CHAPTER 14

UNRAVELING THE MECHANISMS OF RIBOZYME CATALYSIS WITH MULTISCALE SIMULATIONS

TAI-SUNG LEE[1,2], GEORGE M. GIAMBAŞU[1,2], ADAM MOSER[2], KWANGHO NAM[3], CARLOS SILVA-LOPEZ[2], FRANCESCA GUERRA[2], OLALLA NIETO-FAZA[2], TIMOTHY J. GIESE[1], JIALI GAO[2], AND DARRIN M. YORK[1,2]

[1]*Biomedical Informatics and Computational Biology, University of Minnesota, Minneapolis, MN 55455, USA*

[2]*Department of Chemistry, University of Minnesota, 207 Pleasant St. SE, Minneapolis, MN 55455, USA, e-mail: york@umn.edu (D.M. York)*

[3]*Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA*

Abstract: Description of a multiscale simulation strategy we have developed to attack problems of RNA catalysis is presented. Ribozyme systems give special challenges not present in typical protein systems, and consequently demand new methods. The main methodological components are herein summarized, including the assembly of the ***QCRNA*** database, parameterization of the AM1/d-PhoT Hamiltonian, and development of new semiempirical functional forms for improved charge-dependent response properties, methods for coupling many-body exchange, correlation and dispersion into the QM/MM interaction, and generalized methods for linear-scaling electrostatics, solvation and solvent boundary potentials. Results for a series of case studies ranging from noncatalytic reaction models that compare the effect of new DFT functionals, and on catalytic RNA systems including the hairpin, hammerhead and L1 ligase ribozymes are discussed

Keywords: Ribozyme catalysis, multiscale simulation, linear-scaling method, QM/MM, DFT functional

14.1. INTRODUCTION

Over the last several decades, the original notion that the only function of RNA molecules was as messenger intermediates in the pathway from the genetic code to protein synthesis has undergone a revolution. 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 catalyze important biochemical reactions, including protein synthesis itself [1–7]. These discoveries have transformed our view of RNA

D.M. York and T.-S. Lee (eds.), Multi-scale Quantum Models for Biocatalysis, 377–408.
DOI 10.1007/978-1-4020-9956-4_14,

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 [8–10].

Molecular simulations of RNA catalysis, in principle, offer a means of accessing the most intimate mechanistic details that may aid in the interpretation of experiments and provide predictive insight into design [11]. In order to study reactions catalyzed by biological macromolecules such as RNA, simulations are performed using so-called multiscale models. Here, by "multiscale model", we mean 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 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 simplest multiscale model to study enzyme reactions would be a combined quantum mechanical/molecular mechanical (QM/MM) potential [12–18].

Simulations of RNA enzymes, or *ribozymes*, however, are laden with challenges not apparent for most protein enzymes. RNA molecules are highly negatively charged, and exhibit strong, and often specific interactions with solvent [11, 19–21]. 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 [22–24]. These issues are further complicated by the fact that RNA molecules bind divalent metal ions which 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 that in typical protein enzyme systems. The chemistry involved in reactions of prototype ribozymes such as cleavage transesterification involves 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.

14.2. COMPUTATIONAL APPROACH

This section describes the main methodological advances that will be used in subsequent selected applications, including (1) Development of fast semiempirical methods for multiscale quantum simulations, (2) Directions for development of next-generation QM/MM models, and (3) Linear-scaling electrostatic and generalized solvent boundary methods.

14.2.1. Development of Fast Semiempirical Methods for Multiscale Quantum Simulations

The development of new-generation Hamiltonians is greatly benefited by the use and refinement of existing models. Conventional semiempirical Hamiltonians such as MNDO [25, 26], AM1 [27] and PM3 [28] are based on a minimal *s* and *p* orbital basis, and were parametrized to reproduce gas phase geometries and heats of formation for molecules in their ground states. Relatively little consideration was given to models that could accurately describe chemical reactions, especially in the case of phosphate hydrolysis where *d* orbitals are required for an accurate representation of hypervalent phosphorus transition states and intermediates. The *d*-orbital extension of the MNDO method [29, 30] greatly improved the description of hypervalent phosphorus species, but suffered from an extremely poor description of hydrogen bonding as did the original MNDO model. The empirical core-core functions of AM1 and PM3 allowed these models to reasonably predict hydrogen bonding, and if appropriately reparametrized, were demonstrated to be quite accurate for nucleic acid base pairing [31]. Consequently, it seemed that a functional form that merged the core-core functions of AM1 and PM3 with the *d*-orbital formulation of MNDO/d would offer a significantly improved base model that could be easily implemented into any *d*-orbital semiempirical program and thus be readily available to a wide scientific community. However, a prerequisite to performing such a reparameterization is to assemble a consistent set of high-level reference data that considers a wide range of properties and encompasses a broad set of molecules, complexes, potential energy surfaces and chemical reaction mechanisms. This was achieved by the construction of a database of quantum calculations for RNA catalysis (*QCRNA*) based on density-functional theory [32]. Subsequently, the *QCRNA* database was utilized to parameterize an AM1/d-PhoT [33] semiempirical Hamiltonian model for a wide range of phosphoryl transfer reactions and hydrogen bonding interactions. The *QCRNA* database and AM1/d-PhoT model are described in the following subsections, and together provide powerful tools to further the understanding of the mechanisms of ribozyme catalysis.

14.2.1.1. QCRNA*: A Database of Quantum Calculations for RNA Catalysis*

The *QCRNA* database [32, 34] is an on-line resource of ab initio data relevant to phosphorus and RNA chemistry, and was specifically designed to act as the reference

data for the parameterization of new semiempirical models and refinement of existing semiempirical Hamiltonians. The database utilizes a strict computational protocol to determine optimized molecular structures, electronic structure properties, and thermodynamic quantities, including estimates of the solvation free energy and solvent-polarization using continuum models. For brevity we describe the protocol as B3LYP/6-311++G(3df,2p)//B3LYP/6-31++G(d,p); while a detailed explanation of the precise protocol is provided in Ref. 32. One of the database's advantages is that all of the data is obtained using the same computational procedure, whereas the literature data has often been inconsistent, in that disparate theoretical protocols have been employed by different groups that severely limit meaningful quantitative cross-comparison.

There are more than 2000 molecules, 300 molecular complexes, 250 chemical reactions, and 50 potential energy surfaces in the database. The main types of information that can be found in the database include: molecular geometries, electronic structure properties, vibrational frequencies, relative conformational energies, hydrogen bond energies, metal ion binding energies, proton affinities/gas-phase basicities, tautomerization energies, and chemical mechanisms. Many of molecules in the database correspond to stationary points, i.e., local minima and maxima, of chemical mechanisms; however there also exist simple potential energy surfaces involving non-stationary geometries resulting from constrained optimization. Of the mechanisms relevant to RNA catalysis are those that involve: acyclic and cyclic phosphates, phosphate mono-, di- and triester systems, different protonation and charge states, experimentally relevant thio effects, phosphorane pseudorotation reactions, metal-catalyzed reactions, and linear free energy relations that involve different nucleophiles and leaving groups. Data contained in the *QCRNA* database has been used as the basis for numerous publications on biological phosphates and phosphoranes, and phosphoryl transfer reactions [35–46].

The *QCRNA* database is viewable and searchable with a web browser on the internet and it is also contained as a MySQL database that is easily incorporated with parameter optimization software to allow for the rapid development of specific reaction parameters. Molecular structures can be viewed with the JMOL [47, 48] or MOLDEN [49, 50] programs as viewers for chemical MIME types. If the web browser is JAVA-enabled, then the JMOL software will automatically load as a web applet. Both programs allow the structure to be manipulated, i.e., rotated, scaled, and translated, and allow for measurement of internal coordinates, e.g., bond lengths, angles, and dihedral angles. Similarly, animations of the vibrational frequencies are available and can be viewed with either program.

14.2.1.2. AM1/d-PhoT Model for Phosphoryl Transfer Reactions

The AM1/d-PhoT model [33] is a parameterization of a modified AM1/d Hamiltonian developed specifically to model phosphoryl transfer reactions catalyzed by enzymes and ribozymes for use in linear-scaling calculations and combined QM/MM simulations. The model is currently parametrized for H, O, and P atoms to reproduce

high-level density-functional results from the *QCRNA* database [32, 34], including geometries and relative energies of minima, transition states and reactive intermediates, dipole moments, proton affinities and other relevant properties. The model has been tested in the gas phase and in solution using molecular dynamics simulations with a QM/MM potential [33, 51]. The results indicate the method provides significantly higher accuracy than MNDO/d, AM1 and PM3 methods, and for the transphosphorylation reactions is in close agreement with the density-functional calculations at the B3LYP/6–311++G(3df,2p) level with a reduction in computational cost of 3–4 orders of magnitude. The model has recently been applied in QM/MM simulations of the hammerhead [52, 53] and hairpin ribozyme [54, 55] systems.

14.2.2. Directions for Development of Next-Generation QM/MM Models

14.2.2.1. Semiempirical Methods with Improved Charge-Dependent Response Properties

Standard NDDO, MNDO, and tight binding semiempirical models systematically underestimate the polarizability of molecules, and this has been attributed mainly to the use of minimal basis sets [56]. The underestimation of the polarizability is most pronounced in the pathological case where all of the spin orbitals are occupied, such as F^-, in which case a minimal basis calculation is devoid of degrees of freedom and thus lacks the variational parameters to describe polarization. Molecules with physically relevant electronic configurations rarely approach the extreme limit of total occupation of the basis; however, the polarizability of neutral molecules has been observed to be underestimated by approximately 25% when comparing semiempirical calculations with experiment [57]. A more recent and much larger comparison between MNDO/d and B3LYP/6-31++G(3df,2p) polarizabilities for 1132 molecules taken from the *QCRNA* [32] online database indicates that the semiempirical polarizability can be underestimated by approximately 40% [56]. An obvious solution for correcting the polarizabilities is to use a larger orbital basis set; however, this has the disadvantage of increasing the dimensions of the Fock matrices and requires additional semiempirical parametrization. Giese and York [56] (GY) took the alternative approach of including an auxiliary basis of density which explicitly polarizes in response to the external potential. This approach is based on the principle of chemical potential equalization (CPE), which is derivable as a second order Taylor series expansion of the density functional in density response. GY opted to incorporate the CPE model as a post-SCF correction to the semiempirical method, whereby the semiempirical SCF density is used as the reference density in CPE expansion. The auxiliary basis used to describe the density response was chosen to be atom-centered primitive Gaussian dipole functions and the integrals produced from the CPE expansion where performed using a Coulomb approximation.

GY noticed that the polarizability of an atom is approximately an exponential function of charge, and that the polarizability correction provided by the CPE expansion for an isolated atom was equal to the inverse of the Coulomb self energy of the

Gaussian dipole. From this, they developed an expression for the Gaussian exponent as a function of charge, thereby allowing for a charge dependent polarizability correction that directly results from the dilation of the response density. The resulting model (MNDO/d+CPE) was shown to reduce the error in the polarizability of the 1132 molecules down to 3.6% with a mean unsigned error of 0.1%. Furthermore, GY showed that the charge dependence form of the response basis allowed them to reproduce the polarizability of molecules in various charge states.

14.2.2.2. QM/MM Interactions with Coupled Many-Body Polarization, Exchange, and Dispersion

The most commonly applied QM/MM methods utilize a quantum mechanical model combined with an empirical force field which are coupled within the QM Hamiltonian and through a QM/MM interaction energy consisting of electrostatic, bonded, and nonbonded terms. The nonbonded QM/MM interaction is modeled by a simple empirical van der Waals interaction such as a Lennard-Jones 6–12 potential, whose empirical parameters are based on the assignment of atoms to "atom-types". This term is purely empirical and completely neglects explicit coupling to the quantum mechanical electronic degrees of freedom. The lack of explicit quantum mechanical coupling of the van der Waals QM/MM interaction energy is at the root of many problems in QM/MM modeling. Unlike MM atoms, the association of a QM atom to a particular atom-type becomes ambiguous when the QM atom undergoes changes in charge state and/or chemical bonding environment during passage through a reaction coordinate or perturbation parameter. The profiles of reactions involving highly charged species, such as those encountered in phosphate hydrolysis and phosphoryl transfer reactions, can be skewed by the solvent effect if not properly treated. The solvent effect is sensitive to the van der Waals radii used, since these dictate the degree to which solvent can approach ionic substrates.

Giese and York [58] (GY) modified their MNDO/d+CPE model [56] for QM/MM interactions to explicitly treat the charge dependence of the van der Waals forces. In their approach, called OPNQ, the van der Waals correction was applied as a post-SCF correction to the QM/MM energy; however, their approach can easily be incorporated within the SCF procedure. The OPNQ model consists of charge dependent repulsive and dispersion components. The repulsion energy is based on a density overlap model which was motivated in form from the observation that the Hartree-Fock repulsion of rare gas dimers could be reproduced from a parametrized overlap of the unperturbed atomic densities [59, 60]. In the OPNQ model, the atomic density is modeled by an atom-centered spherical Slater function whose exponent was optimized to reproduce the homonuclear dimer density overlap as a function of separation. The Slater exponent is then allowed to vary exponentially with respect to charge, and this introduces an empirical parameter to describe charge dependence. The dispersion model is a traditional multipole expression involving C_6 and C_8 dispersion coefficients which are damped in the short range by the Tang and Toennies function [61, 62]. The expressions for the dispersion coefficients were taken from the

work of Pellenq and Nicholson [63, 64] (PN), whom developed the equations from perturbation theory. The PN equations give heteronuclear dispersion coefficients using 1-body parameters only. GY modified the PN equations, which depend on such things as the dipole polarizability of the atoms, to include their charge dependence.

In their work [58], GY demonstrated that a standard Lennard-Jones model grossly over-predicted the well-depth of rare gas-halide ion dimer potential energy curves when they were parametrized to reproduce the neutral rare gas-halide dimer curves. They further showed that the OPNQ model performed just as badly when the charge dependence of the expressions were ignored, but the potential energy curves for both the neutral and ionic dimers could be simultaneously be reproduced if the charge dependence is considered.

14.2.3. Linear Scaling Electrostatic and Generalized Solvent Boundary Methods

14.2.3.1. Linear Scaling Electrostatics

Linear scaling evaluation of the energy and forces is a prerequisite to the application of the new and improved model Hamiltonians described in the previous sections to simulations of large biomolecules. There are several methodological bottlenecks inherent within ab initio methods, such as diagonalization of the Fock matrix, that prevent linear scaling. The divide-and-conquer algorithm is one method for overcoming the diagonalization problem [65, 66]; however, we here assume that the quantum region is small and relatively independent of the size of the system and that the remainder of the system is composed of molecular mechanical atoms. In this case, only the long-range electrostatic interactions pose a major obstacle to achieving linear scaling. The two main approaches for overcoming this obstacle are Ewald methods for periodic boundary simulations and fast multipole methods for stochastic boundary simulations. The theory of these approaches are fairly standard and are not repeated here. Instead we outline some of our recent contributions in these areas below.

Recently, Nam, Gao and York (NGY) [67] reported a linear scaling semiempirical QM/MM Ewald method and its incorporation into the CHARMM simulation package. In order to take advantage of the optimized Ewald algorithms traditionally used for static point charge distributions, they performed a Mulliken partitioning of the electron density and treated the QM region as a series of point charges. The QM charges are not static, however, and therefore they also developed Fock matrix corrections to obtain a self-consistent wavefunction, which is required to compute the analytic gradients. NGY also examined how much error is introduced into the gradients when the Ewald Fock matrix correction was independent of the SCF cycles.

Giese and York (GY) [68] used the branch-free FMM algorithm of Watson et al. [69] and the recursive bisection ideas of Perez-Jorda and Yang (PJY) [70] to create an adaptive FMM for systems of particles composed of point multipoles, as opposed to the trivial case of point charges (monopoles). GY spent most of their effort in

developing an adaptive termination criteria, i.e., when to stop splitting the system, and evaluating the stability of the adaptive divisioning of the system. More specifically, GY chose to split the system either by: (1) a dividing plane in which the child subsystems lie on either side of the plane, i.e., "fluc-splitting" or (2) dividing the Cartesian rectangular box enclosing the system perpendicular to its largest edge, i.e., "box-splitting". Fluc-splitting gets its name from having chosen the location and orientation of the dividing plane from the center of distribution and the eigenvector of the largest eigenvalue from the 3×3 covariance matrix of the particle distribution, respectively. GY found that box-splitting was superior to fluc-splitting because it did not exhibit errors associated with the creation of subsystem shapes that are only adequately modelled accurately with a large multipole expansion. They also found that an adaptive termination criteria can be constructed to produce near-optimal performance for systems composed of point charges or point multipoles, and for small and large systems of various shapes.

14.2.3.2. Generalized Solvation and Solvent Boundary Methods

In some instances, it is not computationally feasible to treat all of the water molecules explicitly such as in a large simulation cell with periodic boundary conditions. In such cases, recourse must be taken into alternative methods to treat the generalized solvation effects that integrate out the explicit degrees of freedom of the solvent using a continuum or linear-response approach. In this section, we briefly describe the current status of development of such methods in our group, including a smooth COSMO solvation model and a variational electrostatic projection (VEP) method for generalized solvation effects.

Smooth COSMO solvation model. We have recently extended our smooth COSMO solvation model with analytical gradients [71] to work with semiempirical QM and QM/MM methods within the CHARMM and MNDO programs [72, 73]. The method is a considerably more stable implementation of the conventional COSMO method for geometry optimizations, transition state searches and potential energy surfaces [72]. The method was applied to study dissociative phosphoryl transfer reactions [40], and native and thio-substituted transphosphorylation reactions [73] and compared with density-functional and hybrid QM/MM calculation results. The smooth COSMO method can be formulated as a linear-scaling Green's function approach [72] and was applied to ascertain the contribution of phosphate-phosphate repulsions in linear and bent-form DNA models based on the crystallographic structure of a full turn of DNA in a nucleosome core particle [74].

Variational electrostatic projection method. In some instances, the calculation of PMF profiles in multiple dimensions for complex chemical reactions might not be feasible using full periodic simulation with explicit waters and ions even with the linear-scaling QM/MM-Ewald method [67]. To remedy this, we have developed a variational electrostatic projection (VEP) method [75] to use as a generalized solvent boundary potential in QM/MM simulations with stochastic boundaries. The method is similar in spirit to that of Roux and co-workers [76–78], which has been recently

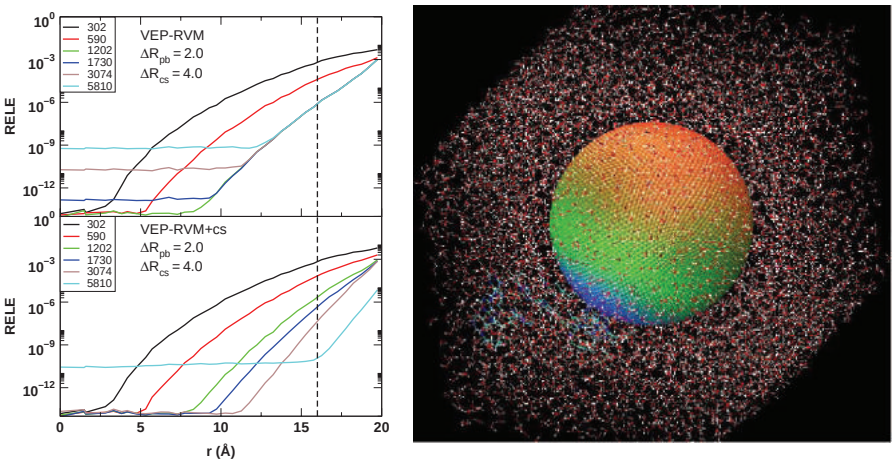


Figure 14-1. Left: Relative errors (RELE) in the force as a function of radial distance from the center of the active dynamical region for the VEP-RVM charge-scaling method [80] for the solvated hammerhead ribozyme at different discretization levels [151] of the ω surface. *Right:* The projected total electrostatic potential due to the fully solvated hammerhead ribozyme projected onto the VEP surface [80]

implemented into QM/MM simulations [79]. We have also developed a charge-scaling implementation of the method [80] that delivers high accuracy (Figure 14-1). Preliminary results suggest that the VEP method is more general and considerably more accurate than methods based on multipole expansions (Figure 14-1).

14.3. SELECTED APPLICATIONS

In this section, we describe a sampling of applications that target different facets of the problems associated with RNA catalysis. The applications range from the study of small model phosphoryl transfer reactions in solution to chemical reactions, metal ion binding and conformational events that occur in ribozyme systems. First, phosphoryl transfer reactions in solution are examined, with the underlying goal of determining reliable model chemistries that capture the essential features of the reaction profile as characterized experimentally. Second, applications to ribozyme are explored that examine different aspects of RNA catalysis. The chemical steps of catalysis are explored in the hairpin ribozyme, a prototype ribozyme that does not have an explicit catalytic divalent metal ion requirement. Next the role of divalent metal ions are explored in a similar reaction catalyzed by the hammerhead ribozyme, whereby changes in conformation and metal ion binding mode have been implicated in proceeding from reactant to transition state. Finally, a very large-scale structural rearrangement is studied in the L1 ligase riboswitch in order to explain the role of conserved residues in stabilizing conformational intermediates. These applications tie together several important factors that provide a broader understanding of the

interplay between chemical and conformational steps, and how they are affected by metal ions and other solvent components to achieve catalysis.

14.3.1. Case Study: Comparison of DFT Functionals on Model Phosphoryl Transfer Reactions

In recent years, density-functional theory has emerged as the computational quantum chemistry method of choice for biological problems of medium size range (up to a few hundreds of atoms) in applications that do not require extensive conformational sampling. The field continues to advance in the accuracy of new functionals, the improvement of algorithms and the functionality and computational performance of software [81].

In the case of the development of new density-functional exchange-correlation functionals, the current climate is one of rapid change. New functionals are being turned out at an increasingly feverish pace. This wave of new functionals makes it difficult for the community to assess their limitations and general reliability over a sufficiently broad range of chemistry. The strategy we have thus far taken is not to continually jump from one DFT functional to another as soon as a new functional appears to have made incremental improvement. Instead, we have remained largely with well-established functionals that have limitations that are well characterized and understood. Moreover, the data which we accumulate in this way has greater impact by allowing cross-comparison with other calculations such as those collected in the *QCRNA* database.

Nonetheless, we find it is important to periodically assess the state of the art and reset our existing gold standard in order to progress. Up to this point, the majority of our database calculations have been performed using the well-established hybrid three parameter exchange functional of Becke [82, 83] coupled with Lee, Yang and Parr (LYP) correlation functional [84]. This functional performs relatively well for phosphate anions, but in general predicts barriers that are systematically too low, in contrast to Hartree-Fock methods that are usually systematically too high. Recently, new insights into the origin of the current limitations of density-functional theory have been discussed [81]. One of the most important caveats that developers have to deal with to improve B3LYP and other functionals of its generation is their poor description of medium and long range correlation that give rise to intermolecular dispersion interactions. Several different approaches to improve dispersion have been proposed such as perturbationally corrected functionals [85], the addition of semiempirical correction terms [86–88], and the use of an exchange-hole dipole moment model [89]. An alternate approach is to reparametrize existing functional forms so as to better model mid-range correlation. By construction these models can not reliably predict the long-range behavior of the dispersion energy, but if appropriately reparametrized, can considerably improve short-range non-bonded interactions. This latter approach has been adopted on a grand scale in the M05 and M06 suite of density functionals [90]. Many of these functionals have been praised for their

accuracy over a wide spectrum of applications [91, 92]. The current front-runner recommended for main group chemistry is the M06-2X functional [93]. In this section we compare B3LYP and M06-2X for the transesterification of a dinucleotide reaction model, and for the pathological case of dissociation of the *p*-nitrophenyl phosphate dianion.

14.3.1.1. Transesterification of a Dinucleotide Model

The transesterification reaction is at the core of the catalytic process in several prototype ribozymes such as the hammerhead and hairpin ribozymes discussed in the next section, and thus the accurate modeling of this reaction is critically important for the study of ribozyme catalysis (Figure 14-3). This reaction involves interaction between the $2'$ alcohol nucleophile of the RNA sugar with a highly negatively charged and polarizable adjacent 3' phosphate group. While in ribozymes it is generally believed that the reaction proceeds via a general or specific acid/base catalytic mechanism with indirect and, in some cases, direct chemical involvement of nucleobase functional groups and metal ions along the reaction coordinate [94, 95].

The transesterification of phosphates in a dianionic state is a concerted transformation sporting a single transition state along the reaction path [96]. Both B3LYP and M06-2X density functionals provide a similar description in terms of energetics and geometry on the 2-D potential energy surface. When modeling reactions occurring in biomacromolecules such RNA, the stationary points are an important but incomplete set of data to analyze. Accurate modeling of ribozymes requires capturing dynamic effects along the chemical step. These effects depend not only on the reactant, transition state and product species, but also on the specific shape and curvature of the multidimensional energy surface connecting reactant and product.

The first and simplest approach to explore the shape of the potential energy surface along the transesterification process is to follow the intrinsic reaction coordinate from the transition state downhill to reactant and product (Figure 14-4). From Figure 14-4 it is clear that the curvatures near the transition states are fairly different for each functional. B3LYP shows a steeper pathway uphill from the ligated dinucleotide model than M06-2X, however, on the other side of the saddle point, the descent to the cleaved product is less pronounced for B3LYP than for M06-2X. Despite the differences in slope shown by these functionals, the curvature of the potential energy surface at the transition state is similar for both profiles, and yield similar imaginary frequencies (-153.22 and $-168.42\,\mathrm{cm}^{-1}$ for B3LYP and M06-2X, respectively). A more thorough approach to explore the potential energy surface can be taken into account by computing a two dimensional surface where the forming/breaking bond lengths are varied independently (Figure 14-5). Despite the overall qualitative similarity of the potential energy surfaces computed with B3LYP and M06-2X, there remain some quantitatively significant differences, emphasizing the need for careful selection of the density functional for chemical reactions where dynamical effects may be important to the reaction rate.

14.3.1.2. Dissociation of p*-Nitrophenyl Phosphate*

It is often convenient experimentally to utilize small molecule substrates such as modified phosphates with enhanced leaving groups as reaction models. These compounds serve as models for RNA transesterification or phosphoryl transfer of phosphate monoesters in kinases and phosphatases. One particularly useful chemical probe for mechanistic studies is the molecule *p*-nitrophenyl phosphate (pNPP). Phosphoryl transfer in pNPP can be easily followed spectrophotometrically, and allows for kinetic isotope effects to be measured at primary (bridge) and secondary (non-bridge) phosphoryl oxygen positions, as well as at the exocyclic nitro group N position [97]. Sulphuryl substitution on the phosphate oxygen has been used to investigate the kinetic and stereochemical aspects of phosphoryl transfer [97, 98].

The experimental barrier for the hydrolysis of dianionic pNPP is estimated to be 29.5 kcal/mol at 39°C, while for the thio-substituted analog *p*-nitrophenyl thiophosphate (pNPTP) an approximate value of 27.9 kcal/mol has been observed [99, 100]. However, preliminary calculations of dianionic pNPP dissociation in solution using QM/MM methods based on the AM1/d-PhoT model predict (incorrectly) a much lower barrier. The origin of the problem can be traced back to the use of the B3LYP functional used to generate the reference data from which AM1/d-PhoT was developed. Despite the fact that B3LYP has been demonstrated to be nearly as accurate as much higher level methods for prediction of *relative* proton affinity values [45], the case of *p*-nitrophenol is somewhat of an anomaly having a proton affinity value in error of −4.2 kcal/mol, suggesting it is an even more enhanced leaving group to the extent that, in solution, dissociation is nearly barrierless.

Figure 14-2 compares the potential energy curve for dianionic pNPP dissociation using B3LYP (red) and M06-2X (blue) for dianionic pNPP (top) and pNPTP (bottom). The potential energy curve for dianionic pNPP dissociation using B3LYP indicates barrierless dissociation in the gas phase, whereas with the M06-2X functional, there is a barrier of 2.3 kcal/mol. The situation is similar comparing the dianionic pNPTP dissociation where again, B3LYP predicts a kinetically insignificant barrier and M06-2X predicts a 4.6 kcal/mol barrier. In solution it is expected that the barrier will be considerably increased due to solvent stabilization of the dianionic transition state. The striking feature is that B3LYP predicts barrierless dissociation in the gas phase, whereas M06-2X predicts a stable reactant species with activation energy barriers of 2.3 and 4.6 kcal/mol. Although further investigation needs to be made of this reaction profile in solution, the present results underscore the need to continue to assess new DFT functionals for their accuracy and predictive capability in order to determine the best, affordable quantum chemistry model from which high-level reference data for phosphoryl transfer reactions can be generated.

14.3.2. Case Study: Chemical Steps of Catalysis in Hairpin Ribozyme

At first glance, it would seem that RNA enzymes, composed of fairly inert nucleobases connected by a sugar-phosphate backbone, are simply not equipped with

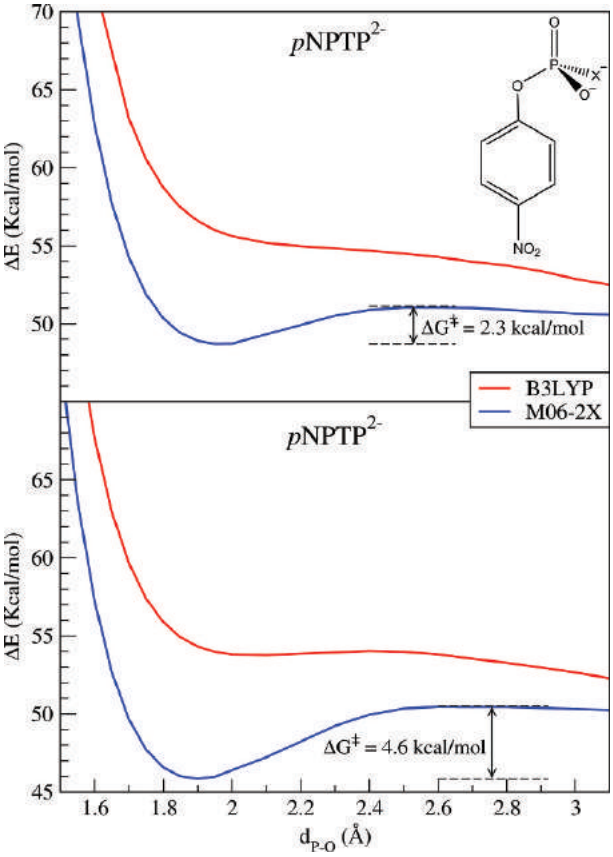


Figure 14-2. Potential energy curves (relative to separated monoanions) for the dissociation of *p*-nitrophenyl phosphate (pNPP) and *p*-nitrophenyl phosphorothioate (pNPTP) in the gas phase

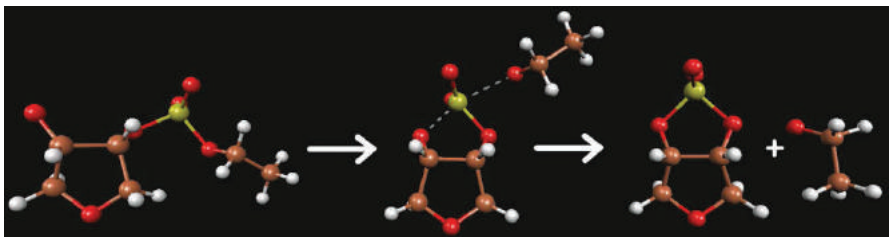


Figure 14-3. Transesterification reaction of the dinucleotide model where the nucleophile-containing ribose sugar is modelled by a tetrahydrofurane structure, whereas the cleaving sugar is further simplified and modelled as a simple primary alcohol (ethanol)

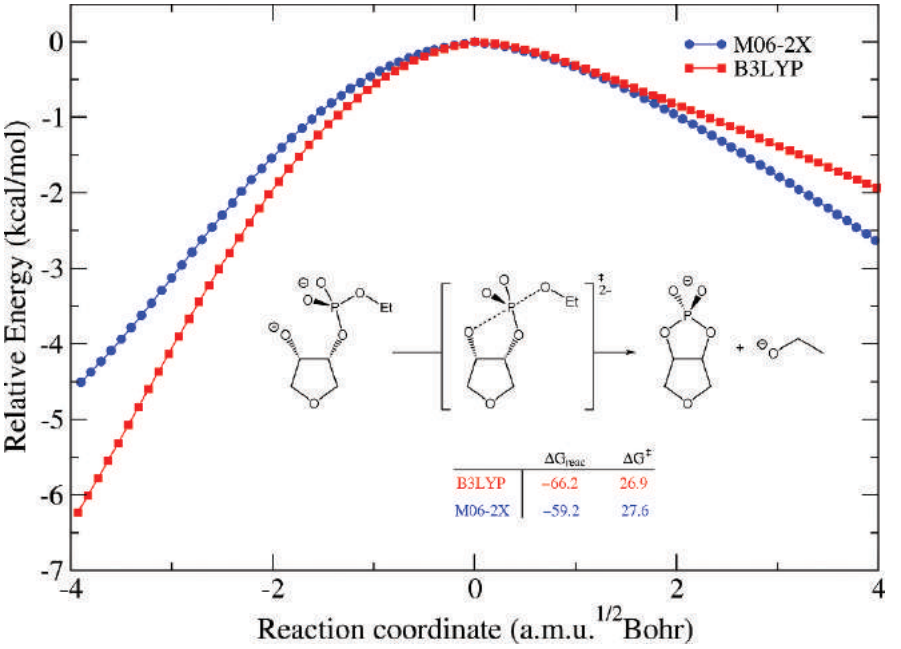


Figure 14-4. Intrinsic reaction coordinate for the transesterification of the dinucleotide model with B3LYP and M06-2X functionals. Relative free energies of reaction and activation are provided in kcal/mol

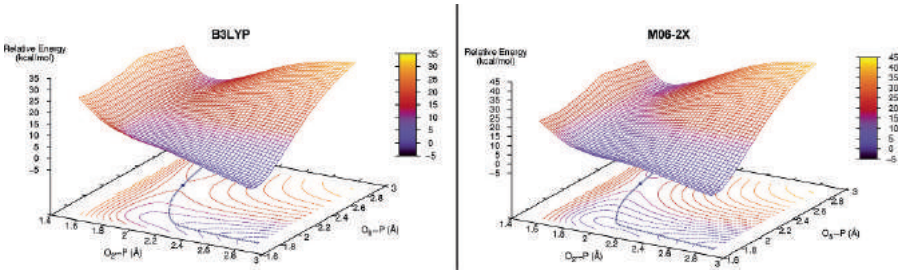


Figure 14-5. Side by side comparison of the two-dimensional potential energy surface for the transesterification reaction computed with B3LYP (*left*) and M06-2X (*right*)

an adequate array of chemical functional groups for effective catalysis. This is in stark contrast to protein enzymes that have a fairly diverse repertoire of amino acids. The central question that has gripped the community that studies RNA enzymes is, simply, by what mechanisms can these molecules achieve catalysis [101]?

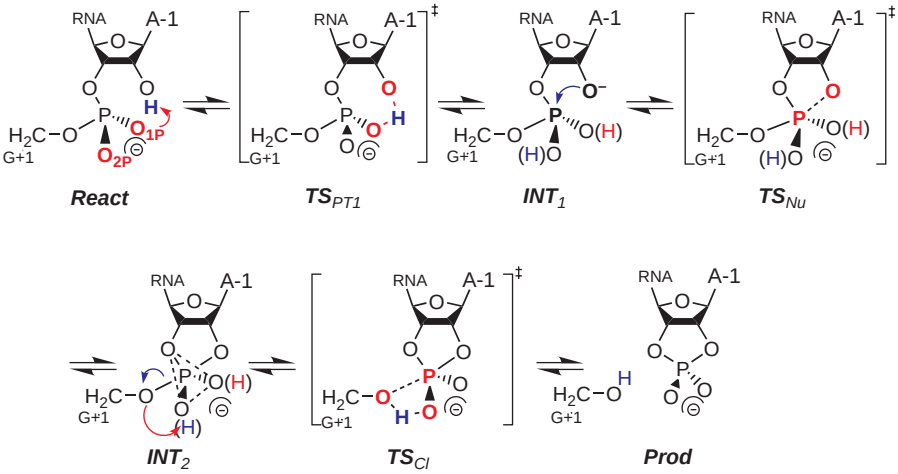
A number of factors have been implicated to be important for RNA catalysis, including the involvement of functional groups of the nucleobases or RNA backbone, divalent metal ions or other solvent components that might provide electrostatic

stabilization or act as general acid and base catalysts [102, 103]. Nonetheless, there currently exists no general consensus as to the origin of the catalytic proficiency exhibited by ribozymes, nor any detailed mechanism that has been unambiguously determined. In several small prototype systems, such as the hammerhead [104], hepatitis delta virus [105], and the L1 ligase [106] ribozymes, metal ions are essential for catalysis as well as RNA folding. The dual role played by metal ions in these systems complicates the identification of the chemical origins of catalysis and the unambiguous determination of detailed mechanism [102].

In contrast, the hairpin ribozyme (HPR) [107, 108], which catalyzes the reversible, site-specific phosphodiester bond cleavage of an RNA substrate, is unique in that the chemical steps of the reaction do not require involvement of a divalent metal ion [107–111]. This lack of an explicit metal ion requirement [112] makes the hairpin ribozyme an ideal target for theoretical studies aimed to characterize the contribution of "generalized solvation" provided by the solvated ribozyme on catalysis.

Here, we demonstrate with combined QM/MM simulations that the electrostatic environment provided by solvated HPR active site lowers the cleavage activation barrier up to 9 kcal/mol relative to the uncatalyzed transphosphorylation barrier in aqueous solution, accounting for the majority of the experimentally observed rate enhancement. Further work [54, 55] has gone on to explore in mode detailed mechanistic scenarios whereby A38 and G8 act as a general acid and base. The present results suggest that the electrostatic environment of the solvated ribozyme active site contributes significantly in achieving 10^6 to 10^7-fold rate enhancement of the phosphodiester cleavage [113–115] relative to the uncatalyzed, but spontaneous cleavage of RNA molecule in aqueous solution [116, 117]. Without the aid of a divalent metal ion, nor direct participation of nucleobase functional groups as a general acid or a general base, the majority of the observed rate enhancement can be realized through specific hydrogen bonding interactions (provided from G8 and other nucleobases) and non-specific electrostatic interactions of the solvated ribozyme active site. In the discussion that follows, the term "electrostatic solvation" is used to discuss the electrostatic component of the "generalized solvation" provided by the ribozyme environment.

The HPR in-line monoanionic mechanism considered in the present work are depicted in Scheme 14-1. This mechanism involves three reaction steps: (1) an initial intramolecular proton transfer from the 2′-hydroxyl group to either the pro-R (O_{1P}) or the pro-S (O_{2P}) non-bridging oxygen atoms of the scissile phosphate group, (2) a nucleophilic attack from the 2′-hydroxyl oxygen at the phosphate center, and (3) an exocyclic bond-cleavage of the leaving group from the phosphate center along with a second intramolecular proton transfer from the phosphate non-bridging oxygen to the leaving group. The proton transfer and nucleophilic substitution steps can occur either in a stepwise or concerted fashion. To explore these possibilities, two-dimensional reaction free energy profiles for the proton transfer and nucleophilic substitution reaction coordinates have been determined using molecular dynamics (MD) free energy simulations with a combined QM/MM potential along with density



Scheme 14-1. General in-line monoanionic mechanism of phosphodiester cleavage transesterification catalyzed by hairpin ribozyme; the first proton transfer ($PT1$), the nucleophilic attack (Nu), and the exocyclic cleavage (Cl) steps are shown, and the O_{1P} and O_{2P} pathways are indicated by *blue* and *red* colored hydrogens, respectively. For the uncatalyzed model reaction in solution, the O_{1P} and O_{2P} pathways are energetically equivalent

functional theory (DFT) corrections to the adiabatic pathways. In addition, the electrostatic solvation free energies are determined for the reactant state, intermediate states, transition states, and product states for both catalyzed and uncatalyzed reactions to address the effects of electrostatic environment provided by the ribozyme on the reaction.

14.3.2.1. Two-Dimensional QM/MM Potential of Mean Force Profiles

Simulations are based on the second transition state analog crystal structure (PDB code 1M5O) [110] and performed using CHARMM [118] (version c32a2). Stochastic boundary MD was performed in a solvated 25-Å sphere centered at the scissile phosphate in the ribozyme active site using the all-atom CHARMM27 nucleic acid force field [119] and TIP3P water model [120], with the AM1/d-PhoT quantum model [33] and GHO method [121] for treatment of the QM/MM boundary. Full details are described elsewhere [55]. Two-dimensional potential umbrella sampling MD simulations [122] were performed, from which the potential of mean force profiles were constructed using the weighted histogram analysis method [123]. The reaction coordinates consist of a nucleophilic substitution coordinate, $\zeta_1 = R(P - O_{5'}) - R(P - O_{2'})$, and a proton transfer coordinate, $\zeta_2 = R(O_{X'} - H_{2'}) - R(O_{NB} - H_{2'})$, where O_{NB} is either O_{1P} or O_{2P}, and $O_{X'}$ is $O_{2'}$ for the first (nucleophilic bond formation) step, and $O_{5'}$ for the second (leaving group bond cleavage) step. Each umbrella window was run for 17 ps of equilibration and 50 ps of configurational sampling. The uncatalyzed model reaction consisted of a molecule

of 2-hydroxyethyl methyl phosphate solvated with a 40-Å^3 cubic box of 2038 water molecules and one Na^+ ion. Simulations were carried out using QM/MM-Ewald scheme [67] at 1 atm and 300 K. The computed free energy values were further refined by density functional adiabatic energies computed at the B3LYP/6-311++G(3df,2p)//B3LYP/6-31++G(d,p) level, in which "//" separates the level for the refined single point energy from the level for geometry optimization. The same geometry optimizations were carried out at the AM1/d-PhoT level in order to derive approximate correction factors. Figure 14-6 shows the two-dimensional free energy profiles for the uncatalyzed model transphosphorylation reaction in aqueous solution, and catalyzed by the HPR along O_{1P} and O_{2P} pathways (Scheme 14-1). Table 14-1 lists free energy values corresponding to stationary points along the minimum free energy path on the surface, determined from the QM/MM free energy simulations, along with corrections at the DFT level to the activation and reaction free energy values.

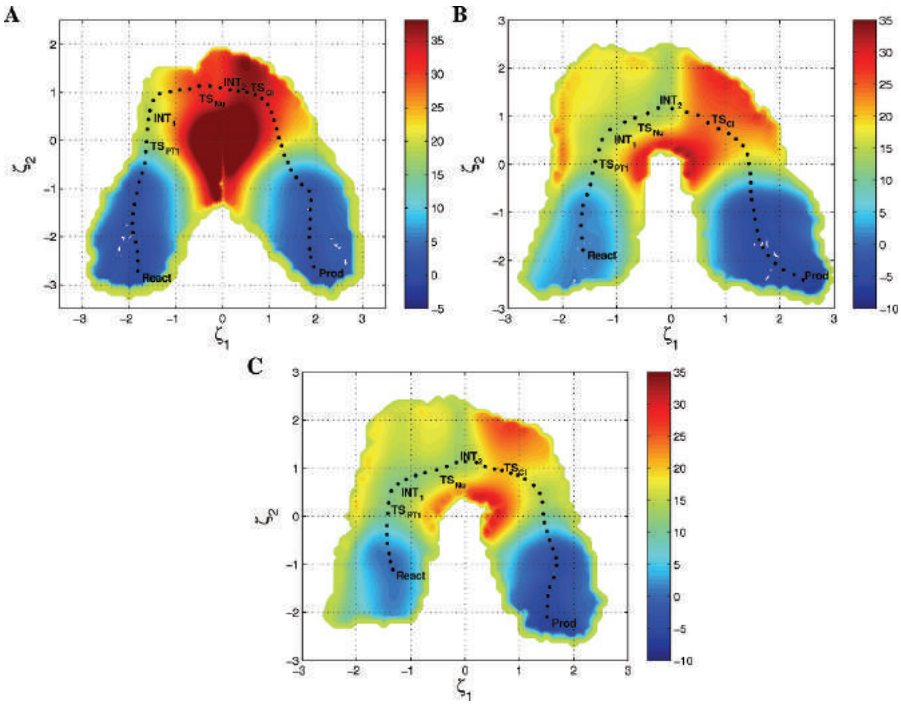


Figure 14-6. Two-dimensional free energy surfaces for in-line monoanionic mechanisms for the (**A**) uncatalytic model reaction in solution, and the catalytic (**B**) O_{1P} and (**C**) O_{2P} pathways in the hairpin ribozyme. ζ_1 is defined as $R_{P-O_{5'}} - R_{O_{2'}-P}$, and ζ_2 is $R_{O_{2'}-H_{2'}} - R_{O_{NB}-H_{2'}}$ for $\zeta_1 < 0.0$ Å and $R_{O_{5'}-H_{2'}} - R_{O_{NB}-H_{2'}}$ for $\zeta_1 > 0.0$ Å, where O_{NB} is for the O_{1P} proton transfer in (**B**), and for the O_{2P} proton transfer in (**A**) and (**C**), respectively. The units for free energies and distances are kcal/mol and Å, respectively

Table 14-1. Calculated reaction free energies and barrier heights (kcal/mol) for uncatalyzed model and catalyzed transesterification reactions in solution and in the hairpin ribozyme[a]

	GB[b]	TS_{PT1}	INT_1	TS_{Nu}[c]	INT_2	TS_{Cl}	Prod
Soln[d]	O_{2P}	19	19	32	31	37	0
				(33)	(33)	(38)	(2)
HPR[e]	O_{1P}	15	14	15	13	25	–7
				(16)	(16)	(27)	(–6)
	O_{2P}	12	11	14	13	21	–5
				(18)	(17)	(21)	(–4)
Expt	Soln[f]		21	32	25	34	
	HPR[g]					∼20–21	∼1

[a]The values are those estimated from the 2-D PMF profiles described in the text and given in Figure 14-6. A DFT correction (in parenthesis) of the semiempirical AM1/d-PhoT model is applied to the intermediate, product, and transition states, respectively, based on active site model calculations.
[b]General base (GB) activating the $O_{2'}$ nucleophile.
[c]A DFT correction on the TS_{Nu} is based on the error of AM1/d-PhoT model at the TS_{PT1+Nu}, in which the nucleophilic attack (TS_{Nu}) is concerted with the proton transfer (TS_{PT1}) in the gas phase.
[d]Uncatalyzed model reaction in solution.
[e]Catalyzed reaction in the hairpin ribozyme.
[f]Values are estimated from references [152] and [153], which combine experimental and computational values for the reaction free energies and activation energies of relevant reactions in solution.
[g]Experimental values for the hairpin ribozyme are taken from references [113–115].

14.3.2.2. Active Site Structure and Mechanism

Figure 14-7 shows representative snapshots of the transition states for the nucleophilic substitution and exocyclic cleavage steps. The overall reaction may be characterized by a sequence of proton transfer, nucleophilic attack, exocyclic cleavage, and proton transfer steps. The rate-limiting step is the exocyclic bond-cleavage of the leaving group from the phosphorus atom, followed by a barrierless proton transfer to the departing $O_{5'}$ alkoxide anion. The same trend is found both for the catalyzed reaction in the ribozyme and the uncatalyzed model reaction in aqueous solution, but the hairpin ribozyme markedly lowers the reaction barriers for each of the three reaction steps and the free energies of the resulting intermediates. The greatest barrier reduction occurs in the nucleophilic attacking step (a net decrease of 17 kcal/mol for the O_{1P} pathway and 18 kcal/mol for the O_{2P} pathway relative to the uncatalyzed reaction). For the rate-limiting step, the free energy barriers are lowered by 12 and 16 kcal/mol along the O_{1P} and the O_{2P} pathways, respectively, while experimental estimation of barrier reduction is 13–14 kcal/mol. After the density functional correction at the B3LYP/6-311++G(3df,2p) level, the overall free energy barrier becomes 27 kcal/mol for the O_{1P} pathway and 21 kcal/mol for the O_{2P} pathway. These results are in accord with the experimental estimate of 20–21 kcal/mol [113–115]. Nonetheless, this does not preclude alternate mechanisms with explicit nucleobase involvement, such as A38 and G8 acting as a general acid and a general base catalyst [110, 124, 125], that could further lower the barrier.

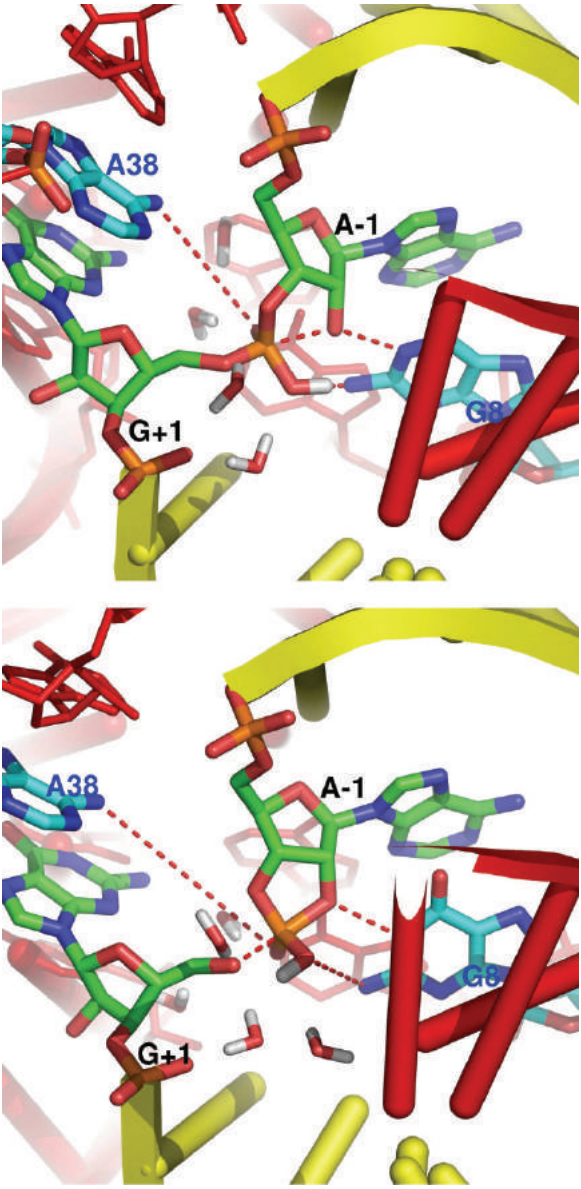


Figure 14-7. Snapshots of the active site structures near the transition state of (*top*) the nucleophilic attack and (*bottom*) the exocyclic cleavage for the in-line monoanionic O_{2P} mechanism of cleavage transesterification in the hairpin ribozyme. The yellow and red colored cartoon is for the substrate and ribozyme strands, respectively, and water molecules interacting with non-bridging oxygens and $O_{5'}$ are shown

These results explore the effects by which the change of the electrostatic environment provided by the hairpin ribozyme relative to that of aqueous solution affects the rate of the transphosphorylation reaction. Since the specific mechanisms explored here do not involve direct intervention of any nucleobases as a general base or general acid in the catalysis, the computed change in the free energy barriers is mostly due to the change of the heterogeneous electrostatic environment in the HPR active site relative to that of bulk solvation by water. The direct electrostatic solvation by the ribozyme and water lowers the overall free energy barrier by 7 and 9 kcal/mol for the two reaction paths corresponding to an initial proton transfer to either of the two non-bridging phosphate oxygen atoms. The results suggest that the non-specific interactions in HPR are sufficient to account for the majority of the observed change of barrier heights without the involvement of a metal ion and general acid-base catalysis by active site nucleobases. The in-line monoanionic mechanism establishes a baseline mechanism that invokes only the generalized solvation and specific hydrogen bonding interactions provided by the ribozyme environment, and provides a departure point for the exploration of alternate mechanisms where participation of nucleobases in the active site play an active chemical role.

14.3.3. Case Study: Role of Divalent Metal Ions in Hammerhead Ribozyme Catalysis

The hammerhead ribozyme (HHR) catalyzes the same type of transesterification reaction as the hairpin ribozyme, which involve the site-specific attack of an activated 2′OH nucleophile to the adjacent 3′ phosphate, resulting in cleavage of the P-O5′ phosphodiester linkage to form a 2′,3′ cyclic phosphate and a 5′ alcohol. However, unlike the hairpin ribozyme that has no divalent metal ion requirement for catalysis, the hammerhead ribozyme, under physiological conditions, requires divalent metal ions to promote its catalytic step [7, 104]. Recent crystallographic studies of a full length HHR have characterized the ground state active site architecture [126] and its solvent structure [127], including the binding mode of a presumed catalytically active divalent metal ion in the active site. These findings have reconciled a long-standing controversy between structural and biochemical studies for this system [128]. It is still not clear, however, what are the roles of the divalent ion in the active site. Recently, large scale molecular dynamics simulations using both molecular mechanics and hybrid QM/MM potentials have been used to explore the structure and dynamics at different states along the catalytic pathway in order to shed light on the possible role of divalent ions in the catalytic mechanism [52, 53].

14.3.3.1. HHR Folds to Form an Electrostatic Negative Metal Ions Recruiting Pocket

One of the recent crystallographic structures of the full length HHR identifies five well-defined divalent ion binding sites, one of them (the C-site) being located in the catalytic pocket and being suggested to have a direct role in catalysis [127].

In order to probe cation occupation in the active site in the absence of Mg^{2+} ions, we examined Na^{+} distributions in the reactant and activated precursor (deprotonated 2OH' nucleophile) states. It has been noted in the recent literature that the modeling of ions in highly charged systems such as HHR affords tremendous challenges with regard to simulation time scales [129]. This section presents the results of series of five 300 ns simulations of the full length HHR, in both the reactant and activated precursor states, in order to ascertain the cation occupation requirement of the active site to maintain catalytic integrity.

The 3D density contour maps for the Na^{+} ion distribution determined over the last 250 ns of simulation (Figure 14-8) show that the overall highest probability Na^{+} occupation sites were concentrated in the active site for both the reactant and activated precursor. This suggests that the HHR folds to form a strong local electronegative pocket that is able to attract and bind Mg^{2+} if present in solution, or recruit a high local concentrations of Na^{+} ions in the absence of Mg^{2+}.

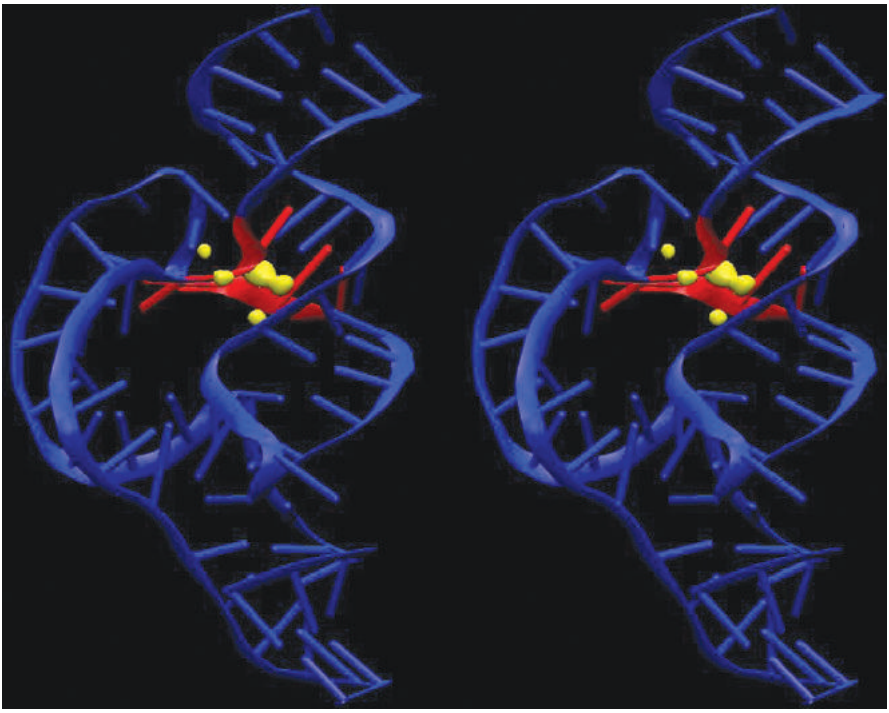


Figure 14-8. The 3D density contour maps (*yellow*) of Na^{+} ion distributions derived from the activated precursor simulation. The hammerhead ribozyme is shown in blue with the active site in red. Only the high-density contour is shown here to indicate the electrostatic recruiting pocket formed in the active site

14.3.3.2. *The Bridging Mg^{2+} Induces a Significant pK_a Shift of the General Acid*

The first published crystal structure of the full length HHR [126] in which there was no solvent or ions resolved showed A9 and the scissile phosphate in close proximity, consistent with the interpretation of thio effect measurements [130], and the G8:$O_{2'}$ and G12:N_1 poised to act as a general acid and base, respectively, as proved in previous photocrosslinking [131] and mutation experiments [132]. Given the strong evidence that Mg^{2+} participates directly in the catalytic process together with the spatial proximity of the A9 and scissile phosphate, made the placement of an Mg^{2+} ion in bridging position a reasonable assumption.

We have explored the role the Mg^{2+} ion placed at the bridging position in the reactant state, the early transition state (ETS), and the late transition state (LTS) [52, 53]. In these studies we have used specifically designed molecular mechanics residues that are able to reproduce the geometry and charge distribution in the early and late transition states of the phosphoryl transfer reaction [46]. The Mg^{2+} ion remained in the bridging position for the entire duration of the three simulations, displaying different potential roles at specific points of the reaction pathway. In the ETS simulation (Figure 14-9, left panel), the Mg^{2+} ion is directly coordinating G8:$O_{2'}$ to induce a possible shift in its pK_a to act as a general acid, while in the LTS

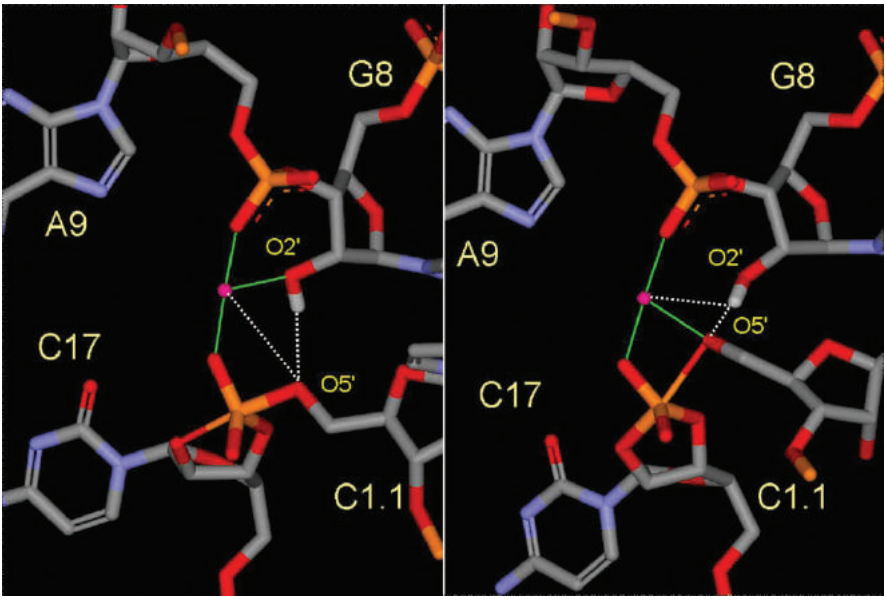


Figure 14-9. Snapshots from the simulations of the early transition state mimic (*left*) and the late transition state mimic (*right*), indicating the Mg^{2+} ion direct coordination (*green lines*) and key hydrogen bonds and indirect Mg^{2+} coordination (*dotted lines*). For clarity, the water molecules are not shown

simulation (Figure 14-9, right panel), the Mg^{2+} ion acts as a potential Lewis acid catalyst to stabilize the leaving group, being poised to assist proton transfer from the G8:$O_{2'}$.

Hybrid QM/MM simulations were performed to further probe the role of the bridging Mg^{2+} ion [52]. The spontaneous proton transfer from the implicated general acid, G8:$O_{2'}$, to the leaving group, C1.1:$O_{5'}$ was observed within the first ns of QM/MM simulation, confirming our assumptions about the role of the bridging Mg^{2+} ion in the catalytic step of HHR. Thus, our simulation results supported the supposition that a single bridging Mg^{2+} ion could assist in the cleavage step in HHR catalysis by acting to increase the acidity of the 2OH' of G8. The Mg^{2+} at the bridging position also preserves the integrity of the active site structure, and may serve as an epicenter in the transition state that coordinates the A9 and scissile phosphates, G8:$O_{2'}$ general acid and C1.1:$O_{5'}$ leaving group.

14.3.3.3. The Accumulation of the Negative Charge in the Precursor State Causes the Migration of the Mg^{2+} Ion from the C-Site to the Bridging Position

Recently, a joint experimental/theoretical study has been reported of the full length hammerhead structure with resolved solvent and metal ions [127]. In this structure a resolved Mn^{2+} ion in the active site was not positioned in a bridging position as postulated in our previous simulations; instead, it binded G10.1:N_7 and A9:O_{2P} (the C-site). We performed simulations with a Mg^{2+} initially placed at the C-site for different stages along the reaction pathway [52]. The results, shown in Figure 14-10, suggest that the Mg^{2+} in fact migrates from the C-site to the bridging position in the transition states and the deprotonated reactant state. This migration is caused by the accumulated negative charge at the cleavage site after the general base step.

In summary, our simulation results draw a possible picture of the roles of Mg^{2+} in supporting the catalytic step of HHR. First, HHR folds to form an electronegative cation recruiting pocket that attracts a Mg^{2+} ion to the C-site. The ion moves to the bridging position between A9 and the scissile phosphate either upon deprotonation of the 2OH' nucleophile, or formation of the dianionic transition state. In this position, the Mg^{2+} ion is poised to provide direct electrostatic stabilization of the transition state and the accumulating negative charge on the leaving group. Moreover, the Mg^{2+} shifts the pK_a of the general acid (G8:$O_{2'}$), and after the proton transfer to the leaving group, reverts back to stabilize the conjugate base. This mechanistic interpretation is supported by the present simulations, and is consistent with a considerable body of experimental work. First, the thio/rescue effect experiments [130] support a mechanism in which a single metal cation bound at the C-site in the ground state acquires an additional interaction with the scissile phosphate in proceeding to the transition state. Second, kinetic studies [133], photocrosslinking experiments [131] and mutational data [132, 134, 135] implicate G8 and G12 as possible general acid and base. Lastly, recent studies involving metal ion titrations suggest that the pK_a of the general acid is down- shifted by around 4–7 pK_a units in a metal-dependent

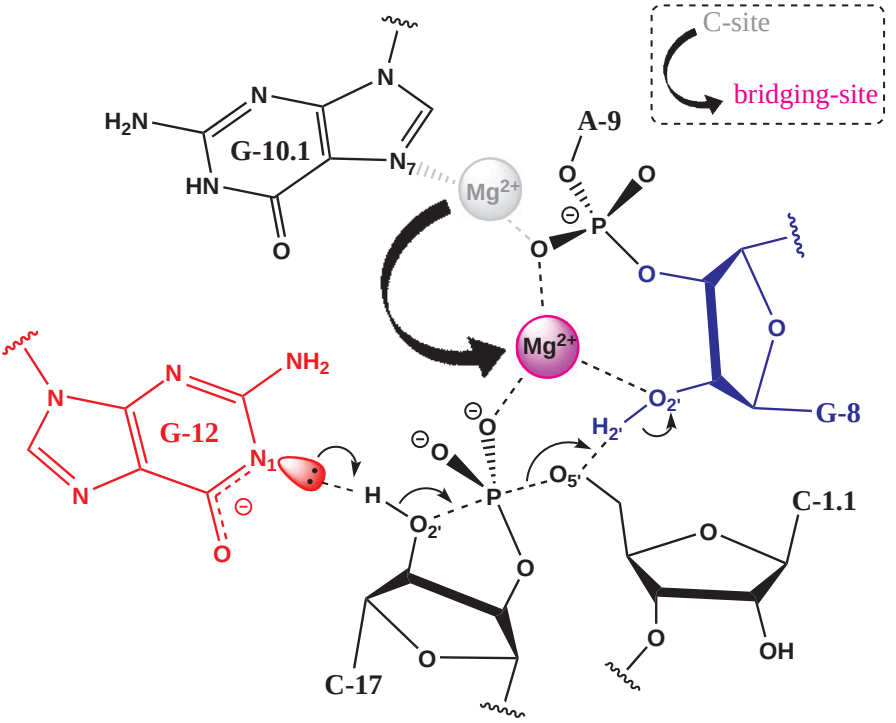


Figure 14-10. A schematic view of the possible migration of the Mg^{2+} ion from the C-site to the bridging position. Spontaneous migration was predicted from the simulation for the transition states and the deprotonated reactant state, with the Mg^{2+} ion initially placed at the C-site

manner, correlated with the metal pK_a [94], and indicate that divalent metals may play a specific chemical role in catalysis [136].

These results represent an important first step in the detailed characterization of the structure, dynamics and free energy profile for the full length HHR catalytic mechanism. Together with experiment, it is the hope that a consensus will emerge that explains the detailed molecular mechanisms of hammerhead ribozyme catalysis, and in doing so may provide new insight into the guiding principles that govern RNA catalysis.

14.3.4. Case Study: Conformational Transition in the L1 Ligase Ribozyme

RNA is characterized by a large and diverse ensemble of conformations that interchange on time scales that range from femtoseconds to microseconds [137]. This conformational variability allows RNA molecules to be designed to allow binding and catalytic activity (to be allosterically controlled) such as in the case of aptamers and aptazymes [138], or have the ability to regulate gene expression by binding small

molecules such as in the case of riboswitches [139]. 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 important conformations, and their time scales, that might play important roles in function [140]. 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.

In this section, we report molecular simulations on the large-scale conformational transition of the L1 ligase ribozyme from an inactive to a catalytically active state. L1 Ligase ribozyme functions as an *allosteric molecular switch* or *aptazyme* and was iteratively optimized by *in vitro* selection to catalyze regioselectively and regiospecifically the $5'$ to $3'$ phosphodiester bond ligation (nucleotidyl transfer reaction) with the possibility to be controlled (activated) by small molecules, peptides and even proteins [141–144]. An unique feature of L1 ligase is its intrinsically flexible non-canonically base paired ligation site, a characteristic possessed only by two other ligase ribozymes [145, 146]. There is no naturally occurring ligase ribozyme and moreover, among the synthesized ligases ribozymes there are only five that accomplish their function in an regiospecific and regioselective way [141, 146–149].

L1 ligase's large intrinsic flexibility was revealed by the recent crystal structure of the ligation product of a reduced size variant with two vastly different conformers, differing by reorientation of one of the stems by around 80 Å, that were resolved in the same asymmetric cell [106]. Based on the presence/absence of specific contacts between distant conserved parts including the ligation site and a totally conserved residue, U38, one of the conformers was postulated to represent the catalytically active or *on* conformation, the other the inactive or *off* one [106].

We have explored the coupled on–off conformational switch in L1 ligase using large scale molecular dynamics simulations for more than 600ns departing from both active and inactive conformations (Figure 14-11). Based on the crystal structure of the two conformers, we identified a limited 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, They were denoted as θ_{18}, θ_{37}, θ_{44}, and η_{38}, and were defined following Ref. [150]. These virtual torsions span two conserved and restricted regions located in the three-way junction and a loop that contains U38, the conserved residue that is postulated to be responsible for allosteric control of the catalytic step [106, 141–144].

The conformational rearrangement in the three-way junction and the U38 loop transition can both be mapped by monitoring the four virtual torsions, and occur on different time scales. The U38 loop transition occurs on the order of tens of nanoseconds, whereas the rearrangement of the three way junction is estimated to occur on a time scale longer than 0.4 μs. On this time scale of our simulations (several hundred of ns) the L1 ligase in its inactive conformation was predicted to cover approximately a third of the complete 80 Å conformational switch (Figure 14-11). Since this transition might correspond to the rate-controlling step of L1 ligase catalysis, it

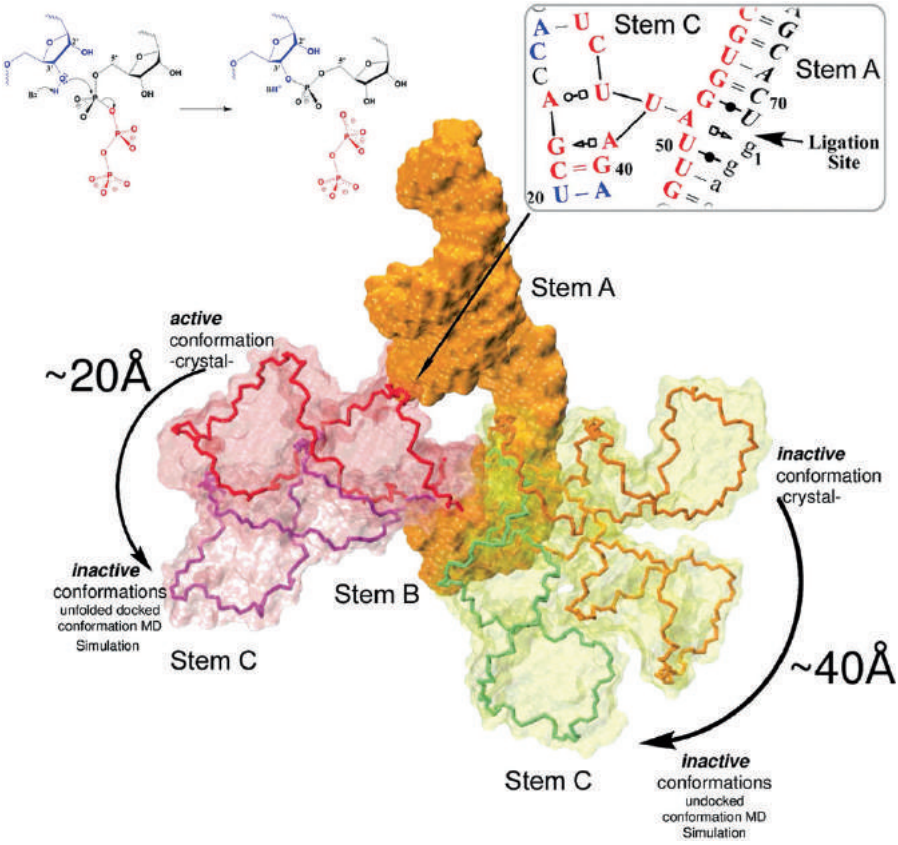


Figure 14-11. Snapshots from the conformational switch path explored in the vicinity of the active conformation (unfolded docked conformation) and starting from the inactive (undocked) conformation found in crystal. Stems A and B were aligned (best fitted) and are shown in *yellow*, different instances of stem C are shown in stick representation with different colors. A schematic of the non-canonical binding scheme of the ligation site is shown in the *right panel* and the general mechanism of ligation in the *left panel*

is of interest to identify the dynamic hinge points and stabilization of conformational intermediates that allow the transition to occur.

The overall fluctuations in the active conformation of the reactant and product states were restricted to a reduced portion of the available conformations due to distant tertiary contacts between U38 loop and the ligation site (RMSD $\sim$ 4.2 Å). In the absence of these tertiary contacts, the fluctuations are significantly larger (RMSD $\sim$ 6.7 Å) in the vicinity of the active conformation or $\sim$ 7.9 Å in the inactive form). The origin of these large variations were traced to fluctuation of the restricted region of the junction, with all the other structural elements remaining close to their starting structures. The large fluctuations observed in the simulations were accompanied by

the formation of new contacts not observed in any of the crystal structure between two conserved portions of the L1 Ligase: stem B and U19. Given their conserved nature we postulate that these contacts have to play an important role in stabilization of intermediary states along the *conformational switch – catalytic step* pathway.

The non-canonically base-paired ligation site shows a high degree of variability, and visits three distinct conformational states characterized by specific hydrogen bonding patterns between GTP1 on one side and G2:U50, U38:A51 and G52:U71 base pairs. The ligation reaction takes place between GTP1 and U71, and simulations were performed from two different initial arrangements of the reactant state differing in the conformation of the GTP1 triphosphate conformation and its ion coordination. Simulations of the ligation site predicted formation of conformational states where the U71:$O_{3'}$ atom, the nucleophile, makes close contacts to a potential general base, GTP1:$O_{2\alpha}$. The formation of these contacts was highly correlated with ligation site being in either the first or the third hydrogen binding pattern and absent when visiting the second one, suggesting that the L1 ligase catalysis might be facilitated by these specific hydrogen bonding patterns which are a direct result of the non-canonically base-paired ligation site and the intrinsic flexibility of the molecule.

The present simulation results have identified important 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 these simulations are a first step toward a detailed understanding of the coupled catalytic/conformational riboswitch mechanism of L1 ligase that may ultimately enhance the future design and engineering of new catalytic riboswitches.

14.4. CONCLUSION

In this chapter, we present a description of a multiscale simulation strategy we have developed to attack problems of RNA catalysis. Ribozyme systems, due to the high degree of charge, strong interaction with ions and other solvent components, and large conformational variations that are coupled with the chemical steps of catalysis, present special challenges not present in typical protein systems, and consequently demand new methods. The main methodological components are herein summarized, including the assembly of the *QCRNA* database, parametrization of the AM1/d-PhoT Hamiltonian, and development of new semiempirical functional forms for improved charge-dependent response properties, methods for coupling many-body exchange, correlation and dispersion into the QM/MM interaction, and generalized methods for linear-scaling electrostatics, solvation and solvent boundary potentials. We then present results for a series of case studies ranging from non-catalytic reaction models that compare the effect of new DFT functionals, and on catalytic RNA systems including the hairpin, hammerhead and L1 ligase ribozymes. The ultimate goal of this work is to develop new multiscale computational tools and bring them to bear on the study of the mechanisms of RNA catalysis. The results may serve to aid in the interpretation of experiments, provide a deeper understanding of

ribozyme mechanisms, and unravel guiding principles for RNA catalysis that may facilitate the design of new technology.

ACKNOWLEDGMENTS

We are grateful for financial support provided by the National Institutes of Health (GM62248 and GM084149 to DY; GM46736 to JG). GMG is grateful for financial support from the University of Minnesota Training Program in Bioinformatics and Computational Biology (BICB). This work was supported in part by the University of Minnesota Biomedical Informatics and Computational Biology program (TL and DY) and by a generous allocation on an IBM Blue Gene BG/P with 4,096 850 MHz CPUs at the IBM Advanced Client Technology Center in Rochester, Minnesota with further thanks to Cindy Mestad, Steven Westerbeck, and Geoffrey Costigan for technical assistance. This research was also performed in part using the Molecular Science Computing Facility (MSCF) in the William R. Wiley Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the U.S. Department of Energy's Office of Biological and Environmental Research and located at the Pacific Northwest National Laboratory, operated for the Department of Energy by Battelle. Computational resources from The Minnesota Supercomputing Institute for Advanced Computational Research (MSI) were utilized in this work.

REFERENCES

1. Gilbert W (1918) Nature 319:618.
2. Scott WG (1996) Biochem Soc Trans 24:604.
3. Gesteland RF, Cech TR, Atkins JF (1999) The RNA World: The nature of modern RNA suggests a Prebiotic RNA, 2nd ed., Cold Spring Harbor Laboratory Press, New York.
4. Yarus M (1999) Curr Opin Chem Biol 3:260.
5. Chen X, Li N, Ellington AD (2007) Chem Biodivers 4:633.
6. Lilley DM (2008) Ribozymes and RNA Catalysis, chap. The Hairpin and Varkud Satellite Ribozymes, 66–91, RSC Biomolecular Series, RSC Publishing, Cambridge.
7. Scott WG (2007) Curr Opin Struct Biol 13:280.
8. Rubenstein M, Tsui R, Guinan P (2004) Drugs of the Future 29:893.
9. Vaish NK, Dong F, Andrews L, Schweppe RE, Ahn NG, Blatt L, Seiwert SD (2002) Nature Biotech 20:810.
10. Breaker RR (2002) Curr Opin Biotechnol 13:31.
11. Biopolymers (2007) Biopolymers 85:169.
12. Gao J, Furlani TR (1995) IEEE Comput Sci Eng 2:24.
13. Gao J (1995) Rev Comput Chem 7:119.
14. Hawkins GD, Zhu T, Li J, Chambers CC, Giesen DJ, Liotard DA, Cramer CJ, Truhlar DG (1998) Combined quantum mechanical and molecular mechanical methods, 201–219, ACS Symposium Series 712, Oxford University Press, New York.
15. Monard G, Merz KM, Jr (1999) Acc Chem Res 32:904.
16. Warshel A (2002) Acc Chem Res 35:385.
17. Warshel A (2003) Annu Rev Biophys Biomol Struct 32:425.

18. Senn HM, Thiel W (2007) Curr Opin Chem Biol 11:182.
19. Norberg J, Nilsson L (2002) Acc Chem Res 35:465.
20. Orozco M, Pérez A, Noy A, Luque FJ (2003) Chem Soc Rev 32:350.
21. Orozco M, Noy A, Prez A (2008) Curr Opin Struct Biol 18:185.
22. TE Cheatham III (2004) Curr Opin Struct Biol 14:360.
23. Chen S-J (2008) Annu Rev Biophys 37:197.
24. Auffinger P, Hashem Y (2007) Curr Opin Struct Biol 17:325.
25. Dewar MJ, Thiel W (1977) J Am Chem Soc 99:4899.
26. Dewar MJ, Thiel W (1977) J Am Chem Soc 99:4907.
27. Dewar MJS, Zoebisch E, Healy EF, Stewart JJP (1985) J Am Chem Soc 107:3902.
28. Stewart JJP (1989) J Comput Chem 10:209.
29. Thiel W, Voityuk AA (1992) Theor Chim Acta 81:391.
30. Thiel W, Voityuk AA (1996) J Phys Chem 100:616.
31. Giese TJ, Sherer EC, Cramer CJ, York DM (2005) J Chem Theory Comput 1:1275.
32. Giese TJ, Gregersen BA, Liu Y, Nam K, Mayaan E, Moser A, Range K, O Nieto Faza, C Silva Lopez, A Rodriguez de Lera, Schaftenaar G, Lopez X, Lee T, Karypis G, York DM (2006) J Mol Graph Model 25:423.
33. Nam K, Cui Q, Gao J, York DM (2007) J Chem Theory Comput 3:486.
34. QCRNA, http://theory.chem.umn.edu/Database/QCRNA.
35. Range K, McGrath MJ, Lopez X, York DM (2004) J Am Chem Soc 126:1654.
36. Mayaan E, Range K, York DM (2004) J Biol Inorg Chem 9:807.
37. CS López, Faza ON, Gregersen BA, Lopez X, AR de Lera, York DM (2004) Chem Phys Chem 5:1045.
38. Range K, Riccardi D, Cui Q, Elstner M, York DM (2005) Phys Chem Chem Phys 7:3070.
39. Liu Y, Gregersen BA, Lopez X, York DM (2005) J Phys Chem B 109:19987.
40. Xu D, Guo H, Liu Y, York DM (2005) J Phys Chem B 109:13827.
41. CS López, Faza ON, AR de Lera, York DM (2005) Chem Eur J 11:2081.
42. Liu Y, Lopez X, York DM (2005) Chem Commun 31:3909.
43. Liu Y, Gregersen BA, Hengge A, York DM (2006) Biochemistry 45:10043.
44. Lopez X, Dejaegere A, Leclerc F, York DM, Karplus M (2006) J Phys Chem B 110:11525.
45. Range K, CS López, Moser A, York DM (2006) J Phys Chem A 110:791.
46. Mayaan E, Moser A, Mackerell AD Jr, York DM (2007) J Comput Chem 28:495.
47. Jmol, http://www.jmol.org.
48. Willighagen E, Howard M (2005) CDK News 2:17.
49. Schaftenaar G, Noordik JH, Molden, http://www.cmbi.ru.nl/molden/molden.html.
50. Schaftenaar G, Noordik JH (2000) J Comput.-Aided Mol Des 14:123.
51. Nam K, Gao J, York DM (2008) Multiscale simulation methods for nanomaterials, Ross RB, Sanat M (eds) Wiley, New York, pp 201–218.
52. Lee T-S, Silva-Lopez C, Martick M, Scott WG, York DM (2007) J Chem Theory Comput 3:325.
53. Lee T-S, Silva Lopez C, Giambasu GM, Martick M, Scott WG, York DM (2008) J Am Chem Soc 130:3053.
54. Nam K, Gao J, York DM (2008) J Am Chem Soc 130:4680.
55. Nam K, Gao J, York D (2008) RNA 14:1501.
56. Giese TJ, York DM (2005) J Chem Phys 123:164108.
57. Matsuzawa N, Dixon DA (1992) J Phys Chem 96:6232.
58. Giese TJ, York DM (2007) J Chem Phys 127:194101.
59. Wheatley RJ, Price SL (1990) Mol Phys 69:507.

60. Piquemal J, Cisneros G, Reinhardt P, Gresh N, Darden TA (2006) J Chem Phys 124:104101.
61. Tang KT, Toennies JP (2003) J Chem Phys 118:4976.
62. Tang KT, Toennies JP (1984) J Chem Phys 80:3726.
63. Pellenq R, Nicholson D (1998) Mol Phys 95:549.
64. Pellenq R, Nicholson D (1999) Mol Phys 96:1001.
65. Yang W (1991) Phys Rev A 44:7823.
66. Khandogin J, Hu A, York DM (2000) J Comput Chem 21:1562.
67. Nam K, Gao J, York DM (2005) J Chem Theory Comput 1:2.
68. Giese TJ, York DM (2008) J Comput Chem 29:1895.
69. Watson MA, P Sałek, Macak P, Helgaker T (2004) J Chem Phys 121:2915.
70. Pérez-Jordá JM, Yang W (1995) Chem Phys Lett 247:484.
71. York DM, Karplus M (1999) J Phys Chem A 103:11060.
72. Khandogin J, Gregersen BA, Thiel W, York DM (2005) J Phys Chem B 109:9799.
73. Gregersen BA, Khandogin J, Thiel W, York DM (2005) J Phys Chem B 109:9810.
74. Range K, Mayaan E, LJ Maher III, York DM (2005) Nucleic Acids Res 33:1257.
75. Gregersen BA, York DM (2005) J Phys Chem B 109:536.
76. Roux B, Beglov D, Im W (1999) Simulation and theory of electrostatic interations in solution, Pratt LR, Hummer G (eds) vol. 492 of Proceedings of the Santa Fe Workshop on Treatment of Electrostatic Interactions in Computer Simulations of Condensed Media, AIP Conference Proceedings, Melville, New York, pp 492–509.
77. Im W, Bernèche S, Roux B (2001) J Chem Phys 114:2924.
78. Banavali NK, Im W, Roux B (2002) J Chem Phys 117:7381.
79. Schaefer P, Riccardi D, Cui Q (2005) J Chem Phys 123:014905.
80. Gregersen BA, York DM (2006) J Comput Chem 27:103.
81. Cohen AJ, P Mori-Sánchez, Yang W (2008) Science 321:792.
82. Becke AD (1988) Phys Rev A 38:3098.
83. Becke AD (1993) J Chem Phys 98:5648.
84. Lee C, Yang W, Parr RG (1988) Phys Rev B 37:785.
85. Grimme S (2006) J Chem Phys 124:034108.
86. Grimme S (2004) J Comput Chem 25:1463.
87. Grimme S (2006) J Comput Chem 27:1787.
88. Schwabe T, Grimme S (2007) Phys Chem Chem Phys 9:3397.
89. Becke AD, Johnson ER (2006) J Chem Phys 124:014104.
90. Zhao Y, Truhlar DG (2008) Theor Chem Acc 120:215.
91. Gu J, Wang J, Leszczynski J, Xie Y, Schaefer HF III (2008) Chem Phys Lett 459:164.
92. Cramer CJ, Gour JR, Kinal A, Wloch M, Piecuch P, Shahi ARM, Gagliardi L (2008) J Phys Chem A 112:3754.
93. Zhao Y, Truhlar DG (2008) Acc Chem Res 41:157.
94. M Roychowdhury-Saha, Burke DH (2006) RNA 12:1846.
95. Thomas JM, Perrin DM (2008) J Am Chem Soc 130:15467.
96. Perreault DM, Anslyn EV (1997) Angew Chem Int Ed 36:432.
97. Hengge AC (2002) Acc Chem Res 35:105.
98. Hengge AC, Cleland WW (1990) J Am Chem Soc 112:7421.
99. Kirby AJ, Jencks WP (1965) J Am Chem Soc 87:3209.
100. Catrina IE, Hengge AC (1999) J Am Chem Soc 121:2156.
101. Takagi Y, Ikeda Y, Taira K (2004) Top Curr Chem 232:213.
102. Lönnberg T, Lönnberg H (2005) Curr Opin Chem Biol 9:665.

103. Sigel RK, Pyle AM (2007) Chem Rev 2007:97.
104. Scott WG (1999) Q Rev Biophys 32:241.
105. Shih I-H, Been MD (2002) Annu Rev Biochem 71:887.
106. Robertson MP, Scott WG (2007) Science 315:1549.
107. Walter NG, Burke JM (1998) Curr Opin Chem Biol 2:24.
108. Lilley DM (1999) Curr Opin Struct Biol 9:330.
109. Doherty EA, Doudna JA (2001) Annu Rev Biophys Biomol Struct 30:457.
110. Rupert PB, Massey AP, Sigurdsson ST, AR Ferré-D'Amaré (2002) Science 298:1421.
111. Bevilacqua PC (2003) Biochemistry 42:2259.
112. Fedor MJ, Williamson JR (2006) Nat Rev Mol Cell Biol 6:399.
113. Nesbitt SM, Erlacher HA, Fedor MJ (1999) J Mol Biol 289:1009.
114. Fedor MJ (2000) J Mol Biol 297:269.
115. Kuzmin YI, Da Costa CP, Cottrell JW, Fedor MJ (2005) J Mol Biol 349:989.
116. Hertel KJ, Peracchi A, Uhlenbeck OC, Herschlag D (1997) Proc Natl Acad Sci USA 94:8497.
117. Li Y, Breaker RR (1999) J Am Chem Soc 121:5364.
118. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) J Comput Chem 4:187.
119. Foloppe N, MacKerell AD, Jr. (2000) J Comput Chem 21:86.
120. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) J Chem Phys 79:926.
121. Gao J, Amara P, Alhambra C, Field MJ (1998) J Phys Chem A 102:4714.
122. Torrie GM, Valleau JP (1977) J Comput Phys 23:187.
123. Kumar S, Bouzida D, Swendsen R, Kollman P, Rosenberg J (1992) J Comput Chem 13:1011.
124. AR Ferré-D'Amaré (2004) Biopolymers 73:71.
125. Cottrell JW, Kuzmin YI, Fedor MJ (2007) J Biol Chem 282:13498.
126. Martick M, Scott WG (2006) Cell 126:309.
127. Martick M, Lee T-S, York DM, Scott WG (2008) Chem Biol 15:332.
128. Scott WG (2007) Biol Chem 388:727.
129. Ponomarev SY, Thayer KM, Beveridge DL (2004) Proc Natl Acad Sci USA 101:14771.
130. Wang S, Karbstein K, Peracchi A, Beigelman L, Herschlag D (1999) Biochemistry 38:14363.
131. Lambert D, Heckman JE, Burke JM (2006) Biochemistry 45:7140.
132. Blount KF, Uhlenbeck OC (2005) Annu Rev Biophys Biomol Struct 34:415.
133. Han J, Burke JM (2005) Biochemistry 44:7864.
134. McKay DB (1996) RNA 2:395.
135. Wedekind JE, McKay DB (1998) Annu Rev Biophys Biomol Struct 27:475.
136. M Roychowdhury-Saha, Burke DH (2007) RNA 13:841.
137. Crothers D (2001) RNA, chap. RNA Conformational Dynamics, Elsevier Science & Technology, Amsterdam, pp 61–71.
138. Bunka DHJ, Stockley PG (2006) Nat Rev Microbiol 4:588.
139. Schwalbe H, Buck J, Frtig B, Noeske J, Whnert J (2007) Angew Chem Int Ed 46:1212. URL http://dx.doi.org/10.1002/anie.200604163.
140. Xia T (2008) Curr Opin Chem Biol 2:1.
141. Robertson MP, Ellington AD (1999) Nature Biotech 17:62.
142. Robertson MP, Knudsen SM, Ellington AD (2004) RNA 10:114.
143. Robertson MP, Ellington AD (2000) Nucleic Acids Res 28:1751.
144. Robertson MP, Ellington AD (2001) Nature Biotech 19:650.
145. Landweber LF, Pokrovskaya ID (1999) Proc Natl Acad Sci USA 96:173.
146. Ekland EH, Szostak JW, Bartel DP (1995) Science 269:364.
147. Rogers J, Joyce GF (1999) Nature 402:323.

148. Jaeger L, Wright MC, Joyce GF (1999) Proc Natl Acad Sci USA 96:14712.
149. Ikawa Y, Tsuda K, Matsumura S, Inoue T (2004) Proc Natl Acad Sci USA 101:13750.
150. Duarte CM, Pyle AM (1998) J Mol Biol 284:1465.
151. Gregersen BA, York DM (2005) J Chem Phys 122:194110.
152. Wolfenden R, Ridgway C, Young G (1998) J Am Chem Soc 120:833.
153. Glennon TM, Warshel A (1998) J Am Chem Soc 120:10234.