UNIVERSITY OF CALIFORNIA, SAN DIEGO

Structure-Function Relationships in Enzymes of the Nucleotidyltransferase Superfamily:

RNA ligation and mRNA capping

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Chemistry

by

Robert Swift

Committee in charge:

Professor J. Andrew McCammon, Chair
Professor Elizabeth Komives
Professor Tadeusz Molinski
Professor Terunaga Nakagawa
Professor Jose Onuchic

2009

The Dissertation of Robert Swift is approved, and it is acceptable

in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2009

DEDICATION

To Sally, Klaus, Mike, Jim, Kalju, and Andy for inspiration.

EPIGRAPH

I think I can, I think I can, I think I can…

The Little Engine That Could

TABLE OF CONTENTS

LIST OF ABREVIATIONS

OB, (oligonucleotide/oligosaccharide) binding

CD, circular dichroism

MD, molecular dynamics

TMD, targeted molecular dynamics

BD, Brownian dynamics

RMSD, root mean square deviation

RMSF, root mean square fluctuation

PDB, protein data bank

PMF, potential of mean force

NMN, nitotinic mononucleotide

dsDNA, double stranded DNA

dsRNA, double stranded RNA

ssRNA, single stranded RNA

PBCV, *Paramecium bursaria Chlorella* virus

TbREL, *Trypanosoma brucei* RNA editing ligase

T4Rnl2, bacteriophage T4 RNA ligase 2

gRNA, guide RNA

LIG1, human DNA ligase 1

AppN, AMP attached to the nicked 5´-PO_4

LIST OF FIGURES

LIST OF TABLES

AKNOWLEDGEMENTS

The PhD is an academic degree, and while it may reflect a certain amount of tenacity in the matriculating individual, it is not awarded to acknowledge any of those human qualities that define who we are. Inspired by its academic nature, I dedicated this dissertation to the teachers and professors who shared their enthusiasm, passion, understanding, and motivated my desire to pursue science.

Here, in the acknowledgements then, I want to briefly thank those who fall outside of the academic spectrum, but who have nevertheless left a lasting, positive impression and shaped who I am as an individual. First, a heartfelt thank you to my mom for her nurturing support and unflagging belief in me; it's a tremendous blessing that I'm daily grateful of. Next, to my stepdad Don, I'm appreciative of his kindness and encouragement, which has often left my world a brighter place. Finally, I am forever grateful to my dad, not only for his love, but also for being a great role model in my life. His example has convinced me of the magic hidden in the persistent pursuit of a goal.

Chapter 2, in full, is a reprint of "Functional and Structural Insights Revealed by Molecular Dynamics Simulations of an Essential RNA Editing Ligase in Trypanosoma brucie", which was published 2007 in PLoS Neglected Tropical Diseases, volume 1, issue 2, e68, by Romie E. Amaro, Robert V. Swift, and J. Andrew McCammon. The dissertation author was the secondary investigator and author of this paper.

Chapter 3, in full, is a reprint of "Toward Understanding the Conformational Dynamics of RNA Ligation", which was published in 2009 in Biochemistry, volume 48, issue, 4, pages 709 to 719, by Robert V. Swift, Jacob Durrant, Rommie E. Amaro, and J.

Andrew McCammon. The dissertation author was the primary investigator and shared first authorship of this paper.

Chapter 4, in full, is a reprint of "Catalytically requisite conformational dynamics in the mRNA-capping enzyme probed by targeted molecular dynamics", which was published in 2008 in Biochemistry, volume 47, issue 13, pages 4102 to 4111, by Robert V. Swift, and J. Andrew McCammon. The dissertation author was the primary investigator and author of this paper.

Chapter 5, in full, is a reprint of the material "Population Shifts and Stochastic Domain Gating in the PBCV-1 mRNA Capping Enzyme" which was published in 2009 in the Journal of the American Chemical Society, volume 131, issue 14, pages 5126 to 5133, by Robert V. Swift, and J. Andrew McCammon. The dissertation author was the primary investigator and author of this paper.

VITA

2003 Associate of Science, Associate of Arts, Santa Barbara City College

2005 Bachelor of Science, University of California, Santa Barbara

2005-2007 Teaching Assistant, Department of Chemistry and Biochemistry, University of California, San Diego

2007 Master of Science, University of California, San Diego

2005-2009 Research Assistant, University of California, San Diego

2009 Doctor of Philosophy, University of California, San Diego

PUBLICATIONS

Amaro R. E.; **Swift R. V.**; McCammon J. A. "Functional and Structural Insights Revealed by Molecular Dynamics Simulations of an Essential RNA Editing Ligase in Trypanosoma brucie." PLoS Journal of Neglected Tropical Disease, Vol. 1, e68, 2007.

Swift R. V.; McCammon J. A. "Catalytically requisite conformational dynamics in the mRNA-capping enzyme probed by targeted molecular dynamics." Biochemistry, Vol. 47, pp. 4102-11, 2008.

Swift R. V.; Durrant J.; Amaro R. E.; McCammon J. A. "Toward Understanding the Conformational Dynamics of RNA Ligation." Biochemistry, Vol. 48, pp. 709-19, 2009.

Swift R. V.; McCammon J. A. "Substrate Induced Population Shifts and Stochastic Gating in the PBCV-1 mRNA Capping Enzyme." Journal of the American Chemical Society, Vol. 131, pp. 5126 – 33, 2009.

Fajer M.; **Swift R. V.**; McCammon J. A. "Using multistate free energy techniques to improve the efficiency of REXAMD." Journal of Computational Chemistry, accepted.

FIELDS OF STUDY

Major Field: Biophysical chemistry

Studies in classical molecular dynamics of biologic macromolecules
Professor J. Andrew McCammon

ABSTRACT OF THE DISSERTATION

Structure-Function Relationships in Enzymes of the Nucleotidyltransferase Superfamily: RNA ligation and mRNA Capping

by

Robert Swift

Doctor of Philosophy in Chemistry

University of California, San Diego, 2009

Professor J. Andrew McCammon, Chair

The nucleotidyltransferase superfamily is composed of ATP and NAD^+ dependent DNA ligases, ATP-dependent RNA ligases and GTP dependent mRNA capping enzymes. Despite a wealth of mutational analysis and structural data, few studies characterize the functional dynamics of these enzymes. Consequently, it is the goal of this dissertation to remedy, at least in part, that short coming. Specifically, we use molecular dynamic methodologies to model dynamic behavior of the ATP-dependent RNA editing ligase, TbREL1, from *Trypanosoma brucei*, the causative agent of African sleeping sickness, and the GTP-dependent mRNA capping enzyme from the *Paramecium bursaria Chlorella* virus, PBCV-1. Chapter two investigates the allosteric and local effects of ATP binding in the active site of TbREL1 and predicts the binding mode of nicked dsRNA substrate by calculating ensemble averaged electrostatic potentials.

Chapter three discusses the functional dynamics of conserved active-site residues that may be important in mediating nicked double stranded RNA ligation, predicts that TbREL1 requires two magnesium ions for optimal catalysis, and provides insight into the experimental observation that TbREL1 requires that the 5´-phosphate nicked strand be RNA. Chapter four describes the use of targeted molecular dynamics to explore the role of conserved residues in mediating a catalytically requisite substrate isomerization event in the PBCV-1 mRNA capping enzyme, discussing the modeled dynamics in context of previously reported mutational analysis. Chapter five examines the impact that the rate and extent of OB domain motion has on various aspects of GTP dependent PBCV-1 mRNA capping enzyme function. Specifically, the rate of OB domain motion is discussed in terms of its effect on the GTP association rate, while the extent of motion provides the context to re-examine a putative mechanism explaining the ability of the enzyme to selectively bind single-stranded mRNA. Furthermore, chapter five discusses the effects of GTP binding in context of the induced fit and population shift models of substrate binding.

Chapter 1

The Nucleotidyltransferase Superfamily and Molecular Dynamics

1.1 The Nucleotidyltransferase Superfamily

The nucleotidyltransferase superfamily is composed of ATP and NAD^+-dependent DNA ligases, ATP-dependent RNA ligases, and GTP dependent mRNA capping enzymes. Each of these enzymes shares unique chemical and structural features that distinguish them from other protein superfamilies[1].

Chemically, each superfamily member is distinguishable by a chemical mechanism initiated by the formation of a nucleotide-enzyme intermediate. Subsequent chemical steps diverge, their exact nature depending upon whether the enzyme is a ligase or an mRNA capping enzyme (fig 1.1). For example, following the formation of the nucleotide-enzyme intermediate, in DNA and RNA ligases nicked polynucleotide binding is followed by nucleoside monophosphate transfer from the enzyme to the 5´-phosphate of the nicked polynucleotide; this links the nucleoside to the 5´-end of the nicked-polynucleotide substrate *via* a phosphodiester bridge, activating it for ligation. During the final step, the nicked 3´-OH attacks the beta phosphate of the phosphodiester bridge,

displacing the nucleoside monophosphate, resulting in polynucleotide ligation. In GTP dependent mRNA capping enzymes, on the other hand, the mechanism is complete upon transfer of the nucleoside monophosphate to the 5´-end of the nascent mRNA.

The chemical homology among superfamily members is mirrored in structural similarity (Fig. 1.2). Each superfamily member shares an N-terminal nucleotidyltransferase domain, which houses a nucleotide binding groove as well as the amino acids responsible for catalysis. The nucleotide binding groove is defined by a set of five conserved motifs (I, III, IIIa, IV, V) arranged in anti-parallel beta sheets connected by short inter-strand loops. Extending from the beta sheet spanned by motif IIIa is a strictly conserved phenylalanine, whose side chain provides extensive pi-stacking interactions with the purine moiety of the nucleotide substrates (Figs. 2.4 & 4.6). Extending from motif IV, opposite the conserved phenylalanine, one of several hydrophobic residues, whose identity depends on the superfamily member, can be found. The phenylalanine and hydrophobic residue stack against the nucleotide purine, forming a hydrophobic sandwich that stabilizes the nucleotide. By minimizing lateral fluctuations of the nucleotide within the active-site, these hydrophobic interactions likely increase the time that the nucleophile and electrophile spend in proximity to one another, thereby improving the reaction rate.

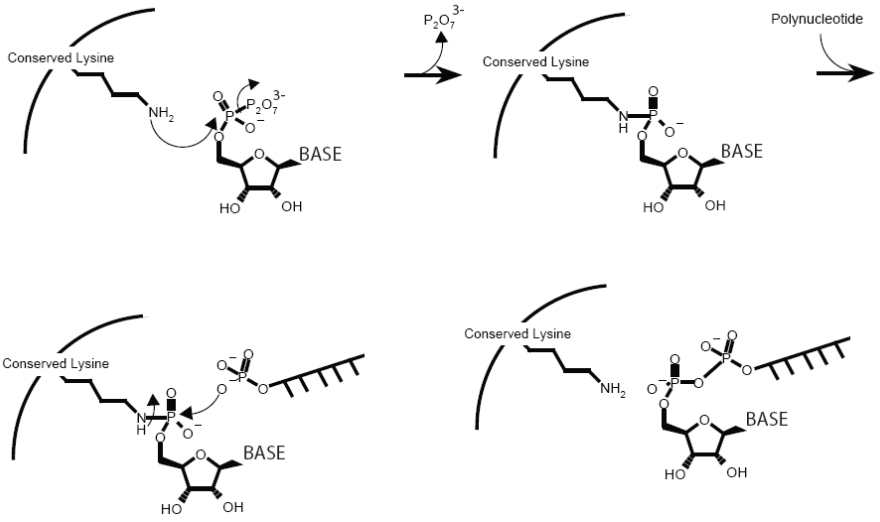


Figure 1.1: Conserved nucleotidyltransferase mechanism. The first two steps of mechanism are displayed. In the case of NAD^+-dependent DNA ligases, $P_2O_7^{3-}$ is replaced with NMN. The third, ligation state is characterized by the attack of 5´-PO_4, displacement of the nucleoside monophosphate and ligation.

Nucleotide specificity determinants are thought to line the deep end of the nucleotide binding groove. Positioned to interact with the exocyclic C6-carbonyl group of guanine, a conserved lysine, found in all GTP-dependent mRNA capping enzymes, is believed to confer GTP specificity. Rather than arise from a strictly conserved interaction, current structural data suggests that ATP specificity may be conferred by interactions that differ depending on enzyme identity. For instance, in the T7 and bacterial ATP-dependent DNA ligases, a conserved glutamate interacts with the exocyclic N6 amine. In the eukaryotic virus that infects the *Paramecium bursaria chlorella* protozoa, this interaction is replaced with the threonine side-chain hydroxyl moiety. In the case of NAD^+ dependent DNA ligases, the nicotinamide group is accommodated by a unique flanking domain, called domain Ia, fused to the nucleotidyltransferase domain[2, 3].

Fusion of different auxiliary, C-terminal domains to the nucleotidyltransferase domain modulates, at least in part, polynucleotide-substrate specificity. ATP-dependent DNA ligases and GTP-dependent mRNA capping enzymes have an oligo-nucleotide/saccharide binding domain (OB domain) fused to their nucleotidylytransferase domains[1, 3, 4]. The OB domain, which consists of a five-stranded antiparallel beta barrel, contains a sixth conserved motif, with an internal RxDK sequence[1, 3, 4]. While these features are present in both DNA ligases and mRNA capping enzymes, the functional roles that they play differ. In DNA ligases, a conserved phenylalanine intercalates the minor groove of nicked dsDNA, distorting nucleotide geometry local to the active site on the 5´-PO_4 side of the nick. In their distorted states, the nucleotides have an A-form geometry, which is believed essential for catalysis. In the mRNA capping enzymes, on the other hand, conserved residues within OB domain motif VI are thought to stabilize the triphosphate substrate prior to catalysis. RNA editing ligases have replaced the OB domain with helical bundles, essential for the transfer of the nucleoside monophosphate to the nicked-terminal 5´-PO_4.

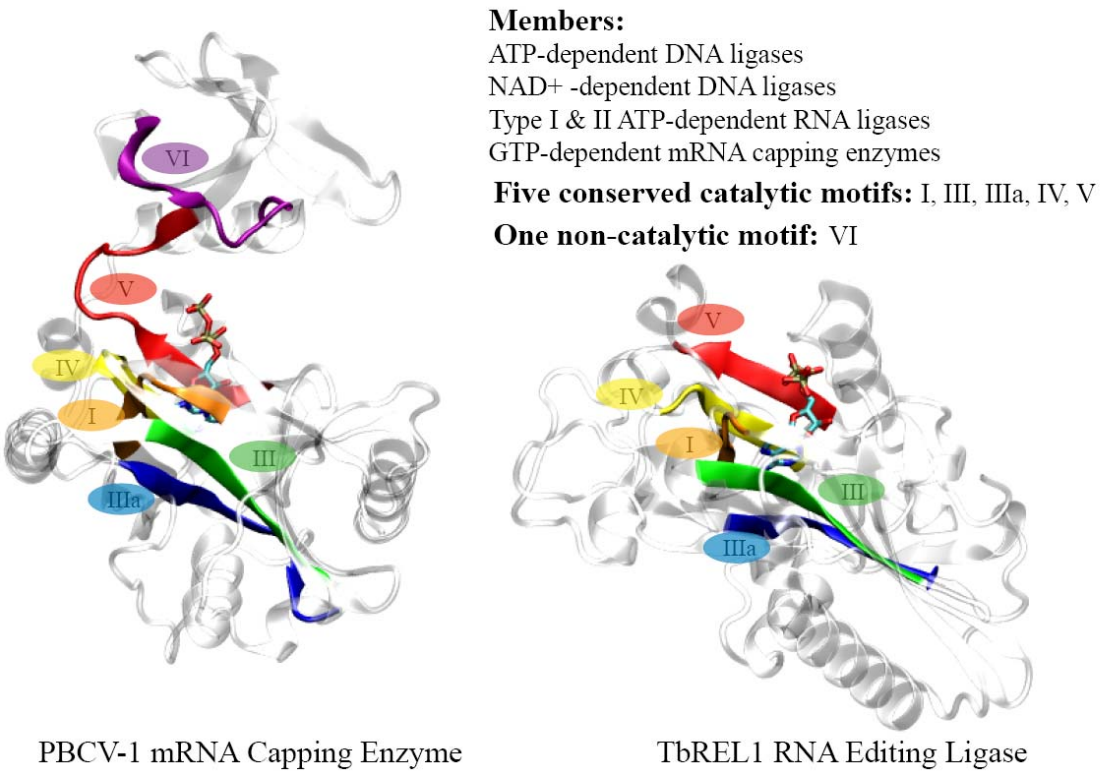


Figure 1.2: **Conserved nucleotidyltransferase motifs**. Members of the nucleotidyltransferase superfamily are listed. The two enzymes that are the principal topic of this dissertation, the PBCV-1 mRNA capping enzyme and the TbREL1 RNA Editing Ligase, are shown in a new cartoon rendering on the left and right, respectively. Five conserved motifs that characterize the active site of all superfamily members are shown by color and labeled 'I' through 'V'. The sixth non-catalytic motif found in all mRNA capping enzymes and ATP-dependent DNA ligase is labeled 'VI'.

While high resolution crystal structures [2, 4-8] and numerous mutational studies [3, 9-12] have been published, to date, only three biophysical studies targeting superfamily members have been undertaken [13-15]. Consequently, it is the primary aim of this dissertation to report the use of molecular dynamics and physical principles to partly fill this deficiency. In particular, we explore and extend previously hypothesized structure-function relationships of two nucleotidyltransferase superfamily members: the PBCV-1 mRNA capping enzyme, a GTP-dependent mRNA capping enzyme from the

Paramecium bursaria chlorella virus; and TbREL1, an ATP-dependent RNA ligase of *Trypanosoma brucei*, the causative agent of African sleeping sickness.

Results may be broken down by chapter: chapter two explores the allosteric influence of ATP binding in TbREL1 dynamics; chapter three uses a homology model to investigate important features of RNA binding and specificity and explores functional dynamics that may be important to step-two of catalysis; chapter four begins discussion of the PBCV-1 mRNA capping enzyme and details possible roles for the RxDK sequence of conserved motif VI during OB domain isomerization, an event thought to mediate rearrangement of the GTP substrate into a catalytically competent conformation; chapter five examines PBCV-1 OB-domain motion in context of the induced fit and population shift models and explores the effects that the extent, and rate, of OB domain motion have on polynucleotide specificity and GTP binding, respectively.

Details of the biological role played by TbREL1 may be found in the introductory sections of chapters two and three, while biologic context for the PBCV-1 mRNA capping enzyme is provided in the introductory sections of chapters four and five. As biologic context for both TbREL1 and the PBCV-1 mRNA capping enzyme is provided elsewhere, it is the goal of the remainder of this introduction to provide a brief overview of the computational technique called molecular dynamics (MD), the principal method of study employed throughout this dissertation.

1.2 Molecular Dynamics

In a perfect world, that is one in which every computational chemists had infinite computing resources at their disposal, all molecules would be treated exactly, using quantum mechanics. In the absence of a time-dependent external field, quantum mechanics requires solving the time-independent Schrödinger equation, $H\Psi = E\Psi$, for E the configuration energy, with H the Hamiltonian operator, and Ψ the state-function of the system, itself a function of all of nuclear and electronic coordinates (the system configuration)[16]. Repeatedly solving the Schrödinger equation for all possible system configurations would yield the corresponding energies. Following solution of the Schrödinger equation, these configuration energies could be used to compute configurational probability densities by using the principles of statistical mechanics[17, 18]. Next, thermodynamic functions, such as the Gibbs and Helmholtz free energies, as well as potentials of mean force[19], which are free energies along a molecular reaction coordinate of interest with respect to the total system free energy, could be calculated from the configurational probability densities. Furthermore, with a potential of mean force in hand, kinetic parameters, which describe rates of change along a given molecular reaction coordinate, may be derived, for example, from numerical solution of the Smoluchowski, or other equations[20, 21]. Thus, in a perfect world, iteratively solving the Schrödinger equation yields both a thermodynamic and kinetic description of a molecular system; and this is all that is needed to provide a mechanistic explanation, at atomic resolution, to molecular phenomena. The world, however, is not perfect. Computational

scientists do not have an infinite pool of computing resources, and the size of biologic macromolecules places them well outside of the scope of currently tractable quantum mechanical calculations. This issue has been resolved, at least partly, with a computational method called molecular dynamics[22, 23].

Molecular dynamics makes the approximation that atoms can be treated as classical, charged spherical particles. These spherical particles interact *via* Columbic and van der Waals forces and are connected by bonds conventionally described by harmonic springs. Similar to the bonds connecting two adjacent atoms, angles between three atoms are assigned a harmonic interaction term, while dihedral angles are treated using periodic potential functions. Because forces can be calculated from potential energies, these interaction potentials are collectively known as the molecular force field. Much careful research has gone into determining proper interaction parameters that closely reproduced experimental observables, and several quality protein force fields are available[24-26]. Realistic water models, explicitly representing each atom of a water molecule, compliment the classical protein force field and reproduce the physical and dynamic properties of a proteins aqueous environment[26, 27] By using the molecular force field to calculate the force acting between every atom within a solvated biomolecular system, and adding to that a random force that duplicates the distribution of thermal energy available at a particular temperature, the atomic motion of the system may be realistically propagated through time in discrete steps. Neglect of electronic degrees of freedom, as well as the simplified form of the force field potential energy function, allows MD simulations to be carried out at tremendous savings relative to quantum mechanical

calculations. With computational cost reduced, functional dynamics may be explored at atomic resolution by saving system configurations at fixed intervals. Furthermore, if it is assumed that the time series of system configurations extracted from the simulation yields the configuration probability density that would be observed at thermal equilibrium, (that is, if the system is assumed to behave ergodically) statistical inference, along with statistical mechanical relations, may be used to yield thermodynamic quantities, thus connecting with experiment.

While molecular dynamics provides an improvement over the computational cost of quantum mechanics, the explicit treatment of all atoms composing the system limits most simulations to lengths of tens to hundreds of nanonseconds. As many interesting biologic events happen on time scales that extend well into the realm of milliseconds[28], the time-scale accessible to molecular dynamics falls short of describing these events by many orders of magnitude. Despite this shortcoming, carefully chosen research questions may be answered using the molecular dynamics method, gaining valuable insight into molecular function at the atomic level (see works within this dissertation, for example). Additionally, improvements in both computer hardware and simulation methodology are continually being made. On the hardware front, for example, the highly parallel architecture of graphical processing units (GPUs) is currently being adapted to efficiently run molecular dynamics software. In the future, GPUs may offer a low cost alternative to the parallel "clusters" of compute nodes that contemporary computational chemistry groups depend on[29]. The spirit that drives hardware advancement also pushes methodological advancement. Many methods that allow aspects of long-time-scale

events to be described have been published. For example, umbrella sampling[30], steered molecular dynamics[31], targeted molecular dynamics, replica-exchange molecular dynamics[32], accelerated molecular dynamics[33], as well as various coupled variations of the preceding (e.g. replica exchange accelerated molecular dynamics[34]), have all been reported in the literature. Taken together, advancements in both hardware and methodology promise to extend time scales accessible to molecular dynamics, paving the way for vibrant and exciting new discoveries in biophysical chemistry.

Chapter 2

Functional and Structural Insights Revealed by Molecular Dynamics Simulations of an Essential RNA Editing Ligase in *Trypanosoma brucei*

Abstract

RNA editing ligase 1 (TbREL1) is required for the survival of both the insect and bloodstream forms of *Trypanosoma brucei*, the parasite responsible for the devastating tropical disease, African sleeping sickness. The type of RNA editing that TbREL1 is involved in is unique to the trypanosomes, and no close human homolog is known to exist. In addition, the high-resolution crystal structure revealed several unique features of the active site, making this enzyme a promising target for structure-based drug design. In this work, two 20 ns atomistic molecular dynamics (MD) simulations are employed to investigate the dynamics of TbREL1, both with and without the ATP substrate present. The flexibility of the active site, dynamics of conserved residues and crystallized water molecules, and the interactions between TbREL1 and the ATP substrate are investigated and discussed in the context of TbREL1's function. Differences in local and global motion upon ATP binding suggest that two peripheral loops, unique to the trypanosomes, may be involved in interdomain signaling events. Notably, a significant structural rearrangement of the enzyme's active site occurs during the apo simulations, opening an

additional cavity adjacent to the ATP binding site that could be exploited in the development of effective inhibitors directed against this protozoan parasite. Finally, ensemble averaged electrostatics calculations over the MD simulations reveal a novel putative RNA binding site, a discovery that has previously eluded scientists. Ultimately, we use the insights gained through the MD simulations to make several predictions and recommendations, which we anticipate will help direct future experimental studies and structure-based drug discovery efforts against this vital enzyme.

2.1 Introduction

The existence and widespread occurrence of several devastating trypanosomal tropical diseases, such as Chagas' disease and African sleeping sickness, cause an estimated 1 million deaths each year in developing countries[35]. In 2005, the completely sequenced genomes of *Trypanosoma brucei*, the causative agent of African sleeping sickness, *T. cruzi*, the causative agent of Chagas disease, and *Leishmania major*, the causative agent of Leishmaniasis, were published, yet, despite these great genomic successes, the need for effective and suitable drugs still remains[36]. Currently available drugs were developed in the first half of the twentieth century and they are toxic, difficult to deliver and often ineffective[37].

The trypanosome pathogens responsible for these diseases all share unique post-transcriptional mRNA editing features, the discovery of which revealed a rich addition to the central dogma of biology, in which information not only passes from DNA to RNA to

protein, but also between different classes of RNA[38]. Through the insertion and deletion of uridylates (U's), the editing process transforms premature mitochondrial RNA (pre-mRNA) to mature mRNA in a multi-protein complex known as the editosome[39, 40]. The exact composition of the editosome complex has yet to be fully characterized, although 20S core complexes have a M_w of 1.6 MDa and appear to be comprised of 16-20 proteins[41]. To further complicate matters, it has recently been demonstrated that at least three different 20S editosomes of heterogeneous composition and distinct specificity are involved in the editing process[42], possibly reflecting compositional changes of this dynamic multicatalyst complex at different stages in the editing process[39].

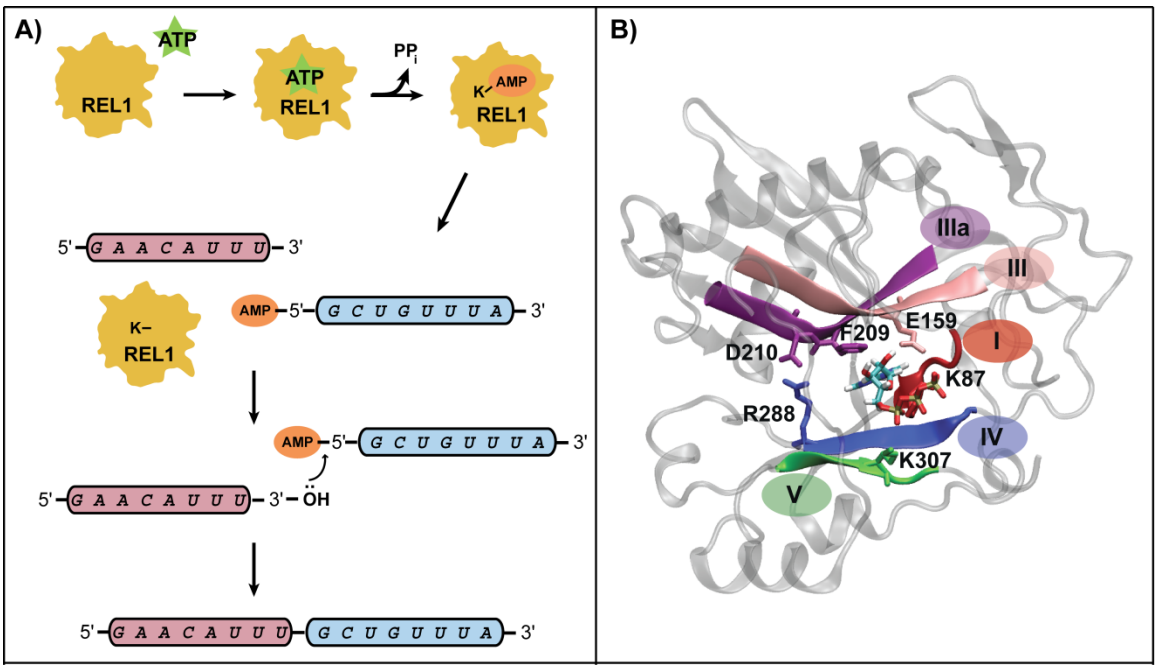


Figure 2.1: RNA editing ligase 1 reaction and structure. A) The three-step ligation reaction catalyzed by RNA editing ligase 1. B) The energy minimized TbREL1 structure is depicted with ATP-bound in the active site. The crystallographic water molecules and explicit solvent are included in the simulations, but omitted in the figure for clarity. The five conserved ligase motifs are colored and labeled, and key conserved residues within each motif are depicted in licorice.

The remarkable mRNA editing process begins in the trypanosomal mitochondrial (mt) DNA, which consists of a topologically linked network of thousands of minicircles and dozens of maxicircles. It is the transcripts of these maxicircles, encoding components of respiratory complexes and energy transduction systems, which undergo extensive RNA editing. The editing process begins when guide RNAs (gRNAs) are transcribed from the minicircles in the mt genome and subsequently base-pair with pre-mRNA sequences through a conserved "anchor sequence"[43, 44]. Endonucleolytic cleavage of the pre-mRNA strand occurs at a point of mismatch between the trans-acting gRNA and its cognate pre-mRNA, by endonucleolytic enzymes that have yet to be characterized. Depending on the type of RNA mismatch, U's are either added, by terminal uridylyl transferase (TUTase), or deleted, by a U-specific 3' exonuclease (ExoUase). The processed RNA fragments are then religated by one of two RNA ligases, Kinetoplastid RNA editing ligase 1 (TbREL1) or RNA editing ligase 2 (TbREL2), generally depending on whether the process is deletion or insertion editing, respectively. Religation of the now completely base-paired double stranded RNA (dsRNA) strands occurs in a three-step process (Fig. 2.1A).

In the first step, the catalytic lysine residue acts in concert with a divalent metal cofactor within the ATP binding pocket of TbREL1 and is autoadenylated, forming a protein-AMP intermediate and releasing pyrophosphate. In the next step, a 5'-5' phosphate linkage is formed when the AMP is transferred to the 5' end of the nicked 3' RNA substrate, which is proximal to the active site. Finally, the ligation process is

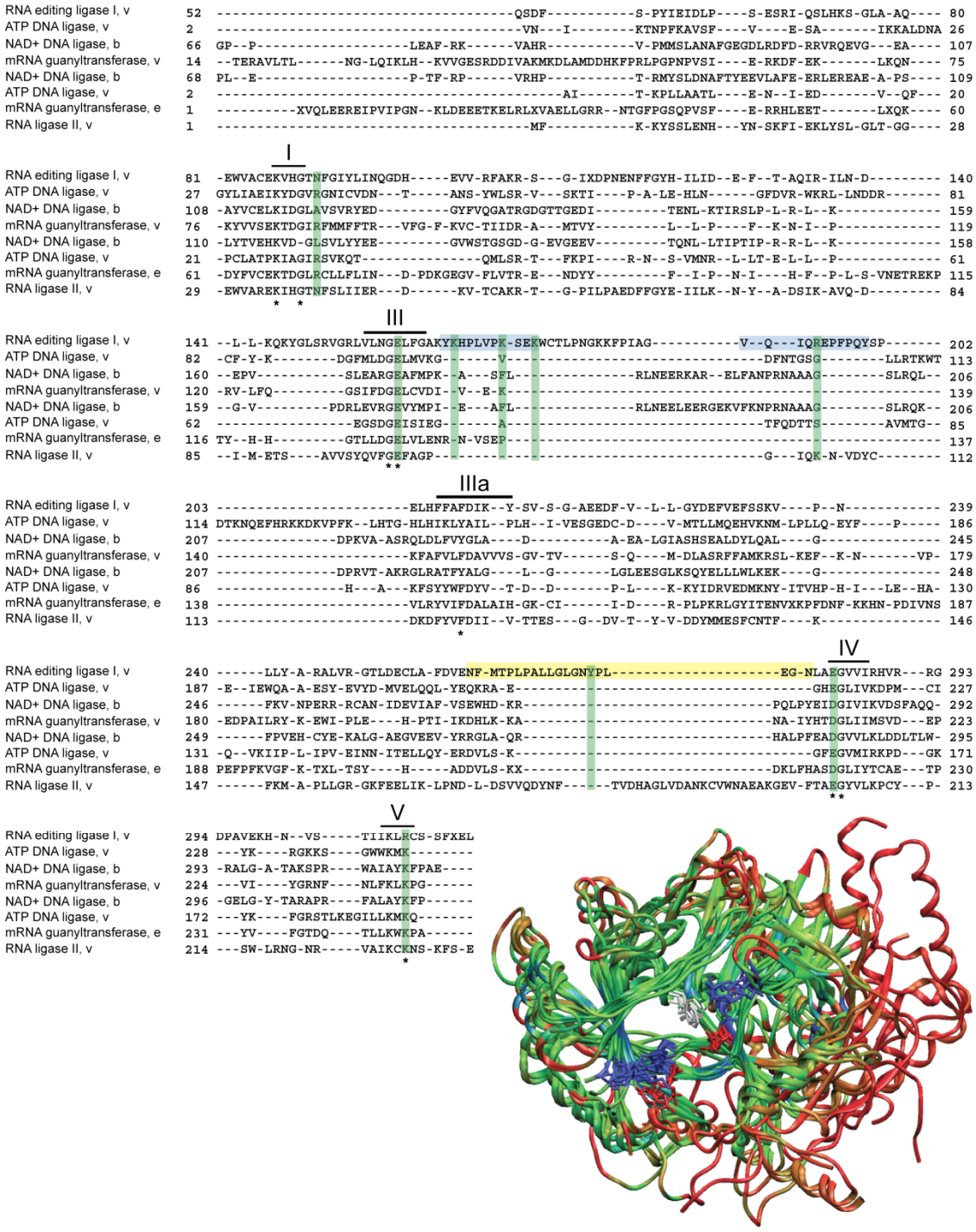
completed when the 3’ hydroxyl group of the other strand displaces the 5’ AMP, resulting in a new phosphodiester bond.

TbREL1 has been shown to play a key role in the vitality of *Trypanosoma brucei*, as it is required for survival of both the insect and bloodstream forms of the pathogen[45, 46]. TbREL1 is comprised of a catalytic N-terminal adenylation domain and a C-terminal domain that facilitates non-covalent interaction with another editosome protein, KREPA2[47]. Interestingly, an oligonucleotide binding (OB-fold) domain, usually associated with DNA ligases and capping enzymes *in cis*, appears to be provided by KREPA2 *in trans* and has been predicted to act as a conformational switch regulating various steps in the editing process[47]. The recently crystallized 1.2 Å resolution structure of the adenylation domain, TbREL1, from *T. brucei* with the bound ATP ligand revealed a number of interesting features of the enzyme, including an unusual water-solvated ATP binding pocket[48]. As the N-terminal domain alone has shown to be capable of RNA ligation activity, further studies characterizing its functional dynamics are relevant. Interestingly, although a single Mg^{2+} ion is clearly coordinated to the protein and ATP in the crystal structure, the protein is not yet adenylated, indicating that the structure is in a pre-catalytic conformation. Furthermore, it has been suggested for this system, and shown experimentally for related systems, that the nucleotidyl transfer reaction proceeds by a two Mg^{2+} mechanism[13]. It has been hypothesized that the first Mg^{2+} binds with high affinity between the ATP β– and γ–phosphate groups, as seen in the TbREL1 crystal structure, and the second magnesium coordinates to a lower-affinity site between the

ATP α-phosphate group and catalytic lysine, subsequently promoting the bond cleavage between the α and β phosphate and the adenylyl transfer[48].

TbREL1 belongs to the covalent nucleotidyl transferase superfamily of enzymes, along with other RNA ligases, mRNA capping enzymes, and DNA ligases[5]. A thorough bioinformatics and phylogenetic analysis of the RNA ligase family shows five well-conserved structural motifs responsible for the three-step nucleic acid repair and strand-joining reaction, a shared overall protein fold, and common evolutionary traces. At the level of the superfamily, the percent identity among the sequences is less than 10%, which renders traditional sequence alignment measures ineffective. However, a structural alignment of an unbiased and nonredundant set of the members of the superfamily, which includes RNA ligase II, mRNA capping enzyme, NAD+ dependent DNA ligase and the ATP dependent DNA ligase, indicates that there is a well-conserved core structure surrounding the nucleotide binding site (Figs. 2.1B & 2.2).

Figure 2.2: Structural alignment of the covalent nucleotidyltransferase superfamily. Structural alignment of the superfamily of enzymes reveals 8 highly conserved residues (marked with asterisks). The enzyme name, domain of life (v – virus, e – eukarya, b – bacteria), and numbering is provided for each member. The five conserved motifs from the RNA ligase family are indicated with lines and roman numerals. The loops predicted to be involved in RNA binding and interdomain signaling are highlighted in blue and yellow, respectively, and the residues suggested for mutation are highlighted in green. The structurally aligned proteins are shown, colored by conservation, with blue and green indicating more conserved regions and red indicating the most variable regions. The eight conserved residues identified in the structural alignment are depicted in licorice and colored by residue type (hydrophobic in white, negatively charge in red, positively charged in blue).


RNA editing ligase I, v
ATP DNA ligase, v
NAD+ DNA ligase, b
mRNA guanyltransferase, v
NAD+ DNA ligase, b
ATP DNA ligase, v
mRNA guanyltransferase, e
RNA ligase II, v
I
III
IIIa
IV
V

Furthermore, this analysis reveals eight highly conserved residues that may play key catalytic roles. The structural differences within the nucleotide binding sites are likely due to the subtle differences in their cofactor specificities. A structural phylogenetic analysis based on the multiple structural alignment of the superfamily indicates that the closest known relative to TbREL1 is T4 phage RNA ligase 2 (Fig. 2.2), and this is consistent with a previous sequence-based evolutionary analysis of the family[5]. Steady-state and pre-steady-state kinetic analysis coupled with strategic mutagenesis of the T4 RNA ligase 2 has established functional roles for the highly conserved binding site residues and mapped many of the important interactions of the RNA ligase active site[12, 49]. The established similarity between these two enzymes is important, as much of this information for the T4 phage system can be used to help interpret, understand, and direct strategic studies for the enzymatic activity of TbREL1.

The existence of a high-resolution ligand-bound crystal structure and unique active site features, coupled with the fact that there is no close human homolog, make TbREL1 an important target for development of inhibitors against these protozoan parasites. In this work, we use all-atom explicit solvent molecular dynamics (MD) simulations to probe the structure, function, and dynamics of TbREL1 on the nanosecond timescale. A comprehensive structural and sequence alignment of all the known superfamily members identifies several key residues that we monitor throughout the simulations. Deeply buried water molecules within the nucleotide binding site and their effect on the mode of ATP-binding are also investigated. A comparison of the principal components for the apo and ATP-bound systems illustrates large local and global

differences in the enzyme motion. Ensemble-averaged electrostatics calculations from the MD simulations reveal a putative RNA binding site near the ligase active site, a finding that has previously eluded scientists. Coupled with the crystal structure and currently available experimental information, the dynamics and structural analysis presented here will likely prove to be salient aspects of future successful drug design efforts against this vital enzyme.

2.2 Materials and Methods

2.2.1 Molecular Dynamics Simulations

Structural coordinates were taken from PDB 1XDN[48]. TbREL1 has a fifty residue N-terminal segment that is believed to be a mitochondrial import signal, which was cleaved before crystallization. The simulated protein contains residues 52-365 of the 469 total residues, thus representing the N-terminal domain. All crystallographically resolved water molecules were retained in the simulations.

In order to prepare the apo-system, the ATP and single Mg^{2+} ion were removed from the active site and replaced by 6 waters using the program Dowser[50]. The selenomethionines used for crystal structure refinement were replaced with methionines. Histidine protonation states were determined using the WHATIF program and manually double-checked. All other hydrogens were added according to the Charmm27 topology

parameters[25] using PSFGEN within NAMD2.5[18]. The protein was immersed in a rectangular box of TIP3P waters[26] providing a 10 Å buffer from the protein to the periodic boundary in each direction. Four sodium atoms were randomly placed at least 5 Å away from the protein in order to neutralize the system's charge. The apo system is comprised of a total of 35,464 atoms.

In order to remove spurious contacts, a set of 26,000 energy minimization steps were carried out. The first 6,000 steps were performed in three 2,000 step cycles. Hydrogen was relaxed during the first 2000 steps, holding all other atoms fixed. Hydrogen, water and ions were relaxed during the next 2000 steps. In the last cycle, the protein backbone was fixed, minimizing all other atoms. No constraints were applied during the last 20,000 steps, freely minimizing all atoms.

Molecular dynamics simulations were carried out for twenty nanoseconds with no constraints and a 1 fs timestep at 1 atm pressure and a temperature of 298.15 K. The temperature bath was maintained by Langevin dynamics while pressure was maintained with the hybrid Nose Hoover - Langevin piston method[50], using period and decay times of 100 and 50 fs, respectively. The Particle Mesh Ewald algorithm was used to treat long-range electrostatics without a cutoff[51]. A multiple time-stepping algorithm was employed, where bonded interactions were computed at every time step, short-range non-bonded interactions were computed every 2 time steps, and full electrostatics were computed every 4 time steps. All minimization and molecular dynamics were carried out using NAMD 2.5. Simulations were performed on our own local cluster, as well as the

San Diego Supercomputer Center's Datastar machine and the National Center for Supercomputing Applications' Cobalt machine. A typical benchmark for the 35,000 atom system on 64 processors on NCSA's Altix platform is 0.12 days per nanosecond of simulation. System configurations were sampled every 500 fs, generating 40,000 coordinate snapshots for subsequent analysis.

The ligand-bound system, with ATP and the single magnesium ion present, was prepared in an identical fashion, except that the crystallized ATP and Mg^{2+} ion were included. An additional 20 ns simulation was performed on the ligand-bound system with identical simulation parameters as described above.

Trajectory analysis was performed with VMD, Matlab, and customized scripts. Ensemble averaged electrostatics calculations were performed in VMD with the PME Electrostatics Plugin. Snapshots every 5 ps over the course of the 20 ns simulation (4000 snapshots) were used in the calculations. All images were created using VMD.

2.3 Results and Discussion

2.3.1 Comparative Dynamics of the Apo and ATP-bound Systems

Two 20 nanosecond simulations were carried out on the apo and ATP-bound TbREL1 systems in order to investigate their stability and dynamical properties. Both

systems were simulated without constraints and reached equilibrium after approximately 7.0 ns of constant pressure and temperature equilibration at 1 atm and 298 K. Correspondingly, the trajectory time from 0 – 7 ns is referred to as the "equilibration phase" and from 7 – 20 ns is referred to as the "dynamics phase" (Fig 2.3).

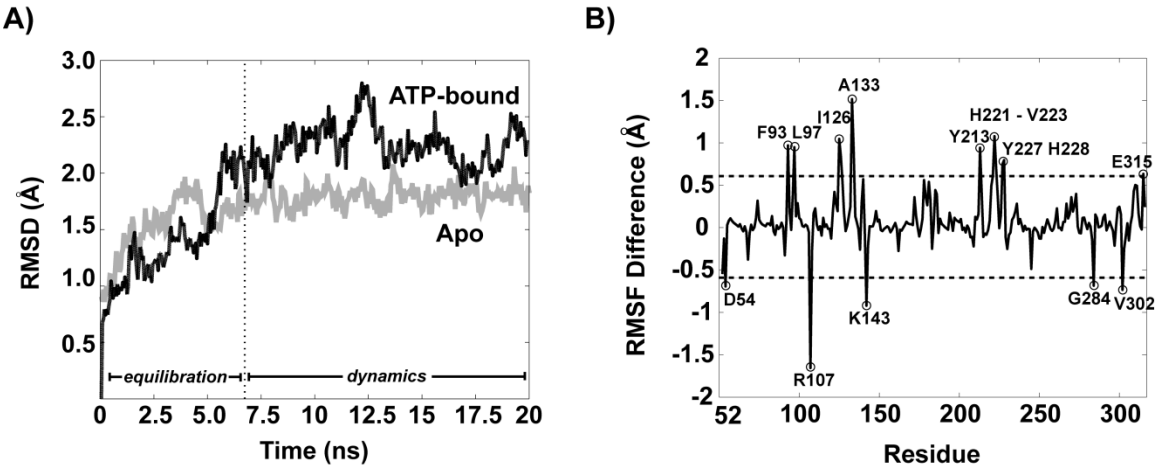


Figure 2.3: RMSD and RMSF difference analysis for apo and ATP-bound systems. A) The overall RMSD of the protein backbone over the 20 ns simulation is shown for the apo (grey) and ATP-bound (black) systems. B) The average C_α RMSF per residue over the dynamics phase for all residues in the apo system is subtracted from the average RMSF per residue for the ATP-bound system, illustrating the relative effect that the presence of ATP has on the amount of fluctuation variance exhibited per residue.

Plots of the time evolution of the root-mean-square-deviation (RMSD), where each trajectory frame was aligned to the initial starting structure in order to remove any rotational or translational motion, indicate that equilibration was achieved for both systems (Fig. 2.3A). The overall root-mean-square-fluctuations (RMSF) per residue was calculated during the dynamics phase for each system based on C_α positions after alignment to the average equilibrated structure, and the resulting RMSF difference plot shows the differences in RMS fluctuations between the two systems (Fig. 2.3B). Our

results indicate that the presence of ATP appreciably affects the system dynamics, providing stabilization to some regions of the protein structure, while causing increased fluctuations in other areas, relative to the apo-system. Notably, the peripheral loop comprised of residues 213 to 223 exhibits marked increases in flexibility when ATP is bound. The residues with the highest destabilization upon ATP binding are found in the middle of helix-2, A133 and Q134. The increase in fluctuation is due to a significant sidechain and backbone conformational transition of Q134 at 7 ns that allows the amine group of Q134 to change hydrogen bonding partners from the backbone carbonyl of N239 to that of E130. Not surprisingly, the RMSF of adjacent A133 is also affected. Among the most stabilized residues are R107 and D54. The formation of a salt bridge between R153 and D54 at 7 ns prevents excessive motion of the D54 sidechain, which would otherwise make hydrogen bonds to solvent water molecules.

2.3.2 Motif Motion

To investigate the differences in local and global motion between the apo and ligand-bound systems, mean RMSF values over each of the five conserved motifs (or residues subsets within each motif) flanking the ATP binding site are reported in Table 2.1. Motifs I and III exhibit nearly the same RMSF fluctuations for both the apo and ATP-bound states, whereas motifs IIIa and IV are stabilized by ATP and motif V shows increased fluctuations.

Table 2.1: **Comparative dynamics of the five conserved motifs and the conserved peripheral loop**. Mean RMSF (in angstroms) over the dynamics phase (7-20 ns) of residues within motifs I-V and the conserved peripheral loop, for the apo and ATP-bound system.

Motif	Residues	Apo RMSF	ATP-bound RMSF
I	K87 – G90	0.52	0.53
III	V155 – G162	0.79	0.75
IIIa	F207 – F209	0.76	0.62
IV	E283 – I287	0.85	0.74
V	I305 – R309	0.42	0.53
	F262 – A282	1.58	2.11

Motif IIIa is one of two motifs exhibiting less motion in the ATP-bound system (Table 2.1). This decrease in RMSD is associated with two stabilizing interactions in motif IIIa: the D210-R288 salt bridge and a pi-stacking interaction between F209 and the adenine ring of ATP (Fig. 2.4).

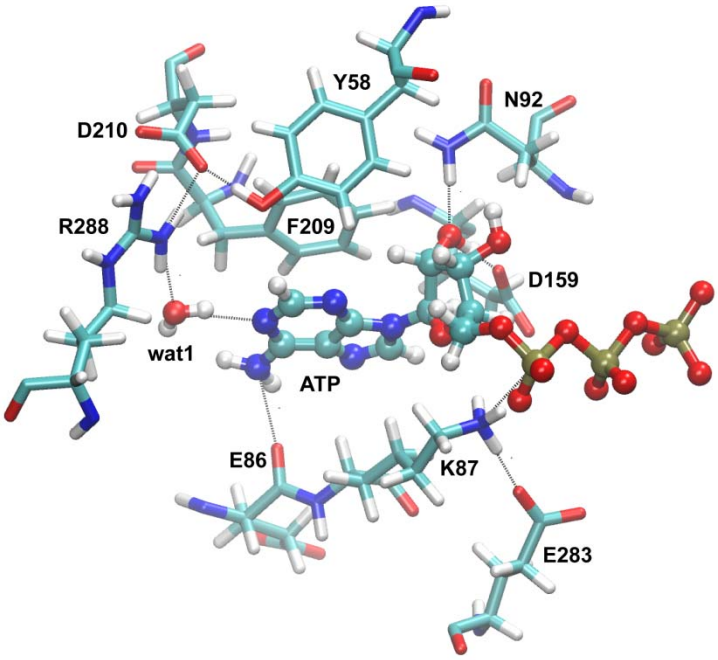


Figure 2.4: Protein-ATP interactions. Interactions between the single water molecule, protein, and ATP, which persist throughout the ATP-bound simulation, are shown.

Throughout the 20 ns ATP-bound system simulation, both of these contacts remain intact (Tables 2.2 & 2.3). Conversely, in the apo system simulation, both of these stabilizing contacts are disrupted (Table 2.3). The D210-R288 salt bridge is disrupted at 5 ns, when D210 forms a salt bridge with nearby R292. Furthermore, in the apo simulation, with no ligand present to stabilize it, F209 undergoes a large swinging motion, eventually resulting in a rearrangement of the active site cavity (Fig. 2.5A).

Motif IV is also stabilized in the ATP-bound system. Significant stabilization of G283 and E284 is the result of an electrostatic network between E284, nearby charged residues K87 and R309, and the polyphosphate tail of ATP.

Motif V is the only motif that is significantly destabilized in the ATP-bound system. The motif is destabilized as L308 and R309 undergo significant C alpha, beta, and gamma rearrangements as R309 twists about its C_β, C_γ, and C_δ bonds to optimize its electrostatic interactions with neighboring E284 and the polyphosphate tail of ATP. The motion of R309 is mainly constrained to its alkyl chain and not its charged end, which is pinned between E284 and the β-phosphate group of ATP (Table 2.3).

2.3.3 TbREL1-ATP Interaction Analysis

An analysis of the interactions between the bound ATP and conserved TbREL1 residues over the course of the MD simulation can be used to gain insight into the

complex dynamics, as well as to predict solution state behavior of the enzyme-substrate complex. Here we describe the conserved ATP-protein interactions, which can be broadly categorized into three groups based on their proximity to the ATP adenosine moiety, ribose, or triphosphate tail.

Table 2.2: ATP-TbREL1 interaction analysis. The interactions between conserved residues in TbREL1 and ATP are reported for the 20 ns simulation (Fig. 2.4 for structural depiction). For the F209 – Adenosine pi-stacking interaction, the distance between the centroid of each of the rings is reported. The mean over the simulation and the standard deviation (in parentheses) are reported, as well as the distance measured in the crystal structure.

Motif	Protein	ATP	Mean (Std Dev)	Crystal
I	K87:NZ	PA	4.04 (0.09)	3.66
III	E159:OE	O2´	2.63 (0.12)	2.66
IIIa	F209	Adenosine	3.87 (0.23)	3.77
IV	R309:NH2	O1B	3.65 (0.65)	2.98
V	K307:NZ	O2A	2.70 (0.15)	3.52
	E86:O	N6	2.88 (0.16)	2.91
	R111:NH2	O3G	4.93 (1.03)	2.81
	N92:ND2	O2´	3.09 (0.26)	2.90

Table 2.3: Protein interaction analysis. Interaction distances are reported between the atoms specified in the first two columns (see Fig. 2.4 for structural depiction). The mean interaction distance and standard deviation is determined over the 20 ns simulation. The final column is the interaction distance reported in the crystal structure.

Protein	Protein	Apo Mean (Std Dev)	ATP-bound Mean (Std Dev)	Crystal
D210:OD2	R288:NH2	5.52 (0.94)	3.22 (0.35)	3.56
D210:OD2	R292:NH1	3.10 (0.45)	3.75 (0.72)	2.65
Y58:OH	D210:OD2*	5.56 (2.49)	3.05 (0.75)	3.95
Y58:OH	R288:NH1	10.00 (2.52)	3.78 (0.30)	4.14

The adenosine moiety of ATP sits deep within the binding site; residues that interact with the adenosine moiety include E86 and F209. The carbonyl oxygen of E86 accepts a hydrogen bond from ATP's amine and helps lock ATP into position, preventing it from translating fore and aft in the binding pocket. The phenyl group of F209 provides important stabilizing pi-pi interactions with the adenosine moiety, further attenuating lateral movement of ATP within the active site. Both the F209 and the E86 interactions have relatively low standard deviations and average values close to those observed in the crystal structure (Table 2.3) demonstrating that these interactions are among the most stable in the protein-substrate complex.

Two conserved residues interacting with the ribose moiety are E159, N92. This pair of residues forms a unique hydrogen bonding pattern with the ribose O2' that was conserved throughout the 20 ns simulation (Table 2.3). The ribose O2' hydroxyl group donates a hydrogen bond to the carboxyl group of E159, while the amine group of N92 donates a hydrogen bond to the ribose O2' (Fig. 2.4). The slight deviation between the bond distances observed in the crystal structure and the average values observed during the simulation is attributable to conformational fluctuations within the ribose ring from the slightly more compact C-3 and C-2 endo conformations to the more extended envelope conformation.

A number of charged residues near the periphery of the binding site, including R309, K307, K87, R111, interact with the polyphosphate of ATP. K307 and R111 primarily stabilize the conserved triphosphate conformation observed over the course of

the 20 ns simulation. The interaction distance between K307 and the O2A of ATP is 0.82 Å shorter than that observed in the crystal structure. This is attributable to rotation about the O5'-P_α bond that brings the O2A closer to the K307 amine. The low standard deviation exhibited throughout the trajectory illustrates the importance of this interaction in maintaining the triphosphate conformation (Table 2.2).

While it is not immediately apparent from the 4.9 Å average R111 NH – O3G interaction distance, R111 also contributes important triphosphate longer-range stabilizing interactions. For the first 1.2 ns, R111 samples conformations local to that found in the crystal structure, with the R111 NH2 group proximal to O3G. Subsequently, R111 rearranges by rotating through the CD-NE bond whereby the R111 NH1 group donates a hydrogen bond to triphosphates bridging O2B while the NH2 group maintains a hydrogen bond with O3G. Following this rearrangement, a rotation about the NE-CZ bond occurs, which exchanges NH1 and NH2 hydrogen bonding partners while preserving the hydrogen bonding pattern. This second rotation moves the NH2 group 5.2 Å from O3G, and because it occurs frequently, is responsible for the high standard deviation and longer average interaction distance (Table 2.2). The K87 electrophile remains at an unreactive 4.0 Å average from the alpha phosphate nucleophile. This agrees with crystal structure data from other superfamily members, which suggest that the triphosphate tail must undergo a conformational rearrangement to properly position the alpha phosphate for catalysis[1]. Because of its relatively distal position and larger standard deviation, we predict R309 plays a secondary role in triphosphate stabilization.

2.3.4 Dynamics of Conserved Residues within the Binding Site

The interactions of several conserved residues within the ATP binding site exhibit altered dynamics in the apo and ATP-bound systems (Table 2.3). Here we provide insight into the stability of the protein complex by examining how these protein interactions are affected by the presence of ATP in the binding pocket.

In the ATP-bound state, D210, R288 and Y58 form a tight hydrogen bonding and electrostatic network at the deep end of the active site (Fig. 2.4). D210 interacts with R292, but because it is already involved in a stable network of three hydrogen bonds, it makes a poor binding partner. As a result R292 undergoes large fluctuations relative to those observed in the D210, R288, Y58 triad (Table 2.3). A significantly different behavior is seen in the absence of ATP. In the absence of the pi-pi stacking interactions provided by the adenosine moiety, F209 swings inward toward the active site cavity, disrupting the D210, R288, Y58 triad. Once the network of interactions is destabilized, thermal fluctuations cause R288 to swing away from D210 causing a 1.96 Å increase in average bond distance relative to that observed in the crystal structure (Table 2.3). D210 subsequently forms a salt bridge with R292 whose average is 0.83 Å shorter than that observed in the crystal structure. This new interaction pattern is conserved throughout more significant active site rearrangements that were observed later in the simulations, as discussed in greater detail below.

2.3.5 ATP-binding Induces Fluctuations at Remote Sites

TbREL1 has several important loop regions on the periphery of the ATP binding site that may be functionally relevant. One region in particular, a helix-loop segment formed by residues F262-A282, is highly conserved among all trypanosomatids and has been hypothesized to be involved in protein-protein association within the editosome complex (Fig. 2.2)[48]. Comparing the backbone RMSF values for the apo and ATP-bound system reveals that ATP binding induces a significant increase in motion within this conserved helix-loop region (Table 2.1). As this region is not directly linked to the ATP binding site, the propagation of this motion must be through cooperative interactions. This is further supported by a principal components analysis (Supplementary Information), which shows that motion in this helix-loop region is primarily accounted for in the most dominant principal components for the ATP bound system, and that the motion is not present in the apo system (Fig. 2.8). These results suggest that these loop regions may play an important role in interdomain signaling or crosstalk upon ATP binding.

2.3.6 Active Site Rearrangement in Apo System

During the apo system simulation, a significant rearrangement of the active site cavity is observed. In the absence of ATP, the active site pocket is initially solvated with water molecules. After 6.5 ns, F209 swings out of the binding pocket towards Y58, making favorable hydrophobic contacts with the sidechain of I305. This structural

reorganization forms a hydrophobic barrier to the deep end of the active site, occluding penetration of solvent water molecules (Fig. 2.5).

A radius of gyration analysis for motifs IIIa and IV indicates that in the absence of ATP the formation of the hydrophobic barrier leads to a slight contraction of the distance between motifs IIIa and IV that comprise the deep end of the binding pocket (Fig. 2.9). Further rearrangement occurs after 12.5 ns when Y58 swings away from motif IIIa, towards the bulk solvent. Over the course of the simulation, the distance between motif IIIa and the strand housing residues 58 to 61 systematically increases from 7 Å initially to almost 13 Å, as measured by the distance between the alpha carbons of Y58 and R111.

These results suggest a strong induced fit effect may occur upon ATP binding. The structural reorganization and increased overall flexibility of TbREL1 without ATP-bound (Table 2.3) may explain the difficulty in crystallizing the apo structure experimentally (personal communication, J. Deng). In addition, the rearrangements observed in the simulations reveal an altered topography near the ATP binding site, which could potentially be exploited in a structure-based inhibitor design scheme in which compounds could be designed that stabilize this inactive conformation. The strategy of targeting a unique inactive conformation was successfully accomplished with Abelson tyrosine kinase and the inhibitory compound Gleevac, which locks the kinase in an inactive conformation with high specificity and affinity[52, 53].

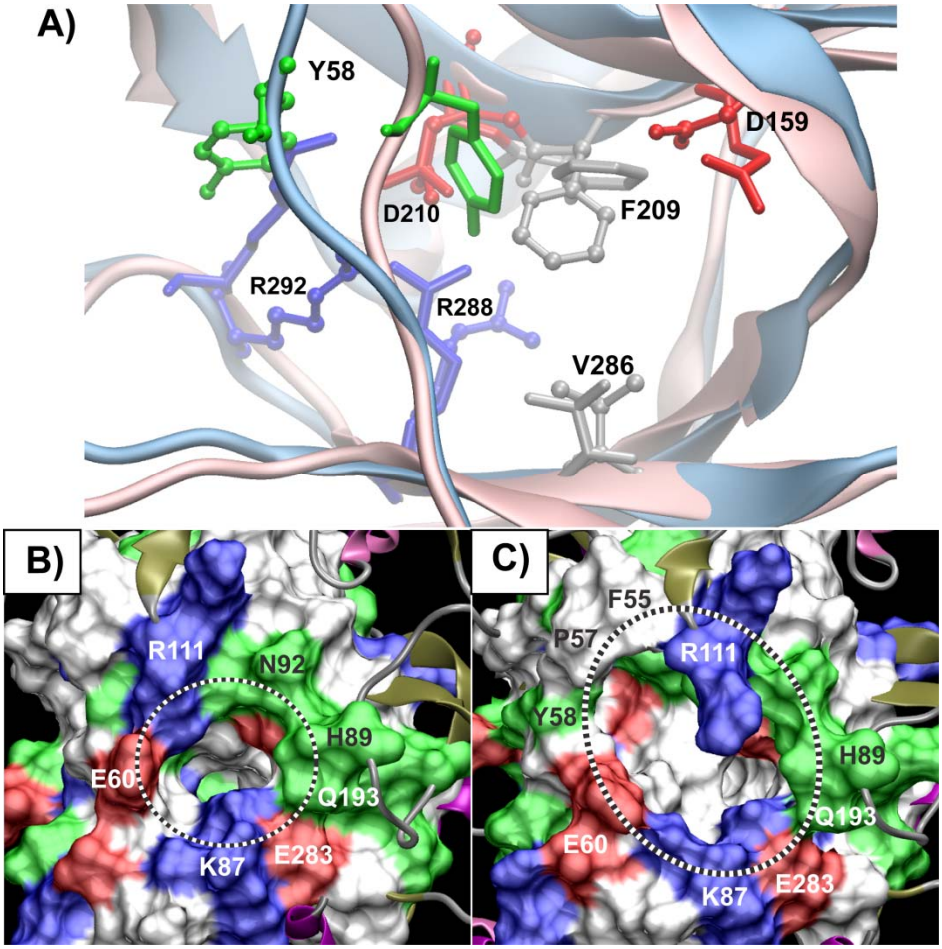


Figure 2.5: ATP binding pocket undergoes significant rearrangement in the apo state simulations. In all panels, residues are colored by residue type: red – acidic, blue – basic, white or gray – hydrophobic, green – polar. A) A structural alignment of a representative structure from the apo simulation (backbone shown in blue cartoon, residues shown in ball and stick) and the ATP-bound simulation (pink cartoon, residues shown stick-only licorice) shows the relative reorientation of key active site residues. B) MSMS surface representation[54] of residues within a 12 Å radius of the ATP binding pocket at the beginning of the apo trajectory. Residues on the periphery are labeled, and the ATP binding pocket is highlighted with a dotted line. C) Same as in (B), but after 20 ns of simulation (structure is same as the blue structure in panel (A)). An additional cavity opens between R111 and E60, exposing several new polar and hydrophobic residues.

2.3.7 Role of Water Molecules in the Binding Site

The TbREL1 crystal structure revealed three deeply buried water molecules coordinating interactions between conserved active site residues and ATP[48]. The simulations presented here include all the crystallized water molecules, allowing us to monitor their behavior over the course of the 20 ns trajectory. For continuity between the crystal structure and this work, we adopt the water molecule naming convention as defined in Ref. Deng et al, 2004. These deeply buried water molecules are a unique feature of TbREL1, therefore it may be advantageous to consider them in the inhibitor design process. In particular, insights into their exchange rates and structural features are of interest.

In the ATP-bound system, wat1, which forms bridging hydrogen bonds between conserved R288 and the N1-adenine group of ATP (Fig. 2.4), remains fairly rigid in its position throughout the equilibration phase. Wat2, which initially forms a hydrogen bond with the backbone carbonyl of conserved V286, immediately changes position to form a hydrogen bond network with wat1, R288 and D210. Once this initial rearrangement occurs, the new configuration remains for the duration of the equilibration phase.

During the dynamics phase of the ATP-bound system, two of the three deeply buried water molecules exchange with bulk solvent. Wat3, which is initially coordinated to wat 1, the backbone carbonyl of F209, and the charged group of R288, is the first of the buried water molecules to exchange. Wat1, the single water molecule that interacts

directly with the adenine ring of ATP, is the only buried water molecule that does not exchange. The exchange of these water molecules and subsequent rearrangement of conserved active site residues may indicate that they help modulate the plasticity of the ATP binding site and that it may be possible to advantageously displace one or more of these water molecules with inhibitor functional groups. Furthermore, during the dynamics phase of the ATP-bound system, additional water molecules gain access to the ATP-binding pocket and lubricate interactions between the ligand and the protein. These additional water molecules contribute to the expansion of the active site cavity as seen in the radius of gyration analysis (Fig. 2.9).

2.3.8 Ensemble Averaged Electrostatics Calculations

To date, the RNA binding region of TbREL1 has eluded scientists. Despite the high resolution crystal structure, a surface electrostatics calculation performed with DelPhi on the crystal structure did not reveal any large positive patches on the surface of the enzyme where RNA binding might occur[48].

To further investigate the electrostatics of TbREL1, we calculated the electrostatic potential around the ATP-bound protein using ensemble averaging over the 20 ns trajectory. This analysis reveals a large positive electrostatic lobe near the ligase active site, clearly indicating where RNA could favorably bind (Fig. 2.6A). Furthermore, a structural alignment of TbREL1 with the related human DNA ligase shows nearly perfect

overlap of its DNA-binding region with the putative RNA binding region predicted here (Fig. 2.6B).

Within the center of the putative RNA binding site are two peripheral loops that are unique to the trypanosomes (Y165 - K175 and V190 – Y200). These loops had previously been speculated to be involved in RNA recognition due to the presence of a positively charged residue within each, K166 and R194, respectively[48]. The ensemble averaged electrostatic potential maps presented here provide compelling evidence that these loops may be involved in RNA binding, and they are particularly encouraging in light of a recent study that showed extensive MD simulations can accurately reproduce experimentally observed changes in protein electrostatic fields[55]. Interestingly, these putative RNA binding loops are absent in the N-terminal domain of homologous T4 RNA ligase 2, which, in the absence of its C-terminal domain, is not able to catalyze the ligation reaction[49], unlike TbREL1, which is still active in the isolated N-terminal domain[48]. The large positive electrostatic potentials substantiate the idea that TbREL1 is equipped with its own RNA binding motifs, which allows it to retain activity even in the absence of a C-terminal OB-fold domain.

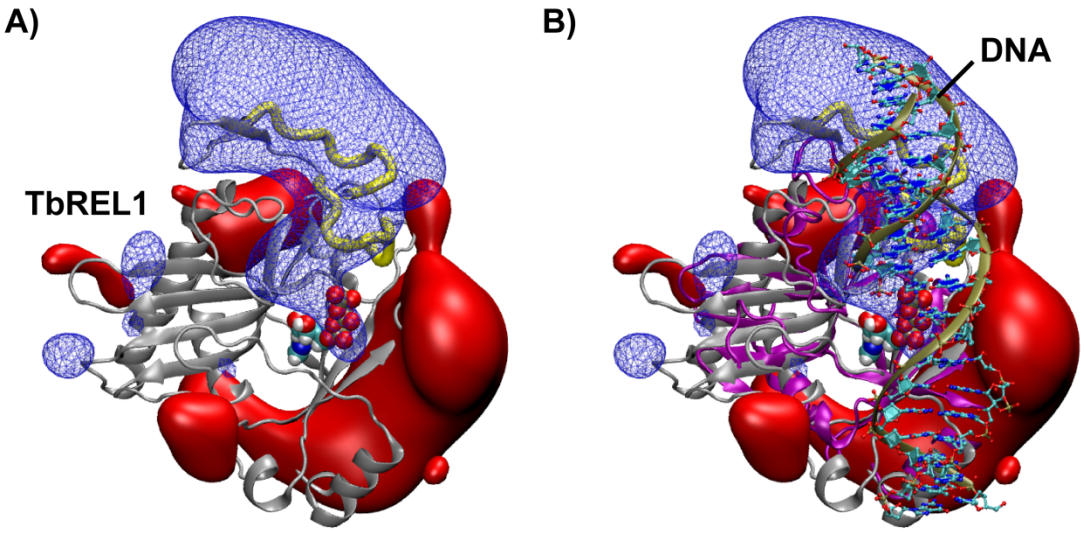


Figure 2.6: Ensemble averaged electrostatics for TbREL1 predict RNA binding site. Negative electrostatic isosurface is shown in red solid, positive isosurface in blue wireframe; isosurfaces are drawn at -70 e/kT and +70 e/kT. A) TbREL1 is shown in silver cartoon with ATP-bound in the active site (depicted in space-filling). Two unique loops in TbREL1 are shown in yellow tube. B) TbREL1 superimposed with the crystal structure of the homologous human DNA ligase (depicted in purple cartoon) shows that the positive electrostatic lobes of TbREL1 overlap well with the location of DNA co-crystallized with human DNA ligase.

2.4 Conclusions and Experimental Predictions

The results presented here allow us to make several predictions that may help motivate future experiments on this vital enzyme. As no single experimental or theoretical approach will answer all questions regarding enzyme function, complex association, and its role in the vitality of the cell, we suggest a variety of approaches to probe the function of TbREL1 and its interactions with the multicomponent editosome complex.

The persistent hydrogen bond interactions formed between N92 and E159 and the ribose groups of ATP (Table 2.2, Fig. 2.4) indicate that these residues may play an important role in ATP binding. Mutagenesis studies of the equivalent of N92 in homologous T4 phage RNA ligase 2 substantiates this hypothesis[12]. We predict that mutation of either or both of these residues will decrease ATP binding, most likely through an increase of the K_m of ATP for this enzyme. As a negative control, we predict that mutation of nonconserved residues far away from the ATP binding pocket should not affect ATP binding, such as N239A or I98A.

The adenylation of TbREL1 is a key component of the overall reaction catalyzed by this enzyme. K87 has been shown to be the adenylated active site residue, as even the charge-conservative mutation K87R abolishes ligation activity[56]. The triphosphate tail was crystallized in an unreactive conformation relative to K87 that persists throughout the duration of the trajectory (Table 2.2), which suggests that the second magnesium, the C-terminal domain, RNA substrate, or some other component of the editosome complex needs to bind before the TbREL1 active site will adopt a catalytic conformation. Repositioning of the polyphosphate tail is necessary in order to bring the alpha-phosphate proximal to the K87 electrophile and in-line for attack. This is also suggested by previously reported structural data on other superfamily members[1]. We predict that nearby R309 plays a secondary role in triphosphate stabilization and that mutation of this residue will reduce the rate of catalysis but not abolish it. The interactions between the non-bridging beta and gamma triphosphate oxygens and the charged end of R309 may help stabilize the polyphosphate tail for the in-line attack. Furthermore, E283, one of

only eight residues conserved in both structure and sequence for the entire superfamily, is likely the catalytic base in the ligation reaction. It is the only negatively charged residue within reasonable proximity of the catalytic K87, and it remains oriented in a position that would promote catalysis throughout the simulations (Table 2.2, Fig. 2.4). This is again substantiated by mutagenesis studies of the equivalent conserved glutamic acid residue in the closely related T4 phage RNA ligase 2[12].

Based on the RMSF (Table 2.1) and principal components analysis (Fig. 2.8), we predict that the conserved peripheral loop comprised of residues F262-A282 may play an important role in interdomain signaling upon ATP binding. It may be possible to test this by engineering a fluorescent probe (e.g. Y275W) in this loop region and then testing whether the fluorescence signal is quenched with the addition of various components of the editosome complex (TUTase, KREPA2, etc), which would indicate that the region is buried and in contact with another editosome protein.

Our simulations and analysis also suggest that the pi-stacking interactions between conserved F209 and the adenosine ring of ATP are of critical importance to stabilizing TbREL1 tertiary structure. Without the ring stacking interactions provided by ATP, the apo protein was greatly destabilized in the vicinity of F209 (Fig. 2.5). We predict that inhibitors that are competitive with ATP will need to exploit this moiety, providing stabilizing interactions to the core protein structure, either through similar pi-stacking interactions or general hydrophobic interactions. In addition, the exchange of

water molecules at the deep end of the ATP binding pocket may indicate that they could be replaced with inhibitor functional groups.

Ensemble averaged electrostatics calculations have allowed us to predict the putative RNA binding region for TbREL1 (Fig. 2.6). Two unique loops, comprised of residues Y165 - K175 and V190 – Y200, are predicted to be near the center of the RNA binding region. We propose that mutation of the positively charged groups in this region, such as K166, K172, K175, or R194, to either neutral groups or negatively charged groups, should alter the electrostatic potential, disrupt the rate of association with the RNA, and reduce the overall rate of catalysis. After the adenylation reaction takes place, the dissociation of the negatively charged pyrophosphate would further increase the positive potential generated in this region.

RNA editing ligase 1 from *T. brucei* is a promising drug target against African sleeping sickness, and it shares close homology to proteins from the parasites responsible for both Chagas' disease and Leishmaniases, which further underscores its importance to global health. It also presents an ideal system to deepen our understanding of the fundamental and fascinating process of RNA editing. Advancing the knowledge of enzymes and biochemical pathways unique to the trypanosomes, such as mitochondrial RNA editing, will play an important role in the efforts to bring these diseases out of neglected status and into the limelight.

2.5 Supporting Information

2.5.1 Principal Component Analysis

Principal component analysis (PCA) is used to re-express large data sets in a basis whose components are orthonormal vectors in the direction of maximum data variance [57]. By expressing a molecular dynamics trajectory as a linear combination of principal components the covariance, or thermal noise, is eliminated and the underlying collective motions become evident[58-60]. Principal component analysis was carried out on the C-alpha backbone for the dynamics phase of the trajectory (from 7-20 ns). Translational and rotational degrees of freedom were removed by fitting each configuration to the structure sampled at 500 fs. A zero-mean covariance matrix, *C*, was generated according to equation 2.1:

$$C = XX^T \tag{2.1}$$

Where the superscript *T* designates the transpose of the matrix *X*, an element of which is given:

$$X_{ij} = \frac{1}{\sqrt{t-1}} x_{ij} - \langle x \rangle \tag{2.2}$$

Where the index i runs from 1 to N, with N denoting the number of residues in the protein. The index j runs from 1 to t, where t is the total number of configurations sampled and $\langle ... \rangle$ denotes the average over all configurations sampled. The matrix C is diagonalized by determining a set of orthonormal eigenvectors, i.e. the principal components, that satisfy the following equality:

$$C = PDP^T \tag{2.3}$$

where P is a $3N$ x $3N$ matrix whose columns are the principal components and D is the diagonalized covariance matrix whose main diagonal contains the corresponding eigenvalues, denoted $\lambda_{i,j}$. That is, p_i, the ith column of P, is a vector in a direction with an associated mean-square deviation, or variance, given by $\lambda_{i,i}$. The root mean square fluctuation of residue j, RMSF_j, along principal component i was calculated using equation 2.4:

$$RMSF_j = \sqrt{\lambda_{i,i}\{(y_{i,j})^2 + (y_{i,j+1})^2 + (y_{i,j+2})^2\}} \tag{2.4}$$

where $y_{i,j}$ is the j^{th} component of p_i.

2.5.2 Radius of Gyration

The radius of gyration (RGYR) was determined using the measure command using the TCL/TK command window within VMD. The formula used to calculate the radius of gyration is given by equation 2.5:

$$rgyr^2 = \frac{\left(\sum_i m(i)(r(i)-\bar{r})^2\right)}{\left(\sum_i m(i)\right)} \qquad 2.5$$

Where the sums run over the selected group of atoms, *m(i)* is the mass of atom *i*, *r(i)* is the position of atoms *i* and $\bar{r}$ is the geometric center of the selected group of atoms.

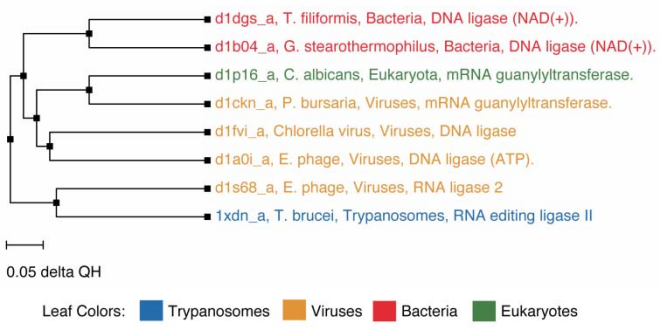


Figure 2.7: Structural phylogeny of the superfamily. A phylogenetic tree based on the multiple structural alignment performed on all covalent nucleotidyl superfamily members (alignment shown in Figure 2.2) is shown above and colored by domain of life. The distance metric used is Q_H and the analysis was performed with the MultiSeq plugin in VMD. Detailed information on the methods used to create the structural phylogenies and their interpretation can be found in Refs.[61, 62].

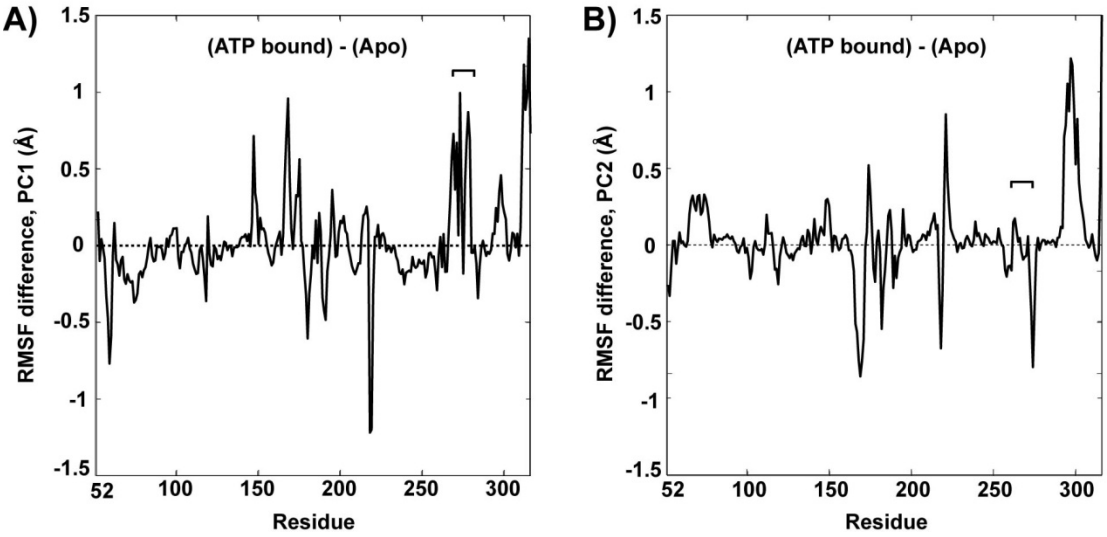


Figure 2.8: Principal component analysis of the apo and ATP-bound systems. A) The RMSF difference per residue due to the first (most dominant) principal component, PC1, indicates the regions of REL1 that are most affected due to the presence of ATP, along the PC1. B) Same as in (A), but for the second most dominant collective motion, PC2. Note that the unique loops predicted to play a role in signaling (I262-A282, denoted with a black bar above the data) exhibit increased fluctuations along the most dominant PC1.

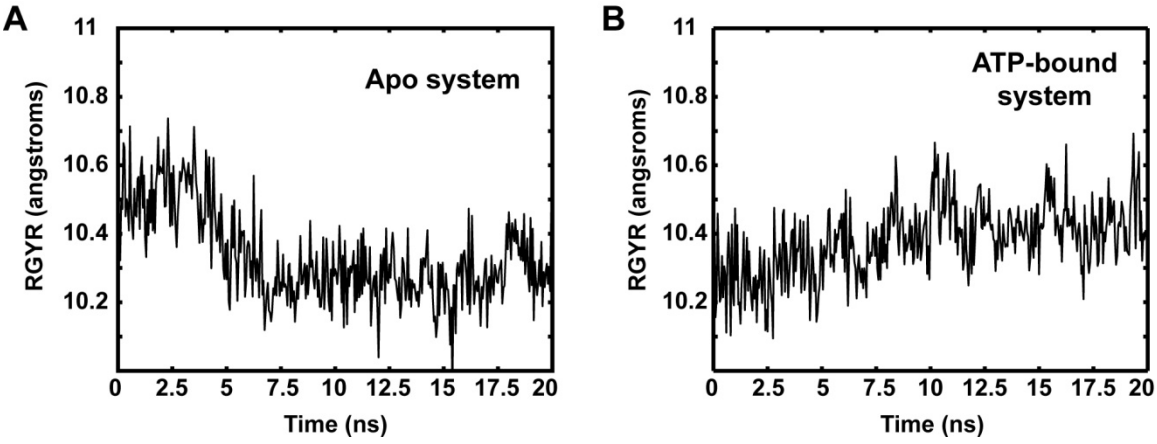


Figure 2.9: Radius of gyration calculated over the 20ns trajectory for the apo-system. (A) and ATP-bound system (B).

This chapter, in full, is a reprint of “Functional and Structural Insights Revealed by Molecular Dynamics Simulations of an Essential RNA Editing Ligase in Trypanosoma brucie”, which was published 2007 in PLoS Neglected Tropical Diseases, volume 1, issue 2, e68, by Romie E. Amaro, Robert V. Swift, and J. Andrew McCammon. This chapter is included with the permission of Romie E. Amaro and J. Andrew McCammon.

Chapter 3

Towards Understanding the Conformational Dynamics of RNA Ligation

Abstract

Members of the genus trypanosoma, which include the pathogenic species *T. brucei* and *T. cruzi*, edit their post-transcriptional mitochondrial RNA via a multi-protein complex called the editosome. In *T. brucei*, the RNA is nicked prior to uridylate insertion and deletion. Following editing, nicked RNA is religated by one of two RNA-editing ligases (*Tb*REL). The current study describes a recent 70 ns molecular dynamics simulation of *Tb*REL1, an ATP-dependent RNA-editing ligase of the nucleotidyltransferase superfamily that is required for the survival of *T. brucei* insect and bloodstream forms. In this work, a model of *Tb*REL1 in complex with its full dsRNA substrate is created based on the homologous relation between *Tb*REL1 and T4 Rnl2. The simulation captures *Tb*REL1 dynamics in the state immediately preceding RNA ligation, providing insights into the functional dynamics and catalytic mechanism of the kinetoplastid ligation reaction. Important features of RNA binding and specificity are revealed for kinetoplastid ligases and the broader nucleotidyltransferase superfamily.

3.1 Introduction

The intricate processing of genetic information is central to biological complexity and to the replication processes that sustain life. The discovery of post-transcriptional RNA editing was a rich addition to our appreciation of the central dogma of biology and is still a vibrant area of research with many unanswered questions. Within the mitochondria of kinetoplastids, a unique form of RNA editing exists whereby uridylates are inserted or deleted[63]. To carry out the editing process, kinetoplastids utilize a multi-protein complex known as the editosome. Though many details remain unresolved, the editosome is known to include a core 20S complex of 1.6 MDa containing 16-20 different proteins. Additionally, at least three distinct 20S complexes, distinguished by their constituent proteins and specificity for insertion or deletion editing, have been detected [42], suggesting that an inherent dynamism in the 20S core may accommodate the varied specificity requirements of different gene products.

The editosome is localized in the mitochondria, where mitochondrial DNA has a topologically linked network of thousands of minicircles, which encode guide RNAs (gRNAs), and dozens of maxicircles, which encode components of respiratory complexes and energy-transduction systems. After transcription, gRNAs base pair with pre-mRNA from maxicircles *via* conserved "anchor sequences"[43, 64]. Following gRNA base pairing, an endonucleolytic component of the editosome cleaves the pre-mRNA at points of nucleotide mismatch between the pre-mRNA and the gRNA. Depending on the type of

mismatch, uridylates are then inserted or removed from the nicked site of the pre-mRNA by a terminal uridydyl transferase or a U-specific 3' exonuclease, respectively[65, 66]. One of two RNA-editing ligases, *Tb*REL1 or *Tb*REL2, then religates the nicked RNA, generally depending on whether the preceding mRNA processing was insertion or deletion, respectively. The religated, "mature" mRNA is then ready for translation. *Tb*REL1 is relevant to structure-based rational drug design and associated structure-function studies both because its catalytically competent N-terminal adenyltransferase domain has been crystallized in complex with ATP at 1.2 Å resolution[48] and because it is required for the survival of *T. brucei* insect and bloodstream forms[45, 67].

An ATP-dependent RNA ligase, *Tb*REL1 belongs to the nucleotidyltransferase superfamily, whose members include ATP and NAD^+ dependent DNA ligases as well as eukaryotic mRNA capping enzymes[1]. Each superfamily member shares five conserved active site structural motifs, along with a conserved catalytic mechanism in which a nucleotide is transferred to the 5´ end of a polynucleotide *via* a covalent enzyme-NMP intermediate. Catalysis of dsRNA by *Tb*REL1 is initiated when a conserved catalytic lysine nucleophile attacks an ATP alpha phosphate, displacing pyrophosphate and forming an enzyme-AMP intermediate. In the second step, nicked dsRNA binds, and AMP is transferred from the enzyme to the $5´PO_4$ end of the nick, forming a 5'-5' phosphoanhydride linkage. In the third step, the 3´OH end, which apposes the $5´PO_4$, attacks the phosphoanhydride linkage, displacing AMP and forming the religated dsRNA product.

Though the three religation steps of *Tb*REL1 are widely acknowledged, the exact dependence on divalent metal co-factors remains in doubt. For example, though the electron-density map of the ATP-bound *Tb*REL1 N-terminal domain clearly shows a Mg^{2+} coordinated between the non-bridging oxygen atoms of the ATP beta and gamma phosphates[48], nucleotide transfer has not yet occurred, suggesting that a single Mg^{2+} may be insufficient to trigger the first chemical step, though the orthogonal position of the lysine nucleophile relative to the alpha-phosphate electrophile, which hinders nucleophilic attack, is an alternative explanation[4]. Additionally, kinetic evidence for related superfamily members suggests that the nucleotidyltransferase step proceeds via a two-Mg^{2+} mechanism[13], while crystal-structure data from both the PBCV-1 capping enzyme and T4 RNA ligase 2 (Rnl2) show that a divalent metal co-factor in the vicinity of the alpha phosphate remains after step one of catalysis[4, 7] and implies that a two metal mechanism may optimize the nucleotide transfer turn over rate. These contrasting results suggest that an allosteric signal, possibly from other protein members of the editosome, may lead to the uptake and coordination of a second Mg^{2+}, thereby triggering nucleotide transfer[48]. The plausibility of ATP-dependent allosteric communication between *Tb*REL1 and other editosome members was demonstrated in molecular dynamics (MD) simulations of *apo* and ATP-bound *Tb*REL1[68]. These simulations suggested that ATP binding induces increased mobility of a putative allosteric loop; similar phenomena may provide the impetus for the uptake of a second Mg^{2+}.

Both sequence and structural phylogenetic analysis of superfamily members indicate that the closest known *Tb*REL1 homolog is the T4 phage Rnl2[48, 68]. Recently,

the structure of full-length Rnl2, in complex with its nucleic-acid substrate, was captured in two different conformations[7]. These two unique conformations illustrate the active-site conformational rearrangements that occur as the complex proceeds from the step-two product, immediately following step-two chemistry, to the step-three substrate, immediately preceding step-three chemistry. Additionally, comparing the crystal structure of human DNA ligase1 (LIG1), in complex with its DNA substrate, to the Rnl2 nucleic-acid complex reveals the salient features of substrate conformational requirements and substrate specificity. In both LIG1 and Rnl2, the nucleotides local to the nick are found in a RNA-like A-conformation. This conformation aligns the 3'OH and apposing 5'PO_4 at the nick site and is likely required for step-three chemistry. Normally in the canonical B-form, the DNA substrate of LIG1 is distorted into the required A-conformation *via* two phenylalanines that intercalate the minor grove. These phenylalanines, conserved in DNA ligases, are absent in RNA ligases like Rnl2 and *Tb*REL1. Further details are presented in the original crystal structure paper[7] and are reviewed elsewhere in the context of other DNA and RNA ligases[69].

Motivated by the close homology between Rnl2 and *Tb*REL1 and the desire to understand the functional dynamics of the kinetoplastid ligation reaction, we investigate the conformational dynamics of *Tb*REL1 in complex with a nicked dsRNA substrate in its canonical A-form. Based on the homologous structural and functional relationships between *Tb*REL1 and T4 Rnl2, we develop a model of a nicked dsRNA-*Tb*REL1 complex (Fig. 3.1) and perform an all-atom explicitly solvated molecular dynamics simulation of the system. The integration of RNA into MD simulations has benefited

from recent improvements in both force-field development and algorithms[25, 70], allowing us to make insights into the dynamics and catalytic mechanism. The simulated system contains AMP attached to the nick 5´PO_4 moiety (AppN), characteristic of the step-two intermediate[7]. The non-nicked strand represents the template, or gRNA, and the nicked strand represents the post-transcriptionally modified pre-mRNA. Fluctuations of conserved interactions between the enzyme and RNA substrate, as well as the dynamics of conserved active-site residues, are examined and discussed in the context of step three of the religation mechanism. Furthermore, we explore the effect of a unique kinetoplastid insert region on dsRNA binding.

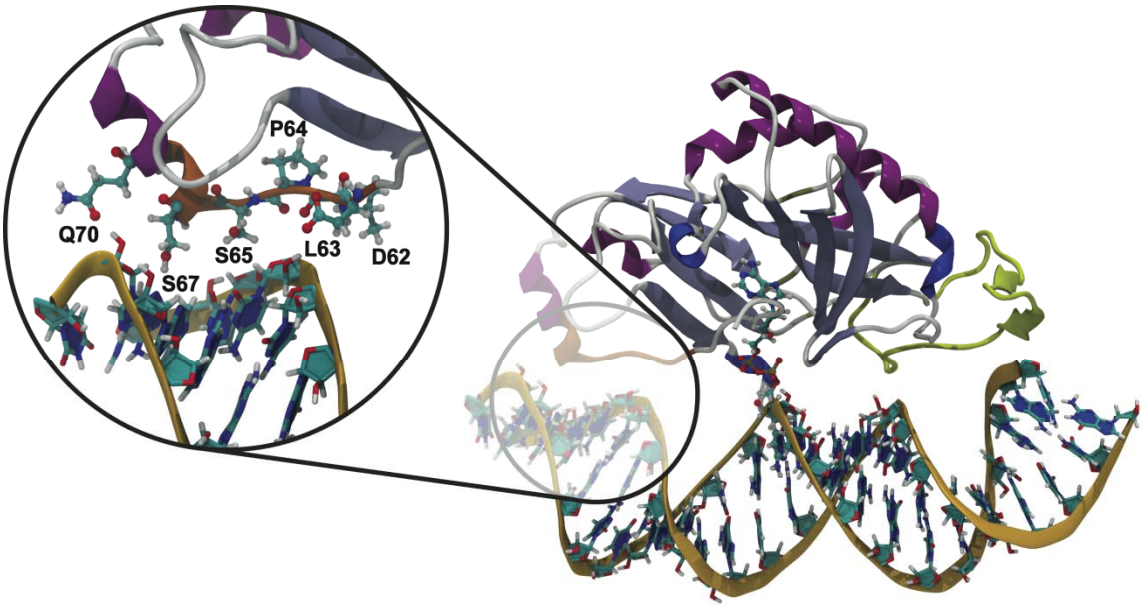


Figure 3.1: Nicked dsRNA bound to *Tb*REL1, poised to catalyze religation. A unique kinetoplastid insert is colored in yellow, and a cluster of residues thought to be important in RNA recognition on the 5'PO_4 side of the nick is shown in orange. The magnified insert shows protein residues that make stabilizing contacts with the RNA.

3.2 Materials and Methods

3.2.1 System preparation.

A model of the step-two intermediate of *Tb*REL1 in complex with A-form dsRNA was built in homology to chain B of the T4 Rnl2 crystal structure (PDBID 2HVR), which captures T4 Rnl2 immediately preceding step-three chemistry[7]. A 12-mer of A-form dsRNA, with a sequence identical to the nicked DNA/RNA hybrid co-crystalized with T4 Rnl2, was generated using the NUCGEN module of Amber 9[71], which creates accurate nucleotide models based on fiber-diffraction studies. The dsRNA was then aligned to the nicked DNA/RNA hybrid by minimizing the root-mean-square-deviation (RMSD) between the two nitrogenous bases of the DNA/RNA hybrid that flanked the nick and the commensurate nitrogenous bases of the model dsRNA. Next, the dsRNA was translated onto the crystal structure of the *Tb*REL1 N-terminus (PDBID 1XDN) by minimizing the RMSD between the AppN adenine moiety of the DNA/RNA hybrid and the adenine moiety of the ATP bound in the *Tb*REL1 active site. A nick was then created in the model dsRNA identical to the nick of the DNA/RNA hybrid bound to T4 Rnl2. The beta and gamma phosphates of the *Tb*REL1-bound ATP were subsequently deleted, and the O-P_{AMP}-O-$P_{5'PO4}$ dihedral angle was brought into agreement with the corresponding dihedral angle of the T4 Rnl2 crystal structure. The resulting AMP is linked through its phosphate to the 5'PO_4 nicked terminal nucleotide; we designate this AMP-phosphate-

phosphate-nucleotide residue AppN according to the nomenclature of Nandakumar and co-workers.

To ensure that the active site of our model reflected the conformation of Rnl2 immediately preceding step-three chemistry, the *Tb*REL1 active-site residue R111 was given a conformation commensurate with its Rnl2 homologue, R55. Specifically, R111 was rotated away from the non-bridging oxygen of the AppN alpha-phosphate and placed within hydrogen-bonding distance of the Y58 backbone carbonyl. The conformations of the other TbREL1 active-site residues were in close agreement with the conformations of the homologous active-site residues of T4 Rnl2 immediately preceding step-three chemistry, and so required no further manipulation. A Mg^{2+} was manually positioned approximately 2 Å from a non-bridging AMP alpha phosphate based on the assumptions that ion dissociation occurs neither during RNA binding nor during subsequent active-site rearrangements and that the *Tb*REL1 catalytic mechanism proceeds via a two-Mg^{2+} mechanism. This Mg^{2+} position was chosen based on the crystal-structure data of both the PBCV-1 capping enzyme[4] and the T4 Rnl2 step-one product[7], as well as on a similarly positioned Ca^{2+} in the crystal structure of T4 Rnl1 bound to the unreactive analog AMPCPP[72]. Five additional water molecules, each approximately 2 Å from the Mg^{2+} ion, were manually positioned to duplicate the expected octahedral coordination complex.

All co-crystallized water molecules in the *Tb*REL1 structure were retained in our model. Selenomethionines used for crystal structure refinement were replaced with

methionines, and WHATIF[73] was used to assign histidine protonation states. Protonation-state assignments were manually verified. With the exception of the non-standard AppN residue, all RNA and protein hydrogens and force-field parameters were assigned according to the Amber ff99SB force field[71] using the xleap module in Amber 9[71]. The non-standard residue AppN is composed of an adenosine linked to a cytidine *via* a phosphoanhydride linkage; consequently, we assigned AppN charges using the RA5 and RC5 residues of the Amber ff99SB force field as a template. The charge of the oxygen bridging the phosphoanhydride linkage was assumed to be identical to that of the oxygen bridging the alpha and beta phosphates in the Carlson ADP parameter set[74]. All missing AppN force-field parameters came from the GAFF force field[49]. Following parameterization, the complex was immersed in a rectangular box of TIP3P waters[26] with a 10 Å buffer from the protein to the edge of the water box in each direction. Sodium and chloride ions were added to the system, bringing it to electric neutrality and yielding a 100 mM salt solution.

3.2.2 MD simulations

Spurious contacts were removed through a four-tier set of 45,000 energy minimization steps. The first 10,000 steps were performed in two 5,000-step cycles. Hydrogens were relaxed during the first 5,000 steps, holding all other atoms fixed. Hydrogens, waters, and ions were relaxed during the next 5,000 steps, holding the protein and nucleic acid fixed. Ten thousand steps of minimization followed in which only the

protein backbone was fixed. No constraints were applied during the last 25,000 steps of minimization.

To equilibrate our model, we performed a one-nanosecond restrained simulation using the minimized system as a starting point. Temperature was introduced at 310 K with harmonic restraints on the protein backbone. Through the gradual relaxation of the backbone restraints, the protein structure was preserved as kinetic energy was introduced. For the first 250 ps, a harmonic restraining force of 4.0 kcal/mol/Å^2 was applied to the backbone atoms. Similarly, for the second, third, and fourth 250 ps segments, the harmonic restraints were consistently weakened to 3.0, 2.0, and 1.0 kcal/mol/Å^2, respectively.

A 70 ns MD simulation followed in which unrestrained dynamics were performed with a one-fs timestep. The duration of the simulation was sufficient to equilibrate the system based on RMSD values (overall alpha carbon [Fig. 3.7] and the 3'-terminus sugar pucker phase angle [Fig 3.3]), while allowing us to sample functionally relevant conformations local to the step-2 intermediate. The temperature bath was maintained by Langevin dynamics at 310 K, and pressure was maintained with the hybrid Nose Hoover-Langevin piston method at 1 atm[50] using period and decay times of 100 and 50 fs, respectively. The Particle Mesh Ewald algorithm was used to treat long-range electrostatics[51]. A multiple time-stepping algorithm was employed where bonded interactions were computed at every time step, short-range non-bonded interactions were computed every two time steps, and full electrostatics were computed every two time

steps. The water hydrogen-oxygen and hydrogen-hydrogen distances were constrained using the SHAKEH algorithm[18] to be within 0.0005 of the nominal lengths. All minimization and MD simulations were carried out with NAMD 2.6[18] using the Amber ff99SB force field[71]. Simulations were performed on the TeraGrid ABE cluster. A typical benchmark for the 57,575-atom system on 128 processors was 0.21 days per nanosecond of simulation. System configurations were sampled every 50 ps, generating 1,400 coordinate snapshots for subsequent analysis.

3.2.3 Pucker analysis

The pseudorotational formalism proposed by Rao, Westhof, and Sundaralingam[75] was used to characterize the puckering of the nick 3' ribose ring. We first defined the following Fourier sums:

$$A = \frac{2}{5}\sum \theta_j \cos\left(\frac{4\pi j}{5}\right) \qquad 3.1$$

$$B = -\frac{2}{5}\sum \theta_j \sin\left(\frac{4\pi j}{5}\right) \qquad 3.2$$

where j ranges from zero to four, and θ_0, θ_1, θ_2, θ_3, θ_4 are the torsion angles about C2'—C3', C3'—C4', C4'—O4', O4'—C1', and C1'—C2', respectively. The pucker phase angle, P, identifies which part of the ring is furthest from planarity. Ribose configurations generally take on the 2' endo pucker typical of DNA or the 3' endo pucker typical of RNA. The possible values of P, which range from zero to 360, correspond to the stages along the so-called pseudorotation pathway. The ribose ring has five atoms, each of

which can be above or below the plane, yielding 10 distinct stages. The envelope conformations correspond to P = 0°, 36°, 72°, etc. and alternate with the twist conformations at P = 18°, 54°, 90°, etc. In the Rao-Westhof-Sundaralingam pseudorotational formalism,

$$P = \tan^{-1}\left(\frac{B}{A}\right) \qquad 3.3$$

The maximal torsion angle or puckering amplitude, θ_m, measures the magnitude of the deviation from planarity[75].

$$\theta_m = \sqrt{A^2 + B^2} \qquad 3.4$$

Polar plots are generally used to represent sugar pucker, with the puckering amplitude and phase angle assigned as the radial and angular coordinate, respectively. Analysis of experimental structures shows that the 2' endo ribose configuration typical of DNA has 137° < P < 194°, and the 3' endo ribose configuration typical of RNA has -1° < P < 34°[76].

3.2.4 Principal Component Analysis

Principal component analysis, or PCA, is a common statistical data-analysis technique whose application to the field of biomolecular simulation was pioneered by Berendsen[77] and Garcia[57]. By diagonalizing a covariance matrix constructed of molecular coordinates sampled over the duration of a MD trajectory and listing the resulting eigenvectors in order of descending eigenvalue magnitude, PCA separates large, low frequency conformational changes from small, high frequency motions. By

considering only the conformational fluctuations of the first few modes, researchers frequently use PCA to filter out large isomerization events and to reduce the complexity of trajectory analysis. In this work, PCA was used to explore the large, low frequency motions of the dsRNA substrate. Following trajectory equilibration, principal component analysis was carried out on the phosphate atoms of the dsRNA substrate, using the ptraj module in Amber 9[71]. Prior to performing PCA, translational and rotational motions were removed by fitting the sampled coordinates to the first equilibrated frame of the trajectory.

3.3 Results and Discussion

3.3.1 Equilibration.

A 70 ns constraint-free simulation was performed on the dsRNA-bound *Tb*REL1 system to examine *Tb*REL1 dynamics. A plot of the time evolution of the RMSD of the protein alpha carbons (Fig. 3.7), in which each trajectory frame was aligned to the initial frame to remove rotational and translational motion, suggested that our preliminary efforts to equilibrate the system prior to our 70 ns simulation were successful. However, subsequent examination of the distance between the nick 3'OH and the phosphorus atom of the 5'-phosphoanhydride linkage suggested that equilibration was not complete until after 38.75 ns. During the initial phase (0-38.75 ns), these two atoms, between which a bond forms following step three, were distant (7.16 Å +/- 1.20). During the second phase

(38.75 ns to 70 ns), however, the nick 3'OH and the nick 5' phosphorus moved closer together (3.95 Å +/- 0.29), as required for catalysis. Correspondingly, the trajectory time from 0 to 38.75 ns is referred to as the "equilibration phase," and from 38.75 to 70 ns as the "dynamics phase." Unless otherwise specified, all analysis was performed on the dynamics phase.

3.3.2 Active-Site Interactions

The atomic interactions within the active site can be divided into three categories: interactions between *Tb*REL1 and the adenine moiety of AppN, interactions between *Tb*REL1 and the ribofuranose of AppN, and interactions between *Tb*REL1 and the pyrophosphate of AppN. Throughout our discussion, we adopt the naming scheme of Nandakumar and coworkers, naming the nicked strands based on their active-site ends (i.e. the 5'PO_4 strand and the 3'OH strand)[7].

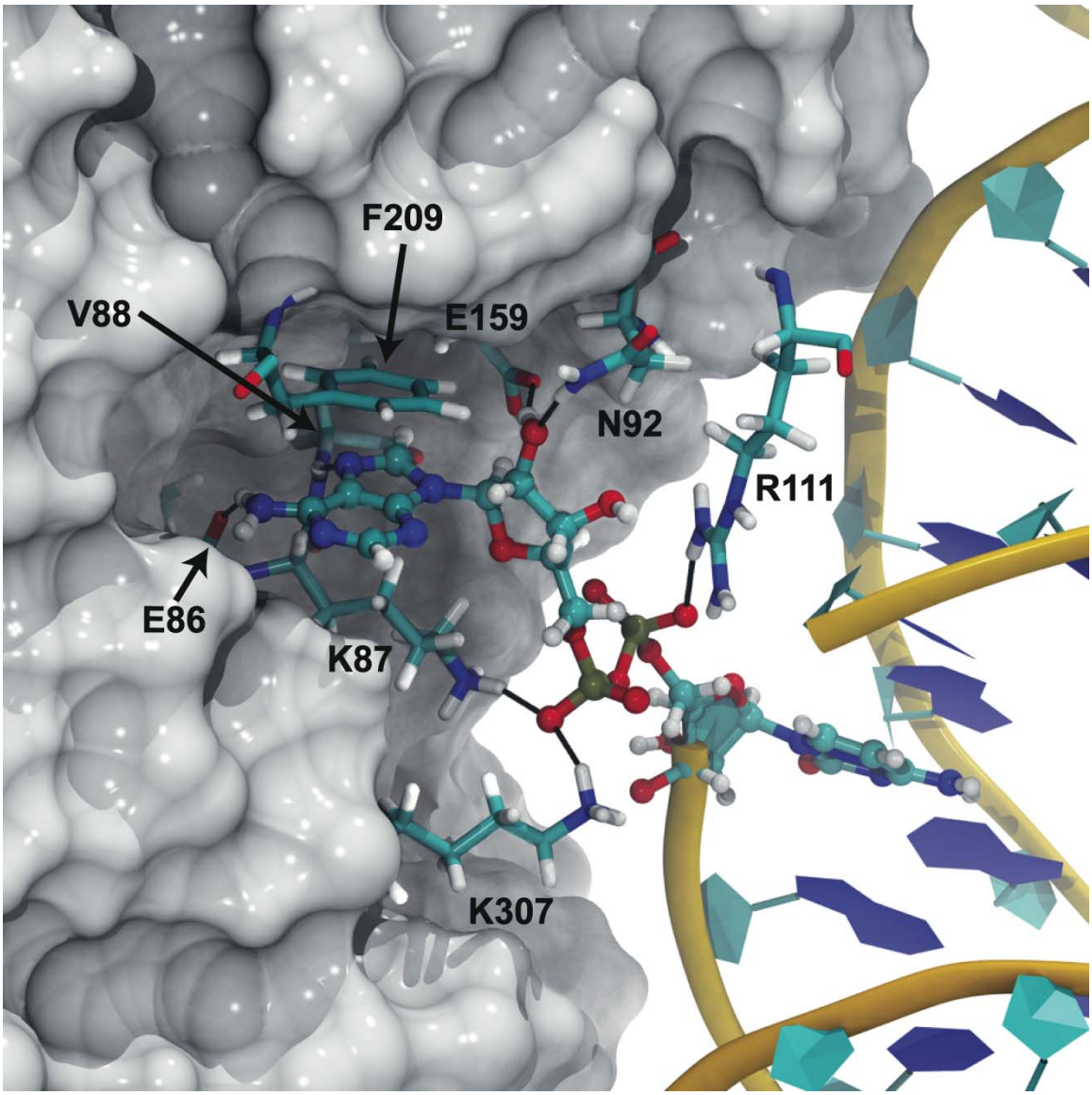


Figure 3.2: The equilibrated structure illustrates the close contacts between key active-site residues (shown in licorice), the AppRNA (shown in ball and stick), and the 3'OH RNA strand.

Deep within the active site, F209, E86, and V88 interact with the adenine moiety of AppN and stabilize its motion (Fig. 3.2). F209 acts as a conserved hydrophobic platform that forms pi-pi interactions with the aromatic adenine moiety of AppN, preventing large lateral fluctuations of this base within the binding pocket. Hydrophobic

stabilization in this position is essential for step-three catalysis in other type II RNA ligases, as demonstrated by the partial recovery of religation activity when a leucine residue was substituted for the F209 homolog in an alanine scan of T4 Rnl2[12]. Our simulation is consistent with this mutagensis study; F209 exhibits minor fluctuations, with an average distance of 4.14 Å between the centers of mass of its phenyl group and the AppN adenine moiety (Table 3.1). The movement of AppN is also restrained by E86, which accepts a hydrogen bond from the adenine N6-amine that holds the E86 O1 (or O2) oxygen within 2.97 Å, on average (Table 3.1). Conserved among type II RNA ligases, E86 is essential in mediating step-three catalysis in T4 Rnl2[12]. The observed distances and fluctuations of both F209 and E86 are similar to those observed in a previous simulation of the ATP-*Tb*REL1 complex[68]. These results, along with the T4 Rnl2 mutational analysis[12], reflect the importance of F209 and E86 in mediating the re-ligation of the adenylated dsRNA substrate and the formation of the adenylated enzyme intermediate.

Stabilizing interactions between the protein and the AppN ribofuranose complement the contacts formed between the adenine moiety and residues deep within the binding pocket. In particular, E159, conserved across all members of the superfamily, and N92, conserved across all type II RNA editing ligases[12], make critical hydrogen bonds with the ribofuranose O2' hydroxyl group (Fig. 3.2, Table 3.1). The close, stable contact made by E159 was observed in a previous simulation of the ATP-bound complex[68] and is consistent with the conservative mutational analysis, which identified E159 as an essential mediator of steps one and three of catalysis[12]. Similarly,

an alanine scan of T4 Rnl2 revealed that the N92 homolog plays a critical role in the religation of the pre-adenylated dsRNA substrate[12]. No isomerization of the AppN ribofuranose moiety was observed during the course of catalysis in the T4 Rnl2 homolog[7]; consequently, the mutational anlaysis results, coupled with the short, stable contacts observed during our simulation, suggest that the N92-ribofuranose interaction helps maintain the ribofuranose conformation throughout the catalytic cycle.

Table 3.1: **Active-site interactions, including protein-AppN interactions, protein-nick interactions, and interactions with the Mg^{2+}-coordinated water molecules**.

	Length [Å] (Std) [Å]
AppN adenine – F209 phenyl group	4.14 (0.22)
E86:O – AppN adenine N6	2.97 (0.20)
E159:OE1 – AppN ribose O2'	2.55 (0.08)
N92:ND2 – AppN ribose O2'	3.30 (0.41)
K87:NZ – AppN:O1P	3.39 (0.42)
K307:NZ – AppN:O2P	3.87 (0.56)
R111:NH1 – AppN:O1P	2.80 (0.11)
Nick 3' HO'3 – nearest non-bridging nick 5'PO_4 oxygen	1.90 (0.64)
R111:CZ - nick O3'	3.95 (0.34)
H89:ND1 – WAT:O	2.85 (0.21)
E159:CD – WAT:O	3.49 (0.27)
E283:CD – WAT:O	3.45 (0.24)
E283:CD – WAT:O	3.44 (0.21)
nick O2' – WAT:O	3.30 (0.53)

Charged residues at the periphery of the binding site, K87, K307, and R111, each interact with the AppN pyrophosphate and stabilize the step-three substrate conformation (Fig. 3.2). The orthogonal orientation of the K87 nucleophile to the alpha phosphate is typical of the step-three substrate conformation and persists throughout the 31.25 ns dynamics phase. The system does not undergo isomerization away from the step-three substrate conformation, indicating that the network of interactions captured in the homologous T4 Rnl2 crystal structure is a reasonable model of the *Tb*REL1 active-site

architecture prior to step three of catalysis. Interestingly, K87 remains within hydrogen-bonding distance of a non-bridging pyrophosphate oxygen during the dynamics phase of the simulation (Table 3.1). This result suggests that the K87 nucleophile helps maintain the protein in the step-three conformation once isomerization into that conformation has occurred. Consequently, K87 seems to perform a dual role during catalysis: during step one it plays the central chemical role as the nucleophile, and during step three it plays a peripheral stabilizing role as a hydrogen-bond donor. To the best of our knowledge, no experimental evidence has been published supporting this hypothesis, but the putative dual role of K87 could be probed using the methods Yin and co-workers employed during their structure-function analysis of T4 Rnl2[12]. Nearby K307 also donates a hydrogen bond to a non-bridging pyrophosphate oxygen, remaining close throughout the dynamics phase of the simulation (Table 3.1). The AppN stabilization that this interaction provides is in agreement with mutational analysis, which has shown that a positively charged hydrogen-bond donor at this position is most efficient for religating preadenylated nicked-dsRNA. Along with K87 and K307, R111 also donates a hydrogen bond to one of the non-bridging pyrophosphate oxygens and remains in close contact throughout the dynamics phase of the simulation (Table 3.1).

3.3.3 Insights into Catalysis

A requirement for step-three catalysis is that the O3' nucleophile and the departing, bridging oxygen of the AppN pyrophosphate take on an apical, or “in-line” arrangement. This apical requirement is achieved only when the terminal ribofuranose of

the nucleotide on the 3'OH side of the nick takes on a 3' endo conformation[78]. The 3' endo conformation is the native conformation of dsRNA, explaining why T4 Rnl2 requires that the strand on the 3'OH side of the nick be RNA while showing no preference for nucleic-acid type on the 5´PO_4 side[78]. *In vitro* religation experiments with *Tb*REL1 alone have likewise demonstrated that the strand on the 3'OH side of the nick must be RNA[47], though inclusion of other components of the editosome complex alters substrate specificity. Here, we assume based on the close homology between *Tb*REL1 and T4 Rnl2 that the RNA requirement on the 3'OH side of the nick arises from the same 3' endo requirement observed in T4 Rnl2. The equilibrium distribution of sugar conformations observed during our simulation should be a reasonable approximation of those available for catalysis during step-three religation.

In order to gain insight into the temporal fluctuations of the required 3' endo conformation, we monitored the ribofuranose pseudorotation angle of the nucleotide at the 3'OH terminal of the nick over the dynamics phase of the simulation (Fig. 3.3). The pseudorotation angle is a measure of the sugar conformational space and includes two stable conformations corresponding to the 2' endo and 3' endo conformations of DNA and RNA, respectively. As noted before, the 2' endo conformation is characterized by a pseudorotation angle between 137^o and 194^o, and the 3' endo conformation is characterized by a pseudorotation angle between -1° and 34°[76]. Consistent with the A-form dsRNA structure, the ribofuranose of the nick terminal 3' hydroxyl nucleotide remains almost exclusively in the 3' endo conformation (15.2^o +/- 10.9^o).

Visual inspection of the O3' nucleophile and the bridging oxygen leaving group of the AppN pyrophosphate demonstrated that while the two atoms remain parallel to one another, they are not strictly in-line; the attacking O3' nucleophile is vertically displaced from the plane of the leaving group (Fig. 3.3). The vertical displacement is a consequence of relaxing steric conflicts in the initial model during minimization and equilibration and implies that the active-site conformation is sub-optimal for religation. Despite this, the time evolution of active-site topography still provides a useful model from which mechanistic insight may be inferred.

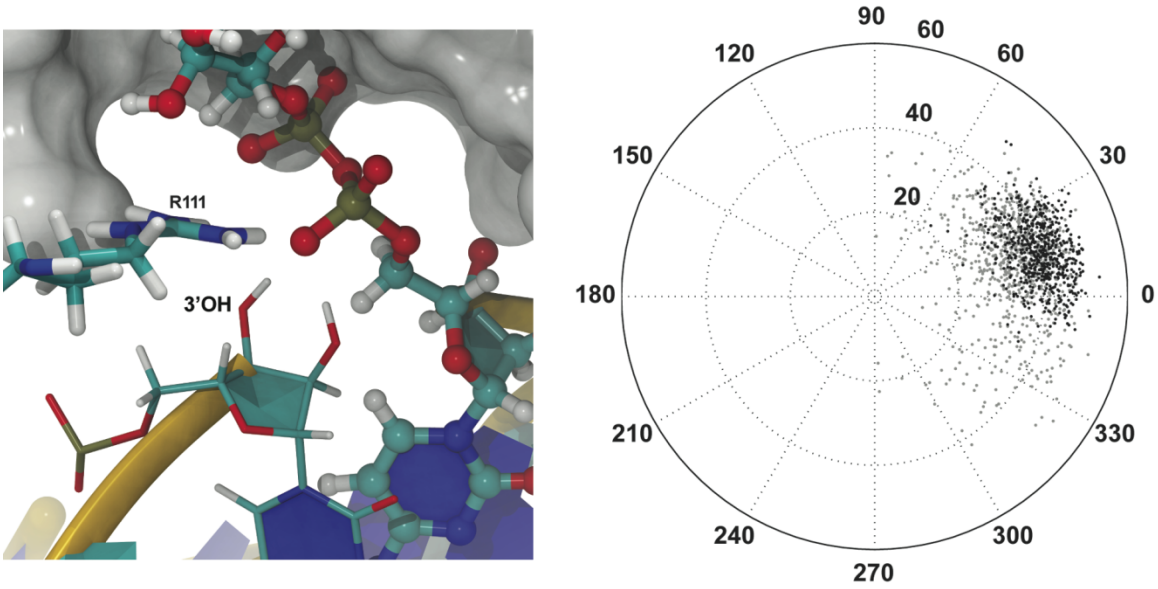


Figure 3.3: The equilibrated ribofuranose pseudorotation angle for the 3'OH nucleotide as depicted within the active site (left panel) and in polar coordinates (right panel, equilibration data in gray, dynamics data in black).

The nicked 3' hydroxyl nucleophile must be deprotonated either prior to, or following, reaction with the beta phosphate of the AppN intermediate. Determining the lowest-energy pathway of proton transfer is complicated and highly dependent on the

electrostatic environment surrounding the acid. However, proton transfer is restrained by the requirement that a new covalent bond forms between the acidic hydrogen and the base. This requirement allows simple hypotheses of proton transfer pathways to be made based on the proximity of the acidic hydrogen to potential general bases. Our simulation demonstrates that the acidic hydrogen of the nicked 3' hydroxyl nucleophile resides at an average distance of 1.90 Å from the nearest negatively charged non-bridging pyrophosphate oxygen (Table 3.1). Furthermore, protein fluctuations bring these two atoms within 1.40 Å of each other, causing significant overlap of their van der Waals radii and suggesting possible proton transfer between these two atoms. As the acidic hydrogen lies between the putative basic oxygen and the nucleophilic O3', it is likely that proton transfer either precedes nucleophilic attack or occurs concertedly with phosphodiester formation. Additionally, the positively charged guanidinium moiety of R111 is proximal to the nick O3' hydroxyl (Table 3.1) and may play a role in stabilizing the O3' as it acquires negative charge during the proton-transfer process. This is consistent with mutational analysis carried out on the T4 Rnl2, which demonstrated that a lysine mutant in the position of the R111 homolog was nearly as efficient in religating preadenylated nicked dsRNA as the wild type enzyme, but that the alanine mutant lacked religation ability[12]. The idea that a non-bridging oxygen of a nucleotide phosphate may serve as a general base has also been proposed as a step in the catalytic mechanism of the PBCV-1 mRNA capping enzyme[79].

3.3.4 Mg^{2+}Coordination

The exact dependence of nicked dsRNA religation on divalent metal cofactors remains uncertain. In our model of the step-three substrate, we assume that *Tb*REL1 step-one chemistry proceeds via a two Mg^{2+} cofactor mechanism. This assumption, along with homology to the crystal structures of other nucleotidyltransferase superfamily members, suggests the Mg^{2+} placement in our model (Fig. 3.4). The first Mg^{2+} cation, which is initially coordinated by the non-bridging oxygens of the AppN beta and gamma phosphates, dissociates with pyrophosphate following the formation of the enzyme nucleotide intermediate in step one. The second Mg^{2+} cation, which is coordinated in the vicinity of the alpha phosphate, remains bound within the active site to mediate later steps of catalysis. With the model so developed, analysis of the network of interactions mediated by the presence of the divalent magnesium, as well as the proximity of the cation to the AppN substrate, suggest possible functional roles for the bound metal in the context of step-three chemistry.

In our model, the magnesium cation is octahedrally coordinated by five water molecules and a non-bridging alpha-phosphate oxygen of the AppN pyrophosphate (Fig. 3.4). None of these water molecules exchanged with bulk solvent during the simulation. Four of the five water molecules participate in a hydrogen bond network with conserved residues in motifs I, III and IV, while the fifth water donates a hydrogen bond to the 2' oxygen of the terminal nucleotide on the nicked 3' hydroxyl strand. This interaction network provides bridging hydrogen bond interactions to conserved active-site residues

and serves to maintain the relative position of the magnesium within the active site, which may be important in transition-state stabilization as the reaction proceeds.

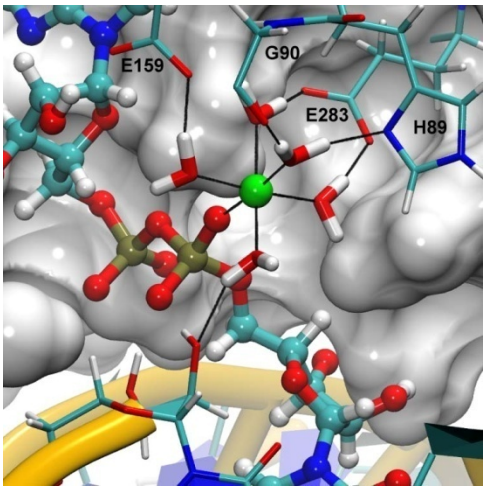


Figure 3.4: **Equilibrated coordination of the active-site magnesium ion**.

In phosphoryl transfer reactions, there are three mechanistically distinct nucleophilic substitution pathways: dissociative, concerted, and addition-elimination[80]. The addition-elimination and concerted pathways are S_N2-like; as the nucleophile approaches the phosphate electrophile, the reacting phosphate passes through a planar, pentacoordinate transition state in which one of the non-bridging oxygens acquires an additional negative charge[80, 81]. In the dissociative, E_1-like pathway, the leaving group first departs, forming a planar metaphosphate, phosphocation intermediate before the nucleophile reacts. In our model of the step-three substrate, the Mg^{2+} cation coordinates one of the non-bridging oxygens of the phosphate electrophile. The magnesium resides at an average distance of 1.83 Å from the non-bridging oxygen, opposite the O3' nucleophile across the pyrophosphate backbone (Fig. 3.4). The proximity of the divalent magnesium and the non-bridging oxygen of the phosphate electrophile may provide

electrostatic stabilization of the excess transition-state negative charge typical of the concerted and addition-elimination pathways, enhancing the turnover rate. Conversely, if step-three chemistry proceeds via a dissociative pathway, the proximity of the positively charged divalent magnesium would destabilize the phosphocation of the metaphosphate intermediate, thereby attenuating the turnover rate. While these considerations are qualitative and non-rigorous, they suggest that the presence of a divalent metal during step-three chemistry favors either a concerted or an addition-elimination pathway.

Having considered the effects of a divalent metal on the turnover rate of the third chemical step, we now consider the network of interactions provided by conserved residues, which stabilize the magnesium within the active site. Within motif I, both the unprotonated delta nitrogen of H89 and the carbonyl oxygen of G90 accept hydrogen bonds from two unique water molecules, each coordinating Mg^{2+} (Fig. 3.4, Table 3.1). Within motif III, one of the negatively charged carboxyl oxygens of E159 accepts a hydrogen bond from a coordinating water molecule. The carboxyl oxygens of E283 in motif IV likewise accept hydrogen bonds from two coordinating water molecules in a similar manner (Table 3.1). When the E159 or E283 homologs are substituted with either aspartate or glutamine in T4 Rnl2, religation activity is completely abolished, indicating a strict glutamate requirement in these positions[12]. Although the cause of this dependence is not immediately apparent, we speculate that neither glutamine, which is neutrally charged, nor aspartate, whose side chain is one methylene group shorter, provide stable hydrogen-bonding partners for the magnesium-coordinating waters.

In addition to the E159 and E283 interactions, one of the Mg^{2+}-coordinating waters also donates a hydrogen bond to the 2' oxygen of the terminal nucleotide on the nicked 3' hydroxyl strand (Fig. 3.4, Table 5.1). This interaction indirectly links the terminal nucleotide of the nicked 3' hydroxyl strand to the AppN moiety and may play a role in properly orienting the terminal nucleotide for step-three chemistry.

3.3.5 RNA-Binding Footprint

While the nucleotide specificity determinants for the T4 Rnl2 and human DNA ligase I superfamily members have been clearly defined[69], the specificity determinants of *Tb*REL1 have not. Available experimental data demonstrates that, *in vitro*, *Tb*REL1 cannot religate nicked nucleic acid in complex with gRNA unless both of the constituent fragments are RNA[47]. Based on studies of the homologous T4 Rnl2, we hypothesize that the RNA requirement on the 3'OH side of the nick may arise at least in part from the need to confer a 3' endo conformation at the 3'OH nicked terminal nucleotide[78]. The absence of religation when DNA is present on the 5' side of the nick suggests that specific *Tb*REL1-RNA interactions play important roles in mediating substrate stability during steps two and three of the religation mechanism.

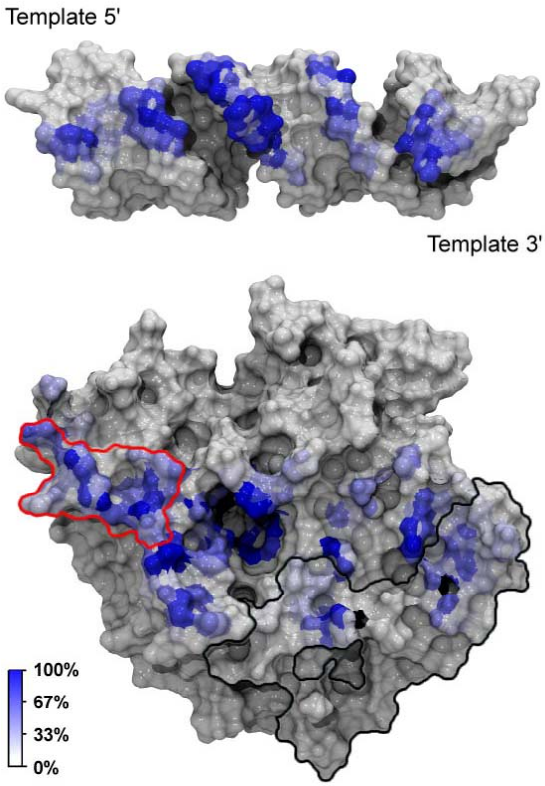


Figure 3.5: RNA binding footoprint. Top panel: the protein dynamic contact footprint projected onto the molecular surface of the RNA. Bottom panel: the RNA dynamic contact footprint projected onto the molecular surface of *Tb*REL1. Both the protein and RNA are shown in gray surface representation with the contacts colored by percent occupancy over the dynamics portion of the simulation. A unique kinetoplastid insert is outlined in black, and a cluster of residues thought to be important in RNA recognition on the 5'PO_4 side of the nick is outlined in red.

In order to identify molecular specificity determinants that mediate dsRNA binding, we generated a RNA binding footprint based on a set of protein-RNA interaction persistence times (Fig. 3.5). This analysis allows us to identify primary contact sites between the RNA and the protein and to discriminate longer-lived interactions from more transient ones. The footprint-like mapping of the protein-RNA interaction persistence times onto the protein and RNA surfaces shows that more extensive interactions occur on the 5'PO_4 side of the nick. Although this contact pattern contrasts with the results of

Nandakumar and co-workers, who report that the closest homolog of *Tb*REL1, T4 Rnl2, makes more extensive contacts with the strands on the 3'OH side of the nick, it supports the experimental observations that *Tb*REL1 requires RNA on the 5'PO_4 strand of the nick for *in vitro* polynucleotide religation. A large, contiguous set of interactions, as well as various individual interactions, makes up the more extensive contacts on the 5'PO_4 side of the nick. Specifically, D62, L63, P64, S65, S67, and Q70 collectively form a contiguous surface, herein called the "recognition sequence," which intercalates the minor grove of the dsRNA on the 5'PO_4 side of the nick (Figs. 3.1 & 3.5). Although the recognition sequence is not conserved at the superfamily level, it is conserved among the Trypanosomatidae type I mitochondrial RNA-editing ligase sequences available in the SWISS-PROT database[82]. As the depth and width of the minor grooves differ between the canonical RNA and DNA forms, this recognition sequence may play a key role in polynucleotide discrimination.

Table 3.2: RNA-protein interactions. Distances and occupancy percentages are calculated over the dynamics phase only. *Side-chain ring center of mass.

Protein	***RNA***	***Length [Å] (Std) [Å]***	***% Time Bond < 3.5 Å***
LEU63:N	C_g-7:O1P	3.84 (0.99)	55.50%
GLN70:NE2	U_n-11:O1P	5.34 (1.78)	19.62%
SER67:OG	A_n-10:N3	4.67 (1.29)	24.4%
SER65:OG	C_g-8:O2'	4.86 (1.43)	18.34%
PRO64:O	U_g-9:O2'	4.14 (0.68)	19.62%
ASP62:OD1	C_g-8:O2'	5.49 (1.64)	20.73%
MET314:O	A_n-1:O2'	3.57 (1.22)	71.29%
GLU119:OE1	C_n+12:O5'	6.37 (4.17)	25.84%
ASN120:ND2	A_n+11:O1P	3.36 (1.74)	88.20%
PHE121*	U_g+4:O2'	7.85 (0.99)	0.00%
PHE122*	G_g3:O2'	5.19 (0.48)	0.00%
ARG194:NH2	G_g+3:O1P	3.09 (0.45)	82.46%
ARG309:NH1	A_n-1:O1P	2.82 (0.12)	99.84%
ARG309:NH1	C_n-2:O2P	2.94 (0.33)	94.90%

In addition to the geometric complementarity conferred by these residues, several specific atomic interactions within this sequence may contribute to polynucleotide discrimination (Table 3.2). The carboxyl group of D62, the carbonyl backbone of P64, the side-chain hydroxyl of S65, and the side-chain amino group of Q70 each make contact with a ribofuransose 2'OH. Two of these four RNA-specific contacts are shared between D62 and S65; roughly nineteen nanoseconds into the dynamics portion of the simulation, the 2'OH interaction with S65 exchanges with D62 and persists throughout the remainder of the simulation. In DNA, the lack of the 2'OH precludes hydrogen-bond interactions with D62, P64, S65, and Q70. These interactions may partly explain why the nucleic acid on the 5'PO_4 side of the nick must be RNA for religation to occur *in vitro*.

The stability conferred by the recognition sequence is immediately apparent upon examining the root-mean-square-fluctuation (RMSF) values and the porcupine plots describing the motion along the first three principal modes (Fig. 3.6). The recognition sequence makes contact with the unnicked gRNA strand at positions -8 and -7, causing a two-fold decrease in mean-square fluctuation relative to that observed at the 5'PO_4 terminal of the gRNA (position -10). At positions -7 to 2, were the gRNA nucleotides are not in contact with the protein, there is a 1.6 fold decrease in mean-square fluctuation. Additionally, the porcupine plots illustrate that the magnitudes of atomic displacements along the first three modes are attenuated in positions -8 and -7 (Fig. 3.6). The relatively low interaction persistence times observed for the recognition sequence (Table 3.2) are readily explained by the relatively large positional fluctuations of the dsRNA at the terminal ends on the 5'PO_4 side of the nick (Fig. 3.6).

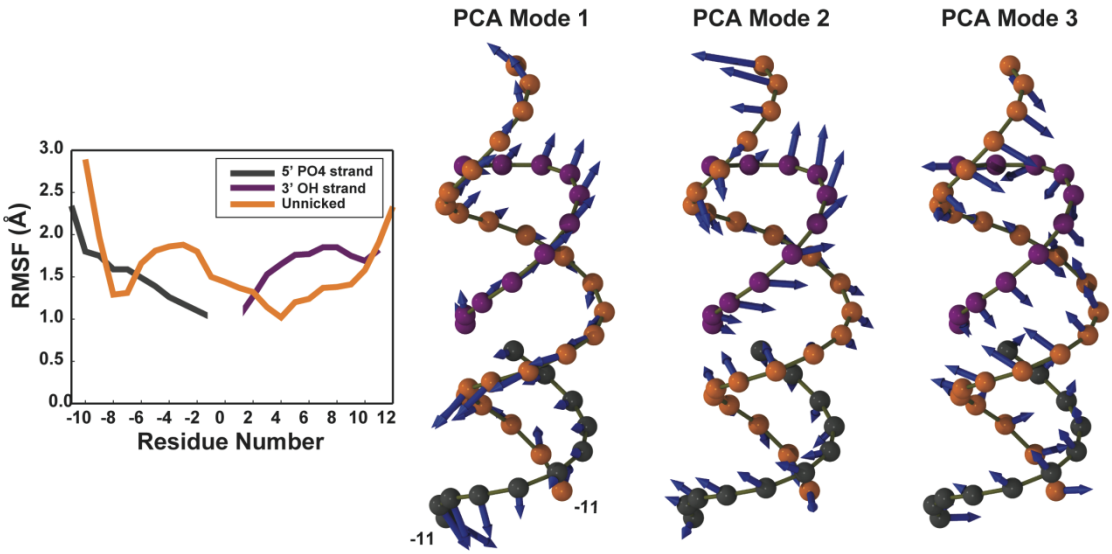


Figure 3.6: RNA flexibility as shown by both the RMSF per residue of the RNA strands and PCA analysis. In both panels, the template strand is shown in orange, the 3'-OH strand in gray, and the 5'-PO_4 strand in purple. PCA vectors were multiplied by a factor of 25 for easier visualization.

Two additional independent *Tb*REL1-RNA interactions play a role in mediating substrate stability on the 5'PO_4 side of the nick. The carbonyl of M314 forms a hydrogen bond with the 2'OH of the penultimate nucleotide of the nicked strand, contributing to RNA specificity (Table 3.2). Additionally, the guanidinium moiety of R309 bridges the penultimate and antepenultimate RNA residues on the 5'PO_4 side of the nick by forming two stable hydrogen bonds with two non-bridging phosphate oxygens. The lack of RNA-specific interactions with R309 suggests that this residue may confer stability on the bound state without affecting substrate specificity.

Aside from the hypothesis that the 3'OH nucleic acid adjacent to the nick must take on the RNA-like 3' endo conformation for catalysis to proceed, the obligate presence

of RNA on the 3'OH side of the nick may also stem from specific RNA-protein interactions that possibly impart stability to the bound complex. The 2'OH of the terminal nucleotide on the 3'OH side of the nick interacts with the carboxyl group of E119 and attenuates fluctuations of the terminal dsRNA, thereby conferring stability to the bound complex (Table 3.2). N120 also provides stabilizing interactions with the RNA phosphate backbone contributing to RNA stability, though probably not specificity (Table 3.2). Two phenylalanine residues, F121, F122 contribute non-specific van der Waals interactions with adjacent ribonucleotides in the gRNA strand on the 3'OH side of the nick.

Mutational religation assays carried out on T4 Rnl2 suggest important roles for both F121 and F122; the F121 homolog is an important contributor to non-specific polynucleotide stabilizing interactions, and the F122 homolog is important in conferring RNA specificity. Additionally, the DNA/RNA-T4 Rnl2 crystal structure shows both the F121 and F122 homologs intercalating the minor groove[7]. In light of the experimentally determined importance of the F121 and F122 homologs and their crystallographically determined positions, which our initial homology model mimicked, it is appropriate to explore why F121 and F122 were not found intercalating the minor groove during the equilibrated, dynamics phase of the trajectory. Roughly 21 ns into the equilibration phase of the simulation, the dsRNA on the 3'OH side of the nick moved away from the protein surface. When it returned, approximately 2 ns later, F121 and F122 no longer penetrated the minor groove. Instead, they made non-specific van der Waals contacts with the two adjacent ribofuranose residues that border the 3'OH side of the minor groove

in which F121 and F122 were originally positioned. While these new positions of F121 and F122 are conserved throughout the remainder of the simulation, approximately 29 ns into the dynamics phase of the simulation the dsRNA on the 3'OH side of the nick again moved away from the protein. These structural fluctuations indicate that the dsRNA undergoes occasional "breathing" motions, moving toward and away from the protein on the nanosecond timescale. During these breathing motions, F121 and F122 may move from positions where they intercalate the minor groove to positions where they make non-specific van der Waals interactions with the ribofuranose moieties flanking the minor groove. This suggestion is consistent with polynucleotide temperature (Debye-Waller) factors reported for the DNA/RNA hybrid in the T4 Rnl2 complex[7] (Supporting Information Fig. 3.8).

3.3.6 RNA Binding Through a Unique Kinetoplastid Insert Region

The kinetoplastids share a unique insert region that roughly spans *Tb*REL1 residues 165 to 200, depending on the exact alignment[7, 48, 68] (Fig. 3.1). Two regions of this insert lack defined secondary structure and contain positively charged residues that were postulated to be involved in RNA binding[48]. Based on a homology model to the T4 Rnl2 structure, Nandakumar et al. hypothesized that the loops may rearrange or adopt secondary structure and help stabilize the 3'OH side of the nicked dsRNA[7]. More recently, ensemble averaged electrostatic calculations over a 20 ns molecular dynamics simulation of the ATP-bound *Tb*REL1 protein without RNA indicated a large lobe of positive electrostatic potential near this region, further substantiating the suggested RNA-

binding role[68]. The electrostatic calculations also suggested that charged residues K166, K172, K175, and/or R194 could play important dsRNA-stabilizing roles[68].

In the current work, we directly monitor the dynamics of this unique *Tb*REL1 insert region and assess its dsRNA-binding capacity. Through a set of transient interactions, the insert region significantly impacts the mobility of the gRNA strand, thus providing a secondary stabilization effect on the nicked 3'OH strand (Figs. 3.5 and 3.6). Despite the prediction that K166, K172, and K175 would play important dsRNA stabilizing roles, the current simulation shows that none of these residues directly contact the RNA. However, because they are commensurate with a large lobe of positive potential[68], they most likely participate in providing the electrostatic environment responsible for directing the diffusional association of the RNA substrate into a productive bimolecular association complex. R194, on the other hand, directly contacts the RNA, forming a persistent salt-bridge with the non-bridging phosphate oxygen atoms of the gRNA strand backbone (Table 3.2). Despite their positive charge, K183 and K184, two residues positioned adjacent to the terminal end of the 3'OH strand, do not make persistent salt bridges or hydrogen bonds with the dsRNA. The RNA/DNA hybrid bound to T4 Rnl2, which was used to construct the simulated model, likewise does not take advantage of the high positive electrostatic potential immediately adjacent to this region; thus we note that the observed behavior may be an artifact of our homology model. Yet, the stability conferred by the kinetoplastid insert on the bound RNA is apparent (Fig. 3.6). The kinetoplastid insert contacts the gRNA sequence from positions +2 to +4,

corresponding to a 0.5 Å drop in RMSF, and contacts the nicked 5'PO_4 strand at positions +10 and +11, attenuating fluctuations at the 3'OH terminal.

3.4 Conclusions

The current molecular dynamics simulation of *Tb*REL1 in complex with a dsRNA substrate has clarified the roles of active-site residues during step-three of catalysis and allowed us to suggest a possible deprotonation pathway involving a non-bridging oxygen of the AppN intermediate. In an effort to help clarify the number of metals required for catalysis, a Mg^{2+} was positioned within the active site based on homology with other superfamily members. By examining the relation of the Mg^{2+} to the reactive residues, we inferred that Mg^{2+} is likely to confer transition-state stabilization during nicked-strand religation and may also help stabilize the ribofuranose of the AppN intermediate. Both of these functional aspects have the potential to impart greater catalytic efficiency; as a result, we conclude that the *Tb*REL1 mechanism likely requires two Mg^{2+} for optimal efficiency. Additionally, by examining the persistence contact times between the protein and RNA during our simulation, we were able to explain, at least in part, why *Tb*REL1 requires RNA on the 5'PO_4 side of the nick *in vitro*[47] and to clarify the role of the unique kinetoplastid insert in RNA binding.

Our analysis allows us to make several predictions that may motivate future experimental work, adding to a growing body of knowledge concerning the structure-

function relationship of RNA-binding residues within this enzyme. The side chains of D62, S65, S67, and Q70 make RNA-specific contacts; as a result, we predict that mutations to these residues will adversely affect the religation rate by disrupting RNA binding. Additionally, the recognition sequence formed by D62, L63, P64, S65, S67, and Q70 collectively form a smooth surface that intercalates the minor grove; consequently, we predict that mutating any of these residues to a large, bulky residue like tryptophan will perturb the protein-surface continuity and attenuate the religation rate by lowering RNA affinity. Positively charged K166, K172, and K175 reside within the unique kinetoplastid insert loop but do not directly contact the RNA. Instead, they are predicted to contribute to the electrostatic environment responsible for mediating RNA association. As a result, we predict that mutating K166, K172, and K175 to alanine will lower the religation activity by decreasing the total charge on the protein. We note that while step-one autoadenylation requires ATP binding, the net charge on ATP is significantly less than that carried by dsRNA, and as a result step-one activity will be significantly less influenced by the decrease in total protein charge.

3.5 Supporting Information

3.5.1 Nucleic-Acid Flexibility

The dsRNA exhibits variable flexibility, as measured by the RMSF of the phosphate atoms (Fig. 3.6). The strand ends are the most flexible. While base pairing between the terminal nucleic acids on the 3'OH side of the nick persists over the course

of the simulation, terminal A-T base-pairing at the 5'PO_4 side of the nick was frequently interrupted, and large fraying events were observed. Fraying events at the dsRNA terminus on the 5'PO_4 side of the nick are reflected in the relatively large RMSF values for positions -9 through -11 on the gRNA strand and position -10 in the nicked strand (Fig. 3.6). The large decrease in fluctuations observed in the gRNA strand from positions -8 to -7 is the result of interactions with the recognition sequence that intercalates the minor groove and confers substrate stability. Moving from the -7 position toward the 3'OH terminal of the gRNA sequence, the RNA helix turn positions the gRNA strand away from the protein, preventing stabilizing interactions and giving rise to the increase in RMSF observed from positions -6 to -2. The decrease in the RMSF of the gRNA strand from positions -1 to 4 is a result of secondary stabilization effects conferred by base-pair coupling to the nicked strands, whose motion is damped by interactions at the periphery of the active site and with the unique kinetoplastid insert region. From positions 4 to 12, the turn of the RNA helix structures the gRNA strand away from the protein surface, preventing the formation of stabilizing interactions and giving rise to increased RMSF. As expected, the nicked strands have little flexibility near the active site where they both come into close contact with the protein surface. The relatively higher RMSF values of the 5'PO_4 nicked strand are attributable to the slightly less extensive interactions on the 3'OH side of the nick (Fig. 3.6).

While the RMSF values provide information about the magnitude of positional fluctuation, they give no information about the direction. To gain an understanding of the relative direction of motion exhibited by each strand during the simulation, a principal

component analysis was performed on the RNA phosphorus atoms (Fig. 3.6). Mode one, which accounts for the largest percentage of the total variance observed during the simulation, illustrates that the terminal ends move away from the long axis of the helix, while the central regions of the helix move roughly parallel to the helical axes but in opposite directions on either side of the active site. The largest amount of motion along mode two is concentrated on the 5'PO_4 nicked strand and describes a bending motion, hinged roughly at the active site, which brings the RNA strand toward the protein surface. Motion along mode three is approximately equally distributed along the length of the RNA strand and describes a contraction of the RNA produced by movement of the strands on either side of the active site toward one another (Fig. 3.6).

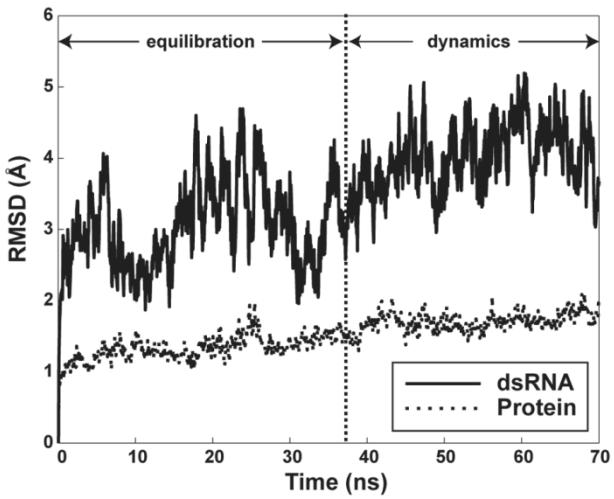


Figure 3.7: RMSD of the protein and RNA strands over the 70 ns simulation.

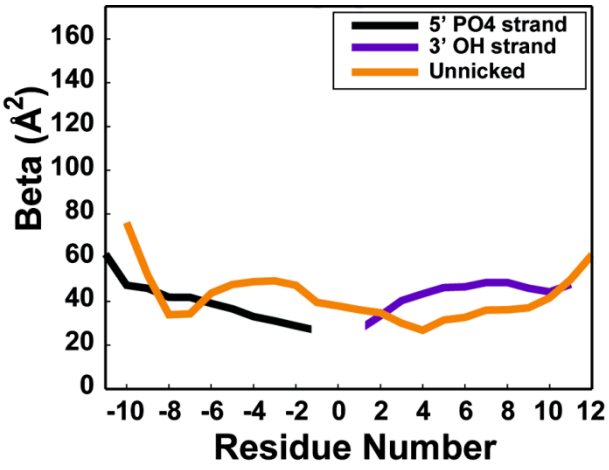


Figure 3.8: Experimental temperature (Debye-Waller or beta) factors.

This chapter, in full, is a reprint of "Toward Understanding the Conformational Dynamics of RNA Ligation", which was published in 2009 in Biochemistry, volume 48, issue, 4, pages 709 to 719, by Robert V. Swift, Jacob Durrant, Rommie E. Amaro, and J. Andrew McCammon. This chapter was included with the permission of Jacob Durrant, Rommie E. Amaro, and J. Andrew McCammon.

Chapter 4

Catalytically Requisite Conformational Dynamics in the mRNA Capping Enzyme Probed by Targeted Molecular Dynamics

Abstract

The addition of a N7-methyl guanosine cap to the 5' end of nascent mRNA is carried out by the mRNA capping enzyme, a two domain protein that is a member of the nucleotidyl transferase superfamily. The mRNA capping enzyme is composed of a catalytic nucelotidyl transferase domain and a non catalytic OB domain. Large scale domain motion triggered by substrate binding mediates catalytically requisite conformational rearrangement of the GTP substrate prior to the chemical step. In this study we employ Targeted Molecular Dynamics (TMD) on the PBCV-1 capping enzyme to probe the global domain dynamics and internal dynamics of conserved residues during the conformational transformation from the open to the closed state. Analysis of the resulting trajectories along with structural and sequence homology to other members of the superfamily allows us to suggest a conserved mechanism of conformational rearrangements spanning all mRNA capping enzymes and all ATP-dependent DNA ligases. Our results suggest that the OB domain moves quasi statically toward the

nucleotidyl transferase domain, pivoting about a short linker region. The approach of the OB domain brings a conserved RxDK sequence within proximity of GTP's triphosphate destabilizing, the unreactive conformation and thereby allowing thermal fluctuations to partition the substrate toward the catalytically competent state.

4.1 Introduction

Conformational flexibility plays an essential role in regulating the functional activity of many proteins and enzymes. The magnitude of protein motion ranges from small local atomic vibrations that occur on a femto- to picosecond time scale, to side chain rotation that occurs on a nanosecond or longer timescale, to larger global domain motion that typically occurs on the time scale of micro- to milliseconds. It is often these large scale global dynamics that play an intrinsically important role in regulating protein and enzyme activity. For instance, global domain motions may be induced by an allosteric effector, influencing downstream functional events. Kern and Zuiderweg[83] offer an excellent review of these types of motions. Alternatively, in enzymatic catalysis, substrate binding may trigger domain closure over the active site resulting in the formation of favorable interactions that facilitate catalysis. This mechanism of substrate-induced protein motion is known as induced fit and was first described by Koshland in 1958[84].

Structural and experimental evidence suggests that mRNA capping enzymes undergo catalytically requisite global domain motion consistent with the induced fit

mechanism prior to the chemical step. These enzymes carry out the second in a set of three enzymatic reactions that result in the formation of a N-7-methyl-GMP cap on the 5' end of nascent mRNA in eukaryotes and their viruses[85]. In the first of these three steps the 5'-phosphate is cleaved from mRNA by a triphosphatase. The resulting 5'-diphosphate end is guanylated by an mRNA capping enzyme in the second step. Finally guanosine's N-7 is methylated by a methyltransferase. Formation of the 5'-cap is the first post-transcriptional processing step and plays an essential role in mRNA splicing, nuclear export, mRNA stability, and translation initiation[86-88]. Additionally, mutational analysis of the encoding gene has demonstrated that the *Saccharomyces cervisiae* capping enzyme is essential in maintaining cell viability[9].

mRNA capping enzymes belong to a larger superfamily of enzymes called the nucleotidyltransferase superfamily whose members include ATP and NAD^+ dependent DNA ligases, ATP dependent RNA ligases, and mRNA capping enzymes. Members are distinguishable by their common structural elements and chemistry. Common structural features include a catalytic N-terminal domain, called the nucleotidyltransferase domain, and a non-catalytic C-terminal domain[1]. The nucleotidyl transferase domain is common to all superfamily members and contains five conserved motifs, I, III, IIIa, IV and V, that form the substrate binding pocket, while the structural elements of the C-terminal domain differ among members. In mRNA capping enzymes and ATP-dependent DNA ligases, the C-terminal domain contains a OB-fold that includes a sixth motif implicated in intermediate formation during catalysis[3] (figure 4.1). All members share a common chemical mechanism whereby a nucleotide is transferred to the 5'-end of a

polynucleotide via a covalent nucleotide-enzyme intermediate[1]. In the mRNA capping enzyme, chemistry is initiated when a conserved lysine in motif I attacks GTP's alpha phosphate, concertedly displacing pyrophosphate and resulting in the formation of a GMP-lys intermediate. Following mRNA binding, GMP is transferred to the 5'-end of the bound mRNA, completing the reaction.

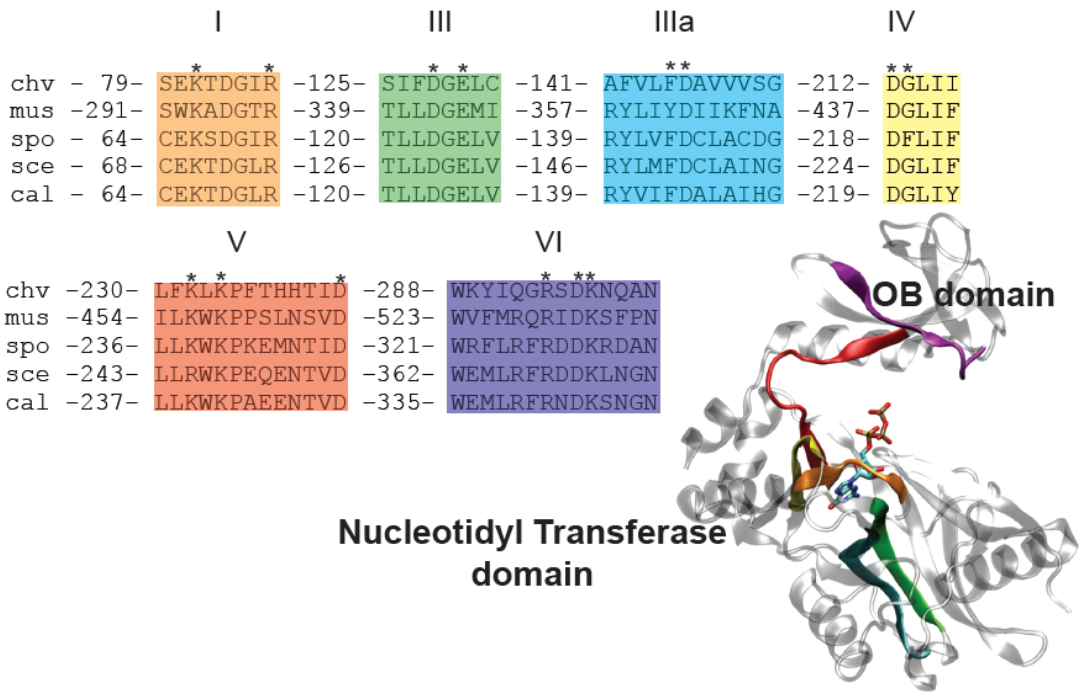


Figure 4.1: Conserved motifs in capping enzymes. Alignments include Chlorella virus PBCV1 (chv), mouse (mus), S. pombe (spo), S. cerevisiae (sce) and C. albicans (cal). The amino acid number preceding the motif is given by the number preceding the motif. Essential residues are denoted with a '*'. The structural locations of the motifs are mapped by color and the N and C-terminal domains are labeled Nucleotidyl transferase domain and OB domain respectively.

Large global rearrangements of the OB-domain prior to the chemical step in catalysis have been implicated by structural evidence[4]. The mRNA capping enzyme from *Paramecium bursaria Chlorella* virus, PBCV-1, has been crystallized with GTP

bound in the active site in two distinct conformations: “open” and “closed” (figure 4.2). In the open state, the OB domain is distal from the nucleotidyl transferase domain: the centers of mass of the two domains are separated by 30.3 Å. The triphosphate of GTP is in an un-reactive conformation, nearly orthogonal to the nucleophilic lysine. In the closed state, the OB-domain is proximal to the active site and the triphosphate has undergone conformational isomerization: the centers of mass of the two domains are separated by 25.5 Å, and the triphosphate is in the catalytically requiste in-line conformation. Subsequent CD and fluorescence studies carried out on the yeast capping enzyme demonstrated that two distinct conformational states exist in the presence and absence of GTP[14]. These results bolster the claim that global rearrangements are an intrinsic part of the capping enzyme’s catalytic repertoire. Additionally, mutational analysis of the mouse capping enzyme identified 16 amino acids essential for activity, fourteen of which were localized in the six conserved motifs[10] (figure 4.1). These results validated an earlier mutational study of the *S. cerevisiae* capping enzyme[48].

While structural and experimental studies have made great progress in illuminating the requisite conformational rearrangements that occur within the mRNA capping enzymes during catalysis, our knowledge of these rearrangements is incomplete. A broader understanding of the process requires knowledge of the internal dynamics that occur during the conformational transition from the open to the closed state. Consequently, the goal of this study is to understand these dynamics. In particular we hope to identify those amino acids essential in mediating the transition from the open, un-reactive state to the closed, reactive state and determine if these residues are consistent

with essential conserved residues found during mutational analysis of the mouse and *S. cerevisiae* capping enzymes[10, 48].

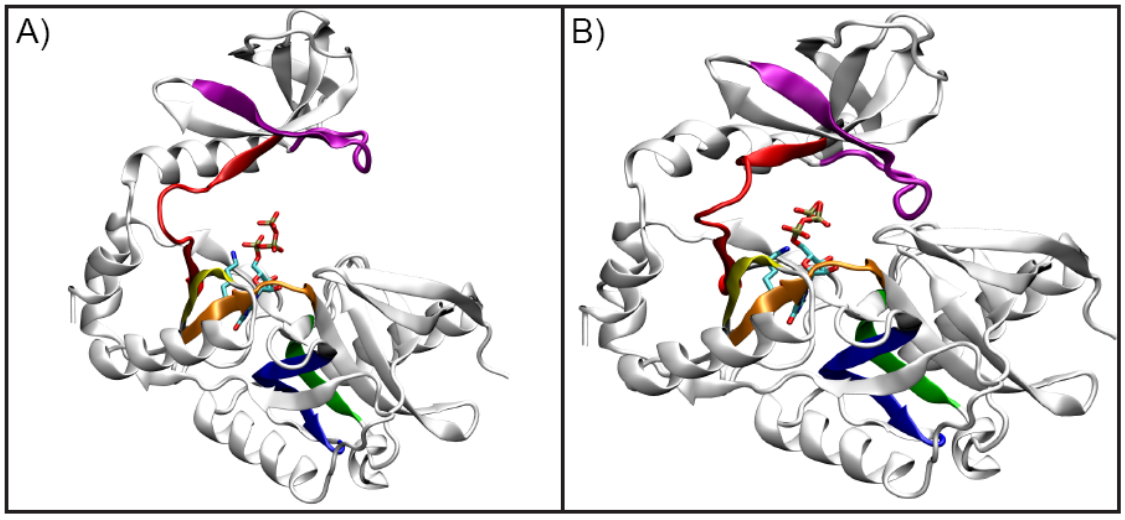


Figure 4.2: Open and closed crystal structures. A) The open crystal structure illustrating the unreactive triphosphate conformation. The motifs are colored as in figure 20 and GTP and nucleophilic K82 are rendered in licorice. The centers of mass of the two domains are separated by 30.3 Å. B) The closed crystal structure illustrating the reactive conformation. The centers of mass of the two domains are separated by 25.5 Å.

Internal protein dynamics may be probed by numerically integrating the equations of motion arising from a protein parameterized with a classical force field, a technique called molecular dynamics[23]. In practice the fastest atomic vibrations limit the size of the numerical time-steps to 1 to 2 femtoseconds. That puts a practical restriction on simulation length to the tens to hundreds of nanoseconds, preventing observation of global domain motions that occur on the micro- to millisecond time scale. In order to access long time scale motions, an energetic bias applied along a reaction coordinate can be added to the classical force field, resulting in forces that drive the enzyme from one conformational state to another. In targeted molecular dynamics (TMD) the energetic

bias is applied in a time-dependent fashion giving rise to a force that quasi-statically decreases the root mean square displacement (RMSD) between a moving and a target structure. While TMD allows us to observe global domain motion during a computationally accessible time scale, it suffers from two drawbacks.

First, there is no guarantee that the minimum potential or free energy pathway will be followed. Changing the initial conditions may result in different pathways. Additionally, thermal fluctuations lead to statistical uncertainty for the pathway determined from any given set of initial conditions[89]. Other, more rigorous chain of state methods, such as the nudged elastic band[90, 91] and the harmonic Fourier bead[92] methods, optimize object functions connecting multiple images of the system between two end points to yield an optimal path. While these methods are more likely to generate the minimum potential, or free energy path, they are more complex to perform, and are computationally more expensive than TMD. A simpler alternative to the chain of states methods, steered molecular dynamics SMD[31], applies a time-dependent linear restraint to the center of mass of a group of atoms, but like TMD is not guaranteed to move the system along the lowest potential or free energy pathway. Furthermore, SMD requires one to define a reasonable reaction coordinate describing the molecular mechanism of interest. This can be a difficult in the high dimensional space occupied by proteins and enzymes.

Second, TMD does not immediately yield the pathway's free energy. If one wishes to calculate a pathway's free energy, a computationally more intensive technique

such as umbrella sampling must be used[93]. The work of Ravindranathan, Gallicchio and Levy provides an excellent example of the umbrella sampling technique to calculate the free energy of domain closure in the two domain ribose binding protein.

Despite TMD's shortcomings, its simple implementation and computational efficiency make it an attractive method to explore the qualitative characteristics of conformational isomerization. Since it was pioneered by Schlitter and coworkers[32], TMD has been used to model conformational transitions in a variety of systems. A few recent examples include the activation of the CCK2R G-protein coupled receptor[94], the coupling of C-loop closure and channel gating in the nicotinic acetylcholine receptor[95], and the activation mechanisms of gelsolin, an actin polymerization regulatory protein[8]. In this work, TMD was used to study the internal dynamics that occur as the PBCV-1 capping enzyme moves from the open, unreactive state to the closed, reactive state.

4.2 **Materials and Methods**

Structural coordinates for the open, unreactive and closed, reactive states were taken from chain A and B, respectively, of the PDB 1CKM. Both the open and closed structures consist of the crystallographically resolved residues 11 – 327. All crystallographic waters were retained in the simulation. Histidine protonation states were specified for the open state using the program WHATIF[73]. The results were manually verified, each protonation state agreeing well with the surrounding micro environment. Because TMD requires that the molecular composition be the same for both the moving

and target structures, the protonation states of the closed target conformation were assigned to match those of the open state. All other hydrogens and protein force field parameters were assigned for both states in AMBER8's xleap module according to the AMBER94 force field[24]. The Carlson parameters were assigned to GTP[74], and both states were immersed in a 10 Å solvent octahedron of pre-equilibrated TIP3P waters[26]. Three sodium ions were added to neutralize the system charge. Each electrically neutral, solvated system comprised an identical number of 40,495 atoms. In order to remove spurious crystal packing contacts, a set of 26,000 minimization steps were independently carried out on both the open and closed states in the Sander module of Amber 8[71]. The first 6,000 steps were performed in three 2,000 step cycles. Hydrogen was relaxed during the first 2,000 steps holding all other atoms fixed. Hydrogen, water and ions were relaxed during the next 2,000 steps. In the last cycle, the protein backbone was held fixed, minimizing all other atoms. Finally, a constraint free minimization was carried out over the last 20,000 steps of which the first 19,500 were steepest descent, while the last 500 used the conjugate gradient algorithm. Following minimization, each system was restrained harmonically with a 5 kcal/mol-$Å^2$ force constant and equilibrated at 300K at constant volume for 30 picoseconds using the langevin thermostat to maintain temperature[96] and the SHAKE algorithm to constrain hydrogens bound to heavy atoms[97]. The restraints were subsequently released and both systems were then equilibrated at 1 atmosphere for an additional 50 picoseconds using a Berendsen barostat[98]. Targeted molecular dynamics was then carried out at NPT as implemented in Amber 8's sander by applying the following time-dependent, harmonic energy bias to all alpha carbons and GTP atoms in the open state (Eq. 4.1):

$$U_{TMD} = \frac{1}{2} Nk(RMSD - \rho(t))^2 \qquad 4.1$$

N gives the number of atoms included in the bias, *k* is the harmonic force constant, *RMSD* is the root mean square distance between a configuration at time point *t* and the target configuration. The reference RMSD value at time *t* is given by *ρ(t)*. By monotonically decreasing *ρ(t)* the moving structure is gradually driven toward the target structure. A truly quasi-static equilibrium transformation occurs in the limit as the force constant goes to zero and the time goes to infinity, consequently smaller force constants and longer simulation times represent more accurate models but come at a greater computational cost. In order to systematically choose the best balance between efficiency and accuracy, four independent simulations with *k* vaues of 5, 4, 3, and 1.5 kcal/mol-Å^2 were carried out over one nanosecond each. We inspected the trajectories in VMD[99] and observed that all four simulations approached the target structure. Desiring the lowest harmonic force constant that produced the sought after transformation, the 1.5 kcal/mol-Å^2 force constant was selected and the simulation time was extended to two nanoseconds. SHAKE was used during the TMD simulation to constrain all hydrogen-heavy atom bonds. Simulations were carried out on the San Diego Supercomputer Center's DataStar machine saving snapshots every 500 fs for analysis. Post-simulation analysis was performed in VMD and Matlab with a variety of customized scripts.

4.3 Results

4.3.1 Quasi-Rigid Domain Motion During the Open to Closed Transformation

TMD was used to probe the internal protein dynamics in the PBCV-1 capping enzyme during the transition from the open, unreactive conformation to the closed, reactive conformation. Figure 4.3A shows a profile of the RMSD as a function of simulation progress. For a given time, *t*, we define progress as the fraction of total simulation length contained in the interval [0, *t*]. It is clear that the profile is a decreasing function without any jumps or discontinuities indicating a smooth transition from the initial to final states. In order to explore the global domain dynamics during the transition between states, the alpha-carbon RMSD of the OB domain was calculated following alignment to the first snapshot of the OB domain and again following a rigid body protein transformation according to the alignment matrix that minimized the RMSD between trajectory snapshots of the nucleotidyltransferase domain and the first snapshot of the nucleotidyltransferase domain. The alpha-carbon RMSD of the nucleotidyltransferase domain was also calculated following alignment to the first snapshot of the nucleotidyltransferase domain. The results are shown in figure 4.3B. Figure 4.3C shows the distance between the centers of mass of the OB and nucleotidyltransferase domains as a function of simulation progress, illustrating that the domains approach one another. These results suggest that both the OB and nucleotidyltransferase domains move as quasi-rigid bodies toward one another about a hinge in the short region linking the two domains. Similar rigid body rotations have been

observed in other two-domain proteins such as adenylate kinase[100], ribose binding protein[93] and transferrin[101] and are reviewed by Gerstein, Lesk and Chothia[102]. Thus the observed quasi-rigid body translations are in good agreement with other structurally similar proteins.

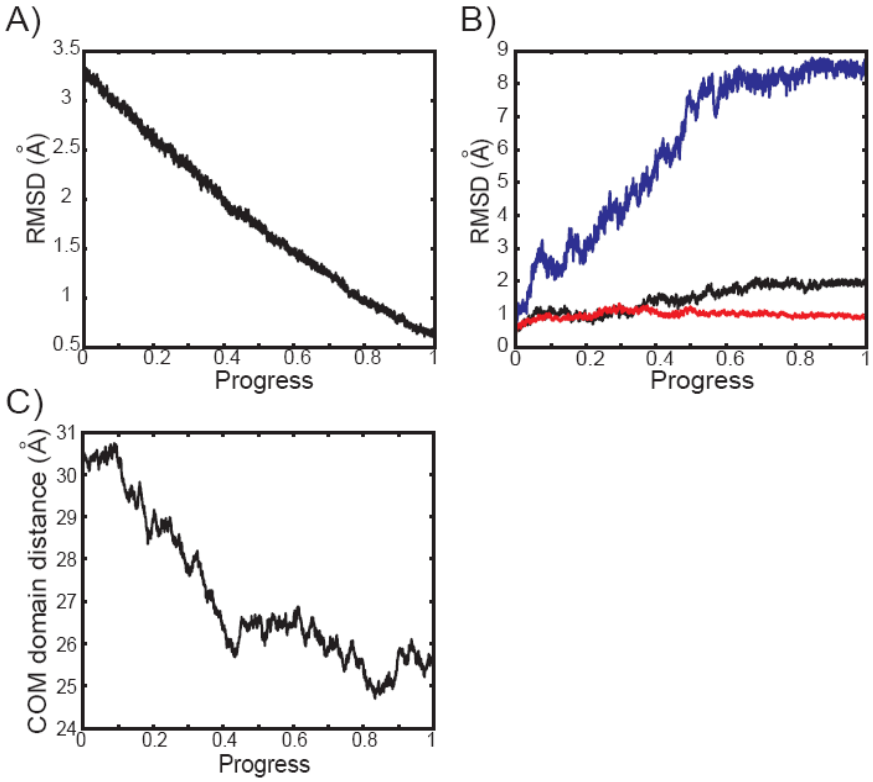


Figure 4.3: RMSD and center of mass to center of mass domain distance. A) Total alpha carbon and GTP RMSD. B) Blue: RMSD of OB domain following structure alignment to reference nucleotidyltransferase domain. Black: RMSD of OB domain following alignment of OB domain to reference OB domain. Red: RMSD of nucleotidyltransferase domain following alignment to reference nucleotidyltransferase domain. C) Distance between centers of mass of the OB and nucleotidyltransferase domains.

4.3.2 A Conserved Active-Site Network Mediates Triphosphate Rearrangement

Figure 4.4 shows four snapshots taken at discontinuously spaced intervals over the trajectory that illustrate key interactions that the enzyme may utilize to stabilize intermediate conformations en route to the reactive state. In figure 4.4A, the simulation has progressed 4 percent toward completion and the enzyme is largely in the same conformational state observed in the open crystal structure. Stabilized by interactions between the gamma-phosphate oxygens and K236, the beta-phosphate oxygens and K82, and the alpha-phoshpate oxygens and K234, the triphosphate is folded back over the K82 nucleophile in an unreactive conformation. The OB domain is distal from the active site in an open conformation.

Figure 4.4B captures a representative conformation 60 percent into the simulation. The OB domain has moved into proximity of the active site bringing R295 and K298 into proximity of the triphosphate tail, which has rotated away from its initial state into a nearly vertical orientation. In this upright position it interacts with K82, K234 and K236 in the nucleotidyl transferase domain, and R295 and K298 in the OB domain.

As the simulation progresses, the triphosphate moves through the vertical orientation observed in figure 4.4B, and into a new state, shown in figure 4.4C. This new state, occurring 78 percent into the simulation, is characterized by interactions between the gamma-phosphate and K298 in the OB domain, and the alpha-phosphate and K234 in the nucleotidyltransferase domain.

The final stages of the simulation entail further rearrangement of the triphosphate tail into the catalytically requisite in-line conformation. A representative conformation of this late phase is shown in figure 4.4D. The simulation has progressed 90 percent to completion, the bulk of the triphosphate rearrangement has occurred, but the final in-line conformation has not been attained.

While figure 4.4 provides a good pictorial representation of the different stages involved in the conformational transformation, examining GTP-side chain bond distances of those residues involved as functions of simulation progress permits a more detailed analysis. These bond distances are illustrated in figure 4.5. Figure 4.5A charts the distance from the K82 nucleophile to the alpha-phosphate electrophile as a function of simulation progress. The distance gradually increases over the first 70 percent of the simulation before discontinuously decreasing as a result of a rotation about the bond connecting ribosyl's O5' oxygen to the alpha-phosphate. The distance decreases over the last 10 percent of the simulation as the triphosphate group attains its final inline conformation. The anomalous large spike that occurs 40 percent into the distance time series results from a rotation about the alpha-phosphate O5' bond. This rotation briefly swings the alpha-phosphate away from K82 before returning to its equilibrium position. Because formation of the enzyme-guanylate intermediate is facilitated by the increasing proximity between the K82 nucleophile and the alpha-phosphate electrophile, the approach of these two atoms over the last 20 percent of the trajectory is consistent with movement into the catalytically competent state.

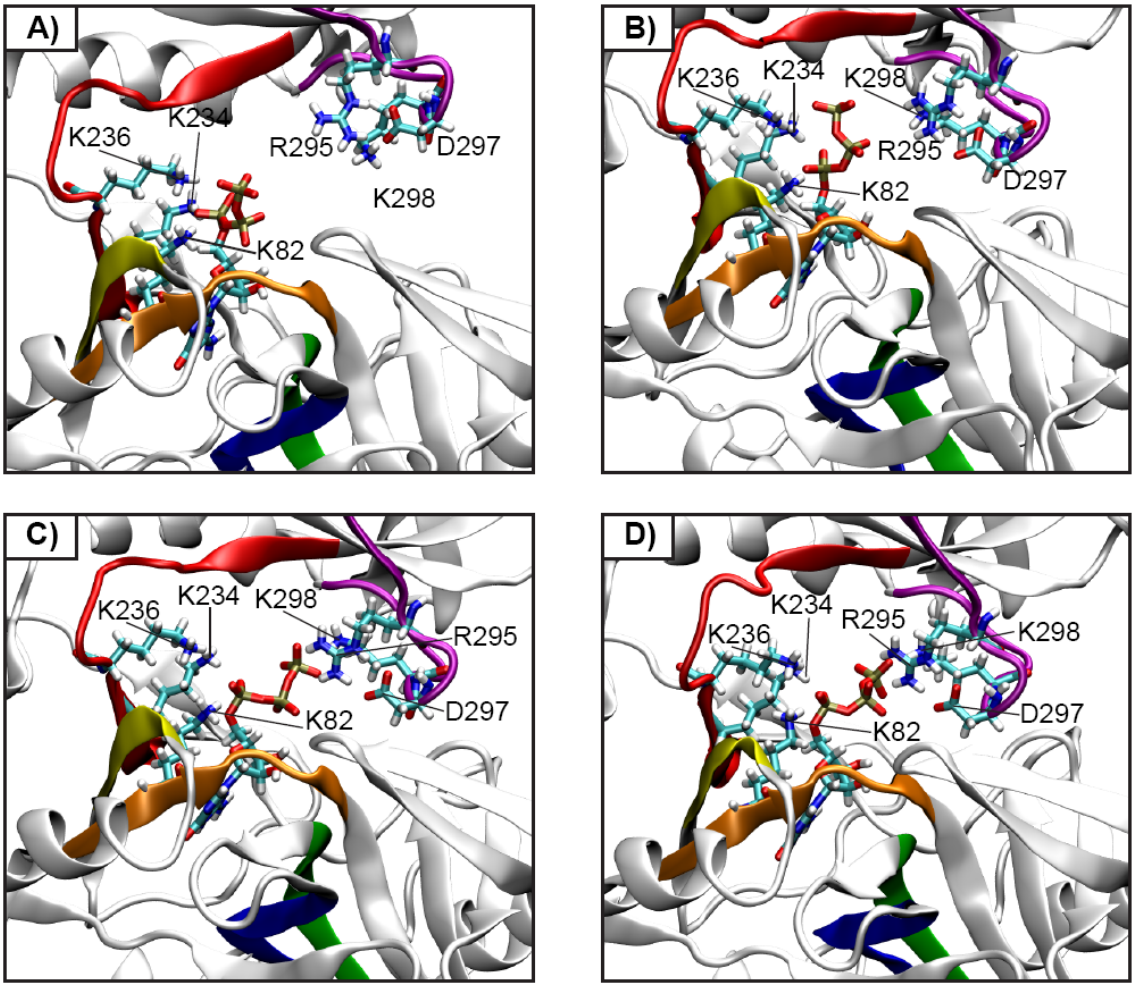


Figure 4.4: Discontinuously spaced snapshots observed during the targeted molecular dynamics simulation at A) 4% simulation progress, B) 60% simulation progress, C) 78% simulation progress, D) 90% simulation progress.

In order to move toward the reactive, closed state, hydrogen bonds stabilizing the enzyme-substrate complex in the unreactive conformation must rupture and transiently be replaced by interactions that allow thermal fluctuations to carry the triphosphate tail of GTP into its final, reactive conformation. Consideration of figures 4.5B through 4.5D gives qualitative insight into how this conformational evolution might progress. Key events that will now be discussed occur 60 percent into the simulation and are pictorially represented in figure 4.4B. Figure 4.5B shows the displacement between the

guanidinium nitrogen of R295 and the non-bridging beta- and gamma-phosphate oxygen atoms. The distances are initially correlated to the displacements separating the centers of mass of the two domains, rapidly decreasing over the first 30 percent of the simulation to meta-stable equilibrium distances of roughly 5.5 Å that are maintained until the simulation is 80 percent complete. Figure 4.5C shows a similar progression for the distance between the amine of K298 and O3G, a non-bridging gamma-phosphate oxygen. Figure 4.5D shows that the distance between K234's amine and O2G decreases over the first 60 percent of the simulation. This decrease is attributable to both the triphosphate tail rotating toward the sidechain and the attractive electrostatic force that result from the approach of the negatively charged triphosphate oxygen toward the positively charged terminal amine of K234. At 60 percent progress, the rapid decrease in the distance between the terminal amine of K234 and O2G is anti-correlated with a 2 Å increase in the distance between K236 and O2G (figure 4.5E) indicating that O2G exchanges hydrogen bonding partners, replacing K236 with K234. Following hydrogen bond exchange, there is brief period of stability that extends as the simulation progresses from 60 to 70 percent completion. After this transient period of stability, both distances increase as R295 and K298 form energetically favorable interactions with triphosphate as it moves into an in-line conformation with respect to the nucleophile.

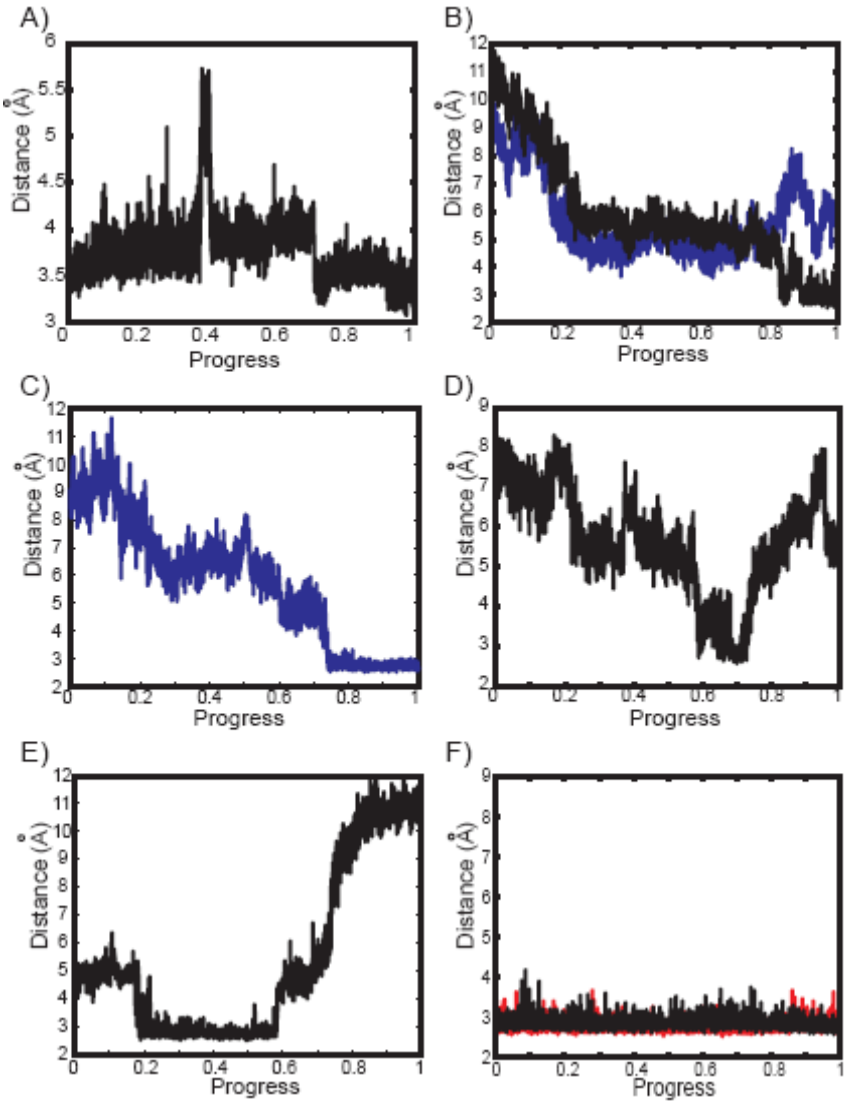


Figure 4.5: Active-site interaction dynamics. A) Distance between K82 nucleophile and alpha-phosphate electrophile. B) Black: distance between R295's NH2 and GTP's O2B. Blue: distance between R295's NH1 and GTP's O1G. C) Distance between K298's side chain nitrogen and GTP's O3G. D) Distance between K234's side chain nitrogen and GTP's O2G. E) Distance between K236's side chain nitrogen and GTP's O2G. F) Blue: distance between K298's side chain nitrogen and D297's OD2. Black: distance between R295's NE2 and D297's OD2. Red: distance between R295's NH2 and D297's OD2.

Figure 4.5F demonstrates the evolution of the interactions between the carboxyl of D297 and the guanidinium of R295 as well as the amine of K298. The red and black curves show that interactions between both nitrogens of the guanidinium moiety of R295

and the carboxyl group of D297 are conserved over the course of the simulation. As a result of these two strong, short interactions, the configurational space available to the R295 sidechain is restrained, limited by the smaller volume of configurational space accessible to the shorter D297 sidechain. As a result, the interacting pair samples a smaller volume of configuration space than either would sample independently. The reduced mobility allows both the guanidinum nitrogens of R295 not involved in contacting the carboxyl group of D297 to readily form two contacts with non-bridging oxygens of the triphosphate tail of GTP (figure 4.5B).

4.3.3 Nucleotidyltransferase Domain Guanosine Stabilization

As illustrated in figure 20, the GTP binding pocket is lined by a set of 5 conserved motifs. Several residues within these motifs make key interactions with GTP's guanosine group that maintain the relative positions and orientations of the ribofuranose ring and guanine base as the OB domain closes over the active site and the triphosphate rotates into a reactive conformation (figure 4.6). Motif IIIa and IV each contribute a hydrophobic side chain that interacts with the guanine forming a stabilizing hydrophobic sandwich. Motiff IIIa contributes F146 that forms pi-stacking interactions with the aromatic guanine. On the opposite side of the guanine plane, I216 is contributed by motif IV. Table 4.1 shows that these interactions undergo small, ~0.2 Å fluctuations around their mean signifying that they constitute a firm binding pocket. In addition the butyl amine of K188 protrudes into the deep end of the active site donating a hydrogen bond to GTP's O6 adding additional stabilization. K188 is independent of the six motifs but as

sequence analysis demonstrates is either conserved or substituted by a potential hydrogen bonding residue in the surveyed capping enzymes[3]. Although not as extensively stabilized as the guanine base, GTP's ribofuranose also makes stabilizing interactions with the surrounding active site. Extending from motif I, the NH1 of R87 donates a hydrogen bond to the O2' of ribofuranose. This interaction is complemented by E131, a motif III residue that accepts a hydrogen bond from ribofuranoses 2' hydroxyl hydrogen. As Table 4.1 demonstrates, both the R87 and E131 interactions are relatively stable around their means. Consequently, they form a steady interaction network with the ribofuranose moiety. The net effect of the five residues is to stabilize the guanosine group. This is a critically important role that constrains the electrophilic alpha-phosphate proximal to the K82 nucleophile thereby allowing the occurrence of the first chemical step following triphosphate conformational isomerization.

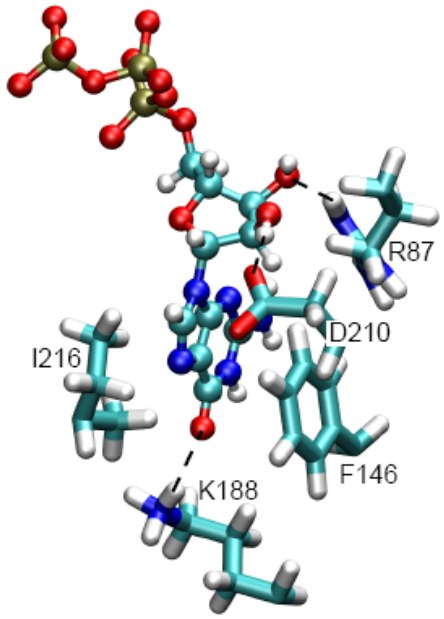


Figure 4.6: Protein-guanosine interactions that persist during the observed global domain motions. The representative conformation occurs 50% into the simulation. For the sake of clarity only protein side chains involved in conserved in interactions with guanonsine are shown.

Table 4.1: **Interactions between active-site residues and the guanosine moiety**. COM indicates Center of Mass. The root mean square deviation (RMSD) values were measured with respect to the first frame of the trajectory. The RMSD value reported is the average over the entire simulation reported in angstroms.

Motif	Protein	GTP	Mean (RMSD)
I	R87:NH1	Ribose: O2'	3.20 (0.44)
III	E131:OE1	Ribose: Ho3'	3.59 (0.53)
IIIa	F146: phenyl COM	Guanine: COM	3.83 (0.24)
IV	I216: center of mass	Guanine: COM	4.24 (0.218)
-	K188: NZ	Guanine: O6	2.93 (0.26)

4.4 Discussion

4.4.1 A Mechanism of Conformational Rearrangement

While the exchange of hydrogen bonding partners observed in the simulation is a result of an external bias applied to GTP, the exchange does provide a useful context within which a model of molecular events, rationalized in terms of only the electrostatic and thermal driving forces available *in vivo*, may be posited. OB domain approach toward the active site is triggered by GTP binding and results in the addition of three potential hydrogen bonds with the triphosphate tail of GTP: two from the guanidinium moiety of R295 and one from the terminal amine of K298. The addition of these interactions may destabilize the pre-existing hydrogen bond between the terminal amine of K236, and the non-bridging gamma-phosphate oxygen, O2G. The hydrogen bond destabilization is likely sufficient enough that thermal fluctuations will eventually rupture the bond thereby allowing thermal fluctuations to rotate the triphosphate toward the

reactive state. Figure 4.4B gives a qualitative, pictorial representation that may resemble the active site during this period of transition. Once the hydrogen bond with K236 is broken, two hydrogen bonds between the non-bridging oxygens on the beta- and gamma-phosphate and the guanidinium of R295 may stabilize the rotated state sufficiently to prevent re-equilibration into the unreactive ground state. Additional thermal fluctuations may result in further rotation of the triphosphate tail toward the terminal amines of K234 and K298. As the distance between these amines and the non-bridging gamma-phosphate oxygenes decreases, an attractive electrostatic force will arise, biasing the triphosphate to fluctuate into the final closed, reactive conformation. This putative model is consistent with independent mutational studies carried out by Sawya and Shuman[10] on the mouse capping enzyme, and Wang and co-workers on the *S. cerevisiae* capping enzyme[48]. Each of these studies showed that mutation of the R295 homolog to alanine, as well as the conservative replacement with lysine, eliminated wild type activity. Both arginine and lysine are positively charged, basic residues; consequently enzyme activity must require a bidenate hydrogen bond donor at this position. TMD simulation results suggest an explanation consistent with this requirement. While a positively charged lysine may be sufficient to destabilize the K236-O2G hydrogen bond, single coordination of the triphosphate is insufficient to stabilize the ensemble of intermediate conformers populated during isomerization.

According to our model of conformational isomerization, destabilization of the unreactive state is principally dependent on interactions between R295 and the triphosphate tail of GTP. Our simulations suggest that these interactions are facilitated

by two conformationally restraining hydrogen bonds between the guanidinuim moiety of R295 and the carboxyl group of D297. These short strong interactions restrain the volume of configurational space accessible to both residues and allow R295 to interact more readily with the triphosphate tail. Accordingly, even though D297 does not interact directly with GTP, it is critically important in influencing the course of the catalytically requisite conformational isomerization. This implication is in good agreement with mutational analysis performed by Sawaya and Shuman on the mouse capping enzyme[10], and Wang and coworkers on the *S. cerevisiae* capping enzyme[48]. These studies independently verified that in both the mouse and *S. cerevisiae* capping enzymes, mutation of the D297 homolog to alanine eliminated wild type activity. Moreover, each study showed that wild type activity could not be recovered by replacement of the D297 homolog with asaparagine and that the conservative replacement with glutamate only partially recovered wild type activity in the mouse capping enzyme. Hydrogen bonds cannot be made between the guanidinium moiety of R295 and the methyl group of alanine and only a single, weak hydrogen bond can be made with the acetamide group of asparagine. As a result, these studies imply that both hydrogen bond interactions between the two guanidinum nitrogens of the R295 moeity and the carboxyl group of either aspartate or glutamate are essential for enzyme activity. This is consistent with the suggested restraining role that D297 plays on R295 dynamics. Furthermore, our simulation results are bolstered by the partial recovery of enzyme activity brought about by conservative glutamate substitution. The glutamate mutant has three additional degrees of conformational freedom so that the glutamate-arginine pair is less restrained than in the wild type enzyme. The increased fluctuations of the guanidinium decrease its

availability to interact with the triphosphate tail of GTP thereby disfavoring conformational interconversion, resulting in decreased overall activity.

While the triphosphate tail of GTP is undergoing dynamic rearrangements, a critical network of conserved interactions identified as essential in alanine scans of the mouse and *S. cerevisiae* capping enzymes, stabilize the the guanosine moiety of GTP. Essential residues include R87 from motif I, E131 from motif III, F142 from motif IV, and K188. The stabilization that they confer is critical in maintaining the relative proximity of the alpha-phosphate of GTP to the terminal amine nucleophile of K82 during the conformational isomerization of the triphosphate tail of GTP.

4.4.2 C-terminal Domain Homology and Mechanistic Conservation

While all members of the nucleotidyl transferase superfamily are structurally characterized by a non-catalytic C-terminal domain and a catalytic N-terminal domain, the structural elements and catalytic role of the C terminal domain differ among superfamily members. Consequently, in order to put the results of this study in broader biological context, it is useful to briefly survey the similarities and differences of C-terminal domains among superfamily members while further dividing the superfamily into three subsets.

NAD^+ dependent DNA ligases constitute the first subset of enzymes and contain a structural domain, 1a, fused to the nucleotidyl transferase domain that interacts with the

NMN leaving group in a manner analogous to the interactions between motif VI and GTP in the capping enzymes[2]. ATP-dependent RNA ligases make up the second subset and can be further broken down into subtypes I and II by the absence of motifs III and IIIa in type I ligases[72]. The C-terminal domains in both types of ligase, are composed of helical bundles that differ in number and topology[7, 72]. Additionally, it was shown that the C-terminal domain in type II ligases is not essential for the formation of the nucleotide enzyme intermediate. The third subset is composed of ATP-dependent DNA ligases and mRNA capping enzymes. In ATP-dependent DNA ligases and mRNA capping enzymes, the C-terminal domain is made up of an OB-fold and a sixth conserved motif with a constituent RxDK sequence that is essential for the formation of the enzyme-nucleotide intermediate[3, 10, 103]. This categorization is in accord with the evolutionary mechanism of biochemical specificity among superfamily members proposed by Shuman and Lima[1]. In their proposal, contemporary biochemical specificity was attained by fusion of structurally different ancillary domains to an ancient fundamental catalytic unit, which is structurally and functionally homologous to the nucleotidyl transferase domain. This categorization allows us to infer that our proposed mechanism of structural rearrangement that occurs prior to the first step of catalysis is conserved across all eukaryotic capping enzymes and ATP-dependent DNA ligases. Our mechanism provides a dynamic context to rationalize the functional roles of residues deemed essential by mutational analysis and complements previous structural explanations. Additionally, the intimate relationship between structure and function implies that similarly conserved mechanisms likely span members of the other two superfamily subcategories.

4.5 Conclusions

The TMD methodology allowed us to generate an ensemble of structures that allowed us to posit a mechanism of conformational isomerization consistent with mutational analysis of the mouse and *S. cerevisiae* capping enzymes. The structural relationships between superfamily members allowed us to infer the generality of the mechanism to include all mRNA capping enzymes and ATP dependent DNA ligases. Despite these extensions to the existing body of knowledge regarding nucleotidyl transferase superfamily members, the study was limited to the first conformational isomerization step prior to the first chemical step. Furthermore, no attempt was made to quantitatively anchor the conformational rearrangements on a thermodynamic footing. Consequently, the study's limitations naturally prompt several additional questions that merit further investigation. First, what is the relationship between global domain motion and internal dynamics during active site remodeling prior to and following subsequent catalytic steps? The *syn*- to *anti*- rotation about the N-glycosidic bond of guanosine after the first chemical step is one example[1]. Second, what are the thermodynamic forces that drive these large global rearrangements and how are they affected by individual catalytic steps during the guanylation of nascent mRNA? For instance, based on the functional requirements of the capping enzyme, we expect that the presence of GTP in the active site will both lower the energetic barrier of C-terminal domain closure and the free energy of the closed state relative to the apo-state. Similarly, it is expected that guanylation of K82 will both destabilize the closed state and lower the energetic barrier

of C-terminal domain opening relative to the apo-state. Detailed analysis of the thermodynamics and internal dynamics that occur throughout the catalytic cycle promise to move us toward a more complete understanding of the fundamental nature of oligonucleotide ligation and mRNA capping.

This chapter, in full, is a reprint of “Catalytically requisite conformational dynamics in the mRNA-capping enzyme probed by targeted molecular dynamics”, which was published in 2008 in Biochemistry, volume 47, issue 13, pages 4102 to 4111, by Robert V. Swift, and J. Andrew McCammon. This chapter was included with the permission of J. Andrew McCammon.

Chapter 5

Substrate Induced Population Shifts and Stochastic Gating in the PBCV-1 mRNA Capping Enzyme

Abstract

The 317 residue PBCV-1 mRNA capping enzyme catalyzes the second enzymatic reaction in the formation of an N-7-methyl-GMP cap on the 5´-end of the nascent mRNA. It is composed of two globular domains bound by a short flexible peptide linker, which have been shown to undergo opening and closing events. The small size and experimentally demonstrated domain mobility make the PBCV-1 capping enzyme an ideally suited model system to explore domain mobility in context of substrate binding. Here, we specifically address the following four questions. Question one: How does substrate binding affect relative domain mobility: is the system better described by an induced fit or population shift mechanism? Question two: What are the gross characteristics of a conformation capable of binding substrate? Question three: Does "domain gating" of the active site affect the rate of substrate binding? Question four: Does the magnitude of receptor conformational fluctuations confer substrate specificity by sterically occluding molecules of a particular size or geometry? We answer these questions using a combination of theory, Brownian dynamics, and molecular dynamics.

Our results show that binding efficiency is a function of conformation, but that isomerization between efficient and inefficient binding conformations does not impact the substrate association rate. Additionally, we show that conformational flexibility alone is insufficient to explain single stranded mRNA specificity. While our results are specific to the PBCV-1 mRNA capping enzyme, they provide a useful context within which the substrate binding behavior of similarly structured enzymes or proteins may be considered.

5.1 Introduction

An enzyme's ability to change conformations in response to substrate binding is fundamentally important to efficient catalysis. For example, a number of enzymes (to name a few, adenylate kinase[104], the PBCV-1 mRNA capping enzyme [4], and DNA ligase [6]) have large accessory domains, which are covalently linked *via* short, flexible peptide hinges to their catalytic domains. When the enzyme binds substrate, these accessory domains close over the active site and provide a network of interactions that facilitate catalysis. The mRNA capping enzyme from the *Paramecium bursaria Chlorella* virus, PBCV-1, is a relatively small, 317 residue, GTP-dependent enzyme composed of two domains that are covalently linked *via* a flexible oligopeptide hinge. Crystal structures of the GTP-bound, or holo, form of the PBCV-1 mRNA capping enzyme show that the non-catalytic-accessory domain may undergo significant movement with respect to the catalytic domain during the catalytic cycle, moving from an "open" to a "closed"

conformation [4] (Figure 5.1). The relatively small size of the capping enzyme and experimentally demonstrated mobility of the accessory domain make it an ideal model system to explore the functional importance of accessory domain mobility in the context of substrate binding. In this spirit, we explore the substrate binding behavior of the PBCV-1 mRNA capping enzyme using a combination of computational and theoretical approaches. While our results are specific to the capping enzyme, they provide a general context within which the substrate binding behavior of similarly structured enzymes and proteins may be considered.

We begin by asking four questions regarding the varying functional roles that domain motion plays during substrate binding. We enumerate these four questions below, introducing them first in a general sense, which may be asked of any two domain, hinged protein, then in context of the PBCV-1 mRNA capping enzyme. We also introduce the methods that we've used to probe their answers.

Question one: how does substrate binding lead to isomerization into a catalytically competent conformation? Two distinct models have been suggested to describe this phenomenon. In the first, proposed by Koshland in 1958 and called the induced fit model, the relationship between substrate binding and enzyme isomerization was explained by suggesting that a substrate free, or apo, enzyme exists in a narrow range of conformations and that substrate binding compels enzyme reorganization into a conformation able to carry out catalysis [84]. The second proposed model makes use of contemporary notions of energy landscapes [105, 106]. Called the pre-existing equilibrium,

or population shift, model of substrate binding, this view suggests that the apo enzyme populates a range of conformations and that substrate binding stabilizes the lowest energy of these, thereby redistributing the pre-existing population [107, 108]. In order to determine the model that better describes substrate binding by the PBCV-1 capping enzyme, we use molecular dynamics (MD) simulations to generate two conformational ensembles, one with GTP bound to the active site and one without GTP bound to the active site. By comparing the distribution of conformations in each state, the more suitable substrate binding model may be inferred.

Question two: if the accessory domain populates a range of conformations, and if the proximity of the accessory domain to the active site mediates the substrate's access to the active site, what defines a binding competent conformation? Conceptually, this question is closely related to the concept of "domain gating" [109-111], which draws an analogy between a hinged accessory domain and a gate. When the mobile-accessory domain prevents substrate binding, the gate is "closed", while the gate is "open" when substrate binding is allowed. Depending upon the specific protein-ligand, or enzyme-substrate system, different conformational requirements must be met before the gate can be considered open and the enzyme or protein can be considered in a "binding competent" conformation. In order to determine what defines a binding competent PBCV-1 mRNA capping enzyme conformation, we obtain and compare Brownian dynamics [112] association rates of GTP to two distinct apo conformations, which differ in the relative distance between the accessory and catalytic domains, that were generated during the MD simulations.

Question three: if obtaining a binding competent conformation is a prerequisite of substrate binding, does the isomerization rate between binding competent and incompetent conformations affect the substrate association rate? This question was addressed by Szabo et al. [109], who, extending the earlier work of McCammon and Northrup [113], developed a general, quantitative method describing stochastically gated diffusional bimolecular association. This method predicts the affects of domain gating on the substrate binding rate in terms of two variables: the domain gating period, and the time required for the initial enzyme-substrate- encounter complex to dissociate. Generally, the domain gating period will be tied to the underlying free energy surface along a reaction coordinate whose value describes the openness of the gate, while the time required for the enzyme-substrate encounter complex to dissociate is a function of the electrostatic interaction between enzyme and substrate and the translational diffusion coefficients of the enzyme and substrate, which are roughly related to the average compactness of each molecule. Using a standard approximation of the time required for dissociation of the enzyme-substrate-encounter complex and an analytic approximation of the domain gating period consistent with our MD results, we apply the Szabo model to evaluate the affects of gating on the GTP association rate.

Question four: do inherent conformational limitations confer substrate specificity by sterically occluding molecules of a certain size or shape from the active site? Fundamentally, this is a question of sterics and is closely related to the magnitude of receptor conformational fluctuations. For example, if the protein receptor does not

sample conformations that accommodate a ligand of a particular size or geometry, binding will be prevented. Based on crystal structure data, such a mechanism was proposed to explain how the PBCV-1 mRNA capping enzyme selectively binds single stranded mRNA while discriminating against double stranded DNA [4]. Using our molecular dynamics results, along with homology to the PBCV-1 DNA ligase, we re-visit this proposal and explore its validity.

In eukaryotes and their viruses, the process of mRNA capping is carried out co-transcriptionally by a set of three enzyme-catalyzed reactions that have been well described [85]. In the first enzymatic reaction, a triphosphatase cleaves the terminal phosphate from the 5´-terminal triphosphate of mRNA, forming a diphosphate end. The terminal diphosphate serves as the substrate for the mRNA capping enzyme, which transfers GMP from GTP to the 5´-end. In the final step, a methyltransferase transfers a methyl group to the N-7 nitrogen of GMP. In the mRNA capping enzyme, chemistry is initiated when a conserved, catalytic lysine attacks the alpha-phosphate of bound GTP, displacing pyrophosphate and forming a GMP-enzyme intermediate. In the second step, mRNA binds, and the GMP is transferred to the 5´-diphosphate end of the mRNA, completing the reaction. Formation of the 5´ cap is the first post-transcriptional modification of nascent mRNA and is essential in mRNA splicing, nuclear export, mRNA stability, and translation initiation [86-88].

The mRNA capping enzyme belongs to the nucleotidyltransferase superfamily whose members also include ATP and NAD^+ dependent DNA ligases and ATP

dependent RNA ligases. Common chemistry and structural elements distinguish the Superfamily [1]. Shared chemistry is characterized by the transfer of a nucleotide to the 5´-end of a polynucleotide *via* a nucleotide-enzyme intermediate. Shared structure is characterized by a catalytic N-terminal domain, or nucleotidyltransferase domain, and a non-catalytic-accessory-C-terminal domain that plays different roles depending on the superfamily member [1]. For example, in both the ATP dependent DNA ligases and the mRNA capping enzymes, the C-terminal domain contains an oligonucleotide/oligosaccharide (OB) binding fold, which is thought to play a role in substrate positioning prior to, and during catalysis [1, 4, 114]. The C and N-terminal domains associate non-covalently in the ATP dependent RNA ligases of kinetoplastids[47]. In all other superfamily members, the domains are covalently linked by a short, flexible oligopeptide.

The “open” and “closed” conformations of the holo state of the PBCV-1 mRNA capping enzyme show large quasi-rigid displacements of the C-terminal, or OB domain, relative to the nucleotidyltransferase domain (Figure 5.1) [4]. In the open state, the centers of mass of the nucleotidyltransferase and OB domains are separated by 30.3 Å, and the GTP triphosphate is in an unreactive position, orthogonal to the nucleophile. In the closed state, the centers of mass of the domains are separated by 25.5 Å, and the triphosphate is in a reactive conformation, in line with the nucleophile. Subsequent opening and closing events are expected to mediate mRNA binding, GMP transfer, and product release [1, 4].

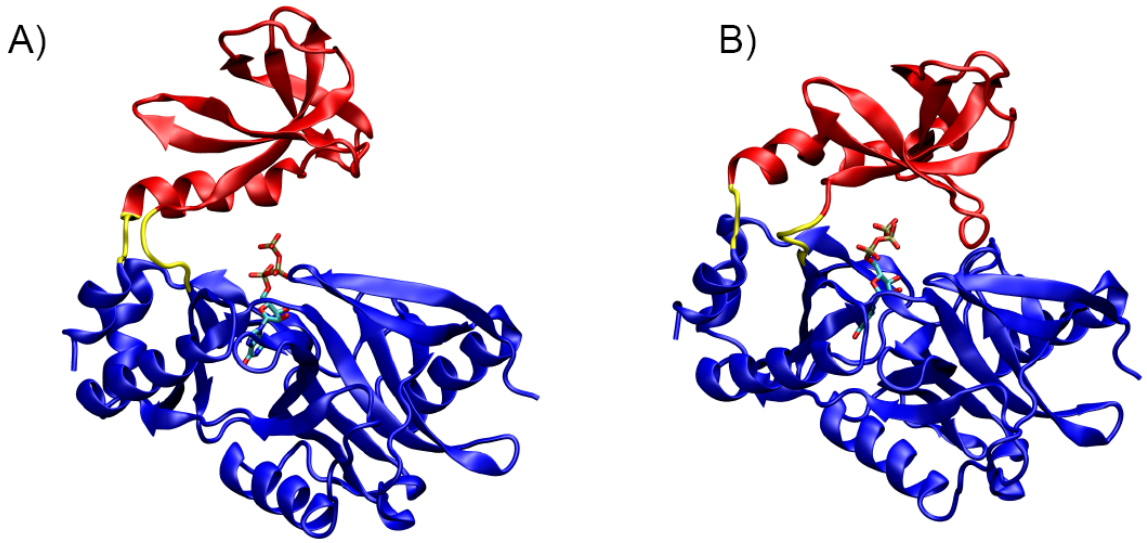


Figure 5.1: Quasi rigid domain displacement. (A) Open crystal structure. The centers of mass of the OB domain (the non-catalytic-accessory domain), colored in red, and the nucleotidyltransferase domain (the catalytic domain), colored in blue, are separated by 30.3 Å. The hinge region is colored in yellow and GTP is rendered in licorice and colored by atom type. (B) Closed crystal structure, coloring is as in (A). The centers of mass of the OB and nucleotidyltransferase domains are separated by 25.5 Å

5.2 Computational Methods and Theory

5.2.1 Molecular Dynamics Simulations

Carrying out several short simulations is a more effective method of covering phase space than carrying out a single long simulation [115]. In order to make use of this observation, six independent 20ns simulations of the PBCV-1 mRNA capping enzyme in the apo state, and six independent 20ns simulations in the holo state, were initiated from the open and closed conformations, respectively, of Protein Data Bank (PDB) code 1CKM [4] (Figure 5.1). For the apo simulations, coordinates were taken from chain A after removing GTP. For the holo simulations, coordinates were taken from chain B of

1CKM. For both simulations, histidine protonation states were assigned using WHATIF [73]. All other hydrogen atoms and protein force field parameters were assigned using the AMBER99 force field [24] in the xleap program of Amber8 [71]. GTP parameters were taken from the Meagher parameter set [74]. Following parameterization, the protein was immersed in a rectangular prism of pre-equilibrated TIP3P waters [26] that provided a 10Å buffer between the protein and the next periodic image in each dimension. Three sodium ions were added in xleap to neutralize the system charge. Following parameterization, conjugate gradient minimization was carried out in four cycles. In the first two thousand step cycle, only hydrogen atoms were free to move, all other atoms were constrained. In the second two thousand step cycle, hydrogen, water and ions were free to move, all other atoms were constrained. In the third two thousand step cycle, the protein backbone was constrained and all other atoms were free to move. In the last ten thousand step cycle, all atoms were free to move. Once minimized, six MD simulations of the holo state, and six MD simulations of the apo state were initiated in the NPT ensemble from different velocity seeds. The holo state was allowed to equilibrate for 50 ps. For the apo simulations, the protein was first constrained, allowing water to equilibrate to the proper density within the active site over 30 ps; this was to prevent spurious domain closure as a result of removing GTP. Following this 30 ps equilibration, the apo state was equilibrated constraint free for 50 ps. All MD simulations were carried out using a 1 fs time step, extracting configurations every 50 fs for analysis. Temperature was maintained at 300K using Langevin dynamics, while the pressure was maintained at 1 atmosphere using the hybrid Nose-Hoover-Langevin method [50] with period and decay times of 100 and 50 fs respectively. Long-range electrostatics were calculated using the

Particle Mesh Ewald method [51]. A multiple time-stepping algorithm was employed to improve integration efficiency. The long range cutoff was set to 14 Å, and bonded interactions were calculated every time step. Short range, non-bonded interactions were calculated every two time steps, and long range electrostatics were calculated every four time steps. Simulations were carried out on NCSA's Abe machine. Benchmarks running on 256 processors ranged from 0.11 days per nanosecond to 0.16 days per nanosecond. All molecular dynamics simulations were carried out using NAMD 2.6 [18].

5.2.2 Brownian Dynamics

Coordinates for the hyper open and open states (defined in the *Induced fit v. population shift* section) were taken from the apo MD simulations retaining all hydrogen atoms. GTP coordinates were taken from chain B of the PDB code 1CKM [4] and hydrogen atoms were assigned according to the Meagher parameter set [74]. The long range electrostatics of the protein were modeled by solving the linearized Poisson-Boltzmann equation using a focusing calculation in the Adaptive Poisson Boltzmann Solver (APBS) [116]. Potential values were assigned to the nodes of a 129 x 129 x 129 node grid with 1 Å spacing between nodes. The long range electrostatics of GTP were calculated by the same method, and potential values were assigned to the nodes of a 65 x 65 x 65 grid with 1 Å spacing between nodes. For both the capping enzyme and GTP, the ionic strength was set to zero, the temperature to 298.15K, the internal molecular dielectric constant to 2.0, and the solvent dielectric constant to 78.4.

The effective charge methodology [117] and desolvation energies [118], as implemented in SDA [112], were used during Brownian dynamics to model the long range electrostatics and desolvation energies, respectively. In GTP, effective charges were positioned on the non-bridging beta and gamma oxygen atoms, as well as on the N2, N3 and O6 atoms. For the desolvation parameter, the alpha value was set to four.

Short-range interactions were treated by an exclusion volume that prevents van der Waals overlap of the capping enzyme and GTP. The excluded volume of the capping enzyme is the region of space contained in a 1.4 Å probe accessible surface. If a BD step carries a surface exposed atom of GTP into the excluded volume of the capping enzyme, the step is repeated with a different random number until GTP no longer enters the excluded volume of the capping enzyme. The excluded volume is pre-calculated for the capping enzyme and stored on a grid with 0.5 Å spacing between nodes. A list of surface GTP atoms is also pre-generated and stored.

Brownian dynamics simulations were carried out using the SDA software [112] which propagates dynamics by solving the Ermak-McCammon equations of motion [119]. For each capping enzyme conformation, 1.6×10^6 trajectories were performed to ensure convergence of the calculated association rates. Each trajectory was run at 298.15K with zero ionic strength. The trajectories were initiated with the GTP and the capping enzyme separated by center to center distance b and were terminated when GTP and the capping enzyme were separated by a center to center distance $c > b$; c was set to 300 Å and b to 100 Å. Diffusion coefficients for GTP and the capping enzyme conformations were

calculated using the program Hydropro [120] setting the temperature to 298.15K, the viscosity to that of water at 298.15K (0.890 cP), the bead radii of the protein to 3.2 Å, and the bead radii of GTP to 2.8 Å.

Rates were calculated using the formalism developed by Northrup, Allison and McCammon [121]. The distance between the side chain nitrogen atom of the K82 nucleophile and the 03G non-bridging oxygen atom on the gamma phosphate of GTP was used as the association criterion.

5.2.3 Gating Effects on GTP Association Kinetics

We begin by assuming domain gating is a two state process in which the domains may exist in either an open or a closed conformation, with opening and closing rates given by k_o and k_c, respectively. The two state assumption allows us to make use of the results of Szabo et al. [109], who showed that the bimolecular association rate, in the presence of stochastic gating, can be written,

$$\frac{1}{k_G} = \frac{1}{k_{UG}} + \frac{k_c}{k_o(k_o + k_c)\hat{k}(k_o + k_c)} \qquad 5.1$$

where k_G is the bimolecular association rate in the presence of gating; k_{UG} is the steady-state-ungated bimolecular association rate to the enzyme fixed in its most open

conformation, $(k_o + k_c)^{-1}$ is the domain gating period, and $\hat{k}(s)$ is the Laplace transform of *s.* Eq. 5.1 reduces to two limiting cases depending on the relative values of the diffusional relaxation time of the system, τ_D, and the domain gating period,

$$\frac{k_G}{k_{UG}} = \frac{k_o}{k_o + k_c}, \ (k_o + k_c)^{-1} >> \tau_D \qquad 5.2$$

$$\frac{k_G}{k_{UG}} = 1, \ (k_o + k_c)^{-1} << \tau_D \qquad 5.3$$

$$\tau_D = r_c^2 / D \qquad 5.4$$

where r_c is the collision distance of the system, approximated as the sum of the smallest radii of spheres which contain each molecule; and *D* is the sum of the translational diffusion coefficients of the substrate and enzyme. Eqns. 5.2 and 5.3 represent substrate association in the presence of "slow" and "fast" gating respectively; hence, domain motion only affects substrate binding kinetics in the slow gating limit.

5.3 Results and Discussion

5.3.1 Induced fit v. population shift

The effect of substrate mediated free energy reorganization on conformational variability has drawn increasing attention and is usefully described by either an induced fit [84] or a population shift model [107, 108]. Examples of both models have been reported in the literature.

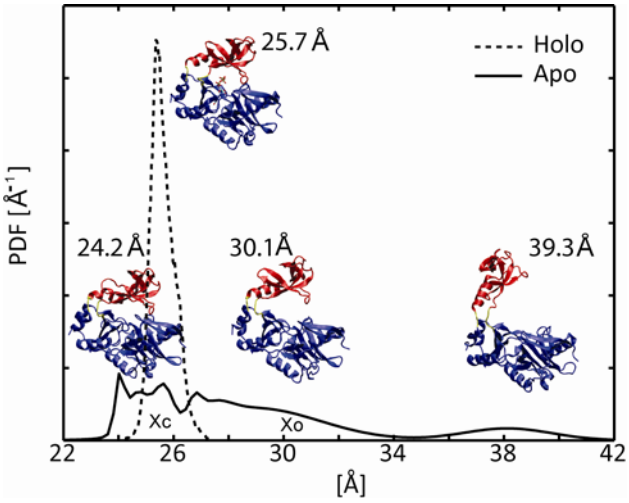


Figure 5.2: **Conformational distributions in the apo and holo states.** Probability distribution functions (PDFs) of the distance separating the centers of mass of the OB and nucleotidyltransferase domains in the MD simulations. Xc and Xo indicate the locations of the closed and open crystal structures, respectively. The holo distribution is represented by a dashed line, and the apo distribution by a solid line. Structures of selected conformations are labeled with the distance separating the centers of mass of their OB and nucleotidyltransferase domains. The three lower structures are representatives of the apo simulation while the single upper structure is a representative of the holo simulations. Coloring is as in figure 5.1.

For example, while approximately the same range of conformational space is sampled in both the apo and holo forms of the two domain glucose and ribose binding proteins, ligand binding biases the conformational populations toward a single conformation and is best described by a population shift model [93, 122]. The same observations hold for adenylate kinase [104]. On the other hand, energy profiles in the presence and absence of acetylcholinesterase demonstrated that loop isomerization in the protein neurotoxin fasciculin-2 does not occur prior to acetylcholinesterase binding; it follows that the system is best described by an induced fit model [123].

In order to determine whether GTP binding by the PBCV-1 capping enzyme is better described by the induced fit or the population shift model of substrate binding, we examine the distribution of conformations of the OB domain relative to the nucleotidyltransferase domain. As we are only interested in the relative motion of the OB domain, we neglect internal domain fluctuations and approximate the OB domain as a rigid body. This approximation is supported by earlier work in which OB-domain secondary structure was retained during isomerization from an open to a closed conformation [114]. Using the rigid body approximation, the position and orientation of the OB domain is described by three positional and three orientational degrees of freedom that are measured with respect to a coordinate system whose origin we center on the center of mass of the nucleotidyltransferase domain. Furthermore, relative to the large fluctuations between the centers of mass of the OB and nucleotidyltransferase domains observed during simulations of the apo state, fluctuations along the two orthogonal positional degrees of freedom, as well fluctuations around the three orientational degrees

of freedom, were modest. This observation is apparent in the three representative conformations from the apo state simulations shown in Figure 5.2. As fluctuations between the centers of mass of the OB and nucleotidyltransferase domains constitute the preponderance of relative domain motion, we approximate the distribution of conformations of the OB domain relative to the nucleotidyltransferase domain by the distribution of distances separating their centers of mass. The probability distribution functions (PDFs) of the distance separating the centers of mass of the domains in the apo and holo trajectories are reported in Figure 5.2. In the apo state, the two domains span a range of conformations and exhibit a bimodal distribution. Conversely, in the holo state, the distribution is localized around a single conformation that resembles the closed, holo crystal structure.

The induced fit model predicts that the apo state samples only small internal fluctuations, such as side-chain reorientations, in the vicinity of a narrow distribution of backbone conformations, which are clearly distinguishable from the predominant backbone conformations sampled in the holo state. Accordingly, the large distribution of distances separating the centers of mass of the OB and nucleotidyltransferase domains observed during simulations of the apo state rule out the induced fit model. Additionally, the narrow distribution of distances observed during simulations of the holo state, when considered in tandem with the distribution observed during simulations of the apo state, implies that the PBCV-1 capping enzyme follows a population shift model of substrate binding. While the possibility that the holo state significantly populates an open conformation at time scales longer than those accessed during the MD simulations can’t

be ignored, this does not negate description by the population shift model (excluding the unlikely exception that the true apo and holo distributions are identical, in which case the population cannot be said to shift).

Population shift substrate binding mechanisms have also been reported for both the glucose[122] and ribose[93] binding proteins, as well as adenylate kinase[124], each of which is structurally similar to the PBCV-1 mRNA capping enzyme. While this correlation suggests that the population shift mechanism may be a commonly occurring substrate binding mechanism among proteins formed by the fusion of two domains *via* a flexible linker, the data set is small and exceptions are likely to be found.

In addition to specifying binding mechanism type, the distributions of distances shown in figure 5.2 allow qualitative insight into binding thermodynamics. For example, the narrowing of the distributions that occurs upon substrate binding implies a reduction of entropy. Assuming spontaneous binding, the process must be driven by a decrease in enthalpy. This is consistent with an extensive network of hydrogen bonds between positively charged active site residues and the negatively charged non-bridging GTP oxygen atoms [4, 114].

Interestingly, the apo simulations show a sparse distribution of states in which the centers of mass of the nucleotidyltransferase and OB domains are separated by 36 Å to 42 Å, 6Å to 12 Å wider than the distance observed in the open crystal structure (Figure 5.1 A). We designate these “hyper-open” states. As the substrate must bind prior to

catalysis, the occurrence of hyper-open conformations may be functionally significant. For instance, if the OB domain is nearby the nucleotidyltransferase domain, GTP binding may be occluded or hindered, and the domains must open to allow facile binding.

5.3.2 Binding Competent Conformations

In order to clarify how extensively the domains must open prior to GTP binding, we carried out two Brownian dynamics simulations. The first simulation measured the association rate to a hyper-open conformation in which the centers of mass of the nucleotidyltransferase and OB domains were separated by 38.8 Å. The second simulation measured the association rate to an open conformation, in which the centers of mass of the two domains were separated by 30.1 Å, a distance similar to that found in the open crystal structure. Table 5.1 reports the association rates at varying distances of encounter complex formation (the "encounter complex" is rigorously defined in the Computational Methods and Theory section).

In the hyper-open conformation, encounter complex formation did not occur at distances shorter than 7.00 Å. In the open conformation, encounter complex formation was not observed at distances shorter than 15.00 Å. Figure 5.3 is a visual comparison of the surfaces that define the shortest encounter complexes formed in the hyper-open and open conformations and indicates that GTP is occluded from the active site while the enzyme is in the open state. These results imply that isomerization into a hyper-open conformation dramatically improves GTP binding efficiency.

Table 5.1: Brownian dynamics GTP association rates as a function of encounter complex distance to two PBCV-1 mRNA capping enzyme conformations. Standard deviations over $1.6x10^6$ simulations are shown in parenthesis. No encounter events are indicated by dashed entries.

Distance (Å)	Hyper-open (M^{-1} s^{-1})	Open (M^{-1} s^{-1})
7.00	$4.2x10^4$ +/- ($7x10^4$)	-
9.00	$4.8x10^4$ +/- ($1x10^5$)	-
11.00	$4.1x10^6$ +/- ($8x10^5$)	-
13.00	$1.9x10^7$ +/- ($3x10^6$)	-
15.00	$1.1x10^9$ +/- ($2x10^7$)	$1.7x10^5$ +/- ($5x10^5$)

The dramatic improvement in binding efficiency following isomerization into a hyper-open conformation is a specific instance of a more general behavior followed by proteins that behave according to the population shift mechanism of substrate binding. This follows from the observation that non-productive substrate binding conformations will be sampled in the apo state of proteins that follow a population shift type mechanism; consequently, isomerization into a conformation conducive to substrate binding enhances binding efficiency.

Neglecting conformational flexibility during BD simulations makes the calculated association rates sensitive to the reaction criteria defining the encounter complex [112, 118, 125]; as a result, our calculated rates cannot be expected to accurately predict the true association rate. However, the conclusion that isomerization into a hyper-open conformation is a prerequisite of GTP binding is based on relative association rates, to

representative hyper-open and open conformations, and so is not affected by the inaccuracy of the reported absolute values.

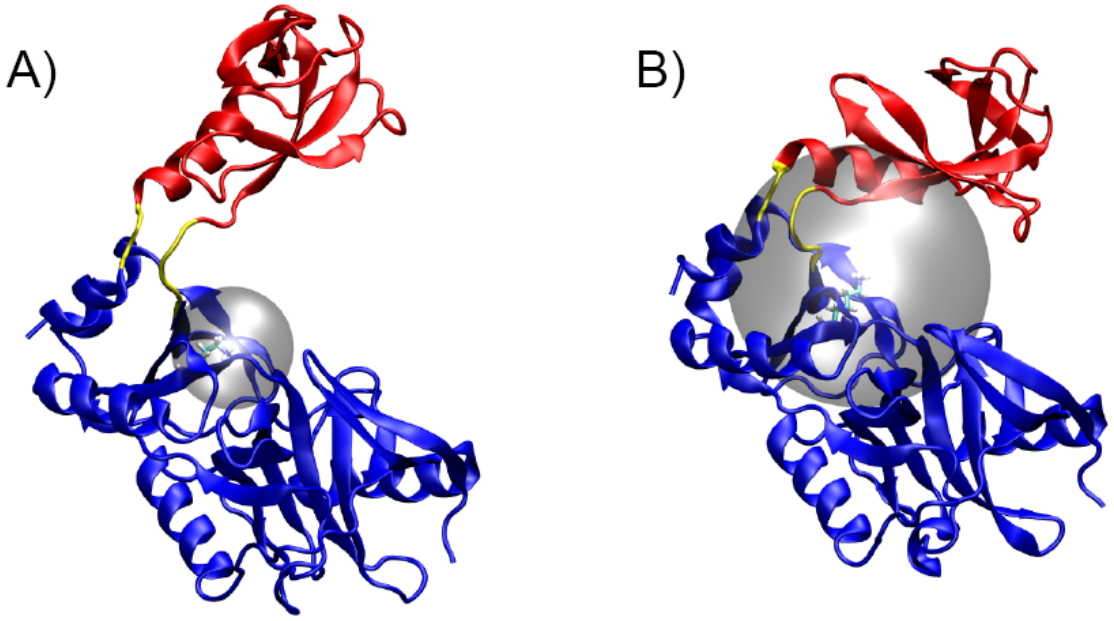


Figure 5.3: Shortest encounter complexes determined by Brownian dynamics simulations. Transparent spherical surfaces define the shortest observed encounter complexes. The volume contained within the spherical surface was inaccessible to GTP. (A) Representative hyper-open conformation. The centers of mass of the OB and nucleotidyltransferase domains are separated by 38.8 Å. (B) Representative open conformation resembling the open crystal structure. The centers of mass of the two domains are separated by 30.1 Å. Coloring is as in figure 5.1.

5.3.3 Domain Gating and GTP Association Kinetics

If isomerization into a hyper-open conformation, or binding competent conformation, is a prerequisite of substrate binding, the isomerization rate may regulate the substrate binding rate. Regulation occurs when an enzyme-substrate encounter complex forms while the enzyme is in a binding incompetent conformation and the

enzyme-substrate encounter complex dissociates faster than isomerization into a binding competent conformation occurs. In this case, called the slow gating limit, binding can only occur if the enzyme-substrate encounter complex forms while the enzyme is in a binding competent conformation, and the association rate is proportional to the fraction of enzymes in the hyper-open, or binding competent conformation (Eq. 5.2). Conversely, regulation does not occur if the enzyme-substrate encounter complex forms when the enzyme is in a binding incompetent conformation and isomerization into a binding competent conformation occurs faster than the enzyme-substrate encounter complex dissociates. In this case, called the fast gating limit, the enzyme isomerizes between binding competent and incompetent conformations many times before the enzyme-substrate encounter complex dissociates. By virtue of the large number of isomerization events that occur prior to dissociation of the enzyme-substrate complex, the likelihood that the substrate re-collides with a binding competent conformation is high, and the substrate binding rate is unaffected (Eq. 5.3).

In order to determine whether the PBCV-1 capping enzyme follows a slow or fast gating scenario, we compare the time scales of diffusional dissociation of the encounter complex, approximated using the diffusional relaxation time of the system (Eq. 5.4), to an analytical approximation of the period of domain gating. We first let r be the distance separating the centers of mass of the nucleotidyltransferase and OB domains. We then define an "open" conformation as one of the binding incompetent conformations characterized by the set of conformations with 24 Å $< r <$ 35 Å and a "hyper-open" conformation as one of the binding competent conformations characterized by the set of

conformations with 35 Å < *r*. The rate of isomerization between the open and hyper-open conformations can be determined by solving the Smoluchowski equation with appropriate boundary conditions [20, 33],

$$\frac{\partial p(r,t)}{\partial t} = \frac{\partial}{\partial r} D(r)\left[\frac{1}{kT}\frac{\partial W(r)}{\partial r} + \frac{\partial}{\partial r}\right] p(r,t) \qquad 5.5$$

where $p(r,t)$ is the probability of finding the domains separated by a distance *r* at a time *t*, *W(r)* is the potential of mean force (PMF) at *r*, and *D(r)* is a position dependent diffusion coefficient. While limited sampling prevents *W(r)* from being accurately determined, the conformational variability of the apo state observed during the six-20 ns simulations implies that the energetic barrier separating the open state from the hyper-open state is roughly comparable to the available thermal energy. To a first approximation then, we take $\partial W(r)/\partial r \approx 0$ for all *r*, which describes diffusive domain motion along a flat free-energy surface. If we further assume that the diffusion coefficient is constant along *r*, then Eq. 5.5 reduces to,

$$\frac{\partial p(r,t)}{\partial t} = D\frac{\partial^2 p(r,t)}{\partial r^2} \qquad 5.6$$

After fixing the center of mass of the nucleotidyltransferase domain at the origin and taking *D* as the sum of the diffusion coefficients of the two domains along *r*, Eq. 5.6 approximately describes relative domain motion. We solve Eq. 5.6 within an infinite

square well potential of length L set to the difference between the maximum and minimum values of r that occur during simulations of the apo state. When r = 0, the enzyme is closed, and when $r = L$ the enzyme is in the most extended hyper-open conformation. We set a absorbing boundary condition $p(0,t) = 0$, and a no flux condition $\left(\partial p/\partial t\right)\big|_{r>L} = 0$. The initial distribution is chosen as a delta-function centered at $r = L$. With these conditions, the solution to Eq. 5.6 is well known, and the average time required for the domains to close is given $\tau = L^2/2D$ [126, 127]. Taking L = 17.5 Å and approximating D as the sum of the translational diffusion coefficients of the two domains, determined using Hydropro, gives a value of τ whose inverse is the closing rate, k_c = $1.5\text{x}10^8$ s^{-1}. By symmetry, the opening and closing rates are equal, and the period of domain gating is $\left(k_o + k_o\right)^{-1} = 3.3\,\text{ns}$. The characteristic diffusional relaxation time of the system is $\tau_D = 130\,\text{ns}$. The ratio of the diffusional relaxation time of the system and the domain gating period is $\tau_D/\left(k_o + k_c\right)^{-1} = 39$ implying that on average, the domains open and close approximately thirty nine times before the encounter complex dissociates, providing many opportunities for the substrate to recollide with, and subsequently bind, a conformation that allows substrate binding. A more detailed analysis of the gating affects on the rate of GTP binding is unlikely to alter the conclusion that the rate of association is of the same order of magnitude as would be expected for the enzyme fixed in its hyper-open state. While crossing of the modest energy barrier during isomerization of the enzyme could slow the gating process somewhat, this is likely to be offset by the longer residence times of GTP in the initial encounter complex due to electrostatic interactions with the enzyme.

5.3.4 Polynucleotide Specificity

When the crystal structure of the PBCV-1 mRNA capping was reported, it was noted that the cleft between the OB and nucleotidyltransferase domains in the mRNA capping enzyme was narrower than the cleft in the DNA ligase homolog [4]. It was subsequently suggested that the narrower cleft of the mRNA capping enzyme serves a functional role, preventing double stranded DNA (dsDNA) binding and conferring single stranded mRNA specificity [4]. In contrast, the observation of hyper-open conformations sampled during the apo MD simulations requires further consideration of this suggestion. To clearly illustrate this, we aligned the PBCV-1 nucleotidyltransferase domain of a hyper open conformation, with domain centers of mass opened 8.8 Å wider than the open crystal structure, to the nucleotidyltransferase domain of the PBCV-1 DNA ligase crystallized in complex with dsDNA [6]. The results are shown in figure 5.4 and indicate that hyper-open conformations of the mRNA capping enzyme may accommodate dsDNA. Consequently, rather than arising from a narrow cleft that occludes dsDNA, mRNA specificity may be conferred by an alternative mechanism.

Comparison of the PBCV-1 DNA ligase and mRNA capping enzyme crystal structures offers a possible explanation of polynucleotide specificity. Figure 5.4B shows that the PBCV-1 DNA ligase contains a latch domain, shown in green, which extends from the OB domain and encircles the DNA substrate. In the DNA ligase, the latch domain is essential for the formation of a stable complex with nicked dsDNA [6].

Accordingly, the absence of a latch domain in the PBCV-1 mRNA capping enzyme implies that mRNA specificity may be imparted by the absence of the latch domain: that is, while conformational flexibility of the capping enzyme accommodates nicked dsDNA binding, the complex is likely unstable, rapidly dissociating before ligation can occur. On the other hand, single stranded mRNA has a smaller diameter than dsDNA and may form a stable complex through interactions with only the nucleotidyltransferase and OB domains.

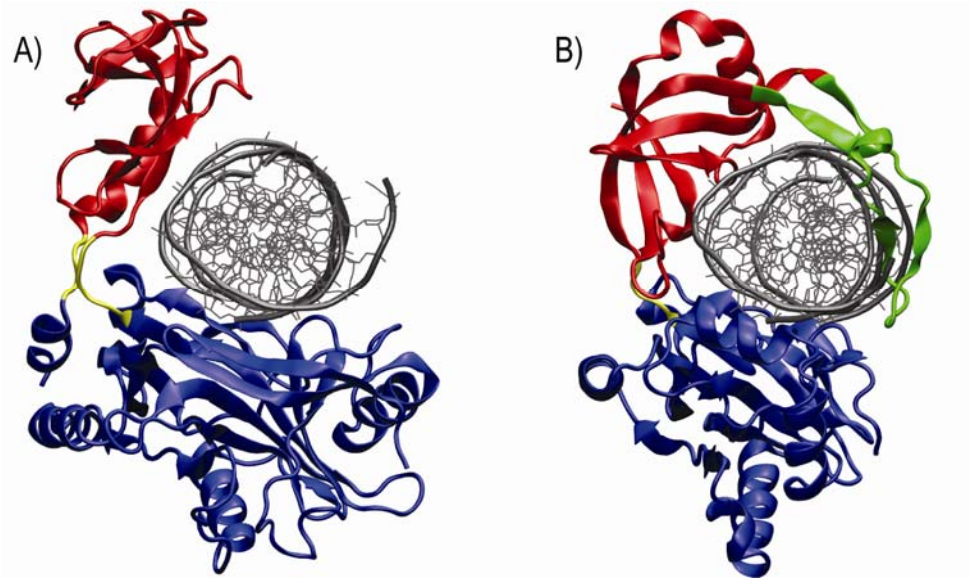


Figure 5.4: Hyper-open conformations may accommodate nicked dsDNA binding. (A) Hyper-open conformation of the PBCV-1 mRNA capping enzyme. The centers of mass of the OB and nucleotidyltransferase domains are separated by 38.8 Å. Coloring is as in figure 5.1. dsDNA was superimposed on the active site by aligning the nucleotidyltransferase domains of the capping enzyme and PBCV-1 DNA ligase. (B) PBCV-1 DNA ligase crystallized in complex with dsDNA. A unique "latch" domain is colored in green. The nucleotidyltransferase domain is shown in blue, the OB domain in red, and the hinge region in yellow.

5.4 **Conclusions**

In this paper we have analyzed several common functional themes related to substrate binding in enzymes and proteins whose tertiary structure is composed of two domains covalently bound by a short, flexible-linking peptide. While our results are specific to the PBCV-1 mRNA capping enzyme, the close relationship between biomolecular structure and function allows them to provide a useful context within which the substrate binding behavior of similarly structured proteins and enzymes may be explored.

By carrying out six independent 20 ns MD simulations of the apo state and six independent 20 ns MD simulations of the holo state of the GTP dependent PBCV-1 mRNA capping enzyme, we have shown that the apo enzyme samples a wide range of conformations and, at the time scales accessed during our simulations, is predisposed towards conformations that resemble the holo closed crystal structure. Rather than inducing a conformational change, substrate binding may further stabilize the more stable apo conformations and shift the equilibrium toward the predominant holo conformation. This is consistent with a population shift mechanism and is similar to results reported for both the ribose and glucose binding proteins [93, 122], as well as the adenylate kinase enzyme [124], each of which shares a tertiary structure similar to that of the PBCV-1 mRNA capping enzyme. While the correlation suggests that the population shift mechanism may be a commonly occurring substrate binding mechanism in proteins with two

domains covalently bound *via* a short, flexible-linking peptide, the data set is limited and exceptions are likely to be found.

Additionally, the MD simulations revealed that the enzyme samples hyper-open conformations in which the centers of mass of the nucleotidyltransferase and OB domains open 6 to12 Å wider than previously reported. As the relative proximity of the OB domain to the nucleotidyltransferase domain mediates the accessibility of the active site, the occurrence of the hyper-open conformations raises questions about their functional significance. To clarify their functionally role, BD simulations were used to show that isomerization into hyper-open conformations dramatically improves GTP binding efficiency by increasing active site accessibility.

The mobility of the OB domain and its role in mediating substrate accessibility leads to the question of whether or not the rate of isomerization into a hyper-open state impacts the substrate association rate. To answer this question, we made an analytic approximation of the domain gating period based on the results of our apo MD simulations. Comparing this result to the time required for the enzyme-substrate encounter complex to dissociate indicated that the substrate likely has many opportunities to recollide with a binding competent conformation prior to diffusing apart to a distance that precludes binding. This analysis indicates that the rate of OB domain isomerization is not likely a factor in determining the rate of substrate association. In other words, interactions provided by the OB domain that are essential for catalysis are gained without

paying a penalty in the form of a decreased substrate association rate. It is possible that other similarly structured enzymes also behave in the same manner.

Finally, by modeling double stranded DNA into the active site of a conformation in which the centers of mass of the OB and nucleotidyltransferase domains were separated by an additional 8.8 Å beyond the separation measured in the crystal structure, we showed that restricted conformational flexibility is insufficient to explain single-stranded mRNA specificity. These results serve as a reminder that as valuable as crystal structure data is, some molecular phenomena must be explained by dynamic arguments.

Chapter 5, in full, is a reprint of the material “Population Shifts and Stochastic Domain Gating in the PBCV-1 mRNA Capping Enzyme” which was published in 2009 in The Journal of the American Chemical Society, volume 131, issue 14, pages 5126 to 5133, by Robert V. Swift, and J. Andrew McCammon. This chapter was included with the permission of J. Andrew McCammon.

Bibliography

1. Shuman, S.; Lima, C. D., The polynucleotide ligase and RNA capping enzyme superfamily of covalent nucleotidyltransferases. *Curr Opin Struct Biol* **2004,** 14, (6), 757-64.

2. Gajiwala, K. S.; Pinko, C., Structural rearrangement accompanying NAD(+) synthesis within a bacterial DNA ligase crystal. *Structure* **2004,** 12, (8), 1449-1459.

3. Sriskanda, V.; Shuman, S., Mutational analysis of Chlorella virus DNA ligase: catalytic roles of domain I and motif VI. *Nucleic Acids Res* **1998,** 26, (20), 4618-25.

4. Hakansson, K.; Doherty, A. J.; Shuman, S.; Wigley, D. B., X-ray crystallography reveals a large conformational change during guanyl transfer by mRNA capping enzymes. *Cell* **1997,** 89, (4), 545-53.

5. Ho, C. K.; Shuman, S., Bacteriophage T4 RNA ligase 2 (gp24.1) exemplifies a family of RNA ligases found in all phylogenetic domains. *Proc Natl Acad Sci U S A* **2002,** 99, (20), 12709-14.

6. Nair, P. A.; Nandakumar, J.; Smith, P.; Odell, M.; Lima, C. D.; Shuman, S., Structural basis for nick recognition by a minimal pluripotent DNA ligase. *Nat Struct Mol Biol* **2007,** 14, (8), 770-8.

7. Nandakumar, J.; Shuman, S.; Lima, C. D., RNA ligase structures reveal the basis for RNA specificity and conformational changes that drive ligation forward. *Cell* **2006,** 127, (1), 71-84.

8. Lee, H. S.; Robinson, R. C.; Joo, C. H.; Lee, H.; Kim, Y. K.; Choe, H., Targeted molecular dynamics simulation studies of calcium binding and conformational change in the C-terminal half of gelsolin. *Biochem Biophys Res Commun* **2006,** 342, (3), 702-9.

9. Mao, X.; Schwer, B.; Shuman, S., Mutational analysis of the Saccharomyces cerevisiae ABD1 gene: cap methyltransferase activity is essential for cell growth. *Mol Cell Biol* **1996,** 16, (2), 475-80.

10. Sawaya, R.; Shuman, S., Mutational analysis of the guanylyltransferase component of Mammalian mRNA capping enzyme. *Biochemistry* **2003,** 42, (27), 8240-9.

11. Wang, L. K.; Ho, C. K.; Pei, Y.; Shuman, S., Mutational analysis of bacteriophage T4 RNA ligase 1. Different functional groups are required for the nucleotidyl transfer and phosphodiester bond formation steps of the ligation reaction. *J Biol Chem* **2003,** 278, (32), 29454-62.

12. Yin, S.; Ho, C. K.; Shuman, S., Structure-function analysis of T4 RNA ligase 2. *J Biol Chem* **2003,** 278, (20), 17601-8.

13. Cherepanov, A. V.; de Vries, S., Kinetic Mechanism of the Mg2+-dependent Nucleotidyl Transfer Catalyzed by T4 DNA and RNA Ligases. *J. Biol. Chem.* **2002,** 277, (3), 1695-1704.

14. Bougie, I.; Parent, A.; Bisaillon, M., Thermodynamics of ligand binding by the yeast mRNA-capping enzyme reveals different modes of binding. *Biochem J* **2004,** 384, (Pt 2), 411-20.

15. Souliere, M. F.; Perreault, J. P.; Bisaillon, M., Kinetic and thermodynamic characterization of the RNA guanylyltransferase reaction. *Biochemistry* **2008,** 47, (12), 3863-74.

16. Griffiths, D., *Quantum Mechanics*. Second ed.; Pearson, Prentice Hall: 2005.

17. McQuarrie D., A., *Statistical Mechanics*. Harper & Row: 1976.

18. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **2005,** 26, (16), 1781-1802.

19. Trzesniak, D.; Kunz, A. P.; van Gunsteren, W. F., A comparison of methods to compute the potential of mean force. *Chemphyschem* **2007,** 8, (1), 162-9.

20. Hanggi, P.; Talkner, P.; Borkovec, M., Reaction-Rate Theory - 50 Years after Kramers. *Reviews of Modern Physics* **1990,** 62, (2), 251-341.

21. Bicout, D. J.; Szabo, A., Electron transfer reaction dynamics in non-Debye solvents. *Journal of Chemical Physics* **1998,** 109, (6), 2325-2338.

22. McCammon, J. A.; Harvey, S., C., *Dynamics of proteins and nucleic acids*. Cambridge University Press: 1987.

23. Adcock, S. A.; McCammon, J. A., Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem Rev* **2006,** 106, (5), 1589-615.

24. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A 2Nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *Journal of the American Chemical Society* **1995,** 117, (19), 5179-5197.

25. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.;

Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M., All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998,** 102, (18), 3586-3616.

26. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *Journal of Chemical Physics* **1983,** 79, (2), 926-935.

27. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J., Interaction Models for Water in Relation to Protein Hydration. *Intermolecular Forces* **1981**, 331-342.

28. Boehr, D. D.; Dyson, H. J.; Wright, P. E., An NMR perspective on enzyme dynamics. *Chem Rev* **2006,** 106, (8), 3055-79.

29. Stone, J. E.; Phillips, J. C.; Freddolino, P. L.; Hardy, D. J.; Trabuco, L. G.; Schulten, K., Accelerating molecular modeling applications with graphics processors. *J Comput Chem* **2007,** 28, (16), 2618-40.

30. Torrie, G. M.; Valleau, J. P., Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977,** 23, 187-199.

31. Isralewitz, B.; Gao, M.; Schulten, K., Steered molecular dynamics and mechanical functions of proteins. *Curr Opin Struct Biol* **2001,** 11, (2), 224-30.

32. Schlitter, J.; Engels, M.; Kruger, P., Targeted molecular dynamics: a new approach for searching pathways of conformational transitions. *J Mol Graph* **1994,** 12, (2), 84-9.

33. Hamelberg, D.; Shen, T.; McCammon, J. A., Relating kinetic rates and local energetic roughness by accelerated molecular-dynamics simulations - art. no. 2411003. *Journal of Chemical Physics* **2005,** 122, (24), -.

34. Fajer, M.; Hamelberg, D.; McCammon, J. A., Replica-Exchange Accelerated Molecular Dynamics (REXAMD) Applied to Thermodynamic Integration. *Journal of Chemical Theory and Computation* **2008,** 4, (10), 1565-1569.

35. Organization, W. H. *WHO Report on Global Surveillance of Epidemic-prone Infectious Diseases*; 2003.

36. Cross, G. A., Trypanosomes at the gates. *Science* **2005,** 309, (5733), 355.

37. Fairlamb, A. H., Chemotherapy of human African trypanosomiasis: current and future prospects. *Trends Parasitol* **2003,** 19, (11), 488-94.

38. Stuart, K.; Allen, T. E.; Heidmann, S.; Seiwert, S. D., RNA editing in kinetoplastid protozoa. *Microbiol Mol Biol Rev* **1997,** 61, (1), 105-20.

39. Stuart, K. D.; Schnaufer, A.; Ernst, N. L.; Panigrahi, A. K., Complex management: RNA editing in trypanosomes. *Trends Biochem Sci* **2005,** 30, (2), 97-105.

40. Simpson, L.; Sbicego, S.; Aphasizhev, R., Uridídine insertion/deletion RNA editing in trypanosome mitochondria: A complex business. *RNA* **2003,** 9, 265-276.

41. Worthey, E. A.; Schnaufer, A.; Mian, I. S.; Stuart, K.; Salavati, R., Comparative analysis of editosome proteins in trypanosomatids. *Nucleic Acids Res* **2003,** 31, (22), 6392-408.

42. Panigrahi, A. K.; Ernst, N. L.; Domingo, G. J.; Fleck, M.; Salavati, R.; Stuart, K. D., Compositionally and functionally distinct editosomes in Trypanosoma brucei. *Rna* **2006,** 12, (6), 1038-49.

43. Blum, B.; Simpson, L., Formation of guide RNA/messenger RNA chimeric molecules in vitro, the initial step of RNA editing, is dependent on an anchor sequence. *Proc Natl Acad Sci U S A* **1992,** 89, (24), 11944-8.

44. Seiwert, S. D.; Heidmann, S.; Stuart, K., Direct visualization of uridylate deletion in vitro suggests a mechanism for kinetoplastid RNA editing. *Cell* **1996,** 84, (6), 831-41.

45. Schnaufer, A.; Panigrahi, A. K.; Panicucci, B.; Igo, R. P., Jr.; Wirtz, E.; Salavati, R.; Stuart, K., An RNA ligase essential for RNA editing and survival of the bloodstream form of Trypanosoma brucei. *Science* **2001,** 291, (5511), 2159-62.

46. Rusche, L. N.; Huang, C. E.; Piller, K. J.; Hemann, M.; Wirtz, E.; Sollner-Webb, B., The two RNA ligases of the Trypanosoma brucei RNA editing complex: cloning the essential band IV gene and identifying the band V gene. *Mol Cell Biol* **2001,** 21, (4), 979-89.

47. Schnaufer, A.; Ernst, N. L.; Palazzo, S. S.; O'Rear, J.; Salavati, R.; Stuart, K., Separate insertion and deletion subcomplexes of the Trypanosoma brucei RNA editing complex. *Mol Cell* **2003,** 12, (2), 307-19.

48. Wang, S. P.; Deng, L.; Ho, C. K.; Shuman, S., Phylogeny of mRNA capping enzymes. *Proc Natl Acad Sci U S A* **1997,** 94, (18), 9573-8.

49. Ho, C. K.; Wang, L. K.; Lima, C. D.; Shuman, S., Structure and mechanism of RNA ligase. *Structure* **2004,** 12, (2), 327-39.

50. Feller, S. E.; Zhang, Y. H.; Pastor, R. W.; Brooks, B. R., Constant-Pressure Molecular-Dynamics Simulation - the Langevin Piston Method. *Journal of Chemical Physics* **1995,** 103, (11), 4613-4621.

51. Darden, T.; York, D.; Pedersen, L., Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *Journal of Chemical Physics* **1993,** 98, (12), 10089-10092.

52. Schindler, T.; Bornmann, W.; Pellicena, P.; Miller, W. T.; Clarkson, B.; Kuriyan, J., Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science* **2000,** 289, (5486), 1938-42.

53. Noble, M. E.; Endicott, J. A.; Johnson, L. N., Protein kinase inhibitors: insights into drug design from structure. *Science* **2004,** 303, (5665), 1800-5.

54. Sanner, M. F.; Olson, A. J.; Spehner, J. C., Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* **1996,** 38, (3), 305-20.

55. Suydam, I. T.; Snow, C. D.; Pande, V. S.; Boxer, S. G., Electric Fields at the Active Site of an Enzyme: Direct Comparison of Experiment with Theory. *Science* **2006,** 313, 200-204.

56. Huang, C. E.; Cruz-Reyes, J.; Zhelonkina, A. G.; O'Hearn, S.; Wirtz, E.; Sollner-Webb, B., Roles for ligases in the RNA editing complex of Trypanosoma brucei: band IV is needed for U-deletion and RNA repair. *Embo J* **2001,** 20, (17), 4694-703.

57. Garcia, A. E., Large-Amplitude Nonlinear Motions in Proteins. *Physical Review Letters* **1992,** 68, (17), 2696-2699.

58. Balsera, M. A.; Wriggers, W.; Oono, Y.; Schulten, K., Principal Component Analysis and Long Time Protein Dynamics. *J. Phys. Chem.* **1996,** 100, (7), 2567-2572.

59. Hayward, S.; Kitao, A.; Go, N., Harmonic and anharmonic aspects in the dynamics of BPTI: a normal mode analysis and principal component analysis. *Protein Sci* **1994,** 3, (6), 936-43.

60. Mongan, J., Interactive essential dynamics. *J Comput Aided Mol Des* **2004,** 18, (6), 433-6.

61. O'Donoghue, P.; Luthey-Schulten, Z., On the evolution of structure in aminoacyl-tRNA synthetases. *Microbiol Mol Biol Rev* **2003,** 67, (4), 550-73.

62. O'Donoghue, P.; Luthey-Schulten, Z., Evolutionary profiles derived from the QR factorization of multiple structural alignments gives an economy of information. *J Mol Biol* **2005,** 346, (3), 875-94.

63. Stuart, K.; Allen, T. E.; Heidmann, S.; Seiwert, S. D., RNA editing in kinetoplastid protozoa. *Microbiology and molecular biology reviews : MMBR* **1997,** 61, (1), 105.

64. Seiwert, S. D.; Heidmann, S.; Stuart, K., Direct visualization of uridylate deletion in vitro suggests a mechanism for kinetoplastid RNA editing. *Cell* **1996,** 84, (6), 831.

65. Simpson, L.; Sbicego, S.; Aphasizhev, R., Uridine insertion/deletion RNA editing in trypanosome mitochondria: a complex business. *RNA (New York, N.Y.)* **2003,** 9, (3), 265.

66. Stuart, K. D.; Schnaufer, A.; Ernst, N. L.; Panigrahi, A. K., Complex management: RNA editing in trypanosomes. *Trends in biochemical sciences* **2005,** 30, (2), 97.

67. Panigrahi, A. K.; Gygi, S. P.; Ernst, N. L.; Igo, R. P., Jr.; Palazzo, S. S.; Schnaufer, A.; Weston, D. S.; Carmean, N.; Salavati, R.; Aebersold, R.; Stuart, K. D., Association of Two Novel Proteins, TbMP52 and TbMP48, with the Trypanosoma brucei RNA Editing Complex. *Molecular and Cellular Biology* **2001,** 21, (2), 380.

68. Amaro, R. E.; Sethi, A.; Myers, R. S.; Davisson, V. J.; Luthey-Schulten, Z. A., A network of conserved interactions regulates the allosteric signal in a glutamine amidotransferase. *Biochemistry* **2007,** 46, (8), 2156-73.

69. Pascal, J. M., DNA and RNA ligases: structural variations and shared mechanisms. *Curr Opin Struct Biol* **2008,** 18, (1), 96-105.

70. Norberg, J.; Nilsson, L., Molecular dynamics applied to nucleic acids. *Acc Chem Res* **2002,** 35, (6), 465-72.

71. Case, D. A.; Darden, T. A.; Cheatham, I., T.E. ; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *Amber 8, University of California, San Francisco*, 2004.

72. El Omari, K.; Ren, J.; Bird, L. E.; Bona, M. K.; Klarmann, G.; LeGrice, S. F. J.; Stammers, D. K., Molecular architecture and ligand recognition determinants for T4 RNA ligase. *Journal of Biological Chemistry* **2006,** 281, (3), 1573-1579.

73. Vriend, G., WHAT IF: A molecular modeling and drug design program. *J. Mol. Graph.* **1990,** 8, 52-56.

74. Meagher, K. L.; Redman, L. T.; Carlson, H. A., Development of polyphosphate parameters for use with the AMBER force field. *J Comput Chem* **2003,** 24, (9), 1016-25.

75. Rao, S. T.; Westhof, E.; Sundaralingam, M., Exact method for the calculation of pseudorotation parameters Ptm and their errors. A comparison of the Altona-Sundaralingam and Cremer-Pople treatment of puckering of five-membered rings. *Acta Cryst.* **1981,** A37, 421-425.

76. Arora, K.; Schlick, T., Deoxyadenosine sugar puckering pathway simulated by the stochastic difference equation algorithm. *Chemical Physics Letters* **2003,** 378, (1-2), 1-8.

77. Andrea Amadei, A. B. M. L. H. J. C. B., Essential dynamics of proteins. *Proteins: Structure, Function, and Genetics* **1993,** 17, (4), 412-425.

78. Nandakumar, J.; Shuman, S., How an RNA ligase discriminates RNA versus DNA damage. *Mol Cell* **2004,** 16, (2), 211-21.

79. Hakansson, K.; Wigley, D. B., Structure of a complex between a cap analogue and mRNA guanylyl transferase demonstrates the structural chemistry of RNA, capping. *Proceedings of the National Academy of Sciences of the United States of America* **1998,** 95, (4), 1505-1510.

80. Qiu, L.; Gulotta, M.; Callender, R., Lactate dehydrogenase undergoes a substantial structural change to bind its substrate. *Biophys J* **2007,** 93, (5), 1677-86.

81. Fersht, A., *Structure and Mechanism in Protein Science*. W. H. Freeman and Company: New York, 1999.

82. Bairoch, A.; Apweiler, R., The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res* **2000,** 28, (1), 45-8.

83. Kern, D.; Zuiderweg, E. R., The role of dynamics in allosteric regulation. *Curr Opin Struct Biol* **2003,** 13, (6), 748-57.

84. Koshland, D. E., Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc Natl Acad Sci U S A* **1958,** 44, (2), 98-104.

85. Shuman, S., Structure, mechanism, and evolution of the mRNA capping apparatus. *Prog Nucleic Acid Res Mol Biol* **2001,** 66, 1-40.

86. Izaurralde, E.; Lewis, J.; McGuigan, C.; Jankowska, M.; Darzynkiewicz, E.; Mattaj, I. W., A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell* **1994,** 78, (4), 657-68.

87. Hamm, J.; Mattaj, I. W., Monomethylated cap structures facilitate RNA export from the nucleus. *Cell* **1990,** 63, (1), 109-18.

88. Shatkin, A. J., mRNA cap binding proteins: essential factors for initiating translation. *Cell* **1985,** 40, (2), 223-4.

89. Apostolakis, J.; Ferrara, P.; Caflisch, A., Calculation of conformational transitions and barriers in solvated systems: Application to the alanine dipeptide in water. *Journal of Chemical Physics* **1999,** 110, (4), 2099-2108.

90. Mills, G.; Jónsson, H., Quantum and thermal effects in H_{2} dissociative adsorption: Evaluation of free energy barriers in multidimensional quantum systems. *Physical Review Letters* **1994,** 72, (7), 1124.

91. Mills, G.; Jónsson, H.; Jacobsen, K. W., *Classical and Quantum Dynamics in Condensed Phase Simulations*. World Scientific, Singapore: 1998.

92. Khavrutskii, I. V.; Arora, K.; Brooks, C. L., 3rd, Harmonic Fourier beads method for studying rare events on rugged energy surfaces. *J Chem Phys* **2006,** 125, (17), 174108.

93. Ravindranathan, K. P.; Gallicchio, E.; Levy, R. M., Conformational equilibria and free energy profiles for the allosteric transition of the ribose-binding protein. *Journal of Molecular Biology* **2005,** 353, (1), 196-210.

94. Marco, E.; Foucaud, M.; Langer, I.; Escrieut, C.; Tikhonova, I. G.; Fourmy, D., Mechanism of activation of a G protein-coupled receptor, the human cholecystokinin-2 receptor. *Journal of Biological Chemistry* **2007,** 282, (39), 28779-28790.

95. Cheng, X.; Wang, H.; Grant, B.; Sine, S. M.; McCammon, J. A., Targeted molecular dynamics study of C-loop closure and channel gating in nicotinic receptors. *PLoS Comput Biol* **2006,** 2, (9), e134.

96. Loncharich, R. J.; Brooks, B. R.; Pastor, R. W., Langevin Dynamics of Peptides - the Frictional Dependence of Isomerization Rates of N-Acetylalanyl-N'-Methylamide. *Biopolymers* **1992,** 32, (5), 523-535.

97. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *Journal of Computational Physics* **1977,** 23, (3), 327-341.

98. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R., Molecular-Dynamics with Coupling to an External Bath. *Journal of Chemical Physics* **1984,** 81, (8), 3684-3690.

99. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **1996,** 14, (1), 33-&.

100. Lou, H. F.; Cukier, R. I., Molecular dynamics of apo-adenylate kinase: A principal component analysis. *Journal of Physical Chemistry B* **2006,** 110, (25), 12796-12808.

101. Rinaldo, D.; Field, M. J., A computational study of the open and closed forms of the N-lobe human serum transferrin apoprotein. *Biophysical Journal* **2003,** 85, (6), 3485-3501.

102. Gerstein, M.; Lesk, A. M.; Chothia, C., Structural Mechanisms for Domain Movements in Proteins. *Biochemistry* **1994,** 33, (22), 6739-6749.

103. Doherty, A. J.; Suh, S. W., Structural and mechanistic conservation in DNA ligases. *Nucleic Acids Research* **2000,** 28, (21), 4051-4058.

104. Henzler-Wildman, K. A.; Thai, V.; Lei, M.; Ott, M.; Wolf-Watz, M.; Fenn, T.; Pozharski, E.; Wilson, M. A.; Petsko, G. A.; Karplus, M.; Hubner, C. G.; Kern, D., Intrinsic motions along an enzymatic reaction trajectory. *Nature* **2007,** 450, (7171), 838-44.

105. Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G., The Energy Landscapes and Motions of Proteins. *Science* **1991,** 254, (5038), 1598-1603.

106. Onuchic, J. N.; Nymeyer, H.; Garcia, A. E.; Chahine, J.; Socci, N. D., The energy landscape theory of protein folding: Insight into folding mechanisms and scenarios. *Advances In Protein Chemistry* **2000,** 53, 87 - 152.

107. Bahar, I.; Chennubhotla, C.; Tobi, D., Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. *Curr Opin Struct Biol* **2007,** 17, (6), 633-40.

108. Kumar, S.; Ma, B.; Tsai, C. J.; Sinha, N.; Nussinov, R., Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci* **2000,** 9, (1), 10-9.

109. Szabo, A.; Shoup, D.; Northrup, S. H.; Mccammon, J. A., Stochastically Gated Diffusion-Influenced Reactions. *Journal of Chemical Physics* **1982,** 77, (9), 4484-4493.

110. Wade, R. C.; Davis, M. E.; Luty, B. A.; Madura, J. D.; Mccammon, J. A., Gating of the Active-Site of Triose Phosphate Isomerase - Brownian Dynamics Simulations of Flexible Peptide Loops in the Enzyme. *Biophysical Journal* **1993,** 64, (1), 9-15.

111. Chang, C. E.; Shen, T.; Trylska, J.; Tozzini, V.; McCammon, J. A., Gated binding of ligands to HIV-1 protease: Brownian dynamics simulations in a coarse-grained model. *Biophys J* **2006,** 90, (11), 3880-5.

112. Gabdoulline, R. R.; Wade, R. C., Brownian dynamics simulation of protein-protein diffusional encounter. *Methods* **1998,** 14, (3), 329-41.

113. Northrup, S. H.; Zarrin, F.; Mccammon, J. A., Rate Theory for Gated Diffusion-Influenced Ligand-Binding to Proteins. *Journal of Physical Chemistry* **1982,** 86, (13), 2314-2321.

114. Swift, R. V.; McCammon, J. A., Catalytically Requisite Conformational Dynamics in the mRNA-Capping Enzyme Probed by Targeted Molecular Dynamics. *Biochemistry* **2008,** 47, (13), 4102-11.

115. Caves, L. S.; Evanseck, J. D.; Karplus, M., Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. *Protein Sci* **1998,** 7, (3), 649-66.

116. Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A., Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America* **2001,** 98, (18), 10037-10041.

117. Gabdoulline, R. R.; Wade, R. C., Effective charges for macromolecules in solvent. *Journal of Physical Chemistry* **1996,** 100, (9), 3868-3878.

118. Elcock, A. H.; Gabdoulline, R. R.; Wade, R. C.; McCammon, J. A., Computer simulation of protein-protein association kinetics: Acetylcholinesterase-fasciculin. *Journal of Molecular Biology* **1999,** 291, (1), 149-162.

119. Ermak, D. L.; Mccammon, J. A., Brownian Dynamics with Hydrodynamic Interactions. *Journal of Chemical Physics* **1978,** 69, (4), 1352-1360.

120. Garcia De La Torre, J.; Huertas, M. L.; Carrasco, B., Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys J* **2000,** 78, (2), 719-30.

121. Northrup, S. H.; Allison, S. A.; Mccammon, J. A., Brownian Dynamics Simulation of Diffusion-Influenced Bimolecular Reactions. *Journal of Chemical Physics* **1984,** 80, (4), 1517-1526.

122. Messina, T. C.; Talaga, D. S., Protein free energy landscapes remodeled by ligand binding. *Biophysical Journal* **2007,** 93, (2), 579-585.

123. Bui, J. M.; McCammon, J. A., Protein complex formation by acetylcholinesterase and the neurotoxin fasciculin-2 appears to involve an induced-fit mechanism. *Proc Natl Acad Sci U S A* **2006,** 103, (42), 15451-6.

124. Arora, K.; Brooks, C. L., 3rd, Large-scale allosteric conformational transitions of adenylate kinase appear to involve a population-shift mechanism. *Proc Natl Acad Sci USA* **2007,** 104, (47), 18496-501.

125. Gabdoulline, R. R.; Wade, R. C., Simulation of the diffusional association of barnase and barstar. *Biophys J* **1997,** 72, (5), 1917-29.

126. Crank, J., *The Mathematics of Diffusion*. second ed.; Oxford Science Publications: 1973.

127. McCammon, J. A.; Karplus, M., Internal motions of antibody molecules. *Nature* **1977,** 268, (5622), 765-6.