COMPUTATIONAL INSIGHT INTO THE BROAD SUBSTRATE SPECIFICITY OF ENZYMES THAT PROCESS NUCLEIC ACIDS

STEFAN LENZ
Bachelor of Science, University of Lethbridge, 2013

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

DOCTOR OF PHILOSOPHY

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

COMPUTATIONAL INSIGHT INTO THE BROAD SUBSTRATE SPECIFICITY OF ENZYMES THAT PROCESS NUCLEIC ACIDS

STEFAN LENZ

Date of Defence: November 8, 2018

Dr. S. Wetmore Supervisor	Professor	Ph.D.
Dr. M. Gerken Thesis Examination Committee Member	Professor	Ph.D.
Dr. M. Roussel Thesis Examination Committee Member	Professor	Ph.D.
Dr. A. Zovoilis Thesis Examination Committee Member	Assistant Professor	Ph.D.
Dr. C. Rowley External Examiner Memorial University of Newfoundland St. John's, Newfoundland and Labrador	Associate Professor	Ph.D.
Dr. N. Thakor Chair, Thesis Examination Committee	Assistant Professor	Ph.D.

Abstract

Many enzymes that bind DNA and RNA possess broad substrate specificity and play diverse roles in biology. Three classes of enzymes with broad substrate specificity are nucleoside hydrolases that salvage nucleic acid building blocks, and alkyladenine DNA glycosylases (AAG) and AlkB enzymes, which repair alkylated and/or deaminated DNA damage. This thesis uses advanced computational techniques to examine how enzymes process structurally diverse substrates. Specifically, structural and energetic information is provided by molecular dynamics (MD) simulations, quantum mechanics (QM) and hybrid quantum mechanics/molecular mechanics (QM/MM) calculations, which provide insight into how each enzyme active site changes to accommodate unique substrates and quantify the impact that these changes have on catalyzed reactions. From these results, atomistic explanations for the activity of these enzymes is obtained, which can be used to develop new treatments for diseases. The computational approach presented can be applied to other enzymes that exhibit broad substrate specificity.

Acknowledgements

I am deeply grateful to my supervisor, Dr. Stacey Wetmore, for her endless support during my time in the lab. I truly cannot express my gratitude enough for the countless opportunities that you have provided for me since I first started in your lab as a second year undergraduate. My sincere thanks to my supervisory committee, Dr. Michael Gerken and Dr. Marc Roussel. I appreciate the helpful comments and encouragement that you have provided throughout my degree. Thank you to Dr. Athan Zovoilis for agreeing to be my external internal examiner and Dr. Nehal Thakor for chairing my defense. A special thank you to my external examiner, Dr. Christopher Rowley, for his valuable time and interest in this thesis.

My work has been supported by both financial and computational resources. Thank you to the Natural Sciences and Engineering Research Council (NSERC), Alberta Innovates – Technology Futures (AI-TF), and the University of Lethbridge for student scholarships. I would also like to thank Westgrid and Compute Canada for computer resources.

I am humbled to have worked alongside remarkable people who have made significant contributions to this thesis and my professional career. Thank you to past and present members of the Wetmore lab: Dr. Emmanuel Naziga, Dr. Jennifer Kellie, Dr. Mohamed Aboelnga, Dr. Preethi Seelam, Dr. Purshotam Sharma, Dr. Shahin Sowlati Hashjin, Briana Boychuk, Cynthia Fonderson, Hanzala Hussain, Johnathan Kohout, Katie Wilson, Lindey Felske, Mohadeseh Majdi Yazdi, Priya Bhutani, Rachael Wells, Rajwinder Kaur, Rebecca Jeong, and Ryan Kung. I would also like to thank Dr. Deyu Li, the Li lab (Ke Bian, Qi Tang, and Rui Qi), Dr. Bongsup Cho, Dr. Ang Cai, and Rachel Carley for welcoming me to the University of Rhode Island. I learned a great deal and had an amazing experience.

Thank you to my friends and family who have provided me with continued support throughout my career. The morning coffees, weekend dinners, and Sunday football were critical for my sanity. Lastly, thank you to Ashley. This thesis and my degree would not have been possible without your love, patience, and understanding. I cannot wait for the next adventure with you.

Table of Contents

Thesis Examination Committee Members ii
Abstract iii
Acknowledgements iv
Table of Contents vi
List of Tables ix
List of Figures x
List of Abbreviations xiii

Chapter 1: Introduction 1
1.1 General Overview 1
1.2 Nucleic Acids 2
1.3 Biosynthesis of RNA and DNA Nucleotides 4
1.4 Purine (Pyrimidine) Salvage Pathway 4
1.5 Nucleic Acid Modifications 6
1.5.1 DNA Damage 7
1.5.1.1 Alkylation 8
1.5.1.2 Deamination 10
1.5.2 Epigenetic Modifications 11
1.6 Essential Roles Played by DNA Repair Enzymes with Broad Substrate Specificity 12
1.6.1 Base Excision Repair 12
1.6.2 Direct Reversal Repair 14
1.7 Computational Approaches Used to Study Enzyme–Substrate Complexes and Catalysis 15
1.8 Thesis Overview 16
1.9 References 18
Chapter 2: Evaluating the Substrate Selectivity of Alkyladenine DNA Glycosylase: The Synergistic Interplay of Active Site Flexibility and Water Reorganization 27
2.1 Introduction 27
2.2 Computational Methods 34
2.3 Results 36
2.3.1 AAG adopts a consistent active site conformation upon binding neutral lesions 36
2.3.2 Reorganization of the AAG active site, including increased solvation, occurs upon binding of adenine and likely prevents excision 39
2.3.3 Guanine binding to AAG results in anti-catalytic active site interactions 42
2.3.4 Despite structural similarity to the canonical purines, methylated lesions are better accommodated in the AAG active site 44
2.4 Discussion 45
2.4.1 Despite similar active site configurations, subtle disparities in DNA–protein interactions lead to different catalytic efficiencies toward the excision of neutral substrates 47
2.4.2 In addition to a key active site hydrogen bond, redistribution of water upon εC binding leads to inhibition of AAG 48
2.4.3 Interactions with the nucleobase amino groups in conjunction with redistribution of key active site residues, including water, allows AAG to discriminate against natural DNA 49
2.4.4 Despite equivalent amino groups as the canonical purines, the methylated lesions are positioned for excision by AAG using unique DNA–protein interactions 51
2.5 Conclusion 52

2.6 References 54
Chapter 3: An QM/MM Study of the Reaction Catalyzed by Alkyladenine DNA Glycosylase: Examination of the Substrate Specificity of a DNA Repair Enzyme 61
3.1 Introduction 61
3.2 Computational Methods 67
3.2.1 MD Simulation Protocol 67
3.2.2 ONIOM(QM:MM) Computational Scheme 68
3.2.3 Reaction Surface Methodology 69
3.2.4 NCI Analysis 71
3.3 Results and Discussion 71
3.3.1 Hypoxanthine Substrate 71
3.3.2 Guanine Nucleotide 78
3.3.3 7MeG Substrate 83
3.4 Conclusion 87
3.5 References 89
Chapter 4: Hydrolytic Glycosidic Bond Cleavage in RNA Nucleosides: Effects of the 2′–Hydroxy Group and Acid–Base Catalysis 96
4.1 Introduction 96
4.2 Computational Methods 100
4.3 Results and Discussion 103
4.3.1 Hydrolysis Facilitated by the $HCOO^{-}\bullet\bullet\bullet H_2O$ Nucleophile 103
4.3.2 Factors that May Facilitate Hydrolysis 108
4.4 Conclusions 120
4.5 References 122
Chapter 5: Structural Explanation for the Tunable Substrate Specificity of Nucleoside Hydrolases: Insights from Molecular Dynamics Simulations 129
5.1 Introduction 129
5.2 Computational Methods 135
5.3 Results and Discussion 138
5.3.1 Despite several catalytically conducive interactions in the active site, lack of nucleobase activation explains the inability of wild-type CU-NH to hydrolyze inosine. 138
5.3.2 The active site conformation of the Thr223Tyr Gln227Tyr CU-NH double mutant does not correlate with the experimentally-observed enhanced activity towards inosine when His239 is modelled as the general acid. 142
5.3.3 An Asp15-Tyr223-I hydrogen-bonding network rationalizes the experimentally-observed activity of the Thr223Tyr Gln227Tyr CU-NH double mutant toward inosine. 145
5.3.4 Interactions between Tyr223 and Tyr227 are required for efficient CU-NH mediated hydrolysis of inosine. 149
5.4 Conclusions 153
5.5 References 154
Chapter 6: An MD and QM/MM Study on the Substrate Scope of AlkB and ALKBH2: Insights into the Oxidative Repair of Etheno Lesions 159
6.1 Introduction 159
6.2 Computational Methods 165
6.3 Results and Discussion 169
6.3.1 Average Dynamic Distance Between the Bound Lesion and Fe(IV)-oxo Group Correlates with Relative Repair Rates for AlkB and ALKBH2. 169

6.3.2 π-interactions Stabilize Etheno Adducts in the Active Sites of AlkB and ALKBH2 172
6.3.3 Catalytic Residues and Water are Aligned for AlkB- or ALKBH2-mediated oxidation of 1,N^6-εA and 3,N^4-εC, but Disrupted for 1,N^2-εG and N^2,3-εG. 175
6.3.4 The Lesion-Dependent Active Site Conformation Enhances AlkB-catalyzed Oxidation of 1,N6-εA and 3,N4-εC, but Significantly Impedes Oxidation of 1,N2-εG and N2,3-εG 178
6.4 Conclusions 182
6.5 References 184
Chapter 7: DNA Repair Enzymes ALKBH2, ALKBH3, and AlkB Oxidize 5-Methylcytosine to 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxylcytosine 190
7.1 Introduction 190
7.2 Experimental and Computational Methods 192
7.3 Results and Discussion 193
7.4 Conclusions 199
7.5 References 200
Chapter 8: Conclusions and Future Directions 203
8.1 Summary 203
8.2 Nucleoside Hydrolases 203
8.2.1 Contributions from Thesis 203
8.2.2 Future Directions 205
8.3 Repair of DNA Damage 205
8.3.1 Contributions from Thesis 205
8.3.2 Future Directions 208
8.4 Final Remarks 211
8.5 References 212
Appendix A: Supplemental Information for Chapter 2 216
Appendix B: Supplemental Information for Chapter 3 229
Appendix C: Supplemental Information for Chapter 4 242
Appendix D: Supplemental Information for Chapter 5 260
Appendix E: Supplemental Information for Chapter 6 286
Appendix F: Supplemental Information for Chapter 7 308

List of Tables

Table 3.1. Relative Potential and Gibbs Energies for the Excision of Hx by AAG 74
Table 3.2. Relative (ΔE) and Gibbs (ΔG) Energies for the Excision of 7MeG by AAG 84
Table 4.1. Comparison of the Calculated Gibbs Barriers and Reaction Energies for the Deglycosylation of Ribose and Deoxyribose Nucleosides Facilitated by the $HCOO^-\bullet\bullet\bullet H_2O$ Nucleophile 106
Table 4.2. Calculated Gibbs Barriers and Reaction Energies for the Deglycosylation of Natural and Corresponding Protonated (Acid–catalyzed) RNA Nucleosides Facilitated by Various (H_2O, OH^-, or $HCOO^-\bullet\bullet\bullet H_2O$) Nucleophiles 110
Table 5.1. Average root-mean-square deviation of the entire CU-NH active site or key active site residues across the MD simulation trajectories 141

List of Figures

Figure 1.1. Structure and numbering of the canonical DNA and RNA nucleobases, and the ribose and 2′-deoxyribose sugars....2
Figure 1.2. Structure of the DNA a) base pairs, b) backbone, and c) double helix....3
Figure 1.3. DNA nucleotides in the anti and syn glycosidic bond conformation. The glycosidic bond conformation is defined by dihedral angle χ with either the Watson-Crick or Hoogsteen hydrogen-bonding face highlighted in red....3
Figure 1.4. Structure of uridine monophosphate and inosine monophosphate formed through nucleotide synthesis pathways....4
Figure 1.5. Simplified schematic of the purine (pyrimidine) salvage pathway (PSP)....5
Figure 1.6. Examples of adducted, deaminated, alkylated, and oxidized DNA lesions of guanine....6
Figure 1.7. Structure of common methylated DNA lesions of A, T, G, and C....9
Figure 1.8. Structure of common etheno DNA lesions....10
Figure 1.9. Structure of deaminated DNA lesions....11
Figure 1.10. Structure of common epigenetic modifications present in DNA....12
Figure 1.11. Schematic representation of the BER pathway....13
Figure 1.12. AlkB-catalyzed repair of 1,N6-εA-containing DNA with the aberrant atoms highlighted....15
Figure 2.1. Structures and chemical numbering of the canonical purines, neutral purine lesions, cationic purine lesions, and 2′-deoxyribose moiety. Nucleobase modifications are highlighted in red....28
Figure 2.2. X-ray crystal structure of AAG bound to εA. Overlay of the AAG active site bound to εA or εC. Key hydrogen bonds with the nucleobase are highlighted with dashed lines....31
Figure 2.3. Overlay of representative MD structures of AAG bound to εA, εC, or Hx highlighting the consistent active site conformation for the neutral lesions, and the common hydrogen bond between the H136 backbone and the bound nucleobase....38
Figure 2.4. Overlay of representative MD structures of AAG bound to Hx or A highlighting the similar position of the nucleotides with respect to several key active site residues, and the discrepancy in the H136 and Y159 orientations. Distribution of the active site water during the MD simulation of AAG bound to A. 40
Figure 2.5. Overlay of representative MD structures of AAG bound to Hx or G highlighting the similarity of the AAG active site conformation for several key active site residues, and the different Y127 and E125 orientations. Distribution of the active site water during the MD simulation of AAG bound to A....43
Figure 2.6. Overlays of representative MD structures of AAG bound to 3MeA or A, or 7MeG and G. 46
Figure 3.1. Structure and chemical numbering of select DNA lesions processed by AAG, as well as the canonical purines (A and G)....62
Figure 3.2. Line diagram of εA bound in the AAG active site adapted from an X-ray crystal structure....64
Figure 3.3. ONIOM scheme used to examine AAG-catalyzed hydrolysis....69
Figure 3.4. ONIOM reaction surfaces for the AAG-mediated excision of Hx, G, and 7MeG....72
Figure 3.5. Structures of the unconstrained ONIOM stationary points corresponding to the AAG-mediated excision of Hx displayed as a line diagram, or NCI plots with promolecular densities....74
Figure 3.6. Structure of the unconstrained ONIOM RC corresponding to the AAG-mediated excision of G displayed as a line diagram....79
Figure 3.7. Overlay of the unconstrained ONIOM G and Hx RC....81
Figure 3.8. NCI plots of the unconstrained ONIOM G RC and representative structure with an elongated glycosidic bond obtained from the G reaction surface, with promolecular densities....82
Figure 3.9. Structures of constrained ONIOM stationary corresponding to AAG-mediated excision of 7MeG displayed as a line diagram....86

Figure 3.10. Overlay of the AAG active site for the RC obtained from constrained ONIOM optimization with 7MeG bound and constrained ONIOM optimizations with G bound. 87
Figure 4.1. Structure and chemical numbering of the canonical RNA and DNA nucleosides. 100
Figure 4.2. Models used in the present study to examine RNA nucleoside deglycosylation. 101
Figure 4.3. Selected B3LYP/6-31G(d) bond lengths and angles in RC, TS, and PC for the deglycosylation of adenine (A), cytosine (C), guanine (G), and uracil (U) containing RNA and DNA nucleosides facilitated by the $HCOO^{-}\bullet\bullet\bullet H_2O$ nucleophile. 104
Figure 4.4. Selected B3LYP/6-31G(d) bond lengths and angles in RC, TS, and PC for the deglycosylation of adenine (A), cytosine (C), guanine (G), and uracil (U) containing RNA nucleosides facilitated by the (unactivated) H_2O nucleophile. 111
Figure 4.5. Selected B3LYP/6-31G(d) bond lengths and angles in RC, TS, and PC for the deglycosylation of adenine (A), cytosine (C), guanine (G), and uracil (U) containing RNA nucleosides facilitated by the OH^{-} nucleophile. 113
Figure 4.6. Selected B3LYP/6-31G(d) bond lengths and angles in RC, TS, and PC for the lowest barrier pathways of the acid-catalyzed deglycosylation of adenine (A–N7H), cytosine (C–O2H(N1)), guanine (G–N7H), and uracil (U–O2H(N3)) containing RNA nucleosides facilitated by the unactivated H_2O nucleophile. 117
Figure 5.1. Structures and chemical numbering of common nucleoside hydrolase substrates, as well as the ribose sugar 131
Figure 5.2. X-ray crystal structure of CU-NH bound to inosine, and IU-NH bound to p-aminophenyliminoribitol highlighting the amino acids interacting with the canonical or modified ribose and or nucleobase 132
Figure 5.3. Representative MD structure of wild-type CU-NH ($Asp15^{-}$ $His239^{+}$) bound to inosine, highlighting active site residues that surround the ribose sugar and Ca^{2+} or the inosine nucleobase and the distribution of active site water. 139
Figure 5.4. Representative MD structure of Thr223Tyr Gln227Tyr CU-NH ($Asp15^{-}$ $His239^{+}$) bound to inosine, highlighting active site residues that surround the ribose sugar and Ca^{2+} or the inosine nucleobase and the distribution of active site water 144
Figure 5.5. Representative MD structure of Thr223Tyr Gln227Tyr CU-NH (Asp15 His239) bound to inosine, highlighting active site residues that surround the ribose sugar and Ca^{2+} or the inosine nucleobase and the distribution of active site water. Hydrogen-bond distance as a function of time, highlighting simultaneous I(N7)···Tyr227(OηH) and Tyr223(Oη)···Asp15(OδH) hydrogen bonds. 147
Figure 5.6. Representative MD structures highlighting the active site residues that surround the inosine nucleobase and the distribution of active site water for Thr223Tyr CU-NH (Asp15 His239), Gln227Tyr CU-NH (Asp15 His239), Thr223Tyr CU-NH ($Asp15^{-}$ $His239^{+}$), or Gln227Tyr CU-NH ($Asp15^{-}$ $His239^{+}$) bound to inosine 152
Figure 6.1. Structures and chemical numbering of four etheno adducts 160
Figure 6.2. Proposed mechanism for the repair of $1,N^6$-εA catalyzed by an AlkB enzyme 161
Figure 6.3. X-ray crystal structures of the AlkB complex bound to $1,N^6$-εA and ALKBH2 complex bound to $1,N^6$-εA. 163
Figure 6.4. Representative MD structures of the AlkB or ALKBH2 complex bound to $1,N^6$-εA, $3,N^4$-εC, $1,N^2$-εG, or $N^2,3$-εG containing DNA, highlighting the average distances between the Fe(IV)-oxo group and the aberrant atoms of the nucleobase 170
Figure 6.5. Representative MD structures of the AlkB or ALKBH2 complex bound to $1,N^6$-εA, $3,N^4$-εC, $1,N^2$-εG, or $N^2,3$-εG containing DNA, highlighting the hydrogen-bonding interactions between the 5′-phosphate and active site residues, and the π-interactions between the nucleobase and aromatic residues 173

Figure 6.6. Overlays of representative MD structures of the AlkB or ALKBH2 complex bound to 1,N^6-εA and 3,N^4-εC, 1,N^2-εG, or N^2,3-εG containing DNA, highlighting the lesion-dependent positions of Asp135, Glu136, and Tyr78 for AlkB, and Glu175 and Tyr122 for ALKBH2 176
Figure 6.7. ONIOM(QM:MM) stationary points corresponding to the first step of AlkB-catalyzed oxidation of 1,N^6-εA, 3,N^4-εC, 1,N^2-εG, and N^2,3-εG 180
Figure 7.1. Reaction pathway of the AlkB family DNA repair enzymes modifying 5mC and 3mC. 191
Figure 7.2. High resolution ESI-TOF MS analyses of 16mer DNA oligonucleotides containing 5mC and oxidized products. The observed m/z values represent the oligonucleotides under their -3 charge state. 5mC; 5mC + ALKBH2; 5mC + ALKBH3; and 5mC + AlkB. 193
Figure 7.3. Product distribution and strand preference for the oxidation of 5mC by the AlkB family enzymes. Reaction products (5hmC, 5fC and 5caC) generated from the reactions of 5mC with the AlkB family enzymes in ds-DNA and ss-DNA. Total product percentage from reactions of the AlkB family enzymes oxidizing 5mC in ds- and ss-DNA. 196
Figure 7.4. Representative molecular dynamics structures of the ALKBH2 or AlkB complex bound to anti-5mC, syn-5mC, or anti-3mC. The distance between the oxo-moiety and methyl groups is highlighted with dashed lines 198

List of Abbreviations

1,N2-εG		1,N2-ethenoguanine
1,N6-εA	εA	1,N6-ethenoadenine
1MeA	1mA	1-methyladenine
3,N4-εC	εC	3,N4-ethenocytosine
3MeA		3-methyladenine
3MeC	3mC	3-methylcytosine
4MeC		4-methylcytosine
5caC		5-carboxylcytosine
5fC		5-formylcytosine
5hmC		5-hydroxymethylcytosine
5mC		5-methylcytosine
6MeA		6-methyladenine
7MeG		7-methylguanine
AAG		alkyladenine DNA glycosylase
AlkA		3-methyladenine DNA Glycosylase II
AP		apyrimidinic/apurinic site
au		atomic units
BER		base excision repair
CG		conjugate gradient
COM		center-of-mass
CTP		cytidine triphosphate
CU-NH		cytidine-uridine nucleoside hydrolase
dA		2′-deoxyadenosine
dC		2′-deoxycytidine
DFT		density functional theory
dG		2′-deoxyguanosine
dsDNA		double-stranded DNA
dT		2′-deoxythymidine
dU		2′-deoxyuridine
EA		1,N6-ethanoadenine
ESI		electrospray ionization
ESP		electrostatic potential
FPG		formamidopyrimidine DNA glycosylase
H		Hoogsteen
hOGG1		human 8-oxoguanine DNA glycosylase
hUNG2		human uracil DNA glycosylase
Hx		hypoxanthine
I		inosine
IMP		inosine monophosphate
IAG-NH		inosine-adenosine-guanosine nucleoside hydrolase
IEF-PCM		integral equation formalism-polarizable continuum model
IC		intermediate complex

IG-NH		inosine-guanosine nucleoside hydrolase
IU-NH		inosine-uridine nucleoside hydrolase
MBD4		methyl-CpG binding domain protein 4
MCPB		metal center parameter builder
MMGBSA		molecular mechanics Generalized Born surface area
MC		Monte Carlo
MD		molecular dynamics
MM		molecular mechanics
mRNA		messenger RNA
MS		mass spectrometry
MutY		adenine DNA glycosylase
N2,3-εG		N2,3-ethenoguanine
NCI		non-covalent interactions
ncRNA		non-coding RNA
NEIL1		endonuclease VIII-like 1
NH		nucleoside hydrolase
ONIOM		Our own N-layered Integrated Molecular Orbital and Molecular Mechanics
pAPIR		p-aminophenyliminoribitol
PC		product complex
PDB		Protein Data Bank
PES		potential energy surface
pol		polymerase
PSP		purine (pyrimidine) salvage pathway
QM		quantum mechanics
RC		reactant complex
RESP		restrained electrostatic potential
RihC		inosine-uridine nucleoside hydrolase
rms		root-mean-square
RMSD	rmsd	root mean square deviation
RNS		reactive nitrogen species
ROS		reactive oxygen species
SD		steepest descent
SMD		Solvation Model based on Density
ssDNA		single-stranded DNA
TDG		thymine DNA glycosylase
TLS		translesion synthesis
TOF		time-of-flight
TS		transition-state complex
UDG		uracil DNA glycosylase
UMP		uridine monophosphate
UTP		uridine triphosphate
WC		Watson-Crick
X		xanthine
α-KG		α-ketoglutarate

Chapter 1: Introduction[a]

1.1 General Overview

Proteins that bind to deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) perform essential roles within the cell. These include transcription factors that bind to the promoter region of DNA to regulate gene expression, DNA repair enzymes that revert harmful modifications, and salvage enzymes that ensure a steady supply of nucleotides for replication and cell division.[1] Often DNA and RNA binding proteins are defined by a high degree of specificity and efficiency. For example, uracil DNA glycosylase (UDG) solely catalyzes the repair of uracil lesions in DNA[2] and adenine DNA glycosylase (MutY) targets adenine paired opposite 8-oxoguanine lesions.[3] Nevertheless, it can be advantageous for DNA and RNA binding enzymes to possess broad substrate specificity (i.e., the ability to catalyze the same reaction on a diverse set of substrates). Three examples of enzymes with broad substrate specificity are nucleoside hydrolases that salvage nucleic acid building blocks,[4-6] and alkyladenine DNA glycosylases (AAG)[7-8] and AlkB enzymes,[9-11] which both repair alkylation damage. Unfortunately, the structural basis for the substrate specificity of these enzymes is not well understood. The purpose of this thesis is to use computational chemistry to provide insight into how AlkB enzymes, AAG, and nucleoside hydrolases recognize and process structurally diverse substrates. To provide the necessary background, the following sections will discuss the structure and function of nucleic acids, nucleotide biosynthesis, the salvage role played by nucleoside hydrolases, modifications to nucleic acids, and the roles played AAG and AlkB enzymes in DNA repair.

1.2 Nucleic Acids

DNA and RNA are biopolymers composed of nucleotide monomers, which consist of a nucleobase attached to a 5-carbon sugar (2′-deoxyribose in DNA and ribose in RNA) and a phosphate

[a]*The Journal of the American Chemical Society reference style is used in this Chapter.*

group (Figure 1.1). Adenine (A), thymine (T), guanine (G), and cytosine (C) are the canonical nucleobases in DNA, while T is substituted by uracil (U) in RNA. A DNA or RNA strand is composed of nucleotides connected via phosphodiester bonds. Two DNA strands align in an anti-parallel configuration and hydrogen bond through their nucleobases (i.e., base pair) to form a double helix structure (Figure 1.2). In contrast, RNA is generally single-stranded and uses intrastrand base pairing to adopt a multitude of structures. The canonical base pairs are G:C and A:T (A:U in RNA), although non-canonical base pairs (e.g., wobble base pairs) can also form, especially in RNA. In DNA, biological information is stored as a sequence of base pairs, which can be transcribed to generate non-coding RNA (ncRNA) or messenger RNA (mRNA) that is translated into proteins. The base-pairing nucleotides are generally in the *anti* glycosidic bond conformation (χ;$\angle$(O4′C1′N1C2) or $\angle$(O4′C1′N9C4) = 180° ± 90°) and hydrogen bond through the Watson-Crick (WC) face of the nucleobases. However, hydrogen-bonding interactions are also possible when the nucleotide is *syn* (χ = 0° ± 90°) through the Hoogsteen (H) face (Figure 1.3). For example, DNA and RNA G-quadruplexes are composed of four guanine bases that hydrogen bond through both the H and WC faces, and have important biological roles, including regulation of transcription.[12]

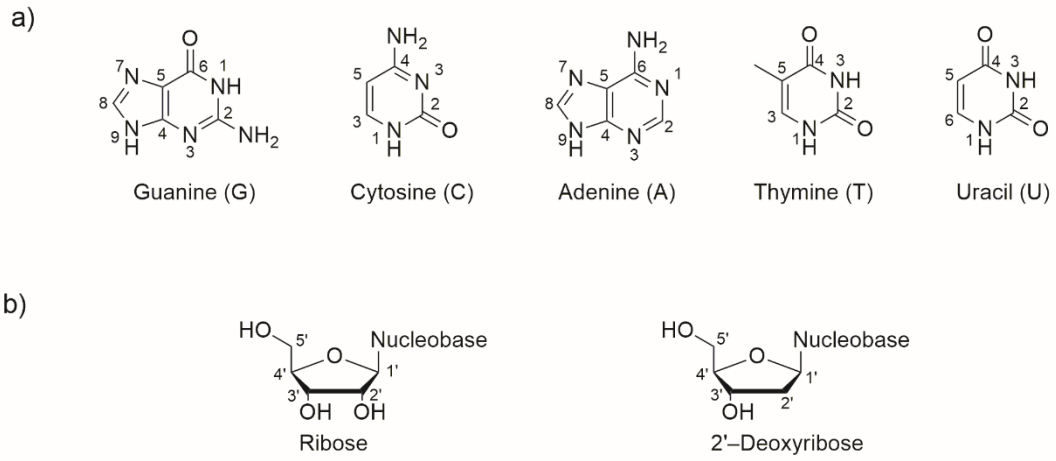


Figure 1.1. Structure and numbering of a) the canonical DNA and RNA nucleobases, and b) the ribose and 2′-deoxyribose sugars.

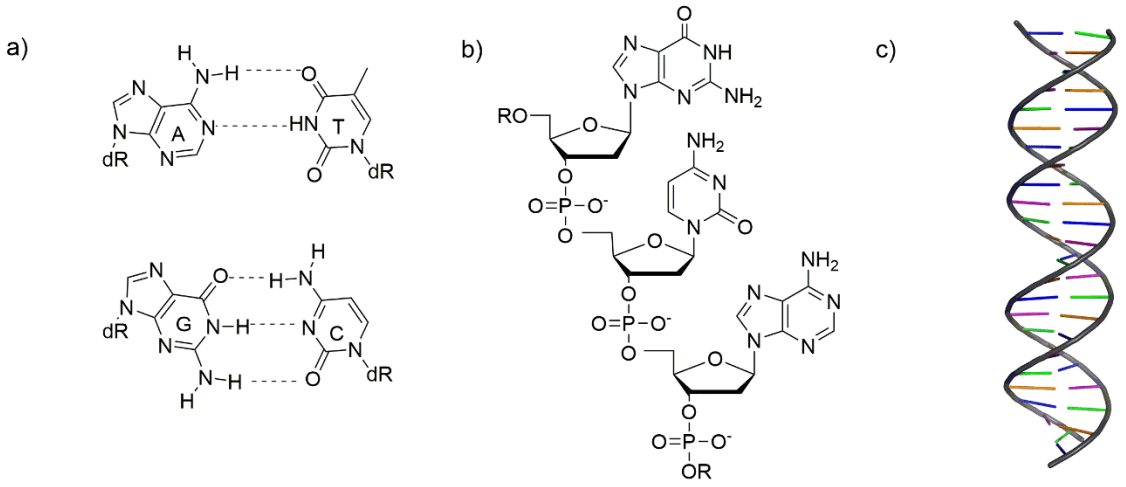


Figure 1.2. Structure of the DNA a) base pairs, b) backbone, and c) double helix.

In addition to being the monomeric subunit of nucleic acid polymers, canonical nucleotides are essential for energy storage (e.g., adenosine triphosphate (ATP)), biosynthetic pathways (e.g., synthesis of glycogen), and signal transduction pathways (e.g., signaling through cyclic adenosine monophosphate (cAMP) pathways).[1] It is essential that a steady supply of nucleotides is available for the cell and therefore the following section will discuss biosynthetic pathways of RNA and DNA nucleotides.

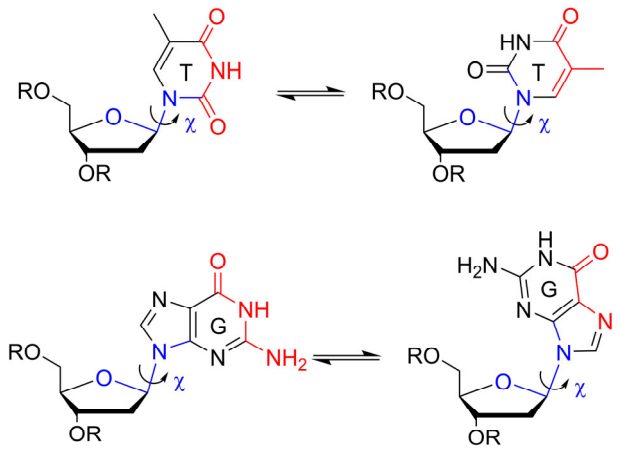


Figure 1.3. DNA nucleotides in the *anti* (left) and *syn* (right) glycosidic bond conformation (right). The glycosidic bond conformation is defined by dihedral angle χ (highlighted in blue) with either the Watson-Crick (*anti*) or Hoogsteen (*syn*) hydrogen-bonding face highlighted in red.

1.3 Biosynthesis of RNA and DNA Nucleotides

Nucleotide biosynthesis can be divided into two major pathways: *de novo* pathways and salvage pathways.[1] In *de novo* pyrimidine synthesis, the U nucleobase is first synthesized and attached to a phosphoribosyl group to form uridine monophosphate (UMP; Figure 1.4), and subsequently phosphorylated twice to generate uridine triphosphate (UTP). UTP can also be aminated to cytidine triphosphate (CTP). In contrast, for *de novo* purine synthesis, the purine group is formed as a ribonucleotide, instead of as a free nucleobase, and is ribophosphorylated through a series of reactions that yields inosine monophosphate (IMP; Figure 1.4). IMP can be converted to adenosine monophosphate or guanosine monophosphate. Salvage pathways can vary depending on the organism; however, they play essential roles for organisms that cannot perform *de novo* synthesis of purines, including protozoa.[13] Within salvage pathways, nucleoside hydrolases act on a diverse set of pyrimidine and purine nucleosides.

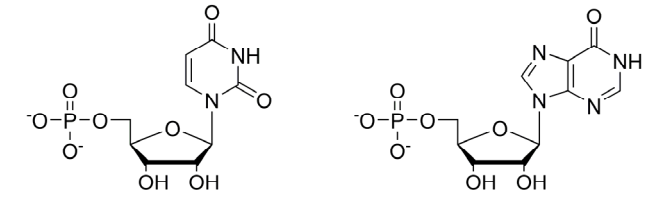


Figure 1.4. Structure of uridine monophosphate and inosine monophosphate formed through nucleotide synthesis pathways.

1.4 Purine (Pyrimidine) Salvage Pathway

Nucleoside hydrolases function as part of the purine (pyrimidine) salvage pathway (PSP; Figure 1.5), which is utilized by several organisms, including parasitic protozoa, to scavenge nucleosides from host organisms.[13-18] These parasites are responsible for numerous diseases, such as African trypanosomiasis (African sleeping sickness), American trypanosomiasis (Chagas disease), and leishmaniasis, which are collectively known as the trypanosomatid diseases.[19-21] Interestingly, parasitic protozoa rely on the PSP to

generate DNA or RNA components since they lack the biochemical infrastructure for *de novo* purine synthesis.[13, 21-22] This pathway differs significantly from the host salvage pathways as a result of phylogenetic separation. The first step of the PSP is transport of ribonucleosides across the cell membrane of the parasite, followed by cleavage of the *N*-glycosidic bond catalyzed by nucleoside hydrolases, and finally ribophosphorylation of the nucleobases to generate the respective 5′-monophosphate nucleoside (Figure 1.5).[13-18] Subsequently, the purine or pyrimidine monophosphates can be interconverted or converted into 2′-deoxyribonucleotides depending on the nucleotide required by the parasitic cell.

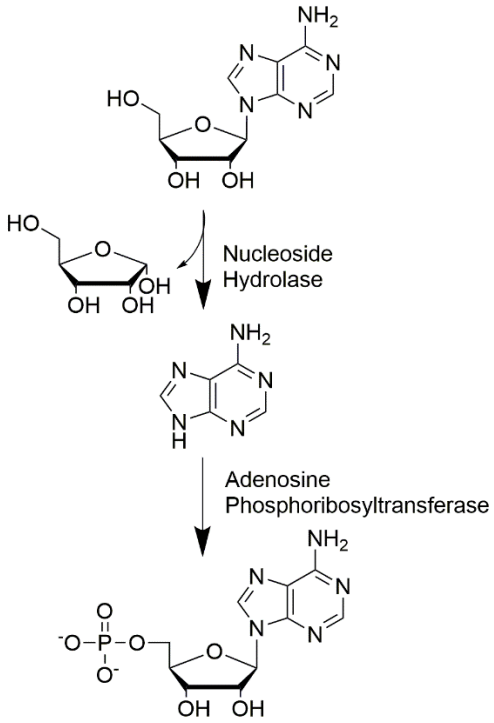


Figure 1.5. Simplified schematic of the purine (pyrimidine) salvage pathway (PSP).

Nucleoside hydrolases are an intriguing target to treat trypanosomatid diseases since they do not have a mammalian analogue.[23-24] Indeed, several compounds have been developed that are anti-parasitic and target nucleoside hydrolase enzymes;[25-30] however, syntheses of these compounds are often multi-step and offer low yields.[31]

These enzymes can be classified based on their substrate specificity as: non-specific inosine-uridine NH (IU-NH), pyrimidine-specific cytidine-uridine NH (CU-NH), purine-specific inosine-adenosine-guanosine NH (IAG-NH), or 6-oxopurine-specific inosine-guanosine NH (IG-NH).[17, 25, 32] Knowledge of the structural basis for the substrate specificity of each nucleoside hydrolase group is important for developing new treatments for trypanosomatid diseases. Interestingly, mutating two specific active site residues to Tyr allows pyrimidine specific CU-NH to hydrolyze inosine, effectively permitting non-specific substrate specificity characteristic of IU-NH.[33] However, the role that these two mutated residues play in the catalytic mechanism is poorly understood. One goal of this thesis is to understand the roles that active site residues play in determining the substrate specificity of nucleoside hydrolases.

Hydrolysis of the glycosidic bond that links nucleobases to sugars is not only an important reaction catalyzed by nucleoside hydrolases as part of the PSP, but is also used to initiate DNA repair. Specifically, DNA glycosylases, including AAG, identify lesions among undamaged DNA and cleave the glycosidic bond that links the nucleobase to the backbone. Like nucleoside hydrolases, AAG possesses broad substrate specificity, which permits repair of many different DNA lesions. In the next section, background is provided about common types of nucleic acid modifications, including several types of DNA damage that are processed by AAG.

1.5 Nucleic Acid Modifications

DNA has evolved as the primary information storage molecule in nature due to its relative stability over RNA.[1] Indeed, RNA can undergo auto-hydrolysis of the phosphodiester bond when the 2′-hydroxy group attacks the phosphorus atom of the 3′-phosphate. Furthermore, the base-pairing interactions in the double helix limit the availability of nucleobases to compounds that can lead to erroneous modifications. Nevertheless, DNA can be chemically modified to serve a wide variety of functions. For example, synthetic nucleic acids can function as biological fluorescence probes,[34-35] and antimicrobial or anticancer agents.[36]

The focus of this thesis is on several nucleic acid modifications that occur naturally to DNA, including damage to nucleobases[37-38] and nucleotide derivatives that regulate gene expression.[39]

1.5.1 DNA Damage

DNA damage can be characterized as bulky or non-bulky modifications (Figure 1.6). Bulky damage includes addition products (adducts) that form when nucleobases are exposed to, for example, polycyclic aromatic hydrocarbons, nitrosamines, or mycotoxins.[40-41] Interstrand and intrastrand crosslinks are other examples of bulky damage that occur when DNA reacts with cisplatin, nitrogen mustards, or UV-activated psoralens.[42] Non-bulky damage arises when DNA is alkylated, oxidized, or deaminated (Figure 1.6).[37-38] Both bulky and non-bulky DNA lesions can stall replication and transcription leading to cell death.[43-44] However, replication may result in substitution (base mispairs) or frameshift (base insertion or deletion) mutations, which can lead to carcinogenesis.[37-38, 45-47] When DNA damage stalls standard replicative polymerases, the damage is often processed (i.e., bypassed) by error-prone translesion synthesis (TLS) DNA polymerases, which often cause mutations.[37, 48-49] Alternatively, DNA damage is reversed by DNA repair enzymes to prevent cytotoxic or mutagenic consequences. Since this thesis examines two enzymes (AAG and AlkB) that function in the repair of non-bulky alkylated and/or deaminated DNA, the formation, replication, and repair of these types of damage will be discussed in more detail below.

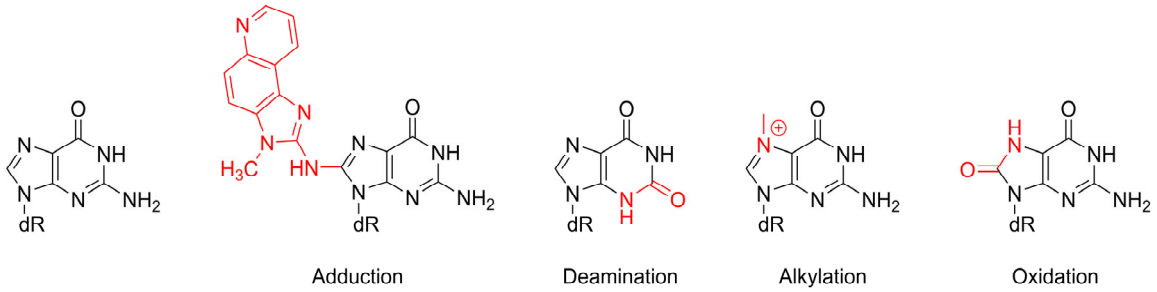


Figure 1.6. Examples of adducted, deaminated, alkylated, and oxidized DNA lesions of guanine.

1.5.1.1 Alkylation

The simplest form of alkylation damage is the transfer of a methyl group to a nucleobase (Figure 1.7). Specifically, exposure to endogeneous *S*-adenosylmethionine or exogeneous formaldehyde, nitrosamines, and nitrosureas generates several cytotoxic and/or mutagenic methyl lesions including N7-methylguanine (7MeG), N3-methyladenine (3MeA), N1-methyladenine (1MeA), and N3-methylcytosine (3MeC).[50-53] Both 7MeG and 3MeA are substrates of AAG,[8] while 1MeA and 3MeC are repaired by AlkB enzymes.[54] However, several deleterious outcomes can result if these lesions are left unrepaired. Specifically, while 7MeG is not inherently mutagenic or cytotoxic, 7MeG depurinates 10^6 times faster than G and generates abasic sites, which are mutagenic (A is preferentially inserted).[55] Alternatively, 7MeG can stall DNA replication or transcription.[56] Neither 1MeA orn 3MeA are not mutagenic; however, each lesion blocks DNA replication leading to cell death.[52-53] Replication of 3MeC also blocks standard polymerases,[57-58] and bypass by TLS polymerases leads to C→A and C→T mutations.[54]

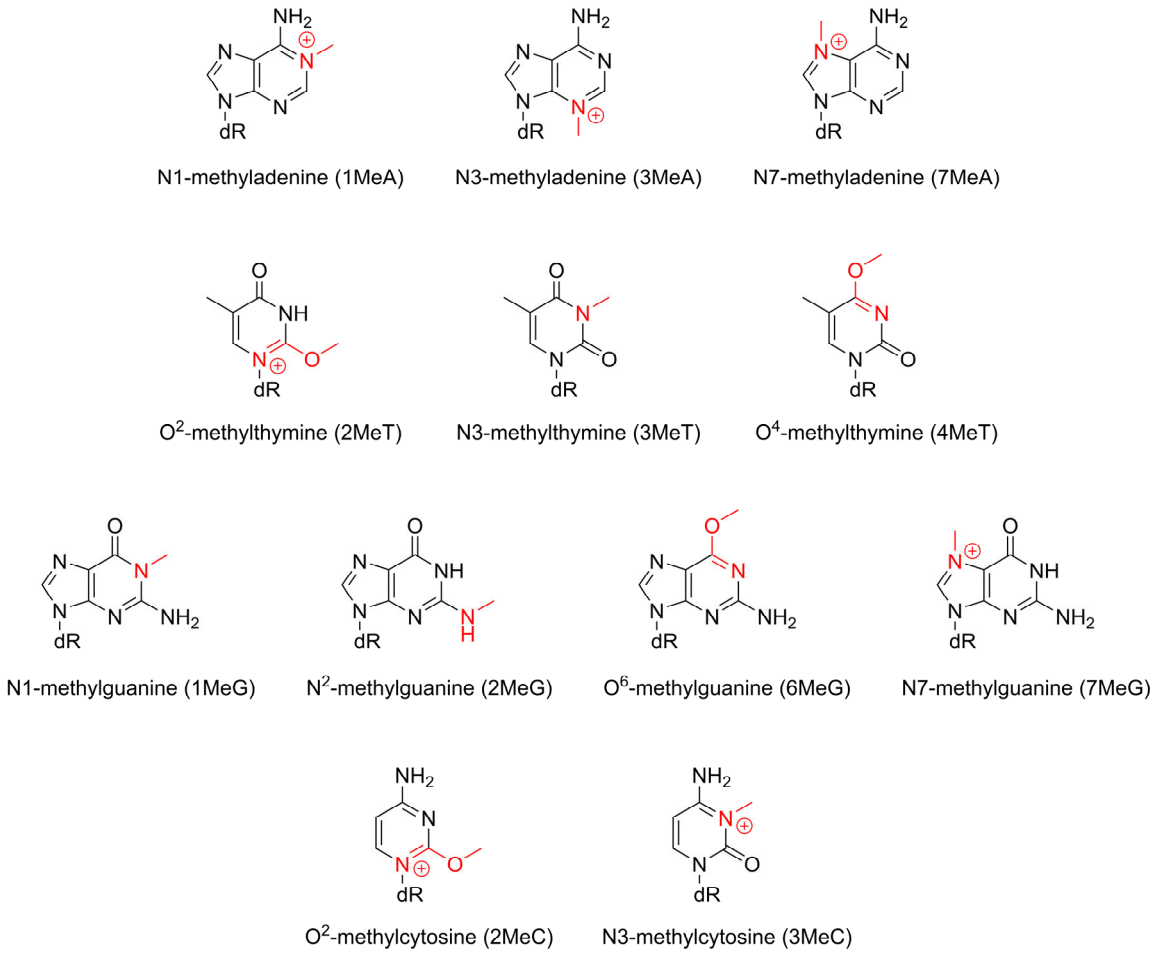


Figure 1.7. Structure of common methylated DNA lesions of A, T, G, and C.

Alkylation damage can also lead to the addition of alkyl groups that are larger than methyl. For example, etheno lesions (Figure 1.8) such as 1,N^6-ethenoadenine (1,N^6-εA), 3,N^4-ethenocytosine (3,N^4-εC), 1,N^2-ethenoguanine (1,N^2-εG), and N^2,3-ethenoguanine (N^2,3-εG) form when DNA is exposed to unsaturated products of lipid peroxidation, or the common industrial agent vinyl chloride and its metabolites.[52, 59] Despite the structural differences between etheno lesions and methylated nucleobases, select lesions from both groups are substrates for DNA glycosylases and AlkB enzymes. Specifically, AAG initiates repair of 1,N^6-εA and 1,N^2-εG,[8] while AlkB repairs 1,N^6-εA, 3,N^4-εC, and 1,N^2-εG,[10, 60-61] and another DNA glycosylase, namely 3-methyladenine DNA glycosylase II (AlkA), initiates N^2,3-εG repair.[62] However, if left unrepaired, each etheno lesion is a replication block for standard DNA polymerases, which

can be cytotoxic.[10, 60] Specifically, successful bypass of 1,N^6-εA by TLS polymerases leads to A→T or A→G substitutions,[10, 63] while bypass of 3,N^4-εC causes C→A and C→T substitutions.[10] Furthermore, G→A, G→T, and G→C mutations arise when either 1,N^2-εG or N^2,3-εG are copied by bypass polymerases.[60] The mutagenic and cytotoxic consequences of the etheno lesions underscores the importance of repair. Therefore, this thesis considers the recognition and repair of many methyl and etheno lesions by DNA repair enzymes.

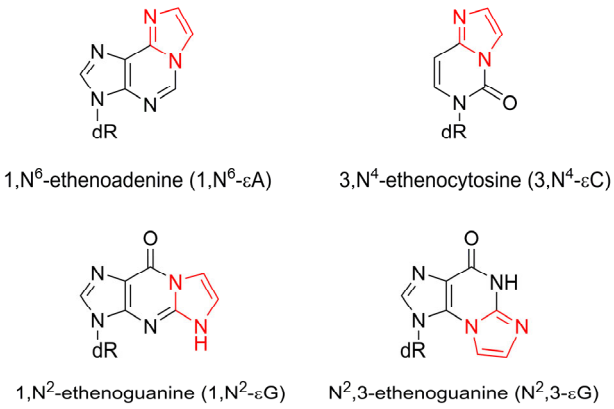


Figure 1.8. Structure of common etheno DNA lesions.

1.5.1.2 Deamination

In addition to alkylation, a common type of DNA damage is oxidative deamination, which replaces a nucleobase amine with a carbonyl. DNA can be oxidatively deaminated by reactive oxygen or nitrogen species (ROS or RNS; e.g., hydroxyl or nitric oxide radicals).[64-65] Cytosine is the most common nucleobase in DNA that is deaminated, which generates U (Figure 1.9).[64, 66] If left unrepaired, replication of U yields C→T transitions mutations.[65, 67] Similarly, deamination of A and G generates hypoxanthine (Hx) and xanthine (X), respectively, which are also mutagenic and result in transition mutations.[64] Deaminated DNA is repaired by several DNA glycosylases including UDG, human uracil DNA glycosylase II (hUNG2), and AAG.[68-73] In this thesis, focus is placed on the AAG-catalyzed repair of Hx.

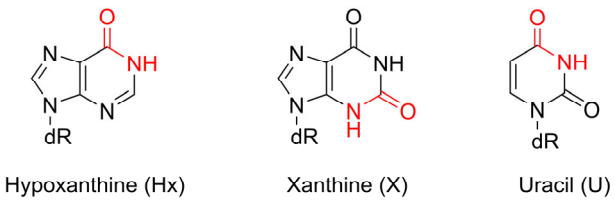
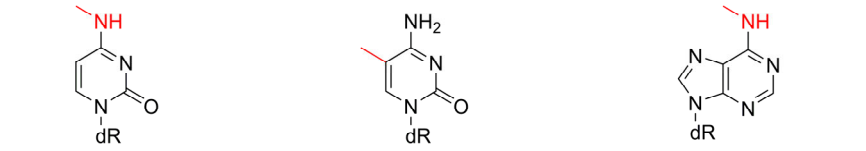


Figure 1.9. Structure of deaminated DNA lesions.

1.5.2 Epigenetic Modifications

Beyond DNA damage, several modifications to canonical DNA are introduced by enzymes to regulate cellular activities.[39] Selective methylation of CpG sites leads to 5-methylcytosine (5MeC, Figure 1.10), which can control gene expression by recruiting reader or effector proteins that bind to 5MeC and perform additional functions.[74-75] Specifically, when 5MeC is present within the promoter regions of DNA, transcription of DNA is repressed.[39] This has been hypothesized to occur by two mechanisms: 1) 5MeC recruits reader proteins thereby blocking the binding of transcription factors, and 2) transcription factors may bind poorly to methylated DNA. However, 5MeC can also enhance transcription if it is present in the transcribed portion of the DNA (i.e., the gene body), which may be due to changes to the interactions between nucleosomes and DNA or alterations to RNA splicing and transcription elongation.[76-77] Two other methyl modifications, 4-methylcytosine (4MeC) and 6-methyladenine (6MeA), also regulate DNA replication and transcription,[78-81] and participate in defense mechanisms against foreign DNA in bacterial systems.[82] The regulation of these modifications within DNA is critical since changes to the epigenetic profiles are associated with several diseases, including cancer.[83-84] In this thesis, AlkB is implicated in the regulation of 5MeC within the genome, which expands its role beyond DNA repair. Since DNA repair is essential to prevent genetic diseases, more detail is provided in the next section on common repair pathways, including base excision repair that uses AAG and direct repair facilitated by AlkB enzymes that are studied in this thesis.



N^4-methylcytosine (4MeC) C5-methylcytosine (5MeC) N^6-methyladenine (6MeA)

Figure 1.10. Structure of common epigenetic modifications present in DNA.

1.6 Essential Roles Played by DNA Repair Enzymes

1.6.1 Base Excision Repair

Base excision repair (BER) is one of several pathways that maintain the genomic integrity of the cell by repairing DNA nucleobase damage (Figure 1.11).[68-73] BER is initiated by DNA glycosylases, which recognize a lesion, extrude the nucleotide into the active site, and cleave the *N*-glycosidic bond connecting the nucleobase to the sugar-phosphate backbone. Two types of DNA glycosylases exist. Monofunctional glycosylases utilize an Asp- or Glu-activated water to facilitate deglycosylation, while bifunctional glycosylases cleave the glycosidic bond with an active site amine and facilitate scission of the 3′- and/or 5′-phosphodiester bonds of the resulting apyrimidinic/apurinic (AP) site. If the AP site is not processed by a glycosylase, AP lyase and AP endonuclease (APE1) cleave the 3′- and 5′- phosphodiester bonds, respectively. A DNA polymerase (generally polymerase β (pol β) in humans) inserts the correct nucleotide using the opposing strand as a template, and a DNA ligase connects the 3′ end of the inserted nucleotide with the rest of the strand.

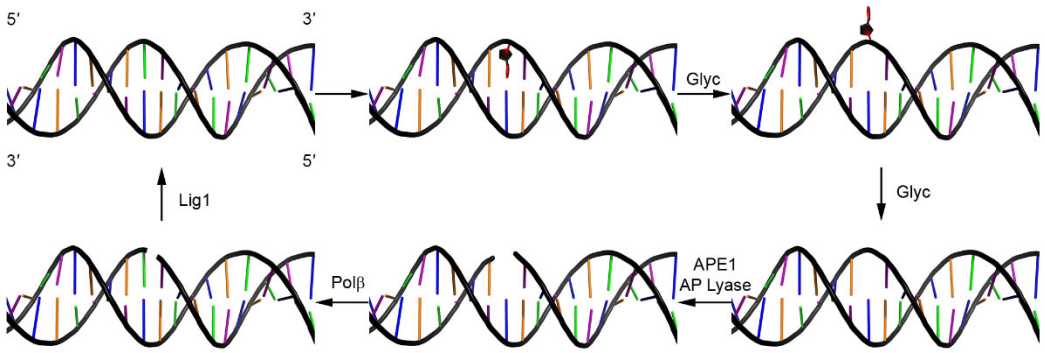


Figure 1.11. Schematic representation of the BER pathway.

Alkyladenine DNA glycosylase (AAG) initiates repair of deaminated or alkylated purines in mammals, plants and bacteria.[8, 69, 85-87] The most studied member of the monofunctional AAG subfamily is human AAG, which has been shown to repair a diverse number of substrates, including Hx, εA, 1,N^2-εG, 3MeA, and 7MeG as discussed in Section 1.5.1.1 and 1.5.1.2.[8, 69, 85-87] This broad substrate specificity is unusual among monofunctional DNA glycosylases that are generally specific to a single lesion.[2-3] AAG does not excise alkylated pyrimidines (e.g., 3MeC) and several pyrimidine lesions inhibit enzyme function (e.g., 3,N^4-ethenocytosine, εC).[88] Interestingly, AAG binds to canonical burines but does not catalyze deglycosylation of these nucleotides, suggesting that AAG can discriminate against the hydrolysis of undamaged DNA.[8] Unfortunately, a structural explanation for the broad, yet discriminatory, substrate specificity remains elusive. It is important to understand the how AAG and other BER proteins function since they maintain the fidelity of the genome, and malfunctioning of BER can lead to the accumulation of damaged DNA and genetic diseases.[89-90] Understanding how AAG functions is especially important since it protects the genome from a vast array of DNA lesions.

1.6.2 Direct Reversal Repair

In addition to BER, alkylated DNA can be repaired through direct repair by the AlkB family of enzymes as discussed in Section 1.5.1.1 and 1.5.2.[9-11, 54, 91-94] These enzymes are Fe(II)- and α-ketoglutarate (αKG)-dependent enzymes that utilize molecular oxygen and decarboxylation of αKG to drive oxidative dealkylation of damaged DNA (Figure 1.12).[95] The *Escherichia coli* AlkB protein is the most studied among this family, although AlkB enzymes are ubiquitous in both prokaryotes and eukaryotes, including 9 human homologs (ALKBH1-8 and fat mass and obesity-associated protein (FTO)).[95-98] Among the human homologues, only ALKBH2 and ALKBH3 are DNA repair proteins with activity comparable to *E. coli* AlkB. In contrast to AAG, which repairs 3MeA and 7MeG lesions,[8] AlkB enzymes target methylation damage that occurs to the WC face of DNA (Figure 1.3).[9, 54, 91-94] However, there is substrate overlap between the two classes of enzymes, since both AlkB and AAG can repair 1,N^6-εA- and 1,N^2-εG-containing DNA. Nevertheless, AlkB enzymes repair 3,N^4-εC, which inhibits AAG. [8, 10-11, 88, 99] Like AAG proteins, AlkB enzymes play critical roles in the repair of damaged DNA and therefore safeguard organisms from genetic diseases, but also repair damaged DNA induced by chemotherapeutics in tumor cells.[89-90] To regulate the activity of AlkB enzymes within cells, an understanding of the full substrate scope and other biological roles played by these enzymes beyond DNA repair is required.

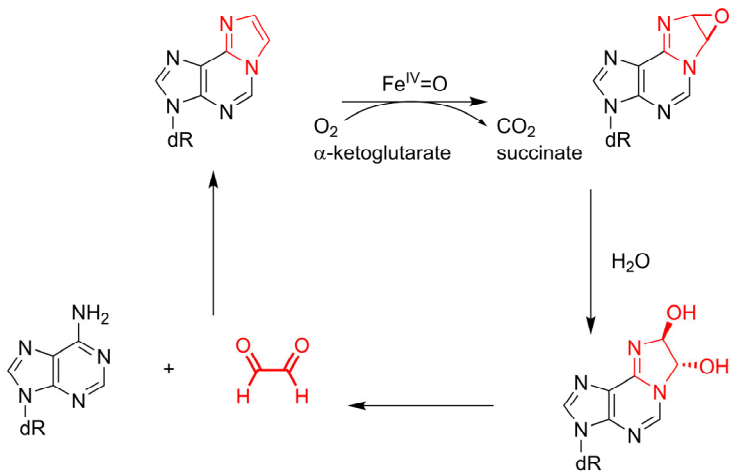


Figure 1.12. AlkB-catalyzed repair of 1,N^6-εA-containing DNA with the aberrant atoms highlighted.

1.7 Computational Approaches Used to Study Enzyme–Substrate Complexes and Catalysis

As highlighted above, the overarching theme of this thesis is to examine the catalytic strategies used by enzymes to achieve broad substrate specificity. Computational chemistry is a valuable tool for studying the structure and function of nucleic acid–protein systems by providing atomic-level details that complement experimental data. For nucleoside hydrolases, QM calculations have identified that π-interactions are critical to stabilize the nucleobase leaving group during deglycosylation catalyzed by IAG-NH.[100] However, the effect of acid, base, or simultaneous acid–base catalysis on deglycosylation of RNA nucleosides is not well understood, and therefore the potential catalytic impact of acidic or basic residues within the enzyme active site is not known. Nevertheless, several forcefield MD and QM/MM studies have determined the catalytic mechanism of select nucleoside hydrolases and clarified the catalytic role of select active site residues.[101-104] However, further atomic details are needed to understand the structural basis for the substrate specificity of different classes of nucleoside hydrolases.

To gain insight into the function of DNA glycosylases, multiscale computational approaches have been employed in the literature. Specifically, density functional theory (DFT) calculations have

investigated the relative glycosidic bond stability among canonical and modified nucleobases,[105-111] and how nucleophile and/or nucleobase activation affects the hydrolytic barrier, which provided insight into the catalytic strategies utilized by DNA glycosylases. For AAG, molecular dynamics (MD) and Monte Carlo (MC) simulations on backbone-constrained systems have provided information about how AAG binds to a subset of substrates,[112] while quantum mechanics (QM)/molecular mechanics (MM) calculations have characterized AAG-mediated deglycosylation of A, 3MeA, and 1,N^6-εA.[113] However, each study utilized constrained computational models, which may not adequately account for the dynamics of substrate binding. Further calculations are also required to understand how AAG binds to additional substrates and discriminates against natural DNA.

In terms of AlkB enzymes, QM cluster models (substrate and active site residues) have revealed the mechanism by which αKG is decarboxylated by AlkB to yield succinate, and subsequent repair of 1MeA and 3MeC.[114-116] A QM/MM study also explored AlkB-catalyzed repair of 1,N^6-εA and proposed structures for repair intermediates identified by mass spectrometry.[117] Nevertheless, knowledge of how AlkB enzymes position structurally diverse alkylated substrates for oxidation remains elusive. Therefore, further structural information is needed to understand the substrate scope of AlkB enzymes and how the catalytic mechanism changes depending on the substrate.

1.8 Thesis Overview

The goal of this thesis is to enhance the understanding of the catalytic strategies utilized by enzymes that possess broad substrate specificity. To this end, DFT, forcefield MD, and QM/MM calculations are utilized to provide significant insights into the function of nucleoside hydrolases, DNA glycosylases, and AlkB enzymes. Specifically, Chapter 2 uses MD simulations to examine the conformation of the AAG active site upon binding of DNA-containing neutral lesions, cationic lesions, and natural DNA, and reveals structural changes to the position and orientation of catalytic residues with respect to the bound

nucleotide. Chapter 3 builds upon the results of Chapter 2 by using ONIOM(QM:MM) calculations to consider how the structural changes impact the catalyzed deglycosylation reaction by performing ONIOM(QM:MM) calculations. The combined results from these Chapters contribute to the understanding of how deaminated and alkylated DNA is repaired. DFT calculations are used in Chapter 4 to characterize the reaction pathways for the hydrolytic deglycosylation of the natural RNA nucleosides, providing fundamental insight into the inherent reactivity of the RNA glycosidic bond and how acid, base, and combined-acid base catalysis affect the reactivity. A structural explanation for the tunable substrate specificity of CU-NH is provided by Chapter 5. Specifically, Chapter 5 builds upon Chapter 4 by examining how the orientations of acidic and basic active site residues with respect to the substrate change depending on the mutational profile of the enzyme. Finally, Chapters 6 and 7 study the AlkB enzymes. Specifically, a combined forcefield MD and QM/MM approach is used in Chapter 6 to examine how AlkB and the human homolog ALKBH2 bind to etheno lesions, and how the position of the substrate with respect to catalytic residues and the metal center permits oxidation. Chapter 7 is a collaborative project that uses mass spectrometry and mutagenesis experiments combined with MD simulations to show that AlkB proteins could play a role in epigenetic regulation by stabilizing 5MeC in the active site to permit oxidation. Finally, Chapter 8 summarizes the how this thesis has contributed to the understanding of structure and function relationships within nucleoside hydrolases, AAG, and AlkB enzymes. This chapter also presents avenues for further investigations that can enhance the knowledge of the roles that enzymes with broad substrate specificities play in biology.

1.9 References

[1] Voet, D.; Voet, J. G., *Biochemistry*. 3 ed.; John Wiley & Sons: 2004.

[2] Drohat, A. C.; Jagadeesh, J.; Ferguson, E.; Stivers, J. T., Role of Electrophilic and General Base Catalysis in the Mechanism of *Escherichia coli* Uracil DNA Glycosylase. *Biochemistry* **1999,** *38* (37), 11866–11875.

[3] Lee, S.; Verdine, G. L., Atomic Substitution Reveals the Structural Basis for Substrate Adenine Recognition and Removal by Adenine DNA Glycosylase. *Proc. Natl. Acad. Sci. USA* **2009,** *106* (44), 18497–18502.

[4] Cui, L. W.; Rajasekariah, G. R.; Martin, S. K., A Nonspecific Nucleoside Hydrolase from *Leishmania donovani*: Implications for Purine Salvage by the Parasite. *Gene* **2001,** *280* (1-2), 153–162.

[5] Versees, W.; Decanniere, K.; Pelle, R.; Depoorter, J.; Brosens, E.; Parkin, D. W.; Steyaert, J., Structure and Function of a Novel Purine Specific Nucleoside Hydrolase from *Trypanosoma vivax*. *J. Mol. Biol.* **2001,** *307* (5), 1363–1379.

[6] Giabbai, B.; Degano, M., Crystal Structure to 1.7 Å of the *Escherichia coli* Pyrimidine Nucleoside Hydrolase YeiK, a Novel Candidate for Cancer Gene Therapy. *Structure* **2004,** *12* (5), 739–749.

[7] Lau, A. Y.; Scharer, O. D.; Samson, L.; Verdine, G. L.; Ellenberger, E., Crystal Structure of a Human Alkylbase-DNA Repair Enzyme Complexed to DNA: Mechanisms for Nucleotide Flipping and Base Excision. *Cell* **1998,** *95* (2), 249–258.

[8] O'Brien, P. J.; Ellenberger, T., Dissecting the Broad Substrate Specificity of Human 3-Methyladenine-DNA Glycosylase. *J. Biol. Chem.* **2004,** *279* (11), 9750–9757.

[9] Falnes, P. Ø.; Rognes, T., DNA Repair by Bacterial AlkB Proteins. *Res. Microbiol.* **2003,** *154* (8), 531–538.

[10] 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*, 855–860.

[11] Mishina, Y.; Yang, C.-G.; He, C., Direct Repair of the Exocyclic DNA Adduct 1,N^6-Ethenoadenine by the DNA Repair AlkB Proteins. *J. Am. Chem. Soc.* **2005,** *127* (42), 14594–14595.

[12] Hänsel-Hertsch, R.; Di Antonio, M.; Balasubramanian, S., DNA G-Quadruplexes in the Human Genome: Detection, Functions and Therapeutic Potential. *Nat. Rev. Mol. Cell Biol.* **2017,** *18*, 279–284.

[13] el Kouni, M. H., Potential Chemotherapeutic Targets in the Purine Metabolism of Parasites. *Pharmacol. Ther.* **2003,** *99* (3), 283-309.

[14] Boitz, J. M.; Ullman, B.; Jardim, A.; Carter, N. S., Purine Salvage in *Leishmania*: Complex or Simple by Design? *Trends Parasitol.* **2012,** *28* (8), 345–352.

[15] Wilson, Z. N.; Gilroy, C. A.; Boitz, J. M.; Ullman, B.; Yates, P. A., Genetic Dissection of Pyrimidine Biosynthesis and Salvage in *Leishmania donovani*. *J. Biol. Chem.* **2012,** *287* (16), 12759–12770.

[16] Carter, N. S.; Yates, P.; Arendt, C. S.; Boitz, J. M.; Ullman, B., Purine and Pyrimidine Metabolism in *Leishmania*. *Drug Targets in Kinetoplastid Parasites* **2008,** *625*, 141–154.

[17] Hammond, D. J.; Gutteridge, W. E., Purine and Pyrimidine Metabolism in the *Trypanosomatidae*. *Mol. Biochem. Parasitol.* **1984,** *13* (3), 243–261.

[18] Murkin, A. S.; Moynihan, M. M., Transition- State-Guided Drug Design for Treatment of Parasitic Neglected Tropical Diseases. *Curr. Med. Chem.* **2014,** *21* (15), 1781–1793.

[19] Fairlamb, A. H., Novel Biochemical Pathways in Parasitic Protozoa. *Parasitology* **1989,** *99*, S93–S112.

[20] Versees, W.; Steyaert, J., Catalysis by Nucleoside Hydrolases. *Curr. Opin. Struct. Biol.* **2003,** *13* (6), 731–738.

[21] Ducati, R. G.; Breda, A.; Basso, L. A.; Santos, D. S., Purine Salvage Pathway in *Mycobacterium tuberculosis*. *Curr. Med. Chem.* **2011,** *18* (9), 1258–1275.

[22] Shrivastav, N.; Li, D. Y.; Essigmann, J. M., Chemical Biology of Mutagenesis and DNA Repair: Cellular Responses to DNA Alkylation. *Carcinogenesis* **2010,** *31* (1), 59–70.

[23] Degano, M.; Gopaul, D. N.; Scapin, G.; Schramm, V. L.; Sacchettini, J. C., Three-Dimensional Structure of the Inosine-Uridine Nucleoside N-Ribohydrolase from *Crithidia fasciculata*. *Biochemistry* **1996,** *35* (19), 5971–5981.

[24] Horenstein, B. A.; Zabinski, R. F.; Schramm, V. L., A New Class of C-Nucleoside Analogues. 1-(*S*)-Aryl-1,4-Dideoxy-1,4-Imino-*D*-Ribitols, Transition State Analogue Inhibitors of Nucleoside Hydrolase. *Tetrahedron Lett.* **1993,** *34* (45), 7213–7216.

[25] Berg, M.; Kohl, L.; Van der Veken, P.; Joossens, J.; Al-Salabi, M. I.; Castagna, V.; Giannese, F.; Cos, P.; Versees, W.; Steyaert, J.; Grellier, P.; Haemers, A.; Degano, M.; Maes, L.; de Koning, H. P.; Augustyns, K., Evaluation of Nucleoside Hydrolase Inhibitors for Treatment of African Trypanosomiasis. *Antimicrob. Agents Chemother.* **2010,** *54* (5), 1900–1908.

[26] Versees, W.; Goeminne, A.; Berg, M.; Vandemeulebroucke, A.; Haemers, A.; Augustyns, K.; Steyaert, J., Crystal Structures of T. Vivax Nucleoside Hydrolase in Complex with New Potent and Specific Inhibitors. *Biochim. Biophys. Acta., Proteins Proteomics* **2009,** *1794* (6), 953–960.

[27] Goeminne, A.; Berg, M.; McNaughton, M.; Bal, G.; Surpateanu, G.; Van der Veken, P.; De Prol, S.; Versees, W.; Steyaert, J.; Haemers, A.; Augustyns, K., N-Arylmethyl Substituted Iminoribitol Derivatives as Inhibitors of a Purine Specific Nucleoside Hydrolase. *Bioorganic Med. Chem.* **2008,** *16* (14), 6752–6763.

[28] Goeminne, A.; McNaughton, M.; Bal, G.; Surpateanu, G.; Van der Veken, P.; De Prol, S.; Versees, W.; Steyaert, J.; Haemers, A.; Augustyns, K., Synthesis and Biochemical Evaluation of Guanidino-

Alkyl-Ribitol Derivatives as Nucleoside Hydrolase Inhibitors. *Eur. J. Med. Chem.* **2008,** *43* (2), 315–326.

[29] Renno, M. N.; Costa Franca, T. C.; Nico, D.; Palatnik-de-Sousa, C. B.; Tinoco, L. W.; Figueroa-Villar, J. D., Kinetics and Docking Studies of Two Potential New Inhibitors of the Nucleoside Hydrolase from *Leishmania donovani*. *Eur. J. Med. Chem.* **2012,** *56*, 301–307.

[30] Franca, T. C. C.; Rocha, M. d. R. M.; Reboredo, B. M.; Renno, M. N.; Tinoco, L. W.; Figueroa-Villar, J. D., Design of Inhibitors for Nucleoside Hydrolase from *Leishmania donovani* Using Molecular Dynamics Studies. *J. Braz. Chem. Soc.* **2008,** *19* (1), 64–73.

[31] Goeminne, A.; McNaughton, M.; Bal, G.; Surpateartu, G.; Van der Veken, P.; De Prol, S.; Versees, W.; Steyaert, J.; Apers, S.; Haemers, A.; Augustyns, K., 1,2,3-Triazolylalkylribitol Derivatives as Nucleoside Hydrolase Inhibitors. *Bioorg. Med. Chem. Lett.* **2007,** *17* (9), 2523–2526.

[32] Versees, W.; Steyaert, J., Catalysis by Nucleoside Hydrolases. *Curr. Opin. Struct. Biol.* **2003,** *13* (6), 731–738.

[33] Iovane, E.; Giabba, B.; Muzzolini, L.; Matafora, V.; Fornili, A.; Minici, C.; Giannese, F.; Degano, M., Structural Basis for Substrate Specificity in Group I Nucleoside Hydrolases. *Biochemistry* **2008,** *47* (15), 4418–4426.

[34] Joo, C.; Balci, H.; Ishitsuka, Y.; Buranachai, C.; Ha, T., Advances in Single-Molecule Fluorescence Methods for Molecular Biology. *Annu. Rev. Biochem* **2008,** *77*, 51–76.

[35] Xu, W.; Chan, K. M.; Kool, E. T., Fluorescent Nucleobases as Tools for Studying DNA and RNA. *Nat. Chem.* **2017,** *9* (11), 1043–1055.

[36] Hernández, D.; Boto, A., Nucleoside Analogues: Synthesis and Biological Properties of Azanucleoside Derivatives. *Eur. J. Org. Chem.* **2014,** *2014* (11), 2201–2220.

[37] Ignatov, A. V.; Bondarenko, K. A.; Makarova, A. V., Non-Bulky Lesions in Human DNA: The Ways of Formation, Repair, and Replication. *Acta Naturae* **2017,** *9* (3), 12–26.

[38] Shafirovich, V.; Geacintov, N. E., Removal of Oxidatively Generated DNA Damage by Overlapping Repair Pathways. *Free Radical Biol. Med.* **2017,** *107*, 53–61.

[39] Chen, K.; Zhao, Boxuan S.; He, C., Nucleic Acid Modifications in Regulation of Gene Expression. *Cell Chem. Biol.* **2016,** *23* (1), 74–85.

[40] Giglia-Mari, G.; Zotter, A.; Vermeulen, W., DNA Damage Response. *Cold Spring Harb. Perspect. Biol.* **2011,** *3* (1), a000745.

[41] Lord, C. J.; Ashworth, A., The DNA Damage Response and Cancer Therapy. *Nature* **2012,** *481* (7381), 287-294.

[42] Cadet, J.; Davies, K. J. A.; Medeiros, M. H. G.; Di Masciod, P.; Wagner, J. R., Formation and Repair of Oxidatively Generated Damage in Cellular DNA. *Free Radical Biol. Med.* **2017,** *107*, 13–34.

[43] Sun, D.; Hurley, L. H., Effect of the (+)-CC-1065-(N3-Adenine)DNA Adduct on *in vitro* DNA Synthesis Mediated by *Escherichia coli* DNA Polymerase. *Biochemistry* **1992,** *31* (10), 2822–2829.

[44] Minca, E. C.; Kowalski, D., Replication Fork Stalling by Bulky DNA Damage: Localization at Active Origins and Checkpoint Modulation. *Nucleic Acids Res.* **2011,** *39* (7), 2610–2623.

[45] Hecht, S. S., Progress and Challenges in Selected Areas of Tobacco Carcinogenesis. *Chem. Res. Toxicol.* **2008,** *21* (1), 160–171.

[46] Hecht, S. S., Tobacco Smoke Carcinogens and Lung Cancer. *J. Natl. Cancer. I.* **1999,** *91* (14), 1194–1210.

[47] Gazinska, P.; Herman, D.; Gillett, C.; Pinder, S.; Mantle, P., Comparative Immunohistochemical Analysis of Ochratoxin a Tumourigenesis in Rats and Urinary Tract Carcinoma in Humans; Mechanistic Significance of P-S6 Ribosomal Protein Expression. *Toxins (Basel)* **2012,** *4* (9), 643.

[48] Wilson, K. A.; Wetmore, S. D., Complex Conformational Heterogeneity of the Highly Flexible O6-Benzyl-Guanine DNA Adduct. *Chem. Res. Toxicol.* **2014,** *27* (7), 1310–1325.

[49] Monti, P.; Traverso, I.; Casolari, L.; Menichini, P.; Inga, A.; Ottaggio, L.; Russo, D.; Iyer, P.; Gold, B.; Fronza, G., Mutagenicity of N3-Methyladenine: A Multi-Translesion Polymerase Affair. *Mutat. Res., Fundam. Mol. Mech. Mutagen.* **2010,** *683*, 50–56.

[50] Culp, L. A.; Dore, E.; Brown, G. M., Methylated Bases in DNA of Animal Origin. *Arch. Biochem. Biophys.* **1970,** *136* (1), 73-79.

[51] IARC *Formaldehyde, 2-Butoxyethanol and 1-Tert-Butoxypropan-2-Ol*; Lyon, Fr, 2006; pp 1-478.

[52] Shrivastav, N.; Li, D.; Essigmann, J. M., Chemical Biology of Mutagenesis and DNA Repair: Cellular Responses to DNA Alkylation. *Carcinogenesis* **2010,** *31* (1), 59–70.

[53] Sedgwick, B., Repairing DNA-Methylation Damage. *Nat. Rev. Mol. Cell Biol.* **2004,** *5* (2), 148–157.

[54] Delaney, J. C.; Essigmann, J. M., Mutagenesis, Genotoxicity, and Repair of 1-Methyladenine, 3-Alkylcytosines, 1-Methylguanine, and 3-Methylthymine in AlkB Escherichia Coli. *Proc. Natl. Acad. Sci. USA* **2004,** *101* (39), 14051–14056.

[55] Obeid, S.; Blatter, N.; Kranaster, R.; Schnur, A.; Diederichs, K.; Welte, W.; Marx, A., Replication through an Abasic DNA Lesion: Structural Basis for Adenine Selectivity. *Embo J.* **2010,** *29* (10), 1738–1747.

[56] Boysen, G.; Pachkowski, B. F.; Nakamura, J.; Swenberg, J. A., The Formation and Biological Significance of N7-Guanine Adducts. *Mutat. Res.-Genet. Toxicol. Environ. Mutag.* **2009,** *678* (2), 76–94.

[57] Boiteux, S.; Laval, J., Mutagenesis by Alkylating Agents: Coding Properties for DNA Polymerase of Poly (dC) Template Containing 3-Methylcytosine. *Biochimie* **1982,** *64* (8), 637–641.

[58] Huff, A. C.; Topal, M. D., DNA Damage at Thymine N-3 Abolishes Base-Pairing Capacity During DNA Synthesis. *J. Biol. Chem.* **1987,** *262* (26), 12843–12850.

[59] 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.

[60] Chang, S.-c.; Fedeles, B. I.; Wu, J.; Delaney, J. C.; Li, D.; Zhao, L.; Christov, P. P.; Yau, E.; Singh, V.; Jost, M.; Drennan, C. L.; Marnett, L. J.; Rizzo, C. J.; Levine, S. S.; Guengerich, F. P.; Essigmann, J. M., Next-Generation Sequencing Reveals the Biological Significance of the N^2,3-Ethenoguanine Lesion in Vivo. *Nucleic Acids Res.* **2015,** *43* (11), 5489–5500.

[61] Li, D.; Delaney, J. C.; Page, C. M.; Yang, X.; Chen, A. S.; Wong, C.; Drennan, C. L.; Essigmann, J. M., Exocyclic Carbons Adjacent to the N6 of Adenine Are Targets for Oxidation by the *Escherichia coli* Adaptive Response Protein AlkB. *J. Am. Chem. Soc.* **2012,** *134* (21), 8896–8901.

[62] Matijasevic, Z.; Sekiguchi, M.; Ludlum, D. B., Release of N^2,3-Ethenoguanine from Chloroacetaldehyde-Treated DNA by *Escherichia coli* 3-Methyladenine DNA Glycosylase II. *Proc. Natl. Acad. Sci. USA* **1992,** *89* (19), 9331–9334.

[63] Pandya, G. A.; Moriya, M., 1,N^6-Ethenodeoxyadenosine, a DNA Adduct Highly Mutagenic in Mammalian Cells. *Biochemistry* **1996,** *35* (35), 11487–11492.

[64] Kow, Y. W., Repair of Deaminated Bases in DNA Editor: Miral Dizdaroglu. *Free Radical Biol. Med.* **2002,** *33* (7), 886–893.

[65] Kreutzer, D. A.; Essigmann, J. M., Oxidized, Deaminated Cytosines Are a Source of C → T Transitions *in vivo*. *Proc. Natl. Acad. Sci. USA* **1998,** *95* (7), 3578–3582.

[66] Lindahl, T.; Nyberg, B., Heat-Induced Deamination of Cytosine Residues in Deoxyribonucleic Acid. *Biochemistry* **1974,** *13* (16), 3405–3410.

[67] Duncan, B. K.; Miller, J. H., Mutagenic Deamination of Cytosine Residues in DNA. *Nature* **1980,** *287*, 560.

[68] Stivers, J. T.; Jiang, Y. L., A Mechanistic Perspective on the Chemistry of DNA Repair Glycosylases. *Chem. Rev.* **2003,** *103* (7), 2729–2759.

[69] Berti, P. J.; McCann, J. A. B., 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.

[70] Zharkov, D. O., Base Excision DNA Repair. *Cell. Mol. Life Sci.* **2008,** *65*, 1544–1565.

[71] Drohat, A. C.; Maiti, A., Mechanisms for Enzymatic Cleavage of the N-Glycosidic Bond in DNA. *Org. Biomol. Chem.* **2014,** *12*, 8367-8378.

[72] Brooks, S. C.; Adhikary, S.; Rubinson, E. H.; Eichman, B. F., Recent Advances in the Structural Mechanisms of DNA Glycosylases. *Biochim. Biophys. Acta., Proteins Proteomics* **2013,** *1834* (1), 247–271.

[73] Schermerhorn, K. M.; Delaney, S., A Chemical and Kinetic Perspective on Base Excision Repair of DNA. *Acc. Chem. Res.* **2014,** *47* (4), 1238–1246.

[74] Meehan, R. R.; Lewis, J. D.; McKay, S.; Kleiner, E. L.; Bird, A. P., Identification of a Mammalian Protein That Binds Specifically to DNA Containing Methylated CpGs. *Cell* **1989,** *58* (3), 499–507.

[75] Li, E.; Beard, C.; Jaenisch, R., Role for DNA Methylation in Genomic Imprinting. *Nature* **1993,** *366*, 362.

[76] Aran, D.; Toperoff, G.; Rosenberg, M.; Hellman, A., Replication Timing-Related and Gene Body-Specific Methylation of Active Human Genes. *Human Molecular Genetics* **2011,** *20* (4), 670–680.

[77] Hellman, A.; Chess, A., Gene Body-Specific Methylation on the Active X Chromosome. *Science* **2007,** *315* (5815), 1141–1143.

[78] Wion, D.; Casadesús, J., N6-Methyl-Adenine: An Epigenetic Signal for DNA–Protein Interactions. *Nat. Rev. Microbiol.* **2006,** *4*, 183.

[79] Luo, G. Z.; He, C., DNA N-6-Methyladenine in Metazoans: Functional Epigenetic Mark or Bystander? *Nat. Struct. Mol. Biol.* **2017,** *24* (6), 503–506.

[80] Luo, G.-Z.; Blanco, M. A.; Greer, E. L.; He, C.; Shi, Y., DNA N(6)-Methyladenine: A New Epigenetic Mark in Eukaryotes? *Nat. Rev. Mol. Cell. Biol.* **2015,** *16* (12), 705–710.

[81] Yu, M.; Ji, L.; Neumann, D. A.; Chung, D. H.; Groom, J.; Westpheling, J.; He, C.; Schmitz, R. J., Base-Resolution Detection of N4-Methylcytosine in Genomic DNA Using 4mC-TET-Assisted-Bisulfite-Sequencing. *Nucleic Acids Res* **2015,** *43* (21), e148.

[82] Collier, J.; McAdams, H. H.; Shapiro, L., A DNA Methylation Ratchet Governs Progression through a Bacterial Cell Cycle. *Procl. Natl. Acad. Sci., U.S.A.* **2007,** *104* (43), 17111–17116.

[83] James, C.; Mazurier, F.; Dupont, S.; Chaligne, R.; Lamrissi-Garcia, I.; Tulliez, M.; Lippert, E.; Mahon, F.-X.; Pasquet, J.-M.; Etienne, G.; Delhommeau, F.; Giraudier, S.; Vainchenker, W.; de Verneuil, H., The Hematopoietic Stem Cell Compartment of Jak2v617f-Positive Myeloproliferative Disorders Is a Reflection of Disease Heterogeneity. *Blood* **2008,** *112* (6), 2429–2438.

[84] Liu, C.; Liu, L.; Chen, X.; Shen, J.; Shan, J.; Xu, Y.; Yang, Z.; Wu, L.; Xia, F.; Bie, P.; Cui, Y.; Bian, X.-w.; Qian, C., Decrease of 5-Hydroxymethylcytosine Is Associated with Progression of Hepatocellular Carcinoma through Downregulation of TET1. *PLOS ONE* **2013,** *8* (5), e62828.

[85] O'Brien, P. J.; Ellenberger, T., Human Alkyladenine DNA Glycosylase Uses Acid-Base Catalysis for Selective Excision of Damaged Purines. *Biochemistry* **2003,** *42* (42), 12418–12429.

[86] Wolfe, A. E.; O'Brien, P. J., Kinetic Mechanism for the Flipping and Excision of 1,N^6-Ethenoadenine by Human Alkyladenine DNA Glycosylase. *Biochemistry* **2009,** *48* (48), 11357–11369.

[87] Wuenschell, G. E.; O'Connor, T. R.; Termini, J., Stability, Miscoding Potential, and Repair of 2'-Deoxyxanthosine in DNA: Implications for Nitric Oxide-Induced Mutagenesis. *Biochemistry* **2003,** *42* (12), 3608–3616.

[88] 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,N^4-Ethenocytosine-Containing DNA. *J. Biol. Chem.* **2011,** *286* (15), 13205–13213.

[89] Knijnenburg, T. A.; Wang, L.; Zimmermann, M. T.; Chambwe, N.; Gao, G. F.; Cherniack, A. D.; Fan, H.; Shen, H.; Way, G. P.; Greene, C. S.; Liu, Y.; Akbani, R.; Feng, B.; Donehower, L. A.; Miller, C.; Shen, Y.; Karimi, M.; Chen, H.; Kim, P.; Jia, P.; Shinbrot, E.; Zhang, S.; Liu, J.; Hu, H.; Bailey, M. H.; Yau, C.; Wolf, D.; Zhao, Z.; Weinstein, J. N.; Li, L.; Ding, L.; Mills, G. B.; Laird, P. W.; Wheeler, D. A.; Shmulevich, I.; The Cancer Genome Atlas Research, N.; Monnat, R. J.; Xiao, Y.; Wang, C., Genomic and Molecular Landscape of DNA Damage Repair Deficiency across the Cancer Genome Atlas. *Cell Reports* **2018,** *23* (1), 239–254.e6.

[90] Hanahan, D.; Weinberg, Robert A., Hallmarks of Cancer: The Next Generation. *Cell* **2011,** *144* (5), 646–674.

[91] 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*, 174.

[92] Aas, P. A.; Otterlei, M.; Falnes, P. Ø.; Vågbø, C. B.; Skorpen, F.; Akbari, M.; Sundheim, O.; Bjørås, M.; Slupphaug, G.; Seeberg, E.; Krokan, H. E., Human and Bacterial Oxidative Demethylases Repair Alkylation Damage in Both RNA and DNA. *Nature* **2003,** *421*, 859.

[93] Koivisto, P.; Duncan, T.; Lindahl, T.; Sedgwick, B., Minimal Methylated Substrate and Extended Substrate Range of *Escherichia coli* AlkB Protein, a 1-Methyladenine-DNA Dioxygenase. *J. Biol. Chem.* **2003,** *278* (45), 44348–44354.

[94] 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. USA* **2002,** *99* (26), 16660-16665.

[95] Muller, T. A.; Hausinger, R. P., Chapter 8 AlkB and Its Homologues - DNA Repair and Beyond. In *2-Oxoglutarate-Dependent Oxygenases*, The Royal Society of Chemistry: 2015; pp 246–262.

[96] Aravind, L.; Koonin, E. V., The DNA-Repair Protein AlkB, EGL-9, and Leprecan Define New Families of 2-Oxoglutarate- and Iron-Dependent Dioxygenases. *Genome Biol.* **2001,** *2* (3), research0007.1-research0007.8.

[97] Kurowski, M. A.; Bhagwat, A. S.; Papaj, G.; Bujnicki, J. M., Phylogenomic Identification of Five New Human Homologs of the DNA Repair Enzyme AlkB. *BMC Genomics* **2003,** *4* (1), 48.

[98] Sundheim, O.; Vågbø, C. B.; Bjørås, M.; Sousa, M. M. L.; Talstad, V.; Aas, P. A.; Drabløs, F.; Krokan, H. E.; Tainer, J. A.; Slupphaug, G., Human ABH3 Structure and Key Residues for Oxidative Demethylation to Reverse DNA/RNA Damage. *The EMBO Journal* **2006,** *25* (14), 3389–3397.

[99] Zdżalik, D.; Domańska, A.; Prorok, P.; Kosicki, K.; van den Born, E.; Falnes, P. Ø.; Rizzo, C. J.; Guengerich, F. P.; Tudek, B., Differential Repair of Etheno-DNA Adducts by Bacterial and Human AlkB Proteins. *DNA Repair* **2015,** *30*, 1–10.

[100] Versees, W.; Loverix, S.; Vandemeulebroucke, A.; Geerlings, P.; Steyaert, J., Leaving Group Activation by Aromatic Stacking: An Alternative to General Acid Catalysis. *J. Mol. Biol.* **2004,** *338* (1), 1–6.

[101] Mazumder-Shivakumar, D.; Bruice, T. C., Computational Study of IAG-Nucleoside Hydrolase: Determination of the Preferred Ground State Conformation and the Role of Active Site Residues. *Biochemistry* **2005,** *44* (21), 7805-7817.

[102] Wu, R.; Gong, W.; Ting, L.; Zhang, Y.; Cao, Z., QM/MM Molecular Dynamics Study of Purine-Specific Nucleoside Hydrolase. *J. Phys. Chem. B* **2012,** *116* (6), 1984-1991.

[103] Mazumder, D.; Bruice, T. C., Exploring Nucleoside Hydrolase Catalysis *in silico*: Molecular Dynamics Study of Enzyme-Bound Substrate and Transition State. *J. Am. Chem. Soc.* **2002,** *124* (49), 14591–14600.

[104] Fornili, A.; Giabbai, B.; Garau, G.; Degano, M., Energy Landscapes Associated with Macromolecular Conformational Changes from Endpoint Structures. *J. Am. Chem. Soc. Society* **2010,** *132* (49), 17570–17577.

[105] Millen, A. L.; Wetmore, S. D., Glycosidic Bond Cleavage in Deoxynucleotides - a Density Functional Study. *Can. J. Chem.* **2009,** *87* (7), 850–863.

[106] Lenz, S. A. P.; Kellie, J. L.; Wetmore, S. D., Glycosidic Bond Cleavage in DNA Nucleosides: Effect of Nucleobase Damage and Activation on the Mechanism and Barrier. *J. Phys. Chem. B* **2015,** *119* (51), 15601–15612.

[107] Navarro-Whyte, L.; Kellie, J. L.; Lenz, S. A. P.; Wetmore, S. D., Hydrolysis of the Damaged Deoxythymidine Glycol Nucleoside and Comparison to Canonical DNA. *PCCP* **2013,** *15* (44), 19343–19352.

[108] Lenz, S. A. P.; Kellie, J. L.; Wetmore, S. D., Glycosidic Bond Cleavage in Deoxynucleotides: Effects of Solvent and the DNA Phosphate Backbone in the Computational Model. *J. Phys. Chem. B* **2012,** *116* (49), 14275–14284.

[109] Przybylski, J. L.; Wetmore, S. D., Modeling the Dissociative Hydrolysis of the Natural DNA Nucleosides. *J. Phys. Chem. B* **2010,** *114* (2), 1104–1113.

[110] Shim, E. J.; Przybylski, J. L.; Wetmore, S. D., Effects of Nucleophile, Oxidative Damage, and Nucleobase Orientation on the Glycosidic Bond Cleavage in Deoxyguanosine. *J. Phys. Chem. B* **2010,** *114* (6), 2319–2326.

[111] Przybylski, J. L.; Wetmore, S. D., Designing an Appropriate Computational Model for DNA Nucleoside Hydrolysis: A Case Study of 2′-Deoxyuridine. *J. Phys. Chem. B* **2009,** *113* (18), 6533–6542.

[112] Alexandrova, A. N., Promiscuous DNA Alkyladenine Glycosylase Dramatically Favors a Bound Lesion over Undamaged Adenine. *Biophys. Chem.* **2010,** *152* (1-3), 118–127.

[113] 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*, 16258–16269.

[114] Liu, H.; Llano, J.; Gauld, J. W., A DFT Study of Nucleobase Dealkylation by the DNA Repair Enzyme AlkB. *J. Phys. Chem. B* **2009,** *113* (14), 4887–4898.

[115] Fang, D.; Lord, R. L.; Cisneros, G. A., Ab Initio QM/MM Calculations Show an Intersystem Crossing in the Hydrogen Abstraction Step in Dealkylation Catalyzed by AlkB. *J. Phys. Chem. B* **2013,** *117* (21), 6410-6420.

[116] Cisneros, G. A., DFT Study of a Model System for the Dealkylation Step Catalyzed by AlkB. *Interdisciplinary Sciences: Comput. Life Sci.* **2010,** *2* (1), 70–77.

[117] 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.

Chapter 2: Evaluating the Substrate Selectivity of Alkyladenine DNA Glycosylase: The Synergistic Interplay of Active Site Flexibility and Water Reorganization[a,b,c]

2.1 Introduction

The DNA nucleobases are susceptible to chemical modification by a number of different intracellular and extracellular sources, which can lead to mutagenic or cytotoxic lesions.[1, 2] Indeed, endogenous reactive oxygen (i.e., hydrogen peroxide or hydroxyl radicals) or nitrogen (i.e., nitric oxide radicals) species can directly react with the DNA nucleobases to form common oxidation lesions.[1-6] For example, adenine reacts with nitric oxide radicals to form hypoxanthine (Hx, Figure 2.1).[7] Alternatively, the DNA nucleobases can be alkylated through a variety of routes. For example, 1,N^6-ethenoadenine (εA, Figure 2.1) and 3,N^4-ethenocytosine (εC, Figure 2.1) are formed from reactions between DNA nucleobases and products of lipid peroxidation.[8, 9] Alternatively, N^3-methyladenine (3MeA, Figure 2.1) arises when A reacts with endogenous S-adenosylmethionine, while N^7-methylguanine (7MeG, Figure 2.1) is formed when DNA is exposed to nitrosamines from tobacco smoke.[10] If left unrepaired, DNA lesions can directly or indirectly cause mutations. For example, replication of DNA in cells with increased εA concentration causes A:T to G:C transition mutations in mammalian cells,[11] which are likely related to the observed inflammation and carcinogenesis associated with intestinal, liver, and brain tissues.[12, 13] Alternatively, the additional steric bulk at N3 of A halts standard DNA replication and mutations are introduced when 3MeA is processed by the notoriously error prone translesion synthesis pathway,[10, 14] while 7MeG depurinates 10^6 times faster than G and leads to an increased presence of abasic sites within the genome.[15]

[a]*Biochemistry* reference style used in this Chapter.
[b]Reprinted with permission from Lenz, S.A.P. and Wetmore, S.D. (2016) Evaluating the Substrate Selectivity of Alkyladenine DNA Glycosylase: The Synergistic Interplay of Active Site Flexibility and Water Reorganization, *Biochemistry 55,* 798-808.
[c]S.A.P.L. performed all calculations. S.A.P.L. and S.D.W. contributed to the design of calculations, interpretation of data, and writing the manuscript.

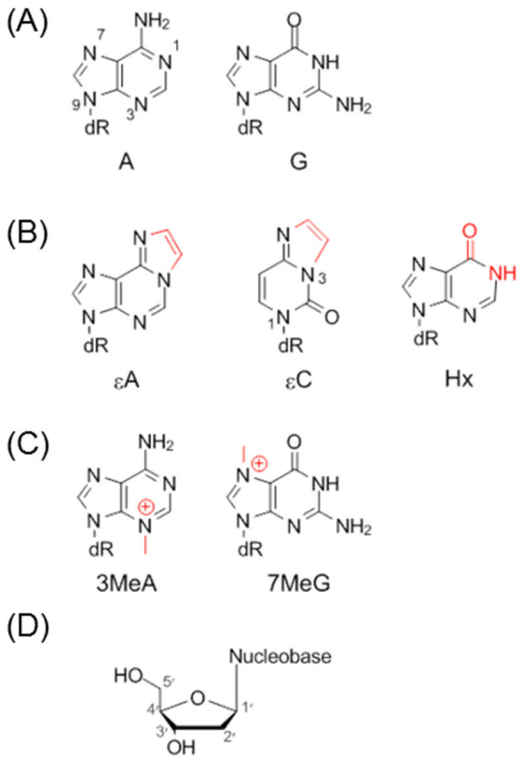


Figure 2.1. Structures and chemical numbering of the (A) canonical purines, (B) neutral purine lesions, (C) cationic purine lesions, and (D) 2′-deoxyribose moiety. Nucleobase modifications are highlighted in red.

The base excision repair (BER) pathway removes lesions to maintain fidelity of DNA.[16-21] The DNA glycosylases initiate the BER pathway by scanning the genome to locate a lesion, flipping the damaged base into the active site, and catalyzing cleavage of the *N*-glycosidic bond connecting the nucleobase to the sugar–phosphate backbone, which generates an abasic site. In humans, alkyladenine DNA glycosylase (AAG) is responsible for the removal of a variety of purine lesions, including Hx, εA, 3MeA, and 7MeG.[22-24] This substrate diversity is unusual among the glycosylase family, with most enzymes targeting a specific lesion.[25-27] Indeed, it is difficult to pinpoint a common structural motif among AAG substrates, which vary in terms of charge, size and position of hydrogen-bond donor/acceptor atoms. As a consequence, AAG sacrifices catalytic efficiency,[22, 23] exhibiting modest catalytic rate enhancements

(10^3 – 10^7) compared to other glycosylases (10^{10} – 10^{26}).[23, 24, 28, 29] Despite the structural diversity among AAG substrates, the enzyme does not catalyze the removal of pyrimidine lesions, with εC inhibiting enzyme activity.[30, 31] Furthermore, although AAG has been shown to bind mismatched A or G, the enzyme exhibits very low catalytic rates when targeting these canonical purines.[23]

The specificity of DNA glycosylases, including AAG, is at least in part dictated by the base-flipping step during which the enzyme inserts an amino acid residue into the DNA helix to facilitate flipping of the damaged nucleotide into the active site.[16-21] For example, since uracil DNA glycosylase (UDG) has a narrow binding pocket that excludes purine substrates, the base-flipping step significantly contributes to the high specificity of this repair enzyme for uracil.[32] AAG activity has also been confirmed to be affected by the base-flipping step.[23] For example, Hx is excised at a higher rate when paired opposite T compared to C,[23] which has been attributed to the lower stability and therefore ease of base flipping in the Hx:T than the Hx:C base pair.[33] Similarly, AAG exhibits greater activity toward the natural purines in destabilized mismatched base pairs than canonical base pairs.[23] Nevertheless, AAG exhibits a catalytic rate several orders of magnitude higher for the excision of damaged nucleobases over mismatched natural purines.[23] Therefore, although base flipping is one important factor, the substrate specificity of AAG is also partially determined by the chemical step. Indeed, the excision of εA is minimally affected by the identity of the opposing base,[23] and the associated rate-limiting step has been determined to be glycosidic bond cleavage.[24]

Due to the observed AAG substrate diversity, it is unsurprising that X-ray crystal structures reveal a lack of direct contacts between the nucleobase of the bound nucleotide and active site amino acids (Figure 2.2).[31, 34, 35] Instead, it has been proposed that AAG primarily uses non-specific π–π interactions (involving Y127, H136 and Y159; Figure 2.2A) to position damaged nucleobases in the active site. A previous computational study suggests that these active site π–π interactions play an additional role, being catalytic toward the removal of neutral nucleobases (εA), but non-catalytic or even anti-catalytic toward

excision of cationic lesions (3MeA).[36] This multi-fold function of π–π interactions may not be surprising given that DNA–protein π–π contacts have been shown to be prevalent across biological systems and provide stability comparable to specific hydrogen-bonding interactions.[37, 38] Another critical AAG active site residue is E125 (Figure 2.2A), which has been proposed to activate the water nucleophile that cleaves the glycosidic bond in the damaged nucleotide via a substitution reaction.[34] Indeed, computational work characterized an S_N2 hydrolysis mechanism facilitated by E125 for the removal of both neutral and cationic AAG substrates.[36] The proximity of E125 to the site of reaction (C1′) is also critical to glycosylase activity since this negatively charged residue stabilizes the positive charge accumulating on the deoxyribose moiety during deglycosylation. Indeed, E125Q and E125A mutations abolish AAG activity.[22] Finally, although it has been proposed that AAG requires a general acid near N7 of the nucleobase for catalysis via leaving group activation, no active site amino acids capable of this function have been identified.[22] Furthermore, despite proposals that a water chain may play the role of the general acid,[35] combined mass spectrometry and computational studies indicate that (full) nucleobase protonation may not be a requirement for base excision.[39, 40]

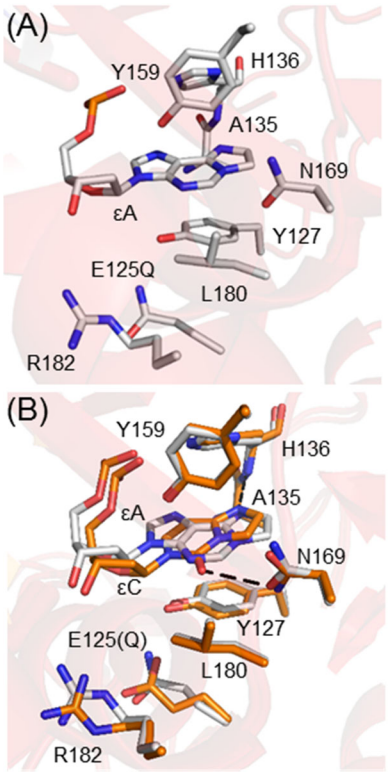


Figure 2.2. (A) X-ray crystal structure of AAG bound to εA (PDB ID: 1EWN).[35] (B) Overlay of the AAG active site bound to εA (gray; PDB ID: 1EWN)[35] or εC (orange; PDB ID: 3QI5).[31] Key hydrogen bonds with the nucleobase are highlighted with dashed lines.

Although important mechanistic information has been acquired to date on AAG, atomic level structural information that explains the broad AAG substrate specificity, and the ability of the enzyme to discriminate against the canonical purines, is still missing. Close comparison of crystal structures of AAG bound to the εA substrate and εC inhibitor (Figure 2.2B) reveals key differences in binding that (at least partially) rationalize the differential AAG activity towards these lesions.[31, 35] Specifically, although both nucleobases form a hydrogen bond with the backbone of H136, εC is pulled further into the active site of the enzyme, and an additional hydrogen bond is formed between εC and the side chain of N169 (Figure 2.2B). The extra hydrogen bond in the case of εC has been proposed to explain the observed two-fold greater binding affinity of AAG for εC compared to εA (that is lost upon N169L or N169I mutation), while

enzyme inhibition was proposed to be caused by the lack of leaving group activation.[31] Interestingly, there are few overall changes to the configuration of AAG active site amino acids upon binding of εA and εC. However, both nucleotides interact with H136 through a similarly positioned hydrogen-bond acceptor (N2 for εC; N6 for εA; Figures 2.2 and Figure A.1 in Appendix A) and not all nucleotides can form this interaction. Indeed, for A it has been proposed that steric clashes between the N6 amino group and the H136 side chain prevent AAG activity at an appreciable rate.[23] Nevertheless, 3MeA possesses the same N6 amino group as A, yet is excised by AAG.[23] This suggests that substrates may bind to the active site in different ways, and active site reorganization may be important for understanding the relative activity that AAG has towards different nucleotides. Indeed, recent experimental evidence indicates that at least part of the AAG active site is disordered prior to substrate binding, and suggests that the final active site conformation may be dictated by the nucleotide bound.[41] However, it is difficult to predetermine the changes that may occur when AAG interacts with chemically diverse nucleotides.

Despite the importance of understanding how different nucleotides bind to AAG, only DNA containing εA or εC have been successfully co-crystallized with AAG to yield structures with intact glycosidic bonds. Indeed, experimental challenges arise since, for example, cationic lesions have short half-lives,[16] while AAG does not bind the canonical purines at an appreciable rate to allow crystallization.[23] However, computational chemistry has proven to be a useful approach to obtain molecular level structural information about biochemical systems (see, for example, references 42-49) including DNA glycosylases.[36, 50-57] Indeed, computational studies have considered certain aspects of the AAG mechanism of action for select nucleotides.[36, 50, 58] Specifically, large-scale ONIOM models were used to gain mechanistic information about AAG activity toward εA, 3MeA, and A,[36] while a range of Monte Carlo (MC)[50] and molecular dynamics (MD)[58] simulation methods were used to consider the binding modes of A, Hx, εA, and 1,N^6-ethanoadenine (EA). Although these studies provided useful information about the

catalytic mechanism or key interactions with AAG for specific substrates, information about differences in the active site conformation upon binding was not explicitly obtained for a range of nucleotides that span the chemical diversity of AAG substrates due to the computational models employed (i.e., ONIOM models based on a single (static) crystal structure or constraints applied to the DNA and protein backbones) or the structural similarity of the substrates considered (i.e., εA and EA can both interact with H136). Therefore, more work is required in order to understand the function of AAG.

To explain the diverse yet discriminatory substrate selectivity of AAG, the present study employs molecular dynamics simulations to evaluate how changes in the chemical structure of bound nucleotides affect the AAG active site conformation and key active site interactions. Specifically, seven nucleotides were examined (Figure 2.1) that are relevant to AAG including neutral substrates (εA, Hx), cationic substrates (3MeA, 7MeG), an inhibitor (εC) and the canonical purines (A and G). By uniquely considering a large range of nucleotides and monitoring active site reorganization following DNA–enzyme complex formation, our study provides the first structural explanations for how AAG can achieve broad substrate diversity, and simultaneously avoid excising the natural purines. Furthermore, each class of nucleotides considered yields important clues regarding the roles played by active site residues, including water, and reveals that these roles can significantly vary depending on the substrate bound. In fact, subtle deviations in active site interactions provide insight into experimentally observed differences in catalytic efficiencies towards excision of various substrates. In addition to affording a more thorough understanding of this critical DNA repair enzyme, our study contributes more broadly to the role that active site reorganization and placement of active site water plays in substrate or inhibitor recognition and binding, and therefore is complementary to previous work on a wide variety of other enzymes including lanthipeptide synthetases,[59] metalloenzymes,[60] aminoacyl tRNA synthetases,[61] and both DNA and RNA polymerases.[62, 63]

2.2 Computational Methods

Two X-ray crystal structures of AAG available in the Protein Data Bank (PDB) were used to initiate MD simulations, namely Δ79AAG, which contains an active site E125Q mutation bound to the εA substrate (PDB ID: 1EWN),[35] and Δ80AAG, which contains the enzyme bound to the εC inhibitor (PDB ID: 3QI5).[31, 35] Both structures were chosen because of their high resolution (2.1 – 2.2 Å), and the fact that the enzyme is bound to a lesion with an intact glycosidic bond (i.e., can be used to probe substrate binding). Both Δ79 and Δ80AAG (henceforth referred to as AAG) were truncated for ease of crystallization, but have catalytic power equivalent to AAG. Although 1EWN contains a single AAG E125Q-εA complex in the unit cell, 3QI5 includes two AAG–εC complexes, with the complex containing the least number of unresolved residues used for our study. 1EWN has unresolved density for residues 80–81, 200–207, 249–254, and 295–298, while 3QI5 has unresolved density for residues 205–206, 265–266, and 294–298. Overlays of the two complexes were used to assign the position and orientation of residues resolved in only one of the structures, while the backbone was added by inspection when a residue was unresolved in both structures. This approach is justified since all unresolved residues are remote from the active site (> 10 Å separation). The LEAP module of AMBER 10 was used to add the unresolved side chains and hydrogen atoms to all residues.[64] The locations of the added side chains were manually inspected using PyMol, and altered to favour hydrogen-bonding interactions and minimize steric clashes. The E125Q mutation in 1EWN was reverted to generate the wild-type enzyme. To consider the impact of different nucleotides, other DNA–protein complexes were generated by replacing εA in the structure generated from 1EWN with Hx, A, G, 3MeA, or 7MeG.

Each DNA–protein system was assigned AMBER parm99SB parameters[65] supplemented with GAFF parameters[66, 67] for non-standard nucleotides. Additionally, Restricted Electrostatic Potential (RESP) charges were assigned to each non-standard nucleoside using the R.E.D.v.III.4 program[68] and

B3LYP/6-31G(d) optimized geometries obtained with Gaussian 09 (Revision C.01 or D.01).[69] Each DNA–protein system was neutralized with Na^+ counter ions and solvated in an explicit TIP3P water box, which ensured that the DNA–protein solute was no closer than 8.0 Å to the boundary of the water box in any direction,[64, 67] and the particle mesh Ewald method was used with an electrostatic cut-off of 12.0 Å. Using a 2 fs time step, the solvent and ions were minimized for 500 steps of steepest descent and 500 steps of conjugate gradient minimization, while applying a 500.0 kcal mol^{-1} $Å^{-2}$ constraint to the protein and DNA. Subsequently, the system was minimized using 1000 steps of steepest descent, followed by 1500 steps of conjugate gradient optimization. The system was then heated to 300 K over 20 ps with a 10 kcal mol^{-1} $Å^{-2}$ restraint placed on the solute. Finally, the entire system was simulated without constraints under NPT conditions (1 atm, 300 K) for 40 ns. All minimization and simulation steps were performed with SANDER and PMEMD,[64, 70] and subsequent analysis was completed using the cpptraj module. Free energy calculations were performed using the molecular mechanics generalized Born surface area (MM-GBSA) method available in AMBER tool.[64] The reported pairwise energies were calculated between each active site amino acid and the bound nucleotide.

To monitor changes in the active site structure, the average root-mean-square deviation (RMSD) and associated standard deviation (σ) in the active site across two simulation trajectories or a single trajectory relative to a static crystal structure were calculated using the position of the heavy atoms in both the side chains and backbones of the bound nucleotide (dX), and active site (E125, Y127, A135, H136, Y159, N169, L180, and R182). Specifically, RMS fitting was performed based on the position of these residues to yield the overall RMSD of the active site, and subsequently the RMSD per residue was determined without refitting. All reported distances were measured and averaged throughout each simulation trajectory. The reported intermolecular separations between nucleobase and amino acid rings correspond to the distances between the center of masses of the aromatic moieties. Structures reported in

the main text were obtained by clustering the position of the bound nucleotide (dX) and the active site residues, and choosing a representative structure from the cluster with the greatest occupancy. Although the reported representative structures yield static pictures, we report the corresponding dynamic data across each trajectory in the Supporting Information. To determine the distribution of water in the active site, a three-dimensional 20 × 20 ×20 Å grid was centered on the active site with 0.50 Å spacing between each point on the grid using Amber Tools.[62] Spheres of different colors are shown on representative structures to reflect the percentage of each trajectory that an oxygen atom of a water molecule occupied a (0.50 × 0.50 × 0.50 Å) grid space, with gray spheres representing 40% occupancy, yellow spheres 60% occupancy, and red spheres 80% occupancy.

2.3 Results

2.3.1 AAG adopts a consistent active site conformation upon binding a range of neutral lesions

MD simulations initiated from X-ray crystal structures with either εA or εC bound to AAG yield active site conformations consistent with the crystallographic data (Figure A.1 in Appendix A). Specifically, the average RMSD across the simulation trajectory relative to the corresponding crystal structure is less than 0.8 Å (Table A.1 in Appendix A). The average RMSD is largest for εA due to reversion of the experimental E125Q mutation in the MD simulation, which yields a E125 position consistent with that observed in the crystal structure of the wild-type enzyme bound to εC (Figure A.1 in Appendix A). When comparing the simulated structures with AAG bound to εA, εC or Hx, the active site conformation does not change significantly regardless of the nucleotide bound (Figure 2.3A), with the average RMSD relative to bound εA being 0.981 Å for Hx and 1.433 Å for εC (Table A.2 in Appendix A). The average RMSD highlights the similarity in the binding of the neutral substrates εA and Hx, while the RMSD is larger in the case of the εC inhibitor because the lesion is further inserted into the active site as previously discussed based on experimental crystal structures.[31, 35] Nevertheless, the relative orientations of the aromatic (Y127,

H136, and Y159) residues and the bound lesions do not change substantially (key torsional angles are within 30°; Table A.3 in Appendix A and Figure 2.3). Furthermore, the relative orientation of the E125 general base with respect to the bound lesion is consistent (within 7°; Table A.3 in Appendix A and Figure 2.3A), and a hydrogen bond is maintained in the E125–Y127 catalytic dyad (> 99% occupancy; Table A.4 in Appendix A) regardless of the neutral lesion considered. However, the average distance between E125 and C1′ deviates by up to 0.830 Å (Table A.5 in Appendix A), being the longest for εA (4.758 Å) and shortest for Hx (3.928 Å). Furthermore, the placement of water within the active site deviates when the neutral lesions are bound to AAG (Figure A.2A-C in Appendix A), with a distinct lack of water between the substrate and the E125 general base for εC. In contrast, there is substantial water density near E125 (> 100% occupancy of the E125(Oε)···H_2O(OH) hydrogen bond; Table A.4 in Appendix A), which could correspond to the nucleophile, when either εA or Hx is bound to AAG. The absence of nucleophilic water may be one factor in εC inhibition of AAG activity.

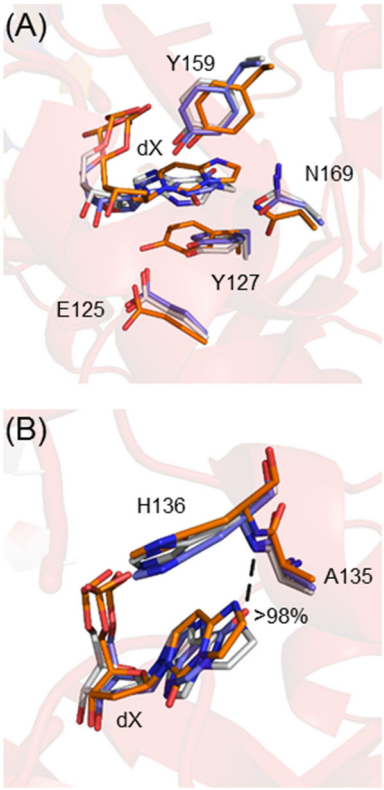


Figure 2.3. Overlay of representative MD structures of AAG bound to εA (gray), εC (orange), or Hx (blue) highlighting (A) the consistent active site conformation for the neutral lesions, and (B) the common hydrogen bond between the H136 backbone and the bound nucleobase.

Despite AAG lacking many direct contacts with the bound lesion, it has been noted based on experimental crystal structures that both εC and εA have a hydrogen-bond acceptor (N4 for εC; N6 for εA; Figure 1.1) well positioned to interact with the NH backbone of H136 (Figure 2.2B).[31, 35] Overlays of representative MD structures obtained with εA, εC, or Hx bound to AAG reveal that O6 of Hx adopts an equivalent acceptor position (Figure 2.3). Indeed, the dX(N2/N4/O6)···H136(NH) hydrogen bond is occupied for over 98% of the simulation trajectory corresponding to each neutral lesion (Table A.4 in Appendix A), and in combination with a dX(O1P)···H136(NεH) interaction, contributes between –29.3 and –41.8 kJ mol^{-1} to the binding free energy between the nucleotide and AAG (Table A.6 in Appendix A). Additionally, the contact with H136 leads to the observed further insertion of εC into the active site because of the smaller distance between the hydrogen-bond acceptor (N2/N4/O6) and the glycosidic bond

(N1/N9) in εC compared to εA and Hx (Figure 2.3A). Other interactions also occur for at least 80% of the simulations, namely the dX(O1P)···H136(NεH) hydrogen bond when εA or εC is bound, and the dX(O3)···R182(NηH) hydrogen bond when Hx is bound (Table A.4 in Appendix A), which further facilitates nucleotide binding by contributing up to –50.2 kJ mol^{-1} to the binding free energy (Table A.6 in Appendix A). However, when εC or Hx is bound to AAG, an additional hydrogen bond forms between the damaged nucleobase and an active site residue (Table A.4 in Appendix A). Specifically, N169 (NεH) hydrogen bonds with O2 of εC (44.8% occupancy and binding energy of –11.4 kJ mol^{-1}) and Y159 (OH) hydrogen bonds with N3 of Hx (see Appendix A; 73.1% occupancy; Table A.4 and binding energy of –10.8 kJ mol^{-1}; Table A.6 and Figure A.3). Despite the formation of a unique hydrogen bond with εC and Hx, the active site conformation adopted by AAG is nearly equivalent when bound to any of these neutral lesions. Since the common hydrogen-bond contact between each damaged nucleobase and H136 likely plays a significant role in stabilizing the consistent active site configuration, it is worthwhile to examine the structural consequences when nucleotides that lack the ability to hydrogen bond with the H136 backbone bind to AAG.

2.3.2 Reorganization of the AAG active site, including increased solvation, occurs upon binding of adenine and likely prevents excision

When A is bound to AAG, the nucleotide cannot form a hydrogen bond with the H136 backbone, since the N6 exocyclic amino group occupies the position of the hydrogen-bond acceptor in the neutral lesions (Figure 2.1). Since the neutral lesions induce a similar AAG active site conformation, we focus our comparison on the binding of A and Hx due to their similar size, and key difference of the C6 substituent. Overlays of representative MD structures with A and Hx bound to AAG (Figure 2.4A) reveal that the relative orientations of E125, Y127, and L180 with respect to the nucleotide are consistent (Figure 2.4A). Specifically, the average RMSD for each of the aforementioned residues is < 0.583 Å (Table A.2 in

Appendix A) and key torsional angles deviate by less than 5˚ (Table A.3 in Appendix A), with a slightly larger (20˚) difference in φ of L180. As discussed for the neutral lesions, the E125(Oε2)···Y127(OH) hydrogen bond is maintained, which contributes to the persistent active site conformation in this region (> 99.6% occupancy; Table A.4 in Appendix A).

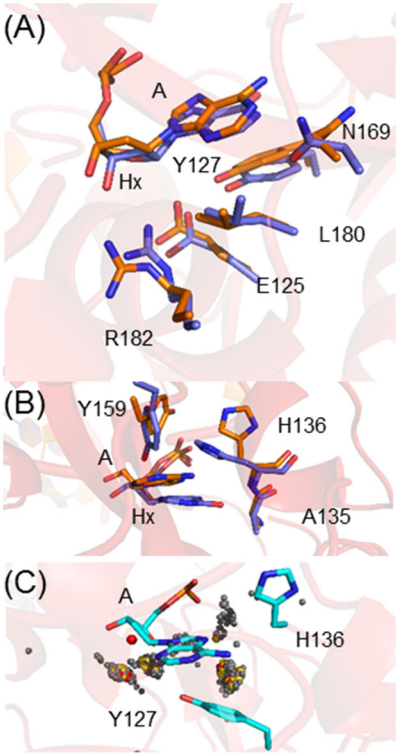


Figure 2.4. Overlay of representative MD structures of AAG bound to Hx (blue) or A (orange) highlighting (A) the similar position of the nucleotides with respect to several key active site residues, and (B) the discrepancy in the H136 and Y159 orientations. (C) Distribution of the active site water during the MD simulation of AAG bound to A (red: 80% occupied; yellow: 60% occupied; gray: 40% occupied).

Despite the similarities discussed above, the average RMSD of the AAG active site with A or Hx bound (RMSD = 2.072 Å) is larger than that between the neutral lesions (Table A.2 in Appendix A). Indeed, the nucleotides adopt different positions in the active site (RMSD of dX is 3.297 Å; Table A.2 in Appendix A and Figure 2.4B). The deviations in the A and Hx bound structures also occur in H136 (RMSD = 2.137 Å) and Y159 (RMSD = 2.404 Å). Additionally, key torsional angles in H136 and Y159 differ by up to 80° depending on the nucleotide bound (Table A.3 in Appendix A). The changes in the backbone

torsional angles result in the Y159 hydroxyl group being directed away from the nucleobase (Figure 2.4B), which is a departure from the Y159 hydroxyl group interacting with N3 of Hx, or the phosphate of εA or εC (Figure 2.3). Furthermore, due to a steric clash between the amino group of A and the backbone N–H of H136, the H136 side chain falls on average 6.325 Å away from the A nucleobase (Figure 2.4B), which is much greater than the distance between H136 and Hx (4.101 Å; Table A.5 in Appendix A). This shift creates a void between A and H136, which is occupied by at least one water molecule for at least 40% of the simulation (Figure 2.4C and A.2D, Appendix A). Water is also found between A and Y127 (Figure 2.4C), with at least two water molecules present between these groups for the duration of the simulation. This water distribution correlates with the approximately 1.0 Å greater average distance between Y127 and the nucleobase, and approximately 1.0 Å greater distance between E125 and C1′, for A compared to Hx (Table A.5 in Appendix A). The placement of active site water between the nucleobase and active site aromatic amino acids was not observed for the neutral lesions (Figure A.2A–C in Appendix A). As a result, the binding free energy between the nucleotide and H136 or Y127 decreases by up to 20.9 kJ mol^{-1} when A is bound to AAG compared to Hx (Table A.6 in Appendix A). We note that the difference in binding reported between A and Hx deviates from a previous computational study that used a constrained DNA–enzyme model and concluded these nucleotides are bound to AAG in a similar fashion,[50] which highlights the importance of considering active-site dynamics following substrate binding. Indeed, our results suggest that the changes in the active site conformation upon substrate binding may be one factor that prevents AAG-mediated excision of the canonical purines.

2.3.3 Guanine binding to AAG results in anti-catalytic active site interactions

To determine whether the conclusions deduced by considering the active site conformation adopted when A is bound within the AAG active site can be generalized to guanine, MD simulations were performed with G bound to AAG. Since the active site reorganization observed upon binding to A arises due to unfavourable interactions with the exocyclic amino group, the role of the amino group in the canonical purines will be further probed by comparing the AAG active site configuration when G or Hx is bound, since these nucleotides solely differ in the presence of the N2 moiety. The active site conformation is very similar regardless of whether G or Hx is bound to AAG (RMSD = 1.647 Å; Table A.2 in Appendix A and Figure 2.5A), with key torsional angles deviating by less than approximately 25° (Table A.3 in Appendix A). Similar to the neutral lesions, the O6 carbonyl of G forms a hydrogen bond with the NH backbone of H136 (> 99% occupancy; Table A.4 in Appendix A). Indeed, the binding free energy between G and H136 (–40.2 kJ mol^{-1}) is similar to that for the neutral lesions (–32.6 to –43.9 kJ mol^{-1}; Table A.6 in Appendix A). Furthermore, there is no active site water intervening between the G nucleobase and the aromatic residues (Figure A.2E in Appendix A).

The major differences in the conformation of the AAG active site when bound to G versus Hx arises in N169 and E125 (Figure 2.5A). Specifically, a significant RMSD for N169 (1.377 Å; Table A.2 in Appendix A) arises due to a hydrogen bond between the N2 amino group of G and the carbonyl of N169 (> 99% occupancy; Table A.4 in Appendix A and Figure 2.5B). This interaction is likely anti-catalytic, since it will destabilize the anionic charge forming on G during the deglycosylation reaction. In the case of E125, binding of G induces a 1.9 Å shift of E125 away from the (C1′) reaction center, which results in a larger E125(Oε)···dX(C1′) distance than seen for any other nucleotide examined (Table A.5 in Appendix A and Figure 2.5B), and is likely unfavourable for catalysis. Furthermore, the hydrogen bond between the side chains of E125 and Y127 that is consistently maintained for all other nucleotides (> 99% occupancy) is

absent when G is bound to AAG (< 1% occupancy; Table A.4 in Appendix A). As a result, the nucleophilic water shifts away from the (C1′) reaction site (Figure 2.5C). Specifically, while water resides between E125 and C1′ when all other nucleotides are bound to AAG (Figure A.2A–D, Appendix A), and therefore is ideally positioned to act as the nucleophile, the closest water to C1′ resides between and interacts with the side chains of E125 (Oε) and Y127 (OH; Figure A.2E in Appendix A) when G is in the AAG active site. The combined effects of the interaction between the amino group of G and N169, changes in the conformation of E125, and shifts in the location of active site water likely impede G excision by AAG.

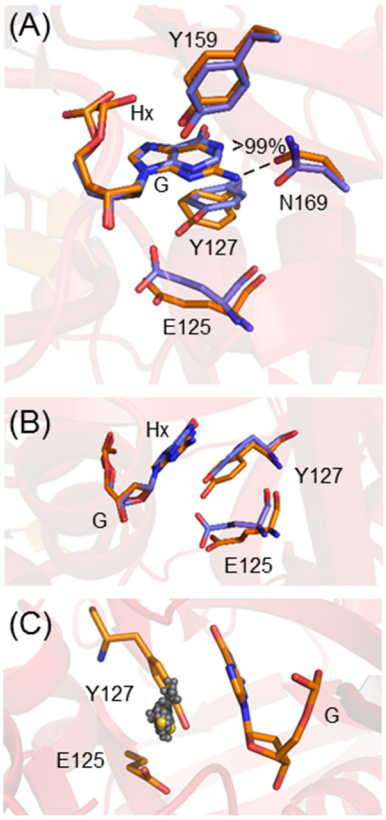


Figure 2.5. Overlay of representative MD structures of AAG bound to Hx (blue) or G (orange) highlighting (A) the similarity of the AAG active site conformation for several key active site residues, and (B) the different Y127 and E125 orientations. (C) Distribution of the active site water during the MD simulation of AAG bound to A (red: 80% occupied; yellow: 60% occupied; gray: 40% occupied).

2.3.4 Despite structural similarity to the canonical purines, methylated lesions are better accommodated in the AAG active site

Although the N6 exocyclic amino group of A disrupts key interactions between the nucleobase and the AAG active site, 3MeA possesses the same substituent, and yet AAG exhibits a significantly higher rate enhancement for removal of this lesion relative to A.[23] Comparison of the AAG active site conformation with 3MeA and A bound reveals a consistent orientation of E125 (Table A.3 in Appendix A and Figure 2.6A), including a hydrogen bond to Y127 (> 99% occupancy; Table A.4 in Appendix A). However, the average distance between E125 and the (C1′) reaction center is greater for 3MeA (by on average 0.533 Å; Table A.5 in Appendix A). This occurs since 3MeA is less inserted into the AAG active site (Figure 2.6A) due to steric repulsion between the N3 methyl group and the L180 side chain. This altered nucleotide position ensures that the steric clash between the A amino group and the H136 backbone does not occur for 3MeA. As a result, the average RMSD between the AAG active site conformations with 3MeA or A bound is 2.592 Å (Table A.2 in Appendix A). The active site residue that most significantly contributes to the large RMSD is H136 (2.266 Å; Table A.2 in Appendix A), with the relative orientation of the side chain (χ_1) differing by approximately 100° (Table A.3 in Appendix A and Figure 2.6A) and the aromatic amino acid being on average closer to 3MeA (5.224 Å) than A (6.325 Å; Table A.5 in Appendix A). Due to these differences in the active site conformation, water does not reside between 3MeA and the aromatic residues (Figure A.2F in Appendix A). Therefore, the contribution of H136 to the binding free energy is greater for 3MeA (–18.0 kJ mol^{-1}) than A (–11.3 kJ mol^{-1}; Table A.6 in Appendix A). Thus, despite the presence of the N6 amino group, L180 favourably positions the cationic 3MeA lesion in the AAG active site for base excision.

As discussed for 3MeA, the methyl group of 7MeG ensures that the lesion is less inserted into the AAG active site than G (Figure 2.6B). Specifically, a steric clash between the methyl substituent and the

H136 side chain prevents 7MeG from adopting the position equivalent to G. Nevertheless, 7MeG and G both form favourable hydrogen bonds with the H136 backbone and N169 side chain (> 99% occupancy; Table A.4 in Appendix A) and binding free energies of up to –41.8 kJ mol^{-1} (Table A.6 in Appendix A). Furthermore, the aromatic active site residues (H136, Y127, and Y159) adopt consistent orientations relative to the nucleobase (Table A.2 in Appendix A and Figure 2.6B) and key backbone torsional angles are within 20° for both nucleotides (Table A.3 in Appendix A). These similarities lead to a small RMSD in the AAG active site conformation when 7MeG or G is bound (0.936 Å; Table A.2 in Appendix A). However, a crucial difference in the AAG active site conformation is the maintenance of the hydrogen bond in the E125–Y127 catalytic dyad for 7MeG (> 99% occupancy; Table A.4 in Appendix A). Furthermore, water is suitably located between E125 and C1′ for nucleophilic attack when 7MeG is bound to AAG (Figure A.2G in Appendix A). Thus, despite the similar position of the N2 amino group in 7MeG and G, the methyl groups in the damaged nucleobases exhibit repulsion with an active site residue (H136), which thereby favourably positions the cationic lesions for catalytic excision by AAG.

2.4 Discussion

The major goal of the present study was to examine the manner in which a number of different nucleotides can be bound within the active site of AAG, and thereby uncover the structural basis for the ability of AAG to excise a structurally diverse set of substrates, while discriminating against the canonical purines. In this light, we examined seven different nucleotides including two neutral substrates (εA and Hx), an inhibitor (εC), the canonical purines (A and G), and two cationic substrates (3MeA and 7MeG). Although previous literature suggests that the base-flipping step is one factor contributing to the substrate diversity of AAG,[23, 24] how the substrates are accommodated within the active site is another factor that determines whether catalysis is possible. Thus, our model provides valuable structural information regarding how different nucleotides bind when forced into the active site. This information is critical due

to the role of other factors in determining the observed AAG substrate diversity and differential catalytic rates, including the chemical step.[23, 24] Our results suggest that active site reorganization upon substrate binding allows the enzyme to exploit water and several active site residues, as well as water, to achieve broad yet discriminatory substrate selectivity and diverse catalytic efficiency.

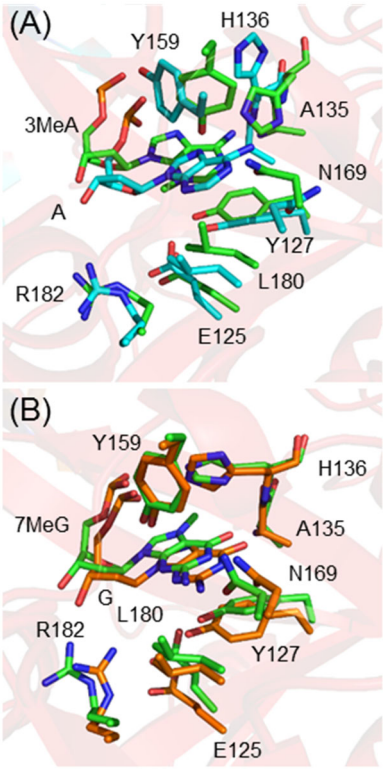


Figure 2.6. Overlays of representative MD structures of AAG bound to (A) 3MeA (green) or A (light blue), or (B) 7MeG (green) and G (orange).

2.4.1 Despite similar active site configurations, subtle disparities in DNA–protein interactions lead to different catalytic efficiencies toward the excision of neutral substrates

A crystal structure of εA bound to AAG (PDB ID: 1EWN) highlights a close contact between N6 of the nucleobase and the NH backbone of H136.[35] Furthermore, superposition of Hx onto εA led to speculation that Hx could also adopt a position in the AAG active site to form an analogous hydrogen bond between O6 of the nucleobase and H136,[23] which was more recently confirmed with Monte Carlo simulations using a constrained DNA–enzyme model.[50] Similarly, a hydrogen bond to H136 is maintained for both εA and Hx throughout our unconstrained MD simulations (Table A.4 in Appendix A). More importantly, our simulations reveal that this interaction results in a consistent orientation of the H136 side chain, as well as Y127 and Y159, relative to the nucleobase regardless of whether εA or Hx is bound to AAG (Tables A.2 and A.3 in Appendix A and Figure 2.3). The uniform relative positioning of the substrate and aromatic amino acids is likely critical for the excision of neutral substrates. Indeed, previous computational work has shown that π–π interactions facilitate removal of εA by AAG.[36] Furthermore, the strength of DNA–protein π–π interactions is comparable to biologically-relevant hydrogen bonds, yet highly dependent on the proximity and orientation of the DNA and protein moieties.[36-38] Other known neutral AAG substrates, including 1,N^2-ethenoguanine and EA,[30] also have a hydrogen–bond acceptor ideally positioned to form an interaction with the H136 backbone, and therefore will likely result in a similar AAG active site conformation upon binding. Indeed, previous computational work illustrates that EA binds to AAG in a manner comparable to εA.[58]

Regardless of the analogous AAG active site conformation upon binding of εA and Hx, experimental studies have reported that AAG binds εA more tightly than Hx, yet excises Hx at a higher rate.[23] Our MD simulations provide a molecular explanation for this intriguing observation in two ways. First, a hydrogen bond forms between the exocyclic hydroxyl group of Y159 and N3 of Hx that is not

observed for εA (Figure 2.3B). This interaction will facilitate Hx excision by stabilizing the negative charge accumulating on the nucleobase upon deglycosylation. Indeed, hydrogen bonding has been shown to be used by other DNA glycosylases in lieu of protonation to stabilize the departing nucleobase.[18, 26, 32] In contrast, Y159 forms a hydrogen bond to the DNA phosphate backbone when εA is bound to AAG (Table A.4 in Appendix A). The stronger interactions with the negatively charged phosphate in the case of εA (Tables A4 and A6 in Appendix A) contributes to the observed stronger binding of εA versus Hx, but will not assist base excision. Second, E125 is closer to the reaction center (C1′) for Hx than εA (Table A.5 in Appendix A). A conserved E or D residue has been shown to be important for base excision facilitated by other DNA glycosylases, including hUNG2,[51, 71] MutY,[25, 52] and hOgg1,[54, 72] with the proposed role being stabilization of the positive charge developing on the sugar during base departure. Indeed, E125Q and E125A AAG mutants lack catalytic activity.[22] Therefore, the closer proximity of E125 to the reaction center when Hx is bound likely also contributes to the observed greater catalytic rate enhancement for this substrate. Overall, despite neutral lesions binding in a similar fashion to AAG, subtle differences in active site interactions due to varied nucleotide composition play a role in dictating relative binding strengths and catalytic efficiencies.

2.4.2 In addition to a key active site hydrogen bond, redistribution of water upon εC binding leads to inhibition of AAG

Experimental studies have attributed the observed two-fold stronger binding of εC over εA to a hydrogen bond between N169 and O2 of εC.[31, 73] Although this is supported by the N169L and N169A mutants exhibiting decreased affinity for the εC lesion,[73] the only structural support for this proposal came from a single static crystal structure of εC bound to the E125Q AAG mutant. Our MD simulations confirm that this hydrogen bond is significantly populated even when the dynamics of the active site are considered (see Appendix A; 45% occupancy; Table A.4 and Figure A.2A), highlighting its importance for binding.

Furthermore, the hydrogen bond between the nucleobase and the backbone of H136 and the relative orientations of the aromatic amino acids discussed for the neutral substrates are maintained when εC is bound to AAG (Figure 2.3), which further underscores the role of these residues in binding neutral nucleotides. The inhibitory behaviour of εC has been attributed to the lack of nucleobase activation due to the absence of a proton acceptor equivalent to the N7 position of εA.[31] However, experimental and computational studies indicate that protonation at the N7 site may not be required for glycosidic bond cleavage in εA[36, 39] or Hx,[39] and neither A nor G are excised appreciably by AAG despite possessing an N7 acceptor site.[23] Furthermore, although the mechanism for εC excision was not explicitly considered, previous computational work speculated that poor nucleophile activation could be a contributing factor to inhibition based on anticipated similarities to the mapped A excision mechanism.[36] In contrast to these proposals, we attribute the inhibitory behaviour of εC to the lack of water near the general base (E125) that could adopt the role of the nucleophile. Indeed, E125 interacts with water for less than 7.0% of the simulation time when εC is bound to AAG (Table A.4 and Figure A.3 in Appendix A). Our new proposal for the role of water reorganization in AAG inhibition complements literature highlighting the diverse function of water in substrate or inhibitor binding for several other enzymes, including kinases,[74] amino acid biosynthetic enzymes,[75] and receptor proteins.[76] Active site water reorganization may also explain AAG inhibition by other pyrimidines, such as 3-methyluracil, 3-methylthymine, and 3-ethyluracil,[30] and should be explored further.

2.4.3 Interactions with the nucleobase amino groups in conjunction with redistribution of key active site residues, including water, allows AAG to discriminate against the canonical purines

For the first time, our work clarifies how AAG uses discrete active site interactions with the amino groups of A and G to discriminate against the undamaged purines. In both cases, binding of the canonical purine to AAG results in reorganization of the active site, including water. In the case of A, repulsion occurs

between the amino group and the backbone of H136, which has been previously proposed based on superposition of A onto εA in a resolved crystal structure of AAG.[23] Our MD simulations reveal that a consequence of this repulsion is alteration of the nature of nucleobase–aromatic amino acid (Y127, H136, and Y159) π–π interactions. Specifically, the relative orientation of the aromatic amino acids with respect to the nucleobase changes, including an increased separation distance (Table A.5 in Appendix A). This observation is consistent with reported disorder in the location of A135 and H136 in the absence of a bound nucleotide,[41] which suggests that at least part of the AAG active site conformation is determined upon nucleotide binding. As a result of these structural changes, water occupies the space between A and the aromatic amino acids. Although the catalytic effects of this water are unknown, the water disrupts the DNA–protein π–π interactions that are critical for base excision.[36]

In the case of G, although it has been proposed that the nucleobase amino group sterically clashes with N169,[77, 78] our simulations reveal a strong hydrogen bond (–24.4 kJ mol^{-1}; Table A.6 in Appendix A) between the N2 amino group of G and the carbonyl of N169 over the entire MD simulation (100% occupancy; Table A.4 in Appendix A and Figure 2.5). However, this interaction is unfavourable for base excision since acting as a hydrogen-bond donor will decreases the ability of G to accommodate the negative charge developing on the nucleobase during deglycosylation. This proposal is supported by experimental evidence that N169S[23] and N169A[77] AAG mutants have increased activity toward G. Concomitant with the formation of the DNA–protein interaction, the AAG active site reorganizes upon binding of G (Figure 2.5). Specifically, the catalytic dyad between the general base (E125) and Y127 observed when all other nucleotides are bound to AAG is disrupted (Table A.4 in Appendix A). Consequently, E125 is positioned far from the (C1′) reaction site and does not adopt an orientation conducive for catalysis (Figure 2.5B), and active site water is no longer ideally positioned to act as a nucleophile (Figure 2.5C). Our findings provide the first direct evidence for the previous proposal that misalignment of the general base and water

nucleophile upon substrate binding may impede catalysis.[36] Furthermore, disruption of active site water upon binding of G mirrors our proposal that the lack of a suitable nucleophile contributes to AAG inhibition by εC. In summary, although reorganization of active site water at least in part prevents the excision of both canonical purines, AAG exploits different DNA–protein interactions to alter the position of the bound nucleotide and active site water to achieve this discrimination.

2.4.4 Despite equivalent amino groups as the canonical purines, the methylated lesions are positioned for excision by AAG using unique DNA–protein interactions

Although the amino groups in the canonical purines form unfavourable interactions with AAG that prevent their excision, methylated purine lesions (3MeA and 7MeG) are excised by AAG even though they retain the amino group of the natural purine equivalent.[23] Our results indicate that activity towards these damaged nucleotides is possible since repulsion between the methyl group on the nucleobase and an active site residue prevents AAG from adopting the same conformation as when the corresponding canonical purine is bound. Indeed, superposition of 3MeA onto εA in a resolved crystal structure of AAG led to speculation of a steric clash between the N3 methyl group and L180.[34] We provide direct evidence for this repulsive interaction. Furthermore, our simulations reveal that this contact causes the 3MeA nucleotide to be less inserted into the active site relative to A. Therefore, the L180 side chain allows AAG to be selective towards 3MeA by enhancing the interactions between the nucleobase and aromatic amino acids, which have been shown to be important for binding.[36] This new proposal is supported by a 2.2-fold reduction in activity toward 3MeA for the L180F AAG mutant,[79] which would exhibit increased repulsion with the N3 methyl group and thereby further prevent productive binding. Although the effect of reducing or eliminating the alkyl chain of L180 has yet to be examined, we propose that such mutations (e.g., alanine or glycine) will also reduce AAG activity by allowing 3MeA to adopt an equivalent position as A.

Analogously, our simulations lead to the new proposal that repulsion between the N7 methyl group and the H136 side chain prevents 7MeG from adopting the same position as G in the AAG active site (Figure 2.6). Indeed, the H136A AAG mutant exhibits reduced activity towards 7MeG.[22] Although this reduced activity is partially caused by the loss of DNA–protein π–π interactions, the H136A AAG mutant only exhibits a 16-fold reduction in the activity toward εA, but a 160-fold reduction in activity toward 7MeG,[22] indicating that H136 exhibits selectivity for the N7 methyl group. Despite 7MeG being less inserted into the AAG active site than G, N169 shifts to maintain the catalytically unfavourable hydrogen-bonding interaction with the (N2) amino group that is observed when G is bound to AAG. However, this interaction likely does not prevent excision due the inherently labile glycosidic bond of cationic lesions. Furthermore, the general base (E125) and active site water are ideally positioned to facilitate excision of 7MeG. Nevertheless, it is intriguing that AAG maintains the ability to discriminate against G at the detriment of forming a catalytically unfavourable interaction with 7MeG. This provides the second example of how AAG sacrifices catalytic efficiency for substrate diversity. Specifically, previous work emphasized that DNA–protein π–π interactions are catalytic towards the removal of neutral substrates, but are anti-catalytic toward the removal of inherently less stable cationic lesions.[36] Most importantly, although AAG exploits unique DNA–protein interactions for each methylated lesion, a common theme emerges of the enzyme using the additional methyl groups to correctly position cationic alkylated lesions for excision.

2.5 Conclusion

By exploring the active site dynamics following binding of a diverse set of nucleotides that span neutral and charged substrates, inhibitors, and canonical purines, we have identified several factors that contribute to the broad yet discriminatory substrate selectivity of AAG. Based on the structural diversity of AAG substrates, it is not surprising that the enzyme uses specific active site residues to target each nucleotide. For example, different residues prevent the formation of a catalytically competent complex

when the canonical purines bind to AAG due to steric repulsion with the nucleobase amino group, while other residues repulse the methyl groups of alkylated lesions to correctly situate the damaged nucleotides for excision. Nevertheless, a conserved interaction with H136 is critical for AAG activity towards neutral lesions, which has the newly proposed role of optimizing DNA–protein π–π interactions that facilitate base excision. Interestingly, reorganization of active site water and the general base that activates the nucleophile is found to be vital for both enzyme inhibition and discrimination against the natural purines, which highlights the importance of water in nucleotide recognition. In summary, AAG uses the inherent flexibility of the active site to form a plethora of unique enzyme–nucleobase interactions with each bound nucleotide that either enhance or mitigate catalysis. Our findings will aid the understanding of the function of other enzymes that use flexible active sites and/or water to control substrate selectivity.

2.6 References

[1] Grando, S. A. (2014) Connections of Nicotine to Cancer, *Nat. Rev. Cancer 14*, 419-429.

[2] Gaillard, H., Garcia-Muse, T., and Aguilera, A. (2015) Replication Stress and Cancer, *Nat. Rev. Cancer 15*, 276-289.

[3] Kanvah, S., Joseph, J., Schuster, G. B., Barnett, R. N., Cleveland, C. L., and Landman, U. (2010) Oxidation of DNA: Damage to Nucleobases, *Acc. Chem. Res. 43*, 280–287.

[4] Sharma, S., Shrivastav, B. R., and Shrivastav, A. (2013) Free Radicals, Antioxidants and Oxidative Stress, *Int. J. Adv. Res. 1*, 252-258.

[5] Zheng, G., Fu, Y., and He, C. (2014) Nucleic Acid Oxidation in DNA Damage Repair and Epigenetics, *Chem. Rev. 114*, 4602-4620.

[6] Cadet, J., and Wagner, J. R. (2014) Oxidatively Generated Base Damage to Cellular DNA by Hydroxyl Radical and One-Electron Oxidants: Similarities and Differences, *Arch. Biochem. Biophys. 557*, 47-54.

[7] Dedon, P. C., and Tannenbaum, S. R. (2004) Reactive Nitrogen Species in the Chemical Biology of Inflammation, *Arch. Biochem. Biophys. 423*, 12–22.

[8] Elghissassi, F., Barbin, A., Nair, J., and Bartsch, H. (1995) Formation of 1,N-6-Ethenoadenille and 3,N-4-Ethenocytosine by Lipid-Peroxidation Products and Nucleic-Acid Bases, *Chem. Res. Toxicol. 8*, 278–283.

[9] Winczura, A., Zdzalik, D., and Tudek, B. (2012) Damage of DNA and Proteins by Major Lipid Peroxidation Products in Genome Stability, *Free Radical Res. 46*, 442-459.

[10] Sedgwick, B. (2004) Repairing DNA-Methylation Damage, *Nat. Rev. Mol. Cell Biol. 5*, 148-157.

[11] Pandya, G. A., and Moriya, M. (1996) 1,N6-Ethenodeoxyadenosine, a DNA Adduct Highly Mutagenic in Mammalian Cells, *Biochemistry 35*, 11487-11492.

[12] Nair, U., Bartsch, H., and Nair, J. (2007) Lipid Peroxidation-Induced DNA Damage in Cancer-Prone Inflammatory Diseases: A Review of Published Adduct Types and Levels in Humans, *Free Radical Biol. Med. 43*, 1109-1120.

[13] Ebrahimkhani, M. R., Daneshmand, A., Mazumder, A., Allocca, M., Calvo, J. A., Abolhassani, N., Jhun, I., Muthupalani, S., Ayata, C., and Samson, L. D. (2014) Aag-Initiated Base Excision Repair Promotes Ischemia Reperfusion Injury in Liver, Brain, and Kidney, *Proc. Natl. Acad. Sci. USA 111*, E4878-E4886.

[14] Monti, P., Traverso, I., Casolari, L., Menichini, P., Inga, A., Ottaggio, L., Russo, D., Iyer, P., Gold, B., and Fronza, G. (2010) Mutagenicity of N3-Methyladenine: A Multi-Translesion Polymerase Affair, *Mutat. Res., Fundam. Mol. Mech. Mutagen. 683*, 50-56.

[15] Boysen, G., Pachkowski, B. F., Nakamura, J., and Swenberg, J. A. (2009) The Formation and Biological Significance of N7-Guanine Adducts, *Mutat. Res. Genet. Toxicol. Environ. Mutagen. 678*, 76–94.

[16] Stivers, J. T., and Jiang, Y. L. (2003) A Mechanistic Perspective on the Chemistry of DNA Repair Glycosylases, *Chem. Rev. 103*, 2729–2759.

[17] Berti, P. J., and McCann, J. A. B. (2006) Toward a Detailed Understanding of Base Excision Repair Enzymes: Transition State and Mechanistic Analyses of N-Glycoside Hydrolysis and N-Glycoside Transfer, *Chem. Rev. 106*, 506–555.

[18] Zharkov, D. O. (2008) Base Excision DNA Repair, *Cell. Mol. Life Sci. 65*, 1544–1565.

[19] Drohat, A. C., and Maiti, A. (2014) Mechanisms for Enzymatic Cleavage of the N-Glycosidic Bond in DNA, *Org. Biomol. Chem. 12*, 8367-8378.

[20] Brooks, S. C., Adhikary, S., Rubinson, E. H., and Eichman, B. F. (2013) Recent Advances in the Structural Mechanisms of DNA Glycosylases, *BBA-Proteins Proteomics 1834*, 247-271.

[21] Schermerhorn, K. M., and Delaney, S. (2014) A Chemical and Kinetic Perspective on Base Excision Repair of DNA, *Acc. Chem. Res. 47*, 1238-1246.

[22] O'Brien, P. J., and Ellenberger, T. (2003) Human Alkyladenine DNA Glycosylase Uses Acid-Base Catalysis for Selective Excision of Damaged Purines, *Biochemistry 42*, 12418–12429.

[23] O'Brien, P. J., and Ellenberger, T. (2004) Dissecting the Broad Substrate Specificity of Human 3-Methyladenine-DNA Glycosylase, *J. Biol. Chem. 279*, 9750–9757.

[24] Wolfe, A. E., and O'Brien, P. J. (2009) Kinetic Mechanism for the Flipping and Excision of 1,N-6-Ethenoadenine by Human Alkyladenine DNA Glycosylase, *Biochemistry 48*, 11357–11369.

[25] Lee, S., and Verdine, G. L. (2009) Atomic Substitution Reveals the Structural Basis for Substrate Adenine Recognition and Removal by Adenine DNA Glycosylase, *Proc. Natl. Acad. Sci. USA 106*, 18497–18502.

[26] Zharkov, D. O., Mechetin, G. V., and Nevinsky, G. A. (2010) Uracil-DNA Glycosylase: Structural, Thermodynamic and Kinetic Aspects of Lesion Search and Recognition, *Mutat. Res. Fundam. Mol. Mech. Mutag. 685*, 11–20.

[27] Kirpota, O. O., Endutkin, A. V., Ponomarenko, M. P., Ponomarenko, P. M., Zharkov, D. O., and Nevinsky, G. A. (2011) Thermodynamic and Kinetic Basis for Recognition and Repair of 8-Oxoguanine in DNA by Human 8-Oxoguanine-DNA Glycosylase, *Nucleic Acids Res. 39*, 4836–4850.

[28] Porello, S. L., Leyes, A. E., and David, S. S. (1998) Single-Turnover and Pre-Steady-State Kinetics of the Reaction of the Adenine Glycosylase Muty with Mismatch-Containing DNA Substrates, *Biochemistry 37*, 14756–14764.

[29] Kuznetsov, N. A., Koval, V. V., and Fedorova, O. S. (2011) Mechanism of Recognition and Repair of Damaged DNA by Human 8-Oxoguanine DNA Glycosylase Hogg1, *Biochem. Moscow 76*, 118–130.

[30] Lee, C.-Y. I., Delaney, J. C., Kartalou, M., Lingaraju, G. M., Maor-Shoshani, A., Essigmann, J. M., and Samson, L. D. (2009) Recognition and Processing of a New Repertoire of DNA Substrates by Human 3-Methyladenine DNA Glycosylase (AAG), *Biochemistry 48*, 1850-1861.

[31] Lingaraju, G. M., Davis, C. A., Setser, J. W., Samson, L. D., and Drennan, C. L. (2011) Structural Basis for the Inhibition of Human Alkyladenine DNA Glycosylase (AAG) by 3,N4-Ethenocytosine-Containing DNA, *J. Biol. Chem. 286*, 13205–13213.

[32] Parikh, S. S., Walcher, G., Jones, G. D., Slupphaug, G., Krokan, H. E., Blackburn, G. M., and Tainer, J. A. (2000) Uracil-DNA Glycosylase-DNA Substrate and Product Structures: Conformational Strain Promotes Catalytic Efficiency by Coupled Stereoelectronic Effects, *Proc. Natl. Acad. Sci. USA 97*, 5083–5088.

[33] Martin, F. H., Castro, M. M., Aboul-ela, F., and Tinoco, I. (1985) Base Pairing Involving Deoxyinosine: Implications for Probe Design, *Nucleic Acids Res. 13*, 8927-8938.

[34] Lau, A. Y., Scharer, O. D., Samson, L., Verdine, G. L., and Ellenberger, E. (1998) Crystal Structure of a Human Alkylbase-DNA Repair Enzyme Complexed to DNA: Mechanisms for Nucleotide Flipping and Base Excision, *Cell 95*, 249–258.

[35] Lau, A. Y., Wyatt, M. D., Glassner, B. J., Samson, L. D., and Ellenberger, T. (2000) Molecular Basis for Discriminating between Normal and Damaged Bases by the Human Alkyladenine Glycosylase, AAG, *Proc. Natl. Acad. Sci. USA 97*, 13573–13578.

[36] Rutledge, L. R., and Wetmore, S. D. (2011) Modeling the Chemical Step Utilized by Human Alkyladenine DNA Glycosylase: A Concerted Mechanism Aids in Selectively Excising Damaged Purines, *J. Am. Chem. Soc. 133*, 16258–16269.

[37] Wilson, K. A., Kellie, J. L., and Wetmore, S. D. (2014) DNA-Protein Pi-Interactions in Nature: Abundance, Structure, Composition and Strength of Contacts between Aromatic Amino Acids and DNA Nucleobases or Deoxyribose Sugar, *Nucleic Acids Res. 42*, 6726-6741.

[38] Wilson, K. A., Wells, R. A., Abendong, M. N., Anderson, C. B., Kung, R. W., and Wetmore, S. D. (2015) Landscape of π–π and Sugar–π Contacts in DNA–Protein Interactions, *J. Biomol. Struct. Dyn.*, 1-14.

[39] Sun, X., and Lee, J. K. (2007) Acidity and Proton Affinity of Hypoxanthine in the Gas Phase Versus in Solution: Intrinsic Reactivity and Biological Implications, *J. Org. Chem. 72*, 6548–6555.

[40] Liu, M., Xu, M., and Lee, J. K. (2008) The Acidity and Proton Affinity of the Damaged Base 1,N-6-Ethenoadenine in the Gas Phase Versus in Solution: Intrinsic Reactivity and Biological Implications, *J. Org. Chem. 73*, 5907–5914.

[41] Setser, J. W., Lingaraju, G. M., Davis, C. A., Samson, L. D., and Drennan, C. L. (2012) Searching for DNA Lesions: Structural Evidence for Lower- and Higher-Affinity DNA Binding Conformations of Human Alkyladenine DNA Glycosylase, *Biochemistry 51*, 382-390.

[42] Perlmutter, J. D., and Sachs, J. N. (2011) Interleaflet Interaction and Asymmetry in Phase Separated Lipid Bilayers: Molecular Dynamics Simulations, *J. Am. Chem. Soc. 133*, 6563-6577.

[43] Riccardi, D., Konig, P., Prat-Resina, X., Yu, H. B., Elstner, M., Frauenheim, T., and Cui, Q. (2006) "Proton Holes" in Long-Range Proton Transfer Reactions in Solution and Enzymes: A Theoretical Analysis, *J. Am. Chem. Soc. 128*, 16302-16311.

[44] Tao, P., Fisher, J. F., Shi, Q., Vreven, T., Mobashery, S., and Schlegel, H. B. (2009) Matrix Metalloproteinase 2 Inhibition: Combined Quantum Mechanics and Molecular Mechanics Studies of the Inhibition Mechanism of (4-Phenoxyphenylsulfonyl)Methylthiirane and Its Oxirane Analogue, *Biochemistry 48*, 9839-9847.

[45] Reblova, K., Sponer, J., and Lankas, F. (2012) Structure and Mechanical Properties of the Ribosomal L1 Stalk Three-Way Junction, *Nucleic Acids Res. 40*, 6290-6303.

[46] Pinto, G. P., Brás, N. F., Perez, M. A. S., Fernandes, P. A., Russo, N., Ramos, M. J., and Toscano, M. (2015) Establishing the Catalytic Mechanism of Human Pancreatic A-Amylase with QM/MM Methods, *J. Chem. Theory Comput. 11*, 2508-2516.

[47] Asami, S., Porter, J. R., Lange, O. F., and Reif, B. (2015) Access to C Alpha Backbone Dynamics of Biological Solids by C-13 T-1 Relaxation and Molecular Dynamics Simulation, *J. Am. Chem. Soc. 137*, 1094-1100.

[48] Pinto, G. P., Ribeiro, A. J. M., Ramos, M. J., Fernandes, P. A., Toscano, M., and Russo, N. (2015) New Insights in the Catalytic Mechanism of Tyrosine Ammonia-Lyase Given by QM/MM and QM Cluster Models, *Arch. Biochem. Biophys. 582*, 107-115.

[49] Ruiz-Barragan, S., Morokuma, K., and Blancafort, L. (2015) Conical Intersection Optimization Using Composed Steps inside the ONIOM(QM:MM) Scheme: CASSCF:UFF Implementation with Microiterations, *J. Chem. Theory Comput. 11*, 1585-1594.

[50] Alexandrova, A. N. (2010) Promiscuous DNA Alkyladenine Glycosylase Dramatically Favors a Bound Lesion over Undamaged Adenine, *Biophys. Chem. 152*, 118–127.

[51] Przybylski, J. L., and Wetmore, S. D. (2011) A QM/QM Investigation of the hUNG2 Reaction Surface: The Untold Tale of a Catalytic Residue, *Biochemistry 50*, 4218–4227.

[52] Brunk, E., Arey, J. S., and Rothlisberger, U. (2012) Role of Environment for Catalysis of the DNA Repair Enzyme MutY, *J. Am. Chem. Soc. 134*, 8608-8616.

[53] Kellie, J. L., Wilson, K. A., and Wetmore, S. D. (2013) Standard Role for a Conserved Aspartate or More Direct Involvement in Deglycosylation? An ONIOM and MD Investigation of Adenine-DNA Glycosylase, *Biochemistry 52*, 8753-8765.

[54] Sadeghian, K., and Ochsenfeld, C. (2015) Unraveling the Base Excision Repair Mechanism of Human DNA Glycosylase, *J. Am. Chem. Soc. 137*, 9824-9831.

[55] Blank, I. D., Sadeghian, K., and Ochsenfeld, C. (2015) A Base-Independent Repair Mechanism for DNA Glycosylase—No Discrimination within the Active Site, *Sci. Rep. 5*, 10369.

[56] Kellie, J. L., Wilson, K. A., and Wetmore, S. D. (2015) An ONIOM and MD Investigation of Possible Monofunctional Activity of Human 8-Oxoguanine-DNA Glycosylase (hOgg1), *J. Phys. Chem. B 119*, 8013-8023.

[57] Kanaan, N., Crehuet, R., and Imhof, P. (2015) Mechanism of the Glycosidic Bond Cleavage of Mismatched Thymine in Human Thymine DNA Glycosylase Revealed by Classical Molecular Dynamics and Quantum Mechanical/Molecular Mechanical Calculations, *J. Phys. Chem. B 119*, 12365-12380.

[58] Guliaev, A. B., Hang, B., and Singer, B. (2002) 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. 30*, 3778–3787.

[59] Thibodeaux, C. J., Ha, T., and van der Donk, W. A. (2014) A Price to Pay for Relaxed Substrate Specificity: A Comparative Kinetic Analysis of the Class II Lanthipeptide Synthetases ProcM and HalM2, *J. Am. Chem. Soc. 136*, 17513-17529.

[60] Kretz, C. A., Dai, M., Soylemez, O., Yee, A., Desch, K. C., Siemieniak, D., Tomberg, K., Kondrashov, F. A., Meng, F., and Ginsburg, D. (2015) Massively Parallel Enzyme Kinetics Reveals the Substrate Recognition Landscape of the Metalloprotease ADAMTS13, *Proc. Natl. Acad. Sci. USA 112*, 9328-9333.

[61] Liu, Z., Vargas-Rodriguez, O., Goto, Y., Novoa, E. M., Ribas de Pouplana, L., Suga, H., and Musier-Forsyth, K. (2015) Homologous Trans-Editing Factors with Broad tRNA Specificity Prevent Mistranslation Caused by Serine/Threonine Misactivation, *Proc. Natl. Acad. Sci. USA 112*, 6027-6032.

[62] Sohl, C. D., Szymanski, M. R., Mislak, A. C., Shumate, C. K., Amiralaei, S., Schinazi, R. F., Anderson, K. S., and Yin, Y. W. (2015) Probing the Structural and Molecular Basis of Nucleotide Selectivity by Human Mitochondrial DNA Polymerase Gamma, *Proc. Natl. Acad. Sci. USA 112*, 8596-8601.

[63] Wang, L. F., Zhou, Y., Xu, L., Xiao, R., Lu, X. Y., Chen, L., Chong, J., Li, H. R., He, C., Fu, X. D.; et al. (2015) Molecular Basis for 5-Carboxycytosine Recognition by RNA Polymerase II Elongation Complex, *Nature 523*, 621-+.

[64] Case, D. A., Darden, T. A., Cheatham, T. E., III, Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Crowley, M., Walker, R. C., Zhang, W.; et al. (2008) Amber Tools, Version 1.0 ed., University of California, San Francisco.

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

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

[67] Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules, *J. Am. Chem. Soc. 117*, 5179–5197.

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

[69] Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A.; et al. (2009) Gaussian 09, Revision A.02 ed., Gaussian, Inc., Wallingford CT.

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

[71] Kavli, B., Sundheim, O., Akbari, M., Otterlei, M., Nilsen, H., Skorpen, F., Aas, P. A., Hagen, L., Krokan, H. E., and Slupphaug, G. (2002) hUNG2 Is the Major Repair Enzyme for Removal of Uracil from U:A Matches, U:G Mismatches, and U in Single-Stranded DNA, with hSMUG1 as a Broad Specificity Backup, *J. Biol. Chem. 277*, 39926–39936.

[72] Norman, D. P. G., Bruner, S. D., and Verdine, G. L. (2001) Coupling of Substrate Recognition and Catalysis by a Human Base-Excision DNA Repair Protein, *J. Am. Chem. Soc. 123*, 359–360.

[73] Gros, L., Maksimenko, A. V., Privezentzev, C. V., Laval, J., and Saparbaev, M. K. (2004) Hijacking of the Human Alkyl-N-Purine-DNA Glycosylase by 3,N-4-Ethenocytosine, a Lipid Peroxidation-Induced DNA Adduct, *J. Biol. Chem. 279*, 17723-17730.

[74] Levinson, N. M., and Boxer, S. G. (2014) A Conserved Water-Mediated Hydrogen Bond Network Defines Bosutinib's Kinase Selectivity, *Nat. Chem. Biol. 10*, 127-132.

[75] Blanco, B., Sedes, A., Peón, A., Otero, J. M., van Raaij, M. J., Thompson, P., Hawkins, A. R., and González-Bello, C. (2014) Exploring the Water-Binding Pocket of the Type II Dehydroquinase Enzyme in the Structure-Based Design of Inhibitors, *J. Med. Chem. 57*, 3494-3510.

[76] Huang, W. J., Gherib, R., and Gauld, J. W. (2012) An Active Site Water Broadens Substrate Specificity in S-Ribosylhomocysteinase (LuxS): A Docking, MD, and QM/MM Study, *J. Phys. Chem. B 116*, 8916-8929.

[77] Connor, E. E., and Wyatt, M. D. (2002) Active-Site Clashes Prevent the Human 3-Methyladenine DNA Glycosylase from Improperly Removing Bases, *Chem. Biol. 9*, 1033–1041.

[78] Connor, E. E., Wilson, J. J., and Wyatt, M. D. (2005) Effects of Substrate Specificity on Initiating the Base Excision Repair of N-Methylpurines by Variant Human 3-Methyladenine DNA Glycosylases, *Chem. Res. Toxicol. 18*, 87–94.

[79] Chen, C. Y., Guo, H. W. H., Shah, D., Blank, A., Samson, L. D., and Loeb, L. A. (2008) Substrate Binding Pocket Residues of Human Alkyladenine-DNA Glycosylase Critical for Methylating Agent Survival, *DNA Repair 7*, 1731–1745.

Chapter 3: An QM/MM Study of the Reaction Catalyzed by Alkyladenine DNA Glycosylase: Examination of the Substrate Specificity of a DNA Repair Enzyme[a,b,c]

3.1 Introduction

Genetic information is stored within the sequence of nucleobases within DNA. Unfortunately, the DNA nucleobases (adenine (A), cytosine (C), guanine (G), and thymine (T)) can be chemically altered, which can change the encoded genetic information.[1, 2] A common damaging event involves reactions between reactive oxygen (ROS) or nitrogen (RNS) species and the nucleobases, which can generate mutations.[1-6] For example, nitric oxide radicals can deaminate A to yield hypoxanthine (Hx; Figure 3.1).[7] Although A forms Watson-Crick pairs with T, Hx preferentially pairs with C and therefore A deamination results in an A→G mutation after two rounds of DNA replication.[8] Nucleobase alkylation is another frequent type of DNA damage that can have negative cellular consequences.[9, 10] For example, nitrosamines can attack N7 of G to form 7-methylguanine (7MeG, Figure 3.1), which depurinates 10^6 times faster than G and leads to cytotoxic abasic sites within DNA duplexes.[11]

Fortunately, many pathways exist that can correct DNA damage,[5, 12-14] including the base excision repair (BER).[9, 15-19] BER is initiated by a DNA glycosylase, which scans DNA to locate modified nucleobases. Upon lesion identification, the DNA glycosylase flips the targeted nucleobase out of the helix and catalyzes the cleavage of the *N*-glycosidic bond linking the nucleobase to the sugar-phosphate backbone. DNA glycosylases have been traditionally divided into two types: 1) monofunctional glycosylases, which solely catalyze *N*-glycosidic bond hydrolysis, and 2) bifunctional glycosylases, which catalyze amine-mediated *N*-glycosidic bond cleavage, and scission of the 3′ and sometimes 5′

[a] *The Journal of Physical Chemistry B* reference style used in this Chapter.
[b] Reprinted with permission from Lenz, S.A.P. and Wetmore, S.D. An QM/MM Study of the Reaction Catalyzed by Alkyladenine DNA Glycosylase: Examination of the Substrate Specificity of a DNA Repair Enzyme, *J. Phys. Chem. B,* ***2017,*** *121,* 11096-11108.
[c] S.A.P.L. performed all calculations. S.A.P.L. and S.D.W. contributed to the design of calculations, interpretation of data, and writing the manuscript.

phosphodiester bonds.[9, 15-19] In human cells, monofunctional alkyladenine DNA glycosylase (AAG) identifies and removes a plethora of nucleobase lesions (Figure 3.1), including Hx and 7MeG, as well as 1,N^6-ethenoadenine (εA) and 3-methyladenine (3MeA).[20-22] This substrate diversity is unusual among monofunctional DNA glycosylases, although some bifunctional DNA glycosylases also possess broad substrate specificity (e.g., NEIL1[23-27] and FPG).[27, 28] AAG achieves diverse activity by sacrificing catalytic power, exhibiting catalytic rate enhancements (10^3 – 10^7)[20, 21] far below other glycosylases (10^{10} – 10^{26}).[21, 22, 29, 30] Despite broad substrate specificity, AAG does not excise damaged pyrimidines, with some pyrimidine lesions inhibiting enzyme function (e.g., 3,N^4-ethenocytosine, εC).[31] Furthermore, AAG exhibits little activity toward either A or G despite being able to bind both canonical purines.[21]

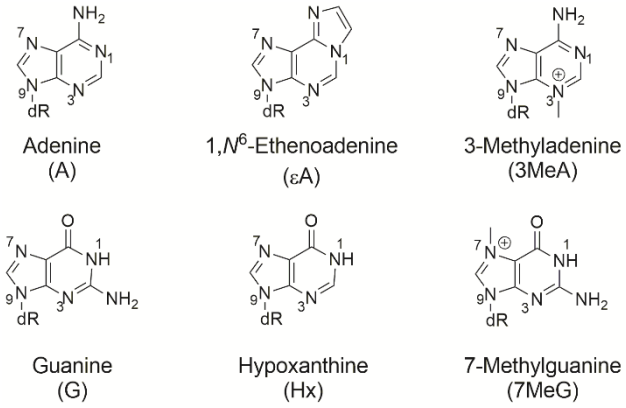


Figure 3.1. Structure and chemical numbering of select DNA lesions processed by AAG, as well as the canonical purines (A and G).

The specificity of AAG may be at least in part dictated by the base-flipping step. Specifically, AAG exhibits greater activity toward Hx when paired opposite T than C,[21] which can be attributed to a lower base-flipping barrier since Hx:T base pairs are less stable than Hx:C base pairs.[32] AAG also exhibits greater activity toward natural purines in less stable mismatched base pairs over canonical complementary pairs, albeit still several orders of magnitude less than the activity toward damaged DNA.[21] However, the AAG

excision rate for εA is not affected by the opposing base, being virtually identical for the equally unstable[33] εA:T and εA:C base pairs.[21] Therefore, while AAG activity can depend on the base paired with the lesion, which may affect the base-flipping step, other factors such as substrate binding and deglycosylation must also dictate the observed substrate specificity.

X-ray structures of AAG–DNA complexes suggest that AAG primarily forms non-specific π–π and T-shaped interactions with nucleobase lesions using aromatic active site residues (Y127, H136, and Y159),[31, 34, 35] with only one hydrogen bond identified between a substrate (i.e., N6 of εA) and the enzyme (i.e., the backbone of H136; Figure 3.2).[34] When coupled with the prevalence and strength of protein–DNA π–π stacking and T-shaped interactions in biology,[36-38] this suggests that π–interactions may be a catalytic driving force in the AAG-catalyzed reaction mechanism. Indeed, a QM/MM computational study revealed that AAG active site π–interactions are catalytic towards the excision of εA.[39] Another key active site amino acid involved in AAG-catalyzed nucleobase excision is E125, which has been proposed to help stabilize the positive charge forming on the sugar moiety upon base departure.[35] Additionally, since E125Q or E125A mutations abolish AAG activity toward εA,[34, 35] E125 has also been proposed to be the general base that accepts a proton from an active site water, which attacks C1′ and displaces the nucleobase. This latter proposal is supported by crystallographic data for AAG bound to εA[34] or εC[31] containing DNA, which reveals E125 interacts with a water molecule that is well positioned for nucleophilic attack. Consistent with this experimental data, E125 maintains hydrogen-bonding interactions with nucleophilic water throughout molecular dynamics (MD) simulations of AAG–substrate complexes,[40] and a QM/MM study has characterized an enzymatically-feasible reaction pathways for εA and 3MeA excision that employ E125 as the general base.[39]

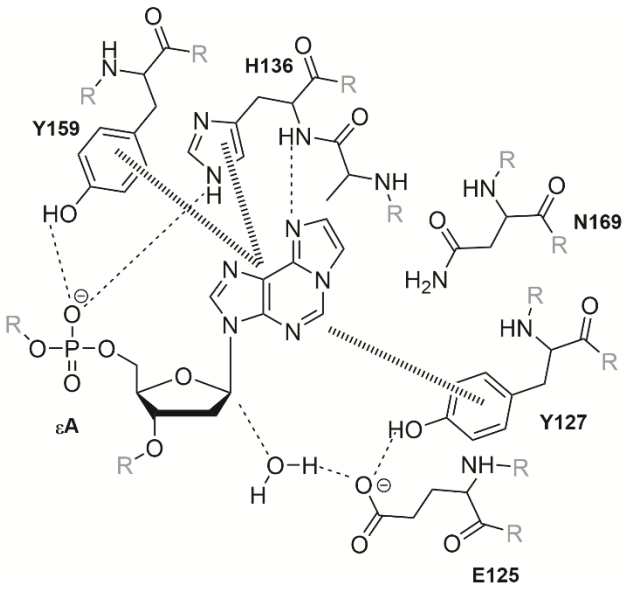


Figure 3.2. Line diagram of εA bound in the AAG active site adapted from an X-ray crystal structure (PDB ID: 1EWN).[34] Hashed bonds denote π–π or T-shaped interactions.

In addition to the roles of the aromatic residues and E125, it has been proposed that AAG may employ a general acid that targets N7 of purine lesions to facilitate nucleobase departure.[34, 35] However, it is unclear whether nucleobase protonation is required for base excision or whether hydrogen-bond donation is sufficient to stabilize the negative charge developing on the departing nucleobase. In fact, combined mass spectrometry and computational studies indicate that protonation is not a requirement for AAG-mediated excision.[41, 42] Furthermore, although protonation of A at N7 is essential for deglycosylation catalyzed by adenine DNA glycosylase (MutY),[43, 44] human uracil DNA glycosylase (hUNG2) uses a strong hydrogen bond between O2 of uracil (U) and a cationic histidine to facilitate base departure.[45-47] While it has been suggested that the H136 side chain could play the role of a general acid by transferring a proton to N7 of the departing nucleobase, a H136A AAG mutant is active.[20] Moreover, the H136A mutation has a greater affect on the AAG activity toward 7MeG compared to εA and therefore, H136 is unlikely to be the general acid for AAG-mediated excision of neutral substrates. No other mutational studies have unveiled a potential general acid. In contrast, a crystal structure of AAG bound to εA-containing DNA

reveals electron density around N7 that is likely due to active site water,[34, 35] which could help stabilize the departing charged nucleobase. Additionally, significant water density exists within the vicinity of N7 during MD simulations on AAG bound to many different purine nucleotides.[40] However, a QM/MM study of AAG-mediated excision of εA characterized a mechanism that retains water density at N7, but is void of N7 protonation, and leads to a barrier in agreement with experimental data for the chemical step.[39] Nevertheless, it is unclear whether this proposed mechanism holds true for other substrates.

Reported similarities between crystal structures of εA (substrate) and εC (inhibitor) bound in the AAG active site suggest that small changes in the active site configuration may allow the enzyme to accommodate a diverse set of lesions.[31, 34] Monte Carlo[48] and MD[49] simulations support this proposal by revealing small alterations in the active site conformation upon binding other nucleotides that have yet to be co-crystallized with AAG (i.e., A, Hx, and 1,N^6-ethanoadenine (EA)). However, these computational studies employed highly constrained models and therefore may not have adequately accounted for potential flexibility within the AAG active site. Indeed, other crystal structures of AAG bound to εC provide evidence for unique pre- and post-lesion-binding conformations, leading to suggestions that the AAG active site is disordered prior to substrate binding and the final conformation may depend on the nucleotide bound.[50] To further probe the effect of different nucleotides on the active site configuration, we previously used MD simulations to examine the active site dynamics when AAG is bound to a large range of nucleotides, including three neutral (εA, εC, and Hx) and two cationic (3MeA and 7MeG) lesions, as well as the canonical purines (A and G).[40] We determined that the positions and orientations of active site residues depend on the bound nucleotide, and proposed that the inherent flexibility of the active site allows AAG to use several enzyme–nucleobase interactions coupled with a redistribution of active site water to achieve broad substrate diversity. However, the catalytic impact of this active site reorganization for a diverse set of substrates is currently unknown. In fact, only one previous computational study has

considered the mechanism of action of AAG.[39] This study used QM/MM models based on a single, static crystal structure of AAG bound to εA to examine excision of this lesion, as well as the structurally-related A and 3MeA nucleobases.[34]

Since large-scale computational studies have been previously used to effectively probe the function of several DNA glycosylases (see, for example, MutY,[44, 51] hOGG1,[52-54] FPG,[55, 56], NEIL1,[57] TDG,[58] and hUNG2[45, 59]), the current study employs a QM/MM protocol to examine the impact of the flexibility of the AAG active site on the mechanism used to excise damaged nucleotides and discriminate against the canonical purines. Specifically, we consider the AAG-mediated excision of Hx, G, and 7MeG since these nucleotides have yet to be co-crystallized with AAG and the corresponding AAG-catalyzed deglycosylation reactions have yet to be characterized. Hx is of particular interest since AAG exhibits the highest rate enhancement for this lesion relative to non-enzymatic deglcosylation.[21] Furthermore, canonical G is not excised by AAG despite structurally deviating from Hx by an amino group, while cationic 7MeG is excised by AAG despite deviating from G by a methyl group. Moreover, Hx, G, and 7MeG are structurally distinct from εA, which has been crystallized within the AAG active site[34] and modeled in previous computational studies.[40, 48, 49] MD simulations have also proposed that Hx, G, and 7MeG induce unique AAG active site conformations compared to εA.[40] When our calculated reaction surfaces are combined with those previously mapped for εA, A and 3MeA excision,[39] we reveal that AAG uses consistent strategies to repair lesions and discriminate against the canonical purines. Furthermore, critical analysis of the active site noncovalent contacts highlights the importance of enzyme–DNA π–π stacking and T-shaped interactions, as well as the crucial role of water, in stabilizing the departing nucleobase. Overall, our study contributes to our understanding of the activity of enzymes with flexible actives sites and broad, yet discriminatory, substrate specificity, such as other DNA glycosylases[26, 55, 56, 60-62] and nucleoside hydrolases.[63-67]

3.2 Computational Methods

3.2.1 MD Simulation Protocol

The starting points used to map AAG-catalyzed reactions were adapted from our previously published MD representative structures.[40] Specifically, the representative structures were obtained by initiating MD simulations from a crystal structure of Δ79 E125Q AAG bound to εA-containing DNA.[34] Unresolved residues (80, 81, 200–207, 249–254, and 295–298) were assigned using a combination of overlays of the crystal structure of AAG bound to εC,[31] adding backbone atoms by inspection, and adding sidechains using the LEaP module of Amber 10.[68] The bound εA nucleobase was modified to Hx, G, or 7MeG and the E125Q mutation was reverted to generate wild type enzyme–DNA complexes. Each unique enzyme–DNA system was assigned AMBER ff99SB parameters,[69] supplemented by GAFF parameters[70] and restrained electrostatic potential (RESP) charges (obtained using HF/6-31G(d) single-point calculations on B3LYP/6-31G(d) geometries) for the damaged nucleobases. Each complex was neutralized with Na^+ atoms and solvated in a TIP3P water box such that at least 8.0 Å exists between the enzyme–DNA complex and the edge of the water box, and the particle mesh Ewald method was used with a cut-off of 12.0 Å. The solvent and ions were first relaxed using 500 steps of steepest descent minimization, and 500 steps of conjugate gradient minimization, with a 500 kcal mol^{-1} $Å^{-2}$ constraint applied to the solute. Subsequently, the constraints were removed and the entire system was subjected to 1000 steps of steepest descent minimization, followed by 1500 steps of conjugate gradient minimization. Next, a 10 kcal mol^{-1} $Å^{-2}$ restraint was placed on the solute as the system was heated to 300 K over 20 ps under NVT conditions. The GPU accelerated PMEMD Amber utility[71, 72] was used to simulate each system for 40 ns under NPT conditions (1 atm and 300 K) with snapshots saved every 0.5 ps. The production MD trajectories were clustered according to the positions of the active site residues (E125, Y127, A135, H136, Y159, and N169) and the bound nucleotide, and the representative structure was identified from the dominant cluster using

an average linkage rmsd algorithm until a 3.0 Å minimum cut-off was reached between clusters or three clusters were obtained.[73] In each simulation, the dominant cluster comprised more than 95% of the simulation (Table B.1 in Appendix B). Cartesian coordinates of each representative structures are available in the Supporting Information. For additional details and analysis of the molecular dynamics simulations see reference 40.

3.2.2 ONIOM(QM:MM) Computational Scheme

To generate ONIOM(QM:MM) models from the MD representative structures, all water molecules more than 2 Å from a DNA or protein atom were removed. The entire system was then optimized using the AMBER force field (parm96) available in Gaussian 09 (revisions C.01 and D.02).[74] Subsequently, the system was separated into three layers: 1) a QM region, 2) a mobile MM region that was allowed to optimize, and 3) a MM region that was fixed to the AMBER minimized structure (Figure 3.3). The QM layer consists of E125, Y127, H136, Y159, N169, and the bound nucleotide (Hx, G, or 7MeG), as well as any water molecules that directly interact with any of these residues (typically 4–5 water molecules). The mobile MM layer includes all amino acid or DNA subunits within 5.000 Å of a QM layer atom, as well as water not included in the QM layer that falls within 3.000 Å of the bound nucleotide or QM amino acid. The fixed MM layer consists of any DNA, protein or water atoms not present in the first two layers.

B3LYP has been shown to yield ONIOM(QM:MM) optimized geometries of the active sites of DNA glycosylases that closely match experimental crystal structures.[75] When the analogous methodology is applied to AAG, we also find an accurate description of the crystallographic active site configuration (Figure B.1 in Appendix B). Nevertheless, applying an empirical correction for dispersion to B3LYP has been shown to yield improved descriptions of π–π interactions.[76] Therefore, due to the rigorous procedure used to generate each reaction surface (see below), B3LYP was initially used to describe the QM layer in

order to map the reaction surface for AAG nucleotide excision. Subsequently, the stationary points identified on the reaction surfaces were re-optimized with B3LYP-D3 in order to afford an accurate description of the active site enzyme–substrate π–interactions. The MM layers were described with the AMBER force field (parm96) available within Gaussian 09 (revisions C.01 and D.01).[74] We note that ONIOM as implemented in Gaussian 09 with AMBER ff96 has been recently used to successfully model other biochemical reactions.[77-79]

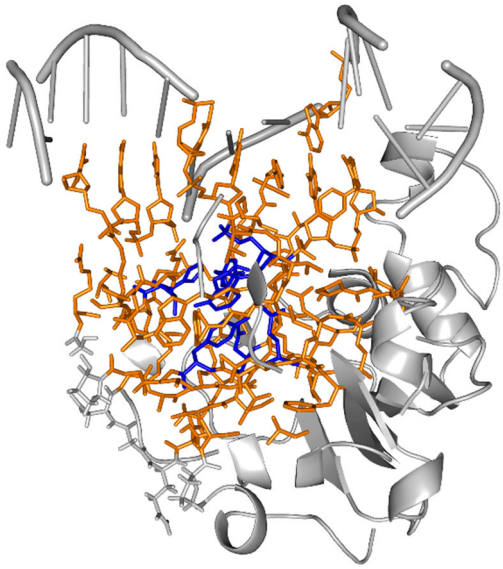


Figure 3.3. ONIOM(B3LYP/6-31G(d):AMBER) scheme used to examine AAG-catalyzed hydrolysis, with the QM layer highlighted in blue, the mobile MM region in orange, and the fixed MM layer in gray. Water molecules are removed for clarity.

3.2.3 Reaction Surface Methodology

To gain insight into the precise deglycosylation mechanism, a procedure consistent with that widely applied in the biochemical modelling literature was used to obtain reactant complexes (RC).[39, 51, 52, 80-82] Specifically, each enzyme–DNA system was first optimized using ONIOM(B3LYP/6-31G(d):AMBER). Subsequently, the RESP charges were recalculated for the QM layer using the electrostatic grid obtained from a gas-phase HF/6-31G(d) single-point calculation on the extracted QM

layer with hydrogen atoms capping truncation points. The system was re-optimized under the ONIOM(B3LYP/6-31G(d):AMBER) scheme using the new charges. This procedure was repeated in an iterative manner until the energy difference between two iterations is less than 5 kJ mol^{-1}.

From the converged RC, reaction potential energy surfaces (PES) were generated using the glycosidic bond (N9–C1′) and (water) nucleophile–DNA (O_{nuc}–C1′) distances as the reaction coordinates. The water molecule closest to C1′ that interacts with the general base (E125) and forms the most linear reaction angle ($\angle(O_{nuc}–C1′–N9)$) was chosen as the nucleophile. Each point on the reaction surface was obtained by holding the two coordinates fixed and relaxing the geometry according to our ONIOM scheme. Once a given point was optimized, the RESP charges were recalculated. Subsequently, one coordinate was adjusted by 0.200 Å to generate the next point on the surface (i.e., the starting structure of each point is based on the converged geometry and recalculated RESP charges of an adjacent point). In total, at least 150 optimizations were used to generate each reaction surface (Figure 3.4). Our meticulous approach ensures continuity along the deglycosylation pathway by minimizing errors that could potentially be introduced by simultaneously adjusting both coordinates or making a large change to one coordinate. Furthermore, recalculation of the RESP charges helps ensure that the charges of the QM layer remain stable across the deglycosylation reaction surfaces.

Once the reaction surface was generated, stationary points were re-optimized at the ONIOM(B3LYP-D3/6-31G(d):AMBER) level, while maintaining restraints on the glycosidic and nucleophilic coordinates. Subsequently, the stationary points were fully optimized with ONIOM(B3LYP-D3/6-31G(d):AMBER) without constraints on the reaction coordinates, and the structure and charges for each stationary point were iteratively converged using the protocol described above for the RC. For each stationary point, frequency calculations were performed at the optimization level of theory. For all reported transition states, visual inspection confirmed that the sole imaginary frequency corresponds to the bond

forming and/or breaking events. Relative Gibbs energies were obtained by adding the unscaled thermal correction to the ONIOM energy for each structure.

3.2.4 NCI Analysis.

The NCIPLOT[83, 84] program was used to analyze the nature of non-covalent interactions (NCI) present in the AAG active site. Specifically, the program identifies areas of low electron density (ρ) and low reduced density gradient, which are characteristic of non-covalent π–π, T-shaped, and hydrogen-bonding interactions. To determine the magnitude of a given interaction, ρ is multiplied by the second eigenvalue (λ_2) of the Hessian matrix. Promolecular densities[83, 84] were determined based on our calculated stationary points along the reaction surfaces when the reduced density gradient is less than or equal to 0.3. In the figures, blue is assigned to negative values of λ_2 (attractive interactions), while red is assigned to positive values of λ_2 (repulsive interactions), with the color ranging from $\lambda_2(\rho) = -0.0500$ au to $\lambda_2(\rho) = 0.0500$ au.

3.3 Results and Discussion

3.3.1 Hypoxanthine Substrate

The ONIOM(B3LYP/6-31G(d):AMBER) reaction surface for the AAG-mediated excision of Hx is shown in Figure 3.4a and the corresponding stationary points in Figure B.2a (Appendix B). The RC is located in the bottom left of the reaction surface, and has a glycosidic bond length of 1.500 Å and a nucleophilic distance of 3.700 Å. The most stable deglycosylation pathway proceeds through a concerted transition state (TS) complex (N9–C1′ distance of 2.500 Å and O_{nuc}–C1′ distance of 2.100 Å), while a corresponding dissociative pathway cannot be identified. A stable enthalpically disfavored product complex (PC) has a N9–C1′ distance of 2.900 Å, and an O_{nuc}–C1′ bond length of 1.5 Å. The constrained AAG–Hx stationary points optimized with ONIOM(B3LYP/6-31G(d):AMBER) and ONIOM(B3LYP-D3/6-31G(d):AMBER) are very similar (see Appendix B, Figure B.2; rmsd < ~0.14 Å; Table B.2). Across

the reaction pathway, the largest structural effect of including the empirical D3 correction in the description of the QM layer is a decrease is the distance between the H136 and Y127 sidechains, and Hx by ~0.3 Å in each stationary point (Figure B.2 in Appendix B).

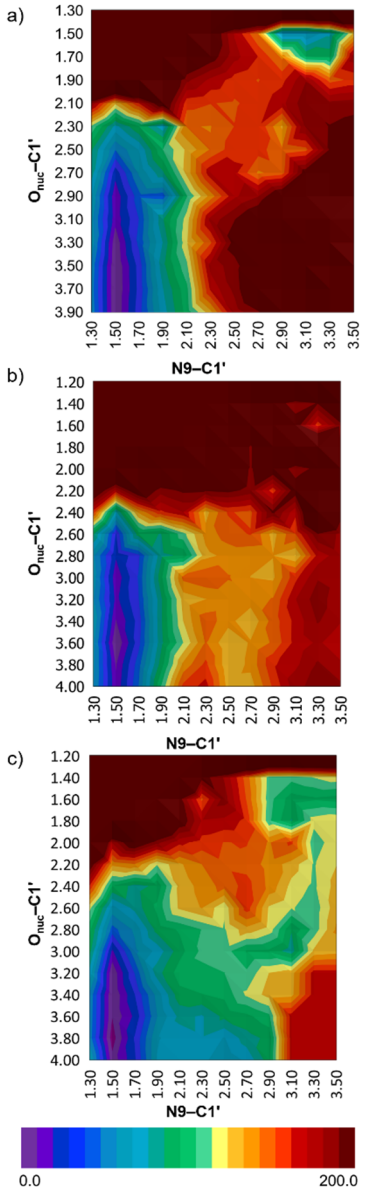


Figure 3.4. ONIOM(B3LYP/6-31G(d):AMBER) reaction surfaces (ΔE, kJ mol^{-1}) for the AAG-mediated excision of a) Hx, b) G, and c) 7MeG. The geometric coordinates N9–C1′ and O_{nuc}–C1′ (Å) correspond to the bond-breaking and bond-forming events, respectively. The energies are reported relative to the reactant complex (purple, bottom left) and each change in color corresponds to a 10 kJ mol^{-1} increase in energy.

To confirm that our constrained stationary points are not artefacts of the PES, we relaxed the stationary points without constraints. The unconstrained ONIOM(B3LYP-D3/6-31G(d):AMBER) glycosidic bond length decreases by 0.019 Å, while the nucleophilic distance decreases by 0.048 Å (Figure 3.5 and Figure B.2b in Appendix B). Other changes in the active site are minimal, leading to a very small rmsd between the constrained and unconstrained B3LYP-D3/6-31G(d) QM layers (0.010 Å, Table B.2 in Appendix B). Upon removing the constraints in the TS, the glycosidic bond length increases by 0.101 Å, while the nucleophilic distance decreases by 0.129 Å (Figure 3.5 and Figure B.2b in Appendix B), and the corresponding rmsd (0.025 Å) is slightly larger than for the RC (Table B.2 in Appendix B). Although the changes in the glycosidic bond (+0.129 Å) and nucleophilic (−0.043 Å) distances upon relaxing the PC are similar to or smaller than those for the TS (Figure 3.5a and Figure B.2b in Appendix B), the rmsd of the active site is largest for the PC (0.164 Å, Table B.2 in Appendix B). The larger geometric changes upon releasing the constraints as the reaction proceeds may be attributed to the flat nature of the reaction surface in both the TS and PC regions, as well as the particularly transient nature of the PC due to the departed nucleobase. Nevertheless, the energetic consequences of these structural changes are minimal, with the barriers and reaction energies for the constrained and unconstrained stationary points falling within ~2.5 kJ mol^{-1} (Table 3.1). Due to the high degree of similarity between the structures and energetics of the constrained and unconstrained stationary points, the remainder of this section will provide a more detailed discussion of the unconstrained ONIOM(B3LYP-D3/6-31G(d):AMBER) Hx deglycosylation pathway (Figure 3.5a).

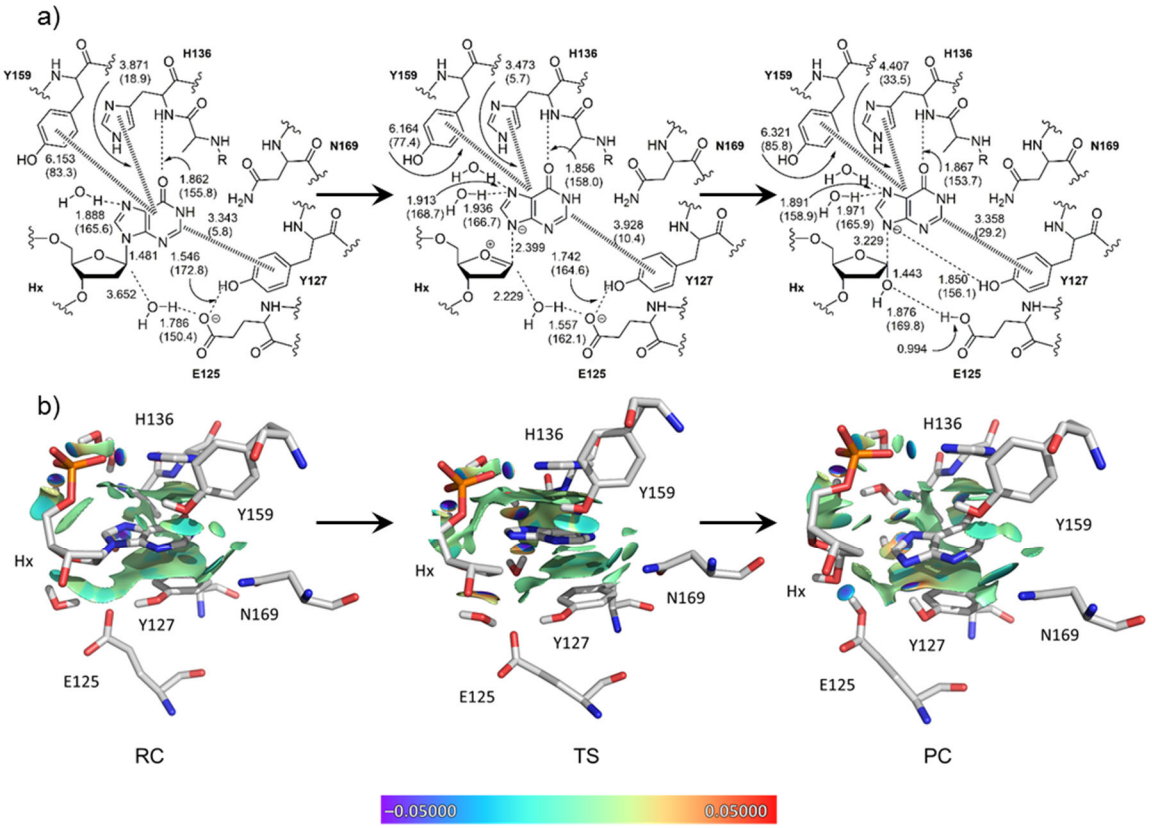


Figure 3.5. Structures of the unconstrained ONIOM(B3LYP-D3/6-31G(d):AMBER) stationary points (RC, TS, and PC) corresponding to the AAG-mediated excision of Hx displayed as a a) line diagram, with distances in Å, angles in deg, and hashed bonds denoting π–π stacking or T-shaped interactions, or as b) NCI plots with promolecular densities reported when the reduced density gradient is less than or equal to 0.3 and the color ranges from −0.05000 au (attractive interaction) to 0.0500 au (repulsive interaction).

Table 3.1. Relative Potential (ΔE) and Gibbs (ΔG) Energies (kJ mol^{-1}) for the Excision of Hx by AAG

	constrained S_N2[a]		unconstrained S_N2[b]	
	ΔE	ΔG	ΔE	ΔG
barrier	150.3	127.0	147.9	127.2
reaction energy	−55.5	−51.4	−57.2	−51.5

[a]Constrained ONIOM(B3LYP-D3/6-31G(d):AMBER) S_N2 stationary points. [b]Unconstrained ONIOM(B3LYP-D3/6-31G(d):AMBER) S_N2 stationary points. [c]Constrained ONIOM(B3LYP-D3/6-31G(d):AMBER) S_N1 stationary points.

An overlay of the AAG–Hx representative MD (AMBER) structure[40] and the unconstrained ONIOM(B3LYP-D3/6-31G(d):AMBER) RC (Figure B.3 in Appendix B) reveals that the structure of the AAG active site does not change significantly when optimized with the ONIOM methodology (rmsd = 0.565 Å), which supports the use of MD (AMBER) to accurately model the structure of other active sites that are similarly lined with aromatic residues. As previously discussed for the MD representative structure,[40] the water nucleophile forms a hydrogen bond with the general base (E125) in the ONIOM(B3LYP-D3/6-31G(d):AMBER) RC, which also hydrogen bonds with Y127 via the same oxygen atom. The only hydrogen bond between the nucleobase and the enzyme is the Hx(O6)···H136(NH) interaction (Figure 3.5a), which is consistent with a εA(N6)···H136(NH) interaction observed in a crystal structure of εA bound in the AAG active site.[34] Both H136 and Y127 form π–π stacking interactions with the nucleobase, with corresponding distances between the centers of mass (defined using all ring and exocyclic substituent atoms) of 3.871 and 3.343 Å, respectively. The angle between the planes of the Hx and H136 rings is 18.9°, while the angle between Hx and Y127 falls closer to planarity (5.8°). Y159 edge interacts with the nucleobase π–system via a T-shaped interaction (angle between the ring planes of 83.3° and inter-ring distance of 6.153 Å), while a single water molecule donates a hydrogen bond to N7 (distance and angle of 1.888 Å and 165.6°, respectively).

To gain more information about the active site AAG–DNA π–interactions, a NCI analysis was performed to obtain the promolecular density, which has been shown to provide a powerful visualization tool for non-covalent interactions in enzymatic systems.[83, 84] The evolution of the density across the reaction pathway corresponding to AAG-mediated excision of Hx is shown in Figure 3.5b, while the evolution of individual enzyme–substrate interactions is provided in Figure B.4 (Appendix B). In the RC, there is significant promolecular density between the deoxyribose sugar and E125 (Figure 3.5b). Furthermore, favorable density exists between the Hx nucleobase, and H136 and Y127, which form amino

acid–nucleobase π–π stacking interactions (Figure 3.5 and Figure B.4a,b in Appendix B), as well as between Y159 and the nucleobase (Figure 3.5b and Figure B.4c in Appendix B), which form a T-shaped interaction. The NCI data also highlights stabilizing hydrogen bonding between water and N7 of Hx, as well as the 5′-phosphate backbone (Figure 3.5b and Figure B.4c in Appendix B).

As the reaction proceeds from the RC to TS, the glycosidic bond increases to 2.399 Å and the nucleophilic distance shortens to 2.229 Å (Figure 3.5a), while the energy increases to ~125 kJ mol^{-1} (ΔG, Table 3.1). As the nucleophile approaches C1′ in the TS, the hydrogen-bond distance between the water nucleophile and E125 shortens by ~0.2 Å relative to the RC and a proton is partially transferred to the general base, while the E125(Oε)···Y127(OH) hydrogen bond elongates by ~0.2 Å. The promolecular density between E125 and the sugar observed in the RC disappears as the nucleophile attacks. Interestingly, the distance between the H136 imidazole ring and the Hx nucleobase decreases by ~0.4 Å and the angle between the ring planes increases in planarity in the TS compared to the RC (Figure 3.5a), while the promolecular density between H136 and Hx increases, reflecting tightening of this contact (Figure 3.5b and Figure B.4a in Appendix B). The tighter π–contact likely plays an important role in stabilizing the negative charge that develops on the nucleobase during the reaction. Conversely, the Y127–Hx π–π stacking interaction elongates by ~0.5 Å, the angle increases by ~5° and the corresponding promolecular density decreases as the reaction proceeds (Figure 3.5 and Figure B.4b in Appendix B). The geometric parameters and density governing the Y159–Hx T-shaped interaction remain relatively consistent between the RC and TS (Figure 3.5 and Figure B.4c in Appendix B). The solvation pattern at N7 of Hx, as well as the 5′ phosphate, changes substantially as the reaction proceeds to the TS. Most notably, although a single water molecule forms a hydrogen bond with N7 of Hx in the RC, two water molecules hydrogen bond to N7 in the TS (Figure 3.4 and Figure B.4d in Appendix B), which further stabilizes the departing nucleobase.

In the PC, the nucleobase is dissociated (N9–C1′ distance = 3.229 Å), the nucleophile forms a bond with C1′ (O_{nuc}–C1′ distance = 1.443 Å), and proton transfer occurs between E125 and the nucleophile (Figure 3.5a). E125 donates a hydrogen bond to O1′, which is highlighted in the only promolecular density between E125 and deoxyribose (Figure 3.5b). The active site expands in response to the departed nucleobase, with the distance between Hx and H136 or Y159 increasing by approximately 1.0 or 0.2 Å compared to the TS, respectively, and the promolecular density decreasing in the PC (Figure B.4a,c in Appendix B). Interestingly, the distance between Y127 and Hx decreases by ~0.6 Å, and a hydrogen bond forms between the hydroxy group of Y127 and N9 of Hx, which stabilizes the (anionic) departed nucleobase. Furthermore, the two hydrogen bonds between water and the nucleobase that stabilize the TS (distances = 1.936 and 1.913 Å; angles = 166.7 and 168.7°) remain in the PC (distances = 1.971 and 1.891 Å; angles = 165.9 and 159.0°; Figure B.4d in Appendix B). Consequently, the Hx deglycosylation reaction is exergonic by ~50 kJ mol^{-1}.

Overall, our data reveals fundamental details about the reaction mechanism for AAG-mediated deglycosylation of Hx. Specifically, although E125 exhibits considerable interactions with the deoxyribose moiety in the reactant complex, we find little evidence that E125 provides significant charge stabilization to the substrate as the reaction proceeds, which contrasts a previous proposal.[35] Instead, the primary role of E125 is to activate the nucleophilic water by accepting a proton. When coupled with the concerted (S_N2) hydrolysis pathway previously characterized for the excision of εA that involves E125 as the general base,[39] our data suggests that AAG may invoke a consistent mechanism for the repair of neutral lesions. Although it has been proposed that H136 may protonate N7 of purine lesions during excision,[20] our data suggests that H136 stabilizes the departing nucleobase through π–π stacking interactions. This important (stabilization), but not critical (activation), role of His136 is supported by experimental mutational studies that revealed a H136A mutation only results in a 40-fold reduction in AAG activity[20] and previous

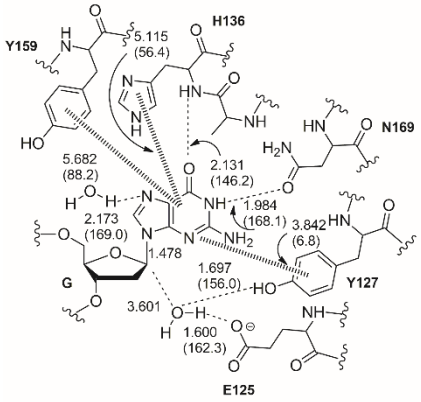
computational work that predicted a H136A mutation slightly increases the εA excision barrier.[39] To complement H136, active site water facilitates base departure through hydrogen bonding at N7, which builds upon the previously proposed roles of active site water in dictating the substrate specificity of AAG.[40] While solvation at N7 increases as the reaction proceeds, our calculations suggest that full nucleobase protonation is not required for excision, which correlates with previous proposals for AAG based on mass spectrometry and computational data,[41, 42] as well as conclusions that other glycosylases use active site hydrogen bonding to facilitate lesion excision (e.g., hUNG2).[45-47] Y159 provides some stabilization to the departing base despite maintaining a consistent position during the reaction, which correlates with the experimental observation that a Y159W AAG mutant is less effective at excising DNA-containing εA.[85] In addition to stabilizing the nucleobase through π–π interactions, Y127 plays an important role by forming a key hydrogen bond to E125 that holds this residue in a catalytically-competent orientation. Furthermore, Y127 donates a hydrogen bond to N9 of the Hx nucleobase in the PC, which provides a potential rationalization for the experimental observation that product release is the rate-determining step in the AAG-mediated repair of Hx.[21]

3.3.2 Guanine Nucleotide

Due to the information about the reaction pathway for AAG-mediated excision of Hx obtained from our methodology, the deglycosylation of G is considered to help understand how AAG discriminates against the canonical purines. A comparison between G and Hx deglycosylation is particularly important since previous MD simulations reveal that the additional amino group of G disrupts the AAG active site configuration upon nucleotide binding.[40] There are minimal differences between the constrained ONIOM(B3LYP/6-31G(d):AMBER), constrained ONIOM(B3LYP-D3/6-31G(d):AMBER) (rmsd = 0.170 Å), and unconstrained ONIOM(B3LYP-D3/6-31G(d):AMBER) optimized RC (rmsd < 0.188 Å,

Table B.2 in Appendix B), as well as the representative MD structure (rmsd with respect to the unconstrained RC equals 0.564 Å, Figure B.5 in Appendix B).

The unconstrained ONIOM(B3LYP-D3/6-31G(d):AMBER) RCs reveal that the AAG active site undergoes several significant changes when bound to G (Figure 3.6) compared to Hx (Figures 3.5a and 3.7). Interestingly, the distance between the aromatic ring of H136 and the nucleobase increases when G is bound (5.115 Å) compared to Hx (3.871 Å). Most notably, N1–H of the G nucleobase donates a hydrogen bond to the sidechain of N169, while no interaction with N169 was observed for Hx. Furthermore, the hydrogen-bonding network that stabilizes the position and orientation of the general base (E125) changes. Specifically, although Y127 forms a tight hydrogen bond with E125 in the Hx RC (distance = 1.546 Å; angle = 172.8°), which correctly positions the general base in close proximity to C1′ (E125(Oε)–C1′ = 4.655 Å), Y127 interacts with the nucleophile in the G RC, which positions E125 further from C1′ (E125(Oε)–C1′ = 6.100 Å). Conversely, the active site π–interactions are consistent, with Y159–G and Y127–G distances of 5.682 and 3.842 Å, respectively, compared to Y159–Hx and Y127–Hx distances of 6.153 and 3.343 Å. Furthermore, Hx and G are both solvated at N7.



RC

Figure 3.6. Structure of the unconstrained ONIOM(B3LYP-D3/6-31G(d):AMBER) RC corresponding to the AAG-mediated excision of G displayed as a line diagram, with distances in Å, angles in deg, and hashed bonds denoting π–π stacking or T-shaped interactions.

Compared to Hx, our calculated reaction surface for G deglycosylation is significantly destabilized relative to the corresponding RC (Figure 3.4b). In fact, no stable excision pathway can be identified on the reaction surface. To evaluate how the active site configuration evolves along a potential deglycosylation pathway and leads to an unstable reaction surface, we more closely examine the unconstrained RC (Figure 3.8a) and a representative complex from the region of the reaction surface corresponding to the TS on the Hx excision surface (termed the reaction surface complex or RSC; N9–C1′ distance = 2.5 Å; nucleophilic distance = 2.2 Å; Figure 3.8b). While E125 and Y127 maintain hydrogen bonding across Hx deglycosylation, E125 does not interact with Y127 in the unconstrained RC for G, but shifts to hydrogen bond with Y127 as the N9–C1′ distance elongates and the O_{nuc}–C1′ shortens (Figure 3.8). Despite the formation of the E125(Oε)···Y127(OH) hydrogen bond, several factors destabilize the enzyme–DNA complex across the reaction surface (Figure B.6 in Appendix B). Specifically, a weak interaction is observed between the H136 side chain and the nucleobase (Figures 8 and Figure B.6a in Appendix B) in the unconstrained RC. Although the corresponding density increases as the glycosidic bond is lengthened and the nucleophile distance decreases, the stacking interaction between G and H136 is minimal compared to the corresponding interaction in the Hx TS (Figures B.4a and B.6a in Appendix B), which decreases the stabilization of the negatively-charged nucleobase by H136 as seen for Hx. An overlay of the Hx TS and G RSC reveals that the position of the departing nucleobase in the active site is not consistent between the two nucleotides as the glycosidic bond elongates due to the N169(Oε)···G(N1H) hydrogen bond (Figure 3.7). Furthermore, favorable density suggests that the N169–G interaction is further stabilized by a N169(Nε)···G(N2H2) hydrogen bond (Figure 3.8). Consequently, the G(O6)···H136(NH) hydrogen bond in the unconstrained RC is broken and a new G(N7)···H136(NH) hydrogen is formed as the glycosidic bond elongates (Figure 3.7). This is in stark contrast to Hx deglycosylation during which a Hx(O6)···H136(NH) hydrogen bond is maintained throughout the pathway (Figure 3.5a). Another outcome of the new G nucleobase orientation is that water does not interact with N7 of G as the glycosidic

bond elongates (Figure B.6d in Appendix B), further limiting stabilization of the departing nucleobase. Nevertheless, the interactions between the nucleobase and Y159 or Y127 remain relatively consistent across the G and Hx deglycosylation pathways.

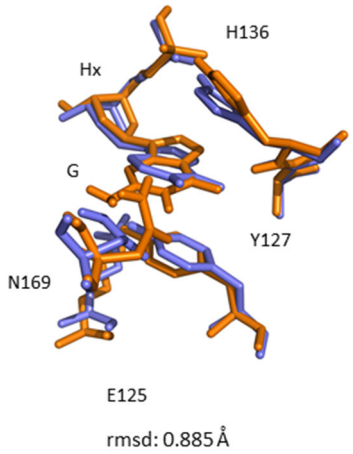


Figure 3.7. Overlay of the unconstrained ONIOM(B3LYP-D3/6-31G(d):AMBER) G (orange) and Hx (blue) RC.

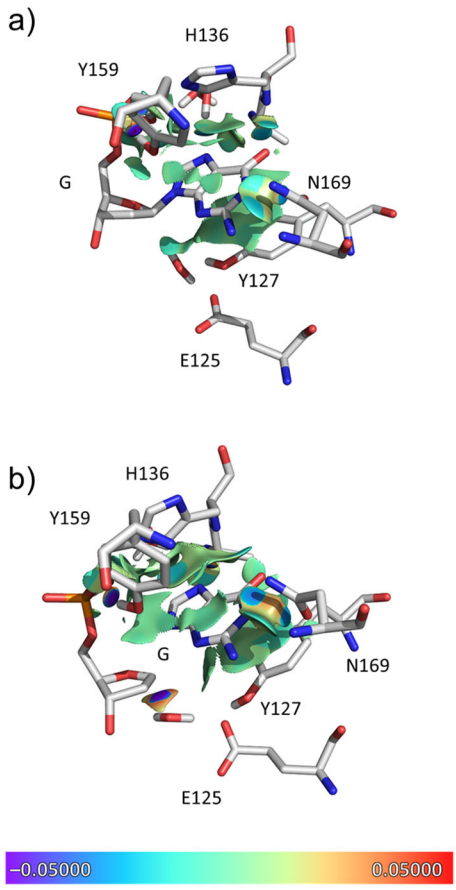


Figure 3.8. NCI plots of the a) unconstrained ONIOM(B3LYP-D3/6-31G(d):AMBER) G RC and b) representative structure with an elongated glycosidic bond obtained from the G reaction surface, with promolecular densities reported when the reduced density gradient is less than or equal to 0.3 and the color ranges from −0.05000 au (attractive interaction) to 0.0500 au (repulsive interaction).

Despite the similar structures of G and Hx, the lack of a stable reaction pathway for G is a direct consequence of subtle differences in the positions of AAG active site residues upon nucleotide binding. The majority of these differences arise from a unique nucleobase position due to a N169–G hydrogen bond with the exocyclic amino group that is absent in Hx. This correlates with the experimental observation that N169S[21] and N169A[86] AAG mutants have higher activity toward G than wild-type AAG. Specifically, since serine and alanine are unlikely to form a similar hydrogen bond with G compared to asparagine, these mutants would likely favorably position the G nucleobase to form a G(O6)···H136(NH) hydrogen bond and thereby maintain a nucleotide position consistent with Hx in wild-type AAG. This provides an

alternative rationalization for the experimental data than previous proposals that suggest the exocyclic amino group of G sterically clashes with N169, which is reduced by S or A.[21, 86] Due to the change in nucleobase position along the reaction pathway compared to Hx, neither the sidechain of H136 nor active site water can stabilize the departing G nucleobase, which helps at least in part rationalize the lack of G excision. Perhaps most importantly, Y127 does not interact with E125 when G is bound in the active site, which displaces the general base and nucleophilic water, and renders a catalytically incompetent reactant complex. When coupled with a previous proposal that misalignment of E125 and the nucleophilic water prevents AAG from excising A,[39] our data point towards a unified structural explanation for the inability of AAG to excise the natural purines.

3.3.3 7MeG Substrate

Despite possessing the same exocyclic amino group as G, 7MeG is efficiently excised by AAG.[21] MD data suggests this differential activity arises since the lesion and G adopt unique positions in the AAG active site.[40] A comparison of the reaction surfaces for the excision of 7MeG and G is necessary to shed light on how these differences in binding affect AAG activity. The reaction surface for AAG-mediated excision of 7MeG is shown in Figure 3.4c and the corresponding refined ONIOM(B3LYP-D3/6-31G(d):AMBER) reaction energetics are summarized in Table 3.2. The RC has a glycosidic bond length of 1.500 Å and a nucleophilic distance of 3.800 Å, and does not significantly differ from the MD representative structure (rmsd = 0.300 Å; Figure B.7 in Appendix B). In contrast to G, a stable excision pathway can be identified for 7MeG. The most stable deglycosylation pathway proceeds through a dissociative (S_N1) transition state (N9–C1′ distance = 2.5 Å; nucleophilic distance = 3.0 Å; relative energy = 117.2 kJ mol^{-1}). The corresponding nearly isoenergetic intermediate complex (IC) is characterized by complete dissociation of the nucleobase (N9–C1′ distance = 3.100 Å; nucleophilic distance = 3.000 Å; relative energy = 103.7 kJ mol^{-1}). Subsequently, an associative TS (N9–C1′ distance = 3.100 Å; nucleophilic distance = 2.000 Å;

relative energy = 115.0 kJ mol^{-1}) leads to an endothermic PC (N9–C1′ distance = 3.100 Å; O_{nuc}–C1′ distance = 1.600 Å; relative energy = 78.5 kJ mol^{-1}). A higher energy concerted (S_N2) TS can also be identified on the surface (N9–C1′ distance = 2.500 Å; nucleophilic distance = 2.200 Å; Figure B.8 in Appendix B; relative energy = 146.0 kJ mol^{-1}). The lower barrier for 7MeG excision relative to Hx excision is consistent with previous literature indicating that positively-charged substrates have lower deglycosylation barriers than neutral substrates.[21, 87] Although the flat reaction surface near the transition state region made attempts to fully relax (unconstrained) stationary points unsuccessful, the similarities in the constrained and unconstrained ONIOM(B3LYP-D3/6-31G(d):AMBER) RC, TS, and PC for Hx (Figure 3.5 and Table B.2 and Figure B.2 in Appendix B) suggest the constrained structures will permit an accurate comparison of the 7MeG and G deglycosylation pathways.

Table 3.2. Relative (ΔE) and Gibbs (ΔG) Energies (kJ mol^{-1}) for the Excision of 7MeG by AAG[a]

	ΔE	ΔG
R	0.0	0.0
$TS_{dissociative}$	113.8	117.2
IC	107.4	103.7
$TS_{associative}$	117.9	115.0
PC	54.0	78.5
$TS_{concerted}$	153.0	146.0

[a]Constrained ONIOM(B3LYP-D3/6-31G(d):AMBER) stationary points.

A steric clash between the N7 methyl group of the nucleobase and the H136 backbone results in 7MeG being less inserted into the AAG active site than G in the constrained ONIOM(B3LYP-D3/6-31G(d):AMBER) RC (Figures 3.6 and 3.9–3.10). As a consequence, while G forms a N169(Oε)···G(N1H) hydrogen bond (Figure 3.6), 7MeG forms a N169(Oε)···7MeG(N2H2) interaction (Figure 3.9 and Figure B.9 in Appendix B), which is maintained throughout the reaction pathway (distance = ~1.8–1.9 Å; angle = 153.0°), and results in less repulsion between the bound nucleotide and

N169. More importantly, in stark contrast to the G RC, a E125(Oε)···Y127(OH) hydrogen bond exists in the 7MeG RC, which results in a catalytically conducive configuration of the general base and nucleophile (Figures 3.9 and 3.10). Furthermore, the E125(Oε)···Y127(OH) hydrogen bond is maintained throughout the 7MeG excision pathway (distance = ~1.5–1.6 Å; angle = ~160–177°) for the RC, TS, and I, albeit elongating in the PC (distance = 1.903 Å; angle = 138.4°; Figure 3.9). This is very similar to the properties of the E125–Y127 dyad along the Hx reaction surface. In terms of key nucleobase interactions in the AAG active site, the H136–nucleobase distance tightens in each TS compared to the RC and IC (by up to ~ 0.6 Å), while the Y159 and Y127–nucleobase distances are tighter in the RC and IC than the dissociative and associative TSs (by up to ~0.2 Å and ~0.3 Å, respectively; Figure 3.9 and Figure B.9 in Appendix B). However, since the 7MeG nucleobase is neutral in the TS and IC, the role of the 7MeG–H136 π–contact may be less important for 7MeG excision compared to Hx excision (i.e., when the nucleobase is negatively charged in the TS). Unlike Hx, the N7 methyl group of 7MeG precludes solvation of the nucleobase at this position. Nevertheless, the inherent positive charge of 7MeG affords nucleobase activation, which when coupled with correct alignment of the general base permits deglycosylation. Indeed, E125 is similarly aligned for excision of 3MeA.[39]

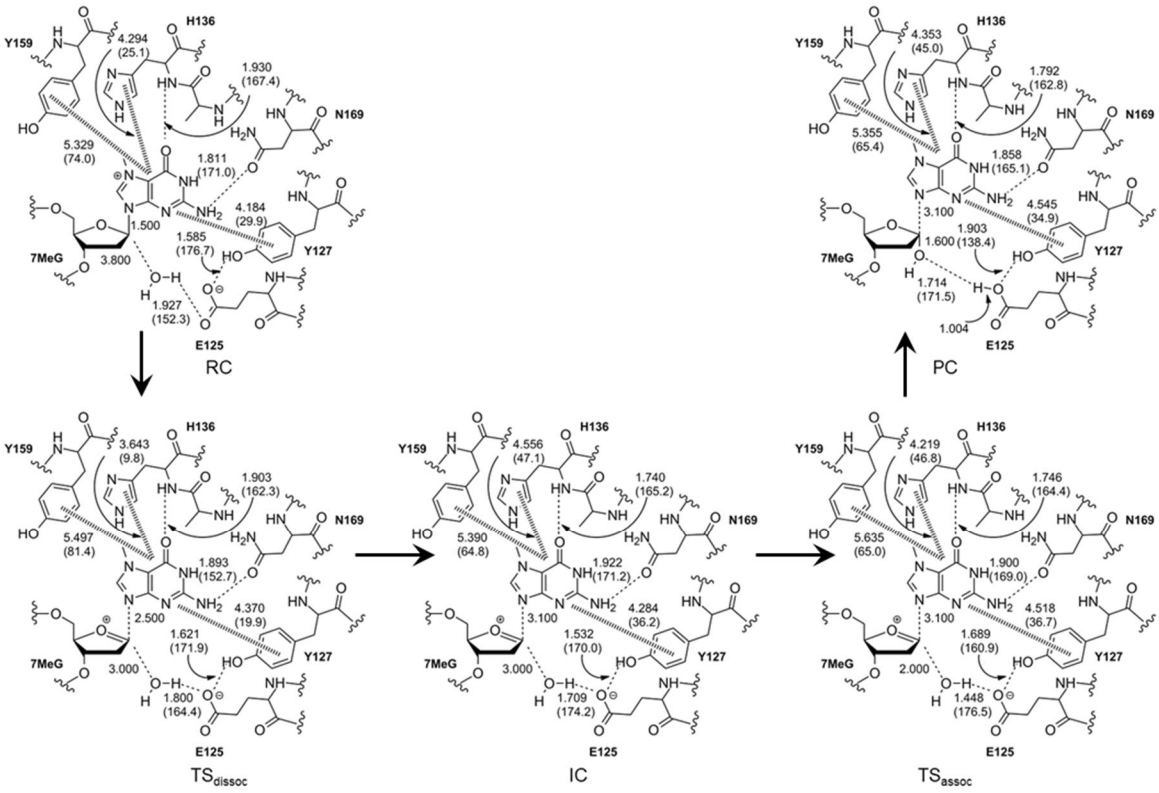


Figure 3.9. Structures of constrained ONIOM(B3LYP-D3/6-31G(d):AMBER) stationary points (RC, TS_{dissoc}, I, TS_{assoc}, and PC) corresponding to AAG-mediated excision of 7MeG displayed as a line diagram, with distances in Å, angles in deg, and hashed bonds denoting π–π stacking or T-shaped interactions.

In summary, when combined with previous MD structural data on the reactant complex for a range of substrates,[40] and a QM/MM study of the AAG-mediated excision of εA, A, and 3MeA,[39] our work emphasizes that subtle deviations in the position of the nucleotide within the active site affect the activity of AAG. Most importantly, the maintenance of the E125–Y127 dyad is critical for AAG function. Specifically, this hydrogen-bonding pattern maintains a catalytically conducive position of the E125 general base and permits deglycosylation of both neutral (Hx and εA) and cationic (7MeG and 3MeA) lesions. In contrast, disruption of the E125–Y127 catalytic dyad upon AAG binding to G or A is a major factor preventing excision of the canonical purines.

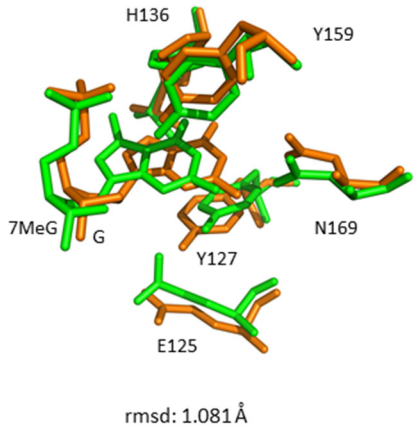


Figure 3.10. Overlay of the AAG active site for the RC obtained from constrained ONIOM(B3LYP-D3/6-31G(d):AMBER) optimization with 7MeG bound (green) and constrained ONIOM(B3LYP/6-31G(d):AMBER) optimizations with G bound (orange).

3.4 Conclusion

Our study extends upon previous experimental and computational work on AAG by providing fundamental information regarding the deglycosylation mechanism for two substrates (DNA-containing neutral Hx and cationic 7MeG) and one non-substrate (G). As a result, we shed light on the broad, yet discriminatory, substrate specificity of this repair enzyme, and clarify the roles of key active site residues. AAG-mediated excision of Hx is facilitated by the correct alignment of the E125 general base due to hydrogen bonding with Y127, and nucleobase activation through active site π–π interactions with aromatic amino acids (Y127, H136, and Y159) and hydrogen bonding with active site water. In contrast, as previously reported from MD simulations,[40] there is significant steric repulsion between the exocyclic amino group of G and N169 in the reactant complex, which disrupts the AAG active site. The present study reveals the consequences of this active site disruption are elimination of the crucial Y127–E125 hydrogen bond throughout the reaction pathway, and displacement of the general base and nucleophilic water. When coupled with reduced active site π–π interactions, and the absence of hydrogen bonds between the

nucleobase and active site water, our data rationalizes the lack of AAG activity toward G. Although 7MeG contains the same exocyclic amino group as G, steric clashes between the N7 methyl group and H136 situate the lesion differently from the canonical base, which preserves the catalytically essential hydrogen bond between the side chains of E125 and Y127, and permits 7MeG deglycosylation. When combined with previous data for the excision of εA, A and 3MeA,[39] our results suggest that AAG can employ a unified mechanism to excise a variety of structurally diverse oxidized and alkylated DNA purines, and yet discriminate against the canonical purines.

3.5 References

[1] Grando, S. A. Connections of nicotine to cancer. *Nat. Rev. Cancer* **2014,** *14*, 419–429.

[2] Gaillard, H.; Garcia-Muse, T.; Aguilera, A. Replication stress and cancer. *Nat. Rev. Cancer* **2015,** *15*, 276–289.

[3] Kanvah, S.; Joseph, J.; Schuster, G. B.; Barnett, R. N.; Cleveland, C. L.; Landman, U. Oxidation of DNA: Damage to nucleobases. *Acc. Chem. Res.* **2010,** *43*, 280–287.

[4] Sharma, S.; Shrivastav, B. R.; Shrivastav, A. Free radicals, antioxidants and oxidative stress. *Int. J. Adv. Res.* **2013,** *1*, 252–258.

[5] Zheng, G.; Fu, Y.; He, C. Nucleic acid oxidation in DNA damage repair and epigenetics. *Chem. Rev.* **2014,** *114*, 4602–4620.

[6] Cadet, J.; Wagner, J. R. Oxidatively generated base damage to cellular DNA by hydroxyl radical and one-electron oxidants: Similarities and differences. *Arch. Biochem. Biophys.* **2014,** *557*, 47–54.

[7] Dedon, P. C.; Tannenbaum, S. R. Reactive nitrogen species in the chemical biology of inflammation. *Arch. Biochem. Biophys.* **2004,** *423*, 12–22.

[8] Hill-Perkins, M.; Jones, M. D.; Karran, P. Site-specific mutagenesis *in vivo* by single methylated or deaminated purine bases. *Mutat. Res. Fund. Mol. Mech. Mut.* **1986,** *162*, 153–163.

[9] 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*, 247–271.

[10] Sedgwick, B. Repairing DNA-methylation damage. *Nat. Rev. Mol. Cell Biol.* **2004,** *5*, 148–157.

[11] Boysen, G.; Pachkowski, B. F.; Nakamura, J.; Swenberg, J. A. The formation and biological significance of N7-guanine adducts. *Mutat. Res. Genet. Toxicol. Environ. Mutagen.* **2009,** *678*, 76–94.

[12] Brosh, R. M. DNA helicases involved in DNA repair and their roles in cancer. *Nat. Rev. Cancer* **2013,** *13*, 542–558.

[13] Marteijn, J. A.; Lans, H.; Vermeulen, W.; Hoeijmakers, J. H. J. Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat. Rev. Mol. Cell Biol.* **2014,** *15*, 465–481.

[14] Berdis, A. J. Mechanisms of DNA polymerases. *Chem. Rev.* **2009,** *109*, 2862–2879.

[15] Stivers, J. T.; Jiang, Y. L. A mechanistic perspective on the chemistry of DNA repair glycosylases. *Chem. Rev.* **2003,** *103*, 2729–2759.

[16] Berti, P. J.; McCann, J. A. B. 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*, 506–555.

[17] Zharkov, D. O. Base excision DNA repair. *Cell. Mol. Life Sci.* **2008,** *65*, 1544–1565.

[18] Drohat, A. C.; Maiti, A. Mechanisms for enzymatic cleavage of the N-glycosidic bond in DNA. *Org. Biomol. Chem.* **2014,** *12*, 8367–8378.

[19] Schermerhorn, K. M.; Delaney, S. A chemical and kinetic perspective on base excision repair of DNA. *Acc. Chem. Res.* **2014,** *47*, 1238–1246.

[20] O'Brien, P. J.; Ellenberger, T. Human alkyladenine DNA glycosylase uses acid-base catalysis for selective excision of damaged purines. *Biochemistry* **2003,** *42*, 12418–12429.

[21] O'Brien, P. J.; Ellenberger, T. Dissecting the broad substrate specificity of human 3-methyladenine-DNA glycosylase. *J. Biol. Chem.* **2004,** *279*, 9750–9757.

[22] Wolfe, A. E.; O'Brien, P. J. Kinetic mechanism for the flipping and excision of 1,N^6-ethenoadenine by human alkyladenine DNA glycosylase. *Biochemistry* **2009,** *48*, 11357–11369.

[23] Melamede, R. J.; Hatahet, Z.; Kow, Y. W.; Ide, H.; Wallace, S. S. Isolation and characterization of endonuclease VIII from *Escherichia coli*. *Biochemistry* **1994,** *33*, 1255–1264.

[24] Jiang, D. Y.; Hatahet, Z.; Melamede, R. J.; Kow, Y. W.; Wallace, S. S. Characterization of *Escherichia coli* endonuclease VIII. *J. Biol. Chem.* **1997,** *272*, 32230–32239.

[25] Dizdaroglu, M.; Burgess, S. M.; Jaruga, P.; Hazra, T. K.; Rodriguez, H.; Lloyd, R. S. Substrate specificity and excision kinetics of *Escherichia coli* endonuclease VIII (Nei) for modified bases in DNA damaged by free radicals. *Biochemistry* **2001,** *40*, 12150–12156.

[26] Krishnamurthy, N.; Zhao, X.; Burrows, C. J.; David, S. S. Superior removal of hydantoin lesions relative to other oxidized bases by the human DNA glycosylase hNEIL1. *Biochemistry* **2008,** *47*, 7137–7146.

[27] Prakash, A.; Doublié, S.; Wallace, S. S. The Fpg/Nei family of DNA glycosylases: Substrates, structures and search for damage. *Prog. Mol. Biol. Transl. Sci.* **2012,** *110*, 71–91.

[28] Boiteux, S.; Gajewski, E.; Laval, J.; Dizdaroglu, M. Substrate-specificity of the *Escherichia coli* Fpg protein (formamidopyrimidine DNA glycosylase) - excision of purine lesions in DNA produced by ionizing-radiation or photosensitization. *Biochemistry* **1992,** *31*, 106–110.

[29] Porello, S. L.; Leyes, A. E.; David, S. S., Single-turnover and pre-steady-state kinetics of the reaction of the adenine glycosylase MutY with mismatch-containing DNA substrates. *Biochemistry* **1998,** *37*, 14756–14764.

[30] Kuznetsov, N. A.; Koval, V. V.; Fedorova, O. S. Mechanism of recognition and repair of damaged DNA by human 8-oxoguanine DNA glycosylase hOgg1. *Biochem.-Moscow* **2011,** *76*, 118–130.

[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,N^4-ethenocytosine-containing DNA. *J. Biol. Chem.* **2011,** *286*, 13205–13213.

[32] Martin, F. H.; Castro, M. M.; Aboul-ela, F.; Tinoco, I. Base pairing involving deoxyinosine: Implications for probe design. *Nucleic Acids Res.* **1985,** *13*, 8927–8938.

[33] Biswas, T.; Clos, L. J.; SantaLucia, J.; Mitra, S.; Roy, R. Binding of specific DNA base-pair mismatches by N-methylpurine-DNA glycosylase and its implication in initial damage recognition. *J. Mol. Biol.* **2002,** *320*, 503–513.

[34] 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*, 13573–13578.

[35] Lau, A. Y.; Scharer, O. D.; Samson, L.; Verdine, G. L.; Ellenberger, E. Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: Mechanisms for nucleotide flipping and base excision. *Cell* **1998,** *95*, 249–258.

[36] Wilson, K. A.; Kellie, J. L.; Wetmore, S. D. DNA–protein π -interactions in nature: Abundance, structure, composition and strength of contacts between aromatic amino acids and DNA nucleobases or deoxyribose sugar. *Nucleic Acids Res.* **2014,** *42*, 6726–6741.

[37] Wilson, K. A.; Wells, R. A.; Abendong, M. N.; Anderson, C. B.; Kung, R. W.; Wetmore, S. D. Landscape of π–π and sugar–π contacts in DNA–protein interactions. *J. Biomol. Struct. Dyn.* **2015**, 1-14.

[38] Wilson, K. A.; Wetmore, S. D. Combining crystallographic and quantum chemical data to understand DNA-protein π-interactions in nature. *Struct. Chem.* **2017,** *28*, 1487–1500.

[39] 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*, 16258–16269.

[40] 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*, 798–808.

[41] Sun, X.; Lee, J. K. Acidity and proton affinity of hypoxanthine in the gas phase versus in solution: Intrinsic reactivity and biological implications. *J. Org. Chem.* **2007,** *72*, 6548–6555.

[42] Liu, M.; Xu, M.; Lee, J. K. The acidity and proton affinity of the damaged base 1,N^6-ethenoadenine in the gas phase versus in solution: Intrinsic reactivity and biological implications. *J. Org. Chem.* **2008,** *73*, 5907–5914.

[43] Lee, S.; Verdine, G. L. Atomic substitution reveals the structural basis for substrate adenine recognition and removal by adenine DNA glycosylase. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106*, 18497–18502.

[44] Brunk, E.; Arey, J. S.; Rothlisberger, U. Role of environment for catalysis of the DNA repair enzyme MutY. *J. Am. Chem. Soc.* **2012,** *134*, 8608–8616.

[45] 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*, 4218–4227.

[46] Kavli, B.; Sundheim, O.; Akbari, M.; Otterlei, M.; Nilsen, H.; Skorpen, F.; Aas, P. A.; Hagen, L.; Krokan, H. E.; Slupphaug, G. hUNG2 is the major repair enzyme for removal of uracil from U:A matches, U:G mismatches, and U in single-stranded DNA, with hSMUG1 as a broad specificity backup. *J. Biol. Chem.* **2002,** *277*, 39926–39936.

[47] Slupphaug, G.; Eftedal, I.; Kavli, B.; Bharati, S.; Helle, N. M.; Haug, T.; Levine, D. W.; Krokan, H. E. Properties of a recombinant human uracil-DNA glycosylase from the UNG gene and evidence that UNG encodes the major uracil-DNA glycosylase. *Biochemistry* **1995,** *34*, 128–138.

[48] Alexandrova, A. N. Promiscuous DNA alkyladenine glycosylase dramatically favors a bound lesion over undamaged adenine. *Biophys. J* . **2010,** *152*, 118–127.

[49] 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*, 3778–3787.

[50] Setser, J. W.; Lingaraju, G. M.; Davis, C. A.; Samson, L. D.; Drennan, C. L. Searching for DNA lesions: Structural evidence for lower- and higher-affinity DNA binding conformations of human alkyladenine DNA glycosylase. *Biochemistry* **2012,** *51*, 382–390.

[51] Kellie, J. L.; Wilson, K. A.; Wetmore, S. D. Standard role for a conserved aspartate or more direct involvement in deglycosylation? An ONIOM and MD investigation of adenine-DNA glycosylase. *Biochemistry* **2013,** *52*, 8753–8765.

[52] Kellie, J. L.; Wilson, K. A.; Wetmore, S. D. An ONIOM and MD investigation of possible monofunctional activity of human 8-oxoguanine-DNA glycosylase (hOgg1). *J. Phys. Chem. B* **2015,** *119*, 8013–8023.

[53] Sadeghian, K.; Ochsenfeld, C. Unraveling the base excision repair mechanism of human DNA glycosylase. *J. Am. Chem. Soc.* **2015,** *137*, 9824–9831.

[54] Šebera, J.; Hattori, Y.; Sato, D.; Řeha, D.; Nencka, R.; Kohno, T.; Kojima, C.; Tanaka, Y.; Sychrovský, V. The mechanism of the glycosylase reaction with hOgg1 base-excision repair enzyme: Concerted effect of Lys249 and Asp268 during excision of 8-oxoguanine. *Nucleic Acids Res.* **2017,** *45*, 5231–5242.

[55] Blank, I. D.; Sadeghian, K.; Ochsenfeld, C. A base-independent repair mechanism for DNA glycosylase—no discrimination within the active site. *Sci. Rep.* **2015,** *5*, 10369.

[56] Sadeghian, K.; Flaig, D.; Blank, I. D.; Schneider, S.; Strasser, R.; Stathis, D.; Winnacker, M.; Carell, T.; Ochsenfeld, C. Ribose-protonated DNA base excision repair: A combined theoretical and experimental study. *Angew. Chem. Int. Ed.* **2014,** *53*, 10044–10048.

[57] Zhu, C.; Lu, L.; Zhang, J.; Yue, Z.; Song, J.; Zong, S.; Liu, M.; Stovicek, O.; Gao, Y. Q.; Yi, C. Tautomerization-dependent recognition and excision of oxidation damage in base-excision DNA repair. *Proc. Natl. Acad. Sci. U. S. A.* **2016,** *113*, 7792–7797.

[58] Kanaan, N.; Crehuet, R.; Imhof, P. Mechanism of the glycosidic bond cleavage of mismatched thymine in human thymine DNA glycosylase revealed by classical molecular dynamics and quantum mechanical/molecular mechanical calculations. *J. Phys. Chem. B* **2015,** *119*, 12365–12380.

[59] Gokey, T.; Hang, B.; Guliaev, A. B. Cadmium(II) inhibition of human uracil-DNA glycosylase by catalytic water supplantation. *Sci. Rep.* **2016,** *6*, 39137.

[60] Sowlati-Hashjin, S.; Wetmore, S. D. Computational investigation of glycosylase and β-lyase activity facilitated by proline: Applications to FPG and comparisons to hOgg1. *J. Phys. Chem. B* **2014,** *118*, 14566–14577.

[61] Sowlati-Hashjin, S.; Wetmore, S. D. Quantum mechanical study of the β- and δ-lyase reactions during the base excision repair process: Application to Fpg. *Phys. Chem. Chem. Phys.* **2015,** *17*, 24696–24706.

[62] Alshykhly, O. R.; Fleming, A. M.; Burrows, C. J. Guanine oxidation product 5-carboxamido-5-formamido-2-iminohydantoin induces mutations when bypassed by DNA polymerases and is a substrate for base excision repair. *Chem. Res. Toxicol.* **2015,** *28*, 1861–1871.

[63] Arivett, B.; Farone, M.; Masiragani, R.; Burden, A.; Judge, S.; Osinloye, A.; Minici, C.; Degano, M.; Robinson, M.; Kline, P. Characterization of inosine-uridine nucleoside hydrolase (RihC) from *Escherichia* coli. *Biochim. Biophys. Acta.* **2014,** *1844*, 656–662.

[64] Cui, L. W.; Rajasekariah, G. R.; Martin, S. K. A nonspecific nucleoside hydrolase from *Leishmania donovani*: Implications for purine salvage by the parasite. *Gene* **2001,** *280*, 153–162.

[65] Wink, P. L.; Sanchez Quitian, Z. A.; Rosado, L. A.; Rodrigues Junior, V. d. S.; Petersen, G. O.; Lorenzini, D. M.; Lipinski-Paes, T.; Saraiva Macedo Timmers, L. F.; de Souza, O. N.; Basso, L. A., et al. Biochemical characterization of recombinant nucleoside hydrolase from *Mycobacterium tuberculosis* H37Rv. *Arch. Biochem. Biophys.* **2013,** *538*, 80–94.

[66] Porcelli, M.; Peiluso, I.; Marabotti, A.; Facchiano, A.; Cacciapuoti, G. Biochemical characterization and homology modeling of a purine-specific ribonucleoside hydrolase from the archaeon *Sulfolobus solfataricus*: Insights into mechanisms of protein stabilization. *Arch. Biochem. Biophys.* **2009,** *483*, 55–65.

[67] Porcelli, M.; De Leo, E.; Marabotti, A.; Cacciapuoti, G. Site-directed mutagenesis gives insights into substrate specificity of *Sulfolobus solfataricus* purine-specific nucleoside hydrolase. *Ann. Microbiol.* **2012,** *62*, 881–887.

[68] Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W., et al. *Amber Tools*, Version 1.0; University of California: San Francisco, 2008.

[69] Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct. Funct. Bioinf.* **2006,** *65*, 712–725.

[70] Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* **2004,** *25*, 1157–1174.

[71] Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born. *J. Chem. Theory Comput.* **2012,** *8*, 1542–1555.

[72] Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *J. Chem. Theory Comput.* **2013,** *9*, 3878–3888.

[73] Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **2013,** *9*, 3084–3095.

[74] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A., et al. *Gaussian 09*, Revision C.01 and D.01; Gaussian, Inc.: Wallingford CT, 2009.

[75] Kellie, J. L.; Wetmore, S. D. Selecting DFT methods for use in optimizations of enzyme active sites: Applications to ONIOM treatments of DNA glycosylases. *Can. J. Chem.* **2013,** *91*, 559–572.

[76] Josa, D.; Rodríguez-Otero, J.; Cabaleiro-Lago, E. M.; Rellán-Piñeiro, M. Analysis of the performance of DFT-D, M05-2X and M06-2X functionals for studying π···π interactions. *Chemical Physics Letters* **2013,** *557*, 170–175.

[77] Dokainish, H. M.; Yamada, D.; Iwata, T.; Kandori, H.; Kitao, A., Electron fate and mutational robustness in the mechanism of (6-4)photolyase-mediated DNA repair. *ACS Catalysis* **2017,** *7*, 4835–4845.

[78] Wang, G.; Chen, Z.; Xu, Z.; Wang, J.; Yang, Y.; Cai, T.; Shi, J.; Zhu, W., Stability and characteristics of the halogen bonding interaction in an anion–anion complex: A computational chemistry study. *J. Phys. Chem. B* **2016,** *120*, 610–620.

[79] Zhuang, C.-B.; Zheng, Q.-C., QM/MM calculations and MD simulations of acetolactate decarboxylase to reveal substrate R/S-acetolactate binding mode and stereoselective catalytic mechansim. *RSC Advances* **2016,** *6*, 9185–91859.

[80] Wang, X. Q.; Hirao, H. ONIOM (DFT:MM) study of the catalytic mechanism of myo-inositol monophosphatase: Essential role of water in enzyme catalysis in the two-metal mechanism. *J. Phys. Chem. B* **2013,** *117*, 833–842.

[81] Abdel-Azeim, S.; Jedidi, A.; Eppinger, J.; Cavallo, L. Mechanistic insights into the reductive dehydroxylation pathway for the biosynthesis of isoprenoids promoted by the IspH enzyme. *Chem. Sci.* **2015,** *6*, 5643–5651.

[82] Geronimo, I.; Paneth, P. A DFT and ONIOM study of C–H hydroxylation catalyzed by nitrobenzene 1, 2-dioxygenase. *Phys. Chem. Chem. Phys.* **2014,** *16*, 13889–13899.

[83] Johnson, E. R.; Keinan, S.; Mori-Sanchez, P.; Contreras-Garcia, J.; Cohen, A. J.; Yang, W. T. Revealing noncovalent interactions. *J. Am. Chem. Soc.* **2010,** *132*, 6498–6506.

[84] Contreras-Garcia, J.; Johnson, E. R.; Keinan, S.; Chaudret, R.; Piquemal, J. P.; Beratan, D. N.; Yang, W. T. NCIPLOT: A program for plotting noncovalent interaction regions. *J. Chem. Theory Comput.* **2011,** *7*, 625–632.

[85] Hendershot, J. M.; Wolfe, A. E.; O'Brien, P. J. Substitution of active site tyrosines with tryptophan alters the free energy for nucleotide flipping by human alkyladenine DNA glycosylase. *Biochemistry* **2011,** *50*, 1864–1874.

[86] Connor, E. E.; Wyatt, M. D. Active-site clashes prevent the human 3-methyladenine DNA glycosylase from improperly removing bases. *Chem. Biol.* **2002,** *9*, 1033–1041.

[87] Lenz, S. A. P.; Kellie, J. L.; Wetmore, S. D., Glycosidic bond cleavage in DNA nucleosides: Effect of nucleobase damage and activation on the mechanism and barrier. *J. Phys. Chem. B* **2015,** *119*. 15601–15612.

Chapter 4: Hydrolytic Glycosidic Bond Cleavage in RNA Nucleosides: Effects of the 2′–Hydroxy Group and Acid–Base Catalysis[a,b,c]

4.1 Introduction

DNA and RNA are stable nucleic acid polymers that participate in a wide variety of cellular functions, including information storage and protein synthesis.[1-8] Cleavage of the glycosidic linkage between a nucleobase and the sugar–phosphate backbone occurs during many cellular pathways. For example, deglycosylation is a step in base excision repair of damaged DNA initiated by DNA glycosylases,[2-4] ribosome inactivation catalyzed by RNA glycosidases,[5-6] and purine or pyrimidine salvage facilitated by RNA nucleoside hydrolases (NH).[7-8] Furthermore, these pathways are often essential for organism survival. For example, breakdown of the base excision repair pathway has been linked to increased incidences of many different types of cancer[2] and ricin-A enzymes that catalyze ribosome inactivation have been identified as potent poisons.[6,9] Similarly, purine or pyrimidine salvage is an important biological process in several organisms including parasitic protozoa, which cause malaria,[10] African and American trypanosomiasis,[7,11-12] and leishmaniasis.[12-13]

Since the glycosidic bonds in nucleic acids are inherently stable, a number of experimental studies have focused on understanding how glycosidic bond cleavage is catalyzed by critical cellular machinery including DNA glycosylases,[14-18] RNA glycosidases,[9,19-21] and RNA NH.[22-25] The associated deglycosylation reactions entail attack by a nucleophile (water or amine) at the anomeric (C1′) carbon, and nucleobase departure. These enzymes have been proposed to utilize diverse catalytic strategies including: 1)

[a] *The Journal of Physical Chemistry B* reference style used in this Chapter.

[b] Reprinted with permission from Lenz, S.A.P., Kohout, J.D., and Wetmore, S.D. Hydrolytic Glycosidic Bond Cleavage in RNA Nucleosides: Effects of the 2′–Hydroxy Group and Acid–Base Catalysis, *J. Phys. Chem. B,* ***2016,*** *120,* 12795-12806.

[c] J.D.W performed preliminary calculations on $HCOO^{-}\bullet\bullet\bullet H_2O$ and OH^{-} mediated deglycosylation. S.A.P.L performed all remaining calculations and data analysis. S.A.P.L. and S.D.W. contributed to the design of calculations, interpretation of data, and writing the manuscript.

stabilization of the negative charge developing on the nucleobase via protonation,[14-15] hydrogen bonding,[9,16,20-21] or π–π interactions;[17,22] 2) stabilization of the positive charge developing on the sugar group via strategic positioning of basic residues (Asp/Glu);[23-24] and 3) activation of the nucleophile that displaces the nucleobase.[18-19,25] While it is clear that each of these strategies can reduce an otherwise prohibitive deglycosylation barrier, the catalytic contribution afforded by each avenue may be dependent on the substrate targeted (e.g., damaged versus undamaged nucleobase, deoxyribose versus ribose sugar, nucleoside versus nucleotide), and the enzyme involved. For example, the DNA glycosylase MutY and RNA hydrolase IU-NH may facilitate deglycosylation via nucleobase protonation,[14-15,23] while DNA-targeting UDG and RNA-targeting CU-NH may achieve catalysis via hydrogen bonding between the departing nucleobase and catalytic active site residues.[16,26-27] On the other hand, AAG (DNA glycosylase) and IAG-NH (RNA NH) have been proposed to stabilize the leaving group via π–π interactions between the departing nucleobase and aromatic amino acids.[17,22,28] To understand the driving force for an enzyme to take advantage of different catalytic approaches, the relative effect of each strategy or combination of strategies on the reaction must be determined.

With the goal to elucidate the general impact of various factors on DNA deglycosylation, our group previously used carefully designed computational models that are not specific to a particular enzyme.[29-37] Initially, the concerted (S_N2) deglycosylation of the uracil nucleoside (a common type of DNA damage) was considered, including the effects of hydrogen-bonding interactions between small molecules (hydrogen fluoride, water, or ammonia) and different acceptor sites of the nucleobase.[29] In a follow-up study, the relative stability of the glycosidic bond in the natural DNA nucleosides was determined, with and without hydrogen-bonding interactions between small molecules and the nucleobases.[30,34-35] The computational model was also scrutinized by examining whether the inclusion of a phosphate moiety changes the reaction pathway (i.e., nucleoside versus nucleotide models), and how the

inclusion of implicit solvent within the optimization routine affects the structures of stationary points and reaction energetics.[35] Furthermore, the effects of different nucleophiles (i.e., amine, proline, or water),[33,38-39] mechanisms (i.e., S_N1 versus S_N2),[31-32,37] and damaged (e.g., 8-oxoguanine and thymine glycol)[31,33,36-37] or activated (i.e., via hydrogen bonding, protonation, or π–π interactions)[34,37] DNA nucleobases were considered. These works were augmented by other contributions in the literature that used small models to investigate the impact of nucleobase oxidation[40-42] and/or activation (i.e., protonation or metal ions)[40-44] on DNA deglycosylation, as well as larger protein–DNA models to focus on specific enzymes (i.e., DNA glycosylases).[45-59]

Despite detailed studies on DNA deglycosylation, little is known about the corresponding reaction in RNA. Although computational work has considered select aspects of specific enzyme-catalyzed reactions,[28, 60-69] such studies do not reveal the intrinsic stability of the glycosidic bond in RNA or comprehensively isolate ways the associated barrier can be reduced across all natural RNA nucleosides. Thus, while enzymes such as RNA glycosidases and nucleoside hydrolases may use at least some of the catalytic strategies outlined above to facilitate deglycosylation, the individual contributions of each approach are currently unclear. Furthermore, while an experimental study has reported greater stability of the glycosidic bond in RNA versus DNA purines under acidic conditions (pH=1), there is currently no comparable data for the pyrimidines nor a comparison of acid and base catalysis.[70] Therefore, the intrinsic effect of 2′–OH on the stability of the glycosidic bond is unknown.

Since work on DNA has shown that small model computational studies can provide important information about nucleic acid deglycosylation and thereby expand our fundamental understanding of the corresponding enzymatic reactions,[29-37] the present work uses computational methods to comprehensively study the hydrolysis of the four canonical RNA nucleosides (A, C, G, and U, Figure 4.1). Specifically, we first examine deglycosylation using formate activated water ($HCOO^-\bullet\bullet\bullet H_2O$) as the nucleophile, since

formate contains the functionality of the catalytic Asp/Glu residue present in the active site of many nucleic acid-targeting enzymes (e.g., DNA glycosylases,[14-18,71] RNA glycosidases,[6,9,19-21] and RNA NH).[22-25,72] Using this model, we compare RNA and DNA hydrolysis to determine the impact of 2′–OH on both the structures of the stationary points and the reaction energetics. Subsequently, to quantify the effects of base and/or acid catalysis, we characterize uncatalyzed RNA nucleoside hydrolysis using a single (unactivated) water molecule as the nucleophile. Since glycosidic bond cleavage barriers have been shown to be dependent on the species accepting a proton from the nucleophilic water,[29] stronger base catalysis than provided by formate is then examined using hydroxide (OH^-) as the nucleophile, which represents the extreme of full water activation. To solely consider acid catalysis, we sequentially protonate each hydrogen-bond acceptor site in the RNA nucleobases using a model containing an unactivated H_2O nucleophile. Finally, to determine whether the effects of base and acid catalysis are additive, we examine the excision of protonated nucleobases facilitated by activated water (i.e., the $HCOO^-\bullet\bullet\bullet H_2O$ or OH^- nucleophile). The current study uncovers the inherent stability of the glycosidic bond in the natural RNA nucleosides, as well as the stability relative to the corresponding DNA analogues. Furthermore, the effects of the reaction environment (i.e., neutral versus basic versus acidic conditions) on the deglycosylation reaction are also revealed. When combined with previous computational and experimental studies, the current work yields valuable insights into strategies that can be exploited by enzymes that target RNA and provides a starting point for developing larger computational models to study enzyme-catalyzed RNA deglycosylation. Furthermore, the present work provides structural insights that can be exploited in the future development of drugs that combat disease by inhibiting RNA-targeting enzymes (e.g., block RNA NH to treat trypanosomatid diseases,[7,11-12] such as American and African trypanosomiasis, and leshmaniasis).

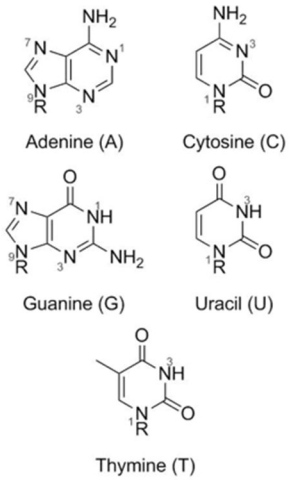


Figure 4.1. Structure and chemical numbering of the canonical RNA (R=ribose) and DNA (R=deoxyribose) nucleosides.

4.2 Computational Methods

Due to the insights acquired from previous computational studies of DNA deglycosylation,[9,14-25] an analogous methodology was used in the present work to gain a fundamental understanding of RNA hydrolysis. This approach will also afford an accurate comparison of RNA and DNA deglycosylation, which was one of the driving forces behind the present study. Specifically, our computational models were built by adding a 2′-hydroxy group to the previously reported[35] DNA nucleoside models that include a $HCOO^-$ •••H_2O nucleophile (Figure 4.2a). The formate in the previous DNA models represents the carboxylate sidechain of Asp/Glu in the active site of many DNA glycosylases that activates the water nucleophile.[14-18] This nucleophile model was retained in the present work due to the similar role Asp/Glu may play in RNA NH and glycosidase-catalyzed reactions.[9,19-25] Additionally, we maintained the methyl caps on the 3′ and 5′ oxygen atoms instead of hydrogen atoms in order to eliminate hydrogen bonding between these groups and the departing nucleobase that is unlikely to occur in enzymatic systems. For example, the 5′- and 3′-hydroxy groups interact with active site residues when RNA substrates are bound to IAG-NH and CU-NH.[24,73] Furthermore, although other RNA-targeting enzymes (such as the ricin-A chain) bind to ribosomal RNA rather than free nucleosides,[74-75] the differences between structures and barriers calculated

using small models with methyl and phosphate caps are minimal for DNA hydrolysis,[35] and therefore our models are also applicable to enzymes that bind to RNA polymers rather than free nucleosides or nucleotides.

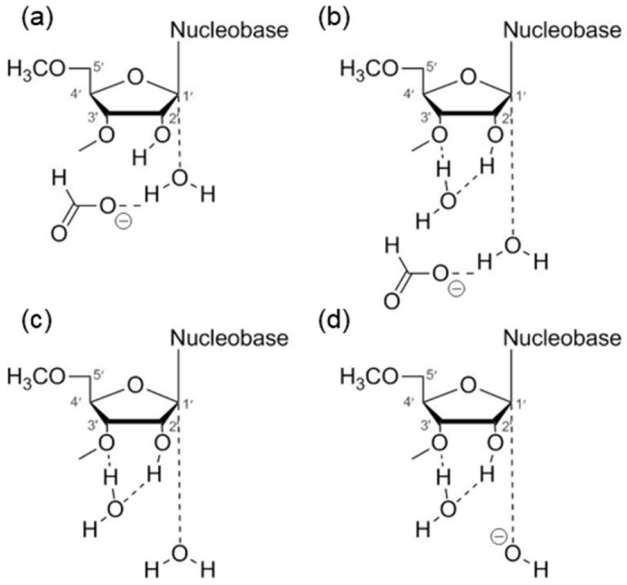


Figure 4.2. Models used in the present study to examine RNA nucleoside deglycosylation.

Using the model described above (Figure 4.2a), a hydrogen bond forms between the $HCOO^-$ anion and 2′–OH (Figure C.1 in Appendix C). This interaction is unlikely to occur in an enzyme-catalyzed mechanism since the 2′–hydroxy group typically interacts with other active site residues. For example, in RNA NH, the 2′–OH oxygen atom is coordinated to a Ca^{2+} ion, while the hydrogen atom is typically occupied by interactions with either another Asp/Glu (that does not activate the nucleophile) or a backbone carbonyl of a neighbouring residue.[26,72] In the case of the ricin-A chain, the 2′-hydroxy group of the bound adenosine monophosphate substrate is in position to donate a hydrogen bond to an active site tyrosine residue.[75] Therefore, to occupy 2′–OH in our models, an explicit water molecule was introduced that bridges between O2′ and O3′ (Figure 4.2b). This choice is also chemically reasonable since a water molecule could occupy this position under experimental conditions used to study the nonenzymatic

hydrolysis reactions. Furthermore, since a previous study from our group determined that the inclusion of solvent in the optimization routine yields structures and barriers in better agreement with experiment,[35] we used the IEF-PCM implicit solvent model with a dielectric constant of water (78) to allow closer comparisons to the experimental data obtained for the hydrolysis reactions.

To examine the effect of base and/or acid catalysis on the reaction barrier, we implemented several small models. First, to provide a point of reference to determine catalytic effects, we characterized the uncatalyzed pathways using an unactivated water nucleophile (Figure 4.2c). Subsequently, to examine the effect of increasing nucleophile activation, we examined pathways catalyzed by fully-activated water (i.e., OH^- nucleophile, Figure 4.2d). Finally, to evaluate the effect of acid catalysis or the combined effect of acid–base catalysis, the hydrogen-bond acceptor sites on the nucleobases (Figure 4.1) were sequentially protonated in models containing an unactivated or activated ($HCOO^-\bullet\bullet\bullet H_2O$ or OH^-) water nucleophile.

In each model considered, S_N2 transition states for RNA nucleoside deglycosylation were optimized at the B3LYP/6-31G(d) level in implicit water (IEF-PCM) using Gaussian 09 default convergence criteria. The choice to solely examine S_N2 mechanisms stems from the small barrier difference previously reported for S_N1 and S_N2 deglycosylation in DNA nucleosides,[37] and this choice allows direct comparisons to our corresponding DNA hydrolysis studies.[35] Subsequently, intrinsic reaction coordinate (IRC) calculations were performed to confirm the connectivity of the transition state (TS) to the corresponding reactant (RC) and product (PC) complexes. The RC and PC were fully optimized to minima, and the identity of each stationary point (RC, TS, and PC) was confirmed through frequency calculations. Scaled (0.9806)[76] zero-point and thermal corrections were obtained from harmonic frequencies performed with IEF-PCM-B3LYP/6-31G(d), and applied to SMD-B3LYP-D3/6-311+G(2df,2p) single-point energies to yield the reported Gibbs energies. The SMD implicit solvation model was implemented since this methodology has been shown to provide more accurate solvated Gibbs

energies.[77] All calculations were performed with the Gaussian 09 program suite (revision A.02, C.01, or D.01).[78]

4.3 Results and Discussion

4.3.1 Hydrolysis Facilitated by the $HCOO^{-}\bullet\bullet\bullet H_2O$ Nucleophile

RNA Deglycosylation. As discussed in the Computational Methods, $HCOO^{-}$ in our initial model (Figure 4.2b) represents the sidechain of an Asp or Glu that is proposed to be the general base in reactions catalyzed by RNA-targeting nucleoside hydrolases[8,22-25] and glycosidases.[6,9,19-21] Structures and key geometric parameters along the $HCOO^{-}\bullet\bullet\bullet H_2O$ hydrolysis pathway of RNA nucleosides are shown in Figure 4.3a, while detailed geometric data is provided in Tables C.1–C.4 (see Appendix C). Regardless of the RNA nucleoside considered, the nucleophilic ($C1'–O_{wat}$) distance shortens, while the glycosidic bond (C1′–N1 for pyrimidines or C1′–N9 for purines) distance elongates until the nucleobase is fully dissociated from the sugar moiety, and a covalent bond is formed between C1′ and the nucleophile. The hydrogen-bond distance between 2′–OH and the bridging water molecule ($O2'–H2'\bullet\bullet\bullet O_{bw}$) is ~1.800 Å ($\angle(H2'O2'O_{bw})$: ~165°) throughout the reaction (Figure 4.3a), which indicates that inclusion of this bridging water will not significantly affect the reported barriers and reaction energies (rxn).

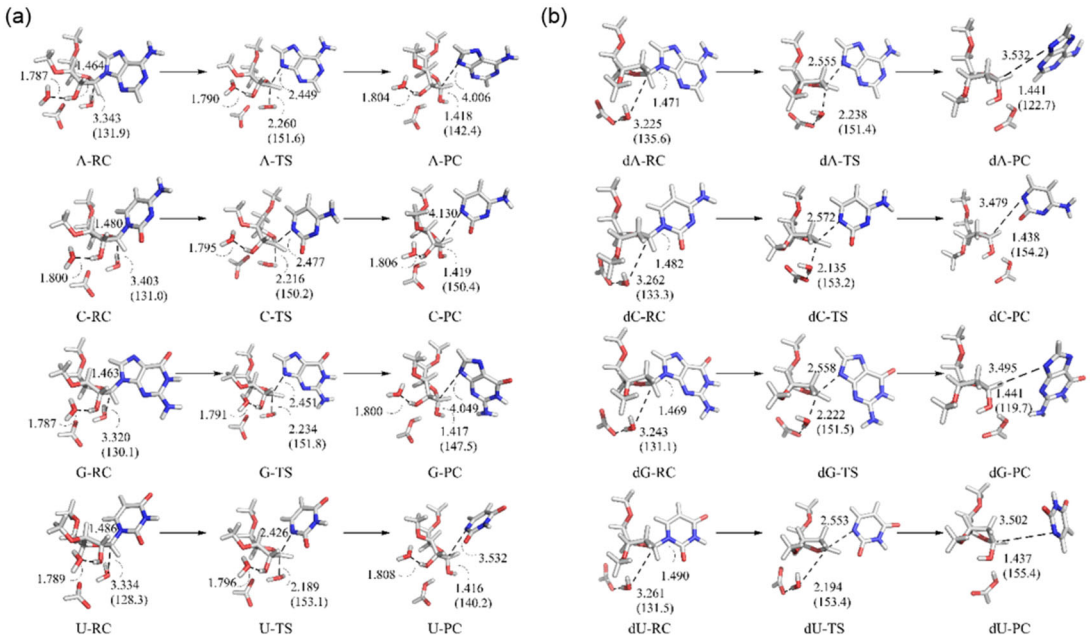


Figure 4.3. Selected B3LYP/6-31G(d) bond lengths (Å) and angles (degrees in parentheses) in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of adenine (A), cytosine (C), guanine (G), and uracil (U) containing (a) RNA and (b) DNA nucleosides facilitated by the $HCOO^{-}\bullet\bullet\bullet H_2O$ nucleophile.

In the RC, the nucleophile is 3.300–3.400 Å away from C1′ and the reaction angle [∠(O_{wat}C1′N1/N9)] is near 130° for each canonical RNA nucleoside (Figure 4.3a). The water nucleophile forms hydrogen bonds with both O4′ and $HCOO^{-}$, while $HCOO^{-}$ also hydrogen bonds with H4′ of the sugar (Tables C.1–C.4 in Appendix C). Therefore, the water nucleophile is well positioned for S_N2 attack. The glycosidic bond length in the TS elongates to 2.400–2.500 Å, while the nucleophilic distance shortens to ~2.200 Å. As the nucleophile approaches the site of reaction, the reaction angle increases to 160°. The nucleobase is fully dissociated in the PC, with a glycosidic bond of 3.500–4.000 Å. Hydrogen bonding is observed between the departed nucleobase and the sugar–nucleophile moiety, which stabilizes the negative charge on the nucleobase (Figure 4.3a). These hydrogen bonds change depending on the identity of the nucleobase, with N3 of the purines or O2 of the pyrimidines interacting with H1′. Full proton transfer

from the water nucleophile to $HCOO^-$ is observed for each reaction, with the O_{wat}–H_{wat}•••O_{HCOO-} distance shortening from ~1.800 Å in the RC to 1.000 Å in the PC (Tables C.1–C.4 in Appendix C).

The reaction barriers and energies for the hydrolytic deglycosylation of the RNA nucleosides range from 152.9–185.6 kJ mol^{-1} (Table 4.1). U has the most labile glycosidic bond (152.9 kJ mol^{-1}), and C the most stable glycosidic bond (185.6 kJ mol^{-1}), while deglycosylation of the purines results in approximately equivalent barrier heights (within 6 kJ mol^{-1}, Table 4.1). This correlates with the hydrolysis of U having the earliest TS (C1′–N1: 2.426 Å; C1′–O_{wat}: 2.189 Å), and C being a poorer leaving group with the latest TS (C1′–N1: 2.477 Å; C1′–O_{wat}: 2.216 Å), as well as the hydrolysis of the purines resulting in nearly identical coordinates for the bond forming and breaking events These results are also supported by the greater calculated N1 acidity for the U compared to C nucleobase, and the near equivalent N9 acidities of the A and G nucleobases.[37] Although the reactions are endothermic (by 55.7–98.1 kJ mol^{-1}), the reaction energies are substantially lower than the barrier heights (by ~100 kJ mol^{-1}), which at least in part reflects stability provided by the hydrogen bonds between the nucleobase and sugar–nucleophile moiety in the PC.

Comparison of RNA and DNA Deglycosylation. In order to determine the effect of the 2′–hydroxy group on nucleic acid hydrolysis, we compare RNA and DNA nucleoside deglycosylation facilitated by the $HCOO^-$•••H_2O nucleophile. To the best of our knowledge, no direct comparisons of the hydrolysis reactions for all canonical RNA and DNA nucleosides has been made to date. Stationary points along the S_N2 reaction pathways of DNA (A, C, G, and U) deglycosylation catalyzed by $HCOO^-$•••H_2O are shown in Figure 4.3b.[35] Although not a canonical DNA nucleobase, U is a common type of DNA oxidation damage. Since a previous study found little difference in the reaction energetics (<5 kJ mol^{-1} difference in the barrier, and <0.1 kJ mol^{-1} difference in the reaction energy) for the U and T DNA hydrolysis pathways,[35] a comparison between the hydrolysis of U in RNA and DNA is made in the present work to permit focus

on the effect of 2′–OH. Additionally, we consider $HCOO^{-}\bullet\bullet\bullet H_2O$ catalyzed pathways since this model is widely applicable to both DNA (e.g., MutY,[79] UDG,[80] or AAG[71]) and RNA (e.g., IAG-NH[22] or ricin[9]) targeting enzymes.

Table 4.1. Comparison of the Calculated Gibbs Barriers and Reaction Energies (kJ mol^{-1}) for the Deglycosylation of Ribose and Deoxyribose Nucleosides Facilitated by the $HCOO^{-}\bullet\bullet\bullet H_2O$ Nucleophile[a]

	RNA		DNA[b]		$\Delta_{2'OH}$[c]	
nucleoside	barrier	rxn	barrier	rxn	barrier	rxn
A	170.9	79.8	137.8	51.4	33.0	–28.3
C	185.6	98.1	148.1	85.0	37.6	–13.1
G	176.9	79.0	140.1	68.9	36.8	–10.0
U	152.9	55.7	119.0	51.4	33.8	–4.3

[a]Relative Gibbs energies obtained from SMD-B3LYP-D3/6-311+G(2df,2p)//IEF-PCM-B3LYP/6-31G(d, including scaled (0.9806) zero-point vibrational and thermal corrections. [b]Structures obtained from ref. 35. [c]The effect of the 2′–OH as determined by the difference between the RNA and DNA deglycosylation barriers or reaction energies.

In both RNA and DNA hydrolysis, the placement of the water nucleophile is nearly equivalent in the RC, with a C1′–O_{wat} distance of ~3.200–3.400 Å. However, the nucleophile orientation changes slightly since water hydrogen bonds with O3′ in DNA and O4′ in RNA (Figure 4.3 and Tables C.1–C.5 in Appendix C). Nevertheless, $HCOO^{-}$ hydrogen bonds with H4′ in both DNA and RNA pathways. The DNA deglycosylation reactions have slightly later TS than RNA deglycosylation, with DNA glycosidic bond lengths being ~0.100 Å greater despite nearly equivalent nucleophile distances (Figure 4.3). A contributing factor to the earlier RNA TS is likely stabilization of the positive charge developing on the sugar by charge donation from the 2′–hydroxy group. As discussed for RNA PC, significant hydrogen bonding occurs between the departing nucleobase (N3 for purines and O2 for pyrimidines) and the sugar–nucleophile moiety (H1′) in DNA PC, which stabilizes the negatively charged nucleobase. However, in the PC for the DNA purines, 1′–OH forms a hydrogen bond with O3′, which shifts the position of HCOOH

to interact with the departed nucleobase, either through the exocyclic amino group of G or H2 of A. These interactions are not seen in RNA PC since the 2′–hydroxyl group prevents 1′–OH from interacting with O3′ and HCOOH maintains a consistent position throughout the reaction. Thus, although the 2′–hydroxy group overall minimally affects the structure of the stationary points, discrete differences prevail between DNA and RNA.

The difference in the RNA and DNA hydrolysis barriers is ~35–40 kJ mol^{-1} ($\Delta_{2'OH}$, barrier, Table 4.1), with the barriers being lower for DNA deglycosylation. Our results are in qualitative agreement with experimental data suggesting that the addition of 2′–OH confers glycosidic stability to RNA purine nucleosides over their DNA counterparts by ~20 kJ mol^{-1}.[70] Differences between the magnitude of the calculated and experimentally-estimated effects of the 2′-hydroxy group arise since the experiments were performed in acidic conditions (pH=1), while our calculations considered base (formate) catalysis. Regardless, the increased stability of RNA over DNA nucleosides may be an evolutionary advantage for organisms. Specifically, RNA is more often exposed to the cellular environment compared to DNA, and therefore the greater stability of the glycosidic bond may be required to prevent premature RNA degradation. Interestingly, the magnitude of the increase in barrier due to 2′–OH depends on the identity of the nucleobase, with an up to 4.6 kJ mol^{-1} deviation ($\Delta_{2'OH}$, Table 4.1). This phenomenon was also observed in the experimental investigation of RNA purine deglycosylation in acidic conditions.[70] Indeed, the previously reported up to 5.5 kJ mol^{-1} increase in hydrolysis barriers for RNA over DNA is in near quantitative agreement with our calculated difference of 3.8 kJ mol^{-1} for the purines ($\Delta_{2'OH}$, Table 4.1). Therefore, 2′–OH can alter both the structures and energies of the stationary points for nucleic acid hydrolysis. Most importantly, since the calculated barriers for RNA hydrolysis by a partially-activated water molecule ($HCOO^{-}\bullet\bullet\bullet H_2O$) are high (>150 kJ mol^{-1}), enzymes that target RNA must use other strategies

to facilitate deglycosylation, and therefore the next section will consider the impact of enhanced base catalysis, acid catalysis, and simultaneous acid-base catalysis.

4.3.2 Factors that May Facilitate Hydrolysis

Uncatalyzed Hydrolysis. Prior to considering ways RNA-targeting enzymes may enhance deglycosylation rates, the uncatalyzed reaction (Figure 4.2c) will be discussed to provide a frame of reference for monitoring various catalytic effects. In the RC for uncatalyzed RNA nucleoside hydrolysis, the unactivated H_2O nucleophile is ~3.400–3.500 Å from C1′ and the [∠(O_{wat}C1′N1/N9)] reaction angle is 125° for each pathway (Figure 4.4). The nucleophile forms an O_{wat}–H_{wat}•••O2′ hydrogen bond with the 2′–hydroxy group, which has a length of 1.833–1.837 Å and an angle of ~170° (Tables C.6–C.9 in Appendix C). Additionally, the O2′–H2′•••O_{bw} hydrogen bond is maintained at ~1.745 Å (angle of ~165°, Figure 4.4 and Tables C.6–C.9 in Appendix C), which is consistent with the $HCOO^-$•••H_2O pathways. Later TS relative to the $HCOO^-$•••H_2O pathways are observed, with the glycosidic bond lengthening to ~2.650 Å, and the nucleophile distance shortening to ~2.000 Å. The hydrogen bonds involving the discrete water molecules (O_{wat}–H_{wat}•••O2′ and O2′–H2′•••O_{bw}) decrease in length going from the RC to the TS by ~0.140 and 0.040 Å, respectively (Tables C.6–C.9 in Appendix C). The tightening of these hydrogen bonds likely reflects the lack of charge stabilization of the sugar cation forming in the TS in this model. Indeed, the O_{wat}–H_{wat}•••O2′ hydrogen bond does not form, and the O2′–H2′•••O_{bw} hydrogen bond maintains a consistent distance (within 0.030 Å), across the $HCOO^-$•••H_2O pathways. In the PC, the associated nucleophile has partially donated a proton to the 2′–hydroxy group, with the O_{wat}–H_{wat}•••O2′ hydrogen bond shrinking to ~1.570 Å and an angle of ~130° (Figure 4.4 and Tables C.6–C.9 in Appendix C). In these complexes, the nucleobase dissociates such that the distance between N1 (pyrimidines) or N9 (purines) and C1′ is 3.000–3.500 Å, which is shorter than observed for $HCOO^-$•••H_2O catalyzed

hydrolysis (3.400–4.000 Å). Nevertheless, there are similar hydrogen bonds between the nucleobase and sugar–nucleophile moiety in the PC for the two nucleophiles.

The barrier heights for the uncatalyzed reactions range from 168.2–202.4 kJ mol^{-1} (H_2O, Table 4.2), which is 15–20 kJ mol^{-1} higher than the barriers for the $HCOO^-\bullet\bullet\bullet H_2O$ catalyzed pathways ($\Delta_{base}(HCOO^-\bullet\bullet\bullet H_2O)$). U exhibits the least nucleophile dependence (15.3 kJ mol^{-1} difference), while A exhibits the greatest nucleophile dependence (18.3 kJ mol^{-1}). The higher calculated barriers correlate with the later TS for the uncatalyzed pathways, and further illustrate that a measure of catalysis is required for RNA hydrolysis. Consistent with the $HCOO^-\bullet\bullet\bullet H_2O$ catalyzed pathways, the barrier height decreases according to the nucleobase as C > A = G > U, which further highlights differences in the intrinsic stability of the glycosidic bond across the RNA nucleosides. Additionally, the reaction energies are only ~10 kJ mol^{-1} more stable than the TS for uncatalyzed hydrolysis, indicating that the PC are transient and re-emphasizing the necessity of catalysis. This contrasts the ~100 kJ mol^{-1} in stability of the $HCOO^-\bullet\bullet\bullet H_2O$ PC relative to the corresponding TS.

Table 4.2. Calculated Gibbs Barriers and Reaction Energies (kJ mol^{-1}) for the Deglycosylation of Natural and Corresponding Protonated (Acid–catalyzed) RNA Nucleosides Facilitated by Various (H_2O, OH^-, or $HCOO^-\bullet\bullet\bullet H_2O$) Nucleophiles[a]

	H_2O			OH^-						$HCOO^-\bullet\bullet\bullet H_2O$						
nucleoside	barrier	Δ_{acid}[b]	rxn	barrier	Δ_{base}[c]	Δ_{acid}[b]	Δ_{sum}[d]	$\Delta_{acid-base}$[e]	$\Delta_{sum}-\Delta_{acid-base}$[f]	rxn	barrier	Δ_{base}[c]	Δ_{acid}[b]	Δ_{sum}[d]	$\Delta_{acid-base}$[e]	$\Delta_{sum}-\Delta_{acid-base}$[f]	rxn
A	189.2	–	183.3	145.8	–43.4	–	–43.4	–43.4	–	–59.9	170.9	–18.3	–	–18.3	–18.3	–	79.8
A–N1H	151.0	–38.2	148.2	112.0	–39.0	–33.8	–81.6	–77.2	–4.4	–99.1	136.0	–14.9	–34.8	–56.6	–53.2	–3.4	34.9
A–N3H	147.5	–41.7	127.2	99.5	–48.0	–46.3	–85.1	–89.7	4.6	–119.9	128.7	–18.8	–42.2	–60.1	–60.6	0.5	26.1
A–N6H	155.9	–33.3	150.2	110.0	–45.9	–35.8	–76.7	–79.2	2.5	–99.4	143.3	–12.7	–27.6	–51.6	–45.9	–5.7	39.6
A–N7H	119.6	–69.6	115.4	80.4	–39.2	–65.5	–113.0	–108.9	–4.2	–147.1	106.5	–13.0	–64.3	–88.0	–82.7	–5.3	9.0
C	202.4	–	190.5	139.9	–62.5	–	–62.5	–62.5	–	–48.9	185.6	–16.8	–	–16.8	–16.8	–	98.1
C–N3H	135.8	–66.6	127.0	92.6	–43.3	–47.3	–129.1	–109.8	–19.3	–114.8	117.5	–18.4	–68.2	–83.3	–84.9	1.6	26.7
C–N4H	151.1	–51.3	144.2	95.9	–55.2	–44.0	–113.8	–106.5	–7.3	–104.5	131.7	–19.5	–54.0	–68.0	–70.7	2.7	42.8
C–O2H(N1)	119.9	–82.5	117.9	81.9	–38.0	–57.9	–145.0	–120.5	–24.5	–143.5	105.8	–14.1	–79.8	–99.2	–96.6	–2.7	12.3
C–O2H(N3)	121.8	–80.6	116.0	78.5	–43.3	–61.4	–143.1	–123.9	–19.2	–132.5	105.0	–16.8	–80.6	–97.3	–97.4	–	10.5
G	194.2	–	182.6	145.4	–48.8	–	–48.8	–48.8	–	–55.1	176.9	–17.4	–	–17.4	–17.4	–	79.0
G–N2H	162.4	–31.8	149.3	117.9	–44.5	–27.5	–80.6	–76.3	–4.3	–93.0	142.3	–20.1	–34.5	–49.2	–51.9	2.7	44.5
G–N3H	154.9	–39.3	145.0	95.4	–59.5	–50.0	–88.2	–98.8	10.7	–128.8	120.8	–34.1	–56.0	–56.7	–73.4	16.7	19.9
G–N7H	126.4	–67.9	113.9	88.6	–37.8	–56.8	–116.7	–105.7	–11.1	–139.7	108.1	–18.2	–68.7	–85.3	–86.1	0.9	13.6
G–O6H(N1)	160.2	–34.0	146.9	120.9	–39.3	–24.5	–82.8	–73.3	–9.5	–87.2	144.7	–15.5	–32.1	–51.4	–49.5	–1.9	41.3
G–O6H(N7)	170.7	–23.5	154.2	117.2	–53.5	–28.2	–72.3	–77.0	4.7	–93.7	146.0	–24.7	–30.9	–40.9	–48.2	7.4	44.4
U	168.2	–	155.9	116.5	–51.8	–	–51.8	–51.8	–	–87.5	152.9	–15.3	–	–15.3	–15.3	–	55.7
U–O2H(N1)	96.3	–71.9	92.1	72.9	–23.5	–43.6	–123.7	–95.4	–28.3	–161.4	61.3	–35.0	–91.6	–87.3	–106.9	19.7	–27.9
U–O2H(N3)	89.7	–78.5	83.0	62.1	–27.6	–54.4	–130.3	–106.2	–24.2	–162.5	80.1	–9.5	–72.7	–93.9	–88.1	–5.8	–20.6
U–O4H(N3)	121.1	–47.2	110.6	71.4	–49.6	–45.0	–98.9	–96.8	–2.1	–141.4	105.8	–15.3	–47.1	–62.5	–62.4	–0.1	6.0
U–O4H(C5)	120.9	–47.3	106.5	74.6	–46.2	–41.8	–99.1	–93.6	–5.5	–137.1	107.4	–13.4	–45.5	–62.7	–60.8	–1.9	6.6

[a]Relative Gibbs energies obtained from SMD-B3LYP-D3/6-311+G(2df,2p)//PCM-B3LYP/6-31G(d) calculations, including scaled (0.9806) zero-point vibrational and thermal corrections. [b]The effect of acid catalysis on RNA nucleoside deglycosylation barriers for a given nucleophile. [c]The effect of base catalysis on RNA nucleoside deglycosylation barriers. [d]Sum of barrier reductions afforded by acid (Δ_{acid}, H_2O nucleophile) and base (Δ_{base}(OH^- or $HCOO^-$)) catalysis on RNA nucleoside deglycosylation barriers. [e]The simultaneous effect of acid and base catalysis calculated using models that contain both water (base) and nucleobase (acid) activation. [f]The additivity of the acid and base catalysis calculated as the difference between Δ_{sum} and $\Delta_{acid-base}$.

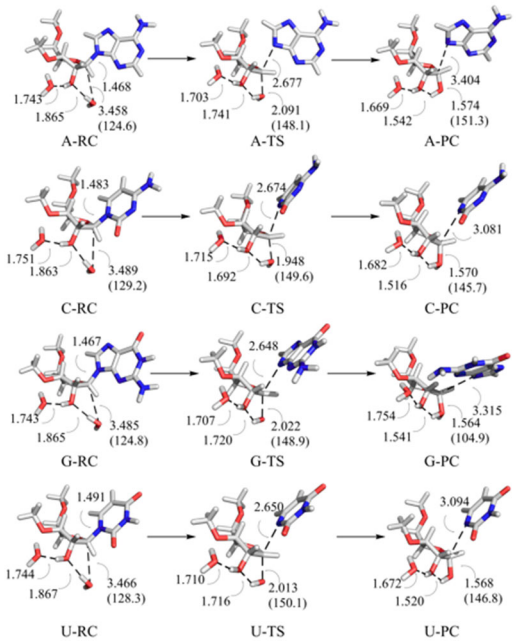


Figure 4.4. Selected B3LYP/6-31G(d) bond lengths (Å) and angles (degrees in parentheses) in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of adenine (A), cytosine (C), guanine (G), and uracil (U) containing RNA nucleosides facilitated by the (unactivated) H_2O nucleophile.

Base Catalysis. Since water nucleophile activation by formate affects the structures and relative energies of stationary points along the uncatalyzed reaction pathway, deglycosylation catalyzed by OH^- is now considered to understand the extreme effect of full water activation (full base catalysis). We note that our pathways do not include the energetic cost associated with deprotonating a water molecule since we are primarily interested in the extreme effect of water activation in the presence of a strong base. Indeed, a previous study compared deglycosylation by water activated using small molecules with a range in basicities (e.g., halogens, $HCOO^-$, and CN^-) and hydrolysis by OH^-, and determined that OH^- yields the lowest deglycosylation barriers.[29] In the OH^- catalyzed pathways, the RC nucleophilic distances are ~2.900–3.100 Å (Figure 4.5 and Tables C.10–C.13 in Appendix C), which is 0.600–0.700 Å shorter than for the

corresponding uncatalyzed reactions. Additionally, the glycosidic bond length in the TS for the OH^- catalyzed pathways is ~1.900–2.000 Å, which represents a 0.500–0.600 Å decrease relative to the uncatalyzed pathways. The OH^- TS also have significantly shorter glycosidic bonds than the $HCOO^-$ •••H_2O catalyzed pathways (by ~0.400 Å). The earlier TS for the OH^- catalyzed reaction likely results from the increased nucleophilic power of fully-activated water over the unactivated water nucleophile. Furthermore, the negative charge of OH^- helps stabilize the positive charge developing on the sugar in the TS. The fact that enhancement in nucleophilic power results in earlier TS has been previously reported for DNA hydrolysis.[30,35] In the PC, the C1′–N1 (C or U) or C1′–N9 (A or G) distance is ~3.500 Å, which is 0.500 Å shorter than observed for the $HCOO^-$•••H_2O catalyzed PC (Figures 4.3 and 4.5). Regardless, the O2′–H2′•••O_{bw} hydrogen bond is maintained, and does not significantly fluctuate along the reaction pathway (Figure 4.5 and Tables C.10–C.13 in Appendix C).

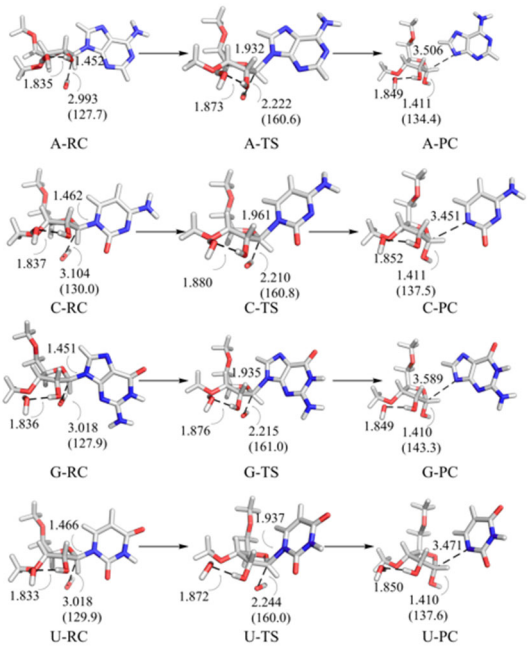


Figure 4.5. Selected B3LYP/6-31G(d) bond lengths (Å) and angles (degrees in parentheses) in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of adenine (A), cytosine (C), guanine (G), and uracil (U) containing RNA nucleosides facilitated by the OH^- nucleophile.

The structural differences between the OH^- catalyzed and uncatalyzed hydrolysis pathways results in ~40–60 kJ mol^{-1} lower barriers when OH^- is the nucleophile ($\Delta_{base}(OH^-)$, Table 4.2). The energetic difference is most pronounced for C (62.5 kJ mol^{-1}) and least pronounced for A (43.4 kJ mol^{-1}), while G (48.8 kJ mol^{-1}) and U (51.4 kJ mol^{-1}) exhibit a similar nucleophile dependence as A. Due to these differences, the relative stability of the glycosidic bond among the canonical RNA nucleosides changes depending on the nucleophile. Specifically, the stability decreases as G = A > C > U when OH^- is the nucleophile, but as C > G > A > U when unactivated H_2O or $HCOO^-\bullet\bullet\bullet H_2O$ is the nucleophile. The reaction energies also change significantly with the level of nucleophile activation. Specifically, the PC are more stable than the RC in the OH^- pathways (reaction energies ~–60 to –90 kJ mol^{-1}, Table 4.2), while

the H_2O (reaction energies ~155 to 190 kJ mol^{-1}, Table 4.2) and $HCOO^{-}\bullet\bullet\bullet H_2O$ (reaction energies ~55 to 100 kJ mol^{-1}, Table 4.2) mediated pathways are endothermic.

The deviations in the reaction pathways and energetics with the nucleophile highlight the importance of nucleophile activation in enzyme-catalyzed RNA deglycosylation, and the possible role of the nucleophile as a rate determinant. Notably, the catalytic rates of RNA nucleoside hydrolases deviate significantly depending on their substrate specificity (i.e., purine specific, pyrimidine specific, or purine–pyrimidine non-specific).[23,81-84] Although the difference in activity may be at least in part attributed to the varied reactivity of the nucleosides, our work shows that the variation in the barrier with nucleoside (up to ~20 kJ mol^{-1}) is less than the variation in the barrier with nucleophile (up to ~60 kJ mol^{-1}, Table 4.2). Therefore, the present work suggests the deviation in catalytic rates exhibited by RNA-targeting enzymes may be a consequence of the position and orientation of the general base, which may affect nucleophile activation and charge stabilization during (for S_N2) or post (for S_N1) nucleobase excision. Nevertheless, despite significant barrier reductions for the fully-activated nucleophile (OH^{-}) over the uncatalyzed deglycosylation pathways, the barriers remain high (up to ~145 kJ mol^{-1}, Table 4.1), and therefore enzymes must use other strategies to further catalyze RNA hydrolysis.

Effects of Acid Catalysis. Direct protonation of the nucleobase has been proposed to facilitate RNA deglycosylation in a number of enzyme-catalyzed reactions. For example, O2 protonation of C by *E. coli* inosine-uridine nucleoside hydrolase (RihC) has been proposed to be an essential step in the catalytic deglycosylation mechanism,[85] while N7 protonation of the purine has been proposed to play a role in the catalytic deglycosylation facilitated by *Crithidia fasciculata* IU-NH.[86-87] In the current study, the effect of acid catalysis is initially considered by investigating full protonation at each possible nucleobase acceptor site (Figure 4.1) for the model with unactivated water as the nucleophile (Figure 4.2b). Protonation rather than hydrogen bonding is considered since protonation will likely result in the greatest reductions in the

hydrolysis barriers. Indeed, when $HCOO^{-}\bullet\bullet\bullet H_2O$ is the nucleophile, protonation of the uracil DNA nucleoside[37] results in at least a 20 kJ mol^{-1} reduction in the deglycosylation barrier calculated when a highly polar acid (HF) hydrogen bonds with U.[35]

Structures for the deglycosylation pathways corresponding to the protonation site that leads to the lowest barrier for each RNA nucleoside are shown in Figure 4.6, and will be discussed as representative examples. Specifically, these pathways include those involving protonation at N7 of A or G, and O2 of C or U. Structural parameters for all reaction pathways can be found in Tables C.6–C.9 (see Appendix C). We note that the orientation of the added proton can vary depending on the heteroatoms neighbouring the nucleobase acceptor site, with the orientation defined in brackets. For example, the proton at O2 of C can be directed either toward N1 (denoted as C–O2H(N1)) or N3 (denoted as C–O2H(N3)). The reactant structures for acid catalysis (Figure 4.6) resemble those for the (uncatalyzed) H_2O-mediated pathways (Figure 4.5), with H_2O positioned for S_N2 attack at C1′. The nucleophilic distance is ~3.200–3.400 Å, which is shorter than observed for the uncatalyzed pathways (~3.400–3.500 Å). As with the uncatalyzed pathways, the nucleophile donates a hydrogen bond to the 2′–hydroxyl group, which donates a hydrogen bond to the bridging water molecule, and these contacts are maintained throughout the reaction pathways. In the TS, the glycosidic bond distance falls between 2.500–2.600 Å, and the nucleophilic distance between ~2.200–2.500 Å, depending on the identity of the nucleobase. The hydrogen bonds between the water nucleophile and 2′–OH, as well as the bridging water molecule and 2′–OH, tighten by up to 0.100 Å in the TS compared to the RC, which helps stabilize the positive-charge developing on the sugar (Figure 4.6 and Tables C.6–C.9 in Appendix C). Similarly, the reaction angle increases from ~125° in the RC to ~150° in the TS. The PC contain fully formed bonds between the water nucleophile and ribose, while the nucleobase has departed, falling 3.100–3.500 Å from C1′ as discussed for the uncatalyzed pathways. The O_{wat}–$H_{wat}\bullet\bullet\bullet$O2′ hydrogen-bond distance decreases by ~0.200 Å from the TS to the PC, indicating partial proton

transfer to the 2′–hydroxyl group. The hydrogen bond between 2′–OH and the bridging water shortens to ~1.650 Å in the PC, which stabilizes the cationic charge forming on ribose as observed for the H_2O-mediated pathways.

The barriers for all of the acid-catalyzed pathways range from 89.7 kJ mol^{-1} to 162.4 kJ mol^{-1} (Table 4.2) and acid catalysis results in a 23.5–82.5 kJ mol^{-1} reduction in the reaction barrier over the corresponding uncatalyzed reactions. U–O2H(N3) has the smallest barrier (89.7 kJ mol^{-1}), which is consistent with the considerably longer nucleophilic distance in the U–O2H(N3) TS (2.538 Å) compared to the other TS (~2.2–2.3 Å; Figure 4.6). Indeed, protonation of O2 results in the largest barrier reduction (~80 kJ mol^{-1}) for both pyrimidines (Δ_{acid}, Table 4.2). Although the orientation of the proton on O2 greatly affects the barrier for U (i.e., a ~35 kJ mol^{-1} dependence on the proton direction), this dependence does not occur for C. The protonation sites that result in the smallest barrier reduction for the pyrimidines is O4 for U (~40 kJ mol^{-1} reduction, Δ_{acid}, Table 4.2), and the exocyclic amino group for C (~50 kJ mol^{-1} reduction, Δ_{acid}, Table 4.2). For the purines, N7 protonation results in the largest barrier reduction (~70 kJ mol^{-1}), while protonation of the exocyclic amino group results in the smallest change in barrier (31.8–33.3 kJ mol^{-1}). When N3 of A or G is protonated, the barrier is reduced by ~40 kJ mol^{-1}. The relative reaction energies of the acid-catalyzed pathways are similar to the uncatalyzed pathways, with the reaction energies being ~80–190 kJ mol^{-1} and falling only ~2–20 kJ mol^{-1} below the barriers. This indicates that protonation reduces the deglycosylation barriers and reaction energies to the same extent. However, acid catalysis does not provide the same stabilization to the PC as base catalysis, which leads to reaction energies of ~50 to 80 kJ mol^{-1} for $HCOO^-\bullet\bullet\bullet H_2O$ or ~–40 to –60 kJ mol^{-1} for OH^-. Therefore, a negatively charged species (e.g., OH^- or $HCOO^-$) is necessary to activate the water nucleophile and neutralize the positive charge forming in the TS, and thereby help stabilize the PC.

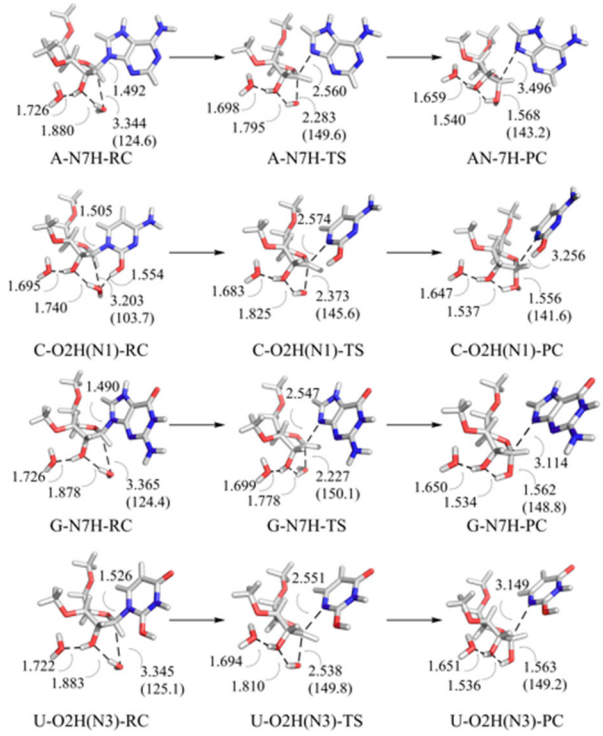


Figure 4.6. Selected B3LYP/6-31G(d) bond lengths (Å) and angles (degrees in parentheses) in reactant complexes (RC), transition state (TS), and product complexes (PC) for the lowest barrier pathways of the acid-catalyzed deglycosylation of adenine (A–N7H), cytosine (C–O2H(N1)), guanine (G–N7H), and uracil (U–O2H(N3)) containing RNA nucleosides facilitated by the unactivated H_2O nucleophile.

Our calculated barrier heights for the most stable acid-catalyzed pathways for A and G hydrolysis are in near quantitative agreement with experimental data.[70] Specifically, the experimentally-determined acid-catalyzed RNA hydrolysis barriers for A and G are 111.7 kJ mol^{-1} and 111.3 kJ mol^{-1}, respectively, which closely match our corresponding calculated barriers of 119.6 kJ and 126.4 kJ mol^{-1}, respectively. Our slightly higher barriers could likely be improved by modeling the reaction in explicit solvent, since the water nucleophile can only transfer a proton to 2′–OH in our model, while transfer to the bulk solvent would likely be more beneficial. Additionally, explicit solvation by even a few water molecules could stabilize the charge separated species and thereby further reduce the barrier, with up to four discrete water molecules having been previously shown to lower the barrier for the deglycosylation of the natural DNA

nucleosides.[31-32] More importantly, the nucleobase protonation sites that result in the greatest barrier reductions match sites that many enzymes are proposed to target. For example, π–π stacking between aromatic amino acids and the nucleobase has been proposed to facilitate solvent-mediated protonation of N7 of the purines by IAG-NH. This proposal correlates with our results that N7 protonation leads to the greatest barrier reduction for deglycosylation of RNA purines. Furthermore, protonation of O2 of C or U has been proposed as part of the activation mechanisms of CU-NH[26] and RihC,[85] which correlates with our findings that O2 is the most favourable activation site for RNA pyrimidines. Thus, our models correctly predict the effects of acid catalysis reported in the experimental literature, and provide valuable insights into nucleobase protonation sites that have the greatest catalytic impact, which rationalizes why enzymes tend to target specific nucleobase acceptor sites.

Combined Effects of Acid and Base Catalysis. In general, the acid-catalyzed pathways afford greater barrier reductions than the base-catalyzed pathways (i.e., $\Delta_{acid} > \Delta_{base}$, Table 4.2). Specifically, acid catalysis results in an up to 82.5 kJ mol^{-1} reduction over the uncatalyzed pathway, while base catalysis results in an up to 62.5 kJ mol^{-1} (OH^-) or 18.3 kJ mol^{-1} ($HCOO^-\bullet\bullet\bullet H_2O$) reduction. Nevertheless, the differences between the acid and base-catalyzed pathways can be minimal, especially for strong base catalysis (i.e., the OH^- nucleophile). Additionally, the reaction energies for the acid-catalyzed pathways are near equivalent to the barriers, suggesting the presence of a negatively charged species (e.g., $HCOO^-$ or OH^-) is essential for favorable reaction thermodynamics. Therefore, both acid and base catalysis may significantly impact the deglycosylation reaction. Indeed, the active sites of enzymes that catalyze nucleic acid deglycosylation typically contain both acidic and basic residues that can facilitate the reaction.[14-18,22-25,45-59] To examine the simultaneous impact of both acid and base catalysis, we modelled the departure of protonated nucleobases assisted by either OH^- or $HCOO^-\bullet\bullet\bullet H_2O$.

The most significant barrier reductions relative to uncatalyzed RNA hydrolysis result when both acid and base catalysis are employed, which leads to an up to ~120 kJ mol^{-1} decrease in the barrier when OH^- is the nucleophile or ~100 kJ mol^{-1} when $HCOO^-\bullet\bullet\bullet H_2O$ is the nucleophile ($\Delta_{acid-base}$, Table 4.2). Regardless, a comparison of the acid-catalyzed pathways for the OH^- or $HCOO^-\bullet\bullet\bullet H_2O$ nucleophile to the corresponding acid-catalyzed pathways for the unactivated water nucleophile reveals that the presence of the base does not change the protonation site that results in the greatest barrier reduction. Specifically, acid catalysis directed at N7 of the purines or O2 of the pyrimidines results in the lowest excision barriers regardless of the nucleophile. Furthermore, regardless of the presence of acid catalysis, the barriers decrease and the reactions become increasingly exothermic with stronger nucleophile activation (i.e., the barriers and reaction energies decrease according to the nucleophile as $H_2O > HCOO^-\bullet\bullet\bullet H_2O > OH^-$, Table C.14 in Appendix C).

While the combined efforts of acid and base catalysis yields the lowest barriers, it is of interest to determine whether the effects of acid and base catalysis on the deglycosylation barriers are additive, less than additive, or synergistic. Therefore, Table 4.2 ($\Delta_{sum} - \Delta_{acid-base}$) compares the sum of the individual acid and base contributions ($\Delta_{sum} = \Delta_{acid} + \Delta_{base}$) to the calculated effects of the simultaneous presence of acid and base catalysis ($\Delta_{acid-base}$, Table 4.2). Whether the combined effects of acid and base catalysis are additive, less than additive, or synergistic depends on the level of nucleophile activation through base catalysis, the nucleobase, and the nucleobase protonation site. For example, when OH^- is the nucleophile, $\Delta_{acid-base} \approx \Delta_{sum}$ (less than ~10 kJ mol^{-1} difference, Table 4.2) for each purine deglycosylation pathway, indicating that the effects of acid and base catalysis are additive. Similarly, for the $HCOO^-\bullet\bullet\bullet H_2O$ deglycosylation pathways, $\Delta_{acid-base} \approx \Delta_{sum}$ in most pathways regardless of the identity of the nucleobase (Table 4.2). Conversely, the effects of acid and base catalysis are significantly diminished when simultaneously present for the C or U hydrolysis pathways facilitated by OH^-, with $\Delta_{acid-base}$ being up to ~30

kJ mol^{-1} less than Δ_{sum}. Finally, the G–N3H and U–O2H(N1) pathways exhibit significant acid–base synergy when $HCOO^{-}\bullet\bullet\bullet H_2O$ is the nucleophile, with $\Delta_{acid-base}$ being 20 kJ mol^{-1} larger than Δ_{sum}. Similar to the barriers, the effects of combined acid–base catalysis on the reaction energies are varied (Table C.14 in Appendix C), with the effects generally being additive ($\Delta_{sum} - \Delta_{acid-base} < 10$ kJ mol^{-1}), but being synergistic (by up to ~35 kJ mol^{-1}) in some instances. Regardless of the additive nature of acid–base catalysis, our data emphasizes that both catalytic approaches are crucial for reducing otherwise prohibitive RNA nucleoside hydrolysis barriers, and both strategies likely play an important role in enzymatic systems.

4.4 Conclusions

The current study used several computational models to examine RNA nucleoside hydrolytic deglycosylation. First, a comparison of RNA and DNA hydrolysis facilitated by the $HCOO^{-}\bullet\bullet\bullet H_2O$ nucleophile revealed the 2′–hydroxy group in RNA nucleosides imparts stability to the glycosidic bond, which agrees with experimental results for purine deglycosylation in acidic conditions,[70] and extends the conclusion to the pyrimidines and basic environments. Second, the effects of base catalysis were evaluated by considering hydrolysis mediated by an OH^{-} nucleophile, which lead to significantly lower barriers than unactivated or $HCOO^{-}$ activated hydrolysis, as well as exothermic reactions. Third, the effect of acid catalysis was considered by protonation of nucleobase hydrogen-bond acceptor sites. Interestingly, the nucleobase protonation sites that result in the greatest barrier reductions are the same sites typically targeted by enzymes that catalyze RNA deglycosylation reactions. However, although acid reduces the hydrolysis barriers, the corresponding product complexes are only slightly more stable than the TS, implying that both acid and base catalysis is important. Indeed, simultaneously invoking both acid and base catalytic approaches is the most effective strategy for reducing RNA deglycosylation barriers. Whether the effect of concomitant acid and base catalysis is additive, less than additive, or synergistic compared to the sum of the individual effects is dependent on the nucleobase, protonation site, and nucleophile. Overall,

this work establishes a fundamental understanding of the intrinsic stability of the RNA glycosidic bond, and the effects of acid and base catalysis on RNA deglycosylation. Our findings provide important mechanistic details for scientists studying enzyme-mediated RNA deglycosylation reactions that can be used as a basis for future large-scale modeling of enzymatic reactions and to design TS analogues that inhibit RNA-targeting enzymes.

4.5 References

[1] Voet, D.; Voet, J. G., *Biochemistry*. 3rd ed.; John Wiley & Sons: 2004.

[2] Wallace, S. S.; Murphy, D. L.; Sweasy, J. B., Base Excision Repair and Cancer. *Cancer Lett.* **2012,** *327* (1–2), 73–89.

[3] Svilar, D.; Goellner, E. M.; Almeida, K. H.; Sobol, R. W., Base Excision Repair and Lesion-Dependent Subpathways for Repair of Oxidative DNA Damage. *Antioxid. Redox Signal.* **2011,** *14* (12), 2491–2507.

[4] Dizdaroglu, M., Oxidatively Induced DNA Damage: Mechanisms, Repair and Disease. *Cancer Lett.* **2012,** *327* (1–2), 26–47.

[5] Das, M. K.; Sharma, R. S.; Mishra, V., Induction of Apoptosis by Ribosome Inactivating Proteins Importance of N-Glycosidase Activity. *Appl. Biochem. Biotechnol.* **2012,** *166* (6), 1552–1561.

[6] Domashevskiy, A. V.; Goss, D. J., Pokeweed Antiviral Protein, a Ribosome Inactivating Protein: Activity, Inhibition and Prospects. *Toxins* **2015,** *7* (2), 274–298.

[7] el Kouni, M. H., Potential Chemotherapeutic Targets in the Purine Metabolism of Parasites. *Pharmacol. Ther.* **2003,** *99* (3), 283–309.

[8] Versees, W.; Steyaert, J., Catalysis by Nucleoside Hydrolases. *Curr. Opin. Struct. Biol.* **2003,** *13* (6), 731–738.

[9] Zhu, Y.; Dai, J.; Zhang, T.; Li, X.; Fang, P.; Wang, H.; Jiang, Y.; Yu, X.; Xia, T.; Niu, L.; et al., Structural Insights into the Neutralization Mechanism of Monoclonal Antibody 6C2 against Ricin. *J. Biol. Chem.* **2013,** *288* (35), 25165–25172.

[10] Lacerda, A. F.; Pelegrini, P. B.; de Oliveira, D. M.; Vasconcelos, E. A. R.; Grossi-de-Sa, M. F., Anti-Parasitic Peptides from Arthropods and their Application in Drug Therapy. *Frontiers in Microbiology* **2016,** *7*, 91.

[11] Murkin, A. S.; Moynihan, M. M., Transition-State-Guided Drug Design for Treatment of Parasitic Neglected Tropical Diseases. *Curr. Med. Chem.* **2014,** *21* (15), 1781–1793.

[12] Barrett, M. P.; Croft, S. L., Management of Trypanosomiasis and Leishmaniasis. *Br. Med. Bull.* **2012,** *104* (1), 175–196.

[13] Boitz, J. M.; Ullman, B.; Jardim, A.; Carter, N. S., Purine Salvage in Leishmania: Complex or Simple by Design? *Trends Parasitol.* **2012,** *28* (8), 345–352.

[14] McCann, J. A. B.; Berti, P. J., Transition-State Analysis of the DNA Repair Enzyme MutY. *J. Am. Chem. Soc.* **2008,** *130* (17), 5789–5797.

[15] Zharkov, D. O.; Mechetin, G. V.; Nevinsky, G. A., Uracil-DNA Glycosylase: Structural, Thermodynamic and Kinetic Aspects of Lesion Search and Recognition. *Mutat. Res.-Fundam. Mol. Mech. Mutag.* **2010,** *685* (1–2), 11–20.

[16] Parikh, S. S.; Walcher, G.; Jones, G. D.; Slupphaug, G.; Krokan, H. E.; Blackburn, G. M.; Tainer, J. A., Uracil-DNA Glycosylase-DNA Substrate and Product Structures: Conformational Strain Promotes Catalytic Efficiency by Coupled Stereoelectronic Effects. *Proc. Natl. Acad. Sci. USA.* **2000,** *97* (10), 5083–5088.

[17] O'Brien, P. J.; Ellenberger, T., Dissecting the Broad Substrate Specificity of Human 3-Methyladenine-DNA Glycosylase. *J. Biol. Chem.* **2004,** *279* (11), 9750–9757.

[18] Koval, V. V.; Knorre, D. G.; Fedorova, O. S., Structural Features of the Interaction between Human 8-Oxoguanine DNA Glycosylase hOGG1 and DNA. *Acta Naturae* **2014,** *6* (3), 52–65.

[19] Chu, A. M.; Fettinger, J. C.; David, S. S., Profiling Base Excision Repair Glycosylases with Synthesized Transition State Analogs. *Bioorg. Med. Chem. Lett.* **2011,** *21* (17), 4969–4972.

[20] Kushwaha, G. S.; Pandey, N.; Sinha, M.; Singh, S. B.; Kaur, P.; Sharma, S.; Singh, T. P., Crystal Structures of a Type-1 Ribosome Inactivating Protein from Momordica Balsamina in the Bound and Unbound States. *Biochim. Biophys. Acta* **2012,** *1824* (4), 679–691.

[21] Gu, Y. J.; Xia, Z. X., Crystal Structures of the Complexes of Trichosanthin with Four Substrate Analogs and Catalytic Mechanism of RNA N-Glycosidase. *Proteins* **2000,** *39* (1), 37–46.

[22] Versees, W.; Loverix, S.; Vandemeulebroucke, A.; Geerlings, P.; Steyaert, J., Leaving Group Activation by Aromatic Stacking: An Alternative to General Acid Catalysis. *J. Mol. Biol.* **2004,** *338* (1), 1–6.

[23] Arivett, B.; Farone, M.; Masiragani, R.; Burden, A.; Judge, S.; Osinloye, A.; Minici, C.; Degano, M.; Robinson, M.; Kline, P., Characterization of Inosine-Uridine Nucleoside Hydrolase (RihC) from Escherichia Coli. *Biochim. Biophys. Acta.* **2014,** *1844* (3), 656–662.

[24] Lovane, E.; Giabba, B.; Muzzolini, L.; Matafora, V.; Fornili, A.; Minici, C.; Giannese, F.; Degano, M., Structural Basis for Substrate Specificity in Group I Nucleoside Hydrolases. *Biochemistry* **2008,** *47* (15), 4418–4426.

[25] Imamura, K.; Averill, A.; Wallace, S. S.; Doublie, S., Structural Characterization of Viral Ortholog of Human DNA Glycosylase NEIL1 Bound to Thymine Glycol or 5-Hydroxyuracil-Containing DNA. *J. Biol. Chem.* **2012,** *287* (6), 4288–4298.

[26] Giabbai, B.; Degano, M., Crystal Structure to 1.7 Å of the *Escherichia Coli* Pyrimidine Nucleoside Hydrolase YeiK, a Novel Candidate for Cancer Gene Therapy. *Structure* **2004,** *12* (5), 739–749.

[27] Iovane, E.; Giabba, B.; Muzzolini, L.; Matafora, V.; Fornili, A.; Minici, C.; Giannese, F.; Degano, M., Structural Basis for Substrate Specificity in Group I Nucleoside Hydrolases. *Biochemistry* **2008,** *47* (15), 4418–4426.

[28] Loverix, S.; Geerlings, P.; McNaughton, M.; Augustyns, K.; Vandemeulebroucke, A.; Steyaert, J.; Versees, W., Substrate-Assisted Leaving Group Activation in Enzyme-Catalyzed N-Glycosidic Bond Cleavage. *J. Biol. Chem.* **2005,** *280* (15), 14799–14802.

[29] Millen, A. L.; Archibald, L. A. B.; Hunter, K. C.; Wetmore, S. D., A Kinetic and Thermodynamic Study of the Glycosidic Bond Cleavage in Deoxyuridine. *J. Phys. Chem. B* **2007,** *111* (14), 3800–3812.

[30] Millen, A. L.; Wetmore, S. D., Glycosidic Bond Cleavage in Deoxynucleotides – a Density Functional Study. *Can. J. Chem.* **2009,** *87* (7), 850–863.

[31] Przybylski, J. L.; Wetmore, S. D., Designing an Appropriate Computational Model for DNA Nucleoside Hydrolysis: A Case Study of 2′-Deoxyuridine. *J. Phys. Chem. B* **2009,** *113* (18), 6533–6542.

[32] Przybylski, J. L.; Wetmore, S. D., Modeling the Dissociative Hydrolysis of the Natural DNA Nucleosides. *J. Phys. Chem. B* **2010,** *114* (2), 1104–1113.

[33] Shim, E. J.; Przybylski, J. L.; Wetmore, S. D., Effects of Nucleophile, Oxidative Damage, and Nucleobase Orientation on the Glycosidic Bond Cleavage in Deoxyguanosine. *J. Phys. Chem. B* **2010,** *114* (6), 2319–2326.

[34] Kellie, J. L.; Navarro-Whyte, L.; Carvey, M. T.; Wetmore, S. D., Combined Effects of π–π Stacking and Hydrogen Bonding on the (N1) Acidity of Uracil and Hydrolysis of 2′-Deoxyuridine. *J. Phys. Chem. B* **2012,** *116* (8), 2622–2632.

[35] Lenz, S. A. P.; Kellie, J. L.; Wetmore, S. D., Glycosidic Bond Cleavage in Deoxynucleotides: Effects of Solvent and the DNA Phosphate Backbone in the Computational Model. *J. Phys. Chem. B* **2012,** *116* (49), 14275–14284.

[36] Navarro-Whyte, L.; Kellie, J. L.; Lenz, S. A. P.; Wetmore, S. D., Hydrolysis of the Damaged Deoxythymidine Glycol Nucleoside and Comparison to Canonical DNA. *PCCP* **2013,** *15* (44), 19343–19352.

[37] Lenz, S. A. P.; Kellie, J. L.; Wetmore, S. D., Glycosidic Bond Cleavage in DNA Nucleosides: Effect of Nucleobase Damage and Activation on the Mechanism and Barrier. *J. Phys. Chem. B* **2015,** *119* (51), 15601–15612.

[38] Sowlati-Hashjin, S.; Wetmore, S. D., Computational Investigation of Glycosylase and β-Lyase Activity Facilitated by Proline: Applications to Fpg and Comparisons to hOGG1. *J. Phys. Chem. B* **2014,** *118* (50), 14566–14577.

[39] Kellie, J. L.; Wetmore, S. D., Mechanistic and Conformational Flexibility of the Covalent Linkage Formed During β-Lyase Activity on an AP-Site: Application to hOGG1. *J. Phys. Chem. B* **2012,** *116* (35), 10786–10797.

[40] Zheng, Y.; Xue, Y.; Yan, G. S., The Influences of Oxidation and Cationization on the N-Glycosidic Bond Stability of 8-Oxo-2′-Deoxyadenosine – A Theoretical Study. *J. Chem. Theory Comput.* **2009,** *8* (6), 1253–1264.

[41] Zheng, Y.; Xue, Y.; Yan, S. G., The Effects of Oxidation and Protonation on the N-Glycosidic Bond Stability of 8-Oxo-2′-Deoxyguanosine: DFT Study. *J. Mol. Struc.-Theochem* **2008,** *860* (1–3), 52–57.

[42] Ebrahimi, A.; Habibi-Khorassani, M.; Bazzi, S., The Impact of Protonation and Deprotonation of 3-Methyl-2′-Deoxyadenosine on N-Glycosidic Bond Cleavage. *PCCP* **2011,** *13* (8), 3334–3343.

[43] Rios-Font, R.; Rodriguez-Santiago, L.; Bertran, J.; Sodupe, M., Influence of N7 Protonation on the Mechanism of the N-Glycosidic Bond Hydrolysis in 2′-Deoxyguanosine. A Theoretical Study. *J. Phys. Chem. B* **2007,** *111* (21), 6071–6077.

[44] Rios-Font, R.; Bertran, J.; Rodriguez-Santiago, L.; Sodupe, M., Effects of Ionization, Metal Cationization and Protonation on 2 '-Deoxyguanosine: Changes on Sugar Puckering and Stability of the N-Glycosidic Bond. *J. Phys. Chem. B* **2006,** *110* (11), 5767–5772.

[45] Osakabe, T.; Fujii, Y.; Hata, M.; Tsuda, M.; Neya, S.; Hoshino, T., Quantum Chemical Study on Base Excision Mechanism of 8-Oxoguanine DNA Glycosylase: Substrate-Assisted Catalysis of the N-Glycosidic Linkage Cleavage Reaction. *Chem-Bio. Infomatics J.* **2004,** *4* (3), 73–92.

[46] Schyman, P.; Danielsson, J.; Pinak, M.; Laaksonen, A., Theoretical Study of the Human DNA Repair Protein hOGG1 Activity. *J. Phys. Chem. A* **2005,** *109* (8), 1713–1719.

[47] Chen, Z. Q.; Zhang, C. H.; Xue, Y., Theoretical Studies on the Thermodynamics and Kinetics of the N-Glycosidic Bond Cleavage in Deoxythymidine Glycol. *J. Phys. Chem. B* **2009,** *113* (30), 10409–10420.

[48] Chen, Z.-q.; Xue, Y., Theoretical Investigations on the Thermal Decomposition Mechanism of 5-Hydroxy-6-Hydroperoxy-5,6-Dihydrothymidine in Water. *J. Phys. Chem. B* **2010,** *114* (39), 12641–12654.

[49] Chen, Z.-Q.; He, Y.-Q.; Guo, L.-F.; Xue, Y., Effects of Explicit and Implicit Solvent Models on the Hydrolysis Cleavage of N-Glycosidic Bond in 8-Oxo-7,8-Dihydro-2′-Deoxyguanosine. *Chinese J. Struct. Chem.* **2014,** *33* (4), 505–512.

[50] Williams, R. T.; Wang, Y., A Density Functional Theory Study on the Kinetics and Thermodynamics of N-Glycosidic Bond Cleavage in 5-Substituted 2′-Deoxycytidines. *Biochemistry* **2012,** *51* (32), 6458–6462.

[51] Delarami, H. S.; Ebrahimi, A.; Bazzi, S.; Khorassani, S. M. H., The Effect of Intramolecular Hydrogen Bond on the N-Glycosidic Bond Strength in 3-Methyl-2′-Deoxyadenosine: A Quantum Chemical Study. *Struct. Chem.* **2015,** *26* (2), 411–419.

[52] 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.

[53] Kellie, J. L.; Wilson, K. A.; Wetmore, S. D., An ONIOM and MD Investigation of Possible Monofunctional Activity of Human 8-Oxoguanine-DNA Glycosylase (hOGG1). *J. Phys. Chem. B* **2015,** *119* (25), 8013–8023.

[54] Brunk, E.; Arey, J. S.; Rothlisberger, U., Role of Environment for Catalysis of the DNA Repair Enzyme MutY. *J. Am. Chem. Soc.* **2012,** *134* (20), 8608–8616.

[55] Alexandrova, A. N., Promiscuous DNA Alkyladenine Glycosylase Dramatically Favors a Bound Lesion over Undamaged Adenine. *Biophys. Chem.* **2010,** *152* (1–3), 118–127.

[56] Sadeghian, K.; Ochsenfeld, C., Unraveling the Base Excision Repair Mechanism of Human DNA Glycosylase. *J. Am. Chem. Soc.* **2015,** *137* (31), 9824–9831.

[57] Blank, I. D.; Sadeghian, K.; Ochsenfeld, C., A Base-Independent Repair Mechanism for DNA Glycosylase—No Discrimination within the Active Site. *Sci. Rep.* **2015,** *5*, 10369.

[58] Kanaan, N.; Crehuet, R.; Imhof, P., Mechanism of the Glycosidic Bond Cleavage of Mismatched Thymine in Human Thymine DNA Glycosylase Revealed by Classical Molecular Dynamics and Quantum Mechanical/Molecular Mechanical Calculations. *J. Phys. Chem. B* **2015,** *119* (38), 12365–12380.

[59] Tehrani, Z. A.; Fattahi, A.; Pourjavadi, A., Interaction of Mg^{2+}, Ca^{2+}, Zn^{2+} and Cu^{+} with Cytosine Nucleosides: Influence of Metal on Sugar Puckering and Stability of N-Glycosidic Bond, a DFT Study. *J. Mol. Struc.-Theochem.* **2009,** *913* (1–3), 117–125.

[60] Chen, N.; Ge, H.; Xu, J.; Cao, Z.; Wu, R., Loop Motion and Base Release in Purine-Specific Nucleoside Hydrolase: A Molecular Dynamics Study. *Biochim. Biophys. Acta.* **2013,** *1834* (6), 1117–1124.

[61] Chen, N.; Zhao, Y.; Lu, J.; Wu, R.; Cao, Z., Mechanistic Insights into the Rate-Limiting Step in Purine-Specific Nucleoside Hydrolase. *J. Chem. Theory. Comput.* **2015,** *11* (7), 3180–3188.

[62] Fornili, A.; Giabbai, B.; Garau, G.; Degano, M., Energy Landscapes Associated with Macromolecular Conformational Changes from Endpoint Structures. *J. Am. Chem. Soc.* **2010,** *132* (49), 17570–17577.

[63] Franca, T. C. C.; Rocha, M. do R. M.; Reboredo, B. M.; Renno, M. N.; Tinoco, L. W.; Figueroa-Villar, J. D., Design of Inhibitors for Nucleoside Hydrolase from Leishmania Donovani Using Molecular Dynamics Studies. *J. Braz. Chem. Soc.* **2008,** *19* (1), 64–73.

[64] Guimaraes, A. P.; Oliveira, A. A.; da Cunha, E. F. F.; Ramalho, T. C.; Franco, T. C. C., Analysis of Bacillus Anthracis Nucleoside Hydrolase via *in Silico* Docking with Inhibitors and Molecular Dynamics Simulation. *J. Mol. Model.* **2011,** *17* (11), 2939–2951.

[65] Mancini, D. T.; Matos, K. S.; da Cunha, E. F. F.; Assis, T. M.; Guimaraes, A. P.; Franca, T. C. C.; Ramalho, T. C., Molecular Modeling Studies on Nucleoside Hydrolase from the Biological Warfare Agent Brucella Suis. *J. Biomol. Struct. Dyn.* **2012,** *30* (1), 125–136.

[66] Ogungbe, I. V.; Erwin, W. R.; Setzer, W. N., Antileishmanial Phytochemical Phenolics: Molecular Docking to Potential Protein Targets. *J. Mol. Graph. Model.* **2014,** *48* (2014), 105–117.

[67] Ogungbe, I. V.; Ng, J. D.; Setzer, W. N., Interactions of Antiparasitic Alkaloids with Leishmania Protein Targets: A Molecular Docking Analysis. *Future Med. Chem.* **2013,** *5* (15), 1777–1799.

[68] Ogungbe, I. V.; Setzer, W. N., In-Silico Leishmania Target Selectivity of Antiparasitic Terpenoids. *Molecules* **2013,** *18* (7), 7761–7847.

[69] Wu, R.; Gong, W.; Ting, L.; Zhang, Y.; Cao, Z., QM/MM Molecular Dynamics Study of Purine-Specific Nucleoside Hydrolase. *J. Phys. Chem. B* **2012,** *116* (6), 1984–1991.

[70] Rios, A. C.; Yu, H. T.; Tor, Y., Hydrolytic Fitness of N-Glycosyl Bonds: Comparing the Deglycosylation Kinetics of Modified, Alternative, and Native Nucleosides. *J. Phys. Org. Chem.* **2015,** *28* (3), 173–180.

[71] Lau, A. Y.; Scharer, O. D.; Samson, L.; Verdine, G. L.; Ellenberger, E., Crystal Structure of a Human Alkylbase-DNA Repair Enzyme Complexed to DNA: Mechanisms for Nucleotide Flipping and Base Excision. *Cell* **1998,** *95* (2), 249–258.

[72] Versees, W.; Decanniere, K.; Pelle, R.; Depoorter, J.; Brosens, E.; Parkin, D. W.; Steyaert, J., Structure and Function of a Novel Purine Specific Nucleoside Hydrolase from Trypanosoma Vivax. *J. Mol. Biol.* **2001,** *307* (5), 1363–1379.

[73] Versees, W.; Goeminne, A.; Berg, M.; Vandemeulebroucke, A.; Haemers, A.; Augustyns, K.; Steyaert, J., Crystal Structures of T. Vivax Nucleoside Hydrolase in Complex with New Potent and Specific Inhibitors. *Biochim. Biophys. Acta.* **2009,** *1794* (6), 953–960.

[74] Monzingo, A. F.; Robertus, J. D., X-Ray Analysis of Substrate Analogs in the Ricin A-Chain Active Site. *J. Mol. Biol.* **1992,** *227* (4), 1136–1145.

[75] Day, P. J.; Ernst, S. R.; Frankel, A. E.; Monzingo, A. F.; Pascal, J. M.; Molina-Svinth, M. C.; Robertus, J. D., Structure and Activity of an Active Site Substitution of Ricin A–Chain. *Biochemistry* **1996,** *35* (34), 11098–11103.

[76] Scott, A. P.; Radom, L., Harmonic Vibrational Frequencies: An Evaluation of Hartree-Fock, Moller-Plesset, Quadratic Configuration Interaction, Density Functional Theory, and Semiempirical Scale Factors. *J. Phys. Chem.* **1996,** *100* (41), 16502–16513.

[77] Marenich, A. V.; Cramer, C. J.; Truhlar, D. G., Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009,** *113* (18), 6378–6396.

[78] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. *Gaussian 09*, Revisions A.02, C.01, and D.01; Gaussian, Inc.: Wallingford, CT, USA, **2009**.

[79] Lee, S.; Verdine, G. L., Atomic Substitution Reveals the Structural Basis for Substrate Adenine Recognition and Removal by Adenine DNA Glycosylase. *Proc. Natl. Acad. Sci. USA.* **2009,** *106* (44), 18497–18502.

[80] Parikh, S. S.; Mol, C. D.; Slupphaug, G.; Bharati, S.; Krokan, H. E.; Tainer, J. A., Base Excision Repair Initiation Revealed by Crystal Structures and Binding Kinetics of Human Uracil-DNA Glycosylase with DNA. *EMBO J.* **1998,** *17* (17), 5214–5226.

[81] Cui, L. W.; Rajasekariah, G. R.; Martin, S. K., A Nonspecific Nucleoside Hydrolase from Leishmania Donovani: Implications for Purine Salvage by the Parasite. *Gene* **2001,** *280* (1–2), 153–162.

[82] Wink, P. L.; Sanchez Quitian, Z. A.; Rosado, L. A.; Rodrigues Junior, V. d. S.; Petersen, G. O.; Lorenzini, D. M.; Lipinski-Paes, T.; Saraiva Macedo Timmers, L. F.; de Souza, O. N.; Basso, L. A.; et al., Biochemical Characterization of Recombinant Nucleoside Hydrolase from Mycobacterium Tuberculosis H37rv. *Arch. Biochem. Biophys.* **2013,** *538* (2), 80–94.

[83] Porcelli, M.; Peiluso, I.; Marabotti, A.; Facchiano, A.; Cacciapuoti, G., Biochemical Characterization and Homology Modeling of a Purine-Specific Ribonucleoside Hydrolase from the Archaeon Sulfolobus Solfataricus: Insights into Mechanisms of Protein Stabilization. *Arch. Biochem. Biophys.* **2009,** *483* (1), 55–65.

[84] Porcelli, M.; De Leo, E.; Marabotti, A.; Cacciapuoti, G., Site-Directed Mutagenesis Gives Insights into Substrate Specificity of Sulfolobus Solfataricus Purine-Specific Nucleoside Hydrolase. *Ann. Microbiol.* **2012,** *62* (2), 881–887.

[85] Hunt, C.; Gillani, N.; Farone, A.; Rezaei, M.; Kline, P. C., Kinetic Isotope Effects of Nucleoside Hydrolase from Escherichia Coli. *Biochim. Biophys. Acta.* **2005,** *1751* (2), 140–149.

[86] Gopaul, D. N.; Meyer, S. L.; Degano, M.; Sacchettini, J. C.; Schramm, V. L., Inosine−Uridine Nucleoside Hydrolase from Crithidia Fasciculata. Genetic Characterization, Crystallization, and Identification of Histidine 241 as a Catalytic Site Residue. *Biochemistry* **1996,** *35* (19), 5963–5970.

[87] Horenstein, B. A.; Parkin, D. W.; Estupinan, B.; Schramm, V. L., Transition-State Analysis of Nucleoside Hydrolase from Crithidia Fasciculata. *Biochemistry* **1991,** *30* (44), 10788–10795.

Chapter 5: Structural Explanation for the Tunable Substrate Specificity of Nucleoside Hydrolases: Insights from Molecular Dynamics Simulations[a,b,c]

5.1 Introduction

Parasitic protozoa are responsible for a host of diseases that affect the developing world including African and American trypanosomiasis (commonly known as African sleeping sickness and Chagas disease, respectively), and leishmaniasis [1-5]. Symptoms of African trypanosomiasis include irregular sleep-wake patterns, psychiatric disorders, and depression [1, 4], while American trypanosomiasis causes inflammation, facial swelling, and organ failure [1]. In addition to inflammation and organ failure, patients with leishmaniasis exhibit skin lesions [5]. Collectively, these ailments comprise the kinetoplastid diseases and primarily affect developing or poverty-stricken countries. However, emigration of infected individuals from endemic countries has increased in the USA, Canada, and Europe [1, 6-8]. Unfortunately, treatments of these diseases are impeded by problems related to drug delivery, efficacy, and host toxicity [1].

Unlike humans, parasitic protozoa lack *de novo* pathways to synthesize purine nucleobases that are essential for nucleic acid synthesis and energy storage [9]. Therefore, parasitic protozoa rely on salvage pathways for their survival [2, 10-14]. Salvage is initiated by uptake of ribonucleosides into the cell via nucleoside transporters, hydrolytic deglycosylation catalyzed by nucleoside hydrolases (NH), and phosphoribosyltransferase-catalyzed ribophosphorylation of the nucleobases. Remarkably, nucleoside hydrolase analogues do not exist in mammals[15, 16] and therefore several anti-kinetoplastid compounds have been developed to target these enzymes. For example, immucillins [17-20], N-arylmethyl-substituted

[a] *Journal of Computer Aided Molecular Design* reference style used in this Chapter.
[b] Adapted from Lenz, S.A.P. and Wetmore, S.D. (2018) Structural Explanation for the Tunable Substrate Specificity of Nucleoside Hydrolases: Insights from Molecular Dynamics Simulations, *J. Comput. Aided Mol. Des. [submitted, JCAM-D-18-00239].*
[c] S.A.P.L. performed all calculations. S.A.P.L. and S.D.W. contributed to the design of calculations, interpretation of data, and writing the manuscript.

iminoribitol [20-23], and nucleoside analogs [24, 25] are anti-parasitic and among the strongest inhibitors of nucleoside hydrolases. However, the syntheses of these compounds are complex and lead to low yields [26]. Due to the adaptability of the salvage pathways, inhibition of one enzyme may not be effective for treating protozoan disease and a combination therapy may be required [19, 20]. Nevertheless, to assist with the rational development of new therapies, knowledge of the substrate binding and activity of nucleoside hydrolases is paramount.

Nucleoside hydrolases can be classified into subgroups based on their substrate specificity, including the non-specific inosine-uridine NH (IU-NH), pyrimidine-specific cytidine-uridine NH (CU-NH), and purine-specific inosine-adenosine-guanosine NH (IAG-NH), as well as 6-oxopurine-specific inosine-guanosine NH (IG-NH) whose substrates include inosine, guanosine, and xanthosine (Figure 5.1) [13, 20, 27]. Nevertheless, all nucleoside hydrolases share several common active site features (see, for example, Figure 5.2). Specifically, several Asp and Glu residues surround ribose, which have been proposed to stabilize the oxocarbenium ion formed in the transition state upon deglycosylation [27]. Furthermore, nucleoside hydrolase actives sites contain a Ca^{2+} ion coordinated to several Asp residues, a water nucleophile, and the 2′– and 3′–OH groups of the bound nucleoside substrate (Figure 5.2a and c). A conserved Ca^{2+}-coordinated Asp acts as a general base (for example, Asp11 in CU-NH and Asp10 in IU-NH; Figure 5.2a and c), which activates the water nucleophile that attacks C1′ and cleaves the *N*-glycosidic bond connecting the nucleobase to ribose [15, 21, 28-36]. The active site is specific toward ribonucleosides since the coordination of both ribose hydroxy groups to Ca^{2+} is crucial for substrate binding [2].

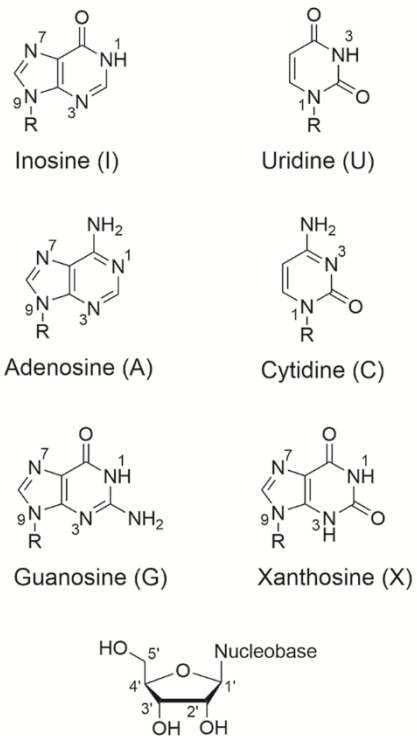


Figure 5.1. Structures and chemical numbering of common nucleoside hydrolase substrates, as well as the ribose sugar (R).

E. coli CU-NH and *C. fasciculata* IU-NH utilize similar active site residues to interact with the nucleobase and thereby facilitate substrate binding, as well as catalyze *N*-glycosidic bond hydrolysis. The similarities between the active sites of CU-NH and IU-NH suggest a conserved role for catalytic residues across both species. For example, in both CU-NH and IU-NH, two proposed catalytic His residues interact with the nucleobase of the bound substrate (Figure 5.2b and d) [15, 33, 35, 37]. Specifically, His82 interacts with the nucleobase through a T-shaped (edge-to-face) interaction in both CU-NH and IU-NH [33, 37]. In the case of IU-NH, the inactivity of a His241Ala mutant confirms a catalytic role for His241 [15, 35], which has been proposed to transfer a proton to N7 of inosine through Tyr229 [15, 33, 37]. Indeed, observed kinetic isotope effects indicate that N7 is protonated during IU-NH-catalyzed hydrolysis [38]. Similarly, for CU-NH, His239 has been proposed to transfer a proton to O2 of C or U [33, 34]. His239 or His241 must be cationic to function as the general acid in CU-NH- or IU-NH-mediated

hydrolysis. Nevertheless, a His239Ala CU-NH mutant has a larger K_M, but equivalent $k_{cat,}$ for U relative to the wild-type enzyme [33, 34], suggesting that while His239 is likely important for substrate binding, other residues may act as the general acid.

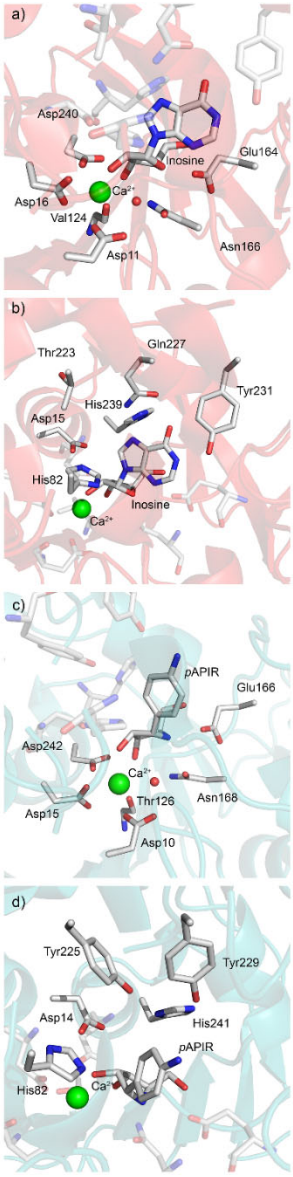


Figure 5.2. X-ray crystal structure of CU-NH bound to inosine (a, b; PDB ID: 3B9X) [33], and IU-NH bound to *p*-aminophenyliminoribitol (c,d; *p*APIR; PDB ID: 2MAS)[37] highlighting the amino acids interacting with the canonical or modified ribose (a and c) or nucleobase (b and d).

Interestingly, (wild-type) CU-NH is unable to process inosine, which has been proposed to arise since His239 is not aligned to transfer a proton to the nucleobase in the X-ray structure (Figure 5.2b). However, the substrate selectivity of CU-NH can be altered. Specifically, mutagenesis experiments revealed that a Thr223Tyr Gln227Tyr CU-NH double mutant has a nearly 50-fold greater k_{cat} for inosine relative to the wild-type enzyme, while the K_M is not significantly altered, which suggests that the mutations primarily affect the hydrolysis reaction [33]. In fact, the reaction rate for the CU-NH double mutant is comparable to that for inosine hydrolysis facilitated by IU-NH [17]. Interestingly, the Thr223Tyr single mutant lowers k_{cat} for inosine relative to the wild-type enzyme [33]. Although the Gln227Tyr mutant increases k_{cat} for inosine relative to the wild-type enzyme, the increase is not as significant as for the Thr223Tyr Gln227Tyr double mutant, which led to proposals that residues 223 and 227 function cooperatively [33]. Indeed, the double mutation introduces a Tyr-Tyr-His motif into the active site, which has been proposed to be characteristic of IU-NH substrate specificity (Figure 5.2d) [15, 37, 39], and was proposed to allow His239 to transfer a proton to inosine via Tyr227 [33]. Unfortunately, there are no X-ray structures of Thr223Tyr and/or Gln227Tyr CU-NH mutants, and therefore the exact impact of these active site alterations on discrete nucleoside–protein interactions that may dictate the substrate specificity are currently unclear.

Computational studies have provided atomic-level details regarding substrate binding and the catalytic mechanisms employed by select nucleoside hydrolases. Specifically, a molecular dynamics (MD) study examined the binding of inosine to IU-NH and the associated catalytic deglycosylation mechanism [40]. It was determined that protonation of inosine at N7 proceeds through a proton shuttle involving Asp14, Tyr229, and His241, which rationalizes the lack of activity of a His241Ala mutant based on the disruption of the hydrogen-bonding network. Another study used MD and free energy calculations to evaluate inhibitor binding and release for CU-NH [41]. This study concluded that His82 is likely neutral

to favour a closed active-site binding conformation. Furthermore, classical and QM/MM MD simulations were used to probe wild-type CU-NH-mediated hydrolysis of U, which identified interactions between the substrate and Asp15, and Gln227 that are critical for the selective binding of U over I [42]. For IAG-NH, *ab initio* calculations indicate that stacking interactions between the bound inosine nucleobase and Trp sidechains (Trp83 and Trp260) facilitate hydrolysis [43]. Subsequent classical and QM/MM MD studies proposed that the most probable deglycosylation mechanism involves protonation of inosine at N7 by water [44, 45]. Docking studies have successfully identified several inhibitors of various nucleoside hydrolases that bind tightly within the active site [24, 25, 46, 47]. These studies highlight the important role of computational modeling in understanding the function of nucleoside hydrolases.

Due to the critical insight obtained to date about nucleoside hydrolase function from computational chemistry, the current study employs classical MD to provide further insight into how the substrate specificity of CU-NH can be modulated. Specifically, we consider the active site dynamics and analyze key nucleoside–enzyme interactions when inosine is bound to the wild-type enzyme, the Thr223Tyr and Gln227Tyr single mutants, and the Thr223Tyr Gln227Tyr double mutant. Through this approach, we uncover a novel route for nucleobase activation in the double mutant, which is likely critical for promoting deglycosylation, and illustrate that the degree of nucleobase activation is dependent on the identity and protonation state of active site residues. As a result, we provide a structural rationale for the observed discrepancy in the inosine hydrolysis rates facilitated by all four wild-type and mutant CU-NH. More generally, this work has several implications for the function of CU-NH and provides insight into differences in the substrate specificities of the various classes of nucleoside hydrolases.

5.2 Computational Methods

MD simulations were initiated from an X-ray crystal structure of *E. coli* CU-NH (YeiK) bound to inosine and Ca^{2+} (PDB ID: 3B9X, denoted the CU-NH–I complex) [33]. Among the four CU-NH–I complexes within the crystal structure, the complex with all residues resolved and that best matches the crystallographic data (based on the R_{free}, Clashscore, and outlier metrics) was chosen as the starting point (chain A). To generate the double mutant complex, Tyr225, Tyr229, and His241 were extracted from a crystal structure of *C. fasciculata* IU-NH bound to a *p*-aminophenyliminoribitol inhibitor (*p*APIR, PDB ID: 2MAS) [15], and superimposed onto Thr223, Gln227, and His239 of the CU-NH–I complex according to the protein backbone. Chemical intuition was used to adjust the position of the sidechain for each new residue to minimize any resulting steric clashes in the active site. The single mutants were similarly derived by replacing only two of the residues of interest in the native CU-NH–I complex with those from the IU-NH crystal structure. Protonation states were initially assigned using H++ [48], and adjusted based on local hydrogen-bonding networks and previous studies. Specifically, His14, His24, and His103 were protonated at Nε, His82 was protonated at Nδ, and His239 was protonated at both Nε and Nδ, while all Asp and Glu residues were modelled as anionic. His82 was modelled as neutral since previous studies indicate that neutral His82 favors a closed active site conformation [41], while His239 was modelled as cationic since this residue has been proposed to play the role of the general acid [33, 34].

Each complex was assigned AMBER parm14SB parameters [49], supplemented with GAFF [50] parameters and Restricted Electrostatic Potential (RESP) charges for the bound inosine nucleoside. To ensure accurate modelling of the metal ion binding site, the Metal Center Parameter Builder (MCPB) [51] utility was used to assign RESP charges to Ca^{2+} and coordinating residues (Asp11, Asp16, Asp240, and the backbone carbonyl of Val124; see Appendix D, Figure D.1 and Table D.1) based on B3LYP/6-31G(d) optimized structures of the Ca^{2+} binding site obtained with Gaussian 09 (revision D.01) [52]. The MCPB

utility was also used to assign force field parameters to the metal ion binding site using the Seminario method (Table D.2 in Appendix D) [53].

Each system was neutralized with Na^+ counter ions, and solvated in a TIP3P water box, while ensuring at least 10.0 Å of water exists between the protein and the water box boundary. The particle mesh Ewald method was employed, with a nonbonded cut-off of 10.0 Å. Constrained (100 kcal mol^{-1} $Å^{-2}$) minimization was performed as follows: (1) relaxation of the solvent and Na^+ ions using 1000 cycles of steepest descent followed by 1000 cycles of conjugate gradient minimization; (2) relaxation of the solute hydrogen atoms using 1000 cycles of steepest descent followed by 1000 cycles of conjugate gradient minimization; and (3) relaxation of the solute using 1000 cycles of steepest descent followed by 1000 cycles of conjugate gradient minimization. Subsequently, the entire system was relaxed without constraints for 1000 cycles of steepest descent followed by 2000 cycles of conjugate gradient minimization. The system was then heated using the Langevin thermostat (collision frequency = 1.0 ps^{-1}) to 310 K over 120 ps using a 1 fs time step and a 25.0 kcal mol^{-1} $Å^{-2}$ restraint on the solute under NVT conditions. The restraint was gradually released (reduced by 5.0 kcal mol^{-1} $Å^{-2}$ every 20 ps) from 25.0 kcal mol^{-1} $Å^{-2}$ to 0.0 kcal mol^{-1} $Å^{-2}$ over 100 ps under NPT conditions (1 atm, 310 K) using a 2 fs time step and SHAKE to constrain bonds involving hydrogen. Each system was further equilibrated for 0.5 ns under the same conditions. Prior to each production simulation, several short (10 to 40 ns) pre-production simulations were performed to explore the accommodation of the substrate within the active site. A representative structure was chosen from these pre-production simulations (using the clustering methodology described below) and three 100 ns replica simulation were initiated from the representative structure for each system, which each employed different initial velocities. Minimization, equilibration, and production steps were performed using the PMEMD module in AMBER 14 [54, 55].

The CPPTRAJ utility [56] was used for all analysis steps and each system was sampled every 50 ps. To evaluate structural changes in the active site, the average root-mean-square deviation (rmsd) and standard deviation (σ) were calculated using the position of heavy atoms relative to the initial crystal structure or another trajectory. Specifically, the overall rmsd fitting was performed based on the position of all heavy atoms in the active site (I, Ca^{2+}, nucleophilic water, Asp11, Asp15, Asp16, His82, Val124, Thr223, Gln227, Tyr231, His239, and Asp240). Subsequently, the rmsd per residue was calculated (with no refitting). The occupancy of hydrogen bonds is reported for the duration of the simulation that the heavy atom distance is < 3.4 Å and the hydrogen-bonding angle is > 120°. To obtain the representative structures displayed in the figures, each simulation was clustered using a hierarchical agglomerative clustering methodology based on the position of active site residues (listed above). Pairwise molecular mechanics generalized Born surface area (MM-GBSA) calculations were performed to estimate the binding interactions between inosine and active site residues. Representative structures discussed in the text match well across the entire trajectory for all 3 replicas (active site rmsds range between ~0.8 and 1.4 Å; Table D.3 in Appendix D). Although the figures depict a representative, static structure, all geometric and energy values are reported as the average across a single production simulation. To provide a visual representation of the placement of water within the active site, a three-dimensional 20 x 20 x 20 Å grid was centered on the bound nucleoside, with 0.50 Å spacing between each point on the grid. In the figures, the red spheres are used to represent an oxygen atom of a water molecule that occupies a grid space for at least 40% of the simulation.

5.3 Results and Discussion

5.3.1 Despite several catalytically conducive interactions in the active site, lack of nucleobase activation explains the inability of wild-type CU-NH to hydrolyze inosine.

An X-ray crystal structure of CU-NH bound to inosine reveals that His239 is misaligned for proton transfer to the inosine nucleobase, which directly correlates with kinetic data that indicates inosine is not a substrate of CU-NH (Figure 5.2b) [33]. To complement this static interpretation, we provide dynamic structural data of CU-NH bound to inosine by examining the wild-type CU-NH–I complex. Our MD structures resemble the native crystal structure, with an average active site rmsd of 1.0 Å (Table 5.1 and Figure D.2 in Appendix D). The sidechains of Asp11, Asp16, and Asp240, the backbone carbonyl of Val124, and the 2′– and 3′–OH groups of inosine ligate to Ca^{2+} (Figure 5.3a), with the coordinating distance ranging between ~2.3 and 2.9 Å (Table D.4 in Appendix D). The shortest coordination distances occur for the Asp and Val residues, while the longest involve the ribose hydroxy groups (Table D.4 in Appendix D). An Asp11(Oδ)···H_2O(OH) hydrogen bond exists (occupancy = 86.7%; Figure 5.3a and Table D.5 in Appendix D) and water is in close proximity to C1′ of inosine (distance = 4.293 Å, Table D.4 in Appendix D), which support proposals that Asp11 is the general base that activates a water nucleophile during nucleoside hydrolysis [33, 34]. The ribose moiety primarily adopts a C4′-endo pucker (216–252°) [57] over the course of the trajectory (Figure D.3 in Appendix D). Hydrogen-bonding interactions between the 2′-, 3′-, and 5′-OH groups of inosine and active site residues stabilize the position of the substrate in the active site (Figure 5.3a). Specifically, the 2′–OH and 5′–OH groups donate a hydrogen bond to Asp240 and Glu164, respectively (occupancy = 100%; Table D.5 in Appendix D), while the 3′–OH group accepts a hydrogen bond from Asn166 (occupancy = 58.8%; Table D.5 in Appendix D). The maintenance of these interactions is unsurprising since the Ca^{2+}-ligating residues are present in each nucleoside hydrolase regardless of substrate specificity [29, 33, 35], which underscores the crucial role

played by these residues in substrate binding. Indeed, it has been proposed that the ribose–protein interactions and ribose pucker likely stabilize the oxocarbenium ion formed in the transition state and thereby promote catalysis [27].

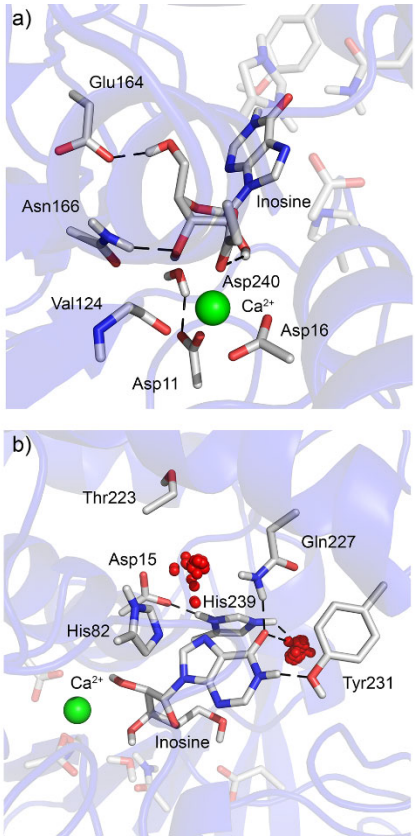


Figure 5.3. Representative MD structure of wild-type CU-NH ($Asp15^{-}$ $His239^{+}$) bound to inosine, highlighting active site residues that surround a) the ribose sugar and Ca^{2+} or b) the inosine nucleobase and the distribution of active site water (depicted as red spheres when occupying a 0.50 Å grid space for at least 40% of the trajectory).

While there are changes to the positions (average rmsd for His82, Thr223, Gln227, Tyr231, and His239 < 1.560 Å; Table 5.1) and orientations (deviation in backbone torsional angles < 56.7°; Figure D.2 and Table D.6 in Appendix D) of the active site residues near the inosine nucleobase, the substrate–protein interactions in the X-ray crystal structure (Figure 5.2b) remain consistent throughout the MD simulations (Figure 5.3b). Specifically, Thr223 does not interact with the bound substrate. In contrast, the sidechain

of His82 interacts with the nucleobase through a T-shaped interaction (average center-of-mass to center-of-mass distance = 5.2 Å and interaction energy = –11.0 kJ mol^{-1}; Tables D.4 and D.7 in Appendix D). Furthermore, Gln227 (NεH) donates a hydrogen bond to O6 of the bound nucleoside (occupancy = 95.2% and interaction energy = –11.9 kJ mol^{-1}; Tables D.5 and D.7 in Appendix D), while Tyr231 accepts a hydrogen bond from N1 of inosine (occupancy = 83.9% and interaction energy = –10.9 kJ mol^{-1}; Tables D.5 and D.7 in Appendix D). His239 interacts with the nucleobase through a T-shaped interaction (distance = 4.927 Å and interaction energy = –13.2 kJ mol^{-1}; Tables D.4 and D.7 in Appendix D). Interestingly, a His239–Asp15 dyad emerges that contains a hydrogen bond between Oδ1 or Oδ2 of Asp15 and NδH of His239 (combined occupancy = 99.0%, Table D.5 in Appendix D). However, no hydrogen bond forms between inosine and His239 (Figure 5.3b), supporting a proposal based on the X-ray crystal structure that His239 is not positioned to directly activate the nucleobase during hydrolysis [33]. Since inosine has been proposed to be activated by water during IAG-NH-mediated deglycosylation [44, 45], we examined the possibility that His239 may transfer a proton to inosine via a water chain. However, water mediates a hydrogen bond between His239 and inosine for only ~1% of the MD trajectory despite high solvation of N7 and O6 of inosine, and NεH of His239 (Figure 5.3b and Table D.5 in Appendix D). Therefore, water-mediated proton transfer to inosine from His239 is also unlikely.

Table 5.1. Average root-mean-square deviation (rmsd) (standard deviation (σ) provided in parentheses) of the entire CU-NH active site or key active site residues across the MD simulation trajectories.[a]

		Active Site	Nucleoside	Asp15	Res223	Res227	Tyr231	His239	His82
		Avg (σ)	Avg (σ)	Avg (σ)	Avg (σ)	Avg (σ)	Avg (σ)	Avg (σ)	Avg (σ)
Crystal structure	Wild ($Asp15^{-}$ $His239^{+}$) [b]	1.003 (0.095)	0.749 (0.139)	1.490 (0.192)	1.468 (0.286)	1.555 (0.280)	1.560 (0.507)	1.194 (0.119)	0.813 (0.610)
Wild ($Asp15^{-}$ $His239^{+}$)[b]	Thr223Tyr Gln227Tyr ($Asp15^{-}$ $His239^{+}$)[b]	0.924 (0.211)	1.716 (0.630)	1.385 (0.195)	-	-	2.782 (1.163)	1.412 (0.549)	1.394 (0.656)
Wild (Asp15 His239)[c]	Thr223Tyr Gln227Tyr (Asp15 His239)[c]	0.866 (0.162)	1.686 (0.650)	1.089 (0.266)	-	-	2.634 (1.143)	1.680 (0.482)	1.436 (0.590)
Thr223Tyr Gln227Tyr ($Asp15^{-}$ $His239^{+}$)[b]	Thr223Tyr Gln227Tyr (Asp15 His239)[c]	1.029 (0.151)	1.756 (0.655)	1.563 (0.211)	2.306 (0.829)	2.588 (1.009)	2.805 (1.125)	1.675 (0.470)	1.373 (0.558)
Wild (Asp15 His239)[c]	Thr223Tyr Gln227Tyr (Asp15 His239)[c]	0.825 (0.230)	1.623 (0.639)	1.110 (0.263)	-	-	2.677 (1.172)	1.517 (0.552)	1.469 (0.656)
Wild ($Asp15^{-}$ $His239^{+}$) [b]	Wild (Asp15 His239)[c]	0.945 (0.192)	1.657 (0.754)	0.952 (0.335)	2.483 (0.895)	2.456 (1.12)	2.824 (1.293)	1.838 (0.487)	1.598 (0.779)
Wild (Asp15 His239)[c]	Thr223Tyr (Asp15 His239)[c]	0.774 (0.186)	1.772 (0.829)	1.062 (0.272)	-	2.611 (1.2)	2.943 (1.408)	1.524 (0.662)	1.459 (0.659)
Wild (Asp15 His239)[c]	Gln227Tyr (Asp15 His239)[c]	0.917 (0.155)	1.764 (0.717)	0.893 (0.324)	2.464 (0.982)	-	2.82 (1.234)	1.374 (0.561)	1.728 (0.505)

[a]rms fit was performed based on the position of all heavy atoms of the active site residues (namely I, Ca^{2+} ion, nucleophilic water, Asp11, Asp15, Asp16, His82, Val124, Res223 (if identities were the same), Res227 (if identities were the same), Tyr231, His239, and Asp240) to yield the average rmsd of the entire CU-NH active site, while the rmsd per residue was determined without refitting. [b]CU-NH with anionic Asp15 and cationic His239. [c]CU-NH with neutral Asp15 and neutral His239.

Overall, our simulations provide the first dynamic structural data that further clarifies why inosine cannot be processed by CU-NH. Specifically, although His239 binds to the inosine nucleobase through T-shaped interactions, His239 is improperly aligned to directly or indirectly transfer a proton to N7 of inosine. This correlates with kinetic data indicating that His239 is important for substrate binding, but not catalytic hydrolysis, of U [33, 34]. Our simulations suggest that neither Thr223, Gln227, nor Tyr231 form catalytic interactions with the substrate upon binding. Indeed, Gln227 interacts with the substrate at O6, which a previous study predicts would have a smaller catalytic effect on the stability of the glycosidic bond of purines than N7 activation [58] and kinetic isotope experiments indicate that N7 is protonated for IU-NH-mediated catalysis of inosine [38]. Furthermore, Tyr231 accepts a hydrogen bond from inosine in the wild-type CU-NH–I complex, which is likely anti-catalytic since it places additional negative charge on the nucleobase. Nevertheless, the interactions between ribose, and Ca^{2+} and active site residues could stabilize the transition state oxocarbenium cation. Furthermore, the consistency of an Asp11(Oδ)···H_2O(OH) hydrogen bond positions a nucleophile for catalysis. Therefore, the non-excision of inosine by wild-type CU-NH is primarily the result of the lack of nucleobase activation.

5.3.2 The active site conformation of the Thr223Tyr Gln227Tyr CU-NH double mutant does not correlate with the experimentally-observed enhanced activity towards inosine when His239 is modelled as the general acid.

The introduction of the Thr223Tyr and Gln227Tyr mutations into the CU-NH active site results in activity toward inosine that is comparable to IU-NH enzymes [17, 33], whose proposed catalytic mechanism involves protonation of N7 by His241 that is possibly mediated by Tyr229 (Figure 5.2d) [38, 40]. Similarly, the CU-NH double mutant has been proposed to facilitate deglycosylation by proton transfer from His239 to N7 of inosine via Tyr227 [33]. The proposed role of Tyr223 is to stabilize the

position of Tyr227 through a hydrogen bond between the Tyr hydroxy groups. Unfortunately, structural data for mutant CU-NH bound to inosine is not available. Therefore, to examine the roles played by His239, Tyr223, and Tyr227, and understand the observed mutant CU-NH activity, we explore the structural dynamics of the Thr223Tyr Gln227Tyr CU-NH double mutant bound to inosine.

In general, the positions of the sugar-binding residues are consistent with wild-type CU-NH (Figure 5.4a and Figure D.4a in Appendix D), with the ψ and φ backbone dihedral angles generally being within 30° (Table D.6 in Appendix D), and the active site rmsd being 0.924 Å (Table 5.1). Additionally, hydrogen bonds are formed between the sugar hydroxy groups and Glu164, Asn166, and Asp240 (occupancy up to 96.3%, Table D.5 in Appendix D). However, the ribose pucker is more flexible in the double mutant than the wild-type enzyme, adopting either the C3′-exo (180°–216°) or C4′-endo conformation (Figure D.3 in Appendix D). An Asp11(Oδ)···H_2O(OH) hydrogen bond is persistent (occupancy = 94.7%, Figure 5.4a and Table D.5 in Appendix D), and water is positioned in the vicinity of C1′ (distance = 4.705 Å; Table D.4 in Appendix D), providing further evidence that Asp11 is the general base and can abstract a proton from a water nucleophile to promote hydrolysis. While the His82 sidechain forms a T-shaped interaction with the nucleobase (distance = 5.034 Å and interaction energy = –10.6 kJ mol^{-1}; Tables D.4 and D.7 in Appendix D), neither Tyr223 nor Tyr227 form significant hydrogen-bonding interactions with inosine (see Appendix D; occupancies < 7 %; Table D.5 and Figure D.4b). Interestingly, a T-shaped interaction is present between the π–system of Tyr223 and edge of Tyr227 (distance = 6.162 Å; Table D.4 in Appendix D and Figure 5.4b), suggesting the positions of these residues are coupled. Additionally, N7 of inosine is minimally solvated (I(N7)···H_2O(OH) occupancy = 6.7%; Figure 5.4 and Table D.5 in Appendix D), and water does not bridge interactions between inosine and Tyr223, Tyr227 or His239 (Figure 5.4b). As discussed for wild-type CU-NH, His239 is misaligned to transfer a proton to the substrate at least in part due to a hydrogen bond between NδH of His239 and Oδ

of Asp15 (occupancy = 100%; Figure 5.4b and Table D.5 in Appendix D). Furthermore, His239 does not interact with Tyr223 or Tyr227, indicating that neither Tyr residue can facilitate proton transfer from His239 to the substrate (Figure 5.4b) as proposed for IU-NH [33]. Nevertheless, His239 interacts with the substrate through a T-shaped interaction (distance = 5.211 Å and interaction energy = –12.8 kJ mol^{-1}; Tables D.4 and D.7 in Appendix D).

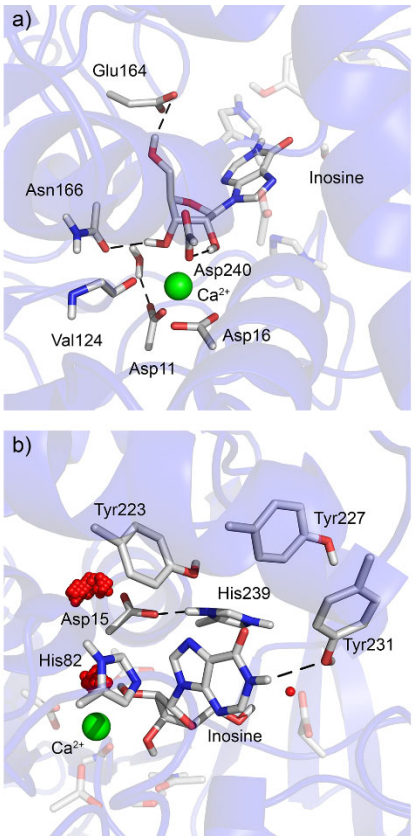


Figure 5.4. Representative MD structure of Thr223Tyr Gln227Tyr CU-NH ($Asp15^{-}$ $His239^{+}$) bound to inosine, highlighting active site residues that surround a) the ribose sugar and Ca^{2+} or b) the inosine nucleobase and the distribution of active site water (depicted as red spheres when occupying a 0.50 Å grid space for at least 40% of the trajectory).

Overall, no additional catalytic interactions exist to activate the nucleobase for departure when the Thr223Tyr Gln227Tyr CU-NH double mutant is bound to inosine compared to the wild-type enzyme, and therefore our simulations do not rationalize the experimentally-observed high activity of the double mutant has toward inosine [33]. However, although His239 was previously proposed to protonate N7 of

inosine via Tyr227 [33], mutational analysis indicates that His239 is an unlikely general acid for the CU-NH-mediated hydrolysis of U and is instead primarily involved in substrate binding [34]. Our results suggest that His239 likely fulfills a similar role when CU-NH binds to inosine. Indeed, His239 forms a T-shaped interaction with the inosine nucleobase, which may facilitate substrate binding, but is unlikely to permit proton transfer to N7. Interestingly, our dynamic structural data reveals a Asp15(Oδ)···His239(NδH) hydrogen bond that is occupied for ~100% of the simulations on both the wild-type enzyme and the Thr223Tyr Gln227Tyr mutant (Table D.5 in Appendix D). However, despite the proximity of Asp15 and His239 in the CU-NH–I crystal structure, these residues do not form a hydrogen bond (Figure 5.2b) [33]. Moreover, when IU-NH is bound to an inhibitor [37], Asp14 and His241 do not interact (Figure 5.2d). This discrepancy between our calculated and the experimental structural data likely arises due to an inconsistent His239 (His241 of IU-NH) protonation state (i.e., the lack of a proton at Nδ in the crystal structures). Therefore, we consider the Thr223Tyr Gln22Tyr CU-NH double mutant bound to inosine with neutral His239 (Nε protonated) in the following section.

5.3.3 An Asp15-Tyr223-I hydrogen-bonding network rationalizes the experimentally-observed activity of the Thr223Tyr Gln227Tyr CU-NH double mutant toward inosine.

Kinetic isotope experiments indicate that N7 of inosine is protonated during IU-NH-mediated hydrolysis [38]. Since the Thr223Tyr and Gln227Tyr mutations to CU-NH introduce the Tyr-Tyr-His active site motif that is characteristic of IU-NH substrate specificity [33], the inosine nucleobase is likely also protonated during hydrolysis catalyzed by the CU-NH double mutant. The previous section provides evidence that His239 is likely neutral. Therefore, other residues must behave as a general acid during hydrolysis of inosine. It is unlikely that His82 will fulfill this role since previous computational studies indicate that His82 must be neutral to favour a closed active site conformation [41]. However, Asp15 is in close proximity to the inosine nucleobase and could act as the general acid when neutral (protonated).

Moreover, in a crystal structure of IU-NH, Asp14 is positioned similarly with respect to the bound inhibitor [37]. Therefore, we hypothesize that Asp15 may be the general acid in the CU-NH double mutant catalyzed deglycosylation of inosine. To explore the effect of the His239 and Asp15 protonation states, we investigate the active site dynamics of the Thr223Tyr Gln227Tyr CU-NH double mutant with neutral Asp15 and His239 bound to inosine.

When Asp15 and His239 are modelled as neutral in the Thr223Tyr Gln227Tyr CU-NH–I complex, the overall active site conformation near ribose is consistent with the wild-type CU-NH–I complex (rmsd = 0.866 Å; Table 5.1 and Figure D.5a in Appendix D). In particular, Glu164, Asn166 and Asp240 stabilize the position of the sugar in the active site through hydrogen-bonding interactions (occupancies up to 100%; Figure 5.5a and Table D.8 in Appendix D). Asp11 consistently hydrogen bonds with a potential nucleophilic water molecule (occupancy = 89.4%; Table D.8 in Appendix D), which is positioned to attack C1′ (distance = 4.022 Å; Table D.4 in Appendix D). Furthermore, the backbone torsional angles of these active site residues deviate by less than 10.0° from the wild-type CU-NH–I complex (Tables D.6 and D.9 in Appendix D). The ribose exhibits similar flexibility regardless of the Asp15 and His239 protonation states, and adopts either a C3′-exo or C4′-endo sugar pucker (Figure D.3 in Appendix D). Additionally, a Tyr231(Oη)···I(N1H) hydrogen bond is formed for the majority of the simulation (89.7%), with a distance of 3.049 Å and angle of 157.7° (Table D.8 in Appendix D). Interestingly, both His82 and His239 stabilize the position of the nucleobase through T-shaped π-interactions (interaction energies = –6.7 and –13.1 kJ mol^{-1}, and distances = 5.483 and 5.102 Å, respectively; Tables D.4 and D.7 in Appendix D).

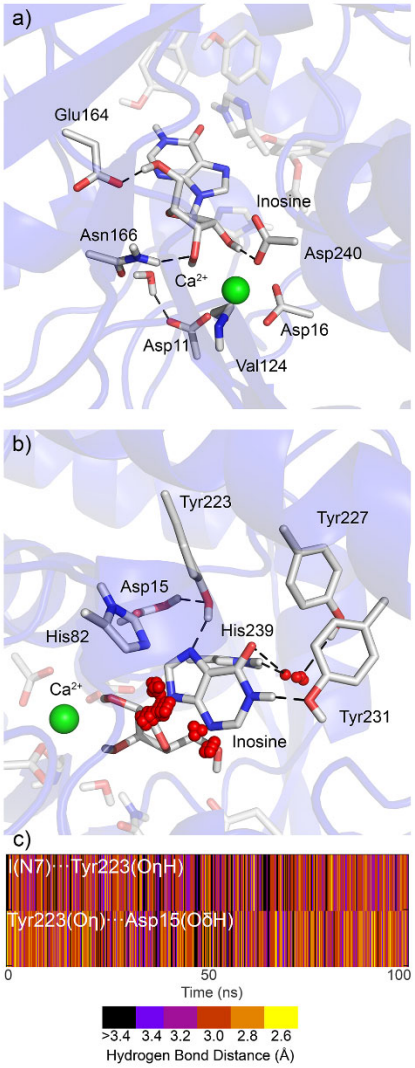


Figure 5.5. Representative MD structure of Thr223Tyr Gln227Tyr CU-NH (Asp15 His239) bound to inosine, highlighting active site residues that surround a) the ribose sugar and Ca^{2+} or b) the inosine nucleobase and the distribution of active site water (depicted as red spheres when occupying a 0.50 Å grid space for at least 40% of the trajectory). c) Hydrogen-bond distance as a function of time, highlighting simultaneous I(N7)···Tyr227(OηH) and Tyr223(Oη)···Asp15(OδH) hydrogen bonds (simultaneous occupancy = 76.4%).

Despite the consistent overall active site conformation discussed above, several interactions occur between the inosine nucleobase and active site residues for the double mutant with neutral Asp15 and His239 that were not present when Asp15 is anionic and His239 is cationic (Figures 4b, 5b, and Figures D.4b, and D.5b in Appendix D). In particular, the positions of Tyr223 and Tyr227 change significantly with

the Asp15 and His239 protonation states (rmsd = 2.306 and 2.588 Å, respectively; Table 5.1). Neutral Asp15 and His239 result in a Tyr223(Oη)···Asp15(OδH) hydrogen bond (occupancy = 80.8%; Table D.8 in Appendix D) and a I(N7)···Tyr223(OηH) interaction (occupancy = 75.6% and interaction energy = –13.6 kJ mol^{-1}; Tables D.7 and D.8 in Appendix D). A time series plot reveals that these two Tyr223 hydrogen bonds are generally formed at the same time (Figure 5.5c). Interestingly, a T-shaped interaction between the π–system of Tyr223 and edge of Tyr227 (distance = 6.055 Å; Figure 5.5b and Table D.4 in Appendix D) is maintained regardless of the Asp15 and His239 protonation states. This contrasts previous proposals that the position of Tyr227 is stabilized by a hydrogen bond with Tyr223 [33].

Both His82 and His239 form π-interactions with the inosine nucleobase (Figure 5.5b). This finding correlates with the ubiquity of π–interactions between nucleic acids and RNA- or DNA-binding proteins and the common proposal that these contacts position substrates or ligands [59, 60]. Furthermore, the catalytic roles of π-interactions are well established in the mechanism of IAG-NH when targeting inosine [21, 29, 30, 43], which uses two Trp residues to stabilize the nucleobase during hydrolysis. In fact, aromatic amino acids also contribute to the catalytic function of enzymes that deglycosylation of DNA, such as AAG [61, 62]. Similarly, the His82 and His239 π-interactions may position the substrate in the active site to promote catalytic deglycosylation by the CU-NH double mutant.

Most importantly, our simulations reveal a Asp15–Tyr223–I hydrogen–bonding network that could result in Asp15 acting as the general acid during hydrolysis by transferring a proton to N7 of inosine through Tyr223. This hydrogen-bonding network cannot form in the wild-type enzyme since neither the position of Thr223 nor Gln227 can mediate a proton transfer to N7 of the substrate. Indeed, when neutral Asp15 and His239 are considered for the wild-type enzyme, Asp15 is not positioned to transfer a proton to inosine and no catalytic interactions with the nucleobase exist (see discussion in the Supporting Information). This is consistent with a previous computational study that concluded Asp15 in wild-type

CU-NH does not play a critical role in substrate binding for inosine [42]. Interestingly, the active site mutations result in the loss of a I(O6)···Gln227(NεH) hydrogen bond, which is replaced by a I(N7)···Tyr223(OηH) hydrogen bond. The predicted similar magnitudes (within 2 kJ mol^{-1}; Table D.7 in Appendix D) suggests these interactions could contribute equally to binding, which directly correlates with the observed similar K_M for inosine in the mutant and wild-type enzymes [33]. However, only the Asp15–Tyr223–I hydrogen-bonding network afforded by the mutant CU-NH could facilitate protonation at N7 of inosine. Indeed, protonation of N7 significantly accelerates deglycosylation of purine nucleosides [45, 58]. Thus, our simulations provide the first structural rationale for the observed nearly 50-fold greater k_{cat} for the double mutant over the wild-type enzyme [33]. Moreover, when our data is combined with previous predictions from computational studies that suggest Asp15 plays a critical role in CU-NH catalyzed hydrolysis of uridine [42], we clarify the catalytic role of Asp15 for both wild-type and mutant CU-NH. Interestingly, Asp14 in IU-NH occupies a similar location relative to Tyr225 [37] as Asp15 of CU-NH relative to Tyr227 (Figure 5.5b). By analogy, it is possible that Asp14 assists protonation of inosine during IU-NH-mediated deglycosylation.

5.3.4 Interactions between Tyr223 and Tyr227 are required for efficient CU-NH mediated hydrolysis of inosine.

Experimental data suggests that both the Thr223Tyr and Gln227Tyr mutations are required to increase the rate of inosine hydrolysis by CU-NH by 50 times over the wild-type enzyme [33]. Furthermore, although the Thr223Tyr mutant catalyzes hydrolysis of inosine at a lower rate than wild-type CU-NH, the Gln227Tyr mutant catalyzes the reaction 7.5 times faster than the wild-type enzyme [33]. To understand the activity of the single mutants relative to the wild-type enzyme and the double mutant, we examined the active site conformational dynamics of the Thr223Tyr and Gln227Tyr CU-NH single mutants bound to inosine with neutral Asp15 and His239.

The overall active site conformation surrounding ribose does not deviate significantly between the single mutants, wild-type CU-NH or the double mutant (Figure D.6 in Appendix D). Indeed, the active site rmsds relative to the wild-type trajectory are minimal (~0.7–0.9 Å; Table 5.1). The ribose pucker is generally maintained at C4′-endo (Figure D.3 in Appendix D), and each complex contains Glu164(Oε)···I(O5′H), I(O3′)···Asn166(NδH), and Asp240(Oδ)···I(O2′H) hydrogen bonds (occupancies up to 100%; Table D.8 in Appendix D). A persistent hydrogen bond is formed between Asp11 and a potentially nucleophilic water (occupancies = ~94%; Table D.8 in Appendix D), which is in proximity to C1′ of inosine (distance = ~3.8–4.4 Å; Table D.4 in Appendix D). Furthermore, the T-shaped interactions between the His82 and His239 imidazole rings, and the inosine nucleobase maintain a distance of 5.2–6.1 Å (Figure 5.6a and b and Table D.4 in Appendix D).

Despite the consistencies in the active site conformations of the single CU-NH mutants discussed above, unique interactions exist between the enzyme and the inosine nucleobase that rationalize observed differences in the catalytic activities [33]. For the Thr223Tyr single mutant, a I(O6)···Gln227(NεH) hydrogen bond exists (occupancy = 93.3%, Figure 5.6a and Table D.8 in Appendix D), which also forms when the wild-type CU-NH is bound to inosine. This interaction is unlikely to provide sufficient nucleobase stabilization to facilitate hydrolysis. Transient Tyr223(OηH)···I(N7) (occupancy = 41.8%; Table D.8 in Appendix D) and Asp15(OδH)···Tyr223(Oη) (occupancy = 76.7%; Figure 5.6a and Table D.8 in Appendix D) hydrogen bonds also occur. However, when the Tyr223(OηH)···I(N7) hydrogen bond is occupied, the Asp15(OδH)···Tyr223(Oη) hydrogen-bond distance is generally larger than 3.4 Å (Figure D.7 in Appendix D). Therefore, unlike the Thr223Tyr Gln227Tyr double mutant (Figure 5.5b), a stable Asp15-Tyr223-I hydrogen-bonding network is likely not persistent enough to afford appreciable nucleobase activation (Figure 5.6a), which explains the observed lack of activity for the Thr223Tyr CU-NH single mutant [33]. Although the CU-NH mutations could change the orientations of active site

residues and distribution of active site water, and thereby affect the preferred protonation states of active site residues [63], our conclusions hold true for alternate His239 and Asp15 protonation states (see discussion in the Appendix D).

In contrast to the Thr223Tyr mutant, a persistent I(N7)···Tyr227(OηH) hydrogen bond occurs in the Gln227Tyr CU-NH active site (occupancy = 86.5%; interaction energy = –13.5 kJ mol^{-1}; distance = 2.892 Å, angle = 159.5°; Figure 5.6b and Tables D.7 and D.8 in Appendix D). Furthermore, a I(O6)···Tyr231(OηH) hydrogen bond exists (occupancy = 59.2%), which is coupled with significant solvation of O6 (occupancy = 81.2%; Table D.8 in Appendix D). Together, these interactions could stabilize the negative charge that develops on the inosine nucleobase during hydrolysis. However, a hydrogen bond does not exist between Tyr227 and Asp15, and therefore full nucleobase protonation is unlikely. Nevertheless, some enzymes that catalyze *N*-glycosidic bond hydrolysis in nucleic acids, such as uracil DNA glycosylase (UDG) [64] and AAG [62], have been proposed to stabilize the departing nucleobase anion through noncovalent interactions, with nucleobase protonation not being explicitly required for bond cleavage. Thus, our predicted nucleobase–amino acid (Tyr227 and Tyr231) active site interactions provide a plausible explanation for the experimental observation that the Gln227Tyr CU-NH single mutant catalyzes hydrolysis of inosine faster than the Thr223Tyr CU-NH mutant and the wild-type enzyme [33]. However, due to the lack of a hydrogen-bonding network to protonate the nucleobase, the Gln227Tyr single mutant is less catalytically efficient than the Thr223Tyr Gln227Tyr CU-NH double mutant. Interestingly, nucleobase activation through hydrogen bonding with active site residues is only achieved when Asp15 and His239 are neutral (see discussion in the Appendix D), which supports our assigned protonation states for both residues. Most importantly, our data on the single mutants suggests that Tyr227 in the CU-NH double mutant plays a critical role in the maintenance of the hydrogen-bonding

network by stabilizing the position of Tyr223 through a T-shaped π-interaction, and further highlights that Tyr223 and Tyr227 function synergistically to enhance the CU-NH hydrolysis rate of inosine.

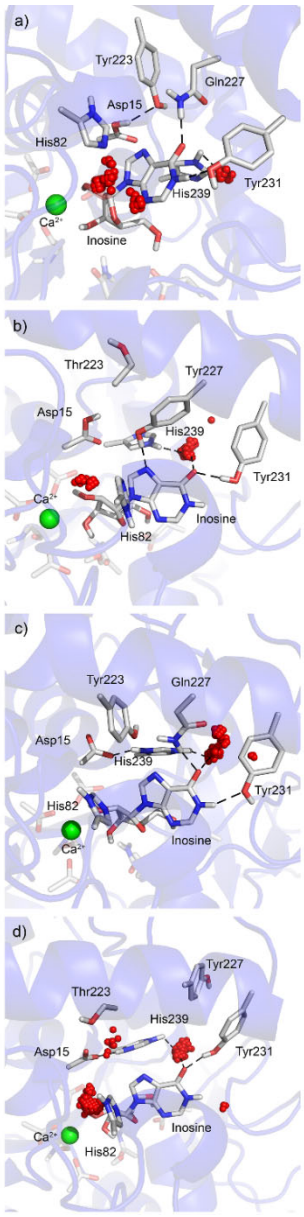


Figure 5.6. Representative MD structures highlighting the active site residues that surround the inosine nucleobase and the distribution of active site water (depicted as red spheres when occupying a 0.50 Å grid space for at least 40% of the trajectory) for a) Thr223Tyr CU-NH (Asp15 His239), b) Gln227Tyr CU-NH (Asp15 His239), c) Thr223Tyr CU-NH ($Asp15^{-}$ $His239^{+}$), or d) Gln227Tyr CU-NH ($Asp15^{-}$ $His239^{+}$) bound to inosine.

5.4 Conclusions

Our molecular dynamics simulation data provides a structural rationalization for the previously reported differential substrate specificity of wild-type and several mutant CU-NH [33]. The interactions between the ribose moiety, and Ca^{2+} and active site residues, as well as the position of nucleophilic water, are catalytically conducive for all wild-type and mutant CU-NH. Furthermore, His82 and His239 form π-interactions with the inosine nucleobase that are likely important for binding. However, unlike previous proposals, His239 is misaligned to transfer a proton to N7 of inosine regardless of the wild-type or mutant CU-NH considered. The resulting lack of nucleobase activation explains the previously observed inactivity of wild-type CU-NH toward inosine. In contrast, an Asp15–Tyr223–I hydrogen-bonding network is uncovered for the Thr223Tyr Gln227Tyr CU-NH double mutant that provides an avenue for nucleobase protonation and inosine hydrolysis, which rationalizes the significant activity of the double mutant toward inosine. This hydrogen-bonding network is stabilized by a critical π-interaction between Tyr223 and Tyr227. Indeed, when Tyr227 is absent, the hydrogen-bonding network is not persistent, which explains the inability of the Thr223Try single mutant to hydrolyze inosine. Similarly, although several hydrogen bonds are observed between inosine and active site residues for the Gln227Tyr single mutant that will partially stabilize the departing nucleobase, these interactions will not result in nucleobase protonation. This directly correlates with the observed rate of the reaction for the Gln227Tyr single mutant being faster than wild-type CU-NH, but slower than the double mutant. Future mutagenesis studies should examine the catalytic role played by Asp15 of the CU-NH double mutant, while computational work should probe the deglycosylation mechanism using a QM/MM methodology, as well as explore the extension of our newly proposed mechanism to other nucleoside hydrolases (IU-NH).

5.5 References

[1] Barrett, M. P., and Croft, S. L. (2012) Management of trypanosomiasis and leishmaniasis, Br. Med. Bull. 104: 175–196.

[2] el Kouni, M. H. (2003) Potential chemotherapeutic targets in the purine metabolism of parasites, Pharmacol. Ther. 99: 283–309.

[3] Delespaux, V., and de Koning, H. P. (2007) Drugs and drug resistance in African trypanosomiasis, Drug. Resist. Updat. 10: 30–50.

[4] Kennedy, P. G. E. (2013) Clinical features, diagnosis, and treatment of human African trypanosomiasis (sleeping sickness), Lancet Neurol. 12: 186–194.

[5] Minodier, P., and Parola, P. (2007) Cutaneous leishmaniasis treatment, Travel Med. Infect. Dis. 5: 150–158.

[6] Garcia, M. N., O'Day, S., Fisher-Hoch, S., Gorchakov, R., Patino, R., Feria Arroyo, T. P., Laing, S. T., Lopez, J. E., Ingber, A., Jones, K. M., et al. (2016) One health interactions of Chagas disease vectors, canid hosts, and human residents along the Texas-Mexico border, PLoS Negl. Trop. Dis. 10: e0005074.

[7] Garcia, M. N., Woc-Colburn, L., Aguilar, D., Hotez, P. J., and Murray, K. O. (2015) Historical perspectives on the epidemiology of human Chagas disease in Texas and recommendations for enhanced understanding of clinical Chagas disease in the southern United States, PLoS Negl. Trop. Dis. 9: e0003981.

[8] Demers, E., Forrest, D. M., and Weichert, G. E. (2013) Cutaneous leishmaniasis in a returning traveller, Can. Med. Assoc. J. 185: 681–683.

[9] Voet, D., and Voet, J. G. (2004) Biochemistry, 3 ed., John Wiley & Sons, New Jersey.

[10] Boitz, J. M., Ullman, B., Jardim, A., and Carter, N. S. (2012) Purine salvage in Leishmania: Complex or simple by design?, Trends Parasitol. 28: 345–352.

[11] Wilson, Z. N., Gilroy, C. A., Boitz, J. M., Ullman, B., and Yates, P. A. (2012) Genetic dissection of pyrimidine biosynthesis and salvage in Leishmania donovani, J. Biol. Chem. 287: 12759–12770.

[12] Carter, N. S., Yates, P., Arendt, C. S., Boitz, J. M., and Ullman, B. (2008) Purine and pyrimidine metabolism in Leishmania, Drug Targets in Kinetoplastid Parasites (Majumder, H., Ed.), pp. 141–154.

[13] Hammond, D. J., and Gutteridge, W. E. (1984) Purine and pyrimidine metabolism in the Trypanosomatidae, Mol. Biochem. Parasitol. 13: 243–261.

[14] Murkin, A. S., and Moynihan, M. M. (2014) Transition-state-guided drug design for treatment of parasitic neglected tropical diseases, Curr. Med. Chem. 21: 1781–1793.

[15] Degano, M., Gopaul, D. N., Scapin, G., Schramm, V. L., and Sacchettini, J. C. (1996) Three-dimensional structure of the inosine-uridine nucleoside N-ribohydrolase from Crithidia fasciculata, Biochemistry 35: 5971–5981.

[16] Horenstein, B. A., Zabinski, R. F., and Schramm, V. L. (1993) A new class of C-nucleoside analogues. 1-(S)-aryl-1,4-dideoxy-1,4-imino-D-ribitols, transition state analogue inhibitors of nucleoside hydrolase, Tetrahedron Lett. 34: 7213–7216.

[17] Shi, W. X., Schramm, V. L., and Almo, S. C. (1999) Nucleoside hydrolase from Leishmania major - cloning, expression, catalytic properties, transition state inhibitors, and the 2.5-angstrom crystal structure, J. Biol. Chem. 274: 21114–21120.

[18] Evans, G. B., Furneaux, R. H., Gainsford, G. J., Schramm, V. L., and Tyler, P. C. (2000) Synthesis of transition state analogue inhibitors for purine nucleoside phosphorylase and N-riboside hydrolases, Tetrahedron 56: 3053–3062.

[19] Berg, M., Van der Veken, P., Goeminne, A., Haemers, A., and Augustyns, K. (2010) Inhibitors of the purine salvage pathway: A valuable approach for antiprotozoal chemotherapy?, Curr. Med. Chem. 17: 2456–2481.

[20] Berg, M., Kohl, L., Van der Veken, P., Joossens, J., Al-Salabi, M. I., Castagna, V., Giannese, F., Cos, P., Versees, W., Steyaert, J., et al. (2010) Evaluation of nucleoside hydrolase inhibitors for treatment of African trypanosomiasis, Antimicrob. Agents Chemother. 54: 1900–1908.

[21] Versees, W., Goeminne, A., Berg, M., Vandemeulebroucke, A., Haemers, A., Augustyns, K., and Steyaert, J. (2009) Crystal structures of T. vivax nucleoside hydrolase in complex with new potent and specific inhibitors, Biochim. Biophys. Acta, Proteins Proteomics 1794: 953–960.

[22] Goeminne, A., Berg, M., McNaughton, M., Bal, G., Surpateanu, G., Van der Veken, P., De Prol, S., Versees, W., Steyaert, J., Haemers, A., et al. (2008) N-arylmethyl substituted iminoribitol derivatives as inhibitors of a purine specific nucleoside hydrolase, Biorg. Med. Chem. 16: 6752–6763.

[23] Goeminne, A., McNaughton, M., Bal, G., Surpateanu, G., Van der Veken, P., De Prol, S., Versees, W., Steyaert, J., Haemers, A., and Augustyns, K. (2008) Synthesis and biochemical evaluation of guanidino-alkyl-ribitol derivatives as nucleoside hydrolase inhibitors, Eur. J. Med. Chem. 43: 315–326.

[24] Renno, M. N., Costa Franca, T. C., Nico, D., Palatnik-de-Sousa, C. B., Tinoco, L. W., and Figueroa-Villar, J. D. (2012) Kinetics and docking studies of two potential new inhibitors of the nucleoside hydrolase from Leishmania donovani, Eur. J. Med. Chem. 56: 301–307.

[25] Franca, T. C. C., Rocha, M. d. R. M., Reboredo, B. M., Renno, M. N., Tinoco, L. W., and Figueroa-Villar, J. D. (2008) Design of inhibitors for nucleoside hydrolase from Leishmania donovani using molecular dynamics studies, J. Braz. Chem. Soc. 19: 64–73.

[26] Goeminne, A., McNaughton, M., et al. (2007) 1,2,3-triazolylalkylribitol derivatives as nucleoside hydrolase inhibitors, Bioorg. Med. Chem. Lett. 17: 2523–2526.

[27] Versees, W., and Steyaert, J. (2003) Catalysis by nucleoside hydrolases, Curr. Opin. Struct. Biol. 13: 731–738.

[28] Vandemeulebroucke, A., De Vos, S., Van Holsbeke, E., Steyaert, J., and Versees, W. (2008) A flexible loop as a functional element in the catalytic mechanism of nucleoside hydrolase from Trypanosoma vivax, J. Biol. Chem. 283: 22272–22282.

[29] Versees, W., Decanniere, K., Van Holsbeke, E., Devroede, N., and Steyaert, J. (2002) Enzyme-substrate interactions in the purine-specific nucleoside hydrolase from Trypanosoma vivax, J. Biol. Chem. 277: 15938–15946.

[30] Versees, W., Decanniere, K., Pelle, R., Depoorter, J., Brosens, E., Parkin, D. W., and Steyaert, J. (2001) Structure and function of a novel purine specific nucleoside hydrolase from Trypanosoma vivax, J. Mol. Biol. 307: 1363–1379.

[31] Arivett, B., Farone, M., Masiragani, R., Burden, A., Judge, S., Osinloye, A., Minici, C., Degano, M., Robinson, M., and Kline, P. (2014) Characterization of inosine-uridine nucleoside hydrolase (RihC) from Escherichia coli, Biochim. Biophys. Acta, Proteins Proteomics 1844: 656–662.

[32] Giannese, F., Berg, M., Van der Veken, P., Castagna, V., Tornaghi, P., Augustyns, K., and Degano, M. (2013) Structures of purine nucleosidase from Trypanosoma brucei bound to isozyme-specific trypanocidals and a novel metalorganic inhibitor, Acta Crystallogr., Sect D: Biol. Crystallogr. 69: 1553–1566.

[33] Iovane, E., Giabba, B., Muzzolini, L., Matafora, V., Fornili, A., Minici, C., Giannese, F., and Degano, M. (2008) Structural basis for substrate specificity in group I nucleoside hydrolases, Biochemistry 47: 4418–4426.

[34] Giabbai, B., and Degano, M. (2004) Crystal structure to 1.7 angstrom of the Escherichia coli pyrimidine nucleoside hydrolase YeiK, a novel candidate for cancer gene therapy, Structure 12: 739–749.

[35] Gopaul, D. N., Meyer, S. L., Degano, M., Sacchettini, J. C., and Schramm, V. L. (1996) Inosine-uridine nucleoside hydrolase from Crithidia fasciculata. Genetic characterization, crystallization, and identification of histidine 241 as a catalytic site residue, Biochemistry 35: 5963–5970.

[36] Vandemeulebroucke, A., Minici, C., Bruno, I., Muzzolini, L., Tornaghi, P., Parkin, D. W., Versees, W., Steyaert, J., and Degano, M. (2010) Structure and mechanism of the 6-oxopurine nucleosidase from Trypanosoma brucei brucei, Biochemistry 49: 8999–9010.

[37] Degano, M., Almo, S. C., Sacchettini, J. C., and Schramm, V. L. (1998) Trypanosomal nucleoside hydrolase. A novel mechanism from the structure with a transition-state inhibitor, Biochemistry 37: 6277–6285.

[38] Horenstein, B. A., Parkin, D. W., Estupinan, B., and Schramm, V. L. (1991) Transition-state analysis of nucleoside hydrolase from Crithidia fasciculata, Biochemistry 30: 10788–10795.

[39] Parkin, D. W., and Schramm, V. L. (1995) Binding modes for substrate and a proposed transition-state analog of protozoan nucleoside hydrolase, Biochemistry 34: 13961–13966.

[40] Mazumder, D., and Bruice, T. C. (2002) Exploring nucleoside hydrolase catalysis in silico: Molecular dynamics study of enzyme-bound substrate and transition state, J. Am. Chem. Soc. 124: 14591–14600.

[41] Fornili, A., Giabbai, B., Garau, G., and Degano, M. (2010) Energy landscapes associated with macromolecular conformational changes from endpoint structures, J. Am. Chem. Soc. 132: 17570–17577.

[42] Fan, F., Chen, N., Wang, Y., Wu, R., and Cao, Z. (2018) QM/MM and MM MD simulations on the pyrimidine-specific nucleoside hydrolase: A comprehensive understanding of enzymatic hydrolysis of uridine, J. Phys. Chem. B 122: 1121-1131.

[43] Versees, W., Loverix, S., Vandemeulebroucke, A., Geerlings, P., and Steyaert, J. (2004) Leaving group activation by aromatic stacking: An alternative to general acid catalysis, J. Mol. Biol. 338: 1–6.

[44] Mazumder-Shivakumar, D., and Bruice, T. C. (2005) Computational study of IAG-nucleoside hydrolase: Determination of the preferred ground state conformation and the role of active site residues, Biochemistry 44: 7805–7817.

[45] Wu, R., Gong, W., Ting, L., Zhang, Y., and Cao, Z. (2012) QM/MM molecular dynamics study of purine-specific nucleoside hydrolase, J. Phys. Chem. B 116: 1984–1991.

[46] Guimaraes, A. P., Oliveira, A. A., da Cunha, E. F. F., Ramalho, T. C., and Franco, T. C. C. (2011) Analysis of Bacillus anthracis nucleoside hydrolase via in silico docking with inhibitors and molecular dynamics simulation, J. Mol. Model. 17: 2939–2951.

[47] Mancini, D. T., Matos, K. S., da Cunha, E. F. F., Assis, T. M., Guimaraes, A. P., Franca, T. C. C., and Ramalho, T. C. (2012) Molecular modeling studies on nucleoside hydrolase from the biological warfare agent Brucella suis, J. Biomol. Struct. Dyn. 30: 125–136.

[48] Anandakrishnan, R., Aguilar, B., and Onufriev, A. V. (2012) H++ 3.0: Automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations, Nucleic Acids Res. 40: W537–W541.

[49] Case, D. A., Darden, T. A., Cheatham, T. E., III, Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Crowley, M., Walker, R. C., Zhang, W., et al. (2008) Amber Tools, Version 1.0 ed., University of California, San Francisco.

[50] Wang, J. M., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and testing of a general amber force field, J. Comput. Chem. 25: 1157–1174.

[51] Li, P., and Merz, K. M. (2016) Mcpb.Py: A python based metal center parameter builder, J. Chem. Inf. Model. 56: 599–604.

[52] Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., et al. (2016) Gaussian 09, Revision D.01 ed., Gaussian, Inc., Wallingford CT.

[53] Seminario, J. M. (1996) Calculation of intramolecular force fields from second-derivative tensors, Int. J. Quantum Chem 60: 1271–1277.

[54] Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S., and Walker, R. C. (2013) Routine microsecond molecular dynamics simulations with amber on GPUs. 2. Explicit solvent particle mesh ewald, J. Chem. Theory Comput. 9: 3878–3888.

[55] Le Grand, S., Götz, A. W., and Walker, R. C. (2013) Spfp: Speed without compromise—a mixed precision model for GPU accelerated molecular dynamics simulations, Comput. Phys. Commun. 184: 374–380.

[56] Roe, D. R., and Cheatham, T. E. (2013) PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data, J. Chem. Theory Comput. 9: 3084–3095.

[57] Altona, C., and Sundaralingam, M. (1972) Conformational analysis of the sugar ring in nucleosides and nucleotides. New description using the concept of pseudorotation, J. Am. Chem. Soc. 94: 8205–8212.

[58] Lenz, S. A. P., Kohout, J. D., and Wetmore, S. D. (2016) Hydrolytic glycosidic bond cleavage in RNA nucleosides: Effects of the 2′-hydroxy group and acid–base catalysis, J. Phys. Chem. B 120: 12795–12806.

[59] Wilson, K. A., Wells, R. A., Abendong, M. N., Anderson, C. B., Kung, R. W., and Wetmore, S. D. (2015) Landscape of π–π and sugar–π contacts in DNA–protein interactions, J. Biomol. Struct. Dyn., 1–14.

[60] Wilson, K. A., Kellie, J. L., and Wetmore, S. D. (2014) DNA-protein pi-interactions in nature: Abundance, structure, composition and strength of contacts between aromatic amino acids and DNA nucleobases or deoxyribose sugar, Nucleic Acids Res. 42: 6726–6741.

[61] Lenz, S. A. P., and Wetmore, S. D. (2016) Evaluating the substrate selectivity of alkyladenine DNA glycosylase: The synergistic interplay of active site flexibility and water reorganization, Biochemistry 55: 798–808.

[62] Rutledge, L. R., and Wetmore, S. D. (2011) Modeling the chemical step utilized by human alkyladenine DNA glycosylase: A concerted mechanism aids in selectively excising damaged purines, J. Am. Chem. Soc. 133: 16258–16269.

[63] Bashford, D., and Karplus, M. (1990) Pka's of ionizable groups in proteins: Atomic detail from a continuum electrostatic model, Biochemistry 29: 10219–10225.

[64] Parikh, S. S., Mol, C. D., Slupphaug, G., Bharati, S., Krokan, H. E., and Tainer, J. A. (1998) Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA, EMBO J. 17: 5214–5226.

Chapter 6: An MD and QM/MM Study on the Substrate Scope of AlkB and ALKBH2: Insights into the Oxidative Repair of Etheno Lesions[a,b,c]

6.1 Introduction

DNA can by modified by endogenous and exogenous agents that alkylate nucleobases and threaten the fidelity of the genome.[1] For example, etheno adducts can form when DNA reacts with metabolites of vinyl chloride or the unsaturated products of lipid peroxidation.[2-3] To date, four etheno adducts have been characterized: 1,N^6-ethenoadenine (1,N^6-εA), 3,N^4-ethenocytosine (3,N^4-εC), 1,N^2-ethenoguanine (1,N^2-εG), and N^2,3-ethenoguanine (N^2,3-εG; Figure 6.1). These lesions are replication blocks for standard DNA polymerases,[4-5] and successful bypass by translesion synthesis polymerases can have mutagenic consequences *in vivo*.[4, 6] Specifically, A→T and A→G substitution mutations occur upon replication of DNA-containing 1,N^6-εA,[4, 6] while C→A and C→T mutations arise when 3,N^4-εC-containing DNA is copied.[4] Both frameshift and substitution (G→A and G→T) mutations result when 1,N^2-εG or 2′-F-N^2,3-εG is replicated, with the 2′-F-N^2,3-εG analogue studied due to the highly labile glycosidic bond of the free N^2,3-εG nucleotide and minimal effect of the 2′ fluoro substituent on repair. [5]

The accumulation of mutations destabilizes the genome, leading to carcinogenesis,[7] and therefore successful repair of mutagenic adducts is paramount.[1] Indeed, deficiencies in the repair of 1,N^6-εA and 3,N^4-εC lesions have been associated with adenocarcinoma lung tumours[8-9] and colorectal carcinoma.[10] Furthermore, N^2,3-εG has been implicated as a cause of vinyl chloride-induced carcinogenesis due to ineffective repair.[5] Several DNA repair pathways act on etheno adducts,[11] including base excision repair (BER). Specifically, alkyladenine DNA glycosylase (AAG) initiates removal of 1,N^6-εA and 1,N^2-εG from

[a] *Journal of the American Chemical Society* reference style used in this Chapter.
[b] Adapted from Lenz, S.A.P., Li, D. and Wetmore, S.D. (2018) An MD and QM/MM Study on the Substrate Scope of AlkB and ALKBH2: Insights into the Oxidative Repair of Etheno Lesions, [*in preparation*].
[c] S.A.P.L. performed all calculations. S.A.P.L. and S.D.W. contributed to the design of calculations, interpretation of data, and writing the manuscript.

DNA,[12-13] while repair of 3,N^4-εC is initiated by methyl-CpG binding domain 4 (MBD4) DNA glycosylase[14-15] and N^2,3-εG is a substrate of 3-methyladenine DNA glycosylase II (AlkA).[16] In 2005, kinetic and mass spectrometry experiments confirmed that *E. coli* AlkB also catalyzes oxidative dealkylation of 1,N^6-εA[4, 17] and 3,N^4-εC,[4] which expanded the substrate scope of AlkB beyond methylated DNA.[18-23] In fact, despite numerous pathways that process alkylated DNA,[11] Delaney and coworkers revealed that only AlkB proficient cells mitigate the mutagenicity of 1,N^6-εA and 3,N^4-εC.[4]

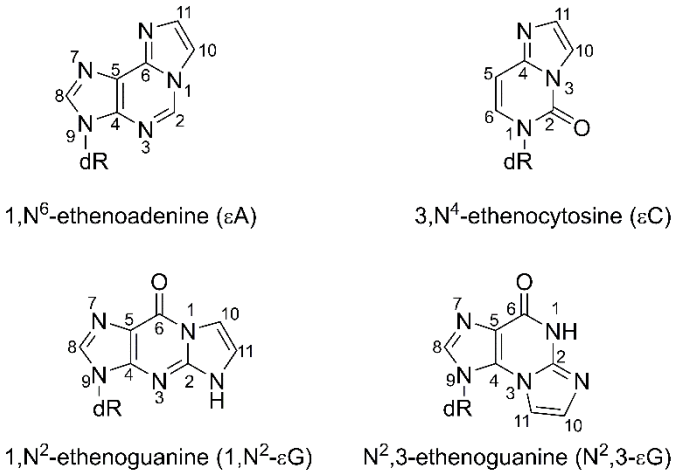


Figure 6.1. Structures and chemical numbering of four etheno adducts.

E. coli AlkB and other members of the AlkB family of enzymes are part of the nonheme α-ketoglutarate- (α-KG)/Fe(II)-dependent dioxygenase superfamily that utilizes molecular oxygen and decarboxylation of α-KG to drive the oxidation of organic molecules (Figure 6.2).[24] Enzymes that are a part of this superfamily perform a diverse number of oxidative transformations and play critical roles within several biological pathways.[25-30] Indeed, AlkB enzymes are ubiquitous throughout life and have been identified in both prokaryotes and eukaryotes, including 9 human homologues (ALKBH1-8 and FTO), although only ALKBH2 and ALKBH3 repair alkylated DNA.[24, 31-33] Among etheno lesions, 1,N^6-εA and 3,N^4-εC are generally dealkylated efficiently by AlkB proteins,[4, 17, 34] while N^2,3-εG is not repaired by *E. coli*

AlkB[5] or other AlkB enzymes including ALKBH2 and ALKBH3.[35] Additionally, kinetic studies performed *in vitro* have determined that only a subset of AlkB proteins are able to remove 1,N^2-εG (including ALKBH2 and ALKBH3, but not *E. coli* AlkB),[34] while *in vivo* sequencing and mass spectrometry experiments suggest that AlkB repairs 1,N^2-εG.[5] Nevertheless, 1,N^2-εG damage is reversed slower than 1,N^6-εA or 3,N^4-εC by AlkB, ALKBH2, and ALKBH3.[34-35]

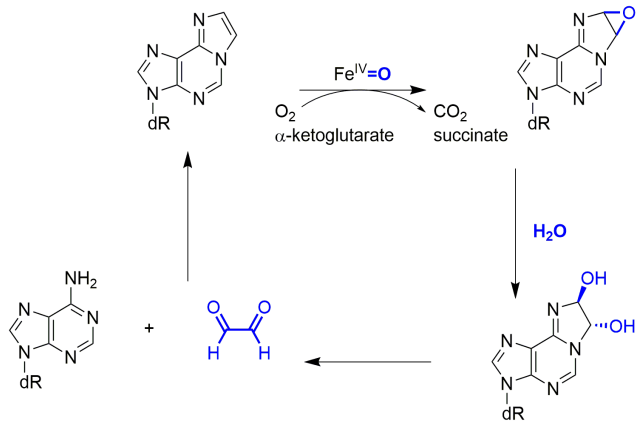


Figure 6.2. Proposed mechanism for the repair of 1,N^6-εA catalyzed by an AlkB enzyme.[4]

AlkB enzymes exhibit different strand preferences when targeting etheno lesions. Specifically, both *E. coli* AlkB and ALKBH3 have higher repair efficiencies for etheno adducts in single-stranded DNA (ssDNA) compared to double-stranded DNA (dsDNA),[34] which correlates with an analogous ssDNA preference for the repair of methyl lesions.[18-23, 33] In contrast, ALKBH2 preferentially repairs both methyl and etheno adducts in dsDNA over ssDNA.[22, 34, 36-37] Crystal structures of AlkB and ALKBH2 bound to DNA explain the strand preference by suggesting unique base flipping strategies for each enzyme.[38] While ALKBH2 uses an aromatic finger residue (Phe102) to facilitate base flipping similar to DNA glycosylases (see, for example, references 38-41), AlkB bends and compresses the DNA backbone to flip the lesion into the active site and stabilizes the extrahelical conformation by forcing the flanking nucleobases into a stacked arrangement.[38] Interestingly, despite conservation of several DNA-binding motifs, the Phe102 finger

residue of ALKBH2 is not conserved in ALKBH3,[33, 38] and an ALKBH3–DNA complex has yet to be crystallized to provide structural clues regarding the base-flipping mechanism. Despite strand preferences for AlkB enzymes, crystal structures of *E. coli* AlkB in complex with ssDNA and dsDNA reveal a remarkably similar active site conformation (rmsd = 0.912 Å),[42] indicating that the DNA context may only impact the base flipping step. Furthermore, AlkB enzymes process 1,N^6-εA and 3,N^4-εC more efficiently than 1,N^2-εG regardless of the ssDNA versus dsDNA contexts,[22, 34, 37] providing evidence that interactions between lesion-containing DNA and the active sites of AlkB enzymes at least in part dictate the substrate specificity.

Crystal structures of AlkB[38, 42-45] and ALKBH2[38, 44-46] bound to cofactors and lesion-containing DNA reveal that these enzymes utilize similar active site architectures to bind substrates (Figure 6.3). Specifically, a 2-His-1-carboxylate (Asp) facial triad that binds Fe(II) and α-KG, which facilitates generation of a high-valent iron-oxo intermediate (Figures 6.2-6.3).[33, 38, 42-46] Several crystal structures have been solved in the presence of the cofactors and Mn(II), which binds to the facial triad in a similar geometry as Fe(II). In AlkB (ALKBH2 numbering in brackets), a hydrogen bond exists between the 5′-phosphate of the lesion and Tyr76 (Ser125), and π–π interactions are present between the nucleobase and His131 (His171) and Trp69 (Phe124). For 1-methyladenine (1MeA) or 3-methylcytosine (3MeC), Asp135 (Glu175) hydrogen bonds with the exocyclic amino group, which likely contributes to the substrate preference for A and C lesions over G and T lesions.[20, 22, 38] However, there is no evidence for direct hydrogen bonds between the nucleobase and active site when AlkB or ALKBH2 are in complex with 1,N^6-εA-containing DNA, although Asp135 (Glu175 in ALKBH2), Glu136, Tyr78 (Tyr122) are positioned to interact with the nucleobase through water-mediated hydrogen-bonding networks.[38, 42-46] Unfortunately, crystal structures of AlkB or ALKBH2 bound to 3,N^4-εC or 1,N^2-εG have yet to be solved and therefore the active site conformation upon binding these lesions is currently unknown.

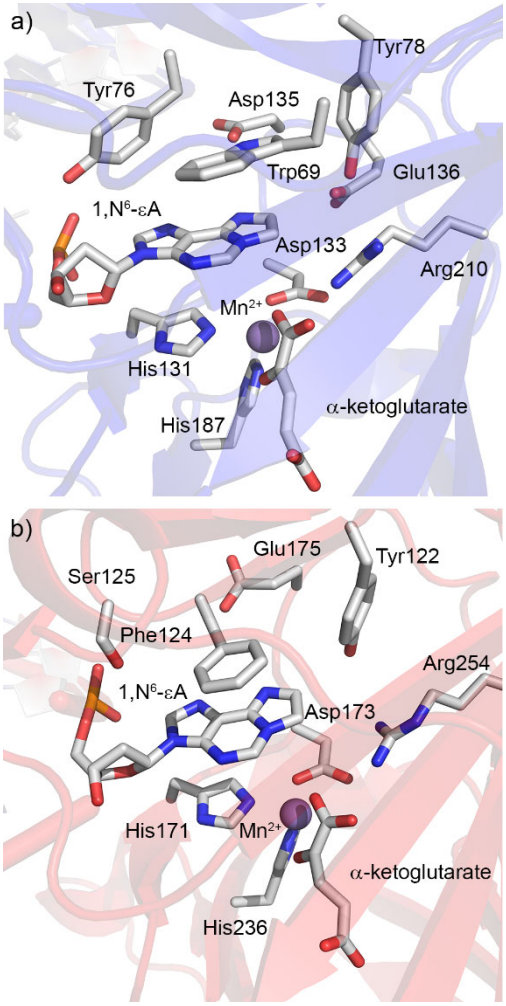


Figure 6.3. X-ray crystal structures of the a) AlkB complex bound to 1,N^6-εA (PDB ID: 3O1P)[43] and b) ALKBH2 complex bound to 1,N^6-εA (PDB ID: 3RZK).[46]

Computational studies have provided valuable insight into the function of AlkB. Specifically, DFT and QM/MM studies have examined *E.* coli AlkB-catalyzed decarboxylation of α-KG to yield succinate and the Fe(IV)–oxo intermediate (Figure 6.2), which confirm a favored quintet spin state for the Fe(IV)-oxo species.[47-49] For the repair of etheno lesions, subsequent oxidation of the aberrant ethene bridge by the oxo ligand has been proposed to proceed through epoxide and glycol intermediates (Figure 6.2).[4-5, 17, 43, 50-51] Indeed, mass spectrometry experiments detected species with the equivalent mass of the putative epoxide and glycol intermediates,[4, 17, 50-51] while the glycol intermediate has been trapped *in crystallo*.[43]

Nevertheless, QM/MM calculations on *E. coli* AlkB-mediated repair of 1,N^6-εA predicted a catalytic mechanism involving oxy zwitterion and glyoxide intermediates.[52] This prediction is consistent with experimental data since the glyoxide intermediate is indistinguishable from the glycol intermediate *in crystallo*, and also proposed aldehyde byproducts have an equivalent molecular mass to the glycol intermediate.

While previous mechanistic and structural studies have provided important insights regarding AlkB- and ALKBH2-mediated repair of 1,N^6-εA, the structural basis for the relative activity of AlkB enzymes toward other etheno lesions is poorly understood. Indeed, to the best of our knowledge, no theoretical study has examined the activity of AlkB on an etheno other than 1,N^6-εA or the activity of ALKBH2 on any etheno adduct. Nevertheless, when 3,N^4-εC, 1,N^2-εG, and N^2,3-εG are superimposed onto a crystal structure of 1,N^6-εA bound by AlkB,[43] N^2,3-εG is the furthest from the metal center (by ~0.5 Å compared to 1,N^6-εA), which was proposed to mitigate repair.[5] However, structural studies of nonheme Fe(II)- and α-KG-dependent enzymes reveal that the distance between the substrate and metal atoms can vary significantly.[53] It is therefore unclear whether a longer distance between N^2,3-εG and Fe(II) could entirely mitigate catalysis or whether there are other factors contributing to the inability of AlkB to oxidize N^2,3-εG. Moreover, NMR studies suggest that AlkB is highly dynamic and that decarboxylation of α-KG to succinate leads to an enhanced dynamic state.[54-55] Thus, significant changes could occur in the position of the lesion relative to the Fe(IV)-oxo moiety and other catalytic active site residues that are not adequately captured by a static structure. It is critical to understand the active site conformation when bound to different lesions to appreciate how AlkB and ALKBH2 repair etheno DNA damage.

The current study combines molecular dynamics (MD) simulations and QM/MM calculations to probe the dynamic active site conformation when AlkB or ALKBH2 are bound to four etheno lesions (1,N^6-εA , 3,N^4-εC, 1,N^2-εG, and N^2,3-εG). We provide the first structural evidence of an AlkB enzyme

bound to the 3,N^4-εC and 1,N^2-εG substrates and the N^2,3-εG non-substrate, which shed light on the lesion-dependent position of catalytic residues and water. To understand the impact of these structural changes on the oxidation mechanism, we characterize the ONIOM(QM:MM) reaction pathway for the first step of AlkB-mediated repair of each etheno lesion. Summed together, our results provide a structural rationale for the observed substrate preference of AlkB enzymes for 1,N^6-εA- and 3,N^4-εC over 1,N^2-εG and explain why N^2,3-εG is not an AlkB or ALKBH2 substrate. More broadly, our data sheds light on the catalytic strategies used by AlkB enzymes to repair alkylated DNA and can be extended to other α-KG/Fe(II)- dependent dioxygenases that play pivotal roles in numerous pathways,[25] including transcription[26-27] and epigenetic regulation,[28] biosynthesis of base J in parasitic trypanosomes,[29] and tRNA modification pathways.[30]

6.2 Computational Methods

In total, 8 unique complexes were considered that contain AlkB or ALKBH2 bound to one of four etheno lesions (namely 1,N^6-εA, 3,N^4-εC, 1,N^2-εG, and N^2,3-εG). X-ray crystal structures of AlkB (PDB IDs: 3O1U and 3O1S)[43] and ALKBH2 (PDB IDs: 3RZK and 3RZJ)[46] bound to lesion-containing DNA were used to build models. In these crystal structures, AlkB is in complex with succinate, Fe(II), and oxidized 1,N^6-εA or 3MeC, while ALKBH2 is in complex with α-KG, Mn(II), and 1,N^6-εA or 3MeC. Starting from the structure of AlkB bound to oxidized 1,N^6-εA, the coordination distance between Fe(II) and the oxygen attached to the nucleobase was shortened to generate the Fe(IV)–oxo AlkB complex. The iron-binding site was then optimized with B3LYP/6-31G(d,p) (H, C, N, and O) and B3LYP/LANL2DZ (Fe(IV)) using Gaussian 16 (revision B.01).[56] Subsequently, the iron-binding site was superimposed onto the metal-binding site of the three other X-ray crystal structures. Starting structures for 3,N^4-εC bound by AlkB or ALKBH2 were generated from the 3MeC-containing crystal structures since these pyrimidine lesions differ by only a single heavy atom. Models of AlkB or ALKBH2 bound to the εG lesions were

generated from the X-ray crystal structures of 1,N^6-εA bound by the corresponding enzyme. Chemical intuition was used to adjust the position of each nucleotide to minimize steric clashes within the active site. Protonation states were initially assigned using H++,[57] and adjusted to preserve local hydrogen-bonding networks. Specifically, for the AlkB complexes, His66, His97, His172, and His197 were protonated at Nε, while His72, His131, and His182 were protonated at Nδ. For the ALKBH2 complexes, His55, His59, His106, His144, and His228 were protonated at Nε, His199 and His220 were protonated at Nδ, and His167 was protonated at both Nε and Nδ. All crosslinking 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 Amber ff14SB forcefield[58] was used to model each complex, with the parameter set assigned to non-standard etheno lesions being supplemented with GAFF parameters.[59] Each etheno adduct was assigned restrained electrostatic potential (RESP) charges calculated based on B3LYP/6-31G(d) optimized geometries (Gaussian 16, revision B.01).[56] The Metal Center Parameter Builder (MCPB.py)[60] 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.[61] Each complex was neutralized with Na^+ counterions and placed in an octahedral TIP3P water box, ensuring that at least 10 Å exists between the protein or DNA and the water box boundary.

The Amber 16 GPU-accelerated and the MPI version of PMEMD[62-63] was used for all equilibration and production steps, while the CPU version of PMEMD was utilized for minimization steps.[64] A 10 Å electrostatic cutoff was implemented using the particle mesh Ewald method. Minimization was performed in several stages, with a 100.0 kcal mol^{-1} $Å^{-2}$ restraint placed on atoms excluded from the minimization. Specifically, the system was minimized with applying 1000 cycles of steepest descent (SD)

and 2000 cycles of conjugate gradient (CG) in the following sequential order: Na^+ counterions and water, complex hydrogen atoms, and complex heavy and hydrogen atoms. Following constrained minimization, the entire system was subjected to 1000 cycles of SD minimization and 2000 cycles of CG unconstrained minimization. Using NVT conditions, the system was heated to 310 K over 120 ps, with a 25.0 kcal mol^{-1} $Å^{-2}$ restraint placed on the solute using the Langevin thermostat and a time step of 1 fs. The applied 25.0 kcal mol^{-1} $Å^{-2}$ restraint was gradually released over 100 ps under NPT conditions, with a 2 fs time step and SHAKE applied to constrain heavy atom–hydrogen bonds. Finally, the complexes were equilibrated during several 100 ns unrestrained simulations to ensure adequate sampling. Representative structures (clustering methodology described below) were chosen from these simulations as starting structures for 1 μs production simulations. The representative structures match well across the trajectory of 3 replicas for each enzyme (active site rmsds range between 0.7 Å and 1.4 Å; Tables E.1 and E.2 in in Appendix E). Coordinates were saved every 5 ps for analysis.

The CPPTRAJ module[65] implemented in Amber 16 was used to analyze the trajectories. Hydrogen bonds were monitored for the duration of each trajectory based on cutoffs for the heavy atom-heavy atom distance of > 3.4 Å and angle of > 120°. A three dimensional 20 x 20 x 20 $Å^3$ grid with 0.5 Å spacing between each point was generated and centered on the bound substrate to visualize the distribution of water in the active site. Red spheres on representative structures reflect an oxygen atom overlapping a grid space for at least 40% of the simulation. The structures in the figures and ONIOM(QM:MM) starting points (*vide infra*) were obtained using a hierarchical agglomerative algorithm to cluster the trajectories according to the positions of the heavy atoms in all key active site residues and the lesion (i.e., Fe(IV), Fe(IV)-ligating residues, and Asp135, Glu136, Trp69, and Arg210 for AlkB or Phe124, Tyr122, Glu175, and Arg254 for ALKBH2, Figure 6.3).

ONIOM (QM:MM) starting points were minimized using a 100.0 kcal mol^{-1} $Å^{-2}$ restraint to fix excluded atoms. Specifically, 2000 cycles of SD minimization and 8000 cycles of CG minimization were sequentially applied to: 1) the solvent and counter ions; 2) the solute hydrogen atoms; 3) the solute excluding active site residues (Figure 6.3); 4) only active site residues; and 5) the entire system without restraints. Subsequently, all water molecules greater than 5 Å from the DNA or protein were truncated. The system was separated into two layers according to the ONIOM formulism: 1) the QM region was treated with B3LYP/6-31G(d,p) for H, C, N, and O atoms and B3LYP/LANL2DZ for Fe, and 2) the MM region was treated with Amber ff14SB supplemented with GAFF and MCPB.py-derived parameters for the bound lesion and iron-ligating site. The QM layer for AlkB contains the sidechains (truncated at the Cα–Cβ bond) of Trp69, His131, Asp133, His187, and Asp135, succinate, Fe(IV), the oxo ligand, the iron-ligated water, the bound nucleotide (truncated at the N9–C1′ bond), and several (typically 4–5) water molecules within 5 Å of the aberrant carbon atoms of the bound lesion. Water greater than 9 Å from the bound nucleotide was held fixed. Each reactant complex (RC) was ONIOM(QM:MM) optimized using electronic embedding (Gaussian 16, revision B.01).[56] Subsequently, a relaxed potential energy surface scan was performed from the optimized RC, by fixing and gradually reducing the reaction coordinate (e.g., distance between the oxo ligand and C10) from ~3.0-3.5 Å to ~1.3 Å in 0.100 Å increments (or 0.05 Å increments around TS regions). For $1,N^6$-εA and $3,N^4$-εC, stationary points were identified from the scans, optimized without constraints, and characterized using frequency calculations. For the εG lesions, constrained stationary points are presented as the unconstrained stationary points had yet to converge at the time of submission. Gibbs energies with scaled (0.9813) zero-point corrections and unscaled thermal corrections are reported in the figures and tables.

6.3 Results and Discussion

6.3.1 Average Dynamic Distance Between the Bound Lesion and Fe(IV)-oxo Group Correlates with Relative Repair Rates for AlkB and ALKBH2

As discussed in the Introduction, the 3,N^4-εC, 1,N^2-εG, and N^2,3-εG nucleotides have been superimposed onto 1,N^6-εA in complex with AlkB, which revealed that the etheno substrates (i.e., 1,N^6-εA, and 3,N^4-εC, 1,N^2-εG) are in close proximity to the Fe(IV)-oxo group, while N^2,3-εG is far from this group.[5] Although this static prediction correlates with kinetic and mass spectrometry data,[4-5, 34] this approach does not adequately account for the reported highly dynamic AlkB complex.[54-55] To more accurately assess how the lesion is positioned with respect to the Fe(IV)-oxo moiety, MD was performed on AlkB and ALKBH2 complexes bound to the aforementioned etheno adducts.

In all modelled structures, the preferred octahedral coordination geometry of iron is maintained, with average distances to Fe(IV) of ~2.2 Å for His residues, ~2.0 Å for water, ~1.9 Å for succinate and Asp, and ~1.6 Å for the oxo ligand (Figure 6.4 and Tables E.3-E.4 in Appendix E). Furthermore, the positions of His131, His187, and Asp133 deviate by only ~0.1 Å for the AlkB complex with 1,N^6-εA compared to the corresponding X-ray crystal structure (Figure E.1 and Table E.3 in Appendix E),[43] while the oxo coordination distance (1.6 Å) is equivalent to that in a previously reported QM/MM structure.[52] These features provide confidence that the iron-coordination step was accurately modelled for assessment of the positioning of each adduct relative to the Fe(IV)-oxo species.

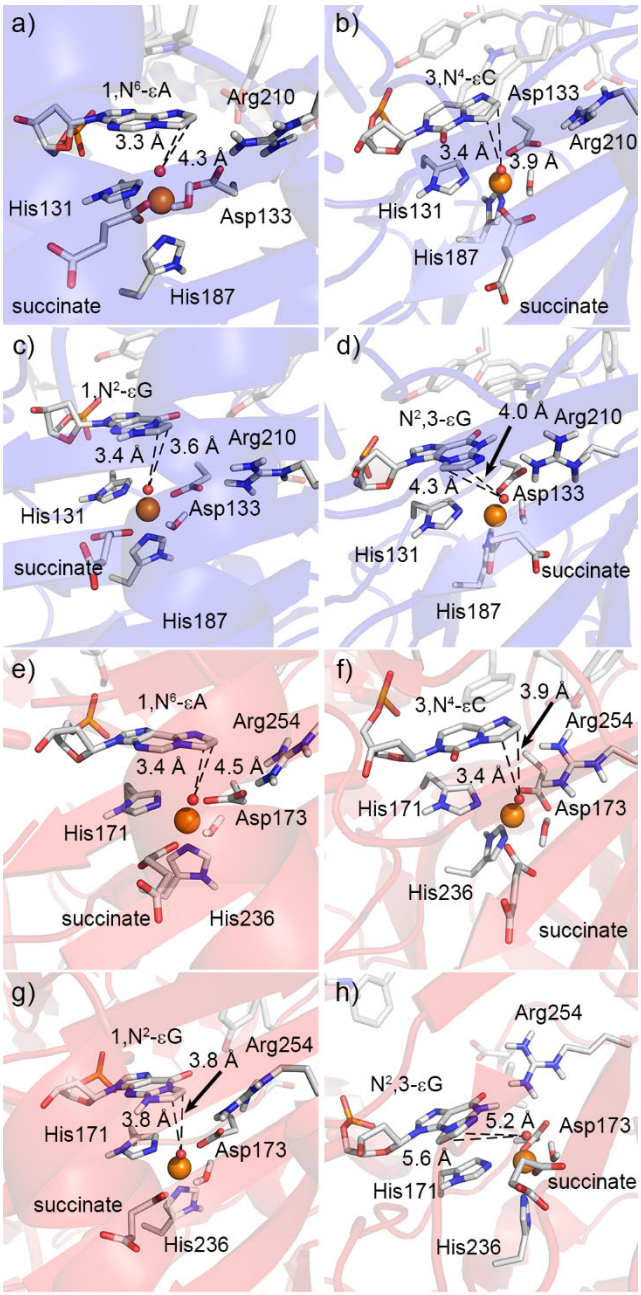


Figure 6.4. Representative MD structures of the AlkB (a-d) or ALKBH2 (e-h) complex bound to 1,N^6-εA (a,e), 3,N^4-εC (b,f), 1,N^2-εG (c,g), or N^2,3-εG (d,h) containing DNA, highlighting the average distances between the Fe(IV)-oxo group and the aberrant atoms of the nucleobase.

For AlkB bound to 1,N^6-εA, 3,N^4-εC, or 1,N^2-εG, the distance between the Fe(IV)-oxo group and closest nucleobase heavy atom (C10 or C11) is 3.3–3.4 Å (Figure 6.4), which matches the equivalent distance reported for a QM/MM reactant of AlkB bound to 1,N^6-εA (3.2 Å).[52] Therefore, the active site is

aligned for the first step of oxidative repair in these AlkB complexes. In contrast, the distance between N^2,3-εG and the Fe(IV)-oxo moiety is longer (4.0 Å; Figure 6.4a-d and Table E.3 in Appendix E). A similar trend occurs for the ALKBH2 complexes, where the substrates are near the oxo ligand (oxo···C10 distance = 3.4–3.8 Å; Figure 6.4e-h), while N^2,3-εG is significantly further away (5.2 Å, Figures 6.4e-h and Table E.4 in Appendix E). Although the substrates are slightly further from the Fe(IV)-oxo moiety in ALKBH2 compared to AlkB, the standard deviation in this distance is also greater for ALKBH2 (±0.4–0.7 Å) compared to AlkB (±0.2–0.4 Å) due to increased flexibility of the bound substrate. This flexibility arises since there are fewer π-interactions between the nucleobase and active site residues in ALKBH2 (i.e. the Tyr76···nucleobase interaction is absent, Figure 6.3). Regardless, the substrates are in close proximity to the Fe(IV)-oxo group for significant portions of the simulations. More importantly, our structural data correlates with the relative AlkB and ALKBH2 repair efficiencies of the etheno lesions,[5, 22, 34, 37] with the shortest distance between Fe(IV)-oxo and the aberrant carbon atoms arising for the strongest substrates $1,N^6$-εA and $3,N^4$-εC, a longer distance for the weaker substrate $1,N^2$-εG, and the longest distance for the non-substrate N^2,3-εG.

A comparison of our structural data to X-ray crystal structures for AlkB and ALKBH2, and other nonheme α-KG/Fe(II)-dependent dioxygenases may provide further clues regarding the correlation between the Fe(IV)–oxo···nucleobase distance and the catalytic activity toward different etheno adducts.[4-5, 34] Since Fe(IV)–oxo intermediates are highly reactive and not possible to trap *in crystallo*, the distance between the metal ion and aberrant carbon atoms must be compared. MD simulations suggest that $1,N^6$-εA and $3,N^4$-εC are positioned ~4.5 Å from the metal center of AlkB or ALKBH2, which matches the corresponding distances (4.3–4.5 Å) in X-ray crystal structures of either enzyme in complex with $1,N^6$-εA.[43, 46] In contrast, $1,N^2$-εG is predicted to fall 5.0 Å from the metal center (Tables E.3 and E.4 in Appendix E), while the corresponding distance for N^2,3-εG is even longer (5.3–6.1 Å; Tables E.3 and E.4

in Appendix E). Nevertheless, the proximity of N^2,3-εG to Fe(IV) is comparable to X-ray crystal structures of other α-KG/Fe(II)-dependent dioxygenases, including JmjC-methyl lysyl demethylases that positions substrates up to 5.8 Å away from the metal ion.[26, 53] Therefore, while the proximity of the lesion to the Fe(IV)-oxo group may be one factor that influences the relative AlkB and ALKBH2 repair rates of etheno lesion-containing DNA, other factors must also play significant roles in dictating whether the DNA–enzyme complex is biologically active.

6.3.2 π-interactions Stabilize Etheno Adducts within the Active Sites of AlkB and ALKBH2

Crystal structures of AlkB or ALKBH2 bound to $1,N^6$-εA suggest no direct hydrogen bonds occur between the nucleobase and active site residues (Figure 6.3).[38, 42-46] However, π–π stacking interactions are formed between the nucleobase and His131/Trp69 in AlkB, and His171 and Phe124 in ALKBH2. Such π-interactions are ubiquitous throughout biology,[67] and have been shown to play a critical role in positioning alkylated DNA adducts in AAG-mediated repair.[68-70] When the dynamics of the repair were examined in the current study, all alkylated nucleobases maintain a π-stacking interaction with the sidechain of His131 in AlkB (His171 in ALKBH2), with a ~3.5–4.5 Å distance (Figure 6.5 and Tables E.3-E.4 in Appendix E). In AlkB, a T-shaped π–interaction also occurs between and each nucleobase (distance = 4.5–5.4 Å, Table E.1) and Tyr76, which matches the crystallographic orientation of Tyr76 when AlkB is in complex with $1,N^6$-εA (Figure 6.3).[43, 45] Interestingly, Tyr76 also forms a direct or water-bridged hydrogen bond with the 5′-phosphate (occupancies up to 71.5%, distances = ~2.8 Å, angle = ~163°; Table E.5 and Figure E.2 in Appendix E), while Ser125 is involved in an equivalent interaction in ALKBH2 (occupancies up to 98.8%, distances = ~2.7 Å, angle = ~164°; Figure 6.5 and Table E.6 in Appendix E). These interactions have been proposed to be important to stabilize a closed and catalytically active configuration of the active site based on X-ray crystal structures of AlkB and ALKBH2 bound to $1,N^6$-εA.[38, 42]

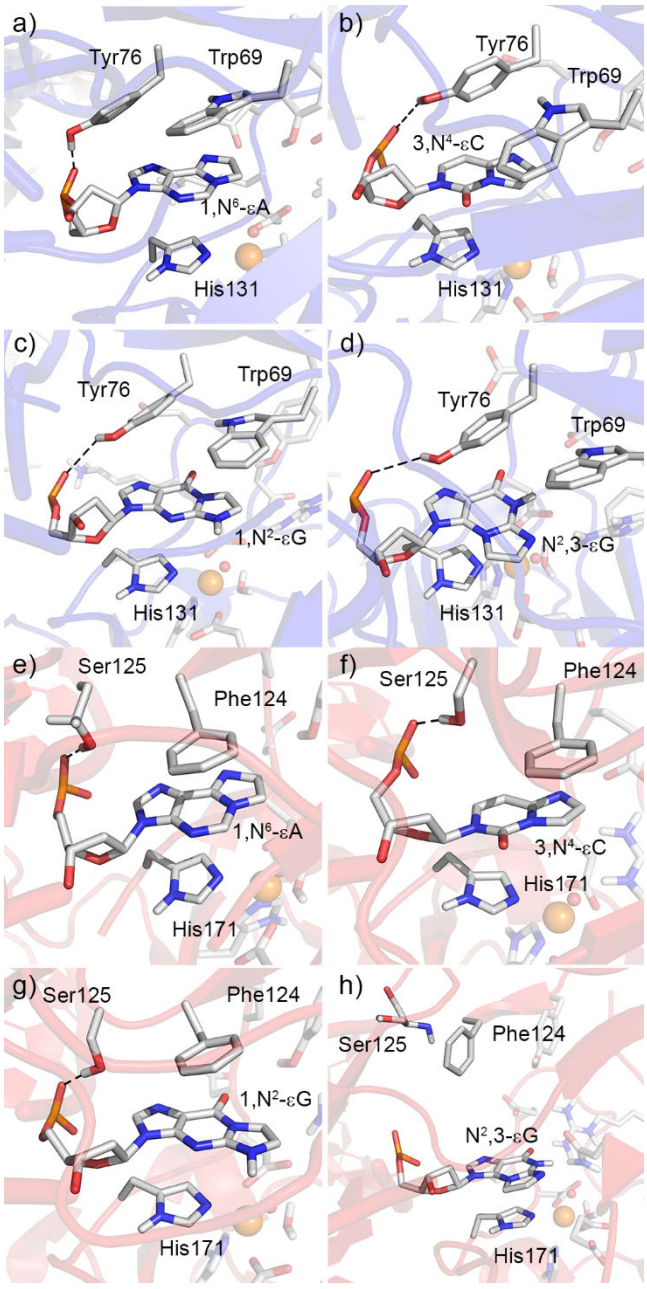


Figure 6.5. Representative MD structures of the AlkB (a-d) or ALKBH2 (e-h) complex bound to 1,N^6-εA (a,e), 3,N^4-εC (b,f), 1,N^2-εG (c,g), or N^2,3-εG (d,h) containing DNA, highlighting the hydrogen-bonding interactions between the 5′-phosphate and active site residues, and the π-interactions between the nucleobase and aromatic residues.

Although the stacking distance between the nucleobase and Trp69 is ~5.4–5.8 Å when AlkB is bound to the etheno substrates (1,N^6-εA, 3,N^4-εC, or 1,N^2-εG), the distance increases to 6.6 Å when AlkB is bound to a non-substrate (N^2,3-εG; Table E.3 in Appendix E). Indeed, the nucleobases of the substrates

overlap with Trp69 significantly, which contrasts with the poor alignment of N^2,3-εG and Trp69 (Figure 5 and Figure E.3 in Appendix E). Similarly, a π–π interaction exists between Phe124 and the nucleobase when ALKBH2 is bound to 1,N^6-εA, 3,N^4-εC, or 1,N^2-εG (distance = ~4.0–4.9 Å), but is generally not present for N^2,3-εG (distance = 8.3 Å; Figure 5 and Tables E.3. and E.4 in Appendix E). For N^2,3-εG bound by ALKBH2, the nucleotide is shifted away from Phe124 due to hydrogen bonds between the nucleobase and Asn169 (occupancy = 70.4%, distance = 3.0 Å , angle = 145°) and Arg254 (occupancy = 74.0%, distance = 2.9 Å, angle = 144°), and disruption of the hydrogen bond between the 5′-phosphate and Ser125 (occupancy = 16.2%, distance = 2.7 Å, angle = 164°; Figures E.3 and E.4 and Table E.6 in Appendix E). When compared to the consistent location of the substrates within the active sites (Figure E.3 in Appendix E), this unique positioning of N^2,3-εG may contribute to the lack of repair activity. Indeed, when the active site configurations for AlkB and ALKBH2 are compared when bound to the same etheno substrate, similar orientations of the nucleobase with respect to Trp69 in AlkB and Phe124 in ALKBH2 suggest these residues fulfill similar roles in alkylated DNA repair (Figure E.5 in Appendix E).

Overall, we propose that π-interactions play key roles in the AlkB and ALKBH2-mediated repair of alkylated DNA, including positioning substrates in catalytically conducive conformations. This suggestion correlates with the critical binding and catalytic roles that π-interactions have been shown to play in AAG-mediated repair of damaged DNA (including 1,N^6-εA)[68-70] and that disruption of these interactions mitigates AAG activity.[69-70] Interestingly, ten-eleven translocation enzyme 2 (TET2) is another α-KG/Fe(II)-dependent dioxygenase that oxidizes DNA and a π-contact is present between Tyr1902 and a bound 5-methylcytosine (distance = 4.1 Å).[28] Our data suggests the π-contact in TET2 may stabilize the position of the 5-methyl group in the vicinity of the Fe(II) center. Thus, π-interactions are not only important for the repair of alkylated DNA, but may also play a broader role in biology by positioning substrates in the active sites of α-KG/Fe(II)-dependent dioxygenases.

6.3.3 Catalytic Residues and Water are Aligned for AlkB- or ALKBH2-mediated oxidation of 1,N^6-εA and 3,N^4-εC, but Disrupted for 1,N^2-εG and N^2,3-εG

Asp135 or Glu136 in AlkB (Glu175 in ALKBH2) have been implicated in the substrate preference of A and C lesions over T and G lesions.[20, 22, 38] Indeed, in crystal structures of AlkB or ALKBH2 in complex with 1,N^6-εA, these residues may interact with the nucleobase through hydrogen-bonding networks involving water. Moreover, the position of Tyr78 (Tyr122 in ALKB2) suggests possible interactions (direct or water-mediated) with the bound substrate. Therefore, these residues may be important for substrate binding and repair.

Our simulations of AlkB and ALKBH2 bound to 1,N^6-εA or 3,N^4-εC reveal that Asp135 and Glu175, and Tyr78 and Tyr122 are equivalently positioned with respect to the lesion (Figure 6.6 and Figure E.6 in Appendix E). Specifically, the distance between the C10–C11 bond midpoint and Asp135 or Glu136 in AlkB is generally between 4.0 and 8.0 Å (Table E.3 and Figure E.7 in Appendix E), while Glu175 is <6 Å away from the lesion for ALKBH2 (Figure E.8 in Appendix E). The location and orientation of these residues relative to 1,N^6-εA matches the corresponding X-ray crystal structure (Figure E.9 in Appendix E).[43, 46] However, slight changes to the positions of these residues arise when either enzyme is bound to 3,N^4-εC, likely due to the smaller nucleobase (Figure 6.6ab and Figure E.9 in Appendix E), which permits increased solvation of the oxo ligand, and may not be favorable for the first step of repair (occupancy = 76%; Tables E.5 and E.6 and Figure E.10 in Appendix E). Nevertheless, a persistent solvent bridge exists between N6 of 1,N^6-εA or N4 of 3,N^4-εC and the carboxylate groups of Asp135 and/or Glu136 in AlkB (Glu175 in ALKBH2) (occupancy = 100%; Tables E.5 and E.6 and Figure E.11 in Appendix E), while Tyr78 (Tyr122) stabilizes the position of Glu136 (Glu175 in ALKBH2) through a water-mediated hydrogen bond. Thus, the solvent-bridged interactions between the active site residues and 1,N^6-εA or N4 of 3,N^4-εC likely plays an important role in substrate binding. Interestingly, Asp135,

Glu136, Tyr78, and water have been proposed play an active role in AlkB-catalyzed repair of 1,N^6-εA, including activation of water to assist the formation of the glycol[50] or glyoxide[52] repair intermediates. Our calculations suggest that these residues (including water) are in position to facilitate repair of 1,N^6-εA and 3,N^4-εC, while Glu175 and Tyr122 may play an equivalent role in the ALKBH2 complexes. Nevertheless, the position of the residues could change significantly during repair due to the highly dynamic nature of the active site of both enzymes.[54-55]

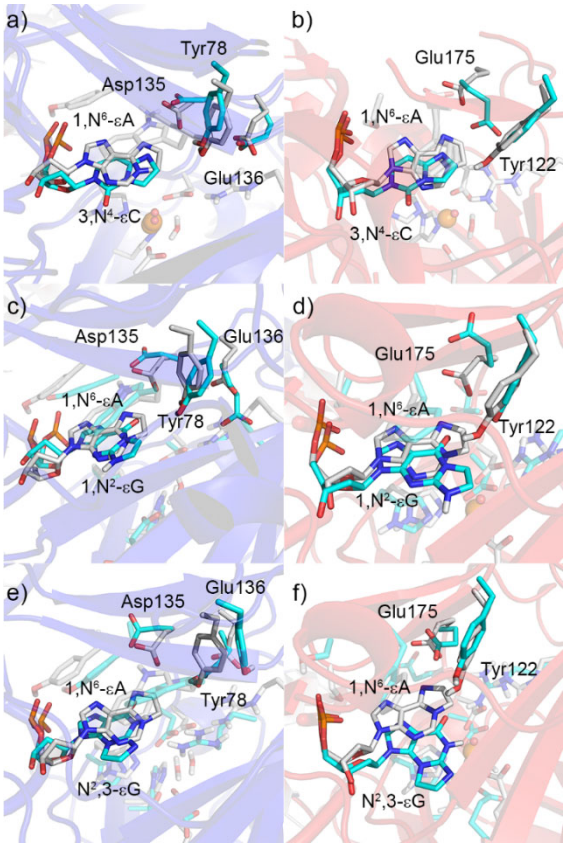


Figure 6.6. Overlays of representative MD structures of the AlkB (a,c,e) or ALKBH2 (b,d,f) complex bound to 1,N^6-εA and 3,N^4-εC (a,b), 1,N^2-εG (c,d), or N^2,3-εG (e,f) containing DNA, highlighting the lesion-dependent positions of Asp135, Glu136, and Tyr78 for AlkB, and Glu175 and Tyr122 for ALKBH2.

For 1,N^2-εG or N^2,3-εG bound by AlkB or ALKBH2, the distances between the aberrant carbon atoms and Asp135, Glu136, and Tyr78 in AlkB (Glu175 and Tyr122 in ALKBH2) increases significantly

(by ~2–8 Å) compared to either enzyme bound to 1,N^6-εA or 3,N^4-εC (Tables E.3-E.4 and Figures 6.6cd and E.6–E.8 in Appendix E). These changes are in part due to electrostatic repulsion between O6 of the nucleobase and the carboxylate atoms of Asp135 or Glu175 (Figure 6.6c-f). For 1,N^2-εG, significant solvation of the oxo ligand is observed, with a water-bridged interaction between the oxo ligand and N2H of the nucleobase, which may mitigate oxidation of the lesion (Figure E.10cg in Appendix E). In contrast, water hydrogen bonds with N1H of the N^2,3-εG and does not solvate the oxo ligand (Figures E.10dh and E.11 in Appendix E). Nevertheless, solvent bridges are observed between O6 of the εG lesions and the carboxylate groups of Asp135 and Glu136 in AlkB or Glu175 in ALKBH2 (occupancy >83%; Tables E.5 and E.6 and Figure E.11 in Appendix E). Furthermore, water is generally >3.5 Å away from the aberrant carbon atoms (Figure E.12 in Appendix E), which indicates that water is not readily available for the formation of glycol or glyoxide intermediates, although the position of the residues could change significantly during repair.[54,55] Interestingly, the aberrant atoms of N^2,3-εG are situated in a hydrophobic region of the AlkB active site , while N^2,3-εG forms a cation–π interaction with Arg100 when bound by ALKBH2 (distance = 6.5 Å, Table E.4 and Figure E.13 in Appendix E). These interactions are likely anti-catalytic due to the anticipated poor stabilization of polar and charged repair intermediates, including the positive charge forming in the oxy zwitterion intermediate previously identified by QM/MM studies on AlkB-catalyzed repair of 1,N^6-εA.[52] Thus, the nonoptimal position of active site residues and water for oxidative repair may be additional factors for the weak activity toward 1,N^2-εG and lack of activity toward N^2,3-εG for AlkB enzymes.[5, 34]

6.3.4 The Lesion-Dependent Active Site Conformation Enhances AlkB-catalyzed Oxidation of 1,N^6-εA and 3,N^4-εC, but Significantly Impedes Oxidation of 1,N^2-εG and N^2,3-εG

QM/MM mechanistic studies have proven to be a valuable tool for assessing the impact of changes in the active site conformations on the catalytic activity.[68, 70] Since MD simulations reveal similar active site conformations for AlkB and ALKBH2 bound to the same lesion and there is more experimental data for AlkB than ALKBH2 to validate our predictions (Figures E.5 and E.6 in Appendix E), we focus AlkB-catalyzed oxidation of the etheno lesions as a representative model for the AlkB family of enzymes. Kinetic data indicates that the overall Gibbs energy barrier for AlkB-catalyzed repair of 1,N^6-εA is ~92 kJ mol^{-1},[4, 17] while QM/MM calculations predict that the rate determining step is one of the final steps of the reaction.[52] Indeed, the first step of oxidation is unlikely to be rate-determining for any of the etheno adducts since the initial step for 1,N^6-εA repair is predicted to be 22.2 kJ mol^{-1}.[52] Since the first step should be rapid and is likely the most affected by changes to the active site conformation, we model this step with ONIOM(QM:MM).

The ONIOM(QM:MM) reactant complexes (RC) closely match the MD representative structures (rmsd <0.375 Å, Figure E.13 in Appendix E), and further highlight the unique lesion-specific active site conformation. For 1,N^6-εA and 3,N^4-εC, the lesion is positioned close to the Fe(IV)-oxo moiety (~3.0–3.2 Å), which is likely critical for oxidative catalysis (Figure 6.7ab). A solvent bridge is present between N6 of 1,N^6-εA (N4 of 3,N^4-εC) and Asp135, which stabilizes the position of the substrate in the active site. As the reaction proceeds from the RC to the transition state complexes (TS), the nucleobase approaches the oxo ligand (oxo···C10 distance = 2.4–2.6 Å), while the oxo coordination distance increases slightly (by ~0.06 Å). The TS is stabilized by tightening of the π-interactions between the nucleobase and His131 (~0.2 Å decrease from the RC). Despite the nucleobase approaching the Fe(IV)-oxo group, the stacking distance between the nucleobase and Trp69 remains unchanged compared to the RC, indicating

that Trp69 remains closely associated with the nucleobase (Figure 6.7ab). Together, this suggests that in addition to facilitating substrate binding, His131 and Trp69 play catalytic roles. In the resulting intermediate complexes (IC), a bond is formed between the oxo ligand and substrate (oxo···C10 distance = ~1.3 Å), while the oxo ligand remains coordinated to iron (distance = ~2.0–2.2 Å). The unconstrained barriers (5.9 kJ mol^{-1} for 1,N^6-εA and 5.4 kJ mol^{-1} for 3,N^4-εC; Figure 6.7ab) are low due to the high reactivity of oxyferryl intermediates and the optimal position of the substrate with respect to the oxo ligand, His131, and Trp69. These barriers vary slightly from the barriers predicted from the constrained pathway (10.0 kJ mol^{-1} for 1,N^6-εA and 5.3 kJ mol^{-1} for 3,N^4-εC). Our QM/MM calculated barrier for 1,N^6-εA is smaller than that previously reported (5.3 kcal mol^{-1});[52] likely since inclusion of Trp69 in our high-level layer stabilizes the charge developing on the nucleobase throughout the reaction. Nevertheless, the interactions between the nucleobase and His131/Trp69 are present in all (MD or QM/MM) enzyme–DNA complexes regardless of the bound substrate, which emphasizes the potential mechanistic roles of these residues.

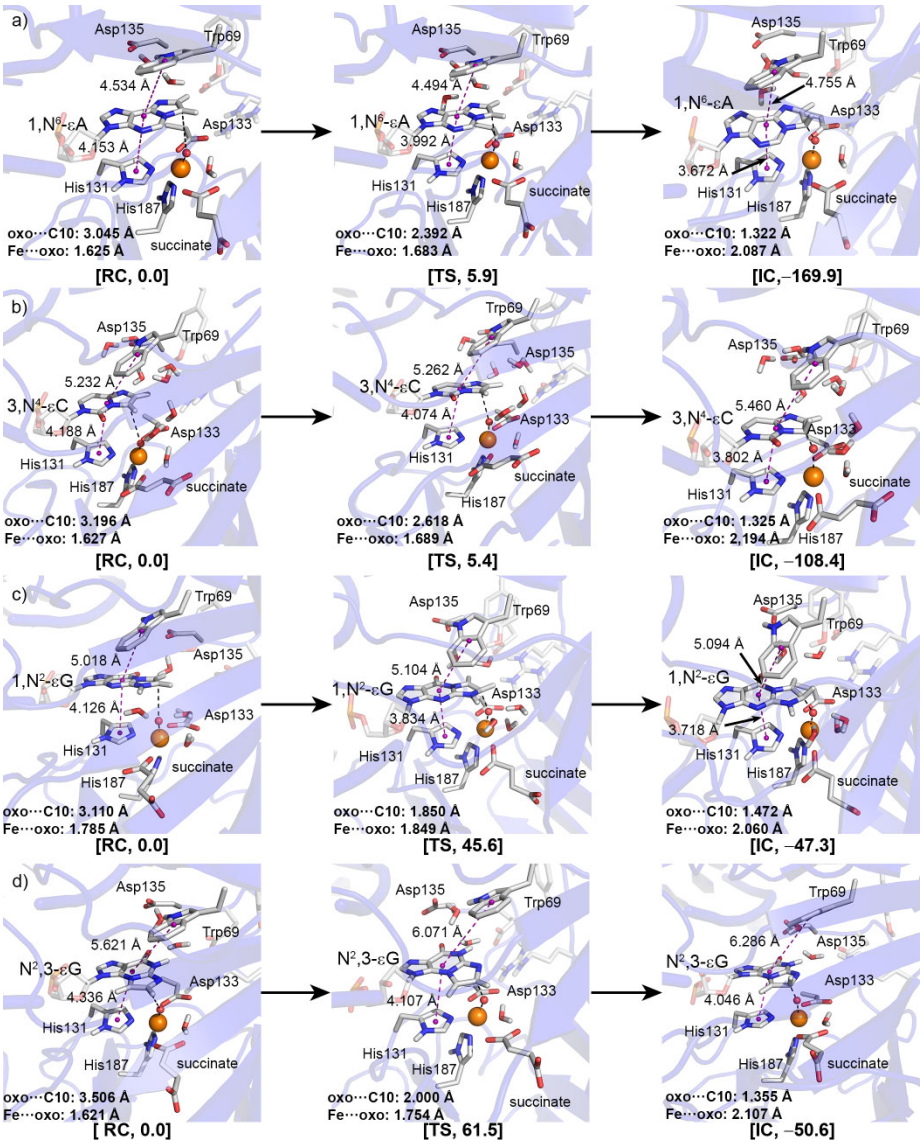


Figure 6.7. ONIOM(QM:MM) stationary points (QM layer in sticks, MM layer in transparent sticks or cartoon) corresponding to the first step of AlkB-catalyzed oxidation of a) 1,N^6-εA, b) 3,N^4-εC, c) 1,N^2-εG, and d) N^2,3-εG. Relative Gibbs energies (kJ mol^{-1}) are displayed in square brackets. Center-of-mass of the nucleobase, Trp69, and His131 are represented with purple spheres.

For 1,N^2-εG, the bound nucleotide is positioned at a similar distance from the Fe(IV)-oxo moiety (3.110 Å) as 1,N^6-εA and 3,N^4-εC in the RC (Figure 6.7c). Moreover, π-contacts between the nucleobase and His131/Trp69 are maintained in the TS, suggesting that these interactions can play critical roles in catalyzing AlkB-mediated oxidation of 1,N^2-εG. However, a solvent bridge is observed between the oxo

ligand and the nucleobase (N2), which increases the coordination distance to Fe(IV) by ~0.2 Å compared to 1,N^6-εA. This likely delocalizes the negative charge of the oxo ligand thereby lowers the oxidizing power of the Fe(IV)-oxo group. Consequently, a later TS occurs for 1,N^2-εG (oxo···C11 distance = ~1.850 Å) with a larger barrier (45.6 kJ mol^{-1}) compared to 1,N^6-εA and 3,N^4-εC. Thus, solvation of the oxo ligand is at least one factor that contributes to the lower activity of AlkB toward 1,N^2-εG.[5]

For N^2,3-εG, the bound nucleotide is further from the oxo ligand (3.506 Å; Figure 6.7d) compared to the RC for any of the substrates. Although a π-contact is formed between the nucleobase and His131, the distance between the nucleobase and Trp69 is longest for N^2,3-εG (5.621 Å; Figure 6.7) indicating that no interaction is present. Similar to 1,N^2-εG, the N^2,3-εG TS is later compared to the TS of the strong etheno substrates (oxo···C11 distance = 1.965 Å) likely due to the lack of stabilization afforded to the nucleobase by elongated π-contacts and the position of the aberrant atoms in a hydrophobic region. Indeed, although the His131···nucleobase distance decreases by ~0.2 Å, the Trp69···nucleobase distance increases by 0.3 Å from RC to TS, which suggests that Trp69 is unable to stabilize the nucleobase in the TS. The barrier for N^2,3-εG oxidation is high (61.5 kJ mol^{-1}) compared to the substrates. Since Fe(IV)-oxo intermediates are highly reactive and the first step of AlkB-mediated repair is expected to occur rapidly, oxidation of N^2,3-εG by AlkB may not be feasible due to the high barrier of this step.

Overall, ONIOM(QM:MM) calculations reveal that subtle changes in the position of the etheno adduct in the AlkB active site can have significant consequences on the first chemical step. For 1,N^6-εA and 3,N^4-εC, relatively low barriers occur because the substrate is situated in close proximity to the Fe(IV)-oxo moiety, and forms tight contacts with His131 and Trp69. The low barriers for 1,N^6-εA and 3,N^4-εC are consistent with the low barrier previously predicted for AlkB (22.2 kJ mol^{-1}).[52] Similar to the strong etheno substrates, 1,N^2-εG falls near the oxo ligand and forms stacking interactions in the active site throughout the reaction. However, a solvent bridge between the oxo ligand and nucleobase lowers the reactivity of the

Fe(IV)-oxo group and thereby increases the barrier. In contrast to the substrates, for N^2,3-εG an interaction between the nucleobase and Trp69 is not present, and the lesion is positioned further from the Fe(IV)-oxo moiety, which leads to the largest barrier (61.5 kJ mol^{-1}). In fact, this barrier is unlikely to be enzymatically feasible, since this initial step is generally accepted to be rapid for αKG- and Fe(II)-dependent enzymes, including AlkB.[52] Thus, our data provides structural and energetic explanations for the observed relative activity of AlkB enzymes toward different etheno lesions.[5, 34] Summed together, our calculations suggest a complex interplay of several factors is critical for successful oxidation, including the placement of the substrate near the Fe(IV)-oxo moiety, maintenance of π-interactions between the nucleobase and aromatic residues, and limited solvation of the oxo ligand.

6.4 Conclusions

This study used a combined classical MD and ONIOM(QM:MM) approach to provide insight into the reported differential activity of AlkB and ALKBH2 toward four etheno adducts. Indeed, this study provides the first structural insight into key active site features of the enzymes bound to an etheno lesion other than 1,N^6-εA. We rationalize experimental data showing that 1,N^6-εA and 3,N^4-εC are strong substrates, 1,N^2-εG is a weak substrate, and N^2,3-εG is a non-substrate, and thereby afford a greater understanding of the repair of this cytotoxic and mutagenic class of DNA damage. We reveal that π-interactions position the substrates (1,N^6-εA, 3,N^4-εC, and 1,N^2-εG) and Fe(IV)-oxo moiety in close proximity, but disrupt active site π-interactions with N^2,3-εG and thereby the position of the lesion is further from the Fe(IV)-oxo group. Furthermore, solvent bridges exist between catalytic residues and N6 of 1,N^6-εA and N4 of 3,N^4-εC, while electrostatic repulsion occurs between the O6 carbonyl of the εG adducts and catalytic Asp or Glu residues. For 1,N^6-εA and 3,N^4-εC, optimum positioning of the lesion with respect to the Fe(IV)-oxo moiety and active site residues permits oxidation. Although 1,N^2-εG binds near the Fe(IV)-oxo group and forms tight π-interactions with aromatic residues, a larger distance between the

lesion and Asp/Glu residues increases solvation of the oxo ligand and thereby mitigates oxidative catalysis. In contrast, disruption of key interactions between N^2,3-εG and active site residues places this lesion in a catalytically unfavorable position, which allows the lesion to escape repair. Together, our findings contribute to the growing body of literature investigating the catalytic strategies used by αKG/Fe(II)-dependent dioxygenases that have diverse biological roles, including epigenetic regulation and post translational modifications.

6.5 References

[1] Knijnenburg, T. A.; Wang, L.; Zimmermann, M. T.; Chambwe, N.; Gao, G. F.; Cherniack, A. D.; Fan, H.; Shen, H.; Way, G. P.; Greene, C. S.; Liu, Y.; Akbani, R.; Feng, B.; Donehower, L. A.; Miller, C.; Shen, Y.; Karimi, M.; Chen, H.; Kim, P.; Jia, P.; Shinbrot, E.; Zhang, S.; Liu, J.; Hu, H.; Bailey, M. H.; Yau, C.; Wolf, D.; Zhao, Z.; Weinstein, J. N.; Li, L.; Ding, L.; Mills, G. B.; Laird, P. W.; Wheeler, D. A.; Shmulevich, I.; The Cancer Genome Atlas Research, N.; Monnat, R. J.; Xiao, Y.; Wang, C., Genomic and Molecular Landscape of DNA Damage Repair Deficiency across the Cancer Genome Atlas. *Cell Reports* **2018,** *23* (1), 239-254.e6.

[2] 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.

[3] Shrivastav, N.; Li, D.; Essigmann, J. M., Chemical Biology of Mutagenesis and DNA Repair: Cellular Responses to DNA Alkylation. *Carcinogenesis* **2010,** *31* (1), 59-70.

[4] 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*, 855.

[5] Chang, S.-c.; Fedeles, B. I.; Wu, J.; Delaney, J. C.; Li, D.; Zhao, L.; Christov, P. P.; Yau, E.; Singh, V.; Jost, M.; Drennan, C. L.; Marnett, L. J.; Rizzo, C. J.; Levine, S. S.; Guengerich, F. P.; Essigmann, J. M., Next-Generation Sequencing Reveals the Biological Significance of the N^2,3-Ethenoguanine Lesion in Vivo. *Nucleic Acids Res.* **2015,** *43* (11), 5489-5500.

[6] Pandya, G. A.; Moriya, M., 1,N^6-Ethenodeoxyadenosine, a DNA Adduct Highly Mutagenic in Mammalian Cells. *Biochemistry* **1996,** *35* (35), 11487-92.

[7] Hanahan, D.; Weinberg, Robert A., Hallmarks of Cancer: The Next Generation. *Cell* **2011,** *144* (5), 646-674.

[8] Tudek, B.; Zdżalik-Bielecka, D.; Tudek, A.; Kosicki, K.; Fabisiewicz, A.; Speina, E., Lipid Peroxidation in Face of DNA Damage, DNA Repair and Other Cellular Processes. *Free Radical Biol. Med.* **2017,** *107*, 77-89.

[9] Speina, E.; Zielińska, M.; Barbin, A.; Gackowski, D.; Kowalewski, J.; Graziewicz, M. A.; Siedlecki, J. A.; Oliński, R.; Tudek, B., Decreased Repair Activities of 1,N^6-Ethenoadenine and 3,N^4-Ethenocytosine in Lung Adenocarcinoma Patients. *Cancer Res.* **2003,** *63* (15), 4351-4357.

[10] Obtułowicz, T.; Winczura, A.; Speina, E.; Swoboda, M.; Janik, J.; Janowska, B.; Cieśla, J. M.; Kowalczyk, P.; Jawien, A.; Gackowski, D.; Banaszkiewicz, Z.; Krasnodebski, I.; Chaber, A.; Olinski, R.; Nair, J.; Bartsch, H.; Douki, T.; Cadet, J.; Tudek, B., Aberrant Repair of Etheno–DNA Adducts in Leukocytes and Colon Tissue of Colon Cancer Patients. *Free Radical Biol. Med.* **2010,** *49* (6), 1064-1071.

[11] Soll, J. M.; Sobol, R. W.; Mosammaparast, N., Regulation of DNA Akylation Damage Repair: Lessons and Therapeutic Opportunities. *Trends Biochem. Sci* **2017,** *42* (3), 206-218.

[12] Lee, C.-Y. I.; Delaney, J. C.; Kartalou, M.; Lingaraju, G. M.; Maor-Shoshani, A.; Essigmann, J. M.; Samson, L. D., Recognition and Processing of a New Repertoire of DNA Substrates by Human 3-Methyladenine DNA Glycosylase (Aag). *Biochemistry* **2009,** *48* (9), 1850-1861.

[13] O'Brien, P. J.; Ellenberger, T., Dissecting the Broad Substrate Specificity of Human 3-Methyladenine-DNA Glycosylase. *J. Biol. Chem.* **2004,** *279* (11), 9750–9757.

[14] Gros, L.; Ishchenko, A. A.; Saparbaev, M., Enzymology of Repair of Etheno-Adducts. *Mutat. Res., Fund. Mol. Mech. Res.* **2003,** *531* (1), 219-229.

[15] Jurado, J.; Maciejewska, A.; Krwawicz, J.; Laval, J.; Saparbaev, M. K., Role of Mismatch-Specific Uracil-DNA Glycosylase in Repair of 3,N4-Ethenocytosine in Vivo. *DNA Repair* **2004,** *3* (12), 1579-1590.

[16] Matijasevic, Z.; Sekiguchi, M.; Ludlum, D. B., Release of N^2,3-Ethenoguanine from Chloroacetaldehyde-Treated DNA by *Escherichia coli* 3-Methyladenine DNA Glycosylase II. *Proc. Natl. Acad. Sci. USA* **1992,** *89* (19), 9331-9334.

[17] Mishina, Y.; Yang, C.-G.; He, C., Direct Repair of the Exocyclic DNA Adduct 1,N^6-Ethenoadenine by the DNA Repair AlkB Proteins. *J. Am. Chem. Soc.* **2005,** *127* (42), 14594-14595.

[18] 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*, 174.

[19] Aas, P. A.; Otterlei, M.; Falnes, P. Ø.; Vågbø, C. B.; Skorpen, F.; Akbari, M.; Sundheim, O.; Bjørås, M.; Slupphaug, G.; Seeberg, E.; Krokan, H. E., Human and Bacterial Oxidative Demethylases Repair Alkylation Damage in Both RNA and DNA. *Nature* **2003,** *421*, 859.

[20] Koivisto, P.; Duncan, T.; Lindahl, T.; Sedgwick, B., Minimal Methylated Substrate and Extended Substrate Range of *Escherichia coli* AlkB Protein, a 1-Methyladenine-DNA Dioxygenase. *J. Biol. Chem.* **2003,** *278* (45), 44348-44354.

[21] Delaney, J. C.; Essigmann, J. M., Mutagenesis, Genotoxicity, and Repair of 1-Methyladenine, 3-Alkylcytosines, 1-Methylguanine, and 3-Methylthymine in AlkB *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **2004,** *101* (39), 14051-14056.

[22] Falnes, P. Ø.; Rognes, T., DNA Repair by Bacterial AlkB Proteins. *Res. Microbiol.* **2003,** *154* (8), 531-538.

[23] 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. USA* **2002,** *99* (26), 16660-16665.

[24] Muller, T. A.; Hausinger, R. P., Chapter 8 AlkB and Its Homologues - DNA Repair and Beyond. In *2-Oxoglutarate-Dependent Oxygenases*, The Royal Society of Chemistry: 2015; pp 246-262.

[25] Hausinger, R. P., Chapter 1 Biochemical Diversity of 2-Oxoglutarate-Dependent Oxygenases. In *2-Oxoglutarate-Dependent Oxygenases*, The Royal Society of Chemistry: 2015; pp 1-58.

[26] Aik, W.; McDonough, M. A.; Thalhammer, A.; Chowdhury, R.; Schofield, C. J., Role of the Jelly-Roll Fold in Substrate Binding by 2-Oxoglutarate Oxygenases. *Curr. Opin. Struct. Biol.* **2012,** *22* (6), 691-700.

[27] Ougland, R.; Lando, D.; Jonson, I.; Dahl, J. A.; Moen, M. N.; Nordstrand, L. M.; Rognes, T.; Lee, J. T.; Klungland, A.; Kouzarides, T.; Larsen, E., AlkBh1 Is a Histone H2a Dioxygenase Involved in Neural Differentiation. *Stem Cells* **2012,** *30* (12), 2672-2682.

[28] 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.

[29] Gommers-Ampt, J. H.; Van Leeuwen, F.; de Beer, A. L. J.; Vliegenthart, J. F. G.; Dizdaroglu, M.; Kowalak, J. A.; Crain, P. F.; Borst, P., B-D-Glucosyl-Hydroxymethyluracil: A Novel Modified Base Present in the DNA of the Parasitic Protozoan *T. Brucei*. *Cell* **1993,** *75* (6), 1129-1136.

[30] Tahiliani, M.; Koh, K. P.; Shen, Y.; Pastor, W. A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L. M.; Liu, D. R.; Aravind, L.; Rao, A., Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by Mll Partner Tet1. *Science* **2009,** *324* (5929), 930-935.

[31] Aravind, L.; Koonin, E. V., The DNA-Repair Protein AlkB, Egl-9, and Leprecan Define New Families of 2-Oxoglutarate- and Iron-Dependent Dioxygenases. *Genome Biol.* **2001,** *2* (3), research0007.1-research0007.8.

[32] Kurowski, M. A.; Bhagwat, A. S.; Papaj, G.; Bujnicki, J. M., Phylogenomic Identification of Five New Human Homologs of the DNA Repair Enzyme AlkB. *BMC Genomics* **2003,** *4* (1), 48.

[33] Sundheim, O.; Vågbø, C. B.; Bjørås, M.; Sousa, M. M. L.; Talstad, V.; Aas, P. A.; Drabløs, F.; Krokan, H. E.; Tainer, J. A.; Slupphaug, G., Human Abh3 Structure and Key Residues for Oxidative Demethylation to Reverse DNA/RNA Damage. *EMBO J.* **2006,** *25* (14), 3389-3397.

[34] Zdżalik, D.; Domańska, A.; Prorok, P.; Kosicki, K.; van den Born, E.; Falnes, P. Ø.; Rizzo, C. J.; Guengerich, F. P.; Tudek, B., Differential Repair of Etheno-DNA Adducts by Bacterial and Human AlkB Proteins. *DNA Repair* **2015,** *30*, 1-10.

[35] Fedeles, B. I.; Singh, V.; Delaney, J. C.; Li, D.; Essigmann, J. M., The AlkB Family of Fe(II)/A-Ketoglutarate Dependent Dioxygenases: Repairing Nucleic Acid Alkylation Damage and Beyond. *J. Biol. Chem.* **2015**.

[36] Fu, D.; Samson, L. D., Direct Repair of 3,N(4)-Ethenocytosine by the Human AlkBh2 Dioxygenase Is Blocked by the Aag/Mpg Glycosylase. *DNA Repair* **2012,** *11* (1), 46-52.

[37] Ringvoll, J.; Moen, M. N.; Nordstrand, L. M.; Meira, L. B.; Pang, B.; Bekkelund, A.; Dedon, P. C.; Bjelland, S.; Samson, L. D.; Falnes, P. Ø.; Klungland, A., AlkB Homologue 2-Mediated Repair of Ethenoadenine Lesions in Mammalian DNA. *Cancer Res.* **2008,** *68* (11), 4142-4149.

[38] Yang, C.-G.; Yi, C.; Duguid, E. M.; Sullivan, C. T.; Jian, X.; Rice, P. A.; He, C., Crystal Structures of DNA/RNA Repair Enzymes AlkB and ABH2 Bound to Dsdna. *Nature* **2008,** *452* (7190), 961-965.

[39] Lau, A. Y.; Scharer, O. D.; Samson, L.; Verdine, G. L.; Ellenberger, E., Crystal Structure of a Human Alkylbase-DNA Repair Enzyme Complexed to DNA: Mechanisms for Nucleotide Flipping and Base Excision. *Cell* **1998,** *95* (2), 249–258.

[40] Parikh, S. S.; Mol, C. D.; Slupphaug, G.; Bharati, S.; Krokan, H. E.; Tainer, J. A., Base Excision Repair Initiation Revealed by Crystal Structures and Binding Kinetics of Human Uracil-DNA Glycosylase with DNA. *EMBO J.* **1998,** *17* (17), 5214–5226.

[41] Lee, S.; Verdine, G. L., Atomic Substitution Reveals the Structural Basis for Substrate Adenine Recognition and Removal by Adenine DNA Glycosylase. *Proc. Natl. Acad. Sci. USA* **2009,** *106* (44), 18497–18502.

[42] Holland, P. J.; Hollis, T., Structural and Mutational Analysis of *Escherichia coli* AlkB Provides Insight into Substrate Specificity and DNA Damage Searching. *PLoS One* **2010,** *5* (1), e8680.

[43] Yi, C.; Jia, G.; Hou, G.; Dai, Q.; Zhang, W.; Zheng, G.; Jian, X.; Yang, C.-G.; Cui, Q.; He, C., Iron-Catalysed Oxidation Intermediates Captured in a DNA Repair Dioxygenase. *Nature* **2010,** *468* (7321), 330-333.

[44] Yu, B.; Hunt, J. F., Enzymological and Structural Studies of the Mechanism of Promiscuous Substrate Recognition by the Oxidative DNA Repair Enzyme AlkB. *Proc. Natl. Acad. Sci. USA* **2009,** *106* (34), 14315-14320.

[45] Yu, B.; Edstrom, W. C.; Benach, J.; Hamuro, Y.; Weber, P. C.; Gibney, B. R.; Hunt, J. F., Crystal Structures of Catalytic Complexes of the Oxidative DNA/RNA Repair Enzyme AlkB. *Nature* **2006,** *439*, 879.

[46] 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*, 671.

[47] Liu, H.; Llano, J.; Gauld, J. W., A DFT Study of Nucleobase Dealkylation by the DNA Repair Enzyme AlkB. *J. Phys. Chem. B* **2009,** *113* (14), 4887-4898.

[48] Fang, D.; Lord, R. L.; Cisneros, G. A., Ab Initio QM/MM Calculations Show an Intersystem Crossing in the Hydrogen Abstraction Step in Dealkylation Catalyzed by AlkB. *J. Phys. Chem. B* **2013,** *117* (21), 6410-6420.

[49] Cisneros, G. A., DFT Study of a Model System for the Dealkylation Step Catalyzed by AlkB. *Interdisciplinary Sciences: Computational Life Sciences* **2010,** *2* (1), 70-77.

[50] Li, D.; Delaney, J. C.; Page, C. M.; Yang, X.; Chen, A. S.; Wong, C.; Drennan, C. L.; Essigmann, J. M., Exocyclic Carbons Adjacent to the N6 of Adenine Are Targets for Oxidation by the *Escherichia coli* Adaptive Response Protein AlkB. *J. Am. Chem. Soc.* **2012,** *134* (21), 8896-8901.

[51] Li, D.; Delaney, J. C.; Page, C. M.; Chen, A. S.; Wong, C.; Drennan, C. L.; Essigmann, J. M., Repair of DNA Alkylation Damage by the *Escherichia coli* Adaptive Response Protein AlkB as Studied by ESI-TOF Mass Spectrometry. *Journal of Nucleic Acids* **2010,** *2010*, 9.

[52] 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.

[53] Aik, W. S.; Chowdhury, R.; Clifton, I. J.; Hopkinson, R. J.; Leissing, T.; McDonough, M. A.; Nowak, R.; Schofield, C. J.; Walport, L. J., Chapter 2 Introduction to Structural Studies on 2-Oxoglutarate-Dependent Oxygenases and Related Enzymes. In *2-Oxoglutarate-Dependent Oxygenases*, The Royal Society of Chemistry: 2015; pp 59-94.

[54] Bleijlevens, B.; Shivarattan, T.; Flashman, E.; Yang, Y.; Simpson, P. J.; Koivisto, P.; Sedgwick, B.; Schofield, C. J.; Matthews, S. J., Dynamic States of the DNA Repair Enzyme AlkB Regulate Product Release. *EMBO reports* **2008,** *9* (9), 872-877.

[55] Bleijlevens, B.; Shivarattan, T.; van den Boom, K. S.; de Haan, A.; van der Zwan, G.; Simpson, P. J.; Matthews, S. J., Changes in Protein Dynamics of the DNA Repair Dioxygenase AlkB Upon Binding of Fe2+ and 2-Oxoglutarate. *Biochemistry* **2012,** *51* (16), 3334-3341.

[56] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Jr., J. A. M.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 16*, Revision B.01; Gaussian, Inc.: Wallingford CT, 2018.

[57] 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-W541.

[58] 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.

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

[60] Li, P.; Merz, K. M., MCPB.Py: A Python Based Metal Center Parameter Builder. *J. Chem. Inf. Model.* **2016,** *56* (4), 599-604.

[61] Seminario, J. M., Calculation of Intramolecular Force Fields from Second-Derivative Tensors. *Int. J. Quantum Chem* **1996,** *60* (7), 1271-1277.

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

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

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

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

[66] Roy, L. E.; Hay, P. J.; Martin, R. L., Revised Basis Sets for the LANL Effective Core Potentials. *J. Chem. Theory Comput.* **2008,** *4* (7), 1029-1031.

[67] Wilson, K. A.; Kellie, J. L.; Wetmore, S. D., DNA-Protein Pi-Interactions in Nature: Abundance, Structure, Composition and Strength of Contacts between Aromatic Amino Acids and DNA Nucleobases or Deoxyribose Sugar. *Nucleic Acids Res.* **2014,** *42* (10), 6726-6741.

[68] 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* (Copyright (C) 2011 U.S. National Library of Medicine.), 16258–16269.

[69] 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.

[70] Lenz, S. A. P.; Wetmore, S. D., QM/MM Study of the Reaction Catalyzed by Alkyladenine DNA Glycosylase: Examination of the Substrate Specificity of a DNA Repair Enzyme. *J. Phys. Chem. B* **2017,** *121* (49), 11096-11108.

Chapter 7: DNA Repair Enzymes ALKBH2, ALKBH3, and AlkB Oxidize 5-Methylcytosine to 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxylcytosine[a,b,c,d]

7.1 Introduction

In mammals, methylation at the 5-position of cytosine (5-methylcytosine, 5mC, Figure 7.1) is the major form of DNA modification and occurs mainly on CpG dinucleotide sites.[1–3] This methylation is achieved and maintained by *S*-adenosylmethionine-dependent methyltransferases.[4] The reverse process termed demethylation of 5mC is first carried out by the ten-eleven translocation proteins (TET1, 2, and 3) through iterative oxidation of 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Figure 7.1). Subsequently, the oxidation products 5fC and 5caC are removed and repaired back to unmethylated cytosine by thymine DNA glycosylase coupled with base excision repair.[1,5–8] The 5mC modification and its three oxidative derivatives play important roles in epigenetic regulations, development and disease: they function in transcriptional regulation, gene silencing and reprogramming, and cell development.[1,5] The misregulations of 5mC and derivatives have been associated with cancer and other diseases.[1,9,10]

[a] *The Journal of the American Chemical Society* reference style is used in this Chapter.
[b] Adapted from 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, *J. Am. Chem. Soc.,* ***2018,*** *submitted [ja-2018-06822v.R1]*.
[c] K.B. and S.L. contributed equally to this work
[d] S.A.P.L. performed molecular dynamics simulations, while experimental work was performed by K.B., Q.T., F.C., R.Q., and M.J. Writing and editing was performed by S.A.P.L., S.D.W., K.B., and D. L.

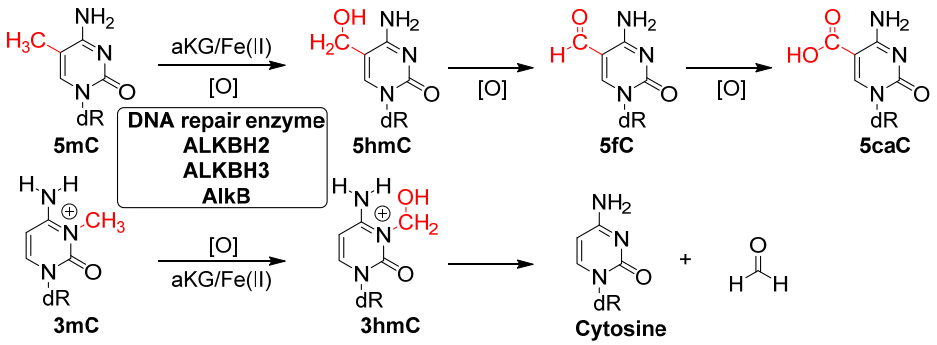


Figure 7.1. Reaction pathway of the AlkB family DNA repair enzymes modifying 5mC and 3mC.

The TET family enzymes belong to the α-ketoglutarate (α-KG)/Fe(II)-dependent dioxygenases. These enzymes have been extensively studied in the last decade for their biological functions, biochemical activities and structural features.[1,5,6,11] In the structures of the TET enzymes, the highly conserved N-terminal β-hairpin-like element for DNA-base recognition and the C-terminal catalytic domain are also found in another family of α-KG/Fe(II)-dependent nucleic acid-modifying dioxygenases, the AlkB family proteins.[5,12,13] In this work, we show that the epigenetic modulator 5mC is *in vitro* modified to 5hmC, 5fC, and 5caC by the DNA repair enzymes in the AlkB family, including human ALKBH2, ALKBH3, and its *E. coli* homolog AlkB (Figure 7.1). The results suggest that these DNA repair enzymes may play a wider role in epigenetic modulation.

Different homologs of the *E. coli* AlkB protein exist in prokaryotic and eukaryotic species; nine homologs exist in human cells (ALKBH1-8 and FTO).[12,14] Among the nine homologs, ALKBH2 and ALKBH3 are DNA repair enzymes that protect the informational integrity of the genome.[5,14–18] They use an α-KG/Fe(II)-dependent mechanism to oxidize aberrant alkyl groups, ultimately restoring the undamaged DNA bases.[14,17,18] The reported substrate scope of AlkB, ALKBH2 and ALKBH3 includes 3-methylcytosine (3mC, Figure 7.1) and 1-methyladenine (1mA), as well as other nitrogen-attached methyl lesions occurring at the Watson-Crick base pairing interface of DNA bases.[15,16,18–20]

Key structural information about the AlkB family enzymes has been obtained from crystal structures of AlkB and ALKBH2 bound to lesion-containing DNA, which reveal that their active sites share several characteristics. Specifically, the AlkB (ALKBH2 in brackets) complex contains a metal center Fe(II) in the wild-type enzymes) coordinated to H131 (H171), H187 (H236), D133 (D173), α-KG, and molecular oxygen.[13,21] Repair of 3mC or 1mA has been proposed to be further enhanced by interactions between D135 of AlkB (E175 of ALKBH2) and the exocyclic amino groups of the lesion.[22] Interestingly, all crystal structures of the AlkB enzymes indicate the lesions are bound in the *anti* glycosidic bond conformation (e.g., for 3mC: χ; $\angle$(O4′C1′N1C2) = 180° ± 90°, Figure F.1a in Appendix F).

Although both 5mC and 3mC carry a methyl modification, the methyl groups are on the opposite sides of the pyrimidine ring. It is reasonable to predict that members of the AlkB enzyme family may be able to oxidize 5mC if the methyl group can be positioned near the catalytic center. The experimental results presented here reveal that *in vitro* AlkB enzymes not only can repair DNA lesions, such as 3mC, but also can modify the epigenetic biomarker 5mC and generate its oxidative derivatives. Our theoretical calculations suggest that AlkB enzymes bind 5mC in the *syn* glycosidic conformation (χ = 0° ± 90°, Figure F.1b in Appendix F) to align the 5-methyl moiety for oxidation, which is similar to how the TET family enzymes bind 5mC. This paper is the first work to demonstrate the ability of the AlkB family enzymes to oxidize a methyl group that is attached to carbon, instead of nitrogen, on a DNA base.

7.2 Experimental and Computational Details

See Sections F.1 and F.2 in Appendix F for details on the experimental and computational procedures.

7.3 Results and Discussion

A 16mer DNA oligonucleotide (5'-GAAGACCTXGGCGTCC-3', X = 5mC) containing 5mC in a CpG dinucleotide context was prepared through solid phase DNA synthesis with the phosphoramidite of 5mC.[23–26] High resolution electrospray ionization time-of-flight (ESI-TOF) MS analysis of the oligonucleotide exhibited an m/z of 1625.281 at its -3 charge state, which is in good agreement with the theoretical m/z 1625.281 expected of the product oligonucleotide (Figure 7.2a and Table F.1 in Appendix F). In the same -3 charge envelope, we also observed the ions of 5mC + Na^+ (1632.608) and 5mC + K^+ (1637.928). The observed m/z values of these species are consistent with the corresponding calculated m/z values (Table F.1 in Appendix F).

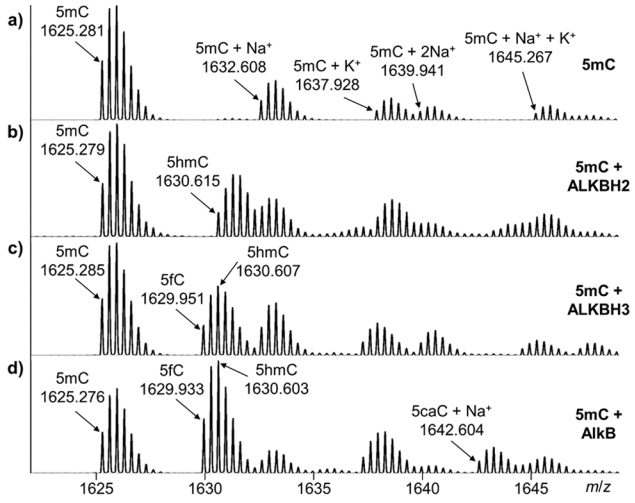


Figure 7.2. High resolution ESI-TOF MS analyses of 16mer DNA oligonucleotides containing 5mC and oxidized products. The observed m/z values represent the oligonucleotides under their -3 charge state. a) 5mC; b) 5mC + ALKBH2; c) 5mC + ALKBH3; and d) 5mC + AlkB.

Previously, our lab has purified the three members of the AlkB enzyme family mentioned above and tested their repair efficiency for 3mC, 1mA and other substrates in both ss- and ds-DNA.[23,24] Similar

procedures were adopted for the modification reactions of 5mC. The reaction conditions and the oligonucleotide-enzyme ratios are similar to those observed in the conversion of 5mC to its oxidative derivatives by the TET proteins reported in the literature.[6–8] For each enzyme, experiments were conducted in triplicate both in the presence and absence of the enzyme with all necessary cofactors at 37°C under both ds- and ss-DNA conditions, and the reaction products were analyzed by high resolution MS to ensure the differentiation of reaction products that have very similar m/z values.[24]

First, we tried to identify the new oxidative products that appeared after the enzymatic reactions. The three oxidative products, 5hmC, 5fC, and 5caC, all formed in the reactions with all three enzymes, but every enzyme had a preference to generate a certain oxidative derivative. Below we use typical examples to demonstrate the formation of a certain product. The MS results of ALKBH2 oxidizing 5mC in ds-DNA (Figure 7.2b) showed a new oligonucleotide species that has an m/z for the monoisotopic peak at 1630.615 at -3 charge state, which corresponds well to the theoretical m/z value of the 16mer oligonucleotide containing 5hmC (1630.612 calculated, Table F.1 in Appendix F). In the reaction of 5mC with ALKBH3 in ss-DNA, another oligonucleotide envelope appeared at m/z 1629.951 (Figure 7.2c), which agrees with the 5fC base in the 16mer oligonucleotide (1629.940 calculated). For the oxidation of 5mC by the AlkB protein in ss-DNA, we observed the peak envelopes of 5hmC (1630.603), 5fC (1629.933) together with a new species with an m/z value of 1642.604, which is close to the 16mer oligonucleotide containing the sodium salt of 5caC (1642.600 calculated, Figure 7.2d and Table F.1 in Appendix F). Other oligonucleotide species containing metal ions, such as Na^+ and K^+, were also observed. The complete assignments of the major species generated from MS analyses are summarized in Figure F.2 and Table F.1 (see Section F.3 in Appendix F).

The product oligonucleotides were digested into single nucleosides, analyzed by LC-MS, and compared with standard nucleosides to confirm the oxidative products generated from enzymatic reactions are indeed 5hmC, 5fC, and 5caC (see the Product oligonucleotides analyses section and Figures F.3 to F.11 in

Appendix F). Also, to make sure the oxidations were carried out by AlkB and its homologs, we generated the catalytically inactive protein variants of AlkB: H131A, D133A, and H187A (Figure F.12 and Table F.2 in Appendix F). The three substituted amino acids in AlkB are the key residues that coordinate the Fe(II) ion.[13,27] The sequences of wild type and variant proteins were confirmed by trypsin digestion with MS analyses (Figure F.12 in Appendix F). None of the AlkB variants showed any detectable oxidative product when reacted with 5mC (Figure F.13 in Appendix F); these observations suggest that the oxidative modifications were carried out by AlkB and not by a contaminating enzyme.

For the oxidation of 5mC, the formation of products 5hmC, 5fC, and 5caC had different distribution patterns for the three enzymes; and the three enzymes had different preferences for oxidation in ss- or ds-DNA reactions (Figure 7.3 and the Product distribution for the oxidation of 5mC section and Tables F.3-F.4 in Appendix F). The overall reactivities of the three AlkB proteins on oxidizing 5mC are similar to the proficiencies of the TET enzymes on modifying 5mC reported in the literature[6–8] and are consistent with theoretical calculations that report higher barriers for each successive oxidation step by TET2.[28] In all of the enzymatic reactions, we only observed the oxidation of 5mC, but not the thymine DNA base, which naturally has a 5-methyl group. The same finding was reported for the TET family enzymes.[8]

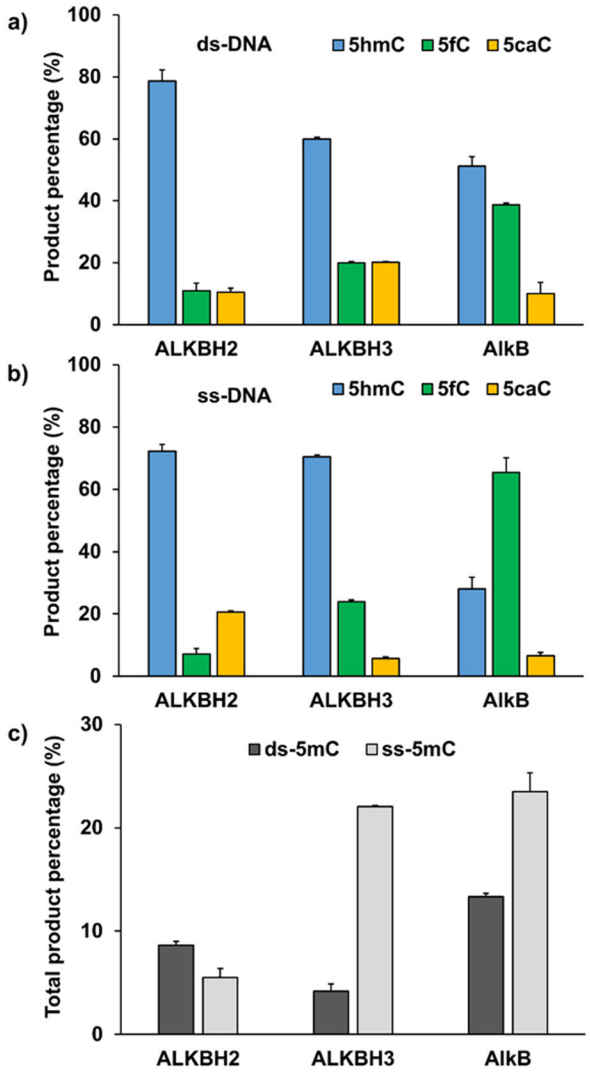


Figure 7.3. Product distribution and strand preference for the oxidation of 5mC by the AlkB family enzymes. Reaction products (5hmC, 5fC and 5caC) generated from the reactions of 5mC with the AlkB family enzymes in a) ds-DNA and b) ss-DNA. c) Total product percentage from reactions of the AlkB family enzymes oxidizing 5mC in ds- and ss-DNA.

To probe the molecular basis by which the AlkB enzymes are able to oxidize 5mC, we performed molecular dynamics (MD) simulations to examine how 5mC is accommodated in the active sites of ALKBH2 and AlkB (Figure 7.4). We found that the 5-methyl of *anti*-5mC in our model is far from the Fe(IV)–oxo moiety (~5.3 Å for ALKBH2 in Figure 7.4a and ~7.9 Å for AlkB in Figure 7.4d; also Figures F.14-F.15, and Tables F.5-F.8 in Appendix F). In contrast, the distances between the Fe(IV)–oxo moiety

and the 5-methyl group in the models with *syn*-5mC bound to ALKBH2 or AlkB are much shorter (~3.8 Å for ALKBH2 in Figure 7.4b and ~3.6 Å for AlkB in Figure 7.4e) and similar to the distances in the structures of ALKBH2 or AlkB bound to their prototypic substrate 3mC in the *anti*-conformation (~3.3 Å for ALKBH2 in Figure 7.4c and ~3.3 Å for AlkB in Figure 7.4f.; See also the Simulations of ALKBH2 and AlkB bound to 3mC, 5hmC, and 5fC section and Figures F.16-F.17 in Appendix F). With *syn*-5mC bound to ALKBH2 and AlkB, hydrogen bonds appear possible between the N^4 amino group of 5mC and active site residues. Specifically, *syn*-5mC interacts with Asp and Glu residues (D174/E175 for ALKBH2, and D135/E136 for AlkB) through water (Tables F.5-F.6 and Figures F.13-F.14 in Appendix F), as well as the Y122 hydroxy group in ALKBH2 (Table F.5 in Appendix F). These interactions likely facilitate oxidative catalysis by positioning the C5 methyl group near the Fe(IV)–oxo moiety (~3.6 – 3.8 Å; Tables F.7-F.8 in Appendix F). Interestingly, a crystal structure of TET2 co-crystallized with 5mC-containing DNA reveals *syn*-5mC in the active site,[29] and the reported χ torsion angle is consistent with that predicted for *syn*-5mC in ALKBH2/AlkB (Figure F.18 and Tables F.7-F.8 in Appendix F). Based on our combined experimental and theoretical data, we propose that the AlkB family enzymes are able to oxidize 5mC bound only in the *syn*-conformation.

To provide insight into the variable activity of ALKBH2 and AlkB for subsequent nucleobase oxidation (Figure 7.3), MD simulations were performed on *syn*-5hmC and 5fC bound in the active sites. For 5hmC, the C5 substituent is further from the Fe(IV)–oxo moiety for ALKBH2 (~4.5 Å; Tables F.7-F.8 in Appendix F) compared to AlkB (3.4 Å), which is consistent with the relative low abundance of the 5fC product for ALKBH2. For 5fC, increased flexibility of the bound nucleobase may permit enhanced catalysis for ALKB2 compared to AlkB. In addition to the insight provided from MD (see Section F.5 in Appendix F for detailed discussions of these simulations and Figures F.19-F.22), several other factors could influence the product distribution including DNA binding and unique base flipping mechanisms used by each enzyme.

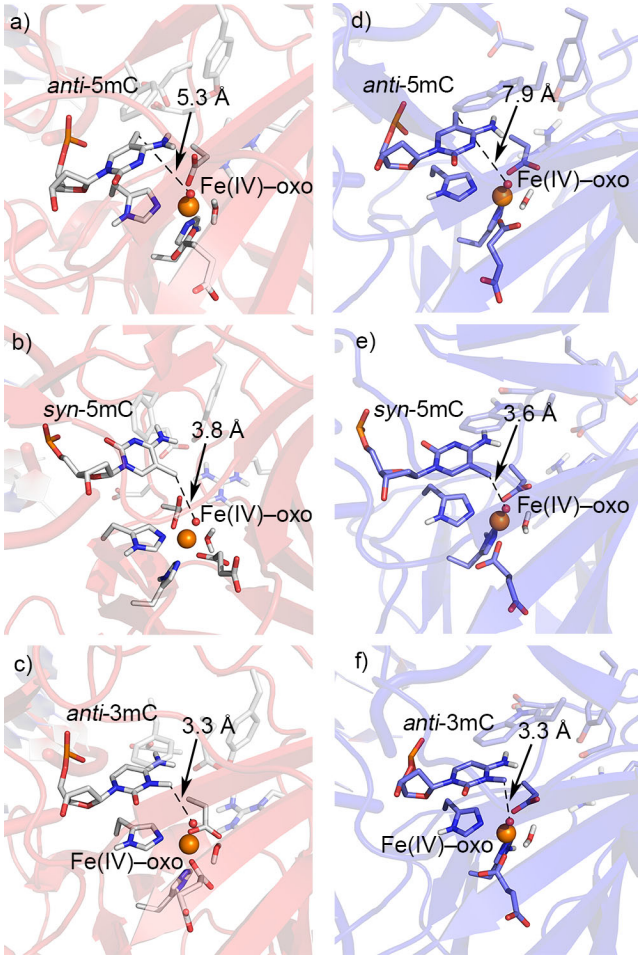


Figure 7.4. Representative molecular dynamics structures of the ALKBH2 (a-c) or AlkB (d-f) complex bound to *anti*-5mC (a,d), *syn*-5mC (b,e), or *anti*-3mC (c,f). The distance between the oxo-moiety and methyl groups is highlighted with dashed lines.

7.4 Conclusions

In this paper, we demonstrated the *in vitro* oxidative modification of 5mC to 5hmC, 5fC, and 5caC by the three AlkB DNA repair enzymes. Thus, AlkB proteins are not only able to repair DNA adducts, such as 3mC, but also can edit the epigenetic modification 5mC and generate the corresponding oxidative derivatives. These observations suggest a possible connection between DNA repair and epigenetic gene modulation. Future investigation includes analyzing the kinetic parameters of the three AlkB enzymes acting on 5mC, confirming the oxidation of 5mC by AlkB enzymes in cells, and probing whether other α-KG/Fe(II)-dependent dioxygenases can oxidize 5mC.

7.5 References

[1] Wu, X.; Zhang, Y. TET-Mediated Active DNA Demethylation: Mechanism, Function and beyond. *Nat. Rev. Genet.* **2017**, *18* (9), 517–534.

[2] Smith, Z. D.; Meissner, A. DNA Methylation: Roles in Mammalian Development. *Nat. Rev. Genet.* **2013**, *14* (3), 204–220.

[3] Li, E.; Zhang, Y. DNA Methylation in Mammals. *Cold Spring Harb. Perspect. Biol.* **2014**, *6* (5), a019133.

[4] Law, J. A.; Jacobsen, S. E. Establishing, Maintaining and Modifying DNA Methylation Patterns in Plants and Animals. *Nat. Rev. Genet.* **2010**, *11* (3), 204–220.

[5] *2-Oxoglutarate-Dependent Oxygenases*; Hausinger, R. P., Schofield, C. J., Eds.; RSC metallobiology; Royal Society of Chemistry: Cambridge, UK, 2015.

[6] Tahiliani, M.; Koh, K. P.; Shen, Y.; Pastor, W. A.; Bandukwala, H.; Brudno, Y.; Agarwal, S.; Iyer, L. M.; Liu, D. R.; Aravind, L.; et al. Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science* **2009**, *324* (5929), 930–935.

[7] Ito, S.; D'Alessio, A. C.; Taranova, O. V.; Hong, K.; Sowers, L. C.; Zhang, Y. Role of Tet Proteins in 5mC to 5hmC Conversion, ES-Cell Self-Renewal and Inner Cell Mass Specification. *Nature* **2010**, *466* (7310), 1129–1133.

[8] Zhang, L.; Chen, W.; Iyer, L. M.; Hu, J.; Wang, G.; Fu, Y.; Yu, M.; Dai, Q.; Aravind, L.; He, C. A TET Homologue Protein from Coprinopsis Cinerea (CcTET) That Biochemically Converts 5-Methylcytosine to 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxylcytosine. *J. Am. Chem. Soc.* **2014**, *136* (13), 4801–4804.

[9] Klungland, A.; Robertson, A. B. Oxidized C5-Methyl Cytosine Bases in DNA: 5-Hydroxymethylcytosine; 5-Formylcytosine; and 5-Carboxycytosine. *Free Radic. Biol. Med.* **2017**, *107*, 62–68.

[10] Cimmino, L.; Aifantis, I. Alternative Roles for Oxidized mCs and TETs. *Curr. Opin. Genet. Dev.* **2017**, *42*, 1–7.

[11] Hashimoto, H.; Zhang, X.; Vertino, P. M.; Cheng, X. The Mechanisms of Generation, Recognition, and Erasure of DNA 5-Methylcytosine and Thymine Oxidations. *J. Biol. Chem.* **2015**, *290* (34), 20723–20733.

[12] Aravind, L.; Koonin, E. V. The DNA-Repair Protein AlkB, EGL-9, and Leprecan Define New Families of 2-Oxoglutarate- and Iron-Dependent Dioxygenases. *Genome Biol.* **2001**, *2* (3), RESEARCH0007.

[13] Yu, B.; Edstrom, W. C.; Benach, J.; Hamuro, Y.; Weber, P. C.; Gibney, B. R.; Hunt, J. F. Crystal Structures of Catalytic Complexes of the Oxidative DNA/RNA Repair Enzyme AlkB. *Nature* **2006**, *439* (7078), 879–884.

[14] Sedgwick, B.; Bates, P. A.; Paik, J.; Jacobs, S. C.; Lindahl, T. Repair of Alkylated DNA: Recent Advances. *DNA Repair* **2007**, *6* (4), 429–442.

[15] 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.

[16] Falnes, P. Ø.; Johansen, R. F.; Seeberg, E. AlkB-Mediated Oxidative Demethylation Reverses DNA Damage in Escherichia Coli. *Nature* **2002**, *419* (6903), 178–182.

[17] Zheng, G.; Fu, Y.; He, C. Nucleic Acid Oxidation in DNA Damage Repair and Epigenetics. *Chem. Rev.* **2014**, *114* (8), 4602–4620.

[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.

[19] Delaney, J. C.; Essigmann, J. M. Mutagenesis, Genotoxicity, and Repair of 1-Methyladenine, 3-Alkylcytosines, 1-Methylguanine, and 3-Methylthymine in alkB Escherichia Coli. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (39), 14051–14056.

[20] Li, D.; Fedeles, B. I.; Shrivastav, N.; Delaney, J. C.; Yang, X.; Wong, C.; Drennan, C. L.; Essigmann, J. M. Removal of N-Alkyl Modifications from N(2)-Alkylguanine and N(4)-Alkylcytosine in DNA by the Adaptive Response Protein AlkB. *Chem. Res. Toxicol.* **2013**, *26* (8), 1182–1187.

[21] 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.

[22] Yi, C.; Jia, G.; Hou, G.; Dai, Q.; Zhang, W.; Zheng, G.; Jian, X.; Yang, C.-G.; Cui, Q.; He, C. Iron-Catalysed Oxidation Intermediates Captured in a DNA Repair Dioxygenase. *Nature* **2010**, *468* (7321), 330–333.

[23] Chen, F.; Tang, Q.; Bian, K.; Humulock, Z. T.; Yang, X.; Jost, M.; Drennan, C. L.; Essigmann, J. M.; Li, D. Adaptive Response Enzyme AlkB Preferentially Repairs 1-Methylguanine and 3-Methylthymine Adducts in Double-Stranded DNA. *Chem. Res. Toxicol.* **2016**, *29* (4), 687–693.

[24] Chen, F.; Bian, K.; Tang, Q.; Fedeles, B. I.; Singh, V.; Humulock, Z. T.; Essigmann, J. M.; Li, D. Oncometabolites D- and L-2-Hydroxyglutarate Inhibit the AlkB Family DNA Repair Enzymes under Physiological Conditions. *Chem. Res. Toxicol.* **2017**, *30* (4), 1102–1110.

[25] Bian, K.; Chen, F.; Humulock, Z. T.; Tang, Q.; Li, D. Copper Inhibits the AlkB Family DNA Repair Enzymes under Wilson's Disease Condition. *Chem. Res. Toxicol.* **2017**, *30* (10), 1794–1796.

[26] Tang, Q.; Cai, A.; Bian, K.; Chen, F.; Delaney, J. C.; Adusumalli, S.; Bach, A. C.; Akhlaghi, F.; Cho, B. P.; Li, D. Characterization of Byproducts from Chemical Syntheses of Oligonucleotides Containing 1-Methyladenine and 3-Methylcytosine. *ACS Omega* **2017**, *2* (11), 8205–8212.

[27] Yu, B.; Hunt, J. F. Enzymological and Structural Studies of the Mechanism of Promiscuous Substrate Recognition by the Oxidative DNA Repair Enzyme AlkB. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (34), 14315–14320.

[28] Lu, J., Hu, L., Cheng, J., Fang, D., Wang, C., Yu, K., Jiang, H., Cui, Q., Xu, Y., and Luo, C. A computational investigation on the substrate preference of ten-eleven-translocation 2 (TET2), *Phys. Chem. Chem. Phys.* **2016** *18*, 4728-4738.

[29] 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.

Chapter 8: Conclusions and Future Directions[a]

8.1 Summary

This thesis used computational techniques to investigate enzymes that process a diverse array of substrates. Broadly, this work can be divided into two major topics: 1) purine or pyrimidine salvage catalyzed by nucleoside hydrolases; and 2) the repair of alkylation damaged catalyzed by either human alkyladenine DNA glycosylase, *E. coli* AlkB, or human ALKBH2. This chapter summarizes the major contributions of each chapter and details future avenues for investigations on each topic.

8.2 Nucleoside Hydrolases

8.2.1 Contributions from Thesis

Chapter 4 used DFT calculations to provide a fundamental understanding of the hydrolytic deglycosylation of the natural RNA nucleosides (A, C, G, and U), offered a comparison to DNA hydrolysis, and examined the effects of acid, base, or simultaneous acid–base catalysis on RNA deglycosylation. Through comparisons to previous work focused on DNA deglycosylation,[1-9] the barriers for RNA hydrolysis were determined to be 30–38 kJ mol^{-1} higher than the corresponding DNA barriers, which is supported by experimental observations that the 2′–OH group stabilizes the glycosidic bond.[10] Both acid and base catalysis individually afford significant reductions to the reaction barriers. Nevertheless, simultaneous acid and base catalysis is the most beneficial way to enhance the reactivity of the glycosidic bonds in RNA. Interestingly, the sites on the nucleobase that resulted in the greatest barrier reductions match those typically protonated by nucleoside hydrolases during catalyzed deglycosylation. Thus, Chapter 4 provides a greater understanding of the instrinsic reactivity of the glycosidic bond in RNA nucleosides, and has implications for the function of RNA-targeting enzymes, including nucleoside hydrolases.

[a]*The Journal of the American Chemical Society reference style is used in this Chapter.*

MD simulations were performed in Chapter 5 to provide missing structural information and key insights into the function of wild-type and mutant CU-NH. Systems that differ in the identity and protonation states of active site catalytic residues were examined to identify key enzyme–substrate interactions that dictate the substrate specificity. Regardless of the wild-type or mutant CU-NH considered, the MD data suggests that inosine binding is facilitated by interactions of the ribose moiety with active site residues and Ca^{2+}, and π-interactions between two His residues (His82 and His239) and the nucleobase. The lack of activity toward inosine exhibited by wild-type CU-NH[11] is explained by no residue being correctly aligned to activate the departing nucleobase, which the small model calculations in Chapter 4 revealed was critical for rapid deglycosylation. In contrast, a hydrogen-bonding network between a newly identified general acid (Asp15) and N7 of the inosine nucleobase is present when the two Tyr mutations are engineered into the active site. Interestingly, the data from Chapter 4 suggests that protonation at N7 of purines has the greatest effect on the barrier compared to other sites of activation. Investigation of the single CU-NH mutants revealed that this hydrogen-bonding network is only maintained when both Tyr mutations are present due to a π-interaction between the residues. These results rationalize previous experiments that show the single Tyr mutants are unable to efficiently hydrolyze inosine[11] and explain how the Tyr residues work synergistically in the double mutant to stabilize the nucleobase leaving group during hydrolysis. Since crystal structures of other nucleoside hydrolases reveal a similar arrangement and identity of active site residues,[12-14] the work in Chapter 5 suggests that new catalytic residues identified by MD simulations may play a conserved role across other nucleoside hydrolase classes.

8.2.2 Future Directions

Chapter 4 characterized the catalytic impact of nucleobase protonation and nucleophile activation on the deglycosylation of the canonical RNA nucleosides. However, additional DFT calculations could be performed to quantify the impact of these catalytic strategies on deglycosylation of inosine and xanthosine, which are common substrates of nucleoside hydrolases.[15] Moreover, since π–π interactions are abundant within nucleoside hydrolase active sites,[11-14] deglycosylation could be modelled in the presence of aromatic amino acid sidechains (benzene, imidazole, phenol, and indole) to examine the catalytic impact of these residues. Indeed, a similar approach has been previously applied to a DNA lesion (2′-deoxyuridine), which substantially lowered the hydrolysis barrier.[16]

This thesis identified that a hydrogen-bonding network between a potential general acid and the substrate is present for mutant CU-NH, but not wild-type CU-NH. To quantify the impact of the hydrogen-bonding network, QM/MM calculations could characterize wild-type and mutant CU-NH-catalyzed hydrolysis. MD simulations could also be performed on IU-NH bound to inosine to establish whether a similar hydrogen-bonding network is present in the active site of that enzyme. Furthermore, other purine nucleosides (xanthosine, guanosine, and adenosine) can be placed within the active site of IU-NH and mutant CU-NH to determine whether the hydrogen-bonding network is present for all purines that are targeted by nucleoside hydrolases. Combined with the data provided by Chapters 4 and 5, this work would yield important insights that could aid development of new treatments for diseases caused by trypanosomatids.

8.3 Repair of DNA Damage

8.3.1 Contributions from Thesis

This thesis provides the first structural explanation for the ability of AAG to catalyze deglycosylation on a diverse set of DNA lesions while discriminating against the natural nucleobases.[17-18]

Specifically, Chapter 2 used MD simulations on seven different neutral (Hx and 1,N^6-εA) or charged (3MeA and 7MeG) substrates, inhibitors (3,N^4-εC), or canonical purines (A or G) to probe how the bound nucleotide affects the conformation of the AAG active site and the roles of active site residues in dictating substrate selectivity. The neutral substrates form a common DNA–protein hydrogen bond, which results in a consistent active site conformation that maximizes π–π interactions between the aromatic residues and the nucleobase required for catalysis. Nevertheless, subtle differences in DNA–enzyme contacts for different neutral substrates explain observed differential catalytic efficiencies.[17-18] This chapter also provides a novel proposal for AAG inhibition by damaged pyrimidines, which is due to the lack of a suitable water nucleophile in the active site, and contrasts previous proposals that inhibition results from lack of nucleobase[19] or nucleophile activation.[20] For the natural purines, the exocyclic amino groups clash with active site residues leading to redistribution of key active site residues, including water, and non-excision. Specifically, water resides between the A nucleobase and the active site aromatic amino acids required for catalysis, while a shift in the position of the general base (Glu125) repositions (potentially nucleophilic) water away from G. Despite sharing common amino groups to natural DNA, the methyl substituents in cationic purine lesions (3MeA and 7MeG) exhibit repulsion with active site residues leading to repositioning of the damaged bases in the active site in a manner that promotes excision. The results from this chapter highlight the complex interplay of many different DNA–protein interactions used by AAG to facilitate BER, as well as the crucial role of the general base and water (nucleophile) positioning.

Chapter 3 used an ONIOM(QM:MM) methodology to provide atomic level details for AAG-catalyzed deglycosylation of Hx, G, or 7MeG. This chapter revealed that the changes to protein–DNA contacts upon binding different substrates identified in Chapter 2 can significantly affect the deglycosylation reaction. Specifically, AAG excises Hx in a concerted mechanism that is facilitated through correct alignment of the (Glu125) general base due to hydrogen bonding with a neighboring aromatic

amino acid (Tyr127). Hx departure is further stabilized by π–π interactions with aromatic amino acids and hydrogen bonds with active site water. Despite a similar structure to Hx, QM/MM calculations confirm that G is not excised due to disruption of the position of the general base, weaker active site π–interactions, and the lack of nucleobase solvation. In contrast, since cationic 7MeG does not occupy the same position within the AAG active site as G due to steric clashes with the additional N7 methyl group, the general base is correctly aligned to permit nucleobase excision as observed for neutral Hx. Thus, combining the results from Chapter 2 detailing the substrate binding mechanism of AAG with previous data on the activity of AAG towards other substrates (εA and 3MeA) and natural A,[20] these chapters provide a clear mechanism by which this critical repair enzyme can remove a structurally diverse set of oxidized and alkylated DNA purines, while discriminating against natural purines.

Since the computational approach used in Chapters 2 and 3 was successful in explaining the substrate preferences of AAG, MD simulations are used in Chapter 6 to provide structural details of the active sites of AlkB or ALKBH2 when bound to etheno lesions, while ONIOM(QM:MM) calculations probes the energetic consequences of the lesion-dependent active site conformation. Specifically, MD simulations reveal that π-interactions are critical to position the nucleobase in a catalytically-active conformation, and these interactions are disrupted only when the non-substrate N^2,3-εG is bound by either AlkB or ALKBH2. Furthermore, the etheno adduct substrates are in close proximity to the Fe(IV)-oxo species, thus facilitating oxidation, while the aberrant carbon atoms of N^2,3-εG are far from the Fe(IV)-oxo group, which explains the non-oxidation of this lesion.[21] For both AlkB and ALKBH2, water, and Asp and Glu residues that are crucial for oxidative repair are optimally positioned when a strong substrate (1,N^6-εA or 3,N^4-εC) is bound, but disrupted when a weak substrate (1,N^2-εG) or non-substrate (N^2,3-εG) is bound. Consequently, ONIOM(QM:MM) calculations reveal relatively small barriers for the first step of the AlkB-mediated oxidative repair of 1,N^6-εA and 3,N^4-εC, a higher barrier for 1,N^2-εG repair, and an

enzymatically unfeasible barrier for N^2,3-εG repair, which correlates with sequencing and kinetic data for repair catalyzed by AlkB enzymes.[21-24]

Chapter 7 provides evidence that expands the biological role of the AlkB DNA repair enzymes. Specifically, *in vitro* mass spectrometry experiments demonstrated that AlkB enzymes oxidize 5mC to 5hmC, 5fC, and 5caC, while MD simulations provide structural data that indicates that AlkB enzymes stabilize the *syn* glycosidic bond conformation through direct and water-mediated hydrogen bonds between the nucleobase and active site residues. This conformation places the methyl group in a position comparable to that of 3MeC, a prototypic substrate of the AlkB proteins.[25] More importantly, this position is near the Fe(IV)-oxo group, which the data in Chapter 6 suggests is critical for successful oxidation. Chapter 7 also revealed that the *syn* conformation of 5hmC and 5fC nucleotides are also stabilized in the AlkB and ALKBH2 active sites and provided an atomistic explanation for the relative product distributions of each enzyme.

8.3.2 Future Directions

While this thesis provided insight into the enzyme active site conformation upon binding damaged and undamaged DNA, the substrate preferences of AAG, AlkB, and ALKBH2 are also influenced by the base-flipping step where the enzyme rotates the initially intrahelical nucleotide by 180° and stabilizes the extrahelical conformation in the active site. This step is not trivial since DNA repair enzymes have the challenge of locating damaged nucleotides among a vast genome of unmodified DNA.[25-26] To accomplish this task, the enzyme may probe the stability of the lesion base pair. For example, AAG exhibits enhanced catalysis toward Hx paired opposite T compared to Hx paired opposite C,[18] which has been proposed to be due to the reduced stability of Hx:T pairs relative to Hx:C pairs.[27] Unfortunately, for many DNA repair enzymes, the base flipping mechanism and the role that the lesion-pair stability plays in this mechanism is not well understood. To shed light on this understudied phenomenon, computational studies can be used

to characterize the stability of the lesion pair and the base-flipping pathway. Specifically, the stability of the lesion pair could be quantified through DFT calculations to determine if the stability could be correlated with the barrier of the base-flipping step. The base-flipping step can be characterized using classical MD and advanced sampling (e.g., umbrella sampling) simulations. Indeed, these methods have been utilized to examine the base-flipping step catalyzed by DNA glycosylases hOgg1 and FPG.[28-29] These studies identified several intermediate structures during the base-flipping step, which is conserved for both enzymes, and allows discrimination between 8-oxoguanine and undamaged G. For AAG, a similar approach can be taken to determine how the activity of the enzyme is influenced by base flipping, and whether AAG discriminates against the natural purines during this step.

Examination of base-flipping would also have important implications for AlkB and ALKBH2. While this thesis established that both AlkB and ALKBH2 can stabilize *syn* 5MeC and the oxidized derivatives in the active site, further study is required to answer several questions. Since 5MeC is in the *anti* conformation paired opposite G,[30] characterization of the base-flipping step using MD simulations and advanced sampling methods could clarify when the nucleotide flips *syn* during substrate binding. Moreover, examination of the base-flipping of T could explain how AlkB enzymes can facilitate oxidation of 5MeC, but do not erroneously oxidize undamaged T to the 5-hydroxymethyluracil lesion. This information is critical to understanding the epigenetic regulation role that AlkB enzymes play within cells.

This thesis considered how the catalytic mechanism of AAG is affected by changes to the active site conformation that occur during binding of substrates and non-substrates. For AAG, when the studies of Hx, G, and 7MeG in this thesis are combined with previous work that examined repair of 1,N^6-εA, 3MeA, and A,[20] a near comprehensive list of substrates have been considered. Nevertheless, further work needs to consider other nucleotides in the active site, including 1,N^2-εG and N^2,3-εG. Specifically, while 1,N^2-εG is a substrate of AAG, N^2,3-εG is not a substrate,[31] and poor repair of N^2,3-εG has been proposed to be a cause

of vinyl chloride-induced carcinogenesis.[21] Thus, additional MD simulations can consider the ethenoguanine lesions in the active site. Based on the results from simulations of A and G, and the location of the etheno atoms, N^2,3-εG may displace the general base (Glu125) leading to non-excision. Moreover, since no human glycosylase has been shown to repair N^2,3-εG, the ethenoguanine adducts should be considered in the active site of *E. coli* DNA glycosylase AlkA since this repair enzyme may excise both lesions,[32] and comparison to AAG may clarify how different DNA glycosylases with overlapping substrate preferences process damaged DNA.

While Chapter 6 revealed the importance of how the etheno adducts are positioned in the active site of AlkB and ALKBH2, further work is required to characterize the full AlkB- or ALKBH2-mediated oxidative dealkylation pathways. By using an ONIOM(QM:MM) approach, the effect of the position of water and of Asp and Glu residues on the catalyzed reaction could be further clarified. Moreover, this approach would yield a comparison between AlkB- and ALKBH2-catalyzed oxidation, which would enhance understanding of the different strategies that these enzymes use to repair alkylated DNA. Additional MD simulations and QM/MM calculations could be performed on AlkB or ALKBH2 bound to larger adducts, including 1,N^6-ethanoadenine,[33] and acrolein- or malondialdehyde-derived adducts.[34] This work would provide additional insight into how AlkB enzymes can bind and repair bulky adducts.

Chapter 7 considered how AlkB and ALKBH2 bind to a prototypic substrate, namely 3MeC. However, further details are required to understand how AlkB enzymes recognize and process methylated DNA. Specifically, AlkB proteins are more active toward A or C methyl lesions compared to T or G methyl lesions, although the structural basis for this difference is not understood. To enhance understanding of how AlkB enzymes process methyl lesions, MD simulations can be initiated on AlkB or ALKBH2 to determine how the methyl group is oriented with respect to the Fe(IV)-oxo group, and probe additional interactions between the substrate and enzyme. While this thesis makes important contributions to

understanding the substrate specificity and biological roles of AlkB enzymes, further study is required to fully clarify the important roles played by these enzymes in biology.

8.4 Final Remarks

This thesis focused on enzymes with broad substrate specificity that functioned in either nucleoside salvage or DNA repair. The findings of each chapter suggest that these enzymes can process structurally diverse substrates by utilizing active sites lined with aromatic amino acids to facilitate substrate binding. Moreover, water can play critical roles in the catalytic mechanism of the enzymes examined, either by mediating proton transfer reactions or stabilizing charged species. The future directions proposed in this chapter can enhance the understanding of enzymes that play critical roles in biological pathways. The insights gained from the computational approach used in this thesis can be applied to other enzymes that use flexible active sites to exhibit diverse substrate specificity. Specifically, a similar methodology can be used to study other enzymes that employ aromatic amino acids to facilitate substrate binding, including cytochrome P450 enzymes, which metabolize a diverse array of potentially toxic compounds,[35-37] and matrix metalloproteinases, which degrade a wide variety of extracellular matrix proteins and regulate cell behavior.[38-39] The overall theme of this thesis is that, when bound by enzymes, subtle differences in the structure of ligands can alter the position and orientation of catalytic residues and water. By using computational chemistry to probe these changes, the atomistic explanations for the differential activity of an enzyme can be obtained.

8.5 References

[1] Millen, A. L.; Archibald, L. A. B.; Hunter, K. C.; Wetmore, S. D., A Kinetic and Thermodynamic Study of the Glycosidic Bond Cleavage in Deoxyuridine. *J. Phys. Chem. B* **2007,** *111* (14), 3800–3812.

[2] Millen, A. L.; Wetmore, S. D., Glycosidic Bond Cleavage in Deoxynucleotides - a Density Functional Study. *Can. J. Chem.* **2009,** *87* (7), 850–863.

[3] Przybylski, J. L.; Wetmore, S. D., Designing an Appropriate Computational Model for DNA Nucleoside Hydrolysis: A Case Study of 2′-Deoxyuridine. *J. Phys. Chem. B* **2009,** *113* (18), 6533–6542.

[4] Przybylski, J. L.; Wetmore, S. D., Modeling the Dissociative Hydrolysis of the Natural DNA Nucleosides. *J. Phys. Chem. B* **2010,** *114* (2), 1104–1113.

[5] Shim, E. J.; Przybylski, J. L.; Wetmore, S. D., Effects of Nucleophile, Oxidative Damage, and Nucleobase Orientation on the Glycosidic Bond Cleavage in Deoxyguanosine. *J. Phys. Chem. B* **2010,** *114* (6), 2319–2326.

[6] Kellie, J. L.; Navarro-Whyte, L.; Carvey, M. T.; Wetmore, S. D., Combined Effects of π–π Stacking and Hydrogen Bonding on the (N1) Acidity of Uracil and Hydrolysis of 2'-Deoxyuridine. *J. Phys. Chem. B* **2012,** *116* (8), 2622–2632.

[7] Lenz, S. A. P.; Kellie, J. L.; Wetmore, S. D., Glycosidic Bond Cleavage in Deoxynucleotides: Effects of Solvent and the DNA Phosphate Backbone in the Computational Model. *J. Phys. Chem. B* **2012,** *116* (49), 14275–14284.

[8] Navarro-Whyte, L.; Kellie, J. L.; Lenz, S. A. P.; Wetmore, S. D., Hydrolysis of the Damaged Deoxythymidine Glycol Nucleoside and Comparison to Canonical DNA. *PCCP* **2013,** *15* (44), 19343–19352.

[9] Lenz, S. A. P.; Kellie, J. L.; Wetmore, S. D., Glycosidic Bond Cleavage in DNA Nucleosides: Effect of Nucleobase Damage and Activation on the Mechanism and Barrier. *J. Phys. Chem. B* **2015,** *119* (51), 15601–15612.

[10] Rios, A. C.; Yu, H. T.; Tor, Y., Hydrolytic Fitness of *N*-glycosyl Bonds: Comparing the Deglycosylation Kinetics of Modified, Alternative, and Native Nucleosides. *J. Phys. Org. Chem.* **2015,** *28* (3), 173–180.

[11] Iovane, E.; Giabba, B.; Muzzolini, L.; Matafora, V.; Fornili, A.; Minici, C.; Giannese, F.; Degano, M., Structural Basis for Substrate Specificity in Group I Nucleoside Hydrolases. *Biochemistry* **2008,** *47* (15), 4418–4426.

[12] Arivett, B.; Farone, M.; Masiragani, R.; Burden, A.; Judge, S.; Osinloye, A.; Minici, C.; Degano, M.; Robinson, M.; Kline, P., Characterization of Inosine-Uridine Nucleoside Hydrolase (RihC) from *Escherichia coli*. *Biochim. Biophys. Acta, Proteins Proteomics* **2014,** *1844* (3), 656–662.

[13] Horenstein, B. A.; Parkin, D. W.; Estupinan, B.; Schramm, V. L., Transition-state Analysis of Nucleoside Hydrolase from *Crithidia fasciculata*. *Biochemistry* **1991,** *30* (44), 10788–10795.

[14] Gopaul, D. N.; Meyer, S. L.; Degano, M.; Sacchettini, J. C.; Schramm, V. L., Inosine–Uridine Nucleoside Hydrolase from *Crithidia fasciculata*. Genetic Characterization, Crystallization, and Identification of Histidine 241 as a Catalytic Site Residue. *Biochemistry* **1996,** *35* (19), 5963–5970.

[15] Berg, M.; Kohl, L.; Van der Veken, P.; Joossens, J.; Al-Salabi, M. I.; Castagna, V.; Giannese, F.; Cos, P.; Versees, W.; Steyaert, J.; Grellier, P.; Haemers, A.; Degano, M.; Maes, L.; de Koning, H. P.; Augustyns, K., Evaluation of Nucleoside Hydrolase Inhibitors for Treatment of African Trypanosomiasis. *Antimicrob. Agents Chemother.* **2010,** *54* (5), 1900–1908.

[16] Kellie, J. L.; Navarro-Whyte, L.; Carvey, M. T.; Wetmore, S. D., Combined Effects of π–π Stacking and Hydrogen Bonding on the (N1) Acidity of Uracil and Hydrolysis of 2′-Deoxyuridine. *J. Phys. Chem. B* **2012,** *116* (8), 2622–2632.

[17] O'Brien, P. J.; Ellenberger, T., Human Alkyladenine DNA Glycosylase Uses Acid-Base Catalysis for Selective Excision of Damaged Purines. *Biochemistry* **2003,** *42* (42), 12418–12429.

[18] O'Brien, P. J.; Ellenberger, T., Dissecting the Broad Substrate Specificity of Human 3-Methyladenine-DNA Glycosylase. *J. Biol. Chem.* **2004,** *279* (11), 9750–9757.

[19] 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,N^4-Ethenocytosine-containing DNA. *J. Biol. Chem.* **2011,** *286* (15), 13205–13213.

[20] 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*, 16258–16269.

[21] Chang, S.-c.; Fedeles, B. I.; Wu, J.; Delaney, J. C.; Li, D.; Zhao, L.; Christov, P. P.; Yau, E.; Singh, V.; Jost, M.; Drennan, C. L.; Marnett, L. J.; Rizzo, C. J.; Levine, S. S.; Guengerich, F. P.; Essigmann, J. M., Next-generation Sequencing Reveals the Biological Significance of the N^2,3-ethenoguanine Lesion i*n vivo*. *Nucleic Acids Res.* **2015,** *43* (11), 5489–5500.

[22] Zdżalik, D.; Domańska, A.; Prorok, P.; Kosicki, K.; van den Born, E.; Falnes, P. Ø.; Rizzo, C. J.; Guengerich, F. P.; Tudek, B., Differential Repair of Etheno-DNA Adducts by Bacterial and Human AlkB Proteins. *DNA Repair* **2015,** *30*, 1–10.

[23] 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*, 855.

[24] Mishina, Y.; Yang, C.-G.; He, C., Direct Repair of the Exocyclic DNA Adduct 1,N^6-Ethenoadenine by the DNA Repair AlkB Proteins. *J. Am. Chem. Soc.* **2005,** *127* (42), 14594–14595.

[25] Muller, T. A.; Hausinger, R. P., CHAPTER 8 AlkB and Its Homologues - DNA Repair and Beyond. In *2-Oxoglutarate-Dependent Oxygenases*, The Royal Society of Chemistry: 2015; pp 246–262.

[26] Lee, A. J.; Wallace, S. S., Hide and Seek: How Do DNA Glycosylases Locate Oxidatively Damaged DNA Bases Amidst a Sea of Undamaged Bases? *Free Radical Biol. Med.* **2017,** *107*, 170–178.

[27] Martin, F. H.; Castro, M. M.; Aboul-ela, F.; Tinoco, I., Base Pairing Involving Deoxyinosine: Implications for Probe Design. *Nucleic Acids Res.* **1985,** *13* (24), 8927–8938.

[28] Bergonzo, C.; Campbell, A. J.; de los Santos, C.; Grollman, A. P.; Simmerling, C., Energetic Preference of 8-oxoG Eversion Pathways in a DNA Glycosylase.. *J. Am. Chem. Soc.* **2011,** *133* (37), 14504-14506.

[29] Li, H.; Endutkin, A. V.; Bergonzo, C.; Fu, L.; Grollman, A.; Zharkov, D. O.; Simmerling, C., DNA Deformation-coupled Recognition of 8-oxoguanine: Conformational Kinetic Gating in Human DNA Glycosylase. *J. Am. Chem. Soc.* **2017,** *139* (7), 2682–2692.

[30] Theruvathu, J. A.; Yin, Y. W.; Pettitt, B. M.; Sowers, L. C., Comparison of the Structural and Dynamic Effects of 5-Methylcytosine and 5-Chlorocytosine in a CpG Dinucleotide Sequence. *Biochemistry* **2013,** *52* (47), 8590–8598.

[31] Lee, C.-Y. I.; Delaney, J. C.; Kartalou, M.; Lingaraju, G. M.; Maor-Shoshani, A.; Essigmann, J. M.; Samson, L. D., Recognition and Processing of a New Repertoire of DNA Substrates by Human 3-Methyladenine DNA Glycosylase (AAG). *Biochemistry* **2009,** *48* (9), 1850–1861.

[32] Matijasevic, Z.; Sekiguchi, M.; Ludlum, D. B., Release of N^2,3-ethenoguanine from Chloroacetaldehyde-treated DNA by *Escherichia coli* 3-methyladenine DNA Glycosylase Ii. *Proc. Natl. Acad. Sci. USA* **1992,** *89* (19), 9331–9334.

[33] Li, D.; Delaney, J. C.; Page, C. M.; Yang, X.; Chen, A. S.; Wong, C.; Drennan, C. L.; Essigmann, J. M., Exocyclic Carbons Adjacent to the N6 of Adenine are Targets for Oxidation by the Escherichia coli Adaptive Response Protein AlkB. *J. Am. Chem. Soc.* **2012,** *134* (21), 8896–8901.

[34] Singh, V.; Fedeles, B. I.; Li, D.; Delaney, J. C.; Kozekov, I. D.; Kozekova, A.; Marnett, L. J.; Rizzo, C. J.; Essigmann, J. M., Mechanism of Repair of Acrolein- and Malondialdehyde-Derived Exocyclic Guanine Adducts by the α-Ketoglutarate/Fe(II) Dioxygenase AlkB. *Chem. Res. Toxicol.* **2014,** *27* (9), 1619–1631.

[35] Mallinson, S. J. B.; Machovina, M. M.; Silveira, R. L.; Garcia-Borràs, M.; Gallup, N.; Johnson, C. W.; Allen, M. D.; Skaf, M. S.; Crowley, M. F.; Neidle, E. L.; Houk, K. N.; Beckham, G. T.; DuBois, J. L.; McGeehan, J. E., A Promiscuous Cytochrome P450 Aromatic O-Demethylase for Lignin Bioconversion. *Nat. Commun.* **2018,** *9* (1), 2487.

[36] Mast, N.; Charvet, C.; Pikuleva, I. A.; Stout, C. D., Structural Basis of Drug Binding to CYP46A1, an Enzyme That Controls Cholesterol Turnover in the Brain. *J. Biol. Chem.* **2010,** *285* (41), 31783-31795.

[37] McDonnell, A. M.; Dang, C. H., Basic Review of the Cytochrome P450 System. *J. Pract. Oncol.* **2013,** *4* (4), 263-–68.

[38] Rodríguez, D.; Morrison, C. J.; Overall, C. M., Matrix Metalloproteinases: What Do They Not Do? New Substrates and Biological Roles Identified by Murine Models and Proteomics. *Biochim. Biophys. Acta* **2010,** *1803* (1), 39–54.

[39] Jabłońska-Trypuć, A.; Matejczyk, M.; Rosochacki, S., Matrix Metalloproteinases (MMPs), the Main Extracellular Matrix (ECM) Enzymes in Collagen Degradation, as a Target for Anticancer Drugs. *J. Enzyme Inhib. Med. Chem.* **2016,** *31* (sup1), 177-183.

Appendix A: Supplemental Information for Chapter 2: Evaluating the Substrate Selectivity of Alkyladenine DNA Glycosylase: The Synergistic Interplay of Active Site Flexibility and Water Reorganization

Table of Contents

Table A.1. Root-mean-square deviation of the AAG active site over the MD simulation trajectory relative to the experimental crystal structure. 217

Table A.2. Average root-mean-square deviation of the select AAG active site residues over two trajectories for the entire production phase.[a] 218

Table A.3. Average backbone and side chain torsional angles for key active site residues during MD simulations of AAG bound to various nucleotides. 219

Table A.4. Summary of important hydrogen bonds formed in the AAG active site during MD simulations of AAG bound to various nucleotides. 221

Table A.5. Average distance and standard deviation between the bound nucleotide and key active site residues throughout the MD simulation. 223

Table A.6. Average pairwise MM–GBSA free energy and standard deviations between the bound nucleotide and key active site amino acids over the MD simulation.[a] 224

Figure A.1. Overlay of crystallographic structure and a representative structure from the corresponding MD trajectory of AAG bound to εA and AAG bound to εC. 225

Figure A.2. Distribution of the active site water during MD simulations when εA, εC, Hx, A, G, 3MeA, or 7MeG is bound to AAG. 226

Figure A.3. Key hydrogen bonds between AAG active site residues and bound εC and Hx. 227

Full Citation for References 63, 64, and 69 228

Table A.1. Root-mean-square deviation (RMSD) of the AAG active site over the MD simulation trajectory relative to the experimental crystal structure.[a]

Simulation	Crystal Structure	RMSD[b]	RMSD (σ)[c]
εA	1EWN	0.404	0.801 (0.063)
εC	3QI5	0.537	0.657 (0.085)

[a]RMSD calculated according to the position of the heavy atoms in the side chains and backbones of dX, E125, Y127, A135, H136, Y159, N169, L180, R182, and sugar and phosphate of the bound nucleotide. [b]Initial RMSD at the start of the production MD simulation, after the initial minimization and equilibration steps. [c]Average RMSD during the production MD simulation.

Table A.2. Average root-mean-square deviation (RMSD) (standard deviation (σ) provided in parentheses) of the select AAG active site residues over two trajectories for the entire production phase.[a]

dX	dX	Active Site	Sugar/Phosphate	dX	E125	Y127	A135
εA	εC	1.433 (0.111)	1.667 (0.219)	-	1.248 (0.117)	2.099 (0.1933)	0.997 (0.234)
εA	Hx	0.981 (0.108)	1.461 (0.315)	1.224 (0.248)	0.576 (0.141)	0.548 (0.145)	0.555 (0.177)
A	Hx	2.072 (0.106)	2.345 (0.117)	3.297 (0.086)	0.577 (0.130)	0.583 (0.136)	0.741 (0.248)
G	Hx	1.647 (0.057)	2.341 (0.151)	3.082 (0.072)	1.151 (0.204)	0.743 (0.193)	0.758 (0.216)
A	3MeA	2.592 (0.218)	2.979 (0.256)	4.238 (0.273)	1.406 (0.133)	1.394 (0.223)	1.552 (0.289)
G	7MeG	0.936 (0.093)	1.251 (0.253)	1.201 (0.211)	0.598 (0.134)	0.672 (0.159)	0.839 (0.202)

dX	dX	A135	H136	Y159	N169	L180	R182
εA	εC	0.997 (0.234)	1.045 (0.320)	1.545 (0.089)	1.478 (0.564)	0.646 (0.294)	0.900 (0.157)
εA	Hx	0.555 (0.177)	0.776 (0.234)	0.831 (0.273)	0.624 (0.195)	0.485 (0.159)	1.672 (0.225)
A	Hx	0.741 (0.248)	2.137 (0.443)	2.404 (0.480)	2.075 (0.202)	0.567 (0.170)	1.450 (0.225)
G	Hx	0.758 (0.216)	0.939 (0.238)	0.883 (0.245)	1.377 (0.296)	0.659 (0.190)	0.726 (0.219)
A	3MeA	1.552 (0.289)	2.859 (0.592)	2.746 (0.566)	1.911 (0.321)	0.776 (0.251)	1.019 (0.336)
G	7MeG	0.839 (0.202)	0.907 (0.288)	0.719 (0.242)	0.705 (0.237)	0.584 (0.172)	1.276 (0.156)

[a]RMS fit was performed based on the position of the active site residues (dX, E125, Y127, A135, H136, Y159, N169, L180, and R182) to yield the overall RMSD of the active site, and subsequently the RMSD per residue was determined without refitting.

Table A.3. Average backbone and side chain torsional angles (standard deviation (σ) provided in parentheses) for key active site residues during MD simulations of AAG bound to various nucleotides.

	Crystal Structure[a]		εC		εA		Hx		A	
Angle[b-d]	1EWN	3QI5	Avg (σ)	%[e]	Avg (σ)	%[e]	Avg (σ)	%[e]	Avg (σ)	%[e]
E125 χ_1	186.1	−79.5	193.3 (8.3)	93.4	192.8 (7.5)	96.8	190.9 (7.3)	99.3	191.9 (8.8)	94.8
E125 φ	−145.5	−134.9	−139.1 (7.3)	81.9	−137.3 (7.7)	89.6	−136.9 (7.7)	90.5	−135.2 (8.4)	93.1
E125 ψ	129.4	140.9	131.8 (8.9)	94.4	138.1 (7.4)	85	136.7 (7.7)	89.3	135.9 (8.3)	87.1
Y127 χ_1	−71.1	−70.2	−74.4 (5.5)	99.7	−47.2 (6.0)	99.7	−47.5 (5.8)	99.7	−52.4 (6.6)	99.9
Y127 φ	−130.9	−118.8	−130.0 (8.7)	98.8	−129.1 (9.2)	99.1	−129.3 (9.1)	99	−126.3 (9.3)	99.7
Y127 ψ	119.7	117.9	121.7 (12.8)	98	136.8 (7.7)	91.1	138.9 (6.9)	87.1	134.3 (9.2)	94.2
H136 χ_1	−78.3	−65.3	−73.0 (11.3)	92.7	−76.4 (10.2)	89.5	−69.2 (9.3)	98.5	193.8 (7.2)	74
H136 φ	−63.7	−68.5	−62.3 (8.9)	99.5	−56.3 (8.3)	99.8	−60.3 (8.9)	99.9	−64.2 (9.3)	99.6
H136 ψ	−25.4	−12.4	−17.2 (7.7)	58.8	−39.4 (6.7)	60.4	−39.8 (7.2)	53.7	−42.3 (7.6)	57
Y159 χ_1	71.4	73.6	69.9 (7.2)	99.6	73.5 (6.7)	98.8	67.2 (7.7)	99.2	176.4 (10.5)	92.8
Y159 φ	−131.8	−125.2	−133.0 (11.3)	89.9	−133.8 (10.7)	90.2	−131.3 (11.5)	92.9	−73.8 (9.9)	76.1
Y159 ψ	162.6	158.9	158.2 (5.9)	65	158.9 (6.2)	71.6	161.5 (7.2)	85.2	−73.8 (9.9)	76.1
N169 χ_1	−57.8	−62.0	−75.5 (12.3)	46.7	−45.3 (9.4)	89.9	−42.8 (9.0)	71.8	177.7 (9.0)	99.7
N169 φ	−122.7	−109.9	−115.6 (12.6)	95.7	−127.2 (11.5)	98.3	−136.4 (9.0)	88	−120.5 (15.0)	88.4
N169 ψ	150.3	155	157.5 (5.3)	54.6	157.9 (5.5)	60.8	139.3 (8.3)	60.1	138.9 (8.1)	68.1
L180 χ_1	164.2	174.5	180.8 (11.9)	95.1	171.2 (6.4)	99.9	169.4 (6.4)	99.9	167.0 (8.3)	97.2
L180 φ	−91.5	−83.6	−101.9 (9.5)	53.8	−101.2 (8.1)	63.5	−99.4 (7.2)	50.5	−80.4 (6.8)	55.3
L180 ψ	134	134.2	136.7 (6.9)	94.8	138.8 (6.5)	90.9	138.1 (6.5)	93.6	136.2 (7.2)	95.5
R182 χ_1	−69.1	−81.9	−75.2 (9.1)	87.6	−64.4 (9.2)	92.3	−77.2 (7.7)	81.2	−61.2 (9.1)	57.6
R182 φ	−109.0	−108.8	−107.1 (9.9)	87	−107.1 (9.5)	91.7	−106.0 (8.8)	89	−106.1 (9.6)	87.5
R182 ψ	−23.6	−27.4	−38.8 (6.9)	55.4	−22.2 (5.7)	52.3	−23.0 (5.1)	51.6	−39.5 (7.1)	65.8

	G		3MeA		7MeG	
Angle[b-d]	Avg (σ)	%[e]	Avg (σ)	%[e]	Avg (σ)	%[e]
E125 χ_1	192.3 (8.6)	65.9	195.4 (7.7)	88.4	192.7 (7.8)	95.9
E125 φ	–136.9 (8.4)	87.2	–140.4 (6.8)	76.9	–138.5 (7.4)	86
E125 ψ	140.4 (6.7)	63.1	133.6 (9.1)	89	137.6 (7.4)	85.8
Y127 χ_1	–57.6 (7.0)	100	–71.0 (6.3)	99.9	–47.9 (5.8)	99.9
Y127 φ	–134.0 (8.7)	96.5	–132.6 (8.1)	98.1	–127.9 (9.2)	99.4
Y127 ψ	131.6 (9.9)	95.8	127.5 (11.6)	95.5	139.0 (6.8)	87.9
H136 χ_1	–77.7 (7.2)	92.7	–54.1 (12.8)	95.3	–60.5 (9.3)	99.8
H136 φ	–60.6 (7.7)	100	–64.1 (8.4)	99.7	–59.5 (8.0)	99.9
H136 ψ	–19.2 (6.8)	75	–17.6 (7.5)	74.2	–19.5 (7.1)	64.5
Y159 χ_1	73.3 (6.9)	98.9	60.0 (12.0)	99.4	71.6 (8.1)	97.9
Y159 φ	–131.1 (11.2)	94.3	–132.2 (12.5)	86.4	–130.7 (11.7)	92.5
Y159 ψ	158.7 (6.4)	62.4	159.0 (6.1)	71.8	161.2 (7.1)	82.6
N169 χ_1	–60.6 (9.8)	99.7	–53.9 (13.8)	76.2	–45.5 (8.2)	94.2
N169 φ	–117.3 (14.9)	90.7	–123.5 (13.0)	96.8	–129.4 (11.1)	96.9
N169 ψ	138.8 (8.8)	68.5	139.4 (7.8)	65.5	162.3 (7.0)	86.2
L180 χ_1	172.0 (7.3)	99.8	172.7 (8.2)	99.5	170.2 (7.2)	99.6
L180 φ	–104.1 (9.4)	75	–106.6 (11.2)	75.6	–79.8 (6.8)	66
L180 ψ	137.6 (7.1)	91.8	137.7 (7.1)	91.2	136.0 (6.8)	97.2
R182 χ_1	–76.8 (8.1)	62.5	–64.8 (8.7)	98.7	188.4 (16.1)	77.8
R182 φ	–104.6 (8.9)	82.7	–113.9 (9.7)	98.6	–105.2 (9.3)	84.3
R182 ψ	–39.8 (7.6)	60	–22.5 (5.3)	60.2	–38.4 (6.2)	65.4

[a]See references 31 and 35 in Chapter 2. [b]χ1 defined as $\angle$(NCαCβCγ). [c]φ defined as $\angle$(C1NCα C2), with φ measured for residue 1. [d]ψ defined as $\angle$(N0CαC1N1), with ψ is measured for residue 1. [d]Percentage of the trajectory that the dihedral angle in the preferred conformation falls within one of the following ranges: –30 to 30; 30 to 90; 90 to 150; 150 to –150; –90 to –150; or –30 to –90.

Table A.4. Summary of important hydrogen bonds formed in the AAG active site during MD simulations of AAG bound to various nucleotides.

Hydrogen Bond	εC		εA		Hx		A	
	%[a]	Dist (Angle)[b]	%[a,c]	Dist (Angle)[b]	%[a]	Dist (Angle)[b]	%[a,c]	Dist (Angle)[b]
N169(Oε)···dX(N2H)	-	-	-	-	-	-	-	-
dX (O2/N1)···N169(NεH)	44.8	3.097 (158.6)	-	-	-	-	23.5	3.009 (159.0)
E125(Oε2)···Y127(OH)	100	2.636 (164.0)	99.6	2.690 (157.5)	99.9	2.670 (157.9)	99.7	2.721 (161.4)
dX(N2/N4/O6)···H136(NH)	98	3.035 (160.2)	99.6	2.996 (164.6)	100	2.828 (164.9)	-	-
dX(O1P) ···Y159(OH)	82.9	2.876 (151.4)	92.3	2.856 (151.0)	5.1	2.859 (151.2)	97.6	2.691 (160.6)
dX(O1P)···H136(NεH)	96.9	2.841 (154.6)	99.5	2.814 (158.4)	54.5	2.884 (151.2)	23.7	2.845 (155.8)
dX(O3)···R182(NηH)	-	-	0.8	3.209 (128.9)	85.7	3.084 (154.8)	1.7	3.075 (126.3)
dX(N3)···Y159(OH)	-	-	-	-	73.1	2.980 (145.2)	-	-
E125(Oε)···H_2O(OH)	7.3	2.668 (163.9)	175.8[b]	-	100	2.698 (164.3)	156.9	2.802 (153.5)
H_2O(O)···dX(N6H)	-	-	-	-	-	-	167.3	3.042 (136.5)
dX(N7)···H_2O(OH)	-	-	8.1	2.976 (152.3)	47.7	3.127 (148.2)	101.6	2.851 (159.2)
H_2O(O)···Y127(OH)	4.4	2.784 (157.3)	56.8	3.230 (144.1)	25.5	3.260 (154.0)	66.2	3.181 (143.0)

Hydrogen Bond	G		3MeA		7MeG	
	%[a,c]	Dist (Angle)[b]	%[a]	Dist (Angle)[c]	%[a,c]	Dist (Angle)[b]
N169(Oε)···dX(N2H)	99.6	2.789 (155.6)	-	-	99.6	2.783 (155.5)
dX (O2/N1)···N169(NεH)	-	-	-	-	-	-
E125(Oε2)···Y127(OH)	-	-	100	2.643 (166.5)	99.7	2.651 (163.8)
dX(N2/N4/O6)···H136(NH)	96.2	2.989 (156.8)	-	-	95.6	3.001 (157.1)
dX(O1P) ···Y159(OH)	55.3	2.859 (152.5)	1.4	2.811 (161.1)	66.2	2.859 (152.5)
dX(O1P)···H136(NεH)	90.2	2.914 (152.6)	-	-	89.1	2.923 (151.7)
dX(O3)···R182(NηH)	9.3	3.045 (126.2)	-	-	8.8	3.045 (125.9)
dX(N3)···Y159(OH)	-	-	-	-	-	-
E125(Oε)···H_2O(OH)	135.5	2.792 (159.2)	98.3	2.726 (160.3)	151.9	2.753 (155.3)
H_2O(O)···dX(N6H)	-	-	-	-	-	-
dX(N7)···H_2O(OH)	-	-	4.9	3.000 (148.8)	-	-
H_2O(O)···Y127(OH)	197.1	2.827 (149.2)	20.9	3.006 (156.9)	56.2	3.232 (142.1)

[a]Percent occupancy of the hydrogen bond, which is defined using cutoffs of a distance less than 3.4 Å and an angle greater than 120°. [b]Hydrogen bonding distance in Å and angle in ° (in parantheses). [c]Greater than 100% occupancy indicates the presence of more than one water hydrogen bonding with an acceptor site.

Table A.5. Average distance and standard deviation (σ) between the bound nucleotide (dX) and key active site residues throughout the MD simulation.

	Y159		H136		Y127		E125(Oε)···dX(C1′)	
dX	Initial[a]	Avg (σ)[b]	Initial[a]	Avg (σ)[b]	Initial[a]	Avg (σ)[b]	Initial[a]	Avg (σ)
εA	5.434	5.613 (0.278)	4.880	4.570 (0.322)	3.675	3.978 (0.202)	6.449	4.758 (0.308)
εC	4.914	5.432 (0.344)	5.365	5.042 (0.555)	3.755	3.650 (0.203)	4.295	4.532 (0.263)
Hx	6.200	5.579 (0.236)	4.804	4.101 (0.304)	3.807	3.621 (0.163)	6.673	3.928 (0.275)
A	5.824	4.376 (0.387)	4.810	6.325 (0.779)	3.505	4.564 (0.252)	6.390	4.921 (0.364)
G	5.114	5.474 (0.291)	4.379	4.858 (0.323)	3.854	3.779 (0.226)	5.068	5.790 (0.651)
3MeA	5.508	6.509 (0.827)	4.380	5.224 (0.333)	3.812	4.251 (0.357)	5.224	5.454 (0.412)
7MeG	5.667	5.834 (0.377)	3.717	4.124 (0.284)	3.726	4.367 (0.193)	6.309	4.564 (0.283)

[a]Initial distance measured at the start of the production MD simulation, after the initial minimization and equilibration steps.

[b]Average distance and standard deviation (between aromatic the amino acids and dX was measured) between the center of masses of the side chains and the endocyclic nucleobase atoms.

Table A.6. Average pairwise MM–GBSA free energy and standard deviations (σ; kJ mol^{-1}) between the bound nucleotide (dX) and key active site amino acids over the MD simulation.[a]

	E125	Y127	H136	Y159	N169	L180	R182
dX	Avg (σ)[a]	Avg (σ)[a]	Avg (σ)[a]	Avg (σ)[a]	Avg (σ)[a]	Avg (σ)[a]	Avg (σ)[a]
εA	–1.9 (1.1)	–16.7 (1.9)	–43.8 (3.8)	–28.9 (6.6)	–9.0 (1.4)	–7.3 (1.5)	–7.3 (5.6)
εC	4.2 (1.9)	–20.5 (2.1)	–41.5 (4.1)	–22.3 (6.6)	–11.4 (5.5)	–13.5 (1.8)	–19.5 (8.0)
Hx	4.4 (3.0)	–18.3 (2.2)	–32.7 (9.0)	–10.8 (4.5)	–5.2 (1.3)	–9.5 (1.2)	–49.7 (7.9)
A	1.9 (1.3)	–13.5 (2.3)	–11.3 (11.4)	–40.7 (6.0)	–10.7 (4.5)	–11.2 (1.5)	–18.8 (8.7)
G	1.7 (1.7)	–20.4 (2.4)	–40.3 (4.0)	–28.1 (6.1)	–24.4 (2.8)	–7.7 (2.0)	–35.4 (14.5)
3MeA	–10.1 (4.8)	–18.6 (3.9)	–18.0 (5.0)	–8.5 (6.9)	–3.6 (3.7)	–5.2 (1.9)	–3.5 (2.1)
7MeG	–7.1 (1.6)	–18.1 (2.1)	–41.4 (5.6)	–21.8 (10.2)	–20.3 (3.4)	–5.9 (1.5)	–23.1 (4.1)

[a]MM–GBSA values calculated throughout the entire production MD simulation.

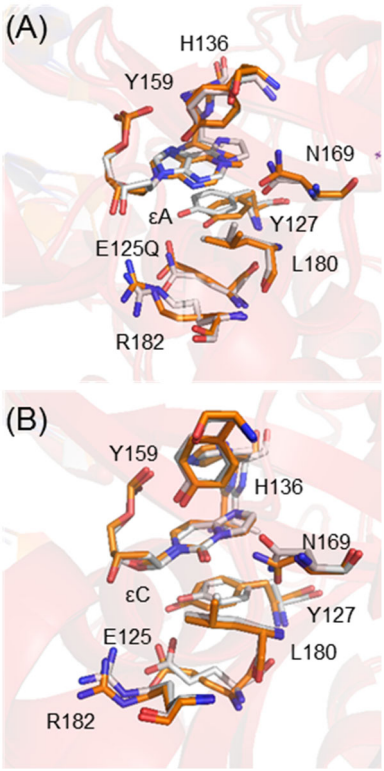


Figure A.1. Overlay of crystallographic structure (gray) and a representative structure from the corresponding MD trajectory (orange) of (A) AAG bound to εA (PDB ID: 1EWN) and (B) AAG bound to εC (PDB ID: 3QI5).

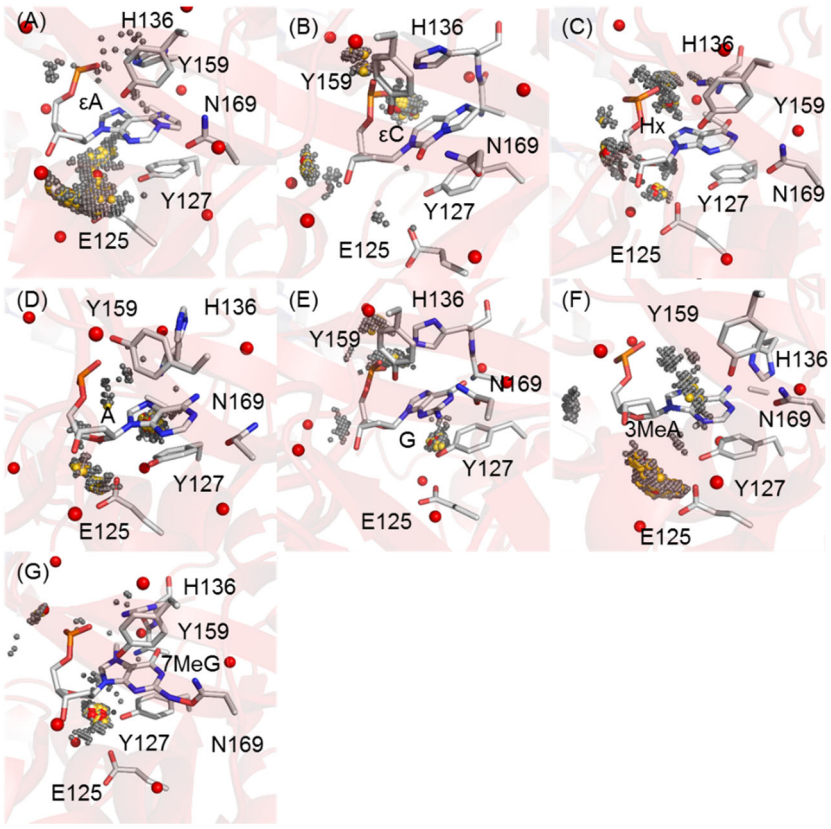


Figure A.2. Distribution of the active site water during MD simulations when (A) εA, (B) εC, (C) Hx, (D) A, (E) G, (F) 3MeA, or (G) 7MeG is bound to AAG. Spheres indicate relative occupancy of different positions across the simulation trajectory (red: 80% occupied; yellow: 60% occupied; gray: 40% occupied).

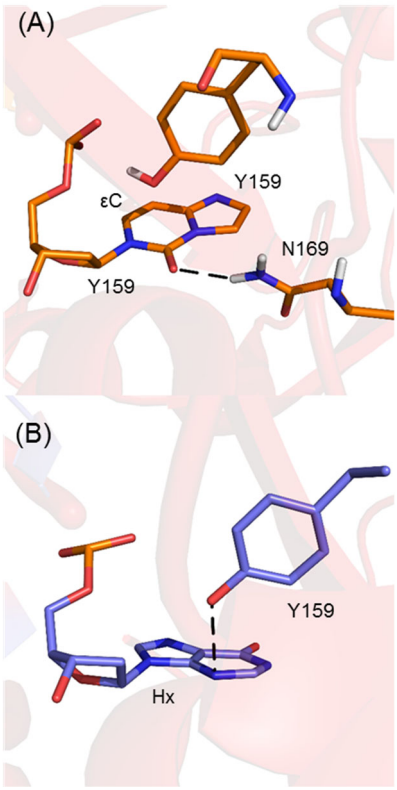


Figure A.3. Key hydrogen bonds between AAG active site residues and bound (A) εC and (B) Hx.

Full Citation for References 63, 64, and 69:

63. Wang, L. F., Zhou, Y., Xu, L., Xiao, R., Lu, X. Y., Chen, L., Chong, J., Li, H. R., He, C., Fu, X. D., and Wang, D. (2015) Molecular basis for 5–carboxycytosine recognition by RNA polymerase II elongation complex, *Nature* 523, 621.

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

69. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Jr., J. A. M.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.; Revision A.02 ed.; Gaussian, Inc.: Wallingford CT, 2009.

Appendix B: Supplemental Information for Chapter 3: An ONIOM(QM:MM) Study of the Reaction Catalyzed by Alkyladenine DNA Glycosylase: Examination of the Substrate Specificity of a DNA Repair Enzyme

Table of Contents

Table B.1. Summary of clustered MD simulations of AAG bound to DNA-containing Hx, G, or 7MeG. 230

Table B.2. Deviations in calculated position of active site residues (rmsd) when AAG is bound to Hx, G, or 7MeG using different ONIOM(QM:MM) methodologies 231

Figure B.1. Overlay of the X-ray crystal structure of AAG bound to DNA-containing εA and corresponding complex optimized with ONIOM(B3LYP/6-31G(d):AMBER). 232

Figure B.2. Structures of constrained stationary points corresponding to AAG-mediated excision of Hx. 233

Figure B.3. Overlay of the unconstrained ONIOM(B3LYP-D3/6-31G(d):AMBER) RC and representative MD (AMBER) structure of Hx bound in the AAG active site. 234

Figure B.4. NCI plots depicting interactions between the nucleotide and H136, Y127, Y159, or water along the AAG-mediated excision of Hx.. 235

Figure B.5. Overlay of the constrained ONIOM(B3LYP-D3/6-31G(d):AMBER) RC and representative MD (AMBER) structure of G bound in the AAG active site. 236

Figure B.6. NCI plot depicting interactions between the nucleotide and H136, Y127, Y159, or water for the G reactant complex and representative reaction surface complex 237

Figure B.7. Overlay of the constrained ONIOM(B3LYP-D3/6-31G(d):AMBER) and representative MD (AMBER) structure of 7MeG bound in the AAG active site. 238

Figure B.8. Structures of constrained ONIOM(B3LYP-D3/6-31G(d):AMBER) concerted TS corresponding to AAG-mediated excision of 7MeG. 239

Figure B.9. NCI plots of the Structures of constrained ONIOM(B3LYP-D3/6-31G(d):AMBER) stationary points corresponding to AAG-mediated excision of 7MeG. 240

Full Citation for Reference 65, 68, and 74 241

Table B.1. Summary of clustered MD simulations of AAG bound to DNA-containing Hx, G, or 7MeG.[a,b]

	Cluster	Occupancy	Avg Distance (σ)[c]	Avg Cluster Distance[d]
Hx	1	97.6%	0.544 (0.107)	0.818
	2	2.3%	0.594 (0.122)	0.846
	3	0.1%	0.487 (0.101)	0.885
G	1	95.8%	0.606 (0.105)	0.897
	2	2.7%	0.521 (0.103)	0.874
	3	0.7%	0.587 (0.124)	0.929
7MeG	1	93.7%	0.517 (0.102)	0.837
	2	6.2%	0.436 (0.069)	0.882
	3	0.1%	0.567 (0.056)	0.913

[a]Each trajectory was clustered according to the positions of the bound nucleotide and active site residues (E125, Y127, A135, H136, Y159, and N169) using a hierarchical agglomerative average linkage rmsd algorithm. [b]Final production simulations were taken from reference 41 in the main text. [c]Average distance (Å) between each point within a cluster. [d]Average distance (Å) between each cluster.

Table B.2. Deviations in calculated position of active site residues (rmsd, Å) when AAG is bound to Hx, G, or 7MeG using different ONIOM(QM:MM) methodologies

Bound Nucleotide	Stationary Point	Calculation Method		rmsd[a]
Hx	RC	Constrained B3LYP[b]	Constrained B3LYP-D3[c]	0.134
Hx	TS	Constrained B3LYP[b]	Constrained B3LYP-D3[c]	0.141
Hx	PC	Constrained B3LYP[b]	Constrained B3LYP-D3[c]	0.143
Hx	RC	Unconstrained B3LYP-D3[d]	Constrained B3LYP-D3[c]	0.010
Hx	TS	Unconstrained B3LYP-D3[d]	Constrained B3LYP-D3[c]	0.025
Hx	PC	Unconstrained B3LYP-D3[d]	Constrained B3LYP-D3[c]	0.164
G	RC	Constrained B3LYP[b]	Constrained B3LYP-D3[c]	0.170
G	RC	Unconstrained B3LYP-D3[d]	Constrained B3LYP[b]	0.188
G	RC	Unconstrained B3LYP-D3[d]	Constrained B3LYP-D3[c]	0.037
7MeG	R	Constrained B3LYP[b]	Constrained B3LYP-D3[c]	0.151
7MeG	$TS_{dissociative}$	Constrained B3LYP[b]	Constrained B3LYP-D3[c]	0.278
7MeG	I	Constrained B3LYP[b]	Constrained B3LYP-D3[c]	0.162
7MeG	$TS_{associative}$	Constrained B3LYP[b]	Constrained B3LYP-D3[c]	0.179
7MeG	PC	Constrained B3LYP[b]	Constrained B3LYP-D3[c]	0.213
7MeG	$TS_{concerted}$	Constrained B3LYP[b]	Constrained B3LYP-D3[c]	0.221

[a]rmsd (Å) of active site residues in the QM layer (including the bound nucleotide, E125, Y127, A135, H136, Y159, and N169). [b]Constrained ONIOM(B3LYP/6-31G(d):AMBER) stationary points. [c]Constrained ONIOM(B3LYP-D3/6-31G(d):AMBER) stationary points. [d]Unconstrained ONIOM(B3LYP-D3/6-31G(d):AMBER) stationary points.

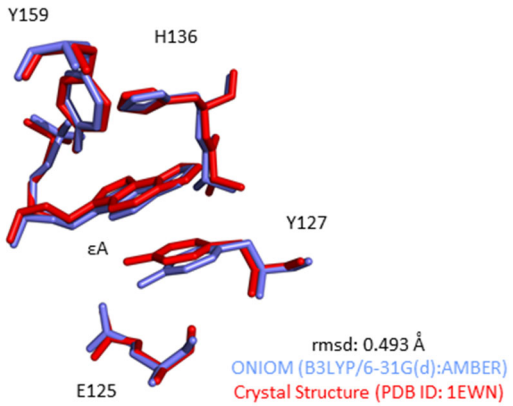


Figure B.1. Overlay of the X-ray crystal structure of AAG bound to DNA-containing εA (red, see reference 34) and corresponding complex optimized with ONIOM(B3LYP/6-31G(d):AMBER) (blue).

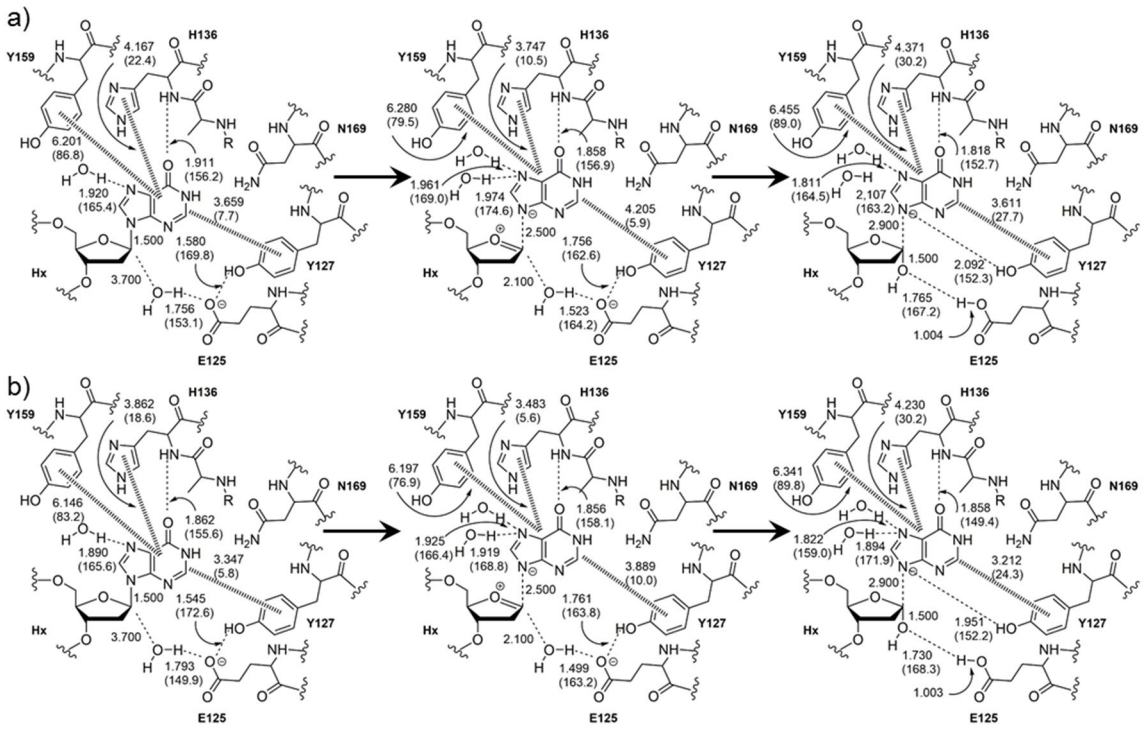


Figure B.2. Structures of constrained a) ONIOM(B3LYP/6-31G(d):AMBER) and b) ONIOM(B3LYP-D3/6-31G(d):AMBER) stationary points (RC, TS, and PC) corresponding to AAG-mediated excision of Hx displayed as line diagrams, with distances in Å, angles in deg, and hashed bonds denoting π–π stacking or T-shaped interactions.

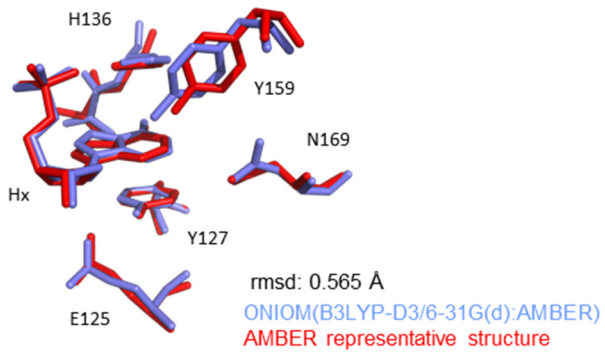


Figure B.3. Overlay of the unconstrained ONIOM(B3LYP-D3/6-31G(d):AMBER) RC (blue) and representative MD (AMBER) structure (red) of Hx bound in the AAG active site.

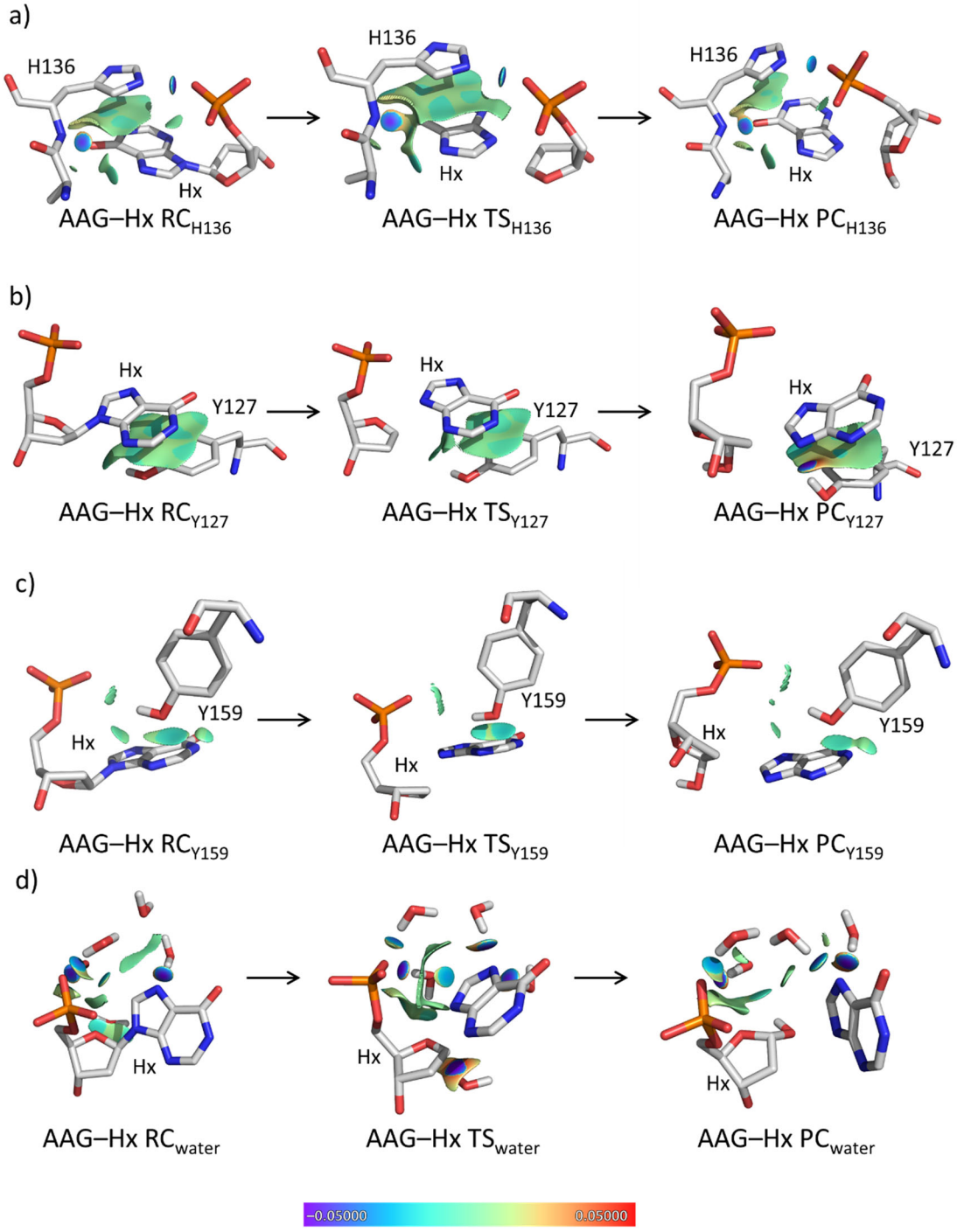


Figure B.4. NCI plots depicting interactions between the nucleotide and a) H136, b) Y127, c) Y159, or d) water along the AAG-mediated excision of Hx. Promolecular densities are reported when the reduced density gradient is less than or equal to 0.3 and the color ranges from –0.0500 au (attractive interaction) to 0.0500 au (repulsive interaction).

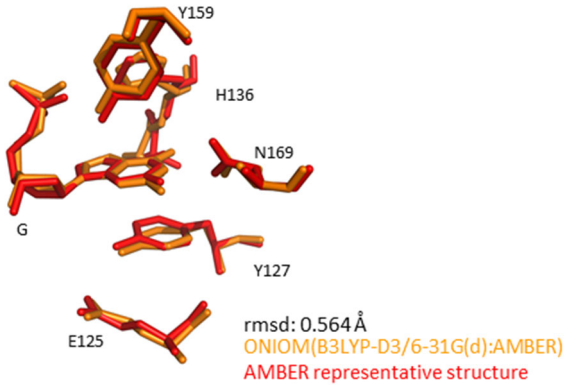


Figure B.5. Overlay of the constrained ONIOM(B3LYP-D3/6-31G(d):AMBER) RC (orange) and representative MD (AMBER) structure (red) of G bound in the AAG active site.

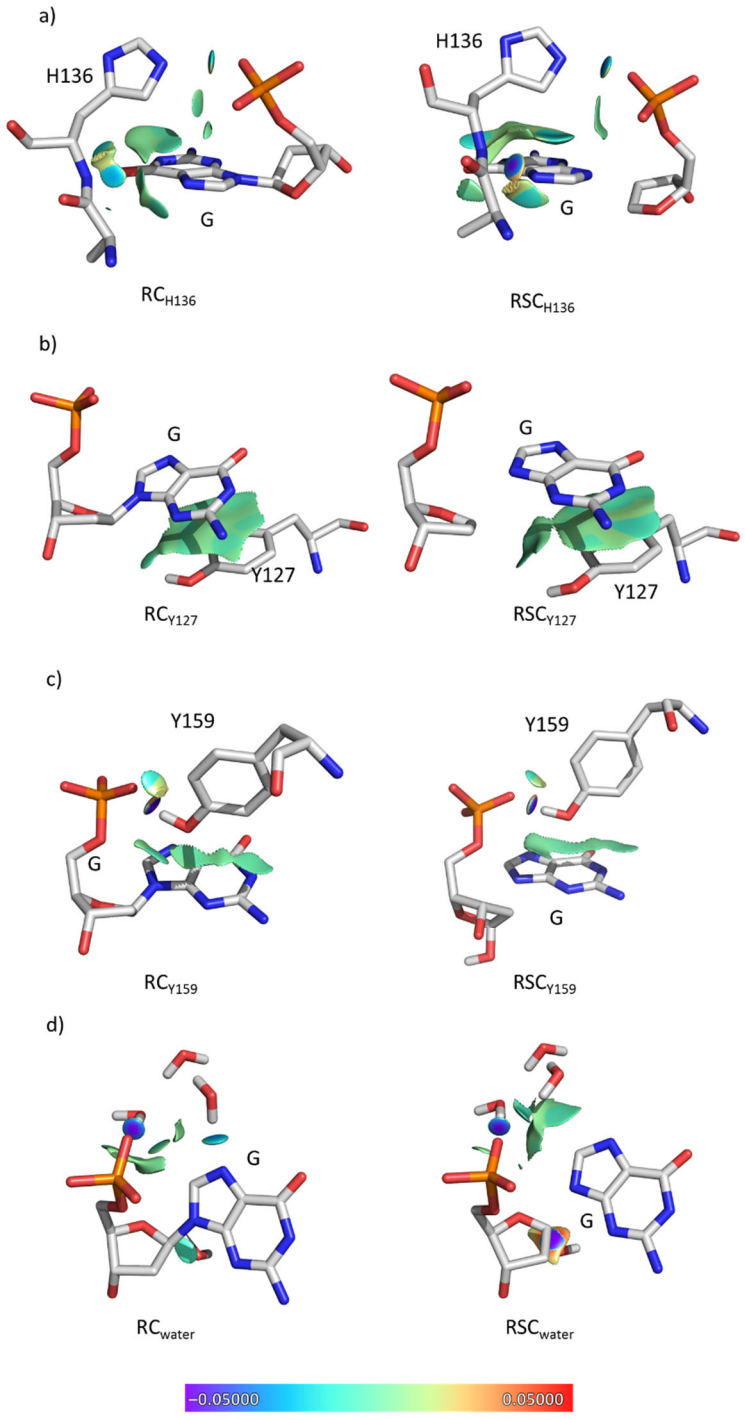


Figure B.6. NCI plot depicting interactions between the nucleotide and a) H136, b) Y127, c) Y159, or d) water for the G reactant complex (RC) and representative structure with an elongated glycosidic bond obtained from the G reaction surface (denoted RSC). Promolecular densities are reported when the reduced density gradient is less than or equal to 0.3 and the color ranges from –0.0500 au (attractive interaction) to 0.0500 au (repulsive interaction).

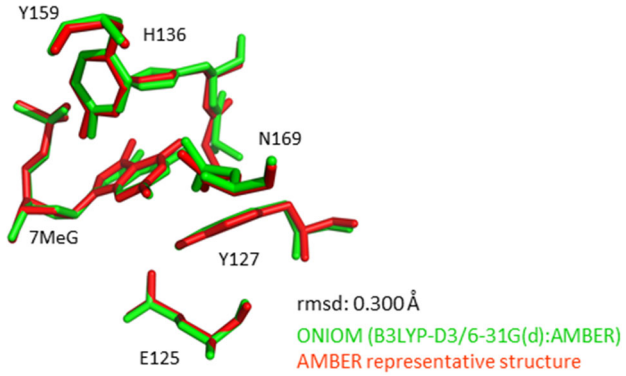


Figure B.7. Overlay of the constrained ONIOM(B3LYP-D3/6-31G(d):AMBER) RC (green) and representative MD (AMBER) structure (red) of 7MeG bound in the AAG active site.

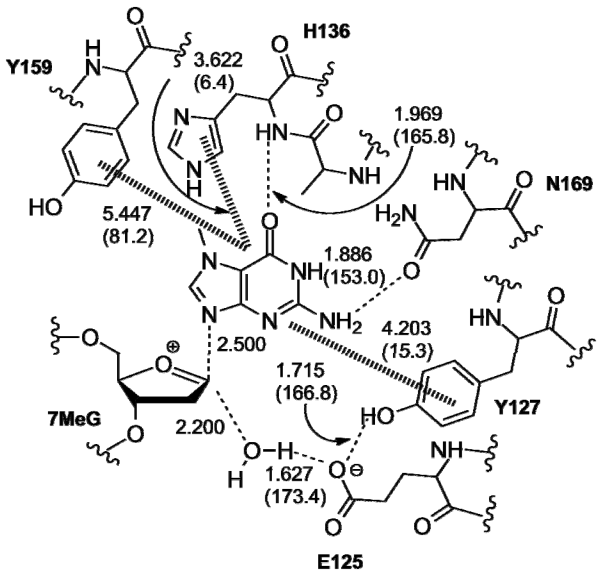


Figure B.8. Structures of constrained ONIOM(B3LYP-D3/6-31G(d):AMBER) concerted TS corresponding to AAG-mediated excision of 7MeG displayed as a line diagram, with distances in Å, angles in deg, and hashed bonds denoting π–π stacking or T-shaped interactions.

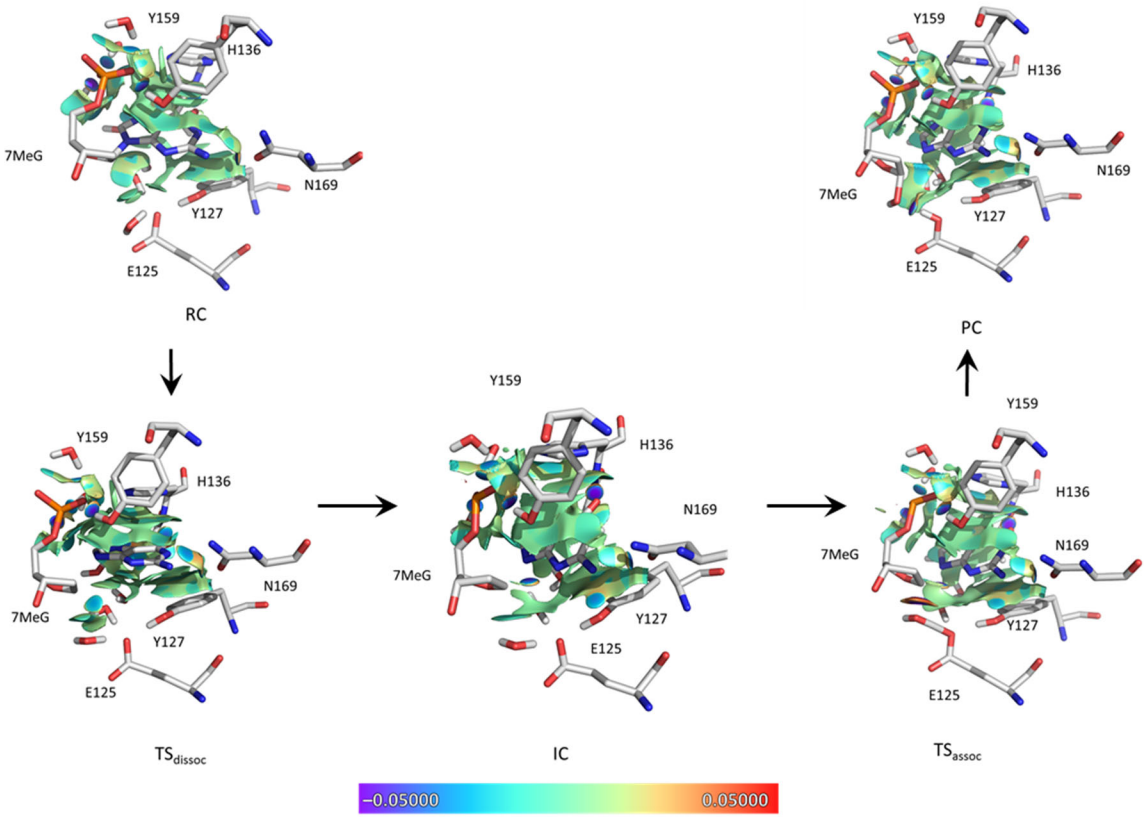


Figure B.9. NCI plots of the constrained ONIOM(B3LYP-D3/6-31G(d):AMBER) stationary points (RC, TS_{dissoc}, IC, TS_{assoc} and PC) corresponding to AAG-mediated excision of 7MeG. Promolecular densities are reported when the reduced density gradient is less than or equal to 0.3 and the color ranges from –0.05000 au (attractive interaction) to 0.0500 au (repulsive interaction).

Full Citation for Reference 65, 68, and 74

[65] Wink, P. L., Sanchez Quitian, Z. A., Rosado, L. A., Rodrigues Junior, V. d. S., Petersen, G. O., Lorenzini, D. M., Lipinski-Paes, T., Saraiva Macedo Timmers, L. F., de Souza, O. N., Basso, L. A., and Santos, D. S. (2013) Biochemical characterization of recombinant nucleoside hydrolase from Mycobacterium tuberculosis H37Rv, *Arch. Biochem. Biophys. 538*, 80-94.

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

[74] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revisions C.01 and D.01; Gaussian, Inc.: Wallingford, CT, USA, 2009.

Appendix C: Supplemental Information for Chapter 4: Hydrolytic Glycosidic Bond Cleavage in RNA Nucleosides: Effects of the 2′–Hydroxy Group and Acid–Base Catalysis

Table of Contents

Figure C.1. Selected B3LYP/6-31G(d) bond lengths and angles in reactant (RC), transition state (TS), and product complexes for the deglycosylation of adenine (A), cytosine (C), guanine (G), and uracil (U) containing RNA nucleosides facilitated by the $HCOO^-\bullet\bullet\bullet H_2O$ nucleophile without a bridging water near the 2′–hydroxy group. Interactions between the $HCOO^-$ and sugar are highlighted. 244

Table C.1. Selected B3LYP/6-31G(d) bond lengths and angles in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of adenine containing RNA nucleosides facilitated by the $HCOO^-\bullet\bullet\bullet H_2O$ nucleophile 245

Table C.2. Selected B3LYP/6-31G(d) bond lengths and angles in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of cytosine containing RNA nucleosides facilitated by the $HCOO^-\bullet\bullet\bullet H_2O$ nucleophile 246

Table C.3. Selected B3LYP/6-31G(d) bond lengths and angles in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of guanine containing RNA nucleosides facilitated by the $HCOO^-\bullet\bullet\bullet H_2O$ nucleophile 247

Table C.4. Selected B3LYP/6-31G(d) bond lengths and angles in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of uracil containing RNA nucleosides facilitated by the $HCOO^-\bullet\bullet\bullet H_2O$ nucleophile 248

Table C.5. Selected B3LYP/6-31G(d) bond lengths and angles in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of DNA nucleosides facilitated by the $HCOO^-\bullet\bullet\bullet H_2O$ nucleophile[a] 249

Table C.6. Selected B3LYP/6-31G(d) bond lengths and angles in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of adenine containing RNA nucleosides facilitated by the H_2O nucleophile 250

Table C.7. Selected B3LYP/6-31G(d) bond lengths and angles in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of cytosine containing RNA nucleosides facilitated by the H_2O nucleophile 251

Table C.8. Selected B3LYP/6-31G(d) bond lengths and angles in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of guanine containing RNA nucleosides facilitated by the H_2O nucleophile 252

Table C.9. Selected B3LYP/6-31G(d) bond lengths and angles in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of uracil containing RNA nucleosides facilitated by the H_2O nucleophile 253

Table C.10. Selected B3LYP/6-31G(d) bond lengths and angles in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of adenine containing RNA nucleosides facilitated by the OH^- nucleophile 254

Table C.11. Selected B3LYP/6-31G(d) bond lengths and angles in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of cytosine containing RNA nucleosides facilitated by the OH^- nucleophile 255

Table C.12. Selected B3LYP/6-31G(d) bond lengths and angles in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of guanine containing RNA nucleosides facilitated by the OH^- nucleophile 256

Table C.13. Selected B3LYP/6-31G(d) bond lengths and angles in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of uracil containing RNA nucleosides facilitated by the OH^- nucleophile 257

Table C.14. Calculated Gibbs Reaction Energies for the Deglycosylation of Natural and Corresponding Protonated (Acid–catalyzed) RNA Nucleosides Facilitated by Various (H_2O, OH^-, or $HCOO^-\bullet\bullet\bullet H_2O$) Nucleophiles 257

Full Citation for Reference 9, 78, and 82 259

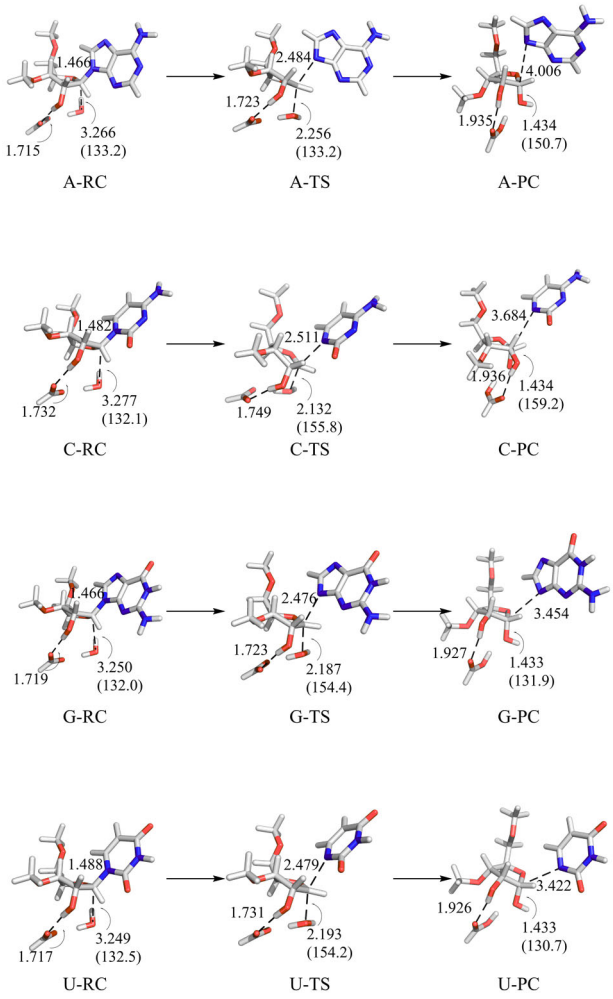


Figure C.1. Selected B3LYP/6-31G(d) bond lengths (Å) and angles (degrees in parentheses) in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of adenine (A), cytosine (C), guanine (G), and uracil (U) containing RNA nucleosides facilitated by the $HCOO^{-}\bullet\bullet\bullet H_2O$ nucleophile without a bridging water near the 2′–hydroxy group. Interactions between the $HCOO^{-}$ and sugar are highlighted.

Table C.1. Selected B3LYP/6-31G(d) bond lengths (Å) and angles (degrees) in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of adenine containing RNA nucleosides facilitated by the $HCOO^{-}\bullet\bullet\bullet H_2O$ nucleophile

		C1′–O_{wat}	C1′–N9	∠(O_{wat}C1′N9)	O_{wat}–$H_{wat}\bullet\bullet\bullet O_{HCOO-}$	∠($O_{wat}H_{wat}O_{HCOO-}$)	O_{bw}–$H_{bw}\bullet\bullet\bullet$O3′	∠($O_{bw}H_{bw}\bullet\bullet\bullet$O3′)	O2′–H2′$\bullet\bullet\bullet O_{bw}$	∠(O2′H2′O_{bw})	χ^a
A	RC	3.343	1.464	131.9	1.763	168.2	1.862	156.2	1.787	166.6	–151.6
	TS	2.260	2.449	151.6	1.639	168.4	1.906	153.8	1.790	162.4	–117.7
	PC	1.418	4.006	142.4	1.000	166.7	1.841	158.3	1.804	164.3	–102.2
A–N1H	RC	3.346	1.477	134.0	1.767	167.3	1.870	155.1	1.776	167.1	–153.0
	TS	2.316	2.412	153.4	1.652	167.9	1.916	153.4	1.780	162.6	–167.9
	PC	1.415	3.493	122.5	0.999	164.1	1.856	156.8	1.793	164.7	–139.8
A–N3H	RC	3.287	1.480	130.2	1.758	167.7	1.877	154.7	1.772	167.0	–152.0
	TS	2.415	2.433	150.2	1.669	168.6	1.920	152.7	1.766	163.1	–154.8
	PC	1.415	3.497	122.9	0.999	163.7	1.856	156.7	1.792	164.6	–139.9
A–N6H	RC	3.350	1.475	134.1	1.768	167.3	1.871	155.1	1.777	167.0	–152.5
	TS	2.309	2.408	153.3	1.645	168.4	1.914	153.4	1.781	162.7	–168.9
	PC	1.416	3.467	115.4	0.999	166.1	1.859	156.7	1.795	164.9	–131.8
A–N7H	RC	3.338	1.488	134.3	1.768	166.4	1.884	154.1	1.768	166.8	–157.5
	TS	2.429	2.396	152.0	1.674	168.7	1.927	152.1	1.770	162.5	–164.3
	PC	1.415	3.484	117.4	0.999	165.9	1.861	156.5	1.792	165.0	–134.5

[a]Dihedral angle defined as ∠(O4′C1′N9C4).

Table C.2. Selected B3LYP/6-31G(d) bond lengths (Å) and angles (degrees) in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of cytosine containing RNA nucleosides facilitated by the $HCOO^{-}\bullet\bullet\bullet H_2O$ nucleophile

		$C1'–O_{wat}$	C1′–N1	$\angle(O_{wat}C1'N1)$	$O_{wat}–H_{wat}\bullet\bullet\bullet O_{HCOO-}$	$\angle(O_{wat}H_{wat}O_{HCOO-})$	$O_{bw}–H_{bw}\bullet\bullet\bullet O3'$	$\angle(O_{bw}H_{bw}\bullet\bullet\bullet O3')$	$O2'–H2'\bullet\bullet\bullet O_{bw}$	$\angle(O2'H2'O_{bw})$	χ^a
C	RC	3.403	1.480	131.0	1.771	168.5	1.861	156.6	1.796	165.8	–159.8
	TS	2.216	2.477	150.2	1.627	169.1	1.896	154.9	1.795	162.5	–133.6
	PC	1.419	4.130	150.4	1.001	167.6	1.840	158.3	1.806	164.5	–138.3
C–N3H	RC	3.322	1.500	131.5	1.762	167.7	1.877	154.8	1.778	166.2	–159.1
	TS	2.316	2.408	152.0	1.652	168.2	1.916	153.5	1.781	162.6	–163.8
	PC	1.415	3.488	143.2	0.999	163.6	1.846	158.2	1.801	164.5	–179.9
C–N4H	RC	3.334	1.496	131.7	1.765	167.5	1.874	155.3	1.781	166.1	–161.1
	TS	2.267	2.422	152.2	1.641	168.1	1.911	154.1	1.787	162.4	–156.7
	PC	1.415	3.507	141.2	0.999	164.1	1.844	158.5	1.805	164.3	175.6
C–O2H(N1)[b]	RC	3.554	1.511	130.1	1.654	175.2	2.009	134.2	1.644	176.6	–172.6
	TS	2.524	2.454	146.8	1.687	168.9	1.939	150.9	1.752	163.4	–165.2
	PC	1.415	3.531	138.1	0.999	164.8	1.857	156.7	1.792	164.6	–162.4
C–O2H(N3)[b]	RC	3.276	1.508	128.5	1.757	167.2	1.882	154.4	1.774	166.2	–158.3
	TS	2.385	2.416	150.7	1.666	168.4	1.923	152.8	1.773	162.7	–164.1
	PC	1.416	3.489	143.0	0.999	165.7	1.850	157.5	1.797	164.5	–169.7

[a]Dihedral angle defined as ∠(O4′C1′N1C2). [b]Proton at O2 is directed toward either N1 or N3.

Table C.3. Selected B3LYP/6-31G(d) bond lengths (Å) and angles (degrees) in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of guanine containing RNA nucleosides facilitated by the $HCOO^{-}\bullet\bullet\bullet H_2O$ nucleophile

		$C1'–O_{wat}$	C1′–N9	$\angle(O_{wat}C1'N9)$	$O_{wat}–H_{wat}\bullet\bullet\bullet O_{HCOO-}$	$\angle(O_{wat}H_{wat}O_{HCOO-})$	$O_{bw}–H_{bw}\bullet\bullet\bullet O3'$	$\angle(O_{bw}H_{bw}\bullet\bullet\bullet O3')$	$O2'–H2'\bullet\bullet\bullet O_{bw}$	$\angle(O2'H2'O_{bw})$	χ^a
G	RC	3.320	1.463	130.1	1.766	168.1	1.863	156.1	1.788	166.2	–150.5
	TS	2.235	2.451	151.8	1.630	168.5	1.903	154.2	1.791	162.6	–128.3
	PC	1.417	4.049	147.5	0.999	164.0	1.840	158.2	1.800	164.7	–124.6
G–N2H	RC	3.291	1.473	129.8	1.760	167.8	1.872	155.2	1.779	166.3	–150.9
	TS	2.283	2.413	153.2	1.638	168.5	1.910	153.6	1.783	162.7	–162.5
	PC	1.415	3.464	139.4	0.999	164.7	1.849	157.5	1.799	164.5	–169.1
G–N3H	RC	3.228	1.477	126.2	1.748	167.1	1.876	154.7	1.774	166.2	–147.4
	TS	2.414	2.433	150.1	1.670	168.4	1.922	152.3	1.770	162.8	–151.3
	PC	1.415	3.471	133.0	0.999	165.0	1.855	156.9	1.794	164.8	–149.3
G–N7H	RC	1.414	3.492	119.6	0.999	165.5	1.858	156.7	1.792	164.4	–118.8
	TS	3.346	1.486	134.5	1.766	167.3	1.881	154.3	1.770	167.4	–157.3
	PC	2.410	2.391	152.0	1.667	168.9	1.922	152.5	1.774	162.3	–164.1
G–O6H(N1)[b]	RC	3.332	1.475	132.2	1.767	167.5	1.873	155.0	1.778	166.7	–152.2
	TS	2.319	2.412	153.0	1.647	168.6	1.913	153.3	1.781	162.5	–164.3
	PC	1.415	3.478	117.1	0.999	165.5	1.858	156.8	1.794	164.6	–131.8
G–O6H(N7)[b]	RC	3.338	1.474	132.6	1.767	167.5	1.872	155.0	1.777	166.7	–152.3
	TS	2.313	2.410	153.1	1.646	168.6	1.913	153.3	1.782	162.5	–163.2
	PC	1.415	3.474	117.1	0.999	165.6	1.858	156.7	1.794	164.6	–133.0

[a]Dihedral angle defined as ∠(O4′C1′N9C4). [b]Proton at O6 is directed toward either N1 or N7.

Table C.4. Selected B3LYP/6-31G(d) bond lengths (Å) and angles (degrees) in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of uracil containing RNA nucleosides facilitated by the $HCOO^{-}\bullet\bullet\bullet H_2O$ nucleophile

		$C1'–O_{wat}$	C1′–N1	$\angle(O_{wat}C1'N1)$	$O_{wat}–H_{wat}\bullet\bullet\bullet O_{HCOO-}$	$\angle(O_{wat}H_{wat}O_{HCOO-})$	$O_{bw}–H_{bw}\bullet\bullet\bullet O3'$	$\angle(O_{bw}H_{bw}\bullet\bullet\bullet O3')$	$O2'–H2'\bullet\bullet\bullet O_{bw}$	$\angle(O2'H2'O_{bw})$	χ^a
U	RC	3.334	1.486	128.3	1.764	167.1	1.866	156.0	1.789	166.0	–156.7
	TS	2.189	2.426	153.1	1.614	168.5	1.900	155.1	1.796	162.4	–166.6
	PC	1.416	3.532	140.2	1.000	165.3	1.840	159.1	1.808	163.7	167.7
U–O2H(N1)[b]	RC	3.196	1.515	119.9	1.738	175.9	1.969	146.9	1.671	155.4	–173.7
	TS	2.716	2.465	140.8	1.708	170.5	1.948	149.7	1.738	162.9	–161.5
	PC	1.410	3.602	135.1	1.003	154.8	1.857	156.3	1.789	165.2	–155.5
U–O2H(N3)[b]	RC	3.250	1.520	128.2	1.754	167.5	1.886	153.8	1.768	166.4	–156.0
	TS	2.510	2.417	148.4	1.687	169.0	1.932	151.9	1.761	162.7	–163.4
	PC	1.414	3.533	138.2	0.999	164.2	1.858	156.7	1.792	164.5	–160.8
U–O4H(N3)[b]	RC	3.307	1.507	131.4	1.760	167.7	1.882	154.5	1.773	166.2	–160.4
	TS	2.375	2.403	151.5	1.665	168.4	1.919	153.1	1.776	162.4	–168.2
	PC	1.417	3.442	145.0	1.000	165.3	1.849	158.0	1.800	164.4	–176.1
U–O4H(C5)[b]	RC	3.313	1.507	131.4	1.761	167.6	1.881	154.4	1.772	166.3	–159.9
	TS	2.378	2.409	151.1	1.665	168.4	1.923	152.9	1.775	162.6	–164.7
	PC	1.416	3.463	144.1	0.999	164.5	1.847	158.0	1.800	164.5	–178.6

[a]Dihedral angle defined as ∠(O4′C1′N1C2). [b]Proton at O2 directed toward N1 or N3 and proton at O4 directed toward either N3 or C5.

Table C.5. Selected B3LYP/6-31G(d) bond lengths (Å) and angles (degrees) in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of DNA nucleosides facilitated by the $HCOO^{-}\bullet\bullet\bullet H_2O$ nucleophile[a]

		C1′–O_{wat}	C1′–N1/N9[b]	∠(O_{wat}C1′N1/N9)[b]	O_{wat}–$H_{wat}\bullet\bullet\bullet O_{HCOO-}$	∠($O_{wat}H_{wat}O_{HCOO-}$)	χ[c]
A	RC	3.225	1.471	135.6	1.756	172.6	–150.1
	TS	2.238	2.555	151.4	1.613	175.4	160.5
	PC	1.441	3.532	122.7	1.003	173.6	105.0
C	RC	3.262	1.486	133.3	1.765	172.1	–158.9
	TS	2.135	2.572	153.2	1.573	175.3	169.9
	PC	1.438	3.479	154.2	1.003	173.0	177.0
G	RC	3.243	1.469	131.1	1.761	172.0	–145.8
	TS	2.222	2.558	151.5	1.608	175.5	158.4
	PC	1.441	3.501	119.7	1.002	173.8	104.3
U	RC	3.244	1.492	131.5	1.761	172.2	–155.0
	TS	2.204	2.546	153.4	1.599	175.3	172.8
	PC	1.438	3.518	155.4	1.003	172.9	176.3

[a]Structures obtained from reference 35. [b]Distance or reaction angle for pyrimidine (N1) or purine (N9) nucleosides. [c]Dihedral angle defined as ∠(O4′C1′N1C2) for pyrimidine or ∠(O4′C1′N9C4) for purine nucleosides.

Table C.6. Selected B3LYP/6-31G(d) bond lengths (Å) and angles (degrees) in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of adenine containing RNA nucleosides facilitated by the H_2O nucleophile

		C1′–O_{wat}	C1′–N9	∠(O_{wat}C1′N9)	O_{wat}–H_{wat}•••O2′	∠($O_{wat}H_{wat}$O2′)	O_{bw}–H_{bw}•••O3′	∠($O_{bw}H_{bw}$•••O3′)	O2′–H2′•••O_{bw}	∠(O2′H2′O_{bw})	χ^a
A	RC	3.458	1.468	124.6	1.865	168.2	1.877	153.3	1.743	165.3	−155.1
	TS	2.091	2.677	148.1	1.741	132.4	1.942	149.5	1.703	162.8	−112.0
	PC	1.574	3.404	142.2	1.542	131.7	1.931	151.3	1.669	157.9	−103.1
A–N1H	RC	3.409	1.481	123.9	1.870	166.3	1.888	152.5	1.732	165.9	−157.2
	TS	2.211	2.617	147.4	1.777	133.2	1.959	148.1	1.699	163.5	−115.0
	PC	1.568	3.408	143.2	1.535	131.8	1.943	150.1	1.659	158.3	−102.9
A–N3H	RC	3.329	1.478	100.2	1.822	163.6	1.894	150.7	1.710	169.3	−156.2
	TS	2.240	2.575	148.2	1.788	133.1	1.960	148.0	1.694	163.9	−155.8
	PC	1.563	3.133	144.7	1.539	131.5	1.961	148.7	1.652	159.1	−159.9
A–N6H	RC	3.410	1.480	123.4	1.869	166.4	1.889	152.4	1.733	165.8	−157.1
	TS	2.210	2.613	147.6	1.776	133.2	1.957	148.3	1.700	163.4	−116.5
	PC	1.568	3.447	142.1	1.537	131.7	1.939	150.5	1.661	158.2	−104.1
A–N7H	RC	3.344	1.492	124.6	1.880	164.2	1.901	151.5	1.726	165.8	−161.5
	TS	2.283	2.560	149.6	1.795	134.0	1.980	146.6	1.698	163.4	−130.7
	PC	1.568	3.496	143.2	1.540	131.6	1.944	150.0	1.659	158.6	−105.2

[a]Dihedral angle defined as ∠(O4′C1′N9C4).

Table C.7. Selected B3LYP/6-31G(d) bond lengths (Å) and angles (degrees) in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of cytosine containing RNA nucleosides facilitated by the H_2O nucleophile

		C1′–O_{wat}	C1′–N1	∠(O_{wat}C1′N1)	O_{wat}–H_{wat}•••O2′	∠($O_{wat}H_{wat}$O2′)	O_{bw}–H_{bw}•••O3′	∠($O_{bw}H_{bw}$•••O3′)	O2′–H2′•••O_{bw}	∠(O2′H2′O_{bw})	χ^a
C	RC	3.489	1.483	129.2	1.863	169.0	1.873	154.2	1.751	164.8	–162.7
	TS	1.948	2.674	149.6	1.692	132.0	1.932	151.4	1.715	160.3	177.5
	PC	1.570	3.081	145.7	1.516	132.8	1.940	151.8	1.682	156.0	166.1
C–N3H	RC	3.394	1.505	127.8	1.875	166.1	1.891	152.5	1.734	165.3	–162.7
	TS	2.148	2.600	150.0	1.758	132.6	1.955	149.0	1.705	162.2	179.5
	PC	1.567	3.111	149.6	1.525	132.2	1.952	150.0	1.661	157.9	175.6
C–N4H	RC	3.415	1.500	128.4	1.872	166.8	1.889	152.8	1.738	164.9	–163.1
	TS	2.142	2.649	147.9	1.762	132.3	1.960	148.2	1.703	163.0	–137.7
	PC	1.576	3.302	142.4	1.555	131.2	1.941	150.4	1.667	158.6	–101.9
C–O2H(N1)[b]	RC	3.203	1.505	103.7	1.740	158.1	1.909	148.7	1.695	170.2	–155.4
	TS	2.373	2.574	145.6	1.825	135.0	1.982	146.0	1.683	165.3	–166.2
	PC	1.556	3.256	141.6	1.538	131.3	1.964	148.1	1.647	159.2	–166.5
C–O2H(N3)[b]	RC	3.369	1.513	127.2	1.878	165.1	1.896	152.0	1.728	165.6	–161.7
	TS	2.212	2.588	148.1	1.780	132.9	1.965	147.9	1.699	163.1	–167.4
	PC	1.564	3.130	148.7	1.533	131.6	1.955	149.3	1.655	158.7	–176.8

[a]Dihedral angle defined as ∠(O4′C1′N1C2). [b]Proton at O2 is directed toward either N1 or N3.

Table C.8. Selected B3LYP/6-31G(d) bond lengths (Å) and angles (degrees) in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of guanine containing RNA nucleosides facilitated by the H_2O nucleophile

		$C1'–O_{wat}$	C1′–N9	∠(O_{wat}C1′N9)	$O_{wat}–H_{wat}•••O2'$	∠($O_{wat}H_{wat}$O2′)	$O_{bw}–H_{bw}•••O3'$	∠($O_{bw}H_{bw}•••O3'$)	O2′–H2′•••O_{bw}	∠(O2′H2′O_{bw})	χ^a
G	RC	3.485	1.467	124.8	1.865	169.0	1.879	153.2	1.743	165.6	–153.3
	TS	2.022	2.648	148.9	1.720	131.9	1.937	150.3	1.707	161.9	166.0
	PC	1.564	3.315	104.9	1.541	132.4	2.000	149.2	1.754	149.2	129.0
G–N2H	RC	3.434	1.478	123.1	1.869	167.4	1.888	152.3	1.734	166.2	–153.5
	TS	2.100	2.606	149.9	1.748	131.9	1.950	148.9	1.700	163.1	–170.9
	PC	1.568	3.082	147.1	1.538	131.6	1.943	150.0	1.658	158.6	–174.1
G–N3H	RC	3.364	1.477	100.8	1.817	162.5	1.893	150.6	1.708	169.9	–157.7
	TS	2.237	2.587	147.6	1.791	132.9	1.960	147.7	1.695	163.9	–158.2
	PC	1.562	3.129	144.2	1.542	131.2	1.951	149.3	1.652	159.4	–162.4
G–N7H	RC	3.365	1.490	124.4	1.878	164.9	1.903	151.3	1.726	166.0	–160.0
	TS	2.227	2.547	150.1	1.778	133.5	1.968	147.8	1.699	163.5	–174.6
	PC	1.562	3.114	148.8	1.534	131.5	1.955	149.1	1.650	159.2	–160.4
G–O6H(N1)[b]	RC	3.412	1.478	124.3	1.873	166.5	1.888	152.2	1.734	165.9	–153.7
	TS	2.133	2.593	149.9	1.756	132.3	1.955	148.6	1.700	163.1	–174.0
	PC	1.565	3.090	146.6	1.541	131.4	1.951	149.4	1.657	158.6	–172.4
G–O6H(N7)[b]	RC	3.429	1.477	124.2	1.872	166.9	1.887	152.2	1.735	165.9	–152.0
	TS	2.128	2.591	150.2	1.754	132.2	1.955	148.8	1.700	163.2	–176.5
	PC	1.565	3.097	147.0	1.542	131.3	1.947	149.6	1.657	158.6	–169.1

[a]Dihedral angle defined as ∠(O4′C1′N9C4). [b]Proton at O6 is directed toward either N1 or N7.

Table C.9. Selected B3LYP/6-31G(d) bond lengths (Å) and angles (degrees) in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of uracil containing RNA nucleosides facilitated by the H_2O nucleophile

		$C1'–O_{wat}$	$C1'–N1$	$\angle(O_{wat}C1'N1)$	$O_{wat}–H_{wat}\bullet\bullet\bullet O2'$	$\angle(O_{wat}H_{wat}O2')$	$O_{bw}–H_{bw}\bullet\bullet\bullet O3'$	$\angle(O_{bw}H_{bw}\bullet\bullet\bullet O3')$	$O2'–H2'\bullet\bullet\bullet O_{bw}$	$\angle(O2'H2'O_{bw})$	χ^a
U	RC	3.466	1.491	128.3	1.867	168.3	1.880	153.5	1.744	165.2	–161.0
	TS	2.013	2.650	150.1	1.716	131.9	1.940	150.4	1.710	161.2	175.8
	PC	1.568	3.094	146.8	1.520	132.6	1.941	151.3	1.672	156.7	168.1
U–O2H(N1)[b]	RC	3.457	1.484	107.3	1.080	170.9	1.955	134.1	1.436	173.5	–157.5
	TS	2.525	2.510	144.0	1.857	138.5	1.988	145.5	1.678	165.4	–165.1
	PC	1.555	3.269	142.2	1.537	131.3	1.967	148.0	1.646	159.2	–163.4
U–O2H(N3)[b]	RC	3.345	1.526	125.1	1.883	163.8	1.903	151.2	1.722	165.9	–158.9
	TS	2.337	2.551	147.3	1.810	134.8	1.977	146.8	1.694	163.7	–165.2
	PC	1.563	3.149	149.2	1.536	131.5	1.963	148.6	1.651	159.0	–172.3
U–O4H(N3)[b]	RC	3.371	1.512	127.6	1.878	165.1	1.897	151.8	1.729	165.3	–163.0
	TS	2.209	2.581	148.7	1.777	133.0	1.961	148.3	1.702	162.8	–169.6
	PC	1.565	3.132	150.0	1.527	132.0	1.951	149.9	1.658	158.1	174.9
U–O4H(C5)[b]	RC	3.368	1.512	127.6	1.877	165.1	1.896	151.9	1.728	165.5	–163.2
	TS	2.209	2.579	148.8	1.777	133.0	1.962	148.2	1.701	162.8	–168.8
	PC	1.565	3.129	150.1	1.528	131.9	1.951	149.9	1.658	158.3	176.3

[a]Dihedral angle defined as ∠(O4′C1′N1C2). [b]Proton at O2 is directed toward either N1 or N3 and proton at O4 is directed toward either N3 or C5.

Table C.10. Selected B3LYP/6-31G(d) bond lengths (Å) and angles (degrees) in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of adenine containing RNA nucleosides facilitated by the OH^- nucleophile

		C1′–O_{wat}	C1′–N9	∠(O_{wat}C1′N9)	O_{bw}–H_{bw}•••O3′	∠($O_{bw}H_{bw}$•••O3′)	O2′–H2′•••O_{bw}	∠(O2′H2′O_{bw})	χ[a]
A	RC	2.993	1.452	127.7	1.850	158.4	1.835	158.1	–136.6
	TS	2.222	1.932	160.6	1.870	161.1	1.873	154.5	–151.3
	PC	1.411	3.506	134.4	1.881	156.7	1.849	153.8	–175.0
A–N1H	RC	2.879	1.460	126.6	1.858	158.0	1.829	158.1	–142.1
	TS	2.288	1.895	159.2	1.876	159.9	1.860	155.2	–152.3
	PC	1.409	3.557	138.8	1.880	156.5	1.845	154.1	–166.3
A–N3H	RC	2.807	1.467	123.1	1.887	152.8	1.750	160.2	165.3
	TS	2.288	1.895	154.8	1.872	155.3	1.762	160.0	173.7
	PC	1.408	3.613	112.8	1.902	151.4	1.758	159.7	149.1
A–N6H	RC	2.883	1.460	126.6	1.858	158.1	1.830	158.0	–142.7
	TS	2.282	1.894	159.3	1.876	160.0	1.862	155.1	–151.6
	PC	1.409	3.593	139.3	1.886	155.9	1.844	154.5	–128.4
A–N7H	RC	2.791	1.469	124.9	1.867	157.4	1.826	158.2	–155.3
	TS	2.340	1.848	159.2	1.880	158.9	1.850	155.6	–157.8
	PC	1.407	3.609	144.7	1.883	156.4	1.840	154.8	–135.3

[a]Dihedral angle defined as ∠(O4′C1′N9C4).

Table C.11. Selected B3LYP/6-31G(d) bond lengths (Å) and angles (degrees) in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of cytosine containing RNA nucleosides facilitated by the OH^- nucleophile

		C1′–O_{wat}	C1′–N1	∠(O_{wat}C1′N1)	O_{bw}–H_{bw}•••O3′	∠($O_{bw}H_{bw}$•••O3′)	O2′–H2′•••O_{bw}	∠(O2′H2′O_{bw})	χ^a
C	RC	3.104	1.479	130.0	1.849	158.2	1.837	158.2	–134.7
	TS	2.210	1.891	160.8	1.862	161.5	1.880	154.0	–153.8
	PC	1.411	3.565	137.5	1.883	156.6	1.853	153.4	–151.7
C–N3H	RC	2.919	1.478	128.6	1.861	157.2	1.824	158.5	–144.0
	TS	2.315	1.909	158.2	1.875	159.6	1.857	155.3	–155.2
	PC	1.409	3.523	139.8	1.885	156.1	1.843	154.4	–153.5
C–N4H	RC	2.945	1.481	129.4	1.858	157.4	1.826	158.6	–147.5
	TS	2.290	1.889	159.0	1.871	160.2	1.863	154.9	–155.3
	PC	1.410	3.736	137.3	1.885	156.2	1.847	153.9	–149.4
C–O2H(N1)[b]	RC	2.736	1.486	121.3	1.910	152.8	1.736	156.5	–172.9
	TS	2.296	1.879	154.5	1.894	155.5	1.761	154.7	–171.9
	PC	1.407	3.582	126.3	1.902	151.9	1.765	159.0	177.2
C–O2H(N3)[b]	RC	2.844	1.462	127.4	1.866	157.1	1.824	158.2	–147.1
	TS	2.352	1.961	157.1	1.878	159.0	1.850	155.7	–154.5
	PC	1.409	3.451	140.1	1.886	156.1	1.842	154.4	–156.1

[a]Dihedral angle defined as ∠(O4′C1′N1C2). [b]Proton at O2 is directed toward either N1 or N3.

Table C.12. Selected B3LYP/6-31G(d) bond lengths (Å) and angles (degrees) in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of guanine containing RNA nucleosides facilitated by the OH^- nucleophile

		C1′–O_{wat}	C1′–N9	∠(O_{wat}C1′N9)	O_{bw}–H_{bw}•••O3′	∠($O_{bw}H_{bw}$•••O3′)	O2′–H2′•••O_{bw}	∠(O2′H2′O_{bw})	χ^a
G	RC	3.018	1.451	127.9	1.851	158.4	1.836	158.1	–132.9
	TS	2.215	1.935	161.0	1.866	161.2	1.876	154.4	–150.7
	PC	1.410	3.589	143.3	1.880	156.9	1.849	153.8	–157.1
G–N2H	RC	2.897	1.458	125.1	1.860	158.0	1.830	157.8	–142.0
	TS	2.273	1.911	159.7	1.869	160.2	1.860	155.0	–154.4
	PC	1.409	3.584	144.7	1.880	156.9	1.847	154.0	–151.7
G–N3H	RC	2.815	1.467	123.6	1.884	152.6	1.742	160.0	168.5
	TS	2.270	1.892	155.7	1.867	155.3	1.755	159.7	175.1
	PC	1.422	5.713	95.4	1.909	150.3	1.738	160.3	–172.8
G–N7H	RC	2.805	1.469	126.1	1.868	157.3	1.827	158.3	–150.7
	TS	2.333	1.859	158.6	1.877	159.2	1.851	155.5	–154.7
	PC	1.408	3.536	143.2	1.883	156.5	1.841	154.5	–170.9
G–O6H(N1)[b]	RC	2.888	1.459	125.3	1.863	157.7	1.830	157.9	–140.9
	TS	2.287	1.898	159.2	1.871	160.0	1.857	155.1	–152.2
	PC	1.409	3.537	138.6	1.884	156.3	1.844	154.2	–171.4
G–O6H(N7)[b]	RC	2.893	1.458	125.6	1.863	157.7	1.830	158.0	–141.0
	TS	2.285	1.900	159.3	1.871	160.0	1.858	155.1	–152.2
	PC	1.410	3.523	138.8	1.885	156.3	1.845	154.3	–172.0

[a]Dihedral angle defined as ∠(O4′C1′N9C4). [b]Proton at O6 is directed toward either N1 or N7.

Table C.13. Selected B3LYP/6-31G(d) bond lengths (Å) and angles (degrees) in reactant (RC), transition state (TS), and product (PC) complexes for the deglycosylation of uracil containing RNA nucleosides facilitated by the OH^- nucleophile

		C1′–O_{wat}	C1′–N1	∠(O_{wat}C1′N1)	O_{bw}–H_{bw}•••O3′	∠($O_{bw}H_{bw}$•••O3′)	O2′–H2′•••O_{bw}	∠(O2′H2′O_{bw})	χ^a
U	RC	3.018	1.466	129.9	1.852	158.2	1.834	158.3	−135.3
	TS	2.244	1.937	160.0	1.866	160.8	1.872	154.4	−153.2
	PC	1.410	3.471	137.6	1.883	156.6	1.850	153.8	−156.9
U–O2H(N1)[b]	RC	2.701	1.488	120.4	1.932	150.0	1.694	157.2	−179.4
	TS	2.333	1.870	152.0	1.905	153.4	1.724	156.2	−176.6
	PC	1.406	3.738	125.9	1.908	150.4	1.739	160.4	176.9
U–O2H(N3)[b]	RC	2.792	1.493	126.1	1.872	156.8	1.821	157.9	−145.9
	TS	2.404	1.850	155.1	1.882	158.3	1.840	156.4	−152.7
	PC	1.408	3.583	144.7	1.881	156.8	1.843	154.5	−143.1
U–O4H(N3)[b]	RC	2.813	1.484	125.9	1.867	157.1	1.825	158.4	−153.2
	TS	2.345	1.870	157.5	1.878	159.2	1.851	155.5	−155.7
	PC	1.408	3.567	144.2	1.882	156.6	1.843	154.5	−148.4
U–O4H(C5)[b]	RC	2.808	1.484	126.3	1.867	157.1	1.825	158.3	−153.6
	TS	2.345	1.869	157.5	1.879	159.2	1.851	155.5	−155.9
	PC	1.409	3.566	140.7	1.886	156.1	1.843	154.3	−150.3

[a]Dihedral angle defined as ∠(O4′C1′N1C2). [b]Proton at O2 is directed toward either N1 or N3, and proton at O4 is directed toward either N3 or C5.

Table C.14. Calculated Gibbs Reaction Energies (kJ mol^{-1}) for the Deglycosylation of Natural and Corresponding Protonated (Acid–catalyzed) RNA Nucleosides Facilitated by Various (H_2O, OH^-, or $HCOO^-\bullet\bullet\bullet H_2O$) Nucleophiles[a]

	H_2O		OH^-						$HCOO^-\bullet\bullet\bullet H_2O$					
nucleoside	rxn	Δ_{acid}[b]	rxn	Δ_{base}[c]	Δ_{acid}[b]	Δ_{sum}[d]	$\Delta_{acid-base}$[e]	$\Delta_{sum}-\Delta_{acid-base}$[f]	rxn	Δ_{base}[c]	Δ_{acid}[b]	Δ_{sum}[d]	$\Delta_{acid-base}$[e]	$\Delta_{sum}-\Delta_{acid-base}$[f]
A	183.3	–	–59.9	–243.2	–	–243.2	–243.2	–	79.8	–103.5	–	–103.5	–103.5	–
A–N1H	148.2	–35.1	–99.1	–247.3	–39.2	–278.2	–282.4	4.2	34.9	–113.3	–44.8	–138.6	–148.3	9.8
A–N3H	127.2	–56.1	–119.9	–247.1	–59.9	–299.2	–303.1	3.9	26.1	–101.1	–53.7	–159.6	–157.2	–2.4
A–N6H	150.2	–33.1	–99.4	–249.6	–39.5	–276.3	–282.7	6.5	39.6	–110.6	–40.1	–136.6	–143.7	7.1
A–N7H	115.4	–67.8	–147.1	–262.6	–87.2	–311.0	–330.4	19.4	9.0	–106.5	–70.8	–171.4	–174.3	3.0
C	190.5	–	–48.9	–239.4	–	–239.4	–239.4	–	98.1	–92.4	–	–92.4	–92.4	–
C–N3H	127.0	–63.4	–114.8	–241.8	–65.9	–302.8	–305.2	2.4	26.7	–100.4	–71.4	–155.8	–163.8	8.0
C–N4H	144.2	–46.3	–104.5	–248.7	–55.6	–285.7	–295.0	9.3	42.8	–101.3	–55.3	–138.7	–147.6	8.9
C–O2H(N1)	117.9	–72.5	–143.5	–261.4	–94.6	–311.9	–334.0	22.0	12.3	–105.7	–85.9	–164.9	–178.2	13.3
C–O2H(N3)	116.0	–74.5	–132.5	–248.5	–83.6	–313.8	–323.0	9.1	10.5	–105.5	–87.6	–166.8	–180.0	13.2
G	182.6	–	–55.1	–237.7	–	–237.7	–237.7	–	79.0	–103.6	–	–103.6	–103.6	–
G–N2H	149.3	–33.3	–93.0	–242.3	–38.0	–270.9	–275.6	4.7	44.5	–104.8	–34.5	–136.9	–138.1	1.2
G–N3H	145.0	–37.6	–128.8	–273.8	–73.7	–275.3	–311.4	36.1	19.9	–125.1	–59.1	–141.2	–162.7	21.5
G–N7H	113.9	–68.7	–139.7	–253.6	–84.6	–306.3	–322.3	16.0	13.6	–100.3	–65.4	–172.3	–169.0	–3.3
G–O6H(N1)	146.9	–35.6	–87.2	–234.2	–32.1	–273.3	–269.8	–3.5	41.3	–105.6	–37.7	–139.2	–141.3	2.1
G–O6H(N7)	154.2	–28.4	–93.7	–247.9	–38.6	–266.0	–276.3	10.2	44.4	–109.8	–34.6	–131.9	–138.1	6.2
U	155.9	–	–87.5	–243.4	–	–243.4	–243.4	–	55.7	–100.2	–	–100.2	–100.2	–
U–O2H(N1)	92.1	–63.8	–161.4	–253.5	–73.9	–307.3	–317.3	10.1	–27.9	–120.0	–83.6	–164.1	–183.8	19.8
U–O2H(N3)	83.0	–72.9	–162.5	–245.4	–74.9	–316.4	–318.4	2.0	–20.6	–103.6	–76.3	–173.2	–176.6	3.4
U–O4H(N3)	110.6	–45.3	–141.4	–252.1	–53.9	–288.7	–297.3	8.6	6.0	–104.6	–49.6	–145.5	–149.9	4.4
U–O4H(C5)	106.5	–49.4	–137.1	–243.6	–49.6	–292.8	–293.0	0.2	6.6	–99.9	–49.1	–149.6	–149.3	–0.3

[a]Relative Gibbs energies obtained from SMD-B3LYP-D3/6-311+G(2df,2p)//PCM-B3LYP/6-31G(d) calculations, including scaled (0.9806) zero-point vibrational and thermal corrections. [b]The effect of acid catalysis on RNA nucleoside deglycosylation reaction energies for a given nucleophile. [c]The effect of base catalysis on RNA nucleoside deglycosylation reaction energies. [d]Sum of reaction energy reductions afforded by acid (Δ_{acid}, H_2O nucleophile) and base (Δ_{base}(OH^- or $HCOO^-$) catalysis on RNA nucleoside deglycosylation reaction energies. [e]The simultaneous effect of acid and base catalysis calculated using models that contain both water (base) and nucleobase (acid) activation. [f]The additivity of acid and base catalysis calculated as the difference between Δ_{sum} and $\Delta_{acid-base}$.

Full Citation for Reference 9, 78, and 82

9. Zhu, Y.; Dai, J.; Zhang, T.; Li, X.; Fang, P.; Wang, H.; Jiang, Y.; Yu, X.; Xia, T.; Niu, L.; Guo, Y.; Teng, M., Structural Insights into the Neutralization Mechanism of Monoclonal Antibody 6c2 against Ricin. *J. Biol. Chem.* **2013,** *288* (35), 25165–25172.

78. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revisions A.02, C.01, and D.01; Gaussian, Inc.: Wallingford, CT, USA, 2009.

82. Wink, P. L.; Sanchez Quitian, Z. A.; Rosado, L. A.; Rodrigues Junior, V. d. S.; Petersen, G. O.; Lorenzini, D. M.; Lipinski-Paes, T.; Saraiva Macedo Timmers, L. F.; de Souza, O. N.; Basso, L. A.; Santos, D. S., Biochemical Characterization of Recombinant Nucleoside Hydrolase from Mycobacterium Tuberculosis H37rv. *Arch. Biochem. Biophys.* **2013,** *538* (2), 80–94.

Appendix D: Supplemental Information for Chapter 5: Structural Explanation for the Tunable Substrate Specificity of Nucleoside Hydrolases: Insights from Molecular Dynamics Simulations

Table of Contents

Discussion of wild-type CU-NH (Asp15 His239) bound to inosine. 262

Discussion of CU-NH single mutants ($Asp15^-$ $His239^+$) bound to inosine. 262

Table D.1. RESP partial charges and atom types of the calcium–ligating residues in the CU–NH active site. 263

Table D.2. GAFF and MCPB.py generated bond and angle parameters for the CU-NH active site..... 265

Table D.3. Average active site root-mean-square deviation (rmsd) (standard deviation (σ) provided in parentheses) of the CU-NH–I complex over each of three 100 ns replicate trajectories relative to the representative structure obtained by clustering replicate 1.[a]. 265

Table D.4. Summary of important average distances (standard deviation (σ) provided in parentheses) in the active site from MD simulations on wild-type and mutant CU-NH bound to inosine.. 265

Table D.5. Summary of occupancies and average geometrical parameters for important hydrogen bonds in the active site from MD simulations on wild-type and mutant CU-NH bound to inosine (I) with anionic Asp15 and cationic His239. 269

Table D.6. Average backbone and side chain torsional angles for key active site residues from MD simulations on wild-type and mutant CU-NH with anionic Asp15 and cationic His239. 270

Table D.7. Average pairwise MM–GBSA free energy and standard deviations between inosine and key active site amino acids over the MD simulation. 272

Table D.8. Summary of occupancies and average geometrical parameters for important hydrogen bonds in the active site during MD simulations on wild-type and mutant CU-NH bound to inosine (I) with neutral Asp15 and neutral His239.. 272

Table D.9. Average backbone and side chain torsional angles for key active site residues from MD simulations on wild-type and mutant CU-NH bound to inosine with neutral Asp15 and neutral His239. 274

Figure D.1. Structure and numbering of the calcium-ligating residues in the CU-NH active site..... 276

Figure D.2. Overlay of X-ray crystal structure and MD representative structure of wild-type CU-NH bound to inosine, highlighting active site residues that surround the ribose sugar and Ca^{2+} or the inosine nucleobase. 277

Figure D.3. Stacked histogram displaying the occupancy of the pseudorotational angle of the ribose moiety of inosine bound within the wild-type or mutant CU-NH active site with cationic His239 and anionic Asp15 ($His239^+$ $Asp15^-$) or neutral His239 and neutral Asp15 (His239 Asp15)..... 278

Figure D.4. Overlay of MD representative structures of wild-type CU-NH ($Asp15^-$ $His239^+$) and Thr223Tyr Gln227Tyr CU-NH ($Asp15^-$ $His239^+$) bound to inosine, highlighting the active site residues that surround the ribose sugar and Ca^{2+} or the inosine nucleobase. 279

Figure D.5. Overlay of MD representative structures of wild-type CU-NH (Asp15 His239) and Thr223Tyr Gln227Tyr CU-NH (Asp15 His239) bound to inosine, highlighting the active site residues that surround the ribose sugar and Ca^{2+} or the inosine nucleobase. 280

Figure D.6. Representative MD structures highlighting the active site residues that surround the ribose moiety and Ca^{2+} for the Thr223Tyr CU-NH ($Asp15^{-}$ $His239^{+}$), Gln227Tyr CU-NH ($Asp15^{-}$ $His239^{+}$), Thr223Tyr CU-NH–I complex (Asp15 His239), or Gln227Tyr CU-NH (Asp15 His239) bound to inosine. 281

Figure D.7. Representative MD structures of Thr223Tyr CU-NH (Asp15 His239) bound to inosine, highlighting hydrogen bonds between Asp15, Tyr223, and the inosine nucleobase. Hydrogen-bonding time series plot, highlighting the lack of simultaneous I(N7)···Tyr227(OH) and Tyr223(O_H)···Asp15(OδH) hydrogen bonds. 282

Figure D.8. Representative MD structures of wild-type CU-NH (Asp15 His239) bound to inosine, highlighting the active site residues that surround the ribose sugar and Ca^{2+} or the inosine nucleobase, and the distribution of active site water. Overlay of the active site from MD representative structures of wild-type CU-NH ($Asp15^{-}$ $His239^{+}$) and wild-type CU-NH (Asp15 His239) bound to inosine..... 283

Full citation for references 6, 20, 22, 26, 49, and 52..... 284

References 284

Discussion of wild-type CU-NH (Asp15 His239) bound to inosine.

As discussed in the main text, we performed simulations on CU–NH bound to inosine with neutral Asp15 and His239. In general, the positions of active site residues do not change considerably upon altering the protonation states of Asp15 and His239 (Figure D.8). Indeed, the average backbone dihedral angles of active site residues surrounding the nucleobase fall within ~23° (Tables D.6 and D.7). The Ca^{2+} coordination sphere (Figure D.8a) is consistent with wild-type CU-NH with cationic Asp15 and anionic His239, and the pucker of the ribose sugar is C4′ endo (Figure D.3). Both the I(O6)···Gln227(NεH) and Tyr231(Oη)···I(N1H) hydrogen bonds are occupied for the majority of the trajectory (occupancy > 79.1 %; Table D.8), while the His82 imidazole group forms a T-shaped interaction with the inosine nucleobase that is likely important for substrate binding (distance = 5.248 Å; and interaction energy = –9.7 kJ mol^{-1}; Tables D.4 and D.7). No hydrogen bond forms between N7 of inosine and Asp15, and no water chain links Asp15 and inosine (Figure D.8b). Thus, when combined with the data on wild-type CU-NH with anionic Asp15 and cationic $His239^{+}$, our data suggests that proton transfer to the departing nucleobase during hydrolysis catalyzed by wild-type CU-NH is unlikely.

Discussion of CU-NH single mutants ($Asp15^{-}$ $His239^{+}$) bound to inosine.

Our simulations suggest that the Thr223Tyr and/or Gln227Tyr mutations influence the orientations of potentially catalytic residues (e.g., Asp15 and His239) and distribution of active site water compared to wild-type CU-NH (Figures 5.3-5.6). Since the pK_a of active site amino acids are sensitive to conformational and solvation changes,[1] we examine the active site dynamics of each single mutant bound to inosine with anionic Asp15 and cationic His239.

(i) Thr223Tyr CU-NH: For Thr223Tyr CU-NH, a His239–Asp15 dyad forms for this active site protonation state that is consistent with wild-type CU-NH, but no hydrogen bond exists between Tyr223 or His239 and N7 of inosine (occupancy ≤ 0.5%, Figure 6c and Table D.5). Additionally, although O6 of inosine is solvated for 88.0% of the trajectory (Table D.5), water does not bridge interactions between His239 and the substrate (Figure 6c). Consequently, as discussed in the main text for neutral Asp15 and His239, there are no viable avenues for leaving group activation by the Thr223Tyr CU-NH single mutant when Asp15 is anionic and His239 is cationic.

(ii) Gln227Tyr CU-NH: For the Gln227Tyr CU-NH mutant, Tyr227 minimally interacts with the substrate when Asp15 is anionic and His239 is cationic (I(N7)···Tyr227(OηH) occupancy = 11.1%; Figure 6d and Table D.5). As observed for neutral His239, cationic His239 does not donate a hydrogen bond to inosine, and therefore is also unlikely to facilitate proton transfer. Interestingly, while a water bridge is present between O6 and His239 (Figure 6d) and could facilitate nucleobase protonation, previous work indicates that N7 protonation will have a greater catalytic effect,[2] and kinetic isotope effects indicate that IU-NH-mediated hydrolysis of inosine proceeds through N7 protonation.[3] Therefore, O6 protonation is likely insufficient to stabilize the nucleobase for departure, and the reaction is unlikely to proceed. Since the simulations on Gln227Tyr CU-NH with neutral Asp15 and His239 (main text) rationalizes the observed 7.5 times faster hydrolysis rate of inosine by the Gln227Tyr single mutant compared to the wild-type enzyme catalyzed hydrolysis,[4] the simulations on the Gln227Tyr mutant with anionic Asp15 and cationic His239 support our proposed neutral protonation state of both residues.

Table D.1. RESP partial charges and atom types of the calcium–ligating residues in the CU–NH active site.

Atom Name	Atom Type	Charge
Inosine		
C5'	CT	0.053
O5'	OH	–0.620
C4'	CT	0.078
O4'	OS	–0.305
C3'	CT	0.054
O3'	Y6	–0.453
C2'	CT	–0.006
O2'	Y7	–0.515
C1'	CT	–0.012
N1	NA	–0.184
C2	CA	0.096
N3	NC	–0.472
C4	CB	0.220
C5	CB	0.316
C6	C	0.279
O6	O	–0.491
N7	NB	–0.529
C8	CK	0.025
N9	N*	–0.002
HO5'	HO	0.433
HO3'	HO	0.254
HO2'	HO	0.348
H5'2	H1	0.076
H5'1	H1	0.076
H1	H	0.278
H8	H5	0.168
H4'	H1	0.163
H3'	H1	0.139
H2'	H1	0.192
H2	H5	0.154
H1'	H2	0.214
Asp16		
N	N	–0.516
H	H	0.361
CA	CX	0.038
HA	H1	0.090
CB	2C	–0.066
HB2	HC	0.056
HB3	HC	0.056
CG	CO	0.083
OD1	Y2	–0.298
OD2	Y3	–0.375
C	C	0.537
O	O	–0.582
Calcium		
Ca^{2+}	CA1	0.494
Water		
O	Y8	–0.176
H1	HW	0.199
H2	HW	0.251
Asp11		
N	N	–0.516
H	H	0.166
CA	CX	0.038
HA	H1	0.049
CB	2C	–0.168
HB2	HC	0.073
HB3	HC	0.073
CG	CO	0.569

OD1	Y1	–0.235
OD2	O2	–0.706
C	C	0.537
O	O	–0.582
Asp240		
N	N	–0.516
H	H	0.317
CA	CX	0.038
HA	H1	0.094
CB	2C	–0.040
HB2	HC	0.086
HB3	HC	0.086
CG	CO	0.274
OD1	O2	–0.575
OD2	Y5	–0.308
C	C	0.537
O	O	–0.582
Val124		
N	N	–0.538
H	H	0.329
CA	CX	0.002
HA	H1	0.123
CB	3C	0.182
HB	HC	–0.020
CG1	CT	–0.141
HG11	HC	0.043
HG12	HC	0.043
HG13	HC	0.043
CG2	CT	–0.128
HG21	HC	0.016
HG22	HC	0.016
HG23	HC	0.016
C	C	0.383
O	Y4	–0.257

Table D.2. GAFF and MCPB.py generated bond and angle parameters for the CU-NH active site.

Parameter	K_r	r_{eq}	Derived from:
M1-Y6	3.9	2.703	Created by Seminario method using MCPB.py
M1-Y7	14.5	2.633	Created by Seminario method using MCPB.py
M1-Y8	0.0	4.017	Created by Seminario method using MCPB.py
Y1-M1	34.0	2.348	Created by Seminario method using MCPB.py
Y2-M1	18.6	2.494	Created by Seminario method using MCPB.py
Y3-M1	19.5	2.471	Created by Seminario method using MCPB.py
Y4-M1	12.4	2.478	Created by Seminario method using MCPB.py
Y5-M1	50.1	2.292	Created by Seminario method using MCPB.py
C -Y4	570.0	1.229	JCC,7,(1986),230; AA,CYT,GUA,THY,URA
CO-Y1	656.0	1.250	Created by Seminario method using MCPB.py
CO-Y2	656.0	1.250	Created by Seminario method using MCPB.py
CO-Y3	656.0	1.250	Created by Seminario method using MCPB.py
CO-Y5	656.0	1.250	Created by Seminario method using MCPB.py
CT-Y6	320.0	1.410	JCC,7,(1986),230; SUGARS
CT-Y7	320.0	1.410	JCC,7,(1986),230; SUGARS
Y6-HO	553.0	0.960	JCC,7,(1986),230; SUGARS,SER,TYR
Y7-HO	553.0	0.960	JCC,7,(1986),230; SUGARS,SER,TYR
Y8-HW	553.0	0.957	! TIP3P water
C -Y4-M1	20.2	125.1	Created by Seminario method using MCPB.py
CO-Y1-M1	35.8	179.6	Created by Seminario method using MCPB.py
CO-Y2-M1	90.7	90.9	Created by Seminario method using MCPB.py
CO-Y3-M1	74.7	92.2	Created by Seminario method using MCPB.py
CO-Y5-M1	50.1	166.4	Created by Seminario method using MCPB.py
M1-Y6-CT	57.9	107.5	Created by Seminario method using MCPB.py
M1-Y6-HO	27.4	100.3	Created by Seminario method using MCPB.py
M1-Y7-CT	53.3	104.6	Created by Seminario method using MCPB.py
M1-Y7-HO	19.9	104.1	Created by Seminario method using MCPB.py
M1-Y8-HW	57.8	125.3	Created by Seminario method using MCPB.py
Y1-M1-Y2	28.4	89.8	Created by Seminario method using MCPB.py
Y1-M1-Y3	28.3	96.6	Created by Seminario method using MCPB.py
Y1-M1-Y4	31.1	82.2	Created by Seminario method using MCPB.py
Y1-M1-Y5	23.1	162.5	Created by Seminario method using MCPB.py
Y1-M1-Y6	30.6	80.9	Created by Seminario method using MCPB.py
Y1-M1-Y7	36.0	81.6	Created by Seminario method using MCPB.py
Y1-M1-Y8	30.7	48.5	Created by Seminario method using MCPB.py
Y2-M1-Y3	101.1	53.5	Created by Seminario method using MCPB.py
Y2-M1-Y4	52.5	80.5	Created by Seminario method using MCPB.py
Y2-M1-Y5	31.4	107.2	Created by Seminario method using MCPB.py
Y2-M1-Y6	40.3	155.9	Created by Seminario method using MCPB.py
Y2-M1-Y7	31.3	140.7	Created by Seminario method using MCPB.py
Y2-M1-Y8	49.7	137.5	Created by Seminario method using MCPB.py
Y3-M1-Y4	52.5	133.9	Created by Seminario method using MCPB.py
Y3-M1-Y5	26.9	96.7	Created by Seminario method using MCPB.py
Y3-M1-Y6	33.5	149.4	Created by Seminario method using MCPB.py
Y3-M1-Y7	31.5	89.3	Created by Seminario method using MCPB.py
Y3-M1-Y8	50.9	115.2	Created by Seminario method using MCPB.py
Y4-M1-Y5	37.6	96.6	Created by Seminario method using MCPB.py
Y4-M1-Y6	25.8	76.2	Created by Seminario method using MCPB.py
Y4-M1-Y7	24.1	135.2	Created by Seminario method using MCPB.py
Y4-M1-Y8	15.9	98.3	Created by Seminario method using MCPB.py
Y5-M1-Y6	30.3	81.9	Created by Seminario method using MCPB.py
Y5-M1-Y7	33.4	87.2	Created by Seminario method using MCPB.py
Y5-M1-Y8	34.4	115.0	Created by Seminario method using MCPB.py
Y7-M1-Y6	18.5	60.1	Created by Seminario method using MCPB.py
Y8-M1-Y6	46.9	42.6	Created by Seminario method using MCPB.py
Y8-M1-Y7	53.1	42.3	Created by Seminario method using MCPB.py
2C-CO-Y1	70.0	117.0	Created by Seminario method using MCPB.py
2C-CO-Y2	70.0	117.0	Created by Seminario method using MCPB.py

2C-CO-Y3	70.0	117.0	Created by Seminario method using MCPB.py
2C-CO-Y5	70.0	117.0	Created by Seminario method using MCPB.py
CT-CT-Y6	50.0	109.5	Created by Seminario method using MCPB.py
CT-CT-Y7	50.0	109.5	Created by Seminario method using MCPB.py
CT-Y6-HO	55.0	108.5	Created by Seminario method using MCPB.py
CT-Y7-HO	55.0	108.5	Created by Seminario method using MCPB.py
CX-C -Y4	80.0	120.4	(was CT-C-O)
H1-CT-Y6	50.0	109.5	changed based on NMA nmodes
H1-CT-Y7	50.0	109.5	changed based on NMA nmodes
HW-Y8-HW	100.0	104.5	TIP3P water
O2-CO-Y1	80.0	126.0	Created by Seminario method using MCPB.py
Y3-CO-Y2	80.0	126.0	Created by Seminario method using MCPB.py
Y4-C -N	80.0	122.9	AA general
Y5-CO-O2	80.0	126.0	Created by Seminario method using MCPB.py

Table D.3. Average active site root-mean-square deviation (rmsd) (standard deviation (σ) provided in parentheses) of the CU-NH–I complex over each of three 100 ns replicate trajectories relative to the representative structure obtained by clustering replicate 1.[a]

	Replicate 1 Avg (σ)	Replicate 2 Avg (σ)	Replicate 3 Avg (σ)
Wild ($Asp15^{-}$ $His239^{+}$) [b]	0.795 (0.168)	0.901 (0.269)	1.293 (0.244)
Thr223Tyr Gln227Tyr ($Asp15^{-}$ $His239^{+}$)[b]	0.898 (0.192)	1.318 (0.155)	1.377 (0.277)
Thr223Tyr Gln227Tyr (Asp15 His239)[c]	0.715 (0.120)	1.011 (0.190)	1.249 (0.255)
Wild (Asp15 His239)[c]	0.750 (0.180)	1.243 (0.311)	1.047 (0.348)
Thr223Tyr (Asp15 His239)[c]	1.126 (0.124)	1.077 (0.115)	1.366 (0.099)
Thr223Tyr ($Asp15^{-}$ $His239^{+}$)[b]	0.739 (0.099)	0.772 (0.083)	1.168 (0.130)
Gln227Tyr (Asp15 His239)[c]	0.859 (0.211)	1.391 (0.329)	1.167 (0.225)
Gln227Tyr ($Asp15^{-}$ $His239^{+}$)[b]	1.287 (0.380)	1.397 (0.309)	1.281 (0.460)

[a]rms fit was performed based on the position of all heavy atoms of the active site residues (namely inosine, Ca^{2+} ion, nucleophilic water, Asp11, Asp15, Asp16, His82, Val124, Res223, Res227, Tyr231, His239, and Asp240) to yield the average rmsd of the entire CU-NH active site, while the rmsd per residue was determined without refitting. [b]CU-NH complex with anionic Asp15 and cationic His239. [c]CU-NH complex with neutral Asp15 and neutral His239.

Table D.4. Summary of important average distances (standard deviation (σ) provided in parentheses) in the active site from MD simulations on wild-type and mutant CU-NH bound to inosine.

	Asp15⁻ His239⁺ [a]				Asp15 His239 [b]			
	Wild-type	Thr223Tyr Gln227Tyr	Thr223Tyr	Gln227Tyr	Wild-type	Thr223Tyr Gln227Tyr	Thr223Tyr	Gln227Tyr
Distance	Avg (σ)	Avg (σ)	Avg (σ)	Avg (σ)	Avg (σ)	Avg (σ)	Avg (σ)	Avg (σ)
Ca^{2+}···Inosine(O2′)	2.787 (0.109)	2.770 (0.108)	2.804 (0.110)	2.800 (0.108)	2.814 (0.111)	2.794 (0.109)	2.789 (0.109)	2.814 (0.110)
Ca^{2+}···Inosine(O3′)	2.921 (0.199)	2.816 (0.178)	2.932 (0.185)	2.946 (0.187)	2.905 (0.181)	3.001 (0.200)	2.888 (0.183)	2.909 (0.182)
Ca^{2+}···Val124(O)	2.285 (0.144)	2.406 (0.140)	2.346 (0.140)	2.314 (0.140)	2.326 (0.139)	2.266 (0.144)	2.404 (0.150)	2.314 (0.143)
Ca^{2+}···Asp11(Oδ1)	2.325 (0.089)	2.329 (0.088)	2.325 (0.093)	2.327 (0.090)	2.319 (0.089)	2.309 (0.090)	2.337 (0.090)	2.336 (0.091)
Ca^{2+}···Asp14(Oδ1)	2.404 (0.086)	2.408 (0.084)	2.403 (0.086)	2.395 (0.085)	2.396 (0.085)	2.399 (0.086)	2.430 (0.085)	2.399 (0.086)
Ca^{2+}···Asp14(Oδ2)	2.429 (0.084)	2.393 (0.084)	2.445 (0.083)	2.437 (0.083)	2.446 (0.083)	2.447 (0.083)	2.427 (0.084)	2.441 (0.083)
Ca^{2+}···Asp238(Oδ2)	2.267 (0.076)	2.270 (0.075)	2.256 (0.076)	2.267 (0.076)	2.260 (0.075)	2.246 (0.075)	2.253 (0.076)	2.260 (0.076)
His82···Inosine[c]	5.170 (0.332)	5.034 (0.454)	5.341 (0.315)	4.966 (0.335)	5.248 (0.293)	5.483 (0.253)	5.569 (0.294)	5.211 (0.346)
His239···Inosine[c]	4.927 (0.288)	5.211 (0.285)	4.998 (0.448)	5.442 (0.389)	5.691 (0.321)	5.102 (0.285)	5.660 (0.339)	6.052 (0.321)
H_2O(O)···Inosine(C1′)	4.293 (0.678)	4.705 (0.513)	3.649 (0.369)	3.825 (0.566)	4.409 (0.602)	4.022 (0.548)	4.369 (0.557)	3.816 (0.535)
Tyr223···Tyr227[c]	-	6.162 (0.732)	-	-	-	6.055 (0.329)	-	-

[a]CU-NH complex with anionic Asp15 and cationic His239. [b]CU-NH complex with neutral Asp15 and neutral His239. [c]Distance between the centers of mass of the amino acid sidechain(s) and/or inosine

Table D.5. Summary of occupancies and average geometrical parameters for important hydrogen bonds in the active site from MD simulations on wild-type and mutant CU-NH bound to inosine (I) with anionic Asp15 and cationic His239.

Hydrogen Bond	Wild			Thr223Tyr Gln227Tyr			Thr223Tyr			Gln227Tyr		
	%[a,b]	Dist[c]	Angle[d]	%[a,b]	Dist[c]	Angle[d]	%[a,b]	Dist[c]	Angle[d]	%[a,b]	Dist[c]	Angle[d]
I(O6)···Gln227(NεH)	95.2%	2.920	160.5	N/A	N/A	N/A	86.4%	3.003	156.8	N/A	N/A	N/A
Tyr231(Oη)···I(N1H)	83.9%	3.045	158.1	62.2%	3.059	153.5	65.3%	3.095	160.2	23.9%	3.080	154.6
Asp11(Oδ)···H_2O(OH)	86.7%	2.736	156.6	94.7%	2.757	155.6	90.6%	2.717	162.4	90.4%	2.725	156.4
Glu164(Oε)···I(O5′H)	100.0%	2.605	167.3	76.1%	2.651	164.0	100.0%	2.621	166.7	100.0%	2.610	167.1
I(O2′)···Asn40(NδH)	86.6%	3.075	156.9	83.1%	3.078	151.5	83.0%	3.079	162.0	89.0%	3.071	159.6
I(O3′)···Asn166(NδH)	58.8%	3.024	146.5	0.2%	3.086	137.1	88.3%	3.075	153.5	90.7%	3.040	148.9
Asn166(Oδ)···I(O3′H)	30.6%	2.969	147.3	49.8%	3.017	154.6	5.1%	3.065	137.9	24.1%	2.983	142.4
Asn166(Oδ)···H_2O(OH)	0.1%	2.961	157.7	0.0%	3.005	151.3	78.2%	2.813	161.7	0.3%	2.905	161.6
Asp240(Oδ)···I(O2′H)	100.0%	3.022	145.0	96.3%	2.990	148.2	100.0%	3.051	146.5	100.0%	3.058	145.1
I(N7)···H_2O(OH)	72.6%	3.030	155.4	6.7%	3.078	140.5	1.1%	3.115	142.8	56.9%	3.006	147.8
I(O6)···H_2O(OH)	70.7%	2.908	149.2	70.9%	2.904	152.4	88.0%	2.853	151.7	135.4%	2.909	149.8
I(N3)···H_2O(OH)	38.0%	2.948	149.7	59.0%	3.032	145.9	8.5%	2.982	146.4	12.5%	2.944	150.0
H_2O(O)···His239(NεH)	183.2%	3.058	142.8	211.8%	3.039	144.9	190.6%	3.047	148.6	140.1%	3.053	143.2
Asp15(Oδ)···His239(NδH)	99.0%	2.763	153.4	100.0%	2.715	152.5	99.1%	2.728	155.0	66.5%	2.786	151.8
I(N7)···Tyr223(OηH)	N/A	N/A	N/A	7.2%	2.969	151.4	0.5%	2.991	134.2	N/A	N/A	N/A
I(N7)···Tyr227(OηH)	N/A	N/A	N/A	4.0%	2.980	152.7	N/A	N/A	N/A	11.1%	2.946	140.2
I(O6)···Tyr231(OηH)	0.8%	2.842	159.6	14.4%	2.815	159.3	3.0%	2.844	158.4	39.0%	2.817	157.0

[a]Percent occupancy of the hydrogen bond, which is defined using cutoffs of a distance less than 3.4 Å and an angle greater than 120°. [b]Greater than 100% occupancy indicates the presence of more than one water hydrogen bonding with an acceptor site.[c]Hydrogen-bonding distance in Å. [d]Hydrogen-bonding angle in degrees.

Table D.6. Average backbone and side chain torsional angles (standard deviation (σ) provided in parentheses) for key active site residues from MD simulations on wild-type and mutant CU-NH with anionic Asp15 and cationic His239.

		Wild		Thr223Tyr Gln227Tyr		Thr223Tyr		Gln227Tyr	
Angle[a-c]	3B9X	Avg (σ)	%[d]	Avg (σ)	%[d]	Avg (σ)	%[d]	Avg (σ)	%[d]
Asp11 ψ	63.3	63.4 (8.4)	99.9	60.8 (7.9)	100.0	64.1 (8.2)	99.9	65.4 (8.6)	99.9
Asp11 φ	−138.0	−132.6 (8.6)	96.1	−131.3 (8.7)	98.0	−132.7 (9.1)	95.0	−133.5 (8.9)	94.3
Asp11 χ1	57.6	66.2 (7.7)	99.9	68.0 (7.2)	100.0	61.7 (7.8)	100.0	62.3 (7.6)	100.0
Asp15 ψ	−27.8	−38.0 (5.5)	71.6	−45.2 (7.3)	97.0	−43.2 (7.2)	93.1	−40.4 (6.5)	83.7
Asp15 φ	−70.8	−74.8 (8.4)	91.2	−75.5 (8.2)	86.9	−68.4 (9.7)	95.6	−70.7 (8.7)	96.9
Asp15 χ1	57.1	−63.3 (10.1)	99.0	−135.7 (10.3)	79.8	−50.5 (9.2)	52.4	−61.4 (10.2)	83.2
Asn40 ψ	−43.5	−38.8 (6.0)	69.3	−38.4 (5.9)	69.7	−40.9 (7.0)	82.3	−38.6 (6.1)	62.9
Asn40 φ	−44.2	−52.3 (9.2)	99.0	−51.8 (8.8)	99.0	−53.5 (9.6)	98.9	−56.3 (10.0)	99.3
Asn40 χ1	−81.7	−70.4 (7.5)	99.2	−70.8 (7.5)	99.0	−72.8 (7.5)	90.7	−69.3 (7.7)	98.9
His82 ψ	−66.0	−9.3 (7.1)	57.6	−7.9 (5.8)	60.8	−7.3 (5.4)	57.4	−8.2 (5.9)	59.7
His82 φ	147.6	−116.2 (12.7)	90.6	−107.4 (10.5)	84.1	−110.7 (10.8)	93.2	−106 (11.0)	69.5
His82 χ1	72.5	−63.0 (9.6)	99.2	−60.9 (7.5)	100.0	−62.5 (7.5)	100.0	−62.7 (7.9)	99.9
Glu164 ψ	149.8	139.1 (6.3)	87.7	140.2 (6.4)	80.3	140.7 (5.6)	84.0	141.0 (5.7)	82.0
Glu164 φ	−91.0	−76.2 (8.2)	80.9	−118.1 (16.2)	67.2	−77.6 (7.9)	70.2	−76.3 (8.2)	80.8
Glu164 χ1	173.0	184.4 (5.9)	100.0	177.9 (9.4)	92.4	185.6 (5.7)	100.0	182.1 (6.3)	100.0
Asn166 ψ	−39.2	−44.0 (7.3)	91.5	−39.6 (6.6)	54.5	−46.0 (7.4)	96.8	−44.5 (7.3)	93.9
Asn166 φ	−52.4	−62.6 (7.8)	99.9	−63.5 (7.5)	100.0	−62.8 (8.0)	99.8	−61.5 (7.4)	100.0
Asn166 χ1	−72.1	−70.3 (7.5)	64.2	202.8 (5.4)	51.6	−68 (7.5)	95.9	−70.5 (7.5)	97.1
Res223 ψ	−22.0	−42.1 (7.2)	86.7	−46.2 (8.3)	93.4	−37.7 (5.5)	63.9	−43.3 (7.7)	84.4
Res223 φ	−71.5	−68.9 (8.1)	99.2	−62.5 (8.3)	99.9	−64.6 (8.4)	99.7	−63.8 (8.5)	99.8
Res223 χ1	−60.9	−59.9 (8.3)	99.8	177.4 (9.5)	99.6	189.0 (8.5)	97.9	181.5 (9.1)	70.0
Res227 ψ	−27.7	−38.5 (5.8)	69.4	−43.6 (7.1)	93.0	−39.1 (6.2)	71.3	−40.3 (6.7)	75.7
Res227 φ	−68.1	−68.9 (8.6)	98.7	−63.4 (8.1)	99.9	−67.3 (8.2)	99.4	−68.0 (8.2)	99.4
Res227 χ1	−57.0	−68.0 (6.7)	99.9	188 (10.4)	87.4	−70.0 (6.4)	99.9	−75.6 (11.6)	41.4
Tyr231 ψ	−5.4	−9.1 (6.6)	60.8	−11.8 (7.2)	74.8	−8.5 (6.4)	56.1	−12.2 (7.3)	77.1
Tyr231 φ	−125.3	−132.5 (10.3)	94.8	−125.1 (12.6)	97.1	−133.8 (9.7)	94.4	−129.6 (11.8)	95.2

Tyr231 χ1	−49.2	−52.8 (8.5)	98.8	−51.6 (9.0)	98.9	−50.9 (8.7)	98.1	−51.8 (9.2)	97.7
His239 ψ	−50.5	−63.8 (9.8)	98.3	−57.6 (9.9)	99.5	−58.0 (10.9)	99.1	−64.3 (11.9)	95.3
His239 φ	−43.2	−75.4 (9.7)	73.7	−110 (12.9)	70.3	−74.0 (11.3)	51.7	−106.7 (12.8)	50.4
His239 χ1	−49.4	−48.5 (7.8)	98.7	−46.3 (7.8)	94.6	−44.7 (7.6)	94.2	−50.7 (8.6)	98.5
Asp240 ψ	−28.0	−22.0 (5.7)	61.2	−22.8 (5.3)	52.7	−22.5 (5.6)	52.4	−21.6 (5.9)	63.7
Asp240 φ	−77.1	−67.0 (8.6)	99.6	−73.2 (8.5)	95.1	−69.8 (9.2)	97.6	−68.5 (9.3)	98.3
Asp240 χ1	−59.3	−53.0 (7.8)	98.6	−63.5 (7.2)	100.0	−57.9 (7.8)	99.9	−55.7 (7.5)	99.7
Inosine χ	−52.1	−41.4 (8.4)	68.4	−48.3 (10.3)	80.0	−41.9 (8.5)	71.7	−46.6 (9.8)	85.6

[a]χ1 defined as $\angle(NC_\alpha C_\beta C_\gamma)$. [b]φ defined as $\angle(C_{-1}NC_\alpha C)$. [c]ψ defined as $\angle(NC_\alpha CN_{+1})$. [d]Percentage of the trajectory that the dihedral angle in the highest occupied conformation falls within one of the following ranges: −30 to 30°; 30 to 90°; 90 to 150°; 150 to −150°; −90 to −150°; or −30 to −90°.

Table D.7. Average pairwise MM–GBSA free energy and standard deviations (σ; kJ mol^{-1}) between inosine and key active site amino acids over the MD simulation.

		Asp11	Asp15	His82	Glu164	Asn166	Res223	Res227	Tyr231	His239	Ca^{2+}
Res227	Protonation State	Avg (σ)	Avg (σ)	Avg (σ)	Avg (σ)	Avg (σ)	Avg (σ)	Avg (σ)	Avg (σ)	Avg (σ)	Avg (σ)
Wild	$Asp15^{-}$ $His239^{+}$ [a]	1.3 (1.8)	−9.8 (3.5)	−11.0 (2.7)	−36.8 (3.6)	−7.7 (4.4)	−0.4 (0.1)	−11.9 (2.2)	−10.9 (3.3)	−13.2 (3.4)	−6.6 (4.1)
Thr223Tyr Gln227Tyr	$Asp15^{-}$ $His239^{+}$ [a]	2.2 (1.8)	−13.3 (4.8)	−10.6 (2.6)	−24.7 (12.3)	−10.0 (7.4)	−4.4 (4.0)	−3.7 (3.2)	−10.3 (4.0)	−12.8 (3.5)	−5.8 (4.6)
Thr223Tyr Gln227Tyr	Asp15 $His239^{b}$	0.7 (1.7)	−7.4 (2)	−6.7 (1.9)	−35.0 (3.4)	−9.4 (3.2)	−13.6 (2.7)	−3.6 (1.7)	−11.2 (2.3)	−13.1 (3.0)	−6.9 (3.8)
Thr223Tyr	Asp15 $His239^{b}$	1.4 (1.5)	−7.0 (1.9)	−7.9 (2.6)	−38.2 (3.6)	−3.3 (3.7)	−7.1 (6.2)	−13.6 (3.2)	−10.5 (2.7)	−9.7 (2.7)	−6.3 (4.2)
Gln227Tyr	Asp15 $His239^{b}$	1.1 (1.4)	−7.8 (2.2)	−8.7 (2.1)	−38.0 (3.5)	−8.6 (3.2)	−0.3 (0.1)	−13.5 (4.0)	−11.4 (4.1)	−8.7 (2.3)	−7.0 (3.9)
Thr223Tyr	$Asp15^{-}$ $His239^{+}$ [a]	1.1 (1.3)	−14.5 (5.7)	−9.9 (2.4)	−35.4 (3.2)	−8.6 (2.5)	−1.0 (1.8)	−14.5 (2.5)	−9.0 (3.3)	−11.6 (3)	−8.6 (4.2)
Gln227Tyr	$Asp15^{-}$ $His239^{+}$ [a]	1.4 (1.6)	−13.1 (4.4)	−11.0 (2.1)	−36.6 (3.6)	−9.0 (2.4)	−0.3 (0.2)	−3.1 (3.8)	−8.2 (5.8)	−9.8 (3.3)	−7.8 (4.5)
Wild	Asp15 $His239^{b}$	0.8 (1.4)	−7.2 (1.8)	−9.7 (2.4)	−37.5 (3.7)	−9.2 (2.9)	−0.4 (0.1)	−12.3 (2.1)	−10.0 (3.2)	−8.7 (2.4)	−6.8 (3.7)

[a]CU-NH complex with anionic Asp15 and cationic His239. [b]CU-NH complex with neutral Asp15 and neutral His239.

Table D.8. Summary of occupancies and average geometrical parameters for important hydrogen bonds in the active site during MD simulations on wild-type and mutant CU-NH bound to inosine (I) with neutral Asp15 and neutral His239.

Hydrogen Bond	Thr223Tyr Gln227Tyr			Thr223Tyr			Gln227Tyr			Wild		
	%[a]	Dist[b]	Angle[c]	%[a]	Dist[b]	Angle[c]	%[a]	Dist[b]	Angle[c]	%[a]	Dist[b]	Angle[c]
I(O6)···Gln227(NεH)	N/A	N/A	N/A	93.3%	2.918	159.7	N/A	N/A	N/A	95.4%	2.923	159.8
Tyr231(Oη)···I(N1H)	89.7%	3.049	157.7	84.4%	3.065	158.7	28.6%	3.146	150.3	79.1%	3.081	161.7
Asp11(Oδ)···H_2O(OH)	89.4%	2.753	156.8	93.7%	2.772	156.0	94.7%	2.717	156.6	94.5%	2.755	156.5
Glu164(Oε)···I(O5′H)	100.0%	2.618	167.1	100.0%	2.604	167.8	100.0%	2.610	167.3	100.0%	2.613	167.5
I(O2′)···Asn40(NδH)	82.9%	3.128	162.5	84.4%	3.097	160.2	88.3%	3.090	160.5	76.2%	3.118	157.4
I(O3′)···Asn166(NδH)	94.0%	2.991	147.0	16.9%	3.044	143.4	88.3%	3.034	149.7	90.2%	3.039	146.0
Asn166(Oδ)···I(O3′H)	22.1%	3.029	138.9	18.7%	3.039	143.3	27.5%	3.042	140.2	35.1%	2.933	146.0
Asn166(Oδ)···H_2O(OH)	3.4%	2.875	159.1	0.2%	3.000	153.9	0.1%	3.127	151.7	0.2%	3.089	154.3
Asp240(Oδ)···I(O2′H)	100.0%	3.018	144.1	100.0%	3.018	146.0	100.0%	3.046	145.4	100.0%	3.025	145.8
I(N7)···H_2O(OH)	0.7%	3.062	143.1	8.0%	3.116	145.3	9.6%	3.068	141.4	60.0%	3.030	151.3
I(O6)···H_2O(OH)	23.0%	2.990	146.8	75.4%	2.913	148.1	81.2%	2.904	151.6	87.2%	2.915	148.5
I(N3)···H_2O(OH)	12.4%	3.105	141.8	37.0%	3.016	145.9	12.9%	2.951	150.9	51.0%	2.950	152.9
Tyr223(Oη)···D15(OδH)	80.8%	2.917	163.2	76.7%	2.957	158.0	N/A	N/A	N/A	N/A	N/A	N/A
I(N7)···Tyr223(OηH)	75.6%	2.926	146.5	41.8%	2.900	158.6	N/A	N/A	N/A	N/A	N/A	N/A
I(N7)···Tyr227(OηH)	0.2%	3.177	138.1	N/A	N/A	N/A	86.5%	2.892	159.5	N/A	N/A	N/A
I(O6)···Tyr231(OηH)	1.1%	2.819	154.9	1.8%	2.829	157.7	59.2%	2.767	160.3	1.0%	2.824	158.6

[a]Percent occupancy of the hydrogen bond, which is defined using cutoffs of a distance less than 3.4 Å and an angle greater than 120°. [b]Greater than 100% occupancy indicates the presence of more than one water hydrogen bonding with an acceptor site. [c]Hydrogen-bonding distance in Å. [d]Hydrogen-bonding angle in degrees.

Table D.9. Average backbone and side chain torsional angles (standard deviation (σ) provided in parentheses) for key active site residues from MD simulations on wild-type and mutant CU-NH bound to inosine with neutral Asp15 and neutral His239.

		Thr223Tyr Gln227Tyr		Thr223Tyr		Gln227Tyr		Wild	
Angle[a-c]	3B9X	Avg (σ)	%[d]	Avg (σ)	%[d]	Avg (σ)	%[d]	Avg (σ)	%[d]
Asp11 ψ	63.3	64.8 (7.9)	99.9	63.0 (8.0)	99.9	65.8 (8.6)	99.8	64.1 (8.1)	99.9
Asp11 φ	−138.0	−133.8 (9)	93.1	−129.7 (9.1)	98.6	−134.8 (8.7)	92.3	−132.2 (9.0)	96.3
Asp11 χ1	57.6	63.3 (7.4)	100.0	68.5 (7.8)	99.8	64.0 (7.8)	100.0	63.2 (7.5)	100.0
Asp15 ψ	−27.8	−41.8 (6.8)	89.0	−43.3 (7.4)	91.2	−39.8 (6.2)	81.6	−38.8 (5.8)	77.1
Asp15 φ	−70.8	−69.6 (9.1)	96.5	−73.6 (8.9)	88.8	−70.4 (8.5)	98.0	−74.2 (7.9)	95.0
Asp15 χ1	57.1	−52.6 (8.7)	99.5	−56.6 (9.8)	99.6	−66.0 (8.4)	99.5	−64.9 (8.7)	99.8
Asn40 ψ	−43.5	−38.0 (5.8)	63.3	−38.7 (6.0)	70.5	−38.7 (6.1)	65.0	−39.2 (6.4)	68.9
Asn40 φ	−44.2	−53.5 (9.4)	98.9	−52.7 (9.2)	99.0	−55.3 (10.6)	99.2	−53.0 (9.2)	99.1
Asn40 χ1	−81.7	−70.6 (7.4)	99.3	−72.1 (7.3)	98.7	−71.7 (8.3)	95.9	−69.9 (7.6)	98.9
His82 χ1	−66.0	−6.9 (5.1)	55.3	−7.4 (5.4)	58.3	−9.3 (6.5)	62.9	−7.9 (5.8)	60.8
His82 φ	147.6	−110.9 (10.7)	94.4	−112.4 (11.1)	92.6	−109.9 (11.2)	86.7	−115.9 (12.4)	92.6
His82 ψ	72.5	−59.8 (6.8)	100.0	−62.7 (7.4)	100.0	−58.5 (9.6)	99.8	−60.9 (8.3)	99.9
Glu164 ψ	149.8	141.8 (5.4)	77.0	138.7 (6.3)	91.4	140.7 (5.7)	83.9	139.4 (6.0)	90.0
Glu164 φ	−91.0	−78.1 (7.8)	69.2	−78.1 (7.8)	63.6	−77.4 (8.0)	70.7	−76.3 (8.2)	80.9
Glu164 χ1	173.0	183.1 (6.2)	100.0	182.7 (6.2)	100.0	184.8 (5.9)	100.0	184.6 (5.7)	100.0
Asn166 ψ	−39.2	−43.4 (7.3)	88.3	−41.5 (6.9)	82.0	−45.9 (7.2)	96.5	−48.4 (7.7)	98.3
Asn166 φ	−52.4	−65.5 (8.0)	99.8	−64.3 (7.7)	99.9	−60.5 (7.8)	100.0	−62.8 (7.9)	99.9
Asn166 χ1	−72.1	−70.5 (7.6)	93.3	201.0 (6.0)	55.6	−68.5 (7.5)	92.0	−69.2 (7.7)	96.9
Res223 ψ	−22.0	−36.7 (5.0)	52.2	−39.7 (6.3)	76.5	−22.0 (6.1)	53.3	−44.6 (8.1)	86.3
Res223 φ	−71.5	−65.3 (7.9)	99.8	−64.2 (8.4)	99.7	−66.1 (8.4)	99.0	−63.3 (8.7)	99.8
Res223 χ1	−60.9	185.0 (8.1)	99.7	182.9 (8.1)	99.9	−61.2 (8.9)	80.3	181.1 (8.6)	89.0
Res227 ψ	−27.7	−41.6 (6.6)	89.6	−40.9 (6.8)	83.0	−22.5 (5.5)	53.2	−38.9 (6.1)	70.9
Res227 φ	−68.1	−67.5 (7.9)	99.6	−65.5 (8.2)	99.6	−71.1 (8.6)	97.3	−68.5 (8.4)	99.1
Res227 χ1	−57.0	195.2 (8.2)	91.9	−69.5 (7.5)	92.7	−72.6 (7.5)	83.4	−67.3 (6.9)	99.9
Tyr231 ψ	−5.4	−11.0 (7.2)	66.0	−9.8 (6.9)	65.2	−10.0 (7.0)	65.3	−11.8 (7.2)	74.8
Tyr231 φ	−125.3	−132 (10.8)	95.3	−132.4 (10.6)	94.3	−133.1 (10.3)	94.2	−132.3 (10.8)	94.4
Tyr231 χ1	−49.2	−54.0 (8.0)	99.7	−54.7 (8.8)	99.3	−49.2 (8.8)	96.7	−52.1 (8.4)	97.8
His239 ψ	−50.5	−98.2 (5.6)	61.2	−58.1 (13.6)	83.0	−54.2 (11.9)	92.8	−54.2 (10.8)	94.7

His239 φ	–43.2	38.9 (10.8)	84.8	–51.4 (10.1)	63.7	–48.9 (9.5)	75.9	–52.3 (11.1)	80.8
His239 χ1	–49.4	–62.4 (9.9)	98.9	–59.3 (9.7)	99.6	–56.7 (9.5)	99.7	–54.4 (3.8)	99.5
Asp240 ψ	–28.0	–22.2 (5.6)	64.4	–19.6 (6.5)	77.7	–21.3 (6.0)	69.6	–20.1 (6.4)	74.2
Asp240 φ	–77.1	–57.4 (9.5)	99.2	–64.8 (10.0)	99.5	–60.7 (9.2)	99.8	–62.0 (9.2)	99.8
Asp240 χ1	–59.3	–48.9 (6.7)	98.7	–47.9 (7.3)	95.7	–51.1 (6.6)	99.2	–50.7 (6.5)	99.5
Inosine χ	–52.1	–19.9 (6.6)	66.7	–21.1 (6.1)	60.5	–45.9 (9.1)	90.6	–40.4 (4.8)	67.9

[a]χ1 defined as ∠($NC_aC_\beta C_\gamma$). [b]φ defined as ∠($C_{-1}NC_aC$). [c]ψ defined as ∠(NC_aCN_{+1}). [d]Percentage of the trajectory that the dihedral angle in the highest occupied conformation falls within one of the following ranges: –30 to 30°; 30 to 90°; 90 to 150°; 150 to –150°; –90 to –150°; or –30 to –90°.

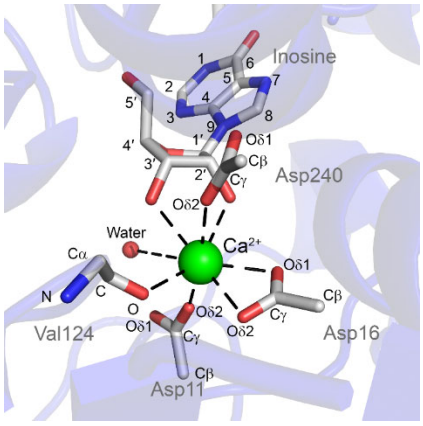


Figure D.1. Structure and numbering of the calcium-ligating residues in the CU-NH active site.

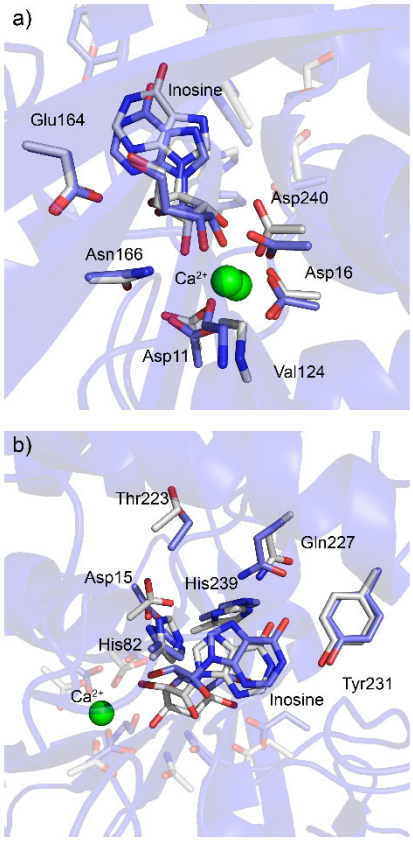


Figure D.2. Overlay of X-ray crystal structure (PDB ID: 3B9X; blue)[33] and MD representative structure (gray) of wild-type CU-NH ($Asp15^{-}$ $His239^{+}$) bound to inosine, highlighting active site residues that surround a) the ribose sugar and Ca^{2+} or b) the inosine nucleobase.

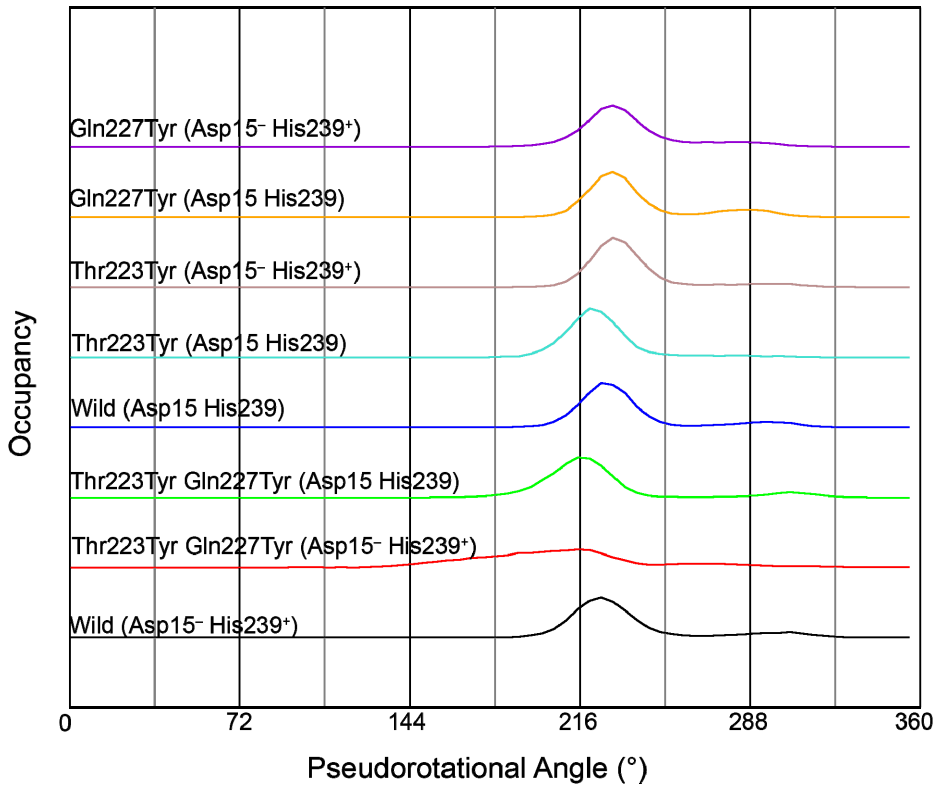


Figure D.3. Stacked histogram displaying the occupancy of the pseudorotational angle (deg.) of the ribose moiety of inosine bound within the wild-type or mutant CU-NH active site with cationic His239 and anionic Asp15 ($His239^{+}$ $Asp15^{-}$) or neutral His239 and neutral Asp15 (His239 Asp15).

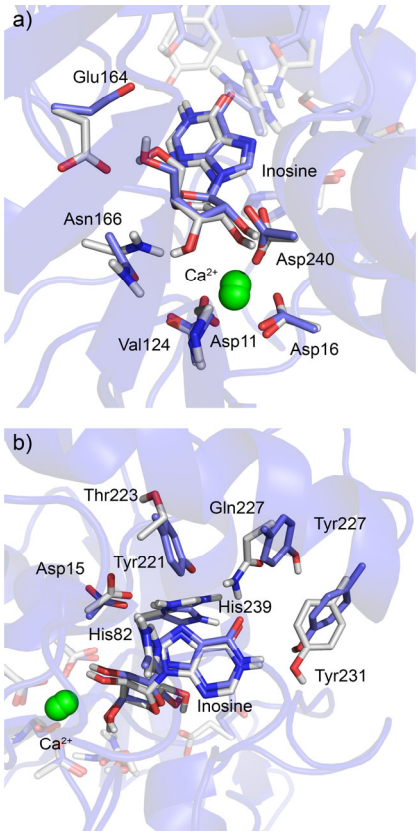


Figure D.4. Overlay of MD representative structures of wild-type CU-NH ($Asp15^{-}$ $His239^{+}$; gray) and Thr223Tyr Gln227Tyr CU-NH ($Asp15^{-}$ $His239^{+}$; blue) bound to inosine, highlighting the active site residues that surround a) the ribose sugar and Ca^{2+} or b) the inosine nucleobase.

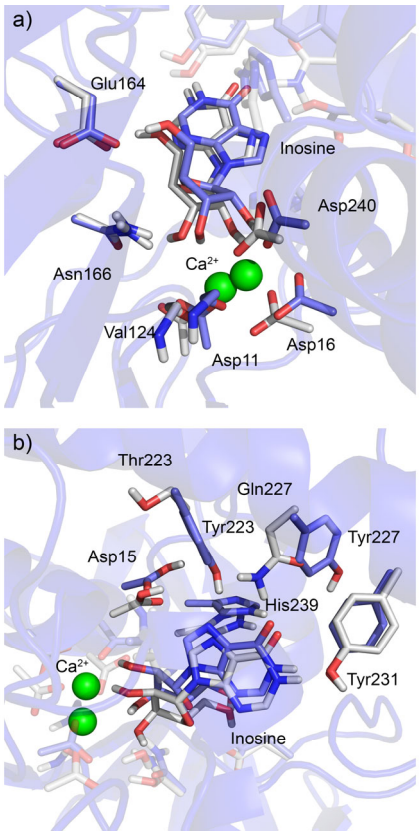


Figure D.5. Overlay of MD representative structures of wild-type CU-NH (Asp15 His239; gray) and Thr223Tyr Gln227Tyr CU-NH (Asp15 His239; blue) bound to inosine, highlighting the active site residues that surround a) the ribose sugar and Ca^{2+} or b) the inosine nucleobase.

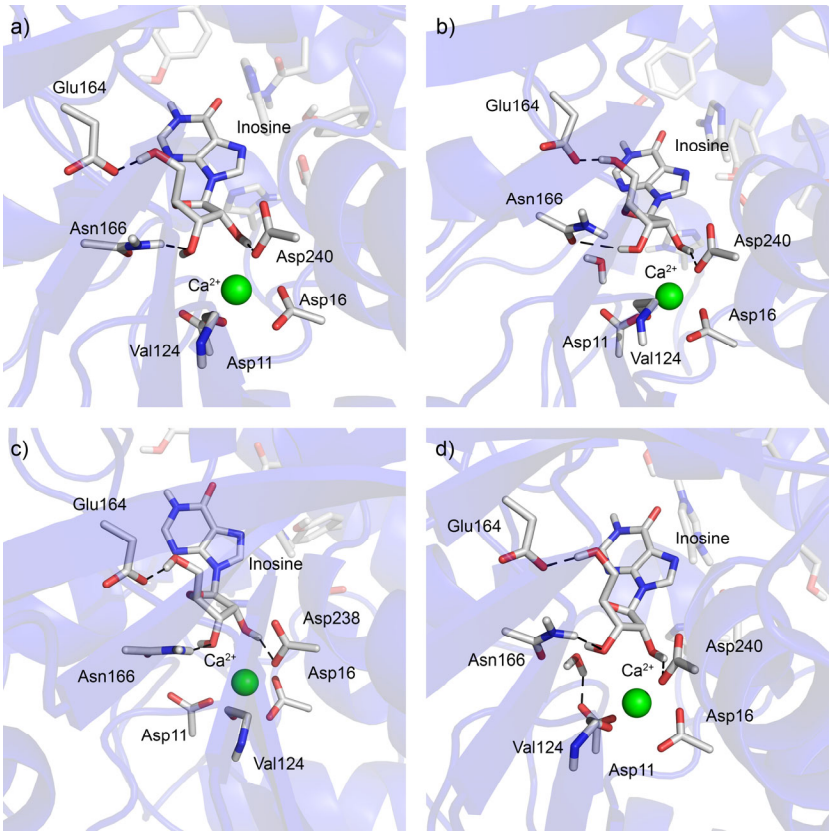


Figure D.6. Representative MD structures highlighting the active site residues that surround the ribose moiety and Ca^{2+} for the a) Thr223Tyr CU-NH ($Asp15^{-}$ $His239^{+}$), b) Gln227Tyr CU-NH ($Asp15^{-}$ $His239^{+}$), c) Thr223Tyr CU-NH–I complex (Asp15 His239), or d) Gln227Tyr CU-NH (Asp15 His239) bound to inosine.

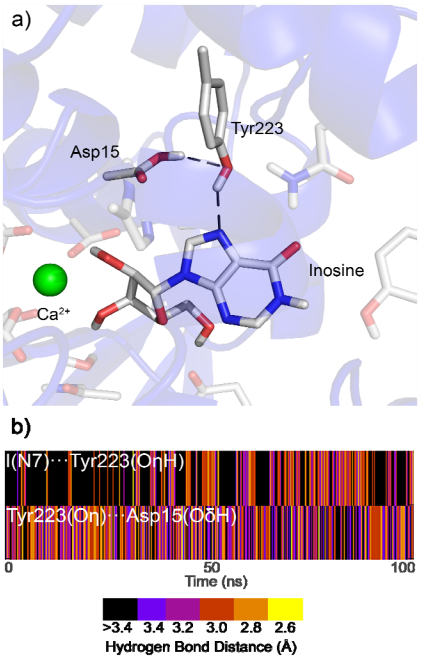


Figure D.7. a) Representative MD structures of Thr223Tyr CU-NH (Asp15 His239) bound to inosine, highlighting hydrogen bonds between Asp15, Tyr223, and the inosine nucleobase. b) Hydrogen-bonding time series plot, highlighting the lack of simultaneous I(N7)···Tyr227(OH) and Tyr223(O_H)···Asp15(OδH) hydrogen bonds (simultaneous occupancy = 23.8%).

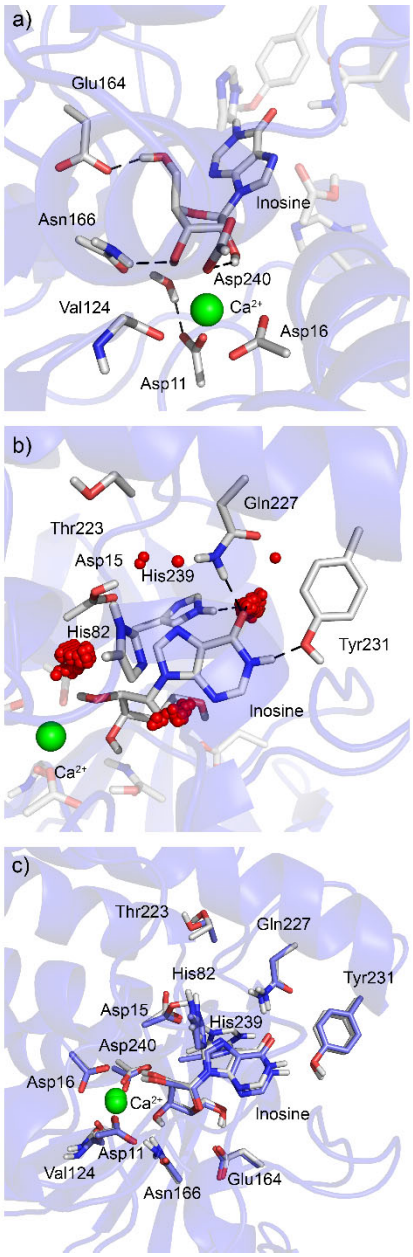


Figure D.8. Representative MD structures of wild-type CU-NH (Asp15 His239) bound to inosine, highlighting the active site residues that surround a) the ribose sugar and Ca^{2+} or b) the inosine nucleobase, and the distribution of active site water (depicted as red spheres when water occupies a 0.50 Å grid space for at least 40% of the trajectory). c) Overlay of the active site from MD representative structures of wild-type CU-NH ($Asp15^{-}$ $His239^{+}$; gray) and wild-type CU-NH (Asp15 His239; blue) bound to inosine.

Full citation for references 6, 20, 22, 26, 49, and 52:

[6] Garcia, M. N., O'Day, S., Fisher-Hoch, S., Gorchakov, R., Patino, R., Feria Arroyo, T. P., Laing, S. T., Lopez, J. E., Ingber, A., Jones, K. M., and Murray, K. O. (2016) One Health Interactions of Chagas Disease Vectors, Canid Hosts, and Human Residents along the Texas-Mexico Border, *PLoS Negl. Trop. Dis. Diseases 10*, e0005074.

[20] Berg, M., Kohl, L., Van der Veken, P., Joossens, J., Al-Salabi, M. I., Castagna, V., Giannese, F., Cos, P., Versees, W., Steyaert, J., Grellier, P., Haemers, A., Degano, M., Maes, L., de Koning, H. P., and Augustyns, K. (2010) Evaluation of Nucleoside Hydrolase Inhibitors for Treatment of African Trypanosomiasis, *Antimicrob. Agents Chemother. 54*, 1900–1908.

[22] Goeminne, A., Berg, M., McNaughton, M., Bal, G., Surpateanu, G., Van der Veken, P., De Prol, S., Versees, W., Steyaert, J., Haemers, A., and Augustyns, K. (2008) *N*-Arylmethyl substituted iminoribitol derivatives as inhibitors of a purine specific nucleoside hydrolase, *Biorg. Med. Chem. 16*, 6752–6763.

[26] Goeminne, A., McNaughton, M., Bal, G., Surpateartu, G., Van der Veken, P., De Prol, S., Versees, W., Steyaert, J., Apers, S., Haemers, A., and Augustyns, K. (2007) 1,2,3-Triazolylalkylribitol derivatives as nucleoside hydrolase inhibitors, *Bioorg. Med. Chem. Lett. 17*, 2523–2526.

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

[52] Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Jr., J. A. M., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, J. M., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, O., Foresman, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J. (2009) Gaussian 09, Revisions D.01 ed., Gaussian, Inc., Wallingford CT.

References

[1] Bashford, D., and Karplus, M. (1990) pKa's of ionizable groups in proteins: atomic detail from a continuum electrostatic model, *Biochemistry 29*, 10219–10225.

[2] Lenz, S. A. P., Kohout, J. D., and Wetmore, S. D. (2016) Hydrolytic Glycosidic Bond Cleavage in RNA Nucleosides: Effects of the 2′-Hydroxy Group and Acid–Base Catalysis, *J. Phys. Chem. B 120*, 12795–12806.

[3] Horenstein, B. A., Parkin, D. W., Estupinan, B., and Schramm, V. L. (1991) Transition-state analysis of nucleoside hydrolase from *Crithidia fasciculata*, *Biochemistry 30*, 10788–10795.

[4] Iovane, E., Giabba, B., Muzzolini, L., Matafora, V., Fornili, A., Minici, C., Giannese, F., and Degano, M. (2008) Structural basis for substrate specificity in group I nucleoside hydrolases, *Biochemistry 47*, 4418–4426.

Appendix E: Supplemental Information for Chapter 6: An MD and QM/MM Study on the Substrate Scope of AlkB and ALKBH2: Insights into the Oxidative Repair of Etheno Lesions

Table of Contents

Table E.1. Average active site root-mean-squared deviation of AlkB bound to 1,N^6-εA, 3,N^4-εC, 1,N^2-εG, or N^2,3-εG over a 1 μs trajectory and two 500 ns replicate trajectories relative to the representative structure obtained by clustering replicate 1. 288

Table E.2. Average active site root-mean-squared deviation of ALKBH2 bound to 1,N^6-εA, 3,N^4-εC, 1,N^2-εG, or N^2,3-εG over a 1 μs trajectory and two 500 ns replicate trajectories relative to the representative structure obtained by clustering replicate 1. 289

Table E.3. Summary of important distances adopted from MD simulations of AlkB bound to 1,N^6-εA, 3,N^4-εC, 1,N^2-εG, or N^2,3-εG. 290

Table E.4. Summary of important distances adopted from MD simulations of ALKBH2 bound to 1,N^6-εA, 3,N^4-εC, 1,N^2-εG, or N^2,3-εG. 291

Table E.5. Summary of important hydrogen bonds from MD simulations on AlkB bound to 1,N^6-εA, 3,N^4-εC, 1,N^2-εG, or N^2,3-εG. 292

Table E.6. Summary of important hydrogen bonds from MD simulations on ALKBH2 bound to 1,N^6-εA, 3,N^4-εC, 1,N^2-εG, or N^2,3-εG. 293

Figure E.1. Overlay of the MD representative structure and the X-ray crystal structure of the AlkB complex bound to 1,N^6-εA-containing DNA, highlighting the metal-ligating residues. 294

Figure E.2. Distribution of active site water during MD simulations on the AlkB complex bound to 1,N^6-εA, 3,N^4-εC, 1,N^2-εG, or N^2,3-εG-containing DNA, highlighting water-mediated interactions between Tyr76 and the 5′-phosphate of the bound nucleotide. 295

Figure E.3. Overlays of representative structures of the AlkB or ALKBH2 complex bound to 1,N^6-εA and N^2,3-εG, or 1,N^6-εA, 3,N^4-εC and 1,N^2-εG, highlighting the position of the lesion nucleobase with respect to Trp69 (AlkB) or Phe124 (ALKBH2). 296

Figure E.4. MD representative structure of the ALKBH2 complex bound to N^2,3-εG, highlighting hydrogen-bonding interactions between active site residues and the bound nucleobase. 297

Figure E.5. Overlays of MD representative structures of the AlkB (gray) and ALKBH2 (cyan) complexes bound to 1,N^6-εA, 3,N^4-εC, 1,N^2-εG, and N^2,3-εG, highlighting the similar orientations of Trp69 (AlkB) and Phe124 (ALKBH2) with respect to the nucleobase. 298

Figure E.6. Overlays of MD representative structures of the AlkB and ALKBH2 complexes bound to 1,N^6-εA, 3,N^4-εC, 1,N^2-εG, and N^2,3-εG, highlighting the similar orientations of Asp135, Glu136, and Tyr78 in AlkB, and Glu175 and Tyr122 in ALKBH2. 299

Figure E.7. Histogram of the distance between the C10 and C11 bond midpoint of the AlkB-bound etheno lesion and Oδ of Asp135 Oε of Glu136 and Oη of Tyr78. 300

Figure E.8. Histogram of the distance between the C10 and C11 bond midpoint of the ALKBH2-bound etheno lesion and Oε of Glu175 and Oη of Tyr78. 301

Figure E.9. Overlay of the MD representative structure and X-ray crystal structure of the AlkB complex bound to 1,N^6-εA, highlighting the positions of Asp135, Tyr78, Glu136, and 1,N^6-εA, and the ALKBH2 complex bound to 1,N^6-εA, highlighting the consistent positions of Glu175, Tyr122, and 1,N^6-εA. 302

Figure E.10. Distribution of active site water during MD simulations of the AlkB complex bound to 1,N^6-εA, 3,N^4-εC, 1,N^2-εG, or N^2,3-εG, highlighting water-mediated interactions between the nucleobase and Asp135, Glu136, and Tyr78 in AlkB or Glu175 and Tyr122 in ALKBH2. 303

Figure E.11. Representative MD structures of the AlkB or ALKBH2 complex bound to N^2,3-εG colored based on the Eisenberg hydrophobicity scale, highlighting interactions between the nucleobase damaged site (C10/C11) and active site residues. 304

Figure E.12. Radius of gyration for the water density as a function of the distance between water and the bond midpoint of C10 and C11 for each etheno lesion bound to AlkB and ALKBH2. 305

Figure E.13. Overlays of AlkB MD representative structures and ONIOM(QM:MM) RC corresponding to the first step of oxidation for 1,N^6-εA, 3,N^4-εC, 1,N^2-εG, and N^2,3-εG. rmsd (Å) is calculated based on the heavy atoms of the bound nucleotide, Trp69, Asp135, His131, His187, succinate, Asp133, and the Fe(IV)-oxo moiety. 306

Full citation for references 1, 10, 28, 56, and 61: 307

Table E.1. Average active site root-mean-squared deviation (rmsd) (standard deviation (σ) provided in parentheses) of AlkB bound to 1,N^6-εA, 3,N^4-εC, 1,N^2-εG, or N^2,3-εG over a 1 μs trajectory and two 500 ns replicate trajectories relative to the representative structure obtained by clustering replicate 1.[a]

Replicate	1,N^6-εA	3,N^4-εC	1,N^2-εG	N^2,3-εG
	Avg (σ)	Avg (σ)	Avg (σ)	Avg (σ)
replicate 1 (500 ns)[b]	0.8 (0.3)	0.9 (0.2)	0.8 (0.2)	0.8 (0.2)
replicate 1 (1 μs)	0.7 (0.2)	0.9 (0.2)	0.7 (0.2)	0.8 (0.2)
replicate 2	1.0 (0.2)	1.0 (0.2)	1.3 (0.1)	1.1 (0.3)
replicate 3	1.2 (0.1)	1.0 (0.4)	1.4 (0.2)	1.0 (0.2)

[a]rms fit was performed based on the position of all heavy atoms of the lesion and active site residues (namely Fe(IV), Fe(IV)-ligating residues, and Asp135, Glu136, Trp69, and Arg210) to yield the average rmsd of the entire AlkB active site. [b]rms performed on the first 500 ns of replicate 1.

Table E.2. Average active site root-mean-squared deviation (rmsd) (standard deviation (σ) provided in parentheses) of ALKBH2 bound to 1,N^6-εA, 3,N^4-εC, 1,N^2-εG, or N^2,3-εG over a 1 μs trajectory and two 500 ns replicate trajectories relative to the representative structure obtained by clustering replicate 1.[a]

Replicate	1,N^6-εA	3,N^4-εC	1,N^2-εG	N^2,3-εG
	Avg (σ)	Avg (σ)	Avg (σ)	Avg (σ)
replicate 1 (500 ns)[b]	0.7 (0.3)	0.7 (0.2)	0.9 (0.2)	0.7 (0.2)
replicate 1 (1 μs)	0.8 (0.3)	0.7 (0.2)	1.1 (0.3)	0.7 (0.2)
replicate 2	1.1 (0.1)	1.5 (0.3)	1.2 (0.2)	0.9 (0.2)
replicate 3	1.1 (0.1)	0.8 (0.1)	1.0 (0.2)	0.8 (0.1)

[a]rms fit was performed based on the position of all heavy atoms of the lesion and active site residues (namely Fe(IV), Fe(IV)-ligating residues, and Phe124, Tyr122, Glu175, and Arg254) to yield the average rmsd of the entire ALKBH2 active site. [b]rms performed on the first 500 ns of the first replicate.

Table E.3. Summary of important average distances from MD simulations on AlkB bound to 1,N^6-εA, 3,N^4-εC, 1,N^2-εG, or N^2,3-εG.

	1,N^6-εA	3,N^4-εC	1,N^2-εG	N^2,3-εG
Distance	Avg (σ)	Avg (σ)	Avg (σ)	Avg (σ)
Nucleobase···His131[a]	3.8 (0.2)	4.1 (0.3)	4.3 (0.3)	4.1 (0.3)
Nucleobase···Tyr76[a]	5.4 (0.4)	4.7 (0.4)	5.0 (0.4)	4.5 (0.6)
Fe(IV)···His131	2.2 (0.1)	2.2 (0.1)	2.2 (0.1)	2.2 (0.1)
Fe(IV)···Asp133	3.4 (0.1)	3.4 (0.1)	3.4 (0.1)	3.4 (0.1)
Fe(IV)···His187	2.2 (0.1)	2.2 (0.1)	2.2 (0.1)	2.2 (0.1)
Fe(IV)···oxo	1.6 (0.0)	1.6 (0.0)	1.6 (0.0)	1.6 (0.0)
Fe(IV)···H_2O	1.9 (0.1)	1.9 (0.1)	1.9 (0.1)	1.9 (0.1)
Fe(IV)···succinate	1.9 (0.1)	1.9 (0.1)	1.9 (0.1)	1.9 (0.1)
Nucleobase···Trp69[a]	5.4 (0.6)	5.8 (0.6)	5.1 (0.3)	6.6 (0.5)
Fe(IV)-oxo···C10	3.3 (0.3)	3.4 (0.3)	3.6 (0.4)	4.0 (0.3)
Fe(IV)-oxo···C11	4.3 (0.3)	3.9 (0.5)	3.4 (0.3)	4.3 (0.3)
Fe(IV)···C10	4.5 (0.2)	4.5 (0.3)	5.0 (0.3)	5.3 (0.3)
Fe(IV)···C11	5.2 (0.3)	4.9 (0.4)	5.0 (0.3)	5.3 (0.3)
C10/C11···Asp135(Oδ)[b]	6.1 (1.4)	8.0 (1.1)	9.6 (0.7)	9.8 (1.3)
C10/C11···Tyr78(Oη)[b]	5.0 (1.0)	5.6 (1.8)	4.6 (0.6)	12.1 (2.5)
C10/C11···Glu136(Oε)[b]	7.1 (0.9)	6.3 (0.9)	8.1 (0.5)	11.4 (3.1)

[a]Distance between the center-of-mass (COM) of the cyclic nucleobase atoms and cyclic aromatic atoms of the active site residue. [b]Distance between C10/C11 bond midpoint and Oδ (Asp135), Oη (Tyr78), and Oε (Glu136).

Table E.4. Summary of important average distances from MD simulations on ALKBH2 bound to 1,N^6-εA, 3,N^4-εC, 1,N^2-εG, or N^2,3-εG.

	1,N^6-εA	3,N^4-εC	1,N^2-εG	N^2,3-εG
Distance	Avg (σ)	Avg (σ)	Avg (σ)	Avg (σ)
Nucleobase···His171[a]	4.5 (0.7)	4.1 (0.3)	4.1 (0.6)	3.5 (0.2)
Fe(IV)···His171	2.2 (0.1)	2.2 (0.1)	2.2 (0.1)	2.2 (0.1)
Fe(IV)···Asp173	3.4 (0.1)	3.4 (0.1)	3.4 (0.1)	3.4 (0.1)
Fe(IV)···His236	2.2 (0.1)	2.3 (0.1)	2.2 (0.1)	2.2 (0.1)
Fe(IV)···oxo	1.6 (0.0)	1.6 (0.0)	1.6 (0.0)	1.6 (0.0)
Fe(IV)···H_2O	1.9 (0.1)	1.9 (0.1)	1.9 (0.1)	1.9 (0.1)
Fe(IV)···succinate	1.5 (0.1)	1.9 (0.1)	1.9 (0.1)	1.9 (0.1)
Nucleobase···Phe124[a]	4.0 (0.4)	4.7 (0.4)	4.9 (1.3)	8.3 (2.1)
Nucleobase···Arg100[a]	8.8 (0.4)	8.3 (0.4)	7.7 (0.9)	6.5 (0.7)
Fe(IV)-oxo···C10	3.4 (0.7)	3.4 (0.4)	3.8 (0.7)	5.2 (0.4)
Fe(IV)-oxo···C11	4.5 (0.6)	3.9 (0.4)	3.8 (0.7)	5.6 (0.4)
Fe(IV)···C10	4.5 (0.7)	4.8 (0.4)	5.1 (0.7)	5.8 (0.4)
Fe(IV)···C11	5.3 (0.7)	5.1 (0.3)	5.6 (0.9)	6.1 (0.3)
C10/C11···Glu175(Oε1)[b]	6.6 (1.9)	9.7 (2.4)	10.8 (1.5)	11.3 (1.8)
C10/C11···Tyr122(Oη)[b]	3.3 (0.4)	4.1 (0.3)	7.0 (1.5)	9.2 (2.1)

[a]Distance between the center-of-mass (COM) of the cyclic nucleobase atoms and cyclic aromatic atoms of the active site residue. [b]Distance between C10/C11 bond midpoint and Oδ (Asp135), Oη (Tyr78), and Oε (Glu136).

Table E.5. Summary of important hydrogen bonds from MD simulations on AlkB bound to 1,N^6-εA, 3,N^4-εC, 1,N^2-εG, or N^2,3-εG.

Hydrogen Bond	1,N^6-εA			3,N^4-εC			1,N^2-εG			N^2,3-εG		
	%[a,b]	Dist[c]	Angle[c]	%[a,b]	Dist[c]	Angle[c]	%[a,b]	Dist[c]	Angle[c]	%[a,b]	Dist[c]	Angle[c]
dX(O1P)···Tyr76(OηH)	71.5%	2.8	163	0.3%	2.8	166	7.0%	2.7	165	0.4%	2.8	165
dX(N2)···Arg210(NηH)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	17.3%	3.1	136
dX(O1P)···H_2O(OH)	100.0%	2.8	157	100.0%	2.8	159	100.0%	2.8	158	100.0%	2.8	159
dX(N4)···H_2O(OH)	N/A	N/A	N/A	100.0%	2.9	156	N/A	N/A	N/A	N/A	N/A	N/A
dX(O2)···H_2O(OH)	N/A	N/A	N/A	72.6%	2.9	154	N/A	N/A	N/A	N/A	N/A	N/A
dX(N6)···H_2O(OH)	100.0%	3.0	155	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
dX(N7)···H_2O(OH)	100.0%	3.0	151	N/A	N/A	N/A	100.0%	3.0	151	35.8%	3.0	153
dX(O6)···H_2O(OH)	N/A	N/A	N/A	N/A	N/A	N/A	100.0%	2.8	155	100.0%	2.9	151
dX(N2)···H_2O(OH)	N/A	N/A	N/A	N/A	N/A	N/A	5.6%	3.2	135	98.1%	2.9	153
H_2O(O)···dX(N2H)	N/A	N/A	N/A	N/A	N/A	N/A	100.0%	3.1	143	N/A	N/A	N/A
H_2O(O)···dX(N1H)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	100.0%	3.1	143
Asp135(Oδ)···H_2O(OH)	100.0%	2.8	158	100.0%	2.8	158	100.0%	2.8	158	100.0%	2.8	158
Glu136(Oε)···H_2O(OH)	100.0%	2.8	158	86.7%	2.7	162	82.7%	2.7	160	100.0%	2.8	159
Tyr78(Oη)···H_2O(OH)	32.3%	3.0	145	39.7%	3.0	152	4.8%	3.0	151	43.9%	3.0	152
H_2O(O)···Tyr76(OηH)	63.1%	3.1	151	100.0%	3.0	158	100.0%	3.0	158	100.0%	3.0	158
oxo(O)···H_2O(OH)	12.4%	3.0	140	75.5%	3.0	143	83.5%	3.0	142	31.2%	3.0	141
H_2O(O)···Tyr78(OηH)	100.0%	3.0	150	100.0%	3.0	153	100.0%	3.0	158	98.2%	3.0	152

[a]Percent occupancy of the hydrogen bond, which is defined using cutoffs of a distance less than 3.4 Å and an angle greater than 120°. [b]Greater than 100% occupancy indicates more than one water hydrogen bonds with an acceptor site. [c]Hydrogen-bonding distance in Å and angle in degrees.

Table E.6. Summary of important hydrogen bonds from MD simulations on ALKBH2 bound to 1,N^6-εA, 3,N^4-εC, 1,N^2-εG, or N^2,3-εG.

Hydrogen Bond	1,N^6-εA			3,N^4-εC			1,N^2-εG			N^2,3-εG		
	%[a,b]	Dist[c]	Angle[c]	%[a,b]	Dist[c]	Angle[c]	%[a,b]	Dist[c]	Angle[c]	%[a,b]	Dist[c]	Angle[c]
dX(O1P)···Ser125(OγH)	97.1%	2.7	163.6	98.8%	2.7	163.3	83.1%	2.7	163.5	16.2%	2.7	164.0
dX(N2)···Asn169(NδH)	N/A	N/A	N/A	N/A	N/A	N/A	6.1%	3.2	138.8	70.4%	3.0	144.9
dX(O6)···Arg254(NηH)	N/A	N/A	N/A	N/A	N/A	N/A	40.9%	2.9	147.8	74.0%	2.9	143.6
dX(O2)···H_2O(OH)	N/A	N/A	N/A	100.0%	2.9	151.9	N/A	N/A	N/A	N/A	N/A	N/A
dX(N4)···H_2O(OH)	N/A	N/A	N/A	40.2%	3.0	157.0	N/A	N/A	N/A	N/A	N/A	N/A
dX(N6)···H_2O(OH)	98.6%	3.0	155.2	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
dX(N7)···H_2O(OH)	65.1%	3.0	151.9	N/A	N/A	N/A	100.0%	3.0	153.1	85.8%	3.0	150.6
dX(O6)···H_2O(OH)	N/A	N/A	N/A	N/A	N/A	N/A	93.9%	2.9	151.9	88.1%	2.9	152.6
dX(N2)···H_2O(OH)	N/A	N/A	N/A	N/A	N/A	N/A	7.7%	3.2	130.9	23.7%	3.0	155.1
Glu175(Oε)···H_2O(OH)	100.0%	2.8	158.0	100.0%	2.8	158.5	100.0%	2.8	158.2	100.0%	2.8	158.0
H_2O(O)···Tyr122(OηH)	100.0%	3.0	150.0	91.0%	3.0	149.9	100.0%	3.0	154.9	100.0%	3.0	152.7

[a]Percent occupancy of the hydrogen bond, which is defined using cutoffs of a distance less than 3.4 Å and an angle greater than 120°. [b]Greater than 100% occupancy indicates more than one water hydrogen bonds with an acceptor site. [c]Hydrogen-bonding distance in Å and angle in degrees.

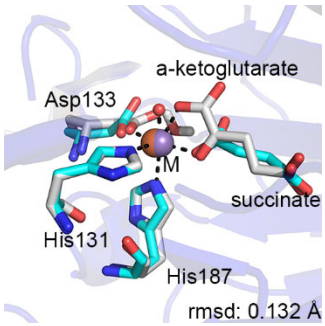


Figure E.1. Overlay of the MD representative structure (cyan) and the X-ray crystal structure (gray; PDB ID: 3O1U) of the AlkB complex bound to 1,N^6-εA-containing DNA, highlighting the metal-ligating residues.

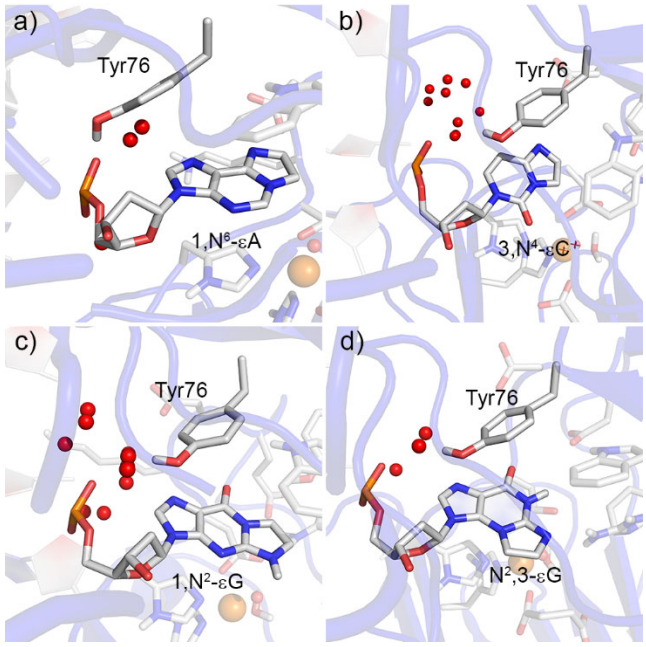


Figure E.2. Distribution of active site water (red spheres) during MD simulations on the AlkB complex bound to a) 1,N^6-εA, b) 3,N^4-εC, c) 1,N^2-εG, or d) N^2,3-εG-containing DNA, highlighting water-mediated interactions between Tyr76 and the 5′-phosphate of the bound nucleotide.

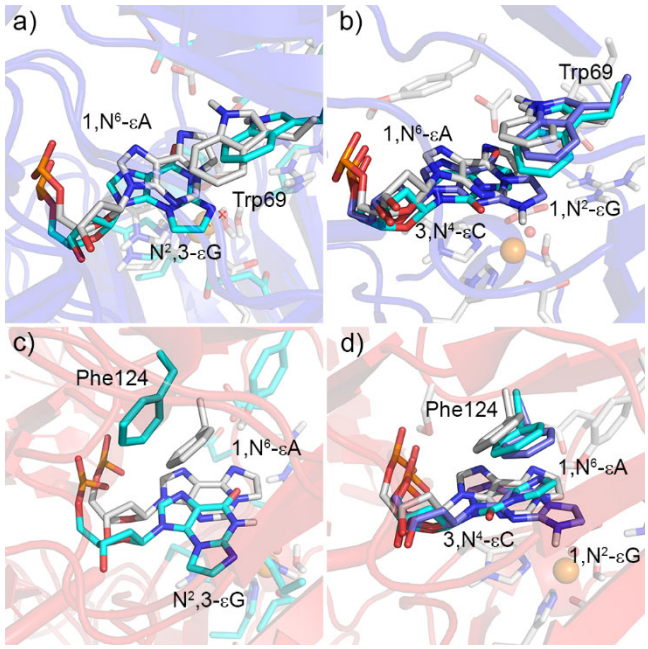


Figure E.3. Overlays of MD representative structures of the AlkB (a,b) or ALKBH2 (c,d) complex bound to a,c) 1,N^6-εA (gray) and N^2,3-εG (cyan), or b,d) 1,N^6-εA (gray), 3,N^4-εC (cyan) and 1,N^2-εG (dark blue), highlighting the position of the lesion nucleobase with respect to Trp69 (AlkB) or Phe124 (ALKBH2).

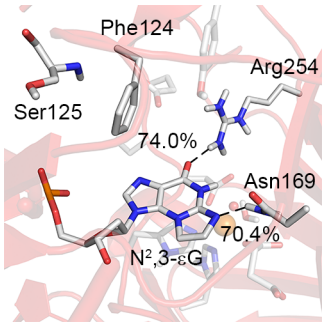


Figure E.4. MD representative structure of the ALKBH2 complex bound to N^2,3-εG, highlighting hydrogen-bonding interactions between active site residues and the bound nucleobase.

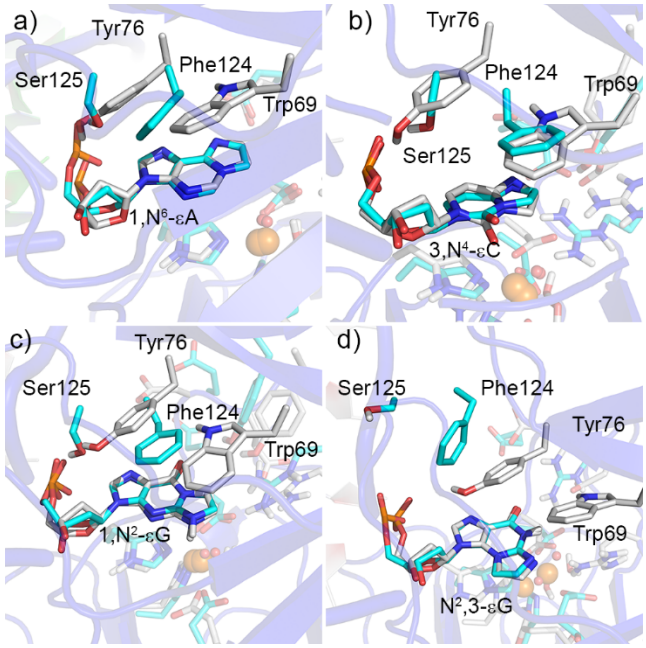


Figure E.5. Overlays of MD representative structures of the AlkB (gray) and ALKBH2 (cyan) complexes bound to a) 1,N^6-εA, b) 3,N^4-εC, c) 1,N^2-εG, and d) N^2,3-εG, highlighting the similar orientations of Trp69 (AlkB) and Phe124 (ALKBH2) with respect to the nucleobase.

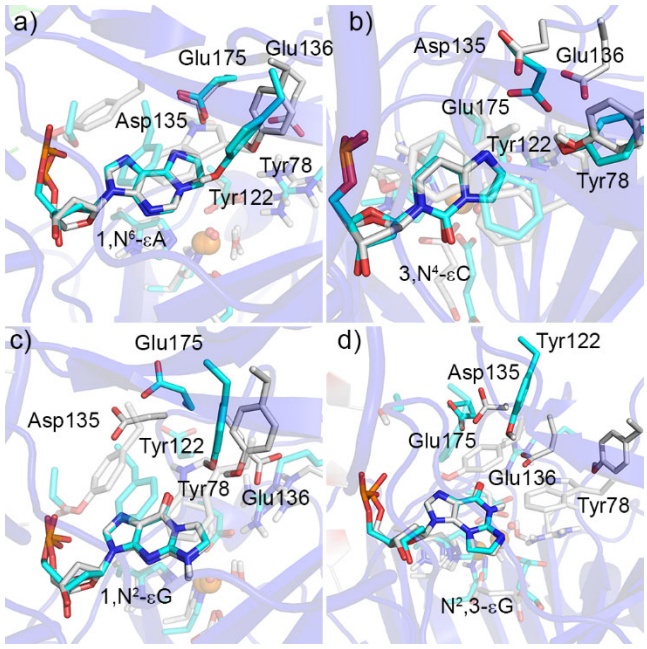


Figure E.6. Overlays of MD representative structures of the AlkB (gray) and ALKBH2 (cyan) complexes bound to a) 1,N^6-εA, b) 3,N^4-εC, c) 1,N^2-εG, and d) N^2,3-εG, highlighting the similar orientations of Asp135, Glu136, and Tyr78 in AlkB, and Glu175 and Tyr122 in ALKBH2.

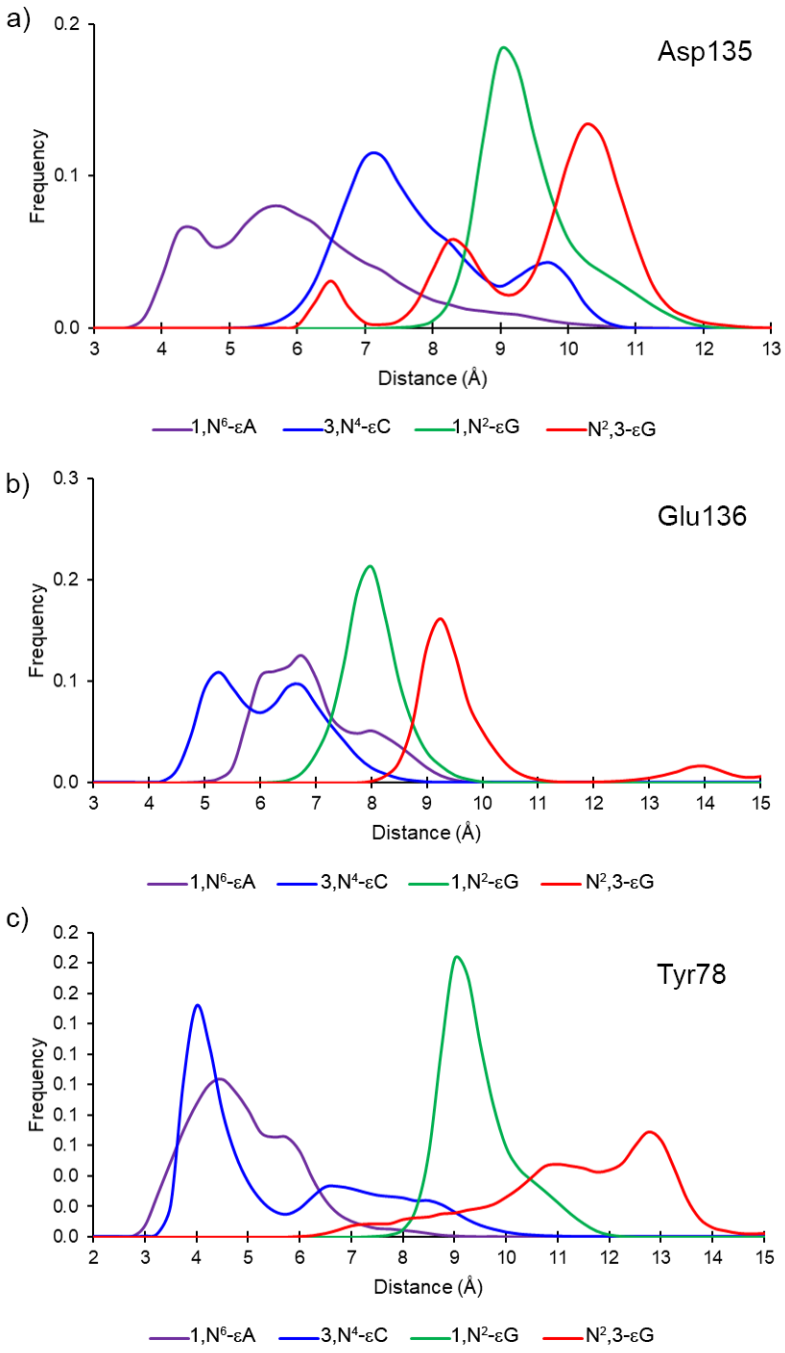


Figure E.7. Histogram of the distance between the C10 and C11 bond midpoint of the AlkB-bound etheno lesion and a) Oδ of Asp135 b) Oε of Glu136 and c) Oη of Tyr78.

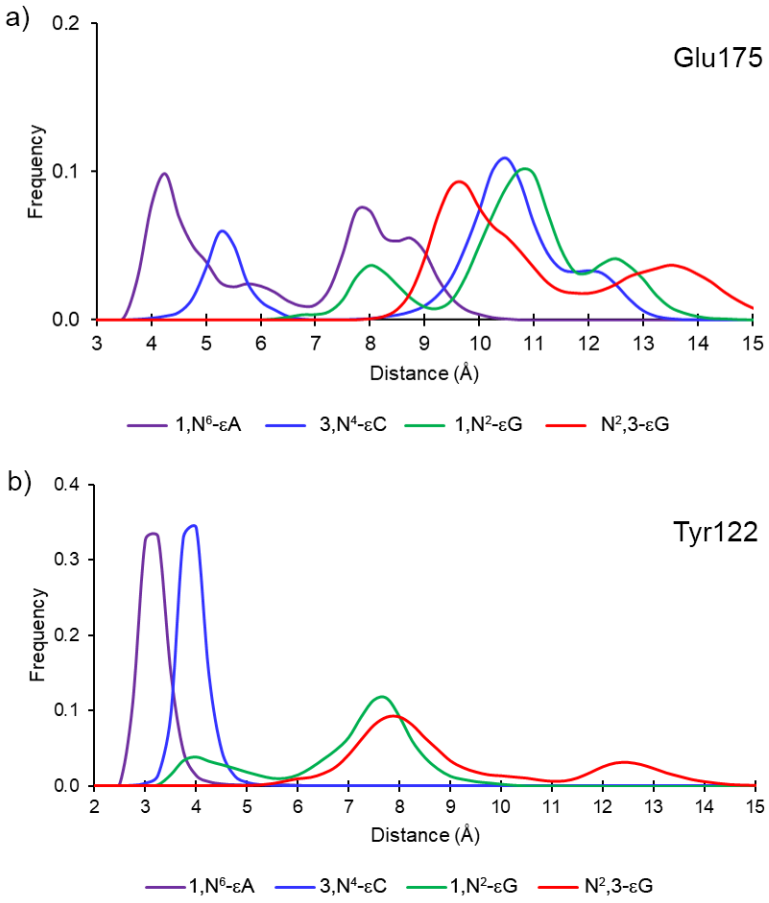


Figure E.8. Histogram of the distance between the C10 and C11 bond midpoint of the ALKBH2-bound etheno lesion and a) Oε of Glu175 and b) Oη of Tyr78.

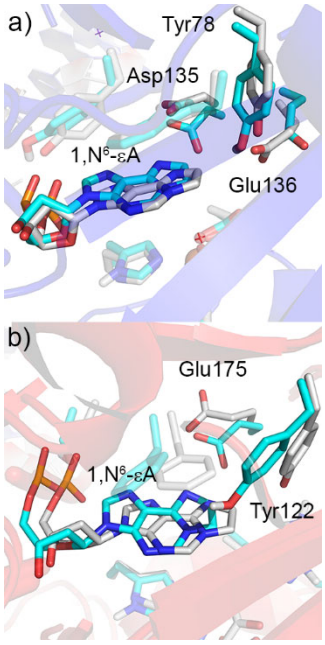


Figure E.9. Overlay of the MD representative structure (cyan) and X-ray crystal structure of a) the AlkB complex (gray; PDB ID: 3O1U) bound to 1,N^6-εA, highlighting the consistent positions of Asp135, Tyr78, Glu136, and 1,N^6-εA, and b) the ALKBH2 complex (gray; PDB ID: 3RZK) bound to 1,N^6-εA, highlighting the consistent positions of Glu175, Tyr122, and 1,N^6-εA.

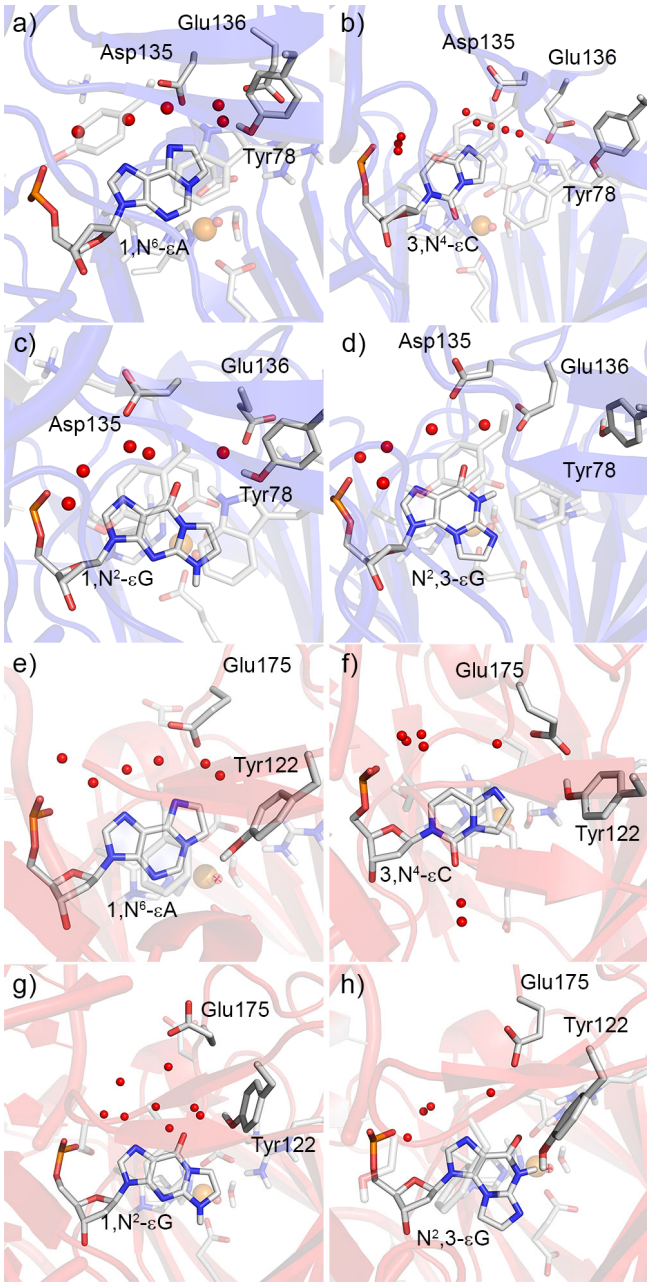


Figure E.10. Distribution of active site water (red spheres) during MD simulations of the AlkB complex bound to a) 1,N^6-εA, b) 3,N^4-εC, c) 1,N^2-εG, or d) N^2,3-εG, highlighting water-mediated interactions between the nucleobase and Asp135, Glu136, and Tyr78 in AlkB or Glu175 and Tyr122 in ALKBH2.

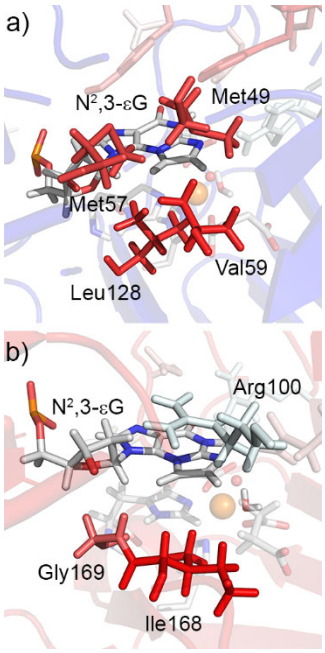


Figure E.11. Representative MD structures of the a) AlkB or b) ALKBH2 complex bound to N^2,3-εG colored based on the Eisenberg hydrophobicity scale, highlighting interactions between the nucleobase damaged site (C10/C11) and active site residues.

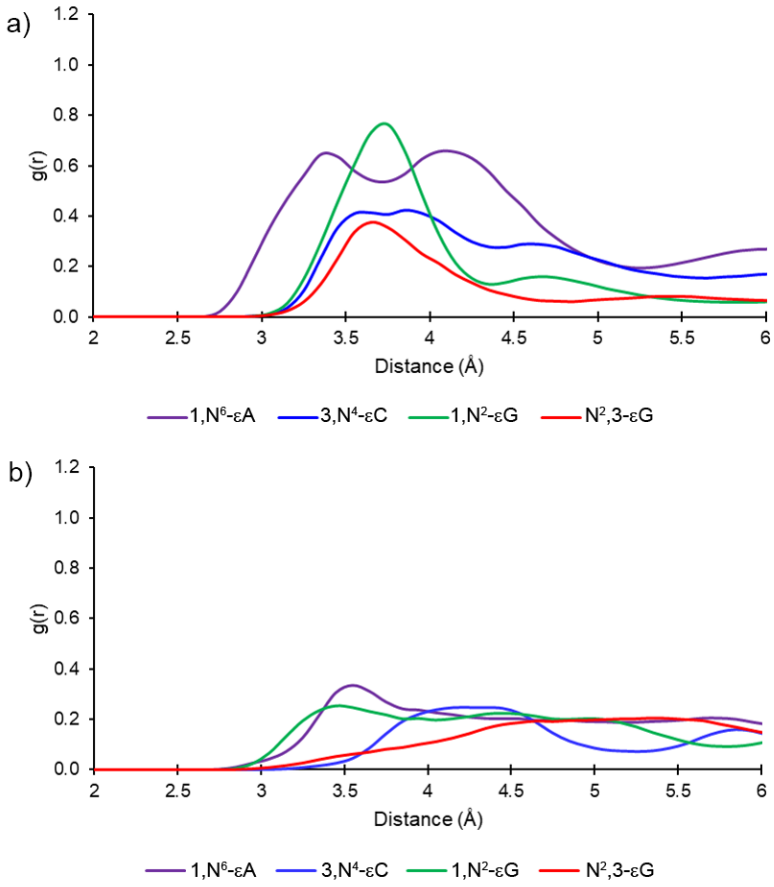


Figure E.12. Radius of gyration for the water density as a function of the distance between water and the bond midpoint of C10 and C11 for each etheno lesion bound to a) AlkB and b) ALKBH2.

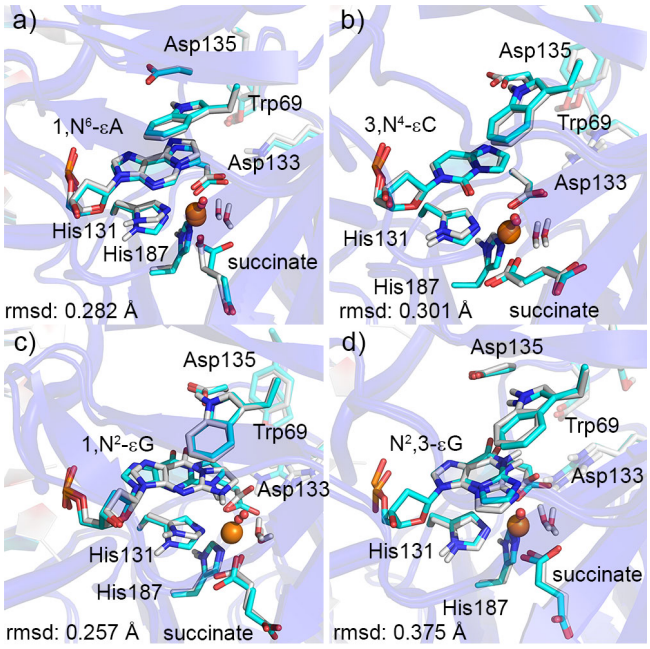


Figure E.13. Overlays of AlkB MD representative structures (gray) and ONIOM(QM:MM) RC (cyan) corresponding to the first step of oxidation for a) 1,N^6-εA, b) 3,N^4-εC, c) 1,N^2-εG, and d) N^2,3-εG. rmsd (Å) is calculated based on the heavy atoms of the bound nucleotide, Trp69, Asp135, His131, His187, succinate, Asp133, and the Fe(IV)-oxo moiety.

Full citation for references 1, 10, 28, 56, and 61:

1. Knijnenburg, T. A.; Wang, L.; Zimmermann, M. T.; Chambwe, N.; Gao, G. F.; Cherniack, A. D.; Fan, H.; Shen, H.; Way, G. P.; Greene, C. S.; Liu, Y.; Akbani, R.; Feng, B.; Donehower, L. A.; Miller, C.; Shen, Y.; Karimi, M.; Chen, H.; Kim, P.; Jia, P.; Shinbrot, E.; Zhang, S.; Liu, J.; Hu, H.; Bailey, M. H.; Yau, C.; Wolf, D.; Zhao, Z.; Weinstein, J. N.; Li, L.; Ding, L.; Mills, G. B.; Laird, P. W.; Wheeler, D. A.; Shmulevich, I.; The Cancer Genome Atlas Research, N.; Monnat, R. J.; Xiao, Y.; Wang, C., Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas. *Cell reports* **2018,** *23* (1), 239-254.e6

10. Obtułowicz, T.; Winczura, A.; Speina, E.; Swoboda, M.; Janik, J.; Janowska, B.; Cieśla, J. M.; Kowalczyk, P.; Jawien, A.; Gackowski, D.; Banaszkiewicz, Z.; Krasnodebski, I.; Chaber, A.; Olinski, R.; Nair, J.; Bartsch, H.; Douki, T.; Cadet, J.; Tudek, B., Aberrant repair of etheno–DNA adducts in leukocytes and colon tissue of colon cancer patients. *Free Radical Biology and Medicine* **2010,** *49* (6), 1064-1071.

28. Ringvoll, J.; Moen, M. N.; Nordstrand, L. M.; Meira, L. B.; Pang, B.; Bekkelund, A.; Dedon, P. C.; Bjelland, S.; Samson, L. D.; Falnes, P. Ø.; Klungland, A., AlkB homologue 2-mediated repair of ethenoadenine lesions in mammalian DNA. *Cancer research* **2008,** *68* (11), 4142-4149.

56. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Jr., J. A. M.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 16*, Revision B.01; Gaussian, Inc.: Wallingford CT, 2018.

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

Appendix F: Supplemental Information for Chapter 7: DNA Repair Enzymes ALKBH2, ALKBH3, and AlkB Oxidize 5-Methylcytosine to 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxylcytosine[a]

Table of Contents

Section F.1. Experimental Procedures 310
Section F.2. Computational Methodology 311
Section F.3. Product Oligonucleotides Analyses 313
Section F.4. Product Distribution for the Oxidation of 5mC 313
Section F.5. Simulations of ALKB2 and AlkB bound to 3mC, 5hmC, and 5fC 314
Figure F.1. Structures and chemical numbering of 3mC and 5mC.. 315
Figure F.2. ESI-TOF analyses of standard oligonucleotide containing 5mC and reaction mixtures. 316
Figure F.3. HPLC analyses of deoxyribonucleoside standards and benzonase digested product oligonucleotides. 317
Figure F.4. MS analysis of standard 5mC deoxyribonucleoside under negative ion mode.. 318
Figure F.5. MS analysis of digested oligonucleotide containing 5mC deoxyribonucleoside under negative ion mode.. 319
Figure F.6. MS analysis of standard 5hmC deoxyribonucleoside under negative ion mode.. 320
Figure F.7. MS analysis of digested oligonucleotide containing 5hmC deoxyribonucleoside under negative ion mode 321
Figure F.8. MS analysis of standard 5fC deoxyribonucleoside under negative ion mode.. 322
Figure F.9. MS analysis of digested oligonucleotide containing 5fC deoxyribonucleoside under negative ion mode.. 323
Figure F.10. MS analysis of standard 5caC deoxyribonucleoside under negative ion mode.. 324
Figure F.11. MS analysis of digested oligonucleotide containing 5caC deoxyribonucleoside under negative ion mode 325
Figure F.12. MS analyses of trypsin digested proteins including the wild type and mutant AlkB enzymes. 326
Figure F.13. ESI-TOF analyses of standard oligonucleotide containing 5mC and reaction mixtures with wild type and mutant proteins.. 327
Figure F.14. Distribution of active site water during MD simulations of ALKBH2 bound to *anti*-5mC or *syn*-5mC. ALKBH2 bound to *syn*-5mC highlighting key hydrogen-bonding interactions. Overlay of MD representative structures of ALKBH2 bound to *anti*- or *syn*-5mC. 328

[a]*The Journal of the American Chemical Society* reference style is used in this Appendix.

Figure F.15. Distribution of active site water during MD simulations of AlkB bound to *anti*-5mC or *syn*-5mC. AlkB bound to *syn*-5mC highlighting key water mediated hydrogen-bonding interactions. Overlay of MD representative structures of AlkB bound to *anti*- or *syn*-5mC. 329

Figure F.16. Distribution of active site water during MD simulations of ALKBH2 bound to *anti*-3mC. ALKBH2 bound to *anti*-3mC. Overlay of MD representative structures of ALKBH2 bound to *syn*-5mC and *anti*-3mC. 330

Figure F.17. Distribution of active site water during MD simulations of AlkB bound to *anti*-3mC. AlkB bound to *anti*-3mC. Overlay of MD representative structures of AlkB bound to *syn*-5mC and *anti*-3mC. 331

Figure F.18. Overlay of the 5mC nucleotide isolated from MD representative structures of AlkB–5mC and ALKBH2–5mC complexes onto a crystal structure of TET2 bound to 5mC-containing DNA, Fe(II), and α-KG. 332

Figure F.19. Structures and chemical numbering of *anti*-3mC, *syn*-5mC, *syn*-5hmC, and *syn*-5fC. 333

Figure F.20. Distribution of active site water during MD simulations of ALKBH2 bound to *syn*-5hmC, ALKBH2 bound to *syn*-5fC, AlkB bound to *syn*-5hmC, or AlkB bound to *syn*-5hmC. 334

Figure F.21. ALKBH2 bound to *syn*-5hmC, and AlkB bound to *syn*-5hmC. Histogram displaying the occupancy of θ during simulations of ALKBH2 or AlkB bound to 5hmC. 335

Figure F.22. *syn*-5fC bound by ALKBH2 or AlkB, highlighting the distance between the C5 substituent and the Fe(IV)-oxo moiety. 336

Table F.1. Calculated and observed monoisotopic molecular weight and m/z value of modified oligonucleotides. 337

Table F.2. Calculated and observed monoisotopic molecular weight and m/z value of peptide fragments digested by trypsin. 338

Table F.3. Individual product percentage of 5hmC, 5fC and 5caC in the total oxidative product mixture from ds- or ss-5mC reactions with ALKBH2, ALKBH3 or AlkB. 338

Table F.4. Percentage of total oxidation product including 5hmC, 5fC and 5caC from ds- or ss-5mC reactions with ALKBH2, ALKBH3 or AlkB. 338

Table F.5. Summary of important hydrogen bonds formed during MD simulations of anti-3mC, anti-5mC, *syn*-5mC, *syn*-5hmC, and *syn*-5fC in the ALKBH2 complex. 339

Table F.6. Summary of important hydrogen bonds formed during MD simulations of *anti*-3mC, *anti*-5mC, *syn*-5mC, *syn*-5hmC, and *syn*-5fC in the AlkB complex. 340

Table F.7. Summary of important distances and dihedral angles adopted during MD simulations of *anti*-3mC, *anti*-5mC, *syn*-5mC, *syn*-5hmC, and *syn*-5fC in the ALKBH2 complex. 341

Table F.8. Summary of important distances and dihedral angles adopted during MD simulations of *anti*-3mC, *anti*-5mC, *syn*-5mC, *syn*-5hmC, and *syn*-5fC in the AlkB complex. 341

References. 341

Section F.1. Experimental Procedures

Synthesis of oligonucleotides containing 5mC. All oligodeoxynucleotides used in this study were synthesized by solid-phase synthesis.[1–4] The 5mC phosphoramidites was purchased from Glen Research. Synthetic oligodeoxynucleotides were purified by reverse-phase HPLC and identified by electrospray ionization mass spectrometry (ESI-MS).

Protein expression and purification. The expression and purification of ALKBH2, ALKBH3 and AlkB proteins were described previously by published papers.[1–4] ALKBH2 and ALKBH3 in storage buffer containing 50 mM N-[tris(hydroxymethyl)methyl]-3-aminopropanesulfonic acid (TAPS), 300 mM NaCl, 10% glycerol, and 1 mM 2-mercaptoethanol, pH 8.0, and AlkB in similar storage buffer (10 mM Tris, 100 mM NaCl, 1 mM 2-mercaptoethanol, 10% glycerol, pH 8.0), were all stored at -80°C.

Enzymatic reaction. The reactions were performed based on previously published procedures.[1–4] A 16mer oligonucleotide containing one 5mC with sequence of 5'-GAAGACCT-5mC-GGCGTCC-3' was used as the substrate. One hundred pmol 16mer 5mC was mixed into reaction buffer (5 μM $Fe(NH_4)_2(SO_4)_2$, 0.93 mM α-ketoglutarate, 1.86 mM ascorbic acid, and 46.5 mM TAPS, pH 7.5) in a total volume of 16 μL. The reactions were started by adding proteins (170 pmol AlkB, 140 pmol ALKBH2 or 128 pmol ALKBH3), and incubated at 37 °C for 1 hour. EDTA was employed to quench the reactions and the reaction mixture was immediately heated up to 80 °C for 5min. A 16/23mer duplex DNA was preannealed as the substrate for dsDNA reaction.[1,4] The 23mer oligonucleotide is complementary to the 16mer sequence plus 7 nucleotides longer with the sequence of 5'-CTGGGACGCCGAGGTCTTCACTG-3'. The rest of the steps of dsDNA reaction were the same as for the ssDNA reaction described above. All reactions were carried out in triplicates.

Oligonucleotide digestion. The procedures of oligonucleotide digestion to deoxyribonucleoside were adopted from published literature.[5] A digestion mixture was premade by adding 250 Units Benzonase (Sigma-Aldrich, MO), 300 mUnits phosphodiesterase I (Sigma-Aldrich, MO) and 200 Units alkaline phosphatase (Sigma-Aldrich, MO) to 5 mL Tris-HCl buffer (20 mM, pH 7.9) containing 100 mM NaCl and 20 mM $MgCl_2$. Reaction of oligonucleotide containing 5mC with AlkB was quenched by heating up to 80°C for 5 min, then digested by adding 50 μL digestion mixture and incubating at 37°C for 6 hours. The nucleoside products were analyzed by HPLC-MS. Samples were chromatographed on a Luna Omega Polar C18 column (150 X 4.6 mm, 5μm, 100Å, Phenomenex, CA) eluted at 1 mL/min with an acetonitrile gradient (1%-15%). ESI triple quadrupole time of flight mass spectrometry was conducted to detect nucleoside signals in the negative ion mode. Standard deoxyribonucleosides (including dA, dT, dC, dG, 5mdC, 5hmdC, 5fdC and 5cadC) were purchased from Berry & Associates (MI).

Protein digestion. The trypsin digestion of wide type AlkB and its mutants were set up according to Trypsin-ultraTM kit (New England Biolabs, MA). Ten μg of AlkB proteins were mixed with 250 ng trypsin

in trypsin-ultra buffer, incubating at 37 °C overnight. The protein fragments were analyzed by HPLC-MS.[1,4] Standard protein fragments (including LFYHGIQPLK and LSLHQDK sequences) were purchased from New England Peptide (MA).

Preparing catalytically inactive mutants of AlkB. The AlkB H131A, D133A, and H187A mutants were generated by QuikChange Mutagenesis (Agilent), using pET24a-AlkBΔN11 as the PCR template and primer pairs encoding the desired mutations. All three mutants were purified essentially as described.[6] Briefly, One Shot BL21 Star (DE3)pLysS *E. coli* cells (Invitrogen) transformed with an AlkB mutant construct were grown at 37 °C until OD_{600} had reached ~0.4, at which point the temperature was lowered to 30 °C and protein production was induced by addition of 1 mM IPTG. Cells were harvested after 4 h and stored at −80 °C until use. For purification, cell pellets were resuspended in lysis buffer (10 mM Tris, pH 7.3, 300 mM NaCl, 2 mM $CaCl_2$,10 mM $MgCl_2$, 5% (v/v) glycerol, 1 mM 2-mercaptoethanol) and lysed by sonication. After clarification by centrifugation, the lysate was loaded onto a Ni-NTA column (Qiagen), the column was washed twice with lysis buffer supplemented with 10 mM and 20 mM imidazole, and bound protein was eluted with lysis buffer supplemented with 70 mM and 250 mM imidazole. Elution fractions containing AlkB, as assessed by SDS-PAGE, were combined and dialyzed for 16 h against 50 mM 2-[[1,3-dihydroxy-2-(hydroxymethyl)propan-2-yl]amino]ethanesulfonic acid (TES), pH 7.1 with or without 20 mM NaCl, and loaded onto a 5 mL HiTrap SP cation exchange column (GE Healthcare). Bound AlkB was eluted with a linear gradient of 0.02 – 1 M NaCl over 12 column volumes (60 mL). Fractions containing pure AlkB were pooled and purity was established by SDS-PAGE.

Section F.2. Computational Methodology

X-ray crystal structures of AlkB (PDB ID: 3O1S)[7] and ALKBH2 (PDB ID: 3RZJ)[8] bound to lesion-containing DNA were chosen as initial structures. The 3O1S crystal structure contains AlkB bound to an oxidized 3mC intermediate-containing DNA, Fe(II), and succinate. The oxidized intermediate was reverted to 3mC, and a water molecule bound to Fe(II) was modelled as an oxo ligand to generate the Fe(IV)–oxo complex. To generate the active Fe(IV)–oxo ALKBH2 complex, the cofactors (Fe(IV), succinate, and oxo ligand) and metal-binding amino acids (H131, D133, and H187) of the AlkB complex were superimposed onto the cofactors (Mn(II) and α-KG) and metal-binding amino acids (H171, D173, and H236) of the ALKBH2 complex. Hydrogen atoms were assigned using the tLEaP AMBER utility.[9] Protonation states of ionizable amino acids were initially assigned using H++[10] and adjusted using chemical intuition. For AlkB, H66, H97, H172, and H197 were assigned epsilon protonation, while H72, H131, and H187 were assigned delta protonation. For ALKBH2, H55, H59, H106, H144, and H228 were assigned epsilon protonation, H199 and H220 were assigned delta protonation, and H167 was modelled as cationic. All crosslinks induced to facilitate crystallization between DNA and protein were removed for both complexes. The G169C, C67S, C165S, and C192S mutations to ALKBH2 were reverted to generate the wild-type enzyme, and the overhanging DNA ends (residues 259, 260, and 284) were truncated. To generate the AlkB–5mC or ALKBH2–5mC complexes, the *anti* and *syn* conformations of the 5mC nucleotide were overlaid onto the 3mC substrate in the ALKBH2- or AlkB-DNA complex using chemical intuition to minimize steric clashes between the bound substrate and active site amino acids.

Each complex was modelled using the AMBER parm14SB force field. Parameters assigned to the nonstandard 5mC and 3mC nucleotides were supplemented with GAFF parameters, and Restrained Electrostatic Potential (RESP) charges. For Fe(IV) and the iron-ligating residues, the Metal Center Parameter Builder[11] was used to assign RESP charges, and bonding, angle, dihedral and non-bonding parameters based on B3LYP/LANL2DZ (Fe(IV)) and B3LYP/6-31G(d,p) (H, C, N, and O) optimized structures (Gaussian 09 revision D.01)[12] of the iron-binding site using the Seminario method.[13] The complexes were neutralized with Na^+ counter ions, and solvated in a water box such that at least 10.0 Å of water exists between the DNA–protein complex and water box boundary.

All minimization, heating, equilibration, and production steps were performed with the GPU-accelerated PMEMD module available in AMBER 14 or AMBER 16.[14-16] For each step, the particle mesh Ewald method was used to employ a 10.0 Å electrostatic cutoff. To minimize each system, 1000 cycles of steepest descent (SD) and subsequent 1000 cycles of conjugate gradient (CG) minimization were first applied to the solvent and Na^+ counter ions, followed by the complex hydrogen atoms, and finally the complex heavy/hydrogen atoms. For the final minimization step, the entire system was subjected to 1000 cycles of SD minimization followed by 2000 cycles of CG minimization. Subsequently, each system was heated to 310 K over 120 ps with a 25.0 kcal mol^{-1} $Å^{-2}$ restraint placed on the solute using the Langevin thermostat with a time step of 1 fs under NVT conditions. The 25.0 kcal mol^{-1} $Å^{-2}$ restraint was reduced to 0.0 kcal mol^{-1} $Å^{-2}$ over 100 ps under constant temperature and pressure conditions (Berendsen barostat) using a 2 fs time step and SHAKE to constrain bonds involving hydrogen, which was followed by 1 ns of unrestrained equilibration. Several pre-production simulations were performed for each system to ensure adequate sampling and accommodation of each substrate within the active site of ALKBH2 or AlkB. Representative structures (clustering methodology below) were chosen from these pre-production simulations to initiate the final 100 ns production simulations. Coordinates were saved from each simulation every 5 ps and analyzed over the same interval.

The CPPTRAJ[17] module was utilized for analysis of each trajectory. The occupancy of hydrogen bonds is reported for the duration of the simulation that the heavy atom distance is < 3.4 Å and the hydrogen-bonding angle is > 120°. To examine the placement of water within the active site, a three dimensional 20 $Å^3$ grid centered on the bound substrate was generated. Dark red spheres are shown on representative structures and reflect an oxygen atom of a water molecule located in a 0.25 $Å^3$ grid space for at least 40% of the simulation. The pre-production and production trajectories were clustered according to the positions of key active site residues (bound substrate, W69, H131, D135 and E136 for AlkB, and bound substrate, F124, H171, D174 and E175 for ALKBH2). A hierarchical agglomerative clustering methodology was used to obtain a 3.0 Å minimum cutoff between clusters or four clusters were obtained. This methodology is a bottom-up approach where each frame starts as a unique cluster and clusters are merged based on the active site rmsd similarity. While the representative structures are static snapshots, we report key structural parameters over the entire trajectory to ensure that conclusions are representative of each simulation trajectory. All distances are reported in Å and all angles are reported in degrees.

Section F.3. Product Oligonucleotides Analyses

It was important to assure that the observed new oligonucleotide species indeed contain the proposed oxidative products. The product oligonucleotides were digested with benzonase into single nucleosides, which were analyzed by LC-MS and compared with standard nucleosides for retention time and molecular weight. In the digestion of the product generated from the AlkB reaction, a nucleoside was eluted out at 7.3 min in HPLC (Figure F.2) and had an observed m/z 240.236 in the negative mode of MS analysis (theoretical m/z 240.239, Figure F.4); these data agree well with the analysis of the standard sample of 5mC nucleoside (7.3 min in HPLC and 240.239 in m/z, Figure F.3). Another nucleoside from digestion had a retention time at 4.9 min in HPLC (Figure F.2) and had an m/z 256.239 in the MS analysis (Figure F.6); these data are comparable to the standard sample of 5hmC nucleoside (5.0 min in HPLC and 256.241 in m/z, theoretical m/z 256.238, Figures F.2 and F.5). Similarly, nucleosides 5fC and 5caC were also discovered in the mixture of the digested reaction product and their identities were confirmed by comparing to standard nucleosides (Figures F.2, F.7 to F.10). These observations support that the oxidative products generated from enzymatic reactions are indeed 5hmC, 5fC, and 5caC.

Section F.4. Product Distribution for the Oxidation of 5mC

For the oxidation of 5mC, the formation of products 5hmC, 5fC, and 5caC had different distribution patterns for the three enzymes; and the three enzymes had different preferences of oxidation in ss- or ds-DNA reactions (Figure 3). The relative amount of each product in the final reaction mixture was quantified according to the abundance of its corresponding ion at -3 charge state in the MS analysis. The product oligonucleotides containing 5hmC was the dominant species for the reactions of ALKBH2 and 3 in both ds- and ss-DNA (Figures 7.3a and 7.3b). For example, 5hmC represents 78.7% of the total amount of the three oxidative products in the reaction of ALKBH2 with 5mC in ds-DNA; 5fC and 5caC are quantified as 10.9% and 10.4% correspondingly (Figure 7.3a and Table F.3). For the oxidation of 5mC by AlkB, 5hmC is the major species (51.2%) in ds-DNA reaction; however, 5fC is the most abundant species (65.3%) in the ss-DNA reaction (Figures 3a and 3b, Table F.3). We were also able to quantify the yield of the oxidative products by comparing them to the total oligonucleotide species. The ratios of the oxidative products show ALKBH2 prefers to oxidize 5mC in the ds-DNA (8.7%) over in ss-DNA (5.5%), and ALKBH3 prefers to modify 5mC in ss-DNA (22.1%) over in ds-DNA (4.2%) (Figure 3c and Table F.4). These results support the strand preference in repair reactions of ALKBH2 (preferring ds-DNA) and ALKBH3 (preferring ss-DNA) reported in the literature.[3] For AlkB, the enzyme oxidized 5mC in ss-DNA (23.5%) more efficiently than in ds-DNA (13.3%) (Figure 3c and Table F.4); these results are similar to the preference of AlkB repairing 3mC in ss-DNA.[3]

Section F.5. Simulations of ALKB2 and AlkB bound to 3mC, 5hmC, and 5fC

To validate that the simulated structures of the ALKBH2 or AlkB complexes are consistent with the observed catalytic activity, we examined the active site conformations adopted upon binding of both enzymes to *anti*-3mC-, *syn*-5hmC-, or *syn*-5fC-containing DNA. Similar DNA–protein interactions form when 3mC and 5mC are bound. Specifically, the 3mC N^4 amino group forms direct hydrogen bonds with the D174 or E175 sidechains when bound to ALKBH2 (Figure F.16b and Table F.5), and direct or water-mediated hydrogen bonds with D135 and E136 when bound to AlkB (Table F.6 and Figure F.17). As a result, the N3 methyl group of 3mC occupies an equivalent active site position as the C5 methyl group of 5mC for both enzymes (Figures S16b and S17b), with a distance between the Fe(IV)–oxo and the 3mC methyl moiety of ~ 3.3 Å (Tables S7-S8) and the *syn*-5mC moiety of ~3.6 Å for AlkB and 3.8 Å for AlkB. Thus, the simulation data are consistent with the proposal that the AlkB family of enzymes is able to oxidize 5mC in the *syn* orientation.

Similarly to ALKBH2 or AlkB bound to 5mC, *syn*-5hmC is stabilized by direct or water-mediated hydrogen bonds between the N^4 amino group of the nucleobase and the carboxylate moieties of D174 and E175 (D135 and E136 in AlkB; Tables S5-S6 and Figure F.20). For 5hmC bound by ALKBH2 or AlkB, hydrogen bonds are formed between the C5 substituent hydroxy group and the Fe(IV)-oxo moiety (occupancy = 34% for ALKBH2 and 52% for AlkB; Tables S5-S6 and Figure F.21). In ALKBH2, an additional hydrogen bond is formed between the C5 substituent hydroxy group and D173 (59%, Table F.10), which results in one orientation of the C5 substituent throughout the simulation (Figure F.21a). In contrast, the C5 substituent hydroxy group does not interact with D133 in AlkB, which results in two conformations of 5hmC within the active site (Figure F.21b-d). More importantly, the C5 substituent is at an optimum distance from the Fe(IV)-oxo moiety in AlkB (~3.4 Å), which matches the prototypic substrate (3mC; ~3.3 Å), while the equivalent distance is longer in the ALKBH2–5hmC complex (~4.5 Å; Tables S7-S8 and Figure 3), which would impede catalysis. This helps explain the observed higher abundance of 5fC for AlkB compared to ALKBH2 catalyzed oxidation, although several other factors could also be significant such as DNA binding and unique base flipping mechanism for each enzyme.

As discussed for 5mC and 5hmC, the *syn* conformation of 5fC is stabilized by hydrogen bonds between the N^4 amino group of the nucleobase and carboxylate sidechains of active site residues (Tables S5-S6 and Figure F.20). When bound to either ALKBH2 or AlkB, 5fC is planar due to an intramolecular hydrogen bond between the N^4 amino group and the carbonyl of the C5 substituent. A hydrogen bond also forms between the carbonyl of the C5 substituent and an active site arginine, which is notably more persistent for AlkB (R210, occupancy = 100.0%) than ALKBH2 (R254, occupancy = 29.2%, Tables S9-S10 and Figure F.20). As a result, the distance between the hydrogen atom of the C5 substituent and the Fe(IV)-oxo group is longer for AlkB (~3.8 Å) compared to ALKBH2 (3.3 Å; Tables S7-S8 and Figure F.22). Although the difference in distance for 5fC bound by ALKBH2 and AlkB is not as significant as observed for 5hmC, QM/MM studies on TET2-catalyzed oxidation of 5mC, 5hmC, and 5fC reveal that the initial hydrogen atom abstraction step is rate limiting, with the barrier increasing as 5mC < 5hmC < 5fC.[19] This suggests that the position of the substituent relative to the Fe(IV)-oxo group is even more crucial in the case of 5fC. Thus, our predicted structures correlate with the lower abundance of 5caC relative to 5fC for AlkB compared to ALKBH2 (Figure 3). Nevertheless, as discussed for 5hmC, nucleotide recognition and the base flipping mechanism could also play significant roles in oxidative catalysis.

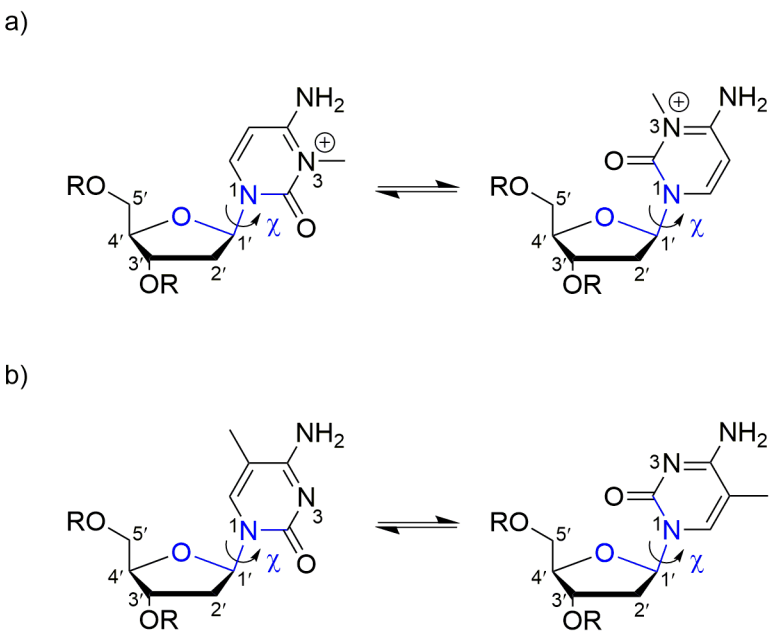


Figure F.1. Structures and chemical numbering of a) 3mC and b) 5mC. The glycosidic bond orientation is defined by dihedral angle χ (blue, ∠O4′C1′N1C2) as either *anti* (left, 180 ± 90°) or *syn* (right, 0 ± 90°).

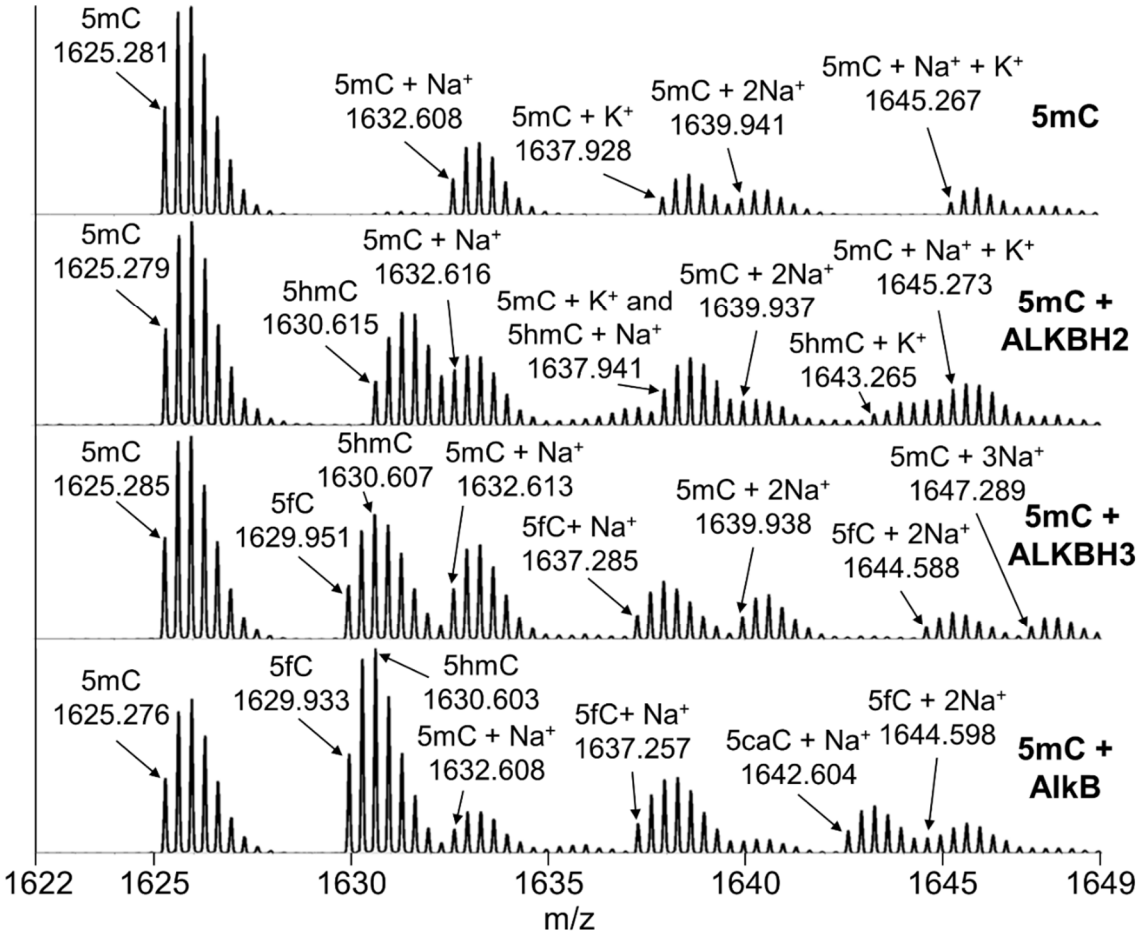


Figure F.2. ESI-TOF analyses of standard oligonucleotide containing 5mC and reaction mixtures. The observed m/z values represent the oligonucleotides under their -3 charge state. The theoretical m/z values of the corresponding species are listed in Table F.1.

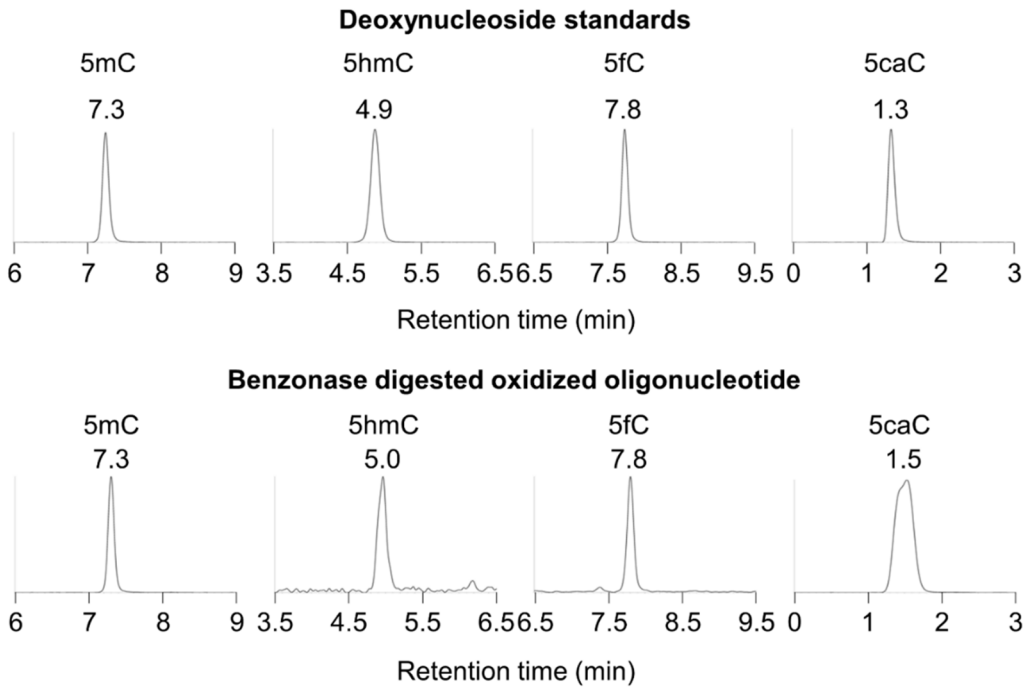


Figure F.3. HPLC analyses of deoxyribonucleoside standards and benzonase digested product oligonucleotides.

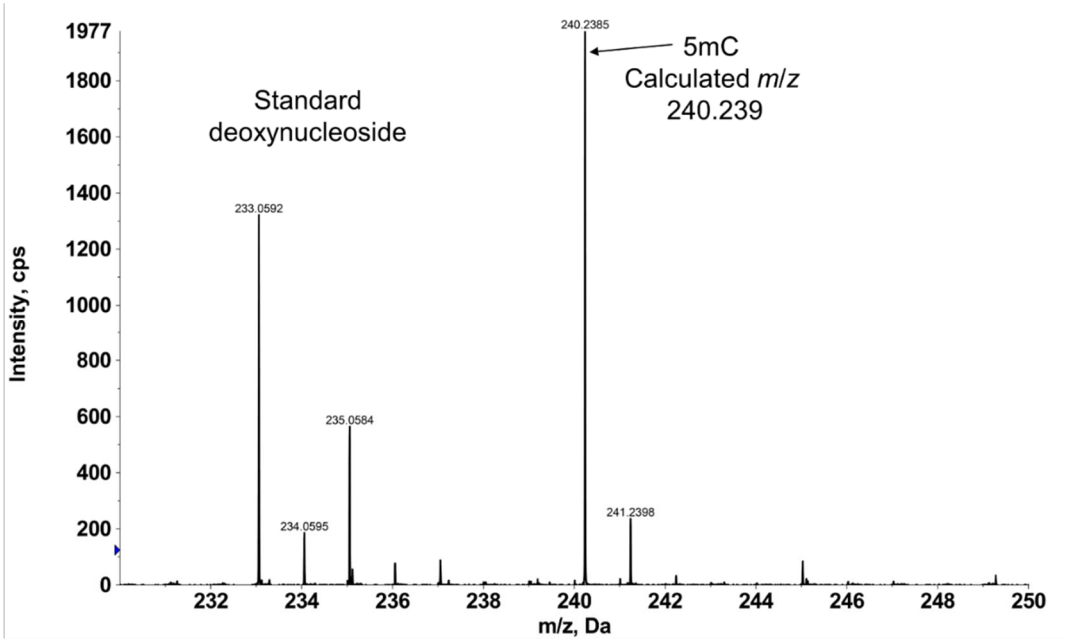


Figure F.4. MS analysis of standard 5mC deoxyribonucleoside under negative ion mode. The theoretical m/z value of the corresponding species is listed next to the observed peak.

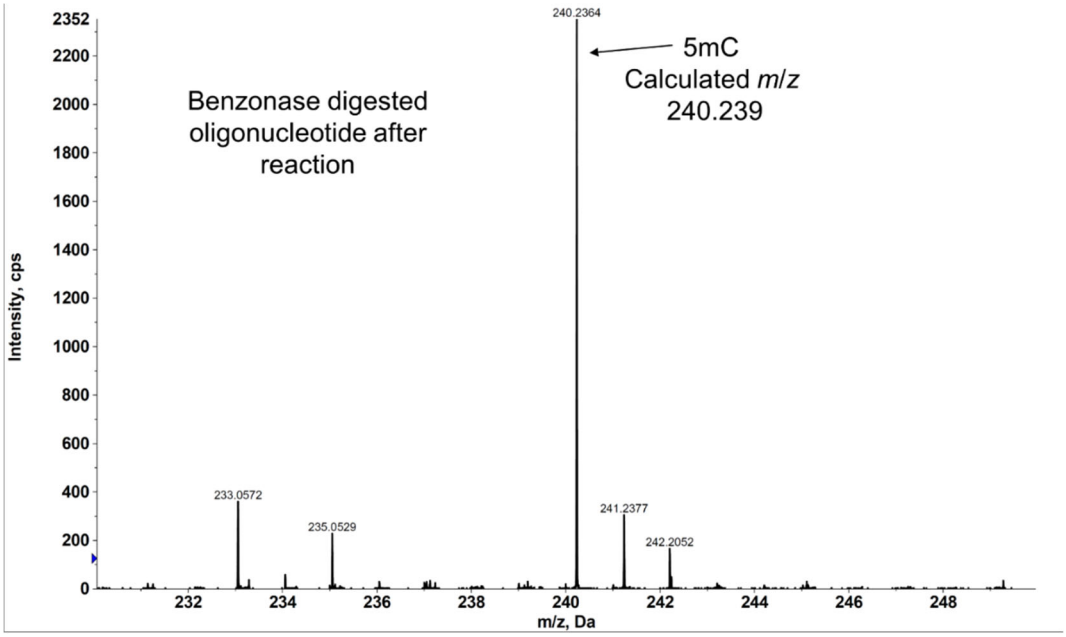


Figure F.5. MS analysis of digested oligonucleotide containing 5mC deoxyribonucleoside under negative ion mode. The theoretical m/z value of the corresponding species is listed next to the observed peak.

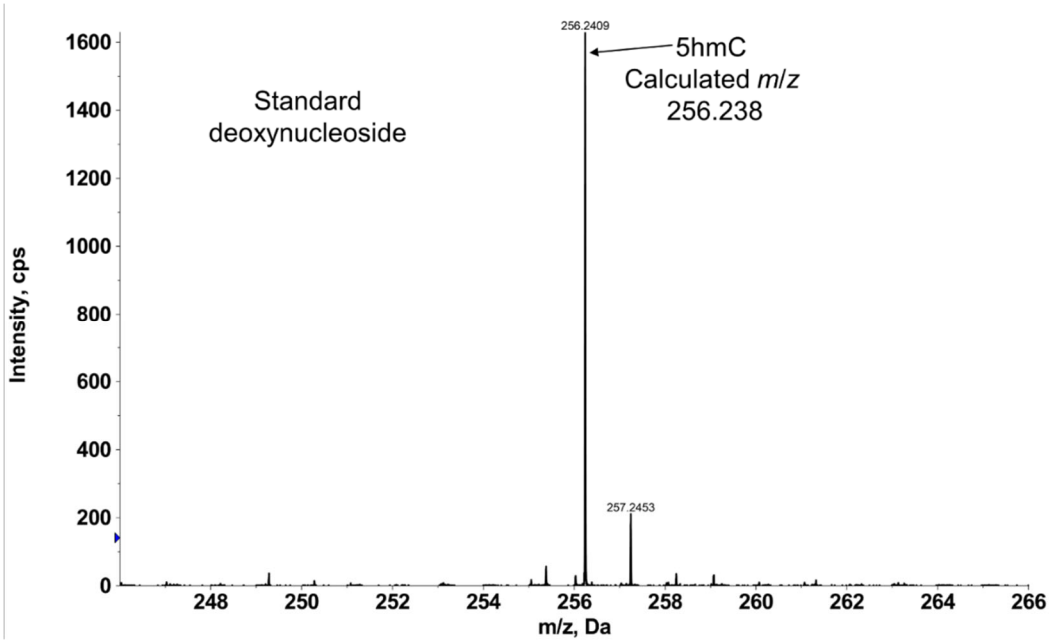


Figure F.6. MS analysis of standard 5hmC deoxyribonucleoside under negative ion mode. The theoretical m/z value of the corresponding species is listed next to the observed peak.

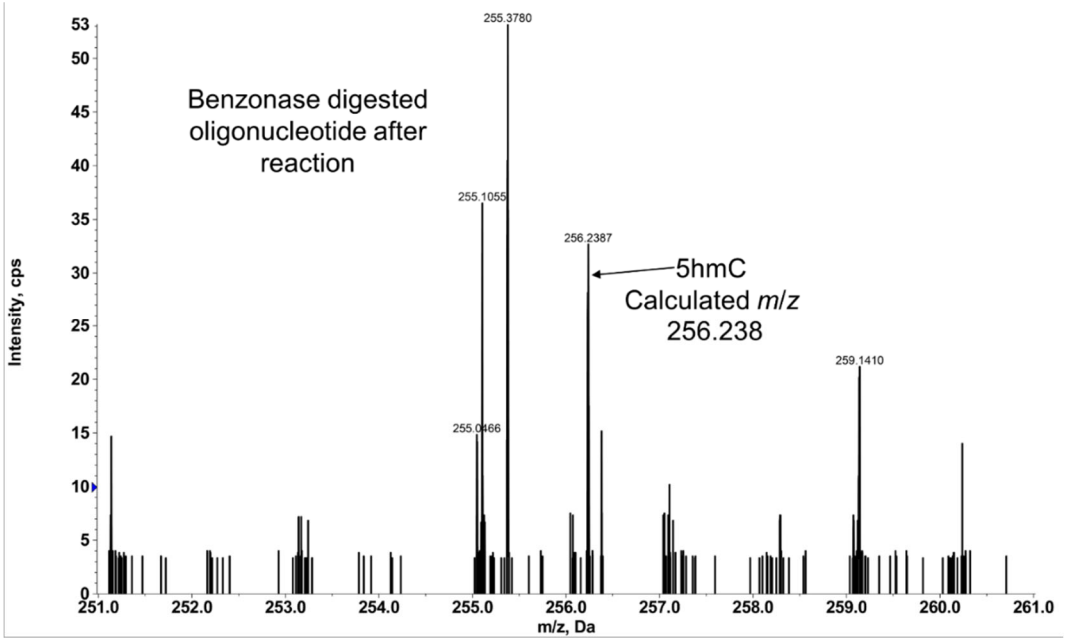


Figure F.7. MS analysis of digested oligonucleotide containing 5hmC deoxyribonucleoside under negative ion mode. The theoretical m/z value of the corresponding species is listed next to the observed peak.

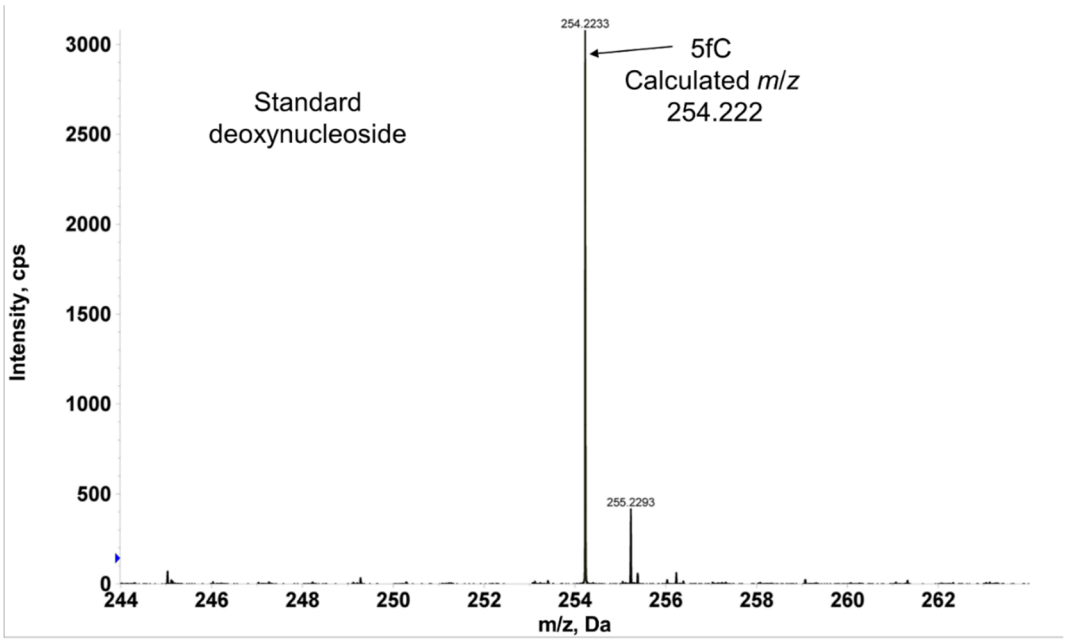


Figure F.8. MS analysis of standard 5fC deoxyribonucleoside under negative ion mode. The theoretical m/z value of the corresponding species is listed next to the observed peak.

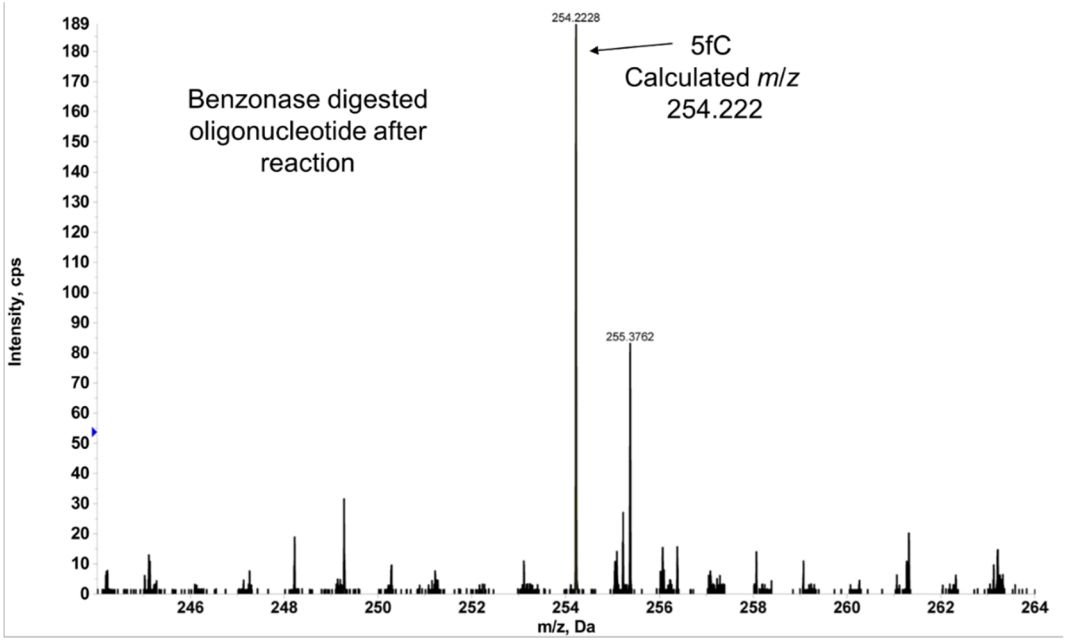


Figure F.9. MS analysis of digested oligonucleotide containing 5fC deoxyribonucleoside under negative ion mode. The theoretical m/z value of the corresponding species is listed next to the observed peak.

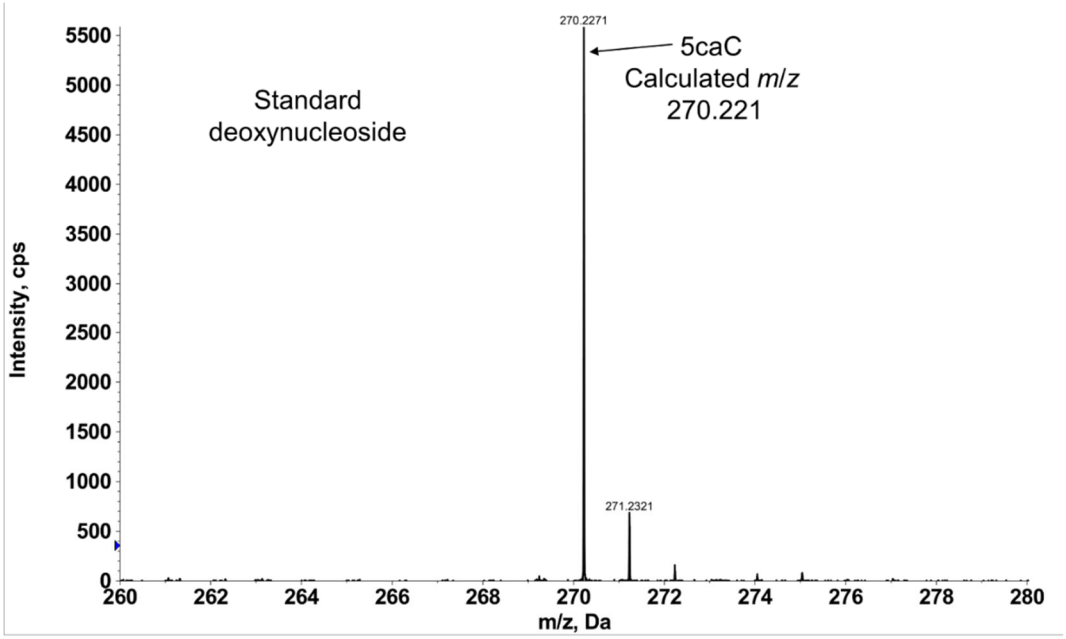


Figure F.10. MS analysis of standard 5caC deoxyribonucleoside under negative ion mode. The theoretical m/z value of the corresponding species is listed next to the observed peak.

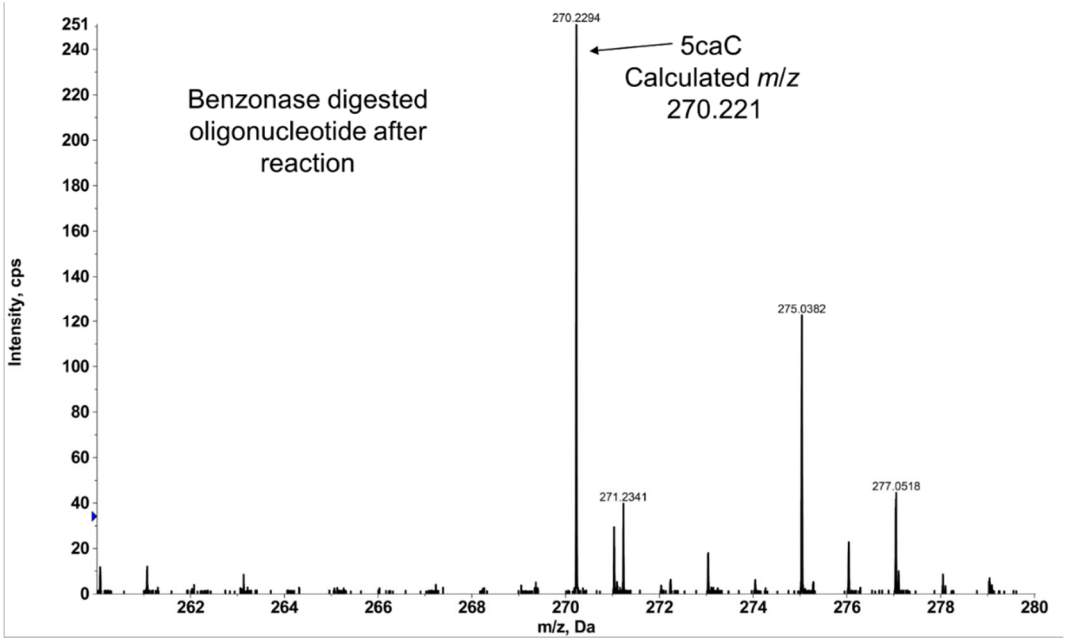


Figure F.11. MS analysis of digested oligonucleotide containing 5caC deoxyribonucleoside under negative ion mode. The theoretical m/z value of the corresponding species is listed next to the observed peak.

AlkB sequence

MLDLFADAEP WQEPLAAGAV ILRRFAFNAA EQLIRDINDV ASQSPFRQMV TPGGYTMSVA
MTNCGHLGWT THRQGYLYSP IDPQTNKPWP AMPQSFHNLC QRAATAAGYP DFQPDACLIN
RYAPGAKLSL HQDKDEPDLR APIVSVSLGL PAIFQFGGLK RNDPLKRLLL EHGDVVVWGG
ESRLFYHGIQ PLKAGFHPLT IDCRYNLTFR QAGKKE

Peptide fragments from trypsin digestion

Mass	Position	Peptide sequence
3415.6230	74-102	QGYLYSPIDPQTNKPWPAMPQSFHNLCQR
2835.2713	48-73	QMVTPGGYTMSVAMTNCGHLGWTTHR
2525.2988	1-23	MLDLFADAEPWQEPLAAGAV ILR
2013.1662	141-160	APIVSVSLGLPAIFQFGGLK
1992.9363	103-121	AATAAGYPDFQPDACLINR
1764.9158	168-183	LLLEHGDVVVWGGESR
1347.6419	36-47	DINDVASQSPFR
1278.6720	25-35	FAFNAAEQLIR
1228.6022	194-204	AGFHPLTIDCR
1214.6811	184-193	LFYHGIQPLK
839.4501	128-134	LSLHQDK
812.4181	205-210	YNLTFR
743.3450	135-140	DEPDLR
605.3173	122-127	YAPGAK
585.3122	162-166	NDPLK

MS analyses of digested protein fragments

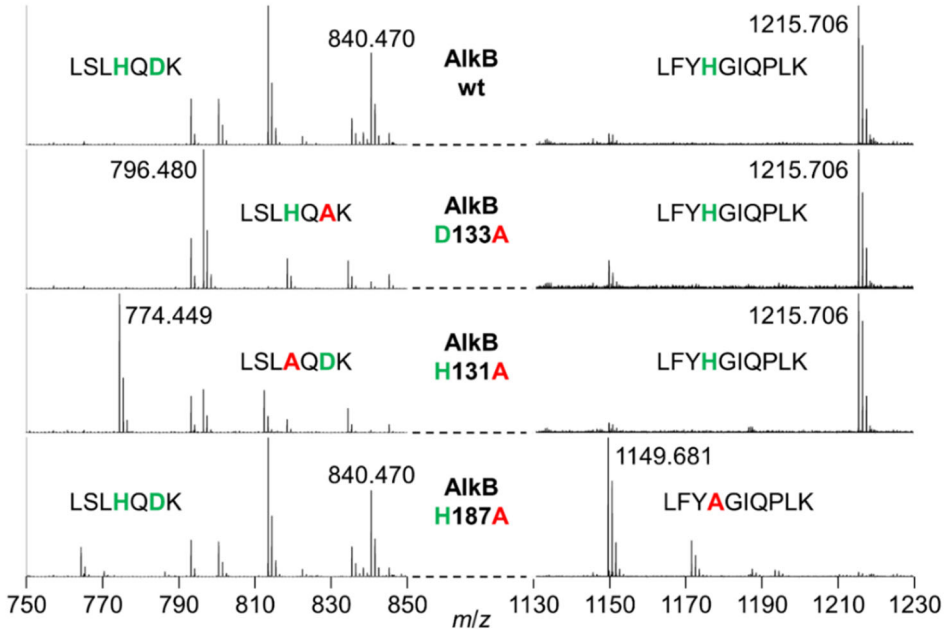


Figure F.12. MS analyses of trypsin digested proteins including the wild type and mutant AlkB enzymes.

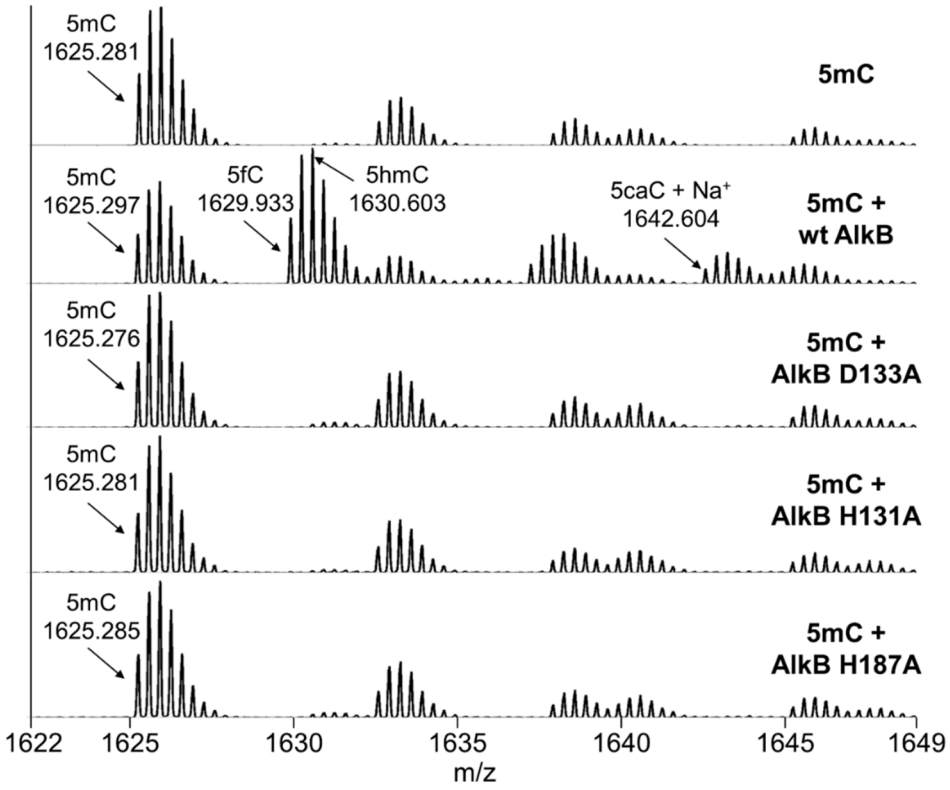


Figure F.13. ESI-TOF analyses of standard oligonucleotide containing 5mC and reaction mixtures with wild type and mutant proteins. The observed m/z values represent the oligonucleotides under their -3 charge state. The theoretical m/z values of the corresponding species are listed in Table F.1.

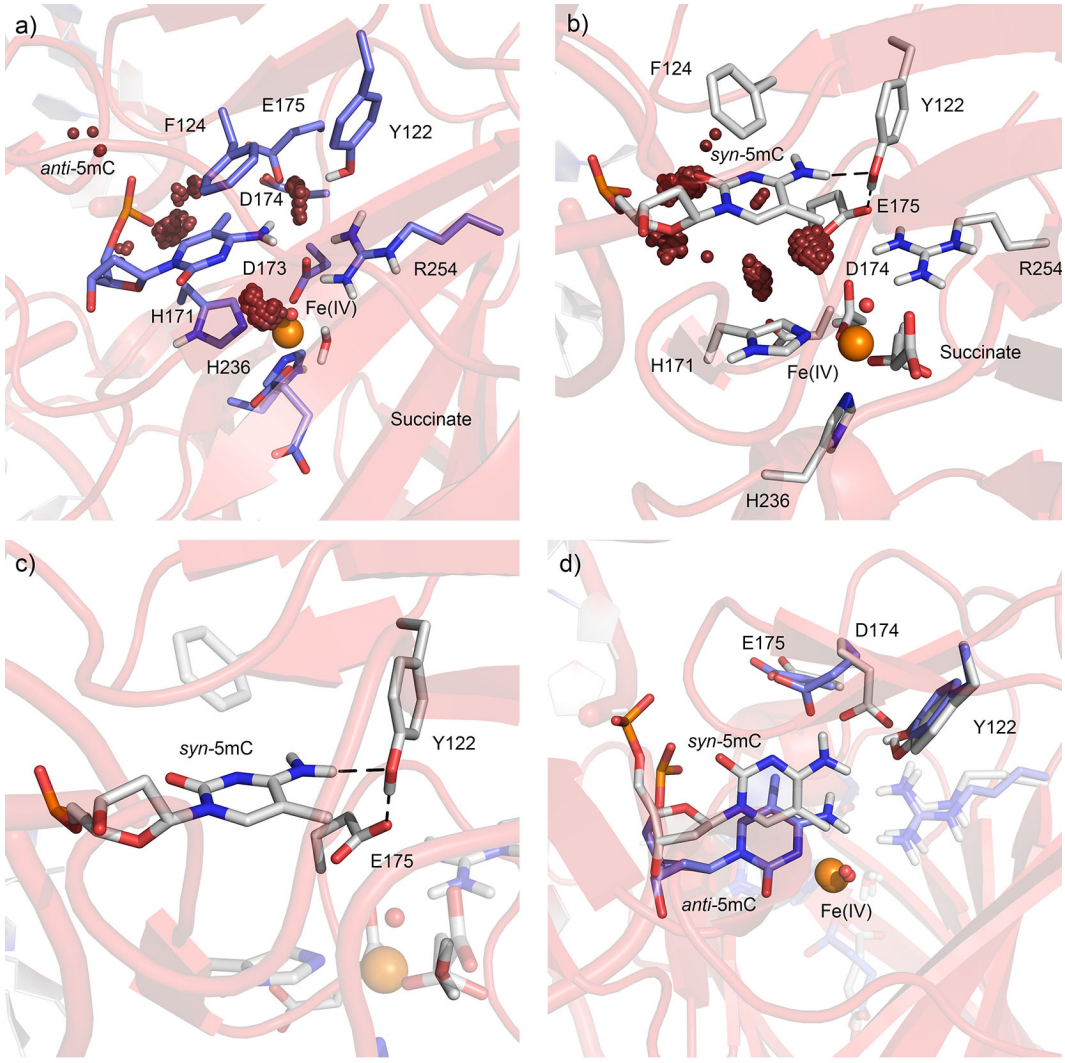


Figure F.14. Distribution of active site water (red spheres) during MD simulations of ALKBH2 bound to a) *anti*-5mC or b) *syn*-5mC. c) ALKBH2 bound to *syn*-5mC highlighting key hydrogen-bonding interactions. d) Overlay of MD representative structures of ALKBH2 bound to *anti*- (blue) or *syn*- (gray) 5mC.

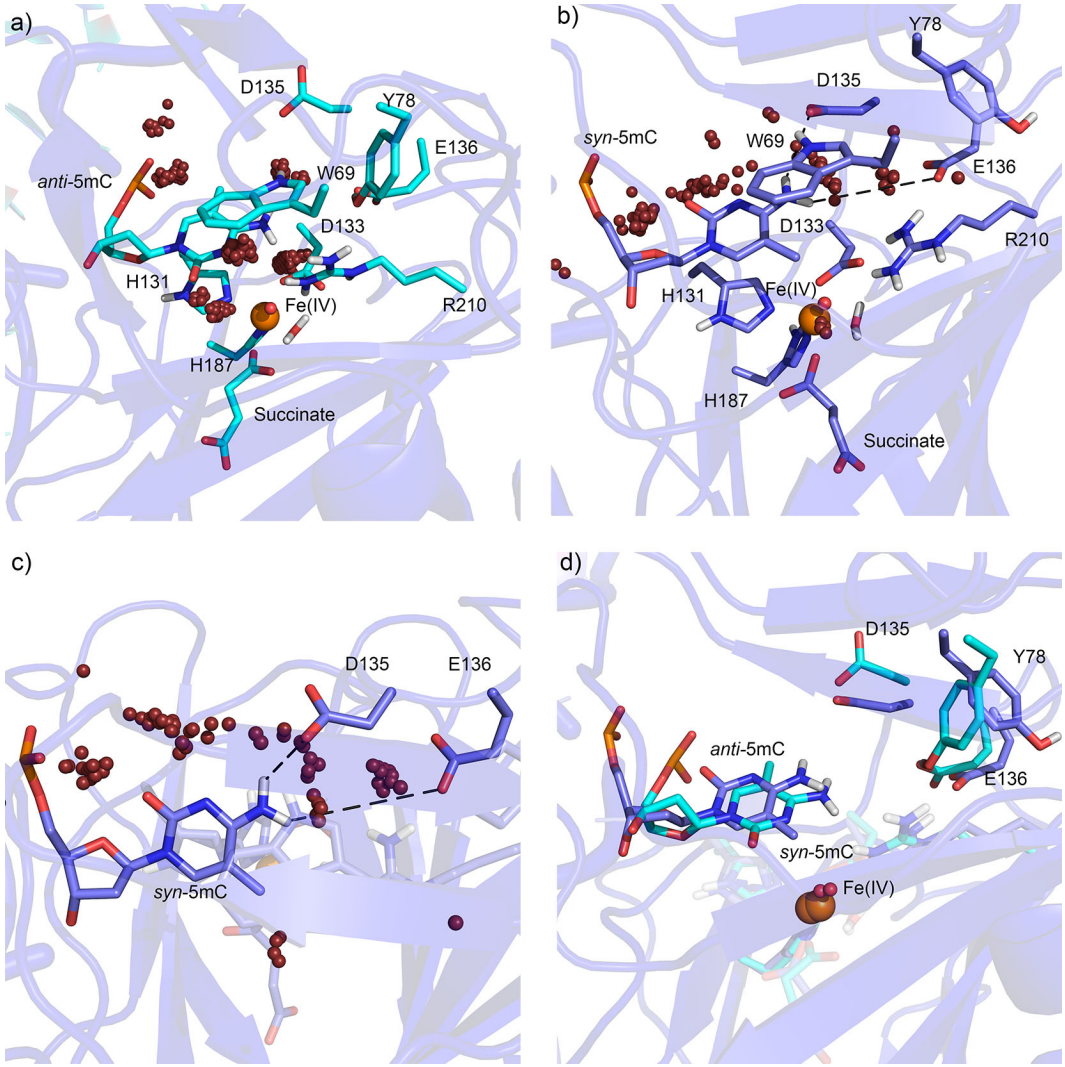


Figure F.15. Distribution of active site water (red spheres) during MD simulations of AlkB bound to a) *anti*-5mC or b) *syn*-5mC. c) AlkB bound to *syn*-5mC highlighting key water mediated hydrogen-bonding interactions. d) Overlay of MD representative structures of AlkB bound to *anti*- (cyan) or *syn*- (blue) 5mC.

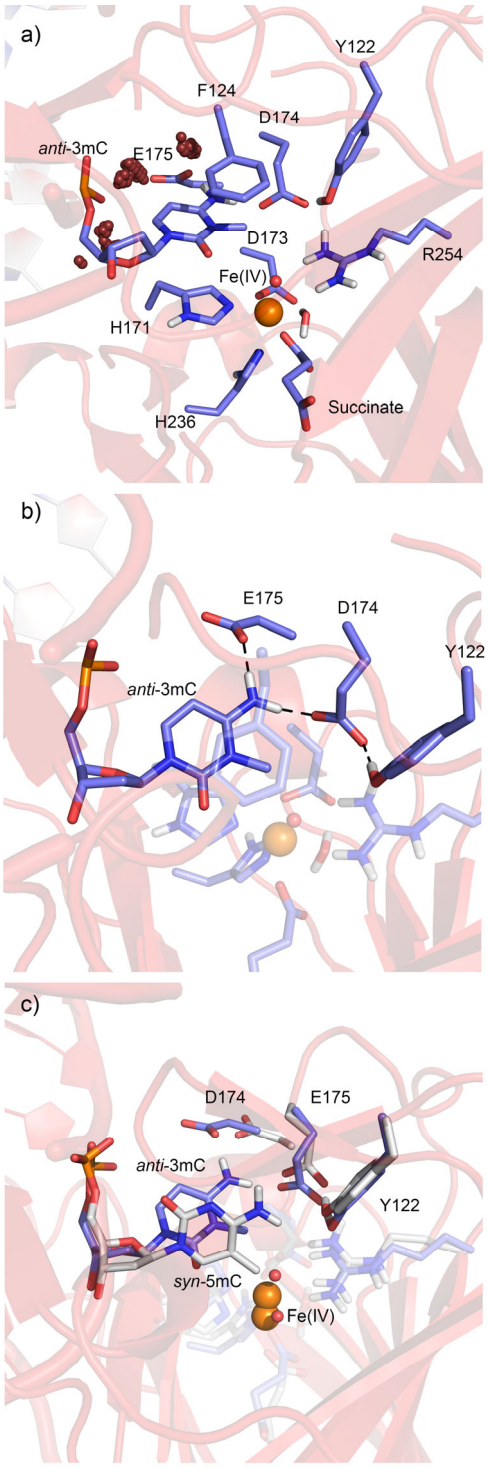


Figure F.16. a) Distribution of active site water (red spheres) during MD simulations of ALKBH2 bound to *anti*-3mC. b) ALKBH2 bound to *anti*-3mC, highlighting key hydrogen-bonding interactions. c) Overlay of MD representative structures of ALKBH2 bound to *syn*-5mC (gray) and *anti*-3mC (blue).

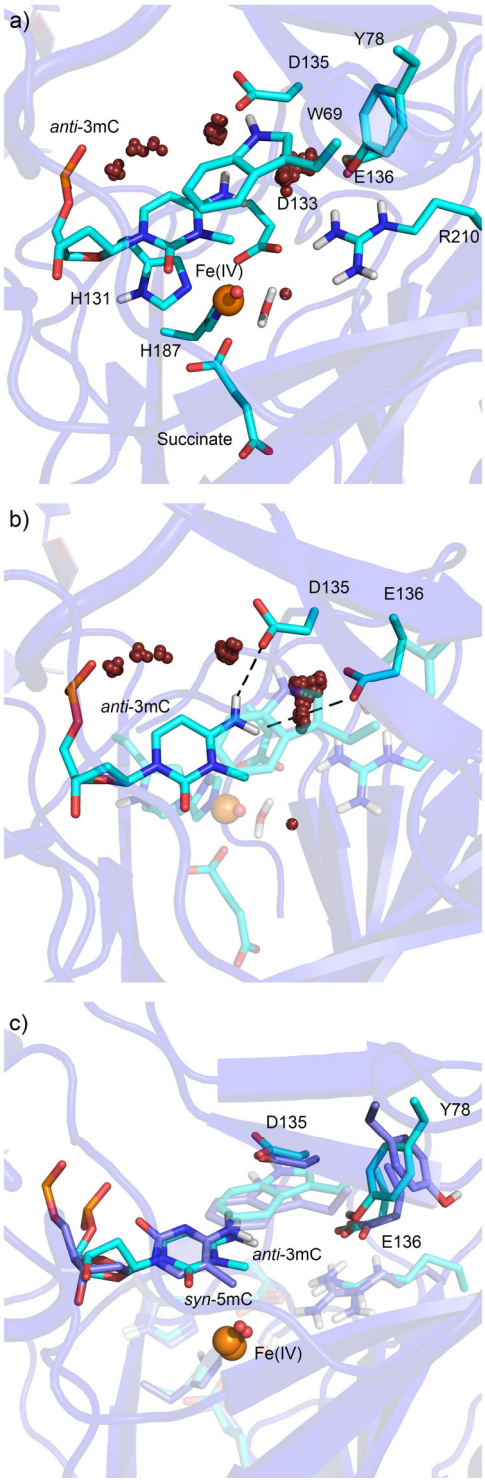


Figure F.17. a) Distribution of active site water (red spheres) during MD simulations of AlkB bound to *anti*-3mC. b) AlkB bound to *anti*-3mC, highlighting key water mediated hydrogen-bonding interactions. c) Overlay of MD representative structures of AlkB bound to *syn*-5mC (blue) and *anti*-3mC (cyan).

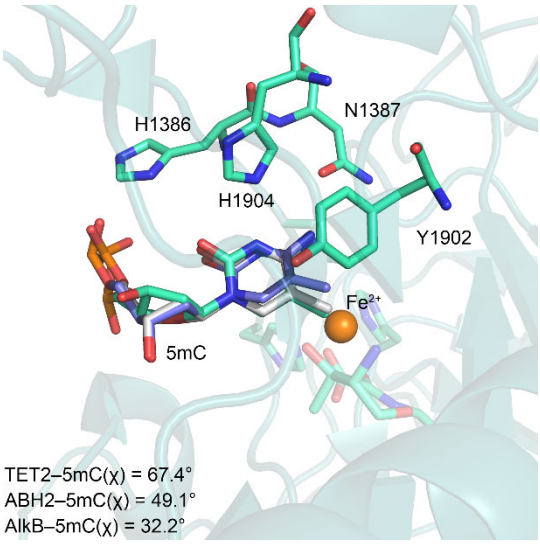


Figure F.18. Overlay of the 5mC nucleotide isolated from MD representative structures of AlkB–5mC (blue) and ALKBH2–5mC (gray) complexes onto a crystal structure of TET2 bound to 5mC-containing DNA, Fe(II), and α-KG (teal; PDB ID: 4NM6),[18] highlighting the similarity in the 5mC glycosidic bond conformation (χ).

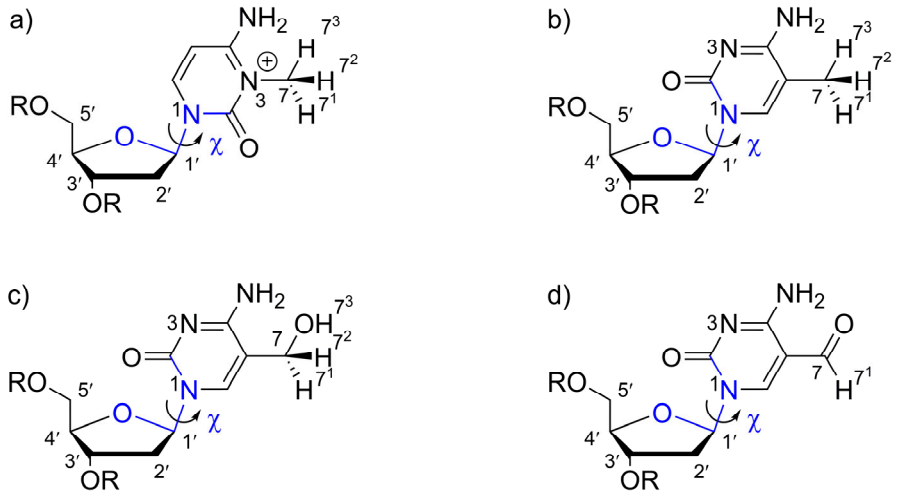


Figure F.19. Structures and chemical numbering of a) *anti*-3mC, b) *syn*-5mC, c) *syn*-5hmC, and d) *syn*-5fC. The glycosidic bond orientation is defined by the χ dihedral angle (blue, $\angle$(O4′C1′N1C2)) as *anti* ($180 \pm 90°$) or *syn* ($0 \pm 90°$).

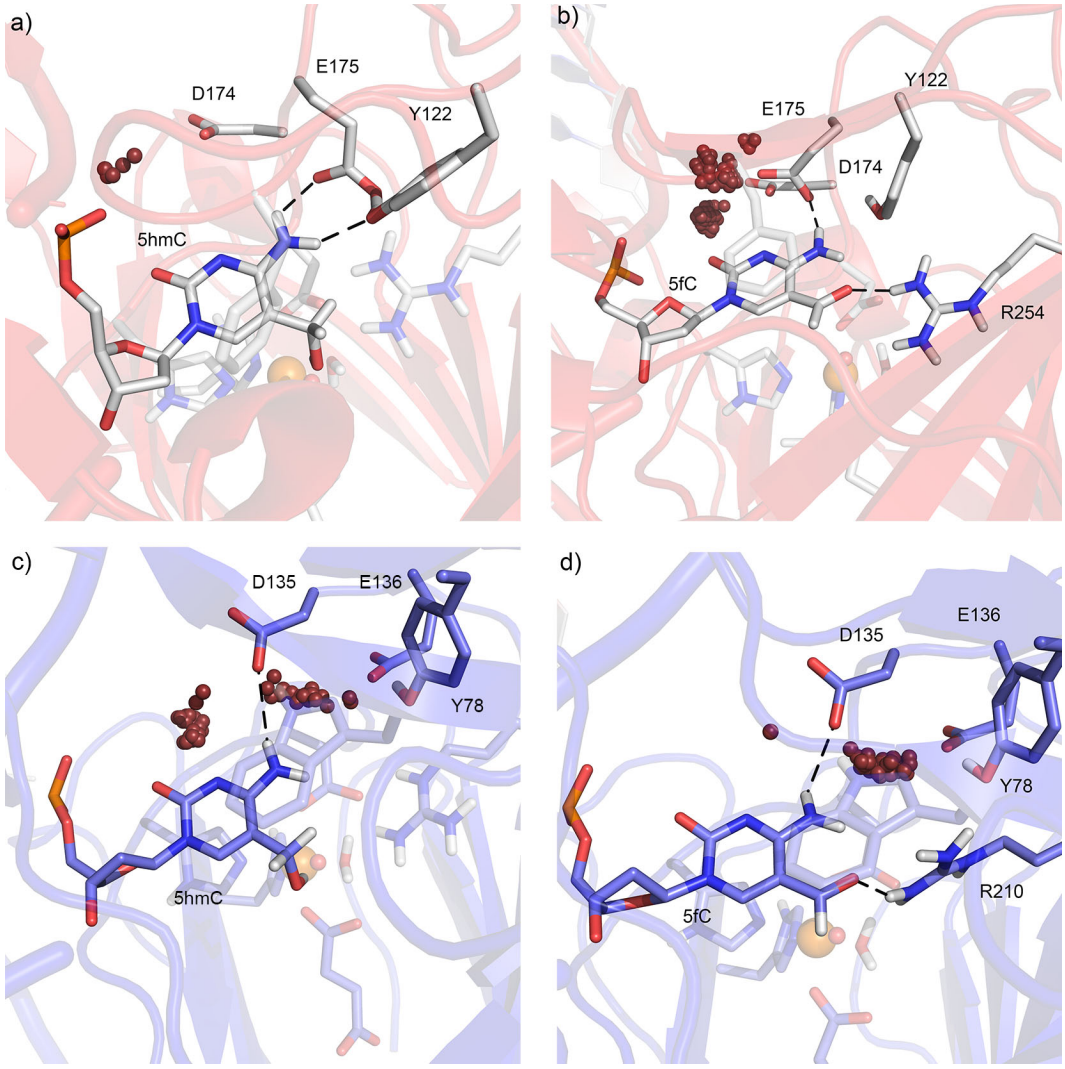


Figure F.20. Distribution of active site water (red spheres) during MD simulations of a) ALKBH2 bound to *syn*-5hmC, b) ALKBH2 bound to *syn*-5fC, c) AlkB bound to *syn*-5hmC, or d) AlkB bound to *syn*-5hmC. Key hydrogen-bonding interactions are highlighted with dashed lines.

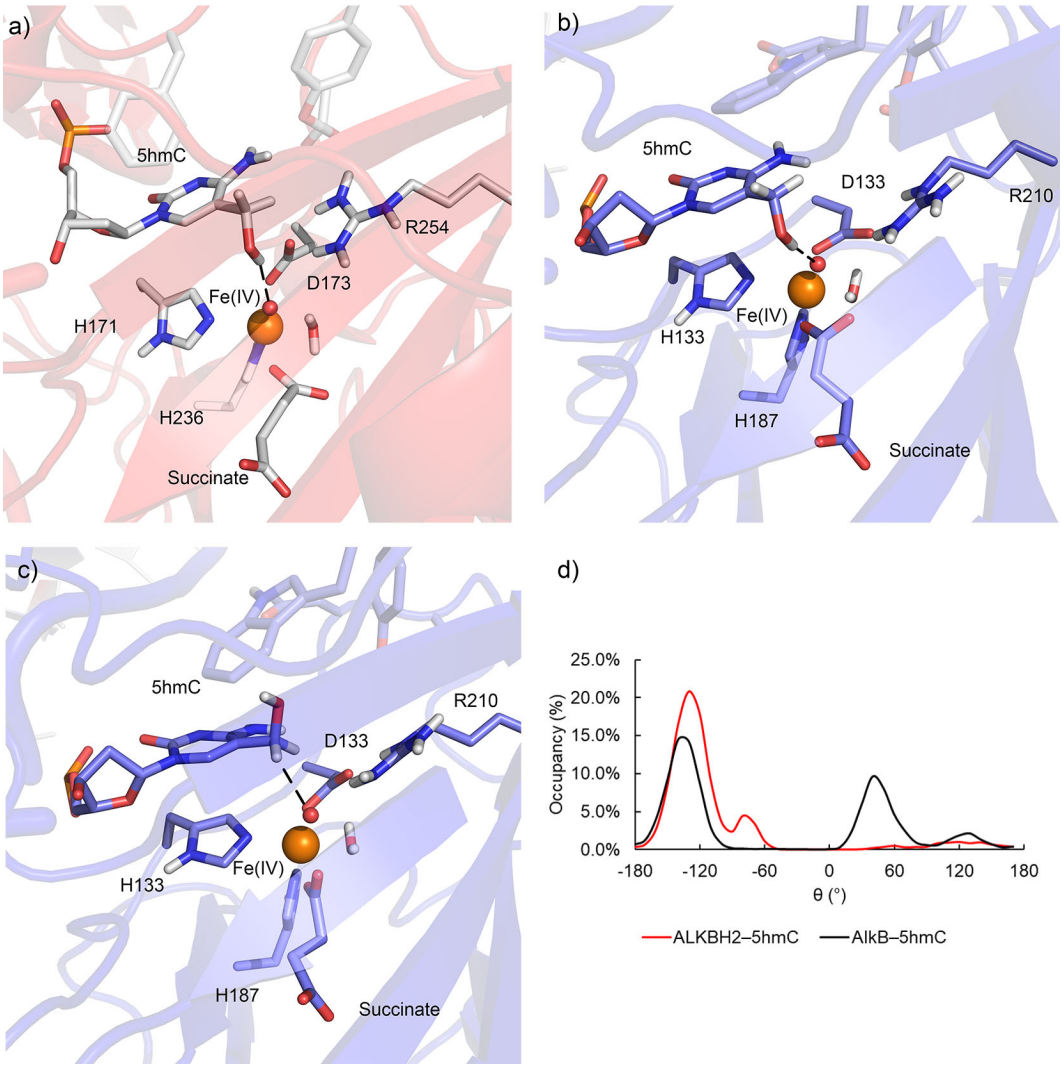


Figure F.21. a) ALKBH2 bound to *syn*-5hmC, and AlkB bound to *syn*-5hmC with b) θ (∠(C4C5C7O7)) ~ –132° and c) θ ~ 50°, highlighting different orientation of the C5 substituent. d) Histogram displaying the occupancy of θ (deg.) during simulations of ALKBH2 or AlkB bound to 5hmC.

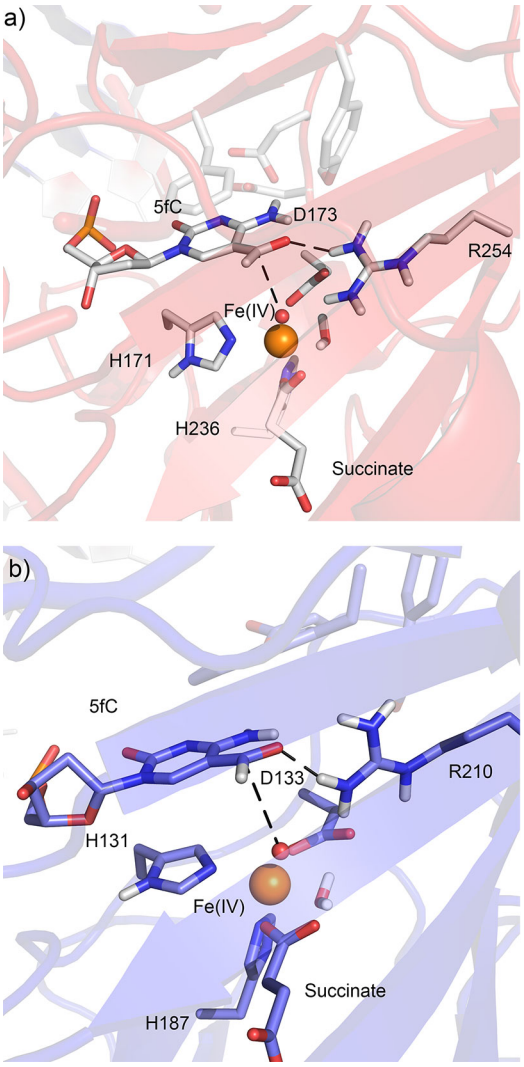


Figure F.22. *syn*-5fC bound by a) ALKBH2 or b) AlkB, highlighting the distance between the C5 substituent and the Fe(IV)-oxo moiety.

Table F.1. Calculated and observed monoisotopic molecular weight and m/z value of modified oligonucleotides. The sequence of the 16mer was 5'-GAAGACCTXGGCGTCC-3', where X indicates the position of 5mC or oxidized bases.

modification	**MW (calculated) of neutral species**	***m/z* (calculated) -3 charge peak**	***m/z* (observed) -3 charge peak**			
			5mC control	ALKBH2	ALKBH3	AlkB
5mC	4878.866	1625.281	1625.281	1625.279	1625.285	1625.276
5mC + Na^+	4900.848	1632.608	1632.608	1632.616	1632.613	1632.608
5mC + K^+	4916.822	1637.933	1637.928	1637.941	1637.942	1637.931
5mC + 2Na^+	4922.830	1639.935	1639.941	1639.937	1639.938	1639.931
5mC + Na^+ + K^+	4938.803	1645.260	1645.267	1645.273	1645.261	1645.263
5hmC	4894.861	1630.612	-	1630.615	1630.607	1630.603
5hmC + Na^+	4916.843	1637.940	-	1637.941	1637.942	1637.931
5hmC + K^+	4932.816	1643.264	-	1643.265	1643.259	1643.262
5fC	4892.845	1629.940	-	1629.936	1629.951	1629.933
5fC + Na^+	4914.827	1637.268	-	1637.288	1637.285	1637.257
5fC + 2Na^+	4936.809	1644.595	-	1644.605	1644.588	1644.598
5caC + Na^+	4930.814	1642.600	-	1642.616	1642.611	1642.604

Table F.2. Calculated and observed monoisotopic molecular weight and m/z value of peptide fragments digested by trypsin.

peptide sequence	**MW (calculated) of neutral species**	***m/z* (calculated) +1 charge peak**	***m/z* (observed) +1 charge peak**			
			AlkB **wt**	AlkB **D133A**	AlkB **H131A**	AlkB **H187A**
LSL**H**Q**D**K	839.450	840.457	840.470	-	-	840.470
LSL**H**Q**A**K	795.460	796.468	-	796.480	-	-
LSL**A**Q**D**K	773.428	774.436	-	-	774.449	-
LFY**H**GIQPLK	1214.681	1215.688	1215.706	1215.706	1215.706	-
LFY**A**GIQPLK	1148.659	1149.667	-	-	-	1149.681

Table F.3. Individual product percentage of 5hmC, 5fC and 5caC in the total oxidative product mixture from ds- or ss-5mC reactions with ALKBH2, ALKBH3 or AlkB.

	enzyme	5hmC	5fC	5caC
ds-DNA	ALKBH2	78.7 ± 3.5	10.9 ± 2.5	10.4 ± 1.4
	ALKBH3	59.9 ± 0.5	20.0 ± 0.3	20.1 ± 0.1
	AlkB	51.2 ± 3.0	38.7 ± 0.5	10.1 ± 3.5
ss-DNA	ALKBH2	72.3 ± 2.2	7.1 ± 1.9	20.6 ± 0.4
	ALKBH3	70.4 ± 0.6	23.9 ± 0.6	5.7 ± 0.6
	AlkB	28.1 ± 3.8	65.3 ± 4.9	6.6 ± 1.1

Table F.4. Percentage of total oxidation product including 5hmC, 5fC and 5caC from ds- or ss-5mC reactions with ALKBH2, ALKBH3 or AlkB.

	ALKBH2	ALKBH3	AlkB
ds-DNA	8.7 ± 0.3	4.2 ± 0.7	13.3 ± 0.3
ss-DNA	5.5 ± 0.9	22.1 ± 0.1	23.5 ± 1.8

Table F.5. Summary of important hydrogen bonds formed during MD simulations of anti-3mC, anti-5mC, *syn*-5mC, *syn*-5hmC, and *syn*-5fC in the ALKBH2 complex.

	3mC		***anti*-5mC**		***syn*-5mC**		***syn*-5hmC**		***syn*-5fC**	
	%[a,b]	**Dist (Angle)[c]**	**%[a,b]**	**Dist (Angle)[c]**	**%[a,b]**	**Dist (Angle)[c]**	**%[a,b]**	**Dist (Angle)[c]**	**%[a,b]**	**Dist (Angle)[c]**
E175(Oδ)···mC(N4H)	99.1%	2.8 (159)	0.0%	N/A	0.3%	3.0 (145)	83.2%	2.9 (151)	100.0%	2.9 (160)
D174(Oδ)···mC(N4H)	94.3%	2.8 (166)	0.0%	N/A	0.0%	N/A	32.0%	2.9 (164)	0.0%	N/A
Y122(O_H)···mC(N4H)	0.0%	N/A	15.9%	3.1 (145)	90.0%	3.0 (146)	5.0%	3.1 (141)	26.2%	3.1 (136)
H_2O(O)···mC(N4H)	55.3%	3.1 (138)	78.9%	3.1 (146)	72.2%	3.1 (137)	>100.0%	3.1 (151)	0.0%	3.2 (126)
mC(O2)···H_2O(OH)	0.1%	3.2 (128)	4.8%	3.1 (143)	94.3%	2.9 (154)	>100.0%	2.8 (150)	56.1%	2.8 (151)
mC(N3)···H_2O(OH)	0.0%	N/A	57.6%	3.0 (154)	62.5%	3.1 (147)	30.0%	3.0 (146)	0.9%	3.2 (138)
E175(Oδ)···Y122(OH)	99.0%	2.7 (165)	0.2%	2.8 (163)	78.7%	2.8 (164)	100.0%	2.9 (163)	41.9%	2.9 (154)
H_2O(O)···Y122(OH)	0.0%	N/A	>100%	3.0 (151)	17.7%	3.2 (129)	10.2%	3.1 (129)	0.0%	N/A
E175(Oε)···H_2O(OH)	0.0%	N/A	>100%	2.8 (159)	>100%	2.7 (160)	>100.0%	2.7 (162)	>100.0%	2.7 (159)
D174(Oδ)···H_2O(OH)	>100%	2.8 (158)	>100%	2.8 (157)	>100%	2.8 (160)	>100.0%	2.8 (157)	>100.0%	2.8 (152)
mC(O5)···R254(NηH)	N/A	N/A	N/A	N/A	N/A	N/A	37.0%	3.1 (140)	29.2%	3.1 (132)
oxo(O)···mC(O5H)	N/A	N/A	N/A	N/A	N/A	N/A	33.9%	2.9 (144)	N/A	N/A
D173(Oδ)···mC(O5H)	N/A	N/A	N/A	N/A	N/A	N/A	58.5%	2.7 (152)	N/A	N/A

[a]Percent occupancy of the hydrogen bond, which is defined using cutoffs of a distance less than 3.4 Å and an angle greater than 120°. [b]Greater than 100% occupancy indicates the presence of more than one water hydrogen bonding with an acceptor site.[c]Hydrogen bonding distance in Å. [d]Angle in ° (in parentheses).

Table F.6. Summary of important hydrogen bonds formed during MD simulations of *anti*-3mC, *anti*-5mC, *syn*-5mC, *syn*-5hmC, and *syn*-5fC in the AlkB complex.

	3mC		***anti*-5mC**		***syn*-5mC**		***syn*-5hmC**		***syn*-5fC**	
	%[a,b]	**Dist (Angle)**[c]	**%**[a,b]	**Dist (Angle)**[c]	**%**[a,b]	**Dist (Angle)**[c]	**%**[a,b]	**Dist (Angle)**[c]	**%**[a,b]	**Dist (Angle)**[c]
D133(Oδ)···mC(N4H)	42.6%	2.9 (147)	0.6%	3.2 (135)	11.8%	3.1 (143)	1.6%	3.2 (132)	0.2%	3.2 (132)
H_2O(O)···mC(N4H)	>100%	3.1 (150)	>100%	3.1 (142)	>100%	3.1 (145)	>100.0 %	3.1 (138)	>100.0 %	3.1 (137)
mC(O2)···H_2O(OH)	32.9%	3.0 (150)	>100%	2.9 (154)	>100%	2.9 (152)	72.9%	3.0 (141)	57.4%	3.0 (143)
mC(N3)···H_2O(OH)	0.0%	N/A	91.0%	2.9 (150)	95.5%	3.0 (152)	>100.0 %	3.0 (156)	>100.0 %	2.9 (155)
E136(Oδ)···Y78(OH)	62.3%	2.6 (163)	58.5%	2.7 (165)	36.0%	2.7 (162)	65.7%	2.7 (165)	96.3%	2.7 (164)
H_2O(O)···Y78(OH)	88.9%	3.1 (150)	25.8%	3.1 (137)	5.6%	3.1 (145)	42.7%	3.1 (135)	43.7%	3.1 (141)
E136(Oε)···H_2O(OH)	>100%	2.8 (158)	>100%	2.7 (160)	>100%	2.9 (152)	97.8%	2.8 (158)	>100.0 %	2.7 (162)
D135(Oδ)···H_2O(OH)	>100%	2.8 (158)	>100%	2.8 (158)	>100%	2.8 (157)	>100.0 %	2.8 (157)	>100.0 %	2.8 (157)
mC(O5)···R210(NηH)	N/A	N/A	N/A	N/A	N/A	N/A	35.1%	2.9 (160)	100.0%	2.8 (155)
Fe(IV)-oxo(O)···mC(O5H)	N/A	N/A	N/A	N/A	N/A	N/A	51.7%	2.8 (152)	N/A	N/A
D133(Oδ)···mC(O5H)	N/A	N/A	N/A	N/A	N/A	N/A	0.0%	N/A	N/A	N/A

[a]Percent occupancy of the hydrogen bond, which is defined using cutoffs of a distance less than 3.4 Å and an angle greater than 120°. [b]Greater than 100% occupancy indicates the presence of more than one water hydrogen bonding with an acceptor site.[c]Hydrogen-bonding distance in Å. [d]Angle in ° (in parentheses).

Table F.7. Summary of important distances and dihedral angles adopted during MD simulations of *anti*-3mC, *anti*-5mC, *syn*-5mC, *syn*-5hmC, and *syn*-5fC in the ALKBH2 complex.

	3mC Avg (σ)	***anti*-5mC** Avg (σ)	***syn*-5mC** Avg (σ)	**syn-5hmC** Avg (σ)	**syn-5fC** Avg (σ)
Fe(IV)–oxo···C(C7)[a]	3.3 (0.2)	5.3 (0.5)	3.8 (0.3)	4.5 (0.8)	3.2 (0.4)
Fe(IV)–oxo···C(H7^1)[a]	3.2 (0.7)	5.5 (0.8)	3.6 (0.6)	4.7 (1.1)	3.3 (0.4)
Fe(IV)–oxo···C(H7^2)[a]	3.2 (0.6)	5.5 (0.8)	3.6 (0.6)	6.9 (0.6)	N/A
Fe(IV)–oxo···C(H7^3)[a]	3.2 (0.7)	5.5 (0.8)	3.6 (0.6)	3.2 (1.4)	N/A
C(χ)[b]	–147 (15)	–143 (19)	49 (11)	-9 (17)	-19 (19)

[a]Modification of either N3 (3mC; Figure F.19) or C5 (5mC, 5hmC, or 5fC; Figure F.19). [b]χ of bound pyrimidine defined as ∠(O4′C1′N1C2) (Figures S18 and S19).

Table F.8. Summary of important distances and dihedral angles adopted during MD simulations of *anti*-3mC, *anti*-5mC, *syn*-5mC, *syn*-5hmC, and *syn*-5fC in the AlkB complex.

	3mC Avg (σ)	***anti*-5mC** Avg (σ)	***syn*-5mC** Avg (σ)	**syn-5hmC** Avg (σ)	**syn-5fC** Avg (σ)
Fe(IV)–oxo···C(C7)[a]	3.3 (0.2)	7.9 (2.0)	3.6 (0.4)	3.4 (0.2)	3.5 (0.4)
Fe(IV)–oxo···C(H7^1)[a]	3.5 (0.7)	8.2 (2.0)	3.6 (0.7)	3.3 (0.6)	3.8 (0.5)
Fe(IV)–oxo···C(H7^2)[a]	3.4 (0.7)	8.2 (2.0)	3.5 (0.7)	5.8 (0.4)	N/A
Fe(IV)–oxo···C(H7^3)[a]	3.4 (0.7)	8.2 (2.0)	3.6 (0.7)	3.3 (1.5)	N/A
C(χ)[b]	–128 (11)	–116 (13)	32 (17)	44 (18)	51 (10)

[a]Modification of either N3 (3mC; Figure F.19) or C5 (5mC, 5hmC, or 5fC; Figure F.19). [b]χ of bound pyrimidine defined as ∠(O4′C1′N1C2) (Figures S18 and S19).

References

[1] Bian, K.; Chen, F.; Humulock, Z. T.; Tang, Q.; Li, D. Copper Inhibits the AlkB Family DNA Repair Enzymes under Wilson's Disease Condition. *Chem. Res. Toxicol.* **2017**, *30* (10), 1794–1796.

[2] Tang, Q.; Cai, A.; Bian, K.; Chen, F.; Delaney, J. C.; Adusumalli, S.; Bach, A. C.; Akhlaghi, F.; Cho, B. P.; Li, D. Characterization of Byproducts from Chemical Syntheses of Oligonucleotides Containing 1-Methyladenine and 3-Methylcytosine. *ACS Omega* **2017**, *2* (11), 8205–8212.

[3] Chen, F.; Tang, Q.; Bian, K.; Humulock, Z. T.; Yang, X.; Jost, M.; Drennan, C. L.; Essigmann, J. M.; Li, D. Adaptive Response Enzyme AlkB Preferentially Repairs 1-Methylguanine and 3-Methylthymine Adducts in Double-Stranded DNA. *Chem. Res. Toxicol.* **2016**, *29* (4), 687–693.

[4] Chen, F.; Bian, K.; Tang, Q.; Fedeles, B. I.; Singh, V.; Humulock, Z. T.; Essigmann, J. M.; Li, D. Oncometabolites D- and L-2-Hydroxyglutarate Inhibit the AlkB Family DNA Repair Enzymes under Physiological Conditions. *Chem. Res. Toxicol.* **2017**, *30* (4), 1102–1110.

[5] Quinlivan, E. P.; Gregory, J. F. DNA Digestion to Deoxyribonucleoside: A Simplified One-Step Procedure. *Anal. Biochem.* **2008**, *373* (2), 383–385.

[6] Frick, L. E.; Delaney, J. C.; Wong, C.; Drennan, C. L.; Essigmann, J. M. Alleviation of 1,N6-Ethanoadenine Genotoxicity by the Escherichia Coli Adaptive Response Protein AlkB. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (3), 755–760.

[7] Yi, C.; Jia, G.; Hou, G.; Dai, Q.; Zhang, W.; Zheng, G.; Jian, X.; Yang, C.-G.; Cui, Q.; He, C., Iron-catalysed oxidation intermediates captured in a DNA repair dioxygenase. *Nature* **2010,** *468*, 330–333.

[8] 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. Biol.* **2012,** *19*, 671–676.

[9] Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W., et al. *AmberTools16*; University of California: San Francisco, 2016.

[10] 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*, W537-W541.

[11] Li, P.; Merz, K. M., Mcpb.Py: A python based metal center parameter builder. *J. Chem. Inf. Mod.* **2016,** *56*, 599–604.

[12] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A., et al. *Gaussian 09*, Revision D.01; Gaussian, Inc.: Wallingford CT, 2016.

[13] Seminario, J. M., Calculation of intramolecular force fields from second-derivative tensors. *Int. J. Quantum Chem.* **1996,** *60*, 1271–1277.

[14] Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C., Routine microsecond molecular dynamics simulations with amber on GPUs. 2. Explicit solvent particle mesh ewald. *J. Chem. Theory. Comput.* **2013,** *9*, 3878–3888.

[15] Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C., Routine microsecond molecular dynamics simulations with amber on GPUs. 1. Generalized born. *J. Chem. Theory. Comput.* **2012,** *8*, 1542–1555.

[16] Le Grand, S.; Götz, A. W.; Walker, R. C., Spfp: Speed without compromise—a mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Phys. Commun.* **2013,** *184*, 374–380.

[17] Roe, D. R.; Cheatham, T. E., PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory. Comput.* **2013,** *9*, 3084–3095.

[18] 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*, 1545-1555.

[19] Lu, J., Hu, L., Cheng, J., Fang, D., Wang, C., Yu, K., Jiang, H., Cui, Q., Xu, Y., and Luo, C. A computational investigation on the substrate preference of ten-eleven-translocation 2 (TET2), *Phys. Chem. Chem. Phys.* **2013** 18, 4728-4738.