Catalytic mechanism and pH dependence of a methyltransferase ribozyme (MTR1) from computational enzymology

Erika McCarthy [ID][1]**, Şölen Ekesan** [ID][1]**, Timothy J. Giese**[1]**, Timothy J. Wilson**[2]**, Jie Deng**[3,4]**, Lin Huang** [ID][3,4]**, David M.J. Lilley** [ID][2] **and Darrin M. York** [ID][1,*]

[1]Laboratory for Biomolecular Simulation Research, Institute for Quantitative Biomedicine and Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ 08854, USA, [2]Nucleic Acid Structure Research Group, MSI/WTB Complex, The University of Dundee, Dundee, Dow Street, Dundee DD1 5EH, UK, [3]Guangdong Provincial Key Laboratory of Malignant Tumor Epigenetics and Gene Regulation, Guangdong–Hong Kong Joint Laboratory for RNA Medicine, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou 510120, China and [4]Medical Research Center, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou 510120, China

ABSTRACT

A methyltransferase ribozyme (MTR1) was selected *in vitro* to catalyze alkyl transfer from exogenous O^6-methylguanine (O^6mG) to a target adenine N1, and recently, high-resolution crystal structures have become available. We use a combination of classical molecular dynamics, *ab initio* quantum mechanical/molecular mechanical (QM/MM) and alchemical free energy (AFE) simulations to elucidate the atomic-level solution mechanism of MTR1. Simulations identify an active reactant state involving protonation of C10 that hydrogen bonds with O^6mG:N1. The deduced mechanism involves a stepwise mechanism with two transition states corresponding to proton transfer from C10:N3 to O^6mG:N1 and rate-controlling methyl transfer (19.4 kcal·mol^{-1} barrier). AFE simulations predict the pK_a for C10 to be 6.3, close to the experimental apparent pK_a of 6.2, further implicating it as a critical general acid. The intrinsic rate derived from QM/MM simulations, together with pK_a calculations, enables us to predict an activity–pH profile that agrees well with experiment. The insights gained provide further support for a putative RNA world and establish new design principles for RNA-based biochemical tools.

GRAPHICAL ABSTRACT

$$k_{int} = \kappa \frac{k_B T}{h} e^{\frac{-\Delta G^{\ddagger}}{k_B T}}$$
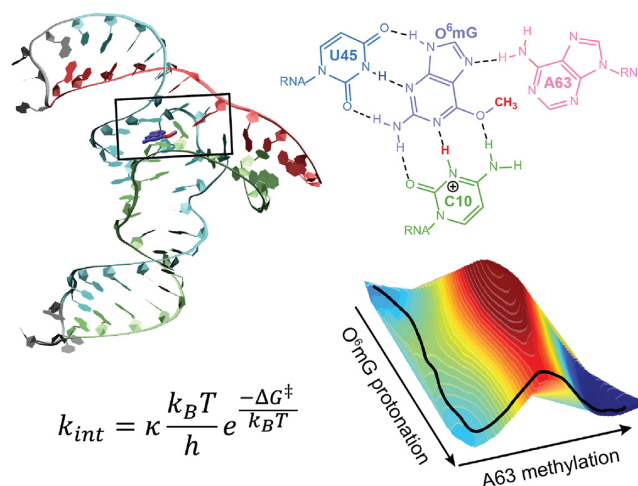
INTRODUCTION

Small self-cleaving ribozymes use unique strategies to maximize the reactivity of their limited chemical library of four nucleobases in order to catalyze phosphoryl transfer reactions (1–5). In the context of the RNA world hypothesis (6,7), these reactions are potentially evolutionary vestiges of chemical functions that have since been overtaken by proteins, begging the question of whether ancient RNAs once had the ability to catalyze more diverse reactions as precursors to modern protein counterparts (8–10), and how their mechanistic strategies may be reflected in modern biology. Evaluating this catalytic range is critical to rationalize evolutionary theories and may also shed light onto relationships between ribozyme classes (3,11), and between ribozymes and riboswitches (12–15).

*To whom correspondence should be addressed. Tel: +1 848 445 5199; Email: Darrin.York@rutgers.edu

Of particular interest from an evolutionary perspective is the ability of RNA molecules to catalyze C–C and C–N bond formation that would be essential for nucleic acid synthesis and early metabolic transformations. *In vitro* selection methods have resulted in the generation of ribozymes that catalyze diverse reactions with potential applications in gene therapy and as biosensors (16,17), including Diels–Alder reaction (18,19), aldol condensation (20), alkylation (16) and Michael addition (21); however, the lack of available structural information has made these reactions difficult to interpret and exploit. Enabled by these experimental data, computational work has explored the mechanism (22) and magnesium-dependent active site conformational selection (23) in the Diels–Alderase ribozyme. Pioneering work by Wilson and Szostak (16) identified the first self-alkylating ribozyme that catalyzed N–C bond formation in a self-biotinylation reaction, which has only recently been structurally characterized (24,25). In addition, guanine N7 methylation has been exhibited by an artificially engineered SAM-dependent methyltransferase ribozyme (SMRZ-1) (26), and cytosine N3 methylation has been demonstrated to some degree in the prequeuosine[1] (preQ$_1$) riboswitch (27).

Recently, a methyltransferase ribozyme (MTR1) has evolved *in vitro* (28) that binds O^6-methylguanine (O^6mG) and catalyzes the methylation of a target adenine (A63) at the N1 position. N^1-Methyladenosine modification has structural and functional roles in tRNA (29) and rRNA (30) as it modifies the Watson–Crick face and carries a positive charge under physiological conditions (31). Crystal structures for MTR1 have been determined (32,33) that, together with biochemical experiments, have provided insights into catalytic mechanism. These studies open the door for rigorous theoretical work aimed at providing an atomic-level interpretation of the current body of experimental data, and ultimately gaining deep insight into mechanism and the factors that regulate activity. Of key importance is the elucidation of the rate-controlling transition state ensemble that, while not directly accessible by experiments, can be investigated computationally (34). Such a predictive-level understanding is instrumental for the design of new ribozyme-based biotechnology and therapeutics, and may further shed light onto inherent differences between engineered ribozymes that have evolved rapidly *in vitro* under artificial selection pressures and biological ribozymes that have evolved gradually *in vivo* through natural selection.

Herein, we employ a computational enzymology approach (35,36) to study the catalytic mechanism of the MTR1 ribozyme. To our knowledge, this is the first in-depth computational mechanistic study of an artificially engineered alkyl transfer RNA enzyme. Molecular dynamics (MD) simulations are used to characterize the structure and dynamics of the ribozyme in solution and identify the dynamical active state ensemble (i.e. the conformation and protonation state capable of catalyzing the chemical reaction). Departing from the active state, combined quantum mechanical/molecular mechanical (QM/MM) simulations are used to explore the chemical mechanism and identify the free energy pathway and transition state ensembles. Alchemical free energy (AFE) simulations are used to examine the electrostatic environment in the active site and predict

pK_a shifts of key nucleotides that plausibly give rise to the experimentally observed activity–pH profiles. Our results suggest a remarkably sophisticated mechanism involving general acid catalysis by an active site cytosine nucleotide (C10), as was suggested by experimental work (32,33), facilitated by electrostatic interactions (5) and dynamical motions (37,38) in the active site. Taken together, the results of this work may provide new insights that facilitate nucleic acid enzyme design efforts and help further our fundamental understanding of the catalytic chemical landscape of ribozymes.

MATERIALS AND METHODS

Here, we provide a summary of the main calculations performed in the present work, the full technical details for which are described in the 'Detailed Computational Methods' section of the Supplementary Data. Departing from the crystal structure of MTR1 with a bound guanine ligand and methylated target adenine [PDB ID 7V9E (32)], we construct a solution-phase model of the wild-type (WT) ribozyme in its reactant (O^6mG-bound) state and performed MM simulations to predict the dynamical ensemble in solution. These MM simulations also provide a starting point from which to identify a plausible 'active state' in solution capable of catalyzing the chemical reaction. The active state is usually a rare state that must satisfy certain geometrical and protonation state requirements (37). In the present case, this involves alignment of the nucleophile (A63:N1), electrophilic carbon (O^6mG:Cm) and leaving group (O^6mG:O6), as well as appropriate protonation state of the implicated general acid (C10:N3).

The system was prepared by solvating the MTR1 [PDB ID 7V9E (32)] crystal structure coordinates in a truncated octahedron (90.2 Å real-space lattice vector lengths) with 18,250 TIP4P-Ew water molecules (39), 133 Na$^+$ ions and 47 Cl$^-$ ions to neutralize the ribozyme charge and achieve a physiological ion concentration of 140 mM NaCl. We used the monovalent ion parameters designed for use with the TIP4P-Ew water model (40). Electrostatics were performed with the particle mesh Ewald method (41,42). We performed five independent 150 ns MD simulations of the full MTR1 (32) using the ff99OL3 (43,44) RNA force field. Solute coordinate heavy-atom root-mean-square deviation (RMSD) with respect to the starting structure was analyzed to determine a single cutoff time that encapsulated the structurally unequilibrated part of the trajectories (50 ns), after which the RMSD was observed to fluctuate stably and analysis could be made over the last 100 ns for each independent simulation (see Supplementary Table S1 and Supplementary Figure S1).

Next, departing from the identified active state ensemble in solution, we performed combined QM/MM simulations with long-ranged electrostatics treated rigorously with the linear-scaling ambient-potential composite Ewald method (45). QM/MM umbrella sampling simulations were used to explore the mechanism of the adenine N1 methylation reaction in the WT ribozyme, and analyzed using a generalized multidimensional variational free energy profile method (46). The reactant state is characterized by a protonated C10 at the N3 position and a methylated (Cm) ligand

oxygen O^6mG:O6. The product state is characterized by a protonated ligand nitrogen O^6mG:N1 and a methylated A63 at the N1 position. The mechanism is explored with two reaction coordinates that represent the proton transfer (ξ_{PT}) and methyl transfer (ξ_{MT}):

$$\xi_{PT} = |R_{C10:N3} - R_H| - |R_{O^6mG:N1} - R_H|, \quad (1)$$

$$\xi_{MT} = |R_{O^6mG:O6} - R_{Cm}| - |R_{A63:N1} - R_{Cm}|. \quad (2)$$

To get an overview of the two-dimensional (2D) free energy surfaces (FES) and enumerate plausible mechanistic pathways, we performed QM/MM umbrella sampling with the DFTB3 (47–49) semi-empirical Hamiltonian using the 3OB-3-1 parameter set (50). The simulations were chosen to form a regular grid of 390 simulations that spanned the ranges $\xi_{PT} \in [-1.5, 1.5]$ and $\xi_{MT} \in [-2.5, 2.5]$ in steps of 0.2 Å. The minimum free energy pathways of the WT and C10U mutant ribozymes were both found to correspond to a stepwise mechanism in which proton transfer precedes the methyl transfer. Alternative pathways in which the order of these steps was reversed or occurred in concert were also identified, but with considerably higher barriers. We thus performed much more computationally intensive PBE0/6-31G* QM/MM umbrella sampling of the reactions using the umbrella sampling finite temperature string method (51) to obtain a more accurate quantitative estimate of the minimum free energy path and rate-limiting transition state barrier.

Finally, we estimate the activity–pH profile of the WT ribozyme using a two-state, noncooperative model (36,52):

$$k_{obs} = k_{int} \cdot f_{obs}(pH) = k_{int}/(1 + 10^{pK_{a,B} - pH}$$
$$+ 10^{pH - pK_{a,A}} + 10^{pK_{a,B} - pK_{a,A}}), \quad (3)$$

where k_{obs} is the observed rate constant, k_{int} is the intrinsic rate constant as shown in Equation (S1) and $f_{obs}(pH)$ is the pH-dependent fraction (probability) the ribozyme is in the active protonation state and written out as the inverse of the partition function (shown in parentheses as the denominator) enumerating the protonation states of the acid (A) and base (B) that affect the activity–pH profile. Specifically, $pK_{a,A}$ and $pK_{a,B}$ refer to the apparent pK_a of the acid and base, respectively. In the present context, the acid is presumed to be the C10 nucleotide. There is no general base catalyst in the reaction; nonetheless, a protonation event of a functionally-important base was identified from the experimental activity–pH profiles through the apparent pK_a values. In the present work, our calculations suggest a plausible candidate for this base apparent pK_a to be A63.

The value of $pK_{a,A}$ (and analogously $pK_{a,B}$) is estimated by correcting the experimental $pK_{a,A}^{expt.}$ in solution with a $\Delta pK_{a,A}$ shift caused by the ribozyme environment:

$$pK_{a,A} = pK_{a,A}^{expt.} + \Delta pK_{a,A}. \quad (4)$$

The $\Delta pK_{a,A}$ shift is calculated from AFE simulations (53,54), evaluated from the difference in free energies associated with deprotonating the C10 nucleotide in the ribozyme and aqueous environments:

$$\Delta pK_{a,A} = \frac{\Delta G_{RNA}(AH^+ \to A) - \Delta G_{aq}(AH^+ \to A)}{RT \ln(10)}. \quad (5)$$

The pK_a values of cytosine and adenine in solution have been experimentally determined to be 4.2 (55) and 3.5 (56), respectively. The recently developed smoothstep softcore potentials (57) were used to create intermediate states that connect the protonated and deprotonated systems, and production statistics were gathered for 5 ns/state using an alchemical enhanced sampling (54,58) approach. Full details of the MM and AFE simulations and other computational methods are provided in the Supplementary Data.

RESULTS AND DISCUSSION

Structure and dynamics of MTR1 in solution

The first goal of the present work is to characterize the dynamical ensemble of the active state of MTR1 in solution, i.e. the substrate-bound state that is competent to carry out the catalytic chemical steps of the reaction. Our simulations depart from the 2.3 Å resolution MTR1 structure 7V9E (32) (the highest resolution structure currently available). The electron density map indicates the presence of a methyl group at the N1 position of A63, suggesting the ribozyme is active in the crystal, or during the crystallization process, and the resolved structure represents the product form. In constructing a model of the substrate-bound reactant state, we considered that in the crystal the N1 of guanine and N3 of C10 are separated by 2.7 Å, indicating a potential hydrogen bond interaction, and consistent with the supposition that a proton was possibly transferred from C10H$^+$ to O^6mG. We thus considered this as a plausible active state model and performed a series of five independent simulations, the last 100 ns of which were shown to be equilibrated and used for analysis (Supplementary Table S1 and Supplementary Figure S1), departing from the crystal structure but with C10:N3 protonated and the methyl group connected to O^6mG as expected for the reactant state. As will be shown below, this protonation in the active state is supported by activation barriers and pK_a shifts predicted from QM/MM and AFE simulations, respectively.

Key results from the MD simulations are summarized in Figure 1. The MTR1 structure [PBD ID 7V9E (32)] is a three-way junction comprising the three helical arms P1, P2 and P3, formed from a single RNA strand connected by addition of a GNRA tetraloop to the end of the P3 helix (Figure 1A and B). The active site is formed at the three-way junction and has a core comprised of three functionally-important nucleotides that bind the O^6mG substrate through hydrogen bond interactions (Figure 1E): C10, U45 and A63. C10 and A63, as will be examined in more detail below, also play chemical roles in catalysis. C10 is stabilized by base stacking with A9 and C11, and A63 is stabilized by base stacking with A40. The O^6mG binding pocket is sandwiched from above and below by planes formed by the A9·A46·A40 base triple and by the C11·G41 Watson–Crick base pair.

The average structures from the simulations are similar to the crystal structure (32) (not including the transferred
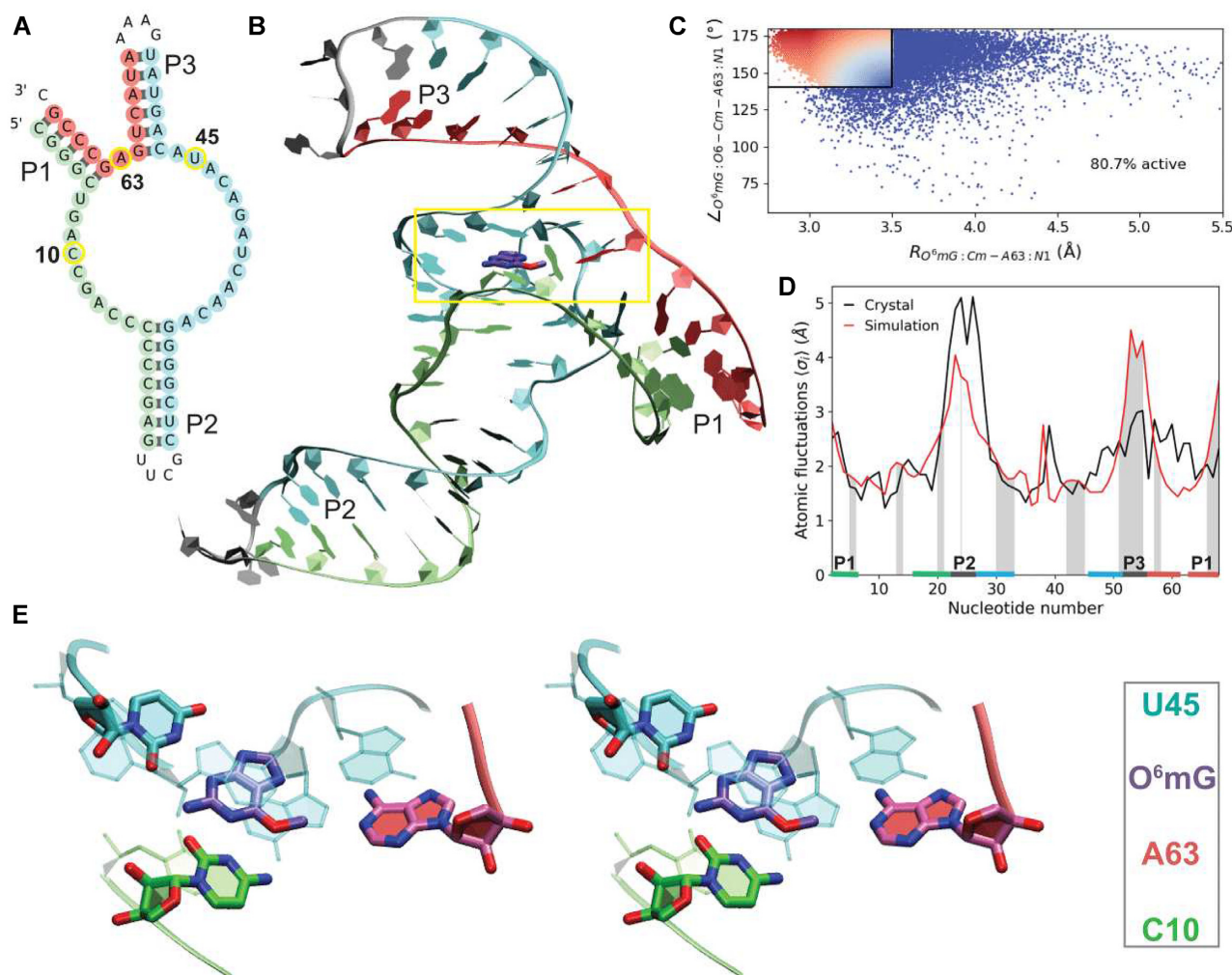
Figure 1. Active structure of MTR1 in solution. (**A**) Secondary structure of MTR1 colored by helical regions P1, P2 and P3 with loop regions in gray. (**B**) Average solution structure over the last 10 ns of simulation with O^6mG (purple) bound. (**C**) Distribution of nucleophile–electrophile angle (O^6mG:O6–Cm–A63:N1) and distance (O^6mG:Cm–A63:N1) colored from least catalytically fit (blue) to most catalytically fit (red). Fitness scores were assigned such that the angle must be between 140° and 180° and distance between 3.50 and 2.75 Å (as denoted by the upper left box), otherwise a score of zero (blue) was given (see Supplementary Data for a detailed explanation of scoring). (**D**) Atomic fluctuations averaged per nucleotide, iterating over i nucleotides ($\langle\sigma_i\rangle$), derived from analysis of 5×100 ns MD simulations (red) and compared with fluctuations estimated from the crystallographic *B*-values (black) with correlation coefficient of 0.62. Helical arms and loop regions are highlighted on the horizontal axis, and shaded regions indicate crystal contacts. (**E**) Stereo view of the active site nucleotides of MTR1 in solution with coloring corresponding to panels (A) and (B).

methyl group), particularly for the functionally-important active site nucleotides with RMSD values ranging from 0.63 to 0.88 Å (Supplementary Table S1). Only one simulation (sim 1) has an overall RMSD >2.9 Å from the crystal structure and appears to be somewhat of an outlier. Figure 1D compares atomic fluctuations (59) ($\langle\sigma_i\rangle$, averaged for each nucleotide i) from the solution simulations and derived from the crystallographic data. The fluctuations are generally similar, with the largest discrepancies in the P1 (nucleotides 65–69) and P3 regions where stabilizing crystal contacts (indicated by shading) are not present in solution. The dynamical ensemble in solution has the largest fluctuations in the P1 (3′ and 5′ paired termini), P2 (loop) and P3 (loop) regions due to lack of base pairing and/or stacking. The ability of RNA force fields to model loop conformations [e.g. hairpins (60,61) and tetraloops (62–64)] is an ongoing challenge. Given the number of crystal contacts

in the P2 and P3 loops in the experimental structure, it is hard to determine whether RNA force field issues might also be a contributing factor to the large fluctuations observed in these regions in solution. Overall, these analyses suggest that the active site structure is preserved, there are no significant conformational changes that occur in the relaxation from the crystalline environment to solution and only moderate deviations in the dynamical fluctuations are observed.

Of key importance for catalysis is the degree to which the substrate and active site nucleotides sample 'active' conformations relevant for the chemical steps of the reaction to occur. In the present case, this involves the nucleophilic attack of the A63:N1 to the O^6mG:Cm, and departure of the O^6mG:O6 leaving group. The requirements for this to occur effectively are that the O^6mG:Cm–A63:N1 distance is sufficiently close, and that the A63:N1–O^6mG:Cm–

O^6mG:O6 angle (θ_{inl}) is 'inline' (sufficiently close to 180°). Previous theoretical work (36–38) has established that a useful heuristic measure of 'inline fitness' of the dynamical ensemble for similar nucleophilic attack reactions is to restrict the nucleophile–electrophile distances within 3.5 Å and the θ_{inl} angle >140°. In the crystal structure of the product form, the A63:Cm–guanine O6 distance is 2.8 Å (suggesting these atoms are in nonbonded contact) and A63:N1–Cm–O^6mG:O6 angle is 167.9°. Figure 1C shows the inline fitness landscape from the solution simulations. The simulations predict that 80.7% of reactant state trajectory frames occupy distances <3.5 Å and angles >140° (as denoted by the upper left box in Figure 1C), suggesting A63 is effectively positioned for reaction a majority of the time. Further analysis of hydrogen bonding patterns in relation to inline fitness can be found in Supplementary Figure S2. Taken together, this provides strong evidence that the dynamical ensemble described here with C10 protonated is representative of a plausible active state in solution that can be used as a departure point for further investigation of the chemical steps of the reaction using multiscale QM/MM simulations.

Catalytic pathway for methyl transfer reaction

Having determined a plausible active state dynamical ensemble in solution, we set out to map out the free energy surface (FES) for the catalytic chemical steps of the reaction and mechanistic pathway. We consider two reaction coordinates: a methyl transfer progress variable that is a difference in the leaving group departure and nucleophilic attack distances [$\xi_{MT} = R(O^6$mG:O6–Cm$) - R($A63:N1–O^6mG:Cm$)$], and a proton transfer progress variable [$\xi_{PT} = R($C10:N3–H$) - R(O^6$mG:N1–H$)$]. The reactant and product states have (ξ_{MT}, ξ_{PT}) values of approximately (−1.5, −1 Å) and (1.5, 1 Å), respectively (see Supplementary Figure S3).

We first exhaustively mapped out the full 2D FES using QM/MM simulations at the DFTB3 level (Supplementary Figure S3) and determined the existence of a low-barrier stepwise pathway (∼17.5 kcal·mol^{-1}) in which proton transfer precedes methyl transfer. We also considered a high-barrier stepwise pathway (∼32.0 kcal·mol^{-1}) in which methyl transfer occurs first, followed by proton transfer as well as a concerted mechanism whereby proton transfer and methyl transfer occur synchronously, and this led to single high barrier of ∼31.8 kcal·mol^{-1}. Although the DFTB3 results are not expected to be quantitatively accurate, the stark difference in barriers strongly suggests that the most likely mechanism corresponds to the low-barrier stepwise path. Informed by these results, we set out to determine the minimum free energy pathway and barriers at the much more computationally intensive and quantitatively accurate *ab initio* QM/MM level.

Toward this end, we performed *ab initio* QM/MM path simulations using the finite temperature umbrella sampling string method departing from the concerted pathway described above as an initial guess so as not to bias the search (see the 'Detailed Computational Methods' section in the Supplementary Data, Supplementary Figure S4 and Supplementary Table S2). The converged *ab initio* QM/MM

string simulations predicted a stepwise pathway similar to that derived from the DFTB3 QM/MM simulations (Figure 2A) and, taken together with the DFTB3 results, eliminated the alternative pathways as feasible; thus, computational effort was not spent to refine this path at the *ab initio* QM/MM level. The minimum free energy pathway (Figure 2A and B) passes through two transition states, the first (TS1) corresponding to the proton transfer with a barrier of 7.2 kcal·mol^{-1} and the second rate-controlling transition state (TS2) corresponding to methyl transfer with a barrier of 19.4 kcal·mol^{-1}. These transition states are illustrated mechanistically and structurally in Figure 2C and D, respectively.

The transition states are separated by a metastable intermediate (I) 2.0 kcal·mol^{-1} higher in free energy than the reactant state. The O^6mG intermediate carries a positive charge and exhibits resonance to Cm that promotes O^6mG:O6–Cm bond cleavage and in turn Cm–A63:N1 bond formation in the subsequent step. Considering these fundamental requirements, the nucleophilicity of adenine and ability of the ligand to stabilize positive charge are essential. This is plausibly reflected in experiments by Scheitl *et al.* in which a split (three-strand) MTR1 construct catalyzes the reaction involving O^6-(*p*-aminomethyl)benzylguanine (ab^{6}G) substrate more rapidly than O^6mG (33). A possible cause for the increase in rate could be that ab^{6}G stabilizes positive charge more effectively through resonance, thus encouraging O^6–Cm bond cleavage (see Supplementary Figure S5).

In order to investigate further the active protonation state of C10 suggested by experiment, we performed *ab initio* QM/MM simulations of the reaction with neutral C10 (deprotonated at N3). In this scenario, there is not an available proton at C10:N3 to be transferred to the O^6mG substrate prior to nucleophilic attack (and hence only the ξ_{MT} reaction coordinate was considered). Calculations of the 1D free energy profile predicted a pathway with a single high-barrier (30.5 kcal·mol^{-1}) transition state. This barrier is >10 kcal·mol^{-1} higher than the corresponding barrier with the active state where C10 is protonated. Taken together with the calculated pK_a shift predictions described below, these results provide strong evidence that the catalytic mechanism requires C10 to be protonated in the active state.

Reconciling crystallographic data in the context of mechanism

In the crystal structure reported by Deng *et al.* (32), the experimental distance between the O^6mG:O6 and A63:N1 is 4.2 Å. The crystal structure represents the product state of the reaction in which the methyl group (Cm) has been transferred to A63:N1. Contemplating the reverse step of the reaction, given that a C–O single bond distance is ∼1.4 Å, this implies the C–N distance in the previous step should be around 2.8 Å. At this distance, however, one does not expect there to be great orbital overlap between the Cm and A63:N1 atoms required to initiate bond formation. This begs the question as to whether a local structural rearrangement from the crystal is required in order to initiate nucleophilic attack.
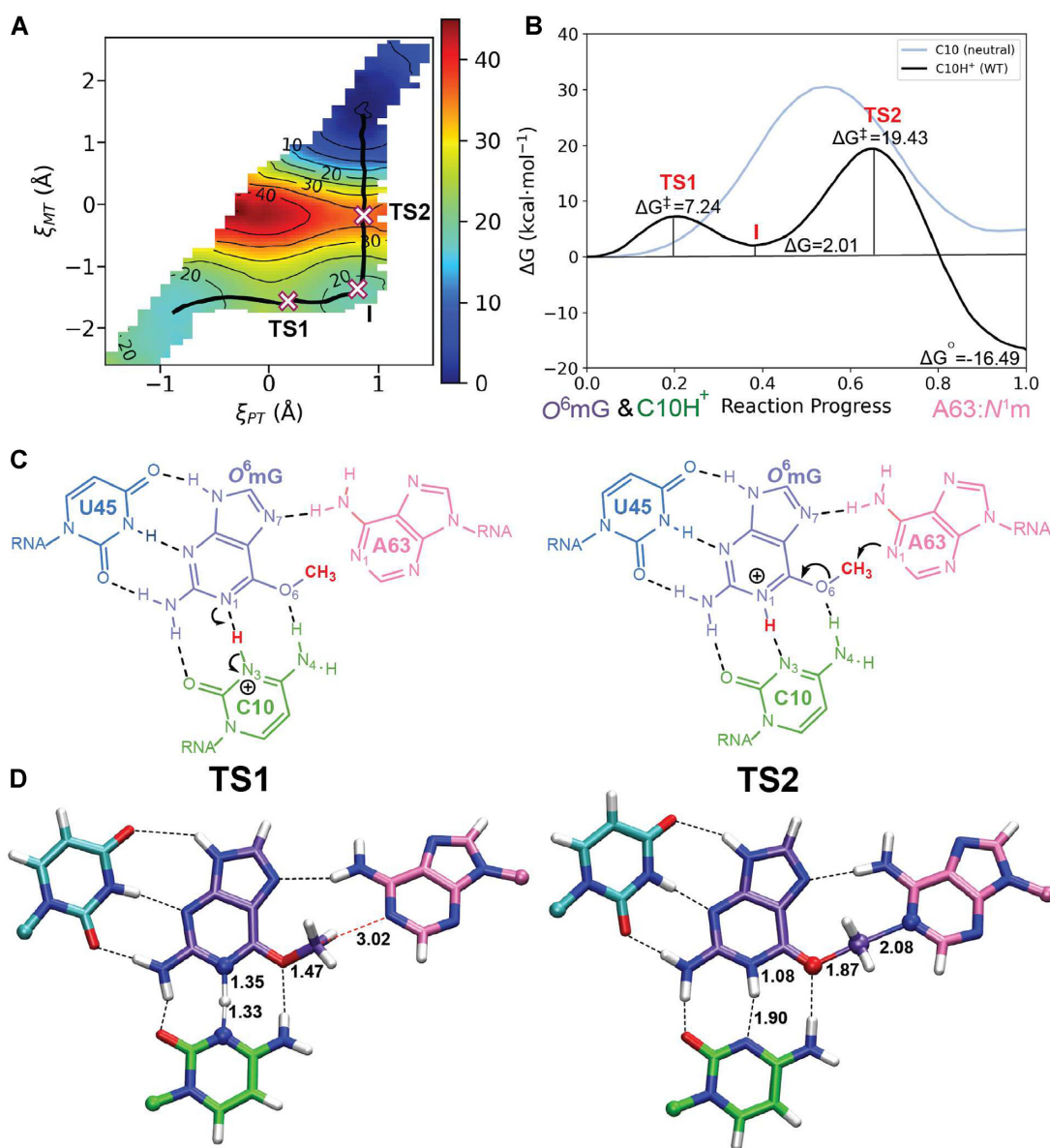
Figure 2. Results of *ab initio* QM/MM simulation of the reaction trajectory of MTR1 departing from the active state in solution. (**A**) 2D free energy landscape from *ab initio* QM/MM simulations with converged, stepwise path overlaid (black) and stationary points marked with crosses. (**B**) Free energy profiles (46) corresponding to the path indicated in panel (A) shown in black and resulting from the 1D reaction coordinate of methyl transfer in the absence of C10 protonation shown in gray. (**C**) Schematic of the active site depicting the transition states and corresponding mechanistic steps. (**D**) Average transition state structures calculated from additional 5 ps sampling at the stationary points found on the minimum free energy path. Bond lengths (Å) corresponding to reaction coordinates are indicated.

In order to reconcile this question, we examined the progression of the O^6mG:O6–A63:N1 distance, as well as the distances of the breaking/forming bonds to the transferred methyl Cm, along the reaction coordinate. These average distances, as well as their standard deviations and 95% confidence intervals, are illustrated in Figure 3 along with the computed free energy profile and dashed lines that reference the product state crystallographic distances (average distances and dynamical ranges are listed in Supplementary Table S3). The rate-controlling transition state, TS2, is associated with bringing O^6mG:O6 and A63:N1 close enough such that orbital overlap between A63:N1 and Cm may initiate nucleophilic attack and chemical bonding. What is im-

mediately clear is that in the product state (reaction progress value of 1.0), the average distances from the simulations are very close to the experimental reference values (highlighted in yellow) and in all cases the range of dynamical fluctuations encompasses the reference values. This indicates that no local rearrangement from the crystal structure is required in the above approach; rather, the distances will undergo the required adjustments as the reaction proceeds away from the minima in the profile and into the transition state region.

Departing from the intermediate, the O^6mG:O6–A63:N1 distance (red) contracts from an average value of 4.32 Å to a value of 3.94 Å as it passes through the rate-controlling
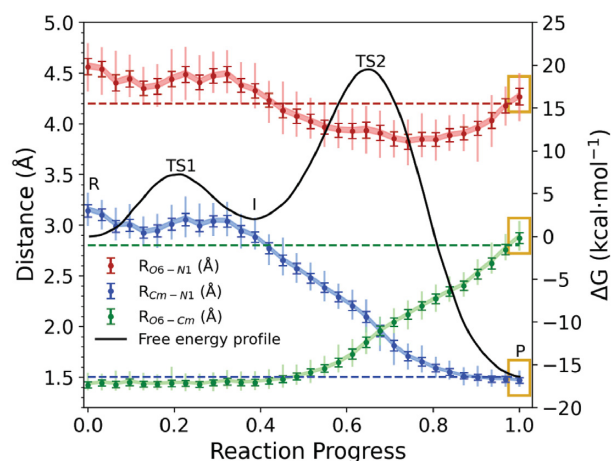
Figure 3. Analysis of methyl transfer distances (Å) in the context of the free energy profile (black) from *ab initio* QM/MM simulations. The distances within O^6mG:O6–A63:N1 (red), Cm–A63:N1 (blue) and O^6mG:O6–Cm (green) were measured for the trajectory average structures of each umbrella window on the converged reaction path. Statistical values such as 95% confidence intervals (shaded), standard deviations (dark error bars) and range of observed values (vertical lines) are traced through the trajectories for each umbrella window. The distances in the crystal structure of the product state are indicated by dashed lines.

transition state (TS2) where it samples a range between 3.71 and 4.17 Å, and then increases again to 4.27 Å as it reaches the product. The minimum value of the O^6mG:O6–A63:N1 average distance is 3.83 Å and occurs slightly after TS2 as the Cm–A63:N1 bond is almost fully formed (1.71 Å) and the O^6mG:O6–Cm bond is partially broken (2.13 Å). In the intermediate, the Cm–A63:N1 distance is 2.88 Å, and in the product the O^6mG:O6–Cm distance is 2.87 Å, very close to the experimental value of 2.8 Å. These values are typical of closely packed nonbonded interatomic C–N and C–O separation distances. We thus conclude that the crystal structure (32) represents a stable product state minimum consistent with simulations, and that the requisite orbital overlap for bond formation is achieved as a result of progression from stable nonbonded equilibrium states to higher free energy states along the reaction profile.

Interpretation of apparent pK_a values in activity–pH profile for MTR1

The measurement of activity–pH profiles provides important data that report on the sensitivity of the observed rate to the protonation state (36,52,65,66). The interpretation of experimental activity–pH profiles of ribozymes is often not straightforward due to complicating factors (36,67). In order to gain mechanistic insights from these data, one must ascertain which protonation events give rise to the shape of the profiles and their maximum rates. This requires identification of nucleotides where specific protonation states are critical for function, and then to ascertain the pK_a values associated with these states in the ribozyme environment.

The experimental activity–pH profile for MTR1 is a bell-shaped curve with apparent pK_a values of 5.0 and 6.2 that

were tentatively assigned to A63 and C10, although without experimental validation (32). Several factors may impact these data. Ideally, experiments would be conducted with a saturating concentration of substrate; however, the low solubility of guanine and associated compounds may preclude this. The substrate concentration used was 50 μM, but the K_d of binding is currently unknown. As the protonation of C10 may plausibly affect the affinity of substrate binding, it is possible the activity–pH profile is influenced by variable substrate affinities. Furthermore, the steep decline in activity below pH 5.5 observed by Scheitl *et al.* (33) suggests that destabilization of the ribozyme structure may influence the apparent pK_a values. These limitations of the experimental approach make direct comparison with predicted pK_a results more problematic, but at the same time highlight why these computational studies are valuable.

The purpose of this section is to provide computational support through QM/MM and AFE simulations to aid in the interpretation of the measured activity–pH profiles, and in particular by providing support for C10 as the general acid apparent pK_a value (32,33). Given the experimental pK_a of O^6mG is quite low (2.35) (68), we eliminated the substrate as a likely candidate to account for either of the observed apparent pK_a values of 5.0 or 6.2. In the previous section, *ab initio* QM/MM simulations predicted the protonation of C10 to be required for catalysis, implicating this nucleotide as a possible candidate for the acid apparent pK_a. To test this hypothesis further, we performed additional *ab initio* QM/MM simulations of the C10U mutant. Uracil has a pK_a value at the N3 position of ∼9.2, upshifted from that of cytosine by 5 units, suggesting it will be much more difficult to abstract a proton from this position (69). The QM/MM simulations (Figure 4A) predict a rate-controlling barrier of 32.6 kcal·mol^{-1}, slightly higher than that predicted for neutral C10 with N3 deprotonated (Table 1). This barrier corresponds to an intrinsic rate constant of ≈10^{-11} min^{-1}, which is well below an experimental detection limit of 10^{-6} min^{-1}, in agreement with the severely impaired activity observed experimentally for this mutation (32). We thus sought to determine the pK_a value of C10:N3 to ascertain the degree to which it corresponded to the experimentally measured apparent pK_a.

We performed AFE simulations to calculate the microscopic pK_a shift of C10 in the substrate-bound form of MTR1 according to the transformations shown in Supplementary Figure S7. Results are listed in Table 2. The value of the acid shift relative to the solution value of 4.2 (55) was determined to be 2.06 when O^6mG is bound, leading to a predicted pK_a value of 6.26 in MTR1, very close to the apparent pK_a value of 6.2 observed experimentally. This pK_a prediction, together with the computational evidence of the preceding section demonstrating the critical role of C10 protonation in the catalytic mechanism, provides strong support for the original experimental supposition (32) that C10 is the most likely candidate from which the apparent pK_a of 6.2 arises. The upshifted pK_a of C10 can be explained by examination of the strong and persistent hydrogen bond with the O^6mG:N1 position, and thus protonation of C10 is likely correlated with substrate binding.

Table 1. Free energy profile data from QM/MM simulations of the WT and C10U mutation with predicted cleavage rates[a]

Species	Free energy (kcal·mol^{-1})				Rate (min^{-1})	
	$\Delta G^{\ddagger}_{TS1}$	ΔG_I	$\Delta G^{\ddagger}_{TS2}$	ΔG_{rxn}	k_{int}(calc.)	k_{int}(expt.)
C10H$^+$ (WT)	7.24(22)	2.01(24)	19.43(17)	$-16.49(07)$	0.45(06)	0.12
C10 (neutral)	N/A	N/A	30.53(14)	4.61(27)	$\approx 10^{-8}$	N/A
C10U	11.56(35)	9.13(43)	32.59(49)	$-2.87(12)$	$\approx 10^{-11}$	$<10^{-6}$

[a]Free energy values (kcal·mol^{-1}) for stationary points in the free energy profiles shown in Figures 2B and 4A. Intrinsic reaction rates (min^{-1}) were estimated from Equation (S1) in the Supplementary Data, where $\Delta G^{\ddagger}$ is the rate-controlling transition state free energy barrier listed in this table and shown in Figure 4. The intrinsic rate for the C10U mutant is below the experimental detection limit of $<10^{-6}$ min^{-1}. Since the experimental intrinsic rate is taken to result from C10H$^+$, the C10 (neutral) experimental rate is not applicable (N/A). Standard error estimates are given in parentheses multiplied by 100, and error bars for the calculated profiles shown in Figure 4B are provided in Supplementary Figure S6.
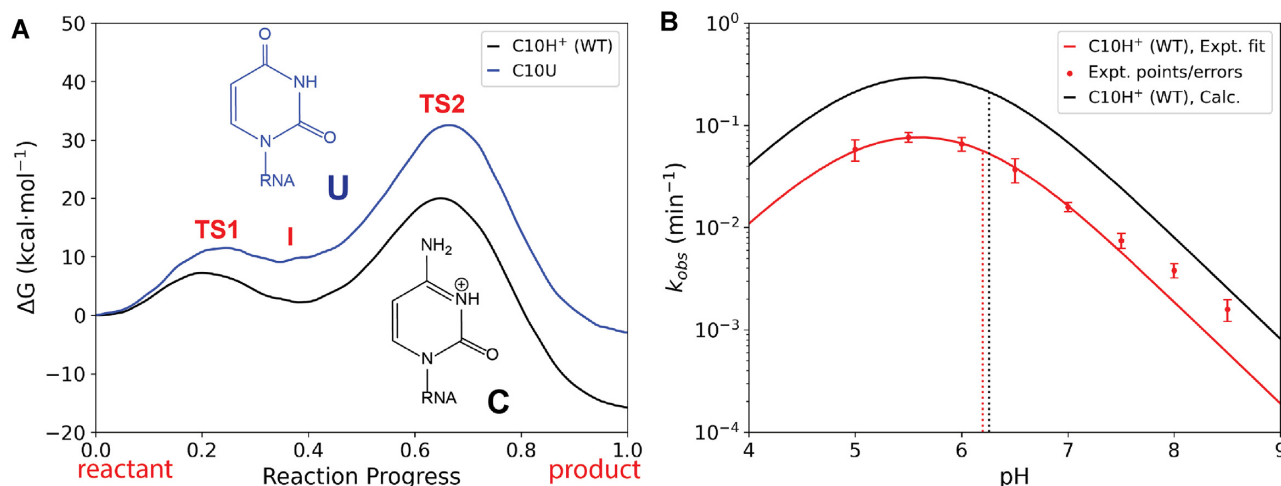

Figure 4. Prediction of activity–pH profiles for C10H$^+$ (WT) based on the free energy barrier from QM/MM and pK_a shift from AFE simulations. (**A**) Free energy profiles corresponding to QM/MM simulations of WT C10H$^+$ (black) and C10U mutation (blue). (**B**) Predicted (black) and experimental (red) WT activity–pH profile calculated using the two-state model shown in Equation (3). Red points and error bars are those of Deng *et al.* (32). Dotted lines indicate the predicted and experimental C10 pK_a values, and the lower pK_a of 5.0 from experiment is assumed. The calculated C10U mutant results were below the detection limit of 10^{-6} min^{-1} (in accord with experiment) and hence not visible on the scale of the plot. Computational error analysis of the free energy profiles is given in Supplementary Figure S6.

Table 2. pK_a values and shifts of the predicted general acid (C10:N3) and nucleophile (A63:N1) nucleotides in the ligand-bound form of MTR1[a]

Nucleotide	Role	Soln.	MTR1 (bound)		Expt.
		pK_a	ΔpK_a	pK_a	App. pK_a
C10	General acid	4.2	2.06(03)	**6.26**	**6.2**
A63	Nucleophile	3.5	$-0.74(62)$	**2.76**	**5.0**

[a]Shifts are calculated relative to experimental solution values (55,56) (Soln.) from AFE simulations for MTR1 bound to the O^6mG substrate. The experimental apparent pK_a values are shown for reference; however, the respective roles were not experimentally validated. Values shown in bold are parameters used to construct the activity-pH profiles in Figure 4. Analysis of enhanced sampling for AFE simulations used to determine the pK_a shifts is provided in Supplementary Table S5 and Supplementary Figure S8.

The origin of the lower apparent pK_a value of 5.0, however, was unclear from experiment. Several naturally occurring ribozymes that use a general acid–base strategy for catalysis exhibit bell-shaped activity–pH profiles similar to that of MTR1. Experiments and computation indicate MTR1 employs a general acid, but not a general base. Nonetheless, the A63:N1 position is the nucleophile in the reaction (i.e. acts as a Lewis base), and thus would be functionally sensitive to and essentially inactivated by protonation at N1. We thus considered A63:N1 as a conceivable candidate as the origin of the unidentified apparent pK_a. Computing the pK_a shift of A63 relative to the solution value of 3.5 (56) in the bound form of the ribozyme results in downshifting by 0.74 pK_a units due to the proximity of the O^6mG methyl group (Table 2). The resulting pK_a value of 2.76 is too low to significantly influence the experimentally observed activity–pH profile, leading us to conclude that the lower apparent pK_a of 5.0 does not arise from protonation of A63:N1.

Alternative explanations for the apparent pK_a of 5.0 could include protonation of one or more of A9, A40 or A46, which form a base triple that stacks with the ligand binding nucleotides; however, it is unclear how these protonation events may impact catalysis. Alternatively, the decrease in activity at low pH may arise from destabilization of the ribozyme or weaker substrate binding.

The AFE simulations thus strongly implicate C10:N3 as the origin of apparent pK_a of 6.2, but do not support the hypothesis that A63:N1 may be the origin of the apparent pK_a of 5.0. Based on the calculated pK_a values, we sought to

predict the full activity–pH profile from computations such that direct comparison could be made with the experimental profile.

In order to predict the full activity–pH profile from Equation (3), in addition to the computed pK_a values (Table 2), one must determine the intrinsic rate (k_{int}), which can be obtained from *ab initio* QM/MM simulations [see Equation (S1) of the Supplementary Data]. The calculated free energy values and estimated intrinsic rates are listed in Table 1, and the resulting computed activity–pH profile is illustrated in Figure 4B. To our knowledge, these results represent one of the first reported activity–pH profiles predicted from a combination of *ab initio* QM/MM and AFE simulations. The calculated and measured profiles are very similar, apart from the divergence at low pH, and are shifted by roughly half of a log unit. This agreement is very encouraging and provides support for the validity of the simulation results and their interpretation.

Mechanism of MTR1 bears similarity to that predicted for *O*[6]-alkylguanine-DNA alkyltransferase protein

The observation of methyl and alkyl transfer activity by RNAs provides support for the RNA world hypothesis by demonstrating that the limited RNA library can include molecules that are capable of diverse chemistry similar to that observed in naturally occurring protein enzymes. Parallel mechanistic concepts emerge from study of *O*[6]-alkylguanine-DNA alkyltransferase (AGT) protein that directly repairs potentially carcinogenic *O*[6]-methylguanine bases in DNA (70). The mechanism of AGT has been studied previously using QM optimization methods on an active site cluster model (71) that suggested a stepwise mechanism involving nucleophile (Cys145) activation through His146 that acts as a water-mediated general base, followed by nucleophilic attack to dealkylate the guanine base. In a later study, QM/MM calculations (nondynamical path optimization) in a spherical partially frozen water droplet (72) supported a mechanism where AGT acts in a stepwise manner of nucleophile (Cys145) activation, followed by methyl transfer that occurs as a two-step migration with a stable intermediate (72). These simulations predicted a rate-controlling barrier of 18.2 kcal·mol^{-1} for the first step of the methyl transfer and differed from the previous cluster model work that predicted the methyl transfer to occur in a single step with higher rate-controlling barrier (23.2 kcal·mol^{-1}). In the present work, we performed *ab initio* QM/MM simulations of a fully explicitly solvated MTR1 system under periodic boundary conditions (45) to determine the free energy profile and pathway. Despite methodological differences between these works, the common theme is that MTR1 and AGT likely exhibit similar catalytic mechanisms (i.e. stepwise mechanisms with rate-controlling methyl transfer steps), highlighting an intriguing connection between these RNA and protein enzymes.

CONCLUSION

Here, we have used a combination of classical MD, QM/MM and AFE simulations to explore the mechanism of methyl transfer from *O*[6]mG to N1 of a specific adenine in the MTR1 ribozyme. The structure and dynamics in solution involve only modest relaxation from the crystal and support an active state in which C10, protonated at the N3 position, forms a key hydrogen bond with the N1 position of the *O*[6]mG substrate. The lowest energy reaction pathway determined by QM/MM free energy simulations suggests a stepwise mechanism that passes through two transition states separated by a metastable intermediate, as opposed to a concerted mechanism. The first transition state (7.2 kcal·mol^{-1} barrier) corresponds to the proton transfer from C10:N3 to *O*[6]mG:N1, whereas the second rate-controlling transition state corresponds to methyl transfer (19.4 kcal·mol^{-1} barrier). Simulations of the reaction with neutral C10 led to a barrier >10 kcal·mol^{-1} higher, highlighting the critical role of this residue acting as an acid in catalysis. In order to aid in the interpretation of experimental activity–pH profiles, AFE simulations predicted the pK_a for C10 to be 6.3, very close to the experimental apparent pK_a of 6.2, strongly implicating this residue as a critical acid, in support of the QM/MM simulation results. Finally, we make mechanistic comparison of MTR1 with AGT. MTR1 has evolved *in vitro* for alkyl transfer activity as would be required for the metabolism of a putative RNA world. It is thus of considerable interest that there are clear mechanistic similarities to naturally occurring protein enzymes like AGT. The results of the current work provide further support for evolutionary theories that suggest life originated in an RNA world, and help to establish design frameworks for new RNA-based biochemical tools.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Seith,D.D., Bingaman,J.L., Veenis,A.J., Button,A.C. and Bevilacqua,P.C. (2018) Elucidation of catalytic strategies of small nucleolytic ribozymes from comparative analysis of active sites. *ACS Catal.*, **8**, 314–327.
2. Bevilacqua,P.C., Harris,M.E., Piccirilli,J.A., Gaines,C., Ganguly,A., Kostenbader,K., Ekesan,Ş. and York,D.M. (2019) An ontology for facilitating discussion of catalytic strategies of RNA-cleaving enzymes. *ACS Chem. Biol.*, **14**, 1068–1076.
3. Lilley,D.M.J. (2019) Classification of the nucleolytic ribozymes based upon catalytic mechanism. *F1000 Res.*, **8**, 1462.
4. Gaines,C.S., Piccirilli,J.A. and York,D.M. (2020) The L-platform/L-scaffold framework: a blueprint for RNA-cleaving nucleic acid enzyme design. *RNA*, **26**, 111–125.
5. Ekesan,Ş., McCarthy,E., Case,D.A. and York,D.M. (2022) RNA electrostatics: how ribozymes engineer active sites to enable catalysis. *J. Phys. Chem. B*, **126**, 5982–5990.
6. Kruger,K., Grabowski,P., Zaug,A., Sands,J., Gottschling,D. and Cech,T. (1982) Self-splicing RNA: autoexcision and autocyclisation of the ribosomal RNA intervening sequence of *Tetrahymena. Cell*, **31**, 147–157.
7. Gilbert,W. (1986) The RNA world. *Nature*, **319**, 618.
8. Johnston,W.K., Unrau,P.J., Lawrence,M.S., Glasner,M.E. and Bartel,D.P. (2001) RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science*, **292**, 1319–1325.
9. Lilley,D.M. and Eckstein,F. (eds) (2008) In: *Ribozymes and RNA Catalysis, RSC Biomolecular Series*. RSC Publishing, Cambridge.
10. Lassila,J.K., Zalatan,J.G. and Herschlag,D. (2011) Biological phosphoryl-transfer reactions: understanding mechanism and catalysis. *Annu. Rev. Biochem.*, **80**, 669–702.
11. Weinberg,C.E., Weinberg,Z. and Hammann,C. (2019) Novel ribozymes: discovery, catalytic mechanisms, and the quest to understand biological function. *Nucleic Acids Res.*, **47**, 9480–9494.
12. Zhang,J., Lau,M.W. and Ferré-D'Amaré,A.R. (2010) Ribozymes and riboswitches: modulation of RNA function by small molecules. *Biochemistry*, **49**, 9123–9131.
13. Baker,J.L., Sudarsan,N., Weinberg,Z., Roth,A., Stockbridge,R.B. and Breaker,R.R. (2012) Widespread genetic switches and toxicity resistance proteins for fluoride. *Science*, **335**, 233–235.
14. Ishida,S., Terasaka,N., Katoh,T. and Suga,H. (2020) An aminoacylation ribozyme evolved from a natural tRNA-sensing T-box riboswitch. *Nat. Chem. Biol.*, **16**, 702–709.
15. Wilson,T.J. and Lilley,D.M.J. (2021) The potential versatility of RNA catalysis. *Wiley Interdiscip. Rev. RNA*, **12**, 1651.
16. Wilson,C. and Szostak,J. (1995) *In vitro* evolution of a self-alkylating ribozyme. *Nature*, **374**, 777–782.
17. Yokobayashi,Y. (2020) High-throughput analysis and engineering of ribozymes and deoxyribozymes by sequencing. *Acc. Chem. Res.*, **53**, 2903–2912.
18. Tarasow,T.M., Tarasow,S.L. and Eaton,B.E. (1997) RNA-catalysed carbon–carbon bond formation. *Nature*, **389**, 54–57.
19. Seelig,B. and Jäschke,A. (1999) A small catalytic RNA motif with Diels–Alderase activity. *Chem. Biol.*, **6**, 167–176.
20. Fusz,S., Eisenfuhr,A., Srivatsan,S.G., Heckel,A. and Famulok,M. (2005) A ribozyme for the aldol reaction. *Chem. Biol.*, **12**, 941–950.
21. Sengle,G., Eisenfuhr,A., Arora,P.S., Nowick,J.S. and Famulok,M. (2001) Novel RNA catalysts for the Michael reaction. *Chem. Biol.*, **8**, 459–473.
22. Zhang Bruice,T.C. (2007) Diels–Alder ribozyme catalysis: a computational approach. *J. Am. Chem. Soc.*, **129**, 1001–1007.
23. Bereźniak,T., Zahran,M., Imhof,P., Jäschke,A. and Smith,J.C. (2010) Magnesium-dependent active-site conformational selection in the Diels–Alderase ribozyme. *J. Am. Chem. Soc.*, **132**, 12587–12596.
24. Krochmal,D., Shao,Y., Li,N.-S., DasGupta,S., Shelke,S.A., Koirala,D. and Piccirilli,J.A. (2022) Structural basis for substrate binding and catalysis by a self-alkylating ribozyme. *Nat. Chem. Biol.*, **18**, 376–384.
25. McDonald,R.I., Guilinger,J.P., Mukherji,S., Curtis,E.A., Lee,W.I. and Liu,D.R. (2014) Electrophilic activity-based RNA probes reveal a self-alkylating RNA for RNA labeling. *Nat. Chem. Biol.*, **10**, 1049–1054.
26. Jiang,H., Gao,Y., Zhang,L., Chen,D., Gan,J. and Murchie,A.I.H. (2021) The identification and characterization of a selected
27. Flemmich,L., Heel,S., Moreno,S., Breuker,K. and Micura,R. (2021) A natural riboswitch scaffold with self-methylation activity. *Nat. Commun.*, **12**, 3877.
28. Scheitl,C. P.M., Ghaem Maghami,M., Lenz,A.-K. and Höbartner,C. (2020) Site-specific RNA methylation by a methyltransferase ribozyme. *Nature*, **587**, 663–667.
29. El Yacoubi,B., Bailly,M. and de Crécy-Lagard,V. (2012) Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annu. Rev. Genet.*, **46**, 69–95.
30. Sharma,S., Watzinger,P., Kötter,P. and Entian,K.-D. (2013) Identification of a novel methyltransferase, Bmt2, responsible for the *N*-1-methyl-adenosine base modification of 25S rRNA in *Saccharomyces cerevisiae. Nucleic Acids Res.*, **41**, 5428–5443.
31. Zhang,C. and Jia,G. (2018) Reversible RNA modification N^1-methyladenosine (m^{1}A) in mRNA and tRNA. *Genomics Proteomics Bioinformatics*, **16**, 155–161.
32. Deng,J., Wilson,T.J., Wang,J., Peng,X., Li,M., Lin,X., Liao,W., Lilley,D.M.J. and Huang,L. (2022) Structure and mechanism of a methyltransferase ribozyme. *Nat. Chem. Biol.*, **18**, 556–564.
33. Scheitl,C. P.M., Mieczkowski,M., Schindelin,H. and Höbartner,C. (2022) Structure and mechanism of the methyltransferase ribozyme MTR1. *Nat. Chem. Biol.*, **18**, 547–555.
34. Garcia-Viloca,M., Gao,J., Karplus,M. and Truhlar,D.G. (2004) How enzymes work: analysis by modern rate theory and computer simulations. *Science*, **303**, 186–195.
35. Panteva,M.T., Dissanayake,T., Chen,H., Radak,B.K., Kuechler,E.R., Giambaşu,G.M., Lee,T.-S. and York,D.M. (2015) Multiscale methods for computational RNA enzymology. In: Chen,S-J. and Burke-Aguero,D.H. (eds). *Methods in Enzymology*. Elsevier, Amsterdam, Vol. **553**.
36. Gaines,C.S., Giese,T.J. and York,D.M. (2019) Cleaning up mechanistic debris generated by twister ribozymes using computational RNA enzymology. *ACS Catal.*, **9**, 5803–5815.
37. Ekesan,Ş. and York,D.M. (2019) Dynamical ensemble of the active state and transition state mimic for the RNA-cleaving 8–17 DNAzyme in solution. *Nucleic Acids Res.*, **47**, 10282–10295.
38. Ganguly,A., Weissman,B.P., Giese,T.J., Li,N.-S., Hoshika,S., Rao,S., Benner,S.A., Piccirilli,J.A. and York,D.M. (2020) Confluence of theory and experiment reveals the catalytic mechanism of the Varkud satellite ribozyme. *Nat. Chem.*, **12**, 193–201.
39. Horn,H.W., Swope,W.C., Pitera,J.W., Madura,J.D., Dick,T.J., Hura,G.L. and Head-Gordon,T. (2004) Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.*, **120**, 9665–9678.
40. Joung,I.S. and Cheatham,T.E. III (2008) Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B*, **112**, 9020–9041.
41. Darden,T., York,D. and Pedersen,L. (1993) Particle mesh Ewald: an N log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092.
42. Essmann,U., Perera,L., Berkowitz,M.L., Darden,T., Lee,H. and Pedersen,L.G. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.*, **103**, 8577–8593.
43. Zgarbová,M., Otyepka,M., Šponer,J., Mládek,A., Banáš,P., Cheatham,T.E. III and Jurečka,P. (2011) Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.*, **7**, 2886–2902.
44. Pérez,A., Marchán,I., Svozil,D., Sponer,J., Cheatham,T.E. III, Laughton,C.A. and Orozco,M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophys. J.*, **92**, 3817–3829.
45. Giese,T.J. and York,D.M. (2016) Ambient-potential composite Ewald method for *ab initio* quantum mechanical/molecular mechanical molecular dynamics simulation. *J. Chem. Theory Comput.*, **12**, 2611–2632.
46. Giese,T.J., Ekesan,Ş. and York,D.M. (2021) Extension of the variational free energy profile and multistate Bennett acceptance ratio methods for high-dimensional potential of mean force profile analysis. *J. Phys. Chem. A*, **125**, 4216–4232.
47. Elstner,M., Porezag,D., Jungnickel,G., Elsner,J., Haugk,M., Frauenheim,T., Suhai,S. and Seifert,G. (1998) Self-consistent-charge

SAM-dependent methyltransferase ribozyme that is present in natural sequences. *Nat. Catal.*, **4**, 872–881.

density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B*, **58**, 7260–7268.

48. Cui,Q., Elstner,M., Kaxiras,E., Frauenheim,T. and Karplus,M. (2001) A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method. *J. Phys. Chem. B*, **105**, 569–585.

49. Gaus,M., Cui,Q. and Elstner,M. (2011) DFTB3: extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB). *J. Chem. Theory Comput.*, **7**, 931–948.

50. Gaus,M., Goez,A. and Elstner,M. (2013) Parametrization and benchmark of DFTB3 for organic molecules. *J. Chem. Theory Comput.*, **9**, 338–354.

51. Rosta,E., Nowotny,M., Yang,W. and Hummer,G. (2011) Catalytic mechanism of RNA backbone cleavage by ribonuclease H from quantum mechanics/molecular mechanics simulations. *J. Am. Chem. Soc.*, **133**, 8934–8941.

52. Dissanayake,T., Swails,J.M., Harris,M.E., Roitberg,A.E. and York,D.M. (2015) Interpretation of pH–activity profiles for acid–base catalysis from molecular simulations. *Biochemistry*, **54**, 1307–1313.

53. Lee,T.-S., Allen,B.K., Giese,T.J., Guo,Z., Li,P., Lin,C., McGee,T.D. Jr, Pearlman,D.A., Radak,B.K., Tao,Y. *et al.* (2020) Alchemical binding free energy calculations in AMBER20: advances and best practices for drug discovery. *J. Chem. Inf. Model.*, **60**, 5595–5623.

54. Lee,T.-S., Tsai,H.-C., Ganguly,A., Giese,T.J. and York,D.M. (2021) Robust, efficient and automated methods for accurate prediction of protein–ligand binding affinities in AMBER drug discovery boost. In: *Free Energy Methods in Drug Discovery: Current State and Future Directions, ACS Symposium Series*. ACS Publications, Washington, DC, Vol. **1397**, pp. 161–204.

55. Shugar,D. and Fox,J.J. (1952) Spectrophotometric studies of nucleic acid derivatives and related compounds as a function of pH. *Biochim. Biophys. Acta*, **9**, 199–218.

56. Harkins,T.R. and Freiser,H. (1958) Adenine–metal complexes. *J. Am. Chem. Soc.*, **80**, 1132–1135.

57. Tsai,H.-C., Lee,T.-S., Ganguly,A., Giese,T.J., Ebert,M.C., Labute,P., Merz,K.M. Jr and York,D.M. (2023) AMBER free energy tools: a new framework for the design of optimized alchemical transformation pathways. *J. Chem. Theory Comput.*, **19**, 640–658.

58. Lee,T.-S., Tsai,H.-C., Ganguly,A. and York,D.M. (2023) ACES: optimized alchemically enhanced sampling. *J. Chem. Theory Comput.*, **19**, 472–487.

59. Ekesan,Ş. and York,D.M. (2019) Framework for conducting and analyzing crystal simulations of nucleic acids to aid in modern force field evaluation. *J. Phys. Chem. B*, **123**, 4611–4624.

60. Giambaşu,G.M., York,D.M. and Case,D.A. (2015) Structural fidelity and NMR relaxation analysis in a prototype RNA hairpin. *RNA*, **21**, 963–974.

61. Deng,N.-J. and Cieplak,P. (2010) Free energy profile of RNA hairpins: a molecular dynamics simulation study. *Biophys. J.*, **98**, 627–636.

62. Bergonzo,C., Henriksen,N.M., Roe,D.R. and Cheatham,T.E. III (2015) Highly sampled tetranucleotide and tetraloop motifs enable evaluation of common RNA force fields. *RNA*, **21**, 1578–1590.

63. Banáš,P., Hollas,D., Zgarbová,M., Jurečka,P., Orozco,M., Cheatham,T.E., Šponer,J. and Otyepka,M. (2010) Performance of molecular mechanics force fields for RNA simulations: stability of UUCG and GNRA hairpins. *J. Chem. Theory Comput.*, **6**, 3836–3849.

64. Mráziková,K., Mlýnský,V., Kührová,P., Pokorná,P., Kruse,H., Krepl,M., Otyepka,M., Banáš,P. and Šponer,J. (2020) UUCG RNA tetraloop as a formidable force-field challenge for MD simulations. *J. Chem. Theory Comput.*, **16**, 7601–7617.

65. Bevilacqua,P.C. (2003) Mechanistic considerations for general acid–base catalysis by RNA: revisiting the mechanism of the hairpin ribozyme. *Biochemistry*, **42**, 2259–2265.

66. Kath-Schorr,S., Wilson,T.J., Li,N.-S., Lu,J., Piccirilli,J.A. and Lilley,D.M.J. (2012) General acid–base catalysis mediated by nucleobases in the hairpin ribozyme. *J. Am. Chem. Soc.*, **134**, 16717–16724.

67. Frankel,E.A. and Bevilacqua,P.C. (2018) Complexity in pH-dependent ribozyme kinetics: dark pK_a shifts and wavy rate–pH profiles. *Biochemistry*, **57**, 483–488.

68. Hornillo-Araujo,A.R., Burrell,A.J.M., Aiertza,M.K., Shibata,T., Hammond,D.M., Edmont,D., Adams,H., Margison,G.P. and Williams,D.M. (2007) The synthesis and properties of tricyclic analogues of S^6-methylthioguanine and O^6-methylguanine. *Nucleosides Nucleotides Nucleic Acids*, **26**, 1099–1102.

69. Moser,A., Range,K. and York,D.M. (2010) Accurate proton affinity and gas-phase basicity values for molecules important in biocatalysis. *J. Phys. Chem. B*, **114**, 13911–13921.

70. Tubbs,J.L., Pegg,A.E. and Tainer,J.A. (2007) DNA binding, nucleotide flipping, and the helix–turn–helix motif in base repair by O^6-alkylguanine-DNA alkyltransferase and its implications for cancer chemotherapy. *DNA Repair*, **6**, 1100–1115.

71. Georgieva,P. and Himo,F. (2008) Density functional theory study of the reaction mechanism of the DNA repairing enzyme alkylguanine alkyltransferase. *Chem. Phys. Lett.*, **463**, 214–218.

72. Hou,Q., Du,L., Gao,J., Liu,Y. and Liu,C. (2010) QM/MM study on the reaction mechanism of O^6-alkylguanine-DNA alkyltransferase. *J. Phys. Chem. B*, **114**, 15296–15300.