COMPUTATIONAL INVESTIGATION OF ENZYME-FACILITATED CLEAVAGE OF THE PHOSPHODIESTER BOND IN NUCLEIC ACIDS

RAJWINDER KAUR

Masters of Science (Chemistry Honors), Panjab University, Chandigarh, 2017

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

COMPUTATIONAL INVESTIGATION OF ENZYME FACILITATED CLEAVAGE OF THE PHOSPHODIESTER BOND IN NUCLEIC ACIDS

RAJWINDER KAUR

Date of Defence: February 5th, 2024

Dr. Stacey Wetmore Supervisor	Professor	PhD
Dr. Michael Gerken Thesis Examination Committee Member	Professor	PhD
Dr. Nehalkumar Thakor Thesis Examination Committee Member	Professor	PhD
Dr. Nathan DeYonker External Examiner The University of Memphis Memphis, Tennessee	Associate Professor	PhD
Dr. Paul Hayes Chair, Thesis Examination Committee	Professor	PhD

DEDICATION

To my sister who gave me unconditional emotional support throughout my academic journey and my late father for supporting my dreams and believing in me without even knowing the meaning of the word ‘chemistry’.

ABSTRACT

The phosphodiester backbone in nucleic acids is remarkably resistant to degradation. Although high stability is essential for storage of genetic information and proper cell function, some circumstances necessitate the cleavage of the nucleic acid backbone. For example, breaking the DNA backbone is critical to repair damage and maintain genetic integrity, while RNA cleavage is necessary for quality control during protein synthesis. Nucleases are enzymes that facilitate the challenging phosphodiester bond cleavage by accelerating the uncatalyzed reaction. Many nucleases utilize metals to enhance catalysis. Despite several experimental studies on enzymes that cleave phosphodiester bonds in nature, the mechanism for bond cleavage used by many enzymes, including the role and/or number of metal ions involved, is still unclear. Computer modeling is a powerful tool to investigate enzyme-catalyzed reaction mechanisms and discern the roles of the metal/s and amino acids involved in the reaction. This thesis uses computational techniques (i.e., quantum mechanics calculations, molecular dynamics simulations, and quantum mechanics–molecular mechanics calculations) to gain an atomic-level understanding of the phosphodiester bond cleavage reaction catalyzed by nucleases, specifically focusing on APE1, I-*Ppo*I, and EndoV. These enzymes are particularly interesting since they either invoke a single metal ion for catalysis, which conflicts with the two-metal mediated mechanism generally proposed for most nucleases, or the metal-dependence is unknown. The mechanistic details uncovered by this thesis will open the door for new and improved applications of these enzymes in the fields of disease diagnostics, genetic engineering, and therapeutics.

PREFACE

This thesis consists of five results chapters, three of which are published in peer reviewed journals, one has been accepted, and the remaining has been submitted for publication. The contributions of all authors will be explained here based on CRediT (Contributor Roles Taxonomy) roles.

Chapter 2 has been published [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]. In this chapter, all authors contributed to the reviewing and editing of the manuscript. I aided conceptualization of the project and performed the computational characterization of the phosphodiester bond cleavage pathways (QM cluster optimizations in gas-phase and single-point calculations in both gas and solvent phase) using models 0, 1, 4, 5, 6, 8–10. In addition, I performed the overall data/results analysis, wrote the first draft of the manuscript, and generated all figures for the publication. D. J. Nikkel performed optimizations (gas-phase) and solvent single-point calculations (gas and solvent phase) using models 2, 3, and 7, investigated the impact of different truncation points using model 0, and contributed to the formal data analysis. M. M. Aboelnga contributed towards conceptualization, project administration, and supervision, while S. D. Wetmore was involved in project conceptualization and administration, supervision, funding and resource acquisition, data visualization and interpretation, and writing the original draft of the manuscript.

Chapter 3 has been published [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]. In this chapter, all authors contributed to the reviewing and editing of the manuscript. I aided conceptualization of the project and performed the computational characterization of the phosphodiester bond cleavage pathways involving Mg^{2+}, Ni^{2+} (square-planar and octahedral), Zn^{2+}, and Ca^{2+}. In addition, I performed the overall data/results analysis, wrote the first draft of the manuscript, and generated all the figures for the publication. D. J. Nikkel performed calculations on the pathway involving Mn^{2+} and contributed to the formal data analysis. M. M. Aboelnga contributed towards conceptualization and project administration, while S. D. Wetmore was involved in project conceptualization and administration, supervision, funding and resource acquisition, data visualization and interpretation, and writing the original draft of the manuscript.

Chapter 4 has been submitted for a publication [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**, Accepted (manuscript ID: CP-ART-12-2023-006201.R1)]. In this chapter, I aided conceptualization of the project and performed the computational characterization of the phosphodiester bond cleavage pathways using QM cluster Model 4 and QM/MM model involving direct Mg–O3′ coordination and QM cluster Models 1–4 and QM/MM model involving indirect Mg–O3′ coordination. In addition, I performed the overall data/results analysis, wrote the first draft of the manuscript, contributed to reviewing and editing, and generated all figures for the publication. Angela Frederickson

performed calculations using QM cluster Models 1–3 involving direct Mg–O3′ coordination, while S. D. Wetmore was involved in project conceptualization and administration, supervision, funding and resource acquisition, data visualization and interpretation, writing the original draft of the manuscript, and reviewing and editing.

Chapter 5 has been published [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]. In this chapter, both authors contributed to the reviewing and editing of the manuscript. I aided conceptualization of the project and conducted all calculations. In addition, I performed the overall data/results analysis, wrote the first draft of the manuscript, and generated all figures for the publication. S. D. Wetmore was involved in project conceptualization and administration, supervision, funding and resource acquisition, data visualization and interpretation, and writing the original draft of the manuscript.

Chapter 6 includes a currently unpublished manuscript [Kaur, R.; Wetmore, S. D., [Mechanism of Nucleic Acid Phosphodiester Bond Cleavage by Human Endonuclease V: MD and QM/MM Calculations Reveal a Versatile Metal Dependence. **2024**, *ACS Cat.* Submitted (manuscript ID: cs-2024-01060f)]. In this chapter, both authors contributed to the reviewing and editing of the manuscript. I aided conceptualization of the project, performed MD simulations, and characterized total 4 phosphodiester bond cleavage pathways. In addition, I performed the overall data/results analysis, wrote the first draft of the manuscript, and generated all figures. S. D. Wetmore was involved in project conceptualization and administration, supervision, funding and resource acquisition, data visualization and interpretation, and writing the original draft of the manuscript.

ACKNOWLEDGEMENTS

First, I would like to express my sincerest gratitude towards my supervisor, Dr. Stacey Wetmore for providing me constant support over the years and guiding me in the right direction when everything seemed too overwhelming. There are not enough words to show my appreciation and I could not have asked for a better supervisor who is so deeply invested in their student's success. I would also like to thank my committee members, Dr. Michael Gerken and Dr. Nehalkumar Thakor for the constructive feedback they provided throughout my graduate school journey. Thank-you to Dr. Paul Hayes for agreeing to be on my PhD thesis defense. Finally, thank-you to Dr. Nathan DeYonker from the University of Memphis for taking your valuable time to participate in my thesis defence. I also would like to thank the funding sources that have supported me throughout my graduate studies including Alberta Innovates – Technology Futures and the University of Lethbridge Board of Governors and School of Graduate Studies for student scholarships.

I had the amazing opportunity to work with many great undergraduate and graduate students at the University of Lethbridge. I want to thank the undergraduate students (Angela Frederickson, Melanie Kebel, Pardeepak Singh, Kohl Yee, Mckenzie Franz, Denisa Moldolovan, and Taylor Adamitz), graduate student (Dylan Nikkel, Rebecca Jeong) and post-doctoral fellow (Dr. Mohamed Aboelnga) who either contributed directly to work in this thesis or to other publications. I would also like to thank all the past and present undergraduate students (Trinity Deak, Mark Lea, Bhadra Pandya, Kohl Yee, Zaki Olvis, and Ryan Fjordbotten), graduate students (Katie Wilson, Nathania Takyi, Dylan Nikkel, Makay Murray, Prakash Chukka, Bimaldeep Singh, Cynthia Fonderson, Katie Wilson,

Ryan Kung, Stefan Lenz, and Shahin Sowlati) and post-doctoral fellows (Dr. Preethi Seelam, Dr. Purshotam Sharma, and Dr. Mohamed Aboelnga) for the enlightening discussions during group meetings and in the lab. I also want to say a big thank you to Priya Bhutani for being my first friend in Canada and helping me get accustomed to Canadian culture and supporting me through the first two years of graduate school. A super special thanks to my besties Briana Boychuk and Rebecca Jeong for providing unwavering support during the most difficult times of my graduate school journey, I would not have made it to the end without your help, care, and motivation. Finally, I would like to thank my family especially my sister and Grandma for all the love and care over the years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A	adenine
ADAR	adenosine deaminases acting on RNA
ADAT	adenosine deaminase acting on transfer RNA
ALS	amyotrophic lateral sclerosis
AP	apyrimidinic/apurinic site
APE1	apurinic/apyrimidinic endonuclease 1
ATP	adenosine triphosphate
BER	base excision repair
BRCA1	breast cancer 1
C	cytosine
CG	conjugate gradient
dA	2′-deoxyadenosine
dC	2′-deoxycytidine
dsDNA	double-stranded DNA
DFT	density functional theory
dG	2′-deoxyguanosine
DNA	deoxyribonucleic acid
dUTP	deoxyuridine triphosphate
EcoRI	*Escherichia coli* restriction endonuclease I
EE	electrostatic embedding
EhoI	*Enterobacter hormaechei* restriction endonuclease
EndoV	Endonuclease V
G	guanine
GAFF	general Amber force field
HCC	hepatocellular carcinoma
HE	homing endonuclease
HEG	homing endonuclease gene

HIV	human immunodeficiency virus
HJ	Holliday junction
HpyAV	*Helicobacter pylori* restriction endonuclease
Hx	hypoxanthine
I	inosine
IC	intermediate complex
IEF-PCM	integral equation formalism-polarizable continuum model
I-*Ppo*I	group I intron-encoded endonuclease from Physarum polycephalum
IRC	intrinsic reaction coordinate
Kpn1	*Klebsiella pneumoniae* restriction endonuclease
MD	molecular dynamics
ME	mechanical embedding
MM	molecular mechanics
MnII	*Moraxella nonliquefaciens* restriction endonuclease II
mRNA	messenger RNA
NER	nucleotide excision repair
NIR	nucleotide incision repair
NMR	nuclear magnetic resonance
ONIOM	our own N-layered integrated molecular orbital and molecular mechanics
PC	product complex
PES	potential energy surface
PDB	protein data bank
PME	particle mesh Ewald
QM	quantum mechanics
RC	reactant complex
RESP	restrained electrostatic potential
RMSD	root-mean-square deviation
RNA	ribonucleic acid

RNase	Ribonuclease
SCID	severe combined immunodeficiency
SD	steepest descent
SepMI	*Staphylococcus epidermidis* restriction endonuclease
ssDNA	single-stranded DNA
T	thymine
T4 Endo VII	phage T4 endonuclease VII
Tma	*Thermotoga maritima*
tRNA	transfer RNA
TS	transition state
U	uracil
UV	ultraviolet
Vvn	*Vibrio vulnificus* nuclease
WC	Watson-Crick

SINGLE AND TRIPLE LETTER CODES FOR THE AMINO ACIDS

A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartate
E	Glu	Glutamate
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
Y	Tyr	Tyrosine

Chapter 1: Introduction

1.1. General Overview

The backbone of nucleic acids contains phosphodiester bonds, which are exceedingly resistant to degradation.[1-2] While this stability is critical for maintaining genome integrity and regulating gene expression, there are many circumstances that necessitate the cleavage of the nucleic acid backbone including repairing damaged DNA and processing RNA in cells. Enzymes are biological catalysts, some of which facilitate this challenging backbone cleavage by increasing the reaction rate by ~10^{17} times over the uncatalyzed reaction.[3] Many enzymes utilize metal ions to accelerate this catalysis. However, despite several experimental studies on enzymes that cleave phosphodiester bonds in nature, the mechanism for bond cleavage used by some enzymes, including the role and/or number of metal ions involved, is still unclear. This thesis investigates the phosphodiester bond hydrolysis pathways of three such enzymes, namely APE1, I-*Ppo*I, and EndoV. A fundamental understanding about the catalytic function of these enzymes can open the door for applications in biotechnology and medicine, including disease diagnostics, genetic engineering, and the design of small molecule inhibitors to target a host of diseases ranging from neurological disorders to cancer.[4-18]

Although experimental studies provide valuable information regarding enzyme function, it is difficult to gain atomic-level details of the mechanism of action and isolate high energy reaction intermediates using experimental techniques. Computational modeling is a powerful tool to investigate enzyme-catalyzed reaction mechanisms, characterize high energy intermediates, and discern the roles of the metal/s and amino acids involved in the reaction. This thesis uses computer modeling to demonstrate how to

accurately model the metal-mediated phosphodiester bond cleavage reaction, and thereby provide answers for the remaining questions related to metal identity and the catalytic mechanisms used by the above-mentioned enzymes. The following sections of the introduction will discuss nucleic acid structure, the processing of nucleic acids, nucleases, the metal dependence of nucleases, and computer modeling of nucleases, as well as provide a brief summary of how a multipronged computational approach was used in this thesis to provide a detailed description of enzyme-facilitated phosphodiester bond cleavage of nucleic acids.

1.2. Nucleic Acid Structure

Nucleic acids are biopolymers consisting of nucleotide subunits. Each nucleotide subunit contains a nucleobase, sugar (2′-deoxyribose in DNA and ribose in RNA), and a phosphate group (Figure 1.1a). Nucleotides are linked together through phosphodiester bonds to form polymers that can fold into a variety of structures. DNA naturally exists as a double-stranded helix (dsDNA) that contains two complementary strands in an antiparallel orientation (Figure 1.2a). Each strand consists of four canonical DNA nucleobases, namely adenine (A), thymine (T), cytosine (C), and guanine (G, Figure 1.1b), which form selective Watson-Crick hydrogen bonds with the nucleobases from the complementary strand. Specifically, A pairs with T through two hydrogen bonds and G pairs with C through three hydrogen bonds (Figure 1.1c). Furthermore, dsDNA has two grooves, one being the deep and wide major groove in which the sugar phosphate backbones are far apart (12 Å wide) and the other being the narrow and shallow minor groove in which the sugar phosphate backbones are close together (6 Å apart, Figure 1.2a).

Like DNA, RNA is also composed of four canonical nucleobases, except T is substituted by uracil (U, Figure 1.1b). Another important structural difference is the presence of a 2′-OH group on the sugar in RNA. Like DNA bases, the nucleobases in RNA can form complementary pairs (A:U and G:C) to yield the secondary structure of RNA, which can fold into three-dimensional structures consisting of tertiary motifs such as coaxially stacked helices, hairpin loops, tetraloops, quadruplexes, and pseudoknots (Figure 1.2b).

Although maintaining nucleic acid structure is important for preserving crucial cell functions, disruptions to the chemical composition and canonical structure are common occurrences. These structural changes can have deleterious outcomes such as altered genetic information and lead to the formation of aberrant proteins. The following section discusses strategies cells use to deal with altered nucleic acids.

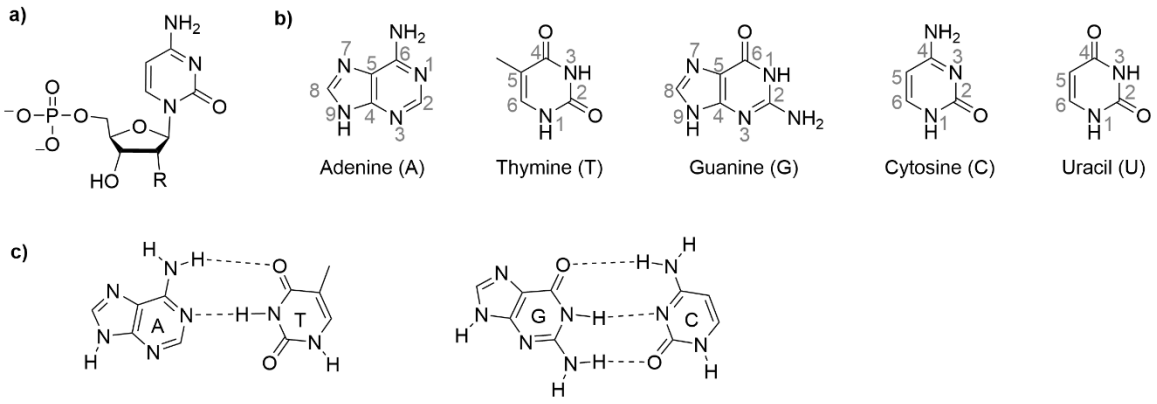

Figure 1.1. a) The structure of the subunits of a nucleic acid including nucleobase, sugar, and phosphate, where R = H for DNA or OH for RNA, b) chemical structure and atom numbering of the canonical DNA and RNA nucleobases, and c) structure of the complementary Watson-Crick base pairs in canonical DNA.

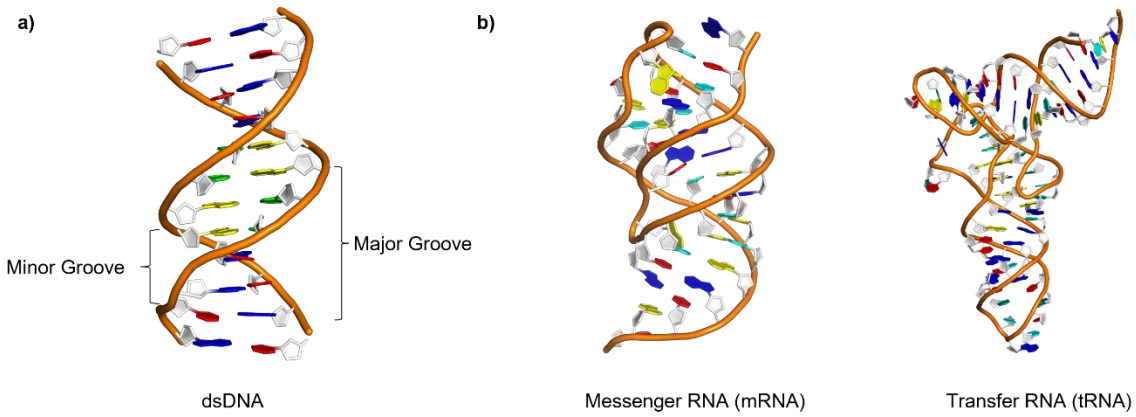

Figure 1.2. The structures of a) a dsDNA helix, showing the major groove and minor groove, and b) RNA with a helical stem connected by two loops (mRNA, left) and four helical stems folding into an L-shape (tRNA, right).

1.3. Processing of Nucleic Acids

DNA can be damaged from various endogenous (such as reactive oxygen species and S-adenosylmethionine) and exogenous (such as pollution and UV light) sources,[19-22] which can result in mutations and cell death.[23] To combat the detrimental effects of DNA damage, cells possess multiple repair pathways including base excision repair (BER) to remove non-bulky lesions,[24] and nucleotide excision repair (NER) to remove bulky adducts or crosslinks.[25] Defects in DNA damage repair pathways have been linked to various diseases including neurodegenerative disorders, cardiovascular diseases, and cancer.[23, 26] In contrast, although modifications to RNA are necessary to create genetic and protein diversity,[27] when the modification balance is tipped such that RNA is over or under modified, codon changes can arise that impact protein translation fidelity, leading to harmful consequences such as neurological disorders[28-30] and cancer.[31-32] Therefore, it becomes important to remove damaged/modified nucleotide sites, a process that involves breaking the phosphodiester backbone.

Although phosphodiester bonds must be cleaved to repair DNA damage and maintain modification balance in RNA, these bonds have an enormous resistance to hydrolysis, taking up to 110 years for RNA and 30 million years for DNA hydrolysis.[1-2] Therefore, specialized enzymes that drastically change the hydrolysis barrier are required to make the P–O bond dissociation feasible. This thesis examines the phosphodiester bond cleavage mechanisms employed by enzymes called nucleases which are discussed in the following section.

1.4. Nucleases

Nucleases form a broad family of enzymes that facilitate the challenging phosphodiester bond cleavage in nucleic acids, enhancing the rate by ~10^{17}-fold over the uncatalyzed reaction.[3, 33] Nucleases can be classified into various categories depending on their function, such as DNA endonucleases (DNA repair),[34] homing endonucleases (genetic exchange), tyrosine or serine recombinases (DNA recombination),[35] topoisomerases (structural alteration of nucleic acids),[36] group I or II introns (RNA splicing),[37-38] ribonucleases (RNA metabolism and interference),[39] 5′ to 3′ RNA exonucleases (RNA processing),[40] and DNases (programmed cell death).[41] Nucleases have been successfully manipulated and reprogrammed for applications in the field of genome engineering and biotechnology.[4, 18, 42-43] Additionally, defective nuclease function has been related to a variety of diseases ranging from auto-immune disorders to cancers.[44-46] Therefore, a deeper understanding of the function of nucleases is critical to create new applications in gene therapy, and design therapeutic solutions to treat a host of diseases. This thesis focuses on understanding the mechanism of action of select nucleases that are involved in DNA repair,[47-48] RNA metabolism,[48-49] and genetic exchange.[50-51] These nucleases are known to

use metal ions to mediate catalysis; however, there are gaps in the literature regarding the number of metal ions required for P–O bond cleavage, which is discussed in detail in the subsequent section.

1.5. Metal Dependence of Nucleases

Although there is a plethora of metal ions that can be used as cofactors in metalloenzymes, Mg^{2+} is considered to be the most common and versatile metal cofactor among nucleases.[33] This specificity for Mg^{2+} mainly arises due to properties like the high natural abundance of this cation in cells, the high charge density due to the small ionic radii, its redox inertness, slow solvent exchange rates, and strict coordination geometry and ligand requirements.[52] The metal co-factors not only influence substrate recognition and binding, but can be directly involved in the chemical step.[53] Although an overwhelming majority of nucleases are known to use metals to enhance catalysis,[33] the role and number of metal ions required for the chemical step has been a topic of constant debate and controversy in the literature.[53-56] The two-metal mediated pathway is the general proposed mechanism for P–O bond cleavage for the majority of the nucleases.[57-59] The first metal (M_A^{2+}) activates the nucleophilic water, while the second metal (M_B^{2+}) promotes leaving group departure either through direct coordination or using a metal-activated water, and the substrate charge stabilization is provided by both metals (Figure 1.3a).[57-59] However, the universality of the two-metal mediated mechanism has been called into question based on studies investigating the metal concentration dependence of nucleases in solution.[53, 60-64] For example, despite two metals being present in the active sites of X-ray crystal structures of several nucleases,[65-69] experimental kinetic studies (cleavage rates obtained as a function of metal ion concentration) indicate these enzymes follow single-metal-mediated

catalysis,[63-64] which underscores the potential for inconsistencies between kinetic and structural data. These experimental observations suggest that although endonucleases can have two metal-binding sites, some enzymes require only one metal for phosphodiester bond cleavage. In these cases, the role of the first metal in nucleophile activation must be taken by a basic amino acid in the active site (Figure 1.3b). This proposal is further corroborated by the emergence of a rising number of crystal structures of nucleases containing a single metal in the active site.[33, 70-75] Nevertheless, the conflicting evidence from X-ray crystallography and solution studies highlights the need for more comprehensive insight into the role and number of metals required in the phosphodiester bond hydrolysis mediated by nucleases.

An abundance of experimental work on nucleases using different methods like X-ray crystallography,[49, 70-74, 76] kinetic,[62, 77-80] and mutational studies[81-86] has provided valuable information regarding the structure of enzymes, enzyme–substrate binding affinity, catalytic rates, and the identity of key residues involved in catalysis. However, these methods cannot give a full picture of the chemical mechanism used for phosphodiester bond hydrolysis. Computer modeling is a powerful tool for capturing atomic-level details of reaction surfaces and characterizing key stationary points, particularly transition states, which are hard to conclusively characterize experimentally. This thesis performs detailed investigations into the chemistry of nucleases to understand their mechanism of action using computer modeling techniques that are briefly described in the following section.

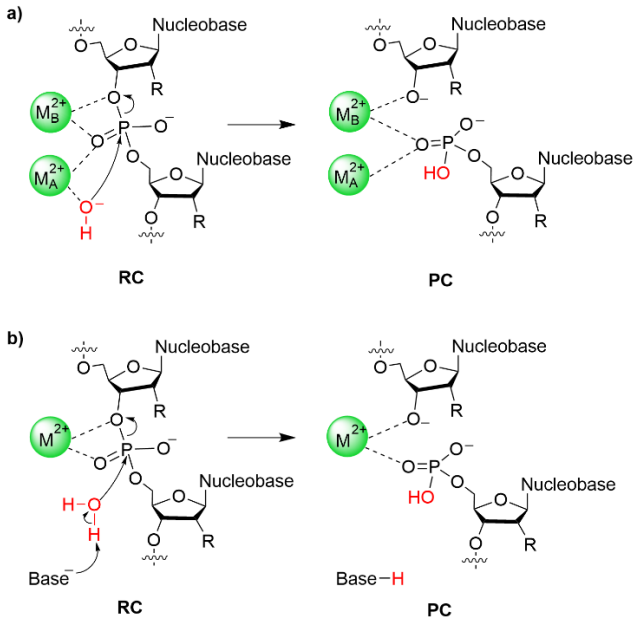

Figure 1.3. The general proposed mechanisms for a) two-metal and b) one-metal mediated phosphodiester bond cleavage (R = H for DNA or OH for RNA).

1.6. Computer Modeling of Nucleases

There are a variety of computational tools available to obtain an atomic-level understanding of nucleases. One such tool is a quantum mechanics (QM) cluster approach in which a limited number of atoms that represent the active site are extracted from the X-ray crystal structure of the enzyme, while the periphery atoms are fixed to the experimental structure coordinates and the model is placed in a dielectric cavity representing the enzyme surroundings.[87-88] All atoms in the cluster model are treated using an accurate density functional theory (DFT) method. Molecular dynamics (MD) simulations are another useful tool that use classical molecular mechanics (MM) principles to propagate a system through time and study the structural dynamics of complex biomolecular systems under physiological conditions.[89] Finally, quantum mechanical–molecular mechanics (QM/MM) techniques are another effective computational approach to characterize enzyme-catalyzed

reaction mechanisms in which the active site of the enzyme is described using high-level QM methods, while the rest of the enzyme−nucleic acid complex is described by cost-effective MM methods.[90-91]

QM cluster modeling, MD simulations, and QM/MM calculations have been successfully applied in the literature to investigate the catalytic mechanism of several two-metal dependent nucleases such as bacterial phosphotriesterase,[92] RNase H,[93-94] BamHI,[95] CRISPR-Cas9,[96] retroviral integrase,[97] HIV-1 reverse transcriptase,[98] and HIV-integrase.[99] These studies have provided insight into how the phosphodiester bond hydrolysis is mediated by two metals and clarified the catalytic role of the metals and select active site residues.

In contrast, theoretical studies on single-metal mediated catalysis are scarce.[100] Only a single computational study exists for a single-metal dependent enzyme that used a combined QM cluster and QM/MM approach on APE1 to elucidate the phosphodiester bond cleavage pathway and reveal the identity of the general base and general acid as well as the role of the metal in the reaction.[100] The fundamental understanding gained from the catalytic mechanism of APE1 can be applied to investigate other single-metal dependent nucleases for which the mechanism has not been firmly established using computational methods or even proposed based on experimental techniques. Overall, this thesis describes a comprehensive investigation of catalytic pathways and the metal-dependence of select nucleases to fully understand the intricacies of the phosphodiester bond cleavage reaction and gain insight into whether nucleases use a unified mechanism, or if different active site compositions result in differential catalytic mechanisms.

1.7. Scope of the Thesis

This thesis uses a multi-scale computational approach to gain an atomic-level understanding of the phosphodiester bond cleavage reaction catalyzed by nucleases. Specifically, Chapter 2 uses a range of cluster models to clarify how to efficiently model the chemistry mediated by nucleases that invoke only a single metal ion. APE1 (Figure 1.4a) was used as a prototypical example due to the availability of abundant experimental crystallographic,[47, 75, 101-102] kinetic,[75, 103] mutational,[75, 101-102, 104] and computational data.[100] This chapter underscores the importance of correct design of the cluster model size, and the DFT functional and basis set combination required for an accurate structural description of the one-metal mediated phosphodiester bond hydrolysis. The DFT functional and basis set combination recommended from this chapter is used for the rest of the chapters in this thesis.

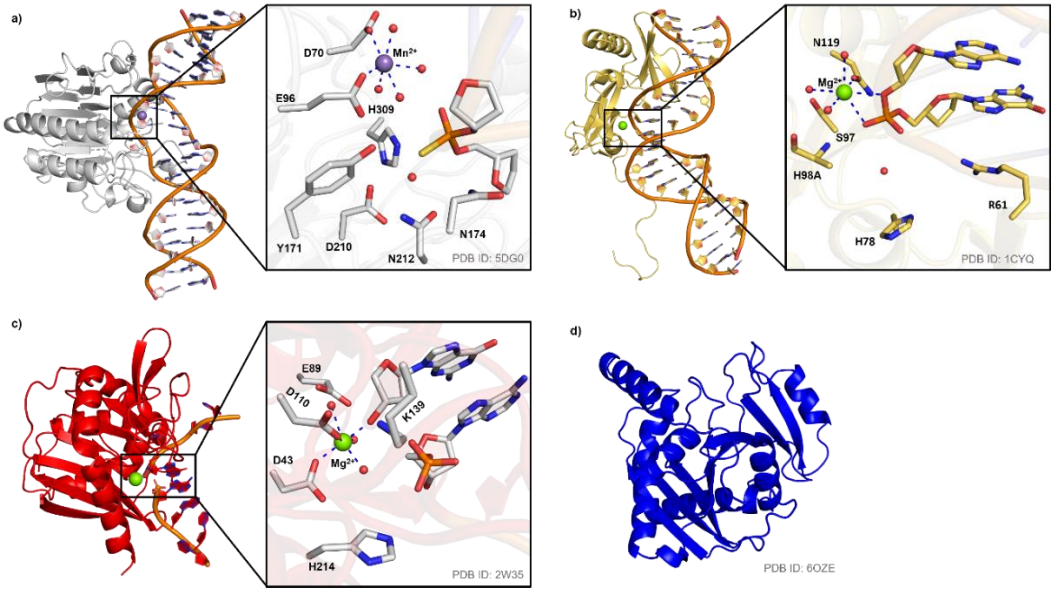

Figure 1.4. Enzymes (left) and the corresponding active sites (right) from the X-ray crystal structures of a: a) RC of thio-substituted APE1 bound to a dsDNA substrate, b) RC of the H98A mutant of I-*Ppo*I bound to a dsDNA substrate, c) PC of the wild-type bacterial EndoV bound to a single-stranded DNA substrate, and d) apo-human EndoV.

Chapter 3 uses a QM/MM approach to understand the metal dependence of APE1. Despite using Mg^{2+} as the native metal cofactor, APE1 has also been shown to remain catalytically active in the presence of several metals, with the rate decreasing as $Mg^{2+} > Mn^{2+} > Ni^{2+} > Zn^{2+}$, while Ca^{2+} completely abolished the catalytic activity.[105-107] However, a chemical explanation for this metal dependence is still missing in the literature. This chapter provides a structural rationalization and reproduces the experimentally-observed trend in the barrier heights for the metal dependence of APE1. Furthermore, this chapter provides clarification on how Ca^{2+} inhibits single-metal mediated APE1.

Chapter 4 uses both QM cluster and QM/MM approaches to elucidate the catalytic pathway used by I-*Ppo*I, which displays a single metal (Mg^{2+}) in the active site of the X-ray crystal structures (Figure 1.4b).[70, 76] I-*Ppo*I belongs to the homing endonuclease (HE) family that transfers genes that encode the HE into a host genome.[51] Although a catalytic mechanism has been proposed based on the crystal structures[70] and mutational data,[77] the roles of the metal and key amino acids need to be clarified in addition to the required computational verification of the experimentally-proposed reaction pathway. Through comparison to Chapter 3, Chapter 4 also sheds light on how a single metal can play diverse roles depending on the active site environment of the enzyme.

Chapter 5 uses a multi-pronged computational approach including MD simulations and QM/MM calculations to provide the first proposed mechanism for bacterial EndoV. Bacterial EndoV is a single metal (Mg^{2+}) containing nuclease (Figure 1.4c)[74] that removes deaminated DNA and RNA nucleobases by cleaving the phosphodiester bond one nucleotide away from a deaminated nucleobase in the 3′ direction.[108] Despite the availability of a crystal structure and abundant kinetic and mutational data,[62, 74, 79, 109] a

catalytic mechanism has not been proposed in the literature. Chapter 5 uncovers a novel catalytic mechanism for bacterial EndoV that uses a previously unseen combination of general base, general acid, and metal-binding architecture for a single-metal dependent nuclease.

Chapter 6 uses MD simulations and QM/MM calculations to provide the first details of the mechanism of action of human EndoV, which removes deaminated RNA nucleobases as a part of RNA metabolism.[110-111] Unlike its bacterial counterpart, the crystal structure of human EndoV (Figure 1.4d) does not contain the substrate or the metal cofactor/s.[49] Therefore, a catalytic mechanism for the phosphodiester bond cleavage has never been proposed. Furthermore, there have been no experimental or computational studies that have examined the role or number of metal/s required for catalysis. The prediction of metal dependence is further complicated by the fact that bacterial EndoV has only one metal (Mg^{2+}) in its active site, while mouse EndoV has two metals (Mg^{2+}) in its active site.[49, 74] Despite the sequence conservation among EndoV species, the usage of a different number of metals to facilitate the same phosphodiester bond hydrolysis reaction raises questions about the number of metal/s and overall mechanistic pathway used by human EndoV. This chapter performs a comprehensive investigation of one- versus two-metal mediated phosphodiester bond cleavage pathways to afford the first proposed mechanism of action for human EndoV.

Finally, Chapter 7 summarizes how this thesis used a multi-pronged computational approach to emphasize the intricate details involved in cleaving the phosphodiester bond in nucleic acids and illustrates that the metal can play different roles to facilitate the same reaction depending on the active site composition of the nuclease. This chapter also

provides possible directions for future investigations of other metal-dependent enzymes that can be built based on the fundamental understanding gained from this thesis.

1.8. References

(1) Mikkola, S.; Lönnberg, T.; Lönnberg, H., Phosphodiester models for cleavage of nucleic acids. *Beilstein J. Org. Chem.* **2018,** *14* (1), 803–837.

(2) Schroeder, G. K.; Lad, C.; Wyman, P.; Williams, N. H.; Wolfenden, R., The time required for water attack at the phosphorus atom of simple phosphodiesters and of DNA. *Proc. Natl. Acad. Sci. U.S.A.* **2006,** *103* (11), 4052–4055.

(3) Radzicka, A.; Wolfenden, R., A proficient enzyme. *Science* **1995,** *267* (5194), 90–3.

(4) Huang, J.; Kirk, B.; Favis, R.; Soussi, T.; Paty, P.; Cao, W.; Barany, F., An endonuclease/ligase based mutation scanning method especially suited for analysis of neoplastic tissue. *Oncogene* **2002,** *21* (12), 1909–1921.

(5) Wang, Z.; Wang, H.-Y.; Feng, H., A simple and reproducible method for directed evolution: combination of random mutation with dITP and DNA fragmentation with endonuclease V. *Mol. Biotechnol.* **2013,** *53*, 49–54.

(6) Kong, X. Y.; Vik, E. S.; Nawaz, M. S.; Berges, N.; Dahl, T. B.; Vågbø, C.; Suganthan, R.; Segers, F.; Holm, S.; Quiles-Jiménez, A., Deletion of Endonuclease V suppresses chemically induced hepatocellular carcinoma. *Nucleic Acids Res.* **2020,** *48* (8), 4463–4479.

(7) Knutson, S. D.; Arthur, R. A.; Johnston, H. R.; Heemstra, J. M., Selective enrichment of A-to-I edited transcripts from cellular RNA using Endonuclease V. *J. Am. Chem. Soc.* **2020,** *142* (11), 5241–5251.

(8) Knutson, S. D.; Arthur, R. A.; Johnston, H. R.; Heemstra, J. M., Direct immunodetection of global A-to-I RNA editing activity with a chemiluminescent bioassay. *Angew. Chem. Int. Ed.* **2021,** *60* (31), 17009–17017.

(9) Chang, Y.; Huang, Z.; Quan, H.; Li, H.; Yang, S.; Song, Y.; Wang, J.; Yuan, J.; Wu, C., Construction of a DNA damage repair gene signature for predicting prognosis and immune response in breast cancer. *Front. Oncol.* **2022,** *12*, 1085632.

(10) Caston, R. A.; Gampala, S.; Armstrong, L.; Messmann, R. A.; Fishel, M. L.; Kelley, M. R., The multifunctional APE1 DNA repair–redox signaling protein as a drug target in human disease. *Drug Discov. Today* **2021,** *26* (1), 218–228.

(11) Caston, R. A.; Shah, F.; Starcher, C. L.; Wireman, R.; Babb, O.; Grimard, M.; McGeown, J.; Armstrong, L.; Tong, Y.; Pili, R., Combined inhibition of Ref-1 and STAT3 leads to synergistic tumour inhibition in multiple cancers using 3D and in vivo tumour co-culture models. *J. Cell. Mol. Med.* **2021,** *25* (2), 784–800.

(12) Hartman, G. D.; Lambert-Cheatham, N. A.; Kelley, M. R.; Corson, T. W., Inhibition of APE1/Ref-1 for neovascular eye diseases: From biology to therapy. *Int. J. Mol. Sci.* **2021,** *22* (19), 10279.

(13) Heisel, C.; Yousif, J.; Mijiti, M.; Charizanis, K.; Brigell, M.; Corson, T. W.; Kelley, M. R., APE1/Ref-1 as a novel target for retinal diseases. *J. Cell. Signal.* **2021,** *2* (2), 133.

(14) Long, K.; Gu, L.; Li, L.; Zhang, Z.; Li, E.; Zhang, Y.; He, L.; Pan, F.; Guo, Z.; Hu, Z., Small-molecule inhibition of APE1 induces apoptosis, pyroptosis, and necroptosis in non-small cell lung cancer. *Cell Death Dis.* **2021,** *12* (6), 1–15.

(15) Rajkumar, A. P.; Christensen, J. H.; Mattheisen, M.; Jacobsen, I.; Bache, I.; Pallesen, J.; Grove, J.; Qvist, P.; McQuillin, A.; Gurling, H. M., Analysis of t (9; 17)(q33. 2; q25. 3) chromosomal breakpoint regions and genetic association reveals novel candidate genes for bipolar disorder. *Bipolar Disord.* **2015,** *17* (2), 205–211.

(16) Kong, X. Y.; Huse, C.; Yang, K.; Øgaard, J.; Berges, N.; Vik, E. S.; Nawaz, M. S.; Quiles-Jiménez, A.; Abbas, A.; Gregersen, I., Endonuclease V regulates atherosclerosis through C-C motif chemokine ligand 2-mediated monocyte infiltration. *J. Am. Heart Assoc.* **2021,** *10* (14), e020656.

(17) Marcaida, M. J.; Muñoz, I. G.; Blanco, F. J.; Prieto, J.; Montoya, G., Homing endonucleases: from basics to therapeutic applications. *Cell. Mol. Life Sci.* **2010,** *67*, 727–748.

(18) Belfort, M.; Bonocora, R. P., Homing endonucleases: from genetic anomalies to programmable genomic clippers. *Methods Mol. Biol.* **2014,** *1123*, 1–26.

(19) Srinivas, U. S.; Tan, B. W. Q.; Vellayappan, B. A.; Jeyasekharan, A. D., ROS and the DNA damage response in cancer. *Redox Biol.* **2019,** *25*, 101084.

(20) Shrivastav, N.; Li, D.; Essigmann, J. M., Chemical biology of mutagenesis and DNA repair: cellular responses to DNA alkylation. *Carcinogenesis* **2010,** *31* (1), 59–70.

(21) Quezada-Maldonado, E. M.; Sánchez-Pérez, Y.; Chirino, Y. I.; García-Cuellar, C. M., Airborne particulate matter induces oxidative damage, DNA adduct formation and alterations in DNA repair pathways. *Environ. Pollut.* **2021,** *287*, 117313.

(22) Rastogi, R. P.; Richa; Kumar, A.; Tyagi, M. B.; Sinha, R. P., Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. *J. Nucleic Acids* **2010,** *2010*, 592980.

(23) Huang, R.; Zhou, P.-K., DNA damage repair: historical perspectives, mechanistic pathways and clinical translation for targeted cancer therapy. *Signal Transduct. Target. Ther.* **2021,** *6* (1), 254.

(24) 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. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2020,** *10* (5), e1471.

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

(26) Jackson, S. P.; Bartek, J., The DNA-damage response in human biology and disease. *Nature* **2009,** *461* (7267), 1071–1078.

(27) Farajollahi, S.; Maas, S., Molecular Diversity through RNA Editing: A Balancing Act. *Trends Genet.* **2010,** *26* (5), 221–230.

(28) Srivastava, P. K.; Bagnati, M.; Delahaye-Duriez, A.; Ko, J.-H.; Rotival, M.; Langley, S. R.; Shkura, K.; Mazzuferi, M.; Danis, B.; van Eyll, J., Genome-wide analysis of differential RNA editing in epilepsy. *Genome Res.* **2017,** *27* (3), 440–450.

(29) Khermesh, K.; D'Erchia, A. M.; Barak, M.; Annese, A.; Wachtel, C.; Levanon, E. Y.; Picardi, E.; Eisenberg, E., Reduced levels of protein recoding by A-to-I RNA editing in Alzheimer's disease. *RNA* **2016,** *22* (2), 290–302.

(30) Hosaka, T.; Tsuji, H.; Kwak, S., RNA editing: a new therapeutic target in amyotrophic lateral sclerosis and other neurological diseases. *Int. J. Mol. Sci.* **2021,** *22* (20), 10958.

(31) Paz, N.; Levanon, E. Y.; Amariglio, N.; Heimberger, A. B.; Ram, Z.; Constantini, S.; Barbash, Z. S.; Adamsky, K.; Safran, M.; Hirschberg, A., Altered adenosine-to-inosine RNA editing in human cancer. *Genome Res.* **2007,** *17* (11), 1586–1595.

(32) Han, L.; Diao, L.; Yu, S.; Xu, X.; Li, J.; Zhang, R.; Yang, Y.; Werner, H. M.; Eterovic, A. K.; Yuan, Y., The genomic landscape and clinical relevance of A-to-I RNA editing in human cancers. *Cancer cell* **2015,** *28* (4), 515–528.

(33) Yang, W., Nucleases: diversity of structure, function and mechanism. *Q. Rev. Biophys.* **2011,** *44* (1), 1–93.

(34) Loenen, W. A.; Dryden, D. T.; Raleigh, E. A.; Wilson, G. G.; Murray, N. E., Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Res.* **2014,** *42* (1), 3–19.

(35) Grindley, N. D.; Whiteson, K. L.; Rice, P. A., Mechanisms of site-specific recombination. *Annu. Rev. Biochem.* **2006,** *75*, 567–605.

(36) Bizard, A. H.; Hickson, I. D., The many lives of type IA topoisomerases. *J. Biol. Chem.* **2020,** *295* (20), 7138–7153.

(37) Lehmann, K.; Schmidt, U., Group II introns: structure and catalytic versatility of large natural ribozymes. *Crit. Rev. Biochem. Mol. Biol.* **2003,** *38* (3), 249–303.

(38) Hausner, G.; Hafez, M.; Edgell, D. R., Bacterial group I introns: mobile RNA catalysts. *Mob. DNA* **2014,** *5* (1), 8.

(39) Bechhofer, D. H.; Deutscher, M. P., Bacterial ribonucleases and their roles in RNA metabolism. *Crit. Rev. Biochem. Mol. Biol.* **2019,** *54* (3), 242–300.

(40) Maraia, R. J.; Lamichhane, T. N., 3′ processing of eukaryotic precursor tRNAs. *Wiley Interdiscip. Rev. RNA* **2011,** *2* (3), 362–375.

(41) Lauková, L.; Konečná, B.; Janovičová, Ľ.; Vlková, B.; Celec, P., Deoxyribonucleases and their applications in biomedicine. *Biomolecules* **2020,** *10* (7), 1036.

(42) Pincas, H.; Pingle, M. R.; Huang, J.; Lao, K.; Paty, P. B.; Friedman, A. M.; Barany, F., High sensitivity EndoV mutation scanning through real-time ligase proofreading. *Nucleic Acids Res.* **2004,** *32* (19), e148.

(43) 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.

(44) Crow, Y. J.; Rehwinkel, J., Aicardi-Goutieres syndrome and related phenotypes: linking nucleic acid metabolism with autoimmunity. *Hum. Mol. Genet.* **2009,** *18* (R2), R130–R136.

(45) Santa, P.; Garreau, A.; Serpas, L.; Ferriere, A.; Blanco, P.; Soni, C.; Sisirak, V., The role of nucleases and nucleic acid editing enzymes in the regulation of self-nucleic acid sensing. *Front. immunol.* **2021,** *12*, 629922.

(46) Balian, A.; Hernandez, F. J., Nucleases as molecular targets for cancer diagnosis. *Biomark. Res.* **2021,** *9* (1), 1–16.

(47) Whitaker, A. M.; Flynn, T. S.; Freudenthal, B. D., Molecular snapshots of APE1 proofreading mismatches and removing DNA damage. *Nat. Commun.* **2018,** *9* (1), 399.

(48) Cao, W., Endonuclease V: an unusual enzyme for repair of DNA deamination. *Cell. Mol. Life Sci.* **2013,** *70* (17), 3145–3156.

(49) 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.

(50) Belfort, M.; Roberts, R. J., Homing endonucleases: keeping the house in order. *Nucleic Acids Res.* **1997,** *25*, 3379–88.

(51) Stoddard, B. L., Homing endonucleases from mobile group I introns: discovery to genome engineering. *Mob. DNA* **2014,** *5*, 1–16.

(52) Cowan, J. A., Metal Activation of Enzymes in Nucleic Acid Biochemistry. *Chem. Rev. (Washington, D. C.)* **1998,** *98* (3), 1067–1087.

(53) 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.

(54) Yang, W., An equivalent metal ion in one- and two-metal-ion catalysis. *Nat. Struct. Mol. Biol.* **2008,** *15* (11), 1228–1231.

(55) Dupureur, C. M., Roles of metal ions in nucleases. *Curr. Opin. Chem. Biol.* **2008,** *12* (2), 250–255.

(56) Dupureur, C. M., An integrated look at metallonuclease mechanism. *Curr. Chem. Biol.* **2008,** *2* (2), 159–173.

(57) 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), 6498–502.

(58) 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.

(59) 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.

(60) Cowan, J. A., Metal-mediated hydrolysis of biological phosphate esters: a critical analysis of the essential metal ion stoichiometry for magnesium-dependent nuclease activation. *J. Biol. Inorg. Chem.* **1997,** *2* (2), 168–176.

(61) Groll, D. H.; Jeltsch, A.; Selent, U.; Pingoud, A., Does the restriction endonuclease *Eco*RV employ a two-metal-ion mechanism for DNA cleavage? *Biochemistry* **1997,** *36* (38), 11389–11401.

(62) Feng, H.; Dong, L.; Cao, W., Catalytic mechanism of endonuclease V: a catalytic and regulatory two-metal model. *Biochemistry* **2006,** *45* (34), 10251–10259.

(63) 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 *Eco*RI family. *J. Mol. Biol.* **2009,** *393* (1), 140–160.

(64) Xie, F.; Qureshi, S. H.; Papadakos, G. A.; Dupureur, C. M., One- and two-metal ion catalysis: global single-turnover kinetic analysis of the PvuII endonuclease mechanism. *Biochemistry* **2008,** *47* (47), 12540–12550.

(65) Viadiu, H.; Aggarwal, A. K., The role of metals in catalysis by the restriction endonuclease Bam HI. *Nat. Struct. Biol.* **1998,** *5* (10), 910–916.

(66) Deibert, M.; Grazulis, S.; Sasnauskas, G.; Siksnys, V.; Huber, R., Structure of the tetrameric restriction endonuclease NgoMIV in complex with cleaved DNA. *Nat. Struct. Biol.* **2000,** *7* (9), 792–799.

(67) Horton, J. R.; Cheng, X., PvuII endonuclease contains two calcium ions in active sites. *J. Mol. Biol.* **2000,** *300* (5), 1049–1056.

(68) Dunten, P. W.; Little, E. J.; Gregory, M. T.; Manohar, V. M.; Dalton, M.; Hough, D.; Bitinaite, J.; Horton, N. C., The structure of SgrAI bound to DNA; recognition of an 8 base pair target. *Nucleic Acids Res.* **2008,** *36* (16), 5405–5416.

(69) Lambert, A. R.; Sussman, D.; Shen, B.; Maunus, R.; Nix, J.; Samuelson, J.; Xu, S.-Y.; Stoddard, B. L., Structures of the rare-cutting restriction endonuclease NotI reveal a unique metal binding fold involved in DNA binding. *Structure* **2008,** *16* (4), 558–569.

(70) Galburt, E. A.; Chevalier, B.; Tang, W.; Jurica, M. S.; Flick, K. E.; Monnat, R. J.; Stoddard, B. L., A novel endonuclease mechanism directly visualized for I-*Ppo*I. *Nat. Struct. Biol.* **1999,** *6* (12), 1096–1099.

(71) 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.

(72) Shen, B. W.; Landthaler, M.; Shub, D. A.; Stoddard, B. L., DNA binding and cleavage by the HNH homing endonuclease I-HmuI. *J. Mol. Biol.* **2004,** *342* (1), 43–56.

(73) Biertümpfel, C.; Yang, W.; Suck, D., Crystal structure of T4 endonuclease VII resolving a Holliday junction. *Nature* **2007,** *449* (7162), 616–620.

(74) Dalhus, B.; Arvai, A. S.; Rosnes, I.; Olsen, Ø. E.; Backe, P. H.; Alseth, I.; Gao, H.; Cao, W.; Tainer, J. A.; Bjørås, M., Structures of endonuclease V with DNA reveal initiation of deaminated adenine repair. *Nat. Struct. Mol. Biol.* **2009,** *16*, 138–143.

(75) Freudenthal, B. D.; Beard, W. A.; Cuneo, M. J.; Dyrkheeva, N. S.; Wilson, S. H., Capturing snapshots of APE1 processing DNA damage. *Nat. Struct. Mol. Biol.* **2015,** *22* (11), 924–931.

(76) Flick, K. E.; Jurica, M. S.; Monnat, R. J.; Stoddard, B. L., DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-*Ppo*I. *Nature* **1998,** *394* (6688), 96–101.

(77) Mannino, S. J.; Jenkins, C. L.; Raines, R. T., Chemical mechanism of DNA cleavage by the homing endonuclease I-*Ppo*I. *Biochemistry* **1999,** *38* (49), 16178–16186.

(78) Eastberg, J. H.; Eklund, J.; Monnat, R.; Stoddard, B. L., Mutability of an HNH nuclease imidazole general base and exchange of a deprotonation mechanism. *Biochemistry* **2007,** *46* (24), 7215–7225.

(79) Feng, H.; Klutz, A. M.; Cao, W., Active site plasticity of endonuclease V from Salmonella typhimurium. *Biochemistry* **2005,** *44* (2), 675–683.

(80) Howard, M. J.; Klemm, B. P.; Fierke, C. A., Mechanistic studies reveal similar catalytic strategies for phosphodiester bond hydrolysis by protein-only and RNA-dependent ribonuclease P. *J. Biol. Chem.* **2015,** *290* (21), 13454–13464.

(81) Crépin, T.; Dias, A.; Palencia, A.; Swale, C.; Cusack, S.; Ruigrok, R. W., Mutational and metal binding analysis of the endonuclease domain of the influenza virus polymerase PA subunit. *J. Virol.* **2010,** *84* (18), 9096–9104.

(82) Friedhoff, P.; Kolmes, B.; Gimadutdinow, O.; Wende, W.; Krause, K. L.; Pingoud, A., Analysis of the mechanism of the *Serratia* nuclease using site-directed mutagenesis. *Nucleic Acids Res.* **1996,** *24* (14), 2632–2639.

(83) Giraud-Panis, M.-J. E.; Lilley, D. M., T4 endonuclease VII: importance of a histidine-aspartate cluster within the zinc-binding domain. *J. Biol. Chem.* **1996,** *271* (51), 33148–33155.

(84) Feng, H.; Dong, L.; Klutz, A. M.; Aghaebrahim, N.; Cao, W., Defining amino acid residues involved in DNA-protein interactions and revelation of 3'-Exonuclease activity in Endonuclease V. *Biochemistry* **2005,** *44* (34), 11486–11495.

(85) Seligman, L. M.; Chisholm, K. M.; Chevalier, B. S.; Chadsey, M. S.; Edwards, S. T.; Savage, J. H.; Veillet, A. L., Mutations altering the cleavage specificity of a homing endonuclease. *Nucleic Acids Res.* **2002,** *30* (17), 3870–3879.

(86) Saravanan, M.; Vasu, K.; Nagaraja, V., Evolution of sequence specificity in a restriction endonuclease by a point mutation. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** *105* (30), 10344–10347.

(87) 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.

(88) Himo, F., Recent trends in quantum chemical modeling of enzymatic reactions. *J. Am. Chem. Soc.* **2017,** *139* (20), 6780–6786.

(89) Leach, A. R., Molecular modelling: principles and applications. Pearson education: 2001.

(90) Chung, L. W.; Hirao, H.; Li, X.; Morokuma, K., The ONIOM method: its foundation and applications to metalloenzymes and photobiology. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012,** *2* (2), 327–350.

(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.; Li, H.-B.; Ding, L.; Morokuma, K., The ONIOM method and its applications. *Chem. Rev.* **2015,** *115* (12), 5678–5796.

(92) Alberto, M. E.; Pinto, G.; Russo, N.; Toscano, M., Triesterase and promiscuous diesterase activities of a di-CoII-containing organophosphate degrading enzyme reaction mechanisms. *Chem. Eur. J.* **2015,** *21* (9), 3736–3745.

(93) 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.

(94) Rosta, E.; Yang, W.; Hummer, G., Calcium inhibition of ribonuclease H1 two-metal ion catalysis. *J. Am. Chem. Soc.* **2014,** *136* (8), 3137–3144.

(95) Mordasini, T.; Curioni, A.; Andreoni, W., Why do divalent metal ions either promote or inhibit enzymatic reactions? The case of BamHI restriction endonuclease from combined quantum-classical simulations. *J. Biol. Chem.* **2003,** *278* (7), 4381–4384.

(96) Casalino, L.; Nierzwicki, L.; 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.

(97) Araujo, A. R.; Ribeiro, A. J.; 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.

(98) Rungrotmongkol, T.; Mulholland, A. J.; Hannongbua, S., QM/MM simulations indicate that Asp185 is the likely catalytic base in the enzymatic reaction of HIV-1 reverse transcriptase. *MedChemComm* **2014,** *5* (5), 593–596.

(99) 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.

(100) 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.

(101) Tsutakawa, S. E.; Shin, D. S.; Mol, C. D.; Izumi, T.; Arvai, A. S.; Mantha, A. K.; Szczesny, B.; Ivanov, I. N.; Hosfield, D. J.; Maiti, B.; Pique, M. E.; Frankel, K. A.; Hitomi, K.; Cunningham, R. P.; Mitra, S.; Tainer, J. A., Conserved structural chemistry for incision activity in structurally non-homologous apurinic/apyrimidinic endonuclease APE1 and endonuclease IV DNA repair enzymes. *J. Biol. Chem.* **2013,** *288* (12), 8445–8455.

(102) He, H.; Chen, Q.; Georgiadis, M. M., High-resolution crystal structures reveal plasticity in the metal binding site of apurinic/apyrimidinic endonuclease 1. *Biochemistry* **2014,** *53* (41), 6520–6529.

(103) Kim, W.-C.; Berquist, B. R.; Chohan, M.; Uy, C.; Wilson, D. M.; Lee, C. H., Characterization of the endoribonuclease active site of human Apurinic/Apyrimidinic Endonuclease 1. *J. Mol. Biol.* **2011,** *411* (5), 960–971.

(104) Kanazhevskaya, L. Y.; Koval, V. V.; Lomzov, A. A.; Fedorova, O. S., The role of Asn-212 in the catalytic mechanism of human endonuclease APE1: Stopped-flow kinetic study

of incision activity on a natural AP site and a tetrahydrofuran analogue. *DNA Repair* **2014,** *21*, 43–54.

(105) Barzilay, G.; Mol, C. D.; Robson, C. N.; Walker, L. J.; Cunningham, R. P.; Tainer, J. A.; Hickson, I. D., Identification of critical active-site residues in the multifunctional human DNA repair enzyme HAP1. *Nat. Struct. Mol. Biol.* **1995,** *2* (7), 561–568.

(106) Schermerhorn, K. M.; Delaney, S., A chemical and kinetic perspective on base excision repair of DNA. *Acc. Chem. Res.* **2015,** *47*, 1238.

(107) Miroshnikova, A. D.; Kuznetsova, A. A.; Vorobjev, Y. N.; Kuznetsov, N. A.; Fedorova, O. S., Effects of mono- and divalent metal ions on DNA binding and catalysis of human apurinic/apyrimidinic endonuclease 1. *Mol. BioSyst.* **2016,** *12* (5), 1527–1539.

(108) Dalhus, B.; Arvai, A. S.; Rosnes, I.; Olsen, O. E.; Backe, P. H.; Alseth, I.; Gao, H.; Cao, W.; Tainer, J. A.; Bjoeraas, M., Structures of endonuclease V with DNA reveal initiation of deaminated adenine repair. *Nat. Struct. Mol. Biol.* **2009,** *16* (2), 138-143.

(109) Feng, H.; Dong, L.; Klutz, A. M.; Aghaebrahim, N.; Cao, W., Defining Amino Acid Residues Involved in DNA−Protein Interactions and Revelation of 3‘-Exonuclease Activity in Endonuclease V. *Biochemistry* **2005,** *44* (34), 11486–11495.

(110) Morita, Y.; Shibutani, T.; Nakanishi, N.; Nishikura, K.; Iwai, S.; Kuraoka, I., Human endonuclease V is a ribonuclease specific for inosine-containing RNA. *Nat. Commun.* **2013,** *4*, 2273.

(111) Vik, E. S.; Nawaz, M. S.; Andersen, P. S.; Fladeby, C.; Bjoras, M.; Dalhus, B.; Alseth, I., Endonuclease V cleaves at inosines in RNA. *Nat. Commun.* **2013,** *4*, 2271.

Chapter 2: 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

Preface: Chapter 2 has been published [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]. In this chapter, all authors contributed to the reviewing and editing of the manuscript. I aided conceptualization of the project and performed the computational characterization of the phosphodiester bond cleavage pathways (QM cluster optimizations in gas-phase and single-point calculations in both gas and solvent phase) using models 0, 1, 4, 5, 6, 8–10. In addition, I performed the overall data/results analysis, wrote the first draft of the manuscript, and generated all figures for the publication. D. J. Nikkel performed optimizations (gas-phase) and solvent single-point calculations (gas and solvent phase) using models 2, 3, and 7, investigated the impact of different truncation points using model 0, and contributed to the formal data analysis. M. M. Aboelnga contributed towards conceptualization, project administration, and supervision, while S. D. Wetmore was involved in project conceptualization and administration, supervision, funding and resource acquisition, data visualization and interpretation, and writing the original draft of the manuscript.

2.1. Introduction

Although the stability of DNA and RNA is necessary for storing and regulating genetic information, the cleavage of phosphodiester bonds in nucleic acids is required for many cellular processes including signaling, energy transduction, protein synthesis, and maintaining genomic integrity.[1-2] As a specific example, the repair of damaged DNA involves breaking the phosphodiester bond near the lesion site,[3] which is essential to combat the effects of exposure to harmful substances (such as tobacco smoke,[4] UV radiation,[5] processed meats,[6] pollution[7] and chemotherapeutics) that lead to cancers,[8] cardiovascular diseases,[9] and neurodegeneration.[10] Despite the reaction being biochemically ubiquitous, the uncatalyzed hydrolysis of nucleic acid phosphodiester bonds is exceedingly slow, with a half-life ranging from 110 years for RNA to 30 million years for DNA under mild conditions.[11]

Nucleases are a broad class of enzymes that can afford 10^{17}-fold rate enhancements over the uncatalyzed cleavage of phosphodiester bonds in nucleic acids.[2] This family of enzymes includes DNA endonucleases,[12] 5′ to 3′ RNA exo and endonucleases,[13] 3′ to 5′ exonucleases,[14] restriction enzymes,[15] polymerases,[16] ribozymes,[17] topoisomerases,[18] sequence-specific recombinases,[19] Holliday junction (HJ) resolvases,[20] and ribonucleases,[21] to name but a few examples. Each enzyme is responsible for recognizing, positioning and stabilizing the substrate, activating the water nucleophile, and promoting leaving group departure.[2, 22] A large number of nucleases also relies on metals to enhance catalysis, with the role of the metal ions being a topic of debate.

The most accepted catalytic pathway for metal-dependent nucleases involves two-metal ions.[23] The first metal initiates the reaction by lowering the pK_a of the nucleophilic

water, while the second metal stabilizes the penta-coordinated intermediate and aids leaving group departure through direct metal ligation (Figure 2.1a). Support for the two-metal mediated mechanism has been obtained for several nucleases, including RNase II,[24] RNase III,[25] RNase D,[26] RNase E,[27] RNase H,[23] BamHI,[2] EcoRV,[28] UvrC,[29] MutL,[30] MutH,[31] Tn5 transposase,[32] T7 endonuclease I,[33] Rrp6p,[34] polymerase η,[35] and ribozymes (hammerhead-ribozyme,[36] group I intron,[37] and group II intron[38]).

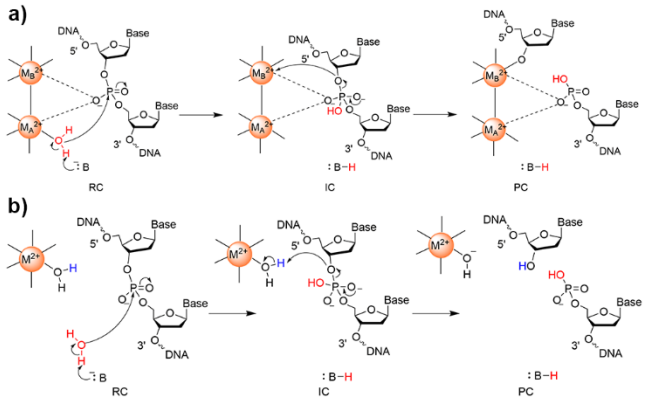

Figure 2.1. The general proposed mechanism for phosphodiester bond hydrolysis by a a) two-metal or b) one-metal dependent nuclease.

To complement literature on nucleases that involve multiple catalytic metals, there is growing evidence that many nucleases employ only a single metal ion for catalysis.[2, 22] To the best of our knowledge, the list of enzymes in this class currently includes human apurinic/apyrimidinic endonuclease (APE1),[39-40] Pvu II endonuclease,[41] *E. coli* restriction endonuclease I (EcoRI),[42] colicin E9,[43] colicin E7,[44] *Vibrio vulnificus* nuclease (Vvn),[45] *Serratia* nuclease,[46] phage T4 endonuclease VII (T4 Endo VII),[47] staphylococcal nuclease,[48] *Moraxella nonliquefaciens* restriction endonuclease II (MnlI),[49] HNH homing endonuclease (I-Hmul),[50] *Klebsiella pneumoniae* restriction endonuclease (Kpn1),[51]

Helicobacter pylori restriction endonuclease (HpyAV),[52] *Enterobacter hormaeche* restriction endonuclease (Ehol),[53] and intron-encoded endonuclease from *Physarum polycephalum* (I-*Ppo*I).[54-55] Experimental kinetics studies, mutational data, and crystal structures suggest that the single metal ion is bound to several water molecules and aspartate, glutamate and/or histidine residues to facilitate leaving group departure, while an active site basic residue initiates the reaction by activating the water nucleophile (Figure 2.1b). For example, an aspartate has been proposed to initiate the reaction facilitated by APE1 (D210, Figure 2.2a),[39-40] a histidine residue may play this role for I-HmuI (H75, Figure 2.2b) and I-*Ppo*I (H98, Figure 2.2c), and glutamate is the base in staphylococcal nuclease[48] (E43, Figure 2.2d). As for other nucleases, additional active site residues in each single metal-dependent enzyme have other important roles, ranging from reorganizing and positioning the substrate[39-40, 46, 55] to stabilizing the leaving group.[44-45, 54-55]

Many studies on a wide variety of enzymes have proven the usefulness of computational approaches to decipher the intricate details of the mechanism of action of enzymes and thereby complement experimental data (see for example [23, 36, 44, 56-61]). Although computational methodologies have provided mechanistic insights for nucleases that use two metal ions to facilitate phosphodiester bond cleavage,[23, 36, 57-59, 62] relatively few studies have considered single-metal dependent nucleases.[40, 44, 48] Quantum mechanical–molecular mechanics (QM/MM) approaches are particularly desirable as they allow for accurate reaction mechanism mapping in the context of the entire enzyme–nucleic acid system and permit direct comparisons to experimental data.[63] However, due to the computational expense of such methods, a more tractable approach is often to initially use density functional theory (DFT) coupled with cluster models in which a limited number of

atoms that represent the active site are extracted from X-ray crystallographic data and the periphery atoms are fixed to the experimental structure.[60, 64] Such cluster models have proven to provide valuable insight by permitting comparison of the structural and energetic features of multiple pathways, and thereby can direct subsequent large-scale QM/MM modeling.[56, 60] Unfortunately, however, it is currently unclear how to accurately design QM cluster models for single-metal dependent nucleases.

Due to the lack of computational studies on the single-metal mediated phosphodiester bond cleavage mechanism and previous successes describing metalloenzymes using cluster modelling approaches,[60-61, 65] the present work uses a range of cluster models coupled with DFT to clarify how to design accurate QM cluster models to study the chemistry facilitated by nucleases that invoke only one metal ion. APE1 is considered as a prototypical single-metal dependent nuclease due to the availability of an abundance of experimental kinetic,[39, 66] mutational[39, 67-69] and crystallographic data,[3, 39, 67-68] as well as previous QM/MM (ONIOM) work that has characterized the atomic level details of the preferred reaction pathway.[40] Furthermore, APE1 plays critical roles in biology, being best known for cleaving the phosphodiester backbone 5' with respect to abasic sites during DNA repair.[39, 67-68] APE1 is also involved in other important cellular processes including interstrand cross-link repair, nucleotide incision repair (NIR), transcription regulation, RNA metabolism, and gene expression,[70-73] while subtle mutations to APE1 have been linked to cancer,[74] ageing,[75] and cardiovascular[76] and neurological diseases.[77] This functional versatility has made APE1 an attractive therapeutic target for ocular diseases and various cancers, with an influx of recent research focusing on designing efficient inhibitors to complement those in clinical trials.[78-80]

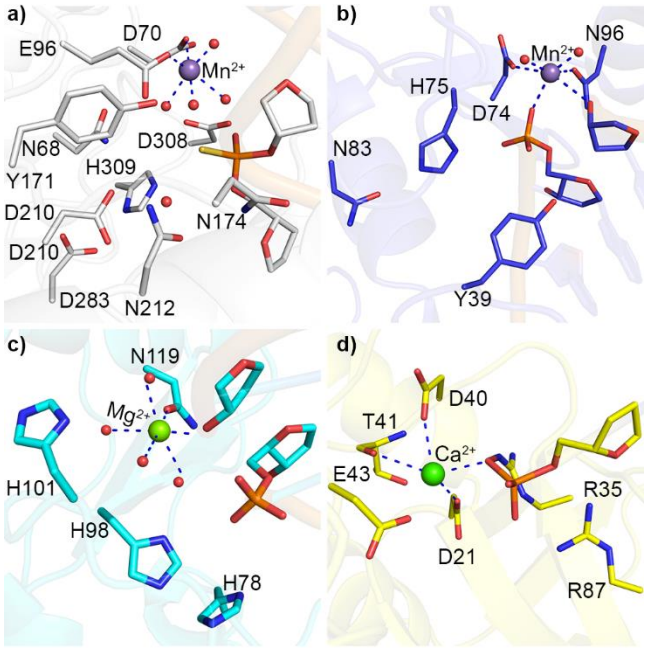

Figure 2.2. Examples of one-metal dependent nucleases that catalyze phosphodiester bond cleavage using different general bases, including a) D210 in APE1 (PDB ID: 5DG0), b) H75 in I-HmuI (PDB ID: 1UE3), c) H98 in I-*Ppo*I (PDB ID: 1A73), and d) E43 in staphyloccocal nuclease (PDB ID: 2SNS).

Starting from a high-resolution X-ray crystal structure of an APE1 reactant mimic (Figure 2.2a),[39] the present work considers the structural and energetic changes that occur during the phosphodiester bond cleavage reaction using a range of DFT functionals, basis sets, and the smallest model that encapsulates the role of the metal center and active site general base. Subsequently, the impact of systematically increasing the model size up to 185 atoms is investigated by including additional active site amino acids and altering the model truncation points. Finally, the effect of the surrounding enzymatic environment is scrutinized using a continuum solvation model. Through this systematic approach, our work provides the impact of DFT functional, cluster model size, and implicit solvation on the single-metal mediated phosphodiester bond cleavage facilitated by APE1 as a representative nuclease as the preferred mechanism has been previously established using

larger QM/MM models.[40] The conclusions drawn regarding how to accurately model the chemistry facilitated by a nuclease that relies on a single metal ion can be generalized to investigate other single-metal dependent enzymes (e.g., Pvu II endonuclease,[41] EcoRI,[42] colicin E9,[43] colicin E7,[44] Vvn,[45] *Serratia* nuclease,[46] T4 Endo VII,[47] staphylococcal nuclease,[48] MnlI,[49] I-Hmul,[50] Kpn1,[51] HpyAV,[52] Ehol,[53] and I-*Ppo*I[54-55]) and to direct future large-scale modeling to further our understanding of this complex chemistry.

2.2. Computational Details

Each cluster model was built from an X-ray crystal structure of a Mn(II)-containing APE1–DNA thiosubstrate reactive complex analogue (PDB ID: 5DG0; 1.8 Å resolution, Figure 2.2a).[39] Our smallest model (69 atoms, Figure 2.3a) includes only the essential features to understand the role of the single metal ion. Specifically, this model incorporates the abasic substrate, D210 as the base that activates the water nucleophile, and the Mg(II) ion with four coordinating waters, D70 and E96. Since no additional water molecules are present in the APE1 active site based on available crystal structures, no additional water molecules were added to any model.[39] In the substrate model, the abasic site is truncated at C3′ and the sugar from the 5′-adjacent nucleotide truncated at C4′, while the associated nucleobase is replaced with a hydrogen atom. Experimental mutational data supports the importance of D210 in APE1 function based on a 10,000-fold decrease in nuclease activity for the D210N mutation.[68] The impacts of D70 and E96 mutations are not as significant, with D70A or D70R resulting in a 26-fold loss of activity and E96Q leading to a 600-fold decrease in activity.[39] However, D70 and E96 are essential for maintaining the octahedral coordination of Mg(II) and therefore are included in all models. Expanded models were

then considered that incorporate active site residues that have been previously proposed to participate in the catalytic mechanism by stabilizing the transition state or aiding leaving group departure as well as residues that may be responsible for positioning key amino acids (namely H309, Y171, N212, N68, N174, D308, and/or D283).[39, 68-69, 81-82] The impact of including the 5′-nucleobase (cytosine) of the substrate was also explored. All amino acids were included in their native protonation states, except H309, which was considered to be cationic based on the experimentally observed pK_a[81] and NMR data.[83] Each amino acid was initially truncated at the β carbon or in other words the boundary was placed between Cβ–Cα bond, Cα was replaced with a hydrogen atom, and the location of Cβ was fixed to the crystallographic coordinates. Since QM cluster studies differ in model truncation points,[84-88] the effects of altering the truncation points to the a carbon were also considered or in other words the boundary was placed between the Cα–C(backbone) bond, the unsaturated carbon was replaced with a hydrogen atom, and the position of Cα was frozen. As a result, all residues in the largest cluster models were truncated at the α carbon except E96, which was consistently truncated at the β carbon. In all cluster models, truncation points were constrained to the respective crystal structure coordinates.

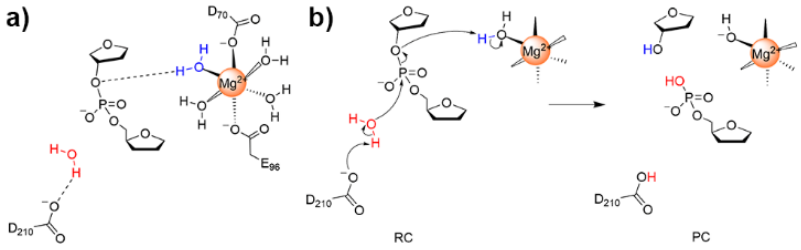

Figure 2.3. Schematic representation of the a) smallest model considered in the present work (69 atoms) and b) the associated concerted mechanism for the single-metal mediated phosphodiester bond hydrolysis.

As described in the Results and Discussion, a variety of density functional and basis set combinations were initially investigated for the smallest model. Based on comparisons of the various DFT methods, B3LYP-D3/6-31G(d,p) was chosen to optimize all remaining systems, with relative Gibbs energies that include thermal corrections (based on the harmonic approximation) determined using M06-2X/6-311+G(2df,2p) single-point calculations. The default convergence criteria were used for optimizations (i.e., max. force = 0.000450, RMS force = 0.00030, max. displacement = 0.001800, and RMS displacement = 0.001200). To investigate the impact of the surrounding environment, a polarized continuum model (IEF-PCM) was used for single-point energy calculations at the M06-2X/6-311+G(2df,2p) level of theory. A dielectric constant of ε=4 was used to represent the enzymatic environment for all continuum solvent calculations, which provides an accurate representation of the APE1 active site which is highly sequestered from water.[89] We note that our models as well as the previous QM/MM (ONIOM) model reference point do not account for the conformational dynamics of the active site, which can be important in enzymatic reactions.

All calculations were performed using Gaussian 09 (revision E.01).[90] Additional details of models and computational methodologies are provided in the Results and Discussion section.

2.3. Results and Discussion

2.3.1. Different Basis Sets and Density Functional Combinations Yield Similar Structural Features, But Large Variations in Reaction Energetics

Previous computational studies have examined nucleic acid phosphodiester bond hydrolysis using different DFT approaches,[91-92] including the widely used B3LYP[93-96] and M06-2X[40, 97-98] functionals. Although rigorous testing of the accuracy of DFT methods compared to CCSD(T)/CBS energetic data has been performed for the hydrolysis of dimethylphosphate in the absence of a metal,[99] it is not clear whether the same conclusions will hold for the reaction catalyzed by a single-metal dependent nuclease. Although the size of our model prohibits testing against gold standard techniques such as CCSD(T), our smallest model allows for comparison of the structural and energetic descriptions of the one-metal assisted hydrolysis characterized by a range of DFT functionals. Specifically, using APE1 as a representative example (Figure 2.3a), our model includes the (truncated) substrate, D210 to activate the water nucleophile, and Mg(II) coordinated to four water molecules, D70 and E96, with all amino acids truncated at the β carbon. This results in a model of 69 atoms with an overall charge of –2. The gas-phase incision mechanism characterized using this model involves a single, concerted step, with D210 abstracting a proton from the nucleophilic water to initiate the reaction at the same time as the nucleophile approaches the phosphate group and a proton is transferred to the leaving group from a metal-activated water to cleave the phosphodiester bond (Figure 2.3b).

Prior to considering a variety of DFT functionals, the impact of the basis set size was investigated using the M06-2X functional based on its ability to accurately describe kinetics involving main-group elements and non-covalent interactions,[100] as well as enzymatic reactions.[101-102] As the Pople basis set used for the geometry optimizations is systematically increased from 6-31G(d) to 6-311+G(2d,2p), the calculated Gibbs energy of activation varies from 91.4 to 142.7 kJ/mol, a variation of 51.3 kJ/mol ($\Delta^{\ddagger}G_{OPT}$, Table 2.1). Specifically, the addition of polarization functions on hydrogen atoms to 6-31G(d)

increases the reaction barrier by 15.7 kJ/mol, while subsequent inclusion of heavy atom diffuse functions or expansion to the triple-zeta variant leads to a negligible energy impact (0.6 – 3.2 kJ/mol). The inclusion of heavy atom diffuse functions to 6-311G(d,p) increases the activation energy more significantly (7.0 kJ/mol). More importantly, further expansion of the basis set to 6-311+G(2d,2p), 6-311+G(2df,p) or 6-311+G(2df,2p) results in much larger and more consistent barriers of approximately 140 kJ/mol.

Table 2.1. M06-2X gas-phase Gibbs activation energy (kJ/mol) for the phosphodiester bond cleavage reaction calculated using a variety of basis sets.[a]

basis set	no. of basis functions	$\Delta^{\ddagger}G_{OPT}$[b]	$\Delta^{\ddagger}G_{SP}$[c]
6-31G(d)	588	91.4	119.1
6-31G(d,p)	693	107.1	140.1
6-31+G(d,p)	829	107.7	123.8
6-311G(d,p)	838	110.3	124.0
6-311+G(d,p)	974	117.3	122.6
6-311+G(2d,2p)	1249	142.7	145.1
6-311+G(2df,p)	1382	140.4	141.1
6-311+G(2df,2p)	1657	140.4	140.4
average		119.7	132.0
stdev.		19.2	10.5

[a]The smallest 69 atom model with a charge of –2 was used (Figure 2.3a). [b]Calculated using M06-2X geometries and energies obtained with the indicated basis set. [c]Calculated using M06-2X/6-311+G(2df,2p) single-point calculations on M06-2X geometries obtained with the indicated basis set.

Careful analysis of key catalytic distances obtained with each basis set suggests minimal geometric differences (Figures A.1– A.2, Tables A.1– A.3, Appendix A), with the largest deviation across basis sets occurring in the distance between O3′ of the leaving group and the proton from the metal-activated water in the reactant complex (1.804 – 2.181 Å, r5, Table A.1, Appendix A). Indeed, when M06-2X/6-311+G(2df,2p) single-point calculations are performed on each optimized structure, the range in the computed Gibbs

energy of activation decreases to 26.0 kJ/mol ($\Delta^{\ddagger}G_{SP}$, Table 2.1). Most importantly, when 6-311+G(2df,2p) single-point calculations are performed on geometries obtained with 6-31G(d,p), the activation barrier is within 0.3 kJ/mol of that obtained using the geometry optimized with the largest basis set considered in the present work (6-311+G(2df,2p)). Additionally, there is less than an 8% deviation in the 6-31G(d,p) and 6-311+G(2df,2p) optimized structures (Tables A.4–A.6, Appendix A). Thus, the 6-31G(d,p) basis set will be used for all subsequent geometry optimizations, while the reaction energetics will be evaluated using 6-311+G(2df,2p) single-point calculations.

The gas-phase concerted mechanism for the phosphodiester bond cleavage facilitated by a single metal was subsequently considered using 18 DFT functionals (Table 2.2 and Figure 2.4). The average Gibbs activation energy is 148.0±17.6 kJ/mol across all functionals. M06-HF, PBEPBE and BP86 result in the smallest barrier (113.5 – 120.3 kJ/mol), while the largest barrier is predicted with M05 (178.1 kJ/mol). Removing these functionals from the comparison results in an average barrier of 152.4±10.0 kJ/mol, which highlights the similar description of the reaction by the remaining methods. Indeed, the structural changes in the stationary points along the reaction pathway across these functionals are minimal (Tables A.7–A.9, Appendix A), suggesting that the differences in the energy barriers arise from the energetic rather than the structural description provided by the heterogeneously designed DFT functionals. Indeed, there is a small variation in M06-2X/6-311+G(2df,2p) energies obtained with structures optimized using the different functionals (Table 2.2 and Figure 2.4, green), while single-point calculations with each functional on a consistent M06-2X/6-31G(d,p) geometry leads to highly varied energetics (Table 2.2 and Figure 2.4, orange). The varied reaction energetics obtained for APE1 are in

line with other DFT studies and arise due to the unique design of each functional, with each having different strengths and weaknesses.[99-100]

Table 2.2. Gas-phase Gibbs activation energies (kJ/mol) for the phosphodiester bond cleavage reaction calculated using a variety of DFT functional combinations.[a]

functional X	X//X[b]	M06-2X//X[c]	X//M06-2X[d]
M06-HF	113.5	139.7	120.5
PBEPBE	119.3	127.5	132.6
BP86	120.3	126.7	133.0
PBEPBE-D3	136.5	143.1	135.9
M06-2X	140.1	140.1	140.1
B3LYP-10%HF	140.5	127.8	147.8
M06-2X-D3	141.6	139.9	141.6
M06-L	146.4	140.0	152.1
MN15	151.0	147.7	135.5
M11	152.4	142.9	151.7
MPWB1K	152.9	145.8	151.6
B3LYP	155.9	127.3	160.2
M06	160.9	144.0	161.7
M11-L	161.4	155.7	152.7
MN12-SX	163.4	149.5	157.5
B3LYP-D3	164.4	141.6	163.5
MN12-L	165.8	160.4	159.1
M05	178.1	150.6	172.3
average	148.0	141.7	148.3
stdev.	17.6	9.7	13.4

[a]All single-point calculations were performed with 6-311+G(2df,2p) and optimizations were performed with 6-31G(d,p) using the smallest 69 atom model with a charge of –2 (Figure 2.3a). [b]Single-point calculations and optimizations were performed with the indicated functional X. [c]M06-2X single-point calculations were performed on geometries obtained with the indicated functional X. [d]Single-point calculations were performed with the indicated functional X on geometries obtained with M06-2X.

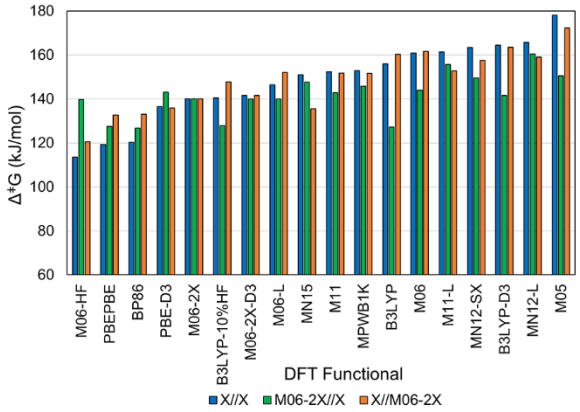

Figure 2.4. The gas-phase Gibbs activation energies for the rate-determining step ($\Delta^{\ddagger}G$) calculated using various DFT functionals (X) for optimizations and/or single-point calculations and the smallest model considered in the present work (see Figure 2.3a). Single-point calculations and optimizations were performed with the indicated functional X (blue), M06-2X single-point calculations were performed on geometries obtained with the indicated functional X (green), and single-point calculations were performed with the indicated functional X on M06-2X geometries (orange).

Overall, the functionals investigated in this section were specifically chosen based on literature supporting their use for phosphate hydrolysis, barrier heights in general, enzymatic reactions, and/or group II metal chemistry.[91-92, 99] Our data suggests that many heterogeneously designed functionals provide a similar description of the structures of key stationary points along the single-metal mediated phosphodiester bond cleavage reaction. Furthermore, although the predicted barrier can vary significantly with functional, several functionals yield similar reaction energetics. Importantly, our data supports the use of standard functionals (M06-2X and B3LYP-D3) for these reactions, which have been widely used in the literature to study other enzymatic mechanisms.[40, 97-98] Regardless, the model used so far is minimal and larger cluster models must be considered to determine the impact of additional amino acids on the reaction mechanism and associated barriers.

2.3.2. Large Fluctuations in Both Structure and Reaction Energetics Occur Upon Model Expansion, Emphasizing That Caution is Required When Designing Cluster Models

Although the minimal model considered in the previous section highlights the importance of both an active site general base (modeled as truncated aspartate) and the metal activated water in the overall nucleic acid phosphodiester bond cleavage reaction, greater information can be obtained about the enzymatic mechanism by systematically expanding the computational model. Due to the size of the models to be considered and the similar performance of most methods on the minimal model, we will evaluate the energies using M06-2X single-point calculations on B3LYP-D3 geometries. We note that comparing the predicted barriers directly to experimental values is not recommended since the full enzyme is not present. Instead, the goal of this work is to determine the minimum number of atoms required for the accurate structural description of the one-metal mediated phosphodiester bond hydrolysis.

2.3.2.1. *Expansion of the Substrate Preserves the Concerted Phosphodiester Bond Hydrolysis Mechanism:* The first place to enhance our computational model is the substrate. Specifically, the substrate model was expanded by adding the nucleobase (cytosine) on the 5′-nucleoside with respect to the abasic site, a hydroxymethyl group to the 5′-terminus, and a hydroxy group to C3′ of the abasic site (Figure 2.5). Cytosine was chosen as the nucleobase to include in the model due to its presence in the crystal structure from which the models were built. As per the smallest model, all amino acids were truncated at the α carbon, which results in a model containing 85 atoms and a charge of –2. The concerted

mechanism previously predicted for the smallest model (Figure 2.3b) is preserved for this expanded model and the barrier increases by 11.8 kJ/mol to a value of 176.3 kJ/mol. This higher barrier arises due to a later transition state (i.e., r4 (P–O3′) is 0.4 Å longer in when the substrate is expanded (Figure A.3, Appendix A)) compared to the smallest model (Figure A.2, blue, Appendix A). The predicted barrier increase is in line with experimental evidence that replacement of the 5′-cystosine with thymine reduces APE1 activity by 5-fold,[39] which emphasizes that the 5′-nucleobase impacts the overall reaction. Therefore, all the subsequent models include this expanded substrate.

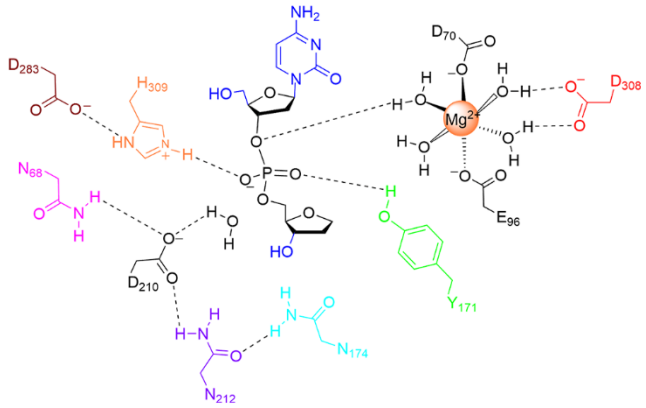

Figure 2.5. Schematic of the expanded computational models considered in the present work: smallest model (black), expanded substrate model (black + dark blue, expanded substrate), model 1 (model 0 + orange, H309), model 2 (model 1 + green, Y171), model 3 (model 1 + purple, N212), model4 (model 1 + pink, N68), model 5 (model 1 + red, D308), model 6 (model 1 + brown, D283), model 7 (model 2 +purple, N212), model 8 (model 7 + cyan, N174), model 9 (model 8 + pink, N68), and model 10 (all residues).

2.3.2.2. *Inclusion of the General Acid (Model 1) Results in a Stepwise Pathway and Significantly Reduces Reaction Barriers:* Following expansion of the substrate, we turn our attention to the inclusion of additional active site residues. Although model 0 contains the

general base to activate the water nucleophile and the metal-activated water to facilitate leaving group departure, substrate charge stabilization by amino acids could greatly impact the reaction pathway. In the case of APE1, this role may be played by positively charged H309, which has been proposed to be important in substrate binding and stabilization of the penta-coordinated transition state/intermediate.[39, 68, 81] When H309 is positioned relative to the substrate as per the APE1 crystal structure (model 1, Figure 2.5), the model charge changes to –1. With model 1, a two-step hydrolysis pathway is characterized (Figure 2.1b). In the first reaction step, D210 maintains the role of the general base that deprotonates the water nucleophile and promotes nucleophilic attack. This yields a phosphorane intermediate that has similar structural features to the corresponding intermediate characterized for the reaction facilitated by two-metal dependent nucleases.[2, 23, 25] In the second reaction step, a metal activated water stabilizes the leaving group by transferring a proton to O3′ of the 5′-nucleotide in the model substrate to cleave the phosphodiester bond. The change from a one-step to a two-step reaction upon inclusion of H309 emphasizes the important role of this residue in substrate stabilization. Indeed, H309 is protonated at the beginning of the reaction and hydrogen bonds with a non-bridging oxygen of the substrate phosphate moiety (Figure A.4, Appendix A). From TS1 to IC, H309 transfers a proton to a non-bridging oxygen of the phosphate group, which in turn stabilizes the IC and PC (Figure A.4, Appendix A). The second step is rate limiting, with a corresponding Gibbs activation energy of 112.8 kJ/mol (Table 2.3), which is 63.5 kJ/mol lower than the barrier in the absence of H309 in the expanded substrate model. This correlates with experimental data indicating a H309N mutation decreases the APE1 catalytic rate by 2,500-fold.[68] This also underscores the critical role of an acid in substrate stabilization for the phosphodiester bond cleavage catalyzed by single-metal dependent nucleases. Therefore, all further models

include H309. We note that a concerted pathway could not be characterized upon inclusion of H309, while the stepwise mechanism could not be isolated in the absence of this important residue.

Table 2.3. Relative gas-phase Gibbs energies (ΔG, kJ/mol) for the phosphodiester bond cleavage reaction calculated using model 1 with the truncation point for select amino acid(s) extended to the α carbon.[a,b]

Residue Expanded to C_α							
D210	H309	D70/E96	RC	TS1	IC	TS2	PC
			0.0	56.7	58.6	112.8	87.0
X			0.0	31.9	30.9	92.2	–69.7
	X		0.0	37.8	37.7	94.4	–33.9
		X	0.0	56.5	57.9	104.7	–78.1
X	X		0.0	35.0	37.5	90.6	–39.2
X		X	0.0	41.4	42.4	89.8	–16.4
	X	X	0.0	47.1	49.1	101.9	50.1
X	X	X	0.0	45.0	36.1	85.7	–26.9

[a]E96 was truncated to the b carbon. [b]Relative energies were obtained using M06-2X/6-311+G(2df,2p) single-point calculations on B3LYP-D3/6-31G(d,p) geometries.

2.3.2.3. *Expansion of Amino Acid Models to Include the α Carbon Enhances Model Flexibility and Results in Significant Structural and Energetic Changes Along the Phosphodiester Bond Hydrolysis Pathway:* It is important to recognize that the truncation points in the above models were fixed to the X-ray crystallographic positions for each amino acid residue (i.e., the α carbon). It is possible that inclusion of a greater portion of the side chain will increase the flexibility of the model and affect the structural description of the reaction along with energy barriers. Furthermore, QM cluster studies in literature differ in the choice of model truncation points.[84-88] Therefore, prior to adding additional amino acids, different combinations of extending the truncation point to the α-carbon for D210, H309, and D70 are examined (Table 2.3), with the length of E96 kept consistent with

D70 in each model since both residues are strongly coordinated to the metal center and likely play similar roles.

Although expansion of key amino acid side chain(s) to C_α does not change the two-step mechanism predicted with model 1 (Table 2.3), large structural and energetic differences occur along the reaction pathway regardless of whether one or multiple amino acids are expanded. Indeed, the barrier associated with the first step can change by up to 24.8 kJ/mol (or 44%) relative to model 1 and that for the second (rate-determining) step can change by up to 27.1 kJ/mol (or 24%) upon expansion of the amino acid residue. Model truncation has even larger effects on the relative energy of the intermediate (27.7 kJ/mol or 47%) and product (165.1 kJ/mol or 190%) complexes. These energetic changes can be understood based on structural adjustments that occur upon amino acid expansion as discussed below.

Structural comparisons of different truncated models reveal that model 1 is heavily constrained (Table A.10, Appendix A), which results in a smaller average distance between the water nucleophile and the phosphate backbone in the reactant complex compared to the other models (i.e., r3 decreases by up to 0.657 Å, Table A.10, Appendix A). Furthermore, the leaving group is unable to move away from the phosphate atom in the product complex for model 1 (i.e., r4 deviates by 0.766 Å compared to other models, Table A.10, Appendix A). When D210 is expanded to C_α (Table 2.3), the first barrier decreases by 24.8 kJ/mol, the second barrier decreases by 20.6 kJ/mol, and the product is highly stabilized. The first reaction barrier is reduced due to a more reactive RC in which the nucleophilic water is better positioned (i.e., r3 is smaller compared to model 1 by 1.8 Å, Figure A.5 and Table A.10, Appendix A). The lower second barrier arises since expanded D210 promotes

enhanced alignment of the substrate with respect to the Mg(II)-activated water (Figure A.5, Appendix A) and leads to a PC that is further along the reaction pathway compared to model 1 (i.e., r4 is 0.4 Å larger and r8 is 0.4 Å smaller, Table A.10, Appendix A). Additionally, the proton is transferred from the phosphate to H309 in the PC upon D210 expansion, leading to an exergonic reaction and regenerating the enzyme for another reaction (Tables A.3 and A.10, Appendix A).

Expansion of H309 to the C_α truncation point also lowers the barriers and stabilizes the product (Table 2.3). First, the enhanced flexibility of the model permits H309 to form a weak hydrogen bond between a C_α–H and the water nucleophile in the RC (Figure A.6a, Appendix A), which situates the nucleophilic water closer to the phosphate moiety (i.e., r3 is smaller by 1.9 Å, Figure A.1, Table A.10, Appendix A) and thereby decreases the barrier for the first reaction step by 18.9 kJ/mol compared to model 1. Second, increased model flexibility allows H309 to reposition with respect to the substrate for the second reaction step (Figure A.6c-e, Appendix A), which lowers the rate-determining barrier by 18.4 kJ/mol (Table 2.3). Despite the energetic benefits of the individual expansion of D210 and H309, the reaction energetics are highly similar when either H309 or both residues are expanded to C_α (Table 2.3), suggesting these effects are not additive.

Finally, since the Mg(II)-coordinating residues (D70 and E96) are far removed from the reaction center, expansion of D70 to C_α and E96 to C_β alone does not significantly change the reaction parameters (Table A.10, Appendix A) or barriers (Table 2.3) compared to model 1, although the final product is more stable. Therefore, these residues likely have structural roles, including maintaining Mg(II) octahedral coordination. When D70/E96 expansion is coupled with larger D210 and/or H309 models, the barriers for the reaction

(Table 2.3) are similar to those for the corresponding model with smaller D70/E96 models. However, the product complex is earlier due to greater motion of the metal (r4, Figure A.8 and Table A.10, Appendix A) and significantly destabilized compared to models with only D70/E96 expanded (by up to 128.2 kJ/mol, Table 2.3). When all residues are expanded to C_α (Table 2.3), the lowest rate-determining barrier is obtained, at least in part because D210 and the nucleophilic water are better aligned with respect to substrate in the RC (i.e., r3 is 1.7 Å smaller, Figure A.8a and Table A.10, Appendix A). Additionally, the product is exergonic (by 26.9 kJ/mol, Table 2.3) due to the later PC (r4 increases by 1.8 Å), H309 regaining a proton, and the reorganization of active site (Figure A.8 and Table A.10, Appendix A).

Overall, the lowest barrier (85.7 kJ/mol, Table 2.3) for the rate-determining step obtained from the expansion of D210, H309, and D70 to C_α falls within the range of the barriers typically observed for metallonucleases facilitated phosphodiester bond hydrolysis reactions (58 – 96 kJ/mol),[47, 103-104] suggesting there is an advantage of model expansion to include C_α of key amino acids. This result underscores the importance of ensuring that residues have the flexibility necessary to capture the chemistry occurring within the enzymatic environment. Indeed, truncation at C_α restricts the computational model and thereby affects the structures and energetics of stationary points along the phosphodiester bond cleavage reaction. Thus, each amino acid residue in all models considered in the remainder of this study is truncated at C_α except for E96, which is truncated at C_β to be consistent in size with D70.

2.3.2.4. *Inclusion of Additional Amino Acid Residues Leads to Significant Active Site Rearrangement, Which Impacts the Reaction Barrier (Models 2 – 9):* Despite model 1 (truncated at C_α) yielding a barrier typical for the enzymatic phosphodiester bond cleavage reactions (58 – 96 kJ/mol),[47, 103-104] the nucleophilic water is hydrogen bound to H309 in the reactant complex (Figure A.8a, Appendix A) and therefore not optimally aligned for the reaction. Indeed, there is no evidence in the X-ray crystal structure of the APE1 reactant complex (PDB ID: 5DG0) of a hydrogen bond between H309 and the nucleophile, with H309 being 4.6 Å from an active site water. Additionally, a hydrogen bond exists between 5′-OH of cytidine and a non-bridging oxygen of the phosphate group throughout the reaction (Figure A.8, Appendix A), which is an artifact of the model lacking a continuing DNA chain. Therefore, there is room for improvement in the cluster model. Indeed, experimental kinetic data reveals that D210 and H309 mutations significantly impact the catalytic power of APE1,[68] and the Y171F mutation leads to a 1,200-fold decrease in the catalytic rate,[39, 68] while the N212A mutation leads to a 70,000-fold rate decrease.[69] The proposed roles of Y171 include positioning of the nucleophile and/or substrate charge stabilization,[39-40] while N212 has been proposed to orient the D210 general base for the reaction.[39, 69, 82] Therefore, the impacts of Y171 and N212 are initially considered by building upon model 1.

Upon addition of Y171 to the cluster model (model 2), the substrate rearranges such that Y171 simultaneously hydrogen bonds to a non-bridging oxygen of the substrate phosphate in the first step and 5′-OH of cytidine throughout the reaction (Figure A.9, Appendix A). This stabilizes the RC compared to model 1, which increases the first reaction barrier by 16.7 kJ/mol (Table 2.4). Even though the second step has a later TS (i.e., r5 is

0.5 Å longer, Table A.11, Appendix A), the second barrier increases by only 5.5 kJ/mol since Y171 rotates to stabilize the leaving group through hydrogen bonding (Figure A.9d, Appendix A). Nevertheless, this active site arrangement might not be possible in the enzymatic environment, suggesting that further model expansion is required.

Adding N212 to the cluster model 1 (model 3) reduces the barrier to 66.3 kJ/mol (Table 2.4). This may arise because of additional hydrogen bonding in the active site, such as interactions between N212 and D210 or the substrate, and a change in metal coordination from octa- to penta-coordinated in the second reaction step (Figure A.10, Appendix A). Similar active site reorganization exists when both Y171 and N212 are present (model 7, Figure A.11, Appendix A), emphasizing the need for more balanced models than models 3 and 7.

Other residues in the active site that might play important roles in the reaction include N68, D308 and D283. Indeed, the N68A mutation reduces APE1 catalytic activity by 600-fold,[39] while D308A and D283A reduce the activity by 5-25 fold[66] and 10-fold,[66, 82] respectively. However, when any one of these residues is added to model 1, significant structural changes occur along the reaction path, which result in high barriers and/or unstable PC (Table 2.4 and A.11, Appendix A). Specifically, when N68 is added to the cluster model (model 4), H309 rotates away from the substrate, which then interacts with E96 (Figure A.12, Appendix A). Upon D308 addition (model 5) to model 1, the net charge of the model changes to –2, and the energy along the step-wise pathway continuously increases from the RC to the PC due to reorganization of the active site (Table 2.4 and Figure A.13, Appendix A). When D283 is included in model 1 (model 6, charge changes to

–2), D210 interacts with the abasic site and is unable to sufficiently activate the nucleophile (Figure A.14, Appendix A).

Due to the recognized importance of Y171 and N212 in the reaction pathway, larger models that include both residues along with additional amino acids were considered. In model 8, N174 was included due to its proposed role in substrate alignment,[39-40, 72] with the N174Q mutation reducing APE1 activity by 9-fold.[105] Addition of N174 to cluster model 7 leads to two additional hydrogen bonds between the substrate and the Mg-ligated water or Y171 that are not possible in the enzymatic system when the substrate is part of a continuous DNA chain (Figure A.15, Appendix A). Similarly, subsequent addition of N68 (model 9) results in a hydrogen-bonding network between N68 and D210 as well as E96, which increases the barrier (Table 2.4), and the system maintains unrealistic artifacts from the previous model (Figure A.16, Appendix A).

Table 2.4. Relative gas-phase Gibbs energies (kJ/mol) for the phosphodiester bond cleavage reaction calculated using models of varying size.[a,b]

model (no. atoms; charge)	Y171	N212	N68	N17	D308	D283	RC	TS1	IC	TS2	PC
model 1 (110; –1)							0.0	45.0	36.1	85.7	–26.9
model 2 (129; –1)	X						0.0	61.7	65.4	91.2	–85.8
model 3 (122; –1)		X					0.0	40.0	38.8	19.4	–68.0
model 4 (122; –1)			X				0.0	143.6	146.8	136.5	35.4
model 5 (120; –2)					X		0.0	53.7	76.1	95.6	113.0
model 6 (120; –2)						X	0.0	195.4	153.8	155.5	34.7
model 7 (141; –1)	X	X					0.0	61.0	62.5	108.5	–56.5
model 8 (153; –1)	X	X		X			0.0	51.5	43.1	70.0	–49.8
model 9 (165; –1)	X	X	X	X			0.0	84.5	88.1	145.2	108.9
model 10 (185; –3)	X	X	X	X	X	X	0.0	79.3	59.7	79.1	–21.7

[a]Relative energies were obtained using M06-2X/6-311+G(2df,2p) single-point calculations on B3LYP-D3/6-31G(d,p) geometries. [b]All models contain D210 and H309. See Figure 2.5 for the schematic of the models. All amino acids were truncated at the α carbon, except for E96 (β carbon truncation).

Extending the cluster model further to include both D308 and D283 (model 10) increases the net charge to –3. This model yields an improved description of the active site chemistry (Figure A.17, Appendix A). In fact, the overall structure, including the hydrogen-bonding network between the amino acids and substrate, and metal coordination, is highly similar to that characterized for the full enzyme–DNA complex with the ONIOM methodology.[40] In agreement with NMR and crystallographic data, H309 remains cationic in the reactant complex (Table A.11 and Figure A.17, Appendix A)[39, 68, 81] and a thermodynamically stable PC is obtained. The excellent agreement between the predicted structural data and previous ONIOM geometries[40] suggests structural convergence occurs at a cluster model with 185 atoms and corroborates previous conclusions that large cluster models containing secondary shell amino acids can yield geometries consistent with QM/MM data.[106-107] This finding is also in line with proposals that the correct representation of various enzymatic reactions occurs with model sizes of 150 – 200 atoms.[56, 60] Indeed, recent QM cluster studies have used 150 – 220 atoms to successfully map the reaction pathways for a range of other enzymes.[61, 108]

Overall, the cluster models considered in the present work highlight the importance of key APE1 active site amino acids. Specifically, D210 acts as a general base that initiates the reaction and H309 is the general acid that stabilizes the charge on the substrate, while second shell amino acids (N212, N68, and D283) help ensure the correct orientation of these critical residues. Equally important is the correct alignment of the substrate for the reaction with respect to these residues, which is facilitated by N174. While D70 and E96 are clearly necessary for maintaining Mg(II) octahedral coordination, D308 is also important for preventing inaccurate penta-coordination seen in smaller models. These

primary amino acids that play the roles of the general base and acid, stabilize the substrate, and maintain the metal octahedral coordination, as well as secondary residues that facilitate these amino acids must be included in the model. This leads to a 185-atom limit for a QM-cluster model of APE1. In addition to these key roles for active site amino acids, the large fluctuations in the predicted barriers emphasize that care must be taken when designing cluster models for phosphodiester bond cleavage, with a minimum model of 185 atoms required to accurately describe the reaction facilitated by APE1. This work corroborates previous literature highlighting the need for cautious model building to achieve accurate representation of enzymatic reactions.[60] Specifically, although smaller models can be a good starting point for early investigations of multiple mechanisms, they can lead to artificial strain and yield incorrect energy profiles.

2.3.2.5. *Accounting for the Broader Enzymatic Environment Does Not Significantly Change the Barrier for the Rate-Determining Step:* Although the previous calculations were performed in the gas phase, literature on other enzymes has illustrated that including the effects of the surrounding environment of the residues included in a cluster model through single-point calculations with continuum solvation models provides a good compromise between accuracy and computational cost.[109-110] Therefore, the impact of solvent-phase single-point calculations on the gas-phase geometries was considered in the present work using the polarizable continuum model and a dielectric constant of 4,[89] which can provide a good representation of active sites that are at least partially sequestered from water and has been successfully used to map other enzymatic reactions using cluster models.[87, 108, 111] Accounting for the broader environment does not significantly change the barrier for the

rate-determining step for most models (less than ~10%, Figure 2.6 and Table A.12, Appendix A). The largest effect occurs for model 5, where the barrier decreases by 29.1 kJ/mol upon inclusion of implicit solvation (~30% change). Most importantly, all previous conclusions regarding the impact of model size based on the trends in the rate-limiting barrier identified in the gas phase hold true in an enzymatic environment (Figure 2.6 and Table A.12, Appendix A), suggesting that models 1 – 10 are large enough to capture the impact of the broader enzymatic environment on catalysis. Indeed, previous DFT studies on other enzymes have emphasized that the magnitude of additional environment effects dissipates for models larger than 100 atoms.[109]

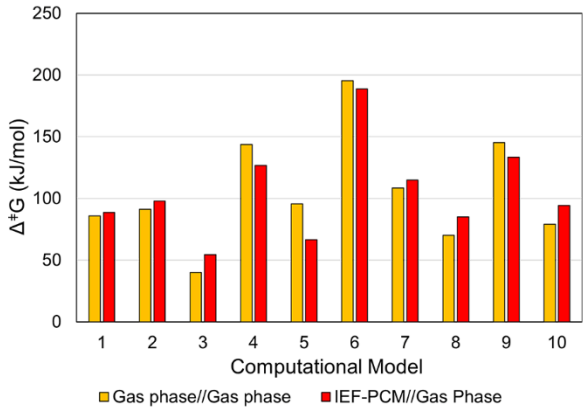

Figure 2.6. Comparison of the M06-2X Gibbs activation energies ($\Delta^{\ddagger}G$) for the rate-determining step obtained from gas-phase (orange) or IEF-PCM (ε=4, red) single-point calculations on gas-phase geometries for models of varying size. Models are defined in Figure 2.5.

2.4. Conclusion

The present study considers the impact of the DFT functional and basis set, cluster model size, and implicit solvation on the single-metal mediated phosphodiester bond cleavage

facilitated by APE1 to unveil how to accurately model such reactions in an efficient manner. APE1 is used as a prototypical example due to its known functional versatility, including important roles in DNA repair and as a therapeutic target to combat several human diseases. As the simplest cluster model illustrates that several functionals and basis sets provide similar descriptions of the phosphodiester bond cleavage reaction, our work highlights that common functionals (e.g., M06-2X and B3LYP-D3) can be combined with 6-31G(d,p) to reliably describe the structures of important stationary points, with the larger 6-311+G(2df,2p) basis set used to obtain more accurate reaction energetics. However, there are clear limitations in smaller cluster models (< 122 atoms), which induce artificial strain and can lead to incorrect metal coordination geometries. Nevertheless, the impact of accounting for the broader enzymatic environment using continuum methods is negligible for larger models. Overall, our work suggests that the general acid and base, residues that maintain the metal octahedral coordination, and amino acids that stabilize the substrate, as well as residues that facilitate these primary amino acids, must be included in the QM cluster model to achieve an accurate structural description of the phosphodiester bond hydrolysis facilitated by APE1, which results in a minimum of 185 atoms. Indeed, although smaller models that contain only the primary amino acids can characterize a two-step (rather than concerted) reaction, larger models that include secondary amino acids are required to maintain the active site architecture and reproduce experimental reaction energetics. This work thereby underscores the importance of correctly choosing the cluster model size for the reliable prediction of possible reaction pathways associated with other single-metal mediated nucleases. Similarly designed cluster models can be used to efficiently characterize different mechanistic possibilities for a range of single-metal dependent nucleases and thereby direct future QM/MM modelling to further explore

whether a single metal is enough for catalytic cleavage of the stable phosphodiester bond in nucleic acids.

2.5. References

(1) Duarte, F.; Åqvist, J.; Williams, N. H.; Kamerlin, S. C., Resolving apparent conflicts between theoretical and experimental models of phosphate monoester hydrolysis. *J. Am. Chem. Soc.* **2015,** *137* (3), 1081–1093.

(2) Yang, W., Nucleases: diversity of structure, function and mechanism. *Q. Rev. Biophys.* **2011,** *44* (1), 1–93.

(3) Whitaker, A. M.; Flynn, T. S.; Freudenthal, B. D., Molecular snapshots of APE1 proofreading mismatches and removing DNA damage. *Nat. Commun.* **2018,** *9* (1), 399.

(4) Hecht, S. S., Tobacco carcinogens, their biomarkers and tobacco-induced cancer. *Nat. Rev. Cancer* **2003,** *3* (10), 733–744.

(5) Helleday, T.; Eshtad, S.; Nik-Zainal, S., Mechanisms underlying mutational signatures in human cancers. *Nat. Rev. Genet.* **2014,** *15* (9), 585–598.

(6) Herrmann, S. S.; Duedahl-Olesen, L.; Christensen, T.; Olesen, P. T.; Granby, K., Dietary exposure to volatile and non-volatile N-nitrosamines from processed meat products in Denmark. *Food Chem. Toxicol.* **2015,** *80*, 137–143.

(7) De Bont, R.; van Larebeke, N., Endogenous DNA damage in humans: A review of quantitative data. *Mutagenesis* **2004,** *19* (3), 169–185.

(8) Alexandrov, L. B.; Nik-Zainal, S.; Wedge, D. C.; Aparicio, S. A.; Behjati, S.; Biankin, A. V.; Bignell, G. R.; Bolli, N.; Borg, A.; Børresen-Dale, A.-L., Signatures of mutational processes in human cancer. *Nature* **2013,** *500* (7463), 415–421.

(9) Bartoli-Leonard, F.; Wilkinson, F. L.; Schiro, A.; Inglott, F. S.; Alexander, M. Y.; Weston, R., Loss of SIRT1 in diabetes accelerates DNA damage-induced vascular calcification. *Cardiovasc. Res.* **2021,** *117* (3), 836–849.

(10) Ainslie, A.; Huiting, W.; Barazzuol, L.; Bergink, S., Genome instability and loss of protein homeostasis: converging paths to neurodegeneration? *Open Biol.* **2021,** *11* (4), 200296.

(11) Williams, N. H.; Takasaki, B.; Wall, M.; Chin, J., Structure and nuclease activity of simple dinuclear metal complexes: quantitative dissection of the role of metal ions. *Acc. Chem. Res.* **1999,** *32* (6), 485–493.

(12) Loenen, W. A.; Dryden, D. T.; Raleigh, E. A.; Wilson, G. G.; Murray, N. E., Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Res.* **2014,** *42* (1), 3-19.

(13) Maraia, R. J.; Lamichhane, T. N., 3′ processing of eukaryotic precursor tRNAs. *Wiley Interdiscip. Rev. RNA* **2011,** *2* (3), 362-375.

(14) Henninger, E. E.; Pursell, Z. F., DNA polymerase ε and its roles in genome stability. *IUBMB life* **2014,** *66* (5), 339–351.

(15) Di Felice, F.; Micheli, G.; Camilloni, G., Restriction enzymes and their use in molecular biology: An overview. *J. Biosci.* **2019,** *44* (2), 1–8.

(16) Jain, R.; Aggarwal, A. K.; Rechkoblit, O., Eukaryotic DNA polymerases. *Curr. Opin. Struct. Biol.* **2018,** *53*, 77-87.

(17) 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.

(18) Bizard, A. H.; Hickson, I. D., The many lives of type IA topoisomerases. *J. Biol. Chem.* **2020,** *295* (20), 7138-7153.

(19) Olorunniji, F. J.; Rosser, S. J.; Stark, W. M., Site-specific recombinases: molecular machines for the Genetic Revolution. *Biochem. J.* **2016,** *473* (6), 673–684.

(20) Lilley, D. M., Holliday junction-resolving enzymes—structures and mechanisms. *FEBS Lett.* **2017,** *591* (8), 1073–1082.

(21) Bechhofer, D. H.; Deutscher, M. P., Bacterial ribonucleases and their roles in RNA metabolism. *Crit. Rev. Biochem. Mol. Biol.* **2019,** *54* (3), 242-300.

(22) 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.

(23) 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.

(24) 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.

(25) 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.

(26) Zuo, Y.; Wang, Y.; Malhotra, A., Crystal Structure of Escherichia coli RNase D, an Exoribonuclease Involved in Structured RNA Processing. *Structure* **2005,** *13* (7), 973–984.

(27) Callaghan, A. J.; Marcaida, M. J.; Stead, J. A.; McDowall, K. J.; Scott, W. G.; Luisi, B. F., Structure of Escherichia coli RNase E catalytic domain and implications for RNA turnover. *Nature* **2005,** *437* (7062), 1187–1191.

(28) Horton, N. C.; Perona, J. J., DNA cleavage by *Eco*RV endonuclease: two metal ions in three metal ion binding sites. *Biochemistry* **2004,** *43* (22), 6841–6857.

(29) Karakas, E.; Truglio, J. J.; Croteau, D.; Rhau, B.; Wang, L.; Van Houten, B.; Kisker, C., Structure of the C-terminal half of UvrC reveals an RNase H endonuclease domain with an Argonaute-like catalytic triad. *EMBO J.* **2007,** *26* (2), 613–622.

(30) Pillon, M. C.; Lorenowicz, J. J.; Uckelmann, M.; Klocko, A. D.; Mitchell, R. R.; Chung, Y. S.; Modrich, P.; Walker, G. C.; Simmons, L. A.; Friedhoff, P.; Guarné, A., Structure of the Endonuclease Domain of MutL: Unlicensed to Cut. *Mol. Cell* **2010,** *39* (1), 145–151.

(31) 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.

(32) Steiniger-White, M.; Rayment, I.; Reznikoff, W. S., Structure/function insights into Tn5 transposition. *Curr. Opin. Struct. Biol.* **2004,** *14* (1), 50–57.

(33) Hadden, J. M.; Déclais, A.-C.; Carr, S. B.; Lilley, D. M. J.; Phillips, S. E. V., The structural basis of Holliday junction resolution by T7 endonuclease I. *Nature* **2007,** *449* (7162), 621–624.

(34) Midtgaard, S. F.; Assenholt, J.; Jonstrup, A. T.; Van, L. B.; Jensen, T. H.; Brodersen, D. E., Structure of the nuclear exosome component Rrp6p reveals an interplay between the active site and the HRDC domain. *Proc. Natl. Acad. Sci. U.S.A.* **2006,** *103* (32), 11898.

(35) Wilson, K. A.; Fernandes, P. A.; Ramos, M. J.; Wetmore, S. D., Exploring the identity of the general base for a DNA polymerase catalyzed reaction using QM/MM: The case study of human translesion synthesis Polymerase η. *ACS Catal.* **2019,** *9* (3), 2543-2551.

(36) Leclerc, F.; Karplus, M., Two-metal-ion mechanism for hammerhead-ribozyme catalysis. *J. Phys. Chem. B* **2006,** *110* (7), 3395-3409.

(37) Stahley, M. R.; Strobel, S. A., RNA splicing: group I intron crystal structures reveal the basis of splice site selection and metal ion catalysis. *Curr. Opin. Struct. Biol.* **2006,** *16* (3), 319–326.

(38) Lilley, D. M. J., Structure, folding and mechanisms of ribozymes. *Curr. Opin. Struct. Biol.* **2005,** *15* (3), 313–323.

(39) Freudenthal, B. D.; Beard, W. A.; Cuneo, M. J.; Dyrkheeva, N. S.; Wilson, S. H., Capturing snapshots of APE1 processing DNA damage. *Nat. Struct. Mol. Biol.* **2015,** *22* (11), 924–931.

(40) 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.

(41) Xie, F.; Qureshi, S. H.; Papadakos, G. A.; Dupureur, C. M., One- and two-metal ion catalysis: global single-turnover kinetic analysis of the PvuII endonuclease mechanism. *Biochemistry* **2008,** *47* (47), 12540–12550.

(42) Kurpiewski, M. R.; Engler, L. E.; Wozniak, L. A.; Kobylanska, A.; Koziolkiewicz, M.; Stec, W. J.; Jen-Jacobson, L., Mechanisms of Coupling between DNA Recognition Specificity and Catalysis in EcoRI Endonuclease. *Structure* **2004,** *12* (10), 1775–1788.

(43) Pommer, A. J.; Cal, S.; Keeble, A. H.; Walker, D.; Evans, S. J.; Kühlmann, U. C.; Cooper, A.; Connolly, B. A.; Hemmings, A. M.; Moore, G. R.; James, R.; Kleanthous, C., Mechanism and cleavage specificity of the H-N-H endonuclease colicin E9. *J. Mol. Biol.* **2001,** *314* (4), 735-749.

(44) Czene, A.; Tóth, E.; Németh, E.; Otten, H.; Poulsen, J.-C. N.; Christensen, H. E.; Rulíšek, L.; Nagata, K.; Larsen, S.; Gyurcsik, B., A new insight into the zinc-dependent DNA-cleavage by the colicin E7 nuclease: a crystallographic and computational study. *Metallomics* **2014,** *6* (11), 2090–2099.

(45) Bueren-Calabuig, J. A.; Coderch, C.; Rico, E.; Jimenez-Ruiz, A.; Gago, F., Mechanistic insight into the catalytic activity of ββα-metallonucleases from computer simulations: *Vibrio vulnificus* periplasmic nuclease as a test case. *Chembiochem* **2011,** *12* (17), 2615–2622.

(46) Friedhoff, P.; Kolmes, B.; Gimadutdinow, O.; Wende, W.; Krause, K. L.; Pingoud, A., Analysis of the mechanism of the serratia nuclease using site-directed mutagenesis. *Nucleic Acids Res.* **1996,** *24* (14), 2632-2639.

(47) Yoon, H.; Zhao, L. N.; Warshel, A., Exploring the Catalytic Mechanism of Cas9 Using Information Inferred from Endonuclease VII. *ACS Catal.* **2019,** *9* (2), 1329–1336.

(48) Aaqvist, J.; Warshel, A., Calculations of free energy profiles for the staphylococcal nuclease catalyzed reaction. *Biochemistry* **1989,** *28* (11), 4680–4689.

(49) Kriukiene, E., Domain organization and metal ion requirement of the Type IIS restriction endonuclease MnlI. *FEBS Lett.* **2006,** *580* (26), 6115–6122.

(50) Shen, B. W.; Landthaler, M.; Shub, D. A.; Stoddard, B. L., DNA binding and cleavage by the HNH homing endonuclease I-HmuI. *J. Mol. Biol.* **2004,** *342* (1), 43-56.

(51) Vasu, K.; Nagamalleswari, E.; Zahran, M.; Imhof, P.; Xu, S.-y.; Zhu, Z.; Chan, S.-H.; Nagaraja, V., Increasing cleavage specificity and activity of restriction endonuclease KpnI. *Nucleic Acids Res.* **2013,** *41* (21), 9812-9824.

(52) Chan, S.-H.; Opitz, L.; Higgins, L.; O'Loane, D.; Xu, S.-Y., Cofactor requirement of HpyAV restriction endonuclease. *PLoS One* **2010,** *5* (2), e9071–e9071.

(53) Belkebir, A.; Azeddoug, H., Metal ion dependence of DNA cleavage by SepMI and EhoI restriction endonucleases. *Microbiol. Res.* **2013,** *168* (2), 99-105.

(54) Galburt, E. A.; Chevalier, B.; Tang, W.; Jurica, M. S.; Flick, K. E.; Monnat, R. J.; Stoddard, B. L., A novel endonuclease mechanism directly visualized for I-*Ppo*I. *Nat. Struct. Mol. Biol.* **1999,** *6* (12), 1096-1099.

(55) Mannino, S. J.; Jenkins, C. L.; Raines, R. T., Chemical mechanism of DNA cleavage by the homing endonuclease I-*Ppo*I. *Biochemistry* **1999,** *38* (49), 16178–16186.

(56) 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.

(57) Casalino, L.; Nierzwicki, L.; 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.

(58) 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.

(59) Rosta, E.; Yang, W.; Hummer, G., Calcium inhibition of ribonuclease H1 two-metal ion catalysis. *J. Am. Chem. Soc.* **2014,** *136* (8), 3137-3144.

(60) Himo, F., Recent trends in quantum chemical modeling of enzymatic reactions. *J. Am. Chem. Soc.* **2017,** *139* (20), 6780–6786.

(61) Alonso-Cotchico, L.; Sciortino, G.; Vidossich, P.; Rodriguez-Guerra Pedregal, J.; Drienovska, I.; Roelfes, G.; Lledos, A.; Marechal, J.-D., Integrated computational study of the Cu-catalyzed hydration of alkenes in water solvent and into the context of an artificial metallohydratase. *ACS Catal.* **2019,** *9* (5), 4616–4626.

(62) 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.

(63) 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.; Li, H.-B.; Ding, L.; Morokuma, K., The ONIOM method and its applications. *Chem. Rev.* **2015,** *115* (12), 5678–5796.

(64) Siegbahn, P. E. M.; Himo, F., The quantum chemical cluster approach for modeling enzyme reactions. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011,** *1* (3), 323-356.

(65) Balhara, R.; Chatterjee, R.; Jindal, G., A computational approach to understand the role of metals and axial ligands in artificial heme enzyme catalyzed C–H insertion. *Phys. Chem. Chem. Phys.* **2021,** *23* (15), 9500–9511.

(66) Kim, W.-C.; Berquist, B. R.; Chohan, M.; Uy, C.; Wilson, D. M.; Lee, C. H., Characterization of the endoribonuclease active site of human Apurinic/Apyrimidinic Endonuclease 1. *J. Mol. Biol.* **2011,** *411* (5), 960–971.

(67) 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.

(68) Tsutakawa, S. E.; Shin, D. S.; Mol, C. D.; Izumi, T.; Arvai, A. S.; Mantha, A. K.; Szczesny, B.; Ivanov, I. N.; Hosfield, D. J.; Maiti, B.; Pique, M. E.; Frankel, K. A.; Hitomi, K.; Cunningham, R. P.; Mitra, S.; Tainer, J. A., Conserved structural chemistry for incision activity in structurally non-homologous apurinic/apyrimidinic endonuclease APE1 and endonuclease IV DNA repair enzymes. *J. Biol. Chem.* **2013,** *288* (12), 8445–8455.

(69) Kanazhevskaya, L. Y.; Koval, V. V.; Lomzov, A. A.; Fedorova, O. S., The role of Asn-212 in the catalytic mechanism of human endonuclease APE1: Stopped-flow kinetic study of incision activity on a natural AP site and a tetrahydrofuran analogue. *DNA Repair* **2014,** *21*, 43–54.

(70) Malfatti, M. C.; Antoniali, G.; Codrich, M.; Tell, G., Coping with RNA damage with a focus on APE1, a BER enzyme at the crossroad between DNA damage repair and RNA processing/decay. *DNA Repair* **2021,** *104*, 103133.

(71) Housh, K.; Jha, J. S.; Yang, Z.; Haldar, T.; Johnson, K. M.; Yin, J.; Wang, Y.; Gates, K. S., Formation and repair of an interstrand DNA cross-link arising from a common endogenous lesion. *J. Am. Chem. Soc.* **2021,** *143* (37), 15344–15357.

(72) Hoitsma, N. M.; Click, T. H.; Agarwal, P. K.; Freudenthal, B. D., Altered APE1 activity on abasic ribonucleotides is mediated by changes in the nucleoside sugar pucker. *Comput. Struct. Biotechnol. J.* **2021,** *19*, 3293–3302.

(73) Liu, T.-C.; Lin, C.-T.; Chang, K.-C.; Guo, K.-W.; Wang, S.; Chu, J.-W.; Hsiao, Y.-Y., APE1 distinguishes DNA substrates in exonucleolytic cleavage by induced space-filling. *Nat. Commun.* **2021,** *12* (1), 1–12.

(74) L Illuzzi, J.; M Wilson III, D., Base excision repair: contribution to tumorigenesis and target in anticancer treatment paradigms. *Curr. Med. Chem.* **2012,** *19* (23), 3922–3936.

(75) Sykora, P.; Wilson, D. M.; Bohr, V. A., Base excision repair in the mammalian brain: Implication for age related neurodegeneration. *Mechanisms of Ageing and Development* **2013,** *134* (10), 440–448.

(76) Thakur, S.; Dhiman, M.; Tell, G.; Mantha, A. K., A review on protein–protein interaction network of APE1/Ref-1 and its associated biological functions. *Cell Biochem. Funct.* **2015,** *33* (3), 101–112.

(77) Stetler, R. A.; Gao, Y.; Leak, R. K.; Weng, Z.; Shi, Y.; Zhang, L.; Pu, H.; Zhang, F.; Hu, X.; Hassan, S., APE1/Ref-1 facilitates recovery of gray and white matter and neurological function after mild stroke injury. *Proc. Natl. Acad. Sci. U.S.A.* **2016,** *113* (25), E3558–E3567.

(78) Caston, R. A.; Gampala, S.; Armstrong, L.; Messmann, R. A.; Fishel, M. L.; Kelley, M. R., The multifunctional APE1 DNA repair–redox signaling protein as a drug target in human disease. *Drug Discov. Today* **2021,** *26* (1), 218–228.

(79) Chen, W.; Wang, S.; Xing, D., New Horizons for the Roles and Association of APE1/Ref-1 and ABCA1 in Atherosclerosis. *J. Inflamm. Res.* **2021,** *14*, 5251.

(80) Heisel, C.; Yousif, J.; Mijiti, M.; Charizanis, K.; Brigell, M.; Corson, T. W.; Kelley, M. R., APE1/Ref-1 as a novel target for retinal diseases. *J. Cell. Signal.* **2021,** *2* (2), 133.

(81) Lowry, D. F.; Hoyt, D. W.; Khazi, F. A.; Bagu, J.; Lindsey, A. G.; Wilson, D. M., Investigation of the Role of the Histidine–Aspartate Pair in the Human Exonuclease III-like Abasic Endonuclease, Ape1. *J. Mol. Biol.* **2003,** *329* (2), 311–322.

(82) Alekseeva, I. V.; Bakman, A. S.; Vorobjev, Y. N.; Fedorova, O. S.; Kuznetsov, N. A., Role of Ionizing Amino Acid Residues in the Process of DNA Binding by Human AP Endonuclease 1 and in Its Catalysis. *J. Phys. Chem. B* **2019,** *123* (45), 9546–9556.

(83) Lipton, A. S.; Heck, R. W.; Primak, S.; McNeill, D. R.; Wilson, D. M.; Ellis, P. D., Characterization of Mg2+ Binding to the DNA Repair Protein Apurinic/Apyrimidic Endonuclease 1 via Solid-State 25Mg NMR Spectroscopy. *J. Am. Chem. Soc.* **2008,** *130* (29), 9332–9341.

(84) Prejano, M.; Marino, T.; Russo, N., QM cluster or QM/MM in computational enzymology: the test case of LigW-decarboxylase. *Front. Chem. (Lausanne, Switz.)* **2018,** *6*, 249/1–249/9.

(85) Cheng, Q.; DeYonker, N. J., QM-Cluster Model Study of the Guaiacol Hydrogen Atom Transfer and Oxygen Rebound with Cytochrome P450 Enzyme GcoA. *J. Phys. Chem. B.* **2021,** *125* (13), 3296–3306.

(86) Planas, F.; McLeish, M. J.; Himo, F., Enzymatic Stetter Reaction: Computational Study of the Reaction Mechanism of MenD. *ACS Catal.* **2021,** *11* (19), 12355–12366.

(87) Su, J.-X.; Chen, S.-L., A Key Piece in the Global N-Cycle: The N–N Bond Formation Presented by Heme-Dependent Hydrazine Synthase. *ACS Catal.* **2021,** *11* (11), 6489–6498.

(88) Lai, R.; Cui, Q., Differences in the Nature of the Phosphoryl Transfer Transition State in Protein Phosphatase 1 and Alkaline Phosphatase: Insights from QM Cluster Models. *J. Phys. Chem. B.* **2020,** *124* (42), 9371–9384.

(89) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H., Electrostatic basis for enzyme catalysis. *Chem. Rev.* **2006,** *106* (8), 3210–3235.

(90) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.;

Farkas, O.; Foresman, J. B.; Fox, D. J.; Gonzalez, J.; Pople, J., Gaussian 09, Revision E. 01, Gaussian, Inc, Wallingford, CT, 2011. *Inc., Wallingford CT* **2009**.

(91) Alberto, M. E.; Pinto, G.; Russo, N.; Toscano, M., Triesterase and promiscuous diesterase activities of a di-CoII-containing organophosphate degrading enzyme reaction mechanisms. *Chem. Eur. J.* **2015,** *21* (9), 3736–3745.

(92) Araujo, A. R.; Ribeiro, A. J.; 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.

(93) Li, Z.; Wu, Y.; Feng, L.-J.; Wu, R.; Luo, H.-B., Ab Initio QM/MM Study Shows a Highly Dissociated SN2 Hydrolysis Mechanism for the cGMP-Specific Phosphodiesterase-5. *J. Chem. Theory Comput.* **2014,** *10* (12), 5448–5457.

(94) Rahimian, M.; Yeole, S. D.; Gejji, S. P., Mechanistic insights for β-cyclodextrin catalyzed phosphodiester hydrolysis. *J. Mol. Model.* **2014,** *20* (4), 2198.

(95) Zhang, X.; Gao, H.; Xu, H.; Xu, J.; Chao, H.; Zhao, C., A density functional theory study of the hydrolysis mechanism of phosphodiester catalyzed by a mononuclear Zn(II) complex. *J. Mol. Catal. A Chem.* **2013,** *368*, 53-60.

(96) Gao, H.; Ke, Z.; DeYonker, N. J.; Wang, J.; Xu, H.; Mao, Z.-W.; Phillips, D. L.; Zhao, C., Dinuclear Zn(II) Complex Catalyzed Phosphodiester Cleavage Proceeds via a Concerted Mechanism: A Density Functional Theory Study. *J. Am. Chem. Soc.* **2011,** *133* (9), 2904–2915.

(97) Acosta-Silva, C.; Bertran, J.; Branchadell, V.; Oliva, A., Phosphoryl Transfer Reaction in RNA: Is the Substrate-Assisted Catalysis a Possible Mechanism in Certain Solvents? *J. Phys. Chem. A* **2017,** *121* (44), 8525–8534.

(98) 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.

(99) Ribeiro, A. J. M.; Ramos, M. J.; Fernandes, P. A., Benchmarking of DFT Functionals for the Hydrolysis of Phosphodiester Bonds. *J. Chem. Theory Comput.* **2010,** *6* (8), 2281–2292.

(100) Mardirossian, N.; Head-Gordon, M., Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Mol. Phys.* **2017,** *115* (19), 2315–2372.

(101) Paul, A.; Mishra, S., Metal–ion promiscuity of microbial enzyme DapE at its second metal-binding site. *J. Biol. Inorg. Chem.* **2021**, 1–14.

(102) Medina, F. E.; Jaña, G. A., QM/MM Study of a VIM-1 Metallo-β-Lactamase Enzyme: The Catalytic Reaction Mechanism. *ACS Catal.* **2022,** *12* (1), 36–47.

(103) Berta, D.; Buigues, P. J.; Badaoui, M.; Rosta, E., Cations in motion: QM/MM studies of the dynamic and electrostatic roles of H^+ and Mg^{2+} ions in enzyme reactions. *Curr. Opin. Struct. Biol.* **2020,** *61*, 198-206.

(104) Yagi, K.; Ito, S.; Sugita, Y., Exploring the Minimum-Energy Pathways and Free-Energy Profiles of Enzymatic Reactions with QM/MM Calculations. *J. Phys. Chem. B* **2021,** *125* (18), 4701–4713.

(105) Redrejo-Rodríguez, M.; Vigouroux, A.; Mursalimov, A.; Grin, I.; Alili, D.; Koshenov, Z.; Akishev, Z.; Maksimenko, A.; Bissenbaev, A. K.; Matkarimov, B. T., Structural comparison of AP endonucleases from the exonuclease III family reveals new amino acid residues in human AP endonuclease 1 that are involved in incision of damaged DNA. *Biochimie* **2016,** *128*, 20–33.

(106) Kumar, D.; Thiel, W.; De Visser, S. P., Theoretical study on the mechanism of the oxygen activation process in cysteine dioxygenase enzymes. *J. Am. Chem. Soc.* **2011,** *133* (11), 3869–3882.

(107) Sundararajan, M.; Neese, F., Detailed QM/MM study of the electron paramagnetic resonance parameters of nitrosyl myoglobin. *J. Chem. Theory Comput.* **2012,** *8* (2), 563–574.

(108) Lai, R.; Cui, Q., What Does the Brønsted Slope Measure in the Phosphoryl Transfer Transition State? *ACS Catal.* **2020,** *10* (23), 13932–13945.

(109) Siegbahn, P. E. M.; Himo, F., Recent developments of the quantum chemical cluster approach for modeling enzyme reactions. *J. Biol. Inorg. Chem.* **2009,** *14* (5), 643–651.

(110) Kazemi, M.; Sheng, X.; Kroutil, W.; Himo, F., Computational study of Mycobacterium smegmatis acyl transferase reaction mechanism and specificity. *ACS Catal.* **2018,** *8* (11), 10698–10706.

(111) Prejanò, M.; Medina, F. E.; Ramos, M. J.; Russo, N.; Fernandes, P. A.; Marino, T., How the destabilization of a reaction intermediate affects enzymatic efficiency: The case of human transketolase. *ACS Catal.* **2020,** *10* (4), 2872–2881.

Chapter 3: The Metal Dependence of Single-Metal Mediated Phosphodiester Bond Cleavage: A QM/MM Study of a Multifaceted Human Enzyme

Preface: Chapter 3 has been published [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]. In this chapter, all authors contributed to the reviewing and editing of the manuscript. I aided conceptualization of the project and performed the computational characterization of the phosphodiester bond cleavage pathways involving Mg^{2+}, Ni^{2+} (square-planar and octahedral), Zn^{2+}, and Ca^{2+}. In addition, I performed the overall data/results analysis, wrote the first draft of the manuscript, and generated all the figures for the publication. D. J. Nikkel performed calculations on the pathway involving Mn^{2+} and contributed to the formal data analysis. M. M. Aboelnga contributed towards conceptualization and project administration, while S. D. Wetmore was involved in project conceptualization and administration, supervision, funding and resource acquisition, data visualization and interpretation, and writing the original draft of the manuscript.

3.1. Introduction

Metalloenzymes are among the most powerful enzymes in human cells, providing $10^{10} – 10^{23}$ fold rate increases over uncatalyzed transformations with the assistance of many different metal cofactors[1] and thereby permitting unique biochemistry. Metallonucleases are particularly interesting as these biocatalysts are responsible for cleaving the highly

stable phosphodiester backbones in nucleic acids, which have estimated half-lives of up to 30 million years under mild conditions.[1] The most well-established catalytic mechanism for phosphodiester bond cleavage involves two Mg^{2+} ions, the first producing an active site hydroxide that attacks the electrophilic phosphorus atom to initiate the reaction and the second stabilizing the charge on the phosphate group to be cleaved through direct ligation.[2-4] Nevertheless, literature is growing that suggests three[2, 4-7] or even a single metal ion[3-4] can support phosphodiester bond scission activity, although the associated catalytic mechanisms are currently debated.

A particularly interesting nuclease is human apurinic/apyrimidinic endonuclease (APE1) due to its significant roles in human health,[8-9] including inflammatory responses, diabetic neuropathy, inflammatory bowel disease, macular edema and degeneration, atherosclerosis, and liver, lung, bladder and pancreatic cancers.[10-15] Indeed, cells genetically engineered to be deficient in APE1 are not viable,[16] overexpression of APE1 has been directly connected to disease,[12] and APE1 has shown promise as a therapeutic target.[14] APE1 is perhaps best known for its role in DNA repair,[17] cleaving the phosphodiester backbone 5' with respect to cytotoxic abasic sites (AP sites)[18] as part of the base excision repair (BER) pathway.[19-20] However, APE1 also exhibits 3′-5′-exonuclease and 3′-phosphatase activity and is involved in many other critical cellular pathways, such as nucleotide incision repair (NIR), interstrand cross-link repair, RNA processing, transcription regulation, and gene expression.[16, 21-26] APE1 is also noteworthy because its activity has been shown to rely on the presence of only a single metal.[9, 19-20, 27] When coupled with its wide-ranging functions, this makes APE1 an interesting case study for

experimental and computational works geared toward understanding catalytic pathways for phosphodiester bond cleavage.

The general proposed two-step APE1 catalytic mechanism[17-20, 28-31] involves attack of a nucleophilic water on the phosphorus reaction center to generate a phosphorane intermediate followed by proton transfer from a metal-activated water to stabilize the leaving group, yielding the 3′-hydroxyl and 5′-phosphate products (Figure 3.1a). Nevertheless, the catalytic pathway remains controversial at least in part due to different metal coordination environments observed in crystal structures. Indeed, although an octahedral coordination geometry consistently occurs that involves E96 and active site waters, the substrate is ligated to the metal in the product complex (PDB ID: 4IEM; Figure 3.1b),[18] while D70 is coordinated to the metal in the apo-enzyme (PDB ID: 4QHE)[20] and reactant analogue (PDB ID: 5DG0; Figure 3.1c).[19] The lack of direct coordination of the substrate to the metal is supported by previous QM cluster calculations, which determined the metal binding architecture present in the crystal structure of the reactant analogue to be the most energetically favorable.[9] There are also varied proposals for the roles of active site amino acid residues, which have recently been reported to be pH dependent.[17] A D210N mutation has the largest impact on APE1 catalytic efficiency (10,000-fold reduction),[18] suggesting that D210 is most likely responsible for activating the water nucleophile, although other roles such as neutralizing the leaving group have been previously proposed.[9, 29] Other significant mutational effects occur for Y171F (1,200-fold reduction),[18-19] which has led to proposals that Y171 positions the substrate and water nucleophile[9, 17] or attacks the phosphate group,[32] and H309N (2,500-fold reduction),[18] which supports suggestions that H309 aids nucleophilic activation or substrate charge stabilization.[9, 18-19, 33] Mutational

effects also occur for N212 (N212A results in a 7,000-fold rate reduction),[34] which may orient and/or stabilize the general base (D210).[9, 17] Smaller rate reductions were seen upon mutation of N174 (N174Q leads to a 9-fold reduction),[35] which has been proposed to align the abasic site,[19] and D308 (D308A reduces the activity by 5–25 fold),[36] which is within hydrogen-bonding distances to metal ligated waters[18-20] and has been proposed to be involved in metal binding or positioning.[31, 37]

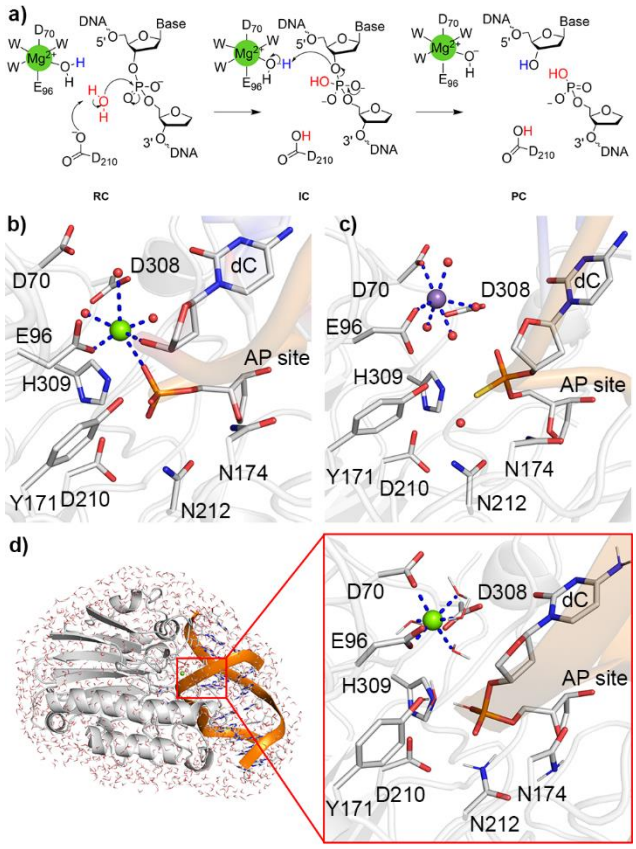

Figure 3.1. a) Proposed stepwise mechanism for APE1-catalyzed phosphodiester bond cleavage. Active site of APE1 from X-ray crystal structure of the b) Mg^{2+}-containing product complex (PDB ID: 4IEM) and c) Mn^{2+}-containing thio-substituted reactant analogue complex (PDB ID: 5G0). d) The QM/MM model (9445 atoms, left), and the QM region (129 atoms, red box, right) used in the present work.

Despite uncertainties regarding the roles of active site residues, a metal cofactor is accepted to be essential for APE1 activity.[38] Although Mg^{2+} is the native metal, APE1 remains active in the presence of several metals, with the catalytic rate decreasing as $Mg^{2+} > Mn^{2+} > Ni^{2+} > Zn^{2+}$.[38-40] The preference for Mg^{2+} and retention of activity in the presence of other metals correlates with the behaviour reported for other proposed single-metal dependent nucleases (*Moraxella nonliquefaciens* restriction endonuclease II (MnII), group I intron-encoded endonuclease from *Physarum polycephalum* (I-*Ppo*l) and HNH homing endonuclease (I-Hmul)),[41-43] while other potential enzymes in this class exhibit enhanced activity with transition metals (Klebsiella pneumoniae restriction endonuclease (Kpn1), *Enterobacter hormaeche* restriction endonuclease (Ehol), and *Helicobacter pylori* restriction endonuclease (HpyAV)).[44-46] In contrast to transition metals, APE1 activity is completely abolished in the presence of Ca^{2+}.[38-40] The replacement of Mg^{2+} by Ca^{2+} has been shown to inhibit catalysis and permit crystallization of many two-metal dependent nucleases,[47-49] a concept supported by computational work.[47-50] While some single-metal dependent nucleases are also inhibited by Ca^{2+} (I-Hmul and *Staphylococcus epidermidis* restriction endonuclease (SepMI)),[42, 46] others maintain catalytic activity albeit at a reduced rate (Kpn1, HpyAV, MnlI, and I-*Ppo*l).[41, 43-45] Although the impact of the metal identity across single-metal dependent nucleases points to the high plasticity of these active sites, the fundamental reason for the observed metal dependence is unclear. Indeed, while computational studies have provided atomic level explanations for the metal dependence of a variety of enzymes,[47, 49, 51-54] including two-metal mediated nucleases,[51-52, 54-55] no such work has been done to date on a single-metal dependent nuclease.

To provide the first atomic-level explanations for the observed metal dependence of a single-metal mediated nuclease, the present work uses computational chemistry to rationalize the metal-dependent activity of APE1. Due to the key insights provided by combined quantum mechanics–molecular mechanics (QM/MM) methodology within the ONIOM formalism for the catalytic mechanism of other enzymes,[56-57] including metalloenzymes,[58-59] QM/MM (ONIOM) is used to map the phosphodiester bond cleavage mechanism catalyzed by APE1 in the presence of a broad spectrum of metals, namely Mg^{2+}, Mn^{2+}, Ni^{2+}, Zn^{2+}, and Ca^{2+}. The structural differences in the characterized stationary points along the reaction pathway facilitated by the various metals rationalize the trend in the associated barrier heights, which correlates with the experimentally reported relative APE1 activity.[38-40] As a result, this work provides computational support for the controversial single-metal mediated phosphodiester bond cleavage mechanism and clarifies the ambiguity regarding the role of the metal and metal identity in this important reaction. This detailed appreciation of the activity of a one-metal dependent nuclease can be more broadly applied to understand other nucleases as well as other enzymes that share similar active sites or mechanistic features (e.g., DNase-I or inositol 5′-phosphatase (IPP) that use metals to catalyze the hydrolysis of substrates ranging from ss/dsDNA, RNA:DNA hybrids, or inositol polyphosphates).[60] This work also paves the way for exploiting APE1 and similar enzymes to engineer artificial metalloenzymes with improved catalytic functions, that facilitate novel chemistry for medicinal applications (e.g., disease diagnosis or drug design), or as biochemical tools (e.g., to detect nucleic acid modifications).

3.2. Computational Methodology

The previously proposed two-step phosphodiester bond cleavage mechanism (Figure 3.1a)[9, 60] was used as the basis to explore the catalytic pathway for APE1 in the presence of five different metals (Mg^{2+}, Mn^{2+}, Ni^{2+}, Zn^{2+}, and Ca^{2+}). Specifically, the QM/MM model was built using a previously reported representative structure from MD simulations,[9] which were initiated from the crystal structure of the APE1–DNA product complex (PDB ID: 4IEM, Figure 3.1b), while modifying the metal coordination geometry to match the preferred architecture identified using QM cluster models.[9] The physiological protonation states of titratable amino acid residues were verified using the propKa server (Table B.1, Appendix B). The QM/MM model includes the entire APE1–DNA system, the metal ion, 15 Na^+ counterions to neutralize the system, and all water molecules for which any atom is within 4 Å of any atom in the DNA–protein complex (total 9445 atoms and 1453 water molecules, Figure 3.1d).

The QM region (129 atoms) contains the metal, the AP site and 5′-dC, five water molecules (including the nucleophile), and all active site residues whose mutation impacts catalysis (i.e., D70, E96, D210, Y171, H309, N212, N174 and D308).[18-19, 29, 31, 33-36] The total charge of the QM region is –2. This represents the most complete QM region for APE1 considered in the literature to date, which is critical for the investigation of the metal dependence of the catalytic activity since accommodation of different sized metals with unique chemical properties may influence the active site configuration.

The full catalytic mechanism was individually mapped for each metal. While Mg^{2+}, Zn^{2+} and Ca^{2+} were considered with a spin multiplicity of 1, Mn^{2+} was treated in the sextet ground state (octahedral coordination geometry), while Ni^{2+} was considered in both the

singlet state (square planar coordination geometry) and triplet state (octahedral). During all geometry optimizations, the QM atoms were described using the B3LYP-D3(BJ) functional, which includes Grimme's D3 empirical dispersion correction with Becke-Johnson damping (D3-BJ), in combination with the 6-31G(d,p) basis set.[61-63] The remaining atoms were included in the MM region, which was treated with the AMBER (ff14SB) force field[64] without constraints on the APE1-DNA complex or water. The QM/MM boundary was described using mechanical embedding (ME), which has proven to be a powerful tool for investigating the catalytic mechanism of many DNA repair,[9, 65] metal-dependent[66-67] and other enzymes.[56, 68] We note that previous computational work on APE1 reported insignificant differences between the geometries and energy barriers obtained with ME versus electronic embedding (EE).[9] The current study implements a larger QM region (addition of D308) and different level of theory compared to the previous literature,[9] which is justified based on improved agreement with the experimentally reported incision rate.[19] The corresponding experimental incision rate constant (36 s^{-1}) was measured at 37°C and pH 7.5 with 100nM DNA substrate and 30 nM of APE1.[19] Additional details of comparisons to the experimental data can be found in the Results and Discussion.

Initially, each reactant complex (RC) was optimized (including the enzyme, DNA, metal, and water). Subsequently, an initial guess for the corresponding transition state (TS) complex was obtained by manually scanning important reaction coordinates. Specifically, in a stepwise approach, the distance between the nucleophilic water and the electrophilic phosphorus of the scissile phosphate moiety was decreased, constrained, and the remainder of the system optimized to identify a guess for TS1. Similarly, the distance between the phosphorus and O3′ leaving group was increased to isolate a guess for TS2. The structures

corresponding to the energy maximum on the resulting potential energy surfaces were then fully optimized. The nature of the fully optimized TS was confirmed using frequency calculations (single imaginary frequency) at the same level of theory. Finally, intrinsic reaction coordinate (IRC) calculations[69-70] were performed to identify the corresponding minima (i.e., RC, product complex (PC) or intermediate complex (IC)). The ICs obtained from the corresponding IRCs for both TS1 and TS2 have similar geometrical and energetic features (Figure B.1, Appendix B), highlighting that the entire pathway is connected. Frequency calculations were performed at same level of theory to confirm the nature of all stationary points and determine thermal corrections to the Gibbs energy. The relative energies for all APE1–DNA models were obtained using ONIOM(M06-2X/6-311+G(2df,p):AMBER) single-point calculations. For comparison, single-point calculations were also performed with ONIOM(B3LYP-D3(BJ)/6-311+G(2df,p):AMBER) for the Mg^{2+}-containing complexes, which resulted in small (< 8 kJ/mol) overall barrier reductions. Although metastable ICs were observed for some metals following the addition of thermal corrections to the single-point energies (Tables B.2–B.3, Appendix B), attempts to isolate TSs corresponding to concerted mechanisms were unsuccessful.

We acknowledge that some uncertainty in the reported relative energies for each pathway arises since a single catalytically relevant active site conformation was considered rather than exploring conformational sampling along the free energy reaction pathway. Nevertheless, the high levels of theory employed coupled with the necessarily large QM region make our QM/MM (ONIOM) approach the most tractable for reliably studying this chemistry. The consistency of our trend in the calculated barriers and the experimentally

observed reduction in enzymatic activity as a function of metal, as well as our structural analysis, further testifies to this point.

The Gaussian 16 program (revision B.01) was used for all calculations.[71]

3.3. Results and Discussion

3.3.1. Octahedral Coordination of a Single Magnesium Permits APE1 to Efficiently Catalyze Phosphodiester Bond Cleavage

Experimental kinetic data suggests that APE1 achieves optimum catalytic activity in the presence of Mg^{2+}.[19-20, 38-40] In the corresponding QM/MM RC (Figures 3.2 and 3.3), D210 is suitably positioned to abstract a proton from an active site water ($r(O_{D210}\cdots H_W)$ = 1.819 Å), which is appropriately aligned with respect to the phosphate backbone to act as the nucleophile ($r(O_W\cdots P)$ = 2.620 Å, Figure 3.3). A catalytically conducive relative orientation of the nucleophilic water and substrate is adopted through hydrogen bonds involving key active site residues. Specifically, N212 hydrogen bonds to D210, Y171 hydrogen bonds to the water nucleophile (Figures 3.2 and B.2, Appendix B), and H309 donates a proton to and hydrogen bonds with a non-bridging oxygen of the substrate at the phosphodiester bond cleavage site. N174 provides further stabilization of the relative nucleophile and substrate arrangement by hydrogen bonding to N212 and interacting with O4′ of the abasic sugar (Figure B.2, Appendix B). Mg^{2+} adopts octahedral coordination that includes D70, E96 and four active site waters, with an average distance across all ligands of 2.082 ± 0.034 Å (Figure B.3, Appendix B). D308 forms hydrogen bonds to Mg^{2+} ligated water molecules (Figure B.3, Appendix B), including a water that interacts with the

bridging oxygen at the substrate cleavage site. The predicted RC correlates with the X-ray crystal structure for the reactant analogue (Figure B.4a, Appendix B).[19]

In the first transition state (TS1), a proton is partially transferred from the nucleophilic water to D210 and the hydrogen bond between N212 and D210 lengthens (by 0.3 Å) as water approaches the phosphate moiety (Figures 3.3 and B.2, Appendix B). The remaining active site interactions are highly similar in the RC and TS1 (RMSD = 0.14 Å, Figure B.2 and B.5a, Appendix B). The overall barrier for nucleophilic attack is 64.8 kJ/mol (Figure 3.4 and Table B.3, Appendix B), which leads to a phosphorane intermediate (Figures 3.3 and B.2, Appendix B). The IC has structural features similar to those characterized for two metal-mediated phosphodiester bond cleavage reactions catalyzed by other enzymes.[2, 72-74] Specifically, the O_W–P bond is shortened and the P–O3′ bond is elongated relative to the TS1, and a proton has been transferred to D210 (Figure 3.3). All other active site features are similar to TS1, including the metal coordination geometry and substrate interactions with H309 and N212 (Figure B.2, Appendix B). Thus, the IC is only slightly (3.9 kJ/mol) more stable than TS1, falling 60.9 kJ/mol above the RC (Figure 3.4 and Table B.3, Appendix B).

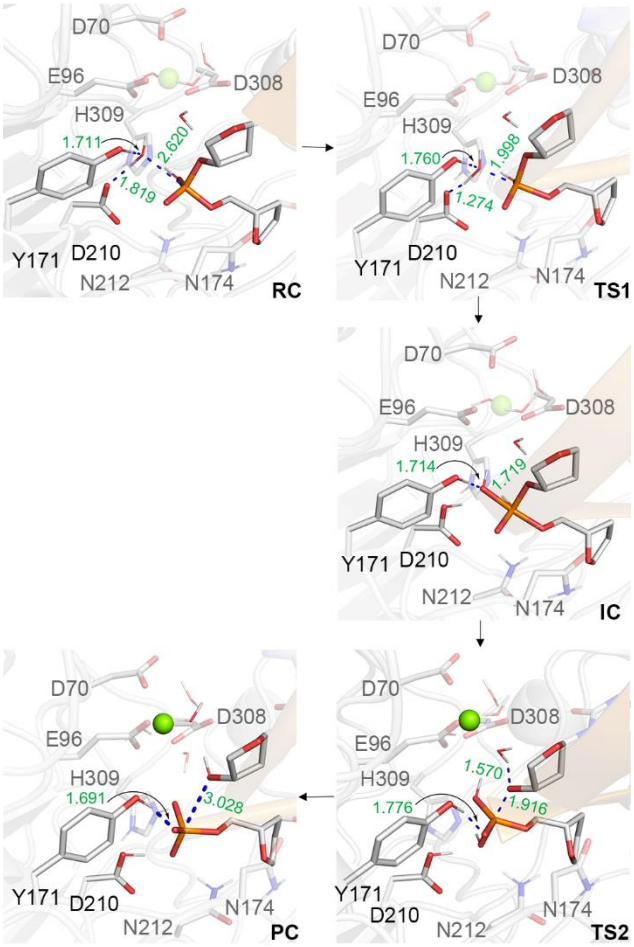

Figure 3.2. QM/MM optimized stationary points along the Mg^{2+}-catalyzed APE1 mechanism with key bond distances (Å). Important residues for each step are highlighted in dark grey tubes, while the remainder of the QM region is shown in light grey tubes.

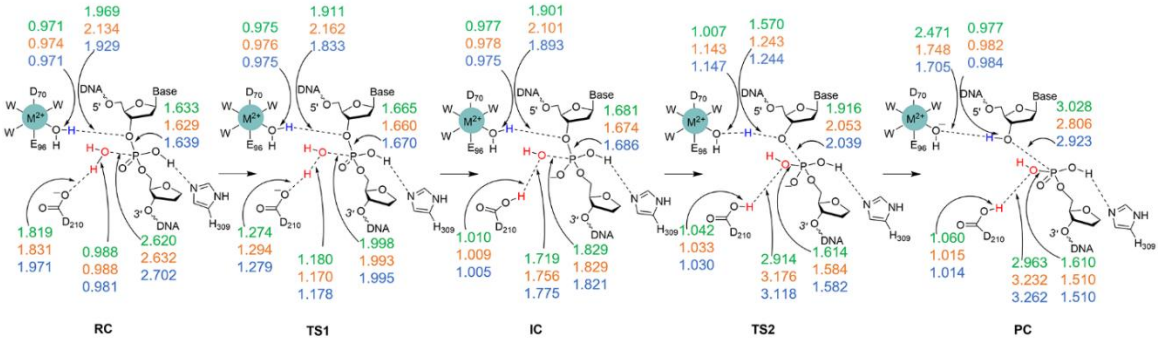

Figure 3.3. Mechanism and key calculated bond distances (Å) for the APE1-catalyzed phosphodiester bond cleavage facilitated by Mg^{2+} (green), Mn^{2+} (orange), or Ni^{2+} (octahedral coordination, blue).

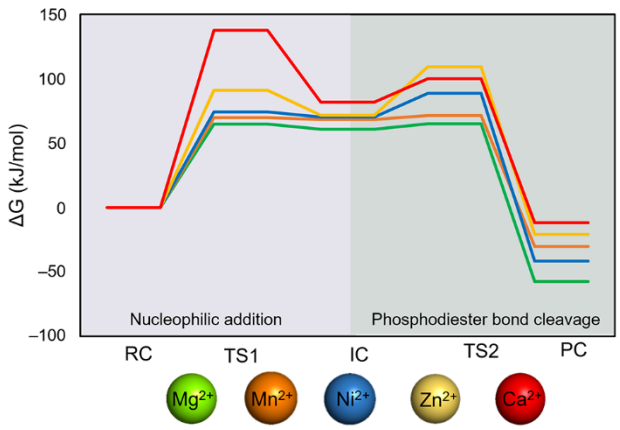

Figure 3.4. Relative Gibbs energy barriers for the APE1-catalyzed phosphodiester bond cleavage facilitated by Mg^{2+} (green), Mn^{2+} (orange), Ni^{2+} (blue, octahedral), Zn^{2+} (yellow), and Ca^{2+} (red).

In the second step of the reaction, the metal activated water transfers a proton to O3′ of the leaving group to cleave the phosphodiester bond. In the corresponding transition state (TS2), the proton on the Mg^{2+}-ligated water shifts toward O3′ (r(Hw···O3′) decreases by 0.3 Å compared to IC), while the P–O3′ bond is partially cleaved (bond lengthens by 0.2 Å, Figure 3.3). Although the substrate position shifts due to the pending bond cleavage, which alters APE1–substrate interactions (active site RMSD = 0.3 Å, Figure B.5c, Appendix B), substrate stabilization is provided through direct hydrogen bonds between nonbridging oxygen atoms and H309, N212 and Y171 (Figure B.2, Appendix B). While D308 no longer hydrogen bonds to the Mg^{2+}-coordinated water responsible for leaving group protonation, the octahedral metal binding architecture is maintained (Figure B.3, Appendix B). The resulting phosphodiester bond cleavage barrier is 65.1 kJ/mol (Figure 3.4 and Table B.3, Appendix B), which leads to a thermodynamically stable product (by 57.4 kJ/mol). The PC maintains the key orientations of active site residues observed in other reaction steps (Figure B.5, Appendix B) and the crystallized product complex (Figure B.4b,

RMSD = 0.894 Å, Appendix B),[18] with the experimentally reported high active site flexibility following the chemical step rationalizing deviations in the position of the metal and E96 side chain.[19-20, 75-76]

Overall, the predicted mechanism for APE1-catalyzed phosphodiester bond cleavage correlates with an abundance of experimental structural,[18-20, 29] mutational,[31, 33-36] and kinetic data,[19, 36] which lends confidence to the proposed mechanism and clarifies the catalytic roles of active site amino acid residues that have been disputed in the literature. Specifically, the metal binding architecture matches that in the crystal structure for the reactant analogue,[19] with D70 and E96 coordinated to Mg^{2+}. Furthermore, although not directly ligated to the metal,[37] D308 helps maintain the correct metal binding architecture by hydrogen bonding to several metal-ligated waters, including the water that protonates the substrate to cleave the phosphodiester bond, which supports the experimentally-observed contribution to catalysis.[36] The relative orientations of key active site residues and the substrate also agree with experimental structural data.[18-19] Although previously proposed to neutralize the leaving group,[31] the predicted role of D210 as the general base is rationalized by its relative position with respect to the substrate (as observed in crystal structure)[29] and the experimentally reported largest mutational effect for this residue.[18] The catalytically conducive orientations of D210, the nucleophilic water and the substrate are maintained through an intricate hydrogen-bonding network established in the early stages of the reaction. Indeed, N212 forms a strong hydrogen bond with D210, while N174 interacts with N212, which confirms the proposed roles for both residues of substrate stabilization and/or positioning.[9, 17, 19, 25, 34] While Y171 aids alignment of the nucleophile in the first reaction step, a change in active site conformation permits a role in substrate

stabilization and positioning in the second stage of the reaction, which is consistent with previous experimental and computational mutational data.[9] Regardless, as previously proposed,[18-19, 33] H309 provides the most significant charge stabilization to the substrate by neutralizing the backbone early in the catalytic pathway, which contrasts the proposal that H309 may activate the nucleophile.[18] Most importantly, the single metal plays a critical role in facilitating leaving group departure in the phosphodiester bond cleavage, which has a predicted barrier (65.1 kJ/mol) that is in close agreement with the estimated experimental barrier for the incision of a DNA substrate (66.8 kJ/mol).[19]

3.3.2. Despite a Similar Size as Magnesium, Subtle Changes in the Active Site Geometry upon Incorporation of Manganese Slightly Increase the Overall Reaction Barrier

Due to their similar size, Mg^{2+} and Mn^{2+} equally prefer to adopt an octahedral coordination geometry. Nevertheless, the chemistry of these metals is inherently different, with Mn^{2+} existing as a sextet and Mg^{2+} as a closed shell singlet.[77] Mn^{2+} can replace the metal in some Mg^{2+}-dependent enzymes, including DNA polymerases[74, 78-79] and endonucleases,[44-45, 80] while retaining the overall catalytic mechanism with similar or enhanced activity. In contrast, the metal in Mn^{2+}-containing systems cannot be replaced with Mg^{2+},[77] mainly due to the higher affinity of Mn^{2+} for N and S-containing ligands, as well as the low energetic cost for Mn^{2+} to change from hexa to penta or tetra-coordination.[77] Although the role of the metal in APE1 is in line with that of many native Mn^{2+}-containing enzymes,[74, 81] the replacement of Mg^{2+} by Mn^{2+} in the APE1 active site slightly decreases the catalytic activity[39-40] and the reasons for this behaviour are unclear.

Upon replacement of Mg^{2+} with Mn^{2+}, most QM/MM structural features along the wild-type (Mg^{2+}) pathway are maintained (active site RMSD < 0.2 Å, Figures 3.3, B.2 and B.6, Appendix B). Nevertheless, subtle structural differences exist. Specifically, although the octahedral coordination at the metal center is maintained in the RC, the average coordination distance is slightly longer for Mn^{2+} than Mg^{2+} (by 0.1 Å; Figure B.3, Appendix B). Furthermore, the distance between the proton in the metal-ligated water and the leaving group increases in TS1 in the presence of Mn^{2+} (by 0.25 Å, Figure 3.3) and remains 0.2 Å longer in the IC. This results in a more significant change when the IC is converted into TS2, with the distance between the leaving group and metal-activated water decreasing by 0.9 Å for Mn^{2+} compared to 0.3 Å for Mg^{2+} (Figures 3.3). As a result, Mn^{2+} leads to a slightly later second transition state (TS2) as evidenced by the (0.14 Å) longer P–O3′ bond and (0.33 Å) shorter hydrogen bond between the Mn^{2+}-ligated water and the substrate (Figure 3.3), which leads to a barrier of 71.5 kJ/mol for the phosphodiester bond cleavage step. The metal–hydroxide distance is also elongated in the PC for Mn^{2+} (by 0.1 Å, Figure B.3, Appendix B), suggesting less charge stabilization and leading to a less exergonic reaction (PC is –30.3 kJ/mol with respect to the RC).

Overall, since Mn^{2+} is a softer metal than Mg^{2+} (see computed charge-to-size ratio, Table B.4, Appendix B), the metal-ligated water is not as strongly activated in the IC, which results in a later transition state for the phosphodiester bond cleavage step. Furthermore, the hydroxide ion is less stabilized (i.e., more weakly coordinated to the metal, Figure 3.3), leading to a less exergonic product complex (Figure 3.4). When coupled with the similarity of the remaining APE1 active site structural features throughout the reaction (Figure B.6, Appendix B), these factors result in a slight (6.4 kJ/mol) increase in the barrier for the

phosphodiester bond cleavage step for the Mn^{2+} over Mg^{2+}-facilitated reaction. Although we acknowledge that these small differences in the computed barrier heights can arise from the choice of the single reactant complex used to map the reaction, our computed trend in barrier heights is consistent with the experimentally reported modest reduction in catalytic rate upon replacement of Mg^{2+} with Mn^{2+}.[39-40]

3.3.3. Regardless of Whether a Square Planar or Octahedral Coordination Geometry is Adopted, Nickel Increases the Barrier for APE1-Catalyzed Phosphodiester Bond Cleavage

Ni^{2+} is a borderline metal that prefers nitrogen and sulfur-containing ligands, but has also been found bound to harder oxygen-containing ligands.[82] Ni^{2+} can exist in both low-spin complexes with square planar coordination and high-spin complexes with octahedral coordination, depending on the ligand environment.[83] Indeed, in Ni^{2+}-dependent enzymes, the metal can maintain a square-planar coordination geometry with nitrogen or sulfur-containing ligands (e.g., [NiFe]-hydrogenase, CO dehydrogenase, methyl-coenzyme M reductase, acetyl-coenzyme A synthase)[82] or an octahedral coordination geometry with oxygen-containing ligands (e.g., glyoxalase I, acidoreductone dioxygenase, urease, *Klebsiella aerogenes* accessory protein (UreE) and *Helicobacter pylori* nickel-dependent repressor (NiKR)).[82-83] When Ni^{2+} replaces Mg^{2+} in the APE1 active site, the catalytic rate decreases.[39-40] To explore the structural and energetic impact of this metal substitution, both the square-planar and octahedral metal binding architectures reported for other enzymes were considered within the APE1–DNA complex.

To adopt the square planar coordination geometry, two water molecules originally coordinated to Mg^{2+} in the APE1 RC become unbound to Ni^{2+} and instead interact with other (low layer) residues (N68 or the 5′-dC with respect to the abasic site, Figure B.7, Appendix B). However, both D70 and E96 remain directly coordinated to the metal center. The metal ligands are bound more tightly to Ni^{2+} than Mg^{2+}, with the average metal coordination distance being 0.2 Å shorter (Figure B.7, Appendix B). Although there are many subtle structural differences in the stationary points along the Mg^{2+} and Ni^{2+}-mediated pathways due to deviations in the metal coordination geometry, the relative orientations of key active site residues and the substrate are generally maintained (Figure B.8 and B.9, Appendix B). As a result, the barriers for nucleophilic attack in the presence of Mg^{2+} and Ni^{2+} are within 2 kJ/mol (Table B.2, Appendix B). However, the leaving group is 0.1–0.3 Å closer to the proton of the metal-ligated water for square-planar Ni^{2+} compared to octahedral Mg^{2+} throughout the reaction (Figure B.8, Appendix B). Furthermore, the distance between the metal-activated water and the leaving group in the second reaction step (i.e., from IC to TS2) decreases by 0.5 Å for Ni^{2+} compared to 0.3 Å for Mg^{2+} (Figures 3.3 and B.10, Appendix B). Overall, these changes result in a rate-limiting barrier of 74.7 kJ/mol in the presence of square-planar coordinated Ni^{2+}, which is 9.6 kJ/mol above the barrier for the wild-type (Mg^{2+}-facilitated) reaction.

Although the square-planar metal binding architecture is well accommodated in the APE1 active site, the large number of oxygen-containing ligands suggest that Ni^{2+} may preferentially adopt an octahedral coordination geometry as found for other enzymes.[82-83] Indeed, the APE1 RC corresponding to octahedral metal coordination (triplet state) is 187.7 kJ/mol more stable than that containing square-planar Ni^{2+} (singlet state), likely due at least

in part to the active site rearrangement required to adopt the square-planar binding configuration. When the octahedral Ni^{2+} coordination geometry is considered, the average metal coordination distance in the RC is only slightly (0.02 Å) shorter than that for Mg^{2+} (Figure B.3, Appendix B). However, D70 no longer hydrogen bonds to a neighboring water molecule (3.037 Å, Figure B.3, Appendix B), a binding configuration previously shown to be less stable for Mg^{2+}.[9] Furthermore, the change in the metal size impacts the active site geometry such that the nucleophile is further from D210 (by 0.15 Å) and the phosphorus center of attack (by 0.1 Å, Figure 3.3). As a result, the barrier for the first reaction step increases by 9.5 kJ/mol in the presence of Ni^{2+} (Figure 3.4 and Table B.3, Appendix B). Although the IC complexes are highly similar (Figure B.11, Appendix B), the TS associated with the second step is later for Ni^{2+} than Mg^{2+}. Specifically, the P–O3′ distance is 0.1 Å longer and the distance between the substrate and metal activated water is 0.3 Å shorter (Figure 3.3). Furthermore, octahedral Ni^{2+} has an even lower charge-to-size ratio than Mn^{2+} (Table B.4, Appendix B), which further reduces the acidity of the metal bound water. These factors collectively increase the barrier for the phosphodiester bond cleavage to 88.8 kJ/mol.

In summary, the APE1 active site can accommodate Ni^{2+} in both a square planar and octahedral coordination geometry. However, the reactant complex with the octahedral metal binding configuration is significantly more stable, emphasizing that the active site environment plays a large role in determining the optimal metal coordination geometry. Regardless of the metal binding architecture, the barrier for the phosphodiester bond cleavage facilitated by Ni^{2+} is enzymatically feasible. Furthermore, the catalytic barrier is

predicted to be up to 23.7 kJ/mol larger in the presence of Ni^{2+} than Mg^{2+}, which is in line with the experimentally observed rate reduction in APE1 activity for Ni^{2+}.[39-40]

3.3.4. Zinc Permits APE1 to Achieve Catalysis by Changing the Metal Environment Along the Reaction Pathway

Zn^{2+} is a much stronger Lewis acid compared to Mg^{2+}. In fact, Zn^{2+} is the metal of choice for many metallohydrolases, such as carbonic anhydrase and β-lactamases.[83] Zn^{2+} is a borderline metal that prefers softer ligands, although it can also bind to O-containing molecules.[83] The coordination sphere of Zn^{2+} is more flexible than Mg^{2+}, varying from 4 to 6. Although Mg^{2+} cannot replace Zn^{2+}, Zn^{2+} has been shown to successfully replace Mg^{2+} in several enzymes including a response regulator protein CheY and catechol-O-methyltransferase.[50-51, 83] The reason for this unique behavior lies in the fact that the majority of Zn^{2+}-dependent enzymes bind the metal in buried or more rigid active sites that result in a tetrahedral metal-binding geometry to relieve coordination strain, an architecture not well-suited to Mg^{2+}.[83] In contrast, Zn^{2+} can adopt an octahedral coordination geometry in catalytic sites, typically those that are solvent exposed.[83-84] As the APE1 active site is also solvent exposed, it may not be surprising that the enzyme maintains activity in the presence of Zn^{2+}.[38, 40] Nevertheless, the experimentally observed reduction in the APE1 catalytic rate upon replacement of Mg^{2+} with Zn^{2+} has yet to be explained.[40]

Unlike the other metals discussed thus far, the structures and energetics along the APE1 catalytic pathway change considerably in the presence of Zn^{2+}. Although the RC complex is overall highly similar for Mg^{2+} and Zn^{2+} (active site RMSD = 0.06 Å, Figure

B.12, Appendix B), Y171 reorientates in TS1 to hydrogen bond with a nonbridging substrate oxygen of the scissile phosphate (Figure B.6, Appendix B) rather than the nucleophile as in the Mg^{2+} complex (Figure B.2, Appendix B). As a result, the metal-ligated water responsible for protonating the leaving group, hydrogen bonds to the water nucleophile to aid active site alignment in the first reaction step (Figure 3.5). These factors contribute to a larger barrier for nucleophilic attack in the presence of Zn^{2+} (91.1 kJ/mol, Table B.3, Appendix B). The structural differences at the metal center persist in the IC, which occurs further along the reaction coordinate compared to that for any other metal discussed thus far. Indeed, the distance between the metal-activated water and the leaving group decreases by 0.9 Å compared to TS1, while the nucleophile O–P bond is formed (1.662 Å) and the substrate P–O bond is elongated (1.733 Å, Figure 3.5). In TS2, Zn^{2+} becomes penta-coordinated due to detachment of D70 (Figures 3.5 and B.13, Appendix B), resulting in significant active site structural rearrangement near the metal and substrate (Figure B.12d, Appendix B). The adoption of similar penta-coordination along the reaction pathway has been reported for other native Zn^{2+}-containing enzymes, including phospholipase C,[85] nuclease P1,[85] and endonuclease IV.[5] This reorganization contributes to the very large rate-determining barrier (109.2 kJ/mol) and only slightly thermodynamically favorable product (–20.9 kJ/mol relative to the RC; Figure 3.4, Table B.3, Appendix B).

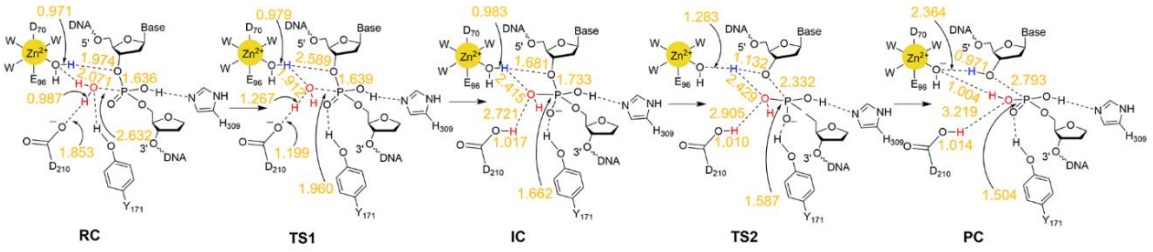

Figure 3.5. Mechanism and key calculated bond distances (Å) for the APE1-catalyzed phosphodiester bond cleavage facilitated by Zn^{2+}.

Overall, in contrast with the vast majority of Zn^{2+}-dependent enzymes, Zn^{2+} adopts an octahedral binding configuration at the start of the APE1 catalyzed reaction, which is consistent with literature suggesting that Zn^{2+} prefers octahedral coordination in more solvent exposed environments and the local architecture of the APE1 catalytic site. However, the metal coordination environment changes to penta-coordinated during the Zn^{2+}-catalyzed reaction. Although this is the first time a decrease in the Zn^{2+} coordination number (from 6 to 5) during a reaction has been reported, this finding parallels changes from tetra to penta-coordination observed along the pathway catalyzed by Zn^{2+}-dependent enzymes.[5, 85] This flexibility in metal coordination rationalizes why softer metals can be accommodated by Mg^{2+}-dependent enzymes. Indeed, although the energetic costs to reorganize the metal coordination environment and active site more broadly result in a high barrier, APE1 is still catalytically active in the presence of Zn^{2+}. The 44.1 kJ/mol larger barrier predicted for Zn^{2+} compared to Mg^{2+}-facilitated phosphodiester bond cleavage correlates with the experimentally observed larger reduction in the APE1 catalytic rate upon metal replacement for Zn^{2+} compared to the other metals discussed above.[40] Importantly, the impact of replacing Mg^{2+} is larger for Zn^{2+} than all metals discussed above despite the larger charge-to-size ratio of Zn^{2+} compared to Ni^{2+} (Table B.4, Appendix B), highlighting the significant effect of the change in coordination number during the reaction as well as the many factors that change the barrier upon metal replacement.

3.3.5. APE1 Activity is Maintained in the Presence of Different Metal Co-factors for a Variety of Reasons, Highlighting the Mg^{2+} Preference yet Promiscuity of the Active Site

As emphasized in the Introduction, many previous experimental studies have highlighted the broad metal specificity for single-metal dependent nucleases,[41-46] including APE1.[38-40] Our pioneering computational study has provided the structural data required to understand this catalytic behaviour. Importantly, our work reveals that an interplay between several factors, each being unique to a specific metal, is responsible for increased, yet catalytically viable, barriers for the APE1 reaction in the presence of transition metals. Specifically, although similar metal size has been used to explain preserved active site conformations and therefore catalytic activity in the presence of Mg^{2+} and Mn^{2+} for a variety of enzymes,[74, 78] our work highlights that the inherent chemistry of Mn^{2+} results in subtle structural changes in the APE1 active site. Indeed, the change in the charge-to-size ratio (Table B.4, Appendix B) decreases the acidity of the metal coordinated water that is responsible for protonating the leaving group, decreases the stability of the PC and increases barrier for the phosphodiester bond cleavage step. In contrast, Ni^{2+} results in greater structural changes in the APE1 active site, disrupting the relative orientation of the water nucleophile, general base, and substrate at the start of the reaction, resulting in a later transition state for the phosphodiester bond cleavage step, and leading to an even greater barrier. Although octahedral coordination is maintained for Mg^{2+}, Mn^{2+} and Ni^{2+} throughout the APE1-catalyzed reaction, incorporation of Zn^{2+} changes the metal coordination geometry along the reaction pathway, which further increases the overall reaction barrier. In fact, the computed trend in the barriers with respect to the active site

metal (Mg^{2+} < Mn^{2+} < Ni^{2+} < Zn^{2+}, Figure 3.4) is consistent with the experimentally reported changes in the APE1 catalytic rates (Mg^{2+} > Mn^{2+} > Ni^{2+} > Zn^{2+}).[38-40] Therefore, in addition to rationalizing the metal dependence of APE1, our work underscores the power of computational studies for understanding metalloenzyme function. Our predicted structural changes induced in the APE1 active site by different metals impact the catalytic barrier complement a recent experimental study[86] that used X-ray crystallographic data to explain changes in the catalytic rate with different metals for carbonic anhydrase II, a single-metal dependent enzyme that catalyzes the reversible hydrolysis of carbon dioxide. Overall, when combined with previous experimental studies, our work emphasizes that the preference of APE1 for Mg^{2+} may arise in part due to greater bioavailability, but likely also because of several competing factors such as the inherent chemistry of the metal (e.g., high charge-to-size ratio, hardness, acidity of the metal coordinated water) and the active site environment (e.g., oxo-ligands, octahedral metal coordination). Indeed, although the calculated metal charge-to-size ratio (Table B.4, Appendix B) correlates with the trend in APE1 barrier reduction for Mg^{2+} > Mn^{2+} > Ni^{2+}, which maintain octahedral coordination throughout the reaction pathway, the change in coordination geometry for Zn^{2+} during the reaction results in the highest barrier.

3.3.6. Changes in the Metal Coordination Geometry Upon Accommodation of Calcium in the APE1 Active Site Affect the Role of the Metal and Completely Inhibit Phosphodiester Bond Cleavage

Although both hard metals that prefer oxygen-containing ligands, Ca^{2+} is significantly larger than Mg^{2+}.[83, 87] Indeed, the coordination number of Ca^{2+} in proteins

varies from 6 to 8,[83] allowing more ligands to bind to the metal center. Furthermore, the higher charge density makes Mg^{2+} a better Lewis acid than Ca^{2+}.[48] Therefore, the majority of native Mg^{2+}-containing enzymes are generally inhibited by Ca^{2+}, including ATPases,[87] polymerases,[78] nucleases,[47-49] dUTPases,[88] adenylyl cyclases,[49] and kinases.[89] Nevertheless, some enzymes, including HpyAV restriction endonuclease[44] and topoisomerase II,[90] show enhanced activity in the presence of Ca^{2+} over Mg^{2+}. In the case of APE1, Ca^{2+} completely abolishes the catalytic activity.[38, 40]

Substitution of Mg^{2+} with Ca^{2+} results in significant rearrangement of the APE1 active site in the RC (RMSD = 0.4 Å, Figure B.14a, Appendix B). This arises, at least in part, because Ca^{2+} adopts a hepta-coordinated geometry. Specifically, Ca^{2+} maintains coordination with the same residues as Mg^{2+} (D70, E96 and 4 water molecules) and the seventh ligand is the nucleophilic water ($r(Ca^{2+}\cdots O_W)$ = 2.725 Å, Figure B.15, Appendix B). The possibility a new water molecule as the nucleophile after the original water nucleophile directly coordinates with Ca^{2+} was discarded as there is no space for an additional water molecule to bind in the active site, which agrees with previous QM/MM free energy calculations for two-metal dependent RNaseH1.[49] Although the average ligand coordination distance to Ca^{2+} (2.479 ± 0.106 Å) is also 0.4 Å longer than that to Mg^{2+}, the coordination of the nucleophilic water to the metal hinders activation by D210, which lies 3.025 Å away from the nucleophile (Figure 3.6). Furthermore, in the transition state corresponding to nucleophilic attack, the Ca^{2+} coordination necessarily changes from the preferred hepta-coordination to octa-coordination (Figure B.15, Appendix B), resulting in greater active site rearrangement along the reaction for Ca^{2+} (RMSD= 0.45 Å, Figure B.16a, Appendix B) compared to Mg^{2+} (RMSD= 0.19 Å, Figure B.14b, Appendix B). As a result,

the barrier for the first reaction step is very high (137.7 kJ/mol, Figure 3.4 and Table B.3, Appendix B). Although the metal binding architecture reverts to hepta-coordination for the remainder of the reaction, D70 becomes unbound from Ca^{2+} in TS2 in favor of direct coordination with the leaving group of the substrate (Figures 3.6 and B.15, Appendix B), which yields a barrier for the second step of 100.1 kJ/mol. In fact, the leaving group remains directly coordinated to Ca^{2+} in the PC, which leads to a nearly thermodynamically neutral reaction (PC falls only 11.9 kJ/mol below RC). This mechanism contrasts protonation of the leaving group by a metal-activated water, but resembles the metal binding architecture in the X-ray crystal structure of the product analogue (PDB ID: 4IEM; Figures 3.1b and B.17, Appendix B) and the role of a similarly positioned metal in the two-metal mediated reaction.[18]

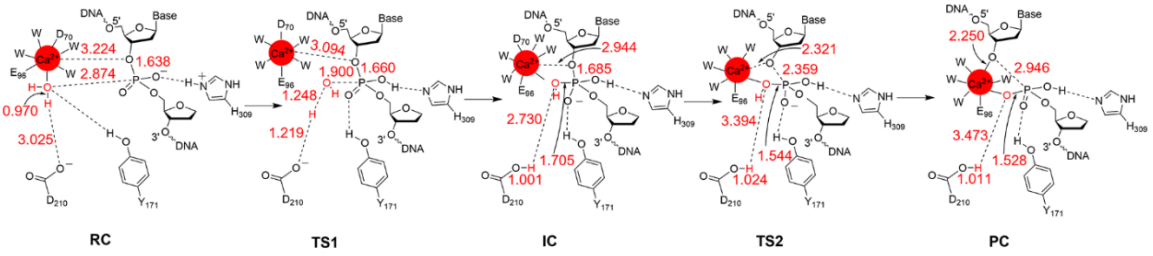

Figure 3.6. Mechanism and key calculated bond distances (Å) for the APE1-catalyzed phosphodiester bond cleavage facilitated by Ca^{2+}.

Overall, weakening of the nucleophile in the reactant complex, the change of the metal coordination geometry throughout the reaction pathway, and the direct ligation of the substrate to the metal during the phosphodiester bond cleavage step (rather than substrate protonation) collectively change the APE1 rate-determining step in the presence of Ca^{2+} compared to Mg^{2+} and the transition metals. Furthermore, the rate-limiting reaction barrier

(137.7 kJ/mol, Figure 3.3 and Table B.3, Appendix B) is not enzymatically feasible. These structural and energetic findings rationalize the experimentally-observed complete abolishment of APE1 activity in the presence of Ca^{2+}.[38, 40, 91] Indeed, crystals of an APE1–DNA complex with the phosphodiester bond intact have been grown in the presence of $CaCl_2$.[76] Furthermore, similar metal hepta-coordination and phosphodiester bond cleavage barriers (~130 kJ/mol) have been computationally predicted for the two-metal mediated bond cleavage facilitated by RNase H[48-49] and BamHI in the presence of calcium.[47] Therefore, the present work extends the concept of Ca^{2+} inhibition of phosphodiester bond cleavage to single-metal dependent nucleases. In fact, the replacement of only a single Mg^{2+} in the two-metal dependent RNase H that is bound with respect to the substrate in a similar way as the metal in APE1 also abolishes catalysis.[49] Thus, our work further emphasizes that one metal is enough to facilitate phosphodiester bond cleavage reactions and supports the proposed role of the single metal in the second step of this difficult reaction.

3.4. Conclusion

In this study, QM/MM calculations were used to map the APE1-faciliated phosphodiester bond cleavage pathway in the presence of a range of metals (Mg^{2+}, Mn^{2+}, Ni^{2+}, Zn^{2+}, and Ca^{2+}) as a prototypical example of a single-metal dependent nuclease. Despite an overall favored octahedral coordination geometry, our work highlights that replacing native Mg^{2+} with Mn^{2+}, Ni^{2+} or Zn^{2+} results in unique structural differences in the active site along the reaction pathway that differentially increase the barrier for the phosphodiester bond cleavage step for a variety of reasons. Nevertheless, all of these metals result in an enzymatically feasible barrier, highlighting the promiscuity of the APE1 active

site towards different metal cofactors. In contrast, Ca^{2+} substitution expands the APE1 metal coordination from octa to heptacoordination, which results in a prohibitively high barrier and enzyme inhibition. The trend in the computed barriers as a function of the various metal co-factors and the inhibition by calcium correlate with the experimentally observed changes in the APE1 catalytic rate, which provides additional support for the single-metal mediated mechanism that is disputed in the literature. This suggests that the preference of APE1 for Mg^{2+} likely arises for a variety of reasons, including greater bioavailability, inherent chemistry of the metal, and the active site environment. The clarified role of the metal and the rationalized dependence of the catalytic activity on the metal identity are critical for exploiting APE1 as a therapeutic target for disease diagnosis or treatment.[92-93] Additionally, previous work on other enzymes emphasizes that this new structural insight can be used to guide the engineering of new proteins with different metal preferences for improved substrate binding[94] or enhanced catalytic activity,[95] or that can even facilitate new chemical transformations.[96]

3.5. References

(1) Mikkola, S.; Lönnberg, T.; Lönnberg, H., Phosphodiester models for cleavage of nucleic acids. *Beilstein J. Org. Chem* **2018,** *14* (1), 803-837.

(2) 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.

(3) Yang, W., Nucleases: diversity of structure, function and mechanism. *Q. Rev. Biophys.* **2011,** *44* (1), 1–93.

(4) 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.

(5) Ivanov, I.; Tainer, J. A.; McCammon, J. A., Unraveling the three-metal-ion catalytic mechanism of the DNA repair enzyme endonuclease IV. *Proc. Natl. Acad. Sci. U.S.A.* **2007,** *104* (5), 1465.

(6) Ho, M.-H.; De Vivo, M.; Dal Peraro, M.; Klein, M. L., Understanding the Effect of Magnesium Ion Concentration on the Catalytic Activity of Ribonuclease H through Computation: Does a Third Metal Binding Site Modulate Endonuclease Catalysis? *J. Am. Chem. Soc.* **2010,** *132* (39), 13702–13712.

(7) Raper, A. T.; Reed, A. J.; Suo, Z., Kinetic Mechanism of DNA Polymerases: Contributions of Conformational Dynamics and a Third Divalent Metal Ion. *Chem. Rev.* **2018,** *118* (12), 6000–6025.

(8) Wu, Z.; Duan, H.; Cheng, Y.; Guo, D.; Peng, L.; Hu, Y.; Hu, J.; Luo, T., A novel ligand swing-mediated active site coordination change of human apurinic/apyrimidinic endonuclease 1: a potential cytotoxic mechanism of nickel ion in the base excision repair. *Chem. Phys.* **2022**, 111456.

(9) 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.

(10) Hartman, G. D.; Lambert-Cheatham, N. A.; Kelley, M. R.; Corson, T. W., Inhibition of APE1/Ref-1 for neovascular eye diseases: From biology to therapy. *Int. J. Mol. Sci.* **2021,** *22* (19), 10279.

(11) Long, K.; Gu, L.; Li, L.; Zhang, Z.; Li, E.; Zhang, Y.; He, L.; Pan, F.; Guo, Z.; Hu, Z., Small-molecule inhibition of APE1 induces apoptosis, pyroptosis, and necroptosis in non-small cell lung cancer. *Cell Death Dis.* **2021,** *12* (6), 1-15.

(12) Lu, X.; Zhao, H.; Yuan, H.; Chu, Y.; Zhu, X., High nuclear expression of APE1 correlates with unfavorable prognosis and promotes tumor growth in hepatocellular carcinoma. *J. Mol. Histol.* **2021,** *52* (2), 219–231.

(13) Mijit, M.; Caston, R.; Gampala, S.; Fishel, M. L.; Fehrenbacher, J.; Kelley, M. R., APE1/Ref-1–one target with multiple indications: Emerging aspects and new directions. *J. Cell. Signal.* **2021,** *2* (3), 151.

(14) Caston, R. A.; Gampala, S.; Armstrong, L.; Messmann, R. A.; Fishel, M. L.; Kelley, M. R., The multifunctional APE1 DNA repair–redox signaling protein as a drug target in human disease. *Drug Discov. Today* **2021,** *26* (1), 218-228.

(15) Chen, W.; Wang, S.; Xing, D., New horizons for the roles and association of APE1/Ref-1 and ABCA1 in atherosclerosis. *J. Inflamm. Res.* **2021,** *14*, 5251.

(16) Malfatti, M. C.; Antoniali, G.; Codrich, M.; Tell, G., Coping with RNA damage with a focus on APE1, a BER enzyme at the crossroad between DNA damage repair and RNA processing/decay. *DNA Repair* **2021,** *104*, 103133.

(17) Alekseeva, I. V.; Bakman, A. S.; Vorobjev, Y. N.; Fedorova, O. S.; Kuznetsov, N. A., Role of Ionizing Amino Acid Residues in the Process of DNA Binding by Human AP Endonuclease 1 and in Its Catalysis. *J. Phys. Chem. B* **2019,** *123* (45), 9546–9556.

(18) Tsutakawa, S. E.; Shin, D. S.; Mol, C. D.; Izumi, T.; Arvai, A. S.; Mantha, A. K.; Szczesny, B.; Ivanov, I. N.; Hosfield, D. J.; Maiti, B.; Pique, M. E.; Frankel, K. A.; Hitomi, K.; Cunningham, R. P.; Mitra, S.; Tainer, J. A., Conserved structural chemistry for incision activity in structurally non-homologous apurinic/apyrimidinic endonuclease APE1 and endonuclease IV DNA repair enzymes. *J. Biol. Chem.* **2013,** *288* (12), 8445-8455.

(19) Freudenthal, B. D.; Beard, W. A.; Cuneo, M. J.; Dyrkheeva, N. S.; Wilson, S. H., Capturing snapshots of APE1 processing DNA damage. *Nat. Struct. Mol. Biol.* **2015,** *22* (11), 924–931.

(20) 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.

(21) Tell, G.; Wilson, D. M.; Lee, C. H., Intrusion of a DNA Repair Protein in the RNome World: Is This the Beginning of a New Era? *Mol. Cell. Biol.* **2010,** *30* (2), 366.

(22) Barchiesi, A.; Bazzani, V.; Jabczynska, A.; Borowski, L. S.; Oeljeklaus, S.; Warscheid, B.; Chacinska, A.; Szczesny, R. J.; Vascotto, C., DNA repair protein APE1 degrades dysfunctional abasic mRNA in mitochondria affecting oxidative phosphorylation. *J. Mol. Biol.* **2021,** *433* (18), 167125.

(23) Housh, K.; Jha, J. S.; Yang, Z.; Haldar, T.; Johnson, K. M.; Yin, J.; Wang, Y.; Gates, K. S., Formation and repair of an interstrand DNA cross-link arising from a common endogenous lesion. *J. Am. Chem. Soc.* **2021,** *143* (37), 15344-15357.

(24) Kuznetsova, A.; Gavrilova, A.; Novopashina, D.; Fedorova, O.; Kuznetsov, N., Mutational and Kinetic Analysis of APE1 Endoribonuclease Activity. *Mol. Biol.* **2021,** *55* (2), 211–224.

(25) Hoitsma, N. M.; Click, T. H.; Agarwal, P. K.; Freudenthal, B. D., Altered APE1 activity on abasic ribonucleotides is mediated by changes in the nucleoside sugar pucker. *Comput. Struct. Biotechnol. J.* **2021,** *19*, 3293-3302.

(26) Liu, T.-C.; Lin, C.-T.; Chang, K.-C.; Guo, K.-W.; Wang, S.; Chu, J.-W.; Hsiao, Y.-Y., APE1 distinguishes DNA substrates in exonucleolytic cleavage by induced space-filling. *Nat. Commun.* **2021,** *12* (1), 1–12.

(27) Manvilla, B. A.; Pozharski, E.; Toth, E. A.; Drohat, A. C., Structure of human apurinic/apyrimidinic endonuclease 1 with the essential Mg2+ cofactor. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2013,** *69* (12), 2555–2562.

(28) Maher, R. L.; Bloom, L. B., Pre-steady-state kinetic characterization of the AP endonuclease activity of human AP endonuclease 1. *J. Biol. Chem.* **2007,** *282* (42), 30577–30585.

(29) Mol, C. D.; Izumi, T.; Mitra, S.; Tainer, J. A., DNA-bound structures and mutants reveal abasic DNA binding by APE1 DNA repair and coordination. *Nature* **2000,** *403* (6768), 451–456.

(30) Mundle, S. T.; Delaney, J. C.; Essigmann, J. M.; Strauss, P. R., Enzymatic Mechanism of Human Apurinic/Apyrimidinic Endonuclease against a THF AP Site Model Substrate. *Biochemistry* **2009,** *48* (1), 19–26.

(31) Erzberger, J. P.; Wilson, D. M., The role of Mg^{2+} and specific amino acid residues in the catalytic reaction of the major human abasic endonuclease: new insights from EDTA-resistant incision of acyclic abasic site analogs and site-directed mutagenesis. *J. Mol. Biol.* **1999,** *290* (2), 447–457.

(32) Mundle, S. T.; Fattal, M. H.; Melo, L. F.; Coriolan, J. D.; O'Regan, N. E.; Strauss, P. R., Novel role of tyrosine in catalysis by human AP endonuclease 1. *DNA Repair* **2004,** *3* (11), 1447–1455.

(33) Lowry, D. F.; Hoyt, D. W.; Khazi, F. A.; Bagu, J.; Lindsey, A. G.; Wilson, D. M., Investigation of the role of the histidine–aspartate pair in the human exonuclease III-like abasic endonuclease, APE1. *J. Mol. Biol.* **2003,** *329* (2), 311-322.

(34) Kanazhevskaya, L. Y.; Koval, V. V.; Lomzov, A. A.; Fedorova, O. S., The role of Asn-212 in the catalytic mechanism of human endonuclease APE1: Stopped-flow kinetic study of incision activity on a natural AP site and a tetrahydrofuran analogue. *DNA Repair* **2014,** *21*, 43–54.

(35) Redrejo-Rodríguez, M.; Vigouroux, A.; Mursalimov, A.; Grin, I.; Alili, D.; Koshenov, Z.; Akishev, Z.; Maksimenko, A.; Bissenbaev, A. K.; Matkarimov, B. T., Structural comparison of AP endonucleases from the exonuclease III family reveals new amino acid residues in human AP endonuclease 1 that are involved in incision of damaged DNA. *Biochimie* **2016,** *128*, 20–33.

(36) Kim, W.-C.; Berquist, B. R.; Chohan, M.; Uy, C.; Wilson, D. M.; Lee, C. H., Characterization of the endoribonuclease active site of human apurinic/apyrimidinic endonuclease 1. *J. Mol. Biol.* **2011,** *411* (5), 960-971.

(37) Masuda, Y.; Bennett, R. A.; Demple, B., Rapid dissociation of human apurinic endonuclease (Ape1) from incised DNA induced by magnesium. *J. Biol. Chem.* **1998,** *273* (46), 30360–30365.

(38) Barzilay, G.; Mol, C. D.; Robson, C. N.; Walker, L. J.; Cunningham, R. P.; Tainer, J. A.; Hickson, I. D., Identification of critical active-site residues in the multifunctional human DNA repair enzyme HAP1. *Nat. Struct. Mol. Biol.* **1995,** *2* (7), 561-568.

(39) Schermerhorn, K. M.; Delaney, S., Transient-state kinetics of apurinic/apyrimidinic (AP) Endonuclease 1 acting on an authentic AP Site and commonly used substrate analogs: The effect of diverse metal ions and base mismatches. *Biochemistry* **2013,** *52* (43), 7669–7677.

(40) Miroshnikova, A. D.; Kuznetsova, A. A.; Vorobjev, Y. N.; Kuznetsov, N. A.; Fedorova, O. S., Effects of mono- and divalent metal ions on DNA binding and catalysis of human apurinic/apyrimidinic endonuclease 1. *Mol. BioSyst.* **2016,** *12* (5), 1527–1539.

(41) Mannino, S. J.; Jenkins, C. L.; Raines, R. T., Chemical mechanism of DNA cleavage by the homing endonuclease I-*Ppo*I. *Biochemistry* **1999,** *38* (49), 16178–16186.

(42) Shen, B. W.; Landthaler, M.; Shub, D. A.; Stoddard, B. L., DNA binding and cleavage by the HNH homing endonuclease I-HmuI. *J. Mol. Biol.* **2004,** *342* (1), 43-56.

(43) Kriukiene, E., Domain organization and metal ion requirement of the Type IIS restriction endonuclease MnlI. *FEBS Lett.* **2006,** *580* (26), 6115-6122.

(44) Chan, S.-H.; Opitz, L.; Higgins, L.; O'Loane, D.; Xu, S.-Y., Cofactor requirement of HpyAV restriction endonuclease. *PLoS One* **2010,** *5* (2), e9071–e9071.

(45) Vasu, K.; Saravanan, M.; Nagaraja, V., Endonuclease active site plasticity allows DNA cleavage with diverse alkaline Earth and transition metal ions. *ACS Chem. Biol.* **2011,** *6* (9), 934–942.

(46) Belkebir, A.; Azeddoug, H., Metal ion dependence of DNA cleavage by SepMI and EhoI restriction endonucleases. *Microbiol. Res.* **2013,** *168* (2), 99-105.

(47) Mordasini, T.; Curioni, A.; Andreoni, W., Why do divalent metal ions either promote or inhibit enzymatic reactions? The case of BamHI restriction endonuclease from combined quantum-classical simulations. *J. Biol. Chem.* **2003,** *278* (7), 4381-4384.

(48) Babu, C. S.; Dudev, T.; Lim, C., Differential role of the protein matrix on the binding of a catalytic aspartate to Mg2+ vs Ca2+: application to ribonuclease H. *J. Am. Chem. Soc.* **2013,** *135* (17), 6541–6548.

(49) Rosta, E.; Yang, W.; Hummer, G., Calcium inhibition of ribonuclease H1 two-metal ion catalysis. *J. Am. Chem. Soc.* **2014,** *136* (8), 3137-3144.

(50) Valdez, C. E.; Smith, Q. A.; Nechay, M. R.; Alexandrova, A. N., Mysteries of metals in metalloenzymes. *Acc. Chem. Res.* **2014,** *47* (10), 3110–3117.

(51) Sparta, M.; Alexandrova, A. N., How metal substitution affects the enzymatic activity of catechol-O-methyltransferase. *PLoS One* **2012,** *7* (10), e47172.

(52) Elton, E. S.; Zhang, T.; Prabhakar, R.; Arif, A. M.; Berreau, L. M., Pb^{2+}-promoted amide cleavage: mechanistic comparison to a Zn^{2+} analogue. *Inorg. Chem.* **2013,** *52* (19), 11480–11492.

(53) Valdez, C. E.; Gallup, N. M.; Alexandrova, A. N., Co^{2+} acireductone dioxygenase: Fe^{2+} mechanism, Ni^{2+} mechanism, or something else? *Chem. Phys. Lett.* **2014,** *604*, 77–82.

(54) Balhara, R.; Chatterjee, R.; Jindal, G., A computational approach to understand the role of metals and axial ligands in artificial heme enzyme catalyzed C–H insertion. *Phys. Chem. Chem. Phys.* **2021,** *23* (15), 9500-9511.

(55) Paul, A.; Mishra, S., Metal–ion promiscuity of microbial enzyme DapE at its second metal-binding site. *J. Biol. Inorg. Chem.* **2021**, 1–14.

(56) Fernandes, H. S.; Ramos, M. J.; Cerqueira, N. M., Catalytic mechanism of the serine hydroxymethyltransferase: a computational ONIOM QM/MM study. *ACS Catal.* **2018,** *8* (11), 10096-10110.

(57) Jiang, C.; He, B.-B.; Zhao, R. L.; Xu, M.-J.; Houk, K.; Zhao, Y.-L., Computational Exploration of How Enzyme XimE Converts Natural S-Epoxide to Pyran and R-Epoxide to Furan. *ACS Catal.* **2021,** *11*, 7928–7942.

(58) Wilson, K. A.; Fernandes, P. A.; Ramos, M. J.; Wetmore, S. D., Exploring the identity of the general base for a DNA polymerase catalyzed reaction using QM/MM: The case study of human translesion synthesis Polymerase η. *ACS Catal.* **2019,** *9* (3), 2543-2551.

(59) Yuan, C.; Zhang, Y.; Tan, H.; Li, X.; Chen, G.; Jia, Z., ONIOM investigations of the heme degradation mechanism by MhuD: the critical function of heme ruffling. *Phys. Chem. Chem. Phys.* **2020,** *22* (16), 8817–8826.

(60) Schein, C. H.; Özgün, N.; Izumi, T.; Braun, W., Total sequence decomposition distinguishes functional modules, "molegos" in apurinic/apyrimidinic endonucleases. *BMC Bioinformatics* **2002,** *3* (1), 37.

(61) Becke, A. D.; Johnson, E. R., A density-functional model of the dispersion interaction. *J. Chem. Phys.* **2005,** *123* (15), 154101.

(62) Johnson, E. R.; Becke, A. D., A post-Hartree–Fock model of intermolecular interactions. *J. Chem. Phys.* **2005,** *123* (2), 024101.

(63) Grimme, S.; Ehrlich, S.; Goerigk, L., Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011,** *32* (7), 1456-1465.

(64) 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.

(65) Sebera, J.; Hattori, Y.; Sato, D.; Reha, D.; Nencka, R.; Kohno, T.; Kojima, C.; Tanaka, Y.; Sychrovsky, 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.

(66) Tolbert, A. E.; Ervin, C. S.; Ruckthong, L.; Paul, T. J.; Jayasinghe-Arachchige, V. M.; Neupane, K. P.; Stuckey, J. A.; Prabhakar, R.; Pecoraro, V. L., Heteromeric three-stranded coiled coils designed using a $Pb^{2+}(Cys)_3$ template mediated strategy. *Nat. Chem.* **2020,** *12* (4), 405–411.

(67) Garcia-Borràs, M.; Kan, S. J.; Lewis, R. D.; Tang, A.; Jimenez-Osés, G.; Arnold, F. H.; Houk, K. N., Origin and control of chemoselectivity in cytochrome c catalyzed carbene transfer into Si–H and N–H bonds. *J. Am. Chem. Soc.* **2021,** *143* (18), 7114–7123.

(68) Brás, N. r. F.; Santos-Martins, D.; Fernandes, P. A.; Ramos, M. J., Mechanistic pathway on human α-glucosidase maltase-glucoamylase unveiled by QM/MM calculations. *J. Phys. Chem. B* **2018,** *122* (14), 3889-3899.

(69) Fukui, K., The path of chemical reactions - the IRC approach. *Acc. Chem. Res.* **1981,** *14* (12), 363-368.

(70) Dykstra, C.; Frenking, G.; Kim, K.; Scuseria, G., Theory and applications of computational chemistry: the first forty years. Elsevier: 2011.

(71) Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Petersson, G.; Nakatsuji, H., Gaussian 16 Rev. B. 01, Wallingford, CT. **2016**.

(72) 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.

(73) Elsässer, B.; Valiev, M.; Weare, J. H., A Dianionic Phosphorane Intermediate and Transition States in an Associative AN+DN Mechanism for the RibonucleaseA Hydrolysis Reaction. *J. Am. Chem. Soc.* **2009,** *131* (11), 3869–3871.

(74) Cisneros, G. A.; Perera, L.; Schaaper, R. M.; Pedersen, L. C.; London, R. E.; Pedersen, L. G.; Darden, T. A., Reaction Mechanism of the ε Subunit of E. coli DNA Polymerase III: Insights into Active Site Metal Coordination and Catalytically Significant Residues. *J. Am. Chem. Soc.* **2009,** *131* (4), 1550–1556.

(75) Lipton, A.; Heck, R.; Primak, S.; McNeill, D.; Wilson Iii, D.; Ellis, P., Characterization of Mg^{2+} binding to the DNA repair protein apurinic/apyrimidic endonuclease 1 via solid-state ^{25}Mg NMR spectroscopy. *J. Am. Chem. Soc.* **2008,** *130* (29), 9332–9341.

(76) Whitaker, A. M.; Flynn, T. S.; Freudenthal, B. D., Molecular snapshots of APE1 proofreading mismatches and removing DNA damage. *Nat. Commun.* **2018,** *9* (1), 399.

(77) Sladek, V.; Tvaroška, I., First-Principles Interaction Analysis Assessment of the Manganese Cation in the Catalytic Activity of Glycosyltransferases. *J. Phys. Chem. B.* **2017,** *121* (25), 6148–6162.

(78) Chaudret, R.; Piquemal, J.-P.; Cisneros, G. A., Correlation between electron localization and metal ion mutagenicity in DNA synthesis from QM/MM calculations. *Phys. Chem. Chem. Phys.* **2011,** *13* (23), 11239–11247.

(79) Koag, M.-C.; Lee, S., Metal-dependent conformational activation explains highly promutagenic replication across O6-methylguanine by human DNA polymerase β. *J. Am. Chem. Soc.* **2014,** *136* (15), 5709–5721.

(80) Vasu, K.; Nagamalleswari, E.; Zahran, M.; Imhof, P.; Xu, S.-y.; Zhu, Z.; Chan, S.-H.; Nagaraja, V., Increasing cleavage specificity and activity of restriction endonuclease KpnI. *Nucleic Acids Res.* **2013,** *41* (21), 9812-9824.

(81) Velazquez-Libera, J. L.; Caballero, J.; Tuñón, I.; Hernandez-Rodriguez, E. W.; Ruiz-Pernía, J. J., On the nature of the enzyme–substrate complex and the reaction mechanism in human Arginase I. A combined molecular dynamics and QM/MM study. *ACS Catal.* **2020,** *10* (15), 8321–8333.

(82) Sydor, A. M.; Zamble, D. B., Nickel Metallomics: General Themes Guiding Nickel Homeostasis. In *Metallomics and the Cell*, Banci, L., Ed. Springer Netherlands: Dordrecht, 2013; pp 375–416.

(83) Dudev, T.; Lim, C., Competition among metal ions for protein binding sites: determinants of metal ion selectivity in proteins. *Chem. Rev.* **2014,** *114* (1), 538–556.

(84) Dudev, T.; Lim, C., Principles governing Mg, Ca, and Zn binding and selectivity in proteins. *Chem. Rev.* **2003,** *103* (3), 773–788.

(85) Liao, R.-Z.; Yu, J.-G.; Himo, F., Phosphate mono-and diesterase activities of the trinuclear zinc enzyme nuclease P1—Insights from quantum chemical calculations. *Inorg. Chem.* **2010,** *49* (15), 6883–6888.

(86) Kim, J. K.; Lee, C.; Lim, S. W.; Adhikari, A.; Andring, J. T.; McKenna, R.; Ghim, C.-M.; Kim, C. U., Elucidating the role of metal ions in carbonic anhydrase catalysis. *Nat. Commun.* **2020,** *11* (1), 1–10.

(87) Grauffel, C.; Dudev, T.; Lim, C., Why Cellular Di/Triphosphates Preferably Bind Mg2+ and Not Ca2+. *J. Chem. Theory Comput.* **2019,** *15* (12), 6992–7003.

(88) Lopata, A.; Jambrina, P. G.; Sharma, P. K.; Brooks, B. R.; Toth, J.; Vertessy, B. G.; Rosta, E., Mutations decouple proton transfer from phosphate cleavage in the dutpase catalytic reaction. *ACS Catal.* **2015,** *5* (6), 3225–3237.

(89) Lu, S. Y.; Huang, Z. M.; Huang, W. K.; Liu, X. Y.; Chen, Y. Y.; Shi, T.; Zhang, J., How calcium inhibits the magnesium-dependent kinase gsk3β: A molecular simulation study. *Proteins* **2013,** *81* (5), 740–753.

(90) Deweese, J. E.; Osheroff, N., The use of divalent metal ions by type II topoisomerases. *Metallomics* **2010,** *2* (7), 450–459.

(91) Beernink, P. T.; Segelke, B. W.; Hadi, M. Z.; Erzberger, J. P.; Wilson, D. M.; Rupp, B., Two divalent metal ions in the active site of a new crystal form of human

apurinic/apyrimidinic endonuclease, APE1: implications for the catalytic mechanism. *J. Mol. Biol.* **2001,** *307* (4), 1023–1034.

(92) Lelyveld, V. S.; Brustad, E.; Arnold, F. H.; Jasanoff, A., Metal-substituted protein MRI contrast agents engineered for enhanced relaxivity and ligand sensitivity. *J. Am. Chem. Soc.* **2011,** *133* (4), 649–651.

(93) Knutson, S. D.; Arthur, R. A.; Johnston, H. R.; Heemstra, J. M., Selective enrichment of A-to-I edited transcripts from cellular RNA using Endonuclease V. *J. Am. Chem. Soc.* **2020,** *142* (11), 5241-5251.

(94) Munoz Robles, V.; Ortega-Carrasco, E.; Alonso-Cotchico, L.; Rodriguez-Guerra, J.; Lledós, A.; Maréchal, J.-D., Toward the computational design of artificial metalloenzymes: From protein–ligand docking to multiscale approaches. *ACS Catal.* **2015,** *5* (4), 2469–2480.

(95) Sreenilayam, G.; Moore, E. J.; Steck, V.; Fasan, R., Metal substitution modulates the reactivity and extends the reaction scope of myoglobin carbene transfer catalysts. *Adv. Synth. Catal.* **2017,** *359* (12), 2076–2089.

(96) Biggs, G. S.; Klein, O. J.; Boss, S. R.; Barker, P. D., Unlocking the Full Evolutionary Potential of Artificial Metalloenzymes Through Direct Metal-Protein Coordination. *Johns. Matthey Technol. Rev.* **2020,** *64* (4), 407–418.

Chapter 4: 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

Preface: Chapter 4 has been submitted for a publication [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**, Accepted (manuscript ID: CP-ART-12-2023-006201.R1)]. In this chapter, I aided conceptualization of the project and performed the computational characterization of the phosphodiester bond cleavage pathways using QM cluster Model 4 and QM/MM model involving direct Mg–O3′ coordination and QM cluster Models 1–4 and QM/MM model involving indirect Mg–O3′ coordination. In addition, I performed the overall data/results analysis, wrote the first draft of the manuscript, contributed to reviewing and editing, and generated all figures for the publication. Angela Frederickson performed calculations using QM cluster Models 1–3 involving direct Mg–O3′ coordination, while S. D. Wetmore was involved in project conceptualization and administration, supervision, funding and resource acquisition, data visualization and interpretation, writing the original draft of the manuscript, and reviewing and editing.

4.1. Introduction

Homing endonucleases (HEs), which are also known as meganucleases, are highly specific DNA cleaving enzymes that are prevalent among bacteria, archaea, and eukarya.[1-2] HEs are known for their ability to self-propagate genetic elements, specifically moving or

transferring the genes that encode them into the host genome.[3] In particular, a homing endonuclease gene (HEG) is transcribed and spliced, and then translated into a HE, which generates double-strand DNA breaks in the uninterrupted host genome that does not contain the intervening sequence.[1] Subsequently, the HEG is incorporated into the host genome through homologous recombination repair in combination with DNA replication, completing the homing process.[1, 3] The human genome is known to be populated with a large number of such mobile genetic elements (~ 40%).[4] Although these insertion elements may help shape the human genome, their insertion has also been linked to human diseases such as hemophilia,[5] breast cancer,[6] and muscular dystrophy.[7]

Due to their ability to create double-strand breaks in complex genomes, HEs have been used for genome engineering,[8] including applications such as gene targeting,[9] cloning,[10] and therapy.[11-12] HE-directed gene therapy has been applied in various applications such as correcting the human XPC gene to treat xeroderma pigmentosum,[13-15] and the mutated Duchenne muscular dystrophy (DMD) gene associated with DMD.[16] Engineered HEs have also been used to target human RAG1 genes, presenting a viable option for treating severe combined immunodeficiency.[17-18] Gene disruption by HEs has been exploited to downregulate the expression of chemokine receptor 5, which carries the HIV infection into cells.[19] HE-based strategies for genetic control of vector-borne diseases have also proven useful for creating transgenic mosquitoes with a high level of sterility, thus confirming the suitability of this approach for controlling diseases like malaria.[20-22] The vast applications of HEs warrant a deeper understanding of their catalytic function.

Like most endonucleases, HEs use metals to facilitate catalysis.[3, 23] Although the majority of nucleases have been proposed to use two metals to cleave phosphodiester

bonds[24-25] and therefore the two-metal mediated reaction has been widely studied,[26-34] evidence has grown that some endonucleases use only one metal for the chemical step.[35-45] Among members of the HE family, I-*Ppo*I (intron-encoded endonuclease from *Physarum polycephalum*) is an example of an enzyme that may use only a single metal for catalysis.[23, 39] Indeed, kinetic studies have demonstrated that Mg^{2+} is essential for I-*Ppo*I catalytic activity[46] and X-ray crystallographic data reveal a single Mg^{2+} bound in the I-*Ppo*I active site (Figure 4.1a–b).[38-39] Nevertheless, despite being one of the most well-characterized members of HE family through experimental biochemical and structural techniques,[20-21, 39, 46-60] no consensus has been reached regarding the I-*Ppo*I catalytic mechanism.

I-*Ppo*I recognizes and binds to the DNA major groove and cleaves across the minor groove to generate 3′-overhangs that are 4-nucleotides long.[3, 61] In the experimentally-proposed I-*Ppo*I mechanism of action, the single metal directly coordinates to the O3′-leaving group and non-bridging oxygen of the scissile phosphate, and thereby promotes leaving group departure through simultaneous substrate stabilization and leaving group protonation by a metal-activated water (Figure 4.1c).[39, 57] Because of the position of H98 with respect to the substrate in the product complex (PC) for wild type I-*Ppo*I (Figure 4.1b)[38] and the complete loss of enzyme activity upon H98A mutation,[57] H98 was proposed to be the general base that activates the water nucleophile to initiate the reaction (Figure 4.1c).[39] The N119A mutation also kills catalytic activity as N119 is the only amino acid coordinated to Mg^{2+} (Figure 4.1a–b). Furthermore, N119 is within hydrogen-bonding distance of the O5′-bridging oxygen of the scissile phosphate in the reactant complex (RC) analogue of the H98A I-*Ppo*I mutant (Figures 4.1a and C.1a, Appendix C), and therefore has been proposed to stabilize the scissile phosphate through hydrogen bonding.[57] Although

the R61A mutation leads to only a slight reduction in I-*Ppo*I activity, R61 has been suggested to stabilize the substrate[39, 57] and may inhibit the reverse reaction by hydrogen bonding with the cleaved phosphate backbone (Figures 4.1b and C.1b, Appendix C).[38-39] Indeed, I-*Ppo*I binds tightly to cleaved DNA, with product release being the proposed rate-determining step.[57] Finally, although the H78A mutation does not significantly impact I-*Ppo*I catalytic activity,[57] H78 is within hydrogen-bonding distance of the water nucleophile in the X-ray crystal structure of the RC analogue of the H98A I-*Ppo*I mutant[39] and close to H98 in the PC for wild type I-*Ppo*I,[39] which suggests that H78 may facilitate the overall P–O bond cleavage process (Figures 4.1 and C.1, Appendix C). Collectively, this data serves as an excellent starting point for modeling, which can provide further insights into the roles of various active site residues, the metal, and metal-ligated water, in the phosphodiester bond cleavage mediated by a representative HE.

Since computational techniques are powerful tools to investigate enzyme-catalyzed reaction mechanisms, characterize high energy intermediates, shed light on the roles of amino acids in the reaction, and compare multiple mechanisms to predict preferred pathways,[25, 62-63] this study uses a combined quantum mechanics (QM) and quantum mechanics-molecular mechanics (QM/MM) approach within the ONIOM formulism to provide atomic-level details of the catalytic pathway used by I-*Ppo*I. Specifically, starting from a high-resolution X-ray crystal structure of the RC analogue of the H98A I-*Ppo*I mutant,[39] several cluster models of varying size are used to map the phosphodiester bond cleavage mechanism using density functional theory (DFT). Subsequently, QM/MM (ONIOM) is used to characterize the phosphodiester bond hydrolysis pathway within the context of the solvated enzyme–DNA complex, which permits comparison of the obtained

Gibbs energy barriers to the experimental values (~86–95 kJ/mol)[56-57] and thereby identification of the preferred catalytic pathway. The fundamental understanding about I-*Ppo*I function obtained from the present work can be used to further explore its potential use in the areas of genome engineering for biotechnological and therapeutic solutions.[8, 11] Furthermore, our improved mechanistic understanding can direct future computational investigations on other HEs, such as I-HmuI[41] and Hpy188I,[44] which share the active site metal-binding arrangement of I-*Ppo*I, but differ in the composition and arrangement of active site amino acids. In addition, this work contributes to the growing body of literature supporting that one metal is enough to facilitate phosphodiester bond cleavage in nucleic acids,[35-37] with comparison to the literature on other single metal-dependent nucleases[40-42, 64-67] highlighting that this chemistry can be achieved in diverse ways.

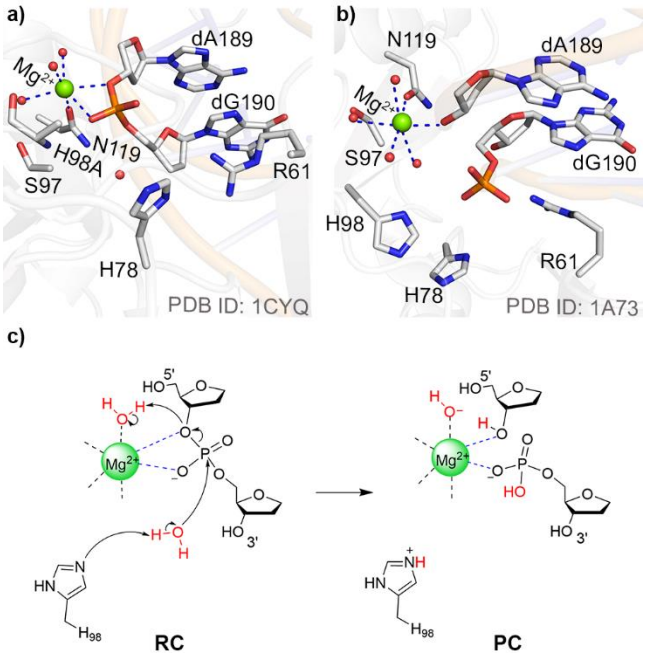

Figure 4.1. I-*Ppo*I active site from an X-ray crystal structure of the a) Mg^{2+}-containing RC for the H98A mutant and b) Mg^{2+}-containing PC for the wild-type enzyme. c) The experimentally proposed phosphodiester bond cleavage pathway for I-*Ppo*I.[39]

4.2. Computational Methodology

4.2.1. QM Calculations

Each QM cluster model was built from a 1.93 Å resolution X-ray crystal structure of the RC analogue of the H98A I-*Ppo*I mutant (PDB ID: 1CYQ; Figure 4.1a), with the alanine mutation at position 98 manually reverted to histidine using PyMOL 2.5.[68] The smallest QM cluster model (Model 1, 92 atoms, Figure 4.2a) includes residues that have been proposed to be essential for catalysis.[46, 57] Specifically, the nucleotides containing the scissile P–O bond (dA189 and dG190), H98 (the proposed general base), and a nucleophilic water were included. Additionally, the model incorporated Mg^{2+} that is directly coordinated to N119 and three water molecules, S97, and a water molecule that may facilitate maintenance of the octahedral metal coordination geometry by simultaneously hydrogen bonding to a Mg^{2+}-coordinated water molecule and a non-bridging phosphate oxygen. An additional water molecule was either hydrogen bonding to another Mg^{2+}-coordinated water molecule (Figure 4.2a) or directly coordinated to Mg^{2+} (Figure C.2a, Appendix C). In the smallest model, the substrate was truncated at O3′ of dG190 and O5′ of dA189, with the truncation points frozen to the crystallographic coordinates and capped with hydrogen atoms. The associated nucleobases (A on the 5′-side and G on the 3′-side of the scissile bond) were also replaced with hydrogen atoms. Each amino acid was truncated at the α carbon, meaning the Cα–C/N(backbone) bonds were cut, the unsaturated C(backbone) replaced with a hydrogen atom, and the location of Cα fixed to the crystallographic coordinates, as previously recommended based on a QM cluster study of a one-metal dependent nuclease.[67] Subsequent cluster models were built from the smallest model. All larger models include an extended substrate that contains the nucleobases in the nucleosides

on either side of the the scissile bond (dA189 and dG190). The 134-atom model also contains H78 (Model 2), the 141-atom model includes R61 (Model 3), and the 163-atom model contains H78, R61, and the 5′-phosphate of dA191 on the 3′-side of dG190 (Model 4, in which the 5′-phosphate is capped and frozen at C5′; Figures 4.2a and C.2a, Appendix C).

Since the inclusion of implicit solvent during optimizations has been shown to minimally impact the energetics of enzyme-catalyzed reactions for cluster models larger than 100 atoms,[67, 69-71] each QM cluster RC was optimized in the gas phase using B3LYP-D3(BJ)/6-31G(d,p) calculations. Subsequently, a guess for a concerted TS was generated by constraining the key reaction parameters involved in the attack of nucleophilic water on the phosphorus reaction center (r(P···O_w) = 1.9–2.2 Å) and cleavage of phosphodiester bond (r(P–O3′) = 1.9–2.2 Å). The distance constraints for the key reaction parameters were chosen based on values reported for a similar phosphodiester bond hydrolysis reaction catalyzed by one-metal dependent APE1.[67] For a stepwise mechanism, the P···O_w distance was constrained (r(P···O_w) = 1.9–2.2 Å) in the first calculation phase to obtain a guess for TS1. After optimizing the corresponding intermediate complex (IC), the P–O3′ distance was then constrained (r(P–O3′) = 1.9–2.2 Å) to obtain a guess for TS2. After fully optimizing each TS, intrinsic reaction coordinate (IRC) calculations[72] were performed to obtain the associated RC, IC, and/or PC, which were then fully optimized. For the one stepwise pathway characterized in the present work, the IRCs from TS1 and TS2 gave rise to structurally and energetically similar ICs (Figure C.3, Appendix C), verifying the pathway is connected. The nature of the optimized stationary points were confirmed using frequency calculations at same level of theory (i.e., minima have all positive and TSs have

one negative frequency), which also afforded the thermal corrections to the Gibbs energy. Furthermore, to account for the surrounding enzymatic environment, IEF-PCM-M06-2X/6-311+G(2df,p) single-point energy calculations were performed on the QM cluster models using a dielectric constant of ε=4 that provides an accurate representation of the enzyme surroundings.[73] Minimal differences in the reported relative energies occur when single-point energy calculations were instead performed using ωB97M-V, which was recently suggested to be among the most reliable functionals for describing enzymatic reactions[74] (Figure C.4 and Table C.1, Appendix C).

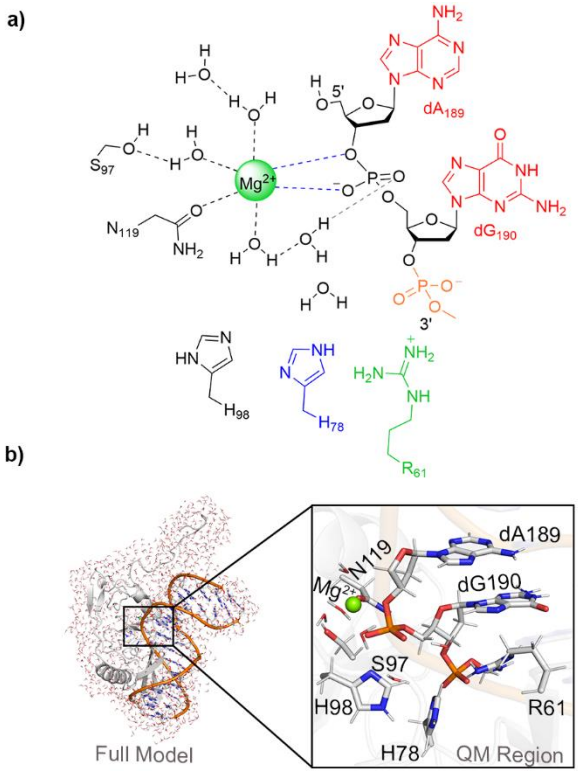

Figure 4.2. a) Schematic of QM cluster models considered in this work: Model 1 (black), Model 2 (Model 1 + H78 (blue) + expanded substrate (red)), Model 3 (Model 1 + R61 (green) + expanded substrate), Model 4 (Model 1 + H78 + R61 + 5′ phosphate moiety of dA191 on the 3′-side of dG190 (orange)). b) The enzyme–DNA QM/MM (ONIOM) model (left) and the corresponding QM region (black box, right).

4.2.2. QM/MM Calculations

QM/MM calculations were performed within the ONIOM formulism due to previous successes of this methodology in providing accurate structural and energetic descriptions for other enzymatic reactions.[63, 65-66, 75-77] The QM/MM model was built using the X-ray crystal structure of the RC analogue of the H98A I-*Ppo*I mutant (PDB ID: 1CYQ, Figure 4.1a) as calculations initiated from an X-ray crystal structure have been successfully used to elucidate the reaction mechanisms for several enzymes.[78-82] The alanine mutation at position 98 in the crystal structure was manually reverted to H98 using PyMOL based on the residue orientation in the crystal structure of the wild type I-*Ppo*I PC (1.80 Å resolution; PDB ID: 1A73, Figure 4.1b). The enzyme–DNA complex was solvated in a 10 Å TIP3P octahedral water box and PROPKA was used to assign the protonation states of titratable amino acids.[83] All water molecules with any atom further than 6 Å from the enzyme–substrate complex were removed. The solvated enzyme–DNA complex was minimized using AMBER18.[84-85]

The QM region for the QM/MM model includes Mg^{2+} along with the directly coordinated three water molecules and N119. Furthermore, the QM layer contains S97, H98, H78, R61, a nucleophilic water, an additional water molecule simultaneously hydrogen bonded to a Mg^{2+}-coordinated water and a non-bridging phosphate oxygen, the nucleotides containing the scissile P–O bond (dA189 and dG190), and the phosphate from dA191 on the 3′-side of dG190. Finally, depending on the metal–substrate binding configuration, an additional water molecule was incorporated into the QM region that either hydrogen bonded to a Mg^{2+}-coordinated water (Figure 4.2b) or directly coordinates to Mg^{2+} (Figure C.2b, Appendix C). The QM/MM model contains 137 atoms in the QM region,

which has a charge of +1. The MM region includes the remaining I-*Ppo*I–DNA complex and the surrounding water molecules. The overall charge of the entire model is –23. The QM/MM boundary was placed between Cα and Cβ of each amino acid, and C4′ and C5′ of dA189 and dA191 of the substrate. Mechanical embedding (ME) was used to describe the boundary in all QM/MM models, while electrostatic embedding (EE) was subsequently employed for the preferred mechanism.

Each RC was fully optimized using B3LYP-D3(BJ)/6-31G(d,p) for the QM region[67] and the AMBER force field (OL15 for DNA[85] and ff14SB for the enzyme[84]) for the MM region. Next, scans of key reaction distances were performed to obtain initial guesses for TSs, where each parameter was increased/decreased in 0.10–0.15 Å increments. Specifically, a TS guess was obtained by successively decreasing the P···O_w distance. The bond was then frozen to the distance corresponding to a maximum on the potential energy surface, and the P–O3′ bond distance was subsequently increased. Full TS optimizations were then performed and IRC calculations run to obtain the associated RC and PC. Frequency calculations were conducted at the same level of theory to confirm the nature of the stationary points (i.e., minima have all positive and TSs have one imaginary frequency) and obtain thermal corrections to the Gibbs energies. Finally, single-point calculations at the ONIOM(M06-2X/6-311+G(2df,p):AMBER) level of theory were used to calculate the relative Gibbs energies.

Each characterized catalytic mechanism was carefully compared to experimental structural,[38-39] mutational,[46, 57] and kinetic data.[56-57] Specifically, the geometries of each optimized RC and PC were assessed in relation to the crystal structures of the RC analogue of the H98A I-*Ppo*I mutant (PDB ID: 1CYQ)[39] and PC for wild type I-*Ppo*I (PDB ID:

1A73), respectively.[38] The observed roles of active site amino acids were correlated with experimental mutational data, including the complete loss of I-*Ppo*I catalytic activity for the H98A and N119A mutants[57] and slight reduction in catalytic activity for the H78A[46, 57] and R61A mutants.[57] The predicted QM/MM activation barriers were compared to the barrier for the overall process (~86–95 kJ/mol) calculated using the experimental k_{cat} (0.046–1 min^{-1}) measured at 37 °C and pH 7.5, with 10–50 nM of a DNA substrate (42-base pair oligonucleotide) and 5–100 pM of I-*Ppo*I.[56-57]

The Gaussian 16 program (revision B.01) was used to perform all QM cluster and QM/MM calculations,[86] with the exception of ωB97M-V QM cluster calculations which were performed using ORCA (version 5.0.4).[87]

4.3. Results and Discussion

4.3.1. The Smallest QM Cluster and Large-scale QM/MM Models Suggest That the Experimentally-proposed Pathway in Which the Metal Aids Leaving Group Departure Through Direct Coordination and Simultaneous Protonation by a Metal-activated Water is not Feasible.

As discussed in the computational details, the smallest QM cluster model of I-*Ppo*I includes all essential residues for catalysis (denoted Model 1). Specifically, Mg^{2+} and three water molecules directly coordinated to the metal in the crystal structure of the RC analogue of the H98A I-*Ppo*I mutant were incorporated since it is known that a single metal is essential for I-*Ppo*I catalytic activity (Figure 4.1a).[38-39, 46] Furthermore, Mg^{2+}-coordinating N119 and the proposed general base (H98) were added as mutation of either residue has

been shown to kill enzymatic activity.[57] Finally, S97 and two additional water molecules were included in Model 1 as these residues may facilitate maintenance of the octahedral metal coordination geometry by hydrogen bonding to Mg^{2+}-coordinated water molecules (Figure 4.2a). This leads to a model containing 92 atoms, with a net charge of +1.

The X-ray crystal structure of the RC analogue of the H98A I-*Ppo*I mutant (Figure 4.1a) was the basis for the proposal that Mg^{2+} assists leaving group departure through direct coordination to O3′ coupled with protonation of the leaving group by a metal-activated water during phosphodiester bond cleavage (Figure 4.1c).[39] Indeed, a water molecule falls within hydrogen-bonding distance of O3′ (3.4 Å, Figure C.1a, Appendix C). However, the corresponding TS could not be characterized using QM cluster Model 1 as the hydrogen-bond distance and angle between the Mg^{2+}-ligated water and the substrate is unfavorable in the optimized RC (Figure C.5a, Appendix C). To ensure that the inability to characterize the experimentally-proposed mechanism was not due to insufficient model size, attempts were made to map the same reaction mechanism using our large-scale QM/MM model. As found for Model 1, there remains a poor relative alignment of the active site water and substrate in the QM/MM optimized RC (Figure C.5b, Appendix C), which prevents successful characterization of the corresponding TS. Together, QM cluster Model 1 and the QM/MM model highlight that simultaneous direct metal coordination to O3′ and leaving group protonation by a metal activated water is not possible during the P–O bond cleavage step within the confines of the I-*Ppo*I active site. Nevertheless, we acknowledge that protonation of the leaving group by metal-activated or bulk water likely occurs after the chemical step as metal migration and active site rearrangement commonly occur in

conjunction with product release for both one- and two-metal dependent nucleases.[24, 29, 35, 37, 88-91]

4.3.2. The Smallest QM Cluster Model Indicates That Leaving Group Protonation by a Metal-ligated Water During P–O bond Cleavage is Infeasible.

When the experimentally-proposed pathway was investigated using QM cluster Model 1 and QM/MM models, a water molecule in the second metal coordination sphere migrated between the metal and O3′ during TS optimizations, while direct metal coordination to the non-bridging oxygen was maintained. This suggests that a water may be aligned to facilitate leaving group protonation during P–O bond cleavage as put forward in the experimentally-proposed pathway[39] if indirect (water-mediated) rather than direct metal coordination to O3′ of the substrate is present. Indeed, a similar indirect metal–substrate binding architecture has been reported for one-metal dependent *Eco*RI (Figure C.6a, Appendix C)[92] and two-metal dependent BamHI (Figure C.6b, Appendix C).[26, 93] Therefore, the reaction was explored using QM Model 1 from a I-*Ppo*I RC that positions a water molecule between Mg^{2+} and O3′ of the scissile phosphate (Figure C.2a, Appendix C). In the optimized RC, H98 hydrogen bonds to a potential water nucleophile, while Mg^{2+} aids substrate stabilization through direct coordination to a non-bridging oxygen of the scissile phosphate in the optimized RC (Figure 4.3). Further substrate stabilization is supplied by a Mg^{2+}-coordinated water through a water chain to the other non-bridging phosphate oxygen (Figure C.7a, Appendix C). The water nucleophile is positioned for attack ($r(O_w \cdots P)$ = 3.549 Å, Figure 4.3b) and the reaction proceeds through a concerted TS in which proton transfer from the nucleophilic water to H98 is coupled with attack at the

phosphorus reaction center ($r(O_w \cdots P)$ = 1.845 Å, Figure 4.3). Simultaneously, the P–O3′ bond lengthens (2.126 Å) and the water between Mg^{2+} and O3′ partially transfers a proton to the leaving group. Together, these structural features lead to a barrier of 182.3 kJ/mol and a highly endergonic PC (by 115.0 kJ/mol, Figure 4.4 and Table C.2, Appendix C). Although direct comparison of QM cluster barriers to experimental data is not recommended due to the missing enzymatic environment,[67] our predicted phosphodiester bond cleavage barrier involving indirect metal coordination to the leaving group is much larger than other P–O bond cleavage barriers evaluated using QM cluster models (~50–95 kJ/mol).[65, 94-96] Therefore, an alternate mechanism must be considered.

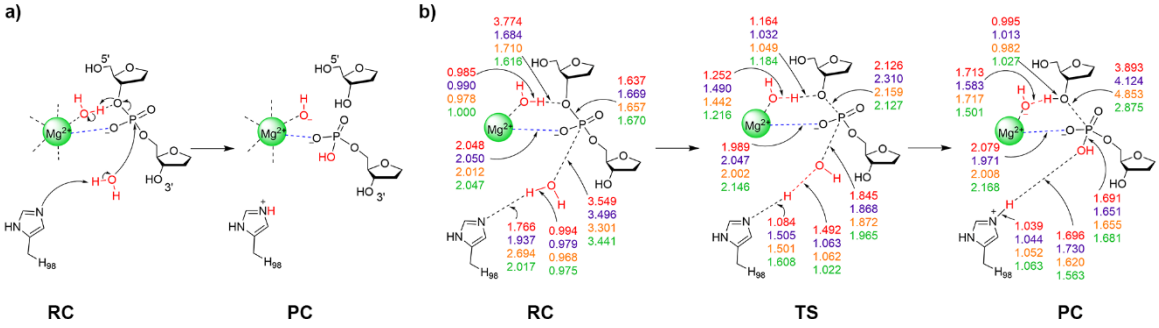

Figure 4.3. a) The proposed reaction pathway and b) key reaction parameters (Å) for each stationary point of the I-*Ppo*I catalyzed phosphodiester bond cleavage involving indirect Mg^{2+} coordination to the leaving group characterized in the present work using QM cluster Model 1 (red), Model 3 (purple), Model 4 (orange), or QM/MM (green). See Figures C.7–C.10 (Appendix C) for additional structural parameters for QM cluster Models 1–4.

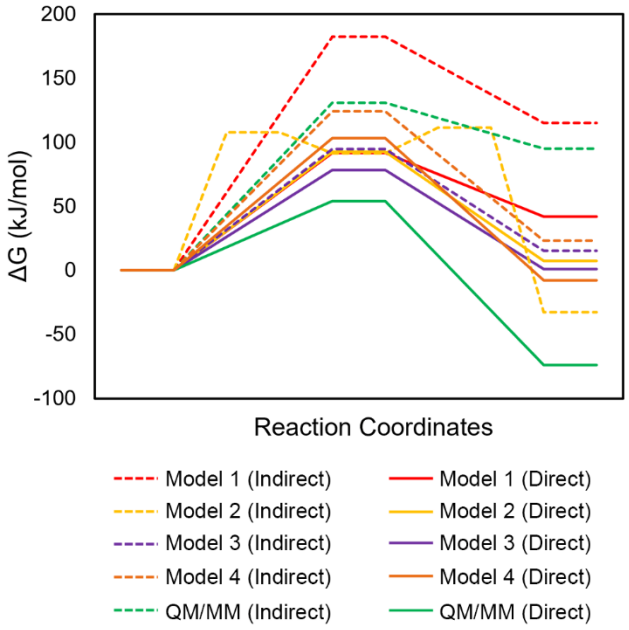

Figure 4.4. Relative Gibbs energies (kJ/mol) for the I-*Ppo*I catalyzed phosphodiester bond cleavage characterized using QM cluster (Models 1 – 4) or QM/MM models. Dashed lines refer to the pathways involving indirect (water-mediated) Mg^{2+} coordination to the leaving group, while solid lines refer to the pathways involving direct Mg^{2+} coordination to the leaving group.

4.3.3. The Smallest QM Model Suggests That Direct Metal Coordination to the Leaving Group is Necessary for P–O Bond Cleavage.

In the crystal structure of the RC analogue of the H98A I-*Ppo*I mutant, Mg^{2+} adopts bidentate coordination to the substrate through the O3′-leaving group and a non-bridging oxygen of the scissile phosphate (Figure 4.1a). A similar coordination has been observed for other one-metal (T4 endonuclease VII (Figure C.6c, Appendix C)[42] and I-HmuI (Figure C.6d, Appendix C)[41]) and two-metal (*Aquifex aeolicus* RNase III (AaRNase III, Figure C.6e, Appendix C)[97] dependent endonucleases. In these cases, direct coordination to the metal has been proposed to aid leaving group departure without the necessity of simultaneous protonation during P–O bond cleavage.[41-42, 97] An I-*Ppo*I RC based on Model

1 can be optimized which preserves the direct Mg^{2+} coordination to the O3′-bridging oxygen as seen in the crystal structure of the RC analogue of the H98A I-*Ppo*I mutant (Figures 4.1 and 4.5a), while N119 interacts with the other bridging oxygen of the scissile phosphate (O5′ of dG190, Figure C.7b, Appendix C). In this active site configuration, the water nucleophile sits close to the phosphorus reaction center ($r(O_w \cdots P)$ is 0.6 Å shorter than in the corresponding RC involving indirect metal coordination to leaving group, Figures 4.3b and 4.5b). The RC evolves into a concerted TS in which the P–O bond cleavage happens earlier than in the first mechanism (r(P···O3′) is 0.3 Å shorter, Figures 4.3 and 4.5). In fact, these structural changes coupled with a strong Lewis acid stabilizing the leaving group (direct Mg^{2+} coordination) reduces the barrier by 90.7 kJ/mol compared to the pathway involving indirect (water-mediated) Mg^{2+}–O3′ coordination (Figure 4.4 and Table C.3, Appendix C). As a result, the new pathway leads to a barrier ($\Delta^{\ddagger}G$ = 91.6 kJ/mol) that falls within the range of those calculated for other metalloenzyme-facilitated phosphodiester bond cleavage reactions using QM cluster models (~50–95 kJ/mol).[65, 94-96]

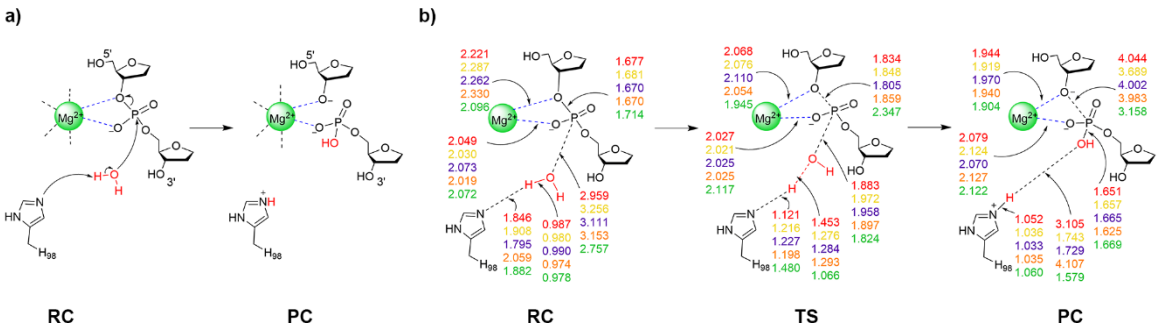

Figure 4.5. a) The proposed reaction pathway and b) key reaction parameters (Å) for each stationary point of the I-*Ppo*I catalyzed phosphodiester bond cleavage involving direct Mg^{2+} coordination to the leaving group characterized in the present work using QM cluster Model 1 (red), Model 2 (yellow), Model 3 (purple), Model 4 (orange), or QM/MM (green). See Figures C.7–C.10 (Appendix C) for additional structural parameters for QM cluster Models 1–4.

Although the experimental proposal that direct coordination to the leaving group and O3′ protonation occur at the same time is not feasible, pathways for the chemical step involving either element can be characterized. The smallest QM-based I-*Ppo*I model reveals that the mechanism involving direct metal–O3′ coordination as seen in the X-ray crystal structure of the RC analogue of the H98A I-*Ppo*I mutant[39] is preferred over the water-mediated coordination necessary for leaving group protonation by a metal-activated water. Nevertheless, H98 activates the water nucleophile and N119 stabilizes the substrate as proposed based on experimental data.[38-39, 57] Despite suggestions from structural (Figure C.1b, Appendix C) and mutational studies that H98 helps stabilize the cleaved product,[57] H98 interacts with S97 rather than the newly formed P–O bond in the PC (Figure C.7b, Appendix C). This is likely an artifact caused by the absence of the broader surrounding enzymatic environment in Model 1, suggesting the need to include additional residue(s) that may help align H98. To ensure the conclusion reached regarding the relative importance of direct metal–O3′ coordination versus leaving group protonation is not the result of model size, both mechanisms successfully characterized for Model 1 will be considered in expanded models moving forward.

4.3.4. H78 Primarily Plays a Structural Role in I-*Ppo*I Catalysis, Positioning the Water Nucleophile and the General Base in QM Model 2.

X-ray crystallographic data reveals that H78 is within hydrogen-bonding distance of the nucleophilic water in the RC analogue of the H98A I-*Ppo*I mutant[39] and H98 in the PC for wild type I-*Ppo*I (Figure C.1, Appendix C),[38] while mutational data shows a slight reduction in I-*Ppo*I catalytic activity upon H78 mutation to alanine.[57] This suggests that

H78 may be important for catalysis and this residue was added to Model 1 (Model 2, 134 atoms). Upon consideration of indirect (water-mediated) Mg^{2+}–O3′ coordination that permits leaving group protonation using Model 2, H78 hydrogen bonds to H98 throughout the reaction, which situates the general base further from the nucleophile in the RC compared to Model 1 ($r(N_{H98}\cdots H_w)$ is 0.8 Å longer, Figures 4.3b, 4.6 and C.8a, Appendix C). Nevertheless, a water molecule is well positioned between Mg^{2+} and the substrate to protonate the leaving group in Model 2 (Figure 4.6). The corresponding phosphodiester bond cleavage progresses through a stepwise mechanism, where TS1 involves proton release from the nucleophilic water to H98 and concomitant attack on the electrophilic phosphorus. TS1 collapses into a stable phosphorane intermediate (16.6 kJ/mol below TS1, Figure 4.4 and Table C.2, Appendix C) from which the P–O bond dissociates in the second step. The rate-determining barrier is 70.8 kJ/mol lower for Model 2 than Model 1 (Figures 4.4, 4.6 and Table C.2, Appendix C). Nevertheless, the barrier for phosphodiester bond hydrolysis involving indirect metal coordination to the leaving group remains high ($\Delta^{\ddagger}G$ = 111.5 kJ/mol, Figure 4.4 and Table C.2, Appendix C).

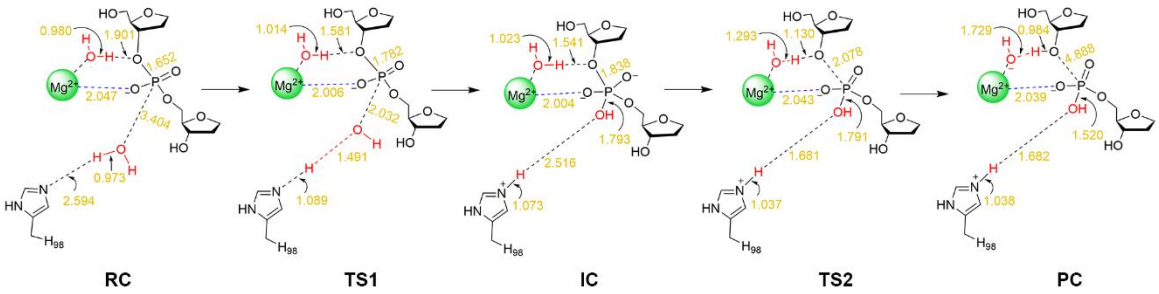

Figure 4.6. Key reaction parameters (Å) for each stationary point of the I-*Ppo*I catalyzed phosphodiester bond cleavage involving indirect Mg^{2+} coordination to the leaving group characterized in the present work using QM cluster Model 2. See Figure C.8a (Appendix C) for additional structural parameters for QM cluster Model 2.

When direct Mg^{2+} coordination to the leaving group is considered with Model 2, H78 hydrogen bonds to the water nucleophile in the RC, which is consistent with experimental structural data (Figures C.1a and C.8b, Appendix C).[39] This results in a 0.3 Å longer nucleophilic distance in the RC compared to Model 1 (Figures 4.3b and 4.5b). In the concerted TS, H78 reorients to hydrogen bond to the general base (H98) as water attacks the phosphorus reaction center (Figure C.8b, Appendix C). Although the energy barriers for Models 1 and 2 are nearly equal (within 1 kJ/mol, Figure 4.4 and Table C.3, Appendix C), the PC for Model 2 is significantly more stable (by 34.4 kJ/mol) due to a hydrogen bond between (protonated) H98 and the newly formed P–O bond (Figure 4.6), which agrees with the experimental proposal that H98 stabilizes the cleaved product (Figure C.1b, Appendix C).[38, 57]

Overall, similar to Model 1, the pathway involving direct metal coordination to the leaving group is preferred over the pathway in which the leaving group is protonated by an intervening water molecule for Model 2. Furthermore, despite a similar activation barrier as Model 1, Model 2 highlights that H78 can position the water nucleophile and the general base (H98), which is consistent with structural[38-39] and mutational data.[57] Nevertheless, H78 interacts with a substrate nucleobase in the 3′ direction with respect to the scissile P–O bond (O6 of dG190) throughout the reaction and the cleaved phosphate in the PC for Model 2 (Figure C.8b, Appendix C), which is not consistent with crystallographic data ($r(N_{H78}\cdots O_{dG190})$ ~11 Å and $r(N_{H78}\cdots O_{non\text{-}bridging})$ ~5 Å, PDB ID: 1CYQ and 1A73)[38-39] due to the presence of an active site residue (R61) between H78 and the substrate (Figure C.1, Appendix C). This suggests further model refinement is required.

4.3.5. R61 Hydrogen Bonds With the 3′-nucleobase With Respect to the Scissile P–O Bond in QM Model 3.

In the crystal structures of the RC analogue of the H98A I-*Ppo*I mutant (PDB ID: 1CYQ)[39] and the PC for wild type I-*Ppo*I (PDB ID: 1A73),[38] R61 hydrogen bonds to the substrate and thereby has been suggested to stabilize the TS and the cleaved phosphate group in the PC (Figure C.1b, Appendix C), which may inhibit the reverse reaction.[38] This proposal is supported by the R61A mutant having slightly reduced I-*Ppo*I catalytic activity.[57] Therefore, prior to expanding Model 2, R61 was added to Model 1 (Model 3, 141 atoms, Figure 4.1). In Model 3, R61 hydrogen bonds to the nucleobase of dG190 regardless of whether indirect or direct Mg^{2+} coordination to the leaving group is considered (Figure C.9, Appendix C). When the indirect (water-mediated) Mg^{2+} coordination that permits leaving group protonation is considered, N119 hydrogen bonds to a metal-ligated water (Figure C.9a, Appendix C), which in turn provides significant stabilization to O3′ in the RC compared to Model 1 ($r(O_w \cdots H_w)$ decreases by 2.1 Å, Figure 4.3b). In the concerted TS, a proton has not yet transferred from the water nucleophile to H98 (Figure 4.3b). These factors substantially decrease the activation barrier compared to Model 1 (by 87.5 kJ/mol, Figure 4.4 and Table C.2, Appendix C). Despite the reasonable activation barrier for phosphodiester bond cleavage (94.8 kJ/mol), there is a 13.2 kJ/mol decrease in the Gibbs energy barrier when direct bidentate metal–substrate coordination is considered (Tables C.2–C.3, Appendix C). Although most of the reaction parameters observed for Model 1 are retained in the RC and TS for Model 3 involving direct metal–O3′ coordination (Figure 4.5b), protonated H98 hydrogen bonds to the newly formed P–O bond in the Model 3 PC as implied by the crystal structure of the wild type I-*Ppo*I PC (Figure C.1 Appendix C),[38]

which yields a 40.7 kJ/mol more stable product (Figure 4.4 and Table C.2, Appendix C). Although the pathway involving leaving group stabilization through direct Mg^{2+} coordination remains the preferred mechanism for Model 3 (Figure 4.4 and Tables C.2–C.3, Appendix C), an interaction between R61 and the 3′-nucleobase (N7 and O6 of dG190) with respect to the scissile P–O bond arises (Figure C.9b, Appendix C) that is absent in the I-*Ppo*I crystal structures (Figures 4.1 and C.1, Appendix C),[38-39] suggesting that the model needs to be further expanded.

4.3.6. Inclusion of Both H78 and R61 as well as the Phosphate on the 3′-side of the Scissile P–O Bond Affords Correct Positioning of the General Base and Stabilization of the Substrate in QM Model 4.

Models 2 and 3 indicate that H78 and R61 each play important roles in the phosphodiester bond cleavage by positioning the general base, water nucleophile, and/or substrate, as well as stabilizing charges in the active site. Furthermore, the crystal structure of the RC analogue of the H98A I-*Ppo*I mutant reveals that R61 is within hydrogen-bonding distance of the phosphate on the 3′-side of the scissile P–O bond (dA191, Figure C.1a, Appendix C).[39] Therefore, H78, R61, and the additional phosphate moiety were simultaneously added to Model 1 (Model 4, 163 atoms) and the reaction pathways were remapped. Regardless of the metal coordination geometry considered using Model 4, R61 hydrogen bonds to the 3′-phosphate and H78 positions the general base (H98) through hydrogen bonding as implied by experimental structural data (Figures 4.1, C.1 and C.10a, Appendix C).[39] Although H98 is situated further from the water nucleophile than in Model 1 with indirect metal–leaving group coordination ($r(N_{H98}\cdots H_w)$ is 0.9 Å longer, Figures

4.3b), the water nucleophile is close to the phosphorus reaction center and a metal-ligated water is well aligned for O3′ protonation. These structural factors collectively reduce the activation barrier by 58.1 kJ/mol compared to Model 1 (Figure 4.4 and Table C.2, Appendix C). Nevertheless, the P–O bond cleavage barrier for the pathway involving water-mediated metal coordination to leaving group remains high ($\Delta^{\ddagger}G$ = 124.3 kJ/mol). On the other hand, when Model 4 with direct metal–O3′ coordination is considered, the majority of Model 1 structural parameters are maintained (Figures C.7b and C.10b), with the exception of a slightly increased nucleophilic attack distance (r(P···O_w) is 0.2 Å longer, Figure 4.5b), which raises the activation barrier for P–O bond cleavage by 11.6 kJ/mol (Figure 4.4 and Table C.3, Appendix C) and stabilizes the PC (by 49.6 kJ/mol).

Regardless of the energetic differences between Models 1 and 4, the phosphodiester bond cleavage pathway involving direct Mg^{2+}–O3′ coordination remains preferred over indirect (water-mediated) Mg^{2+}–O3′ coordination for Model 4. Furthermore, Model 4 highlights that H78, R61, and the 3′-phosphate with respect to the scissile phosphate are necessary to properly position the general base (H98) and stabilize the substrate, as proposed based on experimental structural and mutational data.[38-39, 57] Furthermore, inclusion of these residues overcomes some of the artifacts of the previous QM models including eliminating interactions between H78 or R61 and the substrate nucleobase (dG190, Figure C.10b, Appendix C). However, due to the absence of the surrounding environment, interactions between H78 and R61, and between H78 and the 3′-phosphate exist in Model 4, despite these residues being separated by > 5 Å in the X-ray crystal structures (r(N_{H78}···N_{R61}) and r(N_{H78}···$OP1_{dG190}$), PDB ID: 1CYQ and 1A73).[38-39] Furthermore, interactions between (protonated) H98 and the newly formed P–O bond are

lost in the PC (Figure 4.5b). This underscores the need to characterize the phosphodiester bond cleavage pathway in the presence of the broader enzymatic environment. Indeed, QM cluster models containing multiple shells of amino acids (300–400 atoms) have been shown to be necessary for accurate descriptions of some enzymatic reactions.[82, 98-101] Nonetheless, our QM cluster calculations provide valuable insights into the residues that must be included in the QM region of QM/MM models of I-*Ppo*I.

4.3.7. QM/MM Calculations Further Highlight That the Preferred Mechanism in the Presence of the Surrounding Enzymatic Environment Involves Direct Metal Coordination to the Leaving Group.

As discussed in the computational details, the QM region of the QM/MM model of I-*Ppo*I contains all residues included in QM cluster Model 4. The QM/MM RC involving indirect (water-mediated) Mg^{2+}–O3′ coordination maintains all interactions observed in the I-*Ppo*I active site based on the crystal structures (Figures 4.1, 4.7a and C.1a, Appendix C).[38-39] Specifically, unlike any QM cluster model, H78 simultaneously positions the water nucleophile and orients the general base (H98), N119 provides substrate stabilization by hydrogen bonding to the O5′-bridging oxygen of dG190, and R61 hydrogen bonds with the 3′-phosphate with respect to the scissile phosphate (Figures 4.7a and C.7a–C.10a, Appendix C). As a result, the water nucleophile is in close proximity to the general base (H98) and the electrophilic phosphorus atom (Figure 4.3b), and the QM/MM RC goes through a concerted pathway to cleave the phosphodiester bond. Since most reaction parameters in the TS of Model 4 are preserved (Figure 4.3), the activation barriers are similar (within 6.6 kJ/mol, Figure 4.4 and Table C.2, Appendix C). Therefore, the QM/MM activation barrier

for the pathway involving O3′ protonation by a metal-ligated water (130.9 kJ/mol) is significantly higher than the experimental value for I-*Ppo*I (~86–95 kJ/mol) [56-57] and typical experimental barriers for metallonucleases catalyzed reactions (~58–96 kJ/mol),[45, 102-105] suggesting this pathway is energetically infeasible. This supports the conclusions from all QM cluster models that the water-mediated metal–substrate coordination necessary for leaving group protonation cannot effectively catalyze the P–O bond cleavage reaction.

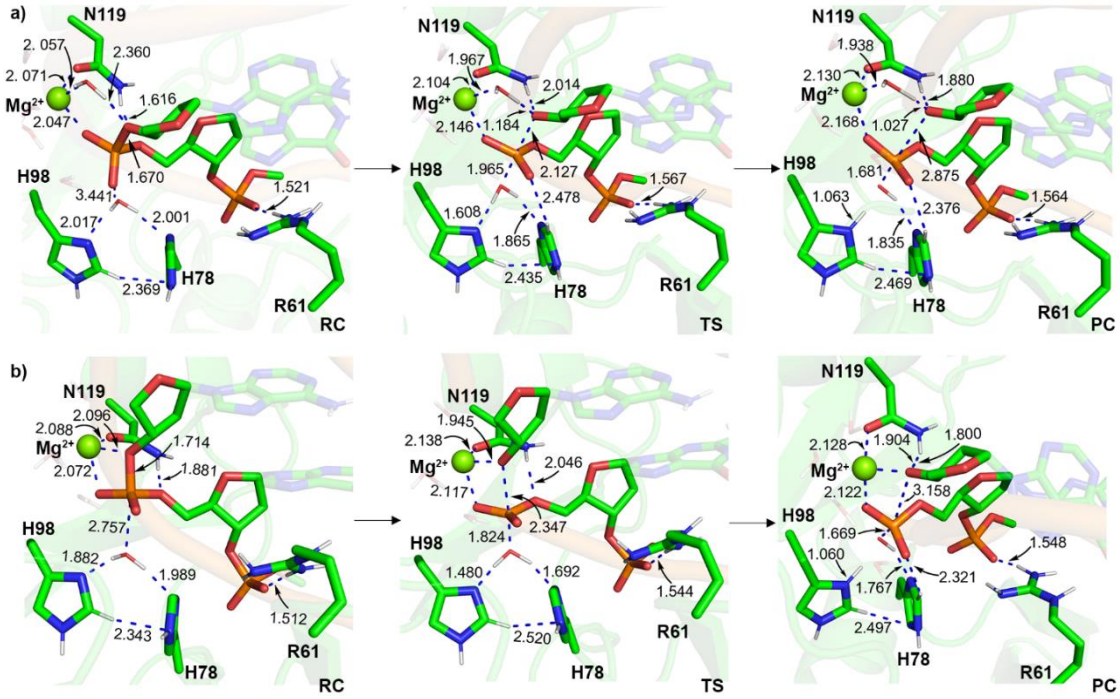

Figure 4.7. Key distances, reaction parameters, and metal coordination distances (Å) in the I-*Ppo*I active site for the phosphodiester bond cleavage pathway characterized using QM/MM with a) indirect or b) direct Mg^{2+} coordination to the leaving group.

When the pathway involving direct coordination of the metal to the leaving group is considered using QM/MM, H78 interacts with both the water nucleophile and the general base (H98), while R61 hydrogen bonds to the phosphate 3′ to the scissile P–O bond in the RC (Figure 4.7b). Despite a smaller nucleophilic attack distance ($r(O_w \cdots P)$ = 2.757 Å) and

the metal being closer to the leaving group (2.096 Å, Figure 4.5b) in the QM/MM RC than any QM cluster RC, the proton has not transferred from the water nucleophile to H98 and the P–O bond cleavage is more advanced (2.347 Å) in the QM/MM TS compared to any QM model. Nevertheless, the barrier for the mechanism involving direct metal–leaving group coordination (54.1 kJ/mol) is lower than that for any other model (by 24.3–38.5 kJ/mol, Figure 4.4 and Table C.3, Appendix C). Furthermore, the direct bidentate metal–substrate binding architecture observed in the crystal structure of the RC analogue of the H98A I-*Ppo*I mutant (Figure 4.1a) results in a 76.8 kJ/mol lower barrier than water-mediated metal coordination to the leaving group (Figure 4.4 and Tables C.2–C.3, Appendix C). Direct metal–O3′ coordination is consistent with X-ray crystallographic data for I-*Ppo*I[38-39] as well as other one-metal dependent endonucleases (e.g., T4 EndoVII (Figure C.6c, Appendix C),[42] I-HmuI (Figure C.6d, Appendix C),[41], and Hpy188I (Figure C.6f, Appendix C)[44]). Furthermore, the alignments of key active site amino acids in the QM/MM optimized RC and PC are consistent with the crystal structures of the RC analogue of the H98A I-*Ppo*I mutant and the PC for wild-type I-*Ppo*I (Figure C.11, Appendix C). We also note that minimal changes occur in the reaction parameters (Figure C.12a, Appendix C) and orientations of active site residues (Figure C.12b, Appendix C) upon reoptimization of the preferred pathway using EE. Although the activation barrier slightly increases (by 23.3 kJ/mol, Table C.3, Appendix C) and the PC is slightly more exergonic (by 28 kJ/mol) for the EE characterized pathway, the computed Gibbs activation barrier for the mechanism involving direct Mg^{2+} coordination to the leaving group ($\Delta^{\ddagger}G$ = 77.4 kJ/mol) remains below the experimental value (~86–94 kJ/mol) for the overall process,[56-57] with product release being the rate-determining step.[57] Thus, our QM/MM models collaborate the conclusion

from QM cluster calculations that direct metal–leaving group coordination is critical for the chemical step facilitated by I-*Ppo*I.

The roles of the amino acid residues uncovered from the preferred QM/MM mechanism are also in line with mutational data.[46, 57] Specifically, H98 initiates the reaction by activating the water nucleophile, which agrees with the experimentally observed complete loss of catalytic activity upon H98A mutation.[57] Our *in silico* H98A mutational results further support this proposal. Specifically, although the reaction parameters and orientation of active site residues are preserved upon H98A mutation (Figures 4.5, 4.7, and C.13, Appendix C), the water nucleophile repositions to be activated by H78 in the QM/MM RC, which substantially increases the activation barrier compared to the wild-type mechanism (by 79.8 kJ/mol, Figure C.13 and Table C.3, Appendix C). In our proposed wild-type I-*Ppo*I mechanism, N119 stabilizes the substrate through hydrogen bonding, while holding the metal in place to facilitate the reaction, which correlates with experimental mutational data showing complete enzyme inactivation upon N119 mutation.[57] Furthermore, H78 and R61 play structural roles including positioning and stabilizing the nucleophilic water, the general base (H98), and the substrate for efficient catalysis, which correlates with reported slight reductions in enzymatic activity upon mutating these residues.[46, 57] Overall, our proposed phosphodiester bond hydrolysis pathway involving direct Mg^{2+} coordination to the leaving group is fully consistent with experimental kinetic,[56-57] structural,[38-39] and mutational data.[46, 57]

4.3.8. Our Proposed Pathway for One-Metal Mediated I-*Ppo*I Catalysis Draws Similarities, Yet Differences from Those Facilitated by Other One- and Two-metal Dependent Nucleases.

For enzyme-catalyzed phosphodiester bond hydrolysis reactions to proceed, a water nucleophile must be activated, the leaving group must depart, and the charge on the substrate must be stabilized.[90, 106] A metal-ligated water has commonly been proposed to act as the general acid, protonating the leaving group of the substrate for several nucleases (e.g., one-metal dependent *Eco*RI (Figure C.6a, Appendix C),[92] APE1,[65] and Vvn,[40] and two-metal dependent *Eco*RV[28] and BamHI (Figure C.6b, Appendix C)[26]). Despite an equivalent proposal in the literature for I-*Ppo*I based on crystallographic data,[39] our QM cluster and QM/MM calculations show that the I-*Ppo*I pathway involving leaving group protonation by a metal-activated water is infeasible. Nevertheless, leaving group departure has also commonly been proposed to be facilitated though direct coordination of the metal to the substrate for various one-metal dependent nucleases (e.g., T4 EndoVII (Figure C.6c, Appendix C),[42] I-HmuI (Figure C.6d, Appendix C),[41] and Hpy188I (Figure C.6f, Appendix C)[44]). Furthermore, experimental structural data for two-metal mediated AaRNase III (Figure C.6e, Appendix C)[97] and computational studies on CRISPR-Cas9[34] have suggested one metal (Mg_B^{2+}) directly aids leaving group departure. Our calculations reveal the preferred P–O bond cleavage pathway for I-*Ppo*I also involves direct Mg^{2+}–O3′ coordination, with the single metal in the I-*Ppo*I active site being equivalently positioned with respect to substrate as the second metal in two-metal mediated endonucleases (Mg_B^{2+}, Figure C.14, Appendix C).

The charge on a non-bridging oxygen of the scissile phosphate has generally been suggested to be stabilized by both metals (M_A^{2+} and M_B^{2+}) for two-metal mediated nucleases (e.g., BamHI[26] and AaRNase III,[97] Figures C.6b and C.6e, Appendix C), while diverse residues have been proposed to adopt this role for single-metal dependent enzymes in addition to the metal. For example, I-HmuI contains Y39,[41] while Hpy188I contains K73 and R84[44] appropriately positioned with respect to the substrate for charge stabilization (Figure C.6d–e, Appendix C). Despite sharing similar metal-binding architectures with I-HmuI and Hpy188I (Figures 4.1, and C.6d and f, Appendix C), our calculations reveal that I-*Ppo*I uniquely uses a water chain between Mg^{2+} and a non-bridging phosphate oxygen to complement the substrate charge stabilization provided by the metal directly coordinated to the other non-bridging oxygen of the scissile phosphate (Figure 4.2a). In terms of nucleophile activation, the first metal (M_A^{2+}) is generally proposed to facilitate nucleophilic attack in the well-accepted mechanism for two-metal dependent nucleases.[34, 97] In contrast, a histidine residue (H98) acts as the general base for I-*Ppo*I, fulfilling the role of the first metal (M_A^{2+}) in a two-metal mediated reaction. The proposed role of histidine as the general base in the I-*Ppo*I mechanism of action is in agreement with the experimental structural[38-39] and mutational data for I-*Ppo*I,[57] as well as experimental crystallographic studies on other single-metal dependent nucleases such as T4 EndoVII (H41, Figure C.6c, Appendix C),[42] I-HmuI (H75, Figure C.6d, Appendix C),[41] and Vvn (H80),[40] and mutational studies on *Serratia* nuclease (H89).[64] Therefore, our calculations coupled with kinetic[46] and structural data[38-39] on I-*Ppo*I supports the proposal that one metal is enough for enzymatic cleavage of phosphodiester bonds.

This is only the second time the feasibility of single-metal mediated catalysis of phosphodiester bond cleavage has been demonstrated for an endonuclease using computational techniques, with previous work focused on one-metal dependent APE1.[65-67] In contrast to I-*Ppo*I, QM cluster calculations on single-metal dependent APE1 revealed that direct bidentate metal–substrate coordination is not feasible.[65] Instead, as confirmed using QM/MM calculations, an indirect metal-binding architecture is preferred in the APE1 active site. Differences between APE1 and I-*Ppo*I arise because of the distinct active site compositions (Figure 4.8). Specifically, the presence of a strong negatively charged general base (D240) coupled with a positively charged histidine (H309) positioned to stabilize the charged substrate allows APE1 to use a metal-coordinated water to facilitate leaving group departure. In contrast, the neutral general base (H98) and the absence of a positively charged residue close to the scissile phosphate forces the single metal of I-*Ppo*I to play a more active role in promoting leaving group departure and substrate charge stabilization through direct bidentate coordination. This comparison underscores how a single metal can play diverse roles depending on the active site environment of the enzyme and how the same phosphodiester bond hydrolysis can be achieved in slightly different ways. The fundamental understanding of the unique I-*Ppo*I chemistry afforded by the present work may inspire studies on other single-metal dependent enzymes (e.g., T4 EndoVII (Figure C.6c, Appendix C),[42] Vvn,[40] and *Serratia* nuclease[64]), as well as other HEs (e.g., I-HmuI[41] and Hpy188I,[44] Figure C.6d and f, Appendix C). Such future work is necessary to understand how nature cleaves the P–O bond in unique ways, depending on the identity of the proposed general base, mode of charge stabilization, and/or number of metals in the active site.

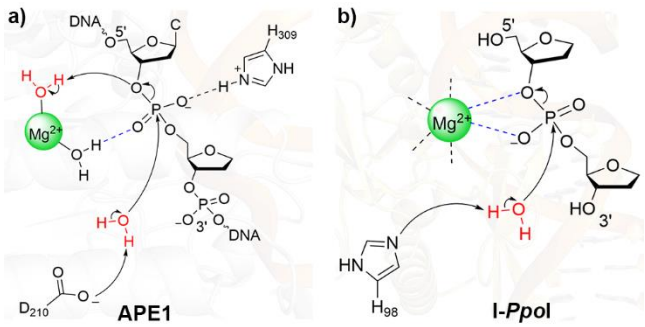

Figure 4.8. Comparison of the proposed phosphodiester bond cleavage pathways for a) APE1[65-67] and b) I-*Ppo*I (present work).

4.4. Conclusion

In the present study, a combined QM and QM/MM approach was used to provide atomic-level details of the catalytic pathway employed by I-*Ppo*I to cleave phosphodiester bonds in DNA. In the experimentally-proposed I-*Ppo*I mechanism of action,[39] a single metal facilitates P–O bond cleavage through direct coordination to the substrate coupled with simultaneous leaving group protonation via a metal-activated water. However, this previously proposed pathway could not be characterized using our smallest QM cluster or QM/MM model due to geometrical constraints imposed by the I-*Ppo*I active site. Although QM cluster and QM/MM models of varying size permitted the characterization of a pathway involving leaving group protonation by a metal-activated water, an indirect (water-mediated) metal binding configuration to the leaving group is required to achieve this chemistry and this mechanism is not energetically feasible regardless of the model size. Instead, QM cluster calculations uncovered a preferred mechanism for phosphodiester bond hydrolysis in which H98 activates the water nucleophile, while the metal provides substrate charge stabilization and promotes leaving group departure through direct coordination. QM/MM calculations verified the feasibility of this preferred pathway within the context

of the solvated enzyme–DNA complex, yielding a mechanism that is fully consistent with experimental structural,[38-39] kinetic,[56-57] and mutational data.[46, 57] Despite our calculations highlighting that simultaneous direct metal coordination to O3′ and leaving group protonation by a metal activated water is not possible during the P–O bond cleavage step within the confines of the I-*Ppo*I active site, we acknowledge that protonation of the leaving group by metal-activated or bulk water likely occurs after the chemical step as metal migration and active site rearrangement are common events in conjunction with product release for both one- and two-metal dependent nucleases.[24, 29, 35, 37, 88-91] The fundamental understanding gained about I-*Ppo*I function from the present work can be used to further explore its potential in the areas of gene therapy and genome engineering for biotechnological and therapeutic solutions. Furthermore, our new mechanistic understanding of I-*Ppo*I can direct computational investigations on other members of the HE family (e.g., I-HmuI[41] and Hpy188I[44]) which share the metal-binding configuration of I-*Ppo*I, but differ in the identities and arrangements of active site amino acids. Additionally, our work will inspire studies on other one-metal dependent nucleases (e.g., T4 EndoVII,[42] Vvn,[40] and *Serratia* nuclease[64]), with the goal to better understand how nature cleaves extremely stable phosphodiester bonds.

4.5. References

(1) Stoddard, B. L., Homing endonucleases from mobile group I introns: discovery to genome engineering. *Mob. DNA* **2014,** *5*, 1–16.

(2) Belfort, M.; Roberts, R. J., Homing endonucleases: keeping the house in order. *Nucleic Acids Res.* **1997,** *25*, 3379–88.

(3) Jurica, M.; Stoddard, B., Homing endonucleases: structure, function and evolution. *Cell. Mol. Life Sci.* **1999,** *55*, 1304–1326.

(4) Guhan, N.; Muniyappa, K., Structural and functional characteristics of homing endonucleases. *Crit. Rev. Biochem. Mol. Biol.* **2003,** *38* (3), 199–248.

(5) Kazazian Jr, H. H.; Wong, C.; Youssoufian, H.; Scott, A. F.; Phillips, D. G.; Antonarakis, S. E., Haemophilia A resulting from de novo insertion of L 1 sequences represents a novel mechanism for mutation in man. *Nature* **1988,** *332* (6160), 164–166.

(6) Hancks, D. C.; Kazazian, H. H., Roles for retrotransposon insertions in human disease. *Mob. DNA* **2016,** *7* (1), 1–28.

(7) Kobayashi, K.; Nakahori, Y.; Miyake, M.; Matsumura, K.; Kondo-Iida, E.; Nomura, Y.; Segawa, M.; Yoshioka, M.; Saito, K.; Osawa, M., An ancient retrotransposal insertion causes Fukuyama-type congenital muscular dystrophy. *Nature* **1998,** *394* (6691), 388–392.

(8) Belfort, M.; Bonocora, R. P., Homing endonucleases: from genetic anomalies to programmable genomic clippers. *Methods Mol. Biol.* **2014,** *1123*, 1–26.

(9) Jasin, M., Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet.* **1996,** *12* (6), 224–228.

(10) Dalgaard, J. Z.; Banerjee, M.; Curcio, M. J., A novel Ty1-mediated fragmentation method for native and artificial yeast chromosomes reveals that the mouse *steel* gene is a hotspot for Ty1 integration. *Genetics* **1996,** *143* (2), 673–683.

(11) Marcaida, M. J.; Muñoz, I. G.; Blanco, F. J.; Prieto, J.; Montoya, G., Homing endonucleases: from basics to therapeutic applications. *Cell. Mol. Life Sci.* **2010,** *67*, 727–748.

(12) Humbert, O.; Davis, L.; Maizels, N., Targeted gene therapies: tools, applications, optimization. *Crit. Rev. Biochem. Mol. Biol.* **2012,** *47* (3), 264–281.

(13) Redondo, P.; Prieto, J.; Munoz, I. G.; Alibés, A.; Stricher, F.; Serrano, L.; Cabaniols, J.-P.; Daboussi, F.; Arnould, S.; Perez, C., Molecular basis of xeroderma pigmentosum group C DNA recognition by engineered meganucleases. *Nature* **2008,** *456* (7218), 107–111.

(14) Arnould, S.; Perez, C.; Cabaniols, J.-P.; Smith, J.; Gouble, A.; Grizot, S.; Epinat, J.-C.; Duclert, A.; Duchateau, P.; Pâques, F., Engineered I-CreI derivatives cleaving sequences from the human XPC gene can induce highly efficient gene correction in mammalian cells. *J. Mol. Biol.* **2007,** *371* (1), 49–65.

(15) Dupuy, A.; Valton, J.; Leduc, S.; Armier, J.; Galetto, R.; Gouble, A.; Lebuhotel, C.; Stary, A.; Pâques, F.; Duchateau, P., Targeted gene therapy of xeroderma pigmentosum cells using meganuclease and TALEN™. *PLoS One* **2013,** *8* (11), e78678.

(16) Popplewell, L.; Koo, T.; Leclerc, X.; Duclert, A.; Mamchaoui, K.; Gouble, A.; Mouly, V.; Voit, T.; Pâques, F.; Cédrone, F., Gene correction of a duchenne muscular dystrophy mutation by meganuclease-enhanced exon knock-in. *Hum. Gene Ther.* **2013,** *24* (7), 692–701.

(17) Grizot, S.; Smith, J.; Daboussi, F.; Prieto, J.; Redondo, P.; Merino, N.; Villate, M.; Thomas, S.; Lemaire, L.; Montoya, G., Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. *Nucleic Acids Res.* **2009,** *37* (16), 5405–5419.

(18) Munoz, I. G.; Prieto, J.; Subramanian, S.; Coloma, J.; Redondo, P.; Villate, M.; Merino, N.; Marenchino, M.; D'Abramo, M.; Gervasio, F. L., Molecular basis of engineered meganuclease targeting of the endogenous human RAG1 locus. *Nucleic Acids Res.* **2011,** *39* (2), 729–743.

(19) Cannon, P.; June, C., CCR5 knockout strategies. *Curr. Opin. HIV AIDS* **2011,** *6* (1), 74.

(20) Klein, T.; Windbichler, N.; Deredec, A.; Burt, A.; Benedict, M., Infertility resulting from transgenic I-*Ppo*I male Anopheles gambiae in large cage trials. *Pathog. Glob. Health* **2012,** *106* (1), 20–31.

(21) Galizi, R.; Doyle, L. A.; Menichelli, M.; Bernardini, F.; Deredec, A.; Burt, A.; Stoddard, B. L.; Windbichler, N.; Crisanti, A., A synthetic sex ratio distortion system for the control of the human malaria mosquito. *Nat. Commun.* **2014,** *5* (1), 3977.

(22) Chan, Y.-S.; Takeuchi, R.; Jarjour, J.; Huen, D. S.; Stoddard, B. L.; Russell, S., The design and in vivo evaluation of engineered I-OnuI-based enzymes for HEG gene drive. *PLoS One* **2013,** *8* (9), e74254.

(23) Galburt, E. A.; Stoddard, B. L., Catalytic mechanisms of restriction and homing endonucleases. *Biochemistry* **2002,** *41* (47), 13851–13860.

(24) 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.

(25) 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.

(26) Mordasini, T.; Curioni, A.; Andreoni, W., Why do divalent metal ions either promote or inhibit enzymatic reactions? The case of BamHI restriction endonuclease from combined quantum-classical simulations. *J. Biol. Chem.* **2003,** *278* (7), 4381–4384.

(27) 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.

(28) Imhof, P.; Fischer, S.; Smith, J. C., Catalytic mechanism of DNA backbone cleavage by the restriction enzyme *Eco*RV: a quantum mechanical/molecular mechanical analysis. *Biochemistry* **2009,** *48* (38), 9061–9075.

(29) 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.

(30) 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.

(31) Araujo, A. R.; Ribeiro, A. J.; 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.

(32) Rungrotmongkol, T.; Mulholland, A. J.; Hannongbua, S., QM/MM simulations indicate that Asp185 is the likely catalytic base in the enzymatic reaction of HIV-1 reverse transcriptase. *MedChemComm* **2014,** *5* (5), 593–596.

(33) Xiao, S.; Klein, M. L.; LeBard, D. N.; Levine, B. G.; Liang, H.; MacDermaid, C. M.; Alfonso-Prieto, M., Magnesium-dependent RNA binding to the PA endonuclease domain of the avian influenza polymerase. *J. Phys. Chem. B* **2014,** *118* (4), 873–889.

(34) Casalino, L.; Nierzwicki, L.; 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.

(35) 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.

(36) 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 *Eco*RI family. *J. Mol. Biol.* **2009,** *393* (1), 140–160.

(37) Xie, F.; Qureshi, S. H.; Papadakos, G. A.; Dupureur, C. M., One- and two-metal ion catalysis: global single-turnover kinetic analysis of the PvuII endonuclease mechanism. *Biochemistry* **2008,** *47* (47), 12540–12550.

(38) Flick, K. E.; Jurica, M. S.; Monnat, R. J.; Stoddard, B. L., DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-*Ppo*I. *Nature* **1998,** *394* (6688), 96–101.

(39) Galburt, E. A.; Chevalier, B.; Tang, W.; Jurica, M. S.; Flick, K. E.; Monnat, R. J.; Stoddard, B. L., A novel endonuclease mechanism directly visualized for I-*Ppo*I. *Nat. Struct. Biol.* **1999,** *6* (12), 1096–1099.

(40) 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.

(41) Shen, B. W.; Landthaler, M.; Shub, D. A.; Stoddard, B. L., DNA binding and cleavage by the HNH homing endonuclease I-HmuI. *J. Mol. Biol.* **2004,** *342* (1), 43–56.

(42) Biertümpfel, C.; Yang, W.; Suck, D., Crystal structure of T4 endonuclease VII resolving a Holliday junction. *Nature* **2007,** *449* (7162), 616–620.

(43) Dalhus, B.; Arvai, A. S.; Rosnes, I.; Olsen, Ø. E.; Backe, P. H.; Alseth, I.; Gao, H.; Cao, W.; Tainer, J. A.; Bjørås, M., Structures of endonuclease V with DNA reveal initiation of deaminated adenine repair. *Nat. Struct. Mol. Biol.* **2009,** *16*, 138–143.

(44) Sokolowska, M.; Czapinska, H.; Bochtler, M., Hpy188I–DNA pre- and post-cleavage complexes—snapshots of the GIY-YIG nuclease mediated catalysis. *Nucleic Acids Res.* **2011,** *39* (4), 1554–1564.

(45) Freudenthal, B. D.; Beard, W. A.; Cuneo, M. J.; Dyrkheeva, N. S.; Wilson, S. H., Capturing snapshots of APE1 processing DNA damage. *Nat. Struct. Mol. Biol.* **2015,** *22* (11), 924–931.

(46) Eastberg, J. H.; Eklund, J.; Monnat, R.; Stoddard, B. L., Mutability of an HNH nuclease imidazole general base and exchange of a deprotonation mechanism. *Biochemistry* **2007,** *46* (24), 7215–7225.

(47) Ellison, E. L.; Vogt, V. M., Interaction of the Intron-Encoded Mobility Endonuclease I-*Ppo*I with its Target Site. *Mol. Cell. Biol.* **1993,** *13* (12), 7531–7539.

(48) Flick, K. E.; McHUGH, D.; Stoddard, B. L.; Heath, J. D.; Stephens, K. M.; Monnat Jr, R. J., Crystallization and preliminary X-ray studies of I-*Ppo*I: A nuclear, intron-encoded homing endonuclease from Physarum polycephalum. *Protein Sci.* **1997,** *6* (12), 2677–2680.

(49) Argast, G. M.; Stephens, K. M.; Emond, M. J.; Monnat Jr, R. J., I-*Ppo*I and I-CreI homing site sequence degeneracy determined by random mutagenesis and sequential in vitro enrichment. *J. Mol. Biol.* **1998,** *280* (3), 345–353.

(50) Lin, J.; Vogt, V. M., I-*Ppo*I, the endonuclease encoded by the group I intron PpLSU3, is expressed from an RNA polymerase I transcript. *Mol. Cell. Biol.* **1998,** *18* (10), 5809–5817.

(51) Wittmayer, P. K.; McKenzie, J. L.; Raines, R. T., Degenerate DNA recognition by I-*Ppo*I endonuclease. *Gene* **1998,** *206* (1), 11–21.

(52) Monnat Jr, R. J.; Hackmann, A. F.; Cantrell, M. A., Generation of highly site-specific DNA double-strand breaks in human cells by the homing endonucleases I-*Ppo*I and I-CreI. *Biochem. Biophys. Res. Comm.* **1999,** *255* (1), 88–93.

(53) Galburt, E. A.; Chadsey, M. S.; Jurica, M. S.; Chevalier, B. S.; Erho, D.; Tang, W.; Monnat Jr, R. J.; Stoddard, B. L., Conformational changes and cleavage by the homing endonuclease I-*Ppo*I: a critical role for a leucine residue in the active site. *J. Mol. Biol.* **2000,** *300* (4), 877–887.

(54) McCarthy, C. B.; Romanowski, V., Digestion of I-*Ppo*I recognition sites in unfavorable sequence contexts achieved by changing the reaction conditions. *Biochem. Genet.* **2006,** *44* (1), 58–65.

(55) Eklund, J. L.; Ulge, U. Y.; Eastberg, J.; Monnat Jr, R. J., Altered target site specificity variants of the I-*Ppo*I His-Cys box homing endonuclease. *Nucleic Acids Res.* **2007,** *35* (17), 5839–5850.

(56) Wittmayer, P. K.; Raines, R. T., Substrate binding and turnover by the highly specific I-*Ppo*I endonuclease. *Biochemistry* **1996,** *35* (3), 1076–1083.

(57) Mannino, S. J.; Jenkins, C. L.; Raines, R. T., Chemical mechanism of DNA cleavage by the homing endonuclease I-*Ppo*I. *Biochemistry* **1999,** *38* (49), 16178–16186.

(58) Turkki, V.; Schenkwein, D.; Timonen, O.; Husso, T.; Lesch, H. P.; Ylä-Herttuala, S., Lentiviral protein transduction with genome-modifying HIV-1 Integrase-I-*Ppo*I fusion proteins: Studies on specificity and cytotoxicity. *Biomed Res. Int.* **2014,** *2014*, 379340.

(59) Windbichler, N.; Papathanos, P. A.; Crisanti, A., Targeting the X chromosome during spermatogenesis induces Y chromosome transmission ratio distortion and early dominant embryo lethality in Anopheles gambiae. *PLoS Genet.* **2008,** *4* (12), e1000291.

(60) Friedhoff, P.; Franke, I.; Krause, K. L.; Pingoud, A., Cleavage experiments with deoxythymidine 3′, 5′-bis-(p-nitrophenyl phosphate) suggest that the homing endonuclease I-*Ppo*I follows the same mechanism of phosphodiester bond hydrolysis as the non-specific *Serratia* nuclease. *FEBS Lett.* **1999,** *443* (2), 209–214.

(61) Chevalier, B. S.; Stoddard, B. L., Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res.* **2001,** *29* (18), 3757–3774.

(62) 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. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2020,** *10* (5), e1471.

(63) Geronimo, I.; Vidossich, P.; Donati, E.; De Vivo, M., Computational investigations of polymerase enzymes: structure, function, inhibition, and biotechnology. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2021**, e1534.

(64) Shlyapnikov, S.; Lunin, V.; Perbandt, M.; Polyakov, K.; Lunin, V. Y.; Levdikov, V.; Betzel, C.; Mikhailov, A., Atomic structure of the Serratia marcescens endonuclease at 1.1 Å resolution and the enzyme reaction mechanism. *Acta Crystallogr. D* **2000,** *56* (5), 567–572.

(65) 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.

(66) 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.

(67) 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.

(68) DeLano, W. L., Pymol: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr* **2002,** *40* (1), 82–92.

(69) Siegbahn, P. E. M.; Himo, F., Recent developments of the quantum chemical cluster approach for modeling enzyme reactions. *J. Biol. Inorg. Chem.* **2009,** *14* (5), 643–651.

(70) Kazemi, M.; Sheng, X.; Kroutil, W.; Himo, F., Computational study of Mycobacterium smegmatis acyl transferase reaction mechanism and specificity. *ACS Catal.* **2018,** *8* (11), 10698–10706.

(71) Acosta-Silva, C.; Bertran, J.; Branchadell, V.; Oliva, A., Phosphoryl Transfer Reaction in RNA: Is the Substrate-Assisted Catalysis a Possible Mechanism in Certain Solvents? *J. Phys. Chem. A* **2017,** *121* (44), 8525–8534.

(72) Fukui, K., The path of chemical reactions - the IRC approach. *Acc. Chem. Res.* **1981,** *14* (12), 363–368.

(73) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H., Electrostatic basis for enzyme catalysis. *Chem. Rev.* **2006,** *106* (8), 3210–3235.

(74) Wappett, D. A.; Goerigk, L., Benchmarking Density Functional Theory Methods for Metalloenzyme Reactions: The Introduction of the MME55 Set. *J. Chem. Theory Comput.* **2023,** *19* (22), 8365–8383.

(75) Castro-Amorim, J.; Oliveira, A.; Mukherjee, A. K.; Ramos, M. J.; Fernandes, P. A., Unraveling the Reaction Mechanism of Russell's Viper Venom Factor X Activator: A Paradigm for the Reactivity of Zinc Metalloproteinases? *J. Chem. Inf. Model.* **2023,** *63*, 4056–4069.

(76) Gérard, E. F.; Mokkawes, T.; Johannissen, L. O.; Warwicker, J.; Spiess, R. R.; Blanford, C. F.; Hay, S.; Heyes, D. J.; de Visser, S. P., How Is Substrate Halogenation Triggered by the Vanadium Haloperoxidase from Curvularia inaequalis? *ACS Catal.* **2023,** *13*, 8247–8261.

(77) Shirazi, J.; Jafari, S.; Ryde, U.; Irani, M., Catalytic Reaction Mechanism of Glyoxalase II: A Quantum Mechanics/Molecular Mechanics Study. *J. Phys. Chem. B* **2023,** *127*, 4480–4495.

(78) Ritacca, A. G.; Rovaletti, A.; Moro, G.; Cosentino, U.; Ryde, U.; Sicilia, E.; Greco, C., Unraveling the reaction mechanism of Mo/Cu CO dehydrogenase using QM/MM calculations. *ACS Catal.* **2022,** *12* (12), 7336–7343.

(79) Fernandes, H. S.; Sousa, S. F.; Cerqueira, N. M., New insights into the catalytic mechanism of the SARS-CoV-2 main protease: An ONIOM QM/MM approach. *Mol. Divers.* **2022,** *26* (3), 1373–1381.

(80) Prejanò, M.; Medina, F. E.; Ramos, M. J.; Russo, N.; Fernandes, P. A.; Marino, T., How the destabilization of a reaction intermediate affects enzymatic efficiency: the case of human Transketolase. *ACS Catal.* **2020,** *10* (4), 2872–2881.

(81) Sánchez, L.; Medina, F. E.; Mendoza, F.; Febres-Molina, C.; Jaña, G. A., Elucidation of the Reaction Mechanism of Cavia porcellus l-Asparaginase: A QM/MM Study. *J. Chem. Inf. Model* **2022,** *63* (1), 270–280.

(82) Bím, D.; Navrátil, M.; Gutten, O.; Konvalinka, J.; Kutil, Z.; Culka, M.; Navrátil, V.; Alexandrova, A. N.; Barinka, C.; Rulisek, L., Predicting effects of site-directed mutagenesis on enzyme kinetics by QM/MM and QM calculations: a case of glutamate carboxypeptidase II. *J. Phys. Chem. B.* **2022,** *126* (1), 132–143.

(83) Rostkowski, M.; Olsson, M. H. M.; Søndergaard, C. R.; Jensen, J. H., Graphical analysis of pH-dependent properties of proteins predicted using PROPKA. *BMC Struct. Biol.* **2011,** *11* (1), 1–6.

(84) 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.

(85) Galindo-Murillo, R.; Robertson, J. C.; Zgarbova, M.; Sponer, J.; Otyepka, M.; Jurecka, P.; Cheatham III, T. E., Assessing the current state of amber force field modifications for DNA. *J. Chem. Theory Comput.* **2016,** *12* (8), 4114–4127.

(86) Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Petersson, G.; Nakatsuji, H., Gaussian 16 Rev. B. 01, Wallingford, CT. **2016**.

(87) Neese, F., The ORCA program system. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012,** *2* (1), 73–78.

(88) 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 APE1. *Proteins* **2007,** *68* (1), 313–323.

(89) Dupureur, C. M., Roles of metal ions in nucleases. *Curr. Opin. Chem. Biol.* **2008,** *12* (2), 250–255.

(90) Yang, W., Nucleases: diversity of structure, function and mechanism. *Q. Rev. Biophys.* **2011,** *44* (1), 1–93.

(91) He, H.; Chen, Q.; Georgiadis, M. M., High-resolution crystal structures reveal plasticity in the metal binding site of apurinic/apyrimidinic endonuclease 1. *Biochemistry* **2014,** *53* (41), 6520–6529.

(92) Jeltsch, A.; Alves, J.; Wolfes, H.; Maass, G.; Pingoud, A., Substrate-assisted catalysis in the cleavage of DNA by the EcoRI and EcoRV restriction enzymes. *Proc. Natl. Acad. Sci. U.S.A.* **1993,** *90* (18), 8499–8503.

(93) Viadiu, H.; Aggarwal, A. K., The role of metals in catalysis by the restriction endonuclease Bam HI. *Nat. Struct. Biol.* **1998,** *5* (10), 910–916.

(94) Gao, H.; Ke, Z.; DeYonker, N. J.; Wang, J.; Xu, H.; Mao, Z.-W.; Phillips, D. L.; Zhao, C., Dinuclear Zn(II) Complex Catalyzed Phosphodiester Cleavage Proceeds via a Concerted Mechanism: A Density Functional Theory Study. *J. Am. Chem. Soc.* **2011,** *133* (9), 2904–2915.

(95) Zhang, X.; Gao, H.; Xu, H.; Xu, J.; Chao, H.; Zhao, C., A density functional theory study of the hydrolysis mechanism of phosphodiester catalyzed by a mononuclear Zn (II) complex. *J. Mol. Catal. A Chem.* **2013,** *368*, 53–60.

(96) Alberto, M. E.; Pinto, G.; Russo, N.; Toscano, M., Triesterase and promiscuous diesterase activities of a di-CoII-containing organophosphate degrading enzyme reaction mechanisms. *Chem. Eur. J.* **2015,** *21* (9), 3736–3745.

(97) 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.* **2007,** *67* (1), 143–154.

(98) Balhara, R.; Chatterjee, R.; Jindal, G., A computational approach to understand the role of metals and axial ligands in artificial heme enzyme catalyzed C–H insertion. *Phys. Chem. Chem. Phys.* **2021,** *23* (15), 9500–9511.

(99) Cheng, Q.; DeYonker, N. J., QM-Cluster Model Study of the Guaiacol Hydrogen Atom Transfer and Oxygen Rebound with Cytochrome P450 Enzyme GcoA. *J. Phys. Chem. B.* **2021,** *125* (13), 3296–3306.

(100) Planas, F.; McLeish, M. J.; Himo, F., Enzymatic Stetter Reaction: Computational Study of the Reaction Mechanism of MenD. *ACS Catal.* **2021,** *11* (19), 12355–12366.

(101) Summers, T. J.; Cheng, Q.; Palma, M. A.; Pham, D.-T.; Kelso, D. K.; Webster, C. E.; DeYonker, N. J., Cheminformatic quantum mechanical enzyme model design: A catechol-O-methyltransferase case study. *Biophys. J.* **2021,** *120* (17), 3577–3587.

(102) Feng, H.; Klutz, A. M.; Cao, W., Active site plasticity of endonuclease V from Salmonella typhimurium. *Biochemistry* **2005,** *44* (2), 675–683.

(103) 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.

(104) Gong, S.; Yu, H. H.; Johnson, K. A.; Taylor, D. W., DNA unwinding is the primary determinant of CRISPR-Cas9 activity. *Cell Rep.* **2018,** *22* (2), 359–371.

(105) Wu, H.; Lima, W. F.; Crooke, S. T., Investigating the structure of human RNase H1 by site-directed mutagenesis. *J. Biol. Chem.* **2001,** *276* (26), 23547–23553.

(106) Eckstein, F., Nucleoside phosphorothioates. *Annu. Rev. Biochem.* **1985,** *54* (1), 367–402.

Chapter 5: Is Metal Stabilization of the Leaving Group Required or Can Lysine Facilitate Phosphodiester Bond Cleavage in Nucleic Acids? A Computational Study of EndoV

Preface: Chapter 5 has been published [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]. In this chapter, both authors contributed to the reviewing and editing of the manuscript. I aided conceptualization of the project and conducted all calculations. In addition, I performed the overall data/results analysis, wrote the first draft of the manuscript, and generated all figures for the publication. S. D. Wetmore was involved in project conceptualization and administration, supervision, funding and resource acquisition, data visualization and interpretation, and writing the original draft of the manuscript.

5.1. Introduction

Deamination of DNA is a common event triggered by exposure to endogenous (e.g., free radicals generated from inflammation or infection) and exogenous (e.g., pollution and UV light) agents.[1-2] On the other hand, deamination of RNA is a key post-transcriptional modification that arises from spontaneous hydrolysis or nitrosative stress, imbalanced purine metabolism, or the action of adenosine deaminases.[3-4] Deamination alters the interactions of a nucleobase with the surrounding nucleotides, which can in turn influence a range of biological processes. For example, deamination of DNA nucleobases can lead to mispairs during replication or transcription,[5] while dysregulation of RNA deamination

results in codon changes and alters protein translation.[6] As a result, high accumulation of deaminated sites in nucleic acids has been linked to various diseases including cancer (glioblastoma)[7] and cardiovascular (atherosclerosis)[8] or neurological (epilepsy,[9] ALS,[10] autism,[11] Alzheimers,[12] and schizophrenia)[13] disorders.

Endonuclease V (EndoV) is a conserved family of enzymes across all domains of life that is involved in removing deaminated nucleobases by facilitating the challenging cleavage of the phosphodiester bond in the nucleic acid backbone.[6, 14] For example, bacterial EndoV breaks the DNA phosphodiester bond one nucleotide away from a deaminated nucleobase in the 3′ direction as part of the alternative excision repair pathway,[15-19] while mammalian EndoV is involved in RNA metabolism.[6] Beyond these roles in the cell, the nicking activity of EndoV has proven to be a powerful tool in biotechnology and medicine. For example, *Thermotoga maritima* (*Tma*) EndoV along with a DNA ligase has been used to detect mutations in the BRCA1 gene linked to breast and ovarian cancers and the p53 gene related to colorectal tumors.[20-22] EndoV in conjunction with a DNA polymerase has also been successfully applied in DNA amplification[23] to rapidly and accurately detect infectious diseases[24] and antimicrobial resistance.[25] Furthermore, EndoV has proven useful to enhance detection of RNA editing sites that are upregulated in brain development, autism spectrum disorder, glioblastoma, and schizophrenia.[26-27] EndoV has been implicated in the development of hepatocellular carcinoma in mice and overexpression of the enzyme in humans leads to an enhanced sensitivity towards hepatocellular carcinoma treatments.[28] EndoV has also been used for directed evolution via *in vitro* DNA shuffling.[29-30] To build upon these applications and

further exploit the enzyme in interdisciplinary applications, the catalytic mechanism used by EndoV must be understood.

EndoV is related to several endonucleases that have a conserved DDD/E motif, including those whose catalytic function relies on two metals such as RNase H (PDB ID: 1ZBI)[31] and FLAP endonuclease 1 (PDB ID: 3Q8M).[32] The two-metal mediated mechanism has been well established and proposed to be the general pathway used by many nucleases.[33-36] In this mechanism (Figure 5.1a), the first metal (Mg_A^{2+}) activates the water nucleophile and the second metal (Mg_B^{2+}) promotes leaving group departure, while both metals likely provide substrate charge stabilization during transition state (TS) formation. Although eukaryotic (mouse) EndoV contains a DDD motif and the active site arrangement resembles that in two-metal containing endonucleases (Figure D.1a, Appendix D),[17] an X-ray structure of bacterial EndoV has uncovered a DDH motif (Figure 5.2a) and proposed the enzyme uses a single metal for catalysis.[19] Although the enzyme must facilitate nucleophile activation, leaving group departure, and substrate (transition state) charge stabilization, the mechanisms of action of nucleases that depend on a single metal for catalysis are not well understood.

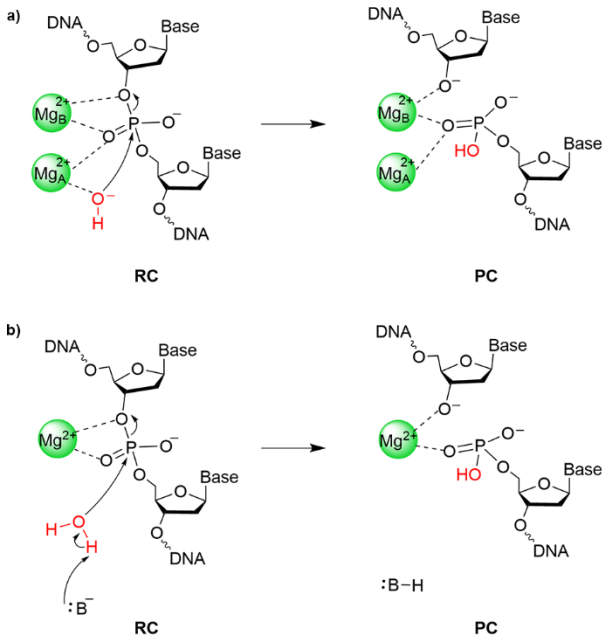

Figure 5.1. The proposed general mechanisms of action for a) two-metal and b) one-metal dependent endonucleases (B = basic amino acid residue).

There is growing evidence that supports the ability of nucleases to exhibit catalytic activity in the presence of a single metal ion.[37-46] Indeed, kinetic studies have investigated the metal concentration dependence of different enzymes and concluded that only one metal is required for their catalytic activity.[37, 45, 47-48] X-ray crystal structures have also revealed a single metal bound in the active site of various nucleases, including human apurinic/apyrimidinic endonuclease (APE1),[49-50] *Vibrio vulnificus* nuclease (Vvn),[41] T4 endonuclease VII,[51] *Eco*RI,[52] homing endonuclease I-HmuI,[53] and intron-encoded endonuclease from *Physarum polycephalum* (I-*Ppo*I, Figure 5.3).[39, 54] This has led to proposals that subtleties in active site architectures allow different enzymes to replace the first metal (Mg_A^{2+}) that is involved in nucleophile activation by an amino acid (Figure 5.1b). For instance, a histidine residue (H98) in I-*Ppo*I has been proposed to activate the water nucleophile, while the single metal directly coordinates with the substrate to provide

TS stabilization and promote leaving group departure (Figure 5.3c).[39, 54-55] In contrast, D210 activates the water nucleophile in the proposed mechanism of action for APE1, while a metal-activated water stabilizes the leaving group and H309 delivers charge stabilization to the scissile phosphate during TS formation (Figures 5.3a and 5.3b), which hints the remaining metal may be involved in the reaction in different ways.[49-50, 56] However, no proposals have been made for the chemical pathway used by single-metal dependent bacterial EndoV.

Although the structures of the reactant complexes (RCs) for the D110N *Tma* EndoV mutant crystallized in the presence of Ca^{2+} (PDB ID: 6OZF, Figure D.2a, Appendix D) and the E89Q mutant crystallized with Mg^{2+} (PDB ID: 6OZG, Figure D.2b, Appendix D) have been characterized,[57] the crystal structure of the product complex (PC, Figure 5.2a) suggests that these structures were obtained by mutating residues critical for metal coordination, which raises questions about the position of the metal in the RC of the wild-type enzyme. On the other hand, the crystal structure of the PC for *Tma* EndoV (PDB ID: 2W35) has provided valuable insights into the base recognition mechanism (Figure 5.2a), including multiple interactions with the phosphate backbone, flipping of the lesion into the recognition pocket, the ability of the PYIP motif to occupy the space left in DNA after lesion flipping, and mainchain interactions between amino acids and the deaminated base that mediate recognition.[17] Furthermore, the PC highlights that the active site contains a single metal, with direct metal–substrate coordination through the O3′ leaving group, and the remainder of the metal octahedral coordination satisfied by 2 aspartates (D43 and D110) and 3 water molecules, while E89 interacts with Mg^{2+}-coordinated waters (Figure 5.2a). H214 and K139 are also uncovered as potentially important residues, falling within

hydrogen-bonding distance of the cleaved substrate. To complement structural information, experimental kinetic data has shown that the H214A mutation reduces catalytic activity by 20%,[47] while mutation of the conserved K155 to alanine in the mouse homologue completely abolishes enzymatic activity, supporting important roles for these residues in catalysis.[57]

Despite the currently available structural information for bacterial EndoV based on the PC, previous literature has revealed that the metal-binding architecture can vary in the crystallized RC and PC of nucleases at least in part because of the well-accepted active site rearrangement that occurs prior to final product formation due to metal ion migration.[31, 45, 48, 58-60] Therefore, crystal structures of PCs have led to incorrect conclusions about the metal coordination geometry and catalytic role for other single-metal mediated endonucleases. For example, although the PC analogue of APE1 (PDB ID: 4IEM, Figure 5.3a) displays direct metal–substrate ligation,[49] the phosphorothioate RC analogue shows indirect Mg^{2+} coordination to the substrate (PDB ID: 5DG0, Figure 5.3b),[50] and QM and QM/MM calculations initiated from the crystal structure of the PC suggest that indirect metal–substrate ligation is preferred.[56] Furthermore, while several computational works have characterized the pathway for various two-metal-dependent nucleases,[61-68] to the best of our knowledge the mechanism of one-metal mediated phosphodiester bond cleavage has only been investigated for APE1.[56, 69-70] Therefore, the catalytic mechanism of one-metal dependent EndoV is unclear, with missing atomic-level information leaving unanswered questions surrounding the identity of the residues that activate the water nucleophile, provide TS charge stabilization, and promote leaving group departure, as well as the role of the single metal.

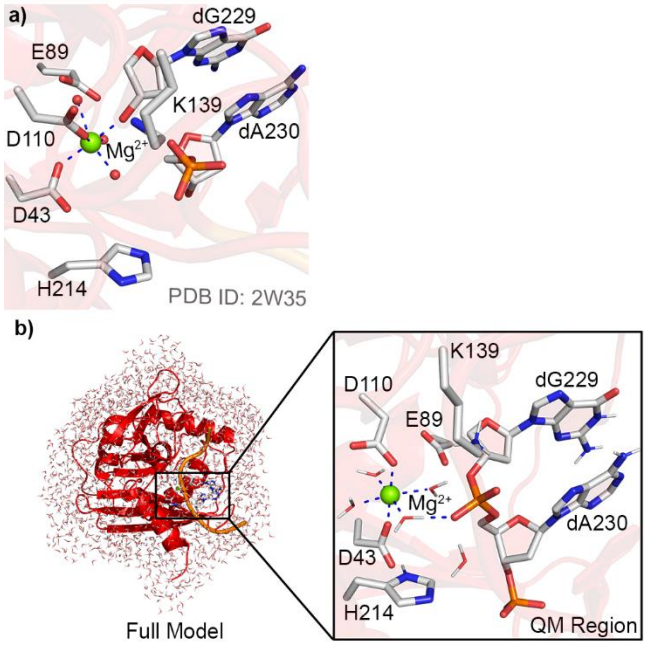

Figure 5.2. a) Active site from an X-ray crystal structure of the Mg^{2+}-containing wild-type PC of *Tma* EndoV bound to a ssDNA substrate and b) a representative enzyme–DNA quantum mechanics-molecular mechanics (QM/MM) model.

Due to the successful characterization of the catalytic pathways of many metalloenzymes using computational techniques[56, 61, 63, 65, 71-76] and the important applications of the EndoV family of enzymes,[20-23, 26-30] the current contribution uses molecular dynamics (MD) simulations and QM/MM calculations within the ONIOM formulism to map the bacterial EndoV mechanism of action. MD simulations were initially performed on the one metal-containing EndoV–DNA RC to gain an understanding of the structural dynamics and identify multiple active site arrangements that correspond to different potential chemical pathways. Based on structural data from the MD simulations, eight unique catalytic mechanisms were characterized for wild-type EndoV using QM/MM that deviate in the metal coordination to the substrate (direct versus indirect), the general base that activates the water nucleophile, and the general acid that facilitates leaving group departure. Furthermore, *in silico* mutations to H214 and K139 were considered to further

explore the roles of these residues in phosphodiester bond hydrolysis. As a result, we propose preferred catalytic pathways that involve previously unseen combinations of a general base, general acid, and metal-binding architecture for a one-metal dependent endonuclease. Our proposed mechanisms are fully consistent with experimental structural,[17] kinetic,[16] and mutational data.[47, 57] As the single metal in the active site of bacterial EndoV is in a similar position with respect to the substrate as the second metal (M_B^{2+}) in many two-metal dependent nucleases (Figures 5.2 and 5.3), our work further confirms that an appropriately positioned active site residue can replace one metal by facilitating water nucleophile activation and thereby emphasizes that one metal is enough for the chemical step. The fundamental understanding gained from this study will aid exploration of new and improved applications of bacterial EndoV in biotechnology[20-23, 29-30] and medicine,[26-28] including the detection of disease inducing gene mutations and the design of therapeutic solutions for ailments such as neurological disorders and carcinomas.

5.2. Computational Methodology

5.2.1. Model Building

The crystal structure of the *Tma* EndoV–DNA product complex (PDB ID: 2W35) in which 3′-O of the substrate is directly ligated to Mg^{2+} and the remainder of the octahedral coordination is fulfilled by 3 water molecules, D43, and D110 (Figure 5.2a) was used to build all starting models. The PC was modified to generate two reactant complexes based on the metal coordination environments observed in X-ray structures of RC mimics for other single-metal mediated endonucleases (Figure 5.3b and 5.3e).[39, 50] Specifically, the

metal directly coordinates with the substrate in one RC and indirectly coordinates with the substrate through a metal-ligated water in the second RC (Figure D.3, Appendix D). The model containing direct metal–substrate ligation was built by forming a phosphodiester bond between the 5′-phosphate of dA230 and 3′-O of dG229, and directly coordinating the non-bridging oxygen of the scissile phosphate to Mg^{2+}. This led to an octahedral Mg^{2+} coordination geometry involving direct ligation to D43, D110, 3′-O of dG229, 5′-phosphate of dA230, and two water molecules. The model containing indirect ligation was similarly built with the exception of replacing 3′-O of dG229 and the 5′-phosphate of dA230 with two additional water molecules directly coordinated to Mg^{2+}. Subsequently, the positions of the metal and coordinating residues were manually adjusted to ensure Mg^{2+} adopts an octahedral coordination geometry that involves four water molecules, D43, and D110, with one Mg^{2+}-ligated water hydrogen bonded to the non-bridging oxygen of the scissile phosphate and another to the O3′-leaving group. We note that Mg^{2+} is known to prefer an octahedral coordination geometry with hard oxygen-containing ligands such as the Asp residues present in the EndoV active site, and this is the standard coordination geometry found in many other Mg^{2+}-dependent enzymes, including nucleases (Figure 5.2). In each model, the non-standard deoxyinosine (dI) residue was considered in its tautomeric enol form since structural data suggests dI undergoes tautomerization upon complexation with EndoV.[14, 17] The physiologically-relevant protonation states of the titratable amino acids at pH 7.0 were assigned using PROPKA.[77] Two *in silico* K139A EndoV mutant models were built by mutating K139 to alanine in QM/MM models in which H214 or a substrate phosphate group was aligned to activate the water nucleophile. The *in silico* H214A EndoV mutant was built by mutating H214 to alanine in the QM/MM model in which a substrate

phosphate group was positioned to activate the water nucleophile (see the results section for further discussion and justification of the mutant models).

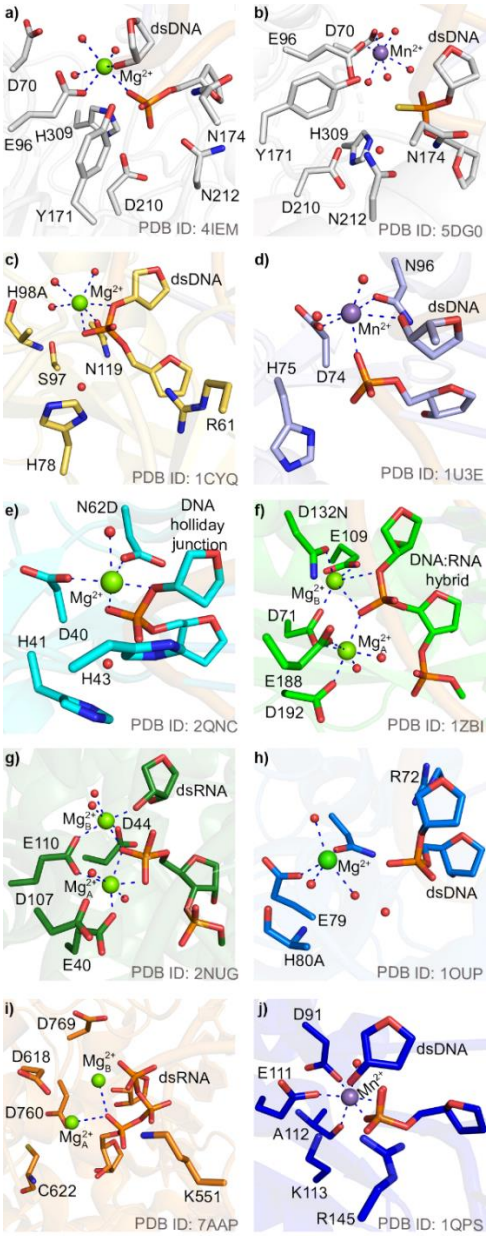

Figure 5.3. Active sites from X-ray crystal structures of a: a) PC of wild-type APE1, b) RC of thio-substituted APE1, c) RC of the H98A mutant of I-*Ppo*I, d) PC of wild-type I-HmuI, e) RC of the N62D mutant of T4 endonuclease VII, f) RC of the D132N mutant of RNase H, g) PC of wild-type *Aquifex aeolicus* RNase III, h) RC of the H80A mutant of Vvn nuclease, i) RC of favipiravir-RTP containing SARS-CoV2 RNA-dependent RNA polymerase, and j) PC of wild-type *Eco*RI.

5.2.2. MD Simulation Protocol

The AMBER ff14SB[78] and OL15[79] force fields were used to describe EndoV and DNA, respectively, and the missing force field parameters for dI were supplemented according to the generalized AMBER force field (GAFF)[80] using ANTECHAMBER 17.3.[81] Partial charges were assigned using the restrained electrostatic potential (RESP) from the R.E.D.v.III.4 program and the B3LYP/6-31G(d) level of theory.[82] The parameters for Mg^{2+} were adopted from Allner *et al.* and the +2 formal charge was maintained.[83] The enzyme–DNA complex was neutralized with Na^{+} ions using the LEaP module of Amber18[84] and solvated in a truncated octahedral TIP3P water box, with a minimum of 10.0 Å between the solute and the box faces. PMEMD module was used to minimize the system in six steps. The first minimization phase involved 2500 steps of the steepest descent and 5000 steps of the conjugate gradient method, with a force constraint of 100 kcal mol^{-1} $Å^{-2}$ applied to the solute while the solvent was minimized. In the second step, solute hydrogen atoms were minimized while the rest of the solute was restrained (100 kcal mol^{-1} $Å^{-2}$). The third step involved solute minimization while applying restraints (100 kcal mol^{-1} $Å^{-2}$) to the solvent and Mg^{2+}-coordination distances to maintain an octahedral coordination geometry. The whole enzyme–DNA complex along with the solvent was minimized while keeping restraints (100 kcal mol^{-1} $Å^{-2}$) only on the Mg^{2+}-coordination distances in the fourth step. The fifth step involved solute minimization while restraining the solvent, and the final step minimized the entire enzyme–DNA complex and surrounding water with no restraints.

The minimized system was heated from 10 to 310 K in 50 K increments using a Langevin thermostat, with a restraint of 25 kcal mol^{-1} $Å^{-2}$ placed on the solute. Furthermore, 200 ps of equilibration was performed with the restraints on the system gradually released

from 20 kcal mol^{-1} $Å^{-2}$ to 1.5 kcal mol^{-1} $Å^{-2}$. A cut-off of 10.0 Å was used to describe the non-bonded interactions, while the particle mesh Ewald (PME) method was used to treat long-range electrostatic interactions in all simulations. Finally, 500 ns MD production simulations were performed in triplicate without restraints under NPT conditions using the periodic boundary condition and applying the SHAKE algorithm to all bonds involving hydrogen atoms.

The resulting trajectories were analyzed using the cpptraj (V18.01) module of the Ambertools 2018 suite.[84] A hydrogen bond was considered to be present when the distance between the two heavy atoms was < 3.4 Å and the hydrogen-bond angle was > 120°. Root-mean-square deviations (RMSDs) for the enzyme, DNA, and enzyme–DNA complex were evaluated based on the backbone atoms of the enzyme and/or DNA with respect to first frame of the production simulation. The RMSD for the active site, which was calculated based on the heavy atoms of D43, D110, H214, K139, E89, the nucleotides connected by the phosphodiester bond being cleaved (dG229 and dA230), and an additional nucleotide (dG231) on the 3′ side of dA230, indicate that all complexes were stable (Figure D.4 and Table D.1, Appendix D). Furthermore, per residue root-mean-square fluctuations (RMSFs), which were calculated with respect to all heavy atoms in the amino acid residue, indicate that metal-coordinating residues (D43 and D110) are considerably stable (Figure D.5, Appendix D). Nevertheless, the EndoV active site was extremely dynamic due to the highly flexible nature of the single-stranded DNA substrate (Figure D.4c–d, Appendix D), which resulted in the enzyme being easily able to adopt and convert between multiple catalytically active conformations throughout the MD simulations. Representative structures for the catalytically-active conformations for different mechanistic possibilities were obtained

using reaction parameter-based criterion (Figure D.6 and Table 5.1, Appendix D). Specifically, the Mg^{2+}-ligated water or K139 was considered to be well aligned to function as the general acid when the hydrogen-bonding distance between the heavy atoms and the O3′ leaving group was < 3.4 Å. Mg^{2+} was considered to be positioned to directly coordinate to the leaving group when the distance to the O3′ leaving group was < 2.5 Å. The substrate phosphate or H214 was considered to be a possible general base when the distance between the non-bridging oxygen of the substrate phosphate or Nε of H214 and the phosphorus reaction centre was < 6.5 Å. A nucleophilic water was deemed posed for the reaction when the distance between oxygen of the water and the phosphorus reaction centre was < 4 Å. A water chain was considered to bridge D43 and the phosphorus reaction center when a water molecule was within 3.4 Å of Oδ of D43 and 3.4 Å of oxygen in another water that was within 4 Å of the phosphorus centre.

5.2.3. QM/MM Calculations

Based on our detailed analysis of the conformational dynamics of the EndoV active site, eight unique MD snapshots were extracted across the three MD simulation replicas to characterize the catalytic mechanism with QM/MM that differ in the combination of the general base and acid, and metal coordination environment (Figure D.6 and Table 5.1, Appendix D). QM/MM calculations were performed using the ONIOM formulism due to this approach successfully characterizing the reaction pathways for many enzymes, including endonucleases.[56, 62, 85-88] Regardless of the metal–substrate coordination geometry, the QM region included D43, D110, and two water molecules directly coordinated to Mg^{2+}, H214, K139, E89, a nucleophilic water, the (dG229, dA230)

nucleotides directly involved in the reaction, and an additional phosphate (dG231) on the 3′ side of dA230 that may activate the nucleophilic water (Figures 5.2b and D.7, Appendix D). In addition to the residues mentioned above, the QM region for models involving indirect metal–substrate ligation included two additional water molecules coordinated to Mg^{2+} to fulfill the octahedral coordination geometry (Figure 5.2b), models for direct bidentate metal–substrate ligation included two additional water molecules that hydrogen bonds to D43 or E89 in the QM region (Figure D.7a, Appendix D), and models with direct monodentate metal–substrate ligation included an additional water directly coordinated to Mg^{2+} to fulfill the octahedral coordination geometry and another water that hydrogen bonds to D43 (Figure D.7b, Appendix D). As a result, all models contain a QM region of 129 atoms and a charge of –2. The MM region contained the remaining EndoV–DNA complex and all water molecules for which any atom is within 6 Å of the enzyme–DNA complex. The MM region had an overall charge of –1.

The QM/MM boundary was placed between C4′ and C5′ of dG231 and dG229 of the DNA substrate, and Cα and Cβ of the amino acids. Mechanical embedding (ME) was used for all QM/MM models due to the large number of pathways investigated and the proven ability of this approach to accurately describe enzymatic reaction mechanisms,[56, 72, 89-92] while electrostatic embedding (EE) was also used for select QM/MM models (see the results and discussion section for additional details).

Each RC generated from the MD snapshots was fully optimized (including EndoV, DNA, water, and Mg^{2+}) using B3LYP-D3(BJ)/6-31G(d,p) to describe the QM region[93-94] and the AMBER force fields (OL15 for DNA and ff14SB for EndoV) to describe the MM region. Subsequently, scans were performed to identify possible concerted and stepwise

pathways for each model. Specifically, for concerted mechanisms, the distance between oxygen of the nucleophilic water and the electrophilic phosphorus of the scissile backbone was reduced by 0.10–0.15 Å stepwise. Subsequently, the P–O distance was frozen to that observed at the potential energy surface maximum, while the distance between the phosphorus and the O3′ leaving group was systematically increased (by 0.10–0.15 Å) to obtain an initial guess for the transition state (TS). For stepwise mechanisms, the distance between the phosphorus reaction centre and the oxygen of the water nucleophile in the QM/MM optimized RC was decreased (by 0.10–0.15 Å) to obtain a guess for TS1, while the distance between the O3′ leaving group and the phosphorous was elongated (by 0.10–0.15 Å) from the corresponding IC to obtain a guess for TS2. Each TS guess was then fully optimized followed by intrinsic reaction coordinate (IRC) calculations[95-96] to identify the associated RC, intermediate complex (IC), and/or PC. The natures of the optimized TS (single imaginary frequency) and corresponding minima were confirmed using frequency calculations at the same level of theory, which also gave the thermal corrections to the Gibbs energy. Finally, the relative Gibbs energies were calculated using single-point calculations at the ONIOM(M06-2X/6-311+G(2df,p):AMBER) level of theory. This choice of methodology was based on previous work that extensively investigated the impact of method on the description of one-metal mediated phosphodiester bond cleavage reactions facilitated by enzymes.[69-70] All calculations were performed using the Gaussian 16 program (revision B.01).[97]

To identify the preferred catalytic pathway, we compare each computed mechanism to experimental structural,[17] kinetic,[16] and mutational data.[47, 57] Specifically, computed geometries were compared to the crystal structure of the EndoV product complex (PDB ID:

2W35).[17] The calculated activation barriers were assessed against the experimental barrier (92.5 kJ/mol) for product generation in the dI-containing substrate, which was calculated using the apparent rate constant (0.1 min^{-1}) for *Salmonella typhimurium* EndoV measured at 37 °C, pH 7.4, 10 nM DNA substrate, and 10 nM of EndoV.[16] Finally, the predicted roles of active site amino acid residues were cross checked against experimental mutational data, including the 20% reduction in the catalytic activity upon introduction of the H214A mutation in bacterial EndoV[47] and the completely abolished catalytic activity upon mutation of the conserved K155 to alanine in the mouse homologue.[57]

Table 5.1. Summary of the phosphodiester bond cleavage pathways characterized in the present work and the reaction parameter-based criterion used to choose corresponding MD snapshots for subsequent QM/MM calculations.[a]

P–O bond cleavage pathway[b]	**General base**			**General acid**			**Reaction parameter-based criterion**
	H214	D43	Phosphate	Mg^{2+}	Mg^{2+}-ligated water	K139	
Direct (H214, Mg^{2+})	X			X			$r(N_{H214}\cdots P) < 6.5$ Å; $r(O3'\cdots Mg^{2+}) < 2.5$ Å
Direct (D43, Mg^{2+})		X		X			$r(O\delta_{D43}\cdots O_{\text{bridging water}}) < 3.4$ Å; $r(O3'\cdots Mg^{2+}) < 2.5$ Å
Direct (H214, Mg^{2+}-ligated water)	X				X		$r(N_{H214}\cdots P) < 6.5$ Å; $r(O3'\cdots H_w\text{–}O) < 3.4$ Å
Direct (Phosphate, Mg^{2+}-ligated water)			X		X		$r(O_{\text{non-bridging}}\cdots P) < 6.5$ Å; $r(O3'\cdots H_w\text{–}O) < 3.4$ Å
Indirect (H214, Mg^{2+}-ligated water)	X				X		$r(N_{H214}\cdots P) < 6.5$ Å; $r(O3'\cdots H_w\text{–}O) < 3.4$ Å
Indirect (Phosphate, Mg^{2+}-ligated water)			X		X		$r(O_{\text{non-bridging}}\cdots P) < 6.5$ Å; $r(O3'\cdots H_w\text{–}O) < 3.4$ Å
Indirect (H214, K139)	X					X	$r(N_{H214}\cdots P) < 6.5$ Å; $r(O3'\cdots H\text{–}N_{K139}) < 3.4$ Å
Indirect (Phosphate, K139)			X			X	$r(O_{\text{non-bridging}}\cdots P) < 6.5$ Å; $r(O3'\cdots H\text{–}N_{K139}) < 3.4$ Å

[a]For all pathways, a nucleophilic water was deemed posed for the reaction when $r(O_w\cdots P) < 4$ Å. [b]Nomenclature for pathways identifies the type of metal coordination (direct versus indirect) and the general base and acid in parentheses.

5.3. Results and Discussion

5.3.1. Despite Being Ubiquitous Among Two-Metal Dependent Endonucleases, Direct Bidentate Metal Ligation to the Substrate Renders EndoV Catalytically Inactive.

In the crystal structure of the PC for EndoV, the O3′ leaving group is directly coordinated to the metal (Figure 5.2a). Two-metal dependent enzymes typically form bidentate metal ligation to both the O3′ bridging oxygen and a non-bridging oxygen of the substrate (see, for example, the active sites of Rnase H (Figure 5.3f)[31] and *Aquifex aeolicus* Rnase III (Figure 5.3g)).[98] Furthermore, similar bidentate metal coordination to the substrate has been reported in the crystallized PC of other single-metal dependent nucleases, including APE1[49] (Figure 5.3a) and I-HmuI (Figure 5.3d),[53] as well as the crystallized RC of single-metal dependent I-*Ppo*I[39] (Figure 5.3c) and T4 endonuclease VII[51] (Figure 5.3e). However, metal ions have also been shown to move after the chemical step catalyzed by both one-metal and two-metal mediated endonucleases, highlighting that active site arrangements can be different before and after phosphodiester bond cleavage and suggesting the metal position in the crystal structure of PCs may be unreliable as a basis for modelling the RC.[31, 45, 48, 58-60] Based on this collection of structural data, a metal binding architecture that involves bidentate Mg^{2+} coordination to the O3′ leaving group and a non-bridging oxygen of the substrate in the RC was initially considered for EndoV (Figure D.3a, Appendix D).

Bidentate coordination of the metal to a non-bridging oxygen and O3′ of the substrate was observed in the post-equilibration RC for EndoV (Figure D.8, Appendix D). In this structure, the metal is 2.4 Å away from the O3′ leaving group and 2.0 Å from a non-bridging oxygen of the phosphate moiety. The remainder of the octahedral metal

coordination is fulfilled by two water molecules, D43, and D110. This metal coordination is maintained in the corresponding QM/MM optimized RC (Figure D.9a, Appendix D). Furthermore, H214 is suitably positioned to activate the nucleophilic water (within 1.871 Å), which is in line for attack on the scissile backbone ($r(O_w \cdots P)$ = 3.330 Å). Additional substrate charge stabilization is supplied by K139, which hydrogen bonds to the second non-bridging oxygen of the phosphate moiety ($r(O_{non\text{-}bridging} \cdots H_{K139})$ = 1.864 Å, Figure D.10a, Appendix D). From this RC, a concerted TS was identified (Figure 5.4a) in which a proton from the nucleophilic water is partially transferred to H214, a new P–O_w bond partially formed (1.936 Å), the P–O3′ bond lengthened (by 0.2 Å), and Mg^{2+} moves 0.3 Å closer to the leaving group. The TS is converted into a PC in which the P–O_w bond is fully formed, the P–O3′ bond cleaved, and the leaving group is more strongly coordinated to Mg^{2+} (Figure D.10a, Appendix D). Although K139 offers substrate charge stabilization early in the reaction through hydrogen bonding to a non-bridging oxygen of the phosphate moiety, a proton is fully transferred to the leaving group in the PC, which correlates with the significant role of K139 implied by experimental mutational data.[57] The roles of histidine in activating the water nucleophile, and the metal in substrate charge stabilization and aiding leaving group departure are consistent with the proposed mechanisms for other one-metal dependent endonucleases such as I-*Ppo*I (H98, Figure 5.3c),[39, 55] I-HmuI (H75, Figure 5.3d),[53] and T4 endonuclease VII (H41, Figure 5.3e).[51, 99] Nevertheless, the activation barrier associated with this EndoV pathway is too high (144.6 kJ/mol), falling over 50 kJ/mol above the experimental barrier,[16] and the reaction is not thermodynamically favorable (by 36.5 kJ/mol, Figures 5.4a and 5.5).

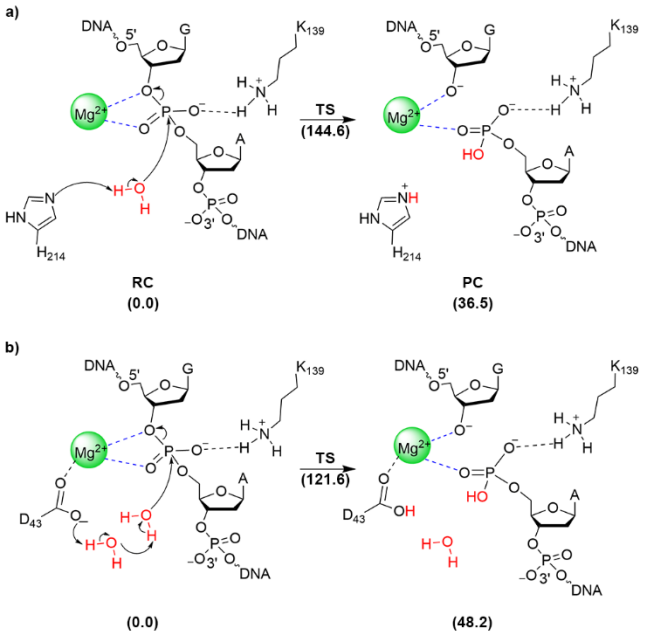

Figure 5.4. The mechanisms characterized in the present work for phosphodiester bond cleavage involving bidentate metal–substrate coordination, with Mg^{2+} aiding leaving group departure and a) H214 or b) D43 via a water chain activating the water nucleophile. Relative Gibbs energies (kJ/mol) are provided in parentheses.

Although H214 is aligned to initiate the reaction in post-equilibration models of the EndoV RC with bidentate metal coordination to the substrate, MD simulations reveal a second active site arrangement in which D43 is suitably oriented to accept a proton from the water nucleophile through a bridging water molecule (Figure D.6b, Appendix D). Nevertheless, the key reaction parameters are similar in the QM/MM RC regardless of the identity of the general base (i.e., metal coordination geometry, nucleophilic attack distance, P–O3′ distance, and K139 hydrogen-bonding distance to the substrate, Figures D.9b and D.10b, Appendix D). Support for a metal-ligated aspartate activating a water nucleophile is found in crystallographic and biochemical data for two-metal mediated RNase P and *Eco*RV.[100-101] Similar to the mechanism with H214 as the general base, QM/MM calculations reveal a S_N2 pathway (Figure 5.4b) in which proton abstraction from the

bridging water molecule by D43 and nucleophilic attack on phosphorus are coupled with a more advanced phosphodiester bond cleavage than the H214-mediated mechanism (P–O3′ bond is ~ 0.5 Å longer), while K139 hydrogen bonds with the O3′ leaving group (Figure D.10, Appendix D). D43-mediated phosphodiester bond hydrolysis results in a lower activation barrier than when H214 acts as the general base (by 23 kJ/mol, Figures 5.4b and 5.5), yet still significantly higher than the experimental barrier (by 29.1 kJ/mol).[16] This reaction pathway also leads to an endergonic PC (by 48.2 kJ/mol) despite the leaving group being more strongly coordinated to the metal (r(O3′···Mg^{2+}) is 0.2 Å shorter, Figure D.10, Appendix D).

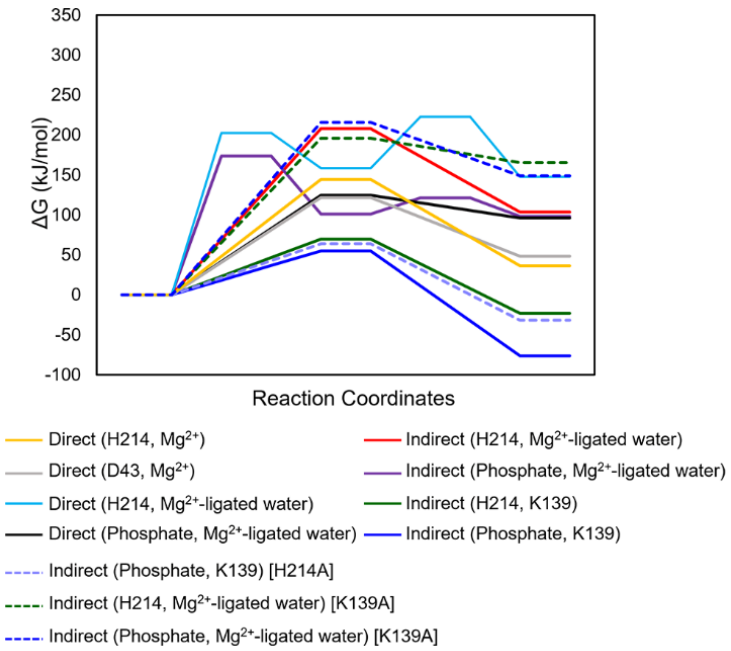

Figure 5.5. The relative Gibbs energies (kJ/mol) for the phosphodiester bond hydrolysis via each chemical pathway characterized in the present work facilitated by wild-type (solid lines) or *in silico* mutated (dashed lines) EndoV. Legend describes the type of metal coordination (direct versus indirect) and identifies the general base and acid in parentheses, and mutants in square brackets.

Overall, regardless of whether H214 or D43 activates the water nucleophile, the reaction barrier corresponding to bidentate metal coordination to the substrate is well above the experimental barrier[16] and the PCs are thermodynamically unfavorable (each by ~ 35–50 kJ/mol). Therefore, despite being proposed for other nucleases,[31, 39, 51, 53-55, 98-101] an EndoV catalytic pathway involving direct bidentate metal coordination to the substrate is unlikely.

5.3.2. Direct Metal Coordination to the Non-bridging Oxygen and Indirect Metal Coordination to the O3′ Leaving Group of the Substrate Results in Unfavorable Reaction Energetics Regardless of the Identity of the General Base.

When the structural dynamics are taken into account by performing MD simulations on the RC containing bidentate metal coordination to the substrate (Figure D.8, Appendix D), Appendix D, a water molecule typically moves between the metal and O3′ leaving group of the substrate despite maintenance of direct coordination to the non-bridging oxygen (Figure D.10a, Appendix D). The remainder of the octahedral Mg^{2+} coordination is fulfilled by D43, D110, and two additional active site water molecules (Figures D.11a and D.12a, Table D.2, Appendix D). Nevertheless, within this metal coordination geometry, the structural dynamics of the active site differentially align other residues with respect to the substrate to afford distinct possible catalytic pathways.

In the first active site conformation uncovered by MD simulations, H214 is well-positioned to activate the water nucleophile and a metal-ligated water is oriented to act as the general acid to stabilize the O3′ leaving group (Figure D.6c, Appendix D). Unlike the

previous mechanism involving same general base (H214) and bidentate metal–substrate coordination (Figure 5.4a), the nucleophilic water lies further from H214 and the phosphorus reaction center in the QM/MM RC ($r(N_{H214}\cdots H_w)$ and $r(O_w\cdots P)$ are each elongated by 0.3 Å, Figures D.9c, D.10a and D.13a, Appendix D). Unlike the first mechanism with the same general base (H214), the phosphodiester bond cleavage pathway follows a S_N1 mechanism when a metal-ligated water acts as the general acid (Figure 5.6a). In the first reaction step, deprotonation of the nucleophilic water proceeds through proton transfer to H214 ($r(N_{H214}\cdots H_w)$ = 1.226 Å) with simultaneous formation of a new P–O_w bond (1.957 Å, Figure D.13a, Appendix D) and a corresponding barrier of 202.4 kJ/mol (TS1, Figures 5.5 and 5.6a). TS1 evolves into a 44.1 kJ/mol more stable penta-coordinated phosphorane intermediate in which the substrate charge stabilization provided by the metal and K139 is maintained. In the second step, the phosphodiester bond starts to dissociate ($r(P\cdots O3' = 2.036$ Å, Figure D.13a, Appendix D) and the proton from the metal-ligated water is partially transferred to the O3′ leaving group. As expected for a weaker general acid, this leads to a rate-determining barrier (222.8 kJ/mol, Figures 5.5 and 5.6a) that is 78.2 kJ/mol higher than the mechanism involving the same base (H214) and direct metal–O3′ coordination. TS2 collapses into a highly unstable PC (by 147.6 kJ/mol) in which the leaving group is protonated by a metal activated water. Although an analogous pathway has been proposed for one-metal dependent I-*Ppo*I (Figure 5.3c),[39, 55] the large rate-determining barrier and thermodynamically unfavorable PC suggest this pathway is unlikely for EndoV.

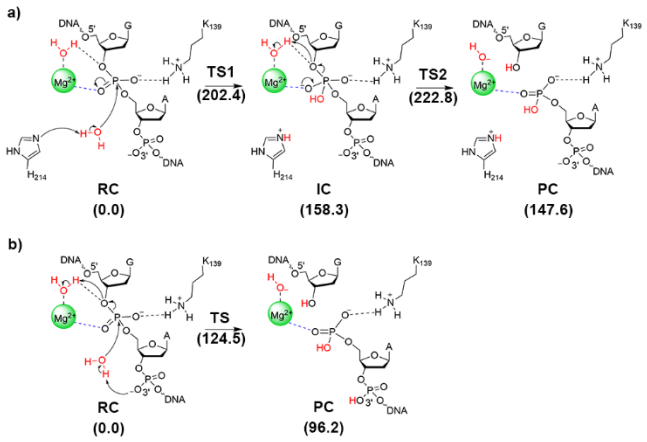

Figure 5.6. The mechanisms characterized in the present work for phosphodiester bond cleavage involving direct metal–substrate ligation to the non-bridging phosphate oxygen, with a metal-ligated water stabilizing the leaving group and a) H214 or b) a substrate phosphate activating the water nucleophile. Relative Gibbs energies (kJ/mol) are provided in parentheses.

During the MD simulations, a water molecule can also be suitably placed with respect to the phosphate moiety in the 3′ direction from the bond cleavage site for nucleophile activation (Figure D.6d, Appendix D). In the corresponding QM/MM RC, switching the general base from H214 to the substrate phosphate in the same metal coordination environment places the nucleophilic water in a more optimal orientation (Figure D.13, Appendix D). While the non-bridging oxygen of the scissile phosphate moiety is significantly more stabilized by K139, the O3′ leaving group is notably less stabilized by metal-ligated water compared to the H214-driven mechanism (Figure D.13, Appendix D). Unlike the corresponding two-step H214-activated mechanism, this reaction occurs in a single step (Figure 5.6b), with the attacking water molecule approaching phosphorus ($r(O_w \cdots P)$ = 1.801 Å) and releasing a proton to the downstream phosphate group in the TS, the P–O3′ bond simultaneously starting to break (2.519 Å), and a proton from a Mg^{2+}-ligated water being partially transferred to the leaving group (Figure D.13b,

Appendix D). This characterized EndoV mediated pathway parallels that reported in a QM/MM MD study of the two-metal mediated RNase H (Figure 5.3f).[31, 66] Additionally, X-ray crystallographic data for two-metal dependent BamHI led to the proposal that a water coordinated to Mn_B^{2+} protonates the leaving group, while the same metal also stabilizes the charge on the substrate through direct coordination (Figure D.1b, Appendix D).[102] In contrast, a QM/MM study on the two-metal mediated *Eco*RV concluded that a similar pathway is infeasible.[101] Although the EndoV pathway affords a significantly lower activation barrier than the reaction initiated by H214 (by 98.3 kJ/mol), the overall energetic requirement is still notably larger than that predicted by experiment (by approximately 32 kJ/mol)[16] and the corresponding PC is not thermodynamically favorable (by 96.2 kJ/mol, Figures 5.5 and 5.6b).

In summary, despite literature precedence for similar proposed reaction pathways for other nucleases,[31, 39, 55, 66, 101-102] EndoV-facilitated phosphodiester bond hydrolysis involving direct coordination to the non-bridging oxygen and water-mediated metal contacts with the O3′ leaving group is not catalytically feasible regardless of the identity of the general base.

5.3.3. Indirect Metal Coordination to the Substrate Leads to a Highly Unfavorable Reaction When a Mg^{2+}-ligated Water Stabilizes the Leaving Group

Although no catalytically feasible pathway could be characterized in the present work that involves direct metal coordination to the substrate, an active site configuration involving indirect metal coordination to the substrate has been observed in the

endonuclease literature.[41, 50, 56, 69] Specifically, a crystal structure of a phosphorothioate RC analogue of one-metal mediated APE1 shows a Mn^{2+}-coordinated water within hydrogen-bonding distance to a non-bridging oxygen of the scissile phosphate (PDB ID: 5DG0, Figure 5.3b).[50] Furthermore, QM calculations on APE1 revealed that indirect metal–substrate ligation is preferred over the direct substrate ligation observed in the crystal structure of the PC (PDB ID: 4IEM, Figure 5.3a), while QM/MM calculations verified this metal binding architecture can facilitate the phosphodiester bond hydrolysis in a manner consistent with experimental data.[56, 69] A similar water-mediated interaction between Mg^{2+} and a non-bridging phosphate oxygen has been detected in the active site of the RC analogue of one-metal dependent Vvn (PDB ID: 1OUP, Figure 5.3h).[41] Based on this literature, indirect metal–substrate coordination was investigated for EndoV using MD simulations in which the octahedral coordination geometry of Mg^{2+} was satisfied by D43, D110, and four active site water molecules (Figures D.11b and D.12b, and Table D.2, Appendix D). Throughout the MD simulations, the metal coordination was maintained and a metal-ligated water provides substrate stabilization by hydrogen bonding to a non-bridging phosphate oxygen (72% occupancy, Figure D.11b and Table D.2, Appendix D).

MD snapshots are present in which H214 is appropriately oriented to activate the nucleophile, while a metal-ligated water is positioned to assist leaving group departure through substrate protonation (Figure D.6e, Appendix D). When the mode of metal coordination to the substrate is switched from direct to indirect in the QM/MM-optimized RC (Figures D.9c and D.9e, Appendix D), a nucleophilic water molecule is better positioned for the reaction ($r(N_{H214}\cdots H_w)$ and $r(O_w\cdots P)$ are each 0.3 Å shorter, Figures D.13a and D.14a, Appendix D). Because of the loss of direct substrate stabilization by the

metal, substrate interactions with K139 become stronger ($r(O_{non\text{-}bridging}\cdots H_{K139})$ is 0.7 Å shorter), while a hydrogen bond with a metal-coordinated water is formed ($r(O\text{–}H_w\cdots O3')$ = 1.743 Å). Unlike the corresponding mechanism involving direct metal–substrate coordination (Figure 5.6a), this reaction follows a single-step mechanism (Figure 5.7a). In the TS, initiation of nucleophilic attack via proton release to H214 ($r(N_{H214}\cdots H_w)$ = 1.074 Å) is coupled with P–O3′ bond cleavage (2.313 Å), while proton transfer from a metal-ligated water to the substrate has yet to occur ($r(O\text{–}H_w\cdots O3')$ = 1.488 Å, Figure D.14a, Appendix D). Although the barrier for the concerted pathway is slightly lower than the barrier for direct metal–substrate ligation (by 15 kJ/mol, Figures 5.5 and 5.7a), and a similar reaction pathway was proposed for single-metal mediated Vvn based on X-ray crystallographic data (Figure 5.3h),[41] the prohibitively high activation barrier (115.3 kJ/mol larger than the experimental barrier)[16] and thermodynamically unfavorable PC (by 103.4 kJ/mol, Figure 5.7a) make this pathway infeasible for bacterial EndoV.

During MD simulations on the RC with indirect metal coordination, a water molecule can also be well-positioned for proton abstraction by the substrate phosphate moiety in the 3′ direction from the scissile phosphate (Figure D.6f, Appendix D). In the corresponding QM/MM RC (Figure D.9f, Appendix D), the water molecule remains optimally placed with respect to the phosphorus reaction center as seen in the corresponding RC involving direct metal ligation (Figures D.13b and D.14b, Appendix D). A Mg^{2+}-coordinated water forms an interaction with a non-bridging oxygen of the scissile phosphate moiety ($r(O\text{–}H_w\cdots O_{non\text{-}bridging})$ = 1.864 Å), while a second metal-ligated water forms a strong hydrogen bond with the leaving group ($r(O\text{–}H_w\cdots O3')$ = 1.704 Å). K139 hydrogen bonds with and has already transferred a proton to a non-bridging oxygen of the

scissile phosphate moiety in the RC directly connected to the first TS for this pathway. Changing the mode of metal coordination from direct to indirect in the phosphate-activated mechanism switches the pathway from a concerted to a stepwise mechanism (Figure 5.7b). In the first step, nucleophilic attack of water at the phosphate ($r(O_w \cdots P)$ = 1.995 Å, Figure D.14b, Appendix D) leads to a high rate-determining barrier (173.6 kJ/mol, TS1, Figures 5.5 and 5.7b). This results in a phosphorane intermediate that is 72.6 kJ/mol more stable than TS1. In the second reaction step, the phosphodiester bond cleavage simultaneously occurs with substrate protonation by a metal-ligated water (Figure D.14b, Appendix D). The barrier is similar to the phosphate-driven mechanism involving direct metal ligation (within 3.1 kJ/mol, Figure 5.5), while the endergonic PC is within 1.9 kJ/mol.

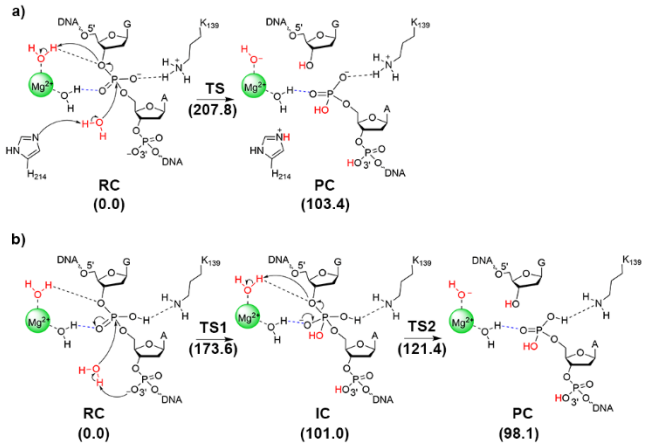

Figure 5.7. The mechanisms characterized in the present work for phosphodiester bond cleavage involving indirect metal–substrate ligation to the non-bridging phosphate oxygen, with a metal-ligated water stabilizing the leaving group and a) H214 or b) a substrate phosphate activating the water nucleophile. Relative Gibbs energies (kJ/mol) are provided in parentheses.

Overall, although phosphodiester bond cleavage pathways involving indirect metal coordination to the substrate have been observed for other endonucleases,[41, 50, 56] a metal-

ligated water cannot effectively stabilize the leaving group for bacterial EndoV regardless of the general base considered.

5.3.4. A Catalytically Feasible Reaction Pathway Occurs When an Active Site Lysine Plays the Major Role of Leaving Group Activation, While a Metal-ligated Water Provides Critical Transition State Stabilization, Regardless of the Route of Water Nucleophile Activation.

The previous sections have underscored that chemical mechanisms in which Mg^{2+} promotes leaving group departure through either direct or indirect (water-mediated) coordination are not catalytically feasible for EndoV regardless of the identity of the general base (H214, the substrate phosphate moiety, or D43) or mode of substrate charge stabilization (direct or indirect metal coordination, or K139). This suggests that there must be another residue that fulfills the role of the general acid. Indeed, MD simulations reveal a new active site conformation in which K139 hydrogen bonds to the O3′ leaving group regardless of the metal–substrate coordination geometry (Figures D.6g–h, Appendix D). A lysine is part of a conserved PD-(D/E)XK motif in the active site of several endonucleases including *Eco*RI,[103] *Eco*RV,[104] PVuII,[105] FokI,[106] and BglI.[107] Although experimental mutational data also suggests that lysine is important for the enzymatic activity of *EcoR*V,[108] PvuII endonuclease,[109] and mouse EndoV,[57] the catalytic role of lysine has not been computationally explored for an endonuclease to date to the best of our knowledge.

Multiple MD frames exist where Mg^{2+} is directly coordinated to the substrate, H214 is properly positioned to activate the water nucleophile, and K139 is hydrogen bonding to

the O3′ leaving group. However, when attempting to characterize the TS from this starting point, the direct metal coordination reverts to indirect coordination, hinting that indirect coordination is preferred. Indeed, MD simulations on the RC involving indirect metal ligation reveal a catalytically plausible active site arrangement in which K139 hydrogen bonds to O3′ and H214 is appropriately placed to activate the water nucleophile (Figure D.6g, Appendix D). When metal-ligated water is substituted by K139 as the general acid, the structural parameters are similar in the QM/MM-optimized RCs, with the substrate stabilization from a Mg^{2+}-ligated water molecule being strengthened ($r(Mg–O_w\cdots O_{non\text{-}bridging})$ decreases by ~ 0.2 Å, Figures D.14a and D.15a, Appendix D). Although the reaction continues to follow a S_N2 mechanism (Figure 5.8a), the reaction parameters in the concerted TS are significantly different. Specifically, despite the approach of the water nucleophile to the phosphorus reaction center being more advanced for K139-mediated reaction, water has not donated a proton to H214 (Figures D.14a and D.15a, Appendix D). In contrast, K139 has almost completely donated a proton to the O3′ leaving group, while the P–O3′ bond cleavage is at the early stage of the reaction. This leads to a significantly lower energy barrier (by 138.0 kJ/mol, Figures 5.5, 5.7a and 5.8a) and the phosphodiester bond dissociation results in the first exergonic PC (by 126.4 kJ/mol). In the PC, the newly formed O–H hydrogen bonds to K139 ($O3'–H\cdots N_{K139}$ = 1.768 Å), while the P–O bond interacts with protonated H214 and a metal-ligated water continues to provide substrate stabilization (Figures D.14a and D.15a, Appendix D).

As the activation barrier (69.8 kJ/mol) is less than the experimental barrier for product generation (92.5 kJ/mol)[16] and an exergonic PC is formed, the catalytic pathway characterized that involves K139 as a general acid is viable for EndoV. Furthermore, the

active site configuration is similar to that observed in the crystal structure of the PC (Figure D.16a, Appendix D). Indeed, although our proposed mechanism requires a slight change in the orientation of K139 compared to the crystal structure of the PC, there is a small (~ 1 Å) difference in heavy atom distance between K139 and the non-bridging oxygen versus the O3′-bridging oxygen of the scissile phosphate in the crystal structure (PDB ID: 2W35).[17] MD simulations verify that the flexible lysine side chain can readily reorientate in EndoV, and active site rearrangements are common to yield the final nuclease products.[35, 45, 48, 58-60] Our proposed pathway also parallels previous enhanced sampling MD simulations and free energy calculations on RNA-dependent RNA polymerase (RdRP), where K551 is not near the leaving group in the crystal structure, but is proposed to form a strong hydrogen bond with and transfer a proton to the leaving pyrophosphate group during product formation.[67]

To ensure that proton transfer from lysine to the substrate is not an artifact of using mechanical embedding to describe the QM/MM boundary, the mechanisms with H214 acting as the general base and a metal-ligated water or K139 as the general acid were reconsidered using electrostatic embedding (Figure D.17, Appendix D). The proton transfer from lysine to the substrate occurs later in both reaction pathways with EE, which results in more endergonic PCs that are directly connected to the TSs. Nevertheless, the final product has yet to form in any reaction pathway investigated in the present work as this likely involves significant active site reorganization coupled with metal migration as discussed for other nucleases.[35, 45, 48, 58-60] Furthermore, the proton from K139 transfers to the O3′-leaving group in the PC under the EE scheme, and other structural features are highly similar along the ME and EE pathways for each mechanism (Figure D.17, Appendix

D). Therefore, there are minimal changes to the predicted barriers upon re-optimization with EE (< 14% change, Figure D.17, Appendix D). Most importantly, the mechanism in which a metal ligated water promotes leaving group departure remains energetically unfavorable, while the mechanism in which K139 stabilizes the O3′-leaving group is feasible, verifying that ME effectively describes the phosphodiester bond hydrolysis reaction considered in this work.

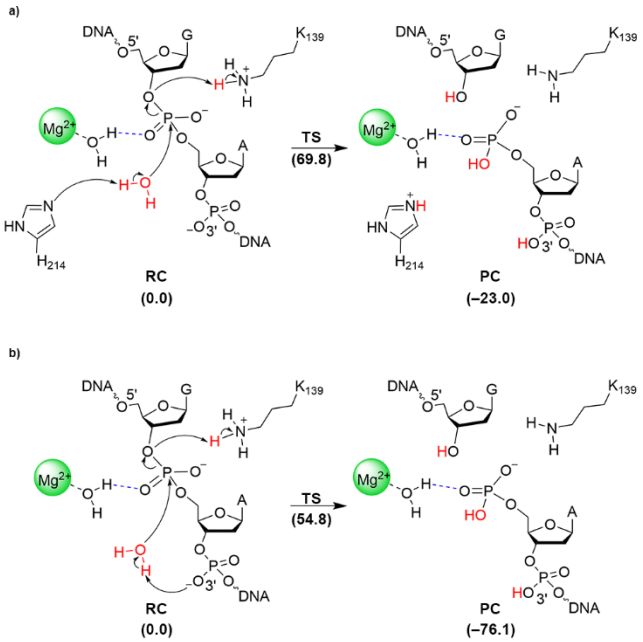

Figure 5.8. The mechanisms characterized in the present work for phosphodiester bond cleavage involving indirect metal–substrate ligation to the non-bridging phosphate oxygen, with K139 stabilizing the leaving group and a) H214 or b) a substrate phosphate activating the water nucleophile. Relative Gibbs energies (kJ/mol) are provided in parentheses.

As seen for the mechanisms characterized that involve a metal-ligated water acting as the general acid, an alternative general base is possible when K139 stabilizes the leaving group. Specifically, in a second active site arrangement observed during MD simulations on the RC in which K139 is positioned close to the O3′ leaving group, a potential

nucleophilic water is oriented toward the 3′-phosphate moiety of the substrate (Figure D.6h, Appendix D). In the corresponding QM/MM-optimized RC, the same water is suitably positioned for in-line attack on the scissile phosphate while being in proximity to the downstream phosphate moiety (Figures D.9h and D.15b, Appendix D). Unlike the previous K139-mediated mechanism with H214 as the base, the leaving group forms a stronger interaction with K139 in the RC ($r(O3'\cdots H_{K139})$ is 0.5 Å shorter), while the substrate charge stabilization from a Mg^{2+}-ligated water molecule is maintained (Figure D.15, Appendix D). Similar to the H214-activated pathway, a S_N2 mechanism was characterized (Figure 5.8b) in which a proton from the nucleophile has not yet transferred to the phosphate group, but water has initiated attack at the phosphorus (Figure D.15, Appendix D). Cleavage of the P–O3′ bond is more advanced in the TS when the substrate phosphate acts as the general base (by ~ 0.3 Å), while K139 protonates the leaving group. As discussed for the H214-mediated mechanism, the activation barrier is catalytically feasible (54.8 kJ/mol, Figures 5.5 and 5.8b) and an exergonic PC (by 76.1 kJ/mol) is obtained that has a similar active site configuration to that observed in the crystal structure of the PC (Figure D.16b, Appendix D), further underscoring the viability of K139 as the general acid for the reaction. Substrate assisted nucleophile activation parallels the proposal reported for one-metal containing *Eco*RI (Figure 5.3j), where a significant reduction in the catalytic activity was observed upon modifying the pro-Rp phosphoryl oxygen of the substrate to H-phosphonate.[52] Although the activation barrier is lower when the substrate phosphate group rather than H214 acts as the general base when the metal indirectly coordinates to the substrate, the difference (15 kJ/mol) is minor and the activation barrier of both catalytic pathways is less than the experimental barrier for product generation (92.5 kJ/mol),[16] suggesting that both pathways are feasible. The possibility of alternate mechanisms correlates with the H214A

EndoV mutation reducing enzymatic activity by only 20%.[47] Furthermore, similar to our proposal, a QM/MM-MD study on the two-metal mediated CRISPR-Cas9 uncovered two alternate mechanisms involving H983 or a substrate phosphate as the general base that fall within the statistical errors of the calculations.[63]

Overall, our data suggest that EndoV can use a unique combination of residues to cleave the phosphodiester bond, namely lysine (K139) stabilizes the leaving group, either H214 or a substrate phosphate moiety activates the nucleophilic water, and the metal plays the important role of stabilizing the transition state formed during the reaction. The metal is also critical for maintaining the correct active site conformation and rigidity to promote the reaction. Although five of the eight MD snapshots used to characterize different chemical pathways in the present work have K139 in the same orientation as the crystal structure (i.e., within hydrogen-bonding distance of a non-bridging phosphate),[17] none of these mechanisms are energetically feasible, which suggests that K139 plays a more active role in the reaction. Indeed, our proposal that K139 is the general acid correlates with experimental mutation of lysine abolishing the catalytic activity of the enzyme in the mouse homologue.[57] Although lysine is rarely observed as a general acid, the proposal that lysine acts as the general acid in the P–O bond cleavage reaction is not completely unprecedented. Specifically, there is experimental evidence that directly confirms the general acid role of a lysine residue in P–O bond cleavage in other metallonucleases (K532 in human topoisomerase I[110] and K167 in vaccinia topoisomerase)[111] as well as computational evidence for the role of lysine as an acid in P–O bond formation by polymerases (K551 in RdRP).[67]

5.3.5. Large Catalytic Impact of *in silico* K139 Mutation Correlates With the Observed Complete Loss of Enzymatic Activity Upon K155 Mutation in the Mouse Homologue of EndoV.

As mentioned previously, the proposed role of K139 as the general acid in the mechanism of action of EndoV correlates with mutational data from the mouse homologue of EndoV, which suggests complete loss of catalytic activity upon K155A mutation,[57] as well as the conservation of this residue across the EndoV family.[14] To further explore the role of lysine in our proposed EndoV catalytic pathways, K139 was mutated to alanine and the reaction pathways were remapped (Figure 5.9a–b). A single-step pathway is characterized for the K139A EndoV mutant regardless of the general base. Although most structural parameters in the RC and PC of wild-type EndoV are preserved upon K139A mutation when H214 activates the water nucleophile, the transition state for the phosphodiester bond cleavage is more advanced for mutant EndoV ($r(O_w \cdots P)$ is 0.5 Å shorter and $r(P \cdots O3')$ is 0.4 Å longer, Figures D.15a and D.18a, Appendix D). Nevertheless, the loss of substrate charge stabilization by K139 raises the activation barrier by 126.0 kJ/mol (Figures 5.5, 5.8a and 5.9a) and the PC becomes endergonic (by 165.3 kJ/mol). When the substrate phosphate acts as the general base in the K139A EndoV mutant (Figure 5.9b), the nucleophilic water falls further from the 3′-phosphate group in the RC than for the corresponding wild-type mechanism (Figures D.15b and D.18b, Appendix D), resulting in a late transition state for nucleophilic attack ($r(O_w \cdots P)$ is reduced by ~ 0.6 Å) and a significantly higher barrier (by 161.1 kJ/mol, Figures 5.5, 5.8b and 5.9b). The TS is then converted into an endergonic PC (unstable by 149.1 kJ/mol). Overall, the predicted energetically unfavorable pathways upon the introduction of the K139A mutation correlates

with the experimental observation that the K155A mutation leads to complete loss of activity for the mouse homologue of EndoV[57] and supports our proposed EndoV reaction mechanisms in which a lysine (K139) facilitates leaving group departure.

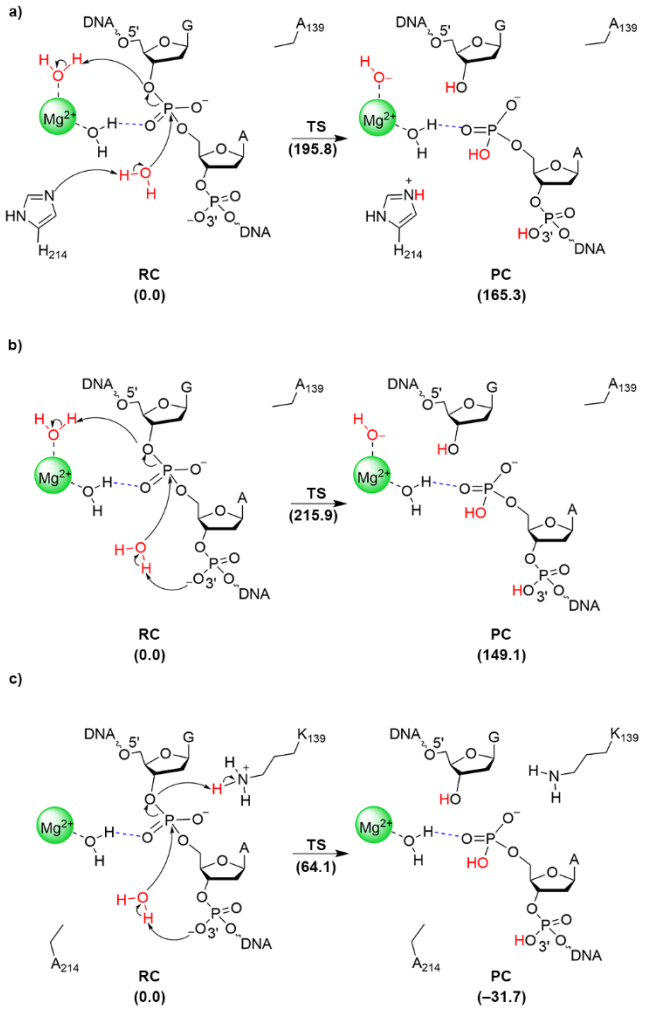

Figure 5.9. The mechanisms characterized in the present work for phosphodiester bond cleavage by the K139A EndoV mutant involving a metal-ligated water stabilizing the leaving group, and a) H214 or b) a substrate phosphate activating the water nucleophile, and c) the H214A mutant involving K139 stabilizing the leaving group and a substrate phosphate activating the water nucleophile. Relative Gibbs energies (kJ/mol) are provided in parentheses.

5.3.6. Small Catalytic Impact of *in silico* H214 Mutation Correlates With a Reported Modest Reduction in Catalytic Activity for the H214A EndoV Mutant.

To further explore our proposed reaction pathways and the experimentally-observed impact of the H214A mutation on EndoV catalytic activity, H214 was mutated to alanine. Since 80% the wild-type catalytic activity is retained in the H214A mutant,[47] it is a reasonable assumption that the active site conformation does not significantly change. Therefore, H214 was directly mutated to alanine in the QM/MM model of the wild-type phosphate-mediated mechanism. Upon introduction of the H214A mutation, active site residues maintain their relative positions with respect to the substrate in the QM/MM optimized RC (Figure D.19a, Appendix D). However, due to the loss of a hydrogen-bonding interaction with H214 (Figure D.19b, Appendix D), the nucleophilic water lies further from the 3′-downstream phosphate group upon removal of H214 ($r(O_{non\text{-}bridging}\cdots H_w)$ is ~ 2 Å longer, Figures D.15b and D.18c, Appendix D). Nevertheless, the mutant RC evolves into a concerted TS (Figure 5.9a) that has very similar structural parameters of the wild-type variant, including the key reaction coordinates ($r(O_w\cdots P)$ is within 0.03 Å and $r(P\cdots O3')$ is within 0.06 Å, Figures D.15b and D.18c, Appendix D). As a result, introduction of the H214A mutation raises the activation barrier by only 9.3 kJ/mol and leads to a 44.4 kJ/mol less exergonic PC that has similar structural parameters (Figures 5.5, 5.8b and 5.9c). Despite helping to align the water nucleophile when not acting as the general base (Figure D.19b, Appendix D), our calculations further highlight that H214 is not critical for the reaction. Indeed, our *in silico* mutational data correlates with the H214A EndoV mutant maintaining 80% of the catalytic activity of the wild-type enzyme.[47]

5.3.7. EndoV Uses a Different Catalytic Pathway Than Proposed for Other Single-Metal Dependent Endonucleases, Yet Further Supports That One-Metal is Enough for Catalysis of Phosphodiester Bond Cleavage.

From the multiple catalytically-active conformations observed during MD simulations, eight distinct pathways for EndoV-facilitated phosphodiester bond cleavage were characterized using QM/MM. In terms of the general base, although crystallographic and biochemical data have revealed a metal-ligated aspartate activates the nucleophilic water for two-metal mediated RNase P (D399)[100] and *Eco*RV (D90),[101] a pathway in which metal-ligated aspartate (D43) activates the nucleophile is not feasible for EndoV. On the other hand, similar to proposals based on crystallographic and biochemical studies that histidine is the general base for one-metal dependent I-*Ppo*I (H98),[39] Vvn (H80),[41] and T4 endonuclease VII (H41),[51] EndoV can use H214 to activate the water nucleophile. However, mutational data indicates these other enzymes exhibit complete loss of catalytic activity upon mutation of the respective histidine,[41, 55, 99, 112] while the H214A bacterial EndoV mutant maintains 80% of the wild-type activity.[47] Indeed, a second pathway involving nucleophile activation by a substrate phosphate moiety 3′ with respect to the cleavage site is feasible for EndoV. A similar substrate-phosphate-mediated mechanism has been suggested based on an experimental study on one-metal dependent *Eco*RI[52] and QM/MM-MD calculations on two-metal dependent RNase H.[66] The possibility of two alternate pathways for EndoV correlates with the activity of the H214A mutant and parallels a proposal from QM/MM-MD that two-metal dependent CRISPR-Cas9 can employ H983 or a substrate phosphate as the general base.[63]

In terms of facilitating leaving group departure, direct coordination of the metal to the substrate has been suggested to stabilize O3′ of the substrate based on X-ray crystallographic studies of single-metal containing I-*Ppo*I[39] and T4EndoVII,[51] and two-metal-containing RNase P[100] and AaRNase III.[98] However, the active site composition of EndoV appears to prevent the metal from efficiently functioning in this capacity. On the other hand, a metal-ligated water has been suggested to protonate the leaving group in two-metal mediated *Eco*RV (QM/MM study)[101] and BamHI (X-ray crystallographic study).[102] A metal-activated water has also been proposed to be the general acid for one-metal mediated APE1 (QM/MM study).[56] Nevertheless, our calculations suggest the corresponding pathway is infeasible for EndoV, emphasizing that different single-metal dependent nucleases function in unique ways due to different active site compositions. Although lysine has not been previously proposed to act as the general acid for reactions catalyzed by endonucleases, DFT/MM-MD calculations suggest that lysine (K551) stabilizes the pyrophosphate leaving group through protonation to facilitate the phosphodiester bond formation catalyzed by a two-metal dependent RNA polymerase.[67] We hypothesize that EndoV similarly uses lysine (K139) as the general acid to promote P–O bond cleavage. When coupled with the ability of EndoV to use either H214 or a substrate phosphate moiety as the general base, our work highlights a novel catalytic pathway for bacterial EndoV to achieve the same phosphodiester bond hydrolysis facilitated by other endonucleases through diverse mechanisms.

The number of metal cofactors required to cleave the backbone of a nucleic acid has been a controversial issue at least in part due to the discrepancy between X-ray crystallographic and solution studies. As one specific example, the crystal structure of

*Eco*RV reveals two metals in the active site,[36] while kinetic data suggests only one metal is essential for catalysis.[46] More broadly, while structural data originally lead to proposals that most endonucleases use a two-metal mediated pathway,[34-35] experimental kinetic study of the metal concentration dependence of enzymatic activity revealed that some enzymes use a single metal for catalysis.[37, 45, 47-48] The proposal that one metal is sufficient for the chemical step is further sustained by the emergence of crystal structures containing a single metal in the nuclease active site.[39, 41, 49-51, 53-54, 113] The present work further corroborates the previous experimental proposals that one metal is enough to cleave the P–O bond in nucleic acids. Indeed, the single metal used by EndoV is in a similar position with respect to the substrate in our calculated structures as the second metal (Mg_B^{2+}) in two-metal mediated endonucleases (Figure D.20, Appendix D), and the role of the first metal (Mg_A^{2+}) in nucleophile activation can be fulfilled by an amino acid (histidine) or the substrate.[48, 59, 114] This idea that EndoV uses a single metal for the catalytic step correlates with kinetic data[47] and the single metal present in the active site in the crystal structure of the enzyme.[17] This work presents only the second example of an endonuclease for which computational methods demonstrate the feasibility of single-metal mediated catalysis.

The multiple pathways compared and contrasted in the present work permitted the prediction of unique chemistry, which highlights that diversity in the active site framework allows different endonucleases to achieve the same challenging phosphodiester backbone cleavage in nucleic acids in different ways. Indeed, despite high sequence conservation among bacterial and mouse EndoV (Figure D.21, Appendix D), time-lapse X-ray crystallography snapshots suggest that mouse EndoV follows a two-metal mediated mechanism,[57] while our calculations reveal that bacterial EndoV follows a single-metal

mediated mechanism for P–O bond cleavage. Sequence alignment of the two enzymes shows that the composition of the active site varies by a single amino acid (H214 in bacterial EndoV versus D240 in mouse EndoV, Figure D.21, Appendix D). This comparison further underscores how a slight difference in the identity of amino acids in the active site of a metalloenzyme can change the required number of metals (or perhaps even a different role of the metals) to facilitate the phosphodiester bond hydrolysis.

5.4. Conclusion

This is the first computational study on an enzyme from the EndoV family that explored different possibilities for the metal binding architecture, the general base, and the general acid for bacterial EndoV, including combinations that have been previously proposed for other nucleases. Specifically, eight different potential catalytic pathways were characterized using a multipronged MD and QM/MM approach. Although the metal can sufficiently stabilize the leaving group either directly or indirectly through a coordinated water for other enzymes, a similar mechanism is unlikely for EndoV. Instead, K139 was identified as the general acid that promotes leaving group departure, while substrate charge stabilization is provided by a metal-coordinated water. This proposed role for K139 is supported by experimental data indicating the K155A mutation leads to complete loss of catalytic activity for the mouse homologue of EndoV.[57] Furthermore, the phosphodiester bond cleavage is energetically unfavorable for the *in silico* K139A mutant. Interestingly, either H214 or a 3′-substrate phosphate can initiate the reaction by activating the water nucleophile, which correlates with experimental data suggesting that the H214A mutant maintains 80% of the catalytic activity of the wild-type enzyme.[47] This previously unseen

combination of general base, general acid, and metal-binding architecture is fully consistent with experimental kinetic,[16] structural,[17] and mutational data[47, 57] and underscores the novelty of the catalytic pathways used by bacterial EndoV to facilitate the same phosphodiester bond hydrolysis as other endonucleases. In the nuclease literature where the role and number of metal ions required for phosphodiester bond cleavage are highly controversial issues, the present work contributes towards the idea that one metal is capable of catalyzing this difficult bond cleavage reaction. The fundamental understanding about the catalytic function of EndoV afforded by the present work will promote exploration of new and improved applications of bacterial EndoV in biotechnology and medicine, including the detection of disease inducing gene mutations or the design of therapeutic solutions for ailments such as neurological disorders and carcinomas.

5.5. References

(1) De Bont, R.; van Larebeke, N., Endogenous DNA damage in humans: A review of quantitative data. *Mutagenesis* **2004,** *19* (3), 169–185.

(2) Gates, K. S., An overview of chemical processes that damage cellular DNA: spontaneous hydrolysis, alkylation, and reactions with radicals. *Chem. Res. Toxicol.* **2009,** *22* (11), 1747-1760.

(3) Farajollahi, S.; Maas, S., Molecular Diversity through RNA Editing: A Balancing Act. *Trends Genet.* **2010,** *26* (5), 221–230.

(4) Alseth, I.; Dalhus, B.; Bjørås, M., Inosine in DNA and RNA. *Curr. Opin. Genet. Dev.* **2014,** *26*, 116–123.

(5) Kow, Y. W., Repair of deaminated bases in DNA. *Free Radic. Biol. Med.* **2002,** *33* (7), 886-893.

(6) Kuraoka, I., Diversity of endonuclease V: from DNA repair to RNA editing. *Biomolecules* **2015,** *5* (4), 2194–2206.

(7) Silvestris, D. A.; Picardi, E.; Cesarini, V.; Fosso, B.; Mangraviti, N.; Massimi, L.; Martini, M.; Pesole, G.; Locatelli, F.; Gallo, A., Dynamic inosinome profiles reveal novel patient stratification and gender-specific differences in glioblastoma. *Genome Biol.* **2019,** *20* (1), 1–18.

(8) Kong, X. Y.; Huse, C.; Yang, K.; Øgaard, J.; Berges, N.; Vik, E. S.; Nawaz, M. S.; Quiles-Jiménez, A.; Abbas, A.; Gregersen, I., Endonuclease V regulates atherosclerosis through C-C motif chemokine ligand 2-mediated monocyte infiltration. *J. Am. Heart Assoc.* **2021,** *10* (14), e020656.

(9) Srivastava, P. K.; Bagnati, M.; Delahaye-Duriez, A.; Ko, J.-H.; Rotival, M.; Langley, S. R.; Shkura, K.; Mazzuferi, M.; Danis, B.; van Eyll, J., Genome-wide analysis of differential RNA editing in epilepsy. *Genome Res.* **2017,** *27* (3), 440–450.

(10) Hosaka, T.; Tsuji, H.; Kwak, S., RNA editing: a new therapeutic target in amyotrophic lateral sclerosis and other neurological diseases. *Int. J. Mol. Sci.* **2021,** *22* (20), 10958.

(11) Tran, S. S.; Jun, H.-I.; Bahn, J. H.; Azghadi, A.; Ramaswami, G.; Van Nostrand, E. L.; Nguyen, T. B.; Hsiao, Y.-H. E.; Lee, C.; Pratt, G. A., Widespread RNA editing dysregulation in brains from autistic individuals. *Nat. Neurosci.* **2019,** *22* (1), 25–36.

(12) Khermesh, K.; D'Erchia, A. M.; Barak, M.; Annese, A.; Wachtel, C.; Levanon, E. Y.; Picardi, E.; Eisenberg, E., Reduced levels of protein recoding by A-to-I RNA editing in Alzheimer's disease. *RNA* **2016,** *22* (2), 290–302.

(13) Breen, M. S.; Dobbyn, A.; Li, Q.; Roussos, P.; Hoffman, G. E.; Stahl, E.; Chess, A.; Sklar, P.; Li, J. B.; Devlin, B., Global landscape and genetic regulation of RNA editing in cortical samples from individuals with schizophrenia. *Nat. Neurosci.* **2019,** *22* (9), 1402–1412.

(14) Cao, W., Endonuclease V: an unusual enzyme for repair of DNA deamination. *Cell. Mol. Life Sci.* **2013,** *70* (17), 3145–3156.

(15) Huang, J.; Lu, J.; Barany, F.; Cao, W., Multiple cleavage activities of endonuclease V from Thermotoga maritima: recognition and strand nicking mechanism. *Biochemistry* **2001,** *40* (30), 8738–8748.

(16) Feng, H.; Klutz, A. M.; Cao, W., Active site plasticity of endonuclease V from Salmonella typhimurium. *Biochemistry* **2005,** *44* (2), 675–683.

(17) Dalhus, B.; Arvai, A. S.; Rosnes, I.; Olsen, Ø. E.; Backe, P. H.; Alseth, I.; Gao, H.; Cao, W.; Tainer, J. A.; Bjørås, M., Structures of endonuclease V with DNA reveal initiation of deaminated adenine repair. *Nat. Struct. Mol. Biol.* **2009,** *16*, 138–143.

(18) Dalhus, B.; Alseth, I.; Bjørås, M., Structural basis for incision at deaminated adenines in DNA and RNA by endonuclease V. *Prog. Biophys. Mol. Biol.* **2015,** *117* (2-3), 134–142.

(19) Hitchcock, T. M.; Gao, H.; Cao, W., Cleavage of deoxyoxanosine-containing oligodeoxyribonucleotides by bacterial endonuclease V. *Nucleic Acids Res.* **2004,** *32* (13), 4071–4080.

(20) Chang, Y.; Huang, Z.; Quan, H.; Li, H.; Yang, S.; Song, Y.; Wang, J.; Yuan, J.; Wu, C., Construction of a DNA damage repair gene signature for predicting prognosis and immune response in breast cancer. *Front. Oncol.* **2022,** *12*, 1085632.

(21) Huang, J.; Kirk, B.; Favis, R.; Soussi, T.; Paty, P.; Cao, W.; Barany, F., An endonuclease/ligase based mutation scanning method especially suited for analysis of neoplastic tissue. *Oncogene* **2002,** *21* (12), 1909–1921.

(22) Pincas, H.; Pingle, M. R.; Huang, J.; Lao, K.; Paty, P. B.; Friedman, A. M.; Barany, F., High sensitivity EndoV mutation scanning through real-time ligase proofreading. *Nucleic Acids Res.* **2004,** *32* (19), e148.

(23) 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.

(24) Louie, M.; Louie, L.; Simor, A. E., The role of DNA amplification technology in the diagnosis of infectious diseases. *Can. Med. Assoc. J.* **2000,** *163* (3), 301–309.

(25) Cockerill III, F. R., Genetic methods for assessing antimicrobial resistance. *Antimicrob. Agents Chemother.* **1999,** *43* (2), 199–212.

(26) Knutson, S. D.; Arthur, R. A.; Johnston, H. R.; Heemstra, J. M., Direct immunodetection of global A-to-I RNA editing activity with a chemiluminescent bioassay. *Angew. Chem. Int. Ed.* **2021,** *60* (31), 17009–17017.

(27) Knutson, S. D.; Arthur, R. A.; Johnston, H. R.; Heemstra, J. M., Selective enrichment of A-to-I edited transcripts from cellular RNA using Endonuclease V. *J. Am. Chem. Soc.* **2020,** *142* (11), 5241–5251.

(28) Kong, X. Y.; Vik, E. S.; Nawaz, M. S.; Berges, N.; Dahl, T. B.; Vågbø, C.; Suganthan, R.; Segers, F.; Holm, S.; Quiles-Jiménez, A., Deletion of Endonuclease V suppresses chemically induced hepatocellular carcinoma. *Nucleic Acids Res.* **2020,** *48* (8), 4463–4479.

(29) Miyazaki, K., Random DNA fragmentation with endonuclease V: application to DNA shuffling. *Nucleic Acids Res.* **2002,** *30* (24), e139.

(30) Wang, Z.; Wang, H.-Y.; Feng, H., A simple and reproducible method for directed evolution: combination of random mutation with dITP and DNA fragmentation with endonuclease V. *Mol. Biotechnol.* **2013,** *53*, 49–54.

(31) Nowotny, M.; Gaidamakov, S. A.; Crouch, R. J.; Yang, W., Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell* **2005,** *121* (7), 1005–1016.

(32) Tsutakawa, Susan E.; Classen, S.; Chapados, Brian R.; Arvai, A. S.; Finger, L. D.; Guenther, G.; Tomlinson, Christopher G.; Thompson, P.; Sarker, Altaf H.; Shen, B.; Cooper, Priscilla K.; Grasby, Jane A.; Tainer, John A., Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily. *Cell* **2011,** *145* (2), 198–211.

(33) 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.

(34) 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.

(35) Yang, W., Nucleases: diversity of structure, function and mechanism. *Q. Rev. Biophys.* **2011,** *44* (1), 1–93.

(36) Horton, N. C.; Perona, J. J., DNA cleavage by *Eco*RV endonuclease: two metal ions in three metal ion binding sites. *Biochemistry* **2004,** *43* (22), 6841–6857.

(37) 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 *Eco*RI family. *J. Mol. Biol.* **2009,** *393* (1), 140–160.

(38) Mol, C. D.; Izumi, T.; Mitra, S.; Tainer, J. A., DNA-bound structures and mutants reveal abasic DNA binding by APE1 DNA repair and coordination. *Nature* **2000,** *403* (6768), 451–456.

(39) Galburt, E. A.; Chevalier, B.; Tang, W.; Jurica, M. S.; Flick, K. E.; Monnat, R. J.; Stoddard, B. L., A novel endonuclease mechanism directly visualized for I-*Ppo*I. *Nat. Struct. Biol.* **1999,** *6* (12), 1096–1099.

(40) Ghosh, M.; Meiss, G.; Pingoud, A.; London, R. E.; Pedersen, L. C., Structural insights into the mechanism of nuclease A, a ββα metal nuclease from Anabaena. *J. Biol. Chem.* **2005,** *280* (30), 27990–27997.

(41) 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.

(42) Raaijmakers, H.; Törö, I.; Birkenbihl, R.; Kemper, B.; Suck, D., Conformational flexibility in T4 endonuclease VII revealed by crystallography: implications for substrate binding and cleavage. *J. Mol. Biol.* **2001,** *308* (2), 311–323.

(43) Vasu, K.; Nagamalleswari, E.; Zahran, M.; Imhof, P.; Xu, S.-y.; Zhu, Z.; Chan, S.-H.; Nagaraja, V., Increasing cleavage specificity and activity of restriction endonuclease KpnI. *Nucleic Acids Res.* **2013,** *41* (21), 9812–9824.

(44) Belkebir, A.; Azeddoug, H., Metal ion dependence of DNA cleavage by SepMI and EhoI restriction endonucleases. *Microbiol. Res.* **2013,** *168* (2), 99–105.

(45) Xie, F.; Qureshi, S. H.; Papadakos, G. A.; Dupureur, C. M., One- and two-metal ion catalysis: global single-turnover kinetic analysis of the PvuII endonuclease mechanism. *Biochemistry* **2008,** *47* (47), 12540–12550.

(46) Groll, D. H.; Jeltsch, A.; Selent, U.; Pingoud, A., Does the restriction endonuclease Eco RV employ a two-metal-ion mechanism for DNA cleavage? *Biochemistry* **1997,** *36* (38), 11389-11401.

(47) Feng, H.; Dong, L.; Cao, W., Catalytic mechanism of endonuclease V: a catalytic and regulatory two-metal model. *Biochemistry* **2006,** *45* (34), 10251–10259.

(48) 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.

(49) Tsutakawa, S. E.; Shin, D. S.; Mol, C. D.; Izumi, T.; Arvai, A. S.; Mantha, A. K.; Szczesny, B.; Ivanov, I. N.; Hosfield, D. J.; Maiti, B.; Pique, M. E.; Frankel, K. A.; Hitomi, K.; Cunningham, R. P.; Mitra, S.; Tainer, J. A., Conserved structural chemistry for incision activity in structurally non-homologous apurinic/apyrimidinic endonuclease APE1 and endonuclease IV DNA repair enzymes. *J. Biol. Chem.* **2013,** *288* (12), 8445–8455.

(50) Freudenthal, B. D.; Beard, W. A.; Cuneo, M. J.; Dyrkheeva, N. S.; Wilson, S. H., Capturing snapshots of APE1 processing DNA damage. *Nat. Struct. Mol. Biol.* **2015,** *22* (11), 924–931.

(51) Biertümpfel, C.; Yang, W.; Suck, D., Crystal structure of T4 endonuclease VII resolving a Holliday junction. *Nature* **2007,** *449* (7162), 616–620.

(52) Jeltsch, A.; Alves, J.; Wolfes, H.; Maass, G.; Pingoud, A., Substrate-assisted catalysis in the cleavage of DNA by the EcoRI and EcoRV restriction enzymes. *Proc. Natl. Acad. Sci. U.S.A.* **1993,** *90* (18), 8499–8503.

(53) Shen, B. W.; Landthaler, M.; Shub, D. A.; Stoddard, B. L., DNA binding and cleavage by the HNH homing endonuclease I-HmuI. *J. Mol. Biol.* **2004,** *342* (1), 43–56.

(54) Flick, K. E.; Jurica, M. S.; Monnat, R. J.; Stoddard, B. L., DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-*Ppo*I. *Nature* **1998,** *394* (6688), 96–101.

(55) Mannino, S. J.; Jenkins, C. L.; Raines, R. T., Chemical mechanism of DNA cleavage by the homing endonuclease I-*Ppo*I. *Biochemistry* **1999,** *38* (49), 16178–16186.

(56) 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.

(57) 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.

(58) Dupureur, C. M., Roles of metal ions in nucleases. *Curr. Opin. Chem. Biol.* **2008,** *12* (2), 250–255.

(59) 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 APE1. *Proteins* **2007,** *68* (1), 313–323.

(60) He, H.; Chen, Q.; Georgiadis, M. M., High-resolution crystal structures reveal plasticity in the metal binding site of apurinic/apyrimidinic endonuclease 1. *Biochemistry* **2014,** *53* (41), 6520–6529.

(61) 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.

(62) Leclerc, F.; Karplus, M., Two-metal-ion mechanism for hammerhead-ribozyme catalysis. *J. Phys. Chem. B* **2006,** *110* (7), 3395–3409.

(63) Casalino, L.; Nierzwicki, L.; 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.

(64) 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.

(65) Rosta, E.; Yang, W.; Hummer, G., Calcium inhibition of ribonuclease H1 two-metal ion catalysis. *J. Am. Chem. Soc.* **2014,** *136* (8), 3137–3144.

(66) 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.

(67) Bignon, E.; Monari, A., Modeling the enzymatic mechanism of the SARS-CoV-2 RNA-dependent RNA polymerase by DFT/MM-MD: an unusual active site leading to high replication rates. *J. Chem. Inf. Model.* **2022,** *62* (17), 4261–4269.

(68) 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.

(69) 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.

(70) 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.

(71) Wilson, K. A.; Fernandes, P. A.; Ramos, M. J.; Wetmore, S. D., Exploring the identity of the general base for a DNA polymerase catalyzed reaction using QM/MM: the case study of human translesion synthesis polymerase η. *ACS Catal.* **2019,** *9* (3), 2543–2551.

(72) Garcia-Borràs, M.; Kan, S. J.; Lewis, R. D.; Tang, A.; Jimenez-Osés, G.; Arnold, F. H.; Houk, K. N., Origin and control of chemoselectivity in cytochrome c catalyzed carbene transfer into Si–H and N–H bonds. *J. Am. Chem. Soc.* **2021,** *143* (18), 7114–7123.

(73) Alonso-Cotchico, L.; Rodríguez-Guerra, J.; Lledós, A.; Maréchal, J.-D., Molecular modeling for artificial metalloenzyme design and optimization. *Acc. Chem. Res.* **2020,** *53* (4), 896-905.

(74) Hu, Q.; Jayasinghe-Arachchige, V. M.; Sharma, G.; Serafim, L. F.; Paul, T. J.; Prabhakar, R., Mechanisms of peptide and phosphoester hydrolysis catalyzed by two promiscuous metalloenzymes (insulin degrading enzyme and glycerophosphodiesterase) and their synthetic analogues. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2020,** *10* (4), e1466.

(75) Geronimo, I.; Vidossich, P.; Donati, E.; De Vivo, M., Computational investigations of polymerase enzymes: Structure, function, inhibition, and biotechnology. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2021,** *11* (6), e1534.

(76) Himo, F., Recent trends in quantum chemical modeling of enzymatic reactions. *J. Am. Chem. Soc.* **2017,** *139* (20), 6780–6786.

(77) Rostkowski, M.; Olsson, M. H. M.; Søndergaard, C. R.; Jensen, J. H., Graphical analysis of pH-dependent properties of proteins predicted using PROPKA. *BMC Struct. Biol.* **2011,** *11* (1), 1–6.

(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.

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

(80) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case," Development and testing of a general amber force field" *J. Comput. Chem.* **2004,** *25* (9), 1157–1174.

(81) 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.

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

(83) Allnér, O.; Nilsson, L.; Villa, A., Magnesium ion–water coordination and exchange in biomolecular simulations. *J. Chem. Theory Comput.* **2012,** *8* (4), 1493–1502.

(84) Case, D.; Ben-Shalom, I.; Brozell, S.; Cerutti, D.; Cheatham III, T.; Cruzeiro, V.; Darden, T.; Duke, R.; Ghoreishi, D.; Gilson, M., AMBER 2018; 2018. *University of California, San Francisco* **2018**.

(85) Elsässer, B.; Fels, G., Atomistic details of the associative phosphodiester cleavage in human ribonuclease H. *Phys. Chem. Chem. Phys.* **2010,** *12* (36), 11081–11088.

(86) 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.

(87) Ni, Z.; Jin, X.; Zhou, P.; Wu, Q.; Lin, X. F., A combination of computational and experimental approaches to investigate the binding behavior of B. sub lipase A mutants with substrate pNPP. *Mol. Inform.* **2011,** *30* (4), 359–367.

(88) Yanagita, H.; Urano, E.; Matsumoto, K.; Ichikawa, R.; Takaesu, Y.; Ogata, M.; Murakami, T.; Wu, H.; Chiba, J.; Komano, J., Structural and biochemical study on the inhibitory activity of derivatives of 5-nitro-furan-2-carboxylic acid for RNase H function of HIV-1 reverse transcriptase. *Bioorg. Med. Chem.* **2011,** *19* (2), 816–825.

(89) Sebera, J.; Hattori, Y.; Sato, D.; Reha, D.; Nencka, R.; Kohno, T.; Kojima, C.; Tanaka, Y.; Sychrovsky, 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.

(90) Brás, N. r. F.; Santos-Martins, D.; Fernandes, P. A.; Ramos, M. J., Mechanistic pathway on human α-glucosidase maltase-glucoamylase unveiled by QM/MM calculations. *J. Phys. Chem. B* **2018,** *122* (14), 3889–3899.

(91) Fernandes, H. S.; Ramos, M. J.; Cerqueira, N. M., Catalytic mechanism of the serine hydroxymethyltransferase: a computational ONIOM QM/MM study. *ACS Catal.* **2018,** *8* (11), 10096–10110.

(92) Tolbert, A. E.; Ervin, C. S.; Ruckthong, L.; Paul, T. J.; Jayasinghe-Arachchige, V. M.; Neupane, K. P.; Stuckey, J. A.; Prabhakar, R.; Pecoraro, V. L., Heteromeric three-stranded coiled coils designed using a Pb (ii)(Cys) 3 template mediated strategy. *Nat. Chem.* **2020,** *12* (4), 405–411.

(93) Becke, A. D.; Johnson, E. R., A density-functional model of the dispersion interaction. *J. Chem. Phys.* **2005,** *123* (15), 154101.

(94) Grimme, S.; Ehrlich, S.; Goerigk, L., Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011,** *32* (7), 1456–1465.

(95) Fukui, K., The path of chemical reactions - the IRC approach. *Acc. Chem. Res.* **1981,** *14* (12), 363–368.

(96) Dykstra, C.; Frenking, G.; Kim, K.; Scuseria, G., Theory and applications of computational chemistry: the first forty years. Elsevier: 2011.

(97) Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Petersson, G.; Nakatsuji, H., Gaussian 16 Rev. B. 01, Wallingford, CT. **2016**.

(98) 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.* **2007,** *67* (1), 143–154.

(99) Giraud-Panis, M.-J. E.; Lilley, D. M., T4 endonuclease VII: importance of a histidine-aspartate cluster within the zinc-binding domain. *J. Biol. Chem.* **1996,** *271* (51), 33148–33155.

(100) Howard, M. J.; Klemm, B. P.; Fierke, C. A., Mechanistic studies reveal similar catalytic strategies for phosphodiester bond hydrolysis by protein-only and RNA-dependent ribonuclease P. *J. Biol. Chem.* **2015,** *290* (21), 13454–13464.

(101) Imhof, P.; Fischer, S.; Smith, J. C., Catalytic mechanism of DNA backbone cleavage by the restriction enzyme *Eco*RV: a quantum mechanical/molecular mechanical analysis. *Biochemistry* **2009,** *48* (38), 9061–9075.

(102) Viadiu, H.; Aggarwal, A. K., The role of metals in catalysis by the restriction endonuclease Bam HI. *Nat. Struct. Biol.* **1998,** *5* (10), 910–916.

(103) Pingoud, A.; Jeltsch, A., Structure and function of type II restriction endonucleases. *Nucleic Acids Res.* **2001,** *29* (18), 3705–3727.

(104) Hiller, D. A.; Rodriguez, A. M.; Perona, J. J., Non-cognate enzyme–DNA complex: structural and kinetic analysis of EcoRV endonuclease bound to the EcoRI recognition site GAATTC. *J. Mol. Biol.* **2005,** *354* (1), 121–136.

(105) Horton, J. R.; Cheng, X., PvuII endonuclease contains two calcium ions in active sites. *J. Mol. Biol.* **2000,** *300* (5), 1049–1056.

(106) Wah, D. A.; Hirsch, J. A.; Dorner, L. F.; Schildkraut, I.; Aggarwal, A. K., Structure of the multimodular endonuclease FokI bound to DNA. *Nature* **1997,** *388* (6637), 97–100.

(107) Newman, M.; Lunnen, K.; Wilson, G.; Greci, J.; Schildkraut, I.; Phillips, S. E., Crystal structure of restriction endonuclease BglI bound to its interrupted DNA recognition sequence. *EMBO J.* **1998,** *17* (18), 5466–5476.

(108) 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.

(109) Nastri, H. G.; Evans, P. D.; Walker, I. H.; Riggs, P. D., Catalytic and DNA binding properties of PvuII restriction endonuclease mutants. *J. Biol. Chem.* **1997,** *272* (41), 25761–25767.

(110) Interthal, H.; Quigley, P. M.; Hol, W. G.; Champoux, J. J., The role of lysine 532 in the catalytic mechanism of human topoisomerase I. *J. Biol. Chem.* **2004,** *279* (4), 2984–2992.

(111) Krogh, B. O.; Shuman, S., Catalytic mechanism of DNA topoisomerase IB. *Mol. Cell* **2000,** *5* (6), 1035–1041.

(112) Friedhoff, P.; Kolmes, B.; Gimadutdinow, O.; Wende, W.; Krause, K. L.; Pingoud, A., Analysis of the mechanism of the *Serratia* nuclease using site-directed mutagenesis. *Nucleic Acids Res.* **1996,** *24* (14), 2632–2639.

(113) Shlyapnikov, S.; Lunin, V.; Perbandt, M.; Polyakov, K.; Lunin, V. Y.; Levdikov, V.; Betzel, C.; Mikhailov, A., Atomic structure of the Serratia marcescens endonuclease at 1.1

Å resolution and the enzyme reaction mechanism. *Acta Crystallogr. D* **2000,** *56* (5), 567–572.

(114) Yang, W., An equivalent metal ion in one- and two-metal-ion catalysis. *Nat. Struct. Mol. Biol.* **2008,** *15* (11), 1228–1231.

Chapter 6: Mechanism of Nucleic Acid Phosphodiester Bond Cleavage by Human Endonuclease V: MD and QM/MM Calculations Reveal a Versatile Metal Dependence

Preface: Chapter 6 includes a currently unpublished manuscript [Kaur, R.; Wetmore, S. D., [Mechanism of Nucleic Acid Phosphodiester Bond Cleavage by Human Endonuclease V: MD and QM/MM Calculations Reveal a Versatile Metal Dependence. **2024**, *ACS Cat.* Submitted (manuscript ID: cs-2024-01060f)]. In this chapter, both authors contributed to the reviewing and editing of the manuscript. I aided conceptualization of the project, performed MD simulations, and characterized total 4 phosphodiester bond cleavage pathways. In addition, I performed the overall data/results analysis, wrote the first draft of the manuscript, and generated all figures. S. D. Wetmore was involved in project conceptualization and administration, supervision, funding and resource acquisition, data visualization and interpretation, and writing the original draft of the manuscript.

6.1. Introduction

Inosine is a key post-transcriptional modification that arises in RNA through various pathways.[1] For example, spontaneous hydrolysis of adenosine (A) or exposure of A to nitrosative stress caused by superoxides and nitric oxide during inflammation can result in inosine.[1-3] Alternatively, imbalanced purine nucleotide metabolism can lead to unexpected inosine triphosphates incorporation into mRNA during transcription, which can result in mutant proteins and inhibit cell viability.[4-5] Although adenosine deaminases also yield inosine as a part of RNA editing to control gene integrity and gene expression,[6-7]

hyperedited RNA can occur, resulting in codon changes and thereby altering protein translation.[1] In fact, although balanced adenosine-to-inosine editing is necessary for generating RNA and protein diversity,[8] dysregulated adenosine-to-inosine editing has been connected to 17 different cancers, ranging from colorectal to breast invasive carcinoma,[9-10] as well as other diseases, such as epilepsy,[11] amyotrophic lateral sclerosis (ALS),[12] autism,[13] Alzheimer's,[14] atherosclerosis,[15] and schizophrenia.[16] Therefore, the amount of inosine in RNA must be carefully regulated.

Endonuclease V (EndoV) is a conserved family of enzymes (Figure 6.1a) that removes deaminated nucleobases, including inosine, by cleaving phosphodiester bonds in nucleic acids.[17-19] Despite highly conserved enzymes generally having similar functions, the bacterial homologue of EndoV is involved in DNA repair,[20] while eukaryotic EndoV is involved in RNA metabolism.[18] EndoV has also been implicated in the development, detection, and treatment of human diseases.[15, 21-27] For example, overexpression of catalytically-inactive EndoV increases sensitivity towards hepatocellular carcinoma treatments.[22] Furthermore, eukaryotic EndoV is overexpressed in carotid atherosclerotic lesions, with enzyme inhibition reducing both plaque size and severity of ischemic strokes.[15] Finally, a genome-wide study of psychiatric disorders demonstrated a significant association between genes that encode EndoV and schizophrenia.[21] To better understand the cellular biochemistry and further enhance disease management, the molecular level details of the mechanism of action used by members of the EndoV family must be uncovered.

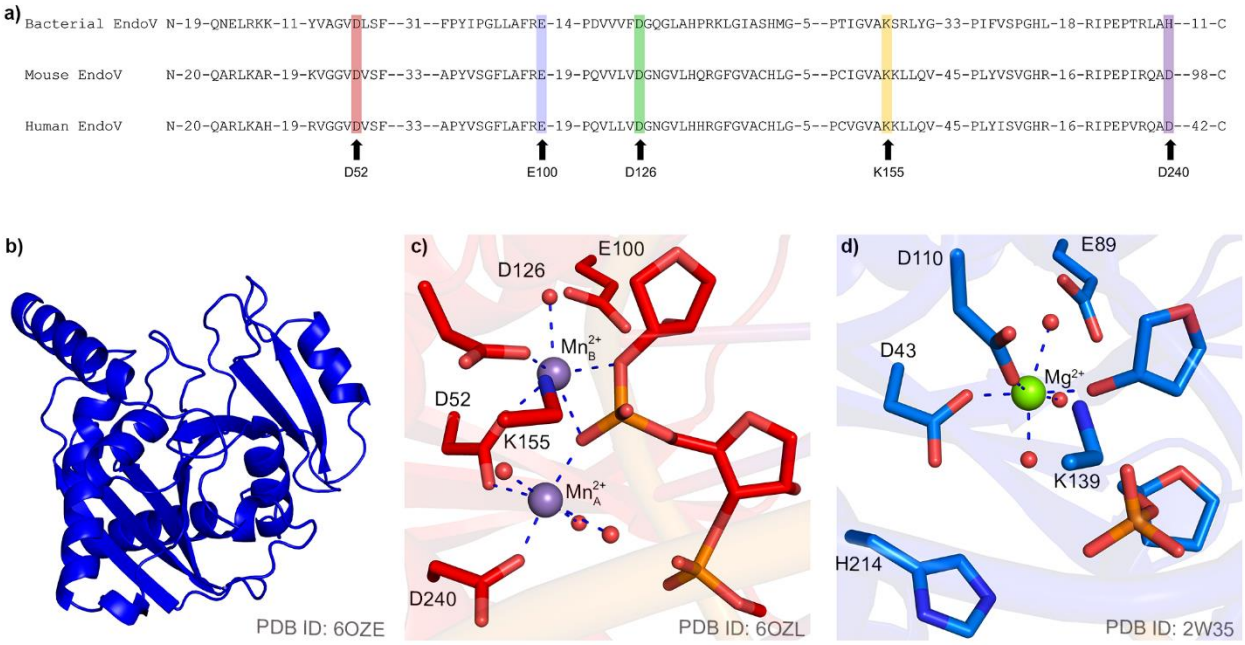

Figure 6.1. a) Sequence alignment of bacterial, mouse, and human EndoV, with key active site amino acids highlighted. The species and respective GenBank accession numbers used for the alignment are *Thermotoga maritima* and Q9X2H9 for bacterial EndoV, *Mus musculus* and Q8C9A2 for mouse EndoV, and *Homo sapiens* and Q8N8Q3 for human EndoV. The X-ray crystal structure of the b) apo-human EndoV, c) active site for the RC of wild-type mouse EndoV bound to a dsRNA substrate, and d) active site for the PC of wild-type bacterial EndoV bound to a ssDNA substrate.

Human EndoV catalyzes the Mg^{2+}-dependent cleavage of the second phosphodiester bond on the 3′-side of inosine to generate the 3′-hydroxy and 5′-phosphate products.[28] However, very few details are available about the corresponding mechanism of action. The role and number of metal/s required to cleave the phosphodiester bond are unknown, with no experimental or computational study having investigated the metal dependence of human EndoV activity to date. In fact, the currently available crystal structure of human EndoV (PDB ID: 6OZE, Figure 6.1b) does not contain the substrate or metal cofactor/s.[28] Furthermore, mutational studies have highlighted the roles of only a few residues, with D52A and Y91A mutations abolishing catalytic activity.[17-18] Comparisons to mouse and bacterial EndoV suggest that Y91 likely occupies the empty space remaining in

DNA following base flipping upon substrate binding,[29] while D52 is likely involved in metal binding in the active site.[28] Due to the scarcity of information, there has not yet been a proposal for the human EndoV phosphodiester bond cleavage pathway.

The prediction of the human EndoV mechanism of action is further complicated by the fact that enzymes from different species have been concluded to use a different number of metals to catalyze the phosphodiester bond cleavage despite high sequence similarity (Figure 6.1a). For mouse EndoV, X-ray crystallographic data reveal two metals in the active site (Figure 6.1c).[28] Indeed, the most widely accepted mechanism for nuclease catalysis based on available structural data involves two metals.[28, 30-50] The first metal (M_A^{2+}) activates the water nucleophile for attack at the scissile phosphate, the second metal (M_B^{2+}) promotes leaving group departure through direct or water-mediated coordination, and both metals provide substrate charge stabilization by coordinating to the non-bridging oxygen of the scissile phosphate (Figure 6.2a).[51] However, the roles and stoichiometry of the metal ions needed for nuclease catalysis has been a controversial issue for almost two decades,[52-66] with proposals for the two-metal mediated mechanism based solely on structural data having been called into question. Indeed, metal concentration dependencies in solution suggest that only one metal may be essential for catalysis by some enzymes despite some endonucleases having two metal-binding sites.[34, 38, 56, 67-68]

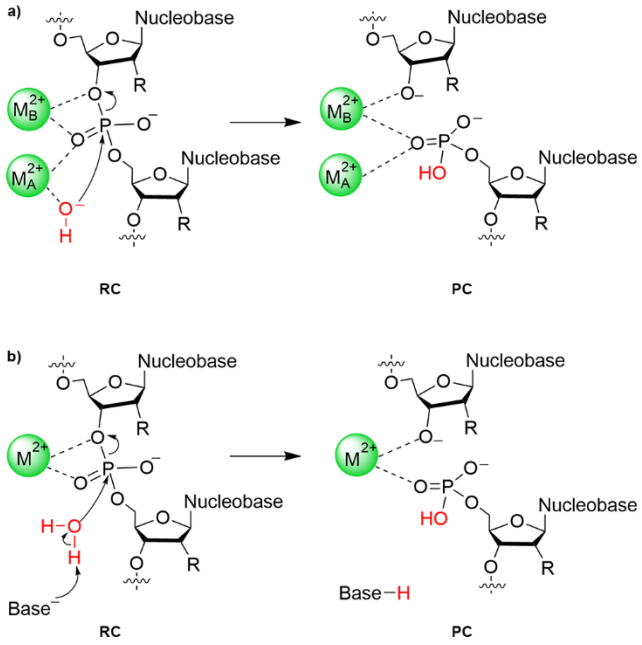

Figure 6.2. The general proposed mechanism for a) two-metal or b) single-metal mediated phosphodiester bond cleavage by an endonuclease (M^{2+} = metal, and R = H for DNA or OH for RNA).

In contrast to mouse EndoV,[28] bacterial EndoV has been proposed to use a single metal to facilitate P–O bond cleavage,[62, 69] at least in part due to the presence of only one metal in the active site in the X-ray crystal structure of the product complex (PC) of the wild-type enzyme bound to a single-stranded DNA (ssDNA) substrate (Figure 6.1d). In single-metal dependent nucleases, the function of the first metal in nucleophile activation in the two-metal mediated pathway (M_A^{2+}) has been proposed to be fulfilled by an active site amino acid (Figure 6.2b),[51, 54-55] such as H214 in bacterial EndoV (Figure 6.1d).[62] According to sequence alignment, human EndoV contains an Asp residue (D240) in an equivalent position as H214 in bacterial EndoV (Figure 6.1a). The suggestion that human EndoV may only require one metal is further supported by other single-metal dependent nucleases that have been proposed to efficiently use an Asp residue coupled with one metal to achieve the P–O bond cleavage, such as APE1 (D210, Figure 6.3a).[57, 63-64] Although an

abundance of experimental structural and kinetic studies have proposed that several nucleases use a single metal to cleave the phosphodiester bond,[29, 70-75] computational studies investigating the catalytic mechanisms used by one-metal mediated nucleases are relatively scarce.[57, 62-64]

To address gaps in the EndoV literature, the present work has used a multi-scale computational approach to decipher the general mechanism of action of human EndoV. Initially, since the only available crystal structure does not contain a substrate or metal cofactor/s (Figure 6.1b),[28] the first model of the human EndoV–substrate complex was constructed. Directed based on existing literature on other nucleases,[28, 32, 70, 75-76] including EndoV from other species,[28-29] molecular dynamics (MD) simulations were then performed in triplicate on complexes with different numbers of active site metals and various metal–substrate binding architectures to gain insights into the structural dynamics of human EndoV. Multiple MD snapshots were chosen that correspond to potentially catalytically-active conformations with differentially aligned active sites as starting points for quantum mechanics-molecular mechanics (QM/MM) calculations within the ONIOM formalism. By characterizing four unique pathways with different numbers of metals (one versus two) and modes of metal–substrate coordination (direct versus indirect), we provide the first atomic-level details of the mechanism of action of human EndoV. Our calculations reveal a rare metal dependence for this enzyme and highlight that nucleases may not all strictly follow traditional one or two-metal dependent mechanisms. As this is only the second computational study of a member of the EndoV family,[62] our work enhances our knowledge of how this broad family of enzymes process nucleic acids in cells. The fundamental details of the chemistry facilitated by human EndoV will be advantageous for future work geared toward further understanding the central role of this enzyme in human disease (e.g.,

cardiovascular disease,[15] cancer,[22] and psychiatric disorders[21]), developing new therapeutics that target EndoV,[15, 21-22] and advancing applications of EndoV in biotechnology.[25-27, 77-79]

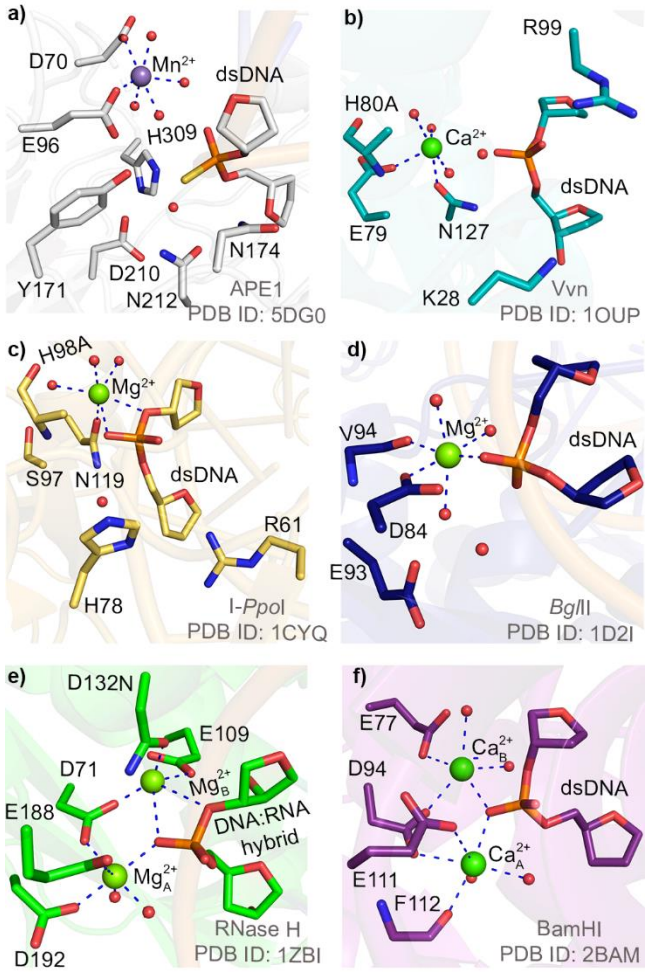

Figure 6.3. X-ray crystal structures of the active site in the RC for: a) APE1 bound to a thio-substituted substrate, b) the H80A mutant of Vvn, c) the H98A mutant of I-*Ppo*I, d) wild-type *Bgl*II, e) the D132N mutant of RNase H, and f) wild-type BamHI.

6.2. Computational Methodology

6.2.1. Model Building

Although the only available crystal structure of human EndoV is of the apoenzyme in the absence of metal cofactor/s (PDB ID: 6OZE),[28] the crystal structure of mouse EndoV contains a double-stranded RNA (dsRNA) substrate and metal ions (PDB ID: 6OZO).[28] Since the protein sequences of mouse and human EndoV, including the active sites, are highly similar (81% sequence identity, Figure 6.1a), a model of the human EndoV–substrate complex was obtained by superimposing the crystal structures with respect to protein heavy atoms (root-mean-square deviation (RMSD) = 1.029 Å, Figure E.1a, Appendix E) and deleting mouse EndoV, while retaining the substrate, two Mn^{2+} ions, and five directly coordinated water molecules. The five unpaired terminal nucleotides from the 5′-side of the inosine-containing RNA strand and 11 nucleotides on the 3′-side of the complementary strand were removed due to their distal location from the active site and their lack of interactions with the enzyme. Mn^{2+} ions were replaced with Mg^{2+} and four mutations in human EndoV (C140A, C225S, C226A, and C228S) were reverted using PyMOL 2.5.[80] PROPKA was used to assign protonation states to titratable amino acids at pH 7.0.[81] Three models with different metal-binding configurations found in either one or two metal-dependent nucleases (Figure 6.3)[28, 32, 70, 75-76] were then generated: 1) Mg_A^{2+} was removed and the remaining Mg^{2+} adopted an octahedral coordination geometry involving D52, D126, and four water molecules, while the metal formed water-mediated contacts with O3′ and a non-bridging oxygen of the scissile phosphate (denoted indirect Mg^{2+}–substrate coordination, Figure E.2a, Appendix E); 2) Mg_A^{2+} was removed and the remaining Mg^{2+} was directly coordinated to both O3′ and a non-bridging oxygen of the scissile

phosphate, as well as D52, D126, and two water molecules (denoted direct Mg^{2+}–substrate coordination, Figure E.2b, Appendix E); and 3) two metals were retained, which are bridged by D52 and a non-bridging oxygen of the scissile phosphate (Figure E.2c, Appendix E), while Mg_A^{2+} was also coordinated to D240 and three water molecules, and Mg_B^{2+} was also coordinated to D126, two water molecules, and O3′ of the scissile phosphate.

6.2.2. MD Simulations

Throughout the MD simulations, the protein and RNA were described using the AMBER ff14SB[82] and ffbsc0χOL3 force fields,[83-84] respectively. Parameters for inosine were supplemented by the generalized AMBER force field (GAFF)[85] using ANTECHAMBER 17.3,[86] while the Mg^{2+} parameters developed by Allnér *et al*.[87] were adopted. Atomic partial charges for inosine were determined using the restrained electrostatic potential (RESP) from the R.E.D.v.III.4 program.[88] The tLEaP module of AMBER18[89] was used to neutralize the system with 22 Na^+ ions, and add 16 Na^+ and $Cl^–$ ions to reach a physiological salt concentration (150 mM), as determined using the SLTCAP calculator.[90] The entire enzyme–RNA complex was solvated in a TIP3P truncated octahedral water box, while maintaining a minimum distance of 10.0 Å between the solute and box faces.

Each system was minimized using 2500 steps of the steepest descent and 2500 steps of the conjugate gradient method in each minimization round and applying a 100 kcal mol^{-1} $Å^{-2}$ force constant to the restrained portions. The first minimization step relaxed the solvent, while restraining the solute. The second step minimized the solute hydrogen atoms,

while restraining the rest of the system. The next minimization step was divided into 9 rounds to enhance substrate binding while maintaining octahedral metal coordination by gradually reducing the solute restraints based on the distance from the active site metal/s (starting with 20 Å sphere and ending at 4 Å in 2 Å increments). Subsequently, the full solute was minimized while restraining the solvent. In the final minimization step, the entire model was minimized with no restraints.

After minimization, the system was heated from 10 to 310 K in 50 K increments using a Langevin thermostat while applying a restraint of 25 kcal mol^{-1} $Å^{-2}$. Next, a 200 ps equilibration run was performed in which the restraint was relaxed from 25 to 1.5 kcal mol^{-1} $Å^{-2}$. Long-range electrostatic interactions were treated with the particle mesh Ewald method and a cut-off of 10 Å from the edge of the solute was used for non-bonding interactions. The SHAKE algorithm was used to restrain all bonds containing a hydrogen atom. While applying the periodic boundary condition, triplicate 500 ns MD production simulations were performed using a NPT ensemble (Berendsen barostat). The hydrogen-bond distances between the heavy atoms in the four base pairs on the terminal end of the 3′-side of inosine were restrained using a force constant of 25 kcal mol^{-1} $Å^{-2}$ to prevent dsRNA unravelling during production simulations.

The CPPTRAJ module (V18.01) of the AMBER 2018 suite was used to analyze the MD trajectories.[89] For all hydrogen bond analysis, a hydrogen bond was considered to be present when the distance between the heavy atoms was < 3.4 Å and the hydrogen-bond angle was > 120°. Backbone RMSD was determined with respect to the first simulation frame. The active site RMSD was calculated using the heavy atoms of D52, D126, D240, E100, K155, the nucleotides containing the scissile phosphate moiety (A250 and U251),

and the 3′-flanking nucleotide of inosine (A252). The RMSDs reveal that all models are stable (Figure E.3 and Table E.1, Appendix E).

MD snapshots of catalytically-active conformations were extracted using strict distance-based criteria (Figure E.4 and Table E.2, Appendix E). In particular, a Mg^{2+}-ligated water was considered to be in position for leaving group protonation when the distance between the water oxygen and O3′ of A250 was < 3.5 Å. Mg^{2+} was considered to be appropriately positioned to facilitate leaving group departure when ≤ 2.5 Å from O3′. A water molecule was considered to be available for nucleophilic attack when the oxygen was < 4.0 Å from the phosphorus reaction centre. D240 was considered to be available to activate the water nucleophile when Oδ of D240 is < 7.5 Å from the phosphorus reaction centre. A water chain was determined to bridge D240 (Oδ) and the phosphorous reaction centre when oxygen of the second water was simultaneously within 3.5 Å of D240 and the oxygen of the nucleophilic water.

6.2.3. QM/MM Calculations

QM/MM was used to map the catalytic pathway starting from each MD snapshot chosen as discussed in the previous section. The ONIOM formalism was used due to previous successes of this methodology in characterizing the chemical pathways for many enzymes,[57, 63, 91-94] including metalloenzymes[91, 95-101] and nucleases.[45, 47-48, 57, 63, 102-103] For mechanisms involving a single metal, the QM region included Mg^{2+} and the directly coordinated D52, D126, and three water molecules, as well as an additional water molecule either directly coordinated to Mg^{2+} or hydrogen bonded to E100 (Figure E.5a–b, Appendix

E). The QM region also includes D240, E100, K155, the nucleotides containing the scissile phosphate moiety (A250 and U251), and an adjacent phosphate on the 3′ side of inosine (A252), leading to a total of 122 QM atoms and a QM region charge of –3. The MM region is composed of the remaining EndoV–RNA complex and any surrounding water molecules with an atom within 6 Å of the enzyme or substrate, which resulted in 2357 (direct Mg^{2+}–substrate coordination) or 2368 (indirect Mg^{2+}–substrate coordination) water molecules and an overall charge of –22. For mechanisms involving two metals, the QM region contained Mg_A^{2+} and Mg_B^{2+}, all coordinated residues, including D52, three water molecules and D240 for Mg_A^{2+}, and 2 water molecules and D126 (in a bidentate manner) as well as the substrate and an additional water molecule or the O3′-leaving group for Mg_B^{2+} (Figure E.5c–d, Appendix E). An additional water hydrogen bonding to D240 was also included in the QM region, along with E100, K155, the nucleotides containing the scissile phosphate moiety (A250 and U251), and an adjacent phosphate on the 3′ side of inosine (A252), which lead to 129 QM atoms and a charge of –1. The MM region was composed as per the single metal models, which led to 2314 water molecules and an overall charge of –20. In all models, the QM/MM boundary was placed between Cα and Cβ for the amino acids, and between C4′ and C5′ for the A250 and A252 nucleotides.

The B3LYP-D3(BJ)/6-31G(d,p) level of theory was used to describe the QM region of each model, while the AMBER force field (ff14Sb for EndoV[82] and ffbsc0χOL3 for RNA[83-84]) was used to describe the MM region. From each optimized reactant complex (RC), scans of key reaction parameters (i.e., distances corresponding to nucleophilic attack $r(O_w\cdots P)$ and P–O bond cleavage r(P–O3′)) were performed to obtain initial guesses for transition states (TSs). Each reaction parameter was increased/decreased in 0.10–0.15 Å

increments. Specifically, for concerted pathways, r(O_w···P) was successively reduced, and the bond was subsequently frozen to the distance corresponding to a maximum on the potential energy surface, while r(P–O3′) was consecutively increased. For stepwise mechanisms, r(O_w···P) was first successively decreased to obtain a guess for TS1, and r(P–O3′) was then elongated in the corresponding intermediate complex (IC, obtained from TS1) to obtain a guess for TS2. Subsequently, full TS optimizations were performed for all models, followed by scans on key reaction distances (i.e., r(O_w···P) and r(P–O3′)) to obtain the corresponding RC, IC, and/or product complex (PC), which were then fully optimized. For stepwise pathways, the scans of key distances (r(O_w···P) and r(P–O3′)) from TS1 and TS2 resulted in structurally and energetically similar ICs (within 6 kJ/mol, Figure E.6, Appendix E), verifying the pathways are connected. The nature of each minimum (all positive frequencies) and TS (single imaginary frequency) was determined using frequency calculations performed at the same level of theory as the optimizations, which also afforded the Gibbs energy corrections. Single-point energy calculations were carried out at the ONIOM(M06-2X/6-311+G(2df,p):AMBER) level of theory to obtain the relative energies.

All QM/MM calculations were performed using Gaussian 16 (revision B.01).[104]

6.3. Results and Discussion

6.3.1. The barrier for the single metal-mediated human EndoV pathway involving indirect (water-mediated) Mg^{2+}–substrate coordination exceeds the typical P–O bond cleavage barrier for nucleases.

As mentioned in the introduction, an X-ray crystal structure of the PC from wild-type bacterial EndoV shows a single metal in the active site (Figure 6.1d).[29] Although bacterial EndoV has been proposed to use H214 as the general base to activate the water nucleophile,[62] sequence alignment indicates that an aspartate residue (D240) replaces histidine in human EndoV (Figure 6.1a). Other single-metal dependent nucleases have been proposed to use aspartate to activate the water nucleophile.[42, 45, 49, 57, 63-64] For example, a crystal structure of a thio-substituted RC analogue of APE1 highlights a single Mn^{2+} ion indirectly coordinated to a non-bridging oxygen of the scissile phosphate through a water molecule, and D210 appropriately positioned to aid P–O bond hydrolysis (Figure 6.3a).[75] QM cluster and QM/MM calculations further verified that APE1 favours water-mediated Mg^{2+} coordination to both a non-bridging oxygen and the O3′-leaving group of the scissile phosphate, and D210 activates the water nucleophile in the preferred mechanism of action.[57] Similar water-mediated coordination occurs between the metal and a non-bridging phosphate oxygen in an RC analogue of single-metal dependent Vvn (Figure 6.3b).[71] Therefore, a human EndoV model involving one Mg^{2+} indirectly (water-mediated) coordinated to the substrate was initially investigated.

During MD simulations, Mg^{2+} maintains an octahedral coordination geometry fulfilled by D52, D126, and four water molecules (Figure 6.4a and Table E.3, Appendix E). A water molecule is consistently found on average 3.4 ± 0.1 Å from the phosphorus reaction

center (Figure 6.4b and Table E.2, Appendix E). Furthermore, D240 is located in proximity to the electrophilic site ($r(O\delta_{D240}\cdots P)$ = 6.8 ± 1.3 Å) and therefore available to initiate hydrolysis for 71% of the simulation time. Despite the flexibility of the active site (Figure E.3, Appendix E), a Mg^{2+}-ligated water molecule is suitably positioned with respect to O3′ to facilitate leaving group departure for 25% of the simulation trajectory (Figure 6.4b and Table E.2, Appendix E). Finally, a second metal-ligated water is aligned to position the substrate and provide charge stabilization through hydrogen bonding to a non-bridging scissile phosphate oxygen throughout the MD simulations (91% occupancy). Thus, the human EndoV–substrate complex frequently adopts a catalytically conducive conformation.

The QM/MM optimized RC from a representative MD snapshot maintains all key interactions necessary for the reaction to proceed (Figures 6.4c–d and E.7a, Appendix E). Furthermore, K155 provides additional substrate charge stabilization, donating a proton to the other non-bridging oxygen of the scissile phosphate and hydrogen bonding with the newly formed H–O group throughout the reaction. From the RC, nucleophilic attack occurs concomitantly with P–O3′ bond dissociation (Figures 6.4c–d and E.7a, Appendix E). The TS occurs before full proton transfer from the water nucleophile to D240 and Mg^{2+}-coordinated water to O3′, and the P–O bond is slightly (~0.15 Å) lengthened. However, despite literature precedence for an indirect (water-mediated) metal–substrate binding architecture in the active site of other nucleases (e.g., APE1[57, 63] and Vvn,[71] Figure 6.3a–b), the energy barrier associated with the corresponding configuration for human EndoV (120.1 kJ/mol, Figure 6.4d) is above typical barriers reported for metalloenzyme facilitated P–O bond cleavage (~58–96 kJ/mol).[28, 75, 105-108] This suggests that this pathway is unlikely,

a conclusion in line with previous MD and QM/MM calculations on bacterial EndoV.[62] Therefore, alternate metal-coordination architectures must be considered.

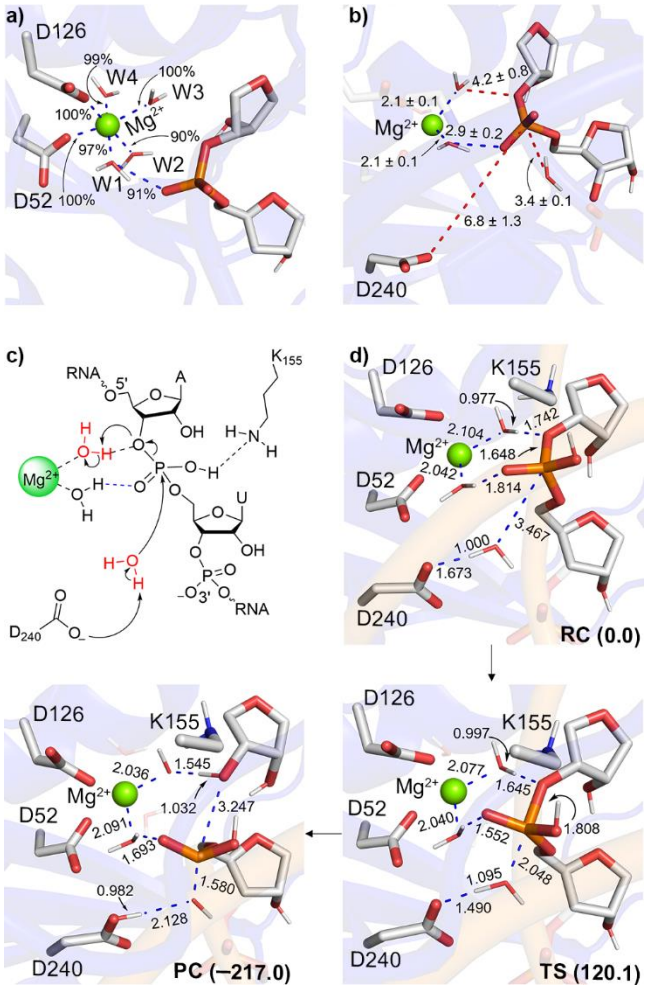

Figure 6.4. a) Average Mg^{2+} coordination distances (Å; occupancies (%) in parentheses) and b) representative MD structure highlighting key reaction parameters (Å) across all MD replicas, as well as the c) proposed reaction pathway and d) key reaction parameters (Å) for each stationary point along the human EndoV-mediated phosphodiester bond cleavage involving a single metal indirectly coordinated to a non-bridging phosphate oxygen of the scissile phosphate. Relative Gibbs energies (kJ/mol) are provided in parentheses. See Figure E.7a for additional reaction parameters.

6.3.2. Charge stabilization of the non-bridging oxygen of the scissile phosphate by direct Mg^{2+} coordination renders the human EndoV one-metal mediated P–O bond cleavage pathway catalytically feasible.

Structural data highlights that single-metal dependent nucleases can adopt many unique metal–substrate binding architectures (Figures 6.3 and E.8, Appendix E).[29, 70-74, 76] The higher than anticipated barrier for P–O bond cleavage in the previous single-metal mediated mechanism for human EndoV could be lowered through enhanced substrate charge stabilization, which could be provided through direct coordination of the metal to the substrate. Indeed, X-ray crystallographic data for various one-metal dependent nucleases, including I-*Ppo*I (Figure 6.3c),[70] Hpy188I (Figure E.8a, Appendix E),[74] T4 endonuclease VII (Figure E.8b, Appendix E),[73] and I-HmuI (Figure E.8c, Appendix E),[72] reveal direct metal coordination to both the O3′-leaving group and a non-bridging oxygen of the scissile phosphate. However, at the beginning of MD production simulations on a human EndoV RC with this metal–substrate ligation, Mg^{2+} immediately loses direct coordination to O3′, with water occupying the space between the metal and the O3′-leaving group (Figures 6.5a–b). Nevertheless, direct metal coordination to a non-bridging phosphate oxygen is maintained throughout the simulation (100% occupancy). The same metal-binding architecture has been seen in crystal structures of single-metal dependent *Bgl*Il (Figure 6.3d)[76] and *Eco*RI (Figure E.8d, Appendix E),[109] as well as computational studies on bacterial EndoV[62] and the endonuclease domain of influenza polymerase.[61] In this active site configuration for human EndoV (Figure 6.5b and Table E.2, Appendix E), Mg^{2+} remains coordinated to the substrate (2.0 ± 0.1 Å) throughout the MD simulations, while the Mg^{2+}-ligated water falls < 3.5 Å from O3′ for 40% of the simulation time. Water

adopts a position for nucleophilic attack (3.5 ± 0.1 Å from electrophilic phosphorus) and D240 is suitably situated to activate the nucleophile ($r(O\delta_{D240}\cdots P)$ = 7.3 ± 0.9 Å). In complement to metal–substrate ligation, the octahedral coordination sphere of Mg^{2+} is filled by three water molecules, D52, and D126 (Figure 6.5a and Table E.3, Appendix E), reflecting the experimentally-reported inactive D52A mutant.[17-18]

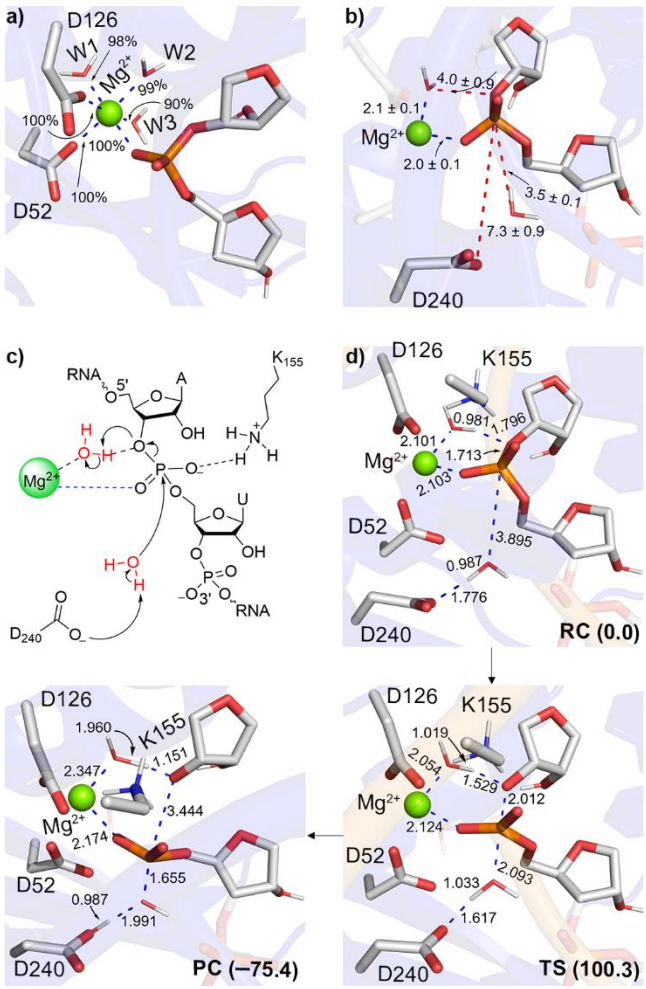

Figure 6.5. a) Average Mg^{2+} coordination distances (Å; occupancies (%) in parentheses) and b) representative MD structure highlighting key reaction parameters (Å) across all MD replicas, as well as the c) proposed reaction pathway and d) key reaction parameters (Å) for each stationary point along the human EndoV-mediated phosphodiester bond cleavage involving a single metal directly coordinated to a non-bridging phosphate oxygen of the

scissile phosphate. Relative Gibbs energies (kJ/mol) are provided in parentheses. See Figure E.7b for additional reaction parameters.

The structural features of the human EndoV RC observed in MD simulations are preserved upon QM/MM optimizations (Figure 6.5c–d and E.7b, Appendix E). Furthermore, the majority of the QM/MM structural parameters along the concerted pathway for P–O bond cleavage are maintained with a change from indirect to direct metal–nonbridging oxygen coordination. The enhanced substrate charge stabilization provided by direct Mg^{2+}–substrate coordination ensures K155 remains cationic throughout the reaction pathway, while providing additional substrate charge stabilization through hydrogen bonding. As a result, the catalytic barrier decreases by 19.8 kJ/mol upon conversion from indirect to direct metal–substrate coordination, leading to a phosphodiester bond cleavage barrier ($\Delta^{\ddagger}G$ = 100.3 kJ/mol) close to the experimentally-observed barriers for a range of nucleases (~58–96 kJ/mol),[28, 75, 105-108] including mouse EndoV (84.5 kJ/mol).[28] The energetic feasibility of the human EndoV pathway involving direct metal–substrate coordination correlates with a more stable active site across the MD simulation trajectories for this metal binding geometry (Figure E.3d and Table E.1, Appendix E) and increased residency of a catalytically-active conformation (Table E.2, Appendix E). The proposal that the catalytic step of human EndoV can proceed in the presence of one metal adds to growing literature suggesting that a single metal may be enough to cleave the phosphodiester backbone of nucleic acids.[51-57, 62-64, 67-68]

6.3.3. Regardless of the mode of leaving group stabilization, the two-metal mediated P–O bond cleavage pathway for human EndoV is catalytically viable.

Although bacterial EndoV has been shown to utilize a single metal for phosphodiester bond cleavage[29, 62, 106] and our calculations suggest that human EndoV can similarly follow a one-metal dependent pathway, the activity of mouse EndoV depends on two catalytic Mg^{2+} ions.[28] Since the active sites of mouse and human EndoV contain the same catalytic residues (Figure 6.1a), and the general two-metal mediated mechanism is widely accepted for nucleic acid backbone cleavage,[28, 30-50] the ability of human EndoV to use two metals to break the P–O bond of dsRNA was considered. Throughout MD simulations, the human EndoV active site was significantly more stable in the presence of two metals compared to a single metal (Figure E.3, Appendix E). The two metals remain ~ 4 Å apart (Figure 6.6a and Table E.3, Appendix E), which correlates with the ~ 3.9 Å separation proposed for two-metal mediated RNA hydrolysis mechanisms.[110] Furthermore, a water molecule is properly positioned with respect to the phosphorus reaction center for nucleophilic attack for 99% of the simulations ($r(O_W \cdots P) = 3.2 \pm 0.1$ Å), while maintaining coordination to Mg_A^{2+} ($r(O_W \cdots Mg_A^{2+}) = 2.1 \pm 0.1$ Å, Figure E.6b, Appendix E). A second water molecule bridges the nucleophile and D240 for 54% of the simulation time (Table S2), which could facilitate nucleophilic attack ($r(O_W \cdots O_{\text{bridging water}}) = 3.0 \pm 0.3$ Å and $r(O\delta_{D240} \cdots O_{\text{bridging water}}) = 2.9 \pm 0.3$ Å). This suggests that two metals are well accommodated in the active site of human EndoV when bound to the dsRNA substrate, while maintaining a catalytically conducive conformation.

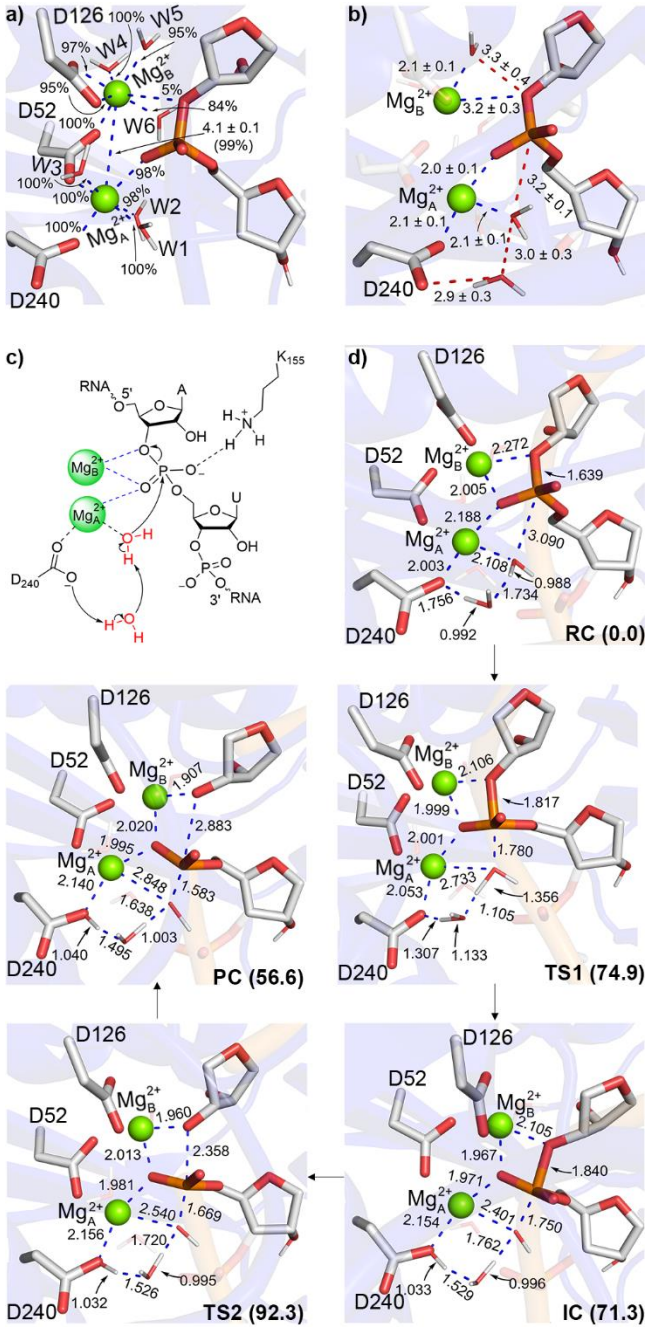

Figure 6.6. a) Average Mg^{2+} coordination distances (Å; occupancies (%) in parentheses) and b) representative MD structure highlighting key reaction parameters (Å) across all MD replicas, as well as the c) proposed reaction pathway and d) key reaction parameters (Å) for each stationary point along the human EndoV-mediated phosphodiester bond cleavage involving two metals with direct Mg_B^{2+}–O3′ coordination. Relative Gibbs energies (kJ/mol) are provided in parentheses. See Figure E.9 for additional reaction parameters.

In terms of metal coordination, the Mg^{2+} ions are bridged by D52 throughout the simulations, correlating with the observation that mutation of this residue to alanine kills enzyme activity.[17-18] The remaining Mg_A^{2+} octahedral coordination is fulfilled by three water molecules, D240, and a non-bridging oxygen of the scissile phosphate. Although two water molecules and D126 (bidentate coordination) are also bound to Mg_B^{2+}, two Mg_B^{2+}–substrate coordination configurations occur (Figure 6.6a and Tables E.2–E.3, Appendix E). First, unlike the single-metal human EndoV model, 75 ns of the MD simulation trajectory maintained direct Mg_B^{2+}–O3′ ligation for the two-metals model. This metal–substrate binding configuration mirrors that observed in X-ray crystal structures of other two-metal mediated nucleases, including RNase H (Figure 6.3e),[30] AaRNase III (Figure E.8e, Appendix E),[31] HindIII (Figure E.8f, Appendix E),[41] and retroviral integrase (Figure E.8g, Appendix E).[37] Second, a water molecular intervenes between Mg_B^{2+} and the O3′-leaving group for the remainder of the simulation time ($r(O3'\cdots O_W) = 3.3 \pm 0.4$ Å, Figure 6.6b). This metal-binding architecture has been observed in X-ray crystallographic data for two-metal dependent BamHI (Figure 6.3f),[32] MutH (Figure E.8h, Appendix E),[40] and BgII (Figure E.8i, Appendix E).[39] Due to the persistence of two metal-binding configurations in the human EndoV active site, which each correspond to architectures previously reported in the nuclease literature, MD snapshots corresponding to both metal–substrate coordination geometries were used to initiate QM/MM calculations.

When an MD snapshot corresponding to direct Mg_B^{2+}–O3′ coordination was considered, the metal binding architecture is generally preserved upon QM/MM optimization, with the exception of Mg_B^{2+} adopting penta-coordination due to monodentate (rather than bidentate) binding of D126 (Figures 6.6c–d and E.9, Appendix E). A similar

Mg_B^{2+} penta-coordination geometry was observed in the crystal structure of RNase H (Figure 6.3e).[30] The phosphodiester bond cleavage follows a stepwise mechanism (Figure 6.6c–d and E.9, Appendix E). In the first step, water nucleophilic attack on the phosphorus reaction center is coupled with proton transfer from the nucleophile to metal-coordinated D240 through a bridging water. While Mg_A^{2+} becomes penta-coordinated in TS1 due to the loss of water nucleophile coordination, Mg_B^{2+} reverts to an octahedral binding configuration due to bidentate coordination of D126. K155 donates a proton to a non-bridging oxygen of the scissile phosphate moiety in TS1 to help stabilize increased charge on the substrate. This change in metal binding architecture and protonation of the substrate parallels that reported in a previous QM/MM MD study of two-metal dependent *Bacillus halodurans* RNase H.[44] These structural features result in a barrier of 74.9 kJ/mol and a slightly (3.6 kJ/mol) stabilized phosphorane intermediate. In the second reaction step, the phosphodiester bond significantly elongates (r(P–O3′) increases by 0.5 Å compared to IC) and the proton on the non-bridging oxygen of the scissile phosphate to return to K155. The associated rate-determining barrier is 92.3 kJ/mol. When an MD snapshot with indirect (water-mediated) metal coordination to the substrate was considered, both metals maintain octahedral coordination geometry throughout the stepwise reaction pathway (Figure 6.7 and E.10, Appendix E). More importantly, most reaction parameters are preserved across both two-metal mediated mechanisms characterized in the present work, leading to similar rate-determining barriers regardless of the Mg_B^{2+}–O3′ coordination mode (within 3.3 kJ/mol, Figures 6.6 and 6.7). Thus, irrespective of whether direct or indirect (water-mediated) Mg_B^{2+} coordination to the O3′ leaving group occurs, our predicted barriers for the two-metal mediated human EndoV pathway fall within the range of the experimental barriers reported for other nucleases (~58–96 kJ/mol),[28, 75, 105-108] including mouse EndoV

(84.5 kJ/mol),[28] and the computed human EndoV–substrate structures are also consistent with X-ray crystallographic data for mouse EndoV (Figure E.11). Overall, our calculations predict that human EndoV can make use of two metal-binding positions in the active site to facilitate P–O bond cleavage.

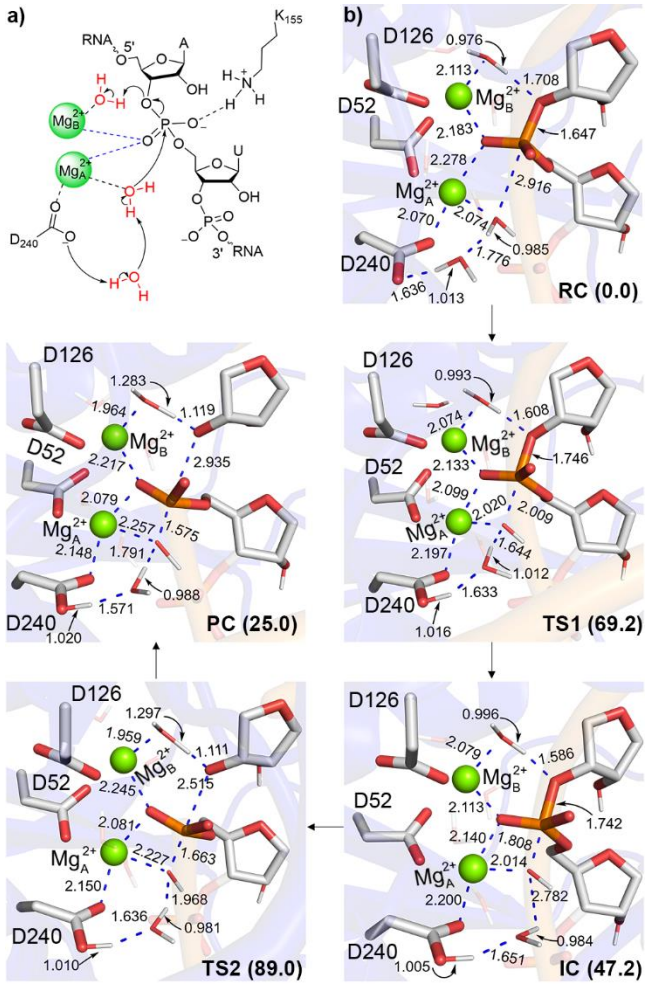

Figure 6.7. a) The proposed reaction pathway and b) key reaction parameters (Å) for each stationary point along the human EndoV-mediated phosphodiester bond cleavage involving two metals with indirect (water-mediated) Mg_B^{2+}–O3′ coordination. Relative Gibbs energies (kJ/mol) are provided in parentheses. See Figure E.10 for additional reaction parameters.

6.3.4. Human EndoV can uniquely use either one or two metals to catalyze cleavage of the highly stable phosphodiester bond.

The present study is an important step toward understanding the catalytic mechanism of human EndoV due to limited existing data. In fact, the only currently available X-ray crystal structure of human EndoV does not contain the substrate or metal cofactor/s (Figure 6.1b)[28] and no kinetic studies have been conducted. As a result, our MD simulations provide the first picture of the human EndoV–substrate complex in the presence of active site metal cofactor/s and show the stability of the complex for a range of metal binding architectures. Our QM/MM calculations highlight that human EndoV can use a single metal to mediate phosphodiester bond cleavage. Indeed, although a mechanism involving indirect (water-mediated) metal coordination to the leaving group and nonbridging oxygen of the substrate results in a barrier well above those observed for other nucleases (Figure 6.4),[28, 75, 105-108] improved substrate charge stabilization afforded by direct metal coordination to the nonbridging oxygen of the scissile phosphate coupled with indirect coordination to the O3′-leaving group renders the one-metal mediated pathway energetically feasible (Figure 6.5). In our proposed single-metal dependent human EndoV mechanism, an active site aspartate residue (D240) acts as the general base to activate the water nucleophile, the metal aids leaving group departure by activating a water molecule that protonates O3′, and an active site lysine (K155) along with the metal facilitate substrate positioning and charge stabilization. Our computed geometries correlate with structural data for other members of the EndoV family (Figure E.11a, Appendix E) and our data clarify the observed loss in catalytic activity upon D52A mutation,[17-18] with D52 being a key metal binding residue.

Our proposal that human EndoV can facilitate the phosphodiester bond cleavage in the presence of a single metal parallels suggestions made for bacterial EndoV,[29, 62, 111] as well as many other nucleases including APE1,[57, 63-64] I-*Ppo*I,[70, 108, 112] Hpy188I,[74] T4 endonuclease VII,[73] I-HmuI,[72] Vvn,[71] *Eco*RI,[109] and *Bgl*ll (Figures 6.3 and E.8, Appendix E).[76] In all cases, one metal is proposed to be enough for the nuclease to facilitate the extremely challenging P–O bond cleavage due to the presence of a strong general base in the active site, namely a histidine (bacterial EndoV, I-*Ppo*I, T4 endonuclease VII, Vvn, I-HmuI), aspartate (APE1, human EndoV), tyrosine (Hpy188I), or a neighboring phosphate moiety of the substrate (*Eco*RI). Nevertheless, the metal–substrate coordination geometry and therefore intricate details of the mechanism can vary between these enzymes due to differences in active site compositions. For example, the weaker general base in bacterial (histidine) compared to human (aspartate) EndoV results in an active site lysine playing a more active role in bacteria (i.e., general acid rather than substrate charge stabilization, Figure 6.8). Alternatively, despite many nucleases containing an active site aspartate residue that likely initiates the reaction (e.g., APE1 (D210),[57, 63-64] RNase P (D399),[42] *Eco*RV (D90),[45] and HIV-1 reverse transcriptase (D185[49])), different active site compositions alter the fine details of the chemical pathway. For example, the availability of a weaker acid to stabilize the charge on the scissile phosphate (lysine versus protonated histidine) renders the indirect metal coordination to the second nonbridging oxygen observed in APE1 unable to promote human EndoV activity, which necessitates direct metal coordination (Figure 6.8). Perhaps more importantly, although active site composition diversity can result in subtle mechanistic deviations, our calculations contribute to the growing body of literature, including computational support,[57, 62-64] that

suggests catalytic cleavage of phosphodiester bonds is feasible in the presence of a single active site metal.[51, 54-55, 67-68]

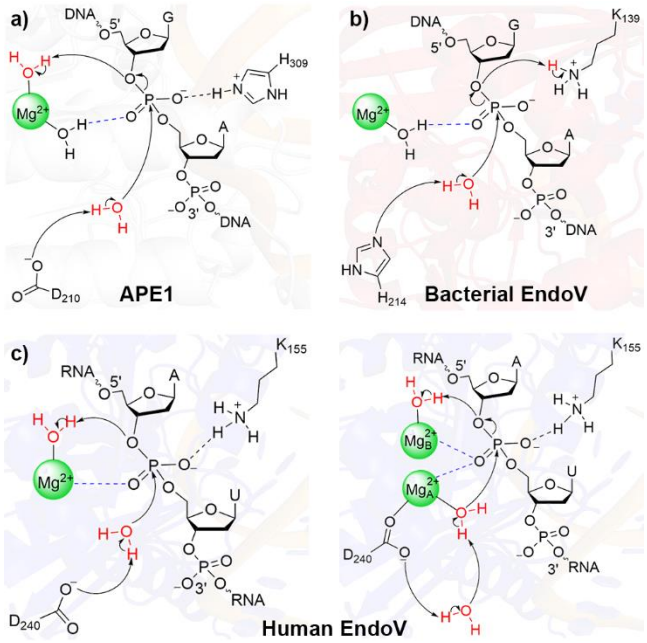

Figure 6.8. The proposed phosphodiester bond cleavage pathway for a) APE1,[57, 63-64] b) bacterial EndoV,[62] and c) one (left) and two-metal mediated (right) human EndoV as characterized in the present work.

To complement one metal mediated catalysis, our calculations suggest that human EndoV can efficiently facilitate phosphodiester bond cleavage in the presence of two metals, regardless of the metal–substrate (leaving group) coordination geometry (Figures 6.6 and 6.7). In these mechanisms, the first metal (Mg_A^{2+}) triggers activation of the nucleophilic water, D240 facilitates proton abstraction from the Mg_A^{2+}-coordinated water nucleophile with help of a bridging water molecule, and the second metal (Mg_B^{2+}) aids leaving group departure through either direct coordination or activation of a water molecule for O3′ protonation. The two metals along with K155 stabilize the charge on the substrate. Both mechanisms are consistent with crystallographic data for mouse EndoV (Figure E.11b–c, Appendix E), which has a similar active site composition (Figure 6.1a).

Furthermore, both metal binding architectures have been reported for other nucleases, with direct metal–leaving group coordination observed for RNase H (Figure 6.3e),[30] AaRNase III (Figure E.9e, Appendix E),[31] HindIII (Figure E.8f, Appendix E),[41] and retroviral integrase (Figure E.8g, Appendix E),[37] and water-mediated Mg_B^{2+}–O3′ coordination observed for BamHI (Figure 6.3f),[32] MutH (Figure E.8h, Appendix E),[40] and Bgll (Figure E.8i, Appendix E).[39] Although direct metal coordination to the leaving group is less persistent in the active site according to MD simulations on the RC, the unique behavior that different metal binding architectures are possible may arise due to the additional charge stabilization afforded by K155 in the human EndoV active site compared to the charge stabilization being restricted to the metals for other two-metal mediated nucleases.[30-32, 37, 40]

While the crystal structure of bacterial EndoV displays a single metal (Figure 6.1d)[29] and mouse EndoV shows two metals (Figure 6.1c)[28] in the active site, our calculations reveal that human EndoV can invoke either one or two metals for catalysis. Although the barrier for single-metal mediated catalysis is slightly larger than that for the two-metal dependent pathways, the final number of metals utilized likely depends on the metal ion concentration among other considerations. This proposal is supported by experimentally-reported metal concentration dependent behaviour of PvuII endonuclease.[67] Specifically, although the X-ray crystal structure of PvuII endonuclease contains two metals in the active site (Figure E.8j, Appendix E),[34] P–O bond cleavage rates as a function of metal ion concentration fit both one-metal and two-metal ion models.[67] Therefore, the enzyme was suggested to use only one metal ion at low metal concentrations, but a more efficient two-metal mediated catalysis under saturating metal ion conditions. Similarly,

while one metal is essential for human EndoV catalytic activity, the enzyme can benefit from the use of two metals due to the presence of two suitable Mg^{2+} binding positions in the active site. To the best of our knowledge, our work is the first computational study to provide direct mechanistic comparisons between one- and two-metal mediated P–O bond cleavage reactions in the same active site architecture. Our work highlights that while some enzymes may use either one[29, 70-74, 76] or two[28, 30-41] metals for phosphodiester bond hydrolysis in nucleic acids (Figures 6.3 and E.8, Appendix E), some enzymes have the ability to exploit both pathways, thus bringing a fresh perspective to the one versus two metal ion controversy. Our calculations thereby emphasize that it is important to bear in mind that there might not be an unequivocal answer for some nucleases regarding the number of metals required to catalyze phosphodiester bond cleavage, with the composition of the enzyme active site and metal concentration playing significant roles in determining the metal ion stoichiometry requirements for catalysis.

6.4. Conclusion

The present work has used MD simulations and QM/MM calculations to afford the first proposed mechanism of action for human EndoV. MD simulations were initially used to reveal the structure of the human EndoV–substrate complex in the presence of metal cofactor/s, which highlighted the stability of the complex in the presence of a range of active site metal binding architectures. Subsequently, four unique phosphodiester bond cleavage pathways were characterized using QM/MM calculations. Although a mechanism involving indirect (water-mediated) coordination of the metal to the substrate results in a barrier well above the experimentally-determined activation energy for a range of other

nucleases,[28, 75, 105-108] direct metal coordination to the nonbridging oxygen of the scissile phosphate coupled with indirect metal coordination to the O3′ leaving group yields an energetically feasible pathway for single-metal mediated catalysis. In our predicted mechanism, D240 activates the nucleophilic water, a Mg^{2+}-ligated water protonates the leaving group, and substrate charge stabilization is provided by both the directly coordinated Mg^{2+} and K155. Nevertheless, our calculations predict that human EndoV can also accommodate two metals in the active site and use both metals to facilitate the P–O bond cleavage. In the two-metal mediated mechanism, the first metal (Mg_A^{2+}) activates the nucleophilic water while D240 abstracts the proton from the Mg_A^{2+}-coordinated water nucleophile with help of a bridging water molecule, and the second metal (Mg_B^{2+}) promotes leaving group departure either through direct coordination or activation of a water molecule to protonate O3′. The two metals along with K155 stabilize the charge on the substrate throughout the reaction. Thus, in line with experimental literature for another nuclease (PvuII endonuclease),[67] while one metal is essential for human EndoV catalytic activity, the enzyme can benefit from the use of two metals, with the final number of metals utilized likely depending on the local metal ion concentration among other considerations. To the best of our knowledge, this is the first computational study that has provided direct evidence that a given endonuclease can utilize both one and two-metal dependent mechanisms to cleave phosphodiester bonds in nucleic acids, thus emphasizing that researchers must bear in mind that there may not be an unequivocal answer regarding metal ion stoichiometry requirements for catalysis and adding a unique perspective to the existing one versus two metal controversy in the literature. The mechanistic details of human EndoV function uncovered in the present work are vital for further discerning the central role of this enzyme in human disease (e.g., cardiovascular disease,[15] cancer,[22] and psychiatric disorders),[21]

developing new therapeutics that target EndoV,[15, 21-22] and refining uses of EndoV in biotechnology.[25-27, 77-79]

6.5. References

(1) Kuraoka, I., Diversity of endonuclease V: from DNA repair to RNA editing. *Biomolecules* **2015,** *5* (4), 2194–2206.

(2) Schouten, K. A.; Weiss, B., Endonuclease V protects *Escherichia coli* against specific mutations caused by nitrous acid. *Mutat. Res.* **1999,** *435* (3), 245–254.

(3) Gott, J. M.; Emeson, R. B., Functions and mechanisms of RNA editing. *Annu. Rev. Genet.* **2000,** *34* (1), 499–531.

(4) Pang, B.; McFaline, J. L.; Burgis, N. E.; Dong, M.; Taghizadeh, K.; Sullivan, M. R.; Elmquist, C. E.; Cunningham, R. P.; Dedon, P. C., Defects in purine nucleotide metabolism lead to substantial incorporation of xanthine and hypoxanthine into DNA and RNA. *Proc. Natl. Acad. Sci. U.S.A.* **2012,** *109* (7), 2319–2324.

(5) Behmanesh, M.; Sakumi, K.; Abolhassani, N.; Toyokuni, S.; Oka, S.; Ohnishi, Y.; Tsuchimoto, D.; Nakabeppu, Y., ITPase-deficient mice show growth retardation and die before weaning. *Cell Death Differ.* **2009,** *16* (10), 1315–1322.

(6) Bass, B. L., RNA editing by adenosine deaminases that act on RNA. *Annu. Rev. Biochem.* **2002,** *71* (1), 817–846.

(7) Nishikura, K., Functions and regulation of RNA editing by ADAR deaminases. *Annu. Rev. Biochem.* **2010,** *79*, 321–349.

(8) Farajollahi, S.; Maas, S., Molecular diversity through RNA editing: a balancing act. *Trends Genet.* **2010,** *26* (5), 221–230.

(9) Han, L.; Diao, L.; Yu, S.; Xu, X.; Li, J.; Zhang, R.; Yang, Y.; Werner, H. M.; Eterovic, A. K.; Yuan, Y., The genomic landscape and clinical relevance of A-to-I RNA editing in human cancers. *Cancer cell* **2015,** *28* (4), 515–528.

(10) Paz, N.; Levanon, E. Y.; Amariglio, N.; Heimberger, A. B.; Ram, Z.; Constantini, S.; Barbash, Z. S.; Adamsky, K.; Safran, M.; Hirschberg, A., Altered adenosine-to-inosine RNA editing in human cancer. *Genome Res.* **2007,** *17* (11), 1586–1595.

(11) Srivastava, P. K.; Bagnati, M.; Delahaye-Duriez, A.; Ko, J.-H.; Rotival, M.; Langley, S. R.; Shkura, K.; Mazzuferi, M.; Danis, B.; van Eyll, J., Genome-wide analysis of differential RNA editing in epilepsy. *Genome Res.* **2017,** *27* (3), 440–450.

(12) Hosaka, T.; Tsuji, H.; Kwak, S., RNA editing: a new therapeutic target in amyotrophic lateral sclerosis and other neurological diseases. *Int. J. Mol. Sci.* **2021,** *22* (20), 10958.

(13) Tran, S. S.; Jun, H.-I.; Bahn, J. H.; Azghadi, A.; Ramaswami, G.; Van Nostrand, E. L.; Nguyen, T. B.; Hsiao, Y.-H. E.; Lee, C.; Pratt, G. A., Widespread RNA editing dysregulation in brains from autistic individuals. *Nat. Neurosci.* **2019,** *22* (1), 25–36.

(14) Khermesh, K.; D'Erchia, A. M.; Barak, M.; Annese, A.; Wachtel, C.; Levanon, E. Y.; Picardi, E.; Eisenberg, E., Reduced levels of protein recoding by A-to-I RNA editing in Alzheimer's disease. *RNA* **2016,** *22* (2), 290–302.

(15) Kong, X. Y.; Huse, C.; Yang, K.; Øgaard, J.; Berges, N.; Vik, E. S.; Nawaz, M. S.; Quiles-Jiménez, A.; Abbas, A.; Gregersen, I., Endonuclease V regulates atherosclerosis through C-C motif chemokine ligand 2-mediated monocyte infiltration. *J. Am. Heart Assoc.* **2021,** *10* (14), e020656.

(16) Breen, M. S.; Dobbyn, A.; Li, Q.; Roussos, P.; Hoffman, G. E.; Stahl, E.; Chess, A.; Sklar, P.; Li, J. B.; Devlin, B., Global landscape and genetic regulation of RNA editing in cortical samples from individuals with schizophrenia. *Nat. Neurosci.* **2019,** *22* (9), 1402–1412.

(17) Morita, Y.; Shibutani, T.; Nakanishi, N.; Nishikura, K.; Iwai, S.; Kuraoka, I., Human endonuclease V is a ribonuclease specific for inosine-containing RNA. *Nat. Commun.* **2013,** *4*, 2273.

(18) Vik, E. S.; Nawaz, M. S.; Andersen, P. S.; Fladeby, C.; Bjoras, M.; Dalhus, B.; Alseth, I., Endonuclease V cleaves at inosines in RNA. *Nat. Commun.* **2013,** *4*, 2271.

(19) Zhang, Z.; Hao, Z.; Wang, Z.; Li, Q.; Xie, W., Structure of human endonuclease V as an inosine-specific ribonuclease. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2014,** *70* (9), 2286–2294.

(20) Cao, W., Endonuclease V: an unusual enzyme for repair of DNA deamination. *Cell. Mol. Life Sci.* **2013,** *70* (17), 3145–3156.

(21) Rajkumar, A. P.; Christensen, J. H.; Mattheisen, M.; Jacobsen, I.; Bache, I.; Pallesen, J.; Grove, J.; Qvist, P.; McQuillin, A.; Gurling, H. M., Analysis of t (9; 17)(q33. 2; q25. 3) chromosomal breakpoint regions and genetic association reveals novel candidate genes for bipolar disorder. *Bipolar Disord.* **2015,** *17* (2), 205–211.

(22) Kong, X. Y.; Vik, E. S.; Nawaz, M. S.; Berges, N.; Dahl, T. B.; Vågbø, C.; Suganthan, R.; Segers, F.; Holm, S.; Quiles-Jiménez, A., Deletion of Endonuclease V suppresses chemically induced hepatocellular carcinoma. *Nucleic Acids Res.* **2020,** *48* (8), 4463–4479.

(23) Knutson, S. D.; Arthur, R. A.; Johnston, H. R.; Heemstra, J. M., Selective enrichment of A-to-I edited transcripts from cellular RNA using Endonuclease V. *J. Am. Chem. Soc.* **2020,** *142* (11), 5241–5251.

(24) Knutson, S. D.; Arthur, R. A.; Johnston, H. R.; Heemstra, J. M., Direct immunodetection of global A-to-I RNA editing activity with a chemiluminescent bioassay. *Angew. Chem. Int. Ed.* **2021,** *60* (31), 17009–17017.

(25) Chang, Y.; Huang, Z.; Quan, H.; Li, H.; Yang, S.; Song, Y.; Wang, J.; Yuan, J.; Wu, C., Construction of a DNA damage repair gene signature for predicting prognosis and immune response in breast cancer. *Front. Oncol.* **2022,** *12*, 1085632.

(26) Huang, J.; Kirk, B.; Favis, R.; Soussi, T.; Paty, P.; Cao, W.; Barany, F., An endonuclease/ligase based mutation scanning method especially suited for analysis of neoplastic tissue. *Oncogene* **2002,** *21* (12), 1909–1921.

(27) Pincas, H.; Pingle, M. R.; Huang, J.; Lao, K.; Paty, P. B.; Friedman, A. M.; Barany, F., High sensitivity EndoV mutation scanning through real-time ligase proofreading. *Nucleic Acids Res.* **2004,** *32* (19), e148.

(28) 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.

(29) Dalhus, B.; Arvai, A. S.; Rosnes, I.; Olsen, Ø. E.; Backe, P. H.; Alseth, I.; Gao, H.; Cao, W.; Tainer, J. A.; Bjørås, M., Structures of endonuclease V with DNA reveal initiation of deaminated adenine repair. *Nat. Struct. Mol. Biol.* **2009,** *16*, 138–143.

(30) Nowotny, M.; Gaidamakov, S. A.; Crouch, R. J.; Yang, W., Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell* **2005,** *121* (7), 1005–1016.

(31) 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.* **2007,** *67* (1), 143–154.

(32) Viadiu, H.; Aggarwal, A. K., The role of metals in catalysis by the restriction endonuclease Bam HI. *Nat. Struct. Biol.* **1998,** *5* (10), 910–916.

(33) Deibert, M.; Grazulis, S.; Sasnauskas, G.; Siksnys, V.; Huber, R., Structure of the tetrameric restriction endonuclease NgoMIV in complex with cleaved DNA. *Nat. Struct. Biol.* **2000,** *7* (9), 792–799.

(34) Horton, J. R.; Cheng, X., PvuII endonuclease contains two calcium ions in active sites. *J. Mol. Biol.* **2000,** *300* (5), 1049–1056.

(35) Dunten, P. W.; Little, E. J.; Gregory, M. T.; Manohar, V. M.; Dalton, M.; Hough, D.; Bitinaite, J.; Horton, N. C., The structure of SgrAI bound to DNA; recognition of an 8 base pair target. *Nucleic Acids Res.* **2008,** *36* (16), 5405–5416.

(36) Lambert, A. R.; Sussman, D.; Shen, B.; Maunus, R.; Nix, J.; Samuelson, J.; Xu, S.-Y.; Stoddard, B. L., Structures of the rare-cutting restriction endonuclease NotI reveal a unique metal binding fold involved in DNA binding. *Structure* **2008,** *16* (4), 558–569.

(37) Hare, S.; Maertens, G. N.; Cherepanov, P., 3′-Processing and strand transfer catalysed by retroviral integrase in crystallo. *EMBO J.* **2012,** *31* (13), 3020–3028.

(38) Hiller, D. A.; Rodriguez, A. M.; Perona, J. J., Non-cognate enzyme–DNA complex: structural and kinetic analysis of EcoRV endonuclease bound to the EcoRI recognition site GAATTC. *J. Mol. Biol.* **2005,** *354* (1), 121–136.

(39) Newman, M.; Lunnen, K.; Wilson, G.; Greci, J.; Schildkraut, I.; Phillips, S. E., Crystal structure of restriction endonuclease BglI bound to its interrupted DNA recognition sequence. *EMBO J.* **1998,** *17* (18), 5466–5476.

(40) 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.

(41) Watanabe, N.; Takasaki, Y.; Sato, C.; Ando, S.; Tanaka, I., Structures of restriction endonuclease HindIII in complex with its cognate DNA and divalent cations. *Acta Crystallogr. D Biol. Crystallogr.* **2009,** *65* (12), 1326–1333.

(42) Howard, M. J.; Klemm, B. P.; Fierke, C. A., Mechanistic studies reveal similar catalytic strategies for phosphodiester bond hydrolysis by protein-only and RNA-dependent ribonuclease P. *J. Biol. Chem.* **2015,** *290* (21), 13454–13464.

(43) Mordasini, T.; Curioni, A.; Andreoni, W., Why do divalent metal ions either promote or inhibit enzymatic reactions? The case of BamHI restriction endonuclease from combined quantum-classical simulations. *J. Biol. Chem.* **2003,** *278* (7), 4381–4384.

(44) 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.

(45) Imhof, P.; Fischer, S.; Smith, J. C., Catalytic mechanism of DNA backbone cleavage by the restriction enzyme *Eco*RV: a quantum mechanical/molecular mechanical analysis. *Biochemistry* **2009,** *48* (38), 9061–9075.

(46) 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.

(47) 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.

(48) Araujo, A. R.; Ribeiro, A. J.; 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.

(49) Rungrotmongkol, T.; Mulholland, A. J.; Hannongbua, S., QM/MM simulations indicate that Asp185 is the likely catalytic base in the enzymatic reaction of HIV-1 reverse transcriptase. *MedChemComm* **2014,** *5* (5), 593–596.

(50) Casalino, L.; Nierzwicki, L.; 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.

(51) 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.

(52) Dupureur, C. M., Roles of metal ions in nucleases. *Curr. Opin. Chem. Biol.* **2008,** *12* (2), 250–255.

(53) Dupureur, C. M., An integrated look at metallonuclease mechanism. *Curr. Chem. Biol.* **2008,** *2* (2), 159–173.

(54) Yang, W., An equivalent metal ion in one- and two-metal-ion catalysis. *Nat. Struct. Mol. Biol.* **2008,** *15* (11), 1228–1231.

(55) 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.

(56) Cowan, J. A., Metal-mediated hydrolysis of biological phosphate esters: a critical analysis of the essential metal ion stoichiometry for magnesium-dependent nuclease activation. *J. Biol. Inorg. Chem.* **1997,** *2* (2), 168–176.

(57) 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.

(58) Pitts, S. L.; Liou, G. F.; Mitchenall, L. A.; Burgin, A. B.; Maxwell, A.; Neuman, K. C.; Osheroff, N., Use of divalent metal ions in the DNA cleavage reaction of topoisomerase IV. *Nucleic Acids Research* **2011,** *39* (11), 4808-4817.

(59) 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.

(60) Onishi, I.; Sunaba, S.; Yoshida, N.; Hirata, F.; Irisa, M., Role of Mg2+ Ions in DNA Hydrolysis by Eco RV, Studied by the 3D-Reference Interaction Site Model and Molecular Dynamics. *J. Phys. Chem. B.* **2018,** *122* (39), 9061-9075.

(61) Xiao, S.; Klein, M. L.; LeBard, D. N.; Levine, B. G.; Liang, H.; MacDermaid, C. M.; Alfonso-Prieto, M., Magnesium-dependent RNA binding to the PA endonuclease domain of the avian influenza polymerase. *J. Phys. Chem. B* **2014,** *118* (4), 873–889.

(62) 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.

(63) 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.

(64) 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.

(65) Onishi, I.; Sunaba, S.; Yoshida, N.; Hirata, F.; Irisa, M., Role of Mg2+ Ions in DNA Hydrolysis by EcoRV, Studied by the 3D-Reference Interaction Site Model and Molecular Dynamics. *J. Phys. Chem. B.* **2018,** *122* (39), 9061-9075.

(66) Raper, A. T.; Reed, A. J.; Suo, Z., Kinetic Mechanism of DNA Polymerases: Contributions of Conformational Dynamics and a Third Divalent Metal Ion. *Chem. Rev.* **2018,** *118* (12), 6000–6025.

(67) Xie, F.; Qureshi, S. H.; Papadakos, G. A.; Dupureur, C. M., One- and two-metal ion catalysis: global single-turnover kinetic analysis of the PvuII endonuclease mechanism. *Biochemistry* **2008,** *47* (47), 12540–12550.

(68) 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 *Eco*RI family. *J. Mol. Biol.* **2009,** *393* (1), 140–160.

(69) Dalhus, B.; Arvai, A. S.; Rosnes, I.; Olsen, O. E.; Backe, P. H.; Alseth, I.; Gao, H.; Cao, W.; Tainer, J. A.; Bjoeraas, M., Structures of endonuclease V with DNA reveal initiation of deaminated adenine repair. *Nat. Struct. Mol. Biol.* **2009,** *16* (2), 138-143.

(70) Galburt, E. A.; Chevalier, B.; Tang, W.; Jurica, M. S.; Flick, K. E.; Monnat, R. J.; Stoddard, B. L., A novel endonuclease mechanism directly visualized for I-*Ppo*I. *Nat. Struct. Biol.* **1999,** *6* (12), 1096–1099.

(71) 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.

(72) Shen, B. W.; Landthaler, M.; Shub, D. A.; Stoddard, B. L., DNA binding and cleavage by the HNH homing endonuclease I-HmuI. *J. Mol. Biol.* **2004,** *342* (1), 43–56.

(73) Biertümpfel, C.; Yang, W.; Suck, D., Crystal structure of T4 endonuclease VII resolving a Holliday junction. *Nature* **2007,** *449* (7162), 616–620.

(74) Sokolowska, M.; Czapinska, H.; Bochtler, M., Hpy188I–DNA pre- and post-cleavage complexes—snapshots of the GIY-YIG nuclease mediated catalysis. *Nucleic Acids Res.* **2011,** *39* (4), 1554–1564.

(75) Freudenthal, B. D.; Beard, W. A.; Cuneo, M. J.; Dyrkheeva, N. S.; Wilson, S. H., Capturing snapshots of APE1 processing DNA damage. *Nat. Struct. Mol. Biol.* **2015,** *22* (11), 924–931.

(76) Lukacs, C. M.; Kucera, R.; Schildkraut, I.; Aggarwal, A. K., Understanding the immutability of restriction enzymes: crystal structure of BglII and its DNA substrate at 1.5 Å resolution. *Nat. Struct. Mol. Biol.* **2000,** *7* (2), 134–140.

(77) Miyazaki, K., Random DNA fragmentation with endonuclease V: application to DNA shuffling. *Nucleic Acids Res.* **2002,** *30* (24), e139.

(78) 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.

(79) Wang, Z.; Wang, H.-Y.; Feng, H., A simple and reproducible method for directed evolution: combination of random mutation with dITP and DNA fragmentation with endonuclease V. *Mol. Biotechnol.* **2013,** *53*, 49–54.

(80) DeLano, W. L., Pymol: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr* **2002,** *40* (1), 82–92.

(81) Rostkowski, M.; Olsson, M. H. M.; Søndergaard, C. R.; Jensen, J. H., Graphical analysis of pH-dependent properties of proteins predicted using PROPKA. *BMC Struct. Biol.* **2011,** *11* (1), 1–6.

(82) 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.

(83) Zgarbová, M.; Otyepka, M.; Šponer, J. i.; Mládek, A. t.; Banáš, P.; Cheatham III, T. E.; Jurecka, 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.

(84) 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.

(85) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case," Development and testing of a general amber force field" *J. Comput. Chem.* **2004,** *25* (9), 1157–1174.

(86) 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.

(87) Allnér, O.; Nilsson, L.; Villa, A., Magnesium ion–water coordination and exchange in biomolecular simulations. *J. Chem. Theory Comput.* **2012,** *8* (4), 1493–1502.

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

(89) Case, D.; Ben-Shalom, I.; Brozell, S.; Cerutti, D.; Cheatham III, T.; Cruzeiro, V.; Darden, T.; Duke, R.; Ghoreishi, D.; Gilson, M., AMBER 2018; 2018. *University of California, San Francisco* **2018**.

(90) Schmit, J. D.; Kariyawasam, N. L.; Needham, V.; Smith, P. E., SLTCAP: A simple method for calculating the number of ions needed for MD simulation. *J. Chem. Theory Comput.* **2018,** *14* (4), 1823–1827.

(91) Geronimo, I.; Vidossich, P.; Donati, E.; De Vivo, M., Computational investigations of polymerase enzymes: structure, function, inhibition, and biotechnology. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2021,** *11* (6), e1534.

(92) Castro-Amorim, J.; Oliveira, A.; Mukherjee, A. K.; Ramos, M. J.; Fernandes, P. A., Unraveling the reaction mechanism of russell's viper venom factor X activator: a paradigm for the reactivity of zinc metalloproteinases? *J. Chem. Inf. Model.* **2023,** *63*, 4056–4069.

(93) Gérard, E. F.; Mokkawes, T.; Johannissen, L. O.; Warwicker, J.; Spiess, R. R.; Blanford, C. F.; Hay, S.; Heyes, D. J.; de Visser, S. P., How is substrate halogenation triggered by the vanadium haloperoxidase from *Curvularia inaequalis*? *ACS Catal.* **2023,** *13*, 8247–8261.

(94) Shirazi, J.; Jafari, S.; Ryde, U.; Irani, M., Catalytic reaction mechanism of glyoxalase II: a quantum mechanics/molecular mechanics study. *J. Phys. Chem. B* **2023,** *127*, 4480–4495.

(95) Perera, L.; Freudenthal, B. D.; Beard, W. A.; Pedersen, L. G.; Wilson, S. H., Revealing the role of the product metal in DNA polymerase β catalysis. *Nucleic Acids Res.* **2017,** *45* (5), 2736–2745.

(96) Medina, F. E.; Jaña, G. A., QM/MM study of a VIM-1 metallo-β-lactamase enzyme: The catalytic reaction mechanism. *ACS Catal.* **2021,** *12* (1), 36–47.

(97) Balhara, R.; Chatterjee, R.; Jindal, G., A computational approach to understand the role of metals and axial ligands in artificial heme enzyme catalyzed C–H insertion. *Phys. Chem. Chem. Phys.* **2021,** *23* (15), 9500–9511.

(98) Chaturvedi, S. S.; Ramanan, R.; Hu, J.; Hausinger, R. P.; Christov, C. Z., Atomic and electronic structure determinants distinguish between ethylene formation and l-arginine hydroxylation reaction mechanisms in the ethylene-forming enzyme. *ACS Catal.* **2021,** *11* (3), 1578–1592.

(99) Paul, A.; Mishra, S., Metal–ion promiscuity of microbial enzyme DapE at its second metal-binding site. *J. Biol. Inorg. Chem.* **2021,** *26*, 569-582.

(100) Garcia-Borràs, M.; Kan, S. J.; Lewis, R. D.; Tang, A.; Jimenez-Osés, G.; Arnold, F. H.; Houk, K. N., Origin and control of chemoselectivity in cytochrome c catalyzed carbene transfer into Si–H and N–H bonds. *J. Am. Chem. Soc.* **2021,** *143* (18), 7114–7123.

(101) Bras, N. F.; Fernandes, P. A.; Ramos, M. J., QM/MM studies on the β-galactosidase catalytic mechanism: hydrolysis and transglycosylation reactions. *J. Chem. Theory Comput.* **2010,** *6* (2), 421–433.

(102) Yanagita, H.; Urano, E.; Matsumoto, K.; Ichikawa, R.; Takaesu, Y.; Ogata, M.; Murakami, T.; Wu, H.; Chiba, J.; Komano, J., Structural and biochemical study on the inhibitory activity of derivatives of 5-nitro-furan-2-carboxylic acid for RNase H function of HIV-1 reverse transcriptase. *Bioorg. Med. Chem.* **2011,** *19* (2), 816–825.

(103) Elsässer, B.; Fels, G., Atomistic details of the associative phosphodiester cleavage in human ribonuclease H. *Phys. Chem. Chem. Phys.* **2010,** *12* (36), 11081–11088.

(104) Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Petersson, G.; Nakatsuji, H., Gaussian 16 Rev. B. 01, Wallingford, CT. **2016**.

(105) Wu, H.; Lima, W. F.; Crooke, S. T., Investigating the structure of human RNase H1 by site-directed mutagenesis. *J. Biol. Chem.* **2001,** *276* (26), 23547–23553.

(106) Feng, H.; Klutz, A. M.; Cao, W., Active site plasticity of endonuclease V from Salmonella typhimurium. *Biochemistry* **2005,** *44* (2), 675–683.

(107) Gong, S.; Yu, H. H.; Johnson, K. A.; Taylor, D. W., DNA unwinding is the primary determinant of CRISPR-Cas9 activity. *Cell Rep.* **2018,** *22* (2), 359–371.

(108) Eastberg, J. H.; Eklund, J.; Monnat, R.; Stoddard, B. L., Mutability of an HNH nuclease imidazole general base and exchange of a deprotonation mechanism. *Biochemistry* **2007,** *46* (24), 7215–7225.

(109) Jeltsch, A.; Alves, J.; Wolfes, H.; Maass, G.; Pingoud, A., Substrate-assisted catalysis in the cleavage of DNA by the *Eco*RI and *Eco*RV restriction enzymes. *Proc. Natl. Acad. Sci. U.S.A.* **1993,** *90* (18), 8499–8503.

(110) 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), 6498–502.

(111) Feng, H.; Dong, L.; Cao, W., Catalytic mechanism of endonuclease V: a catalytic and regulatory two-metal model. *Biochemistry* **2006,** *45* (34), 10251–10259.

(112) Flick, K. E.; Jurica, M. S.; Monnat, R. J.; Stoddard, B. L., DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-*Ppo*I. *Nature* **1998,** *394* (6688), 96–101.

Chapter 7: Conclusions and Future Directions

7.1. Summary

Phosphodiester bond hydrolysis in nucleic acids is a ubiquitous reaction facilitated by enzymes called nucleases, which often use metal ions to achieve catalytic function. However, the mechanism for bond cleavage by some enzymes is still unclear. This thesis has utilized computational techniques to gain an atomic-level understanding of the phosphodiester bond cleavage reaction catalyzed by nucleases, specifically focusing on APE1, I-*Ppo*I, and EndoV. These enzymes are particularly interesting since they either invoke a single metal ion for catalysis, which conflicts with the generally proposed two-metal mediated mechanism for most nucleases, or the metal-dependence is unknown. Although the role and number of metal ions required for phosphodiester bond cleavage are highly controversial issues in the nuclease literature, this thesis has made a significant contribution to the understudied field of one-metal dependent nucleases, supporting the proposal that one metal is enough for catalyzing this difficult bond cleavage reaction. This chapter summarizes the major findings of each chapter and provides details of possible future research avenues.

7.2. Contributions from Thesis

In Chapter 2, quantum mechanics (QM) cluster models were used to provide atomic-level insights into the single-metal mediated phosphodiester bond cleavage of nucleic acids using APE1 as a prototypical example. Initially, the smallest model containing 69 atoms was used to illustrate that common functionals (e.g., B3LYP-D3 and

M06-2X) can be combined with the 6-31G(d,p) basis set to provide a reliable structural description of the phosphodiester bond cleavage reaction. Subsequently, the model size was systematically increased by altering residue truncation points and including additional amino acids to highlight that β-carbon truncation points or small models containing only a few amino acids can introduce model strain and lead to incorrect metal coordination geometry. Nevertheless, the continuum solvent calculations revealed that accounting for the broader enzymatic environment had a negligible impact on energetics of the P–O bond cleavage reaction. Overall, Chapter 2 highlights that a QM cluster model must contain all key residues such as the general base and general acid, residues that provide substrate charge stabilization, and amino acids that help maintain the octahedral metal coordination geometry to achieve an accurate structural description of one-metal mediated phosphodiester bond hydrolysis. In APE1, D240 initiates the reaction by activating the nucleophilic water, a metal-activated water facilitates leaving group departure by protonating the leaving group, and a positively charged H309 stabilizes the charge on the substrate (Figure 7.1a).

Using the DFT functional and basis set combination recommended in Chapter 2, a quantum mechanical−molecular mechanics (QM/MM) approach was used in Chapter 3 to provide a structural rationalization for the experimentally-reported impact of metal substitution on the catalytic activity of APE1.[1-3] Although the catalytic mechanism of APE1 is well established,[4-11] a chemical explanation for the metal dependence of APE1 was missing prior to my work. My calculations reproduced the experimentally-observed trend in the reaction barriers ($Mg^{2+} > Mn^{2+} > Ni^{2+} > Zn^{2+} >> Ca^{2+}$)[1-3] and the structural analysis

revealed an interplay of several factors that allow APE1 to remain catalytically active in the presence of various metals. For instance, Mn^{2+} results in subtle changes to the active site geometry, while Ni^{2+} can adopt both square-planar and octahedral binding architectures in the APE1 active site, and the Zn^{2+} coordination geometry changes from octa- to penta-coordination over the course of the reaction. Nevertheless, the enzymatically feasible barrier resulting from substitution of Mg^{2+} with these transition metals highlights the promiscuity of APE1 towards different metal ions. Furthermore, this chapter provided clarification on how the presence of Ca^{2+} changes the metal coordination geometry from octa- to hepta-coordination so that the water nucleophile directly coordinates to the metal, which in turn changes the rate-determining step of the reaction and results in a prohibitively high energy barrier and thus inhibition of APE1. To the best of my knowledge, this is the first computational study that reproduces and provides a chemical explanation for the experimentally-observed metal dependence of a single-metal dependent nuclease, which provides additional support for the proposal that a single metal is enough for catalysis by some enzymes.

In Chapter 4, a mixed QM cluster and QM/MM approach was used to investigate the catalytic pathway used by another single-metal dependent enzyme, namely I-*Ppo*I. Although a phosphodiester bond cleavage mechanism has been proposed for I-*Ppo*I,[12] the role of the metal and key active site amino acids needed to be clarified in addition to the required computational verification of the experimentally-proposed reaction pathway. Furthermore, although I-*Ppo*I is known to utilize one metal for catalysis like APE1, the water nucleophile was proposed to be activated by a neutral histidine (H98) in I-*Ppo*I in

contrast to a negatively charged aspartate residue (D240) in APE1, which raises questions about the impact of the active site architecture on this relatively new one metal mediated phosphodiester bond cleavage chemistry. In the experimentally-proposed I-*Ppo*I mechanism of action,[12] a single metal facilitates leaving group departure through direct metal coordination to the substrate and leaving group protonation by a metal-ligated water. However, my smallest QM cluster model and a QM/MM model revealed that the previously proposed pathway is infeasible due to geometrical constraints imposed by the I-*Ppo*I active site. Therefore, QM cluster models of varying sizes were used to characterize a similar pathway involving indirect (water-mediated) coordination to the leaving group; however, this mechanism was not energetically feasible regardless of the model size. QM cluster calculations indicate that direct metal coordination to the leaving group is preferred. QM/MM calculations further confirm the preferred mechanism in the presence of the solvated enzyme–DNA complex. In the proposed preferred phosphodiester bond hydrolysis pathway, H98 initiates the reaction by activating the water nucleophile while the metal directly coordinates to the leaving group is preferred and provides substrate charge stabilization (Figure 7.1b). Subsequently, QM/MM calculations were used to confirm the preferred mechanism in the presence of the whole enzyme–DNA complex. Indeed, the calculated QM/MM Gibbs activation barrier for the favored pathway is consistent with the experimental barrier for I-*Ppo*I mediated phosphodiester bond cleavage.[13-14] *In silico* H98A mutational results further support the proposed mechanism.[14-15] My newly proposed catalytic mechanism is in agreement with the experimental structural,[12, 16] kinetic,[13-14] and mutational data.[14-15] Interestingly, this pathway was previously shown to be infeasible for

APE1.[6] Instead, due to a strong negatively charged general base (D240) coupled with efficient charge stabilization from a positively charged H309, the single metal in APE1 plays a role in charge stabilization through indirect coordination to the substrate and protonates the leaving group using a metal-activated water. In contrast, the weaker general base (H98) of I-*Ppo*I combined with the absence of a positively charged residue close to the scissile phosphate forces the single metal to play a more active role in both leaving group departure and substrate stabilization. This highlights how a single metal can play different roles to facilitate the same P–O bond cleavage in nucleic acids due to subtle differences in the active sites of the single-metal dependent nucleases.

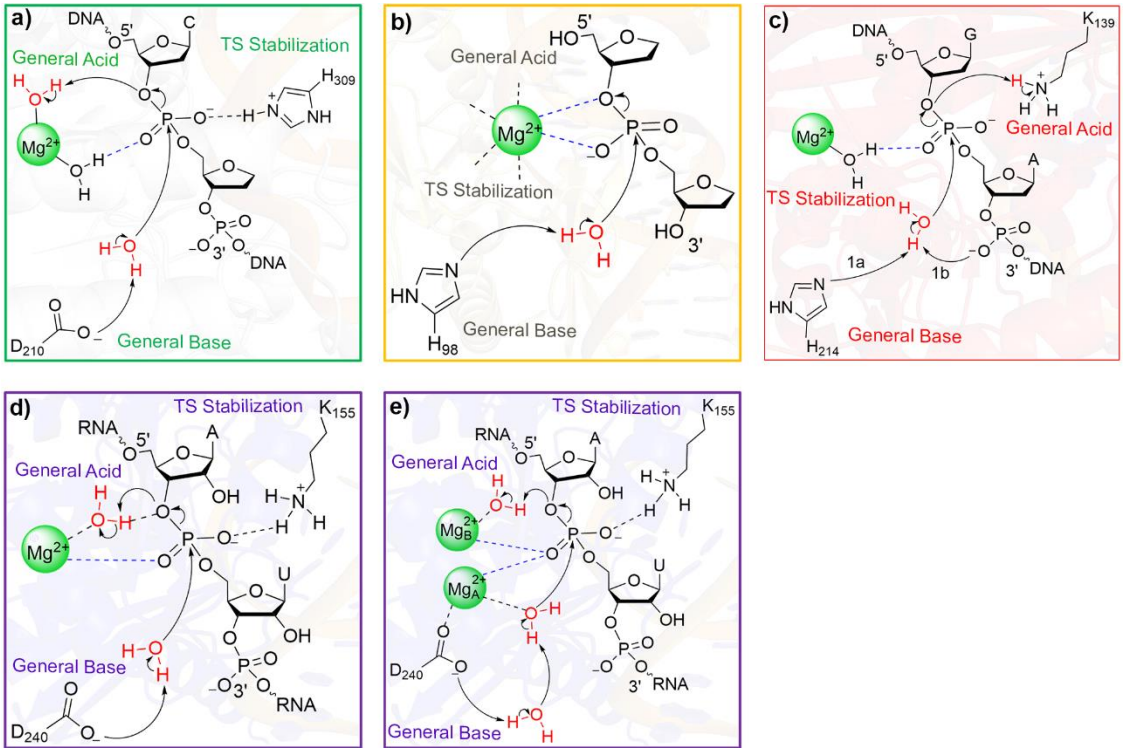

Figure 7.1. The phosphodiester bond cleavage pathways proposed in this thesis, highlighting the residues playing the roles of the general base, the general acid, and stabilizing the TS for a) APE1, b) I-*Ppo*I, c) bacterial EndoV, and d) one-metal or e) two-metal mediated human EndoV.

Chapter 5 used MD simulations and QM/MM calculations to provide the first proposed mechanism of action for bacterial EndoV. Bacterial EndoV contains a histidine (H214) in an equivalent position with respect to the substrate as H98 in I-*Ppo*I which was proposed to be the general base in Chapter 4. On the other hand, bacterial EndoV has a positively charged residue (K139) near the substrate similar to the positively charged H309 in APE1, which was proposed to provide charge stabilization.[6] The similarities of the EndoV active site to both I-*Ppo*I and APE1 brings up the question of whether the role of the metal is the same as either enzyme or different yet again. Despite the abundance of experimental structural,[17] kinetic,[18] and mutational data,[19-20] a catalytic mechanism for bacterial EndoV had not been proposed in the literature before my work. Chapter 5 compares eight distinct catalytic pathways to reveal that the combinations of general base, general acid and metal-binding architecture previously proposed for other metal-dependent nucleases are not feasible for EndoV. Instead, a mechanism using a previously unseen combination of the general base and general acid for a one-metal dependent endonuclease was characterized for EndoV. Specifically, K139 protonates the leaving group, either H214 or a substrate phosphate group activates the water nucleophile, and a metal-coordinated water stabilizes the charged substrate (Figure 7.1c). My proposed catalytic mechanisms are fully consistent with experimental kinetic,[18] structural,[17] and mutational data.[19-20]

Although the catalytic pathway involving direct bidentate metal coordination to the substrate was revealed to be preferred for I-*Ppo*I in Chapter 4, the active site composition of bacterial EndoV appears to prevent the metal from efficiently functioning in the same capacity. In contrast, the preferred metal-binding architecture (indirect metal–substrate

coordination) is same in EndoV and APE1. Nevertheless, there are differences in other key residues that result in different catalytic mechanisms for the two enzymes. Specifically, APE1 uses a powerful general base (D240) to activate the water nucleophile and a metal-coordinated water to facilitate leaving group departure. However, our calculations demonstrated that regardless of the identity of the general base (H214 or a substrate phosphate moiety), a metal-ligated water could not effectively stabilize the leaving group for bacterial EndoV. Instead, due to the presence of a positively charged lysine residue in close proximity to the substrate, the role of facilitating leaving group departure is taken by K139. Collectively, the comparison of bacterial EndoV, I-*Ppo*I, and APE1 further highlights that diversity in the active site framework allows single-metal dependent endonucleases to achieve the same challenging phosphodiester backbone hydrolysis of nucleic acids in diverse ways, with the metal playing different roles (i.e., activating a metal-coordinated water to protonate the leaving group, promoting leaving group departure through direct coordination, or providing stabilization to the charged substrate).

Finally, MD simulations and QM/MM calculations were used to provide the first detailed description of the mechanism of action of human EndoV in Chapter 6. A catalytic mechanism for the phosphodiester bond cleavage has never been proposed for human EndoV, at least in part due to the absence of the substrate or the metal cofactor/s in the X-ray crystal structure of the enzyme.[20] In fact, the role or number of metal/s required for catalysis and the identity of key amino acids involved in the reaction was unknown prior to my work. MD simulations were initially used to reveal the first structure of the human EndoV–substrate complex in the presence of metal ion/s. Subsequently, QM/MM

calculations were used to characterize four unique pathways that vary in metal number (one versus two) and modes of metal–substrate coordination (direct versus indirect (water-mediated)). This chapter reveals that unlike bacterial EndoV, which uses a single metal,[17] and mouse EndoV, which uses two metals,[20] human EndoV can use either one or two metals, with the final number utilized likely depending on the metal cofactor concentration (Figure 7.1d–e). Besides uncovering the roles for key active site amino acids (D240 and K155), our calculations emphasize that, while a single metal is essential for human EndoV-mediated P–O bond cleavage chemistry, the enzyme can benefit from using two metals due to the presence of two suitable active site metal-binding positions. Nevertheless, this chapter further reinforces the proposal made for APE1, I-*Ppo*I, and bacterial EndoV in Chapters 2 – 5 of this thesis that, in presence of a strong general base in the active site, one metal is enough for a nuclease to facilitate the extremely challenging P–O bond cleavage.

The comparison of human EndoV to APE1 indicates that although EndoV has an aspartate in an equivalent position of the general base (D240 in human EndoV and D210 in APE1), the insufficient charge stabilization on the scissile phosphate by the weaker acid (lysine (K155) versus protonated histidine (H309)) necessitates direct metal coordination to the substrate to promote human EndoV activity (Figure 7.1). This emphasizes how similar active sites can require a different role of the metal to mediate phosphodiester bond hydrolysis. On the other hand, the stronger general base in human (aspartate, D240) compared to bacterial (histidine, H214) EndoV obviates the need for a more active role from lysine (i.e., substrate charge stabilization (K155 in human EndoV) rather than the general acid (K139 in bacterial EndoV)). This highlights how a slight difference in the

identity and placement of amino acids in the EndoV active site can make the enzyme prefer one pathway over the other.

Overall, this thesis reveals the astonishing complexity of enzyme-catalyzed phosphodiester bond cleavage reactions emerging from the delicate interplay of several factors such as the identity and placement of key catalytic residues, the metal–substrate binding architecture, and the number and identity of the metal/s in the active site. This thesis represents an important step towards resolving the long-standing one-metal versus two-metal dispute in the literature especially for enzymes that exhibit conflicting experimental results. Further details of possible future research avenues are discussed in the next section.

7.3. Future Directions

The computational methodology developed in Chapter 2 of this thesis could be used to investigate other nucleases that contain a single metal in the active site such as *Vibrio vulnificus* nuclease (Vvn). Vvn is a nuclease that prevents uptake of foreign DNA into cells and is capable of digesting both DNA and RNA.[21] Similar to APE1, crystal structures of Vvn bound to a DNA substrate display different metal coordination environments in the reactant complex (RC) and product complex (PC) of the H80A mutant (Figure 7.2).[22] Specifically, the metal coordination to the non-bridging oxygen of the scissile phosphate moiety is indirect in the RC and direct in the PC (Figure 7.2b). A phosphodiester bond cleavage mechanism has been proposed based on the crystal structure of the PC[22] in which a metal-coordinated water protonates the leaving group and the metal stabilizes the charge

on the substrate. However, this proposal conflicts with the crystal structure of the RC, which displays indirect metal–substrate coordination. This suggests that an alternate chemical pathway involving a different metal-binding configuration than the PC might be possible.

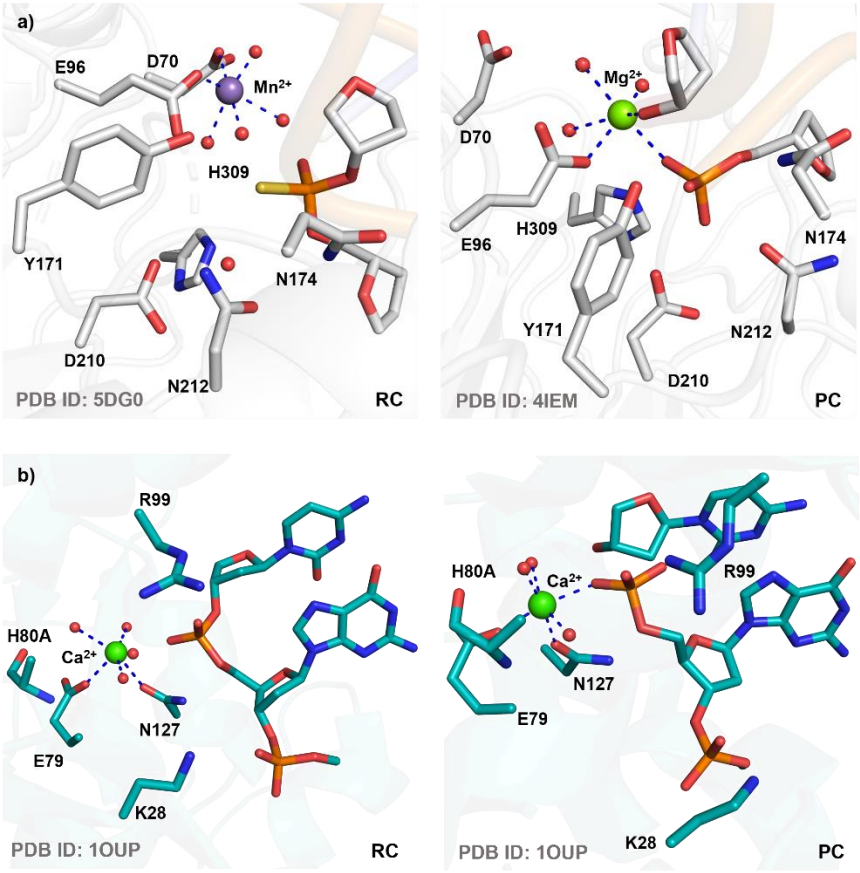

Figure 7.2. Active sites from X-ray crystal structures of the a) Mn^{2+}-containing thio-substituted RC (left) and Mg^{2+}-containing PC (right) for the wild-type APE1, and b) Ca^{2+}-containing RC (left) and PC (right) for the H80A mutant of Vvn.

Based on combined MD simulations and QM/MM calculations, a phosphodiester bond cleavage pathway for Vvn that involves bidentate metal–substrate coordination has

been proposed in which H80 activates the nucleophilic water and the metal promotes leaving group departure through direct coordination and charge stabilizes the substrate.[23] However, this proposal is not consistent with the X-ray crystal structures of the RC and PC of Vvn (Figure 7.2b), neither of which display the bidentate metal–substrate binding configuration suggested in this study. Additionally, a low accuracy semi-empirical method (PM3) was used for the QM region in the QM/MM calculations. Furthermore, the substrate for the starting model was obtained from the crystal structure of the I-*Ppo*I–DNA complex (PDB ID: 1CYQ, Figure 7.3a) rather than using the substrate in the Vvn crystal structures.

Although both Vvn and I-*Ppo*I contain a histidine residue (H98 in I-*Ppo*I and H75 in Vvn) in a similar position with respect to the substrate to activate the water nucleophile, the identity and orientation of the remaining active site residues vary significantly between the enzymes. For example, two amino acids directly coordinate to the metal in Vvn (E79 and N127), while only one coordinates to the metal for I-*Ppo*I (N119). Furthermore, substrate charge stabilization was provided only by the metal in I-*Ppo*I, while Vvn has an additional amino acid (R99) within hydrogen-bonding distance of the scissile phosphate in the PC (Figure 7.2b). These factors coupled with the crystal structures of both RC and PC bound to a substrate being available warrants investigation of the catalytic mechanism of Vvn in a more systematic manner (Figure 7.2b). First, the computational methodology (QM cluster) used in Chapter 2 can be applied to investigate different metal-binding architectures observed in the crystal structures of Vvn as well as I-*Ppo*I. The preferred metal-binding configuration in the active site can subsequently be used to perform QM/MM calculations in the presence of the full enzyme–substrate complex. Finally, comparisons to other single-

metal mediated nucleases can be made to determine if the single metal plays a similar role. Interestingly, unlike APE1 which is inhibited by Ca^{2+},[24] this would only be the first thorough computational study of a nuclease that is catalytically active in the presence of Ca^{2+}, and a comparison between catalytic pathways mediated by Vvn and APE1 could expose the reasons behind the divergent metal dependence.

Chapter 4 investigates the catalytic mechanism of I-*Ppo*I, a member of the His-Cys box homing endonuclease (HE) family. HEs are classified into four families that are distinguished by conserved His-Cys box, H-N-H, GIY-YIG, or LAGLIDADG motifs.[25-26] The crystal structure of wild-type I-HmuI (a member of the H-N-H family of HEs) in complex with cleaved DNA shows direct coordination of the metal to a non-bridging and the O3′-bridging oxygen of the scissile phosphate of the substrate (Figure 7.3b).[27] A catalytic mechanism has been proposed based on the structural data in which a histidine (H75) activates the water nucleophile, while a single metal directly participates in substrate charge stabilization and leaving group departure,[27] similar to the mechanism proposed for I-*Ppo*I in Chapter 4. However, I-HmuI contains a tyrosine residue (Y39) within hydrogen-bonding distance of the scissile phosphate to provide additional charge stabilization and the corresponding residue in I-*Ppo*I is absent. Moreover, R61 that was proposed to stabilize the 3′-downstream phosphate in I-*Ppo*I is replaced by Q54 in I-HmuI.[27] Since this thesis has demonstrated how small changes in the active site composition of enzymes can result in different catalytic mechanisms, QM/MM calculations can be performed on I-HmuI to understand whether the catalytic mechanism between His-Cys box and H-N-H family is

maintained despite differences in the identity of various amino acids involved in the reaction.

Hpy188I is a member of GIY-YIG HE family and displays bidentate metal–substrate coordination in the Na^+-substituted RC (Figure 7.3c)[28] as seen for I-*Ppo*I (Figure 7.3a). However, the identity and orientation of the remaining active site residues is vastly different than I-*Ppo*I (Figure 7.3c).[29] Indeed, Y63 was suggested to be the general base that abstracts a proton from the water nucleophile with the help of another tyrosine (Y88), while K73 and R84 are in position to provide charge stabilization to the scissile phosphate moiety. The leaving group departure was proposed to be aided either through direct metal coordination or protonation by bulk water. Computational characterization of the proposed mechanisms could reveal the preferred catalytic pathway for Hpy188I. Regardless, the mechanism mapping could get complicated as the crystal structure of the Na^+-substituted RC shows direct coordination of Na^+ to E149, while the wild-type PC of Hpy188I displays direct coordination between Ca^{2+} and a non-bridging oxygen of the scissile phosphate, while H76 is the metal binding residue rather than E149 (Figure 7.3c). These conflicting metal-binding arrangements in the RC and PC warrant a deeper investigation of different metal configurations and their impact on the catalytic mechanism of Hpy188I using the QM cluster followed by QM/MM methodology applied in Chapter 4.

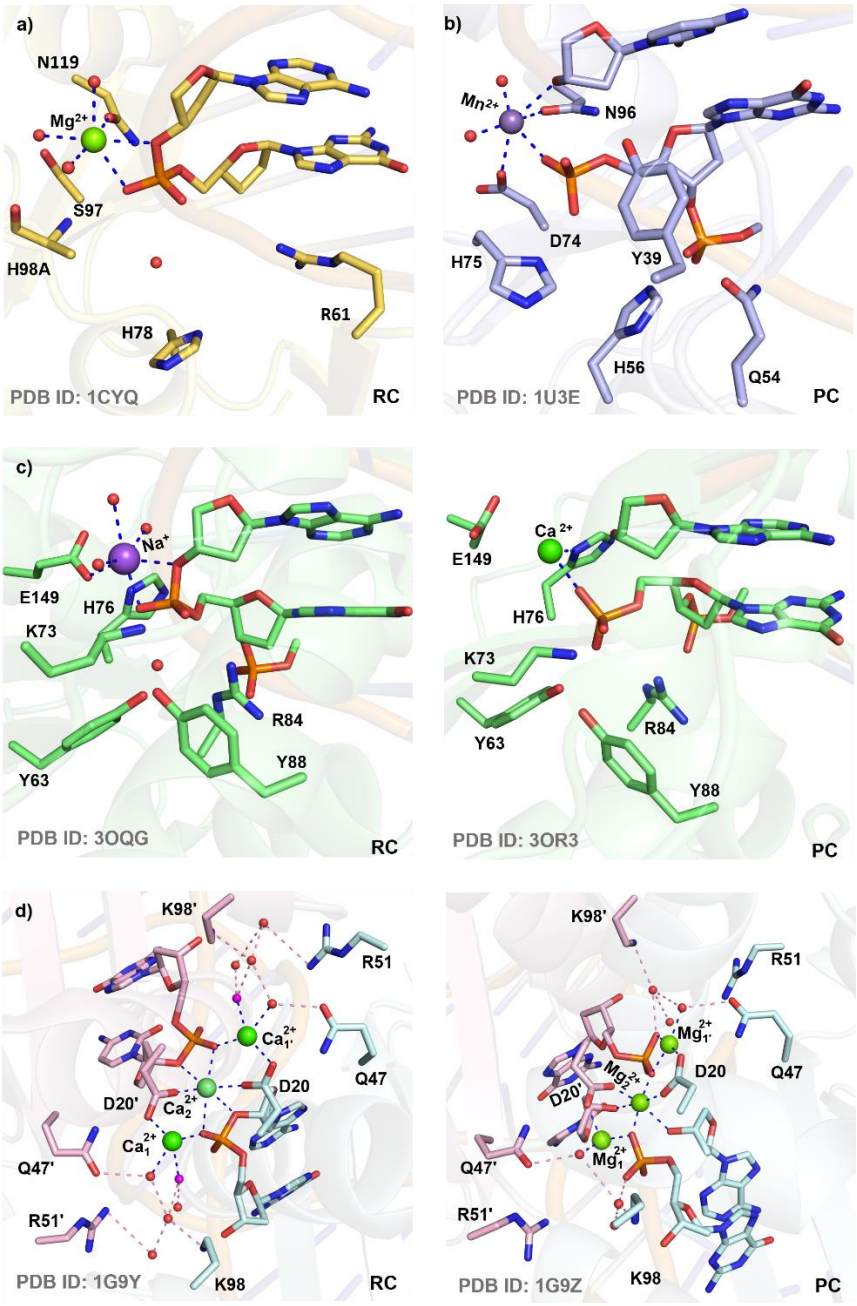

Figure 7.3. Active sites from X-ray crystal structures of the a) Mg^{2+}-containing RC for the H98A mutant of I-*Ppo*I, b) Mn^{2+}-containing PC for the wild-type I-HmuI, c) Na^{+}-substituted RC (left) and Ca^{2+}-containing PC (right) for the wild-type Hpy188I, and d) Ca^{2+}- substituted RC (left) and Mg^{2+}-containing PC (right) for the wild-type I-CreI.

Experimental kinetic studies have revealed that three enzymes from different HE families (I-*Ppo*I, I-HmuI, and Hpy188I) exhibit different patterns for metal dependence.

Specifically, in the presence of Ca^{2+}, Hpy188I is fully active, while I-*Ppo*I shows reduced catalytic activity [13, 28] and I-HmuI is completely inactive.[27] Since the computational methodology in Chapter 3 successfully rationalized the experimentally reported metal dependence of the single-metal dependent APE1, the same approach can be adopted to provide insights into how the same metal (Ca^{2+}) results in such drastically different outcomes in the catalytic activity of these three single-metal mediated enzymes.

Unlike the three above-mentioned HE families that depend on a single metal for the chemical step, the members of the LAGLIDADG HE family (e.g., I-CreI) have completely different active site compositions and metal ion stoichiometry requirements.[30] The crystal structures of Ca^{2+}-substituted RC and Mg^{2+}-containing PC of I-CreI complexed with DNA have been solved.[31-32] Unlike most two-metal dependent endonucleases,[33] the structures of I-CreI illustrate a unique variant of a two-metal endonuclease mechanism where each active site contains one metal ion and a second metal is shared between the two active sites to generate double-strand DNA breaks (Figure 7.3d).[31-32] Although there is no amino acid in position to act as a general base, an extensive network of ordered water molecules (positioned by Q47, R51, and K98) is present in the crystal structure and could assist in deprotonating the metal-ligated water nucleophile. Experimental mutational studies have shown that residues Q47, R51, and K98 are important for catalytic activity and the cleavage of the phosphodiester bonds in the two active sites happens simultaneously.[32] This unique proposal of simultaneous cleavage of two P–O bonds in two active sites that share a common metal makes I-CreI particularly interesting to study. QM/MM calculations could

be utilized to elucidate the catalytic mechanism of I-CreI and reveal how the conventional two-metal mediated mechanism is modified to achieve catalysis.

Although Chapter 5 has provided some valuable insights into the role of Mg^{2+} in the EndoV-catalyzed phosphodiester bond cleavage, future work that examines the impact of changing the metal identity on the structure and energetics of the reaction is required. For example, experimental studies have shown that bacterial EndoV is active in the presence of Mg^{2+}, Mn^{2+}, Ni^{2+}, and Co^{2+}, but completely inactive with Zn^{2+} and Ca^{2+}.[34] Since APE1 and bacterial EndoV both prefer an indirect metal–substrate binding arrangement, the methodology used in Chapter 3 could be applied to understand the metal dependence in bacterial EndoV.

Similar to the nucleases considered in this thesis, NgoMIV is another restriction endonuclease from *Neisseria gonorrhoeae* strain M that uses Mg^{2+} ions to hydrolyze the phosphodiester bond in DNA.[35] However, the crystal structures of a Ca^{2+}-substituted RC and a Mg^{2+}-containing PC of the wild-type Ngo-MIV complexed with DNA reveal a different number of metal ions in the active site (Figure 7.4). Specifically, X-ray crystallographic data shows a single Ca^{2+} in active site of the RC and two Mg^{2+} ions in the PC analogue of NgoMIV. Unlike the roles proposed for lysine in Chapters 5 and 6 as a general acid (K139, bacterial EndoV) or providing substrate charge stabilization (K155, human EndoV), the lysine residue in NgoMIV (K187) is within hydrogen-bonding distance of the water nucleophile. Based on the crystal structure of the NgoMIV PC coupled with structural similarities to the active site of two-metal mediated BamHI, NgoMIV has been conjectured to follow the two-metal mediated mechanism in which K187 activates the

water nucleophile and a metal-activated water promotes leaving group departure in the P–O bond cleavage (Figure 7.4).[35] However, the mechanistic conclusions based on the PC can be incorrect since metal migration and associated active site rearrangements are common before the formation of the final product in nucleases.[10, 36-40] Furthermore, the mechanistic proposal is in conflict with the X-ray crystal structure of the NgoMIV RC (Figure 7.4a), which displays a single metal in the active site. The above-mentioned factors coupled with lack of experimental kinetic data leaves large gaps in our knowledge about the number and role of metal ions involved in NgoMIV activity. The contentious issue of the number of metal ions required for catalysis can be resolved by performing a comprehensive investigation using one-metal and two-metal containing models with different metal-binding configurations. The computational methodology applied in Chapter 6 could be used to determine whether one metal is enough for NgoMIV to achieve catalysis. Further comparison to my work on bacterial and human EndoV will broaden our understanding of the different roles lysine can play in the family of enzymes that cleave phosphodiester bonds in nature.

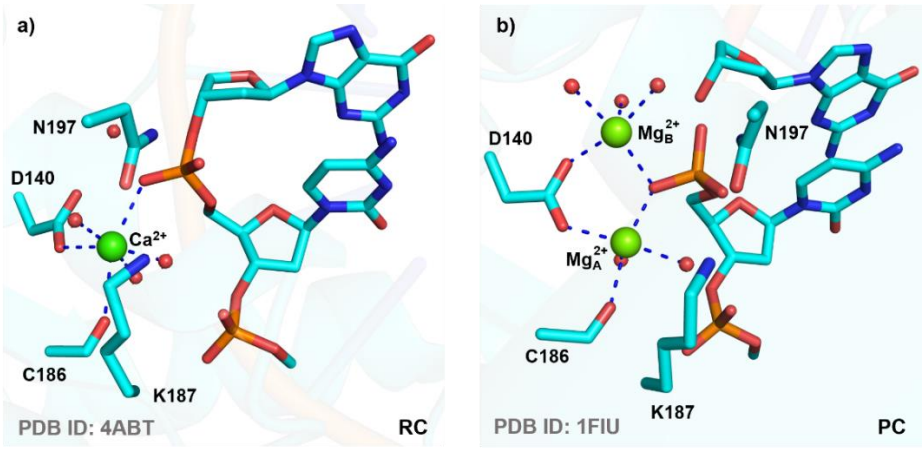

Figure 7.4. Active sites from X-ray crystal structures of the a) Ca^{2+}-substituted and b) Mg^{2+}-containing PC for the wild-type NgoMIV.

Although the current thesis has provided valuable information on single-metal and two-metal mediated nucleases, experimental and theoretical studies in recent years have suggested that a third metal ion may be actively involved in catalysis.[41] However, since the third metal ion in various nucleases was captured during different stages of catalysis, it has become a topic of debate over how or when this metal is recruited from the bulk solvent and what is its exact role in catalysis. For instance, a QM/MM MD study has suggested that the third metal ion acts as an exit shuttle that promotes leaving group departure after DNA hydrolysis by human exonuclease 1.[42] In contrast, another QM/MM MD study on endonuclease IV has revealed that structural rearrangements over the course of phosphodiester bond cleavage in DNA allow the third metal to play multiple mechanistic roles such as facilitating activation of the nucleophilic water in the RC and charge stabilization in the transition state (TS), and stabilization of the leaving group in the PC .[43] Interestingly, an *in crystallo* reaction study on mouse EndoV has revealed that a third metal ion appears along with product formation.[20] A deeper investigation of human EndoV using

the QM/MM MD approach might provide useful insights into the role of a third metal ion in nucleic acid backbone cleavage. Overall, while this thesis contributes to understanding the function of select nucleases, further research is required to decipher the chemistry of other metallonucleases, which will help clarify the important roles played by these enzymes in biology.

7.4. Final Remarks

In conclusion, this thesis provides a comprehensive computational investigation of endonucleases that use either one or two metals to efficiently catalyze the phosphodiester bond hydrolysis in nucleic acids. Specifically, QM calculations, MD simulations, and QM/MM calculations were used to provide atomic-level details of the P–O bond cleavage reaction and reveal a slightly different role of the metal in each enzyme. The mechanistic details uncovered by this work will open the door for applications of nucleases in many areas, including disease diagnostics, genetic engineering, and therapeutic solutions.

7.5. References

(1) Barzilay, G.; Mol, C. D.; Robson, C. N.; Walker, L. J.; Cunningham, R. P.; Tainer, J. A.; Hickson, I. D., Identification of critical active-site residues in the multifunctional human DNA repair enzyme HAP1. *Nat. Struct. Mol. Biol.* **1995,** *2* (7), 561-568.

(2) Schermerhorn, K. M.; Delaney, S., A chemical and kinetic perspective on base excision repair of DNA. *Acc. Chem. Res.* **2015,** *47*, 1238.

(3) Miroshnikova, A. D.; Kuznetsova, A. A.; Vorobjev, Y. N.; Kuznetsov, N. A.; Fedorova, O. S., Effects of mono- and divalent metal ions on DNA binding and catalysis of human apurinic/apyrimidinic endonuclease 1. *Mol. BioSyst.* **2016,** *12* (5), 1527-1539.

(4) Tsutakawa, S. E.; Shin, D. S.; Mol, C. D.; Izumi, T.; Arvai, A. S.; Mantha, A. K.; Szczesny, B.; Ivanov, I. N.; Hosfield, D. J.; Maiti, B.; Pique, M. E.; Frankel, K. A.; Hitomi, K.; Cunningham, R. P.; Mitra, S.; Tainer, J. A., Conserved structural chemistry for incision activity in structurally non-homologous apurinic/apyrimidinic endonuclease APE1 and endonuclease IV DNA repair enzymes. *J. Biol. Chem.* **2013,** *288* (12), 8445-8455.

(5) Freudenthal, B. D.; Beard, W. A.; Cuneo, M. J.; Dyrkheeva, N. S.; Wilson, S. H., Capturing snapshots of APE1 processing DNA damage. *Nat. Struct. Mol. Biol.* **2015,** *22* (11), 924-931.

(6) 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.

(7) 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.

(8) Whitaker, A. M.; Flynn, T. S.; Freudenthal, B. D., Molecular snapshots of APE1 proofreading mismatches and removing DNA damage. *Nat. Commun.* **2018,** *9* (1), 399.

(9) Kim, W.-C.; Berquist, B. R.; Chohan, M.; Uy, C.; Wilson, D. M.; Lee, C. H., Characterization of the endoribonuclease active site of human Apurinic/Apyrimidinic Endonuclease 1. *J. Mol. Biol.* **2011,** *411* (5), 960-971.

(10) He, H.; Chen, Q.; Georgiadis, M. M., High-resolution crystal structures reveal plasticity in the metal binding site of apurinic/apyrimidinic endonuclease 1. *Biochemistry* **2014,** *53* (41), 6520-6529.

(11) Kanazhevskaya, L. Y.; Koval, V. V.; Lomzov, A. A.; Fedorova, O. S., The role of Asn-212 in the catalytic mechanism of human endonuclease APE1: Stopped-flow kinetic study of incision activity on a natural AP site and a tetrahydrofuran analogue. *DNA Repair* **2014,** *21*, 43-54.

(12) Galburt, E. A.; Chevalier, B.; Tang, W.; Jurica, M. S.; Flick, K. E.; Monnat, R. J.; Stoddard, B. L., A novel endonuclease mechanism directly visualized for I-*Ppo*I. *Nat. Struct. Biol.* **1999,** *6* (12), 1096-1099.

(13) Wittmayer, P. K.; Raines, R. T., Substrate binding and turnover by the highly specific I-*Ppo*I endonuclease. *Biochemistry* **1996,** *35* (3), 1076-1083.

(14) Mannino, S. J.; Jenkins, C. L.; Raines, R. T., Chemical mechanism of DNA cleavage by the homing endonuclease I-*Ppo*I. *Biochemistry* **1999,** *38* (49), 16178-16186.

(15) Eastberg, J. H.; Eklund, J.; Monnat, R.; Stoddard, B. L., Mutability of an HNH nuclease imidazole general base and exchange of a deprotonation mechanism. *Biochemistry* **2007,** *46* (24), 7215-7225.

(16) Flick, K. E.; Jurica, M. S.; Monnat, R. J.; Stoddard, B. L., DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-*Ppo*I. *Nature* **1998,** *394* (6688), 96-101.

(17) Dalhus, B.; Arvai, A. S.; Rosnes, I.; Olsen, Ø. E.; Backe, P. H.; Alseth, I.; Gao, H.; Cao, W.; Tainer, J. A.; Bjørås, M., Structures of endonuclease V with DNA reveal initiation of deaminated adenine repair. *Nat. Struct. Mol. Biol.* **2009,** *16*, 138-143.

(18) Feng, H.; Klutz, A. M.; Cao, W., Active site plasticity of endonuclease V from Salmonella typhimurium. *Biochemistry* **2005,** *44* (2), 675-683.

(19) Feng, H.; Dong, L.; Cao, W., Catalytic mechanism of endonuclease V: a catalytic and regulatory two-metal model. *Biochemistry* **2006,** *45* (34), 10251-10259.

(20) 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.

(21) Wu, S. I.; Lo, S. K.; Shao, C. P.; Tsai, H. W.; Hor, L. I., Cloning and characterization of a periplasmic nuclease of Vibrio vulnificus and its role in preventing uptake of foreign DNA. *Appl. Environ. Microbiol.* **2001,** *67* (1), 82-8.

(22) 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.

(23) Bueren-Calabuig, J. A.; Coderch, C.; Rico, E.; Jimenez-Ruiz, A.; Gago, F., Mechanistic insight into the catalytic activity of ββα-metallonucleases from computer simulations: *Vibrio vulnificus* periplasmic nuclease as a test case. *Chembiochem* **2011,** *12* (17), 2615-2622.

(24) 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.

(25) Belfort, M.; Roberts, R. J., Homing endonucleases: keeping the house in order. *Nucleic Acids Res.* **1997,** *25 17*, 3379-88.

(26) Jurica, M.; Stoddard, B., Homing endonucleases: structure, function and evolution. *Cell. Mol. Life Sci.* **1999,** *55*, 1304-1326.

(27) Shen, B. W.; Landthaler, M.; Shub, D. A.; Stoddard, B. L., DNA binding and cleavage by the HNH homing endonuclease I-HmuI. *J. Mol. Biol.* **2004,** *342* (1), 43-56.

(28) Sokolowska, M.; Czapinska, H.; Bochtler, M., Hpy188I–DNA pre- and post-cleavage complexes—snapshots of the GIY-YIG nuclease mediated catalysis. *Nucleic Acids Res.* **2011,** *39* (4), 1554-1564.

(29) Kaminska, K. H.; Kawai, M.; Boniecki, M.; Kobayashi, I.; Bujnicki, J. M., Type II restriction endonuclease R. Hpy188I belongs to the GIY-YIG nuclease superfamily, but exhibits an unusual active site. *BMC Struct. Biol.* **2008,** *8* (1), 1-14.

(30) Chevalier, B. S.; Stoddard, B. L., Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res.* **2001,** *29* (18), 3757-3774.

(31) Chevalier, B. S.; Monnat, R. J.; Stoddard, B. L., The homing endonuclease I-CreI uses three metals, one of which is shared between the two active sites. *Nat. Struct. Biol.* **2001,** *8* (4), 312-316.

(32) Chevalier, B.; Sussman, D.; Otis, C.; Noël, A.-J.; Turmel, M.; Lemieux, C.; Stephens, K.; Monnat, R. J.; Stoddard, B. L., Metal-dependent DNA cleavage mechanism of the I-Cre I LAGLIDADG homing endonuclease. *Biochemistry* **2004,** *43* (44), 14015-14026.

(33) 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.

(34) Wang, Y.; Zhang, L.; Zhu, X.; Li, Y.; Shi, H.; Oger, P.; Yang, Z., Biochemical characterization of a thermostable endonuclease V from the hyperthermophilic euryarchaeon Thermococcus barophilus Ch5. *Int. J. Biol. Macromol.* **2018,** *117*, 17-24.

(35) Deibert, M.; Grazulis, S.; Sasnauskas, G.; Siksnys, V.; Huber, R., Structure of the tetrameric restriction endonuclease NgoMIV in complex with cleaved DNA. *Nat. Struct. Biol.* **2000,** *7* (9), 792-799.

(36) 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 APE1. *Proteins* **2007,** *68* (1), 313-323.

(37) Dupureur, C. M., Roles of metal ions in nucleases. *Curr. Opin. Chem. Biol.* **2008,** *12* (2), 250-255.

(38) Xie, F.; Qureshi, S. H.; Papadakos, G. A.; Dupureur, C. M., One- and two-metal ion catalysis: global single-turnover kinetic analysis of the PvuII endonuclease mechanism. *Biochemistry* **2008,** *47* (47), 12540-12550.

(39) 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.

(40) Yang, W., Nucleases: diversity of structure, function and mechanism. *Q. Rev. Biophys.* **2011,** *44* (1), 1-93.

(41) Yang, W.; Weng, P. J.; Gao, Y., A new paradigm of DNA synthesis: three-metal-ion catalysis. *Cell Biosci.* **2016,** *6*, 1-7.

(42) 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.

(43) Ivanov, I.; Tainer, J. A.; McCammon, J. A., Unraveling the three-metal-ion catalytic mechanism of the DNA repair enzyme endonuclease IV. *Proc. Natl. Acad. Sci. U.S.A.* **2007,** *104* (5), 1465.

Appendix A

Supplementary Information for Chapter 2: 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

Contains Tables A.1–A.12 and Figures A.1–A.17

Table A.1. Key M06-2X distances (in Å) in the RC corresponding to the phosphodiester bond cleavage reaction calculated using the smallest model and different basis sets.[a-c]

basis set	r1	r2	r3	r4	r5	r6	r7	r8
6-31G(d)	1.808	0.982	3.807	1.652	1.987	0.979	1.689	0.995
6-31G(d,p)	1.837	0.976	3.862	1.661	1.823	0.979	1.544	1.030
6-31+G(d,p)	1.838	0.978	3.959	1.651	2.181	0.971	1.786	0.986
6-311G(d,p)	1.817	0.975	3.882	1.652	1.989	0.973	1.735	0.985
6-311+G(d,p)	1.829	0.976	3.944	1.651	2.126	0.968	1.782	0.983
6-311+G(2d,2p)	1.846	0.974	3.957	1.649	1.795	0.976	1.636	0.996
6-311+G(2df,p)	1.846	0.975	3.963	1.644	1.804	0.977	1.621	1.000
6-311+G(2df,2p)	1.845	0.974	3.953	1.644	1.804	0.976	1.617	0.999
average[b]	1.833	0.976	3.916	1.651	1.939	0.975	1.676	0.997
deviation[c]	0.014	0.003	0.058	0.005	0.155	0.004	0.086	0.015

[a]Key distances are defined in Figure A.1. [b]Each distance was averaged across all basis sets. [c]Standard deviation for each distance was calculated across all basis sets.

Table A.2. Key M06-2X distances (in Å) in the TS corresponding to the phosphodiester bond cleavage reaction calculated using the smallest model and different basis sets.[a-c]

basis set	r1	r2	r3	r4	r5	r6	r7	r8
6-31G(d)	1.054	1.454	1.807	2.092	1.288	1.136	1.609	1.012
6-31G(d,p)	1.062	1.412	1.802	2.141	1.206	1.202	1.545	1.019
6-31+G(d,p)	1.079	1.377	1.830	2.022	1.215	1.194	1.499	1.032
6-311G(d,p)	1.077	1.368	1.827	2.056	1.099	1.321	1.549	1.014
6-311+G(d,p)	1.073	1.385	1.839	2.029	1.245	1.165	1.535	1.020
6-311+G(2d,2p)	1.076	1.381	1.832	2.017	1.289	1.130	1.505	1.029
6-311+G(2df,p)	1.078	1.379	1.824	2.012	1.277	1.141	1.502	1.031
6-311+G(2df,2p)	1.074	1.382	1.821	2.015	1.275	1.140	1.501	1.030
average[b]	1.072	1.392	1.823	2.048	1.237	1.178	1.531	1.023
deviation[c]	0.009	0.028	0.013	0.046	0.064	0.064	0.038	0.008

[a]Key distances are defined in Figure A.1. [b]Each distance was averaged across all basis sets. [c]Standard deviation for each distance was calculated across all basis sets.

Table A.3. Key M06-2X distances (in Å) in the PC corresponding to the phosphodiester bond cleavage reaction calculated using the smallest model and different basis sets.[a-c]

basis set	r1	r2	r3	r4	r5	r6	r7	r8
6-31G(d)	1.009	1.633	1.679	3.040	1.029	1.551	1.881	0.989
6-31G(d,p)	1.006	1.634	1.677	3.108	1.039	1.499	1.971	0.983
6-31+G(d,p)	1.008	1.604	1.673	3.135	1.032	1.530	1.856	0.987
6-311G(d,p)	1.005	1.604	1.676	3.025	1.021	1.564	1.898	0.982
6-311+G(d,p)	0.998	1.647	1.667	3.181	1.017	1.582	1.814	0.985
6-311+G(2d,2p)	1.000	1.639	1.657	3.200	1.017	1.578	1.881	0.982
6-311+G(2df,p)	0.997	1.666	1.647	3.391	1.017	1.586	1.831	0.985
6-311+G(2df,2p)	1.002	1.634	1.655	3.374	1.003	1.613	1.725	0.989
average[b]	1.003	1.633	1.666	3.182	1.022	1.563	1.857	0.985
deviation[c]	0.005	0.021	0.012	0.138	0.011	0.036	0.071	0.003

[a]Key distances are defined in Figure A.1. [b]Each distance was averaged across all basis sets. [c]Standard deviation for each distance was calculated across all basis sets.

Table A.4. Percentage change in the key M06-2X distances calculated with respect to the 6-311+G(2df,2p) distances in the RC corresponding to the phosphodiester bond cleavage reaction calculated using the smallest model and different basis sets.[a]

basis set	r1	r2	r3	r4	r5	r6	r7	r8
6-31G(d)	2.00	0.88	3.69	0.48	10.14	0.39	4.44	0.46
6-31G(d,p)	0.44	0.26	2.31	1.04	1.04	0.30	4.50	3.12
6-31+G(d,p)	0.38	0.47	0.15	0.42	20.89	0.53	10.42	1.36
6-311G(d,p)	1.55	0.16	1.80	0.44	10.25	0.28	7.28	1.39
6-311+G(d,p)	0.90	0.20	0.23	0.39	17.80	0.74	10.19	1.64
6-311+G(2d,2p)	0.05	0.01	0.09	0.28	0.54	0.06	1.14	0.27
6-311+G(2df,p)	0.04	0.13	0.24	0.01	0.04	0.15	0.22	0.07
6-311+G(2df,2p)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

[a]Key distances are defined in Figure A.1.

Table A.5. Percentage change in key M06-2X distances calculated with respect to the 6-311+G(2df,2p) distances in the TS corresponding to the phosphodiester bond cleavage reaction calculated using the smallest model and different basis sets.[a]

basis set	r1	r2	r3	r4	r5	r6	r7	r8
6-31G(d)	1.86	5.19	0.75	3.82	0.94	0.36	7.19	1.69
6-31G(d,p)	1.12	2.18	1.05	6.25	5.45	5.45	2.94	1.03
6-31+G(d,p)	0.39	0.35	0.52	0.32	4.77	4.78	0.07	0.24
6-311G(d,p)	0.26	1.03	0.33	2.05	13.87	15.89	3.24	1.50
6-311+G(d,p)	0.12	0.26	0.99	0.71	2.42	2.17	2.28	0.94
6-311+G(2d,2p)	0.12	0.07	0.61	0.10	1.08	0.90	0.29	0.10
6-311+G(2df,p)	0.35	0.24	0.18	0.15	0.12	0.09	0.11	0.10
6-311+G(2df,2p)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

[a]Key distances are defined in Figure A.1.

Table A.6. Percentage change in key M06-2X distances calculated with respect to the 6-311+G(2df,2p) distances in the PC corresponding to the phosphodiester bond cleavage reaction calculated using the smallest model and different basis sets.[a]

basis set	r1	r2	r3	r4	r5	r6	r7	r8
6-31G(d)	0.69	0.08	1.44	9.91	2.65	3.86	8.99	0.06
6-31G(d,p)	0.37	0.01	1.32	7.88	3.61	7.06	4.80	0.62
6-31+G(d,p)	0.60	1.83	1.10	7.08	2.92	5.16	7.58	0.18
6-311G(d,p)	0.25	1.84	1.28	10.33	1.82	3.03	9.99	0.67
6-311+G(d,p)	0.38	0.80	0.75	5.70	1.46	1.91	5.11	0.33
6-311+G(2d,2p)	0.26	0.26	0.13	5.16	1.45	2.15	9.01	0.63
6-311+G(2df,p)	0.53	1.93	0.45	0.52	1.43	1.69	6.13	0.34
6-311+G(2df,2p)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

[a]Key distances are defined in Figure A.1.

Table A.7. Key distances (in Å) in the RC corresponding to the phosphodiester bond cleavage reaction calculated using the smallest model and various DFT functionals.[a,b]

functional	r1	r2	r3	r4	r5	r6	r7	r8
M06-HF	1.845	0.977	3.689	1.640	1.884	0.975	1.441	1.045
PBEPBE	1.794	0.996	4.077	1.713	1.814	0.995	1.490	1.045
BP86	1.792	0.997	4.070	1.714	1.804	0.998	1.484	1.048
PBEPBE-D3	1.815	0.993	3.788	1.701	1.730	0.998	1.564	1.024
M06-2X	1.837	0.976	3.862	1.661	1.823	0.979	1.544	1.014
B3LYP-10%HF	1.775	0.998	3.854	1.690	1.768	0.992	1.556	1.022
M06-2X-D3	1.836	0.977	3.854	1.661	1.821	0.979	1.544	1.013
M06-L	1.869	0.977	3.887	1.681	1.802	0.980	1.638	0.999
MN15	1.842	0.981	3.886	1.656	1.781	0.984	1.591	1.010
M11	1.827	0.978	3.879	1.669	1.807	0.981	1.584	1.009
MPWB1K	1.863	0.968	3.890	1.647	1.878	0.969	1.592	0.994
B3LYP	1.801	0.990	3.909	1.682	1.799	0.984	1.575	1.012
M06	1.862	0.977	3.865	1.666	1.780	0.980	1.650	0.996
M11-L	1.916	0.965	3.893	1.634	1.806	0.967	1.703	0.977
MN12-SX	1.878	0.975	3.967	1.651	1.930	0.976	1.655	0.995
B3LYP-D3	1.807	0.983	3.804	1.680	1.735	0.986	1.586	1.008
MN12-L	1.882	0.973	3.994	1.649	1.898	0.973	1.687	0.995
M05	1.875	0.975	3.874	1.675	1.819	0.977	1.663	0.992
average[c]	1.840	0.981	3.891	1.671	1.816	0.982	1.586	1.011
deviation[d]	0.038	0.010	0.093	0.024	0.053	0.009	0.072	0.020

[a]Key distances are defined in Figure A.1. [b]Geometries were obtained using the 6-31G(d,p) basis set. [c]Each distance was averaged across all functionals. [d]Standard deviation for each distance was calculated across all functionals.

Table A.8. Key distances (in Å) in the TS corresponding to the phosphodiester bond cleavage reaction calculated using the smallest model and various DFT functionals.[a,b]

functional	r1	r2	r3	r4	r5	r6	r7	r8
M06-HF	1.113	1.315	1.893	1.831	1.458	1.043	1.383	1.070
PBEPBE	1.144	1.305	1.916	2.172	1.585	1.039	1.637	1.016
BP86	1.148	1.303	1.923	2.174	1.570	1.045	1.621	1.020
PBEPBE-D3	1.130	1.323	1.897	2.190	1.555	1.043	1.637	1.015
M06-2X	1.062	1.412	1.802	2.141	1.206	1.202	1.545	1.019
B3LYP-10%HF	1.138	1.307	1.915	2.135	1.578	1.035	1.611	1.015
M06-2X-D3	1.063	1.412	1.802	2.137	1.203	1.205	1.542	1.019
M06-L	1.113	1.341	1.893	2.112	1.623	1.013	1.706	0.996
MN15	1.058	1.442	1.801	2.132	1.488	1.042	1.637	1.007
M11	1.057	1.447	1.814	2.231	1.391	1.083	1.638	1.004
MPWB1K	1.037	1.457	1.785	2.160	1.478	1.031	1.635	0.991
B3LYP	1.083	1.389	1.868	2.135	1.485	1.052	1.538	1.023
M06	1.147	1.264	1.873	2.167	1.622	1.014	1.777	0.991
M11-L	1.044	1.424	1.798	1.996	1.616	0.995	1.642	0.988
MN12-SX	1.037	1.486	1.790	2.132	1.478	1.040	1.640	1.001
B3LYP-D3	1.104	1.344	1.887	2.118	1.543	1.030	1.620	1.008
MN12-L	1.112	1.340	1.893	2.111	1.611	1.003	1.757	0.995
M05	1.188	1.224	1.916	2.178	1.673	1.005	1.763	0.985
average[c]	1.099	1.363	1.859	2.125	1.509	1.051	1.629	1.009
deviation[d]	0.045	0.073	0.052	0.087	0.132	0.059	0.093	0.020

[a]Key distances are defined in Figure A.1. [b]Geometries were obtained using the 6-31G(d,p) basis set. [c]Each distance was averaged across all functionals. [d]Standard deviation for each distance was calculated across all functionals.

Table A.9. Key distances (in Å) in the PC corresponding to the phosphodiester bond cleavage reaction calculated using the smallest model and various DFT functionals.[a,b]

functional	r1	r2	r3	r4	r5	r6	r7	r8
M06-HF	1.025	1.513	1.669	3.120	1.049	1.446	1.575	1.012
PBEPBE	1.028	1.607	1.706	3.695	1.063	1.491	1.759	1.005
BP86	1.028	1.605	1.700	4.209	1.085	1.440	1.780	1.003
PBEPBE-D3	1.032	1.553	1.757	4.085	1.106	1.371	2.014	0.980
M06-2X	1.006	1.634	1.677	3.108	1.039	1.499	1.971	0.983
B3LYP-10%HF	1.016	1.632	1.685	4.191	1.058	1.492	1.774	0.997
M06-2X-D3	1.005	1.637	1.676	3.109	1.040	1.496	1.983	0.986
M06-L	1.004	1.673	1.688	3.299	1.015	1.622	1.877	0.984
MN15	1.004	1.694	1.666	3.191	1.025	1.575	1.999	0.987
M11	1.005	1.663	1.682	3.083	1.031	1.552	1.969	0.984
MPWB1K	0.995	1.620	1.662	3.069	1.004	1.607	1.877	0.978
B3LYP	1.003	1.659	1.712	4.077	1.023	1.585	1.771	0.990
M06	1.001	1.696	1.672	3.230	1.013	1.625	1.879	0.982
M11-L	0.983	1.948	1.629	3.023	0.996	1.384	1.829	0.973
MN12-SX	0.998	1.673	1.665	3.096	1.009	1.644	2.031	0.981
B3LYP-D3	1.003	1.637	1.676	4.307	1.035	1.536	1.762	0.993
MN12-L	0.988	1.777	1.655	3.106	1.004	1.665	2.190	0.976
M05	0.994	1.704	1.701	4.043	1.005	1.651	1.796	0.982
average[c]	1.007	1.663	1.682	3.502	1.033	1.538	1.880	0.988
deviation[d]	0.014	0.092	0.027	0.497	0.030	0.091	0.142	0.011

[a]Key distances are defined in Figure A.1. [b]Geometries were obtained using the 6-31G(d,p) basis set. [c]Each distance was averaged across all functionals. [d]Standard deviation for each distance was calculated across all functionals.

Table A.10. Key distances (in Å) for all stationary points corresponding to the phosphodiester bond cleavage reaction with the truncation point for select amino acid(s) in model 1 extended to the α carbon.[a,b]

residue expanded to C_α	D210	H309	D70	RC	TS1	IC	TS2	PC
r1				2.099	1.226	1.041	1.005	0.996
	X			1.874	1.241	1.044	1.006	1.054
		X		1.834	1.226	1.041	1.003	0.994
			X	2.040	1.239	1.044	0.997	0.986
	X	X		1.868	1.234	1.031	1.003	1.034
	X		X	1.919	1.230	1.043	1.006	0.972
		X	X	1.829	1.227	1.041	1.004	0.993
	X	X	X	1.679	1.226	1.031	1.003	0.972
average[c]				1.893	1.231	1.039	1.003	1.000
deviation[d]				0.130	0.006	0.006	0.003	0.029
r2				0.975	1.200	1.495	1.631	1.678
	X			0.986	1.185	1.477	1.631	2.738
		X		0.989	1.200	1.493	1.628	1.762
			X	0.979	1.186	1.475	1.688	1.839
	X	X		0.987	1.192	1.533	1.648	3.107
	X		X	0.984	1.195	1.482	1.626	3.311
		X	X	0.990	1.198	1.491	1.622	1.674
	X	X	X	1.004	1.199	1.525	1.648	3.256
average[c]				0.987	1.194	1.496	1.640	2.421
deviation[d]				0.009	0.006	0.021	0.022	0.750
r3				6.835	2.030	1.848	1.706	1.651
	X			4.987	1.986	1.839	1.708	1.613
		X		4.952	2.029	1.848	1.690	1.664
			X	4.762	1.986	1.838	1.716	1.662
	X	X		5.018	2.032	1.865	1.707	1.607
	X		X	5.079	2.011	1.843	1.704	1.643
		X	X	5.087	2.039	1.849	1.696	1.651
	X	X	X	5.154	2.018	1.841	1.707	1.637
average[c]				5.234	2.016	1.846	1.704	1.641
deviation[d]				0.657	0.021	0.009	0.008	0.021
r4				1.626	1.709	1.747	2.278	2.970
	X			1.634	1.713	1.745	2.263	3.414
		X		1.635	1.709	1.748	2.351	3.360
			X	1.635	1.712	1.745	2.253	3.752
	X	X		1.633	1.705	1.743	2.234	4.569
	X		X	1.637	1.709	1.742	2.255	4.804

residue expanded to C_α	D210	H309	D70	RC	TS1	IC	TS2	PC
		X	X	1.634	1.705	1.741	2.271	3.060
	X	X	X	1.635	1.709	1.749	2.237	4.766
average[c]				1.634	1.709	1.745	2.268	3.837
deviation[d]				0.003	0.003	0.003	0.037	0.766
r5				2.699	1.870	1.783	1.062	1.015
	X			2.644	1.865	1.806	1.066	1.033
		X		2.686	1.863	1.777	1.049	1.027
			X	2.613	1.875	1.810	1.063	1.019
	X	X		2.744	1.880	1.814	1.068	1.021
	X		X	2.698	1.857	1.814	1.060	0.987
		X	X	2.737	1.881	1.810	1.056	1.004
	X	X	X	2.657	1.856	1.790	1.067	0.991
average[c]				2.685	1.868	1.800	1.061	1.012
deviation[d]				0.045	0.010	0.015	0.006	0.017
r6				0.976	0.983	0.991	1.445	1.614
	X			0.992	0.984	0.989	1.435	1.487
		X		0.993	0.984	0.991	1.480	1.514
			X	0.991	0.983	0.989	1.447	1.531
	X	X		0.995	0.983	0.989	1.436	1.576
	X		X	0.991	0.984	0.989	1.448	1.823
		X	X	0.995	0.983	0.989	1.463	1.666
	X	X	X	0.993	0.984	0.991	1.439	1.765
average[c]				0.991	0.983	0.990	1.449	1.622
deviation[d]				0.006	0.001	0.001	0.015	0.121
r7				1.546	0.993	0.990	1.035	1.046
	X			1.468	0.997	0.991	1.035	1.478
		X		1.500	0.992	0.990	1.036	1.078
			X	1.499	0.997	0.991	1.037	1.521
	X	X		1.474	0.990	0.987	1.032	1.544
	X		X	1.553	0.991	0.987	1.034	1.692
		X	X	1.508	0.989	0.986	1.037	1.029
	X	X	X	1.073	1.789	1.785	1.569	1.614
average[c]				1.509	0.992	0.990	1.035	1.375
deviation[d]				0.031	0.003	0.003	0.002	0.276
r8				1.095	1.750	1.780	1.567	1.535
	X			1.083	1.789	1.788	1.558	1.087
		X		1.085	1.750	1.780	1.557	1.474
			X	1.090	1.833	1.826	1.573	1.077
	X	X		1.065	1.813	1.820	1.569	1.074

residue expanded to C_α	D210	H309	D70	RC	TS1	IC	TS2	PC
	X		X	1.081	1.840	1.840	1.554	1.050
		X	X	1.076	1.822	1.748	1.570	1.610
	X	X	X	1.073	1.789	1.785	1.569	1.062
average[c]				1.081	1.798	1.796	1.565	1.246
deviation[d]				0.010	0.035	0.030	0.007	0.246

[a]Key distances are defined in Figure A.1. [b]Geometries were obtained using B3LYP-D3/6-31G(d,p). [c]Each distance averaged across all truncation points for select amino acid(s) extended to the α carbon. [d]Standard deviation for each distance was calculated across all truncation points for select amino acid(s) extended to the α carbon.

Table A.11. Key distances (in Å) in all stationary points corresponding to the phosphodiester bond cleavage reaction calculated using models of varying size.[a-c]

	Models	RC	TS1	IC	TS2	PC
r1	model 1	1.679	1.226	1.031	1.003	0.972
	model 2	1.901	1.116	1.029	1.007	1.055
	model 3	1.651	1.201	1.036	0.992	0.989
	model 4	1.809	1.175	1.082	1.131	1.064
	model 5	1.487	1.098	1.064	1.025	1.025
	model 6	1.710	1.379	1.033	1.013	1.004
	model 7	1.935	1.164	1.033	0.994	1.705
	model 8	1.746	1.265	1.008	1.000	0.989
	model 9	1.938	1.153	1.041	1.011	0.992
	model 10	2.797	1.032	1.009	1.000	1.046
r2	model 1	1.004	1.199	1.525	1.648	3.256
	model 2	0.985	1.342	1.579	1.647	3.019
	model 3	1.003	1.228	1.529	1.773	1.750
	model 4	0.983	1.252	1.399	2.828	1.664
	model 5	1.044	1.378	1.412	1.531	1.531
	model 6	0.989	1.094	2.398	2.558	1.629
	model 7	0.982	1.263	1.530	1.767	0.996
	model 8	0.997	1.167	1.671	1.696	1.736
	model 9	0.981	1.281	1.512	1.611	1.739
	model 10	0.977	1.541	1.670	1.688	1.454
r3	model 1	5.154	2.018	1.841	1.707	1.637
	model 2	5.025	2.089	1.898	1.730	1.610
	model 3	3.694	2.052	1.855	1.722	1.687
	model 4	3.909	1.838	1.790	1.642	1.648
	model 5	3.373	1.975	1.744	1.661	1.661
	model 6	5.314	1.983	1.736	1.674	1.663
	model 7	4.372	2.034	1.856	1.701	1.596
	model 8	4.500	2.105	1.799	1.745	1.664
	model 9	4.021	2.000	1.860	1.755	1.692
	model 10	3.265	1.970	1.748	1.647	1.570
r4	model 1	1.635	1.709	1.749	2.237	4.766
	model 2	1.617	1.672	1.697	2.225	3.503
	model 3	1.660	1.720	1.765	2.067	3.327
	model 4	1.614	1.802	1.830	2.259	3.532
	model 5	1.645	1.708	1.845	2.492	2.492
	model 6	1.650	1.709	1.758	2.288	6.413
	model 7	1.622	1.690	1.716	2.247	5.090
	model 8	1.651	1.720	1.789	2.251	5.241
	model 9	1.632	1.714	1.739	2.361	3.569
r4	model 10	1.665	1.705	1.674	2.323	4.988
	model 1	2.849	1.856	1.790	1.067	0.991

	Models	RC	TS1	IC	TS2	PC
r5	model 2	2.775	2.293	2.265	1.523	1.026
	model 3	1.745	1.723	1.685	1.138	1.018
	model 4	3.698	1.730	1.687	1.426	0.974
	model 5	2.111	1.792	1.598	1.374	1.005
	model 6	2.560	2.395	2.423	1.248	0.987
	model 7	3.205	2.213	2.154	1.085	1.051
	model 8	1.743	1.768	1.690	1.266	0.979
	model 9	2.023	1.902	1.896	1.209	1.029
	model 10	1.877	1.747	1.707	1.487	0.979
r6	model 1	0.974	0.984	0.991	1.439	1.765
	model 2	0.977	0.974	0.978	1.033	1.517
	model 3	0.984	0.989	0.996	1.286	1.533
	model 4	1.003	0.997	1.003	1.063	2.030
	model 5	0.970	0.976	1.006	1.086	1.453
	model 6	0.993	0.982	0.986	1.161	1.809
	model 7	0.981	0.978	0.980	1.360	1.453
	model 8	0.978	0.982	0.992	1.149	3.649
	model 9	0.968	0.975	0.977	1.208	1.534
	model 10	0.970	0.979	0.993	1.034	1.929
r7	model 1	1.522	0.990	0.997	1.033	1.062
	model 2	1.480	0.993	0.987	1.028	1.061
	model 3	1.465	1.008	0.999	1.000	1.559
	model 4	1.007	0.980	0.985	1.014	1.009
	model 5	5.963	0.982	0.990	1.012	1.642
	model 6	3.879	0.993	0.991	1.042	1.743
	model 7	1.477	1.004	0.996	0.991	1.046
	model 8	1.386	1.041	1.020	1.411	1.560
	model 9	1.372	1.027	1.017	1.470	1.570
	model 10	1.731	1.210	1.005	1.051	1.668
r8	model 1	1.076	1.822	1.748	1.570	3.059
	model 2	1.083	1.823	1.848	1.605	1.581
	model 3	1.096	1.678	1.721	1.790	1.067
	model 4	3.749	4.081	1.884	1.620	1.704
	model 5	1.078	4.164	1.771	1.642	1.012
	model 6	1.039	1.931	1.875	1.571	1.039
	model 7	1.086	1.671	1.714	1.768	1.528
	model 8	1.127	1.548	1.607	1.118	1.065
	model 9	1.129	1.565	1.595	1.091	1.058
	model 10	1.034	1.292	1.705	1.535	1.050
	model 1	1.522	0.990	0.997	1.033	1.062
	model 2	1.480	0.993	0.987	1.028	1.061
	model 3	1.465	1.008	0.999	1.000	1.559
	model 4	1.007	0.980	0.985	1.014	1.009
	model 5	5.963	0.982	0.990	1.012	1.642

	Models	RC	TS1	IC	TS2	PC
H_{309}NH…OP1	model 6	3.879	0.993	0.991	1.042	1.743
	model 7	1.477	1.004	0.996	0.991	1.046
	model 8	1.386	1.041	1.020	1.411	1.560
	model 9	1.372	1.027	1.017	1.470	1.570
	model 10	1.731	1.210	1.005	1.051	1.668
H_{309}N…H	model 1	1.076	1.822	1.748	1.570	3.059
	model 2	1.083	1.823	1.848	1.605	1.581
	model 3	1.096	1.678	1.721	1.790	1.067
	model 4	3.749	4.081	1.884	1.620	1.704
	model 5	1.078	4.164	1.771	1.642	1.012
	model 6	1.039	1.931	1.875	1.571	1.039
	model 7	1.086	1.671	1.714	1.768	1.528
	model 8	1.127	1.548	1.607	1.118	1.065
	model 9	1.129	1.565	1.595	1.091	1.058
	model 10	1.034	1.292	1.705	1.535	1.050

[a]Key distances are defined in Figure A.1. [b]Geometries were obtained using B3LYP-D3/6-31G(d,p). [c]Refer to Figure 2.5 for schematic of the models.

Table A.12: Relative solvent-phase Gibbs energies (ΔG, kJ/mol) for the phosphodiester bond cleavage reaction calculated using models of various sizes.[a,b]

	Y171	N212	N68	N174	D308	D283	RC	TS1	IC	TS2	PC
model 1							0.0	57.7	48.1	88.7	–10.3
model 2	X						0.0	74.2	75.7	97.9	–64.6
model 3		X					0.0	54.5	50.5	43.6	–48.1
model 4			X				0.0	126.7	127.6	121.0	4.9
model 5					X		0.0	36.6	56.9	66.5	72.5
model 6						X	0.0	188.6	160.7	155.0	31.0
model 7	X	X					0.0	61.7	62.2	114.8	–40.7
Model 8	X	X		X			0.0	72.1	67.1	85.0	–21.0
model 9	X	X	X	X			0.0	93.3	96.6	133.5	102.5
model 10	X	X	X	X	X	X	0.0	87.6	79.1	94.2	–6.9

[a]Relative energies were obtained from IEF-PCM-M06-2X/6-311+G(2df,2p) (ε=4) single-point calculations on B3LYP-D3/6-31G(d,p) gas-phase geometries. [b]All models contain D210 and H309. Refer to Figure 2.5 for schematic of the models. All amino acids were truncated at the α carbon, with the exception of E96 (β carbon truncation).

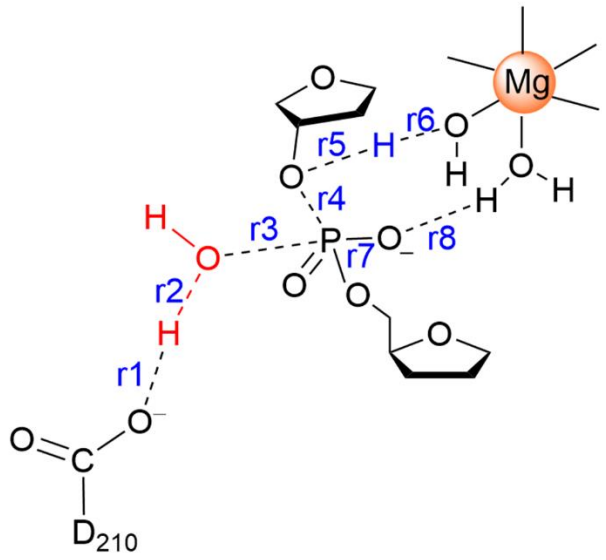

Figure A.1. Nomenclature used for key distances during the phosphodiester bond cleavage reaction in the present work depicted for the smallest model.

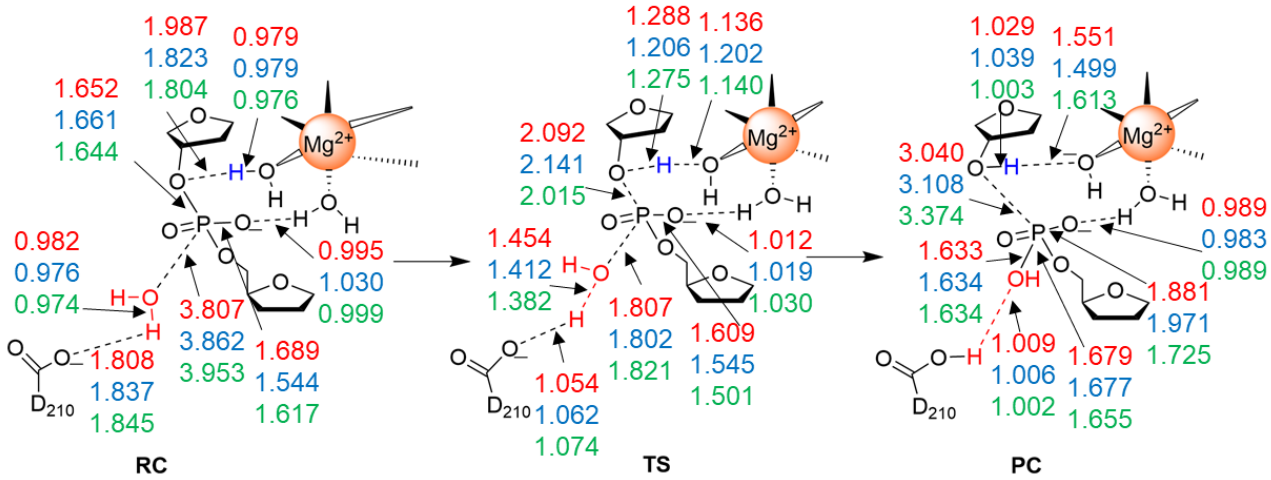

Figure A.2. Important distances (in Å) along the concerted phosphodiester bond cleavage pathway predicted using the smallest model (see Figure 2.3a) optimized with M06-2X and 6-31G(d) (red), 6-31G(d,p) (blue), or 6-311+G(2df,2p) (green).

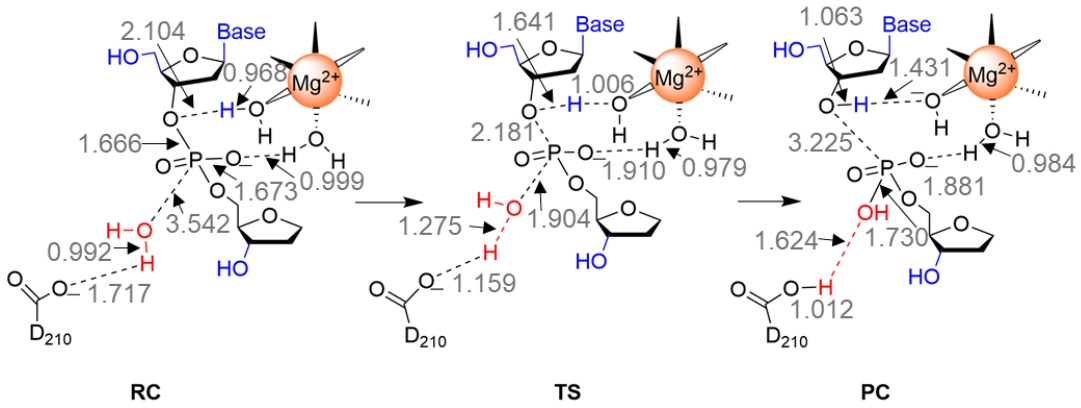

Figure A.3. Important distances (in Å) along the concerted phosphodiester bond cleavage pathway predicted using the expanded substrate model (amino acids truncated at the β carbon).

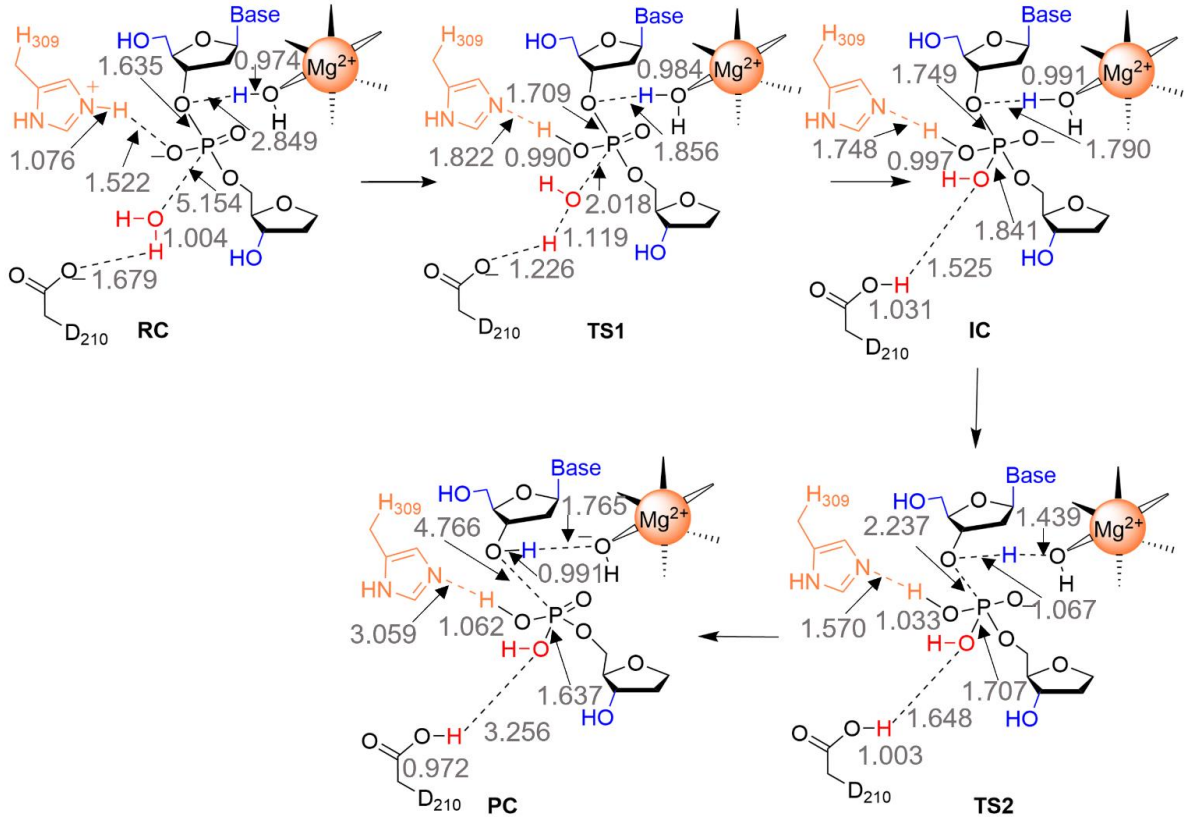

Figure A.4. Important distances (in Å) along the stepwise phosphodiester bond cleavage pathway predicted using model 1.

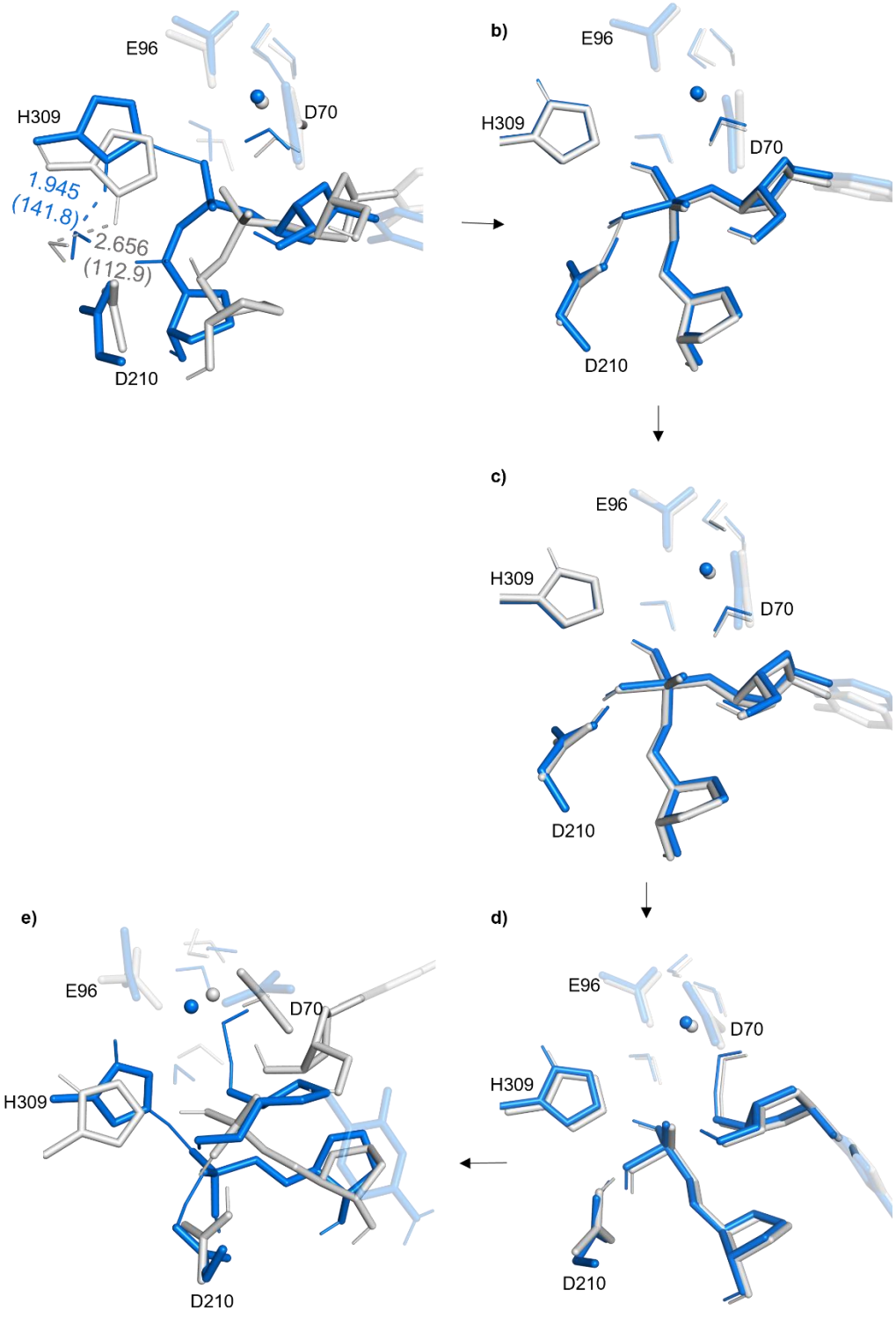

Figure A.5. Overlays of all stationary points along the phosphodiester bond cleavage pathway predicted using model 1 with truncation of D210 at C_β (grey) and truncation at C_α (blue): a) RC (RMSD=1.130 Å), b) TS1 (RMSD=1.021 Å), c) IC (RMSD=1.015 Å), d) TS2 (RMSD=1.024 Å), and e) PC (RMSD=1.645 Å).

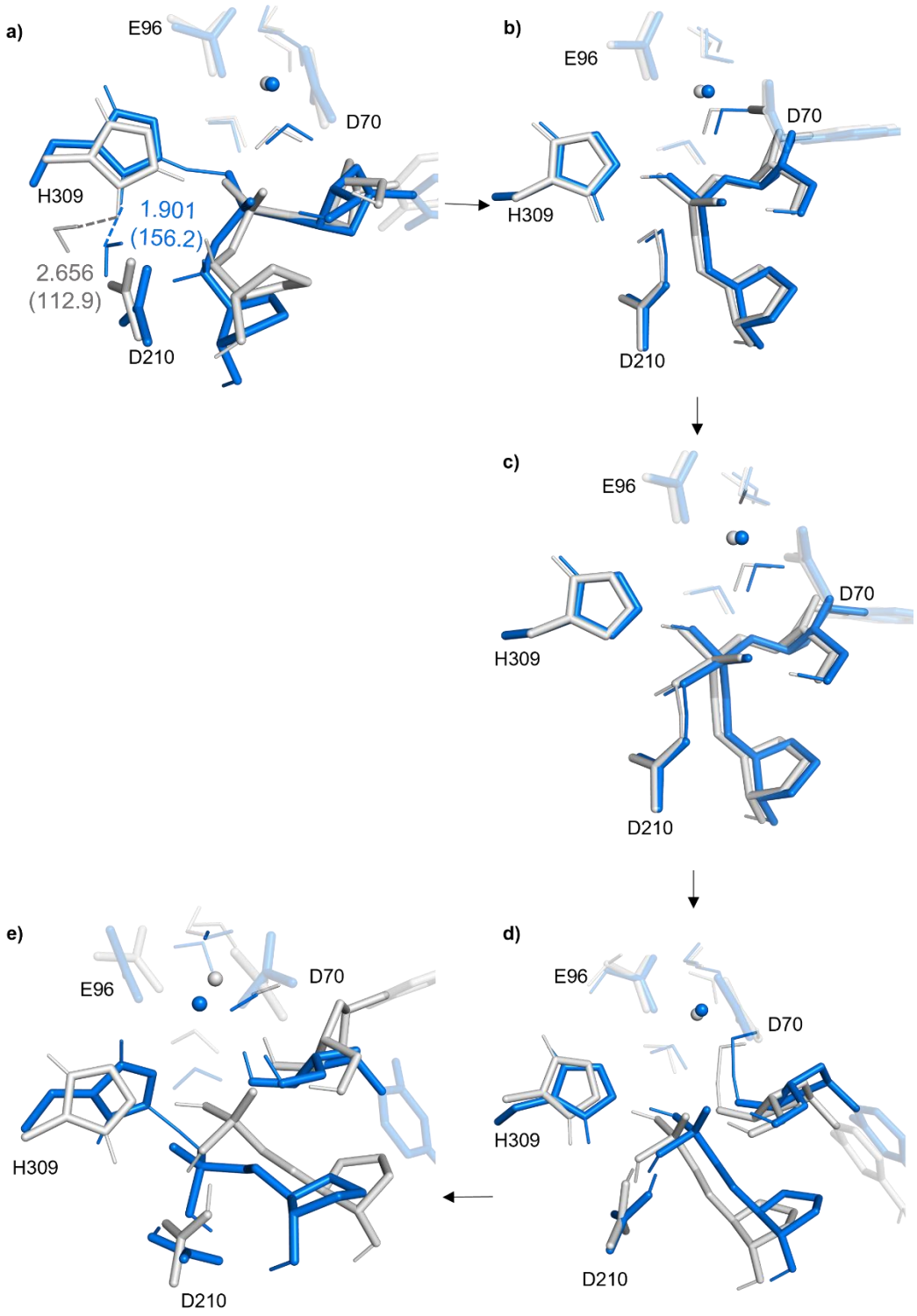

Figure A.6. Overlays of all stationary points along the phosphodiester bond cleavage pathway predicted using model 1 with truncation of H309 at C_β (grey) and truncation at C_α (blue): a) RC (RMSD=1.345 Å), b) TS1 (RMSD=1.149 Å), c) IC (RMSD=1.150 Å), d) TS2 (RMSD=1.185 Å), and e) PC (RMSD=1.537 Å).

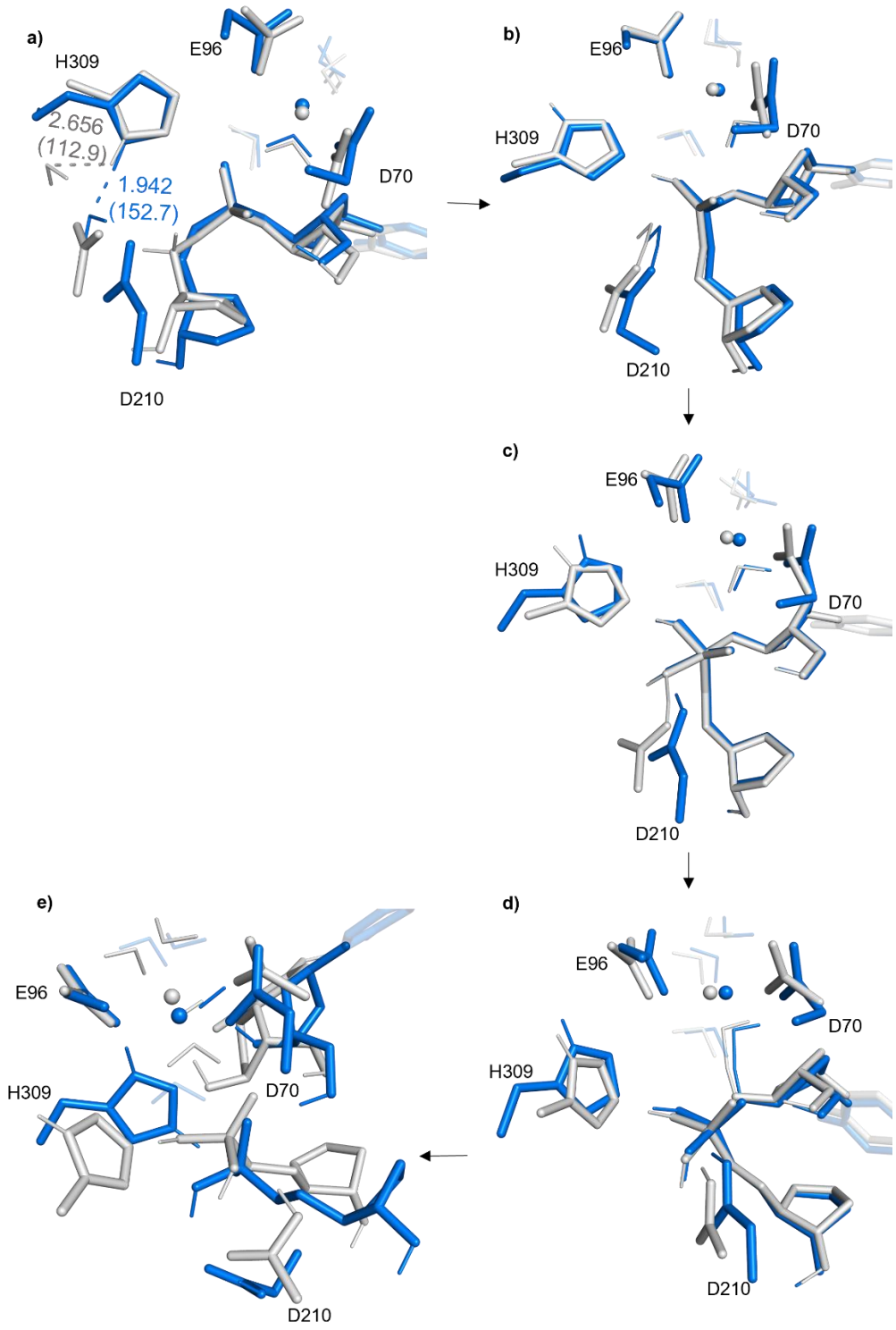

Figure A.7. Overlays of all stationary points along the phosphodiester bond cleavage pathway predicted using model 1 with truncation of all residues at C_β (grey) and truncation at C_α (blue): a) RC (RMSD=1.797 Å), b) TS1 (RMSD=1.453 Å), c) IC (RMSD=1.425 Å), d) TS2 (RMSD=1.462 Å), and e) PC (RMSD=2.015 Å).

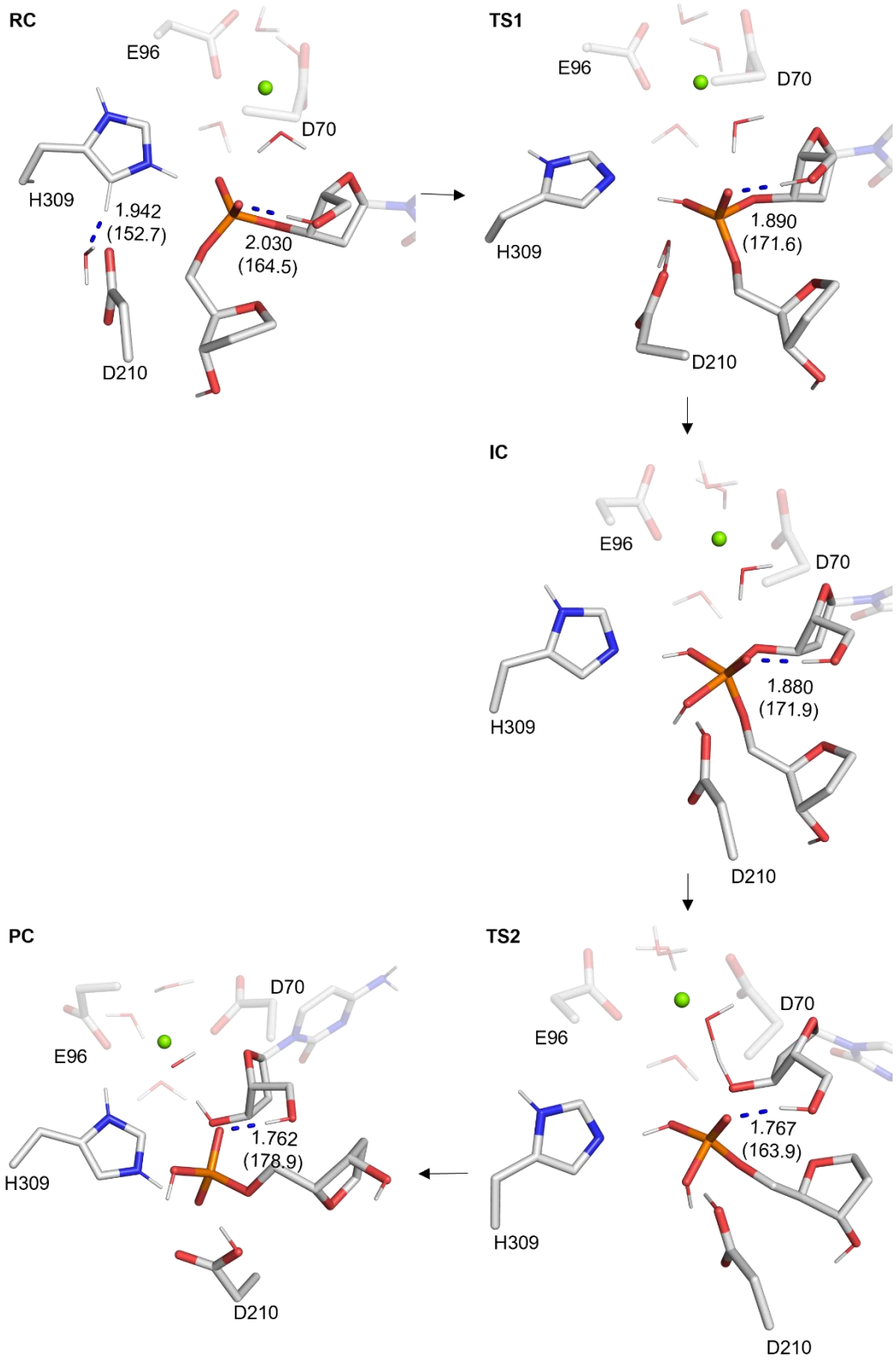

Figure A.8. Key distances (in Å) and angles (in °) along the phosphodiester bond cleavage pathway characterized using model 1 (110 atoms).

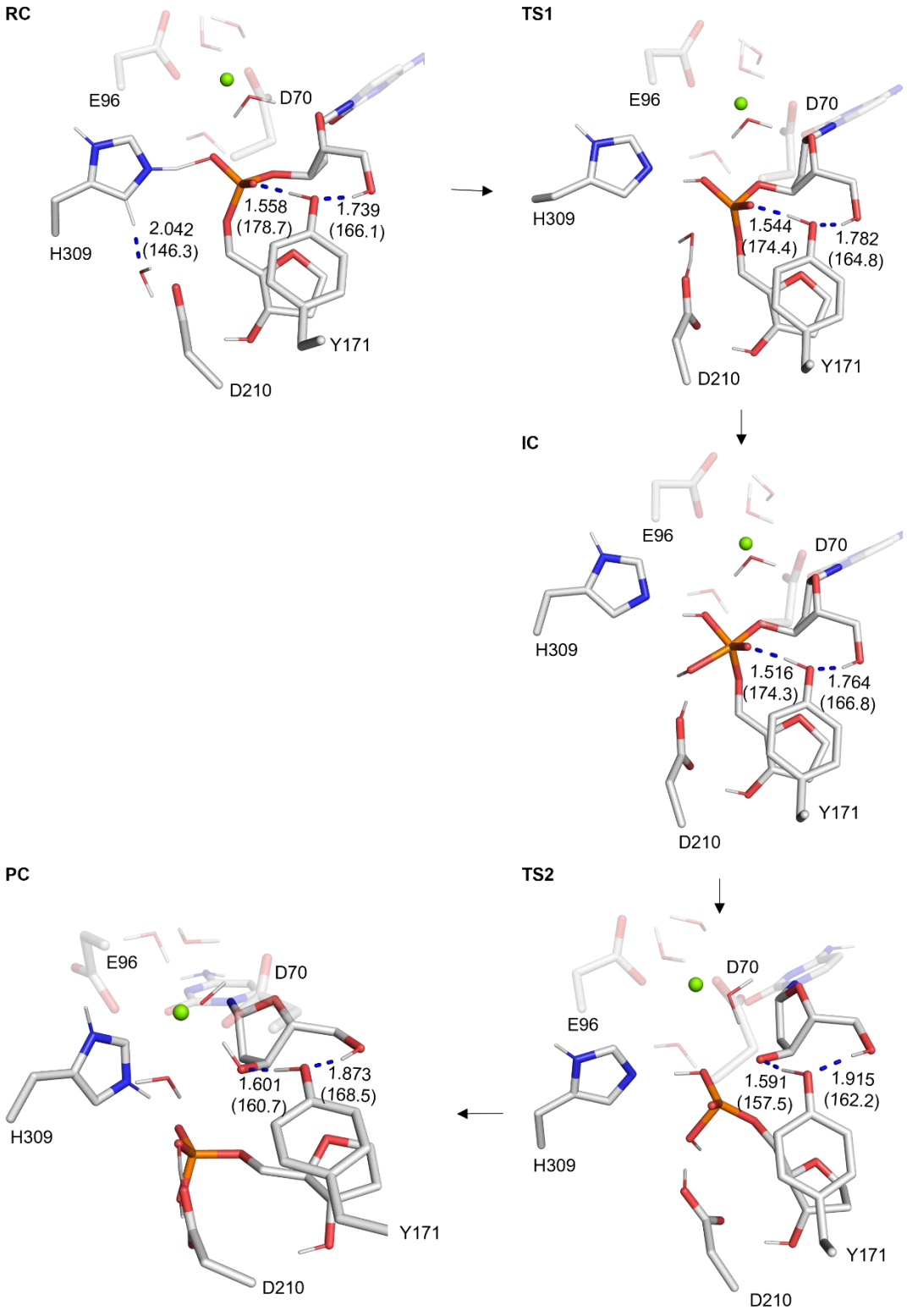

Figure A.9. Key distances (in Å) and angles (in °) along the phosphodiester bond cleavage pathway characterized using model 2 (129 atoms).

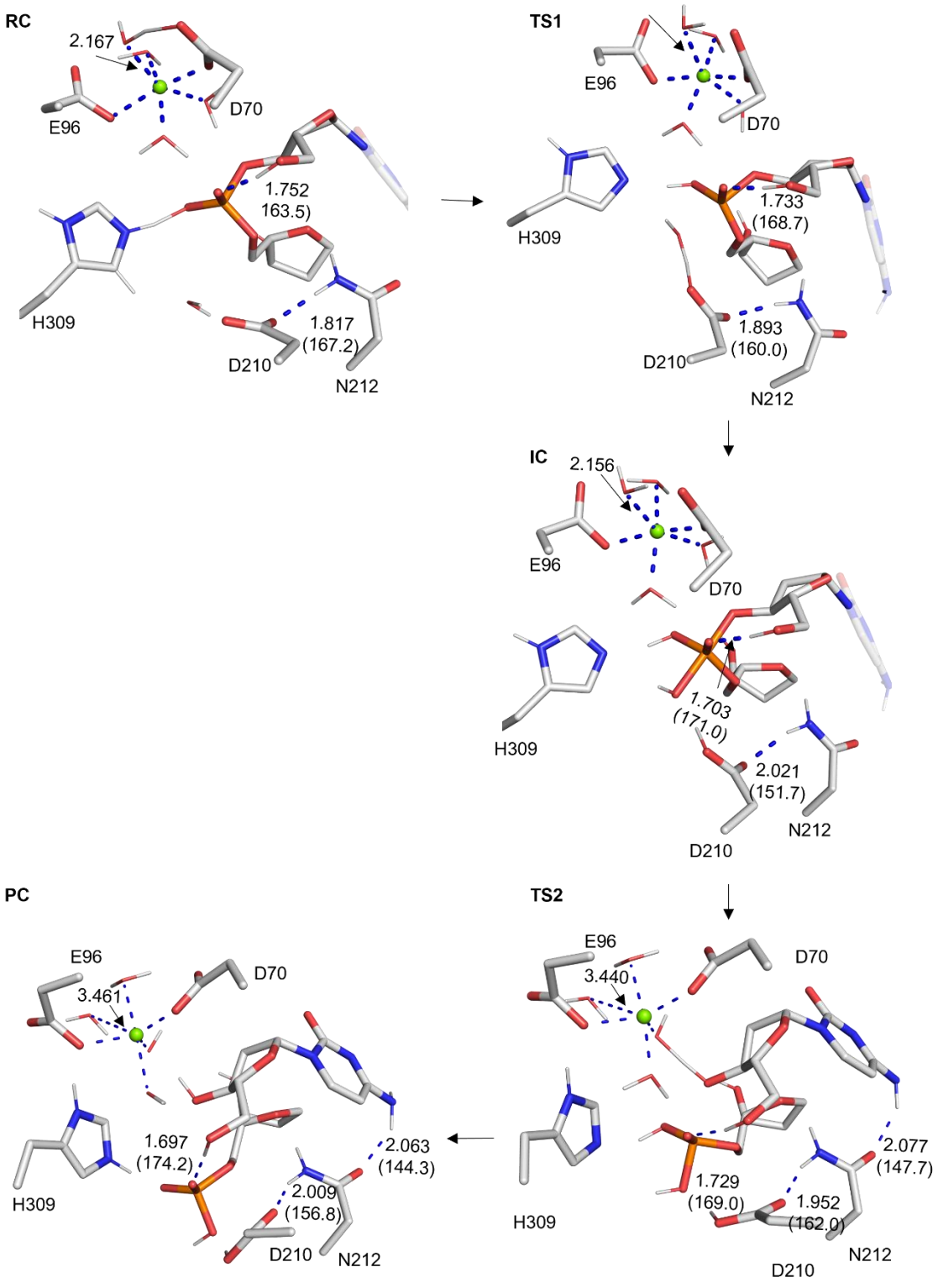

Figure A.10. Key distances (in Å) and angles (in °) along the phosphodiester bond cleavage pathway characterized using model 3 (122 atoms).

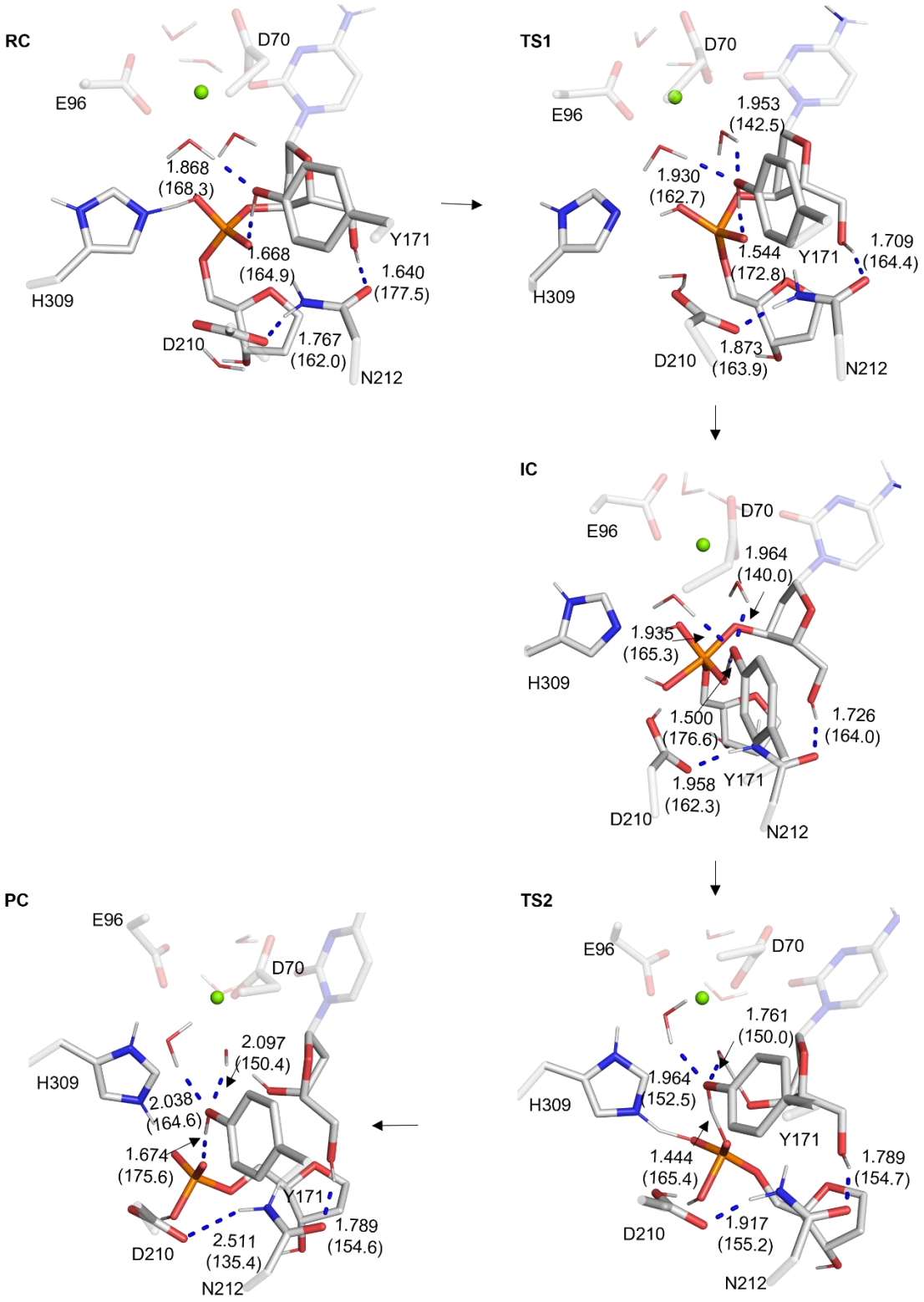

Figure A.11. Key distances (in Å) and angles (in °) along the phosphodiester bond cleavage pathway characterized using model 7 (141 atoms).

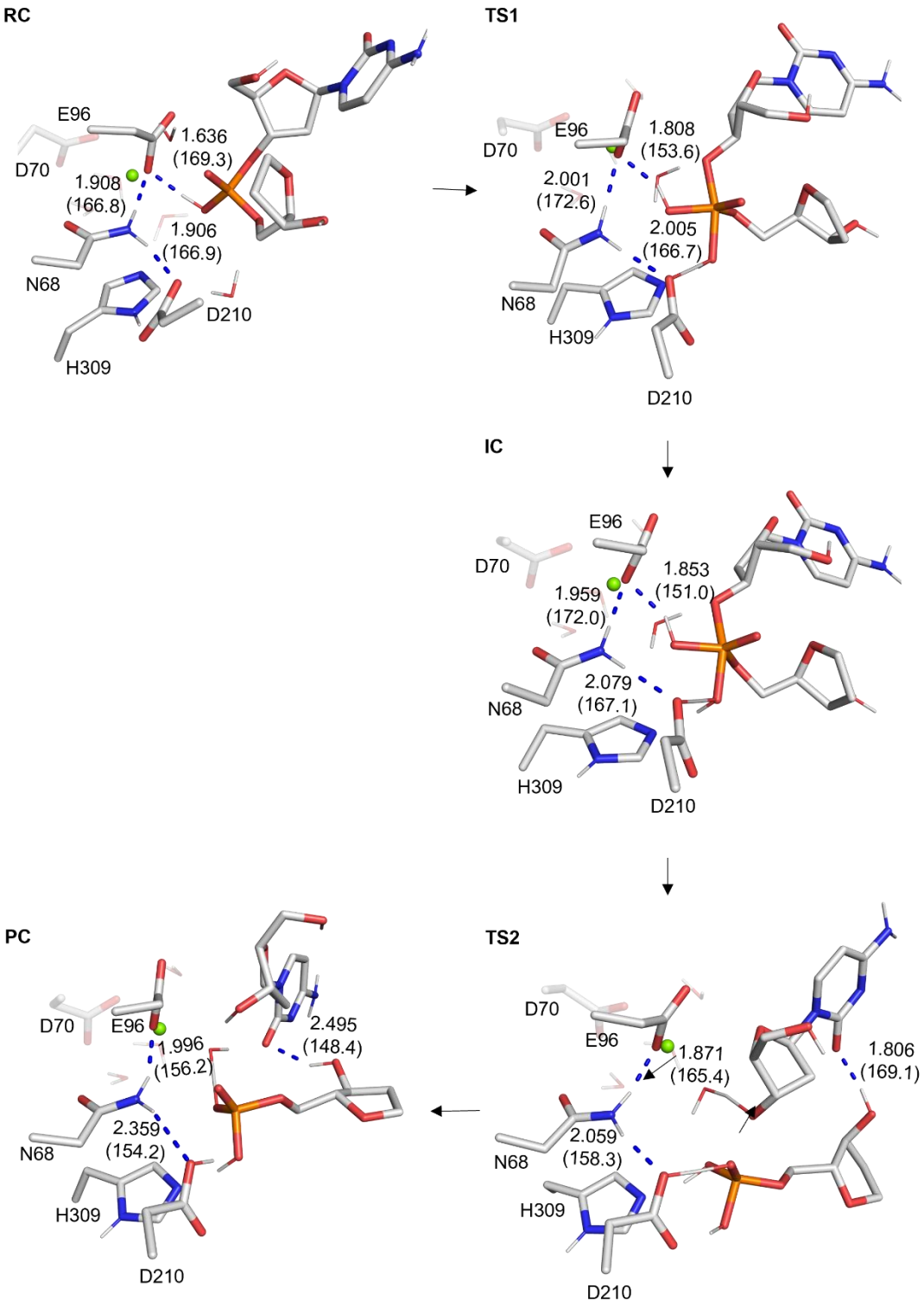

Figure A.12. Key distances (in Å) and angles (in °) along the phosphodiester bond cleavage pathway characterized using model 4 (122 atoms).

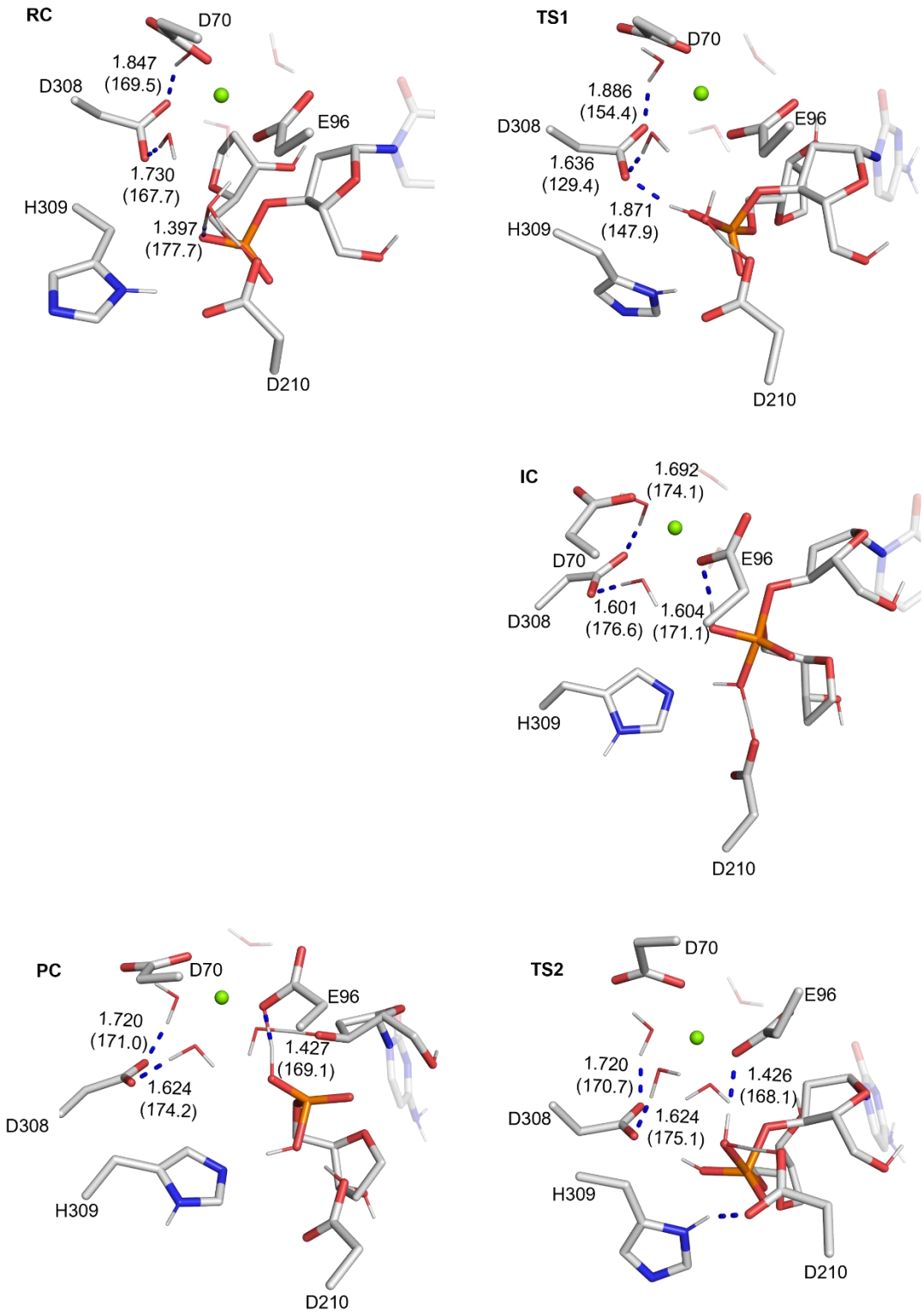

Figure A.13. Key distances (in Å) and angles (in °) along the phosphodiester bond cleavage pathway characterized using model 5 (120 atoms).

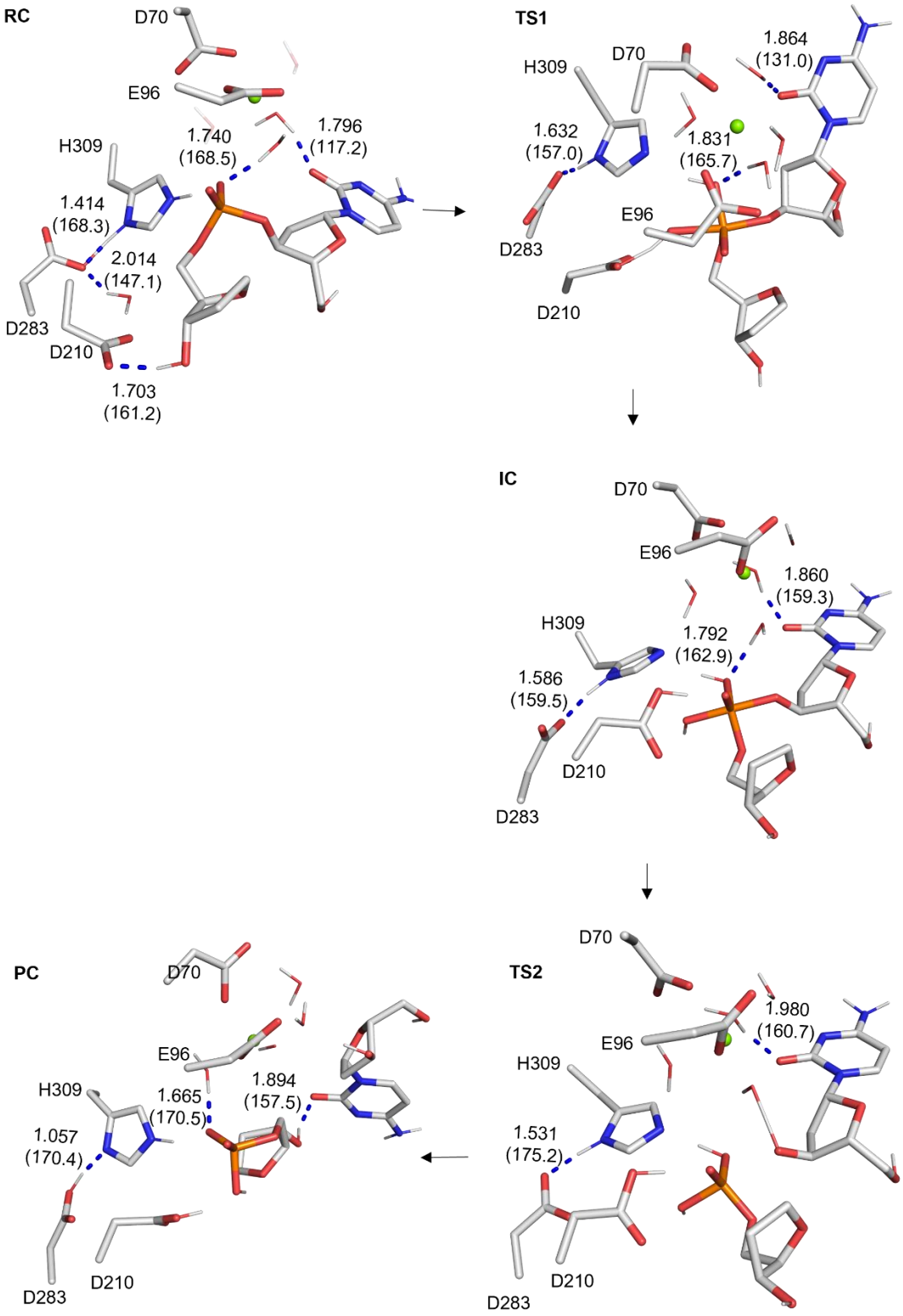

Figure A.14. Key distances (in Å) and angles (in °) along the phosphodiester bond cleavage pathway characterized using model 6 (120 atoms).

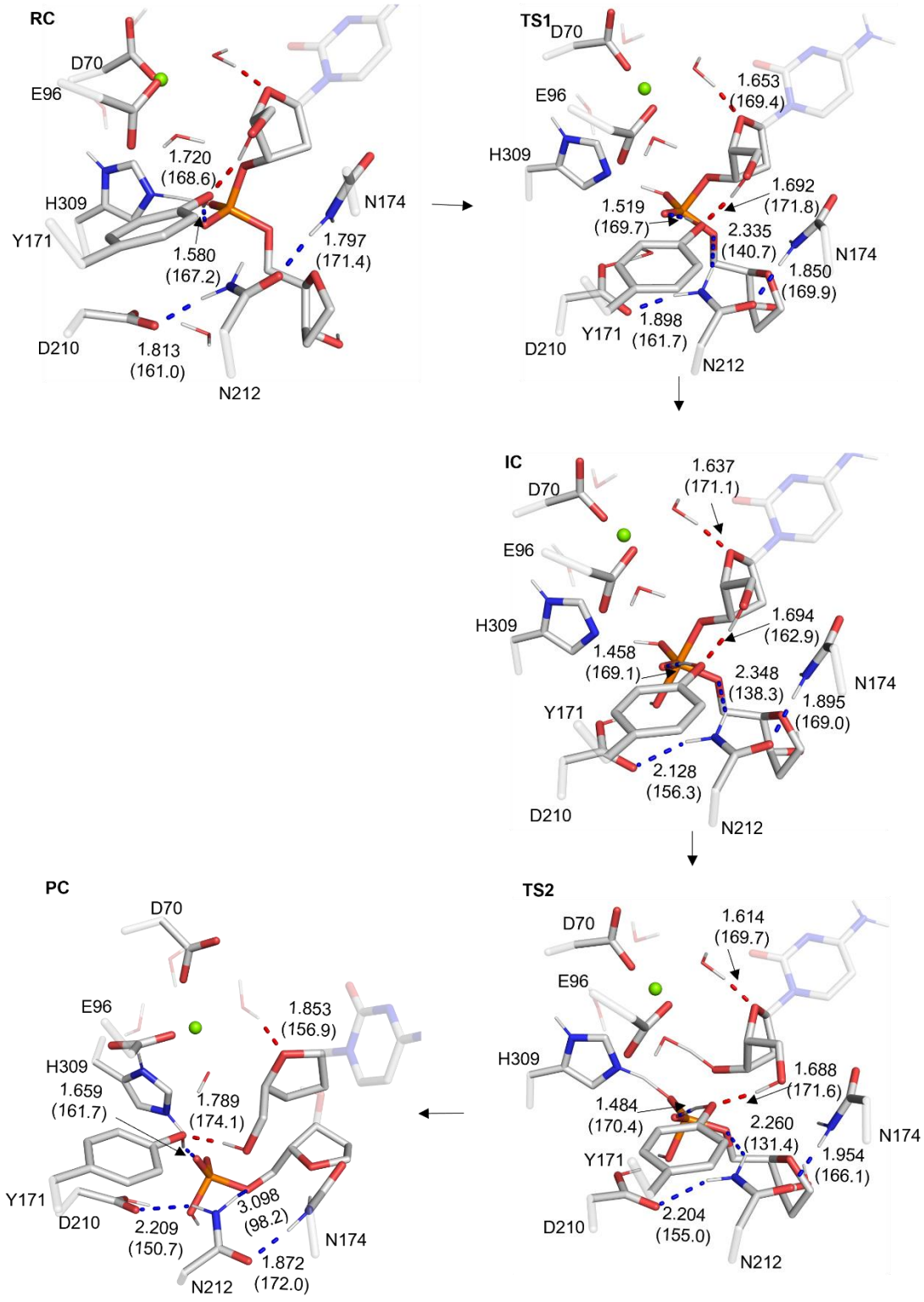

Figure A.15. Key distances (in Å) and angles (in °) along the phosphodiester bond cleavage pathway characterized using model 8 (153 atoms).

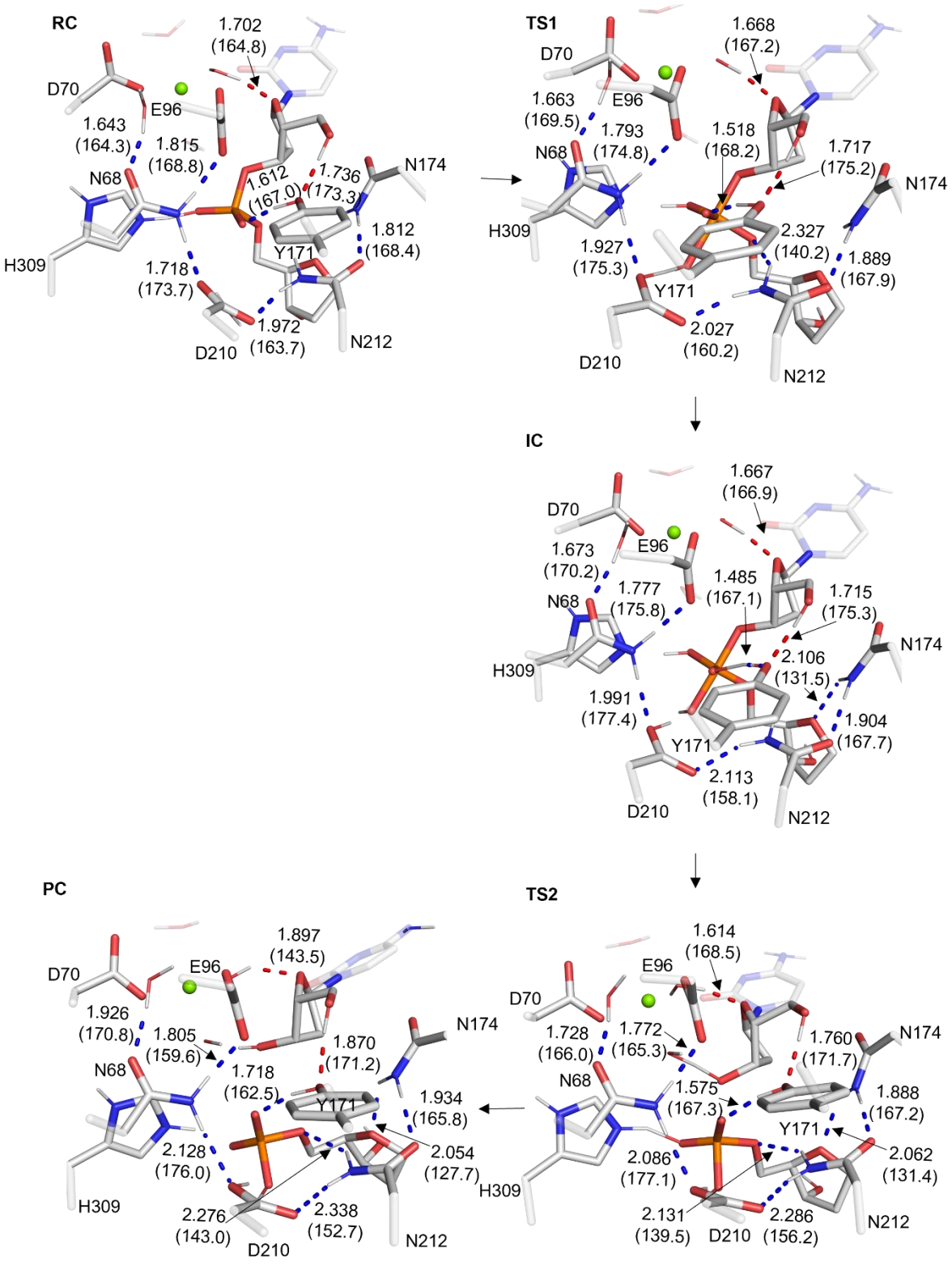

Figure A.16. Key distances (in Å) and angles (in °) along the phosphodiester bond cleavage pathway characterized using model 9 (165 atoms).

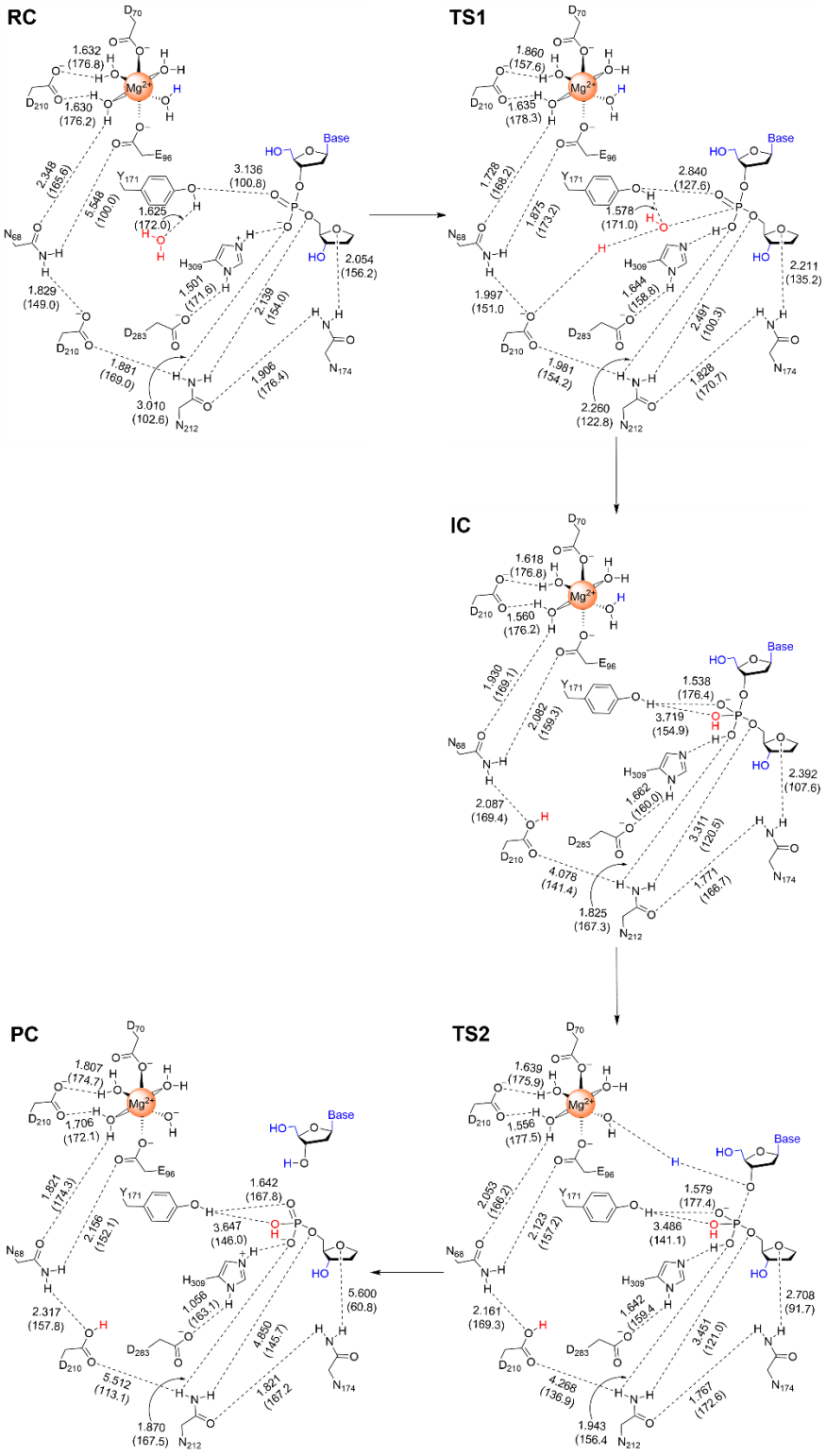

Figure A.17. Key distances (in Å) and angles (in °) along the phosphodiester bond cleavage pathway characterized using model 10 (185 atoms).

Appendix B

Supplementary Information for Chapter 3: The Metal Dependence of Single-Metal Mediated Phosphodiester Bond Cleavage: A QM/MM Study of a Multifaceted Human Enzyme

Contains Tables B.1–B.4 and Figures B.1–B.17

Table B.1. Protonation states of titratable amino acid residues. [a]

Amino Acid	pK_a	Amino Acid	pK_a	Amino Acid	pK_a
D47	5.23	E217	4.28	K77	10.54
D50	1.57	E236	4.90	K78	11.01
D70	7.18	E242	4.58	K79	10.61
D82	3.70	H116	6.55	K85	10.67
D90	3.99	H151	5.97	K98	12.62
D124	3.75	H215	5.55	K103	10.44
D148	3.74	H255	5.79	K125	10.66
D152	4.16	H289	6.02	K141	10.55
D163	3.69	H309	10.35	K194	10.80
D189	5.57	C65	12.59	K197	10.22
D210	7.54	C93	17.79	K203	10.90
D219	2.28	C99	10.49	K224	10.45
D251	4.39	C138	8.98	K227	10.05
D283	4.45	C208	8.17	K228	11.28
D297	3.63	C296	10.66	K276	11.36
D308	3.85	C310	12.74	K299	10.71
E46	4.43	Y45	14.32	K303	10.56
E86	4.66	Y118	12.03	R73	12.59
E87	4.76	Y128	13.35	R136	12.81
E96	4.77	Y144	11.27	R156	16.31
E101	5.15	Y171	17.30	R177	9.65
E107	4.39	Y184	12.19	R181	13.55
E110	4.71	Y257	12.81	R185	13.41
E126	4.53	Y262	9.98	R187	13.12
E149	4.67	Y264	12.86	R193	12.42
E150	4.00	Y269	15.89	R202	12.44
E154	4.14	Y284	14.57	R221	12.56
E161	3.91	Y315	12.86	R237	12.33
E183	4.41	K52	10.49	R254	12.14
E190	4.07	K58	10.72	R274	11.63
E216	4.51	K63	9.82		

[a]The physiological protonation states of titratable amino acid residues were verified using the propKa server.

Table B.2. Relative energy (kJ/mol) for the APE1-catalyzed phosphodiester cleavage facilitated by different metals from the potential energy surface.[a]

Stationary point	Mg^{2+}	Mn^{2+}	Ni^{2+} (s.p.)[b]	Ni^{2+} (oct.)[c]	Zn^{2+}	Ca^{2+}
RC	0.0	0.0	0.0	0.0	0.0	0.0
TS1	37.3	44.0	47.0	33.2	64.1	115.2
IC	28.3	35.9	32.4	28.8	38.6	75.5
TS2	42.7	43.1	61.9	102.8	85.5	89.1
PC	–84.9	–18.8	-37.5	8.2	–41.0	4.4

[a]Relative energies evaluated with ONIOM(B3LYP-D3(BJ)/6-31G(d,p):AMBER). [b]Square planar coordination. [c]Octahedral coordination.

Table B.3. Relative Gibbs energy (kJ/mol) for the APE1-catalyzed phosphodiester cleavage facilitated by different metals.[a]

Stationary point	Mg^{2+}	Mn^{2+}	Ni^{2+} (s.p.)[b]	Ni^{2+} (oct.)[c]	Zn^{2+}	Ca^{2+}
RC	0.0	0.0	0.0	0.0	0.0	0.0
TS1	64.8	69.8	63.1	74.3	91.1	137.7
IC	60.9	68.2	61.9	70.3	71.8	81.8
TS2	65.1	71.5	74.7	88.8	109.2	100.1
PC	–57.4	–30.3	–35.0	–41.5	–20.9	–11.9

[a]Relative energies evaluated with ONIOM(M06-2X/6-311+G(2df,p):AMBER)+Δ_{Gibbs}//ONIOM(B3LYP-D3(BJ)/6-31G(d,p):AMBER). [b]Square planar coordination. [c]Octahedral coordination.

Table B.4. Calculated charge-to-size ratios of different metals in the APE1 reactant complex.

	Mg^{2+}	Mn^{2+}	Ni^{2+} (s.p.)[b]	Ni^{2+} (oct.)[c]	Zn^{2+}	Ca^{2+}
Ionic radius of the metal	0.65	0.75	0.72	0.72	0.74	0.99
Charge on the metal[a]	1.697	1.715	0.980	1.102	1.274	1.763
Charge-to-size ratio	2.611	2.287	1.531	1.361	1.722	1.781

[a]NBO charges evaluated using B3LYP-D3(BJ)/6-31G(d,p) calculations on a model of the RC that includes only the QM region. Hydrogen atom positions at the truncation points were optimized at the same level of theory while constraining the remainder of the model. [b]Square planar coordination. [c]Octahedral coordination.

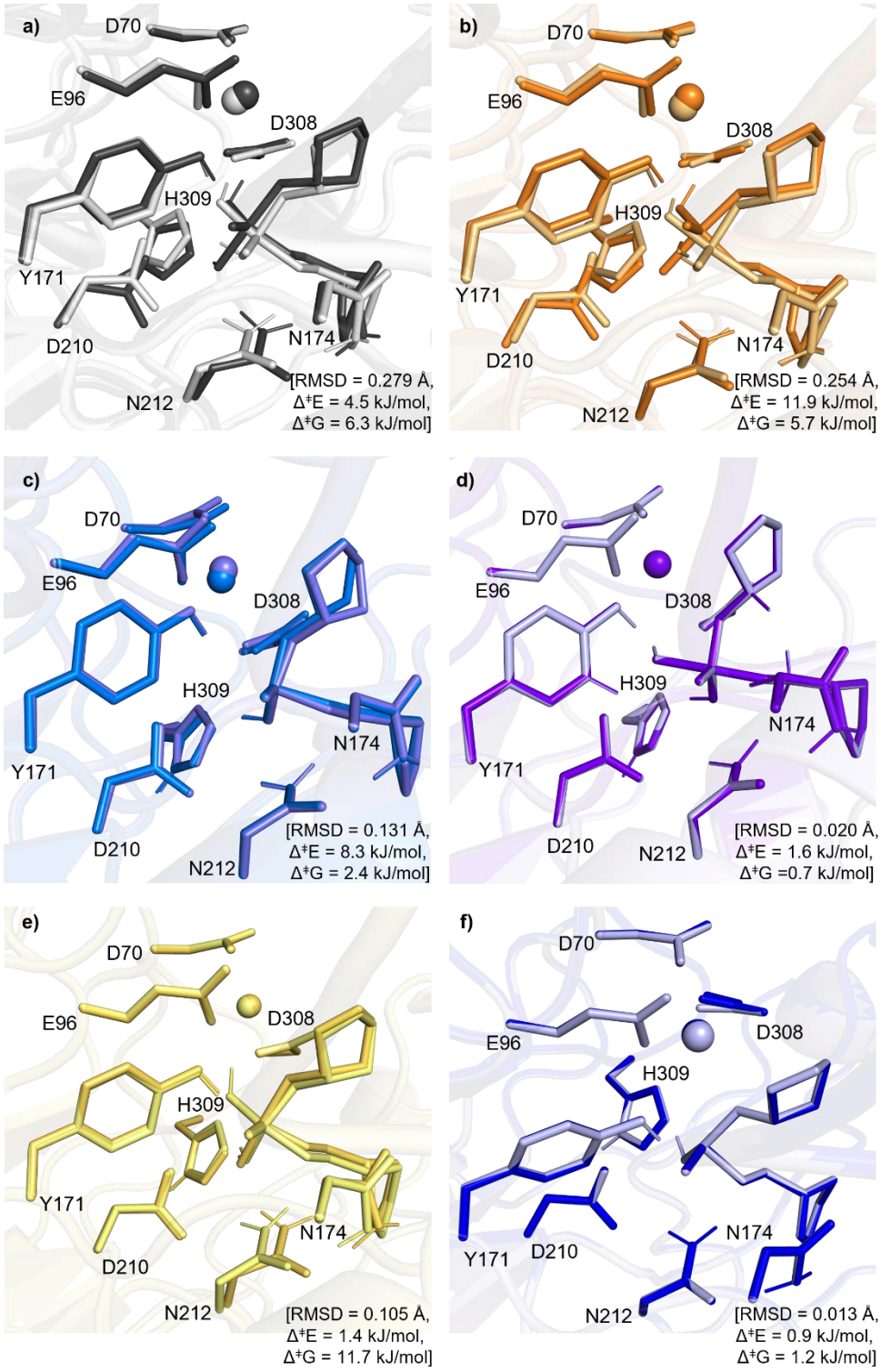

Figure B.1. Overlay of the active site of ICs obtained from IRCs corresponding to TS1 (IC) and TS2 (IC′) for the APE1-catalyzed phosphodiester bond cleavage facilitated by a) Mg^{2+}, b) Mn^{2+}, c) octahedral Ni^{2+}, d) square-planar Ni^{2+}, e) Zn^{2+}, and f) Ca^{2+}. The energy differences were calculated for IC′ with respect to IC.

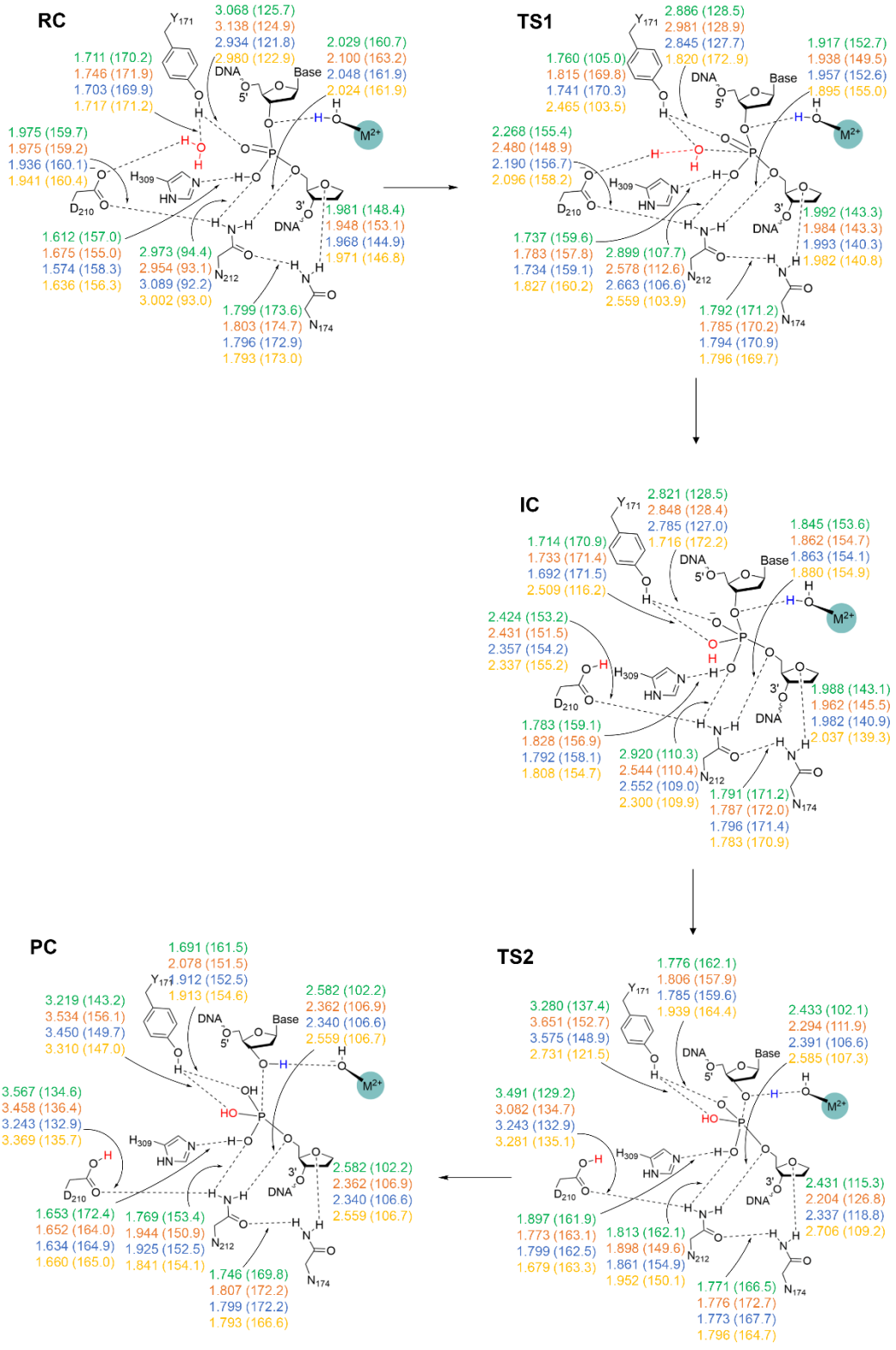

Figure B.2. Active site hydrogen-bond distances (Å) and angles (deg., in parentheses) for the APE1-catalyzed phosphodiester bond cleavage facilitated by Mg^{2+} (green), Mn^{2+} (orange), Ni^{2+} (octahedral coordination, blue) or Zn^{2+} (yellow).

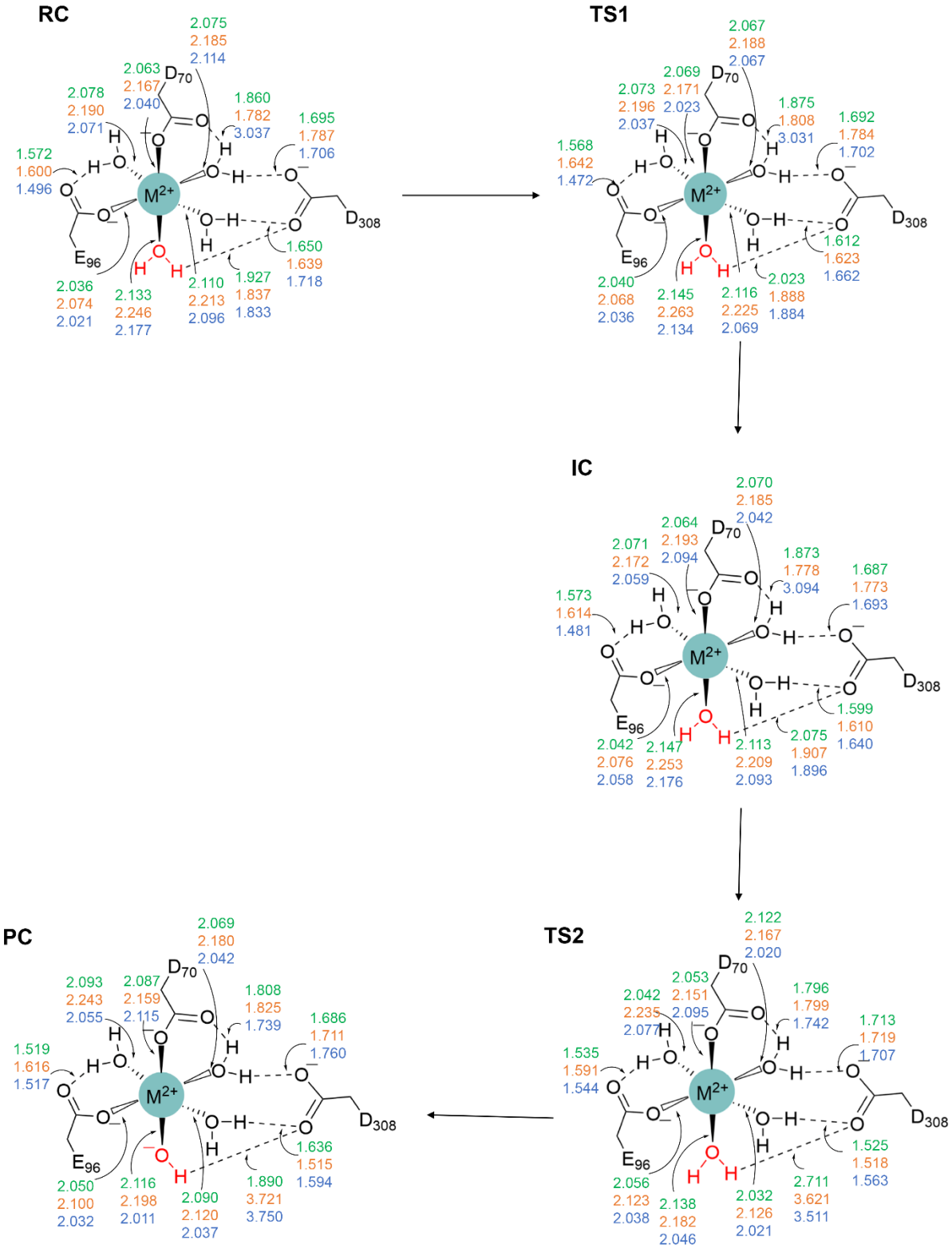

Figure B.3. Metal coordination distances (Å) for the APE1-catalyzed phosphodiester bond cleavage facilitated by Mg^{2+} (green), Mn^{2+} (orange), or Ni^{2+} (octahedral coordination, blue). Leaving group protonating water is shown in red.

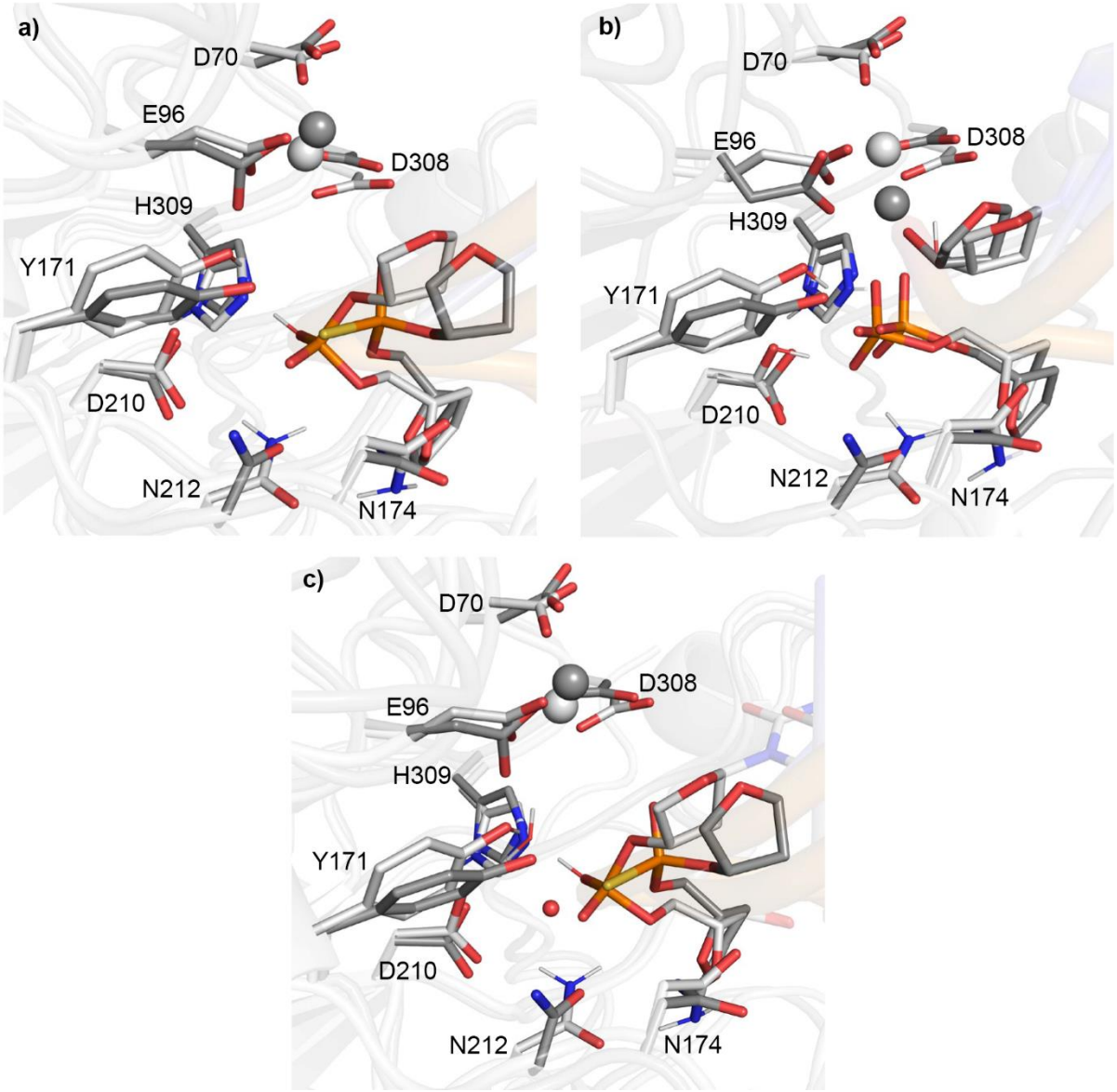

Figure B.4. Overlay of the active site from a) QM/MM Mg^{2+}–RC (light grey) and the crystal structure of Mn^{2+} containing thio-substituted reactant analogue complex (dark grey; PDB ID: 5DG0; RMSD = 0.990 Å), b) QM/MM Mg^{2+}–PC (light grey) and the crystal structure of Mg^{2+} containing product complex (dark grey; PDB ID: 4IEM; RMSD = 0.894 Å), and c) QM/MM Mn^{2+}–RC (light grey) and the crystal structure of Mn^{2+} containing thio-substituted reactant analogue complex (dark grey; PDB ID: 5DG0; RMSD = 0.943 Å).

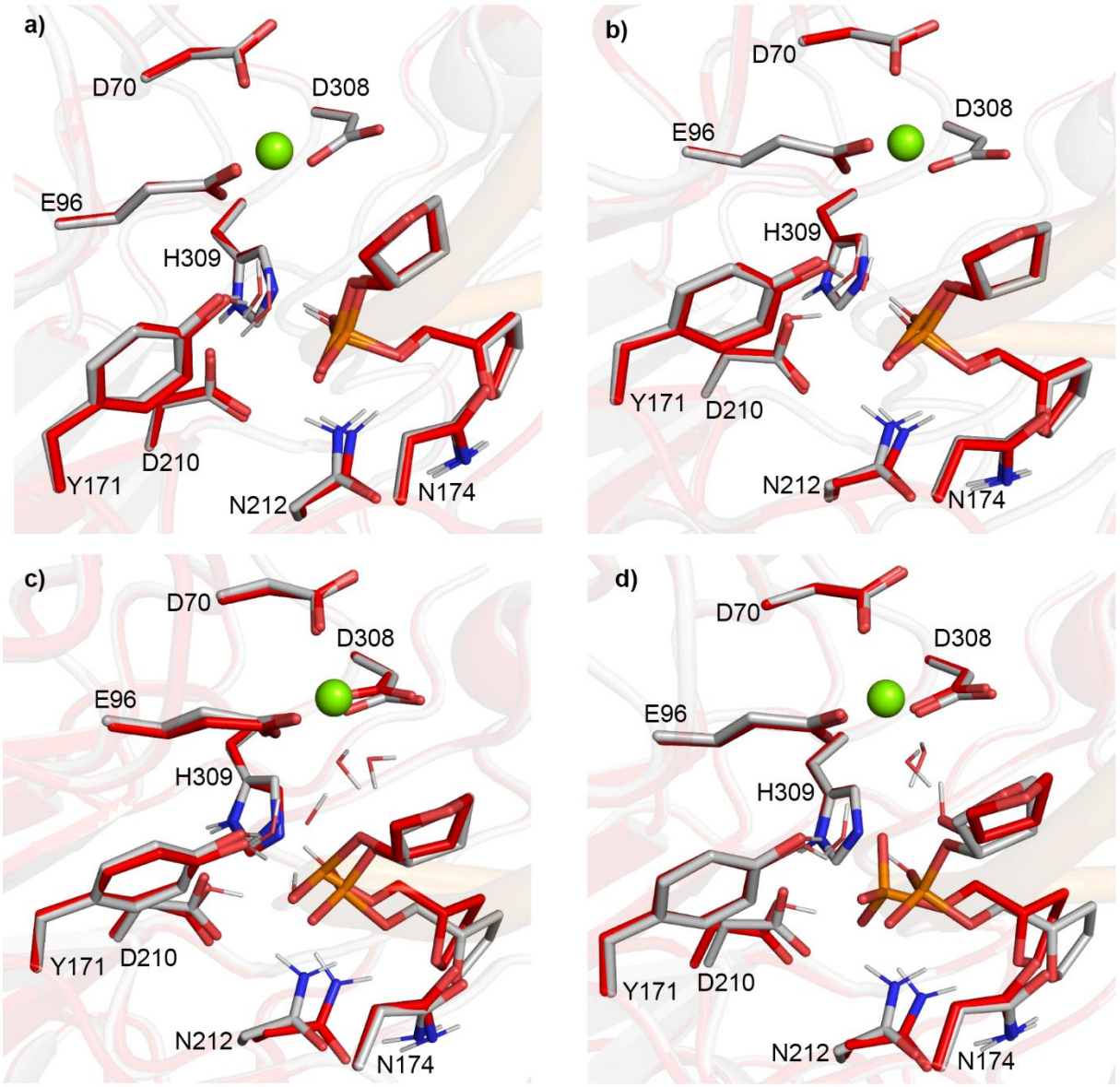

Figure B.5. Overlay of the active site for each stationary point along the APE1-catalyzed phosphodiester bond cleavage facilitated by Mg^{2+} (red) with respect to the RC (grey): a) Mg^{2+}–RC:Mg^{2+}–TS1 (RMSD = 0.138 Å), b) Mg^{2+}–RC:Mg^{2+}–IC (RMSD = 0.143 Å), c) Mg^{2+}–RC:Mg^{2+}–TS2 (RMSD = 0.269 Å), and d) Mg^{2+}–RC:Mg^{2+}–PC (RMSD = 0.321 Å).

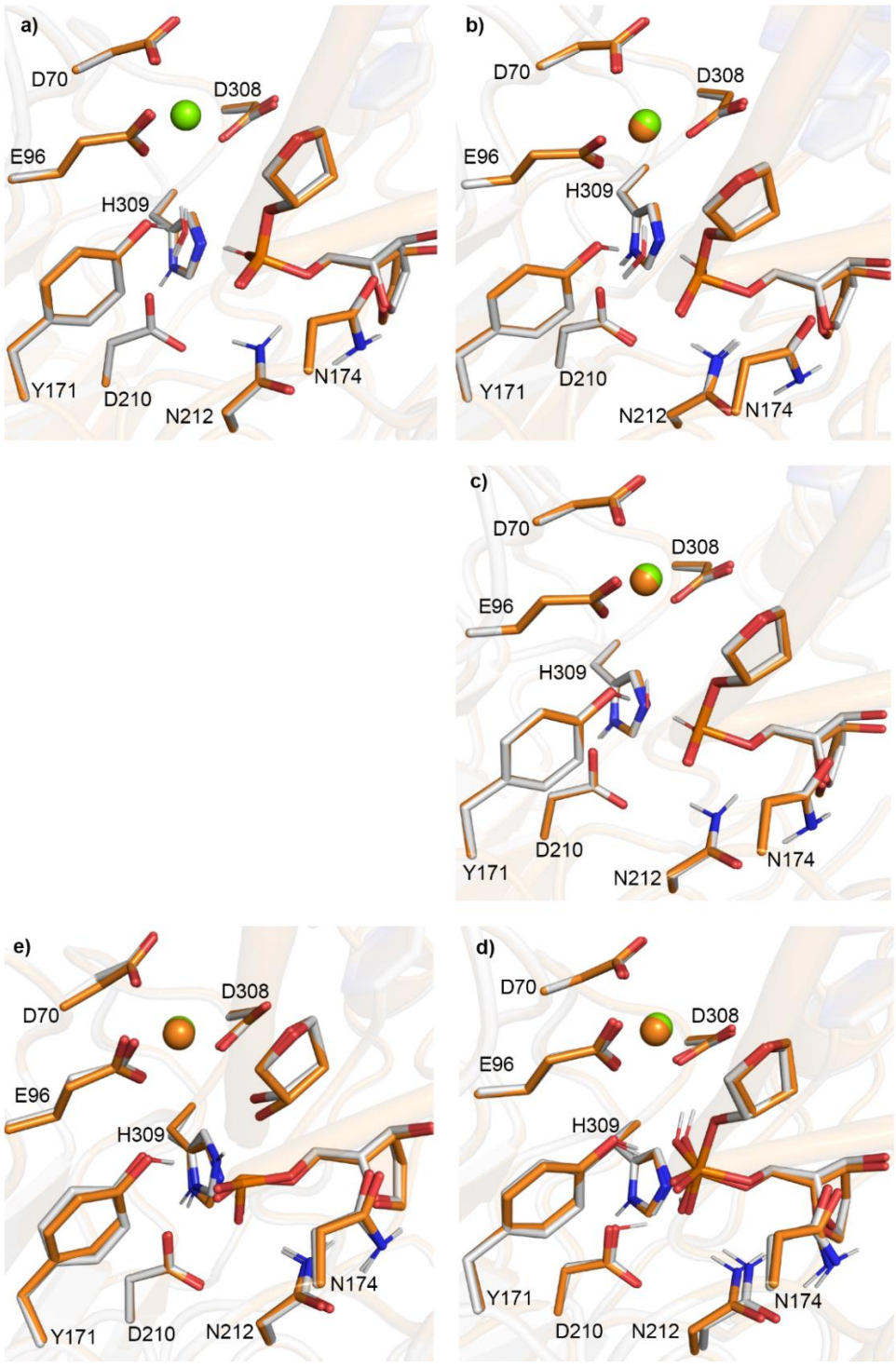

Figure B.6. Overlay of the active site for each stationary point along the APE1-catalyzed phosphodiester bond cleavage facilitated by Mn^{2+} (orange) with respect to the corresponding Mg^{2+} containing structure (grey): a) Mg^{2+}–RC:Mn^{2+}–RC (RMSD = 0.089 Å), b) Mg^{2+}–TS1:Mn^{2+}–TS1 (RMSD = 0.093 Å), c) Mg^{2+}–IC:Mn^{2+}–IC (RMSD = 0.093 Å), d) Mg^{2+}–TS2: Mn^{2+}–TS2 (RMSD = 0.195 Å), and e) Mg^{2+}–PC: Mn^{2+}–PC (RMSD = 0.183 Å).

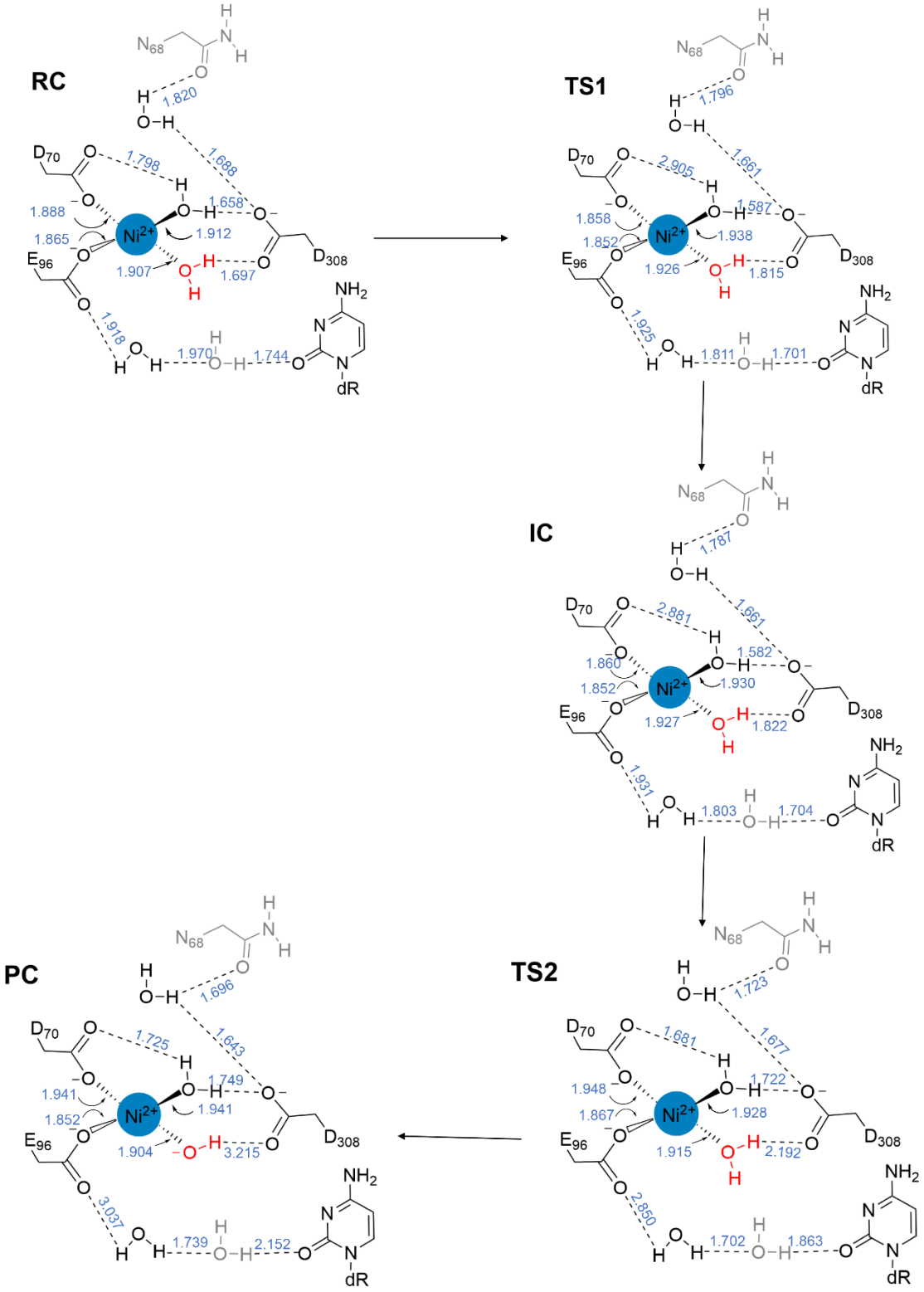

Figure B.7. Metal coordination distances (Å) for the APE1-catalyzed phosphodiester bond cleavage facilitated by square-planar Ni^{2+}. Low layer residues are shown in grey and leaving group protonating water in red.

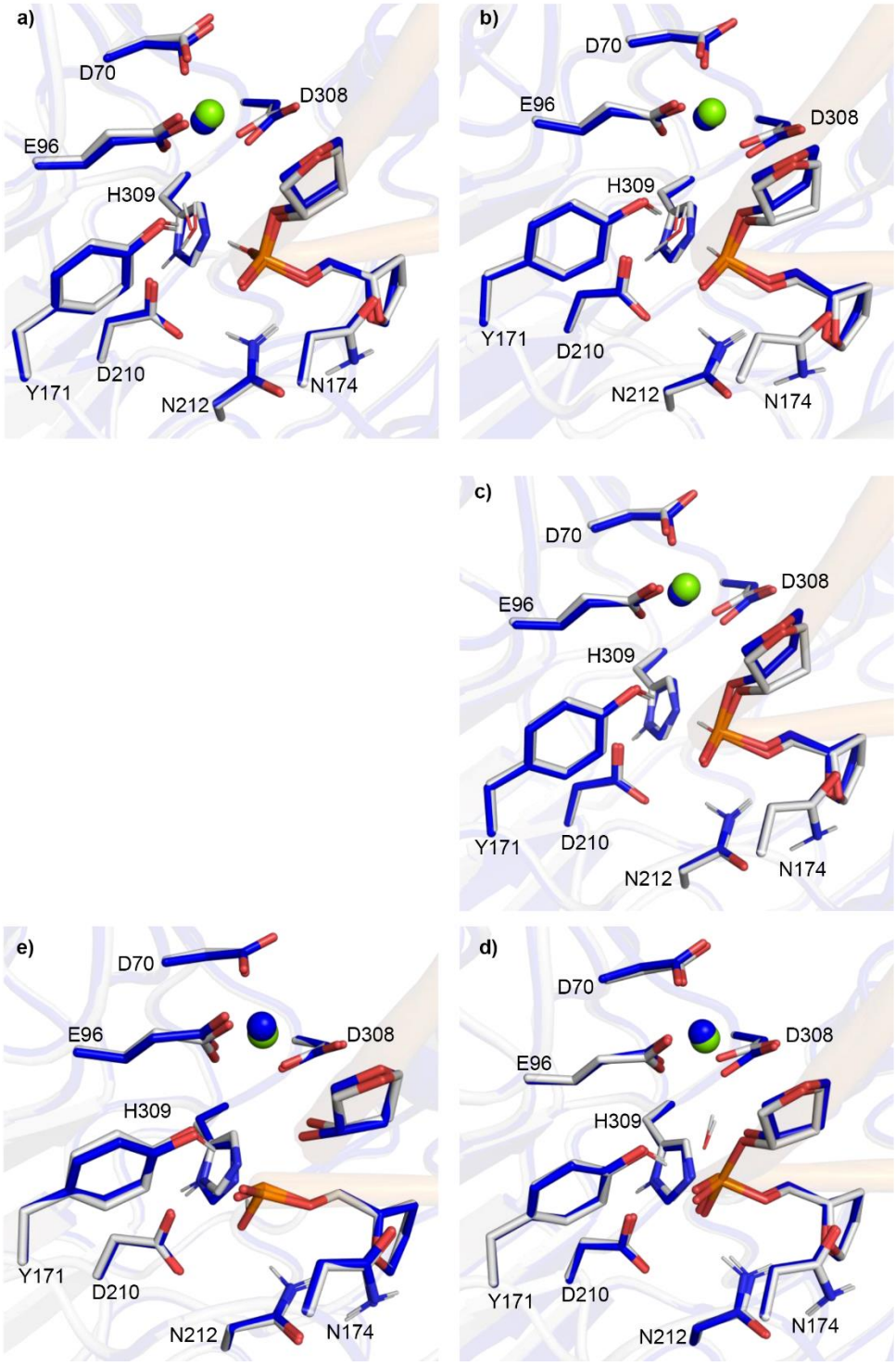

Figure B.8. Overlay of the active site for each stationary point along the APE1-catalyzed phosphodiester bond cleavage facilitated by square-planar Ni^{2+} (blue) with respect to the corresponding Mg^{2+} containing structure (grey): a) Mg^{2+}–RC:Ni^{2+}–RC (RMSD = 0.222 Å), b) Mg^{2+}–TS1:Ni^{2+}–TS1 (RMSD = 0.244 Å), c) Mg^{2+}–IC:Ni^{2+}–IC (RMSD = 0.246 Å), d) Mg^{2+}–TS2:Ni^{2+}–TS2 (RMSD = 0.227 Å), and e) Mg^{2+}–PC: Ni^{2+}–PC (RMSD = 0.289 Å).

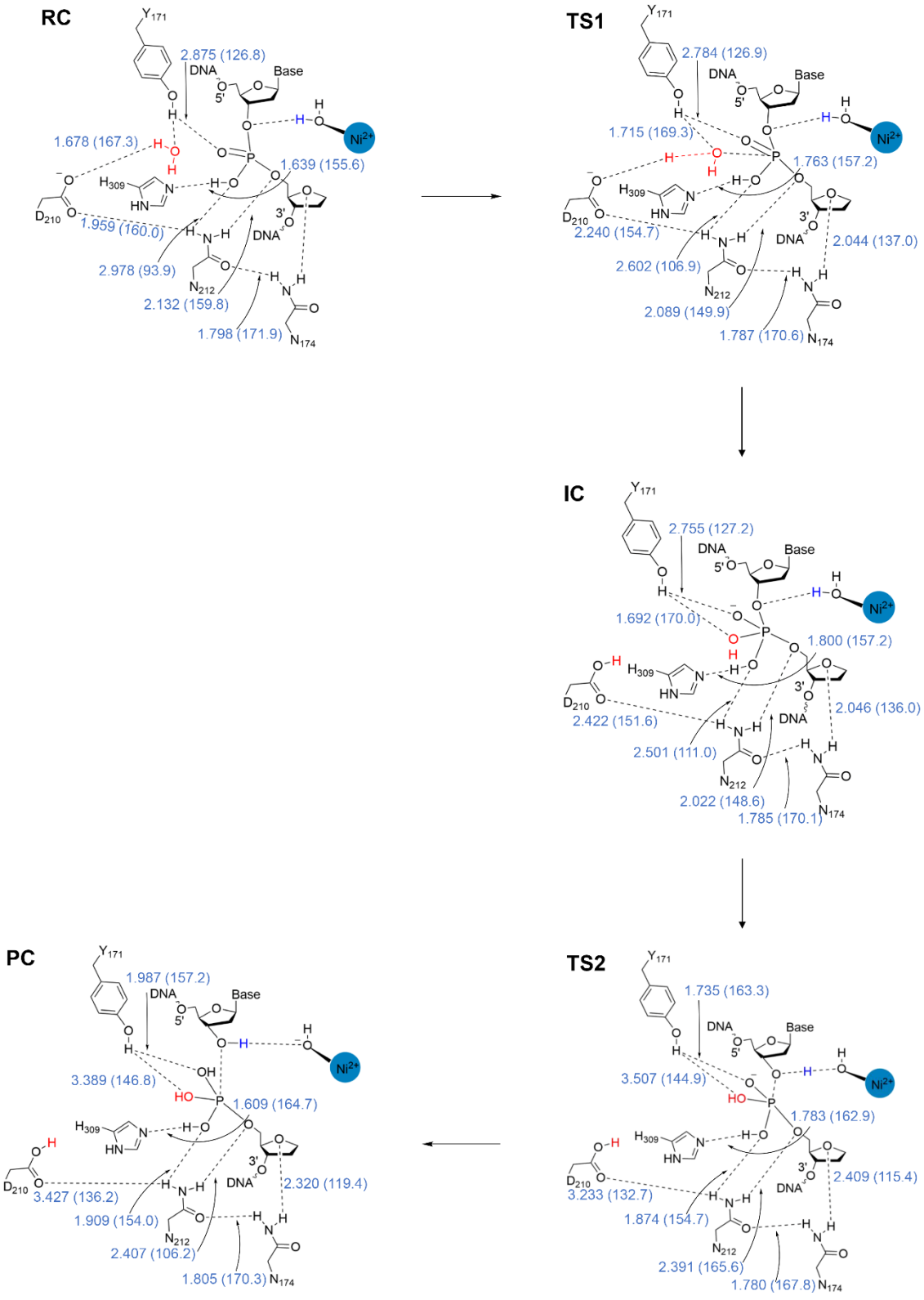

Figure B.9. Active site hydrogen-bond distances (Å) and angles (deg., in parentheses) for the APE1-catalyzed phosphodiester bond cleavage facilitated by square-planar Ni^{2+}.

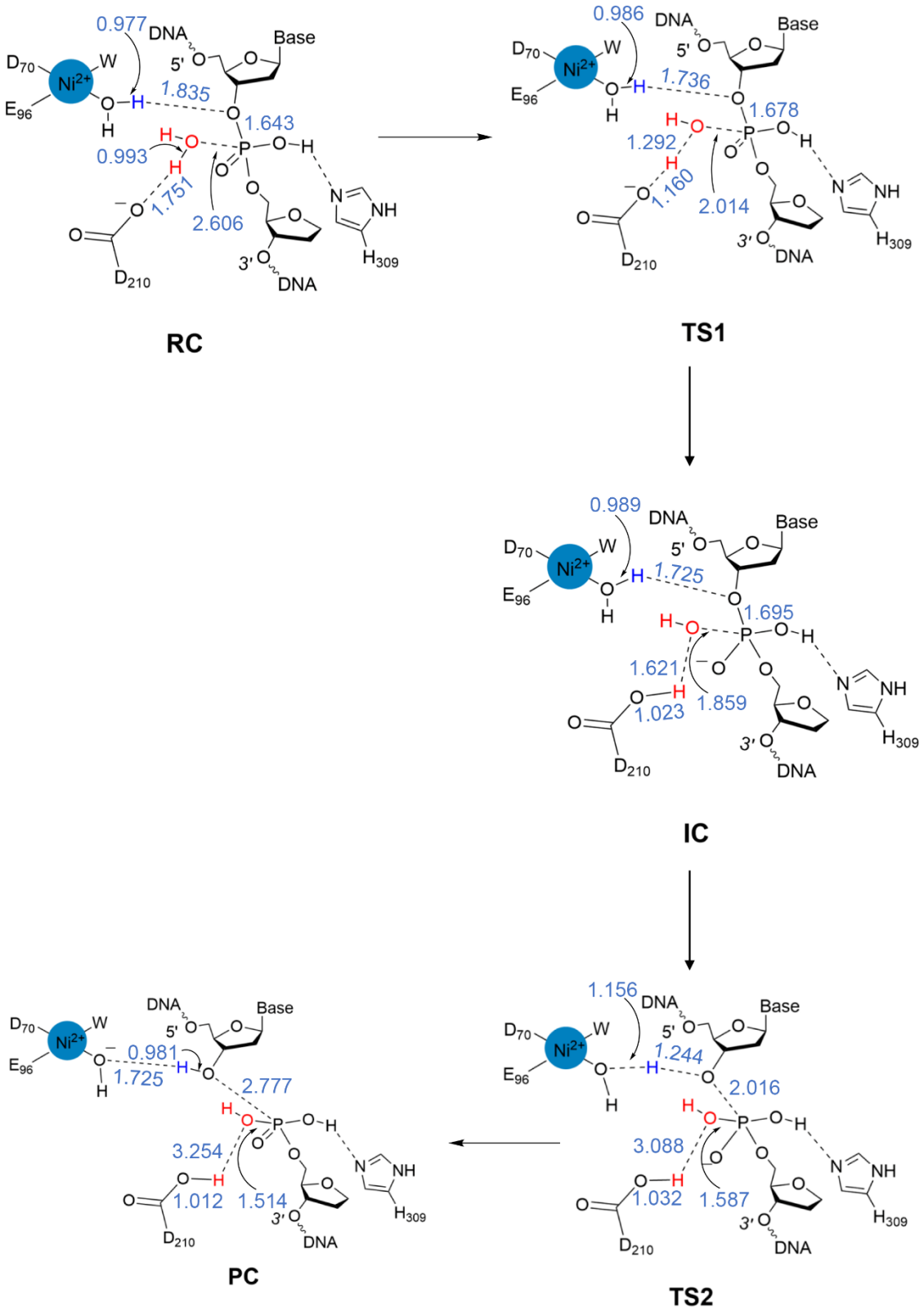

Figure B.10. Mechanism and key calculated bond distances (Å) for the APE1-catalyzed phosphodiester bond cleavage facilitated by square-planar Ni^{2+}.

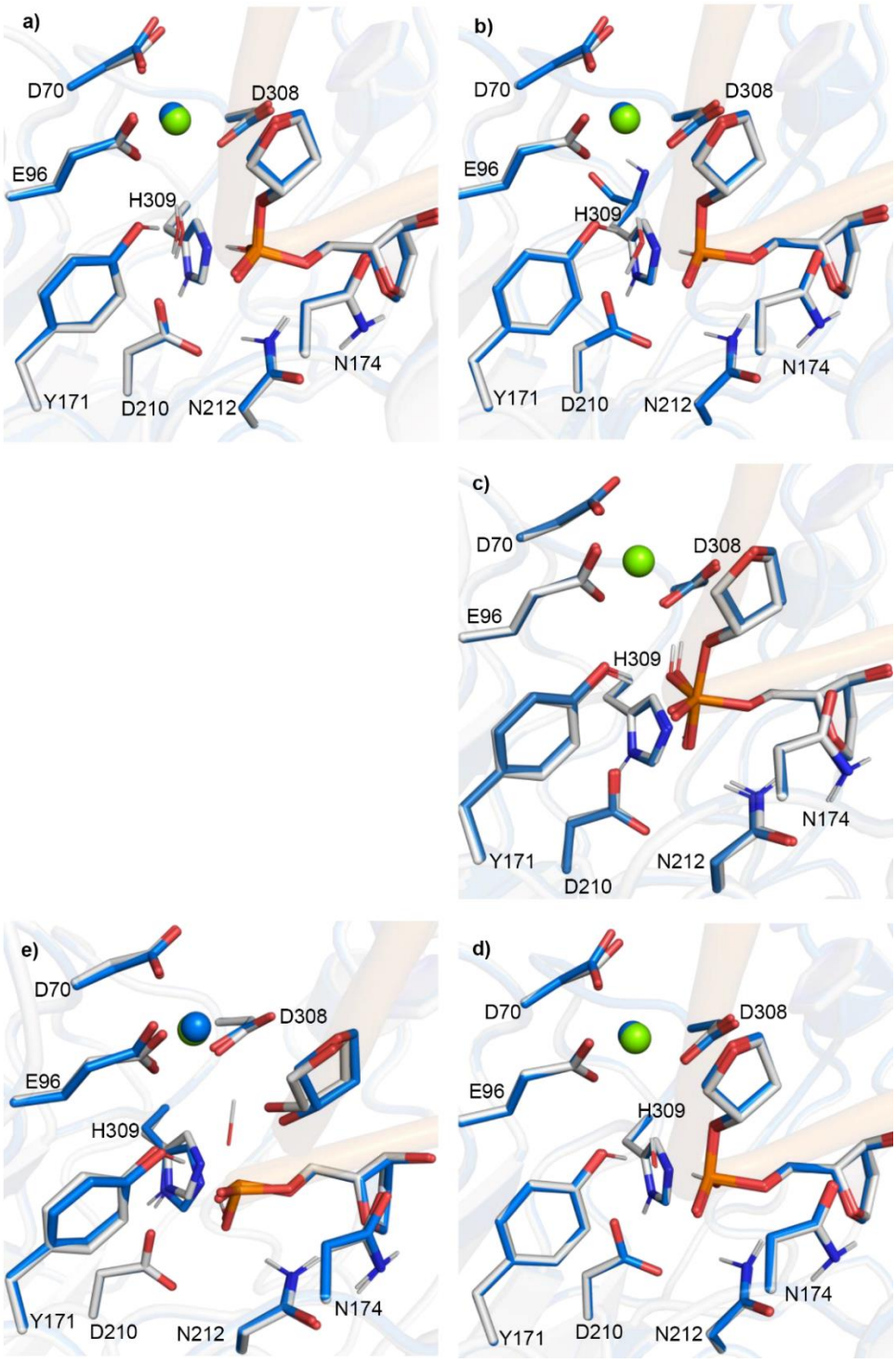

Figure B.11. Overlay of the active site for each stationary point along the APE1-catalyzed phosphodiester bond cleavage facilitated by octahedral Ni^{2+} (blue) with respect to the corresponding Mg^{2+} containing structure (grey): a) Mg^{2+}–RC:Ni^{2+}–RC (RMSD = 0.110 Å), b) Mg^{2+}–TS1:Ni^{2+}–TS1 (RMSD = 0.083 Å), c) Mg^{2+}–IC:Ni^{2+}–IC (RMSD = 0.124 Å), d) Mg^{2+}–TS2:Ni^{2+}–TS2 (RMSD = 0.117 Å), and e) Mg^{2+}–PC: Ni^{2+}–PC (RMSD = 0.211 Å).

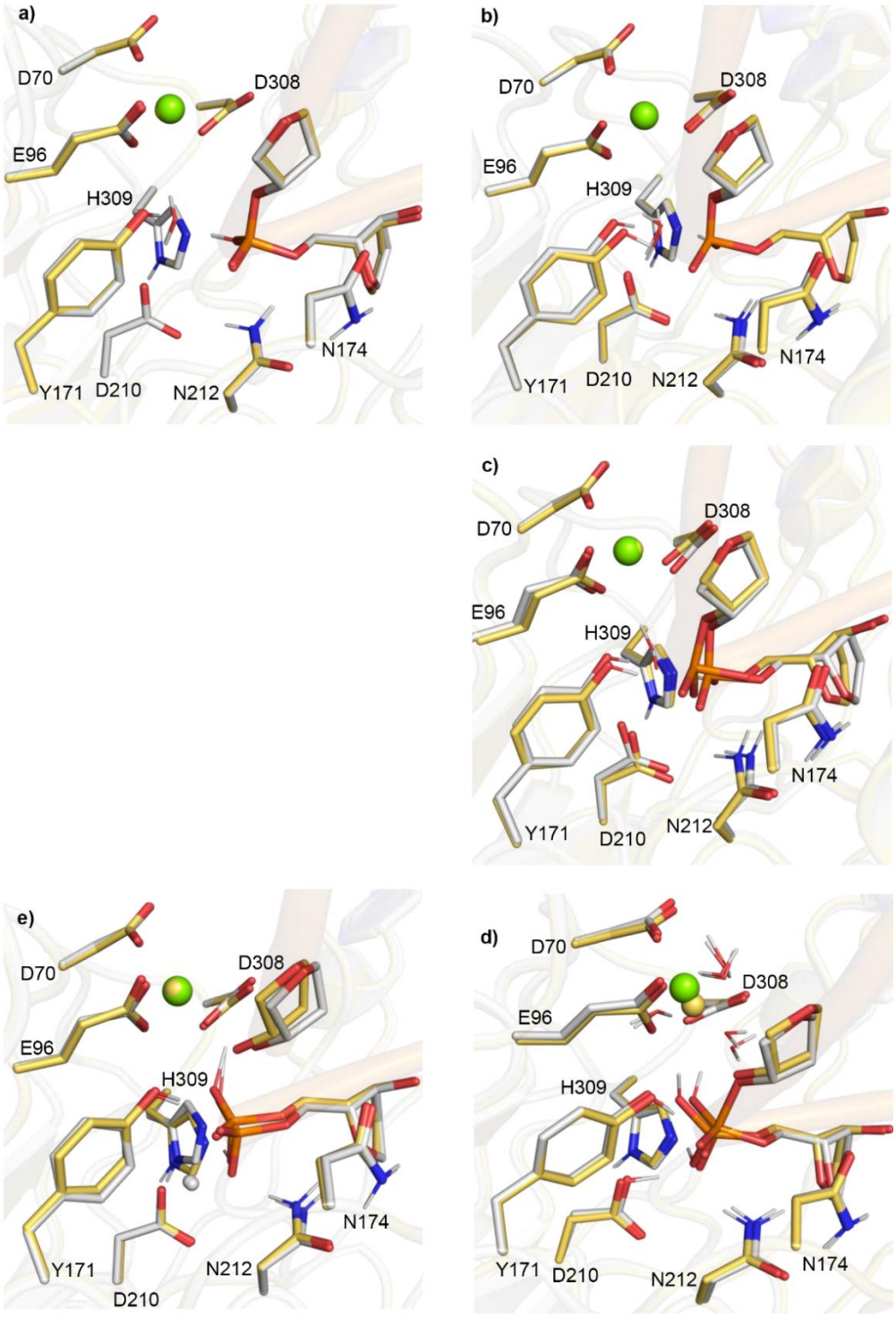

Figure B.12. Overlay of the active site for each stationary point along the APE1-catalyzed phosphodiester bond cleavage facilitated by Zn^{2+} (yellow) with respect to the corresponding Mg^{2+} containing structure (grey): a) Mg^{2+}–RC:Zn^{2+}–RC (RMSD = 0.060 Å), b) Mg^{2+}–TS1:Zn^{2+}–TS1 (RMSD = 0.144 Å), c) Mg^{2+}–IC:Zn^{2+}–IC (RMSD = 0.269 Å), d) Mg^{2+}–TS2:Zn^{2+}–TS2 (RMSD = 0.212 Å) and e) Mg^{2+}–PC: Zn^{2+}–PC (RMSD = 0.199 Å).

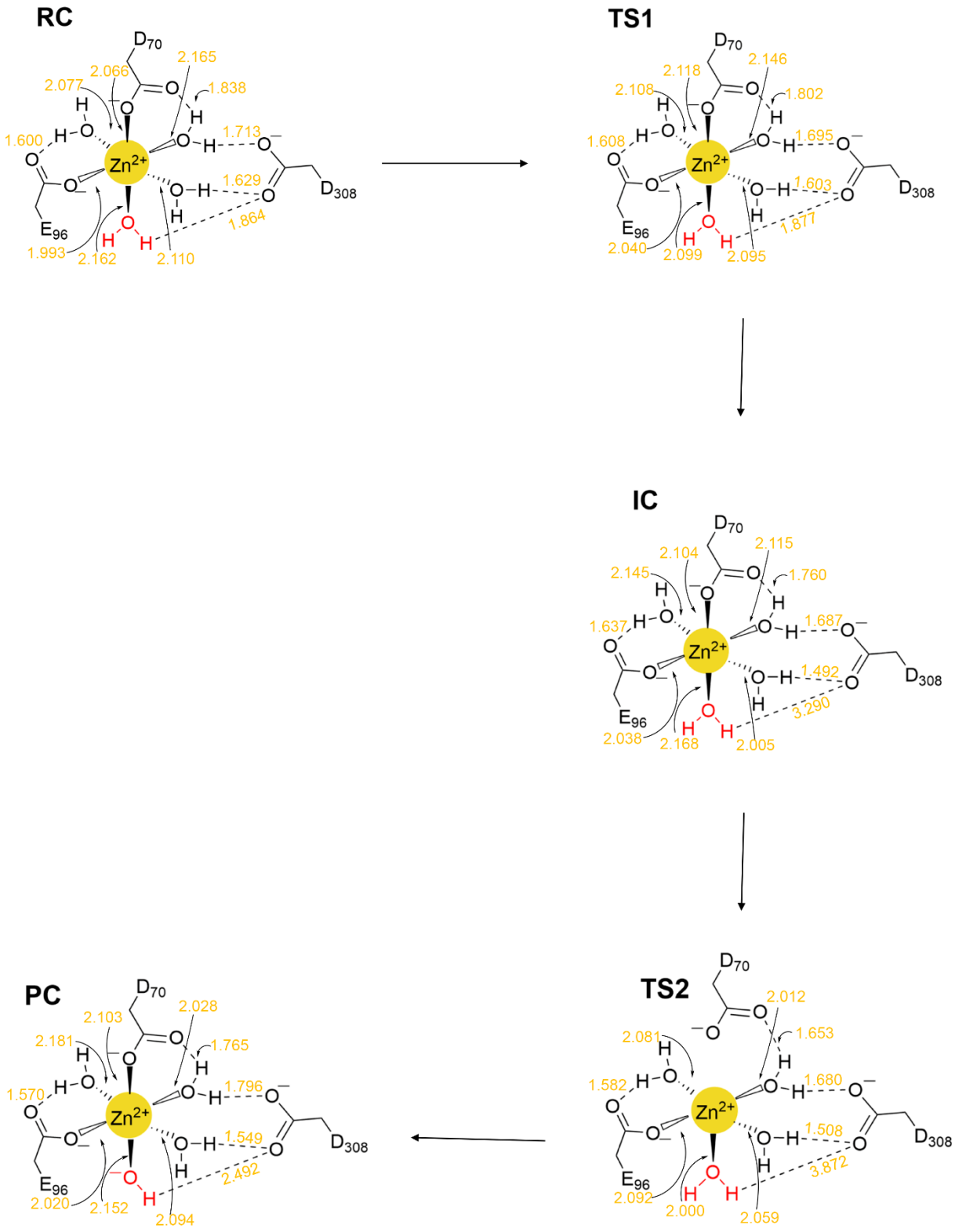

Figure B.13. Metal coordination distances (Å) for the APE1-catalyzed phosphodiester bond cleavage facilitated by Zn^{2+}. Leaving group protonating water is shown in red.

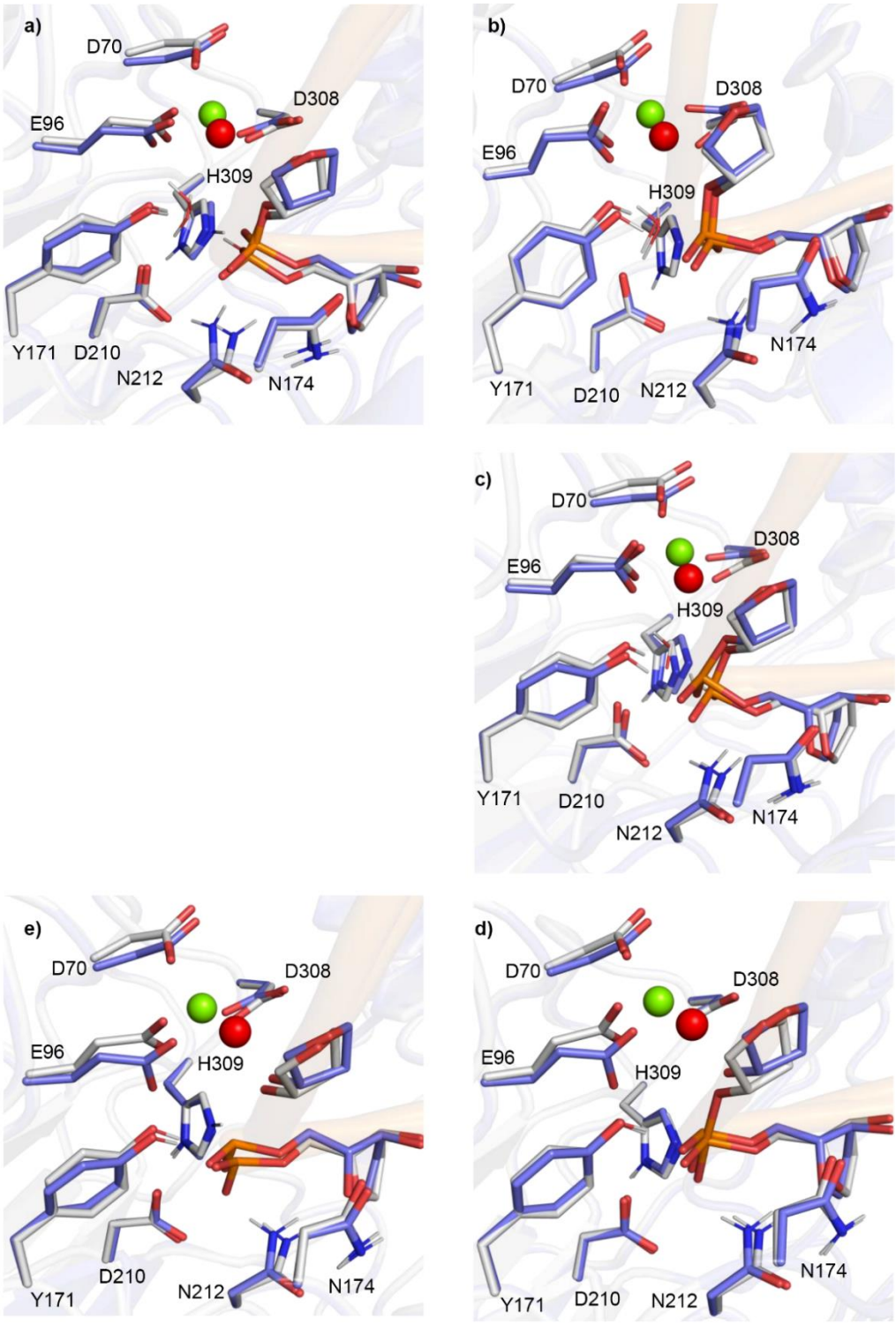

Figure B.14. Overlay of the active site for each stationary point along the APE1-catalyzed phosphodiester bond cleavage inhibited by Ca^{2+} (blue) with respect to the corresponding Mg^{2+} containing structure (grey): a) Mg^{2+}–RC:Ca^{2+}–RC (RMSD = 0.383Å), b) Mg^{2+}–TS1:Ca^{2+}–TS1 (RMSD = 0.191 Å), c) Mg^{2+}–IC:Ca^{2+}–IC (RMSD = 0.347 Å), d) Mg^{2+}–TS2:Ca^{2+}–TS2 (RMSD = 0.170 Å), and e) Mg^{2+}–PC:Ca^{2+}–PC (RMSD = 0.331 Å).

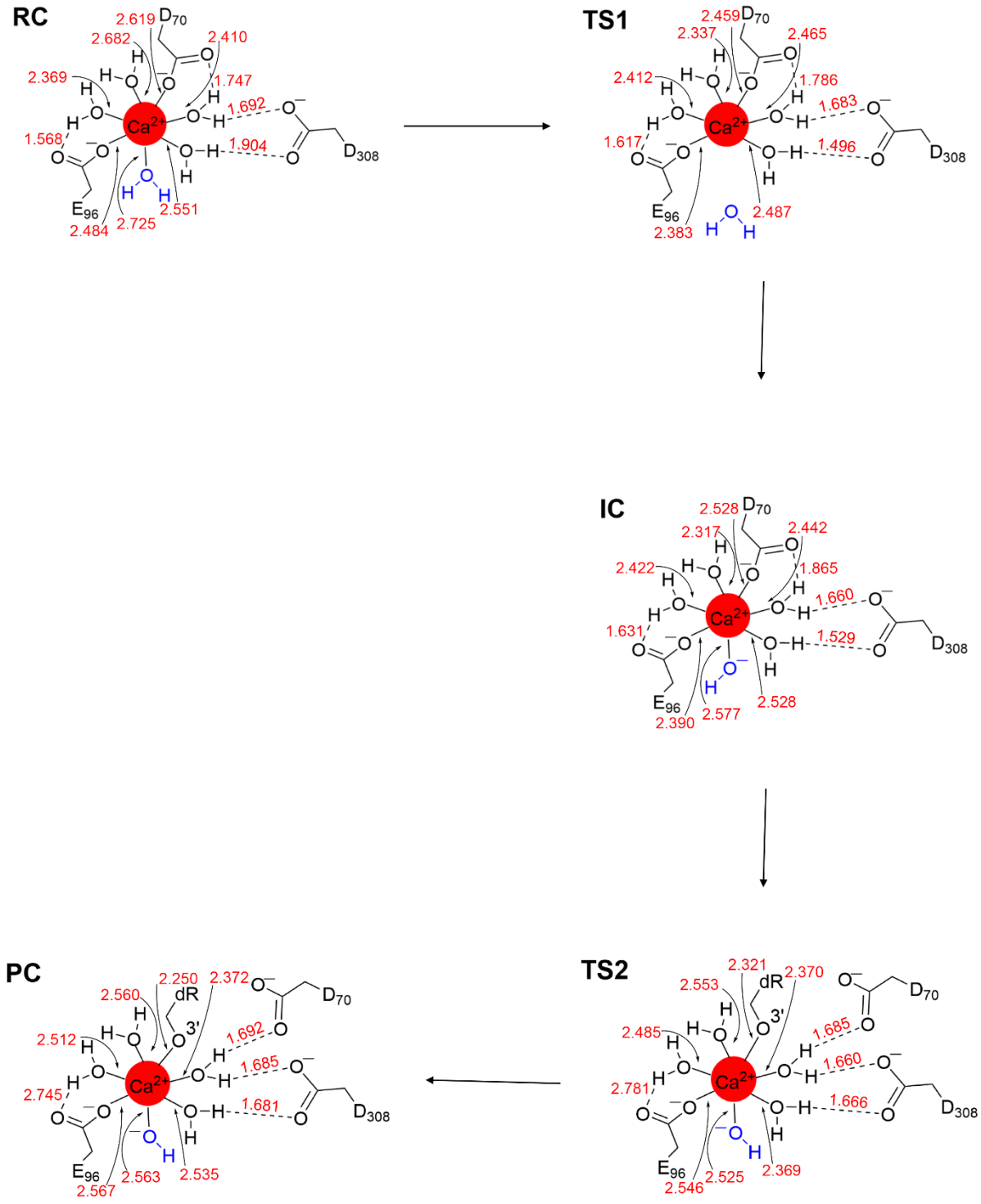

Figure B.15. Metal coordination distances (Å) for the APE1-catalyzed phosphodiester bond cleavage inhibited by Ca^{2+}. Nucleophilic water is shown in blue.

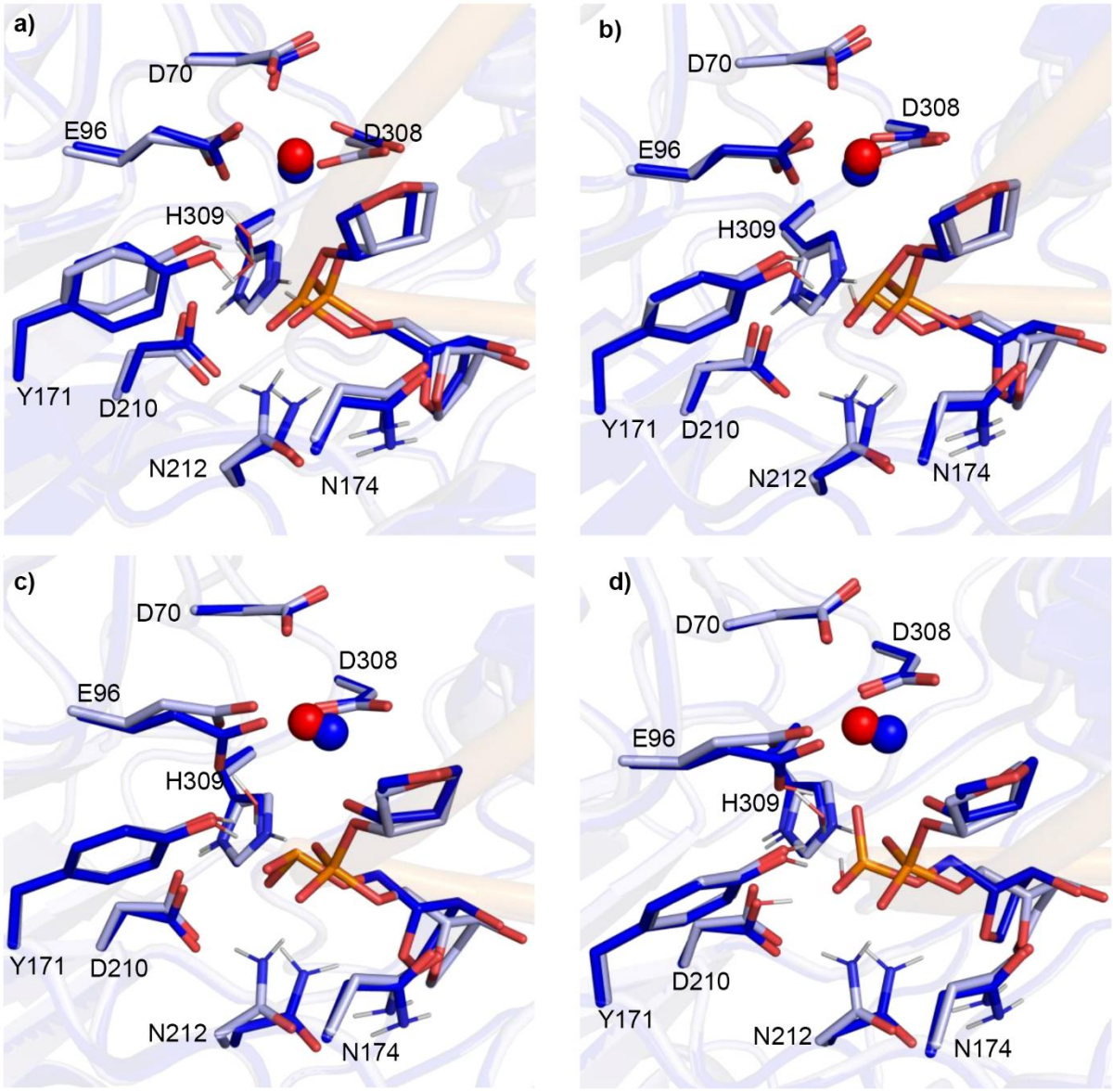

Figure B.16. Overlay of the active site for each stationary point along the APE1-catalyzed phosphodiester bond cleavage inhibited by Ca^{2+} (dark blue) with respect to the RC (light blue): a) Ca^{2+}–RC:Ca^{2+}–TS1 (RMSD = 0.455 Å), b) Ca^{2+}–RC:Ca^{2+}–IC (RMSD = 0.346 Å), c) Ca^{2+}–RC:Ca^{2+}–TS2 (RMSD = 0.192 Å), and d) Ca^{2+}–RC:Ca^{2+}–PC (RMSD = 0.288 Å).

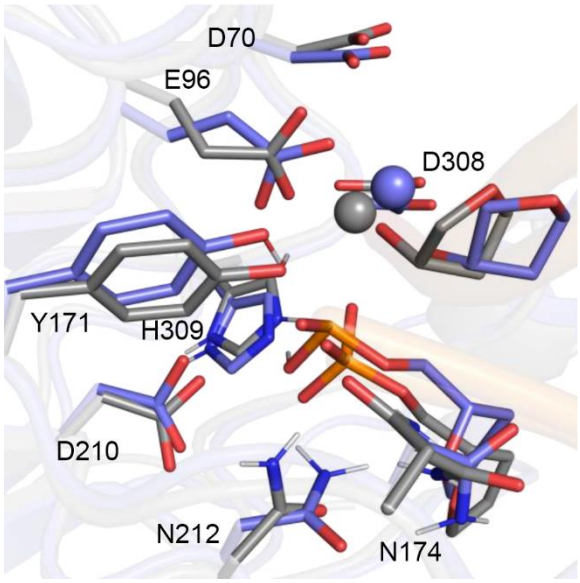

Figure B.17. Overlay of the active site for the QM/MM Ca^{2+}–PC (blue) and the crystal structure of the Mg^{2+} containing product complex (dark grey; PDB ID: 4IEM; RMSD = 0.842 Å).

Appendix C

Supplementary Information for Chapter 4: 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

Contains Tables C.1–C.3 and Figures C.1–C.14

Table C.1. Relative Gibbs energies (kJ/mol) for the I-*Ppo*I facilitated phosphodiester bond cleavage characterized using QM cluster models with M06-2X or ωB97M-V (in parantheses) for the pathways involving either indirect or direct Mg^{2+} coordination to the leaving group. [a,b,c]

	RC	TS1	TS/IC	TS2	PC
Model 1 (Indirect)	0.0		182.3 (179.6)		115.0 (101.3)
Model 1 (Direct)	0.0		91.6 (87.8)		41.8 (39.4)
Model 2 (Indirect)	0.0	107.9 (102.2)	91.3 (85.2)	111.5 (103.1)	–32.8 (–35.4)
Model 2 (Direct)	0.0		92.6 (95.4)		7.4 (10.4)
Model 3 (Indirect)	0.0		94.8 (102.8)		15.4 (31.2)
Model 3 (Direct)	0.0		78.4 (72.6)		1.1 (–16.7)
Model 4 (Indirect)	0.0		124.3 (132.2)		23.2 (24.1)
Model 4 (Direct)	0.0		103.2 (94.5)		–7.8 (–17.9)

[a]Single-point calculations were carried out using Gaussian16 (rev B.01) with IEF-PCM-M06-2X/6-311+G(2df,p) (ε=4). [b]Single-point calculations were carried out using ORCA 5.0.4 with CPCM-ωB97M-V/6-311+G(2df,2p) (ε=4.9). All geometries were optimized in the gas phase using B3LYP-D3(BJ)/6-31G(d,p). [c]Refer to Figure C.2 for a schematic of the QM cluster models 1 – 4.

Table C.2. Relative Gibbs energies (kJ/mol) for the I-*Ppo*I facilitated phosphodiester bond cleavage characterized using QM cluster or QM/MM models for the pathway involving indirect Mg^{2+} coordination to the leaving group.[a,b,c]

	RC	TS1	TS/IC	TS2	PC
Model 1	0.0		182.3 (187.8)		115.0 (132.2)
Model 2	0.0	107.9 (108.7)	91.3 (90.9)	111.5 (113.6)	–32.8 (–33.1)
Model 3	0.0		94.8 (89.6)		15.4 (11.1)
Model 4	0.0		124.3 (128.6)		23.2 (50.1)
QM/MM	0.0		130.9		95.0

[a]Relative energies for the QM cluster models were obtained from IEF-PCM-M06-2X/6-311+G(2df,p) (ε=4) single-point calculations on B3LYP-D3(BJ)/6-31G(d,p) gas-phase geometries. Corresponding gas-phase M06-2X/6-311+G(2df,p) relative energies are provided in parentheses. [b]Relative energies for the QM/MM model were obtained from ONIOM(M06-2X/6-311+G(2df,p):AMBER) single-point calculations on ONIOM(B3LYP-D3(BJ)/6-31G(d,p) :AMBER) geometries. [c]Refer to Figure C.2 for a schematic of the QM cluster models 1 – 4 and QM/MM model.

Table C.3. Relative Gibbs energies (kJ/mol) for the I-*Ppo*I facilitated phosphodiester bond cleavage characterized using QM cluster or QM/MM models for the pathway involving direct Mg^{2+} coordination to the leaving group.[a,b,c]

	RC	TS	PC
Model 1	0.0	91.6 (97.6)	41.8 (54.4)
Model 2	0.0	92.6 (86.4)	7.4 (3.0)
Model 3	0.0	78.4 (80.3)	1.1 (4.5)
Model 4	0.0	103.2 (77.2)	–7.8 (21.3)
QM/MM	0.0	54.1 [77.4]	–73.8 [–101.8]

[a]Relative energies for the QM cluster models were obtained from IEF-PCM-M06-2X/6-311+G(2df,p) (ε=4) single-point calculations on B3LYP-D3(BJ)/6-31G(d,p) gas-phase geometries. Corresponding gas-phase M06-2X/6-311+G(2df,p) relative energies are provided in parentheses. QM/MM relative energies evaluated using electrostatic embedding are provided in square brackets. [b]Relative energies for the QM/MM model were obtained from ONIOM(M06-2X/6-311+G(2df,p):AMBER) single-point calculations on ONIOM(B3LYP-D3(BJ)/6-31G(d,p) :AMBER) geometries. [c]Refer to Figure 4.2 in the main text for a schematic of the QM cluster models 1 – 4 and QM/MM model.

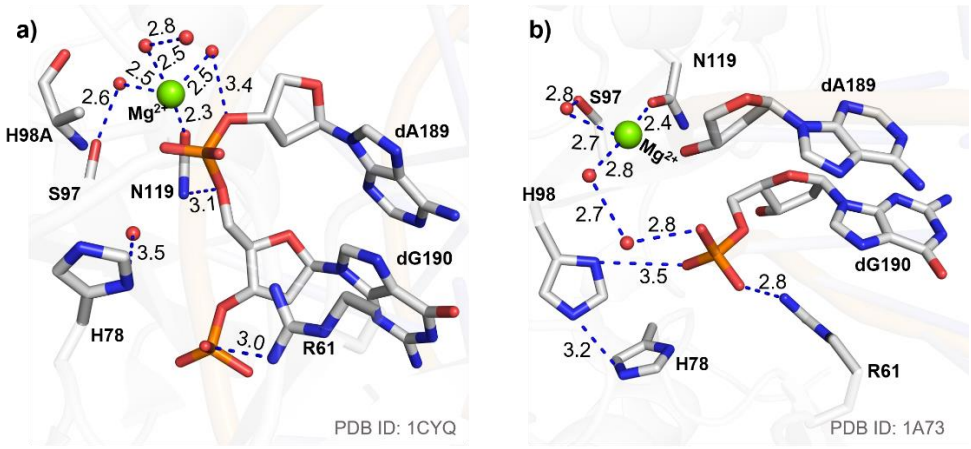

Figure C.1. Key distances (Å) in the I-*Ppo*I active site based on an X-ray crystal structures of a) the Mg^{2+}-containing RC analogue of the H98A mutant or b) the Mg^{2+}-containing PC of the wild-type enzyme.

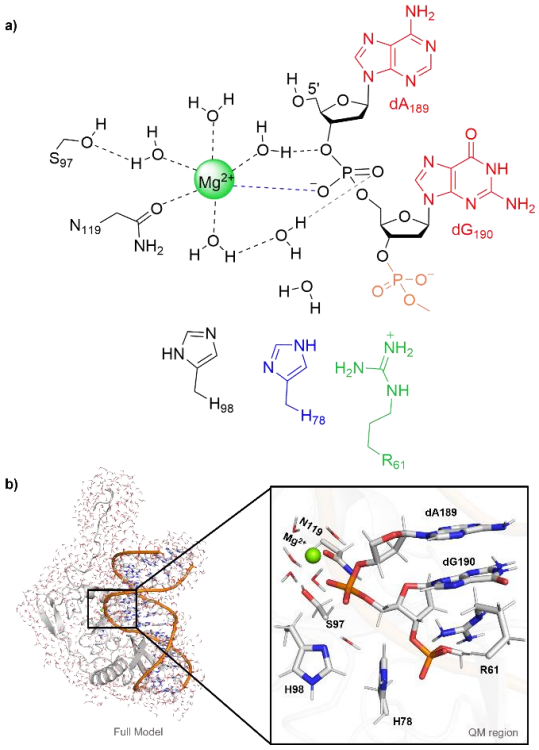

Figure C.2. a) Schematic of QM cluster models involving indirect Mg^{2+} coordination to the leaving group considered in this work: Model 1 (black), Model 2 (Model 1 + H78 (blue) + expanded substrate (red)), Model 3 (Model 1 + R61 (green) + expanded substrate), Model 4 (Model 1 + H78 + R61 + 5′ phosphate moiety of dA191 on the 3′-side of dG190 (orange)). b) The corresponding enzyme–DNA QM/MM (ONIOM) model (left) and the QM region (black box, right).

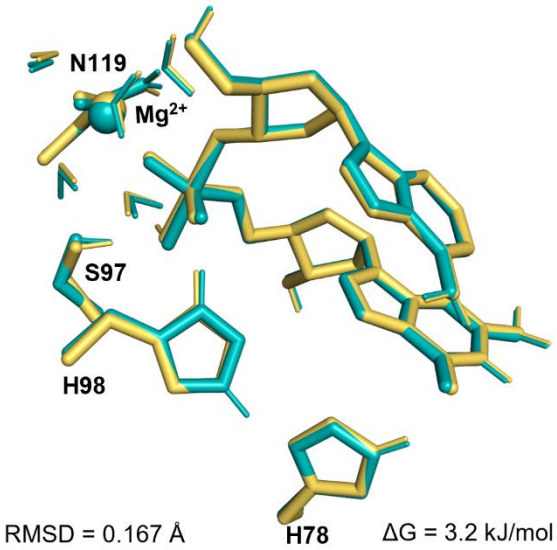

Figure C.3. Overlay of the active sites of ICs obtained from the IRC corresponding to TS1 (IC, yellow) and TS2 (IC′, teal) for the phosphodiester bond cleavage pathway characterized using QM cluster Model 2 with indirect Mg^{2+} coordination to the leaving group. The energy difference was calculated for IC′ with respect to IC.

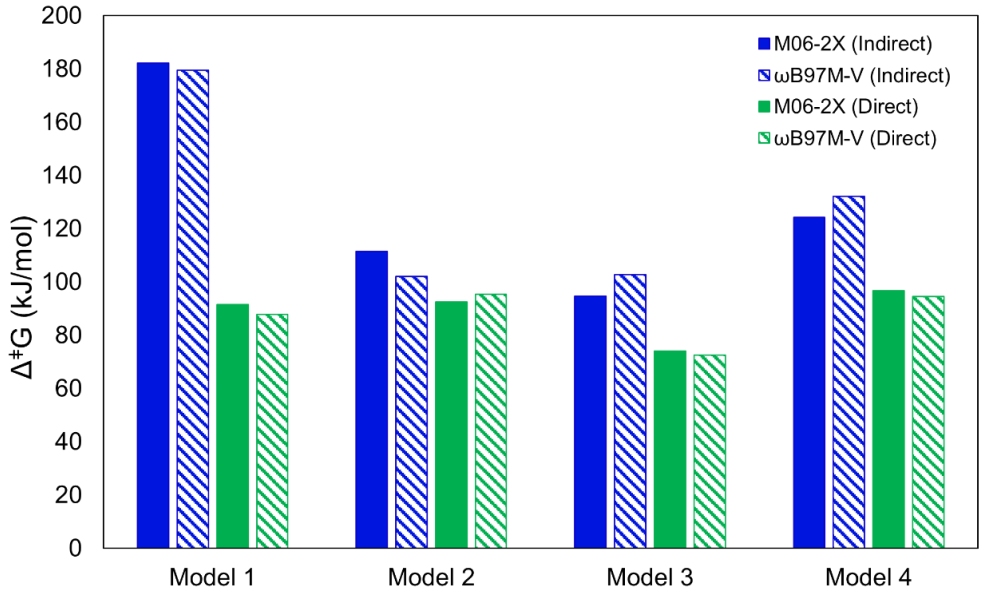

Figure C.4. Comparison of the Gibbs activation energies ($\Delta^{\ddagger}G$) characterized using QM cluster models for the rate-determining step obtained using M06-2X (solid bars) and ωB97M-V (dashed bars) single-point calculations on B3LYP-D3(BJ)/6-31G(d,p) gas-phase geometries for pathways involving indirect (water-mediated, blue) or direct (green) Mg^{2+}–O3′ leaving group coordination.

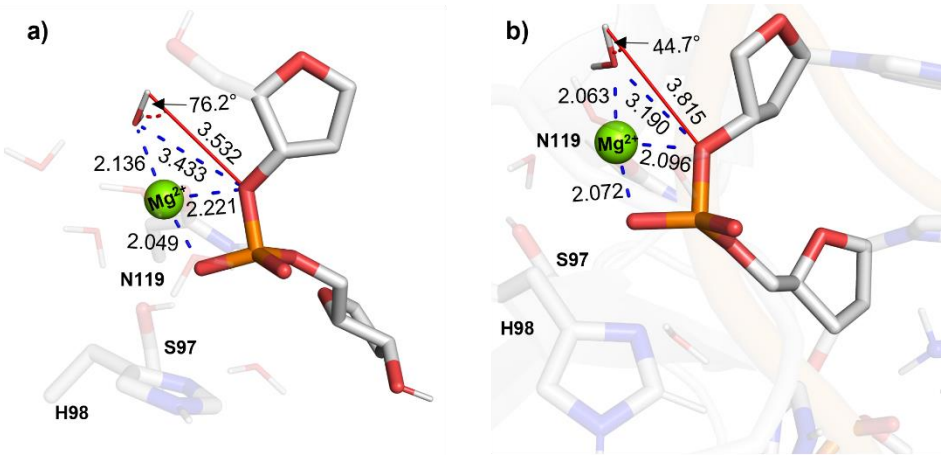

Figure C.5. Key distances (Å) and angles (°) in the I-*Ppo*I active site for the RC optimized using a) QM cluster Model 1 or b) QM/MM model based on the experimentally-proposed phosphodiester bond cleavage pathway.[1]

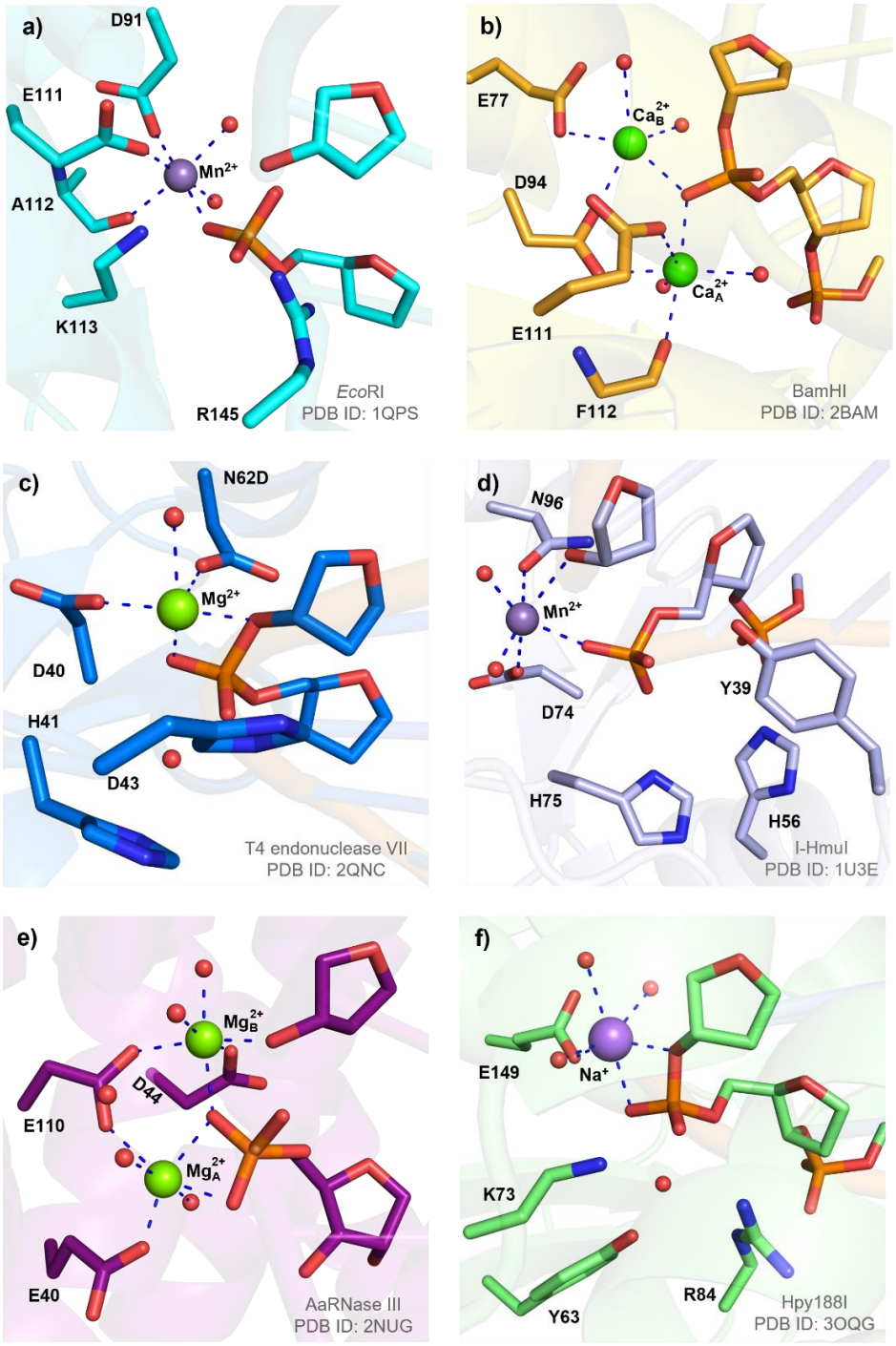

Figure C.6. Active site from an X-ray crystal structure of the a) Mn^{2+}-containing PC of wild-type *Eco*RI bound to a ssDNA substrate, b) Ca^{2+}-containing RC of wild-type BamHI bound to a dsDNA substrate, c) Mg^{2+}-containing RC of the N62D mutant of T4 endonuclease VII bound to a DNA holliday junction, d) Mn^{2+}-containing PC of wild-type I-HmuI bound to a dsDNA substrate, e) Mg^{2+}-containing PC of wild-type AaRNase III bound to a dsRNA substrate, or f) Na^{+}-substituted RC of Hpy188I bound to a dsDNA substrate.

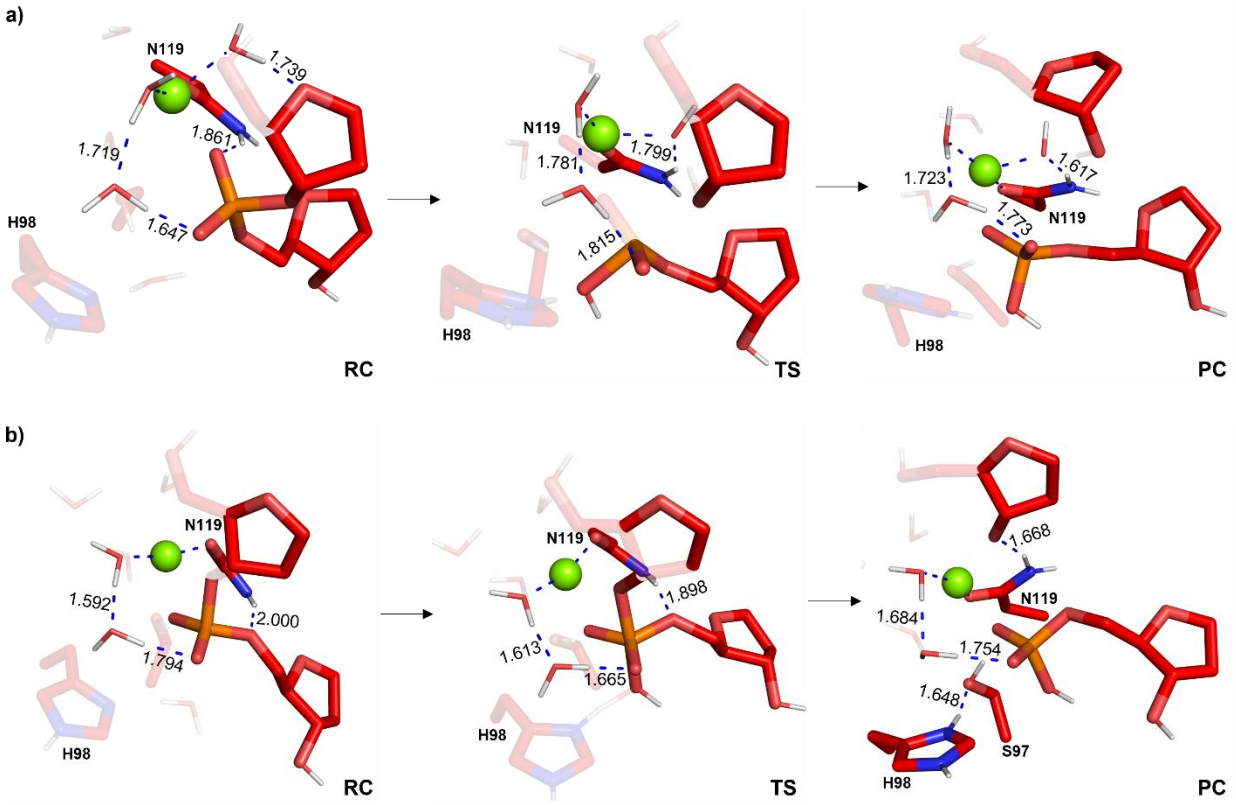

Figure C.7. Key distances (Å) in the I-*Ppo*I active site for the phosphodiester bond cleavage pathway characterized using QM cluster Model 1 with a) indirect or b) direct Mg^{2+} coordination to the leaving group. See main text Figures 4.3b and 4.5b for key reaction parameters for QM cluster Model 1.

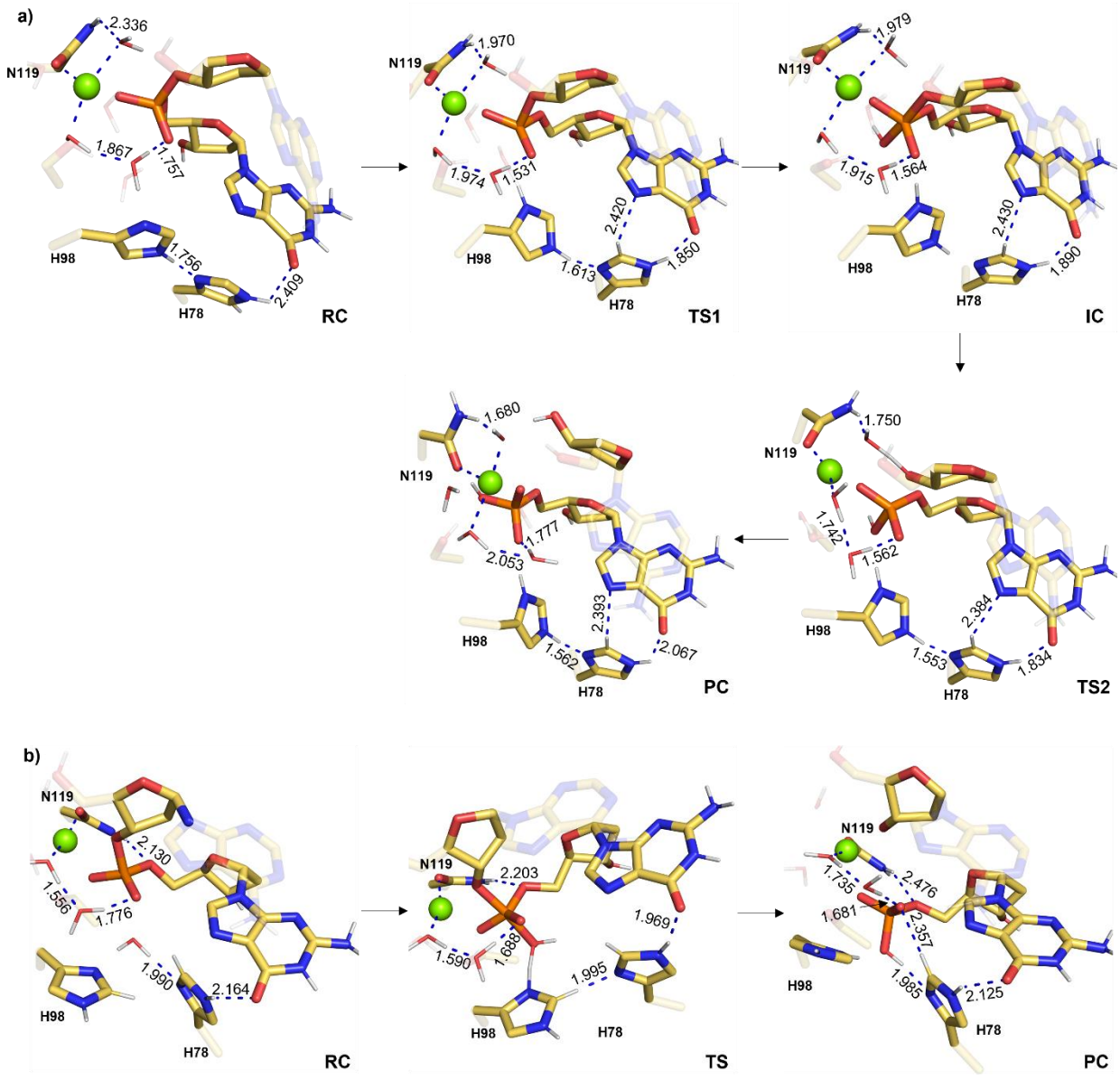

Figure C.8. Key distances (Å) in the I-*Ppo*I active site for the phosphodiester bond cleavage pathway characterized using QM cluster Model 2 with a) indirect or b) direct Mg^{2+} coordination to the leaving group. See main text Figures 4.5b and 4.6 for key reaction parameters for QM cluster Model 2.

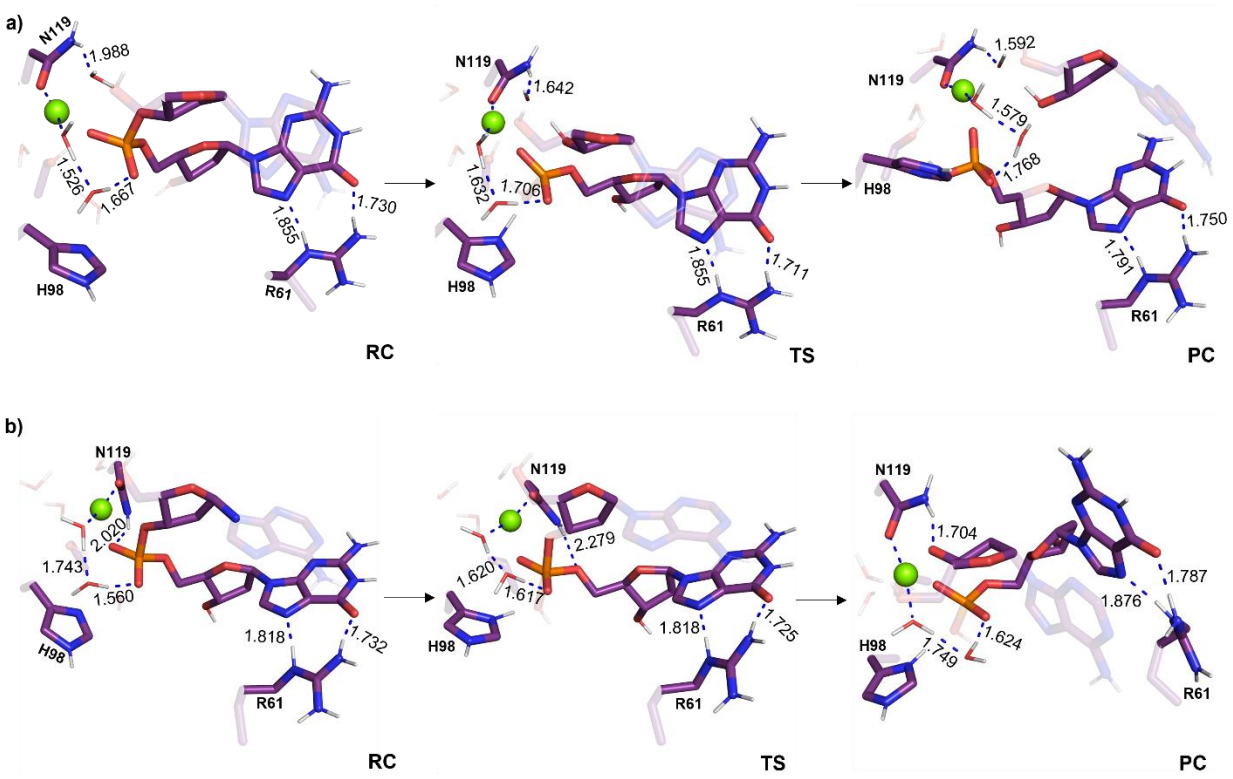

Figure C.9. Key distances (Å) in the I-*Ppo*I active site for the phosphodiester bond cleavage pathway characterized using QM cluster Model 3 with a) indirect or b) direct Mg^{2+} coordination to the leaving group. See main text Figures 4.3b and 4.5b for key reaction parameters for QM cluster Model 3.

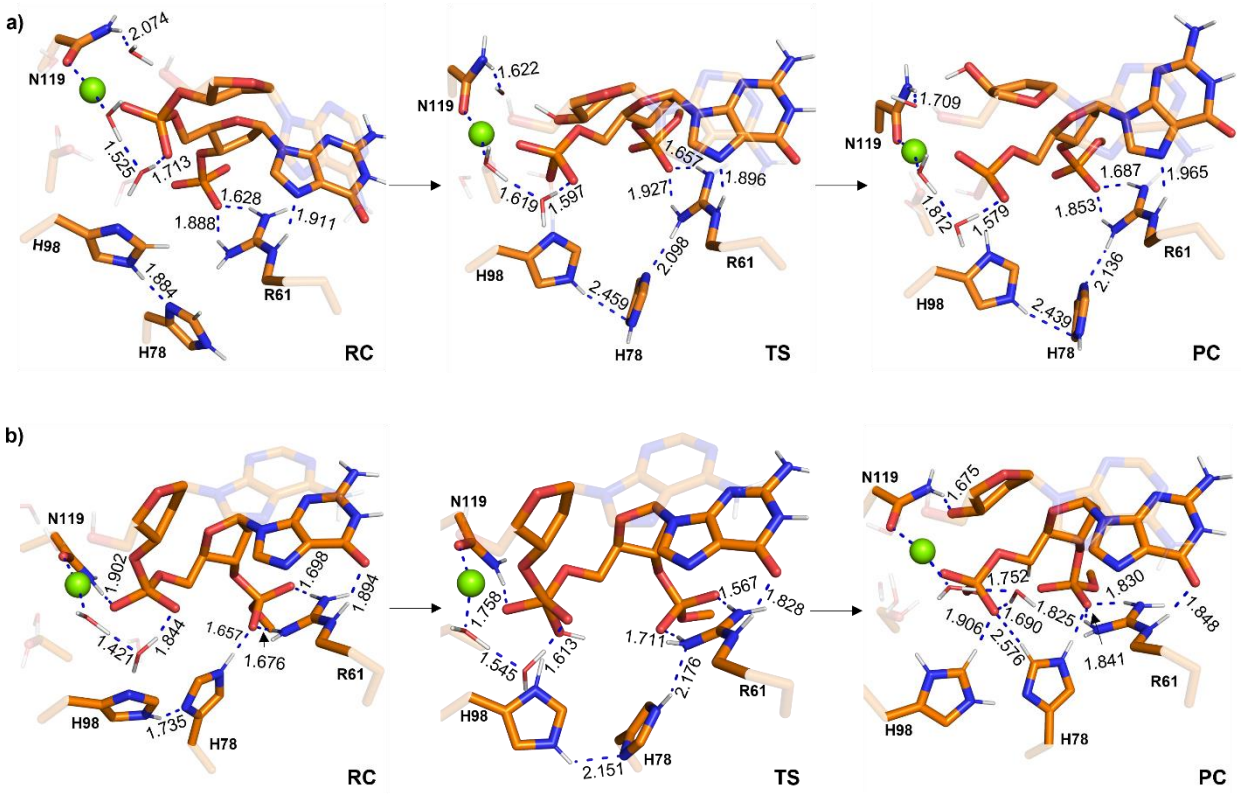

Figure C.10. Key distances (Å) in the I-*Ppo*I active site for the phosphodiester bond cleavage pathway characterized using QM cluster Model 4 with a) indirect or b) direct Mg^{2+} coordination to the leaving group. See main text Figures 4.3b and 4.5b for key reaction parameters for QM cluster Model 4.

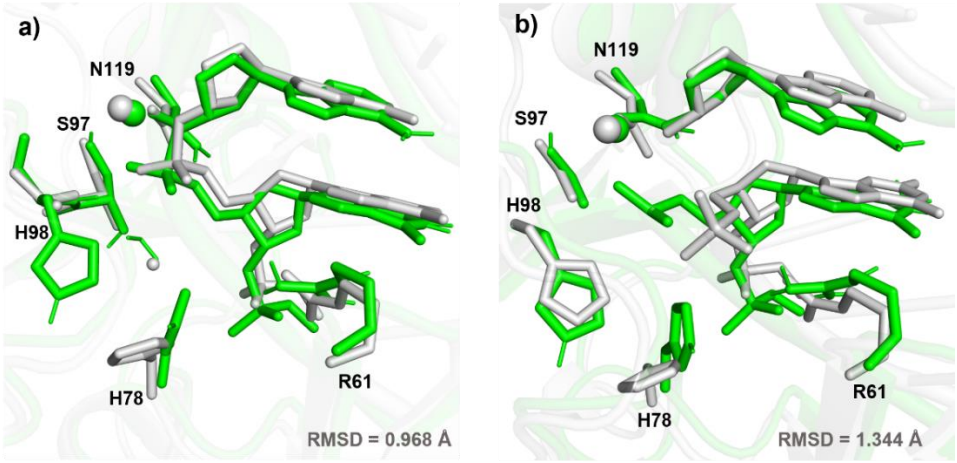

Figure C.11. Overlays of the I-*Ppo*I active site comparing a) the RC from the preferred QM/MM wild-type I-*Ppo*I model (green) and the X-ray crystal structure of the H98A mutant (grey, PDB ID: 1CYQ) and b) the PC from the preferred QM/MM wild-type I-*Ppo*I model (green) and the X-ray crystal structure of wild-type I-*Ppo*I (grey, PDB ID: 1A73).

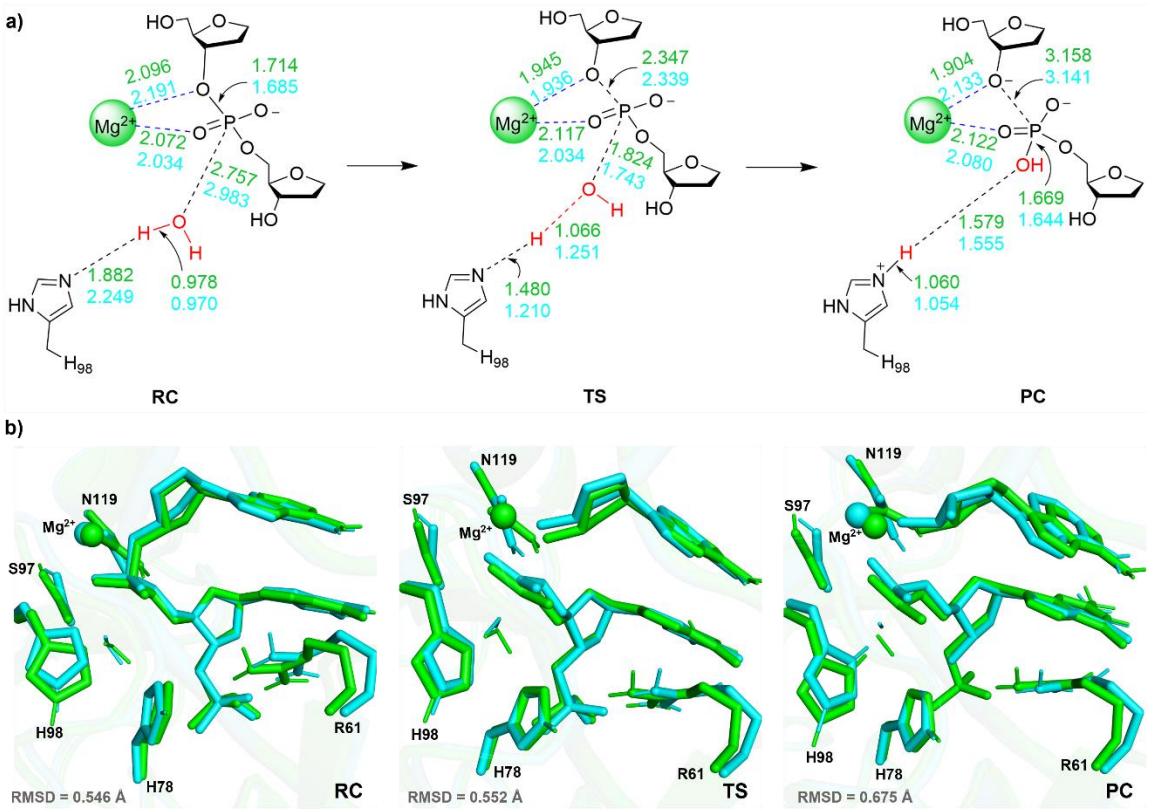

Figure C.12. a) Key reaction parameters (Å) for each stationary point and b) overlays of the active site for the I-*Ppo*I catalyzed phosphodiester bond cleavage involving direct Mg^{2+} coordination to the leaving group characterized in the present work using QM/MM with mechanical (green) or electrostatic (cyan) embedding.

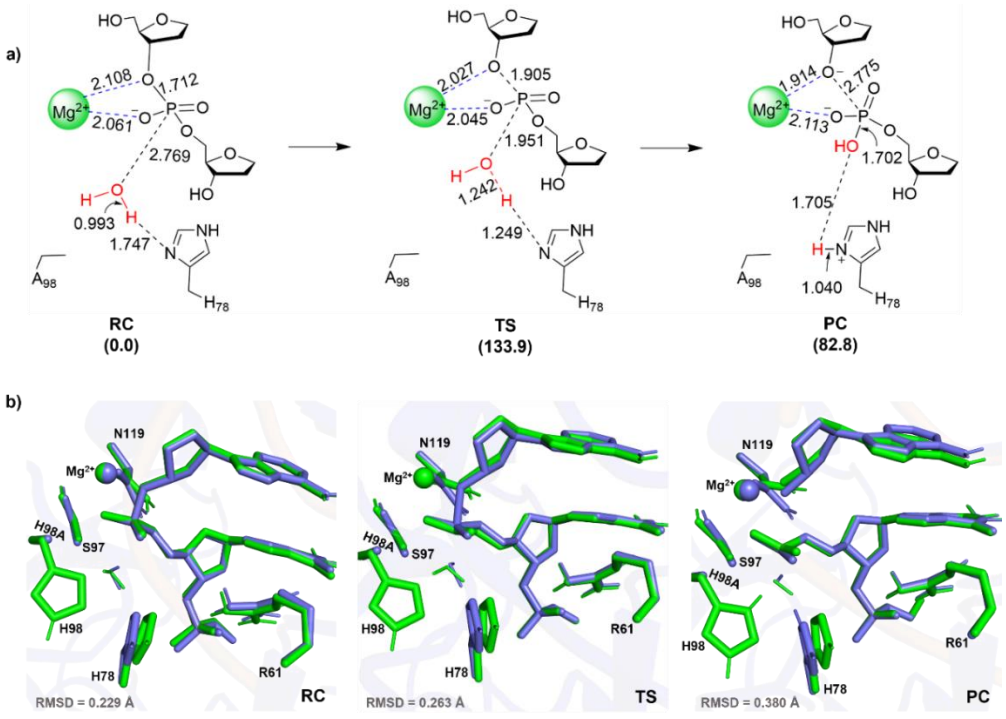

Figure C.13. a) Key reaction parameters (Å) for each stationary point and relative Gibbs energies (kJ/mol, in parantheses) of the phosphodiester bond cleavage by the H98A I-*Ppo*I mutant involving direct Mg^{2+} coordination to the leaving group characterized in the present work using QM/MM and b) the corresponding overlays of the active sites of wild-type (green) and H98A I-*Ppo*I mutant (blue).

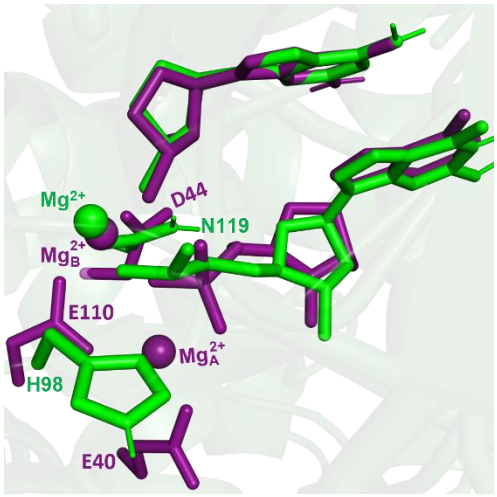

Figure C.14. Overlay of the active sites of two-metal mediated AaRNase III (purple, PDB ID: 2NUG) and the preferred QM/MM PC for wild-type I-*Ppo*I (green), higlighting the similar location of Mg^{2+} with respect to the substrate

Reference

(1) Galburt, E. A.; Chevalier, B.; Tang, W.; Jurica, M. S.; Flick, K. E.; Monnat, R. J.; Stoddard, B. L., A novel endonuclease mechanism directly visualized for I-*Ppo*I. *Nat. Struct. Biol.* **1999,** *6* (12), 1096–1099.

Appendix D

Supplementary Information for Chapter 5: Is Metal Stabilization of the Leaving Group Required or Can Lysine Facilitate Phosphodiester Bond Cleavage in Nucleic Acids? A Computational Study of EndoV

Contains Tables D.1–D.2 and Figures D.1–D.21

Table D.1. Average RMSDs (in Å) with standard deviations (in parentheses) for each 500 ns MD simulation replica.[a]

Direct metal-substrate coordination				
	Replica 1	Replica 2	Replica 3	Replica Average
Backbone of EndoV-DNA complex	1.7 (0.2)	1.6 (0.2)	1.9 (0.5)	1.7 (0.3)
Backbone of EndoV	1.2 (0.2)	1.3 (0.1)	1.1 (0.1)	1.2 (0.2)
Backbone for DNA substrate	2.9 (1.1)	2.9 (0.6)	3.0 (0.9)	2.9 (0.9)
All heavy atom RMSD of active site[b]	1.5 (0.3)	1.6 (0.3)	1.6 (0.4)	1.6 (0.3)
Indirect metal-substrate coordination				
	Replica 1	Replica 2	Replica 3	Replica Average
Backbone RMSD for enzyme-DNA complex	1.6 (0.2)	1.9 (0.4)	1.8 (0.2)	1.8 (0.3)
Backbone RMSD for enzyme	1.2 (0.2)	1.9 (0.4)	1.4 (0.2)	1.5 (0.4)
Backbone RMSD for DNA substrate	3.1 (0.4)	3.3 (1.0)	3.3 (0.7)	3.1 (0.4)
All heavy atom RMSD for the active site[b]	1.9 (0.3)	2.1 (0.6)	1.6 (0.4)	2.0 (0.4)

[a]RMSDs were measured with respect to the first frame of the corresponding MD trajectory. [b]Active site residues D43, E89, D110, K139, H214, Mg, dG229, dA230, and dG231 were included in the analysis.

Table D.2. Average metal-coordination distances (in Å) with standard deviations (in parentheses) and % occupancy across each 500 ns MD simulation replica.[a,b]

Direct metal-substrate coordination								
	Replica 1		Replica 2		Replica 3		Replica Average	
	Average (std. dev.)	% Occ.	Average (std. dev.)	% Occ.	Average (std. dev.)	% Occ.	Average (std. dev.)	% Occ.
Mg^{2+}···D43	2.0 (0.1)	100	2.0 (0.1)	100	2.0 (0.05)	100	2.0 (0.05)	100
Mg^{2+}···D110	1.9 (0.05)	100	2.0 (0.05)	100	2.0 (0.05)	100	1.9 (0.05)	100
Mg^{2+}···OP1	2.0 (0.05)	100	2.0 (0.05)	100	2.0 (0.05)	100	2.0 (0.05)	100
Mg^{2+}···W1	2.1 (0.1)	100	2.1 (0.1)	100	2.1 (0.1)	100	2.1 (0.1)	100
Mg^{2+}···W2	2.1 (0.1)	97	2.1 (0.1)	98	2.1 (0.1)	98	2.1 (0.1)	98
Mg^{2+}···W3	2.1 (0.1)	98	2.1 (0.1)	99	2.1 (0.1)	99	2.1 (0.1)	99
Indirect metal-substrate coordination								
	Replica 1		Replica 2		Replica 3		Replica Average	
	Average (std. dev.)	% Occ.	Average (std. dev.)	% Occ.	Average (std. dev.)	% Occ.	Average (std. dev.)	% Occ.
Mg^{2+}···D43	2.0 (0.05)	100	2.0 (0.05)	100	2.0 (0.05)	100	2.0 (0.05)	100
Mg^{2+}···D110	1.9 (0.05)	100	2.0 (0.05)	100	2.0 (0.05)	100	1.9 (0.05)	100
Mg^{2+}···W1	2.1 (0.1)	100	2.1 (0.1)	100	2.1 (0.1)	99	2.1 (0.1)	99
Mg^{2+}···W2	2.1 (0.1)	100	2.1 (0.1)	99	2.1 (0.1)	99	2.1 (0.1)	99
Mg^{2+}···W3	2.1 (0.1)	100	2.1 (0.1)	99	2.1 (0.1)	100	2.1 (0.1)	100
Mg^{2+}···W4	2.1 (0.1)	100	2.1 (0.1)	99	2.1 (0.1)	100	2.1 (0.1)	99
W1···OP1	2.7 (0.05)	69	2.8 (0.1)	77	2.8 (0.05)	69	2.7 (0.05)	72

[a]Refer to Figure D.11 for a schematic of the metal-coordination distances. [b]Occ. refers to occupancy.

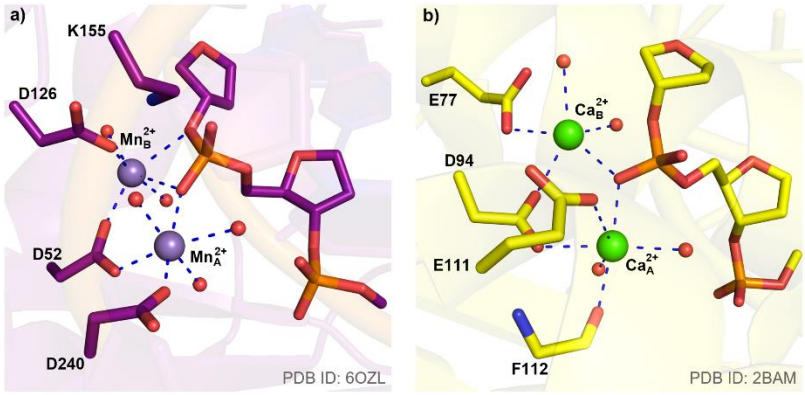

Figure D.1. Active sites from X-ray crystal structures of the a) Mn^{2+}-containing RC of wild-type mouse EndoV bound to a dsRNA substrate and b) Ca^{2+}-containing RC of wild-type BamHI bound to a dsDNA substrate.

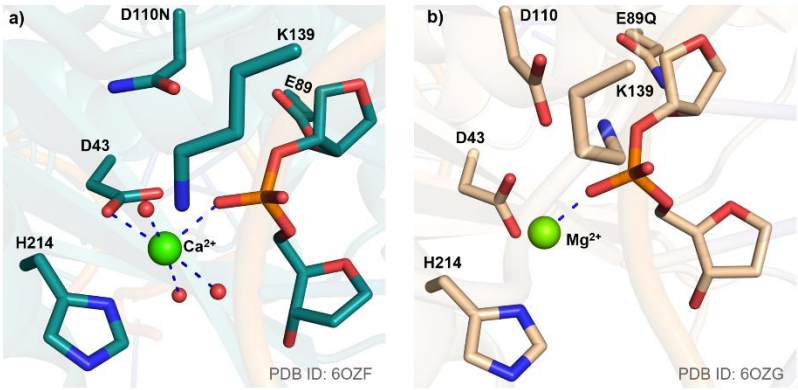

Figure D.2. Active sites from X-ray crystal structures of the RC corresponding to a) the Ca^{2+}-containing D110N mutant and b) the Mg^{2+}-containing E89Q mutant of *Tma* EndoV bound to a ssDNA substrate.

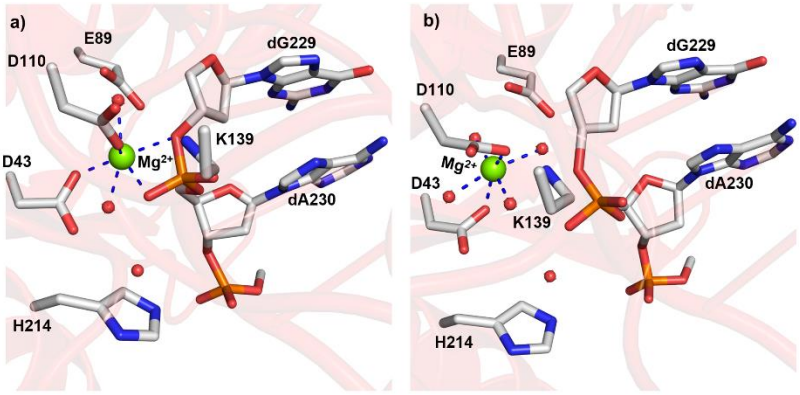

Figure D.3. Starting models for MD minimization built by modifying the crystal structure of the wild-type EndoV PC (PDB ID: 2W35) involving a) bidentate coordination or b) indirect (water mediated) coordination of the metal to the substrate.

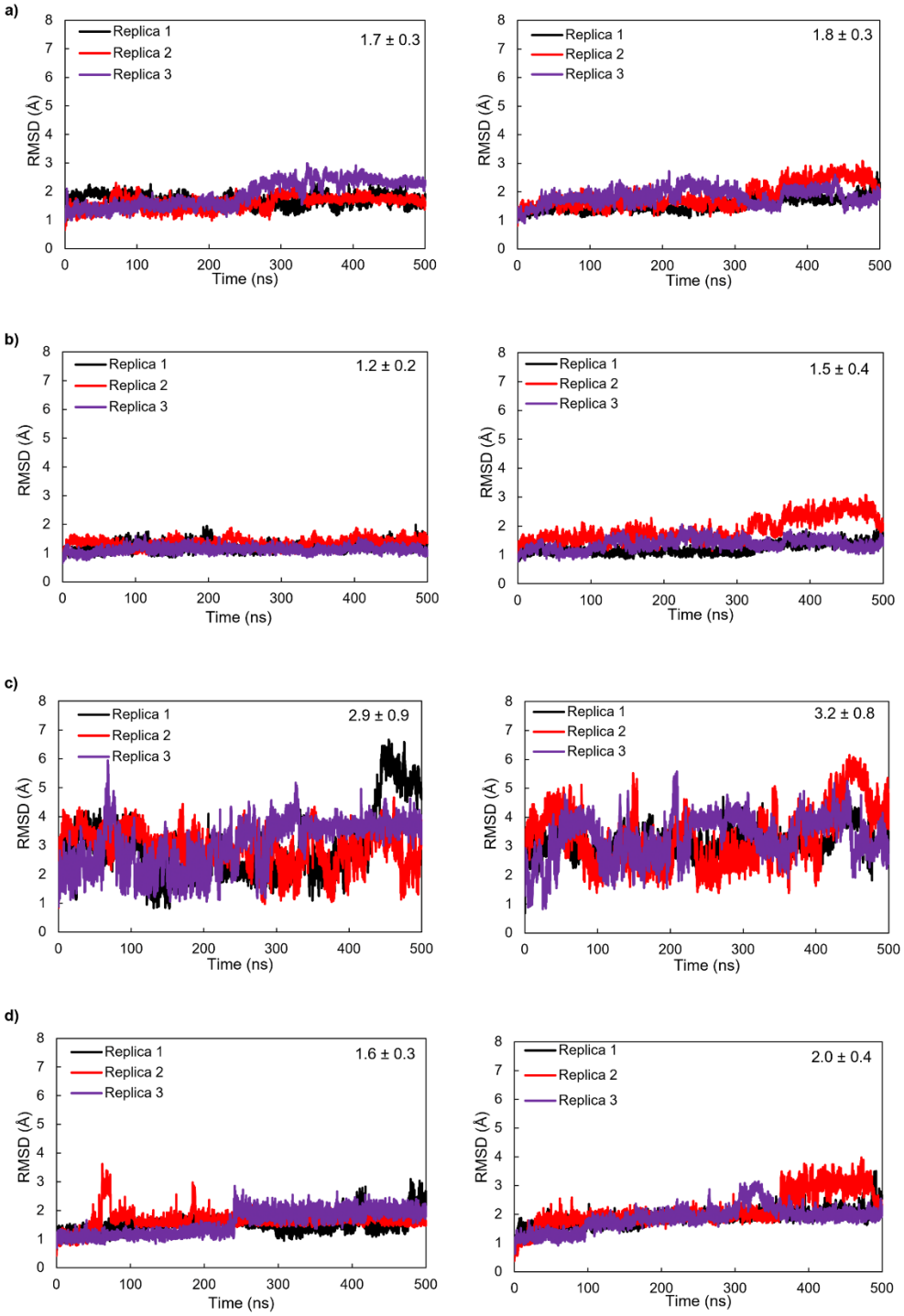

Figure D.4. RMSD (Å) as a function of time for each 500 ns MD simulation replica for the RCs involving direct (left) or indirect (right) metal–substrate coordination with respect to a) the backbone of the entire EndoV–DNA complex, b) the backbone of only the enzyme, c) the backbone of only the DNA substrate, and d) all heavy atoms of the active site residues (D43, E89, D110, K139, H214, Mg^{2+}, dG229, dA230, and dG231). Average RMSD and standard deviation for the corresponding data set is provided in the top right corner.

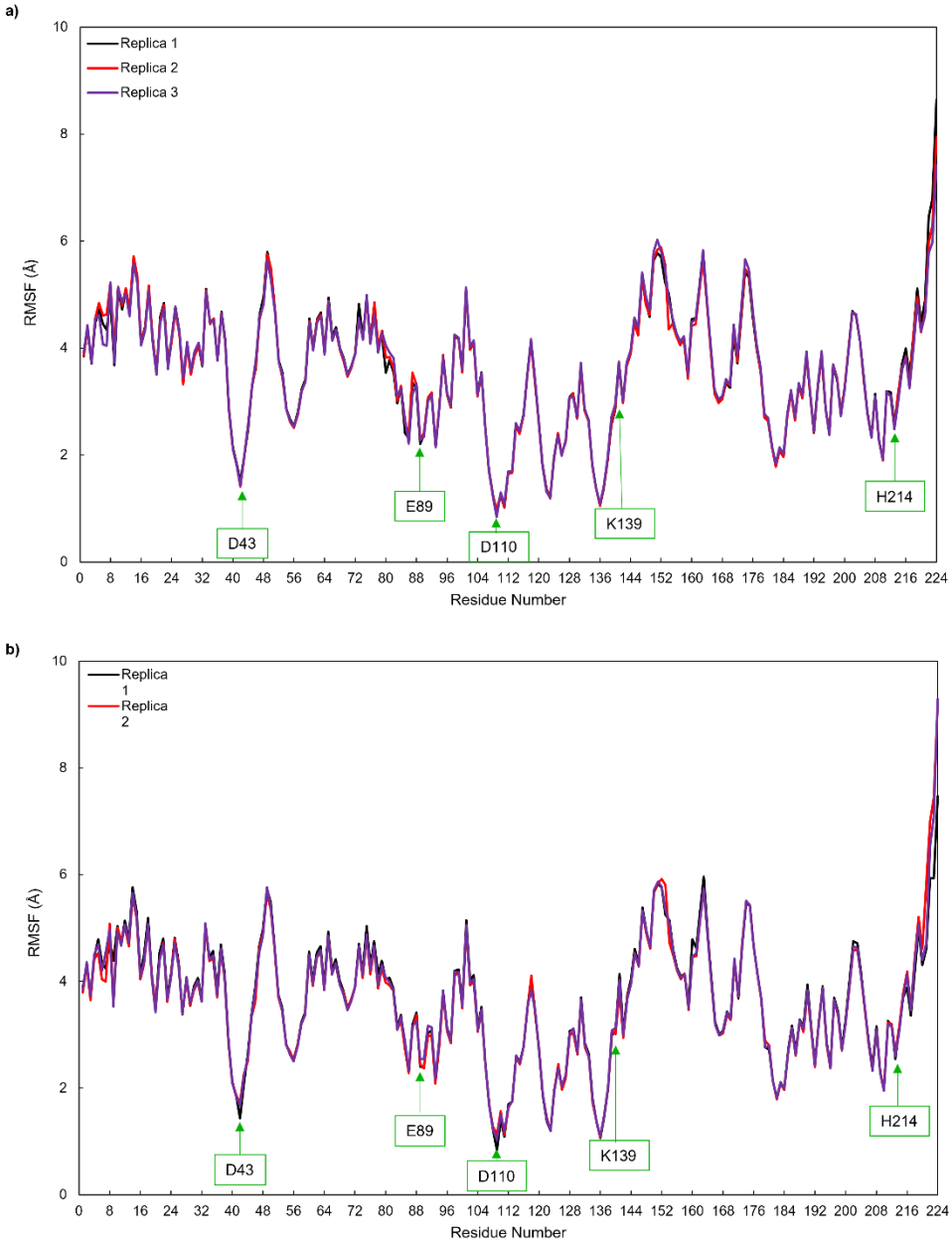

Figure D.5. RMSF (Å) of EndoV amino acid residues for each 500 ns MD simulation replica of the RC involving a) direct or b) indirect metal–substrate coordination. Key active site residues are highlighted using green boxes.

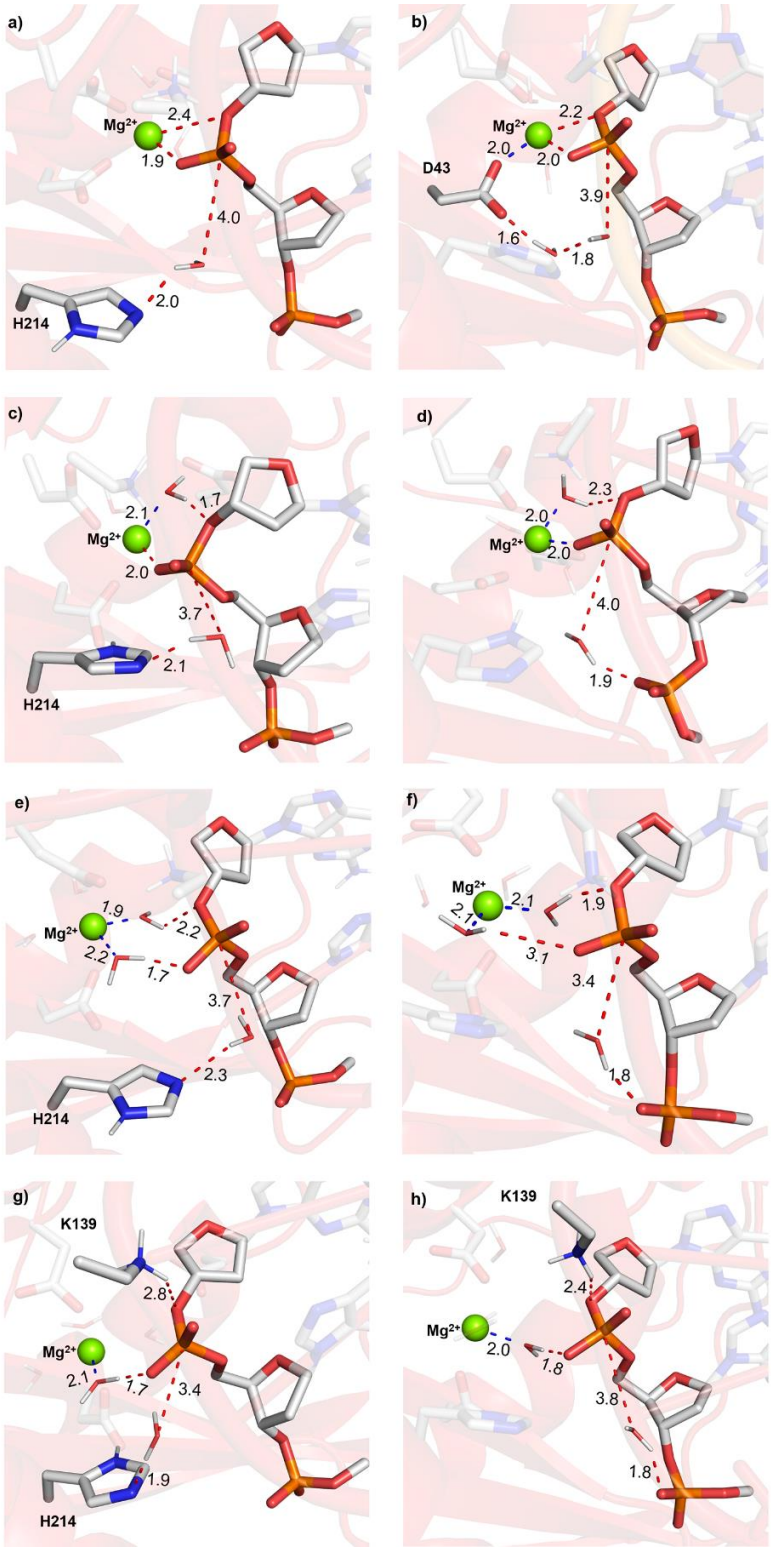

Figure D.6. Representative MD snapshots of the EndoV active site aligned for the phosphodiester bond cleavage reaction, with each panel highlighting the metal coordination to the substrate along with the residues that are aligned to activate the water nucleophile and stabilize the leaving group.

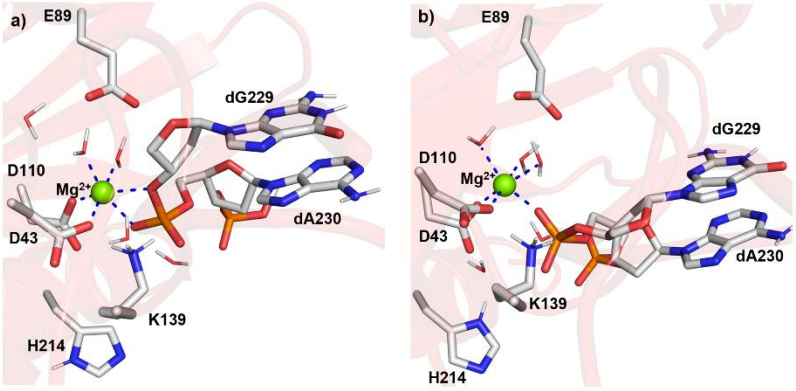

Figure D.7. The QM region (129 atoms) in ONIOM models of RC that contains direct a) bidentate or b) monodentate metal–substrate coordination.

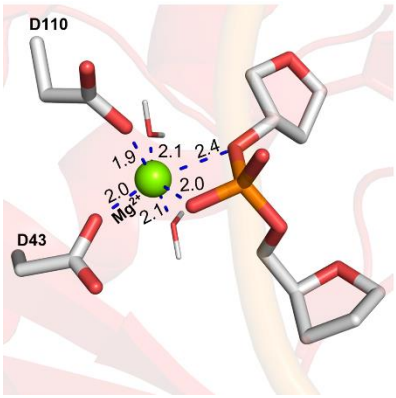

Figure D.8. Metal coordination distances (Å) in the RC from MD equilibrated structure involving bidentate direct metal–substrate coordination.

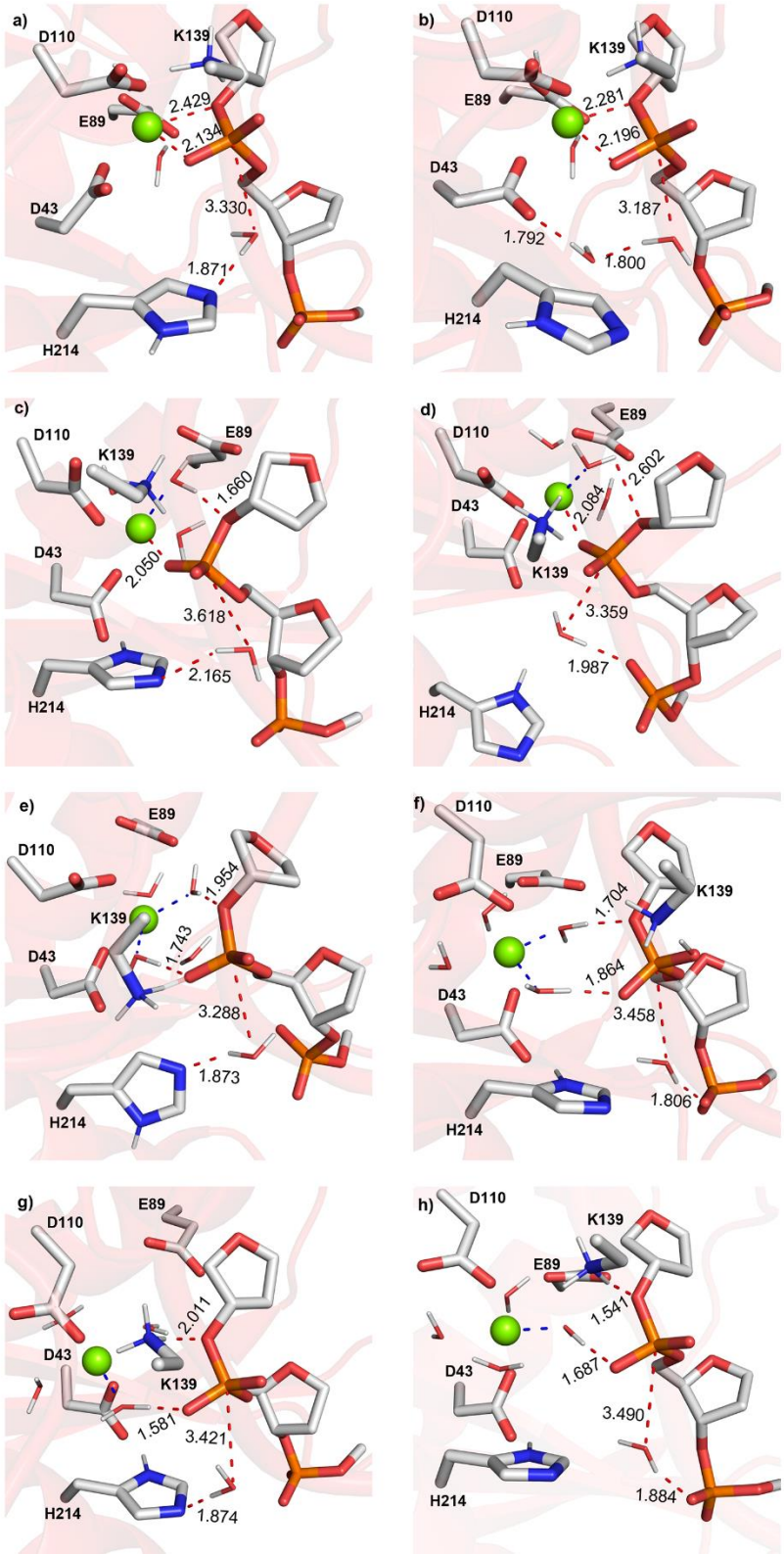

Figure D.9. Structures of the ONIOM(B3LYP-D3(BJ)/6-31G(d,p):AMBERff14SB) optimized RCs obtained from the corresponding representative MD snapshots (Figure D.5, Appendix D) with different metal coordination architectures, general bases, and general acids.

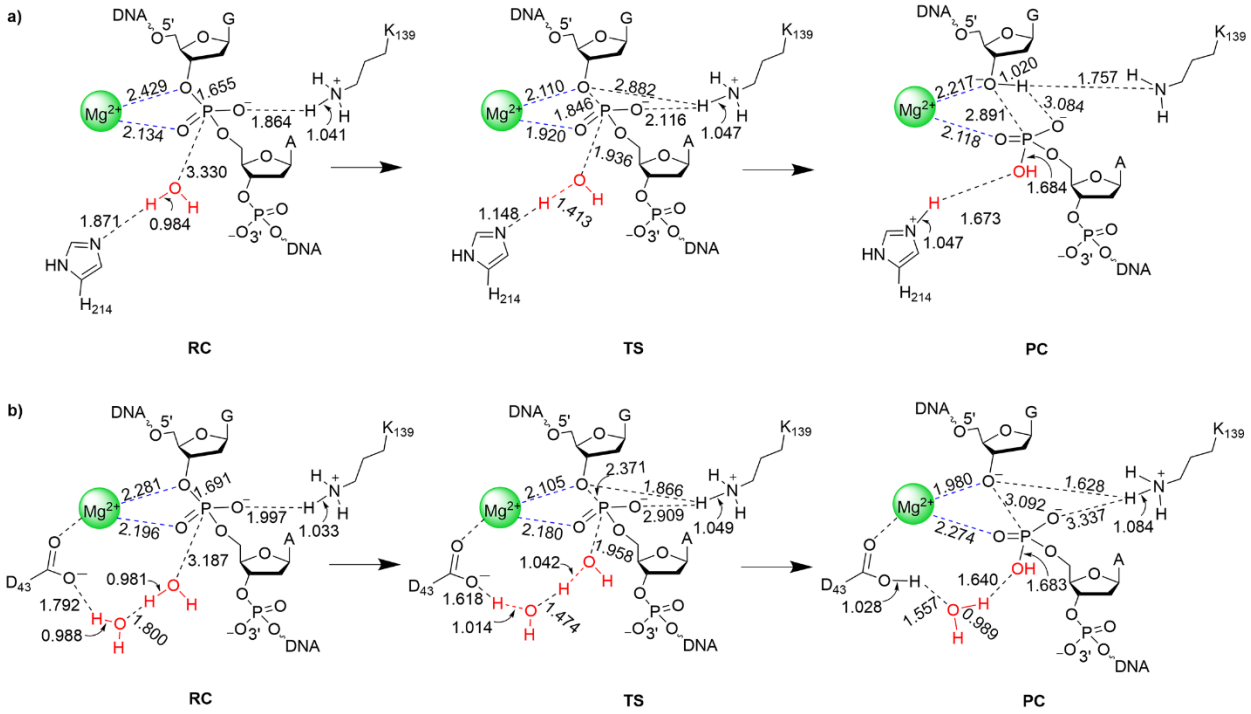

Figure D.10. Mechanism and key ONIOM(B3LYP-D3(BJ)/6-31G(d,p):AMBERff14SB) bond distances (Å) for the EndoV-catalyzed phosphodiester bond cleavage involving bidentate direct metal coordination to the substrate, Mg^{2+} aiding leaving group departure, and a) H214 or b) D43 via a water chain activating the water nucleophile.

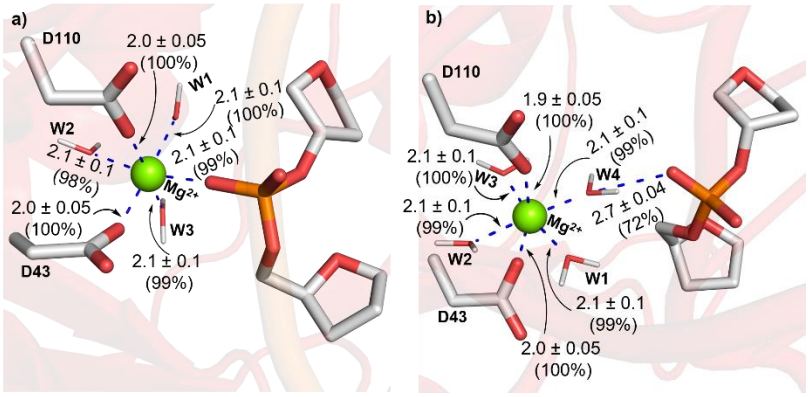

Figure D.11. Average distances with standard deviations (in Å) and occupancies (%, in parentheses) for Mg^{2+} coordination across all MD replicas for the RC involving a) direct or b) indirect metal–substrate coordination.

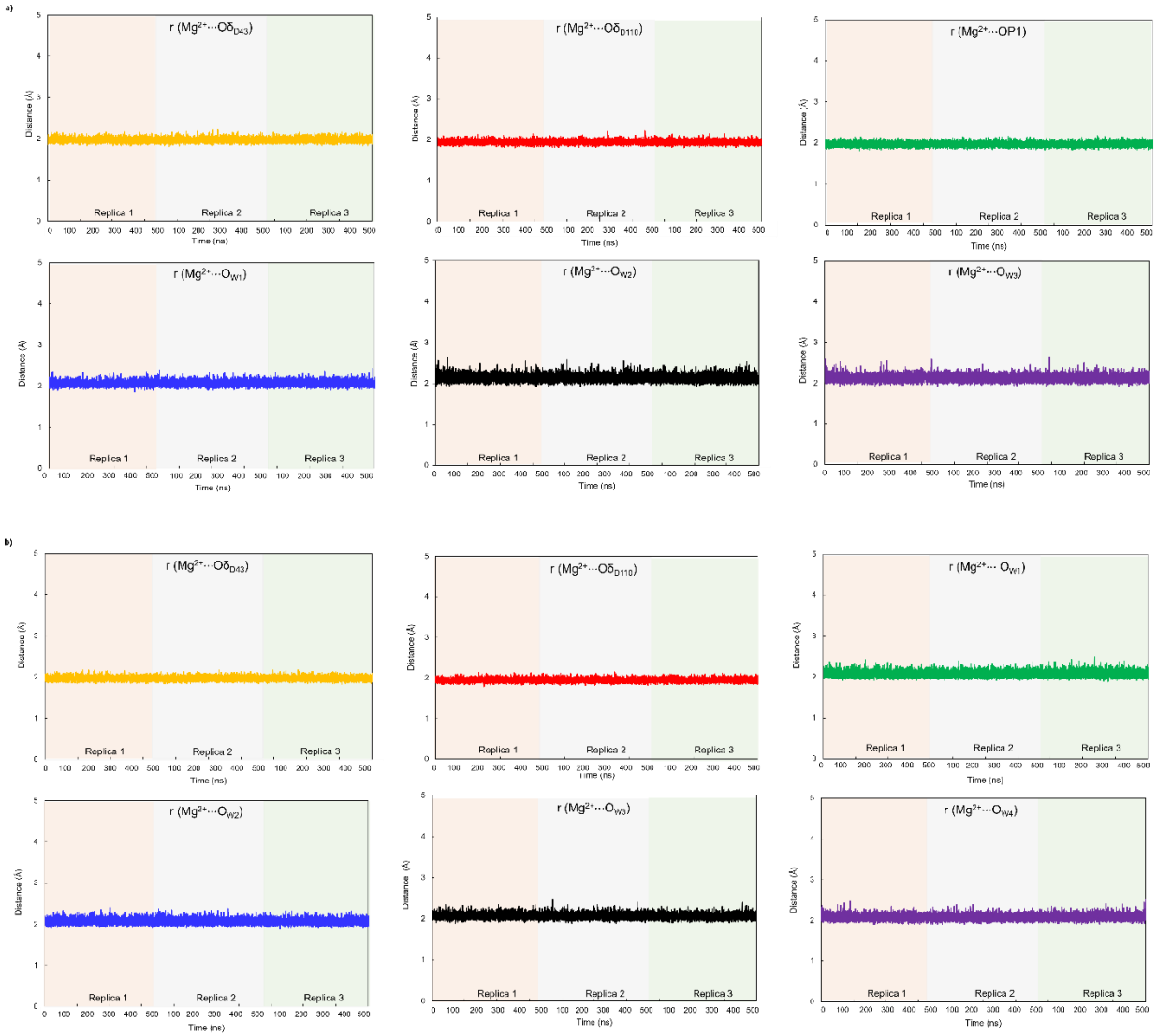

Figure D.12. Mg^{2+}–coordination distances (Å) as a function of time for each 500 ns MD simulation replica of the RCs involving a) direct or b) indirect metal–substrate coordination.

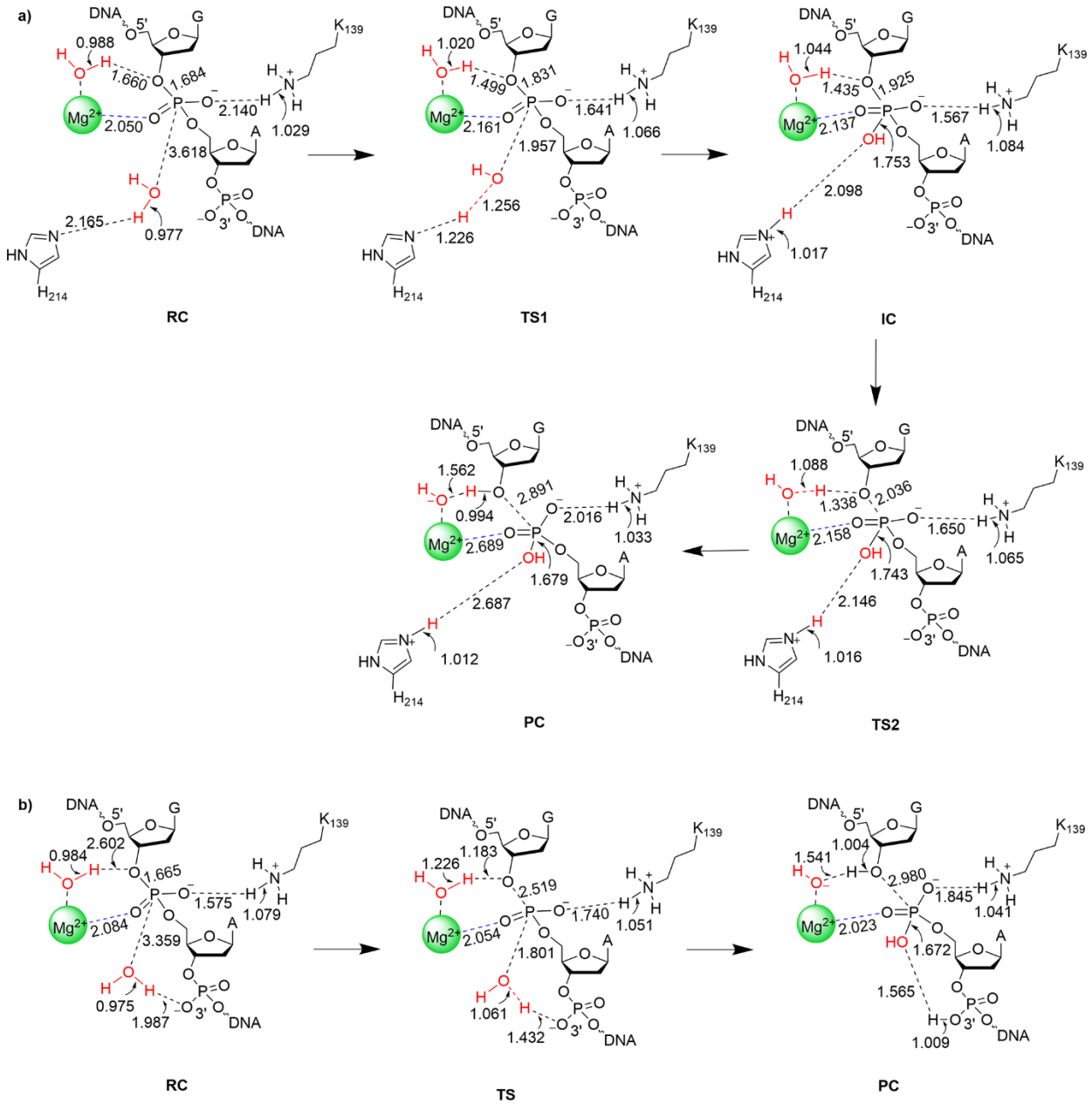

Figure D.13. Mechanism and key ONIOM(B3LYP-D3(BJ)/6-31G(d,p):AMBERff14SB) bond distances (Å) for the EndoV-catalyzed phosphodiester bond cleavage involving direct metal coordination to the nonbridging phosphate oxygen, leaving group stabilization by a Mg^{2+}-ligated water, and a) H214 or b) a substrate phosphate activating the water nucleophile.

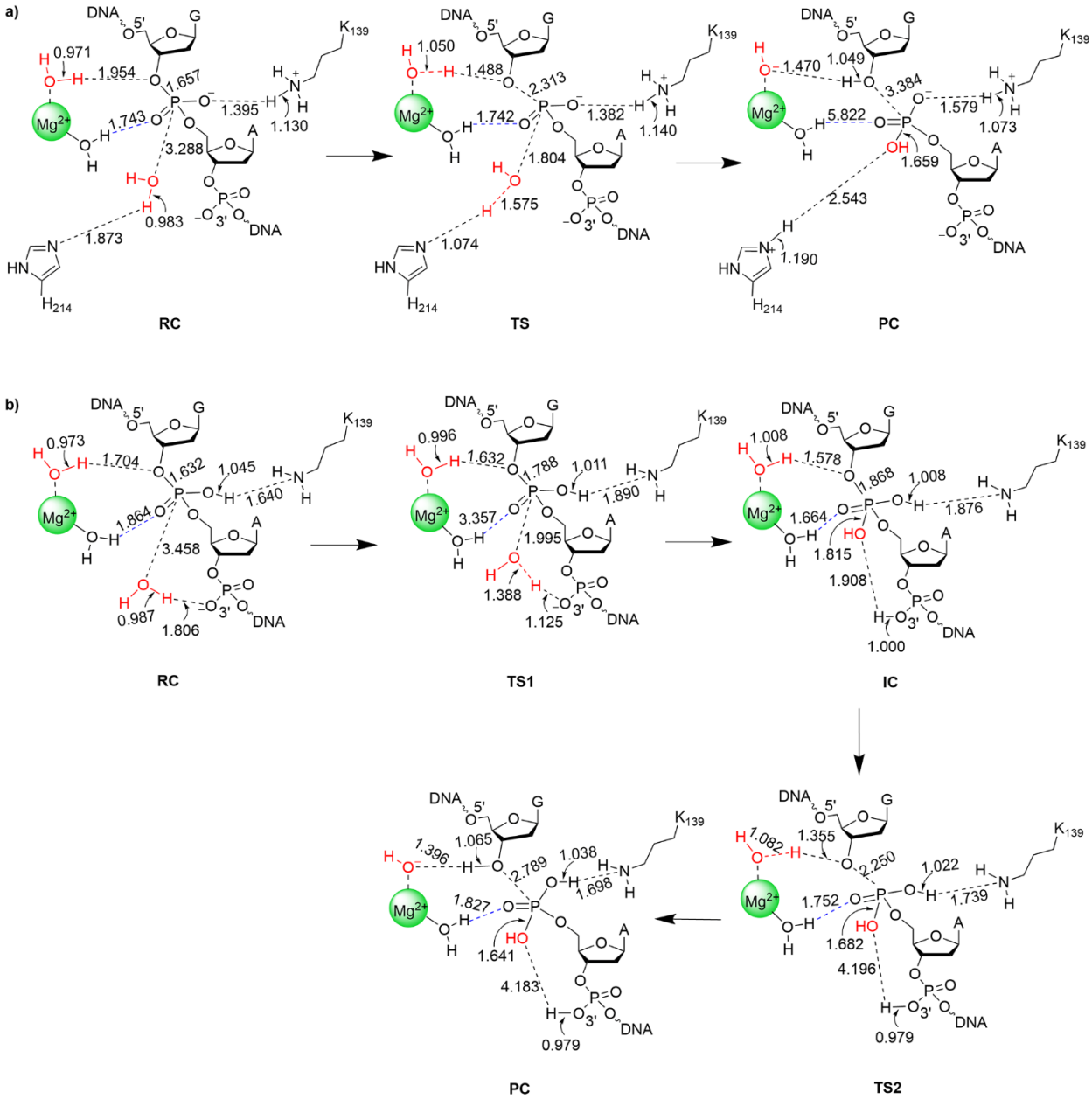

Figure D.14. Mechanism and key ONIOM(B3LYP-D3(BJ)/6-31G(d,p):AMBERff14SB) bond distances (Å) for the EndoV-catalyzed phosphodiester bond cleavage involving indirect metal coordination to the nonbridging phosphate oxygen, leaving group stabilization by a Mg^{2+}-ligated water, and a) H214 or b) a substrate phosphate activating the water nucleophile.

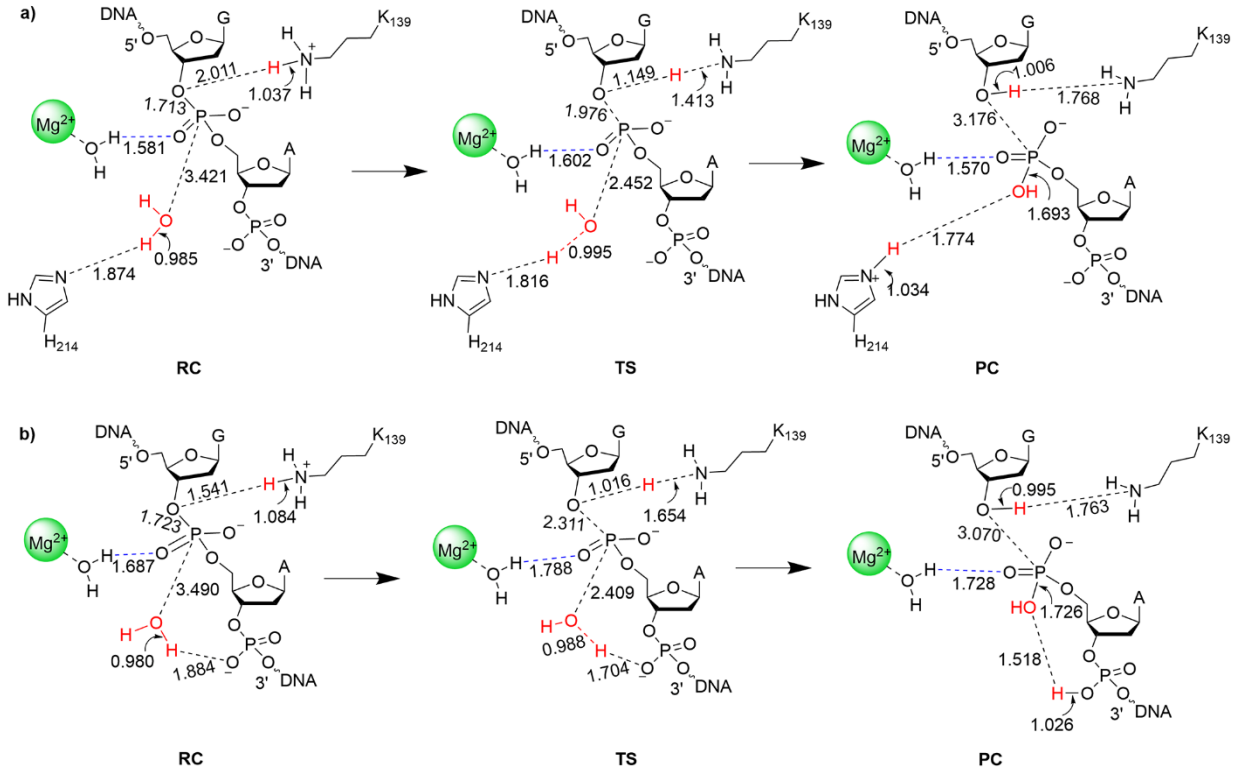

Figure D.15. Mechanism and key ONIOM(B3LYP-D3(BJ)/6-31G(d,p):AMBERff14SB) bond distances (Å) for the EndoV-catalyzed phosphodiester bond cleavage involving indirect metal coordination to the nonbridging phosphate oxygen, leaving group stabilization by K139, and a) H214 or b) a substrate phosphate activating the water nucleophile.

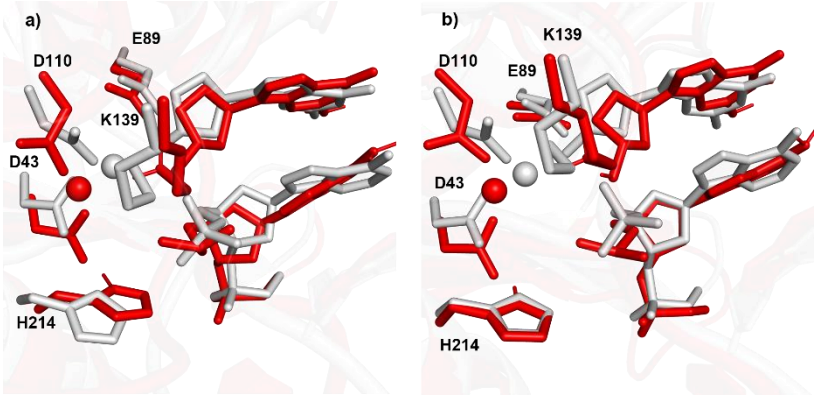

Figure D.16. Overlay of the active sites of the wild-type EndoV PC from an X-ray crystal structure (grey, PDB ID: 2W35) and QM/MM optimized PC (red) involving indirect metal coordination to the nonbridging phosphate oxygen, leaving group stabilization by K139, and a) H214 or b) a substrate phosphate activating the water nucleophile.

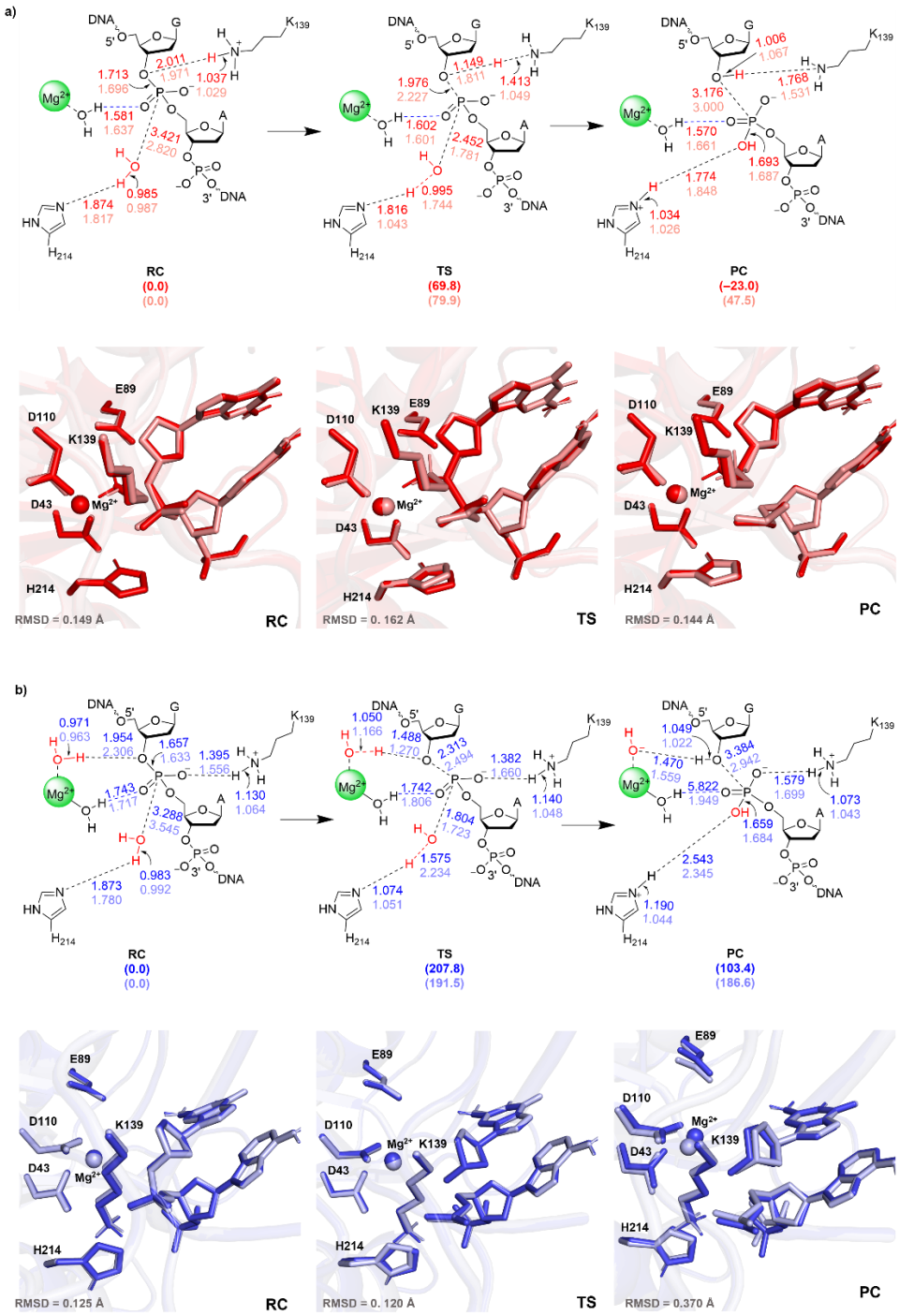

Figure D.17. Comparison of ONIOM(B3LYP-D3(BJ)/6-31G(d,p):AMBERff14SB) data obtained using ME (dark shades) or EE (light shades). Bond distances (top) provided in Å and overlays of the active site (bottom) for the mechanism involving indirect metal coordination to the nonbridging phosphate oxygen, H214 activating the water nucleophile, and leaving group stabilization by a) K139 (red) or b) a Mg^{2+}-ligated water (blue). Relative Gibbs energies (kJ/mol) are provided in parentheses.

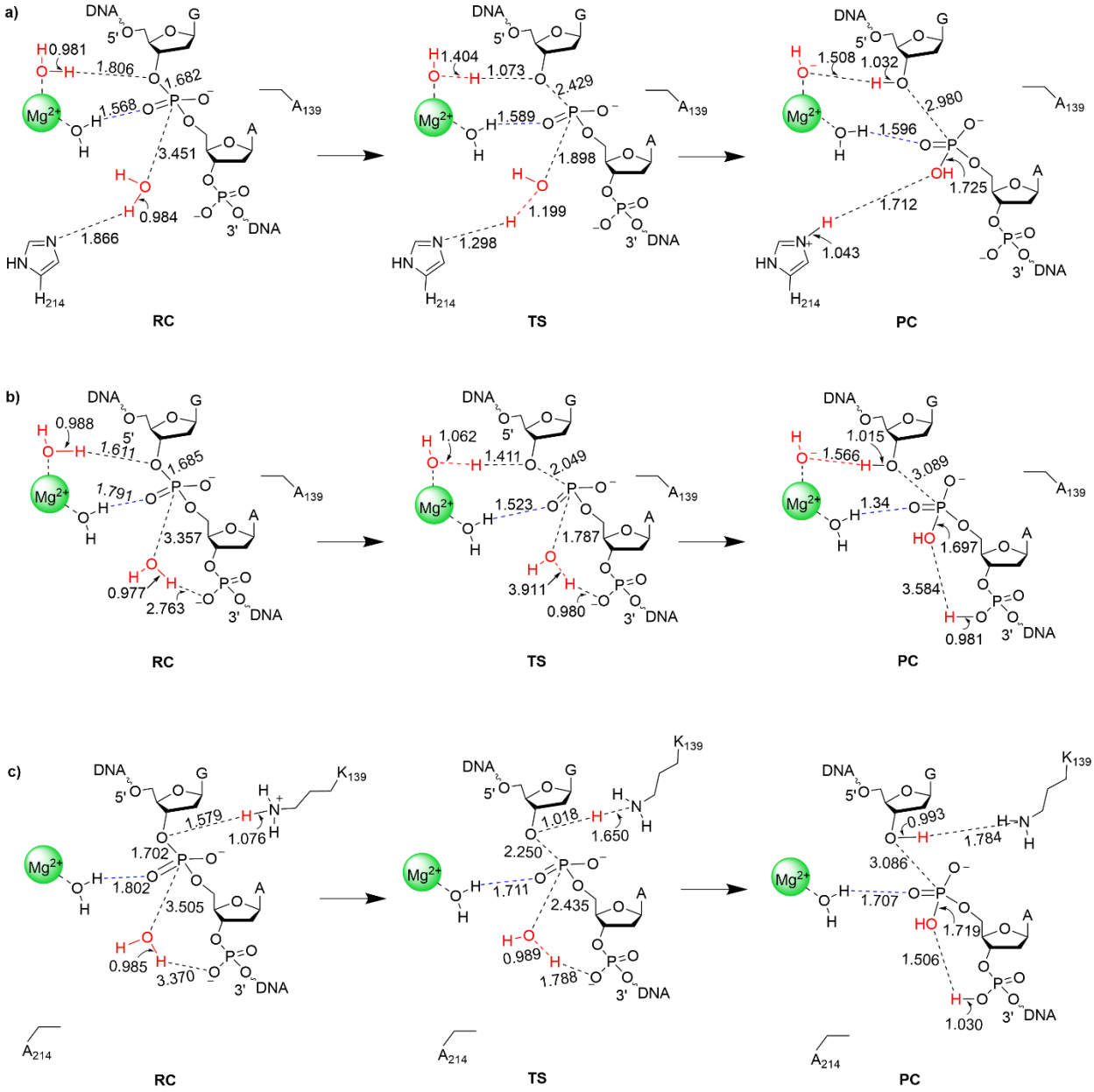

Figure D.18. Mechanism and ONIOM(B3LYP-D3(BJ)/6-31G(d,p):AMBERff14SB) key bond distances (Å) for the phosphodiester bond cleavage by the K139A mutant EndoV involving leaving group stabilization by a Mg^{2+}-ligated water, and a) H214 or b) a substrate phosphate activating the water nucleophile, and c) H214A EndoV mutant involving leaving group stabilization by K139 and a substrate phosphate activating the water nucleophile.

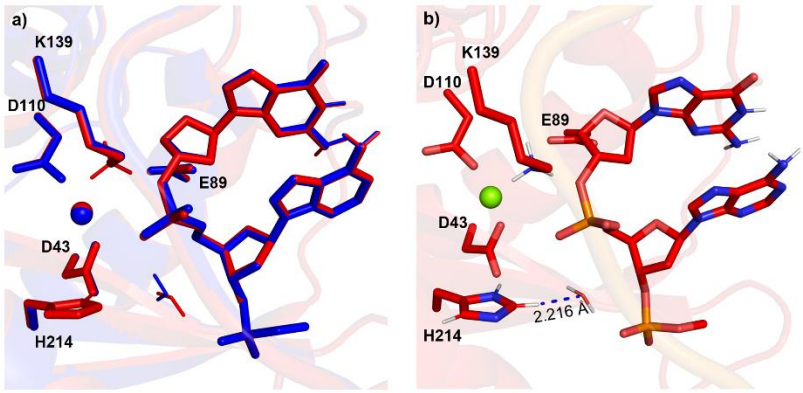

Figure D.19. a) Overlay of the QM/MM optimized RC from the wild-type (red) and H214A EndoV mutant (blue) involving leaving group stabilization by K139 and a substrate phosphate activating the water nucleophile (RMSD = 0.086 Å) and b) QM/MM optimized RC from the wild-type EndoV, highlighting the C-H···O hydrogen bond between H214 and the water nucleophile.

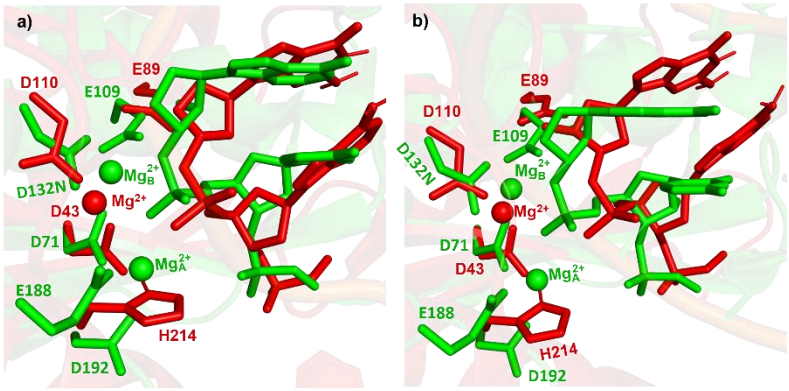

Figure D.20. Overlay of the active sites of two-metal mediated RNase H (green, PDB ID: 1ZBI) and QM/MM optimized RC for wild-type EndoV (red) involving indirect metal ligation to the substrate, K139 stabilizing the leaving group, and a) H214 or b) a substrate phosphate as the general base.

Figure D.21. Sequence alignment of bacterial and mouse EndoV, with key active site amino acids highlighted. The species and the respective GenBank accession numbers used for the alignment are *Thermotoga maritima* and Q9X2H9 for bacterial EndoV, and *Mus musculus* and Q8C9A2 for mouse EndoV.

Appendix E

Supplementary Information for Chapter 6: Mechanism of Nucleic Acid Phosphodiester Bond Cleavage by Human Endonuclease V: MD and QM/MM Calculations Reveal a Versatile Metal Dependence

Contains Tables E.1–E.3 and Figures E.1–E.11

Table E.1. Average RMSDs (Å) across 500 ns MD simulation replicas (standard deviation in parentheses) for the RC involving one metal with indirect (water-mediated) or direct coordination to the non-bridging phosphate oxygen of the scissile phosphate, or two metals.[a]

1 Metal (Indirect Metal–Substrate Coordination)				
	Replica 1	Replica 2	Replica 3	Average
Backbone of EndoV-RNA complex	2.0 (0.2)	2.0 (0.2)	2.1 (0.2)	2.1 (0.2)
Backbone of EndoV	1.5 (0.1)	1.4 (0.2)	1.4 (0.2)	1.4 (0.2)
Backbone of RNA substrate	2.8 (0.4)	3.5 (0.6)	3.6 (0.6)	3.3 (0.5)
All heavy atoms of active site[b]	1.9 (0.2)	2.3 (0.6)	2.1 (0.4)	2.1 (0.4)
1 Metal (Direct Metal–Substrate Coordination)				
	Replica 1	Replica 2	Replica 3	Average
Backbone of EndoV-RNA complex	1.8 (0.2)	2.1 (0.3)	1.7 (0.2)	1.8 (0.2)
Backbone of EndoV	1.5 (0.2)	1.7 (0.2)	1.3 (0.2)	1.5 (0.2)
Backbone of RNA substrate	2.5 (0.4)	2.8 (0.6)	2.6 (0.5)	2.6 (0.5)
All heavy atoms of active site[b]	1.2 (0.1)	1.2 (0.1)	1.4 (0.4)	1.3 (0.2)
2 Metals				
	Replica 1	Replica 2	Replica 3	Average
Backbone of EndoV-RNA complex	1.7 (0.1)	1.8 (0.2)	1.6 (0.2)	1.7 (0.2)
Backbone of EndoV	1.2 (0.1)	1.4 (0.2)	1.1 (0.1)	1.3 (0.1)
Backbone of RNA substrate	2.4 (0.2)	2.3 (0.3)	2.4 (0.5)	2.4 (0.4)
All heavy atoms of active site[b]	1.1 (0.1)	1.0 (0.1)	1.2 (0.1)	1.1 (0.1)

[a]RMSDs were measured with respect to the first frame of the corresponding MD trajectory for each replica. [b]Active site residues D52, E100, D126, K155, D240, Mg^{2+} or Mg_A^{2+} and Mg_B^{2+}, A250, U251, and A252 were included in the analysis.

Table E.2. Percent occupancies for the reaction parameter-based criteria across the MD simulation replicas for the RC involving one metal with indirect (water-mediated) or direct coordination to the non-bridging oxygen of the scissile phosphate, or two metals.[a]

Reaction Parameter-Based Criteria	**1 Metal (Indirect Metal–Substrate Coordination)**			
	Replica 1	Replica 2	Replica 3	Average
$r(O\delta_{D240}\cdots P) < 7.5$ Å	57%	77%	81%	71%
$r(O_W\cdots P) < 4.0$ Å	100%	100%	100%	100%
$r(O3'\cdots O_W) < 3.5$ Å	27%	24%	24%	25%
	1 Metal (Direct Metal–Substrate Coordination)			
	Replica 1	Replica 2	Replica 3	Average
$r(O\delta_{D240}\cdots P) < 7.5$ Å	59%	52%	70%	60%
$r(O_W\cdots P) < 4.0$ Å	100%	100%	100%	100%
$r(O3'\cdots O_W) < 3.5$ Å	38%	43%	38%	40%
	2 Metals			
	Replica 1	Replica 2	Replica 3	Average
$r(O\delta_{D240}\cdots O_{\text{bridging water}}) < 3.5$ Å	55%	48%	60%	54%
$r(O_W\cdots O_{\text{bridging water}}) < 3.5$ Å	55%	48%	60%	54%
$r(O_W\cdots P) < 4.0$ Å	100%	100%	98%	99%
$r(O3'\cdots Mg_B^{2+}) < 2.5$ Å	0.02%	15%	0.3%	5%
$r(O3'\cdots O_W) < 3.5$ Å	98%	90%	66%	85%

[a]See main text Figures 6.4b, 6.5b, and 6.6b for the representative structures displaying the reaction parameters.

Table E.3. Average Mg^{2+} coordination distances with standard deviations (Å, in parentheses) and occupancies (Occ., %) across MD simulation replicas for the RC involving one metal with indirect (water-mediated) or direct coordination to the non-bridging oxygen of the scissile phosphate, or two metals.[a]

1 Metal (Indirect Metal–Substrate Coordination)								
	Replica 1		Replica 2		Replica 3		Average	
	Average (std. dev.)	% Occ.	Average (std. dev.)	% Occ.	Average (std. dev.)	% Occ.	Average (std. dev.)	% Occ.
Mg^{2+}…W1	2.1 (0.1)	100	2.1 (0.1)	98	2.1 (0.1)	93	2.1 (0.1)	97
Mg^{2+}…W2	2.1 (0.1)	100	2.1 (0.1)	99	2.1 (0.1)	100	2.1 (0.1)	100
Mg^{2+}…W3	2.1 (0.1)	100	2.1 (0.1)	100	2.1 (0.1)	100	2.1 (0.1)	100
Mg^{2+}…W4	2.1 (0.1)	99	2.1 (0.1)	100	2.1 (0.1)	99	2.1 (0.1)	99
Mg^{2+}…D52	2.0 (0.05)	100	2.0 (0.1)	100	2.0 (0.1)	100	2.0 (0.1)	100
Mg^{2+}…D126	2.0 (0.05)	100	2.0 (0.1)	100	2.0 (0.05)	100	2.0 (0.1)	100
Mg^{2+}-O_W…O_{P1}[b]	2.9 (0.3)	96	2.9 (0.3)	76	2.9 (0.3)	100	2.9 (0.2)	91
1 Metal (Direct Metal–Substrate Coordination)								
	Replica 1		Replica 2		Replica 3		Average	
	Average (std. dev.)	% Occ.	Average (std. dev.)	% Occ.	Average (std. dev.)	% Occ.	Average (std. dev.)	% Occ.
Mg^{2+}…W1	2.1 (0.1)	98	2.1 (0.1)	99	2.1 (0.1)	98	2.1 (0.1)	98
Mg^{2+}…W2	2.1 (0.1)	98	2.1 (0.1)	98	2.1 (0.1)	100	2.1 (0.1)	99
Mg^{2+}…W3	2.1 (0.1)	91	2.1 (0.1)	81	2.1 (0.1)	99	2.1 (0.1)	90
Mg^{2+}…D52	2.0 (0.05)	100	2.0 (0.05)	100	2.0 (0.05)	100	2.0 (0.05)	100
Mg^{2+}…D126	1.9 (0.05)	100	2.0 (0.1)	100	2.0 (0.05)	100	2.0 (0.1)	100
Mg^{2+}…O_{P1}[b]	2.0 (0.1)	100	2.0 (0.1)	100	2.0 (0.1)	100	2.0 (0.1)	100
2 Metals								
	Replica 1		Replica 2		Replica 3		Replica Average	
	Average	% Occ.	Average	% Occ.	Average	% Occ.	Average	% Occ.

	(std. dev.)		(std. dev.)		(std. dev.)		(std. dev.)	
Mg_A^{2+}…Mg_B^{2+}	4.1 (0.1)	99	4.1 (0.1)	99	4.1 (0.1)	99	4.1 (0.1)	99
Mg_A^{2+}…W1	2.1 (0.1)	100	2.1 (0.1)	100	2.1 (0.05)	100	2.1 (0.1)	100
Mg_A^{2+}…W2	2.1 (0.1)	99	2.1 (0.1)	99	2.1 (0.05)	98	2.1 (0.1)	98
Mg_A^{2+}…W3	2.1 (0.1)	100	2.1 (0.1)	100	2.1 (0.05)	100	2.1 (0.1)	100
Mg_A^{2+}…O_{P1}[b]	2.0 (0.1)	99	2.0 (0.1)	98	2.1 (0.1)	97	2.0 (0.1)	98
Mg_A^{2+}…D52	2.0 (0.05)	100	2.0 (0.05)	100	2.0 (0.1)	100	2.0 (0.05)	100
Mg_A^{2+}…D240	2.0 (0.1)	100	2.0 (0.1)	100	2.0 (0.1)	100	2.0 (0.1)	100
Mg_B^{2+}…W4	2.1 (0.1)	99	2.1 (0.1)	99	2.2 (0.1)	86	2.1 (0.1)	95
Mg_B^{2+}…W5	2.1 (0.1)	100	2.1 (0.1)	100	2.1 (0.1)	100	2.1 (0.1)	100
Mg_B^{2+}…W6	2.2 (0.1)	88	2.2 (0.1)	64	2.1 (0.1)	99	2.1 (0.1)	84
Mg_B^{2+}…D52	1.9 (0.04)	100	1.9 (0.04)	100	1.9 (0.04)	100	1.9 (0.04)	100
Mg_B^{2+}…D126 (Oδ1)	2.0 (0.1)	99	2.0 (0.1)	85	2.0 (0.1)	100	2.0 (0.1)	95
Mg_B^{2+}…D126 (Oδ2)	2.1 (0.1)	98	2.1 (0.1)	97	2.1 (0.1)	95	2.1 (0.1)	97

[a]Refer to main text Figures 6.4a, 6.5a, and 6.6a for a schematic of the metal-coordination distances and nomenclature. [b]O_{P1} refers to a non-bridging oxygen of the scissile phosphate.

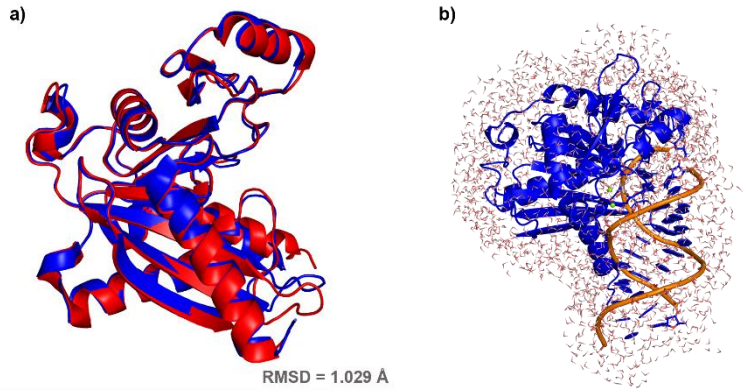

Figure E.1. a) Overlay of X-ray crystal structures of apo-human EndoV (blue, PDB ID: 6OZE) and mouse EndoV (red, PDB ID: 6OZO). b) The full human EndoV–RNA QM/MM model used in the present work.

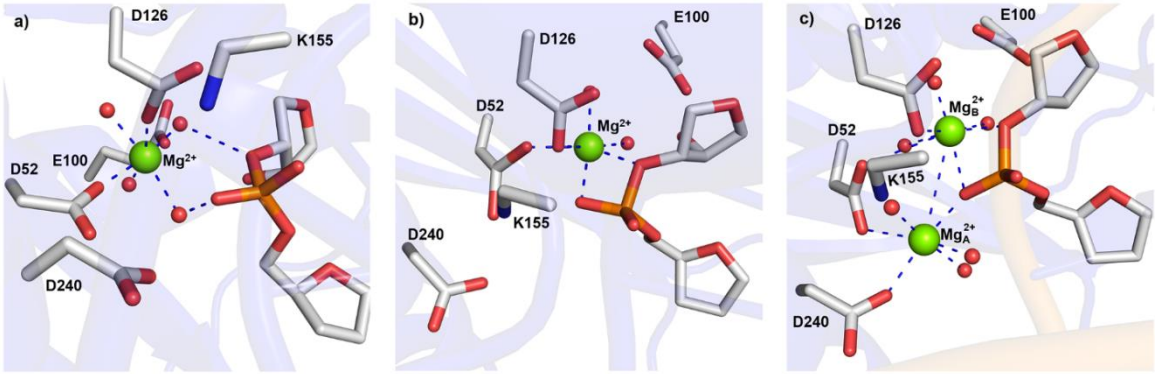

Figure E.2. Active site metal binding architectures used for MD simulations on human EndoV models in the present work that contain one metal with a) indirect (water-mediated) or b) direct metal–substrate coordination, or d) two metals.

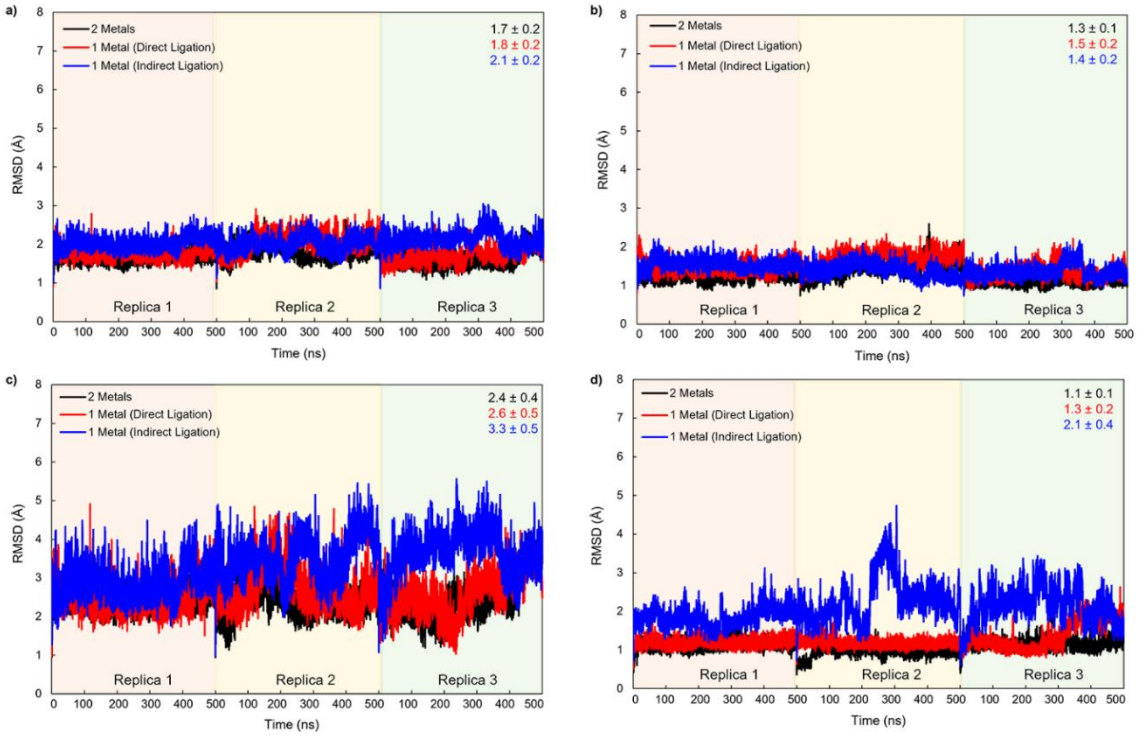

Figure E.3. RMSD (Å) as a function of time for each 500 ns MD simulation replica on the RC containing two metals (black), or one metal with direct (red) or indirect (blue) coordination to the substrate with respect to a) the backbone of the entire enzyme–RNA complex, b) the enzyme backbone, c) the RNA backbone, and d) all heavy atoms of active site residues (D52, E100, D126, K155, D240, Mg^{2+} or Mg_A^{2+} and Mg_B^{2+}, A250, U251, and A252). Average RMSD and standard deviation provided for the corresponding data set (top right corner).

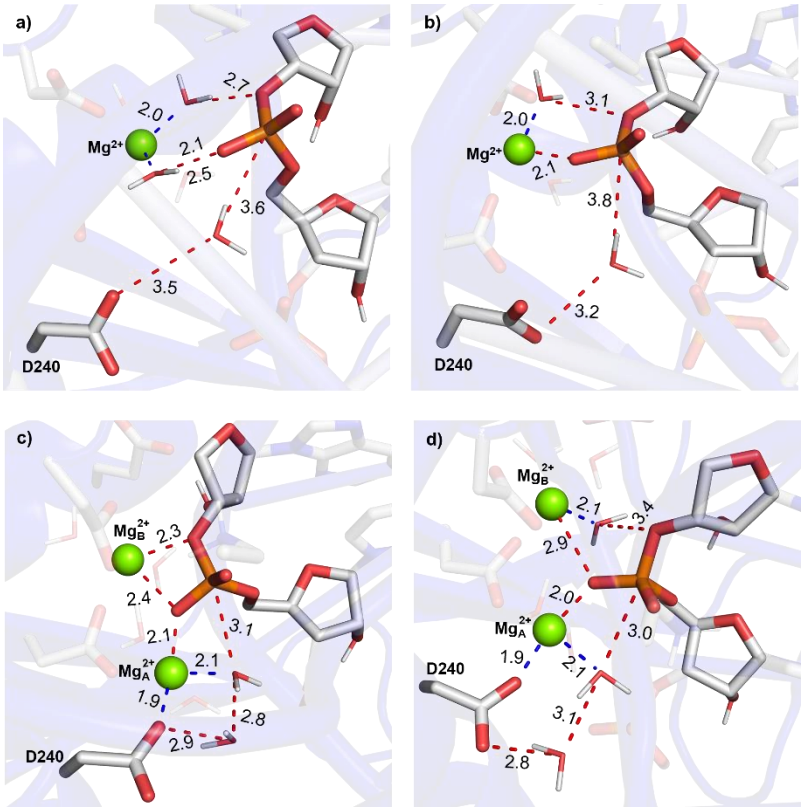

Figure E.4. MD snapshots of the human EndoV active site with different modes of metal–substrate binding and number of metals. Each panel highlights the key distances (Å) for active site residues that are positioned for catalysis.

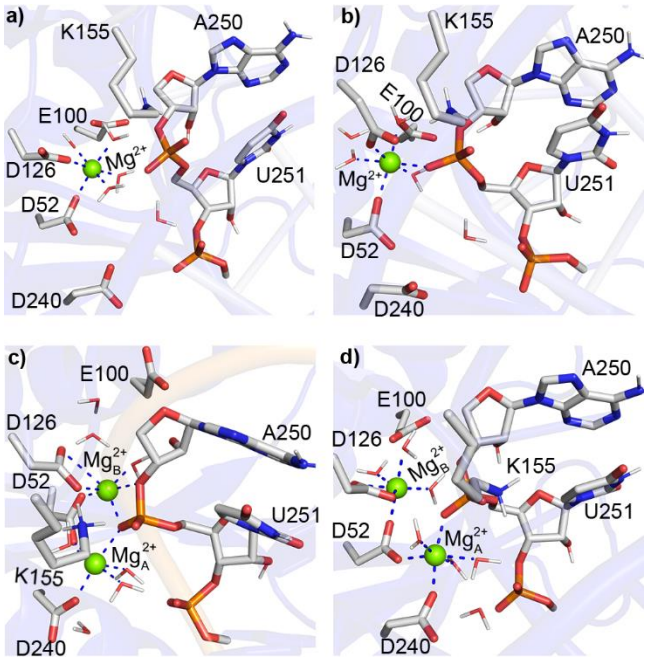

Figure E.5. The QM region of the QM/MM model containing one metal with a) indirect (water-mediated) or b) direct coordination to a non-bridging oxygen of the scissile phosphate, or two metals with c) direct or d) indirect Mg_B^{2+} coordination to the leaving group. See Figure E.1b for the full enzyme–DNA QM/MM model.

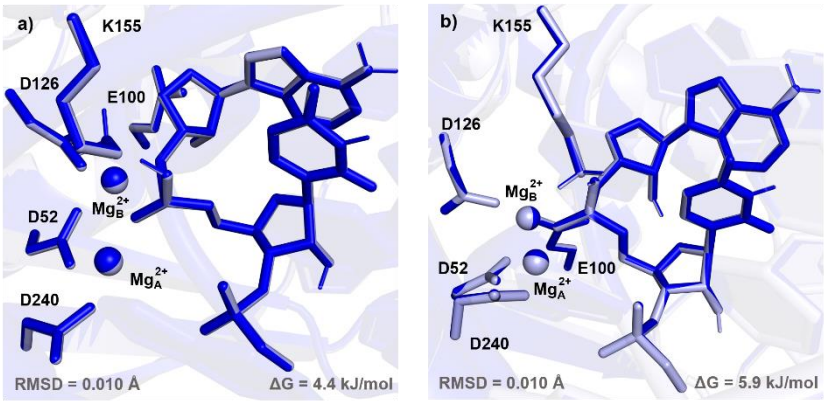

Figure E.6. Overlay of the active sites of the ICs obtained by scanning r(P–O_W) in TS1 (IC, dark blue) and r(P–O3′) in TS2 (IC′, light blue) for the EndoV-mediated phosphodiester bond cleavage mechanisms characterized in the present work involving two metals with a) direct or b) indirect (water-mediated) Mg_B^{2+}–O3′ coordination. The Gibbs energy difference was calculated for IC′ with respect to IC.

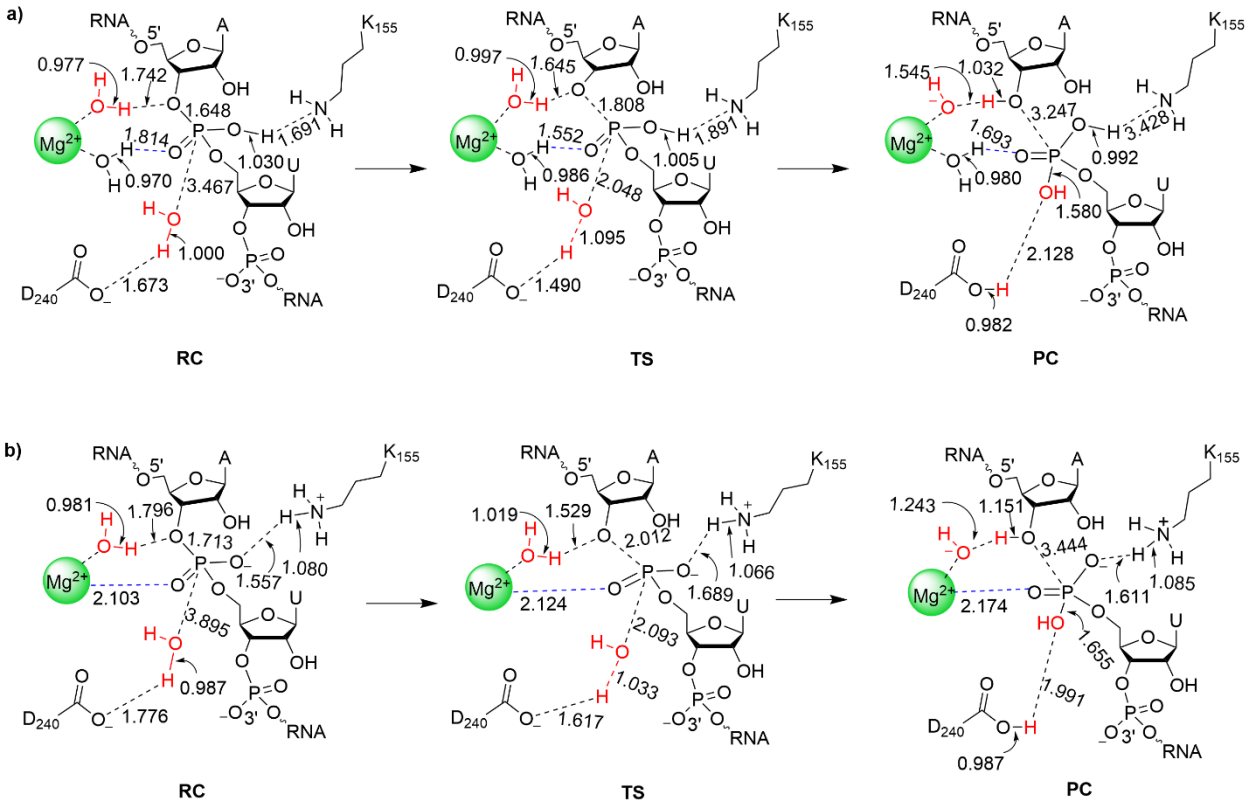

Figure E.7. Mechanism and key ONIOM(B3LYP-D3(BJ)/6-31G(d,p):AMBER) bond distances (Å) for EndoV-mediated phosphodiester bond cleavage involving a single metal with a) indirect (water-mediated) or b) direct coordination to a non-bridging phosphate oxygen of the scissile phosphate.

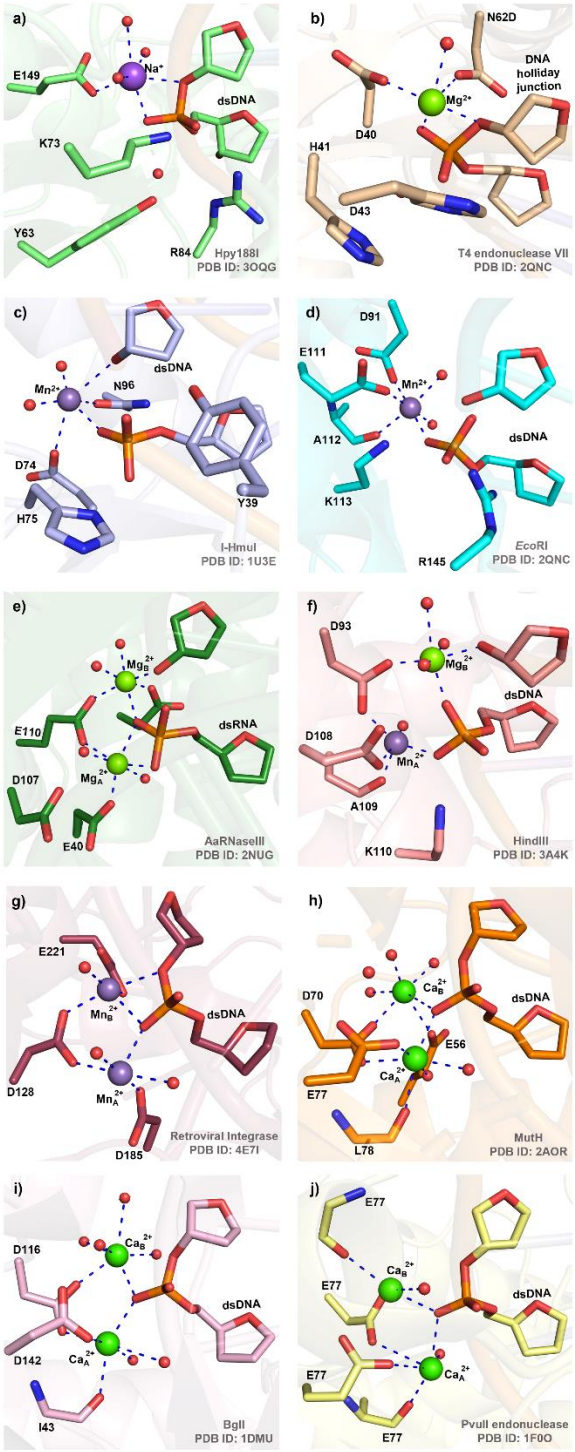

Figure E.8. X-ray crystal structures of the active site in the: a) RC of Na^+-substituted Hpy188I, b) RC of the N62D mutant of T4 endonuclease VII, c) PC of wild-type I-HmuI, d) PC of wild-type *Eco*RI, e) PC of wild-type *Aquifex aeolicus* RNase III, f) PC of wild-type HindIII, g) RC of wild-type retroviral integrase, h) RC of wild-type MutH, i) RC of wild-type BglI, and j) RC of wild-type PvuII endonuclease.

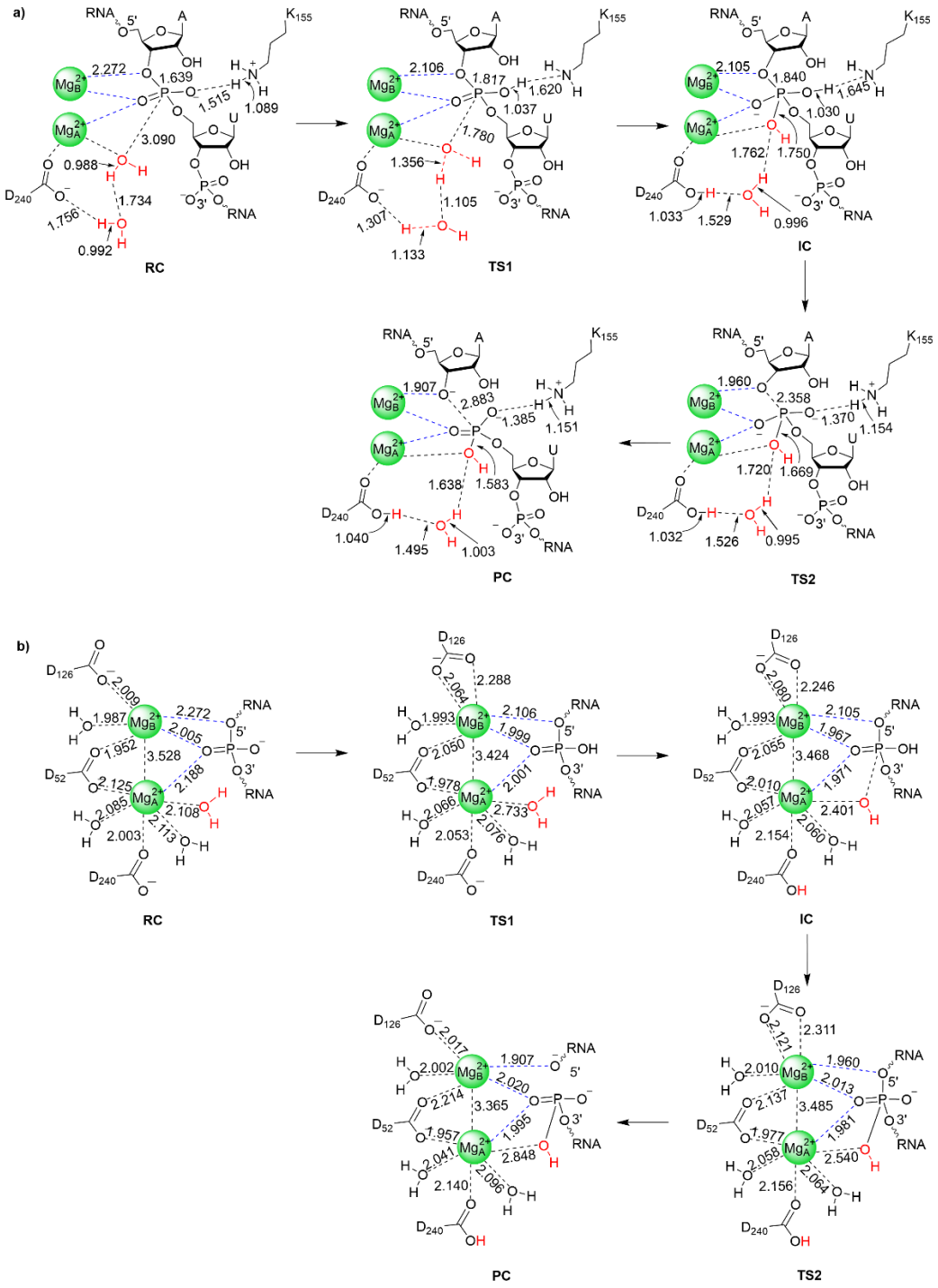

Figure E.9. a) Mechanism and key ONIOM(B3LYP-D3(BJ)/6-31G(d,p):AMBER) bond distances (Å) for EndoV-mediated phosphodiester bond cleavage involving two metals with direct Mg_B^{2+}–O3′ coordination, and b) corresponding metal coordination distances (Å).

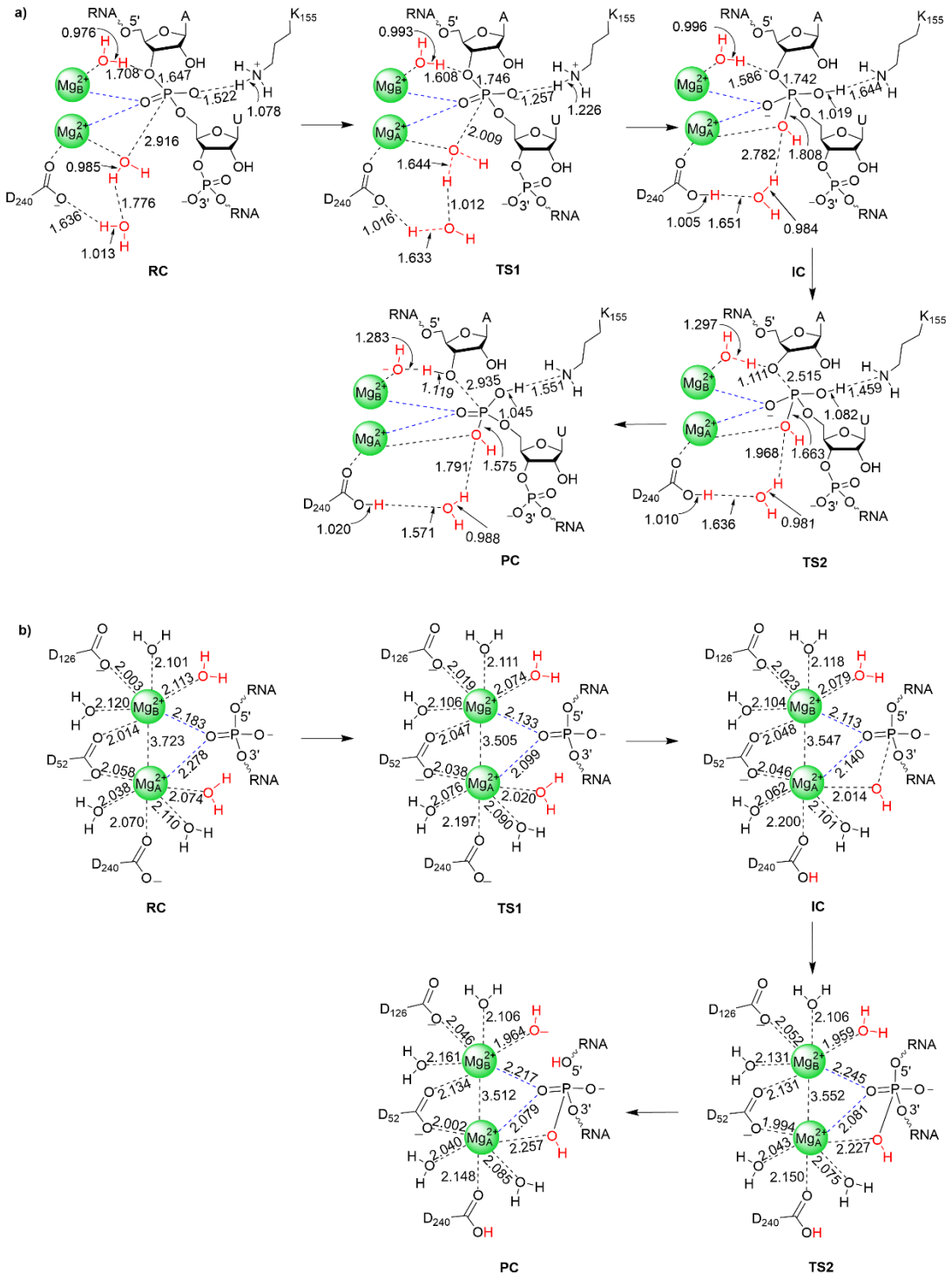

Figure E.10. a) Mechanism and key ONIOM(B3LYP-D3(BJ)/6-31G(d,p):AMBER) bond distances (Å) for EndoV-mediated phosphodiester bond cleavage involving two metals with indirect (water-mediated) Mg_B^{2+}–O3′ coordination, and b) corresponding metal coordination distances (Å).

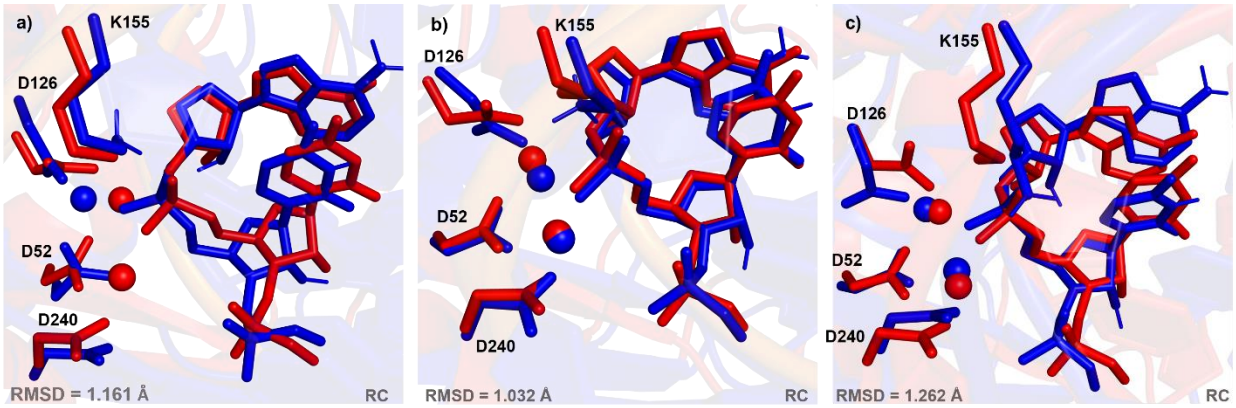

Figure E.11. Overlay of the active sites of the RC from X-ray crystal structures of mouse EndoV (red) and QM/MM optimizations for the feasible human EndoV pathways characterized in the present work involving a) one metal, or two metals with b) direct or c) indirect Mg_B^{2+}–O3′ coordination.