INVESTIGATING THE MECHANISMS OF ACTION OF GLYCOSYLASES AND NUCLEASES USING CLASSICAL AND QUANTUM MECHANICAL COMPUTATIONAL TECHNIQUES

DYLAN NIKKEL

Bachelor of Science (Chemistry Honours), University of Lethbridge, Lethbridge, 2019

A thesis submitted

In partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

THEORETICAL AND COMPUTATIONAL SCIENCE

Department of Chemistry and Biochemistry

University of Lethbridge

LETHBRIDGE, ALBERTA, CANADA

© Dylan Nikkel, 2026

INVESTIGATING THE MECHANISMS OF ACTION OF GLYCOSYLASES AND NUCLEASES USING CLASSICAL AND QUANTUM MECHANICAL COMPUTATIONAL TECHNIQUES

DYLAN NIKKEL

Date of Defense: February 5th, 2026

Dr. Stacey Wetmore Supervisor	Professor	Ph.D.
Dr. Marc Roussel Thesis Examination Committee Member	Professor	Ph.D.
Dr. Trushar Patel Thesis Examination Committee Member	Professor	Ph.D.
Dr. James Gauld External Examiner Memorial University of Newfoundland St. John's, Newfoundland and Labrador	Professor	Ph.D.
Dr. Michael Gerken Chair, Thesis Examination Committee	Professor	Ph.D.

DEDICATION

This thesis is dedicated to my savior Jesus Christ and my parents whose love and support made everything possible.

ABSTRACT

The phosphodiester and glycosidic bonds are essential for the function of nucleic acids, binding nucleotides together and nucleobases to (deoxy)ribose. However, many cellular processes require cleaving these bonds, including DNA repair, RNA processing, and viral defense. To enable these processes to occur at biologically-relevant timescales, cells utilize glycosylases and nucleases to facilitate this chemistry. Due to their important roles in cell survival, glycosylases and nucleases are implicated in many different diseases and disorders, including neurodegeneration, inflammation, and cancer. Despite this, the catalytic mechanism for many of these enzymes is poorly understood, with multiple proposals often present in the literature. In this thesis, computational methodologies, including molecular dynamics (MD) simulations, free energy techniques, quantum mechanics/molecular mechanics (QM/MM) calculations, and QM/MM MD simulations were used to clarify the atomic level details of glycosylase and nuclease catalytic mechanisms. Specifically, the monofunctional glycosylases MutY, MBD4, and AlkA were investigated, focusing on their potential to invoke a novel mechanism that involves a DNA−protein crosslinked intermediate. To complement the work on monofunctional glycosylases, the bifunctional glycosylase hOGG1 was also studied, with this work representing the first characterization of a bifunctional glycosylase β-lyase mechanism. Finally, phosphodiester bond cleavage was further investigated by studying the Dicer mechanism of action, with the elucidated mechanism used to gain insight into the disease-causing nature of DICER1 hotspot mutants. The improved understanding of nuclease and glycosylase mechanisms will allow for the rational design of small molecule inhibitors as therapeutics and pushes forward the use of these enzymes for biotechnological applications.

PREFACE

Some of the research conducted in this thesis includes contributions from collaborators. In Chapter 4, R. Fjordbotten initiated molecular dynamics simulations on the AlkA−substrate complex, performed clustering analysis to obtain representative structures, and constructed initial quantum mechanics/molecular mechanics models under my supervision. B. Mansour mapped the AlkA hydrolysis and crosslinking mechanisms using quantum mechanics/molecular mechanics methods and aided in analyzing the resulting data sets. Additionally, B. Mansour assisted in writing the first draft of the introduction, and the portion of the methodology and results sections that involved quantum mechanics/molecular mechanics calculations. In Chapter 5, T. Deak initiated preliminary molecular dynamics simulations of the hOGG1−substrate complex with a cationic crosslink under my supervision. For all projects, I was involved in project conceptualization, model construction, data collection and analysis, and writing (original draft, review, and editing). For all projects, S. D. Wetmore was involved in project conceptualization, administration, supervision, analysis, funding acquisition, computational resource acquisition, and writing (review and editing).

Chapter 2 of this thesis has been published as Nikkel, D. J.; Wetmore, S. D. Distinctive formation of a DNA–protein cross-link during the repair of DNA oxidative damage: Insights into human disease from MD simulations and QM/MM calculations. *J. Am. Chem. Soc.* **2023**, *145* (24), 13114-13125.

Chapter 3 of this thesis has been accepted for publication as Nikkel, D. J.; Wetmore, S. D. A Tale of Two Mechanisms: Clarification of the Pathway for MBD4 Catalyzed Glycosidic Bond Cleavage Using MD and QM/MM Calculations. *DNA Repair* (accepted Dec. 18, 2025).

ACKNOWLEDGMENTS

To start, I need to thank my supervisor, Dr. Stacey Wetmore, for her invaluable mentorship and support during my time at the University of Lethbridge. She always had my best interests at heart and provided opportunities for growth as both a person and scientist during my stay in her lab and this thesis would be a shadow of itself without her help. I would also like to acknowledge my committee members, Dr. Trushar Patel and Dr. Marc Roussel, for their insightful comments and feedback over the course of my graduate studies.

I have also received support from all fellow members of the Wetmore lab, past and present, during my graduate studies. This includes mentors who supported me when I just joined the lab (Preethi Seelam, Hanzala Hussain, Katie Wilson, Ryan Kung, Stefan Lenz, Priya Bhutani, and Rajwinder Kaur), colleagues who grew with me (Briana Boychuk, Rebecca Jeong, Makay Murray, Nathania Takyi, Cynthia Fonderson, Mark Lea, Angela Frederickson, Bimaldeep Singh, Ridwan Tajudeen, Indu Negi, Austin Pounder, Umer Yaqoob, Basel Mansour), newcomers to the lab (Anita Basu, Clara Cunha, Kari Gaalswyk, Hossein Boroujeni, Parth Rathee) as well as undergraduate students I have had the privilege to mentor (Audra Lutterotti, Trinity Deak, Ryan Fjordbotten). A special mention has to be given to Basel Mansour, Trinity Deak, and Ryan Fjordbotten who I collaborated with and whose research appears in this thesis.

Finally, I have to thank my parents, who have always supported me studying computational chemistry (even if they didn’t understand it). This thesis would not have been possible without them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

1N	(3R,4R)-4-(hydroxymethyl)pyrrolidin-3-ol
3mA	3-methyladenine
3mC	3-methylcytosine
5-caC	5-carboxylcytosine
5-FU	5-fluorouracil
7mG	7-methylguanine
8oG	8-oxoguanine
aa	*Aquifex aeolicus*
AAG	Alkyladenine DNA glycosylase
abMD	Adaptively biased molecular dynamics
AlkA	3-methyladenine DNA glycosylase II
AP	Apurinic/apyrimidinic
APE1	Apurinic/apyrimidinic endonuclease 1
BamHI	Bacillus amyloliquefaciens H
BER	Base excision repair
Cas9	CRISPR-associated protein 9
CpG	Cytosine-phosphate-guanine
CRISPR	Clustered regularly interspaced palindromic repeats
Cryo-EM	Cryogenic electron microscopy
CV	Collective variable
Dicer	Endoribonuclease Dicer
Drosha	Drosha ribonuclease III
DUF	Domain of unknown function
EE	Electronic embedding
Endo VIII	Endonuclease VIII
EndoV	Endonuclease V
FdA	2′-deoxy-2′-fluoro-β-D-arabinoadenosine
FLRC	Fluorine recognition complex
Fpg	Formamidopyrimidine glycosylase
GAFF	general AMBER force field
Gs	*Geobacillus stearothermophilus*
HhH	Helix-hairpin-helix
HIV	Human immunodeficiency virus
hOGG1	8-oxoguanine DNA glycosylase 1
hUNG2	Human uracil DNA glycosylase 2
Hyp	Trans-4-hydroxy-l-proline
IC	Intermediate complex
KIE	Kinetic isotope effects
LRC	Lesion recognition complex
MANS	MBD4 associated neoplasia syndrome
MAP	MUTYH associated polyposis
MBD4	Methyl-CpG binding domain 4
ME	Mechanical embedding
miRNA	MicroRNA
MM	Molecular mechanics
mNEIL3	Mouse Nei like glycosylase 3
Msm	*Mycobacterium smegmatis*
MutH	DNA mismatch repair protein MutH
MutY	Adenine DNA glycosylase
MUTYH	Human adenine DNA glycosylase homologue
NEIL1	Nei like glycosylase 1
Nth	Endonuclease III
O2mC	O2-methylcytosine

O2mT	O2-methylthymine
ONIOM	Our own N-layered Integrated molecular orbital and molecular mechanics
PAZ	Piwi, Argonaute, and Zwille
PC	Product complex
PDB	Protein data bank
PES	Potential energy surface
pre-miRNA	Precursor microRNA
pri-miRNA	Primary microRNA
QM	Quantum mechanics
RC	Reactant complex
RISC	RNA-induced silencing complex
RMSD	Root-mean-square deviation
RNAi	RNA interference
RNase H	Ribonuclease H
RNase III	Ribonuclease III
ROS	Reactive oxygen species
siRNA	Small-interfering RNA
SMD	Steered molecular dynamics
SMUG1	Single-strand selective monofunctional uracil DNA glycosylase
T4 PDG	T4 pyrimidine dimer glycosylase
TDG	Thymine DNA glycosylase
TFAM	Mitochondrial transcription factor A
TI	Thermodynamic integration
TS	Transition state complex
TSAC	Transition state analogue complex
UDG	Uracil DNA glycosylase
UdgX	Uracil DNA glycosylase X
US	Umbrella sampling

Chapter 1: Introduction

1.1. General Overview

Nucleic acids play numerous vital roles in cells including storing genetic information,[1] facilitating protein synthesis and regulation,[2] and defending against viruses.[3] By necessity, the glycosidic bonds that attach nucleobases to (deoxy)ribose and the phosphodiester bonds that connect nucleotides together are stable, with the uncatalyzed hydrolysis having a half-life of 10–40 years[4] and more than 100,000 years,[5] respectively. However, the cell often needs to cleave these bonds for survival. For example, nucleic acid repair, processing, and degradation require breaking glycosidic and phosphodiester bonds. As a result, cells produce many different enzymes that catalyze the cleavage of glycosidic and/or phosphodiester bonds in nucleic acids (Figure 1.1A),[6-9] enabling the reactions to occur on biologically-relevant timeframes.

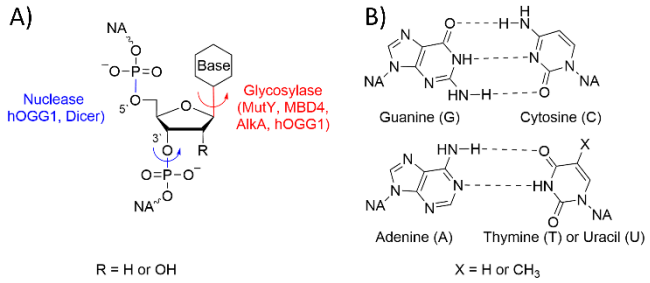


Figure 1.1. A) Structure of a nucleotide highlighting bonds cleaved by the glycosylases and nuclease investigated in this thesis. B) Structure and base pairing of the canonical nucleobases.

Regardless of their important cellular roles, the mechanisms used by many enzymes to facilitate nucleic acid glycosidic and phosphodiester bond cleavage are not well understood. Indeed, enzyme-catalyzed reactions are very fast and key species along reaction pathways are short-lived and challenging to characterize through experimental techniques. Therefore, despite advancements in experimental biochemical tools, critical information about molecular reactions (such as transition state structures and barrier heights) remains difficult to obtain from experiments. Providing a unique, yet complementary, approach to uncovering molecular reactivity, computational

chemistry can characterize high-energy intermediates along chemical pathways[10-13] and uncover enzyme structure in the presence of natural substrates.

This thesis uses a variety of computational techniques ranging from molecular mechanics to quantum mechanical approaches to accurately model the cleavage of glycosidic and phosphodiester bonds in nucleic acids. This will answer questions regarding enzyme active site conformations, roles of catalytically-relevant amino acid residues, and the structure of disease-related mutants. In this thesis, the mechanism for glycosidic bond cleavage, a vital step in the repair of damaged DNA nucleotides,[14] was investigated for the monofunctional glycosylases adenine DNA glycosylase (MutY), methyl-CpG binding domain 4 (MBD4), and 3-methyladenine DNA glycosylase II (AlkA). While the general monofunctional glycosylase mechanism involves a water molecule directly hydrolysing the glycosidic bond (Figure 1.2A), all three of the investigated enzymes had experimental evidence of a potential novel mechanism involving the formation of a DNA–protein crosslinked intermediate. To complement the work done on monofunctional glycosylases, the unknown phosphodiester bond cleavage mechanism of the bifunctional glycosylase 8-oxoguanine DNA glycosylase 1 (hOGG1) was then studied. To contrast the phosphodiester bond cleavage strategy of hOGG1, the RNA processing endoribonuclease Dicer (Dicer) was investigated to determine the mechanism of action and provide a rationale for the reduced activity of specific Dicer mutants. A better understanding of the function of these enzymes will drive the development of treatments for related genetic disorders (e.g. MUTYH associated polyposis (MAP),[15-17] DICER1 syndrome,[18-24] MBD4 associated neoplasia syndrome (MANS)[25]), while also aiding in the development of small molecule inhibitors as therapeutics[26-33] and the discovery of new biotechnological applications.[34-36] The following sections of the introduction will discuss how nucleic acids are processed in cells, the general function of glycosylases and endonucleases, the computational modeling of enzymes, and provide a summary of how various computational techniques were used in this thesis to study the enzymatic cleavage of glycosidic and phosphodiester bonds.

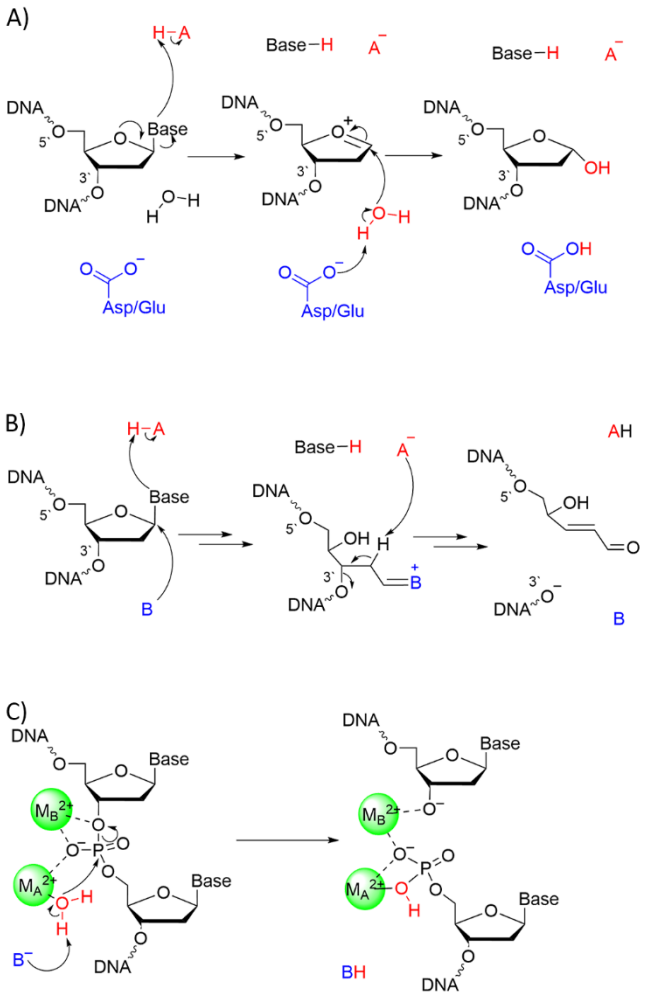


Figure 1.2. General mechanisms for A) glycosidic bond cleavage by a monofunctional glycosylase, B) glycosidic and phosphodiester bond cleavage by a bifunctional glycosylase, and C) phosphodiester bond cleavage by a two-metal-dependent nuclease.

1.2. Nucleic Acid Processing

DNA and RNA are vital biological macromolecules, with DNA primarily acting as information storage,[37] while RNA plays a role in information transportation,[38] catalyzing chemical reactions,[39] aiding the translation of proteins,[40] and regulating gene expression.[41] Nucleic acids are composed of nucleotide subunits consisting of a phosphate, (deoxy)ribose, and nucleobase (adenine (A), thymine (T), guanine (G), cytosine (C), or uracil (U), Figure 1.1). The nucleobase is bound to the sugar through a glycosidic bond to form nucleosides, while nucleotides are connected via phosphodiester bonds. Nucleobases can form hydrogen bonds with each other using the Watson-

Crick−Franklin face of the nucleobase in a complementary arrangement where G pairs with C and A pairs with T (DNA) or U (RNA, Figure 1.1B).

While nucleic acid stability is important, various cellular processes require the cleavage of glycosidic or phosphodiester bonds to ensure cell survival. DNA and RNA repair pathways require the cleavage of glycosidic and/or phosphodiester bonds to remove damaged nucleotides and insert undamaged nucleotides to maintain the integrity of the genetic code.[42, 43] Nucleic acid processing pathways, including splicing,[44] miRNA biogenesis,[45-47] and DNA recombination[48] also require the cleavage of the stable phosphodiester bond. To allow for these processes to occur on biologically-relevant timescales, cells utilize glycosylases and endonucleases to catalyze glycosidic and phosphodiester bond cleavage, allowing for up to a 10^{17}-fold rate enhancement over the uncatalyzed reaction.[5] Glycosylases and endonucleases, both of which are studied in the present thesis, will be discussed in the following two sections.

1.3. Glycosylases

The complementary base pairing between A and T, and G and C in DNA (Figure 1.1B**)** is essential for maintaining genetic integrity during DNA replication and transcription. However, DNA can be damaged by a variety of endogenous (e.g., metabolic products[49] and inflammatory responses[50]) and exogenous (e.g., UV radiation[51] and tobacco smoke[52]) sources, which can result in a change in base complementarity. This can lead to DNA mutations that can cause different disorders including cardiovascular disease,[53] neurodegeneration,[54, 55] and cancer.[18, 56] To prevent mutagenic outcomes, one method cells use to repair small, non-distorting nucleotide modifications (non-bulky damage) is the base excision repair (BER) pathway.[14] BER is initiated by a glycosylase identifying and excising a damaged base by cleaving the glycosidic bond. The resulting abasic site is a substrate for an apurinic/apyrimidinic (AP) endonuclease, which cleaves the backbone before a DNA polymerase and lyase insert the correct nucleotide and seal the nick in the DNA, completing repair.

Glycosylases can broadly be categorized as either monofunctional or bifunctional. Monofunctional glycosylases solely cleave the glycosidic bond. All previous monofunctional glycosylases with characterized mechanisms utilize an active site base to activate a water nucleophile that, in turn, attacks at C1′ to displace the damaged base (**Figure 1.2A**).[7, 57-60] Alternatively, bifunctional glycosylases cleave the glycosidic bond in addition to the 3′ phosphodiester bond, 5′ phosphodiester bond, or both.[61] Catalysis generally proceeds via an active site amine directly attacking at C1′ to form a DNA−protein crosslink and cleave the glycosidic bond while subsequent elimination reactions cleave the phosphodiester bond(s) (**Figure 1.2B)**.[62-65]

Due to their important roles in identifying and removing damaged DNA, glycosylase misfunction is associated with genetic disorders.[15, 25] DNA glycosylases are also attractive targets for cancer therapeutics,[33, 66] with glycosylase inhibitors improving the efficacy of cancer fighting drugs that inflict DNA damage.[67, 68] Despite their importance in cells and therapeutic potential, the mechanism of action is poorly understood for many glycosylases, with contradicting proposals present in the literature. Clarifying the mechanistic pathways for these glycosylases, one of the goals of this thesis, will pave the way for the development of treatments for glycosylase related genetic disorders and aid the rational design of small molecule inhibitors as therapeutics.

1.4. Nucleases

Despite nucleic acid function being dependent on a degradation-resistant backbone, phosphodiester bond cleavage is necessary for a multitude of biological processes including DNA repair,[42] DNA recombination,[48] RNA processing,[45-47] splicing,[44] viral defense,[69, 70] and apoptosis.[71] Nucleases are responsible for facilitating phosphodiester bond cleavage for these processes, with cleavage involving a nucleophile attacking at the phosphorus of the backbone to displace one of the strands.[72] However, the methods nucleases use to catalyze P−O bond cleavage can vary significantly. While two-metal-mediated catalysis was considered the standard pathway (Figure 1.2C),[73] recent experimental and computational studies have revealed a one-metal-mediated mechanism for several enzymes.[74-76] With more mechanistic pathways possible, experimental and

computational studies are necessary to accurately determine nuclease mechanisms rather than assuming the adoption of a generic two-metal pathway. Considering nuclease misfunction is associated with neurodegenerative disorders,[77] delayed development,[78] and cancer,[18] an accurate characterization of nuclease mechanisms is required for the development of therapeutics. Additionally, the atomic level details of nuclease mechanisms of action can be applied to nuclease biotechnologies, with nucleases being utilized for genetic engineering[79, 80] and disease detection.[81, 82]

1.5. Computational Modeling of Enzymes

Most currently available mechanistic information for enzymes that process nucleic acids has been conjectured from structures generated from X-ray crystallography and cryogenic electron microscopy (cryo-EM) coupled with knowledge of the catalytic contributions of different active site residues from mutagenesis experiments; however, these techniques have limitations. While X-ray crystal structures frequently reach atomic resolution, only inferred dynamic information is provided and crystallized enzyme−substrate complexes either contain inactive mutant enzymes or non-functional substrates. Cryo-EM techniques are able to capture enzyme dynamics on microsecond timescales,[83] but this remains insufficient to capture many enzymatic reactions and structures often lack atomic level resolution. Mutagenic studies can identify residues vital for catalysis but cannot elucidate the role of the residue in the mechanism of action. To fill critical gaps left by these methods, computational techniques have been widely adopted to provide atomic level descriptions of enzyme structure and mechanisms.[10-13] In terms of enzyme structure, calculations have the unique advantage of being able to describe wild-type systems in the presence of native substrates without requiring the introduction of modifications to inhibit catalysis that may induce changes in the active site not present in the real system.[84-87]

There are a range of computational methods available with their individual usefulness dependent on the problem being considered. Many steps in biochemical processes require large shifts in biomolecular structure that occur over long timescales and therefore can be investigated

using molecular dynamics (MD) simulations that rely on molecular mechanics (MM) force fields to describe biomolecular structural dynamics.[86, 88-92] However, other aspects of enzyme function, including deciphering the preferred catalytic pathway, require an accurate description of the fine details of electronic configurations and therefore techniques grounded in quantum mechanics (QM) must be applied.[93-97] Although an abundance of information can be obtained by utilizing quantum mechanical methods to models ranging from tens to hundreds of atoms (quantum mechanical cluster methods),[93, 94] the elucidation of the catalytic mechanisms of action of enzymes through comparisons to experimental data requires more realistic representations of the entire nucleic acid–enzyme complex (thousands of atoms).[95-97] Such modeling has become computationally feasible with the development of QM/MM techniques,[96, 98] which divide a system into a QM region (tens to hundreds of atoms) and an MM region (thousands of atoms). This allows models to include entire enzyme complexes while still accounting for electron configurations necessary to elucidate mechanistic pathways. The QM/MM delimitation scheme can be combined with techniques to simulate dynamics to compose QM/MM MD techniques. While MD and QM/MM MD techniques can provide a dynamic description of enzyme−substrate complexes, many dynamic processes, including catalysis and substrate binding/unbinding, do not occur on timescales that can be captured by current computational resources. To study these processes, free energy methods are utilized. These methods add biases to increase sampling for low energy states (e.g., umbrella sampling (US), adaptively biased MD (abMD)) or simulate a (often non-physical) transformation between two states (thermodynamic integration (TI)).

Computational techniques have been shown to be successful in the literature for investigating glycosylase and endonuclease structure and function.[85, 99-120] As an example, one of the best studied glycosylases, uracil DNA glycosylase (UDG), has had extensive computational work done to characterize the catalytic mechanism.[108, 121, 122] MD simulations have given insights into UDG substrate recognition[121] and orientation in the active site,[122] finding that a chromatin environment promotes uracil base-flipping out of a DNA helix for detection by UDG and providing structures for further, quantum mechanical investigations of the catalytic mechanism. QM/MM calculations on a truncated UDG model were employed to map the reaction surface while considering two possible

orientations of a water nucleophile.[108] The resulting reaction surfaces correspond to a highly dissociative mechanism in agreement with experimental kinetic isotope effects (KIE),[123] with significant glycosidic bond lengthening during nucleophilic attack at C1′ of U in the transition state. These QM/MM calculations provided the first evidence that the energetically preferred pathway involves H148 activating the water nucleophile, shattering the long-held belief that D or E was always the final proton acceptor. Subsequent QM/MM calculations on the mechanism confirmed these findings,[106, 122] while also emphasizing the importance of an internal electric field in the active site for facilitating catalysis.[122] In addition to UDG, computational methods have provided key insights into the function of other enzymes, including endonuclease V,[124] AP endonuclease 1,[125] and thymine DNA glycosylase.[99] However, many endonucleases and glycosylases remain poorly understood, with unidentified catalytic mechanisms. This thesis uses a full range of computational methods to provide a comprehensive description of the glycosidic and phosphodiester bond cleavage facilitated by different glycosylases and endonucleases, highlighting the similarities and differences of how enzymes approach this difficult chemistry.

1.6. Scope of Thesis

This thesis applies MD, QM/MM, and free energy techniques to characterize the mechanism of action of different glycosylases and nucleases. Specifically, Chapter 2 uses MD simulations and QM/MM calculations to elucidate the debated mechanism of action for MutY (Figure 1.3A). MutY is a monofunctional glycosylase responsible for removing adenine from A:8-oxoguanine (8oG) mispairs.[126] While previous experimental and computational studies on MutY identified a direct hydrolysis mechanism,[111, 113, 127] more recent crystallographic and NMR studies[128, 129] propose the formation of a DNA−protein crosslink as part of the catalytic pathway. This chapter provides proof that MutY utilizes a crosslinking mechanism, resolving the controversy. Furthermore, this chapter investigates the structure of the N146S MutY mutant, giving insight into why the analogous N224S mutation in MUTYH results in MAP.

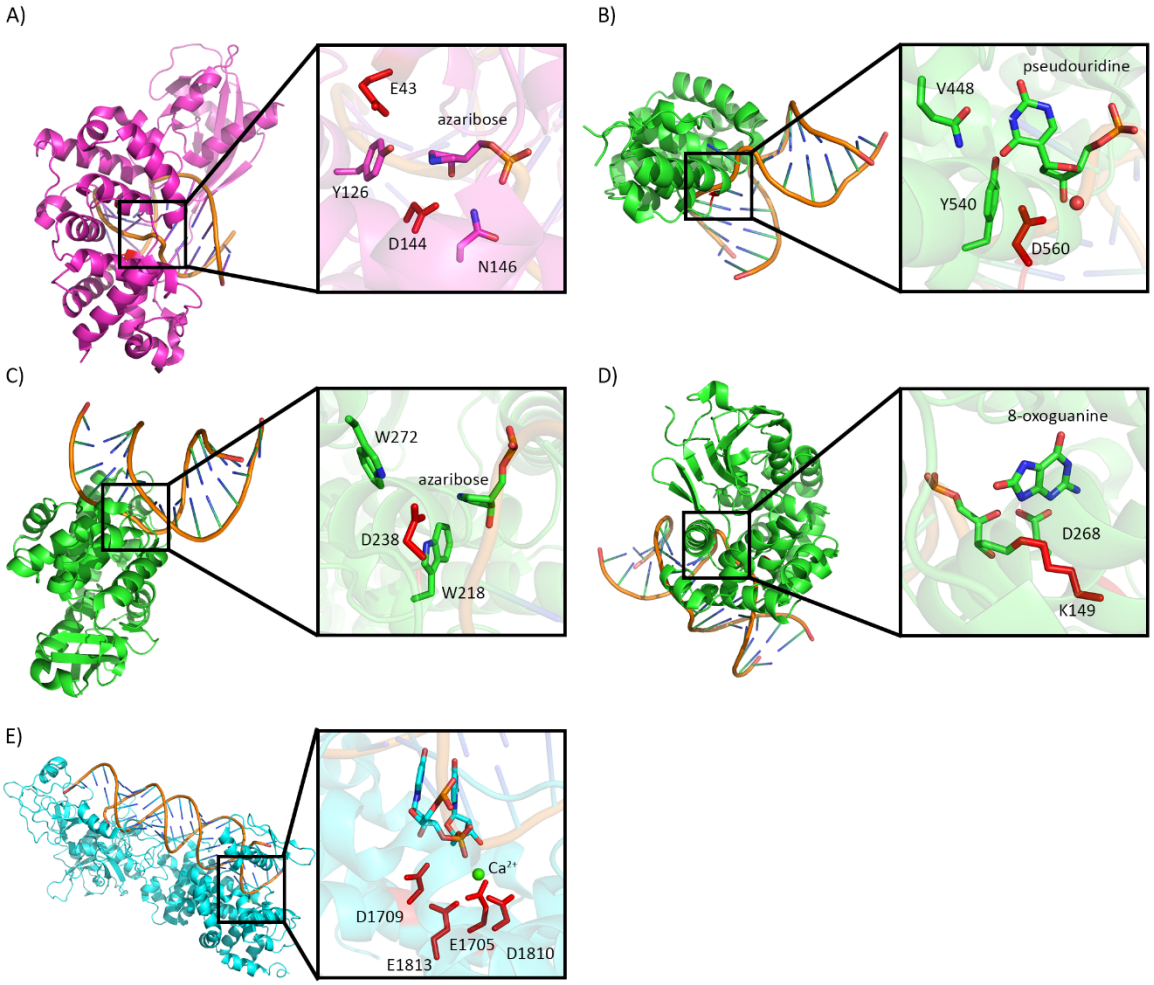


Figure 1.3. Crystal structures highlighting the active sites of A) MutY bound to a DNA transition state analogue inhibitor (PDB ID: 3G0Q), B) MBD4 bound to pseudouridine-containing DNA (PDB ID: 7KZ0), C) AlkA bound to DNA transition state analogue inhibitor (PDB ID: 1DIZ), D) hOGG1−DNA borohydride trapped intermediate (PDB ID: 1HU0), and E) Ca^{2+}-inhibited Dicer bound to substrate RNA (PDB ID: 7XW2). Residues implicated in catalysis are highlighted in red.

MutY was the first monofunctional glycosylase for which a crosslinking mechanism was characterized. However, evidence for crosslink formation exists in other enzymes as well. Specifically, the mechanism of action of MBD4, a monofunctional glycosylase that removes thymine from T:G mispairs resulting from the deamination of 3-methylcytosine (3mC),[130, 131] had been hypothesized in the literature to go through either a direct hydrolysis mechanism or a crosslinking mechanism based on two different crystal structures.[130, 132] To clarify this controversy, Chapter 3 uses abMD and QM/MM techniques to elucidate the mechanism of MBD4 (Figure 1.3B). This

chapter uncovered that MBD4 has a flexible active site, with catalysis proceeding through a direct hydrolysis mechanism to result in an anionic T product. This clarified the conflicting literature regarding MBD4 activity and paves the way for research into understanding and treating MANS.

AlkA is another monofunctional glycosylase that, while initially considered to cleave glycosidic bonds using a direct hydrolysis mechanism,[133] was also proposed to proceed through a DNA−protein crosslinked intermediate based on a crystal structure of AlkA bound to a transition state analogue.[134] Consequently, Chapter 4 uses MD simulations and QM/MM calculations to provide the first structural depiction of AlkA bound to a substrate as well as determine the mechanism of action (Figure 1.3C). AlkA is a bacterial monofunctional glycosylase that targets a wide variety of alkylated purines and pyrimidines, with one of its common substrates being 3-methyladenine (3mA).[134, 135] In this chapter, it was determined that AlkA binds 3mA through a large number of non-specific, hydrophobic contacts, which provides a rationale for AlkA substrate promiscuity. Additionally, catalysis was found to proceed through a direct hydrolysis mechanism, with the provided details of the mechanism promising to aid the development of inhibitors as anti-bacterial agents.[136] With the formation of a crosslinked intermediate investigated for three different enzymes in Chapters 2, 3, and 4, insights can be obtained on why MutY can facilitate crosslink formation while other enzymes cannot and provide a rationale for the apparent uniqueness of this catalytic pathway.

The previous chapters provide an in-depth description of the mechanism of action of monofunctional glycosylases. However, much less is known about bifunctional glycosylase catalysis. Therefore, Chapter 5 uses MD, umbrella sampling, and QM/MM MD simulations to investigate the β-lyase mechanism of hOGG1, a bifunctional glycosylase that removes 8oG damage (**Figure 1.3D**).[61, 137] While the glycosidic bond cleavage portion of the hOGG1 mechanism has been elucidated through several experimental[62, 63] and computational studies,[138-140] conflicting proposals for the β-lyase portion of the mechanism are present in the literature. Specifically, a crystal structure depicting hOGG1 with 8oG bound in the active site lead to a product-assisted mechanism being proposed,[61] while a kinetic study hypothesized D268 catalyzes phosphate

elimination.[141] However, neither proposal could fully account for all available kinetic and mutagenic data.[141-144] This chapter unveils a new mechanism for hOGG1 activity involving an elimination mechanism catalyzed by either D268 or K249, representing the first elucidated β-lyase mechanism for a bifunctional glycosylase.

Bifunctional glycosylases cleave the DNA backbone as part of their activity; however, nucleases are the main class of enzyme that facilitate phosphodiester bond cleavage in cells, with nuclease misfunction often causing disease.[18, 145] Chapter 6 uses thermodynamic integration and MD simulations alongside QM/MM calculations to determine the catalytic mechanism for Dicer and provides a rationale for Dicer mutations that result in DICER1 syndrome (Figure 1.3E). Dicer is a magnesium-dependent endonuclease that is vital for the production of miRNA in cells,[45-47] with mutations in the DICER1 hotspot region resulting in a disorder called DICER1 syndrome.[18] In this chapter, the catalytic Mg^{2+} binding configuration and catalytic mechanism of Dicer are identified, while DICER1 syndrome-causing mutants are revealed to impede functional metal−substrate orientations.

Chapter 7 summarizes how this thesis used a variety of computational techniques to characterize different glycosylase and endonuclease mechanisms. It illustrates the factors involved in the viability of DNA−protein crosslink formation during catalysis for monofunctional glycosylases. Additionally, the differing strategies for nucleic acid backbone cleavage utilized by bifunctional glycosylases and endonucleases are compared. Finally, future directions for research into other glycosylases and endonucleases using the computational methodology developed in this thesis are presented.

1.7. References

(1) Nick McElhinny, S. A.; Gordenin, D. A.; Stith, C. M.; Burgers, P. M.; Kunkel, T. A. Division of labor at the eukaryotic replication fork. *Mol. Cell* **2008**, *30* (2), 137-144. DOI: 10.1016/j.molcel.2008.02.022

(2) Shyu, A. B.; Wilkinson, M. F.; van Hoof, A. Messenger RNA regulation: To translate or to degrade. *EMBO J* **2008**, *27* (3), 471-481. DOI: 10.1038/sj.emboj.7601977

(3) Xu, T.; Li, L. X.; Jia, Y.; Wu, Q.; Zhu, W.; Xu, Z.; Zheng, B.; Lu, X. One microRNA has the potential to target whole viral mRNAs in a given human coronavirus. *Front Microbiol* **2022**, *13*, 1035044. DOI: 10.3389/fmicb.2022.1035044

(4) Schroeder, G. K.; Wolfenden, R. Rates of spontaneous disintegration of DNA and the rate enhancements produced by DNA glycosylases and deaminases. *Biochemistry* **2007**, *46* (47), 13638-13647. DOI: 10.1021/bi701480f

(5) Radzicka, A.; Wolfenden, R. A proficient enzyme. *Science* **1995**, *267* (5194), 90-93. (acccessed 2020/02/25/).

(6) Krokan, H. E.; Bjoras, M. Base excision repair. *Cold Spring Harb Perspect Biol* **2013**, *5* (4), a012583.

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

(8) Tatosyan, K. A.; Ustyantsev, I. G.; Kramerov, D. A. RNA degradation in eukaryotic cells. *Mol. Biol.* **2020**, *54* (4), 485-502. DOI: 10.1134/S0026893320040159

(9) Ha, M.; Kim, V. N. Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* **2014**, *15* (8), 509-524. DOI: 10.1038/nrm3838

(10) Chin, Y. P.; See, N. W.; Jenkins, I. D.; Krenske, E. H. Computational discoveries of reaction mechanisms: Recent highlights and emerging challenges. *Org. Biomol. Chem.* **2022**, *20* (10), 2028-2042. DOI: 10.1039/D1OB02139G

(11) Liang, J.; Zhen, P.; Gan, P.; Li, Y.; Tong, M.; Liu, W. DFT calculation of nonperiodic small molecular systems to predict the reaction mechanism of advanced oxidation processes: Challenges and perspectives. *ACS ES&T Engg.* **2024**, *4* (1), 4-18. DOI: 10.1021/acsestengg.3c00204

(12) Schramm, V. L. Enzymatic transition states, transition-state analogs, dynamics, thermodynamics, and lifetimes. *Annu. Rev. Biochem* **2011**, *80* (Volume 80, 2011), 703-732. DOI: https://doi.org/10.1146/annurev-biochem-061809-100742

(13) Schramm, V. L. Enzymatic transition states and drug design. *Chem. Rev.* **2018**, *118* (22), 11194-11258. DOI: 10.1021/acs.chemrev.8b00369

(14) Caldecott, K. W. Mammalian DNA base excision repair: Dancing in the moonlight. *DNA Repair* **2020**, *93*, 102921. DOI: https://doi.org/10.1016/j.dnarep.2020.102921

(15) Mazzei, F.; Viel, A.; Bignami, M. Role of MUTYH in human cancer. *Mutat. Res., Fundam. Mol. Mech. Mutagen.* **2013**, *743-744*, 33-43. DOI: 10.1016/j.mrfmmm.2013.03.003

(16) Al-Tassan, N.; Chmiel, N. H.; Maynard, J.; Fleming, N.; Livingston, A. L.; Williams, G. T.; Hodges, A. K.; Davies, D. R.; David, S. S.; Sampson, J. R.; et al. Inherited variants of MYH

associated with somatic G:C→t:A mutations in colorectal tumors. *Nat. Genet.* **2002**, *30* (2), 227. DOI: 10.1038/ng828

(17) Robinson, P. S.; Thomas, L. E.; Abascal, F.; Jung, H.; Harvey, L. M. R.; West, H. D.; Olafsson, S.; Lee, B. C. H.; Coorens, T. H. H.; Lee-Six, H.; et al. Inherited MUTYH mutations cause elevated somatic mutation rates and distinctive mutational signatures in normal human cells. *Nat. Commun.* **2022**, *13* (1), 3949. DOI: 10.1038/s41467-022-31341-0

(18) de Kock, L.; Wu, M. K.; Foulkes, W. D. Ten years of DICER1 mutations: Provenance, distribution, and associated phenotypes. *Hum. Mutat.* **2019**, *40* (11), 1939-1953. DOI: 10.1002/humu.23877

(19) Foulkes, W. D.; Priest, J. R.; Duchaine, T. F. DICER1: Mutations, microRNAs and mechanisms. *Nat. Rev. Cancer* **2014**, *14* (10), 662-672. DOI: 10.1038/nrc3802

(20) Robertson, J. C.; Jorcyk, C. L.; Oxford, J. T. DICER1 syndrome: DICER1 mutations in rare cancers. *Cancers* **2018**, *10* (5), 143/141-143/117. DOI: 10.3390/cancers10050143

(21) Rio Frio, T.; Bahubeshi, A.; Kanellopoulou, C.; Hamel, N.; Niedziela, M.; Sabbaghian, N.; Pouchet, C.; Gilbert, L.; O'Brien, P. K.; Serfas, K.; et al. DICER1 mutations in familial multinodular goiter with and without ovarian sertoli-leydig cell tumors. *JAMA* **2011**, *305* (1), 68-77. DOI: 10.1001/jama.2010.1910

(22) Doros, L. A.; Rossi, C. T.; Yang, J.; Field, A.; Williams, G. M.; Messinger, Y.; Cajaiba, M. M.; Perlman, E. J.; K, A. S.; Cathro, H. P.; et al. DICER1 mutations in childhood cystic nephroma and its relationship to DICER1-renal sarcoma. *Mod Pathol* **2014**, *27* (9), 1267-1280. DOI: 10.1038/modpathol.2013.242

(23) Slade, I.; Bacchelli, C.; Davies, H.; Murray, A.; Abbaszadeh, F.; Hanks, S.; Barfoot, R.; Burke, A.; Chisholm, J.; Hewitt, M.; et al. DICER1 syndrome: Clarifying the diagnosis, clinical features and management implications of a pleiotropic tumour predisposition syndrome. *J. Med. Genet.* **2011**, *48* (4), 273-278. DOI: 10.1136/jmg.2010.083790

(24) Schultz, K. A.; Yang, J.; Doros, L.; Williams, G. M.; Harris, A.; Stewart, D. R.; Messinger, Y.; Field, A.; Dehner, L. P.; Hill, D. A. DICER1-pleuropulmonary blastoma familial tumor predisposition syndrome: A unique constellation of neoplastic conditions. *Pathol Case Rev* **2014**, *19* (2), 90-100. DOI: 10.1097/pcr.0000000000000027

(25) Palles, C.; West, H. D.; Chew, E.; Galavotti, S.; Flensburg, C.; Grolleman, J. E.; Jansen, E. A. M.; Curley, H.; Chegwidden, L.; Arbe-Barnes, E. H.; et al. Germline MBD4 deficiency causes a multi-tumor predisposition syndrome. *Am. J. Hum. Genet.* **2022**, *109* (5), 953-960. DOI: https://doi.org/10.1016/j.ajhg.2022.03.018

(26) Schrempf, A.; Slyskova, J.; Loizou, J. I. Targeting the DNA repair enzyme polymerase θ in cancer therapy. *Trends Cancer* **2021**, *7* (2), 98-111. DOI: https://doi.org/10.1016/j.trecan.2020.09.007

(27) Duskova, K.; Lejault, P.; Benchimol, É.; Guillot, R.; Britton, S.; Granzhan, A.; Monchaud, D. DNA junction ligands trigger DNA damage and are synthetic lethal with DNA repair inhibitors in cancer cells. *J. Am. Chem. Soc.* **2020**, *142* (1), 424-435. DOI: 10.1021/jacs.9b11150

(28) Fuso Nerini, I.; Roca, E.; Mannarino, L.; Grosso, F.; Frapolli, R.; D'Incalci, M. Is DNA repair a potential target for effective therapies against malignant mesothelioma? *Cancer Treat. Rev.* **2020**, *90*, 102101. DOI: https://doi.org/10.1016/j.ctrv.2020.102101

(29) Mechetin, G. V.; Endutkin, A. V.; Diatlova, E. A.; Zharkov, D. O. Inhibitors of DNA glycosylases as prospective drugs. *Int. J. Mol. Sci.* **2020**, *21* (9), 3118.

(30) Jiang, M.; Jia, K.; Wang, L.; Li, W.; Chen, B.; Liu, Y.; Wang, H.; Zhao, S.; He, Y.; Zhou, C. Alterations of DNA damage repair in cancer: From mechanisms to applications. *Ann. Transl. Med.* **2020**, *8* (24), 1685.

(31) Gad, H.; Koolmeister, T.; Jemth, A.-S.; Eshtad, S.; Jacques, S. A.; Ström, C. E.; Svensson, L. M.; Schultz, N.; Lundbäck, T.; Einarsdottir, B. O.; et al. MTH1 inhibition eradicates cancer by preventing sanitation of the dNTP pool. *Nature* **2014**, *508* (7495), 215-221. DOI: 10.1038/nature13181

(32) Huber, K. V. M.; Salah, E.; Radic, B.; Gridling, M.; Elkins, J. M.; Stukalov, A.; Jemth, A.-S.; Göktürk, C.; Sanjiv, K.; Strömberg, K.; et al. Stereospecific targeting of MTH1 by (S)-crizotinib as an anticancer strategy. *Nature* **2014**, *508* (7495), 222-227. DOI: 10.1038/nature13194

(33) Helleday, T.; Petermann, E.; Lundin, C.; Hodgson, B.; Sharma, R. A. DNA repair pathways as targets for cancer therapy. *Nat. Rev. Cancer* **2008**, *8* (3), 193-204. DOI: 10.1038/nrc2342

(34) Casacuberta, J. M.; Devos, Y.; du Jardin, P.; Ramon, M.; Vaucheret, H.; Nogué, F. Biotechnological uses of RNAi in plants: Risk assessment considerations. *Trends Biotechnol.* **2015**, *33* (3), 145-147. DOI: https://doi.org/10.1016/j.tibtech.2014.12.003

(35) Moreira-Pinto, C. E.; Coelho, R. R.; Leite, A. G. B.; Silveira, D. A.; de Souza, D. A.; Lopes, R. B.; Macedo, L. L. P.; Silva, M. C. M.; Ribeiro, T. P.; Morgante, C. V.; et al. Increasing anthonomus grandis susceptibility to metarhizium anisopliae through RNAi-induced agrarelish knockdown: A perspective to combine biocontrol and biotechnology. *Pest Manag. Sci.* **2021**, *77* (9), 4054-4063. DOI: https://doi.org/10.1002/ps.6430

(36) Sherman, J. H.; Munyikwa, T.; Chan, S. Y.; Petrick, J. S.; Witwer, K. W.; Choudhuri, S. RNAi technologies in agricultural biotechnology: The toxicology forum 40th annual summer meeting. *Regul. Toxicol. Pharm.* **2015**, *73* (2), 671-680. DOI: https://doi.org/10.1016/j.yrtph.2015.09.001

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

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

(39) Weinberg, C. E.; Weinberg, Z.; Hammann, C. Novel ribozymes: Discovery, catalytic mechanisms, and the quest to understand biological function. *Nucleic Acids Res* **2019**, *47* (18), 9480-9494.

(40) Kozak, M. Initiation of translation in prokaryotes and eukaryotes. *Gene* **1999**, *234* (2), 187-208. DOI: 10.1016/S0378-1119(99)00210-3

(41) Meng, Y.; Meng, L.; Song, Y.; Liu, M.; Zhang, X. Small RNA molecules and regulation of spermatogenesis. *Yichuan* **2011**, *33* (1), 9-16. DOI: 10.3724/SP.J.1005.2011.00009

(42) Loenen, W. A. M.; Dryden, D. T. F.; Raleigh, E. A.; Wilson, G. G.; Murray, N. E. Highlights of the DNA cutters: A short history of the restriction enzymes. *Nucleic Acids Res.* **2013**, *42* (1), 3-19. DOI: 10.1093/nar/gkt990 (acccessed 3/7/2024).

(43) Hindley, H. J.; Gong, Z.; Moradian, S.; Giuliano, M. G.; Sapelkin, A.; Kotta-Loizou, I.; Buck, M.; Engl, C.; Weiße, A. Y. Heterogeneity in responses to ribosome-targeting antibiotics mediated by bacterial RNA repair. *Nat. Commun.* **2025**, *16* (1), 9620. DOI: 10.1038/s41467-025-64759-3

(44) Lehmann, K.; and Schmidt, U. Group II introns: Structure and catalytic versatility of large natural ribozymes. *Crit. Rev. Biochem. Mol. Biol.* **2003**, *38* (3), 249-303. DOI: 10.1080/713609236

(45) Yoshida, T.; Asano, Y.; Ui-Tei, K. Modulation of microRNA processing by Dicer via its associated dsRNA binding proteins. *Non-Coding RNA* **2021**, *7* (3). DOI: 10.3390/ncrna7030057

(46) Zapletal, D.; Kubicek, K.; Svoboda, P.; Stefl, R. Dicer structure and function: Conserved and evolving features. *EMBO Rep.* **2023**, *24* (7), e57215. DOI: https://doi.org/10.15252/embr.202357215

(47) Ambros, V.; Bartel, B.; Bartel, D. P.; Burge, C. B.; Carrington, J. C.; Chen, X.; Dreyfuss, G.; Eddy, S. R.; Griffiths-Jones, S.; Marshall, M.; et al. A uniform system for microRNA annotation. *RNA* **2003**, *9* (3), 277-279. DOI: 10.1261/rna.2183803

(48) Grindley, N. D. F.; Whiteson, K. L.; Rice, P. A. Mechanisms of site-specific recombination. *Annu. Rev. Biochem* **2006**, *75* (1), 567-605. DOI: 10.1146/annurev.biochem.73.011303.073908

(49) Turrens, J. F. Mitochondrial formation of reactive oxygen species. *J. Physiol.* **2003**, *552* (2), 335-344. DOI: https://doi.org/10.1111/j.1469-7793.2003.00335.x

(50) Mittal, M.; Siddiqui, M. R.; Tran, K.; Reddy, S. P.; Malik, A. B. Reactive oxygen species in inflammation and tissue injury. *Antioxid. Redox Singal.* **2014**, *20* (7), 1126-1167. DOI: 10.1089/ars.2012.5149

(51) Tominaga, H.; Kodama, S.; Matsuda, N.; Suzuki, K.; Watanabe, M. Involvement of reactive oxygen species (ROS) in the induction of genetic instability by radiation. *J. Radiat. Res.* **2004**, *45* (2), 181-188. DOI: 10.1269/jrr.45.181 (acccessed 4/22/2025).

(52) Prasad, S.; Gupta, S. C.; Tyagi, A. K. Reactive oxygen species (ROS) and cancer: Role of antioxidative nutraceuticals. *Cancer Lett.* **2017**, *387*, 95-105. DOI: https://doi.org/10.1016/j.canlet.2016.03.042

(53) Chen, Q.; Wang, Q.; Zhu, J.; Xiao, Q.; Zhang, L. Reactive oxygen species: Key regulators in vascular health and diseases. *Br. J. Pharmacol.* **2018**, *175* (8), 1279-1292. DOI: https://doi.org/10.1111/bph.13828

(54) Palmer, J. E.; Wilson, N.; Son, S. M.; Obrocki, P.; Wrobel, L.; Rob, M.; Takla, M.; Korolchuk, V. I.; Rubinsztein, D. C. Autophagy, aging, and age-related neurodegeneration. *Neuron* **2025**, *113* (1), 29-48. DOI: https://doi.org/10.1016/j.neuron.2024.09.015

(55) Gogna, T.; Housden, B. E.; Houldsworth, A. Exploring the role of reactive oxygen species in the pathogenesis and pathophysiology of alzheimer's and parkinson's disease and the efficacy of antioxidant treatment. *Antioxidants* **2024**, *13* (9), 1138.

(56) Mukherjee, A.; Ghosh, K. K.; Chakrabortty, S.; Gulyás, B.; Padmanabhan, P.; Ball, W. B. Mitochondrial reactive oxygen species in infection and immunity. *Biomolecules* **2024**, *14* (6), 670.

(57) Mullins, E. A.; Rodriguez, A. A.; Bradley, N. P.; Eichman, B. F. Emerging roles of DNA glycosylases and the base excision repair pathway. *Trends Biochem. Sci* **2019**, *44* (9), 765-781. DOI: https://doi.org/10.1016/j.tibs.2019.04.006

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

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

(60) 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.

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

(62) Nash, H. M.; Lu, R.; Lane, W. S.; Verdinel, G. L. The critical active-site amine of the human 8-oxoguanine DNA glycosylase, hOGG1: Direct identification, ablation and chemical reconstitution. *Chem. Biol.* **1997**, *4* (9), 693-702. DOI: https://doi.org/10.1016/S1074-5521(97)90225-8

(63) Tyugashev, T. E.; Vorobjev, Y. N.; Kuznetsova, A. A.; Lukina, M. V.; Kuznetsov, N. A.; Fedorova, O. S. Roles of active-site amino acid residues in specific recognition of DNA lesions by human 8-oxoguanine-DNA glycosylase (OGG1). *J. Phys. Chem. B* **2019**, *123* (23), 4878-4887. DOI: 10.1021/acs.jpcb.9b02949

(64) Zheng, J.-H.; Tan, H.-W.; Chen, G.-J. Theoretical study on the mechanism of the DNA repair protein Fpg. *Int. J. Quantum Chem* **2011**, *111* (10), 2454-2463. DOI: https://doi.org/10.1002/qua.22447

(65) McCullough, A. K.; Dodson, M. L.; Lloyd, R. S. Initiation of base excision repair: Glycosylase mechanisms and structures. *Annu. Rev. Biochem* **1999**, *68* (Volume 68, 1999), 255-285. DOI: https://doi.org/10.1146/annurev.biochem.68.1.255

(66) Pushing cancer over the edge. *Chemical & Engineering News Archive* **2013**, *91* (24), 13-18. DOI: 10.1021/cen-09124-cover

(67) He, L.; Yang, H.; Zhou, S.; Zhu, H.; Mao, H.; Ma, Z.; Wu, T.; Kumar, A. K.; Kathera, C.; Janardhan, A.; et al. Synergistic antitumor effect of combined paclitaxel with FEN1 inhibitor in cervical cancer cells. *DNA Repair* **2018**, *63*, 1-9. DOI: https://doi.org/10.1016/j.dnarep.2018.01.003

(68) Wu, T.; Zhu, H.; Zhang, M.; Sun, Y.; Yang, Y.; Gu, L.; Zhang, J.; Mu, D.; Wu, C.; Hu, Z.; et al. FEN1 inhibitor synergizes with low-dose camptothecin to induce increased cell killing via the mitochondria mediated apoptotic pathway. *Gene Ther.* **2022**, *29* (7), 407-417. DOI: 10.1038/s41434-020-00215-9

(69) Sedger, L. M. microRNA control of interferons and interferon induced anti-viral activity. *Mol. Immunol.* **2013**, *56* (4), 781-793. DOI: https://doi.org/10.1016/j.molimm.2013.07.009

(70) Ying, H.; Ebrahimi, M.; Keivan, M.; Khoshnam, S. E.; Salahi, S.; Farzaneh, M. miRNAs; a novel strategy for the treatment of COVID-19. *Cell Biol. Int.* **2021**, *45* (10), 2045-2053. DOI: https://doi.org/10.1002/cbin.11653

(71) Lauková, L.; Konečná, B.; Janovičová, Ľ.; Vlková, B.; Celec, P. Deoxyribonucleases and their applications in biomedicine. *Biomolecules* **2020**, *10* (7), 1036. DOI: 10.3390/biom10071036

(72) Yang, W. Nucleases: Diversity of structure, function and mechanism. *Q Rev Biophys* **2011**, *44* (1), 1-93. DOI: 10.1017/s0033583510000181

(73) Steitz, T. A.; Steitz, J. A. A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. U.S.A* **1993**, *90* (14), 6498-6502. DOI: 10.1073/pnas.90.14.6498

(74) Dupureur, C. M. One is enough: Insights into the two-metal ion nuclease mechanism from global analysis and computational studies. *Metallomics* **2010**, *2* (9), 609-620. DOI: 10.1039/c0mt00013b (acccessed 3/8/2024).

(75) Pingoud, V.; Wende, W.; Friedhoff, P.; Reuter, M.; Alves, J.; Jeltsch, A.; Mones, L.; Fuxreiter, M.; Pingoud, A. On the divalent metal ion dependence of DNA cleavage by restriction endonucleases of the EcoRI family. *J. Mol. Biol.* **2009**, *393* (1), 140-160. DOI: https://doi.org/10.1016/j.jmb.2009.08.011

(76) Nikkel, D. J.; Kaur, R.; Wetmore, S. D. How can one metal power nucleic acid phosphodiester bond cleavage by a nuclease? Multiscale computational studies highlight a diverse mechanistic landscape. *J. Phys. Chem. B* **2025**, *129* (1), 3-18. DOI: 10.1021/acs.jpcb.4c05875

(77) Crow, Y. J.; Rehwinkel, J. Aicardi-goutieres syndrome and related phenotypes: Linking nucleic acid metabolism with autoimmunity. *Hum Mol Genet* **2009**, *18* (R2), R130-136. DOI: 10.1093/hmg/ddp293

(78) Kashiyama, K.; Nakazawa, Y.; Pilz, Daniela T.; Guo, C.; Shimada, M.; Sasaki, K.; Fawcett, H.; Wing, Jonathan F.; Lewin, Susan O.; Carr, L.; et al. Malfunction of nuclease ERCC1-XPF results in diverse clinical manifestations and causes cockayne syndrome, xeroderma pigmentosum, and fanconi anemia. *Am. J. Hum. Genet.* **2013**, *92* (5), 807-819. DOI: https://doi.org/10.1016/j.ajhg.2013.04.007

(79) Belfort, M.; Bonocora, R. P. Homing endonucleases: From genetic anomalies to programmable genomic clippers. In *Homing endonucleases: Methods and protocols*, Edgell, D. R. Ed.; Humana Press, 2014; pp 1-26.

(80) Tang, N.; Ji, Q. Miniature CRISPR-Cas12 systems: Mechanisms, engineering, and genome editing applications. *ACS Chem. Biol.* **2024**, *19* (7), 1399-1408. DOI: 10.1021/acschembio.4c00247

(81) Turner, D. J.; Pingle, M. R.; Barany, F. Harnessing asymmetrical substrate recognition by thermostable EndoV to achieve balanced linear amplification in multiplexed SNP typing. *Biochem. Cell Biol.* **2006**, *84* (2), 232-242. DOI: 10.1139/o06-025

(82) Ye, X.; Li, Z.; Fang, Z.; Shen, N.; Zhou, Y.; Sun, P.; Xu, H.; Liu, Q.; Feng, Y. Thermal activation of argonaute nuclease enables one-pot multiplex detection of viruses. *Sens. Actuators B Chem.* **2024**, *409*, 135587. DOI: https://doi.org/10.1016/j.snb.2024.135587

(83) Harder, O. F.; Barrass, S. V.; Drabbels, M.; Lorenz, U. J. Fast viral dynamics revealed by microsecond time-resolved cryo-EM. *Nat. Commun.* **2023**, *14* (1), 5649. DOI: 10.1038/s41467-023-41444-x

(84) Ryde, U.; Dong, G.; Li, J.; Feldt, M.; Mata, R. A. Computational studies of molybdenum and tungsten enzymes. In *Molybdenum and tungsten enzymes: Spectroscopic and theoretical investigations*, Hille, R., Schulzke, C., Kirk, M. L., Kirk, M. L., Hille, R., Schulzke, C. Eds.; The Royal Society of Chemistry, 2016.

(85) Kaur, R.; Nikkel, D. J.; Wetmore, S. D. Computational studies of DNA repair: Insights into the function of monofunctional DNA glycosylases in the base excision repair pathway. *WIREs Comput. Mol. Sci.* **2020**, *10* (5), e1471. DOI: 10.1002/wcms.1471 (acccessed 2020/09/16).

(86) Dissanayake, U. C.; Roy, A.; Maghsoud, Y.; Polara, S.; Debnath, T.; Cisneros, G. A. Computational studies on the functional and structural impact of pathogenic mutations in enzymes. *Protein Sci.* **2025**, *34* (4), e70081. DOI: https://doi.org/10.1002/pro.70081

(87) Elsässer, B.; Goettig, P. Mechanisms of proteolytic enzymes and their inhibition in QM/MM studies. *Int. J. Mol. Sci.* **2021**, *22* (6), 3232.

(88) Šponer, J.; Banáš, P.; Jurečka, P.; Zgarbová, M.; Kührová, P.; Havrila, M.; Krepl, M.; Stadlbauer, P.; Otyepka, M. Molecular dynamics simulations of nucleic acids. From tetranucleotides to the ribosome. *J. Phys. Chem. Lett.* **2014**, *5* (10), 1771-1782. DOI: 10.1021/jz500557y

(89) van der Kamp, M. W.; Daggett, V. Molecular dynamics as an approach to study prion protein misfolding and the effect of pathogenic mutations. In *Prion proteins*, Tatzelt, J. Ed.; Springer Berlin Heidelberg, 2011; pp 169-197.

(90) Osuna, S.; Jiménez-Osés, G.; Noey, E. L.; Houk, K. N. Molecular dynamics explorations of active site structure in designed and evolved enzymes. *Acc. Chem. Res.* **2015**, *48* (4), 1080-1089. DOI: 10.1021/ar500452q

(91) Fadda, E.; Woods, R. J. Molecular simulations of carbohydrates and protein–carbohydrate interactions: Motivation, issues and prospects. *Drug Discov. Today* **2010**, *15* (15), 596-609. DOI: https://doi.org/10.1016/j.drudis.2010.06.001

(92) Shaikh, S. A.; Li, J.; Enkavi, G.; Wen, P.-C.; Huang, Z.; Tajkhorshid, E. Visualizing functional motions of membrane transporters with molecular dynamics simulations. *Biochemistry* **2013**, *52* (4), 569-587. DOI: 10.1021/bi301086x

(93) Blomberg, M. R.; Borowski, T.; Himo, F.; Liao, R.-Z.; Siegbahn, P. E. Quantum chemical studies of mechanisms for metalloenzymes. *Chem. Rev.* **2014**, *114* (7), 3601-3658. DOI: https://doi.org/10.1021/cr400388t

(94) Himo, F. Recent trends in quantum chemical modeling of enzymatic reactions. *J. Am. Chem. Soc.* **2017**, *139* (20), 6780-6786. DOI: 10.1021/jacs.7b02671

(95) Senn, H. M.; Thiel, W. QM/MM studies of enzymes. *Curr. Opin. Chem. Biol.* **2007**, *11* (2), 182-187. DOI: https://doi.org/10.1016/j.cbpa.2007.01.684

(96) Vennelakanti, V.; Nazemi, A.; Mehmood, R.; Steeves, A. H.; Kulik, H. J. Harder, better, faster, stronger: Large-scale QM and QM/MM for predictive modeling in enzymes and proteins. *Curr. Opin. Struct. Biol.* **2022**, *72*, 9-17. DOI: https://doi.org/10.1016/j.sbi.2021.07.004

(97) Sousa, S. F.; Ribeiro, A. J. M.; Neves, R. P. P.; Brás, N. F.; Cerqueira, N. M. F. S. A.; Fernandes, P. A.; Ramos, M. J. Application of quantum mechanics/molecular mechanics methods in the study of enzymatic reaction mechanisms. *WIREs Comput. Mol. Sci.* **2017**, *7* (2), e1281. DOI: https://doi.org/10.1002/wcms.1281

(98) Tzeliou, C. E.; Mermigki, M. A.; Tzeli, D. Review on the QM/MM methodologies and their application to metalloproteins. *Molecules* **2022**, *27* (9), 2660.

(99) Diao, W.; Peng, J.; Fu, X.; Ye, F.; Wang, Z. Preorganized internal electric field enables nucleophilic attack of a nonactivated water molecule in thymine DNA glycosylase. *J. Am. Chem. Soc.* **2025**, *147* (29), 25919-25930. DOI: 10.1021/jacs.5c08420

(100) Nikkel, D. J.; Wetmore, S. D. Distinctive formation of a DNA–protein cross-link during the repair of DNA oxidative damage: Insights into human disease from MD simulations and QM/MM calculations. *J. Am. Chem. Soc.* **2023**, *145* (24), 13114-13125. DOI: 10.1021/jacs.3c01773

(101) Unno, M.; Morikawa, M.; Sychrovský, V.; Koga, M.; Minowa, N.; Komuro, S.; Shimizu, M.; Fukuta, M.; Tsuyuguchi, F.; Mano, H.; et al. Capturing a glycosylase reaction intermediate in DNA repair by freeze-trapping of a pH-responsive hOGG1 mutant. *Nucleic Acids Res.* **2025**, *53* (14). DOI: 10.1093/nar/gkaf718 (acccessed 8/25/2025).

(102) Rutledge, L. R.; Wetmore, S. D. Modeling the chemical step utilized by human alkyladenine DNA glycosylase: A concerted mechanism aids in selectively excising damaged purines. *J. Am. Chem. Soc.* **2011**, *133* (40), 16258-16269. DOI: 10.1021/ja207181c

(103) Wang, B.; Usharani, D.; Li, C.; Shaik, S. Theory uncovers an unusual mechanism of DNA repair of a lesioned adenine by AlkB enzymes. *J. Am. Chem. Soc.* **2014**, *136* (39), 13895-13901. DOI: 10.1021/ja507934g

(104) Kreppel, A.; Blank, I. D.; Ochsenfeld, C. Base-independent DNA base-excision repair of 8-oxoguanine. *J. Am. Chem. Soc.* **2018**, *140* (13), 4522-4526. DOI: 10.1021/jacs.7b11254

(105) Sadeghian, K.; Ochsenfeld, C. Unraveling the base excision repair mechanism of human DNA glycosylase. *J. Am. Chem. Soc.* **2015**, *137* (31), 9824-9831. DOI: 10.1021/jacs.5b01449

(106) Naydenova, E.; Roßbach, S.; Ochsenfeld, C. QM/MM study of the uracil DNA glycosylase reaction mechanism: A competition between Asp145 and His148. *J. Chem. Theory Comput.* **2019**, *15* (8), 4344-4350. DOI: 10.1021/acs.jctc.8b01305

(107) Melayikandy, S.; Krishnan, A.; Varghese, A.; Jaber Sathik Rifayee, S. B.; Waheed, S. O.; Ramanan, R.; Li, D.; Christov, C. Z.; Karabencheva-Christova, T. G. Catalytic mechanism of the bacterial non-heme Fe(II) and 2-oxoglutarate dependent enzyme AlkB with single-stranded DNA containing complex guanine adducts. *Inorg. Chem.* **2025**, *64* (30), 15650-15666. DOI: 10.1021/acs.inorgchem.5c02176

(108) Przybylski, J. L.; Wetmore, S. D. A QM/QM investigation of the hUNG2 reaction surface: The untold tale of a catalytic residue. *Biochemistry* **2011**, *50* (19), 4218-4227. DOI: 10.1021/bi2003394

(109) 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. DOI: 10.1021/acs.jpcb.7b09646

(110) Wang, W.-J.; Wang, T.; Zhao, Y.; Li, B.-N.; Chen, D.-Z. Theoretical insights into N-glycoside bond cleavage of 5-carboxycytosine by thymine DNA glycosylase: A QM/MM study. *J. Phys. Chem. B* **2024**, *128* (19), 4621-4630. DOI: 10.1021/acs.jpcb.4c00221

(111) 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* (48), 8753-8765. DOI: 10.1021/bi401310w

(112) Diao, W.; Farrell, J. D.; Wang, B.; Ye, F.; Wang, Z. Preorganized internal electric field promotes a double-displacement mechanism for the adenine excision reaction by adenine DNA glycosylase. *J. Phys. Chem. B* **2023**, *127* (40), 8551-8564. DOI: 10.1021/acs.jpcb.3c04928

(113) 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. DOI: 10.1021/ja711363s

(114) Tiwari, S.; Agnihotri, N.; Mishra, P. C. Quantum theoretical study of cleavage of the glycosidic bond of 2′-deoxyadenosine: Base excision-repair mechanism of DNA by MutY. *J. Phys. Chem. B* **2011**, *115* (12), 3200-3207. DOI: 10.1021/jp1109256

(115) 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. DOI: 10.1021/ja301714j

(116) Trasviña-Arenas, C. H.; Dissanayake, U. C.; Tamayo, N.; Hashemian, M.; Lin, W.-J.; Demir, M.; Hoyos-Gonzalez, N.; Fisher, A. J.; Cisneros, G. A.; Horvath, M. P.; et al. Structure of human MUTYH and functional profiling of cancer-associated variants reveal an allosteric network between its [4Fe-4S] cluster cofactor and active site required for DNA repair. *Nat. Commun.* **2025**, *16* (1), 3596. DOI: 10.1038/s41467-025-58361-w

(117) Abyzov, A.; Uzun, A.; Strauss, P. R.; Ilyin, V. A. An AP endonuclease 1–DNA polymerase β complex: Theoretical prediction of interacting surfaces. *PLOS Comput. Biol.* **2008**, *4* (4), e1000066. DOI: 10.1371/journal.pcbi.1000066

(118) Krumkacheva, O. A.; Shevelev, G. Y.; Lomzov, A. A.; Dyrkheeva, N. S.; Kuzhelev, A. A.; Koval, V. V.; Tormyshev, V. M.; Polienko, Y. F.; Fedin, M. V.; Pyshnyi, D. V.; et al. DNA complexes with human apurinic/apyrimidinic endonuclease 1: Structural insights revealed by pulsed dipolar EPR with orthogonal spin labeling. *Nucleic Acids Res.* **2019**, *47* (15), 7767-7780. DOI: 10.1093/nar/gkz620 (acccessed 10/31/2024).

(119) Doss, C. G. P.; NagaSundaram, N. Investigating the structural impacts of I64T and P311S mutations in APE1-DNA complex: A molecular dynamics approach. *PLOS ONE* **2012**, *7* (2), e31677. DOI: 10.1371/journal.pone.0031677

(120) Guliaev, A. B.; Hang, B.; Singer, B. Structural insights by molecular dynamics simulations into specificity of the major human AP endonuclease toward the benzene-derived DNA adduct, pBQ-C. *Nucleic Acids Res.* **2004**, *32* (9), 2844-2852. DOI: 10.1093/nar/gkh594 (acccessed 10/31/2024).

(121) Saravanan, V.; Raouraoua, N.; Brysbaert, G.; Giordano, S.; Lensink, M. F.; Cleri, F.; Blossey, R. The ‘very moment’ when UDG recognizes a flipped-out uracil base in dsDNA. *Sci. Rep.* **2025**, *15* (1), 7993. DOI: 10.1038/s41598-025-91705-6

(122) Diao, W.; Yan, S.; Farrell, J. D.; Wang, B.; Ye, F.; Wang, Z. Preorganized internal electric field powers catalysis in the active site of uracil-DNA glycosylase. *ACS Catal.* **2022**, *12* (20), 12488-12499. DOI: 10.1021/acscatal.2c02886

(123) Werner, R. M.; Stivers, J. T. Kinetic isotope effect studies of the reaction catalyzed by uracil DNA glycosylase: Evidence for an oxocarbenium ion−uracil anion intermediate. *Biochemistry* **2000**, *39* (46), 14054-14064. DOI: 10.1021/bi0018178

(124) Kaur, R.; Wetmore, S. D. Is metal stabilization of the leaving group required or can lysine facilitate phosphodiester bond cleavage in nucleic acids? A computational study of EndoV. *J. Chem Inf. Model.* **2024**, *64* (3), 944-959. DOI: 10.1021/acs.jcim.3c01775

(125) Aboelnga, M. M.; Wetmore, S. D. Unveiling a single-metal-mediated phosphodiester bond cleavage mechanism for nucleic acids: A multiscale computational investigation of a human DNA repair enzyme. *J. Am. Chem. Soc.* **2019**, *141* (21), 8646-8656. DOI: 10.1021/jacs.9b03986

(126) Michaels, M. L.; Tchou, J.; Grollman, A. P.; Miller, J. H. A repair system for 8-oxo-7,8-dihydrodeoxyguanine. *Biochemistry* **1992**, *31* (45), 10964-10968. DOI: 10.1021/bi00160a004

(127) Fromme, J. C.; Banerjee, A.; Huang, S. J.; Verdine, G. L. Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase. *Nature* **2004**, *427* (6975), 652-656. DOI: 10.1038/nature02306

(128) Woods, R. D.; O'Shea, V. L.; Chu, A.; Cao, S.; Richards, J. L.; Horvath, M. P.; David, S. S. Structure and stereochemistry of the base excision repair glycosylase MutY reveal a mechanism similar to retaining glycosidases. *Nucleic Acids Res.* **2016**, *44* (2), 801-810. DOI: 10.1093/nar/gkv1469

(129) Demir, M.; Russelburg, L. P.; Lin, W.-J.; Trasviña-Arenas, Carlos H.; Huang, B.; Yuen, Philip K.; Horvath, Martin P.; David, Sheila S. Structural snapshots of base excision by the cancer-associated variant MutY N146S reveal a retaining mechanism. *Nucleic Acids Res.* **2023**, *51* (3), 1034-1049. DOI: 10.1093/nar/gkac1246 (acccessed 7/24/2025).

(130) Hashimoto, H.; Zhang, X.; Cheng, X. Excision of thymine and 5-hydroxymethyluracil by the MBD4 DNA glycosylase domain: Structural basis and implications for active DNA demethylation. *Nucleic Acids Res.* **2012**, *40* (17), 8276-8284. DOI: 10.1093/nar/gks628 (acccessed 1/14/2025).

(131) Turner, D. P.; Cortellino, S.; Schupp, J. E.; Caretti, E.; Loh, T.; Kinsella, T. J.; Bellacosa, A. The DNA N-Glycosylase MED1 exhibits preference for halogenated pyrimidines and is involved in the cytotoxicity of 5-iododeoxyuridine. *Cancer Res.* **2006**, *66* (15), 7686-7693. DOI: 10.1158/0008-5472.Can-05-4488 (acccessed 1/17/2025).

(132) Pidugu, L. S.; Bright, H.; Lin, W.-J.; Majumdar, C.; Van Ostrand, R. P.; David, S. S.; Pozharski, E.; Drohat, A. C. Structural insights into the mechanism of base excision by MBD4. *J. Mol. Biol.* **2021**, *433* (15), 167097. DOI: https://doi.org/10.1016/j.jmb.2021.167097

(133) Labahn, J.; Schärer, O. D.; Long, A.; Ezaz-Nikpay, K.; Verdine, G. L.; Ellenberger, T. E. Structural basis for the excision repair of alkylation-damaged DNA. *Cell* **1996**, *86* (2), 321-329. DOI: https://doi.org/10.1016/S0092-8674(00)80103-8

(134) Hollis, T.; Ichikawa, Y.; Ellenberger, T. DNA bending and a flip-out mechanism for base excision by the helix-hairpin-helix DNA glycosylase, escherichia coli AlkA. *EMBO J* **2000**, *19* (4), 758-766. DOI: 10.1093/emboj/19.4.758

(135) Dylewska, M.; Dąbrowska, I.; Ćwiek, K.; Padoł, K.; Mielecki, D.; Sokołowska, B.; Poznański, J.; Maciejewska, A. M. AlkA glycosylase and AlkB dioxygenase constitute an effective protective system for endogenously arising acrolein: E. Coli AlkA glycosylase excises acrolein adduct to adenine. *J. Mol. Biol.* **2025**, *437* (2), 168912. DOI: https://doi.org/10.1016/j.jmb.2024.168912

(136) Lim, C. S. Q.; Ha, K. P.; Clarke, R. S.; Gavin, L.-A.; Cook, D. T.; Hutton, J. A.; Sutherell, C. L.; Edwards, A. M.; Evans, L. E.; Tate, E. W.; et al. Identification of a potent small-molecule inhibitor of bacterial DNA repair that potentiates quinolone antibiotic activity in methicillin-resistant staphylococcus aureus. *Biorg. Med. Chem.* **2019**, *27* (20), 114962. DOI: https://doi.org/10.1016/j.bmc.2019.06.025

(137) Kuznetsov, N. A.; Koval, V. V.; Fedorova, O. S. Mechanism of recognition and repair of damaged DNA by human 8-oxoguanine DNA glycosylase hOGG1. *Biochemistry* **2011**, *76* (1), 118-130. DOI: 10.1134/S0006297911010123

(138) Šebera, J.; Trantírek, L.; Tanaka, Y.; Sychrovský, V. Pyramidalization of the glycosidic nitrogen provides the way for efficient cleavage of the N-glycosidic bond of 8-oxoG with the hOGG1 DNA repair protein. *J. Phys. Chem. B* **2012**, *116* (41), 12535-12544. DOI: 10.1021/jp309098d

(139) Šebera, J.; Trantírek, L.; Tanaka, Y.; Nencka, R.; Fukal, J.; Sychrovský, V. The activation of N-glycosidic bond cleavage performed by base-excision repair enzyme hOGG1; theoretical study of the role of Lys 249 residue in activation of G, OxoG and FapyG. *RSC Adv.* **2014**, *4* (83), 44043-44051. DOI: 10.1039/C4RA08278H

(140) 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* (9), 5231-5242. DOI: 10.1093/nar/gkx157 (acccessed 1/8/2024).

(141) Tian, G.; Katchur, S. R.; Jiang, Y.; Briand, J.; Schaber, M.; Kreatsoulas, C.; Schwartz, B.; Thrall, S.; Davis, A. M.; Duvall, S.; et al. Small molecule-mediated allosteric activation of the base excision repair enzyme 8-oxoguanine DNA glycosylase and its impact on mitochondrial function. *Sci. Rep.* **2022**, *12* (1), 14685. DOI: 10.1038/s41598-022-18878-2

(142) Kuznetsov, N. A.; Koval, V. V.; Zharkov, D. O.; Nevinsky, G. A.; Douglas, K. T.; Fedorova, O. S. Kinetics of substrate recognition and cleavage by human 8-oxoguanine-DNA glycosylase. *Nucleic Acids Res.* **2005**, *33* (12), 3919-3931. DOI: 10.1093/nar/gki694 (acccessed 4/16/2025).

(143) Yin, Y.; Sasaki, S.; Taniguchi, Y. Recognition and excision properties of 8-halogenated-7-deaza-2′-deoxyguanosine as 8-oxo-2′-deoxyguanosine analogues and Fpg and hOGG1 inhibitors. *ChemBioChem* **2015**, *16* (8), 1190-1198. DOI: https://doi.org/10.1002/cbic.201402690

(144) Norman, D. P. G.; Chung, S. J.; Verdine, G. L. Structural and biochemical exploration of a critical amino acid in human 8-oxoguanine glycosylase. *Biochemistry* **2003**, *42* (6), 1564-1572. DOI: 10.1021/bi026823d

(145) Zheng, L.; Meng, Y.; Campbell, J. L.; Shen, B. Multiple roles of DNA2 nuclease/helicase in DNA metabolism, genome stability and human diseases. *Nucleic Acids Res.* **2019**, *48* (1), 16-35. DOI: 10.1093/nar/gkz1101 (acccessed 11/20/2025).

Chapter 2: Distinctive Formation of a DNA–Protein Crosslink During the Repair of DNA Oxidative Damage: Insights into Human Disease from MD Simulations and QM/MM Calculations

2.1. Introduction

The generation of reactive oxygen species in human cells from endogenous processes (e.g., mitochondrial aerobic metabolism, and inflammatory responses)[1, 2] or exogenous sources (e.g., pollutants, radiation, tobacco smoke, and heavy metals)[3-5] has been linked to a range of health issues, including cancer, diabetes, cardiovascular disease, chronic inflammation, neurodegeneration (Alzheimer's and Parkinson's diseases), asthma, and male infertility.[6-9] These detrimental effects can result from ROS altering the chemical structure and base-pairing properties of the DNA nucleobases.[10, 11] For example, 8-oxo-7,8-dihydroguanine (8oG), a major ROS product that has been estimated to arise 1000–7000 times per cell per day,[12] frequently mispairs with adenine during replication,[13-15] which ultimately causes G:C to T:A transverse mutations. To prevent these harmful mutations, the human adenine DNA glycosylase homologue (MUTYH) aids the reversal of 8oG:A pairs by catalyzing deglycosylation of 2′-deoxyadenosine (dA) as part of the base excision repair (BER) pathway.[16] MUTYH is highly specific towards both 8oG and the opposing adenine, which prevents damage that could be caused by removing adenine from native T:A base pairs.[17, 18]

There is an urgent need to understand the mechanism of action of DNA repair enzymes such as MUTYH due to direct correlations between defects in their function and human diseases. MUTYH inherited biallelic variations can result in a condition known as MUTYH-Associated Polyposis (MAP), which has symptoms including the development of colon polyps and an increased risk of colorectal cancer.[19-21] Indeed, an N224S MUTYH variant associated with MAP alters a residue close to the active site,[22] which could cause the reduced catalytic rates that have been linked with the disease.[23] However, this is just one example among over eighty MUTYH missense and truncating variants whose functions are unclear.[24, 25] There is also evidence that designing novel pharmaceuticals to target enzymes involved in DNA repair pathways may afford successful

cancer treatments by increasing the efficacy of anticancer drugs targeting DNA-repair defective cancers,[26-33] and highly specific and potent small molecule inhibitors have already been designed for enzymes that share mechanistic features with DNA glycosylases like MUTYH.[34-40] However, to aid the prediction of the dysfunction of MAP-related MUTYH variants in humans and develop possible cancer treatment strategies, the atomic-level details of the chemistry facilitated by MUTYH are required. Indeed, structural information has proven essential for the rational development of small molecule (transition state analogue) inhibitors.[41-44]

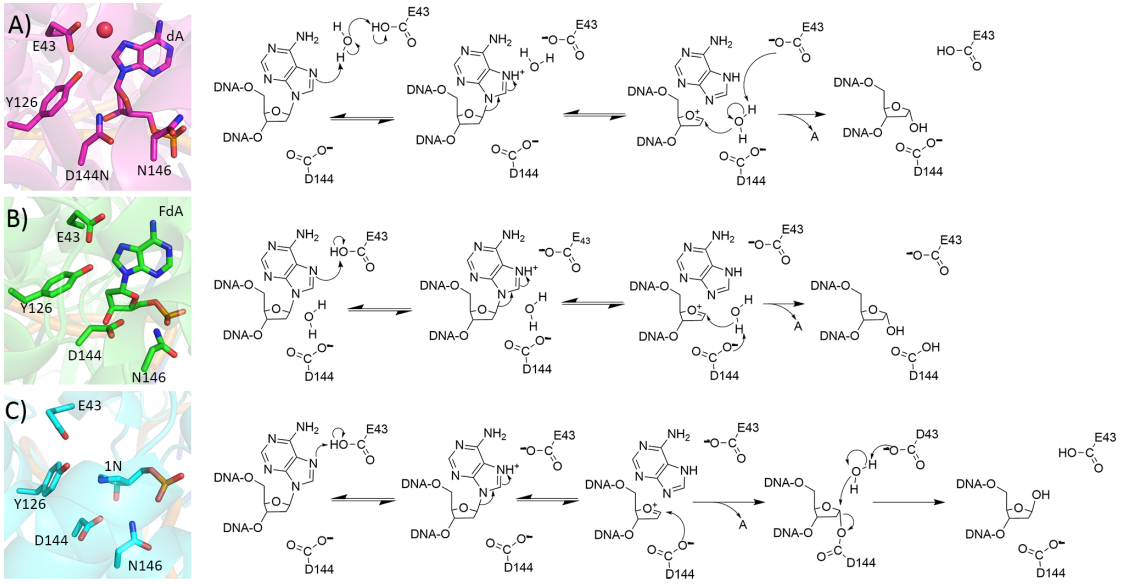


Figure 2.1. Crystal structure (left) and corresponding proposed mechanism (right) of the A) lesion recognition complex (LRC) containing the D144N mutant (PDB ID: 1RRQ), B) fluorine recognition complex (FLRC) containing MutY bound to 2′-β-fluoro-2′-deoxyadenosine (FdA, PDB ID: 3G0Q), and C) transition state analogue complex (TSAC) with MutY bound to a pyrrolidine transition state analogue (1N, PDB ID: 6U7T).

Owing to the same identities and conformations of active site catalytic residues, MUTYH is expected to share a catalytic mechanism with the *Escherichia coli* adenine DNA glycosylase homologue (MutY).[45, 46] Although the function of MutY has been subjected to many experimental studies,[47-57] the deglycosylation mechanism is highly debated in the literature. Mutational data, kinetic isotope effect (KIE) analysis, and a crystal structure of 8oG:A bound to the D144N mutant of *Geobacillus stearothermophilus* (*Gs*) MutY (known as the lesion recognition complex (LRC))[54]

led to the first adopted mechanism (Figure 2.1A).[53] In this proposed pathway, E43 protonates N7 of dA via a bridging water molecule to generate an oxocarbenium ion intermediate and subsequently activates a water nucleophile that attacks C1′ of the dA sugar. Although D144 provides electrostatic charge stabilization of the reaction intermediate in this mechanism, the D144N MutY mutant is catalytically inactive *in vitro*,[51] suggesting a more prominent role for D144. Furthermore, a crystal structure of *Gs* MutY bound to the cleavage resistant substrate analogue 2′-deoxy-2′-fluoro-β-D-arabinoadenosine (FdA; denoted the fluorine lesion recognition complex (FLRC))[55] shows D144 positioned with respect to the substrate to act as the general base. Therefore, in a revised mechanism based on the FLRC, D144 was proposed to activate the water nucleophile following nucleobase departure rather than E43 (Figure 2.1B).[55] This conjectured role for D144 is consistent with that of a conserved aspartate or glutamate residue in the active site of other (monofunctional) DNA glycosylases (e.g., AAG, MBD4, and the UDG family).[58-61]

As for other DNA repair enzymes,[62-71] computational work has provided insights into the mechanism of action of MutY, but has also raised new questions.[53, 58, 72, 73] Indeed, ONIOM(QM:QM) calculations initiated from the FLRC MutY crystal structure provided additional support for the critical and distinct roles of E43 and D144.[73] Furthermore, this work was the first to identify a catalytic role for Y126, namely charge stabilization and positioning of E43, and proposed that D144 explicitly catalyzes nucleobase dissociation by undergoing partial nucleophilic attack of the sugar.[73] A subsequent experimental study substantiated the importance of Y126, at least in part through the crystallization of MutY complexed with the positively charged (3R,4R)-4-(hydroxymethyl)pyrrolidin-3-ol (1N) nucleotide (denoted the transition state analogue complex (TSAC), Figure 2.1C),[56] which revealed a hydrogen-bond network between Y126, E43, and the substrate. Additional support came in the form of a 260-fold decrease in the wild-type MutY catalytic activity for the Y126F mutant.[56] The TSAC structure also indicates that adenine must exit the MutY active site prior to nucleophilic attack of a water molecule,[56] which is consistent with the rapid release of adenine compared to the slower release of the abasic product.[74, 75]

N146 was also implicated as a modulator of catalytic activity via hydrogen bonding with D144,[56] which was recently verified by a measured ~180-fold decrease in adenine excision upon N146S mutation.[57] However, NMR experiments on the MutY methanolysis product[56] and a crystal structure of MutY bound to the enzyme-generated product[57] revealed a β-anomer abasic site with retention of stereochemistry at C1′, which contrasts the inversion of stereochemistry associated with the previously proposed[54, 55] and computationally studied[53, 73] catalytic pathways, as well as the accepted mechanisms for any other monofunctional DNA glycosylase. This structural and kinetic data necessitated a revised mechanism in which D144 directly attacks the oxocarbenium ion in an irreversible reaction step to yield a DNA–protein crosslink (Figure 2.1C). Crosslink formation is consistent with the observed short distance (2.9 Å) between the carboxylate group of D144 and N1′ (the azaribose atom equivalent to C1′ of deoxyribose) and retention of stereochemistry at the anomeric carbon, and would permit active site reorganization, including nucleobase release and alignment of an E43 activated water molecule to hydrolyze the crosslink and regenerate the enzyme in the final chemical step. Although compatible with crystallographic,[54-57] mutational,[47-49, 51] and KIE data,[53] and parallels to mechanisms established for glycosidases,[76, 77] a structural depiction of the DNA–protein crosslink and atomic level details of the crosslink formation mechanism to support the unique double-displacement mechanism for a DNA glycosylase is currently lacking.

To consolidate previous experimental and computational literature on MutY, the present study employs both molecular dynamics (MD) simulations and QM/MM (ONIOM) calculations initiated from DNA–protein complexes that represent different stages of the repair pathway to unequivocally map the MutY catalytic mechanism. Specifically, MD simulations on the reactant complex (RC, modeled from the FLRC) provide initial information about the roles of E43, Y126, N146, and D144 in the first catalytic steps, while MD simulations on the DNA–protein crosslink intermediate complex (IC_{CL}, modeled from the TSAC) afford the first structural information about the crosslinked intermediate and its accommodation in the MutY active site. Subsequently, MD structural data are used to build multiple QM/MM models to provide a dynamic representation of each stage of the repair process and uncover the currently missing atomic level details of the catalytic mechanism, including structures of transition states and computed barrier heights. To gain further insight into

the MutY catalytic strategy and understand the roles of key active site residues, *in silico* mutational studies are used to investigate the structure and function of the Y126F and N146S MutY mutants. Through comparisons to the wealth of available experimental data on MutY,[47-57] this multi-pronged computational approach unveils the structure of key intermediates and provides much-needed support for a unique N-glycosidic bond cleaving mechanism that involves retention of stereochemistry and the formation of a DNA–protein crosslink during DNA repair. In addition to unifying previous experimental[47-57] and computational[53, 58, 72, 73] studies, the atomic level details of the MutY mechanism of action represent an important step in the development of transition state analogue inhibitors as anti-cancer drugs and shed light on the role of a MUTYH mutation associated with MAP. Therefore, this work represents an important step for developing treatment strategies for devastating human diseases.

2.2. Computational Methods

Full details of the computational procedure can be found in Appendix A.

2.2.1. MD simulations

Input models for MD simulations corresponding to the RC and IC_{CL} were generated based on crystal structures of the fluorine recognition complex (FLRC, PDB ID: 3G0Q)[55] and transition state analogue complex (TSAC, PDB ID: 6U7T),[56] respectively. The P164C mutation in the FLRC crystal structure was reverted to proline and 2′-β-fluoro-2′-deoxyadenosine was converted to dA using PyMOL.[78] E43 in the RC was assigned a neutral protonation state based on the relative arrangement of contact partners in the FLRC and LRC crystal structures,[54, 56] and previous PROPKA calculations.[73] In the TSAC model, the proposed DNA–protein crosslink was constructed by replacing azaribose with 2′-deoxyribose and forming a covalent bond between C1′ and the D144 side chain. PyMOL was used to add unresolved residues to the models while avoiding close contacts and to subsequently generate the N146S and Y126F MutY mutants. Each structure was

solvated with TIP3P water molecules in a periodic rectangular prism with sides a minimum of 10 Å from the solute in each direction using the LEaP program in AMBER 2018.[79] Sodium ions were added to neutralize the system and NaCl was added to achieve physiological salt concentration (150 mM). The Amber force field was used throughout, with protein residues described using ff14SB parameters[80] and nucleic acid components modeled using OL15.[81] Parameters for the 8oG lesion were adopted from the literature,[73] while parameters for the sulfur-iron (Fe_4S_4) cluster and the coordinating cysteine residues were generated using the MCPB.py program (Tables A.1, A.2 and Figure A.1, Appendix A).[82] The Amber parameters for the crosslink were supplemented by the general AMBER force field (GAFF)[83] using Antechamber (Table A.1 and Figure A.1, Appendix A)[84] and RESP charge fitting was used to generate charges with the R.E.D.v.III.4 program (Table A.2, Appendix A).[85] Following minimization and equilibration of each model, four 1 μs MD production simulations were performed on each system using AMBER 2018.[79]

Trajectory analysis was conducted across all replicas using the cpptraj program in AMBER 2018.[79] Hydrogen bonds were considered to be present between two residues if the heavy atom distance was less than 3.0 Å and the hydrogen-bonding angle fell between 135° and 180°. D144 was considered in a catalytically conducive position for nucleophilic attack at C1′ of dA in the reactant complex when the $D144_{O\delta}\cdots C1'$ distance was less than 4.0 Å and the $\angle(N9C1'D144_{O\delta})$ angle was greater than 95˚. A water nucleophile was deemed available for hydrolysis of the crosslinked intermediate when located within 3.5 Å of E43 and 4.5 Å of C1′ of deoxyribose. MD data are reported as averages, with the standard deviations calculated across the dataset. Average root-mean-square deviation (RMSD) across each simulation and the data set for each model can be found in the SI (Table A.3, Appendix A). Representative structures were obtained by clustering the entire dataset for each model using the hieragglo algorithm.

2.2.2. QM/MM calculations

Due to the high similarity of the MD replica trajectories, QM/MM models of the DNA–protein complex were built from 10 snapshots, collected at intervals of 4 ns from the beginning of one

trajectory corresponding to each of the wild-type MutY reactant complex (FLRC model) and the crosslinked intermediate (TSAC model). Each model was analyzed to determine whether key residues adopt structural criteria necessary to afford a catalytically conducive complex. Specifically, a direct or water-mediated hydrogen bond between the E43 side chain and N7 of dA is required in the RC, while the distance between E43 and C1′ is less than 5.5 Å in the IC_{CL}. Models that did not satisfy these criteria were discarded. In total, 4 MD snapshots were identified for QM/MM reaction pathway mapping for each catalytic stage. In addition, one QM/MM model was built from the solvated FLRC crystal structure. Y126F MutY mutant models were also constructed by mutating Y126 in the wild-type model initiated from the FLRC crystal structure and one of the TSAC snapshots.

The QM region of the reactant complex (FLRC model) contains 134–137 atoms, with an overall charge of –4. Specifically, the high-level layer contains E43, Y126, D144, N146, V147, dA, R31, E188, E192, and a water molecule hydrogen bonded to E188 (Figure A.2A, Appendix A). Three models also contain a water molecule that bridges E43 and dA in the QM layer. The QM layer for the crosslinked intermediate (TSAC model) contains 111–114 atoms, with an overall charge of –3. Specifically, the high-level layer in the intermediate complex contains E43, Y126, D144, N146, V147, the crosslinked substrate, and S42 (Figure A.2B, Appendix A). The nucleophilic water and 3–4 additional water molecules surrounding E43 were also included in the QM region, with the number of water molecules dependent on the requirements to maintain the E43 conformation with respect to the crosslink as characterized by MD simulations. Although the QM regions for both the FLRC and TSAC models are highly anionic, the large QM region size results in a charge:atom ratio similar to other biosystems that DFT methods have been proven to accurately describe.[86, 87] The MM region for all models includes the remainder of the enzyme and DNA substrate as well as 1300 water molecules, which resulted in an approximately 4 Å deep layer of water.

All QM/MM models were optimized using ONIOM(M06-2X/6-31G(d,p):AMBERff14SB). Transition state structures were isolated by scanning relevant bond distances along the reaction pathway to yield an initial guess and subsequently performing unrestrained transition state

optimizations. Frequency calculations were conducted to characterize the nature of the stationary points. Each pathway was verified to be continuous by mapping the internal reaction coordinate through scanning the nucleophile–electrophile and leaving group distance both to and from each stationary point. The reported reaction Gibbs energies were obtained from ONIOM(M06-2X/6-311+G(2df,p):AMBERff14SB) single-point calculations. The mechanical embedding (ME) scheme was used throughout due to the robustness of this approach and previous successes using the same methodology for various enzymatic reactions.[88-92] The reported ME barrier for the rate-limiting step (enzyme regeneration) was verified to be in close agreement (within 1.5 kJ/mol) with that obtained using the electronic embedding (EE) scheme (see SI for discussion). QM/MM structural and energetic data are reported as averages with standard deviations when models yield similar mechanisms (data for individual models are available in the SI).

All QM/MM calculations were performed using Gaussian 16 (Rev. B.01).[93]

2.3. Results and Discussion

2.3.1. MutY-catalyzed dA deglycosylation coupled with formation of a DNA–protein crosslinked intermediate is energetically feasible, with direct or indirect protonation of the departing nucleobase

For MutY to catalyze dA deglycosylation through DNA–protein crosslink formation, D144 must be positioned to attack C1′ and the departing dA must be stabilized by active site residues in the DNA–MutY reactant complex. During MD simulations of the RC (modeled from the FLRC crystal structure), $D144_{O\delta}$ on average falls within 3.7 ± 0.6 Å of dA (C1′) and the average reaction angle ($\angle$(N9C1′$D144_{O\delta}$)) is 104 ± 19° (Figure 2.2A). As a result, D144 is well aligned for nucleophilic attack to form the DNA–protein crosslink (R($D144_{O\delta}\cdots$C1′) < 4 Å and $\angle$(N9C1′$D144_{O\delta}$) > 95°) for approximately two-thirds of the total simulation time (Table A.4, Appendix A). A persistent hydrogen bond occurs between D144 and N146 throughout the simulations (occupancy of 89%, Figure 2.2A), which is consistent with crystallographic data.[55, 56] This supports the previously hypothesized role

for N146 in positioning D144 based on crystallographic data and the pH dependent activity of the N146S MutY mutant.[55-57] A hydrogen bond also rarely forms between N146 and the DNA backbone (10% occupancy, Table A.4, Appendix A). Y126 and E43 form a strong hydrogen bond (77% occupancy, Figure 2.2A), which correlates with proposals from previous computational[73] and crystallographic[56, 57] work that Y126 is vital for positioning E43 and is possibly involved in active site charge stabilization. While E43 can form a direct hydrogen bond with N7 of dA (Figure 2.2B), this interaction is only present for a small portion of the total simulation time (2% occupancy, Table A.4, Appendix A). Instead, a water molecule frequently bridges E43 and A (34% occupancy, Figure 2.2A), an observation consistent with the E43–dA bridging water resolved in the LRC crystal structure.[54] Additionally, there is a resolved water near E43 in the FLRC that is available to form a

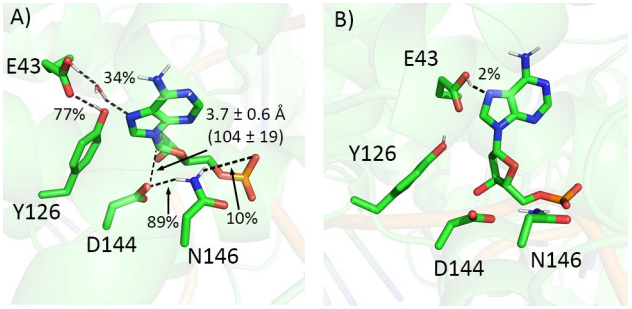


Figure 2.2. MD representative structures of the MutY RC containing A) water bridging E43 and dA, and B) a direct E43···dA(N7) hydrogen bond. Bond distances (Å), attack angle (in parentheses, °), and hydrogen-bond and bridging water occupancies (%) are averaged over the entire simulation time.

bridge when structural dynamics are taken into account.[56] Mechanisms involving direct or water-mediated proton transfer from E43 to the departing A base have been proposed in the literature (Figure 2.1A–B).[53, 55, 57, 72, 73] Taken together, MD simulations of the reactant complex highlight that D144 is properly positioned for nucleophilic attack of dA at C1′, while E43 is available to stabilize the adenine leaving group through either direct or water-mediated proton transfer.

To characterize the mechanism for formation of a DNA–protein crosslink, MD snapshots were extracted for QM/MM calculations and an additional model was built from the FLRC crystal structure. In all five QM/MM characterized reactants (Figure A.3, Appendix A), the D144$_{O\delta}$···C1′

distance (average = 3.125 ± 0.158 Å) and nucleophilic attack angle ($\angle(N9C1'D144_{O\delta})_{avg}$ = 115.9 ± 21.0˚) are optimal for the first reaction step and consistent with the FLRC crystal structure ($R(D144_{O\delta}\cdots C1')$ = 3.2 Å, $\angle(N9C1'D144_{O\delta})$ = 119°). The D144–N146 hydrogen bond is present in all but one model (average distance = 2.786 ± 0.100 Å, average angle = 154 ± 11°), with N146 being close, yet directed away from D144 in the final RC (distance = 3.235 Å, angle = 49.0°). Two models contain a direct hydrogen bond between E43 and N7 of dA (average distance = 2.532 ± 0.001 Å, average angle = 169 ± 7°), while a single water molecule connects the two residues in the remaining three RCs. Most importantly, the optimized QM/MM MutY reactant complexes have consistent (D144) nucleophile positioning with respect to the substrate, while E43 is aligned to facilitate the reaction directly or indirectly.

The slight differences in the active site configuration for each QM/MM optimized RC impact the fine details of the characterized deglycosylation mechanism. In terms of nucleobase activation, barrierless proton transfer from E43 to N7 of dA occurs in RCs containing a direct E43–N7 hydrogen bond (Figure A.4, Appendix A). Although dA is not protonated in RCs containing a water bridge between E43 and dA, proton transfer to dA occurs through the mediating water in a single step (Figure A.5, Appendix A). In the corresponding transition state, the proton is almost completely transferred from E43 to generate a hydrodium-like ion ($R(E43_{H\varepsilon}\cdots WAT_O)_{avg}$ = 1.120 ± 0.115 Å), while proton transfer to the N7 of dA is just beginning ($R(WAT_H\cdots N7)_{avg}$ = 1.427 ± 0.161 Å). The associated barrier is less than 10 kJ/mol (Table A.5, Appendix A). Regardless of the dA protonation pathway, the resulting DNA–MutY complexes (RC_{H+}) have very similar structures, with the average C1′–N9 and N7–H bond lengths and the $D144_{O\delta}\cdots C1'$ distance for structures obtained from direct protonation being within a standard deviation of the values for structures that involve water mediated protonation (Table A.6, Appendix A). Protonation of dA is accompanied by a slight (~0.13 Å) decrease in the $D144_{O\delta}\cdots C1'$ distance (Figures 2.3, A.5, and A.6, Appendix A), which correlates with an increase in the electrophilicity of the substrate. Overall, our data suggest that both direct and water-mediated dA substrate protonation by E43 are valid first steps in the overall MutY mechanism. Furthermore, the predicted ease of dA protonation is consistent with the

experimentally reported lack of correlation between the proton affinities of adenine analogues and MutY catalytic rates.[17]

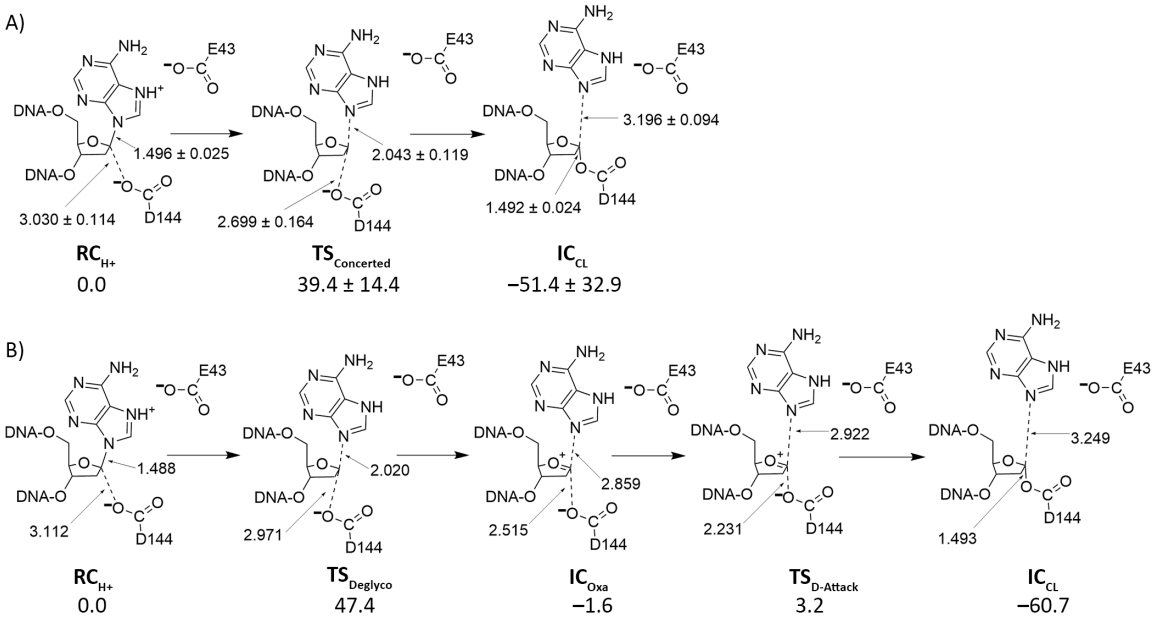


Figure 2.3. QM/MM characterized A) concerted and B) two-step pathways for (protonated) dA deglycosylation through DNA–MutY crosslink formation (IC_{CL}). Bond lengths (Å) and relative Gibbs energies (kJ/mol) are averaged over four models for the concerted pathway.

Following substrate protonation, adenine departure and D144 attack simultaneously occur in four models, with an average energy barrier of 39.4 ± 14.4 kJ/mol (Figures 2.3A and A.6, Table A.7, Appendix A). In the corresponding TSs, D144 is ~0.3 Å closer to C1′ of dA than in the RC_{H+}, while the glycosidic bond stretches by ~0.5 Å, indicating an early transition state. The resulting crosslinked intermediate has a fully formed $D144_{O\delta}$–C1′ bond, while the departed adenine is ~3.2 Å away from the deoxyribose ring ($R(N9\cdots C1')_{avg}$ = 3.196 ± 0.094 Å). In the final model, deglycosylation occurs over two steps, with adenine departure leading to an oxocarbenium ion intermediate (IC_{Oxa}) in the first step and D144 attacking the sugar in the second step (Figure 2.3B). The transition state for formation of the oxocarbenium ion ($TS_{Deglyco}$) is similar to the $TS_{Concerted}$ from the one-step mechanism, with nearly identical C1′–N9 distances (within ~0.02 Å) and D144 positioned slightly further away from C1′ (by ~0.3 Å). Adenine has dissociated from the sugar moiety in the IC_{Oxa} intermediate ($R(N9\cdots C1')$ = 2.859 Å) and D144 falls 2.515 Å from C1′. D144 approaches

C1′ (distance decreases by ~0.3 Å) in the second transition state ($TS_{D\text{-}Attack}$), which results in a crosslinked product that is structurally similar to the IC_{CL} characterized for the one-step mechanism. In the two-step mechanism, the barrier for glycosidic bond cleavage (47.4 kJ/mol) is over 10x larger than the barrier for D144 attack (3.2 kJ/mol).

The average barrier for crosslink formation over all mechanisms is 41.3 ± 13.1 kJ/mol (Table A.7, Appendix A). The predicted energetic cost compares favorably with the barrier for adenine departure reported in other computational studies (30 kJ/mol),[72, 73] and is well below the reported MutY experimental barriers for both adenine release (~80 kJ/mol)[74] and product release (~100 kJ/mol).[74, 75] Additionally, the barrier associated with a D144 nucleophile is much lower than the computationally predicted barrier for attack by a water nucleophile (110.0 kJ/mol),[73] indicating DNA–protein crosslink formation is favored over direct dA hydrolysis. Interestingly, although direct attack of dA at C1′ by D144 was not specifically considered in previous computational studies,[53, 72, 73, 94] partial nucleophilic attack of the substrate by D144 was proposed based on a noticeably short (< 2.5 Å) $D144_{O\delta}\cdots C1'$ distance in a truncated ONIOM(QM:QM) model.[73] Despite supporting the currently characterized mechanism, DNA–protein crosslink formation was not possible in the previously used truncated model, which was constrained outside the active site core to the FLRC crystallographic coordinates throughout the reaction. Overall, our MD simulations reveal that D144 is well positioned for nucleophilic attack at dA, with structural features comparable to those found in the FLRC crystal structure, while ONIOM calculations show crosslink formation is energetically feasible and in fact preferable to direct dA hydrolysis.

2.3.2. The proposed DNA–protein crosslink is well positioned in the MutY active site for hydrolysis by an E43 activated water nucleophile

The crystal structure for the TSAC provides strong evidence for adenine departure following DNA–protein crosslink formation,[56, 57] which permits water to diffuse into the MutY active site. Therefore, the hydrolysis of the crosslink intermediate was modelled from the TSAC crystal structure by initially incorporating the deoxyribose–D144 crosslink and performing MD simulations.

The puckering of the crosslinked sugar is dynamic in the MutY active site. The average Altona–Sundarlingam psuedorotation[95] is 170 ± 31° (Table A.8, Appendix A), which represents a rapid exchange between the C2′-endo and C3′-exo conformations, a shift from the C1′-exo pucker for 1N in the TSAC crystal structure.[56] Water frequently enters the enzyme active site and is positioned between E43 and the crosslink for 43% of the simulation (Figure 2.4A). However, in addition to the local presence of water, crosslink hydrolysis requires E43 and C1′ to be in close proximity. Over the course of the simulations, E43 falls within 5.5–6.5 Å of C1′ to promote water nucleophile activation for over one-third (37%) of the total simulation time (average distance = 6.0 ± 0.3 Å, Figures A.7 and A.8A, Appendix A). In the remainder of the simulation, the $E43_{O\varepsilon}\cdots C1'$ distance is much longer (average distance = 7.4 ± 0.4 Å, Figure A.8B, Appendix A). Finally, in addition to being in close proximity to C1′ and E43, the water nucleophile must be on the 5′ face of the sugar and aligned for attack at the C–O crosslink bond, which can be described by $\angle(E43_{O\varepsilon}C1'D144_{O\delta}) > 125°$. MD simulations reveal three distinct E43–crosslink relative orientations that primarily differ in the $\angle(C2'C1'D144_{O\delta}D144_{C\gamma})$ dihedral angle about the C–O crosslink bond (Figures A.9 and A.10, Appendix A). $\angle(E43_{O\varepsilon}C1'D144_{O\delta})_{avg}$ = 133 ± 11° and the water nucleophile is appropriately positioned opposite D144 in the first (active) conformation ($\angle(C2'C1'\ D144_{O\delta}D144_{C\gamma})_{avg}$ = 140–220°), which occurs for over 81% of the simulation time. In the other two (inactive) conformations ($\angle(C2'C1'D144_{O\delta}D144_{C\gamma})_{avg}$ = 40–120° (8% occupancy) or 280–350° (7% occupancy)), the water nucleophile is not suitably positioned to initiate the reaction ($\angle(E43_{O\varepsilon}C1'D144_{O\delta})$ = 112 ± 16°). When coupled, the proximity and alignment of E43 to the crosslink results in the active site being conducive for catalysis occurs for 27% of the simulation (Figure 2.4A, Table A.8, Appendix A). This active site conformation, particularly the positions of D144, E43, Y126, and N146 relative to the substrate, is structurally similar to the TSAC crystal structure (Figure 2.4B), which supports the structural accuracy of the computationally-predicted crosslink intermediate.

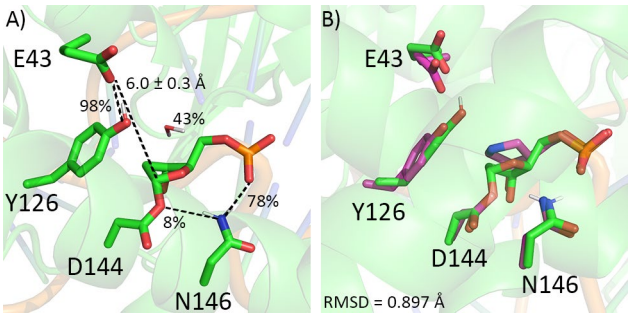


Figure 2.4. A) MD representative structure of IC_{CL} bound in the MutY active site and aligned for hydrolysis. B) Overlay of the MutY active site from the MD representative structure of the IC_{CL} (green) and the TSAC crystal structure (magenta).

N146 has a much more transient interaction with D144 in the crosslinked intermediate (hydrogen bond occupancy = 8%; Figure 2.4A) compared to the RC (89%, Figure 2.2). Although N146 remains within 4 Å of D144 for 37% of the simulation, the residue forms a much stronger hydrogen bond with the DNA backbone in the IC_{CL} (78% occupancy, Table A.8, Appendix A) than RC (10%, Table A.4, Appendix A). This suggests that N146 may not play as significant a role in crosslink hydrolysis, which is consistent with the N146S mutation having little impact on MutY catalysis after adenine release.[57] On the other hand, Y126 maintains a strong hydrogen bond with E43 (98% occupancy, average distance = 2.6 Å, average angle = 164°), pointing to the continued roles of Y126 in positioning E43 and providing charge stabilization during the second reaction step. Overall, although the MutY active site is highly dynamic upon formation of the crosslinked intermediate and therefore does not consistently adopt a catalytically active conformation, the crosslink, E43, water nucleophile, and Y126 are frequently aligned in a manner to promote crosslink hydrolysis and thereby allow product generation.

2.3.3. The barrier for hydrolysis of the DNA–protein crosslink to afford an abasic site product with retention of stereochemistry is consistent with the experimental reaction rate

The QM/MM optimized crosslinked intermediates built from different MD snapshots are very similar (Figure A.11, Appendix A). Specifically, the nucleophilic water is hydrogen bonded to E43

($R(E43_{O\varepsilon}\cdots WAT_H)_{avg}$ = 1.955 ± 0.303 Å), and well positioned with respect to the crosslink ($R(WAT_O\cdots C1')_{avg}$ = 3.019 ± 0.067 Å), which is consistent with the nucleophile observed in the TSAC crystal structure.[56] The $R(D144_{O\delta}–N146_{N\delta})_{avg}$ distance is 3.153 ± 0.042 Å, while Y126 forms a hydrogen bond with E43 in three systems and the water nucleophile in one model. Regardless, all intermediates proceed through a similar single TS (Figure 2.5) in which the crosslink bond is significantly extended ($R(D144_{O\delta}\cdots C1')_{avg}$ = 2.519 ± 0.029 Å), the nucleophile is late in its attack ($R(WAT_O\cdots C1')_{avg}$ = 1.924 ± 0.051 Å), and the proton transfer to E43 is in the initial stages ($R(E43_{O\varepsilon}\cdots WAT_H)_{avg}$ = 1.741 ± 0.426 Å). In the product complex (PC), D144 has dissociated even further from C1′ ($R(D144_{O\delta}\cdots C1')_{avg}$ = 2.798 ± 0.075), resulting in a β-anomer abasic site. The final product is consistent with the NMR analysis of the MutY methanolysis product[56] and the crystal structure of the enzyme-catalyzed product.[57]

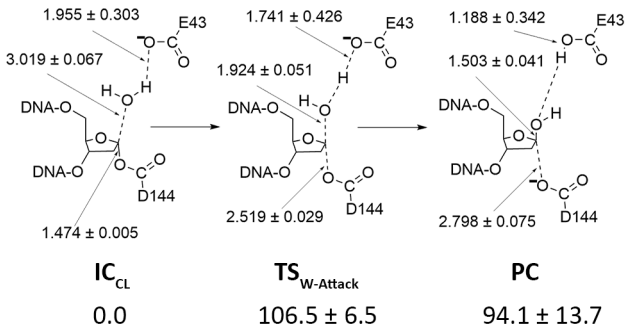


Figure 2.5. QM/MM characterized pathway for DNA–MutY crosslink hydrolysis (enzyme regeneration). Bond distances (Å) and relative Gibbs energies (kJ/mol) are averaged over four models.

The average calculated barrier for crosslink hydrolysis is 106.5 ± 6.5 kJ/mol (Figure 2.5, Table A.9, Appendix A). Although exponential averaging is sometimes used in the literature to determine QM/MM energy barriers,[96] the similarity of our computed barriers for different starting structures results in an exponential average (102.7 kJ/mol) that is within 4 kJ/mol of that obtained through arithmetic averaging. Comparison to the predicted barrier for crosslink formation (41.3 ± 13.1 kJ/mol) indicates that crosslink hydrolysis is the overall rate-determining MutY chemical step. Interestingly, the overall calculated barrier for MutY-catalyzed dA deglycosylation through crosslink

formation (Figure 2.1C) is very similar to that previously characterized for direct dA hydrolysis (110.0 kJ/mol;[73] Figure 2.1B). Furthermore, both processes initially lead to an endothermic product prior to anticipated active site relaxation and product release. Nevertheless, the low initial barrier associated with crosslink formation prevents direct hydrolysis, the accepted mechanism for all other monofunctional DNA glycosylases to date, from being the preferred catalytic pathway. Most importantly, the predicted double-displacement MutY mechanism that involves crosslink formation is fully consistent with an abundance of kinetic, mutagenic, NMR, and crystallographic data,[47-57] including proposed essential roles for D144, E43, Y126, and N146,[51, 54-57, 72, 73] the estimated experimental barrier for the rate-limiting product release (100 kJ/mol),[75] and the retention of stereochemistry in the β-anomer abasic site product.[56, 57] Thus, in addition to unifying previous literature on the MutY mechanism of action, our work provides the first computational support for the previously proposed DNA–protein crosslink in the MutY active site based on a crystalized transition state analogue and NMR data,[56] which showcases an N-glycosidic bond cleaving mechanism that involves retention of stereochemistry. This functionality sets MutY apart from all other monofunctional glycosylases examined to date, which otherwise use direct hydrolysis of N-glycosidic bonds without crosslink formation to yield abasic site products with inversion of stereochemistry.

2.3.4. Y126F MutY mutant hinders the ability of E43 to catalyze both crosslink formation and enzyme regeneration

To provide additional support for the predicted MutY mechanism, as well as more insight into the role of Y126 in the catalytic pathway, the function of the Y126F MutY mutant is considered. Experimental studies have shown that the Y126F mutation causes a 260-fold reduction in MutY

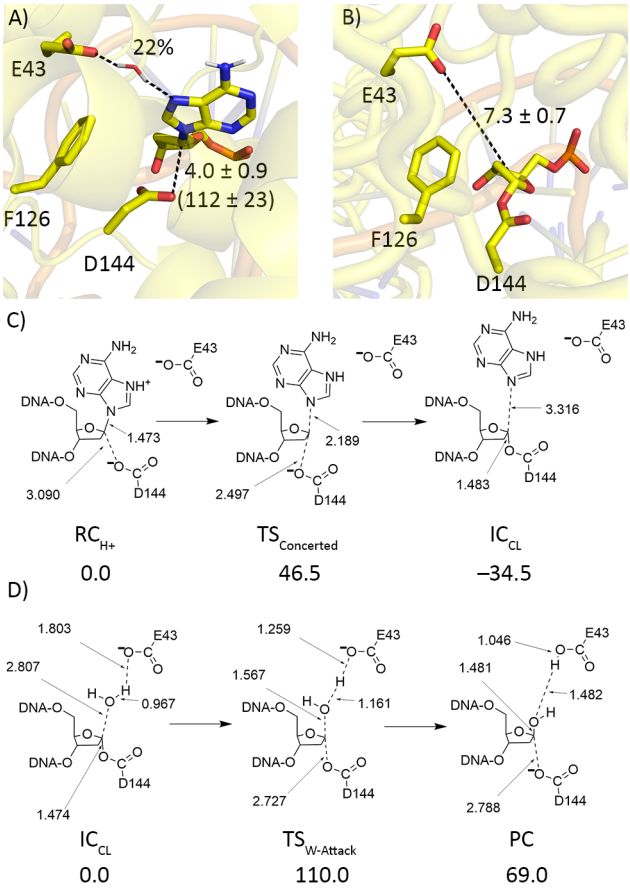


Figure 2.6. A) MD representative structure of the Y126F MutY RC. B) MD representative structure of the Y126F MutY IC$_{CL}$. C) Key ONIOM(M06-2X/6-311+G(2df,p):AMBERff14SB)//ONIOM(M06-2X/6-31G(d,p):AMBERff14SB) bond lengths (Å) and relative Gibbs energies (kJ/mol) for the deglycosylation (crosslink formation) and D) crosslink hydrolysis (enzyme regeneration) pathways for Y126F MutY. Calculations were initiated from the FLRC crystal structure and the first IC$_{CL}$ MD snapshot.

catalytic activity.[56] MD simulations on the Y126F reactant complex (FLRC model) reveal that D144 is similarly positioned with respect to the substrate in mutant and wild-type MutY (Figures 2.6A and A.12A, Appendix A). Specifically, the D144$_{O\delta}$···C1′ distance (4.0 ± 0.9 Å) is slightly longer for the mutant than wild-type MutY (3.7 ± 0.6 Å, Figure 2.2), while ∠(N9C1′D144$_{O\delta}$) is slightly larger (112 ± 23°, mutant; 104 ± 19°, wild-type). As a result, D144 is aligned to attack C1′ of dA for 57% of the simulation on the Y126F mutant (Table A.4, Appendix A), which is comparable to the alignment in the wild-type enzyme (66%). The E43···N7 hydrogen bond has a similar occupancy for Y126F and wild-type MutY (3%), and the bridging water occupancy is slightly reduced in the mutant (from 34%

to 22%). This suggests that the Y126···E43 hydrogen bond plays a minor role in positioning E43 to accommodate a bridging water.

In addition to changes in the RC, MD simulations on the Y126F crosslinked intermediate uncover significant structural destabilization of the active site upon loss of the Y126···E43 interaction (Figure 2.6B and A.12B, Appendix A). Indeed, E43 is only close enough to deoxyribose to activate a water nucleophile for 13% of the simulation for the mutant compared to 37% for wild-type MutY, with the peak corresponding to the active conformation absent from the histogram (Figure A.13, Table A.8, Appendix A). Additionally, E43 is only correctly aligned for nucleophile activation for 19% of the simulation (based on $\angle$(C2′C1′D144$_{O\delta}$D144$_{C\gamma}$)), while the inactive conformation with $\angle$(C2′C1′D144$_{O\delta}$D144$_{C\gamma}$) = 40–120° is adopted for 77% of the simulation time (Figures A.14–15, Table A.8, Appendix A). This combination of factors results in the active site being aligned to facilitate catalysis for only 5%, and water being located between E43 and the crosslink for only 4%, of the simulation time (Figures 2.6B). Together, the MD data suggest that Y126 helps position E43 to facilitate both crosslink formation and crosslink hydrolysis (enzyme regeneration) following adenine departure.

To investigate the energetic impact of the Y126F mutation on the catalytic mechanism, a single pathway was mapped using QM/MM (ONIOM) starting from the FLRC crystal structure for crosslink formation and MD snapshot 1 for enzyme regeneration (Figures 2.6C–D). While the details of the Y126F mutant reaction pathway generally resemble the wild-type MutY mechanism, the Y126F mutant was found to have a later transition state for the rate-limiting crosslink hydrolysis step, with R(D144$_{O\delta}$···C1′) increased by ~0.2 Å and R(WAT$_{O}$···C1′) decreased by ~0.4 Å relative to wild-type MutY. The overall calculated barrier also exhibits a small increase (by 6.8 kJ/mol) for the Y126F mutant compared to wild-type MutY. While we acknowledge our predictions are limited by mapping only a single pathway for mutant MutY, the role of Y126 is consistent across the wild-type models and we anticipate the predicted increase to be representative of the other pathways. Furthermore, the reduced occupancy of a catalytically conducive active site conformation coupled with an increased rate-determining barrier correlates with the reported ~260-fold reduction in MutY catalytic

activity upon Y126F mutation.[56] In addition to highlighting the role of Y126 in positioning and charge stabilizing E43 throughout the reaction, the reproduced Y126F mutational impact further supports the double-displacement mechanisms for MutY, which was originally proposed based on elegant crystallographic, kinetic, spectroscopic, and mutational data.[56, 57] This cross-linking mechanism distinguishes the MutY repair enzyme from others in the monofunctional glycosylase family and provides the first example of retention of stereochemistry upon the enzymatically catalyzed cleavage of an N-glycosidic bond.

2.3.5. Although distinct for monofunctional glycosylases, the MutY mechanism parallels crosslink formation utilized by other enzymes to cleave glycosidic bonds

While most glycosylases have a widely conserved catalytic aspartate/glutamate residue in the active site that is similarly positioned as D144 in MutY (i.e. uracil DNA glycosylases, AAG, and MBD4),[60, 97-99] the residue is widely accepted to activate a water nucleophile for direct hydrolysis of the glycosidic bond in the absence of DNA–protein crosslink formation. Indeed, the MBD4 mechanism of action was initially thought to involve crosslink formation due to close contact between C1′ and the catalytic aspartate (D534) in a crystal structure of the enzyme bound to an abasic site,[100] but this proposal was refuted by recent crystallographic and mutational data.[60] Although a crystal structure of *Mycobacterium smegmatis* uracil DNA glycosylase X (MsmUdgX) complexed with single-stranded DNA contains a DNA–protein crosslink between H109 and C1′ of dU, this crosslink is the final product as part of a suicide inactivation mechanism rather than a transient intermediate.[101, 102] Nevertheless, a crystal structure of 3–methyladenine DNA glycosylase II (AlkA) bound to azaribose contains a similar short distance between D238 and N1′ as found in the MutY TSAC structure (3.2 Å, Figure 2.7A and B),[94] suggesting potential crosslink formation, a proposal yet to be conclusively verified with experimental or computational data. Despite crosslink formation in the MutY mechanism of action being unique among monofunctional glycosylases to

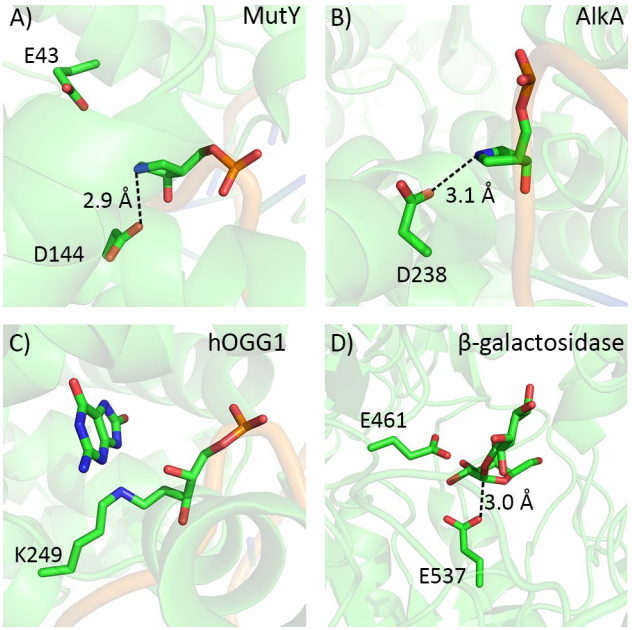


Figure 2.7. Comparison of the active sites from crystal structures of A) the MutY bound to pyrrolidine transition state analogue (PDB ID: 6U7T), B) AlkA bound to azaribose inhibitor (PDB ID: 1DIZ), C) bifunctional hOGG1 bound to a borohydride trapped intermediate (PDB ID: 1HU0), and D) retaining glycosidase β-galactosidase bound to allolactose inhibitor (PDB ID: 4DUW).

date, crosslinks occur during BER involving bifunctional glycosylases, which use an active site amine to facilitate both glycosylase and b/d-lyase activity (Figure 2.7C).[103, 104] Regardless of the distinction between MutY and other repair enzymes, the discovery of a retaining mechanism for MutY draws parallels between BER monofunctional glycosylases and the O-glycosidases, which can also adopt both retaining and inverting mechanisms. Indeed, retaining glycosidases use an aspartate/glutamate residue that is similarly positioned with respect to the substrate as D144 of MutY to form a sugar–protein crosslink as an intermediate in the cleavage of O-glycosidic bonds (Figure 2.7D).[76, 77] Thus, while the utilization of DNA–protein crosslinks by enzymes is well documented, MutY is the only monofunctional glycosylase confirmed to date through previous experimental studies[56, 57] and the present computational work to invoke such a crosslinked intermediate as part of its mechanism of action and the only enzyme yet identified to retain stereochemistry while cleaving N-glycosidic bonds. The distinctive MutY mechanism compared to

other repair enzymes sets the stage for the design of specific and potent small molecule inhibitors as cancer therapeutics.

2.3.6. MAP-causing N146S mutation hinders DNA–protein crosslink formation

The present work has underscored the vital role played by D144 in the MutY mechanism of action, being directly involved in base excision through the formation of a covalent DNA–protein crosslink intermediate. Our calculations suggest that this nucleophilic role is facilitated by N146, which likely holds D144 in the correct position relative to dA through a persistent hydrogen bond (Figure 2.2 and Table A.4, Appendix A). Interestingly, the N146S mutation, which is analogous to the N224S mutation of MUTYH that is associated with MAP in humans,[22] has an ~180-fold reduction in catalytic activity, with the decrease being mainly attributed to an alteration in the chemical steps rather than substrate affinity.[57] To obtain structural information about this cancer-related variant, investigate the catalytic role of N146, allow predictions of the functional consequences of MAP variants, and explain the underlying causes of MAP, MD simulations were performed on the N146S MutY mutant.

MD simulations indicate that S146 rarely forms a hydrogen bond with D144 (30% occupancy, Table A.4, Appendix A) compared to N146 in wild-type MutY (89% occupancy), while the N146 hydrogen bond with the DNA backbone is completely abolished upon mutation. Consequently, D144 is correctly aligned with respect to the substrate for 31% of the simulation compared to 66% for the wild-type enzyme (Figure 2.8A). Furthermore, the N146S mutation affords a unique active site conformation in which D144 is directed away from C1′ of dA (Figure 2.8B) and is too far to participate in catalysis (7.3 ± 1.6 Å). Beyond the mutational impact on D144, there is also a decrease in the Y126–E43 hydrogen bond occupancy (from 77% to 54%) and the E43–dA bridging water occupancy (34% to 22%), which points to the broader impact of the N146S mutation on the active site conformation.

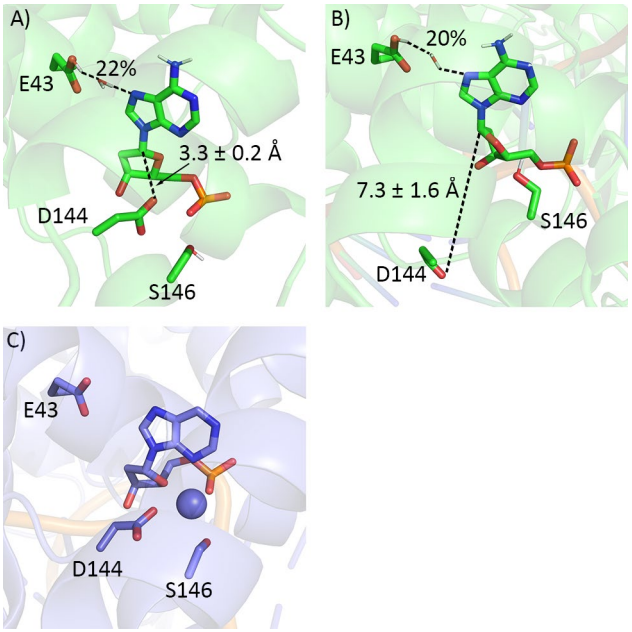


Figure 2.8. MD representative structures of the N146S MutY RC in the A) inactive and B) active conformations. C) The crystal structure of the Ca^{2+}-inhibited N146S MutY bound to purine (PDB ID: 8DVP).

Our simulations correlate with the reduced, but not halted, adenine excision activity of the N146S MutY mutant.[57] Indeed, the predicted active conformation closely resembles a crystal structure of the Ca^{2+}-inhibited N146S MutY mutant bound to purine (Figure 2.8C),[57] where the direct coordination between D144 and Ca^{2+} in the crystal structure likely prevents nucleophilic attack and crosslink formation, resulting in an intact purine glycosidic bond. In contrast, the lack of a Ca^{2+} ion coordinated to D144 in the MD model allows for nucleophilic attack and thus glycosidic bond cleavage. Nevertheless, in the absence of a Ca^{2+} ion, accounting for the structural dynamics results in an inactive conformation being favored (Figure 2.8B). Our findings further underscore the structural significance of N146 for properly positioning D144 in the MutY active site as well as for the overall active site alignment, roles that have been previously proposed based on structural and mutational data.[55-57] Since the bacterial MutY and human homologue MUTYH have highly conserved active sites, with identical positioning of key catalytic residues (D144/222, Y126/204, E43/120, N146/224),[45, 46] the enzymes are expected to share a catalytic mechanism. Therefore, the observed loss of the D144–N146 hydrogen bond and stabilization of a catalytically inactive

conformation upon N146S mutation provides a structural rationalization for the association between the N224S mutation in MUTYH and MAP in humans.[23]

2.4. Conclusion

The present work provides the first computational structural depiction and energetic support for the formation of a DNA–protein crosslinked intermediate during the N-glycosidic bond cleavage by a monofunctional DNA glycosylase to yield a unique abasic site product with retention of stereochemistry, which was previously proposed based on elegant structural, kinetic, spectroscopic, and mutational data.[56, 57] The characterized MutY enzymatic pathway involves protonation of dA at N7 by E43 (either directly or through a bridging water) followed by D144 attack at C1′ to form a DNA–protein crosslink. The low barrier for crosslink formation (< 40 kJ/mol) prevents direct hydrolysis from being the preferred catalytic pathway. Furthermore, MD simulations suggest that the DNA–protein crosslink is well positioned in the MutY active site for subsequent hydrolysis. Indeed, in the second reaction step, Y126 aligns E43 to activate a water nucleophile that hydrolyzes the crosslink and yields a β-anomer abasic site product. Crosslink hydrolysis is the rate-limiting step, with an average calculated barrier of 106.5 ± 6.5 kJ/mol. In addition to highlighting the role of Y126 in positioning and charge stabilizing E43 throughout the reaction, the computationally reproduced Y126F mutational effect further supports the proposed mechanism. Indeed, the predicted double-displacement MutY mechanism is fully consistent with an abundance of experimental data, including the experimentally-estimated barrier for product release (100 kJ/mol),[75] (D144, E43, Y126, and N146) mutational data,[47-49, 51] and retention of stereochemistry in the abasic site product.[56, 57] Although intermediate crosslink formation is used by O-glycosidases to cleave O-glycosidic bonds, as well as other classes of DNA repair enzymes (i.e., bifunctional DNA glycosylases), MutY is the only enzyme known to date that retains stereochemistry while cleaving N-glycosidic bonds.

Beyond resolving a longstanding controversy in the experimental and computational literature surrounding the atomic level details of MutY activity, our calculations highlight that N146 forms a

strong hydrogen bond with D144 and thereby aligns D144 for crosslink formation. Simulations on the N146S mutant further underscore the importance of this residue, revealing a dominant conformation of the MutY active site that is not conducive for catalysis. These structural data confirm the proposed roles of this residue based on the previously reported reduction in activity upon N146 mutation[57] and provides a mechanistic rationalization for why the N224S MUTYH mutant results in MAP. Since cancer cells commonly have compromised DNA repair pathways, DNA repair enzymes are attractive targets for therapeutics as inhibiting enzymes compensating for the reduced repair can creating rising levels of DNA damage and trigger apoptosis or render the cell non-viable.[33, 42] For example, small molecule inhibitors for Flap Endonuclease 1 have been shown to enhance the ability of therapeutics to treat cancer cells.[36, 105, 106] As such, the more detailed understanding of the MutY mechanism of action and the structure of key transition states along the reaction pathway provided by the present work will expedite the future development of specific and potent small molecule inhibitors.

2.5. References

(1) Turrens, J. F. Mitochondrial formation of reactive oxygen species. *J. Physiol.* **2003**, *552* (Pt 2), 335-344. DOI: 10.1113/jphysiol.2003.049478

(2) Mittal, M.; Siddiqui, M. R.; Tran, K.; Reddy, S. P.; Malik, A. B. Reactive oxygen species in inflammation and tissue injury. *Antioxid. Redox Signal.* **2014**, *20* (7), 1126-1167. DOI: 10.1089/ars.2012.5149

(3) Tominaga, H.; Kodama, S.; Matsuda, N.; Suzuki, K.; Watanabe, M. Involvement of reactive oxygen species (ROS) in the induction of genetic instability by radiation. *J. Radiat. Res.* **2004**, *45* (2), 181-188. DOI: 10.1269/jrr.45.181 (acccessed 4/22/2025).

(4) Prasad, S.; Gupta, S. C.; Tyagi, A. K. Reactive oxygen species (ROS) and cancer: Role of antioxidative nutraceuticals. *Cancer Lett.* **2017**, *387*, 95-105. DOI: https://doi.org/10.1016/j.canlet.2016.03.042

(5) Wang, L.; Wise, J. T. F.; Zhang, Z.; Shi, X. Progress and prospects of reactive oxygen species in metal carcinogenesis. *Curr. Pharmacol. Rep.* **2016**, *2* (4), 178-186. DOI: 10.1007/s40495-016-0061-2

(6) Forrester, S. J.; Kikuchi, D. S.; Hernandes, M. S.; Xu, Q.; Griendling, K. K. Reactive oxygen species in metabolic and inflammatory signaling. *Circ. Res.* **2018**, *122* (6), 877-902. DOI: 10.1161/circresaha.117.311401

(7) Chen, Q.; Wang, Q.; Zhu, J.; Xiao, Q.; Zhang, L. Reactive oxygen species: Key regulators in vascular health and diseases. *Br. J. Pharmacol.* **2018**, *175* (8), 1279-1292. DOI: 10.1111/bph.13828

(8) Diehn, M.; Cho, R. W.; Lobo, N. A.; Kalisky, T.; Dorie, M. J.; Kulp, A. N.; Qian, D.; Lam, J. S.; Ailles, L. E.; Wong, M.; et al. Association of reactive oxygen species levels and radioresistance in cancer stem cells. *Nature* **2009**, *458* (7239), 780-783. DOI: 10.1038/nature07733

(9) Helbock, H. J.; Beckman, K. B.; Shigenaga, M. K.; Walter, P. B.; Woodall, A. A.; Yeo, H. C.; Ames, B. N. DNA oxidation matters: The HPLC–electrochemical detection assay of 8-oxo-deoxyguanosine and 8-oxo-guanine. *Proc. Natl. Acad. Sci. U.S.A* **1998**, *95* (1), 288-293. DOI: 10.1073/pnas.95.1.288

(10) Yu, Y.; Cui, Y.; Niedernhofer, L. J.; Wang, Y. Occurrence, biological consequences, and human health relevance of oxidative stress-induced DNA damage. *Chem. Res. Toxicol.* **2016**, *29* (12), 2008-2039. DOI: 10.1021/acs.chemrestox.6b00265

(11) Cadet, J.; Davies, K. J. A.; Medeiros, M. H. G.; Di Mascio, P.; Wagner, J. R. Formation and repair of oxidatively generated damage in cellular DNA. *Free Radical Biol. Med.* **2017**, *107*, 13-34. DOI: https://doi.org/10.1016/j.freeradbiomed.2016.12.049

(12) Collins, A.; Gedik, C.; Vaughan, N.; Wood, S.; White, A.; Dubois, J.; Rees, J.-F.; Loft, S.; Moller, P.; Poulsen, H.; et al. Measurement of DNA oxidation in human cells by chromatographic and enzymic methods. *Free Radic. Biol. Med.* **2003**, *34* (8), 1089-1099. DOI: 10.1016/S0891-5849(03)00041-8

(13) David, S. S.; O'Shea, V. L.; Kundu, S. Base-excision repair of oxidative DNA damage. *Nature* **2007**, *447*, 941. DOI: 10.1038/nature05978

(14) Choi, J.-S.; Dasari, A.; Hu, P.; Benkovic, S. J.; Berdis, A. J. The use of modified and non-natural nucleotides provide unique insights into pro-mutagenic replication catalyzed by polymerase eta. *Nucleic Acids Res.* **2016**, *44* (3), 1022-1035, 10.1093/nar/gkv1509. DOI: 10.1093/nar/gkv1509

(15) Cheng, X.; Kelso, C.; Hornak, V.; de los Santos, C.; Grollman, A. P.; Simmerling, C. Dynamic behavior of DNA base pairs containing 8-oxoguanine. *J. Am. Chem. Soc.* **2005**, *127* (40), 13906-13918. DOI: 10.1021/ja052542s

(16) Michaels, M. L.; Tchou, J.; Grollman, A. P.; Miller, J. H. A repair system for 8-oxo-7,8-dihydrodeoxyguanine. *Biochemistry* **1992**, *31* (45), 10964-10968, 10.1021/bi00160a004. DOI: 10.1021/bi00160a004

(17) Majumdar, C.; McKibbin, P. L.; Krajewski, A. E.; Manlove, A. H.; Lee, J. K.; David, S. S. Unique hydrogen bonding of adenine with the oxidatively damaged base 8-oxoguanine enables specific recognition and repair by DNA glycosylase MutY. *J. Am. Chem. Soc.* **2020**, *142* (48), 20340-20350. DOI: 10.1021/jacs.0c06767

(18) Russelburg, L. P.; O'Shea Murray, V. L.; Demir, M.; Knutsen, K. R.; Sehgal, S. L.; Cao, S.; David, S. S.; Horvath, M. P. Structural basis for finding OG lesions and avoiding undamaged g by the DNA glycosylase MutY. *ACS Chem. Biol.* **2020**, *15* (1), 93-102. DOI: 10.1021/acschembio.9b00639

(19) Mazzei, F.; Viel, A.; Bignami, M. Role of MUTYH in human cancer. *Mutat. Res., Fundam. Mol. Mech. Mutagen.* **2013**, *743-744*, 33-43, 10.1016/j.mrfmmm.2013.03.003. DOI: 10.1016/j.mrfmmm.2013.03.003

(20) Al-Tassan, N.; Chmiel, N. H.; Maynard, J.; Fleming, N.; Livingston, A. L.; Williams, G. T.; Hodges, A. K.; Davies, D. R.; David, S. S.; Sampson, J. R.; et al. Inherited variants of MYH associated with somatic g:C→t:A mutations in colorectal tumors. *Nat. Genet.* **2002**, *30* (2), 227, Article. DOI: 10.1038/ng828

(21) Robinson, P. S.; Thomas, L. E.; Abascal, F.; Jung, H.; Harvey, L. M. R.; West, H. D.; Olafsson, S.; Lee, B. C. H.; Coorens, T. H. H.; Lee-Six, H.; et al. Inherited MUTYH mutations cause elevated somatic mutation rates and distinctive mutational signatures in normal human cells. *Nat. Commun.* **2022**, *13* (1), 3949. DOI: 10.1038/s41467-022-31341-0

(22) Dallosso, A. R.; Dolwani, S.; Jones, N.; Jones, S.; Colley, J.; Maynard, J.; Idziaszczyk, S.; Humphreys, V.; Arnold, J.; Donaldson, A.; et al. Inherited predisposition to colorectal adenomas caused by multiple rare alleles of MUTYH but not OGG1, NUDT1, NTH1 or NEIL 1, 2 or 3. *Gut* **2008**, *57* (9), 1252-1255, 10.1136/gut.2007.145748. DOI: 10.1136/gut.2007.145748

(23) Kundu, S.; Brinkmeyer, M. K.; Livingston, A. L.; David, S. S. Adenine removal activity and bacterial complementation with the human MutY homologue (MUTYH) and Y165C, G382D, P391L and Q324R variants associated with colorectal cancer. *DNA Repair* **2009**, *8* (12), 1400-1410. DOI: https://doi.org/10.1016/j.dnarep.2009.09.009

(24) Cheadle, J. P.; Sampson, J. R. MUTYH-associated polyposis—From defect in base excision repair to clinical genetic testing. *DNA Repair* **2007**, *6* (3), 274-279. DOI: 10.1016/j.dnarep.2006.11.001

(25) Banda, D. M.; Nuñez, N. N.; Burnside, M. A.; Bradshaw, K. M.; David, S. S. Repair of 8-oxoG:A mismatches by the MUTYH glycosylase: Mechanism, metals and medicine. *Free Radic. Biol. Med.* **2017**, *107*, 202-215. DOI: 10.1016/j.freeradbiomed.2017.01.008

(26) Schrempf, A.; Slyskova, J.; Loizou, J. I. Targeting the DNA repair enzyme polymerase θ; in cancer therapy. *Trends Cancer* **2021**, *7* (2), 98-111. DOI: 10.1016/j.trecan.2020.09.007 (acccessed 2022/01/06).

(27) Duskova, K.; Lejault, P.; Benchimol, É.; Guillot, R.; Britton, S.; Granzhan, A.; Monchaud, D. DNA junction ligands trigger DNA damage and are synthetic lethal with DNA repair inhibitors in cancer cells. *J. Am. Chem. Soc.* **2020**, *142* (1), 424-435. DOI: 10.1021/jacs.9b11150

(28) Fuso Nerini, I.; Roca, E.; Mannarino, L.; Grosso, F.; Frapolli, R.; D'Incalci, M. Is DNA repair a potential target for effective therapies against malignant mesothelioma? *Cancer Treat. Rev.* **2020**, *90*, 102101. DOI: https://doi.org/10.1016/j.ctrv.2020.102101

(29) Mechetin, G. V.; Endutkin, A. V.; Diatlova, E. A.; Zharkov, D. O. Inhibitors of DNA glycosylases as prospective drugs. *Int. J. Mol. Sci.* **2020**, *21* (9), 3118.

(30) Jiang, M.; Jia, K.; Wang, L.; Li, W.; Chen, B.; Liu, Y.; Wang, H.; Zhao, S.; He, Y.; Zhou, C. Alterations of DNA damage repair in cancer: From mechanisms to applications. *Ann Transl Med* **2020**, *8* (24), 1685. DOI: 10.21037/atm-20-2920

(31) Gad, H.; Koolmeister, T.; Jemth, A.-S.; Eshtad, S.; Jacques, S. A.; Ström, C. E.; Svensson, L. M.; Schultz, N.; Lundbäck, T.; Einarsdottir, B. O.; et al. MTH1 inhibition eradicates cancer by preventing sanitation of the dNTP pool. *Nature* **2014**, *508* (7495), 215-221. DOI: 10.1038/nature13181

(32) Huber, K. V. M.; Salah, E.; Radic, B.; Gridling, M.; Elkins, J. M.; Stukalov, A.; Jemth, A.-S.; Göktürk, C.; Sanjiv, K.; Strömberg, K.; et al. Stereospecific targeting of MTH1 by (s)-crizotinib as an anticancer strategy. *Nature* **2014**, *508* (7495), 222-227. DOI: 10.1038/nature13194

(33) Helleday, T.; Petermann, E.; Lundin, C.; Hodgson, B.; Sharma, R. A. DNA repair pathways as targets for cancer therapy. *Nat. Rev. Cancer* **2008**, *8* (3), 193-204. DOI: 10.1038/nrc2342

(34) Wallner, O.; Cázares-Körner, A.; Scaletti, E. R.; Masuyer, G.; Bekkhus, T.; Visnes, T.; Mamonov, K.; Ortis, F.; Lundbäck, T.; Volkova, M.; et al. Optimization of n-piperidinyl-benzimidazolone derivatives as potent and selective inhibitors of 8-oxo-guanine DNA glycosylase 1. *ChemMedChem* **2023**, *18* (1), e202200310. DOI: https://doi.org/10.1002/cmdc.202200310

(35) Weilbeer, C.; Jay, D.; Donnelly, J. C.; Gentile, F.; Karimi-Busheri, F.; Yang, X.; Mani, R. S.; Yu, Y.; Elmenoufy, A. H.; Barakat, K. H.; et al. Modulation of ERCC1-XPF heterodimerization inhibition via structural modification of small molecule inhibitor side-chains. *Front. Oncol.* **2022**, *12*, 819172. DOI: 10.3389/fonc.2022.819172

(36) Yang, F.; Hu, Z.; Guo, Z. Small-molecule inhibitors targeting FEN1 for cancer therapy. *Biomolecules* **2022**, *12* (7), 1007. DOI: 10.3390/biom12071007

(37) Visnes, T.; Cázares-Körner, A.; Hao, W.; Wallner, O.; Masuyer, G.; Loseva, O.; Mortusewicz, O.; Wiita, E.; Sarno, A.; Manoilov, A.; et al. Small-molecule inhibitor of OGG1 suppresses proinflammatory gene expression and inflammation. *Science* **2018**, *362* (6416), 834-839. DOI: doi:10.1126/science.aar8048

(38) Donley, N.; Jaruga, P.; Coskun, E.; Dizdaroglu, M.; McCullough, A. K.; Lloyd, R. S. Small molecule inhibitors of 8-oxoguanine DNA glycosylase-1 (OGG1). *ACS Chem. Biol.* **2015**, *10* (10), 2334-2343. DOI: 10.1021/acschembio.5b00452

(39) Bliksrud, Y. T.; Ellingsen, A.; Bjørås, M. Fumarylacetoacetate inhibits the initial step of the base excision repair pathway: Implication for the pathogenesis of tyrosinemia type I. *J. Inherit. Metab. Dis.* **2013**, *36* (5), 773-778. DOI: https://doi.org/10.1007/s10545-012-9556-0

(40) Korycka, A.; Blonski, Z. J.; Robak, T. Forodesine (BCX-1777, Immucillin H) - a new purine nucleoside analogue: Mechanism of action and potential clinical application. *Mini-Rev. Med. Chem.* **2007**, *7* (9), 976-983. DOI: 10.2174/138955707781662636

(41) Schramm, V. L. Transition states, analogues, and drug development. *ACS Chem. Biol.* **2013**, *8* (1), 71-81. DOI: 10.1021/cb300631k

(42) Jarvis, L. M. Pushing cancer over the edge. *Chem. Eng. News*, 2013, 13-18.

(43) Zhang, Z.; Wu, Z.; Shi, X.; Guo, D.; Cheng, Y.; Gao, J.; Liu, L.; Liu, W.; Liang, L.; Peng, L.; et al. Research progress in human AP endonuclease 1: Structure, catalytic mechanism, and inhibitors. *Curr. Protein Peptide Sci.* **2022**, *23* (2), 77-88. DOI: 10.2174/1389203723666220406132737

(44) Abramić, M.; Agić, D. Survey of dipeptidyl Peptidase III inhibitors: From small molecules of microbial or synthetic origin to aprotinin. *Molecules* **2022**, *27* (9), 3006. DOI: 10.3390/molecules27093006

(45) Nakamura, T.; Okabe, K.; Hirayama, S.; Chirifu, M.; Ikemizu, S.; Morioka, H.; Nakabeppu, Y.; Yamagata, Y. Structure of the mammalian adenine DNA glycosylase MUTYH: Insights into the base excision repair pathway and cancer. *Nucleic Acids Res.* **2021**, *49* (12), 7154-7163. DOI: 10.1093/nar/gkab492 (acccessed 8/18/2021).

(46) Luncsford, P. J.; Chang, D.-Y.; Shi, G.; Bernstein, J.; Madabushi, A.; Patterson, D. N.; Lu, A. L.; Toth, E. A. A structural hinge in eukaryotic MutY homologues mediates catalytic activity and Rad9–Rad1–Hus1 checkpoint complex interactions. *J. Mol. Biol.* **2010**, *403* (3), 351-370. DOI: 10.1016/j.jmb.2010.08.045

(47) Lu, A. L.; Yuen, D. S.; Cillo, J. Catalytic mechanism and DNA substrate recognition of *escherichia coli* MutY protein. *J. Biol. Chem.* **1996**, *271* (39), 24138-24143. DOI: 10.1074/jbc.271.39.24138 (acccessed 2022/01/06).

(48) Wright, P. M.; Yu, J.; Cillo, J.; Lu, A. L. The active site of the *escherichia coli* MutY DNA adenine glycosylase. *J. Biol. Chem.* **1999**, *274* (41), 29011-29018. DOI: 10.1074/jbc.274.41.29011

(49) Williams, S. D.; David, S. S. A single engineered point mutation in the adenine glycosylase MutY confers bifunctional glycosylase/AP lyase activity. *Biochemistry* **2000**, *39* (33), 10098-10109. DOI: 10.1021/bi0004652

(50) Chepanoske, C. L.; Lukianova, O. A.; Lombard, M.; Golinelli-Cohen, M.-P.; David, S. S. A residue in MutY important for catalysis identified by photocross-linking and mass spectrometry. *Biochemistry* **2004**, *43* (3), 651-662. DOI: 10.1021/bi035537e

(51) Brinkmeyer, M. K.; Pope, M. A.; David, S. S. Catalytic contributions of key residues in the adenine glycosylase MutY revealed by pH-dependent kinetics and cellular repair assays. *Cell Chem. Biol.* **2012**, *19* (2), 276-286, 10.1016/j.chembiol.2011.11.011. DOI: 10.1016/j.chembiol.2011.11.011

(52) Francis, A. W.; Helquist, S. A.; Kool, E. T.; David, S. S. Probing the requirements for recognition and catalysis in Fpg and MutY with nonpolar adenine isosteres. *J. Am. Chem. Soc.* **2003**, *125* (52), 16235-16242. DOI: 10.1021/ja0374426

(53) 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, 10.1021/ja711363s. DOI: 10.1021/ja711363s

(54) Fromme, J. C.; Banerjee, A.; Huang, S. J.; Verdine, G. L. Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase. *Nature* **2004**, *427*, 652-656, Article. DOI: 10.1038/nature02306 (acccessed 2022/1/15/).

(55) 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* (44), 18497-18502, S18497/18491-S18497/18495, 10.1073/pnas.0902908106. DOI: 10.1073/pnas.0902908106

(56) Woods, R. D.; O'Shea, V. L.; Chu, A.; Cao, S.; Richards, J. L.; Horvath, M. P.; David, S. S. Structure and stereochemistry of the base excision repair glycosylase MutY reveal a mechanism similar to retaining glycosidases. *Nucleic Acids Res.* **2016**, *44* (2), 801-810, 10.1093/nar/gkv1469. DOI: 10.1093/nar/gkv1469

(57) Demir, M.; Russelburg, L. P.; Lin, W.-J.; Trasviña-Arenas, C. H.; Huang, B.; Yuen, P. K.; Horvath, M. P.; David, S. S. Structural snapshots of base excision by the cancer-associated variant MutY N146S reveal a retaining mechanism. *Nucleic Acids Res.* **2023**, *51* (3), 1034-1049. DOI: 10.1093/nar/gkac1246 (acccessed 1/23/2023).

(58) Stivers, J. T.; Jiang, Y. L. A mechanistic perspective on the chemistry of DNA repair glycosylases. *Chem. Rev.* **2003**, *103* (7), 2729-2760. DOI: 10.1021/cr010219b

(59) 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. DOI: 10.1021/cr040461t

(60) Pidugu, L. S.; Bright, H.; Lin, W. J.; Majumdar, C.; Van Ostrand, R. P.; David, S. S.; Pozharski, E.; Drohat, A. C. Structural insights into the mechanism of base excision by MBD4. *J. Mol. Biol.* **2021**, *433* (15), 167097. DOI: 10.1016/j.jmb.2021.167097

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

(62) Rutledge, L. R.; Wetmore, S. D. Modeling the chemical step utilized by human alkyladenine DNA glycosylase: A concerted mechanism aids in selectively excising damaged purines. *J. Am. Chem. Soc.* **2011**, *133* (40), 16258-16269. DOI: 10.1021/ja207181c

(63) Aboelnga, M. M.; Wetmore, S. D. Unveiling a single-metal-mediated phosphodiester bond cleavage mechanism for nucleic acids: A multiscale computational investigation of a human DNA repair enzyme. *J. Am. Chem. Soc.* **2019**, *141* (21), 8646-8656. DOI: 10.1021/jacs.9b03986

(64) Wang, B.; Usharani, D.; Li, C.; Shaik, S. Theory uncovers an unusual mechanism of DNA repair of a lesioned adenine by AlkB enzymes. *J. Am. Chem. Soc.* **2014**, *136* (39), 13895-13901. DOI: 10.1021/ja507934g

(65) Kreppel, A.; Blank, I. D.; Ochsenfeld, C. Base-independent DNA base-excision repair of 8-oxoguanine. *J. Am. Chem. Soc.* **2018**, *140* (13), 4522-4526. DOI: 10.1021/jacs.7b11254

(66) Sadeghian, K.; Ochsenfeld, C. Unraveling the base excision repair mechanism of human DNA glycosylase. *J. Am. Chem. Soc.* **2015**, *137* (31), 9824-9831. DOI: 10.1021/jacs.5b01449

(67) Wang, S. D.; Zhang, R. B.; Eriksson, L. A. Markov state models elucidate the stability of DNA influenced by the chiral 5S-Tg base. *Nucleic Acids Res.* **2022**, *50* (16), 9072-9082. DOI: 10.1093/nar/gkac691

(68) Panigrahi, A.; Vemuri, H.; Aggarwal, M.; Pitta, K.; Krishnan, M. Sequence specificity, energetics and mechanism of mismatch recognition by DNA damage sensing protein Rad4/XPC. *Nucleic Acids Res.* **2020**, *48* (5), 2246-2257. DOI: 10.1093/nar/gkaa078

(69) Lotsof, E. R.; Krajewski, A. E.; Anderson-Steele, B.; Rogers, J.; Zhang, L.; Yeo, J.; Conlon, S. G.; Manlove, A. H.; Lee, J. K.; David, S. S. NEIL1 recoding due to RNA editing impacts lesion-specific recognition and excision. *J. Am. Chem. Soc.* **2022**, *144* (32), 14578-14589. DOI: 10.1021/jacs.2c03625

(70) Donati, E.; Genna, V.; De Vivo, M. Recruiting mechanism and functional role of a third metal ion in the enzymatic activity of 5′ structure-specific nucleases. *J. Am. Chem. Soc.* **2020**, *142* (6), 2823-2834. DOI: 10.1021/jacs.9b10656

(71) Naydenova, E.; Roßbach, S.; Ochsenfeld, C. QM/MM study of the uracil DNA glycosylase reaction mechanism: A competition between Asp145 and His148. *J. Chem. Theory Comput.* **2019**, *15* (8), 4344-4350. DOI: 10.1021/acs.jctc.8b01305

(72) Tiwari, S.; Agnihotri, N.; Mishra, P. C. Quantum theoretical study of cleavage of the glycosidic bond of 2'-deoxyadenosine: Base excision-repair mechanism of DNA by MutY. *J. Phys. Chem. B* **2011**, *115* (12), 3200-3207, 10.1021/jp1109256. DOI: 10.1021/jp1109256

(73) 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* (48), 8753-8765, 10.1021/bi401310w. DOI: 10.1021/bi401310w

(74) Livingston, A. L.; O'Shea, V. L.; Kim, T.; Kool, E. T.; David, S. S. Unnatural substrates reveal the importance of 8-oxoguanine for *in vivo* mismatch repair by MutY. *Nat. Chem. Biol.* **2008**, *4* (1), 51-58. DOI: 10.1038/nchembio.2007.40

(75) McCann, J. A. B.; Berti, P. J. Adenine release is fast in MutY-catalyzed hydrolysis of G:A and 8-oxo-g:A DNA mismatches. *J. Biol. Chem.* **2003**, *278* (32), 29587-29592, 10.1074/jbc.M212474200. DOI: 10.1074/jbc.M212474200

(76) Davies, G. J.; Mackenzie, L.; Varrot, A.; Dauter, M.; Brzozowski, A. M.; Schülein, M.; Withers, S. G. Snapshots along an enzymatic reaction coordinate: Analysis of a retaining β-glycoside hydrolase. *Biochemistry* **1998**, *37* (34), 11707-11713. DOI: 10.1021/bi981315i

(77) Rovira, C.; Males, A.; Davies, G. J.; Williams, S. J. Mannosidase mechanism: At the intersection of conformation and catalysis. *Curr. Opin. Struct. Biol.* **2020**, *62*, 79-92. DOI: https://doi.org/10.1016/j.sbi.2019.11.008

(78) Schrodinger, LLC. The PyMOL molecular graphics system, version 1.8. 2015.

(79) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; III, T. E. C.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; et al. AMBER 2018. **2018**, University of California, San Fransisco.

(80) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696-3713. DOI: 10.1021/acs.jctc.5b00255

(81) Galindo-Murillo, R.; Robertson, J. C.; Zgarbova, M.; Sponer, J.; Otyepka, M.; Jurecka, P.; Cheatham, T. E. Assessing the current state of amber force field modifications for DNA. *J. Chem. Theory Comput.* **2016**, *12* (8), 4114-4127, 10.1021/acs.jctc.6b00186. DOI: 10.1021/acs.jctc.6b00186

(82) Li, P.; Merz, K. M. Mcpb.Py: A python based metal center parameter builder. *J. Chem Inf. Model.* **2016**, *56* (4), 599-604. DOI: 10.1021/acs.jcim.5b00674

(83) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *26* (1), 114, 10.1002/jcc.20145. DOI: 10.1002/jcc.20145

(84) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **2006**, *25* (2), 247-260, 10.1016/j.jmgm.2005.12.005. DOI: 10.1016/j.jmgm.2005.12.005

(85) Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. The R.E.D. Tools: Advances in RESP and ESP charge derivation and force field library building. *Phys. Chem. Chem. Phys.* **2010**, *12* (28), 7821-7839, 10.1039/c0cp00111b. DOI: 10.1039/c0cp00111b

(86) Wei, W.-J.; Liao, R.-Z. QM/MM study of the [4Fe-4S]-dependent (R)-2-hydroxyisocaproyl-CoA dehydratase: Dehydration via a redox pathway with an α-carbonyl radical intermediate. *J. Catal.* **2023**. DOI: https://doi.org/10.1016/j.jcat.2023.01.034

(87) Dong, L.; Liu, Y. Catalytic mechanism of pyridoxal 5′-phosphate-dependent aminodeoxychorismate lyase: A computational QM/MM study. *J. Chem. Inf. Model.* **2023**, *63* (4), 1313-1322. DOI: 10.1021/acs.jcim.2c01593

(88) Sgrignani, J.; Magistrato, A. QM/MM MD simulations on the enzymatic pathway of the human flap endonuclease (hFEN1) elucidating common cleavage pathways to RNase h enzymes. *ACS Catal.* **2015**, *5* (6), 3864-3875. DOI: 10.1021/acscatal.5b00178

(89) Pilbák, S.; Farkas, Ö.; Poppe, L. Mechanism of the tyrosine ammonia lyase reaction—tandem nucleophilic and electrophilic enhancement by a proton transfer. *Chem. Eur. J.* **2012**, *18* (25), 7793-7802. DOI: https://doi.org/10.1002/chem.201103662

(90) Hu, X.; Hu, H.; Melvin, J. A.; Clancy, K. W.; McCafferty, D. G.; Yang, W. Autocatalytic intramolecular isopeptide bond formation in gram-positive bacterial pili: A QM/MM simulation. *J. Am. Chem. Soc.* **2011**, *133* (3), 478-485. DOI: 10.1021/ja107513t

(91) Wu, S.; Xu, D.; Guo, H. QM/MM studies of monozinc β-lactamase CphA suggest that the crystal structure of an enzyme−intermediate complex represents a minor pathway. *J. Am. Chem. Soc.* **2010**, *132* (51), 17986-17988. DOI: 10.1021/ja104241g

(92) Stevens, D. R.; Hammes-Schiffer, S. Exploring the role of the third active site metal ion in DNA polymerase η with QM/MM free energy simulations. *J. Am. Chem. Soc.* **2018**, *140* (28), 8965-8969. DOI: 10.1021/jacs.8b05177

(93) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; et al. *Gaussian 16, revision b.01*; Gaussian, Inc., 2016.

(94) Hollis, T.; Ichikawa, Y.; Ellenberger, T. DNA bending and a flip-out mechanism for base excision by the helix-hairpin-helix DNA glycosylase, escherichia coli alka. *EMBO J* **2000**, *19* (4), 758-766. DOI: 10.1093/emboj/19.4.758

(95) Altona, C.; Sundaralingam, M. Conformational analysis of the sugar ring in nucleosides and nucleotides. New description using the concept of pseudorotation. *J. Am. Chem. Soc.* **1972**, *94* (23), 8205-8212. DOI: 10.1021/ja00778a043

(96) Ribeiro, A. J. M.; Santos-Martins, D.; Russo, N.; Ramos, M. J.; Fernandes, P. A. Enzymatic flexibility and reaction rate: A QM/MM study of HIV-1 protease. *ACS Catal.* **2015**, *5* (9), 5617-5626. DOI: 10.1021/acscatal.5b00759

(97) 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. DOI: 10.1021/bi9910878

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

(99) 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. DOI: 10.1021/acs.jpcb.7b09646

(100) Hashimoto, H.; Zhang, X.; Cheng, X. Excision of thymine and 5-hydroxymethyluracil by the MBD4 DNA glycosylase domain: Structural basis and implications for active DNA demethylation. *Nucleic Acids Res.* **2012**, *40* (17), 8276-8284. DOI: 10.1093/nar/gks628

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

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

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

(104) Dodson, M. L.; Michaels, M. L.; Lloyd, R. S. Unified catalytic mechanism for DNA glycosylases. *J. Biol. Chem.* **1994**, *269* (52), 32709-32712. DOI: https://doi.org/10.1016/S0021-9258(20)30047-8

(105) He, L.; Yang, H.; Zhou, S.; Zhu, H.; Mao, H.; Ma, Z.; Wu, T.; Kumar, A. K.; Kathera, C.; Janardhan, A.; et al. Synergistic antitumor effect of combined paclitaxel with FEN1 inhibitor in cervical cancer cells. *DNA Repair* **2018**, *63*, 1-9. DOI: https://doi.org/10.1016/j.dnarep.2018.01.003

(106) Wu, T.; Zhu, H.; Zhang, M.; Sun, Y.; Yang, Y.; Gu, L.; Zhang, J.; Mu, D.; Wu, C.; Hu, Z.; et al. FEN1 inhibitor synergizes with low-dose camptothecin to induce increased cell killing via the mitochondria mediated apoptotic pathway. *Gene Therapy* **2022**, *29* (7), 407-417. DOI: 10.1038/s41434-020-00215-9

Chapter 3: A Tale of Two Mechanisms: Clarification of the Pathway for MBD4 Catalyzed Glycosidic Bond Cleavage Using MD and QM/MM Calculations

3.1. Introduction

Epigenetic changes to mammalian DNA are essential for regulating development and growth of organisms but also contribute to the progression of diseases.[1, 2] A key epigenetic modification is the methylation of cytosine (C) to 5-methylcytosine (5mC), with approximately 4% of cytosines being methylated in human cells.[3, 4] Cytosine methylation is greatly enriched in cytosine-phosphate-guanine (CpG) dinucleotides, with ~80% containing methylated cytosines in mammalian cell genomes.[4, 5] 5mC is vital for organism survival, with roles in genomic imprinting,[6] X-chromosome inactivation,[7] and gene expression patterns.[3, 8, 9] However, like canonical (unmodified) C,[10] 5mC is vulnerable to deamination damage.[11] While the deamination of cytosine results in the non-canonical DNA base uracil, deamination of 5mC leads to canonical thymine (T). Although this causes T:G mispairs, the formation of a canonical DNA base leads to 5mC deamination avoiding standard DNA repair pathways that handle cytosine deamination, thus requiring specialized strategies to avoid mutagenic outcomes.[12-14]

Methyl-CpG-binding domain protein 4 (MBD4) is responsible for removing T arising from 5mC deamination, as well as several other forms of damaged C, when mispaired with guanine in CpG dinucleotides, which results in an abasic site that is further processed along the base excision repair (BER) pathway.[15-17] Due to the important role MBD4 plays in cells, the enzyme has been implicated in a variety of human health disorders.[18, 19] For example, MBD4 deficiency results in increased levels of CpG to TpG mutations,[18] which can lead to MBD4-associated neoplasia syndrome (MANS) in humans that is associated with an increased risk for colorectal polyposis, acute myeloid leukemia, and uveal melanoma.[19, 20] As a DNA repair enzyme, MBD4 is also a potential target for small molecule inhibitors to enhance the effectiveness of existing cancer treatments,[21-28] with inhibitors previously successfully designed for enzymes with similar functions to DNA glycosylases.[29-35] Indeed, MBD4 has been shown to be primarily responsible for cancer cell resistance to the first-line chemotherapeutic agent 5-fluorouracil (5-FU) and downregulation of

MBD4 increases 5-FU cytotoxicity.[36, 37] To expand our knowledge on MANS and enable the rational design of small-molecule inhibitors to enhance existing cancer therapeutics,[38, 39] atomic level details of the MBD4 catalytic mechanism are required.

There are currently conflicting proposals for the MBD4 catalytic pathway.[16, 40] The first proposed mechanism is based on a crystal structure of mouse MBD4 bound to DNA containing a G:T mispair, which exhibits a DNA–protein crosslinked intermediate bound in the active site (PDB ID: 4EW4, 2.79 Å resolution, Figure 3.1A).[16] The observed crosslink is formed between C1′ of the substrate nucleic acid and D534 (equivalent to D560 in human MBD4). In the proposed human MBD4 crosslinking mechanism, D560 attacks C1′ of deoxyribose, breaking the glycosidic bond and forming the DNA–protein crosslinked intermediate (Figure 3.1C).[16] The crosslink is then hydrolyzed in a subsequent step to form the final abasic site product. This proposal is consistent with the D560G mutation resulting in a 2700-fold reduction in human MBD4 activity[40] and the replacement of D534 in mouse MBD4 with asparagine leading to complete loss of catalytic activity.[16] A crosslinking mechanism would distinguish MBD4 from most other glycosylases, with the only other example involving a crosslinked intermediate being adenine DNA glycosylase (MutY) based on an abundance of experimental and computational evidence.[41-43] Uracil DNA glycosylase X (UdgX) has also been proposed to form a DNA–protein crosslink,[44] but as the product of the reaction, which deactivates the enzyme. While rare in glycosylases, the O-glycosidases commonly cleave glycosidic bonds in carbohydrates through crosslink formation with an active site D/E.[45, 46] However, both MutY[41] and O-glycosidase[45, 46] catalysis through a crosslinked intermediate results in retention of stereochemistry in the product, while NMR and methanolysis experiments highlight inversion of stereochemistry for the product generated by MBD4.[40] This implies that a crosslinking mechanism for MBD4 would require a unique crosslink hydrolysis step compared to these other enzymes.

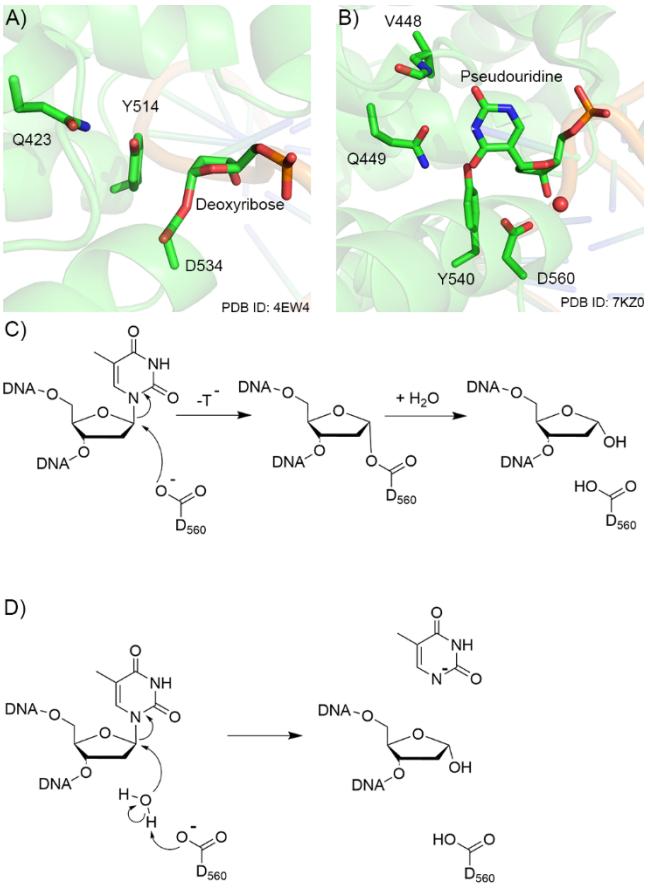


Figure 3.1. Crystal structure for A) mouse MBD4 containing a DNA–protein crosslinked intermediate and B) human MBD4 complexed with pseudouridine-containing DNA. Proposed human MBD4 catalytic mechanisms involving C) a DNA–protein crosslinked intermediate and D) direct hydrolysis of the glycosidic bond.

A more recent crystal structure of MBD4 bound to pseudouridine-containing DNA challenges the proposed crosslinking mechanism (PDB ID: 7KZ0, 1.57 Å resolution, Figure 3.1B).[40] Specifically, in the active site, D560 and a water molecule are positioned near the modified nucleotide in a manner that closely resembles the alignment observed for other DNA glycosylases.[47-50] This resulted in an alternate proposed mechanism in which the glycosidic bond is directly hydrolyzed by a water nucleophile that is activated by D560 to form the abasic site product (Figure 3.1D).[40] The proposed role of D560 in direct nucleotide hydrolysis is aligned with the function of active site D/E residues found in other DNA glycosylases, such as several members

of the UDG family[14, 51] and alkyladenine DNA glycosylase (AAG)[52-54] as well as consistent with the reported inversion of C1′ stereochemistry upon MBD4 function.[40]

Computational methodologies have proven to be invaluable for determining the mechanisms of action of DNA repair enzymes at an atomic level of detail.[42, 51, 53-80] Molecular dynamics (MD) simulations can provide dynamic structural descriptions of enzyme active sites, which cannot be obtained from static crystal structures, and play an integral part in identifying the roles of residues in catalytic mechanisms.[42, 64, 81] In direct complement, quantum mechanics/molecular mechanics (QM/MM) calculations can be used to map enzyme mechanisms[42, 64, 81-90] with a level of accuracy that allows for comparisons to experimental reaction rates.[91] Indeed, these techniques are most effective when used together to synergistically provide the dynamic, structural, and electronic information necessary to confidently predict enzymatic mechanisms. In terms of DNA glycosylases, QM models, MD simulations, and/or QM/MM calculations have been used to study a variety of enzymes (see, for example, Refs. 61 and 74 and references therein). For example, MD simulations and QM/MM calculations have been used to determine how active site D and E residues facilitate crosslink formation and enzyme regeneration in the mechanism used by MutY to remove canonical adenine (A) in mispairs.[42] Alternatively, MD simulations identified residues in thymine DNA glycosylase (TDG) that are important for water nucleophile positioning and product stabilization, while QM/MM simulations characterized necessary substrate reorganization steps and a direct hydrolysis mechanism for the repair of damaged pyrimidines.[55, 61, 74]

Due to discrepancies in the proposed catalytic pathway of MBD4 and previous successes of computational studies in unraveling the mechanism of action of DNA glycosylases,[42, 51, 61, 72, 74] the present study uses a combination of adaptively biased molecular dynamics (abMD) simulations and quantum mechanics/molecular mechanics (QM/MM) calculations to probe the catalytic mechanism for MBD4 glycosylase activity. abMD simulations on the MBD4–substrate complex reveal the high flexibility of D560 and uncover several conformations that may contribute to enzyme activity. Subsequently, QM/MM calculations on multiple snapshots from the abMD simulations are used to characterize seven different excision pathways to explore the relative feasibility of crosslink

formation versus direct DNA hydrolysis. The preferred mechanism identified through this rigorous computational approach is fully consistent with existing experimental data, including the experimental structure of MBD4 bound to pseudouridine-containing DNA,[40] the stereochemistry of the abasic site product,[40] mutagenic data,[16, 40] and the catalytic rate.[40] This work provides foundational knowledge for developing new treatments for MANS[19] and improving cancer therapies, [21-28] especially for 5-FU resistant cancers.[36, 37]

3.2. Computational Methods

3.2.1. abMD simulations

A model to initiate abMD simulations was built based on the crystal structure of human MBD4 bound to pseudouridine (PDB ID: 7KZ0).[40] Specifically, pseudouracil was converted to the T substrate using PyMOL.[92] Protonation states of all residues and the isomers of each histidine were determined using the H++ server.[93] The model was solvated with TIP3P water in a periodic rectangular prism using the LEaP module in AMBER 2018[94] such that the edges are at least 10 Å from the solute. Potassium ions were added to neutralize the system and additional KCl was added to achieve a physiological salt concentration (150 mM). The Amber force field was used, with ff14SB[95] describing the protein residues and OL15[96] modeling the nucleic acid.

The model was initially minimized using 1000 steps of steepest decent followed by 1000 steps of conjugate gradient minimization. First, the solvent was minimized, while a 100 kcal mol^{-1} $Å^{-2}$ restraint was applied to the solute. Next, the restraint was removed from the solute hydrogen atoms. Subsequently, the solute was minimized, while applying a 100 kcal mol^{-1} $Å^{-2}$ restraint to the solvent. Finally, the entire system was minimized with no restraints, using 1000 steps of steepest decent followed by 2000 steps of conjugate gradient minimization. Each system was then heated from 10 K, with the temperature increased by 50 K every 20 ps until 310 K was reached. The restraints on the solute were subsequently decreased from 25 kcal mol^{-1} $Å^{-2}$ to 5 kcal mol^{-1} $Å^{-2}$ at a rate of 5 kcal mol^{-1} $Å^{-2}$ per 20 ps. A final 20 ps equilibration step was completed with a solute restraint of

1.5 kcal mol^{-1} $Å^{-2}$. For these simulations, the Langevin thermostat was used with a collision frequency of 1 ps^{-1} and an NVT ensemble.

After equilibration, five 500 ns abMD simulation replicates were performed. The simulations used two collective variables: 1) the distance between C1′ of thymidine and Oδ of D560 (ranging from 2.5 Å to 18.0 Å), and 2) the angle formed between N1 of thymidine, C1′ of thymidine, and Oδ of D560 (ranging from 0 to 180°; Figure 3.2A). The simulations were conducted at a well-tempered temperature of 5000 K and a flooding timescale of 100 ps. A potential energy surface (PES) for each replica was generated from the simulations using the nfe-umbrella-slice utility in AMBER 2018 (Figure B.1, Appendix B)[94] and averaged to create an overall PES. An anomalous region identifiable by a sudden and extreme change in relative energy compared to adjacent regions of the surface appeared in only two replicas (Figure B.1, Appendix B). Since this likely arose due to limited sampling in those windows (less than 0.04% of the simulation time was spent at that coordinate), these regions were not included in the average. Analysis was performed using the cpptraj program in AMBER 2018.[94] Frames were assigned to conformations for analysis using specific collective variable (CV) criteria (Table B.1, Appendix B). Hydrogen-bond occupancy was determined based on a heavy atom distance of less than 3.2 Å and a hydrogen-bonding angle between 135° and 180°. A water was considered available for a hydrolysis mechanism if simultaneously positioned less than 3.5 Å away from C1′ of T and Oδ of D560.

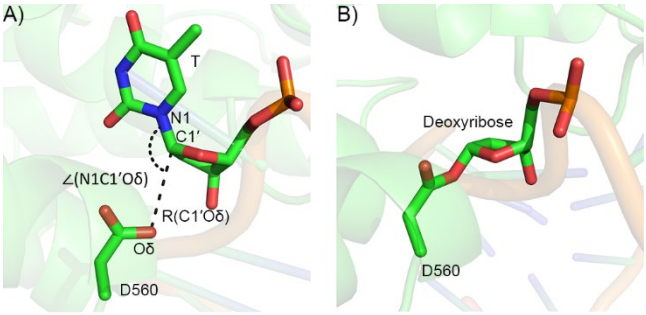


Figure 3.2. A) Depiction of the distance (r(C1′Oδ)) and angle (∠(N1C1′Oδ)) CVs used in abMD simulations. B) Post-equilibration model of the DNA–MBD4 crosslink used in classical MD simulations.

3.2.2. Classical MD Simulations

To create a model of the MBD4–DNA complex containing a DNA–protein crosslink, the crystal structure of pseudouridine-containing DNA bound to human MBD4 (PDB ID: 7KZ0) was first optimized using the QM/MM methodology described in the next section. The crosslink was then formed by removing pseudouridine and repositioning D260 using PyMOL (Figure 3.2B). Parameters for the crosslink were taken from our previous work on MutY,[42] and the minimization, heating, and equilibration procedure described for the abMD simulations was implemented. Five 500 ns MD production simulations were then performed using the same simulation parameters discussed for abMD without invoking the adaptively biased potential. This approach resulted in an MD representative structure that is structurally similar to the crystal structure of the crosslinked intermediate (Figure B.2A, Appendix B).

3.2.3. QM/MM calculations

QM/MM models were generated using representative structures obtained by clustering abMD frames that satisfied CV criteria near the minima (Table B.2, Appendix B). Each model contains the protein and nucleic acid solute as well as any solvent molecule with an atom within 8 Å of a solute atom. Two relative conformations of D560 and T were considered as rationalized in the results section. The model of the first conformation (denoted below, Table B.2, Appendix B) was used to map both the direct hydrolysis and crosslink formation mechanisms. In this model (Figure 3.3A), the QM region (150 atoms) contains L447 truncated at the Cα–Cβ and Cα–N bonds, V448 truncated at the Cα–Cβ bond, Q449 truncated at the Cα–C bond, N467 truncated at the Cα–Cβ and Cα–N bonds, R468 truncated at the Cα–C bond, Y540 truncated at the Cα–Cβ bond, D560 truncated at the Cα–Cβ bond, K562 truncated at the Cα–N bond, L563 truncated at the Cα–Cβ and Cα–C bonds, the thymidine substrate and the 3′ and 5′ phosphate groups truncated at the next C5′ and C3′ atoms, as well as four waters in the active site, which results in a charge of –1. As discussed in the results, two additional QM/MM models were built for the below conformation from points on the abMD PES with r(C1′Oδ) 1 or 2 Å shorter than the value at the below conformational minimum. The model of

the second conformation (denoted side, Table B.2, Appendix B) was used to map a direct hydrolysis mechanism. In this model (Figure 3.3B), the QM region (154 atoms) contains L447 truncated at the Cα–Cβ and Cα–N bonds, V448, Q449 truncated at the Cα–C bond, N467 truncated at the Cα–Cβ and Cα–N bonds, R468 truncated at the Cα–C bond, Y540 truncated at the Cα–Cβ bond, D560 truncated at the Cα–Cβ bond, K562 truncated at the Cα–N bond, L563 truncated at the Cα–Cβ and Cα–C bonds, the thymidine substrate and the 3′ and 5′ phosphate groups truncated at the next C5′ and C3′ atoms, as well as two waters in the active site, which results in a charge of –1. A QM/MM model was also generated from a representative structure obtained from classical MD simulations on the crosslinked intermediate, which was used to map the hydrolysis of the crosslink (Figure 3.3C). The QM region (123 atoms) contains N467 truncated at the Cα–Cβ and Cα–N bonds, R468 truncated at the Cα–C bond, Y540 truncated at the Cα–Cβ bond, the crosslinked D560 truncated at the Cα–Cβ bond, K562 truncated at the Cα–N bond, L563 truncated at the Cα–Cβ and Cα–C bonds, the crosslinked deoxyribose substrate and the 3′ and 5′ phosphate groups truncated at the next C5′ and C3′ atoms, as well as eight waters in the active site, which results in a charge of 0.

All QM/MM models were optimized using ONIOM(M06-2X/6-31G(d,p):AMBERff14SB). Transition states (TS) for the catalytic mechanism were obtained by scanning the formation and cleavage of the relevant bonds to generate a reaction energy curve from which structures corresponding to the maxima were further optimized. Frequency calculations were conducted to characterize the nature of the stationary points using the same level of theory as the optimizations as well as to verify that the isolated imaginary frequency corresponds to the anticipated transition structure, and to obtain Gibbs energy corrections. Each pathway was verified to be continuous by mapping the internal reaction coordinate through scanning the nucleophile–electrophile and leaving group distance both to and from each stationary point. The reported reaction Gibbs energies were obtained from ONIOM(M06-2X/6-311+G(2df,p):AMBERff14SB) single-point calculations. Optimization, frequency, and single-point calculations utilized the mechanical embedding (ME) scheme due to the robustness of this approach and its successes in accurately modeling other enzymatic reactions.[42, 90, 97-101] To ensure that the ME embedding scheme accurately describes the chemistry under investigation, a model was built from the pseudouridine-bound MBD4 crystal

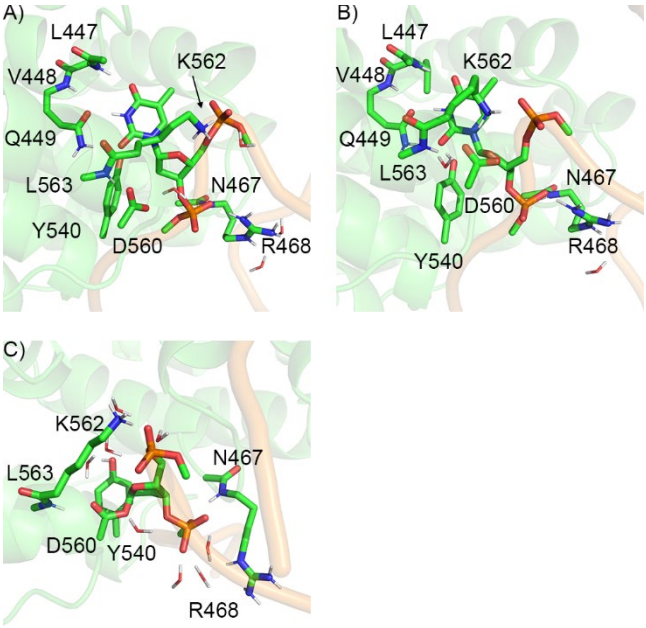


Figure 3.3. High-layer atoms in QM/MM models of the MBD4 active site containing the A) D560 below conformation (D560 positioned on the opposite side of deoxyribose as T), B) D560 side conformation (D560 in the plane of the deoxyribose ring), and C) DNA–protein crosslink.

structure (PDB ID: 7KZ0)[40] by adding an 8 Å layer of water and defining the same the QM region as described for the analogous models built from abMD simulations (below conformation). The model was then used to map the preferred pathway using the ME and electronic embedding (EE) schemes for comparison. Both approaches result in a one-step pathway with energy barriers within 1 kJ/mol, further supporting that ME is sufficient to describe the chemistry facilitated by MBD4 (Figure B.3, Appendix B).

A comparison of computed catalytic pathways to the relevant experimental structural,[40] stereoscopic,[40] mutagenic,[16, 40] and kinetic[40] data was used to determine the preferred mechanism, with calculated barriers being compared to the experimental barrier (92.3 kJ/mol) estimated from a rate constant of 0.11 min^{-1} that was measured at 23°C and a pH of 7.5 under single-turnover conditions[40] using the Eyring equation.

3.3. Results

3.3.1. The catalytically implicated D560 residue is highly dynamic in the MBD4 active site, adopting multiple potentially active conformations that support previously proposed mechanisms

Mutagenic studies have revealed that D560 is a key residue in the MBD4 catalytic mechanism, with the D560G mutation resulting in a 2700-fold reduction in activity[40] and the D534N mutation deactivating mouse MBD4.[16] Nevertheless, preliminary classical MD simulations on the MBD4–substrate complex revealed that D560 is highly flexible in the active site, suggesting the D560 conformation is likely a key factor in determining the catalytic mechanism. Therefore, to fully explore the conformational landscape of the active site and ensure a reliable starting point for subsequent QM/MM calculations, abMD simulations were employed to investigate the conformational dynamics of D560. Both experimentally-proposed MBD4 reaction mechanisms (crosslink formation and direct hydrolysis, Figure 3.1) require a short distance between C1′ of the T deamination product and Oδ of D560 (r(C1′Oδ)), while crosslink formation will be more favorable as the ∠(N1C1′Oδ) angle approaches 180° (Figure 3.2A). Therefore, the PES was mapped as a function of r(C1′Oδ) and ∠(N1C1′Oδ) as the collective variables.

The PES generated from the abMD simulations confirms that D560 is very flexible in the MBD4 active site (Figure 3.4A). Indeed, a large portion of the PES is within 17 kJ/mol of the global minimum and local minima are broad and shallow. Three minima that differ in the relative orientation of D560 and the substrate are within 8 kJ/mol with less than a 10 kJ/mol barrier separating them (Figure 3.4B–D, Table B.3, Appendix B). Despite the significant flexibility of D560, T is consistently positioned through hydrogen bonding with V448 (79–94% occupancy), Q449 (79–96% occupancy), and Y540 (71–84% occupancy, Figure B.4A, Appendix B) in the regions surrounding all minima on the PES.

The global minimum on the PES occurs at ∠(N1C1′Oδ) = ~150° and r(C1′Oδ) = 8.00 Å (Figure 3.4D), which directs the side chain of D560 away from T and into the solvent, precluding participation in the catalytic mechanism (denoted the inactive conformation). The next lowest energy minimum on the PES (ΔG = 4 kJ/mol relative to global minimum; ∠(N1C1′Oδ) = ~110°; r(C1′Oδ) = 5.25 Å; Figure 3.4C) situates D560 in the plane of the deoxyribose ring (denoted side

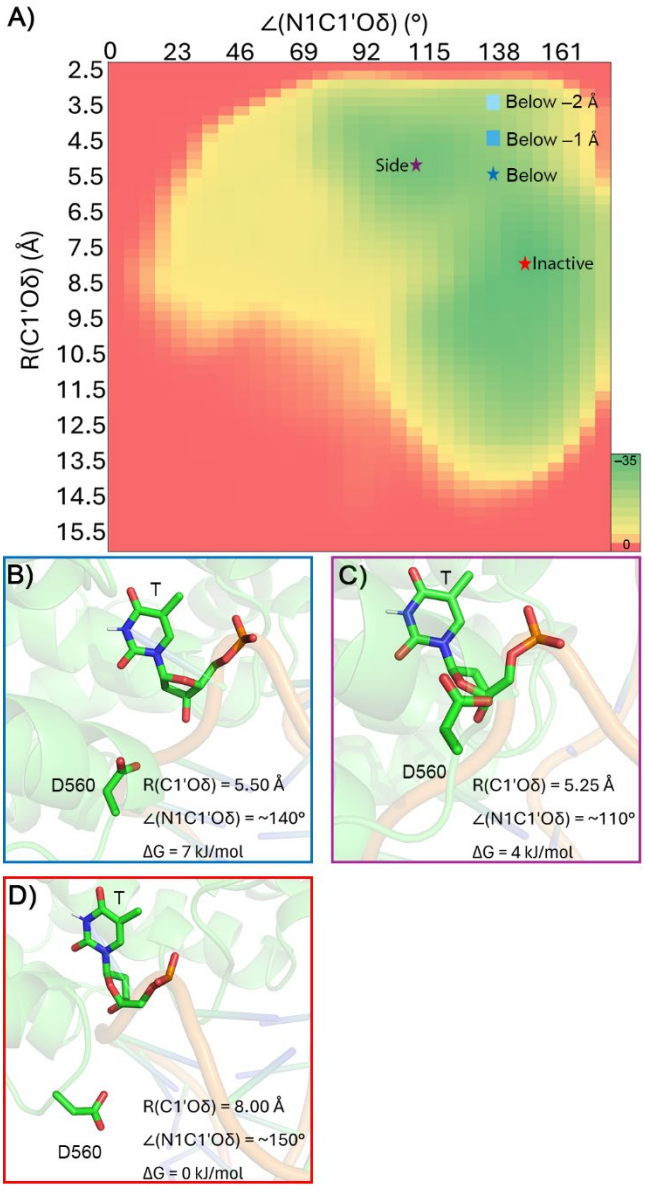


Figure 3.4. A) Average PES (in kJ/mol) generated from abMD simulations, highlighting the three minima (stars) and additional points (squares) used to generate starting structures for QM/MM calculations. Representative structures of the D560 B) below conformation (D560 positioned on the opposite side of deoxyribose as T), C) side conformation (D560 in the plane of the deoxyribose ring), and D) inactive conformation (D560 directed away from T).

conformation). D560 is held in place by transient hydrogen bonds to the backbones of L563 and K562 (< 35% occupancy; Figure B.4B, Appendix B). With O4′ positioned between D560 and C1′, the attack of D560 at C1′ required for crosslink formation is blocked in this conformation. However, D560 could possibly activate a water nucleophile for attack at C1′. Indeed, a water molecule is simultaneously < 3.5 Å from C1′ of T and Oδ of D560 for 64% of the simulation time (Figure 3.5A), suggesting D560 in this conformation could potentially act as a base to activate a water nucleophile. The third minimum (ΔG = 7 kJ/mol above the global minimum; ∠(N1C1′Oδ) = ~140°; r(C1′Oδ) = 5.50 Å; Figure 3.4B) places D560 on the opposite side of deoxyribose as the thymine nucleobase (denoted below conformation). D560 forms sporadic hydrogen bonds with the L562 and K563 backbones and H561 side chain (< 25% occupancy; Figure B.4C, Appendix B). This conformation most closely resembles the D560 orientation in the crystal structure of MBD4 bound to pseudouridine-containing DNA (PDB ID: 7KZ0, Figures 3.1B and B.2B, Appendix B), although D560 is slightly closer to the nonsubstrate pseudouridine in the crystal structure (∠(N1C1′Oδ) = ~140°; r(C1′Oδ) = 3.6 Å). Furthermore, water is present between D560 and C1′ for 50% of the simulation time (Figure 3.5B), suggesting D560 could potentially facilitate nucleotide hydrolysis. Nevertheless, ∠(N1C1′Oδ) = ~140° indicates that direct attack of D560 at C1′ to generate a crosslink is also viable.

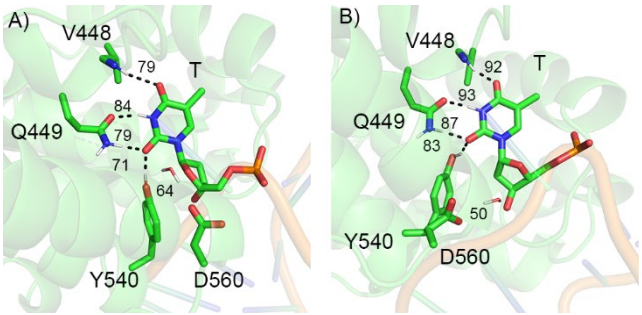


Figure 3.5. MBD4 active site showing hydrogen-bond and water nucleophile occupancies (percentage) from abMD simulations on the D560 A) side conformation (D560 in the plane of the deoxyribose ring), and B) below conformation (D560 positioned on the opposite side of deoxyribose as T).

Overall, the conformational flexibility of D560 characterized with abMD simulations suggests that more than one MBD4 catalytic pathway may be feasible, including a hydrolysis mechanism from either the side or below conformation and a crosslinking mechanism from the below conformation. To characterize the kinetics, thermodynamics, and product stereochemistry for each potential pathway and thereby determine the preferred MBD4 mechanism of action, QM/MM calculations were performed from representative structures corresponding to each conformation.

3.3.2. QM/MM calculations suggest that catalysis is not feasible when D560 is positioned in the plane of deoxyribose

The lowest energy, potentially catalytically active D560 orientation is the side conformation, which has a smaller ∠(N1C1′Oδ) and shorter r(C1′Oδ) compared to the other conformations (Figure 3.4). Although the small ∠(N1C1′Oδ) prevents direct nucleophilic attack of D560 at C1′ due to steric clashes, the short r(C1′Oδ) distance coupled with a water molecule adopting a position conducive for attack at C1′ with activation through proton abstraction by D560 for 64% of the simulation time could afford glycosidic bond hydrolysis from the side conformation (Figure 3.5A). The QM/MM characterized hydrolysis pathway from an abMD representative structure occurs with simultaneous glycosidic bond cleavage and nucleophilic water attack (Figures 3.6, B.5A, and B.6, Appendix B). In the QM/MM optimized RC, the water nucleophile is located 3.019 Å from C1′ and D560 hydrogen bonds with the nucleophile, K562 and L563, while the thymine base hydrogen bonds with V448, Q449, and Y540. The transition state is late, with the glycosidic bond being cleaved (r(C1′–N1) = 3.581 Å) and the nucleophilic water beginning attack (r(O_{WAT}–C1′) = 2.457 Å), but not yet donating a proton to D560. In the PC, the nucleophilic water has added to the nucleotide and transferred a proton to D560, forming an abasic site product with retained stereochemistry at C1′. The detached, anionic thymine is stabilized by strengthened hydrogen bonds with V448, Q449, and Y540. The calculated energy barrier is 148.8 kJ/mol, which is significantly higher than the experimental barrier for MBD4 (92.3 kJ/mol).[40] Additionally, the abasic site product contrasts with the inversion of stereochemistry anticipated based on NMR and methanolysis experiments.[40] Thus, although abMD

simulations suggest the active site conformation with D560 positioned in the same plane as the deoxyribose ring is energetically accessible, this configuration is not conducive for MBD4 activity.

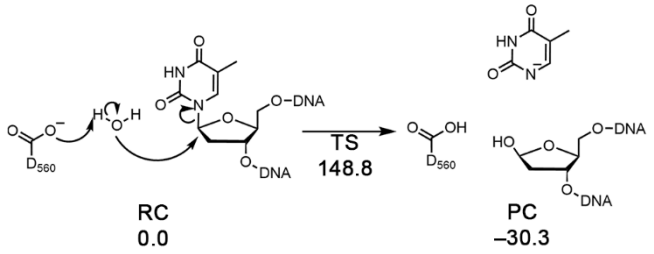


Figure 3.6. MBD4 glycosidic bond hydrolysis mechanism characterized from the D560 side conformation (D560 in the plane of the deoxyribose ring). Relative energies (below) in kJ/mol.

3.3.3. MBD4 does not form a DNA–protein crosslink as part of the catalytic mechanism

Due to an observed crosslink in the crystal structure of mouse MBD4 (PDB ID: 4EW4, Figure 3.1A),[16] the first proposed MBD4 mechanism involves a DNA–protein crosslinked intermediate. The formation of a similar crosslinked species as part of DNA glycosylase activity has been reported for another enzyme, namely MutY,[41-43] while a crosslink has also been proposed as a product (rather than an intermediate) for UdgX.[44] Furthermore, the retaining O-glycosidases consistently generate a crosslinked intermediate as a part of their glycosidic bond cleaving activity,[45, 46] which reinforces crosslink formation as a viable chemical step in DNA glycosylase function. Notably, the below conformation positions D560 on the opposite side of deoxyribose as the thymine base, with a ∠(N1C1′Oδ) that appropriately aligns D560 for attack at C1′ (Figure 3.4), reinforcing the possibility of a crosslinking mechanism for MBD4.

To determine the feasibility of DNA–MBD4 crosslink formation, a QM/MM model was built based on a representative structure of the below conformational minimum on the abMD PES (Figures 3.5B, 3.7A, B.5B, and B.7A, Appendix B). In the QM/MM optimized RC, D560 is located within 4.373 Å of C1′. A late transition state is characterized, with the D560–C1′ bond nearly formed (r(C1′Oδ) = 2.383 Å) and the C1′–N1 bond partially cleaved (r(C1′–N1) = 2.684 Å). This transition state results in a high barrier of 162.1 kJ/mol, more than 10 kJ/mol above that observed for the

QM/MM hydrolysis pathway from the D560 side conformation and significantly higher than the experimental barrier (92.3 kJ/mol).[40] Although the C1′–N9 bond is fully cleaved and the crosslink forms as the reaction proceeds, the crosslinked intermediate is also high in energy (72.1 kJ/mol above the RC). The large calculated barrier and endergonic nature of this crosslinking reaction suggest this pathway is not viable.

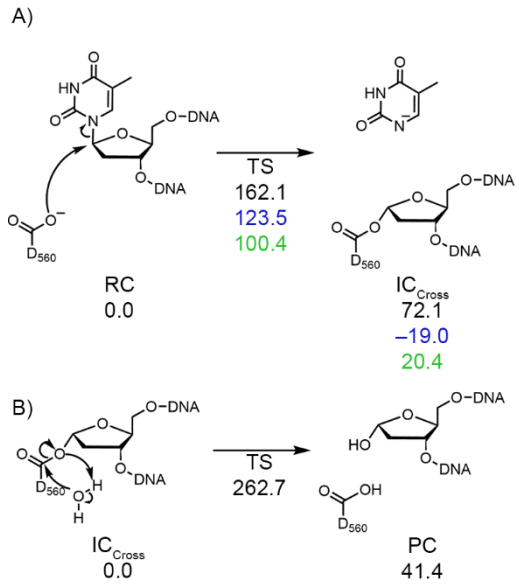


Figure 3.7. A) MBD4 crosslink formation mechanism characterized from the D560 below conformation (D560 on the opposite side of deoxyribose as T). B) MBD4 saponification mechanism for crosslink cleavage. Relative energies (kJ/mol, below) correspond to the minimum (black), 1 Å closer (blue), and 2 Å closer (green) starting points.

It is interesting to note that QM/MM optimization of the RC brings D560 ~1 Å closer to the substrate compared to the abMD representative structure, suggesting a starting structure with an even smaller r(C1′Oδ) could afford more favorable reaction energetics for crosslink formation. Indeed, the corresponding distance in the crystal structure of MBD4 bound to pseudouridine-containing DNA (3.6 Å)[40] is even smaller than the QM/MM optimized value. Furthermore, the abMD PES is shallow near the below conformation, with a structure containing a 1 or 2 Å shorter r(C1′Oδ) falling only 3 or 8 kJ/mol above the minimum, respectively. Since such active site conformational changes would be energetically accessible within biological timeframes, two new QM/MM models

were generated based on abMD representative structures with r(C1′Oδ) 1 Å or 2 Å shorter than the minimum of the below conformation to explore the effects of D560 positioning on the crosslinking mechanism (Figure 3.4A**,** squares).

When the QM/MM starting point was taken from the point on the abMD PES with r(C1′Oδ) 1 Å shorter than the below conformational minimum, a similar reaction pathway is mapped as from the minimum (Figures 3.7, B.5C, and B.7A, Appendix B). Although the resulting crosslinked intermediate is exergonic (–19.0 kJ/mol), the predicted barrier (123.5 kJ/mol) is still larger than the experimental estimate (92.3 kJ/mol).[40] However, when a QM/MM RC is generated from the abMD structure with a 2 Å shorter r(C1′Oδ) compared to the PES minimum, a feasible crosslink formation mechanism was characterized (Figures 3.7A, B.5D, and B.7A, Appendix B**).** Indeed, in agreement with crystallographic data,[40] D560 falls 3.691 Å from C1′ in the QM/MM RC. Furthermore, the transition state results in the lowest barrier among all models considered thus far (100.4 kJ/mol) and is consistent with the experimental value (92.3 kJ/mol).[40] Nevertheless, while the reaction is kinetically feasible, crosslink formation is slightly endergonic (crosslinked intermediate falls 20.4 kJ/mol above the RC).

Following the formation of the crosslinked intermediate, the enzyme must be regenerated to yield an abasic site. Prior to mapping this reaction step, MD simulations were conducted on the crosslinked intermediate. MD simulations reveal a single crosslink conformation (∠(C2′C1′OδCγ) ~200°, Figure B.8, Appendix B), which is consistent with the orientation in the crystal structure of the MBD4 crosslinked intermediate (Figure B.2A, Appendix B). As experiments determined inversion of stereochemistry in the abasic site,[40] nucleophilic attack of water at C1′ to displace D560 will not yield the correct product. Instead, water nucleophilic attack at Cγ through a saponification reaction is required to regenerate the enzyme. This newly proposed pathway is supported by a water density of ~0.8 at a distance of 3 Å from Cγ of D560 across the MD simulations on the crosslinked intermediate (Figure B.9, Appendix B). In the QM/MM optimized RC (Figure 3.7B and B.7B, Appendix B), the water nucleophile is positioned to attack Cγ (r(O_{WAT}–Cγ) = 3.375 Å). The transition state for saponification is very late, with water having almost finished attack (r(O_{WAT}– Cγ)

= 1.595 Å), the Cγ–Oδ bond being extended (r(Cγ –Oδ) = 1.701 Å), and the proton on the water mostly transferred to Oδ of D560. Although the correct abasic site product is formed, the barrier is extremely high (262.7 kJ/mol), likely due to the lack of a nearby active site residue to activate the water nucleophile and stabilize the transition state. Thus, although similar enzymes (MutY,[42] O-glycosidases[45, 46]) use enzyme–substrate crosslinking mechanisms and initial crosslink formation is kinetically accessible for MBD4, the thermodynamically disfavored crosslinked intermediate and kinetically prohibited hydrolysis to generate the final abasic site product suggests that MBD4 does not utilize a crosslinking deglycosylation mechanism.

3.3.4. Preferred MBD4 mechanism of action involves direct hydrolysis of the thymidine glycosidic bond

With the MBD4 crosslinking mechanism being unfeasible, a catalytic pathway in which an activated water molecule directly hydrolyzes the glycosidic bond must be reconsidered. Indeed, a similar mechanism has been successfully characterized for several other monofunctional glycosylases (e.g., hUNG2,[51] AAG,[54] and TDG[72]), which contain an amino acid that activates the water nucleophile that is similarly positioned with respect to the DNA substrate as D560 in the crystal structure of MBD4 bound to pseudouridine-containing DNA (Figure 3.1B).[40] While the below conformation on the abMD PES places D560 on the opposite side of deoxyribose as T for direct attack at C1′, a water molecule also simultaneously exists within 3.5 Å of D560 and C1′ for ~50% of the simulation time (Figure 3.5B). Furthermore, the below active site conformation isolated with abMD simulations closely resembles the crystal structure of MBD4-bound to pseudouridine–containing DNA (Figure B.2B, Appendix B). Therefore, a hydrolysis mechanism was considered using the same three QM/MM models explored for the crosslinking mechanism (i.e., the below minimum, and models with a 1 Å or 2 Å smaller r(C1′Oδ)).

In the QM/MM optimized RC associated with the below minimum on the abMD PES (Figure 3.8A, B.5B, and B.10A, Appendix B), a water nucleophile is positioned to attack C1′ ($r(O_{WAT}–C1')$ = 2.815 Å), while forming a hydrogen bond with D560 ($r(D560–H_{WAT})$ = 1.729 Å). In the transition

state, the water initiates attack at C1′ (r(O_{WAT}–C1′) = 2.129 Å), while the C1′–N1 bond starts to break (r(C1′–N1) = 2.211 Å), but proton transfer to D560 has not yet occurred. The associated barrier is 128.6 kJ/mol, which is significantly larger than the experimental barrier (92.3 kJ/mol)[40] and suggests that catalysis by direct hydrolysis from this starting conformation is not feasible.

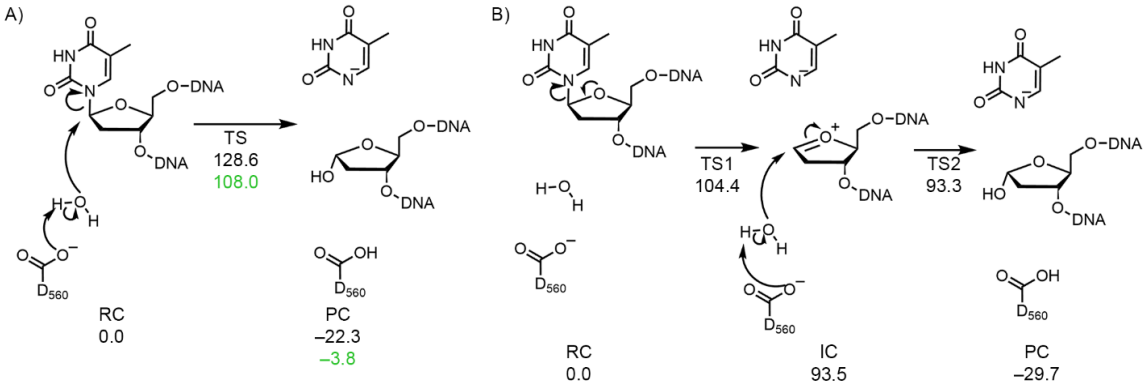


Figure 3.8. MBD4 direct hydrolysis mechanism characterized from the D560 below conformation (D560 on the opposite side of deoxyribose as T) using A) the minimum on the abMD PES (black) and the point with D560 2 Å (green) or B) 1 Å closer to C1′ of the substrate. Relative energies (below) in kJ/mol.

Notably, the distance between D560 and C1′ is large in the QM/MM RC (4.801 Å) and transition state (4.179 Å), indicating a starting structure with D560 positioned closer to C1′ could result in a more feasible pathway. Indeed, the QM/MM model built from the ab MD structure with a 2 Å smaller r(C1′Oδ) compared to the below minimum results in a similar pathway (Figures 3.8A, 3.5D, and B.10A, Appendix B). However, D560 and the water nucleophile fall ~0.7 and 0.3 Å closer to C1′ in the RC, resulting in a 0.5 Å longer C1′–N1 distance in the TS. Furthermore, D560 falls 0.6 Å closer to C1′ in the TS for this pathway than that mapped from the abMD PES below conformational minimum, increasing electrostatic stabilization within the active site. In the slightly exergonic PC, the hydroxy group has added to C1′, thymine has dissociated from the sugar, and D560 has abstracted a proton from the water nucleophile. The corresponding barrier is 108.0 kJ/mol, which is > 20 kJ/mol lower than the previously mapped mechanism.

To further explore the impact of the distance between D560 and the substrate on the feasibility of nucleotide hydrolysis catalyzed by MBD4, QM/MM calculations were performed on a starting

structure with r(C1′Oδ) decreased by 1 Å compared to the below conformational minimum (Figures 3.8B, B.5C and B.10B, Appendix B). QM/MM optimizations reveal the reaction involves an oxocarbenium ion intermediate. Specifically, in the first transition, the C1′–N1 bond is partially broken (r(C1′–N1) = 2.382 Å), while both the nucleophile water (r(O_{WAT}–C1′) = 2.713 Å) and D560 (r(C1′Oδ) = 3.497 Å) move closer to C1′, which provides charge stabilization and results in a barrier of 104.4 kJ/mol. This TS is connected to a high energy (93.5 kJ/mol) intermediate that contains a broken C1′–N1 bond (2.508 Å), which undergoes a barrierless transition to the PC. The rate-limiting barrier for the 1 Å model (104.4 kJ/mol) is very close to that for the 2 Å model (108.0 kJ/mol).

Overall, while direct nucleotide hydrolysis mapped from the below conformational minimum on the abMD PES results in a high barrier, the mechanisms characterized from starting points with 1 Å and 2 Å shorter r(C1′Oδ) are energetically feasible (barriers ~104–108 kJ/mol) and correlate with the experimentally estimated barrier (92.3 kJ/mol),[40] suggesting that catalysis is plausible when r(C1′Oδ) is below a threshold (~4.2 Å according to QM/MM).

3.4. Discussion

3.4.1. Differences in active site residues and leaving group charge result in divergent catalytic pathways for MBD4 and MutY

MBD4 and MutY are monofunctional glycosylases that initiate the excision of mispaired canonical nucleotides using comparable active site architectures. Indeed, crystal structures of each enzyme bound to pseudouridine-containing DNA highlight an aspartate similarly positioned with respect to the substrate (D560 (D534) in human (mouse) MBD4 and D144 in MutY, Figures 3.1B and 3.9A).[40, 102] These aspartate residues have been implicated in catalysis, namely activation of active site water for nucleotide hydrolysis,[40, 75] a proposal supported by the D534N mouse MBD4 and D144N MutY mutants being catalytically inactive.[16, 103] However, a second crystal structure of each enzyme depicts the catalytic aspartate in close proximity to C1′ of deoxyribose or a deoxyribose analogue (Figures 3.1A and 3.9B), leading to the proposed formation of a DNA−protein

crosslink.[16, 104] Therefore, it is interesting to explore whether MBD4 and MutY proceed through the same catalytic mechanism.

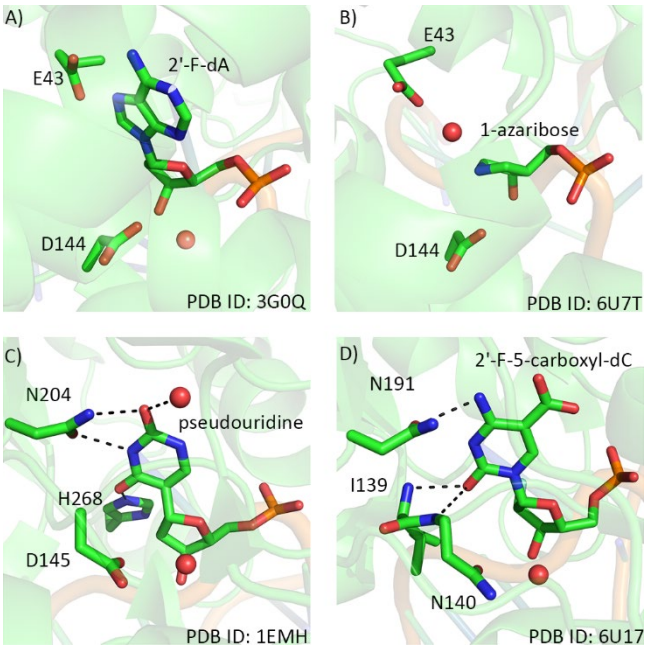


Figure 3.9. Active site from crystal structures of A) MutY bound to 2'-*β*-fluoro-2'-deoxyadenosine, B) MutY bound to a transition state analogue, C) hUNG2 bound to pseudouridine, and D) TDG bound to 2'-fluoro-5-carboxyl-2'-deoxycytidine.

The MutY mechanism of action has been extensively studied with both experimental[41, 43, 102, 105, 106] and computational techniques.[42, 75-80] A combination of experimental crystallographic, NMR, and methanolysis data combined with MD and QM/MM calculations[42, 76] conclusively determined that MutY activity relies on the formation of a DNA−protein crosslinked intermediate. In the MutY mechanism of action, the targeted A is protonated (at N7) by E43 followed by D144 attack at C1′ to form a DNA–protein crosslink. In the second reaction step, E43 activates a water nucleophile that hydrolyzes the crosslink to yield an abasic site with inversion of stereochemistry and regenerate the enzyme. In the case of MBD4, our abMD simulations on the enzyme–substrate complex suggest that D560 can adopt an orientation to promote in-line attack at C1′ (Figure 3.4). Furthermore, QM/MM calculations suggest crosslink formation is kinetically accessible (Figure 3.7) and classical MD simulations reveal the crosslinked intermediate adopts the same dominant

conformation as previously reported for MutY[42] (Figure B.8, Appendix B). However, the crosslinked intermediate is thermodynamically unstable and the barrier for crosslink hydrolysis to generate the reaction product is energetically prohibitive (262.7 kJ/mol), suggesting a crosslinking mechanism is not viable for MBD4.

Two key differences in the crosslinking mechanisms for MBD4 and MutY account for the observed variations in the feasibility of this novel catalytic mechanism for a monofunctional glycosylase. First, an active site residue is available to facilitate MutY but not MBD4 crosslink hydrolysis. Specifically, E43 of MutY activates the water molecule that displaces the crosslinked aspartate.[42] However, since no active site residue can fulfill this role for MBD4, water simultaneously undergoes nucleophilic attack and donates a proton to the leaving group, rationalizing the much higher barrier (by > 150 kJ/mol). Second, the nucleobase leaving group associated with MutY activity is neutral due to A protonation by E43. In contrast, although the leaving group for deglycosylation is stabilized by hydrogen bonds to V448, Q449, and Y540, the departing T in the MBD4 mechanism maintains anionic charge. Therefore, the high-energy crosslinked intermediate would likely be attacked at C1′ by the anionic T prior to crosslink hydrolysis, regenerating the reactant complex. Although the infeasibility of crosslink formation contrasts the crystal structure of the MBD4 crosslinked intermediate (Figure 3.1A),[16] the low resolution of the structure (2.79 Å) may have resulted in D534 being erroneously modeled close to deoxyribose, which is supported by the relatively low electron density between Oδ of D534 and C1′ (Figure B.11, Appendix B).

Overall, while crosslink formation is the preferred mechanism of action for MutY,[41-43, 76] MBD4 activity is unlikely to involve a DNA−protein crosslink due to the lack of active site residues to facilitate crosslink hydrolysis and the more nucleophilic (anionic) leaving group. Thus, MutY remains unique among the monofunctional DNA glycosylase family, relying on DNA–enzyme crosslink formation along the mechanism of action, while generating a final abasic site product with retention of stereochemistry.[42, 76]

3.4.2. MBD4 catalyzes the repair of T:G mismatches through a hydrolysis mechanism similar to that used by other monofunctional DNA glycosylases to target diverse damaged nucleotides

This work provides the first atomic level description of the conformational dynamics of the MBD4 active site when bound to the damaged DNA substrate and the first computational prediction of the MBD4 catalytic mechanism of action. Specifically, abMD simulations reveal the enzyme active site is highly dynamic, with D560 readily sampling a large region of conformational space (Figure 3.4). Although some energetically accessible conformations are not conducive for catalysis, several catalytically-active conformations are thermodynamically and kinetically feasible. QM/MM calculations initiated from different points on the abMD PES to account for active site dynamics characterized seven different reaction pathways. The preferred MBD4 mechanism of action involves direct hydrolysis of the T nucleotide (Figure 3.8 and B.10, Appendix B**)**. In this reaction pathway, D560 activates a water nucleophile, which attacks C1′ of T while cleaving the glycosidic bond to form an abasic site product with inversion of stereochemistry at deoxyribose. In addition to initiating the reaction, D560 provides electrostatic stabilization to the charge build up in the transition state and oxocarbenium intermediate. Consequently, catalysis is only viable when D560 falls below a threshold distance from C1′ of the target nucleotide, which QM/MM calculations suggest is < ~4.2 Å. Although anionic T is a poor leaving group (pKa ~ 9.9),[107] base departure is facilitated by strengthening of an abundance of hydrogen bonds between O2 and O4 of T and V448, Q449, and Y540 (i.e., hydrogen-bonding distances decrease by ~0.2 Å; Figure B.10, Appendix B). These factors result in a computationally-predicted mechanism that is fully consistent with experimental literature, including the reaction rate,[40] retention of stereochemistry in the product,[40] reduced activity of D560 mutants,[16, 40] and crystallographic data.[40]

The proposed MBD4 catalytic pathway, including the roles of key active site residues, mirrors the accepted mechanisms of action of other monofunctional glycosylases known to directly hydrolyze glycosidic bonds with inversion of C1′ stereochemistry, including enzymes that target damaged pyrimidines such as UDG[14] and TDG.[72] For example, D560 and an active site water in

the crystal structure of MBD4 bound to pseudouridine-containing DNA are similarly aligned with respect to the substrate as D145 and water in hUNG2 (PDB IDs: 7KZ0 and 1EMH, Figures 3.1B and 3.9C). In parallel to MBD4, experimental[108, 109] and computational[51] data suggest hUNG2 removes uracil (U) formed in DNA through deamination of C by D145 activating a water molecule for nucleophilic attack at C1′ of the substrate, which results in an abasic site with inversion of stereochemistry.[51] MBD4 and hUNG2 also use similar strategies to stabilize the departing anionic base, with N204, (neutral) H268, and a water molecule forming hydrogen bonds with O2 and O4 of U in the hUNG2 active site.[47, 51] TDG, which targets a variety of U, C, and T adducts and mispairs,[110] also has a similar active site architecture to MBD4 (PDB IDs: 7KZ0 and 6U17, Figures 3.1B and 3.9D). In the crystal structure of TDG bound to 5-carboxylcytosine (5-caC) containing DNA, N140 is positioned as per aspartate in MBD4 and hUNG2 and computational work suggests N140 abstracts a proton from a water nucleophile during catalysis, which is facilitated by charge delocalization across several residues.[72] Furthermore, N191, I139, and N140 form hydrogen bonds with O2 and N4 of 5-caC in the TDG active site to facilitate leaving group departure,[72, 111] highlighting a similar strategy for stabilization of the anionic pyrimidine leaving group as MBD4. Thus, by carefully considering different catalytic pathways for MBD4 and eliminating enzyme–DNA crosslink formation, our computational work unifies the MBD4 catalytic mechanism with strategies employed by other monofunctional glycosylases.

3.5. Conclusion

The present study uses a combination of classical MD, abMD, and QM/MM calculations to characterize seven different potential pathways and propose the preferred MBD4 mechanism of action. Our calculations elucidated a direct hydrolysis reaction where D560 activates a water nucleophile that attacks at C1′, which is facilitated through charge stabilization provided by D560 and several amino acids that hydrogen bond to the departing (anionic) T. This proposal is fully consistent with previous experimental mutagenic,[16] kinetic,[40] crystallographic,[40] and stereoscopic[40] data, clarifies conflicting literature regarding MBD4 forming a DNA−protein crosslinked

intermediate,[16, 40] and relates MBD4 activity with that of other monofunctional DNA glycosylases.[14, 52, 53, 72] The improved understanding of the MBD4 catalytic pathway provided by our work promises to accelerate research into understanding and treating MANS.[19, 20] Additionally, the atomic level description of transient intermediates in the MBD4 mechanism revealed by this work can provide the basis for the development of transition-state mimic inhibitors, which could have applications in cancer treatments,[21-28] especially cancers resistant to first-line therapeutics (5-FU).[36, 37]

3.6. References

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

(2) Berson, A.; Nativio, R.; Berger, S. L.; Bonini, N. M. Epigenetic regulation in neurodegenerative diseases. *Trends Neurosci.* **2018**, *41* (9), 587-598. DOI: https://doi.org/10.1016/j.tins.2018.05.005

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

(4) Singer, B. D. A practical guide to the measurement and analysis of DNA methylation. *Am. J. Respir. Cell Mol. Biol.* **2019**, *61* (4), 417-428. DOI: 10.1165/rcmb.2019-0150TR

(5) Antequera, F. Structure, function and evolution of CpG island promoters. *Cell. Mol. Life Sci.* **2003**, *60* (8), 1647-1658. DOI: 10.1007/s00018-003-3088-6

(6) Li, E.; Beard, C.; Jaenisch, R. Role for DNA methylation in genomic imprinting. *Nature* **1993**, *366* (6453), 362-365. DOI: 10.1038/366362a0

(7) Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev* **2002**, *16* (1), 6-21. DOI: 10.1101/gad.947102

(8) Moore, L. D.; Le, T.; Fan, G. DNA methylation and its basic function. *Neuropsychopharmacology* **2013**, *38* (1), 23-38. DOI: 10.1038/npp.2012.112

(9) Mohn, F.; Weber, M.; Rebhan, M.; Roloff, T. C.; Richter, J.; Stadler, M. B.; Bibel, M.; Schübeler, D. Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol. Cell* **2008**, *30* (6), 755-766. DOI: https://doi.org/10.1016/j.molcel.2008.05.007

(10) Prorok, P.; Alili, D.; Saint-Pierre, C.; Gasparutto, D.; Zharkov, D. O.; Ishchenko, A. A.; Tudek, B.; Saparbaev, M. K. Uracil in duplex DNA is a substrate for the nucleotide incision repair pathway in human cells. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (39), E3695-E3703,SE3695/3691-SE3695/3693. DOI: 10.1073/pnas.1305624110

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

(12) Zharkov, D. O. Base excision DNA repair. *Cell. Mol. Life Sci.* **2008**, *65* (10), 1544-1565. DOI: 10.1007/s00018-008-7543-2

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

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

(15) Cortellino, S.; Turner, D.; Masciullo, V.; Schepis, F.; Albino, D.; Daniel, R.; Skalka, A. M.; Meropol, N. J.; Alberti, C.; Larue, L.; et al. The base excision repair enzyme MED1 mediates DNA damage response to antitumor drugs and is associated with mismatch repair system integrity. *Proc. Natl. Acad. Sci. U.S.A* **2003**, *100* (25), 15071-15076. DOI: doi:10.1073/pnas.2334585100

(16) Hashimoto, H.; Zhang, X.; Cheng, X. Excision of thymine and 5-hydroxymethyluracil by the MBD4 DNA glycosylase domain: Structural basis and implications for active DNA demethylation. *Nucleic Acids Res.* **2012**, *40* (17), 8276-8284. DOI: 10.1093/nar/gks628 (acccessed 1/14/2025).

(17) Turner, D. P.; Cortellino, S.; Schupp, J. E.; Caretti, E.; Loh, T.; Kinsella, T. J.; Bellacosa, A. The DNA N-glycosylase MED1 exhibits preference for halogenated pyrimidines and is involved in the cytotoxicity of 5-iododeoxyuridine. *Cancer Res.* **2006**, *66* (15), 7686-7693. DOI: 10.1158/0008-5472.Can-05-4488 (acccessed 1/17/2025).

(18) Millar, C. B.; Guy, J.; Sansom, O. J.; Selfridge, J.; MacDougall, E.; Hendrich, B.; Keightley, P. D.; Bishop, S. M.; Clarke, A. R.; Bird, A. Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. *Science* **2002**, *297* (5580), 403-405. DOI: doi:10.1126/science.1073354

(19) Palles, C.; West, H. D.; Chew, E.; Galavotti, S.; Flensburg, C.; Grolleman, J. E.; Jansen, E. A. M.; Curley, H.; Chegwidden, L.; Arbe-Barnes, E. H.; et al. Germline MBD4 deficiency causes a multi-tumor predisposition syndrome. *Am. J. Hum. Genet.* **2022**, *109* (5), 953-960. DOI: https://doi.org/10.1016/j.ajhg.2022.03.018

(20) Blombery, P.; Ryland, G. L.; Fox, L. C.; Stark, Z.; Wall, M.; Jarmolowicz, A.; Roesley, A.; Thompson, E. R.; Grimmond, S. M.; Panicker, S.; et al. Methyl-CpG binding domain 4, DNA glycosylase (MBD4)-associated neoplasia syndrome associated with a homozygous missense variant in MBD4: Expansion of an emerging phenotype. *Br. J. Haematol.* **2022**, *198* (1), 196-199. DOI: https://doi.org/10.1111/bjh.18178

(21) Schrempf, A.; Slyskova, J.; Loizou, J. I. Targeting the DNA repair enzyme polymerase θ in cancer therapy. *Trends Cancer* **2021**, *7* (2), 98-111. DOI: https://doi.org/10.1016/j.trecan.2020.09.007

(22) Duskova, K.; Lejault, P.; Benchimol, É.; Guillot, R.; Britton, S.; Granzhan, A.; Monchaud, D. DNA junction ligands trigger DNA damage and are synthetic lethal with DNA repair inhibitors in cancer cells. *J. Am. Chem. Soc.* **2020**, *142* (1), 424-435. DOI: 10.1021/jacs.9b11150

(23) Fuso Nerini, I.; Roca, E.; Mannarino, L.; Grosso, F.; Frapolli, R.; D'Incalci, M. Is DNA repair a potential target for effective therapies against malignant mesothelioma? *Cancer Treat. Rev.* **2020**, *90*, 102101. DOI: https://doi.org/10.1016/j.ctrv.2020.102101

(24) Mechetin, G. V.; Endutkin, A. V.; Diatlova, E. A.; Zharkov, D. O. Inhibitors of DNA glycosylases as prospective drugs. *Int. J. Mol. Sci.* **2020**, *21* (9), 3118.

(25) Jiang, M.; Jia, K.; Wang, L.; Li, W.; Chen, B.; Liu, Y.; Wang, H.; Zhao, S.; He, Y.; Zhou, C. Alterations of DNA damage repair in cancer: From mechanisms to applications. *Ann. Transl. Med.* **2020**, *8* (24), 1685.

(26) Gad, H.; Koolmeister, T.; Jemth, A.-S.; Eshtad, S.; Jacques, S. A.; Ström, C. E.; Svensson, L. M.; Schultz, N.; Lundbäck, T.; Einarsdottir, B. O.; et al. MTH1 inhibition eradicates cancer by preventing sanitation of the dNTP pool. *Nature* **2014**, *508* (7495), 215-221. DOI: 10.1038/nature13181

(27) Huber, K. V. M.; Salah, E.; Radic, B.; Gridling, M.; Elkins, J. M.; Stukalov, A.; Jemth, A.-S.; Göktürk, C.; Sanjiv, K.; Strömberg, K.; et al. Stereospecific targeting of MTH1 by (S)-crizotinib as an anticancer strategy. *Nature* **2014**, *508* (7495), 222-227. DOI: 10.1038/nature13194

(28) Helleday, T.; Petermann, E.; Lundin, C.; Hodgson, B.; Sharma, R. A. DNA repair pathways as targets for cancer therapy. *Nat. Rev. Cancer* **2008**, *8* (3), 193-204. DOI: 10.1038/nrc2342

(29) Wallner, O.; Cázares-Körner, A.; Scaletti, E. R.; Masuyer, G.; Bekkhus, T.; Visnes, T.; Mamonov, K.; Ortis, F.; Lundbäck, T.; Volkova, M.; et al. Optimization of N-piperidinyl-benzimidazolone derivatives as potent and selective inhibitors of 8-oxo-guanine DNA glycosylase 1. *ChemMedChem* **2023**, *18* (1), e202200310. DOI: https://doi.org/10.1002/cmdc.202200310

(30) Weilbeer, C.; Jay, D.; Donnelly, J. C.; Gentile, F.; Karimi-Busheri, F.; Yang, X.; Mani, R. S.; Yu, Y.; Elmenoufy, A. H.; Barakat, K. H.; et al. Modulation of ERCC1-XPF heterodimerization inhibition via structural modification of small molecule inhibitor side-chains. *Front. Oncol.* **2022**, *12*. DOI: 10.3389/fonc.2022.819172

(31) Yang, F.; Hu, Z.; Guo, Z. Small-molecule inhibitors targeting FEN1 for cancer therapy. *Biomolecules* **2022**, *12* (7), 1007.

(32) Visnes, T.; Cázares-Körner, A.; Hao, W.; Wallner, O.; Masuyer, G.; Loseva, O.; Mortusewicz, O.; Wiita, E.; Sarno, A.; Manoilov, A.; et al. Small-molecule inhibitor of OGG1 suppresses proinflammatory gene expression and inflammation. *Science* **2018**, *362* (6416), 834-839. DOI: doi:10.1126/science.aar8048

(33) Donley, N.; Jaruga, P.; Coskun, E.; Dizdaroglu, M.; McCullough, A. K.; Lloyd, R. S. Small molecule inhibitors of 8-oxoguanine DNA glycosylase-1 (OGG1). *ACS Chem. Biol.* **2015**, *10* (10), 2334-2343. DOI: 10.1021/acschembio.5b00452

(34) Bliksrud, Y. T.; Ellingsen, A.; Bjørås, M. Fumarylacetoacetate inhibits the initial step of the base excision repair pathway: Implication for the pathogenesis of tyrosinemia type I. *J. Inherit. Metab. Dis.* **2013**, *36* (5), 773-778. DOI: https://doi.org/10.1007/s10545-012-9556-0

(35) Anna, K.; Jerzy, Z. B.; Tadeusz, R. Forodesine (BCX-1777, Immucillin H) - A new purine nucleoside analogue: Mechanism of action and potential clinical application. *Mini-Rev. Med. Chem.* **2007**, *7* (9), 976-983. DOI: http://dx.doi.org/10.2174/138955707781662636

(36) Suzuki, S.; Iwaizumi, M.; Tseng-Rogenski, S.; Hamaya, Y.; Miyajima, H.; Kanaoka, S.; Sugimoto, K.; Carethers, J. M. Production of truncated MBD4 protein by frameshift mutation in DNA mismatch repair-deficient cells enhances 5-fluorouracil sensitivity that is independent of hMLH1 status. *Cancer Biol. Ther.* **2016**, *17* (7), 760-768. DOI: 10.1080/15384047.2016.1178430

(37) Averill, J. R.; Lin, J. C.; Jung, J.; Jung, H. Novel insights into the role of translesion synthesis polymerase in DNA incorporation and bypass of 5-fluorouracil in colorectal cancer. *Nucleic Acids Res* **2024**, *52* (8), 4295-4312. DOI: 10.1093/nar/gkae102

(38) Schramm, V. L. Transition states, analogues, and drug development. *ACS Chem. Biol.* **2013**, *8* (1), 71-81. DOI: 10.1021/cb300631k

(39) Pushing cancer over the edge. *Chemical & Engineering News Archive* **2013**, *91* (24), 13-18. DOI: 10.1021/cen-09124-cover

(40) Pidugu, L. S.; Bright, H.; Lin, W.-J.; Majumdar, C.; Van Ostrand, R. P.; David, S. S.; Pozharski, E.; Drohat, A. C. Structural insights into the mechanism of base excision by MBD4. *J. Mol. Biol.* **2021**, *433* (15), 167097. DOI: https://doi.org/10.1016/j.jmb.2021.167097

(41) Woods, R. D.; O'Shea, V. L.; Chu, A.; Cao, S.; Richards, J. L.; Horvath, M. P.; David, S. S. Structure and stereochemistry of the base excision repair glycosylase MutY reveal a mechanism similar to retaining glycosidases. *Nucleic Acids Res.* **2016**, *44* (2), 801-810. DOI: 10.1093/nar/gkv1469

(42) Nikkel, D. J.; Wetmore, S. D. Distinctive formation of a DNA–protein cross-link during the repair of DNA oxidative damage: Insights into human disease from MD simulations and QM/MM calculations. *J. Am. Chem. Soc.* **2023**, *145* (24), 13114-13125. DOI: 10.1021/jacs.3c01773

(43) Demir, M.; Russelburg, L. P.; Lin, W.-J.; Trasviña-Arenas, Carlos H.; Huang, B.; Yuen, Philip K.; Horvath, Martin P.; David, Sheila S. Structural snapshots of base excision by the cancer-associated variant MutY N146S reveal a retaining mechanism. *Nucleic Acids Res.* **2023**, *51* (3), 1034-1049. DOI: 10.1093/nar/gkac1246 (acccessed 7/24/2025).

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

(45) Davies, G. J.; Mackenzie, L.; Varrot, A.; Dauter, M.; Brzozowski, A. M.; Schülein, M.; Withers, S. G. Snapshots along an enzymatic reaction coordinate: Analysis of a retaining β-glycoside hydrolase. *Biochemistry* **1998**, *37* (34), 11707-11713. DOI: 10.1021/bi981315i

(46) Rovira, C.; Males, A.; Davies, G. J.; Williams, S. J. Mannosidase mechanism: At the intersection of conformation and catalysis. *Curr. Opin. Struct. Biol.* **2020**, *62*, 79-92. DOI: https://doi.org/10.1016/j.sbi.2019.11.008

(47) 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. U.S.A* **2000**, *97* (10), 5083-5088. DOI: doi:10.1073/pnas.97.10.5083

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

(49) Coey, C. T.; Malik, S. S.; Pidugu, L. S.; Varney, K. M.; Pozharski, E.; Drohat, A. C. Structural basis of damage recognition by thymine DNA glycosylase: Key roles for N-terminal residues. *Nucleic Acids Res.* **2016**, *44* (21), 10248-10258. DOI: 10.1093/nar/gkw768 (acccessed 8/25/2025).

(50) Adhikary, S.; Eichman, B. F. Analysis of substrate specificity of *schizosaccharomyces pombe* Mag1 alkylpurine DNA glycosylase. *EMBO Rep.* **2011**, *12* (12), 1286-1292. DOI: https://doi.org/10.1038/embor.2011.189

(51) Przybylski, J. L.; Wetmore, S. D. A QM/QM investigation of the hUNG2 reaction surface: The untold tale of a catalytic residue. *Biochemistry* **2011**, *50* (19), 4218-4227. DOI: 10.1021/bi2003394

(52) Stivers, J. T.; Jiang, Y. L. A mechanistic perspective on the chemistry of DNA repair glycosylases. *Chem. Rev.* **2003**, *103* (7), 2729-2759. DOI: 10.1021/cr010219b

(53) Rutledge, L. R.; Wetmore, S. D. Modeling the chemical step utilized by human alkyladenine DNA glycosylase: A concerted mechanism aids in selectively excising damaged purines. *J. Am. Chem. Soc.* **2011**, *133* (40), 16258-16269. DOI: 10.1021/ja207181c

(54) 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. DOI: 10.1021/acs.jpcb.7b09646

(55) Diao, W.; Peng, J.; Fu, X.; Ye, F.; Wang, Z. Preorganized internal electric field enables nucleophilic attack of a nonactivated water molecule in thymine DNA glycosylase. *J. Am. Chem. Soc.* **2025**, *147* (29), 25919-25930. DOI: 10.1021/jacs.5c08420

(56) Unno, M.; Morikawa, M.; Sychrovský, V.; Koga, M.; Minowa, N.; Komuro, S.; Shimizu, M.; Fukuta, M.; Tsuyuguchi, F.; Mano, H.; et al. Capturing a glycosylase reaction intermediate in DNA repair by freeze-trapping of a pH-responsive hOGG1 mutant. *Nucleic Acids Res.* **2025**, *53* (14). DOI: 10.1093/nar/gkaf718 (acccessed 8/25/2025).

(57) Blomberg, M. R.; Borowski, T.; Himo, F.; Liao, R.-Z.; Siegbahn, P. E. Quantum chemical studies of mechanisms for metalloenzymes. *Chem. Rev.* **2014**, *114* (7), 3601-3658. DOI: https://doi.org/10.1021/cr400388t

(58) Agbaglo, D. A.; Summers, T. J.; Cheng, Q.; DeYonker, N. J. The influence of model building schemes and molecular dynamics sampling on QM-cluster models: The chorismate mutase case study. *Phys. Chem. Chem. Phys.* **2024**, *26* (16), 12467-12482. DOI: 10.1039/D3CP06100K

(59) Sheng, X.; Himo, F. The quantum chemical cluster approach in biocatalysis. *Acc. Chem. Res.* **2023**, *56* (8), 938-947. DOI: 10.1021/acs.accounts.2c00795

(60) Elsässer, B.; Goettig, P. Mechanisms of proteolytic enzymes and their inhibition in QM/MM studies. *Int. J. Mol. Sci.* **2021**, *22* (6), 3232.

(61) Kaur, R.; Nikkel, D. J.; Wetmore, S. D. Computational studies of DNA repair: Insights into the function of monofunctional DNA glycosylases in the base excision repair pathway. *WIREs Comput. Mol. Sci.* **2020**, *10* (5), e1471. DOI: 10.1002/wcms.1471 (acccessed 2020/09/16).

(62) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Computational enzymatic catalysis – clarifying enzymatic mechanisms with the help of computers. *Phys. Chem. Chem. Phys.* **2012**, *14* (36), 12431-12441. DOI: 10.1039/C2CP41180F

(63) Gherib, R.; Dokainish, H. M.; Gauld, J. W. Multi-scale computational enzymology: Enhancing our understanding of enzymatic catalysis. *Int. J. Mol. Sci.* **2014**, *15* (1), 401-422. DOI: 10.3390/ijms15010401

(64) Aboelnga, M. M.; Wetmore, S. D. Unveiling a single-metal-mediated phosphodiester bond cleavage mechanism for nucleic acids: A multiscale computational investigation of a human DNA repair enzyme. *J. Am. Chem. Soc.* **2019**, *141* (21), 8646-8656. DOI: 10.1021/jacs.9b03986

(65) Wang, B.; Usharani, D.; Li, C.; Shaik, S. Theory uncovers an unusual mechanism of DNA repair of a lesioned adenine by AlkB enzymes. *J. Am. Chem. Soc.* **2014**, *136* (39), 13895-13901. DOI: 10.1021/ja507934g

(66) Kreppel, A.; Blank, I. D.; Ochsenfeld, C. Base-independent DNA base-excision repair of 8-oxoguanine. *J. Am. Chem. Soc.* **2018**, *140* (13), 4522-4526. DOI: 10.1021/jacs.7b11254

(67) Sadeghian, K.; Ochsenfeld, C. Unraveling the base excision repair mechanism of human DNA glycosylase. *J. Am. Chem. Soc.* **2015**, *137* (31), 9824-9831. DOI: 10.1021/jacs.5b01449

(68) Wang, S.-d.; Zhang, R.-b.; Eriksson, L. A. Markov state models elucidate the stability of DNA influenced by the chiral 5S-Tg base. *Nucleic Acids Res.* **2022**, *50* (16), 9072-9082. DOI: 10.1093/nar/gkac691 (acccessed 2/13/2025).

(69) Panigrahi, A.; Vemuri, H.; Aggarwal, M.; Pitta, K.; Krishnan, M. Sequence specificity, energetics and mechanism of mismatch recognition by DNA damage sensing protein rad4/XPC. *Nucleic Acids Res.* **2020**, *48* (5), 2246-2257. DOI: 10.1093/nar/gkaa078 (acccessed 2/13/2025).

(70) Naydenova, E.; Roßbach, S.; Ochsenfeld, C. QM/MM study of the uracil DNA glycosylase reaction mechanism: A competition between Asp145 and His148. *J. Chem. Theory Comput.* **2019**, *15* (8), 4344-4350. DOI: 10.1021/acs.jctc.8b01305

(71) Melayikandy, S.; Krishnan, A.; Varghese, A.; Jaber Sathik Rifayee, S. B.; Waheed, S. O.; Ramanan, R.; Li, D.; Christov, C. Z.; Karabencheva-Christova, T. G. Catalytic mechanism of the bacterial non-heme Fe(II) and 2-oxoglutarate dependent enzyme AlkB with single-stranded DNA containing complex guanine adducts. *Inorg. Chem.* **2025**, *64* (30), 15650-15666. DOI: 10.1021/acs.inorgchem.5c02176

(72) Wang, W.-J.; Wang, T.; Zhao, Y.; Li, B.-N.; Chen, D.-Z. Theoretical insights into N-glycoside bond cleavage of 5-carboxycytosine by thymine DNA glycosylase: A QM/MM study. *J. Phys. Chem. B* **2024**, *128* (19), 4621-4630. DOI: 10.1021/acs.jpcb.4c00221

(73) Nikkel, D. J.; Kaur, R.; Wetmore, S. D. How can one metal power nucleic acid phosphodiester bond cleavage by a nuclease? Multiscale computational studies highlight a diverse mechanistic landscape. *J. Phys. Chem. B* **2025**, *129* (1), 3-18. DOI: 10.1021/acs.jpcb.4c05875

(74) Nikkel, D. J.; Wetmore, S. D. Unlocking the chemistry facilitated by enzymes that process nucleic acids using quantum mechanical and combined quantum mechanics–molecular mechanics techniques. *Pure Appl. Chem.* **2025**, *97* (9), 1065-1089. DOI: doi:10.1515/pac-2025-0507 (acccessed 2025-10-14).

(75) 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* (48), 8753-8765. DOI: 10.1021/bi401310w

(76) Diao, W.; Farrell, J. D.; Wang, B.; Ye, F.; Wang, Z. Preorganized internal electric field promotes a double-displacement mechanism for the adenine excision reaction by adenine DNA glycosylase. *J. Phys. Chem. B* **2023**, *127* (40), 8551-8564. DOI: 10.1021/acs.jpcb.3c04928

(77) 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. DOI: 10.1021/ja711363s

(78) Tiwari, S.; Agnihotri, N.; Mishra, P. C. Quantum theoretical study of cleavage of the glycosidic bond of 2′-deoxyadenosine: Base excision-repair mechanism of DNA by MutY. *J. Phys. Chem. B* **2011**, *115* (12), 3200-3207. DOI: 10.1021/jp1109256

(79) 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. DOI: 10.1021/ja301714j

(80) Trasviña-Arenas, C. H.; Dissanayake, U. C.; Tamayo, N.; Hashemian, M.; Lin, W.-J.; Demir, M.; Hoyos-Gonzalez, N.; Fisher, A. J.; Cisneros, G. A.; Horvath, M. P.; et al. Structure of human MUTYH and functional profiling of cancer-associated variants reveal an allosteric network between its [4Fe-4S] cluster cofactor and active site required for DNA repair. *Nat. Commun.* **2025**, *16* (1), 3596. DOI: 10.1038/s41467-025-58361-w

(81) Kaur, R.; Wetmore, S. D. Is metal stabilization of the leaving group required or can lysine facilitate phosphodiester bond cleavage in nucleic acids? A computational study of EndoV. *J. Chem Inf. Model.* **2024**, *64* (3), 944-959. DOI: 10.1021/acs.jcim.3c01775

(82) Araújo, A. R.; Ribeiro, A. J. M.; Fernandes, P. A.; Ramos, M. J. Catalytic mechanism of retroviral integrase for the strand transfer reaction explored by QM/MM calculations. *J. Chem. Theory Comput.* **2014**, *10* (12), 5458-5466. DOI: 10.1021/ct500570g

(83) Rosta, E.; Nowotny, M.; Yang, W.; Hummer, G. Catalytic mechanism of RNA backbone cleavage by Ribonuclease H from quantum mechanics/molecular mechanics simulations. *J. Am. Chem. Soc.* **2011**, *133* (23), 8934-8941. DOI: 10.1021/ja200173a

(84) Elsässer, B.; Fels, G. Atomistic details of the associative phosphodiester cleavage in human ribonuclease H. *Phys. Chem. Chem. Phys.* **2010**, *12* (36), 11081-11088. DOI: 10.1039/C001097A

(85) Dürr, S. L.; Bohuszewicz, O.; Berta, D.; Suardiaz, R.; Jambrina, P. G.; Peter, C.; Shao, Y.; Rosta, E. The role of conserved residues in the DEDDh motif: The proton-transfer mechanism of HIV-1 RNase H. *ACS Catal.* **2021**, *11* (13), 7915-7927. DOI: 10.1021/acscatal.1c01493

(86) Palermo, G.; Stenta, M.; Cavalli, A.; Dal Peraro, M.; De Vivo, M. Molecular simulations highlight the role of metals in catalysis and inhibition of Type II Topoisomerase. *J. Chem. Theory Comput.* **2013**, *9* (2), 857-862. DOI: 10.1021/ct300691u

(87) Casalino, L.; Nierzwicki, Ł.; Jinek, M.; Palermo, G. Catalytic mechanism of non-target DNA cleavage in CRISPR-Cas9 revealed by *ab initio* molecular dynamics. *ACS Catal.* **2020**, *10* (22), 13596-13605. DOI: 10.1021/acscatal.0c03566

(88) De Vivo, M.; Dal Peraro, M.; Klein, M. L. Phosphodiester cleavage in Ribonuclease H occurs via an associative two-metal-aided catalytic mechanism. *J. Am. Chem. Soc.* **2008**, *130* (33), 10955-10962. DOI: 10.1021/ja8005786

(89) Mones, L.; Kulhánek, P.; Florián, J.; Simon, I.; Fuxreiter, M. Probing the two-metal ion mechanism in the restriction endonuclease BamHI. *Biochemistry* **2007**, *46* (50), 14514-14523. DOI: 10.1021/bi701630s

(90) Kaur, R.; Nikkel, D. J.; Wetmore, S. D. Mechanism of nucleic acid phosphodiester bond cleavage by human Endonuclease V: MD and QM/MM calculations reveal a versatile metal dependence. *J. Phys. Chem. B* **2024**, *128* (39), 9455-9469. DOI: 10.1021/acs.jpcb.4c05846

(91) Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; et al. The ONIOM method and its applications. *Chem. Rev.* **2015**, *115* (12), 5678-5796. DOI: 10.1021/cr5004419

(92) Schrodinger, LLC. The PyMOL molecular graphics system, version 1.8. 2015.

(93) 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* (W1), W537-W541. DOI: 10.1093/nar/gks375 (acccessed 8/9/2023).

(94) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; III, T. E. C.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; et al. AMBER 2018. **2018**, University of California, San Fransisco.

(95) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696-3713. DOI: 10.1021/acs.jctc.5b00255

(96) Galindo-Murillo, R.; Robertson, J. C.; Zgarbova, M.; Sponer, J.; Otyepka, M.; Jurecka, P.; Cheatham, T. E. Assessing the current state of amber force field modifications for DNA. *J. Chem. Theory Comput.* **2016**, *12* (8), 4114-4127. DOI: 10.1021/acs.jctc.6b00186

(97) Sgrignani, J.; Magistrato, A. QM/MM MD simulations on the enzymatic pathway of the human flap endonuclease (hFEN1) elucidating common cleavage pathways to RNase H enzymes. *ACS Catal.* **2015**, *5* (6), 3864-3875. DOI: 10.1021/acscatal.5b00178

(98) Pilbák, S.; Farkas, Ö.; Poppe, L. Mechanism of the tyrosine ammonia lyase reaction—tandem nucleophilic and electrophilic enhancement by a proton transfer. *Chem. Eur. J.* **2012**, *18* (25), 7793-7802. DOI: https://doi.org/10.1002/chem.201103662

(99) Hu, X.; Hu, H.; Melvin, J. A.; Clancy, K. W.; McCafferty, D. G.; Yang, W. Autocatalytic intramolecular isopeptide bond formation in gram-positive bacterial pili: A QM/MM simulation. *J. Am. Chem. Soc.* **2011**, *133* (3), 478-485. DOI: 10.1021/ja107513t

(100) Wu, S.; Xu, D.; Guo, H. QM/MM studies of monozinc β-lactamase CphA suggest that the crystal structure of an enzyme−intermediate complex represents a minor pathway. *J. Am. Chem. Soc.* **2010**, *132* (51), 17986-17988. DOI: 10.1021/ja104241g

(101) Stevens, D. R.; Hammes-Schiffer, S. Exploring the role of the third active site metal ion in DNA polymerase η with QM/MM free energy simulations. *J. Am. Chem. Soc.* **2018**, *140* (28), 8965-8969. DOI: 10.1021/jacs.8b05177

(102) 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* (44), 18497-18502, S18497/18491-S18497/18495. DOI: 10.1073/pnas.0902908106

(103) Brinkmeyer, M. K.; Pope, M. A.; David, S. S. Catalytic contributions of key residues in the adenine glycosylase MutY revealed by pH-dependent kinetics and cellular repair assays. *Chem. Biol. (Oxford, U. K.)* **2012**, *19* (2), 276-286. DOI: 10.1016/j.chembiol.2011.11.011

(104) Russelburg, L. P.; O'Shea Murray, V. L.; Demir, M.; Knutsen, K. R.; Sehgal, S. L.; Cao, S.; David, S. S.; Horvath, M. P. Structural basis for finding OG lesions and avoiding undamaged G by the DNA glycosylase MutY. *ACS Chem. Biol.* **2020**, *15* (1), 93-102. DOI: 10.1021/acschembio.9b00639

(105) Fromme, J. C.; Banerjee, A.; Huang, S. J.; Verdine, G. L. Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase. *Nature* **2004**, *427* (6975), 652-656. DOI: 10.1038/nature02306

(106) Brinkmeyer, M. K.; Pope, M. A.; David, S. S. Catalytic contributions of key residues in the adenine glycosylase MutY revealed by pH-dependent kinetics and cellular repair assays. *Cell Chem. Biol.* **2012**, *19* (2), 276-286. DOI: 10.1016/j.chembiol.2011.11.011

(107) Jonas, J.; Gut, J. Nucleic acid components and their analogs. XXIV. Ultraviolet spectra and dissociation constants of 6-azauracil thioxo derivatives. *Collect. Czech. Chem. Commun.* **1962**, *27*, 1886. DOI: 10.1135/cccc19621886

(108) Parker, J. B.; Stivers, J. T. Uracil DNA glycosylase: Revisiting substrate-assisted catalysis by DNA phosphate anions. *Biochemistry* **2008**, *47* (33), 8614-8622. DOI: 10.1021/bi800854g

(109) Werner, R. M.; Stivers, J. T. Kinetic isotope effect studies of the reaction catalyzed by uracil DNA glycosylase: Evidence for an oxocarbenium ion−uracil anion intermediate. *Biochemistry* **2000**, *39* (46), 14054-14064. DOI: 10.1021/bi0018178

(110) Cortázar, D.; Kunz, C.; Saito, Y.; Steinacher, R.; Schär, P. The enigmatic thymine DNA glycosylase. *DNA Repair* **2007**, *6* (4), 489-504. DOI: https://doi.org/10.1016/j.dnarep.2006.10.013

(111) Pidugu, L. S.; Dai, Q.; Malik, S. S.; Pozharski, E.; Drohat, A. C. Excision of 5-carboxylcytosine by thymine DNA glycosylase. *J. Am. Chem. Soc.* **2019**, *141* (47), 18851-18861. DOI: 10.1021/jacs.9b10376

Chapter 4: Decoding the Catalytic Strategy of AlkA: Insights from MD Simulations and QM/MM Calculations

4.1. Introduction

DNA is constantly attacked by a wide array of damaging agents originating from both within the cell and the environment.[1-3] For example, cellular metabolism generates reactive oxygen species (ROS), which can oxidize nucleotide components,[4] while alkylating agents, ultraviolet (UV) radiation, and diverse chemicals from industrial pollution and lifestyle factors further jeopardize the chemical integrity of DNA.[4-6] The most prevalent DNA damage products include base modifications,[7] abasic sites,[8, 9] single- and double-strand breaks,[10, 11] and crosslinking events.[12] If left unrepaired, this damage can affect the fidelity of transmission of genetic information,[13] leading to chromosomal aberrations,[14] genomic instability,[15] or apoptosis.[16] In multicellular organisms, the accumulation of DNA damage is closely tied to the development of cancer, neurodegenerative diseases, and the aging process.[13, 17, 18]

A common form of DNA damage is the addition of alkyl groups to the DNA bases by, for example, endogenous compounds, tobacco smoke and nitrosamines in food.[19] While all four canonical bases are targeted by alkylating agents,[20] the most common forms of alkylation damage are 7-methylguanine (7mG) and 3-methyladenine (3mA).[19] Although many DNA repair pathways exist in cells,[21-24] alkylation damage is commonly repaired via the base excision repair (BER) pathway.[19, 25, 26] BER of many types of alkylation damage in bacterial cells is initiated by 3-methyladenine DNA glycosylase II (AlkA), which cleaves the glycosidic bond that attaches the nucleobase to the sugar. AlkA is a monofunctional glycosylase belonging to the helix-hairpin-helix (HhH) superfamily.[27, 28] While glycosylases generally recognize a confined spectrum of chemically similar lesions,[29-33] AlkA stands out for its broad substrate specificity.[34, 35] Indeed, while the glycosylase responsible for repairing 3mA in human cells (alklyadenine DNA glycosylase (AAG)) catalyzes the removal of several other damaged purines,[34, 35] AlkA also removes small alkylated pyrimidines (e.g., O2-methylthymine (O2mT) and O2-methylcytosine (O2mC)),[34, 35] large purine adducts (e.g., 1,N^6-α-hydroxypropanoadenine),[27] and undamaged adenine in A:C mispairs.[36] As a

DNA repair enzyme with many targets, AlkA is important for bacterial cell survival, and therefore AlkA inhibitors have potential antibacterial properties.[37] However, inhibitor development is hindered by a lack of information about the AlkA catalytic mechanism, with conflicting proposals present in the literature.[28, 38]

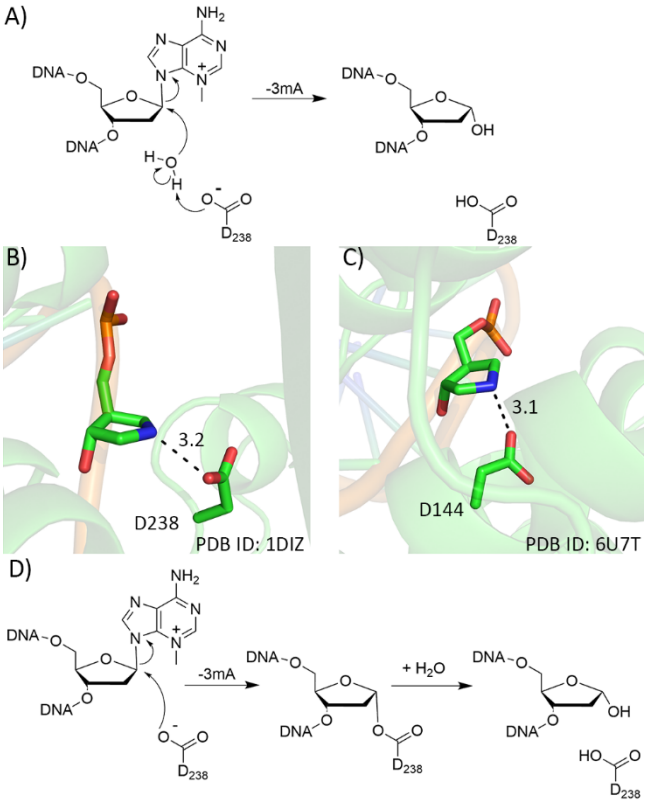


Figure 4.1. A) Proposed direct hydrolysis glycosylation pathway for AlkA.[38, 44] Crystal structure of B) AlkA or C) MutY bound to azaribose−containing DNA. Distances in Å. D) Proposed crosslinking pathway for AlkA.[27, 28]

For the majority of glycosylases, a nearly conserved aspartate or glutamate activates a water molecule for nucleophilic attack at C1' of deoxyribose.[39-43] Since mutagenic studies have revealed that D238 plays an essential role in AlkA activity,[38] a similar mechanism of action was proposed for AlkA catalysis (Figure 4.1A).[38, 44] However, the most recent mechanistic proposal is based on a crystal structure of AlkA bound to DNA containing a transition state mimic (PDB ID: 1DIZ; 2.5 Å resolution), which shows D238 in close proximity to the atom corresponding to C1' of deoxyribose (r(N1′−D238) = 3.2 Å, Figure 4.1B).[28] Therefore, it was proposed that a DNA−protein crosslink

forms along the AlkA mechanism of action (Figure 4.1D).[27, 28] Indeed, a crystal structure of monofunctional adenine DNA glycosylase (MutY) bound to the same transition-state analogue contains a similar alignment of an aspartate (D144) with respect to the substrate (PDB ID: 6U7T; 2.0 Å resolution, Figure 4.1C),[45] and extensive experimental and computational investigations[46-48] revealed MutY utilizes a crosslink formed between D144 and C1′ of the damaged nucleotide as a vital part of catalysis. Nevertheless, it remains unclear whether AlkA cleaves glycosidic bonds through a traditional water nucleophile-based mechanism (Figure 4.1A) or crosslink formation (Figure 4.1D).

Computational methodologies have successfully elucidated the catalytic mechanism of many DNA glycosylases.[40, 42, 43, 47-62] Molecular dynamics (MD) simulations can provide structural details necessary to understand glycosylase substrate binding and active site alignment. For example, MD approaches have been used to determine the substrate binding pathway for 8-oxoguanine DNA glycosylase (hOGG1)[63] and to identify residues necessary for base excision by thymine DNA glycosylase (TDG).[49] To complement structural data obtained from MD simulations, quantum mechanics/molecular mechanics (QM/MM) calculations can provide otherwise unobtainable atomic level details of enzymatic mechanisms. Indeed QM/MM approaches have previously determined that MutY catalysis proceeds through a crosslinked intermediate[47] and identified histidine as the preferred nucleophile-activating residue to catalyze nucleotide hydrolysis by uracil DNA glycosylase (UDG).[55] These select examples represent a small subset of the insightful computational work that has been done to date on members of the DNA glycosylases family.[40, 42, 43, 47-62]

In this study, we use a multipronged computational approach to deepen our understanding of the AlkA mechanism of action. First, all-atom MD simulations on a model of AlkA bound to 3mA-containing DNA are used to identify active site residues responsible for accommodating 3mA in the AlkA binding pocket and explore the active site structural dynamics to identify conformations with an alignment that supports different proposed pathways.[28, 38] To complement these structural insights, QM/MM calculations were then performed from MD representative structures to compare

the thermochemical requirements of the two proposed reaction mechanisms. The feasibility of each pathway was evaluated through comparisons with available experimental kinetic,[64] mutagenic,[38, 65] and structural data.[28] The insights obtained regarding AlkA function will further our understanding of DNA repair pathways, which has implications for the development of cancer therapeutics that target similar pathways in tumor cells,[66-73] and permit the exploitation of the absence of an AlkA human homologue to develop inhibitors as new antibacterial agents.[37]

4.2. Computational Methods

4.2.1. MD simulations

A model of the AlkA−substrate complex was built based on a crystal structure of the enzyme bound to azaribose-containing DNA (PDB ID: 1DIZ, Figure 4.1B).[28] Azaribose was converted to the 3mA nucleotide (Figure 4.2A) and the unresolved side chains of R56, R97, and K133 were added using PyMOL.[74] The system was solvated with TIP3P water in a truncated octahedron water box with edges a minimum of 12 Å from the solute using the LEaP module of AMBER 2018.[75] Sodium ions were used to neutralize the solute charge and NaCl, modeled with Li/Merz ion parameters,[76] was added to bring the system to 150 mM salt concentration as calculated using the SLTCAP ion calculator.[77] The Amber force field was used to model the system, with protein residues described using ff14SB[78] and nucleic acid residues modeled using OL15.[79] Atom charges for the non-standard 3mA nucleotide were calculated using the pyRED server,[80-84] while missing parameters were adapted from the general AMBER forcefield (GAFF)[85] and assigned using Antechamber (Tables C.1 and C.2, Appendix C). Amino acid protonation states were determined using the H++ server.[86]

The model was initially minimized stepwise using 1000 steps of steepest decent followed by 1000 steps of conjugate gradient minimization. First, the solvent was minimized, while a 100 kcal mol^{-1} $Å^{-2}$ restraint was applied to the solute. Next, the restraint was removed from the solute hydrogen atoms. Subsequently, the solute was minimized, while applying a

100 kcal mol^{-1} $Å^{-2}$ restraint to the solvent. Finally, the entire system was minimized with no restraints, using 1000 steps of steepest decent followed by 2000 steps of conjugate gradient minimization. The system was then heated from 10 K to 310 K, with the temperature increased by 50 K every 20 ps. The restraints on the solute were subsequently decreased from 25 kcal mol^{-1} $Å^{-2}$ to 5 kcal mol^{-1} $Å^{-2}$ at a rate of 5 kcal mol^{-1} $Å^{-2}$ per 20 ps followed by an additional 20 ps equilibration step at 1.5 kcal mol^{-1} $Å^{-2}$. To allow the AlkA active site to relax with respect to the added 3mA, an additional 200 ns equilibration step was performed in which all protein residues except for those in close proximity to 3mA (i.e., F18, R22, L125, V128, D238, Y239, L240, W272, and Y273) were restrained with a weight of 1.5 kcal mol^{-1} $Å^{-2}$. After equilibration, ten 500 ns unrestrained MD production simulations were conducted.

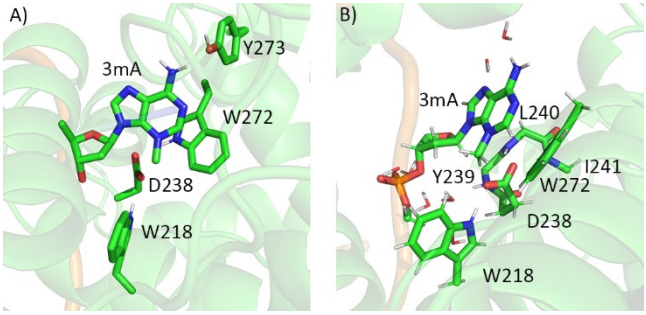


Figure 4.2. A) Post-equilibration structure of AlkA bound to 3mA-containing DNA used as a starting point for MD simulations. B) QM/MM high layer in AlkA−DNA models used to map different catalytic pathways.

Analysis was performed using the cpptraj program in AMBER 2018.[75] Hydrogen-bond occupancies were determined based on a heavy atom distance of less than 3.0 Å and a hydrogen-bond angle between 135° and 180°. The distance between residues in the AlkA substrate binding pocket was calculated from the geometric center of the 3mA purine ring to the geometric center of the tyrosine or tryptophan rings, Cδ1, Cδ2, and Cγ of leucine, Cβ, Cγ1, and Cγ2 of valine, or Nε, Cζ, and Nη of arginine. The binding of 3mA in the AlkA active site was monitored using a pseudodihedral angle formed between N1 and C4′ of 3mA and C4′ and N3 of the 3′-T with respect to the damage site (Figure C.1, Appendix C). Specifically, 3mA was considered to be flipped into the AlkA active site when ∠(N1C4′C4′N3) falls between 120° and 240°, and flipped out of the

enzyme active site towards DNA otherwise, with analysis only being performed on the flipped in portion. A water molecule was determined to be available for hydrolysis if simultaneously within 4 Å of C1′ and Oδ of D238. All simulations were run using AMBER 2018.[75]

4.2.2. QM/MM calculations

To investigate the catalytic mechanism of AlkA, we performed QM/MM calculations starting from a representative MD snapshot extracted by clustering the total dataset based on r(C1′−Oδ) using the heiragglo algorithm as implemented in cpptraj.[75] The QM/MM (ONIOM[87]) approach was used due to its well-established ability to accurately model enzymatic reactions,[88, 89] including those catalyzed by DNA glycosylases.[42, 43, 47-58, 61] The QM layer encompassed the target 3mA nucleotide, the adjacent 3′ phosphate, key active site residues (W218, W272, D238, and the backbones of I241, L240, and Y239), and seven water molecules (119 atoms, Figure 4.2B), yielding a net QM charge of −1. The remaining portion of the AlkA–DNA complex and solvent molecules within 6 Å of the enzyme or DNA (3830 water molecules) formed the MM region. The QM layer was treated with the M06-2X functional and 6-31G(d,p) basis set, while the MM layer was described using the same AMBER force field[75] assignments as implemented in the MD simulations. The mechanical embedding (ME) scheme was employed throughout due to its demonstrated reliability for other enzymes,[47, 90-95] including those that catalyze nucleic acid hydrolysis[95] and form DNA−protein crosslinks.[47] All protein, DNA, and solvent atoms were optimized. Frequency calculations were used to confirm the nature of all stationary points and to obtain thermal corrections for Gibbs energy profiles. Single-point energies were obtained at the ONIOM(M06-2X/6-311+G(2df,p):AMBER14SB) level of theory. Scans of bond distances relevant for the reaction being investigated (i.e., r(O_{WAT}−C1′) for hydrolysis and r(D238Oδ−C1′) for crosslink formation) were used to identify approximate transition state (TS) structures. The energy maxima on the resulting potential energy surfaces were then used as the starting points for transition state optimizations. Transition states were validated by the presence of a single imaginary (negative) frequency. As discussed in the results, single-point calculations were performed on the RC and the TS

corresponding to the direct hydrolysis mechanism in which either W218 or W272 was mutated to alanine or moved from the QM layer to the MM layer (Table C.3, Appendix C). All QM/MM calculations were carried out using Gaussian 16 (Revision B.01).[80] Calculated barriers were compared to the experimental barrier (88.4 kJ/mol) estimated from a rate constant of 0.5 min^{-1} that was measured at 37°C and a pH of 6.0 under single-turnover conditions[64] using the Eyring equation.

4.3. Results

4.3.1. AlkA binds 3mA through non-specific interactions in the active site pocket

While the substrate specificity of AlkA has been carefully investigated,[34] the lack of a crystal structure of AlkA bound to DNA containing a nucleotide substrate means there is limited information regarding how the AlkA active site accommodates damaged nucleotides like 3mA. To gain insight into substrate binding, MD simulations were performed on AlkA bound to 3mA-containing DNA. The calculations reveal that the position of 3mA relative to AlkA is highly dynamic. For a portion of the simulation time (55%), 3mA is partially flipped out of the AlkA active site to integrate into the DNA helix (Figure C.2A, Appendix C). The observed dynamics could arise due to inherent instabilities resulting from the addition of a nucleobase-containing substrate to a crystal structure with an abasic site in the substrate binding pocket, the nucleotide naturally fluctuating between being flipped into the DNA helix and the AlkA active site, or a combination of these two factors. Indeed, although much longer simulated timescales would be required for an accurate comparison, the experimental equilibrium constant for substrate base flipping (K_{flip} = 3)[96] is only slightly larger than the one estimated from the MD simulations ($K_{flip\text{-}calc}$ ~ 1). For the remainder of the simulation time (45%), 3mA remains well accommodated in the AlkA active site.

The damaged nucleobase sits in a pocket formed by V128, W272, R22, Y273, W218, Y239, L240, and Y222 (Figure 4.3A and C.3, Appendix C). L240 and W272 are closest to 3mA, forming one wall of the cavity on the O4′ side of the nucleoside (Figure 4.3B). L240 sits near the 3mA five-

membered ring, while W272 stacks with the six-membered ring, with this relative orientation maintained for the entire time 3mA is accommodated in the active site. Y239 is also positioned on

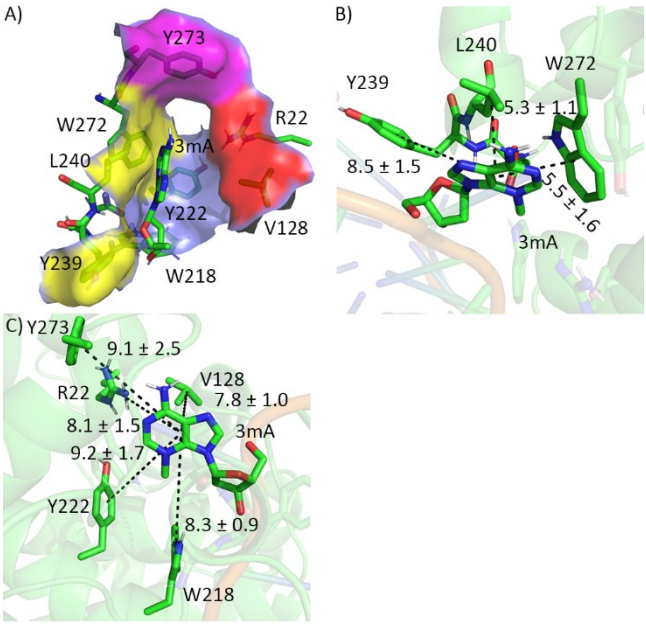


Figure 4.3. A) MD representative structure of the AlkA substrate binding pocket, highlighting the wall closest to 3mA (yellow), the wall farther from 3mA (red), the floor (blue), and the end cap (purple). Distance between 3mA and key residues in the B) closest wall or C) the far wall, floor, and end cap of the binding pocket.

the O4′ side of 3mA. While Y239 does not form close contacts with 3mA for most of the simulation time, Y239 periodically stacks with the six-membered ring of 3mA (7% occupancy, Figure C.2B and C.3, Appendix C). The opposing wall, formed by R22 and V128, sits further from the 3mA nucleobase (~2.5 Å increased distance, Figure 4.3C and C.3, Appendix C). The larger distance may arise due to repulsion between the positively charged 3mA and R22. The floor of the cavity is composed of W218 and Y222, which form hydrophobic interactions with the lesion (Figure 4.3C). The far end of the pocket is occupied by Y273. Unlike other active site residues, which overwhelmingly adopt a single conformation, Y273 adopts three almost evenly occupied orientations over the simulation (Figures 4.4 and C.3, Appendix C). In the first conformation, Y273 falls more than 11 Å from 3mA, which leaves a void in the active site pocket (denoted Y273-disengaged; 37% occupancy, Figure 4.4A). In the second conformation, Y273 is situated closer to

3mA (within ~8 Å), but still too far to interact with the substrate nucleobase (Y273-semi-engaged; 30% occupancy, Figure 4.4B). However, in the third conformation, Y273 stacks with 3mA, capping the active site pocket (Y273-engaged; 33% occupancy, Figure 4.4C). Overall, AlkA accommodates 3mA in the active site through non-specific hydrophobic interactions with many surrounding residues, while Y273 adopts three different conformations that result in different binding pocket sizes.

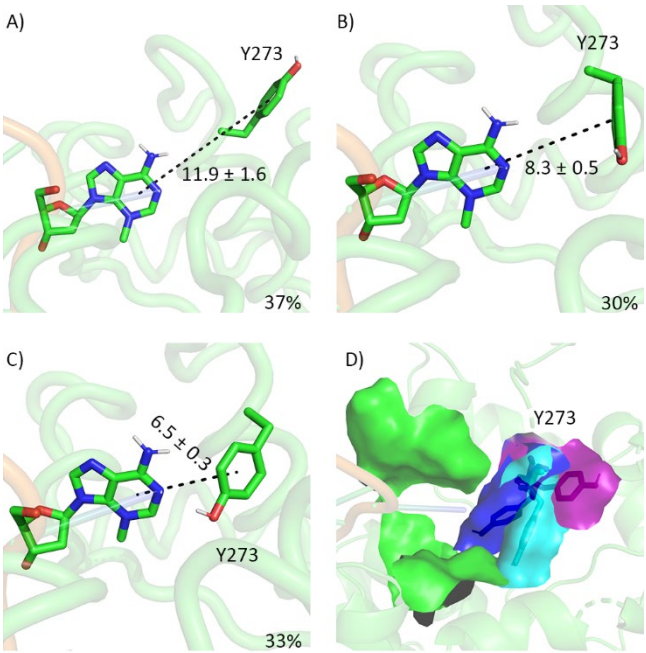


Figure 4.4. Y273 conformations sampled during MD simulations, denoted as the A) disengaged, B) semi-engaged, and C) engaged states. Conformational occupancies in the bottom right corner. D) Overlay of the Y273 disengaged (purple), semi-engaged (cyan), and engaged (blue) conformations, highlighting the change in active site size with Y273 conformation. The rest of the active site pocket is in green space filling.

4.3.2. D238 can be oriented near deoxyribose to allow for either activation of a water nucleophile or formation of a crosslinked intermediate as part of AlkA catalytic activity

D238 has been implicated as a participant in the AlkA catalytic mechanism as the D238N AlkA mutant is inactive.[38] However, the exact role of this residue in catalysis is not understood. A short distance between D238 and the transition state analogue in a crystal structure (PDB ID: 1DIZ) has led to the proposal that D238 forms a crosslink with C1′ as part of catalysis,[27, 28] which parallels the

mechanism reported for MutY.[46, 47] However, it is also common for glycosylases to use an active site aspartate/glutamate to activate a water nucleophile for attack at C1′ to hydrolyze the glycosidic bond,[39-43] which led to the proposal that the mutational impact is due to the role of D238 in nucleophile activation.[38, 44] During MD simulations of AlkA bound to 3mA-containing DNA, D238 remains near C1′ of 3mA (4.1 ± 1.1 Å, Figure 4.5A). This orientation is maintained by highly-occupied D238−W272 (75%) and D238−W218 (54%) hydrogen bonds, as well as less-occupied hydrogen bonds with the backbones of L240 (27%) and I241 (18%, Figure 4.5B). The proposed role of these residues in positioning D238 is consistent with mutation of the tryptophan in archaeal AlkA (W204) corresponding to W218 to alanine abolishing catalytic activity.[65] This highlights that, in addition to forming the active site pocket to bind 3mA, W218 and W272 are key for positioning D238 for catalysis.

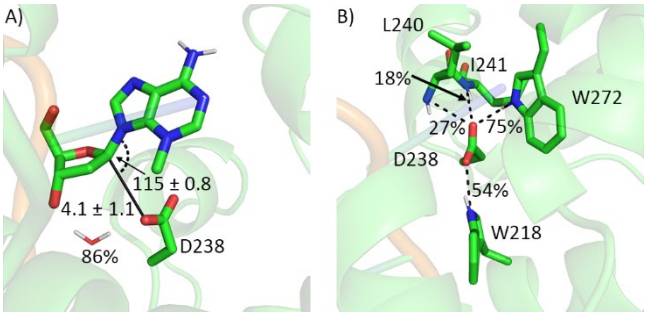


Figure 4.5. A) Relative alignment of D238 and 3mA from MD simulations on AlkA. Distance in Å, angle in degrees, and occupancy refers to the presence of a potential water nucleophile. B) Occupancies of key hydrogen bonds to D238 in the AlkA active site identified from MD simulations.

For a catalytic mechanism where D238 forms a crosslink with C1′ to be feasible, D238 needs to be near C1′ (likely within 4 Å) and positioned on the opposite side of the deoxyribose as the 3mA nucleobase (defined by ∠(N9C1′Oδ) > 95°). When 3mA is positioned in the AlkA active site, these conditions were met for 71% of the simulation time, indicating D238 is well aligned for crosslink formation. On the other hand, a water molecule is located between D238 and C1′, and is therefore available for hydrolysis, for 86% of the simulation time when 3mA is accommodated in the active site (Figure 4.5A). Thus, our classical MD simulations confirm that the active site adopts conformations conducive for both crosslink formation and activation of a water nucleophile for a

large portion of the simulation time. To determine the most energetically feasible pathway, QM/MM calculations were used to map each proposed mechanism.

4.3.3. Although D238 is aligned for attack at C1′, DNA−D238 crosslink formation is kinetically infeasible

Since MD simulations support that D238 is aligned for nucleophilic attack at C1′ of 3mA in AlkA (Figure 4.5), which parallels the accepted mechanism for MutY,[46-48] QM/MM calculations were used to map a crosslink formation mechanism for AlkA (Figures 4.6 and C.4, Appendix C). In the QM/MM reactant complex (RC), D238 is positioned 3.368 Å away from C1′ of 3mA, which is consistent with the crystal structure (3.2 Å from transition state analogue), and suggests a potential nucleophilic role (Figure C.4, Appendix C). D238 forms tight hydrogen bonds with the side chains of W272 and W218, and the backbone of L240, as well as a weaker interaction with the backbone of I241. In the TS, the distance between D238 and C1′ is reduced to 2.429 Å, which is consistent with the initiation of covalent bond formation. The glycosidic bond between C1′ and N9 of 3mA is significantly elongated (2.165 Å) compared to the distance in the RC (1.503 Å). As D238 approaches C1′ in the TS, interactions between D238 and other active site residues are weakened, including hydrogen bonding with W218, L240, and I241 (~0.2 Å, ~1.0 Å, and ~0.7 Å increase in hydrogen-bond length, respectively, Figure C.4, Appendix C). The crosslink is fully formed in the PC, with 3mA completely detaching from deoxyribose and the L240 and I241 hydrogen bonds with D238 are restored. However, the transition state is associated with a very high barrier of 141.9 kJ/mol (Figure 4.6), significantly above the experimental value for AlkA of 88.4 kJ/mol.[64] The large energetic cost likely results from the D238 hydrogen-bonding network in the RC structure, which requires three hydrogen bonds to be weakened to transition to the TS. Thus, crosslink formation is kinetically unfavorable. Although enzyme regeneration would require subsequent crosslink hydrolysis, this step was not considered due to the prohibitively high barrier associated with crosslink formation.

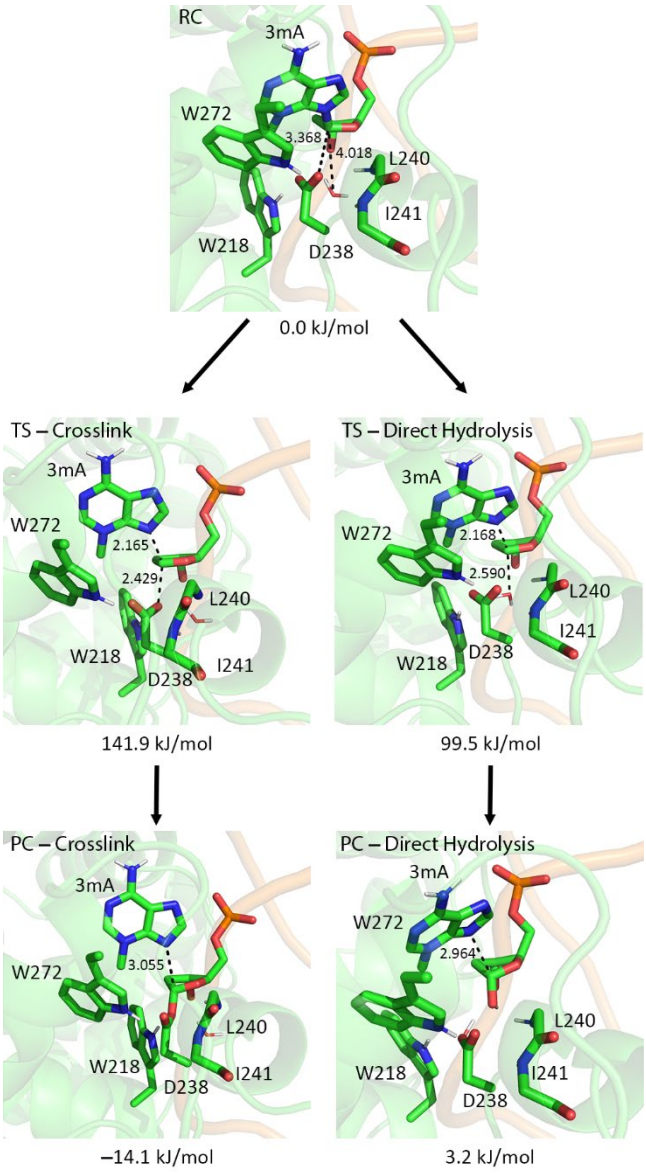


Figure 4.6. QM/MM optimized stationary points and relative Gibbs energy with respect to the RC for the AlkA-catalyzed crosslinking (left) and direct hydrolysis (right) pathways for 3mA excision. Distances in Å. Refer to Figures C.4 and C.5 for more detailed structural information.

4.3.4. AlkA catalysis proceeds through direct hydrolysis of the glycosidic bond via a D238-activated water nucleophile

With crosslink formation disfavored for AlkA, the hydrolysis mechanism widely accepted for other glycosylases[39-43] was considered. Indeed, MD simulations reveal a water nucleophile is frequently positioned between D238 and C1′ in the AlkA active site (86% occupancy when 3mA is

bound in the active site). In the QM/MM RC, a water nucleophile is located near C1′ of 3mA (r(WatO−C1′) = 4.018 Å) and hydrogen bonds to D238 (r(WatH−D238) = 2.926 Å, Figures 4.6 and C.5, Appendix C). In the TS, the nucleophilic water has initiated attack at C1′ (r(WatO−C1′) = 2.590 Å), while the glycosidic bond is partially cleaved (r(C1′−N9) = 2.168, Figure C.5, Appendix C). Unlike the crosslink formation mechanism, D238 interactions with W272, W218, L240, and I241 are maintained at similar lengths in the TS and RC, which results in a barrier of 99.5 kJ/mol. In the PC, the nucleophilic attack by the water molecule has completed (r(WatO−C1′) = 1.435 Å), a proton is donated to D238 (r(WatH−D238) = 1.046 Å), and the glycosidic bond is cleaved (r(C1′−N9) = 2.964 Å).

In glycosylases, it is common for residues that participate in substrate binding to also have a role in base departure.[42, 47] Therefore, we investigated the roles of the residue that forms the closest contact to 3mA (W272) and the residue that is implicated in catalysis from archaeal AlkA (W218) in base departure by mutating the residue to alanine and analyzing the resulting impact on catalysis. While removing W218 from the QM region had a negligible impact on the reaction barrier (less than a 3 kJ/mol difference), removing W272 lowers the barrier by ~10 kJ/mol (Table C.3, Appendix C). This suggests cation–π contacts between W272 and 3mA are key for substrate binding but anticatalytic for base departure.

The calculated energy barrier for the reaction (99.5 kJ/mol, Figure 4.6) is consistent with the experimental barrier (88.4 kJ/mol)[64] and significantly lower than the barrier for crosslink formation (141.9 kJ/mol). Additionally, the prominent role of D238 in the direct hydrolysis mechanism characterized for AlkA is consistent with D238N mutant AlkA losing catalytic activity[38] and mutation of the W218 analogue responsible for positioning D238 in archaeal AlkA stopping catalysis. The characterized transition state for the direct hydrolysis pathway also resembles the crystal structure of AlkA bound to DNA transition state analogue.[28] Considering all these factors, direct hydrolysis of the glycosidic bond represents the preferred catalytic pathway for AlkA activity.

4.4. Discussion

4.4.1. AlkA binds 3mA through hydrophobic interactions in the binding pocket, which aligns the active site to facilitate direct hydrolysis of the glycosidic bond

Most glycosylases target a small group of chemically similar lesions that can be accommodated in the enzyme active site.[29-33, 39] For example, UDG specifically targets uracil (U) in DNA through specific hydrogen bonds between the Watson-Crick face of U and asparagine and histidine residues[39, 97] as well as a series of hydrophobic interactions that ensure only pyrimidines with no C5 substitution can fit in the binding space. On the other hand, AlkA acts on charged and uncharged, purine and pyrimidine substrates that have alkyl damage in different locations.[34] Furthermore, AlkA excises canonical adenine from A:C mispairs, although at a reduced rate compared to damaged nucleobases.[36] The diverse shape, size, and chemical properties of AlkA substrates suggest the presence of an adaptable binding pocket, which must still promote an active site orientation conducive for catalysis.

Our MD simulations provide the first structural insights into the binding of AlkA to a DNA substrate. We reveal that the AlkA substrate binding pocket consists of eight hydrophobic residues that form non-specific interactions with 3mA (Figure 4.3). The closest interactions between the binding pocket and substrate involve L240 and W272. Indeed, W272 likely forms a strong cation–π interaction with 3mA, which contrasts repulsion with the positively charged R22 on the other side of the active site pocket. In fact, mutating W272 to alanine results in an ~10 kJ/mol reduction in the energy barrier, which indicates a strong enough cation–π contact to be anti-catalytic. Our simulations also show that Y273 encloses the substrate binding pocket. However, Y273 is highly flexible, adopting different conformations that adjust the depth of the active site. A conducive active site configuration for catalysis is facilitated by a hydrogen-bonding network between 3mA, D238, W272, W218, L240 and I241. D238 is properly positioned for catalysis primarily through D238−tryptophan hydrogen bonds, with support from hydrogen bonding with the backbones of L240 and I241. This is consistent with the crystal structure of AlkA bound to azaribose-containing DNA where D238 forms hydrogen bonds with those same four residues (PDB ID: 1DIZ, Figure

4.7A).[28] Interestingly, W272 both stacks with 3mA and hydrogen bonds with D238, forming a bridge between the two residues and ensuring a catalytically-conducive orientation of D238 and 3mA. Indeed, D238 was aligned for water nucleophile activation or direct attack at C1′ of 3mA for over 70% the MD simulation time when 3mA is accommodated in the AlkA active site (Figure 4.5A).

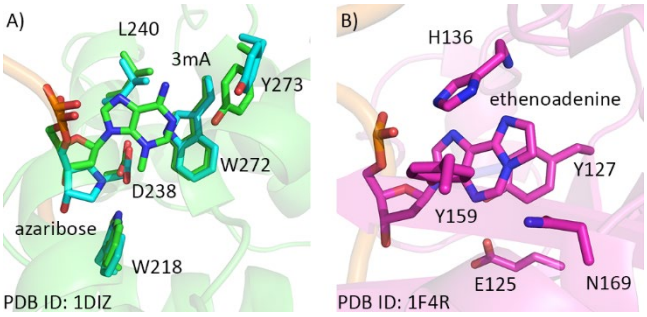


Figure 4.7. A) Overlay of the QM/MM optimized AlkA RC obtained from an MD representative structure (green) and crystal structure of AlkA bound to azaribose-containing DNA (cyan, PDB ID: 1DIZ). B) Crystal structure of AAG bound to ethenoadenine-containing DNA (PDB ID: 1F4R).

Starting from an MD representative structure, QM/MM calculations provide the first computational description of the AlkA mechanism of action. In the preferred mechanism, D238 activates a water nucleophile that attacks at C1′ of 3mA, cleaving the glycosidic bond. This pathway is comparable to the mechanism of action of many other monofunctional DNA glycosylases.[39-43] The energy barrier for the reaction (99.5 kJ/mol, Figure 4.6) is consistent with that estimated from experimental rates (88.4 kJ/mol) for the 3mA substrate.[64] The prominent role for D238 in the mechanism is consistent with D238N mutation preventing AlkA catalytic activity.[38] Additionally, the configuration of active site residues in our calculated RC is consistent with that in the crystal structure of the transition state analogue (Figure 4.7A), including the location of D238 relative to the substrate (3.4 Å vs 3.2 Å) and W272 and W218 with regards to D238.[28] Overall, the direct hydrolysis mechanism characterized in this work is fully consistent with the currently available structural,[28] mutagenic,[38] and kinetic[64] data for AlkA. As an enzyme not present in humans, the improved understanding of the AlkA catalytic mechanism provided by this work promises to aid in the development of transition state analogue inhibitors[98-100] with antibacterial properties.[37] Additionally, the new insights gained about AlkA function broaden our understanding of cellular

DNA repair pathways, which has implications for the development of cancer fighting therapeutics that target BER enzymes.[66-73]

4.4.2. The formation of a DNA−protein crosslink during catalysis remains unique to MutY among the DNA glycosylases

A crystal structure of AlkA bound to azaribose-containing DNA reveals D238 is in close proximity to the transition state analogue (r(N1′−D238) = 3.2 Å, PDB ID: 1DIZ, Figure 4.1B).[28] The crystal structure of MutY bound to azaribose-containing DNA (PDB ID: 6U7T)[45] also uncovers a close contact between an aspartate (D144) and the azaribose (Figure 4.1C). Additionally, both enzymes become catalytically inactive when the aspartate in question is mutated to alanine or asparagine[38, 64, 101] and can cleave canonical adenine from DNA.[102, 103] These factors suggest that AlkA and MutY may utilize similar catalytic mechanisms. However, while MutY forms a crosslink between D144 and C1′ to excise the target base,[46-48] crosslink formation is energetically unfeasible for AlkA (Figure 4.6). Our calculations suggest that this mechanistic difference is a consequence of how the catalytic aspartate is held in the enzyme active site. Specifically, D144 in MutY is positioned for attack by a single hydrogen bond with N146 which is maintained throughout the pathway,[46] while D238 in AlkA is held in place by an intricate network of hydrogen bonds involving four different residues (W218, W272, I241, L240) with three of the interactions weakening or breaking during crosslink formation (Figure C.4, Appendix C**)**. As a result, the energy barrier for crosslink formation in AlkA is increased to the point of being catalytically infeasible. This comparison between AlkA and MutY highlights how seemingly minor changes in the active site of a glycosylase can have large consequences for catalysis. Thus, formation of a DNA−protein crosslink during MutY catalysis[46-48] remains unique among the monofunctional glycosylases.

4.4.3. Key Differences in the AlkA and AAG binding pockets contribute to unique substrate specificity

While AlkA is responsible for repairing many types of alkylation damage in bacteria, AAG repairs several alkylation damage products in human cells.[34, 35] Both enzymes primarily target common purine alkylated bases such as 3mA, 7mG, 1,N^6-ethenoadenine, and hypoxanthine.[34, 35] Correspondingly, the mechanism adopted by the two enzymes is also comparable, with a basic active site residue (E125 in AAG or D238 in AlkA)[43] activating a water nucleophile for attack at C1′ of the damaged nucleotide. Additionally, AlkA and AAG have many similarities in their active site architectures. For AAG, substrate binding involves non-specific π−π interactions between the damaged nucleobase and Y127 and Y159, while W272 plays a similar role in AlkA (Figure 4.7B). Indeed, these residues are similarly anti-catalytic towards excision of charged 3mA, with *in silico* Y127A and Y159A AAG mutants resulting in a ~10 kJ/mol reduction in the barrier to 3mA excision compared to wild-type,[42] while a W272A mutation in AlkA also reduces the hydrolysis energy barrier by ~10 kJ/mol (Table C.3, Appendix C). These aromatic residues form a hydrogen-bonding network to correctly position the catalytic aspartate/glutamate with respect to the substrate. Specifically, Y127 stacks with the AAG substrate while hydrogen bonding with E125, and W272 stacks with the AlkA substrate while hydrogen bonding with D238.

Despite the above similarities, AlkA has a broader set of substrates than AAG, extending to damaged pyrimidines (O2mT, O2mC) and even canonical adenine in A:C pairs.[34-36] The first structure of AlkA bound to a substrate (3mA) containing DNA obtained from our MD simulations provides insight into these differences in substrate specificity. Specifically, in the AAG active site, N169 and the H136 backbone form specific hydrogen-bonding interactions with substrates as seen in the crystal structure of AAG bound to 1,N^6-ethenoadenine-containing DNA (PDB ID: 1F4R, Figure 4.7B)[104] and MD simulations on AAG bound to various substrates.[105] Furthermore, MD simulations highlight that N169 and H136 are responsible for AAG discriminating against natural adenine and guanine bases.[106] Conversely, no AlkA active site residue forms specific hydrogen bonding with 3mA, which is likely a factor in the excision of unmodified adenine from C:A base

pairs.[36] Additionally, AAG does not have a residue that corresponds to Y273 in AlkA that can modulate the size of the binding pocket. This likely contributes to the ability of AlkA to cleave smaller pyrimidines,[34, 35] while AAG is only known to excise purines.[34, 35] Another AlkA residue with no AAG counterpart is R22. Indeed, the presence of a positively-charged residue in the binding pocket of an enzyme with many positively charged substrates appears counterproductive. However, cationic substrates like 3mA have weakened glycosidic bonds,[64] which may permit catalysis even with reduced substrate binding affinity. On the other hand, positively-charged R22 has the potential to form strong cation–π interactions with neutral substrates,[107] facilitating neutral substrate binding. A positively charged amino acid can also stabilize the anionic leaving group resulting from cleaving a less labile glycosidic bond associated with a neutral substrate. This give and take, where an active site residue hinders catalysis for (cationic) substrates that are inherently easier to excise in order to promote activity on (neutral) substrates with more difficult chemistry, could be another factor contributing to the substrate diversity of AlkA.

Overall, the structure of the AlkA−substrate complex, including the composition of the binding pocket, characterized in this work suggests many unique active site features work in tandem to afford the observed substrate promiscuity of AlkA. With an inability of the active site to distinguish between damaged and natural bases, this work supports proposals that AlkA discerns between substrates using base-pair stability[36] and/or glycosidic bond strength.[64] Additional experimental or computational studies on AlkA complexed with diverse substrates (including neutral variants) are needed to conclusively determine the roles of R22 and Y273 in AlkA activity.

4.5. Conclusion

The present study used a combination of MD simulations and QM/MM calculations to reveal how AlkA binds alkylated nucleotides and identify the preferred catalytic mechanism of action. Our MD simulations highlight that the AlkA substrate binding pocket is composed of eight residues (V128, W272, R22, Y273, W218, Y239, L240, and Y222), which form non-specific, largely hydrophobic interactions with the substrate. The lack of specific enzyme−substrate hydrogen

bonding and the ability of Y273 to modulate the binding pocket size is proposed to permit the accommodation of substates of various sizes (purines, pyrimidines) and chemical properties (cationic, neutral). This supports previous proposals that base-pair stability[36] and/or glycosidic bond strength[64] are likely discriminating features for AlkA substrate recognition. Despite being hypothesized to utilize a DNA−protein crosslink to achieve catalysis,[27, 28] our calculations suggest that AlkA-facilitated glycosidic bond cleavage proceeds via a direct hydrolysis mechanism, where D268 activates a water nucleophile that attacks at C1′ of 3mA. The position of D238 required for the reaction is maintained by an intricate hydrogen-bond network involving W272, W218, L240, I241, and the substrate. Our proposed catalytic pathway is fully consistent with currently available experimental mutagenic,[38, 65] kinetic,[64] and structural[28] data. The knowledge gained about AlkA function will aid in the development of inhibitors as antibacterial drugs[37] and pushes forward our understanding of BER in cells.

4.6. References

(1) Carusillo, A.; Mussolino, C. DNA damage: From threat to treatment. *Cells* **2020**, *9* (7), 1665.

(2) Barnes, J. L.; Zubair, M.; John, K.; Poirier, M. C.; Martin, F. L. Carcinogens and DNA damage. *Biochem. Soc. Trans.* **2018**, *46* (5), 1213-1224. DOI: 10.1042/bst20180519 (acccessed 12/17/2025).

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

(4) Rodic, S.; Vincent, M. D. Reactive oxygen species (ROS) are a key determinant of cancer's metabolic phenotype. *Int. J. Cancer* **2018**, *142* (3), 440-448. DOI: https://doi.org/10.1002/ijc.31069

(5) Ackerman, S.; Horton, W. Effects of environmental factors on DNA. 2018; pp 109-128.

(6) Liu, N.; Du, J.; Ge, J.; Liu, S. B. DNA damage-inducing endogenous and exogenous factors and research progress. *Nucleosides Nucleotides Nucleic Acids* **2025**, *44* (12), 969-1001. DOI: 10.1080/15257770.2024.2428436

(7) Basu, A. K. DNA damage, mutagenesis and cancer. *Int. J. Mol. Sci.* **2018**, *19* (4). DOI: 10.3390/ijms19040970

(8) Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **1993**, *362* (6422), 709-715. DOI: 10.1038/362709a0

(9) Cadet, J.; Davies, K. J. A.; Medeiros, M. H. G.; Di Mascio, P.; Wagner, J. R. Formation and repair of oxidatively generated damage in cellular DNA. *Free Radical Biol. Med.* **2017**, *107*, 13-34. DOI: https://doi.org/10.1016/j.freeradbiomed.2016.12.049

(10) Kezuka, Y.; Yoshida, Y.; Nonaka, T. Structural insights into catalysis by βC-S lyase from streptococcus anginosus. *Proteins* **2012**, *80* (10), 2447-2458. DOI: https://doi.org/10.1002/prot.24129

(11) Vítor, A. C.; Huertas, P.; Legube, G.; de Almeida, S. F. Studying DNA double-strand break repair: An ever-growing toolbox. *Front. Mol. Biosci.* **2020**, *Volume 7 - 2020*. DOI: 10.3389/fmolb.2020.00024

(12) Klages-Mundt, N. L.; Li, L. Formation and repair of DNA-protein crosslink damage. *Sci. China Life Sci.* **2017**, *60* (10), 1065-1076. DOI: 10.1007/s11427-017-9183-4

(13) Chatterjee, N.; Walker, G. C. Mechanisms of DNA damage, repair, and mutagenesis. *Environ. Mol. Mutagen.* **2017**, *58* (5), 235-263. DOI: https://doi.org/10.1002/em.22087

(14) Durante, M.; Bedford, J. S.; Chen, D. J.; Conrad, S.; Cornforth, M. N.; Natarajan, A. T.; van Gent, D. C.; Obe, G. From DNA damage to chromosome aberrations: Joining the break. *Mutat. Res. Genet. Toxicol. Environ. Mutagen.* **2013**, *756* (1), 5-13. DOI: https://doi.org/10.1016/j.mrgentox.2013.05.014

(15) Tubbs, A.; Nussenzweig, A. Endogenous DNA damage as a source of genomic instability in cancer. *Cell* **2017**, *168* (4), 644-656. DOI: https://doi.org/10.1016/j.cell.2017.01.002

(16) Boon, N. J.; Oliveira, R. A.; Körner, P.-R.; Kochavi, A.; Mertens, S.; Malka, Y.; Voogd, R.; van der Horst, S. E. M.; Huismans, M. A.; Smabers, L. P.; et al. DNA damage induces p53-independent

apoptosis through ribosome stalling. *Science* **2024**, *384* (6697), 785-792. DOI: doi:10.1126/science.adh7950

(17) Basu, A. K.; Essigmann, J. M. Establishing linkages among DNA damage, mutagenesis, and genetic diseases. *Chem. Res. Toxicol.* **2022**, *35* (10), 1655-1675. DOI: 10.1021/acs.chemrestox.2c00155

(18) Welch, G.; Tsai, L. H. Mechanisms of DNA damage-mediated neurotoxicity in neurodegenerative disease. *EMBO Rep.* **2022**, *23* (6), EMBR202154217. DOI: 10.15252/embr.202154217

(19) Tsao, N.; Schärer, O. D.; Mosammaparast, N. The complexity and regulation of repair of alkylation damage to nucleic acids. *Crit. Rev. Biochem. Mol. Biol.* **2021**, *56* (2), 125-136. DOI: 10.1080/10409238.2020.1869173

(20) Chen, P. S.; Lin, S. C.; Tsai, S. J. Complexity in regulating microRNA biogenesis in cancer. *Exp Biol Med (Maywood)* **2020**, *245* (5), 395-401. DOI: 10.1177/1535370220907314

(21) Li, Z.; Pearlman, A. H.; Hsieh, P. DNA mismatch repair and the DNA damage response. *DNA Repair* **2016**, *38*, 94-101. DOI: https://doi.org/10.1016/j.dnarep.2015.11.019

(22) Spivak, G. Nucleotide excision repair in humans. *DNA Repair* **2015**, *36*, 13-18. DOI: https://doi.org/10.1016/j.dnarep.2015.09.003

(23) Krejci, L.; Altmannova, V.; Spirek, M.; Zhao, X. Homologous recombination and its regulation. *Nucleic Acids Res.* **2012**, *40* (13), 5795-5818. DOI: 10.1093/nar/gks270 (acccessed 11/7/2025).

(24) Woodbine, L.; Gennery, A. R.; Jeggo, P. A. The clinical impact of deficiency in DNA non-homologous end-joining. *DNA Repair* **2014**, *16*, 84-96. DOI: https://doi.org/10.1016/j.dnarep.2014.02.011

(25) Gohil, D.; Sarker, A. H.; Roy, R. Base excision repair: Mechanisms and impact in biology, disease, and medicine. *Int. J. Mol. Sci.* **2023**, *24* (18), 14186.

(26) Nilsen, H.; Krokan, H. E. Base excision repair in a network of defence and tolerance. *Carcinogenesis* **2001**, *22* (7), 987-998. DOI: 10.1093/carcin/22.7.987 (acccessed 8/29/2025).

(27) Dylewska, M.; Dąbrowska, I.; Ćwiek, K.; Padoł, K.; Mielecki, D.; Sokołowska, B.; Poznański, J.; Maciejewska, A. M. AlkA glycosylase and AlkB dioxygenase constitute an effective protective system for endogenously arising acrolein: E. Coli AlkA glycosylase excises acrolein adduct to adenine. *J. Mol. Biol.* **2025**, *437* (2), 168912. DOI: https://doi.org/10.1016/j.jmb.2024.168912

(28) Hollis, T.; Ichikawa, Y.; Ellenberger, T. DNA bending and a flip-out mechanism for base excision by the helix-hairpin-helix DNA glycosylase, escherichia coli AlkA. *EMBO J* **2000**, *19* (4), 758-766. DOI: 10.1093/emboj/19.4.758

(29) Hardeland, U.; Bentele, M.; Jiricny, J.; Schär, P. The versatile thymine DNA-glycosylase: A comparative characterization of the human, Drosophila and fission yeast orthologs. *Nucleic Acids Res.* **2003**, *31* (9), 2261-2271. DOI: 10.1093/nar/gkg344 (acccessed 10/21/2025).

(30) Komakula, S. S. B.; Blaze, B.; Ye, H.; Dobrzyn, A.; Sampath, H. A novel role for the DNA repair enzyme 8-oxoguanine DNA glycosylase in adipogenesis. *Int. J. Mol. Sci.* **2021**, *22* (3), 1152.

(31) Hua, A. B.; Sweasy, J. B. Functional roles and cancer variants of the bifunctional glycosylase NEIL2. *Environ. Mol. Mutagen.* **2024**, *65* (S1), 40-56. DOI: https://doi.org/10.1002/em.22555

(32) Sugimoto, Y.; Masuda, Y.; Iwai, S.; Miyake, Y.; Kanao, R.; Masutani, C. Novel mechanisms for the removal of strong replication-blocking HMCES- and thiazolidine-DNA adducts in humans. *Nucleic Acids Res.* **2023**, *51* (10), 4959-4981. DOI: 10.1093/nar/gkad246 (acccessed 12/17/2025).

(33) Hashimoto, H.; Zhang, X.; Cheng, X. Excision of thymine and 5-hydroxymethyluracil by the MBD4 DNA glycosylase domain: Structural basis and implications for active DNA demethylation. *Nucleic Acids Res.* **2012**, *40* (17), 8276-8284. DOI: 10.1093/nar/gks628 (acccessed 1/14/2025).

(34) Tubbs, J. L.; Tainer, J. A. DNA damage: Alkylation. In *Encyclopedia of biological chemistry (second edition)*, Lennarz, W. J., Lane, M. D. Eds.; Academic Press, 2013; pp 9-15.

(35) Wyatt, M. D.; Allan, J. M.; Lau, A. Y.; Ellenberger, T. E.; Samson, L. D. 3-methyladenine DNA glycosylases: Structure, function, and biological importance. *BioEssays* **1999**, *21* (8), 668-676. DOI: https://doi.org/10.1002/(SICI)1521-1878(199908)21:8<668::AID-BIES6>3.0.CO;2-D

(36) Bowman, B. R.; Lee, S.; Wang, S.; Verdine, G. L. Structure of escherichia coli AlkA in complex with undamaged DNA*. *J. Biol. Chem.* **2010**, *285* (46), 35783-35791. DOI: https://doi.org/10.1074/jbc.M110.155663

(37) Lim, C. S. Q.; Ha, K. P.; Clarke, R. S.; Gavin, L.-A.; Cook, D. T.; Hutton, J. A.; Sutherell, C. L.; Edwards, A. M.; Evans, L. E.; Tate, E. W.; et al. Identification of a potent small-molecule inhibitor of bacterial DNA repair that potentiates quinolone antibiotic activity in methicillin-resistant staphylococcus aureus. *Biorg. Med. Chem.* **2019**, *27* (20), 114962. DOI: https://doi.org/10.1016/j.bmc.2019.06.025

(38) Labahn, J.; Schärer, O. D.; Long, A.; Ezaz-Nikpay, K.; Verdine, G. L.; Ellenberger, T. E. Structural basis for the excision repair of alkylation-damaged DNA. *Cell* **1996**, *86* (2), 321-329. DOI: https://doi.org/10.1016/S0092-8674(00)80103-8

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

(40) Przybylski, J. L.; Wetmore, S. D. A QM/QM investigation of the hUNG2 reaction surface: The untold tale of a catalytic residue. *Biochemistry* **2011**, *50* (19), 4218-4227. DOI: 10.1021/bi2003394

(41) Stivers, J. T.; Jiang, Y. L. A mechanistic perspective on the chemistry of DNA repair glycosylases. *Chem. Rev.* **2003**, *103* (7), 2729-2759. DOI: 10.1021/cr010219b

(42) Rutledge, L. R.; Wetmore, S. D. Modeling the chemical step utilized by human alkyladenine DNA glycosylase: A concerted mechanism aids in selectively excising damaged purines. *J. Am. Chem. Soc.* **2011**, *133* (40), 16258-16269. DOI: 10.1021/ja207181c

(43) 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. DOI: 10.1021/acs.jpcb.7b09646

(44) Schärer, O. D.; Nash, H. M.; Jiricny, J.; Laval, J.; Verdine, G. L. Specific binding of a designed pyrrolidine abasic site analog to multiple DNA glycosylases*. *J. Biol. Chem.* **1998**, *273* (15), 8592-8597. DOI: https://doi.org/10.1074/jbc.273.15.8592

(45) Russelburg, L. P.; O'Shea Murray, V. L.; Demir, M.; Knutsen, K. R.; Sehgal, S. L.; Cao, S.; David, S. S.; Horvath, M. P. Structural basis for finding OG lesions and avoiding undamaged G by the DNA glycosylase MutY. *ACS Chem. Biol.* **2020**, *15* (1), 93-102. DOI: 10.1021/acschembio.9b00639

(46) Woods, R. D.; O'Shea, V. L.; Chu, A.; Cao, S.; Richards, J. L.; Horvath, M. P.; David, S. S. Structure and stereochemistry of the base excision repair glycosylase MutY reveal a mechanism similar to retaining glycosidases. *Nucleic Acids Res.* **2016**, *44* (2), 801-810. DOI: 10.1093/nar/gkv1469

(47) Nikkel, D. J.; Wetmore, S. D. Distinctive formation of a DNA–protein cross-link during the repair of DNA oxidative damage: Insights into human disease from MD simulations and QM/MM calculations. *J. Am. Chem. Soc.* **2023**, *145* (24), 13114-13125. DOI: 10.1021/jacs.3c01773

(48) Diao, W.; Farrell, J. D.; Wang, B.; Ye, F.; Wang, Z. Preorganized internal electric field promotes a double-displacement mechanism for the adenine excision reaction by adenine DNA glycosylase. *J. Phys. Chem. B* **2023**, *127* (40), 8551-8564. DOI: 10.1021/acs.jpcb.3c04928

(49) Diao, W.; Peng, J.; Fu, X.; Ye, F.; Wang, Z. Preorganized internal electric field enables nucleophilic attack of a nonactivated water molecule in thymine DNA glycosylase. *J. Am. Chem. Soc.* **2025**, *147* (29), 25919-25930. DOI: 10.1021/jacs.5c08420

(50) Unno, M.; Morikawa, M.; Sychrovský, V.; Koga, M.; Minowa, N.; Komuro, S.; Shimizu, M.; Fukuta, M.; Tsuyuguchi, F.; Mano, H.; et al. Capturing a glycosylase reaction intermediate in DNA repair by freeze-trapping of a pH-responsive hOGG1 mutant. *Nucleic Acids Res.* **2025**, *53* (14). DOI: 10.1093/nar/gkaf718 (acccessed 8/25/2025).

(51) Kaur, R.; Nikkel, D. J.; Wetmore, S. D. Computational studies of DNA repair: Insights into the function of monofunctional DNA glycosylases in the base excision repair pathway. *WIREs Comput. Mol. Sci.* **2020**, *10* (5), e1471. DOI: 10.1002/wcms.1471 (acccessed 2020/09/16).

(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. DOI: 10.1021/ja507934g

(53) Kreppel, A.; Blank, I. D.; Ochsenfeld, C. Base-independent DNA base-excision repair of 8-oxoguanine. *J. Am. Chem. Soc.* **2018**, *140* (13), 4522-4526. DOI: 10.1021/jacs.7b11254

(54) Sadeghian, K.; Ochsenfeld, C. Unraveling the base excision repair mechanism of human DNA glycosylase. *J. Am. Chem. Soc.* **2015**, *137* (31), 9824-9831. DOI: 10.1021/jacs.5b01449

(55) Naydenova, E.; Roßbach, S.; Ochsenfeld, C. QM/MM study of the uracil DNA glycosylase reaction mechanism: A competition between Asp145 and His148. *J. Chem. Theory Comput.* **2019**, *15* (8), 4344-4350. DOI: 10.1021/acs.jctc.8b01305

(56) Melayikandy, S.; Krishnan, A.; Varghese, A.; Jaber Sathik Rifayee, S. B.; Waheed, S. O.; Ramanan, R.; Li, D.; Christov, C. Z.; Karabencheva-Christova, T. G. Catalytic mechanism of the bacterial non-heme Fe(II) and 2-oxoglutarate dependent enzyme AlkB with single-stranded DNA containing complex guanine adducts. *Inorg. Chem.* **2025**, *64* (30), 15650-15666. DOI: 10.1021/acs.inorgchem.5c02176

(57) Wang, W.-J.; Wang, T.; Zhao, Y.; Li, B.-N.; Chen, D.-Z. Theoretical insights into N-glycoside bond cleavage of 5-carboxycytosine by thymine DNA glycosylase: A QM/MM study. *J. Phys. Chem. B* **2024**, *128* (19), 4621-4630. DOI: 10.1021/acs.jpcb.4c00221

(58) 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* (48), 8753-8765. DOI: 10.1021/bi401310w

(59) 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. DOI: 10.1021/ja711363s

(60) Tiwari, S.; Agnihotri, N.; Mishra, P. C. Quantum theoretical study of cleavage of the glycosidic bond of 2′-deoxyadenosine: Base excision-repair mechanism of DNA by MutY. *J. Phys. Chem. B* **2011**, *115* (12), 3200-3207. DOI: 10.1021/jp1109256

(61) 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. DOI: 10.1021/ja301714j

(62) Trasviña-Arenas, C. H.; Dissanayake, U. C.; Tamayo, N.; Hashemian, M.; Lin, W.-J.; Demir, M.; Hoyos-Gonzalez, N.; Fisher, A. J.; Cisneros, G. A.; Horvath, M. P.; et al. Structure of human MUTYH and functional profiling of cancer-associated variants reveal an allosteric network between its [4Fe-4S] cluster cofactor and active site required for DNA repair. *Nat. Commun.* **2025**, *16* (1), 3596. DOI: 10.1038/s41467-025-58361-w

(63) Shigdel, U. K.; Ovchinnikov, V.; Lee, S.-J.; Shih, J. A.; Karplus, M.; Nam, K.; Verdine, G. L. The trajectory of intrahelical lesion recognition and extrusion by the human 8-oxoguanine DNA glycosylase. *Nat. Commun.* **2020**, *11* (1), 4437. DOI: 10.1038/s41467-020-18290-2

(64) O'Brien, P. J.; Ellenberger, T. The escherichia coli 3-methyladenine DNA glycosylase AlkA has a remarkably versatile active site*. *J. Biol. Chem.* **2004**, *279* (26), 26876-26884. DOI: https://doi.org/10.1074/jbc.M403860200

(65) Jiang, D.; Zhang, L.; Dong, K.; Gong, Y.; Oger, P. Biochemical characterization and mutational studies of a novel 3-methlyadenine DNA glycosylase II from the hyperthermophilic thermococcus gammatolerans. *DNA Repair* **2021**, *97*, 103030. DOI: https://doi.org/10.1016/j.dnarep.2020.103030

(66) Schrempf, A.; Slyskova, J.; Loizou, J. I. Targeting the DNA repair enzyme polymerase θ in cancer therapy. *Trends Cancer* **2021**, *7* (2), 98-111. DOI: https://doi.org/10.1016/j.trecan.2020.09.007

(67) Duskova, K.; Lejault, P.; Benchimol, É.; Guillot, R.; Britton, S.; Granzhan, A.; Monchaud, D. DNA junction ligands trigger DNA damage and are synthetic lethal with DNA repair inhibitors in cancer cells. *J. Am. Chem. Soc.* **2020**, *142* (1), 424-435. DOI: 10.1021/jacs.9b11150

(68) Fuso Nerini, I.; Roca, E.; Mannarino, L.; Grosso, F.; Frapolli, R.; D'Incalci, M. Is DNA repair a potential target for effective therapies against malignant mesothelioma? *Cancer Treat. Rev.* **2020**, *90*, 102101. DOI: https://doi.org/10.1016/j.ctrv.2020.102101

(69) Mechetin, G. V.; Endutkin, A. V.; Diatlova, E. A.; Zharkov, D. O. Inhibitors of DNA glycosylases as prospective drugs. *Int. J. Mol. Sci.* **2020**, *21* (9), 3118.

(70) Jiang, M.; Jia, K.; Wang, L.; Li, W.; Chen, B.; Liu, Y.; Wang, H.; Zhao, S.; He, Y.; Zhou, C. Alterations of DNA damage repair in cancer: From mechanisms to applications. *Ann. Transl. Med.* **2020**, *8* (24), 1685.

(71) Gad, H.; Koolmeister, T.; Jemth, A.-S.; Eshtad, S.; Jacques, S. A.; Ström, C. E.; Svensson, L. M.; Schultz, N.; Lundbäck, T.; Einarsdottir, B. O.; et al. MTH1 inhibition eradicates cancer by preventing sanitation of the dNTP pool. *Nature* **2014**, *508* (7495), 215-221. DOI: 10.1038/nature13181

(72) Huber, K. V. M.; Salah, E.; Radic, B.; Gridling, M.; Elkins, J. M.; Stukalov, A.; Jemth, A.-S.; Göktürk, C.; Sanjiv, K.; Strömberg, K.; et al. Stereospecific targeting of MTH1 by (S)-crizotinib as an anticancer strategy. *Nature* **2014**, *508* (7495), 222-227. DOI: 10.1038/nature13194

(73) Helleday, T.; Petermann, E.; Lundin, C.; Hodgson, B.; Sharma, R. A. DNA repair pathways as targets for cancer therapy. *Nat. Rev. Cancer* **2008**, *8* (3), 193-204. DOI: 10.1038/nrc2342

(74) Schrodinger, LLC. The PyMOL molecular graphics system, version 1.8. 2015.

(75) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; III, T. E. C.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; et al. AMBER 2018. **2018**, University of California, San Fransisco.

(76) Li, P.; Song, L. F.; Merz, K. M., Jr. Systematic parameterization of monovalent ions employing the nonbonded model. *J. Chem. Theory Comput.* **2015**, *11* (4), 1645-1657. DOI: 10.1021/ct500918t

(77) Schmit, J. D.; Kariyawasam, N. L.; Needham, V.; Smith, P. E. SItcap: A simple method for calculating the number of ions needed for MD simulation. *J. Chem. Theory Comput.* **2018**, *14* (4), 1823-1827. DOI: 10.1021/acs.jctc.7b01254

(78) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696-3713. DOI: 10.1021/acs.jctc.5b00255

(79) Galindo-Murillo, R.; Robertson, J. C.; Zgarbova, M.; Sponer, J.; Otyepka, M.; Jurecka, P.; Cheatham, T. E. Assessing the current state of amber force field modifications for DNA. *J. Chem. Theory Comput.* **2016**, *12* (8), 4114-4127. DOI: 10.1021/acs.jctc.6b00186

(80) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; et al. *Gaussian 16 rev. B.01*; Gaussian, Inc., 2016.

(81) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *J. Phys. Chem.* **1993**, *97* (40), 10269-10280. DOI: 10.1021/j100142a004

(82) Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. The R.E.D. Tools: Advances in RESP and ESP charge derivation and force field library building. *Phys. Chem. Chem. Phys.* **2010**, *12* (28), 7821-7839. DOI: 10.1039/c0cp00111b

(83) Vanquelef, E.; Simon, S.; Marquant, G.; Garcia, E.; Klimerak, G.; Delepine, J. C.; Cieplak, P.; Dupradeau, F.-Y. R.E.D. Server: A web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res.* **2011**, *39* (suppl_2), W511-W517. DOI: 10.1093/nar/gkr288 (acccessed 10/11/2025).

(84) *R.E.D. Python: Object oriented programming for amber force fields*; Université de Picardie - Jules Verne, Sanford Burnham Prebys Medical Discovery Institute, 2013.

(85) He, X.; Man, V. H.; Yang, W.; Lee, T.-S.; Wang, J. A fast and high-quality charge model for the next generation general AMBER force field. *J. Chem. Phys.* **2020**, *153* (11). DOI: 10.1063/5.0019056 (acccessed 9/26/2025).

(86) 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* (W1), W537-W541. DOI: 10.1093/nar/gks375 (acccessed 8/9/2023).

(87) Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; et al. The ONIOM method and its applications. *Chem. Rev.* **2015**, *115* (12), 5678-5796. DOI: 10.1021/cr5004419

(88) Vennelakanti, V.; Nazemi, A.; Mehmood, R.; Steeves, A. H.; Kulik, H. J. Harder, better, faster, stronger: Large-scale QM and QM/MM for predictive modeling in enzymes and proteins. *Curr. Opin. Struct. Biol.* **2022**, *72*, 9-17. DOI: https://doi.org/10.1016/j.sbi.2021.07.004

(89) Kar, R. K. Benefits of hybrid QM/MM over traditional classical mechanics in pharmaceutical systems. *Drug Discov. Today* **2023**, *28* (1), 103374. DOI: https://doi.org/10.1016/j.drudis.2022.103374

(90) Sgrignani, J.; Magistrato, A. QM/MM MD simulations on the enzymatic pathway of the human flap endonuclease (hFEN1) elucidating common cleavage pathways to RNase H enzymes. *ACS Catal.* **2015**, *5* (6), 3864-3875. DOI: 10.1021/acscatal.5b00178

(91) Pilbák, S.; Farkas, Ö.; Poppe, L. Mechanism of the tyrosine ammonia lyase reaction—tandem nucleophilic and electrophilic enhancement by a proton transfer. *Chem. Eur. J.* **2012**, *18* (25), 7793-7802. DOI: https://doi.org/10.1002/chem.201103662

(92) Hu, X.; Hu, H.; Melvin, J. A.; Clancy, K. W.; McCafferty, D. G.; Yang, W. Autocatalytic intramolecular isopeptide bond formation in gram-positive bacterial pili: A QM/MM simulation. *J. Am. Chem. Soc.* **2011**, *133* (3), 478-485. DOI: 10.1021/ja107513t

(93) Wu, S.; Xu, D.; Guo, H. QM/MM studies of monozinc β-lactamase CphA suggest that the crystal structure of an enzyme−intermediate complex represents a minor pathway. *J. Am. Chem. Soc.* **2010**, *132* (51), 17986-17988. DOI: 10.1021/ja104241g

(94) Stevens, D. R.; Hammes-Schiffer, S. Exploring the role of the third active site metal ion in DNA polymerase η with QM/MM free energy simulations. *J. Am. Chem. Soc.* **2018**, *140* (28), 8965-8969. DOI: 10.1021/jacs.8b05177

(95) Kaur, R.; Nikkel, D. J.; Wetmore, S. D. Mechanism of nucleic acid phosphodiester bond cleavage by human Endonuclease V: MD and QM/MM calculations reveal a versatile metal dependence. *J. Phys. Chem. B* **2024**, *128* (39), 9455-9469. DOI: 10.1021/acs.jpcb.4c05846

(96) Taylor, E. L.; O'Brien, P. J. Kinetic mechanism for the flipping and excision of 1,N6-ethenoadenine by AlkA. *Biochemistry* **2015**, *54* (3), 898-908. DOI: 10.1021/bi501356x

(97) 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. U.S.A* **2000**, *97* (10), 5083-5088. DOI: doi:10.1073/pnas.97.10.5083

(98) Schramm, V. L. Transition states, analogues, and drug development. *ACS Chem. Biol.* **2013**, *8* (1), 71-81. DOI: 10.1021/cb300631k

(99) Pushing cancer over the edge. *Chemical & Engineering News Archive* **2013**, *91* (24), 13-18. DOI: 10.1021/cen-09124-cover

(100) Zhang, Z.; Wu, Z.; Shi, X.; Guo, D.; Cheng, Y.; Gao, J.; Liu, L.; Liu, W.; Liang, L.; Peng, L.; et al. Research progress in human AP endonuclease 1: Structure, catalytic mechanism, and inhibitors. *Curr Protein Pept Sci* **2022**, *23* (2), 77-88. DOI: 10.2174/1389203723666220406132737

(101) Brinkmeyer, M. K.; Pope, M. A.; David, S. S. Catalytic contributions of key residues in the adenine glycosylase MutY revealed by pH-dependent kinetics and cellular repair assays. *Chem. Biol. (Oxford, U. K.)* **2012**, *19* (2), 276-286. DOI: 10.1016/j.chembiol.2011.11.011

(102) Berdal, K. G.; Johansen, R. F.; Seeberg, E. Release of normal bases from intact DNA by a native DNA repair enzyme. *EMBO J* **1998**, *17* (2), 363-367-367. DOI: https://doi.org/10.1093/emboj/17.2.363 (acccessed 2025/10/22).

(103) Michaels, M. L.; Tchou, J.; Grollman, A. P.; Miller, J. H. A repair system for 8-oxo-7,8-dihydrodeoxyguanine. *Biochemistry* **1992**, *31* (45), 10964-10968. DOI: 10.1021/bi00160a004

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

(105) Lenz, S. A. P.; Wetmore, S. D. Evaluating the substrate selectivity of alkyladenine DNA glycosylase: The synergistic interplay of active site flexibility and water reorganization. *Biochemistry* **2016**, *55* (5), 798-808. DOI: 10.1021/acs.biochem.5b01179

(106) 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. DOI: https://doi.org/10.1074/jbc.M312232200

(107) Kumar, K.; Woo, S. M.; Siu, T.; Cortopassi, W. A.; Duarte, F.; Paton, R. S. Cation–π interactions in protein–ligand binding: Theory and data-mining reveal different roles for lysine and arginine. *Chem. Sci.* **2018**, *9* (10), 2655-2665. DOI: 10.1039/C7SC04905F

Chapter 5: Insights into the Debated Lyase Mechanism of Bifunctional DNA Glycosylases from QM/MM MD Simulations: The Case Study of DNA Oxidative Damage Repair by hOGG1

5.1. Introduction

One of the most common DNA damage pathways in human cells is the oxidation of nucleobases resulting from exposure to reactive oxygen species (ROS).[1] ROS can arise from endogenous (e.g., cell metabolism and inflammatory response)[2, 3] and exogenous (e.g., pollutants, radiation, heavy metals, tobacco, and e-cigarettes)[4-6] sources and are linked to many human health conditions including diabetes,[7, 8] cardiovascular disease,[9, 10] neurodegeneration,[11, 12] and cancer.[13-19] These disorders are a consequence of the damage resulting in changed base-pairing specificity, leading to mutations.[1, 20] The most prevalent form of oxidative DNA damage is 8-oxo-7,8-dihyroguanine (8oG), which is estimated to arise approximately 1000–7000 times per cell per day.[21] 8oG is known to be mutagenic, forming a base pair with adenine and resulting in a G:C to T:A transverse mutation upon several rounds of replication.[22-24] To prevent mutagenic outcomes, human cells use 8-oxoguanine DNA glycosylase (hOGG1) to detect and remove 8oG damage as part of the base excision repair (BER) pathway.[25]

hOGG1 is a bifunctional glycosylase that cleaves the glycosidic bond that attaches the damaged 8oG base to deoxyribose and cleaves the 3′- (and sometimes 5′-) phosphodiester bond, with the resulting enal being a substrate for enzymes that continue the BER pathway.[26] Due to the vital role hOGG1 plays in cells, several experimental[27, 28] and computational[29-31] studies have investigated the initial glycosidic bond cleavage, which points to a mechanism where K249 attacks C1′ to displace the 8oG while D268 catalyzes a ring-opening step that breaks the C1′–O4′ bond, resulting in an imine crosslinked intermediate (Figure 5.1). However, there is little information available about the catalytic mechanism of the rate-limiting β-lyase step.[32] While formation of the phosphate and enal products necessitates hydrolysis of the imine crosslink and elimination of the 3′ phosphate, the order of these steps is unclear. Indeed, even the base that facilitates C2′ proton abstraction to initiate the 3′ elimination has not been conclusively identified, with conflicting proposals present in the literature.[26, 33]

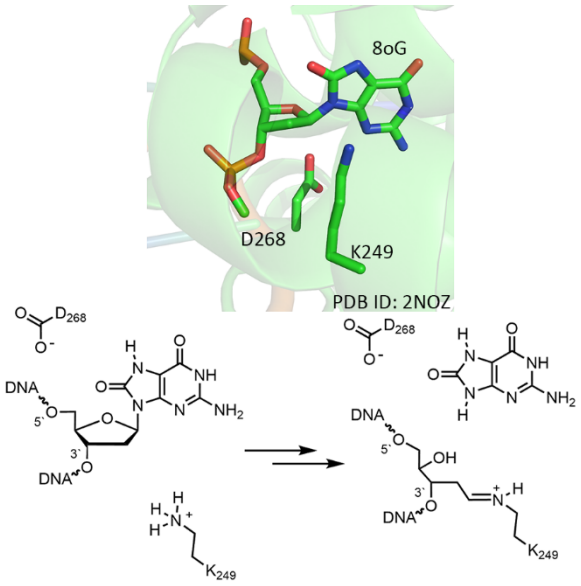


Figure 5.1. Crystal structure of hOGG1−DNA complex (top) and hOGG1 pathway for glycosidic bond cleavage (bottom).

The first proposed mechanism for lyase activity is based on a crystal structure of an hOGG1 borohydride trapped crosslinked intermediate with free 8oG resolved in the active site (PDB ID: 1HU0, Figure 5.2A).[26] In this pathway, the cleaved anionic 8oG from the deglycosylation step abstracts a C2′ hydrogen to catalyze the elimination reaction (Figure 5.2B).[26] This suggestion was supported by the observation that 8oG and 8oG analogues (e.g., 8-bromoguanine) enhance hOGG1 activity.[26] However, it has also been proposed that the 8oG anion would likely either be protonated or leave the active site before the β-lyase reaction can occur.[34] Furthermore, compounds have been found that are not basic and/or are not structurally similar to 8oG but still increase hOGG1 activity, which indicates that 8oG may be an allosteric regulator rather than participating in catalysis.[33] Therefore, in a second proposed mechanism, D268 acts as the base for elimination (Figure 5.2C),[33] which is supported by a 68-fold reduction in catalytic activity upon D268N mutation.[35] Nevertheless, the D268Q mutation does not result in a detectable decrease in activity,[35] indicating D268 may not serve a vital role in catalysis. Therefore, no proposed mechanism to date accounts for all experimental data related to the β-lyase step of hOGG1, and additional research is needed to develop and validate a comprehensive catalytic mechanism.

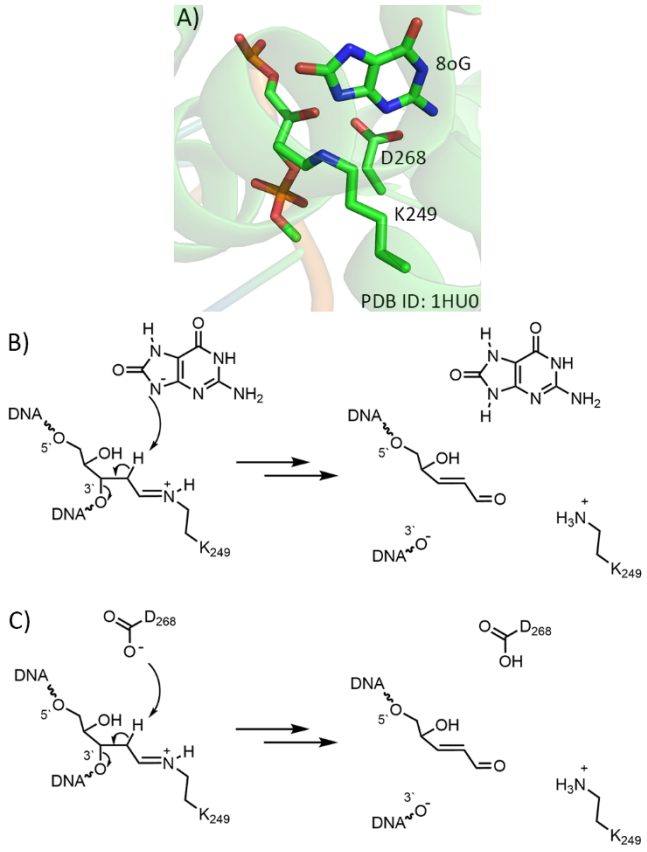


Figure 5.2. A) Crystal structure of the borohydride-trapped hOGG1 crosslinked intermediate with 8oG retained in the active site. Proposed β-lyase pathway catalyzed by hOGG1 involving B) product or C) D268-assisted catalysis.

Computational techniques can elucidate atomic level details of catalytic mechanisms and distinguish the preferred enzymatic pathway among many possibilities.[36-46] Indeed, an abundance of computational studies have considered the mechanism of action of the hydrolysis reaction catalyzed by monofunctional DNA glycosylases.[43, 46-62] Furthermore, molecular dynamics (MD) simulations have been combined with quantum mechanics/molecular mechanics (QM/MM) methods to characterize the MutY catalytic pathway, a monofunctional glycosylase that cleaves the glycosidic bond by forming a crosslinked intermediate.[53] Computational methods have also been used to study the deglycosylation step facilitated by bifunctional glycosylases[58-62] including hOGG1,[31, 46, 63] while hOGG1 substrate recognition and accommodation has been investigated using MD simulations.[28, 64-68] However, computational work on the β-lyase step of any bifunctional

glycosylase remains limited to a single small model QM study (54 atoms), which concluded that the preferred elimination mechanism will be dependent on the enzyme active site.[69]

To fill in gaps in our knowledge of the hOGG1 mechanism of action by capitalizing on the ability of computational techniques to provide key missing information about enzyme function, a combination of long-timescale classical molecular dynamics (MD) simulations and short-timescale quantum mechanics/molecular mechanics MD (QM/MM MD) simulations are used in the present work to explore the β-lyase activity of hOGG1. First, starting from the hOGG1−crosslinked substrate complex, the kinetics of 8oG unbinding from the hOGG1 active site are investigated using umbrella sampling to determine whether catalysis can occur prior to 8oG departure. Next, MD and QM/MM MD umbrella sampling simulations are used to systematically explore the dynamics of the hOGG1 active site in the presence of the crosslinked substrate and to map multiple catalytic pathways that differ in the base facilitating elimination and the order of the elimination and imine hydrolysis steps. In all cases, both neutral and cationic imine crosslinks were considered since the pKa (5−7) suggests both states are feasible in a cellular environment. The preferred mechanism of action correlates with all available experimental kinetic[33, 34, 70] and mutagenic[35] data to date. This is the first time the β-lyase step has been characterized for a bifunctional glycosylase, revealing previously unconsidered roles for active site residues. The improved understanding of the hOGG1 mechanism gained from this work promises to aid in the development of transition state analogue inhibitors[71-73] that can be used to treat inflammatory conditions[74, 75] and cancers,[75-78] and sets the stage for the design of allosteric hOGG1 activators for combating disorders caused by oxidative stress.[79, 80]

5.2. Methods

5.2.1. Classical MD simulations

Models for the hOGG1−crosslinked substrate complex were constructed based on a crystal structure of the borohydride-trapped intermediate with 8oG in the active site (PDB ID: 1HU0, Figure

5.2A).[26] Unresolved residues 80−82 were added using SWISS-MODEL.[81] Four models were generated that differ in the presence of 8oG in the active site and the imine crosslink protonation state. The crystallographic Ca^{2+} ion bound to the DNA backbone was removed in models containing 8oG in the active site. Imine crosslinks were modeled as the E isomer due to general thermodynamic favourability and a small model study showing the imine isomer negligibly impacts the elimination barrier.[69] Each structure was solvated with TIP4P-EW water molecules in a periodic rectangular prism with sides a minimum of 10 Å from the solute using the LEaP program in Amber 20.[82] Sodium ions were added to neutralize the solute, and NaCl was added to achieve a physiological salt concentration (150 mM). The Amber force field was used throughout, with protein residues described using ff14SB[83] and nucleic acid residues described using OL15.[84] Parameters for the cationic and neutral hOGG1−DNA crosslink were supplemented using the general AMBER forcefield (GAFF2),[85] while charges were obtained using the RESP charge fitting procedure in the pyRED program (Table D.1−D.6, Appendix D). Parameters for 8oG were taken from the literature with charges calculated using pyRED (Table D.7, Appendix D).[86]

All models were initially minimized in a stepwise manner, using 1000 steps of steepest decent followed by 1000 steps of conjugate gradient minimization. First, the solvent was minimized, while a 100 kcal mol^{-1} $Å^{-2}$ restraint was applied to the solute. Next, the restraint was removed from the solute hydrogen atoms. Subsequently, all solute atoms were minimized, while applying a 100 kcal mol^{-1} $Å^{-2}$ restraint to the solvent. Finally, the entire system was minimized with no restraints, using 1000 steps of steepest decent followed by 2000 steps of conjugate gradient minimization. Each system was then heated, starting from 10 K with the temperature increased by 50 K every 20 ps until 310 K was reached. The restraints on the solute were subsequently dropped from 25 kcal mol^{-1} $Å^{-2}$ to 5 kcal mol^{-1} $Å^{-2}$ at a rate of 5 kcal mol^{-1} $Å^{-2}$ per 20 ps. A final 20 ps equilibration step was completed while applying a solute restraint of 1.5 kcal mol^{-1} $Å^{-2}$. For these simulations, an NVT ensemble was used, with a collision frequency of 1 ps^{-1} and a Langevin thermostat. Following equilibration, five 1 μs MD production simulations were performed on each system using an NPT ensemble with a pressure of 1 bar (Berendsen barostat). Simulations were run using AMBER 2018 pmemd.cuda.[87]

Analysis was conducted using the cpptraj program in AmberTools20.[82] Unique active site conformations were identified based on all possible combinations of the ∠(C3′C2′C1′Nζ), ∠(C4′C3′C2′C1′), and ∠(D268CγC1′C2′Nζ) dihedral angles, and the ∠(D268CγC1′C2′) angle that occurred over the simulation (see Figures D.1 and D.2, Appendix D). The most occupied conformations that together accounted for at least 80% of the total simulation time were used to identify correlated structural features, permitting reduction in the number of parameters used to define each conformation. The entire simulation was then divided according to the reduced parameter set. Hydrogen-bond occupancies were calculated using a heavy-atom distance of less than 3.0 Å and a hydrogen-bond angle between 135° and 180°.

5.2.2. Umbrella sampling MD simulations

Post-equilibration models of hOGG1 bound to DNA containing a neutral or cationic imine crosslink with 8oG in the active site were used to initiate umbrella sampling simulations to investigate 8oG unbinding from the active site. The same parameters, and initial minimization, heating, and equilibration steps as described for the classical MD simulations were employed. The collective variable (CV) used for umbrella sampling was the distance between the C2′ of the crosslink deoxyribose and N9 of free 8oG. Umbrella windows had restraints centered on CV values that ranged from 3.5 Å to 12.75 Å and increased in increments of 0.25 Å, with each window consisting of a 50 ns simulation undergoing a 15 kcal mol^{-1} $Å^{-2}$ restraint. The structure after 10 ns for every second window was used as the starting point for the next two windows. The potential energy surface was generated using the WHAM method.[88, 89] Simulations were run using AMBER 2018 pmemd.cuda[87] and analysis was performed using the cpptraj program in AmberTools20.[82]

5.2.3. QM/MM MD simulations

Based on data from umbrella sampling MD simulations, QM/MM MD simulations were performed using the neutral and cationic imine crosslink models of hOGG1 with 8oG removed from

the active site. The QM region (44−45 atoms) consists of Asp268 truncated at Cβ, the nucleic acid portion of the crosslink truncated between C4′ and C5′ on the 5′ end of the crosslinking nucleotide and between C5′ and C4′ of the nucleotide 3′ to the crosslinking nucleotide, the protein portion of the crosslink truncated at Cβ, and two water molecules (Figure 5.3). The total charge of the QM region is −1 for the cationic crosslink and −2 for the neutral crosslink models. The QM region is treated using the B3LYP functional[90] due to its success in modeling other enzymes[91-93] including glycosylases.[94] The 6-31G(d) basis set was used with the electronic embedding scheme. Calculations were run using AMBER20[95] interfaced with Gaussian 16 (Rev. C.01).[96] Simulations initially used the same parameters, and minimization, heating, and equilibration steps as the classical MD simulations. After equilibration, the QM region was delimited and an additional 20 ps unrestrained equilibration was run. The last point from the equilibration step was then used as the starting structure for steered MD (SMD) simulations to consider the first reaction step of different catalytic pathways. To model the second reaction step, SMD simulations were initiated from the structure taken from the last frame of an umbrella sampling window to the corresponding product complex. The CVs as well as starting and ending points for each SMD simulation can be found in Table D.8. Frames taken throughout the SMD simulations were used as starting points for umbrella sampling calculations, with the chosen frame having a CV closest to that of the umbrella sampling window. CVs chosen as the center of restraints for umbrella sampling varied depending on the reaction being mapped (see Table D.9, Appendix D). Windows were incremented by 0.1 Å across the CV range considered and consisted of a 6 ps simulation with a restraint weight of 100 kcal mol^{-1} $Å^{-2}$. The potential energy surface was constructed using the WHAM method based on the last 5 ps of each window. Results were compared to an energy barrier of 92 kJ/mol estimated based on an experimental steady-state rate constant of 6.88 h^{-1} at a pH of 7.9 and temperature of 37 °C[70] using the Eyring equation.

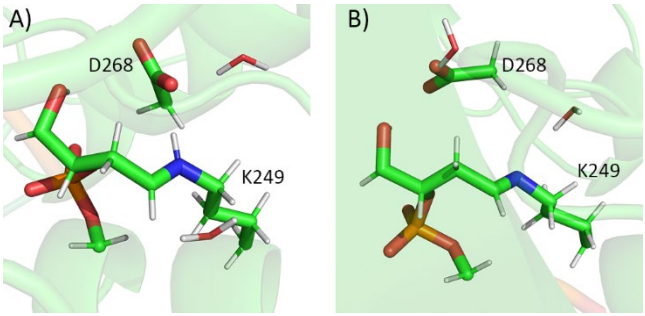


Figure 5.3. QM region for QM/MM MD simulations on models containing a A) cationic (45 atoms) or B) neutral (44 atoms) imine crosslink.

5.3. Results

5.3.1. 8oG rapidly unbinds from the hOGG1 active site, preempting participation in catalysis

One of the first proposals for the mechanism of hOGG1 catalyzed β-lyase activity involved C2′ abstraction by the 8oG anion generated in the preceding glycosylase reaction (Figure 5.2B).[26] This requires 8oG to remain in the active site, which is supported by a crystal structure of an hOGG1 borohydride-trapped intermediate (PDB ID: 1HU0, Figure 5.2A).[26] However, an experimental kinetic study indicated that under *in vivo* conditions 8oG would be unlikely to remain in the active site to participate in catalysis.[34] Therefore, umbrella sampling MD simulations were used to investigate the barrier for 8oG unbinding from the hOGG1 active site for both the neutral and cationic crosslink.

For both the neutral and cationic crosslinks models, the PES generated from umbrella sampling contains a minimum at a CV of 5.6 Å corresponding to the optimal position of 8oG inside the active site (Figure D.3, Appendix D). This is slightly longer than the observed separation in the borohydride-trapped intermediate crystal structure (r(N9−C2′) = 3.9 Å), which could result from the crystal structure containing an amine rather than imine crosslink. The barrier for 8oG anion departure (unbinding complete at a CV of ~12 Å) is low (5−15 kJ/mol depending on crosslink protonation state), with 8oG forming hydrogen bonds with N315, S147, and N43 (Figure D.4, Appendix D) as well as stacking interactions with F319 and H270 (Figure D.5, Appendix D) along the path leaving the active site. The barrier for 8oG unbinding is well below the experimental barrier for hOGG1 catalysis (92 kJ/mol),[70] indicating 8oG will leave the active site before the rate-limiting

elimination reaction. Furthermore, guanine is present in cells at a low concentration (~100 μM)[97] and only two or three bases per million guanines exist as 8oG[98] (expected cellular concentration ~0.25 nM). When coupled with the high pKa of 8oG (9.49 for the N9 proton),[99] this results in a single digit estimate for the number of anionic 8oG present in a cell at neutral pH. Thus, the concentration of anionic 8oG is too low to expect binding in the hOGG1 active site for participation in catalysis. With a low barrier for 8oG unbinding and low cellular concentration of anionic 8oG, it is not feasible for 8oG to be the base for C2′ proton abstraction. This proposal is consistent with hypotheses that 8oG acts as an allosteric activator for hOGG1 activity.[33]

5.3.2. D268 can be positioned to initiate elimination for a neutral crosslink

With 8oG anticipated to depart the enzyme active site, an hOGG1 residue must act as a base for β-elimination. D268 has been previously proposed to adopt this role.[33] Indeed, the D268N mutation results in a 68-fold reduction in catalytic activity.[35] Other enzymes have been reported to use aspartate/glutamate residues as a base for elimination (DesII[100], TFAM[101]) and mutational data suggests a glutamate plays a similar role for bifunctional T4 pyrimidine dimer glycosylase (T4 PDG).[102, 103] Although D268 is a promising general base for elimination, D268 is not close enough to the crosslink substrate for proton abstraction in the crystal structure of the hOGG1 crosslinked intermediate (r(D268−C2′) = 4.9 Å, PDB ID: 1HU0, Figure 5.2A).[26] Nevertheless, this crystal structure contains a reduced amine crosslink rather than the canonical imine substrate and the static nature of the crystal structure may not capture other possible orientations of D268. Therefore, to determine whether D268 can be positioned for C2′−H abstraction, classical MD simulations were initially performed on a model of the neutral imine crosslink.

Across the MD simulations on the hOGG1−DNA neutral crosslinked intermediate, four unique active site conformations were observed, with the primary differentiating criteria being the ∠(C5′C4′C3′C2′) and ∠(C3′C2′C1′Nζ) crosslink dihedral angles (Figure 5.4). In the most-occupied conformation (conformation 1N; 51% occupancy), D268 is positioned in the active site by transient hydrogen bonds to the backbones of V269 and M271, and O4′ of the ring-opened deoxyribose

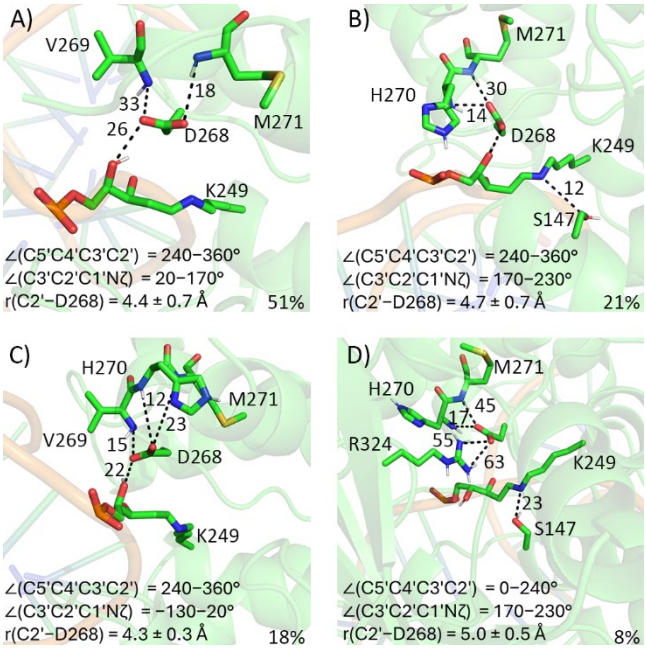


Figure 5.4. Representative structures from MD simulations on hOGG1 models containing a neutral imine crosslink, showing active site orientation, hydrogen-bond occupancies (percentage), key structural parameters (bottom left), and overall occupancy (bottom right) for conformation A) 1N, B) 2N, C) 3N, and D) 4N.

(Figure 5.4A). This places D268 near C2′ of the crosslink (r(C2′−D268) = 4.4 ± 0.7 Å), positioning the residue for proton abstraction (r(C2′−D268) < 4 Å) for the 31% of the time that this conformation is adopted. Water is also present in the active site for crosslink hydrolysis (normalized water density of 1.3 g/mL at a distance of 3.5 Å from the crosslink, Figure D.6A, Appendix D). In the third most occupied conformation (conformation 3N; 18% occupancy), D268 adopts similar hydrogen bonds and orientation with respect to the crosslink (r(C2′−D268) = 4.3 ± 0.3 Å; < 4 Å for 33% of the simulation time) despite having a unique crosslink orientation (Figure 5.4C). Furthermore, water remains available around the crosslink in conformation 3N (normalized water density of 1.1 g/mL at a distance of 3.5 Å, Figure D.6A, Appendix D). In the remaining two conformations (conformations 2N (21%) and 4N (8%)), the active site hydrogen bonding with D268 is altered, which results in a longer distance to C2′ (average r(C2′−D268) = 4.7 – 5.0 Å) and poor positioning for proton abstraction (< 12% occupancy of a suitable orientation). Overall, despite the flexibility of the active site, D268 is positioned to facilitate proton abstraction from C2′ of the neutral crosslink

for a significant portion of the total simulation, suggesting a catalytically-conducive orientation of the hOGG1 active site can be adopted in the presence of a neutral crosslink.

5.3.3. D268 cannot abstract a proton as the first step in the hOGG1 β-lyase activity toward a neutral imine crosslink

To determine whether D268-catalyzed phosphodiester bond cleavage is energetically feasible for the neutral crosslink and the preferred order of the elimination and imine hydrolysis steps, QM/MM MD simulations were initiated from a model built from the dominant active site conformation identified from classical MD simulations (conformer 1N). Specifically, a representative structure was chosen based on conformations with r(C2′–D268) < 4 Å to ensure D268 is optimally aligned for the reaction. From this model, the previously proposed mechanism in which D268 first abstracts a proton from C2′ to initiate the β-lyase activity (Figure 5.2C) was investigated.[33] The proton transfer

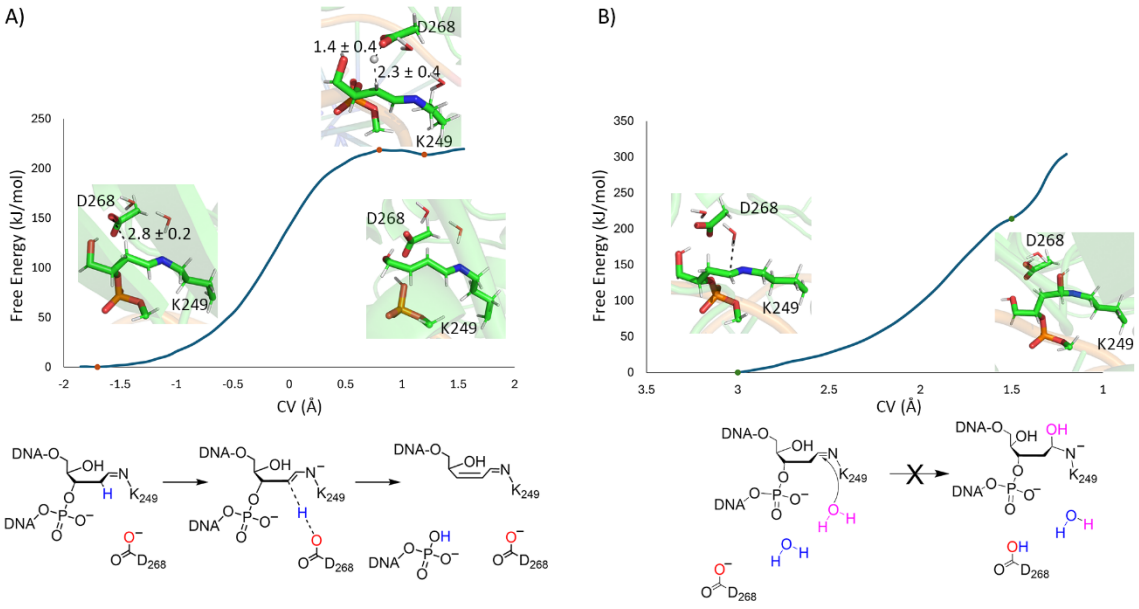


Figure 5.5. Potential energy surface and representative structures of stationary points for the D268-catalyzed A) elimination and B) hydrolysis of a neutral imine crosslink. A representative structure corresponding to a CV of 1.5 Å was shown as a representative complex along the hydrolysis pathway. CV defined in Table D.9.

occurs in one step (Figures 5.5A and D.7, Appendix D). In the reactant complex (CV ~ −1.7 Å), D268 falls 2.8 ± 0.2 Å from H2′. The transition state (CV ~ 0.8 Å) is late, with r(D268−H2′) reduced to 1.4 ± 0.4 Å and the C2′–H2′ bond extended to 2.3 ± 0.4 Å (Figures 5.5 and D.7, Appendix D). In the PC, the C2′–H2′ bond is cleaved, and H2′ is transferred from D268 to O3′ of the phosphate, which cleaves the O3′–C3′ bond. This process has a very large energy barrier (218.8 kJ/mol), significantly exceeding the experimental barrier for catalysis (92 kJ/mol).[70] The high energetic cost potentially arises from charge build up on Nζ over the reaction as evidence by weakening of the Nζ–C1′ bond in the transition state (~0.03 Å increase in bond length, Figure D.7, Appendix D). Since the energy barrier is too high for this process to occur on physiological time scales, D268-assisted elimination is not a feasible pathway for the β-lyase step associated with a neutral crosslink.

5.3.4. Hydrolysis of the neutral imine crosslink is not energetically viable, suggesting a neutral crosslink is not a substrate for hOGG1 β-lyase activity

As water surrounds the neutral crosslink in the hOGG1 active site (Figure D.6A, Appendix D), another possible β-lyase mechanism involves hydrolysis of the imine crosslink as the initial step in the hOGG1 mechanism of action. The feasibility of this pathway was investigated using the same QM/MM MD model as discussed for the previous D268-facilitated elimination. The RC occurs at a C1′–O_{WAT} distance of 3.0 Å, with a second water molecule forming a bridge between the nucleophilic water and D268 (Figure 5.5B and D.8, Appendix D). However, a stable PC could not be characterized, with the relative Gibbs energy steadily increasing as the water approaches C1′. At a CV of 1.5 Å, the C1′–O_{WAT} bond has formed and the proton from the water nucleophile transferred to D268 through the water bridge. Although this corresponds to an area of reduced slope on the PES, no minimum is present and the energy relative to the RC at this point on the PES is over 200 kJ/mol, reinforcing that hydrolysis is not feasible. When combined with the high barrier for D268 proton abstraction discussed in the previous section, the QM/MM MD simulations suggest that a neutral crosslink is not a substrate for hOGG1 β-lyase activity.

5.3.5. D268 is aligned with respect to a cationic crosslink to facilitate elimination

As discussed in section 5.1, the pKa range for a typical imine (5−7) suggests that the DNA–hOGG1 crosslinked species could be cationic at the near neutral pH of human cells. As was done for the neutral crosslink, classical MD simulations were initially performed to explore the conformational landscape of D268 with respect to the cationic imine crosslink. Four unique active site conformations were identified that primarily differ in the ∠(C3′C2′C1′Nζ) dihedral angle (Figure 5.6). The dominant conformation (conformation 1C) was adopted for 60% of the simulation time (Figure 5.6A). In addition to transient hydrogen bonds with O4′ of deoxyribose, and the backbones of V269, H270, and M271, D268 is positioned in the active site by a strong hydrogen bond to Nζ of the crosslink, placing the residue in close proximity to C2′ (r(C2′−D268) = 3.7 ± 0.5 Å; < 4 Å for 82% of the simulation time). Consequently, this conformation is well aligned for D268 to catalyze the

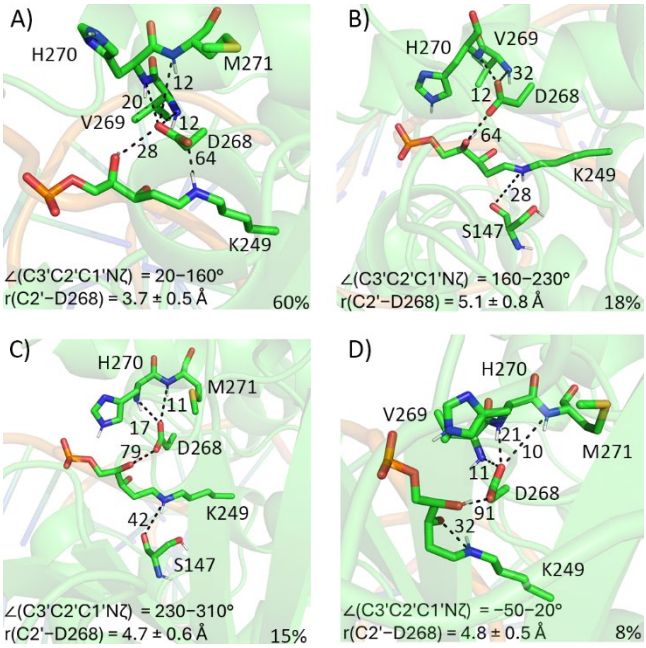


Figure 5.6. Representative structures from MD simulations on hOGG1 models containing a cationic imine crosslink, showing active site structure, hydrogen-bond occupancies, defining parameters (bottom left), and overall occupancy (bottom right) for conformation A) 1C, B) 2C, C) 3C, and D) 4C.

elimination reaction. Furthermore, water is available in the active site in this conformation (normalized water density of 1.6 g/mL at a distance of 3.5 Å from the crosslink, Figure D.6B,

Appendix D). In the remaining three conformations (conformations 2C (18%), 3C (15%) and 4C (8%)), the hydrogen bond between D268 and Nζ of the crosslink is absent, which results in improper D268 positioning for C2′ proton abstraction (average r(C2′−D268) = 4.7 – 5.1 Å). Nevertheless, D268 is aligned for C2′ proton abstraction for a significant portion of the total simulation time, suggesting D268 may help initiate the hOGG1 β-lyase mechanism.

5.3.6. D268 cannot abstract a proton from a cationic imine crosslink as the first step in hOGG1 β-lyase activity

To determine whether β-lyase activity on a cationic imine crosslink is energetically feasible, QM/MM MD calculations mapped potential catalytic pathways, using a model based on a representative structure of the dominant conformer (1C) identified from classical MD simulations. From this model, the previously proposed mechanism in which D268 first abstracts a proton from C2′ to initiate the β-lyase activity was investigated.[33] The proton transfer occurs in one step (Figure 5.7A and D.9, Appendix D). In the RC (CV = −1.55 Å), D268 is 2.6 ± 0.1 Å from H2′ has forms a hydrogen bond with Nζ of the crosslink, as observed in the classical MD simulations. The transition state (CV = 0.30 Å) is earlier than that for the neutral crosslink, with H2′ partially transferred to D268 (r(H2′–D268) = 1.2 ± 0.1 Å; r(C2′–H2′) = 1.5 ± 0.1 Å). As in the elimination mechanism mapped for the neutral crosslink, charge accumulates on Nζ, which weakens the Nζ–C1′ bond in the transition state (~0.04 Å increase in bond length). However, unlike for the neutral system, protonated Nζ can better accommodate the increased electron density, which results in a much lower barrier (140.3 kJ/mol). In the PC (CV = 1.50 Å), H2′ has transferred to D268, while a double bond forms between C2′ and C1′ (r(C2′–C1′) = 1.37 ± 0.04 Å) and the Nζ–C1′ bond is weakened (r(Nζ–C1′) = 1.36 ± 0.04 Å). Since phosphate elimination was not completed, a separate QM/MM MD simulation modeled proton transfer from D268 to O3′ of the phosphate to finish the reaction, which occurred in one step with a low barrier (32 kJ/mol, Figure D.10 and D.11, Appendix D). In the final product, the O3′−C3′ bond is cleaved, the phosphate protonated, the imine in the crosslink regenerated, and a double bond formed between C2′ and C3′.

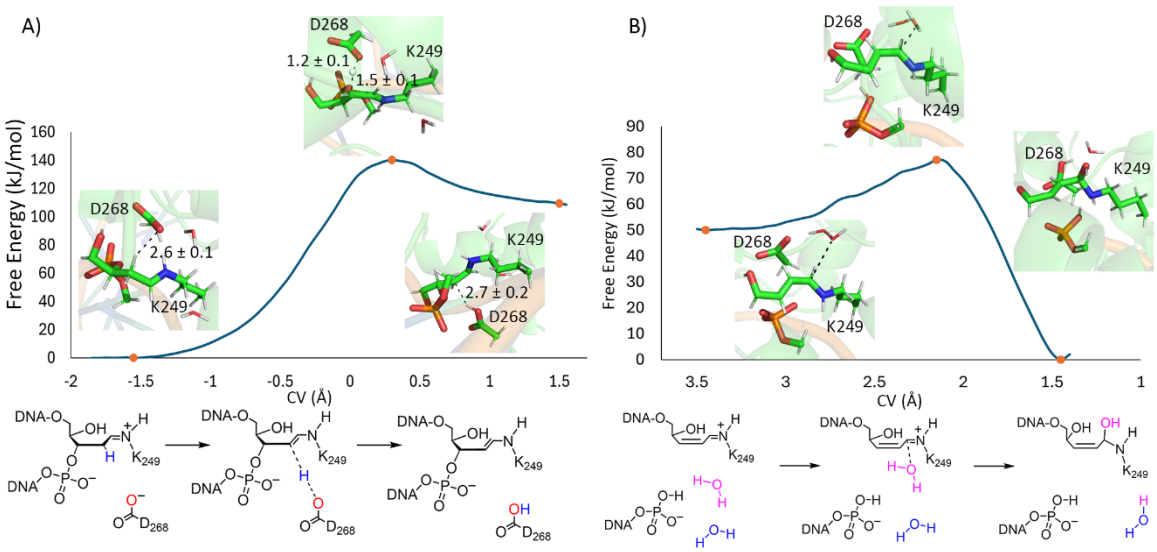


Figure 5.7. Potential energy surface and representative structures of stationary points for the D268-catalyzed A) elimination reaction and B) subsequent hydrolysis of a cationic imine crosslink. CV defined in Table D.9.

Finally, the product from the β-elimination path was used to model the imine hydrolysis step, which yields the enal product (Figure 5.7B and D.12, Appendix D). In the RC (CV = 3.45 Å), the nucleophilic water is positioned to attack C1′, with a water bridging D268 and the nucleophile to facilitate proton transfer (94% occupancy, Figure D.12, Appendix D). In the TS (CV = 2.15 Å), the reduced nucleophilic water−C1′ distance pulls the nucleophile farther from D268, which lowers the population of the water bridge (14% occupancy). In the PC (CV = 1.45 Å), a hydrogen atom from the nucleophilic water has been transferred to D268 through the water bridge and the crosslink has been reduced to an amine (r(Nζ–C1′) = 1.45 ± 0.04). The corresponding energy barrier for this reaction is low (26.5 kJ/mol). Overall, the elimination (140.3 kJ/mol barrier) is the rate-limiting step for this pathway. Although the QM/MM MD barrier is lower than that of any pathway for a neutral crosslink substrate, the energetic cost remains significantly greater than the experimental barrier (92 kJ/mol),[70] which indicates D268-assisted elimination is not a feasible first step for hOGG1 β-lyase activity regardless of the crosslink protonation state.

5.3.7. hOGG1 β-lyase activity initiated by hydrolysis of a cationic imine crosslink followed by D268-assisted elimination is catalytically viable

Although not discussed in the literature to date, there is potential for imine hydrolysis to initiate hOGG1 β-lyase activity. In the QM/MM MD RC (CV ~ 2.90 Å), a water bridges the nucleophilic water and D268 for 54% of the simulation time (Figures 5.8A and D.13, Appendix D). Nucleophilic attack of the water at C1′ results in a TS (CV ~ 1.85 Å) with a weakened C1′–Nζ bond (~0.04 Å increase), which is reduced to an amine in the PC (CV ~ 1.40 Å; r(Nζ–C1′) = 1.46 ± 0.05). The reaction has a barrier of 33.9 kJ/mol, which is below the experimental barrier for hOGG1 β-lyase activity[70] and indicates initial nucleophilic attack of a water is kinetically favorable.

Since conversion of the alcohol hydrolysis intermediate to an aldehyde is expected to be rapid as the initial nucleophilic attack of a water is generally rate-limiting for imine hydrolysis in neutral conditions,[104, 105] the aldehyde product was used as the starting point for the D268-catalyzed elimination step (Figure 5.8B and D.14, Appendix D). D268 is 2.8 ± 0.1 Å away from H2′ in the RC (CV = −1.80 Å), which decreases to 1.2 ± 0.1 Å in the TS (CV = 0.25) in which the H2′−C2′ bond is partially cleaved (r(C2′–H2′) = 1.5 ± 0.1). The aldehyde bond is weakened (~0.05 Å increase in bond length), indicating a buildup of charge that is stabilized by a hydrogen bond to K249 (77% occupancy). The energy barrier for this step is 92.7 kJ/mol. In the PC, H2′ has completely transferred to D268, while the C1′−O1′ bond approaches the distance expected for a C−O single bond (r(C1′−O1′) = 1.35 ± 0.04 Å). Completion of the elimination reaction requires proton transfer from D268 to the O3′ leaving group (Figure 5.8C and D.15, Appendix D), which is associated with a low barrier (19.8 kJ/mol), and results in a product containing a cleaved O3′−C3′ bond and a C2′−C3′ double bond (r(C2′−C3′) = 1.35 ± 0.02 Å).

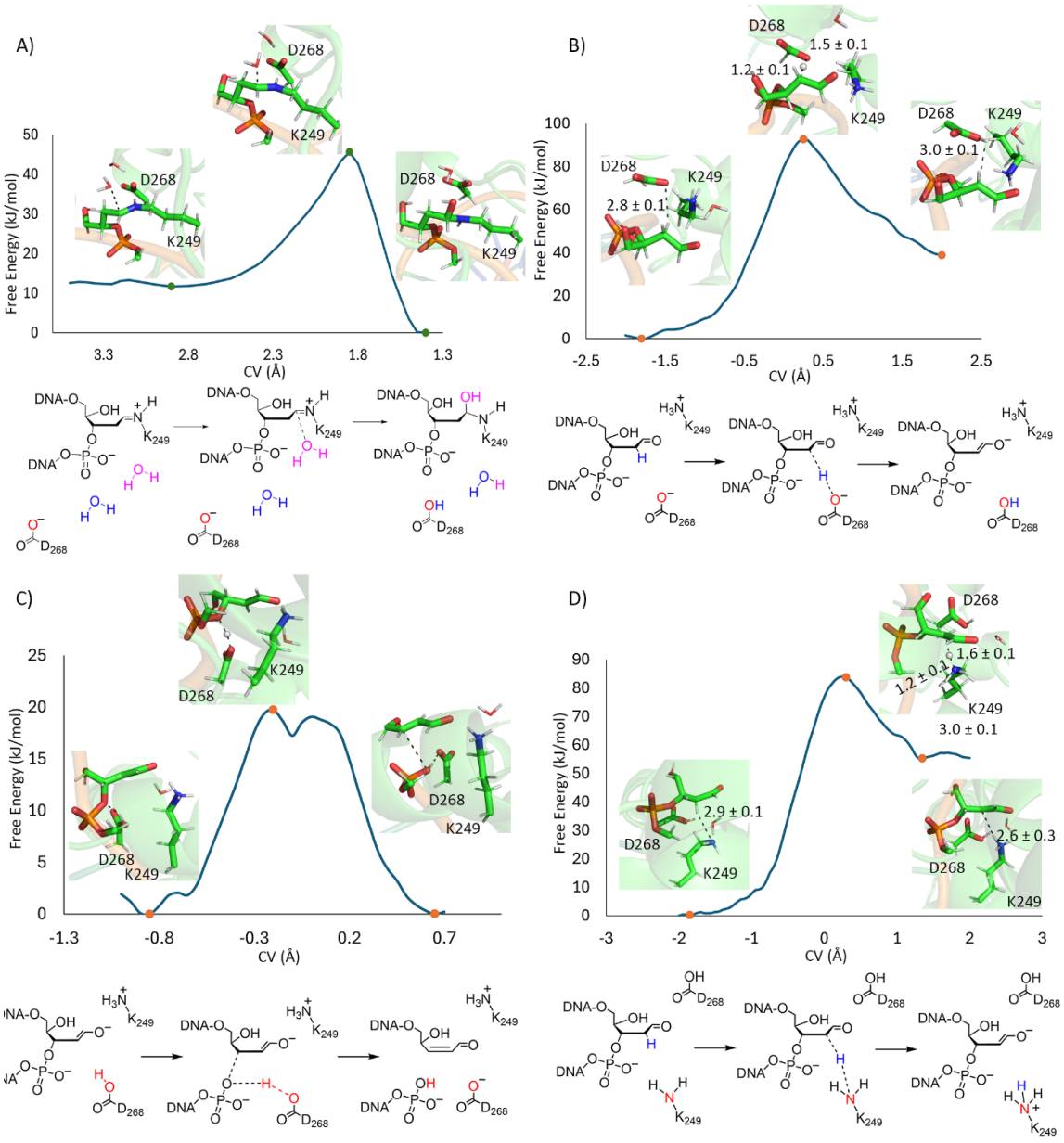


Figure 5.8. Potential energy surface and representative structures of stationary points for the A) D268 catalyzed hydrolysis, and post-hydrolysis B) D268-catalyzed elimination, C) phosphate protonation, and D) K249-catalyzed elimination for a cationic imine crosslink.

The barrier for the rate-limiting elimination step is 92.7 kJ/mol, which is consistent with the experimental barrier (92 kJ/mol)[70] and significantly lower than the barrier for elimination pre-hydrolysis (140.3 kJ/mol). This could be due to the more electronegative oxygen in the aldehyde being able to accommodate the charge buildup in the transition state better than nitrogen in the

imine combined with lysine being released and therefore available to hydrogen bond with the substrate to facilitate catalysis. The pathway with D268-assisted hydrolysis followed by D268-assisted elimination represents the first favorable catalytic mechanism characterized for hOGG1, which is consistent with the experimental barrier (92 kJ/mol),[70] mutagenic data suggesting D268 plays a role in catalysis[35] and proposals that 8oG and related activators are allosteric in nature.[33, 70]

5.3.8. K249 can act as an alternative base for the hOGG1 β-lyase elimination step

Despite the promise of the previously characterized β-lyase pathway consisting of initial imine hydrolysis followed by D268-catalyzed elimination, the proposed central role of D268 is not consistent with a study showing the D268Q hOGG1 mutant has similar catalytic activity to the wild-type enzyme.[35] This suggests another base may facilitate catalysis for the mutant. While no residue other than D268 is positioned for C2′ proton abstraction when hOGG1 is bound to the cationic crosslinked intermediate, K249 is free to act as a base after the kinetically accessible crosslink hydrolysis. While lysine acting as a base for elimination has not been previously proposed for a DNA glycosylase, lysine does fulfill this role in the mechanism of action of other enzymes.[106-110] For example, βC-S lyase uses K234 as the general base for the α,β-elimination of cysteine[106] and SpvC phosphothreonine lyase uses K136 to abstract Hα along the pathway for phosphate elimination.[108] Therefore, the potential of K249 to facilitate phosphate elimination in hOGG1 was investigated using QM/MM MD simulations (Figures 5.8D and D.16, Appendix D).

In the RC (CV ~ −1.85), K249 is 2.9 ± 0.1 Å away from H2′ and partially forms a bond with H2′ in the TS (CV ~ 0.30 Å; r(K249−H2′) = 1.2 ± 0.1 Å), which extends r(C2′−H2′) to 1.6 ± 0.1 Å. As in the mechanism in which D268 acts as the base, the aldehyde bond starts to weaken in the TS (~0.05 Å increased bond length) and is fully cleaved in the PC (CV ~ 1.35 Å), while H2′ is transferred to K249. The aldehyde bond is also weakened further (r(C1′−O1′) = 1.34 ± 0.04 Å). Since the product is the same as that for D268-assisted elimination, protonation of the O3′ leaving group will follow the same low−barrier pathway (19.8 kJ/mol, Figure 5.8C). Therefore, the overall energy

barrier associated with initial crosslink hydrolysis followed by K249-catalyzed elimination is 83.8 kJ/mol (Figure 5.8C), which is similar to the D268-catalyzed elimination barrier (92.7 kJ/mol) and experimental findings (92 kJ/mol).[70] The feasibility of this pathway is also consistent with the D268Q hOGG1 mutant having similar catalytic activity to the wild-type enzyme.[35]

5.4. Discussion

As part of the repair of 8oG damage in DNA, hOGG1 catalyzes the cleavage of both the glycosidic and 3′-phosphodiester bonds.[26] While the glycosidic bond cleavage pathway has been extensively studied,[27-31] the mechanism for hOGG1 β-lyase activity is unknown, with the little experimental data available leading to conflicting proposals.[26, 33] An 8oG anion has been theorized to initiate the hOGG1 β-lyase reaction by acting as a base for the elimination step based on a crystal structure of the borohydride-trapped hOGG1 crosslinked intermediate (Figure 5.2A).[26] However, this pathway is dependent on the 8oG deglycosylation product being tightly bound in the active site or anionic 8oG being available in the cell to bind to the hOGG1−DNA complex as a cofactor. In the present work, umbrella sampling calculations unveiled a low barrier for anionic 8oG departure from the enzyme active site regardless of the crosslink protonation state (5−15 kJ/mol, Figure D.3, Appendix D). Indeed, unbinding was observed to be aided by stacking interactions between the exiting 8oG and F319 and H270 (Figure D.5, Appendix D). Interestingly, a previous MD study of 8oG flipping into the hOGG1 active site identified F319 and H270 as playing roles in the initial binding of the substrate,[68] suggesting that 8oG is accepted and released from the active site through similar mechanisms. With the low barrier for base departure, the only remaining source of anionic 8oG for a product-assisted pathway would be existing free 8oG in the cell. However, cells contain an extremely low concentration of 8oG (~0.25 nM),[97, 98] and the low acidity of N9 of 8oG[99] would result in only a handful of anionic molecules present in an entire cell. Notably, 8oG concentrations during crystallization of an hOGG1 borohydride-trapped intermediate (233 μM assuming every hOGG1 molecule excised one base) were much higher, which likely led to 8oG being retained in the active site in the resolved structure (PDB ID: 1HU0).[26] Our proposal that base

retention is unlikely in a cellular context is further supported by crystal structures of other bifunctional glycosylases containing a covalently-trapped intermediate (e.g., Fpg,[111, 112] Endo VIII,[113] mNEIL3,[114] and Nth[115]), which do not contain an excised base in the active site. Without access to either self-generated or environmental anionic 8oG, a product-assisted mechanism for hOGG1 is unfeasible.

Since our calculations suggest that 8oG cannot directly participate in the catalytic mechanism, the reported activation of hOGG1 by 8oG[26] is unlikely to result from 8oG binding in the active site. Instead, hOGG1 activation is likely allosteric, as has been previously hypothesized based on random screening identifying activators that are structurally distinct from 8oG.[33] Although the allosteric binding site of hOGG1 activators has yet to be discovered, uncovering the nature of 8oG activation paves the way for future research focused on identifying the binding location of hOGG1 activators, which would allow for the rational design of activators with greater efficacy for applications such as therapeutics for treating conditions and diseases related to oxidative stress (heart attack, stroke, Chronic Obstructive Pulmonary Disease).[79, 80, 116-118]

With 8oG being unable to initiate the elimination step, an hOGG1 active site residue must catalyze the β-lyase step. Although not aligned for catalysis in the static crystal structure of the crosslinked intermediate (PDB ID: 1HU0),[26] D268 has been previously theorized to act as a base for elimination.[33] In this work, classical MD simulations reveal for the first time that D268 is positioned to act as a base for β-elimination regardless of the crosslink protonation state. This highlights the importance of accounting for the dynamic nature of enzyme active sites when predicting the roles of key residues. While both crosslink protonation states adopt conformations conducive for catalysis, D268 is correctly aligned over twice as frequently with respect to the cationic crosslink as the neutral crosslink (D268 within 4 Å of C2′ for 52% vs. 24% of the simulation time), mainly due to a persistent hydrogen bond between Nζ of the cationic crosslink and D268. Thus, classical MD simulations provide the first indication that the neutral crosslink may be a less favorable substrate than the cationic variant.

Subsequent QM/MM MD calculations mapping the hOGG1 β-lyase mechanism (Figure 5.5) with a neutral imine substrate uncovered a high energy barrier for both the elimination (218.8 kJ/mol) and crosslink hydrolysis (over 200 kJ/mol) steps. When combined with the inferior D268 alignment in the classical MD simulations, our data indicate a neutral imine crosslink is not a viable substrate for hOGG1 β-lyase activity. This is the first time the correct protonation state of the crosslink substrate has been revealed as vital for the catalytic activity of a bifunctional glycosylase. Nevertheless, it is not uncommon for enzymes, including glycosylases, to require specific substrate protonation states for catalysis to progress (e.g., MutY catalysis involves a cationic adenine intermediate[53, 86] and anionic *trans*-4-hydroxy-l-proline (Hyp) results in the most favorable pathway for Hyp dehydratase activity[119]).[120] Furthermore, our proposal that hOGG1 is only active towards cationic crosslinks mirrors bifunctional glycosylases that form proline−DNA crosslinks, such as Fpg[111, 112] and NEIL1,[60] which also rely on a cationic crosslink for the lyase step.[61, 121] This suggests the requirement of a cationic crosslink-containing substrate for lyase activity may be ubiquitous across the bifunctional glycosylase family, regardless of the crosslinking amino acid residue.

In the preferred catalytic pathway identified in this work (Figure 5.9), β-lyase activity is initiated by hydrolysis of the cationic imine crosslink since immediate C2′ proton abstraction has a prohibitively high energy barrier (~140 kJ/mol). During crosslink hydrolysis, an active site nucleophilic water molecule attacks C1′ of a cationic imine crosslink, which is facilitated by proton donation to D268 through a bridging water. Proton transfers between Nζ, D268, and O1′ complete the hydrolysis step, resulting in an aldehyde product. Subsequently, anionic D268 or neutral K249 abstracts a C2′ proton and a low-barrier proton transfer from D268 to O3′ completes the reaction, with a resulting rate-limiting QM/MM MD barrier (83.8−92.7 kJ/mol) being consistent with the experimental barrier (92 kJ/mol) for hOGG1 β-lyase activity.[70]

This mechanism elucidated in this work reveals, for the first time, that a crosslinking lysine has the potential to act as a base for the elimination step catalyzed by a bifunctional glycosylase, and rationalizes experimental findings that D268 mutation to asparagine or glutamine results in a 0- and 68-fold reduction in hOGG1 catalytic activity respectively.[35] Additionally, the lack of a product

cofactor in catalysis agrees with kinetic data showing compounds structurally distinct from 8oG can behave as hOGG1 activators.[33, 34] It is anticipated that other bifunctional glycosylases that generate an imine crosslink have the potential to adopt similar mechanisms. For example, T4 PDG forms a similar imine crosslink as part of glycosidic bond cleavage and has been hypothesized to use an active site glutamate to facilitate elimination as mutation of E23 to a non-basic residue inhibits catalytic activity,[102, 103] which mirrors findings for hOGG1. However further experimental and computational work is needed to determine whether other bifunctional glycosylases adopt a similar mechanism to hOGG1 or whether the hOGG1 pathway characterized in this work is unique.

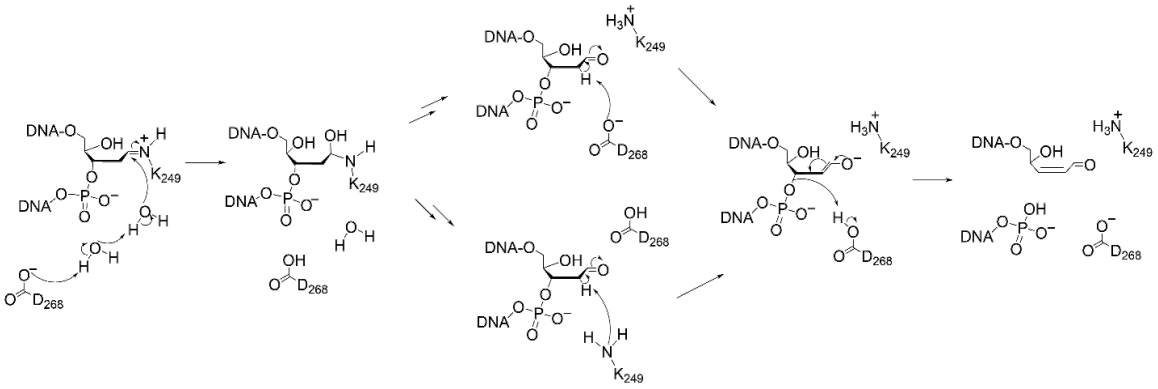


Figure 5.9. Preferred hOGG1 β-lyase pathway elucidated in this study.

Since cancer cells are often under increased levels of oxidative stress,[122, 123] inhibition of enzymes that combat oxidative stress impact cancerous cells more strongly than healthy cells. As the main enzyme responsible for repairing the primary oxidative product in DNA, the development of inhibitors for hOGG1 is an active area of research.[75-78] Indeed, inhibitors for flap endonuclease 1, another enzyme involved in the BER of oxidative DNA damage,[124] have been shown to enhance the ability of other therapeutics to fight cancer cells.[125-127] In addition to treating cancer, hOGG1 inhibition can help alleviate certain pro-inflammatory conditions, with an hOGG1 inhibitor having been proven to decrease the expression of proinflammatory genes in mice.[74, 75] Thus, the improved understanding of the hOGG1 mechanism of action from this study will enable development of rationally-designed transition state analogue inhibitors with high efficacies for treating disorders related to oxidative stress.[71-73]

5.5. Conclusion

While the deglocosylation step of the hOGG1 mechanism of action is well characterized,[27-31] conflicting proposals were present in the literature for the β-lyase step based on limited experimental data.[26, 32, 33] The present work unifies all previous kinetic[33, 34, 70] and mutagenic[35] data on the mechanism of action of hOGG1, characterizing the β-lyase pathway using diverse computational techniques. Umbrella sampling MD simulations revealed that release of the cleaved 8oG base occurs prior to phosphodiester bond cleavage, contrasting previous proposals for product-assisted elimination.[26] Contrary to observations in an experimental crystal structure of the hOGG1−DNA borohydride-trapped intermediate (PDB ID: 1HU0),[26] MD simulations on the hOGG1-crosslinked deglycosylation product uncover that D268 is frequently positioned for C2′ proton abstraction. Comparisons of four QM/MM MD characterized catalytic pathways for β-lyase activity determined that catalysis is only possible when the substrate is a cationic crosslink and hydrolysis of the imine crosslink occurs before elimination, which provides two potential bases for subsequent elimination (D268 and K249) and is consistent with the D268Q hOGG1 mutant retaining catalytic activity.[35] The mechanism characterized in this work supports the proposal that hOGG1 activators are allosteric in nature,[33] laying the groundwork for the identification of the allosteric binding site and the development of hOGG1 activators for treating inflammatory conditions[74, 75] and oxidative stress.[79, 80] The improved understanding of the hOGG1 mechanism of action from this study promises to aid in the development of rationally-designed hOGG1 transition state analogue inhibitors[71-73] to improve the efficacy of cancer therapeutics.[122, 123]

5.6. References

(1) Yu, Y.; Cui, Y.; Niedernhofer, L. J.; Wang, Y. Occurrence, biological consequences, and human health relevance of oxidative stress-induced DNA damage. *Chem. Res. Toxicol.* **2016**, *29* (12), 2008-2039. DOI: 10.1021/acs.chemrestox.6b00265

(2) Turrens, J. F. Mitochondrial formation of reactive oxygen species. *J. Physiol.* **2003**, *552* (2), 335-344. DOI: https://doi.org/10.1111/j.1469-7793.2003.00335.x

(3) Mittal, M.; Siddiqui, M. R.; Tran, K.; Reddy, S. P.; Malik, A. B. Reactive oxygen species in inflammation and tissue injury. *Antioxid. Redox Singal.* **2014**, *20* (7), 1126-1167. DOI: 10.1089/ars.2012.5149

(4) Tominaga, H.; Kodama, S.; Matsuda, N.; Suzuki, K.; Watanabe, M. Involvement of reactive oxygen species (ROS) in the induction of genetic instability by radiation. *J. Radiat. Res.* **2004**, *45* (2), 181-188. DOI: 10.1269/jrr.45.181 (acccessed 4/22/2025).

(5) Prasad, S.; Gupta, S. C.; Tyagi, A. K. Reactive oxygen species (ROS) and cancer: Role of antioxidative nutraceuticals. *Cancer Lett.* **2017**, *387*, 95-105. DOI: https://doi.org/10.1016/j.canlet.2016.03.042

(6) Wang, L.; Wise, J. T. F.; Zhang, Z.; Shi, X. Progress and prospects of reactive oxygen species in metal carcinogenesis. *Curr. Pharmacol. Rep.* **2016**, *2* (4), 178-186. DOI: 10.1007/s40495-016-0061-2

(7) Fatehi-Hassanabad, Z.; Chan, C. B.; Furman, B. L. Reactive oxygen species and endothelial function in diabetes. *Eur. J. Pharmacol.* **2010**, *636* (1), 8-17. DOI: https://doi.org/10.1016/j.ejphar.2010.03.048

(8) Rendra, E.; Riabov, V.; Mossel, D. M.; Sevastyanova, T.; Harmsen, M. C.; Kzhyshkowska, J. Reactive oxygen species (ROS) in macrophage activation and function in diabetes. *Immunobiology* **2019**, *224* (2), 242-253. DOI: https://doi.org/10.1016/j.imbio.2018.11.010

(9) Chen, Q.; Wang, Q.; Zhu, J.; Xiao, Q.; Zhang, L. Reactive oxygen species: Key regulators in vascular health and diseases. *Br. J. Pharmacol.* **2018**, *175* (8), 1279-1292. DOI: https://doi.org/10.1111/bph.13828

(10) Murphy, E.; Liu, J. C. Mitochondrial calcium and reactive oxygen species in cardiovascular disease. *Cardiovasc. Res.* **2022**, *119* (5), 1105-1116. DOI: 10.1093/cvr/cvac134 (acccessed 8/6/2025).

(11) Patten, D. A.; Germain, M.; Kelly, M. A.; Slack, R. S. Reactive oxygen species: Stuck in the middle of neurodegeneration. *J. Alzheimer's Dis.* **2010**, *20*, 357-367. DOI: 10.3233/JAD-2010-100498

(12) Bhatt, S.; Puli, L.; Patil, C. R. Role of reactive oxygen species in the progression of alzheimer's disease. *Drug Discov. Today* **2021**, *26* (3), 794-803. DOI: https://doi.org/10.1016/j.drudis.2020.12.004

(13) Diehn, M.; Cho, R. W.; Lobo, N. A.; Kalisky, T.; Dorie, M. J.; Kulp, A. N.; Qian, D.; Lam, J. S.; Ailles, L. E.; Wong, M.; et al. Association of reactive oxygen species levels and radioresistance in cancer stem cells. *Nature* **2009**, *458* (7239), 780-783. DOI: 10.1038/nature07733

(14) Shinmura, K.; Yokota, J. The OGG1 gene encodes a repair enzyme for oxidatively damaged DNA and is involved in human carcinogenesis. *Antioxid. Redox Singal.* **2001**, *3* (4), 597-609. DOI: 10.1089/15230860152542952

(15) Wu, K.; El Zowalaty, A. E.; Sayin, V. I.; Papagiannakopoulos, T. The pleiotropic functions of reactive oxygen species in cancer. *Nat. Cancer* **2024**, *5* (3), 384-399. DOI: 10.1038/s43018-024-00738-9

(16) Nakamura, H.; Takada, K. Reactive oxygen species in cancer: Current findings and future directions. *Cancer Sci.* **2021**, *112* (10), 3945-3952. DOI: https://doi.org/10.1111/cas.15068

(17) Shi, Y.; Xu, W.; Zhang, X. Association of the hOGG1 Ser326Cys polymorphism with gynecologic cancer susceptibility: A meta-analysis. *Biosci. Rep.* **2020**, *40* (12). DOI: 10.1042/bsr20203245 (acccessed 4/23/2025).

(18) Xu, Z.; Yu, L.; Zhang, X. Association between the hOGG1 Ser326Cys polymorphism and lung cancer susceptibility: A meta-analysis based on 22,475 subjects. *Diagn. Pathol.* **2013**, *8* (1), 144. DOI: 10.1186/1746-1596-8-144

(19) Leu, M.; Riebeling, T.; Dröge, L. H.; Hubert, L.; Guhlich, M.; Wolff, H. A.; Brockmöller, J.; Gaedcke, J.; Rieken, S.; Schirmer, M. A. 8-oxoguanine DNA glycosylase (OGG1) cys326 variant: Increased risk for worse outcome of patients with locally advanced rectal cancer after multimodal therapy. *Cancers* **2021**, *13* (11), 2805.

(20) Cadet, J.; Davies, K. J. A.; Medeiros, M. H. G.; Di Mascio, P.; Wagner, J. R. Formation and repair of oxidatively generated damage in cellular DNA. *Free Radical Biol. Med.* **2017**, *107*, 13-34. DOI: https://doi.org/10.1016/j.freeradbiomed.2016.12.049

(21) Collins, A.; Gedik, C.; Vaughan, N.; Wood, S.; White, A.; Dubois, J.; Rees, J.-F.; Loft, S.; Moller, P.; Poulsen, H.; et al. Measurement of DNA oxidation in human cells by chromatographic and enzymic methods. *Free Radical Biol. Med.* **2003**, *34* (8), 1089-1099. DOI: 10.1016/S0891-5849(03)00041-8

(22) David, S. S.; O'Shea, V. L.; Kundu, S. Base-excision repair of oxidative DNA damage. *Nature* **2007**, *447* (7147), 941-950. DOI: 10.1038/nature05978

(23) Choi, J.-S.; Dasari, A.; Hu, P.; Benkovic, S. J.; Berdis, A. J. The use of modified and non-natural nucleotides provide unique insights into pro-mutagenic replication catalyzed by polymerase eta. *Nucleic Acids Res.* **2016**, *44* (3), 1022-1035. DOI: 10.1093/nar/gkv1509

(24) Cheng, X.; Kelso, C.; Hornak, V.; de los Santos, C.; Grollman, A. P.; Simmerling, C. Dynamic behavior of DNA base pairs containing 8-oxoguanine. *J. Am. Chem. Soc.* **2005**, *127* (40), 13906-13918. DOI: 10.1021/ja052542s

(25) Kuznetsov, N. A.; Koval, V. V.; Fedorova, O. S. Mechanism of recognition and repair of damaged DNA by human 8-oxoguanine DNA glycosylase hOGG1. *Biochemistry* **2011**, *76* (1), 118-130. DOI: 10.1134/S0006297911010123

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

(27) Nash, H. M.; Lu, R.; Lane, W. S.; Verdinel, G. L. The critical active-site amine of the human 8-oxoguanine DNA glycosylase, hOGG1: Direct identification, ablation and chemical reconstitution. *Chem. Biol.* **1997**, *4* (9), 693-702. DOI: https://doi.org/10.1016/S1074-5521(97)90225-8

(28) Tyugashev, T. E.; Vorobjev, Y. N.; Kuznetsova, A. A.; Lukina, M. V.; Kuznetsov, N. A.; Fedorova, O. S. Roles of active-site amino acid residues in specific recognition of DNA lesions by human 8-oxoguanine-DNA glycosylase (OGG1). *J. Phys. Chem. B* **2019**, *123* (23), 4878-4887. DOI: 10.1021/acs.jpcb.9b02949

(29) Šebera, J.; Trantírek, L.; Tanaka, Y.; Sychrovský, V. Pyramidalization of the glycosidic nitrogen provides the way for efficient cleavage of the N-glycosidic bond of 8-oxoG with the hOGG1 DNA repair protein. *J. Phys. Chem. B* **2012**, *116* (41), 12535-12544. DOI: 10.1021/jp309098d

(30) Šebera, J.; Trantírek, L.; Tanaka, Y.; Nencka, R.; Fukal, J.; Sychrovský, V. The activation of N-glycosidic bond cleavage performed by base-excision repair enzyme hOGG1; theoretical study of the role of Lys 249 residue in activation of G, OxoG and FapyG. *RSC Adv.* **2014**, *4* (83), 44043-44051. DOI: 10.1039/C4RA08278H

(31) 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* (9), 5231-5242. DOI: 10.1093/nar/gkx157 (acccessed 1/8/2024).

(32) Dalhus, B.; Forsbring, M.; Helle, I. H.; Vik, E. S.; Forstrøm, R. J.; Backe, P. H.; Alseth, I.; Bjørås, M. Separation-of-function mutants unravel the dual-reaction mode of human 8-oxoguanine DNA glycosylase. *Structure* **2011**, *19* (1), 117-127. DOI: https://doi.org/10.1016/j.str.2010.09.023

(33) Tian, G.; Katchur, S. R.; Jiang, Y.; Briand, J.; Schaber, M.; Kreatsoulas, C.; Schwartz, B.; Thrall, S.; Davis, A. M.; Duvall, S.; et al. Small molecule-mediated allosteric activation of the base excision repair enzyme 8-oxoguanine DNA glycosylase and its impact on mitochondrial function. *Sci. Rep.* **2022**, *12* (1), 14685. DOI: 10.1038/s41598-022-18878-2

(34) Kuznetsov, N. A.; Koval, V. V.; Zharkov, D. O.; Nevinsky, G. A.; Douglas, K. T.; Fedorova, O. S. Kinetics of substrate recognition and cleavage by human 8-oxoguanine-DNA glycosylase. *Nucleic Acids Res.* **2005**, *33* (12), 3919-3931. DOI: 10.1093/nar/gki694 (acccessed 4/16/2025).

(35) Norman, D. P. G.; Chung, S. J.; Verdine, G. L. Structural and biochemical exploration of a critical amino acid in human 8-oxoguanine glycosylase. *Biochemistry* **2003**, *42* (6), 1564-1572. DOI: 10.1021/bi026823d

(36) Senn, H. M.; Thiel, W. QM/MM studies of enzymes. *Curr. Opin. Chem. Biol.* **2007**, *11* (2), 182-187. DOI: https://doi.org/10.1016/j.cbpa.2007.01.684

(37) Vennelakanti, V.; Nazemi, A.; Mehmood, R.; Steeves, A. H.; Kulik, H. J. Harder, better, faster, stronger: Large-scale QM and QM/MM for predictive modeling in enzymes and proteins. *Curr. Opin. Struct. Biol.* **2022**, *72*, 9-17. DOI: https://doi.org/10.1016/j.sbi.2021.07.004

(38) Sousa, S. F.; Ribeiro, A. J. M.; Neves, R. P. P.; Brás, N. F.; Cerqueira, N. M. F. S. A.; Fernandes, P. A.; Ramos, M. J. Application of quantum mechanics/molecular mechanics methods in the study of enzymatic reaction mechanisms. *WIREs Comput. Mol. Sci.* **2017**, *7* (2), e1281. DOI: https://doi.org/10.1002/wcms.1281

(39) Blomberg, M. R.; Borowski, T.; Himo, F.; Liao, R.-Z.; Siegbahn, P. E. Quantum chemical studies of mechanisms for metalloenzymes. *Chem. Rev.* **2014**, *114* (7), 3601-3658. DOI: https://doi.org/10.1021/cr400388t

(40) Agbaglo, D. A.; Summers, T. J.; Cheng, Q.; DeYonker, N. J. The influence of model building schemes and molecular dynamics sampling on QM-cluster models: The chorismate mutase case study. *Phys. Chem. Chem. Phys.* **2024**, *26* (16), 12467-12482. DOI: 10.1039/D3CP06100K

(41) Sheng, X.; Himo, F. The quantum chemical cluster approach in biocatalysis. *Acc. Chem. Res.* **2023**, *56* (8), 938-947. DOI: 10.1021/acs.accounts.2c00795

(42) Elsässer, B.; Goettig, P. Mechanisms of proteolytic enzymes and their inhibition in QM/MM studies. *Int. J. Mol. Sci.* **2021**, *22* (6), 3232.

(43) Kaur, R.; Nikkel, D. J.; Wetmore, S. D. Computational studies of DNA repair: Insights into the function of monofunctional DNA glycosylases in the base excision repair pathway. *WIREs Comput. Mol. Sci.* **2020**, *10* (5), e1471. DOI: 10.1002/wcms.1471 (acccessed 2020/09/16).

(44) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Computational enzymatic catalysis – clarifying enzymatic mechanisms with the help of computers. *Phys. Chem. Chem. Phys.* **2012**, *14* (36), 12431-12441. DOI: 10.1039/C2CP41180F

(45) Gherib, R.; Dokainish, H. M.; Gauld, J. W. Multi-scale computational enzymology: Enhancing our understanding of enzymatic catalysis. *Int. J. Mol. Sci.* **2014**, *15* (1), 401-422. DOI: 10.3390/ijms15010401

(46) 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. DOI: 10.1021/acs.jpcb.5b04051

(47) Nikkel, D. J.; Wetmore, S. D. Unlocking the chemistry facilitated by enzymes that process nucleic acids using quantum mechanical and combined quantum mechanics–molecular mechanics techniques. *Pure Appl. Chem.* **2025**, *97* (9), 1065-1089. DOI: doi:10.1515/pac-2025-0507 (acccessed 2025-10-14).

(48) Przybylski, J. L.; Wetmore, S. D. A QM/QM investigation of the hUNG2 reaction surface: The untold tale of a catalytic residue. *Biochemistry* **2011**, *50* (19), 4218-4227. DOI: 10.1021/bi2003394

(49) Das, R.; Vázquez-Montelongo, E. A.; Cisneros, G. A.; Wu, J. I. Ground state destabilization in uracil DNA glycosylase: Let's not forget "tautomeric strain" in substrates. *J. Am. Chem. Soc.* **2019**, *141* (35), 13739-13743. DOI: 10.1021/jacs.9b06447

(50) Luo, N.; Mehler, E.; Osman, R. Specificity and catalysis of uracil DNA glycosylase. A molecular dynamics study of reactant and product complexes with DNA. *Biochemistry* **1999**, *38* (29), 9209-9220. DOI: 10.1021/bi990262h

(51) Gokey, T.; Hang, B.; Guliaev, A. B. Cadmium(II) inhibition of human uracil-DNA glycosylase by catalytic water supplantation. *Sci. Rep.* **2016**, *6* (1), 39137. DOI: 10.1038/srep39137

(52) Naydenova, E.; Roßbach, S.; Ochsenfeld, C. QM/MM study of the uracil DNA glycosylase reaction mechanism: A competition between Asp145 and His148. *J. Chem. Theory Comput.* **2019**, *15* (8), 4344-4350. DOI: 10.1021/acs.jctc.8b01305

(53) Nikkel, D. J.; Wetmore, S. D. Distinctive formation of a DNA–protein cross-link during the repair of DNA oxidative damage: Insights into human disease from MD simulations and QM/MM calculations. *J. Am. Chem. Soc.* **2023**, *145* (24), 13114-13125. DOI: 10.1021/jacs.3c01773

(54) Kanaan, N.; Imhof, P. Interactions of the DNA repair enzyme human thymine DNA glycosylase with cognate and noncognate DNA. *Biochemistry* **2018**, *57* (39), 5654-5665. DOI: 10.1021/acs.biochem.8b00409

(55) Naydenova, E.; Dietschreit, J. C. B.; Ochsenfeld, C. Reaction mechanism for the N-glycosidic bond cleavage of 5-formylcytosine by thymine DNA glycosylase. *J. Phys. Chem. B* **2019**, *123* (19), 4173-4179. DOI: 10.1021/acs.jpcb.8b11706

(56) Amara, P.; Serre, L.; Castaing, B.; Thomas, A. Insights into the DNA repair process by the formamidopyrimidine-DNA glycosylase investigated by molecular dynamics. *Protein Sci.* **2004**, *13* (8), 2009-2021. DOI: https://doi.org/10.1110/ps.04772404

(57) Guliaev, A. B.; Hang, B.; Singer, B. Structural insights by molecular dynamics simulations into differential repair efficiency for ethano-A versus etheno-A adducts by the human alkylpurine-DNA N -glycosylase. *Nucleic Acids Res.* **2002**, *30* (17), 3778-3787. DOI: 10.1093/nar/gkf494 (acccessed 4/23/2025).

(58) Liu, M.; Zhang, J.; Zhu, C.; Zhang, X.; Xiao, W.; Yan, Y.; Liu, L.; Zeng, H.; Gao, Y. Q.; Yi, C. DNA repair glycosylase hNEIL1 triages damaged bases via competing interaction modes. *Nat. Commun.* **2021**, *12* (1), 4108. DOI: 10.1038/s41467-021-24431-y

(59) 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 Engl* **2014**, *53* (38), 10044-10048.

(60) 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* (28), 7792-7797. DOI: doi:10.1073/pnas.1604591113

(61) Zheng, J.-H.; Tan, H.-W.; Chen, G.-J. Theoretical study on the mechanism of the DNA repair protein Fpg. *Int. J. Quantum Chem* **2011**, *111* (10), 2454-2463. DOI: https://doi.org/10.1002/qua.22447

(62) Kreppel, A.; Blank, I. D.; Ochsenfeld, C. Base-independent DNA base-excision repair of 8-oxoguanine. *J. Am. Chem. Soc.* **2018**, *140* (13), 4522-4526. DOI: 10.1021/jacs.7b11254

(63) Dračínský, M.; Šála, M.; Klepetářová, B.; Šebera, J.; Fukal, J.; Holečková, V.; Tanaka, Y.; Nencka, R.; Sychrovský, V. Benchmark theoretical and experimental study on 15N NMR shifts of oxidatively damaged guanine. *J. Phys. Chem. B* **2016**, *120* (5), 915-925. DOI: 10.1021/acs.jpcb.5b11428

(64) Kim, H.; Pak, Y. Free energy landscape of lesion recognition by human 8-oxoguanine DNA N-glycosylase 1: Mechanistic insights into detection and processing of 8-oxoguanine in DNA. *J. Phys. Chem. B* **2024**, *128* (49), 12133-12142. DOI: 10.1021/acs.jpcb.4c07073

(65) Sowlati-Hashjin, S.; Wetmore, S. D. Structural insight into the discrimination between 8-oxoguanine glycosidic conformers by DNA repair enzymes: A molecular dynamics study of human oxoguanine glycosylase 1 and formamidopyrimidine-DNA glycosylase. *Biochemistry* **2018**, *57* (7), 1144-1154. DOI: 10.1021/acs.biochem.7b01292

(66) Lukina, M. V.; Koval, V. V.; Lomzov, A. A.; Zharkov, D. O.; Fedorova, O. S. Global DNA dynamics of 8-oxoguanine repair by human OGG1 revealed by stopped-flow kinetics and molecular dynamics simulation. *Mol. Biosyst.* **2017**, *13* (10), 1954-1966. DOI: 10.1039/C7MB00343A

(67) Naômé, A.; Schyman, P.; Laaksonen, A.; Vercauteren, D. P. Molecular dynamics simulation of 8-oxoguanine containing DNA fragments reveals altered hydration and ion binding patterns. *J. Phys. Chem. B* **2010**, *114* (14), 4789-4801. DOI: 10.1021/jp1000539

(68) Shigdel, U. K.; Ovchinnikov, V.; Lee, S.-J.; Shih, J. A.; Karplus, M.; Nam, K.; Verdine, G. L. The trajectory of intrahelical lesion recognition and extrusion by the human 8-oxoguanine DNA glycosylase. *Nat. Commun.* **2020**, *11* (1), 4437. DOI: 10.1038/s41467-020-18290-2

(69) 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. DOI: 10.1021/jp306344g

(70) Yin, Y.; Sasaki, S.; Taniguchi, Y. Recognition and excision properties of 8-halogenated-7-deaza-2'-deoxyguanosine as 8-oxo-2'-deoxyguanosine analogues and Fpg and hOGG1 inhibitors. *ChemBioChem* **2015**, *16* (8), 1190-1198. DOI: https://doi.org/10.1002/cbic.201402690

(71) Schramm, V. L. Transition states, analogues, and drug development. *ACS Chem. Biol.* **2013**, *8* (1), 71-81. DOI: 10.1021/cb300631k

(72) Pushing cancer over the edge. *Chemical & Engineering News Archive* **2013**, *91* (24), 13-18. DOI: 10.1021/cen-09124-cover

(73) Zhang, Z.; Wu, Z.; Shi, X.; Guo, D.; Cheng, Y.; Gao, J.; Liu, L.; Liu, W.; Liang, L.; Peng, L.; et al. Research progress in human AP endonuclease 1: Structure, catalytic mechanism, and inhibitors. *Curr Protein Pept Sci* **2022**, *23* (2), 77-88. DOI: 10.2174/1389203723666220406132737

(74) Mabley, J. G.; Pacher, P.; Deb, A.; Wallace, R.; Elder, R. H.; Szabo, C. Potential role for 8-oxoguanine DNA glycosylase in regulating inflammation. *FASEB J.* **2005**, *19* (2), 290-292. DOI: 10.1096/fj.04-2278fje

(75) Visnes, T.; Cázares-Körner, A.; Hao, W.; Wallner, O.; Masuyer, G.; Loseva, O.; Mortusewicz, O.; Wiita, E.; Sarno, A.; Manoilov, A.; et al. Small-molecule inhibitor of OGG1 suppresses proinflammatory gene expression and inflammation. *Science* **2018**, *362* (6416), 834-839. DOI: doi:10.1126/science.aar8048

(76) Wallner, O.; Cázares-Körner, A.; Scaletti, E. R.; Masuyer, G.; Bekkhus, T.; Visnes, T.; Mamonov, K.; Ortis, F.; Lundbäck, T.; Volkova, M.; et al. Optimization of N-piperidinyl-benzimidazolone derivatives as potent and selective inhibitors of 8-oxo-guanine DNA glycosylase 1. *ChemMedChem* **2023**, *18* (1), e202200310. DOI: https://doi.org/10.1002/cmdc.202200310

(77) Donley, N.; Jaruga, P.; Coskun, E.; Dizdaroglu, M.; McCullough, A. K.; Lloyd, R. S. Small molecule inhibitors of 8-oxoguanine DNA glycosylase-1 (OGG1). *ACS Chem. Biol.* **2015**, *10* (10), 2334-2343. DOI: 10.1021/acschembio.5b00452

(78) Jaruga, P.; Kant, M.; Luzadder, M. M.; Lloyd, R. S.; Boldogh, I.; Dizdaroglu, M. Inhibition by 4-(4-bromo-2-oxo-3h-benzimidazol-1-yl)-N-(4-iodophenyl)piperidine-1-carboxamide (TH5487) of the activity of human 8-oxoguanine DNA glycosylase-1 (OGG1) for the excision of 2,6-diamino-4-hydroxy-5-formamidopyrimidine, 4,6-diamino-5-formamidopyrimidine, and 8-oxoguanine from oxidatively damaged DNA. *Biochemistry* **2025**, *64* (8), 1788-1796. DOI: 10.1021/acs.biochem.4c00419

(79) Baptiste, B. A.; Katchur, S. R.; Fivenson, E. M.; Croteau, D. L.; Rumsey, W. L.; Bohr, V. A. Enhanced mitochondrial DNA repair of the common disease-associated variant, Ser326Cys,

of hOGG1 through small molecule intervention. *Free Radical Biol. Med.* **2018**, *124*, 149-162. DOI: https://doi.org/10.1016/j.freeradbiomed.2018.05.094

(80) Cardozo-Pelaez, F.; Sanchez-Contreras, M.; Nevin, A. B. C. Ogg1 null mice exhibit age-associated loss of the nigrostriatal pathway and increased sensitivity to MPTP. *Neurochem. Int.* **2012**, *61* (5), 721-730. DOI: https://doi.org/10.1016/j.neuint.2012.06.013

(81) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A P.; Rempfer, C.; Bordoli, L.; et al. Swiss-model: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46* (W1), W296-W303. DOI: 10.1093/nar/gky427 (acccessed 8/9/2023).

(82) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; III, T. E. C.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; et al. AMBER 2020. **2020**, University of California, San Fransisco.

(83) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696-3713. DOI: 10.1021/acs.jctc.5b00255

(84) Galindo-Murillo, R.; Robertson, J. C.; Zgarbova, M.; Sponer, J.; Otyepka, M.; Jurecka, P.; Cheatham, T. E. Assessing the current state of amber force field modifications for DNA. *J. Chem. Theory Comput.* **2016**, *12* (8), 4114-4127. DOI: 10.1021/acs.jctc.6b00186

(85) He, X.; Man, V. H.; Yang, W.; Lee, T.-S.; Wang, J. A fast and high-quality charge model for the next generation general AMBER force field. *J. Chem. Phys.* **2020**, *153* (11). DOI: 10.1063/5.0019056 (acccessed 9/26/2025).

(86) 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* (48), 8753-8765. DOI: 10.1021/bi401310w

(87) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; III, T. E. C.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; et al. AMBER 2018. **2018**, University of California, San Fransisco.

(88) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13* (8), 1011-1021. DOI: 10.1002/jcc.540130812

(89) Grossfield, A. WHAM: The weighted histogram analysis method, version 2.1.0, http://membrane.Urmc.Rochester.Edu/wordpress/?Page_id=126.

(90) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.* **1994**, *98* (45), 11623-11627. DOI: 10.1021/j100096a001

(91) Sahrawat, A. S.; Polidori, N.; Kroutil, W.; Gruber, K. Deciphering the unconventional reduction of C═N bonds by old yellow enzymes using QM/MM. *ACS Catal*. **2024**, *14* (3), 1257-1266. DOI: 10.1021/acscatal.3c04362

(92) González, R. D.; Carvalho, A. T. P. Papain-mediated conjugation of peptide nucleic acids to delivery peptides: A density functional theory/molecular mechanics metadynamics study in aqueous and organic solvent. *J. Phys. Chem. B* **2024**, *128* (31), 7500-7512. DOI: 10.1021/acs.jpcb.4c02294

(93) Sousa, J. P. M.; Ramos, M. J.; Fernandes, P. A. QM/MM study of the reaction mechanism of l-tyrosine hydroxylation catalyzed by the enzyme CYP76AD1. *J. Phys. Chem. B* **2024**, *128* (39), 9447-9454. DOI: 10.1021/acs.jpcb.4c05209

(94) Wang, W.-J.; Wang, T.; Zhao, Y.; Li, B.-N.; Chen, D.-Z. Theoretical insights into N-glycoside bond cleavage of 5-carboxycytosine by thymine DNA glycosylase: A QM/MM study. *J. Phys. Chem. B* **2024**, *128* (19), 4621-4630. DOI: 10.1021/acs.jpcb.4c00221

(95) Lee, T. S.; Allen, B. K.; Giese, T. J.; Guo, Z. Y.; Li, P. F.; Lin, C.; McGee, T. D.; Pearlman, D. A.; Radak, B. K.; Tao, Y. J.; et al. Alchemical binding free energy calculations in amber20: Advances and best practices for drug discovery. *J. Chem Inf. Model*. **2020**, *60* (11), 5595-5623. DOI: 10.1021/acs.jcim.0c00613

(96) *Gaussian 16 rev. C.01*; Wallingford, CT, 2016.

(97) Traut, T. W. Physiological concentrations of purines and pyrimidines. *Mol. Cell. Biochem*. **1994**, *140* (1), 1-22. DOI: 10.1007/BF00928361

(98) D'Augustin, O.; Huet, S.; Campalans, A.; Radicella, J. P. Lost in the crowd: How does human 8-oxoguanine DNA glycosylase 1 (OGG1) find 8-oxoguanine in the genome? *Int. J. Mol. Sci*. **2020**, *21* (21), 8360.

(99) Jang, Y. H.; Goddard, W. A.; Noyes, K. T.; Sowers, L. C.; Hwang, S.; Chung, D. S. First principles calculations of the tautomers and pka values of 8-oxoguanine: Implications for mutagenicity and repair. *Chem. Res. Toxicol*. **2002**, *15* (8), 1023-1035. DOI: 10.1021/tx010146r

(100) Hou, X.; Feng, J.; Franklin, J. L.; Russo, R.; Guo, Z.; Zhou, J.; Gao, J.-M.; Liu, H.-w.; Wang, B. Mechanistic insights from the crystal structure and computational analysis of the radical SAM deaminase DesII. *Adv. Sci*. **2024**, *11* (33), 2403494. DOI: https://doi.org/10.1002/advs.202403494

(101) Zhao, W.; Xu, W.; Tang, J.; Kaushik, S.; Chang, C.-E. A.; Zhao, L. Key amino acid residues of mitochondrial transcription factor A synergize with abasic (AP) site dynamics to facilitate AP-lyase reactions. *ACS Chem. Biol*. **2023**, *18* (5), 1168-1179. DOI: 10.1021/acschembio.3c00047

(102) Golan, G.; Zharkov, D. O.; Grollman, A. P.; Dodson, M. L.; McCullough, A. K.; Lloyd, R. S.; Shoham, G. Structure of T4 pyrimidine dimer glycosylase in a reduced imine covalent

complex with abasic site-containing DNA. *J. Mol. Biol.* **2006**, *362* (2), 241-258. DOI: https://doi.org/10.1016/j.jmb.2006.06.059

(103) Hori, N.; Doi, T.; Karaki, Y.; Kituchi, M.; Ikehara, M.; Ohtsuka, E. Participation of glutamic acid 23 of T4 endonuclease V in the β-elimination reaction of an abasic site in a synthetic duplex DNA. *Nucleic Acids Res.* **1992**, *20* (18), 4761-4764. DOI: 10.1093/nar/20.18.4761 (acccessed 9/15/2025).

(104) Shi, K.; Vasconez, N.; Corcelli, S. A.; Smith, B. D. Inhibited hydrolysis of 4-(N,N-dimethylamino)aryl imines in a weak acid by supramolecular encapsulation inside cucurbit[7]uril. *J. Org. Chem.* **2025**, *90* (24), 8309-8318. DOI: 10.1021/acs.joc.5c00774

(105) Cordes, E. H.; Jencks, W. P. The mechanism of hydrolysis of schiff bases derived from aliphatic amines. *J. Am. Chem. Soc.* **1963**, *85* (18), 2843-2848. DOI: 10.1021/ja00901a037

(106) Kezuka, Y.; Yoshida, Y.; Nonaka, T. Structural insights into catalysis by βC-S lyase from streptococcus anginosus. *Proteins* **2012**, *80* (10), 2447-2458. DOI: https://doi.org/10.1002/prot.24129

(107) Bertoldi, M.; Cellini, B.; D'Aguanno, S.; Voltattorni, C. B. Lysine 238 is an essential residue for α,β-elimination catalyzed by *treponema denticola* cystalysin *. *J. Biol. Chem.* **2003**, *278* (39), 37336-37343. DOI: 10.1074/jbc.M305967200 (acccessed 2025/09/15).

(108) Smith, G. K.; Ke, Z.; Hengge, A. C.; Xu, D.; Xie, D.; Guo, H. Active-site dynamics of SpvC virulence factor from salmonella typhimurium and density functional theory study of phosphothreonine lyase catalysis. *J. Phys. Chem. B* **2009**, *113* (46), 15327-15333. DOI: 10.1021/jp9052677

(109) Tai, C.-H.; Cook, P. F. Pyridoxal 5'-phosphate-dependent α,β-elimination reactions: Mechanism of O-acetylserine sulfhydrylase. *Acc. Chem. Res.* **2001**, *34* (1), 49-59. DOI: 10.1021/ar990169l

(110) Rege, V. D.; Kredich, N. M.; Tai, C.-H.; Karsten, W. E.; Schnackerz, K. D.; Cook, P. F. A change in the internal aldimine lysine (K42) in O-acetylserine sulfhydrylase to alanine indicates its importance in transimination and as a general base catalyst. *Biochemistry* **1996**, *35* (41), 13485-13493. DOI: 10.1021/bi961517j

(111) Gilboa, R.; Zharkov, D. O.; Golan, G.; Fernandes, A. S.; Gerchman, S. E.; Matz, E.; Kycia, J. H.; Grollman, A. P.; Shoham, G. Structure of formamidopyrimidine-DNA glycosylase covalently complexed to DNA*. *J. Biol. Chem.* **2002**, *277* (22), 19811-19816. DOI: https://doi.org/10.1074/jbc.M202058200

(112) Fromme, J. C.; Verdine, G. L. Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine DNA glycosylase MutM. *Nat Struct Biol* **2002**, *9* (7), 544-552.

(113) Zharkov, D. O.; Golan, G.; Gilboa, R.; Fernandes, A. S.; Gerchman, S. E.; Kycia, J. H.; Rieger, R. A.; Grollman, A. P.; Shoham, G. Structural analysis of an *escherichia coli* endonuclease VIII covalent reaction intermediate. *EMBO J* **2002**, *21* (4), 789-800. DOI: https://doi.org/10.1093/emboj/21.4.789

(114) Huskova, A.; Dinesh, D. C.; Srb, P.; Boura, E.; Veverka, V.; Silhan, J. Model of abasic site DNA cross-link repair; from the architecture of NEIL3 DNA binding domains to the X-structure model. *Nucleic Acids Res.* **2022**, *50* (18), 10436-10448. DOI: 10.1093/nar/gkac793 (acccessed 10/16/2025).

(115) Fromme, J. C.; Verdine, G. L. Structure of a trapped endonuclease III-DNA covalent intermediate. *EMBO J.* **2003**, *22* (13), 3461-3471. DOI: 10.1093/emboj/cdg311

(116) da Silva, A. L. G.; da Rosa, H. T.; Karnopp, T. E.; Charlier, C. F.; Ellwanger, J. H.; Moura, D. J.; Possuelo, L. G.; Valim, A. R. d. M.; Guecheva, T. N.; Henriques, J. A. P. Evaluation of DNA damage in COPD patients and its correlation with polymorphisms in repair genes. *BMC Med. Genet.* **2013**, *14* (1), 93. DOI: 10.1186/1471-2350-14-93

(117) Li, H.; Horke, S.; Förstermann, U. Oxidative stress in vascular disease and its pharmacological prevention. *Trends Pharmacol. Sci.* **2013**, *34* (6), 313-319. DOI: https://doi.org/10.1016/j.tips.2013.03.007

(118) Chamorro, Á.; Dirnagl, U.; Urra, X.; Planas, A. M. Neuroprotection in acute stroke: Targeting excitotoxicity, oxidative and nitrosative stress, and inflammation. *Lancet. Neurol.* **2016**, *15* (8), 869-881. DOI: https://doi.org/10.1016/S1474-4422(16)00114-9

(119) Yang, Z.; Kulik, H. J. Protein dynamics and substrate protonation states mediate the catalytic action of trans-4-hydroxy-l-proline dehydratase. *J. Phys. Chem. B* **2021**, *125* (28), 7774-7784. DOI: 10.1021/acs.jpcb.1c05320

(120) Drago, V. N.; Campos, C.; Hooper, M.; Collins, A.; Gerlits, O.; Weiss, K. L.; Blakeley, M. P.; Phillips, R. S.; Kovalevsky, A. Revealing protonation states and tracking substrate in serine hydroxymethyltransferase with room-temperature X-ray and neutron crystallography. *Commun. Chem.* **2023**, *6* (1), 162. DOI: 10.1038/s42004-023-00964-9

(121) McCullough, A. K.; Dodson, M. L.; Lloyd, R. S. Initiation of base excision repair: Glycosylase mechanisms and structures. *Annu. Rev. Biochem* **1999**, *68* (Volume 68, 1999), 255-285. DOI: https://doi.org/10.1146/annurev.biochem.68.1.255

(122) Perillo, B.; Di Donato, M.; Pezone, A.; Di Zazzo, E.; Giovannelli, P.; Galasso, G.; Castoria, G.; Migliaccio, A. ROS in cancer therapy: The bright side of the moon. *Exp. Mol. Med.* **2020**, *52* (2), 192-203. DOI: 10.1038/s12276-020-0384-2

(123) Khan, A. Q.; Rashid, K.; AlAmodi, A. A.; Agha, M. V.; Akhtar, S.; Hakeem, I.; Raza, S. S.; Uddin, S. Reactive oxygen species (ROS) in cancer pathogenesis and therapy: An update on the role of ROS in anticancer action of benzophenanthridine alkaloids. *Biomed. Pharmacother.* **2021**, *143*, 112142. DOI: https://doi.org/10.1016/j.biopha.2021.112142

(124) Liu, P.; Qian, L.; Sung, J.-S.; de Souza-Pinto, N. C.; Zheng, L.; Bogenhagen, D. F.; Bohr, V. A.; Wilson Iii, D. M.; Shen, B.; Demple, B. Removal of oxidative DNA damage via FEN1-dependent long-patch base excision repair in human cell mitochondria. *Mol. Cell. Biol.* **2008**, *28* (16), 4975-4987. DOI: 10.1128/MCB.00457-08

(125) Yang, F.; Hu, Z.; Guo, Z. Small-molecule inhibitors targeting FEN1 for cancer therapy. *Biomolecules* **2022**, *12* (7), 1007.

(126) He, L.; Yang, H.; Zhou, S.; Zhu, H.; Mao, H.; Ma, Z.; Wu, T.; Kumar, A. K.; Kathera, C.; Janardhan, A.; et al. Synergistic antitumor effect of combined paclitaxel with FEN1 inhibitor in cervical cancer cells. *DNA Repair* **2018**, *63*, 1-9. DOI: https://doi.org/10.1016/j.dnarep.2018.01.003

(127) Wu, T.; Zhu, H.; Zhang, M.; Sun, Y.; Yang, Y.; Gu, L.; Zhang, J.; Mu, D.; Wu, C.; Hu, Z.; et al. FEN1 inhibitor synergizes with low-dose camptothecin to induce increased cell killing via the mitochondria mediated apoptotic pathway. *Gene Ther.* **2022**, *29* (7), 407-417. DOI: 10.1038/s41434-020-00215-9

Chapter 6: How Do DICER1 Syndrome Mutations Disrupt Catalysis? Unveiling Dicer Metal Binding Architecture and Mechanism of Action using MD and QM/MM Techniques

6.1. Introduction

RNA interference is a vital pathway in human cells for gene regulation and defense against viral infections (Figure 6.1).[1] In this process, non-coding microRNAs (miRNA) or small-interfering RNAs (siRNA) are loaded into an Argonaute protein and associated binding partners to form the RISC.[1, 2] The RISC then targets messenger RNA (mRNA) that is complementary to the RNA guide strand and degrades the mRNA or otherwise prevents its translation.[1, 2] miRNA production is highly regulated along the entire transcription process.[3, 4] The final step of miRNA biogenesis is the cleavage of two phosphodiester bonds in the double-stranded precursor miRNA (pre-miRNA) by an endonuclease called Dicer to produce the approximately 22 nucleotide long mature miRNA duplex.[3, 5, 6] Due to cell requirements of processing an extremely diverse array of miRNAs,[7] Dicer has a broad substrate specificity, targeting RNA with a variety of sequences at similar rates.[8-10]

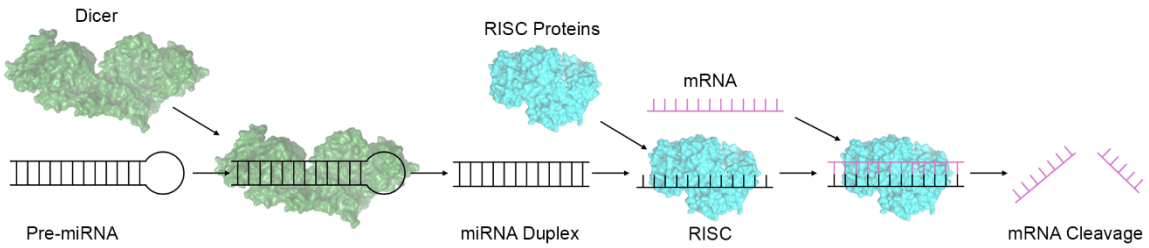


Figure 6.1. The RNA interference (RNAi) pathway in which pre-miRNA is cleaved by Dicer to produce miRNA and mRNA is cleaved by the RISC to silence genes.

miRNA has been implicated in a multitude of physiological processes in humans including cell differentiation,[11] carcinogenesis,[12-14] cardiovascular disease,[15] the immune system,[16] and viral defense.[17, 18] As a result, Dicer misfunction and miRNA dysregulation are associated with a wide variety of diseases and disorders.[11, 19-24] Furthermore, Dicer mutations correlate with the development of lung, thyroid, and brain cancers, including early childhood variants,[21, 25-30] in a genetic disorder called DICER1 syndrome. DICER1 syndrome arises from mutations to the

DICER1 hotspot region in the catalytic RNase IIIb domain,[25, 31] with the impacted residues being proposed to bind catalytic Mg^{2+} ions that facilitate phosphodiester bond cleavage.[31-34] These mutations impair the processing of miRNA-5′ strands,[31] which could compromise tumor suppressor miRNA and lead to the predisposition for cancer.[22, 35-37] Despite the direct connection between Dicer catalysis and DICER1 syndrome, the atomic details of the mechanism behind Dicer–mediated phosphodiester bond cleavage are poorly understood.

Dicer is a large, magnesium-dependent,[34, 38] multi-domain protein consisting of the dsRNA-binding, DExD/H-box helicase, DUF283, platform, PAZ, and two catalytic RNase III (labeled a and b) domains (Figure E.1, Appendix E).[32, 33] However, a truncated form of Dicer that only contains the RNase IIIb domain retains catalytic activity.[38] All nucleases function under the same general principle, with a nucleophile attacking the phosphorus atom of the phosphate moiety and the enzyme stabilizing charge build up on the phosphate and leaving group (Figure 6.2).[39, 40] However, the strategies different enzymes use to activate a nucleophile, stabilize intermediates, and facilitate leaving group departure vary significantly, with many endonucleases, including Dicer, utilizing metal cofactors.[39, 40]

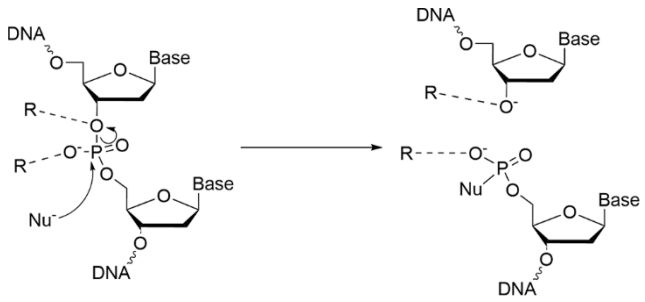


Figure 6.2. General mechanism for nucleases. Nu represents a generic nucleophile and R represents a metal ion or protein residue.

While Mg^{2+} ions are well accepted to be necessary for Dicer catalytic acivity,[34, 38] the metal coordination in the active site is unclear. An early cryo-EM structure of Dicer bound to an RNA substrate (PDB ID: 5ZAL, Figure 6.3A) lacks Mg^{2+} ions and the RNA is not aligned in the active site of either RNase III domain.[33] Although RNA is bound in the catalytic active site in a later cryo-EM

structure of Dicer crystallized in the presence of Ca^{2+} (PDB ID: 7XW2, Figure 6.3B),[32] calcium inhibition of enzymatic activity suggests that the active site conformation is not conducive for catalysis. A structure of the apo-RNase IIIb domain crystallized in the presence of magnesium shows two ions bound in the active site separated by 7.6 Å (PDB ID: 2EB1, Figure 6.3C).[38] This distance is too large for both metal ions to coordinate to the substrate near the bond cleavage site, as commonly observed for other nucleases including the homologous *aquifex aeolicus* (aa) RNase III (Figure 6.3D).[41-50] However, catalysis could still be possible through the observed human Dicer Mg^{2+} configuration if one metal adopts indirect (water mediated) coordination. Indeed, indirect metal coordination to the substrate leaving group occurs for many two-metal-dependent nucleases[45, 51-53] and several one-metal-dependent nucleases cleave the phosphodiester bond using a single Mg^{2+} that is indirectly coordinated to the substrate.[51, 54-56] Nonetheless, metal locations could shift upon substrate binding as reported for the EcoRV endonuclease.[57]

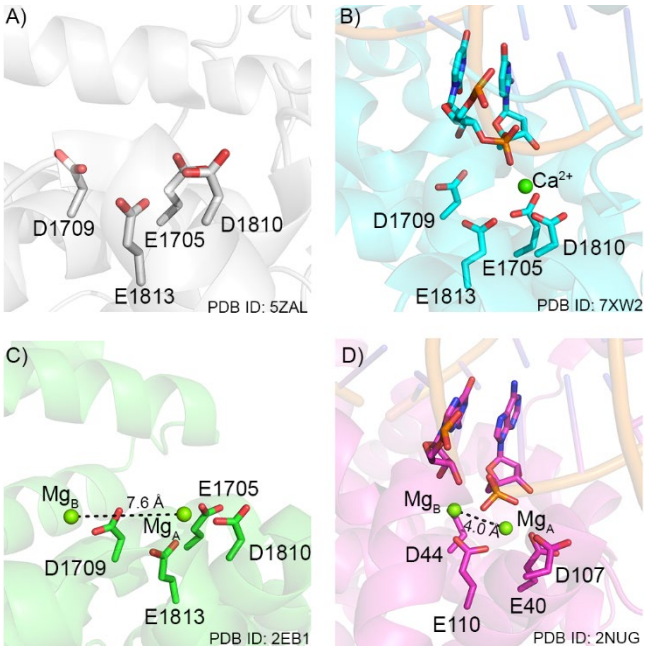


Figure 6.3. Cryo-EM structures of the active site of A) human Dicer without the substrate bound in the catalytic active site, and B) Ca^{2+}–inhibited human Dicer bound to RNA. X-ray crystal structure of the active site of C) the human Dicer RNase IIIb domain homodimer in the absence of substrate and D) aa-RNase III bound to substrate RNA.

In addition to questions surrounding the coordination of active site Mg^{2+} ions, the roles of active site residues are also unclear. K1806 has been implicated to have an important function in Dicer activity as mutation of the analogously positioned residue in mouse Dicer inhibits catalysis (Figure E.2A and B, Appendix E).[58] A similarly positioned lysine residue can also be found in the active site of other nucleases (Figure E.2C−F, Appendix E),[45, 51, 57-62] which mutational experiments (EcoRV[57, 61] and MutH[60]) and computational studies (EndoV[51]) suggest to be important for catalysis. However, the function of this lysine in nucleases is undetermined, with proposed roles for the residue including stabilizing charge build up in the transition state,[51, 61] facilitating DNA bending,[57] and mediating metal cofactor binding.[59] Without a clear picture of Mg^{2+} ion and substrate binding and the roles of functionally important active site residues, questions remain regarding how Dicer cleaves phosphodiester bonds and how DICER1 syndrome related mutations disrupt this function.

Computational chemistry techniques have proven to be invaluable for discerning enzyme structures and catalytic mechanisms,[63-69] including for nucleases.[41, 47, 51, 54, 56, 70-85] Specifically, molecular dynamics (MD) simulations have aided in the characterization of enzyme structural features, including metal binding architectures. For example, MD simulations uncovered different apurinic/apyrimidinic endonuclease 1 (APE1)–pol-β[86] and APE1–substrate[87-89] interactions to help identify residues responsible for binding active site Mg^{2+} ions and reveal the broader catalytic activity.[90-92] Quantum mechanics/molecular mechanics (QM/MM) calculations have also been used to elucidate the mechanisms of action for nucleases.[41, 47, 51, 70, 72, 78-81] Indeed, accurate QM/MM energy barriers enable comparison to experimental data and the identification of the preferred mechanism of action among many possibilities.[93] For example, QM/MM calculations on bacterial EndoV uncovered the preferred (indirect) Mg^{2+} ion−substrate coordination and the role of lysine as the general acid.[56]

Due to previous insights computational methods have provided for nucleases, the present study employs a combination of MD simulations and QM/MM calculations to uncover the binding mode of the substrate and Mg^{2+} ions in the Dicer active site, the catalytic mechanism, and the function of Dicer mutants. Specifically, MD simulations were initiated from a model of the Dicer

catalytic RNase IIIb domain bound to an RNA substrate to determine the relative positioning of the substrate, Mg^{2+}, and key active site residues. Two models were considered that differ in the initial magnesium ion placement and coordination. Subsequently, snapshots from MD simulations on each model were used to initiate QM/MM calculations under the ONIOM formalism, modeling seven different pathways to identify the preferred catalytic mechanism. Armed with knowledge of the ion configuration and roles of critical active site residues, MD simulations were used to determine the impact of select DICER1 syndrome causing mutants (D1709N, D1810Y, E1705K, E1813D, E1813G, and G1809R)[25, 31] on the structural dynamics of the catalytically-conducive enzyme–substrate complex. Overall, this work provides the first description of the Dicer catalytic mechanism of action, clarifies the role of a catalytic lysine that may be conserved in other endonucleases, and provides a structural rationalization for the connection between Dicer mutants and DICER1 syndrome.[25, 31] This information can be used in the future to develop innovative treatments for Dicer1 syndrome and powerful new RNAi–based therapeutics[94-96] and biotechnologies.[97-99]

6.2. Computational Methods

6.2.1. MD simulations

An initial model for wild-type Dicer bound to RNA was built using the crystal structure of the Dicer RNase IIIb domain homodimer (PDB ID: 2EB1, Figure 6.3C).[38] The substrate was added by aligning the homodimer with the crystal structure of the homologous *aa*-RNase III bound to RNA (PDB ID: 2NUG, Figures 6.3D and E.3A, Appendix E)[49] and removing the bacterial enzyme. Side chain conformations were minimized to reduce close contacts between the RNA substrate and Dicer using PyMOL.[100] The side chain for partially resolved E142 was added using PyMOL.[100] The unresolved amino acid chain consisting of residues N122 to E141 was generated using SWISS-MODEL.[101] Protonation states for amino acid side chains were calculated using the H++ server.[102, 103] From this starting structure, two models were generated that differ in the position and coordination geometry of the two active site Mg^{2+} ions. In the first model (denoted RIIIb), the Mg^{2+} ions are retained from the Dicer RNase IIIb homodimer crystal structure (PDB ID: 2EB1,

Figure 6.4A),[38] which are separated by 7.6 Å. Specifically, Mg_A^{2+} is coordinated to E1705, D1810, E1813, a nonbridging phosphate oxygen, and two water molecules, while Mg_B^{2+} is coordinated to D1709 and 5 water molecules. In the second model (denoted aaRIII), the Mg^{2+} ions were retained from the *aa*–RNase III crystal structure (PDB ID: 2NUG, Figure 6.4B) and are separated by 4.0 Å.[49] Specifically, Mg_A^{2+} is coordinated to E1705, D1810, E1813, a nonbridging phosphate oxygen, and two water molecules, while Mg_B^{2+} is coordinated to E1813, D1709, a bridging and non-bridging phosphate oxygen, and two water molecules.

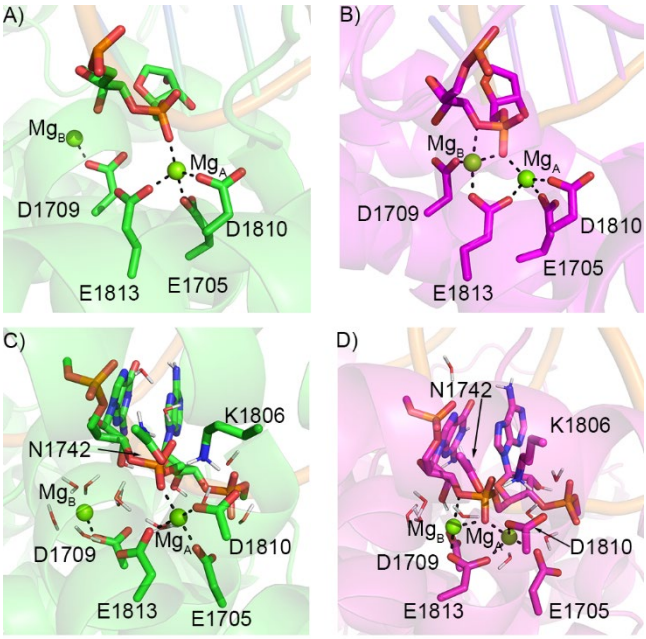


Figure 6.4. Post-equilibration Dicer model constructed using the Mg^{2+} ion positioning from the A) RNase IIIb homodimer (denoted RIIIb) and B) *aa*-RNase III (denoted aaRIII). High-layer atoms in the QM/MM C) RIIIb and D) aaRIII models.

Each model was solvated with TIP3P water molecules in a periodic rectangular prism with sides a minimum of 10 Å from the solute in each direction using the LEaP program in AMBER 2018.[104] Sodium ions were added to neutralize the system and NaCl was added to achieve a physiological salt concentration (150 mM). The Amber force field was used throughout, with protein residues described using ff14SB parameters[105] and nucleic acid components modeled using χOL3.[106, 107]

Both the RIIIb and aaRIII models were initially minimized in a stepwise manner, using 1000 steps of steepest decent followed by 1000 steps of conjugate gradient minimization. First, the solvent was minimized, while a 100 kcal mol^{-1} $Å^{-2}$ restraint was applied to the solute. Next, the restraint was removed from the solute hydrogen atoms. Subsequently, all solute atoms were minimized, while applying a 100 kcal mol^{-1} $Å^{-2}$ restraint to the solvent. Finally, the entire system was minimized with no restraints, using 1000 steps of steepest decent followed by 2000 steps of conjugate gradient minimization. Each system was then heated from 10 K, with the temperature increased by 50 K every 20 ps until 310 K was reached. The restraints on the solute were subsequently dropped from 25 kcal mol^{-1} $Å^{-2}$ to 5 kcal mol^{-1} $Å^{-2}$ at a rate of 5 kcal mol^{-1} $Å^{-2}$ per 20 ps. A final 20 ps equilibration step was completed while applying a solute restraint of 1.5 kcal mol^{-1} $Å^{-2}$. For these simulations, an NVT ensemble was used, with a collision frequency of 1 ps^{-1} and a Langevin thermostat. As an additional equilibration step for the aaRIII model, a 100 ns production simulation was performed where the distances between Mg_A^{2+} and the phosphate non-bridging oxygen, Mg_B^{2+} and the phosphate non-bridging oxygen, and Mg_B^{2+} and O3′ were constrained to be less than 2.4 Å, which allowed the system to relax in the presence of the added Mg^{2+} ions and substrate. Following equilibration, five 500 ns MD production simulations were performed on each system. The additional equilibration step and production simulations employed an NPT ensemble with the Langevin thermostat, Berendsen barostat (1 bar), and a collision frequency of 3 ps^{-1}. Minimization, equilibration, and production simulations were performed using AMBER 2018 pmemd.cuda.[104] Heavy atom RMSDs from the simulations confirm system stability (Figure E.4, Appendix E). Overlays between the post equilibration aaRIII and RIIIb models and the Ca^{2+} inhibited crystal structure show comparable Mg^{2+}−binding residue and substrate positioning (Figure E.3B and C, Appendix E), which indicates there were no large structural shifts upon addition of the substrate and ions from the bacterial crystal structure.

Directed by calculations on the wild-type enzyme, D1709N, D1810Y, E1705K, E1813D, E1813G, and G1809R Dicer mutants were generated by substituting the appropriate residue in the post-equilibration aaRIII model using the PyMOL mutagenesis tool.[100] Each mutated side chain was manually manipulated to remove steric clashes. Similarly, a wild-type aaRIII model with a

hydroxide nucleophile was generated by replacing a water bound to Mg_A^{2+} with a hydroxide ion. MD simulations were run on each system using the same protocol as implemented for the wild-type models. Heavy atom RMSDs show similar behavior between replicates for each model (Figure E.5, Appendix E).

Trajectory analysis was conducted on the same, single active site across all replicas using the cpptraj program in AMBER 2018.[104] Hydrogen-bond occupancy was determine using a heavy atom distance of less than 3.0 Å and a hydrogen-bonding angle between 135° and 180°. A residue was considered coordinated to Mg^{2+} when the distance to the metal was less than 3.0 Å. A water was considered in position to attack the phosphate moiety when the distance between the phosphorus atom and water oxygen ($r(O_{WAT}–P)$) was less than 4.2 Å and the $\angle(O3'PO_{WAT})$ was greater than 135°. Representative structures shown in the figures were obtained by clustering using the heiragglo algorithm based on the all-atom RMSD of the active site Mg^{2+} ions, E1705, D1709, D1810, E1813, and the nucleotide containing the phosphate to be cleaved. RESP charge fitting was used to generate charges for the hydroxide ion with the R.E.D.v.III.4 program (Table E.1, Appendix E),[108] with other atomic parameters being the same as a TIP3P water molecule.

6.2.2. QM/MM methodology

QM/MM models were generated starting from MD representative structures for each wild-type system (RIIIb and aaRIII). First, directed by the MD analysis (see results), MD simulation data were clustered (heiragglo algorithm) based on the distance between O3′ of the leaving group and Mg_B^{2+}. Since mutation of the lysine analogously positioned to K1806 in the structurally similar mouse Dicer (K1790, Figure E.2B, Appendix E) inhibits catalysis,[58] a second round of clustering based on the K1806–substrate phosphorus distance was performed to obtain the final representative structure used for QM/MM modeling. The QM region of the RIIIb model contains 172 atoms (E1705, D1709, N1742, K1806, D1810, and E1813, 2 Mg^{2+} ions, 7 metal-coordinated waters, and 6 additional active site waters) and an overall charge of –2 (Figure 6.4C). The QM layer of the aaRIII model similarly has an overall charge of –2 and contains 160 atoms (E1705, D1709, N1742, K1806, D1810, and

E1813 all truncated at the Cα–Cβ bond, the G and A residues 5′ and 3′ with respect to the phosphate being cleaved (truncated at C3′ and C5′, respectively), 2 Mg^{2+} ions, 4 metal-coordinated waters, and 5 additional active site waters; Figure 6.4D). The aaRIII model with a hydroxide nucleophile was generated by converting one Mg^{2+}–bound water into a hydroxide ion and adding an additional water molecule to the QM layer to help stabilize the more highly charged active site, resulting in a QM region with 162 atoms and a charge of –3. In all models, the MM region contained the remaining protein and substrate atoms as well as any water molecules with an atom that falls within 8 Å of any solute atom (Figure E.6, Appendix E).

All QM/MM calculations invoked the ONIOM scheme.[109] Optimizations were performed using the M06-2X functional[110] with the 6-31G(d,p) basis set to treat the QM region, while the MM region was modeled with the same force field used for the MD simulations. This approach was chosen due its robustness and successes accurately modeling other enzymatic reactions,[76, 82, 83, 111-114] including metal–dependent enzymatic phosphodiester bond cleavage or formation.[76, 82, 83] Guesses for transition states (TS) were isolated from energy maxima obtained by scanning key reaction parameters (i.e., those corresponding to P–O bond formation (r(O_{WAT}–P)) and cleavage (r(O3′–P))), which were subsequently subjected to unrestrained transition state optimizations. Frequency calculations were conducted to characterize the nature of all stationary points and verify that the isolated imaginary frequency was associated with the anticipated transition structure. Frequencies were treated as positive if greater than −10 cm^{-1}. Each pathway was verified to be continuous by mapping the internal reaction coordinate through scanning the nucleophile–electrophile and leaving group distances both to and from each stationary point. The reported reaction Gibbs energies were obtained from ONIOM(M06-2X/6-311+G(2df,p):AMBERff14SB) single-point calculations. Optimizations were performed with the mechanical embedding (ME) scheme, while single-point calculations were performed with the electronic embedding (EE) scheme. To ensure ME optimizations do not affect the accuracy of the reaction surface, the mechanistic pathway for the aaRIII model with a water nucleophile was fully optimized with EE. Structures optimized with ME and EE are highly similar (active site all atom RMSD < 0.3 Å, Figure E.7A, Appendix E) and the resulting energy barriers are within 11 kJ/mol (Figure E.7B, Appendix E), supporting the robustness

of the ME approach for predicting the structures of this system. QM/MM activation barriers were compared to the estimated barrier for the catalytic step (~80–100 kJ/mol) calculated from experimental k_{cat} values ($1 \times 10^{-1} - 7 \times 10^{-5}$ s^{-1}) measured at 37 °C under single and multiple turnover conditions with the pre-let-7a-1 pre-miRNA substrate.[115-117]

All QM/MM calculations were performed using Gaussian 16 (Rev. B.01).[118]

6.2.3. Free energy of hydroxide binding

The Gibbs energy difference for binding a hydroxide ion compared to a water molecule at Mg_A^{2+} for the aaRIII configuration was determined using alchemical thermodynamic integration. Specifically, the endpoints were defined as the Dicer–RNA complex with the open coordination on Mg_A^{2+} filled with waters and the Dicer–RNA complex with an Mg_A^{2+}–bound water replaced by a hydroxide ion. The aaRIII post-equilibration structure was used as a starting point. The transition from one endpoint to the other was done in 5 steps, each consisting of a 100 ns simulation. The λ values for each step were 0.04691, 0.23076, 0.5, 0.76923, and 0.95308. The final energy difference was calculated using numerical integration. This same process was done using a box of 2548 water molecules, with one endpoint being a pure water box and the other endpoint having a single water replaced with OH^-. For both systems, the simulations were repeated in the opposite direction. The resulting Gibbs energy values are within 0.3 kJ/mol, showing negligible hysteresis. Block averaging was used to estimate the error in the energy. Simulations were run using the same parameters and programs as the MD simulations.

6.3. Results and Discussion

6.3.1. Indirect coordination of Mg_B^{2+} to the substrate does not yield sufficient charge stabilization of the leaving group to permit catalysis

With only a single crystal structure of Dicer containing Mg^{2+} ions bound in the active site (PDB ID: 2EB1, Figure 6.3C),[38] the feasibility of the Mg^{2+} ion configuration presented in this crystal structure upon substrate binding was considered. However, the large Mg^{2+} separation prevents both Mg^{2+} ions from directly coordinating to the substrate, with only Mg_A^{2+} being positioned for direct coordination with the scissile phosphate backbone in the Dicer−substrate complex. Nevertheless, other two-metal-dependent endonucleases are active with indirect Mg_B^{2+}−leaving group coordination[45, 51-53] and single-metal-dependent nucleases have been shown to use indirect metal−substrate coordination to cleave the phosphodiester bond.[51, 54-56] This includes the case of APE1, where an indirectly coordinated Mg^{2+} ion facilitates catalysis by having a Mg^{2+}−bound water donate a proton to the O3′ leaving group during the catalytic mechanism.[54, 77] Therefore, a model was constructed that maintains the RIIIb ion configuration to explore whether indirect Mg_B^{2+}−substrate coordination can be maintained in a dynamic environment and can effectively catalyze phosphodiester bond cleavage.

Throughout MD simulations on the RIIIb model, Mg_A^{2+} remained coordinated to E1705, D1810, E1813, a non-bridging phosphate oxygen of the substrate, and two water molecules, while Mg_B^{2+} is coordinated to D1709 and five water molecules (Figure 6.5). Furthermore, Mg_A^{2+} consistently positioned a nucleophilic water for attack at the phosphate moiety ($r(O_{WAT}–P) = 3.6 \pm 0.2$ Å for 97% of the simulation time). However, two dominant active site conformations emerged that primarily differ in the distance between Mg_B^{2+} and O3′ of the scissile phosphate, and the orientation of the Mg_B^{2+}–coordinated D1709 (∠(OδCγCβCα), Figure E.8, Appendix E). The first, lower occupancy conformation has a long Mg_B^{2+}−substrate distance ($r(Mg_B^{2+}–O3') = 7.0 \pm 0.4$ Å; 27% occupancy; Figure 6.5A), which does not allow for coordination between Mg_B^{2+} and the substrate. Accordingly, a Mg_B^{2+}–activated water is positioned to stabilize the substrate for less than 1% of the time in this conformation, suggesting this active site architecture is not conducive for catalysis (denoted

Mg_B^{2+}–NC conformation). The second, higher occupancy conformation positions Mg_B^{2+} closer to the substrate (r(Mg_B^{2+}–O3′)=4.7 ± 0.3 Å; 73% occupancy; Figure 6.5B). Although direct Mg_B^{2+}–substrate coordination remains absent, this metal binding arrangement permits water to bridge the metal and substrate to afford indirect metal–substrate stabilization as has been reported for other nucleases (denoted Mg_B^{2+}–IC conformation).[51, 54, 56, 76] Indeed, a Mg_B^{2+}–coordinated water is within 3.0 Å of O3′ for 100% of the time this conformer is adopted, which could potentially protonate the leaving group during departure as seen for APE1.[54] Interestingly, a K1806–substrate hydrogen bond rarely forms during the simulations regardless of the Mg_B^{2+} orientation (occupancy <2%), which raises questions regarding the viability of this Mg_B^{2+} metal binding architecture since mutation of the corresponding residue in mouse Dicer significantly impacts catalysis.[58] Nevertheless, QM/MM calculations were performed on the RIIIb model to probe the corresponding catalytic mechanism, with the QM/MM starting structure built from a representative MD snapshot in the Mg_B^{2+}–IC conformation with a K1806–non-bridging phosphate oxygen hydrogen bond.

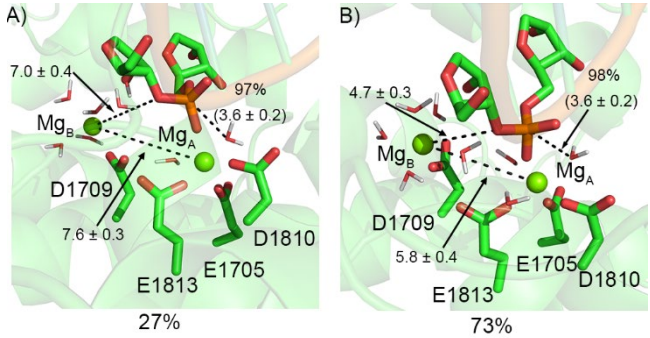


Figure 6.5. MD representative structures of the A) Mg_B^{2+}–NC and B) Mg_B^{2+}–IC conformations of the Dicer RIIIb active site. Occupancies over the course of the simulation provided (below). Distances are reported in Å.

The QM/MM optimized RC retains the Mg^{2+} coordination observed in the MD simulations, including indirect coordination between Mg_B^{2+} and O3′ of the substrate (r($O3'–H_{WAT}$) = 2.211 Å, Figures 6.6A, E.9A and E.10, Appendix E). The water nucleophile bound to Mg_A^{2+} is positioned to attack the phosphate moiety (r(O_{WAT}–P) = 3.392 Å). K1806 hydrogen bonds with a non-bridging phosphate oxygen of the substrate, which is maintained throughout the reaction. The first transition state is late, with water far into nucleophilic attack (r(O_{WAT}–P) = 1.963 Å) and a proton on the

nucleophilic water being transferred to an unbound water (r(O_{WAT}–H_{WAT}) = 1.771 Å). The resulting hydronium ion is stabilized through hydrogen bonds to the phosphate backbone and E1705. This transition state results in a very high barrier (253.4 kJ/mol). The first TS leads to an intermediate complex, with large leaving group and nucleophilic water–phosphate distances (r(O_{WAT}–P) = 1.820 Å and r(O3′–P) = 1.795 Å). When single-point and Gibbs energy corrections are taken into account, the IC is ~16 kJ/mol higher in energy than TS1 (Figure 6.6A and Table E.2, Appendix E).

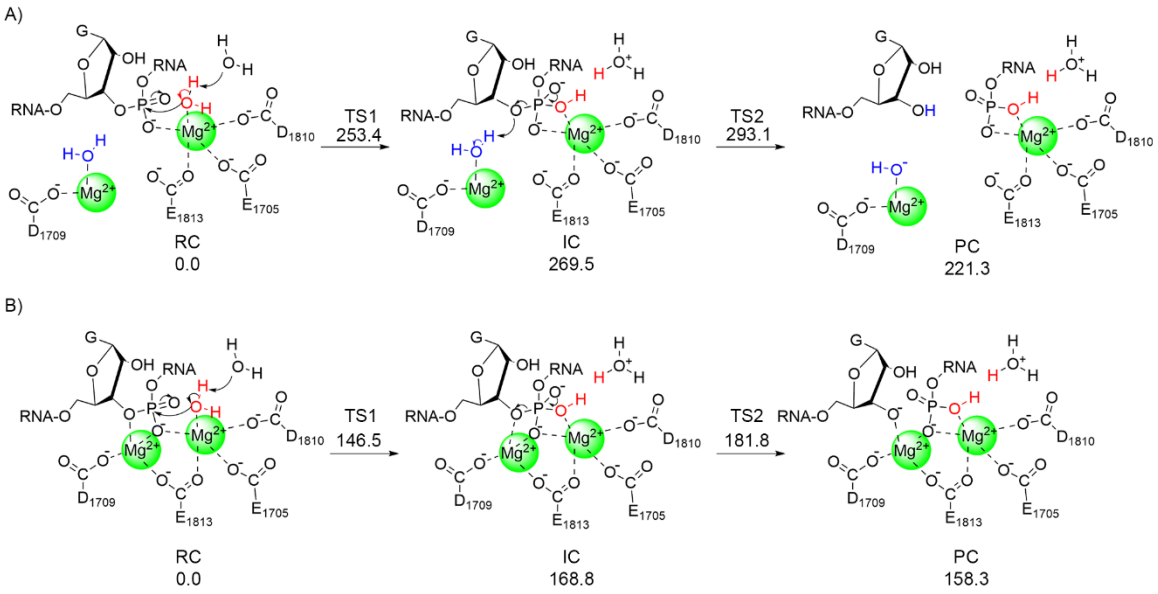


Figure 6.6. ONIOM(M06-2X/6-311+G(2df,p):AMBERff14SB)//ONIOM(M06-2X/6-31G(d,p):AMBERff14SB) calculated Dicer mechanism of action involving a water nucleophile for the A) RIIIb and B) aaRIII models. Relative energies reported in kJ/mol. Water molecules coordinated to Mg^{2+} ions that do not participate in the reaction are omitted for clarity (see Figure E.9 for detailed coordination).

The second transition state is also late, with the nucleophile–phosphate bond almost formed (O_{WAT}–P distance = 1.684 Å) and the leaving group–phosphate bond significantly extended (r(O3′–P) = 2.404 Å). Additionally, O3′ of the substrate has partially abstracted a proton from a Mg^{2+}–coordinated water (r(O3′–H_{WAT}) = 1.399 Å), which helps stabilize the charge on the leaving group. The K1806–phosphate hydrogen-bond strengthens slightly over the course of the reaction, with the hydrogen-bond distance decreasing from 2.772 Å in the RC to 2.649 Å in TS2. In the PC, the proton from the Mg_B^{2+}–bound water is fully transferred to the substrate (r(O3′···H_{WAT}) = 0.984 Å), the nucleophile–phosphate bond formed (r(O_{WAT}–P) = 1.623 Å), and the leaving group–phosphate bond

cleaved (r(O3′–P) = 2.751 Å). The energy barrier associated with TS2 is rate limiting and extremely high (293.1 kJ/mol, Figure 6.6A and Table E.2, Appendix E), likely at least in part due to inadequate stabilization of the charge developing on the phosphate moiety due to the lack of direct Mg_B^{2+}–substrate coordination.

As the QM/MM predicted rate-determining barrier is ~200 kJ/mol above the estimated experimental barrier (80–100 kJ/mol),[115-117] a Dicer catalytic mechanism involving indirect Mg_B^{2+} coordination to the substrate is unfeasible. This conclusion is further supported by the low occupancy of the K1806–substrate hydrogen bond during the MD simulations, with K1806 expected to play a role in catalysis.[58] Thus, despite the predicted catalytic mechanism for other endonucleases involving similar indirect metal–substrate coordination,[51, 54, 56, 76] this metal binding geometry in the RNase IIIb domain of Dicer is not conducive for catalysis. This confirms a previous proposal that the addition of RNA shifts the location of active site Mg^{2+} ions to yield a functional form.[38]

6.3.2. Dicer phosphodiester bond cleavage facilitated by direct Mg_B^{2+}–substrate coordination and a Mg_A^{2+}–bound water nucleophile is kinetically unfavoured

The high barrier associated with indirect metal–substrate coordination suggests insufficient O3′–leaving group stabilization is provided by the water bridging the metal and substrate. Therefore, a new model was considered that mirrors the coordination geometry observed in the homologous *aa*-RNase III (PBD ID: 2NUG, Figure 6.3D),[49] which decreases the distance between the Mg^{2+} ions and allows both ions to directly coordinate to the substrate to stabilize charge build up on the phosphate and facilitate leaving group departure along the nuclease mechanism. Indeed, an analogous metal–substrate binding geometry has been reported for other nucleases,[41-50] including HIV RNase H,[79] CRISPR-Cas9,[70] BamHI,[81] and the homologous *aa*−RNase III.[49]

To explore direct metal−substrate coordination in the Dicer active site, MD simulations were initiated from the aaRIII conformer, with Mg_A^{2+} coordinated to E1705, D1810, E1813, a non-bridging

phosphate oxygen of the substrate, and two water molecules, while Mg_B^{2+} is coordinated to D1709, E1813, a bridging and a non-bridging phosphate oxygen of the substrate, and two water molecules. The simulations highlight the conformational flexibility of the Dicer active site, with three distinct active site conformers adopted that differ in Mg^{2+} ion coordination to the substrate (Figure E.11, Appendix E). For the majority of the MD simulation time (61%) both Mg^{2+} ions are bound to the substrate, mirroring the aaRIII starting structure, denoted the two Mg^{2+}−bound conformation, or $2Mg^{2+}$−B (Figure 6.7A). This ensures a water nucleophile is correctly positioned with respect to the substrate to initiate the reaction (100% occupancy; $r(O_{WAT}–P)$ = 3.4 ± 0.2 Å) and the active site is aligned to afford leaving group stabilization ($r(Mg_B^{2+}–O3')$ = 2.5 ± 0.2 Å). This conformation maintains the Mg^{2+} ions within 3.5 ± 0.1 Å, which is comparable to the distance reported for other enzymes that invoke a two-metal-mediated mechanism.[41-50, 119] Additionally, K1806, which is known to be critical for the activity of mouse Dicer,[58] hydrogen bonds with the substrate phosphate moiety for 59% of the time this conformation is adopted, suggesting this dominant conformer is aligned for catalysis. A second metal coordination geometry is adopted for 34% of the total trajectory in which the coordination of Mg_A^{2+} to the non-bridging phosphate oxygen is replaced with coordination to water (designated the Mg_A^{2+}−unbound conformation or Mg_A^{2+}−U, Figure 6.7B). The loss of direct Mg_A^{2+}–substrate coordination would prevent nucleophilic attack at the substrate as Mg_A^{2+} is too far from the phosphorus ($r(Mg_A^{2+}–P)$ = 4.0 ± 0.2 Å). Additionally, this conformer is associated with the loss of Mg_B^{2+}–O3′ coordination, which would significantly diminish leaving group stabilization. Therefore, this active site configuration is not catalytically viable. For the remaining simulation time (5% occupancy across the total simulation), direct coordination of Mg_B^{2+} to the bridging and non-bridging phosphate oxygens is replaced by coordination to an additional water molecule and D1713 (denoted the Mg_B^{2+}−unbound conformation or Mg_B^{2+}−U, Figure 6.7C) and the K1806 hydrogen bond with the scissile phosphate is absent. While indirect Mg_B^{2+}–substrate coordination is maintained in this conformer, our previous QM/MM calculations highlight that indirect coordination is not sufficient to enable catalysis. Therefore, only the dominant $2Mg^{2+}$−B conformation has the potential to be conducive for phosphodiester bond hydrolysis.

To determine whether the dominant conformation is catalytically feasible, QM/MM calculations were performed from the corresponding representative structure. In the optimized RC, the metal ions maintain the coordination observed in the MD simulations (Figures 6.6B, E.7C and E.9B, Appendix E). K1806 hydrogen bonds to a non-bridging phosphate oxygen, with this interaction being maintained and slightly strengthened throughout the reaction (r($O_{Phosphate}$–K1806) is reduced

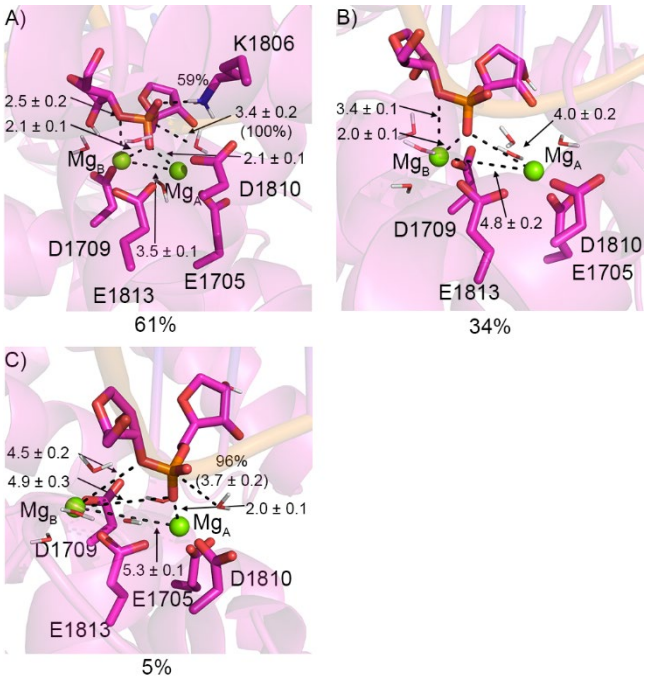


Figure 6.7. MD representative structures of the A) $2Mg^{2+}$–B, B) Mg_A^{2+}–U, and C) Mg_B^{2+}–U conformations of the Dicer active site for the aaRIII model with a water nucleophile. Occupancies over the course of the simulation provided (below).

from 2.651 Å in RC to 2.607 Å in TS2). The water nucleophile is positioned to attack phosphorus (r(O_{WAT}–P) = 2.911 Å). As observed for the RIIIb model, TS1 is late (r(O_{WAT}–P) = 2.040 Å) and the nucleophilic water has lost a proton (r(O_{WAT}–H_{WAT}) = 1.677 Å) to generate a hydronium ion that is stabilized by hydrogen bonding to the phosphate backbone and E1705. This first reaction step has a high barrier (146.5 kJ/mol, Figure 6.6B and Table E.2, Appendix E). The IC has long phosphorus distances to the leaving group and nucleophile (r(O3′–P) = 1.831 Å and r(O_{WAT}–P) = 1.924 Å) and is higher in energy than TS1 after accounting for single-point and Gibbs corrections (168.8 kJ/mol, Table E.2, Appendix E). The second transition state is also late (r(O3′–P) = 2.383 Å), with the

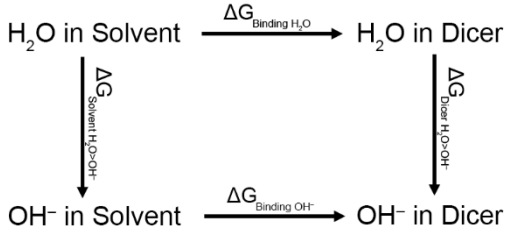
nucleophile attack nearly complete (r(O_{WAT}–P) = 1.717 Å). This leads to a product complex in which the backbone is cleaved (r(O3′–P) = 2.720 Å) and the nucleophile–phosphate bond is fully formed (r(O_{WAT}–P) = 1.676 Å). TS2 is associated with the rate-limiting step, with a calculated barrier of 181.8 kJ/mol (Figure 6.6B).

Although the barrier for phosphodiester bond catalysis involving direct Mg_B^{2+}−substrate coordination is over 100 kJ/mol lower than that involving indirect Mg_B^{2+}−substrate coordination, the predicted barrier remains significantly higher than the experimental estimates (~80–100 kJ/mol),[115-117] likely a consequence of the lack of a strong base for nucleophile activation. Therefore, despite this second mechanism mirroring the catalytically favourable pathway elucidated for other two-metal-dependent nucleases (e.g., HIV RNase H,[79] CRISPR-Cas9,[70] and EndoV[51]), this mechanism is unfeasible for Dicer.

6.3.3. Dicer phosphodiester bond cleavage is catalytically favorable with direct Mg_B^{2+}–substrate coordination and a hydroxide nucleophile

Although the Dicer mechanism involving a water nucleophile that parallels that used by many two-metal-dependent nucleases[47, 70, 78, 79, 82] is associated with a high energy barrier, other nucleophiles are available in a cell. Specifically, hydroxide ions are also available in a cellular environment, albeit existing at a much lower concentration than water, and are much stronger nucleophiles than water. Indeed, hydroxide nucleophiles are known to be utilized by other two-metal-dependent nucleases,[41, 81, 84, 85, 119-121] including cases where a water nucleophile has proven to be insufficient for catalysis.[81, 84, 119] Additionally, some hydroxide–dependent nucleases, such as Ribonuclease H[41] and BamHI,[81] have similar active site architectures as Dicer, with 4 D/E residues binding two Mg^{2+} ions in the active site.

Scheme 1. Thermodynamic cycle used to calculate the $\Delta\Delta G_{Binding}$ for replacing water with hydroxide in the Dicer active site.



Thermodynamic integration was used (Scheme 1) to determine the energetic feasibility of binding a hydroxide ion in the Dicer active site as has been done for other enzymes,[84, 122, 123] including nucleases.[84, 122] Our calculations predict that hydroxide binding is favorable, with a binding free energy of –21.7 ± 1.8 kJ/mol (Figure E.12, Appendix E). However, the free energy penalty (ΔG) for binding a low concentration ligand needs to be taken into account. With the initial volume (V_i) calculated assuming a 1 x 10^{-7} M concentration of hydroxide in solution and a water density of 0.997 g/mL, the final volume (V_f) calculated based on two potential hydroxide binding sites on Mg_A^{2+}, and a temperature of 310 K, the free energy penalty (ΔG) is estimated to be 50.1 kJ/mol (equation 1). The resulting, total energy cost for hydroxide binding (28.4 kJ/mol) is significantly less than the experimental barrier for catalysis (80–100 kJ/mol),[115-117] suggesting that binding of a nucleophilic hydroxide ion to Mg_A^{2+} in the Dicer active site is energetically viable.

$$\Delta G = -k_B T \ln\left(\frac{V_f}{V_i}\right) \quad (1)$$

To investigate whether a pathway involving a hydroxide ion nucleophile is catalytically feasible, a QM/MM model was built by replacing a Mg_A^{2+}–bound water in the QM/MM optimized RC with hydroxide (Figure E.9C, Appendix E). Due to charge repulsion, the hydroxide nucleophile is positioned slightly further from the phosphate in the optimized RC ($r(O_{OH}–P)$ = 3.077 Å) compared to the water nucleophile (2.911 Å, Figure E.13, Appendix E). K1806 forms a hydrogen bond with a non-bridging phosphate oxygen throughout the reaction, which strengthens as the reaction progresses ($r(O_{Phosphate}–K1806)$ decreases by ~0.13 Å). Unlike the previously characterized late TS

associated with a water nucleophile, the first transition state corresponding to hydroxide ion attack is early, with the hydroxide nucleophile just starting to form a bond to phosphorus ($R(O_{OH}–P)$ = 2.406 Å). TS1 results in a low barrier of 39.2 kJ/mol. The phosphorane intermediate contains a slightly extended phosphorus–hydroxide bond ($r(O_{OH}–P)$ = 1.805 Å) and an elongated phosphorus–leaving group distance (r(O3′–P) = 1.940 Å). Unlike models involving a water nucleophile, the IC is slightly thermodynamically stable compared to TS1 when single-point and Gibbs energy corrections are applied, falling 36.8 kJ/mol above the RC (Figure 6.8). TS2 is also earlier than the analogous transition state for the pathway involving a water nucleophile, with a long phosphorus–nucleophile distance ($r(O_{OH}–P)$ = 1.800 Å) and a short phosphorus–leaving group distance (r(O3′–P) = 1.967 Å). The stronger nucleophile coupled with direct Mg_B^{2+}–O3′ coordination for leaving group stabilization leads to a low calculated rate-limiting barrier of 45.0 kJ/mol (Figure 6.8). When combined with the previously predicted penalty for hydroxide binding in the active site, the total energetic cost for the reaction is 73.4 kJ/mol.

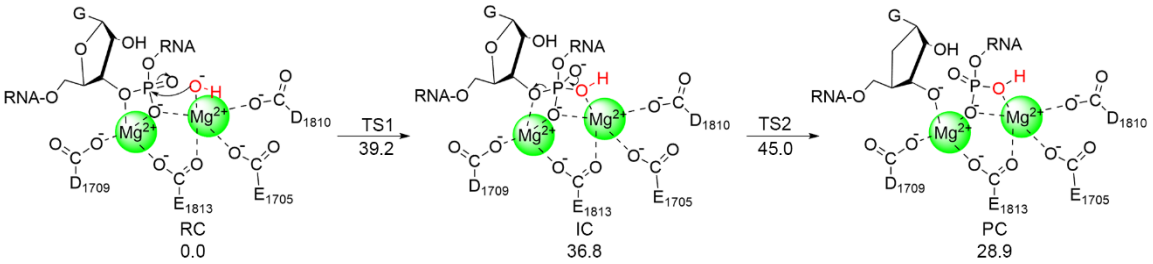


Figure 6.8. ONIOM(M06-2X/6-311+G(2df,p):AMBERff14SB)//ONIOM(M06-2X/6-31G(d,p):AMBERff14SB) calculated Dicer mechanism of action for the aaRIII model with a hydroxide nucleophile. Relative energies reported in kJ/mol. Water molecules coordinated to Mg^{2+} ions that do not participate in the reaction are omitted for clarity.

Since a hydroxide nucleophile results in a catalytically viable barrier, MD simulations were performed on the corresponding RC to ensure the hydroxide nucleophile bound to Mg_A^{2+} does not significantly change the active site structural dynamics. Two distinct active site conformations were observed (Figure 6.9). In the major conformation (71% occupancy), Mg_A^{2+} is coordinated to E1705, D1810, E1813, a non-bridging phosphate oxygen of the substrate, one water molecule, and the hydroxide nucleophile, while Mg_B^{2+} is coordinated to D1709, E1813, a bridging and a non-bridging

phosphate oxygen of the substrate, and two water molecules (Figure 6.9A). The hydroxide nucleophile is well positioned for nucleophilic attack ($r(O_{OH}–P)$ = 3.4 ± 0.2 Å; 100% occupancy) and a hydrogen bond between the substrate and K1806 is formed for 39% of the simulation time. In the minor conformation (29% occupancy), the coordination of Mg_A^{2+} to the non-bridging phosphate oxygen and Mg_B^{2+} to the bridging phosphate oxygen are replaced with coordination to water, and the K1806–substrate hydrogen bond is absent (Figure 6.9B). These conformations and occupancies mirror the $2Mg^{2+}$−B and Mg_A^{2+}−U conformations observed for the models with a water nucleophile (Figures 6.7 and 6.9). Indeed, the representative structures of the dominant $2Mg^{2+}$−B conformation for a water or hydroxide nucleophile are highly similar (active site RMSD = 0.682 Å, Figure E.14, Appendix E). Thus, in addition to being energetically viable, binding of the hydroxide nucleophile in the Dicer active site does not significantly change the active site structural dynamics.

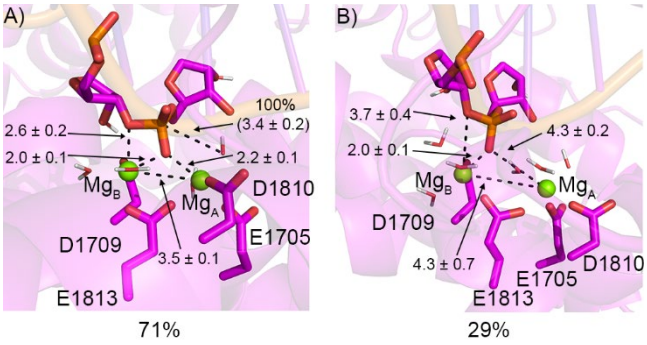


Figure 6.9. MD representative structures of the A) $2Mg^{2+}$−B and B) Mg_A^{2+}−U conformations of the Dicer aaRIII model with a hydroxide nucleophile. Occupancies over the course of the simulation provided (below).

Overall, the predicted barrier for the Dicer catalytic mechanism involving a hydroxide nucleophile (73.4 kJ/mol) is consistent with the experimental barrier (~80-100 kJ/mol),[115-117] supporting that a hydroxide nucleophile bound to Mg_A^{2+} can facilitate the reaction when combined with Mg_B^{2+} (direct) coordination to D1709, E1813, and the substrate leaving group. An analogous two-metal mediated mechanism has been characterized for other phosphodiester bond-cleaving enzymes,[41, 81, 84] such as Ribonuclease H[41] and BamHI,[81] which similarly use 4 D/E residues to bind two Mg^{2+} ions in the active site and require a hydroxide nucleophile for activity. It is notable that our predicted catalytically active Mg^{2+} configuration (Figure 6.9A) differs from that found in the crystal

structure of the human Dicer RNase IIIb domain in the absence of a substrate (Figure 6.3C), suggesting that the Mg^{2+} ions migrate upon substrate binding. This is not entirely unexpected as changes in enzyme structure upon substrate binding are common,[124-126] including changes in the metal binding configuration.[57, 127-129] For example, the *Eco*RV endonuclease, Mg^{2+} ions change binding sites along the course of the reaction to facilitate nuclease activity.[57] Furthermore, there are other examples in the literature highlighting that the Mg^{2+} ion binding configuration in a crystal structure differs from that in the catalytically active state.[54, 75, 130]

A lysine with unknown function similarly positioned next to the substrate as K1806 in Dicer is a common feature of endonucleases (Figure E.2C−F, Appendix E).[38, 45, 51, 57-61] Our results clarify the role of K1806 in catalysis, forming a hydrogen bond with the scissile phosphate that strengthens as the reaction proceeds to stabilize the transition states and intermediate in the catalytic mechanism. This is consistent with mutations to the analogous residue in mouse Dicer inhibiting catalysis,[58] and suggests conserved lysine residues in other endonuclease active sites may have a similar role.

Overall, our combined MD and QM/MM approach provides the first structural description of the Dicer RNase III domain bound to the substrate RNA and Mg^{2+} in a catalytically–conducive conformation, confirming that Mg^{2+} ion binding differs when Dicer is bound or unbound to substrate. Our calculations also provide the first characterization of the Dicer catalytic mechanism and clarify the role of the active site lysine, which can be applied to other endonucleases. The details obtained on the function of active site residues in Dicer obtained in the study will aid the rational design of artificial endonucleases. Additionally, the characterized catalytic mechanism will facilitate the rational design of modified siRNA as RNAi-based therapeutics.[94-96]

6.3.4. DICER1 syndrome related mutations disrupt Mg^{2+} stabilization of the substrate in the Dicer active site

As an enzyme that plays an integral role in miRNA biogenesis, Dicer is closely involved in maintaining human health.[11, 19-24] As such, mutations in the catalytic core of the Dicer RNase IIIb domain increase susceptibility to disease,[21, 25-30] a genetic disorder called DICER1 syndrome. Although these mutations are known to impair miRNA processing,[31] a structural rationale for reduced enzymatic activity associated with DICER1-syndrome-causing mutants is not currently available. With a new, detailed understanding of the structure of the wild-type Dicer–substrate complex, active site Mg^{2+} ion configuration, and catalytic mechanism, DICER1-syndrome-causing mutants can be investigated to understand the mechanism of their pathogenic outcomes. In the present study, the most common D1709N, D1810Y, E1705K, E1813D, E1813G, and G1809R DICER1 syndrome mutants are investigated, which also cover all DICER1 hotspots.[25] MD simulations initiated from the catalytically–active metal configuration (aaRIII, $2Mg^{2+}$−B) suggest that none of these mutations cause large-scale changes to the Dicer active site. Indeed, the nucleophile positioning for the mutants is comparable to the wild-type enzyme (99–100% occupancy, r(O_{WAT}–P) = 3.4–3.5 Å). Furthermore, the same three Mg^{2+}−binding conformations observed for wild-type Dicer are sampled for the mutants, including the dominant catalytically active ($2Mg^{2+}$−B) conformation, and two inactive conformations with substrate unbinding with a metal (Mg_A^{2+}−U and Mg_B^{2+}−U); Figures E.10 and E.15, Appendix E). However, the relative stability of the active site conformations significantly differs from the wild-type model in a mutant-dependent manner (Table 6.1).

Table 6.1. Conformation occupancies for wild-type Dicer and DICER1-syndrome-causing mutants observed in MD simulations.[1]

Conformation[2]	Wild-Type[3]	D1709N	G1809R	E1813G	E1705K	D1810Y	E1813D
$2Mg^{2+}$−B	61−71%	6%	1%	<1%	28%	2%	44%
Mg_A^{2+}−U	29−34%	94%	99%	20%	52%	41%	21%
Mg_B^{2+}−U	0−5%	0%	0%	80%	20%	57%	35%

[1]Occupancies are a percentage of frames across all replicas with a specific Mg^{2+} ion coordination (Figures 6.7, 6.9, 6.10, and E.12).
[2]$2Mg^{2+}$−B represents the conformation with both metals bound to the substrate, while the substrate is unbound to Mg_A^{2+} (Mg_A^{2+}−U) or Mg_B^{2+} (Mg_B^{2+}−U) in the other conformations.
[3]Occupancies for wild-type Dicer include both water nucleophile and hydroxide nucleophile models.

Despite N1709 having the ability to adopt similar coordination to Mg_B^{2+} through a side chain oxygen and G1809 not coordinating to a metal in the wild-type enzyme, the D1709N and G1809R mutants change the overall charge of the active site. As a result, the mutations significantly change Dicer–Mg^{2+} coordination. Specifically, for both mutants, Mg_A^{2+} coordination to the substrate phosphate moiety is almost entirely absent throughout the simulation (94–99% Mg_A^{2+}–U, Figure 6.10A and B). The resulting active site conformation would render the nucleophilic attack required for catalysis infeasible. The significantly decreased occupancy of the catalytically competent

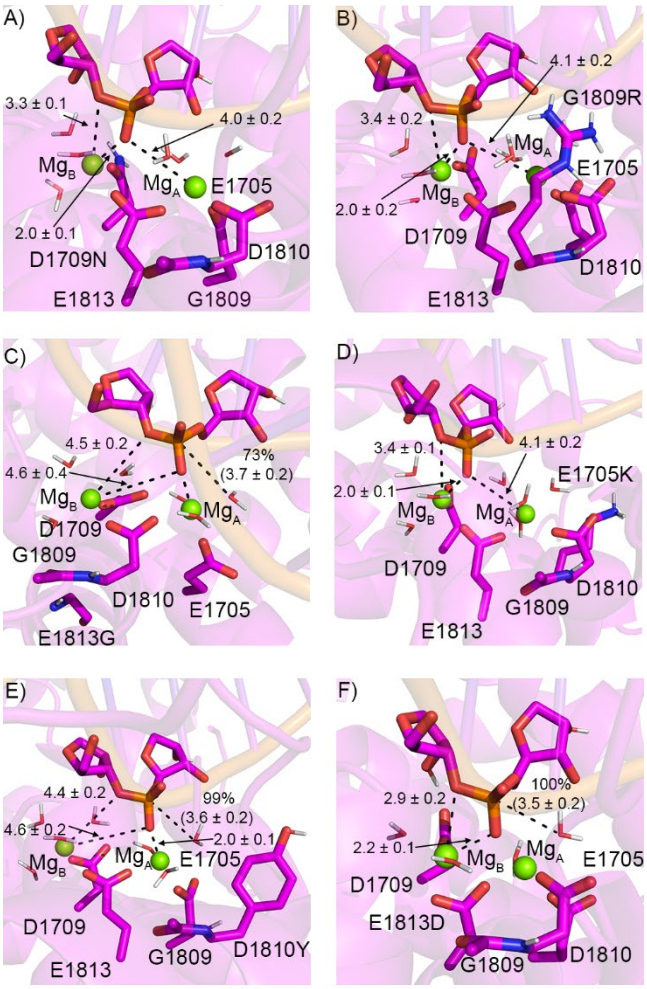


Figure 6.10. MD representative structures of dominant conformations for the A) D1709N, B) G1809R, C) E1813G, D) E1705K, E) D1810Y, and F) E1813D Dicer mutants. Distances are in Å, and occupancy of correctly positioned nucleophile is reported as a percentage.

$2Mg^{2+}$−B conformation compared to wild-type Dicer rationalizes the reduced activity for these DICER1 syndrome Dicer mutants.[131-133]

The E1813G, E1705K, and D1810Y mutants alter residues that our calculations highlight coordinate to Mg_A^{2+}. While the D1810Y mutant has similar nucleophile positioning in the dominant conformation as the wild-type enzyme (99% occupancy, $r(O_{WAT}–P)$ = 3.6 ± 0.2 Å), the E1813G and E1705K mutants lead to less favorable nucleophile positioning (73–86%, $r(O_{WAT}–P)$ = 3.4–3.5 Å), reflecting the impact on the Mg_A^{2+} environment. Indeed, the E1705K and D1810Y mutants result in a moderate increase in the Mg_A^{2+}−U conformation compared to wild-type Dicer (increased to 41–52%). All three mutants increase the occupancy of the Mg_B^{2+}−U conformation (~20–80%, Table 6.1 and Figure 6.10C–E), a conformation rarely sampled by wild-type Dicer (~5%, Figure 6.7), highlighting that both Mg_B^{2+} and Mg_A^{2+} positioning are impacted. The weakened Mg_B^{2+}–substrate coordination will lead to reduced catalytic activity as direct coordination between Mg_B^{2+} and the leaving group was shown to be vital for favorable phosphodiester bond cleavage. Coinciding with the increased population of the other conformations, there is a drastic reduction in the occupancy of the $2Mg^{2+}$−B conformation compared to wild type (~1–28% $2Mg^{2+}$−B occupancy, Table 6.1 and Figure 6.10C–E). These structural changes are likely the main causes of observed reduced RNA processing capabilities.[131-133]

As aspartate and glutamate have anionic side chains, E1813D is the only mutation considered that maintains the overall charge of the active site. As a result, E1813D is the only mutant that maintains the $2Mg^{2+}$−B conformation for a significant portion of the simulation (44%, Table 6.1 and Figure 6.10F). This starkly contrasts the E1813G mutant, which results in the lowest $2Mg^{2+}$−B conformation occupancy among all mutants investigated, despite impacting the same residue. Nevertheless, the coordination of Mg_B^{2+} to the substrate is weakened upon E1813D mutation, with the average Mg_B^{2+}–O3′ coordination distance increasing from 2.5 ± 0.2 Å for wild-type Dicer to 2.9 ± 0.2 Å (Figure 6.10F) due to the shorter aspartate side chain pulling Mg_B^{2+} away from the leaving group. Indeed, the occupancy of the $2Mg_B^{2+}$−B conformation decreases to 44% from over 61% and the occupancy of the Mg_B^{2+}−U conformation increases to 35% from less than 5% for wild-type Dicer,

highlighting that both Mg_B^{2+}−substrate and Mg_B^{2+}−Dicer interactions are impacted. Since our QM/MM calculations highlighted the requirement of sufficient leaving group stabilization for a viable catalytic mechanism, this weakened Mg_B^{2+}–O3′ interaction is likely primarily responsible for the observed negative impact of the E1813D mutation on Dicer activity,[131-133] although a slight reduction in Mg_A^{2+}−substrate binding is also observed across simulations upon mutation.

Taken together, our results provide critical insight into the molecular underpinnings of DICER1 syndrome. Indeed, the associated mutants all disrupt the active site Mg^{2+} ion binding configuration, particularly metal−substrate coordination that QM/MM calculations reveal are critical for catalysis. Nevertheless, the changes in metal binding configuration can take on different forms, namely disrupting Mg_A^{2+} coordination to the substrate (higher occupancy of Mg_A^{2+}−U conformation for D1709N and G1809R), breaking Mg_B^{2+} coordination to the substrate (higher occupancy of Mg_B^{2+}−U conformation for E1813G, E1705K, and D1810Y) or weakening Mg_B^{2+}–leaving group stabilization (longer Mg_B^{2+}−O3′ distance for E1813D). By affording the first structural descriptions of the active site for DICER1 syndrome mutants, our simulations explain why cells with DICER1 syndrome hotspot mutations have reduced miRNA biogenesis.[131-133] This information is invaluable for the rational design of DICER1 syndrome treatments that reactivate the enzyme.

6.4. Conclusions

The present work provides a structural description of the catalytically-conducive conformation of the Dicer RNase IIIb domain bound to the RNA substrate, including the first depiction of the catalytically-active Mg^{2+} binding configuration, and characterizes the previously unknown catalytic mechanism. In the functional wild-type Dicer conformation, both Mg^{2+} ions directly coordinate to the RNA substrate, with Mg_A^{2+} also coordinating to E1705, D1810, E1813, a water molecule and a hydroxide ion, and Mg_B^{2+} also coordinating to D1709, E1813, and two water molecules. This preferred metal binding architecture is distinct from the Mg^{2+} ion coordination geometry in the crystal structure of the homodimer in the absence of the substrate (PDB ID: 2EB1),[38] suggesting that Mg^{2+} ion migration upon substrate binding is critical for catalytic activity. The characterized

preferred catalytic mechanism for Dicer phosphodiester bond cleavage uses a hydroxide ion bound to Mg_A^{2+} as a nucleophile to attack the scissile phosphate, breaking the bond between the phosphate and the O3′ leaving group in two steps. The leaving group is stabilized by direct coordination to Mg_B^{2+}, while both Mg^{2+} ions and K1806 stabilize the charge build up on the phosphate moiety during the reaction. The calculated barrier for the reaction (73.4 kJ/mol) is consistent with experimental estimates (~80–100 kJ/mol),[115-117] and the proposed role for K1806 is consistent with the mutation of the analogous residue reducing mouse Dicer activity.[58] A similarly positioned active site lysine has been identified in other endonucleases,[38, 45, 51, 57-61] with this work suggesting they may have comparable functions in catalysis. However, more research is needed to determine if the substrate−stabilizing role observed for K1806 is conserved in the broader endonuclease family.

While six mutants were found to disrupt the Dicer active site in ways that would inhibit catalysis, the exact impact varies. Indeed, across the mutations, the mutant can disrupt Mg_A^{2+}–substrate coordination (D1709N and G1809R), Mg_B^{2+}–substrate coordination (E1813G, E1705K, and D1810Y), or weakened leaving group stabilization (E1813D). These active site changes would prevent catalysis by inhibiting leaving group departure or preventing attack of the nucleophile, explaining the reduced Dicer catalytic activity associated with DICER1 syndrome.[131-133] Overall, the characterization of the Dicer active site structure, catalytic mechanism, and mutant function represent important steps in the development of treatments for DICER1 syndrome and other Dicer-related disorders,[11, 19-24] and promises to accelerate the rational-design of new, powerful RNAi-based therapeutics with better efficacy[94-96] and the development of new RNAi based biotechnologies.[97-99]

6.5. References

(1) Iwakawa, H.-o.; Tomari, Y. Life of RISC: Formation, action, and degradation of RNA-induced silencing complex. *Mol. Cell* **2022**, *82* (1), 30-43. DOI: https://doi.org/10.1016/j.molcel.2021.11.026

(2) Hammond, S. M.; Bernstein, E.; Beach, D.; Hannon, G. J. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature* **2000**, *404* (6775), 293-296. DOI: 10.1038/35005107

(3) Yoshida, T.; Asano, Y.; Ui-Tei, K. Modulation of microRNA processing by Dicer via its associated dsRNA binding proteins. *Non-Coding RNA* **2021**, *7* (3). DOI: 10.3390/ncrna7030057

(4) Ha, M.; Kim, V. N. Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* **2014**, *15* (8), 509-524. DOI: 10.1038/nrm3838

(5) Zapletal, D.; Kubicek, K.; Svoboda, P.; Stefl, R. Dicer structure and function: Conserved and evolving features. *EMBO Rep.* **2023**, *24* (7), e57215. DOI: https://doi.org/10.15252/embr.202357215

(6) Ambros, V.; Bartel, B.; Bartel, D. P.; Burge, C. B.; Carrington, J. C.; Chen, X.; Dreyfuss, G.; Eddy, S. R.; Griffiths-Jones, S.; Marshall, M.; et al. A uniform system for microRNA annotation. *RNA* **2003**, *9* (3), 277-279. DOI: 10.1261/rna.2183803

(7) Kozomara, A.; Griffiths-Jones, S. miRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* **2014**, *42* (D1), D68-73. DOI: 10.1093/nar/gkt1181

(8) Dadhwal, G.; Samy, H.; Bouvette, J.; El-Azzouzi, F.; Dagenais, P.; Legault, P. Substrate promiscuity of Dicer toward precursors of the let-7 family and their 3'-end modifications. *Cell Mol Life Sci* **2024**, *81* (1), 53. DOI: 10.1007/s00018-023-05090-2

(9) Scarfiello, E.; Eichlinger, J.; Meister, G. The double-stranded microRNA precursor. *Postepy Biochem.* **2024**, *70* (1), 57-61.

(10) Le, C. T.; Nguyen, T. D.; Nguyen, T. A. Two-motif model illuminates dicer cleavage preferences. *Nucleic Acids Res.* **2024**, *52* (4), 1860-1877. DOI: 10.1093/nar/gkad1186 (acccessed 9/5/2024).

(11) Bartel, D. P. Metazoan microRNAs. *Cell* **2018**, *173* (1), 20-51. DOI: https://doi.org/10.1016/j.cell.2018.03.006

(12) Messina, S. The RAS oncogene in brain tumors and the involvement of let-7 microRNA. *Mol. Biol. Rep.* **2024**, *51* (1), 531. DOI: 10.1007/s11033-024-09439-z

(13) Kian, R.; Moradi, S.; Ghorbian, S. Role of components of microRNA machinery in carcinogenesis. *Exp Oncol* **2018**, *40* (1), 2-9.

(14) Smolarz, B.; Durczyński, A.; Romanowicz, H.; Szyłło, K.; Hogendorf, P. miRNAs in cancer (review of literature). *Int. J. Mol. Sci.* **2022**, *23* (5). DOI: 10.3390/ijms23052805

(15) Wang, Z.; Luo, X.; Lu, Y.; Yang, B. miRNAs at the heart of the matter. *J. Mol. Med.* **2008**, *86* (7), 771-783. DOI: 10.1007/s00109-008-0341-3

(16) Singh, R. P.; Massachi, I.; Manickavel, S.; Singh, S.; Rao, N. P.; Hasan, S.; Mc Curdy, D. K.; Sharma, S.; Wong, D.; Hahn, B. H.; et al. The role of miRNA in inflammation and autoimmunity. *Autoimmun. Rev.* **2013**, *12* (12), 1160-1165. DOI: https://doi.org/10.1016/j.autrev.2013.07.003

(17) Sedger, L. M. microRNA control of interferons and interferon induced anti-viral activity. *Mol. Immunol.* **2013**, *56* (4), 781-793. DOI: https://doi.org/10.1016/j.molimm.2013.07.009

(18) Ying, H.; Ebrahimi, M.; Keivan, M.; Khoshnam, S. E.; Salahi, S.; Farzaneh, M. miRNAs; A novel strategy for the treatment of COVID-19. *Cell Biol. Int.* **2021**, *45* (10), 2045-2053. DOI: https://doi.org/10.1002/cbin.11653

(19) Peng, Y.; Croce, C. M. The role of microRNAs in human cancer. *Signal Transduct Target Ther* **2016**, *1*, 15004. DOI: 10.1038/sigtrans.2015.4

(20) Chiosea, S.; Jelezcova, E.; Chandran, U.; Luo, J.; Mantha, G.; Sobol, R. W.; Dacic, S. Overexpression of Dicer in precursor lesions of lung adenocarcinoma. *Cancer Res.* **2007**, *67* (5), 2345-2350. DOI: 10.1158/0008-5472.Can-06-3533 (acccessed 6/26/2023).

(21) Foulkes, W. D.; Priest, J. R.; Duchaine, T. F. DICER1: Mutations, microRNAs and mechanisms. *Nat. Rev. Cancer* **2014**, *14* (10), 662-672. DOI: 10.1038/nrc3802

(22) Chen, P. S.; Lin, S. C.; Tsai, S. J. Complexity in regulating microRNA biogenesis in cancer. *Exp Biol Med (Maywood)* **2020**, *245* (5), 395-401. DOI: 10.1177/1535370220907314

(23) Dedes, K. J.; Natrajan, R.; Lambros, M. B.; Geyer, F. C.; Lopez-Garcia, M. A.; Savage, K.; Jones, R. L.; Reis-Filho, J. S. Down-regulation of the miRNA master regulators Drosha and Dicer is associated with specific subgroups of breast cancer. *Eur. J. Cancer* **2011**, *47* (1), 138-150. DOI: https://doi.org/10.1016/j.ejca.2010.08.007

(24) Klein, S.; Lee, H.; Ghahremani, S.; Kempert, P.; Ischander, M.; Teitell, M. A.; Nelson, S. F.; Martinez-Agosto, J. A. Expanding the phenotype of mutations in DICER1: Mosaic missense mutations in the RNase IIIb domain of DICER1 cause glow syndrome. *J Med Genet* **2014**, *51* (5), 294-302. DOI: 10.1136/jmedgenet-2013-101943

(25) de Kock, L.; Wu, M. K.; Foulkes, W. D. Ten years of DICER1 mutations: Provenance, distribution, and associated phenotypes. *Hum. Mutat.* **2019**, *40* (11), 1939-1953. DOI: 10.1002/humu.23877

(26) Robertson, J. C.; Jorcyk, C. L.; Oxford, J. T. DICER1 syndrome: DICER1 mutations in rare cancers. *Cancers* **2018**, *10* (5), 143/141-143/117. DOI: 10.3390/cancers10050143

(27) Rio Frio, T.; Bahubeshi, A.; Kanellopoulou, C.; Hamel, N.; Niedziela, M.; Sabbaghian, N.; Pouchet, C.; Gilbert, L.; O'Brien, P. K.; Serfas, K.; et al. DICER1 mutations in familial multinodular goiter with and without ovarian sertoli-leydig cell tumors. *JAMA* **2011**, *305* (1), 68-77. DOI: 10.1001/jama.2010.1910

(28) Doros, L. A.; Rossi, C. T.; Yang, J.; Field, A.; Williams, G. M.; Messinger, Y.; Cajaiba, M. M.; Perlman, E. J.; K, A. S.; Cathro, H. P.; et al. DICER1 mutations in childhood cystic nephroma and its relationship to DICER1-renal sarcoma. *Mod Pathol* **2014**, *27* (9), 1267-1280. DOI: 10.1038/modpathol.2013.242

(29) Slade, I.; Bacchelli, C.; Davies, H.; Murray, A.; Abbaszadeh, F.; Hanks, S.; Barfoot, R.; Burke, A.; Chisholm, J.; Hewitt, M.; et al. DICER1 syndrome: Clarifying the diagnosis, clinical features and management implications of a pleiotropic tumour predisposition syndrome. *J. Med. Genet.* **2011**, *48* (4), 273-278. DOI: 10.1136/jmg.2010.083790

(30) Schultz, K. A.; Yang, J.; Doros, L.; Williams, G. M.; Harris, A.; Stewart, D. R.; Messinger, Y.; Field, A.; Dehner, L. P.; Hill, D. A. DICER1-pleuropulmonary blastoma familial tumor predisposition syndrome: A unique constellation of neoplastic conditions. *Pathol Case Rev* **2014**, *19* (2), 90-100. DOI: 10.1097/pcr.0000000000000027

(31) Vedanayagam, J.; Chatila, W. K.; Aksoy, B. A.; Majumdar, S.; Skanderup, A. J.; Demir, E.; Schultz, N.; Sander, C.; Lai, E. C. Cancer-associated mutations in DICER1 RNase IIIa and IIIb domains exert similar effects on miRNA biogenesis. *Nat. Commun.* **2019**, *10* (1), 3682. DOI: 10.1038/s41467-019-11610-1

(32) Lee, Y.-Y.; Lee, H.; Kim, H.; Kim, V. N.; Roh, S.-H. Structure of the human DICER–pre-miRNA complex in a dicing state. *Nature* **2023**, *615* (7951), 331-338. DOI: 10.1038/s41586-023-05723-3

(33) Liu, Z.; Wang, J.; Cheng, H.; Ke, X.; Sun, L.; Zhang, Q. C.; Wang, H.-W. Cryo-EM structure of human Dicer and its complexes with a pre-miRNA substrate. *Cell* **2018**, *173* (5), 1191-1203.e1112. DOI: https://doi.org/10.1016/j.cell.2018.03.080

(34) Provost, P.; Dishart, D.; Doucet, J.; Frendewey, D.; Samuelsson, B.; Rådmark, O. Ribonuclease activity and RNA binding of recombinant human Dicer. *EMBO J* **2002**, *21* (21), 5864-5874. DOI: https://doi.org/10.1093/emboj/cdf578

(35) Murray, M. J.; Bailey, S.; Raby, K. L.; Saini, H. K.; de Kock, L.; Burke, G. A. A.; Foulkes, W. D.; Enright, A. J.; Coleman, N.; Tischkowitz, M. Serum levels of mature microRNAs in DICER1-mutated pleuropulmonary blastoma. *Oncogenesis* **2014**, *3* (2), e87-e87. DOI: 10.1038/oncsis.2014.1

(36) Chen, J.; Wang, Y.; McMonechy, M. K.; Anglesio, M. S.; Yang, W.; Senz, J.; Maines-Bandiera, S.; Rosner, J.; Trigo-Gonzalez, G.; Grace Cheng, S.; et al. Recurrent DICER1 hotspot mutations in endometrial tumours and their impact on microRNA biogenesis. *J. Pathol.* **2015**, *237* (2), 215-225. DOI: https://doi.org/10.1002/path.4569

(37) Jakymiw, A.; Patel, R. S.; Deming, N.; Bhattacharyya, I.; Shah, P.; Lamont, R. J.; Stewart, C. M.; Cohen, D. M.; Chan, E. K. Overexpression of Dicer as a result of reduced let-7 microrna levels contributes to increased cell proliferation of oral cancer cells. *Genes Chromosomes Cancer* **2010**, *49* (6), 549-559. DOI: 10.1002/gcc.20765

(38) Takeshita, D.; Zenno, S.; Lee, W. C.; Nagata, K.; Saigo, K.; Tanokura, M. Homodimeric structure and double-stranded RNA cleavage activity of the C-terminal RNase III domain of human Dicer. *J. Mol. Biol.* **2007**, *374* (1), 106-120. DOI: 10.1016/j.jmb.2007.08.069

(39) Yang, W. Nucleases: Diversity of structure, function and mechanism. *Q Rev Biophys* **2011**, *44* (1), 1-93. DOI: 10.1017/s0033583510000181

(40) Nikkel, D. J.; Kaur, R.; Wetmore, S. D. How can one metal power nucleic acid phosphodiester bond cleavage by a nuclease? Multiscale computational studies highlight a diverse mechanistic landscape. *J. Phys. Chem. B* **2025**, *129* (1), 3-18. DOI: 10.1021/acs.jpcb.4c05875

(41) De Vivo, M.; Dal Peraro, M.; Klein, M. L. Phosphodiester cleavage in Ribonuclease H occurs via an associative two-metal-aided catalytic mechanism. *J. Am. Chem. Soc.* **2008**, *130* (33), 10955-10962. DOI: 10.1021/ja8005786

(42) Frazão, C.; McVey, C. E.; Amblar, M.; Barbas, A.; Vonrhein, C.; Arraiano, C. M.; Carrondo, M. A. Unravelling the dynamics of RNA degradation by Ribonuclease II and its RNA-bound complex. *Nature* **2006**, *443* (7107), 110-114. DOI: 10.1038/nature05080

(43) Beese, L. S.; Steitz, T. A. Structural basis for the 3'-5' exonuclease activity of escherichia coli DNA polymerase I: A two metal ion mechanism. *EMBO J* **1991**, *10* (1), 25-33. DOI: 10.1002/j.1460-2075.1991.tb07917.x

(44) Yang, W.; Steitz, T. A. Recombining the structures of HIV integrase, RuvC and RNase H. *Structure* **1995**, *3* (2), 131-134. DOI: https://doi.org/10.1016/S0969-2126(01)00142-3

(45) Lee, J. Y.; Chang, J.; Joseph, N.; Ghirlando, R.; Rao, D. N.; Yang, W. MutH complexed with hemi- and unmethylated DNAs: Coupling base recognition and DNA cleavage. *Mol. Cell* **2005**, *20* (1), 155-166. DOI: https://doi.org/10.1016/j.molcel.2005.08.019

(46) Palermo, G.; Cavalli, A.; Klein, M. L.; Alfonso-Prieto, M.; Dal Peraro, M.; De Vivo, M. Catalytic metal ions and enzymatic processing of DNA and RNA. *Acc. Chem. Res.* **2015**, *48* (2), 220-228. DOI: 10.1021/ar500314j

(47) Rosta, E.; Nowotny, M.; Yang, W.; Hummer, G. Catalytic mechanism of RNA backbone cleavage by Ribonuclease H from quantum mechanics/molecular mechanics simulations. *J. Am. Chem. Soc.* **2011**, *133* (23), 8934-8941. DOI: 10.1021/ja200173a

(48) Yang, W.; Lee, J. Y.; Nowotny, M. Making and breaking nucleic acids: Two-Mg^{2+}-ion catalysis and substrate specificity. *Mol. Cell* **2006**, *22* (1), 5-13. DOI: https://doi.org/10.1016/j.molcel.2006.03.013

(49) Gan, J.; Shaw, G.; Tropea, J. E.; Waugh, D. S.; Court, D. L.; Ji, X. A stepwise model for double-stranded RNA processing by Ribonuclease III. *Mol. Microbiol.* **2008**, *67* (1), 143-154. DOI: 10.1111/j.1365-2958.2007.06032.x (acccessed 2020/02/20).

(50) Steitz, T. A.; Steitz, J. A. A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. U.S.A* **1993**, *90* (14), 6498-6502. DOI: 10.1073/pnas.90.14.6498

(51) Kaur, R.; Nikkel, D. J.; Wetmore, S. D. Mechanism of nucleic acid phosphodiester bond cleavage by human Endonuclease V: MD and QM/MM calculations reveal a versatile metal dependence. *J. Phys. Chem. B* **2024**, *128* (39), 9455-9469. DOI: 10.1021/acs.jpcb.4c05846

(52) Viadiu, H.; Aggarwal, A. K. The role of metals in catalysis by the restriction endonuclease Bam HI. *Nat. Struct. Mol. Biol.* **1998**, *5* (10), 910-916. DOI: 10.1038/2352

(53) Newman, M.; Lunnen, K.; Wilson, G.; Greci, J.; Schildkraut, I.; Phillips, S. E. V. Crystal structure of restriction endonuclease *Bgl*I bound to its interrupted DNA recognition sequence. *EMBO J* **1998**, *17* (18), 5466-5476-5476. DOI: https://doi.org/10.1093/emboj/17.18.5466 (acccessed 2025/08/18).

(54) Aboelnga, M. M.; Wetmore, S. D. Unveiling a single-metal-mediated phosphodiester bond cleavage mechanism for nucleic acids: A multiscale computational investigation of a human DNA repair enzyme. *J. Am. Chem. Soc.* **2019**, *141* (21), 8646-8656. DOI: 10.1021/jacs.9b03986

(55) Li, C. L.; Hor, L. I.; Chang, Z. F.; Tsai, L. C.; Yang, W. Z.; Yuan, H. S. DNA binding and cleavage by the periplasmic nuclease Vvn: A novel structure with a known active site. *EMBO J* **2003**, *22* (15), 4014-4025. DOI: https://doi.org/10.1093/emboj/cdg377

(56) Kaur, R.; Wetmore, S. D. Is metal stabilization of the leaving group required or can lysine facilitate phosphodiester bond cleavage in nucleic acids? A computational study of EndoV. *J. Chem Inf. Model.* **2024**, *64* (3), 944-959. DOI: 10.1021/acs.jcim.3c01775

(57) Horton, N. C.; Perona, J. J. DNA cleavage by EcoRV endonuclease: Two metal ions in three metal ion binding sites. *Biochemistry* **2004**, *43* (22), 6841-6857. DOI: 10.1021/bi0499056

(58) Du, Z.; Lee, J. K.; Tjhen, R.; Stroud, R. M.; James, T. L. Structural and biochemical insights into the dicing mechanism of mouse Dicer: A conserved lysine is critical for dsRNA cleavage. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (7), 2391-2396. DOI: 10.1073/pnas.0711506105

(59) Zhao, H.; Lin, Z.; Lynn, A. Y.; Varnado, B.; Beutler, J. A.; Murelli, R. P.; Le Grice, Stuart F. J.; Tang, L. Two distinct modes of metal ion binding in the nuclease active site of a viral DNA-packaging terminase: Insight into the two-metal-ion catalytic mechanism. *Nucleic Acids Res.* **2015**, *43* (22), 11003-11016. DOI: 10.1093/nar/gkv1018 (acccessed 2/25/2025).

(60) Wu, T.-h.; Loh, T.; Marinus, M. G. The function of Asp70, Glu77 and Lys79 in the escherichia coli MutH protein. *Nucleic Acids Res.* **2002**, *30* (3), 818-822. DOI: 10.1093/nar/30.3.818 (acccessed 2/25/2025).

(61) Selent, U.; Rueter, T.; Koehler, E.; Liedtke, M.; Thielking, V.; Alves, J.; Oelgeschlaeger, T.; Wolfes, H.; Peters, F.; Pingoud, A. A site-directed mutagenesis study to identify amino acid residues involved in the catalytic function of the restriction endonuclease EcoRV. *Biochemistry* **1992**, *31* (20), 4808-4815. DOI: 10.1021/bi00135a010

(62) Wu, J.; Samara, N. L.; Kuraoka, I.; Yang, W. Evolution of inosine-specific Endonuclease V from bacterial DNase to eukaryotic RNase. *Mol. Cell* **2019**, *76* (1), 44-56.e43. DOI: 10.1016/j.molcel.2019.06.046

(63) Blomberg, M. R.; Borowski, T.; Himo, F.; Liao, R.-Z.; Siegbahn, P. E. Quantum chemical studies of mechanisms for metalloenzymes. *Chem. Rev.* **2014**, *114* (7), 3601-3658. DOI: https://doi.org/10.1021/cr400388t

(64) Agbaglo, D. A.; Summers, T. J.; Cheng, Q.; DeYonker, N. J. The influence of model building schemes and molecular dynamics sampling on QM-cluster models: The chorismate mutase case study. *Phys. Chem. Chem. Phys.* **2024**, *26* (16), 12467-12482. DOI: 10.1039/D3CP06100K

(65) Sheng, X.; Himo, F. The quantum chemical cluster approach in biocatalysis. *Acc. Chem. Res.* **2023**, *56* (8), 938-947. DOI: 10.1021/acs.accounts.2c00795

(66) Elsässer, B.; Goettig, P. Mechanisms of proteolytic enzymes and their inhibition in QM/MM studies. *Int. J. Mol. Sci.* **2021**, *22* (6), 3232.

(67) Kaur, R.; Nikkel, D. J.; Wetmore, S. D. Computational studies of DNA repair: Insights into the function of monofunctional DNA glycosylases in the base excision repair pathway. *WIREs Comput. Mol. Sci.* **2020**, *10* (5), e1471. DOI: 10.1002/wcms.1471 (acccessed 2020/09/16).

(68) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Computational enzymatic catalysis – clarifying enzymatic mechanisms with the help of computers. *Phys. Chem. Chem. Phys.* **2012**, *14* (36), 12431-12441. DOI: 10.1039/C2CP41180F

(69) Gherib, R.; Dokainish, H. M.; Gauld, J. W. Multi-scale computational enzymology: Enhancing our understanding of enzymatic catalysis. *Int. J. Mol. Sci.* **2014**, *15* (1), 401-422. DOI: 10.3390/ijms15010401

(70) Casalino, L.; Nierzwicki, Ł.; Jinek, M.; Palermo, G. Catalytic mechanism of non-target DNA cleavage in CRISPR-Cas9 revealed by *ab initio* molecular dynamics. *ACS Catal.* **2020**, *10* (22), 13596-13605. DOI: 10.1021/acscatal.0c03566

(71) Alberto, M. E.; Pinto, G.; Russo, N.; Toscano, M. Triesterase and promiscuous diesterase activities of a Di-Co^{II}-containing organophosphate degrading enzyme reaction mechanisms. *Chem. Eur. J.* **2015**, *21* (9), 3736-3745. DOI: https://doi.org/10.1002/chem.201405593

(72) Araújo, A. R.; Ribeiro, A. J. M.; Fernandes, P. A.; Ramos, M. J. Catalytic mechanism of retroviral integrase for the strand transfer reaction explored by QM/MM calculations. *J. Chem. Theory Comput.* **2014**, *10* (12), 5458-5466. DOI: 10.1021/ct500570g

(73) Rahimian, M.; Yeole, S. D.; Gejji, S. P. Mechanistic insights for β-cyclodextrin catalyzed phosphodiester hydrolysis. *J. Mol. Model.* **2014**, *20* (4), 2198. DOI: 10.1007/s00894-014-2198-4

(74) Xia, H.; Zhang, W.; Yang, Y.; Zhang, W.; Purchase, D.; Zhao, C.; Song, X.; Wang, Y. Degradation mechanism of tris(2-chloroethyl) phosphate (TCEP) as an emerging contaminant in advanced oxidation processes: A DFT modelling approach. *Chemosphere* **2021**, *273*, 129674. DOI: https://doi.org/10.1016/j.chemosphere.2021.129674

(75) Kaur, R.; Frederickson, A.; Wetmore, S. D. Elucidation of the catalytic mechanism of a single-metal dependent homing endonuclease using QM and QM/MM approaches: The case study of I-*Ppo*I. *Phys. Chem. Chem. Phys.* **2024**, *26* (11), 8919-8931. DOI: 10.1039/D3CP06201E

(76) Kaur, R.; Nikkel, D. J.; Aboelnga, M. M.; Wetmore, S. D. The impact of DFT functional, cluster model size, and implicit solvation on the structural description of single-metal-mediated DNA phosphodiester bond cleavage: The case study of APE1. *J. Phys. Chem. B* **2022**, *126* (50), 10672-10683. DOI: 10.1021/acs.jpcb.2c06756

(77) Kaur, R.; Aboelnga, M. M.; Nikkel, D. J.; Wetmore, S. D. The metal dependence of single-metal mediated phosphodiester bond cleavage: A QM/MM study of a multifaceted human enzyme. *Phys. Chem. Chem. Phys.* **2022**, *24* (47), 29130-29140. DOI: 10.1039/D2CP04338F

(78) Elsässer, B.; Fels, G. Atomistic details of the associative phosphodiester cleavage in human ribonuclease H. *Phys. Chem. Chem. Phys.* **2010**, *12* (36), 11081-11088. DOI: 10.1039/C001097A

(79) Dürr, S. L.; Bohuszewicz, O.; Berta, D.; Suardiaz, R.; Jambrina, P. G.; Peter, C.; Shao, Y.; Rosta, E. The role of conserved residues in the DEDDh motif: The proton-transfer mechanism of HIV-1 RNase H. *ACS Catal.* **2021**, *11* (13), 7915-7927. DOI: 10.1021/acscatal.1c01493

(80) Palermo, G.; Stenta, M.; Cavalli, A.; Dal Peraro, M.; De Vivo, M. Molecular simulations highlight the role of metals in catalysis and inhibition of Type II Topoisomerase. *J. Chem. Theory Comput.* **2013**, *9* (2), 857-862. DOI: 10.1021/ct300691u

(81) Mones, L.; Kulhánek, P.; Florián, J.; Simon, I.; Fuxreiter, M. Probing the two-metal ion mechanism in the restriction endonuclease BamHI. *Biochemistry* **2007**, *46* (50), 14514-14523. DOI: 10.1021/bi701630s

(82) Sgrignani, J.; Magistrato, A. QM/MM MD simulations on the enzymatic pathway of the human flap endonuclease (hFEN1) elucidating common cleavage pathways to RNase H enzymes. *ACS Catal.* **2015**, *5* (6), 3864-3875. DOI: 10.1021/acscatal.5b00178

(83) Stevens, D. R.; Hammes-Schiffer, S. Exploring the role of the third active site metal ion in DNA polymerase η with QM/MM free energy simulations. *J. Am. Chem. Soc.* **2018**, *140* (28), 8965-8969. DOI: 10.1021/jacs.8b05177

(84) Ribeiro, A. J. M.; Ramos, M. J.; Fernandes, P. A. The catalytic mechanism of HIV-1 integrase for DNA 3′-end processing established by QM/MM calculations. *J. Am. Chem. Soc.* **2012**, *134* (32), 13436-13447. DOI: 10.1021/ja304601k

(85) Fothergill, M.; Goodman, M. F.; Petruska, J.; Warshel, A. Structure-energy analysis of the role of metal ions in phosphodiester bond hydrolysis by DNA Polymerase I. *J. Am. Chem. Soc.* **1995**, *117* (47), 11619-11627. DOI: 10.1021/ja00152a001

(86) Abyzov, A.; Uzun, A.; Strauss, P. R.; Ilyin, V. A. An AP endonuclease 1–DNA polymerase β complex: Theoretical prediction of interacting surfaces. *PLOS Comput. Biol.* **2008**, *4* (4), e1000066. DOI: 10.1371/journal.pcbi.1000066

(87) Krumkacheva, O. A.; Shevelev, G. Y.; Lomzov, A. A.; Dyrkheeva, N. S.; Kuzhelev, A. A.; Koval, V. V.; Tormyshev, V. M.; Polienko, Y. F.; Fedin, M. V.; Pyshnyi, D. V.; et al. DNA complexes with human apurinic/apyrimidinic endonuclease 1: Structural insights revealed by pulsed dipolar EPR with orthogonal spin labeling. *Nucleic Acids Res.* **2019**, *47* (15), 7767-7780. DOI: 10.1093/nar/gkz620 (acccessed 10/31/2024).

(88) Doss, C. G. P.; NagaSundaram, N. Investigating the structural impacts of I64T and P311S mutations in APE1-DNA complex: A molecular dynamics approach. *PLOS ONE* **2012**, *7* (2), e31677. DOI: 10.1371/journal.pone.0031677

(89) Guliaev, A. B.; Hang, B.; Singer, B. Structural insights by molecular dynamics simulations into specificity of the major human AP endonuclease toward the benzene-derived DNA adduct, pBQ-C. *Nucleic Acids Res.* **2004**, *32* (9), 2844-2852. DOI: 10.1093/nar/gkh594 (acccessed 10/31/2024).

(90) Oezguen, N.; Schein, C. H.; Peddi, S. R.; Power, T. D.; Izumi, T.; Braun, W. A “moving metal mechanism” for substrate cleavage by the DNA repair endonuclease APE-1. *Proteins* **2007**, *68* (1), 313-323. DOI: https://doi.org/10.1002/prot.21397

(91) Oezguen, N.; Mantha, A. K.; Izumi, T.; Schein, C. H.; Mitra, S.; Braun, W. MD simulation and experimental evidence for Mg^{2+} binding at the B site in human AP endonuclease 1. *Bioinformation* **2011**, *7* (4), 184-198. DOI: 10.6026/97320630007184

(92) Batebi, H.; Dragelj, J.; Imhof, P. Role of AP-endonuclease (Ape1) active site residues in stabilization of the reactant enzyme-DNA complex. *Proteins* **2018**, *86* (4), 439-453. DOI: https://doi.org/10.1002/prot.25460

(93) Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; et al. The ONIOM method and its applications. *Chem. Rev.* **2015**, *115* (12), 5678-5796. DOI: 10.1021/cr5004419

(94) Setten, R. L.; Rossi, J. J.; Han, S.-p. The current state and future directions of RNAi-based therapeutics. *Nat. Rev. Drug Discov.* **2019**, *18* (6), 421-446. DOI: 10.1038/s41573-019-0017-4

(95) Weng, Y.; Xiao, H.; Zhang, J.; Liang, X.-J.; Huang, Y. RNAi therapeutic and its innovative biotechnological evolution. *Biotechnol. Adv.* **2019**, *37* (5), 801-825. DOI: https://doi.org/10.1016/j.biotechadv.2019.04.012

(96) Merritt, W. M.; Bar-Eli, M.; Sood, A. K. The dicey role of Dicer: Implications for RNAi therapy. *Cancer Res.* **2010**, *70* (7), 2571-2574. DOI: 10.1158/0008-5472.Can-09-2536 (acccessed 12/10/2024).

(97) Casacuberta, J. M.; Devos, Y.; du Jardin, P.; Ramon, M.; Vaucheret, H.; Nogué, F. Biotechnological uses of RNAi in plants: Risk assessment considerations. *Trends Biotechnol.* **2015**, *33* (3), 145-147. DOI: https://doi.org/10.1016/j.tibtech.2014.12.003

(98) Moreira-Pinto, C. E.; Coelho, R. R.; Leite, A. G. B.; Silveira, D. A.; de Souza, D. A.; Lopes, R. B.; Macedo, L. L. P.; Silva, M. C. M.; Ribeiro, T. P.; Morgante, C. V.; et al. Increasing anthonomus grandis susceptibility to metarhizium anisopliae through RNAi-induced agrarelish knockdown: A perspective to combine biocontrol and biotechnology. *Pest Manag. Sci.* **2021**, *77* (9), 4054-4063. DOI: https://doi.org/10.1002/ps.6430

(99) Sherman, J. H.; Munyikwa, T.; Chan, S. Y.; Petrick, J. S.; Witwer, K. W.; Choudhuri, S. RNAi technologies in agricultural biotechnology: The toxicology forum 40th annual summer meeting. *Regul. Toxicol. Pharm.* **2015**, *73* (2), 671-680. DOI: https://doi.org/10.1016/j.yrtph.2015.09.001

(100) Schrodinger, LLC. The PyMOL molecular graphics system, version 1.8. 2015.

(101) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A P.; Rempfer, C.; Bordoli, L.; et al. Swiss-model: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46* (W1), W296-W303. DOI: 10.1093/nar/gky427 (acccessed 8/9/2023).

(102) 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* (W1), W537-W541. DOI: 10.1093/nar/gks375 (acccessed 8/9/2023).

(103) Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A. H++: A server for estimating p ka s and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* **2005**, *33* (suppl_2), W368-W371. DOI: 10.1093/nar/gki464

(104) Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; III, T. E. C.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; et al. AMBER 2018. **2018**, University of California, San Fransisco.

(105) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696-3713. DOI: 10.1021/acs.jctc.5b00255

(106) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers. *Biophys. J.* **2007**, *92* (11), 3817-3829. DOI: https://doi.org/10.1529/biophysj.106.097782

(107) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. E.; Jurečka, P. Refinement of the cornell et al. Nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.* **2011**, *7* (9), 2886-2902. DOI: 10.1021/ct200162x

(108) Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. The R.E.D. Tools: Advances in RESP and ESP charge derivation and force field library building. *Phys. Chem. Chem. Phys.* **2010**, *12* (28), 7821-7839. DOI: 10.1039/c0cp00111b

(109) Vreven, T.; Morokuma, K. The ONIOM (our own N-layered integrated molecular orbital + molecular mechanics) method for the first singlet excited (s1) state photoisomerization

path of a retinal protonated schiff base. *J. Chem. Phys.* **2000**, *113* (8), 2969-2975. DOI: 10.1063/1.1287059

(110) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120* (1), 215-241. DOI: 10.1007/s00214-007-0310-x

(111) Pilbák, S.; Farkas, Ö.; Poppe, L. Mechanism of the tyrosine ammonia lyase reaction—tandem nucleophilic and electrophilic enhancement by a proton transfer. *Chem. Eur. J.* **2012**, *18* (25), 7793-7802. DOI: https://doi.org/10.1002/chem.201103662

(112) Hu, X.; Hu, H.; Melvin, J. A.; Clancy, K. W.; McCafferty, D. G.; Yang, W. Autocatalytic intramolecular isopeptide bond formation in gram-positive bacterial pili: A QM/MM simulation. *J. Am. Chem. Soc.* **2011**, *133* (3), 478-485. DOI: 10.1021/ja107513t

(113) Wu, S.; Xu, D.; Guo, H. QM/MM studies of monozinc β-lactamase CphA suggest that the crystal structure of an enzyme−intermediate complex represents a minor pathway. *J. Am. Chem. Soc.* **2010**, *132* (51), 17986-17988. DOI: 10.1021/ja104241g

(114) Nikkel, D. J.; Wetmore, S. D. Distinctive formation of a DNA–protein cross-link during the repair of DNA oxidative damage: Insights into human disease from MD simulations and QM/MM calculations. *J. Am. Chem. Soc.* **2023**, *145* (24), 13114-13125. DOI: 10.1021/jacs.3c01773

(115) Ma, E.; MacRae, I. J.; Kirsch, J. F.; Doudna, J. A. Autoinhibition of human Dicer by its internal helicase domain. *J. Mol. Biol.* **2008**, *380* (1), 237-243. DOI: https://doi.org/10.1016/j.jmb.2008.05.005

(116) Chakravarthy, S.; Sternberg, S. H.; Kellenberger, C. A.; Doudna, J. A. Substrate-specific kinetics of Dicer-catalyzed RNA processing. *J. Mol. Biol.* **2010**, *404* (3), 392-402. DOI: https://doi.org/10.1016/j.jmb.2010.09.030

(117) Bouvette, J.; Korkut, D. N.; Fouillen, A.; Amellah, S.; Nanci, A.; Durocher, Y.; Omichinski, J. G.; Legault, P. High-yield production of human Dicer by transfection of human HEK293-EBNA1 cells grown in suspension. *BMC Biotechnology* **2018**, *18* (1), 76. DOI: 10.1186/s12896-018-0485-3

(118) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; et al. *Gaussian 16 rev. B.01*; Gaussian, Inc., 2016.

(119) Drusin, S. I.; Rasia, R. M.; Moreno, D. M. Study of the role of Mg^{2+} in dsRNA processing mechanism by bacterial RNase III through QM/MM simulations. *J. Biol. Inorg. Chem.* **2019**. DOI: 10.1007/s00775-019-01741-7

(120) Jackson, C.; Kim, H.-K.; Carr, P. D.; Liu, J.-W.; Ollis, D. L. The structure of an enzyme–product complex reveals the critical role of a terminal hydroxide nucleophile in the bacterial phosphotriesterase mechanism. *Biochim. Biophys. Acta - Proteins Proteom.* **2005**, *1752* (1), 56-64. DOI: https://doi.org/10.1016/j.bbapap.2005.06.008

(121) Morris, E. R.; Caswell, S. J.; Kunzelmann, S.; Arnold, L. H.; Purkiss, A. G.; Kelly, G.; Taylor, I. A. Crystal structures of SAMHD1 inhibitor complexes reveal the mechanism of water-mediated dNTP hydrolysis. *Nat. Commun.* **2020**, *11* (1), 3165. DOI: 10.1038/s41467-020-16983-2

(122) Hong, M.-K.; Ribeiro, A. J. M.; Kim, J.-K.; Ngo, H.-P.-T.; Kim, J.; Lee, C. H.; Ahn, Y.-J.; Fernandes, P. A.; Li, Q.; Ramos, M. J.; et al. Divalent metal ion-based catalytic mechanism of the nudix hydrolase Orf153 (YmfB) from escherichia coli. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2014**, *70* (5), 1297-1310. DOI: doi:10.1107/S1399004714002570

(123) Carvalho, A. T. P.; Fernandes, P. A.; Ramos, M. J. The catalytic mechanism of RNA Polymerase II. *J. Chem. Theory Comput.* **2011**, *7* (4), 1177-1188. DOI: 10.1021/ct100579w

(124) Gutteridge, A.; Thornton, J. Conformational changes observed in enzyme crystal structures upon substrate binding. *J. Mol. Biol.* **2005**, *346* (1), 21-28. DOI: https://doi.org/10.1016/j.jmb.2004.11.013

(125) Newman, M.; Murray-Rust, J.; Lally, J.; Rudolf, J.; Fadden, A.; Knowles, P. P.; White, M. F.; McDonald, N. Q. Structure of an XPF endonuclease with and without DNA suggests a model for substrate recognition. *EMBO J* **2005**, *24* (5), 895-905. DOI: https://doi.org/10.1038/sj.emboj.7600581

(126) Kim, C.-Y.; Park, M. S.; Dyer, R. B. Human flap endonuclease-1: Conformational change upon binding to the flap DNA substrate and location of the Mg^{2+} binding site. *Biochemistry* **2001**, *40* (10), 3208-3214. DOI: 10.1021/bi002100n

(127) Marsden, S. R.; Wijma, H. J.; Mohr, M. K. F.; Justo, I.; Hagedoorn, P.-L.; Laustsen, J.; Jeffries, C. M.; Svergun, D.; Mestrom, L.; McMillan, D. G. G.; et al. Substrate induced movement of the metal cofactor between active and resting state. *Angew. Chem. Int. Ed.* **2022**, *61* (49), e202213338. DOI: https://doi.org/10.1002/anie.202213338

(128) Dollins, D. E.; Xiong, J.-P.; Endo-Streeter, S.; Anderson, D. E.; Bansal, V. S.; Ponder, J. W.; Ren, Y.; York, J. D. A structural basis for lithium and substrate binding of an inositide phosphatase. *J. Biol. Chem.* **2021**, *296*. DOI: 10.1074/jbc.RA120.014057 (acccessed 2025/10/03).

(129) Mullen, G. P.; Serpersu, E. H.; Ferrin, L. J.; Loeb, L. A.; Mildvan, A. S. Metal binding to DNA polymerase I, its large fragment, and two 3‘,5‘-exonuclease mutants of the large fragment. *J. Biol. Chem.* **1990**, *265* (24), 14327-14334. DOI: https://doi.org/10.1016/S0021-9258(18)77304-3

(130) He, H.; Chen, Q.; Georgiadis, M. M. High-resolution crystal structures reveal plasticity in the metal binding site of apurinic/apyrimidinic Endonuclease I. *Biochemistry* **2014**, *53* (41), 6520-6529. DOI: 10.1021/bi500676p

(131) Rakheja, D.; Chen, K. S.; Liu, Y.; Shukla, A. A.; Schmid, V.; Chang, T.-C.; Khokhar, S.; Wickiser, J. E.; Karandikar, N. J.; Malter, J. S.; et al. Somatic mutations in DROSHA and DICER1 impair microRNA biogenesis through distinct mechanisms in Wilms tumours. *Nat. Commun.* **2014**, *5* (1), 4802. DOI: 10.1038/ncomms5802

(132) Fernández-Martínez, L.; Villegas, J. A.; Santamaría, Í.; Pitiot, A. S.; Alvarado, M. G.; Fernández, S.; Torres, H.; Paredes, Á.; Blay, P.; Balbín, M. Identification of somatic and germ-line DICER1 mutations in pleuropulmonary blastoma, cystic nephroma and rhabdomyosarcoma tumors within a DICER1 syndrome pedigree. *BMC Cancer* **2017**, *17* (1), 146. DOI: 10.1186/s12885-017-3136-5

(133) Apellaniz-Ruiz, M.; Segni, M.; Kettwig, M.; Gluer, S.; Pelletier, D.; Nguyen, V.-H.; Wagener, R.; Lopez, C.; Muchantef, K.; Soglio, D. B.-D.; et al. Mesenchymal hamartoma of the liver and DICER1 syndrome. *N. Engl. J. Med.* **2019**, *380* (19), 1834-1842. DOI: 10.1056/nejmoa1812169

Chapter 7: Conclusions and Future Directions

7.1. Summary

The cleavage of nucleic acid glycosidic and phosphodiester bonds is required for a variety of processes necessary for cell survival, including DNA repair[1] and RNA processing.[2-4] These bond cleavages are well accepted to be catalyzed by glycosylases and nucleases. However, the mechanisms behind the reactions are often unclear. This thesis utilized a collection of computational techniques based on classical and quantum mechanics to obtain insights into protein−substrate structures and characterize the mechanism of action for different glycosylases and nucleases. Specifically, the monofunctional glycosylases MutY, MBD4, and AlkA were investigated to clarify controversies regarding the formation of a DNA−protein crosslink during catalysis (Figure 7.1). A MAP-related MutY mutant was also studied to gain insight into the

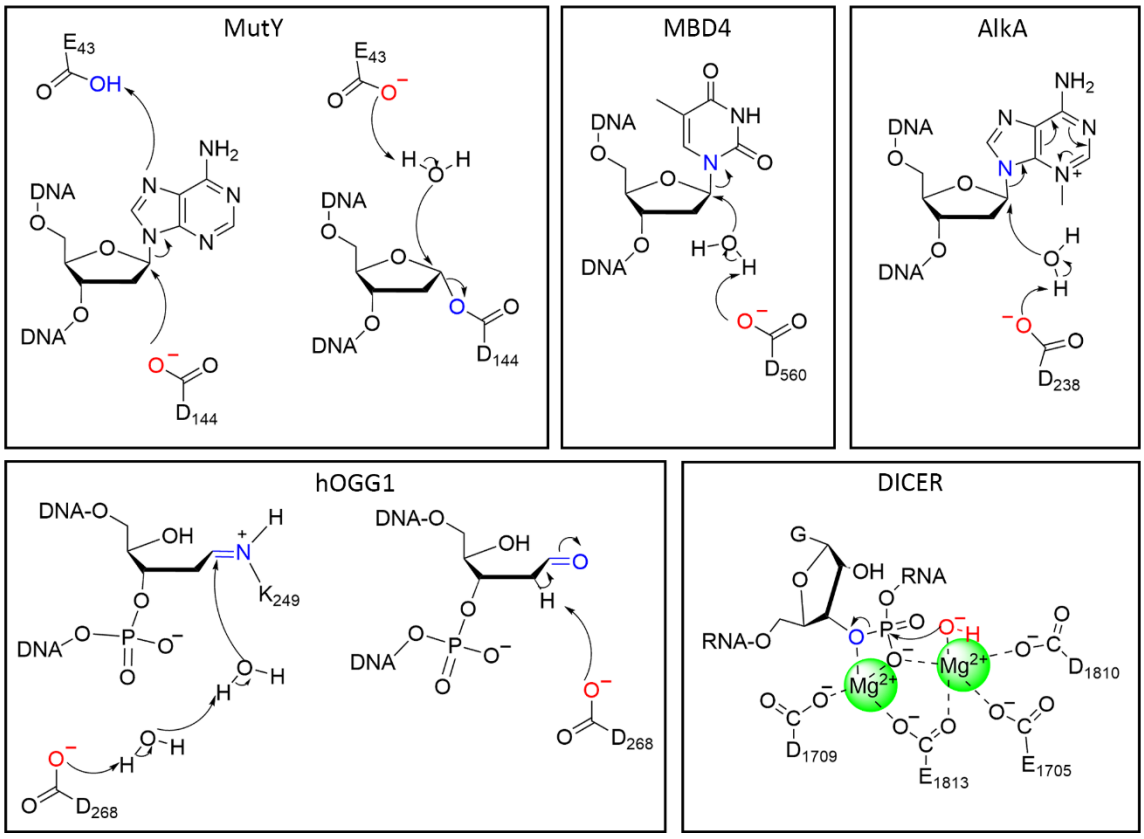


Figure 7.1. Summary of the key aspects of the proposed mechanisms for the five enzymes investigated in this thesis.

molecular underpinnings of this human disorder. Additionally, this thesis characterized the β-lyase mechanism of hOGG1, a first for the bifunctional glycosylase family. Finally, this thesis uncovered the atomic level details of the two-metal-mediated phosphodiester bond cleavage facilitated by Dicer, which also provided insight into the etiology of DICER1 syndrome-causing mutants. This chapter provides a summary of the findings on each enzyme and highlights possible future avenues of research.

7.2. Contributions From Thesis

In Chapter 2, MD simulations and QM/MM techniques initiated from DNA–protein complexes representing different stages of the repair pathway were used to map the catalytic mechanism of wild-type MutY. This multi-pronged computational approach characterized a DNA–protein crosslinking mechanism that is consistent with all previous experimental data[5-11] and represents a distinct pathway across the broad class of monofunctional glycosylase repair enzymes. In addition to clarifying how the crosslink is stabilized by the enzyme and hydrolyzed to release an abasic site, my calculations rationalize why crosslink formation is favored over immediate glycosidic bond hydrolysis, the accepted mechanism for all other monofunctional DNA glycosylases to date.[12-16] Investigation of Y126F and N146S MutY uncovered the role of these residues in catalysis and rationalized the connection between analogous MUTYH mutants and MAP.[17] In addition to advancing our knowledge of the chemistry associated with a devastating disorder, the structural information gained about the distinctive MutY mechanism compared to other repair enzymes represents an important step for the development of specific and potent small molecule inhibitors as cancer therapeutics.

While MutY was the first glycosylase discovered to form a DNA−protein crosslink as part of catalysis, other enzymes were previously hypothesized to utilize a similar mechanism. In Chapter 3, the possibility of a MBD4 forming a crosslink was explored, as evidenced by a close contact in a mouse MBD4 crystal structure.[18] MBD4 is responsible for removing T mispaired with G in CpG motifs,[18-20] resulting in MBD4 being implicated in MBD4-associated neoplasia syndrome[21, 22] and

cancer resistance to 5-fluorouracil treatment.[23, 24] To provide atomic level structural details of the active site conformation and establish the mechanistic pathway, this chapter used a combination of abMD simulations and QM/MM calculations to map the MBD4 catalytic mechanism. Although my data indicated that the catalytic D560 residue is flexible in the active site, only one conformation facilitates 5mC excision. In contrast to evidence for the formation of a DNA−protein crosslinked intermediate,[18] my modeling suggested catalysis is only viable through a deglycosylation mechanism that involves a D560-activated water nucleophile attacking C1′ of T, with a network of hydrogen bonds stabilizing the transition state and facilitating nucleobase departure. This proposal is fully consistent with experimental mutagenic,[18] kinetic,[25] crystallographic,[25] and stereoscopic [25] data, and aligns the MBD4 catalytic pathway with UDG[26] and several other monofunctional DNA glycosylases.[12-16] By furthering our knowledge of MBD4 catalysis, this work aids in the future development of treatments for MBD4-related genetic disorders and the rational design of transition state mimic inhibitors to enhance existing cancer therapies.

AlkA is a monofunctional glycosylase responsible for initiating the repair of many alkylated bases as part of the BER pathway in bacterial cells.[27, 28] Like MBD4, AlkA was previously hypothesized to form a DNA−protein crosslinked intermediate in the catalytic pathway based on a close contact between D238 and a transition state analogue in a crystal structure.[28] In Chapter 4, a model of AlkA bound to 3mA was constructed as no structure of the enzyme−substrate complex was known. Subsequently, combined MD and QM/MM approaches investigated AlkA interactions with 3mA, and mapped multiple catalytic pathways for 3mA excision. MD simulations revealed that eight residues comprise the AlkA active site, binding the substrate through non-specific, hydrophobic contacts. Substrate promiscuity is likely further promoted by Y273, which can change conformation to modulate the binding pocket size. Despite the previous proposal of DNA−protein crosslink formation,[28] QM/MM calculations suggested that catalysis proceeds through a direct hydrolysis mechanism in which D238 activates a water nucleophile that attacks at C1′, consistent with experimental kinetic,[29] mutagenic,[30, 31] and structural data.[28] Proper D238 positioning with respect to the substrate is supported by a network of interactions between the substrate, W218,

W272, and D238. The insight into AlkA substrate binding and catalysis gained from this study can push forward the development of small molecule inhibitors as antibacterial agents.

Together, the details of the catalytic mechanism uncovered for MutY, MBD4, and AlkA allow for insights into the function of monofunctional glycosylases, specifically regarding the formation of a DNA−protein crosslink. While all three enzymes have been hypothesized to use a crosslinking mechanism of action, MutY remains unique among the monofunctional glycosylases, using the direct attack of D144 at C1′ to displace the damaged nucleobase rather than activating a water nucleophile, which is the preferred mechanism of the other two enzymes. Elucidating the mechanisms of MBD4 and AlkA revealed how common active site features of glycosylases can impede crosslink formation. For MBD4, the crosslink was thermodynamically unstable, with the anionic thymine leaving group being a potent nucleophile for potential reattack at C1′. Conversely, in AlkA, crosslink formation was prevented through numerous hydrogen-bonding interactions with the catalytic D238, which restricted the residue from approaching C1′ and resulted in crosslink formation being kinetically infeasible. Additionally, neither the AlkA nor MBD4 active site contained a viable base that could activate a water molecule to hydrolyze the crosslink and eventually regenerate the enzyme. Meanwhile, MutY had the perfect combination of features to make crosslink formation possible. Since adenine is protonated prior to glycosidic bond cleavage, the leaving group is neutral and not a strong nucleophile, making reattack at C1′ by the departed A unlikely. Additionally, the residue (E43) that protonates adenine is perfectly positioned to activate a water nucleophile for enzyme regeneration. Finally, D144 is positioned in the active site through only a single hydrogen bond with N146, allowing adequate flexibility for nucleophilic attack. With all of these features needing to occur simultaneously, it is unsurprising that MutY is unique among the DNA monofunctional glycosylases in forming the crosslinked intermediate. The discoveries in this thesis will aid in the development for inhibitors of these enzymes as anticancer or antibacterial drugs and allow for the prediction of a monofunctional glycosylase mechanism based on active site features.

While MutY is unique among the monofunctional glycosylases in forming a DNA−protein crosslink, crosslink formation is ubiquitous among the bifunctional glycosylase family.[32-36] However, relatively little was known about the β-lyase step during bifunctional glycosylase activity, including for the enzyme that initiates the repair of the major DNA damage product (8oG) in humans, hOGG1.[37] In Chapter 5, a combination of classical MD and QM/MM MD simulations were utilized to identify the preferred mechanism of action for the rate-limiting β-lyase step of hOGG1 activity. MD simulations revealed that the cleaved 8oG glycosylation product rapidly leaves the active site, precluding product-assisted elimination and supporting the allosteric nature of 8oG activators.[38] Instead, lyase activity was discovered to occur in two steps, namely the hydrolysis of a cationic imine crosslink followed by the D268-catalyzed elimination of the phosphate group. K249 was identified as an alternative base for elimination, rationalizing how the enzyme can accommodate D268 mutation.[39] This represents the first ever characterized β-lyase mechanism for a bifunctional glycosylase, which unifies the literature surrounding hOGG1 lyase activity. The structural details uncovered about the mechanism can push forward the development of small molecule inhibitors and activators as disease therapeutics.

Alongside bifunctional glycosylases, nucleases catalyze the cleavage of nucleic acid backbones and play vital roles in gene regulation[40] and viral defense.[41, 42] Specifically, RNAi involves several phosphodiester bond cleavage steps and has been exploited in therapeutic solutions to selectively prevent protein translation,[43] allowing inhibition of otherwise undruggable targets. Dicer plays an integral role in RNAi by generating cleaved RNA products that are incorporated into the RISC,[2-4] with Dicer misfunction being associated with a variety of disorders[44-50] such as DICER1 syndrome.[51, 52] In Chapter 6, a combination of MD simulations and QM/MM techniques were used to determine the binding pattern of two Mg^{2+} ions known to be critical for Dicer activity,[53, 54] characterized the Dicer catalytic mechanism for phosphodiester bond cleavage, and explored the misfunction of Dicer mutants associated with DICER1 syndrome.[51, 52] My data suggested catalysis is only feasible when a hydroxide ion nucleophile is bound to an active site Mg^{2+} ion and both active site Mg^{2+} ions are directly coordinated to the RNA substrate. Phosphodiester bond hydrolysis proceeded through a two-step mechanism with a phosphorane

intermediate, which is stabilized by direct Mg^{2+}–substrate coordination and a hydrogen bond with K1806. MD simulations suggested that six known DICER1 syndrome causing mutants[51, 52] impede catalysis by inducing unique active site disruptions that inhibit Mg^{2+} ion coordination to the substrate. Considering that RNAi-based therapeutics require processing by Dicer to be active and often contain significant nucleotide modifications, the knowledge of Dicer catalysis unveiled in this work will allow for the rational design of modified siRNA drugs that remain Dicer compatible. Additionally, this work uncovered the molecular underpinnings of DICER1 syndrome which opens the door for further research on this disease.

This thesis provides insight into the different ways enzymes facilitate phosphodiester bond cleavage and how attempting to predict catalytic mechanisms solely based on features of crystal structures can lead to erroneous results. My work on hOGG1 revealed the first characterized β-lyase mechanism for a bifunctional glycosylase. This mechanism conflicted with the previously assumed product-assisted mechanism based on a crystal structure that showed the cleaved substrate retained in the enzyme active site. On the other hand, this thesis identified a traditional two-metal-mediated mechanism for Dicer despite crystal structures showing different metal arrangements. Together, this thesis further emphasizes that elucidating enzyme mechanisms requires the constructive interplay between experimental and computational techniques. Details of future research projects that could provide insight into the function of other members of the glycosylase and endonuclease families are provided in the next section.

7.3. Future Directions

The work done in this thesis has provided critical insights into the function of different glycosylases and nucleases. However, the mechanisms of many enzymes in these classes remain unexplored and opportunities for experimental and computational studies to corroborate or expand on these findings are present. Namely, the work done in Chapter 4 of this thesis has conclusively identified the catalytic mechanism of AlkA and uncovered important active site features for binding 3mA, but questions remain regarding how AlkA accommodates pyrimidine and neutral substrates.

In Chapter 4, it is hypothesized that the flexibility of Y272 may modify binding pocket size to match the substrate being bound and R22 may play a bigger role in the catalysis of neutral substrates than seen for 3mA. To test these hypotheses, MD simulations on AlkA bound to substrates of different sizes could be performed, with the previously constructed AlkA model used as a starting point. 1N^6-ethenoadenine and O2mC would be ideal test cases,[55] representing the extremes of AlkA substrate accommodation with a large three-ringed and a small single-ringed system. Observations on how Y272 interacts with each substrate would reveal whether this residue is responsible for the wide range of AlkA substrate sizes. Additionally, since both of these bases are neutral, simulations would be able to identify whether AlkA behaves differently when binding neutral and cationic substrate (e.g., determine whether R22 is more involved in neutral substrate binding compared to cationic substrate binding). The computational work could be combined with crystallographic or cryo-EM structures of AlkA bound to different substrates to obtain a holistic understanding of AlkA substrate binding. Mutagenic studies to examine the activity of AlkA in the absence of Y272 and R22 and with various substrates could also provide insight into how these residues impact AlkA substrate promiscuity. Overall, these proposed studies would comprehensively identify the features that allow AlkA to target alkylation damage of various sizes, both neutral and cationic, improving our fundamental understanding of AlkA function as well as identifying features that could allow controlled substrate promiscuity in the development of artificial enzymes.

The methodology used to study monofunctional glycosylases in Chapters 2, 3, and 4 of this thesis can be applied to other glycosylases where the mechanism of action is unknown. One glycosylase of particular interest is single-strand selective monofunctional uracil DNA glycosylase (SMUG1). SMUG1 is responsible for removing uracil and uracil oxidized at the C5 position (e.g. 5-hydroxymethyluracil, 5-hydroxyuracil, 5-formyluracil, and 5fU) from double-stranded and single-stranded DNA.[56] Indeed, SMUG1 is considered the primary enzyme responsible for 5fU removal *in vivo*,[57] which indicates that inhibition of this enzyme could increase the efficacy of 5fU cancer treatments. SMUG1 substrate binding is facilitated by F98 and N163 which form specific interactions with the substrate nucleobase.[58] Moreover, mutagenic studies show N85 and H239

play an important role in catalysis (**Figure 7.2A**).[58, 59] Despite this, the mechanism of action for SMUG1 is unknown. While crystal structures of *Xenopus laevis* SMUG1 exist (PDB ID: 1OE6, **Figure 7.2A**),[56] there is no crystal structure of human SMUG1 or SMUG1 bound to a DNA substrate. As such, homology modeling combined with methodology employed for Dicer in Chapter 6 would be needed to obtain an initial model for the SMUG1−substrate complex. Subsequent MD and QM/MM calculations would identify a catalytic mechanism consistent with available experimental data for the removal of uracil analogues from DNA.

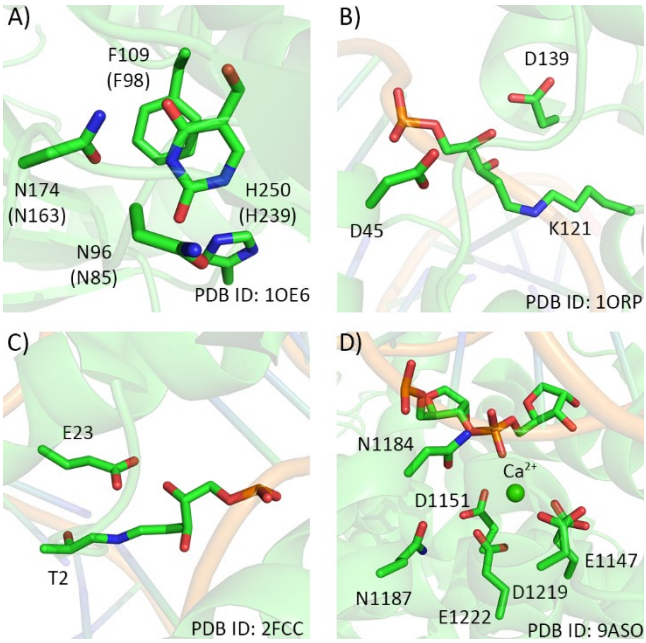


Figure 7.2. Crystal structure highlighting the active site of A) *Xenopus laevis* SMUG1 bound to 5-hydroxyuracil (analogous human residues in parenthesis), and B) Nth and C) T4 PDG borohydride-trapped intermediates, highlighting a DNA−protein crosslink. D) Cryo-EM structure of Ca^{2+}-inhibited Drosha bound to substrate pri-miRNA.

In Chapter 5, it was revealed that product-assisted elimination was not a viable pathway for the hOGG1 catalyzed β-lyase step and a mechanism for hOGG1 was characterized where D268 acts as a base for the rate-limiting elimination. K249 was also hypothesized to act as the base for elimination in the case of the D268Q hOGG1 mutant to account for the negligible change in catalytic activity.[39] While QM/MM MD simulations showed that K249 is a strong enough base to catalyze the reaction, further MD simulations investigating K249 as the general base in the D268Q hOGG1

mutant could confirm whether K249 is in position to abstract the C2′ proton or if the change in active site results in other residues repositioning to afford this role. Follow up QM/MM or QM/MM MD calculations based on the MD results could compare different possible pathways to uncover the atomic level details of how hOGG1 retains activity upon D268 mutation. This would underscore that hOGG1 does not require a cofactor in the active site for β-lyase activity and set the stage for studies to identify the allosteric binding site of 8oG in hOGG1. Molecular docking simulations would be able to identify a constellation of potential allosteric binding locations. This could be confirmed through cross-correlation network analysis[60] and principal component analysis[61] on long-timescale MD simulations with 8oG bound in the site to identify the interactions and dynamic motions in the enzyme that could transmit binding information from the allosteric site to the active site. Identifying the allosteric binding site of hOGG1 would have massive implication in the design of hOGG1 activators as therapeutics for oxidative stress.[62-64]

The work in Chapter 5 represents the first time the β-lyase step of a bifunctional DNA glycosylase was characterized. However, it is still unclear whether this mechanism is generalizable to other bifunctional glycosylases that form similar imine crosslinks or whether this mechanism is specific to hOGG1. Nth is a bifunctional glycosylase responsible for repairing oxidative pyrimidine damage in bacterial cells.[65] Both Nth and hOGG1 form an imine crosslink with a lysine side chain as part of their catalytic mechanisms (K121 for Nth).[66] Additionally, Nth has two aspartate residues positioned near the crosslink (D45 and D139) that could have a similar role to D268 in hOGG1.[66] With a crystal structure of the crosslinked intermediate available (PDB ID: 1ORP, **Figure 7.2B**),[67] the same methodology used in Chapter 5 can be applied to identify the catalytic mechanism for Nth. Based on whether Nth has the same, similar, or different catalytic mechanism compared to hOGG1, the generalizability of the hOGG1 β-lyase mechanism for other bifunctional glycosylases that form a DNA−lysine crosslink can be inferred. This would set the stage for the study of bifunctional glycosylases that form a DNA−protein crosslink with residues other than lysine to determine if a similar catalytic mechanism to hOGG1 is utilized. For example, T4 PDG is a bifunctional glycosylase responsible for repairing cyclobutane pyrimidine dimers and forms a crosslink using the N-terminus of T2.[68] Additionally, T4 PDG has a glutamate (E23)[68] in the enzyme

active site which is positioned similarly to D268 in hOGG1 and could have a similar role in catalysis (PDB ID: 2FCC, **Figure 7.2C**). Indeed, mutation of E23 severely reduces T4 PDG catalytic activity.[68] Studying these enzymes would significantly expand our understanding of bifunctional glycosylases and either allow for the proposal of a general bifunctional glycosylase mechanism or show that the β-lyase step is enzyme dependent.

In Chapter 6, the mechanism of Dicer, an enzyme vital for RNAi, was characterized. However, Dicer is not the only enzyme involved in miRNA processing. To generate the pre-miRNA that acts as a substrate for Dicer, drosha ribonuclease III (Drosha) cleaves the backbone of primary miRNA (pri-miRNA). Similar to Dicer, Drosha misfunction is associated with disease, with mutations to residues likely involved in metal binding related to Wilms tumours, the most common childhood genitourinary tract cancer.[69] Currently, the catalytic mechanism of Drosha is unknown, although structural similarities with Dicer suggest it could utilize a similar mechanism.[70] A recent cryo-EM structure depicts Drosha bound to substrate pri-miRNA (PDB ID: 9ASO, **Figure 7.2D**).[71] However, calcium ions were introduced to inhibit catalysis, which could impact substrate binding as was seen for the Ca^{2+}−inhibited cryo-EM structure of Dicer.[72] Therefore, the same methodology used to determine the Dicer Mg^{2+}–binding pattern and mechanism can be applied to Drosha. Additionally, *in silico* mutations can identify how the D1151Y and E1147K mutants associated with Wilm's tumours prevent pre-miRNA biogenesis.[69] Considering that Dicer mutations are also associated with Wilm's tumours,[69] the results found in Chapter 6 combined with a similar study on Drosha would provide a broad understanding of how disruptions in RNAi-related enzymes result in Wilm's tumours and represent a step forward in developing treatments of this disorder.

7.4. Concluding Remarks

In conclusion, this thesis provides a comprehensive description of the various methods enzymes use to cleave glycosidic and phosphodiester bonds in nucleic acids. Specifically, the characterized mechanisms of action for MutY, MBD4, and AlkA uncovered a unique mechanism involving the formation of a DNA−protein crosslink for MutY and identified active site characteristics

that promote or hinder crosslink formation in the glycosylase family. The analysis of hOGG1 provided the first description of a bifunctional glycosylase β-lyase mechanism. Finally, this work clarified the catalytic mechanism of Dicer and provided a rationale for the reduced miRNA biogenesis for DICER1 syndrome-related mutants. Overall, the improved description of enzyme mechanisms provided in this thesis will aid in the development of transition state analogue inhibitors as therapeutics and improves our understanding of RNAi and RNAi-related disorders. Additionally, the methodology outlined provides a foundation for future work investigating the mechanisms of glycosylases and endonucleases that remain unclear.

7.5. References

(1) Caldecott, K. W. Mammalian DNA base excision repair: Dancing in the moonlight. *DNA Repair* **2020**, *93*, 102921. DOI: https://doi.org/10.1016/j.dnarep.2020.102921

(2) Yoshida, T.; Asano, Y.; Ui-Tei, K. Modulation of microRNA processing by Dicer via its associated dsRNA binding proteins. *Non-Coding RNA* **2021**, *7* (3). DOI: 10.3390/ncrna7030057

(3) Zapletal, D.; Kubicek, K.; Svoboda, P.; Stefl, R. Dicer structure and function: Conserved and evolving features. *EMBO Rep.* **2023**, *24* (7), e57215. DOI: https://doi.org/10.15252/embr.202357215

(4) Ambros, V.; Bartel, B.; Bartel, D. P.; Burge, C. B.; Carrington, J. C.; Chen, X.; Dreyfuss, G.; Eddy, S. R.; Griffiths-Jones, S.; Marshall, M.; et al. A uniform system for microRNA annotation. *RNA* **2003**, *9* (3), 277-279. DOI: 10.1261/rna.2183803

(5) McCann, J. A. B.; Berti, P. J. Adenine release is fast in MutY-catalyzed hydrolysis of G:A and 8-oxo-G:A DNA mismatches. *J. Biol. Chem.* **2003**, *278* (32), 29587-29592. DOI: 10.1074/jbc.M212474200

(6) Lu, A. L.; Yuen, D. S.; Cillo, J. Catalytic mechanism and DNA substrate recognition of *escherichia coli* MutY protein. *J. Biol. Chem.* **1996**, *271* (39), 24138-24143. DOI: 10.1074/jbc.271.39.24138 (acccessed 2022/01/06).

(7) Wright, P. M.; Yu, J.; Cillo, J.; Lu, A. L. The active site of the *escherichia coli* MutY DNA adenine glycosylase. *J. Biol. Chem.* **1999**, *274* (41), 29011-29018. DOI: 10.1074/jbc.274.41.29011

(8) Williams, S. D.; David, S. S. A single engineered point mutation in the adenine glycosylase MutY confers bifunctional glycosylase/AP lyase activity. *Biochemistry* **2000**, *39* (33), 10098-10109. DOI: 10.1021/bi0004652

(9) Brinkmeyer, M. K.; Pope, M. A.; David, S. S. Catalytic contributions of key residues in the adenine glycosylase MutY revealed by pH-dependent kinetics and cellular repair assays. *Cell Chem. Biol.* **2012**, *19* (2), 276-286. DOI: 10.1016/j.chembiol.2011.11.011

(10) Woods, R. D.; O'Shea, V. L.; Chu, A.; Cao, S.; Richards, J. L.; Horvath, M. P.; David, S. S. Structure and stereochemistry of the base excision repair glycosylase MutY reveal a mechanism similar to retaining glycosidases. *Nucleic Acids Res.* **2016**, *44* (2), 801-810. DOI: 10.1093/nar/gkv1469

(11) Demir, M.; Russelburg, L. P.; Lin, W.-J.; Trasviña-Arenas, C. H.; Huang, B.; Yuen, P. K.; Horvath, M. P.; David, S. S. Structural snapshots of base excision by the cancer-associated variant MutY N146S reveal a retaining mechanism. *Nucleic Acids Res.* **2023**, *51* (3), 1034-1049. DOI: 10.1093/nar/gkac1246 (acccessed 1/23/2023).

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

(13) Mullins, E. A.; Rodriguez, A. A.; Bradley, N. P.; Eichman, B. F. Emerging roles of DNA glycosylases and the base excision repair pathway. *Trends Biochem. Sci* **2019**, *44* (9), 765-781. DOI: https://doi.org/10.1016/j.tibs.2019.04.006

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

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

(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* (2), 506–555.

(17) Dallosso, A. R.; Dolwani, S.; Jones, N.; Jones, S.; Colley, J.; Maynard, J.; Idziaszczyk, S.; Humphreys, V.; Arnold, J.; Donaldson, A.; et al. Inherited predisposition to colorectal adenomas caused by multiple rare alleles of MUTYH but not OGG1, NUDT1, NTH1 or NEIL 1, 2 or 3. *Gut* **2008**, *57* (9), 1252-1255. DOI: 10.1136/gut.2007.145748

(18) Hashimoto, H.; Zhang, X.; Cheng, X. Excision of thymine and 5-hydroxymethyluracil by the MBD4 DNA glycosylase domain: Structural basis and implications for active DNA demethylation. *Nucleic Acids Res.* **2012**, *40* (17), 8276-8284. DOI: 10.1093/nar/gks628 (acccessed 1/14/2025).

(19) Cortellino, S.; Turner, D.; Masciullo, V.; Schepis, F.; Albino, D.; Daniel, R.; Skalka, A. M.; Meropol, N. J.; Alberti, C.; Larue, L.; et al. The base excision repair enzyme MED1 mediates DNA damage response to antitumor drugs and is associated with mismatch repair system integrity. *Proc. Natl. Acad. Sci. U.S.A* **2003**, *100* (25), 15071-15076. DOI: doi:10.1073/pnas.2334585100

(20) Turner, D. P.; Cortellino, S.; Schupp, J. E.; Caretti, E.; Loh, T.; Kinsella, T. J.; Bellacosa, A. The DNA N-glycosylase MED1 exhibits preference for halogenated pyrimidines and is involved in the cytotoxicity of 5-iododeoxyuridine. *Cancer Res.* **2006**, *66* (15), 7686-7693. DOI: 10.1158/0008-5472.Can-05-4488 (acccessed 1/17/2025).

(21) Palles, C.; West, H. D.; Chew, E.; Galavotti, S.; Flensburg, C.; Grolleman, J. E.; Jansen, E. A. M.; Curley, H.; Chegwidden, L.; Arbe-Barnes, E. H.; et al. Germline MBD4 deficiency causes a multi-tumor predisposition syndrome. *Am. J. Hum. Genet.* **2022**, *109* (5), 953-960. DOI: https://doi.org/10.1016/j.ajhg.2022.03.018

(22) Blombery, P.; Ryland, G. L.; Fox, L. C.; Stark, Z.; Wall, M.; Jarmolowicz, A.; Roesley, A.; Thompson, E. R.; Grimmond, S. M.; Panicker, S.; et al. Methyl-CpG binding domain 4, DNA glycosylase (MBD4)-associated neoplasia syndrome associated with a homozygous missense variant in MBD4: Expansion of an emerging phenotype. *Br. J. Haematol.* **2022**, *198* (1), 196-199. DOI: https://doi.org/10.1111/bjh.18178

(23) Suzuki, S.; Iwaizumi, M.; Tseng-Rogenski, S.; Hamaya, Y.; Miyajima, H.; Kanaoka, S.; Sugimoto, K.; Carethers, J. M. Production of truncated MBD4 protein by frameshift mutation in DNA mismatch repair-deficient cells enhances 5-fluorouracil sensitivity that is independent of hMLH1 status. *Cancer Biol. Ther.* **2016**, *17* (7), 760-768. DOI: 10.1080/15384047.2016.1178430

(24) Averill, J. R.; Lin, J. C.; Jung, J.; Jung, H. Novel insights into the role of translesion synthesis polymerase in DNA incorporation and bypass of 5-fluorouracil in colorectal cancer. *Nucleic Acids Res* **2024**, *52* (8), 4295-4312. DOI: 10.1093/nar/gkae102

(25) Pidugu, L. S.; Bright, H.; Lin, W.-J.; Majumdar, C.; Van Ostrand, R. P.; David, S. S.; Pozharski, E.; Drohat, A. C. Structural insights into the mechanism of base excision by MBD4. *J. Mol. Biol.* **2021**, *433* (15), 167097. DOI: https://doi.org/10.1016/j.jmb.2021.167097

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

(27) Dylewska, M.; Dąbrowska, I.; Ćwiek, K.; Padoł, K.; Mielecki, D.; Sokołowska, B.; Poznański, J.; Maciejewska, A. M. AlkA glycosylase and AlkB dioxygenase constitute an effective protective system for endogenously arising acrolein: E. Coli AlkA glycosylase excises acrolein adduct to adenine. *J. Mol. Biol.* **2025**, *437* (2), 168912. DOI: https://doi.org/10.1016/j.jmb.2024.168912

(28) Hollis, T.; Ichikawa, Y.; Ellenberger, T. DNA bending and a flip-out mechanism for base excision by the helix-hairpin-helix DNA glycosylase, escherichia coli AlkA. *EMBO J* **2000**, *19* (4), 758-766. DOI: 10.1093/emboj/19.4.758

(29) O'Brien, P. J.; Ellenberger, T. The escherichia coli 3-methyladenine DNA glycosylase AlkA has a remarkably versatile active site*. *J. Biol. Chem.* **2004**, *279* (26), 26876-26884. DOI: https://doi.org/10.1074/jbc.M403860200

(30) Labahn, J.; Schärer, O. D.; Long, A.; Ezaz-Nikpay, K.; Verdine, G. L.; Ellenberger, T. E. Structural basis for the excision repair of alkylation-damaged DNA. *Cell* **1996**, *86* (2), 321-329. DOI: https://doi.org/10.1016/S0092-8674(00)80103-8

(31) Jiang, D.; Zhang, L.; Dong, K.; Gong, Y.; Oger, P. Biochemical characterization and mutational studies of a novel 3-methlyadenine DNA glycosylase II from the hyperthermophilic thermococcus gammatolerans. *DNA Repair* **2021**, *97*, 103030. DOI: https://doi.org/10.1016/j.dnarep.2020.103030

(32) Gilboa, R.; Zharkov, D. O.; Golan, G.; Fernandes, A. S.; Gerchman, S. E.; Matz, E.; Kycia, J. H.; Grollman, A. P.; Shoham, G. Structure of formamidopyrimidine-DNA glycosylase covalently complexed to DNA*. *J. Biol. Chem.* **2002**, *277* (22), 19811-19816. DOI: https://doi.org/10.1074/jbc.M202058200

(33) Fromme, J. C.; Verdine, G. L. Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine DNA glycosylase MutM. *Nat Struct Biol* **2002**, *9* (7), 544-552.

(34) Zharkov, D. O.; Golan, G.; Gilboa, R.; Fernandes, A. S.; Gerchman, S. E.; Kycia, J. H.; Rieger, R. A.; Grollman, A. P.; Shoham, G. Structural analysis of an *escherichia coli* endonuclease VIII covalent reaction intermediate. *EMBO J* **2002**, *21* (4), 789-800. DOI: https://doi.org/10.1093/emboj/21.4.789

(35) Huskova, A.; Dinesh, D. C.; Srb, P.; Boura, E.; Veverka, V.; Silhan, J. Model of abasic site DNA cross-link repair; from the architecture of NEIL3 DNA binding domains to the X-structure model. *Nucleic Acids Res.* **2022**, *50* (18), 10436-10448. DOI: 10.1093/nar/gkac793 (acccessed 10/16/2025).

(36) Fromme, J. C.; Verdine, G. L. Structure of a trapped endonuclease III-DNA covalent intermediate. *EMBO J.* **2003**, *22* (13), 3461-3471. DOI: 10.1093/emboj/cdg311

(37) Kuznetsov, N. A.; Koval, V. V.; Fedorova, O. S. Mechanism of recognition and repair of damaged DNA by human 8-oxoguanine DNA glycosylase hOGG1. *Biochemistry* **2011**, *76* (1), 118-130. DOI: 10.1134/S0006297911010123

(38) Tian, G.; Katchur, S. R.; Jiang, Y.; Briand, J.; Schaber, M.; Kreatsoulas, C.; Schwartz, B.; Thrall, S.; Davis, A. M.; Duvall, S.; et al. Small molecule-mediated allosteric activation of the base excision repair enzyme 8-oxoguanine DNA glycosylase and its impact on mitochondrial function. *Sci. Rep.* **2022**, *12* (1), 14685. DOI: 10.1038/s41598-022-18878-2

(39) Norman, D. P. G.; Chung, S. J.; Verdine, G. L. Structural and biochemical exploration of a critical amino acid in human 8-oxoguanine glycosylase. *Biochemistry* **2003**, *42* (6), 1564-1572. DOI: 10.1021/bi026823d

(40) Iwakawa, H.-o.; Tomari, Y. Life of RISC: Formation, action, and degradation of RNA-induced silencing complex. *Mol. Cell* **2022**, *82* (1), 30-43. DOI: https://doi.org/10.1016/j.molcel.2021.11.026

(41) Sedger, L. M. microRNA control of interferons and interferon induced anti-viral activity. *Mol. Immunol.* **2013**, *56* (4), 781-793. DOI: https://doi.org/10.1016/j.molimm.2013.07.009

(42) Ying, H.; Ebrahimi, M.; Keivan, M.; Khoshnam, S. E.; Salahi, S.; Farzaneh, M. miRNAs; A novel strategy for the treatment of COVID-19. *Cell Biol. Int.* **2021**, *45* (10), 2045-2053. DOI: https://doi.org/10.1002/cbin.11653

(43) Traber, G. M.; Yu, A.-M. The growing class of novel RNAi therapeutics. *Mol. Pharmacol.* **2024**, *106* (1), 13-20. DOI: 10.1124/molpharm.124.000895 (acccessed 2025/12/08).

(44) Peng, Y.; Croce, C. M. The role of microRNAs in human cancer. *Signal Transduct Target Ther* **2016**, *1*, 15004. DOI: 10.1038/sigtrans.2015.4

(45) Bartel, D. P. Metazoan microRNAs. *Cell* **2018**, *173* (1), 20-51. DOI: https://doi.org/10.1016/j.cell.2018.03.006

(46) Chiosea, S.; Jelezcova, E.; Chandran, U.; Luo, J.; Mantha, G.; Sobol, R. W.; Dacic, S. Overexpression of Dicer in precursor lesions of lung adenocarcinoma. *Cancer Res.* **2007**, *67* (5), 2345-2350. DOI: 10.1158/0008-5472.Can-06-3533 (acccessed 6/26/2023).

(47) Foulkes, W. D.; Priest, J. R.; Duchaine, T. F. DICER1: Mutations, microRNAs and mechanisms. *Nat. Rev. Cancer* **2014**, *14* (10), 662-672. DOI: 10.1038/nrc3802

(48) Chen, P. S.; Lin, S. C.; Tsai, S. J. Complexity in regulating microRNA biogenesis in cancer. *Exp Biol Med (Maywood)* **2020**, *245* (5), 395-401. DOI: 10.1177/1535370220907314

(49) Dedes, K. J.; Natrajan, R.; Lambros, M. B.; Geyer, F. C.; Lopez-Garcia, M. A.; Savage, K.; Jones, R. L.; Reis-Filho, J. S. Down-regulation of the miRNA master regulators Drosha and Dicer is associated with specific subgroups of breast cancer. *Eur. J. Cancer* **2011**, *47* (1), 138-150. DOI: https://doi.org/10.1016/j.ejca.2010.08.007

(50) Klein, S.; Lee, H.; Ghahremani, S.; Kempert, P.; Ischander, M.; Teitell, M. A.; Nelson, S. F.; Martinez-Agosto, J. A. Expanding the phenotype of mutations in DICER1: Mosaic missense mutations in the RNase IIIb domain of DICER1 cause glow syndrome. *J Med Genet* **2014**, *51* (5), 294-302. DOI: 10.1136/jmedgenet-2013-101943

(51) Vedanayagam, J.; Chatila, W. K.; Aksoy, B. A.; Majumdar, S.; Skanderup, A. J.; Demir, E.; Schultz, N.; Sander, C.; Lai, E. C. Cancer-associated mutations in DICER1 RNase IIIa and IIIb domains exert similar effects on miRNA biogenesis. *Nat. Commun.* **2019**, *10* (1), 3682. DOI: 10.1038/s41467-019-11610-1

(52) de Kock, L.; Wu, M. K.; Foulkes, W. D. Ten years of DICER1 mutations: Provenance, distribution, and associated phenotypes. *Hum. Mutat.* **2019**, *40* (11), 1939-1953. DOI: 10.1002/humu.23877

(53) Provost, P.; Dishart, D.; Doucet, J.; Frendewey, D.; Samuelsson, B.; Rådmark, O. Ribonuclease activity and RNA binding of recombinant human Dicer. *EMBO J* **2002**, *21* (21), 5864-5874. DOI: https://doi.org/10.1093/emboj/cdf578

(54) Takeshita, D.; Zenno, S.; Lee, W. C.; Nagata, K.; Saigo, K.; Tanokura, M. Homodimeric structure and double-stranded RNA cleavage activity of the C-terminal RNase III domain of human Dicer. *J. Mol. Biol.* **2007**, *374* (1), 106-120. DOI: 10.1016/j.jmb.2007.08.069

(55) Tubbs, J. L.; Tainer, J. A. DNA damage: Alkylation. In *Encyclopedia of biological chemistry (second edition)*, Lennarz, W. J., Lane, M. D. Eds.; Academic Press, 2013; pp 9-15.

(56) Zhang, Z.; Shen, J.; Yang, Y.; Li, J.; Cao, W.; Xie, W. Structural basis of substrate specificity in geobacter metallireducens smug1. *ACS Chem. Biol.* **2016**, *11* (6), 1729-1736. DOI: 10.1021/acschembio.6b00164

(57) An, Q.; Robins, P.; Lindahl, T.; Barnes, D. E. 5-fluorouracil incorporated into DNA is excised by the smug1 DNA glycosylase to reduce drug cytotoxicity. *Cancer Res.* **2007**, *67* (3), 940-945. DOI: 10.1158/0008-5472.Can-06-2960 (acccessed 12/4/2025).

(58) Iakovlev, D. A.; Alekseeva, I. V.; Vorobjev, Y. N.; Kuznetsov, N. A.; Fedorova, O. S. The role of active-site residues phe98, his239, and arg243 in DNA binding and in the catalysis of human uracil–DNA glycosylase smug1. *Molecules* **2019**, *24* (17), 3133.

(59) Matsubara, M.; Tanaka, T.; Terato, H.; Ohmae, E.; Izumi, S.; Katayanagi, K.; Ide, H. Mutational analysis of the damage-recognition and catalytic mechanism of human smug1 DNA glycosylase. *Nucleic Acids Res.* **2004**, *32* (17), 5291-5302. DOI: 10.1093/nar/gkh859 (acccessed 12/4/2025).

(60) Yu, H.; Dalby, P. A. Chapter two - A beginner's guide to molecular dynamics simulations and the identification of cross-correlation networks for enzyme engineering. In *Methods in enzymology*, Tawfik, D. S. Ed.; Vol. 643; Academic Press, 2020; pp 15-49.

(61) Greenacre, M.; Groenen, P. J. F.; Hastie, T.; D'Enza, A. I.; Markos, A.; Tuzhilina, E. Principal component analysis. *Nature Reviews Methods Primers* **2022**, *2* (1), 100. DOI: 10.1038/s43586-022-00184-w

(62) Baptiste, B. A.; Katchur, S. R.; Fivenson, E. M.; Croteau, D. L.; Rumsey, W. L.; Bohr, V. A. Enhanced mitochondrial DNA repair of the common disease-associated variant, Ser326Cys, of hOGG1 through small molecule intervention. *Free Radical Biol. Med.* **2018**, *124*, 149-162. DOI: https://doi.org/10.1016/j.freeradbiomed.2018.05.094

(63) Cardozo-Pelaez, F.; Sanchez-Contreras, M.; Nevin, A. B. C. Ogg1 null mice exhibit age-associated loss of the nigrostriatal pathway and increased sensitivity to MPTP. *Neurochem. Int.* **2012**, *61* (5), 721-730. DOI: https://doi.org/10.1016/j.neuint.2012.06.013

(64) Hamm, M. L.; Gill, T. J.; Nicolson, S. C.; Summers, M. R. Substrate specificity of Fpg (MutM) and hOGG1, two repair glycosylases. *J. Am. Chem. Soc.* **2007**, *129* (25), 7724-7725. DOI: 10.1021/ja0716453

(65) Dizdaroglu, M.; Laval, J.; Boiteux, S. Substrate specificity of the escherichia coli endonuclease III: Excision of thymine- and cytosine-derived lesions in DNA produced by radiation-generated free radicals. *Biochemistry* **1993**, *32* (45), 12105-12111. DOI: 10.1021/bi00096a022

(66) Fromme, J. C.; Verdine, G. L. Structure of a trapped endonuclease III–DNA covalent intermediate. *EMBO J* **2003**, *22* (13), 3461-3471. DOI: 10.1093/emboj/cdg311

(67) Golan, G.; Zharkov, D. O.; Grollman, A. P.; Dodson, M. L.; McCullough, A. K.; Lloyd, R. S.; Shoham, G. Structure of T4 pyrimidine dimer glycosylase in a reduced imine covalent complex with abasic site-containing DNA. *J. Mol. Biol.* **2006**, *362* (2), 241-258. DOI: https://doi.org/10.1016/j.jmb.2006.06.059

(68) Hori, N.; Doi, T.; Karaki, Y.; Kituchi, M.; Ikehara, M.; Ohtsuka, E. Participation of glutamic acid 23 of T4 endonuclease V in the β-elimination reaction of an abasic site in a synthetic duplex DNA.

Nucleic Acids Res. **1992**, *20* (18), 4761-4764. DOI: 10.1093/nar/20.18.4761 (acccessed 9/15/2025).

(69) Rakheja, D.; Chen, K. S.; Liu, Y.; Shukla, A. A.; Schmid, V.; Chang, T.-C.; Khokhar, S.; Wickiser, J. E.; Karandikar, N. J.; Malter, J. S.; et al. Somatic mutations in DROSHA and DICER1 impair microRNA biogenesis through distinct mechanisms in Wilms tumours. *Nat. Commun.* **2014**, *5* (1), 4802. DOI: 10.1038/ncomms5802

(70) Aharoni, R.; Dror, T. Dynamical comparison between Drosha and Dicer reveals functional motion similarities and dissimilarities. *PLoS One* **2019**, *14* (12). DOI: https://doi.org/10.1371/journal.pone.0226147

(71) Garg, A.; Shang, R.; Cvetanovic, T.; Lai, E. C.; Joshua-Tor, L. The structural landscape of microprocessor-mediated processing of pri-let-7 miRNAs. *Mol. Cell* **2024**, *84* (21), 4175-4190.e4176. DOI: https://doi.org/10.1016/j.molcel.2024.09.008

(72) Lee, Y.-Y.; Lee, H.; Kim, H.; Kim, V. N.; Roh, S.-H. Structure of the human DICER–pre-miRNA complex in a dicing state. *Nature* **2023**, *615* (7951), 331-338. DOI: 10.1038/s41586-023-05723-3

Appendix A

Supplementary Information for Chapter 2: Distinctive Formation of a DNA–Protein Cross-Link During the Repair of DNA Oxidative Damage: Insights into Human Disease from MD Simulations and QM/MM Calculations

Contains Tables A.1−A.9, Figures A.1−A.16, and full computational methods

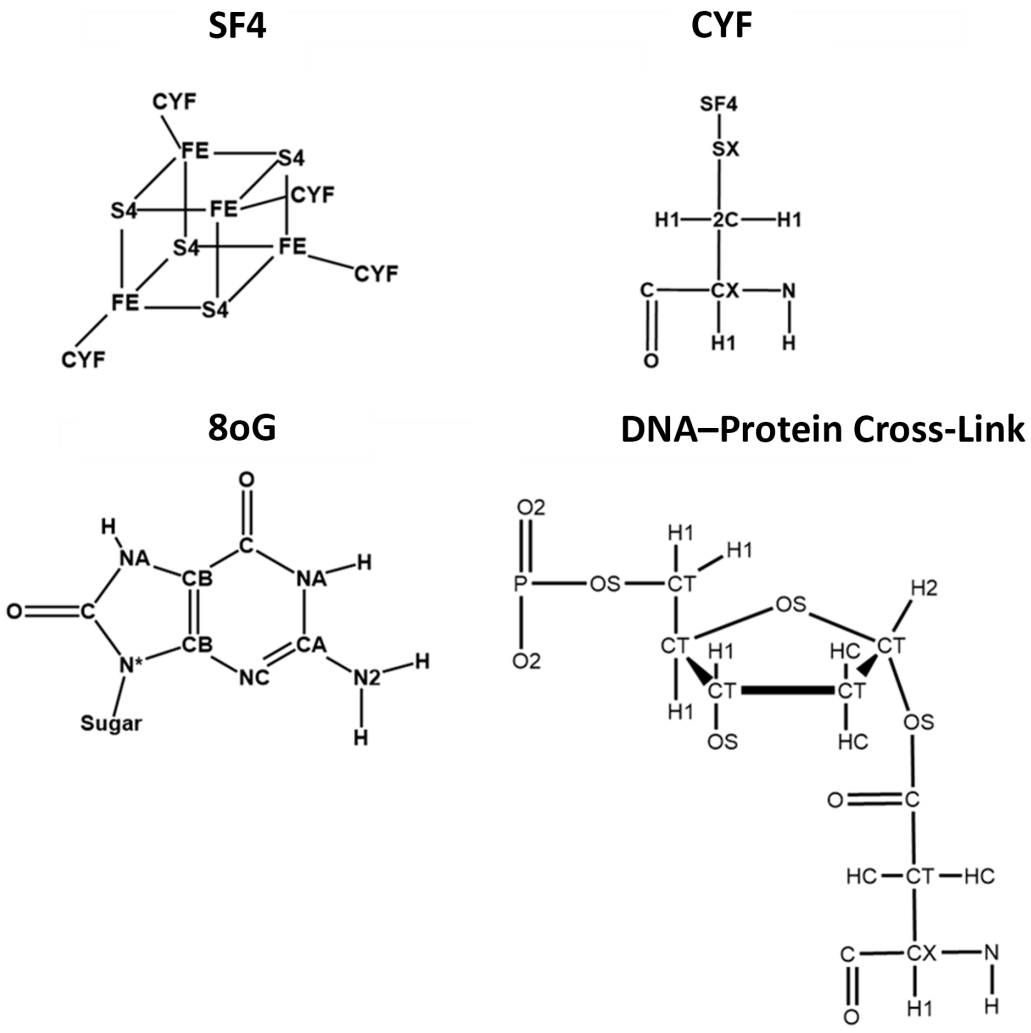


Figure A.1. Atom types used in the present work for the Fe_4S_4 cluster (SF4), Fe_4S_4-coordinated cysteine (CYF), 8oG, and the DNA–protein cross-link.

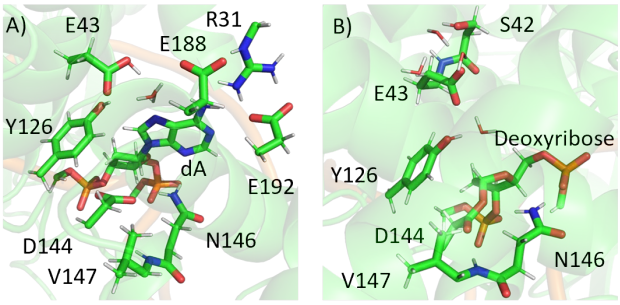


Figure A.2. Representative QM layers used in ONIOM calculations of the A) RC and B) IC_{CL}. The QM layer of other RC may contain one less water molecule, while the QM layer of other IC_{CL} may contain one more water molecule.

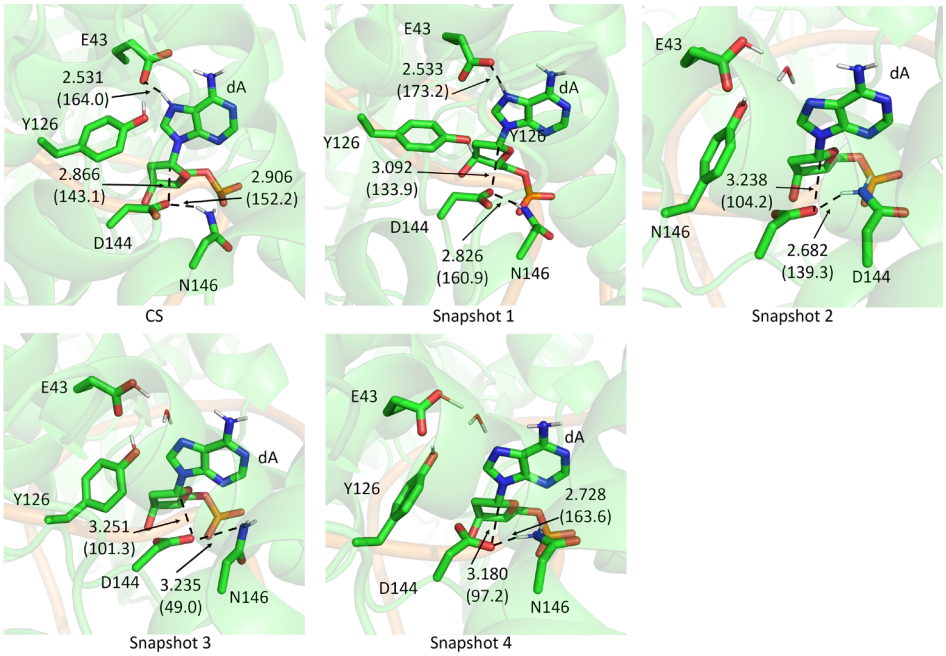


Figure A.3. Key ONIOM(M06-2X/6-31G(d,p):AMBERff14SB) hydrogen-bonding distances (Å) and angles (in parentheses, °) as well as D144–C1′ distance and ∠(N9C1′D144$_{O\delta}$) (in parentheses) for the RC (snapshots 2–4) and RC_{H+} (crystal structure and snapshot 1) models.

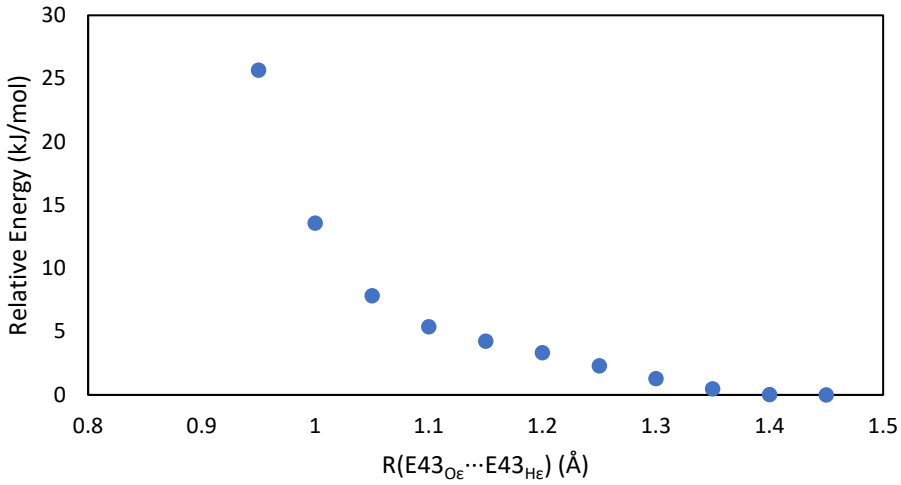


Figure A.4. Relative ONIOM(M06-2X/6-31G(d,p):AMBERff14SB) energy (kJ/mol) as a function of the $E43_{O\varepsilon}\cdots E43_{H\varepsilon}$ distance prior to deglycosylation. Data generated based on the FLRC MD snapshot 1, which is representative of FLRC models containing a direct E43···N7 hydrogen bond.

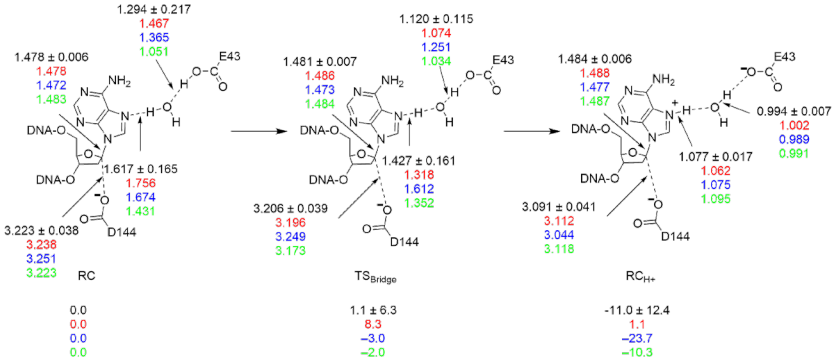


Figure A.5. Key ONIOM(M06-2X/6-31G(d,p):AMBERff14SB) distances (Å) and average relative Gibbs energies (kJ/mol) during water-mediated protonation at N7 of dA initiated from MD snapshot 2 (red), MD snapshot 3 (blue), and MD snapshot 4 (green). The average value is in black.

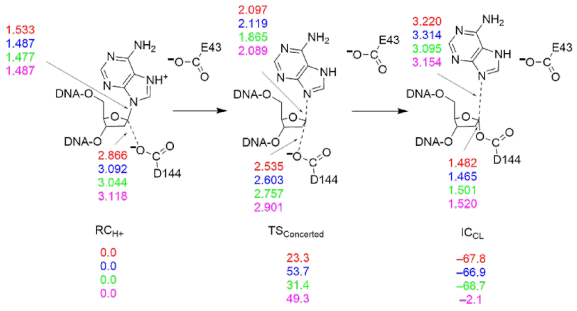


Figure A.6. Key ONIOM(M06-2X/6-31G(d,p):AMBERff14SB) bond lengths (Å) for the one-step deglycosylation pathway initiated from the FLRC crystal structure (red), MD snapshot 1 (blue), MD snapshot 2 (green), and MD snapshot 3 (purple). Average values available in the main text Figure 2.3A.

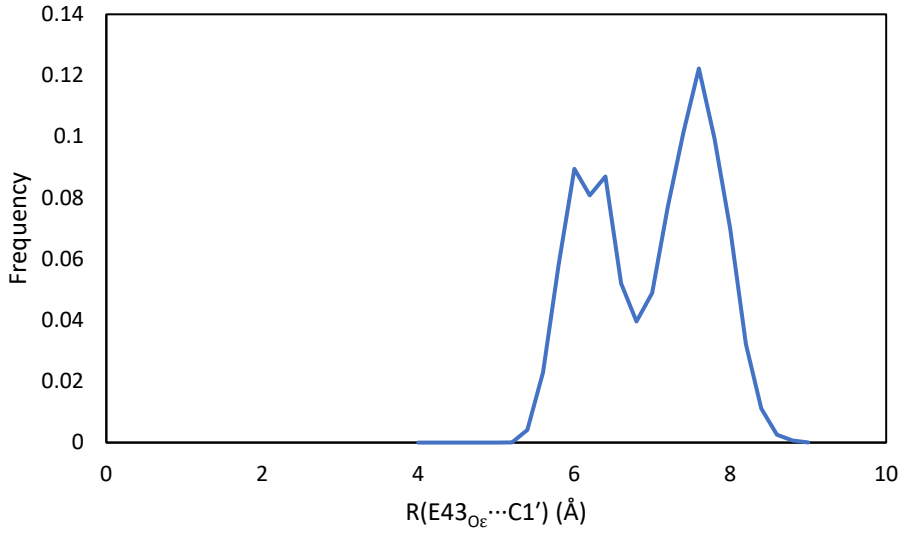


Figure A.7. Histogram of the distance (Å) between C1′ of dA and closest $E43_{O\varepsilon}$ over four MD simulation replicas of the DNA–MutY cross-linked intermediate (IC_{CL}) highlighting the peaks corresponding to the active (5.5–6.5 Å) and inactive (7–8.5 Å) conformations.

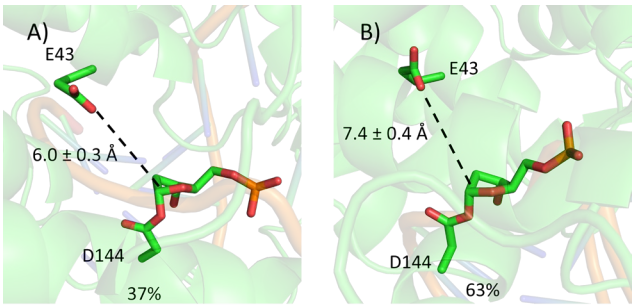


Figure A.8. MD representative structures and occupancies of the A) active ($R(E43_{O\varepsilon}\cdots C1')$ = 5.5–6.5 Å) and B) inactive ($R(E43_{O\varepsilon}\cdots C1')$ = 7.5–8.5 Å) E43 conformations of IC_{CL}. The histogram of $R(E43_{O\varepsilon}\cdots C1')$ can be found in Figure S7.

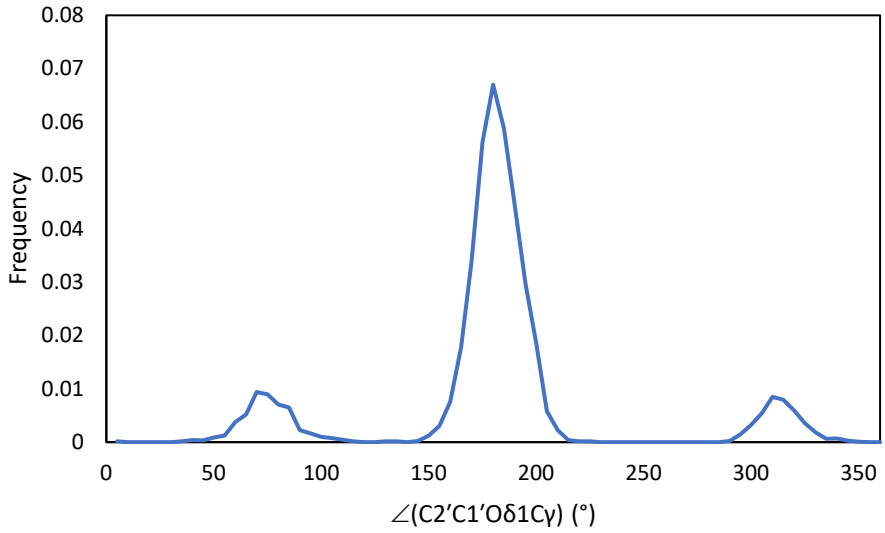


Figure A.9. Histogram of the ∠(C2′C1′Oδ1Cγ) dihedral over four MD simulation replicas for the IC_{CL}, highlighting the peaks corresponding to the active (140–220°) and inactive (40–120°, 280–350°) conformations.

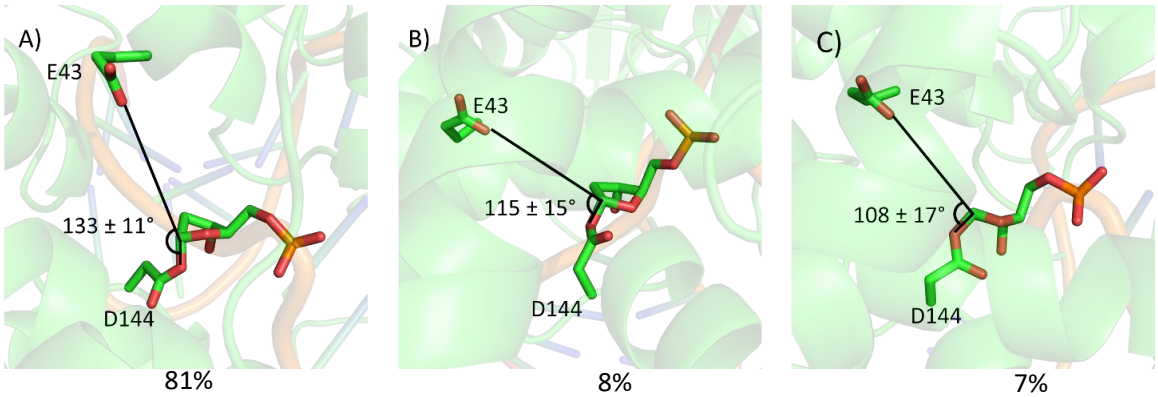


Figure A.10. Representative structures of the A) active (∠(C2′C1′Oδ1Cγ) = 140–220°) and inactive B) ∠(C2′C1′Oδ1Cγ) = 40–120° or C) 280–350° cross-link conformations from MD simulations. The histogram of ∠(C2′C1′Oδ1Cγ) can be found in Figure A.9.

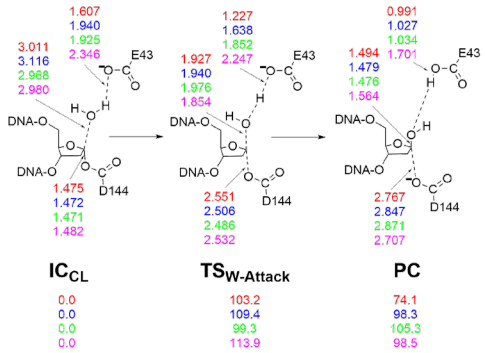


Figure A.11. Key ONIOM(M06-2X/6-31G(d,p):AMBERff14SB) bond lengths (Å) for the cross-link hydrolysis (enzyme regeneration) pathway initiated from the MD snapshot 1 (red), MD snapshot 2 (blue), MD snapshot 3 (green), and MD snapshot 4 (purple). Average values available in the main text Figure 2.5.

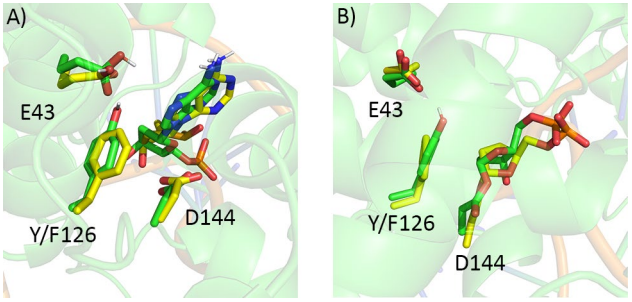


Figure A.12. Overlay of the MutY active site from the MD representative structures of the wild-type (green) and Y126F (yellow) MutY A) RC and B) IC_{CL}.

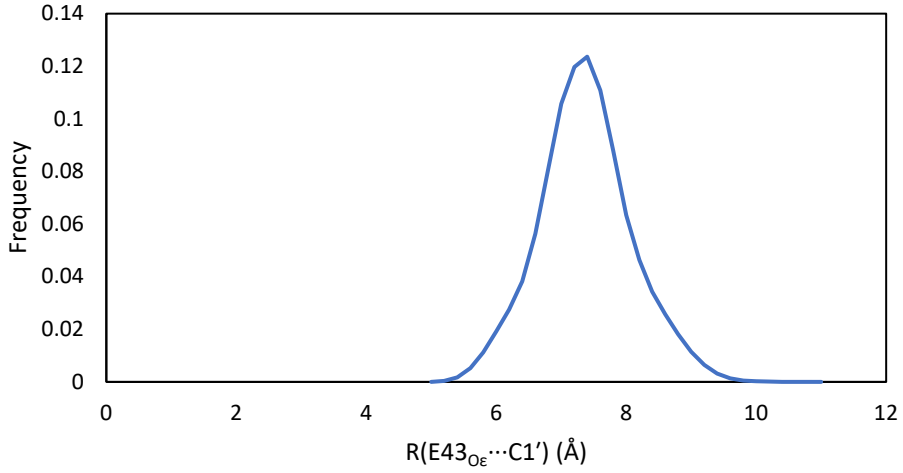


Figure A.13. Histogram of the distance (Å) between C1′ of dA and closest $E43_{O\varepsilon}$ over four MD simulation replicas of the Y126F DNA–MutY cross-linked intermediate (IC_{CL}).

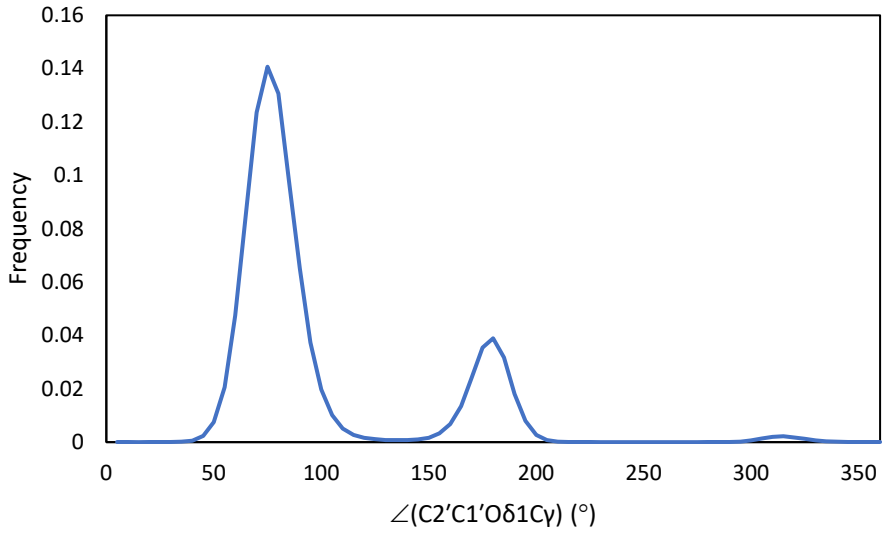


Figure A>14. Histogram of the ∠(C2′C1′Oδ1Cγ) dihedral over four MD simulation replicas for the Y126F MutY IC_{CL}, highlighting the peaks corresponding to the active (140–220°) and inactive (40–120°, 280–350°) conformations.

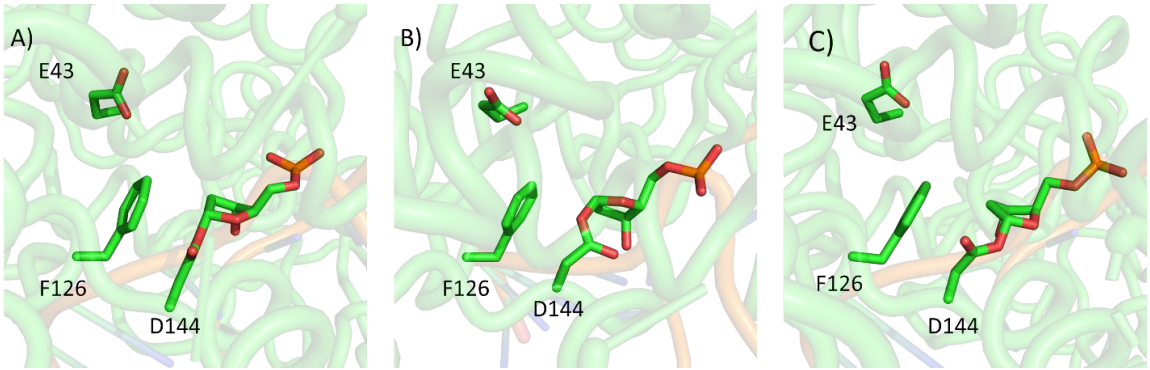


Figure A.15. Representative structures of the A) active (∠(C2′C1′Oδ1Cγ) = 140–220°) and inactive (B) ∠(C2′C1′Oδ1Cγ) = 40–120° or C) 280–350°) cross-link conformations from Y126F MD simulations. The histogram of ∠(C2′C1′Oδ1Cγ) can be found in Figure A.14.

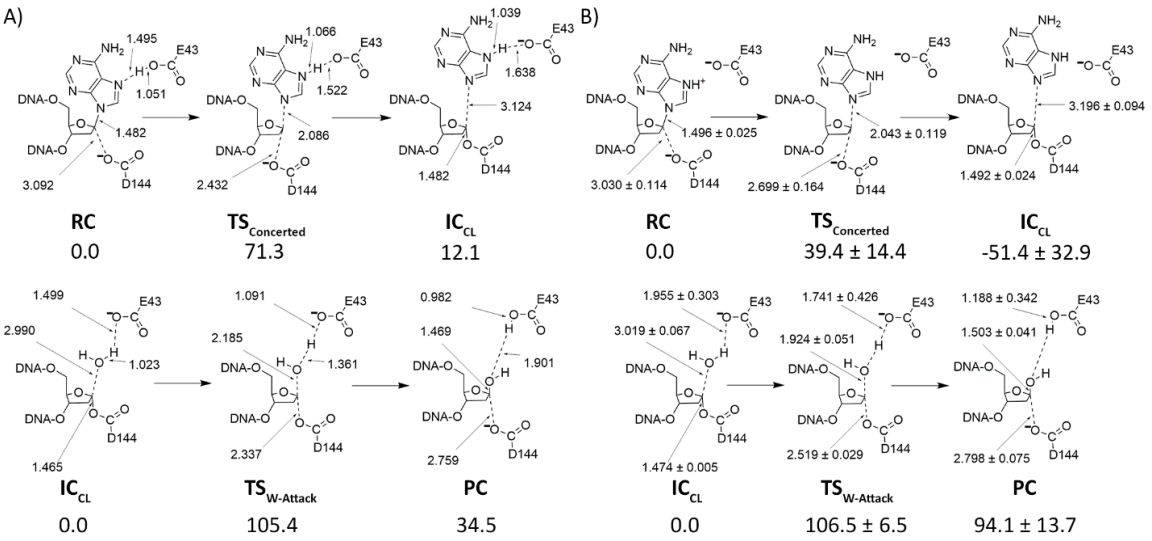


Figure A.16. ONIOM(M06-2X/6-311+G(2df,p):AMBERff14SB)//ONIOM(M06-2X/6-31G(d,p):AMBERff14SB) bond distances (Å) and relative Gibbs energies (kJ/mol) for the MutY catalytic mechanism using the A) EE or B) ME scheme. ME data averaged over relevant models is provided.

Table A.1. Parameters for non-standard residues used in the present work.[a]

Bond Parameters				
Residue	Bond	K_r (kcal mol^{-1} Å^{-2})	R_{eq} (Å)	
8oG	CB-NA	411.1	1.3910	
SF4	FE-S4	56.8	2.1569	
SF4	SX-FE	49.0	2.2387	
SF4	2C-SX	237.0	1.8100	
Angle Parameters				
Residue	Angle	K_θ (kcal mol^{-1} radian^{-2})	θ_{eq} (deg.)	
8oG	C-N*-CB	0.000	125.09	
8oG	C-NA-CB	0.000	125.09	
8oG	C-CB-NA	69.987	114.11	
8oG	CB-CB-NA	69.830	121.38	
8oG	H-NA-CB	47.620	119.28	
SF4	2C-SX-FE	72.070	106.05	
SF4	FE-S4-FE	38.510	70.58	
SF4	SX-FE-S4	44.100	112.31	
SF4	S4-FE-S4	36.370	106.42	
SF4	CX-2C-SX	50.000	108.60	
SF4	H1-2C-SX	50.000	109.50	
Dihedral Parameters				
Residue	Dihedral	$V_n/2$ (kcal mol^{-1})	γ (deg.)	n
8oG	C -CB-NA-C	0.625	180	2
8oG	C -CB-NA-H	0.625	180	2
8oG	CB-CB-NA-C	0.625	180	2
8oG	CB-CB-NA-H	0.625	180	2
SF4	2C-SX-FE-S4	0.000	0	3
SF4	CX-2C-SX-FE	0.000	0	3
SF4	H1-2C-SX-FE	0.000	0	3
SF4	FE-S4-FE-S4	0.000	0	3

SF4	N -CX-2C-SX	0.033	0	4
SF4	N -CX-2C-SX	0.251	0	3
SF4	N -CX-2C-SX	0.486	180	2
SF4	N -CX-2C-SX	0.154	0	1
SF4	SX-2C-CX-C	0.075	0	4
SF4	SX-2C-CX-C	0.251	0	3
SF4	SX-2C-CX-C	0.337	180	2
SF4	SX-2C-CX-C	0.269	180	1
SF4	SX-FE-S4-FE	0.000	0	3
SF4	S4-FE-S4-FE	0.000	0	3

Improper Torsion Parameters

Residue	Dihedral	$V_n/2$ (kcal mol^{-1})	γ (deg.)	n
8oG	C -CA-NA-H	1.0	180	2
8oG	C -CB-N*-CT	1.1	180	2
8oG	N*-NA-C -O	10.5	180	2
8oG	C -CB-NA-H	1.0	180	2
8oG	C -CB-CB-NA	1.1	180	2
8oG	CB-NA-C -O	10.5	180	2
8oG	CB-N*-CB-NC	1.1	180	2

Van der Waals Parameters

Residue	Non-Bond	R* (Å)	ε (kcal mol^{-1})
SF4	FE	1.386	0.013571
SF4	SX	2.000	0.250000
SF4	S4	2.000	0.250000

[a]8oG parameters were adapted from the literature.[1]

Table A.2. Partial charges assigned to non-standard residues in the present work. [a]

Residue	Atom Name	Charge
Fe-S Cluster	Fe	–0.6991
Fe-S Cluster	S4	0.8228
Cys Coordinating to Fe-S cluster	N	–0.4157
Cys Coordinating to Fe-S cluster	H	–0.2719
Cys Coordinating to Fe-S cluster	Cα	0.0213
Cys Coordinating to Fe-S cluster	Hα	0.1124
Cys Coordinating to Fe-S cluster	Cβ	0.0660
Cys Coordinating to Fe-S cluster	Hβ	–0.0485
Cys Coordinating to Fe-S cluster	Sγ	–0.6120
Cys Coordinating to Fe-S cluster	C	0.5973
Cys Coordinating to Fe-S cluster	O	–0.5679
Cross-Link	N	–0.5163
Cross-Link	Cα	0.0381
Cross-Link	Hα	0.0880
Cross-Link	Cβ	0.5431
Cross-Link	Hβ	–0.1224
Cross-Link	Cγ	0.5865
Cross-Link	Oδ-sp^3	–0.4639
Cross-Link	Oδ-sp^2	–0.5811
Cross-Link	OP	–0.7761
Cross-Link	P	1.1659
Cross-Link	C5′	–0.0069
Cross-Link	H5′	0.0754
Cross-Link	C4′	0.1536
Cross-Link	H4′	0.0843

Cross-Link	O4′	–0.4612
Cross-Link	C3’	0.4033
Cross-Link	H3′	0.0002
Cross-Link	O3′	–0.5232
Cross-Link	C2′	–0.4531
Cross-Link	H2′	0.1387
Cross-Link	C1′	0.4409
Cross-Link	H1′	0.1177
8oG	P	1.1659
8oG	C5'	0.0375
8oG	O5'	-0.5323
8oG	C4'	-0.0056
8oG	O4'	-0.3058
8oG	C3'	0.3691
8oG	O3'	-0.5470
8oG	C2'	-0.1054
8oG	C1'	0.2073
8oG	N1	-0.4376
8oG	C2	0.6542
8oG	N2	-0.8517
8oG	N3	-0.5066
8oG	C4	0.1209
8oG	C5	0.0649
8oG	C6	0.4688
8oG	O6	-0.5488
8oG	N7	-0.5408
8oG	C8	0.5044
8oG	O8	-0.5150
8oG	N9	-0.0236
8oG	OP1	-0.7761
8oG	OP2	-0.7761
8oG	H5'	0.0781
8oG	H5''	0.0781

8oG	H4'	0.0625
8oG	H3'	0.0044
8oG	H2'	0.0481
8oG	H2''	0.0481
8oG	H1'	0.0344
8oG	H22	0.3878
8oG	H21	0.3878
8oG	H1	0.3492
8oG	H7	0.4010

[a]8oG parameters were adapted from the literature.[1]

Table A.3. Average RMSD (Å) with standard deviation in parentheses for each MD simulation and across the complete data set for wild-type and mutant MutY.[a]

Model	Replicate 1	Replicate 2	Replicate 3	Replicate 4	Overall[d]
Wild-type RC[b]	2.5 (0.2)	2.6 (0.2)	2.6 (0.2)	2.5 (0.2)	2.6 (0.2)
Y126F RC[b]	3.0 (0.3)	3.5 (0.4)	3.3 (0.2)	3.3 (0.5)	3.2 (0.4)
N146S RC[b]	3.2 (0.3)	2.7 (0.3)	2.8 (0.2)	3.1 (0.2)	2.9 (0.3)
Wild-type IC_{CL}[c]	2.1 (0.2)	2.1 (0.2)	2.3 (0.2)	2.1 (0.2)	2.1 (0.2)
Y126F IC_{CL}[c]	2.3 (0.2)	2.7 (0.2)	3.2 (0.4)	3.3 (0.5)	2.9 (0.5)

[a]RMSD is relative to the first frame of the corresponding simulation. [b]Reactant complex. [c]DNA–protein cross-linked intermediate. [d]Calculated over all replicas.

Table A.4. Average key distances (Å), angles (°), and occupancies (occ., %) across four MD simulation replicas for the wild-type and mutant MutY RC.

Structural Feature	Wild-Type			Y126F			N146S		
R(D144$_{O\delta}$···C1')	3.7 ± 0.6			4.0 ± 0.9			5.8 ± 2.3		
∠(N9C1'D144$_{O\delta}$)	104 ± 19			112 ± 23			133 ± 25		
Reaction Alignment[a]	66			57			31		
Bridging Water Occ[b]	34			22			21		
Hydrogen Bond	Occ.[c]	Dist	Angle	Occ.[c]	Dist	Angle	Occ.[c]	Dist	Angle
E43···N7	2	2.8	164	3	2.8	157	3	2.8	165
Y126···E43	77	2.7	161	-	-	-	54	2.7	162
X146···D144 [d]	89	2.8	160	63	2.8	159	30	2.7	164
N146···Backbone[e]	10	2.8	157	9	2.8	158	-	-	-

[a]The active site was considered aligned for catalysis when R(D144$_{O\varepsilon}$···C1′) was less than 4 Å and ∠(N9C1′ D144$_{O\delta}$) was greater than 95˚. [b]A bridging water was considered present when a water molecule simultaneously hydrogen bonds to D144 and N7 of dA. [c]Residues were considered to be hydrogen bonded when the heavy atom distance was less than 3.0 Å and the hydrogen-bond angle fell between 135° and 225°. [d]X146=N146 for wild-type and Y126F MutY, and S146 for N146S. [e]Backbone refers to the 5′ phosphate group with respect to the adenine substrate.

Table A.5. Relative Gibbs energy (kJ/mol) for water-mediated protonation at N7 of dA for different QM/MM models.[a,b]

Model	RC	TS_{Bridge}	RC_{H+}
Snapshot 2	0.0	8.3	1.1
Snapshot 3	0.0	–3.0	–23.7
Snapshot 4	0.0	–2.0	–10.3
Average	0.0	1.1 ± 6.3	–11.0 ± 12.4

[a]Energies were calculated using ONIOM(M06-2X/6-311+G(2df,p):AMBERff14SB)//ONIOM(M06-2X/6-31G(d,p):AMBERff14SB) and the ME scheme. [b]Reaction is depicted in Figure S5.

Table A.6. Comparison of important distances (Å) for QM/MM optimized RC_{H+} complexes produced via direct or water-mediated protonation at N7 of dA.[a b]

Structural Parameter	Direct Protonation				Water-mediated Protonation				
	CS	Snapshot 1	Average[c]		Snapshot 2	Snapshot 3	Snapshot 4	Average[d]	
R(D144$_{O\delta}$···C1')	2.866	3.092	2.979	± 0.160	3.112	3.044	3.118	3.091	± 0.041
R(C1'–N9)	1.533	1.487	1.510	± 0.033	1.488	1.477	1.487	1.484	± 0.006
R(N7–H)	1.086	1.111	1.096	± 0.018	1.062	1.075	1.095	1.077	± 0.017

[a]Systems were optimized using ONIOM(M06-2X/6-31G(d,p):AMBERff14SB). [b]Structures depicted in Figures 3, S5 and S6. [c]Average of data from CS and MD snapshot 1 models. [d]Average of data from MD snapshot 2, 3, and 4 models.

Table A.7. Relative Gibbs energy (kJ/mol) barrier for dA deglycosylation and for different RC QM/MM models.[a,b]

Model	RC_{H+}	Rate Limiting TS[c]	IC_{CL}
Crystal Structure[d]	0.0	23.3	–67.8
Crystal Structure EE[d,e]	0.0	71.3	12.1
Snapshot 1	0.0	53.7	–66.9
Snapshot 2	0.0	47.4	–60.7
Snapshot 3	0.0	31.4	–68.7
Snapshot 4	0.0	49.3	–2.1
Average One-step[f]	0.0	39.4 ± 14.4	–51.4 ± 32.9
Overall Average[g]	0.0	41.2 ± 13.1	–53.0 ± 28.7

[a]Energies were obtained from ONIOM(M06-2X/6-311+G(2df,p):AMBERff14SB)//ONIOM(M06-2X/6-31G(d,p):AMBERff14SB). ME scheme was used unless otherwise indicated. [b]Reaction is depicted in Figures 3 and S6. [c]$TS_{Concerted}$ for CS, snapshot 1, snapshot 3 and snapshot 4 models. $TS_{Deglyco}$ for snapshot 2 model. [d]Model was initiated based on the FLRC crystal structure. Snapshot 2 yields a two-step mechanism shown in Figure 3B. [e]EE scheme used. [f]Average relative energy over all models that yield a one-step deglycosylation reaction using the ME scheme. [g]Average relative energy over all models that use the ME scheme.

Table A.8. Important average distances (Å), angles (°), and occupancies (occ., %) across four MD simulation replicas for the wild-type and mutant DNA–MutY IC_{CL}.

Structural Feature	Wild-type	Y126F
∠(C2'C1'$O_\delta C_\gamma$) Active Conformation Occ.[a]	81	19
E43 Active Conformation Occ.[b]	37	13
Aligned Active Site Occ.[c]	27	5
∠($E43_{O\varepsilon}$C1'$D144_{O\delta}$)[d]	133 ± 11	143 ± 11
Nucleophilic Water Occ.[e]	43	4
Altona–Sundarlingam Psuedorotation Angle	170 ± 31	187 ± 33

Hydrogen Bond	Occ.[f]	Dist	Angle	Occ.[f]	Dist	Angle
Y126···E43	98	2.6	164	X	X	X
N146···D144	8	2.9	153	6	2.9	157
N146···Backbone[g]	78	2.8	161	29	2.8	161

[a]In the active conformation, ∠(C2′C1′$O_\delta C_\gamma$) falls between 140° and 220°. [b]E43 was considered to be in the active conformation when R(C1′···E43) is less than 6.5 Å. [c]Active site is considered aligned for catalysis when both ∠(C2′C1′$O_\delta C_\gamma$) and E43 are in the active conformation. [d]Average angle present when E43 is in the active conformation. [e]A water nucleophile was considered present when within 3.5 Å of E43 and 4.5 Å of C1′ of deoxyribose. [f]Residues were considered to be hydrogen bonded when the heavy atom distance was less than 3.0 Å and the hydrogen-bond angle fell between 135° and 180°. [g]Backbone refers to the 5′ phosphate group with respect to the cross-link.

Table A.9. Relative Gibbs energy (kJ/mol) for DNA–MutY cross-link hydrolysis (enzyme regeneration) for different QM/MM models.[a,b]

Model	IC_{CL}	$TS_{W\text{-}Attack}$	PC
Snapshot 1	0.0	103.2	74.1
Snapshot 1 EE[c]	0.0	105.4	34.5
Snapshot 2	0.0	109.4	98.3
Snapshot 3	0.0	99.3	105.3
Snapshot 4	0.0	113.9	98.5
Average ME[d]	0.0	106.5 ± 6.5	94.1 ± 13.7

[a]Energies were calculated using ONIOM(M06-2X/6-311+G(2df,p):AMBERff14SB)//ONIOM(M06-2X/6-31G(d,p):AMBERff14SB). ME scheme was used unless otherwise indicated. [b]Reaction depicted in Figures 5 and S11. [c]EE scheme used. [d]Average relative energy over all models using the ME scheme.

Full Computational Methods

MD Model Building

Input models for MD simulations corresponding to the RC and IC_{CL} were generated based on crystal structures of the fluorine recognition complex (FLRC, PDB ID: 3G0Q)[2] and transition state analogue complex (TSAC, PDB ID: 6U7T),[3] respectively. The P164C mutation in the FLRC crystal structure was reverted to proline and 2′-β-fluoro-2′-deoxyadenosine was converted to dA using PyMOL.[4] E43 in the RC was assigned a neutral protonation state based on the relative arrangement of contact partners in the FLRC and LRC crystal structures,[3, 5] and previous PROPKA calculations.[1] In the TSAC model, the proposed DNA–protein cross-link was constructed in GaussView 6.0.16[6] by replacing azaribose with 2′-deoxyribose, and forming a covalent bond between C1′ and the O_δ of D144. The K230, K231, T232, A233, D288, N289, Y290, and G291 residues in the FLRC model, and the Y290, G291, L292, N293, and N294 residues in the TSAC model were manually added and adjusted using PyMOL to avoid steric clashes with the remainder of the DNA–enzyme complex. Additionally, R11 in the FLRC and R7, E52, K103, E116, R202, K228, K231, D277, K279, Q283, Q295, E298, R320, E328, E329, and D337 in the TSAC had partially unresolved side chains, which were completed in the models using the PyMOL builder tool[4] and then visually moved to prevent steric clashes. The N146S and Y126F MutY mutants were subsequently generated from the

previously built wild-type MutY models using PyMOL.[4] The mutation sites were visually inspected to eliminate close atomic contacts. Each structure was solvated with TIP3P water molecules in a periodic rectangular prism with sides a minimum of 10 Å from the solute in each direction using the LEaP program in AMBER18. In total, 14688 TIP3P water molecules were added in the FLRC models and 13258 water molecules included in the TSAC models.[7] Sodium ions were added to neutralize the system and NaCl was inserted to achieve physiological salt concentration (150 mM) as calculated using the SLTCAP calculator.[8] As a result, 52 sodium ions and 23 chloride ions were added to the FLRC models, and 50 sodium ions and 26 chloride ions were included in the TSAC models. The Amber force field was used throughout, with protein residues described using ff14SB parameters[9] and nucleic acid components modeled using OL15.[10] Parameters for the 8oG lesion were adopted from the literature,[1] while parameters for the sulfur-iron (SF4) cluster and the coordinating cysteine (CYF) residues were generated using the MCPB.py program (Tables S1, S2 and Figure S1), with the system optimized using HF/6-31G(d) as part of the protocol.[11] The sulfur iron cluster consisted of two Fe(II) and two Fe(III) ions, with an overall charge of −2. The Amber parameters for the cross-link were supplemented by the general AMBER force field (GAFF)[12] using Antechamber (Table S1 and Figure S1).[13] RESP charge fitting was used to generate charges with the R.E.D.v.III.4 program[14] (Table S2) using models built in GaussView 6.0.16[6] and optimized using HF/6-31G(d) to be consistent with the original Amber protocol.

MD Simulation Protocol

The FLRC and TSAC models were initially minimized in a stepwise manner, using 1000 steps of steepest decent followed by 1000 steps of conjugate gradient minimization. First, the solvent was minimized, while a 100 kcal mol^{-1} $Å^{-2}$ restraint was applied to the solute. Next, the restraint was removed from the solute hydrogen atoms. Subsequently, the solute was minimized, while applying a 100 kcal mol^{-1} $Å^{-2}$ restraint to the solvent. Finally, the entire system was minimized with no restraints, using 1000 steps of steepest decent followed by 2000 steps of conjugate gradient minimization. Each system was then heated from 10 K, with the temperature increased by 50 K

every 20 ps until 310 K. The restraints on the solute were subsequently dropped from 25 kcal mol^{-1} $Å^{-2}$ to 5 kcal mol^{-1} $Å^{-2}$ at a rate of 5 kcal mol^{-1} $Å^{-2}$ per 20 ps. A final 20 ps equilibration step was completed with a solute restraint of 1.5 kcal mol^{-1} $Å^{-2}$. For these simulations, the Langevin thermostat was used with a collision frequency of 1 ps^{-1} and an NVT ensemble. As an additional equilibration step for the TSAC model, a 100 ns MD simulation step was conducted in which a 10 kcal mol^{-1} $Å^{-2}$ force constant was applied to the DNA–protein cross-link to allow the remainder of the system to relax in response to the addition of the DNA–protein cross-link. Following minimization, heating, and equilibration, four 1 μs MD production simulations were performed on each system. The additional equilibration step and production simulations employed the Langevin thermostat with a collision frequency of 3 ps^{-1} and an NPT ensemble with the pressure set to 1 bar. Minimization, equilibration, and production simulations were all performed using AMBER 2018 pmemd.cuda.[7]

MD Analysis:

Trajectory analysis was conducted across all replicas using the cpptraj program in AMBER 2018.[7] Hydrogen bonds were considered to be present between two residues if the heavy atom distance was less than 3.0 Å and the hydrogen-bonding angle fell between 135° and 180°. A water was considered to bridge two residues when the water simultaneously forms a hydrogen bond to both residues. D144 was considered to be in a catalytically conducive position for nucleophilic attack at C1′ of dA in the reactant complex when the $D144_{O\delta}\cdots C1'$ distance was less than 4.0 Å and the ∠($N9C1'D144_{O\delta}$) was greater than 95˚. A water nucleophile was deemed aligned for hydrolysis of the cross-linked intermediate when located within 3.5 Å of E43 and 4.5 Å of C1′ of deoxyribose. Representative structures for figures were obtained by clustering across entire data sets using the cpptraj program and the heiragglo algorithm. The representative structures of the FLRC with a direct E43···N7 hydrogen bond or a bridging water (Figure 2) were obtained by clustering the wild-type FLRC MD simulation trajectories based on the minimum E43···N7 distance. The representative structure of the TSAC conformation that was aligned for hydrolysis (Figure 4)

was obtained by first clustering based on the ∠(C2′C1′D144$_{O\delta}$D144$_{C\gamma}$) dihedral angle and subsequently the E43–C1′ distance. The representative structure of the reactant complex for the Y126F mutant aligned for hydrolysis (Figure 6A) was obtained by clustering the MD simulation trajectories based on the D144–C1′ distance. The representative structure of the intermediate complex for the Y126F mutant in the inactive conformation (Figure 6B) was obtained by clustering the MD simulation trajectories based on the RMSD of the active site residues (E43, F126, D144, N146 and the cross-linking deoxyribose). The representative structures for the reactant complex for the N146S mutant in the active (Figure 8A) and inactive conformations (Figure 8B) were obtained by clustering the MD simulation trajectories based on the D144–C1′ distance. The representative structures of the three cross-link conformations of the wild-type (Figure S10) and Y126F (Figure S15) MutY intermediate complexes were obtained by clustering the MD simulation trajectories based on the ∠(C2′C1′D144$_{O\delta}$D144$_{C\gamma}$) dihedral angle. The representative structures of the intermediate complexes with active and inactive conformations of E43 (Figure S8) were obtained by clustering the MD simulation trajectories of the wild-type TSAC models based on the E43–C1′ distance.

QM/MM Methodology

Due to the high similarity of the MD replica trajectories, QM/MM models of the DNA–protein complex were built from 10 snapshots, collected at intervals of 4 ns from the beginning of one trajectory corresponding to each of the wild-type MutY reactant complex (FLRC model) and the cross-linked intermediate (TSAC model). Each model was analyzed to determine whether key residues adopt structural criteria necessary to afford a catalytically conducive complex. Specifically, a direct or water-mediated hydrogen bond between the E43 side chain and N7 of dA is required for the FLRC model, while the distance between E43 and C1′ of less than 5.5 Å is required in the cross-linked intermediate (TSAC model). Models that did not satisfy these criteria were discarded. In total, 4 MD snapshots were identified for QM/MM reaction pathway mapping for each catalytic stage. In addition, one QM/MM model was built from the solvated FLRC crystal structure as

common modeling practice in the literature.[15-19] From the wild-type MutY models, Y126F MutY mutant models were constructed by mutating Y126 from the wild-type models initiated from the FLRC crystal structure and the TSAC snapshot 1.

The QM region of the reactant complex (FLRC model) contains 134–137 atoms, with an overall charge of –4. Specifically, the high-level layer contains E43 truncated at the Cα–Cβ bond, Y126 truncated at the Cα–Cβ bond, D144 truncated at the Cα–C and Cα–N bonds, N146 truncated at the Cα–C bond, V147 truncated at the Cα–N bond, dA truncated at the C–O bonds of C3′ and C5′, R31 truncated at the Cγ–Cδ bond, E188 truncated at the Cα–Cβ bond, E192 truncated at the Cα–Cβ bond, and a water molecule hydrogen bonded to E188 (Figure S2). Three models also contain a water molecule that bridges E43 and dA in the QM layer. The QM layer for the cross-linked intermediate (TSAC model) contains 111–114 atoms, with an overall charge of –3. Specifically, the high-level layer in the intermediate complex contains E43 truncated at the Cα–C bond, Y126 truncated at Cα–Cβ bond, D144 truncated at the Cα–C and Cα–N bonds, N146 truncated at the Cα–C bond, V147 truncated at the Cα–N bond, the cross-linked substrate truncated at the C–O bonds of C3′ or C5′, and S42 truncated at the Cα–N bond (Figure S2). The nucleophilic water and 3 additional water molecules surrounding E43 were also included in the QM region. Three models contained an additional water molecule in the QM region near E43 to maintain the E43 conformation with respect to the cross-link as characterized by MD simulations. The MM region for all models includes the remainder of the enzyme and DNA substrate as well as the closest 1300 water molecules to the solute, which results in an approximately 4 Å deep layer of water.

All QM/MM models were optimized using ONIOM(M06-2X/6-31G(d,p):AMBERff14SB) and the mechanical embedding (ME) scheme. Transition state structures were isolated by scanning relevant bond distances along the reaction pathway to yield an initial guess and subsequently performing unrestrained transition state optimizations. Specifically, the C1′–N9 bond distance was scanned in the FLRC models while the C1′–WAT_O distance was scanned in the TSAC models. Frequency calculations were conducted to characterize the nature of stationary points and verify that the isolated imaginary frequency was associated with the anticipated transition structure. Each

pathway was verified to be continuous by mapping the internal reaction coordinate through scanning the nucleophile–electrophile and leaving group distance both to and from each stationary point. The reported reaction Gibbs energies were obtained from ONIOM(M06-2X/6-311+G(2df,p):AMBERff14SB) single-point calculations.

The ME scheme was used throughout due to the robustness of this approach, the number of models and pathways considered in the present work, and previous successes using the same methodology for various enzymatic reactions.[20-24] To verify that the ME scheme accurately models MutY, both the deglycosylation and enzyme regeneration mechanistic pathways were mapped using the electronic embedding (EE) scheme (Figure S16, Tables S7 and S9). Due to the similar structural and energetic characteristics of the mechanistic pathways modeled using ME, only one starting structure for each of the FLRC and TSAC models was used for the EE calculations. For the deglycosylation reaction step, the FLRC crystal structure was used as the starting point for EE calculations as typically done in the literature. When modeled with EE, the RC contains the same direct hydrogen bond between the E43 residue and the N7 of adenine as observed with ME optimizations. Although E43 retains the proton in the EE derived RC (Figure S16), the proton is transferred concurrently with deglycosylation along the EE characterized pathway, with the proton being fully transferred in the transition state. We note that differences in the proton location with ME and EE have been previously reported for other DNA repair enzymes.[25] The EE deglycosylation transition state is later than that obtained using the ME scheme, with D144 being further along in attack and the glycosidic bond being stretched by an additional 0.3–0.4 Å, although the reaction still falls between A_ND_N and $D_N{*}A_N$ mechanisms (Figure S16). This results in an ~50 kJ/mol increase in the EE barrier compared to the corresponding ME model. However, the barrier to cross-link formation is still well below the calculated barrier for direct hydrolysis.[1] The cross-link hydrolysis step was modeled using EE based on snapshot 1 from the MD simulations on the TSAC model and the resulting pathway closely matches the ME mechanism. All relevant structural features are very similar, with the C1′−nucleophile, C1′−leaving group, and proton transfer distances being within 0.2 Å of each other in the ME and EE models (Figure S16). More importantly, cross-link hydrolysis remains the rate determining step for the reaction. Moreover, the overall energy barrier

calculated with EE (105.4 kJ/mol) is within 3 kJ/mol of that predicted using ME. Thus, the MutY mechanisms modeled with EE and ME have the same overall pathway, similar structural features, and nearly identical rate-limiting barriers, which validates the use of ME for a greater range of models in the main text to permit analysis of averaged data from different reactant complexes.

References:

(1) 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* (48), 8753-8765. DOI: 10.1021/bi401310w

(2) 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* (44), 18497-18502, S18497/18491-S18497/18495. DOI: 10.1073/pnas.0902908106

(3) Woods, R. D.; O'Shea, V. L.; Chu, A.; Cao, S.; Richards, J. L.; Horvath, M. P.; David, S. S. Structure and Stereochemistry of the Base Excision Repair Glycosylase MutY Reveal a Mechanism Similar to Retaining Glycosidases. *Nucleic Acids Res.* **2016**, *44* (2), 801-810. DOI: 10.1093/nar/gkv1469

(4) Schrodinger, LLC. The Pymol Molecular Graphics System, Version 1.8. 2015.

(5) Fromme, J. C.; Banerjee, A.; Huang, S. J.; Verdine, G. L. Structural Basis for Removal of Adenine Mispaired with 8-Oxoguanine by MutY Adenine DNA Glycosylase. *Nature* **2004**, *427*, 652-656. DOI: 10.1038/nature02306

(6) Roy Dennington, T. A. K., John M. Millam. Gaussview Version 6.0.16. Semichem Inc. Shawnee Mission KS, 2019.

(7) Case, D. A.; Ben-Shalom, I. Y.; et al. AMBER 2018. **2018**, University of California, San Fransisco.

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

(9) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14sb: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99sb. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696-3713. DOI: 10.1021/acs.jctc.5b00255

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

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

(12) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *26* (1), 114. DOI: 10.1002/jcc.20145

(13) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* **2006**, *25* (2), 247-260. DOI: 10.1016/j.jmgm.2005.12.005

(14) Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P. The R.E.D. Tools: Advances in Resp and Esp Charge Derivation and Force Field Library Building. *Phys. Chem. Chem. Phys.* **2010**, *12* (28), 7821-7839. DOI: 10.1039/c0cp00111b

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

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

(17) Kreppel, A.; Blank, I. D.; Ochsenfeld, C. Base-Independent DNA Base-Excision Repair of 8-Oxoguanine. *J. Am. Chem. Soc.* **2018**, *140* (13), 4522-4526. DOI: 10.1021/jacs.7b11254

(18) Sadeghian, K.; Ochsenfeld, C. Unraveling the Base Excision Repair Mechanism of Human DNA Glycosylase. *J. Am. Chem. Soc.* **2015**, *137* (31), 9824-9831. DOI: 10.1021/jacs.5b01449

(19) Naydenova, E.; Roßbach, S.; Ochsenfeld, C. QM/MM Study of the Uracil DNA Glycosylase Reaction Mechanism: A Competition between Asp145 and His148. *J. Chem. Theory Comput.* **2019**, *15* (8), 4344-4350. DOI: 10.1021/acs.jctc.8b01305

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

(21) Pilbák, S.; Farkas, Ö.; Poppe, L. Mechanism of the Tyrosine Ammonia Lyase Reaction—Tandem Nucleophilic and Electrophilic Enhancement by a Proton Transfer. *Chem. Eur. J.* **2012**, *18* (25), 7793-7802. DOI: 10.1002/chem.201103662

(22) Hu, X.; Hu, H.; Melvin, J. A.; Clancy, K. W.; McCafferty, D. G.; Yang, W. Autocatalytic Intramolecular Isopeptide Bond Formation in Gram-Positive Bacterial Pili: A QM/MM Simulation. *J. Am. Chem. Soc.* **2011**, *133* (3), 478-485. DOI: 10.1021/ja107513t

(23) Wu, S.; Xu, D.; Guo, H. QM/MM Studies of Monozinc B-Lactamase CphA Suggest That the Crystal Structure of an Enzyme−Intermediate Complex Represents a Minor Pathway. *J. Am. Chem. Soc.* **2010**, *132* (51), 17986-17988. DOI: 10.1021/ja104241g

(24) Stevens, D. R.; Hammes-Schiffer, S. Exploring the Role of the Third Active Site Metal Ion in DNA Polymerase H with QM/MM Free Energy Simulations. *J. Am. Chem. Soc.* **2018**, *140* (28), 8965-8969. DOI: 10.1021/jacs.8b05177

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

Appendix B

Supplementary Information for Chapter 3: A Tale of Two Mechanisms: Clarification of the Pathway for MBD4 Catalyzed Glycosidic Bond Cleavage Using MD and QM/MM Calculations

Contains Figures B.1−B.11 and Tables B.1−B.3

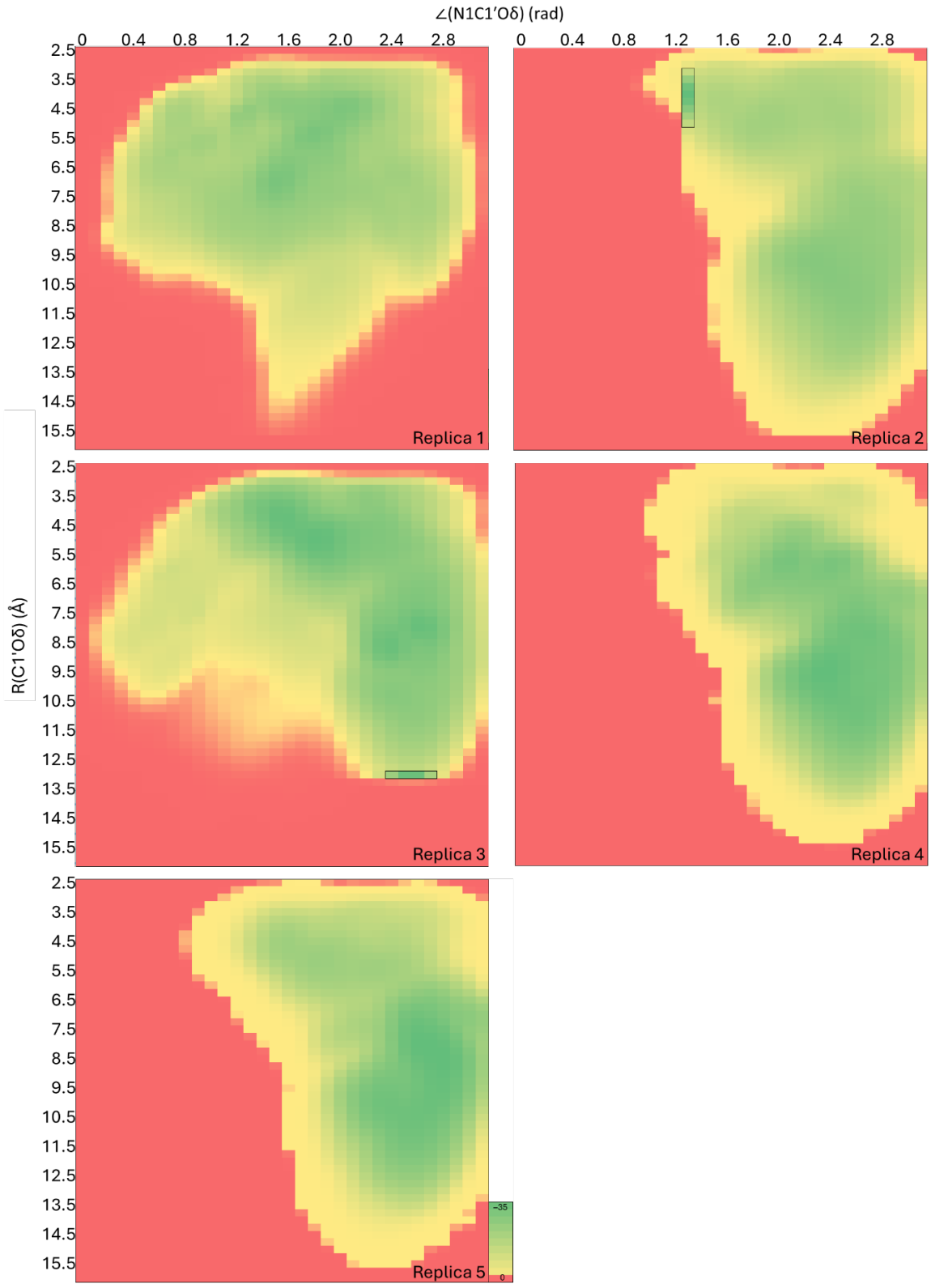


Figure B.1. PES generated from the individual abMD replicates. The two anomalous regions that were not included in the average PES are boxed (replicas 2 and 3). The legend is in kJ/mol.

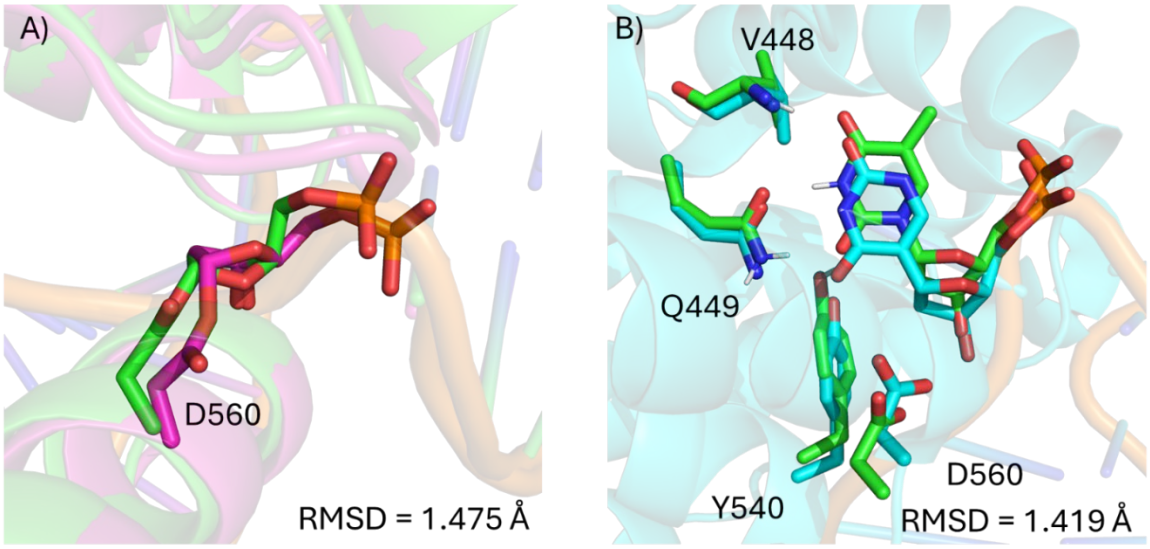


Figure B.2. Overlays of A) MD representative structure of the human MBD4 crosslinked intermediate (green) and the crystal structure of the mouse MBD4 crosslinked intermediate (PDB ID: 4EW4, magenta), and B) abMD representative structure (below conformation, green) and the crystal structure of MBD4 bound to pseudouridine-containing DNA (PDB ID: 7KZ0, cyan).

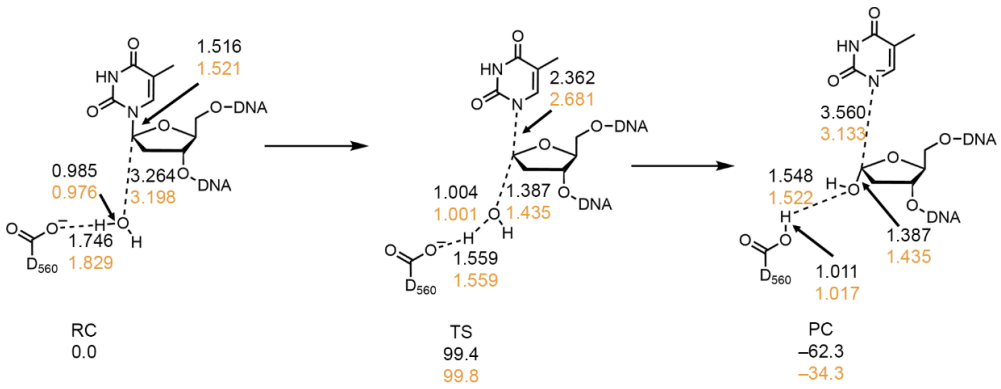


Figure B.3. One-step QM/MM hydrolysis mechanism mapped from the crystal structure of human MBD4 bound to pseudouridine-containing DNA (PDB ID: 7KZ0) using the ME (black) and EE (orange) schemes. Distances are in Å and relative energies (below) are in kJ/mol.

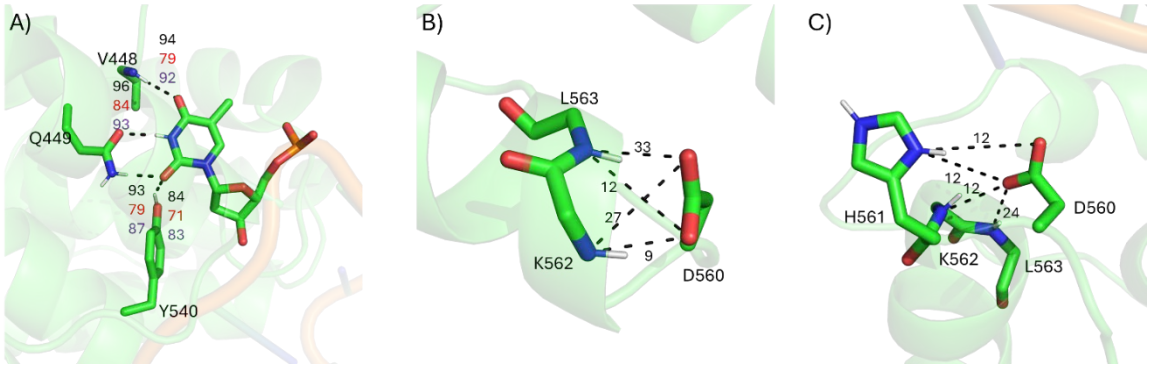


Figure B.4. A) Hydrogen-bond occupancies between MBD4 active site residues and the substrate thymine from abMD simulations for the D560 below conformation (D560 positioned on the opposite side of the deoxyribose as T; purple), the side conformation (D560 in the plane of the deoxyribose ring; red), and the inactive conformation (D560 directed away from T; black). Hydrogen-bond occupancies between D560 and nearby MBD4 active site residues from abMD simulations for the B) side and C) below conformations. Occupancies are percentages.

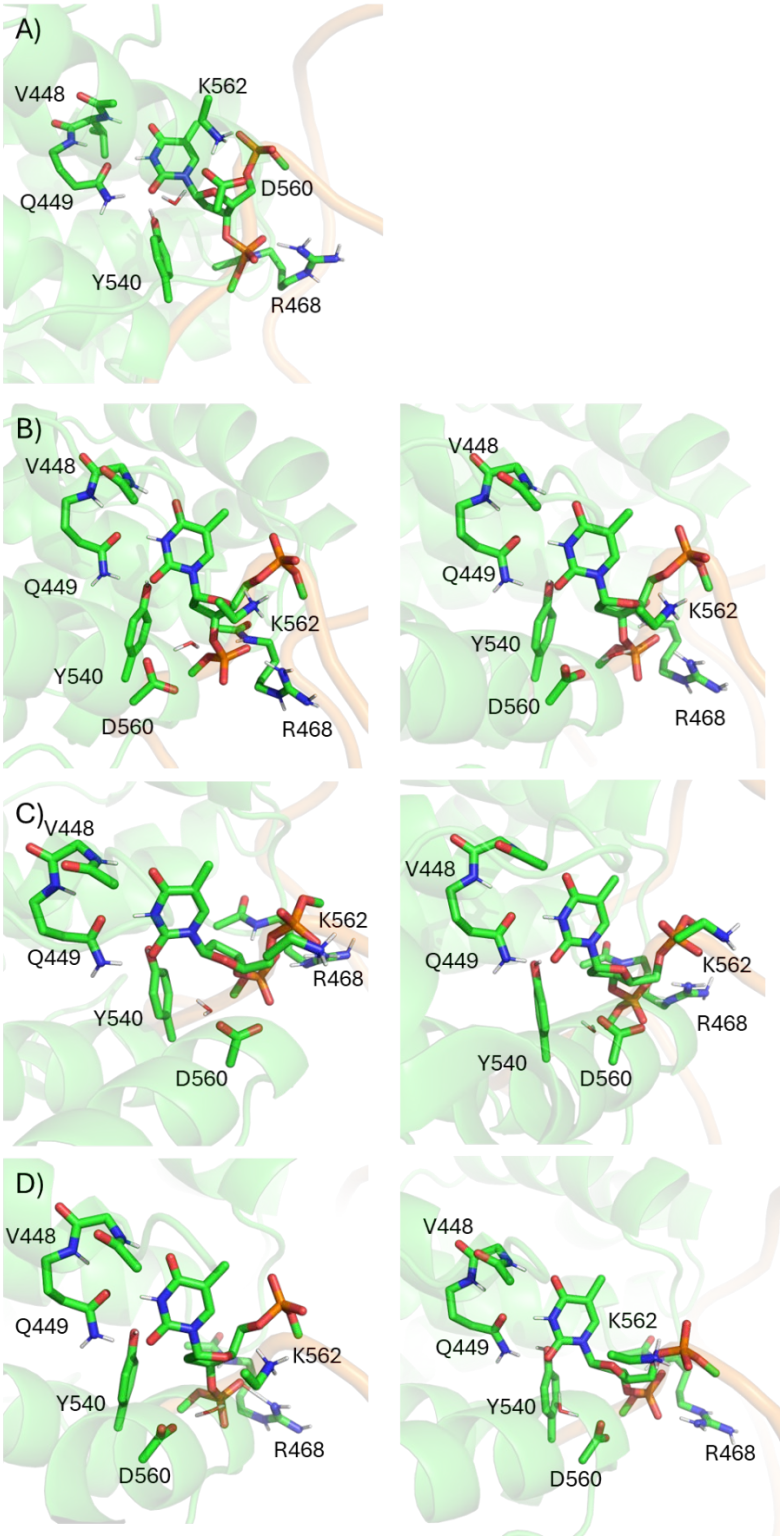


Figure B.5. QM/MM optimized reactant complexes of the D560 A) side conformation (D560 in the plane of the deoxyribose ring) and below conformation (D560 on the opposite side of deoxyribose as T) from the B) minimum, C) 1 Å closer, and D) 2 Å closer starting points when mapping the direct hydrolysis (left) and crosslinking (right) mechanisms.

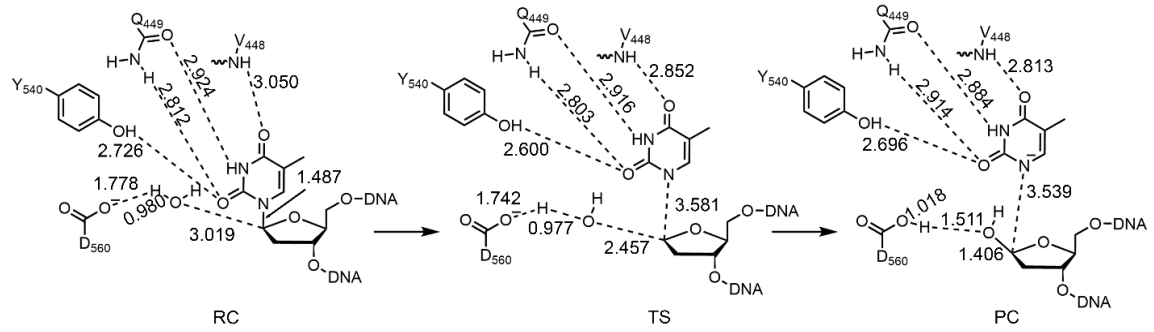


Figure B.6. MBD4 hydrolysis mechanism characterized from the D560 side conformation (D560 in the plane of the deoxyribose ring). Hydrogen-bond distances (Å) to the thymine base are heavy atom−heavy atom distances.

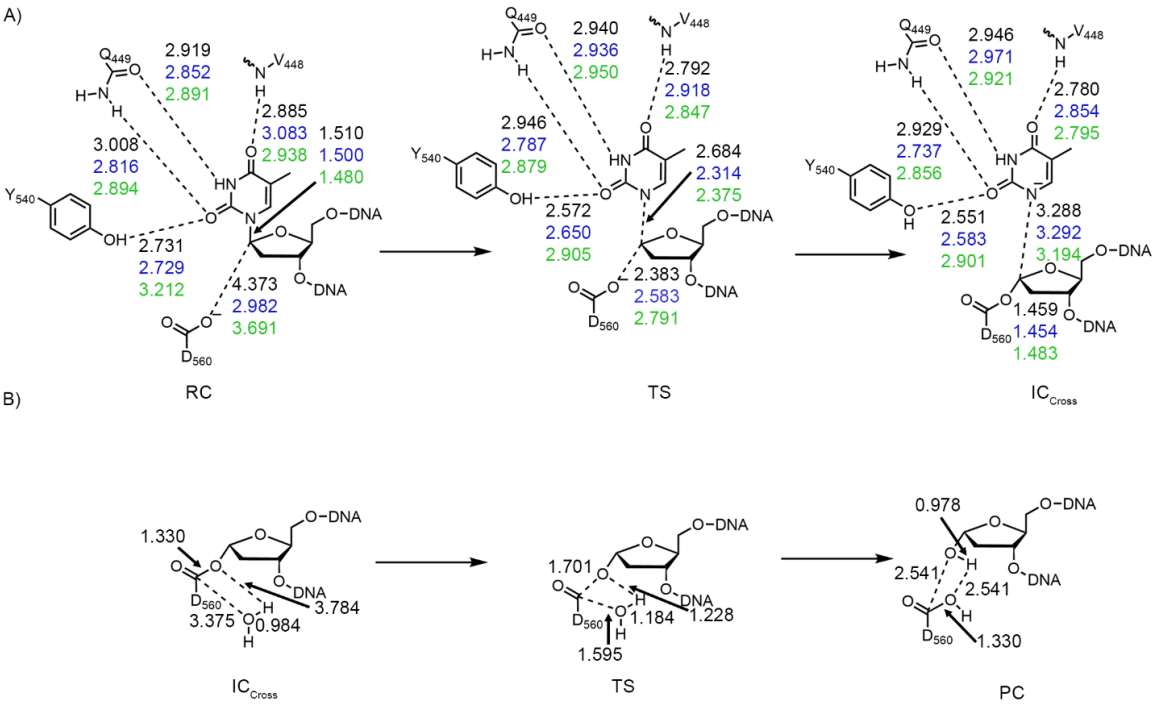


Figure B.7. A) MBD4 crosslinking mechanism characterized from the D560 below conformation (D560 on the opposite side of the deoxyribose as T) for the minimum (black), 1 Å closer (blue), and 2 Å closer (green) starting points. B) Crosslink saponification mechanism. Hydrogen-bond distances to the thymine base (Å) are heavy atom−heavy atom distances.

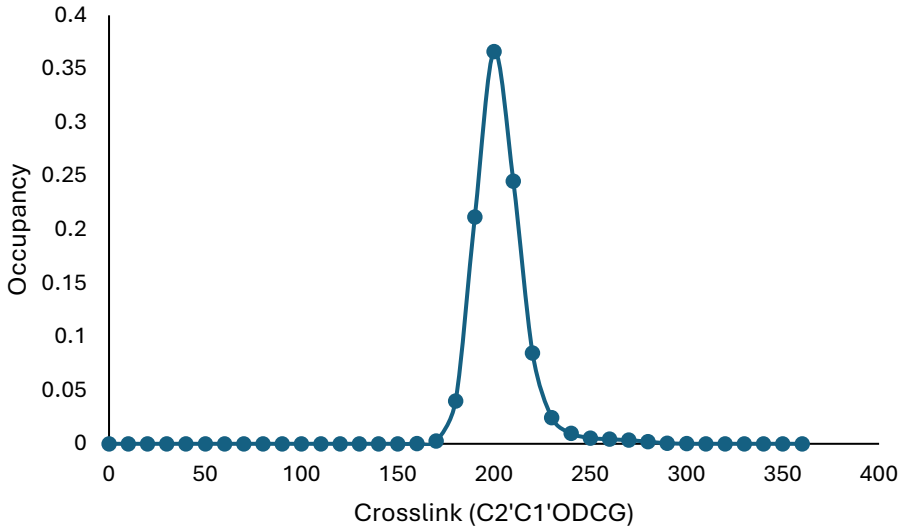


Figure B.8. Histogram of ∠(C2′C1′OδCγ) in the DNA−protein crosslink from classical MD simulations on the MBD4 crosslinked intermediate.

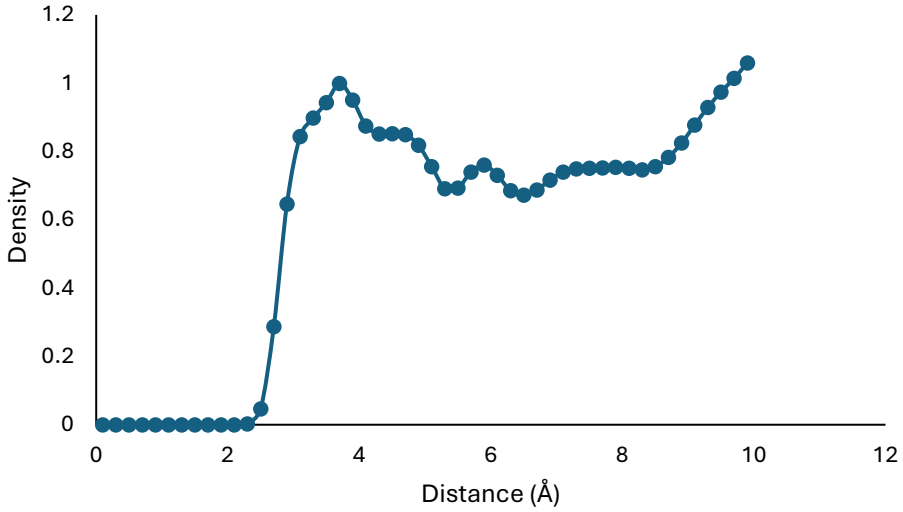


Figure B.9. Water radial density distribution around D560(Cγ) from classical MD simulations on the MBD4 crosslinked intermediate.

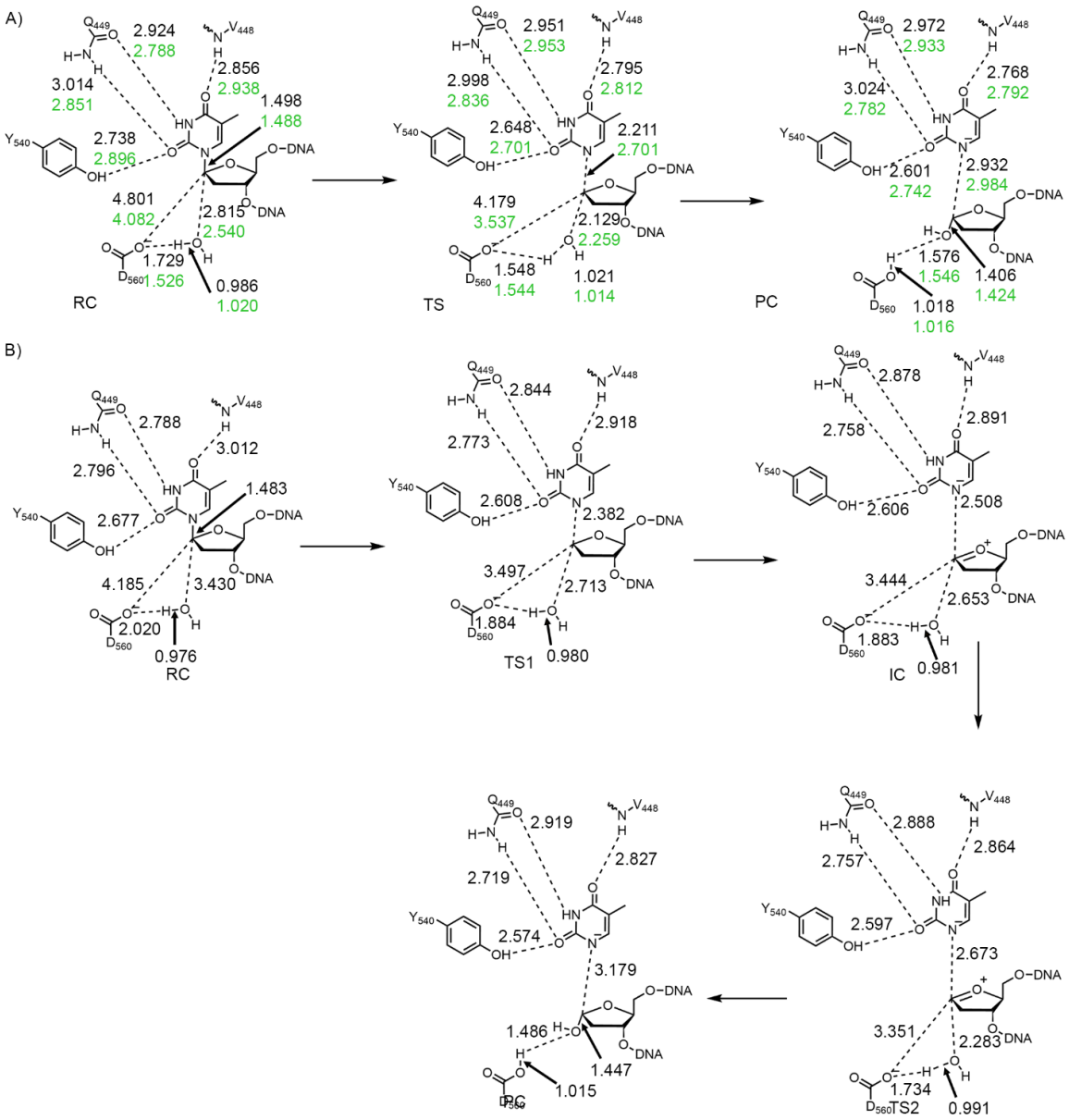


Figure B.10. MBD4 direct hydrolysis mechanism characterized from the D560 below conformation (D560 on the opposite side of deoxyribose as T) using A) the minimum on the abMD PES (black) and the point with D560 2 Å (green) or B) 1 Å closer to C1′ of the substrate. Distances in Å. Hydrogen-bond distances to the thymine base are heavy atom−heavy atom distances.

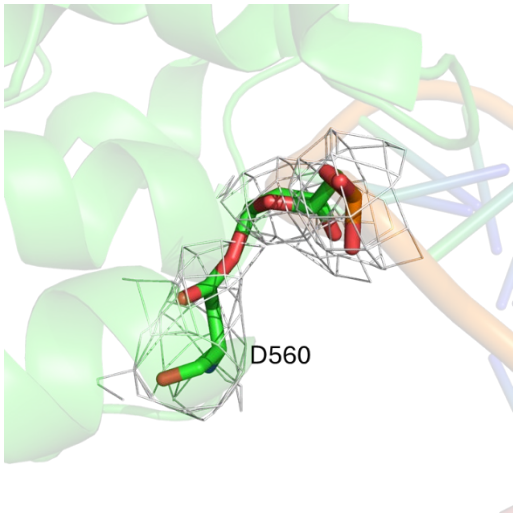


Figure B.11. Electron density map for the crystal structure of the mouse MBD4 crosslinked intermediate (PDB ID: 4EW4), highlighting the low electron density between D560 and deoxyribose.

Table B.1. Criteria for the abMD frames used to analyze structural features of each MBD4 conformation.

Model	r(C1′Oδ) (Å)	∠(N1C1′Oδ) (rad)
Side Conformation Minimum	5.00 – 5.75	1.8 – 2.1
Below Conformation Minimum	5.25 – 6.00	2.3 – 2.6
Inactive Conformation Minimum	7.75 – 8.50	2.5 – 2.8

Table B.2. Selection criteria used to cluster abMD frames[a] to obtain representative structures for each MBD4 conformation as QM/MM starting points.

Model	r(C1′Oδ) (Å)	∠(N1C1′Oδ) (rad)
Side Conformation Minimum	5.25 – 5.50	1.9 – 2.0
Below Conformation Minimum	5.50 – 5.75	2.4 – 2.5
Below Conformation 1 Å Closer	4.50 – 4.75	2.4 – 2.5
Below Conformation 2 Å Closer	3.50 – 3.75	2.4 – 2.5
Inactive Conformation Minimum	8.00 – 8.25	2.6 – 2.7

[a] Frames containing the R468−DNA backbone interaction observed in the crystal structure of MBD4 bound to pseudouridine-containing DNA (PDB ID: 7KZ0) were used for clustering.

Table B.3. Minima on abMD PES corresponding to each MBD4 conformation.[a]

Model	r(C1′Oδ) (Å)	∠(N1C1′Oδ)	
		(rad)	(°)
Side Conformation Minimum	5.25	1.9	109
Below Conformation Minimum	5.50	2.4	138
Inactive Conformation Minimum	8.00	2.6	149

[a]See Figure 3.4 in main text for PES and structures of minima.

Appendix C

Supplementary Information for Chapter 4: Decoding the Catalytic Strategy of AlkA: Insights from MD Simulations and QM/MM Calculations

Contains Figures C.1−C.5 and Tables C.1−C.3

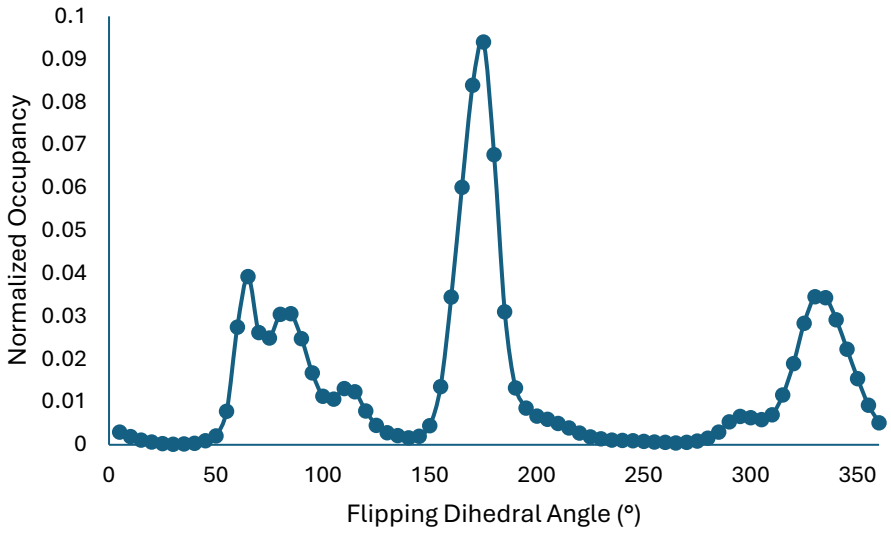


Figure C.1. Histogram of the dihedral angle formed between N1 and C4′ of 3mA and C4′ and N3 of the 3′−T with regards to the damaged site used to monitor 3mA binding in the AlkA active site over MD simulations.

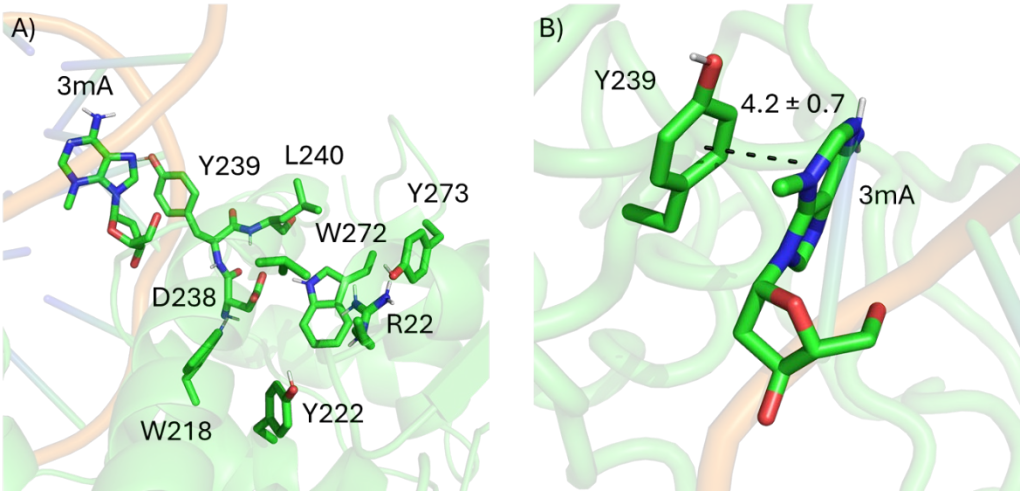


Figure C.2. MD representative structure of A) 3mA flipped out of the AlkA active site and B) the Y239 conformation that permits cation−π interactions with the 3mA nucleobase.

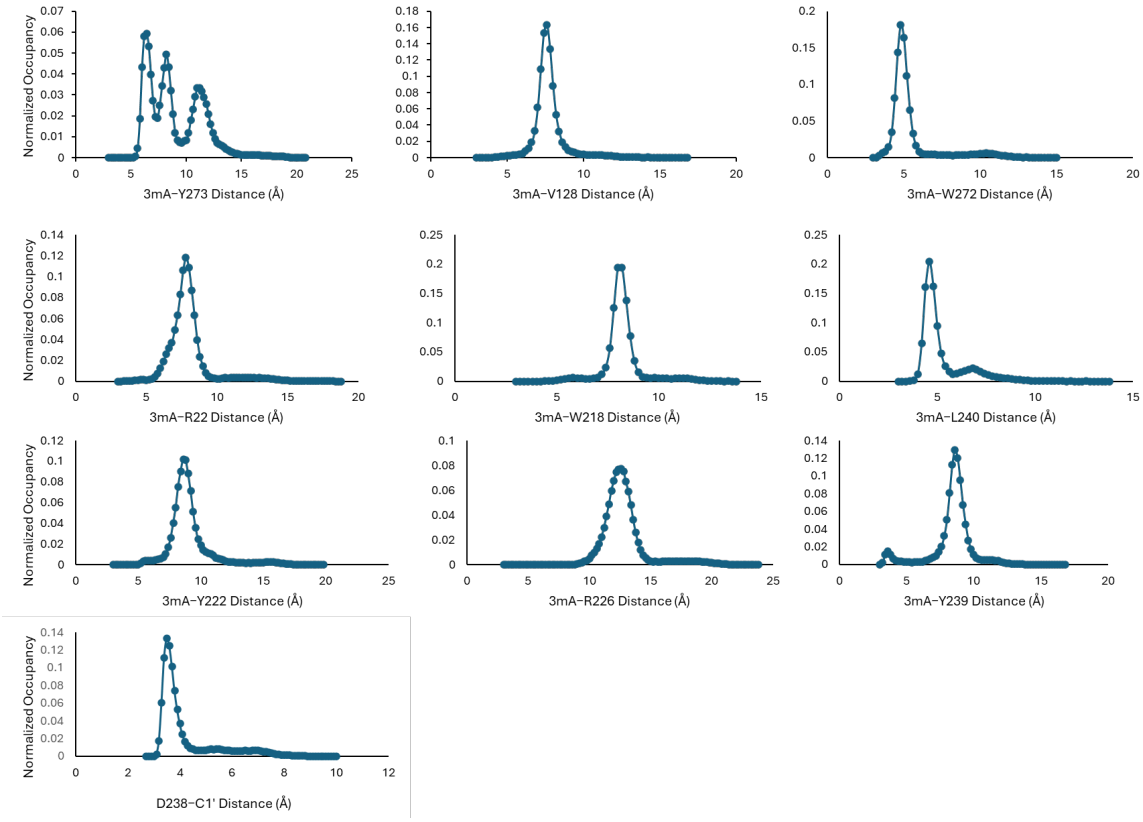


Figure C.3. Distance between 3mA and different binding pocket residues over MD simulations on the AlkA bound to 3mA-containing DNA.

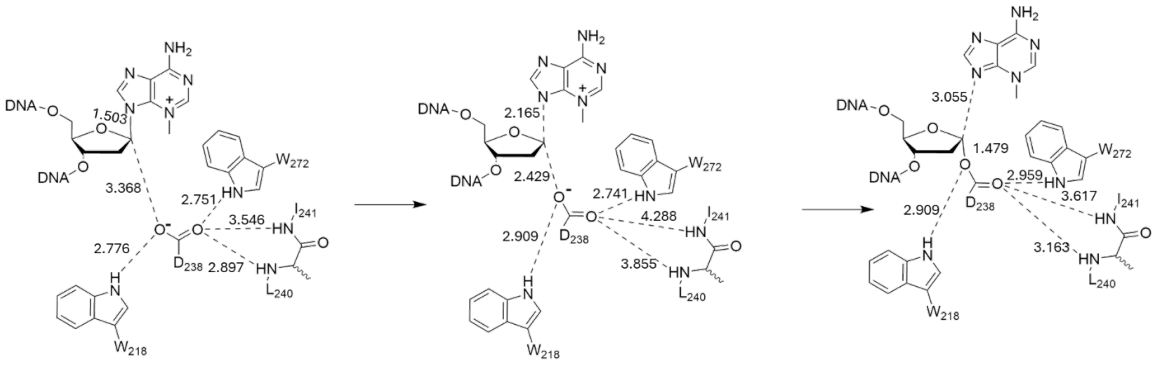


Figure C.4. QM/MM AlkA crosslinking mechanism characterized from an MD representative structure. Distances in Å. Hydrogen-bonding distances are heavy atom−heavy atom distances.

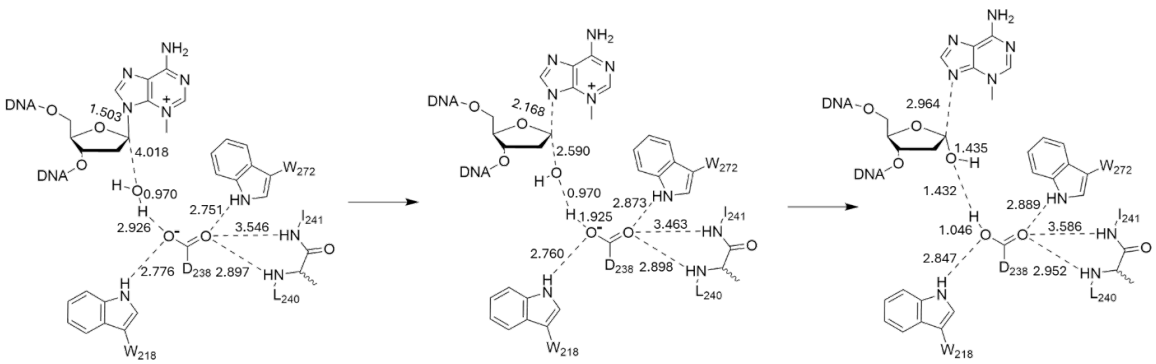


Figure C.5. QM/MM AlkA direct hydrolysis mechanism characterized from an MD representative structure. Distances in Å. Hydrogen-bonding distances not to a water molecule are heavy atom−heavy atom distances.

Table C.1. Amber atom types and charges assigned to 3mA during MD simulations.[a]

Name	Atom Type	Charge	Name	Atom Type	Charge
O3	OS	−0.5160	N7	NB	−0.5529
P	P	1.2120	C5	CB	0.1592
OP1	O2	−0.7920	C6	CA	0.6290
OP2	O2	−0.7920	N6	N2	−0.7729
O5′	OS	−0.4875	H61	H	0.4158
C5′	CJ	0.0096	H62	H	0.4158
H5′	H1	0.0885	N1	NC	−0.5714
H5″	H1	0.0885	C2	CQ	0.2546
C4′	CT	0.0884	H2	H5	0.1727
H4′	H1	0.0882	N3	na[b]	−0.0148
O4′	OS	−0.2948	C3	CT	−0.1518
C1′	CT	−0.0057	H31	H1	0.1142
H1′	H2	0.1530	H32	H1	0.1142
N9	N*	−0.0180	H33	H1	0.1142
C8	C2	0.2062	C4	CB	0.0967
H8	H5	0.1852	C3′	C7	0.1646
C2′	CT	−0.0372	H3′	H1	0.0882
H2′	HC	0.0741	H2″	HC	0.0741

[a]Atom types are from the Amber OL13 force field unless otherwise specified. [b]General amber force field atom type.

Table C.2. Amber non-standard parameters for 3mA.

Bonding				
Bond	Force Constant (kcal mol^{-} Å^{-2})		Equilibrium Value (Å)	
CQ-na	425.8		1.38	
CB-na	420.5		1.384	
CT-na	327.7		1.463	
Angles				
Angle	Force Constant (kcal mol^{-1} rad^{-2})		Equilibrium Value (°)	
N*-CB-na	69.8		127.09	
CB-CB-na	69.1		118.34	
NC-CQ-na	70.6		121.95	
CB-na-CQ	67.4		113.15	
CQ-na-CT	61.9		126.46	
H5-CQ-na	49.6		121.55	
H1-CT-na	49.8		108.78	
CB-na-CT	62.3		124.36	
Dihedral Angles				
Dihedral	Divider	Barrier (kcal/mol)	Phase (°)	Periodicity
na-CB-N*-C2	2	6.6	180	2
na-CB-N*-CA	2	9.6	180	2
NB-CB-CB-na	4	14.5	180	2
CA-CB-CB-na	4	14.5	180	2
na-CQ-NC-CA	2	9.5	180	2
NC-CQ-na-CB	4	6.8	180	2
NC-CQ-na-CT	4	6.8	180	2
N*-CB-na-CQ	4	1.2	180	2
CB-CB-na-CQ	4	1.2	180	2
H1-CT-na-CQ	6	0	0	2
H5-CQ-na-CB	4	6.8	180	2
H5-CQ-na-CT	4	6.8	180	2
H1-CT-na-CB	6	0	0	2
N*-CB-na-CT	4	1.2	180	2

CB-CB-na-CT	4	1.2	180	2

Improper Dihedral Angles			
Improper Dihedral	Barrier (kcal/mol)	Phase (°)	Periodicity
H5-NC-CQ-na	1.1	180	2
CB-CQ-na-CT	1.1	180	2
CB-N*-CB-na	1.1	180	2

Table C.3. QM/MM energy barrier for AlkA-catalyzed direct hydrolysis of the 3mA glycosidic bond to investigate the roles of W272 or W218.

Residue of Interest	Energy Barrier when moved to MM layer[a]	Energy Barrier When Mutated to Alanine[a]
W218	101.8 kJ/mol	94.1 kJ/mol
W272	88.8 kJ/mol	88.5 kJ/mol

[a]Relative energies obtained by mutating W218 or W272 to alanine in the QM/MM optimized RC and TS, optimizing the position of the added hydrogen atom while the rest of the system is fixed, and subsequently performing a single-point calculation. Optimizations and single-points used the same level of theory as calculations on the wild-type.

Appendix D

Supplementary Information for Chapter 5: Insights Into the Debated Lyase Mechanism of Bifunctional DNA Glycosylases From QM/MM MD Simulations: The Case Study of DNA Oxidative Damage Repair by hOGG1

Contains Tables D.1−D.16 and Figures D.1−D.9

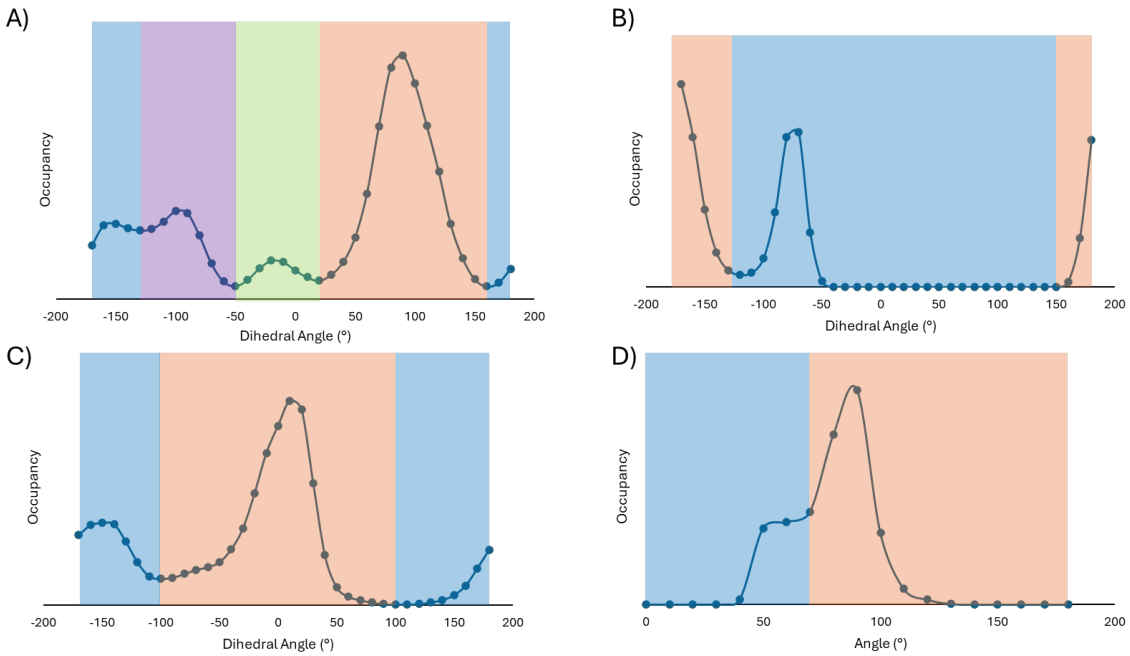


Figure D.1. Histograms of A) ∠(C3′C2′C1′Nζ), B) ∠(C4′C3′C2′C1′), C) ∠(D268CγC1′C2′Nζ), and D) ∠(D268CγC1′C2′) across MD simulations on hOGG1 bound to a cationic imine crosslink. The first (orange), second (blue), third (purple), and fourth (green) most occupied conformations are highlighted when applicable.

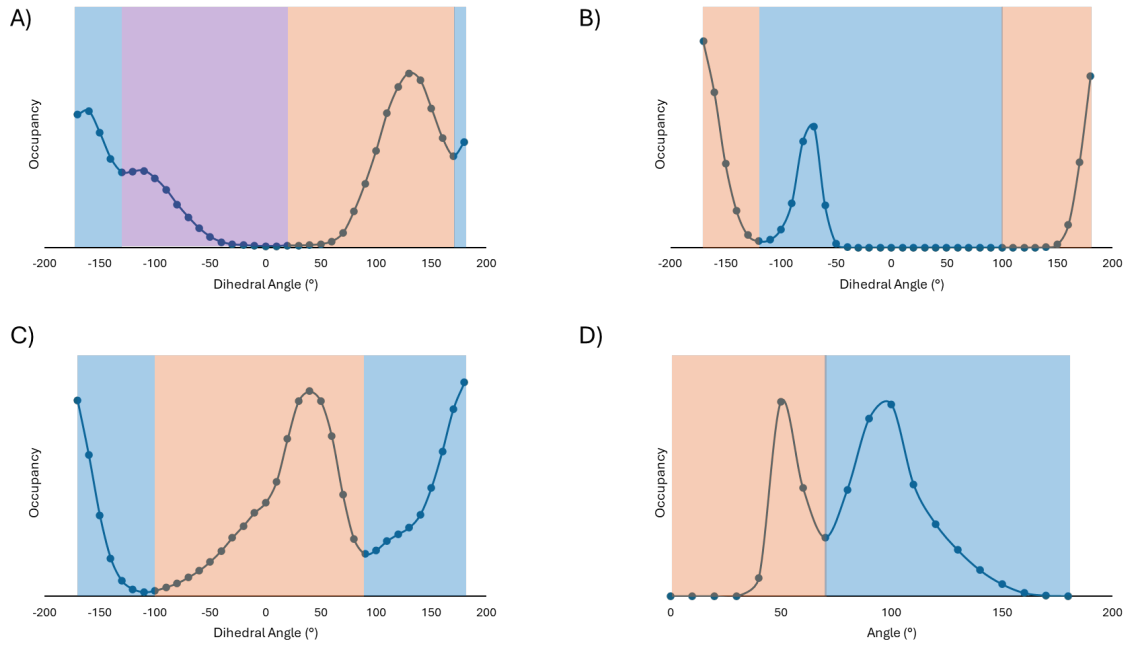


Figure D.2. Histograms of A) ∠(C3′C2′C1′Nζ), B) ∠(C4′C3′C2′C1′), C) ∠(D268CγC1′C2′Nζ), and D) ∠(D268CγC1′C2′) across MD simulations on hOGG1 bound to the neutral imine crosslink. The first (orange), second (blue), and third (purple) most occupied conformations are highlighted when applicable.

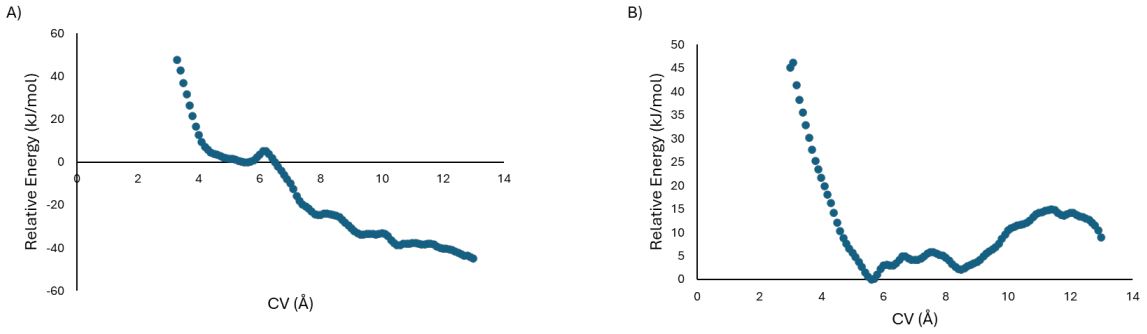


Figure D.3. Potential energy surface from umbrella sampling MD simulations, mapping 8oG unbinding from the hOGG1 active site containing a A) cationic or B) neutral imine crosslink.

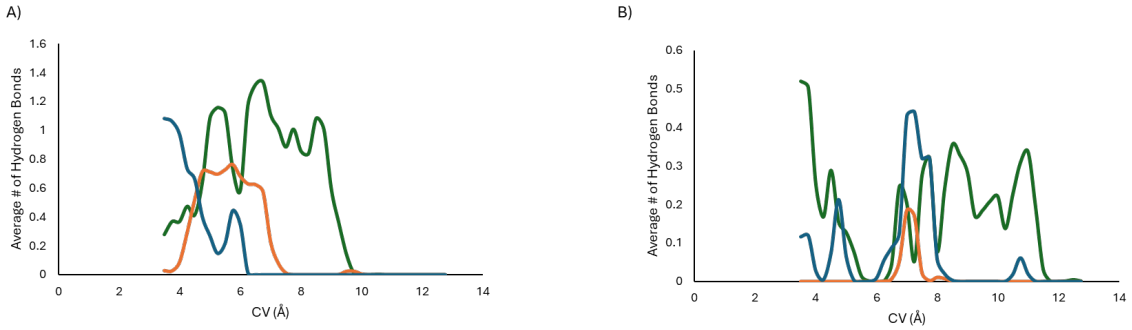


Figure D.4. Average number of hydrogen bonds formed between 8oG and N315 (green), S147 (blue), or N43 (orange) over the course of the umbrella sampling to model 8oG unbinding from the hOGG1 active site for a A) cationic or B) neutral imine crosslink.

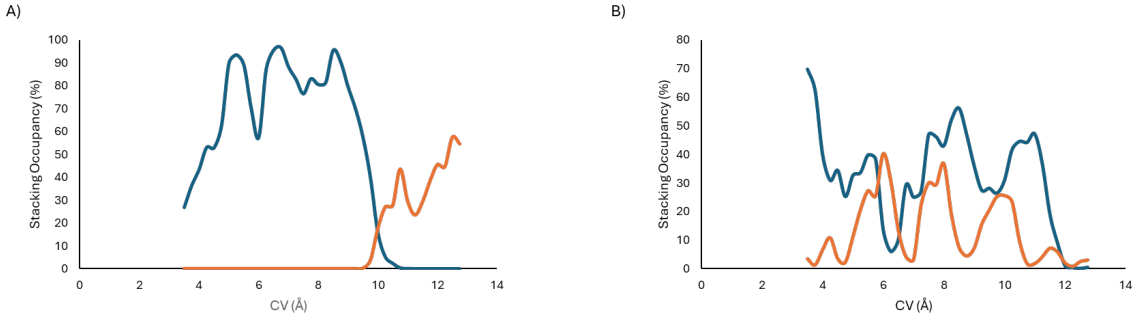


Figure D.5. Occupancy of stacking interactions between 8oG and F319 (blue) or H270 (orange) over the course of umbrella sampling to model 8oG unbinding from the hOGG1 active site for a A) cationic or B) neutral imine crosslink.

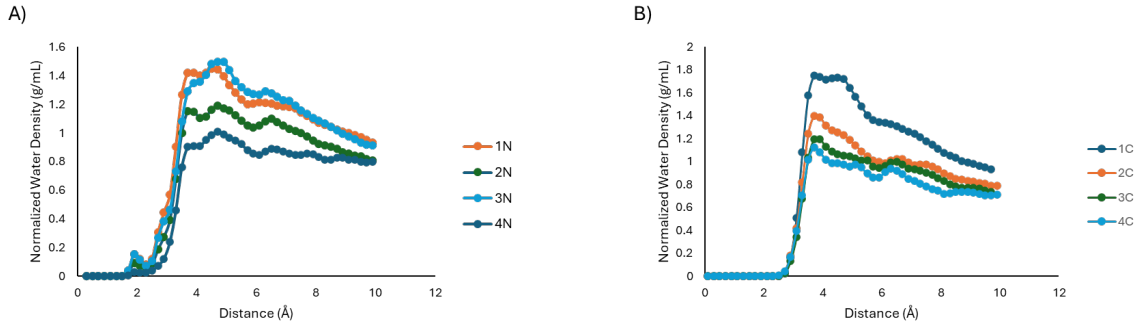


Figure D.6. Normalized water density plots across classical MD simulations on hOGG1 bound to a A) neutral or B) cationic imine crosslink. Distance is between water and the Cβ, Cγ, Cδ, Cε, Nζ, C1′ and C2′ atoms of the imine crosslink.

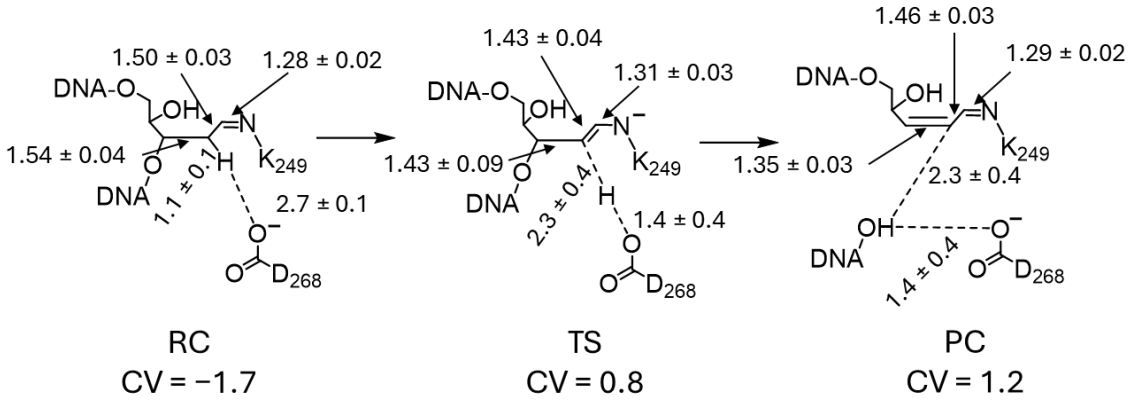


Figure D.7. Bond lengths (Å) in the stationary points along the QM/MM MD elimination reaction for hOGG1 bound to a neutral imine crosslink.

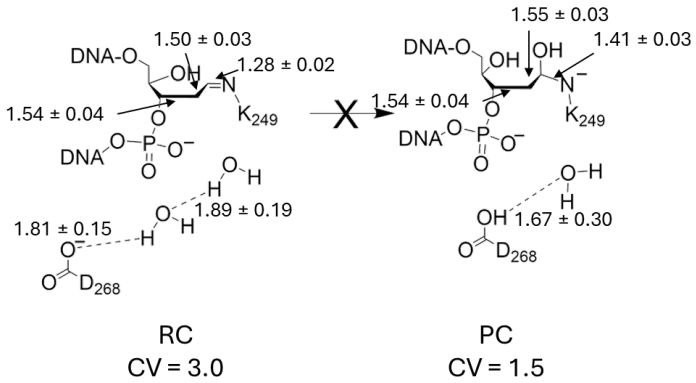


Figure D.8. Bond lengths (Å) in the RC and theoretical PC along the QM/MM MD elimination reaction for hOGG1 bound to a neutral imine crosslink. A representative structure corresponding to a CV of 1.5 Å was shown as a representative complex for the hydrolysis reaction.

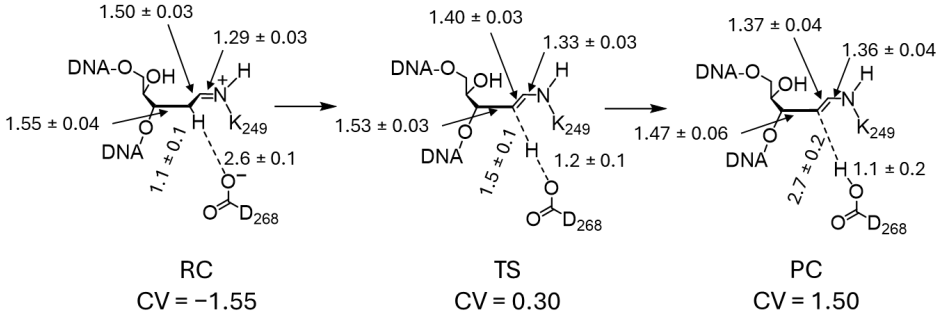


Figure D.9. Bond lengths (Å) for the stationary points along the QM/MM MD elimination reaction for hOGG1 bound to a cationic imine crosslink.

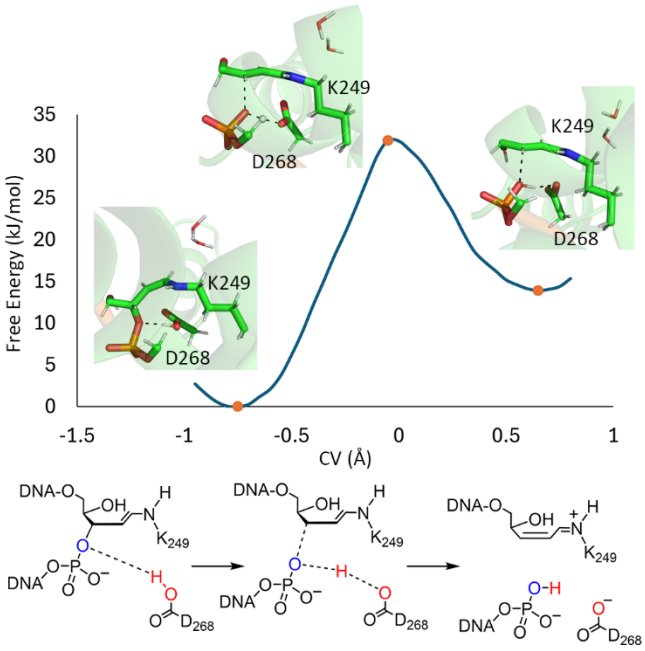


Figure D.10. Potential energy surface and representative structures of stationary points for the phosphate protonation of a cationic imine crosslink post H2′ abstraction. CV defined in Table D.9.

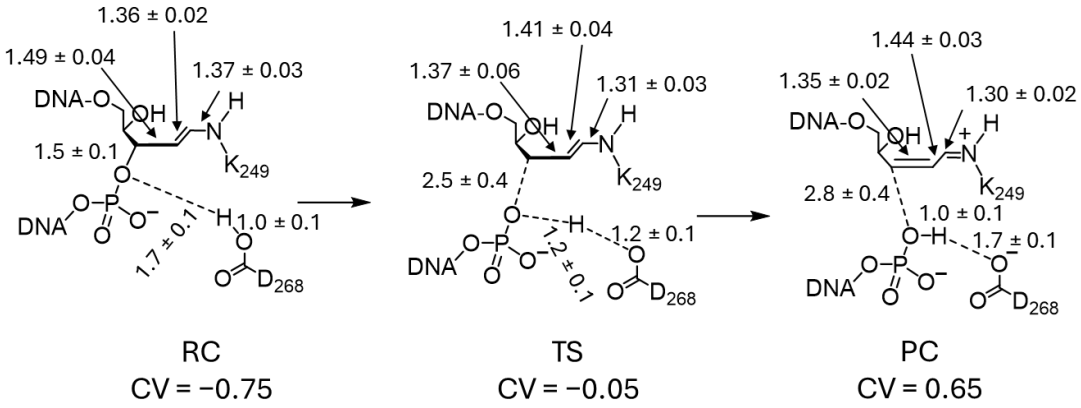


Figure D.11. Bond lengths (Å) for the stationary points along the QM/MM MD phosphate protonation reaction for hOGG1 bound to a cationic imine crosslink post H2′ abstraction.

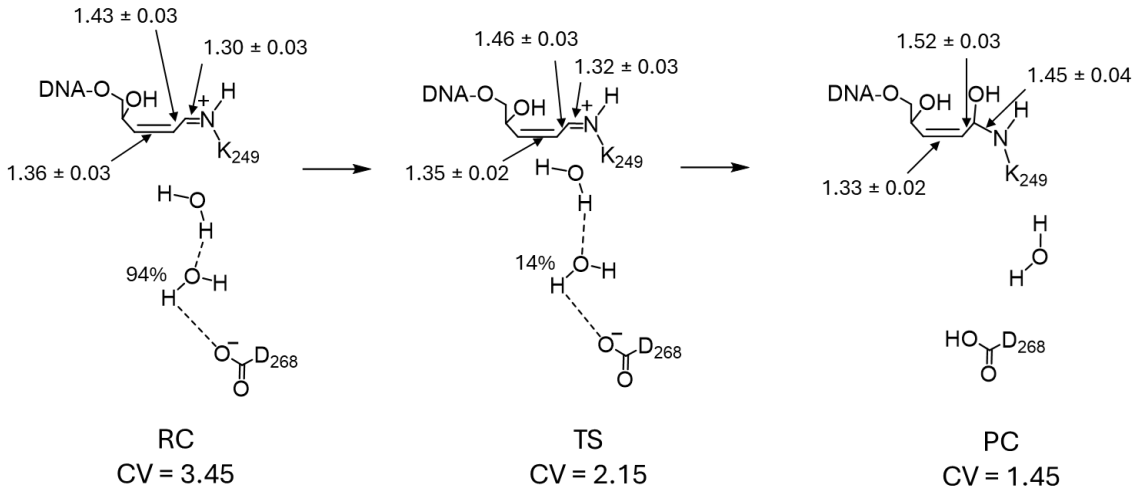


Figure D.12. Bond lengths (Å) and bridging water occupancies for the stationary points along the QM/MM MD hydrolysis reaction for hOGG1 bound to a post-elimination cationic imine crosslink.

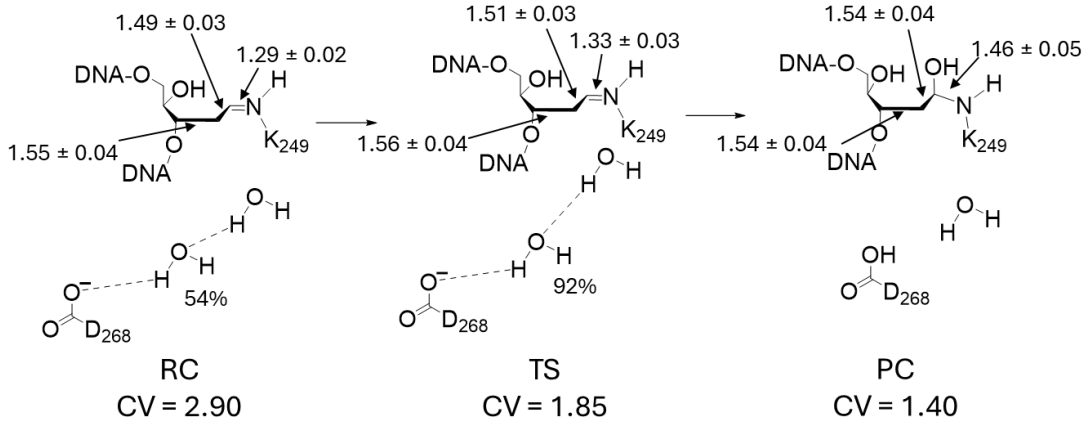


Figure D.13. Bond lengths (Å) and bridging water occupancies for the stationary points along the QM/MM MD hydrolysis reaction for hOGG1 bound to a cationic imine crosslink.

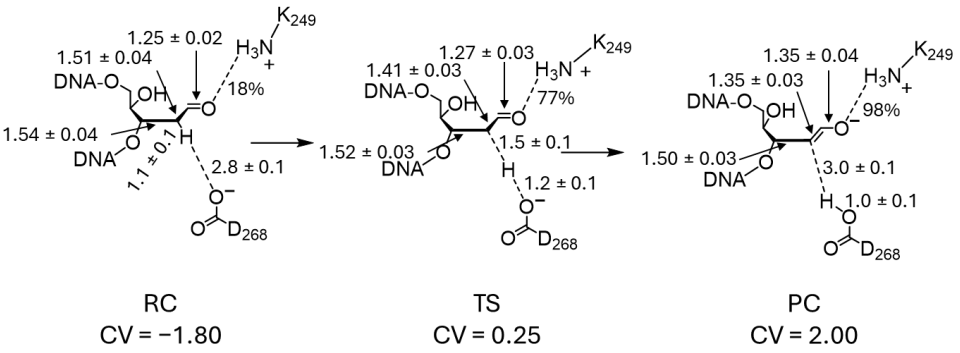


Figure D.14. Bond lengths (Å) and hydrogen bond occupancies for the stationary points along the QM/MM MD D268-catalyzed elimination reaction for hOGG1 bound to a hydrolyzed crosslink.

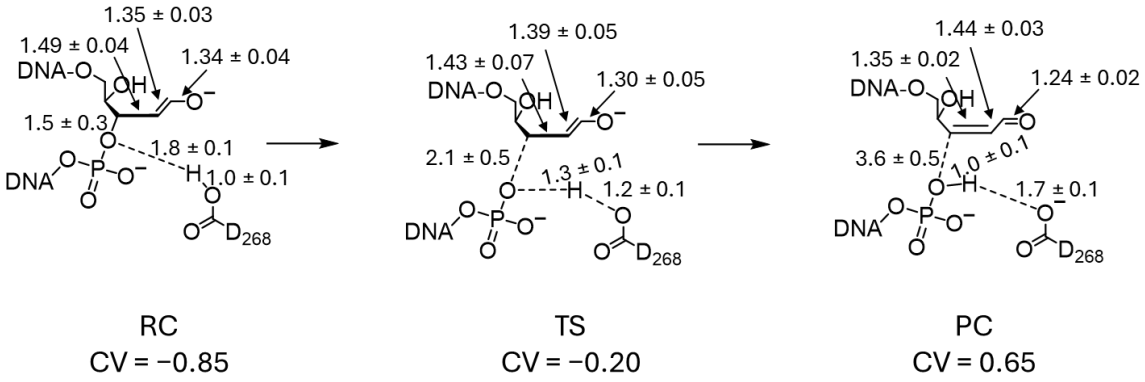


Figure D.15. Bond lengths (Å) for the stationary points along the QM/MM MD phosphate protonation reaction for hOGG1 bound to a hydrolyzed crosslink after the H2′ abstraction.

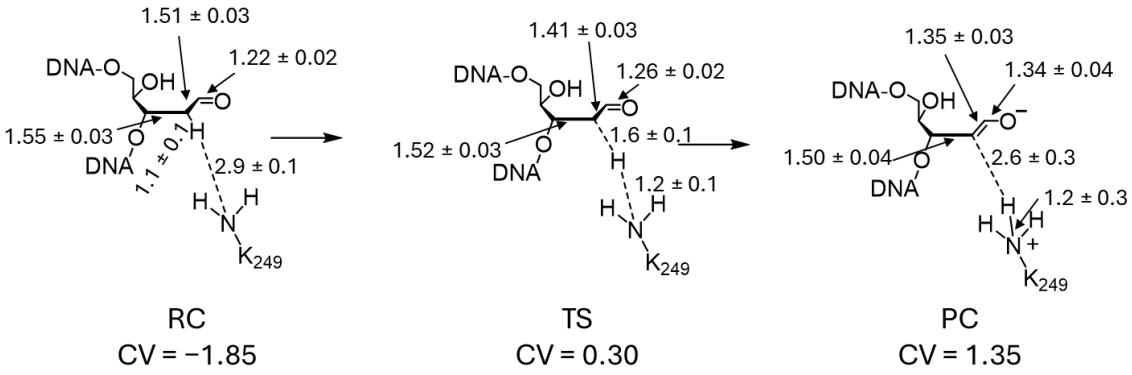


Figure D.16. Bond lengths (Å) and hydrogen bond occupancies for the stationary points along the QM/MM MD K249-catalyzed elimination reaction for hOGG1 bound to a hydrolyzed crosslink.

Table D.1. Amber atom types and charges for a neutral imine crosslink.[a]

Atom	Atom Type	Atom Charge	Atom	Atom Type	Atom Charge
P	P	1.214	C2'	CT	−0.147
C5'	CT	−0.014	C1'	c2[b]	0.26
O5'	OS	−0.524	O3'	OS	−0.514
C4'	CT	0.073	OP1	O2	−0.791
O4'	OH	−0.618	OP2	O2	−0.791
C3'	CT	0.000	HO4'	HO	0.430
H1'	h4[b]	0.089	H4'	H1	0.123
H2'	HC	0.086	H5'	H1	0.100
H3'	H1	0.177	H2'	HC	0.086
H5''	H1	0.100	N	N	−0.365
CA	CX	−0.180	CB	CT	0.010
C	C	0.680	CG	CT	−0.009
O	O	−0.600	CD	CT	0.011
CE	CT	0.094	HG1	HC	0.016
NZ	n2[b]	−0.537	HG2	HC	0.016
H	H	0.242	HD1	HC	0.020
HA	H1	0.147	HD2	HC	0.020
HB1	HC	0.010	HE1	H1	0.039
HB2	HC	0.010	HE2	H1	0.039

[a]Atoms corresponding to lysine use ff14SB atom types, while atoms corresponding to the abasic site use OL15 atom types. [b]General Amber forcefield atom type.

Table D.2. Amber atom types and charges for a cationic imine crosslink.[a]

Atom	Atom Type	Atom Charge	Atom	Atom Type	Atom Charge
C5'	CT	−0.0026	CA	CX	0.0614
O5'	OS	−0.5092	C	C	0.6271
C4'	CT	0.1082	O	O	−0.5516
O4'	OH	−0.7056	CB	CT	−0.0852
C3'	CT	0.0881	CG	CT	−0.0086
O3'	OS	−0.5093	CD	CT	−0.0141
C2'	CT	−0.1375	CE	CT	0.0223
C1'	c2[b]	0.2222	NZ	n2[b]	−0.2761
H1'	h4[b]	0.2425	H	H	0.2619
H2'	HC	0.1224	HA	H1	0.0745
H2''	HC	0.1224	HB1	HC	0.0465
H3'	H1	0.1475	HB2	HC	0.0465
H4'	H1	0.1410	HG1	HC	0.0237
HO4'	HO	0.4799	HG2	HC	0.0237
H5'	H1	0.0944	HD1	HC	0.0255
H5''	H1	0.0944	HD2	HC	0.0255
P	P	1.2133	HE1	H1	0.0853
OP1	O2	−0.7918	HE2	H1	0.0853
OP2	O2	−0.7918	HZ	Hn[b]	0.3426
N	N	−0.4447			

[a]Atoms corresponding to lysine use ff14SB atom types, while atoms corresponding to the abasic site use OL15 atom types. [b]General Amber forcefield atom type.

Table D.3. Amber non-standard bond parameters for the imine crosslinked systems.

Bond	Force Constant (kcal/mol/Å^2)	Equilibrium Value (Å)
CT-n2	259.9	1.466
CT-c2	255.6	1.51

Table D.4. Amber non-standard angle parameters for the imine crosslinked systems.

Angle	Force Constant (kcal/mol/rad^2)	Equilibrium Value (°)
CT-CT-n2	84.1	108.8
CT-n2-hn	45.9	118.4
CT-n2-c2	68.5	115.3
H1-CT-n2	61.1	109.81
CT-c2-n2	84	123.43
CT-CT-c2	65.5	111.56
CT-c2-h4	45.8	119.02

Table D.5. Amber non-standard dihedral parameters for the imine crosslinked systems.

Dihedral	Divider	Barrier (kcal/mol)	Phase (°)	Periodicity
OS-CT-CT-c2	9	1.4	0	3
CT-CT-CT-n2	9	1.4	0	3
CT-CT-n2-hn	6	0	0	3
CT-CT-n2-c2	6	0	0	3
HC-CT-CT-n2	9	1.4	0	3
CT-c2-n2-CT	1	10.37	180	2
H1-CT-n2-hn	6	0	0	3
H1-CT-n2-c2	1	0.165	180	3
CT-CT-c2-n2	6	0	0	2
CT-CT-c2-h4	6	0	0	2
H1-CT-CT-c2	9	1.4	0	3
HC-CT-c2-n2	6	0	0	2
HC-CT-c2-h4	6	0	0	2
CT-CT-CT-c2	9	1.4	0	3
OS-CT-CT-c2	9	1.4	0	3
OS-CT-CT-c2	9	1.4	0	3

Table D.6. Amber non-standard improper dihedral parameters for the imine crosslinked systems.

Improper Dihedral	Barrier (kcal/mol)	Phase (°)	Periodicity
CT-h4-c2-n2	1.1	180	2

Table D.7. Amber atom types and charges for anionic 8oG. [a]

Atom	Atom Type	Atom Charge	Atom	Atom Type	Atom Charge
N1	NA	-0.7641	C5	CB	-0.2651
H1	H	0.3955	C6	C	0.6438
C2	CA	0.8596	O6	O	-0.6762
N2	N2	-0.9227	N7	NA	-0.6292
H21	H	0.3594	H7	H	0.3614
H22	H	0.3594	C8	C	0.8682
N3	NC	-0.749	O8	O	-0.740
C4	CB	0.7151	N9	N*	-0.8161

[a] Atom types correspond to the OL15 forcefield.

Table D.8. Start and end points for steered MD simulations on the hOGG1−substrate complex used to identify starting points for QM/MM MD umbrella sampling calculations.

Simulation	CV	Starting CV (Å)	Ending CV (Å)
Neutral Crosslink Elimination	H2'−D268Oδ Distance	2.8	0.8
Neutral Crosslink Hydrolysis	WatO−C1' Distance	3.0	1.0
Cationic Crosslink Elimination	H2'−D268Oδ Distance	2.8	0.9
Cationic Crosslink Post-Elimination Proton Transfer	D268Hδ−O3' Distance	1.8	0.9
Cationic Crosslink Hydrolysis	WatO−C1' Distance	3.5	1.0
Post-elimination Cationic Crosslink Hydrolysis	WatO−C1' Distance	3.5	1.4
Hydrolyzed Crosslink D268-catalyzed Elimination	r(H2'−D268Oδ)−r(C2'−H2')	−1.5	1.5
Hydrolyzed Crosslink K249-catalyzed Elimination	H2'−D268Oδ Distance	2.7	1.0
Hydrolyzed Crosslink Post-elimination Proton Transfer	D268Hδ−O3' Distance	2.0	0.9

[a]Simulations used a harmonic constant of 200 (kcal/mol)/Å^2.

Table D.9. CV and range of CV windows for every umbrella sampling calculation modeling the β-lyase reaction in the hOGG1−substrate complex.

Simulation	CV	CV Range (Å)
Neutral Crosslink Elimination	r(H2'−D268Oδ)−r(C2'−H2')	−1.8−1.5
Neutral Crosslink Hydrolysis	WatO−C1' Distance	1.0−3.0
Cationic Crosslink Elimination	r(H2'−D268Oδ)−r(C2'−H2')	−1.8−1.5
Cationic Crosslink Post-elimination Proton Transfer	r(D268Hδ−D268Oδ)−r(D268Hδ −O3')	−0.9−0.8
Cationic Crosslink Hydrolysis	WatO−C1' Distance	1.4−3.5
Post-elimination Cationic Crosslink Hydrolysis	WatO−C1' Distance	1.4−3.5
Hydrolyzed Crosslink D268-catalyzed Elimination	r(H2'−D268Oδ)−r(C2'−H2')	−2.0−2.0
Hydrolyzed Crosslink K249-catalyzed Elimination	r(H2'−D268Oδ)−r(C2'−H2')	−2.0−2.0
Hydrolyzed Crosslink Post-elimination Proton Transfer	r(D268Hδ−D268Oδ)−r(D268Hδ −O3')	−1.0−0.7

Appendix E

Supplementary Information for Chapter 6: How Do DICER1 Syndrome Mutations Disrupt Catalysis? Unveiling Dicer Metal Binding Architecture and Mechanism of Action using MD and QM/MM Techniques

Contains Figures E.1−E.15 and Tables E.1−E.2

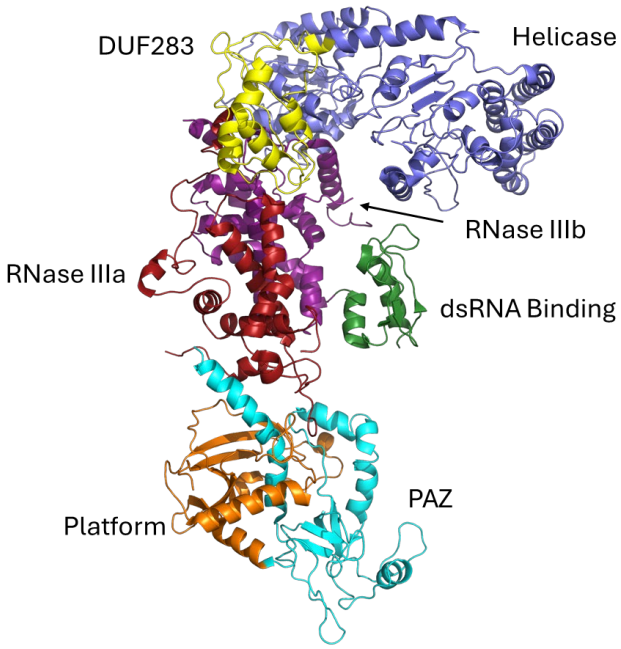


Figure E.1. Dicer structure, highlighting the various domains.

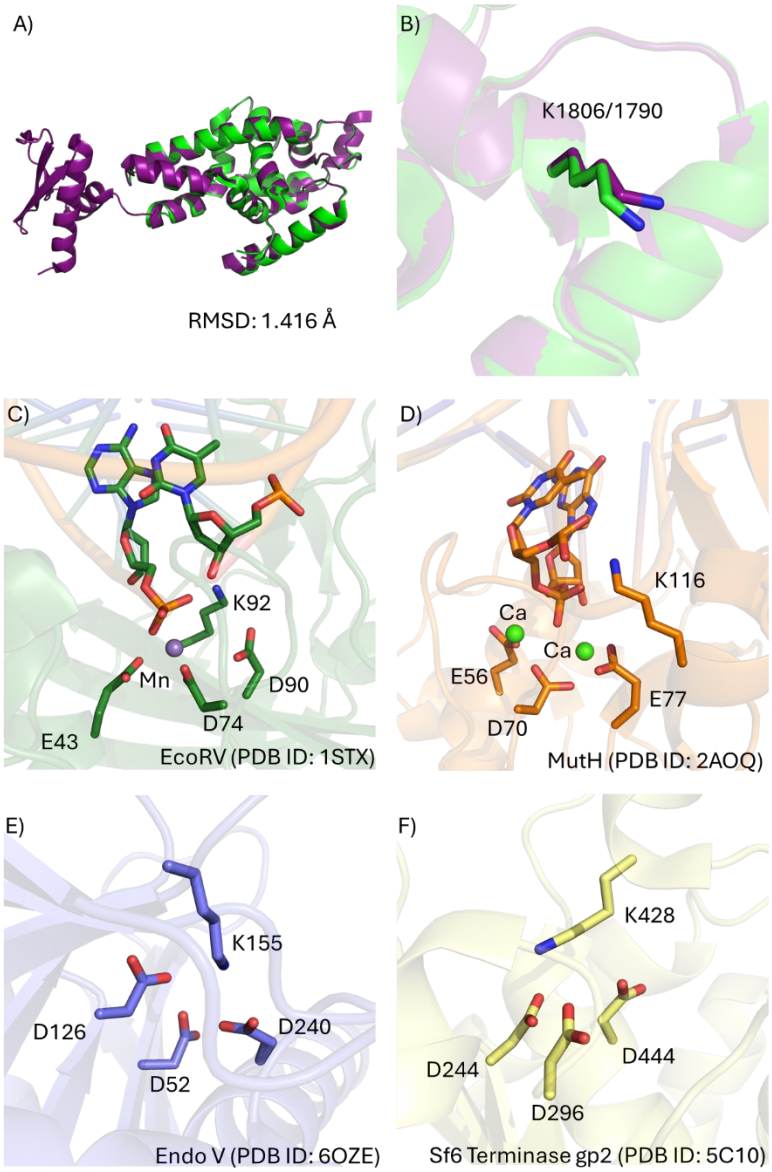


Figure E.2. Overlay of mouse (purple, PDB ID: 3C4B) and human (green, PDB ID: 2EB1) RNase IIIb domains of Dicer highlighting A) the overall structural similarity and B) the similar positioning of K1806 and K1790. Crystal structures of the active sites of C) the K38A mutant EcoRV bound to DNA, D) MutH bound to DNA, E) human Endo V, and F) the gp2 subunit of Sf6 showing a similarly positioned active site lysine (K428) as in Dicer (K1806).

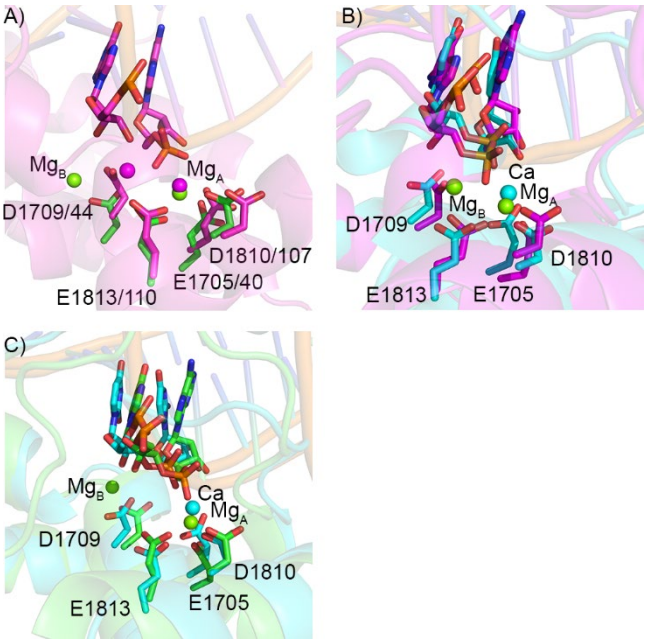


Figure E.3. A) Overlay of *aa*-RNase III (magenta) and Dicer RNase IIIb homodimer (PDB ID: 2EB1, green), highlighting key active site residues and differing locations of Mg^{2+}. Overlays of the cryo-EM structure of Ca^{2+}–inhibited Dicer (cyan, PDB ID: 7XW2) and B) the post-equilibration aaRIII model (magenta), and C) the post-equilibration RIIIb model (green).

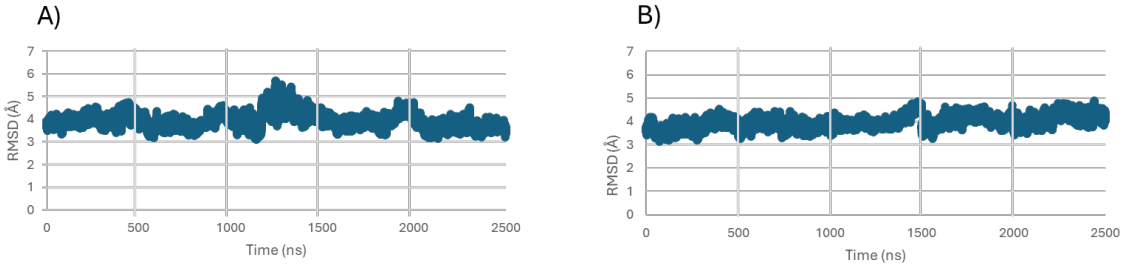


Figure E.4. All-heavy-atom RMSD relative to the first frame over the MD simulations for the A) aaRIII and B) RIIIb wild-type Dicer models. The five 500 ns replicates composing each simulation are joined back-to-back.

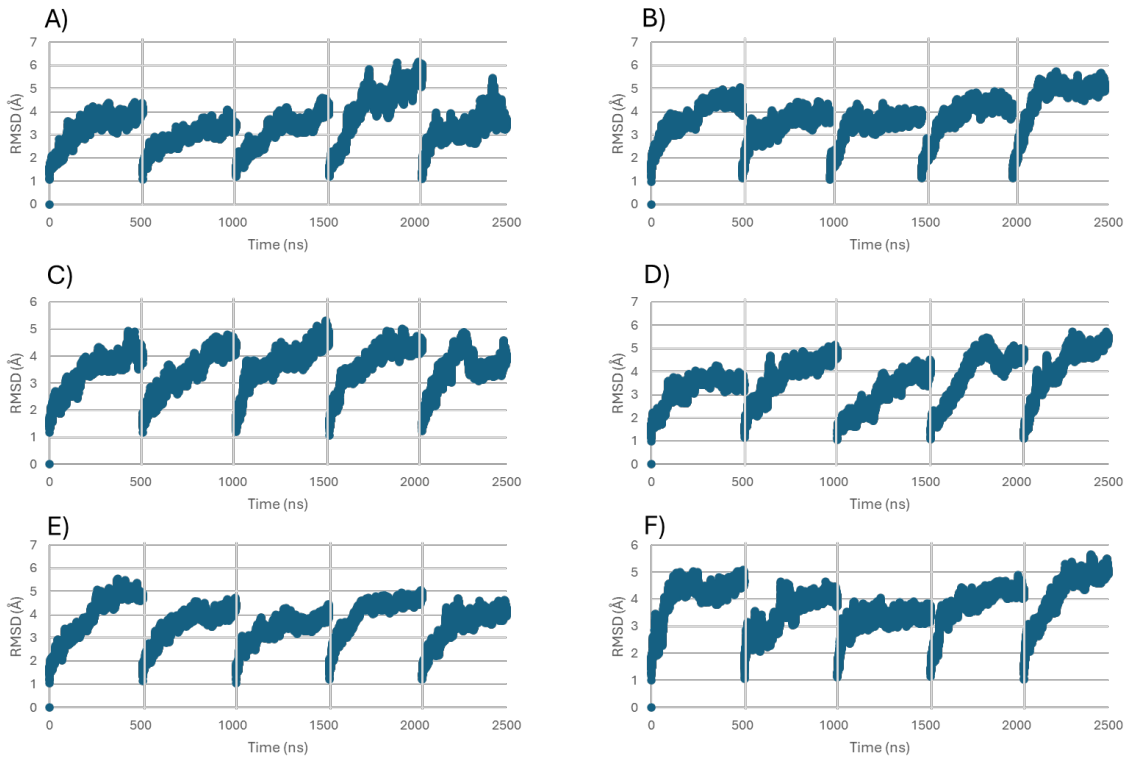


Figure E.5. All-heavy-atom RMSD relative to the first frame over the MD simulations for the A) D1709N, B) D1810Y, C) E1813D, D) E1705K, E) E1813G, and F) G1809R Dicer mutants. The five 500 ns replicates composing each simulation are joined back-to-back.

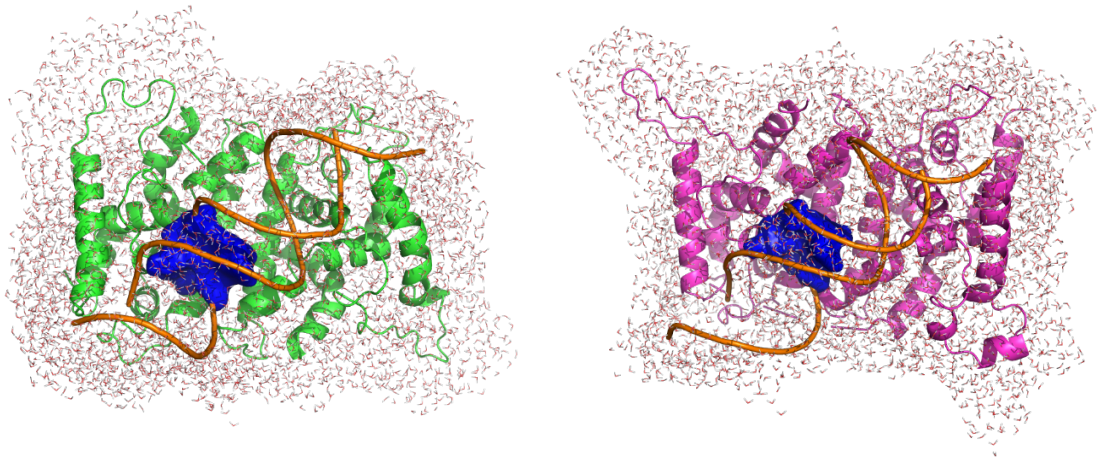


Figure E.6. QM/MM RIIIb (left) and aaRIII (right) models used to calculate the Dicer catalytic pathway. QM region highlighted in blue.

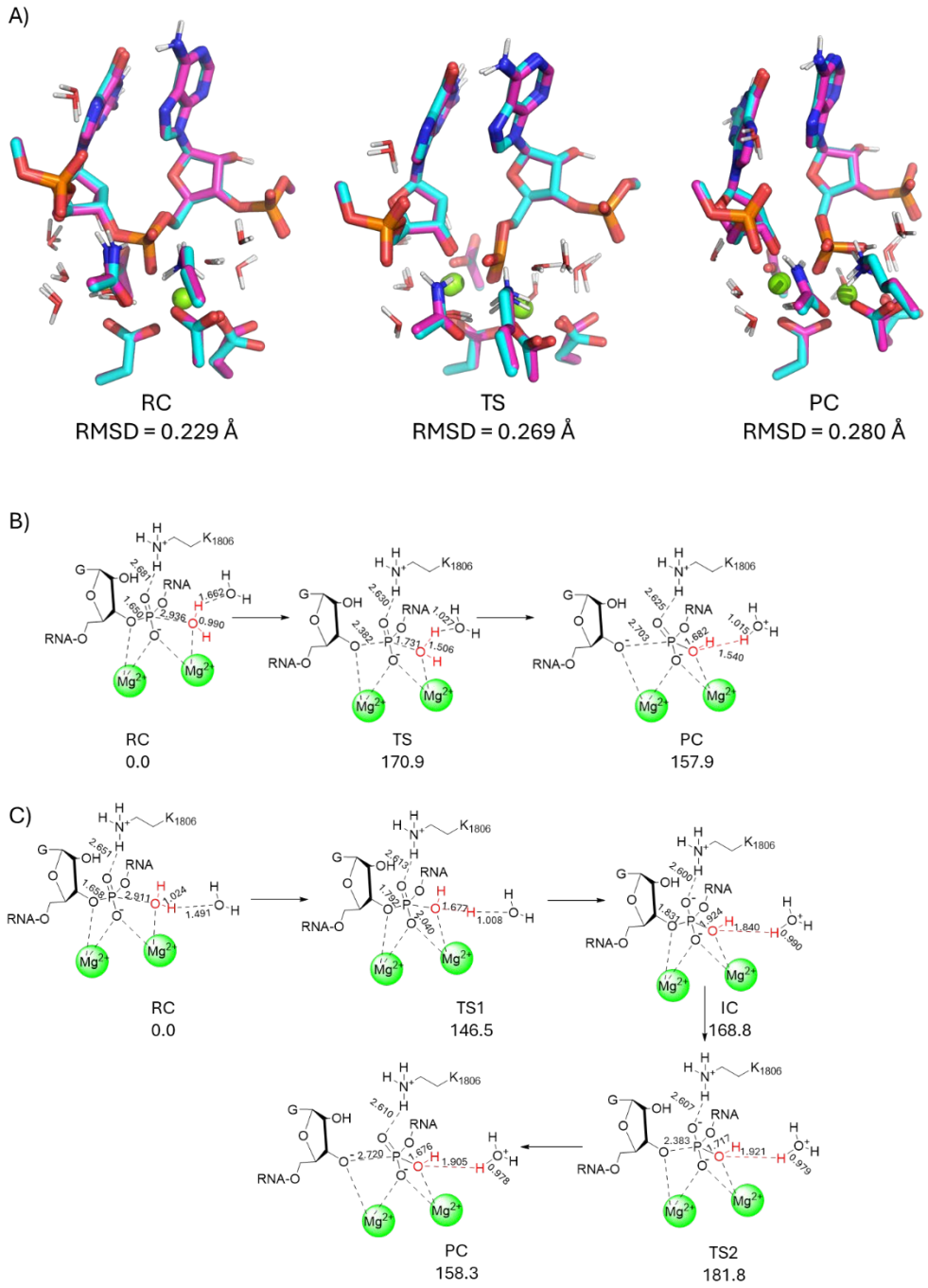


Figure E.7. A) Overlay of key ONIOM(M06-2X/6-311+G(2df,p):AMBERff14SB)//ONIOM(M06-2X/6-31G(d,p):AMBERff14SB) stationary points for the Dicer catalytic mechanism obtained using the aaRIII model employing the mechanical (magenta) and electronic (cyan) embedding schemes during geometry optimization. Calculated Dicer mechanism using B) electronic embedding and C) mechanical embedding during optimizations for the aaRIII model. Distances and RMSDs are reported in Å. Relative energies are reported in kJ/mol.

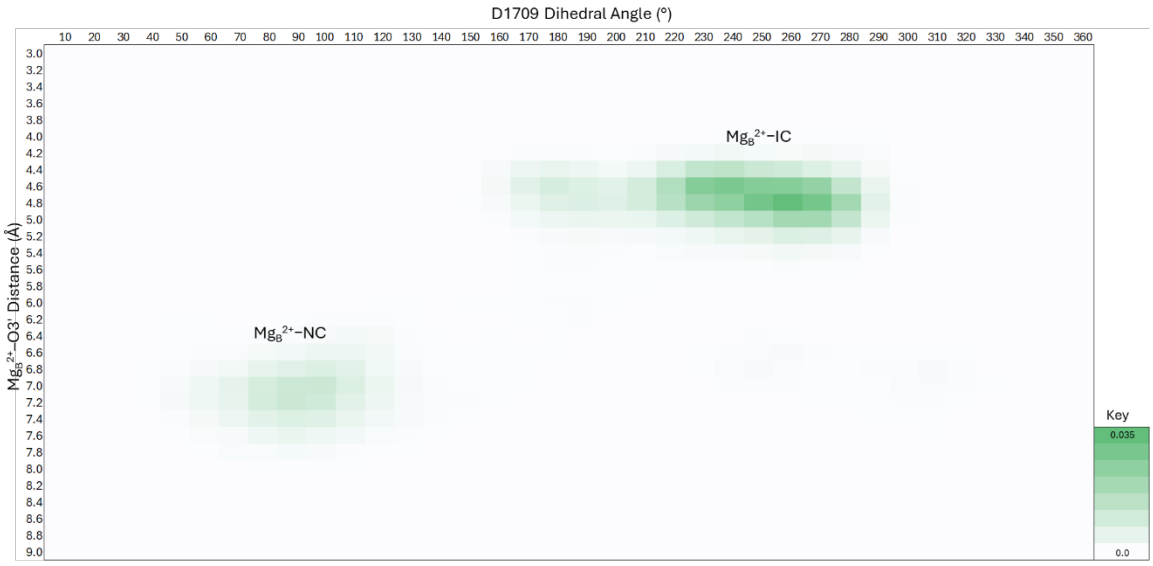


Figure E.8. Histogram of the fractional occupancy of the distance (Å) between Mg_B^{2+} and the substrate O3′ leaving group, and the ∠(OδCγCβCα) dihedral angle in D1709 (°) over the MD simulations for the Dicer RIIIb model. MD representative structures of conformations can be found in Figure 6.5.

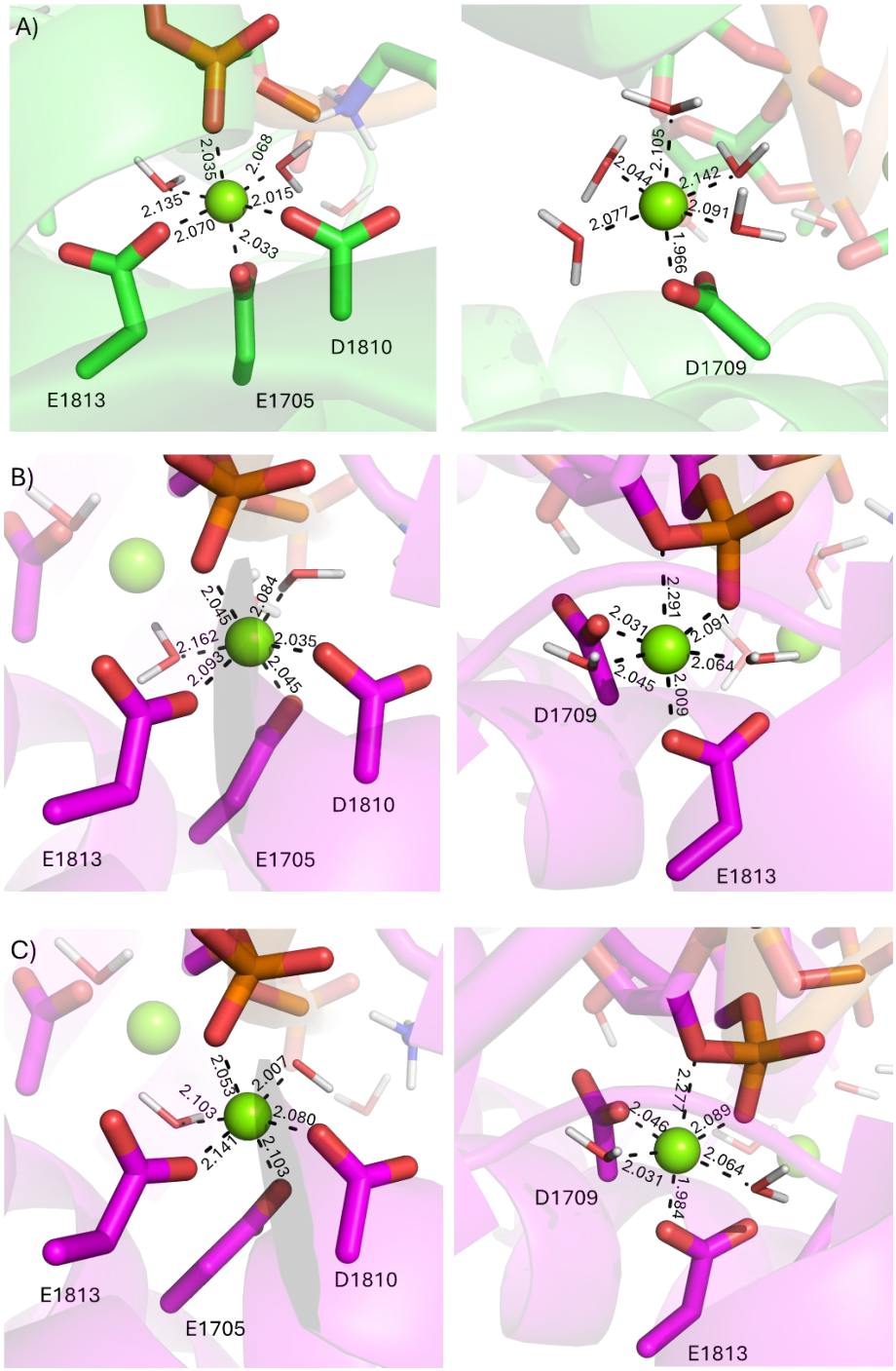


Figure E.9. ONIOM(M06-2X/6-31G(d,p):AMBERff14SB) optimized reactant complexes, highlighting Mg_A^{2+} (left) and Mg_B^{2+} (right) coordination for the A) RIIIb and B) aaRIII models involving a water nucleophile, and C) aaRIII model involving a hydroxide nucleophile.

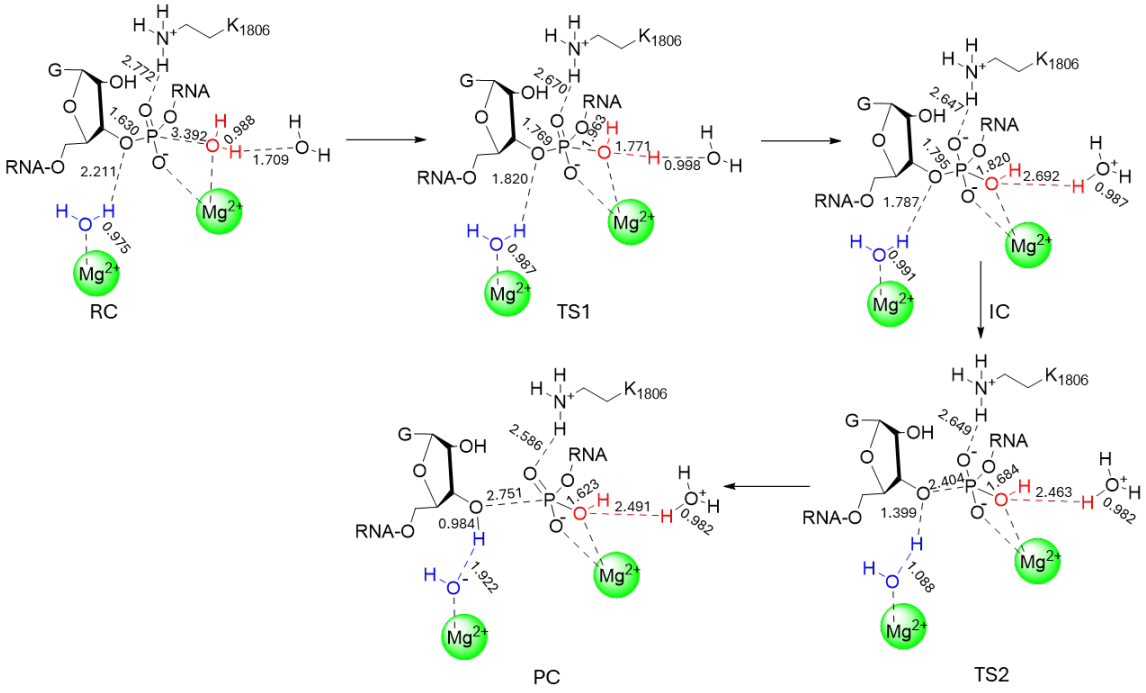


Figure E.10. ONIOM(M06-2X/6-311+G(2df,p):AMBERff14SB)//ONIOM(M06-2X/6-31G(d,p):AMBERff14SB) calculated Dicer catalytic mechanism involving a water nucleophile obtained using the RIIIb model. Distances reported in Å. Water molecules coordinated to Mg^{2+} ions that do not participate in the reaction are omitted for clarity.

Figure E.11. Histogram of the fractional occupancies of the distance between Mg_A^{2+} and Mg_B^{2+} and the substrate non-bridging phosphate oxygen over the MD simulations with the aaRIII model highlighting three conformations. MD representative structures of the three conformations can be found in Figure 6.7.

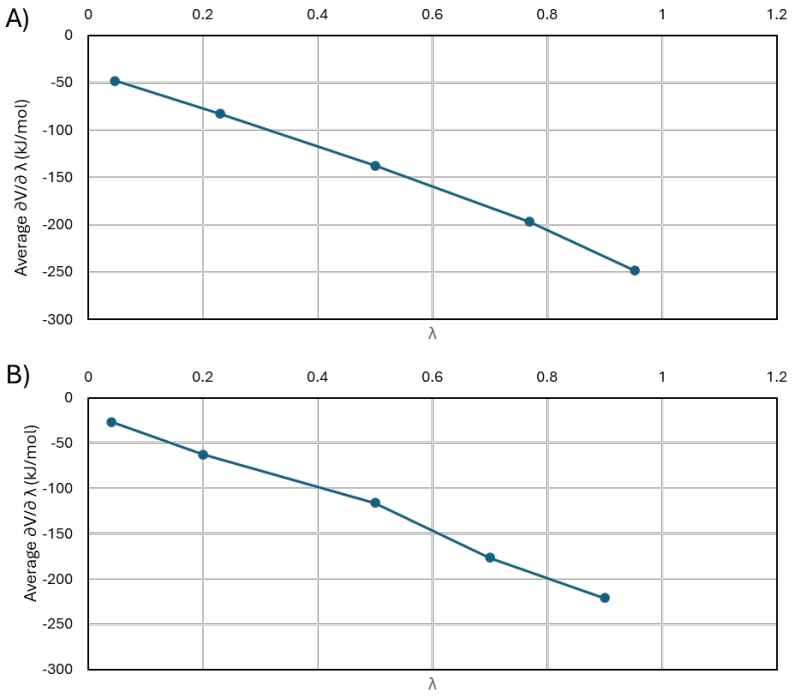


Figure E.12. Average ∂V/∂λ from thermodynamic integration simulations at different λ values for the alchemical transformation of water into hydroxide in A) Dicer and B) bulk solvent.

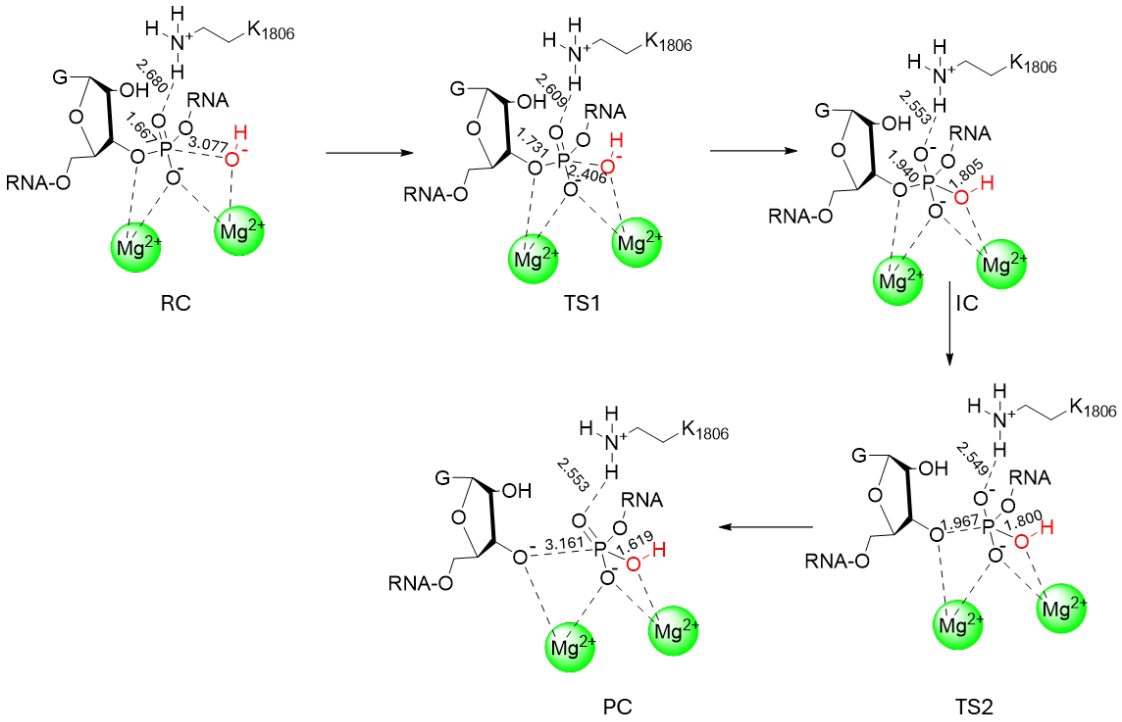


Figure E.13. ONIOM(M06-2X/6-311+G(2df,p):AMBERff14SB)//ONIOM(M06-2X/6-31G(d,p):AMBERff14SB) calculated Dicer catalytic mechanism involving a hydroxide nucleophile obtained using the aaRIII model. Distances reported in Å.

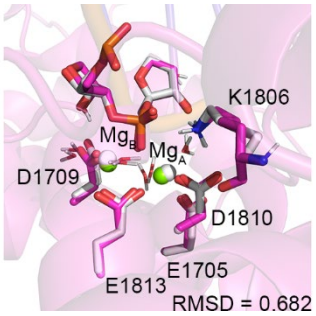


Figure E.14. Overlay of MD representative structures of the active ($2Mg^{2+}$−B) conformation of the Dicer aaRIII model containing a water (magenta) and hydroxide (white) nucleophile coordinated to Mg_A^{2+}. RMSD is reported in Å.

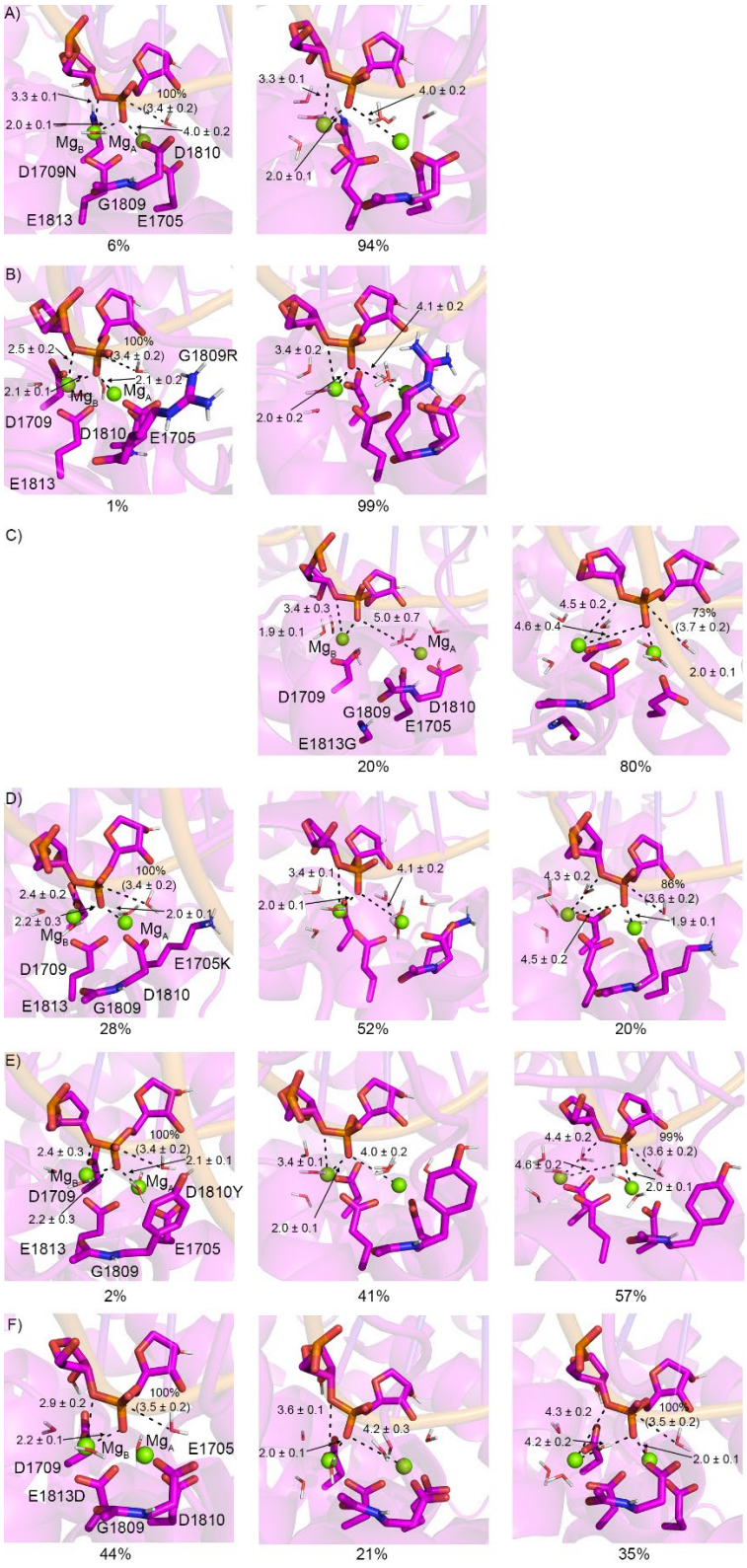


Figure E.15. MD representative structures for the $2Mg^{2+}$−B (left), Mg_A^{2+}−U (middle), and Mg_B^{2+}−U (right) conformations for the A) D1709N, B) G1809R, C) E1813G, D) E1705K, E) D1810Y, and F) E1813D Dicer mutants.

Table E.1. RESP charges for a hydroxide ion calculated at the HF/6-31G(d) level of theory.

Atom Name	Charge
O	–1.2049
H1	0.2049
H2	0.0000

Table E.2. Relative energies of key stationary points for the RIIIb and aaRIII Dicer models.

Structure	RC	TS1	IC	TS2	PC
RIIIb Uncorrected[a]	0.0	126.1	124.1	154.8	69.8
RIIIb Corrected[b]	0.0	253.4	269.5	293.1	221.3
aaRIII Uncorrected (Water Nucleophile)[a]	0.0	62.2	61.4	77.6	74.5
aaRIII Corrected (Water Nucleophile)[b]	0.0	146.5	168.8	181.8	158.3
aaRIII Uncorrected (Hydroxide Nucleophile)[a]	0.0	15.1	–12.8	–12.8	–90.1
aaRIII Corrected (Hydroxide Nucleophile)[b]	0.0	39.2	36.8	45.0	28.9

[a]Energies were calculated using ONIOM(M06-2X/6-31G(d,p):AMBERff14SB) with the ME scheme. [b]Energies were calculated using ONIOM(M06-2X/6-311+G(2df,p):AMBERff14SB)//ONIOM(M06-2X/6-31G(d,p):AMBERff14SB). The EE scheme was used for single-point calculations and a Gibbs energy correction was included.