%	3.162	144.6
dTTP(O4)···G*(O13H)	NO[b]			24%	2.751	162.6
dTTP(O3')···Phe18(NH)	90%	3.089	163.4	59%	3.123	162.8
dTTP(O4)···Arg61(NH1H)	28%	2.828	157.2	NO[b]		
dTTP(Oβ1)···Arg61(NH1H)	NO[b]			57%	2.936	159
dTTP(Oβ1)···Arg61(NH2H)	NO[b]			36%	3.092	142.7
dTTP(Oβ2)···Phe17(NH)	92%	3.154	159.9	73%	3.155	159.9
Wat2(O)···Val50(NH)	50%	2.982	163.1	50%	2.94	164
dTTP(Oβ1)···Wat2(OH)	51%	2.71	162.2	71%	2.743	160.2
dTTP(Oβ3)···Wat1(OH)	41%	2.695	162.2	NO[b]		
Wat1(O)···Val50(NH)	40%	2.955	163.2	NO[b]		
Wat2(O)···Arg55(NH2H)	NO[b]			23%	3.004	133.4
dTTP(Oβγ)···Arg55(NH2H)	99%	2.913	150.5	94%	2.948	146.6
dTTP(Oγ3)···Tyr52(OH)	76%	2.597	165	29%	2.584	164.3
dTTP(Oγ3)···Lys231(NZH)	NO[b]			52%	2.84	155.5
dTTP(Oγ3)···Cys16(NH)	95%	2.98	165.5	98%	2.958	165.5
dTTP(Oγ2)···Arg55(NH1H)	59%	3.002	154	51%	3.082	147.2
dTTP(Oγ3)···Arg55(NH1H)	65%	3.003	147.3	84%	2.98	148.9
dTTP(Oγ3)···Arg55(NH2H)	62%	2.988	149.5	77%	2.956	149.1

[a]Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of < 120°. [b]Not Observed.

Table D.9. Occupancies (%), average heavy atom distances (Å), and average angles (deg.) for the hydrogen bonds between the lesion or opposing dATP and the surrounding DNA:polymerase η complex across the entire MD simulation on the polymerase η insertion complex.

	POB-G:ATP			PHB-G:ATP		
	Occupancy	Average Distance	Average Angle	Occupancy	Average Distance	Average Angle
3′-T(O4)···A(N6H)	93%	3.008	158.1	89%	3.016	155.5
3′-T(N3H)···A(N1)	100%	2.645	165.1	100%	2.933	161.6
dATP(N6H')···G*(N1)	29%	3.200	148.0	17%	3.156	147.3
dATP(N7)···G*(N2H)	55%	2.993	161.8	19%	3.020	154.6
dATP(N6H')···G*(N3)	12%	3.092	146.1	18%	3.110	148.3
dATP(O3')···Phe18(NH)	81%	3.143	167.5	74%	3.171	167.2
dATP(N1)···Arg61(NH1H)	18%	2.982	156.4	12%	2.892	158.8
dATP(Oβ2)···Phe17(NH)	90%	3.165	161.7	91%	3.142	161.1
dATP(Oα1)···Wat1(OH)	40%	2.808	143.8	NO[b]		
Wat1(O)···Arg55(NH2H)	17%	2.955	131.7	NO[b]		
Wat2(O)···Arg55(NH2H)	NO[b]			23%	2.974	131.9
dATP(Oβ1)···Wat2(OH)	15%	2.681	164.1	70%	2.682	163.2
Wat2(O)···Val50(NH)	23%	2.969	162.3	69%	2.976	163.1
dATP(Oβ1)···Wat3(OH)	41%	2.664	162.7	36%	2.82	142.9
Wat3(O)···Val50(NH)	40%	2.967	163.3	NO[b]		
dATP(Oβγ)···Arg55(NH1H)	98%	2.949	151.2	93%	2.968	147.1
dATP(Oγ3)···Tyr52(OH)	91%	2.609	167.3	74%	2.613	165.7
dATP(Oγ3)···Lys231(NZH)	62%	3.056	155.6	41%	2.987	155.7
dATP(Oγ1)···Lys231(NZH)	67%	2.898	140.4	56%	2.911	143.4
dATP(Oγ3)···Cys16(NH)	99%	2.941	164	99%	2.917	165.4
dATP(Oγ3)···Arg55(NH1H)	38%	3.213	139.2	68%	2.997	150.2
dATP(Oγ3)···Arg55(NH2H)	38%	3.045	148.8	54%	3.069	143.1
dATP(Oγ2)···Arg55(NH2H)	91%	2.906	159.8	84%	2.972	153.3

[a]Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of < 120°. [b]Not Observed.

Table D.10. Coordination of the active site Mg^{2+} ions across the entire MD simulation trajectory for the polymerase η insertion complex corresponding to various lesion replication outcomes.[a]

	Mg^{2+} (binding)	Mg^{2+} (catalytic)
G:dCTP	6/6	6/6
POB-G:dCTP	5/6 (Oα2, 56%)[b]	6/6
PHB-G:dCTP	5/6 (Oα2, 82%)[b]	6/6
POB-G:dTTP	5/6 (Oα2, 67%)[b]	6/6
PHB-G:dTTP	5/6 (Oα2, 51%)[b]	6/6
POB-G:dATP	6/6	6/6
PHB-G:dATP	5/6 (Oα2, 85%)[b]	6/6

[a]Coordination is considered to be present if the distance is < 2.5 Å for > 95% of the simulation. [b]Atom that does not meet coordination criteria and percentage of simulation it meets the coordination criteria is given in brackets.

Table D.11. Atom types and charges for POB-G and PHB-G.

Atom Name	Atom Type	POB-G Charge	PHB-G Charge
P	P	1.1617	1.1624
O5'	OS	-0.4974	-0.4949
OP1	O2	-0.7662	-0.7629
OP2	O2	-0.7662	-0.7698
O3'	OS	-0.5408	-0.5444
C5'	CT	-0.0247	0.001
H5'1	H1	0.0806	0.072
H5'2	H1	0.0806	0.072
C4'	CT	0.148	0.1731
H4'	H1	0.0874	0.1016
O4'	OS	-0.3754	-0.4053
C1'	CT	0.1532	0.2091
H1'	H2	0.0809	0.0431
C3'	CT	0.2336	0.1433
H3'	H1	0.0363	0.0719
C2'	CT	-0.0405	-0.0565
H2'1	HC	0.0287	0.0445
H2'2	HC	0.0287	0.0445
N9	N*	-0.0791	-0.0438
N1	NC	-0.4172	-0.7314
C2	CQ	0.6269	0.8253
N2	N2	-0.9021	-0.9516
H21	H	0.3758	0.4031
H22	H	0.3758	0.4031
N3	NC	-0.5135	-0.588
C4	CB	0.2216	0.2658
C5	CB	0.1513	0.0834
C6	CA	0.3917	0.6332
O6	OS	-0.3159	-0.4136
N7	NB	-0.5899	-0.5995
C8	CK	0.2068	0.1512
H8	H5	0.1357	0.1618
C9	CT	0.0249	0.1154
H91	H1	0.0718	0.0654
H92	H1	0.0718	0.0654
C10	CT	0.0199	0.0326
H101	HC	0.0373	0.0313
H102	HC	0.0373	0.0313
C11	CT	-0.0959	-0.0029
H111	HC	0.0497	0.0125
H112	HC	0.0497	0.0125
C13	CA	-0.0585	-0.0502
C14	CA	-0.0378	-0.0548
H14	HA	0.1577	0.135
C15	CA	-0.2635	-0.3221
H15	HA	0.1481	0.1541
C16	CA	0.2165	0.2986
H16	H4	0.0964	0.07
C17	CA	0.2337	0.2679
H17	H4	0.0431	0.0862
N18	NC	-0.4886	-0.5909
C12	C (CT)[a]	0.4018	0.1062
O12	O (OH)[a]	-0.4907	-0.6278
H12'	(HO)[a]	N/A	0.4061
H12	(H1)[a]	N/A	0.0545

[a]Atom types for PHB-G are given in brackets

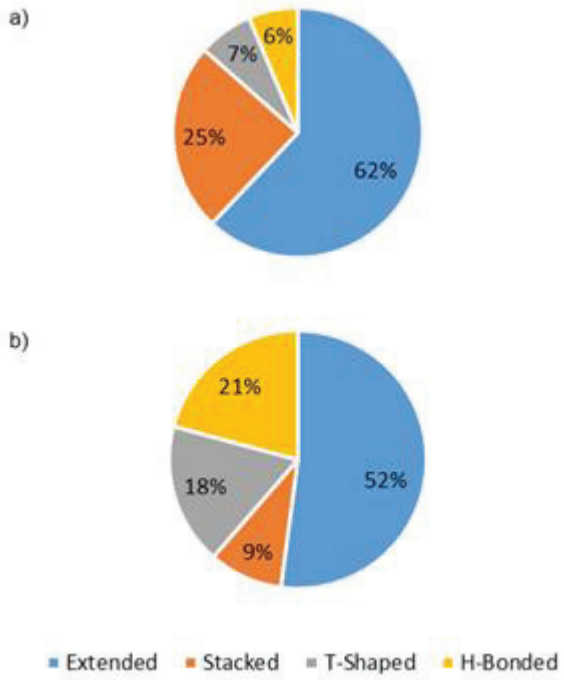

Figure D.1. Distribution of the a) POB-G and b) PHB-G nucleobase conformations resulting from the DFT conformational search.

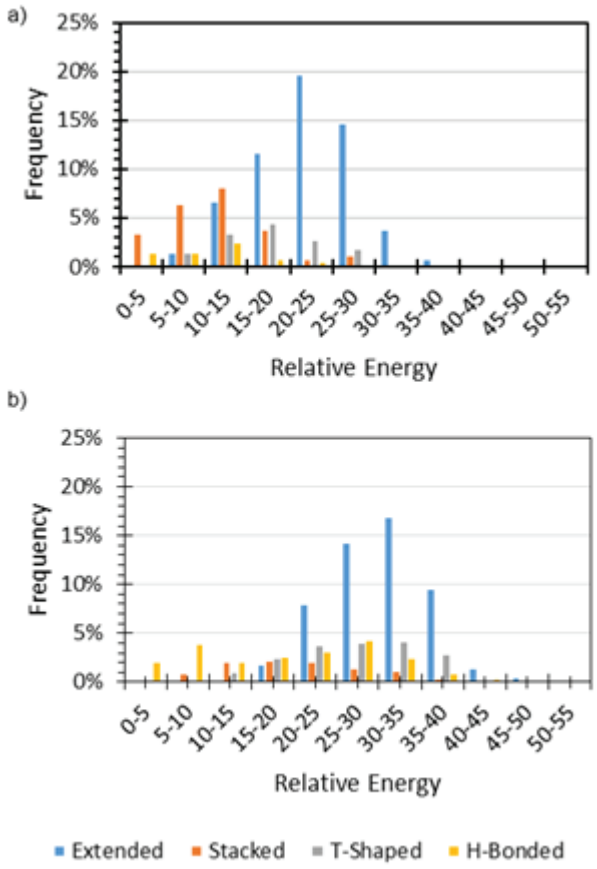

Figure D.2. Distribution of the B3LYP-D3(BJ)/6-311+G(2df,2p) relative energies (kJ/mol) of the a) POB-G and b) PHB-G nucleobase conformations for each structural category.

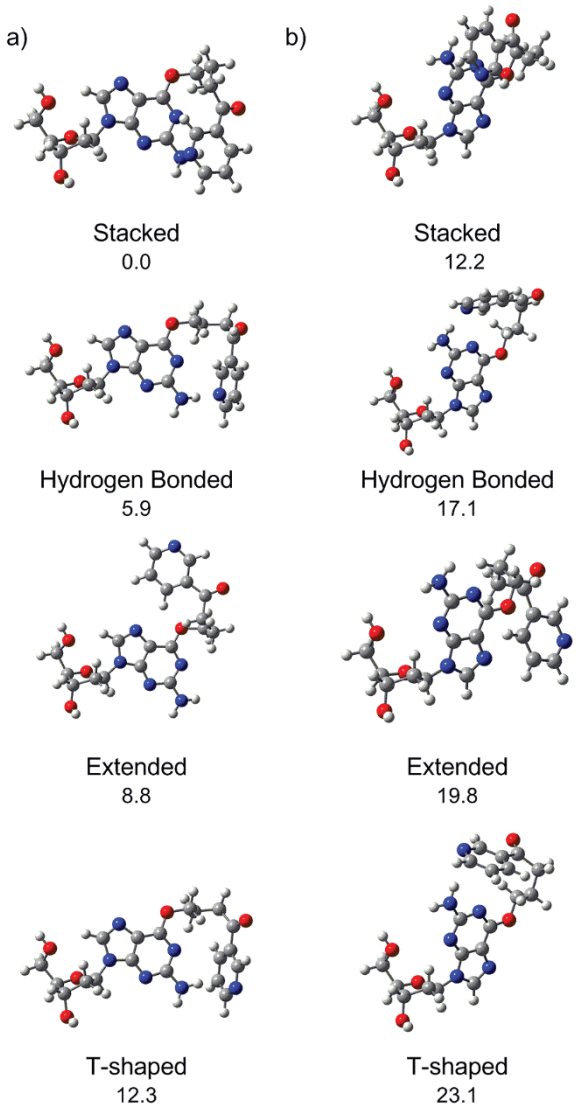

Figure D.3. For each structural category B3LYP-D3(BJ)/6-311+G(2df,2p)//B3LYP-D3(BJ)/6-31G(d) a) *anti* and b) *syn* structures of the POB-G nucleoside, as well as the relative energies (kJ/mol).

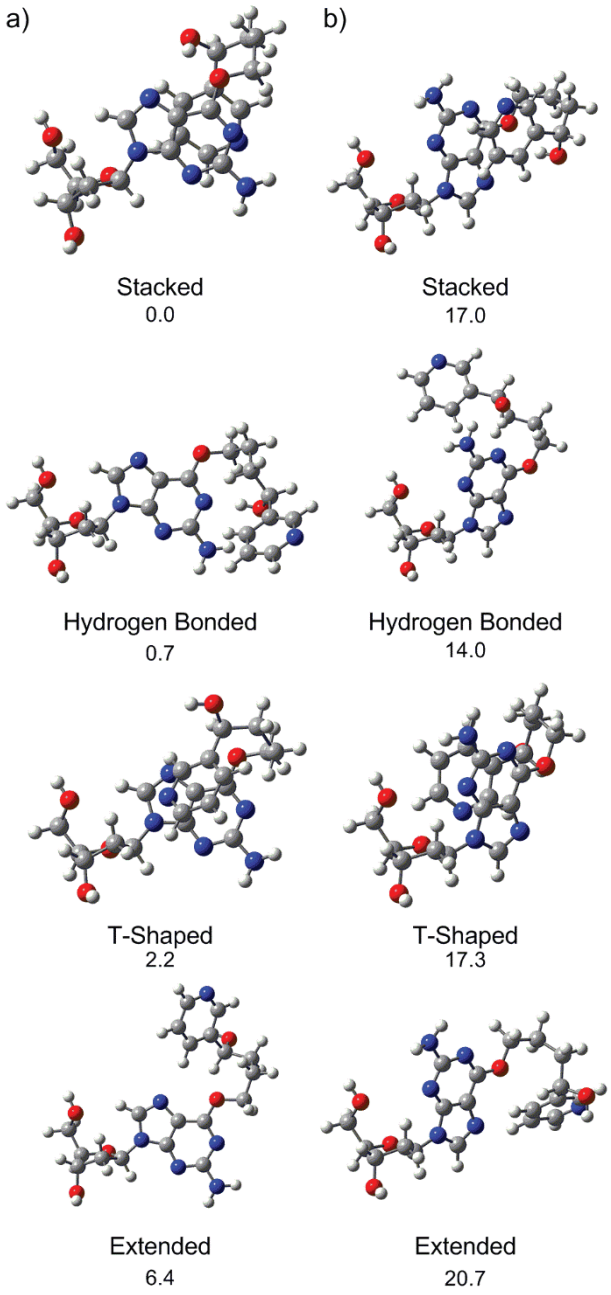

Figure D.4. For each structural category B3LYP-D3(BJ)/6-311+G(2df,2p)//B3LYP-D3(BJ)/6-31G(d) a) *anti* and b) *syn* structures of the PHB-G nucleoside, as well as the relative energies (kJ/mol).

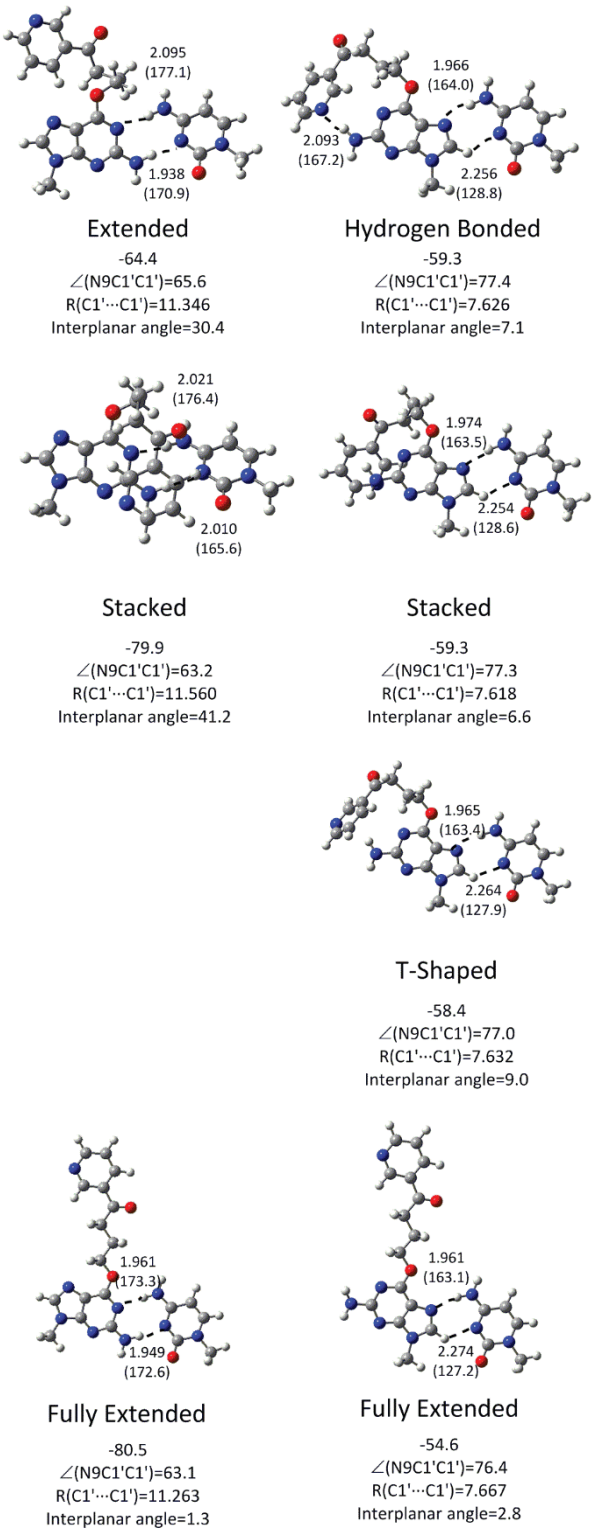

Figure D.5. B3LYP-D3(BJ)/6-311+G(2df,2p)//M06-2X/6-31G(d) structures (distances in Å and angles in deg.) and binding energy (kJ/mol) for dimers between the Watson-Crick (left) or Hoogsteen (right) face of various POB-G conformations and C.

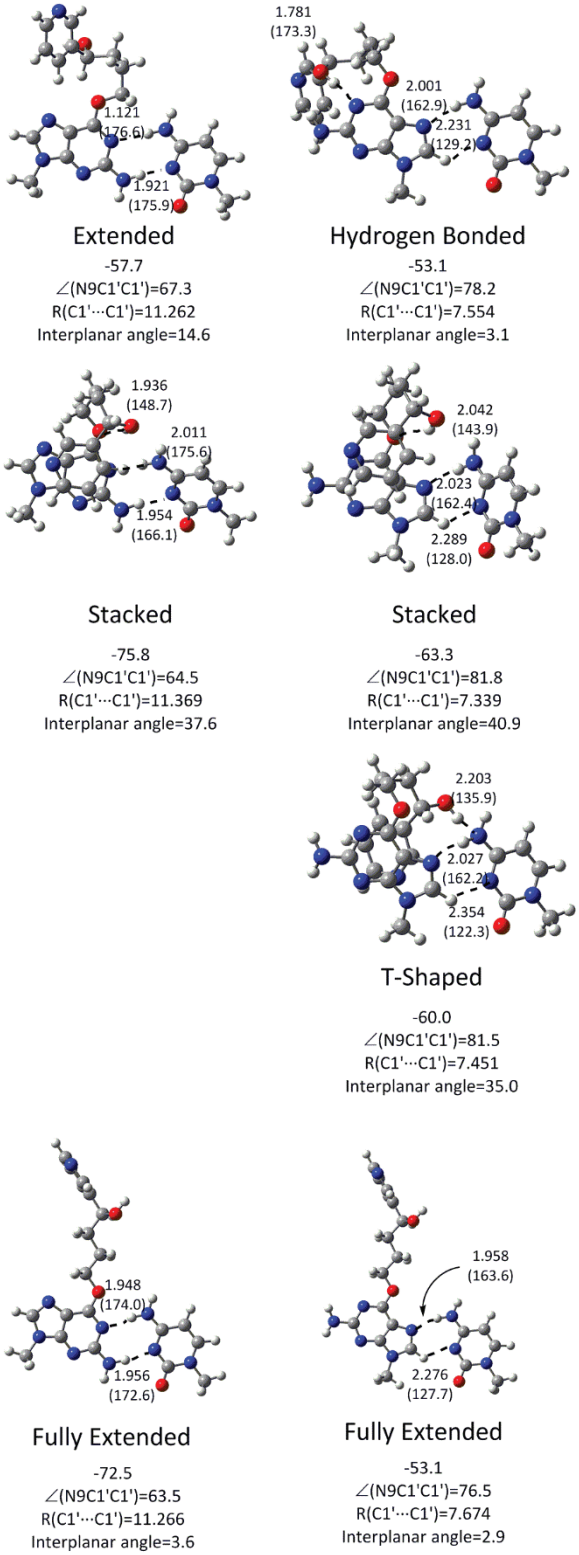

Figure D.6. B3LYP-D3(BJ)/6-311+G(2df,2p)//M06-2X/6-31G(d) structures (distances in Å and angles in deg.) and binding energy (kJ/mol) for dimers between the Watson-Crick (left) or Hoogsteen (right) face of various PHB-G conformations and C.

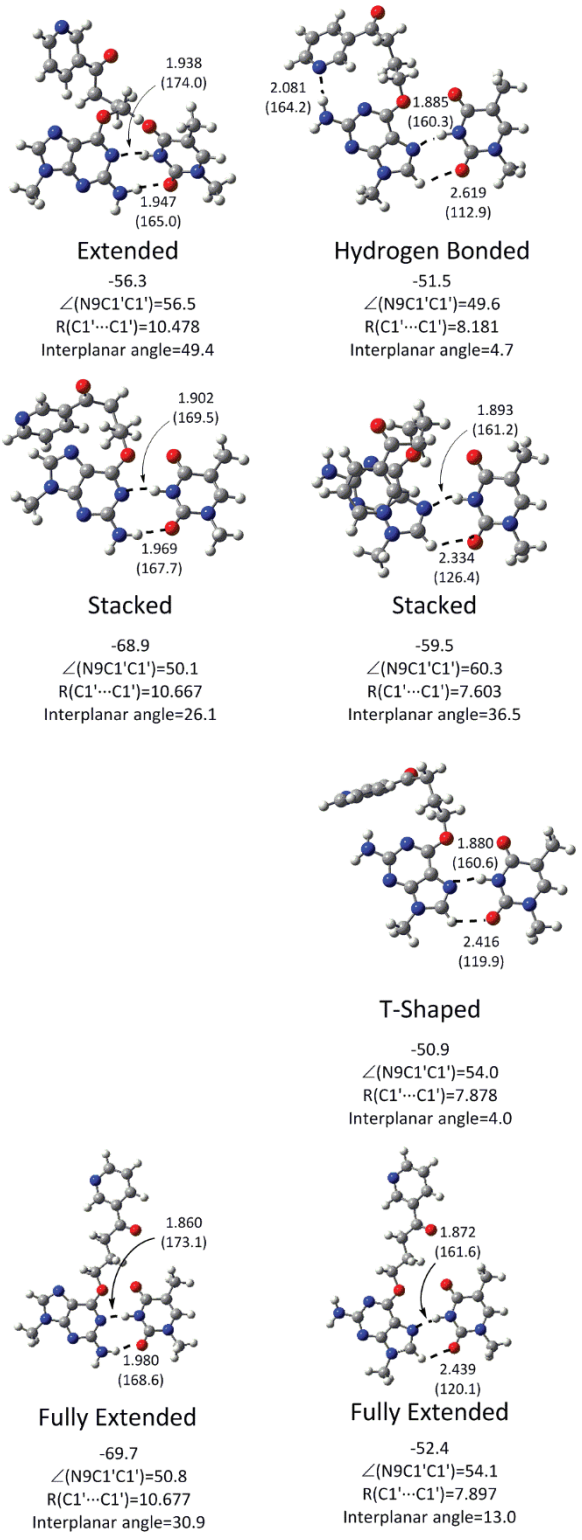

Figure D.7. B3LYP-D3(BJ)/6-311+G(2df,2p)//M06-2X/6-31G(d) structures (distances in Å and angles in deg.) and binding energy (kJ/mol) for dimers between the Watson-Crick (left) or Hoogsteen (right) face of various POB-G conformations and T.

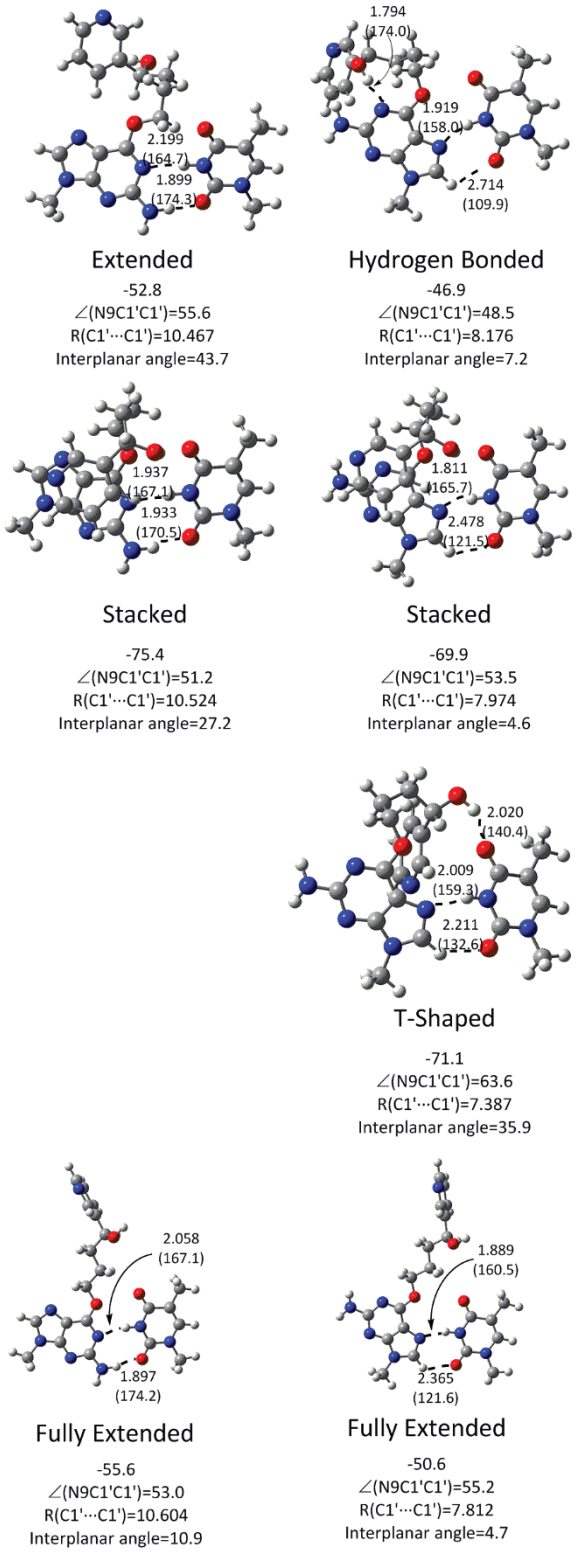

Figure D.8. B3LYP-D3(BJ)/6-311+G(2df,2p)//M06-2X/6-31G(d) structures (distances in Å and angles in deg.) and binding energy (kJ/mol) for dimers between the Watson-Crick (left) or Hoogsteen (right) face of various PHB-G conformations and T.

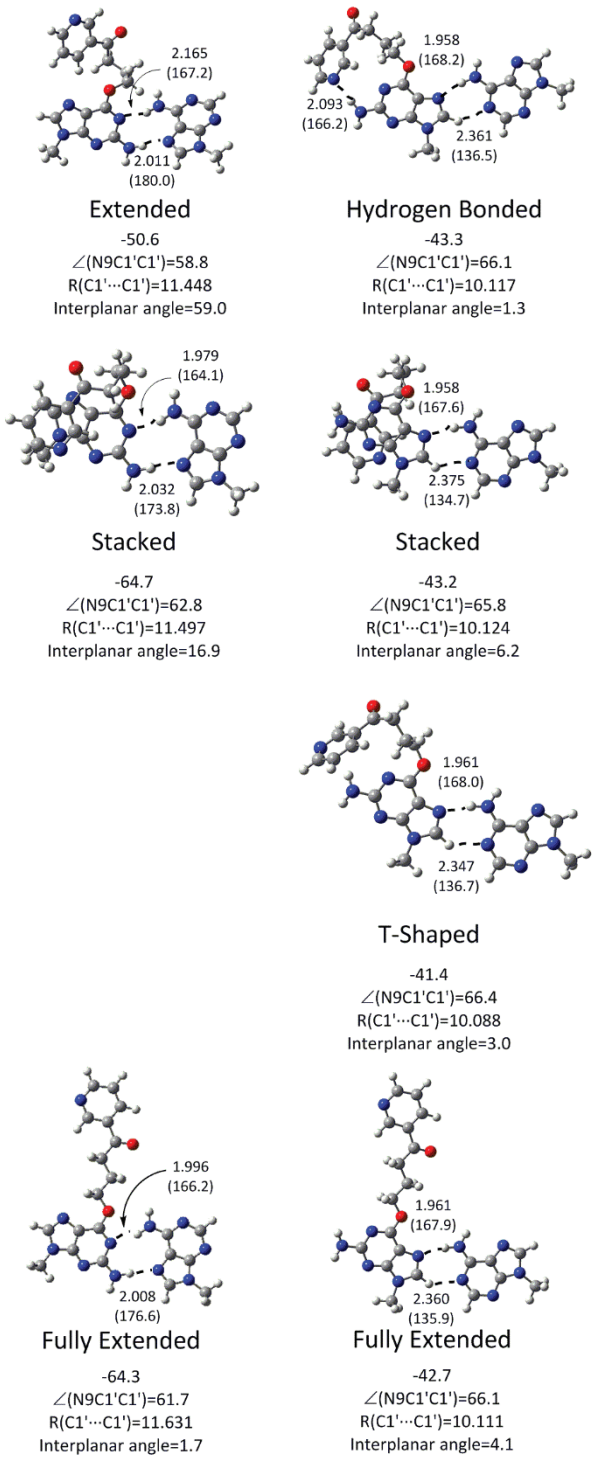

Figure D.9. B3LYP-D3(BJ)/6-311+G(2df,2p)//M06-2X/6-31G(d) structures (distances in Å and angles in deg.) and binding energy (kJ/mol) for dimers between the Watson-Crick (left) or Hoogsteen (right) face of various POB-G conformations and A.

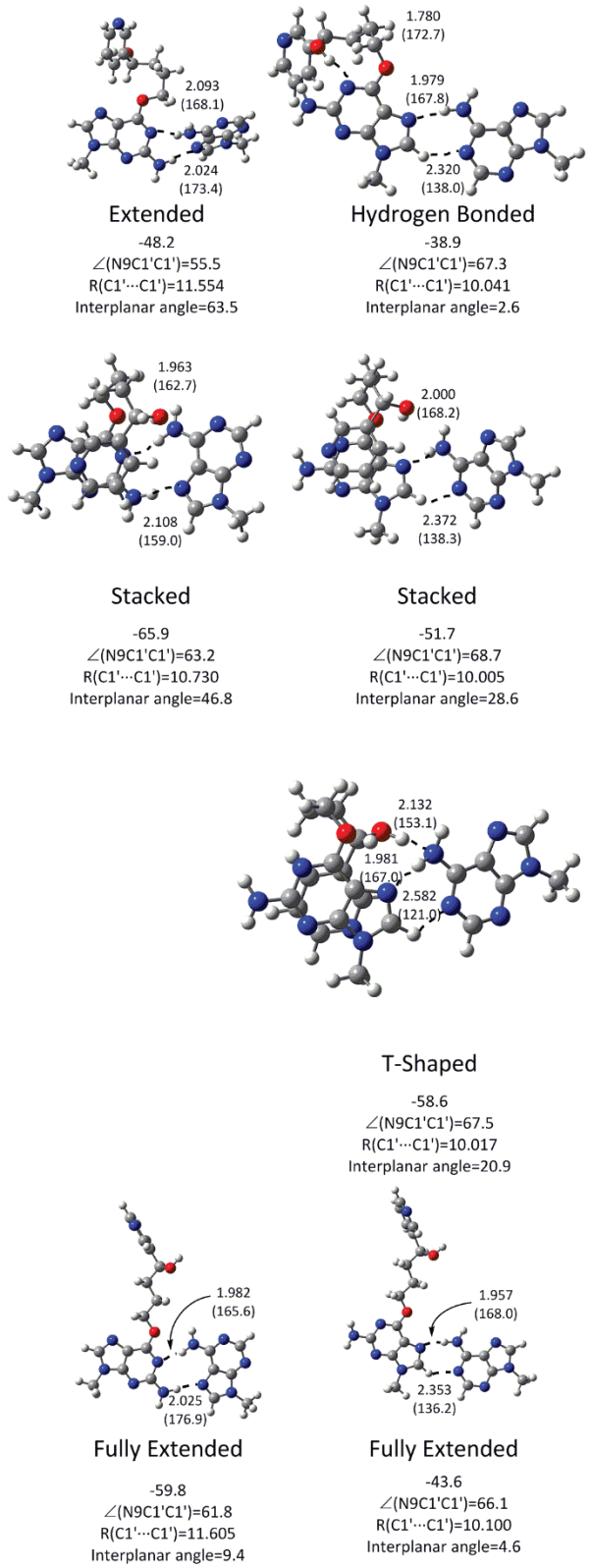

Figure D.10. B3LYP-D3(BJ)/6-311+G(2df,2p)//M06-2X/6-31G(d) structures (distances in Å and angles in deg.) and binding energy (kJ/mol) for dimers between the Watson-Crick (left) or Hoogsteen (right) face of various PHB-G conformations and A.

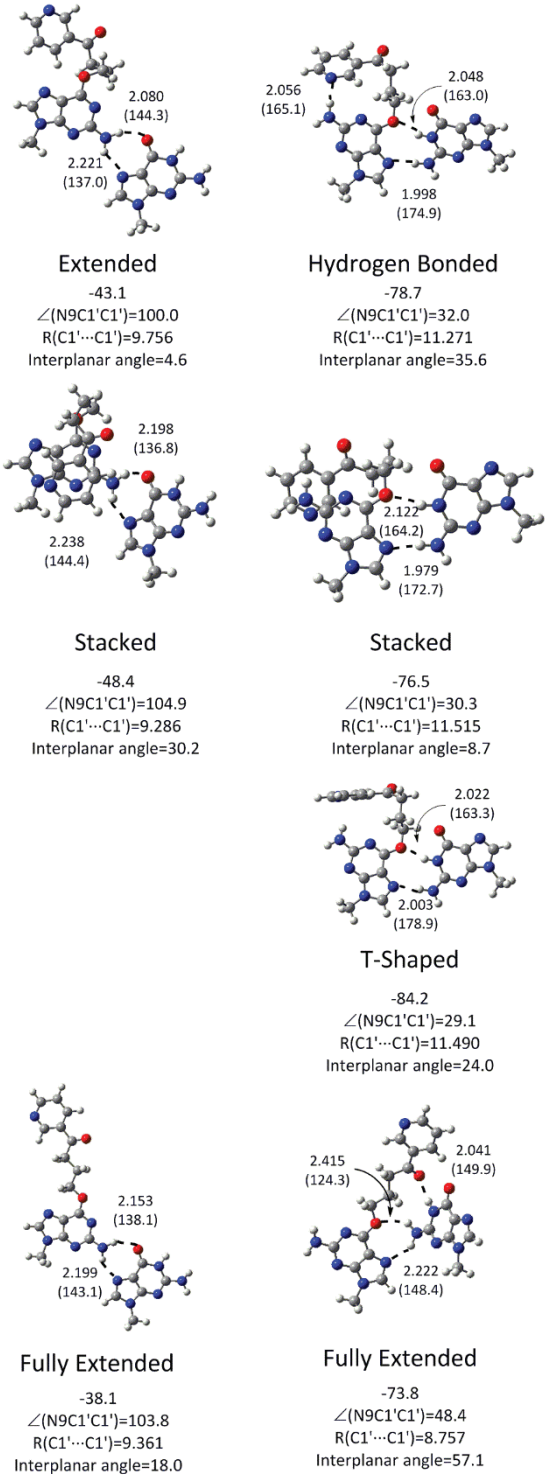

Figure D.11. B3LYP-D3(BJ)/6-311+G(2df,2p)//M06-2X/6-31G(d) structures (distances in Å and angles in deg.) and binding energy (kJ/mol) for dimers between the Watson-Crick (left) or Hoogsteen (right) face of various POB-G conformations and G.

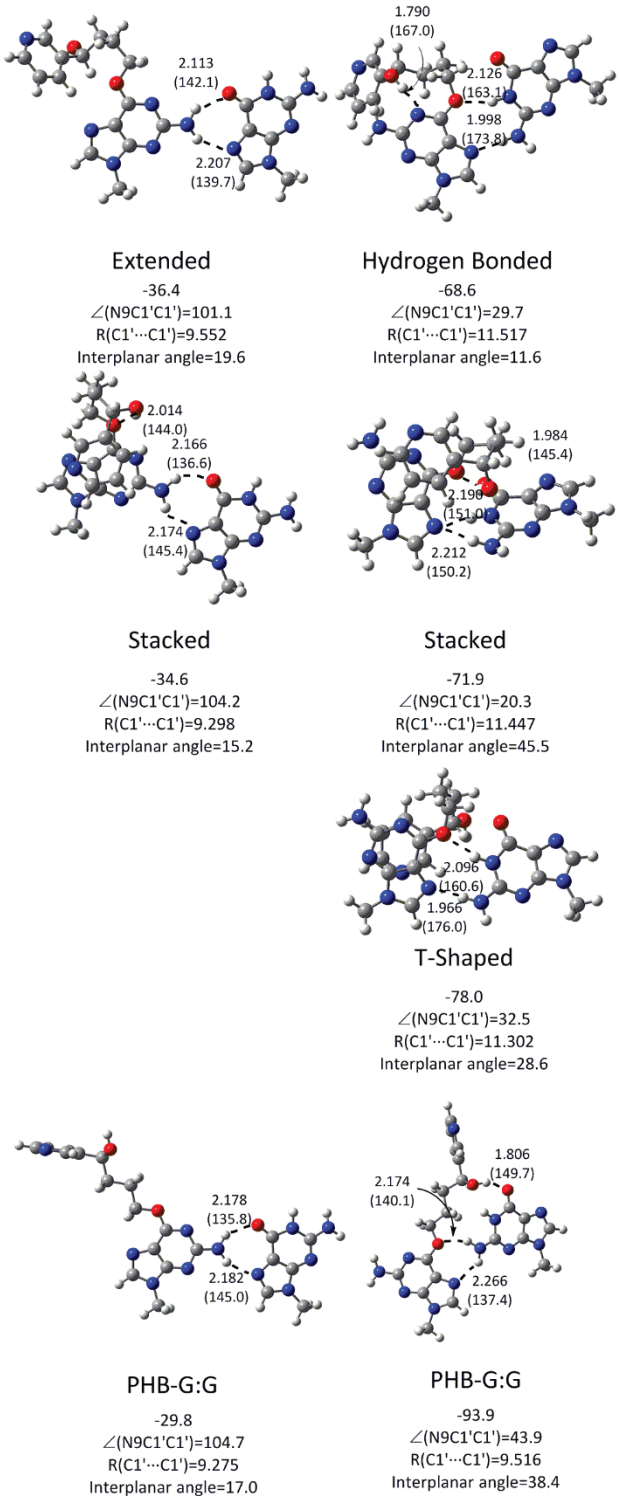

Figure D.12. B3LYP-D3(BJ)/6-311+G(2df,2p)//M06-2X/6-31G(d) structures (distances in Å and angles in deg.) and binding energy (kJ/mol) for dimers between the Watson-Crick (left) or Hoogsteen (right) face of various PHB-G conformations and G.

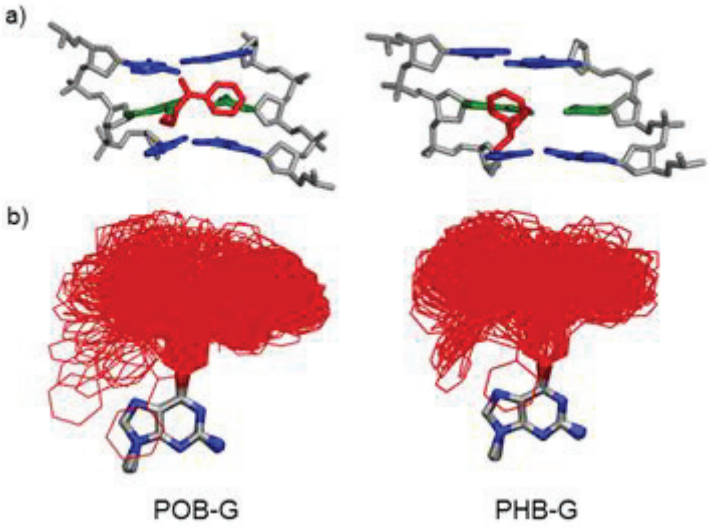

Figure D.13. a) MD representative structures obtained for POB-G (left) and PHB-G (right) adducted DNA based on an initial stacked lesion conformation opposite C. b) Overlay (based on G ring atoms) of lesion conformations adopted throughout the MD simulation, highlighting the deviation in bulky moiety orientation (red).

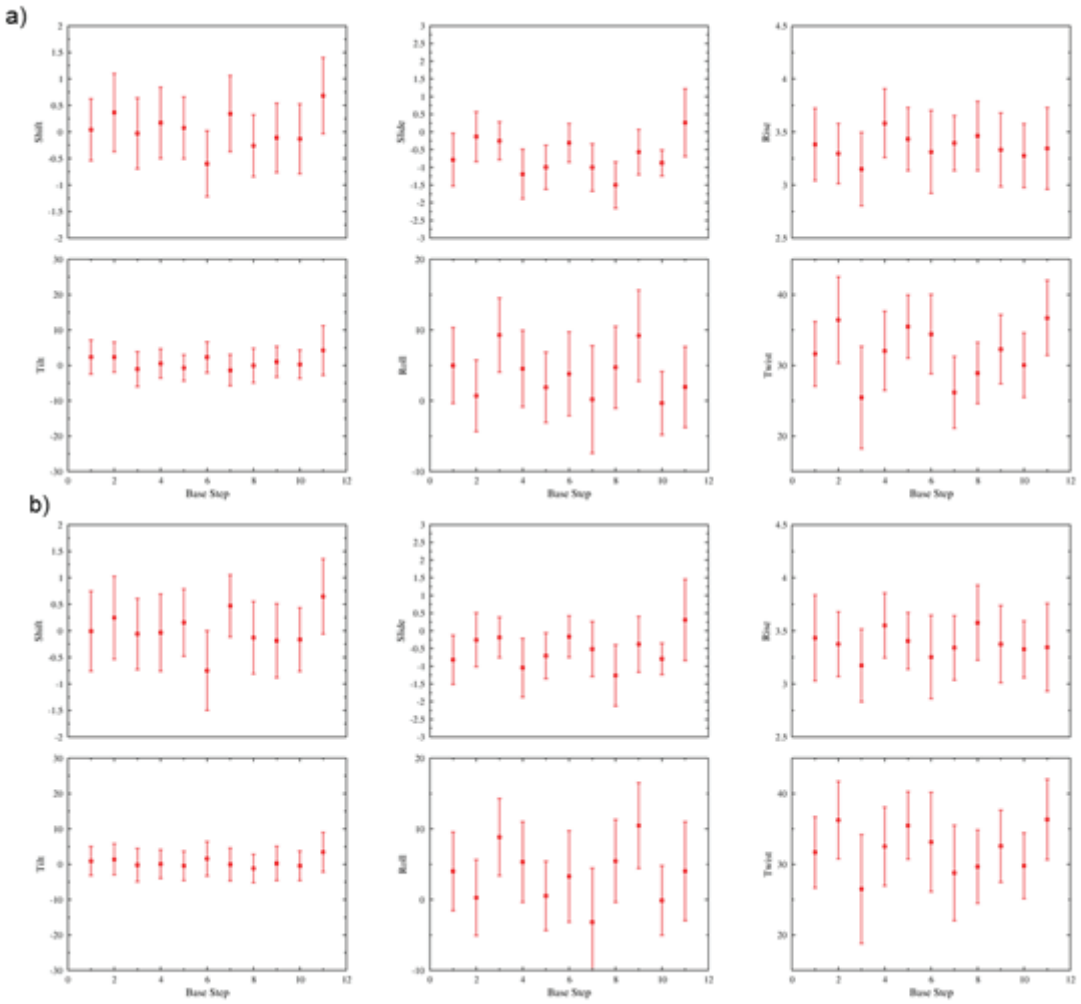

Figure D.14. Average base step parameters from MD simulations on a) POB-G or b) PHB-G adducted DNA with the lesion in the extended conformation paired opposite C.

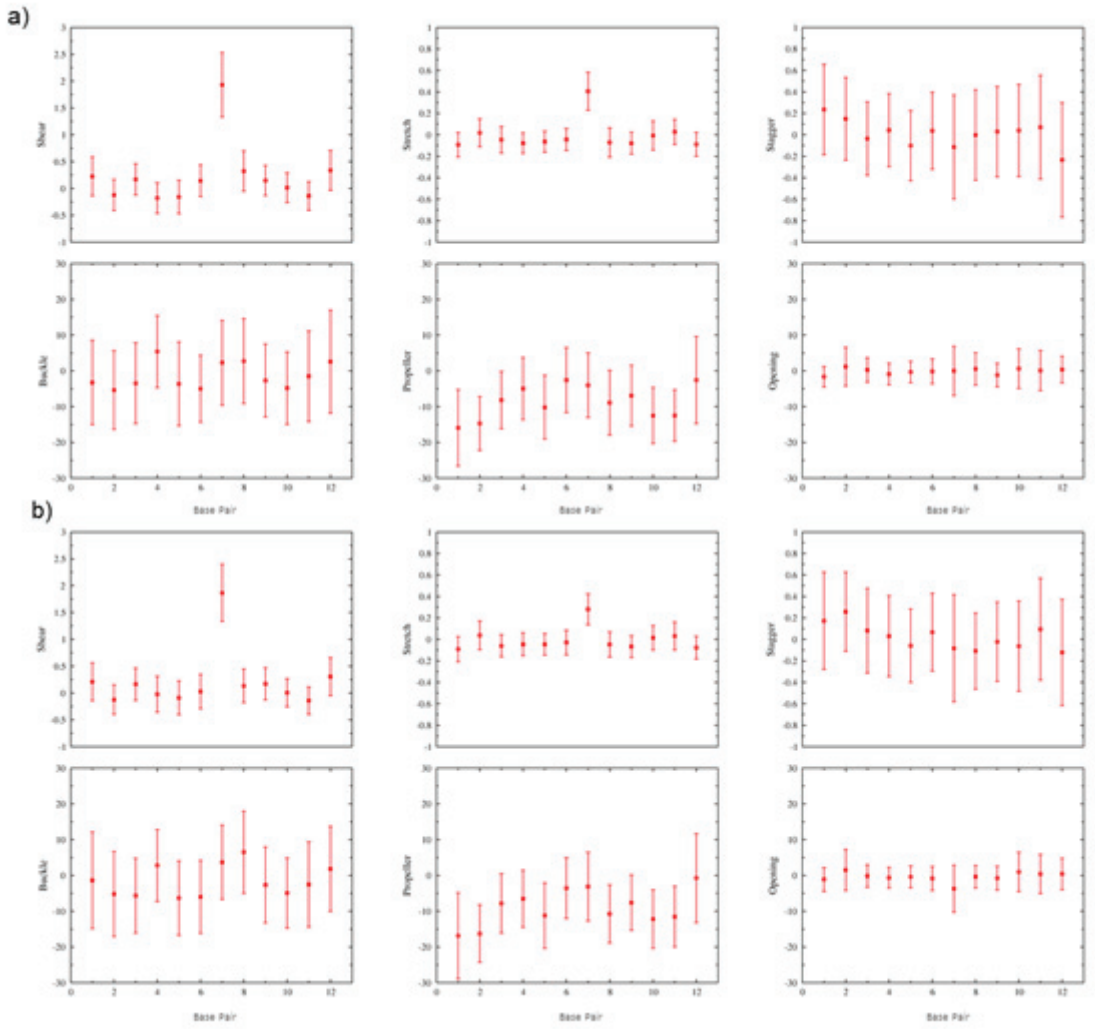

Figure D.15. Average base pair parameters for a) POB-G or b) PHB-G adducted DNA with the lesion in the extended conformation paired opposite C.

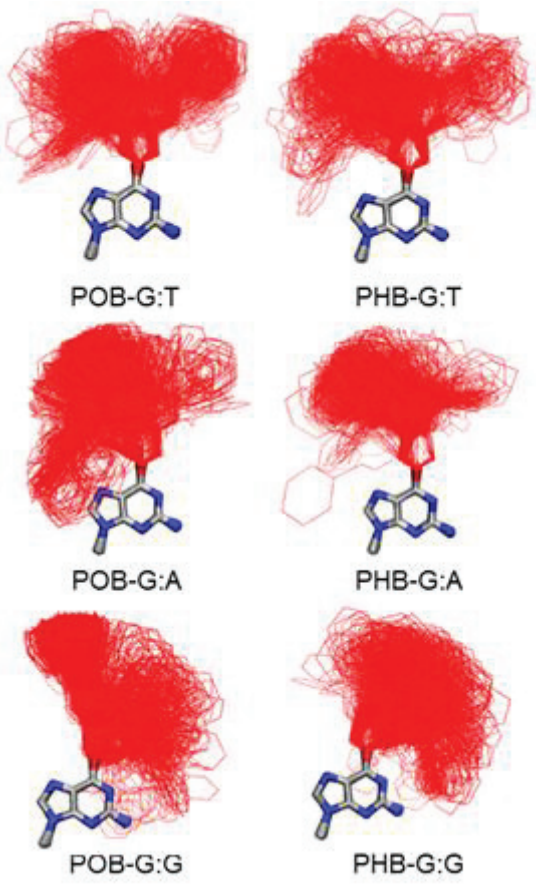

Figure D.16. Overlay (based on G ring atoms) of lesion conformations adopted throughout MD simulations on DNA containing POB-G(left) or PHB-G (right) paired opposite T (top), A (middle) and G (bottom), highlighting the deviation in bulky moiety orientation (red).

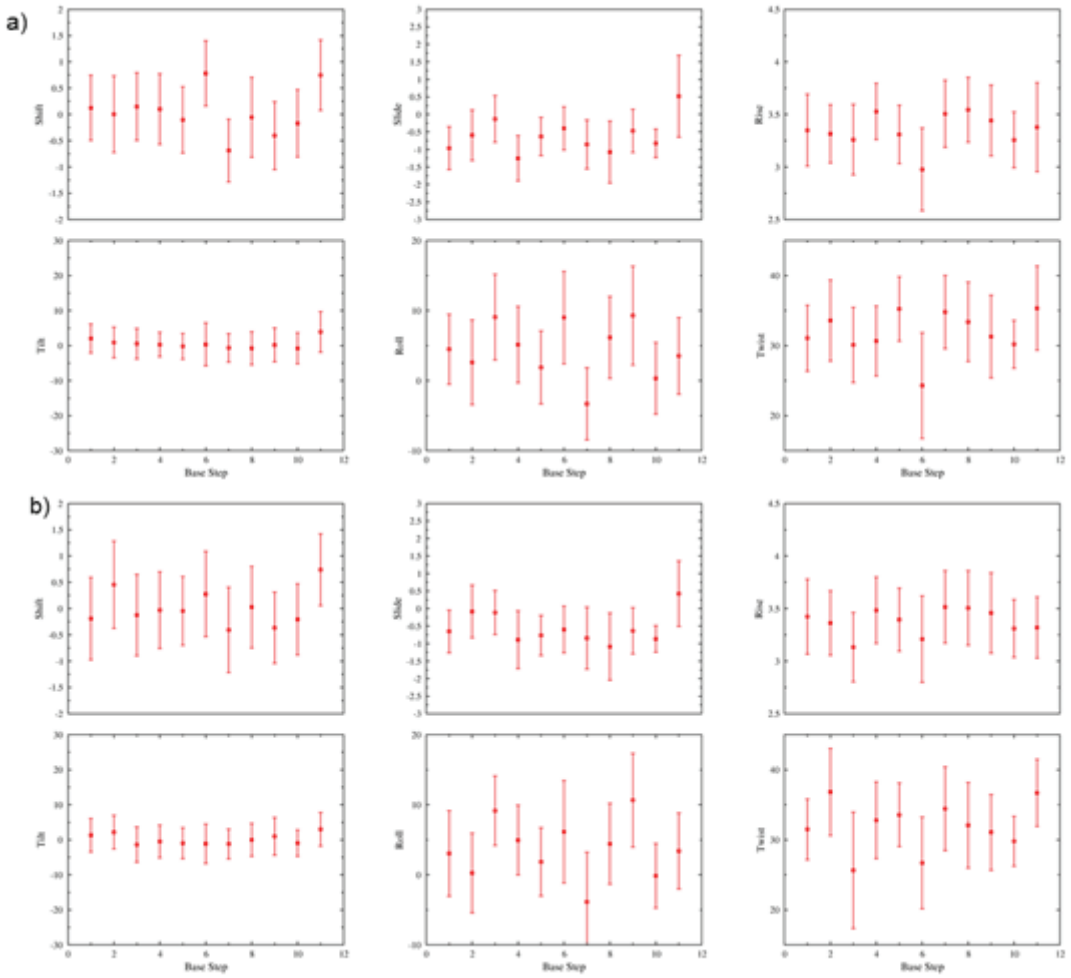

Figure D.17. Average base step parameters from MD simulations on a) POB-G or b) PHB-G adducted DNA with the lesion in the extended conformation paired opposite T.

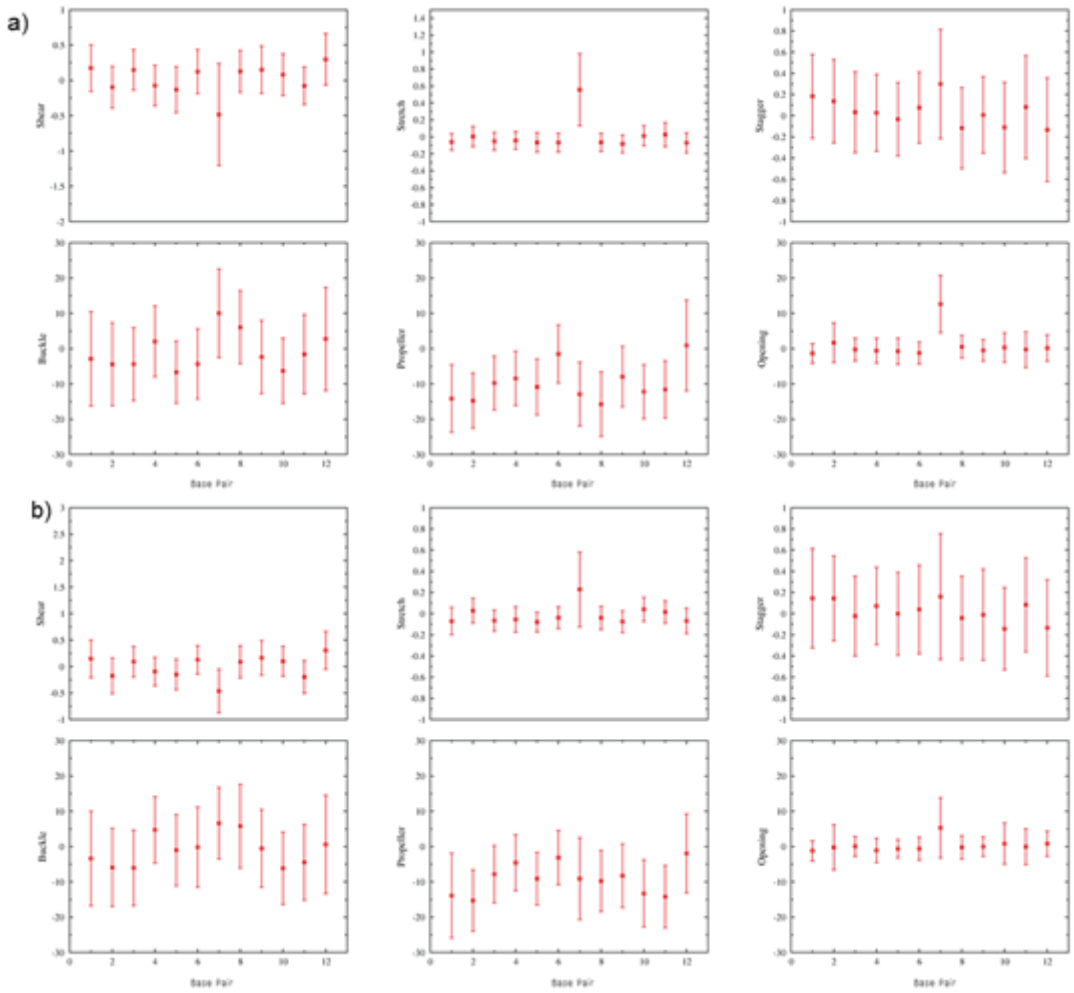

Figure D.18. Average base pair parameters for a) POB-G or b) PHB-G adducted DNA with the lesion in the extended conformation paired opposite T.

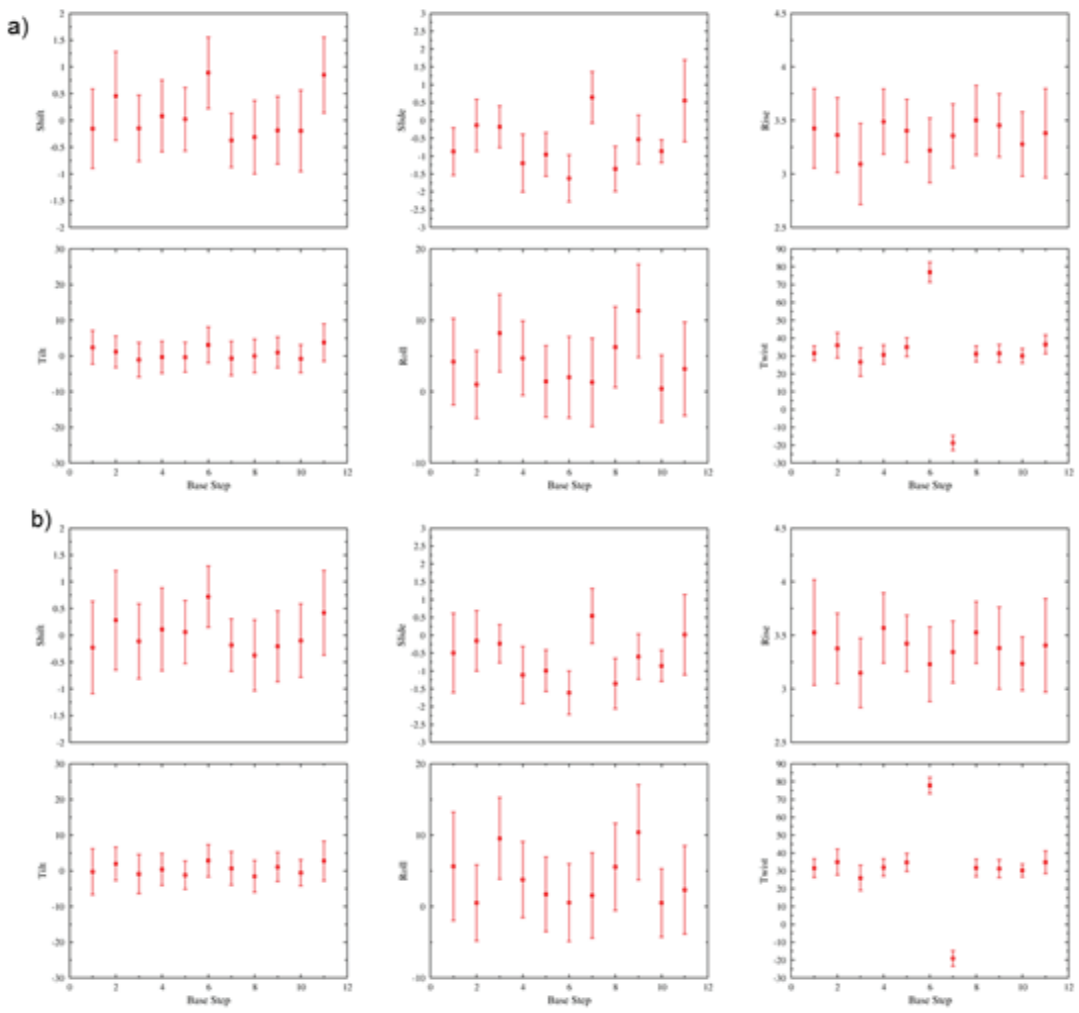

Figure D.19. Average base step parameters for a) POB-G or b) PHB-G adducted DNA with the lesion in the extended conformation paired opposite A.

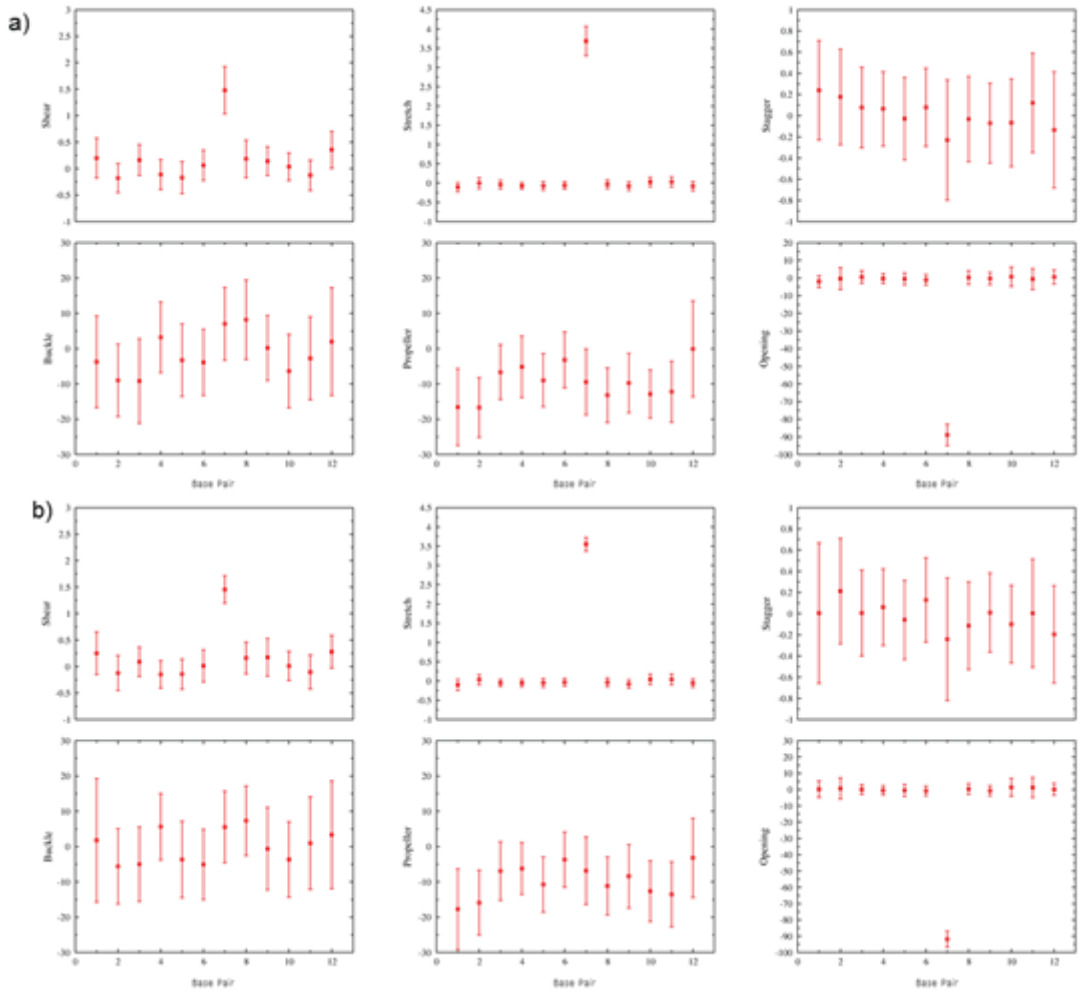

Figure D.20. Average base pair parameters for a) POB-G or b) PHB-G adducted DNA with the lesion in the extended conformation paired opposite A.

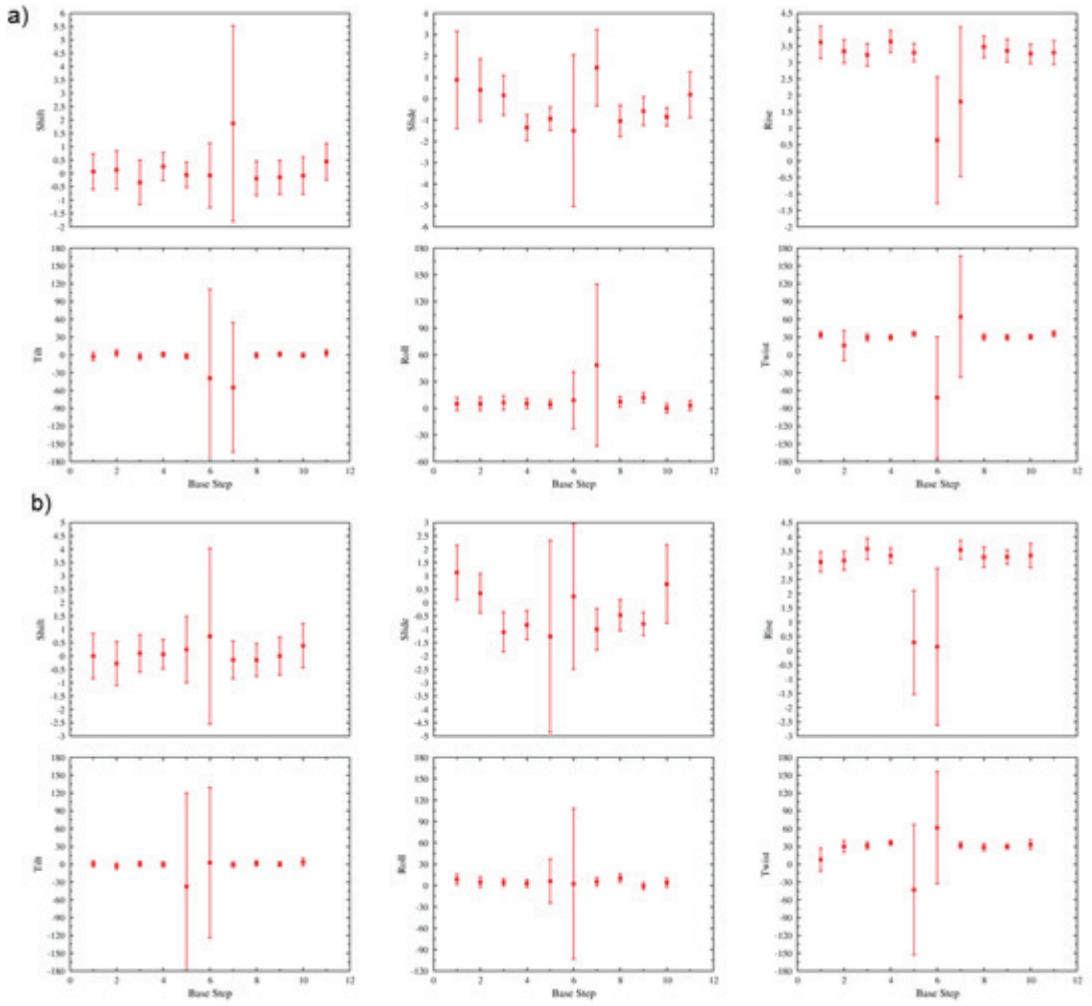

Figure D.21. Average base step parameters for a) POB-G or b) PHB-G adducted DNA with the lesion in the extended conformation paired opposite G.

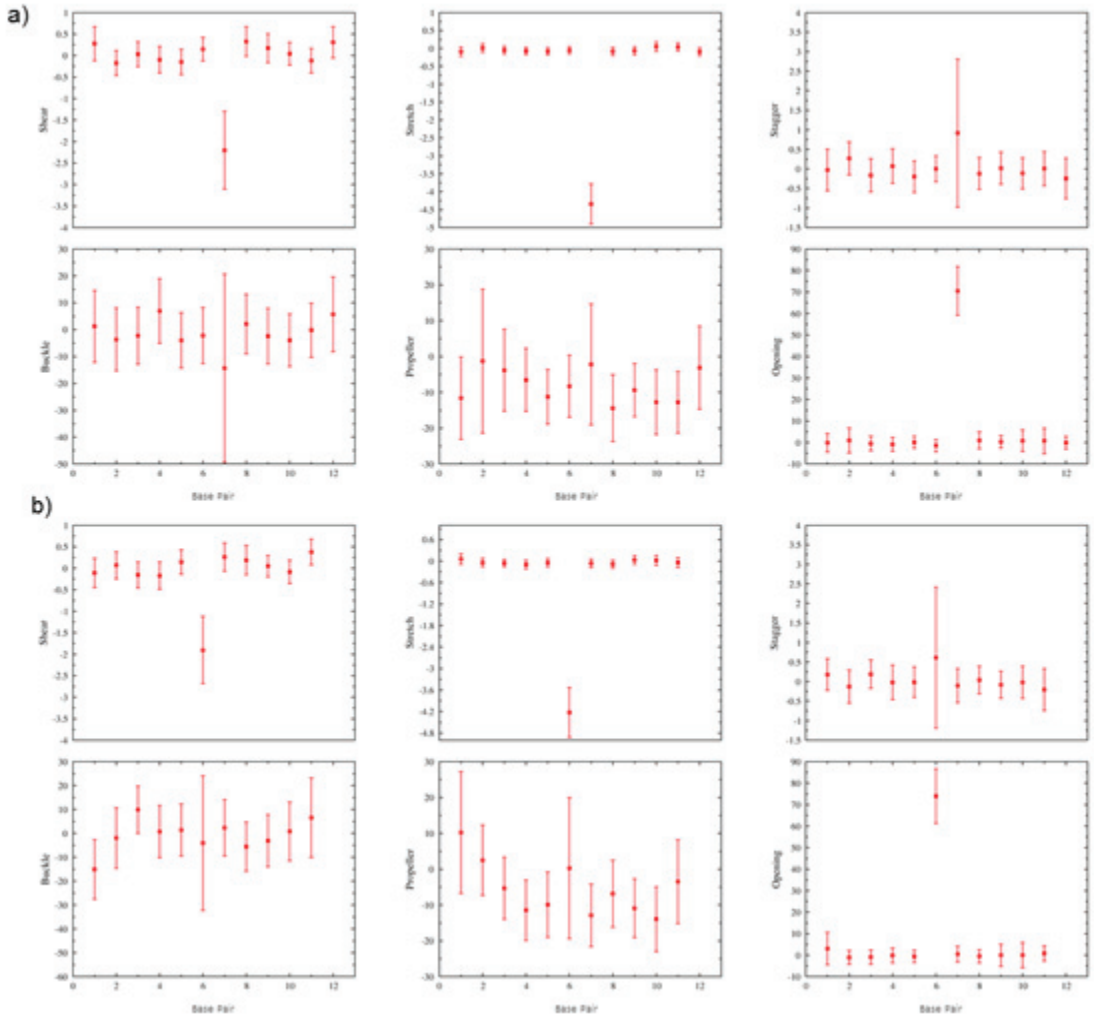

Figure D.22. Average base pair parameters for a) POB-G or b) PHB-G adducted DNA with the lesion in the extended conformation paired opposite G.

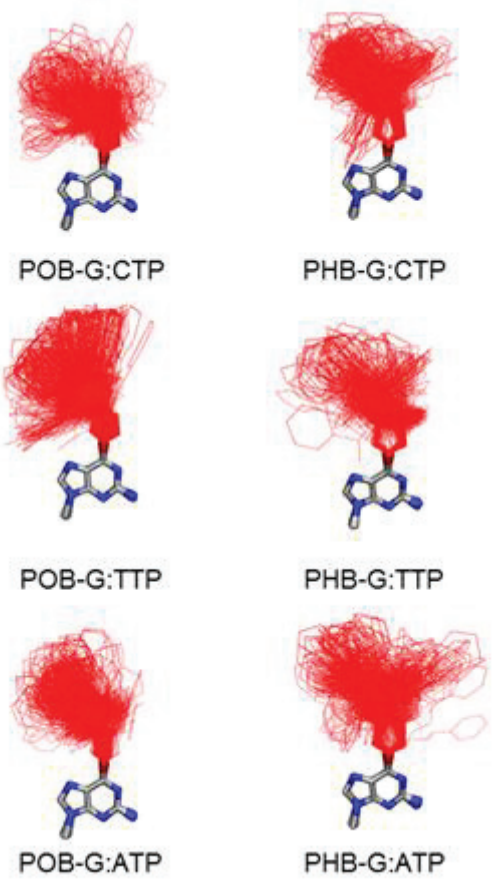

Figure D.23. Overlay (based on G ring atoms) of lesion conformations adopted throughout the MD simulations on polymerase η complexes corresponding to dCTP (top), dTTP (middle), or dATP (bottom) insertion opposite POB-G (left) or PHB-G (right), highlighting the deviation in bulky moiety orientation (red).

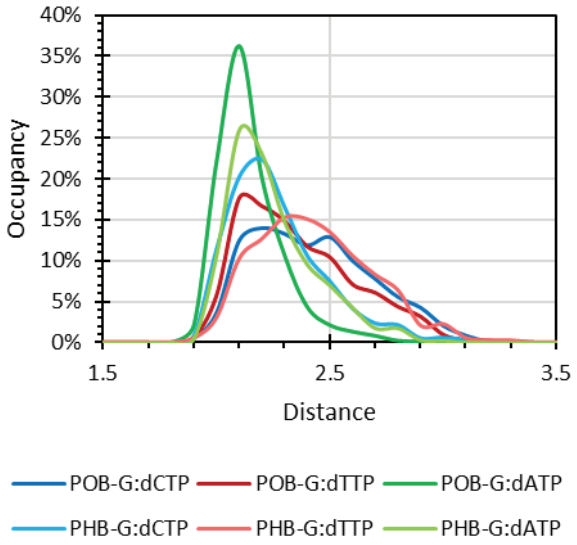

Figure D.24. Distance between the binding Mg^{2+} ion and coordinating Oα during MD simulations on the insertion of a dNTP opposite POB-G or PHB-G by polymerase η. Full coordination was considered to occur if the distance was < 2.5 Å for > 95% of the simulation.

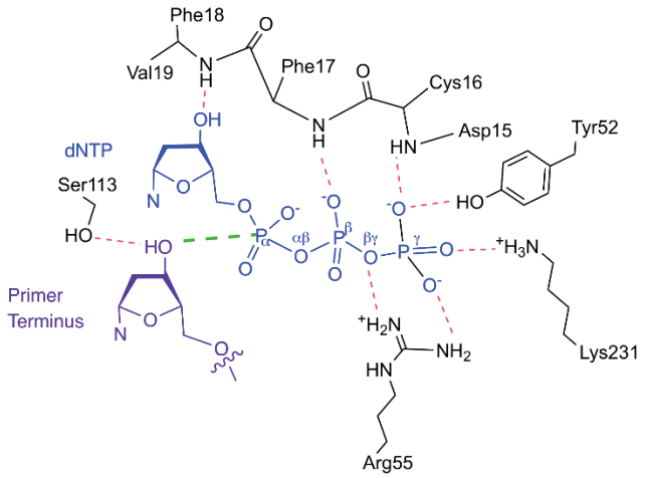

Figure D.25. Hydrogen-bonding interactions with the dNTP in the polymerase η active site observed in the crystal structure corresponding to dATP incorporation opposite T (PDB ID: 4ECS).

Appendix E

Supplementary Information for Chapter 6: Evidence for the replication of a *syn* DNA pyrimidine: A computational investigation of a highly mutagenic tobacco-derived thymine lesion

Contains full computational details, Figures E.1–E.13 and Tables E.1–E.14.

Full Computational Details

Nucleobase. To parallel a previous study on POB-G (1), we initially investigated the flexibility of the POB-T nucleobase based on a fully extended orientation (i.e., α′, β′, γ′, δ′, ϵ′, and ρ′ = 180°, Figure 6.1a). The rotational barriers with respect to α′, β′, γ′, δ′, ϵ′, and ρ′ were determined by altering each dihedral angle in 10° increments from 0 to 360° using B3LYP-D3(BJ)/6-31G(d). Due to the absence of steric constraints, this model will provide the lowest energy rotational barriers within the bulky moiety. Subsequently, to gain a more comprehensive picture of the possible orientations of the POB-T nucleobase, a usage directed conformational search with respect to α′, β′, γ′, δ′, ϵ′, and ρ′ as implemented in the HyperChem program (2) was performed using AMBER99 (see Table E.13 for charges and atom types). During the search, a maximum of 100,000 iterations or 1000 optimizations were performed. Any retained conformations were required to be within 42 kJ/mol of the lowest energy conformation, have an energy difference from any other conformation of > 0.2 kJ/mol, and have a heavy atom root-mean-square deviation from any other conformation of < 0.25 Å. All 282 POB-T conformers isolated from this conformational search were subsequently optimized using B3LYP-D3(BJ)/6-31G(d) and the relative energies were determined using B3LYP-D3(BJ)/6-311++G(2df,2p). All duplicate conformations following the B3LYP-D3(BJ) analysis (defined based on an energy difference from any other conformation of < 0.2 kJ/mol, and a difference in α′, β′, γ′, δ′, ϵ′, and ρ′ from any other conformation of < 0.2°) were removed from the data set. The 164 remaining conformations were visually categorized based on the type of interaction between the bulky moiety and the adducted T. The barrier about the glycosidic bond in the POB-T nucleoside was determined by scanning with respect to χ in 10° increments from 0

to 360° using B3LYP-D3(BJ)/6-31G(d). In the scan, a fully extended nucleoside model was considered that contains the 2′-*endo* sugar pucker, and the O3′ hydroxy group oriented and the O5′ hydroxy group fixed to reflect the geometry of a non-terminal nucleoside (i.e., ∠(HC3′O3′H) ≈ −60° and ∠(C4′C5′O5′H) = 180°). Stationary points from the scan were fully optimized using B3LYP-D3(BJ)/6-31G(d) and the relative energies were calculated with B3LYP-D3(BJ)/6-311++G(2df,2p) and include zero-point vibrational energy corrections scaled by 0.9670.

Nucleobase Dimers. Based on previous work on the POB-G adduct (1), M06-2X/6-31G(d) was used to optimized the hydrogen-bonding pairs between T or POB-T and the canonical nucleobases. The sugar–phosphate backbone of DNA was replaced by a methyl group to reduce computational costs and yet permit analysis of changes in the base pair width (C1′–C1′ distance) and base pair opening ∠(N9C1′C1′). Additionally, the base pair distortion was monitored by measuring the angle between the two nucleobase planes (denoted interplanar angle). The interaction energies were calculated with B3LYP-D3(BJ)/6-311++G(2df,2p) and include zero-point vibrational energy corrections scaled by 0.9670. The deformation energy was not considered in the reported binding strengths due to the significant flexibility of the POB-T nucleobase.

All DFT calculations were performed using Gaussian 09 (revision D.01) (3).

Polymerase and DNA. Initial structure for MD simulations on the pol κ replication complexes were obtained from chains A, C, and D for a crystal structure corresponding to 2'-deoxy-5'-O-[(R)-hydroxy[[(R)-hydroxy(phosphonooxy)phosphoryl]amino]phosphoryl]

adenosine (dAMPNPP) insertion opposite T for the extension past a lucidin guanine adduct derivative (PDB ID: 5W2C). Glycerol, 1,2-ethanediol, and di(hydroxyethyl)ether were removed from the system. Missing N-terminal residues (Met–24 – Gly30), a loop (Met225 – Gln281) and C-terminal residues (Glu519 – Gln526) were not included in the model. Indeed, previous studies on pol κ mutants, as well as work on the pol κ bypass of the N2-furfuryl-G and N6-furfuryl-A lesions has shown that the exclusion of these residues does not affect the activity of the enzyme (4, 5).

Initial structures for MD simulations on the pol η replication complexes were obtained from a crystal structure corresponding to the insertion of dATP opposite T (PDB ID: 4ECS). Missing residues (Thr155–Glu159) were added by hand using GaussView. Glycerol, Ca^{2+}, and pyrophosphate were removed from the system. In cases where multiple orientations of amino acids were present in the experimental data, the orientation that best aligns the residue to interact with the surrounding amino acids was selected, while the highest occupied orientation was chosen in cases were the amino acid was not positioned to interact with a surrounding residue.

The conformation of POB-T in the polymerase active site was initially investigated by overlaying the nucleobase conformations identified in the DFT conformational search onto the template T nucleobase in the crystal structure (according to the heavy atoms) in the absence of an incoming dNTP (preinsertion complex). For each overlay, the closest heavy atom distance between the lesion and the surrounding DNA or protein was measured. Although this approach does not account for rearrangement of the active site to accommodate different lesion orientations, the resulting structures are reasonable starting

points for MD simulations. Representative lesion orientations were then used to initiate MD simulation to understand the dynamics of POB-T binding in the pol η and κ active sites, and further explore conformations that can be accommodated by the enzymes. Additionally, directed by DFT calculations on nucleobase dimers, pol κ insertion complexes with a pairing dGTP or dATP opposite *anti*-POB-T, and pol η insertion complexes with dCTP, dTTP or dATP paired opposite *syn*-POB-T were modeled. Finally, based on the polymerase replication complexes, MD simulations were initiated on the 5′-CTGTGTCT*GCAGCG 14-mer DNA duplex, with T*=*syn*-POB-T opposite T, to understand the structure and dynamics of the post-replication helix.

The resulting complexes were prepared for simulation using the tleap module of AMBER 14 (6). Specifically, missing atoms, including hydrogen atoms, were added to generate the standard protonation states of all DNA and protein residues. Furthermore, NaCl was added to the water box to yield an approximate concentration of 0.15 M for DNA–polymerase complexes, while DNA systems were neutralized with Na^+. Mg^{2+} ions were included in the DNA–polymerase insertion complexes. All systems were solvated in a TIP3P octahedral water box such that the solute was at least 10.0 Å from the edge of the box. All canonical amino acids, nucleotides, and solvent were modeled with AMBER ff14SB parameters. Parameters for POB-T were assigned according to the GAFF and AMBER ff14SB force fields using ANTECHAMBER 1.4 (7), and partial charges for the lesions were developed using RESP charge fitting from an HF/6-31G(d) calculation by the R.E.D.v.III.4 program (8, 9) to be consistent with the AMBER force field (Table E.13, Appendix E). The parameters for the dNTPs were adapted from the literature (10-12), Mg^{2+} was modeled

using the parameters from Allner et al. (13), and Na^+ and Cl^- were modeled using the monovalent ion parameters from Joung and Cheatham (14).

All systems were minimized stepwise, with the first phase minimizing the water for 2500 steps of steepest descent and 2500 steps of conjugte gradient minimization, while applying a 50 kcal/(mol $Å^2$) restraint on the rest of the system. Next, the hydrogen atoms were minimized for 4000 steps of steepest descent and 4000 steps of conjugte gradient minimization, with a 50 kcal/(mol $Å^2$) restraint on the heavy atoms. Subsequently, the systems were minimized with a 50 kcal/(mol $Å^2$) restraint on the amino acid backbone for 12500 steps of steepest descent and 12500 steps of conjugate gradient minimization. Finally, the entire system was minimized without restraints for 5000 steps of steepest descent and 5000 steps of conjugate gradient minimization. For DNA systems, restraints were placed on the hydrogen-bonding distances in the terminal base pairs to prevent strand unraveling. The systems were then heated from 10 to 310 K in 6 steps, each increasing the temperature by 50 K over 10 ps using a 1 fs time step. Heating was performed with a 10 kcal/(mol $Å^2$) restraint on the solute using the Langevin thermostat (γ=1.0). The systems were then equilibrated with a constraint weight of 20, 15, 10, 5 and 1 kcal/(mol $Å^2$) on the solute. Each constraint weight was simulated for 20 ps using a 2 fs time step. Subsequently, 20 ns unconstrained pre-production simulations were performed to identify geometries that were distinctly different from the initial structures and cannot be easily converted during standard MD sampling, with particular attention being paid to the active site or lesion site (including lesion orientation and hydrogen bonding) for polymerase-DNA and DNA systems. Structures of the polymerase-DNA complexes with a hydrogen-bonding geometry and lesion orientation most conducive for the reaction were selected as initial structures for

the final 100 ns production simulations, which were run in triplicate. Since the replicas lead to negligible all-atom root-mean-square deviation (rmsd < 2.5 Å; Table E.12, Appendix E), data from one replica was discussed throughout the main text.

Analysis of the MD simulations was performed on frames spaced by 0.1 ns. A representative structure based on the lesion orientation and hydrogen bonding was chosen using the average-linkage clustering algorithm. All MD simulations were performed using AMBER 14 (6). The strengths of the interactions between the dNTP and lesion were calculated using frames spaced by 1 ns, B3LYP-D3(BJ)/6-311++G(d,p), hydrogen-capped nucleobases (hydrogen atoms optimized with M06-2X/6-31G(d)) and Gaussian 09 (revision D.01) (3). A smaller basis set was used for calculating the interaction energies relative to the basis set used in the nucleobase dimer calculations due to the number of structures considered. Nevertheless, the difference between nucleobase dimer interaction strengths calculated with B3LYP-D3(BJ)/6-311++G(2df,2p) and B3LYP-D3(BJ)/6-311++G(d,p) is ~1–8 kJ/mol (Table E.14, Appendix E).

References

1. Wilson KA, Szemethy KG, & Wetmore SD (2017) Conformational flexibility and base-pairing tendency of the tobacco carcinogen O6-[4-oxo-4-(3-pyridyl)butyl]guanine. *Biophys. Chem.* 228:25-37.

2. HyperChem ((Hypercube, Inc., Gainesville, Florida 32601, USA), Professional 7.5.

3. Frisch MJ*, et al.* (2013) Gaussian 09 (Gaussian, Inc., Wallingford CT), Revision D.01.

4. Antczak NM*, et al.* (2018) Characterization of nine cancer-associated variants in human DNA polymerase κ. *Chem. Res. Toxicol.*

5. Antczak NM, Packer MR, Lu X, Zhang K, & Beuning PJ (2017) Human y-family DNA polymerase κ is more tolerant to changes in its active site loop than its ortholog Escherichia coli dinb. *Chem. Res. Toxicol.* 30(11):2002-2012.

6. Case DA*, et al.* (2014) Amber 14 (University of California, San Francisco).

7. Wang J, Wang W, Kollman PA, & Case DA (2006) Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* 25:247–260.

8. Dupradeau F-Y*, et al.* (2010) The r.E.D. Tools: Advances in resp and esp charge derivation and force field library building. *Phys. Chem. Chem. Phys.* 12(28):7821-7839.

9. Vanquelef E*, et al.* (2011) R.E.D. Server: A web service for deriving resp and esp charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res.* 39(suppl 2):W511-W517.

10. Zhang L*, et al.* (2006) 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.* 34(11):3326-3337.

11. Perlow RA & Broyde S (2002) Toward understanding the mutagenicity of an environmental carcinogen: Structural insights into nucleotide incorporation preferences. *J. Mol. Biol.* 322(2):291-309.

12. Zhang L, Shapiro R, & Broyde S (2005) Molecular dynamics of a food carcinogen−DNA adduct in a replicative DNA polymerase suggest hindered nucleotide incorporation and extension. *Chem. Res. Toxicol.* 18(9):1347-1363.

13. Allnér O, Nilsson L, & Villa A (2012) Magnesium ion–water coordination and exchange in biomolecular simulations. *J. Chem. Theory Comput.* 8(4):1493-1502.

14. Joung IS & Cheatham TE (2008) Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* 112(30):9020-9041.

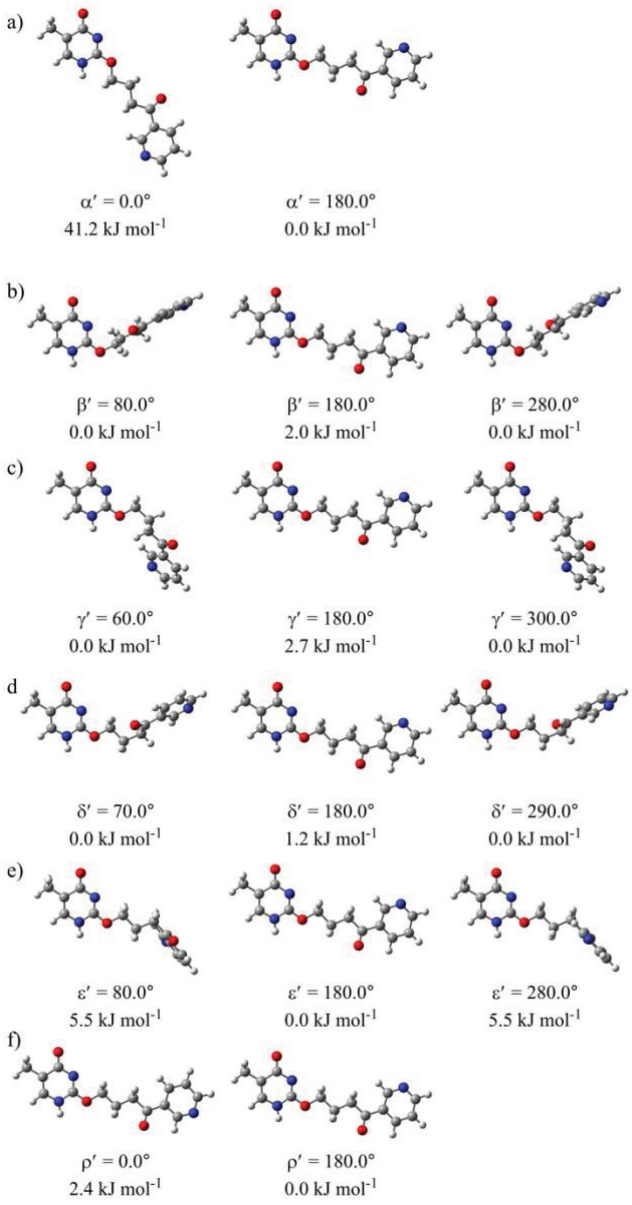

Figure E.1. B3LYP-D3(BJ)/6-31G(d) minimum energy structures for rotation about the a) α′, b) β′, c) γ′, d) δ′, e) ϵ′, and f) ρ′ dihedral angles initiated from a fully extended POB-T nucleobase model. See Figure 6.1 for definitions of the dihedral angles.

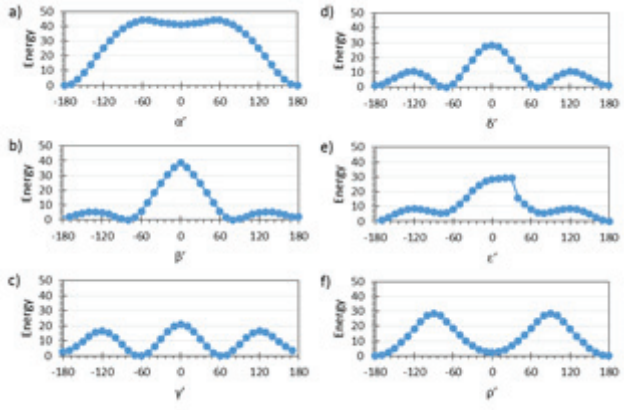

Figure E.2. B3LYP-D3(BJ)/6-31G(d) potential energy surfaces (kJ/mol) for rotation about important POB-T dihedral angles (deg.). See Figure 6.1 for definitions of dihedral angles.

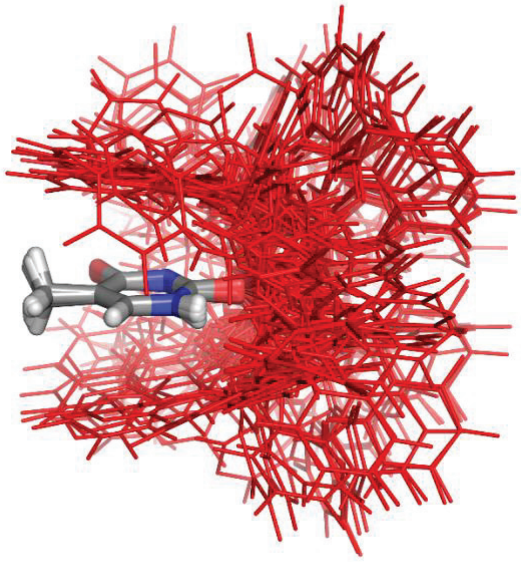

Figure E.3. Overlay of all unique conformers isolated in the conformational search (based on damaged T heavy atoms), with the bulky moiety in red and the T nucleobase carbon atoms in grey.

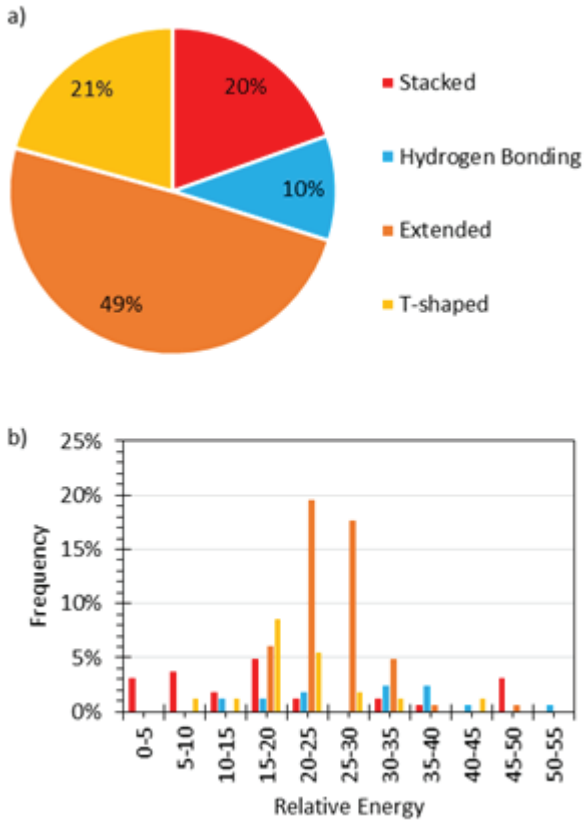

Figure E.4. a) Distribution of the POB-T nucleobase conformations resulting from the DFT conformational search and b) distribution of the B3LYP-D3/6-311++G(2df,2p) relative energies (kJ/mol) for each structural category. Energies are reported relative to the overall lowest energy structure.

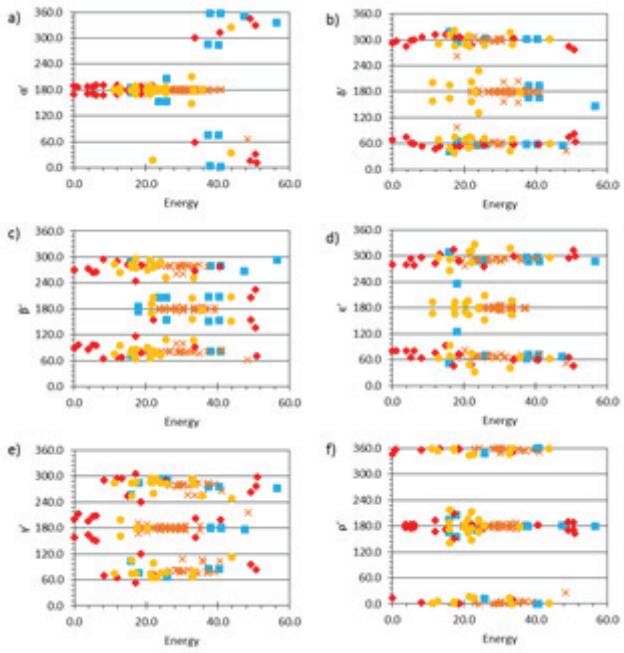

Figure E.5. Energetic effects (B3LYP-D3/6-311++G(2df,2p), kJ/mol) of different bulky moiety orientations in the isolated stacked (red), T-shaped (yellow), hydrogen bonded (blue), or extended (orange) structures from the conformational search.

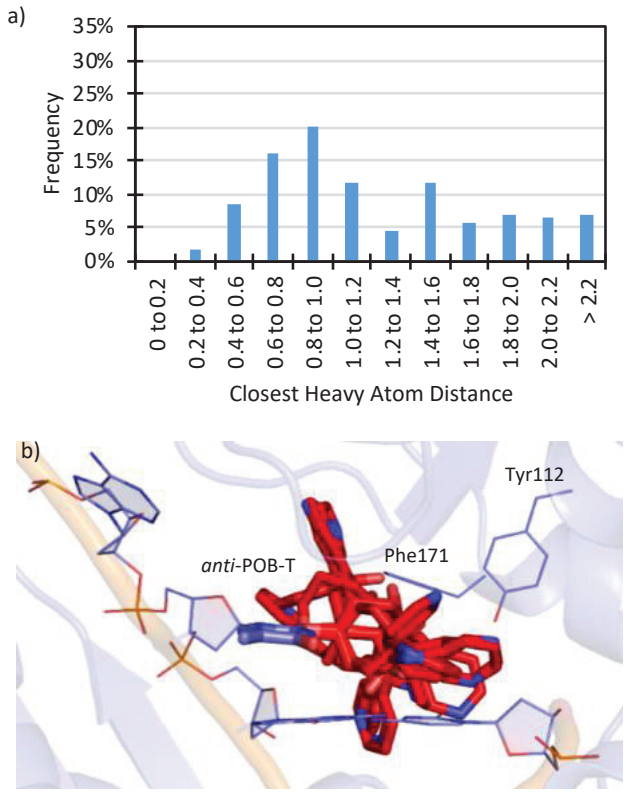

Figure E.6. a) Closest heavy atom distance and b) orientations of the POB-T bulky moiety from the DFT conformational search that fit in the pol κ (PDB ID: 5W2C) active site, with the lesion in the *anti* orientation.

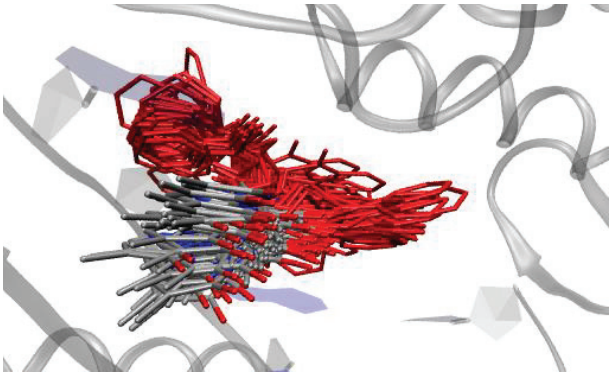

Figure E.7. Conformations of POB-T adopted throughout the MD simulation on the pre-insertion complex when *anti*-POB-T is bound in the pol κ active site.

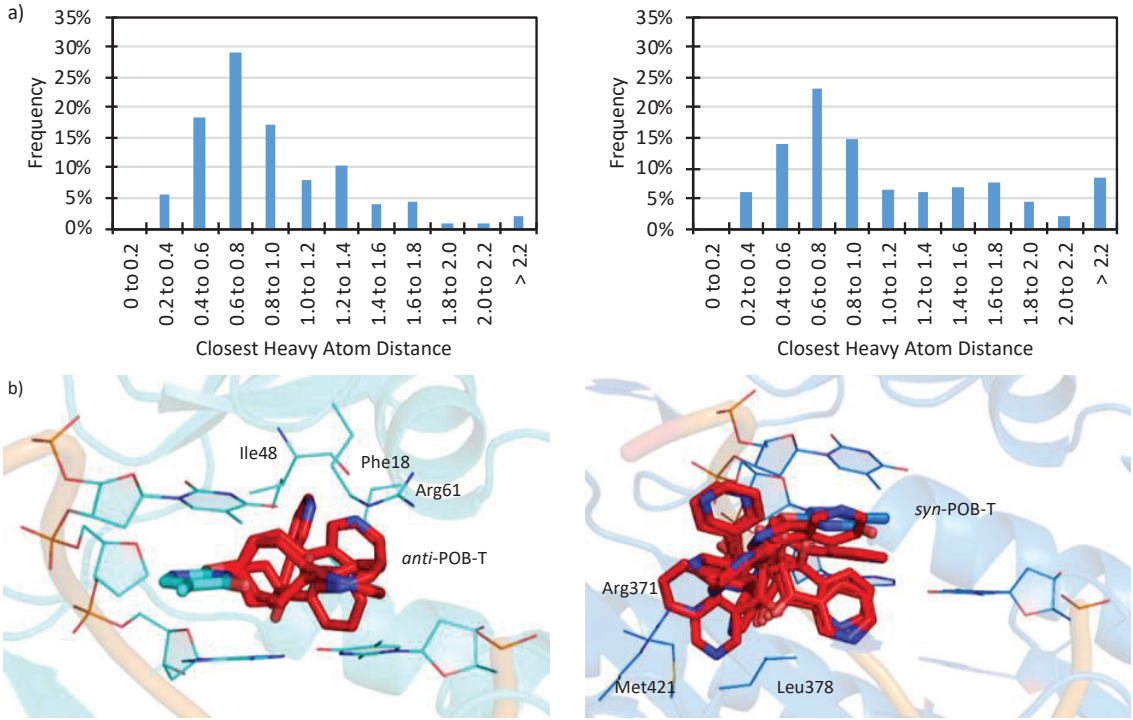

Figure E.8. a) Closest heavy atom distance and b) orientations of the POB-T bulky moiety from the DFT conformational search that fit in the pol η (PDB ID: 4ECS) active site, with the lesion in the *anti* orientation (left), or pol η active site (PDB ID: 4ECS), with the lesion in the *syn* orientation (right).

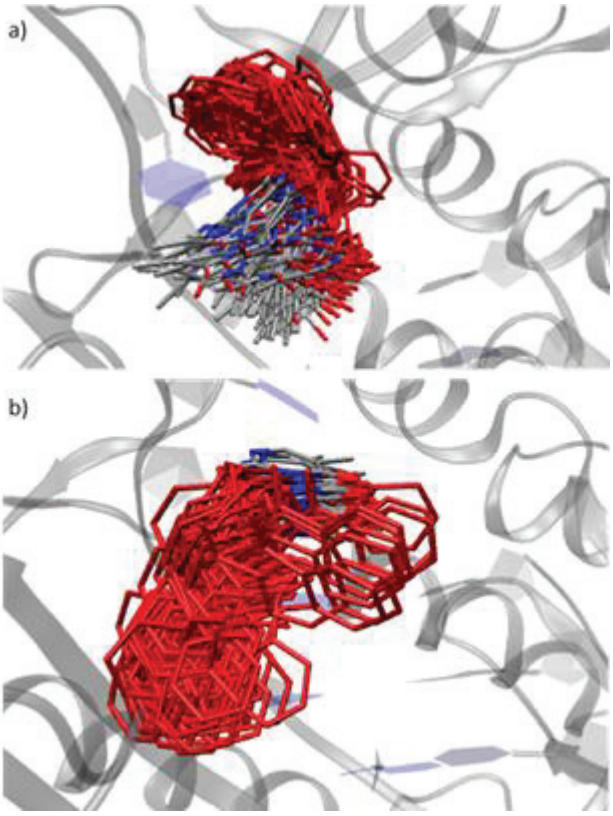

Figure E.9. Conformations of POB-T adopted throughout the MD simulation on the pre-insertion complex when a) *anti*-POB-T or b) *syn*-POB-T is bound in the pol η active site.

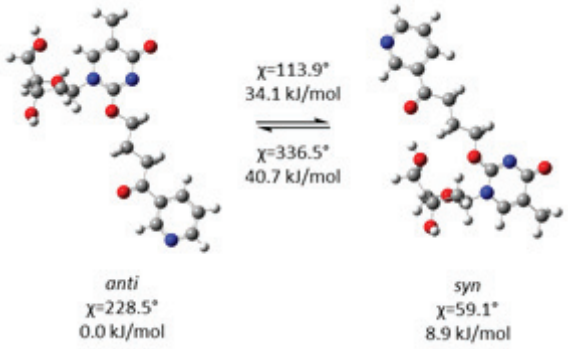

Figure E.10. B3LYP-D3/6-311++G(2df,2p)//B3LYP-D3(BJ)/6-31G(d) structures and relative energies (kJ/mol) for rotation about the POB-T glycosidic bond.

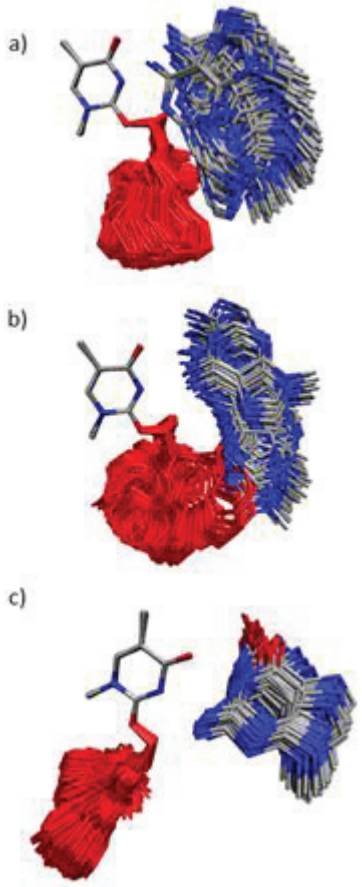

Figure E.11. Overlay (based on damaged T heavy atoms) of lesion conformations adopted throughout the MD simulation on the pol κ complex for the insertion of a) *anti*-dATP, b) *syn*-dATP or c) dGTP opposite POB-T, highlighting the deviation in the bulky moiety (red) and pairing base orientations.

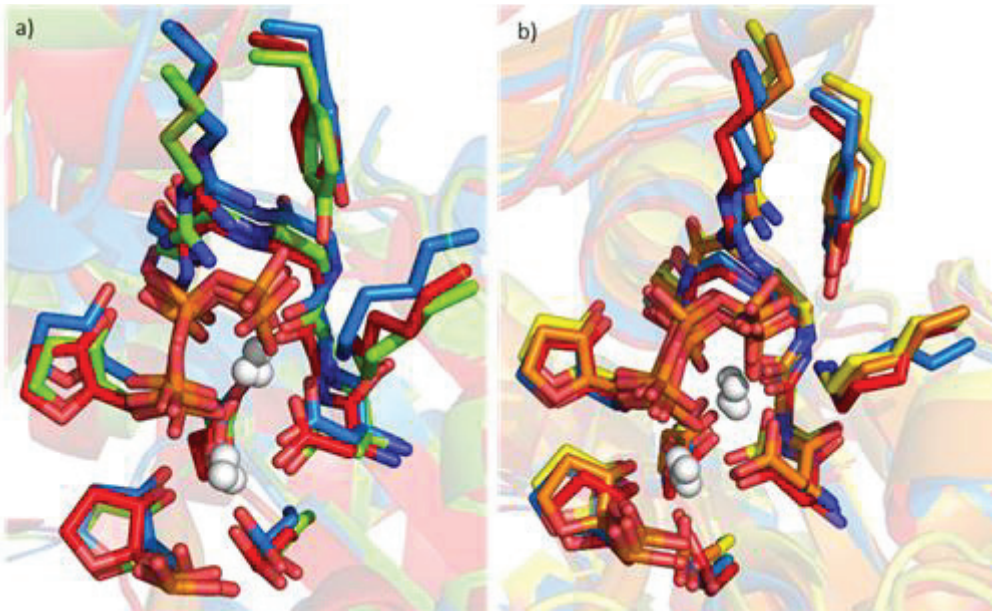

Figure E.12. a) Overlay (based on pol κ backbone) of *anti*-dATP (blue), *syn*-dATP (red) and dGTP (green) insertion complexes for the replication of POB-T by pol κ, highlighting the interactions between the enzyme and the dNTP backbone. b) Overlay (based on pol η backbone) of *anti*-dATP (blue), *syn*-dATP (red), dCTP (yellow), and dTTP (orange) insertion complexes for the replication of POB-T by pol η, highlighting the interactions between the enzyme and the dNTP backbone.

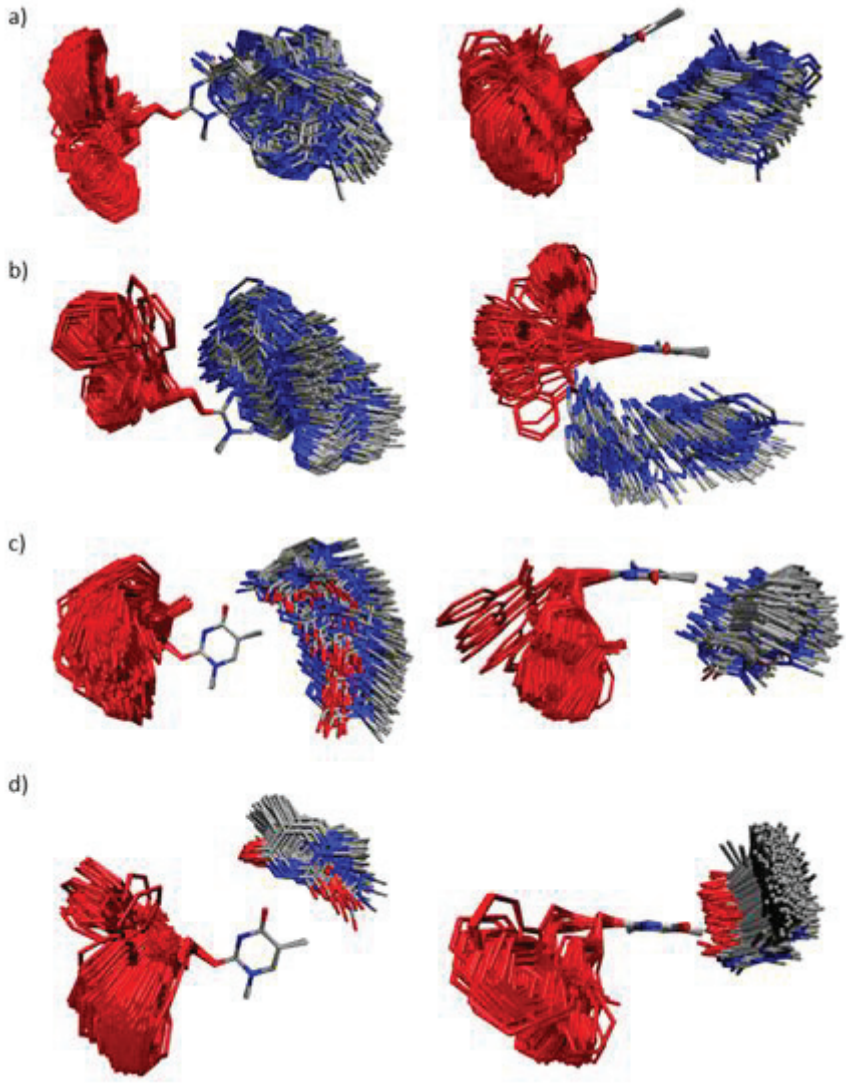

Figure E.13. Overlay (based on damaged T heavy atoms) of lesion conformations adopted throughout the MD simulation on the pol η complex for the insertion of a) *anti*-dATP, b) *syn*-dATP, c) dCTP, or d) dTTP opposite *syn*-POB-T, highlighting the deviation in the bulky moiety (red) and pairing base orientations.

Table E.1. Occupancies (%), average heavy atom distances (Å), and average angles (deg.) for active site hydrogen bonds in the pol κ complex for insertion of *anti*-dATP opposite POB-T.[a]

	Occupancy	Average Distance	Average Angle
3′-dG(O6)···dC(N4H)	93%	2.972	157.1
3′-dG(N1H)···dC(N3)	99%	2.991	161.6
3′-dG(N2H)···dC(O2)	100%	2.895	159.3
POB-T(N14)···Ser102(OγH)	27%	2.980	155.4
POB-T(O11)···Ser107(OγH)	29%	2.832	157.6
dATP(O3')···Tyr82(NH)	86%	3.152	166.2
dATP(Oβ1)···Phe81(NH)	95%	3.143	156.9
dATP(Oβ2)···Thr108(OG1)	100%	2.675	165.4
dATP(Oα)···Wat-Mg^{2+}(OH)	82%	2.702	145.6
dATP(Oβγ)···Arg114(NH2H)	57%	3.075	148.3
dATP(Oγ3)···Tyr111(OH)	42%	2.621	165.1
dATP(Oγ3)···Lys241(NZH)	85%	2.829	147.4
dATP(Oγ1)···Ala80(NH)	99%	2.943	164.8
dATP(Oγ1)···Arg114(NH2H)	88%	2.933	152.4
dATP(Oγ2)··· Arg114(NH1H)	99%	2.917	158.2

[a]Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of < 120°.

Table E.2. Occupancies (%), average heavy atom distances (Å), and average angles (deg.) for active site hydrogen bonds in the pol κ complex for insertion of *syn*-dATP opposite POB-T.[a]

	Occupancy	Average Distance	Average Angle
3′-dG(O6)···dC(N4H)	97%	2.961	160.8
3′-dG(N1H)···dC(N3)	99%	2.959	163.5
3′-dG(N2H)···dC(O2)	100%	2.869	161.3
POB-T(N14)···Ser102(OγH)	63%	2.939	157.4
dATP(O3')···Tyr82(NH)	78%	3.179	164.9
dATP(Oβ1)···Phe81(NH)	94%	3.161	156.6
dATP(Oβ2)···Thr108(OG1)	100%	2.669	165.7
dATP(Oα)···Wat-Mg^{2+}(OH)	80%	2.708	146.5
dATP(Oβγ)···Arg114(NH2H)	93%	2.969	158.2
dATP(Oγ3)···Tyr111(OH)	100%	2.618	167.2
dATP(Oγ3)···Lys241(NZH)	87%	2.815	144.9
dATP(Oγ1)···Ala80(NH)	100%	2.902	165.1
dATP(Oγ1)···Arg114(NH2H)	100%	2.914	158.5
dATP(Oγ2)··· Arg114(NH1H)	51%	3.118	142.0

[a]Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of < 120°.

Table E.3. Occupancies (%), average heavy atom distances (Å), and average angles (deg.) for active site hydrogen bonds in the pol κ complex for insertion of dGTP opposite POB-T.[a]

	Occupancy	Average Distance	Average Angle
3′-dG(O6)···dC(N4H)	99%	2.957	161.9
3′-dG(N1H)···dC(N3)	100%	2.980	161.4
3′-dG(N2H)···dC(O2)	99%	2.917	161.3
POB-T(O4)···GTP(N1H)	99%	2.891	158.6
POB-T(N3)···GTP(N2H)	92%	3.061	151.3
POB-T(N14)···Ser102(OγH)	13%	3.077	148.0
POB-T(O11)···Ser107(OγH)	10%	3.013	143.8
dCTP(O3')···Tyr82(NH)	62%	3.208	165.7
dCTP(Oβ1)···Phe81(NH)	100%	3.011	157.0
dCTP(Oβ2)···Thr108(OG1)	99%	2.730	162.1
dTTP(Oα)···Wat-Mg^{2+}(OH)	72%	2.795	148.1
dTTP(Oγ1)···Wat1(OH)	41%	2.618	165.1
dCTP(Oβγ)···Arg114(NH2H)	55%	3.020	143.4
dCTP(Oγ1)···Tyr111(OH)	59%	2.590	167.1
dCTP(Oγ1)···Lys241(NZH)	83%	2.987	151.7
dCTP(Oγ3)···Lys241(NZH)	77%	2.829	144.6
dCTP(Oγ1)···Ala80(NH)	37%	3.213	159.6
dCTP(Oγ2)··· Arg114(NH1H)	55%	2.877	163.3

[a]Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of < 120°.

Table E.4. Coordination of the active site Mg^{2+} ions across the simulation for the pol κ and η insertion complexes corresponding to various lesion replication outcomes.[a]

Enzyme	Base Pair	Mg^{2+} (binding)	Mg^{2+} (catalytic)
pol κ	T:dATP	6/6	6/6
	POB-T:*anti*-dATP	6/6	6/6
	POB-T:*syn*-dATP	6/6	6/6
	POB-T:dGTP	5/6 dGTP(Oα) 65%	6/6
pol η	T:dATP	6/6	6/6
	POB-T:*anti*-dATP	6/6	6/6
	POB-T:*syn*-dATP	6/6	6/6
	POB-T:dCTP	6/6	5/6 dCTP(Oα) 72%
	POB-T:dTTP	6/6	5/6 dTTP(Oα) 75%

[a]Atom is considered coordinated if the distance is < 2.5 Å for > 90% of the simulation.

Table E.5. Occupancies (%), average heavy atom distances (Å), and average angles (deg.) for active site hydrogen bonds in the pol κ complex for insertion of *anti*-dATP opposite T.[a]

	Occupancy	Average Distance	Average Angle
3′-dG(O6)···dC(N4H)	99%	2.944	163.3
3′-dG(N1H)···dC(N3)	100%	2.966	161.9
3′-dG(N2H)···dC(O2)	100%	2.864	160.9
dATP(O3')···Tyr82(NH)	93%	3.128	166.6
dATP(Oβ1)···Phe81(NH)	94%	3.154	155.7
dATP(Oβ2)···Thr108(OG1)	100%	2.670	166.1
dATP(Oα)···Wat-Mg^{2+}(OH)	86%	2.705	147.0
dATP(Oβγ)···Arg114(NH2H)	61%	2.971	158.5
dATP(Oγ3)···Tyr111(OH)	63%	2.686	164.1
dATP(Oγ3)···Lys241(NZH)	65%	2.880	145.6
dATP(Oγ1)···Ala80(NH)	99%	2.909	166.5
dATP(Oγ1)···Arg114(NH2H)	77%	3.014	148.4
dATP(Oγ2)··· Arg114(NH1H)	99%	2.926	160.0

[a]Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of < 120°.

Table E.6. Occupancies (%), average heavy atom distances (Å), and average angles (deg.) for active site hydrogen bonds in the pol η complex for insertion of *anti*-dATP opposite T.[a]

	Occupancy	Average Distance	Average Angle
3′-dT(O4)···dA(N6H)	96%	2.989	159.8
3′-dT(N3H)···dA(N1)	99%	2.964	163.3
dATP(N7)···Arg61(N1H)	38%	2.958	159.1
dATP(O3′)···Phe18(NH)	87%	3.144	167.5
dATP(Oβ2)···Phe17(NH)	90%	3.182	161.3
dATP(Oβ1)···Wat1(OH)	68%	2.659	163.6
Wat1(O)···Val50(NH)	65%	2.969	163.5
dATP(Oα1)···Wat-Mg^{2+}(OH)	46%	2.800	142.5
dATP(Oγ3)···Tyr52(OH)	99%	2.607	164.7
dATP(Oγ1)···Lys231(NZH)	66%	2.924	141.0
dATP(Oγ3)···Lys231(NZH)	73%	3.001	157.2
dATP(Oγ3)···Cys16(NH)	99%	2.928	164.6
dATP(Oβγ)···Arg55(NH2H)	98%	2.917	152.7
dATP(Oγ3)···Arg55(NH1H)	92%	2.920	158.9

[a]Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of < 120°.

Table E.7. Occupancies (%), average heavy atom distances (Å), and average angles (deg.) for active site hydrogen bonds in the pol η complex for insertion of *anti*-dATP opposite *syn*-POB-T.[a]

	Occupancy	Average Distance	Average Angle
3′-T(O4)···A(N6H)	93%	2.983	153.6
3′-T(N3H)···A(N1)	100%	2.924	163.6
Ser116(Oγ)···3′-T(O3′H)	86%	2.944	155.0
dATP(N7)···Arg61(N1H)	90%	2.949	157.3
dATP(O3′)···Phe18(NH)	88%	3.134	167.3
dATP(Oβ2)···Phe17(NH)	94%	3.140	161.2
dATP(N1)···Wat1(OH)	21%	2.977	150.9
dATP(Oβ1)···Wat2(OH)	45%	2.662	163.7
Wat2(O)···Val50(NH)	43%	3.013	162.3
3′-T(OP)···Wat-Mg^{2+}(OH)	26%	2.930	156.2
dATP(Oα1)···Wat-Mg^{2+}(OH)	51%	2.781	144.0
dATP(Oγ3)···Tyr52(OH)	100%	2.618	166.5
dATP(Oγ1)···Lys231(NZH)	79%	2.898	142.5
dATP(Oγ3)···Lys231(NZH)	70%	3.054	155.6
dATP(Oγ3)···Cys16(NH)	99%	2.989	163.6
dATP(Oβγ)···Arg55(NH2H)	99%	2.919	152.7
dATP(Oγ3)···Arg55(NH1H)	97%	2.869	161.1

[a]Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of < 120°.

Table E.8. Occupancies (%), average heavy atom distances (Å), and average angles (deg.) for active site hydrogen bonds in the pol η complex for insertion of *syn*-dATP opposite *syn*-POB-T.[a]

	Occupancy	Average Distance	Average Angle
3′-T(O4)···A(N6H)	90%	3.028	151.3
3′-T(N3H)···A(N1)	100%	2.940	162.9
Ser116(Oγ)···3′-T(O3′H)	31%	2.997	152.5
dATP(Oα)···Arg61(N1H)	38%	2.894	155.0
dATP(O3′)···Phe18(NH)	79%	3.162	166.9
dATP(Oβ2)···Phe17(NH)	92%	3.176	161.0
dATP(Oα1)···Wat-Mg^{2+}(OH)	29%	2.818	141.9
dATP(Oγ3)···Tyr52(OH)	100%	2.603	166.8
dATP(Oγ1)···Lys231(NZH)	82%	2.890	144.5
dATP(Oγ3)···Lys231(NZH)	54%	3.085	155.8
dATP(Oγ3)···Cys16(NH)	99%	2.918	164.6
dATP(Oβγ)···Arg55(NH2H)	93%	2.911	160.0
dATP(Oγ2)···Arg55(NH1H)	97%	2.949	150.0
dATP(Oγ3)···Arg55(NH1H)	48%	3.032	149.1

[a]Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of < 120°.

Table E.9. Occupancies (%), average heavy atom distances (Å), and average angles (deg.) for active site hydrogen bonds in the pol η complex for insertion of dCTP opposite *syn*-POB-T.[a]

	Occupancy	Average Distance	Average Angle
3′-T(O4)···A(N6H)	88%	3.015	157.8
3′-T(N3H)···A(N1)	99%	2.960	158.7
dCTP(O3')···Phe18(NH)	94%	3.072	161.2
Ser113(Oγ)···3′-dT(O3′H)	90%	2.869	158.6
dCTP(Oα1)···Arg61(NH1H)	78%	2.837	154.5
dCTP(Oβ2)···Phe17(NH)	76%	3.224	160.4
dCTP(Oβ1)···Wat1(OH)	87%	2.676	163.8
Wat1(O)···Val50(NH)	87%	2.927	163.1
dCTP(Oβ1)···Wat2(OH)	53%	2.866	157.1
Wat2(O)···Arg61(NH1H)	54%	2.891	161.6
3′-T(OP)···Wat-Mg^{2+}(OH)	10%	2.967	148.8
dCTP(Oβγ)···Arg55(NH2H)	98%	2.930	149.7
dCTP(Oγ3)···Tyr52(OH)	100%	2.573	167.0
dCTP(Oγ3)···Lys231(NZH)	44%	3.079	157.4
dCTP(Oγ1)···Lys231(NZH)	81%	2.900	146.4
dCTP(Oγ3)···Cys16(NH)	99%	2.933	165.7
dCTP(Oγ3)···Arg55(NH2H)	52%	3.056	142.7
dCTP(Oγ3)···Arg55(NH1H)	72%	3.011	146.3
dCTP(Oγ2)···Arg55(NH1H)	69%	2.986	157.9

[a]Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of < 120°.

Table E.10. Occupancies (%), average heavy atom distances (Å), and average angles (deg.) for active site hydrogen bonds in the pol η complex for insertion of dTTP opposite *syn*-POB-T.[a]

	Occupancy	Average Distance	Average Angle
3′-T(O4)···A(N6H)	98%	2.941	159.2
3′-T(N3H)···A(N1)	98%	3.036	161.6
dTTP(O3')···Phe18(NH)	89%	3.086	163.3
Ser113(Oγ)···3′-dT(O3′H)	39%	2.970	152.6
dTTP(Oα1)···Arg61(NH1H)	14%	2.946	152.8
dTTP(O4)···Arg61(NH1H)	39%	2.827	156.4
dTTP(Oβ2)···Phe17(NH)	87%	3.177	161.2
dTTP(O4)···Wat1(OH)	33%	2.827	155.5
POB-T(O4)···Wat1(OH)	16%	2.935	140.7
dTTP(Oβ1)···Wat2(OH)	74%	2.682	161.9
Wat2(O)···Val50(NH)	36%	2.973	163.0
dTTP(Oβ1)···Wat3(OH)	37%	2.678	161.7
Wat3(O)···Val50(NH)	37%	2.960	163.5
dTTP(Oγ3)···Wat4(OH)	17%	2.625	164.4
Wat4(O)···Tyr52(OH)	8%	2.722	165.0
dTTP(Oα)···Wat-Mg^{2+}(OH)	31%	2.838	140.6
dTTP(Oβγ)···Arg55(NH2H)	98%	2.883	155.3
dTTP(Oγ3)···Tyr52(OH)	56%	2.582	166.7
dTTP(Oγ3)···Lys231(NZH)	68%	2.884	156.8
dTTP(Oγ1)···Lys231(NZH)	22%	2.984	138.8
dTTP(Oγ3)···Cys16(NH)	92%	3.033	164.5
dTTP(Oγ3)···Arg55(NH2H)	46%	3.041	145.3
dTTP(Oγ3)···Arg55(NH1H)	55%	3.006	149.6
dTTP(Oγ2)···Arg55(NH1H)	73%	2.995	155.5

[a]Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of < 120°.

Table E.11. Occupancies (%), average heavy atom distances (Å), and average angles (deg.) for hydrogen bonds in the adducted, and 3′ and 5′ flanking, base pairs across the entire MD simulation trajectory for POB-T adducted DNA with the lesion paired opposite T.

	Occupancy	Average Distance	Average Angle
5′-C(O2)···G(N2H)	99%	2.871	163.0
5′-C(N3)···G(N1)	99%	2.963	164.8
5′-C(N4H)···G(O6)	97%	2.940	160.3
3′-C(O2)···G(N2H)	99%	2.885	164.1
3′-C(N3)···G(N1)	99%	2.960	165.3
3′-C(N4H) ··G(O6)	97%	2.918	160.9
POB-T(O4) ··T(N3H)	92%	2.896	161.6
T(O2) ··POB-T(C7H)	14%	3.277	127.4

[a]Hydrogen-bonding occupancies are based on a distance cutoff of < 3.4 Å and an angle cutoff of < 120°.

Table E.12. Heavy atom rmsd (Å) with respect to the representative structure for Trial 1.

		Trial 1		Trial 2		Trial 3	
		Entire System	Active Site	Entire System	Active Site	Entire System	Active Site
pol κ	*anti*-POB-T[a]	1.936±0.332	1.346±0.295	2.289±0.194	1.770±0.214	2.373±0.239	1.578±0.132
	anti-POB-T:*anti*-dATP[a]	2.304±0.359	1.390±0.538	2.178±0.147	0.863±0.123	2.398±0.307	1.001±0.177
	anti-POB-T:*syn*-dATP[a]	1.864±0.372	0.670±0.161	2.404±0.206	0.863±0.140	2.816±0.365	0.832±0.104
	anti-POB-T:*anti*-dGTP[a]	1.834±0.186	1.050±0.452	2.164±0.142	1.119±0.157	2.119±0.234	0.853±0.123
pol η	*anti*-POB-T[a]	2.189±0.273	1.231±0.319	2.472±0.312	1.432±0.209	2.157±0.132	1.526±0.164
	syn-POB-T[a]	2.075±0.244	1.488±0.262	2.751±0.436	1.647±0.140	2.401±0.126	1.657±0.193
	anti-POB-T:*anti*-dATP[a]	1.910±0.355	0.975±0.141	2.987±0.199	1.298±0.149	2.535±0.146	1.132±0.136
	syn-POB-T:*anti*-dATP[a]	1.948±0.285	0.700±0.201	2.187±0.160	1.230±0.129	2.712±0.194	1.542±0.106
	syn-POB-T:*syn*-dATP[a]	2.109±0.356	0.959±0.263	2.424±0.317	1.439±0.354	2.562±0.158	1.513±0.118
	syn-POB-T:*anti*-dCTP[a]	2.169±0.244	1.015±0.140	2.480±0.103	0.961±0.121	2.632±0.201	1.262±0.254
	syn-POB-T:*anti*-dTTP[a]	1.865±0.204	1.020±0.340	2.401±0.198	0.910±0.139	2.920±0.220	1.345±0.328
DNA	*syn*-POB-T:*anti*-dTTP[a]	2.073±0.378	1.249±0.249	1.999±0.399	1.216±0.251	2.093±0.427	1.447±0.339

[a]100 ns pol η or κ insertion complex or adducted DNA simulation.

Table E.13. Atom types and charges for POB-T.

Atom Name	Atom Type	Charge
P	P	1.1623
O5'	OS	-0.4914
OP1	O2	-0.7616
OP2	O2	-0.7706
O3'	OS	-0.5532
C5'	CT	0.0105
H5'1	H1	0.0681
H5'2	H1	0.0681
C4'	CT	0.1734
H4'	H1	0.0965
O4'	OS	-0.3982
C1'	CT	0.2113
H1'	H2	0.0585
C3'	CT	0.1826
H3'	H1	0.0609
C2'	CT	-0.0839
H2'1	HC	0.0486
H2'2	HC	0.0486
N1	N*	-0.1376
C6	CM	-0.2167
H6	H4	0.2227
C5	CM	-0.0481
C7	CT	-0.2565
H71	HC	0.0828
H72	HC	0.0828
H73	HC	0.0828
C4	C	0.7946
O4	O	-0.609
N3	NC	-0.7289
C2	CA	0.6597
O2	OS	-0.2901
C8	CT	0.0419
H81	H1	0.0654
H82	H1	0.0654
C9	CT	0.0684
H91	HC	0.0007
H92	HC	0.0007
C10	CT	0.0492
H101	HC	-0.0082
H102	HC	-0.0082
C11	C	0.4364
O11	O	-0.4909
C12	CA	-0.1437
C13	CA	0.2872
H13	H4	0.0733
C17	CA	-0.0146
H17	HA	0.1505
C16	CA	-0.301
H11	HA	0.1509
C15	CA	0.3295
H15	H4	0.0643
N14	NC	-0.5864

Table E.14. Comparison of B3LYP-D3(BJ) binding energies (kJ/mol) calculated with 6-311++G(2df,2p) and 6-311++G(d,p) for base pairs between the Watson-Crick (WC) face or Hoogsteen (H) face of POB-T and the canonical nucleobases.[a]

	B3LYP-D3(BJ)/ 6-311++G(2df,2p)	B3LYP-D3(BJ)/ 6-311++G(d,p)	Deviation
WC-POB-T:WC-A	–71.3	–78.9	7.6
WC-POB-T:H-A	–68.8	–75.8	7.0
WC-POB-T:WC-T	–70.6	–75.5	4.9
WC-POB-T:WC-C	–80.7	–86.5	5.8
WC-POB-T:WC-G	–93.6	–96.8	3.2
H-POB-T:WC-A	–36.5	–37.8	1.3
H-POB-T:H-A	–35.3	–36.5	1.2
H-POB-T:WC-T	–35.9	–36.9	1.0
H-POB-T:WC-C	–47.7	–49.7	2.0

[a]Single-point calculations were performed on M06-2X/6-31G(d) optimized geometries.

Appendix F

Supplementary Information for Chapter 7: Exploring the Identity of the General Base for a DNA Polymerase Catalyzed Reaction Using QM/MM: The Case Study of Human TLS Polymerase η

Contains Figures F.1–F.8.

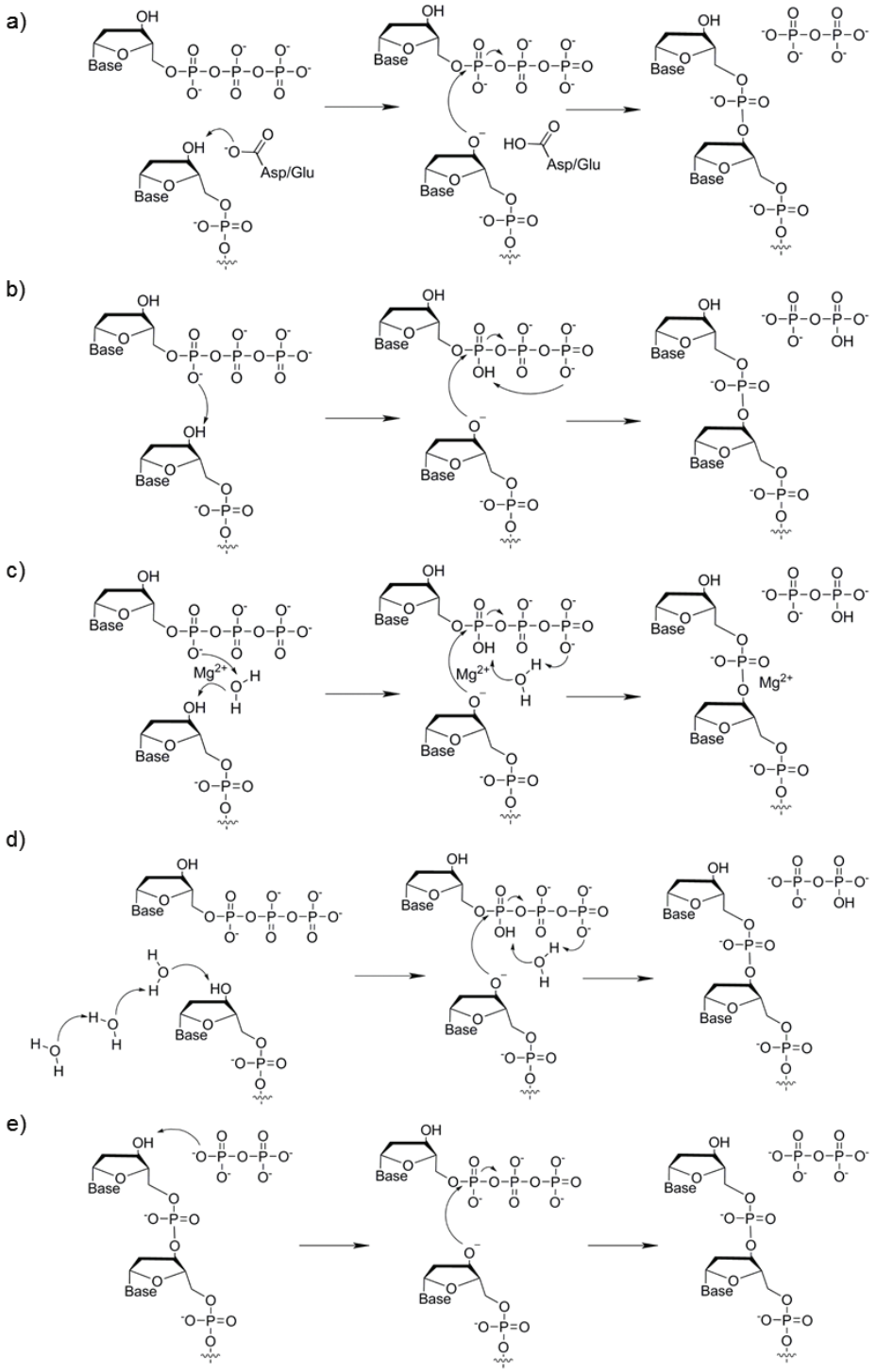

Figure F.1. Examples of the proposed reaction mechanisms for the nucleotide addition by a DNA polymerase using a) a conserved Asp/Glu, b) the α-phosphate of the dNTP, c) water coordinated to the catalytic Mg^{2+} ion (WMSA mechanism), d) bulk solvent, or e) pyrophosphate (SAM) acting as the general base for proton abstraction.

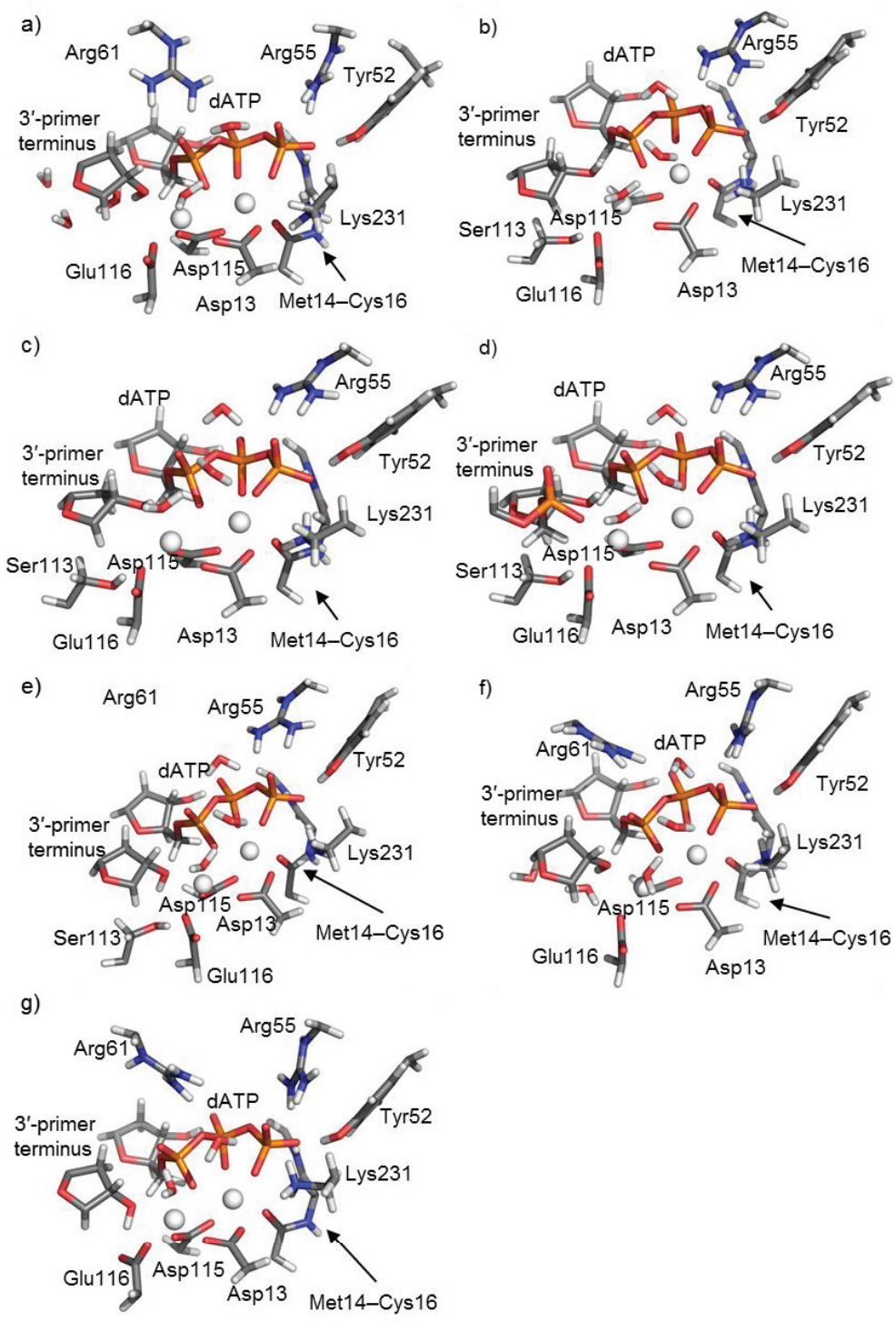

Figure F.2. High-level QM region used in ONIOM calculations for the proton abstraction assisted by a) bulk solvent, b) the α-phosphate of the dNTP, c) water coordinated to the catalytic Mg^{2+}, d) water coordinated to the catalytic Mg^{2+} with the 5′-phospate group in the QM region, e) Glu116 assisted by Ser113, f) Glu116 assisted by water in the Ser113Ala mutant, or g) Glu116 directly.

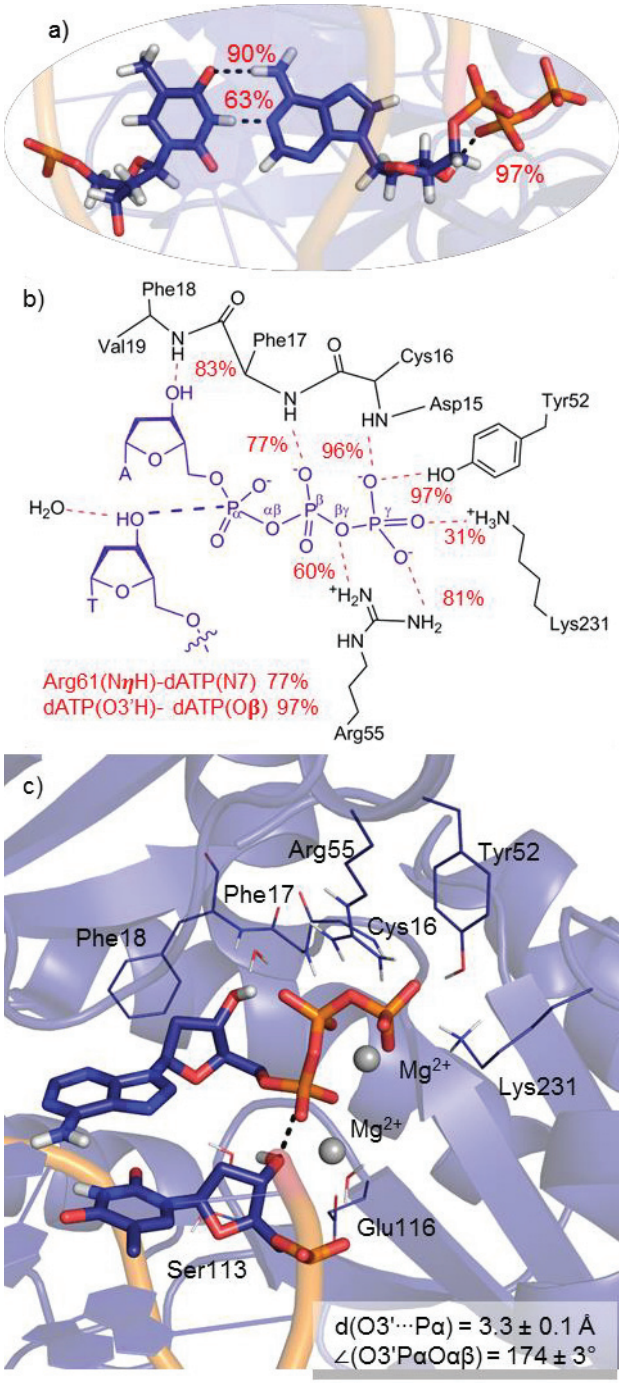

Figure F.3. Hydrogen bonding between the incoming dATP and a) the pairing dT or b) the surrounding enzyme, as well as c) the average reaction distance and angle for the nucleotidyl transfer reaction, derived from a classical MD simulation on the reactant complex.

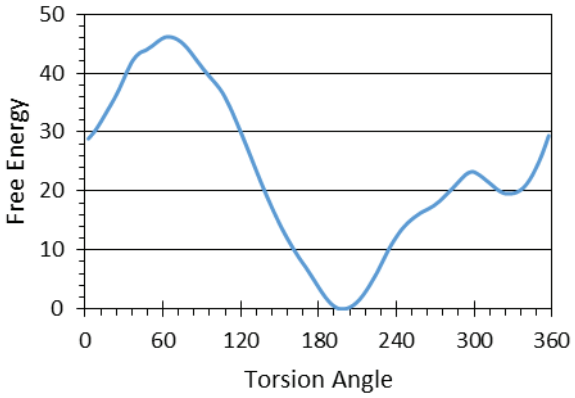

Figure F.4. Free energy profile (kJ/mol) of the ∠(H3′–C3′–O3–HO3′) torsion angle (degrees) of the 3′-primer terminus.

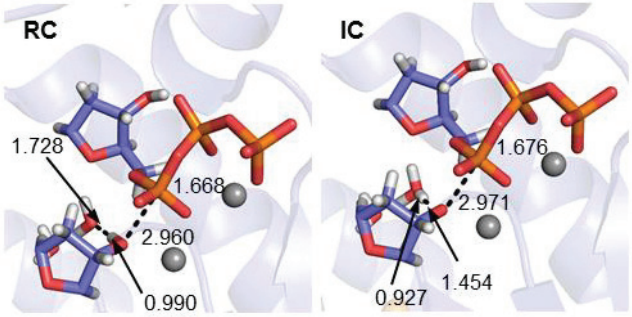

Figure F.5. ONIOM unconstrained reactant (RC) and constrained intermediate (IC) complexes for the polymerase η nucleotidyl transfer reaction involving bulk water as the general base.

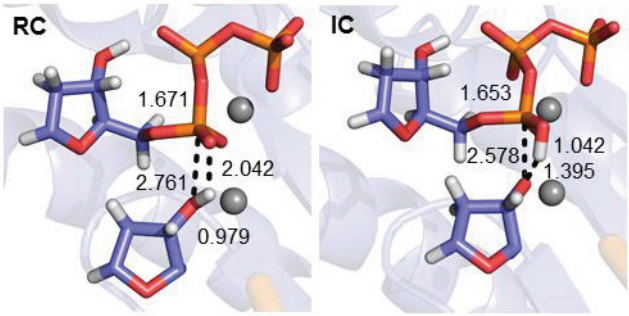

Figure F.6. ONIOM unconstrained reactant (RC) and constrained intermediate (IC) complexes for the polymerase η nucleotidyl transfer reaction involving the dATP α-phosphate as the general base.

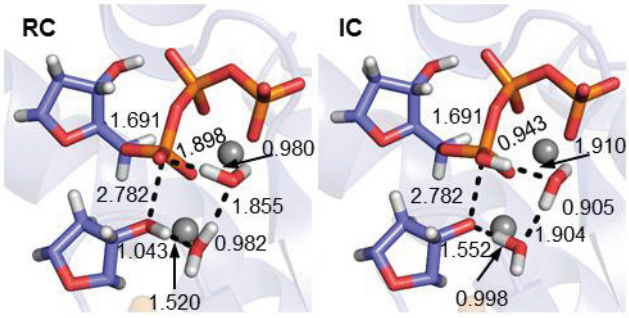

Figure F.7. ONIOM unconstrained reactant (RC) and constrained intermediate (IC) complexes for the polymerase η nucleotidyl transfer reaction involving the Mg^{2+} coordinated water in combination with the dATP α-phosphate as the general base.

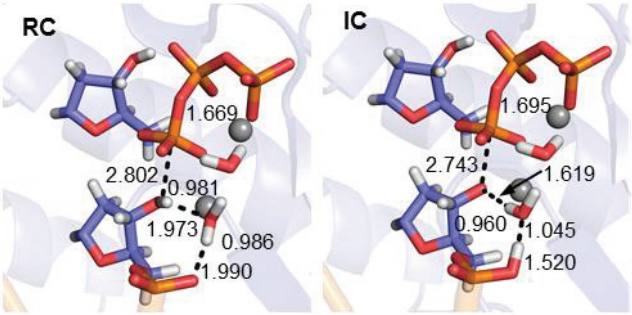

Figure F.8. ONIOM unconstrained reactant (RC) and constrained intermediate (IC) complexes for the polymerase η nucleotidyl transfer reaction involving the Mg^{2+} coordinated water in combination with the 3′-primer terminus 5′-phospate moiety as the general base.