INVESTIGATING THE CATALYTIC MECHANISMS OF DNA REPAIR ENZYMES USING CLASSICAL AND QUANTUM MECHANICAL COMPUTATIONAL TECHNIQUES

ANGELA IVY FREDERICKSON

Bachelor of Science, University of Lethbridge, 2023

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

MASTER OF SCIENCE

in

CHEMISTRY

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

INVESTIGATING THE CATALYTIC MECHANISMS OF DNA REPAIR ENZYMES USING CLASSICAL AND QUANTUM MECHANICAL COMPUTATIONAL TECHNIQUES

ANGELA IVY FREDERICKSON

Date of Defence: April 21st, 2026

Dr. Stacey Wetmore	Professor	Ph.D.
Supervisor		
Dr. Trushar Patel	Professor	Ph.D.
Thesis Examination Committee Member		
Dr. Paul Hayes	Professor	Ph.D.
Thesis Examination Committee Member		
Dr. Jean-Denys Hamel	Assistant	Ph.D.
Chair, Thesis Examination Committee	Professor	

DEDICATION

To the people, playlists, and takeout meals that carried me through — and to the person I became along the way.

"Still I am learning."— Michelangelo

ABSTRACT

DNA is continuously subjected to chemical damage arising from endogenous metabolic processes and environmental stressors. If left unrepaired, the resulting DNA lesions can disrupt genome stability and interfere with essential cellular processes. In some cases, damage occurs to bases that also serve regulatory roles, meaning that repair must restore not only the chemical structure of DNA, but also the epigenetic information encoded within the genome. To counteract these threats, cells employ a diverse network of DNA repair enzymes that recognize and process chemically-modified bases. Defects in DNA repair pathways have been linked to mutagenesis, disease development, and altered cellular function. Despite their biological importance, the catalytic mechanisms of many DNA repair enzymes remain poorly understood, with multiple and sometimes conflicting mechanistic proposals present in the literature. In this thesis, computational methodologies including molecular dynamics (MD) simulations and quantum mechanics/molecular mechanics (QM/MM) calculations were used to clarify and identify atomic-level details of the catalytic mechanisms involved in glycosidic bond cleavage and oxidative DNA repair. Specifically, the monofunctional glycosylase UdgX was investigated, focusing on its unique suicide inactivation mechanism involving the formation of a covalent DNA–protein crosslinked intermediate during uracil excision. Mutational variants were also investigated to determine how changes to the active site environment shift the catalytic pathway from crosslink to hydrolysis. To complement this work, monofunctional alkyladenine glycosylase (AAG) was also studied to better understand how lesion identity influences catalytic competency and determines when repair is supported or inhibited despite the enzyme's broad substrate specificity. In particular, the understudied etheno lesion 3,N^4-etheno-5-methylcytosine (ε5mC) was examined alongside the canonical substrates 3,N^4-

ethenocytosine (εC) and 1,N^6-ethenoadenine (εA). Finally, the oxidative repair enzyme ALKBH2 was investigated to characterize how etheno lesions, including εC and ε5mC, are recognized and repaired through direct reversal repair, and to determine whether chemical modifications to these lesions influences the catalytic mechanism and repair outcome. Together this work provides a comprehensive mechanistic framework for understanding how DNA repair enzyme's function and reveals how subtle variations in active site architecture and solvation dictate pathway selection at the molecular level.

PREFACE

Some of the research in this thesis includes contributions from collaborators. In chapter 2, Dylan Nikkel completed parameter set up of the iron-sulfur cluster for UdgX models and provided guidance during model building. I performed molecular dynamics simulations for all UdgX systems, followed by data/results analysis. I generated representative molecular dynamics structures that were used as starting points for quantum mechanics/molecular mechanics calculations of the crosslink and hydrolysis mechanisms, which were then mapped by Basel Mansour. Additionally, Basel Mansour assisted in drafting the quantum mechanics/molecular mechanics calculations methods section. In chapter 3, I was responsible for all aspects of model building and parameterization of non-standard nucleobases, as well as conducting molecular dynamics simulations and performing data/results analysis. For all projects, I was involved in project conceptualization, model construction, data collection and analysis, and writing (original draft, review, and editing). Stacey D. Wetmore was involved in project conceptualization and administration, supervision, data visualization and interpretation, funding acquisition, computational resource acquisition, and review/editing.

ACKNOWLEDGEMENTS

I would first like to express my sincere gratitude to my supervisor, Dr. Stacey Wetmore, for your guidance, support, and mentorship throughout my graduate studies. Your insight, encouragement, and commitment to scientific rigor have profoundly shaped my development as a researcher. I am especially grateful for the opportunities you provided, your patience during challenging stages of this work, and your continuous belief in my ability to grow and succeed. I would also like to thank my committee members, Dr. Trushar Patel and Dr. Paul Hayes, for their support throughout both my undergraduate and graduate studies, and for their thoughtful feedback, questions, and insight over the years. I am also grateful to Dr. Jean-Denys Hamel for chairing my defence.

I am so grateful to my fellow members of the Wetmore lab, past, present, and new, for creating a collaborative, fun, and supportive environment (Mark Lea, Ridwan Tajudeen, Rebecca Jeong, Indu Negi, Makay Murray, Austin Pounder, Briana Boychuk, Clara Cunha, Anita Basu, Umer Yaqoob, Kari Gaalswyk, Hossein Boroujeni, Parth Rathee). I would especially like to thank Rajwinder Kaur for overseeing my first computational chemistry project. A special acknowledgement goes to Dylan Nikkel, and Basel Mansour whose collaboration, guidance, and willingness to share their knowledge greatly influenced my development as a computational chemist. To the dear friends I met during graduate school, Mark, Rebecca, Indu, Makay, and Navya Sehgal, thank you for your kindness, advice, and unwavering support. The friendships formed during this time will always be meaningful to me.

I am also thankful for the opportunity to serve as a teaching assistant in the Department of Chemistry and Biochemistry. This experience reaffirmed my love for chemistry, as it

allowed me to connect with students, share my enthusiasm for the subject, and contribute to making chemistry more accessible and enjoyable. I am grateful to my teaching coordinators at the University of Lethbridge, Wayne Lippa, John Eng, Dr. Greg Patenaude, and Dr. Ying Zheng, for their guidance, mentorship, and support throughout my teaching experience.

Finally, I would like to thank my family and loved ones. To my Omi, for your unwavering love, your gentle spirit, and for always knowing how to make everything feel okay. To my better half, Erick, thank you for being my greatest supporter, and for your patience, love, and reassurance throughout this journey. To my sister, Carley, thank you for the endless encouragement, for celebrating every milestone with me, and for always keeping me grounded. Thank you, Zeus, for all the puppy love on the hard days. To my parents, thank you for always believing in me and for supporting my educational journey every step of the way. Dad, you instilled in me perseverance, resilience, and a strong work ethic. Mom, your constant love and encouragement gave me the strength and confidence to overcome every challenge. This achievement would not have been possible without all of you.

USE OF GENERATIVE AI

I attest that no generative AI was used for any part of the following thesis, from the research proposals, script writing, data assembly and analysis, figure creation, writing, or editing.

TABLE OF CONTENTS

DEDICATION ..... **iii**

ABSTRACT ..... **iv**

PREFACE ..... **vi**

ACKNOWLEDGEMENTS ..... **vii**

USE OF GENERATIVE AI ..... **ix**

TABLE OF CONTENTS ..... **x**

LIST OF TABLES ..... **xiv**

LIST OF FIGURES ..... **xv**

LIST OF ABBREVIATIONS ..... **xix**

Chapter 1 : Introduction **1**

1.1. General Overview 1

1.2. Nucleic Acid Structure and Function 2

1.3. DNA Modifications 4

1.4. Enzymes That Process Nucleic Acids 9

1.5. Base Excision Repair and Glycosylases 10

1.6. Direct Reversal Repair and AlkB Family Enzymes 14

1.7. Implications of Enzyme Misfunction for Human Health 17

1.8. Computational Approaches Used to Study the Chemistry of Enzyme–Substrate Complexes 19

1.9. Thesis Overview 20

1.10. References 21

Chapter 2 : A Computational Investigation of the Uracil DNA Glycosylase *Msm*UdgX Catalytic Mechanism of Action Using MD and QM/MM Techniques27

2.1. Introduction 27

2.2. Computational Methods 34

2.2.1. MD Simulation Protocol 34

2.2.2. MD Simulation Analysis 35

2.2.3. QM/MM Calculations 36

2.3. Results and Discussion 38

2.3.1. MD Simulations Reveal Active Site Alignment for Crosslink Formation in the HID, but Not HIE, H109 Tautomer 38

2.3.1.1. HIE 38

2.3.1.2. HID 43

2.3.2. QM/MM of HID109 Wild-Type Displays Concerted Mechanism of Action for Irreversible Crosslink Formation 47

2.3.3. MD Simulations of UdgX Mutants Reveal Active Site Architecture that Points to a Hydrolysis Pathway Selection 52

2.3.4. QM/MM Comparison Shows Feasibility of Hydrolysis for Mutants with Possibility of Crosslinking for H109E 59

2.3.4.1. H109A Displays Energetically Feasible Hydrolysis Mechanism 59

2.3.4.2. H109E Displays an Energetically Feasible Crosslink Mechanism 64

2.4. Conclusions 69

2.5. References 71

Chapter 3 : A Computational Investigation of the Substrate Specificity in Alkyladenine DNA Glycosylase and ALKBH2 Towards 3,N^4–Etheno-5-methylcytosine 75

3.1. Introduction 75

3.2. Computational Methods 80

3.2.1. MD Simulation Protocol 80

3.2.2. MD Simulation Analysis 84

3.3. Results and Discussion 85

3.3.1. AAG Dynamics and Nucleotide-Dependent Active Site Organization 85

3.3.1.1. AAG Active Site Conformations Display Subtle Differences Across Repair and Inhibitory Nucleotides 85

3.3.1.2. Aromatic Stacking Interactions and the E125–Y127 Catalytic Dyad Geometry are Preserved Despite Differences in Nucleotide Positioning 89

3.3.1.3. Nucleotide-Dependent Hydrogen-Bonding to H136 and N169 Impacts the Insertion of εC and ε5mC Inhibitors 92

3.3.1.4. εC and ε5mC Disrupt Active Site Solvation and Catalytic Water Positioning 95

3.3.2. ALKBH2 98

3.3.2.1. π-Stacking and S125 Interaction with 5′-Phosphate Backbone Promotes an Alignment that is Catalytically Conducive in ALKBH2 101

3.3.2.2. Fe^{IV}–oxo Octahedral Coordination is Conserved Across εC and ε5mC Complexes 103

3.3.2.3. Nucleotide-Dependent Hydrogen-Bond Networks Stabilize Active Site Organization in ALKBH2 104

3.3.2.4. Dynamic Distance to C7 and C8 Enables Oxidative Repair of εC and ε5mC in ALKBH2 106

3.4. Conclusions 107

3.5. References 109

Chapter 4 : Conclusions and Future Directions 114

4.1. Summary 114

4.2. Contributions from Thesis 116

4.3. Future Directions 119

4.4. References 121

Appendix A 124

Appendix B 135

LIST OF TABLES

Table 3.1. Average E125–C1′ distance in AAG active site, with a standard deviation of 0.464 Å between substrates..... 97

LIST OF FIGURES

Figure 1.1. Structure and numbering of A) the canonical DNA nucleobases, and B) 2′-deoxyribose. 4

Figure 1.2. Structure of the DNA A) base pairs, B) single strand, and C) double helix. ... 4

Figure 1.3. Examples of A) the deaminated DNA lesion of cytosine, and B) the alkylated DNA lesion of guanine, with modifications highlighted in red. 6

Figure 1.4. Structure of common methylated DNA lesions, with modifications highlighted in red. 7

Figure 1.5. Structure of common etheno DNA lesions, with modifications highlighted in red. 8

Figure 1.6. Structure of common epigenetic modifications, with modification of the base highlighted in red. 9

Figure 1.7. Schematic representation of the BER pathway using the representative example of D deamination to U. 12

Figure 1.8. AlkB-catalyzed repair of 3,N^4-εC-containing DNA, with aberrant atoms highlighted in blue. 14

Figure 1.9. α-KG decarboxylated to succinate generated during the catalytic cycle of α-KG dependent enzymes. 15

Figure 2.1. X-ray crystal structures of *msm*UdgX for the wild-type DNA–protein crosslink (PDB ID: 6IOD) (green, left), the H109A mutant unbound to DNA (PDB ID: 8IIG) (pink, middle), and the H109E mutant DNA–protein crosslink (PDB ID: 6L6S) (orange, right), highlighting active site residues. 30

Figure 2.2. The proposed mechanisms of action of UdgX involving H109 forming a DNA–protein crosslink from two different tautomers. A) HIE109 tautomer where Nε is protonated and activated by E52 forming a crosslink with Nε via a concerted mechanism. B) HID109 tautomer where Nδ is protonated and activated by E52, enabling Nε to form the crosslink via an S_N1-like mechanism. 32

Figure 2.3. The percent occupancy of UdgX HIE109 tautomer clustered according to the distance of C1′ to Nε measured over the course of the MD simulation replicates. Cluster 1 displays the highest occupancy at 75% of the simulation time based on the closest alignment of H109 and C1′ (left). Cluster 2 displays an occupancy of 23% and much weaker alignment between H109 and C1′ (middle). Cluster 3 displays the weakest and furthest alignment of H109 and C1′ which occupies only 2% of the simulation time (right). 39

Figure 2.4. HIE representative dihedral angles plotted on a 360° range to describe the possible conformations observed. HIE (∠(ND1–CG–Cβ–Cα); RING) (x-axis) and (∠(CG–Cβ–Cα–N); ARM) (y-axis) dihedral angles plotted to show highest populated conformations for all 5 replicates (left). RING (orange) and ARM (blue) dihedral angles on a histogram highlighting the percent occupancy of each position over all trajectories and replicates (right). 40

Figure 2.5. UdgX active site of wild-type HIE tautomer with corresponding C1'–H109 distance, stabilizing residues, and water density. A) The percent occupancies of hydrogen bonds observed for structurally supportive residues over the course of MD simulations. B) High density 0.6 grade grid iso-mesh plot of water that frequents the active site over the simulation of all quintuplicates. 42

Figure 2.6. The percent occupancy of UdgX HID109 tautomer clustered according to the distance of C1′ to Nε measured over the course of the MD simulation replicates. Cluster 1 displays the highest occupancy at 49% of the simulation time based on the closest alignment of H109 and C1′ (left). Cluster 2 displays an occupancy of 35% and weaker alignment between H109 and C1′ (middle). Cluster 3 displays the weakest and furthest alignment of H109 and C1′ which only occupies about 15% of the simulation time (right). 44

Figure 2.7. HID representative dihedral angles plotted on a 360° range to describe the possible conformations observed. HID (∠(ND1–CG–Cβ–Cα); RING) (x-axis) and (∠(CG–Cβ–Cα–N); ARM) (y-axis) dihedral angles plotted to show highest populated conformations for all 5 replicates (left). RING (orange) and ARM (blue) dihedral angles on a histogram highlighting the percent occupancy of each position over all trajectories and replicates (right). 45

Figure 2.8. UdgX active site of wild-type the HID tautomer with corresponding C1'–H109 distance and stabilizing residues. (A) The percent occupancies of hydrogen bonds displayed by structurally supportive residues over the course of MD simulations. (B) High density 0.45 Å mesh plot of water in position to aid nucleophilic attack of E52. 46

Figure 2.9. UdgX HID active site, highlighting an MD representative structure of bridging water molecules located in high water density areas. Bridging water molecules were found between E52 (OE1 and OE2) and the 5′-phosphate backbone, while also hydrogen bonding to Nδ–H of H109..... 47

Figure 2.10. Catalytic mechanism of action of HID109 wild type using QM/MM calculations. Detailed distance and hydrogen-bonding interactions of the RC, TS, and PC. Accompanied by a Gibbs energy level diagram displaying the energy barrier to reach an intermediate complex and detailed mechanism of action. 49

Figure 2.11. Representative structure from MD simulations of the H109A, H109E, H109S, H109Q, H109K and H109D mutants. (A) H109A (C) H109E (E) H109S (G) H109Q (I) H109K (K) H109D; water density mesh plot to display water in the active site over the course of all replicates to display water availability in the system. (B) H109A (D) H109E (F) H109S (H) H109Q (J) H109K (L) H109D; structure and stabilizing residues and the percentage occupancy of hydrogen bonds over the course of the simulations. 56

Figure 2.12. Catalytic mechanism of action of H109A mutant undergoing hydrolysis reaction using QM/MM calculations. Accompanied by a Gibbs energy level diagram displaying energy barrier to reach intermediate complex. 61

Figure 2.13. Catalytic mechanism of action of H109E mutant undergoing crosslink formation reaction using QM/MM calculations. Accompanied by a Gibbs energy diagram displaying the energy barrier to reach intermediate complex. 66

Figure 3.1. Structures of A) common etheno DNA lesions (εA, εC) and B) a newly identified DNA etheno lesion (ε5mC), with modifications highlighted in red. 76

Figure 3.2. X-ray crystal structure of AAG bound to A) εC-containing DNA (PDB ID: 3QI5) and B) εA-containing DNA (PDB ID: 1EWN). 80

Figure 3.3. X-ray crystal structure of ALKBH2 bound to 3MeC-containing DNA (PDB ID: 3RZJ). 81

Figure 3.4. Proposed S_N1-like hydrolysis mechanism for excision of εA by AAG, with stacking interactions highlighted in blue. Initial cleavage of the C1′–N9 glycosidic bond generates an oxocarbenium-like intermediate, followed by nucleophilic attack of an activated water molecule at C1′. The catalytic residue E125 facilitates water activation, resulting in hydrolysis of the glycosidic bond, protonation of the departing εA base, and formation of an abasic (AP) site. 86

Figure 3.5. Representative structures and structural overlays highlighting differences in nucleobase positioning within the AAG active site. Individual representative MD structures of AAG bound to A) εA (green), B) εC (pink), and C) ε5mC (cyan), highlighting key active site residues. D) Overlay of the εA- (PDB ID: 1EWN) and εC-bound (PDB ID: 3QI5) crystal structures demonstrating that εC adopts a slightly deeper insertion into the active site pocket relative to εA. E) Overlay of representative MD structures for the εA and εC systems showing that the deeper positioning of εC observed experimentally is maintained during the simulations. F) Overlay of representative MD structures for the εA and ε5mC systems illustrating that ε5mC adopts a binding mode similar to that of εC despite the presence of the additional methyl group. 88

Figure 3.6. Representative MD structures of AAG bound to A) εA (green), B) εC (pink), or C) ε5mC (cyan) containing DNA, highlighting π–π-stacking distance (Å) and percent occupancy of the nucleobase with H136 and Y127. 90

Figure 3.7. Representative MD structures of AAG bound to A) εA (green), B) εC (pink), or C) ε5mC (cyan) containing DNA, highlighting hydrogen-bonding interactions between E125 and L180, and E125–Y127 catalytic dyad maintained for >98% of simulations for all nucleotides. 92

Figure 3.8. Representative MD structures of AAG bound to A) εA (green), B) εC (pink), or C) ε5mC (cyan) containing DNA, highlighting a common hydrogen bond between the H136 backbone and the bound nucleobase. 93

Figure 3.9. Representative MD structures of AAG bound to A) εA (green), B) εC (pink), or C) ε5mC (cyan) containing DNA, highlighting hydrogen-bonding interactions from Y159 or N169 and the nucleotide. 94

Figure 3.10. Representative MD structures of AAG bound to A) εA (green), B) εC (pink), or C) ε5mC (cyan) containing DNA, highlighting water density around E125 and the nucleobase and illustrated using an isolated water density mesh plot. 96

Figure 3.11. Proposed direct reversal repair pathway of etheno lesions to form an epoxide when starting from the understudied 3,N^4–etheno–5–methylcytosine (ε5mC) DNA lesion and the resulting repaired canonical structure of 5–methylcytosine (5MeC). 99

Figure 3.12. Representative MD structures of ALKBH2 bound to A) εC (burgundy) or B) ε5mC (blue) containing DNA, highlighting the π-Interactions with a commonly observed phosphate backbone hydrogen bond in ALKBH2 repaired substrates. 102

Figure 3.13. Representative MD structures of ALKBH2 bound to A) εC (burgundy) or B) ε5mC (blue) containing DNA, highlighting the Fe^{IV}-oxo center and the coordinated H171, H236, D173, succinate, water, and oxygen and their corresponding coordination distances. 104

Figure 3.14. Representative MD structures of ALKBH2 bound to A) εC (burgundy) or B) ε5mC (blue) containing DNA, highlighting the nucleobase (C7 and C8 position) to Fe^{IV}-oxo distance and surrounding hydrogen-bonding interactions. 105

LIST OF ABBREVIATIONS

1MeA	N1-methyladenine
3MeA	N3-methyladenine
3MeC	N3-methylcytosine
4MeC	N4-methylcytosine
5MeC	5-methylcytosine
6MeA	N6-methyladenine
7MeG	N7-methylguanine
1,N^6-εA	1,N^6-ethenoadenine
3,N^4-εC	3,N^4-ethenocytosine
εA	ethenoadenine
εC	ethenocytosine
ε5mC	3,N^4-etheno-5-methylcytosine
AAG	Alkyladenine DNA glycosylase
ALKBH2	AlkB homolog 2
ALKBH3	AlkB homolog 3
AlkB	Alkylation repair dioxygenase
α-KG	Alpha-ketoglutarate
APE1	Apurinic/apyrimidinic endonuclease 1
AP site	Apurinic/apyrimidinic site
BER	Base excision repair
DNA	Deoxyribonucleic acid
ds-/ss-DNA	Double-stranded DNA/Single-stranded DNA
DRR	Direct reversal repair
FTO	Fat mass and obesity-associated protein (AlkB homolog)
GAFF	General AMBER force field
HIE / HID	Histidine tautomer (ε-protonated / δ-protonated)
MD	Molecular dynamics
MM	Molecular mechanics
MBD4	Methyl-CpG binding domain 4
Msm	*Mycobacterium smegmatis*
MutY	Adenine DNA glycosylase
NEIL1	Nei like glycosylase 1
ONIOM	Our own N-layered Integrated molecular orbital and molecular mechanics
PC	Product complex
PDB	Protein Data Bank
QM	Quantum mechanics
QM/MM	Quantum mechanics / molecular mechanics hybrid method
RC	Reactant complex
RMSD	Root-mean-square deviation
SN_1	Unimolecular nucleophilic substitution
TS	Transition state
UDG	Uracil DNA glycosylase
UdgX	Uracil DNA glycosylase X
UNG	Uracil-N-glycosylase

Chapter 1 : Introduction

1.1. General Overview

Nucleic acids are fundamental biomolecules essential for the storage,[1] transmission,[2] and regulation of genetic information in all living organisms.[3] Deoxyribonucleic acid (DNA) serves as the primary storage of genetic information and encoding instructions for protein synthesis and cellular function.[4, 5] DNA is subject to a range of chemical modifications that influence gene expression, structural stability, and cellular responses to environmental or metabolic stimuli.[4] Chemical modifications to DNA can arise through regulated enzymatic processes,[6] such as the development of epigenetic markers,[7-10] or through spontaneous and environmentally-induced reactions that generate DNA damage.[11, 12] The structure and function of DNA can be influenced by these chemical changes, which must be accurately interpreted and processed by cellular enzymes.[13] For example, enzymes can install or remove epigenetic markers or resolve chemically-damaged nucleobases to repair DNA.[14]

This thesis is focused on elucidating the atomic-level mechanisms of DNA repair enzymes. The DNA repair pathways in human cells include base excision repair (BER)[15] and direct reversal repair (DRR).[16, 17] Each pathway involves several enzymes and misfunction of these enzymes is related to genome instability and human diseases, including cancer.[18] DNA repair enzymes can also influence the cellular response to chemotherapeutic agents.[19] However, the substrate specificity and catalytic mechanisms of many DNA repair enzymes are not fully understood. For example, the uracil DNA glycosylase (UDG) enzyme UdgX has been shown to exhibit mechanistic behaviour that differs from other members of the UDG family.[20] As another example, alkyladenine DNA

glycosylase (AAG) and AlkB family enzymes repair a similar set of diverse substrates yet also have distinct differences in their repair capabilities.[21, 22] Although it is difficult to gain atomic-level information about enzyme function solely from experimental techniques, computational chemistry can provide detailed information about enzyme structural dynamics, and catalytic mechanisms. By enabling atomic-level investigation of enzyme structure, conformational dynamics, and reaction pathways, computational methods provide access to transition states and mechanistic details that are often inaccessible experimentally.

This thesis uses computational chemistry to provide a powerful complementary approach to experimental structural and biochemical methods to address questions regarding the function of DNA repair enzymes. The central objective is to elucidate how select DNA repair enzymes, namely UdgX, AAG, and ALKBH2, achieve substrate specificity and catalytic efficiency. In this chapter, the biochemical background is provided, including an overview of the structure of DNA, non-bulky DNA nucleobase modifications, the major repair pathways used to resolve such lesions, the enzymes of interest in this work, and the computational tools that can be used to study enzyme function.

1.2. Nucleic Acid Structure and Function

DNA is a polymer composed of nucleotide monomers, each consisting of a nucleobase, a deoxyribose sugar, and a phosphate group (Figure 1.1).[4] Under physiological conditions DNA most commonly adopts the B–form helix however, alternative conformations such as A or Z form can arise depending on sequence context, hydration state, and ionic

environment. The B, A, and Z conformations differ in helical geometry, base orientation, and groove dimensions as B–DNA is a right-handed helix with well-defined major and minor grooves, A–DNA is more compact with a deeper major groove and bases tilted relative to the helix axis, and Z–DNA adopts a left-handed zigzag backbone with alternating syn- or anti- glycosidic conformations. The canonical DNA nucleobases are adenine (A), guanine (G), cytosine (C), and thymine (T), which pair through Watson-Crick-Franklin hydrogen bonding between complementary bases (A–T and G–C) to form antiparallel double helices (Figure 1.2).[4, 5] In addition to base pairing, hydrophobic stacking interactions contribute significantly to the duplex stability and regulate the accessibility of the nucleobases for chemical modification.[4] The glycosidic bond linking the nucleobase to the sugar is a critical structural feature of DNA. Although this bond is generally stable, the glycosidic bond is susceptible to cleavage under certain chemical or enzymatic conditions, particularly when the canonical nucleobases are chemically altered.[23, 24] Such chemical alterations may arise from either spontaneous damage or regulated enzymatic modification, underscoring the dual role of DNA chemistry in genome maintenance and epigenetic regulation.

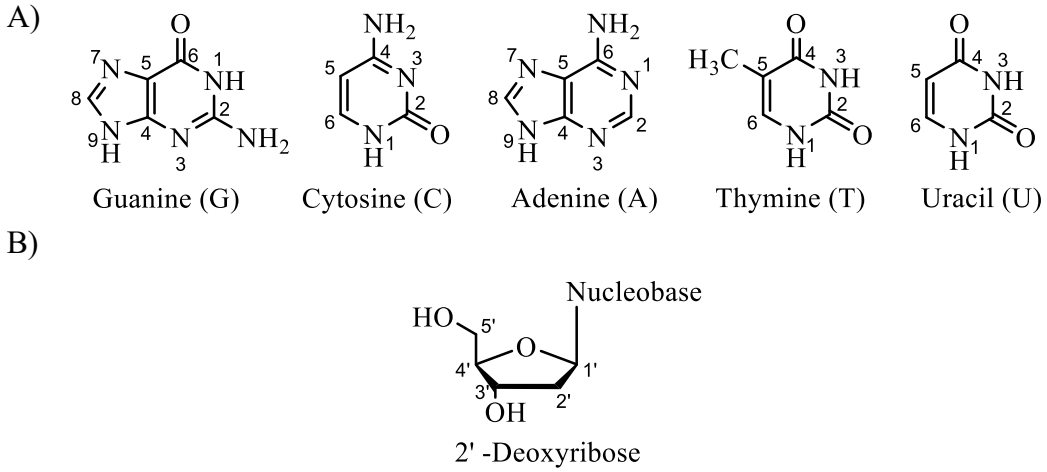


Figure 1.1. Structure and numbering of A) the canonical DNA nucleobases, and B) 2′-deoxyribose.

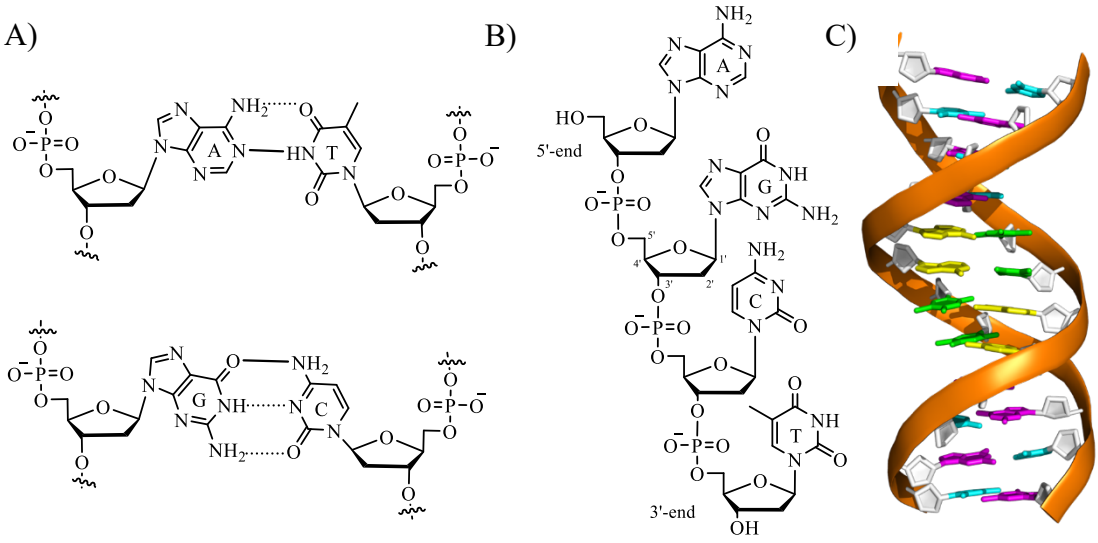


Figure 1.2. Structure of the DNA A) base pairs, B) single strand, and C) double helix.

1.3. DNA Modifications

DNA can undergo a wide range of chemical modifications, which may arise from normal cellular metabolism, environmental exposure, or programmed enzymatic processes. These modifications can be broadly categorized as DNA damage and epigenetic

modifications, depending on their origin and biological consequences.[25] While DNA damage generally compromises genomic integrity and requires repair, epigenetic modifications function as regulatory signals that influence gene expression, chromatin organization, and cellular identity.[25]

Some common types of DNA damage include nucleobase alkylation, oxidation, and deamination (Figure 1.3).[13, 26] These lesions can originate from both endogenous and exogenous sources. Endogenously, DNA damage arises from normal cellular processes such as the generation of reactive oxygen and nitrogen species during oxidative metabolism, spontaneous hydrolytic reactions, and replication-associated errors.[18, 27] These processes can lead to base oxidation, depurination, and deamination events. Exogenously, DNA is exposed to a range of environmental agents, including ultraviolet (UV) and ionizing radiation, environmental toxins, and chemical mutagens such as alkylating and nitrosylation agents, which can induce base modifications, strand breaks, and crosslinks.[11, 12, 27] One of the most prevalent forms of DNA damage is base deamination, a spontaneous hydrolytic reaction in which an exocyclic amine is replaced by a carbonyl group.[28, 29] Cytosine is particularly susceptible to deamination, yielding uracil through the hydrolytic loss of its amino group.[30] This process occurs under physiological conditions and is promoted by heat, fluctuations in pH, and reactive oxygen or nitrogen species generated from normal cellular metabolism. If left unrepaired, uracil pairs with adenine during replication, resulting in C to T transition mutations.[13, 31] Deamination of C to U therefore represents a major source of point mutations in genomic DNA.

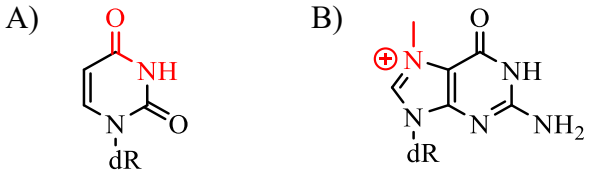


Figure 1.3. Examples of A) uracil produced by the deamination of cytosine, and B) the alkylated DNA lesion of guanine, with modifications highlighted in red.

In addition to deamination, alkylation is a common source of DNA damage.[32] The simplest form of alkylation involves the transfer of a methyl group to a nucleobase, which can occur through exposure to endogenous sources such as S-adenosylmethionine or exogenous formaldehyde, nitrosamines, and nitrosoureas.[22, 33] These reactions generate a variety of methylated lesions, including N7–methylguanine (7MeG), N3–methyladenine (3MeA), N1–methyladenine (1MeA), and N3–methylcytosine (3MeC) (Figure 1.4).[22, 32] While these lesions differ in their chemical structure, they can all disrupt DNA stability and processing.[21, 22] Although 7MeG is not directly mutagenic or cytotoxic, it undergoes depurination approximately 10^6-fold faster than unmodified guanine.[22] This results in the formation of abasic sites that are strongly mutagenic due to preferential adenine insertion during replication. Alternatively, 7MeG can impede DNA replication or transcription.[22] In contrast, 1MeA and 3MeA are not directly mutagenic but strongly block replication DNA polymerases, resulting in cytotoxic replication arrest.[21, 22] Similarly, 3MeC interferes with Watson-Crick-Franklin hydrogen bonding and effectively blocks DNA replication, leading to increased mutagenesis if not repaired.[22] These methylated lesions illustrate how relatively small chemical changes can have severe biological consequences. Beyond simple methyl adducts, alkylation damage also includes bulkier lesions, such as etheno adducts formed when DNA reacts with lipid peroxidation products or industrial carcinogens

including vinyl chloride (Figure 1.5).[21, 22, 34] These lesions distort base-pairing interactions and are both cytotoxic and mutagenic, further underscoring the chemical diversity of alkylation damage in the genome.

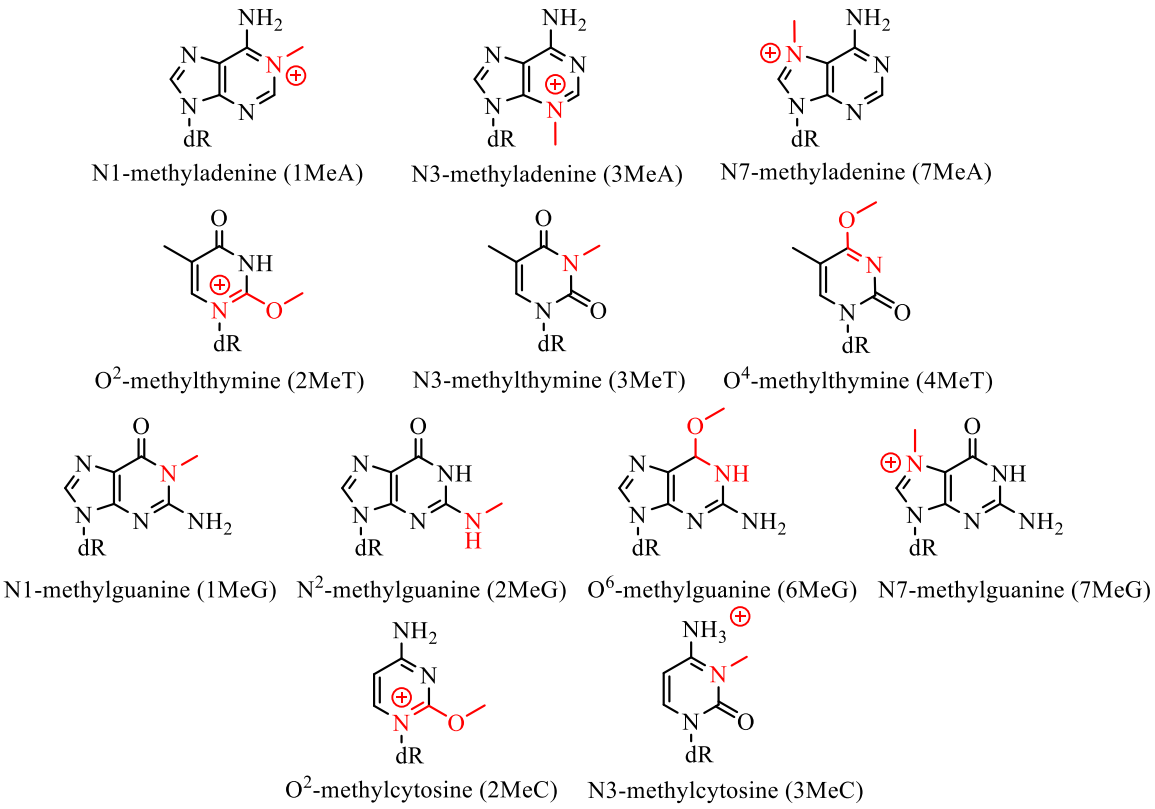


Figure 1.4. Structure of common methylated DNA lesions, with modifications highlighted in red.

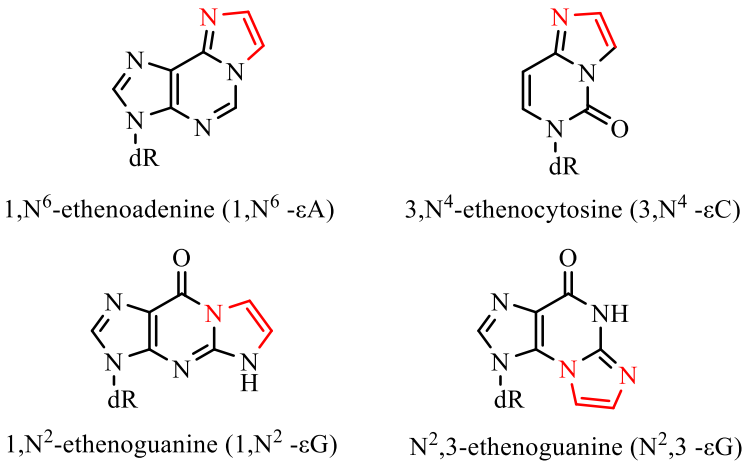
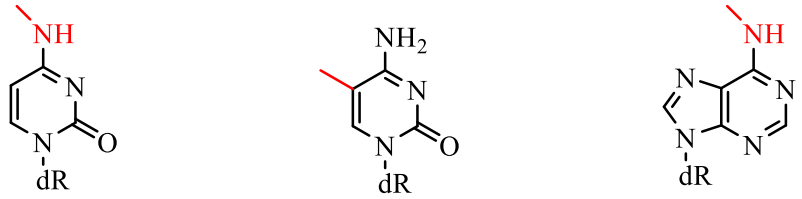


Figure 1.5. Structure of common etheno DNA lesions, with modifications highlighted in red.

In contrast to damage-induced modifications, DNA is also subject to enzymatically-installed epigenetic modifications that play essential regulatory roles.[35] Cytosine methylation at the C5 position to form 5MeC is a central epigenetic mark involved in transcriptional regulation, genomic imprinting, and chromatin structure (Figure 1.6).[25, 35, 36] Additional methylated bases, such as N4–methylcytosine (4MeC) and N6–methyladenine (6MeA), are well-established epigenetic signals in prokaryotic systems and have emerging regulatory roles in eukaryotes.[37] These modifications do not 'damage' DNA but instead encode biological information that must be accurately maintained across cell divisions. Epigenetic modifications can however become substrates for further chemical damage, which complicates the distinction between modification and lesion formation. For example, deamination of 5MeC yields thymine (as opposed to C deaminating to U), generating T:G mispairs, avoiding standard DNA repair pathways and therefore posing a unique mutagenic threat.[14]



N^4-methylcytosine (4MeC) C^5-methylcytosine (5MeC) N^6-methyladenine (6MeA)

Figure 1.6. Structure of common epigenetic modifications, with modification of the base highlighted in red.

Together, these diverse DNA modifications highlight the chemical vulnerability of the genome and the necessity for precise molecular mechanisms to interpret, process, and resolve altered nucleobases. The following section introduces the classes of enzymes that recognize and process modified DNA, focusing on how cells discriminate between structural similar lesions and regulatory modifications to maintain both genomic and epigenetic integrity.

1.4. Enzymes That Process Nucleic Acids

To preserve the genomic stability of DNA, cellular machinery must distinguish damaged or modified nucleotides from undamaged DNA. The integrity and function of nucleic acids are maintained by a diverse array of enzymes that participate in processes such as replication, repair, modification, and degradation.[38] These enzymes ensure proper gene expression, prevent mutations, and regulate cellular homeostasis.[6] Dysregulation of these enzymes can lead to severe genetic and epigenetic consequences. Therefore, these enzymes must balance specificity and flexibility to recognize chemically-diverse lesions and avoid unnecessary processing of canonical bases. Among the various pathways employed to

process modified nucleic acids, BER[15] plays a central role in resolving small, non-bulky DNA lesions through the action of DNA glycosylases,[39] which initiate repair by selectively excising damaged nucleobases.[40] Additional details of BER and glycosylases are provided in the next section.

1.5. Base Excision Repair and Glycosylases

BER is a conserved DNA repair pathway that is responsible for correcting small, non-helix-distorting lesions that arise from spontaneous base hydrolysis, oxidation and alkylation.[13, 41] These lesions frequently alter the hydrogen-bonding or electronic properties of nucleobases without significantly perturbing the overall structure of the DNA duplex, making them difficult to detect. BER provides a targeted mechanism that can restore genomic integrity by selectively removing damaged bases while preserving the surrounding DNA structure.[42] The BER pathway is initiated by DNA glycosylases by scanning for damaged nucleotides.[39] This initial step is both lesion-specific and rate-limiting as it commits the damaged nucleotides to repair and determines consecutive downstream processing.

DNA glycosylases that initiate BER may be monofunctional or bifunctional.[13] Monofunctional glycosylases typically utilize an Asp- or Glu-activated water to facilitate deglycosylation through nucleotide hydrolysis, while bifunctional glycosylases cleave the glycosidic bond using an active site amine to form a crosslink and facilitate scission of the 3′- and/or 5′-phosphodiester bonds of the resulting AP site.[13, 26] The function of the glycosylase generates an AP site which is cleaved by an apurinic or apyrimidic

endonuclease, like APE1[43, 44] in eukaryotes, which represents a chemically unstable intermediate that must be processed quickly in order to prevent strand breakage or mutagenesis (Figure 1.7).[40, 44-46] In bifunctional glycosylase-initiated BER, strand incision may occur at the same time as base removal through various unique elimination reactions. Subsequent repair steps involve gap filling by a DNA polymerase using the complementary strands as a template. In the case of monofunctional glycosylases, the AP endonuclease-mediated incision generates a 3′–hydroxyl and a 5′–deoxyribose phosphate (5′–dRP) terminus.[44] The 5′–dRP group is subsequently removed by the lyase activity of DNA polymerase β, which then inserts the correct nucleotide opposite the lesion.[47] The repair process is completed by DNA ligase, which seals the remaining break in the sugar–phosphate backbone, fully restoring DNA strand continuity.[47] While BER pathways exhibit mechanistic diversity, they are fundamentally governed by the accuracy and catalytic efficiency of the initiating DNA glycosylases, emphasizing their importance.

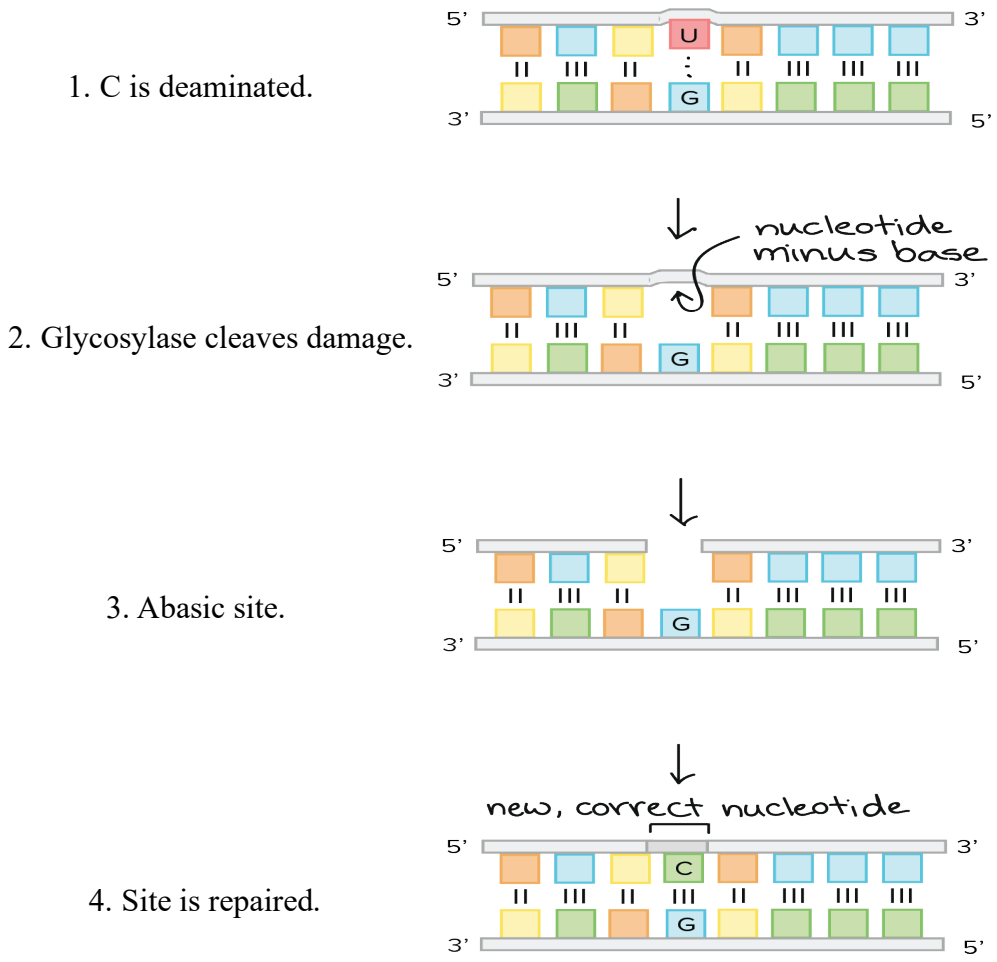


Figure 1.7. Schematic representation of the BER pathway using the representative example of D deamination to U.

UDGs represent a major class of BER-initiating enzymes and are responsible for removing uracil from DNA, uracil that arises via deamination primarily from misincorporation during replication. Canonical UDGs are highly specific and catalyze hydrolytic cleavage of the N–glycosidic bond. In contrast, UdgX is a mechanistically distinct member of the UDG family that excises uracil through direct nucleophilic attack by an active site histidine, forming a stable covalent DNA–protein crosslink.[20] This unusual

reactivity commits the lesion to excision while simultaneously inactivating the enzyme, representing a fundamentally different solution to uracil removal within the BER process.

AAG plays a role in BER by initiating repair of a very broad spectrum of alkylated and deaminated purine lesions. AAG is a monofunctional glycosylase notable for its unusually broad substrate specificity. AAG can excise a variety of alkylated and deaminated purines, including hypoxanthine and several etheno adducts.[48] Unlike many glycosylases that are highly selective for a single substrate, AAG can accommodate chemically diverse lesions within a single active site. However, this variety of substrate scope is paired with some exclusions. AAG does not excise alkylated pyrimidines such as 3MeC, and certain lesions, including 3,N^4–ethenocytosine (εC) can bind to AAG yet inhibit catalysis.[48] These observations suggest that lesion discrimination by AAG is caused by slight differences in substrate positioning, electronic activation, or transition state stabilization. Together UDG and AAG enzymes highlight the diversity of catalytic strategies that exist within BER, ranging from highly specific lesion recognition to broad substrate accommodation within a single active site.

Despite extensive biochemical and structural characterization of DNA glycosylases, key questions remain regarding how active site architecture, enzyme dynamics, and electronic activation collectively govern lesion recognition and catalytic outcome. In particular, the molecular basis for broad versus narrow substrate specificity, the role of nucleophiles in glycosidic bond cleavage, and the influence of protein dynamics on active site stabilization remain incompletely understood and mechanistically unresolved. These unresolved questions form the basis for the computational investigations presented in this thesis, with

a focus on UdgX-mediated uracil excision and comparative analysis of AAG substrate recognition.

1.6. Direct Reversal Repair and AlkB Family Enzymes

In contrast to BER, which removes damaged nucleobases through glycosidic bond cleavage and replacement, DRR directly removes damage from DNA without excising the base or disrupting the sugar–phosphate backbone. AlkB family enzymes use the DRR pathway and play a central role in repairing alkylated and etheno-modified DNA. AlkB enzymes are Fe (II)- and α-ketoglutarate (α-KG) dependent dioxygenases that catalyze oxidative dealkylation of damaged nucleobases (Figure 1.8).[17, 48] Instead of cleaving the glycosidic bond, these enzymes hydroxylate the alkyl substituent, producing an unstable intermediate that decomposes to release glyoxal or a related small molecule, generating the canonical base.[21] This chemistry allows AlkB enzymes to directly reverse the damage while preserving the integrity of the DNA strand.

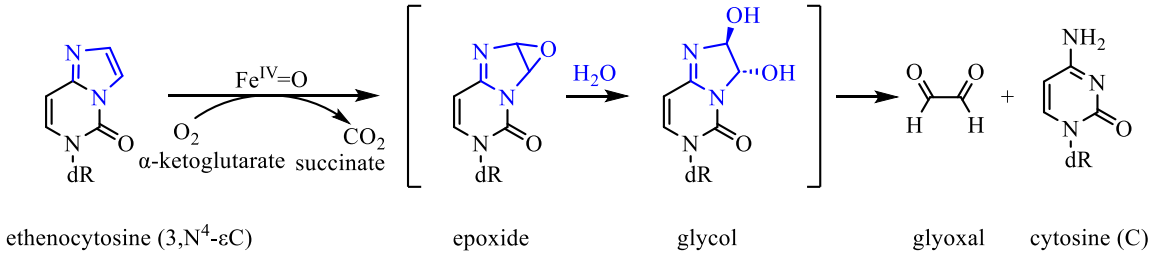


Figure 1.8. AlkB-catalyzed repair of 3,N^4-εC-containing DNA, with aberrant atoms highlighted in blue.

The catalytic cycle of AlkB enzymes requires molecular oxygen, Fe (II), and αKG, which is decarboxylated to succinate during turn over (Figure 1.9).[48] Activation of oxygen

generates a high-valent iron–oxo species that abstracts a hydrogen atom from the alkyl group attached to the nucleobase.[21] The efficiency of this process depends on precise substrate positioning because productive oxidation requires alignment of the damaged functional group with the reactive Fe–O center. Substrate selectivity in AlkB enzymes is therefore ultimately controlled by the structural and dynamic features of the binding pocket, which define substrate binding modes and enforce productive alignment. The *Escherichia coli* (E. coli) AlkB protein is the most extensively characterized member of this family, but AlkB homologs are widespread across prokaryotes and eukaryotes.[21] In humans, only two homologs, ALKBH3 and ALKBH2, exhibit the same robustness within DNA repair activity.[49, 50] ALKBH2, in particular is associated with the repair of lesions within double-stranded DNA and plays a key role in maintaining genomic stability.

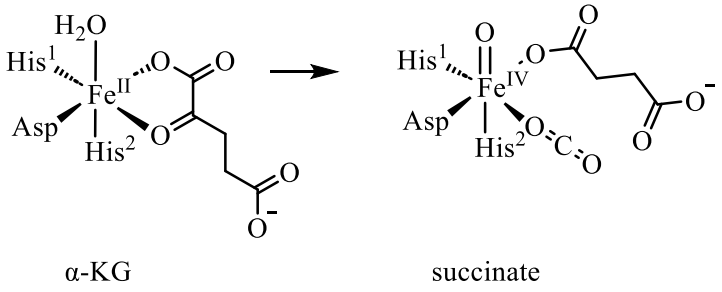


Figure 1.9. α-KG decarboxylated to succinate generated during the catalytic cycle of α-KG dependent enzymes.

A defining distinction between DRR and BER lies in substrate selection and lesion tolerance. While AAG initiates BER by excising a broad range of alkylated and deaminated purines, AlkB enzymes preferentially target alkylation damage that perturbs the Watson–Crick-Franklin face of the nucleobase, such as 1MeA, 3MeC, and several etheno adducts.[48, 50, 51] Importantly, there is partial overlap in substrate scope between these

pathways. For example, both AAG and AlkB enzymes can process certain etheno lesions such as, 1,N^6–ethenoadenine (1,N^6–εA) and 3,N^4–ethenocytosine (3,N^4–εC).[48] However, AlkB enzymes uniquely repair lesions such as εC, which inhibit AAG catalyzed excision.[21, 22, 48] This difference underscores that substrate binding alone is insufficient to predict repair outcome, and that mechanistic constraints implemented by each pathway are also important.

In addition to their role in repairing alkylated DNA lesions, AlkB family enzymes have been proposed to act on DNA modifications containing etheno lesions, which when repaired are associated with epigenetic regulation. An example of this overlap is the formation of 3,N^4–etheno–5–methylcytosine (ε5mC), which arises when the epigenetic modification 5MeC undergoes lipid peroxidation-induced alkylation.[50] Oxidative processing of methylated cytosines suggests a potential involvement of AlkB enzymes in a more dynamic DNA modification pathway, which complicates the distinction between repair and regulatory processes. ε5mC is chemically distinct from both canonical alkylation damage and unmodified 5MeC, as the methyl group remains while introducing a bulky etheno bridge that disrupts base pairing.[50] Importantly, ε5mC has been shown to inhibit oxidation by TET family enzymes, interfering with active DNA demethylation pathways and posing a threat to epigenetic regulation.[50] Recent experimental studies have demonstrated that AlkB family enzymes, including human ALKBH2, ALKBH3, and FTO, are capable of repairing ε5mC through oxidative removal of the etheno moiety and restoring it to the original 5MeC base.[50] Notably, ε5mC is not a substrate for base excision repair glycosylase AAG,[50] highlighting a unique role for DRR in preserving epigenetic information. This leads to one of the goals of this thesis, namely to describe why AAG, a

glycosylase with broad substrate specificity, does not repair ε5mC despite processing other etheno lesions. The ability of ALKBH2 to repair ε5mC is also mechanistically intriguing, as this substrate combines steric bulk, altered electronic structure, and epigenetic functionality within a single nucleobase. While biochemical and cellular assays have established that ALKBH2 can process ε5mC, the structural basis for how this lesion is accommodated within the active site remains unclear. An objective of this thesis is to use molecular dynamics (MD) simulations to elucidate how ALKBH2 accommodates and processes ε5mC by characterizing substrate binding modes, active site interactions, and conformational dynamics that are not readily accessible through experimental approaches alone.

1.7. Implications of Enzyme Misfunction for Human Health

Dysregulation of the processes involved in modifying nucleic acids can have severe implications for human health. For example, failure of BER to properly repair damaged DNA has severe biological consequences. Persistent lesions may block replicative polymerases, promote error-prone translesion synthesis, or lead to accumulation of abasic sites that destabilize the genome.[52, 53] Defects in BER proteins, including loss of function mutations, altered expression levels, or impaired lesion recognition and catalytic activity, are associated with increased mutation rates, genomic instability, and susceptibility to cancer.[19, 48] Mechanistically, deficiencies in glycosylases or downstream BER enzymes can result in the persistence of mutagenic lesions or toxic repair intermediates, such as abasic sites and strand breaks, which can interfere with replication and transcription or trigger cell death pathways. At the same time, hyperactivation of BER has been implicated in

resistance to chemotherapeutic agents that rely on DNA alkylation for cytotoxicity,[19] highlighting the dual role of BER in genome protection and therapeutic response. Alternatively, deficiencies in enzymes responsible for repairing specific types of DNA damage, such as those in the BER pathway, can increase susceptibility to diseases like M. tuberculosis,[20] characterized by lung infections that can attack any part of the body including the kidneys, spine, and brain.[54-56] Furthermore, disruptions in epigenetic modifications, including abnormal DNA methylation patterns or histone modifications, are strongly linked to disease development and progression.[57] In particular, aberrant DNA methylation, such as global or site-specific methylation at promoter regions, can lead to genomic instability or inappropriate silencing of tumor suppressor genes.[58] Dysregulation of histone modifications, including acetylation and methylation, can further alter chromatin structure and accessibility, disrupting normal transcriptional programs.[57, 58] These epigenetic alterations are frequently observed in cancer and other complex diseases, highlighting the importance of maintaining precise control over nucleic acid modification and repair processes.[57] Additionally, elevated AlkB activity can counteract the effectiveness of alkylating chemotherapeutic agents, contributing to drug resistance in tumor cells.[58] As a result of the connection between deficiencies in nucleic acid processing and human disease, it is important for us to understand the molecular mechanisms by which repair enzymes recognize and process damaged DNA. This is therefore critical for both fundamental biology and therapeutic strategies aimed at restoring proper nucleic acid function as well as insights into the drug resistance mechanisms.

1.8. Computational Approaches Used to Study the Chemistry of Enzyme–Substrate Complexes

Since processes like the BER and DRR pathways involve a series of intermediates and specific chemical steps, many aspects of these pathways remain difficult to resolve experimentally. In general, enzymes have complex molecular structures and dynamic behaviors, which can be difficult to investigate in traditional 'wet' laboratories. Computational chemistry provides a powerful framework that allows us to look deeper and more closely at enzyme function and serves as a tool to decipher the mechanisms of enzyme function.[13, 14, 29] Computer simulations permit exploration of the atomic-level details of enzyme–substrate interactions, including the structural dynamics of enzyme complexes and reaction mechanisms.[29] Through MD simulations, we can observe how enzymes undergo conformational changes and how the enzyme interacts with the substrate.[29, 43] Once a representative structure of the reactant complex (RC) is obtained from MD simulations, the catalytic reaction can be investigated using combined quantum mechanics (QM) and molecular mechanics (MM) approaches, where the atomic structure and relative energy of the RC, transition state (TS), and product complex (PC) can be obtained.[14, 29, 43, 59] Data collected from computer simulations complement experimental findings while offering a unique perspective. Specifically, many drugs are small molecule inhibitors that work well because they are TS mimics, and therefore having the details of the reaction pathway aids in drug design.[60] Despite growing computational interest in enzyme-mediated nucleic acid modification, fundamental questions remain regarding the atomistic origins of substrate specificity, pathway selection, transition-state stabilization, and the role of conformational dynamics in directing repair outcomes. This work provides new atomistic

insight into the dynamic and electronic factors that govern DNA repair catalysis, thereby helping to resolve longstanding mechanistic uncertainties surrounding enzyme specificity and reaction energetics. Specifically, MD and QM/MM are especially valuable for understanding how enzymes such as UdgX, AAG, and ALKBH2 achieve broad substrate specificity while maintaining stringent control over catalysis, and why certain lesions escape repair despite apparent binding competence.

1.9. Thesis Overview

The purpose of this thesis is to use computational chemistry to provide insight into the catalytic repair mechanism and substrate recognition strategies of DNA repair enzymes, with a focus on the DNA glycosylases UdgX and AAG, as well as the DRR enzyme ALKBH2. By utilizing MD simulations and QM/MM calculations, this works aims to elucidate how these enzymes accommodate either very specific or structurally diverse substrates and how subtle changes in lesion chemistry influence repair outcomes.

Chapter 2 focuses on UdgX, an unconventional UDG that excises uracil through a covalent crosslinking mechanism. Building on previously proposed reaction pathways,[30, 61, 62] this chapter utilizes MD simulations to characterize the dynamic behaviour of the active site, identify accessible conformational states, and evaluate the roles of the histidine protonation and position in catalysis. QM/MM models seeded from MD snapshots help characterize the reaction pathway, providing insight into the mechanism of uracil excision and the electronic requirements for covalent DNA–protein crosslink formation.

Chapter 3 extends this computational investigation to AAG and ALKBH2, examining how these enzymes accommodate structurally-diverse alkylated and etheno lesions despite using different repair chemistry. MD simulations are used to explain why ALKBH2 can repair ε5mC, whereas AAG can not. This chapter begins with a review of prior computational studies of AAG in complex with εA or εC, providing a basis for understanding why εA is repaired while εC inhibits catalysis. These results are then compared with new simulations of ALKBH2 bound to εC and ε5mC to elucidate why, in contrast to AAG, ALKBH2 can process both substrates. Structural and dynamic analysis including atomic distances, π-stacking interactions between the nucleobase and active site residues, and water alignment, are used to rationalize differences in substrate accommodation and repair competency.

Together, these studies aim to establish unifying principles that explain how DNA repair enzymes balance substrate specificity, catalytic efficiency, and mechanistic diversity, collectively providing insight into both genome maintenance and broader design principles that command enzyme evolution.

1.10. References

(1) Travers, A.; Muskhelishvili, G. DNA structure and function. *FEBS J.* **2015**, *282* (12), 2279-2295. DOI: 10.1111/febs.13307.

(2) Sekine, S.-i.; Ehara, H.; Kujirai, T.; Kurumizaka, H. Structural perspectives on transcription in chromatin. *Trends Cell Biol.* **2024**, *34* (3), 211-224. DOI: 10.1016/j.tcb.2023.07.011.

(3) Kozak, M. Initiation of translation in prokaryotes and eukaryotes. *Gene* **1999**, *234* (2), 187-208. DOI: 10.1016/s0378-1119(99)00210-3.

(4) Minchin, S.; Lodge, J. Understanding biochemistry: structure and function of nucleic acids. *Essays Biochem.* **2019**, *63* (4), 433-456. DOI: 10.1042/ebc20180038.

(5) Watson, J. D.; Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171* (4356), 737-738. DOI: 10.1038/171737a0.

(6) Jacobs, A. L.; Schär, P. DNA glycosylases: in DNA repair and beyond. *Chromosoma* **2012**, *121* (1), 1-20. DOI: 10.1007/s00412-011-0347-4.

(7) Berson, A.; Nativio, R.; Berger, S. L.; Bonini, N. M. Epigenetic Regulation in Neurodegenerative Diseases. *Trends Neurosci.* **2018**, *41* (9), 587-598. DOI: 10.1016/j.tins.2018.05.005.

(8) Handy, D. E.; Castro, R.; Loscalzo, J. Epigenetic modifications: basic mechanisms and role in cardiovascular disease. *Circ.* **2011**, *123* (19), 2145-2156. DOI: 10.1161/circulationaha.110.956839.

(9) Dolinoy, D. C.; Weidman, J. R.; Jirtle, R. L. Epigenetic gene regulation: Linking early developmental environment to adult disease. *Reprod. Toxicol.* **2007**, *23* (3), 297-307. DOI: 10.1016/j.reprotox.2006.08.012.

(10) Kumar, S.; Chinnusamy, V.; Mohapatra, T. Epigenetics of Modified DNA Bases: 5-Methylcytosine and Beyond. *Front. Genet.* **2018**, *9*, 640. DOI: 10.3389/fgene.2018.00640.

(11) Barnes, J. L.; Zubair, M.; John, K.; Poirier, M. C.; Martin, F. L. Carcinogens and DNA damage. *Biochem. Soc. Trans.* **2018**, *46* (5), 1213-1224. DOI: 10.1042/bst20180519.

(12) Hakem, R. DNA-damage repair; the good, the bad, and the ugly. *EMBO J.* **2008**, *27* (4), 589-605. DOI: 10.1038/emboj.2008.15.

(13) Kaur, R.; Nikkel, D. J.; Wetmore, S. D. Computational studies of DNA repair: Insights into the function of monofunctional DNA glycosylases in the base excision repair pathway. *WIREs Comput. Mol. Sci.* **2020**, *10* (5), e1471. DOI: 10.1002/wcms.1471.

(14) Nikkel, D. J.; Wetmore, S. D. A tale of two mechanisms: Clarification of the pathway for MBD4 catalyzed glycosidic bond cleavage using MD and QM/MM calculations. *DNA Repair* **2026**, *157*, 103917. DOI: 10.1016/j.dnarep.2025.103917.

(15) Krokan, H. E.; Bjørås, M. Base excision repair. *Cold Spring Harb. Perspect. Biol.* **2013**, *5* (4), a012583. DOI: 10.1101/cshperspect.a012583.

(16) Duncan, T.; Trewick, S. C.; Koivisto, P.; Bates, P. A.; Lindahl, T.; Sedgwick, B. Reversal of DNA alkylation damage by two human dioxygenases. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (26), 16660-16665. DOI: 10.1073/pnas.262589799.

(17) Müller, T. A.; Hausinger, R. P. AlkB and Its Homologues – DNA Repair and Beyond. In *2-Oxoglutarate-Dependent Oxygenases*, Schofield, C., Hausinger, R. Eds.; The Royal Society of Chemistry, 2015; pp 246-262.

(18) Jin, B.; Robertson, K. D. DNA methyltransferases, DNA damage repair, and cancer. *Adv. Exp. Med. Biol.* **2013**, *754*, 3-29. DOI: 10.1007/978-1-4419-9967-2_1.

(19) Torgovnick, A.; Schumacher, B. DNA repair mechanisms in cancer development and therapy. *Front. Genet.* **2015**, *6*, 157. DOI: 10.3389/fgene.2015.00157.

(20) Tu, J.; Chen, R.; Yang, Y.; Cao, W.; Xie, W. Suicide inactivation of the uracil DNA glycosylase UdgX by covalent complex formation. *Nat. Chem. Biol.* **2019**, *15* (6), 615-622. DOI: 10.1038/s41589-019-0290-x.

(21) Lenz, S. A. P.; Li, D.; Wetmore, S. D. Insights into the Direct Oxidative Repair of Etheno Lesions: MD and QM/MM Study on the Substrate Scope of ALKBH2 and AlkB. *DNA Repair* **2020**, *96*, 102944. DOI: 10.1016/j.dnarep.2020.102944.

(22) Lenz, S. A. P.; Wetmore, S. D. Evaluating the Substrate Selectivity of Alkyladenine DNA Glycosylase: The Synergistic Interplay of Active Site Flexibility and Water Reorganization. *Biochemistry* **2016**, *55* (5), 798-808. DOI: 10.1021/acs.biochem.5b01179.

(23) Mullins, E. A.; Rodriguez, A. A.; Bradley, N. P.; Eichman, B. F. Emerging Roles of DNA Glycosylases and the Base Excision Repair Pathway. *Trends Biochem. Sci.* **2019**, *44* (9), 765-781. DOI: 10.1016/j.tibs.2019.04.006.

(24) Dizdaroglu, M.; Coskun, E.; Jaruga, P. Repair of oxidatively induced DNA damage by DNA glycosylases: Mechanisms of action, substrate specificities and excision kinetics. *Mutat. Res.* **2017**, *771*, 99-127. DOI: 10.1016/j.mrrev.2017.02.001.

(25) Kumar, S.; Chinnusamy, V.; Mohapatra, T. Epigenetics of Modified DNA Bases: 5-Methylcytosine and Beyond. *Front. Genet.* **2018**, *9*, 640. DOI: 10.3389/fgene.2018.00640.

(26) Stivers, J. T.; Jiang, Y. L. A Mechanistic Perspective on the Chemistry of DNA Repair Glycosylases. *Chem. Rev.* **2003**, *103* (7), 2729-2760. DOI: 10.1021/cr010219b.

(27) Carusillo, A.; Mussolino, C. DNA Damage: From Threat to Treatment. *Cells* **2020**, *9* (7). DOI: 10.3390/cells9071665.

(28) Majumdar, C.; McKibbin, P. L.; Krajewski, A. E.; Manlove, A. H.; Lee, J. K.; David, S. S. Unique Hydrogen Bonding of Adenine with the Oxidatively Damaged Base 8-Oxoguanine Enables Specific Recognition and Repair by DNA Glycosylase MutY. *J. Am. Chem. Soc.* **2020**, *142* (48), 20340-20350. DOI: 10.1021/jacs.0c06767.

(29) Nikkel, D. J.; Wetmore, S. D. Distinctive Formation of a DNA–Protein Cross-Link during the Repair of DNA Oxidative Damage: Insights into Human Disease from MD Simulations and QM/MM Calculations. *J. Am. Chem. Soc.* **2023**, *145* (24), 13114-13125. DOI: 10.1021/jacs.3c01773.

(30) Liang, C.; Yang, Y.; Ning, P.; Chang, C.; Cao, W. Structural and functional coupling in cross-linking uracil-DNA glycosylase UDGX. *Biosci. Rep.* **2024**, *44* (1). DOI: 10.1042/bsr20231551.

(31) Duncan, B. K.; Miller, J. H. Mutagenic deamination of cytosine residues in DNA. *Nature* **1980**, *287* (5782), 560-561. DOI: 10.1038/287560a0.

(32) Fahrer, J.; Christmann, M. DNA Alkylation Damage by Nitrosamines and Relevant DNA Repair Pathways. In *Int. J. Mol. Sci.*, 2023; Vol. 24, p 4684.

(33) Sedgwick, B. Repairing DNA-methylation damage. *Nat. Rev. Mol. Cell Biol.* **2004**, *5* (2), 148-157. DOI: 10.1038/nrm1312.

(34) Guliaev, A. B.; Hang, B.; Singer, B. Structural insights by molecular dynamics simulations into differential repair efficiency for ethano-A versus etheno-A adducts by the human alkylpurine-DNA N -glycosylase. *Nucleic Acids Res.* **2002**, *30* (17), 3778-3787. DOI: 10.1093/nar/gkf494.

(35) Breiling, A.; Lyko, F. Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics Chromatin* **2015**, *8* (1), 24. DOI: 10.1186/s13072-015-0016-6.

(36) Nora M. Al Aboud, C. T., Ishwarlal Jialal. Genetics, Epigenetic Mechanism. *StatPearls* **2023**.

(37) Rodriguez, F.; Yushenova, I. A.; DiCorpo, D.; Arkhipova, I. R. Bacterial N4-methylcytosine as an epigenetic mark in eukaryotic DNA. *Nat. Commun.* **2022**, *13* (1), 1072. DOI: 10.1038/s41467-022-28471-w.

(38) Berti, P. J.; McCann, J. A. Toward a detailed understanding of base excision repair enzymes: transition state and mechanistic analyses of N-glycoside hydrolysis and N-glycoside transfer. *Chem. Rev.* **2006**, *106* (2), 506-555. DOI: 10.1021/cr040461t.

(39) Brooks, S. C.; Adhikary, S.; Rubinson, E. H.; Eichman, B. F. Recent advances in the structural mechanisms of DNA glycosylases. *Biochim. Biophys. Acta* **2013**, *1834* (1), 247-271. DOI: 10.1016/j.bbapap.2012.10.005.

(40) Kladova, O. A.; Kuznetsova, A. A. The Link Between Human Alkyladenine DNA Glycosylase and Cancer Development. *Int. J. Mol. Sci.* **2025**, *26* (15), 7647. DOI: 10.3390/ijms26157647.

(41) Liu, M.; Doublié, S.; Wallace, S. S. Neil3, the final frontier for the DNA glycosylases that recognize oxidative damage. *Mutat. Res.* **2013**, *743-744*, 4-11. DOI: 10.1016/j.mrfmmm.2012.12.003.

(42) Fromme, J. C.; Bruner, S. D.; Yang, W.; Karplus, M.; Verdine, G. L. Product-assisted catalysis in base-excision DNA repair. *Nat. Struct. Biol.* **2003**, *10* (3), 204-211. DOI: 10.1038/nsb902.

(43) Kaur, R.; Nikkel, D. J.; Aboelnga, M. M.; Wetmore, S. D. The Impact of DFT Functional, Cluster Model Size, and Implicit Solvation on the Structural Description of Single-Metal-Mediated DNA Phosphodiester Bond Cleavage: The Case Study of APE1. *J. Phys. Chem. B* **2022**, *126* (50), 10672-10683. DOI: 10.1021/acs.jpcb.2c06756.

(44) DeHart, K. M.; Hoitsma, N. M.; Thompson, S. H.; Borin, V. A.; Agarwal, P. K.; Freudenthal, B. D. APE1 active site residue Asn174 stabilizes the AP site and is essential for catalysis. *J. Biol. Chem.* **2025**, *301* (10), 110655. DOI: 10.1016/j.jbc.2025.110655.

(45) Jia, Q.; Zeng, H.; Tu, J.; Sun, L.; Cao, W.; Xie, W. Structural insights into an MsmUdgX mutant capable of both crosslinking and uracil excision capability. *DNA Repair* **2021**, *97*, 103008. DOI: 10.1016/j.dnarep.2020.103008.

(46) Aboelnga, M. M.; Wetmore, S. D. Unveiling a Single-Metal-Mediated Phosphodiester Bond Cleavage Mechanism for Nucleic Acids: A Multiscale Computational Investigation of a Human DNA Repair Enzyme. *J. Am. Chem. Soc.* **2019**, *141* (21), 8646-8656. DOI: 10.1021/jacs.9b03986.

(47) García-Díaz, M.; Bebenek, K.; Kunkel, T. A.; Blanco, L. Identification of an Intrinsic 5′-Deoxyribose-5-phosphate Lyase Activity in Human DNA Polymerase λ: A POSSIBLE ROLE IN BASE EXCISION REPAIR*. *J. Biol. Chem.* **2001**, *276* (37), 34659-34663. DOI: 10.1074/jbc.M106336200.

(48) Fu, D.; Samson, L. D. Direct repair of 3,N4-ethenocytosine by the human ALKBH2 dioxygenase is blocked by the AAG/MPG glycosylase. *DNA Repair* **2012**, *11* (1), 46-52. DOI: 10.1016/j.dnarep.2011.10.004.

(49) Kanazhevskaya, L. Y.; Gorbunov, A. A.; Lukina, M. V.; Smyshliaev, D. A.; Zhdanova, P. V.; Lomzov, A. A.; Koval, V. V. The Role of Key Amino Acids of the Human Fe(II)/2OG-Dependent Dioxygenase ALKBH3 in Structural Dynamics and Repair Activity toward Methylated DNA. *Int. J. Mol. Sci.* **2024**, *25* (2), 1145.

(50) Ma, J.; Qi, R.; Harcourt, Emily M.; Chen, Y.-T.; Barbosa, Giovannia M.; Peng, Z.; Howarth, S.; Delaney, S.; Li, D. 3,N4-Etheno-5-methylcytosine blocks TET1-3 oxidation but is repaired by ALKBH2, 3 and FTO. *Nucleic Acids Res.* **2024**, *52* (20), 12378-12389. DOI: 10.1093/nar/gkae818.

(51) Ringvoll, J.; Nordstrand, L. M.; Vågbø, C. B.; Talstad, V.; Reite, K.; Aas, P. A.; Lauritzen, K. H.; Liabakk, N. B.; Bjørk, A.; Doughty, R. W.; et al. Repair deficient mice reveal mABH2 as the primary oxidative demethylase for repairing 1meA and 3meC lesions in DNA. *EMBO J.* **2006**, *25* (10), 2189-2198. DOI: 10.1038/sj.emboj.7601109.

(52) Thompson, P. S.; Cortez, D. New insights into abasic site repair and tolerance. *DNA Repair* **2020**, *90*, 102866. DOI: 10.1016/j.dnarep.2020.102866.

(53) Shachar, S.; Ziv, O.; Avkin, S.; Adar, S.; Wittschieben, J.; Reissner, T.; Chaney, S.; Friedberg, E. C.; Wang, Z.; Carell, T.; et al. Two-polymerase mechanisms dictate error-free and error-prone translesion DNA synthesis in mammals. *EMBO J.* **2009**, *28* (4), 383-393. DOI: 10.1038/emboj.2008.281.

(54) Srinath, T.; Bharti, S. K.; Varshney, U. Substrate specificities and functional characterization of a thermo-tolerant uracil DNA glycosylase (UdgB) from Mycobacterium tuberculosis. *DNA Repair* **2007**, *6* (10), 1517-1528. DOI: 10.1016/j.dnarep.2007.05.001.

(55) Sang, P. B.; Srinath, T.; Patil, A. G.; Woo, E. J.; Varshney, U. A unique uracil-DNA binding protein of the uracil DNA glycosylase superfamily. *Nucleic Acids Res.* **2015**, *43* (17), 8452-8463. DOI: 10.1093/nar/gkv854.

(56) Yang, J.; Zhang, L.; Qiao, W.; Luo, Y. Mycobacterium tuberculosis: Pathogenesis and therapeutic targets. *MedComm* **2023**, *4* (5), e353. DOI: 10.1002/mco2.353.

(57) Cheng, Y.; He, C.; Wang, M.; Ma, X.; Mo, F.; Yang, S.; Han, J.; Wei, X. Targeting epigenetic regulators for cancer therapy: mechanisms and advances in clinical trials. *Signal Transduct. Target. Ther.* **2019**, *4* (1), 62. DOI: 10.1038/s41392-019-0095-0.

(58) Li, Q.; Zhu, Q. The role of demethylase AlkB homologs in cancer. *Front. Oncol.* **2023**, *13*, 1153463. DOI: 10.3389/fonc.2023.1153463.

(59) Kaur, R.; Frederickson, A.; Wetmore, S. D. Elucidation of the catalytic mechanism of a single-metal dependent homing endonuclease using QM and QM/MM approaches: the case study of I-PpoI. *Phys. Chem. Chem. Phys.* **2024**, *26* (11), 8919-8931. DOI: 10.1039/d3cp06201e.

(60) Li, Q.; Kang, C. Mechanisms of Action for Small Molecules Revealed by Structural Biology in Drug Discovery. *Int. J. Mol. Sci.* **2020**, *21* (15). DOI: 10.3390/ijms21155262.

(61) Ahn, W.-C.; Aroli, S.; Kim, J.-H.; Moon, J. H.; Lee, G. S.; Lee, M.-H.; Sang, P. B.; Oh, B.-H.; Varshney, U.; Woo, E.-J. Covalent binding of uracil DNA glycosylase UdgX to abasic DNA upon uracil excision. *Nat. Chem. Biol.* **2019**, *15* (6), 607-614. DOI: 10.1038/s41589-019-0289-3.

(62) Aroli, S.; Woo, E.-J.; Gopal, B.; Varshney, U. Mutational and structural analyses of UdgX: insights into the active site pocket architecture and its evolution. *Nucleic Acids Res.* **2023**, *51* (13), 6554-6565. DOI: 10.1093/nar/gkad486.

Chapter 2 : A Computational Investigation of the Uracil DNA Glycosylase *Msm*UdgX Catalytic Mechanism of Action Using MD and QM/MM Techniques

Contributions of Authors: Dylan Nikkel completed parameter set up of the iron-sulfur cluster for UdgX models. I (Angela Frederickson) performed molecular dynamics simulations for all UdgX systems, followed by data/results analysis, writing and editing, and generated representative molecular dynamics structures for quantum mechanics/molecular mechanics calculations, which were mapped by Basel Mansour. Stacey D. Wetmore was involved in project conceptualization and administration, supervision, data visualization and interpretation, funding acquisition, computational resource acquisition, and review/editing.

2.1. Introduction

The chemical stability and proper functioning of nucleic acids are essential for maintaining genetic information across all forms of life.[1, 2] However, DNA can be damaged naturally or from environmental or cellular stress.[3-5] Common forms of DNA damage include oxidation, alkylation, methylation, and deamination of nucleobases.[6, 7] A specific example is the deamination of cytosine to form uracil, which disrupts standard Watson-Crick-Franklin base pairing. This deamination reaction occurs spontaneously through hydrolysis, reactive nitrogen species, or oxidative processes.[6] If left unrepaired, the presence of uracil in DNA can cause G/C to A/T transition mutations during the replication process.[6, 8, 9]

To counteract the presence of uracil modifications, cells rely on the base excision repair (BER) pathway. The first step in the BER pathway is carried out by a DNA glycosylase. Monofunctional DNA glycosylases cleave the N–glycosidic bond to form an apurinic/apyrimidic (AP) or abasic site, which is subsequently processed by other enzymes to complete the repair and preserve genomic integrity.[10-13] Uracil DNA glycosylases (UDG) are monofunctional glycosylases that recognize and excise uracil from DNA by cleaving

the glycosidic bond found across diverse organisms, including viruses.[6, 14] UDGs can be grouped into six families based on sequence similarity and substrate preference.[14, 15] Despite low sequence conservation, all UDGs share a common α/β–fold and two key catalytic motifs: motif A, involved in activating a water molecule for nucleophilic attack, and motif B, which facilitates DNA intercalation.[10-13] Among members of the UDG family, Family 1 UNG (uracil–N–glycosylase), exemplified by *Escherichia coli* UNG, are highly efficient and display relatively narrow specificity, and were the first DNA glycosylases discovered.[10-13] Other families display more specialized activity. For example, Family 2 includes thymine DNA glycosylase (TDG) and mismatch-specific UDG (MUG) target T:G mismatches, while Family 3 (e.g. SMUG1) exhibits broader substrate tolerance and can excise uracil from both single- (ss-DNA) and double-stranded DNA (ds-DNA).[10-13] Families 4 and 5, common in thermophiles, contain [4Fe–4S] clusters and differ from Family 1 and 2 in substrate specificity and efficiency.[10-13] Family 4 also recognizes both ss-DNA and ds-DNA, while Family 5 only recognizes ds-DNA.[10] The most recently identified, Family 6 UDGs, primarily excise hypoxanthine.[13, 16] Together, these structural and functional differences reflect evolutionary adaptation while preserving a core repair mechanism.

Structural studies of UDGs from Families 1 to 3 suggest UDGs typically remove damaged or mismatched U through an S_N1-like hydrolysis mechanism that involves flipping the target base out of the DNA helix and cleaving the N–glycosidic bond.[17, 18] In the standard S_N1-like mechanism, glycosidic bond cleavage is initiated by departure of the of the damaged base through a dissociative transition state with significant oxo-carbenium-ion character at C1′.[11, 12] Nucleophilic attack by water occurs subsequently, hydrolyzing the resulting abasic intermediate and preparing the DNA for downstream repair.[12]

However, not all monofunctional glycosylases follow this classical mechanism.[12] For example, the human adenine DNA glycosylase homologue (MutY) removes adenine (A) mispaired with oxidative lesions such as 8–oxo–7,8-dihydroguanine (8oG) by forming a transient covalent intermediate with DNA and highlights an alternative catalytic pathway.[19, 20] A *mycobacterium smegmatis* UDG (*msm*UdgX) was recently identified as another notable exception as it has been proposed to form a stable DNA–protein crosslink; however, in this case the crosslink inactivates the enzyme rather than functioning as a transient intermediate.[10, 12, 18]

Although UdgX exhibits a distinct biochemical outcome, it retains several structural and catalytic features characteristic of the broader UDG superfamily. Structurally, UdgX shares similarities with Family 1 UDGs conserved canonical fold yet also resembles the Family 4 ability to recognize ss-DNA. Additionally, the [4Fe–4S] cluster occupies a position analogous to that observed in Family 4 UDGs; however, in UdgX one of the conserved cysteine ligands is replaced by histidine, suggesting evolutionary adaptation of the metal-binding motif. Additionally, active site residues such as E52, Q53, F65, N91, and H178 are highly conserved, supporting a shared substrate-binding strategy with Families 1 and 4.[10] Despite these similarities, UdgX diverges from canonical UDG behavior, emphasizing how subtle structural modifications within a conserved scaffold can lead to specialized enzymatic function.

UdgX is distinguished by a unique signature motif (AAGGKRRIH) and an unusual ability to form a stable covalent complex with the abasic (AP) site following uracil excision even under denaturing conditions.[12, 21] This unusual, irreversible crosslinking behavior is not observed in any of the six known UDG families, leading to the classification of UdgX as a member of a new UDG superfamily.[11, 12] This covalent linkage is formed through a process known as suicide inactivation, in which a catalytic histidine (H109) located in an extended motif 3 loop forms a stable bond with the C1′ of deoxyribose.[18] Structural and

mutational studies have shown that this reaction depends on the catalytic dyad H109 and E52, as well as on the precise positioning of the uracil base, which is supported by conserved residues such as F65 and N91 that are likely critical for stabilizing the reaction intermediate.[11, 12] Unlike conventional DNA glycosylases that dissociate from DNA after base excision, UdgX remains covalently bound to DNA, suggesting a potential role in protecting or marking damaged sites.[12, 18] Further insights from crystal structures and site-directed mutants (e.g., H109S, E52N, and R184A) have confirmed the importance of E52 and R184 in catalysis and DNA binding.[11, 13] UdgX also features a unique double-loop domain that encloses the DNA-binding pocket, restricting product release and reinforcing its single-turnover or "one-shot" behavior.[10] Nevertheless, the precise mechanistic roles of several active site residues remain unclear, leaving open questions about how UdgX coordinates uracil excision with irreversible crosslinking.

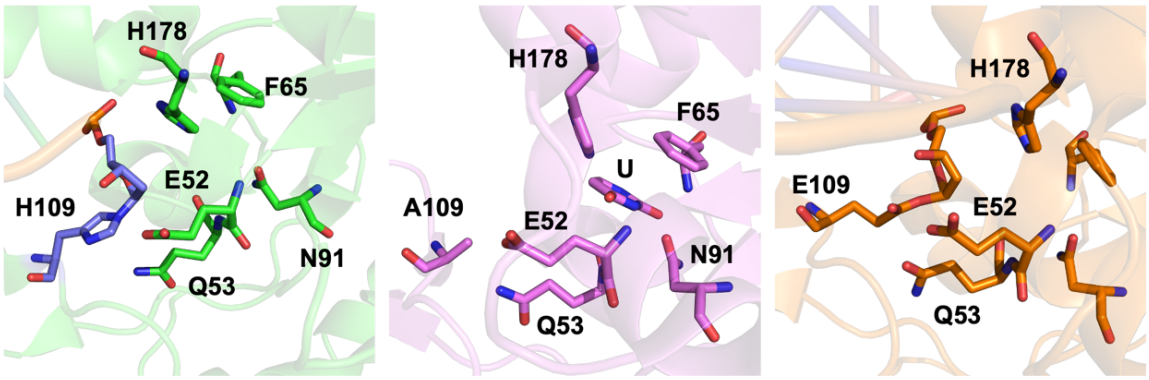


Figure 2.1. X-ray crystal structures of *msm*UdgX for the wild-type DNA–protein crosslink (PDB ID: 6IOD) (green, left), the H109A mutant unbound to DNA (PDB ID: 8IIG) (pink, middle), and the H109E mutant DNA–protein crosslink (PDB ID: 6L6S) (orange, right), highlighting active site residues.

The only available crystal structure of the UdgX–DNA wild-type enzyme is of the covalently crosslinked protein in complex with single-stranded DNA (PDB ID: 6IOD)[18] (Figure 2.1). Experimental studies proposed potential mechanisms for uracil excision

through crosslink formation by UdgX. Two crosslinking mechanistic pathways that differ in histidine tautomer were proposed. Both mechanisms involve the same key active site residues: H109, E52, Q53, F65, N91 and H178.[11-13] Specifically, in the first mechanism H178 facilitates uracil departure by hydrogen bonding to O2 of uracil, which increases the bond distance between C1′ of the sugar and the uracil base, while at the same time E52 is proposed to activate H109 in the HIE (Nε) tautomer for nucleophilic attack of C1′ by Nε, in a concerted mechanism (Figure 2.2).[11] The second mechanism is proposed to be a step-wise unimolecular substitution where uracil anion formation is supported by H178 and N91 hydrogen bonding to uracil, followed by E52 activating the H109 nucleophile that now adopts the HID (Nδ) tautomer.[13] Supporting the functional role of H109, mutational studies have shown that H109A (PDB ID: 8IIG), H109S (PDB ID: 8IIT), H109Q (PDB ID: 8IIP), H109G (PDB ID: 8IIL), H109K (PDB ID: 8IIN), or H109C (PDB ID: 8III) abolishes the crosslinking activity,[11] resulting in a standard hydrolytic uracil excision common to all other members of the UDG family. Interestingly, experimental mutation studies have also reported that H109E can undergo hydrolysis as well as crosslink formation where X-ray crystal structures have been resolved of the pre– and post–excision state without DNA, and the crosslink product bound with DNA (PDB ID: 6L6S, 6L5A and, 6L5B).[10] As this experimental evidence supports plausible reaction pathways for both tautomers, additional information is required to determine the unique mechanism of action of UdgX.

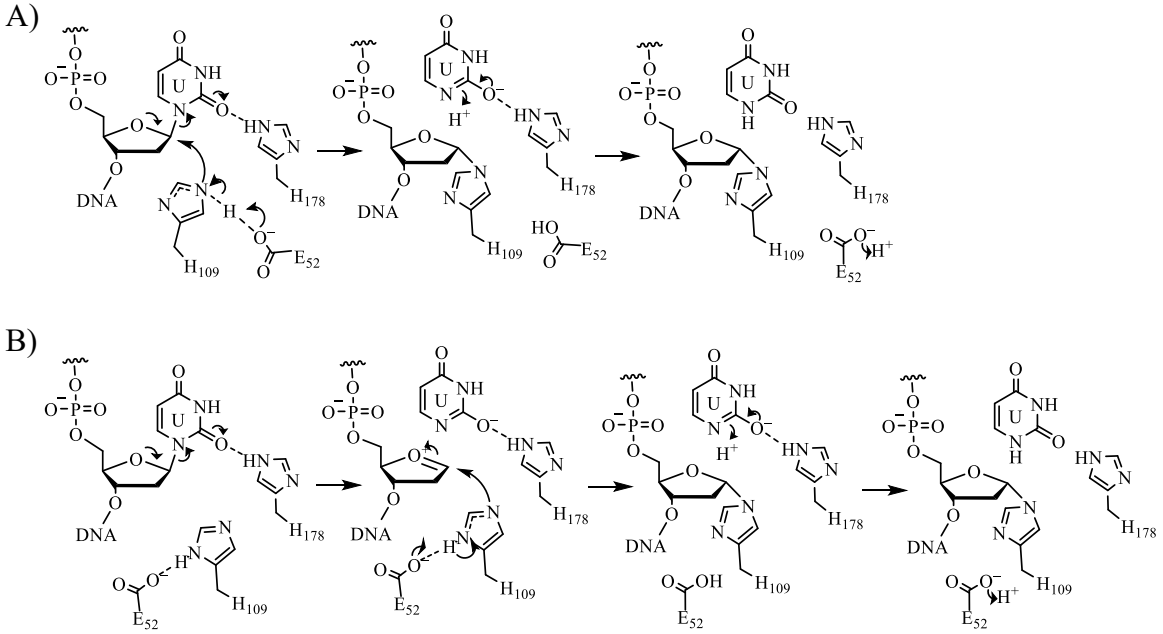


Figure 2.2. The proposed mechanisms of action of UdgX involving H109 forming a DNA–protein crosslink from two different tautomers. A) HIE109 tautomer where Nε is protonated and activated by E52 forming a crosslink with Nε via a concerted mechanism. B) HID109 tautomer where Nδ is protonated and activated by E52, enabling Nε to form the crosslink via an S_N1-like mechanism.

Due to the demonstrated success of computational studies in elucidating catalytic pathways of DNA glycosylases, this work aims to deepen our mechanistic understanding of UdgX. Previous molecular dynamics (MD) simulation and quantum mechanics and molecular mechanics (QM/MM) investigations of repair enzymes such as MutY,[19] TDG,[22] MBD4,[23] or canonical UDG[24] have provided critical insight into substrate positioning, nucleophile activation, and the electronic requirements of N–glycosidic bond cleavage.[6, 24, 25] Past studies have reconciled discrepancies between biochemical observations and structural data. Building on this foundation, atomistic MD simulations will be used herein to explore conformational states of UdgX that support catalytically-competent active site alignment, with particular emphasis on evaluating alternative tautomers of H109 and

defining the roles of conserved residues implicated in DNA–protein crosslink formation. Complementing these dynamic structural analyses, QM/MM calculations will be performed to characterize the thermochemistry of the reaction mechanism, including optimization of the reactant complex (RC), transition state (TS), and product complex (PC) for both wild-type UdgX and select mutants (i.e., H109A and H109E). By integrating dynamic sampling with electronic structure calculations, the proposed reaction pathways can be directly evaluated against available crystallographic,[10, 18] biochemical,[12] and mutational data.[10, 11] The reaction mechanism identified from these calculations is consistent with available structural and biochemical data, including experimentally observed crosslink formation[10-13, 18] and mutational phenotypes,[11, 13] and provides a mechanistic framework for rationalizing changes in catalytic activity arising from H109A and H109E substitutions.[11, 13, 18] A computational approach is particularly valuable for UdgX, where the unusual covalent DNA–protein intermediate complicates experimental characterization of transient states along the reaction coordinate. Furthermore, this work has implications beyond mechanistic clarification; in pathogens such as mycobacteria, where reliance on base excision repair is heightened and UdgX homologs are absent in humans, defining the structural and energetic determinants of catalysis may inform the design of selective antibacterial strategies. With a better understanding of the mechanism of action, the energetic features governing nucleophile activation and intermediate stabilization can be used to rationalize mutational effects, guide future experimental design, and establish a framework for targeting specialized DNA repair pathways. Together, these studies aim to provide a detailed atomic-level description of UdgX catalysis while highlighting how conserved glycosylase scaffolds evolve distinct chemical strategies.

2.2. Computational Methods

2.2.1. MD Simulation Protocol

An initial model of the wild-type RC was generated by replacing the sugar phosphate backbone (ORP) structure (PDB ID: 6IOD)[18] (Figure 2.1) with deoxyuridine using PyMOL 3.0.3.[26] The covalent bond to H109 was removed by repositioning the histidine ~2.2 Å from the C1′ of the nucleotide. Residues with partially unresolved side chains (K105, D166, K187, E189, R190, and E191) were manually added using the mutagenesis tool in PyMOL and refined to prevent steric clashes. Protonation states were assigned using the H++ webserver[27] for titratable residues with all residues in the expected state at pH 7.0 except HIE3, HIE98, HIE159, HIE180, HIP13, and HIP165. H109 was modeled in both the Nε (HIE) and the Nδ (HID) protonation tautomers. H109 in the pre-minimized wild-type model was mutated to H109A, H109S, H109E, H109Q, H109D, and H109K and visually inspected to ensure structural compatibility using the mutagenesis package in PyMOL. Each protein–DNA complex was solvated in TIP3P water molecules using the LEaP program in Amber21[28] to form a truncated rectangular box of water molecules, maintaining a minimum distance of 12 Å between the solute and box edge. The system was then neutralized with sodium ions, and additional NaCl was added to approximate physiological salt conditions (150 mM), with ion counts determined using the SLTCAP calculator[29] (see Table A.1, Appendix A, for water and ion counts).

All simulations used the AMBER ff14SB force field[30] for proteins and the OL15 force field[31] for DNA. Parameters for the [4Fe–4S] cluster and its coordinating residues (three cysteines and one histidine) were derived using the MCPB.py[32] module, employing Hartree-Fock (HF) calculations with the 6-31G(d) basis set. Missing parameters were

supplemented using the general AMBER force field (GAFF2)[33] and ions were modelled with Li/Merz[34] ion parameters (see Table A.2, Appendix A, for parameter details).

Energy minimization for each model was performed in four steps using pmemd of Amber21:[35] (1) solvent and ions were minimized with 100 kcal·mol^{-1}·Å^{-2} restraints on the solute, (2) restraints were lifted from solute hydrogen atoms, (3) solute atoms were minimized with restraints on the solvent, and (4) a final unrestrained minimization of the entire system was completed (1000 steps steepest descent followed by 2000 steps conjugate gradient). Heating to 310 K was achieved over six 50 K increments while applying 25 kcal·mol^{-1}·Å^{-2} restraints to the solute. Five-stage equilibration was performed with decreasing restraint strengths on the solute from 25 to 20, 15, 10, 5, and 1.5 kcal·mol^{-1}·Å^{-2}. Finally, 1000 ns unrestrained MD production simulations were performed in quintuplicate using pmemd.cuda of Amber21.[35] Simulations employed the NPT ensemble (310 K, 1 bar) with isotropic pressure scaling, a 10 Å cutoff for nonbonded interactions, the periodic boundary conditions, and TIP3P water density fixed at ~1 g mL^{-1}.

2.2.2. MD Simulation Analysis

Trajectory frames were saved every 20 ps for subsequent analysis, yielding approximately 20,000 frames per system across all replicas. All trajectory analyses were performed using the cpptraj module implemented in Amber21.[35] Custom analysis scripts were developed for each system to quantify key structural descriptors of the active site to evaluate residue alignment for catalysis, including interatomic distances, bond angles, and dihedral angles. Hydrogen-bond analyses were conducted using cpptraj, with hydrogen bonds defined by a heavy atom-heavy atom distance ≤ 3.4 Å and a donor-hydrogen-

acceptor angle between 130° and 180°. Structural stability over the course of the 1000 ns simulation was assessed by calculating the root-mean-square deviation (RMSD) backbone of the enzyme–DNA complex, the enzyme or DNA, or all active site (H109, E52, Q53, N91, F65, and H178) heavy atoms. Representative structures for each system were obtained by using the cpptraj cluster command based on distance or RMSD metrics tailored to the specific system. Additional analyses were performed to evaluate water occupancy within the active site throughout the production simulations with the criteria being less than 3.4 Å away from E52 oxygens or C1′ of dU. Water density maps were also generated from the MD trajectories using cpptraj module of the AmberTools package by combining all replicates and structurally aligning the trajectory to the first frame using the heavy atoms of the protein. Water density was computed by binning the positions of water oxygen atoms from all frames onto a three-dimensional grid with a spacing of 0.45 Å encompassing the protein (E52, Q53, H109, F65, H178, and dU). For each grid voxel, water occupancy was averaged across the trajectory to yield a normalized three-dimensional water density distribution. The resulting density maps were visualized as isodensity meshes in PyMOL to identify persistent hydration sites within the active site.

2.2.3. QM/MM Calculations

QM/MM calculations were performed starting from snapshots obtained from the MD representative structures. Our own N–layered integrated molecular orbital and molecular mechanics (ONIOM) method was employed due to its proven effectiveness in modeling the chemistry of DNA-target enzymes, including other monofunctional glycosylases (MutY, MBD4, and TDG).[19, 22, 23, 25] For the wild-type mechanism, the QM region included

the nucleotide bearing the target uracil base (U216), the adjacent 3′ phosphate (A217), and the active site residues H109, E52, Q53, and H178, along with two active site water molecules (71 atoms) (Figure A.3, Appendix A). The net charge of the QM region was −2. The MM region comprised the remaining UdgX–DNA complex and surrounding water molecules within 6 Å of the enzyme or DNA (1465 water molecules and an overall charge of –13). For the H109A mutant, the QM region was similar, except H109 was replaced by A109, and one water molecule was in the active site in the MD representative structure (70 atoms and −2 charge). The H109E mutant structure was taken from a snapshot of the optimized QM/MM RC structure, therefore QM region was nearly identical to the wild-type as H109 was replaced by E109 (70 atoms and −3 charge). The MM region of the mutant models contained 1618 water molecules and the same overall charge (−14). The QM/MM boundary was defined between the Cα and Cβ atoms of amino acids and between the C4′ and C5′ atoms of nucleotides A216 and A217. Hydrogen link atoms were used to cap the QM region at these boundaries. The QM region was treated at the M06-2X/6-31G(d,p) level of theory, while the MM region was described using the same AMBER force field as employed for the MD simulations. A mechanical embedding (ME)[19, 36, 37] scheme was used based on the robustness of this approach and its successes in accurately modeling other enzymatic reactions, including comparable results to the electrostatic embedding (EE) scheme for other glycosylases that use crosslinking (MutY)[19] or hydrolysis (MBD4)[23] mechanisms. Transition states were found by scanning the bond between N_{HIS} and C1′. All atom geometry optimizations were carried out for each system. Frequency calculations were performed to verify the nature of stationary points and obtain thermal corrections for the Gibbs energy single-point calculations. Transition states were

validated by a single imaginary frequency and intrinsic reaction coordinate (IRC) analysis. Relative energies were obtained using ONIOM(M06-2X/6-311+G(2df,p):AMBER14SB). ME scheme used throughout for all QM/MM calculations performed using Gaussian 16 (Revision B.01).[38]

2.3. Results and Discussion

2.3.1. MD Simulations Reveal Active Site Alignment for Crosslink Formation in the HID, but Not HIE, H109 Tautomer

For UdgX to catalyze dU cleavage through DNA–protein crosslink formation, H109 must be positioned to attack C1′ while the departing dU must be stabilized by surrounding active site residues in the DNA–UdgX complex. As described in the methods, HIE and HID are the two tautomers evaluated in this context. Previous studies have suggested plausible reaction pathways for both H109 tautomers that converge on an Nε–C1′ covalent-linkage, consistent with the available crystal structure containing the crosslinked product.[18] To assess the effects of different tautomers and identify the catalytically active state, MD simulations of the DNA–UdgX RC were performed.

2.3.1.1. HIE

The HIE tautomer was first investigated by evaluating the active site conformation. The trajectory was clustered based on the distance between Nε in H109 and C1′ in dU (Figure 2.3). For HIE, there is a single broad peak spanning distances of 3.2 to 8.2 Å. This indicates substantial conformational variation in the relative position of H109 and dU. The dominant

cluster (Cluster 1) accounted for approximately 75% of the trajectory, with the protonated nitrogen (Nε) oriented toward C1′ at an average distance of 4.3 ± 0.4 Å. Although E52 was proposed to hydrogen-bond with the protonated nitrogen (HE2),[13] this interaction was only present for 21% of Cluster 1. Instead, E52 persistently formed a nonpolar hydrogen-bond with the neighboring C–H (HE1). Cluster 2, comprising ~23% of the simulation, showed an increase in the Nε–C1′ distance to 5.7 ± 0.3 Å and disrupted E52–HD2 and E52–HE1 hydrogen-bonding interactions, while Cluster 3 (~2%) displayed a further extended Nε–C1′ separation of 6.8 ± 0.5 Å and similarly disrupted hydrogen-bonding to H109.

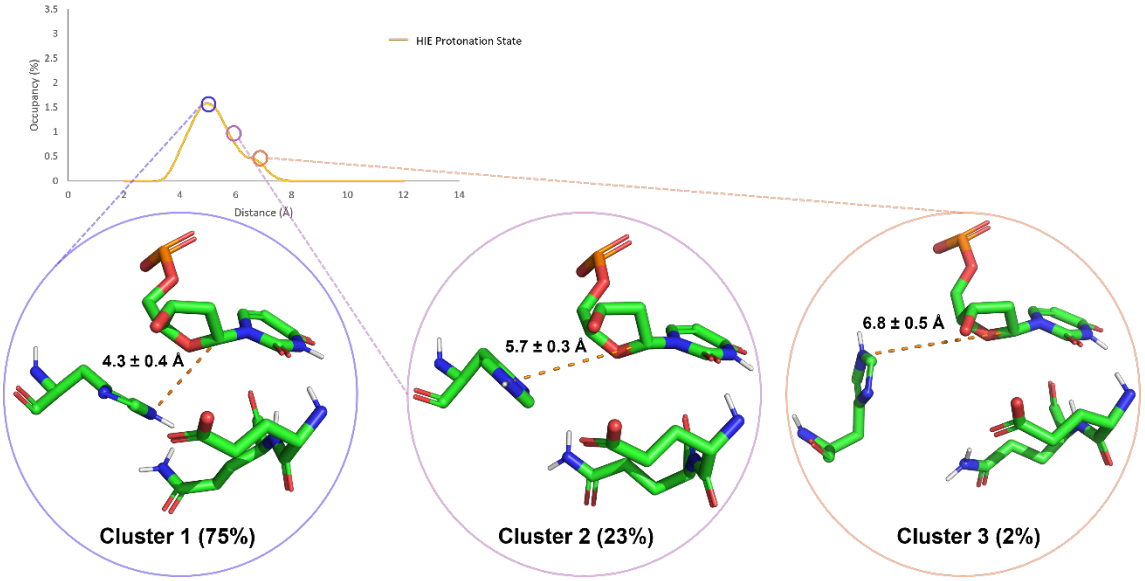


Figure 2.3. The percent occupancy of UdgX HIE109 tautomer clustered according to the distance of C1′ to Nε measured over the course of the MD simulation replicates. Cluster 1 displays the highest occupancy at 75% of the simulation time based on the closest alignment of H109 and C1′ (left). Cluster 2 displays an occupancy of 23% and much weaker alignment between H109 and C1′ (middle). Cluster 3 displays the weakest and furthest alignment of H109 and C1′ which occupies only 2% of the simulation time (right).

Two dihedral angles of H109 were analyzed to characterize the variable conformations of the imidazole ring (∠(ND1–CG–Cβ–Cα); RING) and the side-chain linker (∠(CG–Cβ–Cα–N); ARM). Two principal ring orientations were observed corresponding to a ring tilt of ± 71° (Figure 2.4). The side-chain linker (ARM) adopted three general conformers that determined whether the ring was tilted away from dU or toward dU. This yields a total of six conformational states, indicating substantial conformational freedom within the H109 side chain. Together, these observations reflect a dynamic ensemble of conformations over the simulations.

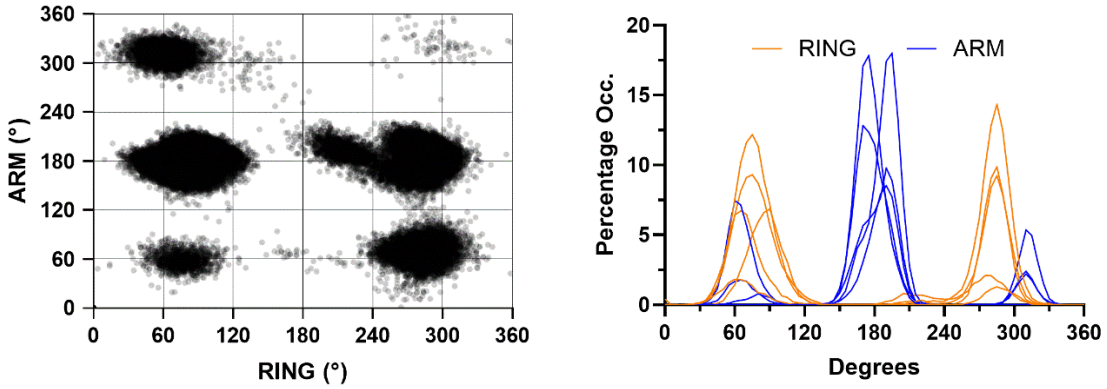


Figure 2.4. HIE representative dihedral angles plotted on a 360° range to describe the possible conformations observed. HIE (∠(ND1–CG–Cβ–Cα); RING) (x-axis) and (∠(CG–Cβ–Cα–N); ARM) (y-axis) dihedral angles plotted to show highest populated conformations for all 5 replicates (left). RING (orange) and ARM (blue) dihedral angles on a histogram highlighting the percent occupancy of each position over all trajectories and replicates (right).

Experimental crystallographic data has identified other key residues in the active site including E52, Q53, F65, N91, and H178.[12, 18] UdgX, within the UDG superfamily, retains conserved structural motifs and active site residue interactions associated with uracil processing that are also present in Family 4 UDG homologs.[10-13] Specifically, similar to Family 4 UDGs, [13] the uracil base in the active site is recognized by a bidentate interaction

with N91, involving a hydrogen bond from uracil N3–H to the side chain oxygen of N91 paired with a reciprocal hydrogen bond from N91 ND2–H to uracil O2. Additional hydrogen bonds are formed from the side chain of H178 (N–H) to uracil O2. Finally, the backbone amide of F65 (N–H) donates a hydrogen bond to uracil O4, while the F65 aromatic side chain forms a π-stacking interaction with the uracil base. MD simulations on the DNA–UdgX complex suggest these interactions are preserved, each with an occupancy of >86% (Figure 2.5). These hydrogen bonds corroborate the involvement of direct side-chain contacts from H178 and N91 in uracil recognition and leaving group departure. In contrast, the backbones of residues E52 and Q53 have been previously implicated in uracil base-recognition through backbone-mediated hydrogen bonding.[10, 11] Persistent interactions between the backbone of E52 (94%) or Q53 (63%) and dU were observed. Moreover, our simulations indicate that the influence of E52 and Q53 on H109 is less direct than suggested by experimentally proposed details. Specifically, E52 participates in non-classical C–H···O hydrogen bonding with the imidazole C–H of H109 and OE1 of E52 (25%), while occasionally forming a conventional N–H···O interaction with the histidine Nε atom for (16%) (Figure A.4, Appendix A). Although E52–H109 coupling has been proposed to be a key element of a catalytic dyad to activate the nucleophile (Figure 2.2),[11, 12] the dynamic behavior observed here suggests a more nuanced role in modulating the active site geometry. This suggests E52 more subtly contributes to active site stabilization by regulating the local electrostatic environment and aiding H109 crosslink alignment rather than directly anchoring the catalytic histidine through a strong hydrogen bond. Q53 also appears to indirectly support H109 positioning, reinforcing the geometry of the R–

loop and the active site rather than forming dominant side-chain hydrogen bonds to H109, as proposed experimentally.

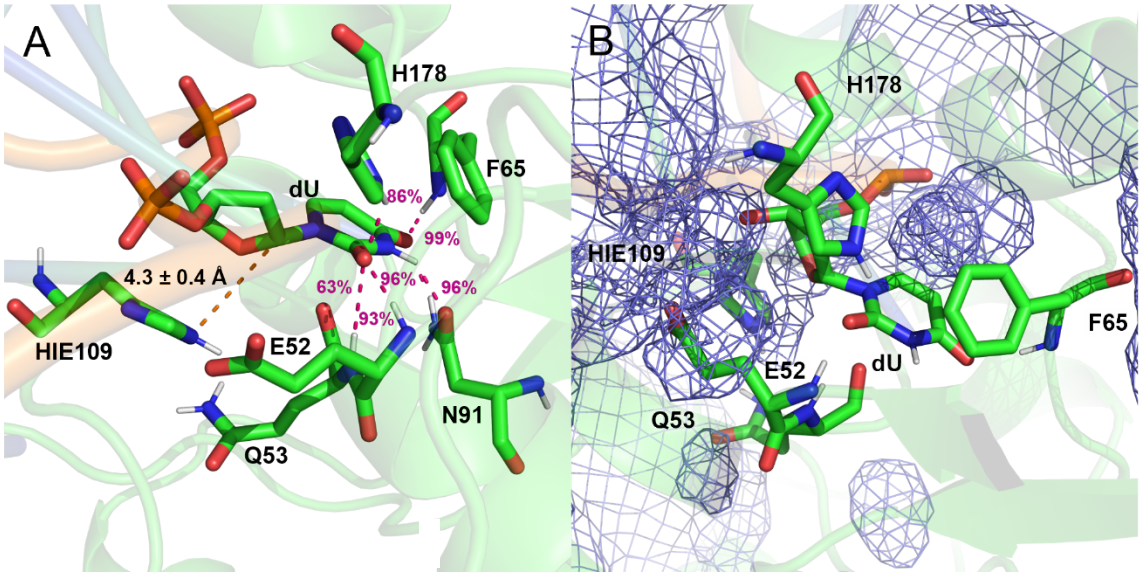


Figure 2.5. UdgX active site of wild-type HIE tautomer with corresponding C1'–H109 distance, stabilizing residues, and water density. A) The percent occupancies of hydrogen bonds observed for structurally supportive residues over the course of MD simulations. B) High density 0.6 grade grid iso-mesh plot of water that frequents the active site over the simulation of all quintuplicates.

Since the active site was not aligned for crosslink formation, the distribution of active site water was examined to determine whether hydrolysis was a viable alternative as all other UDGs known to date catalyze hydrolysis.[39-41] Structured water molecules frequently act as bridging partners between amino acid residues and substrates, completing interaction networks that protein alone cannot provide. Analysis across all simulation replicates showed a water occupancy of approximately 98% between E52 and C1′ of dU. However, the water near C1′ and within the immediate vicinity of the uracil base in the HIE system was disrupted and exhibited a higher density surrounding E52 (Figure 2.5B). Collectively, the lack of clear water positioning, inconsistent E52 interactions with H109, and imidazole

flexibility does not align with the previously proposed mechanism in which E52 acts as a general base to deprotonate the attacking H109 nitrogen and facilitate nucleophilic attack (Figure 2.2A). As a result, the UdgX–DNA complex in the HIE state is likely not catalytically conducive for protein–DNA crosslink and was not considered further.

2.3.1.2. HID

The HID tautomer was next examined, which aligns with proposals from prior experimental studies (Figure 2.2B).[11, 12] MD simulations of the UdgX–DNA complex containing H109 in the HID tautomer were clustered according to the Nε–C1′ distance. Compared to the HIE model, HID clustering shows three distinct peaks. The first narrow peak (Cluster 1) corresponds to Nε falling within 3.5 ± 0.2 Å of dU (C1′) for 49% of the simulation time (Figure 2.6). It has been proposed that E52 may act as a nucleophilic activator when H109 adopts the HID tautomer;[11] however, this behavior was not supported by the present simulations. Specifically, across all five replicas, no hydrogen-bonding was observed between either carboxylate oxygen of E52 (OE1 or OE2) and the protonated Nδ atom of H109. Instead, as found for the HIE system, E52 primarily interacts with the imidazole ring through hydrogen bonding to a neighboring C–H group, an interaction observed predominantly in Cluster 1. In Cluster 2, which represents approximately 35% of the sampled conformations, disruption of the nonpolar E52–HE1 interaction coincides with an increase in the Nε–C1′ distance of 5.3 ± 0.4 Å. In Cluster 3 (∼15% of the ensemble), this separation increases further to 7.8 ± 0.8 Å, accompanied by continued loss of stabilizing contacts between E52 and H109. The Cluster 2 and 3 configurations are geometrically unfavorable for nucleophilic attack. Despite the lack of direct E52-mediated

activation, the HID tautomer intrinsically satisfies the electronic requirements for nucleophilic participation in Cluster 1. Deprotonation at Nδ allows for resonance delocalization of negative charge across the imidazole ring, a feature that enhances nucleophilicity at Nε. Consistent with this electronic configuration, transient Nε–C1′ distances below 4 Å were observed (Figure A.5, Appendix A), which are necessary for crosslink alignment, indicating that HID109 can intermittently adopt geometries potentially compatible with DNA–protein crosslink formation.

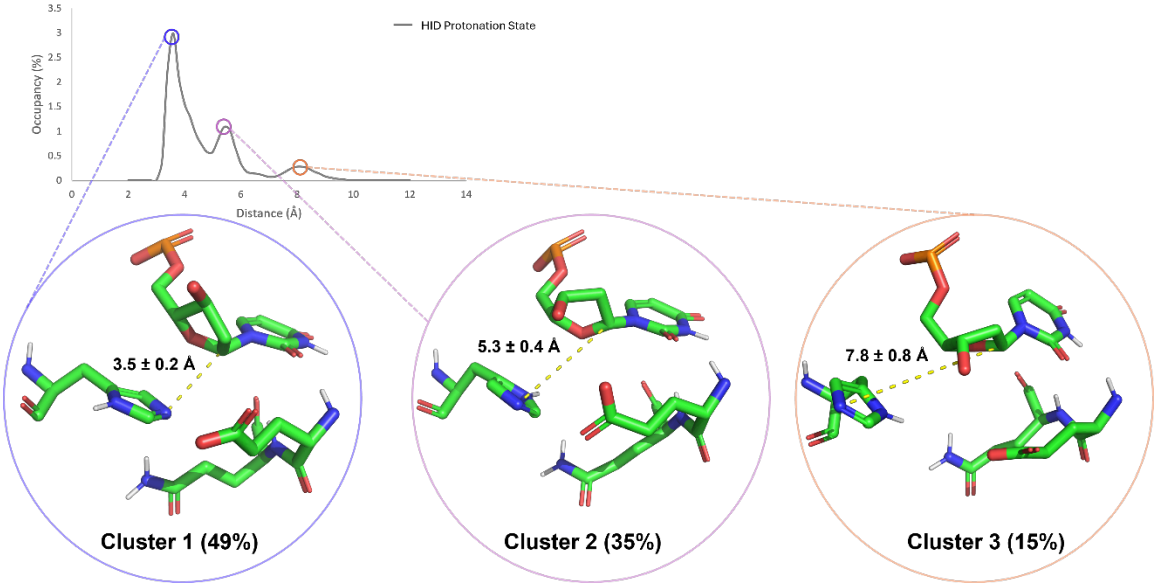


Figure 2.6. The percent occupancy of UdgX HID109 tautomer clustered according to the distance of C1′ to Nε measured over the course of the MD simulation replicates. Cluster 1 displays the highest occupancy at 49% of the simulation time based on the closest alignment of H109 and C1′ (left). Cluster 2 displays an occupancy of 35% and weaker alignment between H109 and C1′ (middle). Cluster 3 displays the weakest and furthest alignment of H109 and C1′ which only occupies about 15% of the simulation time (right).

Dihedral angle analysis distinguishes the conformational behavior of the HID state (Figure 2.7). The imidazole RING dihedral samples two dominant orientations corresponding to two dominant orientations centered near ~90° and ~270°, corresponding

to opposite tilts of the imidazole ring relative to the side chain, with minimal occupancy of intermediate angles. The ARM dihedral primarily occupies the conformation near ~180°, with smaller populations in ~60 and ~300° regions. The two-dimensional distribution reveals discrete clusters corresponding to specific combinations of ring and linker orientations, indicating that the HID side chain adopts a limited set of preferred rotameric states. Together, these features indicate that the HID state samples a less broad conformational ensemble, allowing H109 to adopt orientations that facilitate proper catalytic alignment within the active site.

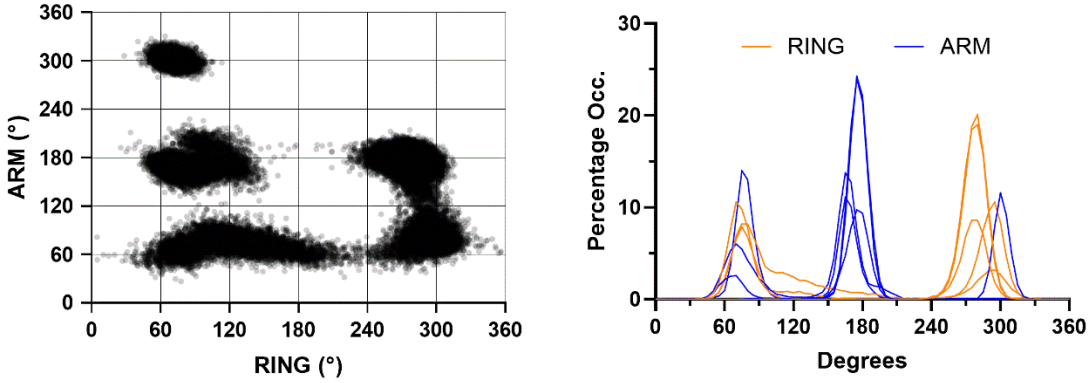


Figure 2.7. HID representative dihedral angles plotted on a 360° range to describe the possible conformations observed. HID (∠(ND1–CG–Cβ–Cα); RING) (x-axis) and (∠(CG–Cβ–Cα–N); ARM) (y-axis) dihedral angles plotted to show highest populated conformations for all 5 replicates (left). RING (orange) and ARM (blue) dihedral angles on a histogram highlighting the percent occupancy of each position over all trajectories and replicates (right).

Hydrogen bonding between residues H178, N91, and F65 and the departing uracil was present at a slightly higher percent for HID, compared to the other tautomer, with occupancies of 92%, 93%, and 99% of the simulation, respectively (Figure 2.8A) (see Table A.4, Appendix A). The E52 backbone amide (N–H) hydrogens bond with dU for 94% of the simulation. Unique to HID, Q53 appears to play a larger role in supporting H109

alignment in the active site by hydrogen bonding with Nε for 95% of the simulation at an average distance of 1.9 Å. These interactions suggest that the HID tautomer promotes a more aligned and pronounced hydrogen bonding network within the active site compared to the HIE tautomer.

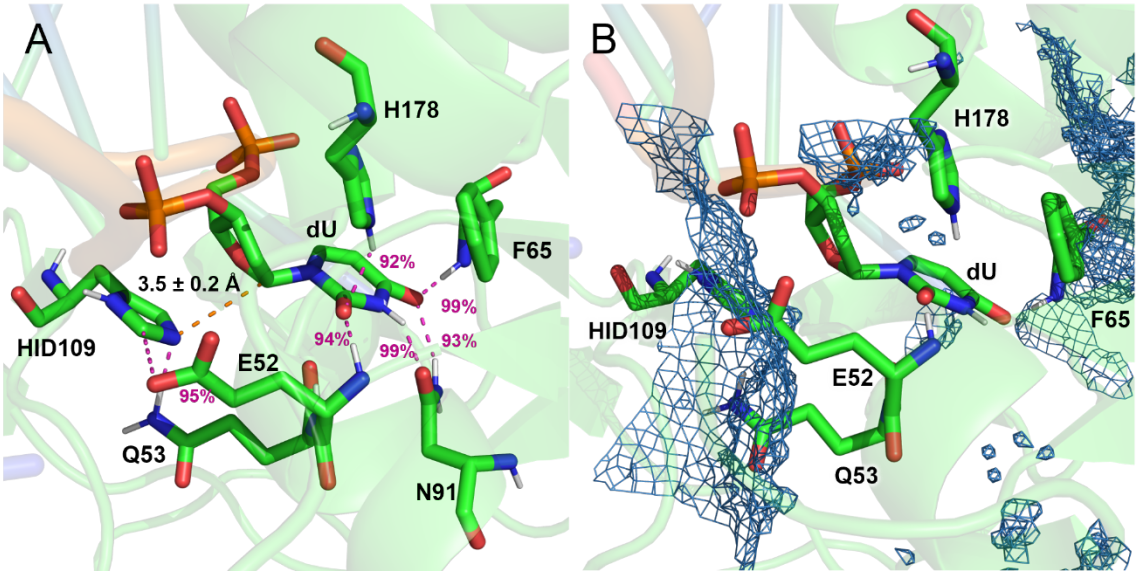


Figure 2.8. UdgX active site of wild-type the HID tautomer with corresponding C1'–H109 distance and stabilizing residues. (A) The percent occupancies of hydrogen bonds displayed by structurally supportive residues over the course of MD simulations. (B) High density 0.45 Å mesh plot of water in position to aid nucleophilic attack of E52.

Because structured water networks frequently contribute to substrate positioning in enzymatic active sites and water was found in the active site for the HIE model, we examined whether water molecules participate in aligning key residues involved in UdgX catalysis. Water density analysis determined there was a high density of water between E52 and H109 (Figure 2.8B). Notably, the area surrounding dU was generally void of water, potentially making room for the leaving group to depart. Indeed, water molecules occupy the region surrounding E52, the 5′–phosphate backbone, and H109 throughout the simulation. Specifically, a bridging water network was observed between the phosphate

backbone and the oxygens of E52 (OE1 and OE2), with near-continuous occupancy (~100%) (Figure 2.9). This ordered water network may contribute to stabilizing the active site geometry and promoting alignment of H109 into a catalytically competent orientation for crosslink formation. Collectively, the combination of H109–dU distance and reduced imidazole flexibility suggest HID is the preferred tautomer over HIE.

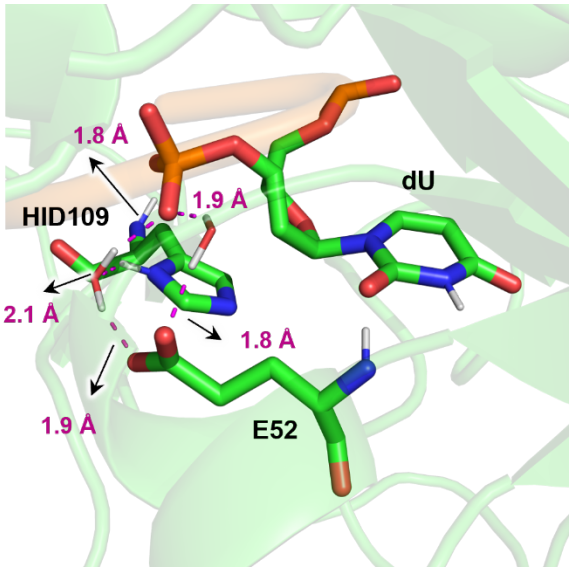


Figure 2.9. UdgX HID active site, highlighting an MD representative structure of bridging water molecules located in high water density areas. Bridging water molecules were found between E52 (OE1 and OE2) and the 5′-phosphate backbone, while also hydrogen bonding to Nδ–H of H109.

2.3.2. QM/MM of HID109 Wild-Type Displays Concerted Mechanism of Action for Irreversible Crosslink Formation

A representative MD snapshot of the HID (Cluster 1) model was used to build a QM/MM model to map the catalytic mechanism of UdgX (Figure 2.10). In the RC, the imidazole ring of H109 is positioned to initiate nucleophilic attack at C1′ of dU. This

process is assisted by a water molecule that is readily available between H109 and E52 in the MD representative structure, and the side chain of E52. Specifically, E52 accepts a proton from the water molecule, which in turn facilitates deprotonation of H109 at Nδ. This is supported by a hydrogen-bond distance of 1.869 Å between the H–OH and the E52 side chain (OE1), and a 2.062 Å hydrogen-bond distance between the same proton and Nδ of H109. The Nδ–H bond in H109 is 1.020 Å. E52 not only assists the activation of H109 but also forms a hydrogen-bond with C–H in the H109 imidazole ring (HE1) (2.494 Å). This interaction anchors and orients H109 precisely for nucleophilic attack at dU, corroborating observations from MD simulation (Figure 2.10). At the same time, the backbone of E52 (N–H) and the side chain of H178 (N–H) structurally stabilizes uracil via separate hydrogen-bonds to O2 (1.868 Å and 2.062 Å). Additionally in the low layer, N91, and F65, are well positioned to hydrogen bond and π–π stack with the uracil base. These results reinforced the MD findings, highlighting the essential roles of H109, E52, Q53, and H178 in the reaction (see Figure A.9, Appendix A, for more details).

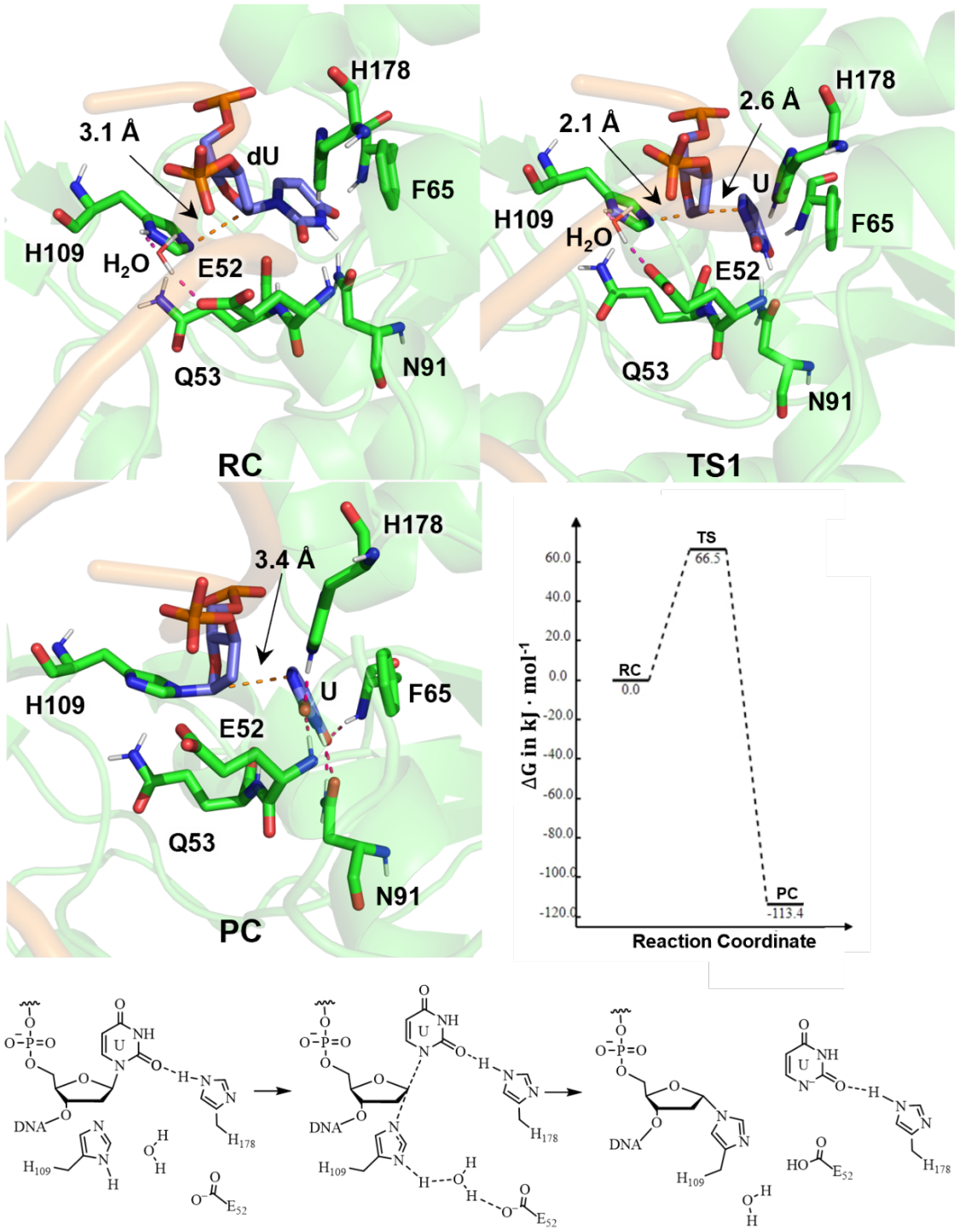


Figure 2.10. Catalytic mechanism of action of HID109 wild type using QM/MM calculations. Detailed distance and hydrogen-bonding interactions of the RC, TS, and PC. Accompanied by a Gibbs energy level diagram displaying the energy barrier to reach an intermediate complex and detailed mechanism of action.

In the TS, the departing uracil moiety is stabilized by a dense hydrogen-bonding network involving H178–Nε (1.756 Å), the E52 backbone N–H (1.916 Å), and the N91 side-chain of N–H and carbonyl oxygen (1.916 Å and 1.741 Å, respectively). Compared to the RC, where the H178–Nε to uracil distance was slightly longer (2.062 Å), the shortened distance in the TS underscores the role of H178 in stabilizing the developing oxyanion. Similarly, the hydrogen-bond distances between E52, H_2O, and H109 become tighter in the TS, with (r(E52–H_2O)) decreasing from 1.869 Å to 1.711 Å and (r(H_2O–H109–Nδ)) from 2.067 Å to 1.685 Å. The bond between C1′ and the leaving nitrogen (N1) also significantly lengthens from 1.526 Å to 2.573 Å.

In the PC, the glycosidic bond has been cleaved, and a covalent crosslink has formed between H109 and the C1′ carbon of the sugar moiety. The hydrogen-bond between H178 and the oxyanion of the uracil base becomes even stronger (0.031 Å shorter), enhancing stabilization of the departing U anion. Meanwhile, E52 completes deprotonation of H109 through the bridging water molecule. The overall barrier for crosslink formation is 66.5 $kJ \cdot mol^{-1}$ and the irreversible PC falls −113.4 $kJ \cdot mol^{-1}$ lower in Gibbs energy than the RC (Figure 2.10). Compared to experimental kinetic data, wild-type UdgX exhibits a second-order rate constant (k_2/K_m) of 1.4×10^4 M^{-1} s^{-1}, corresponding to an apparent activation Gibbs energy of approximately 52–54 $kJ \cdot mol^{-1}$ using a 1 M standard state at 37 °C.[13]

Early experimental studies describe the catalytic mechanism of the UdgX crosslink formation to be S_N1-like based largely on the stabilization of the uracil base in the active site, the formation of a long-lived covalent DNA–protein complex, and the absence of detectable reaction intermediates. [12, 13] Since direct characterization of the chemical transition state is experimentally inaccessible, this S_N1-like determination was inferred

from the overall reaction outcomes and comparisons to canonical UDGs, which are known to stabilize positive charge development at the anomeric carbon during glycosidic bond cleavage.[11] This interpretation reflects the dissociative nature of the reaction rather than definitive evidence for a fully stepwise pathway involving a stable intermediate. Subsequent structural studies and biochemical analysis proposed that UdgX catalysis proceeds through a concerted mechanism, in which glycosidic bond cleavage and H109 crosslink formation are mechanistically coupled.[13] These studies showed that the enzyme active site is preorganized so that H109 is positioned in close proximity to C1′ prior to uracil departure. This was corroborated by mutational studies that suggest any disruption of H109 abolishes the covalent crosslink formation without preventing uracil excision. Our computational potential energy surface provides direct mechanistic resolution of these different experimental proposals. We identify a single dominant TS structure in which the elongated C1′ glycosidic bond and the formation of the H109–C1′ crosslink simultaneously occur, demonstrating that the rate-limiting step encompasses both the bond breaking and forming events. Thus, while earlier studies correctly identify the dissociative nature of UdgX catalysis,[11] our computational results refine the picture by demonstrating a concerted transition state.

Most monofunctional DNA glycosylases promote glycosidic bond cleavage through hydrolysis, ultimately generating an abasic site that is further processed within the BER pathway. Mechanistically, these enzymes typically favour dissociative reaction pathways that allow formation of an oxocarbenium-like transition state, which is subsequently trapped by a nucleophilic water molecule. Studies on enzymes such as MutY[19] illustrate that even when covalent DNA–protein crosslink interactions are possible, these species

function as transient intermediates that are resolved to regenerate the enzyme and enable catalytic turnover. In contrast, UdgX adopts a distinct catalytic method in which nucleophilic attack by H109 is mechanistically coupled to glycosidic bond cleavage, leading to formation of a stable irreversible covalent DNA–protein crosslink rather than productive hydrolysis. This divergence reflects differences in active site preorganization and solvent accessibility that push the reaction pathway toward irreversible enzyme–DNA trapping. UdgX represents an example of catalytic tuning in which active site preorganization redirects a conserved glycosylase scaffold from hydrolytic repair toward irreversible lesion trapping, generating a persistent DNA–protein crosslink that may serve to mark and stabilize sites of damage for downstream processing.

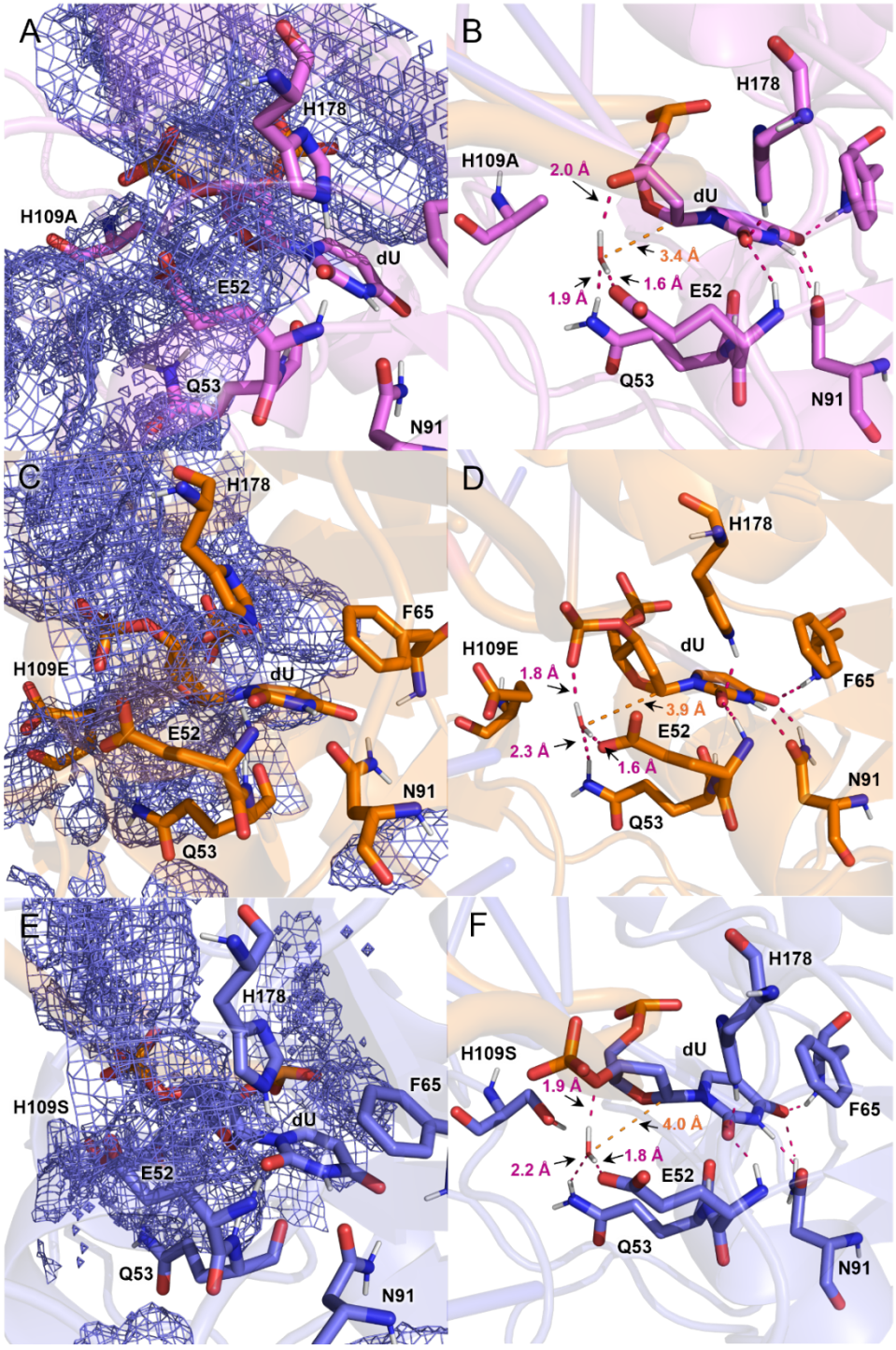
2.3.3. MD Simulations of UdgX Mutants Reveal Active Site Architecture that Points to a Hydrolysis Pathway Selection

Previous experimental investigations using biochemical uracil excision assays, steady-state kinetic analysis, and X-ray crystallography characterization of H109 mutants have established that substitution of H109 abolishes DNA crosslink formation and redirects UdgX activity towards a canonical hydrolysis mechanism.[11] Systematic structural and kinetic analysis further revealed a relationship between active site cavity volume and uracil excision activity, in which modest cavity enlargement initially supports increased hydrolysis activity (G and S mutants), followed by a decline in activity as the cavity size increases (C and A mutants).[11] These effects were attributed to changes in the ability of the active site to localize and position a nucleophilic water molecule near C1′ of dU, highlighting the sensitivity of mutant UdgX activity on solvent organization.

To understand changes in the structural dynamics upon mutation, representative structures of UdgX–DNA complexes H109A, H109E, H109S, H109Q, H109D, and H109K were obtained by RMSD-based clustering of MD replicates. As an initial assessment of crosslink competency, distances between the mutated 109 residue and C1′ atom of dU were measured and ranged from 5.1–6.4 Å revealing a loss of the precise positioning required for covalent bond formation. A factor of this geometric separation comes from the cavity formed between the mutant side-chain, E52, Q53, and C1′ of dU, consistent with structural trends previously reported.[11] The distance between E52 (OE2) and C1′ was monitored as a structural descriptor of active site accessibility to facilitate comparison with experimental data, across all mutants the active site cavity revealed 5.1–5.5 Å (Table A.6, Appendix A).

Given the large cavity volume, solvent accessibility was assessed. All mutants displayed a highly hydrated active site, including water occupancy surrounding deoxyribose and the uracil leaving group. Water is also consistently located between E52 and C1′ of dU across all mutants (∼100% occupancy) (Figure 2.11). The increased cavity volume permits rapid solvent exchange measured by identifying all water molecules that occupy the space (75–274 molecules varying between mutants), resulting in a more dynamic and less ordered hydration environment. In correlation to experimental interpretations, the enlargement of the cavity from mutations like H109A destabilizes the ordered water network. The cavity was large enough in H109A to accommodate and stabilize a water molecule for nucleophilic attack, while not being too large to affect hydrogen bonding between water, E52 (OE2), and Q53 (N–H) necessary for hydrolysis. A representative H109A structure revealed water is positioned at 3.3 ± 0.3 Å from C1′ of dU, stabilized through hydrogen

bonding with E52 and Q53, and further supported by interactions with the phosphate backbone (Figure 2.11B). Hydrogen bonding that supports the uracil leaving group remained largely conserved relative to the wild-type enzyme (>92% occupancy), including bidentate interactions between uracil (O3 and N4) and N91, as well as contacts between O2 of uracil with E52 and H178, and backbone hydrogen bonds with F65. These dynamics represent a well hydrated and potentially well aligned active site architecture across the H109 mutants (H109A, H109E, H109S, H109Q, H109K, and H109D), with H109A serving as a representative model for solvent-mediated catalysis.


A
H178
H109A
dU
E52
Q53
N91
B
H178
H109A
2.0 Å
dU
3.4 Å
1.6 Å
1.9 Å
E52
Q53
N91
C
H178
F65
H109E
dU
E52
N91
Q53
D
H178
H109E
dU
1.8 Å
F65
3.9 Å
E52
2.3 Å
1.6 Å
N91
Q53
E
H178
F65
H109S
dU
E52
Q53
N91
F
H178
H109S
dU
1.9 Å
F65
4.0 Å
1.8 Å
2.2 Å
E52
Q53
N91

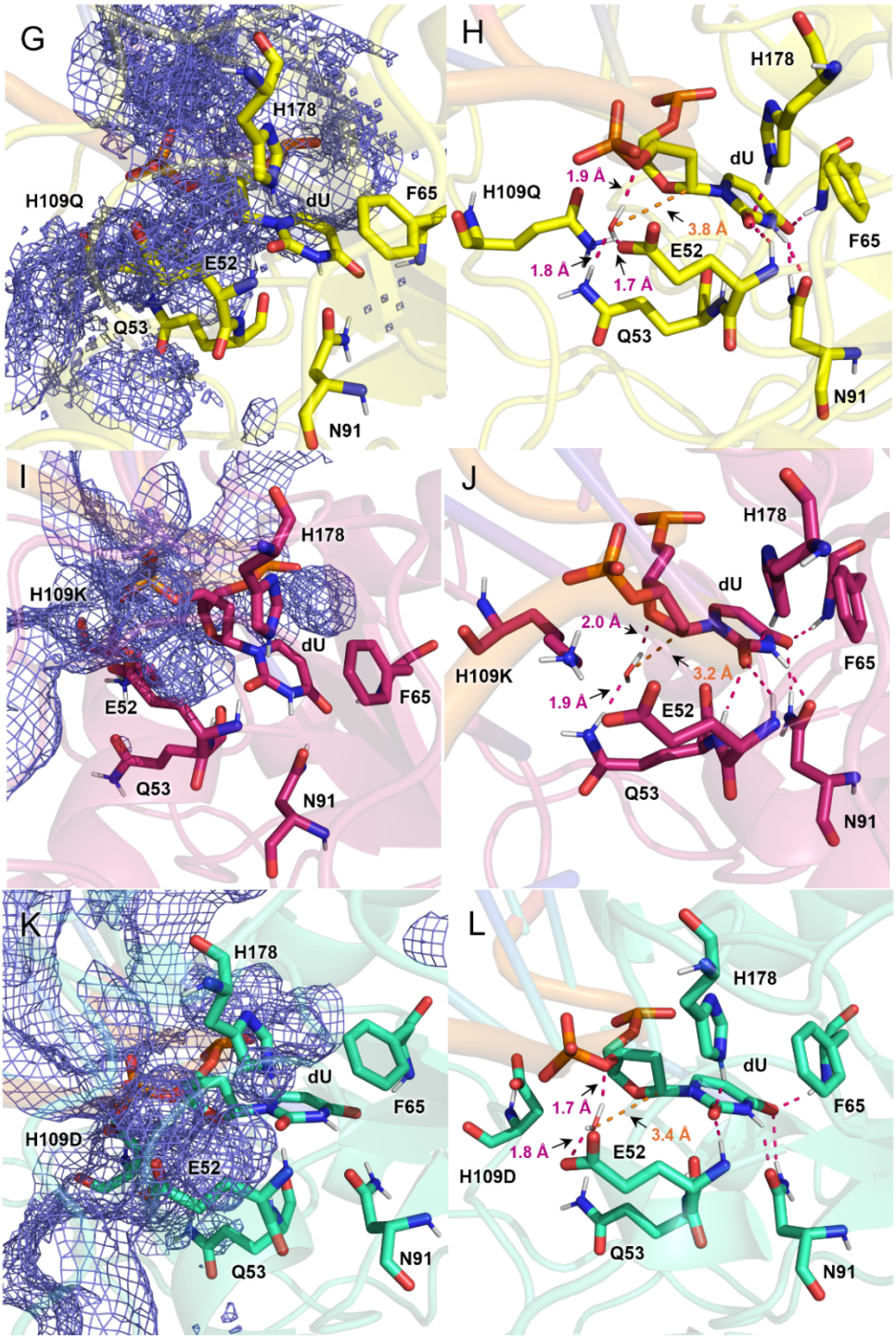


Figure 2.11. Representative structure from MD simulations of the H109A, H109E, H109S, H109Q, H109K and H109D mutants. (A) H109A (C) H109E (E) H109S (G) H109Q (I) H109K (K) H109D; water density mesh plot to display water in the active site over the course of all replicates to display water availability in the system. (B) H109A (D) H109E (F) H109S (H) H109Q (J) H109K (L) H109D; structure and stabilizing residues and the percentage occupancy of hydrogen bonds over the course of the simulations.

In contrast to the majority of H109 mutations that undergo hydrolysis mechanisms, H109E has been experimentally reported to have limited crosslinking abilities, as evidenced by detection of both DNA–bound and free enzyme populations in biochemical assays.[10] Structural studies were capable of capturing H109E in crosslinked, pre-excision and post-excision states (PDB ID: 6L6S and 6L5A/B) (Figure A.6, Appendix A), suggesting that this mutant may occupy a mechanistic intermediate between canonical hydrolysis and covalent crosslink formation.[10] Despite these observations, MD simulations initiated from multiple starting points failed to identify conformations that could support crosslink formation (see Appendix A for details). Similar to the other mutants, <100 ns into the MD production simulation, the E109–C1′ distance exceeded 4 Å, forming the solvent accessible cavity (Figure 2.11D and Figure A.7, Appendix A). Furthermore, H109E consistently adopted conformations that supported the positioning of a nucleophilic water molecule in the active site, resembling the hydrolysis competent alignment observed for H109A. In the MD representative structure (Figure 2.11D), the E109 side chain rotates away from the C1′, increasing solvent accessibility and allowing a water molecule to occupy the catalytic pocket at a distance of approximately 3.9 Å from C1′. This water molecule was supported by hydrogen-bonding to OE2 of E52 (1.6 Å) and to a non-bridging oxygen of the 5′-phosphate backbone (1.8 Å), while also forming a hydrogen-bond with Q53 (2.3 Å). Together, these interactions position the water molecule for potential nucleophilic attack on the anomeric carbon following glycosidic bond cleavage. Consistent with hydrolysis conformations of other H109 mutants, the departing uracil base remains flipped into the active site through interactions with the backbone of E52 and F65, as well as hydrogen bonding with the side chains of H178 and N91. These results suggest that

while H109E is predisposed to undergo hydrolysis within the time-scale accessible in MD simulation, its experimentally observed crosslinking capability may arise from electronically driven effects that are not fully captured by dynamic structure.

Overall, MD simulations of the H109 mutants demonstrate that disruption of the histidine nucleophile consistently reorganizes the active site architecture toward a solvent-accessible, hydrolysis-competent conformation. Loss of precise positioning of residue 109 with the C1′ atom results in formation of an enlarged cavity that promotes increased hydration and dynamic solvent movement. Across all mutants, persistent water occupancy between E52 and C1′, coupled with conserved hydrogen-bonding interactions stabilizing the uracil leaving group, supports a structural environment favorable for water-mediated deglycosylation. In particular, H109A exemplifies how moderate cavity enlargement still allows proper localization of a nucleophilic water molecule while preserving the hydrogen-bonding network necessary for hydrolysis. Although H109E has been experimentally shown to retain limited crosslinking ability, MD simulations indicate that its dominant conformation resembles that of hydrolysis-competent mutants, suggesting that crosslink formation in this variant may depend more strongly on electronic factors than on structural preorganization. Collectively, these findings highlight that reaction pathway selection in UdgX is governed not solely by residue identity at position 109, but by the balance between active site geometry, solvent organization, and hydrogen-bonding support.

2.3.4. QM/MM Comparison Shows Feasibility of Hydrolysis for Mutants with Possibility of Crosslinking for H109E

Representative structures of H109A and H109E were selected for QM/MM calculations to explicitly compare the activation barriers associated with hydrolysis (H109A) and crosslink formation (H109E) in comparison to the H109 wild-type enzyme. These mutants enable direct evaluation of whether crosslink activity is electronically feasible for H109E despite its apparent structural flexibility, while providing insight to the framework for linking active site structure, solvent organization, and catalytic energetics across divergent reaction pathways.

2.3.4.1. H109A Displays Energetically Feasible Hydrolysis Mechanism

A representative MD snapshot of the H109A model was used to build a QM/MM model to map the catalytic mechanism (Figure 2.12). In the RC, E52 is positioned to initiate nucleophilic attack by positioning a water molecule in alignment to attack at C1′ of dU. This process occurs because a water molecule is readily available between E52 and C1′ in the MD representative structure. Specifically, E52 accepts a proton from the water molecule, which in turn facilitates and initiates hydrolysis. This is supported by a hydrogen-bond distance of 1.482 Å between the H–OH and the E52 side chain (OE2), a 2.096 Å hydrogen-bond distance between the water and the 5′-phosphate backbone, and a 1.739 Å hydrogen-bond between Q53 (N–H) and the water oxygen. The H–OH bond distance to C1′ is 3.061 Å and the C1′ to uracil bond of the RC is 1.537 Å. Within the H109A hydrolysis pathway, the uracil base is stabilized by an extensive hydrogen-bonding network involving conserved active site residues. The O2 position of the uracil moiety

forms interactions with the backbone amides of E52 (2.084 Å) and Q53 (2.277 Å), as well as the Nε atom of H178 (1.856 Å), contributing to anchoring of the nucleobase within the active site. Additional support is provided by low level residues N91, which engages both the uracil base and surrounding functional groups through hydrogen-bonds to the nucleobase (1.831 Å) and the O4 position (1.889 Å), and the aromatic residue F65 through a backbone hydrogen-bond with O4 (1.800 Å), suggesting a combined hydrogen-bonding and stacking contribution to substrate orientation. Together, this dense interaction network supports positioning of the uracil moiety in the absence of H109 nucleophilic participation and likely assists in maintaining a catalytically competent alignment for water-mediated hydrolysis. Notably, the same network was observed in the wild-type enzyme highlighting the essential roles of E52, F65, N91, and H178 in the reaction.

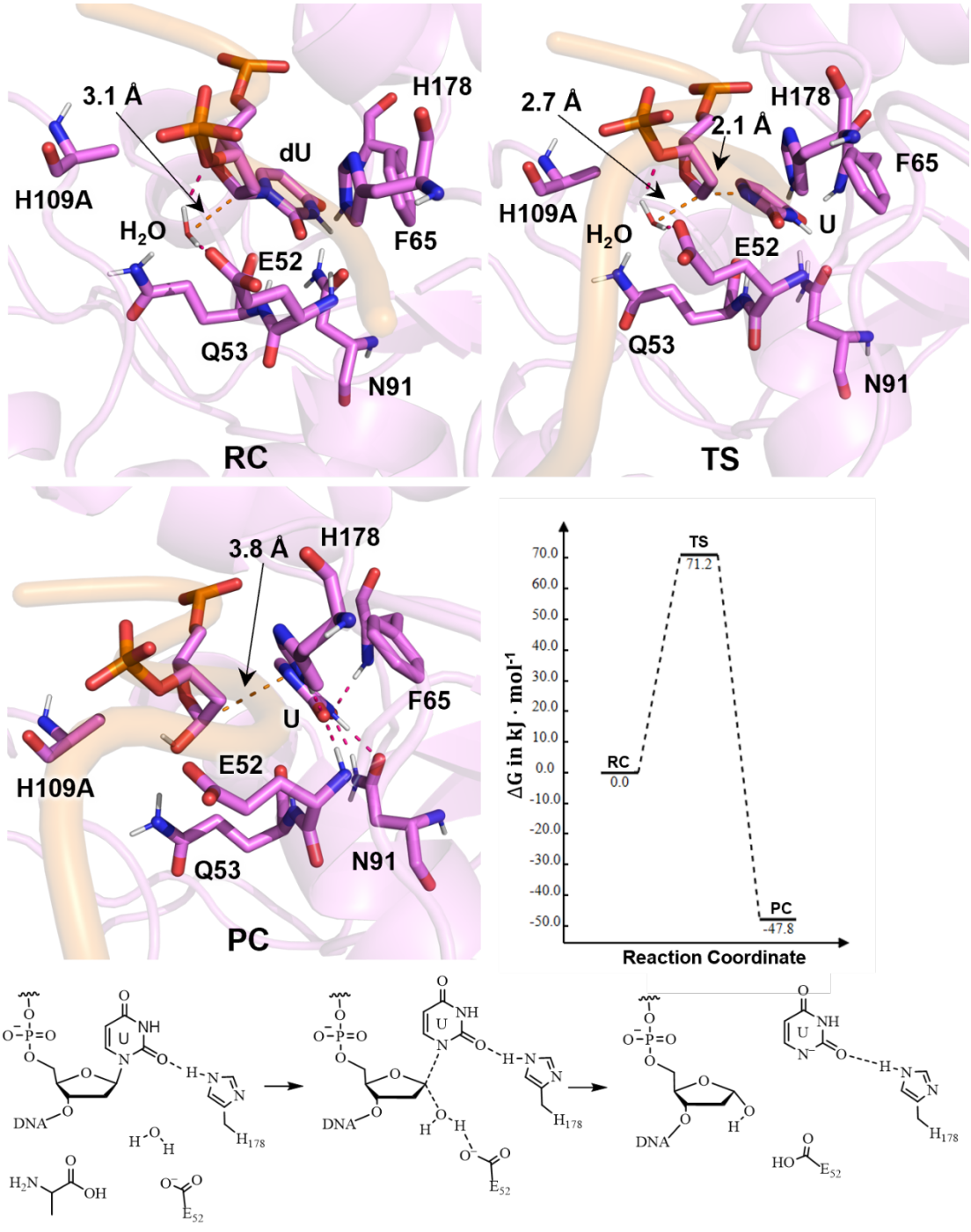


Figure 2.12. Catalytic mechanism of action of H109A mutant undergoing hydrolysis reaction using QM/MM calculations. Accompanied by a Gibbs energy level diagram displaying energy barrier to reach intermediate complex.

In the TS, the uracil moiety is stabilized by a dense hydrogen-bonding network involving H178–Nε (1.681 Å), the E52 backbone N–H (1.996 Å), and the N91 side-chain of N–H and carbonyl oxygen (1.806 Å and 1.890 Å, respectively). The water and Q53 hydrogen-bond is lengthened by 0.200 Å. Compared to the RC, the H178–Nε to uracil distance is shortened in the TS (1.681 Å), which underscores the role of H178 in adding in leaving group departure. Concurrently, the attacking water molecule approaches the C1′ carbon (2.670 Å) while remaining hydrogen-bonded to the phosphate backbone (2.067 Å), E52 OE2 (1.479 Å), and Q53 N–H (1.741 Å). This indicates a structured solvent network that supports nucleophilic alignment. Consistent with progression toward hydrolysis, the forming O–C1′ interaction shortens relative to the RC, while the C1′–N1 glycosidic bond elongates to 2.100 Å, reflecting advancement toward bond cleavage.

In the PC, the glycosidic bond has been cleaved, and the sugar moiety has been hydrated, resulting in clear separation between the ribose C1′ and the departing uracil base producing an abasic site. Consistent with product formation, the C1′–N1 distance increases substantially to 3.822 Å, confirming dissociation of the nucleobase from the sugar. The ordered catalytic water network observed in the TS was no longer present, as hydrogen-bonds between the water molecule and the phosphate backbone, E52 OE2, and Q53 N–H are absent in the product state, indicating completion of nucleophilic attack and collapse of the transition-state hydration environment. Despite this solvent reorganization, the uracil anion remains stabilized within the active site through a conserved hydrogen-bonding network involving the E52 backbone (1.944 Å) and the H178 (1.606 Å) side chain with uracil O2. Notably, the interaction between H178–Nε and the uracil O2 further shortens

relative to the TS. Additionally, low layer N91 (1.735 Å and 1.943 Å) and F65 (1.913 Å) interactions with uracil are maintained.

The overall barrier for hydrolysis is 71.2 kJ·mol^{-1} and the PC falls −47.8 kJ·mol^{-1} lower in Gibbs energy than the RC (Figure 2.12). Experimentally, the H109A mutant exhibits diminished catalytic efficiency relative to wild-type UdgX (k_{cat}/K_m = 0.131 (x10^{-3}) min^{-1} nM^{-1} at 37 °C), corresponding to an apparent activation Gibbs energy of ∼56.7 kJ·mol^{-1} estimated using the Eyring equation.[11, 13] The higher experimental barrier likely reflects that k_{cat} reports an overall rate-limiting process that can include steps beyond the intrinsic chemical barrier captured in the computed profile.

The QM/MM reaction profiles for the H109A mutant (Figure 2.12) revealed a stark difference in energetics and mechanistic outcomes compared to the wild-type UdgX, highlighting the catalytic significance of H109. In the wild-type model, the initial nucleophilic attack by H109 proceeds through a TS at 66.5 kJ·mol^{-1}, leading to a highly stabilized PC (relative Gibbs energy of −113.4 kJ·mol^{-1}). In contrast, the H109A mutant relies on a water molecule that is activated by E52 to attack at C1′ of dU. Although the overall pathway resembles that of the wild-type, the absence of H109 leads to a less stabilized product (PC: −47.8 kJ·mol^{-1}) and an albeit slightly larger activation barrier for TS (71.2 kJ·mol^{-1}). This variability likely arises from minor changes in key stabilizing interactions: such as the π–π stacking with F65 and weaker hydrogen-bonds with E52 and Q53. Specifically, in the H109A model, the hydrogen-bond distance between E52–backbone NH and the substrate increase to 1.996 Å, from 1.916 Å in the wild-type. Although a slight strengthening is observed in the F65–backbone NH interaction (1.800 Å

vs. 1.845 Å in the wild-type), it is not sufficient to compensate for the destabilization caused by the absence of H109.

These results underscore the variable ability in wild-type and mutant UdgX catalytic function. The wild-type enzyme not only facilitates a low-barrier nucleophilic attack and stabilizes key intermediates but also ensures completion of the reaction through an energetically favorable pathway. However, the reaction can also proceed with uracil cleaving activity when UdgX is mutated. The H109A mutant proceeds via a complete hydrolysis mechanism, reinforcing the variable nature of the 109 position to proceed with uracil cleaving activity. Importantly, these computational findings are in strong agreement with experimental observations.[11]

2.3.4.2. H109E Displays an Energetically Feasible Crosslink Mechanism

Although MD simulations of the H109E mutant predominantly sampled conformations consistent with hydrolysis, experimental evidence demonstrates that the mutant is capable of both hydrolysis and covalent crosslink formation. To investigate the feasibility of the crosslink pathway, an alternative modeling strategy was therefore employed. A representative structure for the H109E mutant was generated by mutating H109 to glutamate within the optimized QM/MM RC structure of the HID model and subsequently mapping the crosslinking pathway (Figure 2.13 and Figure A.8, Appendix A). In the RC, the glutamate side chain is positioned near the anomeric C1′ of the uracil nucleoside, with an E–C1′ distance of 3.497 Å, while the glycosidic bond remains intact (C1′–N1 = 1.554 Å). The uracil base is stabilized through the same conserved hydrogen-bonding network as the wild-type model, involving E52, Q53, N91, F65, and H178. Specifically, backbone

interactions from E52 (1.934 Å) and Q53 (2.084 Å) to uracil O2, along with hydrogen-bonds involving N91 (1.809 Å and 1.953 Å) and F65 (1.900 Å), maintain proper substrate alignment. In contrast to the wild-type RC, the H178–uracil O2 interaction is longer (2.844 Å), suggesting reduced stabilization of the uracil oxyanion prior to activation. Supporting interactions between Q53 N–H (1.804 Å) or water with E109 OE2 (2.043 Å) further help position the side chain for nucleophilic attack.

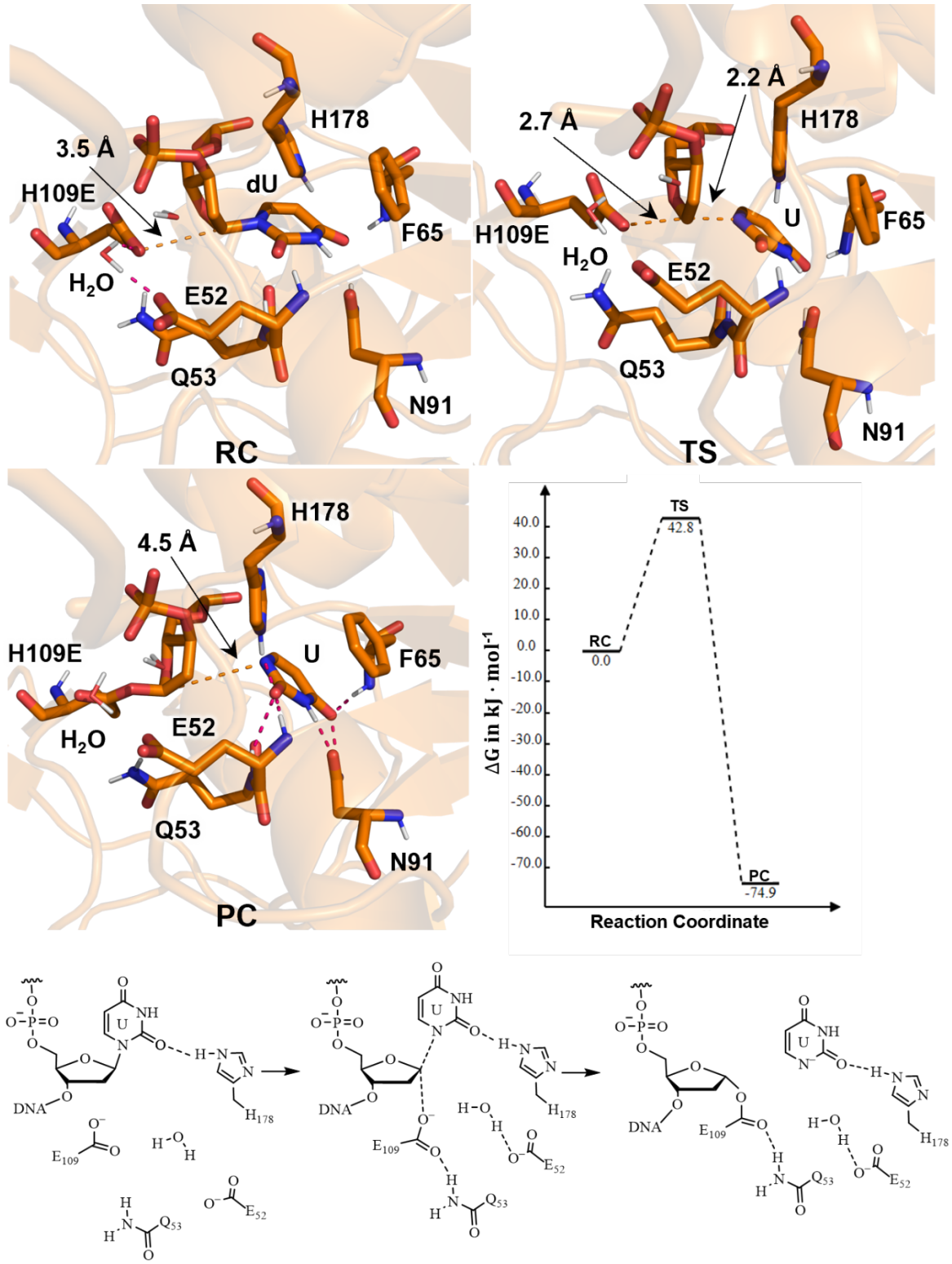


Figure 2.13. Catalytic mechanism of action of H109E mutant undergoing crosslink formation reaction using QM/MM calculations. Accompanied by a Gibbs energy diagram displaying the energy barrier to reach intermediate complex.

In the TS, substantial rearrangement occurs as the glutamate approaches the anomeric center (E–C1′ = 2.671 Å) while the glycosidic bond elongates to 2.177 Å, indicating progression toward crosslink formation. The uracil moiety is stabilized by a dense hydrogen-bonding network involving H178–Nε (1.879 Å), E52 backbone N–H (1.862 Å), and N91 (1.797 Å and 1.926 Å). Compared to the RC, the marked shortening of the H178 interaction highlights its role in stabilizing developing oxyanion character during bond cleavage. Interactions involving Q53 and F65 remain conserved (2.387 Å and 1.859 Å, respectively), maintaining stacking and hydrogen-bonding interactions that preserve substrate positioning throughout the reaction coordinate. In the PC, the glycosidic bond is fully cleaved (C1′–N1 = 4.494 Å) and a covalent linkage is formed between the glutamate side chain and C1′ (E–C1′ = 1.432 Å), demonstrating that the H109E mutant is capable of crosslink formation despite replacement of the histidine nucleophile. The hydrogen-bonding network surrounding uracil in the RC persists, with strong interactions involving E52 (1.957 Å), N91 (1.756 Å and 1.891 Å), F65 (1.882 Å), and H178 (1.798 Å), suggesting continued stabilization of the departing uracil anion. Additionally, a Q53 N–H to E52 OE1 interaction (1.652 Å) emerges in the PC, indicating reorganization of the local hydrogen-bonding environment following crosslink formation. Together, these results demonstrate that although H109E alters the active site chemistry, the conserved residue network can still support a crosslinking mechanism analogous to the wild-type pathway. The overall barrier for crosslink formation is 42.8 kJ·mol^{-1} and the PC falls −74.9 kJ·mol^{-1} lower in Gibbs energy than the RC (Figure 2.13). Experimental kinetic analyses show that the H109E mutant retains substantial catalytic activity relative to wild-type UdgX, despite a modest reduction in overall efficiency (Figure 2.13). For the natural G/U substrate, H109E

displays a k_2/K_m of 4.3×10^3 M^{-1} s^{-1}, corresponding to an approximately threefold decrease compared to wild-type. A similar trend is observed for the A/U substrate, where H109E exhibits a k_2/K_m of 3.9×10^3 M^{-1} s^{-1}, which correspond to apparent activation Gibbs energies of approximately ~55.6 kJ·mol^{-1}, representing a ~2-fold reduction relative to wild-type.[13, 42] The calculated barrier for H109E likely underestimates the experimentally observed kinetic barrier because QM/MM describes the intrinsic chemistry from a preorganized crosslink–competent reactant complex, whereas experiment reflects cost of accessing this rare state from a predominantly hydrolysis-favoring ensemble. In particular, rearrangement of E109 toward C1′, exclusion of active site water, and reorganization of the local electrostatic environment are expected to contribute to the effective barrier measured experimentally.

QM/MM reaction profiles for the H109E mutant further support its retained catalytic competence. The computed free energy barrier for the nucleophilic attack of E109 at C1′ is 42.8 kJ·mol^{-1}, which is comparable to, and slightly lower than, the corresponding barrier in the wild-type model (66.5 kJ·mol^{-1}). This indicates that the intrinsic chemical step remains energetically accessible following mutation to glutamate. However, a major qualitative difference emerges in the stability of the covalent product. While the wild-type enzyme strongly stabilizes the PC (−113.4 kJ·mol^{-1}), the H109E mutant forms a significantly less stabilized PC (−74.9 kJ·mol^{-1}). This substantially shallower energy well suggests reduced trapping of the covalent intermediate, implying a weakened “suicide” character despite preservation of the catalytic pathway.

The experimental and computational data indicate that mutation of H109 to glutamate (E) perturbs the energetics of product stabilization far more than the TS barrier for

nucleophilic attack. QM/MM calculations suggest that, once a productive enzyme–substrate complex is formed, the chemical step in H109E proceeds with an efficiency comparable to the wild-type, consistent with the modest 2–3-fold reduction in k_2/K_m observed experimentally. The reduced stabilization of the covalent intermediate provides a mechanistic explanation for the slight loss in overall efficiency, while the enhanced substrate binding explains why H109E remains one of the most active H109 variants. Overall, these results demonstrate that H109E preserves much of the catalytic functionality of H109 while attenuating intermediate trapping, offering insight into how H109 fine-tunes both chemical reactivity and the lifetime of the covalent intermediate in UdgX.

2.4. Conclusions

In the UdgX wild-type model, MD simulations suggest that persistent hydrogen bonding occurs between H178, N91, F65, and dU in both tautomer models, which is consistent with crystallographic data and prior mechanistic proposals. These conserved active site interactions support a role in stabilizing the leaving group and facilitating elongation of the N–glycosidic bond during uracil excision. Members of the UDG family are recognized to have active site residues that are consistent across enzymes, as such it is often observed that glutamic acid (E) is part of their catalytic mechanism. While E52 has previously been proposed to act as a catalytic dyad with H109 to directly activate the nucleophile in the wild-type enzyme, the present simulations reveal a more complex mechanism involving a bridging water molecule. Importantly, this solvent-mediated triad (E52, H_2O, and H109) is preferentially stabilized when H109 adopts the HID tautomer.

Over all simulations, HID maintains improved geometric alignment for nucleophilic attack, presenting a proton-transfer pathway that would otherwise be sterically inaccessible relative to the HIE state, which displayed less optimal orientations and reduced coupling between proton transfer and nucleophilic alignment. From the conformations determined through MD simulations, E52 likely acts as the initial general base, polarizing the bridging water which in turn activates the H109 nucleophile. The water acts as both a geometric and catalytic residue by relaying proton transfer and aligning H109 for nucleophilic attack on the C1′ position of the abasic sugar during dU excision. Therefore, this bridging water molecule functions as an essential element of the catalytic mechanism, coupling residue protonation dynamics with precise positioning of the nucleophile to drive crosslink formation that is unique to UdgX. These results refine existing mechanistic models by highlighting the irreversible protein–DNA crosslink pathway that defines wild-type UdgX catalysis and establishes a framework in which solvent-mediated catalysis plays a central role in distinguishing UdgX from other members of the UDG superfamily.

Collectively, mutational analysis provided key mechanistic insight into the catalytic requirements of UdgX. MD simulations on a series of active site variants (H109A, H109S, H109E, H109Q, H109D, and H109K) support hydrolytic activity. Specifically, across all mutants the hydrolysis-competent active site architecture observed during MD simulations is consistent with experimental observations, indicating that disruption of the H109 nucleophile does not fully abolish catalytic alignment but instead shifts the reaction pathway toward water-mediated deglycosylation. These results support the hypothesis that the conserved residue network surrounding E52, Q53, N91, F65, and H178 plays a dominant role in substrate positioning independent of histidine chemistry. QM/MM

calculations on the H109A variant as a representative hydrolysis pathway revealed how loss of the imidazole nucleophile favors water-driven glycosidic bond cleavage. In contrast, the H109E mutant exhibited dual catalytic potential, retaining the ability to support both hydrolysis as supported by MD and covalent crosslink formation as supported by QM/MM. Consequently, QM/MM analysis of the crosslinking pathway in H109E was performed to determine how substitution with a glutamate side chain influences nucleophile activation and reaction energetics following active site reorganization, including reorientation of E109 toward C1′ and displacement of the hydrolysis-supporting water molecules. Together, these mutant studies demonstrate that UdgX catalytic behavior is not solely dictated by residue identity at position 109 but rather emerges from the cooperative organization of the active site hydrogen-bonding network and electrostatic environment.

2.5. References

(1) Minchin, S.; Lodge, J. Understanding biochemistry: structure and function of nucleic acids. *Essays Biochem.* **2019**, *63* (4), 433-456. DOI: 10.1042/ebc20180038.

(2) Houlihan, G.; Arangundy-Franklin, S.; Holliger, P. Exploring the Chemistry of Genetic Information Storage and Propagation through Polymerase Engineering. *Acc. Chem. Res.* **2017**, *50* (4), 1079-1087. DOI: 10.1021/acs.accounts.7b00056.

(3) Carusillo, A.; Mussolino, C. DNA Damage: From Threat to Treatment. *Cells* **2020**, *9* (7). DOI: 10.3390/cells9071665.

(4) Barnes, J. L.; Zubair, M.; John, K.; Poirier, M. C.; Martin, F. L. Carcinogens and DNA damage. *Biochem. Soc. Trans.* **2018**, *46* (5), 1213-1224. DOI: 10.1042/bst20180519.

(5) Hakem, R. DNA-damage repair; the good, the bad, and the ugly. *EMBO J.* **2008**, *27* (4), 589-605. DOI: 10.1038/emboj.2008.15.

(6) Kaur, R.; Nikkel, D. J.; Wetmore, S. D. Computational studies of DNA repair: Insights into the function of monofunctional DNA glycosylases in the base excision repair pathway. *WIREs Comput. Mol. Sci.* **2020**, *10* (5), e1471. DOI: 10.1002/wcms.1471.

(7) Stivers, J. T.; Jiang, Y. L. A Mechanistic Perspective on the Chemistry of DNA Repair Glycosylases. *Chem. Rev.* **2003**, *103* (7), 2729-2760. DOI: 10.1021/cr010219b.

(8) Nora M. Al Aboud, C. T., Ishwarlal Jialal. Genetics, Epigenetic Mechanism. *StatPearls* **2023**.

(9) Kumar, S.; Chinnusamy, V.; Mohapatra, T. Epigenetics of Modified DNA Bases: 5-Methylcytosine and Beyond. *Front. Genet.* **2018**, *9*, 640. DOI: 10.3389/fgene.2018.00640.

(10) Jia, Q.; Zeng, H.; Tu, J.; Sun, L.; Cao, W.; Xie, W. Structural insights into an MsmUdgX mutant capable of both crosslinking and uracil excision capability. *DNA Repair* **2021**, *97*, 103008. DOI: 10.1016/j.dnarep.2020.103008.

(11) Aroli, S.; Woo, E.-J.; Gopal, B.; Varshney, U. Mutational and structural analyses of UdgX: insights into the active site pocket architecture and its evolution. *Nucleic Acids Res.* **2023**, *51* (13), 6554-6565. DOI: 10.1093/nar/gkad486.

(12) Ahn, W.-C.; Aroli, S.; Kim, J.-H.; Moon, J. H.; Lee, G. S.; Lee, M.-H.; Sang, P. B.; Oh, B.-H.; Varshney, U.; Woo, E.-J. Covalent binding of uracil DNA glycosylase UdgX to abasic DNA upon uracil excision. *Nat. Chem. Biol.* **2019**, *15* (6), 607-614. DOI: 10.1038/s41589-019-0289-3.

(13) Liang, C.; Yang, Y.; Ning, P.; Chang, C.; Cao, W. Structural and functional coupling in cross-linking uracil-DNA glycosylase UDGX. *Biosci. Rep.* **2024**, *44* (1). DOI: 10.1042/bsr20231551.

(14) Pearl, L. H. Structure and function in the uracil-DNA glycosylase superfamily. *Mutat. Res.* **2000**, *460* (3), 165-181. DOI: 10.1016/S0921-8777(00)00025-2.

(15) Schormann, N.; Ricciardi, R.; Chattopadhyay, D. Uracil-DNA glycosylases—Structural and functional perspectives on an essential family of DNA repair enzymes. *Protein Sci.* **2014**, *23* (12), 1667-1685. DOI: 10.1002/pro.2554.

(16) Lee, H.-W.; Dominy, B. N.; Cao, W. New Family of Deamination Repair Enzymes in Uracil-DNA Glycosylase Superfamily *J. Biol. Chem.* **2011**, *286* (36), 31282-31287. DOI: 10.1074/jbc.M111.249524.

(17) Slupphaug, G.; Mol, C. D.; Kavli, B.; Arvai, A. S.; Krokan, H. E.; Tainer, J. A. A nucleotide-flipping mechanism from the structure of human uracil–DNA glycosylase bound to DNA. *Nature* **1996**, *384* (6604), 87-92. DOI: 10.1038/384087a0.

(18) Tu, J.; Chen, R.; Yang, Y.; Cao, W.; Xie, W. Suicide inactivation of the uracil DNA glycosylase UdgX by covalent complex formation. *Nat. Chem. Biol.* **2019**, *15* (6), 615-622. DOI: 10.1038/s41589-019-0290-x.

(19) Nikkel, D. J.; Wetmore, S. D. Distinctive Formation of a DNA–Protein Cross-Link during the Repair of DNA Oxidative Damage: Insights into Human Disease from MD Simulations and QM/MM Calculations. *J. Am. Chem. Soc.* **2023**, *145* (24), 13114-13125. DOI: 10.1021/jacs.3c01773.

(20) Majumdar, C.; McKibbin, P. L.; Krajewski, A. E.; Manlove, A. H.; Lee, J. K.; David, S. S. Unique Hydrogen Bonding of Adenine with the Oxidatively Damaged Base 8-Oxoguanine Enables Specific Recognition and Repair by DNA Glycosylase MutY. *J. Am. Chem. Soc.* **2020**, *142* (48), 20340-20350. DOI: 10.1021/jacs.0c06767.

(21) Sang, P. B.; Srinath, T.; Patil, A. G.; Woo, E. J.; Varshney, U. A unique uracil-DNA binding protein of the uracil DNA glycosylase superfamily. *Nucleic Acids Res.* **2015**, *43* (17), 8452-8463. DOI: 10.1093/nar/gkv854.

(22) Wang, W. J.; Wang, T.; Zhao, Y.; Li, B. N.; Chen, D. Z. Theoretical Insights into N-Glycoside Bond Cleavage of 5-Carboxycytosine by Thymine DNA Glycosylase: A QM/MM Study. *J. Phys. Chem. B* **2024**, *128* (19), 4621-4630. DOI: 10.1021/acs.jpcb.4c00221.

(23) Nikkel, D. J.; Wetmore, S. D. A tale of two mechanisms: Clarification of the pathway for MBD4 catalyzed glycosidic bond cleavage using MD and QM/MM calculations. *DNA Repair* **2026**, *157*, 103917. DOI: 10.1016/j.dnarep.2025.103917.

(24) Przybylski, J. L.; Wetmore, S. D. A QM/QM Investigation of the hUNG2 Reaction Surface: The Untold Tale of a Catalytic Residue. *Biochemistry* **2011**, *50* (19), 4218-4227. DOI: 10.1021/bi2003394.

(25) Nikkel, D. J.; Wetmore, S. D. Unlocking the chemistry facilitated by enzymes that process nucleic acids using quantum mechanical and combined quantum mechanics-molecular mechanics techniques. *Pure Appl. Chem.* **2025**, *97* (9), 1065-1089. DOI: 10.1515/pac-2025-0507.

(26) Schrodinger. *The Pymol Molecular Graphics System*, Version 3.0.4; Schrodinger LLC, 2020.

(27) Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* **2012**, *40* (Web Server issue), W537-541. DOI: 10.1093/nar/gks375.

(28) Case, D. A.; Aktulga, H. M.; Belfon, K.; Cerutti, D. S.; Cisneros, G. A.; Cruzeiro, V. W. D.; Forouzesh, N.; Giese, T. J.; Götz, A. W.; Gohlke, H.; et al. AmberTools. *J. Chem. Inf. Model.* **2023**, *63* (20), 6183-6191. DOI: 10.1021/acs.jcim.3c01153.

(29) Schmit, J. D.; Kariyawasam, N. L.; Needham, V.; Smith, P. E. SLTCAP: A Simple Method for Calculating the Number of Ions Needed for MD Simulation. *J. Chem. Theory Comput.* **2018**, *14* (4), 1823-1827. DOI: 10.1021/acs.jctc.7b01254.

(30) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696-3713. DOI: 10.1021/acs.jctc.5b00255.

(31) Galindo-Murillo, R.; Robertson, J. C.; Zgarbová, M.; Šponer, J.; Otyepka, M.; Jurečka, P.; Cheatham, T. E., III. Assessing the Current State of Amber Force Field Modifications for DNA. *J. Chem. Theory Comput.* **2016**, *12* (8), 4114-4127. DOI: 10.1021/acs.jctc.6b00186.

(32) Li, P.; Merz, K. M., Jr. MCPB.py: A Python Based Metal Center Parameter Builder. *J. Chem. Inf. Model.* **2016**, *56* (4), 599-604. DOI: 10.1021/acs.jcim.5b00674.

(33) He, X.; Man, V. H.; Yang, W.; Lee, T. S.; Wang, J. A fast and high-quality charge model for the next generation general AMBER force field. *J. Chem. Phys.* **2020**, *153* (11), 114502. DOI: 10.1063/5.0019056.

(34) Li, P.; Song, L. F.; Merz, K. M., Jr. Systematic Parameterization of Monovalent Ions Employing the Nonbonded Model. *J. Chem. Theory Comput.* **2015**, *11* (4), 1645-1657. DOI: 10.1021/ct500918t.

(35) Case, D. A. B.-S., I. Y.; Brozell, S. R.; Cerutti, D. S.; III, T. E. C.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; et al. *AMBER* **2018**.

(36) Sgrignani, J.; Magistrato, A. QM/MM MD Simulations on the Enzymatic Pathway of the Human Flap Endonuclease (hFEN1) Elucidating Common Cleavage Pathways to RNase H Enzymes. *ACS Catal.* **2015**, *5* (6), 3864-3875. DOI: 10.1021/acscatal.5b00178.

(37) Kaur, R.; Nikkel, D. J.; Wetmore, S. D. Mechanism of Nucleic Acid Phosphodiester Bond Cleavage by Human Endonuclease V: MD and QM/MM Calculations Reveal a Versatile Metal Dependence. *J. Phys. Chem. B* **2024**, *128* (39), 9455-9469. DOI: 10.1021/acs.jpcb.4c05846.

(38) Frisch, M. J. T., G. W.; et al. . Gaussian 16, Revision B.01. *Gaussian, Inc* **2016**.

(39) Xiao, G.; Tordova, M.; Jagadeesh, J.; Drohat, A. C.; Stivers, J. T.; Gilliland, G. L. Crystal structure of Escherichia coli uracil DNA glycosylase and its complexes with uracil and glycerol: structure and glycosylase mechanism revisited. *Proteins* **1999**, *35* (1), 13-24.

(40) Pedersen, H. L.; Johnson, K. A.; McVey, C. E.; Leiros, I.; Moe, E. Structure determination of uracil-DNA N-glycosylase from Deinococcus radiodurans in complex with DNA. *Acta Crystallogr. D Biol. Crystallogr.* **2015**, *71* (Pt 10), 2137-2149. DOI: 10.1107/s1399004715014157.

(41) Leiros, I.; Moe, E.; Lanes, O.; Smalås, A. O.; Willassen, N. P. The structure of uracil-DNA glycosylase from Atlantic cod (Gadus morhua) reveals cold-adaptation features. *Acta Crystallogr. D Biol. Crystallogr.* **2003**, *59* (Pt 8), 1357-1365. DOI: 10.1107/s0907444903011144.

(42) Liang, C. Biochemical Analyses of UDGX-A Crosslinking Uracil-Dna Glycosylase. *Diss. Abstr. Int.* **2023**, (3494).

Chapter 3 : A Computational Investigation of the Substrate Specificity in Alkyladenine DNA Glycosylase and ALKBH2 Towards 3,N^4–Etheno-5-methylcytosine

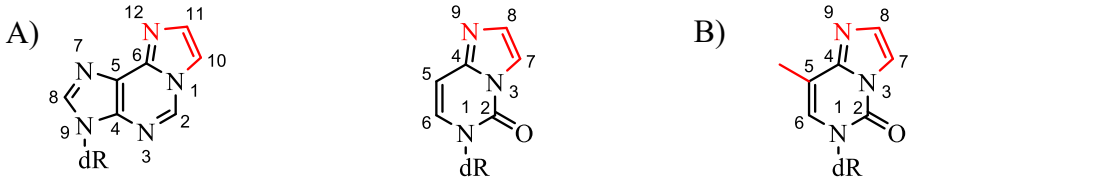
Contributions of Authors: I (Angela Frederickson) was responsible for all aspects of model building and parameterization of non-standard nucleobases, conducting molecular dynamics simulations and performing data/results analysis, writing, and editing. Stacey D. Wetmore was involved in project conceptualization and administration, supervision, data visualization and interpretation, funding acquisition, computational resource acquisition, and review/editing.

3.1. Introduction

5-Methylcytosine (5MeC) is a major epigenetic modification in mammalian genomes and occurs predominantly in CpG dinucleotide sites (i.e., the cytosine that precedes guanine residues).[1, 2] 5MeC plays an important role in regulating gene expression, chromatin organization, and the maintenance of cellular identity.[1-3] Because CpG methylation serves as a stable yet dynamically interpreted regulatory signal, the chemical integrity of 5MeC is critical for proper epigenetic control.[2, 4] Therefore, chemical modification of this base has consequences that extend beyond genome stability and into epigenetic regulation.

Endogenous cellular processes associated with oxidative stress and inflammation, particularly lipid peroxides, generate reactive electrophilic species capable of modifying DNA bases and forming exocyclic etheno adducts.[3, 5, 6] When these electrophiles target cytosine within CpG contexts, cytosine is first converted to 5-methylcytosine (5MeC), which can subsequently be modified by electrophilic lipid-peroxidation products to form the exocyclic lesion 3,N^4-etheno-5-methylcytosine (ε5mC) (Figure 3.1), resulting in direct damage to an established epigenetic signal.[3, 7, 8] This modification introduces a bulky

etheno bridge while retaining the C5 methyl group, producing a base that is chemically distinct from both unmodified 5MeC and classic alkylation damage in the form of etheno adducts.



1,N^6-ethenoadenine (1,N^6 -εA) 3,N^4-ethenocytosine (3,N^4 -εC) 3,N^4-etheno-5-methylcytosine (ε5mC)

Figure 3.1. Structures of A) common etheno DNA lesions (εA, εC) and B) a newly identified DNA etheno lesion (ε5mC), with modifications highlighted in red.

Exocyclic etheno cytosine lesions, including 3,N^4-ethenocytosine (εC) and ε5mC (Figure 3.1), disrupt Watson-Crick-Franklin base pairing and compromise genomic integrity by introducing steric bulk into and altering the electronic structure of the nucleobase.[3, 5] Experimental studies have shown that these lesions are replication blocking and mutagenic, promoting base substitution mutations when bypassed by replicative DNA polymerases.[5, 6] The biological consequences of ε5mC are further amplified because this lesion also interferes with epigenetic regulation by blocking oxidation of 5MeC by family ten-eleven translocation (TET) methylcytosine dioxygenases, therefore disrupting active DNA demethylation pathways.[3, 9, 10]

Cells rely on several DNA repair pathways to counteract the harmful effects of etheno adduct formation, most prominently base excision repair (BER) and direct reversal repair (DRR).[11-13] In the BER pathway, alkyladenine DNA glycosylase (AAG) initiates repair by

recognizing damaged nucleobases and catalyzing cleavage of the N–glycosidic bond, generating an abasic site that is subsequently processed by downstream enzymes.[14] AAG is notable for its unusually broad substrate specificity and efficiently excises a range of alkylated purine lesion, including ethenoadenine (εA), 3–methyladenine (3MeA), and 7–methylguanine (7MeG).[14, 15] Despite this apparent versatility, AAG does not excise cytosine-derived etheno lesions.[3, 16] Biochemical and structural studies demonstrated that εC inhibits repair by forming a tightly bound yet catalytically inactive complex rather than undergoing AAG-catalyzed excision, and recent work has shown that ε5mC is likely not a substrate for AAG or other BER glycosylases.[3, 14] The inability of AAG to repair these cytosine-derived lesions is particularly intriguing given the broad substrate specificity of AAG.[17] This raises the question of how cells resolve cytosine-derived etheno damage that inhibit glycosylase-initiated repair. Molecular dynamics (MD) simulations suggest that binding of εC induces subtle yet catalytically unfavourable reorganization of the AAG active site, including changes in residue positioning (E125, Y127, A135, H136, Y159, N169, L180, and R182) and active site solvation that prevent productive glycosidic bond cleavage.[14] These observations indicate that nucleotide binding alone is insufficient to predict repair outcome and that mechanistic constraints imposed by the BER pathway may play a decisive role in lesion processing.

In contrast to BER, the AlkB family of α-ketoglutarate (α-KG)-dependent dioxygenases repairs alkylated and etheno-modified bases through DRR rather than base excision.[12, 18] The literature describes Fe(II)/α-KG-dependent dioxygenase ALKBH2 as being capable of repairing both εC (a frequently studied lesion) and ε5mC (an understudied lesion) through the DRR mechanism, restoring the etheno lesion to canonical cytosine and 5MeC,

respectively.[19] Experimental studies have established that ALKBH2 efficiently processes these lesions; however, the structural and dynamic basis for this selectivity remains unclear.[16] Both bacterial AlkB and the human homolog ALKBH2 catalyze oxidative dealkylation that restores the native base without cleavage of the sugar–phosphate backbone.[12, 18, 20, 21] Past experimental investigations have established that ALKBH2 efficiently repairs εC in double-stranded DNA, placing this enzyme as a contributor to the cellular response to etheno cytosine damage.[16] More recently, ALKBH2 has been shown to also repair ε5mC both *in vitro* and *in vivo*, converting this lesion back to the original 5MeC base.[3] This activity separates ALKBH2 from BER glycosylases and highlights a unique role for DRR in preserving epigenetic information following chemical damage. Notably, AlkB family enzymes have also been reported to oxidize the epigenetic modification 5MeC to its oxidized derivatives 5-hydroxymethycytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC), further suggesting a broader connection between AlkB-mediated repair and epigenetic regulation.[9, 19, 22] Computational studies indicate that productive ALKBH2 catalysis depends on precise nucleotide positioning within the active site, where alignment of the aberrant carbon atoms with the reactive Fe^{IV}–oxo species is required for effective oxidation.[5] However, despite clear biochemical evidence for εC and ε5mC repair by ALKBH2, the structural basis for how these lesions are accommodated, particularly in comparison to the inability of AAG to process them, remains incompletely understood.

Collectively, these studies reveal clear functional divergence between BER and DRR pathways in the processing of cytosine etheno lesions. Despite the broad substrate specificity of AAG, this enzyme fails to excise εC and ε5mC, whereas ALKBH2 efficiently

repairs both lesions through oxidative reversal. The molecular determinants that contribute to this division remain unclear as it is not well understood how differences in active site architecture, nucleotide positioning, solvation, and electronic structure of the nucleotides dictates the ability of these enzymes to process and repair. Since experimental approaches provide limited access to details of substrate recognition on a molecular level, MD simulations offer a complementary perspective. MD simulation studies have already proven valuable for elucidating substrate specificity, active site reorganization, and solvation effects in both AAG[11, 14, 23, 24] and ALKBH2[5, 19, 25, 26] among other DNA repair enzymes (MutY,[27] MBD4,[28] TDG,[29] and UdgX[30]) that have provided critical insight.

Building on these foundations, the present work employs MD simulations to directly compare the accommodation of εA, εC, and ε5mC by distinct repair enzymes, addressing a central mechanistic question highlighted by experimental observations.[17, 31, 32] By examining differences in nucleotide (εA, εC, and ε5mC) positioning, hydrogen-bonding networks, solvation patterns, and interactions with key catalytic residues, these simulations aim to identify structural features that favour productive catalysis of εA by AAG, εC and ε5mC by ALKBH2, while εC and ε5mC lead to inhibition in AAG. Understanding the structural basis for this functional divergence has broader implications for how DNA repair pathways compete to process chemically similar lesions and how repair outcomes influence the persistence of mutagenic damage in the genome. In particular, elucidating the molecular determinants that govern lesion recognition and catalytic competency may help rationalize experimentally-observed repair efficiencies and provide insight into how conserved repair scaffolds evolve specialized chemical strategies. Ultimately, this work highlights how computational approaches can bridge gaps between structural, biochemical, and

mechanistic observations while providing a framework for interpreting how subtle differences in enzyme architecture dictate pathway selection in DNA repair.

3.2. Computational Methods

3.2.1. MD Simulation Protocol

A total of 5 unique complexes were considered that contain AAG or ALKBH2 bound to DNA containing one of two etheno lesions (namely εC and ε5mC), as well as AAG bound to εA. X-ray crystal structures available for AAG (PDB ID: 3QI5 and 1EWN)[17, 31] (Figure 3.2) and ALKBH2 (PDB ID: 3RZJ)[32] (Figure 3.3) bound to lesion-containing DNA were used to build all models. In these crystal structures the E125Q AAG mutant is in complex with the εC inhibitor and ALKBH2 is in complex with α-KG, a Mn(II) metal substitution, and 3MeC. These structures were chosen because of their high-resolution (2.2–2.5 Å) and their usage in previous studies.[14]

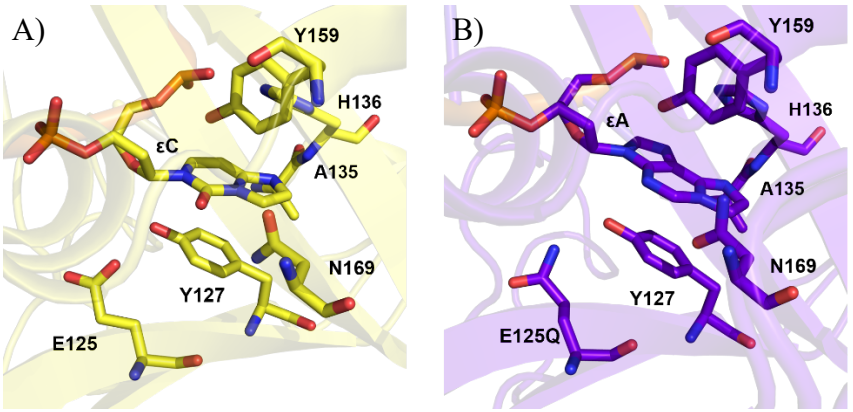


Figure 3.2. X-ray crystal structure of AAG bound to A) εC-containing DNA (PDB ID: 3QI5) and B) εA-containing DNA (PDB ID: 1EWN).

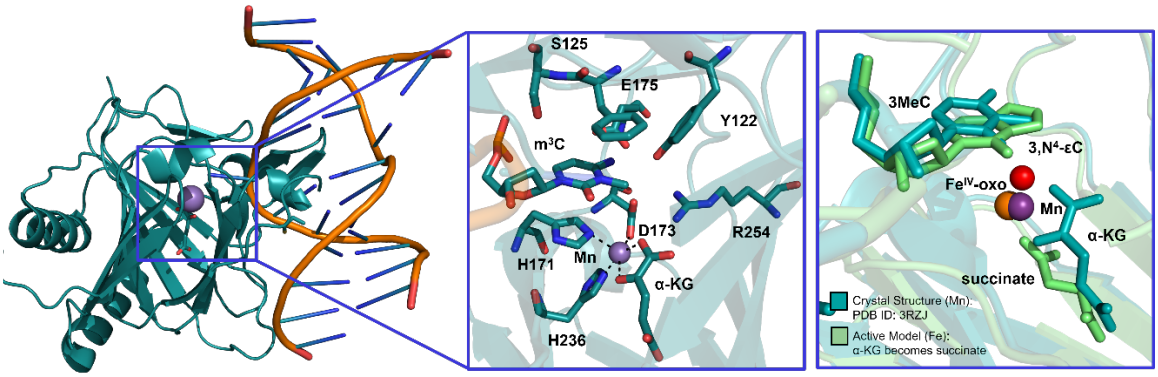


Figure 3.3. X-ray crystal structure of ALKBH2 bound to 3MeC-containing DNA (PDB ID: 3RZJ).

Although 3QI5 includes two AAG–εC complexes, the complex containing the fewest unresolved residues was used for our study. An initial model for AAG was generated by resolving residues 203–207 and 265–268 using the SWISS-MODEL[33] server. Overlays of the sequence generator and both complexes were used to ensure proper position and orientation of residues resolved in the AAG model, which is justified because all unresolved residues are remote from the active site (>10 Å separation). Residues with partially unresolved side chains (201, 249, and 253) were manually added using the PyMOL[34] mutagenesis tool and refined to favor hydrogen-bonding interactions and prevent steric clashes. Hydrogen atoms and appropriate protonation states (at pH 7.0) were assigned using the H++ webserver[35] for titratable residues HIE57 and HIE187. Other DNA–protein complexes were generated by replacing εC in the structure generated from 3QI5 with εA or ε5mC.

For the ALKBH2 models, previous studies provided a reference structure of the coordination distance between Fe(II) and the reactive oxygen species to generate the Fe^{IV}–

oxo complex.[5] The iron-binding site was superimposed onto the metal-binding site (Mn(II)) of the crystal structure and α-KG was replaced with succinate. An initial model for ALKBH2 was generated by resolving missing residues using the Swiss Model webserver and mutations for crystallization were removed. Starting structures for εC bound by ALKBH2 were generated from the 3MeC-containing crystal structure followed by the addition of a methyl group to generate a model with ε5mC bound in ALKBH2. Residues with partially unresolved side chains (67, 165, 169, and 192) were manually added using the PyMOL mutagenesis tool and refined to favor hydrogen-bonding interactions and prevent steric clashes. Protonation states were initially assigned using the H++ webserver and adjusted to preserve local hydrogen-bonding networks. All mutations necessary for crystallization were reverted to the wild-type phenotype, while the overhanging 5′ and 3′ termini of the bound DNA were truncated. The MCPB.py[36] program was utilized to assign bonding, angle, dihedral, and non-bonding parameters based on B3LYP (6-31G(d,p) for H, C, N, and O, and LANL2DZ for Fe^{IV}) optimized structures of the iron-binding site according to the Seminaro method,[37] as adopted from previous computational studies.[5]

Each DNA–protein complex (5) was solvated in a truncated octahedral box of TIP3P water molecules, maintaining a minimum distance of 12 Å between the solute and box edge. Water molecules were added using the Leap program in Amber23. Each system was then neutralized with sodium ions and additional Na^+ and Cl^- ions were added to approximate physiological salt conditions (150 mM) as determined using the SLTCAP[38] calculator (Table B1, Appendix B). The AMBER ff14SB force field[39] was used to model proteins and the OL15 force field[40] for DNA in each complex and parameters for the nonstandard etheno lesions being supplemented with GAFF2 parameters[41] and ions were

modelled with Li/Merz ion parameters.[42] Additionally, nonstandard nucleosides residues were constructed using the pyRED[43] server for εA, εC, and ε5mC (Table B2, Appendix B), which was used to derive restricted electrostatic potential (RESP) charges following established Amber force field protocols.

Energy minimization was performed in four steps using pmemd of Amber23[44] for the AAG and the ALKBH2 models: (1) solvent and ions were minimized with 100 kcal·mol^{-1}·Å^{-2} restraints on the solute, (2) restraints were lifted from solute hydrogen atoms, (3) solute atoms were minimized with restraints on the solvent, and (4) a final unrestrained minimization of the entire system was completed (1000 steps steepest descent followed by 2000 steps conjugate gradient). Heating to 310 K was achieved over six 50 K increments, applying a 25 kcal·mol^{-1}·Å^{-2} restraint to the solute. Five-stage equilibration was performed with decreasing restraint strengths on the solute from from 25 to 20, 15, 10, 5, and 1.5 kcal·mol^{-1}·Å^{-2}. The equilibrated structure then underwent 1000 ns unrestrained MD production simulations in quintuplicate using Amber23. Simulation timescales in the present work extended beyond those reported in earlier AAG studies, which employed ∼40 ns MD trajectories to characterize active site behavior, while more recent AlkB family simulations utilized extended microsecond-scale sampling to capture nucleotide-dependent dynamics.[5, 14] Simulations employed the NPT ensemble (310 K, 1 bar) with isotropic pressure scaling, a 10 Å cutoff for nonbonded interactions, periodic boundary conditions, and TIP3P water density fixed at ∼1 g mL^{-1}.

3.2.2. MD Simulation Analysis

Frames were saved every 20 ps for analysis, resulting in 20,000 frames per system across all replicas for analysis. The cpptraj program implemented in Amber23 was used to analyze the geometric data obtained from trajectories. A script personalized to the system was used to evaluate key distances, hydrogen-bonding, and stacking interactions of the active site to access the alignment of residues. The hydrogen bonds were measured using the cpptraj command and considered present when the heavy atom-to-heavy atom distance was $\leq$ 3.4 Å and hydrogen-bond angle was 130°–180°. π–stacking interactions were quantified using cpptraj by defining planes for aromatic systems using selected ring atoms (nucleotide) and calculating the distance between ring atom groups together with the angle between plane-normal vectors. The root-mean-square deviation (RMSD) of the enzyme–DNA complex backbone, enzyme backbone, DNA backbone, and active site (E125, Y127, A135, H136, Y159, N169, L180, R182, and dX; AAG and Y122, F124, S125, H171, D173, E175, H236, R254, and dX; ALKBH2) heavy atoms were separately analyzed to confirm structural convergence of the systems over the 1000 ns simulations. RMSD values for the systems across all five independent simulations show a modest increase during the initial equilibration period followed by stabilization, with fluctuations occurring within a relatively narrow range (Figure B.1 and B.2, Appendix B). The absence of a continuous increase in RMSD over time indicates that the system rapidly reaches a stable conformational ensemble and does not undergo large-scale structural drift during the simulations. A representative structure was obtained using the cluster function of cpptraj based on RMSD values of the active site for all AAG systems (E125, Y127, A135, H136, Y159, N169, L180, R182, and dX) or important distances specific to the ALKBH2 systems

(C7 and C8 of the nucleotide to the Fe^{IV}–oxo). For all mechanisms involving hydrolysis, analysis was conducted to assess the occupancy of water in the active site over the production simulations. Water density maps were generated from MD trajectories using cpptraj of AmberTools by combining all replicates and structurally aligned to the first frame using the heavy atoms of the protein residues. The water density was then calculated by binning the positions of water oxygen atoms from all frames onto a three-dimensional grid with a spacing of 0.45 Å that encompassed the protein. For each grid voxel, the water occupancy was averaged to yield a normalized water density distribution and visualized as isodensity meshes in PyMOL to highlight persistent hydration sites.

3.3. Results and Discussion

3.3.1. AAG Dynamics and Nucleotide-Dependent Active Site Organization

3.3.1.1. AAG Active Site Conformations Display Subtle Differences Across Repair and Inhibitory Nucleotides

The AAG active site is composed of several conserved residues that position the flipped lesion and facilitate cleavage of the N–glycosidic bond during base excision repair. In crystal structures of AAG bound to damaged DNA, aromatic residues Y127, H136, and Y159 form π–π interactions with the flipped nucleobase, while residues such as A135 and N169 help shape the binding pocket and stabilize nucleotide binding (Figure 3.4).[17, 31] The catalytic residue E125 is proposed to act as a general base that activates a water molecule for nucleophilic attack on the C1′ of the deoxyribose while being stabilized by Y127, enabling hydrolysis of the N–glycosidic bond and release of the damaged base assisted by

the E125–Y127 catalytic dyad.[17, 31] Consistent with this mechanism, AAG efficiently excises εA lesions through hydrolysis, whereas εC binds within the active site but inhibits catalysis, forming a stable abortive complex (Figure 3.4).[11, 24]

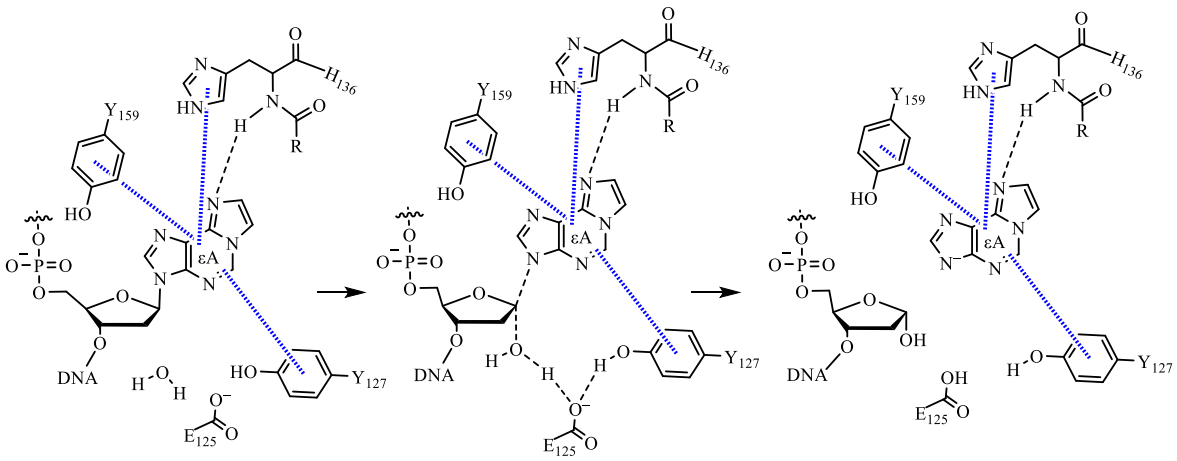


Figure 3.4. Proposed S_N1-like hydrolysis mechanism for excision of εA by AAG, with stacking interactions highlighted in blue. Initial cleavage of the C1′–N9 glycosidic bond generates an oxocarbenium-like intermediate, followed by nucleophilic attack of an activated water molecule at C1′. The catalytic residue E125 facilitates water activation, resulting in hydrolysis of the glycosidic bond, protonation of the departing εA base, and formation of an abasic (AP) site.

MD simulations initiated from X-ray crystal structures with either εA or εC bound to AAG yield an active site conformation that is consistent with crystallographic models. Active site (E125, Y127, A135, H136, Y159, and N169) RMSD values calculated relative to the corresponding crystal structures throughout the trajectories, with average deviations of 1.18 Å for the εA system, 1.25 Å for the εC system, and 1.14 Å for the ε5mC system (Figure B.3, Appendix B). These consistently low RMSD values indicate that the catalytic pocket remains structurally stable and closely aligned with the experimentally-observed conformations during the simulations. To evaluate where differences in AAG repair activity arise from structural alignment within the active site, representative structures from MD

simulations of AAG bound to εA, εC, or ε5mC were subsequently investigated (Figure 3.5). Focusing on the relative position of the modified nucleobase within the active site and the surrounding catalytic residues (E125, Y127, A135, H136, Y159, and N169). Across all three systems, the overall alignment of the AAG active site remains highly conserved. Representative structures reveal close alignment of the protein backbone and side chains, indicating that nucleotide identity does not induce major conformational changes (Figure 3.5). This observation is consistent with experimental X-ray crystal structure[17] studies and previous MD simulations[14] indicating that AAG accommodates diverse substrates within a largely pre-existing active site architecture, rather than undergoing extensive nucleotide-dependent active site remodeling. Quantitative analysis of active site RMSD values previously reported[14] further supports this conclusion, with minimal deviation observed between εA, εC, and ε5mC bound complexes.

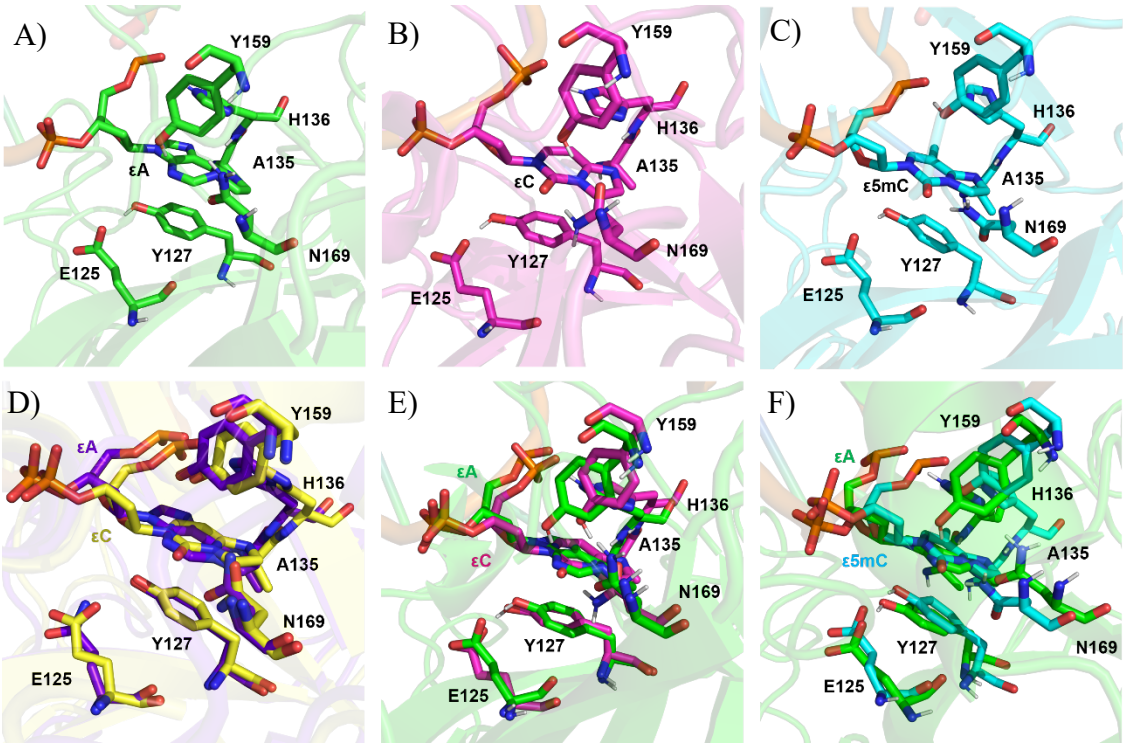


Figure 3.5. Representative structures and structural overlays highlighting differences in nucleobase positioning within the AAG active site. Individual representative MD structures of AAG bound to A) εA (green), B) εC (pink), and C) ε5mC (cyan), highlighting key active site residues. D) Overlay of the εA- (PDB ID: 1EWN) and εC-bound (PDB ID: 3QI5) crystal structures demonstrating that εC adopts a slightly deeper insertion into the active site pocket relative to εA. E) Overlay of representative MD structures for the εA and εC systems showing that the deeper positioning of εC observed experimentally is maintained during the simulations. F) Overlay of representative MD structures for the εA and ε5mC systems illustrating that ε5mC adopts a binding mode similar to that of εC despite the presence of the additional methyl group.

Despite the conserved global alignment, subtle differences in nucleobase positioning within the active site are apparent. In particular, εC adopts a binding orientation that places the nucleobase slightly deeper within the active site pocket relative to εA (Figure 3.5E). This difference in positioning is consistent with crystallographic observations, where overlays of the εA- and εC-bound structures show that εC sits further within the catalytic pocket (Figure 3.5D).[17, 31] The ε5mC-bound system closely mirrors the behaviour of εC in

the simulations, exhibiting a similar degree of insertion despite the presence of the additional methyl substituent (Figure 3.5F). This similarity between εC and ε5mC suggests that methylation does not substantially alter how the etheno-modified cytosine base is accommodated within the AAG active site. Notably, both εC and ε5mC adopt a more deeply embedded binding mode than the repaired εA while preserving the overall active site geometry. These observations suggest that inhibitory binding does not arise from gross structural incompatibility but rather from subtle differences in how the nucleobase is positioned within an otherwise conserved catalytic environment. Together, these findings establish a structural framework for interpreting subsequent analyses of stacking interactions, hydrogen bonding networks, and active site hydration that likely contribute to the observed differences in repair outcomes.

3.3.1.2. Aromatic Stacking Interactions and the E125–Y127 Catalytic Dyad Geometry are Preserved Despite Differences in Nucleotide Positioning

The relative orientations and π–stacking interactions between the damaged nucleobase and aromatic residues within the AAG active site were monitored by measuring center of mass–center of mass distances and angular alignment (Figure 3.6). Across all three AAG systems, stacking interactions are observed between the nucleobase and aromatic residues lining the active site pocket. In particular, the flipped nucleobase is positioned adjacent to the aromatic side chains of Y127 and H136, forming π–π stacking interactions that stabilize the lesion within the catalytic pocket (Figure 3.6). These interactions position the damaged base within an aromatic environment that helps maintain the flipped-out conformation required for productive binding and/or catalysis. The prevalence of these stacking

interactions throughout the trajectories is quantified by their percent occupancy over the total simulation time.

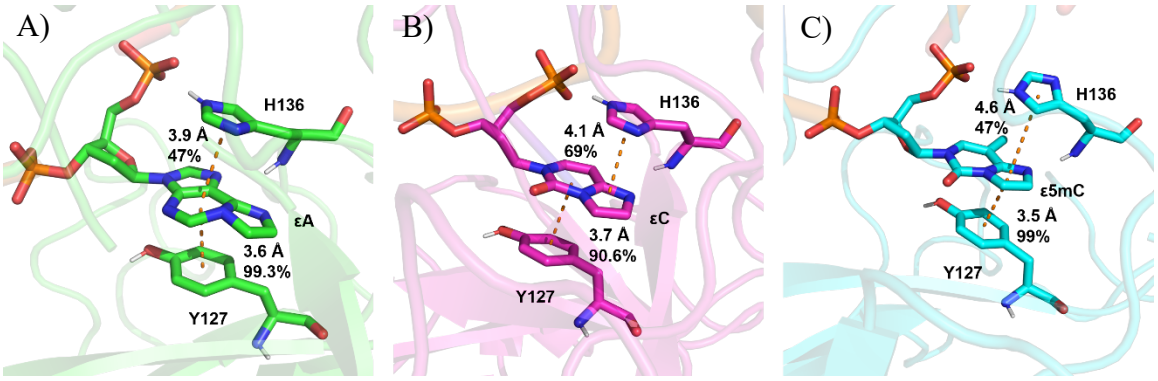


Figure 3.6. Representative MD structures of AAG bound to A) εA (green), B) εC (pink), or C) ε5mC (cyan) containing DNA, highlighting π–π-stacking distance (Å) and percent occupancy of the nucleobase with H136 and Y127.

In the εA-bound system, the substrate maintains persistent stacking with Y127, for an average distance of 3.6 ± 0.7 Å and an occupancy of 99.3%, and H136 for an average distance of 3.9 ± 1.5 Å and an occupancy of 47% of the trajectory. Despite the inhibitory nature of εC, these stacking interactions are maintained at levels like those observed for εA. Particularly, εC exhibits π–π stacking with Y127 throughout the simulation with an average distance of 3.7 ± 0.3 Å and an occupancy of 90.6%. With a slightly higher occupancy observed for εC and H136 of 69%. The ε5mC-bound system displays π–π stacking occupancy closely resembling εA where Y127 persisted for 99% and H136 for 47% of the simulation, though observed at a further center-of-mass distances < 4.6 Å, indicating that the presence of the methyl substituent does causes slight stacking disruption for only H136 within the active site. Together, these results demonstrate that all three nucleotides engage the aromatic platform formed by Y127 and H136, and that loss of

catalytic activity toward εC and ε5mC is likely not attributed to weakened or transient π–π stacking interactions.

In addition to nucleotide stacking, the integrity of the E125–Y127 catalytic dyad was evaluated, as this interaction is considered a key determinant in orienting E125 for nucleophile activation during glycosidic bond cleavage in the hydrolysis mechanism proposed for AAG.[11, 24] Hydrogen bonding between (Y127) O–H···O (E125) was monitored over the simulation. Across εA-, εC-, or ε5mC-bound systems, the E125–Y127 hydrogen bond is maintained for approximately 64%, 53%, and 50% of the simulation time, respectively (Figure 3.7). This persistent interaction is slightly higher for εA suggesting an increase in the formation of the catalytic dyad with the repaired substrate, although the presence of εC or ε5mC disrupts the occupancy of this interaction by 10%. Additional hydrogen bonding between E125 and L180 is observed for >98% of the simulation for all nucleotides. L180 contributes to the formation of a hydrophobic binding pocket that accommodates the flipped base. Consistent with crystallographic observations, L180 maintains close contact with E125 for all nucleotides investigated, suggesting that differences in catalytic outcome does not arise from loss of hydrophobic interactions within the binding cavity.[17, 31]

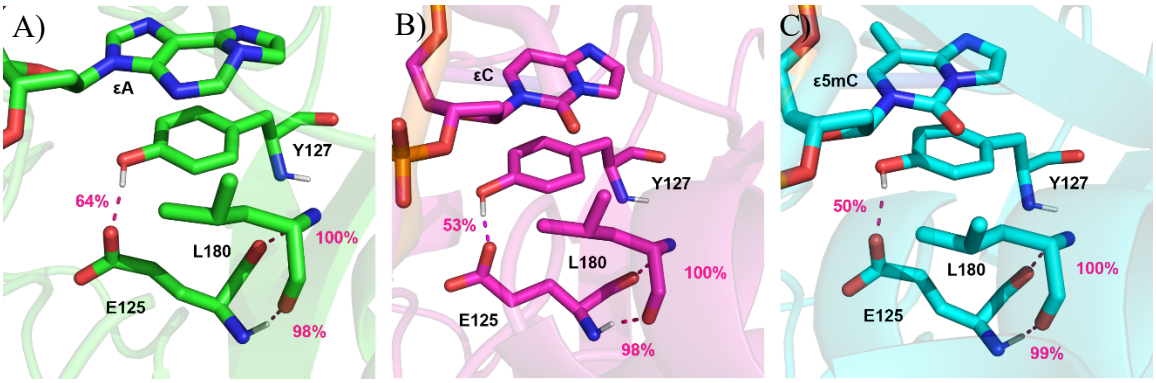


Figure 3.7. Representative MD structures of AAG bound to A) εA (green), B) εC (pink), or C) ε5mC (cyan) containing DNA, highlighting hydrogen-bonding interactions between E125 and L180, and E125–Y127 catalytic dyad maintained for >98% of simulations for all nucleotides.

3.3.1.3. Nucleotide-Dependent Hydrogen-Bonding to H136 and N169 Impacts the Insertion of εC and ε5mC Inhibitors

A persistent hydrogen bond between the N–H backbone of H136 and neutral nucleotides is known to form in AAG bound systems based on experimental X-ray crystal structures.[17] In all three systems, the backbone amide of H136 forms hydrogen bonds with the flipped base (Figure 3.8). However, the nature and persistence of this interaction differ depending on the nucleotide present. For εA, the H136 backbone forms two hydrogen bonds with the nucleobase at N7 and N12. The N7 interaction is observed at an average distance of 2.3 Å with a 25% occupancy, while the N12 hydrogen bond is 2.4 Å and persists for 94% of the trajectory. These interactions position εA within the active site while maintaining conformational flexibility consistent with productive catalysis. In contrast, εC presents only a single hydrogen-bond acceptor capable of interacting with the H136 backbone. This interaction is highly persistent, with a donor–acceptor distance of 2.3 Å and an occupancy

of 93%. The ε5mC system displays a similarly strong interaction, with a hydrogen-bond distance of 2.1 Å and an occupancy of 85%.

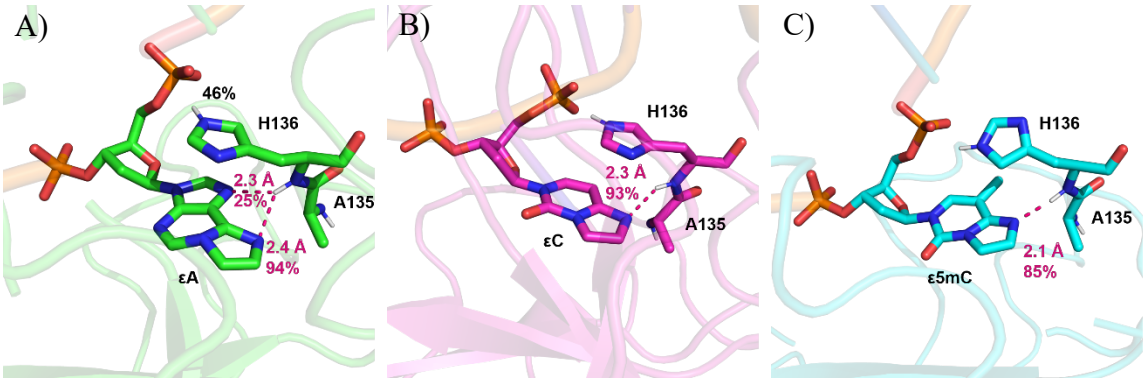


Figure 3.8. Representative MD structures of AAG bound to A) εA (green), B) εC (pink), or C) ε5mC (cyan) containing DNA, highlighting a common hydrogen bond between the H136 backbone and the bound nucleobase.

The increased persistence and shorter average distance of the H136 backbone hydrogen bond in εC and ε5mC is observed within the active site pocket. Because this backbone interaction is directly coupled to the aromatic stacking interaction involving H136, increased hydrogen-bond persistence may contribute to pulling of the cytosine-derived etheno lesions into the active site. These findings are consistent with crystallographic observations showing that εC is tightly accommodated within an otherwise intact active site architecture.[31]

Pronounced nucleotide-dependent differences are also observed in interactions with Y159 and N169 (Figure 3.9). In the εA-bound system, the only direct hydrogen bond between the nucleobase and the protein involves Y159, where the Y159 O–H donates a hydrogen bond to N3 of the base (31% occupancy). This relatively modest and transient

hydrogen-bonding pattern suggests that εA is stabilized without excessive anchoring within the pocket. In contrast, in the εC-bound system, the side chain of N169 forms a hydrogen bond between its N–H group and O2 of the nucleobase for 48% of the trajectory, while the Y159 hydroxyl group forms an additional hydrogen bond with N169 for 28% of the trajectory. The presence of the N169 interaction represents a key distinction from εA and is consistent with crystallographic evidence indicating that N169 contributes to εC binding.[31] The ε5mC-bound system exhibits the highest hydrogen-bond occupancy involving N169, with the N169 N–H donating a hydrogen bond to O2 of the nucleobase for 84% of the trajectory. In addition, Y159 forms a hydrogen bond between its hydroxyl group and the 5′-phosphate backbone (OP2) for 36% of the trajectory. The high occupancy of the interaction with N169 suggests strong binding of ε5mC within the active site. This increased support correlates with the observed deeper insertion of the nucleotide and reduced flexibility relative to εA.

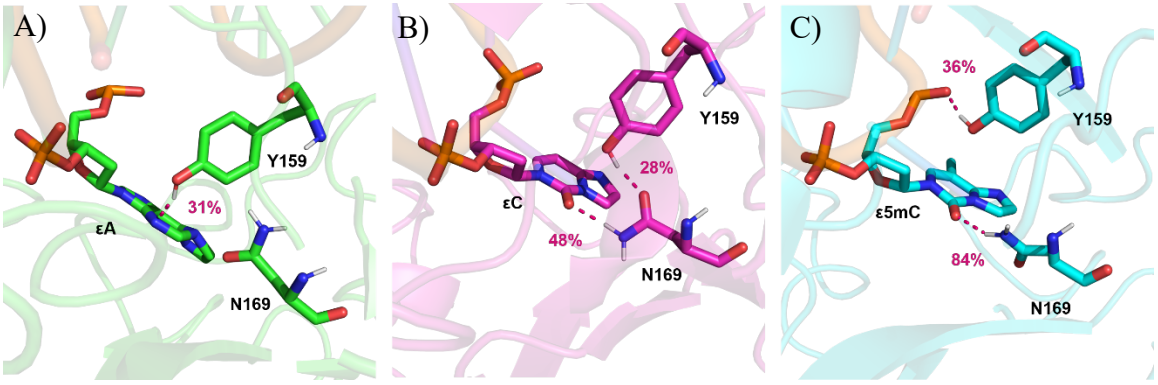


Figure 3.9. Representative MD structures of AAG bound to A) εA (green), B) εC (pink), or C) ε5mC (cyan) containing DNA, highlighting hydrogen-bonding interactions from Y159 or N169 and the nucleotide.

Crystallographic analysis of the εC-bound complex demonstrated that interaction with N169 may contribute to a conformation incompatible with efficient catalysis.[31] The present simulations extend this finding by showing that N169 engagement is not merely a static feature but is dynamically sustained throughout the trajectory. Importantly, because no structural information is available for ε5mC bound to AAG, these simulations provide the first molecular description of how this lesion is accommodated within the active site. The results indicate that ε5mC adopts a binding mode similar to that of εC, with persistent N169 interactions that stabilize the nucleobase within the catalytic pocket. This suggests that the inhibitory binding mode identified for εC may extend to ε5mC, providing a structural rationale for why AAG fails to efficiently excise this lesion. Notably, while π–π stacking interactions and the E125–Y127 catalytic dyad remain intact across all nucleotides, cytosine-derived etheno lesions enhance hydrogen bonding to H136 and especially N169. This increased support appears to promote tight binding within the pocket, potentially affecting the ability of the repair enzyme to initiate glycosidic bond cleavage. Collectively, these results indicate that nucleotide discrimination in AAG arises not from disruption of catalytic architecture but from hydrogen-bonding networks that favour alignment of εC and ε5mC within the active site. These interactions appear to promote inhibitory binding modes that impede progression to BER-mediated excision despite preservation of the catalytic framework.

3.3.1.4. εC and ε5mC Disrupt Active Site Solvation and Catalytic Water Positioning

To evaluate whether nucleotide-dependent differences in active site hydration contribute to the observed differences in repair outcome, the water density around the catalytic residue

E125 and the flipped base was analyzed for AAG bound to εA, εC, and ε5mC. Representative isolated water-density mesh plots were computed for each system after alignment to the protein heavy atoms, using a 0.45 Å grid encompassing the AAG active sites. Isodensity meshes were visualized to compare solvent access around the lesion and E125 across nucleotides (Figure 3.10).

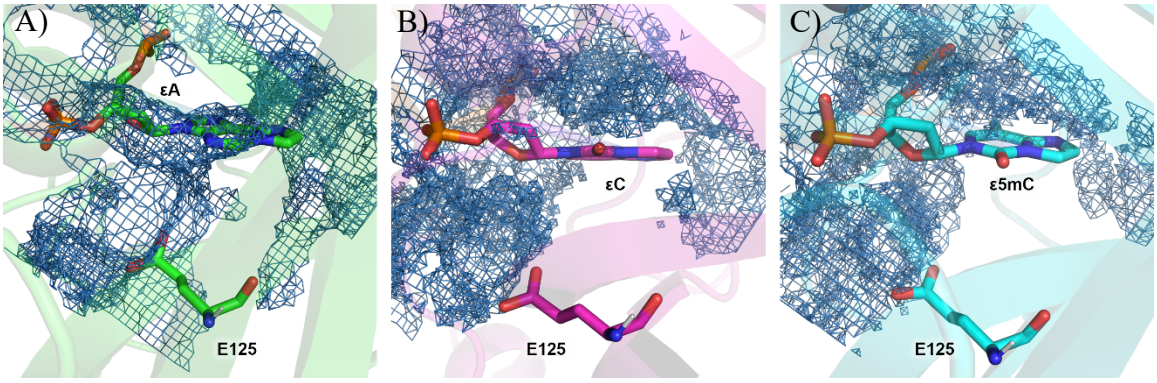


Figure 3.10. Representative MD structures of AAG bound to A) εA (green), B) εC (pink), or C) ε5mC (cyan) containing DNA, highlighting water density around E125 and the nucleobase and illustrated using an isolated water density mesh plot.

In the εA-bound system, the water-density plot indicates extensive hydration of the active site, with high water density surrounding the flipped εA base and extending into the region proximal to E125 (Figure 3.10). Notably, the mesh encompasses the carboxylate oxygens of the E125 side chain (OE1/OE2), indicating that the catalytic residue remains solvent exposed throughout the trajectory. This hydration pattern is compatible with a hydrolysis mechanism in which water must access the active site and be positioned for activation by E125 during glycosidic bond cleavage.

In contrast, both the εC- and ε5mC-bound systems exhibit a disruption of active site water density surrounding the nucleobase (Figure 3.10). In these complexes, water

molecules are distributed away from the immediate vicinity of the flipped base, resulting in reduced solvent density within the catalytic pocket compared to the εA-bound system. Additionally, the region around the E125 carboxylate shows reduced water density compared to the εA system, suggesting diminished solvent access to the catalytic activator. This difference is consistent with the deeper insertion and tighter anchoring of cytosine-derived etheno lesions observed in the simulations, which restricts solvent penetration into the binding cavity.

Table 3.1. Average E125–C1′ distance in AAG active site, with a standard deviation of 0.464 Å between substrates.

E125–C1’	Distance (Å) ± (δ^1)
εA	6.37 ± 1.2
εC	5.73 ± 1.1
ε5mC	5.47 ± 1.0

[1]Standard deviation of each data set.

To further quantify catalytic geometry relevant to hydrolytic activation, the distance between E125 and the nucleotide C1′ atom was measured across trajectories (Table 3.1). The εA system exhibits the largest average E125–C1′ distance of 6.37 ± 1.2 Å. In comparison, εC shows a reduced distance of 5.73 ± 1.1 Å, and ε5mC is further reduced to 5.47 ± 1.0 Å. This trend is consistent with previous glycosylase studies where hydrolysis activity is impacted by the ability of water to come in and out of the active site relative to the distance between the nucleophilic activator and C1′.[45] Importantly, the shorter E125–

C1′ distance does not imply improved catalytic competency. It is instead providing further insight when interpreted alongside the water-density analysis detailing that the closer approach of E125 to the sugar in εC and ε5mC coincides with decreased solvent between E125 and C1′ the catalytic pocket.

These results suggest that AAG-mediated excision is associated with a hydrated active site capable of supporting water entry, exchange, and positioning near both the nucleotide and the E125 carboxylate. While εC and ε5mC are accommodated within a conserved active site and retain key stacking and dyad interactions, the cytosine-derived lesions promote a more inserted binding conformation that disrupts the local solvation environment required for hydrolysis. This solvent disruption provides insights as to why these lesions bind tightly, yet fail to undergo productive glycosidic bond cleavage.[17, 31] The results presented on AAG provide a dynamic explanation for why εC and ε5mC behave as tightly bound inhibitors despite the ability of AAG to excise a wide range of chemically diverse lesions. The following section investigates how the oxidative repair enzyme ALKBH2 accommodates and processes these same nucleotides.

3.3.2. ALKBH2

Previous biochemical,[3, 10] structural,[19, 32] and computational studies[5] have provided important insight into the catalytic mechanism of ALKBH2 and related AlkB-family dioxygenases. Crystallographic analyses and kinetic investigations have shown that these enzymes employ a Fe(II)/α-KG-dependent oxidative mechanism to reverse alkyl and etheno DNA damage, while computational studies have further characterized the electronic

requirements for substrate oxidation.[3, 5, 18, 19, 32] Building on these findings, the proposed catalytic pathway for the understudied etheno lesion ε5mC (Figure 3.11) begins following nucleotide binding, where coordination of Fe(II) and α-KG within the ALKBH2 active site enables activation of molecular oxygen and formation of a high-valent Fe^{IV}–oxo species.[3, 46] Structural studies of ALKBH2–DNA complexes reveal that the active site is organized around a conserved two histidine and one aspartate metal coordination motif, in which H171, D173, and H236 coordinate the catalytic iron center together with succinate (the active form of α-KG) and a water molecule, forming the metal environment responsible for dioxygen activation.[32] The epoxide intermediate then initiates hydroxylation of the etheno bridge, producing an unstable glycol secondary intermediate. Subsequent breakdown releases glyoxal as a by-product while restoring the 5MeC base.[3]

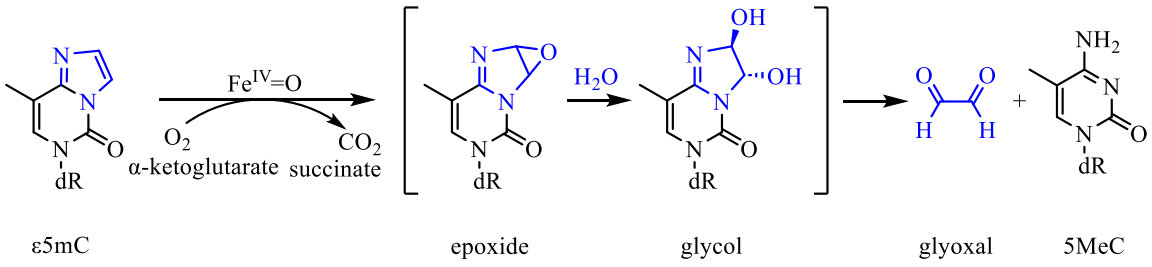


Figure 3.11. Proposed direct reversal repair pathway of etheno lesions to form an epoxide when starting from the understudied 3,N^4–etheno–5–methylcytosine (ε5mC) DNA lesion and the resulting repaired canonical structure of 5–methylcytosine (5MeC).

This oxidative repair mechanism differs fundamentally from the BER pathway employed by AAG. Rather than requiring nucleophilic attack at the glycosidic bond, ALKBH2 relies on precise positioning of the damaged carbon atoms (C7 and C8) relative to the Fe^{IV}–oxo center to enable hydrogen abstraction and oxidative cleavage.

Crystallographic analyses further show that substrate recognition in ALKBH2 is achieved through a combination of aromatic stacking and hydrogen-bonding interactions that position the flipped lesion within the catalytic pocket.[32] The flipped nucleobase is held within an aromatic environment formed primarily by F124 and H171, which help position the base for catalysis, while Y122 forms hydrogen-bond interactions with the nucleobase that help orient the lesion within the active site.[32] Additional interactions with the DNA backbone, including contacts involving S125, help anchor the flipped nucleotide within the duplex context. Furthermore, residues near the metal center, including R254, contribute to positioning of the catalytic oxo species and coordinating water molecules that participate in the oxidative reaction.[32]

Oxidation occurs only when the aberrant alkyl or etheno group is properly aligned with the Fe^{IV}–oxo center, therefore productive catalysis requires that the reactive nucleobase atoms approach the metal center within a narrow geometric window. Structural analyses have shown that the modified carbon atoms of damaged bases must be positioned in close proximity to the catalytic iron-oxo intermediate for efficient oxidation to occur.[32] Previous structural and computational studies of AlkB family enzymes have demonstrated that substrate specificity is largely dictated by geometric alignment of the damaged nucleobase relative to the catalytic Fe^{IV}–oxo center, as efficient substrates position their reactive carbon atoms near the oxidizing metal–oxo species while non-substrates place these atoms at larger distances or disrupt key active-site interactions.[3, 5, 16, 21, 24, 32] Experimental studies have established that ALKBH2 efficiently repairs both εC and ε5mC;[3] however, the structural basis for this activity remains poorly understood, as crystal structures of ALKBH2 bound to these cytosine-derived etheno lesions are not currently available. In

particular, it is unclear how the enzyme accommodates the additional methyl substituent of ε5mC while maintaining the geometric alignment between the reactive nucleobase atoms and the Fe^{IV}–oxo center required for productive oxidation. Understanding how ALKBH2 preserves these catalytic geometries despite structural differences between the lesions is therefore critical for rationalizing its repair activity.

3.3.2.1. π-Stacking and S125 Interaction with 5′-Phosphate Backbone Promotes an Alignment that is Catalytically Conducive in ALKBH2

Crystal structures of ALKBH2 bound to alkylated nucleotides (3MeC), reveal that direct hydrogen bonds between the nucleobase and active site residues are minimal, with nucleotide positioning dictated primarily by aromatic stacking interactions within the catalytic active site.[32] The structural stability of the ALKBH2–DNA complexes was evaluated by monitoring the all-heavy-atom RMSD relative to the first frame of each trajectory (Figure B.2, Appendix B). Across the 1 μs simulations, both εC- and ε5mC-bound systems remain structurally stable with RMSD values fluctuating around approximately 1.8–3.2 Å relative to the starting crystal structure. While individual replicates exhibit modest shifts in RMSD over time, these changes remain within a narrow range and do not indicate large-scale conformational rearrangements of the protein–DNA complex. Collectively, these results indicate that ALKBH2 maintains a stable global position throughout the simulations, suggesting that differences observed in subsequent analyses arise from local variations in nucleotide positioning and active-site interactions rather than from global structural instability.

In these structures, π–π interactions between the flipped base and residues H171 and F124 provide key stabilizing contacts that orient the lesion relative to the metal center. Such interactions are ubiquitous in DNA repair enzymes and have been shown to play a central role in positioning alkylated modifications during AAG-mediated repair, suggesting a conserved structural strategy for stabilizing flipped bases across repair pathways.[5] Consistent with these observations, π-stacking interactions were maintained throughout the MD simulations for both εC- and ε5mC-bound ALKBH2 systems (Figure 3.12). For εC, stacking between the base and F124 was observed at an average center-of-mass distance of 3.9 Å with 55% occupancy, while stacking with H171 occurred at a distance of 4.1 Å with 47% occupancy. In the ε5mC system, stacking with F124 became more persistent, occurring at a distance of 3.6 Å for 87% of the trajectory, and interaction with H171 was maintained at 4.4 Å for 49% of the simulation. These distances fall within the ~3.5–4.5 Å range previously reported for productive AlkB family stacking interactions, including those observed for εA and εC nucleotides for a reduced MD simulation duration.[5]

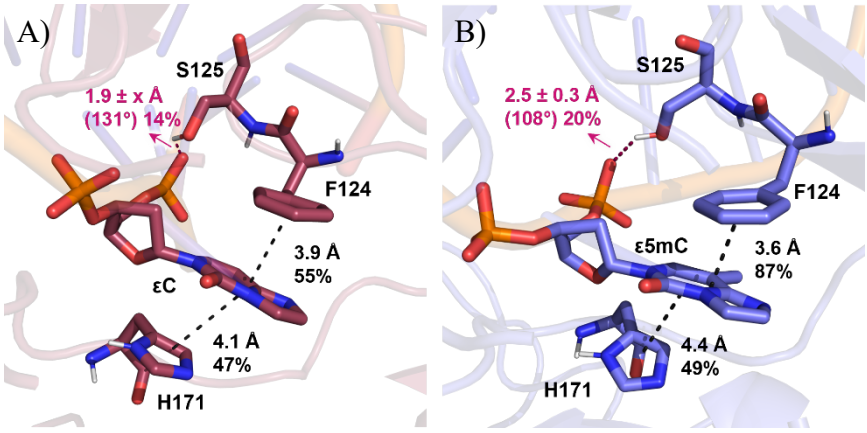


Figure 3.12. Representative MD structures of ALKBH2 bound to A) εC (burgundy) or B) ε5mC (blue) containing DNA, highlighting the π-Interactions with a commonly observed phosphate backbone hydrogen bond in ALKBH2 repaired substrates.

In addition to aromatic stacking, the S125 side chain (OH) forms a hydrogen bond with the 5′-phosphate backbone (OP2) in both εC and ε5mC complexes. Crystal structures of AlkB family enzymes have proposed that this interaction contributes to organization of a closed, catalytically competent active site configuration.[32] The presence of S125 hydrogen bonding in the present MD simulations suggests that nucleobase anchoring in ALKBH2 differs fundamentally from AAG-mediated binding, where excessive hydrogen bonding restricts catalytic flexibility. Instead, the combination of π–stacking and S125 interaction appears to guide the lesion into a geometry of a flipped base into the pocket containing the Fe^{IV}–oxo center, in which alignment is required for oxidative chemistry. These observations support a model in which π–interactions play a dual role in ALKBH2-mediated repair by stabilizing the flipped base. Similar π-contacts have been observed in other Fe(II)/α-KG-dependent dioxygenases, such as TET2, where aromatic residues help orient methylated cytosines for oxidation.[9]

3.3.2.2. Fe^{IV}–oxo Octahedral Coordination is Conserved Across εC and ε5mC Complexes

To determine whether nucleotide alignment is impacted by variations in the catalytic metal center, the coordination of Fe^{IV}–oxo species was monitored throughout the simulations. Across all DNA–ALKBH2 structures, the iron center maintained a preferred octahedral coordination geometry consistent with that observed in crystallographic studies of AlkB family enzymes.[32] Average coordination distances remained consistent across all trajectories, with Fe^{IV}–H171/H236 distances of 2.0–2.3 Å, Fe^{IV}–succinate/water/A173 interactions ∼1.9 Å, and an Fe^{IV}–oxo bond length of ∼1.6 Å (Figure 3.13). Variations

between εC and ε5mC systems were minimal, with a standard deviation of 0.1 Å indicating that nucleotide identity does not perturb the geometry of the catalytic center. The Fe^{IV}–oxo bond lengths observed in the present study closely match that of previously reported quantum mechanics and molecular mechanics (QM/MM) investigations on the bacterial AlkB enzyme, which further supports the structural validity of the modeled active site.[46]

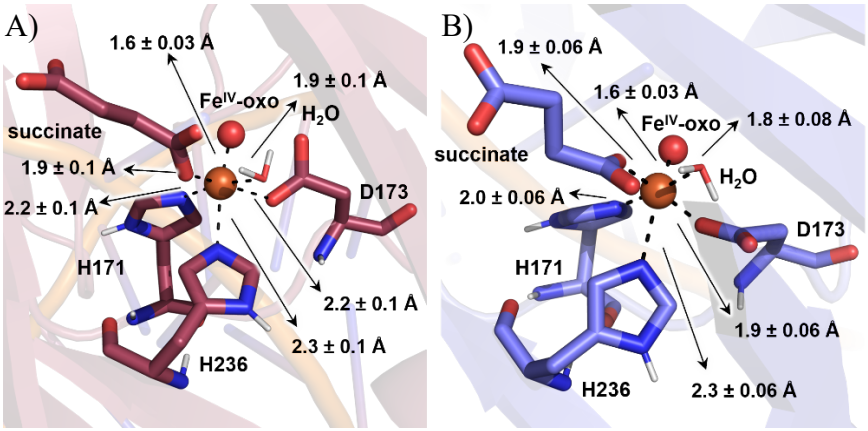


Figure 3.13. Representative MD structures of ALKBH2 bound to A) εC (burgundy) or B) ε5mC (blue) containing DNA, highlighting the Fe^{IV}-oxo center and the coordinated H171, H236, D173, succinate, water, and oxygen and their corresponding coordination distances.

3.3.2.3. Nucleotide-Dependent Hydrogen-Bond Networks Stabilize Active Site Organization in ALKBH2

While crystal structures of ALKBH2 typically show minimal direct hydrogen bonding between active site residues and the oxo ligand, the present simulations reveal transient interactions that may stabilize catalytically relevant conformations. One notable interaction involves the side chain of R254, which forms a hydrogen bond (NH1) to the Fe^{IV}–oxo oxygen for approximately 31% of the simulation in the εC-bound and for 49% (NH1) and 67% (NH2) of the trajectory in the ε5mC-bound (Figure 3.14). This interaction is not

prominently observed in static crystal structures, suggesting that R254 may participate in dynamic electrostatic organization of the reactive oxo species during catalysis. The increased occupancy observed in the ε5mC-bound complex may contribute to maintaining alignment despite the larger conformational variability of this nucleotide.

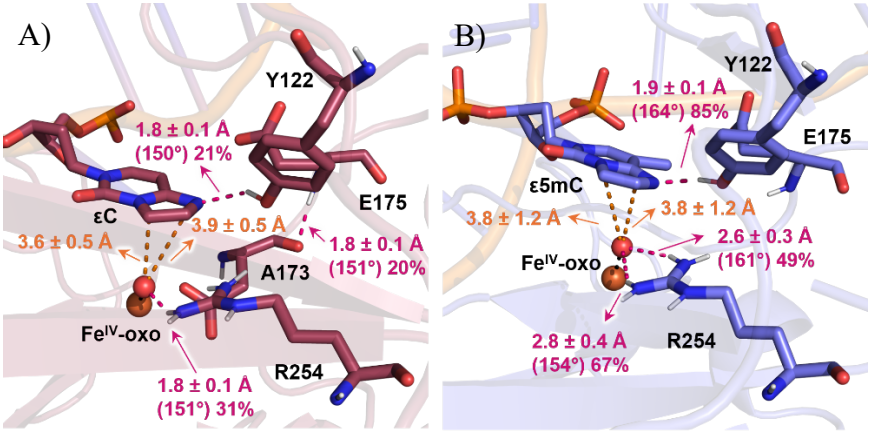


Figure 3.14. Representative MD structures of ALKBH2 bound to A) εC (burgundy) or B) ε5mC (blue) containing DNA, highlighting the nucleobase (C7 and C8 position) to Fe^{IV}-oxo distance and surrounding hydrogen-bonding interactions.

Additional nucleotide specific interactions were observed between Y122 and the nucleobase. A hydrogen bond between Y122 (OH) and the N9 position of the nucleotide is present for ~21% of the εC simulation and increases to 85% occupancy in the ε5mC system. This interaction may help guide rotational positioning of the nucleobase relative to the iron center, potentially compensating for steric effects introduced by the methyl modification. Hydrogen-bond networks were also identified involving residues coordinating the metal center. In the εC-bound system, interactions between A173 and E175 occur at ~81% occupancy, with additional hydrogen bonding of A173 to the water

ligand coordinated to iron observed for ~71% of the simulation. These interactions may contribute to maintaining the integrity of the octahedral coordination sphere while permitting subtle adjustments in nucleotide positioning. Rather than disrupting metal geometry, these transient networks appear to stabilize the catalytic configuration while allowing dynamic sampling of the available conformations.

3.3.2.4. Dynamic Distance to C7 and C8 Enables Oxidative Repair of εC and ε5mC in ALKBH2

Reactive C7 and C8 atoms of the lesion and the Fe^{IV}–oxo center maintain alignment for approximately 68% of the trajectory for εC and 91% occupancy for ε5mC, indicating that both nucleotides frequently adopt geometries compatible with oxidative chemistry (alignment of C7 and C8 with the lesion).

In the εC-bound system, the structure displays reactive carbon distances of approximately 3.6 Å (C7) and 3.9 Å (C8) from the Fe^{IV}–oxo center, with a standard deviation of ~0.5 Å (Figure 3.14). This relatively narrow distribution indicates that εC remains consistently positioned near the reactive oxo species during the simulation, representing a positive oxidative alignment. In the ε5mC system, the corresponding distances are approximately 3.8 Å (C7) and 3.8 Å (C8); however, the associated standard deviation is larger (~1.2 Å), indicating greater dynamical variation in reactive-carbon positioning. Importantly, this increased variance does not contradict observed repair competency.[5] Instead, it suggests that ε5mC samples a broader ensemble of geometries, including catalytically competent states, and is therefore more dynamic. Given that ε5mC contains additional steric and electronic complexity relative to εC, the increased

conformational variation may reflect the dynamic adjustments required to accommodate the methyl substituent while still allowing productive alignment for oxidation (Figure B.2, Appendix B).

Collectively, these results provide insights as to how ALKBH2 can repair both εC and ε5mC. Specifically, each lesion frequently adopts conformations that satisfy the geometric requirements for oxidation, even though ε5mC exhibits greater dynamical variability. This provides a dynamic explanation for the experimentally-observed repair of ε5mC and highlights the importance of dynamic sampling of ALKBH2-mediated oxidative reversal in revealing catalytically competent states that are not evident from static structural models of bulky, non-canonical lesions.

3.4. Conclusions

The combined MD simulation data presented here provide a mechanistic explanation for the contrasting activity of AAG and ALKBH2 toward cytosine-derived etheno lesions. Although AAG maintains a structurally conserved active site architecture across εA, εC, and ε5mC complexes, the cytosine-derived nucleotides adopt more deeply inserted into the active site binding mode that are stabilized by persistent hydrogen bonding and reduced active site solvation. These interactions restrict catalytic hydration and prevent productive positioning of the glycosidic bond relative to the nucleophilic activator E125, resulting in tightly bound inhibitor complexes. The inability of AAG to repair εC and ε5mC therefore arises not from failure to recognize or bind the lesions, but rather from containing the nucleotide in a hydrolytically inactive configuration. On the other hand, MD simulations

show ALKBH2 maintains a catalytically-competent active site and employs aromatic stacking interactions to guide nucleotides into the ALKBH2 active site in an alignment compatible with oxidative repair. Rather than forming rigid hydrogen-bonding networks, εC and ε5mC remain dynamically positioned within the active site, frequently sampling conformations in which the reactive C7 and C8 atoms approach the Fe^{IV}–oxo center at distances consistent with productive oxidation. This dynamic positioning enables ALKBH2 to access catalytically-competent geometries required for oxidative attack by the Fe^{IV}–oxo species, thereby facilitating efficient direct reversal of cytosine-derived etheno lesions.

Together, these findings reveal fundamentally distinct mechanisms for lesion handling between BER glycosylases and AlkB family dioxygenases. Where AAG relies on precise catalytic hydration and base activation that can be inhibited by overly restrictive positioning of the nucleotide within the active site, ALKBH2 leverages dynamic nucleotide positioning and oxidative chemistry to accommodate structurally diverse lesions, including bulky etheno-modified cytosines. The ability of ALKBH2 to repair εC and ε5mC therefore emerges not from stronger binding interactions, but from maintaining a balance between stabilization and mobility that enables repeated sampling of oxidation-competent states. These results provide a unified structural framework for understanding pathway partitioning between BER and direct-reversal repair mechanisms for etheno lesions.

More broadly, this work provides a framework for understanding how cells resolve chemically complex lesions that occur on epigenetically modified bases such as 5MeC. By revealing how cytosine-derived etheno lesions inhibit BER while remaining competent substrates for ALKBH2-mediated repair, these findings offer insight into how genome

maintenance systems preserve both genetic and epigenetic information following chemical damage. Such mechanistic insight will help guide future studies investigating lesion processing by related repair enzymes and may inform efforts to predict or manipulate repair pathway choice in contexts where DNA damage and epigenetic regulation intersect.

3.5. References

(1) Moore, L. D.; Le, T.; Fan, G. DNA Methylation and Its Basic Function. *Neuropsychopharmacology* **2013**, *38* (1), 23-38. DOI: 10.1038/npp.2012.112.

(2) Kim, M.; Costello, J. DNA methylation: an epigenetic mark of cellular memory. *Exp. Mol. Med.* **2017**, *49* (4), e322-e322. DOI: 10.1038/emm.2017.10.

(3) Ma, J.; Qi, R.; Harcourt, Emily M.; Chen, Y.-T.; Barbosa, Giovannia M.; Peng, Z.; Howarth, S.; Delaney, S.; Li, D. 3,N4-Etheno-5-methylcytosine blocks TET1-3 oxidation but is repaired by ALKBH2, 3 and FTO. *Nucleic Acids Res.* **2024**, *52* (20), 12378-12389. DOI: 10.1093/nar/gkae818.

(4) Kumar, S.; Chinnusamy, V.; Mohapatra, T. Epigenetics of Modified DNA Bases: 5-Methylcytosine and Beyond. *Front. Genet.* **2018**, *9*, 640. DOI: 10.3389/fgene.2018.00640.

(5) Lenz, S. A. P.; Li, D.; Wetmore, S. D. Insights into the Direct Oxidative Repair of Etheno Lesions: MD and QM/MM Study on the Substrate Scope of ALKBH2 and AlkB. *DNA Repair* **2020**, *96*, 102944. DOI: 10.1016/j.dnarep.2020.102944.

(6) Delaney, J. C.; Smeester, L.; Wong, C.; Frick, L. E.; Taghizadeh, K.; Wishnok, J. S.; Drennan, C. L.; Samson, L. D.; Essigmann, J. M. AlkB reverses etheno DNA lesions caused by lipid oxidation in vitro and in vivo. *Nat. Struct. Mol. Biol.* **2005**, *12* (10), 855-860. DOI: 10.1038/nsmb996.

(7) Nair, J. G., R. W.; Nair, U.; Owen, R. W.; Hull, W. E.; Bartsch, H. Identification of 3,N4-Etheno-5-Methyl-2′-Deoxycytidine in Human DNA: A New Modified Nucleoside Which May Perturb Genome Methylation. *Chem. Res. Toxicol.* **2012**, *25* (1), 162-169. DOI: 10.1021/tx200392a.

(8) Nair, U.; Bartsch, H.; Nair, J. Lipid peroxidation-induced DNA damage in cancer-prone inflammatory diseases: a review of published adduct types and levels in humans. *Free Radical Biol. Med.* **2007**, *43* (8), 1109-1120. DOI: 10.1016/j.freeradbiomed.2007.07.012.

(9) Hu, L.; Li, Z.; Cheng, J.; Rao, Q.; Gong, W.; Liu, M.; Shi, Y. G.; Zhu, J.; Wang, P.; Xu, Y. Crystal Structure of TET2-DNA Complex: Insight into TET-Mediated 5mC Oxidation. *Cell* **2013**, *155* (7), 1545-1555. DOI: 10.1016/j.cell.2013.11.020.

(10) Peng, Z.; Ma, J.; Christov, C. Z.; Karabencheva-Christova, T.; Lehnert, N.; Li, D. Kinetic Studies on the 2-Oxoglutarate/Fe(II)-Dependent Nucleic Acid Modifying Enzymes from the AlkB and TET Families. *DNA* **2023**, *3* (2), 65-84. DOI: 10.3390/dna3020005.

(11) Kaur, R.; Nikkel, D. J.; Wetmore, S. D. Computational studies of DNA repair: Insights into the function of monofunctional DNA glycosylases in the base excision repair pathway. *WIREs Comput. Mol. Sci.* **2020**, *10* (5), e1471. DOI: 10.1002/wcms.1471.

(12) Falnes, P.; Johansen, R. F.; Seeberg, E. AlkB-mediated oxidative demethylation reverses DNA damage in Escherichia coli. *Nature* **2002**, *419* (6903), 178-182. DOI: 10.1038/nature01048.

(13) Müller, T. A.; Hausinger, R. P. AlkB and Its Homologues – DNA Repair and Beyond. In *2-Oxoglutarate-Dependent Oxygenases*, Schofield, C., Hausinger, R. Eds.; The Royal Society of Chemistry, 2015; pp 246-262.

(14) Lenz, S. A. P.; Wetmore, S. D. Evaluating the Substrate Selectivity of Alkyladenine DNA Glycosylase: The Synergistic Interplay of Active Site Flexibility and Water Reorganization. *Biochemistry* **2016**, *55* (5), 798-808. DOI: 10.1021/acs.biochem.5b01179.

(15) Wang, L.; Xi, K.; Zhu, L.; Da, L.-T. DNA Deformation Exerted by Regulatory DNA-Binding Motifs in Human Alkyladenine DNA Glycosylase Promotes Base Flipping. *J. Chem. Inf. Model.* **2022**, *62* (13), 3213-3226. DOI: 10.1021/acs.jcim.2c00091.

(16) Fu, D.; Samson, L. D. Direct repair of 3,N4-ethenocytosine by the human ALKBH2 dioxygenase is blocked by the AAG/MPG glycosylase. *DNA Repair* **2012**, *11* (1), 46-52. DOI: 10.1016/j.dnarep.2011.10.004.

(17) Lau, A. Y.; Wyatt, M. D.; Glassner, B. J.; Samson, L. D.; Ellenberger, T. Molecular basis for discriminating between normal and damaged bases by the human alkyladenine glycosylase, AAG. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97* (25), 13573-13578. DOI: doi:10.1073/pnas.97.25.13573.

(18) Fedeles, B. I.; Singh, V.; Delaney, J. C.; Li, D.; Essigmann, J. M. The AlkB Family of Fe(II)/α-Ketoglutarate-dependent Dioxygenases: Repairing Nucleic Acid Alkylation Damage and Beyond. *J. Biol. Chem.* **2015**, *290* (34), 20734-20742. DOI: 10.1074/jbc.R115.656462.

(19) Bian, K.; Lenz, S. A. P.; Tang, Q.; Chen, F.; Qi, R.; Jost, M.; Drennan, C. L.; Essigmann, J. M.; Wetmore, S. D.; Li, D. DNA repair enzymes ALKBH2, ALKBH3, and AlkB oxidize 5-methylcytosine to 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine in vitro. *Nucleic Acids Res.* **2019**, *47* (11), 5522-5529. DOI: 10.1093/nar/gkz395.

(20) Trewick, S. C.; Henshaw, T. F.; Hausinger, R. P.; Lindahl, T.; Sedgwick, B. Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. *Nature* **2002**, *419* (6903), 174-178. DOI: 10.1038/nature00908.

(21) Duncan, T.; Trewick, S. C.; Koivisto, P.; Bates, P. A.; Lindahl, T.; Sedgwick, B. Reversal of DNA alkylation damage by two human dioxygenases. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (26), 16660-16665. DOI: 10.1073/pnas.262589799.

(22) Breiling, A.; Lyko, F. Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics Chromatin* **2015**, *8* (1), 24. DOI: 10.1186/s13072-015-0016-6.

(23) Rutledge, L. R.; Wetmore, S. D. Modeling the Chemical Step Utilized by Human Alkyladenine DNA Glycosylase: A Concerted Mechanism Aids in Selectively Excising Damaged Purines. *J. Am. Chem. Soc.* **2011**, *133* (40), 16258-16269. DOI: 10.1021/ja207181c.

(24) Guliaev, A. B.; Hang, B.; Singer, B. Structural insights by molecular dynamics simulations into differential repair efficiency for ethano-A versus etheno-A adducts by the human alkylpurine-DNA N -glycosylase. *Nucleic Acids Res.* **2002**, *30* (17), 3778-3787. DOI: 10.1093/nar/gkf494.

(25) Waheed, S. O.; Ramanan, R.; Chaturvedi, S. S.; Lehnert, N.; Schofield, C. J.; Christov, C. Z.; Karabencheva-Christova, T. G. Role of Structural Dynamics in Selectivity and Mechanism of Non-heme Fe(II) and 2-Oxoglutarate-Dependent Oxygenases Involved in DNA Repair. *ACS Cent. Sci.* **2020**, *6* (5), 795-814. DOI: 10.1021/acscentsci.0c00312.

(26) Berger, M. B.; Walker, A. R.; Vázquez-Montelongo, E. A.; Cisneros, G. A. Computational investigations of selected enzymes from two iron and α-ketoglutarate-dependent families. *Phys. Chem. Chem. Phys.* **2021**, *23* (39), 22227-22240. DOI: 10.1039/d1cp03800a.

(27) Nikkel, D. J.; Wetmore, S. D. Distinctive Formation of a DNA–Protein Cross-Link during the Repair of DNA Oxidative Damage: Insights into Human Disease from MD Simulations and QM/MM Calculations. *J. Am. Chem. Soc.* **2023**, *145* (24), 13114-13125. DOI: 10.1021/jacs.3c01773.

(28) Nikkel, D. J.; Wetmore, S. D. A tale of two mechanisms: Clarification of the pathway for MBD4 catalyzed glycosidic bond cleavage using MD and QM/MM calculations. *DNA Repair* **2026**, *157*, 103917. DOI: 10.1016/j.dnarep.2025.103917.

(29) Wang, W. J.; Wang, T.; Zhao, Y.; Li, B. N.; Chen, D. Z. Theoretical Insights into N-Glycoside Bond Cleavage of 5-Carboxycytosine by Thymine DNA Glycosylase: A QM/MM Study. *J. Phys. Chem. B* **2024**, *128* (19), 4621-4630. DOI: 10.1021/acs.jpcb.4c00221.

(30) Frederickson, A.; Nikkel, D.; Mansour, B.; Wetmore, S. D. Computational Investigation of the Uracil DNA Glycosylase MsmUdgX Catalytic Mechanism of Action Using MD and QM/MM Techniques. Lethbridge, U. o., Ed.; 2026; pp 1-31.

(31) Lingaraju, G. M.; Davis, C. A.; Setser, J. W.; Samson, L. D.; Drennan, C. L. Structural Basis for the Inhibition of Human Alkyladenine DNA Glycosylase (AAG) by 3,N4-Ethenocytosine-containing DNA*. *J. Biol. Chem.* **2011**, *286* (15), 13205-13213. DOI: 10.1074/jbc.M110.192435.

(32) Yi, C.; Chen, B.; Qi, B.; Zhang, W.; Jia, G.; Zhang, L.; Li, C. J.; Dinner, A. R.; Yang, C. G.; He, C. Duplex interrogation by a direct DNA repair protein in search of base damage. *Nat. Struct. Mol. Biol.* **2012**, *19* (7), 671-676. DOI: 10.1038/nsmb.2320.

(33) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46* (W1), W296-w303. DOI: 10.1093/nar/gky427.

(34) Schrodinger. *The Pymol Molecular Graphics System*, Version 3.0; Schrodinger LLC, 2020.

(35) Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* **2012**, *40* (Web Server issue), W537-541. DOI: 10.1093/nar/gks375.

(36) Li, P.; Merz, K. M., Jr. MCPB.py: A Python Based Metal Center Parameter Builder. *J. Chem. Inf. Model.* **2016**, *56* (4), 599-604. DOI: 10.1021/acs.jcim.5b00674.

(37) Seminario, J. M. Calculation of intramolecular force fields from second-derivative tensors. *Int. J. Quantum Chem* **1996**, *60* (7), 1271-1277.

(38) Schmit, J. D.; Kariyawasam, N. L.; Needham, V.; Smith, P. E. SLTCAP: A Simple Method for Calculating the Number of Ions Needed for MD Simulation. *J. Chem. Theory Comput.* **2018**, *14* (4), 1823-1827. DOI: 10.1021/acs.jctc.7b01254.

(39) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696-3713. DOI: 10.1021/acs.jctc.5b00255.

(40) Zgarbová, M.; Šponer, J.; Otyepka, M.; Cheatham, T. E., 3rd; Galindo-Murillo, R.; Jurečka, P. Refinement of the Sugar-Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. *J. Chem. Theory Comput.* **2015**, *11* (12), 5723-5736. DOI: 10.1021/acs.jctc.5b00716.

(41) He, X.; Man, V. H.; Yang, W.; Lee, T. S.; Wang, J. A fast and high-quality charge model for the next generation general AMBER force field. *J. Chem. Phys.* **2020**, *153* (11), 114502. DOI: 10.1063/5.0019056.

(42) Li, P.; Song, L. F.; Merz, K. M., Jr. Systematic Parameterization of Monovalent Ions Employing the Nonbonded Model. *J. Chem. Theory Comput.* **2015**, *11* (4), 1645-1657. DOI: 10.1021/ct500918t.

(43) F. Wang, J.-P. B., P. Cieplak & F.-Y. Dupradeau. *R.E.D. Python: Object oriented programming for Amber force fields*. 2013. (accessed.

(44) Case, D. A.; Aktulga, H. M.; Belfon, K.; Cerutti, D. S.; Cisneros, G. A.; Cruzeiro, V. W. D.; Forouzesh, N.; Giese, T. J.; Götz, A. W.; Gohlke, H.; et al. AmberTools. *J. Chem. Inf. Model.* **2023**, *63* (20), 6183-6191. DOI: 10.1021/acs.jcim.3c01153.

(45) Aroli, S.; Woo, E.-J.; Gopal, B.; Varshney, U. Mutational and structural analyses of UdgX: insights into the active site pocket architecture and its evolution. *Nucleic Acids Res.* **2023**, *51* (13), 6554-6565. DOI: 10.1093/nar/gkad486.

(46) Wang, B.; Usharani, D.; Li, C.; Shaik, S. Theory Uncovers an Unusual Mechanism of DNA Repair of a Lesioned Adenine by AlkB Enzymes. *J. Am. Chem. Soc.* **2014**, *136* (39), 13895-13901. DOI: 10.1021/ja507934g.

Chapter 4 : Conclusions and Future Directions

4.1. Summary

DNA is continuously subjected to spontaneous hydrolysis, oxidative stress, alkylation, and deamination, all of which generate lesions that alter the hydrogen bonding, electronic structure, or steric properties of nucleobases.[1, 2] If left unrepaired, these modifications can result in mutagenesis, replication arrest, or disruption of regulatory epigenetic markers.[3] DNA repair enzymes play a central role in preserving genomic stability by recognizing and processing chemically modified nucleobases.[4-8] These enzymes are particularly interesting since they must discriminate between structurally similar canonical bases, damaged lesions, and regulatory epigenetic modifications, often within highly conserved structural scaffolds.[9] This thesis used computational tools to understand the function of enzymes involved in the base excision repair (BER) and direct reversal repair (DRR) pathway.

BER is responsible for correcting small, non-bulky lesions through multiple steps.[10-12] This pathway is initiated by DNA glycosylases, which cleave the N–glycosidic bond to generate an abasic site that is subsequently processed by downstream repair enzymes.[1, 13] The chemistry of BER requires precise positioning of catalytic residues to activate a nucleophile and stabilize charge developed during glycosidic bond cleavage. Because this step commits the lesion to repair, substrate recognition and catalytic alignment must be tightly controlled. Although many glycosylases are highly substrate specific, others, such as alkyladenine DNA glycosylase (AAG), display broader substrate tolerance.[14, 15] However, even broad-specificity glycosylases impose strict geometric and electronic constraints on substrates to achieve productive catalysis. In contrast, DRR restores chemically modified bases without cleaving the glycosidic bond or disrupting the sugar–

phosphate backbone.[16] Members of the AlkB family of Fe(II)/α-ketoglutarate (α-KG)-dependent dioxygenases catalyze oxidative dealkylation reactions that directly reverse alkylation and etheno lesions.[17] Rather than generating an abasic intermediate, these enzymes hydroxylate the aberrant substituent, allowing spontaneous decomposition to regenerate the canonical nucleobase.[16] This chemistry requires precise alignment of the lesion relative to a reactive iron–oxo species and therefore imposes distinct structural and electronic requirements compared to BER.[18, 19] Although BER and DRR can target overlapping classes of lesions, the determinants that dictate pathway selection remain incompletely understood. In some cases, structurally similar lesions are processed by one pathway but not the other, despite apparent binding competence. These observations suggest that substrate discrimination arises not solely from binding interactions, but from compatibility with the underlying chemical mechanism of each repair pathway.

Although repair mechanisms have been proposed, details about the exact chemistry and repair pathway remain incompletely understood. In particular, it is often unclear why certain lesions are excised by glycosylases, oxidatively reversed by dioxygenases, or left unrepaired despite apparent binding competence. This thesis has contributed to a clearer picture of how active site architecture, substrate positioning, and mechanistic requirements collectively determine catalytic outcomes in representative BER and DRR enzymes. The primary objective of this thesis is to elucidate how msmUdgX, AAG, and ALKBH2 achieve substrate specificity and catalytic outcomes within these distinct repair frameworks, and to define how subtle differences in active site composition influence reaction pathway selection at the molecular level. This chapter summarizes the major findings for each chapter and provides details of possible future research avenues.

4.2. Contributions from Thesis

In Chapter 2, a mixed molecular dynamics (MD) simulation and quantum mechanics and molecular mechanics (QM/MM) approach was used to investigate the catalytic mechanism of uracil DNA glycosylase (UDG) *msm*UdgX. Unlike canonical UDG family members that catalyze hydrolytic N–glycosidic bond cleavage, UdgX forms a stable covalent DNA–protein crosslink through nucleophilic attack by H109 at C1′ of deoxyribose. MD simulations demonstrated that the HID tautomer promotes reduced imidazole flexibility, favorable Nε–C1′ alignment, and a structured hydrogen-bonding network involving E52, Q53, F65, N91, and H178 that supports nucleophilic attack. While HIE exhibited increased conformational variability that did not support crosslink formation. Subsequent QM/MM calculations identified a concerted transition state in which glycosidic bond cleavage and H109–C1′ bond formation occur simultaneously. The calculated free energy barrier (66.5 kJ/mol) is consistent with experimentally-derived activation parameters,[20] providing strong support for a concerted dissociative mechanism. This refines earlier experimental interpretations that described the reaction as S_N1-like, demonstrating that bond breaking and forming events are mechanistically coupled.

To further define pathway selection, H109 mutants were examined to evaluate how small changes in active site composition alter catalytic outcome. MD simulations of H109A revealed a loss of precise nucleophile positioning, with increased average distances between residue 109 and C1′ and enhanced solvent accessibility within the active site. Persistent water occupancy near C1′ was observed across replicates, consistent with a hydrolytic mechanism analogous to canonical UDGs. The enlarged cavity generated by

alanine substitution disrupts the organized hydrogen-bonding network observed in the wild-type enzyme and permits stabilization of a nucleophilic water molecule, explaining the experimentally-observed redirection toward hydrolysis.[11] The H109E mutant presented a more complex behavior. MD simulations indicated that substitution with glutamate increases active site polarity and solvent accessibility relative to the wild-type enzyme, again allowing water molecules to sample positions proximal to C1′. This geometry supports hydrolytic uracil excision, consistent with experimental observations that H109E can undergo hydrolysis.[10, 21] However, QM/MM calculations revealed that crosslink formation remains energetically feasible in this mutant, with formation of a covalent linkage between the glutamate side chain and C1′ proceeding through a concerted transition state analogous to the wild-type. Although altered electrostatics and side-chain geometry modulate alignment and activation, the calculated barrier supports the possibility of crosslink formation, rationalizing the experimentally-observed dual reactivity of H109E.[10, 13, 20, 21]

In Chapter 3, MD simulations of AAG and ALKBH2 repair enzymes were performed to provide detail explanation as to why AAG cannot repair 3,N^4–etheno–5–methylcytosine (ε5mC) by cleaving the nucleobase prior to BER, whereas ALKBH2 efficiently restores this lesion through oxidative DRR. MD simulations of AAG bound to εC, εA, and ε5mC demonstrated that substrate binding alone does not guarantee productive catalysis. Although εC and ε5mC can occupy the AAG active site, subtle changes in positioning, hydrogen-bonding networks, and active site solvation disrupt the geometry required for glycosidic bond activation. In particular, disrupted water organization and shifts in key catalytic residues prevent proper alignment for hydrolytic cleavage. These results reinforce

that the broad substrate scope of AAG is constrained by strict geometric requirements for reaction stabilization. In contrast, simulations of ALKBH2 revealed that εC and ε5mC adopt conformations that maintain productive alignment between the lesion and the reactive Fe^{IV}–oxo center. Dynamic distance analysis demonstrated that repair competency correlates with maintenance of a catalytically favorable iron coordination. Stabilizing π–stacking interactions and persistent hydrogen-bond networks further support oxidative activation. Importantly, ALKBH2 accommodates the additional methyl group of ε5mC without disrupting alignment, explaining its ability to preserve epigenetic information during repair.

Overall, this thesis highlights the mechanistic complexity of DNA repair enzymes and the interplay between active site architecture, tautomeric form, hydrogen-bonding networks, and solvent accessibility in determining catalytic outcome. The results demonstrate that subtle variations in nucleophile identity, residue positioning, and hydration can redirect reaction pathways within a conserved structural scaffold, as observed for UdgX and its mutants. Moreover, comparison of BER and DRR enzymes illustrates that these pathways impose fundamentally distinct mechanistic constraints. Whereas glycosylases such as UdgX and AAG require precise geometric and electronic alignment to facilitate N–glycosidic bond cleavage or covalent crosslink formation, ALKBH2 relies on accurate substrate positioning relative to a reactive iron–oxo species to enable oxidative reversal. This comparative analysis demonstrates that substrate specificity arises not solely from lesion binding, but from compatibility with pathway-specific chemical requirements. Further details of potential future research avenues are discussed in the following section.

4.3. Future Directions

This thesis demonstrates that subtle changes in active site composition, tautomeric form, and solvation can fundamentally alter catalytic outcome. These findings open several avenues for future investigation. First, QM/MM calculations could be extended to additional Chapter 2 UdgX mutants to quantify how incremental changes in cavity volume and hydration influence the free energy landscape of hydrolysis versus crosslink formation. Explicit free energy simulations, such as umbrella sampling, would provide further insight into competing reaction pathways and solvent-mediated activation. Second, comparative QM/MM analysis of Chapter 3 AAG bound to εC or ε5mC could directly quantify the energetic penalties associated with non-productive alignment, complementing the MD-based structural interpretation presented here.

Similarly, modeling of the full oxidative cycle of ALKBH2, including hydrogen abstraction and rebound steps, would provide a more complete description of ε5mC reversal. Future mechanistic studies employing QM/MM calculations would complement the present MD analyses by providing atomistic insight into the oxidative reaction pathway of ALKBH2 and how active site structure governs catalytic chemistry. In particular, QM/MM approaches would allow direct investigation of how changes in nucleotide positioning, active-site geometry, and interactions with catalytic residues influence key steps of the oxidative repair process, including hydrogen abstraction, formation of the ferryl intermediate, and subsequent rebound chemistry that restores the canonical base. By explicitly modeling the electronic structure of the reacting system, QM/MM calculations could characterize intermediate states, electronic rearrangements, and transition structures

that cannot be resolved through classical MD simulations alone, thereby providing a more complete description of the oxidative repair mechanism for the ε5mC modification.

The computational framework employed in this thesis can also be applied to other DNA glycosylases and Fe(II)/α-KG-dependent dioxygenases to investigate how conserved enzyme scaffolds evolve distinct catalytic functions. In particular, members of the NEIL family of bifunctional DNA glycosylases represent promising targets for further mechanistic investigation. These enzymes remove oxidatively damaged bases and subsequently cleave the DNA backbone through formation of a transient covalent intermediate during the base excision repair pathway. However, aspects of their catalytic mechanism and substrate specificity remain incompletely understood, particularly for NEIL3 (PDB ID: 8B9N),[22, 23] which has been reported to exhibit distinct substrate preferences and activity toward lesions in single-stranded DNA. Computational studies could therefore help clarify how substrate positioning and active site interactions influence catalytic activity in these enzymes. Similarly, additional members of the AlkB family of Fe(II)/α-KG-dependent dioxygenases, such as ALKBH3 (PDB ID: 8JNK)[24] and FTO, present attractive targets for mechanistic comparison. Although these enzymes share a conserved oxidative demethylation mechanism, they exhibit distinct substrate preferences and biological roles. Further computational investigations could help elucidate how subtle variations in active site architecture and substrate positioning influence oxidative repair chemistry across this enzyme family.

The ability to define transition states and catalytic determinants at atomic resolution has implications for inhibitor design. Since many small-molecule inhibitors function as transition-state mimics,[25] the mechanistic insights obtained here may inform the

development of selective modulators of both bacterial and human repair enzymes. In particular, UdgX presents unique active site pockets and a suicide crosslinking mechanism that could be exploited for the design of selective inhibitors or covalent probes, although such modulators remain to be discovered. Likewise, members of the human AlkB family have emerged as therapeutic targets because their oxidative demethylase activity contributes to tumor progression and resistance to alkylating chemotherapeutic agents, highlighting the value of mechanistic understanding for future inhibitor design.[26]

Overall, this thesis establishes a mechanistic framework linking active site architecture, solvation dynamics, and electronic activation to catalytic outcome in DNA repair enzymes. By integrating dynamic and quantum mechanical approaches, this thesis contributes to a deeper understanding of how repair enzymes discriminate between chemically similar lesions and preserve both genomic and epigenetic integrity.

4.4. References

(1) Kaur, R.; Nikkel, D. J.; Wetmore, S. D. Computational studies of DNA repair: Insights into the function of monofunctional DNA glycosylases in the base excision repair pathway. *WIREs Comput. Mol. Sci.* **2020**, *10* (5), e1471. DOI: 10.1002/wcms.1471.

(2) Stivers, J. T.; Jiang, Y. L. A Mechanistic Perspective on the Chemistry of DNA Repair Glycosylases. *Chem. Rev.* **2003**, *103* (7), 2729-2760. DOI: 10.1021/cr010219b.

(3) Nora M. Al Aboud, C. T., Ishwarlal Jialal. Genetics, Epigenetic Mechanism. *StatPearls* **2023**.

(4) Minchin, S.; Lodge, J. Understanding biochemistry: structure and function of nucleic acids. *Essays Biochem.* **2019**, *63* (4), 433-456. DOI: 10.1042/ebc20180038.

(5) Houlihan, G.; Arangundy-Franklin, S.; Holliger, P. Exploring the Chemistry of Genetic Information Storage and Propagation through Polymerase Engineering. *Acc. Chem. Res.* **2017**, *50* (4), 1079-1087. DOI: 10.1021/acs.accounts.7b00056.

(6) Carusillo, A.; Mussolino, C. DNA Damage: From Threat to Treatment. *Cells* **2020**, *9* (7). DOI: 10.3390/cells9071665.

(7) Barnes, J. L.; Zubair, M.; John, K.; Poirier, M. C.; Martin, F. L. Carcinogens and DNA damage. *Biochem. Soc. Trans.* **2018**, *46* (5), 1213-1224. DOI: 10.1042/bst20180519.

(8) Hakem, R. DNA-damage repair; the good, the bad, and the ugly. *EMBO J.* **2008**, *27* (4), 589-605. DOI: 10.1038/emboj.2008.15.

(9) Kumar, S.; Chinnusamy, V.; Mohapatra, T. Epigenetics of Modified DNA Bases: 5-Methylcytosine and Beyond. *Front. Genet.* **2018**, *9*, 640. DOI: 10.3389/fgene.2018.00640.

(10) Jia, Q.; Zeng, H.; Tu, J.; Sun, L.; Cao, W.; Xie, W. Structural insights into an MsmUdgX mutant capable of both crosslinking and uracil excision capability. *DNA Repair* **2021**, *97*, 103008. DOI: 10.1016/j.dnarep.2020.103008.

(11) Aroli, S.; Woo, E.-J.; Gopal, B.; Varshney, U. Mutational and structural analyses of UdgX: insights into the active site pocket architecture and its evolution. *Nucleic Acids Res.* **2023**, *51* (13), 6554-6565. DOI: 10.1093/nar/gkad486.

(12) Ahn, W.-C.; Aroli, S.; Kim, J.-H.; Moon, J. H.; Lee, G. S.; Lee, M.-H.; Sang, P. B.; Oh, B.-H.; Varshney, U.; Woo, E.-J. Covalent binding of uracil DNA glycosylase UdgX to abasic DNA upon uracil excision. *Nat. Chem. Biol.* **2019**, *15* (6), 607-614. DOI: 10.1038/s41589-019-0289-3.

(13) Liang, C.; Yang, Y.; Ning, P.; Chang, C.; Cao, W. Structural and functional coupling in cross-linking uracil-DNA glycosylase UDGX. *Biosci. Rep.* **2024**, *44* (1). DOI: 10.1042/bsr20231551.

(14) Lenz, S. A. P.; Wetmore, S. D. Evaluating the Substrate Selectivity of Alkyladenine DNA Glycosylase: The Synergistic Interplay of Active Site Flexibility and Water Reorganization. *Biochemistry* **2016**, *55* (5), 798-808. DOI: 10.1021/acs.biochem.5b01179.

(15) Wang, L.; Xi, K.; Zhu, L.; Da, L.-T. DNA Deformation Exerted by Regulatory DNA-Binding Motifs in Human Alkyladenine DNA Glycosylase Promotes Base Flipping. *J. Chem. Inf. Model.* **2022**, *62* (13), 3213-3226. DOI: 10.1021/acs.jcim.2c00091.

(16) Ma, J.; Qi, R.; Harcourt, Emily M.; Chen, Y.-T.; Barbosa, Giovannia M.; Peng, Z.; Howarth, S.; Delaney, S.; Li, D. 3,N4-Etheno-5-methylcytosine blocks TET1-3 oxidation but is repaired by ALKBH2, 3 and FTO. *Nucleic Acids Res.* **2024**, *52* (20), 12378-12389. DOI: 10.1093/nar/gkae818.

(17) Kuznetsov, N. A.; Kanazhevskaya, L. Y.; Fedorova, O. S. DNA Demethylation in the Processes of Repair and Epigenetic Regulation Performed by 2-Ketoglutarate-Dependent DNA Dioxygenases. In *Int. J. Mol. Sci.*, 2021; Vol. 22, p 10540.

(18) Müller, T. A.; Hausinger, R. P. AlkB and Its Homologues – DNA Repair and Beyond. In *2-Oxoglutarate-Dependent Oxygenases*, Schofield, C., Hausinger, R. Eds.; The Royal Society of Chemistry, 2015; pp 246-262.

(19) Delaney, J. C.; Smeester, L.; Wong, C.; Frick, L. E.; Taghizadeh, K.; Wishnok, J. S.; Drennan, C. L.; Samson, L. D.; Essigmann, J. M. AlkB reverses etheno DNA lesions caused by lipid oxidation in vitro and in vivo. *Nat. Struct. Mol. Biol.* **2005**, *12* (10), 855-860. DOI: 10.1038/nsmb996.

(20) Liang, C. Biochemical Analyses of UDGX-A Crosslinking Uracil-Dna Glycosylase. *Diss. Abstr. Int.* **2023**, (3494).

(21) Tu, J.; Chen, R.; Yang, Y.; Cao, W.; Xie, W. Suicide inactivation of the uracil DNA glycosylase UdgX by covalent complex formation. *Nat. Chem. Biol.* **2019**, *15* (6), 615-622. DOI: 10.1038/s41589-019-0290-x.

(22) Liu, M.; Doublié, S.; Wallace, S. S. Neil3, the final frontier for the DNA glycosylases that recognize oxidative damage. *Mutat. Res.* **2013**, *743-744*, 4-11. DOI: 10.1016/j.mrfmmm.2012.12.003.

(23) Klima, M., Boura, E., Silhan, J. Crystal structure of NEI domain of mouse NEIL3 trapped in covalent complex with ssDNA with abasic site. *Protein Data Bank* **2026**, *PDB ID 8B9N*. DOI: 10.2210/pdb8b9n/pdb.

(24) Zhang, L.; Duan, H.-C.; Paduch, M.; Hu, J.; Zhang, C.; Mu, Y.; Lin, H.; He, C.; Kossiakoff, A. A.; Jia, G.; et al. The Molecular Basis of Human ALKBH3 Mediated RNA N1-methyladenosine (m1A) Demethylation. *Angew. Chem. Int. Ed.* **2024**, *63* (7), e202313900. DOI: 10.1002/anie.202313900.

(25) Yuen, P. K.; Green, S. A.; Ashby, J.; Lay, K. T.; Santra, A.; Chen, X.; Horvath, M. P.; David, S. S. Targeting Base Excision Repair Glycosylases with DNA Containing Transition State Mimics Prepared via Click Chemistry. *ACS Chem. Bio.* **2019**, *14* (1), 27-36. DOI: 10.1021/acschembio.8b00771.

(26) Li, Q.; Zhu, Q. The role of demethylase AlkB homologs in cancer. *Front. Oncol.* **2023**, *13*, 1153463. DOI: 10.3389/fonc.2023.1153463.

Appendix A

Supplemental Information for Chapter 2: Computational Investigation of the Uracil DNA Glycosylase MsmUdgX Catalytic Mechanism of Action Using MD and QM/MM Techniques

Contains Figures A.1–A.9 and Tables A.1–A.6

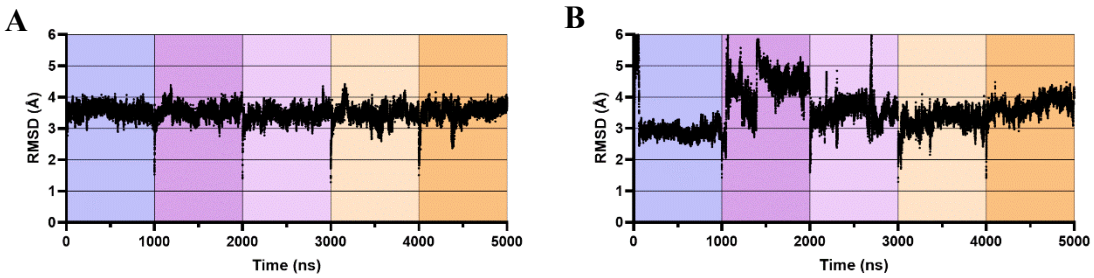


Figure A.1. All-heavy-atom RMSD relative to the first frame over the MD simulations for the wild-type systems A) HID and B) HIE. The five 1000 ns replicates composing each simulation are joined back-to-back.

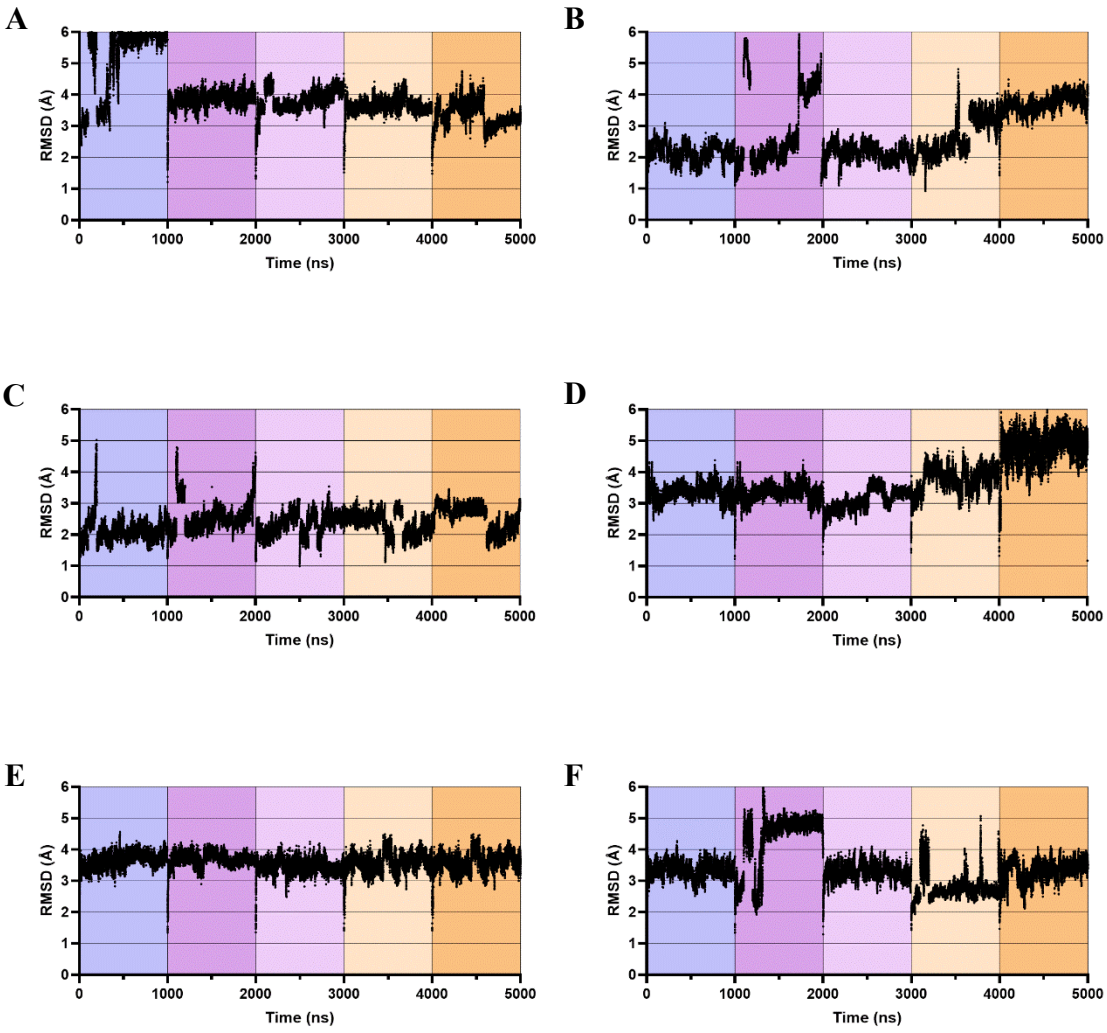


Figure A.2. All-heavy-atom RMSD relative to the first frame over the MD simulations for the 109 mutants A) H109A, B) H109S, C) H109E, D) H109Q, E) H109K, and F) H109D. The five 1000 ns replicates composing each simulation are joined back-to-back.

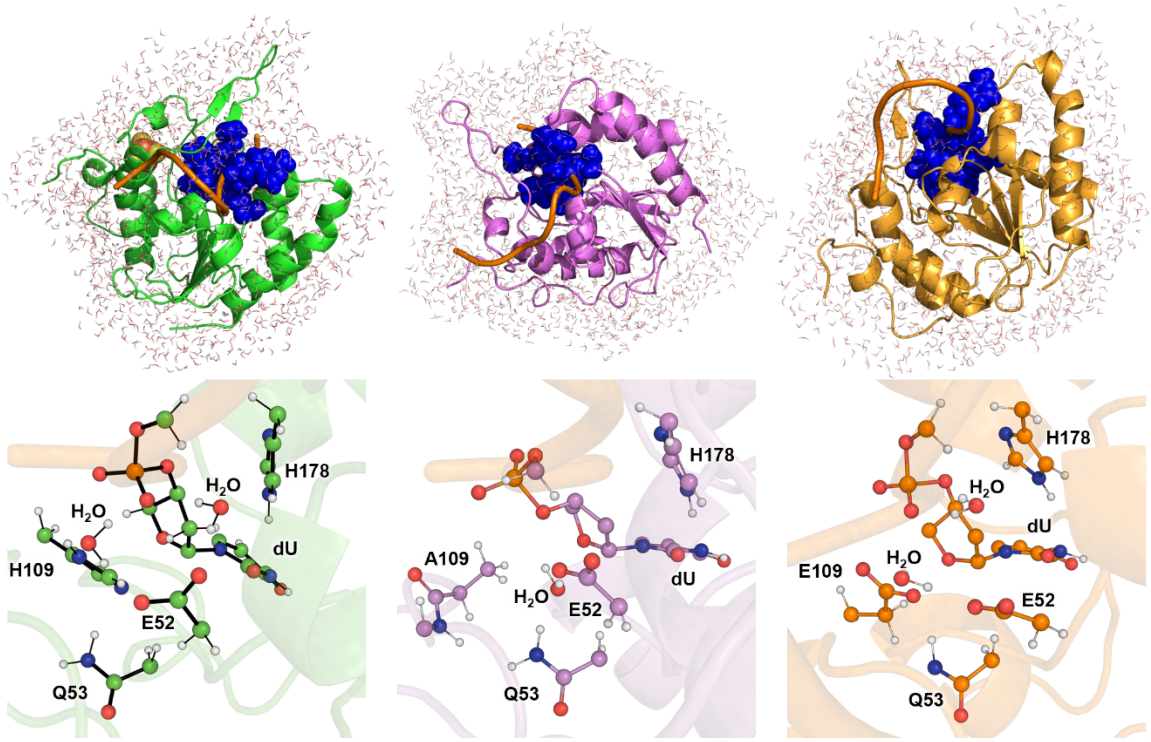


Figure A.3. QM/MM wild-type (green, top left), H109A (pink, top middle), and H109E (orange, top right) models used to calculate the UdgX catalytic pathway with QM region highlighted in blue. QM region atoms for wild-type (bottom left), H109A (bottom middle), and H109E (bottom right) models relative energies obtained using ONIOM(M06-2X/6-311+G(2df,p):AMBER14SB).

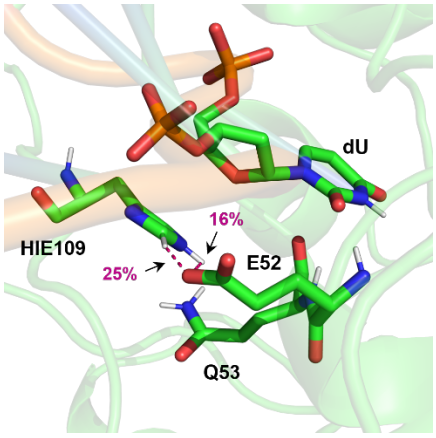


Figure A.4. UdgX HIE109 tautomer displaying occasional interaction between H109 protonated amide (Nε–H) or imidazole C–H and OE2 of E52, highlighting reduced occupancies.

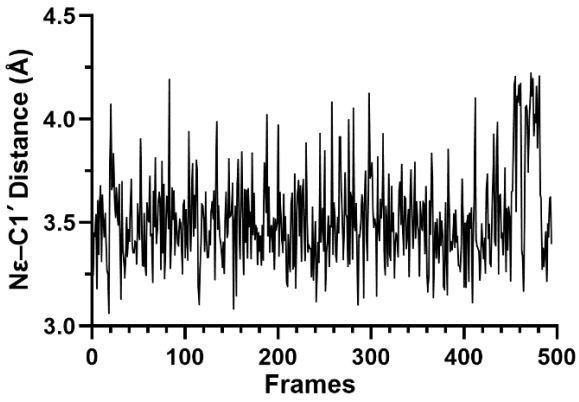


Figure A.5. The HID system displays transient Nε–C1′ distances below 4 Å, which are necessary for crosslink alignment, reported as function of distance over simulation frames (every 100 frames for Cluster 1, 49% of the 1 ns simulation).

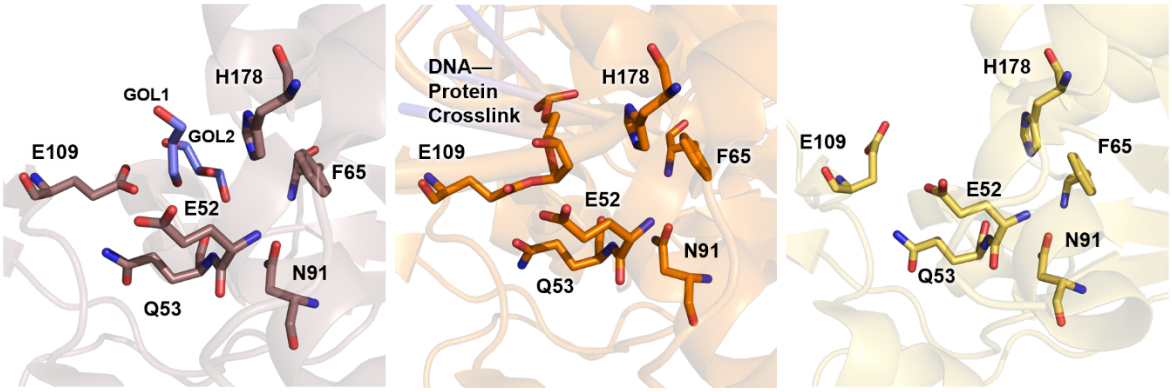


Figure A.6. X-ray crystal structures of H109E in the pre-excision state unbound to DNA with 2 glycerol molecules that were occupying the uracil-binding pocket (brown, left), the DNA–protein crosslink state (orange, middle), and the post-excision state unbound to DNA (yellow, right) (PDB ID: 6L6S and 6L5A/B).

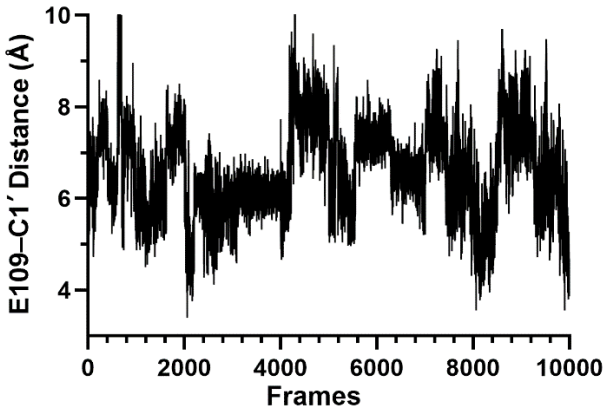


Figure A.7. MD simulations of the H109E system displays the E109–C1′ distance exceeded 4 Å, creating a solvent accessible cavity for the hydrolysis mechanism. Reported as a function of distance over simulation frames (every 10 frames for all 5 replicates).

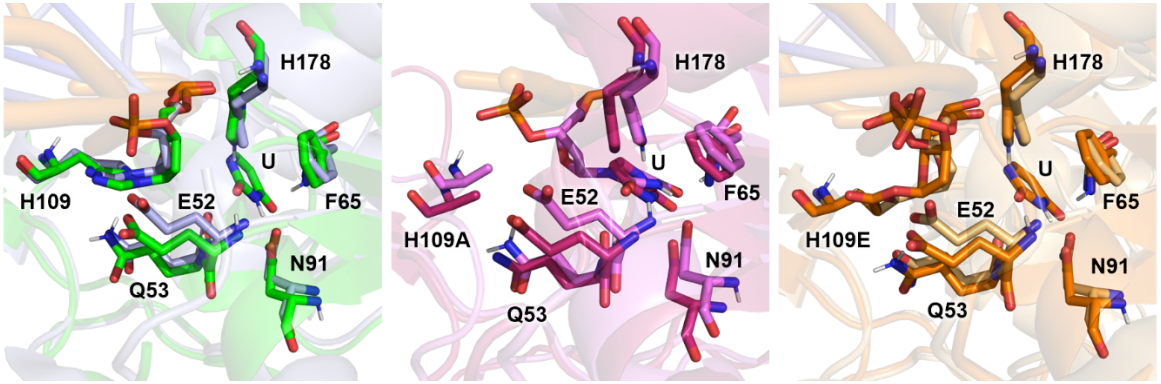


Figure A.8. QM/MM models overlayed onto their respective crystal structures for the wild-type model (green) and PDB ID: 6IOD (pale blue) (left), the H109A mutant (pink) and PDB ID: 8IIG (dark pink) (middle), and the H109E mutant (orange) and PDB ID: 6L6S (pale orange) (right).

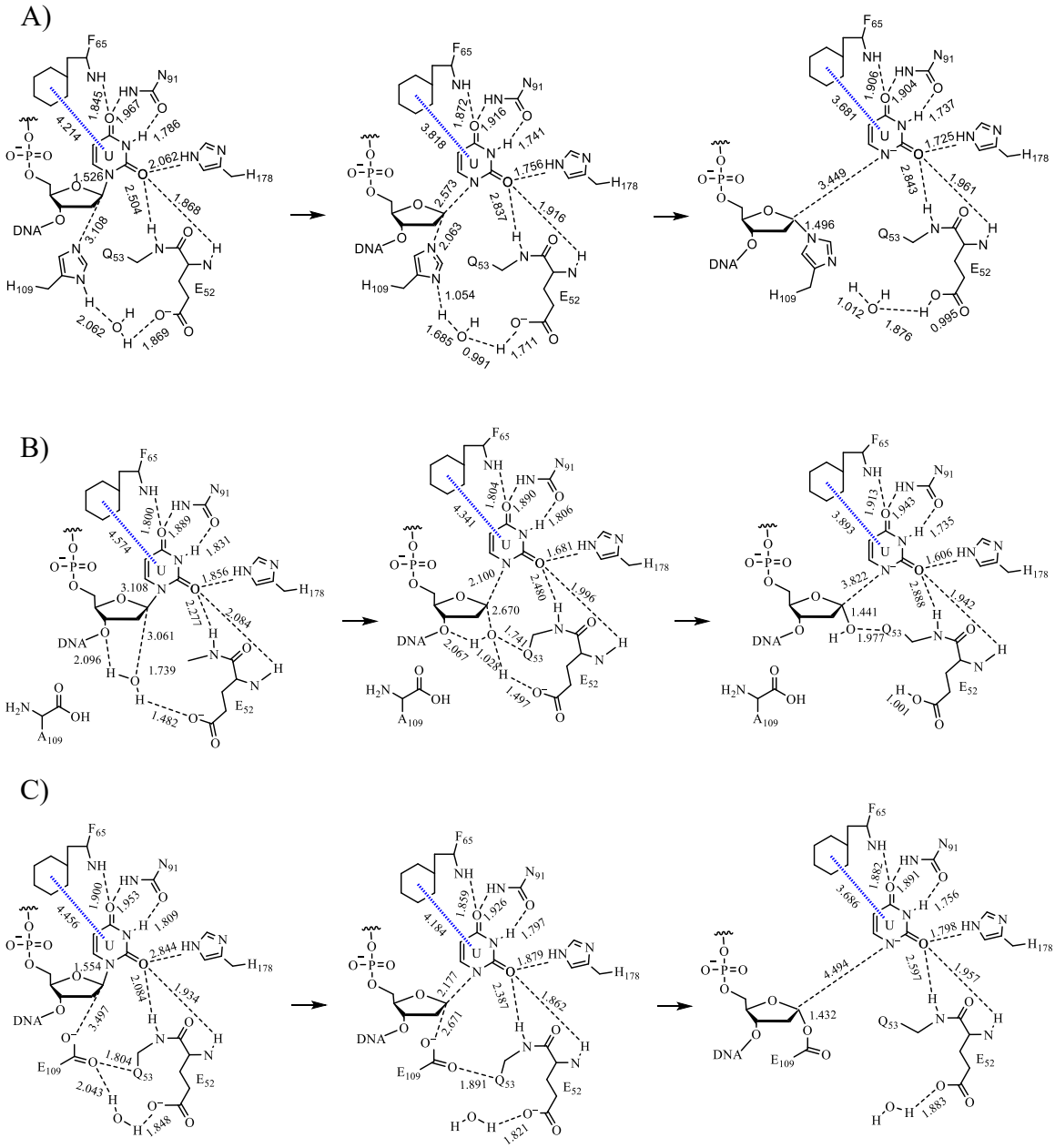


Figure A.9. QM/MM mechanism of action of the UdgX A) wild-type crosslink formation, B) H109A hydrolysis reaction, and C) H09E crosslink formation, highlighting important distances and hydrogen-bond interaction distances. The distance is measured in Å and π-stacking interactions are highlighted in blue.

Table A.1. The number of explicit water molecules and counterions (Na^+ and Cl^-) included in each MD simulation system reported after solvation of the enzyme–DNA complexes in a TIP3P rectangular box and additional NaCl was added to approximate physiological salt conditions after neutralization of the system charge.

Protein–Complex	109 Modification	# of Water Molecules	# of Na^+ Ions	# of Cl^- Ions
UdgX Wild-type	HID	11315	35	22
	HIE	11315	35	22
UdgX Mutants	A	11245	35	22
	E	11946	37	23
	S	11946	37	24
	Q	11245	35	22
	K	11245	34	22
	D	11245	35	21

Table A.2. Atom types defined for the Fe_4S_4 cluster of UdgX generated using MCPB.py.

Atom Type	Element / Chemical Environment	Mass (amu)	Source
M1–M4	Fe^{2+} ions in Fe_4S_4 cluster	55.85	IOD model (Li et al., JCTC 2013)
Y1–Y4, Y6–Y8	Thiolate sulfur (cystine coordination)	32.06	GAFF2
Y5	sp^2 nitrogen in imidazole-like ring of Fe4S4-coordinated histidine	14.01	OPLS-derived

Table A.3. Metal-ligand bond parameters and representative angle parameters involving the Fe_4S_4 cluster.

Bond	Force Constant (kcal mol^{-1} Å^{-2})	Equilibrium Distance (Å)
Y1–M1	98.6	2.276
Y2–M1	43.4	2.404
Y2–M3	43.4	2.404
Y2–M4	43.4	2.404
Y3–M1	39.1	2.435
Y3–M2	52.6	2.381
Y3–M4	39.1	2.435
Y4–M1	39.1	2.435
Y4–M2	52.6	2.381
Y4–M3	39.1	2.435
Y5–M2	59.4	2.095
Y6–M2	52.6	2.381
Y6–M3	39.1	2.435
Y6–M4	39.1	2.435
Y7–M3	98.6	2.276
Y8–M4	98.6	2.276
Angle	**Force Constant (kcal mol^{-1} rad^{-2})**	**Equilibrium Angle (°)**
Y1–M1–Y2	34.6	123.93
Y1–M1–Y3	27.69	117.78
Y2–M1–Y3	53.47	95.62
Y3–M2–Y4	25.67	103.37
Y5–M2–Y6	34.6	114.59
Y6–M3–Y4	47.34	100.69
Y7–M3–Y2	34.6	123.93
Y8–M4–Y2	34.6	123.93

Table A.4. Proper dihedral and improper torsion parameters involving the metal coordinated Fe_4S_4.

Dihedral	**Barrier Height (kcal mol^{-1})**	**Phase (°)**	**Periodicity**
X–CC–Y5–X	4.8	180	2
X–Y5–CR–X	10	180	2
Improper	**Force Constant**	**Phase (°)**	**Periodicity**
CT–CW–CC–Y5	1.1	180	2

Table A.5. Non-bonded parameters involving the metal coordinated Fe_4S_4.

Non-Bonded	**R_{min}/2 (Å)**	**ε (kcal mol^{-1})**
M1–M4	1.409	0.0172
Y1–Y4, Y6–Y8	1.983	0.282
Y5	1.824	0.17

Table A.6. Distances between E52 sidechain oxygens (OE1 and OE2) and C1′ of uracil was measured across all mutants. Increased separation is interpreted as reflecting enlargement of the active site cavity and enhanced capacity to accommodate a nucleophilic water molecule.

E52–C1′	Distance (Å) ± (δ^1)
A	5.52 ± 0.7
E	5.30 ± 0.9
S	5.40 ± 0.6
Q	5.11 ± 0.7
K	5.07 ± 0.6
D	5.17 ± 0.7

[1]Standard deviation of E52–C1′ distance for each mutant model.

Full MD Model Details of the H109E Mutant

The H109E mutant is mechanistically distinct from the other H109 variants, as experimental studies indicate that it is capable of both hydrolysis activity and covalent DNA–protein crosslink formation. Structural evidence for crosslinking competency is further supported by the crystal structure of the enzyme–DNA complex (PDB ID: 6L6S). In the present simulations, H109 mutations were introduced into the pre-minimized wild-type model. However, the H109E variant did not sample crosslink-competent

conformations during five independent 1000 ns MD trajectories. Instead, the glutamate side chain consistently reoriented away from the C1′ reaction center and toward the solvent, resulting in increased active site hydration and alignment more consistent with a hydrolytic mechanism. To explore whether this behavior was dependent on the initial side chain orientation, alternative H109E conformers were generated using the mutagenesis tool in PyMOL. In all cases, the glutamate side chain relaxed away from the anomeric carbon during MD equilibration, yielding configurations that favoured water accessibility. Additional starting models were constructed by manually positioning and structural alignment matching the crosslinked crystal structure (PDB ID: 6L6S), using both pre- and post-minimized wild-type structures as templates. None of these configurations maintained catalytically competent geometries for crosslink formation during MD simulations. These observations suggest that the crosslink-competent state of the H109E mutant may correspond to a transient conformational ensemble that is not sufficiently sampled on the microsecond timescale accessible to classical MD simulations. To directly evaluate the feasibility of the crosslinking reaction pathway, the experimentally resolved crosslinked structure was therefore used as the starting point for QM/MM calculations to characterize the associated reaction energetics.

Appendix B

Supplemental Information for Chapter 3: A Computational Investigation of the Substrate Specificity in Alkyladenine DNA Glycosylase and ALKBH2 Towards 3,N^4–Etheno-5-methylcytosine

Contains Figures B.1–B.3 and Tables B.1–B.5

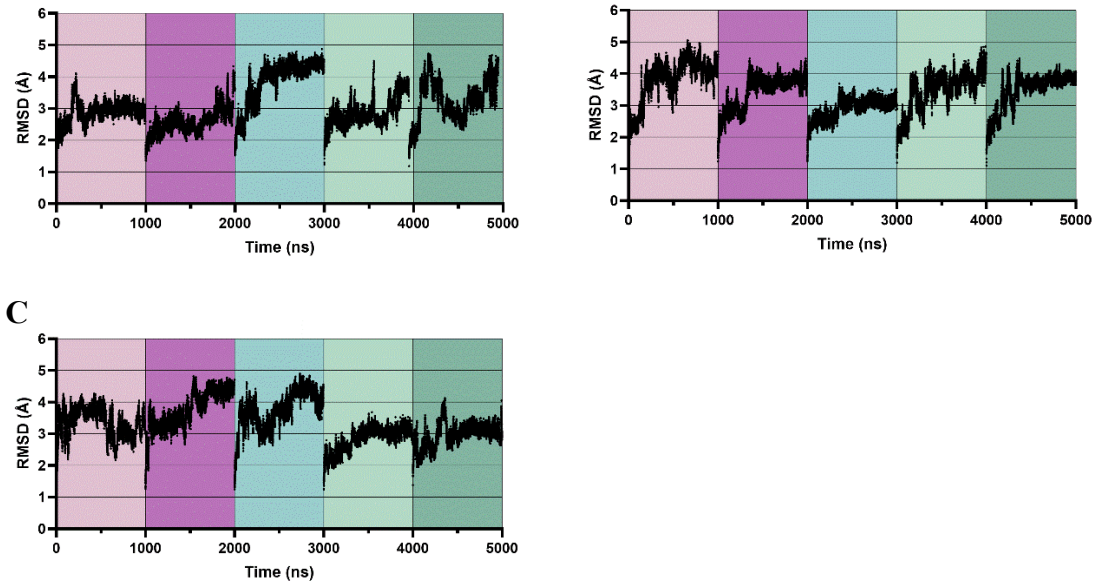


Figure B.1. All-heavy-atom RMSD relative to the first frame over the MD simulations for the AAG models bound to A) εA, B) εC, and C) ε5mC. The five 1000 ns replicates composing each simulation are joined back-to-back.

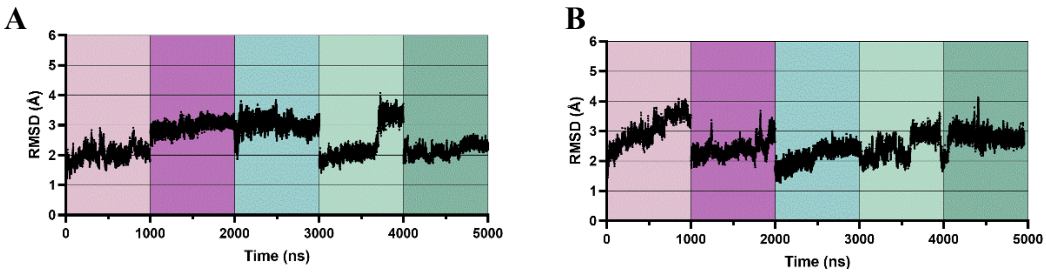


Figure B.2. All-heavy-atom RMSD relative to the first frame over the MD simulations for the ALKBH2 models bound to A) εC and B) ε5mC. The five 1000 ns replicates composing each simulation are joined back-to-back.

A) B) C)

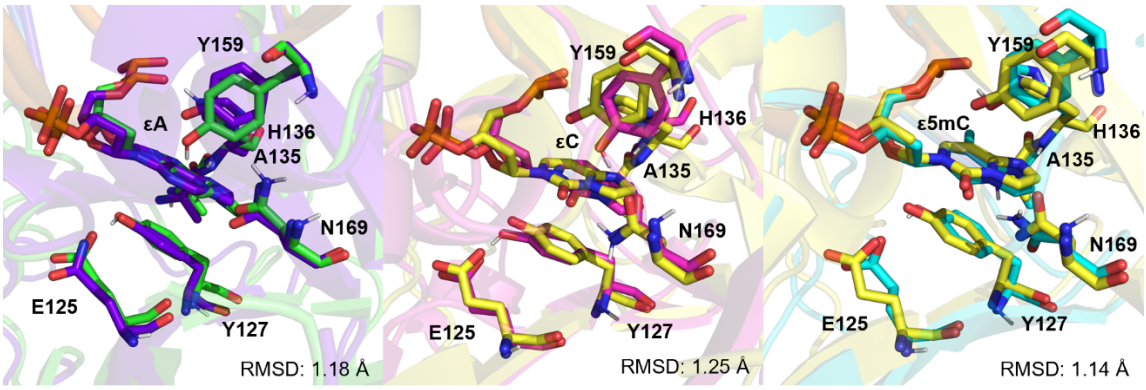


Figure B.3. MD representative structures of AAG systems overlayed with their corresponding crystal structure A) εA with 1EWN, and B) εC or C) ε5mC with 3QI5, highlighting active site residues and their corresponding RMSD value reported in Å.

Table B.1. The number of explicit water molecules and counterions (Na^+ and Cl^-) included in each simulation system are reported after solvation of the enzyme–DNA complexes and neutralization of the system charge.

Protein–Complex	**Nucleotide Bound**	**# of Water Molecules**	**# of Na^+ Ions**	**# of Cl^- Ions**
AAG	εA	12340	39	23
	εC	12339	39	23
	ε5mC	12341	39	23
ALKBH2	εC	9138	28	16
	ε5mC	9132	28	16

Table B.2. Atom types assigned for nonstandard etheno nucleotides.

Lesion	Atom Type	Chemical Environment
εA	N*	sp2 ring nitrogen
εA	CN / CK / CQ	aromatic bridge carbons
εA	NB / NC	aromatic nitrogen
εC	NB	aromatic nitrogen in fused ring
εC	CM / CB / CV	aromatic carbons in fused ring
ε5mC	N*	aromatic nitrogen
ε5mC	CA / CB	aromatic carbons
ε5mC	CK	etheno bridge carbon

Table B.3. Additional bond, angle, dihedral, improper torsion angle and nonbonded parameters.

Lesion	Bond	k (kcal mol^{-1} Å^{-2})	Req (Å)	Analogy
εA	CN–N*	436	1.374	CB–N*
εA	CN–CN	518	1.371	CC–CW
εC	C–NB	374.6	1.42	c–n2
εC	NB–CB	441.1	1.369	cc–nc
ε5mC	CA–N*	436	1.374	CB–N*
ε5mC	CK–CK	549	1.35	CM–CM
Lesion	**Angle**	**k (kcal mol^{-1} rad^{-2})**	**θeq (°)**	
εA	CN–CN–N*	70.9	107.9	
εA	N*–CN–NC	71.2	129.1	
εC	O–C–NB	72.9	122.5	
εC	C–NB–CB	66.3	121	

ε5mC	CA–N*–CT	70	121.2	
ε5mC	N*–CB–NB	70	126.2	
Lesion	**Dihedral**	**Barrier (kcal mol⁻¹)**	**Phase (°)**	**Periodicity**
εA	N*–CN–NC–CQ	9.6	180	2
εA	CN–CN–N*–CK	6.6	180	2
εC	N*–C–NB–CB	4.15	180	2
εC	NB–CB–CM–HA	6.65	180	2
ε5mC	CA–N*–CK–CK	6.6	180	2
ε5mC	CK–CK–NB–X	26.6	180	2
Lesion	**Improper**	**k**	**Phase**	**Periodicity**
εA	CK–CN–N*–CT	1.1	180	2
εC	N*–O–C–NB	1.1	180	2
ε5mC	C–CA–N*–CT	1.1	180	2
Atom Type	**$R_{min}/2$ (Å)**	**ε (kcal mol⁻¹)**	**Source**	
Aromatic C (CN/CA/CB/CK)	1.908	0.086	GAFF	
Aromatic N (N*/NB/NC)	1.824	0.17	GAFF	
Carbonyl O	1.661	0.21	AMBER	
Aliphatic C	1.908	0.109	AMBER	
H (aromatic/aliphatic)	1.29–1.49	0.015	AMBER	

Table B.4. Atom types assigned to the iron metal center coordination.

Atom Type	Element	Chemical Environment	Source
M1	Fe^{3+}	Catalytic iron ion	IOD metal model
Y1	N	Histidine-like nitrogen donor	nc / n2
Y2	O	Carboxylate oxygen donor	o
Y3	N	Histidine-like nitrogen donor	nc / n2
Y4	O	Ether/neutral oxygen donor	OPLS ether
Y5	O	Carbonyl oxygen donor	OPLS carbonyl
Y6	O	Coordinating water oxygen	TIP3P

Table B.5. All metal bond, angle, dihedral, improper torsion angle and nonbonded parameters.

Bond	Force Constant (kcal mol^{-1} $Å^{-2}$)	Req (Å)	
Fe–O(ether) (M1–Y4)	397.9	1.643	
Fe–O(carbonyl) (M1–Y5)	62.1	1.901	
Fe–Ow (M1–Y6)	45.9	2.055	
Fe–N(His) (M1–Y1)	72.5	1.991	
Fe–O(carboxylate) (M1–Y2)	79.4	1.982	
Fe–N(His) (M1–Y3)	22.2	2.215	
Angle	**Force Constant (kcal mol^{-1} rad^{-2})**	**θeq (°)**	
CO–Y2–Fe	130.96	124.71	
CR–Y1–Fe	123.78	131.03	

CR–Y3–Fe	86.89	123.37	
Fe–Y1–CV	125.31	121.63	
Fe–Y5–C	84.62	138.78	
Fe–Y6–Hw	34.23	96.72	
Y1–Fe–Y2	135.61	86.46	
Y1–Fe–Y3	119.84	93.65	
Y2–Fe–Y5	63.16	166.21	
Y3–Fe–Y6	92.26	88.39	
Dihedral	**Barrier (kcal mol^{-1})**	**Phase (°)**	**Periodicity**
X–CR–Y1–X	10	180	2
X–CR–Y3–X	10	180	2
X–CV–Y1–X	4.8	180	2
X–CV–Y3–X	4.8	180	2
Improper	**Force Constant**	**Phase (°)**	**Periodicity**
X–X–C–Y5	10.5	180	2
X–Y5–C–O	1.1	180	2
X–O2–CO–Y2	10.5	180	2
Atom Type	**$R_{min}/2$ (Å)**	**ε (kcal mol^{-1})**	**Source**
Fe^{3+}	1.386	0.0136	IOD model (Li et al.)
N donor (Y1, Y3)	1.824	0.17	AMBER
O donor (Y2, Y5)	1.661	0.21	AMBER
Ether O (Y4)	1.684	0.17	OPLS
Water O (Y6)	1.768	0.152	TIP3P