Monovalent (K^+/Na^+) and Divalent (Mg^{2+}/Ca^{2+}) metal ion selectivity in Group II intron and Cas1 protein

A Thesis

Submitted in partial Fulfilment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

by

Abhishek Kumar

Under supervision of

Dr. Priyadarshi Satpati

Department of Biosciences and Bioengineering

Indian Institute of Technology Guwahati

Guwahati-781039, Assam, India

Dedicated to my family and friends

Indian Institute of Technology Guwahati
Department of Biosciences and Bioengineering

DECLARATION

I do hereby declare that the content embodied in this thesis entitled "**Monovalent (K^+/Na^+) and Divalent (Mg^{2+}/Ca^{2+}) metal ion selectivity in Group II intron and Cas1 protein**" is the result of investigations carried out by me in the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati for the award of degree of Doctor of Philosophy, under the supervision of Dr. Priyadarshi Satpati.

As per the general norms of reporting research findings, due acknowledgments have been made wherever the research findings of other researchers have been cited in this thesis.

Abhishek Kumar

Date: 14/12/2021

Abhishek Kumar
(156106033)

Indian Institute of Technology Guwahati

Department of Biosciences and Bioengineering

CERTIFICATE

It is certified that the work described in this thesis entitled "**Monovalent (K^+/Na^+) and Divalent (Mg^{2+}/Ca^{2+}) metal ion selectivity in Group II intron and Cas1 protein**", by Mr. Abhishek Kumar (Roll No. 156106033) for the award of degree of Doctor of Philosophy is an authentic record of the results obtained from the research work carried out under my supervision in the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, India. This work has not been submitted elsewhere for the award of any degree or diploma.

Dr. Priyadarshi Satpati
(Thesis Supervisor)

Date: 14/12/2021

ACKNOWLEDGEMENT

Finally, at the end of it, what seems to be an unending road towards completing my thesis, I would like to thank everyone whose criticisms, comments, questions, support, and encouragement, both personal and intellectual, have left their imprint on this work.

First and foremost, I would like to thank my research supervisor, Dr. Priyadarshi Satpati, for his continuous support, precious advice, direction, and supervision of the research that enabled me to complete the Ph.D. thesis. He provided me with intellectual freedom in my work and endless intellectual discussions to improve the quality of work in all my pursuits. I was fortunate enough to work as a student under his supervision.

I am also grateful to the members of my doctorate committee, Prof. Shankar Prashad Kanaujia (Chairman), Dr. Shrisha Nagotu, Dr. Anki Reddy Katha, and Prof. Bhubneswar Mondal for their timely and valuable comments, which helped me better the work for my Ph.D. thesis.
I owe my thanks to the present and former Head of the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, for providing me the necessary facilities to fulfil my Ph.D. thesis objectives.

In addition, I want to thank the Bioinformatics Infrastructure Facility (BIF) and Supercomputing facility (PARAM-ISHAN) at IIT Guwahati for providing the computing facility. I would like to thank the Ministry of Human Resource Development and Indian Institute of Technology Guwahati for providing financial assistance during my Ph.D. tenure.

I will be eternally thankful to the faculty that taught me the essential principles in my undergraduate and postgraduate studies, including Dr. T. Rathinavelan, Dr. A. Roy, Dr. B.K. Patel, Dr. N.K. Raghavendra, Dr. V. Sharma, and Mr. G.K. Singh.

It was a pleasure to work with my lab members Dr. Amit, Suvankar, Gaurav, Vignesh, Pallav, Debaditya, Tiasha, Sonali, and Ayaz. I thank them for all the discussions and all the fun we have had throughout my Ph.D. tenure.

I am deeply indebted to all my friends: Sandeep, Dr. Kuldeep, Krishna, Arti, Ponoop, Nitesh, Vinay, and others for providing a soothing environment and helping in the critical situations during my Ph.D., whenever I needed. I must acknowledge all my school, university friends, and Ph.D. batch mates for their love, encouragement, and support.

I would also like to thank Dr. Sunanda Chatterjee and the December Trip group for all of the wonderful experiences.

My heartfelt gratitude and appreciation go to my parents and family for their blessings, love, patience, support, and understanding during this juncture.

Synopsis

Metal ions play crucial roles in biological processes (e.g., catalytic activity, structural stability, transportation, homeostasis, muscle contraction, etc.). Thus, metal-ion selectivity is essential for the function and stability of the biomolecules. This dissertation explores the use of combined classical molecular dynamics (MD) and ab initio quantum chemical calculations to estimate the energetics of metal ion selectivity (K^+ vs. Na^+, Mg^{2+} vs. Ca^{2+}) in group II intron and cas1 protein. Considering X-ray structures of metal-ion bound biomacromolecule (group II intron and cas1 protein) as a template, the strength of selectivity ($\Delta\Delta G$= Free energy difference between cognate vs. near-cognate metal binding to the active site of the biomolecule) was estimated using MD free energy simulations employing appropriate thermodynamic cycles. Classical force-fields are limited by the fact that it does not include electronic polarizability explicitly. Instead of using computationally expensive polarizable force-field, quantum chemical ab initio calculations were performed on the reduced model of the metal ion binding pocket to incorporate the polarization effect. The adopted methodology (combining classical MD and quantum chemical calculation) is not only computationally cheap but also successful in establishing a direct link between the estimated energetics and the 3D structures of cognate and near-cognate metal complexes. This thesis is divided into six chapters.

Chapter 1 consists of a brief overview of the group II intron and cas1 protein, methodology for molecular dynamics simulations, thermodynamic cycles, and statistical methods adopted in this thesis for estimating free energies were elaborated.

Group II introns are enzymes that undergo self-splicing and remove themselves from pre-messenger RNA. X-ray structures of group II intron of *Oceanobacillus iheyensis* at various stages of the self-splicing pathway (pre-hydrolytic, post-hydrolytic, and ligand-free state) revealed an intricate atomic interaction network in the active site of the intron. Biochemical studies confirmed that a heteronuclear metal ion cluster consisting of four metal ions (K1, K2 sites with K^+ and M1, M2 sites with Mg^{2+}) are crucial for function (Figure 1). Monovalent ion K^+ is very important for the function, and buffers containing only Na^+ were shown to cease self-splicing activity of group II intron. Similarly, the substitution of Mg^{2+} by Ca^{2+} also results in the loss of enzymatic activity. The X-ray structures provide good models for quantitative estimation of the energetics associated with monovalent (K$^+$ vs. Na$^+$) and divalent (Mg^{2+} vs. Ca^{2+}) selectivity by group II intron.

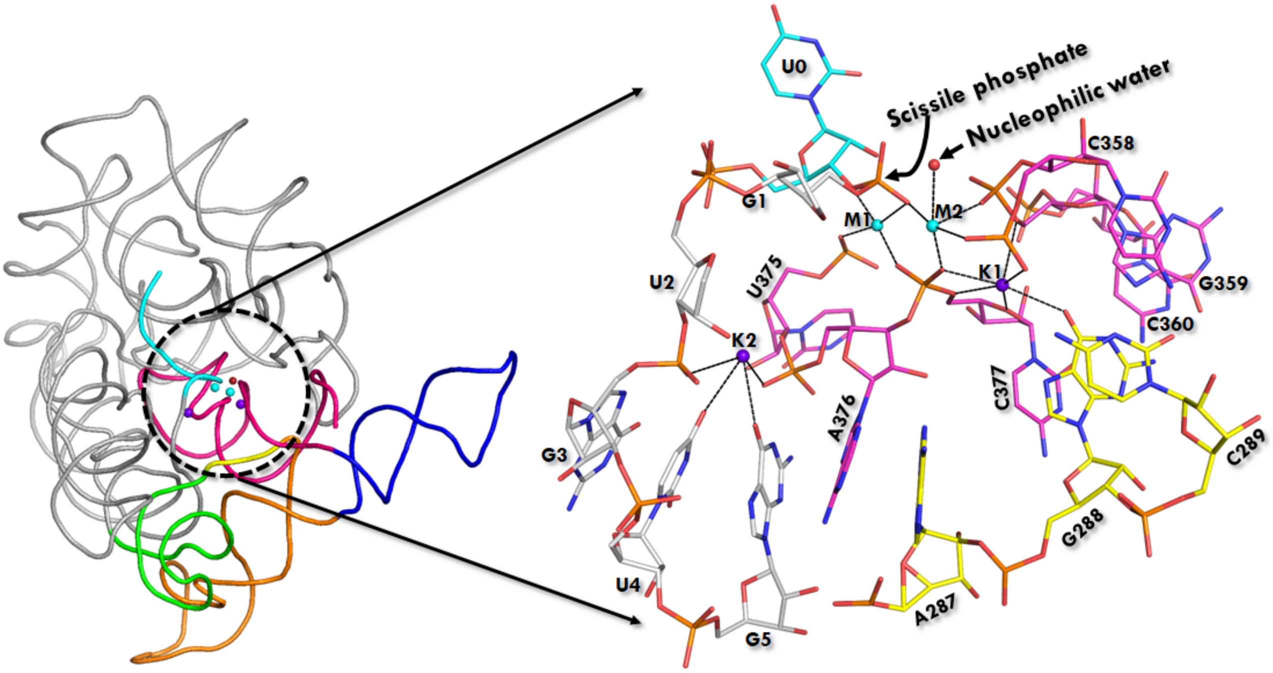


Figure 1. The architecture of group II intron. Cartoon representation of pre-hydrolytic state (PDB: 4FAQ): Different colour represents different domains. D1 (grey), D2 (green), D3 (orange), D4 (blue), D5 (pink), J2/3 (yellow), 5’ exon (cyan), monovalent metal ions K1, K2 (purple sphere) and divalent metal ions M1, M2 (cyan sphere). The dashed circle highlights the active site region considered for the MD study. Close up showing the active site elements. The nucleophilic water (red sphere) is placed for the inline attack to the scissile phosphate. The key RNA residues around the metal ions are shown in line representation.

Chapter 2 discusses the principle of monovalent ion (K$^+$ vs. Na$^+$) selectivity in the active site of group II intron at different stages of splicing (i.e., pre-hydrolytic state, post-hydrolytic state, and ligand-free state) of *Oceanobacillus iheyensis* group II introns. X-ray structures (pre-hydrolytic, post-hydrolytic, ligand/exon free intron) were used as the template for MD simulations. It has been shown that the strength of monovalent ion selectivity ($\Delta\Delta G$: K$^+$ versus Na$^+$ binding) varies along the self-splicing pathway (Figure 2). The ligand-free and pre-hydrolytic states are highly selective for K$^+$ over Na$^+$, and the selectivity is lost in the post-hydrolytic state. The simulations revealed that Na$^+$ in the free and pre-hydrolytic state is trapped with an unsatisfied first coordination shell in the active site, which is responsible for large discrimination (~ 6 -7 Kcal/mol). The Na^+ in the active site of the post-hydrolytic state allows water entry and satisfies the Na^+ bonding requirement, resulting in low discrimination (~ 1 kcal/mol). The results not only give insights into the experimentally unresolved Na^+ bound complexes but also provide the link between structures and computed energetics.

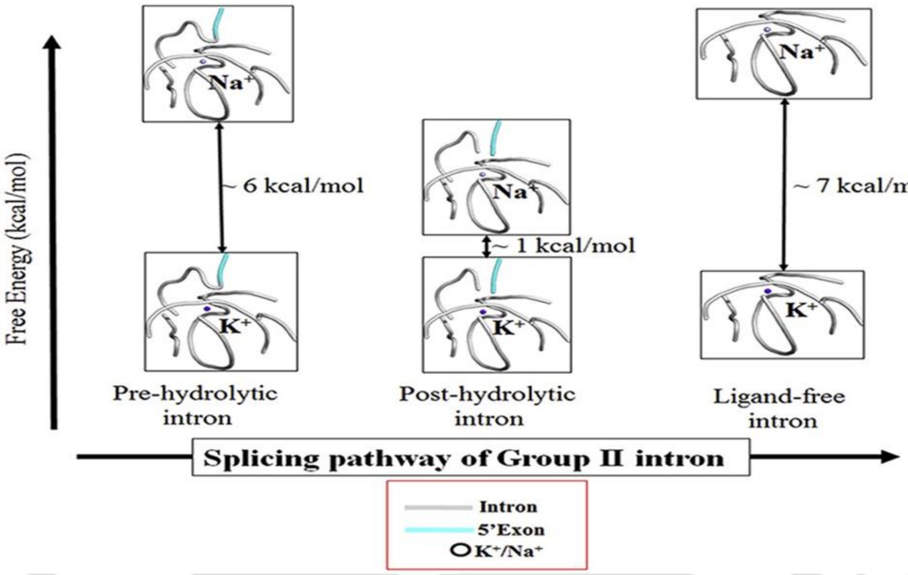


Figure 2. Monovalent ion selectivity (K^+ vs. Na^+) in group II intron at various stages of splicing. Intron (grey), exon (cyan), and ion (sphere).

Chapter 3 discusses the energetics of divalent metal ion (Mg^{2+} vs. Ca^{2+}) selectivity in the active site of the group II intron. Combined classical MD and quantum chemical calculations confirm that the group II intron prefers Mg^{2+} relative to Ca^{2+} (Figure 3. a, b).

Mg^{2+} bound active site is compact and dry. Substitution of Ca^{2+} results in water entry (wet) and loosens up the M1, M2 pocket due to the larger size of Ca^{2+}. The orientation of nucleophilic water relative to scissile phosphate is strikingly different between Mg^{2+} and Ca^{2+} bound pre-hydrolytic state (Figure 3. c). Furthermore, the scissile phosphate oxygens form direct interactions with two Mg^{2+} (M1, M2) in the pre-hydrolytic state. Ca^{2+} substituted pre-hydrolytic state suggests only one direct interaction between the scissile phosphate oxygen and Ca^{2+} (M1). Mg^{2+} or Ca^{2+} bound post-hydrolytic state suggests a direct interaction between the exon (U0) and M1, whereas the intron part (G1) is highly dynamic in the MD trajectories (Figure 3. d). Simulations suggest that the nature of the divalent metal-ion (Mg^{2+} vs. Ca^{+2}) does not disrupt the overall architecture/integrity of group II intron but could perturb the local environment, which might be sufficient to modulate the catalytic activity (Figure 3. c). The absence of exon part in the ligand-free state allows maximum water exposure of M1, M2 sites.

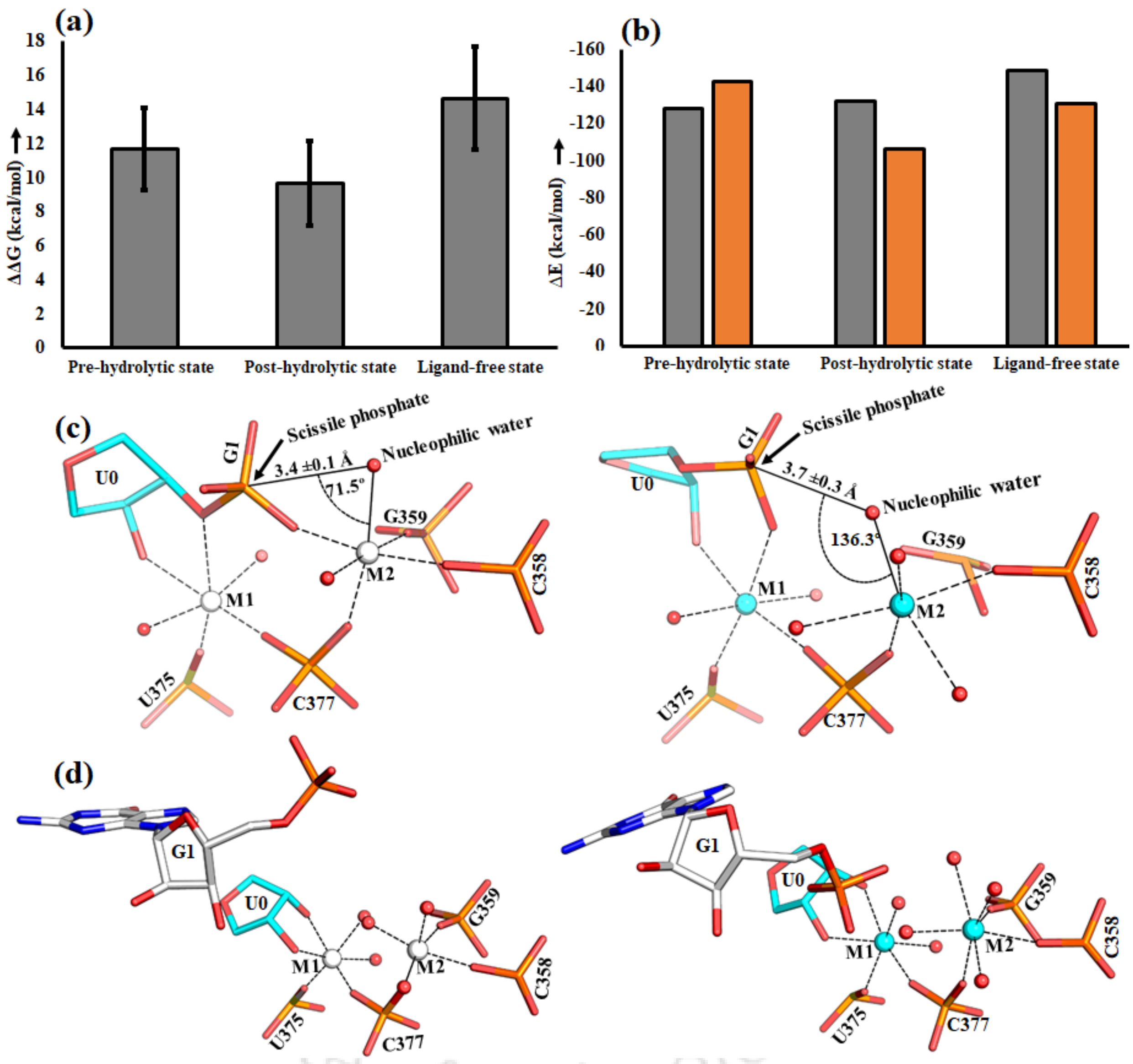


Figure 3. (a) Relative binding free energy (ΔΔG, estimated by classical MD simulation) of divalent ion (Mg^{2+} vs. Ca^{2+} in M1, M2 sites) binding to group II intron at various stages of splicing. (b) ΔE was estimated from quantum chemical calculations (B3LYP/6-311+G* (grey bar) and M06–2X/6-311++G** (orange bar) level of theory). ΔE approximately captures the relative divalent cation preference in the group II intron. Negative ΔE indicates Mg^{2+} preference relative to Ca^{2+}. (c) MD structures of Pre-hydrolytic intron active site and waters around the divalent ions (M1, M2). Mg^{2+}-bound intron, where M1, M2 =Mg^{2+} (left), and Ca^{2+}-

bound intron (right), where M1, M2 =Ca^{2+}. For clarity, hydrogens and nucleotide bases were not shown. (d) MD structures of Post-hydrolytic intron active site and waters around the divalent (M1, M2) ions. (left) Mg^{2+}-bound active site. (right) Ca^{2+}-bound active site. For clarity, hydrogens and nucleotide bases except G1 were not shown. G1 is away from divalent metal ion sites (M1, M2).

In **Chapter 4,** we discussed divalent (Mg^{2+} vs. Ca^{2+}) and monovalent (K^+ vs. Na^+) ion selectivity in the active site of lariat group II intron of a eukaryotic brown algae *Pylaiella littoralis* in the lariat-3′ exon (pre-hydrolytic state) and post-hydrolytic state, respectively (Figure 4). The results indicate that the intron is selective between Mg^{2+} and Ca^{2+}, favouring the former but non-selective for monovalent ions (K^+ versus Na^+). Water accessibility is found to be associated with the strength of selectivity.

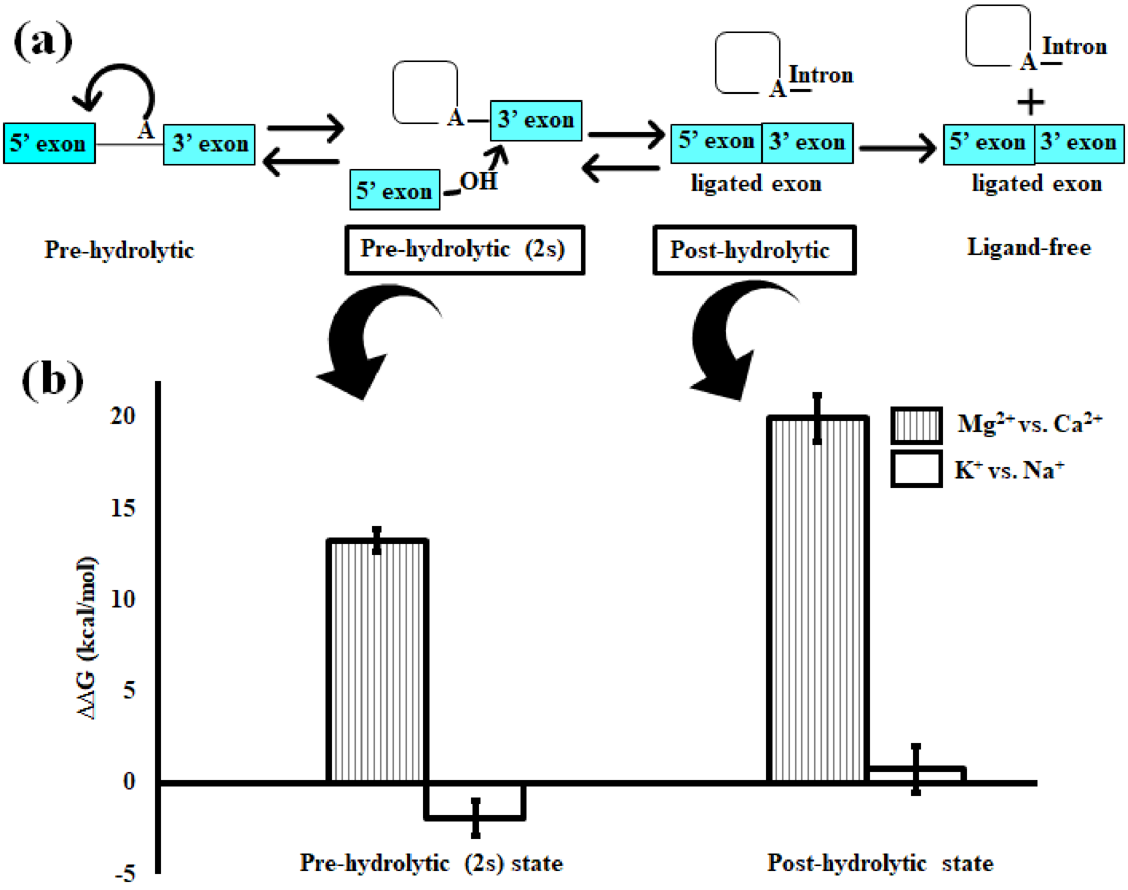


Figure 4. (a) Schematic representation of the branched mechanism of splicing in group II intron. (b) Relative binding free energies of Mg^{2+} vs. Ca^{2+} (vertical stripes bar) and K^+ vs. Na^+ (white bar) binding to group II intron. Error bar (in s.e.m).

In **Chapter 5**, divalent ion selectivity (Mg^{2+} vs. Ca^{2+}) by Cas1 protein (Figure 5.1) has been discussed. Cas1 is a universally conserved essential metalloenzyme of the CRISPR-Cas immune system of prokaryotes that can cut and integrate a part of viral DNA into its host genome with the help of other proteins. The integrated DNA acts as a memory of viral infection, which can be transcribed to RNA and stop future infection by recognition followed by protein-mediated degradation of the viral DNA. Experiments confirmed that Cas1-mediated DNA degradation is hindered by metal ion-substitution, metal ion chelation, or mutation of the binding pocket residues (e.g., E190A, H254A, D265A, D268A, etc.).

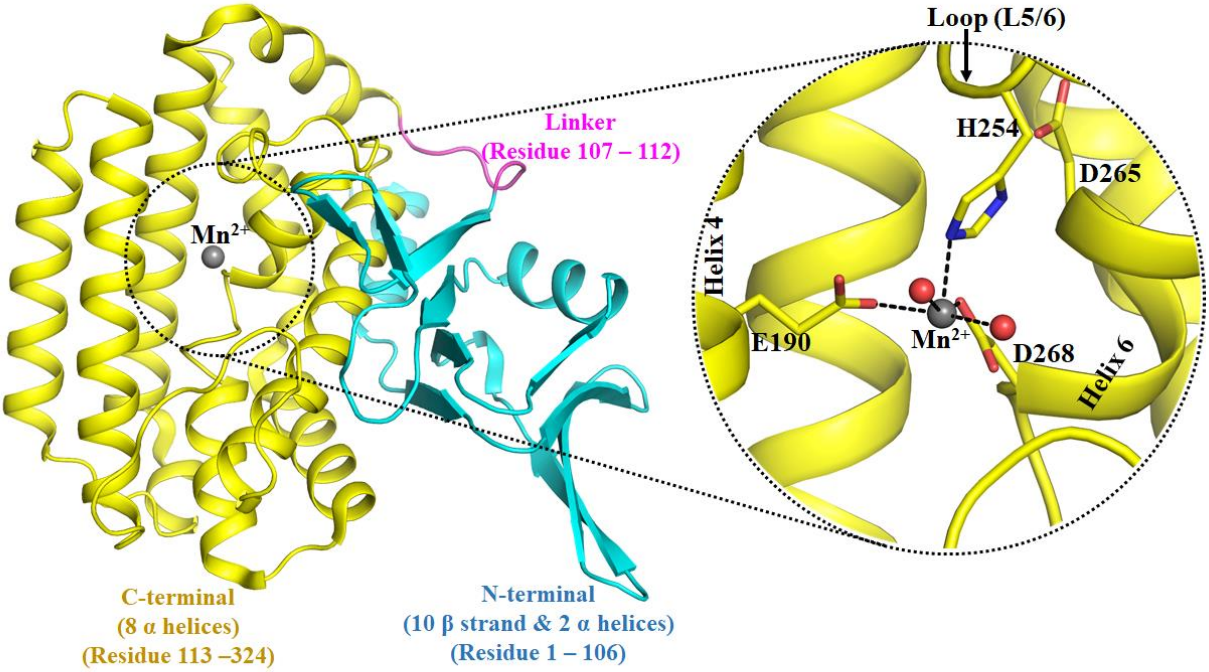


Figure 5.1. Cartoon representation of Cas1 protein (PDB: 3GOD). Cas1 is composed of two structurally distinct domains: N-terminal and C-terminal domains connected by a linker loop. Zoomed-in view of the divalent Mn^{+2} ion (grey sphere) binding pocket (broken circle). Key amino-acid residues of Cas1 are highlighted with sticks and crystal waters (within 2.5 Å of Mn^{2+}) in the red-sphere. The direct interactions with Mn^{+2} are shown as dashed lines.

Cas1 is active towards dsDNA degradation with both Mn^{2+} and Mg^{2+} but not with Ca^{2+}. Thus, efficient divalent-metal ion discrimination is essential for Cas1 mediated endonuclease activity. Using the crystal structure of Cas1 protein as a template, we modelled and performed classical molecular dynamics simulations with Mg^{2+}, Ca^{2+} bound wild-type (WT), and mutant Cas1 proteins and estimated the divalent metal ion selectivity. Classical molecular dynamics free energy calculations suggest that Cas1 imposes an energy penalty of ~ 4-5 kcal/mol for Ca^{2+}

binding relative to Mg^{2+} (Figure 5.2a). Quantum chemical calculations were performed on the reduced binding pocket models to incorporate the effect of polarizability. Quantum chemical calculations indicate: (1) Mg^{2+} binding is preferred over Ca^{2+}, in line with the results obtained from classical MD simulations, (2) Mn^{2+} is the most preferred ion in the WT Cas1 binding pocket among all the three ions studied in this work (i.e., Mg^{2+}, Ca^{2+}, Mn^{2+}) (Figure 5.2b). The metal ion binding pocket of Cas1 is dry when bound to Mg^{2+}. Ca^{2+} substitution results in a single water entry (wet) in the ion-binding pocket (Figure 5.2 c). Mutation of conserved amino acids that coordinates with the divalent metal ion also results in water exposure of the metal ion binding pocket. The strength of divalent metal ion selectivity depends on the wetness of the binding pocket. Wet binding pocket is less selective, and dry binding pocket is more selective in favor of Mg^{2+} binding relative to Ca^{2+}.

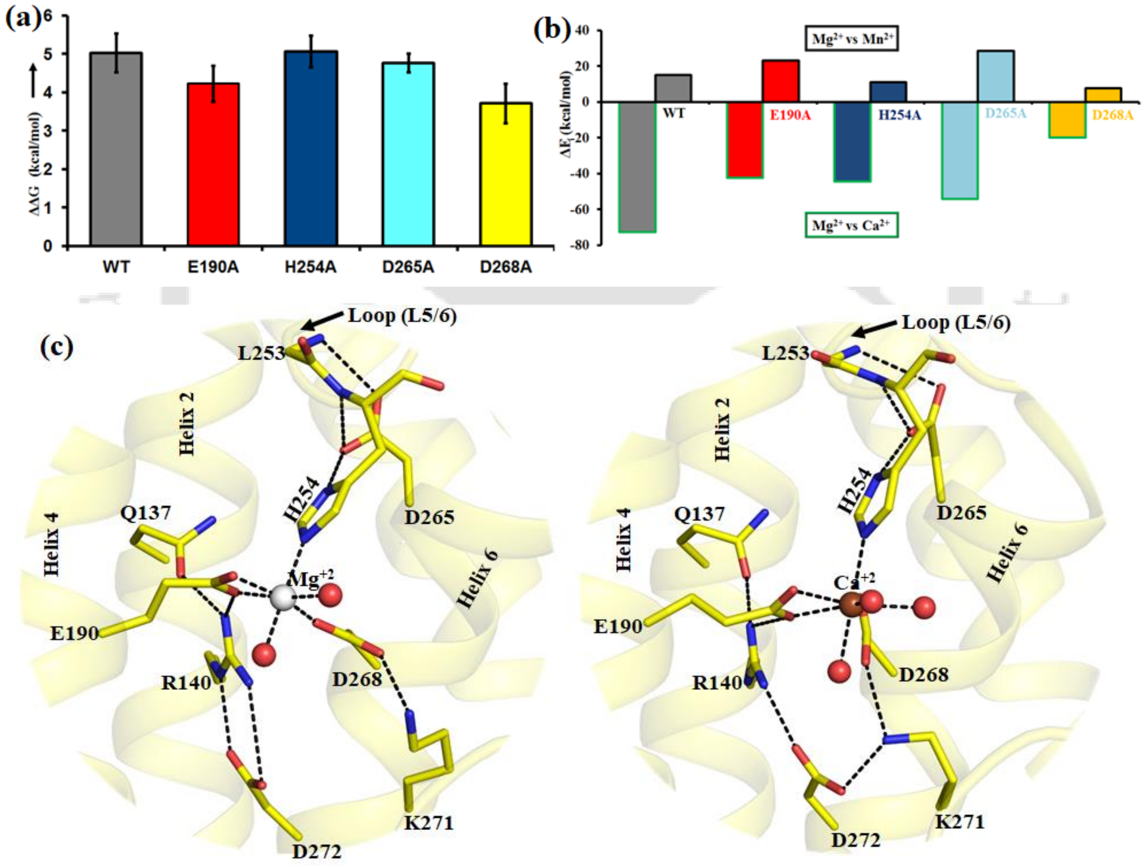


Figure 5.2 (a) Estimated energetics (ΔΔG; using classical molecular dynamics free energy simulations) of divalent metal ion selectivity: Mg^{2+} vs. Ca^{2+} binding to wild-type (WT) and mutant (E190A, H254A, D265A, and D268A) Cas1 protein. Positive ΔΔG indicates Mg^{2+} binding is preferred relative to Ca^{2+} in the binding pocket of Cas1. (b) Relative energies ΔE were estimated from quantum chemical calculations (M06–2X/6-311++G** level of theory).

ΔE approximately captures the relative divalent cation preference in the Cas1 active site. Positive ΔE indicates Mn^{2+} preference relative to Mg^{2+}. Negative ΔE indicates Mg^{2+} preference relative to Ca^{2+}. Divalent metal ion preference in the Cas1 active site follows the order $Mn^{2+} > Mg^{2+} > Ca^{2+}$. Bar colours represents wild-type and mutants of Cas1. (c) MD structures of wild-type Cas1 protein. (left) Mg^{2+}-bound active site, (right) Ca^{2+}-bound active site. Key amino acids are shown in the sticks, and the interaction network is shown in broken lines. Hydrogens were not shown for clarity. Waters (red sphere), Mg^{2+} (white sphere) and Ca^{2+} (dark red sphere).

In **Chapter 6**, the overall conclusion from this thesis work and the future scope of this current work is discussed. We hope the hypothesis that emerged from our study will encourage experimental verification.

Contents

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Appendices

List of Figures

List of Tables

Abbreviations

DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
CRISPR	Clustered regularly interspaced short palindromic repeats
WT	Wild type
MD	Molecular dynamics
MDFE	Molecular Dynamics Free Energy
mRNA	Messenger RNA
PBC	Periodic Boundary Condition
PDB	Protein data bank
PME	Particle mesh Ewald
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
TC	Thermodynamic Cycle
TI	Thermodynamic Integration
vdw	Van der Waals
LJ	Lennard-Jones

Chapter 1

Introduction, Methods, and Objectives

1.1 Introduction

Metal ions play crucial roles in biological processes (i.e., enzymatic or catalytic activity, structural stability, transportation, homeostasis, muscle contraction, etc.); thus, recognition of cognate metal ions is crucial for function (**Mustafi et al., 2003; Klein et al., 2004; Rolfs et al. 1999; Agranoff et al., 1998; Farah et al., 1995**). Accuracy of group II intron splicing, structural stability of ribosome and ribozyme, the catalytic activity of CRISPR-Cas proteins are a few examples that rely on metal ion selectivity (**Sigel, 2005; Klein et al., 2004; Denesyuk et al., 2015; Palermo, 2019**). The strength of metal ion selectivity depends on its ability to discriminate cognate metal ions from non-cognate analogues. The discriminatory strength is linked with the frequency of errors and is thus directly related to the accuracy of the biological processes. Structural studies (NMR, Cryo-EM, X-ray) highlight the key atomic interaction network crucial for metal ion binding. In contrast, biochemical studies reveal the kinetics and thermodynamics associated with metal ion binding. However, the link between structural and biochemical studies is lacking. Molecular dynamics (MD) free energy simulation is a popular approach for bridging the gap to some extent.

In this dissertation, we performed classical molecular dynamics free energy simulations (**Kollman, 1993; Tsui et al., 2001**) and electronic structure calculations to study metal ion selectivity (K^+ versus Na^+, Mg^{2+} versus Ca^{2+}) in RNA (group II intron) and protein (Cas1). X-ray structures of the biomolecules were used as a template, and the thermodynamics of the metal ion selectivity was estimated. The estimated strength of selectivity was linked to the 3D structures of the cognate and non-cognate complexes of intron/Cas1.

1.2 Group II intron and splicing

1.2.1 Background

Eukaryotic RNA generally contains various protein-coding regions and non-coding regions known as exons and introns, respectively. Exons are joined together by removing introns results in mature mRNA, which is then translated by the ribosome, resulting in protein synthesis in cells (**Figure 1.1**) **(Pyle et al., 2006; Clancy, 2008)**. Thus, gene expression fidelity is entirely reliant on precise RNA processing. Splicing can form a range of unique proteins by varying the exon composition in the mature mRNA, known as alternative splicing (**Pan et al., 2008; Wang et al., 2008**).

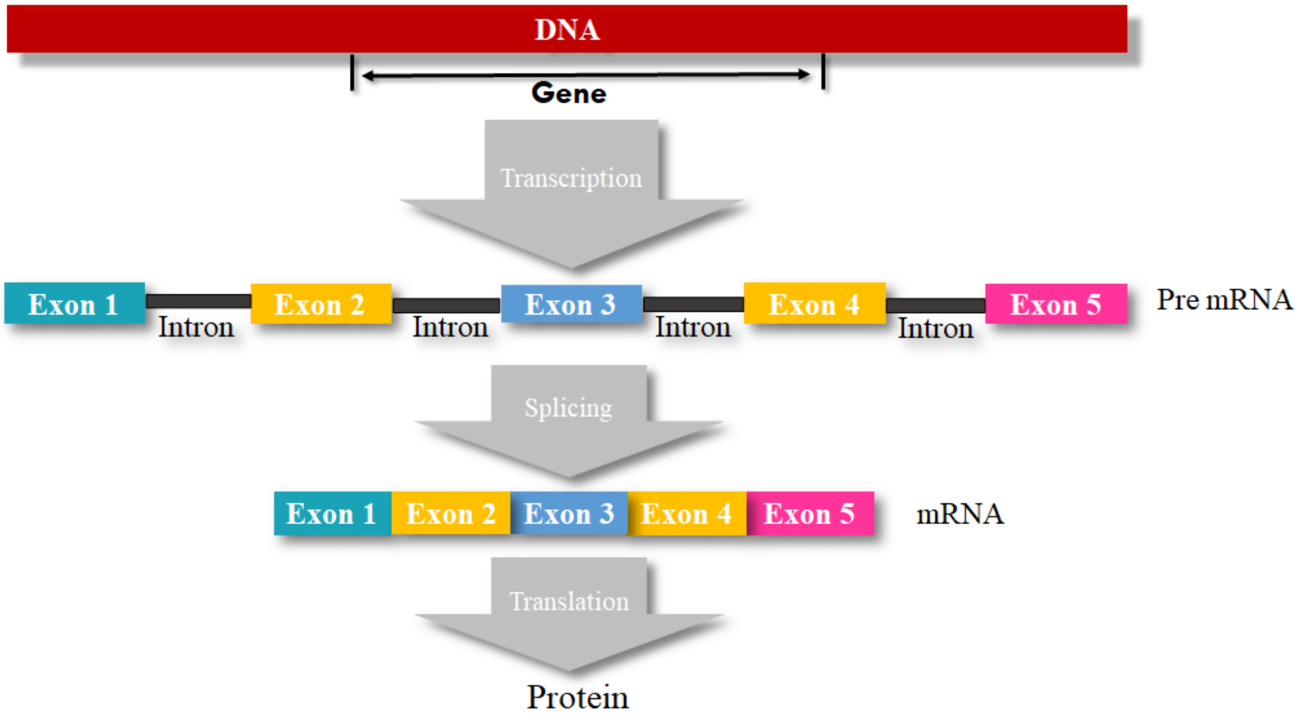


Figure 1.1. Schematic representation of RNA splicing

At the molecular level, splicing is critical for understanding genomic expression and contributing to the well-being of society. When splicing goes wrong, it can cause serious health problems for humans, such as muscular dystrophy, myotonia, retinitis pigments, and cystic fibrosis (**Cooper et al., 2008; Scotti et al., 2016**). Splicing is an advantageous intervention point for disease therapies since it occurs at an early stage in gene expression and does not affect the DNA. Understanding the idea of splicing can serve as a framework for the development of medications that may be effective in the prevention of disease (**Cooper et al., 2009; Scotti et al., 2016**).

It is needless to say that splicing is a fundamental process of gene expression. Introns are divided into four categories based on their splicing mechanism, i.e., tRNA, mRNA, group I, and group II introns. mRNA and tRNA introns are spliced out by ribonuclear protein complexes (spliceosome) and protein enzymes (kinase), respectively. Group I and II introns, on the other hand, do not require additional proteins because the folded intron can catalyze its own splicing, which is referred to as self-splicing (**Pyle et al., 2006**). Chemically, splicing for mRNA and group II introns is identical, as both processes require the nucleotide adenosine. However, group II introns undergo splicing without involving proteins. In contrast, group I and II introns undergo self-splicing but involve different nucleotides, guanosine for group I and adenosine for the II introns. Furthermore, group II introns can be considered an intermediate between group I introns and spliceosomes. Group II Intron is an excellent model system for studying the 3D structure, catalysis, and evolutionary aspect of the splicing process in all domains of life (**Mattick, 1994; Dai et al., 2002**).

1.2.2 Mechanism of group II intron splicing

There are two mechanisms for group II introns to catalyze their splicing through two successive transesterification reactions: (i) hydrolytic mechanism (ii) branched mechanism. During hydrolytic mechanism, nucleophilic water attacks as the nucleophile in the first step of splicing at the 5’ splice site, producing linear intron intermediate. In the second step, 3′ OH of the cleaved 5′ exon is the nucleophile and attacks the 3′-splice site, resulting in exon ligation and excision of a linear intron **(Figure 1.2 a)**. While, in branched mechanism, the 2′ OH of the bulged adenosine in DVI acts as the nucleophile to attack the 5′-splice site in the first step of splicing, producing an intron lariat intermediate. In the second step, the 3′ OH of the cleaved 5′ exon is the nucleophile and attacks the 3′-splice site, resulting in exon ligation and excision of a lariat intron **(Figure 1.2 b)** (**Marcia et al., 2012**).

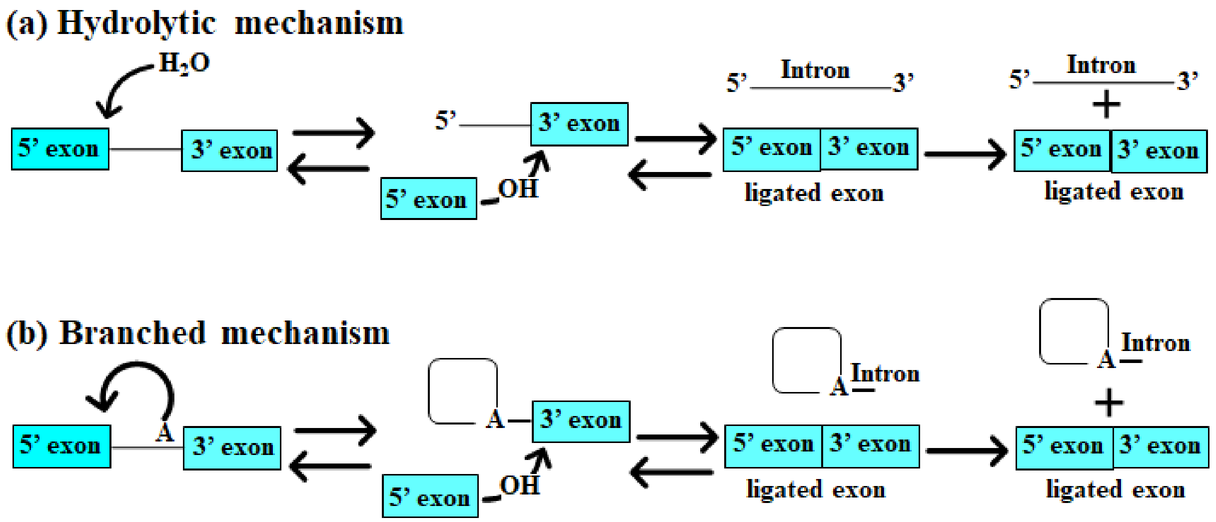


Figure 1.2. Splicing mechanism of group II intron. (a) Hydrolytic splicing pathway. The 5' splice site is attacked by the water molecule as the nucleophile in the first step of splicing, forming a free 5' exon. In the second step, 3' OH of the free 5' exon attacks the 3' spice site and form a ligated exon and linear intron. (b) The branched splicing pathway, where bulged adenosine nucleotide attacks at the 5' splice in the first step of splicing, form a free 5' exon and 3' lariat exon intermediate. The second step of splicing takes place as in the hydrolytic splicing pathway to produce ligated exon and lariat intron.

1.2.3 Structural studies of Group II intron

1.2.4 Hydrolytic splicing

Group II introns are large ribozymes composed of 400-1000 nucleotides in size, which can catalytically cut and religate themselves. The secondary structure of group II introns consists of 6 domains (D1-DVI). Group II intron with five domains (D1-D5) was structurally characterized for *Oceanobaciellus iheyensis* at various stages of the splicing pathway (e.g., pre-hydrolytic (PDB: 4FAQ), post-hydrolytic (PDB: 4FAR), and ligand-free state (PDB: 4E8M) state) (**Marcia et al., 2012**). The high flexibility of domain 6 (D6) was speculated to hinder structural characterization and thus removed for structure determination. Structural studies have helped us to visualize the splicing machine in atomic details and provide an excellent model for computational analysis. The remarkable set of tertiary interactions along with metal ions are responsible for the highly compact and globular shape of the functional form of group II intron.

1.2.5 Pre-hydrolytic state

Following transcription, group II intron folds spontaneously into a stable configuration, resulting in the formation of a highly reactive ribozyme in the presence of Mg^{2+}. Thus, in order to characterize the pre-hydrolytic state of group II intron, Ca^{2+} was used instead of Mg^{2+} in the buffer. Ca^{2+} is a known inhibitor of Mg^{2+} dependent enzymes; several endonuclease structures have previously been captured in their pre-catalytic state by substituting Ca^{2+} for Mg^{2+} (**Marcia et al., 2012**). The folded structure of group II intron in its pre-hydrolytic state is shown in **Figure 1.3**. D1 is the largest domain onto which all other domains dock. D5 is embedded into the catalytic core and is most important for catalysis.

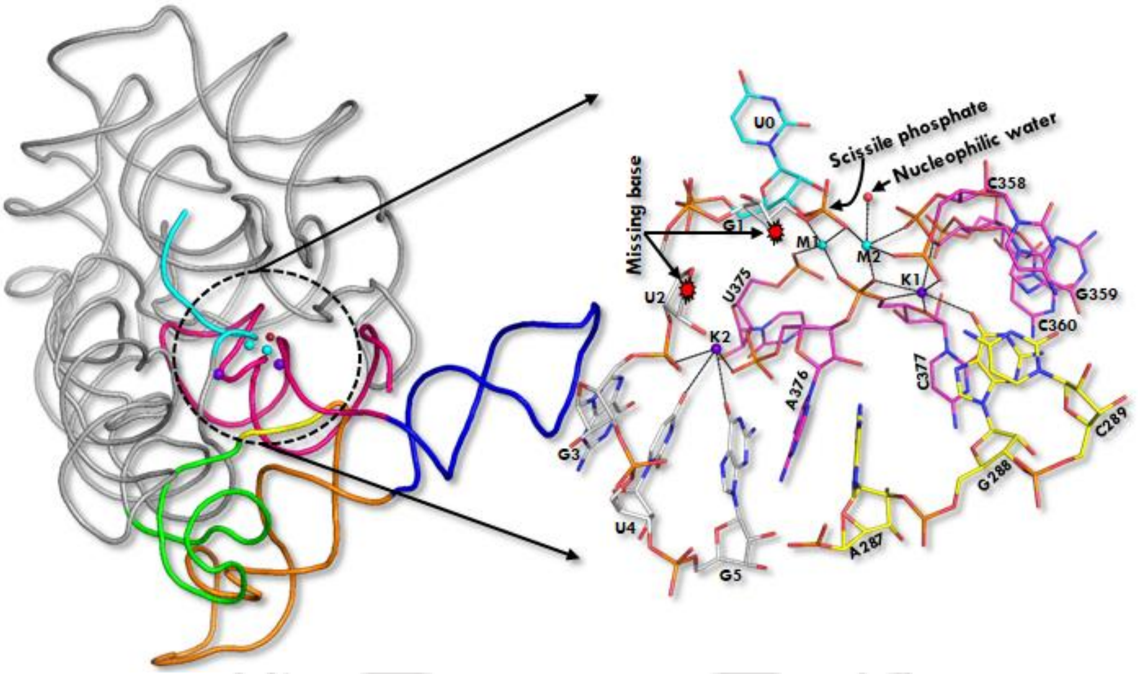


Figure 1.3. 3D architecture of group II intron. Cartoon representation of pre-hydrolytic state (PDB: 4FAQ): D1 (grey), D2 (green), D3 (orange), D4 (blue), D5 (pink), J2/3 (yellow), 5' exon(cyan), monovalent metal ions (K1= K^{+}, K2= K^{+}, purple sphere) and divalent metal ions (M1 = Ca^{2+}, M2= Ca^{2+}, cyan sphere). A dashed circle highlights the active site considered for the MD study. Close-up showing the active site elements in the pre-hydrolytic state (immediately before hydrolysis, PDB: 4FAQ). The nucleophilic water (red sphere) is placed for the inline attack to the scissile phosphate. The most important RNA nucleotides are shown in line representation.

It is remarkable that despite the large size of group II introns, only ~35 nucleotides of D5 are conserved among different forms of life. It is also interesting that the linker region (three nucleotides) connecting D2 and D3 (J2/3) is also a major player in catalysis and highly conserved. Three highly conserved motifs, i.e., ‘‘catalytic triad’’ of Domain 5 (CYT358, GUA359, CYT360), “two nucleotide bulge” of Domain 5 (ADE376, CYT377), and “Junction between domain 2 and domain 3” (ADE287, GUA288, CYT289), two monovalent ions K1 and K2, and two divalent ions (M1 and M2) form the catalytic pocket. The presence of highly negative phosphates from multiple nucleotides in a congested region leads to strong metal ion binding sites inside the intron core. Scissile phosphate linking U0 (exon-part) and G1 (intron-part) define the exon-intron boundary (**Figure 1.3**). The nucleophilic water (red sphere) close to the scissile phosphate is coordinated by the divalent metal ion at the M2 position (**Figure 1.3**). Catalysis is aided by metal ions in the heteronuclear core, which acts as an important cofactor. Group II intron is unique as the monovalent ions are also crucial (**Figure 1.3**) along with divalent metal ions (M1, M2) for catalysis. K1 and K2 were believed to rigidify the divalent ion binding sites and stabilize critical reaction intermediates by moulding the RNA backbone within the conserved core (**Sigel et al., 2005; Toor et al., 2008; Marcia et al., 2012**). The M1 and M2 divalent ions are firmly anchored in place by the extremely dehydrated K1 ion, which could explain the stability of the intron in the absence of bound oligonucleotide. As a result, K1 is expected to control conformational toggling between the first and second splicing processes. K2, on the other hand, serves a twofold purpose by guaranteeing proper alignment of the 5′-splice junction and stabilizing the post-catalytic state following the initial splicing process (**Marcia et al., 2012**).

1.2.6 Post-hydrolytic and ligand free state

Two key structures of the group II intron have been characterized after splicing, (1) spliced exon part is still linked with intron (post-hydrolytic state) and (2) Exon free group II intron (ligand-free state) (**Figure 1.2)**. The folded structure of group II intron at the post-hydrolytic state is shown in **Figure 1.4 a.** The 5′-exon hydrolysis was stimulated by catalytically active Mg^{2+} ions, which were used to crystallize the post-hydrolytic state. The dehydrated metal ion K1 maintains the distinctive triple helix conformation of active sites.

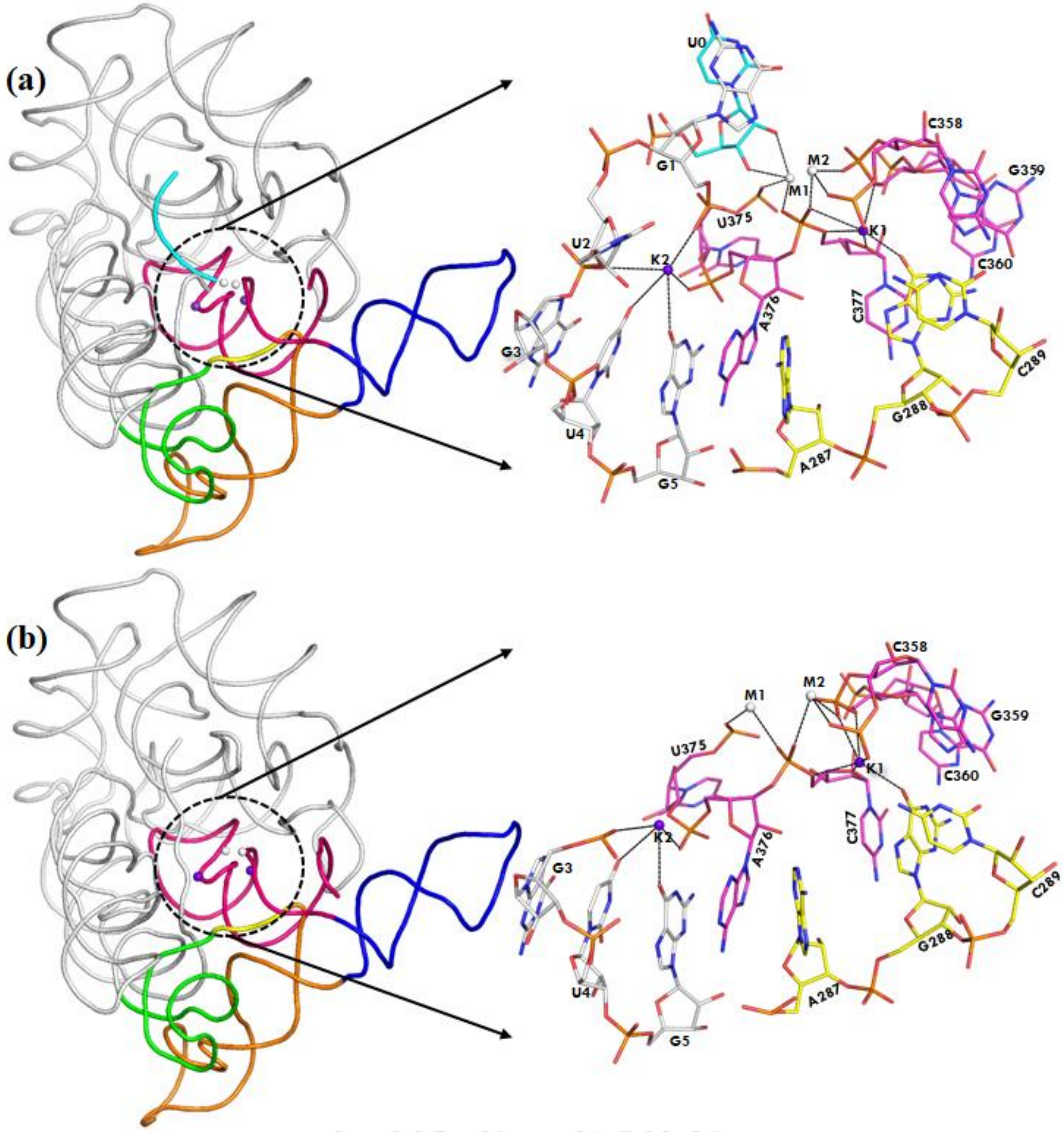


Figure 1.4. 3D architecture of group II intron (a) Cartoon representation of post-hydrolytic state (PDB: 4FAR): D1 (grey), D2 (green), D3 (orange), D4 (blue), D5 (pink), J2/3 (yellow), 5’ exon (cyan), monovalent metal ions K1, K2 (purple sphere) and divalent metal ions M1, M2 (grey sphere). (b) ligand-free state close-up showing the active site elements in the ligand-free state (PDB: 4E8M).

K2, on the other hand, serves a twofold purpose by ensuring that the 5′-splice junction is properly aligned and stabilizing the post-catalytic state. M2 coordinates with three intron phosphates (C358, G359, and C377) in the post-hydrolytic state, whereas M1 interacts with the intron (U375, C377) and spliced exon (2′, and 3′-ribose-OH of U0). Structures of ligand-free and ligand-bound introns are nearly identical, which is common in catalytic RNAs (ribozymes) and protein enzymes that catalyze two-metal ion phosphor-diester cleavage. The interaction network in the post-hydrolytic and ligand-free intron is more or less similar, except for the fact that water molecules occupy the exon position in the latter (**Figure 1.4 b**).

1.2.7 Branched splicing

Recently crystal structures of a eukaryotic group IIB intron (branched lariat state) were reported (**Chan et al., 2018**). Splicing involving branched lariat group II intron is evolutionary conserved and critical for the reliability of 5′ splice-site selection because it is composed of a 2′-5′ phosphodiester link between a bulged adenosine and the 5′ end of the intron. In the brown alga *Pylaiella littoralis (P.li.LSUI2)*, the structures of lariat group II intron was determined in the pre-hydrolytic (PDB:6CHR) and post-hydrolytic states (PDB:6CIH) (**Chan et al., 2018**).

1.2.8 Pre-hydrolytic and Post-hydrolytic state

The experimentally resolved structures (**Chan et al., 2018**) of lariat group II introns (containing six domains D1-D6) was reported for the pre-hydrolytic stage (PDB 6CHR, 2nd structure from the left described schematically in **Figure 1.2 b**) and the post-hydrolytic state (PDB 6CIH, 3rd structure from the left described schematically in **Figure 1.2 b**) (**Chan et al., 2018**). In the pre-hydrolytic state, M1 interacts with the 3’-OH of the 5’-exon, thus positioning the 2’-OH in the correct orientation for the nucleophilic attack to the 3’-spice-site at the intron-exon boundary (**Figure 1.5 a**). In the post-hydrolytic state, ligated exons are non-covalently bound to the lariat II intron (**Figure 1.5 b**). M1 was proposed to form electrostatic interaction with the ligated exon in the post-hydrolytic state (**Figure 1.5 b**) (**Chan et al., 2018**). Monovalent ion (K1) interacts with D5 and D2 in the pre-hydrolytic state.

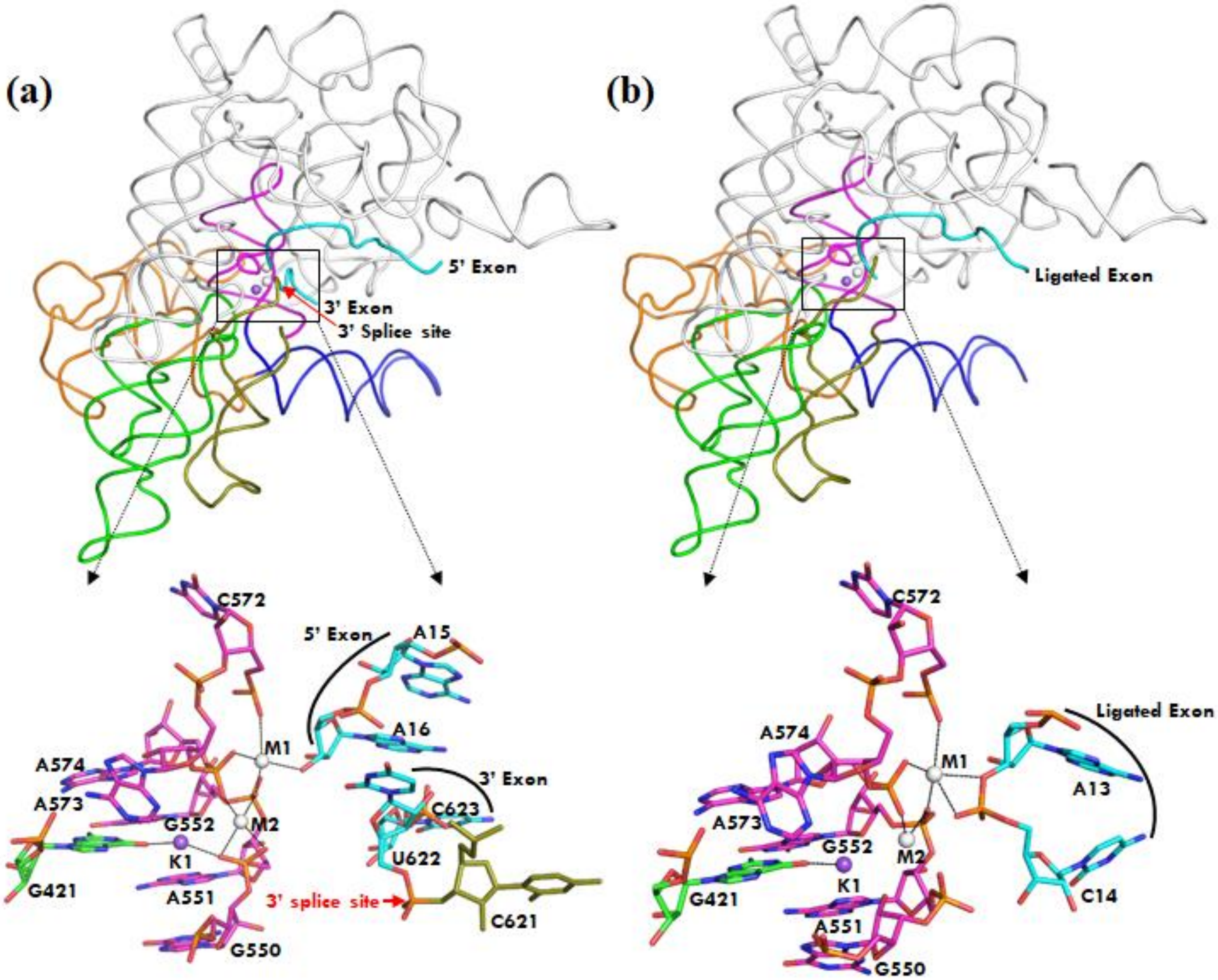


Figure 1.5. Group II intron 3D architecture (a) Pre-hydrolytic state (2s) as seen in a cartoon (PDB: 6CHR): D1 (grey), D2 (green), D3 (orange), D4 (blue), D5 (pink), D6 (olive green), 5' and 3' exons (cyan), and M1, M2 (grey sphere), and K1 (purple sphere). A close-up of the active site constituents in their pre-hydrolytic state is shown in the sticks. (b) Post-hydrolytic state as seen in a cartoon (PDB: 6CIH): D1 (grey), D2 (green), D3 (orange), D4 (blue), D5 (pink), D6 (olive green), 5' exon (cyan), and M1, M2 (grey sphere), and K1 (purple sphere). A close-up of the active site constituents in their post-hydrolytic state is shown in the sticks.

1.2.9 Biochemical study of group II intron

Biochemical experiments of the group II intron of *Oceanobacillius iheyensis* (**Marcia et al., 2012**) revealed that the activity of group II intron depends on the nature of bound divalent and monovalent ions. Maximum splicing activity was observed in metal ion combinations (K^+/Mg^{2+}). Interestingly, the activity was severely compromised in the presence of other metal ion combinations (Cs^+, Na^+, Li^+, Ca^{2+} Sr^{2+}, and Ba^{2+}), which may be due to non-optimal size resulting in poor fitting in the ion binding pocket of the intron (**Marcia et al., 2012**).

1.3 Cas1 protein and its metal ion selectivity

1.3.1 Background

The CRISPR-Cas system is a defense mechanism against invading viruses found in bacteria and archaea (**Van der Oost et al., 2009**). It helps the host to identify, recognize and battle invading pathogens. CRISPR stands for clustered regularly interspaced short palindromic repeats found in the genome of prokaryotes and archaea (**Deveau et al., 2010**). The loci of CRISPR consists of short repeats (30-40 bp long) interspersed with short variable sequences (30-40 bp long) called 'spacers'. Such spacer sequences are homologous with short plasmid and viral genetic sequences (**Van der Oost et al., 2009; Deveau et al., 2010; Horvath et el., 2010**). CRISPR-Cas system synthesizes the RNA from the spacer sequences and, involving other CRISPR associated proteins recognize the foreign invading DNA using the complementarity principle and catalyze the hydrolysis of the foreign DNA, thus, provides adaptive immunity to the host (**Marraffini et al., 2010**). CRISPR-Cas host immunity is a three-stage process: (i) adaptation or spacer acquisition, (ii) crRNA biogenesis, and (iii) interference. In the adaptation stage, spacer sequence will be recognized from the viral DNA and integrated into the host genome by Cas proteins (**Fineran et al., 2012; Savitskaya et al., 2013**), where Cas1 and Cas2 proteins are of crucial importance (**Nuñez et al., 2014**). During crRNA biogenesis, long precursor RNA or pre-crRNA will be transcribed and subsequently processed into small RNA fragments (complementary to viral DNA), known as crRNA. In the final stage of interference, the crRNAs, together with another segment tracrRNA, bind to the specialized protein (viz., Cas9 nuclease protein), forming a crRNA-Cas9 complex and are guided for its degradation towards the target foreign nucleic acid (plasmid or viral genome) (**Swarts et al., 2012; Barrangou et al., 2013**). Cas9 results in the degradation of foreign DNA by binding it to complementary crRNA regions and introducing double-stranded breaks (**Jiang et al., 2017**). Cas1 is one of the universally conserved protein and a common feature of the CRISPR immune system. Cas1 is a metal-dependent endonuclease that can cut foreign DNA and produces small DNA fragments for integration into the host genome. Cas1, together with Cas2 (the other universally conserved CRISPR protein), recognizes, cleave, and integrates the foreign DNA into the host CRISPR loci. Thus, Cas1 is essential for the spacer acquisition/adaptation step (**Wiedenheft et al., 2009**).

1.3.2 Structural and Biochemical studies

The 3-dimensional structure of Cas1 protein from *Pseudomonas aeruginosa* reveals that Cas1 protein is organized into N-terminal and C-terminal domain via linker loop. The N-terminal domain contains 10 β-strand and 2 α-helices, while the C-terminal domain contains 8 α-helices connected with loops (**Wiedenheft et al., 2009**). The metal ion-binding domain is present in the α-helical C-terminal domain, where the universally conserved residues (E190, H254, D265) and a strongly conserved residue at position 268 (D or E) cluster around the manganese (Mn^{2+}, cognate divalent metal ion) **(Figure 1.6)**. Cas1 proteins form an asymmetrical homodimer with an N-terminal domain followed by a helical C-terminal catalytic domain. Experiments confirmed that at physiological KCl concentrations, Cas1 shows the highest nuclease activity on dsDNA with Mn^{2+} than with Mg^{2+}. Moreover, Cas1-mediated ssDNA cleavage is exclusively supported with Mn^{2+} (**Wiedenheft et al., 2009**). Cas1 protein is a metal-dependent, DNA-specific endonuclease that generates 80-base pair double-stranded DNA fragments (**Wiedenheft et al., 2009**). Cas1 mediated DNA degradation activity is severely compromised by mutation of the binding pocket residues as well as metal ion substitution (**Wiedenheft et al., 2009**).

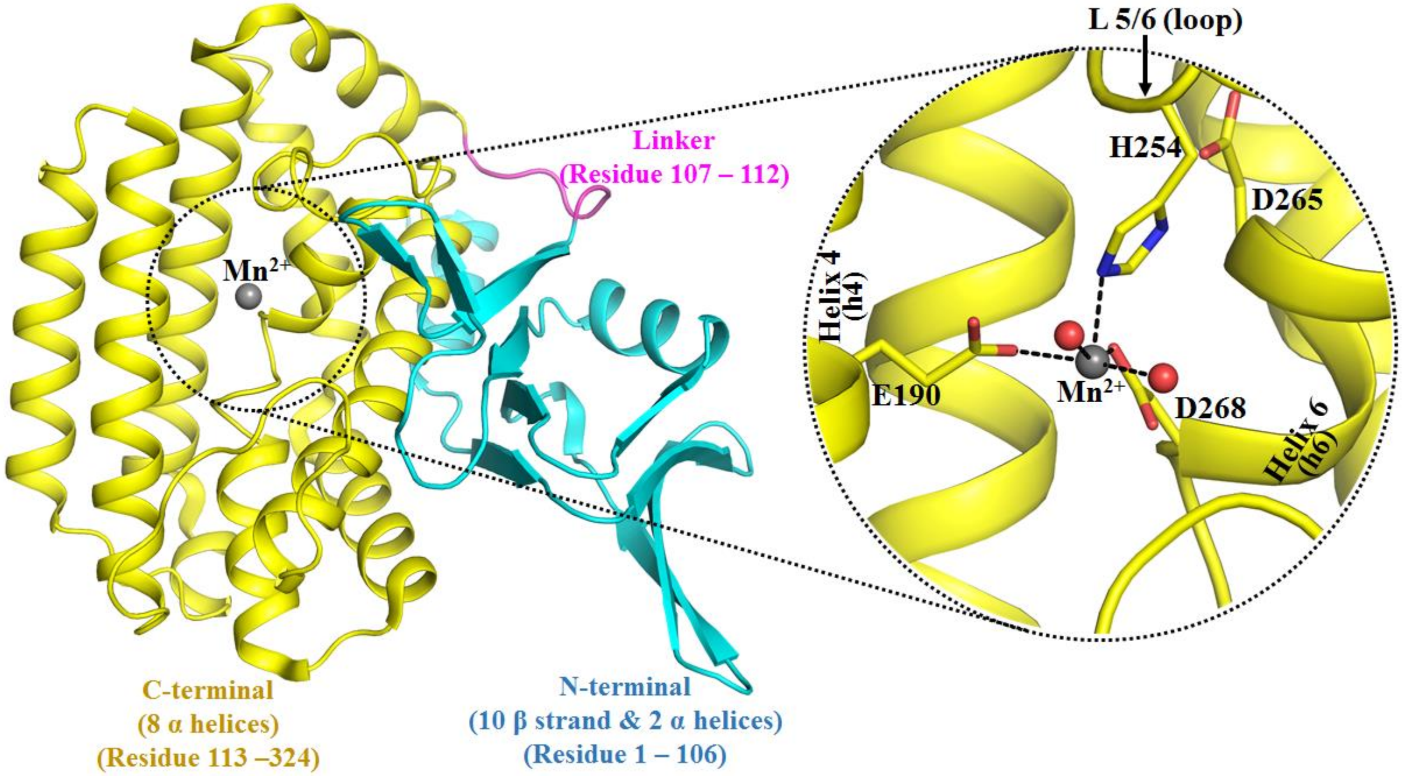


Figure 1.6. Cartoon representation of Cas1 protein (PDB 3GOD). Cas1 is composed of two structurally distinct domains: N-terminal (cyan, β-strand domain, residue 1-106) and C-terminal (yellow, α-helical domain, residue 113-324) domains connected by a linker loop (magenta, residue 107-112). Zoomed-in view of the divalent Mn^{+2} ion (grey sphere) binding pocket (broken circle). Key amino-acid residues of Cas1 [E190 of helix 4, D268 of helix 6, and H254 of L5/6 (loop that connects helix 6 and helix 5)] are highlighted with sticks and crystal waters in the red spheres. Direct interactions with Mn^{+2} are shown as dashed lines.

1.4 Methods

1.4.1 Software and Computational methodology

In this thesis, PyMOL (**DeLano et al., 2002**), UCSF Chimera (**Pettersen et al., 2004**), VMD (**V**isual **M**olecular **D**ynamics) (**Humphrey et al., 1996**) were used for the visualization of biomolecules. CHARMM (**C**hemistry at **HAR**vard **M**olecular **M**echanics) (**Brooks et al., 1983; Brooks et al., 2009**) was used to prepare setup files for performing MD simulations. NAMD

(**NA**noscale **M**olecular **D**ynamics) (**Phillips et al., 2005**) was used to run MD simulations. Gaussian16 **(Frisch et al., 2016)** was used for electronic structure calculations. Plots were generated by the Grace plotting tool (https://plasma-gate.weizmann.ac.il/Grace/). We used our own TCL and Linux bash scripts for analyzing the MD trajectories. The computations were performed at Supercomputing facility IIT Guwahati (Param Ishan) and BSL (Biomolecular Simulation Lab of Department of Biosciences and Bioengineering, IIT Guwahati).

1.4.2 Classical Molecular dynamics Simulations

MD simulation is a popular choice for elucidating dynamics and energetics associated with biological processes, widely used in chemistry, molecular biology, materials science, and drug discovery (**Durrant et al., 2011; Borhani et al., 2012; Dror et al., 2012; Raczyński et al., 2013**). In this thesis, we used experimentally (X-ray, Cryo-EM) characterized structures and templates for performing classical MD simulations. The thermodynamic properties, viz., free-energies, were estimated from the MD trajectories by employing popular free energy perturbation (FEP) and thermodynamic integration (TI) methodologies.

The classical MD simulation method is based on Newton's second law (Equation 1), expressed in its most simplified form as follows:

$$F_i = m_i a_i = m_i \ddot{r}_i \qquad (1)$$

In which F_i is the force acting on particle i. m_i and a_i are the mass and acceleration of particle "*i*". The acceleration $a_i = \ddot{\boldsymbol{r}}_i$, the second derivative of the position (r) of particle "*i*" with respect to time. The gradient of the potential energy (U(r)) with respect to position determine the force F_i (Equation 2).

$$F_i = -\nabla_i U(r) \qquad (2)$$

Combining equation (1) and equation (2) given equation 3.

$$\frac{-dU(r)}{dr_i} = m_i \frac{d^2 r_i}{dt^2} \qquad (3)$$

Verlet integration algorithm (**Verlet, 1967**) is a popular approach for solving Equation 3. In this approach, the positions at two different times are expanded using Taylor expansion about r(t) (Equation 4, 5), where Δt is the time-step:

$$r(t+\Delta t) = r(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 + \frac{1}{6}b(t)\Delta t^3 + \cdots \tag{4}$$

$$r(t-\Delta t) = r(t) - v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 - \frac{1}{6}b(t)\Delta t^3 \cdots \tag{5}$$

Adding equation (4) and (5) gives equation (6)

$$r(t+\Delta t) = 2r(t) - r(t-\Delta t) + a(t)\Delta t^2 \tag{6}$$

Note, velocity is not required (Equation 6) for updating the positions (trajectory). However, velocities are useful for estimating the kinetic energy (hence, total energy), velocity auto-correlation functions etc. Velocities were estimated by subtracting equations (4) and (5), as given in equation 7:

$$v(t) = \frac{1}{2\Delta t}\left[\, r(t+\Delta t) - r(t-\Delta t)\right] \tag{7}$$

The Verlet algorithm has the advantages: (1) position integration is very accurate, and (2) it is independent of the velocity term, which makes it simple and reduces storage requirements. However, the downside of the Verlet algorithm is that it does not include an explicit velocity term in the equation. In the Verlet integration scheme, the errors associated with the estimated position, i.e., r(t) is of the order of Δt^4, as the higher-order term of the Taylor series was ignored. Thus, small value for Δt is required for accuracy but limits sampling. In practice, Δt ~ 1 fs- 2fs is used for performing MD simulations of biomolecules.

A variation of the Verlet algorithm has been developed, known as the leap-frog algorithm (**Hockney, 1970**). A mid-step 'leap frog Verlet' scheme has been proposed (equation 8, 9).

$$r(t+\Delta t) = r(t) + v\left(t+\frac{1}{2}\Delta t\right)\Delta t \tag{8}$$

$$v\left(t+\frac{1}{2}\Delta t\right)=v\left(t-\frac{1}{2}\Delta t\right)+a(t)\Delta t \qquad (9)$$

Thus, position update requires current positions r(t), accelerations a(t), and the mid-step velocities $\boldsymbol{v\left(t-\frac{1}{2}\Delta t\right)}$. In practice, equation 9 is implemented first and then equation 8 estimate the position. The velocities at time t is estimated by using equation 10:

$$v(t)=\frac{1}{2}\left[v\left(t+\frac{1}{2}\Delta t\right)+v\left(t-\frac{1}{2}\Delta t\right)\right] \qquad (10)$$

The velocities are calculated first, and then the positions are calculated using these velocities in this algorithm approach. As a result, the velocities jump over the positions to determine their values. The leap-frog algorithm has the advantage of improving velocity evaluation; however, the downside is that positions and velocities are not synchronized; thus, it is still not satisfactory.

Another popular approach is the Velocity Verlet algorithm, which is the improved variation of the Verlet algorithm. Velocity Verlet algorithm is a precision-preserving algorithm that stores positions, velocities, and acceleration all at the same time (**Swope et al., 1982**) (Equation 11, 12, 13).

$$r(t+\Delta t)=r(t)+v(t)\Delta t+\frac{1}{2}a(t)\Delta t^2 \qquad (11)$$

$$v(t+\Delta t)=v(t)+\frac{1}{2}[a(t)+a(t+\Delta t)]\Delta t \qquad (12)$$

The velocities at time t+Δt is approximated using equation (13).

$$v(t+\Delta t)=v\left(t+\frac{1}{2}\Delta t\right)+\frac{1}{2}a(t+\Delta t)\Delta t \qquad (13)$$

The velocity Verlet method has the advantages of providing a more accurate evaluation of velocities and kinetic energy, as well as being numerically stable. Newton’s Law is deterministic; once the positions and velocities of each atom are known, the system’s state can be predicted at any time in the past or future.

1.4.3 Force Fields

The potential energy function is the key factor for estimating the forces (Equation 2). Force field is a simplified representation in which atoms are represented as a van der Waals sphere having a partial charge, and bonds are defined as springs of finite stiffness. The potential energy of interaction between atoms consists of bonded and nonbonded interactions. The typical form of the potential energy function (U_{total}) is given in Equation 14.

$$U_{total} = \sum_{bonds} K_b\,(b-b_0)^2 + \sum_{angles} K_\theta\,(\theta-\theta_0)^2 + \sum_{dihedrals} K_\chi\,(1+cos(n\chi-\sigma)^2) \qquad (14)$$

$$+\sum_{nonbondedpairs,ij}\left(\varepsilon_{ij}\left[\left(\frac{R_{min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{min,ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_o \varepsilon r_{ij}}\right)$$

The first three terms of equation 14 correspond to bonded interactions (**Figure 1.7**), whereas the last two terms denote non-bonded interactions. The first term accounts for energy associated with the bonds, where k_b is the spring constant and (b-b_0) is the net displacement from equilibrium bond-length (b_0). Energy cost associated with deviation from equilibrium angle θ_0 is considered in the second term, where k_θ is the force constant and ($\theta-\theta_0$) is the angular deviation. The third term defines a torsional potential that models the dihedral rotation, where K_χ = force constant, χ = dihedral angle, n = periodicity, σ = phase. Force field parameters correspond to bond, angle and dihedral terms are K_b, b_0, K_θ, θ_0, K_χ, χ, n, σ. The fourth and fifth term describe the non-covalent interactions. The fourth term is the interaction potential (Lennard-Jones potential: LJ) that mimic the interaction between neutral atoms, consist of attractive r^{-6} term (Van der Waals term) and repulsive r^{-12} term (prevents atoms from pervading at very short distances) between the atom-pair “i” and “j”. The firth term (Coulomb interaction potential) describes the interaction between charged atoms (repulsive/attractive depending on whether the sign of partial charges q_i and q_j) where R_{min}, ε, q are the force-field parameters. Force-field parameters are obtained from experiments or quantum chemical calculations and kept constant in the classical force fields. In this thesis, we used the CHARMM force field **(MacKerell Jr et al., 1995; MacKerell Jr et al., 1998)**.

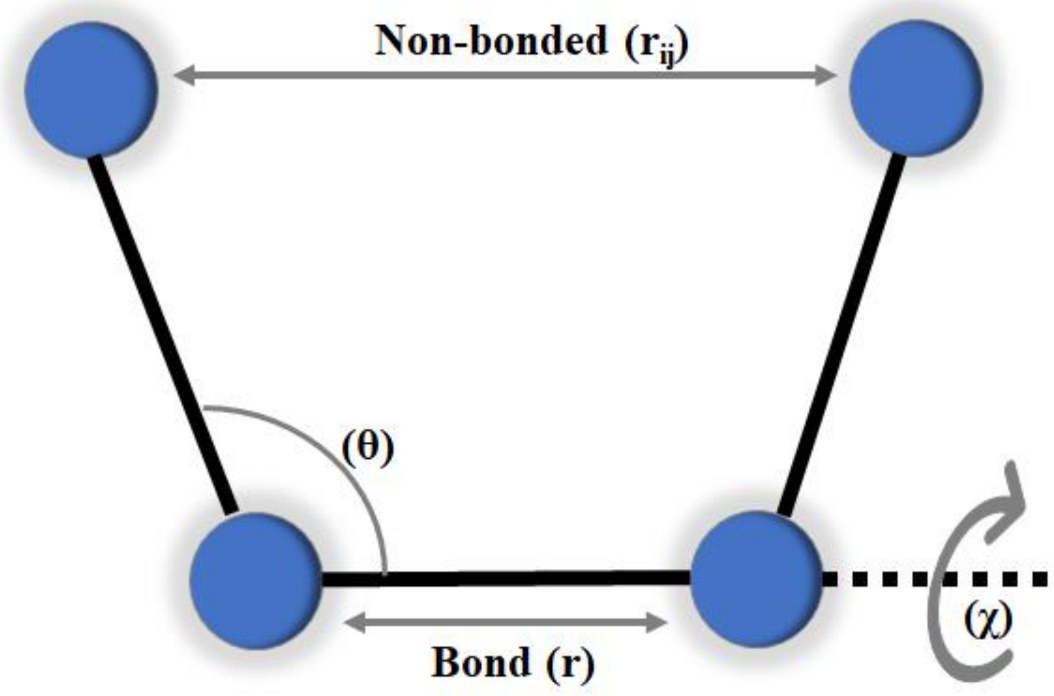


Figure 1.7: Schematic representation of bond, angle, dihedral, and non-bonded terms.

1.4.4 Solvation and water model

In this thesis work, we overlaid a box of equilibrated water and kept the biomolecule at the centre. The popular TIP3P water model (**Durell et al., 1994**) was used for performing classical MD simulations in an explicit water box. This simple, rigid, and non-polarizable three-site model remains the most commonly used water model in MD simulations of RNA/DNA/protein. The simplest water models treat the water molecule as rigid by constraining the bond lengths using the SHAKE algorithm (**Ryckaert et al., 1977**). The solvated simulation system is neutralized by adding counter-ion.

1.4.5 Minimization and production dynamics

Structures of solvated electrically neutral biomolecular systems are typically subjected to energy minimization before running MD simulations. The objective of the minimization is to remove strains (steric overlaps) from the initial template structure and approach to the local minima in the energy landscape. We used experimentally resolved structures (PDB) or homology modelled structures for the MD simulation model. First, the biomolecule is spherically truncated, and our region of the truncated sphere (buffer region) was harmonically retrained related to its template structures. The missing residues or atoms (if any) were built, and hydrogen atoms were added. The spherically truncated model was then solvated with a box of explicit water and subjected to minimization using popular energy minimization techniques such as steepest descents (**Lee et al., 2002; Lee et al., 2003**) and/or adopted basis Newton-Raphson (**Chu et al., 2003**) methods. The energy minimized simulation box was then subjected to unrestrained (except buffer region) MD simulation resulting in production MD trajectories.

1.4.6 Temperature and Pressure Control

Langevin piston (**Feller et al., 1995**) is used to control the temperature in MD simulations. Momenta of all atoms in the system are altered to keep the temperature at a specified value. The Nose-Hoover algorithm (**Martyna et al., 1994**) was used to regulate pressure. The volume of the simulation box size is regulated to keep the pressure constant. The instantaneous pressure of a biomolecular system fluctuates frequently; however, the average pressure of all particles in the system represents the total pressure of the system with less fluctuations (**Hoover et. al., 1982; Nosé, 1984; Hoover, 1985**).

1.4.7 Truncation of short-range Van der Waals Interactions

Van der Waals interactions are short-range decay fast ~ r^{-6} with distance. As a result, the cut-off for LJ potential is the most used choice for MD simulations to save resources and time. However, in order to avoid abrupt truncation of the LJ energy, MD simulations frequently use a switching function that truncates the potential energy smoothly to truncate the LJ interactions progressively as a function of distance. We have used 12Å-16Å as cut-off where switching function was used for the last 2Å.

1.4.8 Periodic Boundary Condition

Periodic boundary conditions were employed in our MD simulations to remove surface effects and include bulk effect. Thus, the simulation box was periodic and replicated in all special directions. **Figure 1.8** depicts periodic boundary conditions in two dimensions (for simplicity), where each box is surrounded by eight neighbors. If a particle moves outside the central cell during the simulation, the same is introduced from the neighboring, ensuring that the number of particles in the central cell remains constant (**Figure 1.8**).

1.4.9 Long-range Electrostatic Interaction

The most expensive aspect of a MD simulation is calculating long-range interactions (decay as r^{-1}) Coulomb interaction energy. Thus, electrostatic interactions need to be computed without truncation. We used the particle mesh Ewald (PME) method (**Darden et al., 1993**) for estimating the electrostatic interactions that require periodic boundary conditions. PME estimated electrostatics consist of two parts: (1) short-range real space component (2) Long-range Fourier component, which requires periodicity and charge neutrality.

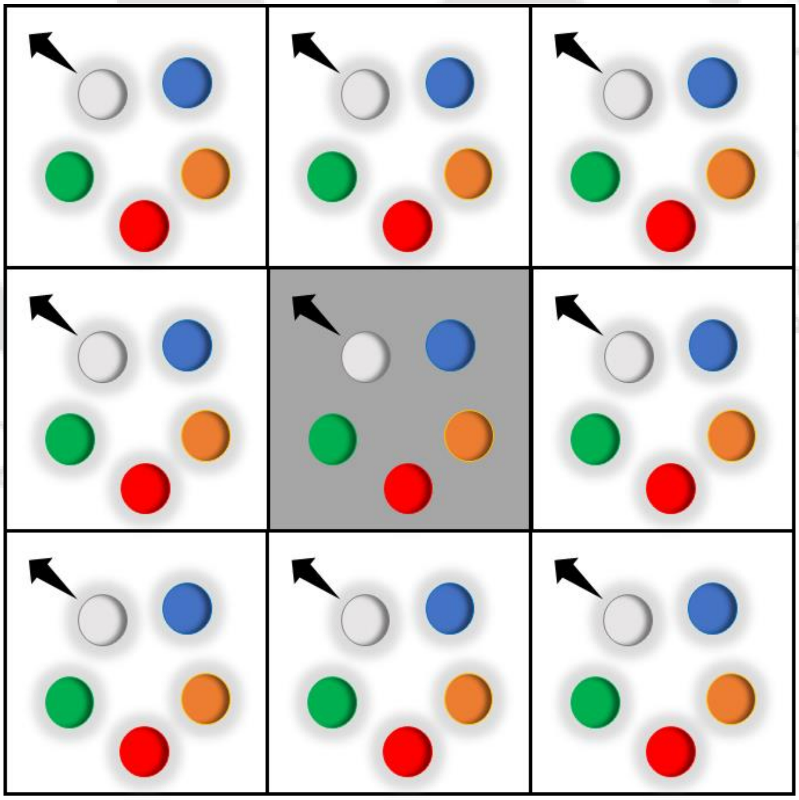


Figure 1.8. Schematic representation of Periodic boundary conditions in two-dimension. The center box (gray) is shown along with its first periodic images (white).

1.4.10 MD setup adopted in this thesis

The overview of our MD setup is described in **Figure 1.9**. The template structure of the biomolecule was either taken from the Protein Data Bank or from homology modelling. Typically, 25Å radii sphere, centered at the region of interest, was cut from the selected biomolecular structures and considered for MD simulations (**Figure 1.9**). Non-hydrogen atoms of the biomolecule in the outer region between 22Å-25Å from the sphere's center ("buffer region") were harmonically restrained to their experimentally determined positions. The restraints in the buffer region were increased gradually from 3.0 to 5.0 kcal/mol/Å^2 as one moves closer to the outer boundary. The inner 22Å radius shell was fully flexible in our MD. A cubic water box (edge length= 80Å) was overlaid. The waters that overlapped with the biomolecule were removed. The calculations were repeated by considering larger truncated biomolecular sphere into a larger water-box to ensure that the calculations were size-independent. Periodic boundary conditions using the Particle Mesh Ewald method (**Darden et al., 1993**) for long-range electrostatics were used for calculating full electrostatics. The overall charge of the simulation system was neutralized by scaling down the partial charges of the biomolecule or adding counter ions. Van der Waals interaction was truncated at 16Å cutoff distance. Temperature and pressure were maintained at 300K/310K and 1 bar. Temperature and pressure were controlled by using Langevin dynamics for non-hydrogen atoms with a coupling coefficient of 5 ps^{-1} and Langevin piston using the Nose-Hoover method, respectively. The CHARMM36 force field (**Mackerell et al., 1998; MacKerell Jr. et al., 1995**) with TIP3P water model (**Jorgensen et al., 1983**) was used. CHARMM (**Brooks et al., 1983; Brooks et al., 2009**) and NAMD (**Phillips et al., 2005**) programs were used to run MD. The system was initially equilibrated by a series of short MD runs with a total length of ~300 ps. The system was heated up to 310K in the first phase of equilibration and then kept fixed throughout the production MD trajectories. At the initial stage of equilibration, harmonic restraint was applied to the heavy atoms of the inner region (within 22Å) at their experimentally resolved positions. At the final stage of equilibration, the restraints were completely removed from the inner region (within 22Å). Root mean square deviations (RMSD) between MD structures from production dynamics and template structure was estimated to measure structural deviation. Plateau in the RMSD vs. time plot indicates structural convergence. The trajectory segment, which corresponds to structural convergence, is used to estimate root mean square fluctuation (RMSF), that measure the flexibility at different region of the biomolecule. RMSD or RMSF is estimated only for the heavy atoms of the biomolecule except for the buffer region.

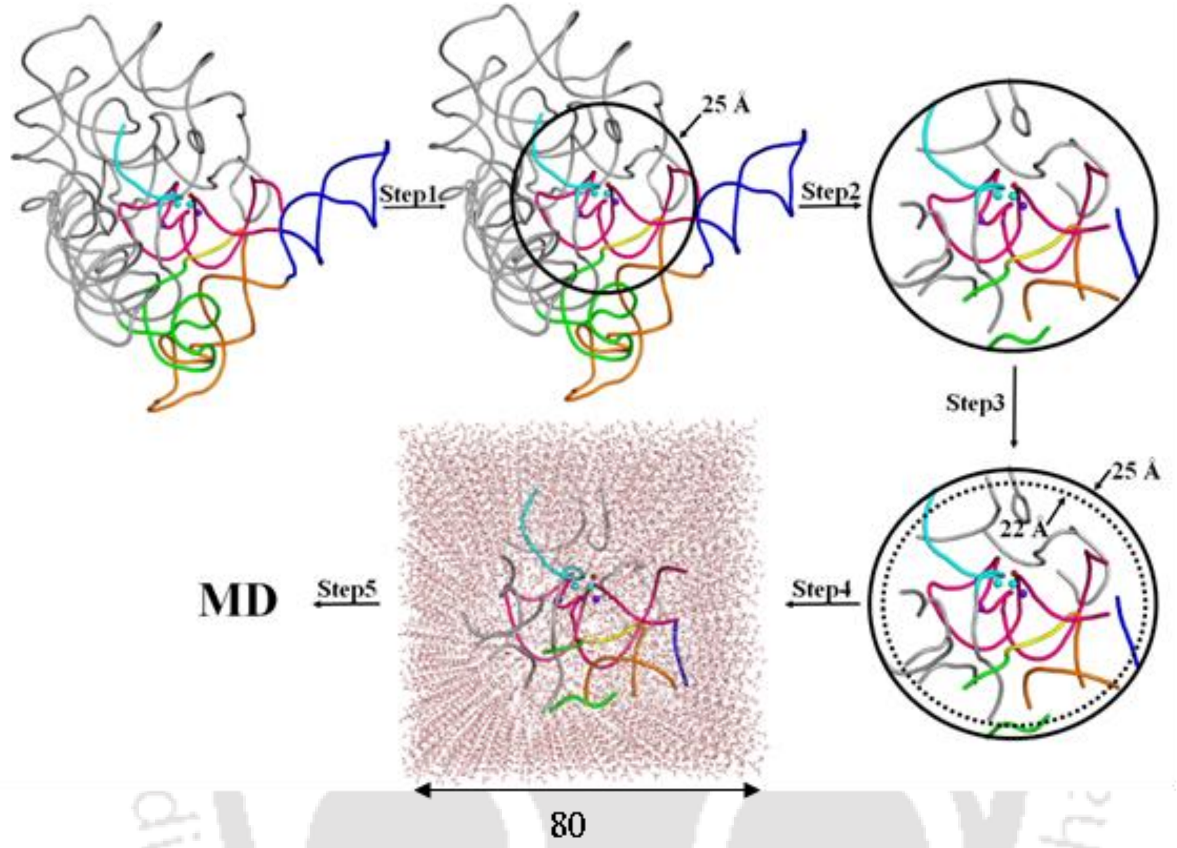


Figure 1.9. Typical MD setup. Step1: Selection of 25 Å radii sphere centered at P atom of cytosine 377 (PDB: 4FAQ). Step 2: Truncated model (25 Å sphere) for molecular dynamics simulation. Step 3: Heavy atoms of the "buffer region" (22Å-25Å) are harmonically restrained to their experimentally determined positions. Step 4: Overlay water box of edge length 80Å, minimize, equilibrate, and proceed to MD.

1.4.11 Thermodynamic Cycle and Relative Binding Energy

Binding free energy differences of two different cations binding to a biomolecule (RNA/protein) were estimated by using an appropriate thermodynamic cycle (**Figure 1.10)**. The vertical legs correspond to the real binding event (L or L' binding to the biomolecule). In contrast, horizontal legs correspond to the alchemical transformation of L → L', either in complex with the biomolecule (upper horizontal leg) or free in water (lower horizontal leg). Note, the horizontal legs cannot be realized experimentally.

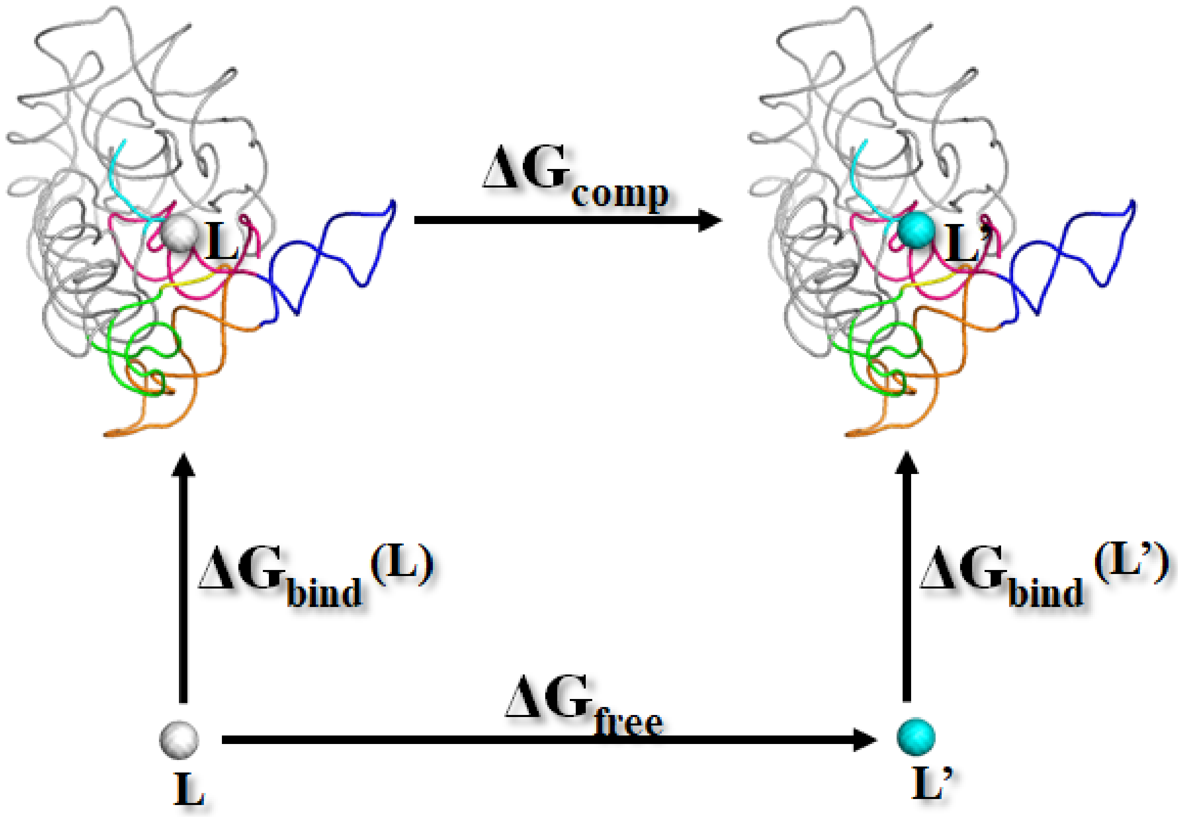


Figure 1.10: Thermodynamic cycle and relative binding free energy. Vertical legs correspond to the binding, horizontal legs correspond to the alchemical transformation (L → L') either in the complex (upper arm) or free in water (lower arm).

We compute the free energy change associated with the horizontal arms (ΔG_{comp}, ΔG_{free} of **Figure 1.10**) and estimate the relative binding free energy as,

$$\Delta\Delta G = \Delta G_{comp} - \Delta G_{free} = \Delta G_{bind}(L') - \Delta G_{bind}(L) \qquad (15)$$

A hybrid energy function (U) with a coupling coordinate λ was used to represent a mixture of two endpoint states (U = λ U(L) + (1-λ) U(L')) for a particular horizontal leg (**Figure 1.10**). **McCammon, 1991; Mordasini et al., 2000**). The coupling coordinate λ connects the two endpoints (L and L'). Coupling coordinates, λ was used to modify the van der Waals energy terms

(L and L' only differ by size) and alchemically transform L into L' (as λ goes from 1 to 0). Free energy derivative was calculated as $\partial G/\partial\lambda = \langle \partial U/ \partial\lambda \rangle_\lambda$, where brackets "< >" represents averaging over MD trajectory for a particular value of λ (**Kirkwood, 1935**). We used equally spaced λ values between 1 and 0 (1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, and 0.0). The free energy derivative $\langle \partial U/ \partial\lambda \rangle_\lambda$ at each window was computed from a finite-difference estimate. Each λ window lasted for 2-5 nanoseconds. The estimated derivative versus time plot was numerically integrated to estimate free energy. Each trajectory segment is divided into two equal batches. and the uncertainties of the free energy derivatives at each λ value were estimated by taking the difference between the batch averages, and the same is reported as the statistical error. For each biomolecule, we performed multiple runs and the average over those runs. Free energy calculation for each replica was based on 22-55 ns of data collection averaged over 5-15 replicas with different initial velocities. The sign of ΔΔG indicate which ligand is favorable (ΔΔG = +ve indicate L is preferred), whereas the magnitude indicates the strength of preference.

1.4.12 Electronic Structure calculations

Classical Force-field does not include explicit polarizability effect; thus, estimation of relative binding free energies (ΔΔG) using classical MD simulations may not be appropriate, especially for divalent ions for which many-body effect is known to be crucial. Thus, in order to check if the sign of the estimated ΔΔG is correct, we estimated energies using ab initio calculations of the reduced metal-ion binding pocket (which explicitly includes polarization and approximately captures the essence of the relative free energy of binding to some extent). The objective was not to estimate highly accurate ΔΔG from ab initio calculations but to focus on a quantity that can indicate divalent-metal-ion preference in a computationally inexpensive way. The reduced binding pocket model (divalent metal ion and key residues around the ion) was embedded in an implicit aqueous dielectric (SMD solvent model) and was to density functional energy calculation in two different levels of theory: (1) the M06-2X/6-311++G** (**Zhao et al., 2008**) and (2) B3LYP/6-31+G* (**Becke 1988; Lee et al., 1988**). Gaussian16 program (**Frisch et al., 2016**) was used for the ab initio calculations.

1.3 Objectives

1. Estimation of the energetics (ΔΔG or recognition) of monovalent (K^+ vs. Na^+) and divalent (Mg^{2+} vs. Ca^{2+}) metal ion selectivity in Cas1 protein and at various stages of Group II intron.

2. Establish the link between the estimated energetics and the structures of correct and incorrect metal ion bound biomolecule (Group II intron/Cas1 protein).

3. Estimate the effect of mutation of the metal ion binding pocket on the metal-ion recognition in Cas1 protein.

4. Compare the structure and dynamics of cognate and near-cognate metal-ion complexes of the biomolecule (Group II intron/Cas1 protein).

5. Understand the role of polarizability in divalent metal-ion (Mg^{2+}, Ca^{2+}, Mn^{2+}) selectivity using quantum chemical calculations.

Chapter 2

Monovalent metal ion (K^+ vs. Na^+) selectivity in Group II intron

Group II introns are ribozymes which can catalyze its own splicing and relegate itself. They share common structural and are evolutionarily related to eukaryotic spliceosome. Hence, group II introns are an excellent model system in understanding the mechanism of RNA-splicing in gene expression. Recent advancement of structural studies has provided X-ray structures of group II intron at different stages (Pre-hydrolytic, post-hydrolytic and free intron) of splicing pathway and revealed heteronuclear metal ion cluster (two potassium; K1 and K2, two magnesium; M1 and M2) in the active site as common structural feature of group II introns. It is believed that these four metal ions are very crucial for catalysis. K^+ is very important for the function and buffers containing only Na^+ can destroy the function of group II introns. Despite the available 3D structures of different stages of self-splicing, the energetic origin of K^+ selectivity over Na^+ is not known. However, these structures provide sufficiently good models for directly computing the energetics of K^+ vs Na^+ discrimination that is required for understanding self-splicing. Here, we report extensive (~ 550 ns) of structure-based molecular dynamics free energy simulations that quantitatively calculated K1 vs Na1 discrimination. We show that the strength of discrimination varies along the self-splicing pathway. The free and post-hydrolytic states are highly selective for K1 and the selectivity is lost in the post-hydrolytic state. The simulations further reveal that Na1 in the free and pre-hydrolytic state is trapped with an unsatisfied first coordination shell in the active site, which is responsible for large discrimination. The Na1 in the active site of post-hydrolytic state allow water entry and satisfy the Na1 bonding requirement, resulting in low discrimination. The results not only give insights into the experimentally unresolved Na1 bound complexes but also provide the link between structures and computed energetics.

2.1 Background

RNA transcribed from the DNA often contains both the protein coding regions (exons) and protein non-coding regions (introns) (**Clancy, 2008**). Removal of introns is essential for the production of mature mRNA leading to desired protein synthesis in cells. The process by which introns are removed is called splicing. Splicing is a fundamental process of gene expression and directly related to the fitness of the organism. Introns could be divided into 4 broad categories based on their distinct mechanism of splicing i.e, tRNA, nuclear mRNA, group I and group II introns (**Clancy, 2008**). tRNA and nuclear mRNA introns are spliced out by protein enzymes (protein kinase) and ribonuclear protein complexes (spliceosome) respectively. Group I and group II introns are special as the folded intron itself can catalyze splicing. The splicing process is chemically similar to group II introns and spliceosome (**Pyle et al., 2006**), but the former undergoes splicing in the absence of any additional proteins. It is remarkable that spliced out free group II intron can readily integrate into new genomic locations of DNA and RNA leading to diversification of genome (**Dai et al., 2002**). Group II introns are a very good model system for understanding the role of introns in gene expression, application in biotechnology (**García-Rodríguez et al., 2011**), structure, chemistry and evolutionary perspective in all domains of life (**Mattick, 1994**).

As discussed in **Chapter 1**, Biochemical and structural studies of *Oceanobacillus iheyensis* group II introns (**Pyle, 2006; Toor et al., 2008; and Marcia et al., 2012**) has given fundamental insight into the mechanism of RNA splicing. Group II intron splicing is a two-step process; Step 1: nucleophilic cleavage of the 5’ exon and liberation of free 5’ exon, Step 2: 3’-OH of the 5’ exon terminus attacks and connects the 3’ splice site. Group II introns consist of 6 domains (D1-D6). However, in order to determine the structure at different stages of the first step of splicing, Marcia *et. al* (**Marcia et al., 2012**) studied construct of *O. iheyensis* group II introns (containing only D1-D5) with and without short 5’ exon (as ligand). All the constructs were fully active and important in understanding the mechanism of self-splicing. Near-atomic resolution (~3Å) X-ray structures have revealed (**Marcia et al., 2012**) intricate interaction network that encloses the catalytic pocket within its center.

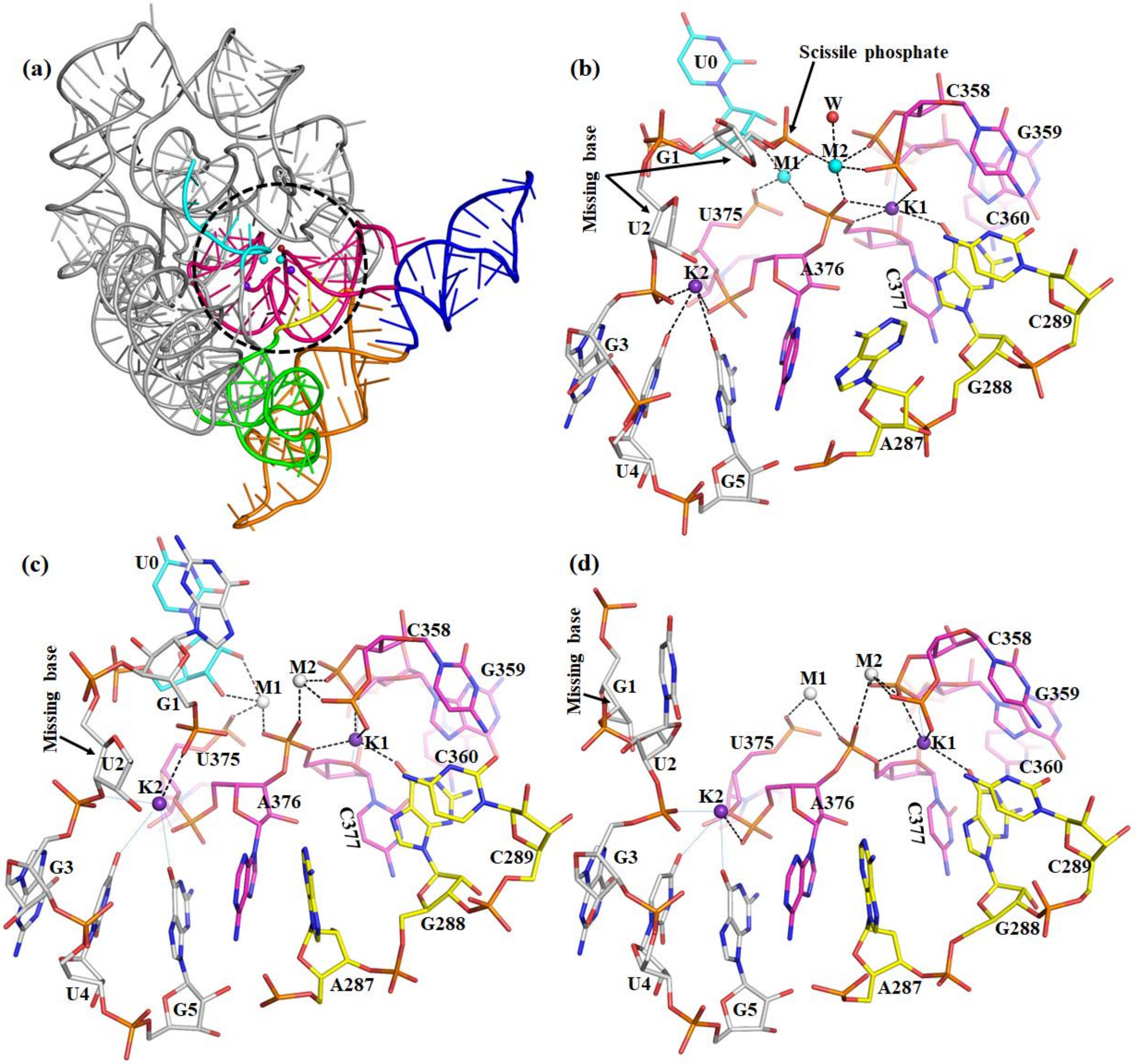


Figure 2.1. 3D architecture of group II intron (a) Cartoon representation of pre-hydrolytic state (PDB 4FAQ): D1 (grey), D2 (green), D3 (orange), D4 (blue), D5 (pink), J2/3 (yellow), 5’ exon (cyan), monovalent metal ions K1, K2 (purple sphere) and divalent metal ions M1, M2 (cyan sphere). Active site considered for MD study is highlighted by a dashed circle. (b) Close up showing the active site elements in pre-hydrolytic state (immediately before hydrolysis, PDB 4FAQ). The nucleophilic water (red sphere) is placed for inline attack to the scissile phosphate. Most important RNA nucleotides are shown in line representation. (c) post hydrolytic state, immediately after hydrolysis. 5’ exon is cleaved and free phosphate interacts with K2. (d) Free intron without the 5’ exon.

Three major conformations along the first step of splicing pathway have been structurally resolved i.e, pre-hydrolytic (the stage prior to hydrolysis, PDB 4FAQ), post-hydrolytic (stage immediately after hydrolysis and with bound 5’ exon, PDB 4FAR) and free intron (PDB 4E8M) state (**Marcia et al., 2012**). The folded structure of group II intron in its pre-hydrolytic state is shown in **Figure 2.1 a**. D1 is the largest domain onto which other domains dock. D5 is embedded into the catalytic core and most important for catalysis. It is remarkable that despite the large size of group II introns only ~35 nucleotides of D5 is conserved among different forms of life.

It is also interesting that the linker region (three nucleotides) connecting D2 and D3 (J2/3) is also a major player in catalysis and highly conserved (**Figure 2.1**). Three highly conserved motifs, i.e, ‘‘catalytic triad’’ of D5 (CYT358, GUA359, CYT360), “two-nucleotide bulge” of D5 (ADE376, CYT377) and “J2/3 junction” (ADE287, GUA288, CYT289) form the catalytic pocket (**Figure 2.1 b-d**). Presence of highly negative phosphates from multiple nucleotides in a congested region leads to the creation of a strong metal ion binding sites inside the intron core. Experimental studies (**Sigel et al., 2000; Gordon et al., 2001; Kruschel et al., 2008; Swisher et al., 2002; Su et al., 2003; Su et al., 2005; Erat et al., 2007; Podar et al., 1995; Basu et al., 1998; Conn et al., 2002; and Lambert et al., 2009**) shown many metal ions e.g, Mg^{2+} and K^+ ions facilitate folding and catalysis of group II introns. Four metal ions consisting two divalent metal ions (M1, M2) and two monovalent metal ions (K1, K2) are very crucial for structural stability and function of the group II intron. Monovalent ions are known to dictate the splicing reaction pathway (**Jarrell et al., 1998; Erat et al., 2008**) and modulating alternative conformations of the group II introns (**Marcia et al., 2012**). The pre-hydrolytic state (**Figure 2.1 a, b**) was crystallized with Ca^{2+} which is known to inhibit hydrolysis (**Marcia et al., 2012; Erat et al., 2008; Viadiu et al., 1998**). The phosphodiester linkage (scissile phosphate at the 5′-splice site) is activated by the M1-M2-K1 metal cluster, where M2 coordinated nucleophilic water (w) is correctly positioned for inline attack (**Figure 2.1 a, b**). The K1 site is specifically important as it has been suggested (**Marcia et al., 2012**) to have a direct role in stabilizing the M2 binding pocket. It has been concluded that monovalent ion binding sites are very selective for K^+ under physiological conditions.

The following questions related to monovalent ion binding remains unanswered: (a) How strongly K^+ is preferred with respect to other monovalent ion say Na^+ at different stages of splicing? (b) How is discrimination strength related to the 3D structures of the K^+ and Na^+ bound complexes?

Medium resolution (~3Å) X-ray structures (**Marcia et al., 2012**) at different stages of splicing, now provide sufficiently good models for structure-based computational studies, and here we report extensive molecular dynamics free energy simulations based on these X-ray structures to decipher the energetics of monovalent ion selectivity (K1 vs Na1), thereby providing the link between the structure and energetics. These calculations involve computing the change in binding affinity for K1/Na1 to the intron active site upon K1→ Na1 mutation, using the appropriate thermodynamic cycle described in **Figure 2.3**.

2.2 Methods

2.2.1 Molecular dynamics procedure

Structure of *Oceanobacillus iheyensis* group IIC intron structure in the free, pre-catalytic and post-catalytic state was taken from the Protein Data Bank entry 4E8M (resolution 3.5 Å), 4FAQ (resolution 3.11Å) and 4FAR (resolution 2.86 Å) respectively (**Marcia et al., 2012**). Total ~48100 number of atoms are considered in our simulations. The total number of water molecules were ~14700. Our molecular dynamics setup is described in **Figure 1.11** (**see Chapter 1; method section for details**). Root-mean-square deviation (RMSD) of the heavy atoms (within 22Å of simulation sphere) of the introns with respect to the X-ray structures (PDB: 4E8M, 4FAQ and 4FAR) are given in **appendix Figure 2.1**.

Average RMSD was calculated for the heavy atoms, within 22Å of the simulation sphere, averaging over the 2-5 ns MD trajectory with 2ps interval. MD structures are very similar to the X-ray structures (**see appendix Figure 2.1**). For each replica we performed ~100-122 ns of molecular dynamics run. The overall charge of the simulation was neutralized by scaling down the partial charges of the phosphate backbone of RNA.

Reported X-ray structures were partly disordered (see missing bases of **Figure 2.1**) and few bases were missing in the PDB. Nucleotides which were well within the 25Å truncated model were build and considered for MD simulation. The orientation of ADE287 in the pre-hydrolytic state (PDB 4FAQ, **Figure 2.1 b**) is different (Watson-Crick edge is rotated by ~ 90° away from the K1 binding site) from the post-catalytic (PDB 4FAR, **Figure 2.1 c**) and free intron (PDB 4E8M) structure

(**Figure 2.1 d**). The beta factor of ADE287 base was also large ~240 in the X-ray structure 4FAQ (**Figure 2.1 b**). MD simulations of pre-catalytic state starting with the template PDB 4FAQ (**Figure 2.1 b**) show structural distortion and scissile phosphate moved away from its crystallographic position (**Figure 2.2**). Then we replaced the ADE287 coordinates of pre-catalytic state with the coordinates of the same in the free state. The resulting pre-catalytic state show almost identical interaction pattern (as seen in the X-ray) in our MD (**Figure 2.2**).

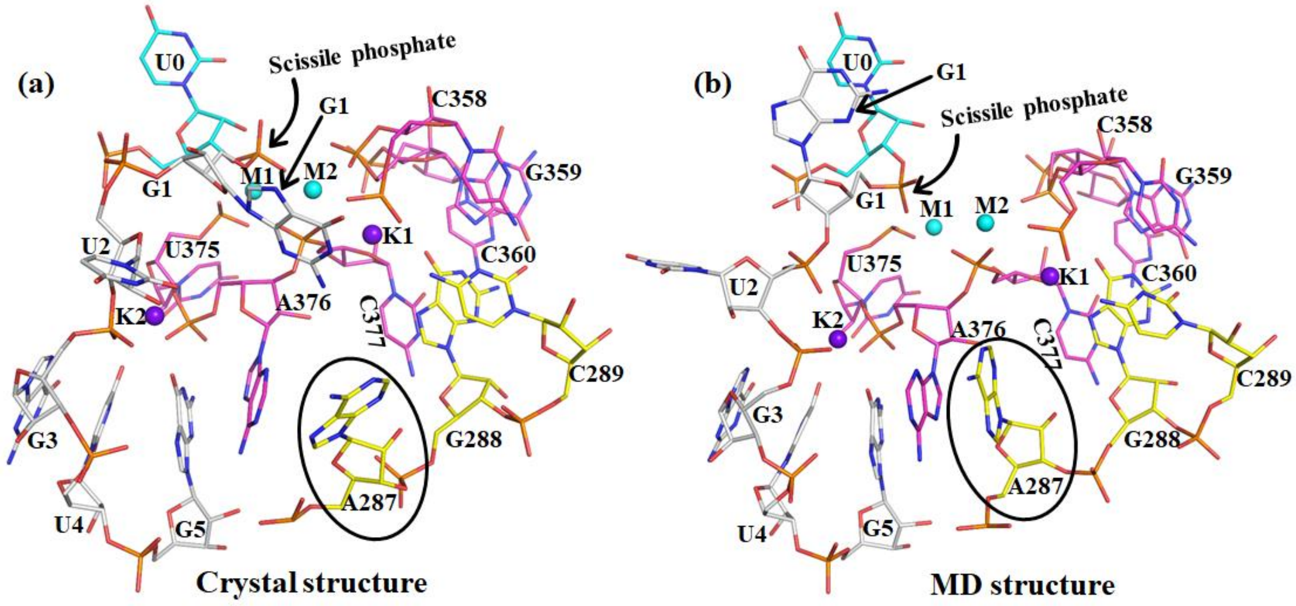


Figure 2.2. Base of ADE287 was poorly resolved (Average beta factor ~240) in the pre-hydrolytic intron state, PDB 4FAQ. (a) Active site in the X-ray structure. (b) Active site in the MD. Structural distortion observed, highlighted with arrow and circle.

2.2.2 Free Energy Calculations

Relative binding free energies (ΔΔG) of $K1^+$/$Na1^+$ binding to intron were calculated by alchemically transforming a K^+ into Na^+ following the horizontal legs of the thermodynamic cycle in **Figure 2.3**. The vertical legs correspond to K^+/Na^+ binding. On the other hand, horizontal legs correspond to the alchemical transformation of $K^+ \rightarrow Na^+$ which cannot be realized experimentally. We computed the free energy change associated with the change in van der Waals radii (horizontal arms of **Figure 2.3**) and calculated the relative binding free energy as, $\Delta\Delta G_{bind} = \Delta G_{comp} - \Delta G_{free}$

= $\Delta G_{bind}(Na^+) - \Delta G_{bind}(K^+)$. Each λ window lasted for 2-5 nanoseconds and the data from last 1ns of each simulation was used for averaging. Free energy change was calculated using numerical integration method. Free energy calculation for each replica was based on 22-55 ns of data collection averaged over 5-6 replicas with different initial velocities. Excellent agreement between different MD runs was obtained (**see appendix Table 2.1**).

Overall a total of about 0.55 µs of molecular dynamics free energy simulations have been done to get good convergence and reasonable statistical error (1-2 kcal/mol), comparable to the earlier reported force field uncertainty (**Satpati et al., 2011**). The λ vs free energy derivatives are given in **appendix Table 2.2**.

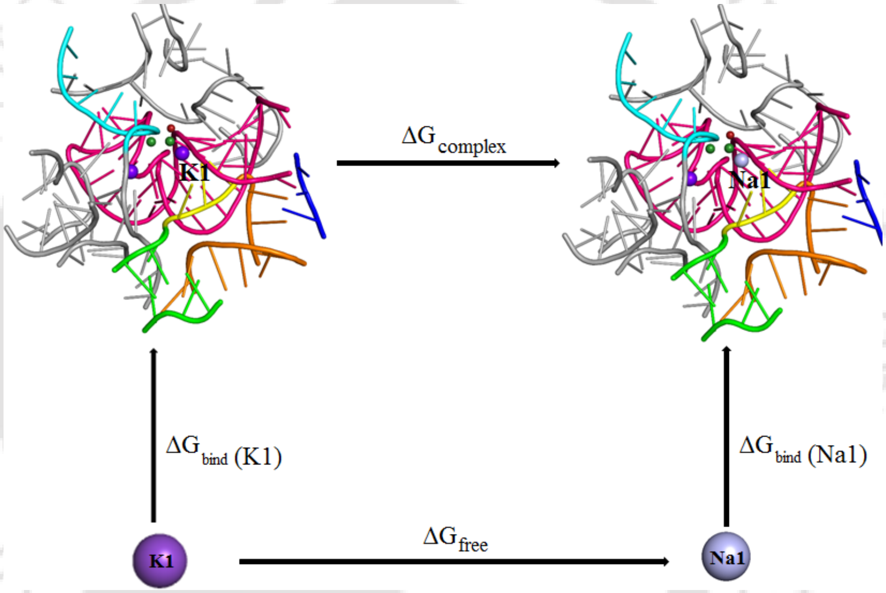


Figure 2.3. Thermodynamic cycle for K^+/Na^+ ion binding to the K1 binding site of intron. Vertical legs correspond to ion binding; horizontal legs correspond to the alchemical transformation (mutation) of K^+ into Na^+ in the K1 binding site, either in the intron bound (above) or free (below) in solution. The horizontal legs of the thermodynamic cycle have been computed by MD simulations. The binding free energy difference is $\Delta\Delta G_{bind} = \Delta G_{comp} - \Delta G_{free} = \Delta G_{bind}(Na1) - \Delta G_{bind}(K1)$.

2.3 Results

The results are organized as follows. First, the relative binding free energies of K^+/Na^+ at different stages (Ligand free, Pre-hydrolytic and Post-hydrolytic intron) of self-splicing group II intron are presented. Second, the structural insights from MD simulations of K^+/Na^+ in water and in complex with group II introns are discussed and compared with the template X-ray structures. Finally, the link between the relative binding free energies and structures is discussed and concluded.

2.3.1 Structure-Based Energetics for K^+ vs Na^+ Recognition

Previously described (**Satpati et al., 2014; Åqvist et al., 2000**) molecular dynamics free energy (MDFE) calculations were performed utilizing the X-ray structures of the *Oceanobacillus iheyensis* group II intron at different stages of splicing e.g, Free, pre-hydrolytic, and post-hydrolytic states (PDB accession numbers 4E8M, 4FAQ, and 4FAR respectively). To evaluate the relative binding free energies, we performed $K^+ \rightarrow Na^+$ mutations in the active site of group II intron and free in the water. The relative binding free affinities were evaluated by calculating the free energy change along the horizontal arms of the thermodynamic cycle described in **Figure 2.3**. K^+ vs Na^+ binding free energy differences to the active site of Group II intron at different stages of splicing are summarized in **Figure 2.4**. Group II intron clearly imposes a very high energetic penalty of about 6-7.5 kcal/mol for Na^+ binding, which corresponds to a probability of 10^{-5}-10^{-6} relative to K^+ binding for the pre-catalytic and free state. Interestingly post catalytic state of group II intron discriminate weakly between K^+ and Na^+ in its active site, preferring the former by only 1 kcal/mol. Each free energy calculations have been repeated several times with different initial velocities. The coupling parameter (λ) vs free energy derivatives along the $K^+ \rightarrow Na^+$ transformation is given in **appendix Figure 2.3**. The area under the λ vs free energy derivatives curve gives the free energy change and different runs are in excellent agreement with each other (**see appendix Figure 2.3**). The uncertainty of overall free energy change (**see appendix Table 2.1**) and free energy derivatives at each of the λ points (**see appendix Table 2.2**) is well within the acceptable statistical uncertainty.

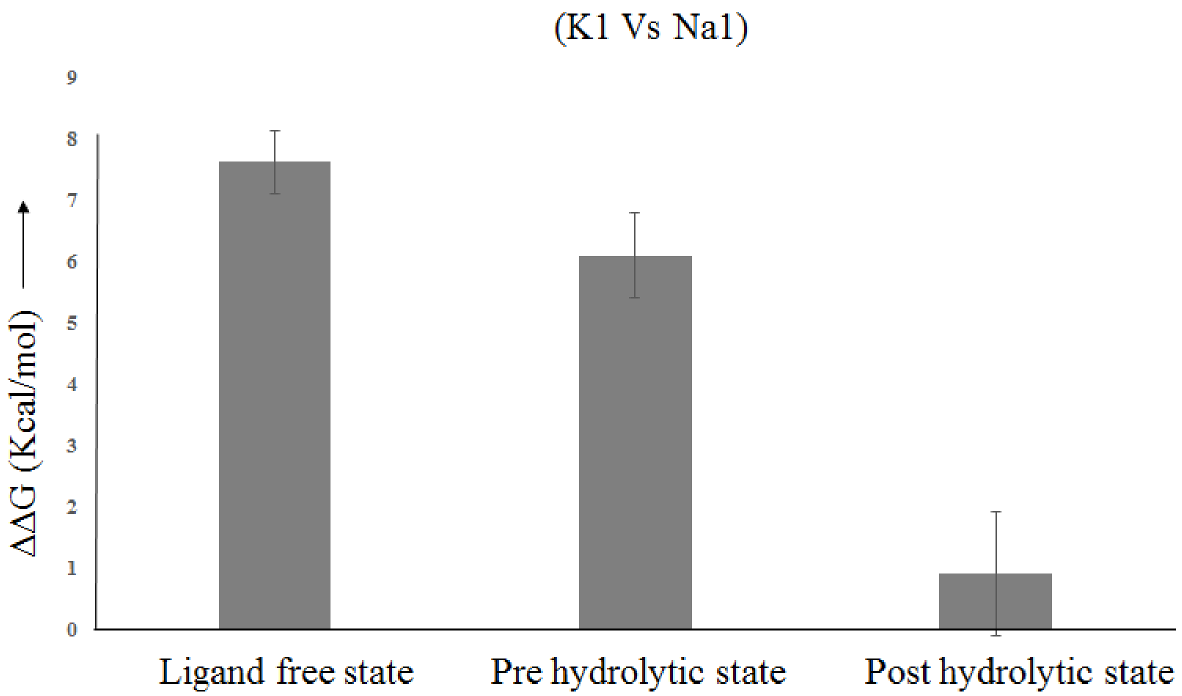


Figure 2.4. Computed binding free energy difference of intron upon K1 to Na1 mutation at different stages of splicing. Free energies are in kcal/mol. Statistical uncertainty is in the parentheses. The MD trajectories were divided into two equal halves and the difference between the computed ΔG's from the two halves is reported as uncertainty in the parenthesis. The uncertainties for ΔΔG's were calculated by propagating the uncertainties of individual ΔG's.

2.3.2 Robust structural features from MD and its comparison with X-ray Structures

Simulated structures agree well with the X-ray structures with an RMSD (of heavy atoms) of about 1.0 Å- 1.6 Å (**see appendix Figure 2.1**). MD simulations show following robust features:

(a) Consideration of free K^+/Na^+ in water is essential for understanding the monovalent ion binding to group II intron. In water K^+/Na^+ coordinates with 6/5 water molecules with an average distance between 2.7- 3.0Å/2.2- 2.4Å (**Figure 2.5**).

(b) Six coordination of K1 in the active site of group II intron is satisfied by forming direct interaction with a single water molecule and rest 5 interaction with the group II intron (**Figure 2.6 a, 2.7 a, and 2.8 a**). It should be noted that X-ray structures (**Figure 2.1**) did not report water molecule in the K1 binding pocket.

(c) Interestingly, MD trajectories show that the distance between O6(G288) −K1 is smaller (also less fluctuating) with respect to other coordination distances of K1 (**Table 2.1**). This suggests that G288 is a key player in proper positioning K1 in the binding pocket.

The above structural features are consistent and observed in all the MD replicas. The MD structures of the intron core and positions of four metal ions M1, M2, K1, and K2 are almost identical to their corresponding X-ray structures (**see appendix Figure 2.1 b-d and Table 2.1**).

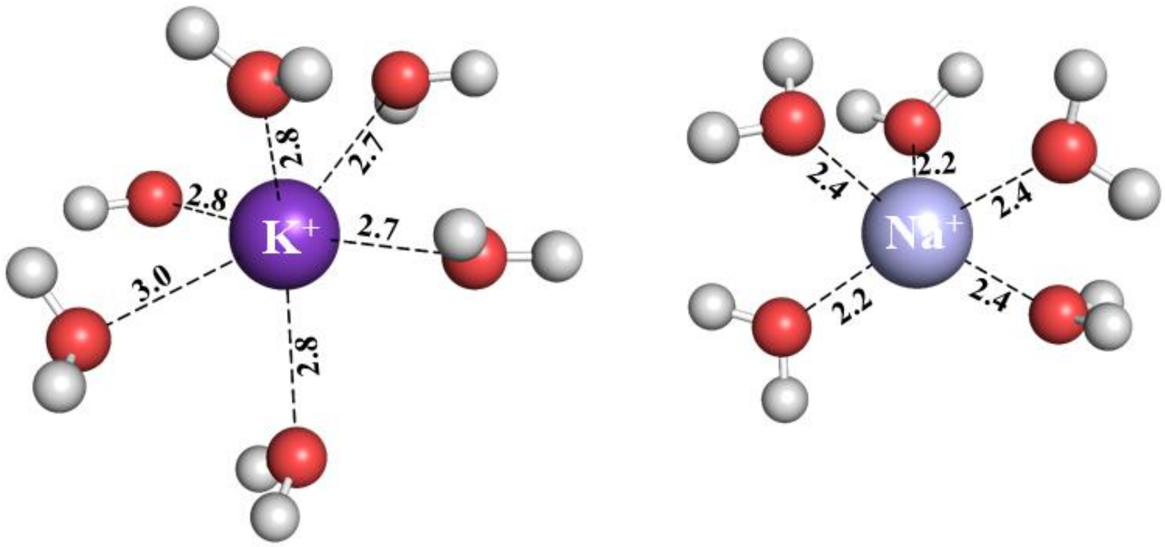


Figure 2.5. MD insight into the first coordination shell of free K^+/Na^+ in water. Interaction distances are shown as broken line and the distances are in angstrom.

2.3.3 Ligand free group II intron with K^+/Na^+: MD vs X-ray

MD structures of K^+ (K1) bound to the ligand free intron agree well with corresponding X-ray structure; PDB 4E8M (**Table 2.1**). The rms deviation of the main chain and side chain heavy atoms are 1.02 ± 0.07Å and 1.12 ± 0.06Å, respectively, with respect to 4E8M. The positions of four metal ions (K1, K2, M1, and M2) are very stable throughout the MD trajectory (**Table 2.1**). MD simulations show that CYT358, GUA359, GUA288, CYT377 and one H_2O form the binding pocket for K1 (**Figure 2.6 a, and Table 2.1**). The water molecule is positioned by the N7 of G288 for favorable interaction with K1 (**Figure 2.6 a**). Interaction distance between a K1 and G288 was observed to be shortest with a mean K1-G288(06) distance of 2.61±0.11Å. M2 interacts with phosphates of CYT358, GUA359 and CYT377 with a mean M2-O distance of 1.84±0.05Å, 1.88±0.05Å and 1.88±0.05Å respectively and fulfill 6 coordination by interacting with 3 water molecules (**Figure 2.6 a**).

M1 form direct interaction with the phosphates of CYT377, URA375, and 4 water molecules. It is interesting to note that there is only a single water molecule coordinating with K1 in the intron binding pocket. The desolvation penalty associated with the binding of K1 from water to the intron active site is compensated by establishing a direct interaction with the intron backbone and sidechain. It should be noted that K1 in its free and intron bound form always satisfy its 6 coordination with average heavy atom distances between 2.7- 3.0Å (**Table 2.1**). It can be argued that the position of K1 in the intron active site might be more favorable due to the presence of negatively charged binding partners.

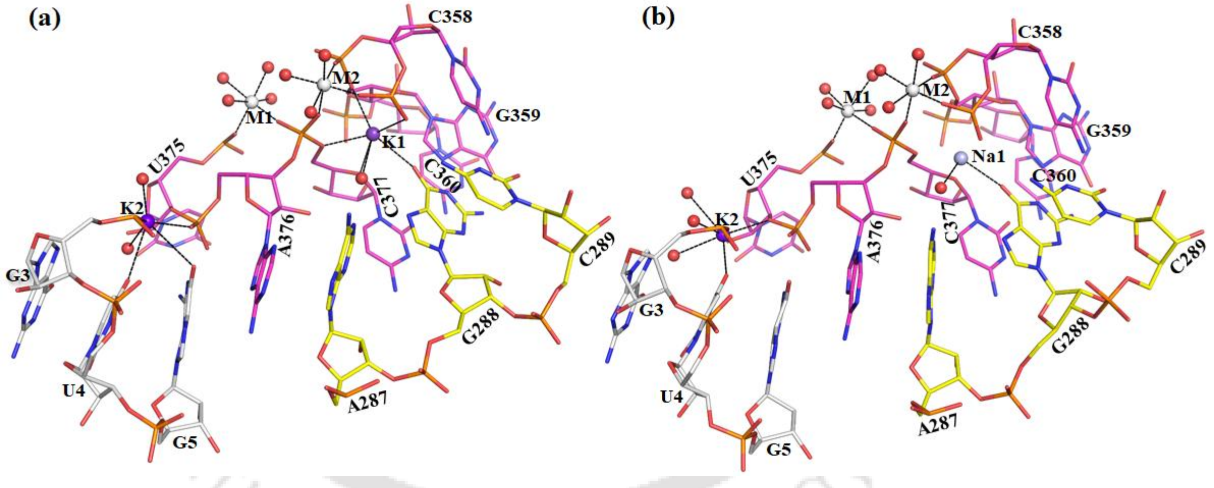


Figure 2.6. Close up of the binding pocket from molecular dynamics simulation (a) K1 bound ligand free intron (b) Na1 bound ligand free intron. (M1, M2) = Mg^{2+}.

The mutation of K1 into Na1 in the intron active site provides molecular insight into the thermodynamically hidden high energy state. Na1 is trapped with two interactions (with heavy atom distances ≤ 2.4Å) with G288(O6) and H_2O in the intron active site (**Figure 2.6 b**). It should be noted that the Na1 in the free intron active site is trapped in the dry desolvated pocket with unsatisfied coordination shell. This leads to very high discrimination ΔΔG ~ 7 kcal/mol disfavoring Na1 in the intron active site. The large discrimination strength could be utilized towards structural instability and preferential dissociation of ions from the higher energy intron conformation.

2.3.4 Pre-catalytic group II intron with K^+/Na^+: MD vs X-ray

In the pre-hydrolytic state 5' exon is present and covalently linked with the intron. The rms deviation of the main chain and side chain heavy atoms are 1.1 ± 0.07Å and 1.14 ± 0.06Å, respectively, with respect to the template pre-catalytic intron conformation (PDB 4FAQ). The interaction pattern of K1 and M2 in the pre-hydrolytic state is almost identical to the free state (**Table 2.1 and Figure 2.7 a**). M1 forms monodentate coordination with two waters, CYT377, URA375 and bidentate coordination with URA0.

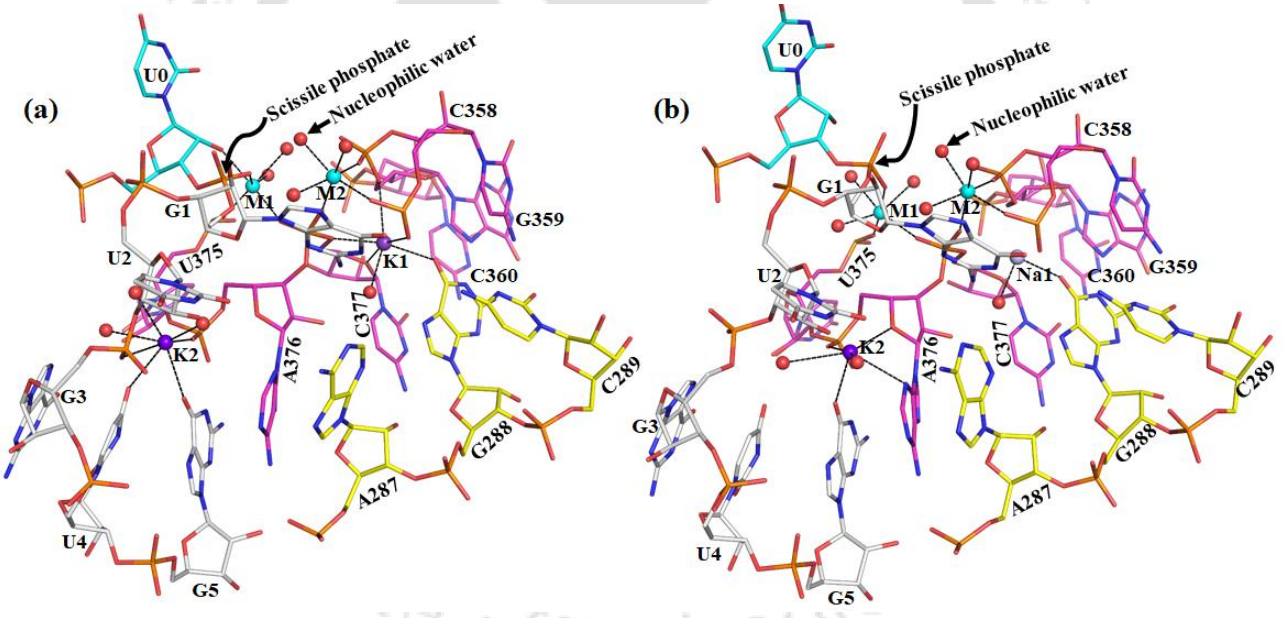


Figure 2.7. Close up of the binding pocket from molecular dynamics simulation (a) K1 bound pre-hydrolytic state (b) Na1 bound pre-hydrolytic state. (M1, M2) = Ca^{2+}.

The scissile linkage is positioned by M1 and the nucleophilic water is positioned by M2 for the in-line attack. None of our MD trajectories show M2-scissile phosphate interaction as proposed in X-ray structure (**Figure 2.1 b**). The interaction pattern described in the X-ray structures are very much stable throughout our MD trajectory (**Table 2.1**). The binding site of K1 is dry as seen in the ligand free intron. Na1 in the pre-catalytic state (**Figure 2.7 b**) is trapped in almost in a similar fashion as in free intron state and show higher discrimination ΔΔG ~ 6 kcal/mol favoring K1 for the same reason. It should be noted that Na^+ bound in the active site of the intron has not been experimentally observed. This might be due to the higher energy and low Boltzmann weight.

3.3.5 Post-catalytic group II intron with K^+/Na^+: MD vs X-ray

The rms deviation of the main chain and side chain heavy atoms are 1.1 ± 0.07Å and 1.14 ± 0.06Å, respectively, with respect to the template post-catalytic intron conformation (PDB 4FAR). The network of interactions between ions (K1, M1, and M2) and intron in K1 bound post-catalytic state is very stable in all our MD simulations. K1, M1 and M2 fulfil their coordination sphere by coordinating with 1, 2 and 3 water molecules respectively (**Figure 2.8 a**).

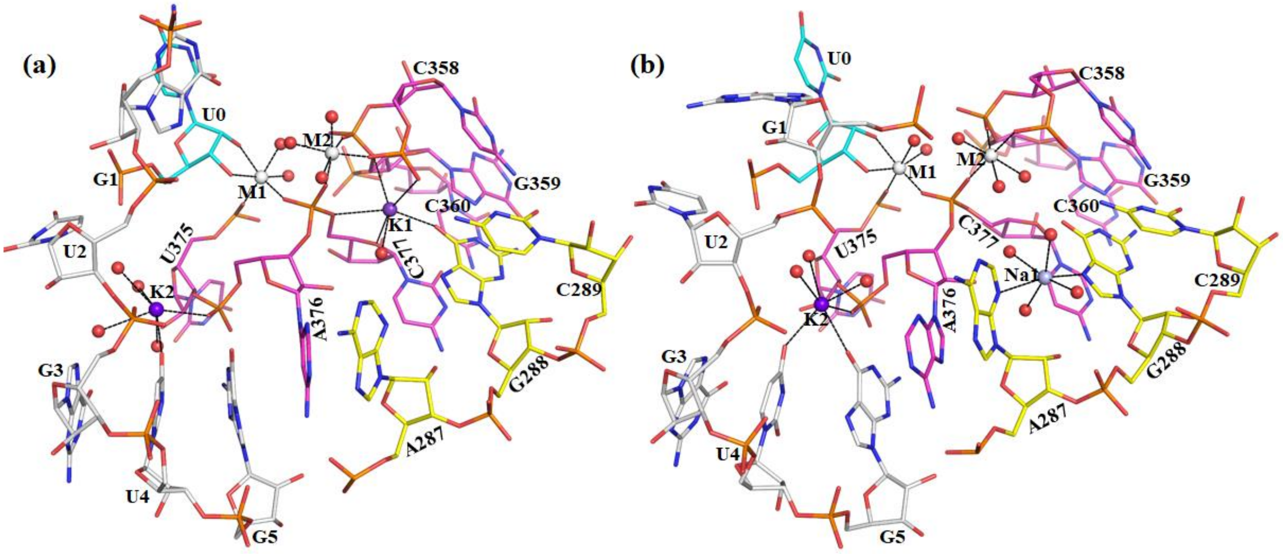


Figure 2.8. Close up of the binding pocket from molecular dynamics simulation (a) K1 bound post-hydrolytic state (b) Na1 bound post-hydrolytic state. (M1, M2) = Mg^{2+}.

Mutation of K1 into Na1 always open up the binding pocket and water molecules enter into the binding site for coordinating with Na1 (**Figure 2.8 b**). This suggests that the K1 binding pocket in the post catalytic state is loose and easy water accessibility leads to smaller discrimination (ΔΔG ~ 1 kcal/mol) between K1 and Na1. Structural parameters are given in **Table 2.1**.

Table 2.1. Selected interatomic distances averaged over the MD trajectories. Standard deviations are in the parentheses. Distances are in angstrom. Residues/groups absent in the X-ray/MD structures are indicated with a cross.

	Ligand-free state			**Pre-hydrolytic state**			**Post-hydrolytic state**		
Interacting Pair	**X-ray**	**MD**		**X-ray**	**MD**		**X-ray**	**MD**	
	4E8M	**K1**	**Na1**	**4FAQ**	**K1**	**Na1**	**4FAR**	**K1**	**Na1**
K1/Na1-C377:O1P	3.54	3.24 (0.20)	3.40 (0.22)	2.52	3.37 (0.23)	3.49 (0.24)	3.03	3.51 (0.22)	6.17 (0.29)
K1/Na1-C377:O5’	2.84	2.77 (0.11)	2.58 (0.19)	2.59	2.75 (0.11)	2.63 (0.19)	2.70	2.84 (0.15)	6.35 (0.25)
K1/Na1-C377:O4’	3.22	3.05 (0.18)	2.77 (0.25)	3.09	2.93 (0.17)	2.61 (0.22)	3.09	2.87 (0.15)	5.83 (0.31)
K1/Na1-G288:O6	2.80	2.61 (0.11)	2.36 (0.12)	2.80	2.64 (0.11)	2.38 (0.13)	2.48	2.62 (0.11)	4.03 (0.36)
K1/Na1-C358:O2P	2.54	2.74 (0.16)	2.65 (0.42)	2.95	2.71 (0.15)	2.58 (0.36)	2.93	2.68 (0.13)	5.89 (0.68)
K1/Na1-G359:O5’	3.25	2.88 (0.17)	3.10 (0.33)	3.16	2.95 (0.22)	3.01 (0.31)	3.06	3.03 (0.24)	8.42 (0.31)
K1/Na1-A287:N3	X	X	X	X	X	X	X	X	2.58 (0.17)
K1/Na1-G288:N7	X	X	X	X	X	X	X	X	2.46 (0.10)
K1/Na1-TIP3	X	2.73 (0.14)	2.32 (0.09)	X	2.66 (0.11)	2.33 (0.09)	X	2.70 (0.11)	2.48 (0.51)
K1/Na1-TIP3	X	X	X	X	X	X	X	X	2.39 (0.15)
K1/Na1-TIP3	X	X	X	X	X	X	X	X	2.44 (0.21)
K1/Na1-TIP3	X	X	X	X	X	X	X	X	2.46 (0.45)
M2-C377:O1P	2.78	1.88 (0.05)	1.88 (0.05)	1.92	2.15 (0.06)	2.14 (0.05)	2.15	1.88 (0.05)	1.85 (0.04)
M2-C358:O1P	2.44	1.84 (0.05)	1.83 (0.04)	2.21	2.13 (0.05)	2.12 (0.04)	1.84	1.84 (0.04)	1.85 (0.04)
M2-G359:O2P	1.95	1.88 (0.05)	1.88 (0.05)	2.41	2.13 (0.05)	2.12 (0.04)	1.98	1.89 (0.05)	1.85 (0.04)
M2-TIP3	X	2.02 (0.07)	2.02 (0.07)	X	2.31 (0.11)	2.27 (0.07)	X	2.02 (0.07)	2.01 (0.07)
M2-TIP3	X	2.02 (0.07)	2.02 (0.07)	X	2.27 (0.08)	2.23 (0.06)	X	2.01 (0.07)	2.03 (0.07)
M2-TIP3	X	1.99 (0.07)	2.00 (0.07)	X	2.25 (0.06)	2.27 (0.06)	X	2.02 (0.07)	2.00 (0.06)
M1-C377:O2P	2.94	1.86 (0.04)	1.86 (0.04)	2.17	2.13 (0.05)	2.12 (0.05)	2.20	1.84 (0.04)	1.85 (0.04)
M1-U375:O1P	2.56	1.84 (0.04)	1.85 (0.05)	2.04	2.11 (0.04)	2.10 (0.04)	1.85	1.84 (0.04)	1.82 (0.04)
M1-U0:O2’	X	X	X	2.49	2.14 (0.05)	2.13 (0.05)	2.63	2.08 (0.08)	2.11 (0.09)
M1-U0:O3’	X	X	X	2.03	2.43 (0.31)	4.50 (0.19)	2.47	2.04 (0.07)	2.08 (0.09)
M1-TIP3	X	2.01 (0.07)	2.02 (0.07)	X	2.26 (0.07)	2.24 (0.06)	X	2.00 (0.06)	1.99 (0.06)
M1-TIP3	X	2.00 (0.07)	2.00 (0.06)	X	2.27 (0.07)	2.33 (0.07)	X	2.00 (0.06)	1.99 (0.06)
M1-TIP3	X	2.03 (0.07)	2.02 (0.07)	X	X	2.46 (0.06)	X	X	X
M1-TIP3	X	1.99 (0.06)	1.99 (0.06)	X	X	X	X	X	X
M2-M1	3.69	5.05 (0.11)	5.07 (0.11)	4.31	5.07 (0.15)	5.04 (0.14)	4.45	5.05 (0.11)	5.35 (0.11)
K1-M2	4.42	4.23 (0.18)	4.45 (0.21)	3.83	4.65 (0.19)	4.75 (0.20)	4.26	4.39 (0.19)	6.11 (0.36)
K1-M1	6.42	6.45 (0.14)	6.48 (0.20)	6.14	6.66 (0.19)	6.66 (0.20)	5.72	6.52 (0.18)	9.51 (0.20)
K2-K1	10.90	12.36 (0.77)	12.89 (0.26)	10.77	11.31 (0.49)	10.03 (0.62)	10.47	11.52 (1.05)	10.81 (0.42)

2.4 Discussion

We have calculated K1/Na1 binding free energy differences to the group II intron at different stages (Free, pre-catalytic and post catalytic state) along the first step of self-splicing pathway. The calculated strength of discrimination, ΔΔG ~6-7 kcal/mol suggests that K1 binding to free and pre-catalytic state is strongly favored with respect to Na1. On the other hand, small ΔΔG ~1 kcal/mol suggests that K1 binding to post-catalytic state is weakly favored with respect to Na1. Since the magnitude of the relative preferences (ΔΔG) is not known experimentally, the calculated MDFE values cannot be confirmed or disproved. Certainly, the signs are correct and corroborates the experiment (**Marcia et al., 2012**). Free K^+/Na^+ ions in water show 6/5 coordination in their first coordination shell. The average distance between K^+/Na^+ and oxygen of waters are between 2.7-3.0Å/2.2-2.4Å. In the intron bound state K1 satisfy its 6 coordination by establishing a direct interaction with 1 water and 5 intron atoms with an average distance of 2.7-3.0Å. The same is true for free, pre and post catalytic intron conformation. However, Na1 in the free and pre-catalytic state is locked within the catalytic pocket with only 2 interactions with an average distance 2.2-2.4Å. Due to the smaller atomic radii of the trapped Na1 (in the free and pre-catalytic intron active site) the interactions are optimized by forming 2 strong interactions with an average distance 2.2-2.4Å and 2-3 weak interaction with an average distance >2.4Å (**Table 2.1 and Figure 2.6 b, 2.7 b**). This leads to the energetic discrimination of 6-7 kcal/mol in favor of K1. The high energy Na1 bound free, pre-catalytic states are thermodynamically hidden and could not be resolved experimentally. Our simulation has given atomic insight into those states. The large discrimination strength of 6-7 kcal/mol could be utilized towards structural instability and preferential dissociation of ions from the higher energetic complexes. Indeed, the crystal structure (**Marcia et al., 2012**) obtained of group II intron in presence of Na^+ and Mg^{2+} has a very different active site conformation without having bound metal ions **(see appendix Figure 2.2**). The post-catalytic K1 site is loose and the mutation of K1 by Na1 leads to water entry and compensate the interaction loss due to its smaller size into the active site, leading to loss of discrimination. The difference in the discrimination strength is related to the strict confinement of the monovalent ion in the binding site. It is interesting to note that K1-GUA288(O6) distance is the smallest (**Table 2.1 and Figure 2.6, 2.7, 2.8**) and the single water molecule coordinating with K1 in the binding site is positioned by G288:N7 in all the simulations, suggesting the importance of G288 in stabilizing the K1 in the binding pocket. The direct interaction between GUA288:O6 and K1 is very specific to guanine and excludes any other nucleotide at that position. It is worth mentioning the ADE287 block the water entry into the K1 binding site and the dryness of the binding pocket leads to discrimination. The conservation of J2/3 junction supports the hypothesis.

2.5 Conclusion

Free energy simulations demonstrated that K1/Na1 relative binding free energetics varies between different conformational states (free, pre-catalytic, post-catalytic) of group II intron. Free and pre-catalytic intron states are highly selective towards K1 binding and rejecting Na1. It is interesting to note that small change in ionic radii could boost the discrimination strength by 6-7 kcal/mol in the free and pre-catalytic intron active site. Trapping of Na1 in the free and pre-catalytic intron binding site with unsatisfied coordination sphere is responsible for strong discrimination. In the post-catalytic intron conformation, the K1/Na1 binding site opens up and allow waters to coordinate with Na1, leading to loss of selectivity. The role of G288 is twofold (a) proper positioning of K1 into the active site by establishing direct interaction (b) holding a single water molecule for solvating K1 in the dry desolvated binding pocket. Though the high-selectivity states (Na1 bound ligand-free, pre-hydrolytic intron states) are hidden from structural and binding studies but evidently exist and this computational study has provided insight into the same.

Chapter 3

Divalent metal ion (Mg^{2+} vs. Ca^{2+}) selectivity in Group II intron

Group II introns are enzymes which undergo self-splicing and remove itself from pre-messenger RNA. X-ray structures of group II intron of *Oceanobacillus iheyensis* at various stages of the self-splicing pathway (Pre-hydrolytic, post-hydrolytic, and ligand-free state) revealed intricate atomic interaction network in the active site of the intron. It has been confirmed that a heteronuclear metal ion cluster consisting of four metal ions (K1, K2 sites with K^+ and M1, M2 sites with Mg^{2+}) are crucial for function. Substitution of Mg^{2+} by Ca^{2+} results in loss of enzymatic activity. The X-ray structures not only opens up the possibility of modelling Mg^{2+} and Ca^{2+} bound active site of group II intron and quantitatively estimate the energetics of Mg^{2+} vs Ca^{2+} preference but also explore the relative structural and dynamical differences in response to divalent metal ion substitution. Thus, using X-ray structures as a template we performed molecular dynamics simulations to compare structural and dynamical differences between Mg^{2+} and Ca^{2+} bound active site of group II intron at various stages of the splicing pathway (i.e, Pre-hydrolytic, post-hydrolytic, and ligand-free state). Quantitative estimation of Mg^{2+} vs Ca^{2+} selectivity at the M1, M2 sites confirmed Mg^{2+} preference at intron active sites relative to Ca^{2+}. Ca^{2+} is relatively more hydrated in the intron active site relative to Mg^{2+}. The local environment (bound nucleophilic water, interaction with scissile phosphate) around M2 is strikingly different between Mg^{2+} and Ca^{2+} bound pre-hydrolytic state. In the post-hydrolytic state, the exon part of the hydrolysis product is involved in direct interaction with the M1, whereas the intron part is highly flexible in our MD trajectories. Solvent exposure of M1, M2 sites are least in the pre-hydrolytic state, highest in the ligand-free state, and intermediate in the post-hydrolytic state.

3.1 Background

Eukaryotic RNA usually contains multiple protein-coding (exons) and non-coding (introns) regions **(Clancy, 2008)**. Ligation of exons by excision of introns results in mature mRNA, which is then translated by the ribosome, resulting in protein synthesis in cells. Thus, the fidelity of gene expression relies on accurate RNA processing. The process by which introns are removed and exons are joined together is known as splicing. Several human diseases (e.g, dilated cardiomyopathy, spinomuscular atrophy, myotonic dystrophy, cystic fibrosis, cancer etc.) are linked to incorrect RNA splicing **(Scotti et al., 2016; Cooper et al., 2009)**. Thus, understanding the mechanism of splicing is of great interest for fundamental science as well as its relevance to social welfare. Splicing can create a range of unique proteins by varying the exon composition (known as alternative splicing) in the mature mRNA. Based on the mechanism of splicing, introns are broadly classified into four groups i.e, tRNA, mRNA, group I and group II introns **(Clancy, 2008)**. mRNA and tRNA introns are spliced out by ribonuclear protein complexes (spliceosome) and protein enzymes (kinase) respectively. On the contrary, group I and II introns do not require proteins, the folded RNA/intron itself can catalyze its splicing (self-splicing). Splicing for mRNA and group II introns are chemically similar, as both involve nucleotide adenosine for splicing, however, the later undergo splicing without involving proteins (**Pyle et al., 2006**). On the other hand, group I and II introns undergo self-splicing but involving different nucleotides, guanosine for the former and adenosine for the later. Thus, group II introns could be considered as intermediate between group I introns and spliceosome. The study of group II intron as a model system is an excellent choice for understanding 3D-structure, catalysis and evolutionary perspective of the splicing process in all domains of life (**Mattick, 1994**). Spliced out free group II intron was also known to integrate into DNA and RNA resulting in genome diversification (**Dai et al., 2002**).

As explained in Chapter 1, group II introns consist of six-domains (D1-D6). Construct of group II introns (containing only five domains D1-D5) was structurally characterized for *Oceanobaciellus iheyensis* at various stages of splicing (e.g. pre-hydrolytic (4FAQ), post-hydrolytic (4FAR), and ligand-free state (4E8M) state) (**Marcia et al., 2012**). The high flexibility of domain 6 (D6) was speculated to hinder structural characterization and thus removed for structure determination. The constructs were shown to be active (**Marcia et al., 2012**). The pre-hydrolytic state of group II intron is shown in **Figure 3.1 a, b**. Different color represents the different domains. D1 (grey) is the largest domain onto which another domains dock. D2 and D3 are relatively short.

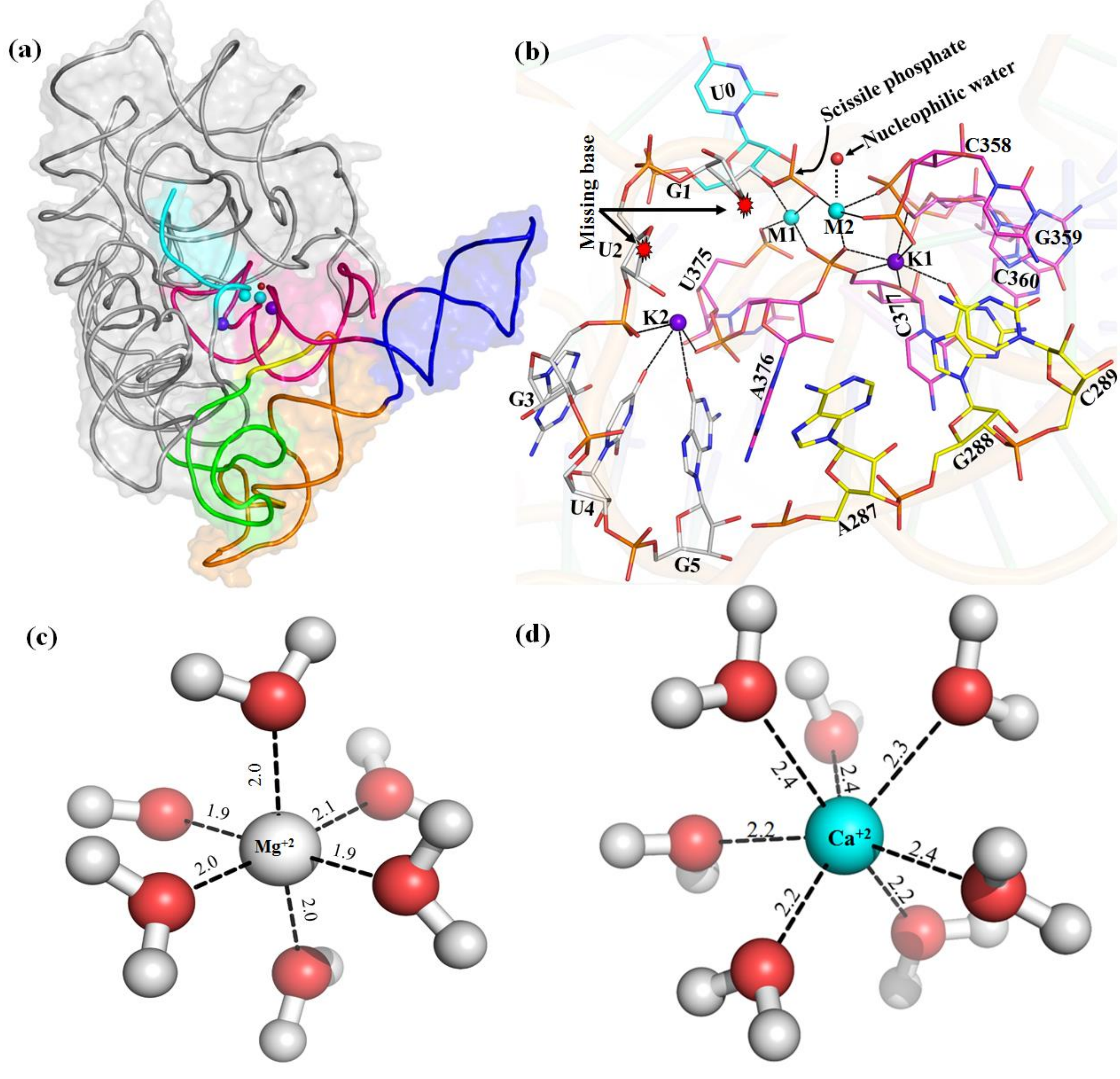


Figure 3.1. (a) Pre-hydrolytic state (PDB 4FAQ) of the construct of group II intron (D1-D5) in the cartoon. (D1: grey, D2: green, D3: orange, D4: blue, D5: pink, Junction between D2 and D3 or J2/3: yellow, 5'-exon: cyan, K^+ ions K1, K2: purple sphere, Divalent (M1, M2) Ca^{2+}: cyan sphere, Mg^{2+}: white sphere). (b) Zoomed in view of the four-metal (K1, K2, M1, M2)

heteronuclear active site from PDB 4FAQ. Nucleophilic water, scissile phosphate are marked. Nucleotide bases (G1, U2) were disordered/unresolved in the PDB 4FAQ. MD structure of first coordination shell of (c) free Mg^{2+} (d) Ca^{2+} in water.

D5 (red) is highly conserved (~35 nucleotides long) throughout evolution and constitute the active site of group II intron. The active site of group II intron is formed by the residue C358, G359, C360 (catalytic triad of D5), A376, C377 (bulge of D5), A287, G288, C289 (junction between D2 and D3), two monovalent ions (K1 and K2), and two divalent ions (M1 and M2). The presence of multiple negatively charged nucleotide phosphates in the compact active site creates hetero-nuclear metal ion (M1, M2, K1, K2) binding sites. Scissile phosphate linking U0 (exon-part) and G1 (intron-part) define the exon-intron boundary (**Figure 3.1 b**). Conservation of bulge and the catalytic triad of D5 and D2, D3 junction was proved to be important for catalysis **(Sigel et al., 2000; Gordon et al., 2001)**. To structurally characterize the pre-hydrolytic state (**Figure 3.1 b**) of group II intron, the crystallization was done in the presence of Ca^{+2}. The presence of Mg^{2+} facilitates hydrolysis of exon-intron linkage resulting post-hydrolytic splicing product. Hetero-nuclear metal ion cluster (M1, M2, K1, and K1) require for the structural integrity of the active site of group II intron. The divalent metal ion at the M2 position coordinates the nucleophilic water (red sphere) close to the scissile phosphate (**Figure 3.1 b**). K1 and K2 are crucial for the structural integrity of the catalytic pocket (**Marcia et al., 2012**).

Experimental studies (**Marcia et al., 2012; Sigel et al., 2000; Gordon et al., 2001; Pyle, 2016; Toor et al., 2008; Chan et al., 2018; Costa et al., 2016**) established the fact that group II introns are highly selective for monovalent (K1, K2: K^+) and divalent ion (M1, M2: Mg^{2+}) for catalysis. Previously, using computer simulations we have quantitatively estimated the relative binding affinity (K^+ vs Na^+) at the K1 site of group II intron (**Kumar et al., 2018**). The results suggest that K^+ at the K1 site is always preferred but the strength of preference is weaker for post-hydrolytic state relative to ligand-free and pre-hydrolytic state (**Kumar et al., 2018**). X-ray structures (**Marcia et al., 2012**) of group II introns at various stages of the splicing pathway has revealed the intricate interaction network in the active site of group II intron in atomic details. However, the structure of Mg^{2+} bound pre-hydrolytic state and divalent-metal ion selectivity (Mg^{2+} vs Ca^{2+}) by group II intron at various stages are unknown. Available X-ray structures now provide models for structure-based molecular dynamics simulations of group II intron.

Using X-ray structures (PDB 4E8M, 4FAQ, and 4FAR) as a template (**Marcia et al., 2012**), we modelled Mg^{2+} and Ca^{2+} active sites of group II intron and performed MD simulations. In this paper, we have discussed: (a) Structural and dynamical comparison of Mg^{2+} and Ca^{2+} bound group II intron at various stages of splicing. (b) Relative preference of divalent metal ions (Mg^{2+} vs Ca^{2+}) at various stages of splicing.

3.2 Methods

3.2.1 Molecular dynamics setup

Crystal structures of group II intron of *Oceanobacillus iheyensis* in its ligand-free state (PDB 4E8M, resolution 3.5Å), pre-hydrolytic state (PDB 4FAQ, resolution 3.11Å), and post-hydrolytic state (PDB 4FAR, resolution 2.86Å) were retrieved from the Protein Data Bank (**Marcia et al., 2012**). Spherical region of 25Å radius, centred at the P atom of the residue C377, was truncated from the X-ray structures and solvated with a cubic water box with 80Å of edge length. Intron nucleotide bases G1, U2 of pre-hydrolytic state (PDB 4FAQ) and G1 of ligand-free state were disordered/unresolved in the X-ray structure (**Figure 3.1 b**). Thus, we first build those RNA bases and energy minimized relative to the rest of experimentally determined atomic positions. Total ~48100 number of atoms are considered in our simulation and the total number of water molecules were ~14700. (**see Chapter 1; methods section for details**)

3.2.2 Free energy calculations

Binding free energy difference between Mg^{2+} and Ca^{2+} in the catalytic pocket was computed by alchemically transforming Mg^{2+} into Ca^{2+} in complex with RNA and free in the water, as described in the thermodynamic cycle (**Figure 3.2**). Vertical legs of **Figure 3.2** are Mg^{2+}/Ca^{2+} binding and horizontal legs correspond to alchemical/unphysical transformation of Mg^{2+} into Ca^{2+}. The free energy change associated with the horizontal legs (ΔG_{comp}, ΔG_{free}) was calculated (**see appendix Table 3.1 and 3.2**). The relative binding free energy was then computed using $\Delta\Delta G = \Delta G_{comp} - \Delta G_{free}$ (**see appendix Table 3.3**). The methodology of free energy estimation adopted in this work has been described in detail previously in **Chapter 1**.

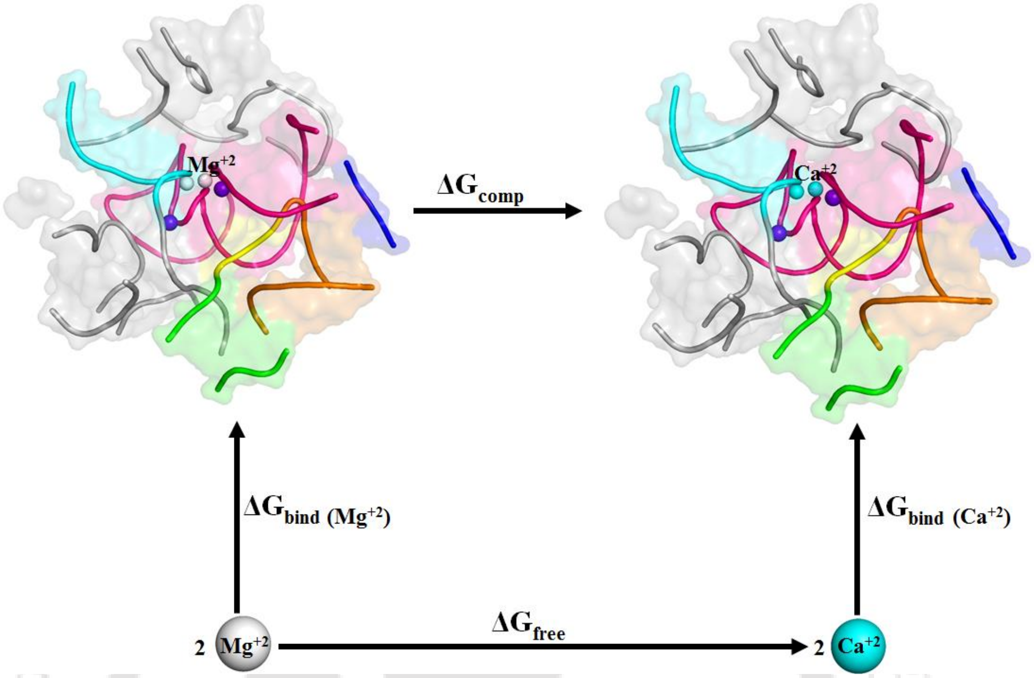


Figure 3.2. Thermodynamic cycle for Mg^{2+} vs Ca^{2+} ion binding to the divalent ion binding sites M1, M2 of group II intron active sites. Vertical legs correspond to divalent cation binding; horizontal legs correspond to the alchemical transformation of Mg^{2+} into Ca^{2+} in the M1, M2 binding sites, either in the intron bound (above) or free (below) in solution. Free energy associated with the horizontal legs of the thermodynamic cycle has been computed by MD simulations and thermodynamic integration. The binding free energy difference is $\Delta\Delta G = \Delta G_{comp} - \Delta G_{free} = \Delta G_{bind}(Ca^{2+}) - \Delta G_{bind}(Mg^{2+})$.

3.2.3 Electronic structure calculation

Ab initio quantum calculations were performed for Mg^{2+} and Ca^{2+} bound active site of group II intron in pre-hydrolytic, post-hydrolytic, and ligand-free state. Representative MD structures of group II intron were selected and divalent metal ions M1, M2 and its nearby nucleotides were extracted and embedded into a dielectric continuum representing bulk water. Nucleotide phosphates were directly coordinated to M1, M2. Thus, the nitrogenous base is replaced by a methyl group to reduce the computational cost. Structures were prepared for pre, post, and ligand-free intron state (**Figure 3.3**) with a variety of interaction modes, including two to four direct interactions between M1, M2, and phosphate oxygens.

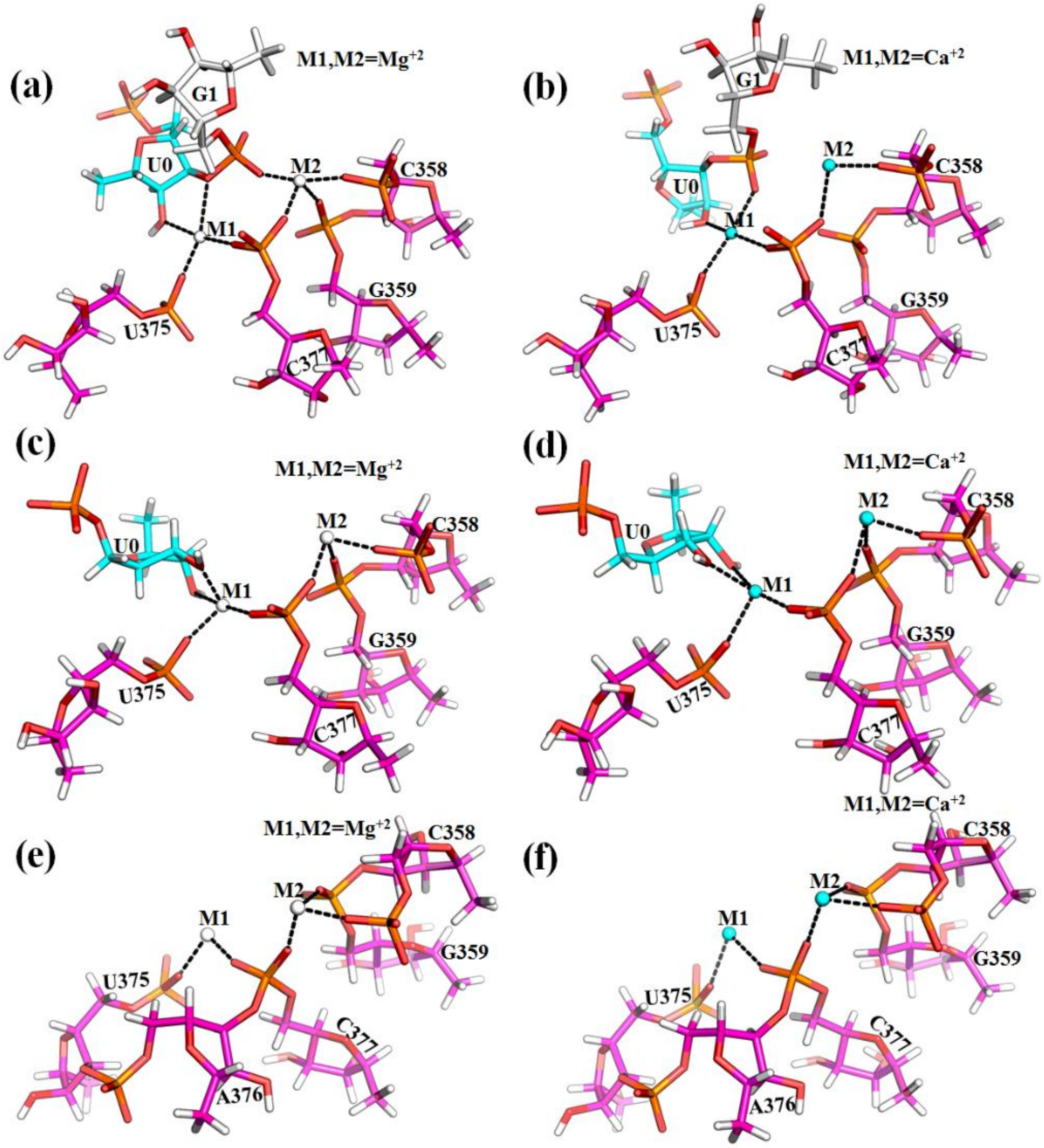


Figure 3.3. Local environment around M1, M2 sites considered for quantum chemical calculations. (a,b) Pre-hydrolytic state. U0-G1 is exon-intron boundary. (c,d) pos-hydrolytic state and (e,f) ligand-free state. Hydrogens are explicitly shown as white sticks. Nucleotide bases replaced by methyl group.

Representative local environment around M1, M2 embedded in aqueous dielectric was subjected to single point density functional energy calculation using two different functionals (1) B3LYP functional (**Becke 1988; Lee et al. 1988**) in combination with 6-31+G* basis set and (2) M06-2X functional (**Zhao et al., 2008**) in combination with 6-311++G** basis set with solvation treated by the implicit SMD solvent model. Calculations were performed using the Gaussian16 program (**Frisch et al., 2016**). Single point energy calculations were also performed by considering only M1 and M2 separated by 30Å in an aqueous dielectric. We focus on a quantity (Energy difference of product and reactant of eq. 3.1) that approximates the relative preference (Mg^{2+} vs Ca^{2+}) at M1, M2 in the intron active site by using the following equation:

$$[Mg^{2+}, Mg^{2+} - water] + [Ca^{2+}, Ca^{2+} - intron] \rightarrow [Ca^{2+}, Ca^{2+} - water] + [Mg^{2+}, Mg^{2+} - intron] \quad (3.1)$$

In eq (3.1), $[Ca^{2+}, Ca^{2+} - intron]$ and $[Mg^{2+}, Mg^{2+} - water]$ represents Ca^{2+} bound (M1, M2 sites) intron active site and intron free Mg^{2+} embedded into the implicit aqueous dielectric. Single point energy differences evaluated between the right-hand and left-hand side of eq (3.1) approximate the relative divalent cation preference at the intron active site M1, M2. A positive energy difference from eq (3.1) implies a Ca^{2+}-preference whereas a negative value implies Mg^{2+} preference at the intron active site.

3.3 Results

3.3.1 MD vs X-ray structures of group II intron at various stage of splicing

Structural similarly between MD and template X-ray structures of group II intron at various stages (ligand-free state, pre-hydrolytic state, and post-hydrolytic state) of splicing was analyzed by computing root mean square deviation (RMSD) of heavy atoms (within 22Å radius) relative to their template X-ray structures (4E8M, 4FAQ, and 4FAR) (**Marcia et al., 2012**). Average RMSD was found to be small, 1.29 ± 0.08Å, 1.33 ± 0.11Å, and 1.26 ± 0.08Å for pre-hydrolytic, post-hydrolytic, ligand-free state respectively. The structural convergence was confirmed from RMSD vs time plots (**see appendix Figure 3.1 a**). Local environment around hetero-nuclear metal ion cluster (M1, M2, K1, K2) observed in the X-ray structures agree well with the MD structures (**see appendix Figure 3.1 b-d and Table 3.1**). MD simulations of free Mg^{2+}/Ca^{2+} in explicit waters (**Figure 3.1 c, d**) reveals that Mg^{2+} and Ca^{2+} coordinates with 6 and 7 water molecules respectively. Mg^{2+}-water and Ca^{2+}-water average distances are 1.9-2.1Å and 2.2-2.4Å respectively.

3.3.2 Active site of pre-hydrolytic group II intron with bound Mg^{2+}/Ca^{2+}

MD structure of Mg^{2+} bound pre-hydrolytic complex (**Figure 3.4 a**) reveals the following features: (1) Both the Mg^{2+} (M1 and M2) establish direct interaction with the scissile phosphate. (2) Hexa-coordination of M2 is satisfied by coordinating with four phosphates (G1, C358, G359, and C377) and 2 water. (3) M2 coordinates the nucleophilic water and positioned it above the plane of scissile phosphate (distance P-O =3.4Å, angle $\angle$P-O-M2 = 71.5^0). (4) M1 coordinates with two phosphates (U375, C377), two ribose hydroxyl group (2' and 3'-OH of U0), and 2 water. We may conclude that the M1=M2=Mg^{2+} pocket in the pre-hydrolytic state is compact and dry (only two waters in the first coordination sphere of M1, M2, see **Table 3.2**).

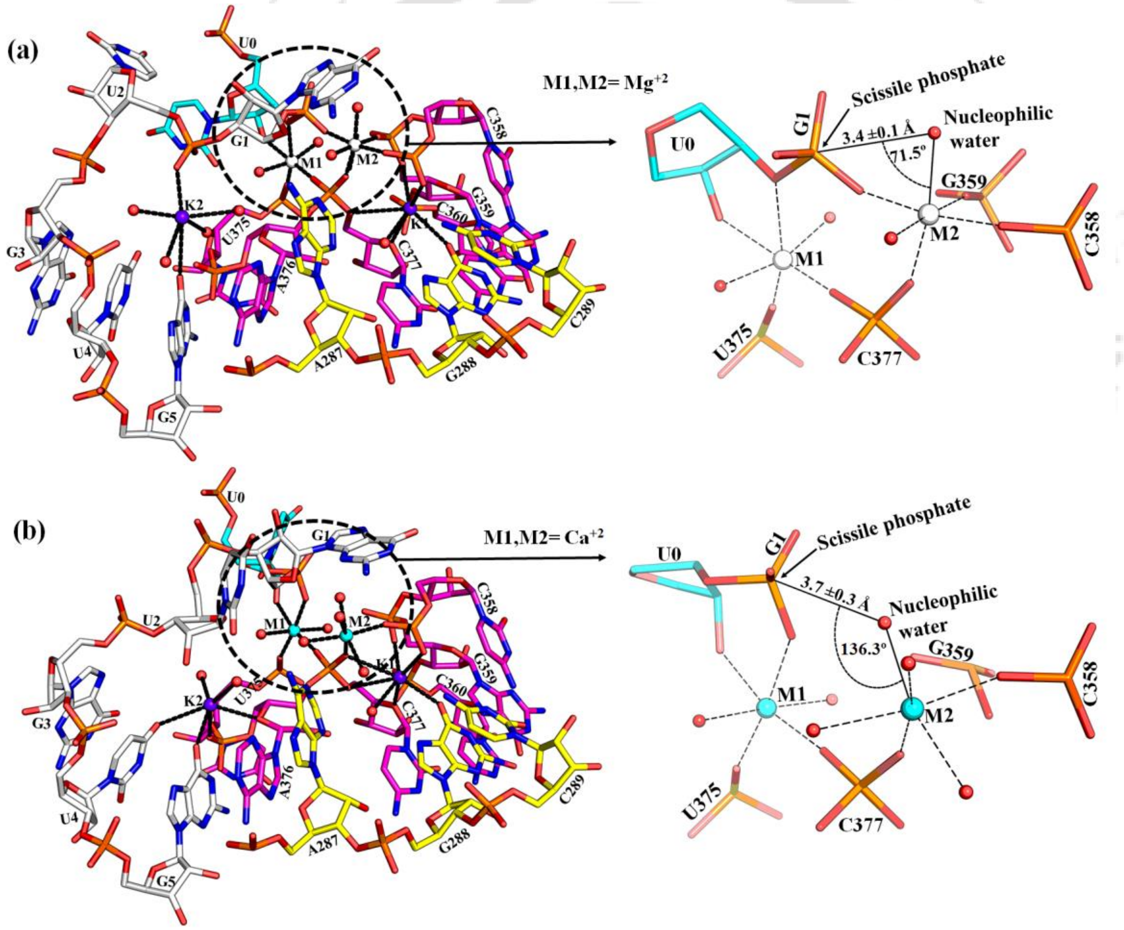


Figure 3.4. MD structures of Pre-hydrolytic intron active site and waters around the divalent ions (M1, M2). (U0, Cyan sticks) - (G1, Gray sticks) is the exon-intron boundary. Zoomed-in view of the local interaction network around (M1, M2) on the right: (a) Mg^{2+}-bound (b) Ca^{2+}-bound. For clarity hydrogens and nucleotide bases (in the zoomed-in view) were not shown.

Table 3.1. Selected interatomic distances (in Å) averaged over the MD trajectories. O1P, O2P: Terminal oxygens of phosphate backbone.

Interacting pair	**Pre-hydrolytic state**			**Post-hydrolytic state**			**Ligand-free state**		
	X-ray (Ca^{2+})	**Mg^{+2}**	**Ca^{+2}**	**X-ray (Mg^{2+})**	**Mg^{+2}**	**Ca^{+2}**	**X-ray (Mg^{2+})**	**Mg^{+2}**	**Ca^{+2}**
K1-C377:O5’	2.59	2.91 ±0.18	2.85 ± 0.16	2.70	2.92 ± 0.18	2.81 ± 0.15	2.84	2.79±0.12	2.84 ± 0.14
K1-C377:O4’	3.09	2.86 ± 0.15	2.89 ± 0.18	3.09	2.82 ± 0.14	2.87 ± 0.18	3.22	3.03 ±0.09	3.04 ± 0.20
K1-G288:O6	2.80	2.63 ± 0.12	2.65 ± 0.12	2.48	2.64 ± 0.13	2.68± 0.14	2.80	2.64 ±0.12	2.70± 0.14
K1-C358:O2P	2.95	2.75 ± 0.16	2.95 ± 0.25	2.93	2.71 ± 0.14	2.83± 0.22	2.54	2.74 ±0.16	2.80± 0.19
K1-G359:O5’	3.16	2.90 ± 0.18	4.75 ± 0.17	3.06	2.99 ± 0.22	3.08 ± 0.32	3.25	2.88 ±0.17	2.85 ± 0.16
K1-G359:O2P	2.95	3.94 ± 0.22	2.54 ± 0.09	3.31	4.35 ± 0.24	4.10 ± 0.30	2.65	4.10 ±0.20	3.94 ± 0.26
K1-TIP3	X	2.68 ± 0.12	2.84 ± 0.20	X	2.67 ± 0.12	2.81± 0.19	X	2.75 ±0.15	2.70 ± 0.13
M2-C377:O1P	1.92	1.89 ± 0.05	2.15 ± 0.20	2.15	1.88 ± 0.05	2.17± 0.06	2.78	1.88 ±0.05	2.17 ± 0.06
M2-C358:O1P	2.21	1.85± 0.05	2.11± 0.05	1.84	1.84 ± 0.04	2.13 ± 0.06	2.44	1.84 ±0.05	2.12 ± 0.05
M2-G359:O2P	2.41	1.89 ± 0.05	4.64 ± 0.27	1.98	1.90 ± 0.05	2.21 ± 0.16	1.95	1.88 ±0.05	2.20 ± 0.08
M2-G1:O2P	2.21	1.88 ± 0.05	4.53 ± 0.30	4.52	8.34 ± 1.30	5.44 ± 0.95	12.61	16.42±0.63	15.16 ±0.82
M2-TIP3	2.48	2.04 ± 0.07	2.26 ± 0.07	1.71	2.02 ± 0.08	2.31± 0.10	1.96	2.02 ±0.07	2.31 ± 0.09
M2-TIP3	X	2.02 ± 0.07	2.25 ± 0.06	2.46	2.01 ± 0.07	2.30 ± 0.10	X	2.00 ±0.06	2.34 ± 0.10
M2-TIP3	X	X	2.13 ± 0.08	X	2.02 ± 0.05	2.32± 0.12	X	2.02 ±0.08	2.29 ± 0.08
M2-TIP3	X	X	2.23 ± 0.06	X	X	2.35 ± 0.11	X	X	2.35 ± 0.09
M1-C377:O2P	2.17	1.85 ± 0.05	2.13 ± 0.05	2.20	1.86 ± 0.04	2.11 ±0.05	2.94	1.86 ±0.04	2.13 ± 0.04
M1-U375:O1P	2.04	1.82 ± 0.04	2.10 ± 0.04	1.85	1.83 ± 0.04	2.10 ±0.05	2.56	1.85 ±0.04	2.11 ± 0.04
M1-U0:O2’	2.49	2.08 ± 0.08	2.35 ± 0.09	2.63	2.12 ± 0.10	2.36 ±0.07	X	X	X
M1-U0:O3’	2.03	2.40 ± 0.08	3.65 ± 0.15	2.47	2.06 ± 0.08	2.28 ±0.08	X	X	X
M1-G1:O2P	2.49	3.61± 0.16	2.12 ± 0.05	4.61	5.98± 1.36	7.65 ± 1.09	10.01	14.0 ±0.62	12.95±0.82
M1-TIP3	X	1.98 ± 0.06	2.25 ± 0.06	2.60	1.98± 0.06	2.24 ± 0.06	X	2.02 ±0.07	2.33± 0.08
M1-TIP3	X	1.95 ± 0.06	2.22 ± 0.06	X	2.01± 0.06	2.26 ± 0.06	X	2.00 ±0.07	2.35± 0.07
M1-TIP3	X	X	X	X	X	2.23 ± 0.09	X	1.99 ±0.06	2.22 ± 0.07
M1-TIP3	X	X	X	X	X	X	X	2.01 ±0.07	2.30 ± 0.08
M1-TIP3	X	X	X	X	X	X	X	X	2.35± 0.06
K1-M2	3.83	4.18 ± 0.20	4.36 ± 0.20	4.26	4.40 ± 0.21	4.54± 0.24	4.42	4.23 ±0.19	4.59 ± 0.21
K1-M1	6.14	6.62 ± 0.18	6.69 ± 0.18	5.72	6.57± 0.18	6.69 ± 0.18	6.42	6.52 ±0.15	6.75± 0.20
M2-M1	4.31	4.89 ± 0.11	5.67 ± 0.16	4.45	5.00 ± 0.12	5.67 ± 0.16	3.69	5.07 ±0.12	5.11 ± 0.17
K1-K2	10.77	11.34 ±1.22	10.18 ±0.40	10.47	10.48 ±0.62	10.18 ±0.40	10.90	13.03±0.3	13.02±0.28

MD structure of the catalytic pocket of Ca^{2+} bound pre-hydrolytic complex (**Figure 3.4 b**) show that the M2 binding pocket is less compact relative to its Mg^{2+} analog. Characteristic features (**Table 3.1, and Table 3.2**) of M2 pocket are: (1) wet M2 pocket, coordinates with four waters and two phosphates (C358, C377) (2) Scissile phosphate is away from Ca^{2+} (at M2 position). (3) The nucleophilic water is 3.7Å away from the phosphorus of scissile phosphate and the orientation of the same water (angle ∠P-O-M2 = 136.3^0) is strikingly different relative to its Mg^{2+} analog. The orientation of nucleophilic water is a key factor for facile hydrolysis. MD structures of M2 binding pocket of Ca^{2+} bound pre-hydrolytic complex differ from its X-ray template, showing disruption of M2-G359, M2-G1:O2P interactions (**Table 3.1, and Figure 3.4 b**). The observed difference in the M2 local environment is a robust feature confirmed by multiple independent MD replicas. Thus, it is unclear if the crystal environment and/the limited resolution of the X-ray structure results in the less accurate microenvironment around M2, or the limited accuracy of force-filed is the sole cause of observed deviation.

Table 3.2. Number of waters in the first coordination shell around M1 and M2. Number of waters as a function of distance from M1, M2 sites

Intron State	**M1 site**		**M2 site**	
	Mg^{2+}	Ca^{2+}	Mg^{2+}	Ca^{2+}
Pre-hydrolytic	2	2	2	4
Post-hydrolytic	2	3	3	4
Ligand-free	4	5	3	4

3.3.3 Active site of post-hydrolytic group II intron with bound Mg^{2+}/Ca^{2+}

Mg^{2+} bound post-hydrolytic state (**Figure 3.5 a**) shows the Mg^{2+} (in M2 site) coordinates with three intron phosphates (C358, G359, and C377) and three waters in its first coordination shell. Mg^{2+} (in M1 site) establishes four direct interactions with the intron (residues U375, C377, and 2',3'-ribose-OH of U0) and two interactions with waters (**Figure 3.5 a**). The local interaction network around four metal center (M1, M2, K1, and K2) observed form MD structures are in excellent agreement with the X-ray structure (**Table 3.1**). The substitution of Ca^{2+} in M1 and M2 sites results in the entry of an additional water molecule (**Figure 3.5 b, and Table 3.2**). The post-hydrolytic intron part (residue G1) is observed to be free, with no direct interactions with M1/M2. Whereas, direct interaction between exon part (residue U0) and M1 was stable throughout our MD trajectories (**Figure 3.5 a, b, and Table 3.1**).

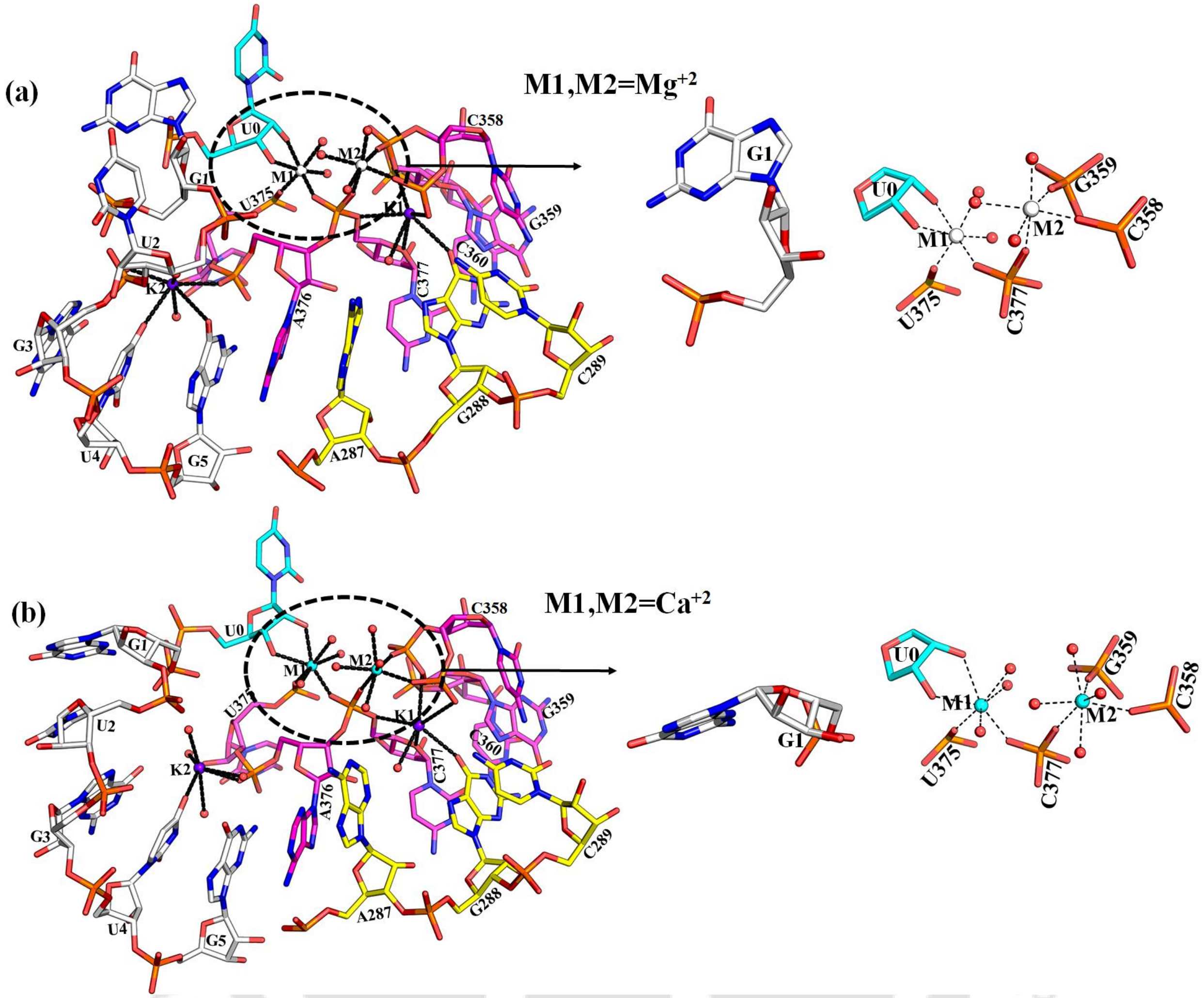


Figure 3.5. MD structures of Post-hydrolytic intron active site and waters around the divalent (M1, M2) ions. (a) Mg^{2+}-bound active site. (b) Ca^{2+}-bound active site. For clarity hydrogens and nucleotide bases except G1 (in the zoomed-in view) were not shown. G1 is away from divalent metal ion sites (M1, M2).

3.3.4 Active site of ligand-free group II intron with bound Mg^{2+}/Ca^{2+}

The cognate metal ion cluster (K1, K2, M2, and M1) and its local environment were stable during MD simulations and in excellent agreement with the X-ray structure (**Table 3.1**). Mg^{2+} (in M2 site) was stabilized by the 3 intron phosphates (residue C358, G359, and C377) and 3 water molecules (**Figure 3.6 a**). Mg^{2+} (in the M1 site) interacts with two intron phosphates (residue

C377, and U375) and 4 water molecules. Absence of exon part in the ligand-free state results in water exposure of M1 site. Substitution of Ca^{2+} in M1 and M2 sites of the ligand-free intron state, results in the entry of additional water molecules (**Figure 3.6 b, Table 3.1**), as seen in post-hydrolytic state (**Figure 3.5 b**). Maximum solvent exposer of M1, M2 sites are observed in the ligand-free state (**Table 3.2**) relative to its pre and post-hydrolytic analog. Monovalent ions (K1, K2) are hexa-coordinated in the active site of the intron that includes 5 direct interactions with the intron (backbone and sidechain) and 1 interaction with a water molecule (**Table 3.1**). The same interaction network is stable and independent of the nature of divalent (M1, M2: Mg^{2+} or Ca^{2+}) metal ions in the active site of group II intron (**Table 3.1**).

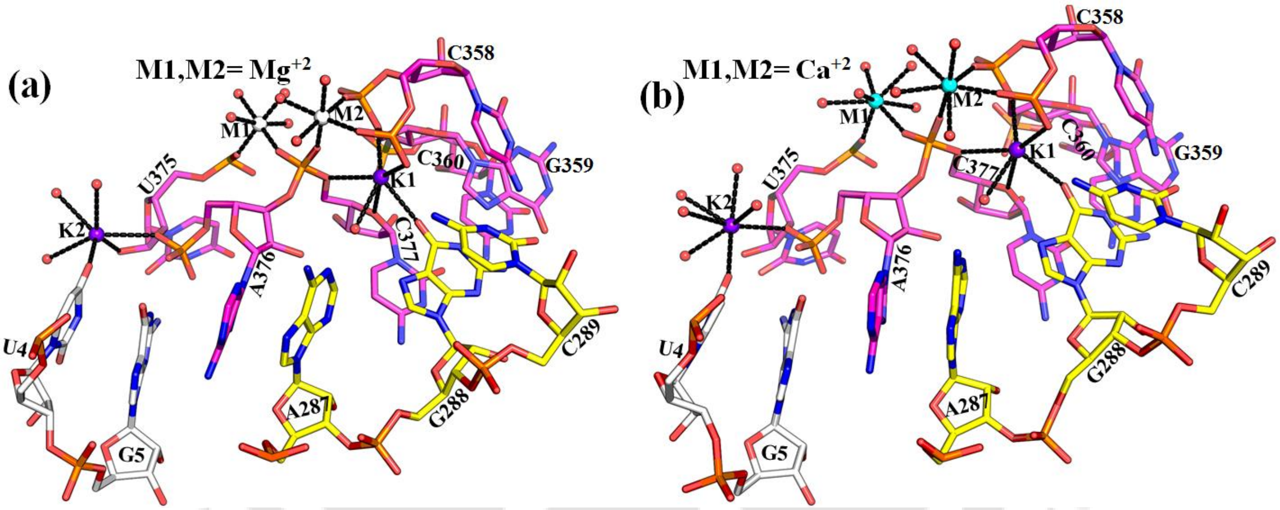


Figure 3.6. MD structures of ligand-free intron active site and waters around the divalent (M1, M2) ions. (a) Mg^{2+}-bound (b) Ca^{2+}-bound.

3.3.5 Energetics of Mg^{2+} vs. Ca^{2+} binding to group II intron active site

To compute the binding free energy difference of group II intron to Mg^{2+} and Ca^{2+}, we conducted classical molecular dynamics simulations in combination with thermodynamic integration (**Figure 3.2**) of group II intron in complex with M1=M2= Mg^{2+} or M1=M2=Ca^{2+} at different stages of splicing pathway (i.e, Pre, post, ligand-free). These calculations involve computing the change in binding affinity for divalent metal ions (M1, M2) in the active site of group II intron upon

Mg^{2+}/Ca^{2+} mutations in both the location (M1 and M2) (**see appendix Table 3.3**). Binding free energy difference of group II intron for Mg^{2+} and Ca^{2+}, favouring the former is summarized in **Figure 3.7**. Results indicate that group II intron imposes a very high energetic penalty of > 9.5 kcal/mol for Ca^{2+} binding at its active site (M1, M2 locations).

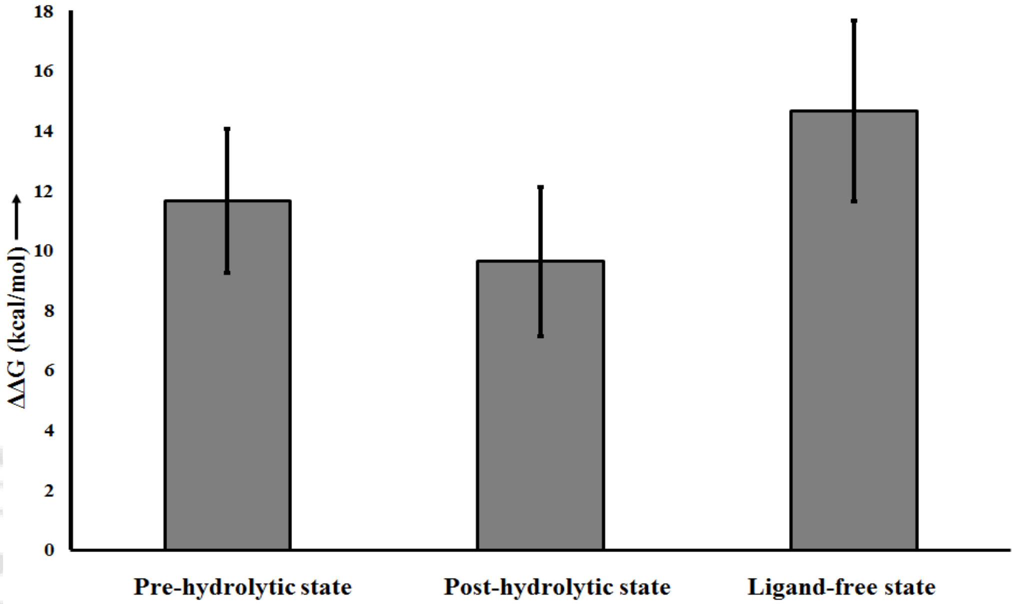


Figure 3.7. Relative binding free energy (ΔΔG) of divalent ion (Mg^{2+} vs Ca^{2+} in M1, M2 sites) binding to group-II intron at various stages of splicing. Free energies are in kcal/mol (error bars, 1 s.e.m).

The large discriminatory power of group II intron in favor of Mg^{2+} relative to Ca^{2+} suggests highly selective catalytic pocket. Ab initio single point energy difference between right and left-side of eqn (1) approximates the competition between Mg^{2+} and Ca^{2+} for M1, M2 intron site (**Figure 3.8**).

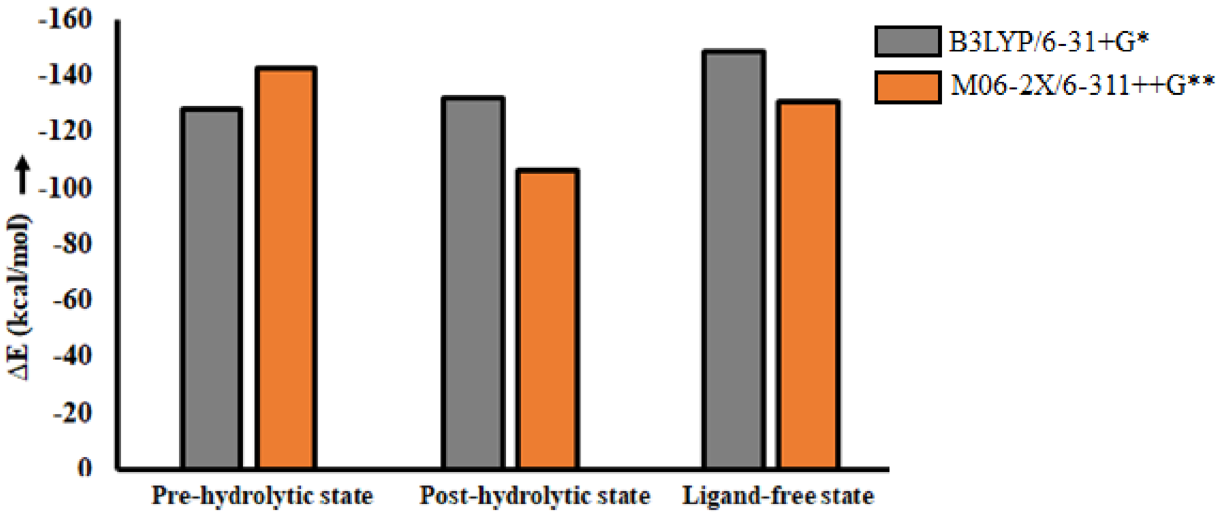


Figure 3.8. Quantum chemical single point energies (ΔE) calculated for divalent ion (Mg^{2+} vs. Ca^{2+} in M1, M2 sites) binding to group II intron at various stages of splicing. Energies are in kcal/mol.

The energy difference (in kcal/mol) at B3LYP/6-31+ g* level of theory for the pre, post, and ligand-free intron states are -128.3, -132.3, and -148.7 kcal/mol respectively (**see appendix Table 3.4**). M06-2X/6-311++G** calculations estimated the energy difference of -142.3, -106.1, and -130.7 kcal/mol for the pre, post, and ligand-free intron states respectively. Large negative energy difference implies Mg^{2+} is favored at M1, M2 sites relative to Ca^{2+} in the intron active site at various stages of the splicing pathway.

3.4. Discussion

Based on the X-ray structures of Ca^{2+} bound Pre-hydrolytic and Mg^{2+} bound post/ligand-free states, MD simulations were performed for both Mg^{2+} and Ca^{2+} bound group II introns at various stages of the splicing pathway. Mg^{2+} (in M1 and M2 sites) of pre-hydrolytic state coordinates with only two waters and rest four coordination with the intron (**Table 3.2, Figure 3.4 a**). Scissile phosphate is directly coordinated to two divalent metal ions (M1, M2) in the pre-hydrolytic state (**Figure 3.4 a**), thus, expected to increase the partial positive charge on the phosphorous, resulting facile nucleophilic attack. M1 holds the nucleophilic water (**Figure 3.4 a**). Mg^{2+} bound post-hydrolytic state shows an increase in water coordination at the M2 position relative to the pre-hydrolytic analog (**Table 3.2**). After hydrolysis, the scissile phosphate-M2 interaction (as in the pre-hydrolytic state, **Figure 3.4 a**) is lost, which is compensated by the recruitment of water molecule (**Figure 3.5 a**). The local environment of M1 is essentially the same for pre and post-hydrolytic states (**Table 3.2, Figure 3.4 a and 3.5 a**). U0 and G1 are parts of exon and intron respectively. Interaction between M1 and ribose hydroxyl (2' and 3') of U0 (**Figure 3.4 a and 3.5 a**) suggests that M1 might be the key in positioning the exon part correctly during reverse-splicing. Total waters around Mg^{2+} (M1 and M2) in the ligand-free state is maximum (Total seven waters, **Table 3.2**). Based on the presence of waters around Mg^{2+} (M1 and M2, **Table 3.2**) we may conclude that the pre-hydrolytic state is dry and ligand-free state is wet. The substitution of Ca^{2+} into the intron active site (M1, M2) increases the wetness (**Table 3.2, appendix Figure 3.2**). The large size of Ca^{2+} relative to Mg^{2+} opens up divalent metal ion binding pocket. Furthermore, in the pre-catalytic state Ca^{2+} (M2) is away from scissile phosphate and the orientation of associated nucleophilic water altered related to its Mg^{2+} analog. It can be argued that the drastic alteration of the local environment around M2 might be responsible for hindering catalysis. Averaged RMSF (root mean square fluctuation) is a popular choice to access the flexibility of biomolecules. Average RMSF (**Figure 3.9**) of nucleotide residues of group II intron suggests that the RNA flexibility of Mg^{2+} and Ca^{2+} bound group II intron at various states e.g. pre-hydrolytic, post-hydrolytic, and ligand-free state appears to be similar with a marginal difference in most residues. It should be noted that U0-G1 is the exon-intron boundary in the pre-hydrolytic state. After hydrolysis i.e in the post-hydrolytic state, the highest flexibility of the intron part (G1) was evident from the average RMSF (**Figure 3.9**).

Recent QM/MM molecular dynamics simulations (**Casalino et al., 2016**) propose that the proton of the nucleophilic water is taken up by the bulk water and activates its attack on the scissile phosphate. Classical molecular dynamics free energy calculations suggest that Mg^{2+} is strongly preferred related to Ca^{2+} at various stages of the splicing pathway (**Figure 3.7**). Large selectivity ($\Delta\Delta G > 9.5$ kcal/mol, **Figure 3.7**) could be utilized for preferential dissociation of Ca^{2+} from the catalytic pocket (M1, M2) of group II intron. Higher charge density on Mg^{2+} relative to Ca^{2+} expected to result in tighter binding of the former in the negatively charge pocket. Thus, the use of polarizable force fields (**Jing et al., 2018**) may be more appropriate in the evaluation of the relative binding free energy between Ca^{2+} and Mg^{2+}, which could explicitly include many-body polarization effects. Quantum chemical calculations, also support Mg^{2+} preference over Ca^{2+} in the intron active site (**Figure 3.8**).

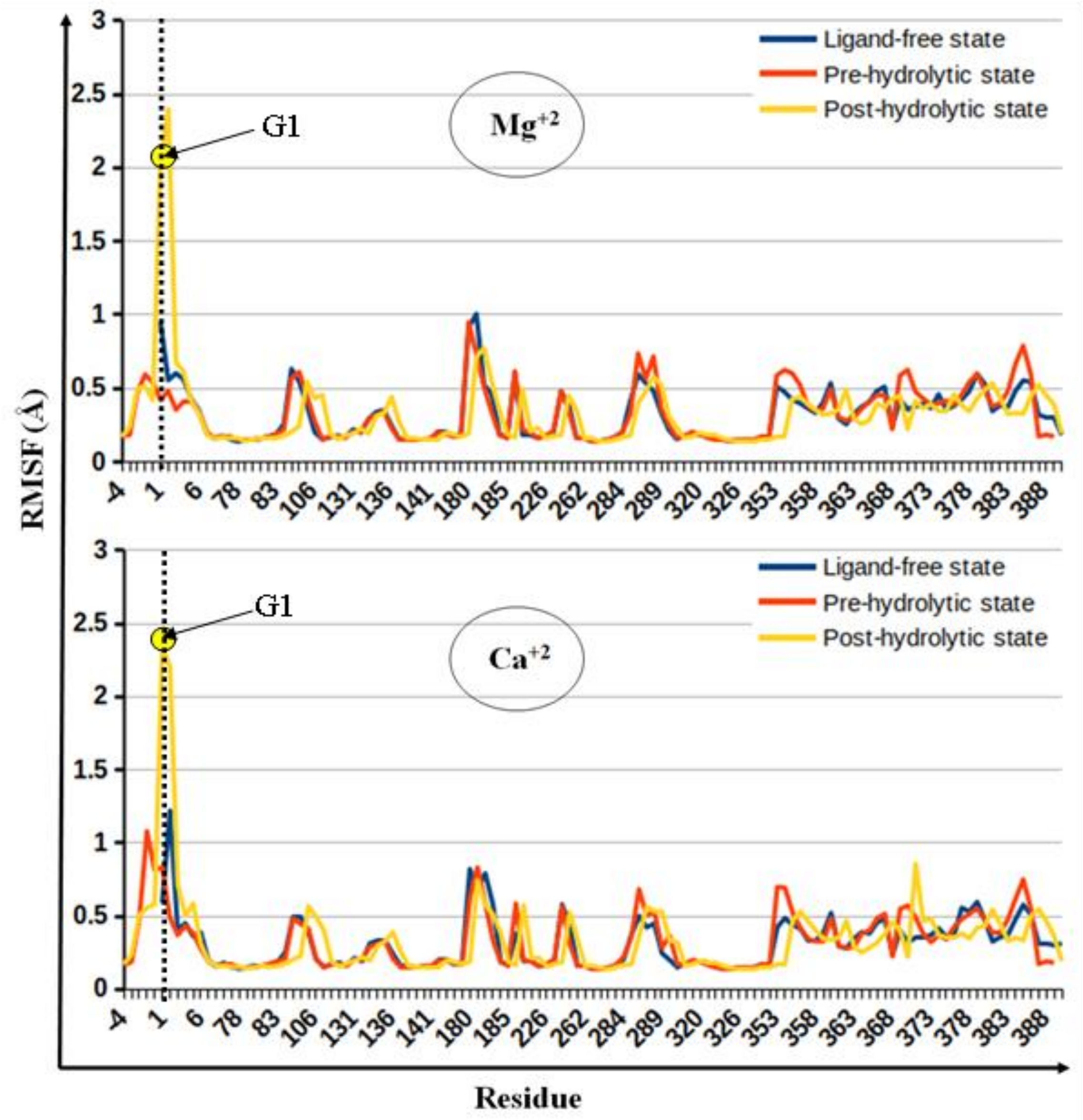


Figure 3.9. (a) Average root-mean-square fluctuation (RMSF) of the heavy atoms of RNA nucleotides of group II introns. Graphs are shown for M1=M2=Mg^{2+} bound (upper) and M1=M2=Ca^{2+} bound (lower) group II intron. Average RMSF was computed from the last 2 ns of post equilibrated MD trajectory. Right-side (intron-part) and left-side (exon-part) of vertical dotted arrow. G1 of the post-hydrolytic state is marked explicitly.

3.5. Conclusion

Intron active site for divalent metal ions (M1, M2) requires efficient discrimination between Mg^{2+} and Ca^{2+}. Our computer simulations show that intron imposed very high energy penalty for Ca^{2+} for M1, M2 sites. Mg^{2+} bound active site is compact and dry. Substitution of Ca^{2+} results in water entry (wet) and loosen up the M1, M2 pocket due to the larger size of Ca^{2+}. The orientation of nucleophilic water relative to scissile phosphate is strikingly different between Mg^{2+} and Ca^{2+} bound pre-hydrolytic state. Furthermore, the scissile phosphate oxygens form direct interactions with two Mg^{2+} (M1, M2) in the pre-hydrolytic state. Ca^{2+} substituted pre-hydrolytic state suggests only one direct interaction between scissile phosphate oxygen and Ca^{2+} (M1). Mg^{2+} or Ca^{2+} bound post-hydrolytic state suggests a direct interaction between the exon (U0) and M1, whereas, the intron part (G1) is highly dynamic in the MD trajectories. The absence of exon part in the ligand-free state allows maximum water exposure of M1, M2 sites.

Chapter 4

Lariat intron and metal ion selectivity (Mg^{2+} versus Ca^{2+}, K^+ versus Na^+)

Group II intron ribozymes are metalloenzymes that can catalyze self-splicing. Recently, the crystal structures of eukaryotic group IIB lariat intron from a brown algae *Pylaiella littoralis* have been reported for two intermediate states (pre-hydrolytic (2s) and post-hydrolytic) along the self-splicing pathway. Three characteristic metal ion binding sites (M1, M2 sites for catalytic Mg^{2+} ions, and K1 site for K^+) in the catalytic pocket of the lariat intron have been identified and proposed to be crucial for self-splicing. Using the X-ray structures as a template, we quantitatively estimated the energetics of divalent (Mg^{2+} versus Ca^{2+}) and monovalent (K^+ versus Na^+) ion selectivity in the active site of lariat intron and established a direct link between the energetics and structures of lariat intron (bound to cognate and near-cognate metal ions). The strength of divalent metal-ion selectivity was noticeably high in the post-hydrolytic state ($\Delta\Delta G \sim 20$ kcal mol^{-1}) relative to its pre-hydrolytic (2s) state ($\Delta\Delta G \sim 13$ kcal mol^{-1}). Quantum chemical calculations ensured that the sign of the estimated divalent metal ion selectivity was correct. M1- binding pocket was relatively less solvent-exposed in the case of post-hydrolytic state relative to pre-hydrolytic (2s) state, which boosted the metal ion selectivity for the former. Surprisingly, lariat intron was found to be non-selective between monovalent ions (K^+ versus Na^+), in contrast to bacterial linear group II intron, which was previously reported to be selective between monovalent ions in the pre-hydrolytic state ($\Delta\Delta G \sim 6$ kcal mol^{-1}). The interaction network in the first coordination shell of Ca^{2+} in the M1 binding pocket was different relative to Mg^{2+}. Mg^{2+} substitution by Ca^{2+} in the M1 site resulted in the substitution of a single M1-RNA interaction by M1-water interaction. In the pre-hydrolytic (2s) state, Ca^{2+} substitution completely disrupted the M1…5'-exon interaction; thus, the nature of divalent metal ion seemed to be critical for catalysis, in line with the general experimental observation related to group II intron. The interaction network in the M2 site was found to be independent of the nature of divalent metal ions (Mg^{2+} or Ca^{2+}). The Monovalent ion was loosely bound in the wet binding pocket (K1 site) of the lariat intron; thus, the substitution of K^+ by Na^+ could not significantly alter the free energy of the complex, resulting in weak discrimination.

4.1 Background

The branched self-splicing pathway of group II intron is a two-step process (**Figure 4.1a**) (**van der Veen et al., 1986; Peebles et al., 1986**). In the first step, the nucleophilic attack of adenosine to the 5'-splice site yields a lariat-3'-exon intermediate and 5'-exon (Pre-hydrolytic state). The 3'-OH of free 5'-exon attacks the 3'-splice-site (lariat-3'-exon intermediate) in the second step, resulting in lariat intron and ligated exons (Post-hydrolytic state). Errors in splicing can alter the ligated exon and cause diseases in humans (**Cooper et al., 2009; Scotti et al., 2016**). Recently, X-ray structure (resolution ~ 3.7Å) of the eukaryotic group II intron from the brown algae *Pylaiella littoralis* (*P.li.LSUI2*) in the pre-hydrolytic (2s) (PDB 6CHR) and post-hydrolytic (PDB 6CIH) state has been reported (**Chan et al., 2018; Robart et al., 2014**). Active site architecture of group II intron is highly conserved and functionally similar to the higher eukaryotes that catalyze splicing using a large protein-RNA complex called the spliceosome (**Fica et al., 2013; Galej et al., 2016; Hang et al., 2015; Bertram et al., 2017**). The catalytic group II intron consists of six domains (Domain 1-6, **Figure 4.1b**). Domain 6 contains the bulged adenosine nucleophile, and domain 5 forms the active site for splicing after binding to divalent (M1 and M2 sites, **Figure 4.1b**) and monovalent (K1, **Figure 4.1b**) metal ions. The compact special arrangement of the negatively charged RNA backbone creates the hetero-nuclear ion cluster (M1, M2, and K1 **Figure 4.1b**) in the active site of the lariat intron. X-ray structures (PDB 6CHR, 6CIH) contain Mg^{2+} (M1 and M2 sites) and Na^+ (K1 site) in the active site of the lariat intron.

Group II introns were broadly classified into three classes: IIA, IIB, and IIC (**Michel et al., 1989; Toor et al., 2001**). However, X-ray structures were reported for group IIC intron from bacterium *Oceanobacillus iheyensis* (O. iheyensis) (**Toor et el., 2008; Costa et al., 2016**) and eukaryotic group IIB intron from algae *Pylaiella littoralis* (*P.li.LSUI2*). Most primitive group IIC class undergo splicing through hydrolysis and release linear intron as the product (hydrolytic mechanism) (**Sigel et al., 2000**). In contrast, group IIB introns release lariat intron and evolutionary close to the spliceosome.

Mg^{2+} and K^+ were reported to be the cognate metal ions in the active site of group II intron. Thus, metal-ion selectivity is crucial for splicing in the group II intron (**Sigel et al., 2000; Gordon et al., 2001; Toor et al., 2006**). Previously, we have reported (**Kumar et al., 2018; Kumar et al., 2020**) structure-based computational study and estimated the strength of metal-ion selectivity (K^+ versus Na^+, Mg^{2+} versus Ca^{2+}) in the active site of bacterial group IIC intron (*Oceanobacillus iheyensis*) at various stages of the hydrolytic splicing process. The results showed that group IIC intron was strongly selective for monovalent (K^+ favored over Na^+) as well as divalent (Mg^{2+} favored over Ca^{2+}) metal ion in the active site. Moreover, the strength of metal-ion selectivity was depended on the solvent accessibility of the metal-ion binding pocket; dry-pocket enhanced selectivity.

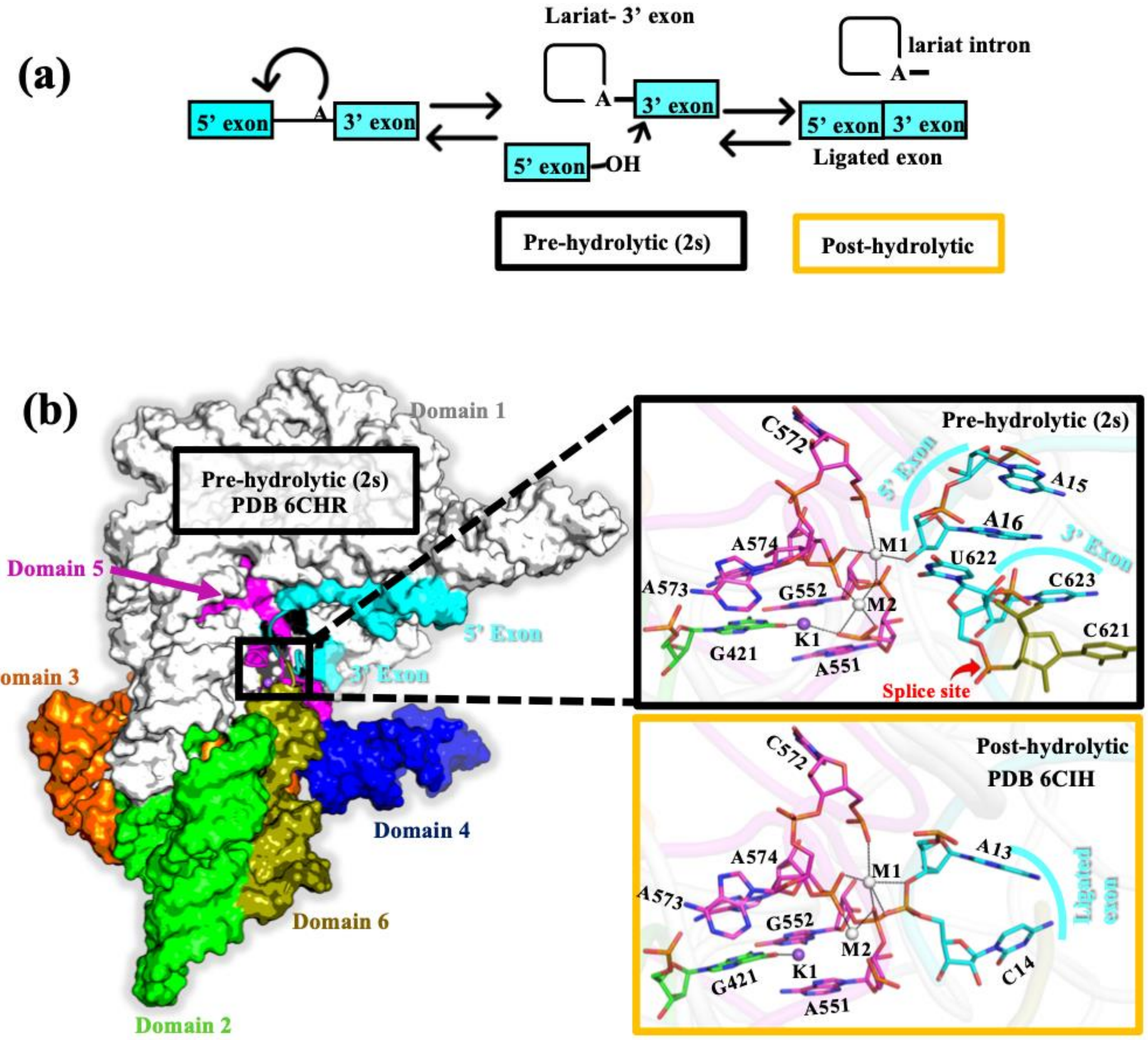


Figure 4.1. (a) Scheme of splicing pathway in the lariat group II intron. (b) X-ray structure of lariat-3' exon with the 5'-exon in the pre-hydrolytic state (2s) (PDB 6CHR) shown in surface representation. Color coding: Different regions of group II intron (D1: grey, D2: green, D3: orange, D4: blue, D5: magenta, D6: olive), 5'-exon and 3'- exon: cyan, Monovalent Na^+ ion (K1 site): purple sphere, and Divalent ions (M1, M2 sites) are Mg^{2+}: grey sphere. Zoomed-in view of the

catalytic pocket in the pre-hydrolytic (2s) (black rectangle) and post-hydrolytic state (orange rectangle, taken from PDB 6CIH).

Recently reported near-atomic resolution (~3.7 Å) X-ray structures (pre-hydrolytic (2s): PDB 6CHR, post-hydrolytic: PDB 6CIH) (**Chan et al., 2018**) of the lariat group IIB intron (*P.li.LSUI2*) not only revealed the atomic interaction network in the active site but also provided the reasonable model for computational analysis. The thermodynamics of metal-ion selectivity in the active site of the lariat intron is unknown. Using X-ray structures (PDB 6CHR, 6CIH) as a template, we modelled Mg^{2+}/Ca^{2+} (in the M1 and M2 sites) and K^+/Na^+ bound (K1 site) in the active site of the lariat group II intron and subjected those models to classical MD simulations and quantum chemical calculations.

In this chapter, we have estimated the relative divalent (Mg^{2+} versus Ca^{2+}) and monovalent (K^+ versus Na^+) binding affinities (ΔΔG) using classical MD simulations and attempted to establish a direct link between the estimated energetics with 3D structures. The sign of the MD estimated divalent metal ion selectivity (ΔΔG) was shown to be correct by employing quantum chemical calculations, which included polarization effect explicitly. The results obtained for the lariat group II intron were compared with the previous findings related to linear group II intron. The results showed that the lariat group II intron of *P.li.LSUI2* was strongly selective for divalent metal-ion (favoring Mg^{2+} relative to Ca^{2+} in M1, M2 sites), but more or less unselective between monovalent ions (K^+ and Na^+ in the K1 site) contrary to the linear intron. The strength of the metal-ion discrimination by lariat intron was found to be correlated with the solvent accessibility as reported previously for the linear intron.

4.2 Computational methods

Classical MD simulations were used to estimate the relative binding free energy (cognate vs. non-cognate). Overall, a total of ~2 μs of MD simulations (over multiple independent replicas) were performed to estimate the free energy difference. In order to check if the sign of MD estimated relative binding free energy is correct, ab initio quantum calculations were performed (that considered the polarization effects explicitly) by considering reduced metal-ion binding pockets from MD structures.

4.2.1 Molecular Dynamics Simulations setup

X-ray structures (**Chan et al., 2018**) were taken from PDB. Missing atoms were modelled, and a 25Å radius spherical region (cantered at P atom of residue ADE 576, metal ion binding pocket) was truncated and then solvated with a cubic water box with 80 Å of edge length (**see details in Chapter 1 method section**).

4.2.2 Relative Binding Free Energy Calculation

Binding free energy difference between Mg^{2+}/Ca^{2+} and K^{+}/Na^{+} in the catalytic pocket was computed by alchemically transforming $Mg^{2+} \rightarrow Ca^{2+}$ and $K^{+} \rightarrow Na^{+}$ in complex with RNA and free in the water, as described in the thermodynamic cycle (**Figure 4.2**). Vertical legs correspond to divalent metal ion binding; horizontal legs correspond to the alchemical transformation of $Mg^{2+} \rightarrow Ca^{2+}$ in the M1, M2 binding sites, either in the RNA bound (above) or free (below) in solution. MD simulations and thermodynamic integration have computed free energy associated with the horizontal legs of the thermodynamic cycle. The binding free energy difference is $\Delta\Delta G = \Delta G_{comp} - \Delta G_{free} = \Delta G_{bind}(Ca^{2+}) - \Delta G_{bind}(Mg^{2+})$; the same methodology was used to compute the energetics for $K^{+} \rightarrow Na^{+}$ transformation. The detailed methodology of free energy estimation adopted in this work has described previously in **Chapter 1**.

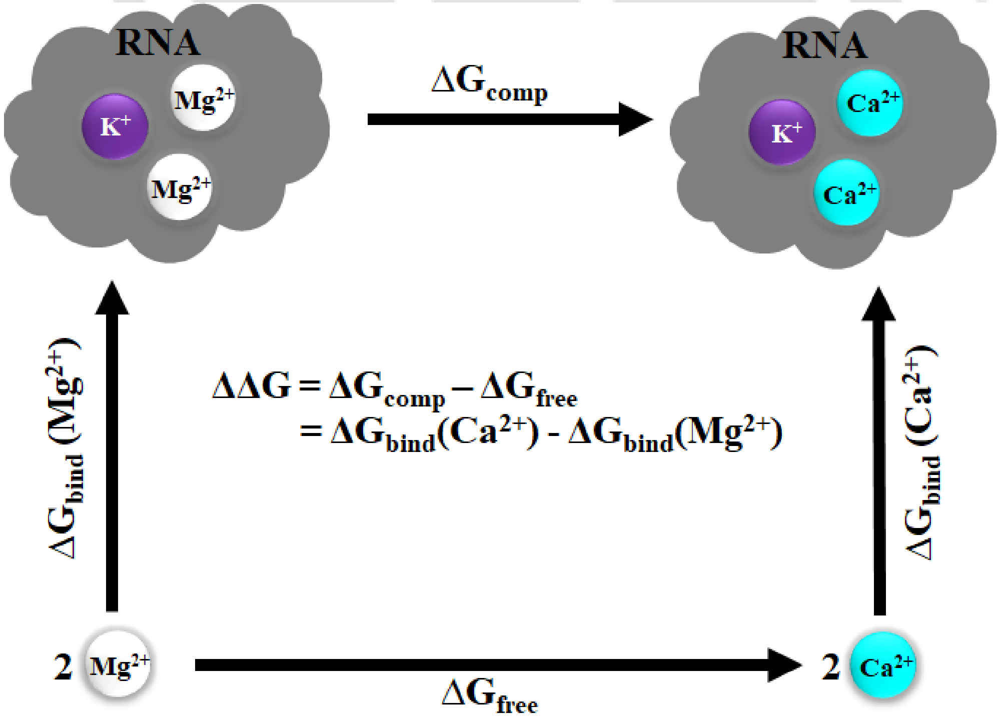


Figure 4.2. Thermodynamic cycle for Mg^{2+} vs. Ca^{2+} ion binding to the divalent ion binding sites M1, M2 of lariat group II intron active sites.

4.2.3 Ab Initio Quantum Calculations

Classical Force-field does not include explicit polarizability effect; thus, estimation of relative binding free energies (ΔΔG) using classical MD simulations may not be appropriate, especially for divalent ions for which many-body effect is known to be crucial. Thus, in order to check if the sign of the estimated ΔΔG is correct, we estimated energies using ab initio calculations of the reduced metal-ion binding pocket. The goal was not to obtain an extremely precise estimate of ΔΔG from electronic structure computations but to concentrate on a quantity that may be used to indicate divalent-metal-ion preference in a computationally efficient manner. The simplified binding pocket model was embedded in an implicit aqueous dielectric and was used to calculate density functional energies at two different theoretical levels: (1) the M06-2X/6-311++G** (**Zhao et al., 2008**) and (2) B3LYP/6-31+G* (**Becke 1988; Lee et al., 1988**). The ab initio calculations were performed using the Gaussian16 programme (**Frisch et al., 2016**). By employing the following equation (eq. 4.1), we concentrated on a quantity that approximates the relative preference (Mg^{2+} vs. Ca^{2+}) at M1, M2 in the intron active site:

$$[2Mg^{2+} - water] + [2Ca^{2+} - intron] \rightarrow [2Ca^{2+} - water] + [2Mg^{2+} - intron] \quad (4.1)$$

In eq (4.1), $[2Ca^{2+} - intron]$ and $[2Mg^{2+} - water]$ denote Ca^{2+} binding (at M1, M2 sites) intron active site and intron free Mg^{2+} embedded into the implicit aqueous dielectric. The differences in single-point energy between the right-hand and left-hand sides of eq (4.1) approximate the relative preference for divalent cations in the intron active sites M1, M2. A positive energy difference from eq (4.1) indicates a Ca^{2+} preference at the intron active site, whereas a negative value indicates the Mg^{2+} preference.

4.3 Results

Classical molecular dynamic (MD) simulations can efficiently sample the phase space of the large biomolecules, thus a popular choice for estimating free energies. Metal ion recognition is a key component of biological function and structural stability (**Schnabl et al., 2010; Dudev et al.,**

2013). In this chapter, we used classical MD and quantum chemical calculations to estimate the monovalent (K^+ versus Na^+) and divalent (Mg^{2+} versus Ca^{2+}) metal ion selectivity at two states (pre-hydrolytic (2s) and post-hydrolytic) of eukaryotic lariat group II intron along the splicing pathway. Based on the X-ray structures, we first modeled cognate (Mg^{2+}, K^+) and near-cognate (Ca^{2+}, Na^+) metal ion bound lariat intron and carried out classical MD simulations to those models. Then we estimated the binding free energy difference between the cognate and near-cognate metal ions in the group II intron and established the link between the estimated energetics with the biomolecular structures (cognate and near-cognate metal ion bound intron). Ions were modelled with the fixed parameters (charge, van der Walls) in the classical force field that does not explicitly include polarization. The absence of explicit polarization in the classical forcefield imposed challenges, especially for divalent metal ions (viz., Mg^{2+}, Ca^{2+}, Mn^{2+}) (**Kumar et al., 2021; Jing et al., 2018; Hu et al., 2011; Neves et al., 2013; Banci, 2003**). Polarizable force-field seemed to be necessary for the correct modeling of divalent ions in the negatively charged protein binding pocket (**Jing et al., 2018**). Instead of using computationally expensive polarizable force fields, we performed simple density functional theory calculations on the reduced ion binding pocket models (using MD structures) and approximately evaluated the divalent metal ion preference (Mg^{2+} versus Ca^{2+}) in the lariat intron.

4.3.1 Energetics of metal-ion selectivity by lariat group II intron

To estimate the metal-ion discrimination strength by lariat intron quantitatively, we performed molecular dynamics free energy simulations of cognate (Mg^{2+}, K^+) and non-cognate (Ca^{2+}, Na^+) metal-ion-lariat intron complexes, where the Mg^{2+}/K^+ ion was alchemically transformed into Ca^{2+}/Na^+ in the intron as well as free in the water. These calculations used crystallographic structures (**Chan et al., 2018**) of eukaryotic group IIB lariat intron (from a brown algae *Pylaiella littoralis (P.li.LSUI2)*) at the pre-hydrolytic (2s) and post-hydrolytic state of the splicing pathway (**Figure 4.1a,b**), as the starting structures. MD estimated relative binding free energies were summarized in **Figure 4.3**. It was evident that lariat intron imposed a very high energetic penalty against Ca^{2+} (ΔΔG; Pre-hydrolytic (2s) state ~13 kcal mol^{-1} and Post-hydrolytic state ~20 kcal mol^{-1}), thus, strongly favoring Mg^{2+}.

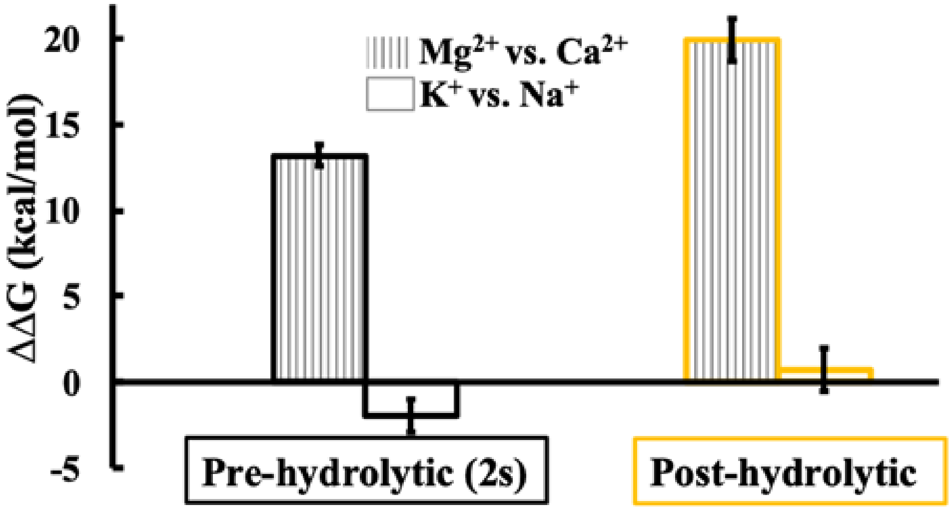


Figure 4.3. Estimated relative binding free energy (ΔΔG) of the divalent (Mg^{2+} vs. Ca^{2+} in M1, M2 sites) and the monovalent (K^+ vs. Na^+ in K1 site) ion binding to lariat group II intron at pre-hydrolytic (2s) and post-hydrolytic states of splicing. Free energies are in kcal mol^{-1} (error bars, 1 s.e.m). The postive value of ΔΔG indicates Mg^{2+} (K^+) is favoured relative to Ca^{2+} (Na^+) binding in the active site of the lariat intron. The error was calculated by propagating the standard error of the mean associated with the averaged ΔG^{comp} and ΔG^{free}.

Indeed, the catalytic activity of group II introns was reported (**Schnabl et al., 2010; Fedorova et al., 2007**) to be dependent on the nature of the metal ion, significantly compromised if the Mg^{2+} (cognate) ion was replaced with near-cognate Ca^{2+}, in line with the estimated binding affinity differences. Surprisingly, lariat intron was found to be weakly selective between monovalent ions (K^+ and Na^+). Moreover, the pre-hydrolytic (2s) state weakly favored Na^+ relative to the K^+ ion (ΔΔG ~ -2 kcal mol^{-1}). Previously we reported the metal-ion selectivity for linear group II intron of bacterium (*Oceanobacillus iheyensis*) and the estimated strength of divalent metal-ion selectivity (ΔΔG: between Mg^{2+} and Ca^{2+}, favoring former) for the pre-hydrolytic and post-hydrolytic states were ~12 kcal mol^{-1} and ~10 kcal mol^{-1} respectively (**Kumar et al., 2018; Kumar et al., 2020**). Thus, the discriminatory power of lariat-intron (favoring Mg^{2+} relative to Ca^{2+} binding) was found to be higher than the linear intron in bacteria, particularly in the post-hydrolytic state. On the other hand, the pre-hydrolytic state of linear group II intron of *Oceanobacillus iheyensis* was found to be strongly selective between K^+ and Na^+ (ΔΔG ~ 6 kcal mol^{-1} in favor of K$^+$) in comparison to the lariat group II intron of *Pylaiella littoralis* (ΔΔG ~ - 2 kcal mol^{-1} in favor of Na$^+$). The post-hydrolytic state of both the linear and lariat group II intron discriminated weakly between K$^+$ and Na$^+$, preferred the former by only ~ 1 kcal mol^{-1}.

Recently, it has been shown that the many-body polarization effect governed the Mg^{2+} versus Ca^{2+} specificity in the negatively charged pocket of calcium-binding proteins (**Jing et al., 2018**). Despite higher charged density on Mg^{2+}, negatively-charged pockets of calcium-binding proteins can selectively bind to Ca^{2+} against a relatively higher concentration of Mg^{2+} (**Dudev et al., 2013; Gifford et al., 2007**). Thus, consideration of explicit polarization effect might be necessary for estimating the binding preference. Note classical force field does not include polarization effect explicitly, thus the magnitude and sign of the estimated divalent metal binding free energy differences (ΔΔG, Mg^{2+} versus Ca^{2+}) in the negatively charged pocket of lariat intron using classical MD simulations were questionable. Thus, in order to consider the polarization effect in a computationally inexpensive way, we performed ab initio calculations on the reduced metal-ion binding pocket ($^\#$) of lariat group II intron. Single-point energy calculation was performed and the energy difference (ΔE) between the right- and left- sides of eqn. 4.1 were estimated using two level of theories (i.e., B3LYP/6-31+G* and MO6-2X/6-31++G**). Note ΔE included the polarization effect explicitly and approximately estimated the Mg^{2+} versus Ca^{2+} preference in the catalytic pocket of lariat intron. The negative sign of ΔE implied that Mg^{2+} was preferred over Ca^{2+} in the lariat intron. The estimated ΔE's are given in **Figure 4.4**.

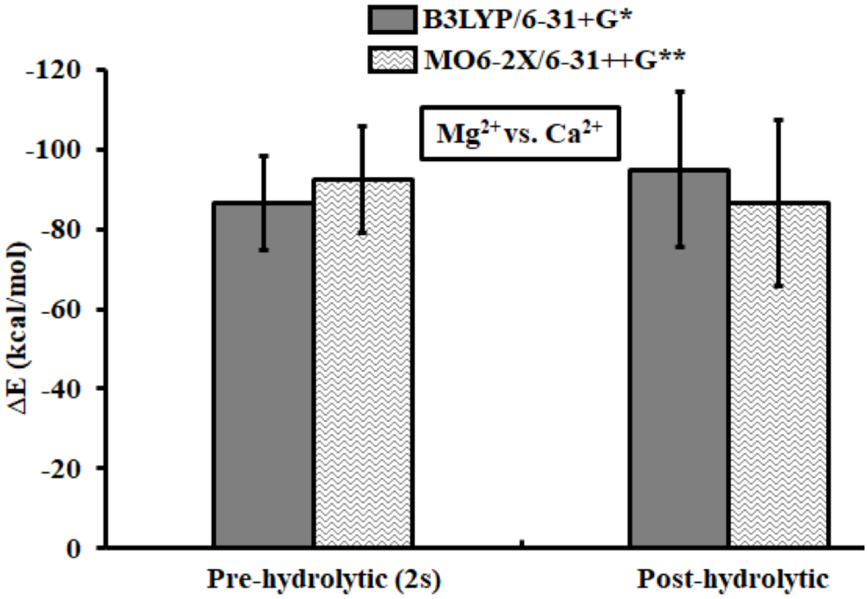
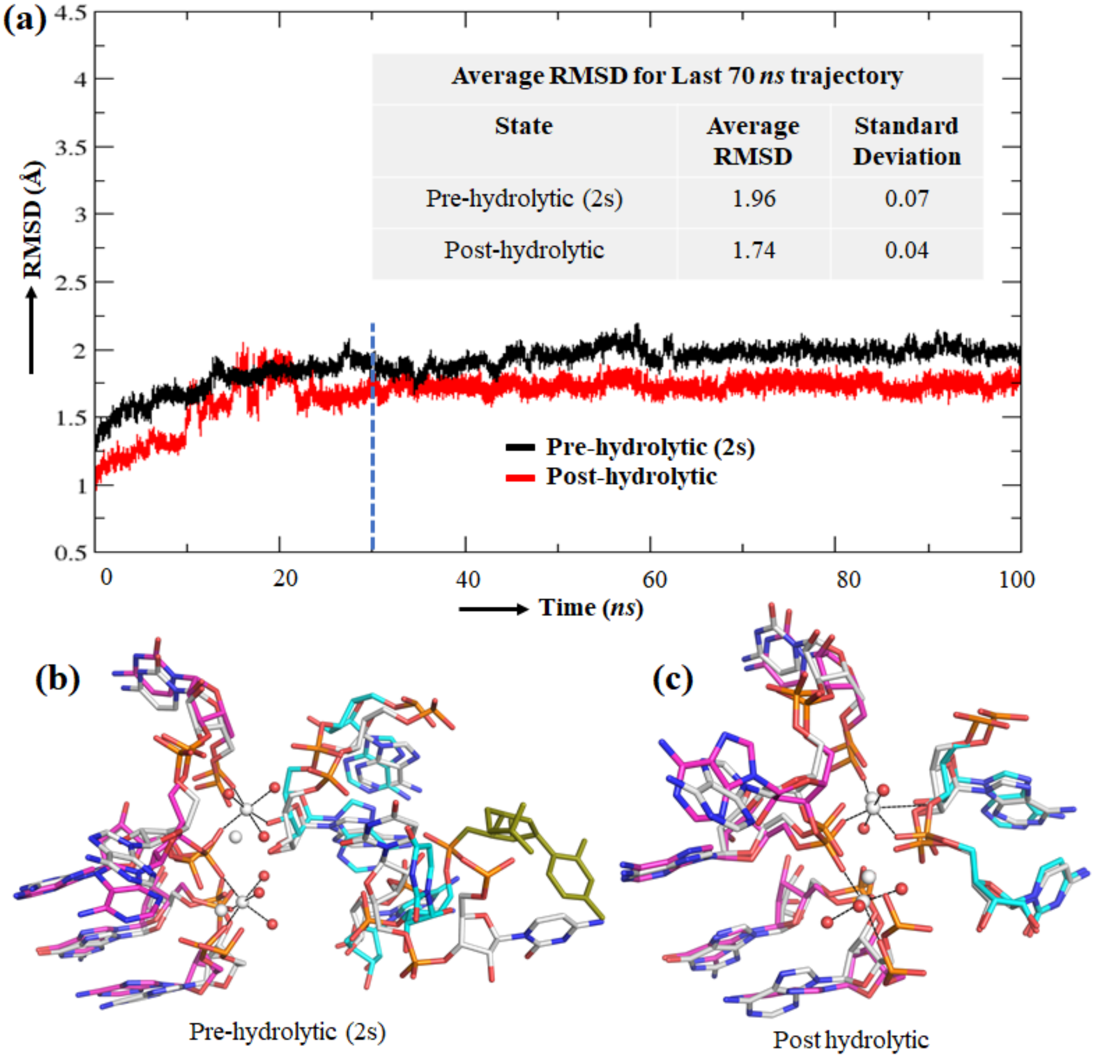


Figure 4.4. Single point relative energies (ΔE, see eqn. 4.1) estimated from quantum chemical calculations for the reduced catalytic pocket model (indicated by *) of the pre-hydrolytic (2s) and post-hydrolytic states by employing two levels of theories (M06–2X/6-311++G** and B3LYP/6-31+G*). ΔE approximately captures the divalent metal ion preference in the lariat group II intron. The negative sign of ΔE confirms that lariat intron prefers Mg^{2+} binding relative to Ca^{2+}.

The large negative ΔE's confirmed that Mg^{2+} was energetically favored relative to Ca^{2+} in the reduced pocket of lariat intron, thus, ensuring that the sign of the estimated ΔΔG's was correct. Note ΔE was just an indicator (limited by the small size of the reduced binding pocket that does not consider entropy effect) of metal-ion preference, which can capture the free energy to some extent. However, a simple indicator (viz., ΔE of eqn. 4.1) was computationally inexpensive for ensuring if the sign of ΔΔG estimated from MDFE was correct, thus, bypassing the use of computationally expensive polarizable force-field for free-energy estimation. The same methodology (estimating ΔΔG using MD and ΔE using quantum chemical calculations) was successfully used in similar contexts, including ion selectivity by linear group II intron (**Kumar et al., 2020**) and Cas1 protein (**Kumar et al., 2021**).

4.3.2 Comparison MD and X-ray Structure

To compare the overall structure of the cognate ion (Mg^{2+}, K^+) bound lariat intron between template X-ray and MD structures, we estimated root-mean-square deviation (RMSD) for the heavy atoms and given in **Figure 4.5**. The observed plateau after 30 ns of MD implied that the structural convergence was attained. Small average RMSD of 1.74Å (post-hydrolytic) and 1.96Å (pre-hydrolytic (2s)) with a standard deviation of maximum 0.07Å indicated that the template X-ray structure closely resembled the MD structures; thus, confirming the sampling of the desired minima in the energy hypersurface. To access the flexibility, residue-wise root-mean-square fluctuation (RMSF) was estimated from the last 50 ns of 100 ns MD trajectory (**Figure 4.5**).



Average RMSD for Last 70 *ns* trajectory		
State	**Average RMSD**	**Standard Deviation**
Pre-hydrolytic (2s)	1.96	0.07
Post-hydrolytic	1.74	0.04

Figure 4.5. RMSD vs. Time plot from 100ns trajectory: (a) Root-mean-square deviation of heavy atoms within 22Å radius of the active site pocket of lariat group II intron relative to template X-ray structures. Pre-hydrolytic (2s) (black), post-hydrolytic (red). Structural comparison [MD (coloured) and X-ray (Grey)]: (b) Pre-hydrolytic (2s), (c) post-hydrolytic.

The RMSF (**Figure 4.6**) illustrated that the pre-hydrolytic (2s) was relatively more flexible than the post-hydrolytic state, particularly at the intron terminals. Very high B-factor (average > 300) associated with exon-intron boundary (residue 619-624) and unresolved G1 in the X-ray structure of the pre-hydrolytic (2s) state was in line with the observed high flexibility in the MD trajectory. The divalent metal ion binding pocket residues of lariat intron were found to be rather rigid (RMSF

~ 1 Å), and the flexibility was found to be more or less identical between the pre-hydrolytic (2s) and post-hydrolytic state of the lariat intron (**Figure 4.6**).

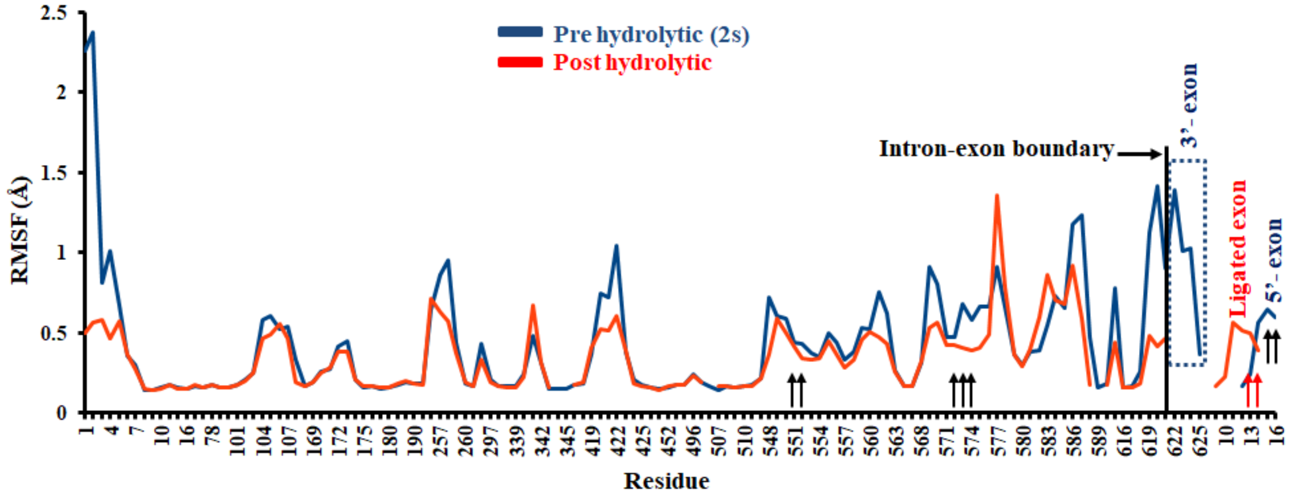


Figure 4.6. Root-mean-square fluctuation (RMSF) of the heavy atoms of RNA nucleotides of lariat group II introns. Average RMSF was computed from the last 50 ns of post equilibrated MD trajectory. Right-side (exon-part) and left-side (intron-part) of vertical line (black line). The metal-ion binding pocket residues are marked with black/red arrows.

4.3.3 Mg^{2+} and Ca^{2+} bound (M1, M2 site) pre-hydrolytic (2s) lariat group II intron

Overall the MD structure (Mg^{2+} bound) was in good agreement with the crystal structure (**Figure 4.5**). However, the distance between the two Mg^{2+} ions (M1 and M2 site) in the MD structure was found to be increased (~5 Å) relative to the X-ray structure (3.7 Å, PDB 6CHR). Moreover, the 3'-exon part was found deviated significantly from the experimentally reported location (**Figure 4.5 b**). The precise location of the 3'-exon part (U622, C623, and A624) in the X-ray structure (PDB 6CHR) was limited by the poor resolution (High average B factor of ~300), thus, explaining the distinctly different orientation of the 3'-exon observed in the MD structure.

Mg^{2+} in the M1 and M2 sites satisfied its 6 coordination by involving 3 water molecules, 2 phosphate-oxygens (C572 and A574), and one ribose hydroxyl (A16) of the RNA (**Figure 4.7**). Ca^{2+} substitution in the place of Mg^{2+} (**Figure 4.7 b**) resulted in (1) disruption of M1…A16 interaction (2) single additional water entry in the M1 site. The average interaction distances between the divalent metal ion (Mg^{2+}, Ca^{2+}) and the coordinating atoms in the first coordination shell around M1 (**Figure 4.7 c**) and M2 site (**Figure 4.7 d**) showed the increase in the inter-atomic distances (due to the obvious larger size of Ca^{2+}). Note the interaction network in the first coordination sphere in the M2 site was independent of the nature of the divalent metal ion.

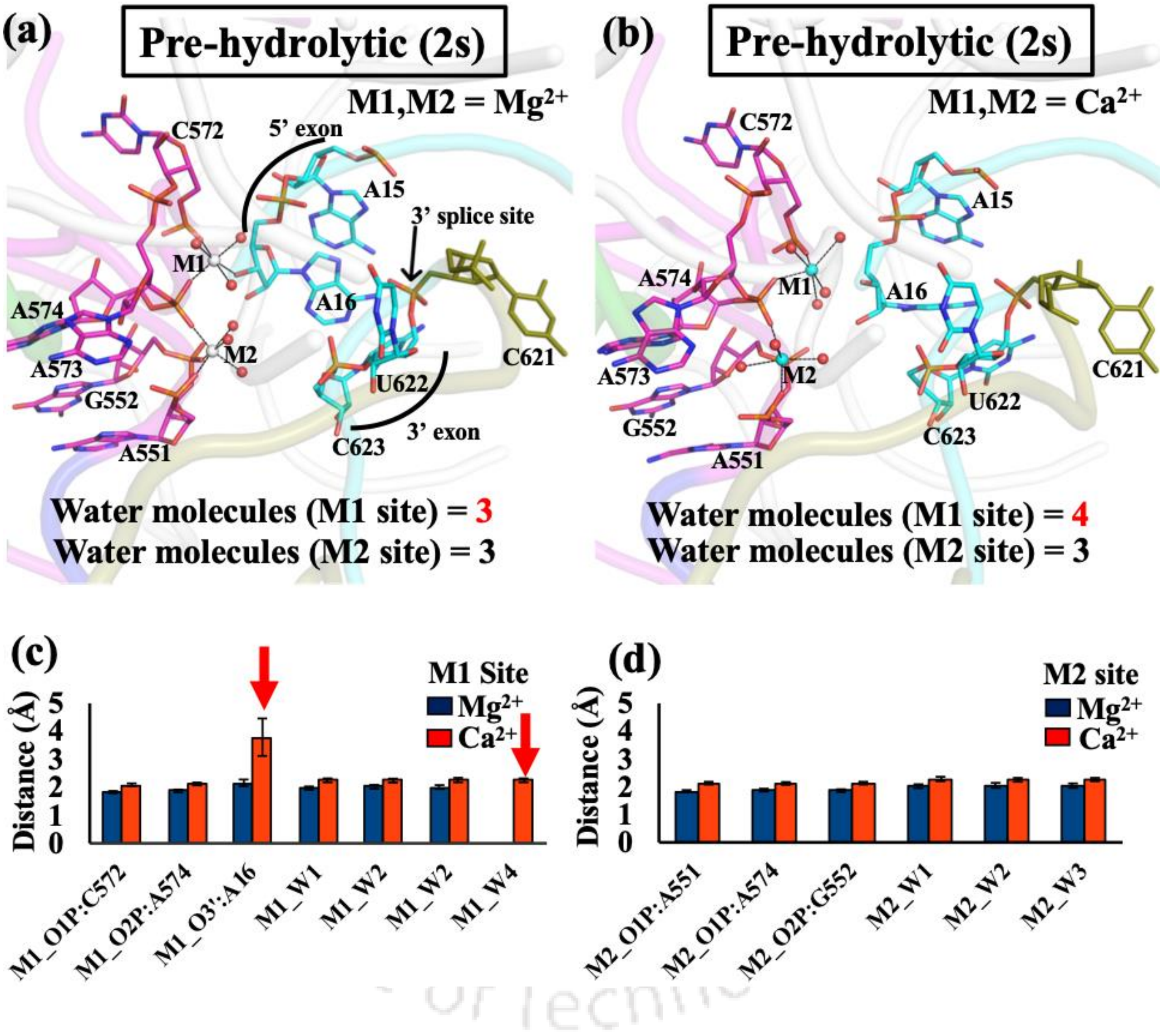


Figure 4.7. Representative MD structures of pre-hydrolytic (2s) state of lariat intron. Zoomed-in view of the catalytic pocket: (a) Mg^{2+} bound, (b) Ca^{2+} bound. Water molecules coordinating with the divalent metal ions (M1, M2) are shown explicitly as red spheres. Black dashed line indicates direct interactions. MD trajectory averaged key interaction distances of Mg^{2+}/Ca^{2+} at (c) M1 site (d) M2 site. Standard deviation is given as error bar.

4.3.4 Mg^{2+} and Ca^{2+} bound (M1, M2 site) post-hydrolytic lariat group II intron

MD structure of the catalytic pocket of Mg^{2+} bound lariat intron (post-hydrolytic state) was in excellent agreement (particularly the ligated exon part) with the X-ray structure (**Figure 4.5**). However, the distance between two Mg^{2+} ions was found to be larger (~5 Å) than the template X-ray structure (4.04 Å, PDB 6CIH). Mg^{2+} in the M1 pocket satisfied its six coordination (**Figure 4.8 a**) by involving two water molecules and three phosphates of the RNA backbone (A13, C572, A574). Ca^{2+} substitution in the M1 pocket allowed an additional water entry (**Figure 4.8 b**).

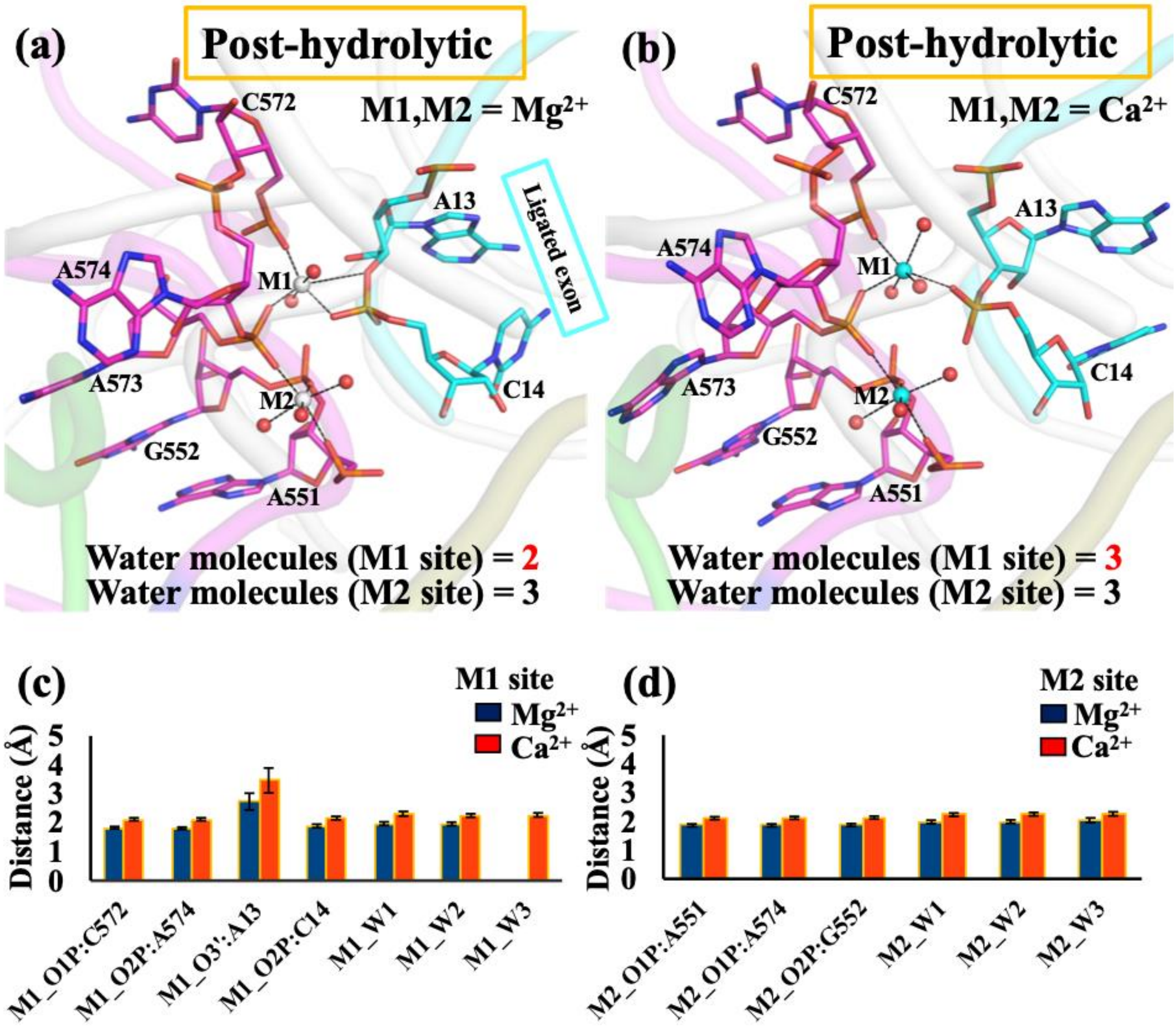


Figure 4.8. Representative MD structures of post-hydrolytic state of lariat intron. Zoomed-in view of the catalytic pocket: (a) Mg^{2+} bound, (b) Ca^{2+} bound. Water molecules coordinating with the divalent metal-ions (M1, M2) are shown explicitly as red spheres. Black dashed line indicates

direct interactions. MD trajectory averaged key interaction distances of Mg^{2+}/Ca^{2+} at (c) M1 site (d) M2 site. Standard deviation is given as error bar.

The atomic interaction network in the M2 site was found to be independent of metal ions (**Figure 4.8 a, b**), the same as pre-hydrolytic (2s) (**Figure 4.7 a, b**). Interestingly, the M1 site in the post-hydrolytic state was relatively dry (2/3 water molecules coordinating Mg^{2+}/Ca^{2+}) relative to the pre-hydrolytic (2s) state (3/4 water molecules coordinating Mg^{2+}/Ca^{2+}). Moreover, the negative charge density around the M1 site of the post-hydrolytic state was relatively higher compared to pre-hydrolytic (2s) state. Water entry in the M1 site and the increase in the interaction distances in response to $Mg^{2+} \rightarrow Ca^{2+}$ mutation were further supported by the distance plots (**Figure 4.8 c, d**).

4.3.5 Na^+ /K^+ bound (K1 site) lariat group II intron

Monovalent ion in the K1 binding pocket was found to be loosely bound, indicated by the multiple modes of interactions from various independent MD runs (supported by different water coordination, **Figure 4.9**). K1 site of the post-hydrolytic state was found to be relatively more hydrated (up to 5 water molecules, **Figure 4.9 c, d**) relative to the pre-hydrolytic state (up to 2 water molecules, **Figure 4.9 a, b**). Multiple modes of interactions in the K1 site of the post-hydrolytic state were observed from MD simulations (**appendix Figure 4.1, 4.2**). Post-hydrolytic K1 binding was loose and solvent accessible, resulting in poor selectivity ($\Delta\Delta G < 1$ kcal mol−1) between K^+ and Na^+. The same argument held true for the pre-hydrolytic (2s) state, which was found to be weakly discriminatory in favor of Na^+ binding against K^+.

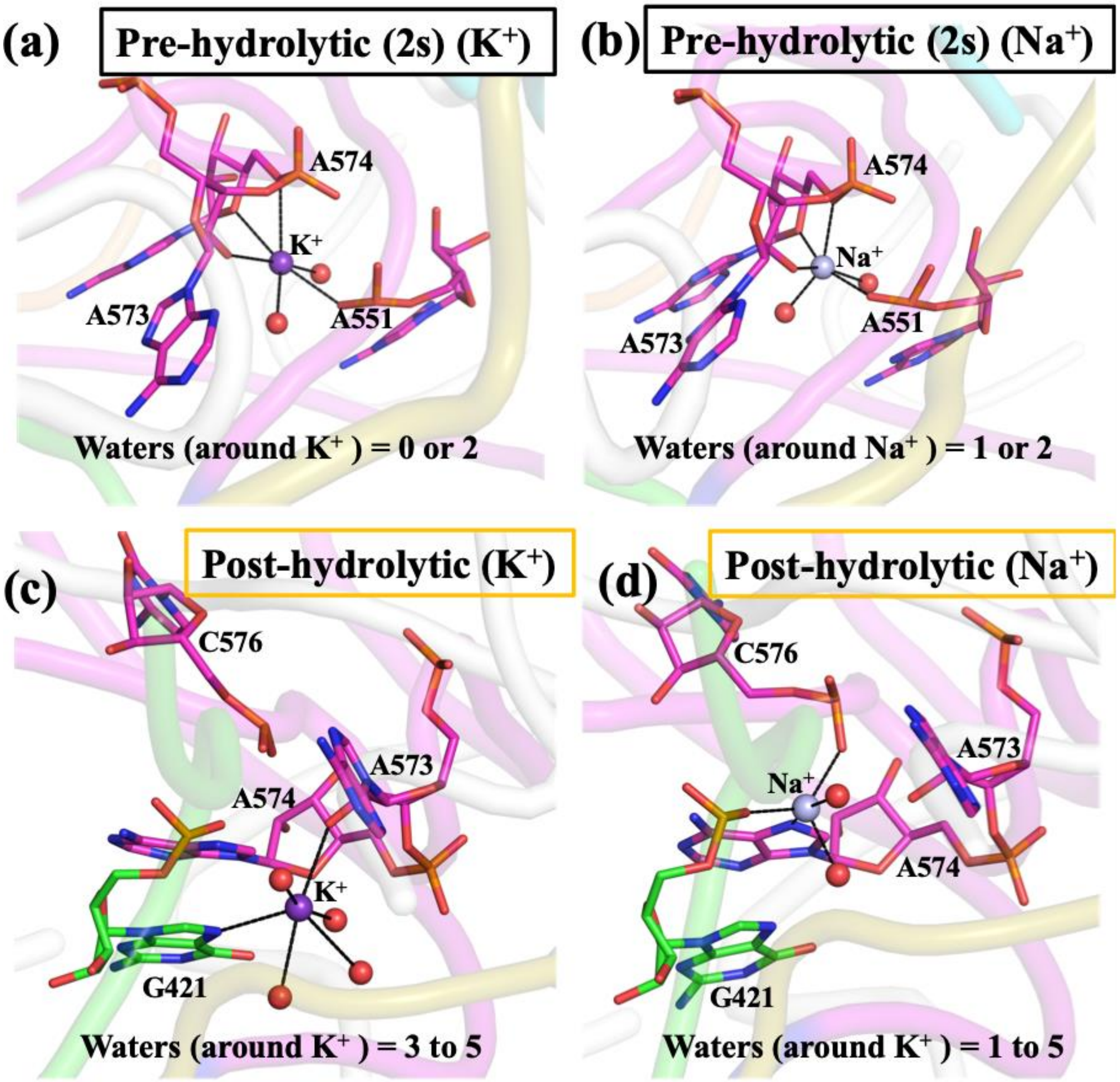


Figure 4.9. Zoom-in view of monovalent ion binding pocket (K1 site) of the pre-hydrolytic (2s) (upper) and post-hydrolytic (lower) state of the lariat intron. K^+ bound structures on the left-side. Na^+ bound structures at the right-side. Coordinating water molecules are explicitly shown as red spheres. The black-dashed line indicates direct interaction in the first coordination shell of K1.

4.4. Discussion

X-ray structure-based molecular dynamics free energy simulation was used to estimate the metal ion selectivity (Mg^{2+} versus Ca^{2+}, K^+ versus Na^+) in the catalytic pocket of the eukaryotic lariat group II intron (Figure 2). MD simulations showed that lariat group II intron was strongly selective between divalent metal-ions, favoring Mg^{2+} against Ca^{2+} [ΔΔG (Pre-hydrolytic) ~13 kcal mol^{-1}, ΔΔG (Post-hydrolytic) ~20 kcal mol^{-1}]. Interestingly, lariat intron weakly discriminated between monovalent ions (K^{+} versus Na^{+}) [ΔΔG (Pre-hydrolytic) ~ -1.96 kcal mol^{-1}, ΔΔG (Post-hydrolytic) ~0.72 kcal mol^{-1}]. Recently, we had estimated the divalent (Mg^{2+} versus Ca^{2+}) (**Kumar et al., 2020**) and monovalent (K^{+} versus Na^{+}) (**Kumar et al., 2020**) ion selectivity in the active site of bacterial linear group II intron at various stages (pre-hydrolytic, post-hydrolytic, free) of self-splicing pathway. The strength of divalent (monovalent) ion selectivity for the pre-hydrolytic and post-hydrolytic state of the bacterial group II intron was reported to be ~12 (6) and ~10 (1) kcal mol^{-1}, respectively. It was evident that eukaryotic lariat intron was non-selective between K^{+} and Na^{+}, in contrast to bacterial linear group II intron. Moreover, eukaryotic lariat intron was more selective between Mg^{2+} and Ca^{2+} (particularly in the post-hydrolytic state), compared to bacterial linear group II intron. MD structures of Mg^{2+} and Ca^{2+} bound lariat intron suggests six coordination of divalent metal ion in the M1 and M2 site of the catalytic pocket (**Figure 4.7, Figure 4.8**). The atomic radii of Ca^{2+} were larger than Mg^{2+}; thus, the average distances between the divalent-metal ion and oxygen of water/RNA were found to be smaller for Mg^{2+} (between 1.9-2.1Å) relative to Ca^{2+} (between 2.2-2.4 Å). The first coordination shell of the divalent metal-ion in the M1 site was altered in response to $Mg^{2+} \rightarrow Ca^{2+}$ substitution (coordinating water molecules increased 3→4 and disrupted M1…A16 interaction), whereas the coordination network in the M2 site seemed to be independent of the nature of the metal ion (**Figure 4.7, 4.8**). The atomic interaction network around the M1 and M2 sites was a robust feature of the simulations, confirmed by multiple independent MD trajectories. Thus, it can be argued that the divalent metal ion was strongly bound to the catalytic pockets (M1 and M2). The M1 site of the pre-hydrolytic (2s) state was found to be wetter (water molecules = 3/4) than the post-hydrolytic state (water molecules = 2/3). The higher discriminatory power of the post-hydrolytic state (ΔΔG ~20 kcal mol^{-1}) relative to pre-hydrolytic (2s) state (ΔΔG ~13 kcal mol^{-1}) seemed to be linked with the wetness of the binding pocket, particularly M1 site. Wet binding pocket (pre-hydrolytic (2s)) discriminated weakly, whereas dry binding pocket (post-hydrolytic) discriminated strongly between Mg^{2+} and Ca^{2+}, favored the former. The monovalent ion was found to be loosely bound in the K1 site (supported by the various possible coordination modes of K^{+}/Na^{+} in the MD structures) with high degree of solvent exposure. Water molecules were found to satisfy the binding requirement of K^{+}/Na^{+} in the K1 pocket involving water molecules ranging from 0 to 5.

The various possible mode of interaction indicated the loosely bound water exposed K1 pocket. Thus, $K^+ \rightarrow Na^+$ mutation in the loose and wet K1 pocket of the lariat intron always ensured the bonding requirement of the monovalent ion, leading to loss of discrimination. Thus, we proposed that the metal-ion selectivity in the lariat group II intron depended on the solvent accessibility of the metal-ion binding pocket; dry binding pocket ensured strong discrimination, and wetness off-sets the discriminatory ability. Water-shielding has been reported to boost the discriminatory power in various biomolecular recognition events, including metal-ion selectivity in bacterial group II intron (**Kumar et al., 2018; Kumar et al., 2020**), mRNA decoding (**Satpati et al., 2014; Satpati et al., 2014; Kumar et al., 2017**), tRNA recognition by synthetases (**Kumar et al., 2019**), etc.

Mg^{2+} (ionic radii = 0.86 Å) was smaller than Ca^{2+} (ionic radii = 1.14 Å) (**Shannon, 1976**), thus, charge density was higher for the former. High charged density usually resulted in stronger binding to negatively charged binding pockets (**Jing et al., 2018**). Thus, it was expected that the use of classical force field (which only included polarization effects implicitly) for binding affinity estimation will always have bias Mg^{2+} binding against Ca^{2+} in the metal ion binding pockets. Indeed, it has been shown that many-body effects ensured favorable Ca^{2+} binding against Mg^{2+} to the calcium-binding proteins (**Jing et al., 2018; Kohagen et al., 2014**). Thus, estimation of divalent metal ion selectivity (Mg^{2+} and Ca^{2+}) demanded calculations using expensive polarizable force fields. Clearly, the estimated divalent metal ion selectivity ($\Delta\Delta G$: Mg^{2+} versus Ca^{2+}, **Figure 4.3**) required justification. To verify if the sign of the estimated $\Delta\Delta G$ (Mg^{2+} versus Ca^{2+}, **Figure 4.3**) was correct, we estimated a quantity "ΔE" (eqn. 1) using density functional theory, that can explicitly consider polarization and approximately estimated the relative free energy of binding (Mg^{2+} versus Ca^{2+}). Estimated ΔE was limited by the size (reduced binding pocket models from MD structures), solvation model (implicit dielectric), and absence of entropic effect (**See Methods Chapter 1**). But ΔE estimation was computationally inexpensive and could be considered as an indicator of divalent metal-ion preference. Estimated ΔE (**Figure 4.4**) implied that Mg^{2+} in the catalytic pocket was preferred against Ca^{2+}, in line with the classical MD estimations (**Figure 4.3**). The magnitude of the metal-ion preference in the lariat intron pocket ($\Delta\Delta G$) was experimentally unknown. Thus, the estimated $\Delta\Delta G$'s (**Figure 4.3**) cannot be disproved or confirmed. However, the sign of the estimated $\Delta\Delta G$'s corroborated the experiment (**Sigel, 2005**). Experiments confirmed that Mg^{2+} was the cognate metal ion and key for the catalytic activity of lariat group II intron.

Notably, the structures of the Ca^{2+}-bound lariat intron were uncharacterized experimentally to date. Thus, the MD structures provided an insight into the experimentally uncharacterized states and encouraged experimental verification.

4.5. Conclusion

Using the X-ray structures (pre-hydrolytic (2s) and post-hydrolytic state) of eukaryotic lariat group II intron, we carried out classical molecular dynamics free energy simulations with cognate (Mg^{2+}, K^{+}) and near-cognate (Ca^{2+}, Na^{+}) intron structures and estimated the divalent and monovalent ion selectivity ($\Delta\Delta G$) in the binding pocket (**Figures 4.3**). Eukaryotic lariat group II intron strongly discriminated between Mg^{2+} and Ca^{2+}, favored the ($\Delta\Delta G > 10$ kcal mol^{-1}, **Figure 4.3**). Density function calculations (**Figure 4.4**) indicated that the sign of the MD estimated divalent metal-ion discriminatory power ($\Delta\Delta G$) was correct, thus confirming the Mg^{2+} preference against Ca^{2+} in the catalytic pocket of lariat intron. Notably, the lariat intron showed weak monovalent ion discrimination (K^{+} versus Na^{+}, **Figure 4.3**). The divalent metal ions (M1, M2 sites) were strongly bound in the dry catalytic pocket, whereas the monovalent ion (K1 site) was loosely associated in the solvent accessible wet binding pocket of the lariat intron. The solvent-accessibility of the ion binding pocket determined the strength of metal-ion selectivity. Dry binding pockets ensured high selectivity, whereas loosely bound wet binding pockets diminished the discriminatory power of the lariat group II intron.

Chapter 5

Divalent metal ion (Mg^{2+} vs. Ca^{2+}) selectivity in CRISPR-Cas system associated Cas1 protein

Cas1 is a universally conserved essential metalloenzyme of the CRISPR immune system of prokaryotes that can cut and integrate a part of viral DNA to its host genome with the help of other proteins. The integrated DNA acts as a memory of viral infection which can be transcribed to RNA and stop future infection by recognition followed by protein-mediated degradation of the viral DNA. It has been proposed that the presence of single manganese (Mn^{2+}) ion in a conserved divalent metal-ion binding pocket (key residues: E190, H254, D265, D268) of Cas1 is crucial for its function. Cas1-mediated DNA degradation was proposed to be hindered by metal substitution, metal chelation, or mutation of the binding pocket residues. Cas1 is active towards dsDNA degradation with both Mn^{2+} and Mg^{2+}. X-ray structures of Cas1 revealed an intricate atomic interaction network of the divalent metal ion-binding pocket and open up the possibility of modelling related metal ions (viz., Mg^{2+}, Ca^{2+}) in the binding pocket of wild-type and mutated Cas1 protein for computational analysis, which includes (1) quantitative estimation of the energetics of the divalent-metal ion preference (2) explore the structural and dynamical aspects of the protein in response to divalent metal ion substitution or amino-acid mutation. Using X-ray structure of Cas1 protein from *Pseudomonas aeruginosa* as a template (PDB 3GOD), we performed (~2.23μs) classical molecular dynamics simulations to compare structural and dynamical differences between Mg^{2+} and Ca^{2+} bound binding pocket of wildtype (WT) and mutant (E190A, H254A, D265A, D268A) Cas1. Results suggest that Cas1 prefers Mg^{2+} binding relative to Ca^{2+} and the preference is strongest for WT and weakest for D268A mutant. Quantum chemical calculations indicate that Mn^{2+} is most preferred relative to both Mg^{2+}, Ca^{2+} in the wild-type Cas1. Substitution of Mg^{2+} by Ca^{2+} does not alter the interaction network between Cas1 and divalent metal ion but increase the wetness of the binding pocket by introducing a single water molecule in the first coordination shell of the latter. The strength of metal ion preference seems to be dependent on the solvent accessibility of the binding pocket.

5.1 Background

Prokaryotes are known (**Barrangou et al., 2007; Brouns et al., 2008; Barrangou et al., 2014**) to acquire immunity by integrating short fragments (24 - 48 nt) of viral DNA into its genome intercalated by short repeating elements (CRISPR: clustered regularly interspaced short palindromic repeats). Insilico studies (**Mojica et al., 2000; Makarova et al., 2006**) first identified a distinct repetitive DNA feature common in the prokaryotic genome called CRISPR. CRISPR repeats contain “unique” spacer sequences in between. It has been established (**Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005**) that these “unique” spacer sequences are identical to foreign genetic elements (e.g. viruses and plasmids). Bacteria can transcribe RNA from those “unique” spacer sequences and readily neutralize the infecting virus by recognizing through DNA-RNA complementarity and fragmenting the viral DNA by involving associated proteins (CRISPR-associated proteins). CRISPR-Cas system acquires adaptive immunity in three stages: adaptation, expression and interference (**Van der Oost et al., 2009; Garneau et al., 2010; Nuñez et al., 2014; Jackson et al., 2017**). A fragment of foreign DNA (“unique” spacer) is selected and inserted into the CRISPR array during adaptation, which serves as a memory of viral infection. In the expression and interference step, RNA is expressed from the integrated spacer followed by the fragmentation of the foreign DNA. Each of these steps involves multiple diverse sets of proteins (**Makarova et al., 2006; Haft et al., 2005; Jansen et al., 2002**). CRISPR-Cas systems have received considerable attention from the theoretical/computational community in the recent past (**Haft et al., 2005; Martynov et al., 2017; Pinello et al., 2016; Zheng et al., 2017; Zeng et al., 2018; Xu et al., 2017; Estarellas et al., 2015; Wan et al., 2019; Palermo et al., 2016; Palermo et al., 2017**). Bioinformatics approach has identified different Cas protein families (**Haft et al., 2005**) and estimated the optimum number of nucleotide spacers that maximizes prokaryotic immunity (**Martynov et al., 2017**). A computational tool “CRISPResso” has been developed to analyse CRISPR editing experiments (**Pinello et al., 2016**). Coarse-grained modelling (**Zheng et al., 2017**) and dynamics (**Zeng et al., 2018**) of CRISPR-Cas9 systems were studied. The statistical mechanical approach proposed a strong correlation between Cas9 mediated DNA cleavage efficiency and the stability of the DNA-RNA loop in the protein complex (**Xu et al., 2017**). The first molecular dynamics (MD) simulations addressed the structure and dynamics of the Csy4-RNA complex (Csy4 is another protein component of the CRISPR system) and highlighted the limitations of the simulation techniques in studying protein-RNA systems (**Estarellas et al., 2015**).

MD simulations of Cas1-Cas2 (protein complex) binding to duplex DNA (**Wan et al., 2019**) indicated that the presence of PAM sequence (a short 2-6 base-pair DNA sequence) in the target DNA improves the stability of the Cas1-Cas2-DNA complex. Classical MD simulations (**Palermo et al., 2016; Palermo et al., 2017**) of the CRISPR/Cas9 system revealed the importance of the HNH domain of Cas9 in the concerted editing mechanism. Thus, computational studies are a useful complement to experiments in understanding the mechanism of the CRISPR-Cas system in atomic details. CRISPR-associated protein 1 (Cas1) is a universally conserved (and hallmark) protein of the CRISPR immune system. Cas1 is a metal-dependent (Mn^{2+}) endonuclease that can cut foreign DNA and produces small DNA fragments for integration to host-genome (**Wiedenheft et al., 2009**). Cas1, together with Cas2 (another universally conserved CRISPR protein) recognizes, cleave, and integrates the foreign DNA into the host CRISPR loci. Thus, Cas1 is essential for spacer acquisition/adaptation step (**Brouns et al., 2008; Nuñez et al., 2014; Martynov et al., 2017**). Experiments confirmed that at physiological KCl concentrations, Cas1 shows the highest nuclease activity on dsDNA with Mn^{2+} than with Mg^{2+} (**Wiedenheft et al., 2009**). Moreover, Cas1-mediated ssDNA cleavage is exclusively supported with Mn^{2+} (**Wiedenheft et al., 2009**). The structural study of the Cas1 protein revealed the intricate atomic interaction network in the conserved divalent metal ion-binding pocket. Biochemical studies showed that metal-ion substitution, mutation of conserved metal ion-binding residues, metal ions chelation inhibits Cas1-mediated DNA fragmentation (**Wiedenheft et al., 2009**).

As discussed in Chapter 1, the structure of Cas1 protein from *Pseudomonas aeruginosa* reveals that Cas1 is organized into N-terminal (10 β-strand and 2 α-helix) and C-terminal (8α-helices connected with loops) domain connected via loop-linker (**Figure 5.1**). The metal ion-binding domain is present in the α-helical C-terminal domain, where the universally conserved residues (E190, H254, D265) and a strongly conserved residue at position 268 (D or E) cluster around the manganese (Mn^{2+}, cognate divalent metal ion).

Despite experimental and theoretical studies, several important questions remain unanswered related to Cas1, which are essential for understanding the role of Cas1 in the CRISPR adaptation step of prokaryotic immunity. The energetics of divalent metal ion selectivity is not known. The link between metal-ion selectivity and 3D structure is missing. Furthermore, the structures of related divalent metal-ion bound (Mg^{2+}, Ca^{2+}) Cas1 are not yet experimentally resolved. Using the

X-ray structure as a template we have performed computational analysis and quantitatively estimated the energetics of divalent metal-ion selectivity by Cas1. Results confirm Mn^{2+} is the most preferred ion in the Cas1 binding pocket. Moreover, Cas1 is selective between Mg^{2+} and Ca^{2+}, favouring the former. The strength of ion selectivity seems to be linked with the wetness of the metal-ion binding pocket.

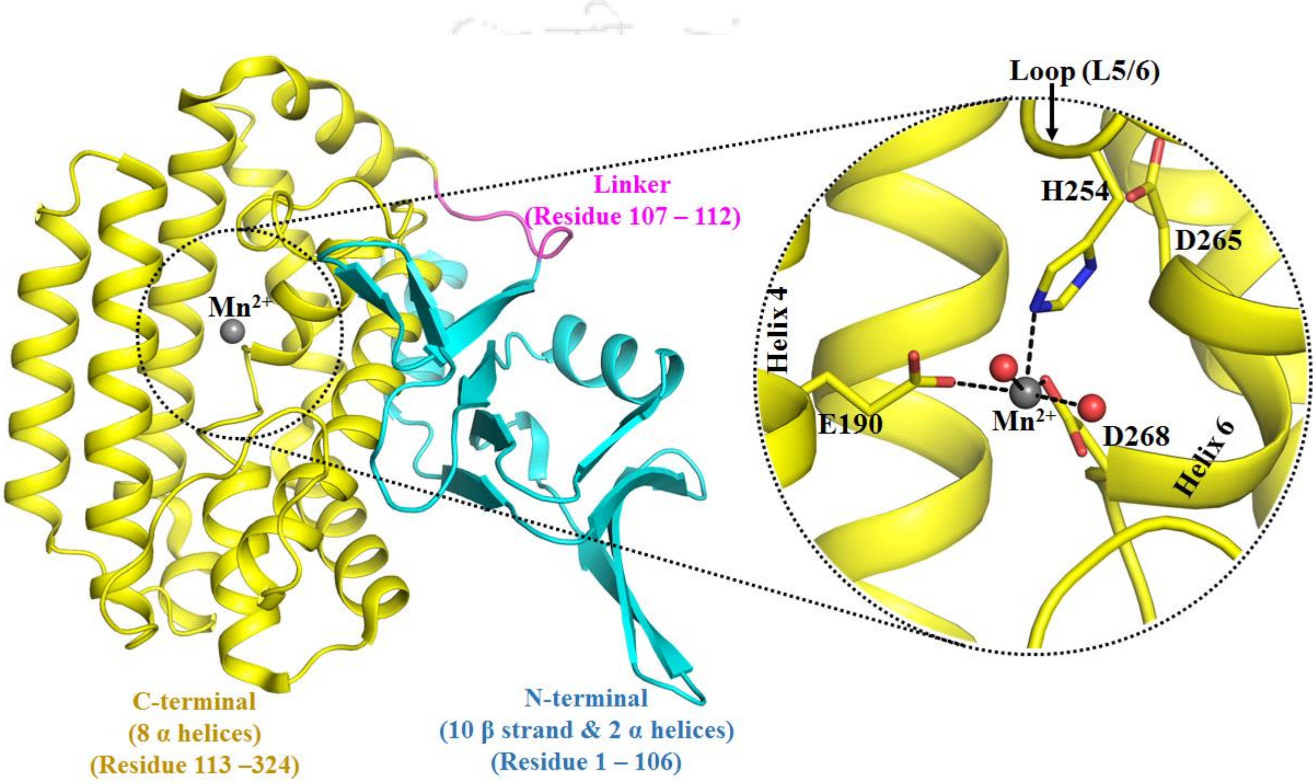


Figure 5.1. 3D architecture of Cas1 protein. Cartoon representation of Cas1 protein (PDB:3GOD) (N-terminal: cyan, linker: magenta, C-terminal: yellow, divalent ion (Mn^{+2}): grey sphere). Key residues involved in stabilizing interactions across the divalent metal ion (Mn^{+2}) are marked and highlighted by sticks.

5.2 Methods

5.2.1 Molecular dynamics procedure

The structure of Cas 1 protein was taken from the PDB entry 3GOD (resolution 2.17 Å) (**Wiedenheft et al., 2009**), and a 25 Å radii sphere, centred at the backbone oxygen of L269 was truncated. Template PDB 3GOD (**Wiedenheft et al., 2009**) consists of manganese (Mn^{2+}) ion at the binding pocket. Mg^{2+}, Ca^{2+} bound Cas1 was modelled by substituting Mn^{2+}. Mutants of Cas1 were modelled by *in silico* mutation of single amino-acid sidechain (E190A, H254A, D265A, and D268A) of the Cas1 protein into the template structure (PDB 3GOD, and the resulting models were subjected to MD. Total ~49178 atoms were considered in our simulations.

For each of the simulation models, several independent replicas were performed (from minimum 6 to a maximum of 7). These independent 6-7 trajectories were divided into two groups: (A) for structural analysis (B) for free energy calculation. 2 trajectories were used for (A), which include 300/350 ps equilibration followed by 150 ns production dynamics for each trajectory. Thus, for each simulation model total 300 ns production run was considered for structural analysis. In total 1500 ns of production MD simulations were considered for structural analysis of wild-type and mutant Cas1 protein. The rest 4-5 replicas were used for (B), discussed in detail in the next section. The overall charge of the simulation was neutralized by adding counter ions (5 chloride ions were added for WT and H254A mutant, 4 chloride ions were added for E190A, D265A, and D268A mutants). Root-mean-square deviation (RMSD) of the heavy atoms (within 22 Å of simulation sphere) of the Cas1 protein relative to the X-ray structure (PDB: 3GOD) was estimated to compare MD structures with its template X-ray structure. Average RMSD and RMSF were estimated by averaging over the last 100 ns of 150 ns production MD.

5.2.2 Protocol for binding free energy calculation

An appropriate thermodynamic cycle (**Figure 5.2**) was used to estimate the relative binding affinity (ΔΔG) of Mg^{2+}/Ca^{2+} binding to Cas1. The vertical arms of the cycle (**Figure 5.2**) correspond to Mg^{2+}/Ca^{2+} binding. Whereas, horizontal arms correspond to the alchemical transformation of Mg^{2+} into Ca^{2+} (cannot be realized experimentally); in complex with the Cas1 protein (upper arm) or free in water (lower arm). We have computed the free energy change associated with the horizontal alchemical arm of the cycle, which includes the free energy change associated with the change in van der Waals radii during $Mg^{2+} \rightarrow Ca^{2+}$ transformation. Free energy calculation for each replica was based on 88-110 ns of data collection averaged over 4-5 replicas with different initial velocities. Each replica includes 22 ns production dynamics for free energy calculation. Computed free energies from different MD runs are in excellent agreement (**appendix Table 5.1**). A total of 704 ns of molecular dynamics free energy simulations have been performed to get good convergence and reasonable statistical error (< 1 kcal/mol). Free energy derivatives at different λ values are given in **appendix Table 5.1, 5.2** and support convergence. Positive $\Delta\Delta G$ refers to favourable Mg^{2+} binding relative to Ca^{2+} (**appendix Table 5.3**).

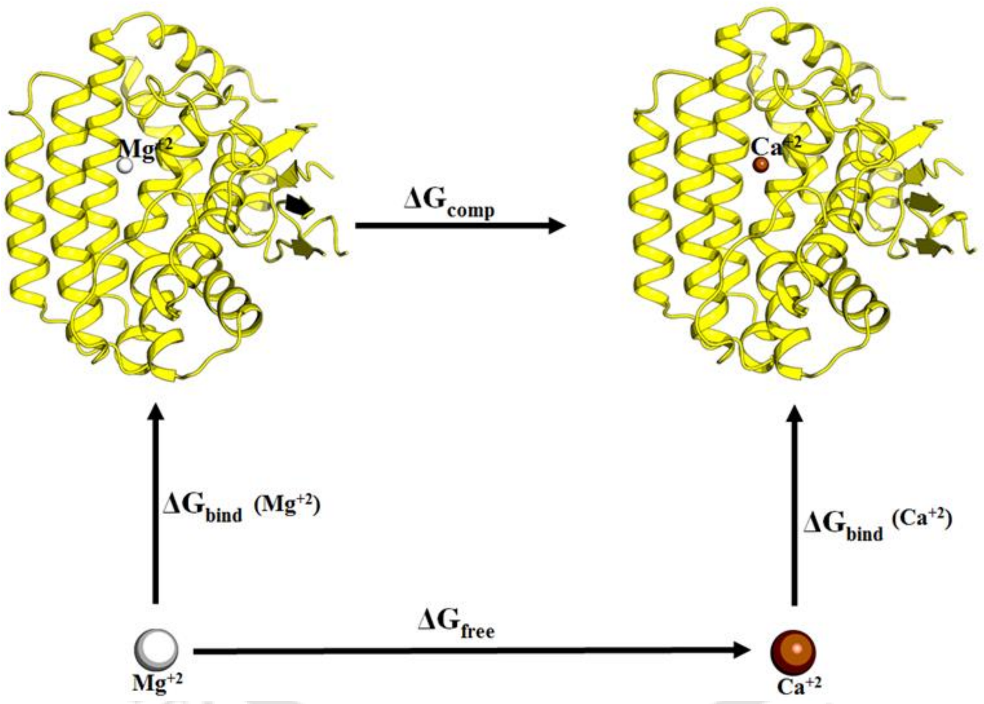


Figure 5.2. Thermodynamic cycle for Mg^{2+} vs Ca^{2+} ion binding to the divalent ion binding sites of Cas1 protein. Vertical legs correspond to divalent ion binding; horizontal legs correspond to the alchemical transformation of Mg^{2+} into Ca^{2+} in the cas1 binding sites, either in the Cas1 bound (above) or free (below) in solution. Free energy associated with the horizontal legs of the thermodynamic cycle has been computed by MD simulations and thermodynamic integration. The binding free energy difference is $\Delta\Delta G = \Delta G_{comp} - \Delta G_{free} = \Delta G_{bind}(Ca^{2+}) - \Delta G_{bind}(Mg^{2+})$.

5.2.3 Ab initio quantum calculations

Representative MD snapshot from the production dynamics of the Cas1 protein was selected. Then divalent metal-ion and its amino-acid residues of the first coordination shell, coordinating waters were taken (reduced binding pocket model) and considered for electronic structure calculations. Methyl/hydrogen capping has been done for amino-acid residues (**appendix Figure 5.1**). The reduced binding pocket model was embedded in aqueous dielectric and was subjected to single point density functional energy calculation in two different levels of theory (1) M06-2X functional (**Zhao et al., 2008**) in combination with 6-311++G** basis set (2) B3LYP functional (**Becke 1988; Lee et al., 1988**) with 6-31+G* basis. Solvation was included by implicit SMD solvent model (**Marenich et al., 2009**). Gaussian16 program (**Frisch et al., 2016**) was used for abinitio calculations. Free Mg^{2+}/Ca^{2+} in an aqueous dielectric was also subjected to abinitio energy calculation. We estimated the single point energy difference (ΔE) of the right and the left-hand side of eqn (5.1, 5.2):

$$[Mg^{2+} - water] + [Mn^{2+} - Cas1] \rightarrow [Mn^{2+} - water] + [Mg^{2+} - Cas1] \quad (5.1)$$

$$[Mg^{2+} - water] + [Ca^{2+} - Cas1] \rightarrow [Ca^{2+} - water] + [Mg^{2+} - Cas1] \quad (5.2)$$

$[M^{2+} - Cas1]$ and $[M^{2+} - water]$ represents M^{2+} (Mn^{2+}/ Mg^{2+}/Ca^{2+}) in the reduced binding pocket of Cas1 and free in water respectively. Single point energy differences (ΔE, **see appendix Table 5.4**) evaluated between the right-hand and left-hand side of eqn (5.1) and eqn (5.2) approximates the relative divalent cation preference (Mn^{2+} vs Mg^{2+}) and (Mg^{2+} vs Ca^{2+}) respectively, into the Cas1 binding pocket. A positive energy difference (ΔE) from eqn (5.1) implies Mn^{2+} preference over Mg^{2+}, whereas, negative energy difference (ΔE) from eqn (5.2) implies a Mg^{2+} preference over Ca^{2+} at the Cas1 active site. Transition metal Mn^{2+} ion (d^5 electronic configuration) may have various possible spin states (high-spin state with S = 5/2, low-spin state = 1/2 or 3/2). However, low-spin states are expected to be energetically disfavoured due to half-filled d-shell configurations. We performed ab initio calculations using unrestricted B3LYP functional considering of all the three spin states of Mn^{2+} (S= 5/2, 3/2, and 1/2). The results confirmed that

Mn^{2+} in the high-spin state (S=5/2) is preferred in the free as well as in the truncated binding pocket model of Cas1 (results not shown). Thus, the Mn^{2+} in the high-spin state was considered for estimating ΔE using equation (5.1).

5.3 Results

Classical force fields are the popular choice of computational chemistry/biochemistry because they allow efficient phase space sampling of the biological systems at atomic details and estimate thermodynamic properties (viz. free energies). Divalent ions, which are fundamental units of biochemistry, are described as a set of constant parameters in the classical force field. Fixed parameter description in the force field especially for the transition metal ions (viz. Mn^{2+}) represents a challenge **(Comba et al., 1995; Banci 2003; Hu et al., 2011; Zhang et al., 2012; Neves et al., 2013**). Force-field modification seems to be necessary for transition metals as the parameters strongly depend not only on the oxidation state of the transition metal ion but also on the nature of the residues of the metal ion-binding pocket (**Neves et al., 2013**). In the paper, using Mn^{2+} bound Cas1 as template we first modelled Mg^{2+} and Ca^{2+} bound Cas1 and subjected those models to classical MD simulation and estimated the relative preference of Mg^{2+} and Ca^{2+}. Rather than attempting to simulate Mn^{2+} bound Cas1 using classical MD or computationally demanding polarizable force-field, we performed quantum chemical calculations (considering a simple reduced binding pocket-model) to approximately evaluate the Mn^{2+} preference relative to Mg^{2+}/Ca^{2+} in the Cas1 binding pocket.

5.3.1 MD insight into the structural convergence and overall flexibility of the protein

Root mean square deviation (RMSD) is a straightforward and popular choice for overall structural comparison. RMSD of the protein-heavy atoms relative to template X-ray structure was estimated (**Figure 5.3**) to evaluate the convergence of the simulations and structural similarity between X-ray and MD structures. Within 50ns, RMSD reaches a plateau and fluctuates around 1.21 Å with a standard deviation of 0.04 Å. The plateau of the RMSD plot indicates structural convergence. Moreover, small average RMSD of 1.21-1.28 Å indicates that the MD structures closely resemble the experimental structure, thus sampling the desired minima of the potential energy hypersurface was confirmed.

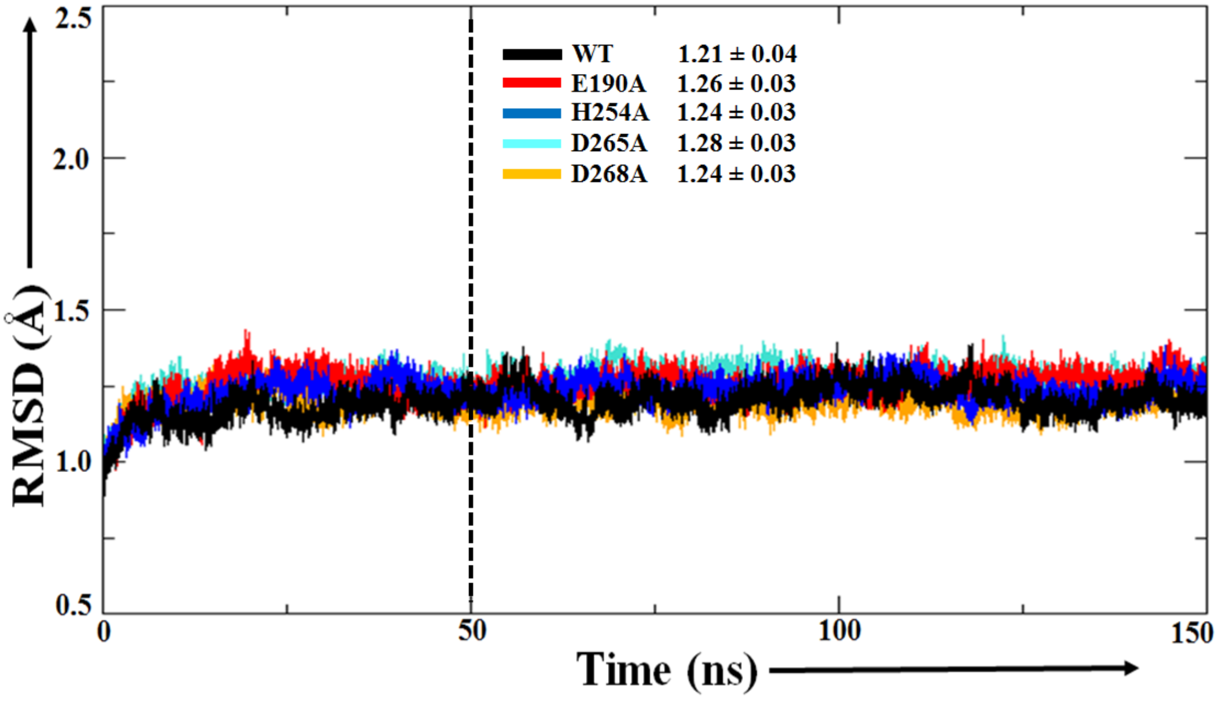


Figure 5.3. Root-mean-square deviation of heavy atoms within 22Å of unrestrained simulation sphere with respect to X-ray structure. WT (black), E190A (red), H254A (blue), D265A (cyan), and D268A (yellow).

The flexibility of the amino-acid residues of Cas1 protein during MD simulations was evaluated by estimating the trajectory averaged root-mean-square fluctuations (RMSF) (**Figure 5.4**). The RMSF colour scale (**Figure 5.4 a-f**) cannot illustrate the distinct difference in flexibility between WT and mutated Cas1. But they do highlight that the divalent metal binding pocket is rather rigid but indicates increased flexibility at various regions (helix 2, loop connecting helix 2 and 3, helix 5, loop connecting helix 5 and 6: L5/6) in the E190A mutant Cas1 relative to its wild-type analogue.

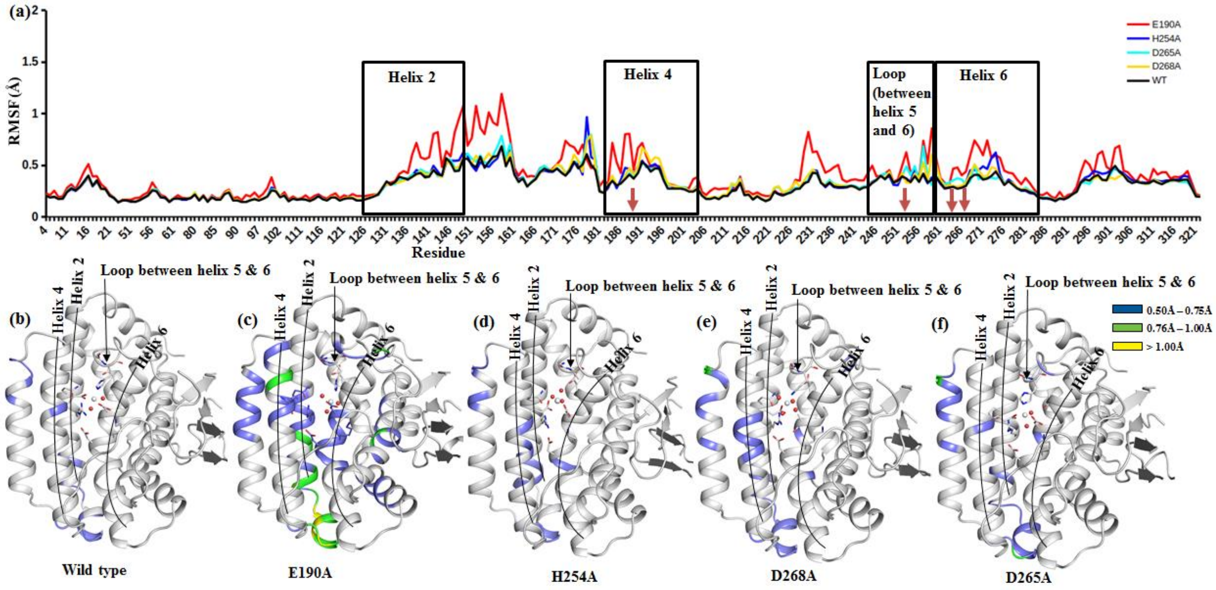


Figure 5.4. Cas1 protein active site flexibility upon mutations. (a) Root-mean-square fluctuation (RMSF) of the Cα atom. Graphs are shown for WT (black), E190A (red), H254A (blue), D265A (cyan), and D268A (yellow). RMSF is shown for last 100 ns of production MD trajectory. The RMSF are color-coded according to the scale (0.50-0.75: blue, 0.76-1.0: green, and >1: yellow). RMSFs by color are shown for WT (b), E190A (c), H254A (d), D268A (e), and D265A (f).

5.3.2 Energetics of ion selectivity

To understand the energetics of divalent metal ion selectivity (Mg^{2+} vs Ca^{2+}), we performed classical MD simulations couple to free energy estimations using the thermodynamic integration (TI) methodology. The calculations involve estimation of the change in divalent metal ion binding affinity upon $Mg^{2+} \rightarrow Ca^{2+}$ mutation in the Cas1 binding pocket and free in water (**Figure 5.2**). Further, the effect of Cas1 mutations (E190A, H254A, D265A, and D268A) in divalent metal ion selectivity (Mg^{2+} vs Ca^{2+}) was also examined. The results (**Figure 5.5**) suggest that: (1) Mg^{2+} is favoured over Ca^{2+} by ~ 4-5 kcal/mol in Cas1 protein. The energetic penalty may result in facile Ca^{2+} dissociation relative to Mg^{2+} from the Cas1 binding pocket, ensuring selectivity. (2) The strength of discrimination is highest for WT and lowest for the D268A mutant of Cas1. However, Mg^{2+} vs Ca^{2+} selectivity in protein binding pocket estimated by classical force-field is tricky (**Neves et al., 2013; Jing et al., 2018**). Due to the smaller size and higher charge density on Mg^{2+} relative to Ca^{2+}, one can expect the former to bind relatively tightly to the negatively charged side-chain of the protein binding pocket as in Cas1. Gel electrophoresis and urea denaturation experiments (**Cruz-León et al., 2020**) was performed to measure the folding equilibrium of the Tetrahymena ribozyme in the presence of various divalent metal cations (Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+}). The results suggest stability increases with the charge density of the cation and the binding affinity, therefore, follows a direct Hofmeister ordering $Mg^{2+} > Ca^{2+} > Sr^{2+} > Ba^{2+}$. The catalytic activity of introns was shown to be compromised significantly by replacing small amounts of Mg^{2+} (cognate) with near-cognate Ca^{2+} (**Schnabl et al., 2010; Fedorova et al., 2007**). The above experimental studies (**Cruz-León et al., 2020; Schnabl et al., 2010; Fedorova et al., 2007**) confirm preferable Mg^{2+} binding relative to Ca^{2+} and support the "high charge density resulting in high-affinity binding concept. However, the concept cannot be generalized for all. Although Mg^{2+} is a better charge acceptor, calcium-binding proteins with highly charged binding pockets can selectively bind Ca^{2+} against a much higher concentration of Mg^{2+} (**Dudev et al., 2014; Gifford et al., 2007; Schwaller, 2010**). It has been shown (**Jing et al., 2018**) that Mg^{2+} vs Ca^{2+} selectivity of proteins is dictated by the many-body polarization effect; thus, consideration of electronic polarization might be appropriate. Thus, Ca^{2+} versus Mg^{2+} selectivity seems to be dependent on two factors (i) geometric constraint and (2) the many-body effect. However, the same is not been proved for other divalent metal ions (**Jing et al., 2018**). Note, classical force-field is limited by the fact that it does not include electronic polarizability explicitly. Instead of using computationally expensive polarizable force-field, we performed quantum-chemical ab-initio calculations on the reduced model of the divalent metal ion binding pocket (including divalent metal ion and its 1st and 2nd coordination shell from selected MD snap-shots, **appendix Figure 5.1**). We computed the single-point ab initio energy difference (ΔE) between right and left- side of Eqn (5.1), which approximates the divalent metal ion preference (Mg^{2+} vs Ca^{2+}, and Mn^{2+} vs Mg^{2+}) in the Cas1 binding pocket. Positive ΔE indicates that the left side of Eqn (5.1) is energetically more stable, i.e, Mn^{2+} in Cas1 pocket is energetically more favourable relative to Mg^{2+}. The abinitio result obtained from M06–2X/6-311++G** level of theory is shown in **Figure 5.6**.

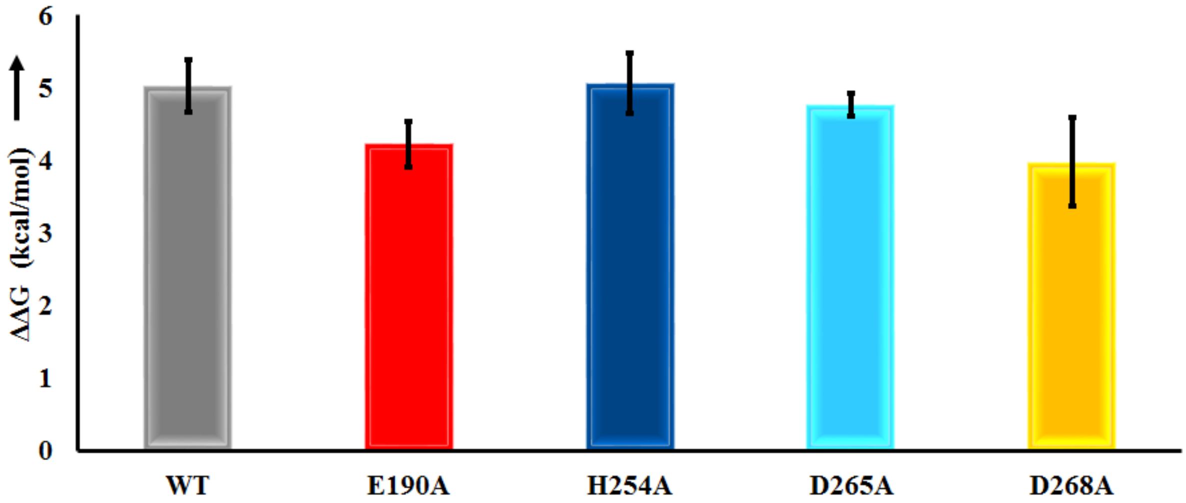


Figure 5.5. Calculated energetics of divalent ion (Mg^{2+} vs Ca^{2+}) binding to cas1 protein for wild-type and different mutations (E190A, H254A, D265A, and D268A). Error bars represented by the standard error of the mean.

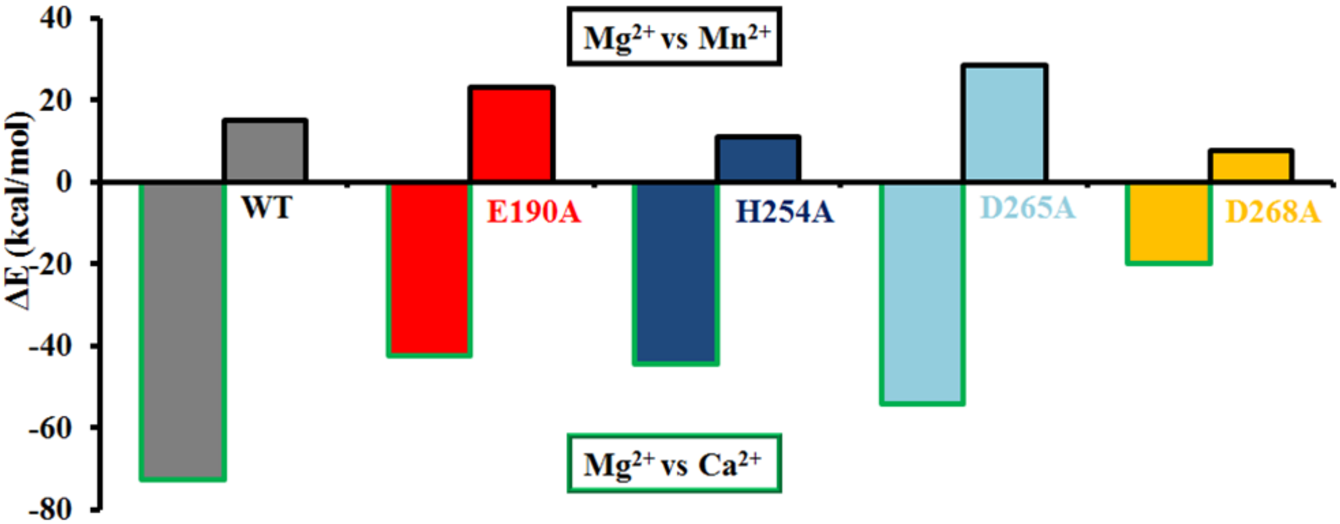


Figure 5.6. Quantum chemical single point energies calculated (Mn^{2+} vs Mg^{2+}, and Ca^{2+}) bound Cas1 for wild-type and mutated.

Large negative ΔE implies Mg^{2+} is energetically more favoured relative to Ca^{2+} at Cas1 active site. Note, ΔE is an approximation to the divalent metal ion selectivity (limited by small size of the binding pocket model as well as the absence of entropic effect), thus should be viewed just as an indicator of metal-ion selectivity. Representative snapshots (Mg^{2+} and Ca^{2+} bound Cas1) from the molecular dynamics trajectories were selected for modelling of simplified binding-pocket for quantum chemical analysis. Mn^{2+} bound WT reduced binding pocket model was generated by using the X-ray structure (PDB 3GOD) as template. Results (**Figure 5.6**) indicate that (1) divalent metal ion preference of Cas1 follow the order $Mn^{2+} > Mg^{2+} > Ca^{2+}$, and (2) Mg^{2+} vs Ca^{2+} discrimination is highest for WT Cas1 and least for D268A mutant. The strength of divalent metal ion binding to a given binding pocket (ligand) depends on several factors, viz., the ionic radii of the metal, charge-accepting ability of the metal, coordination number of the metal, crystal field stabilization energy of the metal-ligand complex, the binding pocket geometry (the 3D architecture and the nature of the ligands around the metal), etc (**Dudev et al., 2014; Cotton et al., 1988; Kuppuraj et al., 2009**). Large ionic radii decrease the charge density on the metal, resulting in less favourable metal-binding free energy. However, between two cations of the same formal charge and similar ionic radii, the metal ion with the better electron-accepting ability (Lewis acid) binds more strongly to the same ligand. The coordination number of the metal alters the metal-ligand distances (**Dudev et al., 2000**). For example, high-level ab initio calculations (**Pavlov et al., 1998**) show that the average Mg…O distances for tetrahydrated (1.98 Å) and hexahydrated (2.08 Å) differ by 0.10Å. The radius (and coordination number) of Mn^{2+}, Mg^{2+}, and Ca^{2+} are 0.97Å (6), 0.86Å (6), and 1.14Å (7) respectively (**Dudev et al., 2014; Shannon, 1976**). Thus, divalent manganese has a radius somewhat larger (by 0.11Å) than that of magnesium, and smaller than that of calcium. The largest size of Ca^{2+} can be correlated with the weak binding affinity. But the question is why does Cas1 prefer to bind to Mn^{2+} rather than Mg^{2+}? Structural analyses of Mg^{2+} and Mn^{2+} binding pockets (**Bock et al., 1999**) highlighted the striking similarity (viz., both Mn^{2+} and Mg^{2+} prefer 6 coordination number) and differences (Mn^{2+} will bind more readily to a site containing nitrogen in addition to oxygen, whereas Mg^{2+} exclusively prefers oxygen in its binding pocket). Nitrogen is more polarizable than oxygen. Therefore, Mn^{2+} could be classified as "softer" than Mg^{2+}, implying that Mn^{2+} is relatively more polarizable than Mg^{2+}. Thus, although Mn^{2+} and Mg^{2+} have the same charge and similar ionic radii (differing by only 0.11Å), Mn^{2+} can accept more charge from the binding pocket residues (better Lewis acid) than Mg^{2+}, and the favourable charge-transfer result in preferable Mn^{2+} binding relative to Mg^{2+}. Thus, it appears that Cas1 ensures favourable binding of Mn^{2+} over Mg^{2+} by inserting nitrogen-containing histidine (H254 of **Figure 5.1**) in the catalytic pocket.

5.3.3 MD insight into the Mg^{2+} and Ca^{2+} bound binding pocket of WT Cas1

The overall structure of Mg^{2+} bound WT Cas1 is very similar to its template Mn^{2+} bound X-ray structure (RMSD= 1.21 ± 0.04 Å, **Figure 5.3 and Figure 5.7 a-c**). However, an inspection of the metal ion binding pocket reveals residues (Q137, E190, D272, and K271) rearrange the interaction network (1st and 2nd coordination shell) around Mg^{2+}/Ca^{2+} (**Figure 5.7 c-d**) relative to the X-ray structure (**Figure 5.7 b**).

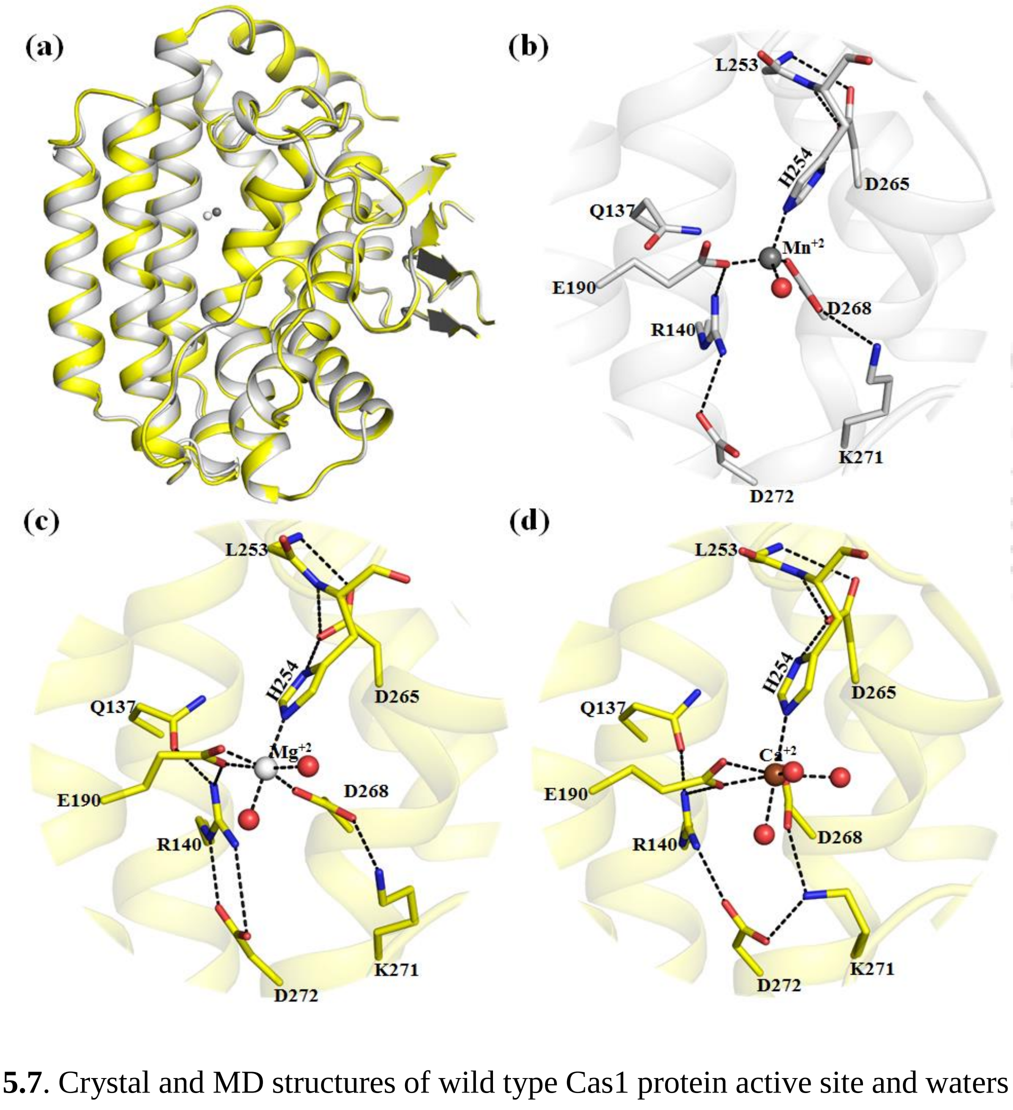


Figure 5.7. Crystal and MD structures of wild type Cas1 protein active site and waters around the divalent ions (Mn^{2+}, Mg^{2+}, and Ca^{2+}) (a) Crystal (grey) vs MD (yellow) structure (b) Mn^{2+} bound crystal structure (c) Mg^{2+}-bound MD structure (d) Ca^{2+}-bound MD structure. For clarity hydrogens (in the zoomed-in view) were not shown.

Robust features observed from the simulations of Mg^{2+} bound WT Cas1 are (1) Bifurcated E190-Mg^{2+} interaction, (2) R140-E190 and K271-D268 salt-bridge interactions. It can be argued that Mg^{2+} substation results in more compact divalent metal ion binding pocket, which include Mg^{2+} coordination with three protein residues (E190, D268, H254) and two water molecules in its first coordination sphere (**Figure 5.7 c**), (3) Mg^{2+} seems to anchor two helices and a loop (helix 4, helix 6, and L5/6), (4) Mg^{2+} is surrounded by shells of alternating charges (shell A: negatively charged including residues E190, D268; shell B: positively charged including residues R140, K271), and (5) Side-chain interaction involving Q137 and R140 residues. Ca^{2+} substitution does not disrupt the interaction network observed as in the case of Mg^{2+} bound binding pocket but results in an additional water entry in the binding pocket (**Figure 5.7 d**).

5.3.4 MD insight into the Mg^{2+} and Ca^{2+} bound binding pocket of mutated Cas1

Mutation of the residues (E190A, H254A, D268A) directly coordinating with the Mg^{2+} disrupt the helix-loop cross talk (**Figure 5.7, and 5.8**) and increases the wetness of the metal ion-binding pocket by introducing additional water molecules in the first coordination shell of the Mg^{2+} ion (**Figure 5.8 a, c, and e**). Ca^{2+} substitution in the mutated Cas1 introduces additional water in the first coordination shell of Ca^{2+} (**Figure 5.8 b, d, and f**), keeping the interaction network intact as observed in the case of Mg^{2+}. Irrespective of the nature of divalent metal ion, MD simulations confirm stable hydrogen-bonding interaction (**Figure 5.7 c-d, 5.8 a-f, 5.9 a-b, and Table 5.1**) between side-chains of (1) D265: H254 (except for H254A, and D265A mutant), (2) R140: E190 (except for E190A), (3) K271: D268 (except for D268A), and (4) Q137: R140. It seems residues within the first coordination shell of the divalent metal ion were held in place by the second-shell residues (R140, D265, and K271). R140 side-chain is oriented by Q137 and D272 residues. Simulations with H254A mutant show H-bonding involving side-chain of D265 and backbone of A254 (**Figure 5.8 c and d**). Second shell mutation (D265A) does not alter the first-shell interaction network of the divalent metal ion, resemble WT binding pocket (**Figure 5.9**).

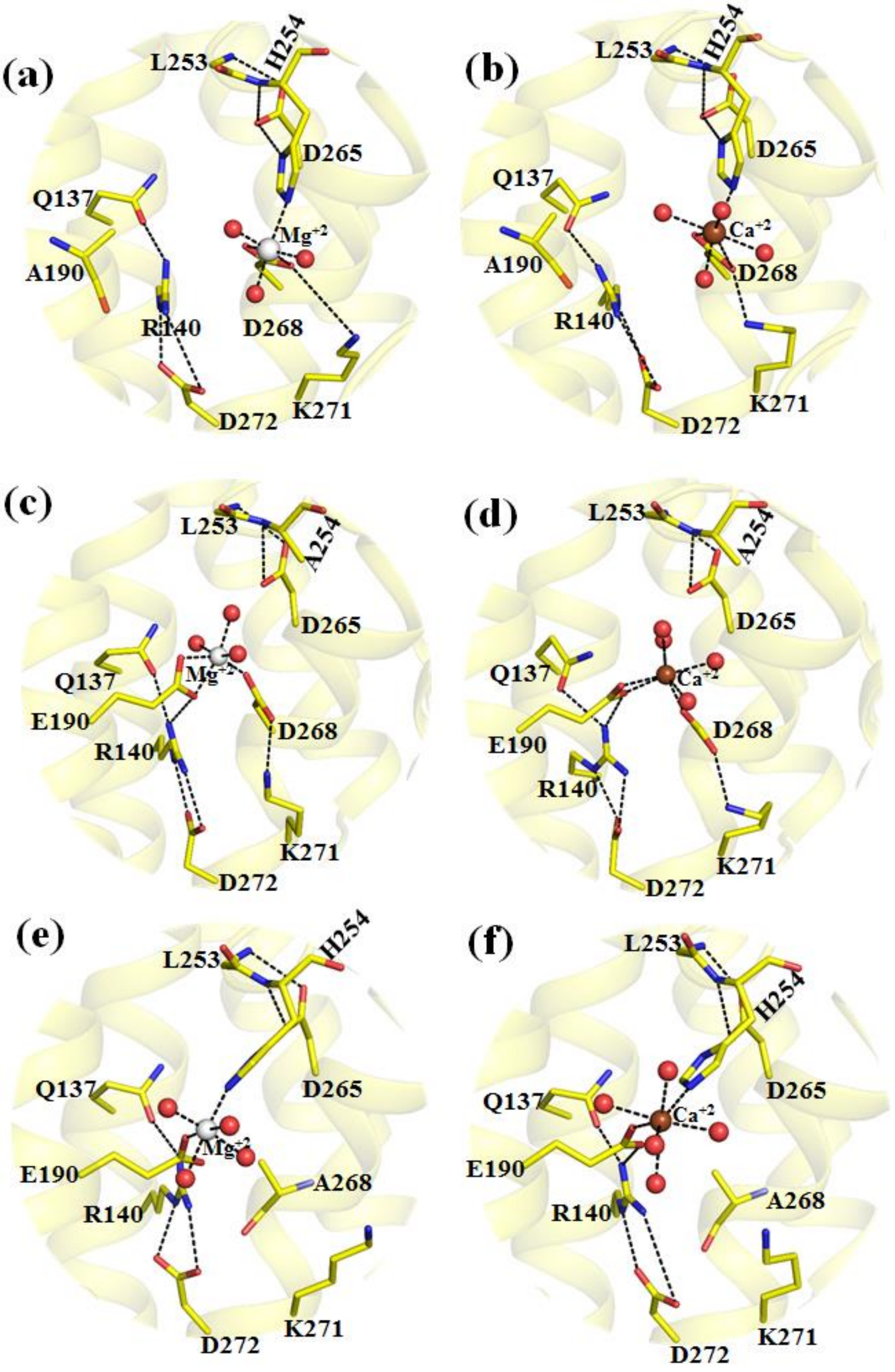

(a)
L253
H254
D265
Q137
Mg+2
A190
R140
D268
D272
K271
(b)
L253
H254
D265
Q137
Ca+2
A190
D268
R140
K271
D272
(c)
L253
A254
D265
Mg+2
Q137
E190
D268
R140
K271
D272
(d)
L253
A254
D265
Ca+2
Q137
D268
E190
R140
K271
D272
(e)
H254
L253
Q137
D265
Mg+2
E190
A268
R140
K271
D272
(f)
L253
H254
D265
Q137
Ca+2
E190
A268
R140
D272
K271

Figure 5.8. MD insight into the Mg^{2+} & Ca^{2+} bound Cas1 active site (mutated) (a,b) E190A mutated Mg^{2+} & Ca^{2+} bound Cas1 active site (c,d) H254A mutated Mg^{2+} & Ca^{2+} bound Cas1 active site (e,f) D268A mutated Mg^{2+} & Ca^{2+} bound cas1 active site. For clarity hydrogens (in the zoomed-in view) were not shown.

Divalent metal ion solvation depends on the nature of the Cas1 binding pocket. The divalent metal ion is in a relatively dry binding pocket (with 2-3 water molecules in the first coordination sphere) for WT and D265A mutant (**Figure 5.7 c, 5.9, and Table 5.1**), relative to its first shell mutant analogues (E190A, H254A, D268A) in which metal-ion coordinates with 3-5 water molecules (**Figure 5.8, Table 5.1**). The smaller size of Mg^{2+} in the binding site results in compact and dry (**Table 5.1**) binding pocket. Ca^{2+} substitution leads to bigger and relatively wet (allowing an additional water entry) binding pocket.

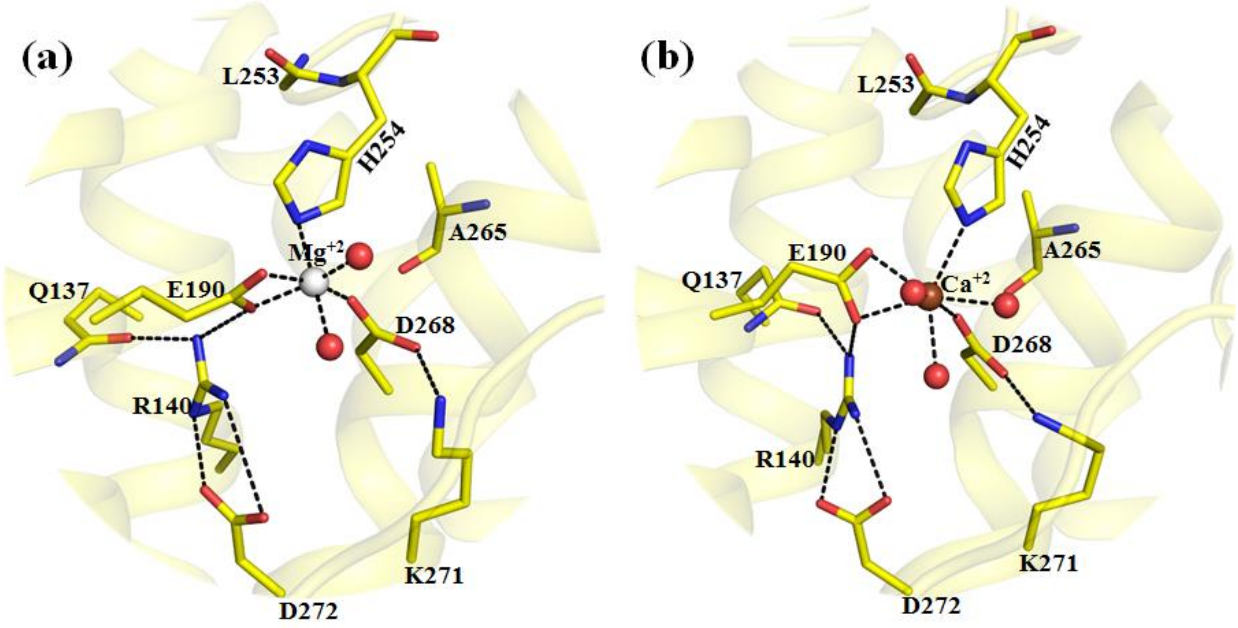


Figure 5.9. MD structures of D265A mutant cas1 site and waters around the divalent ions (a) Mg^{2+}-bound (b) Ca^{2+}-bound.

Table 5.1. Selected interatomic distances (in Å) in the active site of Cas1 protein.

Interacting residues	Wild type		E190A		H254A		D265A		D268A	
	Mg^{2+}	Ca^{2+}	Mg^{2+}	Ca^{2+}	Mg^{2+}	Ca^{2+}	Mg^{2+}	Ca^{2+}	Mg^{2+}	Ca^{2+}
Mg^{2+}/Ca^{2+} : OD2-Asp268	1.81(0.04)	2.22(0.07)	1.88(0.05)	2.11(0.05)	1.82(0.04)	2.13(0.06)	1.80(0.03)	2.12(0.05)	-	-
Mg^{2+}/Ca^{2+} : NE2-His254	2.10(0.07)	2.36(0.08)	2.15(0.09)	2.33(0.07)	-	-	2.12 (0.08)	2.34 (0.06)	2.12 (0.08)	2.38(0.08)
Mg^{2+}/Ca^{2+} : OE2-Glu190	1.89(0.06)	2.23(0.08)	-	-	1.89(0.06)	2.25(0.10)	1.90 (0.06)	2.24 (0.08)	1.82 (0.04)	2.13(0.06)
Mg^{2+}/Ca^{2+} : OE1-Glu190	1.93(0.07)	2.27(0.12)	-	-	1.94(0.07)	2.24(0.08)	1.93 (0.07)	2.27 (0.09)	-	-
Mg^{2+}/Ca^{2+} : OD1-Asp268	-	-	1.91(0.06)	2.47(0.12)	-	-	-	-	-	-
Glu190-OE1 : NH1Arg140	2.72(0.10)	2.77(0.21)	-	-	2.73(0.11)	2.71(0.11)	2.77 (0.13)	2.74 (0.11)	2.72 (0.10)	2.78(0.16)
Arg140-NE : OD2-Asp272	2.77(0.14)	2.94(0.23)	2.79(0.14)	2.79(0.15)	2.72(0.10)	2.77(0.14)	2.77 (0.14)	3.05 (0.20)	2.80 (0.16)	2.80(0.16)
Arg140-NH2 : OD1Asp272	2.72(0.12)	2.66(0.09)	2.72(0.11)	2.80(0.20)	2.71(0.16)	2.76(0.18)	3.10 (0.25)	2.89 (0.41)	2.76 (0.14)	2.77(0.15)
Arg140-NH1 : OE1Gln137	2.76(0.11)	2.75(0.10)	2.98(0.33)	2.96(0.36)	2.89(0.16)	2.78(0.13)	2.88 (0.23)	2.83 (0.20)	2.78 (0.12)	2.80(0.13)
Asp268-OD1 : NZ-Lys271	2.68(0.10)	2.76(0.17)	4.68(0.54)	2.76(0.31)	2.70(0.13)	2.67(0.10)	2.68 (0.12)	2.67 (0.09)	-	-
Asp265-OD1 : ND1His254	2.69(0.09)	2.67(0.08)	2.69(0.10)	2.71(0.10)	-	-	-	-	2.99 (0.26)	3.01(0.25)
Asp265-OD1 : N-His254	2.83(0.18)	2.78(0.10)	3.10(0.24)	3.29(0.27)	-	-	-	-	3.01 (0.20)	3.05(0.22)
Asp265-OD1 : N-Ala254	-	-	-	-	2.95(0.16)	2.96(0.16)	-	-	-	-
Asp265-OD2 : N-Leu253	2.97(0.18)	3.13(0.22)	-	-	2.76(0.10)	2.80(0.12)	-	-	2.77 (0.10)	2.80(0.11)
Asp272-OD2 : NZ-Lys271	-	2.71(0.11)	-	-	-	-	-	-	-	-
Mg^{2+}/Ca^{2+} : TIP3	2.10(0.11)	2.35(0.10)	2.03(0.08)	2.24(0.06)	2.00(0.07)	2.26(0.07)	1.96(0.06)	2.34(0.09)	1.99(0.06)	2.36(0.08)
Mg^{2+}/Ca^{2+} : TIP3	1.96(0.06)	2.30(0.08)	1.95(0.06)	2.24(0.06)	1.97(0.06)	2.34(0.09)	2.09(0.10)	2.28(0.07)	2.00(0.06)	2.30(0.09)
Mg^{2+}/Ca^{2+} : TIP3	-	2.41(0.15)	1.95(0.06)	2.25(0.07)	2.03(0.07)	2.27(0.07)	-	2.33 (0.09)	2.00 (0.07)	2.31(0.09)
Mg^{2+}/Ca^{2+} : TIP3	-	-	-	2.23(0.06)	-	2.30(0.08)	-	-	2.03(0.07)	2.33(0.13)
Mg^{2+}/Ca^{2+} : TIP3	-	-	-	-	-	-	-	-	-	2.33(0.12)

5.4 Discussion

Universally conserved Cas1 protein of CRISPR immune systems is a metal-dependent endonuclease that produces double-stranded DNA fragments from viral DNA and integrates the same to a bacterial genome. Cas1 catalysed DNA fragmentation is inhibited by the mutation in the metal ion-binding residues. Moreover, the nature of metal ion is crucial for Cas1 function. Cleavage of dsDNA by Cas1 is only possible if Mn^{2+} or Mg^{2+} is bound to the Cas1. Note, Cas1 exhibits the highest dsDNA cleavage activity only with Mn^{2+}. Despite extensive structural and biochemical studies (**Wiedenheft et al., 2009; Nuñez et al., 2015; Wang et al., 2015; Xiao et al., 2017; Shannon, 1976; Wright et al., 2017**), the energetic origin of metal ion specificity of Cas1 protein is not known. Using X-ray structure (**Wiedenheft et al., 2009)** of Cas1 protein (Mn^{2+} bound) as a template, we performed MD simulations of Mg^{2+} and Ca^{2+} bound Cas1 (WT and mutants) and quantified their relative binding affinities (**Figure 5.5**). The results suggest that Mg^{2+} is preferred over Ca^{2+} by ~ 4-5 kcal/mol in Cas1 protein. Smaller size of Mg^{2+} (ionic radii = 0.86Å) relative to Ca^{2+} (ionic radii = 1.14Å) (**Shannon, 1976**) results in higher charge density in Mg^{2+}. A divalent metal ion with higher charge density is known to bind strongly in the negatively charged protein binding pocket (**Jing et al., 2018**). Thus, the use of classical force field (which does not explicitly include many-body polarization effects) in estimating the relative binding affinity of divalent metal ions (Mg^{2+} versus Ca^{2+}) might be inappropriate. Thus, computationally expensive polarizable force fields (**Jing et al., 2018**) may be necessary for relative binding free energy estimation of divalent metal ions (Mg^{2+} and Ca^{2+}) to Cas1. Indeed, it has been shown (**Kohagen et al., 2014**) that Ca^{2+} binding free energies to calmodulin protein estimated by MD simulations can be greatly improved by the inclusion of polarization effect via charge rescaling. Thus, to check if the sign of our estimated relative binding free energies (ΔΔG, **Figure 5.5**), we define a quantity "ΔE" (i.e., single-point abinitio energy difference between the right and left side of eq (5.1, 5.2), See methods) which explicitly include polarization and approximately capture the essence of the relative free energy of binding to some extent. Note, ΔE is limited due to the simplified description of metal ion binding pocket (Divalent metal ion with its first coordination shell embedded into a water dielectric). The objective was not to estimate ΔΔG accurately from abinitio calculations but to focus a quantity that can indicate divalent metal ion preference in a computationally inexpensive way. ΔE as an indicator of metal ion preference and likely to justify the sign of ΔΔG estimated from classical MD simulations. Results obtained from quantum chemical calculations (**Figure 5.6**) suggest that Mg^{2+} indeed energetically preferred relative to Ca^{2+}, as suggested by classical MD simulations. Moreover, Mn^{2+} is preferred relative to both Mg^{2+} and Ca^{2+} in the binding pocket of WT Cas1 (**Figure 5.6**). Note, Cas1 is known to be more active with Mn^{2+} than with Mg^{2+} at physiological conditions, in line with our estimated relative binding preference. The Cas1 protein has a relatively rigid metal ion-binding core (**Figure 5.7**) and softer outer regions, especially h2, h4, and h6 and loop between h5 and h6 (**Figure 5.4**). Simulation suggests that the first coordination shell of Mg^{2+} and Ca^{2+} in the Cas1 binding pocket are more or less similar, except the presence of single extra water molecule for the later (**Figure 5.7, 5.8, and 5.9**). Mutation of residues of the first-coordination shell (E190A, H254A, and D268A) of the divalent metal ion results in water exposure relative to WT Cas1 (**Figure 5.8**). D265A mutation does not disrupt the first shell interaction network of the divalent metal ion as observed in WT Cas1. MD simulations of free Mg^{2+} and Ca^{2+} in explicit water suggest that the ions are coordinated with 6 and 7 water molecules respectively in its first coordination shell (**appendix Figure 5.2**). The average distance between metal ion and oxygen of waters are between 1.9-2.1Å (for Mg^{2+}) and 2.2-2.4Å (for Ca^{2+}) respectively, consequence of larger atomic radii of the later. Note, the structures of Cas1 bound to Mg^{2+}/Ca^{2+} has not been characterized experimentally till now. The coordination number of the divalent metal ion is the same as those observed in the free water simulation (6 for Mg^{2+} and 7 for Ca^{2+}). The number of water molecules coordinating with the metal ion depends on the nature of the divalent metal ion and its coordinating amino-acid residues (**Table 5.1, Figure 5.7, 5.8, and 5.9**). WT and D265A mutants are dry which include only 2/3 water molecules coordinating Mg^{2+}/Ca^{2+}. Whereas D268A mutant shows maximum wetness, which includes 4/5 water molecules coordinating Mg^{2+}/Ca^{2+}. E190A and H254A mutants could be considered as of intermediate wetness, consist of 3/4 water molecules coordinating with Mg^{2+}/Ca^{2+}. The magnitude of divalent metal ion preference (favouring Mg^{2+} over Ca^{2+}) estimated from quantum chemical calculations follow the order (WT & D265A) > (E190A & H254A) > D268A (**Figure 5.5 and 5.6**). Classical MD also suggests the discriminatory power (in favour of Mg^{2+} binding relative to Ca^{2+}) of WT Cas1 (dry binding pocket) is highest and lowest for D268A mutant (wet binding pocket). Thus, it can be argued that the wetness of the binding pocket is linked to divalent metal ion selectivity; dryness results in high selectivity. Indeed Mg^{2+} and Ca^{2+} have comparable binding affinities with acetate in water ($\Delta G^0_{bind} \sim -1.73$ versus -1.6 kcal/mol at 300K) (**Xu et al., 2017**). The similarity of binding affinities (between Mg^{2+} and Ca^{2+}) may be due to the dielectric screening of the water (**Liu et al., 2017**). The various divalent metal ion (viz., Mn^{2+}, Mg^{2+}, Ca^{2+} etc.) binding affinities to Cas1 catalytic pocket are experimentally unknown. Thus, our estimated divalent metal ion selectivity (ΔΔG's) cannot be proved or disproved. We believe our work will encourage experimental verification. The effects of metal ions on the structure and stability of a protein could be investigated using a combination of circular dichroism and protein melting experiments (**Sissi et al., 2005**). Cas1 in various buffers (differing only by divalent metal ion Mn^{2+}, Mg^{2+}, Ca^{2+}) might be subjected to CD experiments and thermal denaturation experiments. The secondary structure content and Cas1 stability at various buffers might be monitored in the CD and protein melting experiments respectively. CD and NMR experiments as a function of metal (Mn^{2+} or Mg^{2+} or Ca^{2+})

titration to Cas1 protein might also provide the Cas1:metal dissociation constant (**Ye et al., 2001**). In principle, the K_d values will be obtained by fitting the titration curves from the CD signals and NMR chemical shift changes. However, experimental verification might be challenging because our estimated metal ion selectivity is related to only a specific site of the Cas1 protein. The role of additional active site metals cannot be ruled out but has not yet been identified by experimental studies. It is believed that single Mn^{2+} in the Cas1 binding pocket coordinates the nuclease activity (**Yang et, 2008**). Computer simulation suggests that Mg^{2+} is preferred over Ca^{2+} by Cas1. Moreover, divalent metal ion substitution (Mg^{2+} by Ca^{2+}) or mutations (E190A, H254A, and D265A) alter the solvent accessibility of the divalent metal ion binding pocket, which might be linked to its selectivity.

5.5 Conclusion

Efficient divalent-metal ion discrimination is essential for Cas1 mediated endonuclease activity. Cas1 is known to be active with Mn^{2+} and Mg^{2+} (highest activity with the former) but loses its activity with Ca^{2+}. Mutation of the Cas1 protein residues in the divalent metal ion-binding site inhibits Cas1-mediated DNA degradation. Using the crystal structure of Cas1 protein as a template, we modelled and performed classical molecular dynamics simulations with Mg^{2+}, Ca^{2+} bound WT, and mutant Cas1 proteins (**Figure 5.7, 5.8, and 5.9**), and attempted to estimate the divalent metal ion selectivity in the binding pocket. Classical molecular dynamics free energy calculations suggest that Cas1 imposes energy penalty of ~ 4-5 kcal/mol for Ca^{2+} binding relative to Mg^{2+} (**Figure 5.5**). Quantum chemical calculations indicate: (1) Mg^{2+} binding is preferred over Ca^{2+}, in line with the results obtained from classical MD simulations, (2) Mn^{2+} is the most preferred ion in the WT Cas1 binding pocket among all the three ions studied in this work (i.e. Mg^{2+}, Ca^{2+}, Mn^{2+}) (**Figure 5.6**). The metal ion binding pocket of Cas1 is dry when bound to Mg^{2+}. Ca^{2+} substitution results in a single water entry (wet) in the ion-binding pocket (**Figure 5.7 c, and d**). Mutation of conserved amino acids that coordinates with the divalent metal ion also results in water exposure of the metal ion binding pocket (**Figure 5.8, and 5.9**). The strength of divalent metal ion selectivity depends on the wetness of the binding pocket. Wet binding pocket is less selective, and dry binding pocket is more selective in favor of Mg^{2+} binding relative to Ca^{2+}.

Chapter 6

Overall conclusion and future perspective

Metal ions play key roles in a variety of cellular processes and serve a variety of functions within their respective biological processes. Recent advancements in the structural characterization of metal ion bound biomolecular complexes (RNA/protein) and the ever-increasing computational power, make structure-based computational analysis as a powerful tool for understanding the biology in terms of structure, dynamics and energetics in atomic detail. In this dissertation, structure-based classical molecular dynamics free energy calculations and ab initio quantum calculations were used to determine the thermodynamics of metal ion selectivity in RNA (group II intron) and protein (Cas1) systems. An attempt has been made to correlate computed energetics (strength of metal ion selectivity) with 3D structures or biomolecular structures.

In the **Chapter 1**, a brief overview of the group II intron, Cas1 protein and the methodology adopted in this thesis was discussed. Objectives of the thesis is described at the end of this chapter.

Chapter 2 discussed the principle of monovalent metal ion selectivity (K^+ vs. Na^+) in group II intron at various stages of the splicing pathway (pre-hydrolytic, post-hydrolytic, and ligand-free group-II intron) of a bacteria *Oceanobaciellus iheyensis*. We found that the strength of monovalent ion selectivity (K^+ versus Na^+ binding to intron: $\Delta\Delta G$) changes along the self-splicing pathway. The ligand-free and pre-hydrolytic stages are highly selective for K^+ over Na^+, whereas post-hydrolytic show weak selectivity. Na^+ in the free and pre-hydrolytic states was shown to be trapped in the dry pocket with an unsatisfied first coordination shell, resulting in strong discrimination ($\Delta\Delta G$ =6-7 kcal/mol) in favor of K^+. On the other hand, the post-hydrolytic state allows water entry and fulfills the Na^+ bonding requirement, resulting in minimal discrimination ($\Delta\Delta G$ ~1 kcal/mol).

Chapter 3 provides insights into the energetics of divalent metal ion selectivity (Mg^{2+} vs. Ca^{2+}) in bacterial (*Oceanobaciellus iheyensis*) group II intron at various stages of splicing pathway (pre-hydrolytic, post-hydrolytic, and ligand-free group-II intron). Estimated energetics show that group II intron imposes a very high energetic penalty (9-14 kcal/mol) for Ca^{2+} binding in all three stages of splicing. The orientation of nucleophilic water relative to scissile phosphate is strikingly different between Mg^{2+} and Ca^{2+} bound pre-hydrolytic state. The nature of the divalent metal-ion (Mg^{2+} or Ca^{+2}) does not disrupt the overall architecture/integrity of group II intron but could perturb the local environment, which might be sufficient to modulate the catalytic activity. Furthermore, ab initio calculations ensured that the sign of ΔΔG (Mg^{2+} vs. Ca^{2+}) estimated from classical MD is correct.

Chapter 4 discussed the divalent (Mg^{2+} vs. Ca^{2+}) and monovalent (K^+ vs. Na^+) ion selectivity in group II intron of a eukaryotic brown algae *Pylaiella littoralis*. Two states of lariat intron were considered for this study, i.e., pre-hydrolytic and post-hydrolytic states. Estimated thermodynamics (ΔΔG) suggests that lariat group II intron is strongly selective for divalent-metal ions (ΔΔG ~ 12-19 kcal/mol, Mg^{2+} versus Ca^{2+}) but non-selective for monovalent ions (ΔΔG ~ 0.0 kcal/mol, K^+ versus Na^+). Divalent metal ion binding pocket was found to be compact and dry. Thus, high selectivity was ensured by trapping the Ca^{2+} in the binding pocket with the unsatisfied first-shell coordination sphere. Moreover, the direct interaction between Mg^{2+} and ribose sugar of the 5'-exon (observed in the X-ray and MD structures) was disrupted in response to divalent-metal ion substitution by Ca^{2+}. The bonding requirement of the monovalent ion (K^+ or Na^+) is satisfied in the lariat intron resulting in weak selectivity.

Chapter 5 has explored the energetics of divalent metal ion selectivity (Mg^{2+} vs. Ca^{2+}) in the CRISPR-Cas system associated Cas1 protein. Classical molecular dynamics free energy calculations suggest that Cas1 imposes the free energy penalty (ΔΔG ~ 4-5 kcal/mol) for Ca^{2+} binding relative to Mg^{2+}. Quantum chemical calculations indicate: (1) Mg^{2+} binding is preferred over Ca^{2+}, in line with the results obtained from classical MD simulations, (2) Mn^{2+} is the most preferred ion in the WT Cas1 binding pocket among all the three ions studied in this work (i.e., Mg^{2+}, Ca^{2+}, Mn^{2+}). The metal ion binding pocket of Cas1 is dry when bound to Mg^{2+}. Ca^{2+} substitution results in a single water entry in the ion-binding pocket. Mutation of conserved amino acids that coordinates with the divalent metal ion also results in water exposure of the metal ion binding pocket. The strength of divalent metal ion selectivity depends on the wetness of the binding pocket. Wet binding pocket is less selective, and dry binding pocket is more selective in favor of Mg^{2+} binding relative to Ca^{2+}.

In this dissertation, we established that structure-based molecular dynamics free energy calculations are a viable and reliable methodology for answering some of the most fundamental concerns underlying metal ion selectivity in biological systems comprising proteins and RNAs. Water accessibility of the metal-ion binding pocket seems to be linked with the strength of recognition by biomolecules viz., group II intron, and Cas1 protein. Dry binding pocket results in high selectivity, whereas water accessibility diminishes the strength of recognition. Our approach of combining classical MD and quantum chemical calculations is relatively less expensive and successful in establishing the link between 3D atomic structures (cognate or near-cognate complexes) of the biomolecule and free energy differences. The same structure-based simulation strategy could be used to estimate the strength of recognition for related RNA and protein in general.

Appendices

Appendix 2

Table 2.1. Calculated free energetics (in kcal/mol) and simulation time scale (in *ns*) for the alchemical transformation of K1 → Na1 in free and intron bound conformations.

System	Alchemical transformation	Simulation length (Time in nano seconds)			ΔG
		Equilibration	Free Dynamics	Free energy calculations by Thermodynamic integration	
In solution	$K^+ \rightarrow Na^+$	0.28	2	22	-23.47 (0.41)
		0.28	2	22	-23.45 (0.35)
		0.32	2	22	-23.36 (0.56)
		0.32	2	22	-23.28 (0.28)
		0.32	2	22	-23.51 (0.37)
		Average over 5 replicas			**-23.41±0.39**
Ligand free state	$K^+ \rightarrow Na^+$	0.28	2	22	-15.91 (0.17)
		0.49	2	22	-16.38 (0.35)
		0.32	2	22	-15.55 (0.30)
		0.32	2	22	-15.73 (0.50)
		0.32	2	22	-15.57 (0.29)
		0.38	5	55	-15.56 (0.59)
		Average over 6 replicas			**- 15.78±0.36**
Pre-hydrolytic state	$K^+ \rightarrow Na^+$	0.32	2	22	-17.66 (0.59)
		0.32	2	22	-17.41 (0.65)
		0.63	2	22	-16.70 (0.44)
		0.63	2	22	-17.54 (1.02)
		0.32	2	22	-17.23 (0.30)
		Average over 5 replicas			**-17.30±0.60**
Post-hydrolytic state	$K^+ \rightarrow Na^+$	0.20	2	22	-23.66 (0.73)
		0.26	2	22	-22.65 (0.90)
		0.24	2	22	-21.63 (1.05)
		0.30	2	22	-22.03 (0.48)
		Average over 4 replicas			**-22.49±0.79**

Table 2.2: Free energy derivative ∂G/∂λ vs λ plot for alchemical transformation $K^+ \rightarrow Na^+$. Free energies in kcal/mol. The MD trajectories were divided into two equal halves and the difference between the computed ∂G/∂λ's from the two halves is reported as uncertainty in the parenthesis at each λ points.

$K^+ \rightarrow Na^+$ in solution

λ	**Run 1**	**Run 2**	**Run 3**	**Run 4**	**Run 5**
1.0	07.25 (0.06)	07.12 (0.20)	07.28 (0.02)	07.28 (0.27)	07.20 (0.12)
0.9	08.97 (0.22)	08.95 (0.06)	08.78 (0.19)	08.91 (0.02)	08.98 (0.23)
0.8	10.52 (0.00)	11.05 (0.22)	10.95 (0.07)	10.80 (0.36)	11.05 (0.30)
0.7	13.23 (0.11)	13.60 (0.31)	13.33 (0.33)	13.63 (0.00)	13.17 (0.70)
0.6	16.49 (0.11)	16.41 (0.31)	16.46 (0.71)	15.99 (0.38)	16.12 (0.21)
0.5	20.00 (0.29)	20.23 (0.90)	20.01 (0.18)	19.89 (0.02)	20.25 (0.22)
0.4	24.41 (0.32)	24.24 (0.19)	24.19 (1.02)	24.12 (0.70)	24.52 (0.24)
0.3	29.87 (1.23)	29.76 (0.00)	29.17 (0.00)	30.11 (0.56)	30.40 (0.62)
0.2	36.35 (0.28)	35.90 (0.10)	36.24 (0.38)	35.08 (0.28)	36.34 (0.90)
0.1	44.38 (1.72)	44.27 (0.88)	43.91 (1.35)	43.88 (0.52)	43.88 (0.52)
0.0	53.73 (0.24)	53.01 (0.75)	53.92 (2.01)	53.56 (0.02)	53.59 (0.02)
ΔG	**-23.47 (0.41)**	**-23.45 (0.35)**	**-23.36 (0.56)**	**-23.28 (0.28)**	**-23.51 (0.37)**

$K^+ \rightarrow Na^+$ in ligand-free group II intron

λ	**Run 1**	**Run 2**	**Run 3**	**Run 4**	**Run 5**	**Run 6**
1.0	06.06 (0.06)	05.95 (0.15)	05.11 (0.11)	05.65 (0.47)	05.94 (0.08)	05.98(0.15)
0.9	07.33 (0.16)	07.15 (0.21)	05.86 (0.51)	06.68 (0.03)	07.01 (0.00)	06.62(0.18)
0.8	08.25 (0.22)	08.54 (0.04)	07.32 (0.06)	08.62 (0.03)	07.73 (0.29)	08.57(0.37)
0.7	09.51 (0.11)	09.86 (0.13)	09.00 (0.53)	10.16 (0.21)	09.23 (0.70)	09.73(0.53)
0.6	11.59 (0.10)	12.33 (0.16)	11.22 (0.24)	11.83 (0.55)	11.41 (0.38)	11.74(0.29)
0.5	13.94 (0.05)	14.48 (0.38)	14.08 (0.08)	13.53 (0.00)	13.84 (1.01)	13.30(0.20)
0.4	16.81 (0.12)	17.88 (0.27)	16.24 (0.03)	16.20 (0.52)	15.76 (0.20)	16.31(1.02)
0.3	19.55 (0.12)	20.40 (0.82)	19.88 (0.65)	19.27 (0.43)	19.37 (0.31)	19.23(1.21)
0.2	23.30 (0.36)	23.66 (0.19)	24.31 (0.43)	23.94 (0.82)	22.40 (0.05)	22.58(0.09)
0.1	28.70 (0.34)	28.78 (0.14)	28.73 (0.15)	27.67 (0.70)	29.41 (0.17)	27.56(1.24)
0.0	34.17 (0.26)	35.42 (1.37)	32.54 (0.53)	33.15 (1.78)	33.18 (0.10)	33.99(1.22)
ΔG	**-15.91(0.17)**	**-16.38(0.35)**	**-15.55(0.30)**	**-15.73(0.50)**	**-15.57(0.29)**	**15.56 (0.59)**

$K^{+} \rightarrow Na^{+}$ in pre-hydrolytic group II intron state

λ	Run 1	Run 2	Run 3	Run 4	Run 5
1.0	06.71 (0.16)	06.38 (0.24)	06.09 (0.04)	06.49 (0.50)	06.63 (0.19)
0.9	07.84 (0.13)	07.22 (0.17)	07.65 (0.07)	07.83 (0.26)	07.86 (0.29)
0.8	09.35 (0.00)	09.37 (0.25)	08.72 (0.42)	09.23 (0.24)	08.83 (0.07)
0.7	11.30 (0.36)	11.25 (1.04)	10.86 (0.09)	10.56 (0.19)	11.47 (0.06)
0.6	13.83 (0.17)	13.34 (0.05)	12.61 (0.76)	12.31 (0.34)	13.15 (0.11)
0.5	15.34 (0.68)	15.51 (0.38)	15.26 (0.10)	14.80 (0.31)	15.36 (0.20)
0.4	18.20 (0.74)	18.29 (0.38)	17.78 (0.29)	17.82 (0.45)	18.08 (0.18)
0.3	21.30 (0.30)	21.83 (0.40)	20.34 (1.02)	20.83 (0.70)	22.02 (0.43)
0.2	27.23 (0.83)	25.99 (1.74)	24.68 (0.19)	24.89 (0.18)	25.25 (0.24)
0.1	32.10 (2.12)	29.57 (1.02)	28.95 (0.25)	28.94 (0.16)	29.15 (0.93)
0.0	33.51 (1.04)	37.16 (1.56)	34.17 (1.66)	49.88 (7.94)	35.64 (0.63)
ΔG	**-17.66 (0.59**	**-17.41 (0.65)**	**-16.70 (0.44)**	**-17.54 (1.02)**	**-17.23 (0.30)**

$K^{+} \rightarrow Na^{+}$ in post-hydrolytic group II intron state

λ	Run 1	Run 2	Run 3	Run 4
1.0	06.06 (0.22)	07.07 (0.22)	05.69 (0.02)	05.29 (0.03)
0.9	07.20 (0.53)	09.74 (0.76)	07.96 (0.09)	06.10 (0.01)
0.8	10.49 (0.24)	10.94 (0.59)	09.11 (0.04)	07.69 (0.55)
0.7	13.63 (0.35)	13.93 (0.12)	12.45 (1.21)	10.16 (0.77)
0.6	18.13 (1.26)	16.33 (0.25)	16.87 (0.62)	17.21 (0.61)
0.5	19.67 (0.62)	18.47 (0.37)	19.59 (1.26)	21.43 (0.26)
0.4	25.63 (0.45)	24.09 (3.08)	23.04 (3.19)	25.48 (0.80)
0.3	31.28 (1.04)	29.25 (3.09)	27.64 (1.45)	30.24 (0.25)
0.2	36.73 (0.90)	32.08 (0.38)	34.18 (1.36)	34.89 (0.50)
0.1	45.15 (1.90)	42.41 (0.39)	37.67 (2.25)	39.93 (0.64)
0.0	51.37 (0.61)	51.39 (0.72)	49.81 (0.16)	49.00 (0.86)
ΔG	**-23.66(0.73)**	**-22.65(0.90)**	**-21.63(1.05)**	**-22.03(0.48)**

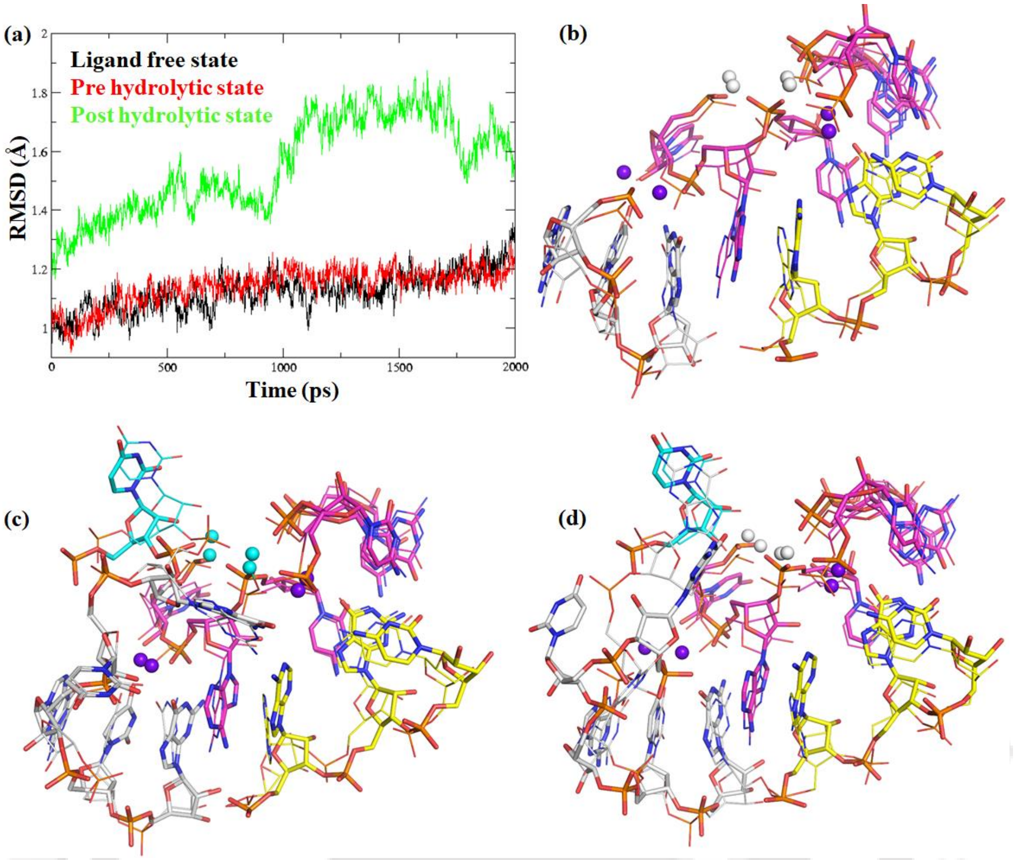


Figure 2.1. Root-mean-square deviation and MD (sticks) vs X-ray (line) structural comparison (a) RMSD of heavy atoms within 22Å of unrestrained simulation sphere with respect to X-ray structure. RMSD is shown for 2ns MD trajectory averaging over 2ps interval. (b) Ligand Free intron (c) intron in Pre-hydrolytic state (d) intron in Post-hydrolytic state.

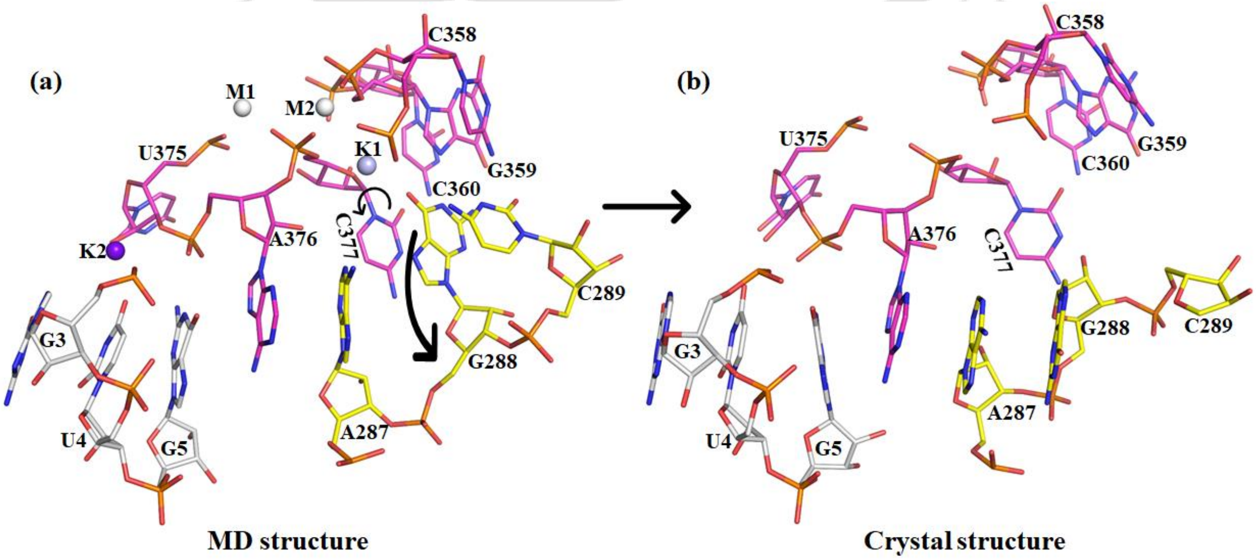


Figure 2.2. (a) Active site of Na1 bound group II intron: MD insight of the high energy state. (b) X-ray structure of group II intron in presence of Na^+ and Mg^{2+}. No ions were bound to the X-ray structure and different conformation of CYT377 and ADE287 is visible (marked with arrow in the MD structure).

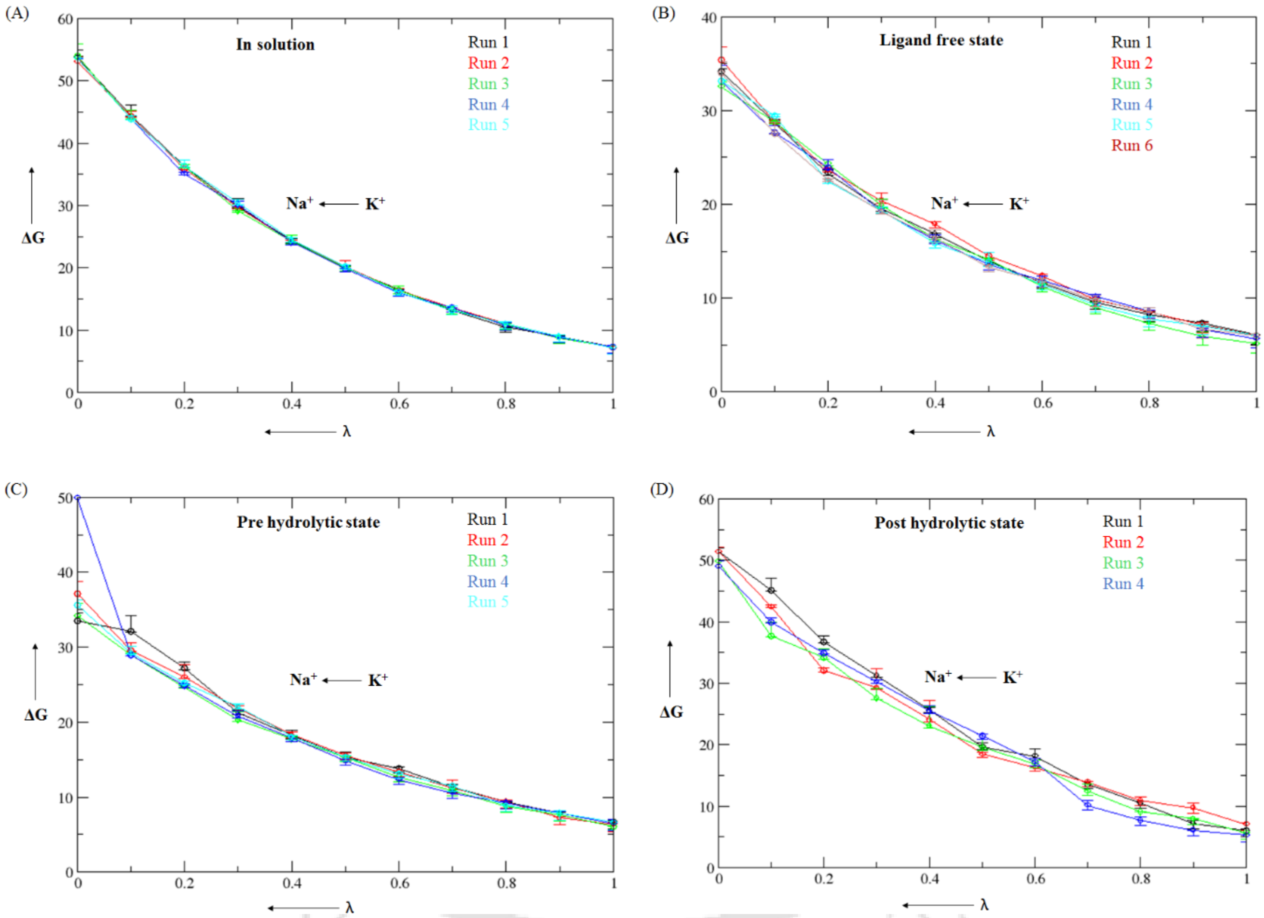


Figure 2.3. Plot of free energy derivative vs λ (coupling coordinate) for the alchemical transformation of $K^+ \rightarrow Na^+$ in the free and intron bound state for different replicas (denoted with different color). Data for different replicas are shown in different color. Uncertainties are shown as vertical bars (often too small to be seen). Total number of λ points = 11.

Appendix 3

Table 3.1. Alchemical coordinate λ was defined and connects end states i.e, λ =1 corresponds to Mg^{2+} and λ =0 corresponds to Ca^{2+}. Free energy derivative $\partial G/\partial \lambda$ was calculated for alchemical transformation $Mg^{+2} \rightarrow Ca^{+2}$ for 11λ points between 1 and 0. Free energy derivative was calculated as $\partial G/\partial \lambda = \langle U(\lambda=1) – U(\lambda=1)\rangle_{\lambda}$, where < > represents averaging over MD trajectory at a specific value of λ. Alchemical transformation associated free energies (in kcal/mol) was calculated by numerically integrating $\partial G/\partial \lambda$ vs λ plot. The MD trajectories were divided into two equal halves and the difference between the computed $\partial G/\partial \lambda$'s from the two halves is reported as uncertainty in the parenthesis at each λ point. Derivative at each λ is given below:

$Mg^{2+} \rightarrow Ca^{2+}$ in solution

λ	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6
1.0	-153.84(0.16)	-154.32(0.71)	-155.68(0.08)	-154.70 (0.53)	-154.85 (0.80)	-154.39 (1.78)
0.9	-121.86(0.33)	-122.17(0.07)	-121.06(1.49)	-121.23 (0.73)	-121.03 (1.34)	-121.43 (0.32)
0.8	-100.52(0.68)	-100.19(1.13)	-99.68(0.81)	-100.28 (0.08)	-99.83 (0.92)	-100.40 (0.10)
0.7	-84.57(0.42)	-84.15(0.54)	-84.29(0.36)	-83.57 (0.01)	-84.65 (0.40)	-84.29 (0.34)
0.6	-72.25(0.05)	-71.84(0.67)	-72.20(0.27)	-71.67 (0.01)	-72.30 (1.51)	-71.62 (0.46)
0.5	-62.09(0.44)	-62.06(0.40)	-61.58(0.72)	-61.76 (0.14)	-61.75 (0.58)	-61.55 (0.90)
0.4	-53.08(0.77)	-53.75(0.95)	-53.00(0.62)	-52.40 (2.23)	-52.69 (1.68)	-53.13 (0.08)
0.3	-44.73(0.69)	-44.51(1.38)	-45.29(0.74)	-45.04 (0.72)	-44.16 (0.90)	-45.47 (0.20)
0.2	-34.74(1.64)	-34.47(0.15)	-34.13(1.97)	-34.11 (0.56)	-34.65 (2.56)	-34.70 (0.35)
0.1	-27.67(1.53)	-26.93(0.24)	-27.41(0.26)	-27.44 (0.44)	-26.79 (0.33)	-26.77 (0.29)
0.0	-24.05(0.38)	-23.55(0.43)	-23.64(0.63)	-23.17 (0.40)	-23.30 (0.50)	-23.09 (0.36)
ΔG	**69.05(0.64)**	**68.90(0.60)**	**68.83(0.72)**	**68.64 (0.50)**	**68.69 (1.04)**	**68.81 (0.47)**

$Mg^{2+} \rightarrow Ca^{2+}$ in ligand-free group II intron

λ	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6
1.0	-337.13(2.75)	-340.12(0.32)	-341.38(1.83)	-343.90(0.26)	-339.65(0.79)	-341.13(1.95)
0.9	-261.11(0.11)	-265.20(1.66)	-263.86(1.85)	-267.75(0.70)	-262.94(0.44)	-265.16(1.24)
0.8	-212.40(2.08)	-218.11(1.84)	-216.58(1.47)	-217.95(1.84)	-216.62(0.25)	-216.85(0.41)
0.7	-179.76(0.18)	-182.23(0.77)	-181.29(0.91)	-184.14(1.45)	-181.98(0.82)	-182.26(0.71)
0.6	-153.82(0.70)	-155.40(0.20)	-154.74(1.54)	-155.02(1.28)	-154.39(0.08)	-154.48(0.54)
0.5	-131.70(0.61)	-134.20(1.09)	-133.33(1.35)	-132.82(1.32)	-133.12(0.39)	-134.36(0.82)
0.4	-113.93(0.19)	-115.53(0.34)	-116.00(0.68)	-115.94(0.57)	-115.91(0.45)	-115.47(0.38)
0.3	-99.14(0.54)	-101.48(1.82)	-100.76(2.18)	-100.30(1.34)	-101.00(0.16)	-99.49(0.07)
0.2	-84.39(3.46)	-84.96(1.89)	-84.43(2.42)	-84.84(1.04)	-86.72(1.40)	-86.64(1.50)
0.1	-69.28(1.85)	-72.81(2.65)	-70.70(2.62)	-73.54(1.15)	-72.35(1.07)	-72.88(0.13)
0.0	-59.74(1.11)	- 61.41(0.10)	-60.75(0.04)	-60.59(0.83)	-61.59(0.02)	-61.35(0.09)
ΔG	**150.40(1.23)**	**153.07(1.15)**	**152.28(1.53)**	**153.45(1.07)**	**152.56(0.53)**	**152.08(0.71)**

Ca^{2+} → Mg^{2+} in pre-hydrolytic group II intron

λ	Run 1	Run 2	Run 3	Run 4	Run 5
1.0	63.41 (1.07)	63.28 (1.03)	62.95 (0.58)	58.58 (0.58)	59.05 (0.41)
0.9	70.90 (0.07)	72.20 (1.50)	72.36 (0.64)	71.25 (0.19)	66.95 (0.13)
0.8	82.28 (0.15)	81.38 (0.64)	82.64 (0.15)	80.44 (1.26)	86.45 (0.47)
0.7	92.34 (1.79)	94.45 (0.98)	94.14 (0.75)	93.30 (0.90)	99.55 (0.06)
0.6	106.69 (1.25)	113.91 (1.97)	108.08 (0.28)	105.91 (0.37)	115.85 (0.44)
0.5	129.99 (0.60)	130.41 (0.52)	124.51 (0.27)	121.39 (0.56)	130.69 (1.00)
0.4	150.83 (0.98)	152.28 (0.25)	143.75 (0.51)	143.67 (0.21)	153.05 (0.70)
0.3	179.38 (2.65)	181.17 (1.31)	168.73 (1.31)	169.14 (2.51)	185.08 (1.31)
0.2	210.48 (2.03)	220.17 (2.22)	213.44 (0.91)	211.92 (0.20)	222.41 (2.46)
0.1	271.95 (2.88)	265.77 (1.33)	272.63 (2.96)	267.37 (0.02)	273.85 (0.04)
0.0	332.86 (0.39)	361.01 (2.06)	355.15 (2.20)	353.57 (0.21)	355.09 (1.14)
ΔG	**-149.30 (1.26)**	**-152.39 (1.25)**	**-148.93 (0.96)**	**-147.04 (0.64)**	**-154.10 (0.74)**

Mg^{2+} → Ca^{2+} in post-hydrolytic group II intron

λ	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6
1.0	-332.98(1.18)	-330.67(0.11)	-330.63(2.68)	-327.40(3.23)	-331.10(0.96)	-330.22(1.08)
0.9	-256.39(0.20)	-256.87(1.78)	-253.85(0.15)	-255.90(3.52)	-256.89(0.05)	-254.07(2.45)
0.8	-210.14(0.09)	-210.79(1.86)	-209.61(2.78)	-211.27(0.85)	-210.33(0.45)	-207.58(1.25)
0.7	-177.11(2.03)	-176.37(1.07)	-175.60(0.06)	-177.55(1.95)	-175.62(1.28)	-176.34(1.34)
0.6	-152.64(2.02)	-153.60(0.03)	-149.11(0.23)	-150.66(0.26)	-148.17(1.59)	-148.42(0.31)
0.5	-130.54(0.00)	-131.53(2.12)	-132.42(0.06)	-128.12(0.89)	-128.41(1.88)	-132.30(2.18)
0.4	-113.59(0.67)	-114.16(1.30)	-113.72(0.71)	-113.24(0.07)	-109.67(4.55)	-109.71(2.10)
0.3	-97.62(0.57)	-97.42(3.80)	-98.29(0.89)	-101.10(0.63)	-94.99(0.46)	-94.60(3.00)
0.2	-83.86(1.69)	-83.43(0.29)	-81.45(0.76)	-84.40(0.29)	-78.59(0.02)	-79.65(0.46)
0.1	-71.38(0.84)	-65.66(1.78)	-63.46(3.16)	-64.48(3.49)	-58.44(1.02)	-66.26(2.90)
0.0	-61.31(1.35)	-54.26(1.44)	-52.50(0.72)	-53.94(1.56)	-53.35(0.26)	-60.56(1.03)
ΔG	**149.04(0.97)**	**148.23(1.42)**	**146.91(1.11)**	**147.74(1.52)**	**145.33(1.14)**	**146.43(1.64)**

Table 3.2. Free energy calculations. (a) Calculated free energetics (in kcal/mol) and simulation time scale (in ns) for the alchemical transformation of $Mg^{2+} \rightarrow Ca^{2+}$ in free and intron bound conformations.

System	Alchemical transformation	Replica	Equilibration Time (ns)	MD Free energy calculations Time (ns)	Estimated ΔG (kcal/mol)
(A) Free in water	$Mg^{+2} \rightarrow Ca^{+2}$	Run 1	2.28	22	69.05 (0.64)
		Run 2	2.28	22	68.90 (0.60)
		Run 3	2.32	22	68.83 (0.72)
		Run 4	2.32	22	68.64 (0.50)
		Run 5	2.32	22	68.69 (1.04)
		Run 6	2.38	22	68.81 (0.47)
					Average = 68.82 ±0.13
(B) Ligand-free intron state	$Mg^{+2} \rightarrow Ca^{+2}$	Run 1	2.28	22	150.40 (0.47)
		Run 2	2.28	22	153.07 (0.30)
		Run 3	2.32	22	152.28 (1.12)
		Run 4	2.32	22	153.45 (0.14)
		Run 5	2.32	22	152.56 (0.14)
		Run 6	2.38	22	152.08 (0.46)
					Average = 152.30 ±0.97
(C) Pre-hydrolytic intron	$Ca^{+2} \rightarrow Mg^{+2}$	Run 1	0.28	22	-149.30 (1.58)
		Run 2	0.28	22	-152.39 (0.26)
		Run 3	0.32	22	-148.93 (0.84)
		Run 4	0.38	22	-154.10 (0.23)
		Run 5	0.32	22	-147.04 (0.38)
					Average = - 150.34 ±2.53
(D) Post hydrolytic -state	$Mg^{+2} \rightarrow Ca^{+2}$	Run 1	0.28	22	149.04 (0.56)
		Run 2	0.28	22	148.23 (0.55)
		Run 3	0.32	22	146.91 (0.29)
		Run 4	0.32	22	147.74 (0.92)
		Run 5	0.32	22	145.33 (0.65)
		Run 6	0.38	22	146.43 (0.97)
					Average = 147.28 ±1.21

Table 3.3. Calculated Mg^{2+} vs Ca^{2+} relative binding free energy (**ΔΔG** in kcal/mol) to group II intron active sites M1, M2.

State (Alchemical path)	**ΔG_{Comp}**	**ΔG_{free}**	**ΔΔG**
Pre-hydrolytic state (2 $Mg^{2+} \rightarrow$ 2 Ca^{2+})	150.34 ±2.53	137.64 ±0.13	**11.91 (2.49)**
Post-hydrolytic state (2 $Mg^{2+} \rightarrow$ 2 Ca^{2+})	147.28 ±1.21		**9.64 (3.01)**
Ligand –free state (2 $Mg^{2+} \rightarrow$ 2 Ca^{2+})	152.30 ±0.97		**14.66 (2.40)**

Table 3.4. Quantum chemical single point energies calculated for free and intron bound divalent ions embedded into a water dielectric.

		Energy (Au)				**E4+E3-(E1+E2)*627.5095 = ΔE (in kcal/mol)**
System	**Theory**	**(E1) [Mg^{+2}, Mg^{+2}-Wat]**	**(E2) [Ca^{+2},Ca^{+2}-intron]**	**(E3) [Ca^{+2}, Ca^{+2}-Wat]**	**(E4) [Mg^{+2}, Mg^{+2}-intron]**	
Pre	b3lyp/631+g*	-399.6619832	-7823.4542225	-1354.9648355	-6868.3557733	-128.26
	M062X/6-311++g**	-399.624579	-7822.9552468	-1355.0242504	-6867.7823557	-142.31
Post	b3lyp/631+g*	-399.6619832	-6795.8063715	-1354.9648355	-5840.71433	-132.28
	M062X/6-311++g**	-399.624579	-6795.4221387	-1355.0242504	-5840.1914875	-106.06
Free	b3lyp/631+g*	-399.6619832	-6644.376168	-1354.9648355	-5689.3103142	-148.72
	M062X/6-311++g**	-399.624579	-6644.082263	-1355.0242504	-5688.8908034	-130.65

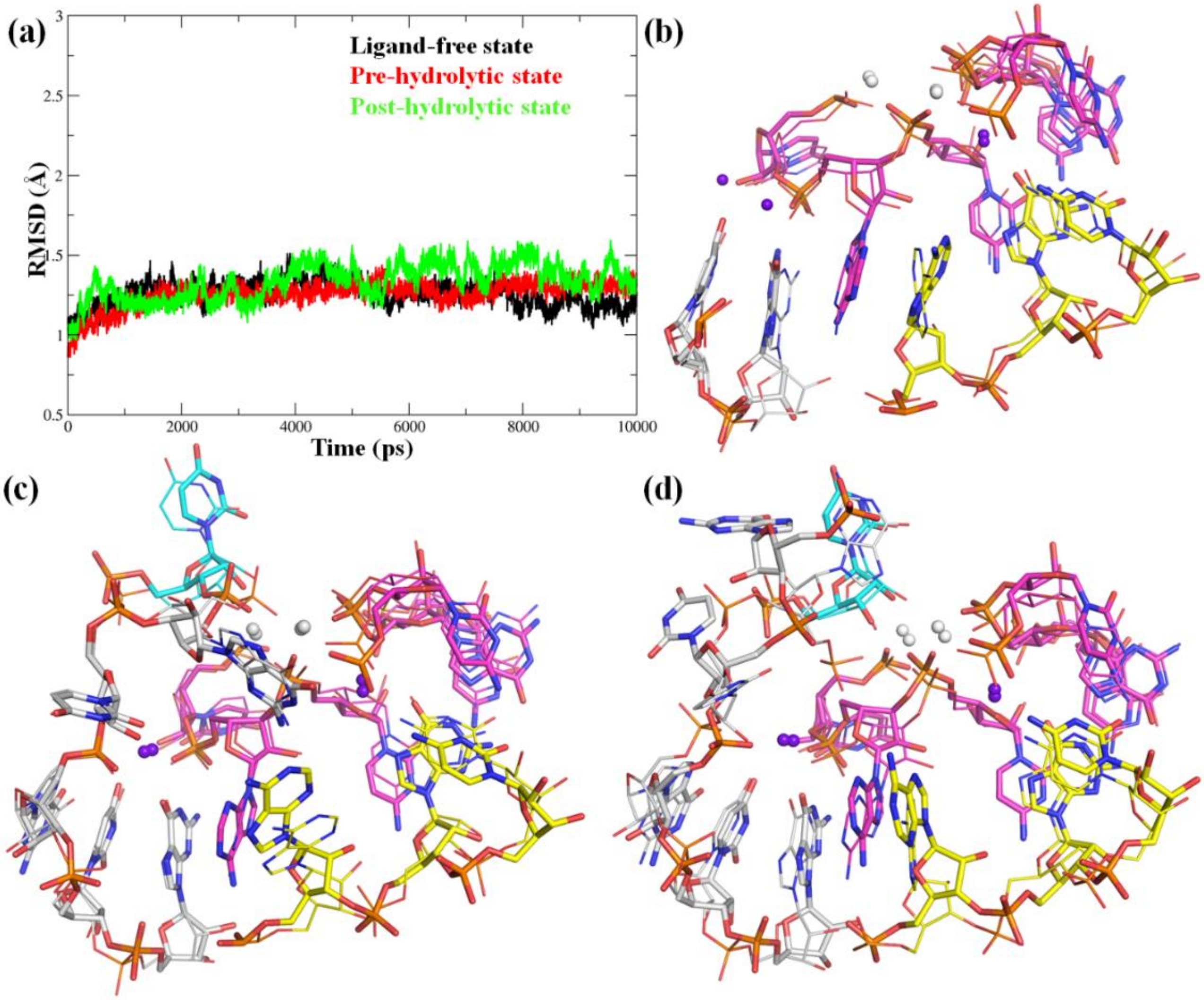


Figure 3.1. (a) The convergence of MD simulations, RMSD vs Time from 10ns trajectory. Structural comparison around M1, M2 sites of group II intron: X-ray (line) and MD structure (sticks), (b) ligand-free state (c) pre-hydrolytic state (d) post-hydrolytic state.

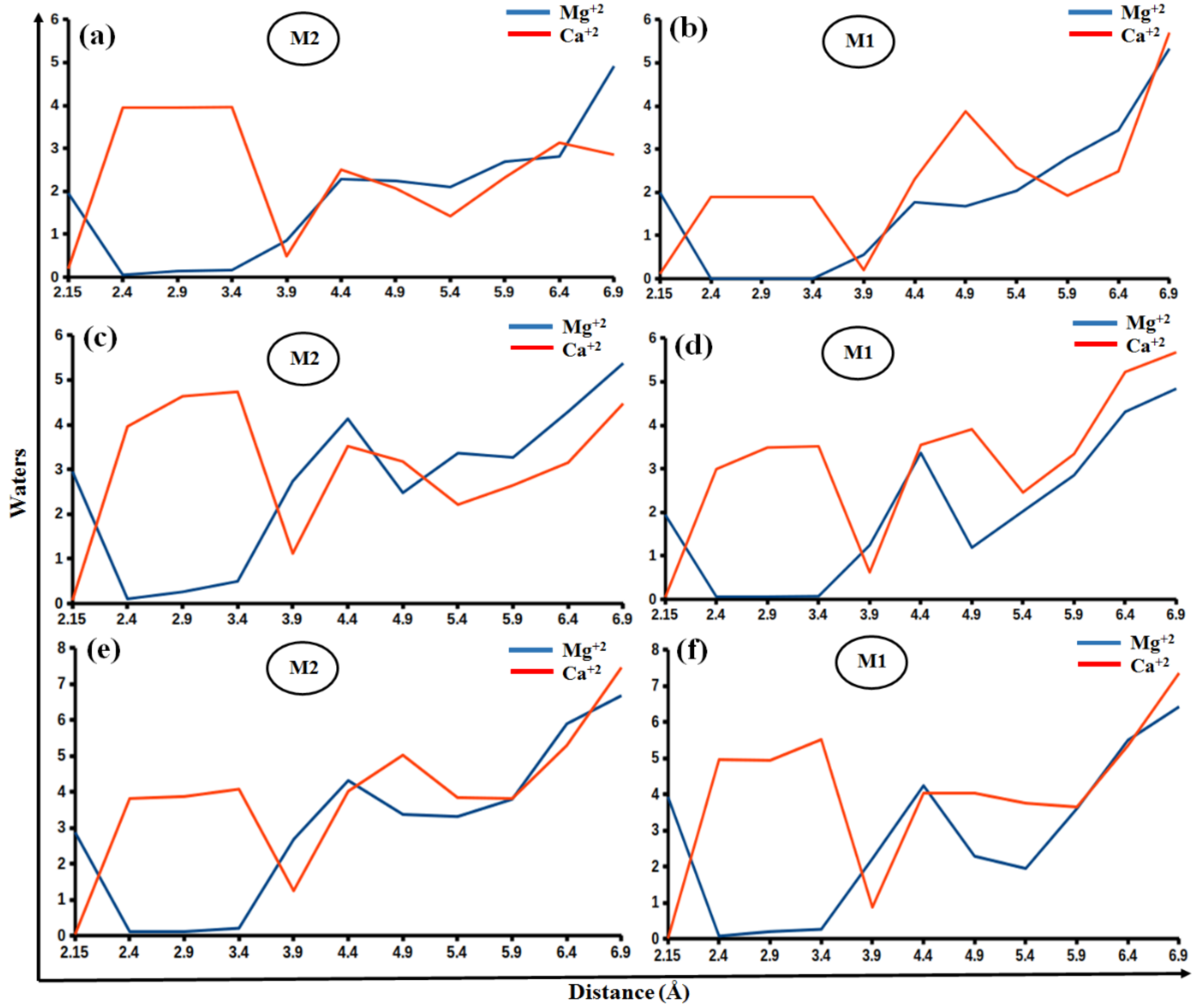


Figure 3.2. Number of waters around M2 (left) and M1 (right) sites. (a,b) Pre-hydrolytic state (c,d) post-hydrolytic state and (e,f) ligand-free state

Appendix 4

Table 4.1. Estimated free energetics (in Kcal/mol) and simulation time scale (in *ns*) for the alchemical transformation of $Mg^{2+} \rightarrow Ca^{2+}$, and $K^{+} \rightarrow Na^{+}$ in water (Free) and in complex with lariat group II intron at pre-hydrolytic and post-hydrolytic state.

Alchemical transformation	System	Replica	MD Free energy calculations Time (ns)	Estimated ΔG (Kcal/mol)
Mg^{2+} → Ca^{2+}	Free in water	Replica 1	25	68.90 (0.41)
		Replica 2	25	68.83 (0.56)
		Replica 3	25	68.64 (0.62)
		Replica 4	25	68.69 (0.67)
		Replica 5	25	68.81 (0.49)
			Average	**68.77 ±0.11**
	Pre-hydrolytic state	Replica 1	25	150.56 (1.57)
		Replica 2	25	150.38 (1.52)
		Replica 3	25	150.74 (1.67)
		Replica 4	25	150.22 (2.23)
		Replica 5	25	151.68 (1.10)
		Replica 6	25	151.52 (0.79)
			Average	**150.85 (0.61)**
	Post-hydrolytic state	Replica 1	25	156.45 (1.78)
		Replica 2	25	157.69 (1.40)
		Replica 3	25	155.18 (2.56)
		Replica 4	25	158.53 (1.57)
		Replica 5	25	157.91 (1.41)
		Replica 6	25	157.88 (1.56)
			Average	**157.27 (1.23)**
K^{+} → Na^{+}	Free in water	Replica 1	25	-23.47 (0.41)
		Replica 2	25	-23.45 (0.35)
		Replica 3	25	-23.36 (0.56)
		Replica 4	25	-23.28 (0.28)
		Replica 5	25	-23.51 (0.37)
			Average	**-23.41 (0.09)**
	Pre-hydrolytic state	Replica 1	25	-22.53 (0.72)
		Replica 2	25	-20.09 (0.41)
		Replica 3	25	-21.57 (0.55)
		Replica 4	25	-20.61 (1.56)
		Replica 5	25	-21.49 (0.38)
		Replica 6	25	-22.40 (0.77)
			Average	**-21.45 (0.96)**
	Post-hydrolytic state	Replica 1	25	-23.56 (0.98)
		Replica 2	25	-22.84 (0.94)
		Replica 3	25	-22.90 (1.43)
		Replica 4	25	-25.77 (1.08)
		Replica 5	25	-25.37 (1.15)
		Replica 6	25	-24.35 (0.56)
			Average	**-24.13 (1.25)**

Table 4.2. Alchemical coordinate λ was defined that connects end states i.e, λ =1 corresponds to Mg^{2+}/K^{+} and λ =0 corresponds to Ca^{2+}/Na^{+}. Free energy derivative ∂G/∂λ was calculated for alchemical transformation $Mg^{2+} \rightarrow Ca^{2+}/K^{+} \rightarrow Na^{+}$ for 11λ points between 1 and 0. Free energy derivative was calculated as $\partial G/\partial\lambda = \langle U(\lambda=1) - U(\lambda=1)\rangle_{\lambda}$, where < > represents averaging over MD trajectory at a specific value of λ. Alchemical transformation associated free energies (in Kcal/mol) was calculated by numerically integrating ∂G/∂λ vs λ plot. The MD trajectories were divided into two equal halves and the difference between the computed ∂G/∂λ's from the two halves is reported as uncertainty in the parenthesis at each λ point. Derivative at each λ is given below:

$Mg^{2+} \rightarrow Ca^{2+}$ free in water

λ	Replica 1	Replica 2	Replica 3	Replica 4	Replica 5
1.00	-154.32 (0.71)	-155.68 (0.08)	-154.70 (0.53)	-154.85 (0.80)	-154.39 (1.78)
0.90	-122.17 (0.07)	-121.06 (1.49)	-121.23 (0.73)	-121.03 (1.34)	-121.43 (0.32)
0.80	-100.19 (1.13)	-99.68 (0.81)	-100.28 (0.08)	-99.83 (0.92)	-100.40 (0.10)
0.70	-84.15 (0.54)	-84.29 (0.36)	-83.57 (0.01)	-84.65 (0.40)	-84.29 (0.34)
0.60	-71.84 (0.67)	-72.20 (0.27)	-71.67 (0.01)	-72.30 (1.51)	-71.62 (0.46)
0.50	-62.06 (0.40)	-61.58 (0.72)	-61.76 (0.14)	-61.75 (0.58)	-61.55 (0.90)
0.40	-53.75 (0.95)	-53.00 (0.62)	-52.40 (2.23)	-52.69 (1.68)	-53.13 (0.08)
0.30	-44.51 (1.38)	-45.29 (0.74)	-45.04 (0.72)	-44.16 (0.90)	-45.47 (0.20)
0.20	-34.47 (0.15)	-34.13 (1.97)	-34.11 (0.56)	-34.65 (2.56)	-34.70 (0.35)
0.10	-26.93 (0.24)	-27.41 (0.26)	-27.44 (0.44)	-26.79 (0.33)	-26.77 (0.29)
0.00	-23.55 (0.43)	-23.64 (0.63)	-23.17 (0.40)	-23.30 (0.50)	-23.09 (0.36)
ΔG	**68.90 (0.41)**	**68.83 (0.56)**	**68.64 (0.62)**	**68.69 (0.67)**	**68.81 (0.49)**

$Mg^{2+} \rightarrow Ca^{2+}$ pre-hydrolytic state

λ	Replica 1	Replica 2	Replica 3	Replica 4	Replica 5	Replica 6
1.00	-335.63 (3.37)	-335.68 (5.39)	-334.78 (3.67)	-338.06 (7.88)	-340.30 (0.38)	-352.04 (0.76)
0.90	-257.91 (3.01)	-257.09 (2.04)	-261.65 (2.03)	-256.23 (3.94)	-261.06 (0.39)	-260.39 (2.40)
0.80	-210.25 (0.62)	-212.41 (2.59)	-215.20 (0.78)	-210.22 (0.71)	-214.50 (0.29)	-213.68 (1.72)
0.70	-174.19 (0.49)	-180.32 (0.51)	-179.85 (2.47)	-174.01 (0.28)	-181.21 (1.66)	-178.65 (0.60)
0.60	-150.06 (1.75)	-150.73 (0.96)	-149.41 (5.45)	-150.70 (0.67)	-151.32 (2.33)	-153.47 (0.12)
0.50	-129.55 (1.46)	-130.64 (0.65)	-127.41 (1.80)	-129.31 (0.86)	-132.07 (0.89)	-132.65 (1.23)
0.40	-115.88 (0.26)	-114.35 (0.73)	-109.92 (0.20)	-110.80 (1.10)	-113.77 (0.00)	-112.11 (0.31)
0.30	-101.74 (0.39)	-93.27 (2.35)	-99.20 (0.53)	-100.43 (1.29)	-100.31 (0.86)	-96.80 (0.07)
0.20	-88.25 (0.25)	-86.26 (0.09)	-86.93 (0.50)	-89.29 (0.42)	-86.81 (1.48)	-84.47 (0.36)
0.10	-76.86 (1.60)	-77.43 (0.31)	-76.60 (0.61)	-79.08 (1.14)	-75.45 (1.33)	-74.16 (0.49)
0.00	-66.20 (4.13)	-66.91 (1.14)	-67.54 (0.36)	-66.21 (6.22)	-60.13 (2.50)	-65.60 (0.68)
ΔG	**150.56 (1.57)**	**150.38 (1.52)**	**150.74 (1.67)**	**150.22 (2.23)**	**151.68 (1.10)**	**151.52 (0.79)**

$Mg^{2+} \rightarrow Ca^{2+}$ post-hydrolytic state

λ	Replica 1	Replica 2	Replica 3	Replica 4	Replica 5	Replica 6
1.00	-367.73 (2.92)	-367.97 (4.29)	-375.10 (2.43)	-373.72 (3.58)	-365.99 (2.08)	-355.53 (4.87)
0.90	-277.77 (1.59)	-278.79 (0.86)	-284.41 (2.30)	-279.84 (0.08)	-277.56 (0.46)	-275.85 (0.77)
0.80	-224.61 (4.11)	-224.59 (1.26)	-229.78 (2.72)	-226.46 (0.28)	-226.29 (2.52)	-223.92 (1.85)
0.70	-187.08 (1.76)	-188.54 (2.11)	-173.30 (4.17)	-189.90 (0.47)	-188.32 (0.11)	-185.83 (3.30)
0.60	-159.54 (2.00)	-160.21 (0.63)	-146.71 (1.31)	-159.40 (1.92)	-160.00 (1.31)	-158.89 (0.89)
0.50	-135.76 (2.11)	-136.91 (0.45)	-130.55 (2.41)	-135.63 (0.91)	-138.28 (0.04)	-137.04 (1.37)
0.40	-115.40 (0.47)	-117.41 (0.59)	-113.34 (4.45)	-118.19 (1.10)	-117.55 (0.15)	-116.73 (1.30)
0.30	-100.35 (2.27)	-98.72 (3.56)	-94.73 (2.77)	-97.85 (5.29)	-103.04 (1.42)	-102.22 (0.91)
0.20	-79.69 (0.43)	-83.83 (0.69)	-85.22 (4.70)	-84.10 (0.47)	-80.14 (1.84)	-89.60 (0.84)
0.10	-70.82 (1.05)	-72.03 (0.16)	-74.38 (0.83)	-74.02 (1.08)	-74.78 (2.52)	-77.41 (0.68)
0.00	-59.20 (0.92)	-63.65 (0.77)	-63.60 (0.10)	-65.94 (2.05)	-60.13 (3.11)	-67.12 (0.38)
ΔG	**156.45 (1.78)**	**157.69 (1.40)**	**155.18 (2.56)**	**158.53 (1.57)**	**157.91 (1.41)**	**157.88 (1.56)**

$K^{+} \rightarrow Na^{+}$ free in water

λ	Replica 1	Replica 2	Replica 3	Replica 4	Replica 5
1.00	07.25 (0.06)	07.12 (0.20)	07.28 (0.02)	07.28 (0.27)	07.20 (0.12)
0.90	08.97 (0.22)	08.95 (0.06)	08.78 (0.19)	08.91 (0.02)	08.98 (0.23)
0.80	10.52 (0.00)	11.05 (0.22)	10.95 (0.07)	10.80 (0.36)	11.05 (0.30)
0.70	13.23 (0.11)	13.60 (0.31)	13.33 (0.33)	13.63 (0.00)	13.17 (0.70)
0.60	16.49 (0.11)	16.41 (0.31)	16.46 (0.71)	15.99 (0.38)	16.12 (0.21)
0.50	20.00 (0.29)	20.23 (0.90)	20.01 (0.18)	19.89 (0.02)	20.25 (0.22)
0.40	24.41 (0.32)	24.24 (0.19)	24.19 (1.02)	24.12 (0.70)	24.52 (0.24)
0.30	29.87 (1.23)	29.76 (0.00)	29.17 (0.00)	30.11 (0.56)	30.40 (0.62)
0.20	36.35 (0.28)	35.90 (0.10)	36.24 (0.38)	35.08 (0.28)	36.34 (0.90)
0.10	44.38 (1.72)	44.27 (0.88)	43.91 (1.35)	43.88 (0.52)	43.88 (0.52)
0.00	53.73 (0.24)	53.01 (0.75)	53.92 (2.01)	53.56 (0.02)	53.59 (0.02)
ΔG	**-23.47 (0.42)**	**-23.45 (0.36)**	**-23.36 (0.57)**	**-23.28 (0.28)**	**-23.51 (0.37)**

$K^+ \rightarrow Na^+$ pre-hydrolytic state

λ	Replica 1	Replica 2	Replica 3	Replica 4	Replica 5	Replica 6
1.00	6.27 (0.07)	6.75 (0.19)	6.67 (1.12)	4.47 (0.28)	6.38 (0.06)	6.17 (0.18)
0.90	7.39 (0.30)	8.11 (0.31)	7.11 (0.26)	6.08 (1.80)	7.38 (0.61)	7.71 (0.38)
0.80	11.02 (0.60)	10.12 (0.26)	8.69 (0.78)	7.00 (1.27)	11.59 (0.17)	10.42 (0.22)
0.70	13.31 (0.45)	14.51 (0.45)	10.75 (0.01)	8.51 (1.29)	13.03 (0.33)	12.32 (0.63)
0.60	16.87 (0.21)	16.33 (0.29)	15.01 (0.01)	9.55 (0.51)	14.36 (0.02)	14.96 (0.10)
0.50	21.51 (0.22)	18.27 (0.74)	19.71 (0.30)	22.38 (0.88)	14.84 (0.03)	18.95 (0.20)
0.40	23.27 (0.06)	21.43 (0.04)	22.35 (0.64)	26.24 (0.80)	22.96 (0.94)	26.38 (0.72)
0.30	27.62 (2.63)	24.49 (0.14)	27.57 (0.03)	28.47 (1.74)	27.91 (0.74)	30.06 (1.20)
0.20	35.50 (0.88)	29.81 (0.85)	34.31 (0.71)	30.67 (2.27)	34.99 (0.96)	37.32 (1.38)
0.10	41.11 (0.50)	33.07 (0.15)	41.65 (0.03)	39.90 (0.44)	39.41 (0.17)	37.97 (1.81)
0.00	48.99 (2.03)	42.81 (1.11)	50.26 (2.18)	50.02 (5.90)	50.32 (0.13)	49.50 (1.70)
ΔG	**-22.53 (0.72)**	**-20.09 (0.41)**	**-21.57 (0.55)**	**-20.61 (1.56)**	**-21.49 (0.38)**	**-22.40 (0.77)**

$K^+ \rightarrow Na^+$ post-hydrolytic state

λ	Replica 1	Replica 2	Replica 3	Replica 4	Replica 5	Replica 6
1.00	7.88 (0.18)	7.07 (0.26)	7.58 (0.33)	7.46 (0.37)	6.33 (0.30)	7.38 (0.17)
0.90	9.15 (0.76)	8.43 (0.68)	8.00 (0.42)	9.90 (0.99)	8.05 (0.69)	9.01 (0.79)
0.80	11.05 (1.37)	10.75 (0.57)	10.29 (0.02)	11.75(0.14)	10.25 (0.68)	11.50 (0.29)
0.70	13.87 (2.68)	12.67 (0.10)	12.71 (0.70)	14.72 (1.62)	14.54 (1.16)	13.93 (0.03)
0.60	17.70 (0.97)	16.23 (0.31)	16.00 (1.00)	18.38 (0.91)	16.09 (0.22)	17.18 (0.78)
0.50	21.06 (1.59)	18.63 (1.95)	19.36 (2.52)	22.13 (0.54)	25.92 (0.38)	19.89 (0.21)
0.40	25.28 (0.06)	24.49 (0.86)	25.24 (0.60)	24.69 (1.52)	27.89 (2.82)	24.83 (0.46)
0.30	28.87 (0.94)	29.05 (0.35)	28.34 (3.21)	30.98 (0.18)	34.26 (1.74)	29.03 (0.39)
0.20	36.37 (0.26)	35.28 (1.44)	36.21 (0.32)	41.10 (0.81)	34.32 (0.04)	37.38 (1.12)
0.10	42.87 (0.80)	42.34 (0.86)	42.53 (1.07)	50.41 (1.86)	49.84 (3.13)	48.47 (0.76)
0.00	50.81 (0.37)	54.01 (2.99)	52.94 (5.54)	59.71 (2.97)	58.60 (1.51)	57.03 (1.16)
ΔG	**-23.56 (0.98)**	**-22.84 (0.94)**	**-22.90 (1.43)**	**-25.77 (1.08)**	**-25.37 (1.15)**	**-24.35 (0.56)**

Table 4.3. Calculated Mg^{2+} vs. Ca^{2+} and K^{+} vs. Na^{+} relative binding free energy (**ΔΔG** in Kcal/mol) to the lariat group II intron active site.

Alchemical path	**State**	**ΔG_{Comp}**	**ΔG_{free}**	**ΔΔG (kcal/mol)**
$Mg^{2+} \rightarrow Ca^{2+}$	Pre-hydrolytic state	150.85 ± 0.61	137.64 ± 0.13	13.21 ± 0.62
	Post-hydrolytic state	157.57 ± 1.23		19.93 ± 1.24
$K^{+} \rightarrow Na^{+}$	Pre-hydrolytic state	-21.45 ± 0.96	- 23.41 ± 0.09	-1.96 ± 0.96
	Post-hydrolytic state	-24.13 ± 1.25		0.72 ± 1.25

Table 4.4. Single point energies estimated from quantum chemical calculations.

System	**Theory**	**Energy (in Hartree)**				**E4+E3-(E1+E2) *627.5095 = ΔE (in kcal/mol)**
		(E1) [Mg^{2+}-Wat]	**(E2) E2:[Ca^{2+}-intron]**	**(E3) E3:[Ca^{2+}-Wat]**	**(E4) [Mg^{2+}-intron]**	
Pre-hydrolytic state	B3LYP/6-31+G*	-934.883365	-7028.0286755	-1890.0150122	-6073.0453898	-93.10
	MO62X/6-31++G**	-934.8501293	-7027.799285	-1890.1025138	-6072.6601375	-71.06
Post-hydrolytic state	B3LYP/6-31+G*	-858.4338742	-9006.9311672	-1813.5575499	-8051.9359819	-80.63
	MO62X/6-31++G**	-858.3921152	-9006.6318331	-1813.6450037	-8051.4867215	-67.63

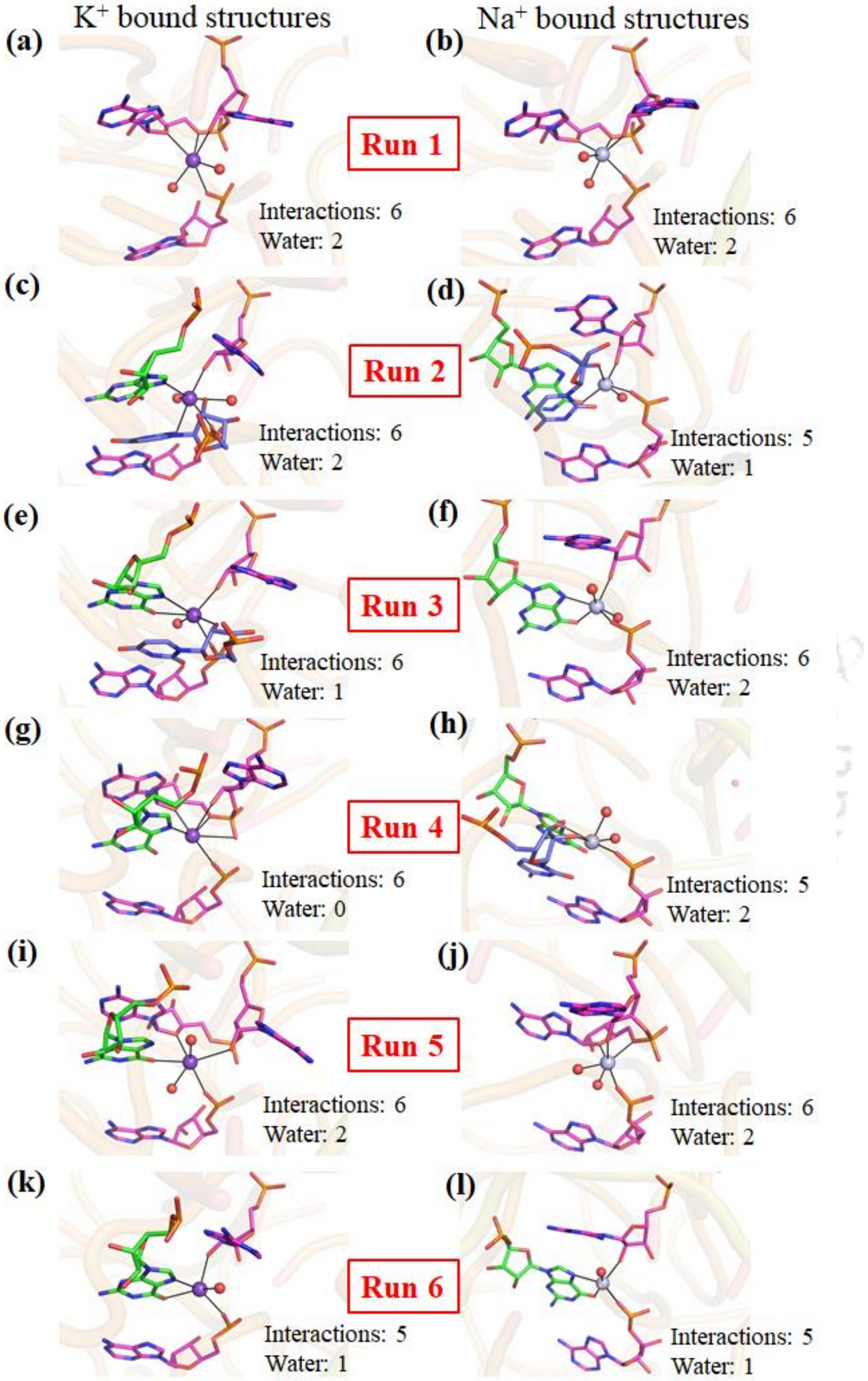


Figure 4.1. MD insight of multiple replicas in the pre-hydrolytic (2s) state (a,c,e,g,i and k) K^+ bound structures (b,d,f,h,j, and l) Na^+ bound structures.

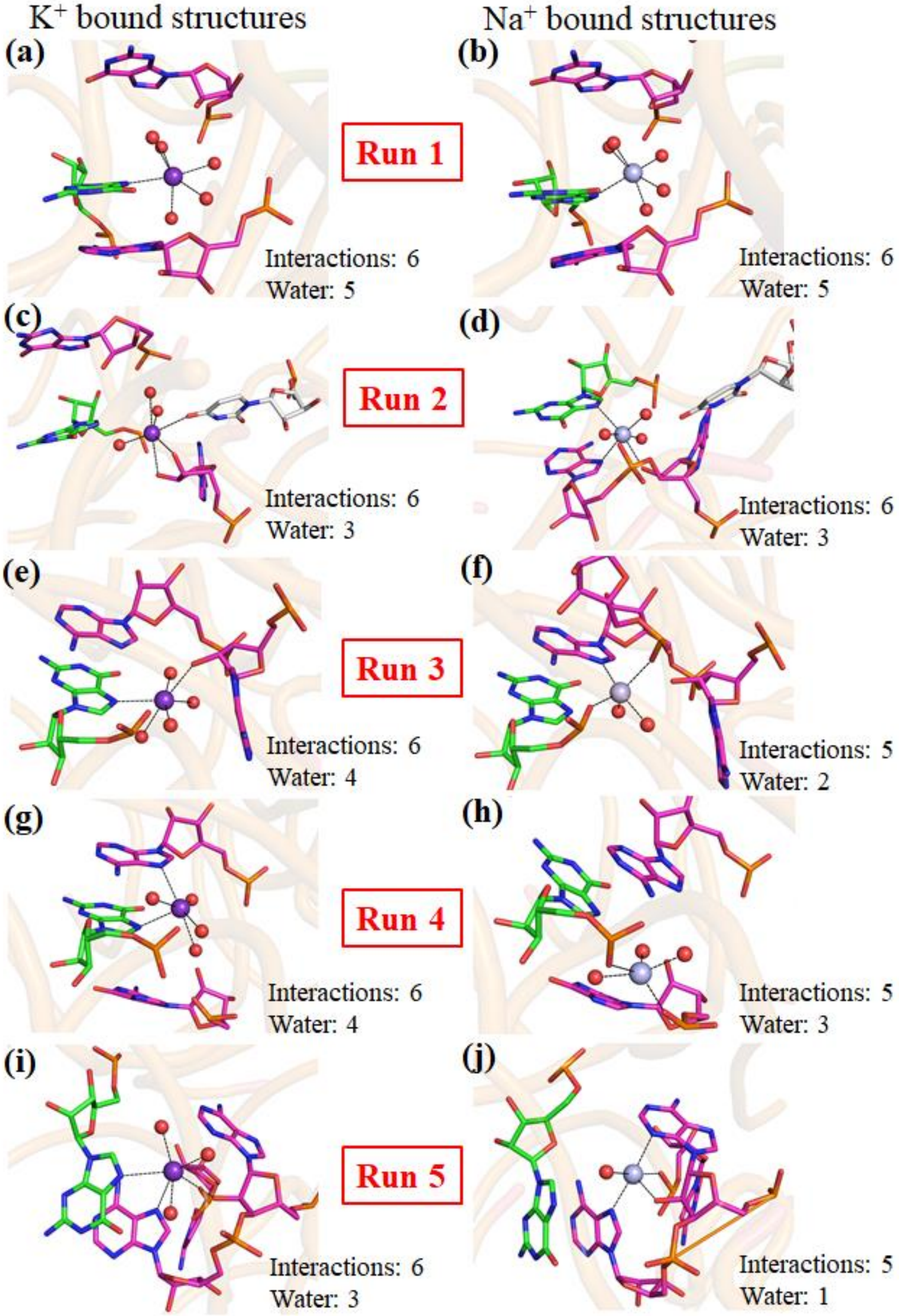


Figure 4.2. MD insight of multiple replicas in the post-hydrolytic state (a,c,e,g, and i) K^+ bound structures (b,d,f,h, and j) Na^+ bound structures.

Appendix 5

Table 5.1. Free energy calculations. (a) Calculated free energetics (in kcal/mol) and simulation time scale (in *ns*) for the alchemical transformation of $Mg^{2+} \rightarrow Ca^{2+}$ in free and Cas1 bound conformations.

System	**Alchemical transformation**	**Replica**	**Equilibration Time (ns)**	**MD Free energy calculations Time (ns)**	**Estimated ΔG (kcal/mol)**
Free	$Mg^{2+} \rightarrow Ca^{2+}$	Run 1	2.28	22	68.90 (0.41)
		Run 2	2.28	22	68.83 (0.56)
		Run 3	2.32	22	68.64 (0.62)
		Run 4	2.32	22	68.69 (0.67)
		Run 5	2.32	22	68.81 (0.49)
					Average = 68.77 (0.11)
. Wild type	$Mg^{2+} \rightarrow Ca^{2+}$	Run 1	2.28	22	73.67 (1.08)
		Run 2	2.28	22	73.97 (1.54)
		Run 3	2.32	22	73.54 (0.56)
		Run 4	2.32	22	73.22 (0.94)
		Run 5	2.32	22	74.55 (0.84)
					Average = 73.79 (0.50)
E190A	$Mg^{2+} \rightarrow Ca^{2+}$	Run 1	2.28	22	72.37 (1.30)
		Run 2	2.28	22	73.20 (0.88)
		Run 3	2.32	22	72.98 (0.95)
		Run 4	2.32	22	72.81 (0.58)
		Run 5	2.32	22	73.57 (0.78)
					Average = 72.99 (0.45)
H254A	$Mg^{2+} \rightarrow Ca^{2+}$	Run 1	0.28	22	73.36 (0.62)
		Run 2	0.28	22	74.37 (0.99)
		Run 3	0.32	22	73.76 (0.92)
		Run 4	0.38	22	73.56 (0.86)
		Run 5	0.32	22	74.09 (0.55)
					Average = 73.83 (0.40)
D265A	$Mg^{2+} \rightarrow Ca^{2+}$	Run 1	0.28	22	73.26 (0.64)
		Run 2	0.28	22	73.51 (0.98)
		Run 3	0.32	22	73.75 (0.95)
		Run 4	0.32	22	73.37 (1.23)
		Run 5	0.32	22	73.75 (1.10)
					Average = 73.53 (0.22)

D268A	$Mg^{2+}\rightarrow Ca^{2}$	Run 1	0.28	22	72.50 (1.05)
		Run 2	0.28	22	73.11 (2.05)
		Run 3	0.32	22	72.45 (1.45)
		Run 4	0.32	22	71.88 (0.94)
		Run 5	0.32	22	73.74 (0.74)
					Average = 72.74 (0.71)

Table 5.2. Alchemical coordinate λ was defined and connects end states i.e, λ =1 corresponds to Mg^{2+} and λ =0 corresponds to Ca^{2+}. Free energy derivative $\partial G/\partial\lambda$ was calculated for alchemical transformation $Mg^{2+} \rightarrow Ca^{2+}$ for 11λ points between 1 and 0. Free energy derivative was calculated as $\partial G/\partial\lambda = \langle U(\lambda=1) - U(\lambda=1)\rangle_{\lambda}$, where < > represents averaging over MD trajectory at a specific value of λ. Alchemical transformation associated free energies (in kcal/mol) was calculated by numerically integrating $\partial G/\partial\lambda$ vs λ plot. The MD trajectories were divided into two equal halves and the difference between the computed $\partial G/\partial\lambda$'s from the two halves is reported as uncertainty in the parenthesis at each λ point. Derivative at each λ is given below:

$Mg^{2+}\rightarrow Ca^{2+}$ free in water

λ	Run 1	Run 2	Run 3	Run 4	Run 5
1.00	-154.32 (0.71)	-155.68 (0.08)	-154.70 (0.53)	-154.85 (0.80)	-154.39 (1.78)
0.90	-122.17 (0.07)	-121.06 (1.49)	-121.23 (0.73)	-121.03 (1.34)	-121.43 (0.32)
0.80	-100.19 (1.13)	-99.68 (0.81)	-100.28 (0.08)	-99.83 (0.92)	-100.40 (0.10)
0.70	-84.15 (0.54)	-84.29 (0.36)	-83.57 (0.01)	-84.65 (0.40)	-84.29 (0.34)
0.60	-71.84 (0.67)	-72.20 (0.27)	-71.67 (0.01)	-72.30 (1.51)	-71.62 (0.46)
0.50	-62.06 (0.40)	-61.58 (0.72)	-61.76 (0.14)	-61.75 (0.58)	-61.55 (0.90)
0.40	-53.75 (0.95)	-53.00 (0.62)	-52.40 (2.23)	-52.69 (1.68)	-53.13 (0.08)
0.30	-44.51 (1.38)	-45.29 (0.74)	-45.04 (0.72)	-44.16 (0.90)	-45.47 (0.20)
0.20	-34.47 (0.15)	-34.13 (1.97)	-34.11 (0.56)	-34.65 (2.56)	-34.70 (0.35)
0.10	-26.93 (0.24)	-27.41 (0.26)	-27.44 (0.44)	-26.79 (0.33)	-26.77 (0.29)
0.00	-23.55 (0.43)	-23.64 (0.63)	-23.17 (0.40)	-23.30 (0.50)	-23.09 (0.36)
ΔG	**68.90 (0.41)**	**68.83 (0.56)**	**68.64 (0.62)**	**68.69 (0.67)**	**68.81 (0.49)**

WT

λ	Run 1	Run 2	Run 3	Run 4	Run 5
1.00	-181.04(2.79)	-177.85(3.28)	-180.21(0.67)	-178.56(0.12)	-177.89(0.89)
0.90	-135.02(1.60)	-135.37(1.15)	-136.79(0.37)	-134.86(0.55)	-137.79(2.32)
0.80	-108.67(0.26)	-110.07(0.08)	-110.24(1.35)	-108.61(0.29)	-110.62(0.35)
0.70	-93.02(0.16)	-91.31(1.82)	-92.87(0.91)	-91.35(0.21)	-91.32(1.10)
0.60	-78.09(1.07)	-77.07(0.54)	-76.46(1.22)	-77.95(0.26)	-73.66(0.24)
0.50	-64.65(0.30)	-66.22(2.60)	-52.82(0.15)	-61.91(5.90)	-64.70(0.93)
0.40	-47.35(2.22)	-48.75(5.55)	-53.85(0.67)	-46.96(0.22)	-55.66(0.03)
0.30	-40.69(0.72)	-40.28(0.61)	-40.64(0.18)	-41.81(0.08)	-40.15(0.46)
0.20	-35.15(0.87)	-35.56(0.54)	-35.87(0.50)	-35.20(0.38)	-36.06(1.08)
0.10	-30.65(1.28)	-31.87(0.07)	-31.48(0.15)	-30.99(0.27)	-32.17(0.80)
0.00	-25.60(0.62)	-28.52(0.66)	-28.39(0.00)	-26.54(1.98)	-28.71(1.01)
ΔG	**73.67(1.08)**	**73.97(1.54)**	**73.54(0.56)**	**73.22(0.94)**	**74.55(0.84)**

E190A

λ	Run 1	Run 2	Run 3	Run 4	Run 5
1.00	-166.46(0.31)	-169.00(2.96)	-167.31(1.88)	-167.47(1.26)	-167.11(0.58)
0.90	-128.96(0.36)	-127.66(0.91)	-127.90(0.42)	-127.09(1.18)	-128.76(0.97)
0.80	-103.84(0.40)	-100.64(1.18)	-102.05(0.52)	-102.60(0.18)	-100.41(0.58)
0.70	-84.79(0.31)	-86.34(0.64)	-85.67(2.41)	-86.40(0.18)	-86.19(0.40)
0.60	-72.40(1.41)	-73.43(1.02)	-73.49(0.51)	-74.48(0.19)	-73.59(1.64)
0.50	-58.96(6.16)	-63.99(0.11)	-64.24(0.43)	-63.85(0.32)	-63.50(1.47)
0.40	-52.30(0.20)	-55.53(1.31)	-53.02(0.24)	-55.82(0.97)	-55.20(0.22)
0.30	-47.28(0.56)	-47.61(0.55)	-46.09(0.29)	-41.38(0.32)	-48.49(0.22)
0.20	-38.92(2.80)	-41.24(0.28)	-40.89(1.54)	-42.22(0.26)	-42.72(0.47)
0.10	-37.15(0.96)	-36.38(0.43)	-36.56(0.73)	-36.19(0.28)	-36.97(0.80)
0.00	-31.56(0.84)	-29.28(0.25)	-32.32(1.43)	-28.62(1.17)	-32.55(1.15)
ΔG	**72.37(1.30)**	**73.20(0.88)**	**72.98(0.95)**	**72.81(0.58)**	**73.57(0.78)**

H254A

λ	Run 1	Run 2	Run 3	Run 4	Run 5
1.00	-172.67(0.65)	-190.60(0.68)	-180.41(1.14)	-173.02(2.45)	-183.43(0.03)
0.90	-131.78(0.25)	-136.28(1.10)	-136.65(2.09)	-137.29(2.35)	-140.18(1.85)
0.80	-114.68(0.10)	-112.12(1.30)	-113.06(1.30)	-109.04(1.10)	-113.70(0.99)
0.70	-88.26(0.14)	-91.59(0.83)	-89.24(0.28)	-92.79(0.33)	-94.42(0.78)
0.60	-79.43(0.43)	-73.65(1.47)	-75.14(0.27)	-73.51(0.14)	-71.41(0.37)
0.50	-58.09(3.84)	-59.27(1.22)	-60.34(2.61)	-62.57(0.18)	-55.09(0.89)
0.40	-47.48(0.67)	-50.55(2.13)	-47.62(0.25)	-48.85(1.29)	-47.45(0.14)
0.30	-41.88(0.00)	-41.45(0.62)	-41.51(1.05)	-41.35(0.25)	-42.40(0.30)
0.20	-37.76(0.30)	-36.62(0.23)	-36.39(0.02)	-36.88(0.69)	-36.77(0.29)
0.10	-33.08(0.41)	-32.59(0.12)	-32.76(0.63)	-32.04(0.39)	-32.93(0.08)
0.00	-29.48(0.04)	-28.58(0.18)	-29.23(0.45)	-29.18(0.33)	-29.55(0.28)
ΔG	**73.36(0.62)**	**74.37(0.99)**	**73.76(0.92)**	**73.56(0.86)**	**74.09(0.55)**

D265A

λ	Run 1	Run 2	Run 3	Run 4	Run 5
1.00	-178.54 (0.06)	-178.68 (1.36)	-176.56 (2.41)	-173.33 (0.33)	-179.36 (2.15)
0.90	-132.78 (1.42)	-132.54 (0.13)	-133.62 (2.07)	-136.52 (1.07)	-135.83 (1.95)
0.80	-107.79 (1.99)	-108.67 (1.67)	-108.38 (0.31)	-109.23 (1.37)	-108.08 (0.37)
0.70	-90.29 (0.63)	-90.14 (0.89)	-90.99 (0.30)	-89.09 (1.03)	-89.30 (0.94)
0.60	-77.43 (0.65)	-73.84 (1.39)	-72.91 (0.96)	-76.35 (1.52)	-73.82 (2.39)
0.50	-64.95 (0.21)	-63.79 (0.08)	-62.55 (0.23)	-64.57 (0.33)	-63.68 (2.19)
0.40	-47.74 (0.38)	-50.16 (1.14)	-54.44 (0.25)	-51.47 (4.38)	-53.80 (1.12)
0.30	-40.77 (0.67)	-44.05 (0.55)	-42.66 (0.76)	-40.61 (0.49)	-40.56 (0.12)
0.20	-35.68 (0.11)	-37.84 (2.46)	-36.96 (0.65)	-35.38 (0.47)	-36.28 (0.05)
0.10	-31.73 (0.08)	-30.81 (0.62)	-32.73 (0.95)	-31.20 (0.42)	-32.16 (0.46)
0.00	-28.35 (0.82)	-27.73 (0.52)	-27.79 (1.49)	-25.09 (2.06)	-28.46 (0.27)
ΔG	**73.26 (0.64)**	**73.51 (0.98)**	**73.75 (0.95)**	**73.37 (1.23)**	**73.75 (1.10)**

D268A

λ	Run 1	Run 2	Run 3	Run 4	Run 5
1.00	-161.36 (2.91)	-169.56 (1.44)	-160.37 (1.23)	-159.80 (1.20)	-169.97 (1.94)
0.90	-125.10 (1.72)	-131.19 (2.25)	-126.33 (0.32)	-124.65 (0.01)	-130.66 (0.08)
0.80	-100.98 (0.48)	-105.78 (1.00)	-103.63 (0.16)	-101.52 (0.43)	-105.64 (1.40)
0.70	-85.66 (1.07)	-85.79 (0.10)	-86.65 (0.61)	-85.28 (0.01)	-85.78 (1.37)
0.60	-73.19 (0.49)	-73.53 (0.19)	-73.59 (0.68)	-73.19 (0.14)	-74.45 (1.32)
0.50	-64.05 (0.02)	-63.31 (0.77)	-63.81 (0.97)	-63.89 (0.72)	-62.23 (0.16)
0.40	-54.91 (0.18)	-55.53 (1.24)	-54.94 (0.40)	-55.04 (0.15)	-54.83 (0.94)
0.30	-46.96 (0.46)	-45.86 (4.40)	-48.12 (0.38)	-47.45 (0.07)	-47.14 (0.18)
0.20	-41.79 (0.34)	-40.37 (2.05)	-39.92 (4.80)	-41.60 (1.62)	-40.97 (0.49)
0.10	-36.07 (1.72)	-31.32 (5.59)	-33.93 (3.81)	-33.34 (3.77)	-35.39 (0.25)
0.00	-31.17 (2.20)	-27.12 (3.47)	-26.74 (2.56)	-25.76 (2.11)	-30.52 (0.02)
ΔG	**72.50 (1.05)**	**73.11 (2.05)**	**72.45 (1.45)**	**71.88 (0.94)**	**73.74 (0.74)**

Table 5.3. Calculated Mg^{2+} vs Ca^{2+} relative binding free energy (**ΔΔG** in kcal/mol) to Cas1 active sites.

State (Alchemical path) ($Mg^{2+} \rightarrow Ca^{2+}$)	**ΔG_{Comp}**	**ΔG_{free}**	**ΔΔG (kcal/mol)**
Wild type	73.79 ±0.50	68.77 ±0.11	**5.02 ± 0.36**
E190A	72.99 (0.45)		**4.22 ± 0.31**
H254A	73.83 (0.40)		**5.06 ± 0.41**
D265A	73.53 (0.22)		**4.76 ± 0.16**
D268A	72.74 (0.71)		**3.97 ± 0.61**

Table 5.4. Quantum chemical single point energies calculated for free and Cas1 bound divalent ions embedded into a water dielectric.

System	**Theory**	**Energy (in hartree)**				**E4+(E3/E3') - (E1+E2/E2')* 627.5095 = ΔE (in kcal/mol)**
		(E1) **[Mn^{2+} -Wat]**	**(E2)** **E2:[Mg^{2+} -cas1]** **E2':[Ca^{2+} -cas1]**	**(E3)** **E3:[Mg^{2+} -Wat]** **E3':[Ca^{2+} -Wat]**	**(E4)** **[Mn^{2+} -cas1]**	
WT	B3LYP/6-31+G*	-1379.983466	-3618.4970701 -4095.9472798	-429.2065101 -906.7979605	-4569.326185	**-32.73** **-121.36**
E190A	B3LYP/6-31+G*	-1456.4217949	-3467.4438553 -3944.9611905	-505.6644207 -983.2119464	-4418.2977688	**-60.58** **-79.52**
H254A	B3LYP/6-31+G*	-1456.4236473	-3469.8844016 -3947.3572744	-505.6665397 -983.2275078	-4420.6821449	**-25.50** **-80.80**
D268A	B3LYP/6-31+G*	-1532.8575864	-3583.1949785 -4060.730279	-582.1054835 -1059.6698246	-4534.027688	**-50.58** **-68.80**
D265A	B3LYP/6-31+G*	-1379.9817044	-3430.336254 -3907.8292782	-429.2034479 -906.7987892	-4381.169419	**-34.45** **-98.66**

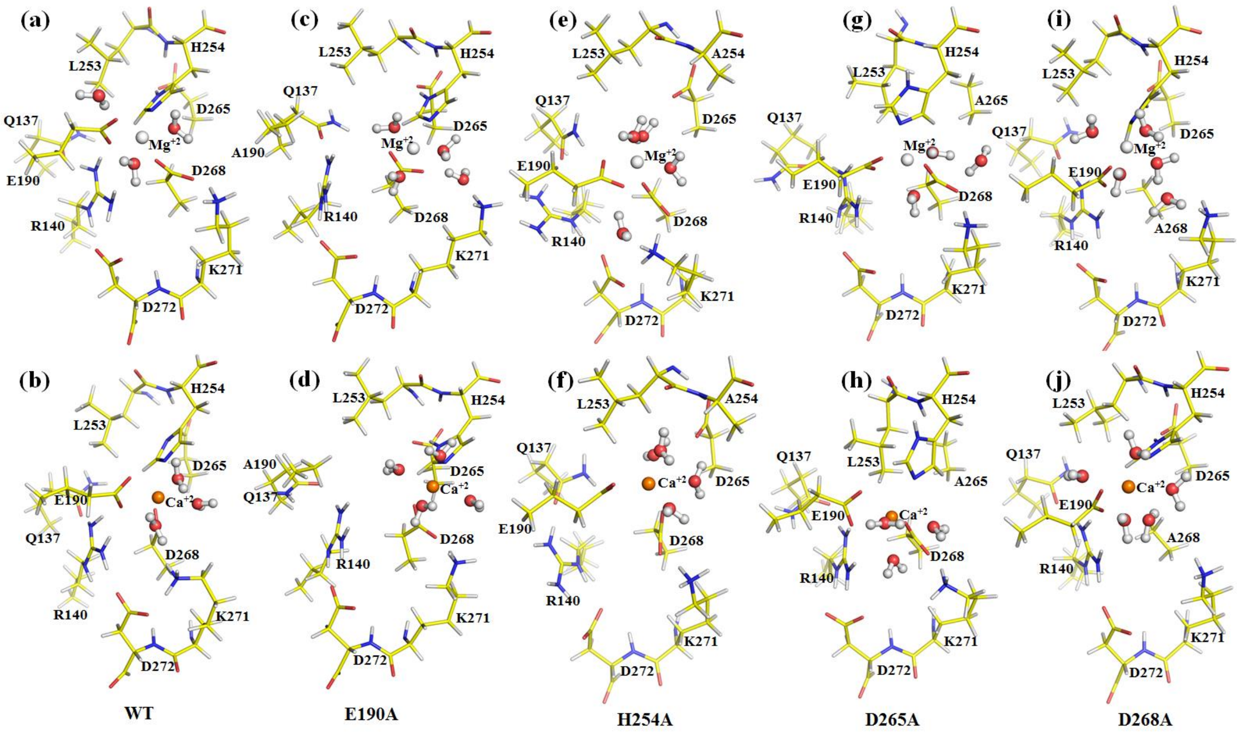


Figure 5.1. Local environment around divalent metal ion binding sites (Mn^{+2}, Mg^{+2}, and Ca^{+2}) considered for quantum chemical calculations. WT (a,b), E190A (c,d), H254A (e,f), D265A (g,h), and D268A (i,j).

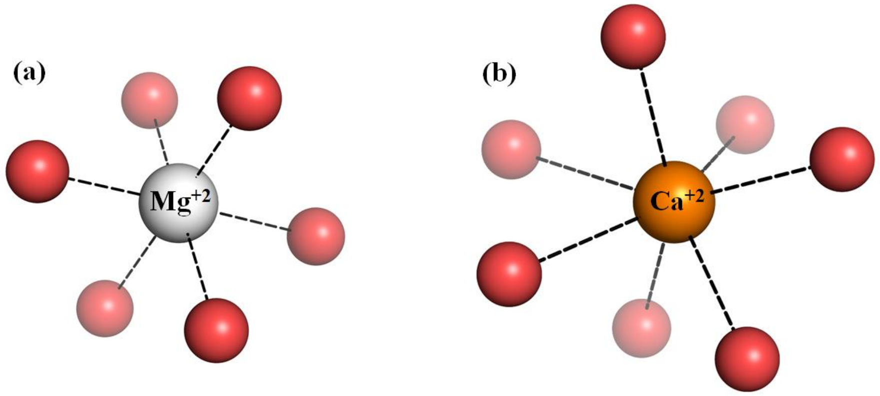


Figure 5.2. MD insight into the first coordination shell of free Mg^{2+} and Ca^{2+} in water. Interaction distances are shown as broken lines.

References

Agranoff, D. D., & Krishna, S. (1998). Metal ion homeostasis and intracellular parasitism. *Molecular microbiology*, *28*(3), 403-412.

Allen, M. P., & Tildesley, D. J. (2017). *Computer simulation of liquids*. Oxford university press.

Åqvist, J., & Luzhkov, V. (2000). Ion permeation mechanism of the potassium channel. *Nature*, *404*(6780), 881-884.

Banci, L. (2003). Molecular dynamics simulations of metalloproteins. *Current opinion in chemical biology*, *7*(1), 143-149.

Banerjee, R., Kumar, A., Satpati, P., & Nagotu, S. (2021). Mimicking human Drp1 disease-causing mutations in yeast Dnm1 reveals altered mitochondrial dynamics. *Mitochondrion*, S1567-7249.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., ... & Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, *315*(5819), 1709-1712.

Barrangou, R. (2013). CRISPR‐Cas systems and RNA‐guided interference. *Wiley Interdisciplinary Reviews: RNA*, *4*(3), 267-278.

Barrangou, R., & Marraffini, L. A. (2014). CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. *Molecular cell*, *54*(2), 234-244.

Basu, S., Rambo, R. P., Strauss-Soukup, J., Cate, J. H., Ferré-D, A. R., Strobel, S. A., & Doudna, J. A. (1998). A specific monovalent metal ion integral to the AA platform of the RNA tetraloop receptor. *Nature structural biology*, *5*(11), 986-992.

Becke, A. D. (1988). Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical review A*, *38*(6), 3098.

Banci, L. (2003). Molecular dynamics simulations of metalloproteins. *Current opinion in chemical biology*, *7*(1), 143-149.

Bertram, K., Agafonov, D. E., Liu, W. T., Dybkov, O., Will, C. L., Hartmuth, K., ... & Lührmann, R. (2017). Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature*, *542*(7641), 318-323.

Bock, C. W., Katz, A. K., Markham, G. D., & Glusker, J. P. (1999). Manganese as a replacement for magnesium and zinc: functional comparison of the divalent ions. *Journal of the American Chemical Society*, *121*(32), 7360-7372.

Bogusz, S., Cheatham III, T. E., & Brooks, B. R. (1998). Removal of pressure and free energy artifacts in charged periodic systems via net charge corrections to the Ewald potential. *The Journal of chemical physics*, 108(17), 7070-7084.

Bolotin, A., Quinquis, B., Sorokin, A., & Ehrlich, S. D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, *151*(8), 2551-2561.

Borhani, D. W., & Shaw, D. E. (2012). The future of molecular dynamics simulations in drug discovery. *Journal of computer-aided molecular design*, *26*(1), 15-26.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. A., & Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry*, *4*(2), 187-217.

Brooks, B. R., Brooks III, C. L., Mackerell Jr, A. D., Nilsson, L., Petrella, R. J., Roux, B., ... & Karplus, M. (2009). CHARMM: the biomolecular simulation program. *Journal of computational chemistry*, *30*(10), 1545-1614.

Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J., Snijders, A. P., ... & Van Der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, *321*(5891), 960-964.

Casalino, L., Palermo, G., Rothlisberger, U., & Magistrato, A. (2016). Who activates the nucleophile in ribozyme catalysis? An answer from the splicing mechanism of group II introns. *Journal of the American Chemical Society*, *138*(33), 10374-10377.

Chan, R. T., Peters, J. K., Robart, A. R., Wiryaman, T., Rajashankar, K. R., & Toor, N. (2018). Structural basis for the second step of group II intron splicing. *Nature communications*, *9*(1), 1-10.

Chan, R. T., Robart, A. R., Rajashankar, K. R., Pyle, A. M., & Toor, N. (2012). Crystal structure of a group II intron in the pre-catalytic state. *Nature structural & molecular biology*, *19*(5), 555-557.

Chu, J. W., Trout, B. L., & Brooks, B. R. (2003). A super-linear minimization scheme for the nudged elastic band method. *The Journal of chemical physics*, *119*(24), 12708-12717.

Clancy, S. (2008). RNA splicing: introns, exons and spliceosome. *Nature Education*, *1*(1), 31.

Cruz-León, S., & Schwierz, N. (2020). Hofmeister Series for Metal-Cation–RNA Interactions: The Interplay of Binding Affinity and Exchange Kinetics. *Langmuir, 36*(21), 5979-5989.

Comba, P., Ströhle, M., & Hambley, T. W. (1995). The Directionality of d-Orbitals and Molecular-Mechanics Calculations of Octahedral Transition-Metal Compounds. *Helvetica chimica acta*, *78*(8), 2042-2047.

Conn, G. L., Gittis, A. G., Lattman, E. E., Misra, V. K., & Draper, D. E. (2002). A compact RNA tertiary structure contains a buried backbone–K^+ complex. *Journal of molecular biology*, *318*(4), 963-973.

Cooper, T. A., Wan, L., & Dreyfuss, G. (2009). RNA and disease. *Cell*, *136*(4), 777-793.

Costa, M., Walbott, H., Monachello, D., Westhof, E., & Michel, F. (2016). Crystal structures of a group II intron lariat primed for reverse splicing. *Science*, *354*(6316).

Costa, M., Walbott, H., Monachello, D., Westhof, E., & Michel, F. (2016). Crystal structures of a group II intron lariat primed for reverse splicing. *Science*, *354*(6316), aaf9258.

Cotton, F. A., Wilkinson, G., Murillo, C. A., & Bochmann, M. (1999). *Advanced inorganic chemistry*. John Wiley and Sons, Inc.

Denesyuk, N. A., & Thirumalai, D. (2015). How do metal ions direct ribozyme folding? *Nature chemistry*, *7*(10), 793-801.

Dai, L., & Zimmerly, S. (2002). The dispersal of five group II introns among natural populations of Escherichia coli. *Rna*, *8*(10), 1294-1307.

Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *The Journal of chemical physics*, 98(12), 10089-10092.

DeLano, W. L. (2002). Pymol: An open-source molecular graphics tool. *CCP4 Newsletter on protein crystallography*, *40*(1), 82-92.

Deveau, H., Garneau, J. E., & Moineau, S. (2010). CRISPR/Cas system and its role in phage-bacteria interactions. *Annual review of microbiology*, *64*, 475-493.

Dror, R. O., Dirks, R. M., Grossman, J. P., Xu, H., & Shaw, D. E. (2012). Biomolecular simulation: a computational microscope for molecular biology. *Annual review of biophysics*, *41*, 429-452.

Dudev, T., & Lim, C. (2000). Tetrahedral vs octahedral zinc complexes with ligands of biological interest: a DFT/CDM study. *Journal of the American Chemical Society*, *122*(45), 11146-11153.

Dudev, T., & Lim, C. (2014). Competition among metal ions for protein binding sites: determinants of metal ion selectivity in proteins. *Chemical reviews*, *114*(1), 538-556.

Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC biology*, *9*(1), 1-9.

Durell, S. R., Brooks, B. R., & Ben-Naim, A. (1994). Solvent-induced forces between two hydrophilic groups. *The Journal of Physical Chemistry*, *98*(8), 2198-2202.

Erat, M. C., & Sigel, R. K. (2007). Determination of the intrinsic affinities of multiple site-specific Mg^{2+} ions coordinated to domain 6 of a group II intron ribozyme. *Inorganic chemistry*, *46*(26), 11224-11234.

Erat, M. C., & Sigel, R. K. (2008). Divalent metal ions tune the self-splicing reaction of the yeast mitochondrial group II intron Sc. ai5γ. *JBIC Journal of Biological Inorganic Chemistry*, *13*(6), 1025-1036.

Estarellas, C., Otyepka, M., Koča, J., Banáš, P., Krepl, M., & Šponer, J. (2015). Molecular dynamic simulations of protein/RNA complexes: CRISPR/Csy4 endoribonuclease. *Biochimica et Biophysica Acta (BBA)-General Subjects*, *1850*(5), 1072-1090.

Farah, C. S., & Reinach, F. C. (1995). The troponin complex and regulation of muscle contraction. *The FASEB Journal*, *9*(9), 755-767.

Fedorova, O., & Zingler, N. (2007). Group II introns: structure, folding and splicing mechanism. *Biological Chemistry*, 388 (7) 665-678.

Feller, S. E., Zhang, Y., Pastor, R. W., & Brooks, B. R. (1995). Constant pressure molecular dynamics simulation: the Langevin piston method. *The Journal of chemical physics*, *103*(11), 4613-4621.

Fica, S. M., Tuttle, N., Novak, T., Li, N. S., Lu, J., Koodathingal, P., ... & Piccirilli, J. A. (2013). RNA catalyses nuclear pre-mRNA splicing. *Nature*, *503*(7475), 229-234.

Fineran, P. C., & Charpentier, E. (2012). Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. *Virology*, *434*(2), 202-209.

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., ... & Fox, D. J. (2016). Gaussian 16.

Galej, W. P., Wilkinson, M. E., Fica, S. M., Oubridge, C., Newman, A. J., & Nagai, K. (2016). Cryo-EM structure of the spliceosome immediately after branching. *Nature*, *537*(7619), 197-201.

García-Rodríguez, F. M., Barrientos-Durán, A., Díaz-Prado, V., Fernández-López, M., & Toro, N. (2011). Use of RmInt1, a group IIB intron lacking the intron-encoded protein endonuclease domain, in gene targeting. *Applied and environmental microbiology*, *77*(3), 854-861.

Garneau, J. E., Dupuis, M. È., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., ... & Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, *468*(7320), 67-71.

Gifford, J. L., Walsh, M. P., & Vogel, H. J. (2007). Structures and metal-ion-binding properties of the Ca2+-binding helix–loop–helix EF-hand motifs. *Biochemical Journal*, *405*(2), 199-221.

Gordon, P. M., & Piccirilli, J. A. (2001). Metal ion coordination by the AGC triad in domain 5 contributes to group II intron catalysis. *Nature structural biology*, *8*(10), 893-898.

Haft, D. H., Selengut, J., Mongodin, E. F., & Nelson, K. E. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS computational biology*, *1*(6), e60.

Hang, J., Wan, R., Yan, C., & Shi, Y. (2015). Structural basis of pre-mRNA splicing. *Science*, *349*(6253), 1191-1198.

Hockney, R. W. (1970). The potential calculation and some applications. *Methods Comput. Phys.*, *9*, 136.

Horvath, P., & Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. *Science*, *327*(5962), 167-170.

Hoover, W. G., Ladd, A. J., & Moran, B. (1982). High-strain-rate plastic flow studied via nonequilibrium molecular dynamics. *Physical Review Letters*, *48*(26), 1818.

Hoover, W. G. (1985). Canonical dynamics: Equilibrium phase-space distributions. *Physical review A*, *31*(3), 1695.

Hu, L., & Ryde, U. (2011). Comparison of methods to obtain force-field parameters for metal sites. *Journal of Chemical Theory and Computation*, *7*(8), 2452-2463.

Hummer, G., Pratt, L. R., & Garcia, A. E. (1996). Free energy of ionic hydration. *The Journal of Physical Chemistry*, 100(4), 1206-1215.

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *Journal of molecular graphics*, *14*(1), 33-38.

Jackson, S. A., McKenzie, R. E., Fagerlund, R. D., Kieper, S. N., Fineran, P. C., & Brouns, S. J. J. (2017). CRISPR-Cas: adapting to change. *Science* 356: eaal5056.

Jansen, R., Embden, J. D. V., Gaastra, W., & Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular microbiology*, *43*(6), 1565-1575.

Jarrell, K. A., Peebles, C. L., Dietrich, R. C., Romiti, S. L., & Perlman, P. S. (1988). Group II intron self-splicing. Alternative reaction conditions yield novel products. *Journal of Biological Chemistry*, *263*(7), 3432-3439.

Jiang, F., & Doudna, J. A. (2017). CRISPR–Cas9 structures and mechanisms. *Annual review of biophysics*, *46*, 505-529.

Jing, Z., Liu, C., Qi, R., & Ren, P. (2018). Many-body effect determines the selectivity for Ca^{2+} and Mg^{2+} in proteins. *Proceedings of the National Academy of Sciences*, *115*(32), E7495-E7501.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2), 926-935.

Kirkwood, J. G. (1935). Statistical mechanics of fluid mixtures. *The Journal of chemical physics*, 3(5), 300-313.

Klein, D. J., Moore, P. B., & Steitz, T. A. (2004). The contribution of metal ions to the structural stability of the large ribosomal subunit. *Rna*, *10*(9), 1366-1379.

Kohagen, M., Lepsik, M., & Jungwirth, P. (2014). Calcium binding to calmodulin by molecular dynamics with effective polarization. *The journal of physical chemistry letters*, *5*(22), 3964-3969.

Kollman, P. (1993). Free energy calculations: applications to chemical and biochemical phenomena. *Chemical reviews*, *93*(7), 2395-2417.

Kruschel, D., & Sigel, R. K. (2008). Divalent metal ions promote the formation of the 5′-splice site recognition complex in a self-splicing group II intron. *Journal of inorganic biochemistry*, *102*(12), 2147-2154.

Kumar, A., Basu, D., & Satpati, P. (2017). Structure-Based Energetics of Stop Codon Recognition by Eukaryotic Release Factor. Journal of chemical information and modeling, 57(9), 2321-2328.

Kumar, A., & Satpati, P. (2018). Principle of K^+/Na^+ selectivity in the active site of group II intron at various stages of self-splicing pathway. *Journal of Molecular Graphics and Modelling*, *84*, 1-9.

Kumar, A., & Satpati, P. (2018). Energetics of preferential binding of retinoic acid-inducible gene-I to double-stranded viral RNAs with 5′ Tri-/Diphosphate over 5′ Monophosphate. *ACS omega*, *3*(4), 3786-3795.

Kumar, A., Åqvist, J., & Satpati, P. (2019). Principles of tRNAAla Selection by Alanyl–tRNA Synthetase Based on the Critical G3· U70 Base Pair. *ACS omega*, 4(13), 15539-15548.

Kumar, A., & Satpati, P. (2020). Mg^{2+} vs Ca^{2+} bound active site of group II intron–A MD Study. *Journal of Molecular Graphics and Modelling*, *97*, 107546.

Kumar, A., & Satpati, P. (2020). Why double-stranded RNA with 5′-monophosphate is a poor binder to retinoic acid-inducible gene-I with respect to 5′-hydroxyl analogue?. *Journal of Biomolecular Structure and Dynamics*, *38*(13), 4048-4055.

Kumar, A., & Satpati, P. (2021). Divalent-Metal-Ion Selectivity of the CRISPR-Cas System-Associated Cas1 Protein: Insights from Classical Molecular Dynamics Simulations and Electronic Structure Calculations. *The Journal of Physical Chemistry B*, *125*(43), 11943-11954.

Kuppuraj, G., Dudev, M., & Lim, C. (2009). Factors governing metal− ligand distances and coordination geometries of metal complexes. *The journal of physical chemistry B*, *113*(9), 2952-2960.

Lambert, D., Leipply, D., Shiman, R., & Draper, D. E. (2009). The influence of monovalent cation size on the stability of RNA tertiary structures. *Journal of molecular biology*, *390*(4), 791-804.

Lee, C., Yang, W., & Parr, R. G. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical review B*, *37*(2), 785.

Lee, M. S., Salsbury Jr, F. R., & Brooks III, C. L. (2002). Novel generalized Born methods. *The Journal of chemical physics*, *116*(24), 10606-10614.

Lee, M. S., Feig, M., Salsbury Jr, F. R., & Brooks III, C. L. (2003). New analytic approximation to the standard molecular volume definition and its application to generalized Born calculations. *Journal of computational chemistry*, *24*(11), 1348-1356.

Lee, J., Cheng, X., Swails, J. M., Yeom, M. S., Eastman, P. K., Lemkul, J. A., ... & Im, W. (2016). CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *Journal of chemical theory and computation*, *12*(1), 405-413.

Liu, Y., Sengupta, A., Raghavachari, K., & Flood, A. H. (2017). Anion binding in solution: beyond the electrostatic regime. *Chem*, *3*(3), 411-427.

Lind, C., Esguerra, M., Jespers, W., Satpati, P., Gutierrez-de-Terán, H., & Åqvist, J. (2019). Free energy calculations of RNA interactions. *Methods*, 162, 85-95.

MacKerell Jr, A. D., Wiorkiewicz-Kuczera, J., & Karplus, M. (1995). An all-atom empirical energy function for the simulation of nucleic acids. *Journal of the American Chemical society*, 117(48), 11946-11975.

MacKerell Jr, A. D., Bashford, D., Bellott, M. L. D. R., Dunbrack Jr, R. L., Evanseck, J. D., Field, M. J., ... & Karplus, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The journal of physical chemistry B*, 102(18), 3586-3616.

Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I., & Koonin, E. V. (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology direct*, *1*(1), 1-26.

Marenich, A. V., Cramer, C. J., & Truhlar, D. G. (2009). Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *The Journal of Physical Chemistry B*, *113*(18), 6378-6396.

Marcia, M., & Pyle, A. M. (2012). Visualizing group II intron catalysis through the stages of splicing. *Cell*, *151*(3), 497-507.

Martyna, G. J., Tobias, D. J., & Klein, M. L. (1994). Constant pressure molecular dynamics algorithms. *The Journal of chemical physics*, *101*(5), 4177-4189.

Martynov, A., Severinov, K., & Ispolatov, I. (2017). Optimal number of spacers in CRISPR arrays. *PLoS computational biology*, *13*(12), e1005891.

Marraffini, L. A., & Sontheimer, E. J. (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature*, *463*(7280), 568-571.

Mattick, J. S. (1994). Introns: evolution and function. *Current opinion in genetics & development*, *4*(6), 823-831.

McCammon, J. A. (1991). Free energy from simulations: Current Opinion in Structural Biology 1991, 1: 196–200. Current Opinion in Structural Biology, 1(2), 196-200.

Michel, F., Kazuhiko, U., & Haruo, O. (1989). Comparative and functional anatomy of group II catalytic introns—a review. *Gene*, *82*(1), 5-30.

Mojica, F. J., Díez-Villaseñor, C., Soria, E., & Juez, G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular microbiology*, *36*(1), 244-246.

Mojica, F. J., Díez-Villaseñor, C., García-Martínez, J., & Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution*, *60*(2), 174-182.

Mordasini, T. Z., & McCammon, J. A. (2000). Calculations of relative hydration free energies: a comparative study using thermodynamic integration and an extrapolation method based on a single reference state. *The Journal of Physical Chemistry B*, 104(2), 360-367.

Mustafi, D., Bekesi, A., Vertessy, B. G., & Makinen, M. W. (2003). Catalytic and structural role of the metal ion in dUTP pyrophosphatase. *Proceedings of the National Academy of Sciences*, *100*(10), 5670-5675.

Neves, R. P., Sousa, S. F., Fernandes, P. A., & Ramos, M. J. (2013). Parameters for molecular dynamics simulations of manganese-containing metalloproteins. *Journal of chemical theory and computation*, *9*(6), 2718-2732.

Nosé, S. (1984). A molecular dynamics method for simulations in the canonical ensemble. *Molecular physics*, *52*(2), 255-268.

Nuñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V., Davies, C. W., & Doudna, J. A. (2014). Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nature structural & molecular biology*, *21*(6), 528-534.

Nunez, J. K., Harrington, L. B., Kranzusch, P. J., Engelman, A. N., & Doudna, J. A. (2015). Foreign DNA capture during CRISPR–Cas adaptive immunity. *Nature*, *527*(7579), 535-538.

Palermo, G., Miao, Y., Walker, R. C., Jinek, M., & McCammon, J. A. (2016). Striking plasticity of CRISPR-Cas9 and key role of non-target DNA, as revealed by molecular simulations. *ACS central science*, *2*(10), 756-763.

Palermo, G., Miao, Y., Walker, R. C., Jinek, M., & McCammon, J. A. (2017). CRISPR-Cas9 conformational activation as elucidated from enhanced molecular simulations. *Proceedings of the National Academy of Sciences*, *114*(28), 7260-7265.

Palermo, G. (2019). Structure and dynamics of the CRISPR–Cas9 catalytic complex. *Journal of chemical information and modeling*, *59*(5), 2394-2406.

Pan, Q., Shai, O., Lee, L. J., Frey, B. J., & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature genetics*, *40*(12), 1413-1415.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, *25*(13), 1605-1612.

Peebles, C. L., Perlman, P. S., Mecklenburg, K. L., Petrillo, M. L., Tabor, J. H., Jarrell, K. A., & Cheng, H. L. (1986). A self-splicing RNA excises an intron lariat. *Cell*, *44*(2), 213-223.

Pinello, L., Canver, M. C., Hoban, M. D., Orkin, S. H., Kohn, D. B., Bauer, D. E., & Yuan, G. C. (2016). Analyzing CRISPR genome-editing experiments with CRISPResso. *Nature biotechnology*, *34*(7), 695-697.

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., ... & Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of computational chemistry*, *26*(16), 1781-1802.

Podar, M., Perlman, P. S., & Padgett, R. A. (1995). Stereochemical selectivity of group II intron splicing, reverse splicing, and hydrolysis reactions. *Molecular and cellular biology*, *15*(8), 4466-4478.

Pourcel, C., Salvignol, G., & Vergnaud, G. (2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, *151*(3), 653-663.

Pyle, A. M., & Lambowitz, A. M. (2006). Group II introns: ribozymes that splice RNA and invade DNA. *COLD SPRING HARBOR MONOGRAPH SERIES*, *43*, 469.

Raczyński, P., Gorny, K., Pabiszczak, M., & Gburski, Z. (2013). Nanoindentation of biomembrane by carbon nanotubes–MD simulation. *Computational materials science*, *70*, 13-18.

Robart, A. R., Chan, R. T., Peters, J. K., Rajashankar, K. R., & Toor, N. (2014). Crystal structure of a eukaryotic group II intron lariat. *Nature*, *514*(7521), 193-197.

Rolfs, A., & Hediger, M. A. (1999). Metal ion transporters in mammals: structure, function and pathological implications. *The Journal of Physiology*, *518*(1), 1-12.

Ryckaert, Jean-Paul, Giovanni Ciccotti, and Herman JC Berendsen. "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes." *Journal of computational physics* 23.3 (1977): 327-341.

Satpati, P., Clavaguéra, C., Ohanessian, G., & Simonson, T. (2011). Free energy Simulations of a GTPase: GTP and GDP binding to archaeal initiation factor 2. *The Journal of Physical Chemistry B*, 115(20), 6749-6763.

Satpati, P., Sund, J., & Åqvist, J. (2014). Structure-based energetics of mRNA decoding on the ribosome. *Biochemistry*, *53*(10), 1714-1722.

Kumar, A., Basu, D., & Satpati, P. (2017). Structure-Based Energetics of Stop Codon Recognition by Eukaryotic Release Factor. *Journal of Chemical Information and Modeling*, *57*(9), 2321-2328.

Savitskaya, E., Semenova, E., Dedkov, V., Metlitskaya, A., & Severinov, K. (2013). High-throughput analysis of type IE CRISPR/Cas spacer acquisition in E. coli. *RNA biology*, *10*(5), 716-725.

Schnabl, J., & Sigel, R. K. (2010). Controlling ribozyme activity by metal ions. *Current opinion in chemical biology*, *14*(2), 269-275.

Schwaller, B. (2010). Cytosolic Ca2+ buffers. *Cold Spring Harbor perspectives in biology*, *2*(11), a004051.

Scotti, M. M., & Swanson, M. S. (2016). RNA mis-splicing in disease. *Nature Reviews Genetics*, *17*(1), 19-32.

Shannon, R. D. (1976). Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta crystallographica section A: crystal physics, diffraction, theoretical and general crystallography*, *32*(5), 751-767.

Sigel, R. K., Vaidya, A., & Pyle, A. M. (2000). Metal ion binding sites in a group II intron core. *Nature structural biology*, *7*(12), 1111-1116.

Sigel, R. K. (2005). Group II intron ribozymes and metal ions–a delicate relationship. *European journal of inorganic chemistry*, *2005*(12), 2281-2292.

Sissi, C., Marangon, E., Chemello, A., Noble, C. G., Maxwell, A., & Palumbo, M. (2005). The effects of metal ions on the structure and stability of the DNA gyrase B protein. *Journal of molecular biology*, *353*(5), 1152-1160.

Swarts, D. C., Mosterd, C., Van Passel, M. W., & Brouns, S. J. (2012). CRISPR interference directs strand specific spacer acquisition. *PloS one*, *7*(4), e35888.

Swisher, J. F., Su, L. J., Brenowitz, M., Anderson, V. E., & Pyle, A. M. (2002). Productive folding to the native state by a group II intron ribozyme. *Journal of molecular biology*, *315*(3), 297-310.

Su, L. J., Brenowitz, M., & Pyle, A. M. (2003). An alternative route for the folding of large RNAs: apparent two-state folding by a group II intron ribozyme. *Journal of molecular biology*, *334*(4), 639-652.

Su, L. J., Waldsich, C., & Pyle, A. M. (2005). An obligate intermediate along the slow folding pathway of a group II intron ribozyme. *Nucleic acids research*, *33*(21), 6674-6687.

Swope, W. C., Andersen, H. C., Berens, P. H., & Wilson, K. R. (1982). A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *The Journal of chemical physics*, *76*(1), 637-649.

Toor, N., Haushner, G., & Zimmerly, S. (2001). Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *Rna*, *7*(8), 1142-1152.

Toor, N., Robart, A. R., Christianson, J., & Zimmerly, S. (2006). Self-splicing of a group IIC intron: 5′ exon recognition and alternative 5′ splicing events implicate the stem–loop motif of a transcriptional terminator. *Nucleic acids research*, *34*(22), 6461-6471.

Toor, N., Keating, K. S., Taylor, S. D., & Pyle, A. M. (2008). Crystal structure of a self-spliced group II intron. *Science*, 320(5872), 77-82.

Tsui, V., & Case, D. A. (2001). Calculations of the absolute free energies of binding between RNA and metal ions using molecular dynamics simulations and continuum electrostatics. *The Journal of Physical Chemistry B*, *105*(45), 11314-11325.

Van der Oost, J., Jore, M. M., Westra, E. R., Lundgren, M., & Brouns, S. J. (2009). CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends in biochemical sciences*, *34*(8), 401-407.

Van der Veen, R., Arnberg, A. C., Van der Horst, G., Bonen, L., Tabak, H. F., & Grivell, L. A. (1986). Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing in vitro. *Cell*, *44*(2), 225-234.

Verlet, L. (1967). Computer" experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical review*, *159*(1), 98.

Viadiu, H., & Aggarwal, A. K. (1998). The role of metals in catalysis by the restriction endonuclease Bam HI. *Nature structural biology*, *5*(10), 910-916.

Wan, H., Li, J., Chang, S., Lin, S., Tian, Y., Tian, X., ... & Hu, J. (2019). Probing the behaviour of Cas1-Cas2 upon protospacer binding in CRISPR-Cas systems using molecular dynamics simulations. *Scientific reports*, *9*(1), 1-16.

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., ... & Burge, C. B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature*, *456*(7221), 470-476.

Wiedenheft, B., Zhou, K., Jinek, M., Coyle, S. M., Ma, W., & Doudna, J. A. (2009). Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure*, *17*(6), 904-912.

Wright, A. V., Liu, J. J., Knott, G. J., Doxzen, K. W., Nogales, E., & Doudna, J. A. (2017). Structures of the CRISPR genome integration complex. *Science*, *357*(6356), 1113-1118.

Xiao, Y., Ng, S., Nam, K. H., & Ke, A. (2017). How type II CRISPR–Cas establish immunity through Cas1–Cas2-mediated spacer integration. *Nature*, *550*(7674), 137-141.

Xu, X., Duan, D., & Chen, S. J. (2017). CRISPR-Cas9 cleavage efficiency correlates strongly with target-sgRNA folding stability: from physical mechanism to off-target assessment. *Scientific reports*, *7*(1), 1-9.

Yang, W. (2008). An equivalent metal ion in one-and two-metal-ion catalysis. *Nature structural & molecular biology*, *15*(11), 1228-1231.

Ye, Y., Lee, H. W., Yang, W., Shealy, S. J., Wilkins, A. L., Liu, Z. R., ... & Yang, J. J. (2001). Metal binding affinity and structural properties of an isolated EF-loop in a scaffold protein. *Protein engineering*, *14*(12), 1001-1013.

Zeng, Y., Cui, Y., Zhang, Y., Zhang, Y., Liang, M., Chen, H., ... & Lou, J. (2018). The initiation, propagation and dynamics of CRISPR-SpyCas9 R-loop complex. *Nucleic acids research, 46*(1), 350-361.

Zhang, S. T., Yan, H., Wei, M., Evans, D. G., & Duan, X. (2012). Valence force field for layered double hydroxide materials based on the parameterization of octahedrally coordinated metal cations. *The Journal of Physical Chemistry C, 116*(5), 3421-3431.

Zhao, Y., & Truhlar, D. G. (2008). The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theoretical chemistry accounts*, *120*(1), 215-241.

Zheng, W. (2017). Probing the structural dynamics of the CRISPR-Cas9 RNA-guided DNA-cleavage system by coarse-grained modeling. *Proteins: Structure, Function, and Bioinformatics*, *85*(2), 342-353.

List of Publications:

1. **Kumar, A.**, & Satpati*, P. (2018). Principle of K^+/Na^+ selectivity in the active site of group II intron at various stages of self-splicing pathway. *Journal of Molecular Graphics and Modelling*, *84*, 1-9.
2. **Kumar, A.**, Mukherjee, D., & Satpati*, P. (2019). Mutations in Parkinson's Disease Associated Protein DJ-1 Alter the Energetics of DJ-1 Dimerization. *Journal of Chemical Information and Modeling*, *59*(4), 1497-1507.
3. **Kumar, A.**, & Satpati*, P. (2020). Mg^{2+} vs Ca^{2+} bound active site of group II intron–A MD study. *Journal of Molecular Graphics and Modelling*, *97*, 107546.
4. Banerjee, R., **Kumar, A.**, Satpati, P., & Nagotu*, S. (2021). Mimicking human Drp1 disease-causing mutations in yeast Dnm1 reveals altered mitochondrial dynamics. *Mitochondrion*, S1567-7249.
5. **Kumar, A.**, & Satpati*, P. (2021). Divalent Metal-ion Selectivity of CRISPR-Cas System Associated Cas1 Protein- Insights from Classical Molecular Dynamics Simulations and Electronic Structure Calculations. *The Journal of Physical Chemistry B*, *125*, 43, 11943-11954.
6. **Kumar, A.**, & Satpati*, P. (2021). Structure based energetics of ion selectivity (Mg^{2+} vs Ca^{2+} and K^+ vs Na^+) in the active site of lariat group II intron of a brown algae *Pylaiella littoralis (**submitted**)*.

Workshops attended:

1. Completed 3 days workshop on "**Advanced Workshop on Molecular Docking, Virtual Screening & Computational Biology**" Organized by Indian Institute of Information Technology Allahabad, Prayagraj, 211015, India from 13 -15 March 2019.

2. Completed 3 days workshop on "**Artificial Intelligence (AI) in Healthcare Engineering**" Organized by NECBH, Department of CSE, Department of EEE IIT Guwahati, Guwahati, 781039, India from 28 -30 March 2019.