Theoretical studies of RNA catalytic activity

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE
SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

George Mădălin Giambaşu

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHYLOSOPHY

Adviser: Darrin M. York
August 2010

ACKNOWLEDGEMENTS

To my grandparents.

ABSTRACT

RNA possess a diverse ensemble of conformations that interchange on time scales that range from femtoseconds to milliseconds. This conformational variability has a profound effect on RNA function, and has been exploited in RNA molecules such as aptamers, aptazymes, and riboswitches that have been engineered to be allosterically controlled.

Central to this work is the questions of how RNA conformational variability affects with its catalytic activity. We employ a set of diverse theoretical and simulation tools, such as molecular dynamics, free energy calculations, and path sampling that greatly extend the capability of structural and molecular biology experiments to reveal atomic level details of RNA energetics and dynamics.

Our systems of interest are two prototypical catalytic RNA systems: L1 ligase and Hammerhead ribozymes, both of which are postulated to reach their catalytically active states through complex mechanisms that involve large domains movements, change in base pairing patterns and direct participation of divalent metal ions. In the case of L1 ligase, we reveal the structural basis of its allosteric control that involves an 80 angstrom swing of one of its stems, as well as the role of a flexible active site toward providing catalytic selectivity. In the case of Hammerhead ribozyme, we characterize the conformation and dynamics of several constructs in different ionic environments that lead to catalysis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

IDENTIFICATION OF THE HINGE POINTS OF L1 LIGASE RIBOZYME

The L1 ligase is an *in vitro* selected ribozyme that uses a non-canonically base-paired ligation site to catalyze regioselectively and regiospecifically the 5′ to 3′ phosphodiester bond ligation, a reaction relevant to origin of life hypotheses that invoke an RNA World scenario. The L1 ligase crystal structure revealed two different conformational states that were proposed to represent the active and inactive forms. It remains an open question as to what degree these two conformers persist as stable conformational intermediates in solution, and along what pathway are they able to inter-convert. To explore these questions, we have performed a series of molecular dynamics simulations in explicit solvent of the inactive–active conformational switch in L1 ligase. Four simulations were performed departing from both conformers in both the reactant and product states, in addition to a simulation where local unfolding in the active state was induced. From these simulations, along with crystallographic data, a set of 4 virtual torsion angles that span two evolutionarily conserved and restricted regions were identified as dynamical hinge points in the conformational switch

transition. The ligation site visits three distinct states characterized by hydrogen bond patterns that are correlated with the formation of specific contacts that may promote catalysis. The insights gained from these simulations contribute to a more detailed understanding of the coupled catalytic/conformational switch mechanism of L1 ligase that may facilitate the design and engineering of new catalytic riboswitches.

1.1   Introduction

The RNA World hypothesis proposes that life originated from catalytic RNA molecules capable of replication via template-dependent assembly of RNA fragments of nucleotide monomers. To do so, RNA replicases would be required to catalyze the regiospecific condensation of a $3'$ ribose hydroxyl upon the $5'$ α phosphate of an RNA or nucleotide $5'$ triphosphate, to yield the ubiquitous $3'$ to $5'$ phosphodiester linkage that is characteristic of all extant genomic nucleic acids. However, no known naturally occurring ribozyme catalyzes this phosphodiester assembly reaction. The concern that RNA might be inherently incapable of catalyzing this reaction was put to rest in 1993 with the first *in vitro* evolution of a ribozyme ligase. [1] Subsequently, several other ribozyme ligases have been produced using *in vitro* selection techniques, [2, 3, 4, 5, 6, 7, 8, 9] including a small subset that specifically catalyze regiospecific $3'$ to $5'$ phosphodiester linkages characteristic of all extant RNA and DNA polymerases. The L1 ligase ribozyme [3] is one such example, and is unusual in that it uses an intrinsically flexible non-canonically base paired ligation site (Fig. 1). [10, 2] The crystal structure of the L1 ligase product has been solved recently [11], providing "a glimpse of biology's first enzyme" [12].

In addition to its potential relevance to the origin of life, the L1 ligase ribozyme, presumably as a fortuitous consequence of *in vitro* selection, is an allosteric ribozyme molecular switch [13, 14, 15]. It is possible to take advantage of this property to engineer new L1 ligase ribozymes by further selections that enable derivatives to be controlled by small molecules, peptides, or even proteins, in effect, creating an artificial allosteric ribozymes

[3, 16, 17, 18].

The intrinsic flexibility of the L1 ligase, was verified from the crystallized ligation product, in which two crystallographically independent conformations, both products of self-ligation, were resolved in the same asymmetric unit. These conformers differed in the orientation of one of the stems (stem C) by a movement of the stem tip by around 80 Å [11]. Based on the presence/absence of specific contacts between the ligation site and evolutionarily conserved regions of the U_{38} loop of stem C, it was proposed that the conformers represented catalytically active "*on*" and inactive "*off* " states. [11]

Here, we address how changes in the L1 ligase conformation affect its function and catalytic activity. Generally, RNA switching and folding pathways occur along a rugged free energy landscape, characterized by multiple minima, low transition barriers and kinetic traps. [19, 20, 21, 22, 23] The ground states of RNA molecules are thus best characterized by an ensemble of conformers. The factors that can shift the equilibrium between conformers include formation/disruption of tertiary contacts, binding of ligands, metal ions (typically Mg^{2+}) or proteins by induced fit or tertiary structure capture. [19, 20, 21, 22, 23]

The minimal set of 8 torsion angles per nucleotide affords flexibility to the polynucleotide backbone, and has been used as the basis for the classification of different RNA folds. [24, 25, 26, 27] An alternative two-dimensional representation of the RNA backbone conformations can be achieved using a reduced set of so-called *virtual torsion angles* formed along the virtual bonds between $C_{4'}$ – P atoms that have been shown to be able to discriminate between major RNA folds. [28, 29]

In the current work, we identify dynamical hinge points of the L1 ligase ribozyme using large scale molecular dynamics simulations. We depart from an analysis of the two crystallized conformers, and show using over 600 ns of molecular dynamics simulations that the transition between *on* and *off* conformational states can be almost entirely described by changes in only four virtual torsion angles. The insights gained from these simulations are a first step toward a detailed understanding of the coupled catalytic-conformational

riboswitch mechanism of L1 ligase.

1.2 Four virtual torsions can be used to distinguish between the two crystallized conformers

1.2.1 Restricted parts of the L1 ligase junction differ between the docked and undocked conformers.

Fig. 2 illustrates the crystallographic presumably active "docked" and inactive "undocked" conformers. Superposition of the three individual stems (A, B and C), as found in the crystal structure, show small deviations (RMSD 1.71, 0.71 and 2.06 Å, respectively) between the docked and undocked conformational states (Table 5). Moreover, the base pair hydrogen bonding patterns are almost completely conserved between conformers (Table 2), the main differences being located at the terminal GAAA tetraloops at the ends of the stems. Most of the structural variation is observed in the junction region connecting the stems where the RMSD between crystal conformers is 8.30 Å. This variation in the junction propagates to a large scale swing of stem C by around 80 Å. When stems A and B together of the docked and undocked conformers are superimposed with RMS fitting, the combined RMSD for stems A and B is 4.9 Å, whereas the RMSD for stem C (based on the same superposition) is 48.4 Å. This indicates that the relative orientations of stems A and B are similar between the two crystallographic conformers, whereas the orientation of stem C relative to stems A and B is vastly different.

1.2.2 Reorientation of stem C is affected mainly by changes in four virtual torsion angles: θ_{18} , θ_{37}, θ_{44}, η_{38}.

The two crystallographic conformers can be distinguished by changes in their virtual torsion angles (Fig. 2). [28, 29] [*] A survey of the 142 virtual torsion angles in the L1 ligase (2 torsions per each of the 71 nucleotides) reveals that only four show a significant change

---

[*] The backbone real torsions analysis shown in Fig. 4 as well as the and ribose pucker analysis shown Table 3 is definitely a complex problem that doesnot show clear trends for the two crystallized conformations.

($> 45°$): θ_{18}, θ_{37}, θ_{44}, η_{38} (Fig. 3). These torsions span regions that contain the evolutionarily conserved residues: U_{37}, U_{38}, A_{39} for θ_{37} and η_{38} and the 5-base motif (C_{39}=G_{18} G_{37} A_{38} C_{17} / U_{39}=A_{18} U_{37} G_{38} U_{17}) and the neighboring U_{19} for θ_{18} and θ_{44} (for a detailed statistical analysis of the conserved residues see reference [18]). Both U_{38} and U_{19} are of particular interest in the current work: U_{38} contributes to the docking of stem C into stem A in the active conformation and mutation data demonstrates that it is critical for catalysis, and U_{19} is evolutionarily conserved (97%), although its role in catalysis remains unclear. [3, 16, 17, 18]

It is of interest to note that the shortest pathway (in the space of the virtual torsion angles) to affect a transition between the docked and undocked conformers would involve changes in θ_{18} and θ_{44} by arcs of 82.4 and 128.2 degrees, respectively. This transition, however, would *not* be possible as the path would require stem C to pass through regions of stem A. Rather, the sterically allowed pathway would require the transition to proceed along the longer complementary arcs of 277.6 and 231.8 degrees (Fig. 2).

1.3 Design and overall structure and dynamics of the simulations

The molecular dynamics simulation that have been carried out in this study are summarized in Table 4. These include: product in the docked conformation (Prod-D), product in the undocked conformation (Prod-D-U), product departing from the docked conformation with forced unfolding (Prod-D-UF), reactant precursor in the undocked conformation with extended triphosphate (Prec-D-XTP), reactant precursor in the undocked conformation with Mg^{2+}-coordinated polymerase-like triphosphate (Prec-D-MgTP).

1.3.1 Simulations of the docked and undocked product indicate a large reorientation of stem C.

A useful index to aid in the quantification of the positional deviation of one structure relative to another is to use an RMS fitting procedure to superimpose the structures, and

measure the resulting average RMSD. In order to isolate the precise nature of the positional deviations between the two structures, it is useful to define separately a fitting set (FS) of atoms used to define the superposition, and a possibly different measure set (MS) of atoms from which the RMSD is averaged. We designate this type of generalized RMSD as RMSD(FS,MS). Note, the RMSD(FS,MS) corresponds to the standard RMSD when the FS and MS are the same. Table 5 lists the RMSD(FS,MS) for various fitting sets and measuring sets in order to help quantify the structural deviation between crystallographic structures and structures from the MD simulations.

In the simulation of the docked product (Prod-D), the overall heavy-atom RMSD with respect to the initial structure (based on the docked crystal structure) was ~4.2 Å, and that of the individual stems ranged from ~1.8-3.0 Å (Table 5). On the contrary, the undocked product simulations (Prod-U) showed a much higher overall heavy-atom RMSD of ~7.9 Å relative to the starting structure (based on the undocked crystal structure), despite having RMSD values for the individual stems (2.5-3.3Å) similar to that of the docked simulation. This indicates that the structure in the simulation exhibits significant variation. The relatively low value of the RMSD(AB,AB) of 3.4 Å indicates that the relative orientation of stems A and B remain similar to the starting structure, whereas the RMSD(AB,C) value of 19.1 Å indicates that it is the orientation of stem C relative to stems A and B that exhibits large variations. Whereas the Prod-D simulation exhibits fairly small structural variations in the junction region as indicated by the RMSD(J,J) value of 2.1 Å, the prod-U simulation exhibits much larger variations having a RMSD(J,J) value of 4.1 Å. Together, these results indicate that restricted regions of the junction in the Prod-U simulation exhibit large local conformational variations, and as will be characterized in further detail below, act as hinge points for the re-orientation of stem C.

6

1.3.2 Simulations of unfolding of the docked conformer.

The unfolding procedure of the docked conformer is described in detail in the METHODS section. The Prod-D-UF simulation shows a larger overall RMSD relative to the crystallographically derived initial structure of 6.7 Å, whereas the corresponding RMSD for the Prod-D simulation is 4.2 Å. The RMSD of the individual stems, on the other hand, range from ~1.6 to 3.3 Å in the Prod-D-UF simulation. Whereas the orientation of stem B relative to stems A and C is fairly small [RMSD(AC,B) value of 5.5 Å], and similar to the value of 5.4 Å in the Prod-D simulation, the orientations of stem A relative to stem C is markedly larger as expected [RMSD(AB,C) and RMSD(BC,A) values of 13.5 and 16.7 Å, respectively], are intermediate between those of the Prod-D and Prod-U simulations. These movements of the stems are accompanied by a modest increase in the average RMSD of the junction from 2.1 Å in the Prod-D simulation to 2.3 Å in the Prod-D-UF simulation, and a more notable increase in the fluctuations (e.g., standard deviation of the RMSD in Table 5).

It is apparent from the Prod-U and Prod-D-UF simulations that when stem A and C are not docked (stabilized by tertiary interactions between stems) L1 ligase exhibits a higher degree of conformational flexibility. The origin of this flexibility can be traced to increased fluctuations within the junction region with individual stems keeping their internal structure close to that found in crystal.

1.3.3 Design of two arrangements of the ligation site in the reactant state.

It has been proposed that involvement of two Mg^{2+} ions enhances recognition or specificity of the ligation reaction. [30] L1 ligase is a Mg^{2+} dependent ribozyme. Generally, the ligation reaction consists of the formation of a phosphodiester bond between the $O_{3'}$ of the terminal nucleotide of the nucleic acid strand and the incoming P_α atom of a nucleotide triphosphate. The reaction proceeds through a pentacovalent transition state/intermediate with inversion of the P_α center stereochemistry, and results in the release of pyrophosphate.

7

To explore the structure and dynamics of the reactant precursor state of the L1 ligase, we designed two possible models based on the L1 ligase product structure in the docked conformation. These models differed only in the conformation and metal ion coordination of the extra GTP_1 pyrophosphate moiety (G_1 in the product state). The first precursor simulation (Prec-D-XTP) was based on an extended starting conformation of the pyrophosphate. The second precursor simulation (Prec-D-MgTP) was based on a homology model derived from the crystal structure of DNA polymerase beta as follows. The motivating premise for the Prec-D-MgTP model was that, in order to accommodate highly charged substrates such as GTP, catalytic sites require two Mg^{2+} ions. [30] We observed a high degree of similarity of the disposition of the charged groups and Mg^{2+} between the L1 ligase product structure and crystal structures of several protein polymerases. We have included an additional Mg^{2+} ion and used homology modeling to replicate the conformation of the pyrophosphate based on the crystal structure of human DNA polymerase beta [31] (PDB ID 1BPY). The similarity between the mechanism of transition state stabilization of the RNA and protein polymerases has been pointed out recently based on the analysis of the crystallized catalytic core of a in vitro optimized class I ligase ribozyme. [8, 9] We have followed the naming convention for the two Mg^{2+} ions following reference [30] with the ion closest to the reaction site labeled A and the other labeled B. Data presented in table 5 indicate that the overall structure and dynamics of these two simulations followed a similar trend as the Prod-D simulation.

1.4 *Intrinsic flexibility of the U_{38} loop*

The observation that U_{38}, a conserved residue critical for catalytic activity in the L1 ligase family, is docked into the ligation site and makes a canonical base pair with a constituent of the ligation site A_{51} in the docked-active conformation, whereas in the undocked-inactive conformation is positioned ~40 Å away from the site, has led to the postulate that the former is more likely representative of a catalytically active state. [11] The simulation

results presented here provide insight into the transition between docked and undocked conformations observed in the crystal and their implications for catalysis. Of particular focus in this section is the conformational change of the U_{38}-containing loop (designated henceforth as the "U_{38} loop").

1.4.1 The conformational change of the U_{38} loop can be tracked by the θ_{37} and η_{38} virtual torsions

In the crystal structures, the conformation of the U_{38} loop varies considerably between the docked and undocked conformers, showing two different conformations. Two virtual torsions that span the loop, θ_{37} and η_{38}, show a large variation (~120.0 deg) between the two crystal conformers, whereas all the other virtual torsions in the immediate base-paired vicinity show much smaller variation ($\leq$ 20.0 deg) (Fig. 3). This is suggestive that during the undocked $\rightleftharpoons$ docked conformational transition, the U_{38} loop is highly flexible along the θ_{37} and η_{38} virtual torsions. In particular, the loop has to transition from a state where U_{38} is buried into stem C and makes a sugar-base hydrogen bond with A_{23} (Table 2) into an open conformation that will expose U_{38} in solution giving it the possibility to dock into stem A. The docking consists in the formation of a canonical base $U_{38}=A_{52}$ and several phosphate-Mg^{2+} contacts, supported by A_{39} and G_{40} of stem C on one side and G_1 of stem A on the other.

1.4.2 The simulations span the full range of the L1 ligase conformational transition involving the U_{38} loop.

In the Prod-D, Prec-D-XTP and Prec-D-MgTP simulations, U_{38} remains base-paired with A_{51}, and the overall behavior of the θ_{37} and η_{38} virtual torsions are similar. For the Prod-D simulation, the average values for the θ_{37} and η_{38} torsions (220.3 and 27.0 deg, respectively) stay within 43 degrees of the values found in the docked crystal conformer (243.8 and -15.5 deg, respectively). The θ_{37} and η_{38} torsions exhibit larger fluctuations with respect to other nearby virtual torsion angles (e.g., the set of 4 torsions upstream from η_{38} and 4 torsions downstream from θ_{37}). In the Prod-D simulation, the standard deviation for θ_{37} and η_{38}

virtual torsions were 36.4 and 60.0 deg, respectively, whereas the corresponding values for the nearby virtual torsions remained within 16.7 deg. Similar trends were observed for Prec-D-XTP and Prec-D-MgTP simulations (data not shown).

In the Prod-U simulation, the θ_{37} and η_{38} virtual torsions follow a similar trend with respect to their undocked conformer values. Averages were 192.1 and 128.7 deg, respectively, compared to 135.9 and 105.9 deg in the undocked crystal structure. The standard deviations (32.6 and 28.2 deg, respectively) were lower than in than in the Prod-D simulation. For these set of simulations (Prod-D, Prec-D-XTP, Prec-D-MgTP and Prod-U) U_{38} fluctuates relatively close to its starting geometry in two approximately separate states indicated with different colors in Fig. 6b (lower panel). These two regions are labeled D (solvent exposed/docked) and U (buried), respectively, and two representative structures are depicted in Fig. 6a. As shown by the correlation plots of θ_{37} and η_{38} virtual torsions (Fig. 6b, upper panel) in the Prod-D-UF simulation the U_{38} loop is able to span the conformational space that covers almost exactly the same area that includes the reunion of the regions sampled during Prod-D and Prod-U simulations, encompassing the conformations found in crystal.

1.5 Three-way junction flexibility: U_{19} makes specific contacts with stem B

The high conformational variability of the L1 ligase backbone in the crystallized conformers suggests that the junction region is the main structural element that is responsible for the large flexibility of the riboswitch. Two virtual torsions that span the junction have to travel 277.6 deg, θ_{18}, and 231.8 deg, θ_{44}, to switch from docked to undocked conformation (Fig. 3)

Three and four-way junctions constitute extremely important scaffolds for RNA riboswitches or aptamers. [32] The flexibility in the junction plays a key role in catalytic

activity in the L1 ligase family. [33] It was found that there are only two pentuple nucleotide structural motifs (C_{39}=G_{18} G_{37} A_{38} C_{17} or U_{39}=A_{18} U_{37} G_{38} U_{17}) located in the junction region that support catalysis. [33] These motifs are preserved during the evolutionary optimization of L1 ligase constructs, but it is yet unclear how they can affect the dynamics of the ligase in order to promote catalysis.

Additionally, U_{19} is a conserved residue in all the L1 ligase constructs. However, in both crystallographic conformers U_{19} is oriented toward the exterior of the L1 ligase body without making any contacts. The lack of clear structural basis that explains the conserved character of U_{19} in either the docked or undocked crystal structures motivates the question as to whether this residue might be important in facilitating the undocked $\rightleftharpoons$ docked conformational switch. This section explores the possible role of U_{19} in stabilizing intermediate states in this transition.

1.5.1 Relatively larger fluctuations in the three-way junction suggest the presence of the hinge points.

In the Prod-D and Prod-U simulations, the virtual torsions associated with the residues that form the junction region (J) are characterized by an residue-average standard deviation (RASD) of 16.1 deg and 18.5 deg, respectively. The reader is pointed to the METHODS section for the definition of the set of specific residues that are included in junction region (J) and the exact definition of the virtual torsions and how they are associated with a specific RNA residue. The values of RASD for the entire set of 142 virtual torsions were 14.2 and 15.4 deg, respectively. As a second reference the RASD for a base paired structure, like stem A, is ~ 12.5 Å for the Prod-U simulation. These values suggest that the junction possesses a higher degree of intrinsic flexibility relative to other regions of the L1 ligase. The major contributors (SD $\geq$ 30.0) to these large values of RASD are θ_{44} (SD 48.7 deg) and η_{45} (SD 71.0 deg) for the Prod-D simulation, and θ_{18} (SD 56.8 deg), η_{19} (SD 30.7 deg), θ_{19} (SD 37.8 deg), η_{20} (SD 53.6 deg), and θ_{20} (SD 30.8 deg) for the Prod-U simulation. It's worth noting that through this two sets we find θ_{44} and θ_{19} - two virtual torsions out of a

total of four that can be used to distinguish between the two crystallized conformers.

1.5.2 Exploring connections between the fluctuations of the junction and the conformational switch.

In the Prod-D, Prec-D-MgTP and Prec-D-XTP simulations the the overall structure of L1 Ligase fluctuates around the common starting docked conformation, with an overall RMSD of 4.21 Å(Table 5). The Prod-D-UF simulation samples conformations in the vicinity of the docked conformer in addition to exploring conformations along the path toward the undocked conformation. The overall RMSD (Table 5) 6.66 Åis larger than that of the Prod-D simulation. As shown in Fig. 5 and stem C is oscillating away from its docked conformation and closer to the set of conformations derived from the Prod-U simulation. RMSD(AB,C) for the Prod-D-UF simulation is 13.51 Å, whereas the same quantity for the Prod-D simulation is 7.16 Å. In the case of the Prod-D-UF simulation, θ_{44} and η_{45} virtual torsions have the largest standard deviations in the junction region ($\sim$ 30.0 deg) .

A question that emerges from this data is what is the connection between these relatively large fluctuations of some of the virtual torsions that span the junction region and the transition from the inactive to the active conformation. During Prod-U simulation θ_{18} is observed to transition through a series of two relatively distant clusters (Fig. 8) that are disposed exactly on the same path that would allow the undocked $\rightleftharpoons$ docked conformational transformation to occur via changes in the backbone virtual torsions apparent from the crystal structure. The transition along θ_{18} is not unique in the junction region during the Prod-U simulation, the other angles that show large SD during Prod-U simulations populated more than one state and were partially correlated with θ_{18}.

The conserved U_{19} interacts specifically with stem B, and accounts for its possible role in the stabilization of intermediate states on the conformational transition pathway. For the duration of the Prod-D, Prod-D-UF and precursors simulations, U_{19} remains exposed to the solvent, with no specific contacts with any part of the L1 ligase. However, during the Prod-U simulations, a series of interaction patterns involving U_{19} and stem B that

correlate with the transition along the θ_{18} virtual torsion were observed. Their succession is presented from different perspectives in Fig. 7. We identify two major structural arrangements, labeled A and C (Fig. 7). A transition structure, labeled B, was chosen to be approximately equidistant from A and C in the θ_{18} virtual torsion space. Once θ_{18} passes through the transition structure B (vertical dashed line) after only ~10 ns, U_{19} starts to make contacts with stem B through a series of hydrogen bonding patterns that interchange during the course of the simulation. These interaction patterns include base–base (U_{19}:N_3–C_{16}:O_2, U_{19}:O_2–G_9:N_2, U_{19}:N_3–C_{10}:O_2), base–backbone (U_{19}:$O_{2'}$–C_{16}:O_2), and backbone–backbone (U_{19}:$O_{2'}$–C_{17}:$O_{2'}$, C_{17}:$O_{2'}$–U_{20}:O_{1P}) contacts. Given the conserved nature of the interacting partners, it is possible that these base–base interactions play an important role in the stabilization of the intermediate states during transitions between the two conformations, and contribute in part to the discrimination in the evolutionary optimization process between active and inactive constructs.

1.6 High variability of the non-canonically based paired ligation site has implications on catalysis

1.6.1 The anatomy of the ligation site in the two crystal conformers is different.

L1 ligase catalyzes the phosphodiester bond formation between GTP_1:P_α and U_{71}:$O_{3'}$. Generally this type of nucleotidyl transfer reaction proceeds through a pentacovalent transition state or metastable intermediate that causes the inversion of the P_α stereochemistry, and results in the release of pyrophosphate. [34, 35] The non-canonically base-paired ligation site on stem A is built from 3 base-pairs: U_{50}•G_2 (cis-WC/WC), $A_{51}\square\rightarrow G_1$ (trans-Hoogsten/sugar edge), G_{52}•U_{71}(cis-WC/WC) (Fig. 5, right inset). In the crystallographic docked product conformation A_{51} is also involved in a canonical base pair with U_{38} giving rise to base triple U_{38}–$A_{51}\square\rightarrow G_1$. This interaction along with the interactions between G_1, A_{39} and G_{40} phosphates with a Mg^{2+} ion are responsible for mediating the docking of stem C into stem A. The formation of the base triple U_{38}–$A_{51}\square\rightarrow G_1$ disrupts the hydrogen bond

between $G_1:O_{2'}$ and $A_{51}:N_6$. This hydrogen bond is present in the undocked conformer crystal structure and is typical for a trans-Hoogsten/sugar edge base pair. [36] The distance $G_1:O_{2'}$ and $A_{51}:N_6$ is 3.2 Å in the undocked conformer and 4.8 Å in the docked conformer (Table 2).

1.6.2 Dynamics of the ligation site in solution simulations.

During the Prod-D, Prec-XTP, Prec-MgTP simulations the ligation site shows a strong variability, spanning a series of three conformational clusters characterized by the formation of specific hydrogen bond interactions between G_1/GTP_1 and the rest of the ligation site. G_1/GTP_1 oscillates between (Fig. 11): formation of hydrogen bond with G_{52}, $GTP_1/G_1:N2–G_{52}:O6$ (cluster 1), a base triple with A_{51} and U_{38} that is identical to that found in crystallized docked conformer (cluster 2), and a base triple with U_{50} and G_2 (cluster 3).

The average values and standard deviations of the hydrogen bond distances for each cluster are shown in Table 6. We also indicate which contact yields a hydrogen bond in each simulation. The hydrogen bond pattern in the Prod-U simulation follows a different trend with respect to the Prod-D, Prec-XTP, Prec-MgTP simulations, occupying only one of the clusters (cluster 2), identical to that found in undocked crystal conformer, that corresponds to a typical trans-Hoogsten/sugar edge base pair between A_{51} and G_1.

One of the structural features that confers an increased flexibility for G_1/GTP_1 and its ability to interact with different parts of the ligation site for the Prod-D, Prec-XTP, Prec-MgTP simulations resides in the way $GTP_1/G_1:O_{2'}$ interacts with the ligation site and the solvent. During Prod-D, Prec-XTP, Prec-MgTP simulations $GTP_1/G_1:O_{2'}$ shows no tendency to recover the $G_1:O_{2'}–A_{51}:N_6$ hydrogen bond present in the undocked crystal and Prod-U simulation (Table 6), but instead interacts alternately with bulk water molecules or $G_2:O_{2P}$. Docking of stem C into stem A causes the rupture of the $G_1:O_{2'}–A_{51}:N_6$ hydrogen bond in the active state, and has the effect of inducing a conformational variability in the

ligation site of the product and precursor/reactant state.

1.6.3 Specific conformations of the catalytic site support the first stage of the catalytic event.

We explored the impact of the variability of the ligation site on the first step of the catalytic reaction: the deprotonation of $U_{71}:O_{3'}$ atom. This was assessed by analyzing the possible hydrogen bonds formed with a potential general bases near the ligation site. Several QM/MM studies on enzymatic systems that catalyze the nucleotidyl transfer reaction have hypothesized three routes for the nucleophile ($U_{71}:O_{3'}$ in the case of L1 Ligase) deprotonation. [37, 38, 39] The general base might be: (i) one of the non-bridging oxygens on of the P_α atom of the triphosphate, (ii) a carboxylate placed of an ASP residue or (iii) a water molecule that functions as a proton shuttle.

During the simulations, the nucleophile $U_{71}:H_{O3'}$ is involved in contacts with $U_{71}:O_{2'}$ and $GTP_1:O_{2P\alpha}$, of which $GTP_1:O_{2P\alpha}$ appears to be a strong candidate for a general base. These hydrogen bonds are mainly formed from structures in cluster 1 and 3, with negligible probability for cluster 2, as shown in Fig. 11. The variability of the ligation site (i.e. the capacity to visit several conformational clusters) makes it possible to induce the formation of specific contacts that might promote the chemical step in catalysis (i.e., the contact between $U_{71}:H_{O3'}$ and $GTP_1:O_{2P\alpha}$). This is due to the non-canonical base pairing of the ligation site. Since more than one conformation might favor the deprotonation of $U_{71}:O_{3'}$ (both cluster 1 and 3 are viable), the L1 ligase ligation site may favor on multiple catalytic pathways.

1.6.4 Overall arrangement and dynamics of the ligation site accounts for the regioselectivity and regiospecificity of the L1 Ligase.

The arrangement of the ligation site accounts for the regiospecificity of the L1 Ligase, equivalent to its capability to specifically catalyze the 5′–3′ and not 5′–2′ or 5′–5′ phosphodiester bond formation (Table 6). First, for all duration of the simulations $U_{71}:O_{3'}$ remained exclusively directly bounded to the Mg^{2+} ion A, fact that will inevitably decrease

15

its pK_a and make it easily to be deprotonated and to further attack the $GTP_1:P_\alpha$. On the other hand, the other candidate for the nucleophile role in ligation, $U_{71}:O_{2'}$ makes no contacts with any Mg^{2+} ion or other potential pK_a lowering entity having thus a lower inclination to be deprotonated and to further attack $GTP_1:P_\alpha$. Second, Mg^{2+} ion A supports the regiospecificity by maintaining the proximity of the $U_{71}:O_{3'}$ and $GTP_1:P_\alpha$ by electrostatic interactions, so once deprotonated $GTP_1:P_\alpha$ will be close to attack. $U_{71}:O_{2'}$ is positioned at a larger distance from $GTP_1:P_\alpha$ and, consequently, has a decreased probability to come in the proximity of $GTP_1:P_\alpha$.

Mg^{2+} ion A and B have a major role in supporting the docking of stems A and C by compensating the repelling electrostatic interactions of the negatively charged backbone. In one of our. Along with that, Mg^{2+} ion A is possibly perturbing the $U_{71}:O_{3'}$ pK_a to increase its nucleofilicity At this point, there is no clear indication if the presence/ absence of Mg^{2+} ion B might affect the reactive pathway/pathways. There are postulates for similar proteic systems that support a direct role in the electrostatic stabilization of the transition state.[30]

1.7 Conclusion

RNA has a diverse ensemble of conformations that interchange on time scales that range from femtoseconds to microseconds [40]. This conformational variability allows RNA molecules to be allosterically controlled such as in the case of aptamers and aptazymes [41], or riboswitches [42]. However, due to RNA's rugged conformational landscape, structural biology methods such as X-ray crystallography and NMR face challenges to capture an accurate, complete picture of the range of conformations that might play important roles in function [20]. On the other hand, molecular simulation methods provide a wealth of detail into both structure and dynamics, and offer a powerful tool to complement structural biology, biochemical and biophysical experiments. We report here the first insight obtained from molecular simulations on the conformational transition of the synthetic L1 ligase ribozyme.

We have identified a reduced set of four virtual torsions out of a total of 142 that can be used to distinguish between the active and inactive conformations found in crystal. These virtual torsions span two evolutionarily conserved and restricted regions located in the three-way junction and a U_{38} loop, the conserved residue that is postulated to be responsible for allosteric control of the catalytic step [3, 16, 17, 18, 11] and have been used in the present work as an index to follow the evolution of the conformational rearrangement and monitor important transitions. On the time scale of our simulations we were able to observe approximately two thirds of the complete stem C's 80 Å conformational switch. Since this transition might correspond to the rate-controlling step in L1 ligase catalysis, it is of interest to identify the dynamic hinge points and key interactions that provide stabilization of conformational intermediates that allow the transition to occur. The origin of the large-scale reorientation was traced to variations of a restricted region of the junction. Progress along the transition coordinate was accompanied by the formation of new contacts, not apparent in the crystal structures, between two evolutionarily conserved portions of the L1 ligase: the CG stack of stem B and U_{19} of stem C. These contacts might play an important role in stabilization of intermediate states along the transition pathway.

Simulations were performed from two different initial arrangements of the reactant state differing in the conformation of the GTP_1 triphosphate and its Mg^{2+} coordination. We observed that the non-canonically base-paired ligation site shows a high degree of variability, visiting three distinct conformational states characterized by specific hydrogen bonding patterns. The U_{71}: $O_{3'}$ nucleophile was observed to make direct hydrogen bond interactions with the $O_{2\alpha}$ of GTP_1, in support of the role of this residue as a potential general base. These interactions arise as a direct result of the non-canonically base-paired nature of the ligation site and its intrinsic flexibility.

The present simulation results have identified important dynamical hinge points in the conformational transition from inactive to active forms of the L1 ligase, and characterized interactions that stabilize intermediates along the transition pathway. The insights gained

from this work advance our knowledge of the coupled catalytic/conformational riboswitch mechanism of L1 ligase and may have broader implications for prebiotic RNA enzymes.

1.8 Methods

General system setup. Initial structures used in the simulations were based on a 2.6 Å resolution crystal structure of two co-crystallized L1 ligase ligation product conformers (PDBID:2OIU) [11]. A series of 5 simulations were carried out using as starting structures the two separate conformers (chain P and Q in the PDB file), corresponding to the reputed active state with stem C docked with stem A, and in an alternate presumably inactive un-docked state where stem C is positioned approximately 80 Å away in the crystal.

Since the crystal structure contained the ligated product of L1 ligase, for the reactant state simulations, an additional pyrophosphate moiety not present in the crystal structure was added in the specified conformation (Table 4). The ribozyme was immersed in an orthogonal cell (a=90.0 Å, b=90.0 Å, c=90.0 Å) of pre-equilibrated TIP3P [43] waters so as to ensure a 15 Å buffer between any RNA atom and the boundary of the cell.

The ribozyme molecule was rotated so that the longest molecular axis was aligned with the longest water cell axis and it was centered about the site, and pruned such that any water molecule within 2.8 Å from the solute was removed. All water molecules and Mg^{2+} ions resolved in the crystal structure and within 7 Å of each conformer were kept in the starting structures. The ionic atmosphere consisted of Na^+ and Cl^- ions that were added at random positions at least 5.0 Å away from any solute to neutralize the system and reach the physiologic concentration of 0.14 M.

Simulation protocol. Simulations were performed with the NAMD simulation package (version 2.6) [44] using the all-atom Cornell *et al.* force field (parm99) [45], generated with the AMBER 10 package [46, 47, 48] and TIP3P water model.[43] Periodic boundary conditions were used along with the isothermal-isobaric ensemble (*NPT*) at 1 atm and 300 K using Nosé-Hoover-Langevin pressure piston control [49, 50] with a decay period of 100.0

fs and a damping time scale of 50 fs, and the Langevin thermostat with a damping coefficient of 0.1 ps^{-1}. The smooth particle mesh Ewald (PME) method [51, 52] was employed with a B-spline interpolation order of 6 and the default κ value used in NAMD. The FFT grid points used for the lattice directions were chosen using ~1.0 Å spacing. Non-bonded interactions were treated using an atom-based cutoff of 12.0 Å with switching of non-bond potential beginning at 10.0 Å. Numerical integration was performed using the leap-frog Verlet algorithm with 1 fs time step. [53] Covalent bond lengths involving water hydrogens were constrained using the SHAKE algorithm. [54]

Equilibration procedure. The positions of the solute atoms, including the crystallographic waters and Mg^{2+} ions (if any), were kept fixed, letting the water and counter ion and coion positions to equilibrate for 15 ns, after 5000 steps of initial minimization. The solute atom positions were energy-optimized and then were allowed to move under harmonic restraints over 5 ns. The initial harmonic force constant was 5.0 kcal mol^{-1} and exponentially released over 5 ns. The whole system was energy-optimized and unconstrained dynamics simulation began from 30 K under constant pressure of 1 atm. The temperature was increased to 300 K in steps of 5 K, each step taking 10 ps. The same equilibration process was applied for each simulation. The motions and relaxation of solvent and counterions are notoriously slow to converge in nucleic acid simulations, [55] and careful equilibration is critical. For each simulation, a total of 20 ns of equilibration (15 ns of water/ion relaxation and 5 ns of solvent and structure relaxation) was carried out before production of the trajectory used for analysis.[55]

Analysis. Analysis of the trajectories was carried out using VMD (version 1.8.6). [56] L1 Ligase Stem A (A) was defined to comprise residues 1 to 7 and 45 to 71, stem B (B) residues 8 to 17, and stem C (C) residues 18 to 44 and the junction (J) residues 6 to 10 and 15 to 22 and 39 to 47.

For structural comparison, we do not include the three GAAA tetraloops since they were introduced in L1 ligase structure for crystallization purposes. [11] The notation used

19

to denote base pairing follows that of reference [36]. The definition of virtual torsions used in the present work follows that of reference [28], and are designated as follows: η_n is the virtual torsion angle defined by the atoms C4'$_{n-1}$, P$_n$, C4'$_n$ and P$_{n+1}$; θ_n is the virtual torsion angle defined by the atoms P$_n$, C4'$_n$, P$_{n+1}$ and C4'$_{n+1}$. The coefficient n stands for the position in the sequence (residue number) of the nucleotide that contains that specific atom.

Clustering of structures from the trajectories was performed based on hydrogen bond pattern. For this purpose, it was useful to define a smooth hydrogen bond probability function to use as an index to aid in the definition of the cluster. The cluster probability, $P_i(t)$, that the simulation structure at time t belongs to cluster i is defined as:

$$P_i(t) = \frac{\dfrac{1}{N_i} \sum_{j=1}^{N_i} HBP_{ij}(t)}{\sum_{k=1}^{clusters} \dfrac{1}{N_k} \sum_{j=1}^{N_k} HBP_{kj}(t)} \tag{1}$$

where N_i is the total number of hydrogen bonds associated with cluster i, $HBP_{ij}(t)$ is the probability of formation of the hydrogen bond j in cluster i at time t and is defined by

$$HBP_{ij}(t) = \{1 + exp[c\Delta E_{ij}(t)]\}^{-1} \tag{2}$$

where $\Delta E_{ij}(t) = E[r_{ij}(t), \theta_{ij}(t)] - E[r_0, \theta_0]$, and $E[r, \theta]$ is defined as:

$$E[r, \theta] = -\frac{\cos \theta}{r^2} \tag{3}$$

For hydrogen bonds of the form D-H$\cdots$A where D and A are the hydrogen bond donor and acceptor, respectively, r in the above equation is the H$\cdots$A distance in Å, and θ is the D-H$\cdots$A angle in degrees. The constants r_0, θ_0 and c were chosen empirically to be 3.2 Å, $\pi/3$ radians and 100 Å^2.

Unfolding the docked conformer. To explore the conformational switch pathway in the vicinity of the docked conformation we have conducted a simulation in which we induced the unfolding of the docked product by removing key interactions responsible for docking

stem C into stem A. These interactions include a Mg^{2+} ion that forms an inter-stem coordination, and two hydrogen bonds between the canonical U_{38} and A_{51} base pair. Departing from the docked crystallographic structure, the Prod-D-UF simulation was prepared analogously to the Prod-D simulation except without the inter-stem Mg^{2+} ion. Once solvent was equilibrated, a harmonic force constant of 30.0 kcal/mol/$Å^2$ was gradually applied over 5.0 ns to force the distance between U_{38}:N3 and A_{51}:N1 from 3.5 Å to 8.0 Å in order to gently break the interaction between stems.
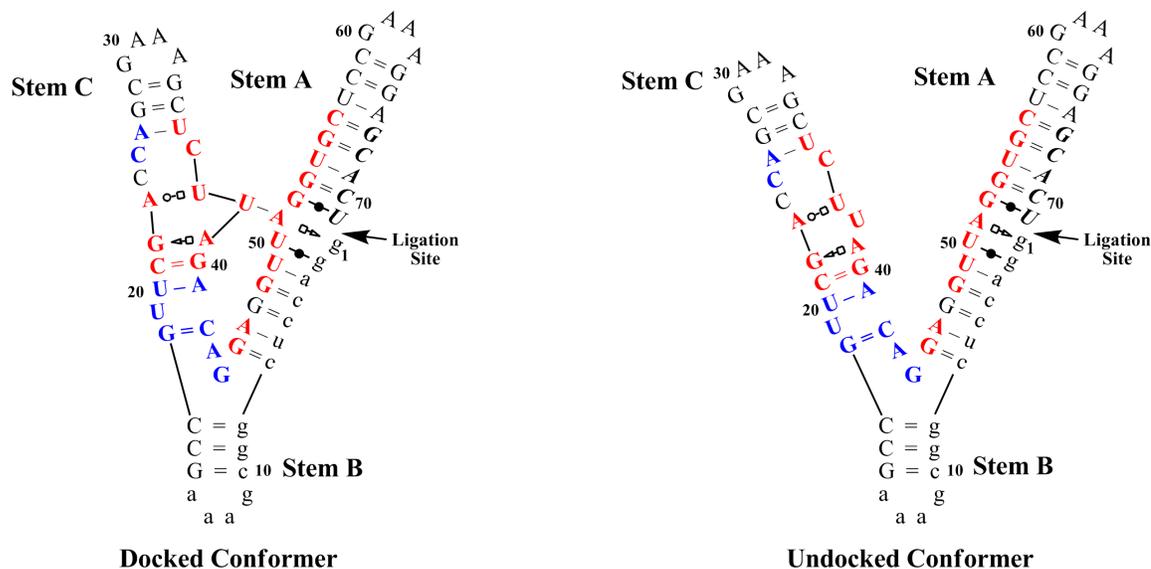
Figure 1: Secondary structure of the L1 ligase ribozyme conformers as found in the crystal structure [11]. With the exception of the interactions responsible for docking stem C into stem A (represented by U_{38}–A_{51} base pair found in the docked conformer) and several hydrogen bonds located mainly in the terminal GAAA tetraloops, the internal base pairing scheme of L1 ligase is conserved between the two conformers (see also Table 2). Notations for base-pairing schemes follow reference [36]. Lower case letters indicate regions that were not targeted for the in vitro evolutionary optimization. Residue in blue were ∼ 80 % conserved within the L1 ligase family, residues in red were invariant. The substrate is italicized.
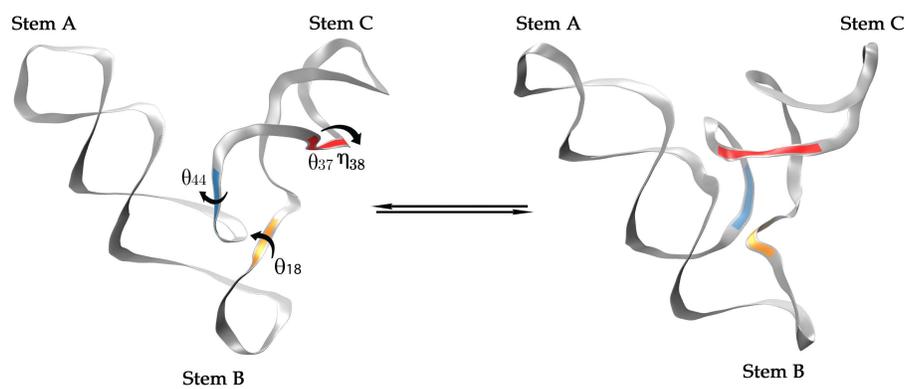
Figure 2: Ribbon representations of the crystallographic presumed inactive undocked (left) and active docked (right) conformations of the L1 ligase.Changes in four virtual torsions (labeled and colored) distinguish between the conformers. The regions that do not change significantly are shown in gray.
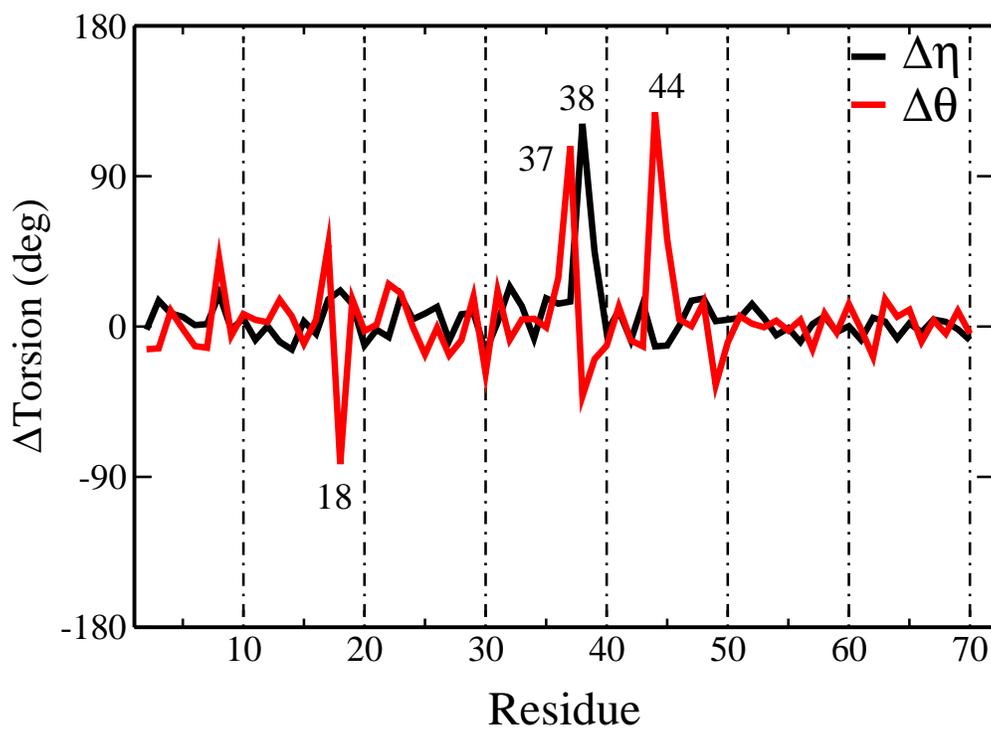
Figure 3: Variations of the virtual torsions ($\Delta Torsion$, where $Torsion$ indicates either the η or θ) between the docked (active) and undocked (inactive) conformers as found in the crystal structure. The $\Delta\eta$ (black) and $\Delta\theta$ (red) are shown. Only four virtual torsions, θ_{19}, θ_{37}, η_{38} and η_{44}, show significant deviations (80.0 deg or more), whereas all other virtual torsions show relatively minor changes (typically less than 25.0 deg).
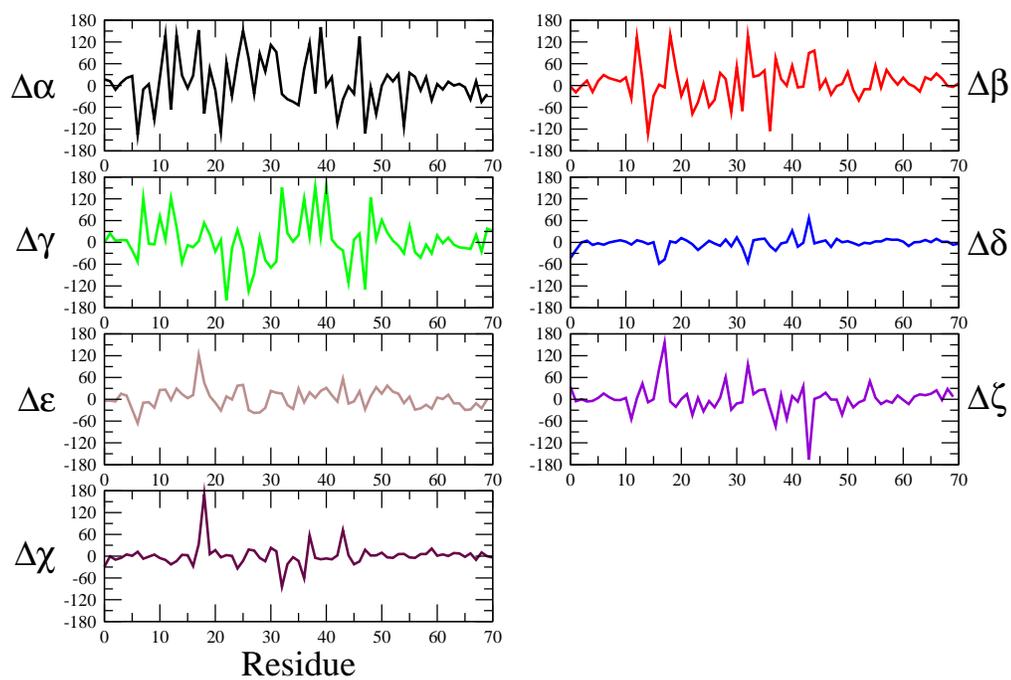
Figure 4: Variations of the real torsions ($\Delta Torsion$, where $Torsion$ can be α, β, γ, ϵ, δ, ζ, χ) between the docked (active) and undocked (inactive) conformers as found in the crystal structure.
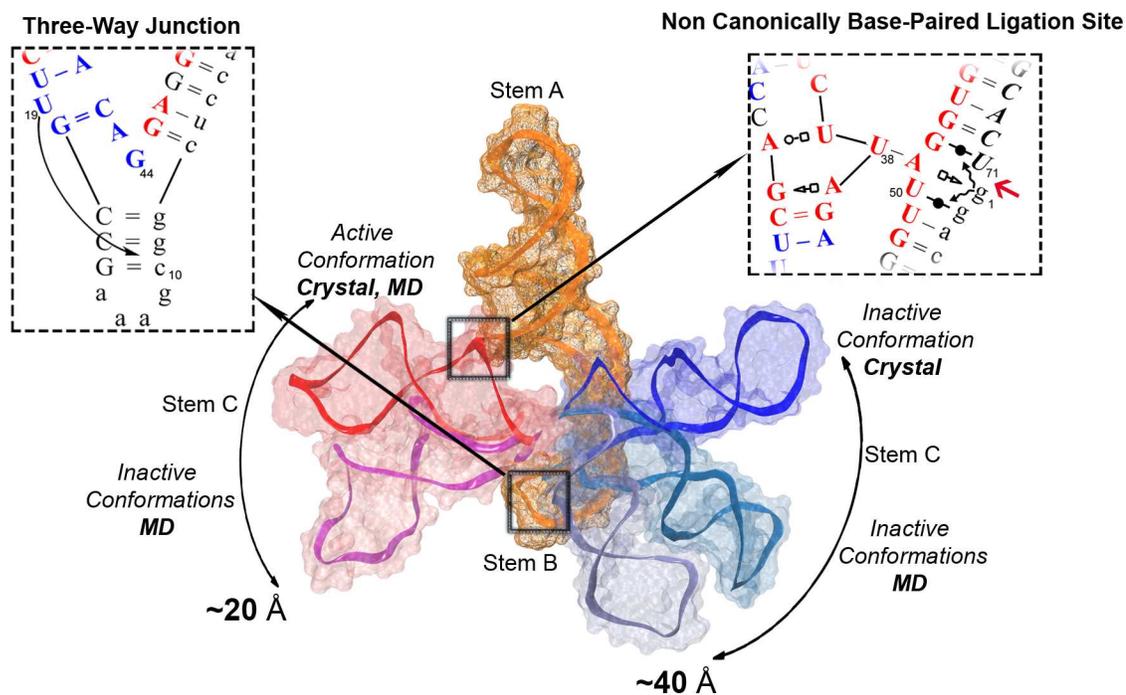
Figure 5: Representative snapshots from the simulations illustrating the coupled on-off conformational switch–catalytic pathway starting from the active and inactive conformations, both in precursor and product states. (middle) Representative snapshots from the simulations are shown with stem A and B (in yellow wire frame surface) aligned and stem C in ribbons with transparent surfaces. (left panel) Contacts important for stabilization of intermediate states that involve the conserved U_{19} and stem B. (right panel) Interaction patterns for three states observed in the simulations between G_1/GTP_1 and the non-canonically base-paired ligation site. The states that do not appear in the crystal structure are indicated with curly arrows.

(a) (b)
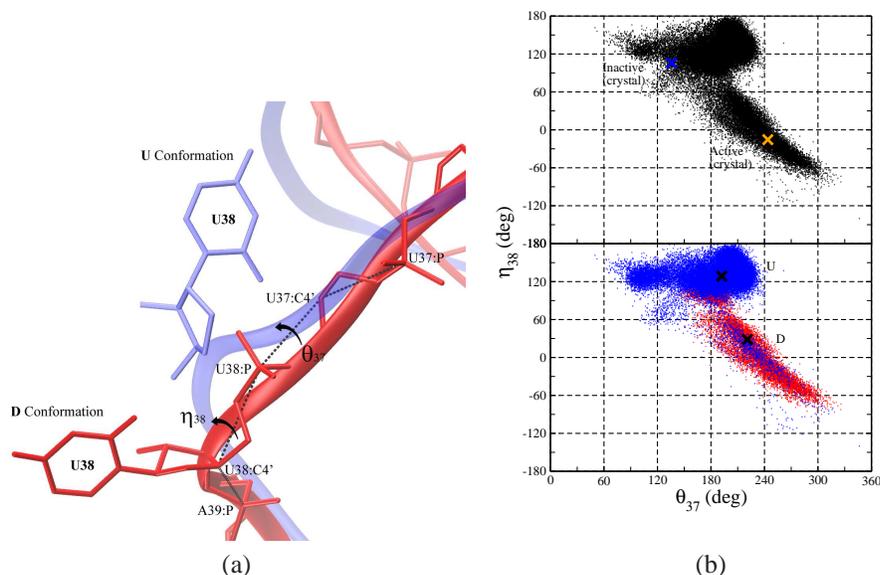
Figure 6: The U_{38} loop is responsible for allosteric control of the catalytic step by transitioning from a "closed" conformation, labeled "U conformation", (specific, as shown by our simulations, to the inactive/undocked conformers) to an "open" conformation, labeled "D conformation" (specific to the active/docked conformations). (a) Representative conformations for the U and D conformations. (b) Upper panel: During Prod-D-UF simulation, where unfolding of the docked state was induced by removal of key inter-stem interactions, the complete transition between the two states was observed by monitoring θ_{37} and η_{38} virtual torsions. The values of the θ_{37} and η_{38} found in the two crystallized conformers are shown with blue and orange crosses. Lower Panel: The θ_{37}/η_{38} space sampled in the Prod-D-UF overlaps with regions sampled during the Prod-D (red) and Prod-U (blue) simulations. The center of each distribution is marked with black crosses. For additional details on the distribution of the two virtual torsions during the Prod-D and Prod-U simulations relative to the Prod-D-UF

Table 1: Hydrogen Bonds in Ligase L1 Crystal Structure Common to the Docked and Undocked Conformers

Donnor	Acceptor	d_P^c	d_Q^c	Location
G_2:N1	U_{50}:O2	2.71	2.76	$G_2 \bullet U_{50}$
U_{50}:N3	G_2:O6	2.89	2.73	$G_2 \bullet U_{50}$
U_{49}:N3	A_3:N1	2.92	2.74	$A_3 - U_{49}$
A_3:N6	U_{49}:O4	3.30	2.81	$A_3 - U_{49}$
G_{48}:N1	C_4:N3	2.87	2.70	$C_4 - G_{48}$
G_{48}:N2	C_4:O2	2.78	2.55	$C_4 - G_{48}$
C_4:N4	G_{48}:O6	2.93	3.08	$C_4 - G_{48}$
G_{47}:N1	C_5:N3	2.92	2.84	$C_5 - G_{47}$
G_{47}:N2	C_5:O2	2.82	2.75	$C_5 - G_{47}$
C_5:N4	G_{47}:O6	2.93	2.89	$C_5 - G_{47}$
A_{46}:N6	U_6:O4	3.09	3.05	$U_6 - A_{46}$
U_6:N3	A_{46}:N1	2.83	2.72	$U_6 - A_{46}$
G_{45}:N1	C_7:N3	2.75	2.89	$C_7 - G_{45}$
G_{45}:N2	C_7:O2	2.72	2.72	$C_7 - G_{45}$
C_7:N4	G_{45}:O6	2.77	2.92	$C_7 - G_{45}$
G_8:N1	C_{17}:N3	2.89	2.66	$G_8 - C_{17}$
G_8:N2	C_{17}:O2	2.69	2.67	$G_8 - C_{17}$
C_{17}:N4	G_8:O6	3.05	2.57	$G_8 - C_{17}$
G_9:N1	C_{16}:N3	2.88	2.87	$G_9 - C_{16}$
G_9:N2	C_{16}:O2	2.85	2.78	$G_9 - C_{16}$
C_{16}:N4	G_9:O6	2.84	2.87	$G_9 - C_{16}$
C_{10}:N4	G_{15}:O6	2.73	2.81	$C_{10} - G_{15}$
G_{15}:N1	C_{10}:N3	2.69	2.75	$C_{10} - G_{15}$
G_{15}:N2	C_{10}:O2	2.54	2.58	$C_{10} - G_{15}$
G_{11}:N2	A_{14}:N7	2.92	2.82	$G_{11} \bullet A_{14}$
C_{42}:N4	G_{18}:O6	2.89	2.97	$G_{18} - C_{42}$
G_{18}:N1	C_{42}:N3	2.82	2.82	$G_{18} - C_{42}$
G_{18}:N2	C_{42}:O2	2.80	2.55	$G_{18} - C_{42}$
U_{20}:N3	A_{41}:N1	3.19	2.66	$U_{20} - A_{41}$
G_{40}:N1	C_{21}:N3	2.98	2.74	$C_{21} - G_{40}$
C_{21}:N4	G_{40}:O6	2.89	2.72	$C_{21} - G_{40}$
G_{40}:N2	C_{21}:O2	3.12	2.74	$C_{21} - G_{40}$
A_{39}:N6	G_{22}:N3	3.16	3.21	$G_{22} \bullet A_{39}$
G_{22}:N2	A_{39}:N7	3.16	2.89	$G_{22} \bullet A_{39}$
A_{23}:N6	U_{37}:O2	3.21	2.77	$A_{23} \bullet U_{37}$
G_{27}:N1	C_{34}:N3	2.89	3.26	$G_{27} - C_{34}$
G_{27}:N2	C_{34}:O2	2.44	3.31	$G_{27} - C_{34}$
C_{34}:N4	G_{27}:O6	3.30	3.31	$G_{27} - C_{34}$
C_{28}:N4	G_{33}:O6	3.03	3.35	$C_{28} - G_{33}$
U_{71}:N3	G_{52}:O6	2.89	2.77	$G_{52} \bullet U_{71}$
G_{52}:N1	U_{71}:O2	2.62	2.65	$G_{52} \bullet U_{71}$
G_{53}:N1	C_{70}:N3	2.86	2.70	$G_{53} - C_{70}$
G_{53}:N2	C_{70}:O2	2.70	2.45	$G_{53} - C_{70}$
C_{70}:N4	G_{53}:O6	3.02	2.64	$G_{53} - C_{70}$
U_{54}:N3	A_{69}:N1	2.87	2.78	$U_{54} - A_{69}$
A_{69}:N6	U_{54}:O4	3.23	2.88	$U_{54} - A_{69}$
G_{55}:N1	C_{68}:N3	2.88	2.90	$G_{55} - C_{68}$
G_{55}:N2	C_{68}:O2	2.61	2.94	$G_{55} - C_{68}$
C_{68}:N4	G_{55}:O6	3.00	2.93	$G_{55} - C_{68}$
C_{56}:N4	G_{67}:O6	3.05	3.06	$C_{56} - G_{67}$
G_{67}:N1	C_{56}:N3	2.96	2.96	$C_{56} - G_{67}$
A_{66}:N6	U_{57}:O4	3.11	3.20	$U_{57} - A_{66}$
C_{58}:N4	G_{65}:O6	2.98	2.90	$C_{58} - G_{65}$
G_{65}:N1	C_{58}:N3	2.86	2.94	$C_{58} - G_{65}$
G_{65}:N2	C_{58}:O2	2.71	3.00	$C_{58} - G_{65}$
C_{59}:N4	G_{64}:O6	2.70	2.83	$C_{59} - G_{64}$
G_{64}:N1	C_{59}:N3	2.84	2.92	$C_{59} - G_{64}$
G_{64}:N2	C_{59}:O2	2.74	2.91	$C_{59} - G_{64}$

Table 2: Hydrogen bonds specific to the docked and undocked crystallographic conformers. The main differences occur at the ends of the stems in the GAAA tetraloops, consistent with conserved internal base pairing within the interior of the stems. Notation for canonical and non-canonical base pairing schemes follow reference [36].

Chain P - Undocked Conformer		Chain Q - Docked Conformer	
HB Donnor–Acceptor	Location	**HB Donnor–Acceptor**	Location
G_1:O2'–A_{51}:N6	$A_{51}\square{\rightarrow}G_1$	G_{11}:N2–A_{14}:O2P	GAAA (stem B)
G_{22}:O2'–A_{39}:N6	$G_{22}{\leftarrow}\square A_{39}$	G_{11}:O2'–A_{13}:N7	GAAA (stem B)
A_{23}:N6–U_{38}:O2'	–	A_{13}:N6–G_{11}:O2'	GAAA (stem B)
G_{33}:N1–C_{28}:N3	$C_{28}{=}G_{33}$	G_{22}:N1–U_{37}:O2'	–
G_{33}:N2–C_{28}:O2	$C_{28}{=}G_{33}$	U_{37}:N3–A_{23}:N7	$A_{23}\square{-}{\circ}U_{38}$
G_{29}:N2–A_{32}:O2P	GAAA (stem C)	U_{38}:N3–A_{51}:N1	$U_{38}{-}A_{51}$
G_{29}:N2–A_{32}:N7	GAAA (stem C)	A_{51}:N6–U_{38}:O2	$U_{38}{-}A_{51}$
A_{31}:N6–G_{29}:O2'	GAAA (stem C)	G_{60}:N2–A_{63}:N7	GAAA (stem C)
G_{60}:O2'–A_{62}:N7	GAAA (stem A)	G_{60}:N2–A_{63}:O2P	GAAA (stem C)
		A_{62}:N6–G_{60}:O2'	GAAA (stem C)
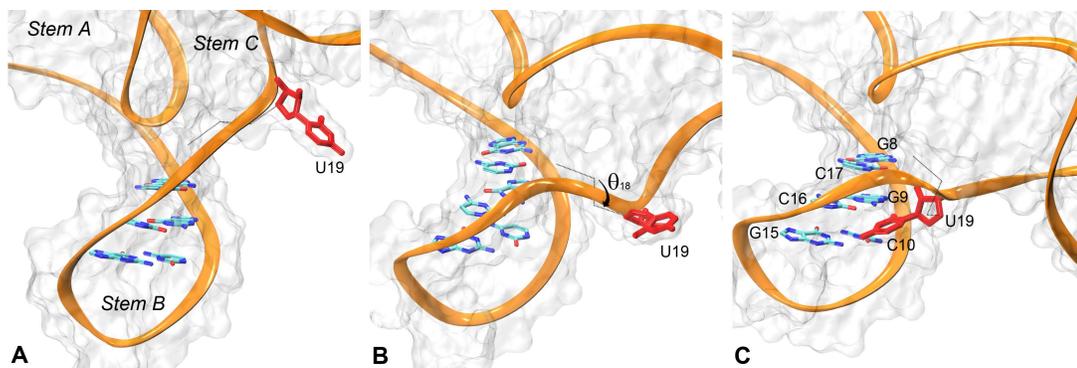
Figure 7: Exploring connections between the flexibility of the three-way junction and the L1 ligase conformational switch. Representative structures for the conformational transition along θ_{18} toward the active conformation that brings U_{19} in the close proximity of stem B. (A) Structure is similar to the undocked crystal conformation ($\theta_{18} \approx$ -110 deg). (B) Structure is the transition structure between (A) and (C) ($\theta_{18} \approx$ 170 deg). (C) Structure shows a set of specific tertiary interactions between the conserved U_{19} and stem B ($\theta_{18} \approx$ 90 deg).
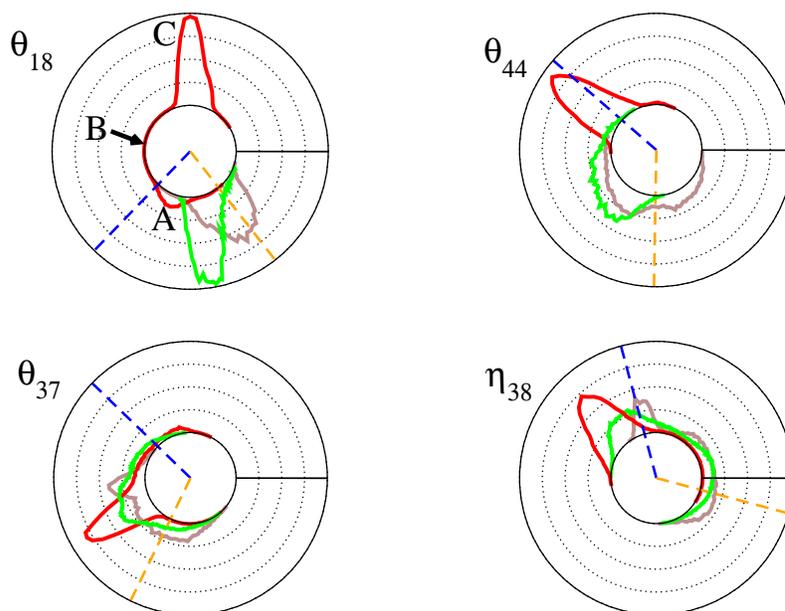
Figure 8: Exploring the L1 ligase flexibility in the space of the four virtual torsions angles for the Prod-D (black), Prod-U (red) and Prod-D-UF (green) simulations. The values of the the virtual torsions in the two crystal conformers are indicated with vertical dashed lines: orange and blue for the docked and undocked conformers, respectively. (Upper Left) As suggested by the structures of the two crystallized conformers θ_{18} has to travel an arc of length 277.6 degrees. Moving along the complimentary shorter arc is prevented by steric clashes between stem A and C. This is confirmed by the Prod-U simulation in red that shows θ_{18} sampling space along the longer arc towards to docked structure. (Upper right) θ_{44} spans two distinct regions in the Prod-D and Prod-U simulations. Similarly with θ_{18}, θ_{44} must travel on an arc of 231.8 degrees in order for the docked $\rightleftharpoons$ undocked transformation to take place. (Lower Left and Right) θ_{37} and η_{38} can be used follow the interchange between the two states of the U_{38} loop, as suggested by the crystal structures and confirmed by the simulations. If during the Prod-D and Prod-U the two torsions cover two distinct regions, during the Prod-D-UF simulation the interchange between the two states is observed. For a detailed depiction of the two states on the correlation plot of the two virtual along with their structures see Fig. 6
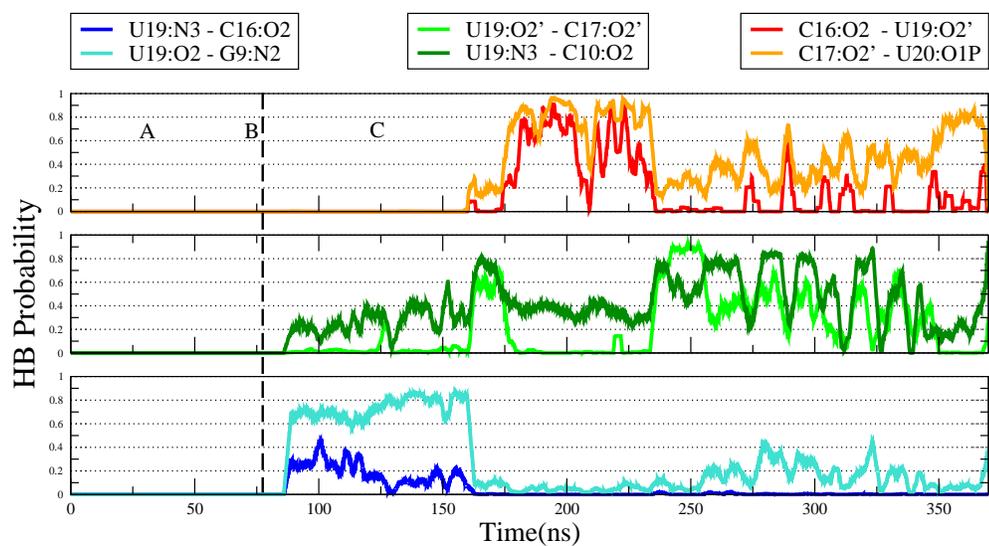
Figure 9: Time series of specific hydrogen bond probabilities between U_{19} and stem B. The dashed line represents the moment at which the backbone undergoes a transition along the θ_{18} virtual torsion angle passing from structure A to structure C illustrated in Fig. 7.
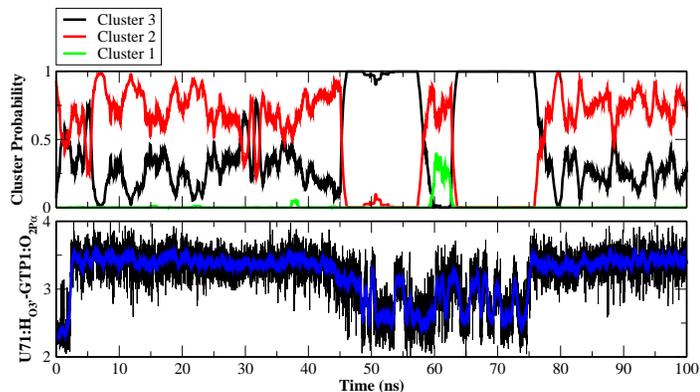
(a)
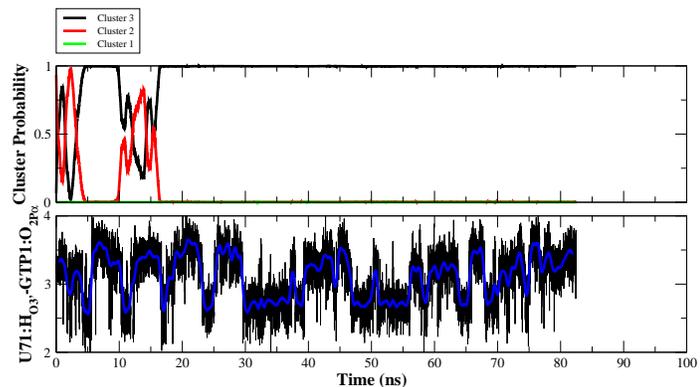
(b)

Figure 10: The non-canonically base-paired ligation site exhibits a high degree of conformational variability, passing through a series of three states (clusters 1, 2 and 3) characterized by specific hydrogen bond patterns between GTP_1/G_1 and the ligation site. The 3D arrangement of the ligation site in the case of the Prec-MgTP simulations is shown in Fig. 11. Here, the time evolution of probabilities of the three cluster states, and their correlation with the formation of specific $U71:H_{O3'} - GTP1:O_{2P\alpha}$ interactions, are shown for the Prec-D-MgTP (a) and Prec-D-XTP (b) simulations.
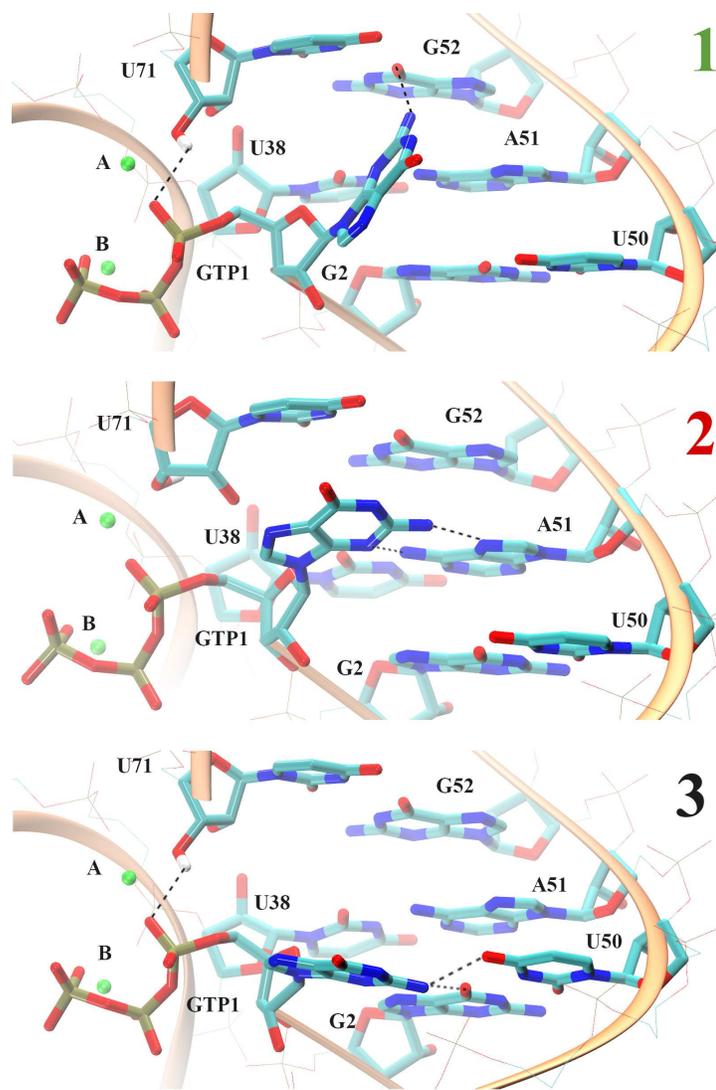
Figure 11: The non-canonically base-paired ligation site exhibits a high degree of conformational variability, passing through a series of three states (1, 2 and 3) characterized by specific hydrogen bond patterns between GTP_1/G_1 and the ligation site. Here, the 3D arrangement of the ligation site in the case of the Prec-MgTP simulations is shown. An identic set of states is identified also in the product state during the Prec-D-XTP and Prod-D simulations. The correlation between the U71:$H_{O3'}$–GTP1:$O_{2P\alpha}$ distance can be seen in Fig. 11

Table 3: Comparison of the sugar pucker amplitudes (P) and phase angles (T_m) for the two crystallized conformations of L1L. D stands for the docked conformer and U for the undocked conformer.

Res.	$P^D(T_m^D)$	$P^U(T_m^U)$	Pucker D	Pucker U	Res.	$P^D(T_m^D)$	$P^U(T_m^U)$	Pucker D	Pucker U
G_1	10.7(44.0)	147.2(31.5)	**C3'-endo**	**C2'-endo**	U_{37}	-139.8(22.7)	158.1 (39.7)	**C4'-endo**	**C2'-endo**
G_2	15.4(46.6)	-9.6(37.6)	**C3'-endo**	**C2'-exo**	U_{38}	126.4 (25.4)	147.5 (40.2)	**C1'-exo**	**C2'-endo**
A_3	15.2(45.3)	3.9(45.4)	C3'-endo	C3'-endo	A_{39}	6.3 (42.0)	8.9 (37.1)	C3'-endo	C3'-endo
C_4	27.3(49.2)	17.2(51.7)	C3'-endo	C3'-endo	G_{40}	13.7 (45.0)	18.6 (41.8)	C3'-endo	C3'-endo
C_5	15.3(40.7)	4.6(42.8)	C3'-endo	C3'-endo	A_{41}	-34.6 (37.9)	33.9 (50.7)	C2'-exo	C2'-exo
U_6	27.9(47.6)	24.2(46.3)	C3'-endo	C3'-endo	C_{42}	27.4 (48.7)	15.5 (36.5)	C3'-endo	C3'-endo
C_7	23.9(47.3)	25.1(44.2)	C3'-endo	C3'-endo	A_{43}	25.5 (42.1)	22.2 (41.7)	C3'-endo	C3'-endo
G_8	11.4(43.7)	13.2(45.8)	C3'-endo	C3'-endo	G_{44}	162 (37.8)	81.4 (55.1)	**C2'-endo**	**O4'-endo**
G_9	10.8(39.8)	10.9(43.4)	C3'-endo	C3'-endo	G_{45}	3.2 (45.1)	-13.2 (36.7)	**C3'-endo**	**C2'-exo**
C_{10}	17.6(41.8)	16.4(45.5)	C3'-endo	C3'-endo	A_{46}	14.4 (49.3)	4.6 (47.2)	C3'-endo	C3'-endo
G_{11}	19.1(47.3)	22.8(46.3)	C3'-endo	C3'-endo	G_{47}	11 (49.5)	10.4 (53.3)	C3'-endo	C3'-endo
A_{12}	18.6(38.7)	10.8(42.6)	C3'-endo	C3'-endo	G_{48}	9.9 (45.7)	-20.7 (41.4)	**C3'-endo**	**C2'-exo**
A_{13}	9.8 (23.2)	14.8(41.7)	C3'-endo	C3'-endo	U_{49}	-7.2 (39.4)	27.2 (46.2)	**C2'-exo**	**C3'-endo**
A_{14}	3.1 (40.8)	-0.1(49.3)	**C3'-endo**	**C2'-exo**	U_{50}	0.4 (45.0)	22.5 (45.4)	C3'-endo	C3'-endo
G_{15}	2.5 (43.6)	-1.8(35.3)	**C3'-endo**	**C2'-exo**	A_{51}	-20 (41.4)	7.9 (36)	**C2'-exo**	**C3'-endo**
C_{16}	16.7(41.7)	24.2(47.7)	C3'-endo	C3'-endo	G_{52}	5.5 (46.1)	8.1 (45.7)	C3'-endo	C3'-endo
C_{17}	26.3(40.5)	157.1(42.8)	**C3'-endo**	**C2'-endo**	G_{53}	14.7 (43.3)	14.4 (43.5)	C3'-endo	C3'-endo
G_{18}	66.7(18.4)	159.5(38.2)	**C4'-exo**	**C2'-endo**	U_{54}	16.6 (47.5)	24.1 (49)	C3'-endo	C3'-endo
U_{19}	93.9(40.7)	27.6(33.8)	**O4'-endo**	**C3'-endo**	G_{55}	6.7 (43.7)	6.1 (43)	C3'-endo	C3'-endo
U_{20}	7.4(40.2)	8.9(42.6)	C3'-endo	C3'-endo	C_{56}	5 (43.9)	24.6 (46.5)	C3'-endo	C3'-endo
C_{21}	0.5(43.9)	28.8(51.5)	C3'-endo	C3'-endo	U_{57}	20.7 (42.3)	6.7 (44.2)	C3'-endo	C3'-endo
G_{22}	6.4(42.1)	28.7(40.4)	C3'-endo	C3'-endo	C_{58}	18.2 (42.9)	3.9 (45.6)	C3'-endo	C3'-endo
A_{23}	0.6(47.1)	3.8(44.0)	C3'-endo	C3'-endo	C_{59}	10.3 (48.8)	12.8 (44.5)	C3'-endo	C3'-endo
C_{24}	33.2(53.8)	17.2(36.8)	C3'-endo	C3'-endo	G_{60}	17.1 (41.2)	15.9 (48.3)	C3'-endo	C3'-endo
C_{25}	18.0(46.6)	2.6(44.0)	C3'-endo	C3'-endo	A_{61}	7.5 (35.0)	9.1 (30.8)	C3'-endo	C3'-endo
A_{26}	10.7(45.2)	12.7(45.4)	C3'-endo	C3'-endo	A_{62}	17.8 (44.4)	13.4 (36.1)	C3'-endo	C3'-endo
G_{27}	7.2(39.6)	-10.3(43.2)	**C3'-endo**	**C2'-exo**	A_{63}	17.4 (41.1)	-1.5 (40)	**C3'-endo**	**C2'-exo**
C_{28}	14.6(50.7)	-9.2(47.9)	**C3'-endo**	**C2'-exo**	G_{64}	3.1 (41.2)	19.6 (46.4)	C3'-endo	C3'-endo
G_{29}	13.7(48.4)	3.5(53.2)	C3'-endo	C3'-endo	G_{65}	12.7 (43.8)	12.4 (46.2)	C3'-endo	C3'-endo
A_{30}	26.0(34.2)	-1.7(33.0)	**C3'-endo**	**C2'-exo**	A_{66}	14.7 (48.1)	14.2 (43.9)	C3'-endo	C3'-endo
A_{31}	9.5 (32.5)	23.1(47.1)	C3'-endo	C3'-endo	G_{67}	-0.9 (43.0)	6.8 (41.4)	**C2'-exo**	**C3'-endo**
A_{32}	21.3(40.1)	0.8(33.2)	C3'-endo	C3'-endo	C_{68}	13.3 (49.5)	11.3 (40.6)	C3'-endo	C3'-endo
G_{33}	-19.2(36.4)	166.0(53.0)	C2'-exo	C2'-endo	A_{69}	10 (43.4)	9.4 (40.9)	C3'-endo	C3'-endo
C_{34}	5.0(46.6)	1.6(47.2)	C3'-endo	C3'-endo	C_{70}	12.6 (43.1)	21.7 (43.5)	C3'-endo	C3'-endo
U_{35}	-1.2(42.9)	-1.9(46.9)	C2'-exo	C2'-exo	U_{71}	1.9 (39.5)	7.9 (43.1)	C3'-endo	C3'-endo
C_{36}	11.9(39.1)	18.3(48.7)	C3'-endo	C3'-endo					

Table 4: Summary of simulations presented in the current work.

Label	Sequence/Structure	Ions	Length
Prod-D	Chain Q (PDBID:2oiu)	6 Mg^{2+} from crystal structure	100 ns
Prod-U	Chain P (PDBID:2oiu)	no Mg^{2+}	370 ns
Prec-D-XTP	Chain Q (PDBID:2oiu) precursor, TP in extended conformation	6 Mg^{2+} from crystal structure	80 ns
Prec-D-MgTP	Chain Q (PDBID:2oiu) precursor, TP in two-Mg^{2+}-bound conformation as in DNA polymerase β structure (PDBID:1BPY)	7 Mg^{2+}, 6 from crystal structure	100 ns
Prod-D-UF	Chain P (PDBID:2oiu) forced unfolding	5 Mg^{2+} from crystal, deleted 1 Mg^{2+} to induce unfolding	60 ns

Table 5: RMSD(FS,MS) values comparing crystallographic and simulation structures. The RMSD(FS,MS) is defined by superimposing two structures so as to minimize the RMSD based on a "fitting set" of atoms (FS), and then measuring the average RMSD of a "measuring set" of atoms (MS). In this way, one can use the RMSD(FS,MS) to quantify the internal structural deviations of each of the L1 ligase stems A, B and C and junction J, and also assess the relative orientation of stems. For example, if the RMSD(AB,AB) and RMSD(C,C) is small, but RMSD(AB,C) is large, this indicates that the internal structure of all the stems, as well and the relative orientation of stems A and B, are generally unchanged, whereas there is a large re-orientation of stem C relative to stems A and B. The reference coordinates used to superimpose the simulation structures were the crystallographically-derived structures also used as the starting points of the simulations (see METHODS). Standard deviations are shown in parentheses. The RMSD(FS,MS) for the crystallographic docked and undocked conformers are also compared (Crystal). The FS and MS sets were defined as follows: (A) stem A, residues 1-7 and 45-71; (B) stem B, residues 8-17; (C) stem C, residues 18-44; and (J) the junction, residues 6-10, 15-22 and 39-47. Note that the definition of stems A, B and C involve atoms sets that do not intersect, and the union of these sets constitute the entire L1 ligase. The definition of the junction, on the other hand, is a subset of atoms that intersects with parts of all of the stems. Hence the set "ABC" implies the entire L1 ligase. The average values and standard deviations from the simulations were calculated from the entire trajectory, excluding the first 10 ns, with data collected every ps. All values in the table are in Å.

Simulation (MS,FS)	Prod-D	Prec-D-MgTP	Prec-D-XTP	Prod-D-UF	Prod-U	Crystal (ref. U)
(A,A)	3.0(0.7)	2.7(0.6)	2.9(0.6)	3.3(0.6)	2.5(0.6)	1.7
(B,B)	1.8(0.6)	1.2(0.3)	1.2(0.2)	1.6(0.4)	3.3(0.6)	0.7
(C,C)	2.3(0.4)	2.6(0.3)	2.5(0.4)	3.2(0.5)	2.9(0.4)	2.1
(AB,C)	7.2(1.6)	4.9(1.3)	5.8(1.6)	13.5(4.0)	19.1(4.8)	48.4
(AC,B)	5.4(1.9)	4.2(1.3)	6.2(2.2)	5.5(1.6)	11.3(2.4)	8.5
(BC,A)	8.2(1.9)	5.0(1.5)	4.9(1.4)	16.7(5.9)	19.3(6.3)	44.0
(AB,AB)	3.7(0.8)	2.8(0.6)	3.1(0.7)	3.5(0.7)	3.4(0.6)	4.9
(AC,AC)	4.1(0.7)	3.1(0.5)	3.4(0.6)	7.0(1.9)	7.6(1.8)	15.1
(BC,BC)	3.0(0.6)	3.0(0.5)	3.7(0.8)	3.8(0.7)	5.8(1.0)	7.8
(ABC,ABC)	4.2(0.7)	3.1(0.5)	3.5(0.6)	6.7(1.8)	8.0(1.7)	14.3
(J,J)	2.1(0.3)	2.0(0.3)	2.3(0.3)	2.3(0.4)	4.1(0.5)	8.3

Table 6: Important contacts in L1 ligase in the non-canonically base paired ligation site: precursor and product simulations. Simulations derived from the docked conformer show a higher degree of variability by populating 3 distinct states than those derived from the undocked conformer that exhibit a unimodal distribution. These three states correspond to the three clusters showed in Fig. 11 and are characterized by specific hydrogen bonding interactions. Contacts that during a specific simulation yield a hydrogen bond are labeled with a $\star$. Interatomic distance averages and standard deviations (in parentheses) are listed in Å.

Contact/(Cluster)	Prec-D-MgTP	Prec-D-XTP	Prod-D	Prod-U
GTP_1/G_1:N2–G_{52}:O6 (1)	5.5 (1.9) $\star$	6.3 (0.9)	6.0 (1.1)$\star$	4.0 (0.5)
GTP_1/G_1:N3–A_{51}:N6 (2)	5.4 (0.8) $\star$	4.1 (1.1) $\star$	6.2 (1.2)$\star$	3.1 (0.2)$\star$
GTP_1/G_1:N2–A_{51}:N7 (2)	3.8 (1.3) $\star$	5.6 (1.0) $\star$	5.2 (1.3)$\star$	3.0 (0.1)$\star$
GTP_1/G_1:O2'–A_{51}:N6(2)	6.3 (0.9) $\star$	7.4 (0.8) $\star$	7.4 (1.1)$\star$	3.0 (0.2)$\star$
GTP_1/G_1:N2–G_2:O6 (3)	3.4 (0.8) $\star$	3.0 (0.3) $\star$	3.3 (0.9)$\star$	3.6 (0.5)
GTP_1/G_1:N2–U_{50}:O4 (3)	3.7 (1.0) $\star$	3.5 (1.1) $\star$	3.7 (1.5)$\star$	5.2 (1.0)

CHAPTER II

MAPPING L1 LIGASE CONFORMATIONAL SWITCH MECHANISM

L1 Ligase (L1L) is a synthetic ribozyme ligase that function as a molecular switch. In Chapter 1 we have identified the dynamical hinge points from molecular dynamics and crystal structure analysis. The presented simulations were not able to cover the entire conformational space that is spanned during the activation of L1L. Here, we present a enhanced sampling technique that uses concepts from network theory to focus the sampling towards unexplored regions of the activation process of L1L. We find that this process comprises three states and discuss the impact of conserved nucleotides to support the transformation between those.

2.1   Introduction

In this chapter we use molecular dynamics simulations to map the conformational switch mechanism that supports L1L catalysis by employing an enhanced sampling technique that makes use of network theory concepts.

We build our results around the findings of Chapter 1 and point the reader there for an extensive discussion of L1L significance and for an in depth analysis that leads to the identification of the L1L hinge points that support its switching mechanism. In short, L1L is a molecular switch and the first synthetic ligase ribozyme that has been crystallized[11]. L1L is known to catalyze the 5' to 3' ligation regiospecifically and regioselectively making use of an non-canonically base-paired ligation site[3, 16, 17, 18]. Its crystal structure revealed two largely different conformers that were associated with the active and inactive states based on some specific tertiary contacts between conserved regions and so sketched

up some possible routes of L1L allosteric control[11].

A limited set of four virtual torsions was identified to differentiate between the two crystallized conformers. Further, this set of virtual torsion constituted a starting basis for the identification of the hinge points that might support the hypothesized large conformational switch of L1L. The hinge points were located on the conserved L1L three–way junction and the U_{38} loop. Moreover, the non-canonically base-paired active site was shown to possess different dynamical patterns in the active and inactive states based on a hydrogen bonding interacting patterns analysis. The active state was observed to be able to cover three distinct sub-states that have different capacities to initiate the catalytic process, whereas the inactive state dynamics was unimodal [57] and Chapter 1.

The molecular dynamics simulations presented in Chapter 1 were not able to cover the entire conformational space along which the switching mechanism of L1L may take place. Even the fact that the hinge points are rather localized in restricted regions of the L1L structure was not able to overcome the complexity of the transformation and of the relationship between the sampled structures. Here we use network theory to devise an enhanced sampling technique that aims to focus the sampling towards regions of the conformational space that might reside on the L1L switching pathway and that are under-sampled.

The rest of the chapter is structured as follows: in Section 2.2 we present the design of the enhanced sampling technique, in Section 2.3 we characterize the overall switching pathway and find that L1L follows a three states two steps mechanism, in Section 2.4 and Section 2.5 we discuss the structural details and the implications of the two steps of the switching mechanism.

2.2 Using network theory to focus the sampling towards the unexplored regions of the conformational space

We have observed in Chapter 1 that very long molecular dynamics simulations are insufficient to sample the whole conformational space covered during the active $\rightleftharpoons$ inactive

conformational switch even with unfolding/undocking the L1L active conformer. The presented set of long MD simulations were able to cover almost two thirds of the overall proposed activation mechanism of L1L that consits of a swing of more than 80 Å of stem C tip, followed by the docking of stem C's U_{38} loop into the active site. More precisely, the issue was that the trajectories that originated from either the undocked crystallized conformer (labeled Prod-U in Chapter 1) or the unfolded docked conformer (labeled Prod-D-UF in Chapter 1) oscillate with relative high amplitudes in the proximity of their starting points without advancing towards the other attractor/attractors. In order to enhance the exploration of the structural details of the active $\rightleftharpoons$ inactive conformational switch of L1L we employ here a methodology that focuses the sampling on under-sampled regions of the conformational space that might be crossed to complete the L1L switching.

We call the method FSN, an abbreviation from **F**ocused **S**ampling on **N**etworks. The method creates a network to represent the relationships/connectivity between the sampled structures and aims to guide the sampling, using unbiased molecular dynamics simulations, towards regions of the conformational space that are under-sampled and can be reside on the most probable transition routes that connect all the nodes of the network.

When applied to large scale conformational changes, network theory [58, 59, 60, 61, 62, 63] has the advantage of representing the relationships between intrinsically multidimensional sampled structures as a 2D or 3D graph based on some local proximity-measure and hence is not limited by the projection of the entire dynamics on few overall geometric order parameters . The possible order parameters may include RMSD from a reference structure, internal coordinates like distances, angles, torsions, number of native contacts, etc. In general, the creation of a conformational space network, a special case of a network where the connectivity between the nodes is based on interchangeability along molecular dynamics trajectories, is an *a posteriori* analysis procedure. [58, 59, 60, 61, 62, 63]

The use of unbiased simulations for extensive sampling has the advantage of not considering incompletely relaxed or unrealistically perturbed structures in the final structural

analysis of the simulations. Also, the use of biasing potentials implies the *a priori* definition and identification of a set of global order parameters or reaction coordinates along which the structures are dragged. This has a high probability to limit or bias the identification of the complete set of order parameters that define the dynamics of the system. [64, 65, 66, 67, 68, 69] These methods are efficient in terms of readily obtaining possible paths but it has been argued that for example, in the case of using targeted molecular dynamics - a method that uses the RMSD as a guiding order parameter - the resulting structures are perturbed towards unrealistic and statistically irrelevant parts of the conformational space[62]. On the other hand, the absence of a guiding biasing potential requires more sampling as it is statistically less probable to cross free energy barriers. To overcome this problem, at least partially, we propose to guide our sampling by launching *swarms of trajectories* [70, 71, 72] from points that are identified as probable crossing points from the network. *Swarms of trajectories* have been used before in conjunction with other path sampling methods, such as those included under the name of *string methods*. [70, 73, 74, 75, 76, 77, 78, 71, 79, 80] These iterative methods represent the transition of interest as a walk along a chain of states that is parametrized by a curvilinear pathway. The pathway is optimized by initiating a set of molecular dynamics trajectories (i.e. swarms of trajectories) from uniformly distributed points along the path, under the constraint that the newly generated path connects the predefined end points. These end points usually coincide with two stable/metastable states that the path search has to connect. These approaches are attractive for solving the sampling issues enumerated here, however they are limited to optimizing one single pathway whereas, given the size of our system, we have to consider a multitude of pathways.

The **FSN** method is designed as an iterative procedure. Each iteration the following operations are accomplished (see also Fig. 14):

- building the network from the existing sampled structures. The nodes of the network represent each of the clusters (typically, on the order of a hundred) obtained

from a partitional incremental k-way clustering procedure (see Section 2.7) of the entire set of sampled structures (tipically, on the order of one milion). The structures are represented through the characteristic vector of stem C (see Section 2.7). The connectivity of the network is realized based on the proximity measured by a cosine-similarity function of the structures such that each node is connected to its closest eight neighbors.

- Identification of the under-sampled nodes of the network. The sampling density is estimated by determining the number of sampled structures that lie within a threshold value from the centroid of each cluster.

- Identification of the major possible traversal points of the network. For each node a *traversal count* is computed by counting how many times the node is traversed by all the shortest paths between all the possible pairs of nodes (see Section 2.7 for details).

- Identification of the swarms of trajectories launching points. Five nodes of the network are chosen to be the swarms of trajectories launching points based on their lower ranking in the sampling density and higher ranking in the traversal counts.

The procedure is repeated untill sufficient sampling is achieved [*].

2.3 Identification of the main attractors along the L1L switching pathway

Identification of the main free energy attractors along the L1L switching pathway gives the possibility to characterize (with a certain precision dictated by the amount of sampling and the size of the system) the overall shape of the free energy surface that is spanned during this process as well as to identify the main routes that connect these attractors. The

---

[*] In Fig. 16 we show how focused swarms of trajectories can be used to map a conformational transition on (Φ, Ψ) map of the solvated alanine dipeptide.

strategy that is used here for the identification of the main attractors involves partitioning the sampled structures using a two-step process. First, we partition the data set using clustering techniques based on *structural proximity*. Second, the determined clusters are further grouped based on their *interchangeability along molecular dynamics trajectories*.

Clustering is a machine learning technique [81, 82, 83] that can segment a given data set into subsets or classes based on some similarity–dissimilarity measure. Given that we can define such a measure, we can use different clustering strategies to group the large amount of structural data that we obtained from the current simulations. Ones of the major problems of traditional clustering algorithms is that the number of clusters that the data set has to be partitioned into has to be predetermined and the fact that the overall objective similarity/dissimilarity functions will increase/decrease monotonically with the number of clusters. This means that without any other constrain clustering would not be able to group the sampled data into a reasonable number of chemically relevant states: on the contrary, it would prefer (in terms of the value of its objective function) to partition the data set in as much as possible number of clusters. To overcome this arbitrarily of the choice of the number of clusters we have devised a strategy that builds on the particularities of the L1L dynamics. Specifically, we will use the fact that the docked/active conformation is the most rigid and less flexible conformational state that can be considered as indivisible in any number of substates with respect to the large fluctuations of stem C that are necessary to support the switching mechanism. So we will choose as the optimum number of clusters the one at which all the structures spanned during the docked/active simulations reside mainly in a single cluster.

For the clustering stage a hybrid partitional-agglomerative clustering algorithm is employed, that is described in detail in the Methods section. Fig. 13a shows the hierarchical tree/dendrogram whose leaves are a initial set of intermediary clusters obtained from an 100-way incremental partitional algorithm. The clusters that are spanned/sampled during the docked product simulations are colored in blue. Fig. 13b assimilates the same set of

intermediary clusters assimilated with the nodes of a (in this case, complex) conformational space network, in which the linkages are created between clusters interchange along molecular dynamics simulations. Fig. 17a shows a less complex hierarchy that was obtained during the agglomerative stage of clustering that was stopped when all the states spanned by the docked/active conformer simulation were grouped/agglomerated in a single cluster. In the same figure it can be observed that the agglomeration process would keep the integrity of this cluster for 10 more iterations. Moreover, the other 11 clusters show a complex structural similarity relationships to each other.

A conformational space network is built with these 12 clusters (Fig. 17b) and assuming that the groups of nodes that interconnect more than with the other groups can be considered as forming an attractor (fact pointed out in [60]) it can be observed that two more major states individualize, colored in yellow and orange. In fact, the component nodes of each of these two groups show also a strong structural similarity with each other. Indeed, as shown by the structural–similarity–based tree the merging of the "yellow" and "orange" states occurs at the last stage of the agglomerative clustering procedure.

We will denote these three major attractors as A ("Active", colored in blue in Fig. 17b), I_1 ("Inactive Intermediate 1", yellow) and I_2 ("Inactive Intermediate 2", orange). A closer analysis of these 3 main states shows that they mainly originate from the docked/active conformation (state A), undocked/unfolded conformation (state I_1) and crystallize inactive conformation (state I_2) and that the L1L switching follows the following 3-states/2-steps mechanism: A $\rightleftharpoons$ I_1 $\rightleftharpoons$ I_2 whose details will be discussed in the following sections.

2.4 A$\rightleftharpoons$$I_1$: insights into the allosteric control mechanism

The A to I_1 transitions corresponds to the docking/undocking of stem C, through U_{38} loop, from the L1L non-canonically base paired active site. From both biochemical and crystallographic data [3, 16, 17, 18, 11] it has been advanced the hypothesis that this stem-loop interaction is one of the ingredients of the way L1L ligase catalytic activity is controlled.

Indeed, the activation process of L1L begins with binding a DNA effector, followed by the binding of the substrate and then by the formation of the aforementioned stem-loop contact supported by an 80 Å swing of stem C. [3, 16, 17, 18, 11]

The question that immediately follows is what is the role of the tertiary interaction that supposedly activates the L1L? Otherwise said, how does the realization of the stem-loop interaction contribute to the allosteric control of L1L catalytic activity? In Chapter 1 we have made a first step towards answering this question and described how the dynamics of the active site shows two completely different variability patterns in the active and inactive forms. More precisely, the catalytic site was shown to be very flexible, spanning three conformational states in its active form, whereas the dynamics of the inactive form was unimodal.

We hypothesized in Chapter 1 that the different variability pattern are caused by the lack of a hydrogen bond that is typical to a trans Hoogsten/sugar face base pair formed by A_{51} and G_1/GTP_1 (denoted here as $A_{51} \square\!\!\rightarrow G_1$ [36]). We have to note that A_{51} and G_1 are the ones that directly intermediate the interaction of the catalytic site with stem C, through a canonical base pair and some phosphate–Mg^{2+} interactions, respectively.

Here, we investigate the effect of slowly inducing the undocking U_{38} loop from the active site, as described in Section 2.7. Fig. 15a compares the distributions of the three hydrogen bonds that define a standard $A_{51} \square\!\!\rightarrow G_1$ base pair in the case of the docked/active and undocked/inactive conformations as well as their evolution after the removal of the tertiary contact between stem A and stem C. It can be easily observed that the distributions of these three hydrogen bonds reflect the aforementioned different dynamical patterns of the two active and inactive states characterized in Chapter 1 . If in the case of the docked/active the distributions are multi-modal, suggesting the existence of several conformational states, in the case of the inactive/undocked state the distributions are unimodal. The evolution in time of these three hydrogen bonding distances recorded after the disruption of the tertiary contacts between stem A and stem C shows that after almost 50 ns the active site adopts

the inactive/undocked dynamical pattern, with the formation of the complete typical trans Hoogsten/sugar face base pair between A_{51} and G_1 depicted in Fig. 15b C . Prior to this moment the active site oscillates between two states that can undoubtedly be identified based on their hydrogen binding patterns as conformational clusters 2 and 3 of the active site in the precursor state (compare Fig. 15b A and B and Fig. 11).

The analysis presented here suggests there is a direct and clear link between the specific dynamical pattern of the active site and the tertiary contacts with stem C. The formation of these contacts induces/allows the active site to adopt three main conformational states that have been shown to correlate differently with initiation of the first steps of the catalytic process, namely the deprotonation of $U_{71}:O_{3'}$ (see Chapter 1) .

Interestingly, the two states that positively correlate with the $U_{71}:O_{3'}$ deprotonation require the disruption of the incomplete trans Hoogsten/sugar face base pair between A_{51} and G_1 that is found in the docked crystal conformation and also as one of the states of the active site in our simulations.

This clearly suggests the role of the tertiary contacts between the active site and stem C: to weaken the the $A_{51} \square\!\!\rightarrow G_1$ base pair so that it can be easily disrupted, to promote the first step of the catalytic process, i. e. the deprotonation of $U_{71}:O_{3'}$.

2.5 $I_1 \rightleftharpoons I_2$: the mechanism of stem C 80 Å swing

$I_2 \rightleftharpoons I_1$ conformational transition corresponds to the swing of almost 80 Å of stem C from the vicinity of the inactive conformation isolated in the crystal structure that is brings U_{38} loop in close proximity of the active site. This mechanism was suggested by structures of the two crystallized conformers [11] and which supporting hinge points were identified using virtual torsions analysis in conjunction with large scale molecular dynamics simulations to reside in restricted regions of the L1L three-way junction. The simulations that sampled the I_1 and I_2 states were characterized in Chapter 1, under the names Prod-D-UF and Prod-U, respectively. Here we focus our attention on what are the structural factors that

support the transition and integrate these findings within the existing experimental data.

The location of the interface between the two states is marked as a green node on the conformational space network represented in Fig. 17c. Some sample structures representative for this node are shown in Fig. 18b. Given the complexity of L1L crystal structure, there are, without any doubt, many order parameters that can be used to characterize the structures at this point. Here, we look at the correlation between θ_{44} values and the relative orientation of L1L stem C with respect to stem A, i.e. the orientation of the stem C characteristic vector. Otherwise said, we analyze the distribution of θ_{44} as function of the distance (or depth) from the $I_1 - I_2$ interface node. The regions of the L1L structure that are spanned by the virtual torsions θ_{44} and θ_{18} were identified as ones of the main L1L hinge points (see Chapter 1). θ_{18} region was signaled to have a large flexibility, spanning at least four states(see Fig. 8), however not necessarily correlated with the stem C orientation (data not shown). Previous simulation have shown that θ_{44} region was also flexible (see Fig. 8)., but were not able to sample a part of the torsion space that was hypothesized as necessary for the stem C swing to take place.

In Fig. 19a the distribution of the θ_{44} virtual torsion angle is shown as function of the depth from the $I_1 - I_2$ interface node of the conformational space network. The *depth 1* selection contains the interface node direct neighbors and the *depth 2* selection contains the direct neighbors as well the second layer of neighbors of the interface node. It can be observed that, as expected, the number of samples in increasing from the *depth 1* to *depth 2* selections, however the region of the torsion space located approximately between -10 and 0 degrees remains approximately with the same number of samples. This indicates that θ_{44} traverses this specific interval only when the structures are located at the interface node and its immediate neighbors and not when walking further away. Otherwise said, in order to transition from the I_1 to I_2, θ_{44} has to pass through this interval. This confirms our hypothesis drawn from the analysis of the crystallized conformers of L1L that stated that θ_{44} has to travel along an arc of 231.8 degrees that includes the -10 to 0 degrees interval.

In Fig. 19a we show some representative conformations of the part of the three–way junction that is spanned by θ_{44} along the stem C swing. It can be observed the G_{44} and A_{43} undergo a significant repositioning with respect to A_{46} and G_{45} that are the constituents of the highly canonical and, consequently, relatively rigid stem A. We have to point out that G_{44} and A_{43} are part of a conserved motif that is high sensitiveness to mutations located on the three way junction[3, 16, 17, 18].

2.6   Conclusion

L1 Ligase allosterically controlled catalytic mechanism is a complex process that requires the participation of several flexible constitutive structural motifs. In fact, each step of the catalytic process of the specific L1L construct studied in this work requires a conformational transition event. The complexity of such processes range from that of a stem swing for 80 Å, loop closing/opening up to changes in base pairing pattern. This wide variety of complexity cannot be captured by dragging the structure along some predefined simple order parameters, but instead, as we propose here, can be sampled efficiently using networks theory and unbiased molecular dynamics trajectories.

This work brings a deeper insight into the structural factors that sustain the allosteric control of L1L catalysis. First, we use a novel clustering strategy in conjunction with conformational space analysis to show that the switching mechanism of L1L comprises three successive major states. Second, the docking of stem C, through the U_{38} loop, in the active site has the role of disrupting the A_{51} □→ G_1 base pair to allow the active site to adopt two conformations that have a high probability to initiate the first step of the catalytic process. Third, highly conserved region of the L1L three way junction that is spanned by the θ_{44} virtual torsion can be shown to directly correlate with the stem C swing from the I_2 to I_1 states.

This work is a second major step towards understanding the L1L function and lays the ground for a quantitative treatment of the effect of specific mutations and for future

contributions towards tuning the L1L function.

2.7   Methods

Description of simulations. The molecular dynamics simulations set-up is described in detail in Section 1.8. With the exeption of the precursors simulations and unfolding of the docked conformer, all the data presented in Chapter 1 has been used for the current analysis. All the new simulations presented in this chapter use a periodic box of rhombo-dodecahedral simmetry with and edge legth of approximately 120 Å, making the number of atoms conatained in the system to be approximately 104,000.

  Swarms of trajectories. Every itteration step of the FSN procedure, a set of 5 launching points of the swarms of trajectories are chosen. From each of the launching points a molecular dynamics trajectory is started using the coordinates and velocities saved at that specific point during the previous FSN itterations.

  Order parameters. In this work the dynamics of L1L ribozyme is described in terms of relative position of its stems. This appears to be a reasonable choice given the conservation of the three-dimesional structure of each individual stem. We have signalled in Chapter 1 that the major hinge points are located in the restricted region of the junction. For each one of the three stems we can define a characteristic vector v_{char}^{stem} in the following mannor: 1) we choose two ajacent canonical base pairs (i and j) and for each we define a vector that is determined by the positions of the C1' atoms (v_i, v_j) 2) the characteristic vector is defined as the cross product of the two vectors: $v_{char}^{stem} = v_i \times v_j$. Prior to the calculation of this vectors the all structures are rms-fitted against the structure of stem A, as found in the undocked crystallized conformer. Across this work we make use of the characteristic vector of stem C (see Fig. 12). The two base-pairs that were chosen to define v_{char}^{C} were U_{20}-A_{41} and C_{21}=G_{40}.

  Unfolding/undocking of stem C from the active site in the docked conformation simulation The unfolding simulation of the docked conformer was prepared as the one described

in Chapter 1, with the exeption that the biasing potential that was used to disrupt the canonical base pair between U_{38} and A_{51} was a step like potential of 5 Kcal/mole that was turned on gradually from 3.5 Å to 5.0 Å during a 15 ns simulation period. During this period the base pairing of the active site was restrained using harmonic potentials (spritong constants were 5 Kcal/molÅ^2) centered at the equilibrium values specific to the hydrogen binding pattern found in crystal.

Clustering. All clustering calculations were realized with CLUTO [84] software. The k-way incremental partitional clustering scheme used here is realized using k-1 repeated bisections, followed by a global optimization of the solution. The algorithm maximizes the cosine similarity function between each data point and the centroid of the cluster that is assigned to. The clustering procedure is repeated 20 times to avoid local minimums of the overall similarity function.[85] During agglomerative clustering each of the objects are initially assigned to their own cluster and then pairs of clusters are repeatedly merged based on the same similarity function until all the initial objects are grouped in one cluster. A hybrid scheme that in the first stage clusters the data using a partitional scheme and in the second uses an agglomerative scheme is called an constrained agglomerative algorithm and have been shown to efficiently lead to better solutions than partitional or agglomerative schemes alone. [86]

Networks manipulation, rendering and analysis was done with NAViGaTOR [87, 88] and yED graph editor (yFiles software, Tubingen, Germany). Shortest path searches between all the nodes of the network as well as the determination of the traversal counts were done using a Floyd-Warshall alogorithm [89, 90] as implemented in NAViGaTOR.
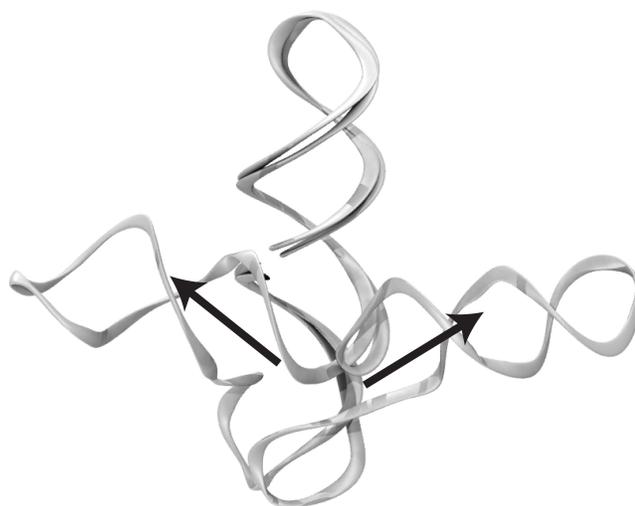
Figure 12: The main order parameter that is used in Chapter 2 to monitor L1L dynamics is the characteristic vector of stem C (see Section 2.7), which aims to capture the relative motion of L1L stem C with respect to stem A. The vector is depicted here with a black arrow for both docked and undocked crystallized conformers. The angle formed by these two instances of the vector is approximately 165 degrees.
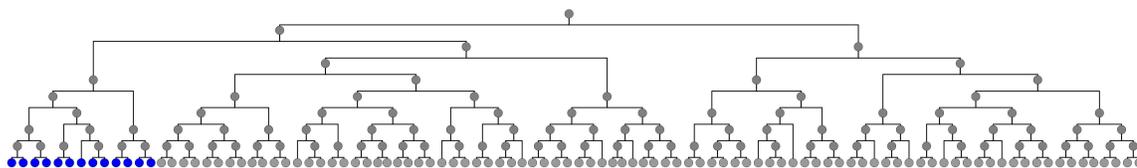
(a)
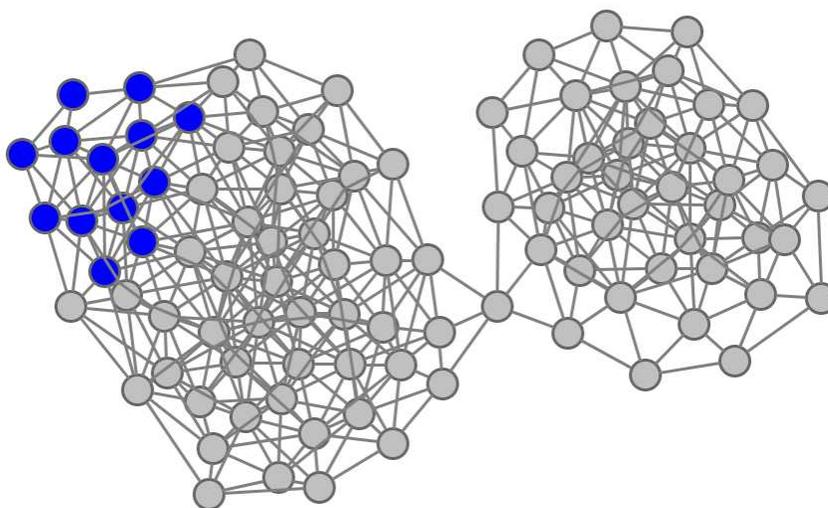
(b)

Figure 13: (a) The structures sampled in this work can be organized in complex structural–similarity–based hierarchy that is obtained by employing a two stage clustering technique described in Section 2.7. We use this hierarchy to identify major states along the L1L switching pathway by imposing the external restraint that the clustering should be stopped at the point when all the docked conformations sampled with molecular dynamics (shown in blue) are grouped within a single cluster (see Fig. 17). (b) The conformational space network obtained from using the leaves of the hierarchical tree as nodes where the edges are constructed based on interchangeability of clusters during molecular dynamics trajectories obtained in this work.

(a)   (b)

(c)   (d)

(e)   (f)
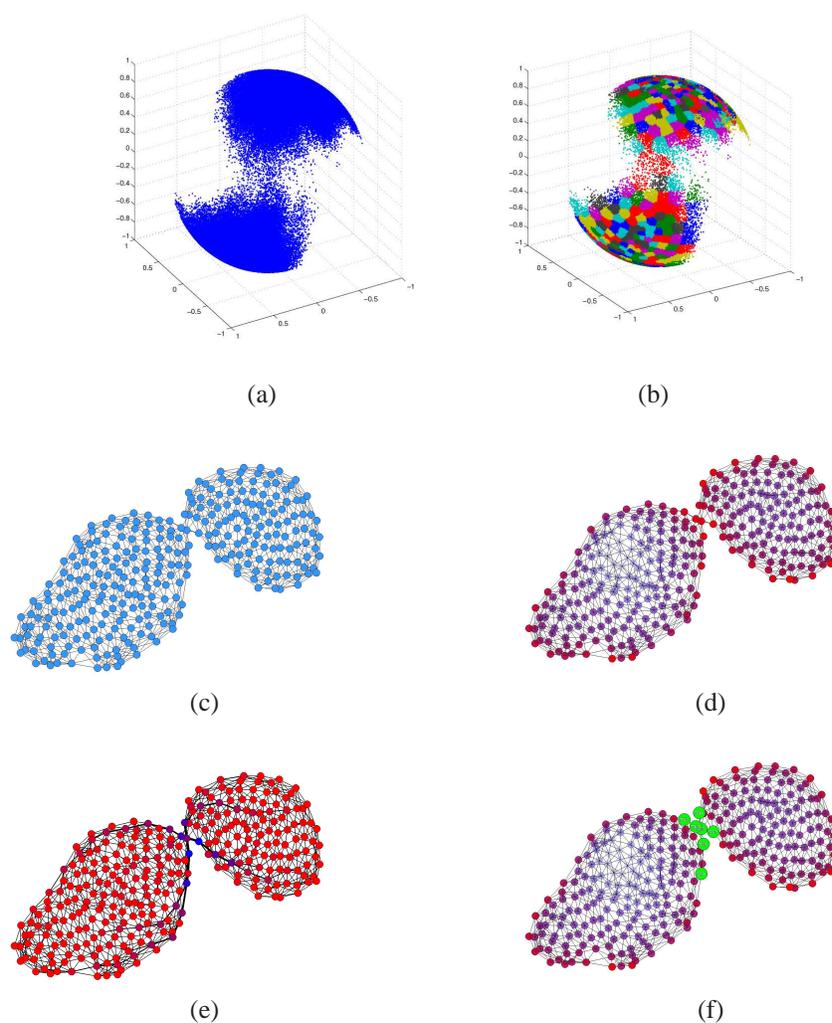
Figure 14: To focus the sampling around relatively unexplored/under–sampled regions of the conformational space that might reside along the L1L switching pathway we designed an iterative strategy (FSN – Focused Sampling on Networks) that is described in detail in Section 2.2. The projection of the structures on a set of order parameters (see (a)) is clustered using an incremental partitional 300-way clustering technique. In (b) the data points belonging to the same cluster have the same color. (c) A network is built for which each one of the previously determined clusters is assimilated with a node. Each node being connected to its closest 8 neighbors using a cosine similarity/distance-based function. (d) For each of the clusters a sampling density is calculated that is displayed here using red-blue colormap, the red corresponding to low densities whereas the blue corresponds to the high densities. (e) Each of the network nodes is associated a traversal count, i.e. number of times it is crossed by all the shortest paths along the graph between all the nodes of the graph. A similar red-blue colormap is used here, with red corresponding to lower values of the traversal counts and blue corresponding to a higher value of the same quantity. (f) The swarm of trajectories launching points (here colored in green) are chosen as the nodes that have a low sampling density and have a high traversal counts.

(a)

(b)
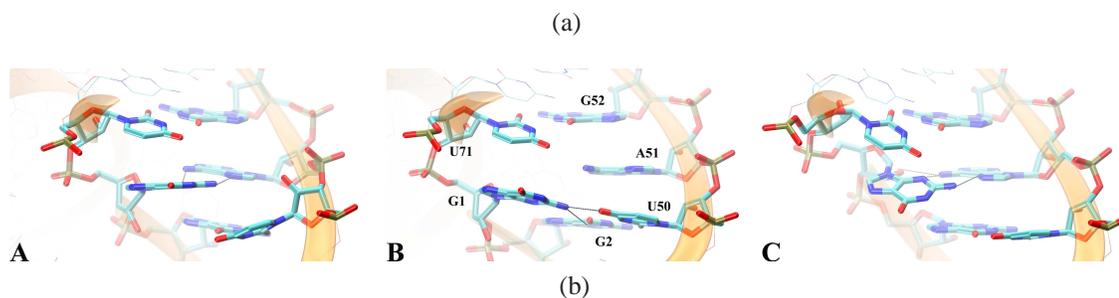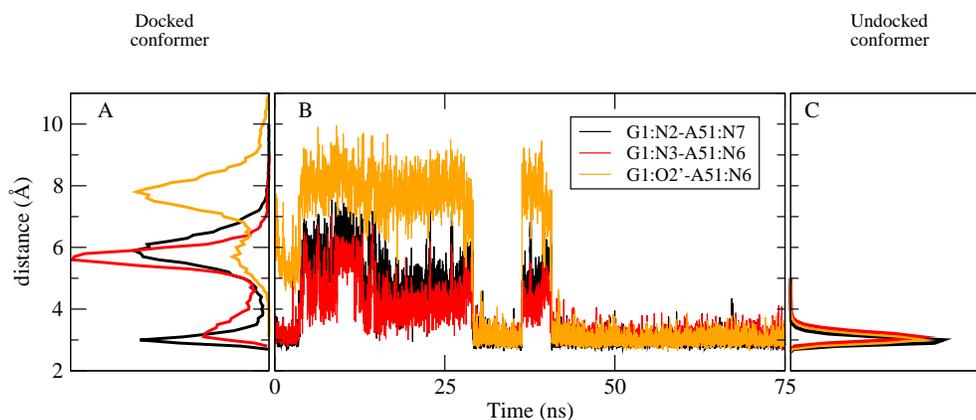
Figure 15: After removing the distant contacts between U_{38} and the active site, the latter recovers the dynamical signature of the inactive/undocked conformer. (a) The change in the dynamical signature is followed in terms of the time series evolution of three important hydrogen bonding distances that are specific to complete $A_{51} \square\mapsto G_1$ (trans-Hoogsten/sugar) edge base pair. On the left the multi-modal distribution of these three hydrogen bonding distances specific to the active/docked conformation simulations both in the product as well as in the precursor state. On the right, the same distributions, that are now unimodal in nature, obtained during the molecular dynamics simulations of the inactive/undocked conformer simulations. In the middle the time series evolution of the hydrogen bonding distances after the unfolding/undocking biasing potential has been removed. (b) Three representative snapshots along the unfolding/undocking trajectory. A and B are representative for the active conformation and are the two major conformations identified also in the active precursor simulation (see Chapter 1 and Fig. 11). C is representative for the undocked/inactive conformation, yielding a complete $A_{51} \square\mapsto G_1$ base pair.
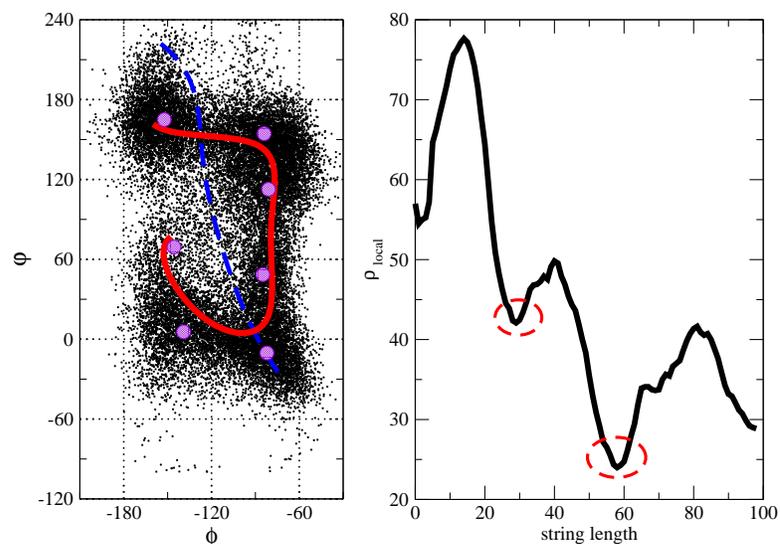
Figure 16: The conformational transition in the [-180;0] × [-120;240] interval of (Φ, Ψ) map of the solvated alanine dipeptide is mapped using the string method with swarms of trajectories. [70, 73, 74, 75, 76, 77, 78, 71, 79, 80] For comparison see the potential of mean force for this interval in [91]. As opposed to the original implementation where the swarms of trajectories were launched from uniformly distributed points along the string, here the swarms launching points are located around the local minima of the samples density along the string, thus focusing the sampling towards the under-sampled regions of the conformational space. Each iteration, a string is generated using a smoothing spline from a set of points that represent the centroids of the clusters obtained by applying a partitional clustering scheme to the existing sampled data. Based on the samples density along the string, 10 swarms launching points are identified. 10 NVT-ensemble molecular dynamics trajectories are initiated in which the Φ and Ψ angles are restrained with a harmonic potential around the identified swarms launching points. In a subsequent step, the restraints are removed and the 10 simulations are continued and the projection of the sampled structures on the (Φ, Ψ) map is added to the existing data. (Left) The samples collected during the string calculation are depicted with black dots. The centroids of the clusters are depicted with violet filled circles. The dashed blue line is the initial guess for the string, whereas the red curve represents the final string. Note that, in this implementation, there is no need to specify the location of the string end points. (Right) The local sample density (ρ_{local}) along the string. The local minima of ρ_{local} are encircled. The string length is parametrized from 0 to 100.
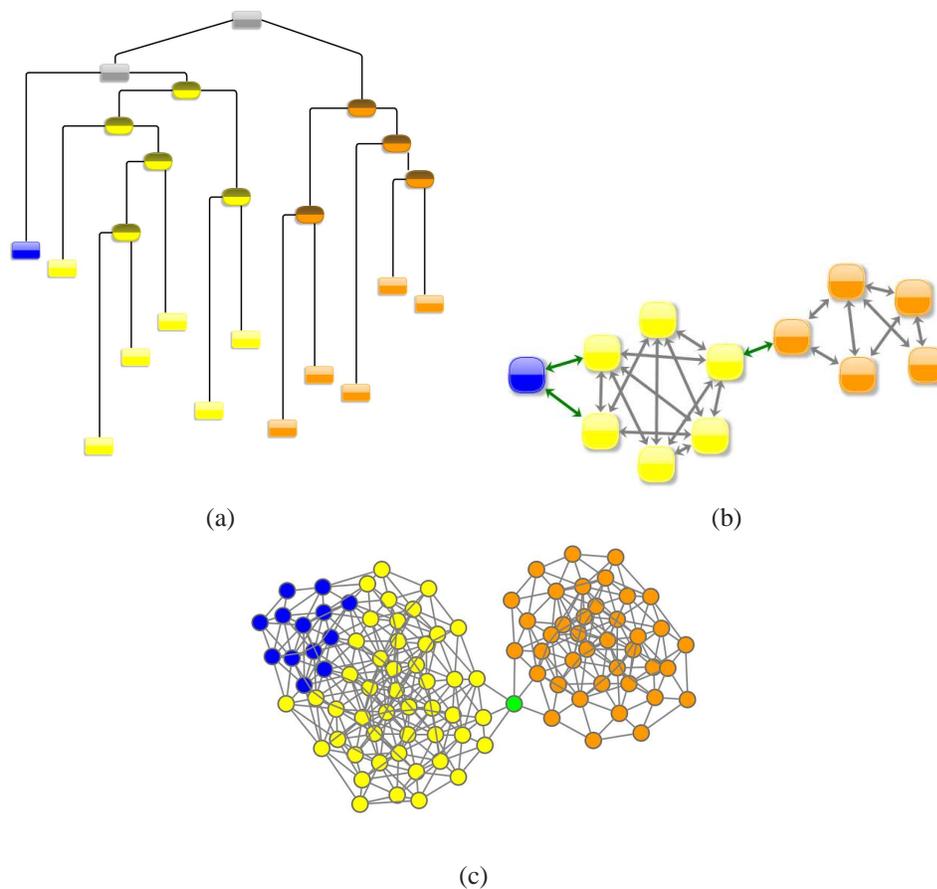
Figure 17: (a) Structural similarity based hierarchical tree/dendrogram obtained by agglomerative clustering from the complete hierarchy presented in 13b. The agglomeration process is stopped when all the initial clusters spanned by the docked/active conformation are merged/agglomerated into one single cluster, colored here in blue. The nodes of the hierarchical tree are ordered on the vertical axis in order of their discovery during the agglomeration process. (The lower the node resides the earlier it is discovered.) (b) A conformational space network obtained by considering the 12 leaves of the hierarchical tree as the nodes and linking the nodes that interchange during molecular dynamics simulations. Based on the connectivity pattern three major conformational states can be identified/isolated. The edges/links of the graph that support the transitions between these three states are colored in green. (c) The conformational space network formed by the initial 100 clusters with nodes colored based on their inclusion into each of the three identified major conformational states.

(a)

(b)
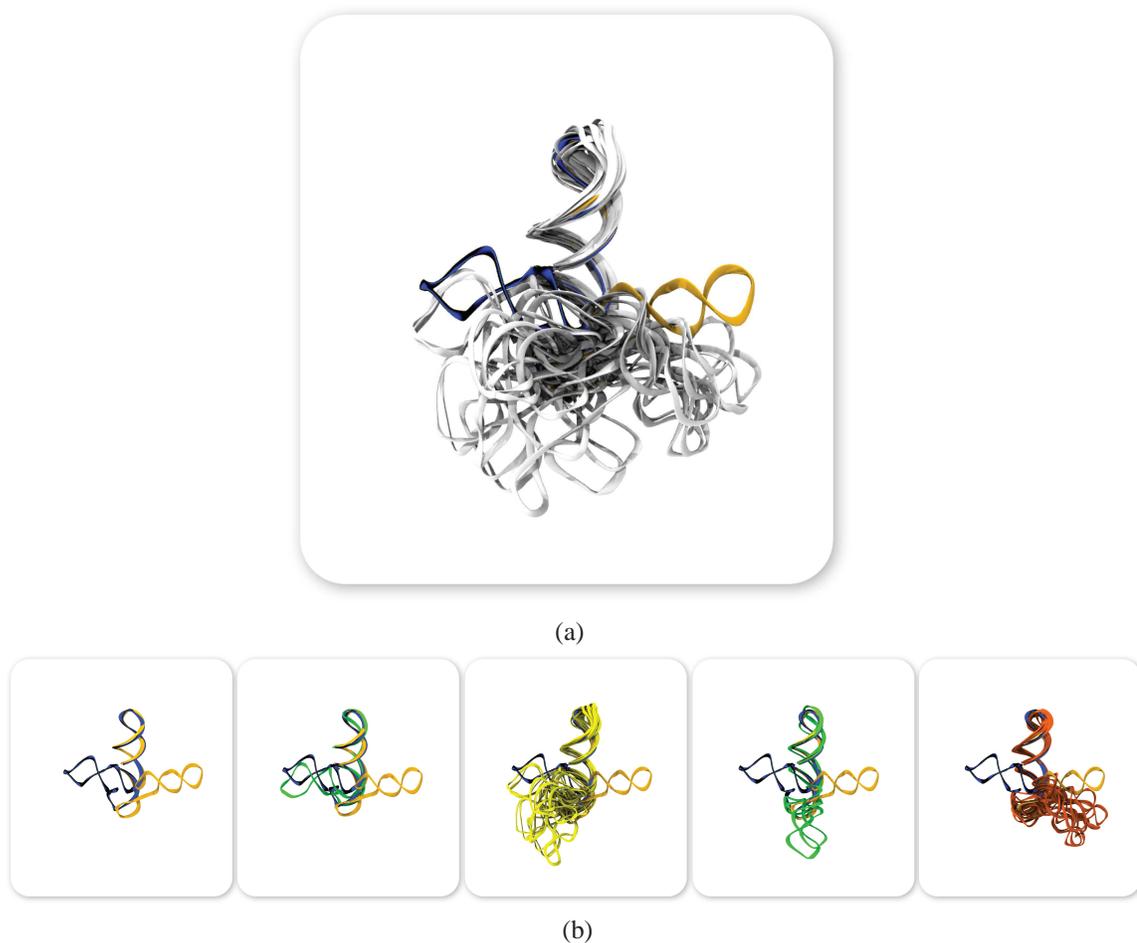
Figure 18: The clustering strategy allows the classification of the structures obtained from molecular dynamics simulations, represented here by the centroids obtained from a 20-way partitional clustering of the sampled structures that are colored in gray in (a) along with the crystallized docked conformer colored in blue and the crystallized undocked conformer colored in orange. As shown in Section 2.3 three major conformational states are identified. In (b) the the structures that belong to each of the three states are colored with blue (A), yellow (I_1) and orange (I_2) respectively. In green some representative structures located at the boundary/interface between A and I_1 and I_1 and I_2 are depicted.

(a)

A   B   C
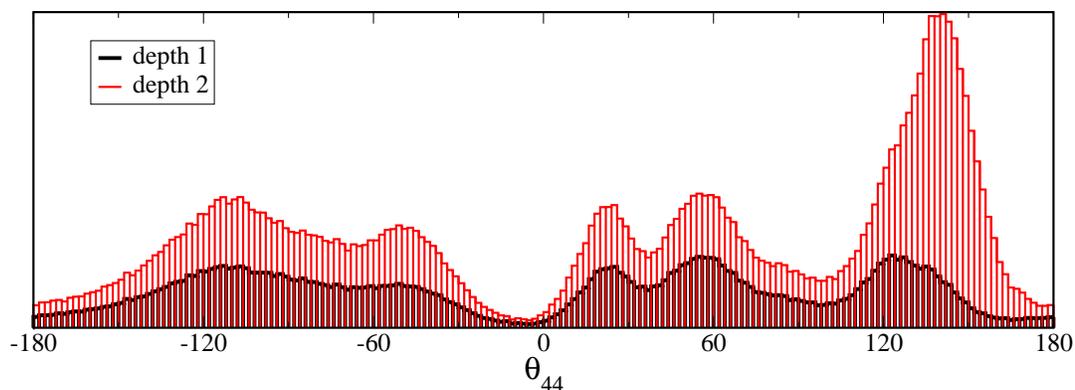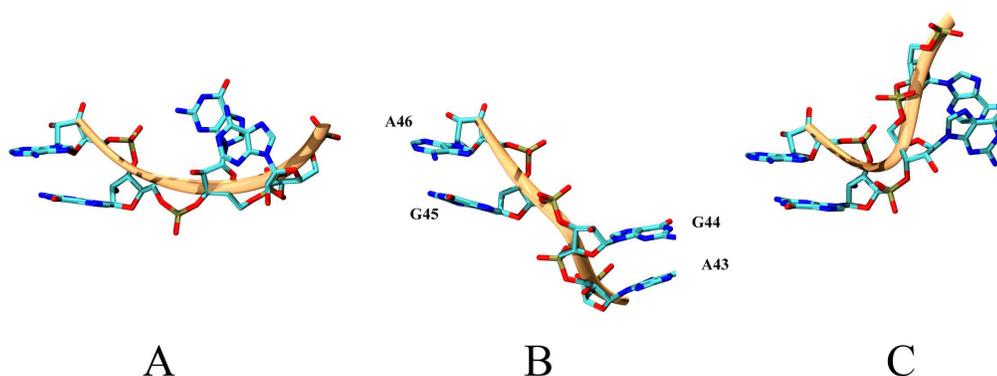
(b)

Figure 19: The unnormalized distribution of θ_{44} in the vicinity of the the $I_1 - I_2$ interface node (marked with green in Fig. 17c) shows that θ_{44} has to traverse the -10 to 0 degree interval to allow the transition from I_1 to I_2 to take place. (a) The distributions of θ_{44} for two selection of nodes of the conformational space network are shown: in black the *depth 1* selection, that includes the interface node as well as its neighbors and in red the *depth 2* selection that includes the *depth 1* selection as well as the second layer of neighbors of the interface node. The distribution of θ_{44} in the vicinity of the docked and undocked crystal structures is presented in Fig. 8. (b) Representative conformations of the region spanned by θ_{44} of the three-way junction along the conformational switch pathway. A: conformation specific to the vicinity of the active conformation, B: conformation located on the -10:0 interval of θ_{44}, C: conformation specific to the vicinity of the crystallized active conformation.

CHAPTER III

THE ROLE OF METAL ION BINDING AND CONFORMATIONAL TRANSITIONS IN HAMMERHEAD RIBOZYME

The hammerhead ribozyme (HHR) is a small archetype ribozyme that undergoes a conformational transition into a catalytically active conformation in a step that is concerted with changes in metal ion binding in the active site. Here, we present a summary of recent advances in the application of molecular simulation methods to study the mechanisms of HHR catalysis. We focus on the nature and role of conformational transitions and metal ion binding on structure and catalytic activity of HHR.

3.1 Introduction

The original notion that the only function of RNA molecules was that of a messenger intermediate along the pathway from the genetic code to protein synthesis has undergone

---

This chapter compiles material published in: T. S. Lee, G. M. Giambaşu, C. P. Sosa, M. Martick, W. G. Scott and D. M. York *Threshold occupancy and specific cation binding modes in the hammerhead ribozyme active site are required for active conformation*, J. Mol. Biol., 388, 195-206, (2009); T. S. Lee, C. Silva-Lopez, G. M. Giambaşu, M. Martick, W. G. Scott, and D. M. York, *Role of Mg²⁺ in hammerhead ribozyme catalysis from molecular simulation*, J. Am. Chem. Soc., 130, 3053-3064, (2008).; T. S. Lee, G. M. Giambaşu, A. Moser, K. Nam, C. Silva-Lopez, F. Guerra, O. Nieto-Faza, T. J. Giese, J. Gao and D. M. York *Unraveling the mechanisms of ribozyme catalysis with multi-scale simulations*, Multi-scale Quantum Models for Biocatalysis: Modern Techniques and Applications., D. M. York and T. S. Lee eds., Springer Verlag, New York, 2009; T. S. Lee, G. M. Giambaşu, and D. M. York *Insights into the Role of Conformational Transitions and Metal Ion Binding in RNA Catalysis from Molecular Simulations*, Annual Reviews in Computational Chemistry, (in press).

a revolution in the past two and a half decades. The role of RNA in cellular function is now known to be considerably more diverse, ranging from regulation of gene expression and signalling pathways to catalyzing important biochemical reactions, including protein synthesis itself.[92, 93, 94, 95, 15, 96, 97] These discoveries have transformed our view of RNA as a simple messenger to one more profoundly central in the evolution of life forms, our understanding and appreciation of which is still in its infancy. Ultimately, the elucidation of the mechanisms of RNA catalysis will yield a wealth of new insights that will extend our understanding of biological processes and facilitate the design of new RNA-based technologies. [98, 99, 100]

Simulations of biological systems at the atomic level could potentially offer access to the most intimate mechanistic details that may aid in the interpretation of experiments and provide predictive insight into relevant drug design or therapeutic efforts.[101] A quantum mechanical description is ultimately required for reliable study of chemical reactions, including reactions catalyzed by biological macromolecules such as RNA, but at the same time, a high-level fully quantum mechanical treatment of these systems in molecular simulations is not yet feasible. A practical approach involves the use of so-called "multiscale models" that require only a small, typically localized region of the system to be treated with the most computationally costly quantum mechanical methods. The term multiscale model here implies the integration of a hierarchy of models that work together to provide a computationally tractable representation of a complex biochemical reaction in a realistic environment. As a specific example, for enzyme systems, one typically treats the reactive chemical events with a sufficiently accurate high-level quantum mechanical model, the microscopic solvent fluctuations and changes in molecular conformation using molecular mechanical force field model, and the macroscopic dielectric relaxation using a continuum solvation model. The most simple and widely applied multiscale model to study enzyme reactions is the use of a combined quantum mechanical/molecular mechanical (QM/MM) potential. [102, 103, 104, 105, 106]

RNA catalysis simulations are particularly laden with challenges not apparent for most other biological systems, such as protein enzymes. RNA molecules are highly negatively charged, and exhibit strong, and often specific interactions with solvent.[107, 108, 101, 109] This requires special attention to the microscopic *in silico* model that requires consideration of a very large number of solvent molecules and counter and co-ions to be included. Electrostatic interactions need to be treated rigorously without cut-off, and long simulation times are typically needed to insure that the ion environment is properly equilibrated. [110, 111, 112] These issues are further complicated by the fact that RNA molecules bind divalent metal ions that play an important role in folding, and in many instances, also contribute actively to the catalytic chemical steps. The highly-charged nature of RNA and its interaction with divalent metal ions and other solvent components makes inclusion of explicit electronic polarization in the molecular models much more important than in typical protein enzyme systems. The chemistry involved in reactions of prototype ribozymes such as cleavage transesterification involve large changes in local charge state and hybridization around phosphorus, exacerbating the need to design QM/MM methods that can reliably model hypervalent states of phosphorus. There is a need to design new models that circumvent the need for "atom-type" parameters to be assigned to the QM system in order to compute QM/MM interactions, as the "atom-type" can change as a reaction proceeds. Finally, there is a growing precedent that many ribozyme reactions may involve large changes in conformation and metal ion binding along the reaction coordinate, creating the need to develop extremely fast semiempirical quantum models that can be practically applied in conjunction with long-time simulations to adequately sample relevant configurations and create multidimensional free energy surfaces along multiple reaction coordinates.

The hammerhead ribozyme (HHR) [113, 114] is an archetype system to study RNA catalysis. [115, 116] HHR catalyzes the site-specific attack of an activated 2'OH nucleophile to the adjacent 3' phosphate, resulting in cleavage of the $P\text{-}O_{5'}$ phosphodiester linkage to form a 2',3' cyclic phosphate and a 5' alcohol. A detailed understanding of

the structure-function relationships in the hammerhead ribozyme [113, 97] will ultimately aid in the understanding of other cellular RNA catalysts such as the ribosome. The HHR has gained attention as a potential anti-HIV- 1 therapeutic agent [117, 118], an inhibitor of *BCR-ABLi1* gene expression, [119] an inhibitor of hepatitis-B virus gene expressions, [120, 121] and as a tool in drug design and target discovery for other diseases. [98, 122] Very recently, a discontinuous hammerhead ribozyme motif has been found embedded in the 3' untranslated regions of a mammalian messenger RNA, suggesting a possible role in post-transcriptional gene regulation.[123] However, the detailed reaction mechanism of HHR is still elusive despite significant experimental and theoretical work.[113, 115, 124, 125]

One aspect of the catalytic mechanism that has perplexed the community involves the specific role of divalent metal ions in catalysis. Specifically, one of the main puzzles involves the apparent inconsistency between the interpretation of thio substitution [126, 127] and mutational [125] experiments with available crystallographic structural information of the minimal hammerhead sequence.[128, 129, 130] Biochemical experiments have been interpreted to suggest that a pH-dependent conformational change must precede or be concomitant with the catalytic chemical step, including a possible metal ion bridge between the A9 and scissile phosphates. This is inconsistent with crystallographic data for the minimal hammerhead motif, [128, 129, 130] where A9 and scissile phosphates are found to be ~20 Å apart. Moreover, the function of the 2'OH group of G8 remains unclear from this data. [113, 97] Recent crystallographic studies of a full length HHR have characterized the ground state active site architecture [131] and its solvent structure, [132] including the binding mode of a presumed catalytically active divalent metal ion in the active site. These findings, together with molecular simulation studies, [133, 134, 135, 136] have reconciled a long-standing controversy between structural and biochemical studies for this system.[137]

In this chapter, we summarize our recent efforts to unveil the detailed mechanisms

of HHR catalysis using molecular dynamics simulations and hybrid quantum mechanical/molecular mechanical (QM/MM) calculations. We emphasize the characterization of metal binding modes at different stages along the reaction coordinate, the role of metal ion interactions on structure and activity, and the origin of mutational effects on catalysis. Detailed analysis of these binding modes affords insight into the role of divalent metal ions in catalysis. Moreover, long time molecular dynamics simulations allow characterization of the electrostatic environment of the hammerhead ribozyme and the ability to recruit cationic charge, that when threshold occupancy is achieved, leads to formation of catalytically active conformations. Further study of hammerhead ribozyme mutations provide an atomic level interpretation of the origin of mutational effects.

3.2   Metal Binding Modes

Molecular dynamics (MD) simulations were set up to explore the metal binding modes in the hammerhead ribozyme and their relation to structure and catalysis. A recent joint crystallographic/molecular simulation study [132] of the solvent structure of the full length HHR indicates that, in the reactant state prior to activation of the nucleophile, an Mn^{2+} ion is coordinated to the O_{2P} atom of the A9 phosphate and the N7 atom of G10.1. This binding site is designated as the "C-site". An alternate binding site, where a divalent metal ion bridges $A9:O_{2P}$ and the scissile phosphate $(C1.1:O_{2P})$ and is designated as the "B-site", has been inferred from thio/rescue effects [126, 127] and also predicted from molecular simulations [133, 134]. In the absence of divalent metal ions, HHR activity can be recovered by high concentrations of monovalent ions. [138, 139, 140] The specific metal ion binding modes at different stages along the HHR reaction coordinate, and their relation to formation of catalytically active structures has not yet been determined. Here we report results from a series of molecular dynamics simulations that aim to provide atomic level insight into these questions.

To explore the divalent and monovalent metal ion binding modes (Fig. 20) and their

relation to formation of catalytically active, in-line attack conformations in both the neutral reactant and activated precursor (deprotonated 2'OH nucleophile) states, we set up the following series of simulations:

1. **RT-C-Mg**, the reactant state with Mg^{2+} at the C-site.

2. **RT-B-Mg**, the reactant state with Mg^{2+} at the bridging position.

3. **dRT-C-Mg**, the activated precursor with Mg^{2+} at the C-site.

4. **dRT-B-Mg**, the activated precursor with Mg^{2+} at the bridging position.

5. **RT-Na**, the reactant state in the absence of Mg^{2+} (presence only of NaCl).

6. **dRT-Na**, the activated precursor in the absence of Mg^{2+} (presence only of NaCl).

In the dRT-C-Mg simulation involving the activated precursor with Mg^{2+} initially placed at the C-site, the Mg^{2+} ion quickly (less than 200 ps) migrates into the B-site position, as observed in a previous simulation study [134], and afterward exhibits nearly identical behavior as if the Mg^{2+} ion was initially placed at the B-site (dRT-B-Mg). Hence, we only extended the dRT-B-Mg simulation to 300 ns, and designate it simply as dRT-Mg. All 5 simulations were carried out to 300 ns in a background of 0.14 M NaCl. The equilibration for each simulation was monitored by the root-mean-square-deviation (RMSD), and was observed to reach a steady state after 30-50 ns . Hence all analysis was performed over the last 250 ns of trajectories for each simulation.

3.2.1 Threshold occupancy and specific coordination patterns of ions are highly correlated with catalytically active conformations

The formation of an in-line attack conformation, characterized by the "in-line fitness"[141] has been used as a measure for the likelihood of activity of a ribozyme.[142] To understand the relationship between the active in-line conformation and the metal ion binding patterns, we first examine the metal occupation and coordination numbers in the active site. Four key

coordination sites are used for characterizing metal binding (Figure 20): $G8:O_{2'}$, $A9:O_{2P}$, $C1.1:O_{2P}$, and $C17:O_{2'}$. The averages, denoted by $\langle\rangle$, of the following three quantities were calculated for both Mg^{2+} and Na^+: N: the number of ions with at least one coordination to any one of the four key coordination sites, CN: the total coordination number of ions with at least one coordination to any one of coordination sites, and NB: the number of ions which coordinate to at least two of the four coordination sites.

The active conformation can be examined by the distribution of the nucleophile in-line attack angle and distance. Their distributions from all simulations are shown in Figure **??** and manifest two k-mean clusters: [143] cluster A (red) containing structures in the near in-line attack conformations, and cluster B (blue) containing structures not in the near in-line attack conformations, with the exception of the dRT-Mg simulation, where only cluster A is found.

The metal occupation and coordination numbers were then analyzed for each cluster and the results are listed in Table 8. During all simulations, when present in the active site, one Mg^{2+} ion remained stably bound at full occupancy ($\langle N_{Mg^{2+}}\rangle$=1.00). In the reactant state simulations RT-C-Mg and RT-B-Mg, there is no significant population of Na^+ ions in the active site($\langle N_{Na^+}\rangle \leq 0.1$), which suggests that the Mg^{2+} ion is sufficient to neutralize the local charge of the A9 and scissile phosphates. In the RT-C-Mg (Mg^{2+} at the C-site) simulation, the Mg^{2+} ion directly coordinates only $A9:O_{2P}$ of the RNA ($\langle CN_{Mg^{2+}}\rangle$=1.00), and thus not involved in a bridge ($\langle NB_{Mg^{2+}}\rangle$=0.00). Cluster A (in-line conformation) represents approximately 21% of the sampled data over the last 250 ns of simulation, whereas cluster B (not in-line conformation) represents the remaining 79%.

The RT-B-Mg (Mg^{2+} at the bridging position) simulation, on the other hand, is dramatically different. In this simulation, the Mg^{2+} ion directly coordinates both $A9:O_{2P}$ and the scissile phosphate $C1.1:O_{2P}$ of the RNA ($\langle CN_{Mg^{2+}}\rangle$=2.00) as a stable bridge ($\langle NB_{Mg^{2+}}\rangle$=1.00). Cluster A, containing a high degree of in-line near attack conformations, represents the vast majority of the sampled data (over 99%), whereas cluster B is observed less than 1% of the

time. This suggests that a bridging Mg^{2+} ion contributes to stabilization of catalytically active conformations in the reactant simulations. This feature is even more pronounced in the dRT-Mg simulations (activated precursor state with Mg^{2+}). There is only a single cluster with in-line conformation ($\theta=155°$). For dRT-Mg, an additional negative charge in the active site is arising from deprotonation of the nucleophile, and a single Na^+ ion is observed in the active site ($\langle N_{Na^+} \rangle=0.97$) that makes direct coordination to only one RNA ligand ($\langle CN_{Na^+} \rangle=1.01$) with essentially no bridging interactions.

On the other hand, the simulations without an active site Mg^{2+} ion are considerably different than those with the divalent ion present. In the reactant RT-Na simulation, cluster A is dominant (87% of the time) and shows a high degree of in-line conformations ($\theta=153°$). Most of cluster A population contains a single Na^+ ion in the active site ($\langle N_{Na^+} \rangle=1.15$), less frequently two Na^+ ions, and the average number of bridging Na^+ ions is 0.88. Cluster B, on the other hand, has a slightly higher active site Na^+ occupation ($\langle N_{Na^+} \rangle=1.38$), but a lower average number of bridging ions (0.66).

In the activated precursor dRT-Na simulation, the average Na^+ occupancy ($\langle N_{Na^+} \rangle$) increases to approximately 3 and 2.5 for clusters A and B, respectively. Cluster B (not in-line conformation) is the dominate population, occurring 76% of the time. Cluster A (in-line conformation) occurs 24% of the time. A striking feature that distinguishes cluster A from B is that it exhibits a very high degree of bridging ion character, in addition to higher Na^+ occupancy. For cluster A, and the average number of these ions that coordinate at least two RNA ligands ($\langle NB_{Na^+} \rangle$) is 2.68 while the number is only 1.36 for cluster B.

These results suggest that the bridging coordination patterns are highly correlated with formation of in-line conformations for both cases with and without Mg^{2+} ions. Besides the above ion occupation and coordination number analysis, we further look into the specific binding patterns for both cases with or without the Mg^{2+} ion.

3.2.2 A bridging Mg^{2+} ion maintains rigid coordination patterns that stabilize in-line attack conformations

In this section we compare the effect of different Mg^{2+} binding modes in both the neutral reactant and activated (deprotonated 2'OH) precursor states on the active site structure and fluctuations. Table 9 lists the averages of key in-line indexes, the A9/scissile phosphate-phosphate distance and Mg^{2+} coordination distances for the RT-C-Mg, RT-B-Mg and dRT-Mg simulations. Figure 20 shows a general schematic view of the active site metal ion coordination from the simulations. The distances and standard deviations indicate that the Mg^{2+} ion retains rigid coordination with the phosphate oxygens over the course of the simulation, being directly coordinated to $A9:O_{2P}$ in all simulations. In the RT-C-Mg simulation, the Mg^{2+} ion coordinates $G10.1:N_7$ indirectly through one of four inner-sphere water molecules. However, this coordination pattern is not highly conducive to formation of an in-line attack conformation. The RT-B-Mg simulation, on the other hand, shows a more rigid Mg^{2+} coordination with both the A9 and scissile phosphate oxygens, and sustains a considerable population of in-line attack conformations. These results suggest that the coordination pattern found in the RT-B-Mg simulation is able to stabilize in-line attack conformations more readily than Mg^{2+} binding at the C-site as in the RT-C-Mg simulation. The dRT-Mg simulation is similar to the RT-B-Mg simulation with regard to exhibiting rigid coordination with the A9 and scissile phosphate oxygens and stabilization of in-line attack conformations.

With the Mg^{2+} ion at the bridging position (RT-B-Mg and dRT-Mg simulations), there is considerably reduced interaction with $G10.1:N_7$, which are compensated by interactions with the $C17:O_{2'}$ that occur through two water molecules in the inner sphere of the Mg^{2+} ion. This interaction is most pronounced in the dRT-Mg simulation where the $C17:O_{2'}$ is deprotonated. In the ground state reactant simulations with Mg^{2+} (RT-C-Mg and RT-B-Mg), no Na^+ ions were observed to infiltrate the active site. In the activated precursor simulation, dRT-Mg, a single Na^+ ion was observed to be bound at high occupancy to the

deprotonated C17:$O_{2'}$ in a manner similar to the M_3 position in Figure 20.

3.2.3 Specific Na$^+$ binding patters are correlated with formation of in-line attack conformations

In this section we explore the monovalent metal ion binding modes that are correlated with formation of catalytically active in-line attack conformations. For the simulations with no Mg^{2+} ions (RT-Na and dRT-Na) in the active site, binding of the Na$^+$ ions to the coordination sites exhibits larger variation, and exchange events occur giving rise to a fairly broad array of coordination patterns. In order to characterize the distribution and frequency of this array of coordination patterns, a binary-coded coordination index is used (Figure 20). This index is defined as follows. When a Na$^+$ is within a cutoff distance (3.0 Å) to a ligand, it is classified as bound to the ligand, and assigned a unique coordination score for binding to that particular ligand. The coordination scores for the four possible coordination sites (ligands) are 1 for G8:$O_{2'}$, 2 for A9:O_{2P}, 4 for C1.1:O_{2P}, and 8 for C17:$O_{2'}$. The coordination index of an ion is the sum of all coordination scores from its bound sites. In this way, the coordination pattern of a Na$^+$ can be uniquely represented by a single number. For example, an index of 12 means a Na$^+$ directly coordinates to both C1.1:O_{2P} and C17:$O_{2'}$ simultaneously (4+8=12). Through this coordination index, the coordination patterns of Na$^+$ ions in the active site can be traced as a time series over the course of the simulation as shown in Figure 21. Distinct colors are also used to distinguish individual Na$^+$ ions present in the active site such that transitions between coordination patterns (indexes) can be monitored.

Figure 21 shows that in the RT-Na simulation, two Na$^+$ ions (colored as red and green) are present in the active site and, most of time, they both have a coordination index of 6, indicating binding to both A9:O_{2P} and C1.1:O_{2P} at the same time (2+4=6, refer to Figure 20). Hence, two Na$^+$ ions collectively act like a single bridging Mg^{2+} ion to hold the negatively charged A9 and scissile phosphates together to maintain an in-line conformation. During the period from approximately 210-240 ns, only one Na$^+$ ion (red) with coordination index

69

of 6 is present in the active site. During this period the in-line angle drops suddenly from around 155 to 120 degrees. In those periods, the in-line conformation is no longer held.

In the dRT-Na simulation, the in-line angle is less well preserved than in the dRT-Mg simulation. This is consistent with the lower activity of the ribozyme in the absence of Mg^{2+}. Nonetheless, there are several periods (e.g., 25-50 ns and 210-270 ns) where an in-line conformation is visited, and again we observe a high correlation between the Na^+ ion coordination index and in-line conformation. When less than three Na^+ ions bind to the active site ligands, the in-line conformation is no longer held, which happens during most of the simulation. During the periods of simulation where three Na^+ ions bind to different ligand sites simultaneously and the in-line angle comes to a ready-to-react value ($\geq$ 150 degrees).

Figure 22 illustrates the different Na^+ binding patterns for clusters A, (defined in Table 8, in-line conformation, lower panels) and cluster B (not in-line conformation, upper panels) from the dRT-Na simulation. The in-line cluster A clearly exhibits three Na^+ bridges that involve $C17:O_{2'}/C1.1:O_{2P}$, $C1.1:O_{2P}/G8:O_{2'}$, and $C1.1:O_{2P}/A9:O_{2P}$. On the other hand, for cluster B, the first two of these bridges are absent with the third one being significantly less pronounced.

The above analysis suggests that the compensation of the negative charges of these three coordination sites, as well as the bridging binding patterns of Na^+ to bring them together, are necessary to keep the in-line conformation in the deprotonated activated precursor state, although the binding patterns are not as rigid as those of Mg^{2+}.

3.2.4   HHR folds to form a cation recruiting pocket in the active site

In this section we examine the preferential occupation of cations in the HHR active site. The 3D density contour maps for the Na^+ ion distribution determined over the last 250 ns of simulation (Figure 23) show that the average Na^+ ion density at a medium contour level (left panels, Figure 23) is located near the RNA's phosphate backbone, whereas at high

contour level (right panels, Figure 23) the highest probability Na^+ occupation sites were all concentrated in the active site for both the reactant and activated precursor. No explicit Na^+ ions were initially placed in the active site, and Na^+ ion exchange events were observed to occur.

These results suggests that the HHR folds to form a strong local electronegative pocket that attracts cations from solution (e.g., either Mg^{2+}, if present, or Na^+). As discussed previously, threshold occupancy of cationic charge and specific metal ion binding patterns (in particular, bridging coordination of the A9 and scissile phosphates), stabilize the active site and facilitate formation of in-line attack conformations. As analyzed by NMR, a similar case has been observed in the tetraloop-receptor complex, where the divalent ions were experimentally found to be located at strong electronegative positions formed by the RNA fold. [144]

Together with the known divalent metal ion binding at the C-site, these results provoke the speculation that perhaps the active sites of some ribozymes such as the HHR have evolved to form electrostatic cation binding pockets that facilitate catalysis. In the case of the HHR, this speculation is further supported by the simulated correlation of cation binding mode with formation of active conformations discussed in detail in the previous sections.

3.2.5 Simulation protocols

Initial structures used in the simulations were based on a 2.0 Å crystal structure with divalent Mn^{2+} ions and solvent resolved (PDB: 2OEU) [132], with Mn^{2+} ions replaced by native Mg^{2+} ions in the simulations. The positions of hydrogen atoms were determined using the VMD program (version 1.8.6). [56] Simulations were performed in a cubic cell of 60 Å by 60 Å by 100 Å filled with pre-equilibrated TIP3P [43] waters. The ion atmosphere consisted of Na^+ and Cl^- ions that were added to neutralize the system and reach a concentration of 0.14 M. The resulting system contains 36,534 atoms: 11,463 water molecules, 5

Mg^{2+}, 83 Na$^+$, 29 Cl$^-$, and 2,021 RNA atoms.

Simulations were performed with the NAMD simulation package (version 2.6) [44] using the all-atom Cornell *et al.* force field (parm94) [45] in CHARMM format, provided in the AMBER 9 package [145, 47]. Periodic boundary conditions were used along with the isothermal-isobaric ensemble (*NPT*) at 1 atm and 298 K using extended system pressure algorithm [146] with effective mass of 500.0 *amu* and Nosé-Hoover thermostat [147, 148] with effective mass of 1000.0 kcal/mol-ps^2, respectively. The smooth particle mesh Ewald (PME) method [51, 52] was employed with a B-spline interpolation order of 6 with κ value of 0.2579. FFT grid points of 60, 60, and 100 were used for the lattice directions x, y, and z, respectively. Non-bonded interactions were treated using an atom-based cutoff of 12 Å with switching of non-bond potential beginning at 10 Å. Numerical integration was performed using the leap-frog Verlet algorithm with 1 fs time step. [53]

Covalent bond lengths involving hydrogen were constrained using the SHAKE algorithm. [54]

The following equilibration procedures (total 10 ns) were applied to the system prior the production simulations. The positions of the solute atoms, including the Mg^{2+} ion, were restrained by a harmonic potential of 50 kcal/mole/Å^2 in the equilibration stages.

Pre-annealing stage: Water and ion molecules were first energy-optimized for 2,000 steps then underwent a constant volume simulation annealing: The temperature was increased from 0 K to 298K at the rate of 1 K per ps. The system then was kept at 298 K for 500 ps.

Solvent annealing stage: First step: The temperature increased from 298K to 600 K at the rate of 1 K/ps, then was kept at 600 K for 500 ps with constant volume. Second step: The temperature decreased from 600 K to 298K at the rate of 1K/ps, then was kept at 298K for 1,500 ps with constant volume. Third step: The system was kept at 298K for 3,000 ps at constant pressure (1 atm). The whole annealing stage was repeated twice before the post-annealing stage.

Solute relaxation stage: After the annealing stage, the solute atoms were energy-optimized

and then were allowed to move under harmonic restraints over 500 ps simulation at 298 K with a constant pressure of 1 atm. The harmonic force constant (in kcal mol^{-1} Å^{-2}) on each heavy atom was obtained from the empirical formula $k_i = 25 + 2 \times 10^3 / B_i$ where k_i is the force constant for atom i and B_i is the corresponding crystallographic B-value. The restraints were exponentially released over 500 ps with a half-life decay parameter of 100 ps. At the end of the 500 ps simulation, the restraints were reduced to about 3 percent of the initial restraint values.

Production simulation: After the 10 ns of solvent equilibration, the whole system was energy-optimized and unconstrained dynamics simulation began from 0 K under constant pressure of 1 atm. The temperature was increased to 298 K at the rate of 1 K/ps and then kept fixed at 298 K. The same equilibration process was applied for each simulation. At the first 10 ns production simulation, two harmonic restraints of 20 kcal mol^{-1} Å^{-2} were added to keep the Mg^{2+} ion binding to G10.1:N$_7$ and A9:O$_{2P}$ position. Another three harmonic restraint of 20 kcal mol^{-1} Å^{-2} were used: the distances between G8:H$_{O2P}$ and C1.1:O$_{P5}$, and between G12:H$_1$ and C17:O$_{2'}$, were kept around 1.8 Å to ensure the initial hydrogen bonding; the distance between A9:O$_{2P}$ and C1.1:O$_{2P}$ was kept at 4.3 Å (crystal distance). After 10 ns, all restraints were removed. The motions and relaxation of solvent and counterions are notoriously slow to converge in nucleic acid simulations, [55] and careful equilibration is critical for reliable simulations.

In summary, for each simulation, a total of 20 ns of equilibration (10 ns of solvent/ion relaxation and 10 ns of solvent and structure relaxation) and carried out to 300 ns. Analysis was performed over the last 250 ns with data collected every 10 ps.

3.3 *Simulations along the reaction coordinate*

In the previous section, we showed that the active site Mg^{2+} ion prefers to occupy the C-site and B-site in the reactant state and in the deprotonated active precursor, respectively. Starting with these two possible sites, simulations have performed for transition state mimics to

explore the possible roles of the Mg^{2+} ion in the chemical reaction step, including four MD simulations of the reactant state in protonated (RT) or activated/deprotonated (dRT) form, two simulations of the early transition state (ETS) and two simulations of the late transition state (LTS) with the Mg^{2+} ion initially placed at C-Site (c-) or B-site (b-), and finally two additional QM/MM simulations of the early and late transition state.

3.3.1 Molecular Dynamics studies of transition state mimics

In the transition state mimic simulations with Mg^{2+} initially placed at the C-site, in both early and late transition states (c-ETS and c-LTS), the Mg^{2+} ion migrates from the C-site to the bridging position in less than 0.5 ns and remains at the B-site for the remainder of the simulation, as postulated previously. [133] This migration is likely facilitated by deprotonation of the 2'OH of C17 (the nucleophile), and the accumulation of negative charge that is formed in moving toward the transition state. These results are consistent with thio/rescue effect experiments indicating that both the A9 phosphate oxygens and pro-R_P scissile phosphate oxygen, respectively, exhibit a stereospecific kinetic thio effect in the presence of Mg^{2+} that can be rescued by Cd^{2+} ions. [126]. Since both c-ETS and c-LTS resulted in the migration of Mg^{2+} to the B-site, we will focus only on the early and late transition state mimic simulations with Mg^{2+} initially placed at the B-site (b-ETS and b-LTS) here.

In b-ETS and b-LTS simulations, the distance between the A9 and scissile phosphates keeps around 4 Å and the Mg^{2+} coordination between the C1.1 and A9 phosphate oxygens keeps an axial-axial position along the whole simulations (Table 10). The distance between A9 and scissile phosphates in the crystallographic structure is around 4.3 Å, which is well suited for Mg^{2+}-bridging coordination. [149] A similar situation is found in the deprotonated reactant state simulation in the previous section discussing the Mg^{2+} binding modes.

Specifically, in the early TS mimic simulations, where the nucleophilic $O_{2'}$ and leaving group $O_{5'}$ are equidistant from the phosphorus, the Mg^{2+} ion becomes directly coordinated

to the 2'OH of G8, and is positioned closer to the $O_{5'}$ leaving group. Both early TS mimic simulations with Mg^{2+} initially placed at the bridging position and the C- site position showed very similar results. The coordination of the Mg^{2+} ion in the early TS mimic simulations is consistent with a role of shifting the pK_a of the 2'OH in G8 so as to act as a general acid.

In the late TS mimic simulations, both with Mg^{2+} initially placed at the bridging position and the C-site position, a transition occurs whereby the Mg^{2+} coordination with the 2'OH of G8 is replaced by direct coordination with the leaving group $O_{5'}$ In this way, the Mg^{2+} may provide electrostatic stabilization of the accumulating charge of the leaving group (i.e., a Lewis acid catalyst). [124] At the same time, the 2'OH of G8 forms a hydrogen bond with the leaving group $O_{5'}$ and is positioned to act as a general acid catalyst.

3.3.2   Metal-assisted proton transfer in the general acid step

The classical MD simulations in the previous section suggest that in the early TS, the Mg^{2+} ion is positioned to shift the pK_a of the 2'OH of G8 to act as a general acid, and in the late TS, the Mg^{2+} ion can act as a Lewis acid catalyst to stabilize the leaving group and is poised to assist proton transfer from the 2'OH of G8. The possible roles inferred from purely classical MD simulations are supported by the QM/MM results as a similar binding pattern is observed. The Mg^{2+} ion is bonded to G8:$O_{2'}$ in the early transition state mimic and switches to C1.1:$O_{5'}$ of the leaving group in the late transition state mimic. In the late transition state mimic, the G8:$H_{O_{2'}}$ is strongly hydrogen-bonded to C1.1:$O_{5'}$.

In fact, the proton transfer from G8:$O_{2'}$ to C1.1:$O_{5'}$ occurs spontaneously within 1 ns in the QM/MM simulation of late transition state mimicr. At the start of the simulation, C1.1:$O_{5'}$ is tightly bound to Mg^{2+} while G8:$O_{2'}$ is around 3 Å away from Mg^{2+}. As the simulation proceeds, G8:$O_{2'}$ moves closer to Mg^{2+} and eventually binds to Mg^{2+} as it gives up its proton to C1.1:$O_{5'}$. After being protonated, the binding between C1.1:$O_{5'}$ and Mg^{2+} becomes weaker and their direct coordination becomes broken. This QM/MM simulation

confirms that Mg^{2+} enhances the acidity of $G8:O_{2'}$ and facilitates proton transfer to the leaving group.

3.3.3 Simulation protocols

Molecular dynamics siumulations were set up with protocols identical to decribed in Section 3.2.5 with the exception that the simulations reported in this section were performed with CHARMM [150] (version c32a2) using the all-atom CHARMM27 nucleic acid force field [151, 152] with extension to reactive intermediate models (e.g., transition state mimics) [149].

QM/MM simulations on the early and late transition state mimics were set up as follows. Initial structures were taken from snapshots of the classical molecular dynamics simulations after 2 ns production simulation, and the $O_{2'}$-P and P-$O_{5'}$ distances of the scissile phosphate were harmonically restrained with a force constant of 1000 kcal/mol/$Å^2$ and equilibrium distances of 2.010 and 1.850 Å, respectively, for the early TS mimic and 1.856 and 2.382 Å, respectively, for the late TS mimic. The system is partitioned into a quantum mechanical (QM) region constituting the active site that is represented by the AM1/d-PhoT Hamiltonian[153] and the modified AM1 magnesium parameters of Hutter and co-workers. [154] The total number of solute and solvent atoms, setup of periodic boundary conditions, etc., was identical to the classical simulations. The QM subsystem was defined as the 43 atoms around the active site, and included the scissile and A9 phosphates, parts of the nucleophilic and leaving ribose rings, and Mg^{2+} ion and coordinated waters. The generalized hybrid orbital (GHO) method [155] is used to cut a covalent bond to divide the system into QM and MM region. Full electrostatic interactions were calculated using a recently introduced linear-scaling QM/MM-Ewald method. [156]

3.4 Simulations of mutations of key residues

In this section, we report the results obatined from molecular dynamics simulations of the native and mutated full length hammerhead ribozymes in the reactant state and in an activated precursor state (C17:$O_{2'}$ deprotonated). Each simulation has a production trajectory of 60 ns. There exists a wealth of experimental mutational effect data, and simulations of key mutations can offer a detailed rationalization of these effects in terms of structure and dynamics, forge a closer connection between theory and experiment, and provide deeper insight into mechanism. Simulations were performed with the C3U, G8A, and G8I single mutants, and a C3U/G8A double-mutant that exhibits an experimental rescue effect[157].

Simulation-derived key active site structural parameters are provided in Table 7, and representative hydrogen-bond base pairing at the C3-G8 positions are shown in Figure 24. In addition a set of control simulations were performed on a benign U7C mutation, and of the wild type with the active site Mg^{2+} ion removed. An implicit assumption herein is that the mutated sequences fold to a native-like structure.

Mg^{2+} can migrate to a bridging position in the activated precursor state. In all reactant-state simulations, the Mg^{2+} stays near A9:O_{2P} and G10.1:N_7. In all activated precursor state (deprotonated C17:$O_{2'}$) simulations, after few hundred ps, the Mg^{2+} migrates into a bridging position between A9:O_{2P} and C1.1:O_{2P}, and reduces the distance (d_0 in Table 7) by 1 Å relative to the reactant-state simulations. The C17:$O_{2'}$ has significant in-line fitness for nucleophilic attack on C1.1:P. In all simulations except G8A, the G8:$O_{2'}$ is hydrogen bonded to the leaving group, C1.1:$O_{5'}$ and positioned to act as the general acid.

C3U mutation disrupts the active site in the reactant state. The C3U mutation reduces the catalytic rate by a factor ~3×10^{-4}. [158] The C3U mutation disrupts the normal Watson-Crick hydrogen bonding with G8 (Figure 24), causing a base shift that disrupts the active site structure in the reactant state. The distance between the A9 and scissile phosphate increases more than 3.5 Å and breaks key hydrogen bonds between the $O_{2'}$ nucleophile of C17 and N_1 of G12 (the implicated general base), and between the $O_{5'}$

leaving group of C1.1 and $H_{2'}$ of G8 (the implicated general acid). These perturbations in the reactant state would prevent activation of the nucleophile and progress toward the transition state. Experimental evidence shows that C3U indeed reduces the rate constant by more than three orders of magnitude. [158]

G8A mutation disrupts the positioning of G8:$O_{2'}$ as general acid in the activated precursor state. The G8A mutation reduces the catalytic rate by a factor ≤ 0.004.[159] Simulation results indicate the G8A mutation considerably weakens the base pair with C3 with only one weak hydrogen bond that remains intact (Figure 24). In the reactant state simulation, G8A does not appear to dramatically alter the active site contacts relative to the wild-type simulation, with the exception of the A8:$N_1 \cdots$ C3:N_3 distance which increases due to a shift in the hydrogen bond pattern (Figure 24). In the activated precursor state, however, the hydrogen bond positioning between G8:$H_{2'}$ and C1.1:$O_{5'}$ is significantly disrupted relative to the wild- type simulation. The G8A mutation shifts the conserved 2'OH of G8 away from the ideal general acid position, and can possibly block the general acid step of the reaction. Mutation of G8 to 2-aminopurine (AP) [125, 160] or to 2,6-diaminopurine (diAP),[160] which are expected to have similarly weakened hydrogen bonding as the G8A mutation reduces the reaction rate by over three orders of magnitude.

G8I and C3U/G8A mutations are relatively benign. Whereas the relatively isosteric C3U and G8A mutations lead to considerably reduced catalytic rates, the G8I [125, 160] and C3U/G8A [157] mutations affect the rate by less than an order of magnitude. The C3U/G8A double-mutation and G8I single mutation simulations indicate that the hydrogen bond network retains the overall base positions relative to the wild-type simulation and suggest that these two mutations do not significantly alter any of the active site indexes that would affect activity relative to the wild-type simulation (Table 7).

Structural deviations that give rise to mutation effects can occur at different stages along the reaction path. Although the canonical Watson-Crick hydrogen bond network is altered significantly in both C3U and G8A mutations, the simulations suggest that the

origin of the mutational effect on the ribozyme kinetics can occur at different stages along the reaction path. In the reactant state, the Mg^{2+} ion is bound between the $G10.1:N_7$ and $A9:O_{2P}$. The large base-pair shift that occurs in the C3U mutation simulation results in compromise of the active site structure, including the loss of interactions between the proposed general base and nucleophile. In the activated precursor state, the Mg^{2+} ion occupies a bridging position between $A9:O_{2P}$ and the scissile phosphate. The G8A mutation, which is very weakly hydrogen bonded, does not sustain a catalytically viable position of the general acid.

Hydrogen bonding between nucleobases in the 3 and 8 positions is necessary but not sufficient to preserve active site structural integrity. The G8I and C3U/G8A mutations that largely preserve a stable base-pair hydrogen bonded scaffold lead to relatively benign mutations. A C3G/G8C base-pair switch mutation that preserves hydrogen bonded base pairing partially rescues activity relative to the single mutations, although still reduces activity 150-200 fold. [131, 161] Recent analysis of all base-pair mutations indicate considerable variation in activity, but all of the non-native mutations at this position are considerably less active. [161] The present simulation results offer the prediction that whereas both C3U [158] and G8diAP[160] single mutations are observed experimentally to reduce catalytic activity by several orders of magnitude, a correlated C3U/G8diAP double mutation, which retains base-pair hydrogen bonding, should exhibit a partial rescue effect in the hammerhead ribozyme.

3.4.1 Simulation protocols

Molecular dynamics simulations were set up with protocols identical to that described in Section 3.2.5. The mutants are generated from the VMD package [56].

3.5 Conclusion

In this chapter, we summarized our progress toward the understanding of hammerhead ribozyme catalysis through a multiscale simulation strategy. This strategy employs long-time molecular dynamics simulations using a classical MM force field in explicit solvent, and specialized MM residues for metal ion interactions and reactive intermediates. Additional shorter time simulations using a combined QM/MM potential we employed that use a recently-developed semiempirical QM model for phosphoryl transfer reactions that was derived from high-level density-functional calculations of reactions important in RNA catalysis. Long-range ionic interactions were treated rigorously with linear-scaling electrostatic methods for periodic systems.

Simulation results for the hammerhead ribozyme paint a picture of the HHR catalysis that includes a novel role for a catalytic metal ion. The HHR folds to form an electrostatic negative pocket to recruit a threshold occupation of cationic charge, either a Mg^{2+} ion or multiple monovalent ions when Mg^{2+} ions are not present. The position and coordination pattern of these ions are important for formation of active in-line attack conformations. In the case of single Mg^{2+} ion bound in the active site, the Mg^{2+} ion initially stays at the C-site in the reactant state and migrates to a bridging position (the B-site) after the nucleophile $(C17:O_{2'})$ is deprotonated. As the reaction proceeds, the Mg^{2+} ion can stabilize the accumulating charge of the leaving group, and bind to the general acid $(G8:O_{2'})$, significantly increasing its ability to act as a general acid catalyst to transfer a proton to the leaving group $(C1.1:O_{5'})$. Our QM/MM studies demonstrate that the Mg^{2+} ion not only facilitates the protonation of the leaving $O_{5'}$, but it also plays an important role in the final dissociation step of the catalysis. The mutational simulation results are consistent with observed mutational data and suggest that the active site fold is well-tuned for the reaction and most disruptions due to mutations have severe impact the HHR catalysis that can occur at different stages of the reaction.

The present work uses molecular simulations to provide a deeper understanding into

how structure and dynamics affect catalysis in the hammerhead and L1 ligase ribozymes. In particular, the role of conformational transitions and metal ion binding are explored. The insight gained from these studies provide guiding principles into catalysis of a archetype ribozyme and ultimately facilitate the design of new RNA-based biomedical technology. Still much work, both experimental and theoretical, is needed to obtain a consensus view of the detailed mechanism, for which the present study provides important groundwork and progress.
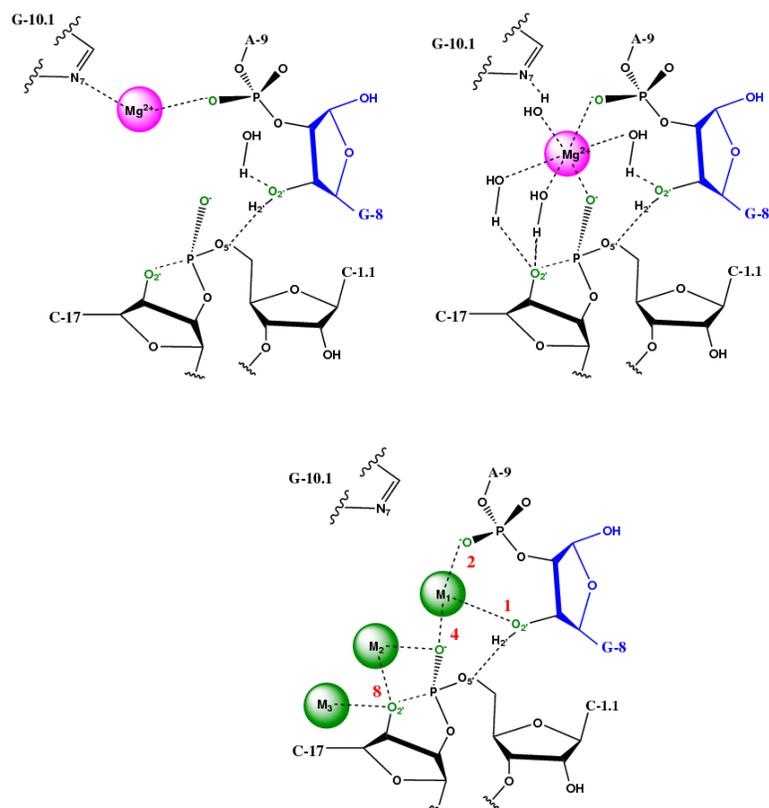
Figure 20: Schematic view of the coordination sites in the hammerhead ribozyme active site. *Upper left*: The coordination pattern of Mg^{2+} in the "C-site" coordinated to G10.1"N_7 and A9:O_{2P}. *Upper Right*: The coordination pattern of Mg^{2+} in the "B-site" bridging A9:O_{2P} and C1.1:O_{2P} of the scissile phosphate. *Lower*: Coordination sites for Na^+ in the hammerhead ribozyme active site found in the RT-Na and dRT-Na simulations. Red numbers next to the coordination sites are the scores used to calculate the coordination index (see text). M_1 involves direct binding to A9:O_{2P} and C.1:O_{2P} and indirect binding to G10.1:N_7 through a water molecule. M_2 involves direct binding to C17:$O_{2'}$ and C.1:O_{2P}. M_3 involves direct binding to C17:$O_{2'}$ and is positioned toward the outside of the active site.
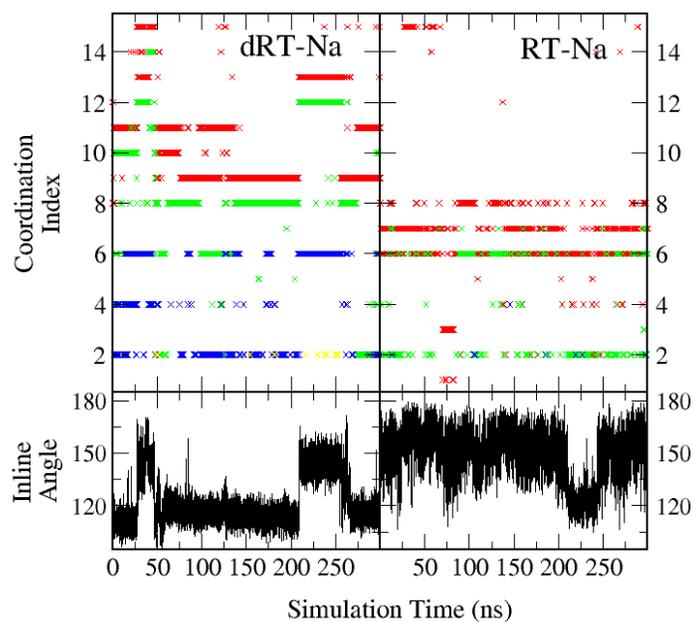
Figure 21: Plot of the in-line attack angle ($O_{2'}$–P–$O_{5'}$) in degrees and the coordination index of Na^+ ions for the dRT-Na (left) and RT-Na (right) simulations. The coordination index is defined as follows: when a Na^+ ion has a distance less than a 3.0-Å cutoff value to a ligand, it is defined as bound to that ligand for indexing purposes. When an ion is bound, the scores for the four possible coordination sites are 1 for G8:$O_{2'}$, 2 for A9:O_{2P}, 4 for C1.1:O_{2P}, and 8 for C17:$O_{2'}$. The coordination index of a single Na^+ ion is the sum of all scores from its bound sites. Individual Na^+ ions are tracked using different colors (red, green, blue and yellow). Data obtained from the last 250 ns are shown in steps of 500 ps.
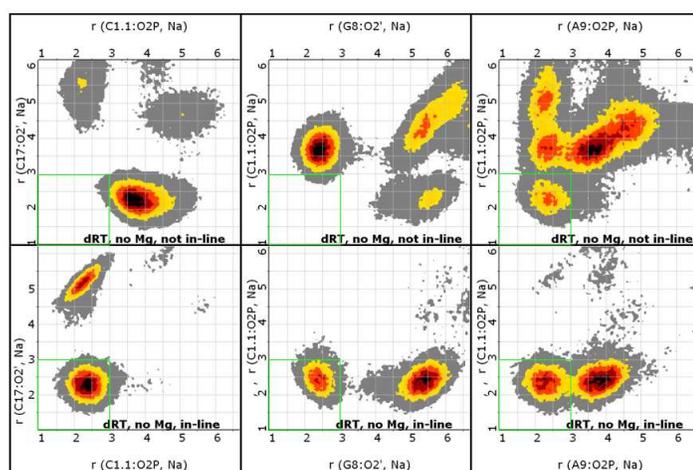
Figure 22: Two dimensional radial distribution function of Na$^+$ ions in the active site for the activated precursor simulation without without Mg^{2+} present in the active site (dRT-Na). The lower panels shows results for cluster A that contains population members that are in active in-line conformations, and the upper panels show results for cluster B that are not in-line (see Table 8). The axes are the distances (in Å) to different metal coordination sites. The green lines indicate the regions where Na$^+$ ions have distances less than 3.0 Å to both sites indicated by the axes.
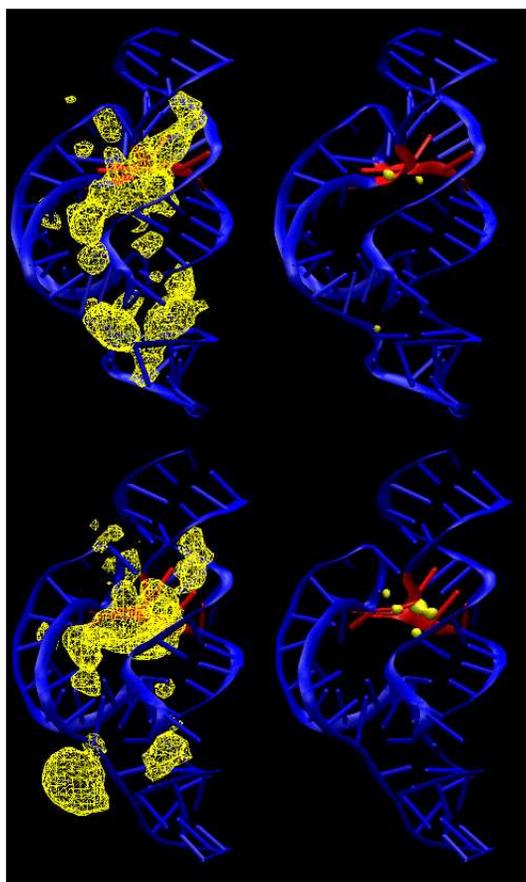
Figure 23: The 3D density contour maps (yellow) of Na^+ ion distributions derived from the RT-Na (upper panels) and dRT-Na simulations (low panels) at different isodensity contour levels (left panels: 0.1; right panels: 1.0). The hammerhead ribozyme is shown in blue with the active site highlighted red. The figure shows that, although the Na^+ ions distribute around the RNA phosphate backbone (left panels), the hammerhead ribozyme folds to form a local electronegative recruiting pocket that attracts a highly condensed distribution of the Na^+ ions (left panels) both in the reactant state and the deprotonated activated precursor state (deprotonated $C17:O_{2'}$) simulations.
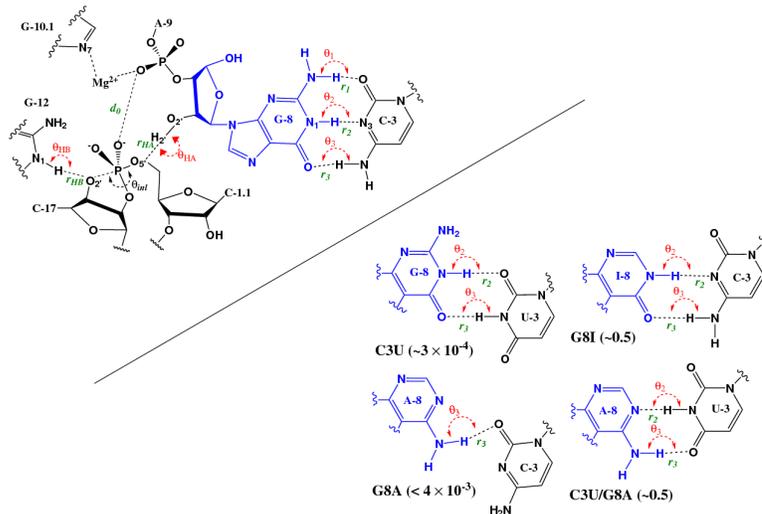
Figure 24: *Upper:* Active site of the full length hammerhead RNA using the canonical minimal sequence numbering scheme described in references [128] and [132]. *Lower:* Representative hydrogen-bonding of the C3:G8 base pair observed from mutant simulations. Experimental relative catalytic rates of mutant versus wild type minimal sequence ribozymes (k_{mut}/k_{wt}) are shown in parentheses (C3U from Ref. [158], G8A from Ref. [159], C3U/G8A from Ref. [157], and G8I from Ref. [125]), and may differ for the full length sequence.

86

Table 7: Characterization of the active site structure and fluctuations. Analysis was performed over the last 25 ns (10 ps sampling). U7C and d-U7C are considered as control simulations as it has been shown U7C mutation has almost no effect on the catalysis [162]. Distance and angles (Figure 24) are in Å and degrees, respectively. Standard deviations (SD) are listed in parenthesis divided by the decimal precision of the average (e.g., if the number is reported to two digits of decimal precision, the SD is divided by 0.01). Boldface font is used to highlight key quantities that are significantly altered with respect to the wild type (WT) simulation upon mutation and that are discussed in the text.

	WT	NoMg	U7C	C3U	G8A	C3U/G8A	G8I
d_0	3.98(40)	3.40(30)	4.06(40)	**7.80(66)**	4.22(34)	3.94(39)	4.27(59)
r_{Nu}	3.98(34)	3.18(12)	4.08(17)	**4.16(14)**	3.2(10)	3.59(48)	3.77(44)
θ_{inl}	128.2(116)	159.6(82)	121.7(65)	127.1(67)	156.0(70)	141.3(169)	131.6(152)
F*	0.29(15)	0.72(10)	0.24(6)	0.23(4)	0.68(9)	0.49(23)	0.37(20)
r_1	2.00(17)	2.12(23)	1.98(17)	–	–	–	–
θ_1	162.9(89)	163.4(89)	163.7(85)	–	–	–	–
r_2	2.01(10)	2.04(13)	2.00(10)	2.25(26)	–	2.04(21)	2.01(14)
θ_2	160.9(90)	162.7(83)	161.0(88)	157.0(127)	–	160.7(104)	162.0(88)
r_3	1.88(12)	1.86(11)	1.89(12)	1.90(13)	2.13(39)	2.06(22)	1.98(19)
θ_3	164.2(85)	161.8(98)	164.8(81)	163.1(88)	149.4(151)	161.2(101)	163.6(90)
r_{NN}†	2.97(9)	3.01(12)	2.96(9)	**3.61(21)**	**5.33(66)**	3.00(19)	2.98(13)
r_{HB}	2.10(26)	1.96(14)	2.11(27)	**3.51(66)**	2.01(16)	2.38(46)	2.05(26)
θ_{HB}	152.1(138)	154.9(90)	150.7(136)	**121.2(155)**	164.2(85)	147.1(181)	155.6(124)
r_{HA}	2.83(44)	4.61(94)	2.92(55)	**7.82(58)**	2.97(33)	3.27(64)	2.83(86)
θ_{HA}	118.2(168)	90.7(506)	112.6(205)	**45.2(164)**	119.3(150)	130.6(198)	118.5(344)
%‡	24.4	0.0	22.4	**0.0**	31.2	13.2	38.4
	d-WT§	d-NoMg	d-U7C	d-C3U	d-G8A	d-C3U/G8A	d-G8I
d_0	2.96(12)	3.51(76)	2.94(12)	2.94(13)	2.95(12)	2.93(12)	2.93(13)
r_{Nu}	3.54(17)	3.85(44)	3.59(15)	3.67(16)	3.80(18)	3.62(16)	3.58(18)
θ_{inl}	158.6(78)	**139.2(185)**	156.2(69)	154.2(74)	146.6(88)	155.5(73)	156.6(73)
F*	0.52(10)	**0.34(18)**	0.49(8)	0.45(8)	0.38(8)	0.48(8)	0.50(9)
r_1	1.90(14)	1.94(15)	2.02(17)	–	–	–	–
θ_1	160.1(103)	162.6(91)	163.7(85)	–	–	–	–
r_2	1.98(11)	1.99(10)	2.00(10)	1.92(15)	–	2.08(17)	1.95(11)
θ_2	163.1(84)	161.6(85)	162.6(86)	160.5(109)	–	160.8(101)	164.3(82)
r_3	2.02(20)	1.91(13)	1.88(11)	1.93(15)	**2.82(44)**	1.96(17)	2.11(22)
θ_3	163.3(93)	164.7(83)	163.8(85)	161.9(94)	**106.9(116)**	160.3(105)	163.2(95)
r_{NN}†	2.96(10)	2.96(9)	2.97(9)	**3.75(20)**	**6.58(46)**	3.04(15)	2.93(11)
r_{HB}	2.16(55)	1.97(18)	1.95(15)	1.94(14)	1.89(12)	1.95(14)	1.99(16)
θ_{HB}	152.0(151)	154.9(90)	155.7(96)	158.5(90)	155.5(93)	156.0(99)	
r_{HA}	2.82(95)	**3.28(130)**	2.31(51)	2.99(99)	**4.10(63)**	2.61(91)	2.47(52)
θ_{HA}	126(398)	**92.4(559)**	143.4(277)	108.7(397)	**82.8(215)**	130.6(385)	138.6(286)
%‡	62.8	45.6	83.6	43.6	**0.8**	68.8	77.2

*In-line fitness index.[141] †The $N_3 \cdots N_1$ distance between nucleobases in the 3 and 8 position. ‡The hydrogen bond contact percentage of the general acid with the leaving group defined as the percentage of the snapshots in which $r_{HA} \leq 3.0$ Å and $\theta_{HA} \geq 120^o$. §The notation "d-" denotes the activated precursor state simulations having the $C17:O_{2'}$ deprotonated.

Table 8: Coordination patterns of Mg^{2+} and Na^+ ions in active site. Distances and angles (Figure 20) are in Å and degrees, respectively. The average values, denoted as $\langle \cdots \rangle$, are obtained by averaging over all snapshots in the cluster. R is the in-line attack distance (C17:$O_{2'}$ to C1.1:P). θ is the in-line attack angle (between C17:$O_{2'}$, C1.1:P, and C1.1:$O_{5'}$). N is the number of ions with at least one coordination to any one of the four coordination sites. CN is the total coordination number of all ions with at least one coordination to any one of the four coordination sites. NB is the number of ions which coordinate to at least two of the four coordination sites.

	Cluster	Percentage	R	θ	$\langle N_{Mg^{2+}} \rangle$	$\langle CN_{Mg^{2+}} \rangle$	$\langle NB_{Mg^{2+}} \rangle$	$\langle N_{Na^+} \rangle$	$\langle CN_{Na^+} \rangle$	$\langle NB_{Na^+} \rangle$
RT-C-Mg	A	20.78%	3.30	144.05	1.00	1.00	0.00	0.05	1.00	0.00
	B	79.22%	4.13	122.66	1.00	1.00	0.00	0.03	1.00	0.00
RT-B-Mg	A	99.54%	3.27	151.10	1.00	2.00	1.00	0.00	1.00	0.00
	B	0.46%	4.00	129.76	1.00	2.00	1.00	0.09	1.00	0.00
RT-Na	A	86.72%	3.23	152.91	-	-	-	1.15	1.99	0.88
	B	13.28%	4.12	122.82	-	-	-	1.38	1.54	0.66
dRT-Mg	A	100.00%	3.64	154.89	1.00	2.00	1.00	0.97	1.01	0.01
dRT-Na	A	23.99%	3.50	144.72	-	-	-	2.96	2.29	2.68
	B	76.01%	4.30	115.16	-	-	-	2.46	1.69	1.36

Table 9: Characterization of the Mg^{2+} coordination in the active site. Analysis was performed over the last 250 ns (10 ps sampling frequency). Distances and angles (Figure 20) are in Å and degrees, respectively. Standard deviations (SD) are listed in parenthesis divided by the decimal precision of the average (e.g., if the average is reported to two digits of decimal precision, the SD is divided by 0.01). R is the in-line attack distance (C17:$O_{2'}$ to C1.1:P), and θ is the in-line attack angle (between C17:$O_{2'}$, C1.1:P, and C1.1:$O_{5'}$). $O--O$ is the distance between A9:O_{2P} and C1.1:O_{2P}, and other distances are between the Mg^{2+} and the indicated ligand site.

	R	θ	O–O	A9:O_{2P}	C1.1:O_{2P}	C17:$O_{2'}$	G8:$O_{2'}$	G10:N_7
RT-C-Mg	4.01(34)	126.5(119)	4.14(49)	2.01(4)	4.40(30)	6.04(90)	5.76(46)	4.19(31)
RT-B-Mg	3.28(12)	151.2(79)	2.95(13)	2.02(5)	2.04(5)	4.25(24)	4.57(30)	4.38(25)
dRT-Mg	3.64(17)	155.0(80)	2.94(13)	2.01(4)	2.03(5)	3.76(17)	4.62(62)	5.05(26)

Table 10: Comparison of crystallographic and simulation data for selected heavy-atom distances (Å) in the hammerhead active site. Results are from simulations with Mg^{2+} initially placed at the bridging position for the reactant state (b-RT), the early transition state mimic (b-ETS), and the late transition state mimic (b-LTS). Average values are shown with standard deviations in the parenthesis (divided by the decimal precision). X-ray structures used for comparison include the full length hammerhead RNA crystallographic structure at 2.2 Å resolution (2GOZ)[131] that was also used in this paper as the starting structure, and the 2.0 Å resolution structure with resolved Mn^{2+} sites and solvent (2OEU). [132]

	X-ray		Simulation		
	2GOZ	2OEU	b-RT	b-ETS	b-LTS
$C1.1{:}O_{2P} \cdots A9{:}O_{2P}$	4.33	4.28	3.36(49)	4.00(60)	4.01(70)
$Mg \cdots G8{:}O_{2'}$	3.04	3.14	3.97(102)	2.24(13)	3.21(23)
$Mg \cdots C1.1{:}O_{5'}$	3.84	4.01	4.22(21)	3.68(35)	2.09(50)
$G8{:}O_{2'} \cdots C1.1{:}O_{5'}$	3.19	3.51	4.29(77)	4.41(65)	2.91(17)
$C17{:}O_{2'} \cdots C1.1{:}P$	3.18	3.30	3.61(23)	1.89(12)	1.76(40)
$G12{:}N1 \cdots C17{:}O_{2'}$	3.54	3.26	3.02(27)	3.14(28)	2.97(13)
$A9{:}N6 \cdots G12{:}N3$	2.63	3.22	3.27(58)	3.15(21)	3.17(21)
$A9{:}N6 \cdots G12{:}O_{2'}$	3.21	2.98	3.36(86)	3.01(18)	2.99(16)
$A9{:}N7 \cdots G12{:}N2$	2.90	2.90	3.42(93)	3.85(44)	3.66(33)

Table 11: Simulations performed in Section 3.2. Each simulation is given an abbreviation (Abbrev.) that is referenced in the text. The simulations differ in the protonation state of the 2'OH nucleophile (State:2'OH) being either in the "neutral reactant", or "deprotonated precursor" state. The simulations also differed from the presence/absence and initial placement of the Mg^{2+} ions in the active site. In the presence of a Mg^{2+} ion in the active site, initial placement could be in the "C-site" coordinated to O_2P of the A9 phosphate and N7 of G10.1, or in the "B-site" position bridging between the O_2P atoms of the A9 and scissile phosphates. In the "dRT-C-Mg" simulation of the deprotonated precursor with Mg^{2+} initially placed in the C-site, the Mg^{2+} ion quickly migrated to a B-site position, and exhibited dynamics essentially equivalent to the dRT-B-Mg simulation, and thus was not carried out further or analyzed. All simulations were carried out in 0.14 M NaCl at 298 K and 1 atm pressure, and with the exception of the "dRT-C-Mg" simulation mentioned above, were carried out to 300 ns, the last 250 ns of which was used for analysis.

Abbrev.	State:2'OH	Mg^{2+}	[NaCl]	Temp., Press.	simulation	analysis
RT-C-Mg	neutral reactant	C-site	0.14 M	298 K, 1 atm	300 ns	250 ns
RT-B-Mg	neutral reactant	B-site	0.14 M	298 K, 1 atm	300 ns	250 ns
dRT-C-Mg	deprotonated precursor	C-site	0.14 M	298 K, 1 atm	< 1 ns	——
dRT-B-Mg	deprotonated precursor	B-site	0.14 M	298 K, 1 atm	300 ns	250 ns
RT-Na	neutral reactant	absent	0.14 M	298 K, 1 atm	300 ns	250 ns
dRT-Na	deprotonated precursor	absent	0.14 M	298 K, 1 atm	300 ns	250 ns

REFERENCES

[1] David P. Bartel and Jack W. Szostak. Isolation of new ribozymes from a large pool of random sequences. *Science*, 261(5127):1411–1418, Sep 1993.

[2] Eric H. Ekland, Jack W. Szostak, and David P. Bartel. Structurally Complex and Highly Active RNA Ligases Derived from Random RNA Sequences. *Science*, 269:364–369, 1995.

[3] Michael P. Robertson and Andrew D. Ellington. In vitro selection of an allosteric ribozyme that transduces analytes to amplicons. *Nature Biotech.*, 17:62–66, 1999.

[4] Jeff Rogers and Gerald F. Joyce. A ribozyme that lacks cytidine. *Nature*, 402:323–325, 1999.

[5] Luc Jaeger, Martin C. Wright, and Gerald F. Joyce. A complex ligase ribozyme evolved in vitro from a group I ribozyme domain. *Proc. Natl. Acad. Sci. USA*, 96:14712–14717, 1999.

[6] Yoshia Ikawa, Kentaro Tsuda, Shigeyoshi Matsumura, and Tan Inoue. De novo synthesis and development of an RNA enzyme. *Proc. Natl. Acad. Sci. USA*, 101:13750–13755, 2004.

[7] Kathleen E. McGinness and Gerald F. Joyce. In search of an RNA replicase ribozyme. *Chem. Biol.*, 10(1):5–14, Jan 2003.

[8] Sarah C. Bagby, Nicholas H. Bergman, David M. Shechner, Catherine Yen, and David P. Bartel. A class I ligase ribozyme with reduced Mg^{2+} dependence: Selection, sequence analysis, and identification of functional tertiary interactions. *RNA*, 15:2129–2146, 2009.

[9] David M. Shechner, Robert A. Grant, Sarah C. Bagby, Yelena Koldobskaya, Joseph A. Piccirilli, and David P. Bartel. Crystal Structure of the Catalytic Core of an RNA-Polymerase Ribozyme. *Science*, 326:1271–1275, 2009.

[10] Laura F. Landweber and Irina D. Pokrovskaya. Emergence of a dual-catalytic RNA with metal-specific cleavage and ligase activities: The spandrels of RNA evolution. *Proc. Natl. Acad. Sci. USA*, 96:173–178, 1999.

[11] Michael P. Robertson and William G. Scott. The structural basis of ribozyme-catalyzed RNA assembly. *Science*, 315:1549–1550, 2007.

[12] Gerald F. Joyce. A Glimpse of Biology's First Enzyme. *Science*, 315:1507–1508, 2007.

[13] Kristin A. Marshall and Andrew D. Ellington. Training ribozymes to switch. *Nature Struct. Biol.*, 6(11):992–994, 1999.

[14] A. D. Ellington and J. W. Szostak. In vitro selection of rna molecules that bind specific ligands. *Nature*, 346(6287):818–822, Aug 1990.

[15] Xi Chen, Na Li, and Andrew D. Ellington. Ribozyme Catalysis of Metabolism in the RNA World. *Chem. Biodivers.*, 4:633–655, 2007.

[16] Michael P. Robertson, Scott M. Knudsen, and Andrew D. Ellington. In vitro selection of ribozymes dependent on peptides for activity. *RNA*, 10:114–127, 2004.

[17] Michael P. Robertson and Andrew D. Ellington. Design and optimization of effector-activated ribozyme ligases . *Nucleic Acids Res.*, 28:1751–1759, 2000.

[18] Michael P. Robertson and Andrew D. Ellington. In vitro selection of nucleoprotein enzymes. *Nature Biotech.*, 19:650–655, 2001.

[19] James R. Williamson. Induced fit in RNA-protein recognition. *Nature Struct. Biol.*, 7(10):834–837, 2000.

[20] T. Xia. Taking femtosecond snapshots of RNA conformational dynamics and complexity. *Curr. Opin. Chem. Biol.*, 2:1–8, Sep 2008.

[21] N. Leulliot and G. Varani. Current topics in RNA-protein recognition: control of specificity and biological function through induced fit and conformational capture. *Biochemistry*, 40(27):7947–7956, 2001.

[22] Gregory Bokinsky and Xiaowei Zhuang. Single-molecule RNA folding. *Acc. Chem. Res.*, 38:566–573, 2005.

[23] D. K. Treiber and J. R. Williamson. Exposing the kinetic traps in RNA folding. *Curr. Opin. Struct. Biol.*, 9(3):339–345, 1999.

[24] Bohdan Schneider, Zdeněk Morávek, and Helen M. Berman. RNA conformational classes. *Nucleic Acids Res.*, 32:1666–1677, 2004.

[25] Laura J. W. Murray, W. Bryan Arendall, David C. Richardson, and Jane S. Richardson. RNA backbone is rotameric. *Proc. Natl. Acad. Sci. USA*, 100(24):13904–13909, Nov 2003.

[26] Gregory E. Sims and Sung-Hou Kim. Global mapping of nucleic acid conformational space: dinucleoside monophosphate conformations and transition pathways among conformational classes. *Nucleic Acids Res.*, 31(19):5607–5616, Oct 2003.

[27] Jane S. Richardson, Bohdan Schneider, Laura W. Murray, Gary J. Kapral, Robert M. Immormino, Jeffrey J. Headd, David C. Richardson, Daniela Ham, Eli Hershkovits, Loren Dean Williams, Kevin S. Keating, Anna Marie Pyle, David Micallef, John Westbrook, Helen M. Berman, and R. N. A. Ontology Consortium. RNA backbone:

consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *RNA*, 14(3):465–481, Mar 2008.

[28] C. M. Duarte and A. M. Pyle. Stepping through an RNA structure: A novel approach to conformational analysis. *J. Mol. Biol.*, 284(5):1465–1478, Dec 1998.

[29] Leven M. Wadley, Kevin S. Keating, Carlos M. Duarte, and Anna Marie Pyle. Evaluating and learning from RNA pseudotorsional space: quantitative validation of a reduced representation for RNA structure. *J. Mol. Biol.*, 372(4):942–957, Sep 2007.

[30] Wei Yang, Jae Young Lee, and Marcin Nowotny. Making and breaking nucleic acids: two-mg2+-ion catalysis and substrate specificity. *Mol Cell*, 22(1):5–13, Apr 2006.

[31] M. R. Sawaya, R. Prasad, S.H. Wilson, J. Kraut, and H. Pelletier. Crystal structures of human DNA polymerase beta complexed with gapped and nicked DNA: evidence for an induced fit mechanism. *Biochemistry*, 36(37):11205–11215, 1997.

[32] A. Lescoute and E. Westhof. Topology of three-way junctions in folded RNAs. *RNA*, 12(1):83–93, 2006.

[33] M. P. Robertson, J. R. Hesselberth, and A. D. Ellington. Optimization and optimality of a short ribozyme ligase that joins non-Watson-Crick base pairings. *RNA*, 7:513–523, 2001.

[34] T. A. Steitz and J. A. Steitz. A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. USA*, 90:6498–6502, 1993.

[35] Thomas A. Steitz. DNA- and RNA-dependent DNA polymerases. *Curr. Opin. Struct. Biol.*, 3:31–38, 1993.

[36] Neocles B. Leontis, Jesse Stombaugh, and Eric Westhof. The non-watson-crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.*, 30:3497–3531, 2002.

[37] Y. Wang and T. Schlick. Quantum mechanics/molecular mechanics investigation of the chemical reaction in Dpo4 reveals water-dependent pathways and requirements for active site reorganization. *J. Am. Chem. Soc.*, 130(40):13240–13250, 2008.

[38] Ping Lin, Lars C. Pedersen, Vinod K. Batra, William A. Beard, Samuel H. Wilson, and Lee G. Pedersen. Energy analysis of chemistry for correct insertion by DNA polymerase β. *Proc. Natl. Acad. Sci. USA*, 103:13294–13299, 2006.

[39] Lihua Wang, Xinyun Yu, Po Hu, Suse Broyde, and Yingkai Zhang. A Water-Mediated and Substrate-Assisted Catalytic Mechanism for Sulfolobus solfataricus DNA Polymerase IV. *J. Am. Chem. Soc.*, 129:4731–5737, 2007.

[40] D.M. Crothers. *RNA*, chapter RNA Conformational Dynamics, pages 61–71. Elsevier Science & Technology, December 2001.

[41] D. H. J. Bunka and P. G. Stockley. Aptamers come of age – at last. *Nat. Rev. Microbiol.*, 4(8):588–596, Aug 2006.

[42] Harald Schwalbe, Janina Buck, Boris Frtig, Jonas Noeske, and Jens Whnert. Structures of rna switches: insight into molecular recognition and tertiary structure. *Angew Chem Int Ed*, 46(8):1212–1219, 2007.

[43] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926–935, 1983.

[44] James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kaleé, and Klaus Schulten. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.*, 26:1781–1802, 2005.

[45] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, D. M Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.*, 117:5179–5197, 1995.

[46] D.A. Case, T.A. Darden, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, M. Crowley, R.C. Walker, W. Zhang, K.M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, , and P.A. Kollman. *AMBER 10*. University of California, San Francisco, San Francisco, 2002.

[47] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. R. Ross, T.E. Cheatham, III, S. De-Bolt, D. Ferguson, G. Seibel, and P. Kollman. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structure and energetic properties of molecules. *Comput. Phys. Commun.*, 91:1–41, 1995.

[48] David A. Case, Thomas E. Cheatham III, Tom Darden, Holger Gohlke, Ray Luo, Kenneth M. Merz, Alexey Onufriev, Carlos Simmerling, Bing Wang, and Robert J. Woods. The Amber biomolecular simulation programs. *J. Comput. Chem.*, 26:1668–1688, 2005.

[49] Glenn J. Martyna, Douglas J. Tobias, and Michael L. Klein. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.*, 101:4177–4189, 1994.

[50] Scott E. Feller, Yuhong Zhang, Richard W. Pastor, and Bernard R. Brooks. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.*, 103:4613–4621, 1995.

[51] Ulrich Essmann, Lalith Perera, Max L. Berkowitz, Tom Darden, Lee Hsing, and Lee G. Pedersen. A smooth particle mesh Ewald method. *J. Chem. Phys.*, 103(19):8577–8593, 1995.

[52] Celeste Sagui and Thomas A. Darden. Molecular Dynamics Simulations of Biomolecules: Long-Range Electrostatic Effects. *Annu. Rev. Biophys. Biomol. Struct.*, 28:155–179, 1999.

[53] M.P. Allen and D.J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, Oxford, 1987.

[54] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.*, 23:327–341, 1977.

[55] Sergei Y. Ponomarev, Kelly M. Thayer, and David L. Beveridge. Ion motions in molecular dynamics simulations on DNA. *Proc. Natl. Acad. Sci. USA*, 101(41):14771–14775, 2004, http://www.pnas.org/cgi/reprint/101/41/14771.pdf.

[56] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD: Visual Molecular Dynamics. *J. Mol. Grapics*, 14:33–38, 1996.

[57] George M. Giambaşu, Tai-Sung Lee, Carlos P. Sosa, Michael P. Robertson, William G. Scott, and Darrin M. York. Identification of dynamical hinge points of the L1 ligase molecular switch. *RNA*, 16(4):769–780, Apr 2010.

[58] Francesco Rao. Local Transition Gradients Indicating the Global Attributes of Protein Energy Landscapes. *J. Phys. Chem. Lett*, 1(10):1580–1583, May 2010.

[59] Francesco Rao and Amedeo Caflisch. The protein folding network. *J. Mol. Biol.*, 342:299–306, 2004.

[60] Amedeo Caflisch. Network and graph analyses of folding free energy surfaces. *Curr. Opin. Struct. Biol.*, 16(1):71–78, Feb 2006.

[61] D. Gfeller, P. De Los Rios, A. Caflisch, and F. Rao. Complex network analysis of free-energy landscapes. *Proc. Natl. Acad. Sci. USA*, 104:1817–1822, 2007.

[62] Frank Noé and Stefan Fischer. Transition networks for modeling the kinetics of conformational change in macromolecules. *Curr. Opin. Struct. Biol.*, 18:154–162, 2008.

[63] Diego Prada-Gracia, Jesús Gómez-Garde nes, Pablo Echenique, and Fernando Falo. Exploring the free energy landscape: from dynamics to networks and back. *PLoS Comput. Biol.*, 5(6):e1000415, Jun 2009.

[64] Christoph Dellago, Peter G. Bolhuis, and Phillip L. Geissler. Transition path sampling. *Adv. Chem. Phys.*, 123:1–78, 2002.

[65] Peter G. Bolhuis, David Chandler, Christoph Dellago, and Phillip L. Geissler. Transition Path Sampling: Throwing Ropes Over Rough Mountain Passes, in the Dark. *Annu. Rev. Phys. Chem.*, 53:291–318, 2002.

[66] Christoph Dellago, Peter G. Bolhuis, Félix S. Csajka, and David Chandler. Transition path sampling and the calculation of rate constants. *J. Chem. Phys.*, 108:1964–1977, 1998.

[67] J. Rogal and P.G. Bolhuis. Multiple state transition path sampling. *J. Chem. Phys.*, 129(22):224107, Dec 2008.

[68] C. Dellago, P.G. Bolhuis, and P.L. Geissler. Transition Path Sampling Methods. *Lect. Notes Phys.*, 703:349–391, 2006.

[69] Peter Bolhuis, Christoph Dellago, and David Chandler. Reaction coordinates of biomolecular isomerization. *Proc. Natl. Acad. Sci. USA*, 97:5877–5882, 2000.

[70] Wenxun Gan, Sichun Yang, and Benoit Roux. Atomistic view of the conformational activation of Src kinase using the string method with swarms-of-trajectories. *Biophys. J.*, 97(4):L8–L10, Aug 2009.

[71] Albert C. Pan, Deniz Sezer, and Benoit Roux. Finding transition pathways using the string method with swarms of trajectories. *J. Phys. Chem. B*, 112(11):3432–3440, Mar 2008.

[72] Albert C. Pan and Benoit Roux. Building Markov state models along pathways to determine free energies and rates of transitions. *J. Chem. Phys.*, 129(6):064107, Aug 2008.

[73] Luca Maragliano and Eric Vanden-Eijnden. On-the-fly string method for minimum free energy paths calculation. *Chem. Phys. Lett.*, 446(1-3):182–190, September 2007.

[74] Weiqing Ren, Eric Vanden-Eijnden, Paul Maragakis, and Weinan E. Transition pathways in complex systems: Application of the finite-temperature string method to the alanine dipeptide. *J. Chem. Phys.*, 123(13):134109, Oct 2005.

[75] Luca Maragliano, Alexander Fischer, Eric Vanden-Eijnden, and Giovanni Ciccotti. String method in collective variables: minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.*, 125(2):24106, Jul 2006.

[76] Eric Vanden-Eijnden and Maddalena Venturoli. Revisiting the finite temperature string method for the calculation of reaction tubes and free energies. *J. Chem. Phys.*, 130(19):194103, 2009.

[77] Eric Vanden-Eijnden. Some recent techniques for free energy calculations. *J. Comput. Chem.*, 30(11):1737–1747, 2009.

[78] Weinan E, Weiqing Ren, and Eric Vanden-Eijnden. Finite temperature string method for the study of rare events. *J. Phys. Chem. B*, 109(14):6688–6693, Apr 2005.

[79] Weinan E, Weiqing Ren, and Eric Vanden-Eijnden. String method for the study of rare events. *Phys. Rev. B*, 66(5):052301, August 2002.

[80] Thomas F. Miller, Eric Vanden-Eijnden, and David Chandler. Solvent coarse-graining and the string method applied to the hydrophobic collapse of a hydrated chain. *Proc. Natl. Acad. Sci. U. S. A.*, 104(37):14559–14564, Sep 2007.

[81] J. Han, M. Kamber, and A. K. H. Tung. Spatial Clustering Methods in Data Mining: A Survey. In H. Miller and J. Han, editors, *Geographic Data Mining and Knowledge Discovery*, pages 1–29. Taylor and Francis, 2001.

[82] A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: a review. *ACM Comput Surv*, 31(3):264–323, 1999.

[83] Ying Zhao and George Karypis. Data clustering in life sciences. *Mol Biotechnol*, 31(1):55–80, September 2005.

[84] George Karypis. CLUTO: A Clustering Toolkit (release 2.1.2), 2006.

[85] Ying Zhao and George Karypis. Criterion Functions for Document Clustering: Experiments and Analysis. *Machine Learning*, 55:311–331, 2004.

[86] Ying Zhao, George Karypis, and Usama Fayyad. Hierarchical Clustering Algorithms for Document Datasets. *Data Mining and Knowledge Discovery*, 10(2):141–168, March 2005.

[87] Kevin R Brown, David Otasek, Muhammad Ali, Michael J McGuffin, Wing Xie, Baiju Devani, Ian Lawson van Toch, and Igor Jurisica. NAViGaTOR: Network Analysis, Visualization and Graphing. *Bioinformatics*, 25(24):3327–3329, Dec 2009.

[88] Michael J McGuffin and Igor Jurisica. Interaction techniques for selecting and manipulating subgraphs in network visualizations. *IEEE Trans Vis Comput Graph*, 15(6):937–944, 2009.

[89] Stephen Warshall. A Theorem on Boolean Matrices. *JACM*, 9(1):11–12, 1962.

[90] Robert W. Floyd. Algorithm 97: Shortest path. *Commun ACM*, 5(6):345, 1962.

[91] Eric Darve, David Rodrguez-Gmez, and Andrew Pohorille. Adaptive biasing force method for scalar and vector free energy calculations. *J Chem Phys*, 128(14):144120, Apr 2008.

[92] Walter Gilbert. The RNA World. *Nature*, 319:618, 1918.

[93] William G. Scott. Molecular palaeontology: understanding catalytic mechanisms in the RNA world by excavating clues from a ribozyme three-dimensional structure. *Biochem. Soc. Trans.*, 24(3):604–608, 1996.

[94] Raymond F. Gesteland, Thomas R. Cech, and John F. Atkins. *The RNA World: The Nature of Modern RNA Suggests a Prebiotic RNA*. Cold Spring Harbor Laboratory Press, New York, 2nd edition, 1999.

[95] Michael Yarus. Boundaries for an RNA world. *Curr. Opin. Chem. Biol.*, 3:260–267, 1999.

[96] David M. Lilley. *Ribozymes and RNA Catalysis*, chapter The Hairpin and Varkud Satellite Ribozymes, pages 66–91. RSC Biomolecular Series. RSC Publishing, Cambridge, 2008.

[97] William G. Scott. Ribozymes. *Curr. Opin. Struct. Biol.*, 17:280–286, 2007.

[98] M Rubenstein, R. Tsui, and P Guinan. A review of antisense oligonucleotides in the treatment of human disease. *Drugs of the Future*, 29:893–909, 2004.

[99] Narendra K. Vaish, Fand Dong, Lori Andrews, Rebecca E. Schweppe, Natalie G. Ahn, Lawrence Blatt, and Scott D. Seiwert. Monitoring post-tranlation modification of proteins with allosteric ribozymes. *Nature Biotech.*, 20:810–815, 2002.

[100] Ronald R. Breaker. Engineered allosteric ribozymes as biosensor components. *Curr. Opin. Biotechnol.*, 13(1):31–39, 2002.

[101] S. Elizabeth McDowell, Nad'a Špačková, Jiří Šponer, and Nils G. Walter. Molecular Dynamics Simulations of RNA: An In Silico Single Molecule Approach. *Biopolymers*, 85:169–184, 2007.

[102] J. Gao. Methods and applications of combined quantum mechanical and molecular mechanical potentials. *Rev. Comput. Chem.*, 7:119–185, 1995.

[103] Gerald Monard and Kenneth M. Merz, Jr. Combined Quantum Mechanical/ Molecular Mechanical Methodologies Applied to Biomolecular Systems. *Acc. Chem. Res.*, 32:904–911, 1999.

[104] Arieh Warshel. Molecular dynamics simulations of biological reactions. *Acc. Chem. Res.*, 35:385–395, 2002.

[105] Arieh Warshel. Computer simulations of enzyme catalysis: methods, progress, and insights. *Annu. Rev. Biophys. Biomol. Struct.*, 32:425–443, 2003.

[106] Hans Martin Senn and Walter Thiel. QM/MM studies of enzymes. *Curr. Opin. Chem. Biol.*, 11:182–187, 2007.

[107] Jan Norberg and Lennart Nilsson. Molecular Dynamics Applied to Nucleic Acids. *Acc. Chem. Res.*, 35:465–472, 2002.

[108] Modesto Orozco, Alberto Pérez, Agnes Noy, and F. Javier Luque. Theoretical methods for the simulation of nucleic acids. *Chem. Soc. Rev.*, 32:350–364, 2003.

[109] Modesto Orozco, Agnes Noy, and Alberto Prez. Recent advances in the study of nucleic acid flexibility by molecular dynamics. *Curr Opin Struct Biol*, 18(2):185–193, Apr 2008.

[110] Thomas E. Cheatham III. Simulation and modeling of nucleic acid structure, dynamics and interactions. *Curr. Opin. Struct. Biol.*, 14:360–367, 2004.

[111] Shi-Jie Chen. Rna folding: conformational statistics, folding kinetics, and ion electrostatics. *Annu Rev Biophys*, 37:197–214, 2008.

[112] Pascal Auffinger and Yaser Hashem. Nucleic acid solvation: from outside to insight. *Curr. Opin. Struct. Biol.*, 17(3):325–333, Jun 2007.

[113] William G. Scott. Biophysical and biochemical investigations of RNA catalysis in the hammerhead ribozyme. *Q. Rev. Biophys.*, 32:241–294, 1999.

[114] K. F. Blount and O. C. Uhlenbeck. The hammerhead ribozyme. *Biochem Soc Trans*, 30(Pt 6):1119–1122, Nov 2002.

[115] William G. Scott. RNA catalysis. *Curr. Opin. Struct. Biol.*, 8(6):720–726, 1998.

[116] Elizabeth A. Doherty and Jennifer A. Doudna. Ribozyme Structures and Mechanisms. *Annu. Rev. Biophys. Biomol. Struct.*, 30:457–475, 2001.

[117] Nava Sarver, Edouard M. Cantin, Pairoj S. Chang, John A. Zaia, Paula A. Ladne, Delilah A. Stephens, and John J. Rossi. Ribozymes as potential anti-HIV-1 therapeutic agents. *Science*, 247:1222–1225, 1990.

[118] A. Michienzi, L. Cagnon, I. Bahner, and J. J. Rossi. Ribozyme-mediated inhibition of hiv 1 suggests nucleolar trafficking of hiv-1 rna. *Proc Natl Acad Sci U S A*, 97(16):8955–8960, Aug 2000.

[119] D. S. Snyder, Y. Wu, J. L. Wang, J. J. Rossi, P. Swiderski, B. E. Kaplan, and S. J. Forman. Ribozyme-mediated inhibition of bcr-abl gene expression in a philadelphia chromosome-positive cell line. *Blood*, 82(2):600–605, Jul 1993.

[120] M. Weinberg, M. Passman, M. Kew, and P. Arbuthnot. Hammerhead ribozyme-mediated inhibition of hepatitis b virus x gene expression in cultured cells. *J Hepatol*, 33(1):142–151, Jul 2000.

[121] Y. Feng, Y. Y. Kong, Y. Wang, and G. R. Qi. Intracellular inhibition of the replication of hepatitis b virus by hammerhead ribozymes. *J Gastroenterol Hepatol*, 16(10):1125–1130, Oct 2001.

[122] Masayuki Sano and Kazunari Taira. Hammerhead ribozyme-based target discovery. *Methods Mol Biol*, 360:143–153, 2007.

[123] Monika Martick, Lucas H. Horan, Harry F. Noller, and William G. Scott. A discontinuous hammerhead ribozyme embedded in a mammalian messenger RNA. *Nature*, 454(7206):899–902, Aug 2008.

[124] Yasuomi Takagi, Yutaka Ikeda, and Kazunari Taira. Ribozyme Mechanisms. *Top. Curr. Chem.*, 232:213–251, 2004.

[125] Kenneth F. Blount and Olke C. Uhlenbeck. The Structure-Function Dilemma of the Hammerhead Ribozyme. *Annu. Rev. Biophys. Biomol. Struct.*, 34:415–440, 2005.

[126] Shenglong Wang, Katrin Karbstein, Alessio Peracchi, Leonid Beigelman, and Daniel Herschlag. Identification of the Hammerhead Ribozyme Metal Ion Binding Site Responsible for Rescue of the Deleterious Effect of a Cleavage Site Phosphorothioate. *Biochemistry*, 38(43):14363–14378, 1999.

[127] Ken-ichi Suzumura, Yasuomi Takagi, Masaya Orita, and Kazunari Taira. NMR-Based Reappraisal of the Coordination of a Metal Ion at the Pro- Rp Oxygen of the A9/G10.1 Site in a Hammerhead Ribozyme. *J. Am. Chem. Soc.*, 126(47):15504–15511, 2004.

[128] William G. Scott, James B. Murray, John R. P. Arnold, Barry L. Stoddard, and Aaron Klug. Capturing the structure of a catalytic RNA intermediate: The Hammerhead Ribozyme. *Science*, 274:2065–2069, 1996.

[129] James B. Murray, Daniel P. Terwey, Lara Maloney, Alexander Karpeisky, Nassim Usman, Leonid Beigelman, and William G. Scott. The Structural Basis of Hammerhead Ribozyme Self-Cleavage. *Cell*, 92:665–673, 1998.

[130] James B. Murray, Hanna Szöke, Abraham Szöke, and William G. Scott. Capture and Visualization of a Catalytic RNA Enzyme-Product Complex Using Crystal Lattice Trapping and X-Ray Holographic Reconstruction. *Mol. Cell*, 5:279–287, 2000.

[131] Monika Martick and William G. Scott. Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell*, 126(2):309–320, 2006.

[132] Monika Martick, Tai-Sung Lee, Darrin M. York, and William G. Scott. Solvent structure and hammerhead ribozyme catalysis. *Chem. Biol.*, 15:332–342, 2008.

[133] Tai-Sung Lee, Carlos Silva-Lopez, Monika Martick, William G. Scott, and Darrin M. York. Insight into the role of Mg^{2+} in hammerhead ribozyme catalysis from x-ray crystallography and molecular dynamics simulation. *J. Chem. Theory Comput.*, 3:325–327, 2007.

[134] Tai-Sung Lee, Carlos Silva Lopez, George M. Giambasu, Monika Martick, William G. Scott, and Darrin M. York. Role of Mg^{2+} in Hammerhead ribozyme catalysis from molecular simulation. *J. Am. Chem. Soc.*, 130(10):3053–3064, 2008.

[135] Tai-Sung Lee and Darrin M. York. Origin of mutational effects at the C3 and G8 positions on Hammerhead Ribozyme catalysis from molecular dynamics simulations. *J. Am. Chem. Soc.*, 130(23):7168–7169, 2008.

[136] Tai-Sung Lee, George M Giambau, Carlos P Sosa, Monika Martick, William G Scott, and Darrin M York. Threshold occupancy and specific cation binding modes in the hammerhead ribozyme active site are required for active conformation. *J Mol Biol*, 388:195–206, Mar 2009.

[137] William G. Scott. Morphing the minimal and full-length hammerhead ribozymes: implications for the cleavage mechanism. *Biol. Chem.*, 388:727–735, 2007.

[138] James B. Murray, Attila A. Seyhan, Nils G. Walter, John M. Burke, and William G. Scott. The hammerhead, hairpin and VS ribozymes are catalytically proficient in monovalent cations alone. *Chem. Biol.*, 5:587–595, 1998.

[139] Edward A. Curtis and David P. Bartel. The hammerhead cleavage reaction in monovalent cations. *RNA*, 7:546–552, 2001.

[140] Jessica L. O'Rear, Shenglong Wang, Andrew L. Feig, Leonid Beigelman, Olke C. Uhlenbeck, and Daniel Herschlag. Comparison of the hammerhead cleavage reactions stimulated by monovalent and divalent cations. *RNA*, 7:537–545, 2001.

[141] Garrett A. Soukup and Ronald R. Breaker. Relationship between internucleotide linkage geometry and the stability of RNA. *RNA*, 5:1308–1325, 1999.

[142] Rhonda A. Torres and Thomas C. Bruice. The Mechanism of Phosphodiester Hydrolysis: Near In-line Attack Conformations in the Hammerhead Ribozyme. *J. Am. Chem. Soc.*, 122:781–791, 2000.

[143] J. MacQueen. Some methods for classification and analysis of multivariate observations. In L. M. Le Cam and J. Neyman, editors, *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability - Vol. 1*, pages 281–297. University of California Press, Berkeley, CA, USA, 1967.

[144] Jared H Davis, Trenton R Foster, Marco Tonelli, and Samuel E Butcher. Role of metal ions in the tetraloop-receptor complex as analyzed by nmr. *RNA*, 13(1):76–86, Jan 2007.

[145] D. A. Case, D. A. Pearlman, J. W. Caldweel, T. E. Cheatham III, J. Wang, W. S. Ross, C. L. Simmerling, T. A. Darden, K. M. Merz, r. V. Stanton, A. L. Cheng, J. J. Vincent, M. Crowley, V. Tsui, H. Gohlke, R. J. Radmer, Y. Duan, J. Pitera, I. Massova, G. L. Seibel, U. C. Singh, P. K. Weiner, and P. A. Kollman. *AMBER 7*. University of California San Francisco, San Francisco, 2002.

[146] H. C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, 72:2384–2393, 1980.

[147] Shuichi Nosé and M. L. Klein. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.*, 50(5):1055–1076, 1983.

[148] William G. Hoover. Canonical dynamics: Equilibration phase-space distributions. *Phys. Rev. A*, 31(3):1695–1697, 1985.

[149] Evelyn Mayaan, Adam Moser, Alexander D. MacKerell Jr., and Darrin M. York. CHARMM Force Field Parameters for Simulation of Reactive Intermediates in Native and Thio-Substituted Ribozymes. *J. Comput. Chem.*, 28:495–507, 2007.

[150] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. Charmm: a program for macromolecular energy minimization and dynamics calculations. *J. Comput. Chem.*, 4:187–217, 1983.

[151] Nicolas Foloppe and Alexander D. MacKerell, Jr. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.*, 21:86–104, 2000.

[152] Alexander D. MacKerell, Jr. and Nilesh K. Banavali. All-Atom Empirical Force Field for Nucleic Acids: II. Application to Molecular Dynamics Simulations of DNA and RNA in Solution. *J. Comput. Chem.*, 21:105–120, 2000.

[153] Kwangho Nam, Qiang Cui, Jiali Gao, and Darrin M. York. Specific reaction parametrization of the AM1/d Hamiltonian for phosphoryl transfer reactions: H, O, and P atoms. *J. Chem. Theory Comput.*, 3:486–504, 2007.

[154] Michael C. Hutter, Jeffrey R. Reimers, and Noel S. Hush. Modeling the Bacterial Photosynthetic Reaction Center. 1. Magnesium parameters for the semiempirical AM1 method developed using a genetic algorithm. *J. Phys. Chem. B*, 102:8080–8090, 1998.

[155] Jiali Gao, Patricia Amara, Cristobal Alhambra, and Martin J. Field. A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations. *J. Phys. Chem. A*, 102:4714–4721, 1998.

[156] Kwangho Nam, Jiali Gao, and Darrin M. York. An Efficient Linear-Scaling Ewald Method for Long-Range Electrostatic Interactions in Combined QM/MM Calculations. *J. Chem. Theory Comput.*, 1(1):2–13, 2005.

[157] Rita Przybilski and Christian Hammann. The tolerance to exchanges of the WatsonCrick base pair in the hammerhead ribozyme core is determined by surrounding elements . *RNA*, 13:1625–1630, 2007.

[158] Narayan Baidya and Olke C. Uhlenbeck. A kinetic and thermodynamic analysis of cleavage site mutations in the hammerhead ribozyme. *Biochemistry*, 36:1108–1114, 1997.

[159] D. E. Ruffner, G. D. Stormo, and O. C. Uhlenbeck. Sequence Requirements of the Hammerhead RNA Self-Cleavage Reaction. *Biochemistry*, 29:10695–10712, 1990.

[160] Joonhee Han and John M. Burke. Model for General Acid-Base Catalysis by the Hammerhead Ribozyme: pH-Activity Relationships of G8 and G12 Variants at the Putative Active Site. *Biochemistry*, 44:7864–7870, 2005.

[161] Jennifer A. Nelson and Olke C. Uhlenbeck. Minimal and extended hammerheads utilize a similar dynamic reaction mechanism for catalysis. *RNA*, 14(1):43–54, Jan 2008.

[162] Alex B. Burgin, Jr., Carolyn Gonzalez, Jasenka Matulic-Adamic, Alexander M. Karpeisky, Nassim Usman, James A. McSwiggen, and Leonid Beigelman. Chemically Modified Hammerhead Ribozymes with Improved Catalytic Rates. *Biochemistry*, 35:14090–14097, 1996.