Organic & Biomolecular Chemistry

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Org. Biomol. Chem.*, 2022, **20**, 6219

Who stole the proton? Suspect general base guanine found with a smoking gun in the pistol ribozyme†

Şölen Ekesan * and Darrin M. York *

The pistol ribozyme (Psr) is one among the most recently discovered classes of small nucleolytic ribozymes that catalyze site-specific RNA self-cleavage through 2′-*O*-transphosphorylation. The Psr contains a conserved guanine (G40) that in crystal structures is in a position suggesting it plays the role of the general base to abstract a proton from the nucleophile to activate the reaction. Although some functional data is consistent with this mechanistic role, a notable exception is 2-aminopurine (2AP) substitution which has no effect on the rate, unlike similar substitutions across other so-called "G + M" and "G + A" ribozyme classes. Herein we postulate that an alternate conserved guanine, G42, is the primary general base, and provide evidence from molecular simulations that the active site of Psr can undergo local refolding into a structure that is consistent with the common "L-platform/L-scaffold" architecture identified in G + M and G + A ribozyme classes with Psr currently the notable exception. We summarize the key currently available experimental data and present new classical and combined quantum mechanical/molecular mechanical simulation results that collectively suggest a new hypothesis. We hypothesize that there are two available catalytic pathways supported by different conformational states connected by a local refolding of the active site: (1) a primary pathway with an active site architecture aligned with the L-platform/L-scaffold framework where G42 acts as a general base, and (2) a secondary pathway with the crystallographic active site architecture where G40 acts as a general base. We go on to make several experimentally testable predictions, and suggest specific experiments that would ultimately bring closure to the mystery as to "who stole the proton in the pistol ribozyme?".

Received 2nd February 2022,
Accepted 13th April 2022

DOI: 10.1039/d2ob00234e

rsc.li/obc

1. Introduction

The ground-breaking discovery that molecules of RNA[1–3], and later molecules of DNA[4], can act as catalysts has inspired a generation of scientists to explore new applications in biotechnology and medicine,[5] and gain a deeper understanding of biology[6] and the origins of life itself.[7,8] It remains a fascinating puzzle how these molecules, with their limited set of building blocks and lack of diversity of chemical functional groups, are able to catalyze complex reactions with rates that often rival those of more common protein enzymes. Nonetheless, these remarkable biomolecules are able to find ways of taking advantage of their environments and folding into 3D structures able to scaffold and position residues in the active site, recruit solvent components and shift pK_as to participate in catalysis and achieve rate enhancements over a million-fold.[9–11]

The intriguing mystery behind the catalytic capabilities of nucleic acids has attracted many a scientific explorer, experimental and theoretical alike. Experiments provide critical insight into truths of nature through measurements, however these measurements often provide somewhat veiled answers to the intended questions. Nowhere is this more apparent than in studies of enzyme catalysis, where the critical entity that controls the reaction rate is a transient mirage nearly impossible to observe directly – the transition state ensemble.[12] Theoretical methods, on the other hand, promise an almost limitless wealth of atomic level detail into catalytic pathways and the transition states they pass through. Nonetheless, as these methods must ultimately take recourse into approximate models to be made practical, results should be viewed with an appropriate level of skepticism until validated experimentally.

We have long endeavored to break down barriers between experiment and theory[13–16] and forge new paths for interdisciplinary collaborations to study mechanisms of biocatalysis. The vast majority of collaborative papers reported in the literature in this area have used theory and computation mainly to

Laboratory for Biomolecular Simulation Research, Institute for Quantitative Biomedicine and Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ 08854, USA.
E-mail: solen.ekesan@rutgers.edu, Darrin.York@rutgers.edu

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d2ob00234e

aid in the interpretation of experiments. We feel it is critical that theory break out of this traditional supporting role and evolve into a tool to gain predictive insight that inspires new experiments and guides new technology design.[17,18]

To achieve this, we strived to develop design principles for nucleic acid enzymes that can be used to make experimentally testable predictions,[19] and in the present work, we put these principles, and our models, humbly to the test to make a new hypothesis.

Hypothesis: *The primary general base that abstracts a proton from the nucleophile in the pistol ribozyme to initiate the reaction is not G40 as implicated by numerous crystal structures, but rather G42 that hitherto has not been a suspect in the mechanism*.

Herein we discuss currently available experimental data supporting and/or refuting G40 or G42 as the general base, and report a new model three-dimensional structure of the pistol ribozyme that supports G42 as the general base. The model was obtained through molecular simulations that involved local refolding of the active site into a common "L-platform/L-scaffold" framework[19] without disturbance to other structural components. We present the model evidence and reasoning behind our hypothesis in the form of an unsolved mystery:

"Who stole the proton in the pistol ribozyme?"

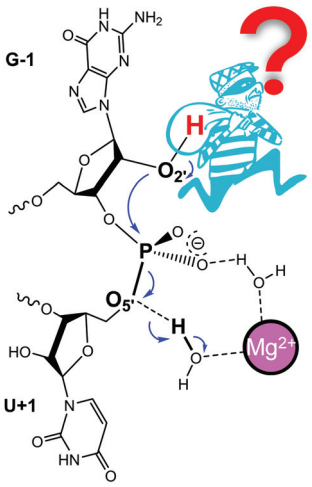

Fig. 1 RNA cleavage *via* 2′-*O*-transphosphorylation by the pistol ribozyme. It is generally accepted that a Mg^{2+} bound at the active site acts as the general acid (δ catalysis) and also provides electrostatic stabilization (β catalysis); and further that nucleophile activation (γ catalysis) is achieved by an active site guanine residue (deprotonated at the N1 position). Up until now, it has generally been assumed that G40 acts as the general base to abstract the nucleophile proton, but the current work brings this assumption into question, and forms a hypothesis that perhaps an alternative suspect guanine, G42, acts as the proton bandit.

Background: the facts in the case

RNA cleavage catalyzed by the pistol ribozyme

The pistol ribozyme (Psr) catalyzes site-specific RNA strand cleavage *via* 2′-*O*-transphosphorylation[20,21] that involves a nucleophilic attack of a 2′-hydroxyl group to the adjacent scissile phosphate and departure of the 5′O leaving group to produce a 2′,3′-cyclic phosphate (Fig. 1). The ribozyme draws from four fundamental catalytic strategies[15,22] to achieve speed limits[23] of roughly million-fold rate enhancement for the reaction: (1) alignment of the nucleophile, phosphorus and leaving group (α catalysis), (2) activation of the 2′OH nucleophile by abstraction of the proton by a general or specific base (γ catalysis), (3) electrostatic stabilization of the localized negative charge on the non-bridging phosphoryl oxygens (NPOs) in the transition state (β catalysis), and (4) facilitation of O5′ leaving group departure by donation of a proton from a general or specific acid.

Pioneering structural and functional studies by Breaker,[24,25] Micura and Patel,[26–28] Lilley and Wilson,[29] and others,[30] and supported by theoretical studies,[31–33] have made clear that a metal ion plays an important role in the acid step of the reaction, and further established a functional link with the hammerhead ribozyme[31] and the 8–17 DNAzyme,[17] although ultimately their mechanisms are distinct.

The pistol ribozyme belongs to a set of ribozymes classes, including the hairpin, Varkud satellite, twister, *glmS*, and hammerhead ribozymes, that utilize a guanine residue as a general base.[11] This strategy for nucleophile activation is also shared by artificially engineered DNAzymes, including the 8–17 DNAzyme.[34–36] Abundant structural evidence, along with some functional data, implicates G40 as the general base in Psr catalysis, and up until this point, this role has been generally accepted (or at least not loudly debated). A recent computational study has reported the possibility of specific base mechanism, as they observed better inline fitness when G40 was protonated and hydrogen bonded to O2′.[37] Proton abstraction by solvent has been proposed for ribozymes before,[38] and discussed as a viable mechanism for the HDV ribozyme[39] where the active site lacks a guanine positioned to act as the base. While a specific base mechanism in Psr can, as yet, not defintively be excluded, as discussed below the current body of evolutionary and functional data suggest the strong possibility that one or more guanine residues play a role in general base catalysis.

We now summarize the experimental evidence in support of G40 as the general base, and make new computational predictions, supported by some functional data, that raise suspicion about a different possible nucleobase of interest, G42, as the proton bandit.

G40, a highly conserved residue, plays a functional role

When the pistol ribozyme class was first revealed by comparative genomics,[24,25] G40 was identified early on as a highly conserved residue that might play an important functional role. The conservation of G40 is 100% and a double mutant G40U/C41A showed no activity.

A comparison of relative cleavage rates (k/k') for different ribozyme variants involving mutants and chemical modifications to the presumed general base guanine across different ribozyme classes is shown in Table 1. Mutation of G to A has a

Table 1 Summary of relative cleavage rates (k/k') for different ribozyme variants involving mutants and chemical modifications to the presumed general base guanine across different ribozyme classes and 8–17 DNAzyme[a]

	pK_a	VS G638	Hairpin (ref. 40) G8	Twister (ref. 41) G33	Hammerhead G12	8–17 DNAzyme G14	Pistol G40
G	9.4 (ref. 42)	1	1	1	1	1	1
A	3.8 (ref. 42)	7300 (ref. 42)	304	263	12 500 (ref. 43)	>2000 (ref. 44)	NA (ref. 26)
I	8.7 (ref. 42)	27 (ref. 16)	12	430	333 (ref. 45)	75 (ref. 34)	8 (ref. 29), 229 (ref. 28)
diAP	5.1 (ref. 42)	1600 (ref. 42)	6			10 (ref. 34)	
P	2.1 (ref. 42)	8000 (ref. 42)					
6sG	8.1 (ref. 16)	2.7 (ref. 16)				2 (ref. 44)	0.79 (ref. 29)
2AP	3.8 (ref. 42)	8500 (ref. 42)	130	72	1000 (ref. 45)	500 (ref. 34)	**0.98** (ref. 28)

[a] Relative cleavage rates, reported as fold decrease (k/k'), are with respect to the wild type guanine (G), modified to adenine (A), inosine (I), 2,6-diaminopurine (diAP), purine (P), 6-thioguanine (6sG), and 2-aminopurine (2AP). Value marked as "NA" was reported as no cleavage and rates were not determined. Also shown for reference in the second column are the pK_a values for the free nucleobases in aqueous solution.

large effect for all ribozymes, with the smallest effects occurring for the hairpin and twister ribozymes, neither of which has an explicit divalent metal ion requirement for catalysis. Inosine substitution has largely varying effects ranging from 12-fold in the hairpin to 430 in twister, with the Psr G40I values of 8 (ref. 29) and 229 (ref. 28). The 6-thioguanine (6sG) substitution effects were relatively small ranging from 0.8 to 2.7 for the 3 systems that were measured. Due to the lack of 6sG data for the other systems, it is not clear whether Psr's speed up compared to the slow down in VS and 8–17 DNAzyme is suspicious. Overall, there has not been a great deal of functional studies involving G40 in Psr, but surveying the available data, with the exception of 2-aminopurine (2AP) which will be discussed in more detail below, the Psr G40 variants do not appear to be outliers with regard to analogous variants across other ribozyme classes and the 8–17 DNAzyme.

G40 is found near the scene of the crime

In order to act as a general base, a guanine residue in ionized form (deprotonated at N1) must be in close proximity to the nucleophile and positioned in such a way to be able to abstract the proton from the 2′OH nucleophile. An abundance of valuable crystallographic data is available for the pistol ribozyme,[26,28,29,46] and in all available crystal structures (Table 2), the G40:N1 is within 4 Å of the C2′ – the attachment point of the 2′O nucleophile (which has been artificially removed to prevent the reaction in the crystals). In some ribozymes, the general base guanine is observed to interact with one of the NPOs of the scissile phosphate through the N2 exocyclic amine. In all available crystal structures, G40:N2 is within 4 Å of one of the NPOs (usually the pro-R_P). These distances are consistent with the corresponding distances in the HHr crystal structure[47] (Table 2). Taken together, this suggests that G40 is in a viable position to act as general base in the crystal structures, and by induction, is strongly suggestive that G40 acts in this capacity in Psr.

G40 does not fit the criminal profile

As a result of a series of studies that applied a "computational enzymology" approach,[18] often in collaboration with experimental groups,[15,32,50] we developed an L-platform/L-scaffold framework,[19] extending the L-platform motif originally described by Suslov and coworkers,[51] as a blueprint for site-specific RNA-cleaving nucleic acid enzymes that employ an active site guanine as a general base. The generic framework is illustrated in Fig. 2, as well as specific examples of ribozymes and a DNAzyme that belong to the G + M ribozyme class that contain a divalent metal ion at the active site that plays a chemical role in catalysis.[11] The "L" of the L-platform is formed from a four-residue motif, N-1 and three sequential residues from the enzyme strand, arranged such that first two residues stack on top of and the third is shifted over to base-pair N-1 forming the base of the "L" (see Fig. 2.) The conserved general base guanine is sandwiched between the first and third residues of the base stack, where it can be held vertically in position adjacent to the scissile phosphate and poised to activate the nucleophile. The L-scaffold helps to stabilize the L-platform though base-pairing interactions, and contains two key constituent elements: the L-anchor and the L-pocket. The

Table 2 Comparison of G40 and G42 active site distances (Å) to the nucleophile and scissile phosphate in crystal structures of Psr[a]

		Nucleobase of interest			
		G40		G42	
PDB ID	Ref.	N1–C2′	N2-R_P	N1–C2′	N2-R_P
5K7C	26	3.85	3.89	6.41	5.15
5KTJ	46	3.50	2.70	6.48	7.11
6R47	29	3.58	(3.93)	6.39	5.70
6UEY	28	3.23	3.35	6.39	5.73
		*2.45		*7.33	

[a] Shown are the distances between the G:N1 and N-1:C2′ in Å. The value in parentheses for 6R47 indicates the N2-S_P distance (the N2-R_P distance is 5.10 Å), which is the only case where the S_P is the NPO closest to the N2 atom. An asterisk "*" indicates distances to the O2′ (which for most structures are absent in order to deactivate the ribozyme for crystallization) only the 6UEY structure has O2′ in the crystal as it was for the vanadate transition mimic structure. For reference, the corresponding values in full-length HHr crystal structure[47] for the general base G12 are 4.24, 3.26 and 5.24 Å for N1–C2′, N1–O2′ and N2-R_P, respectively.

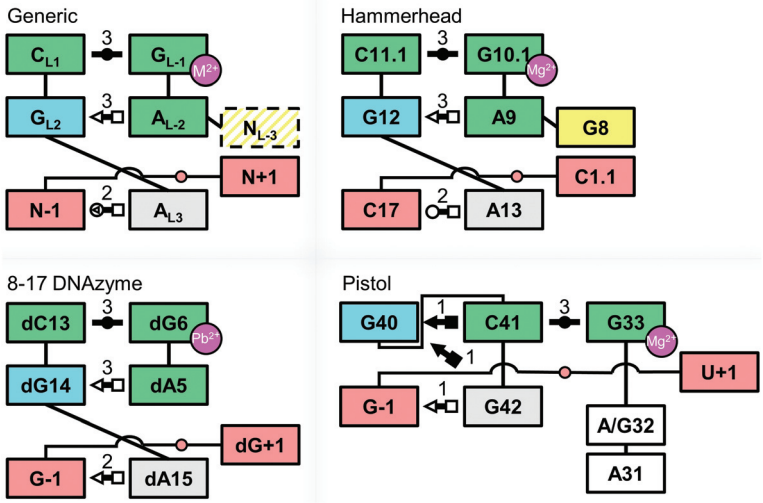

Fig. 2 Symbolic secondary structure representation of the L-platform/L-scaffold framework[19] for G + M classes of nucleic acid enzymes. A generic (ideal) representation is shown in addition to specific frameworks for the hammerhead ribozyme[47] (HHr), 8–17 DNAzyme[35] (8–17dz) and the pistol ribozyme[26,28,29,46] (Psr) based on crystallographic data. The Psr crystal structures with G40 positioned as the general base do not fit into this framework. Base-pair interactions are indicated by symbols and abbreviations for edge-to-edge base-pairing families taken from the work of Leontis and Westhof[48] and Leontis *et al.*,[49] where circle, square and triangle refer to Watson–Crick, Hoogsteen and sugar edges; and blank and filled-in depictions refer to *trans* and *cis* orientations of the nucleobases, respectively. Residues flanking the scissile phosphate are denoted in pink, general base in blue, general acid in yellow, general base anchoring and metal ion pockets in green, N-1 anchoring in gray, and residues without a role in white. Number of hydrogen bonds each base pair can make in the specified type of interaction is reported near each symbol. In the generic case, the general base is anchored by three hydrogen bonds whereas the interactions holding G40 in Psr in all crystal structures can only afford a single hydrogen bond.

L-anchor forms a non-canonical base pair with the general base guanine and anchors it in a position laterally within the base stack to activate the nucleophile. The L-pocket is a guanine residue that enables the formation of a divalent metal ion binding site through interaction at the O6/N7 positions. Structurally, the L-pocket is immediately 3′ to the L-anchor and stacks on top of it, and forms a base-pair with the residue immediately 5′ to the general base and stacked on top of it at the top of the "L".

Both the hammerhead ribozyme and 8–17 DNAzyme adhere closely to the L-platform/L-scaffold framework with minor variation (Fig. 2), however the pistol ribozyme with G40 designated as the general base does not. The L-pocket residue in Psr is G33, where the catalytic Mg^{2+} ion has been established to bind[29] as indicated by a roughly 10^4-fold decrease in rate in the G33(7cG) mutant where the N7 position was replaced by carbon to inhibit metal ion binding to that position. In the crystal structures, the L-pocket G33 forms a base pair with C41, but does not stack well with A/G32 immediately 5′ to it, which in the L-platform/L-scaffold framework should serve as the L-anchor. Instead G42 stacks midway C41 and G33. The general base G40, also not numerically 3′ to C41 as the motif has, appears to be somewhat wedged between C41 and G42. In essence, from the crystal structures, Psr does not fit the L-platform/L-scaffold framework at all.

And then we noticed Psr did not fit another pattern, this time for functional data. All the other G + M and G + A ribozymes are sensitive to 2AP modification at the general base position, with the effects ranging from 72-fold in the twister ribozyme to 8500-fold in the VS ribozyme. In the G + M class Psr is a member of, hammerhead and 8–17 DNAzyme have 1000-fold and 500-fold effects, respectively. In fact, recently the rate reduction of 2AP has been used as a tool for getting the full pH-rate profile of 8–17 DNAzyme in presence of the faster divalent metal ion Pb^{2+}, still yielding a 50-fold decrease at the fastest point for the 2AP variant at pH 6.5.[52] The G40(2AP) variant in Psr reported by Teplova *et al.*[28] not only did not significantly decrease the rate, but slightly increased it at pH 7.5 (k/k' = 0.98, highlighted as bold in Table 1). This data point was clearly a huge outlier across ribozyme classes.

Results and discussion

QM/MM simulations suggest G40(2AP) does not have the means to abstract the proton

The significant rate decrease with 2AP variants as seen in the rest of the G + M and G + A class ribozymes is expected, as the pK_a value of 2AP is 3.8, roughly the same as adenine, representing a downshift of 5.6 pK_a units from guanine which has a pK_a value of 9.4. This amounts to approximately 7.5 kcal mol^{-1} in free energy for which it would be more favorable to protonate an ionized guanine (G^-) at the N1 position as opposed to 2AP at the same position. At pH 7 (assuming unshifted pK_a values), 2AP substitution of the general base guanine would amount to an approximately 1500-fold reduction in rate. Simply put, 2AP is chemically unfit to act as an effective general base.

Enzyme environments and active sites are known to shift pK_as of nucleobases so we set to find out if Psr can shift the pK_a of G40(2AP) to be as or more favorable than the wild type G40 and explain the slight positive experimental rate effect. We calculated the free energy profile[53] and reaction barrier for the nucleophile proton abstraction by G40$^-$ and G40(2AP) through *ab initio* quantum mechanical/molecular mechanical (QM/MM) simulations for Psr in solution (Fig. 3). Our results suggest G40(2AP) to be 7.5 kcal mol^{-1} higher in free energy, and inherently unstable; *i.e.*, there is no stable minimum for the abstracted proton to be covalently bound to the N1 of G40 (2AP) as opposed to transferring back to the nucleophile. Taken together, we felt there was very compelling evidence that G40 might not act as the primary general base in Psr.

These calculations were carried out with the active site held in the active state established by experimental data,[29] with Mg^{2+} direct coordinating G33:N7 (and five water molecules), where it is far enough from the nucleophile activation step (>5.5 Å from both N1 and O2′) that it is not expected to chemically affect the nucleophile activation. For this reason the Mg^{2+} ion and the directly coordinating water molecules were not included in the QM region.

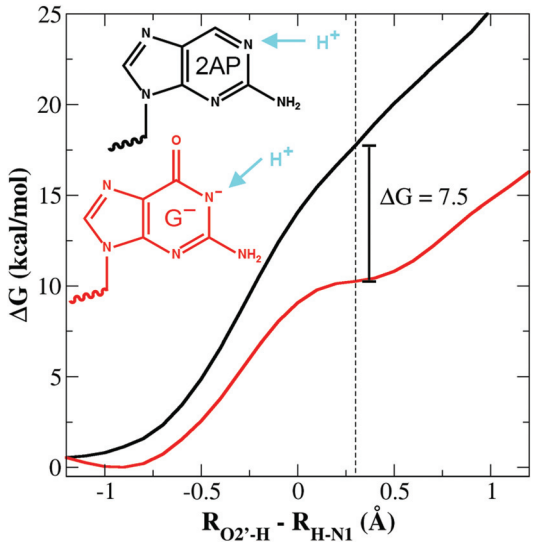

Fig. 3 Free energy profile for abstraction of the nucleophile proton by G40 and G40(2AP) along a proton transfer reaction coordinate that is the differences in the O2′–H and H–N1 distances. Reaction coordinate values <−0.5 Å correspond to proton fully bonded to O2′. The transfer occurs between −0.5 to 0.3 Å. The dotted line marks the position where the proton becomes fully bonded to the N1 (*i.e.* fully transferred), which in the case of G40(2AP) is not predicted to be a stable stationary point (meaning the proton would spontaneously transfer back to the nucleophile). Points beyond 0.3 Å correspond to activated O2′ losing the stabilizing hydrogen bond. The difference in free energy between G40 and G40(2AP) is 7.5 kcal mol^{-1}, which is consistent with the expected free energy difference estimated from relative pK_a values for guanine (9.4) and 2-aminopurine (3.8): ΔG = (9.4−3.8)/(k_BT) ≈ 7.5 kcal mol^{-1} (where k_B is the Boltzmann constant and T = 298.15 K). Profile was produced by QM/MM free energy calculations performed on Psr in solution using PBE0/6-31G* functional. See Computational methods for further details.

G42 emerges as a nucleobase of interest

If not G40 then who? If Psr were to follow the L-platform/L-scaffold motif, the general base would be immediately 3′ from C41 at position 42, which is another highly conserved guanine in Psr[24,25] and could potentially fit into the framework in the role of the general base (Fig. 2). If G42 were actually the general base in Psr, then it should be stacked with C41, and form a non-canonical base-pair with A/G32, the position immediately 5′ of and stacked under the L-pocket G33 as depicted in Fig. 4 (top center). Clearly, from the crystallographic data, this would require local refolding of the residues in the active site. We thus became intrigued by the question: *What if G42 was the general base in the pistol ribozyme?*

The first stage of our investigation involved data collection. We set out to examine what functional data had been reported for G42 to see whether it would pass a "litmus test" in terms of expected sensitivity to chemical modifications. In the paper of Ren *et al.*,[26] both G40A and G42A mutants were studied, and in both cases no activity was reported. The only other relevant G42 single variant we could find was, reported by Wilson, Lilley and co-workers,[29] for G42(2AP) which showed a large 700-fold effect on rate, and implicates either the O6 or a proton at the N1 position as having functional importance. This value also falls within the range of the G + M classes (Fig. 2). Hence, given the limited data available, G42 seemed to pass the "litmus test" as the possible general base in Psr.

We thus set out to further investigate our hypothesis by using molecular simulations to see if we could induce local refolding of the active site into a stable conformation aligned with the L-platform/L-scaffold architecture that implicated G42 as the general base.

Computer simulations predict that G42 had the motive, means and opportunity

We investigated local refolding of the Psr active site to examine whether we could achieve a stable structure that fit the L-platform/L-scaffold architecture with G42 in the role of the general base. We made rearrangements by hand and through employing distance and angle restraints in accord with the base-pair interactions of the generic L-platform/L-scaffold (details are provided in computational methods). After many revisions, and structure refinement using molecular dynamics simulations, we were able to achieve a stable structural ensemble in accord with the L-platform/L-scaffold framework that we designate as the model L-P/S (Fig. 4).

In the model L-P/S structure, G42 is positioned to act as the general base, and is stabilized by base-stacking interactions with C41 and G-1, and anchored in position by A32. The generic L-platform/L-scaffold would have the residue immediately 3′ of general base (U/G43) anchor N-1. However when we tried to achieve this in our refolding attempts we only got strained high-energy structures. Upon investigation we saw that A31 preferred to stay stacked under A32. Therefore we let A31 fall into place of anchoring G-1, and U/G43 quickly responded to stably stack under it. A31, yet another highly con-

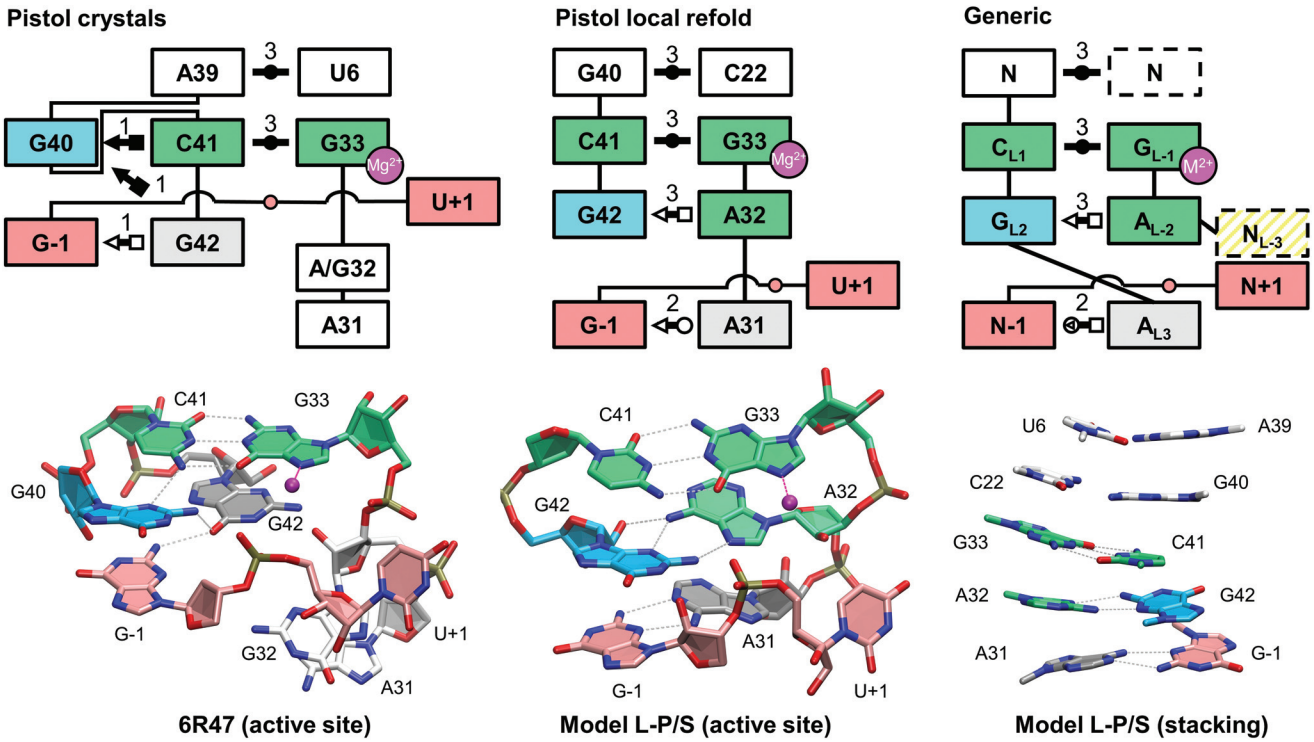

Fig. 4 Symbolic secondary (top) and 3D atomic (bottom) structure representations of the Psr active sites for crystals and model fold for L-platform/L-scaffold motif (model L-P/S). The symbolic secondary structures are extended to include additional level of stacked bases to help visualize the local refolding. In the crystal structures (left) G40 is most closely positioned to act as the general base. Proposed theoretical model from molecular dynamics simulations (middle) depict a locally re-folded active site that fits the generic L-platform/L-scaffold architecture (right), and G42 plays the role of the general base. The model L-P/S structure is the average of last 10 ns of simulation. A rotated 3D structure is also included to show the stacking of the residues including the G40–C22 and A39–U6 base pairs atop the L-platform/L-scaffold.

served residue in Psr,[24,25] was not positioned in the crystals to suggest any essential structural or functional role that would explain its conservation. Having A31 anchor G-1 also resulted in maintaining the crystallographically observed enzyme substrate base pairing for U/G43 and all other substrate residues, with the exception of G-1. This refolding tidied up the active site and is in agreement with L-platform/L-scaffold framework. But what about G40?

We looked into finding an alibi for G40 and saw three possible options for its whereabouts: (1) maintaining its relative position to C41 and G42 from the crystals, (2) stacking between A39 and C41 as would be its position in a stem, and (3) forming AUG base-triple with A39 and U6. Only the first two were stable. In the first position the exocyclic amine of G40 hydrogen bonds to S_P of the scissile phosphate. In this position, one might expect this hydrogen bond to play a role in β catalysis to stabilize the charge on the phosphate, but thio substitution experiments at S_P indicate no such effect is observed,[29] suggesting any hydrogen bond (or direct Mg^{2+} ion coordination) at that site is not essential. In the second position, when stacking between A39 and C41 and alone, the O6 of G40 formed an outer-sphere contact with the Mg^{2+} ion. In this scenario, one might infer that the O6 plays an important role in forming the Mg^{2+} ion binding pocket, but again experiments indicate when O6 is replaced with S (6sG), Psr is actually faster,[29] again suggesting there is no hydrogen bond or direct Mg^{2+} ion coordination at the O6 position that is essential for activity.

We then increased our search perimeter, and looked for other residues that in the crystals seemed to not have a key structural or functional role. In *all* of the available crystal structures the position 22 (C or U) right after the highly conserved A-minor motif (A19, A20, A21),[24,25] was unpaired and pointing out to the solvent. The possibility of forming canonical G–C (or non-canonical G–U) base pairs between the Watson–Crick edges of these two positions (*i.e. cis* Watson–Crick/Watson–Crick[49]) was very intriguing. By visual inspection it was not evident that this C22 residue could move over and stably base-pair with G40 without affecting the A-minor motif or disrupting the G5·C12·G24 base-triple where G5 and R24 (G or C) are also conserved.[24,25] To our pleasant surprise, we were able to find a stable structure with G40–C22 base paired and stacked between A39–U6 and C41–G33 pairs. Initially the A-minor motif seemed perturbed with A21 pointing out at the solvent, but it was able to reorient and point back at the stem. With C22 changing orientation, A23 also had to change in concert, leaving the stable A-minor motif intact. Despite all this rearrangement, the model L-P/S structure we arrived at maintains all the important structural motifs seen in the crystals, tidies up the active site, and leaves no nucleobase

unpaired. It is remarkable that Psr has the flexibility to allow such refolding to form a new stable active site structure that does not disrupt the rest of the architecture.

Critical assessment of the model. It should be noted that for achieving these local structural rearrangements we applied a manual knowledge-based approach through use of experimentally motivated distance and angle restraints followed by MD refinement. In this way, we have used restraints as a way to guide the system to form the interactions of the L-platform/L-scaffold framework.[19] The use of half harmonic potentials (*i.e.* one sided penalty) gave the system the flexibility to find favorable free energy basins that satisfied the restraint conditions, oftentimes not incurring any restraint penalty by the end of the simulation. Care was given to ensure the structure we arrived at was a local free energy minima by running multiple independent unrestrained simulations for 100–200 ns to check that the system was in a stable fold (full details are provided in the ESI†). Nonetheless, it should be emphasized that the structural re-arrangement did not occur spontaneously, and the final structure we describe herein not only depends on the force field models employed (described in Computational details), but may also not be the only structure that satisfies the restraint conditions imposed.

What if … G40 and G42 had a conspiracy and there are *two available general base mechanisms*?

Up until this point, there are two suspects, G40 and G42, implicated as the general base in the pistol ribozyme. A summary of the *pros* and *cons* in the case against each suspect is summarized as follows.

G40 as the general base. • *Pros*: G40 is found in all the crystal structures at the scene of the crime, and mutational data and activity-pH profiles are largely consistent with its role as a general base.

• *Cons*: G40(2AP) variant has slightly increased rate, a distinct outlier across G + M and G + A ribozyme classes that is further contested by QM/MM simulations presented here, for which no clear mechanistic explanation is apparent.

G42 as the general base. • *Pros*: G42(2AP) variant has large effect on rate consistent with other G + M and G + A ribozyme classes, fits the theoretically predicted L-platform/L-scaffold design framework, and is supported by simulations that identify a stable refolded active site where G42 is positioned to act as general base.

• *Cons*: G42 is not observed crystallographically at the scene of the crime.

There is thus compelling evidence that argues for either suspect nucleobase to act as the general base in the reaction, begging the question: *Is there a single explanation that fits all the currently available data?*

The wealth of high-quality and consistent structural data available demonstrates beyond a reasonable doubt that there exists a stable thermally accessible active site structure where G40 is in a position to act as general base. Nonetheless, RNA is known to exhibit a high degree of conformational heterogeneity,[54–57] and the active state of ribozymes capable of performing catalysis are often not the most probable ground states in solution.[18,31,32,58] Simulation results presented here support the supposition that there exists a plausible alternative active site structure that positions G42 as the general base, fits the L-platform/L-scaffold framework and has a catalytically relevant occupancy in solution. In this scenario, it is possible that both states are accessible and can contribute to catalysis *via* different pathways. The 2AP experiment results for G40 and G42 suggest that the primary (most efficient) pathway leads through the predicted L-scaffold/L-platform state with G42 acting as general base. This does not, however, preclude the possibility that the crystallographically observed state can also convey catalytic activity with a reduced rate through a different pathway where G40 acts as the general base.

In this scenario, G40 and G42 may be conspirators in establishing multiple routes to general base abstraction of the nucleophile proton. A rough free energy estimate of 2AP substitution for the general base guanine based on assumed unshifted pK_a values (ΔpK_a = 9.4–3.8 = 5.6, adjusted to pH 7 is 7–3.8 = 3.2) would suggest a roughly 1500-fold effect on rate at pH 7. The insensitivity of the observed rate to the G40(2AP) variant is consistent with the G40 pathway not being the primary catalytic route for the wild type under normal conditions. The observed 700-fold effect of G42(2AP) substitution[29] is somewhat lower than the corresponding substitution in VSr (8500-fold) and HHr (1000-fold). This observation is consistent with the supposition that the G42(2AP) variant might eliminate the primary G42 pathway, enabling the crystallographically observed G40 pathway to become rate-controlling (albeit with reduced rate). While this "conspiracy theory" is more complicated than a single dominant pathway, it is plausible, is consistent with all existing data, and can be experimentally tested as described below.

Burden of proof is upon experiments to demonstrate G42's guilt beyond a reasonable doubt

Having achieved what we feel is a plausible and stable active site structure that fits the L-platform/L-scaffold framework and positions G42 to play the role of general base, we turn to experiments to ultimately confirm or refute our hypothesis that G42 acts as the general base in Psr either as the sole pathway, or as the primary pathway with an alternative secondary pathway available with reduced rate that is consistent with crystallographic data where G40 plays the role of the general base. Toward this end, we suggest several experiments with predicted outcomes that may distinguish these different scenarios and shed light into this mechanistic mystery.

2′-Bromoacetamide affinity label experiments. Affinity labels have been used in the past to study general base catalysis in RNase A,[59] and later adapted by Thomas and Perrin[60] for ribozymes and applied to the hairpin ribozyme[60] and hammerhead[61] ribozymes. In this technique, an electrophilic 2′-bromoacetamide affinity label is designed to replace the nucleophilic 2′OH group. This affinity label can undergo an alkylation reaction with a nearby Lewis base, the most likely candidate being the catalytic general base, which is subsequently identified by

footprinting analysis. Hence, such affinity label experiments on the pistol ribozyme would enable the determination of whether G40 and/or G42 became alkylated at the N1 position. If there is a clear discrimination, this strongly implicates the alkylated residue as the general base. Alternatively, if both alkylated products are observed, it would support the hypothesis that both G40 and G42 pathways are available.

Activity-pH profiles and linear free energy relationships (LFER). Measurement of the activity-pH profiles, *i.e.*, $\log(k_{obs})$ *versus* pH, for acid–base catalysts provide insight into the nature of the protonation events that are correlated with rate.[62–64] The fundamental assumption in the interpretation of activity-pH profiles is that the rate is proportional to the probability of observing the system in an "active state", which due to the protonation state requirements of the general acid and base, is pH-dependent. Data for k_{obs} can be fitted to simple single or double ionization models[63] (or more rigorous models[18,64]) to determine "apparent pK_a" values, the most straight-forward interpretation of which is that they should be closely related to the actual pK_a values of the general acid and base. Hence, systematic modulation of the pK_a of the general base residue should correlate with the associated changes in the apparent pK_a values. Among the most subtle modifications are "isofunctional" purine analogs that preserve the guanine Watson–Crick face and exocyclic substituents, but provide a range of pK_a values.[16] Ideally, although often not practical, an additional 5′-thio substitution is made at the leaving group position (introducing an enhanced leaving group) so as to ensure that the rate-controlling step of the reaction is shifted to the general base step.

In this case, the intrinsic rate constant should depend log-linearly on the pK_a of the general base.[65] This is known as a "linear free energy relationship" (LFER). A plot of k_{cat} *versus* pK_a of the modified general base yields a so-called Brønsted plot,[66] the slope of which corresponds to a Brønsted β coefficient that measures the sensitivity of the reaction to base strength and reflects the extent of proton transfer in the rate-controlling transition state.[67] Hence, the measurement of activity-pH profiles and LFERs using isofunctional purine analogs to substitute alternatively for G40 and G42 in Psr, and identification of which of these positions exhibits the expected pH sensitivity, would provide convincing evidence that this residue acts as the general base.

In the scenario that both G40 and G42 pathways are available, the interpretation of activity-pH and LFER data becomes more complicated in that at some point, there can be an inversion of the G42 and G40 pathways as being rate-controlling. For example, in the case of 2AP substitution, if there were only one general base, one might expect a greatly reduced rate accompanied by a large down-shift of the apparent pK_a such as is observed in the 8–17 DNAzyme[34] where the wild type apparent pK_a is shifted from 9.0 to 5.5 in the G14(2AP) variant. Alternatively, if there is an alternative pathway available through G40, one might expect that the shape of the activity-pH profile for G42(2AP) to remain largely the same, but with correspondingly lower k_{obs} values. This could be further explored by examining the pH-dependent kinetics of the G40(2AP)/G42(2AP) double substitution. If these modifications are independent, one would expect the result to closely resemble that of the G42(2AP) variant since the G40(2AP) variant had negligible effect (0.98). However, if the observed 700-fold effect was due to a secondary G40 pathway, then one would expect the double substitution to have a greater effect on the rate, and possibly force the reaction to proceed through a specific base mechanism. Hence, more detailed functional studies that quantitatively examine activity-pH profiles for doubly substituted variants would produce important new insight into the Psr mechanism.

Further experiments. It is possible that further experiments such as the measurement of kinetic isotope effects of isotopically labelled guanine (*e.g.*, ^{18}O at the O6 position and ^{15}N at the N1 position) at G40 and alternatively at G42 might provide aditional insight. These atomic positions would be expected to undergo changes in bonding environment upon protonation/deprotonation at the N1 position if acting as a general base in the transition state, and thus exhibit a kinetic isotope effect.

Conclusion

Having investigated a series of experimental and computational clues, we are ready to make our final accusation as to the crime of "who stole the proton in the pistol ribozyme?":

We accuse G42 of abstracting the proton in the model L-platform/L-scaffold active site using a pK_a tuned N1.

There remains solid structural and functional evidence that support the generally accepted alternative that G40 plays the role of the general base. However, this evidence remains circumstantial and is called into question by the lack of sensitivity to 2AP substitution at the G40 position (having large downshift in pK_a at the N1 position,) that is inconsistent with the rates of analogous variants across G + M and G + A ribozyme and 8–17 DNAzyme classes, and the instability and barriers predicted by QM/MM calculations reported in this study. Our L-platform/L-scaffold design framework points to G42 as a plausible general base, which is supported by preliminary functional data and computational models presented herein. Specifically, we show that local refolding of the active site can lead to a stable architecture that fits the L-platform/L-scaffold framework, and anchors G42 in a position to act as general base. Further, quantum mechanical simulations suggest that $G40^-$ is able to abstract the nucleophile proton with significantly lower barrier than the corresponding 2AP variant, consistent with the expected G to 2AP pK_a shift. Ultimately, the hypothesis that is consistent with all the current data is that there are two available catalytic pathways enabled by different thermally accessible conformational states connected by a local refolding of the active site: (1) a primary pathway with an active site architecture aligned with the L-platform/L-scaffold framework where G42 acts as a general base, and (2) a secondary pathway with the crystallographic active site architecture where G40 acts as a general base. Our computationally tested

hypothesis suggests a number of experimentally testable predictions that, it is our hope, might inspire new collaborative experimental/theoretical work in order to confirm or refute. In this way, we hope to prove beyond a reasonable doubt that both G42 and G40 conspired to commit the crime, and solve the mystery as to who stole the proton in the pistol ribozyme.

Computational methods

Classical molecular dynamics simulations

Pistol system was built departing from 6R47 crystal structure.[29] Residue 33, reported as A in the PDB, was changed to G as is in the crystal sequence. C30 was mutated to U30 as that position is observed as U in >90% of the sequences.[24] G32 was mutated to A, as most of the mutational studies are performed with A in position 32. The nucleophile, which was not explicitly present in the crystal structure, was added based on internal coordinates of the sugar ring. Ions were added to balance the system charge and achieve a bulk ion concentration of 0.14 M NaCl. The solvated system was equilibrated (as described in ref. 17), and simulated for 100 ns, for overall system relaxation in solution.

In simulations investigating the local refolding, G42 was deprotonated at N1 position to enable productive hydrogen bonding with the nucleophile. In all simulations nucleophile to G42:N1 hydrogen bond, and O2′–P–O5′ inline angle were restrained to maintain an active state in order to arrive at a fold that can accommodate it. To influence local refolding, distance and angle restraints were employed between various residues to achieve base pairing and stacking, including but not limited to G42–A32 and G40–C22 base pairs and G40 C41 and G42 stacking. Additional restraints were used to maintain significant base pairs in the structure such as C41–G33. A complete list of restraints and simulation steps taken to achieve the model L-P/S fold are available in ESI,† as well as a separate pdb file for the final structure, an average of the last 10 ns, with residue numbers corresponding to the 6R47 crystal structure.[29]

The classical MD simulations were carried out using AMBER20,[68] employing ff99OL3 RNA force field,[69,70] TIP4P/Ew water model[71] and corresponding Na^+/Cl^- ions[72] and 12-6-4 [73] Mg^{2+} ions specifically tuned to have balanced interactions between solvent and nucleic acids.[74,75] The RNA force field has been reviewed and shown to be remakably robust,[76] and the Mg^{2+} ion parameters are able to reproduce structural, thermodynamic, kinetic and mass transport properties in bulk solution,[74] as well as experimental site specific binding free energies to nucleic acids derived from potentiometric pH titration data.[75]

Simulations were performed under periodic boundary conditions at 300 K using an 12 Å nonbond cutoff and PME electrostatics.[77,78] The Langevin thermostat with $\gamma = 5$ ps^{-1} and Berendsen isotropic barostat with $\tau = 1$ ps were used to maintain a constant pressure and temperature. A 1 fs time step was used along with the SHAKE algorithm to fix hydrogen bond lengths.[79]

QM/MM simulations

Nucleophile activation proton transfer reaction was investigated using the crystal fold of the pistol ribozyme, with G40 positioned to act as the base, departing from the solvent relaxed structure described above. Two systems, $G40^-$ and G40 (2AP) were created. The QM region comprised 30 atoms including: the scissile phosphate, N-1 sugar, and the nucleobase of residue 40 as G^- or 2AP. The QM region was treated with the PBE0/6-31G* hybrid density functional method.[80,81] The remainder of the system was treated with the molecular mechanical force field, described above.

The reaction coordinate was defined as the difference in distances $R_{H–O2'}$–$R_{N1–H}$, and was evaluated from −1.2 to 1.2 Å with 0.1 Å steps yielding total of 25 umbrella window QM/MM simulations. The production simulations were performed at constant volume and temperature (298 K) using a Langevin thermostat. A 100 kcal mol^{-1} Å^{-2} umbrella potential force constant was used, and each simulation was performed for 20 ps using a 1 fs timestep. The reaction coordinate was saved every 25 frames. The Lennard-Jones potential was cutoff at 10 Å and a long-range tail correction was used to model the LJ interactions beyond the cutoff. The ambient potential composite Ewald method was used to compute the *ab initio* QM/MM electrostatic interactions.[82] The DFT molecular quadrature grid is a union of atomic grids composed of Laguerre radial and Lebedev angular quadrature points.[83–87] The atomic grids of hydrogen are composed of 30 radial shells, whose largest radius is 12 Å, and the number of angular points vary as a function of radius to yield a net total of 4272 atomic grid points. The heavy atoms are composed of 60 radial shells, whose largest radius is 16 Å, and the angular points are pruned to yield 5580 atomic grid points. The molecular quadrature grid is the union of atomic grids partitioned using the Becke's decomposition scheme.[88] These grids have been shown to adequately conserve energy in PBE0/6-31G* QM/MM simulations using a 1 fs timestep.[82] The free energy surfaces were generated by analyzing the reaction coordinate time series using the variational free energy profile (vFEP) method[89,90] using in the FE-ToolKit software.[53]

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors thank Timothy Wilson and David Lilley for numerous insightful discussions, and Virgil Yumuşak-Kabuklu Kaplumbağa for encouragement and moral support. The authors are grateful for financial support provided by the National Institutes of Health (No. GM62248 to DMY), and to Alan Grossman for providing early-stage seed support for our research through the Grossman Innovation Prize. Computational resources that have contributed to the research

results reported within this work were provided by the Office of Advanced Research Computing (OARC) at Rutgers, the State University of New Jersey, the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation Grant ACI-1548562 (specifically, the resources EXPANSE and COMET at SDSC through allocation TG-CHE190067), and the Texas Advanced Computing Center (TACC, **https://www.tacc.utexas.edu**) at The University of Texas at Austin, specifically the Frontera Supercomputer.

References

1 T. R. Cech, A. J. Zaug and P. J. Grabowski, *In vitro* splicing of the ribosomal RNA precursor of *Tetrahymena*: involvement of a guanosine nucleotide in the excision of the intervening sequence, *Cell*, 1981, **27**, 487–496.
2 K. Kruger, P. Grabowsky, A. Zaug, J. Sands, D. Gottschling and T. Cech, Self-splicing RNA: autoexcision and autocyclisation of the ribosomal RNA intervening sequence of *Tetrahymena*, *Cell*, 1982, **31**, 147–157.
3 C. Guerrier-Takada, K. Gardiner and T. Maresh, The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme, *Cell*, 1983, **35**, 849–857.
4 R. R. Breaker and G. F. Joyce, A DNA enzyme that cleaves RNA, *Chem. Biol.*, 1994, **1**, 223–229.
5 C. M. Kim and C. D. Smolke, Biomedical applications of RNA-based devices, *Curr. Opin. Biomed. Eng.*, 2017, **4**, 106–115.
6 R. R. Breaker and G. F. Joyce, The Expanding View of RNA and DNA Function, *Chem. Biol.*, 2014, **21**, 1059–1065.
7 T. R. Cech, The RNA Worlds in Context, *Cold Spring Harbor Perspect. Biol.*, 2012, **4**, 006742–006742.
8 J. D. Sutherland, Opinion: Studies on the origin of life—the end of the beginning, *Nat. Rev. Chem.*, 2017, **1**, 0012.
9 W. L. Ward, K. Plakos and V. J. DeRose, Nucleic acid catalysis: metals, nucleobases, and other cofactors, *Chem. Rev.*, 2014, **114**, 4318–4342.
10 D. M. J. Lilley, How RNA acts as a nuclease: some mechanistic comparisons in the nucleolytic ribozymes, *Biochem. Soc. Trans.*, 2017, **45**, 683–691.
11 D. M. J. Lilley, Classification of the nucleolytic ribozymes based upon catalytic mechanism, *F1000Research*, 2019, **8**, 1462.
12 M. Garcia-Viloca, J. Gao, M. Karplus and D. G. Truhlar, How enzymes work: Analysis by modern rate theory and computer simulations, *Science*, 2004, **303**, 186–195.
13 H. Gu, S. Zhang, K.-Y. Wong, B. K. Radak, T. Dissanayake, D. L. Kellerman, Q. Dai, M. Miyagi, V. E. Anderson, D. M. York, *et al.*, Experimental and computational analysis of the transition state for ribonuclease A-catalyzed RNA 2′-*O*-transphosphorylation, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 13002–13007.
14 M. E. Harris, J. A. Piccirilli and D. M. York, Enzyme transition states from theory and experiment, *Biochim. Biophys. Acta*, 2015, **1854**, 1727–1728. PMID: 26302659.
15 P. C. Bevilacqua, M. E. Harris, J. A. Piccirilli, C. Gaines, A. Ganguly, K. Kostenbader, Ş. Ekesan and D. M. York, An Ontology for Facilitating Discussion of Catalytic Strategies of RNA-Cleaving Enzymes, *ACS Chem. Biol.*, 2019, **14**, 1068–1076.
16 A. Ganguly, B. P. Weissman, T. J. Giese, N.-S. Li, S. Hoshika, R. Saieesh, S. A. Benner, J. A. Piccirilli and D. M. York, Confluence of theory and experiment reveals the catalytic mechanism of the Varkud satellite ribozyme, *Nat. Chem.*, 2020, **12**, 193–201.
17 Ş. Ekesan and D. M. York, Dynamical ensemble of the active state and transition state mimic for the RNA-cleaving 8-17 DNAzyme in solution, *Nucleic Acids Res.*, 2019, **47**, 10282–10295.
18 C. S. Gaines, T. J. Giese and D. M. York, Cleaning Up Mechanistic Debris Generated by Twister Ribozymes Using Computational RNA Enzymology, *ACS Catal.*, 2019, **9**, 5803–5815.
19 C. S. Gaines, J. A. Piccirilli and D. M. York, The L-platform/L-scaffold framework: a blueprint for RNA-cleaving nucleic acid enzyme design, *RNA*, 2020, **26**, 111–125.
20 D. L. Kellerman, D. M. York, J. A. Piccirilli and M. E. Harris, Altered (transition) states: mechanisms of solution and enzyme catalyzed RNA 2′-O-transphosphorylation, *Curr. Opin. Chem. Biol.*, 2014, **21**, 96–102.
21 H. Chen, T. J. Giese, M. Huang, K.-Y. Wong, M. E. Harris and D. M. York, Mechanistic Insights into RNA Transphosphorylation from Kinetic Isotope Effects and Linear Free Energy Relationships of Model Reactions, *Chem. – Eur. J.*, 2014, **20**, 14336–14343.
22 G. M. Emilsson, S. Nakamura, A. Roth and R. R. Breaker, Ribozyme speed limits, *RNA*, 2003, **9**, 907–918.
23 R. R. Breaker, G. M. Emilsson, D. Lazarev, S. Nakamura, I. J. Puskarz, A. Roth and N. Sudarsan, A common speed limit for RNA-cleaving ribozymes and deoxyribozymes, *RNA*, 2003, **9**, 949–957.
24 Z. Weinberg, P. B. Kim, T. H. Chen, S. Li, K. A. Harris, C. E. Lünse and R. R. Breaker, New classes of self-cleaving ribozymes revealed by comparative genomics analysis, *Nat. Chem. Biol.*, 2015, **11**, 606–610.
25 K. A. Harris, C. E. Lünse, S. Li, K. I. Brewer and R. R. Breaker, Biochemical analysis of pistol self-cleaving ribozymes, *RNA*, 2015, **21**, 1852–1858.
26 A. Ren, N. Vusurovic, J. Gebetsberger, P. Gao, M. Juen, C. Kreutz, R. Micura and D. Patel, Pistol Ribozyme Adopts a Pseudoknot Fold Facilitating Site-specific In-line Cleavage, *Nat. Chem. Biol.*, 2016, **12**, 702–708.
27 S. Neuner, C. Falschlunger, E. Fuchs, M. Himmelstoss, A. Ren, D. J. Patel and R. Micura, Atom-Specific Mutagenesis Reveals Structural and Catalytic Roles for an Active-Site Adenosine and Hydrated Mg^{2+} in Pistol Ribozymes, *Angew. Chem., Int. Ed.*, 2017, **56**, 15954–15958.
28 M. Teplova, C. Falschlunger, O. Krasheninina, M. Egger, A. Ren, D. J. Patel and R. Micura, Crucial Roles of Two Hydrated Mg^{2+} Ions in Reaction Catalysis of the Pistol Ribozyme, *Angew. Chem., Int. Ed.*, 2020, **59**, 2837–2843.

29 T. J. Wilson, Y. Liu, N. S. Li, Q. Dai, J. A. Piccirilli and D. M. Lilley, Comparison of the structures and mechanisms of the pistol and hammerhead ribozymes, *J. Am. Chem. Soc.*, 2019, **141**, 7865–7875.
30 Y. Lihanova and C. E. Weinberg, Biochemical analysis of cleavage and ligation activities of the pistol ribozyme from Paenibacillus polymyxa, *RNA Biol.*, 2021, **18**, 1858–1866.
31 K. Kostenbader and D. M. York, Molecular simulations of the pistol ribozyme: unifying the interpretation of experimental data and establishing functional links with the hammerhead ribozyme, *RNA*, 2019, **25**, 1439–1456.
32 A. Ganguly, B. P. Weissman, J. A. Piccirilli and D. M. York, Evidence for a Catalytic Strategy to Promote Nucleophile Activation in Metal-Dependent RNA-Cleaving Ribozymes and 8-17 DNAzyme, *ACS Catal.*, 2019, **9**, 10612–10617.
33 N. N. Joseph, R. N. Roy and T. A. Steitz, Molecular dynamics analysis of Mg^{2+}-dependent cleavage of a pistol ribozyme reveals a fail-safe secondary ion for catalysis, *J. Comput. Chem.*, 2020, **41**, 1345–1352.
34 M. Cepeda-Plaza, C. E. McGhee and Y. Lu, Evidence of a general acid–base catalysis mechanism in the 8–17 DNAzyme, *Biochemistry*, 2018, **57**, 1517–1522.
35 H. Liu, X. Yu, Y. Chen, J. Zhang, B. Wu, L. Zheng, P. Haruehanroengra, R. Wang, S. Li, J. Lin, *et al.*, Crystal Structure of an RNA-Cleaving DNAzyme, *Nat. Commun.*, 2017, **8**, 2006–2015.
36 M. Cepeda-Plaza and A. Peracchi, Insights into DNA catalysis from structural and functional studies of the 8–17 DNAzyme, *Org. Biomol. Chem.*, 2020, **18**, 1697–1709.
37 N. Serrano-Aparicio, K. Świderek, I. Tuñón, V. Moliner and J. Bertran, Theoretical Studies of the Self Cleavage Pistol Ribozyme Mechanism, *Top. Catal.*, 2022, **65**, 505–516.
38 L. Casalino, G. Palermo, U. Rothlisberger and A. Magistrato, Who activates the nucleophile in ribozyme catalysis? An answer from the splicing mechanism of group II introns, *J. Am. Chem. Soc.*, 2016, **138**, 10374–10377.
39 A. Ganguly, P. Thaplyal, E. Rosta, P. C. Bevilacqua and S. Hammes-Schiffer, Quantum Mechanical/Molecular Mechanical Free Energy Simulations of the Self-Cleavage Reaction in the Hepatitis Delta Virus Ribozyme, *J. Am. Chem. Soc.*, 2014, **136**, 1483–1496.
40 R. Pinard, K. J. Hampel, J. E. Heckman, D. Lambert, P. A. Chan, F. Major and J. M. Burke, Functional involvement of G8 in the hairpin ribozyme cleavage mechanism, *EMBO J.*, 2001, **20**, 6434–6442.
41 Y. Liu, T. J. Wilson, S. A. McPhee and D. M. J. Lilley, Crystal structure and mechanistic investigation of the twister ribozyme, *Nat. Chem. Biol.*, 2014, **10**, 739–744.
42 D. Jaikaran, M. D. Smith, R. Mehdizadeh, J. Olive and R. A. Collins, An important role of G638 in the cis-cleavage reaction of the Neurospora VS ribozyme revealed by a novel nucleotide analog incorporation method, *RNA*, 2008, **14**, 938–949.
43 A. Mir, J. Chen, K. Robinson, E. Lendy, J. Goodman, D. Neau and B. L. Golden, Two Divalent Metal Ions and Conformational Changes Play Roles in the Hammerhead Ribozyme Cleavage Reaction, *Biochemistry*, 2015, **54**, 6369–6381.
44 A. Peracchi, M. Bonaccio and M. Clerici, A mutational analysis of the 8–17 deoxyribozyme core, *J. Mol. Biol.*, 2005, **352**, 783–794.
45 K. F. Blount and O. C. Uhlenbeck, The Structure-Function Dilemma of the Hammerhead Ribozyme, *Annu. Rev. Biophys. Biomol. Struct.*, 2005, **34**, 415–440.
46 L. A. Nguyen, J. Wang and T. A. Steitz, Crystal structure of Pistol, a class of self-cleaving ribozyme, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, 1021–1026.
47 M. Martick, T.-S. Lee, D. M. York and W. G. Scott, Solvent structure and hammerhead ribozyme catalysis, *Chem. Biol.*, 2008, **15**, 332–342.
48 N. B. Leontis and E. Westhof, Geometric nomenclature and classification of RNA base pairs, *RNA*, 2001, **7**, 499–512.
49 N. B. Leontis, J. Stombaugh and E. Westhof, The non-watson-crick base pairs and their associated isostericity matrices, *Nucleic Acids Res.*, 2002, **30**, 3497–3531.
50 R. N. Sengupta, S. N. Van Schie, G. Giambaşu, Q. Dai, J. D. Yesselman, D. York, J. A. Piccirilli and D. Herschlag, An active site rearrangement within the Tetrahymena group I ribozyme releases nonproductive interactions and allows formation of catalytic interactions, *RNA*, 2016, **22**, 32–48.
51 N. B. Suslov, S. DasGupta, H. Huang, J. R. Fuller, D. M. J. Lilley, P. A. Rice and J. A. Piccirilli, Crystal Structure of the Varkud Satellite Ribozyme, *Nat. Chem. Biol.*, 2015, **11**, 840–846.
52 C. Cortés-Guajardo, F. Rojas-Hernández, R. Paillao-Bustos and M. Cepeda-Plaza, Hydrated metal ion as a general acid in the catalytic mechanism of the 8-17 DNAzyme, *Org. Biomol. Chem.*, 2021, **19**, 5395–5402.
53 T. J. Giese, Ş. Ekesan and D. M. York, 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*, 2021, **125**, 4216–4232.
54 S. Alam, V. Grum-Tokars, J. Krucinska, M. L. Kundracik and J. E. Wedekind, Conformational Heterogeneity at Position U37 of an All-RNA Hairpin Ribozyme with Implications for Metal Binding and the Catalytic Structure of the S-Turn, *Biochemistry*, 2005, **44**, 14396–14408.
55 L. Zhao and T. Xia, Probing RNA conformational dynamics and heterogeneity using femtosecond time-resolved fluorescence spectroscopy, *Methods*, 2009, **49**, 128–135.
56 H. Shi, A. Rangadurai, H. A. Assi, R. Roy, D. A. Case, D. Herschlag, J. D. Yesselman and H. M. Al-Hashimi, Rapid and accurate determination of atomistic RNA dynamic ensemble models using NMR and structure prediction, *Nat. Commun.*, 2020, **11**, 5531.
57 L. Levintov and H. Vashisth, Role of conformational heterogeneity in ligand recognition by viral RNA molecules, *Phys. Chem. Chem. Phys.*, 2021, **23**, 11211–11223.
58 Ş. Ekesan and D. M. York, Framework for conducting and analyzing crystal simulations of nucleic acids to aid in modern force field evaluation, *J. Phys. Chem. B*, 2019, **123**, 4611–4624.

59 C. F. Hummel, M. R. Pincus, P. W. Brandt-Rauf, G. M. Frei and R. P. Carty, Reaction of (bromoacetamido)nucleoside affinity labels with ribonuclease A: evidence for steric control of reaction specificity and alkylation rate, *Biochemistry*, 1987, **26**, 135–146.
60 J. M. Thomas and D. M. Perrin, Active Site Labeling of G8 in the Hairpin Ribozyme: Implications for Structure and Mechanism, *J. Am. Chem. Soc.*, 2006, **128**, 16540–16545.
61 J. M. Thomas and D. M. Perrin, Probing general acid catalysis in the hammerhead ribozyme, *J. Am. Chem. Soc.*, 2009, **131**, 1135–1143.
62 P. C. Bevilacqua, Mechanistic considerations for general acid-base catalysis by RNA: Revisiting the mechanism of the hairpin ribozyme, *Biochemistry*, 2003, **42**, 2259–2265.
63 S. Kath-Schorr, T. J. Wilson, N.-S. Li, J. Lu, J. A. Piccirilli and D. M. J. Lilley, General Acid-Base Catalysis Mediated by Nucleobases in the Hairpin Ribozyme, *J. Am. Chem. Soc.*, 2012, **134**, 16717–16724.
64 E. A. Frankel and P. C. Bevilacqua, Complexity in pH-Dependent Ribozyme Kinetics: Dark pK_a Shifts and Wavy Rate-pH Profiles, *Biochemistry*, 2018, **57**, 483–488.
65 W. P. Jencks, General acid-base catalysis of complex reactions in water, *Chem. Rev.*, 1972, **72**, 705–718.
66 W. P. Jencks, A Primer for the Bema Hapothle. An Empirical Approach to the Characterization of Changing Transition-State Structures, *Chem. Rev.*, 1985, **85**, 511–527.
67 E. V. Anslyn and D. A. Dougherty, *Modern Physical Organic Chemistry*, University Science Books, Sausalito, CA, 2006.
68 D. A. Case, K. Belfon, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham III, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke and G. Giambasu, *et al.*, *AMBER 20*, University of California, San Francisco, San Francisco, CA, 2020.
69 A. Pérez, I. Marchán, D. Svozil, J. Sponer, T. E. Cheatham III, C. A. Laughton and M. Orozco, Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers, *Biophys. J.*, 2007, **92**, 3817–3829.
70 M. Zgarbová, M. Otyepka, J. Šponer, A. Mládek, P. Banáš, T. E. Cheatham III and P. Jurečka, Refinement of the Cornell *et al.* nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles, *J. Chem. Theory Comput.*, 2011, **7**, 2886–2902.
71 H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura and T. Head-Gordon, Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew, *J. Chem. Phys.*, 2004, **120**, 9665–9678.
72 I. S. Joung and T. E. Cheatham III, Determination of alkali and, halide monovalent ion parameters for use in explicitly solvated biomolecular simulations, *J. Phys. Chem. B*, 2008, **112**, 9020–9041.
73 P. Li, B. P. Roberts, D. K. Chakravorty and K. M. Merz Jr., Rational design of Particle Mesh Ewald compatible Lennard-Jones parameters for +2 metal cations in explicit solvent, *J. Chem. Theory Comput.*, 2013, **9**, 2733–2748.
74 M. T. Panteva, G. M. Giambaşu and D. M. York, Comparison of structural, thermodynamic, kinetic and mass transport properties of Mg^{2+} ion models commonly used in biomolecular simulations, *J. Comput. Chem.*, 2015, **36**, 970–982.
75 M. T. Panteva, G. M. Giambasu and D. M. York, Force Field for Mg^{2+}, Mn^{2+}, Zn^{2+}, and Cd^{2+} Ions that have Balanced Interactions with Nucleic Acids, *J. Phys. Chem. B*, 2015, **119**, 15460–15470.
76 M. A. Ditzler, M. Otyepka, J. Sponer and N. G. Walter, Molecular dynamics and quantum mechanics of RNA: conformational and chemical change we can believe in, *Acc. Chem. Res.*, 2010, **43**, 40–47.
77 T. Darden, D. York and L. Pedersen, Particle mesh Ewald: An N log(N) method for Ewald sums in large systems, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
78 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, L. Hsing and L. G. Pedersen, A smooth particle mesh Ewald method, *J. Chem. Phys.*, 1995, **103**, 8577–8593.
79 J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes, *J. Comput. Phys.*, 1977, **23**, 327–341.
80 J. P. Perdew, M. Ernzerhof and K. Burke, Rationale for mixing exact exchange with density functional approximations, *J. Chem. Phys.*, 1996, **105**, 9982–9985.
81 C. Adamo and G. E. Scuseria, Accurate excitation energies from time-dependent density functional theory: Assessing the PBE0 model, *J. Chem. Phys.*, 1999, **111**, 2889–2899.
82 T. J. Giese and D. M. York, Ambient-Potential Composite Ewald Method for ab Initio Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulation, *J. Chem. Theory Comput.*, 2016, **12**, 2611–2632.
83 V. I. Lebedev, Values of the nodes and weights of quadrature formulas of Gauss-Markov type for a sphere from the ninth to seventeenth order of accuracy that are invariant with respect to an octahedron group with inversion, *Zh. Vychisl. Mat. Mat. Fiz.*, 1975, **15**, 48–54.
84 V. I. Lebedev, Spherical quadrature formulas exact to orders 25-29, *Sib. Math. J.*, 1977, **18**, 99–107.
85 V. I. Lebedev and A. Skorokhodov, Quadrature formulas of orders 41, 47, and 53 for the sphere, *Russ. Acad. Sci. Dokl. Math.*, 1992, **45**, 587–592.
86 V. I. Lebedev, A quadrature formula for the sphere of 59th algebraic order of accuracy, *Russ. Acad. Sci. Dokl. Math.*, 1995, **50**, 283–286.
87 V. I. Lebedev and D. N. Laikov, A quadrature formula for the sphere of the 131st algebraic order of accuracy, *Russ. Acad. Sci. Dokl. Math.*, 1999, **59**, 477–481.
88 A. D. Becke, A multicenter numerical integration scheme for polyatomic molecules, *J. Chem. Phys.*, 1988, **88**, 2547–2553.
89 T.-S. Lee, B. K. Radak, A. Pabis and D. M. York, A new maximum likelihood approach for free energy profile construction from molecular simulations, *J. Chem. Theory Comput.*, 2013, **9**, 153–164.
90 T.-S. Lee, B. K. Radak, M. Huang, K.-Y. Wong and D. M. York, Roadmaps through free energy landscapes calculated using the multidimensional vFEP approach, *J. Chem. Theory Comput.*, 2014, **10**, 24–34.

