

Role of tautomerism in RNA biochemistry

VIPENDER SINGH,^{1,2,3} BOGDAN I. FEDELES,^{1,2,3} and JOHN M. ESSIGMANN^{1,2,3}

¹Department of Chemistry, ²Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

³Center for Environmental Health Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

ABSTRACT

Heterocyclic nucleic acid bases and their analogs can adopt multiple tautomeric forms due to the presence of multiple solvent-exchangeable protons. In DNA, spontaneous formation of minor tautomers has been speculated to contribute to mutagenic mispairings during DNA replication, whereas in RNA, minor tautomeric forms have been proposed to enhance the structural and functional diversity of RNA enzymes and aptamers. This review summarizes the role of tautomerism in RNA biochemistry, specifically focusing on the role of tautomerism in catalysis of small self-cleaving ribozymes and recognition of ligand analogs by riboswitches. Considering that the presence of multiple tautomers of nucleic acid bases is a rare occurrence, and that tautomers typically interconvert on a fast time scale, methods for studying rapid tautomerism in the context of nucleic acids under biologically relevant aqueous conditions are also discussed.

Keywords: RNA; tautomerism; ribozymes; riboswitches

INTRODUCTION

Nucleic acid bases can exist in multiple tautomeric forms due to the presence of solvent-exchangeable protons. Tautomers are structural isomers that differ from one another based on the position of proton(s) and double bonds (Antonov 2014). The presence of multiple tautomers is expected to increase the structural and chemical diversity of nucleic acid bases, as shown in Figure 1. Structural diversification by tautomerism of canonical bases, their metabolic products and related analogs, can be utilized by many RNA enzymes and aptamers to execute some of their biological functions (Cochrane et al. 2007; Klein et al. 2007; Cochrane and Strobel 2008a; Thore et al. 2008; Gilbert et al. 2009; Wilcox et al. 2011; Singh et al. 2014). This review focuses on the importance of tautomerism in catalysis of small self-cleaving ribozymes and in the recognition of certain ligand analogs by the purine and TPP riboswitches. Additionally, this review also discusses methods that can be utilized for studying tautomerism under biologically relevant aqueous conditions, an area that has traditionally posed many experimental challenges, because typically, minor tautomeric forms of nucleic acid bases are rare and interconvert rapidly in aqueous solution.

The term tautomerism was first used by Laar in 1886 to describe a dynamic equilibrium between two compounds containing weakly bonded hydrogen atoms (Laar 1886). This type of tautomerism that occurs from repositioning or

movement of protons is called prototropic tautomerism. Nucleic acid bases have carbonyl and amino functional groups, which contain solvent-exchangeable (weakly bonded) hydrogen atoms that can participate in keto–enol and amino–imino types of tautomerism (Fig. 1; Watson and Crick 1953; Topal and Fresco 1976). Specifically, adenine has the ability to adopt amino and imino tautomeric forms involving the exocyclic group at the 6-position; uracil and thymine have carbonyl functional groups that can participate in keto–enol tautomerism. Guanine and cytosine have both amino and carbonyl groups, thus they can exhibit both amino–imino and keto–enol types of tautomerism (Fig. 1). Of all possible tautomeric forms of the canonical nucleobases, the keto- and amino-forms predominate under physiological conditions, and are thus considered “major” tautomers. The imino- and enol forms are considered “minor” tautomers and are typically very rare. This review covers specifically the instances where the minor tautomers of nucleobases and base analogs have been shown to play critical roles in biochemical processes pertaining to, or involving RNA molecules.

Although all nucleic acid bases can potentially adopt minor tautomeric forms, only the tautomeric equilibria of guanine have been proposed to play essential roles in RNA biochemistry. Specifically, minor tautomeric and/or ionized forms of catalytic guanosines have been implicated in the acid–base catalysis of the autolytic cleavage reactions of self-cleaving ribozymes, such as hairpin, hammerhead, and

Corresponding author: jessig@mit.edu

Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.048371.114>. Freely available online through the RNA Open Access option.

© 2014 Singh et al. This article, published in *RNA*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

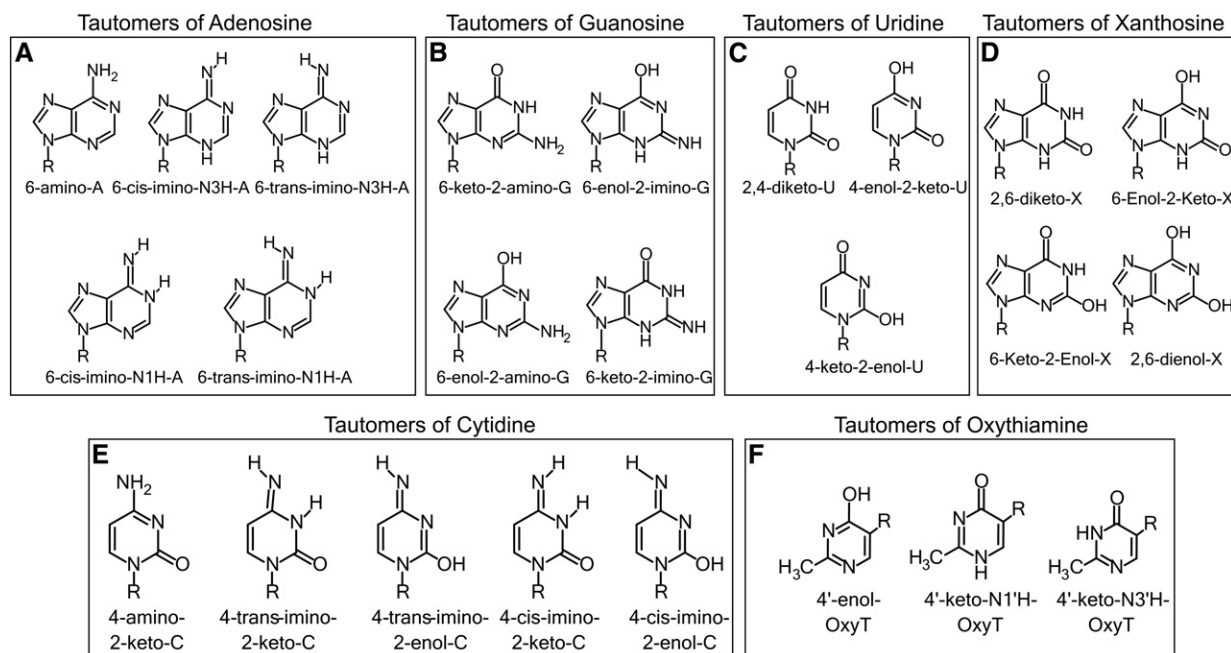


FIGURE 1. Tautomers of (A) adenosine, (B) guanosine, (C) uridine, (D) xanthosine, (E) cytidine, and (F) oxythiamine.

glmS (Cochrane and Strobel 2008a). Tautomerism has also been invoked as a molecular rationale for the interaction of certain base analogs with RNA aptamers. For example, the purine riboswitch (Gilbert et al. 2009), and the thiamine pyrophosphate (TPP) riboswitch (Thore et al. 2008; Gilbert et al. 2009; Singh et al. 2014) are proposed to bind to the minor tautomeric forms of their nonnative ligands, xanthine, and oxythiamine pyrophosphate (OxyTPP), respectively.

Studying tautomerism in aqueous conditions is challenging (Singh et al. 2014). While tautomerism of nucleic acid bases has been studied in aprotic solvents, gas phase or excited state conditions, the conclusions derived are less relevant for RNA biochemistry. Under these conditions, tautomeric equilibria are significantly altered, and the relative proportion of major and minor tautomeric forms is changed, even for the canonical nucleobases (Nir et al. 2002). Challenges of studying tautomerism in aqueous conditions include the fast rates of the tautomeric equilibria, the low abundance of minor tautomers, and the high chemical and structural similarity, between the corresponding minor and major tautomeric species. Traditional spectroscopic methods, although sensitive to certain chemical and structural properties, have been of limited use for studying tautomerism under aqueous conditions (Peng and Tokmakoff 2012; Singh et al. 2014). However, recent important developments in spectroscopic methods are beginning to overcome historical limitations and are starting to allow comprehensive characterization of tautomeric equilibria of base analogs free in solution or in the context of nucleic acids (Singh et al. 2014). Such methods (discussed later in this review) are based on variable temperature NMR and 2D IR spectroscopies, binding isotope effect

measurements combined with theoretical calculations (Singh et al. 2014), and high-resolution crystallography (Bebenek et al. 2011; Wang et al. 2011; Demeshkina et al. 2012).

This review article examines the following topics: (i) the role of tautomerism of catalytic guanines in the catalysis of hammerhead, hairpin, and *glmS* small self-cleaving ribozymes; (ii) the importance of ligand tautomerism in the recognition of xanthine by the purine riboswitch and OxyTPP by the TPP riboswitch; and (iii) current methods available for studying tautomerism of nucleobases and analogs free in solution and in RNA context under biologically relevant aqueous conditions. We also discuss the contribution of nucleobase tautomerism in creating structural and functional diversity in RNA biochemistry.

BIOCHEMICAL FUNCTIONS AND MECHANISMS OF CHEMICAL DIVERSIFICATION IN RNA

Despite the limited complexity of the RNA building blocks (the four canonical ribonucleotides), RNA enzymes and aptamers are capable of efficiently performing a wide range of biochemical functions (for review, see Cochrane and Strobel 2008a; Hiller and Strobel 2011). Small self-cleaving ribozymes promote acid-base-catalyzed autolytic reactions at rates comparable to protein enzymes (Nakano et al. 2000; Bevilacqua et al. 2004; Cochrane and Strobel 2008a). These ribozymes were originally found in the genomes of many RNA viruses, where they are thought to participate in resolving products of rolling circle replication, but studies have shown that they are more widely distributed (Ferred'Amare and Scott 2010). In addition to naturally occurring

ribozymes, artificial ribozymes have also been obtained through selection; such artificial ribozymes can catalyze a wide range of chemistries including carbon–carbon bond formation (Diels–Alder reaction) (Seelig and Jäschke 1999), isomerization (Prudent et al. 1994), redox chemistry (Tsukiji et al. 2003), and even small molecule biosynthesis (Huang et al. 2000). In addition, RNA molecules also play diverse regulatory roles in many biological processes. For example, RNA aptamers, riboswitches and T-box RNAs recognize their target ligands with high affinities (micromolar to picomolar) and specificities to regulate expression of genes involved in the biosynthesis and transport of their respective ligands. Additionally, the active site of the ribosome, which catalyzes the peptidyl transferase reaction, is entirely composed of RNA (Nissen et al. 2000).

To increase their chemical versatility, RNA enzymes and aptamers use multiple mechanisms. One such mechanism involves nucleobase tautomerism and/or nucleobase ionization. Many small-cleaving ribozymes and riboswitches are believed to utilize tautomerism and/or ionization to execute their functions. Another general mechanism involves covalent modifications of the RNA nucleotides with alkyl group and other functional groups, which greatly expand the chemical repertoire of the oligoribonucleotides. For example, tRNA molecules have been shown to feature over a hundred of different modified nucleosides (Su et al. 2014).

RNA aptamers or riboswitches also harness the chemical proficiency of cofactors such as flavin mononucleotide (Winkler et al. 2002a), S-adenosylmethionine (Winkler et al. 2003), lysine (Sudarsan et al. 2003), glycine, thiamine pyrophosphate (TPP), and glucosamine-6-phosphate to perform their biochemical functions. In addition, divalent cations such as Mg^{2+} and backbone phosphate oxygens also play key functional roles in catalysis of many ribozymes (Cochrane and Strobel 2008a; Donghi and Schnabl 2011). Biochemical roles of cofactors and metal ions in RNA biochemistry have been studied elsewhere (for review, see Cochrane and Strobel 2008b; Donghi and Schnabl 2011). Furthermore, as indicated above, chemical modifications of nucleobases (such as the ones observed in tRNA and rRNA) are also known to contribute to an increased chemical versatility of RNA, as detailed in Björk et al. (1999) and Czerwoniec et al. (2009). In contrast, tautomerism as a mechanism of chemical diversification in RNA has not been well studied.

MECHANISM OF TAUTOMERIZATION OF RNA BASES

In aqueous solution, the interconversion among different prototropic tautomers is acid–base catalyzed. Using model systems for nucleic acid bases, it has been shown that the mechanism involves water-mediated exchange of proton(s) between the donor and the acceptor atoms, with equilibrium being established on a nanosecond time scale (Peng and Tokmakoff 2012; Peng et al. 2013). Acid catalyzed interconversion involves formation of a cation intermediate through

protonation, followed by deprotonation at a different location to generate another tautomeric form. Base catalyzed tautomeric interconversions involve deprotonation to form an anion, followed by protonation at another location to generate a neutral alternative tautomeric form. Ionized (cationic or anionic) states of nucleic acid bases have been well characterized in many RNA systems (Colominas et al. 1996; Bevilacqua et al. 2004; Das and Piccirilli 2005; Perrotta et al. 2006; Gong et al. 2007, 2011; Suydam and Strobel 2008; Viladoms et al. 2011; Wilcox et al. 2011; Viladoms and Fedor 2012; Wilcox and Bevilacqua 2013). Given the observed mechanism of tautomerization, such ionized states suggest the existence of tautomeric equilibria, as they can be intermediates in the acid- or base-mediated conversion of one tautomeric form to another.

Tautomeric equilibria are influenced by chemical and physical factors including the presence of metals, temperature (Peng et al. 2013; Singh et al. 2014) and pH (CS Peng, V Singh, BI Fedeles, D Li, T Amariuta, JM Essigmann, and A Tokmakoff, in prep.). The effect of metals on nucleobase tautomerism has been reviewed previously (Lippert and Gupta 2009), while the influence of temperature on tautomeric equilibria is discussed later in this review. Given the acid–base catalytic mechanism of tautomerism in aqueous solutions, pH is expected to dictate the relative amounts of different tautomers at equilibrium. Therefore, the concentration of ionized intermediates that promote tautomerism, regardless of their charge state, is maximal when the pH is close to the pK_a of the functional groups involved in tautomerism. Consequently, under physiological conditions (pH 7), compounds with pK_a 's ~ 7 are expected to display a maximum diversity of tautomers. However, the functional groups in RNA bases have unperturbed pK_a values that are either significantly lower than neutrality ($pK_a = 4.2$ for N3 of cytidine and $pK_a = 3.5$ for N1 of adenosine), or higher than the neutral pH ($pK_a = 9.2$ for N1 of guanosine and $pK_a = 9.2$ for N3 of uridine) (Bevilacqua et al. 2004). Therefore, under physiological conditions, the tautomeric equilibria are shifted toward the keto and amino forms, which are the predominant (major) tautomeric forms observed. As a result, the proportion of minor tautomers of canonical nucleic acid bases is small around physiological pH. The presence of only one major tautomeric form at physiological pH (pH ~ 7) is critical for maintaining the genomic information integrity during DNA and RNA replication and the structural integrity of RNA enzymes and aptamers.

If the pK_a 's of nucleic acid bases, however, are perturbed toward physiological pH, tautomerism is expected to play a more significant role in the biochemistry of nucleic acids. An example of a nucleoside analog with a $pK_a \sim 7.0$ is 5-aza-5,6-dihydro-2'-deoxycytidine (KP1212) (Li et al. 2014; CS Peng, V Singh, BI Fedeles, D Li, T Amariuta, JM Essigmann, and A Tokmakoff, in prep.), an anti-HIV agent currently in phase-IIa clinical trials. KP1212 has been shown to exist in as many as five different tautomeric forms (Li

et al. 2014). One of the consequences of having multiple tautomeric forms is ambiguous base-pairing, as different tautomeric forms are expected to have different and distinct base-pairing preferences. Indeed, KP1212 is a mutagenic compound, causing G to A and A to G mutations, with this property being successfully exploited as an antiviral strategy (Li et al. 2014). Considering that pK_a shifts of up to ~ 4 units toward the physiological pH have been observed in RNA enzymes (Legault and Pardi 1997; Ravindranathan et al. 2000; Gong et al. 2007; Wilcox and Bevilacqua 2013), formation of minor tautomers of nucleic acid bases is expected to occur to a significant extent and play a functional role in these systems.

Besides the prototropic tautomerism that stems from proton repositioning, other types of prototropic tautomerism exist including annular (Minkin et al. 2000; Alkorta et al. 2006), valence (Sato et al. 2007) and ring-chain (Baker et al. 1924) tautomerism. The annular tautomerism explores a quantum mechanical property of a proton in which the proton can simultaneously reside on multiple positions on a heterocycle, such as a purine or pyrimidine. Valence tautomerism involves electronic exchanges without the repositioning of protons and is more common in transition metal complexes (Sato et al. 2007). The ribose or deoxyribose in nucleosides can also participate in ring-chain tautomerism (Jones 1963), a type of tautomerism in which one of the tautomeric forms is cyclic and the interconversion is mediated by the transfer of a proton. An example of ring-chain tautomerism enables mutarotation of sugars (Baker et al. 1924). As the functional importance of valence, annular and ring-chain tautomerism in nucleic acids remains to be established, this review only focuses on examples of prototropic tautomerism that arises from the repositioning of protons in the context of RNA systems.

ROLE OF TAUTOMERISM IN RNA BIOCHEMISTRY

Self-cleaving ribozymes such as hammerhead, hairpin, and *glmS* are proposed to utilize minor tautomeric or ionized forms of catalytic guanosines to perform their catalytic function (Figs. 4, 5, below; Cochrane and Strobel 2008a). Tautomerism has also been proposed to influence recognition of xanthine and OxyTPP by the purine and TPP riboswitches, respectively (Thore et al. 2008; Gilbert et al. 2009; Singh et al. 2014). Although not discussed below, tautomeric preference of DNA and RNA polymerases is another area where formation of minor tautomeric forms may have important physiological consequences; specifically, the minor tautomeric forms of canonical nucleobases have been long speculated to induce spontaneous mutations, both during DNA replication (Watson and Crick 1953; Topal and Fresco 1976; Bebenek et al. 2011; Wang et al. 2011) and during transcription (Shugar and Kierdaszuk 1985). Considering different tautomers of nucleic acid bases and analogs have different chemical properties and base-pairing preferences (Topal and

Fresco 1976; Peng et al. 2013; Li et al. 2014; Singh et al. 2014), formation of alternative tautomeric forms is expected to influence RNA biochemistry.

Role of tautomerism in recognition of ligands by riboswitches

Structural and biochemical studies have predicted that purine and the thiamine pyrophosphate riboswitches bind to their target ligand analogs by preferentially recognizing their minor tautomeric forms (Thore et al. 2008; Gilbert et al. 2009; Singh et al. 2014). Riboswitches are structured RNA regulatory elements found in the noncoding region of many bacterial mRNAs (Roth and Breaker 2009; Breaker 2011). They bind to small molecules to regulate the expression of the downstream genes. Below we review studies where specific minor tautomeric forms of xanthine and OxyTPP ligands are proposed to be recognized preferentially by the purine and the TPP riboswitches, respectively.

Recognition of xanthine by the purine riboswitch

The purine riboswitch regulates the expression of genes involved in the metabolism of adenine and guanine (Mandal et al. 2003; Mandal and Breaker 2004). In addition to binding to normal purines, this riboswitch has been shown to also bind purine metabolites (hypoxanthine, xanthine) and synthetic analogs (such as 2,6-diaminopurine) (Gilbert et al. 2009). Structural and biochemical studies of xanthine bound to the purine riboswitch indicated that the heterocyclic base binds as the 2-enol minor tautomer (Gilbert et al. 2009). Crystal structures of the riboswitch with various ligands show that a single pyrimidine at the 74-position is critical for determining ligand specificity (Fig. 2A; Batey et al. 2004; Serganov et al. 2004; Gilbert et al. 2006a, 2009). The xanthine ligand forms a Watson–Crick base pair with C74 (Gilbert et al. 2009). Conserved residues at the U51, U47, and U22 positions in the aptamer form part of the binding pocket and contribute to ligand recognition (Fig. 2). The carbonyl oxygens (O2) of C74 and U51 form hydrogen bonds with the 2-amino functional group of the native ligand guanine (Serganov et al. 2004). Hypoxanthine does not have the 2-amino group, thus binding to the purine riboswitch with almost 200-fold less affinity compared with guanine (Batey et al. 2004). Besides the 2-amino functional group, the 6-keto, and N1 groups are also recognized by the riboswitch to form a Watson–Crick type of base-pairing (Batey et al. 2004; Serganov et al. 2004; Gilbert et al. 2006b).

The carbonyl oxygens of C74 and U51 surround the 2-position of ligands to create a highly negatively charged binding pocket. Therefore, the riboswitch strongly prefers ligands that have hydrogen bond donor groups at their 2-position, such as the amine in guanine or 2,6-diaminopurine. Xanthine, with an oxygen at the 2-position was not expected to bind to the riboswitch, yet it binds albeit weakly with a K_D of 32 μ M (Gilbert et al. 2009). The crystal structures of the

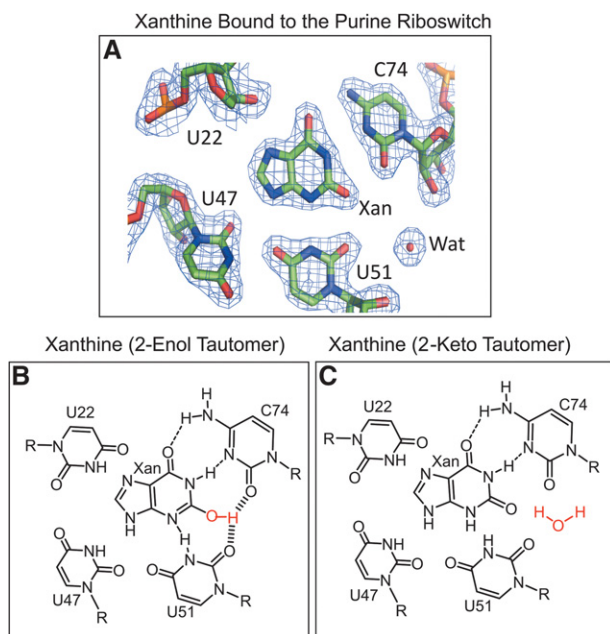


FIGURE 2. (A) The $2F_o - F_c$ omit map of xanthine bound to the purine riboswitch (coloring scheme: carbons, green; nitrogens, blue; and oxygens, red). (B) Proposed interactions for the 2-enol tautomer of xanthine bound to the purine riboswitch. (C) Proposed interactions for the 2-keto tautomer of xanthine bound to the purine riboswitch.

riboswitch bound to xanthine or guanine reveal no significant structural differences between guanine and xanthine binding (Fig. 2; Gilbert et al. 2009). To rationalize the biochemical basis of xanthine recognition, it was proposed that the 2-enol minor tautomeric form of xanthine would partially alleviate the repulsive interactions that exist due to the presence of three negatively charged oxygens in close proximity. This proposal was also consistent with the biochemical observation that xanthine binds the riboswitch at pH 6 ($K_D = 33 \mu\text{M}$), but no binding is detected at pH 8.5, because enol tautomers are expected to be more stable at lower pH (Gilbert et al. 2009). On the basis of these observations, it was speculated that the purine riboswitch preferentially binds to the 2-enol tautomer of xanthine (Gilbert et al. 2009). However, no direct evidence exists showing the presence of minor tautomeric forms of xanthine either in isolation or bound to the purine riboswitch. As it is shown in the Methods section of this review, techniques based on ^{18}O binding isotope effects combined with the density functional theory (DFT) calculation could be used to characterize the tautomeric form of xanthine bound to the riboswitch. The tautomeric equilibria of

the unbound form of xanthine could also be studied using 2D IR and variable temperature NMR methods. We applied a combination of these methods to characterize the tautomeric form(s) of oxythiamine in the context of the TPP riboswitch, as described in the next section (Singh et al. 2014).

Recognition of oxythiamine pyrophosphate by the thiamine pyrophosphate riboswitch

Our work on the tautomerism of OxyTPP in the context of the TPP riboswitch utilized a combination of spectroscopic, biochemical, and theoretical techniques to identify all three tautomeric forms of oxythiamine (OxyT) in the unbound form and the tautomeric form of oxythiamine pyrophosphate (OxyTPP) in the bound form (Singh et al. 2014). The TPP riboswitch binds to TPP, providing negative regulation of the expression of genes involved in the biosynthesis and transport of this essential vitamin (Winkler et al. 2002b, 2004). Crystal structures of the TPP riboswitch from thale cress (*Arabidopsis thaliana*) in complex with various ligands have been solved showing that the G28 within the riboswitch is critical for determining ligand specificity (Thore et al. 2006, 2008). The X-ray structure of the riboswitch with the TPP ligand shows that the amino group at the 4'-position of TPP acts as a hydrogen bond donor to the N3 position of G28 (Fig. 3A; Thore et al. 2006). OxyTPP, a presumably natural metabolite of TPP, features a keto group instead of the 4'-amino functionality and could play a role in maintaining TPP homeostasis via its interaction with the TPP riboswitch (Singh et al. 2014). Structural studies of OxyTPP with the TPP riboswitch indicate that, under crystalline conditions, its hydrogen bonding interactions to

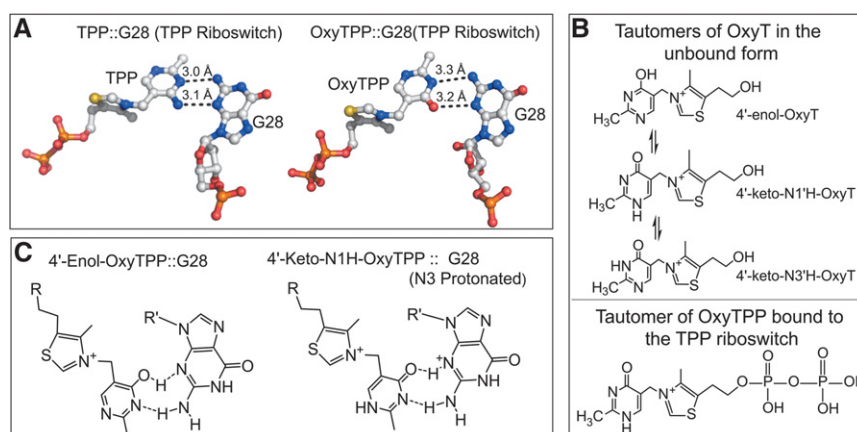


FIGURE 3. (A) Interactions of thiamine and oxythiamine pyrophosphate (TPP and OxyTPP, respectively) with the G28 of the TPP riboswitch (generated using the pdb 3D2X; coloring scheme: nitrogens are shown in blue, carbons in white, oxygens in red, and sulfur in yellow). (B) (Top) Tautomeric forms of oxythiamine (OxyT), as identified using 2D IR and variable temperature NMR spectroscopy (shown in Fig. 6B); (bottom) tautomeric form of OxyTPP bound to the TPP riboswitch as determined using binding isotope effects and DFT calculations (shown in Fig. 6B). (C) Proposed interactions for the keto and enol forms of OxyTPP with the G28 of the TPP riboswitch.

G28 are almost identical to those of TPP (Fig. 3A; Thore et al. 2008). Since the 4'-position of OxyTPP can only act as a hydrogen bond donor in its enol form, it was speculated that OxyTPP binds to the riboswitch as an enol tautomer (Thore et al. 2008).

Spectroscopic approaches based on variable temperature NMR and 2D IR spectroscopies were used to show that OxyT, under physiologically relevant aqueous conditions, exists in three possible tautomeric forms: 4'-keto-N1'H-OxyT, 4'-keto-N3'H-OxyT, and 4'-enol-OxyT (Fig. 3B; Singh et al. 2014). Application of these methods also allowed quantitation of the relative distribution of tautomers at equilibrium. The spectroscopic data showed that 4'-keto-N1'H-OxyT is the predominant tautomeric state in the unbound form followed by 4'-keto-N3'H-OxyT and 4'-enol-OxyT (Fig. 6B, below). Experimental binding isotope effects (BIE) measured under aqueous conditions, combined with DFT calculations showed that the TPP riboswitch actually binds to OxyTPP in its keto tautomeric form, namely 4'-keto-N1'H-OxyT (Fig. 3B). The presence of the keto tautomeric form of OxyTPP, however, is expected to form unfavorable interactions with the residues in the binding pocket of the riboswitch, a finding consistent with the observation that OxyTPP binds 33 times weaker than TPP (Singh et al. 2014). Because the TPP cellular concentration is tightly regulated, the lower binding affinity of OxyTPP allows activation of the riboswitch in response to oxidative damage to the TPP pool.

The presence of similar hydrogen bonding interactions for the keto form of OxyTPP and TPP ligands with the G28 of the TPP riboswitch suggested that the N3 of G28 is protonated in its interaction with OxyTPP (Fig. 3C; Singh et al. 2014). Protonation of N3 of G28 would either generate an ionized cationic form or a neutral minor tautomeric form produced through subsequent deprotonation of the cationic G28 elsewhere on the base. Further work is needed to determine whether these forms (cationic or minor tautomeric) of G28 exist and/or interact with the 4'-keto-OxyTPP ligand. The elucidation of the details of this interaction will significantly enhance our understanding of the role played by the structural diversification through tautomerism/ionization of the canonical nucleic acid bases.

Tautomerism in small autolytic ribozymes catalysis

Small self-cleaving (autolytic) ribozymes catalyze nucleolytic intramolecular self-scission reactions. Well-studied examples of small self-cleaving ribozymes include hammerhead, hairpin, *glmS*, Varkud Satellite (VS), Hepatitis delta virus (HDV), HDV-like, and twister ri-

bozymes. With the exception of *glmS*, small autolytic ribozymes do not require participation of an external coenzyme to catalyze the self-cleavage reaction (Cochrane and Strobel 2008a). The self-cleavage reaction is driven by the base-catalyzed activation of a key 2'-hydroxyl nucleophile. The activation is followed by an internal transesterification reaction, in which the nucleophilic 2'-oxygen attacks the adjacent scissile 3'-phosphate to form two pieces of RNA, one containing the 2',3'-cyclic phosphate and the other the 5'-OH functional group (Cochrane and Strobel 2008a).

Tautomerism and/or ionization of catalytic guanines have been invoked in the initial step of the activation of the 2'-hydroxyl nucleophile. A catalytic guanine with unprotonated N1 has been proposed to participate in the reaction as a general base to activate the 2'-hydroxyl for nucleophilic attack. Consistent with this mechanism, structural and biochemical studies of hammerhead, hairpin, and *glmS* ribozymes show that the N1 of catalytic G is in close proximity, within hydrogen bonding distance, to the 2'-hydroxyl nucleophile adjacent to the scissile phosphate (Figs. 4A, 5; Rupert and Ferré-D'Amaré 2001; Martick and Scott 2006). A similar mechanism, based on biochemical observations, has also been proposed for the VS ribozyme; however, in the absence of crystal structure information, the type of nucleic acid base (adenine or guanine) in close structural proximity to the catalytic 2'-hydroxyl in the VS ribozymes is uncertain.

Although the N1 of G has been implicated in the activation of the catalytic 2'-hydroxyl, the discrepancy in pK_a 's between the N1 of G and the 2'-OH nucleophile has been difficult to explain. The N1 of guanosine has a pK_a of ~ 10 , and therefore at physiological pH, it is expected to be fully protonated and thus, not chemically suitable to deprotonate the 2'-OH, which has a pK_a of ~ 13 (Veliky et al. 2001). However, a tautomeric and/or ionized form of catalytic G in which the

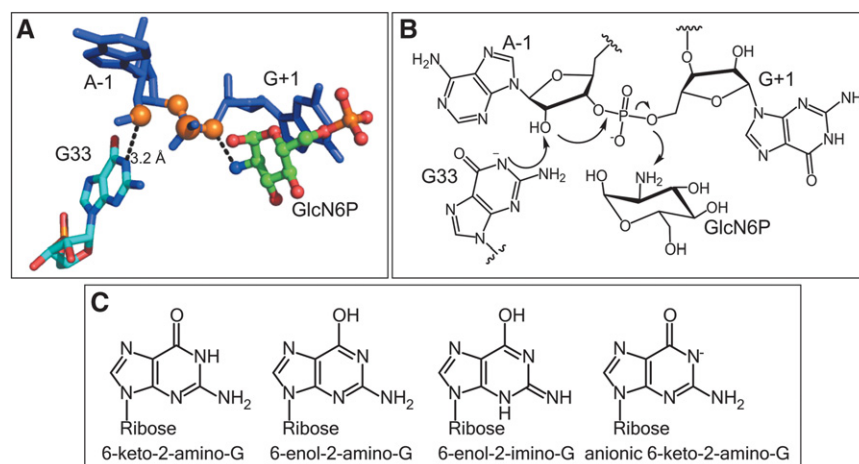


FIGURE 4. (A) The active site of the *glmS* ribozyme in the presence of glucosamine-6-phosphate (GlcN6P), showing the N1 of catalytic guanosine (G33) in close proximity to the 2'-OH nucleophile (generated using pdb 2NZ4). (B) Schematic of the autolytic reaction catalyzed by the *glmS* ribozyme. (C) Guanine and its tautomeric or ionized forms in which the N1 position is not protonated.

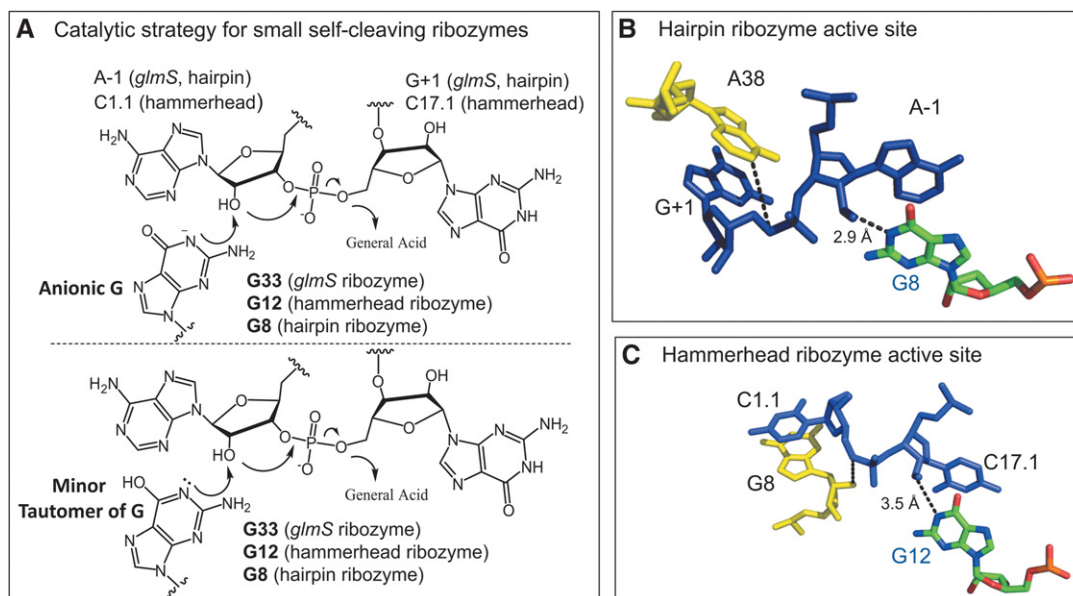


FIGURE 5. (A) General catalytic mechanism used by self-cleaving ribozymes—*glmS*, hairpin, and hammerhead with either the anionic form (top) or the minor tautomeric form (bottom) of catalytic guanosines participating in the reaction as a general base. (B) Structure of the active site of the pre-cleaved hairpin ribozyme, emphasizing the role of G8 in the activation of the 2'-OH nucleophile (generated from pdb 1M5K). (C) Active site of the hammerhead ribozyme (generated from pdb 3ZD5). A similar mechanism is also used by the *glmS* ribozyme. The active site structure is shown in Figure 4A. Coloring scheme: The nucleotides flanking the scissile phosphate are in blue; the nucleotides that interact with the reactive atoms are shown in yellow. Hydrogen bonds between the catalytic guanines (shown in green [carbon], blue [nitrogen], and red [oxygen]) and the 2'-OH nucleophile are shown as dashed lines.

N1 is unprotonated is anticipated to be more nucleophilic, and thus chemically capable of activating the 2'-hydroxyl nucleophile. Moreover, it is possible that the pK_a of N1 would be perturbed in a minor tautomeric form, because of different electronic distribution around N1.

As explained earlier, tautomerization and/or ionization of G to generate a form with unprotonated N1 would occur to a greater extent at pK_a 's around the physiological pH. Significant alterations in pK_a 's are not uncommon and changes of up to 4 units have been observed in RNA enzymes (Legault and Pardi 1997; Wilcox and Bevilacqua 2013). Furthermore, in the context of RNA, minor tautomeric forms of nucleosides are expected to be more stable than the ionized (anionic) forms. Formation of minor tautomer through protonation at an alternative site would stabilize the active (unprotonated N1) anionic form of the catalytic guanosine, by removing the negative charge of an anionic intermediate, which in a polyanionic system like RNA would be unfavorable. The next section discusses specific experimental evidence suggestive of the involvement of the N1 of catalytic G, through ionization or tautomerism, in the chemistry of small self-cleaving ribozymes.

Potential for tautomerism in the *glmS* ribozyme-catalyzed nucleolytic reaction

The *glmS* riboswitch/ribozyme binds to glucosamine-6-phosphate (GlcN6P) to regulate expression of the gene that

encodes for glutamine-fructose-6-phosphate amidotransferase, the biosynthetic enzyme for GlcN6P (Milewski 2002; Winkler et al. 2004; Collins et al. 2007; McCown et al. 2012). The *glmS* riboswitch is also a ribozyme that utilizes the self-cleavage reaction as a mode of genetic regulation. The reaction is ligand-dependent, the ligand participating in the reaction as a general acid (McCarthy et al. 2005; Cochrane et al. 2007). Its reaction mechanism is similar to RNase A and involves an attack of the vicinal 2'-OH on the scissile phosphate to generate two fragments of RNA, one containing the 2',3' cyclic phosphate and the other having a free 5'-OH group (Klein and Ferré-D'Amaré 2006; Cochrane et al. 2007; Cochrane and Strobel 2008a). Structural and biochemical analysis supports a mechanism in which the transphosphorylation reaction is initiated by the base-catalyzed activation of the vicinal 2'-hydroxyl nucleophile by the N1 of a catalytic guanine at the 33-position (G33) (Klein and Ferré-D'Amaré 2006; Cochrane et al. 2007; Klein et al. 2007). The amine of GlcN6P, through fine tuning of its pK_a (Davis et al. 2011; Gong et al. 2011; Wilcox et al. 2011), participates in the transesterification reaction as a general acid by donating a proton to the 5'-oxygen of the leaving group RNA fragment (Klein and Ferré-D'Amaré 2006; Cochrane et al. 2007; Viladoms and Fedor 2012). Brønsted coefficient of 0.7 measured for the 5'-oxygen supports a late transition state in which its protonation is rate-limiting (Viladoms and Fedor 2012).

The G33 residue of *glmS* is a critical residue for the *glmS* catalyzed self-cleavage reaction (Cochrane et al. 2007; Klein

et al. 2007; Cochrane and Strobel 2008a), because the N1 position of G33 is structurally in close proximity to the 2'-OH nucleophile (Fig. 4A). Replacement of the G33 with any other canonical nucleobase significantly reduces the rate of the self-cleavage reaction by a factor of 10^3 – 10^5 (Cochrane et al. 2007; Klein et al. 2007). Given the functional importance of G33 in the cleavage reaction and the structural proximity to the 2'-hydroxyl group, it was proposed that the N1 of G33 directly activates the 2'-OH nucleophile (Cochrane et al. 2007). However, as mentioned earlier the N1 of guanine has a $pK_a \sim 10$, making it a poor base to abstract a proton from the 2'-OH, with a pK_a of 13. Deprotonation of the N1 of G33 through formation of either the minor tautomeric form (6-enol) or the anionic form, as shown in Figures 4C and 5A, could generate a form of G33 more suitable to participate in the *glmS* catalyzed reaction as a general base (Cochrane et al. 2007; Cochrane and Strobel 2008a). Deprotonation of N1 of G either through the formation of an anionic state or a minor tautomeric form would however require significant alteration in the pK_a of the G33. Given that ionized states of catalytic guanosines have been identified (Wilcox et al. 2011), we suspect that minor tautomers are also formed and play a role in the mechanism of self-cleaving ribozymes; however, identifying such minor tautomers in the context of RNA has been experimentally challenging. Further work is needed to characterize the functional state of the catalytic G33 guanosine in the *glmS* ribozyme. Methods based on binding isotope effects, which are described later in this review, could be useful to characterize the functional state of the G33 catalytic guanosine.

Possible role of tautomerism in catalysis of hammerhead ribozyme reaction

The hammerhead ribozyme is a subcategory of satellite RNAs originally found in the genomes of plant viruses, but now known to be present in many eukaryotic genomes (McKay 1996; Hammann et al. 2012). It participates in catalyzing the cleavage and ligation of RNA as part of the rolling circle replication to resolve the replicated products (McKay 1996), and may contribute to RNA processing such as RNA-catalyzed ligation and *trans*-cleavage (Hammann et al. 2012).

Structural and biochemical studies of the hammerhead ribozyme are consistent with the observation that the N1 of the catalytic guanosine G12 is in close proximity to the 2'-hydroxyl nucleophile and plays an important role in its activation (Fig. 5C; McKay 1996; Martick and Scott 2006; Thomas and Perrin 2008). The N1 of the catalytic guanosine is within hydrogen bonding distance (3.5 Å) from the 2'-OH nucleophile (Martick and Scott 2006). G12 exhibits a log-linear trend of activity with pH that does not plateau even at the highest pH, indicating general base catalysis (Han and Burke 2005). Replacement of G12 with any other canonical nucleobase results in 10^2 - to 10^5 -fold decrease in the reaction rate (Ruffner et al. 1990) with maximal reduction of 10^5 -fold

observed for the adenine substitution (Chi et al. 2008). Structural studies have shown that substitution of G12 to A does not perturb its structure, suggesting that the decrease in the rate of cleavage is primarily from the pK_a shift, which changes from ~ 9.5 to ~ 3.5 upon G12A substitution (Chi et al. 2008). In addition, even though the A12 (with unprotonated N1) is a much weaker base than G12 (protonated N1) because of its lower pK_a , G12A substitution results in a better alignment of N1 of the purine with the 2'-OH nucleophile (3.5 Å for G versus 2.5 Å for A), indicating the preference for the deprotonated N1 as a general base catalyst (Chi et al. 2008). Furthermore, substituting G12 with analogs that specifically disrupt the pK_a of N1 of G, such as inosine, diaminopurine, or 2-aminopurine, affects the overall rate of the reaction with maximum reduction of 10^3 observed with 2-aminopurine (Han and Burke 2005), which suggests a direct role of the N1 of G in the chemistry of hammerhead ribozyme. In addition, replacement of the nucleophilic 2'-hydroxyl group with the electrophilic 2'-bromoacetamide group results in alkylation of the N1 position of G12 in a pH and Mg^{2+} -dependent manner that is consistent with the direct involvement of the N1 of G12 in the cleavage reaction (Thomas and Perrin 2008).

Taken together, the biochemical and structural data are consistent with a mechanism in which the N1 of the G12 residue directly participates as a general base in the hammerhead ribozyme-catalyzed self-cleavage reaction. Similarly to the mechanism discussed earlier for *glmS* ribozyme, the unprotonated form of N1 is more suitable to participate in the reaction as a general base, which can be an anionic G form, generated through deprotonation at N1, or a minor tautomer of G, generated by the subsequent reprotonation of anionic G at another site. It remains to be determined the extent to which the anionic form or the minor tautomeric form of the catalytic guanosine participate in the catalytic step of the cleavage reaction.

Possible tautomerism in hairpin ribozyme catalysis

The hairpin ribozymes like hammerhead constitute a subset of RNA satellites in the genomes of RNA viruses and are thought to catalyze processes that precede genomic RNA replication (Fedor 2000). They catalyze reversible cleavage of sequence-specific regions of RNA to form a 2',3'-cyclic phosphate and a 5'-OH terminus.

Similar to hammerhead and *glmS* ribozymes, the first step in the hairpin ribozyme-catalyzed cleavage reaction involves activation of a 2'-OH nucleophile by a catalytic guanosine (G8 for hairpin) (Fedor 2000; Bevilacqua and Yajima 2006). The N1 of G8 is within hydrogen bonding distance from the 2'-hydroxyl nucleophile (Fig. 5B; Rupert and Ferré-D'Amaré 2001). The G8 residue is essential for catalysis, replacement of G8 with an abasic site resulting in a 10^3 -fold reduction in the rate of cleavage reaction (Kuzmin et al. 2004). Perturbation of pK_a of the N1 of G8 through its substitution

with either inosine or diaminopurine changes the pH profile of the self-cleavage reaction (Pinard et al. 2001), suggesting a direct role of the N1 of G8 in the cleavage reaction (Pinard et al. 2001). However, the decrease in rate observed with the abasic substitution did not influence the pH dependence of the reaction, indicating that the apparent pK_a could be due to another source (Kuzmin et al. 2004). While the catalytic role of the G8 residue, as a general base, in catalysis of hairpin ribozyme has been established, further studies are needed to examine whether G8 participates in the reaction as a minor tautomer or in its anionic form.

Varkud Satellite (VS) ribozyme catalysis

The Varkud Satellite (VS) ribozyme is the largest auto-nucleolytic ribozyme found in the mitochondrial satellite plasmid of *Neurospora* species, playing a role in the processing and resolution of replication intermediates (Griffiths 1995; Lilley 2004). It catalyzes reversible site-specific cleavage and ligation reactions (Saville and Collins 1990). Since the crystal structure of VS ribozyme is not yet available, the mechanism of VS ribozyme catalysis remains uncertain. Biochemical data suggest the involvement of a site-specific adenosine (A756) and guanosine (G638) in acid–base catalyses. Substitution of A756 with any other base results in a 300-fold or more reduction of the catalytic activity of the VS ribozyme, whereas G638 to A substitution results in a 10^4 -fold reduction in rate (Lafontaine et al. 2001; Sood and Collins 2002). Complete removal of A756 causes 900-fold reduction in activity, without affecting ribozyme folding suggesting that it directly participates in the ribozyme catalysis (Lafontaine et al. 2001, 2002). Biochemical cross-linking experiments using 4-thiouridine also demonstrated a direct interaction between A756 and the cleavage site (Hiley et al. 2002). Nucleotide analog interference mapping (NAIM) studies, using a series of ionization sensitive adenosine and cytosine analogs, indicated that ionization of A756 plays a direct role in the VS ribozyme-catalyzed ligation reaction (Jones and Strobel 2003). Although NAIM studies are consistent with the ionization of A756 during catalysis, systematic studies to structurally identify the ionized or tautomeric form of A756 have been lacking. Methods based on BIEs could be adapted to study amino–imino tautomerism of A756 in the VS ribozyme. The VS ribozyme is particularly suitable for studying amino–imino tautomerism in RNA, as it uses adenosine in its catalysis. For this purpose, double labeled adenosines, [6-N15, 1'-C14] and [6-N14, 1'-H3], can be used to characterize the tautomeric form A756 in the active site of the VS ribozyme by using a strategy that is briefly described later in this review.

Methods for studying tautomerism in nucleic acids

Because tautomers are structurally similar and interconvert on less than a nanosecond time scale in aqueous conditions (Bensaude et al. 1977; Colominas et al. 1996; Peng et al.

2013), it is challenging to directly observe and quantitate tautomeric distributions using conventional spectroscopic methods. It is especially difficult to study tautomerism under aqueous equilibrium conditions, because tautomerization is mediated by water molecules (Sayle 2010; Singh et al. 2014).

NMR spectroscopy, although sensitive to tautomerization, is challenging to use for identifying tautomers under physiologically relevant aqueous conditions. At room temperature in aqueous solution, the proton exchange rates between tautomers are faster than the NMR time scale (Kühne et al. 1979). Low (subzero) temperature conditions are required to decrease the rate of tautomerization to NMR time scale to allow separation of chemical shifts from individual tautomers (Fig. 6B; Singh et al. 2014). These conditions, however, require the use of nonaqueous aprotic solvents, such as dimethylformamide, ethylacetate, or acetone, that remain liquid up to -61°C , -83°C , and -95°C , respectively. An additional caveat of nonaqueous solvents with low dielectric constants is that they often stabilize minor tautomeric forms even for naturally occurring nucleic acid bases (Jaworski 1990; Mons et al. 2002; Bakker et al. 2004); in contrast, under aqueous conditions, canonical nucleobases exist predominantly in their keto- and amino major tautomeric forms (Peng et al. 2011).

Theoretical quantum mechanical calculations are relevant for predicting distribution of tautomers based on their relative thermodynamic stability (Brown et al. 1989; Colominas et al. 1996). However, for a given set of conditions (i.e., gas phase or implicit solvent), predictions of calculated energies can vary significantly, as much as 5 kcal/mol, depending on the level of theory (Moreno and Miller 1990; Mata et al. 2010) and basis sets. Moreover, depending on the system, the correlation between theoretical and experimental values may not have been fully tested. Nevertheless, theoretical calculations are useful in formulating hypotheses, which can then be experimentally tested.

Vibrational 1D FTIR and Raman spectroscopies are sensitive to identifying tautomeric forms of nucleic acid bases under aqueous conditions. Using Raman spectroscopy on single RNA crystals (Raman crystallography) affords excellent signal-to-noise ratios, and because the measurements are carried out at room temperature, in nascent crystals that contain up to 70% solvent, they capture many of the biologically relevant properties of RNA in solution (Gong et al. 2009). However, unambiguous peak assignments from different tautomers, in 1D spectra, are often challenging due to the congested nature of the 1D FTIR or Raman spectra. Nevertheless, FTIR has been used for studying tautomerism under aqueous conditions. Using variable temperature FTIR, the presence of tautomeric equilibria of OxyT was established under aqueous conditions (Fig. 6B; Singh et al. 2014). Tautomeric equilibria are also dependent on the temperature; therefore, the change in intensity of a vibrational mode is indicative of the presence of multiple tautomeric forms corresponding to that mode. For example, the change

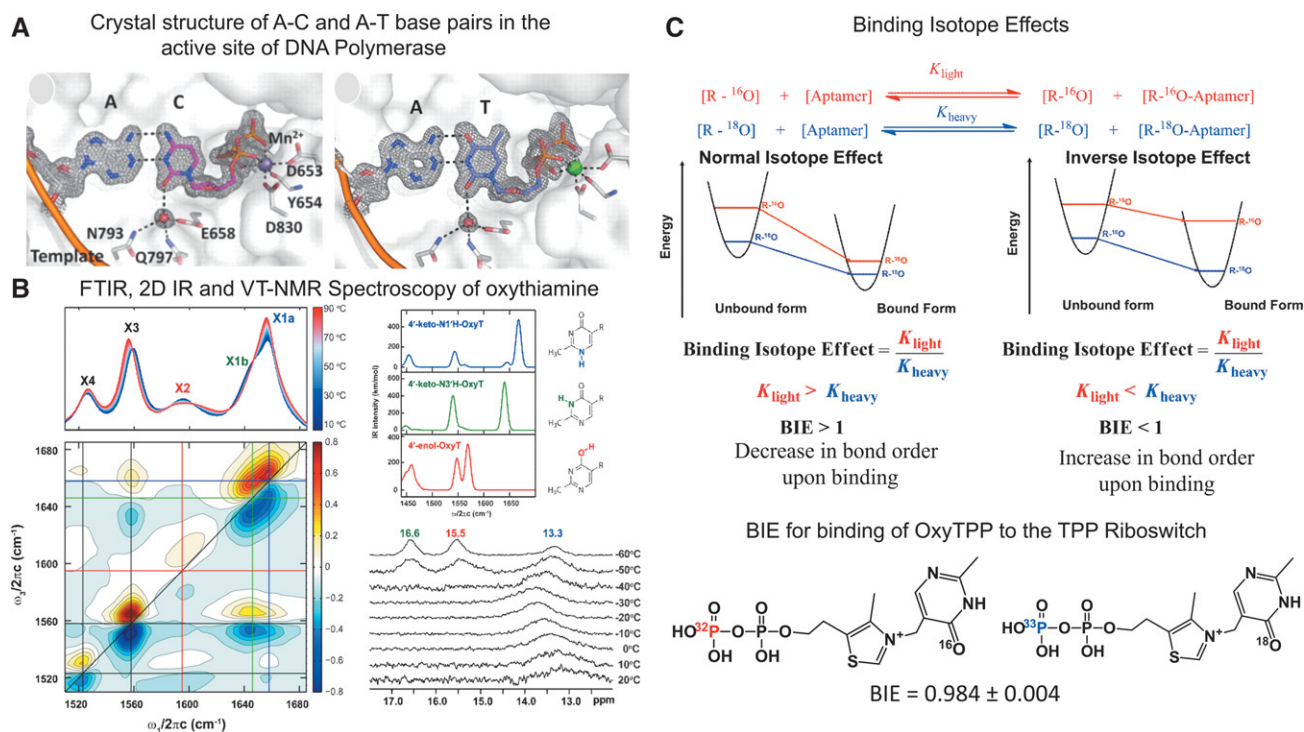


FIGURE 6. Examples of methods used for studying tautomerism of nucleic acid bases and their analogs. (A) Observation of a rare tautomeric form of a mismatched base pair in the active site of a DNA polymerase—figure showing a comparison of C-A and T-A base pairs placed at the polymerase insertion site (figure courtesy Wang et al. 2011 with permission from Lorena Beese © 2011). Composite omit maps (gray) at 1.5σ are shown around the base pairs and the anchored water molecules. Dashed lines in black indicate hydrogen bonds. (B) 1D FTIR, 2D IR, and variable temperature NMR data of oxythiamine. 1D FTIR and 2D IR were used to identify all three tautomeric forms of OxyT under aqueous conditions. Shown are color-coded horizontal/vertical lines that connect the diagonal FTIR peaks to cross-peaks of coupled vibrations that can be correlated to each of the three tautomeric states. Variable temperature NMR for OxyT (color coded) in DMF identifying each of the three tautomeric forms for oxythiamine. (C) Scheme showing physical basis for interpreting ^{18}O binding isotope effects. A BIE of <1 indicates increased bond order upon binding of a ligand to its target or tighter binding (or greater stabilization) of the ligand carrying the heavier isotope and vice versa. (Bottom) Magnitude of ^{18}O BIE measured for the binding of OxyTPP to the TPP riboswitch. The inverse value was consistent with the keto tautomer of OxyTPP binding in the pocket of the TPP riboswitch.

in the intensity of the carbonyl stretching mode with temperature is indicative of the presence of tautomeric forms (e.g., enols) that are in equilibrium with the keto form. Although variable temperature FTIR is useful for identifying the presence of tautomeric equilibria, and thus the presence of multiple tautomeric forms, relative quantification of individual tautomers using 1D methods have proven challenging because the 1D IR spectra are often congested, different tautomeric forms typically featuring many overlapping vibrational modes.

2D IR spectroscopy is an emerging, powerful method that has been successfully used to characterize tautomeric distributions of small molecules under aqueous equilibrium conditions (Hamm and Zanni 2011; Peng and Tokmakoff 2012b; Peng et al. 2013). 2D IR spectroscopy, an optical analog of 2D NMR spectroscopy, uses sequences of ultrafast IR pulses to characterize vibrational couplings, which appear as cross-peaks in 2D IR spectra (Hamm and Zanni 2011). The distinct cross-peaks originating from different tautomers in 2D IR spectra enable unambiguous peak assignments that are often not possible in the congested 1D FTIR spectra. Furthermore, the intrinsic picosecond time-resolution means that these

measurements can potentially characterize the time scale of tautomer exchange processes (Peng and Tokmakoff 2012b). For example, time-resolved 2D IR spectroscopy has been used to determine that keto–enol tautomeric equilibria in nucleobase models and nucleoside analogs establish on nanosecond time scales in aqueous solutions (Peng and Tokmakoff 2012; Singh et al. 2014; CS Peng, V Singh, BI Fedeles, D Li, T Amariuta, JM Essigmann, and A Tokmakoff, in prep).

Spectroscopic approaches such as 2D IR and low temperature NMR spectroscopy were recently applied to characterize the tautomeric forms of oxythiamine under aqueous conditions (Fig. 6B; Singh et al. 2014). These methods are directly useful for studying tautomerism of naturally occurring nucleic acid bases and their analogs, free in solution. However, complex RNA systems such as ribozymes and riboswitches are currently outside the realm of application of these methods. In large RNA systems, it is challenging to unambiguously isolate the signal from a particular base in the background of many bases with same/similar chemical properties; additionally minor tautomers are expected to be present in small amounts.

X-ray crystallography is a powerful technique for structural characterization of biomolecules and, at high resolution, it can be used effectively for studying tautomerism in RNA biosystems. Using a high-resolution crystal structure of the Dpo4 DNA polymerase bound to a double stranded oligonucleotide template containing a mismatched base pair, it was recently shown that the structure of the mismatched base pair is consistent with the presence of a minor tautomer (Fig. 6A; Wang et al. 2011). Similarly, such mismatched base pairs that could involve minor tautomeric forms have been observed in a crystal structure of DNA polymerase λ (Bebenek et al. 2011), and in the codon–anticodon interaction inside the 70S ribosome (Demeshkina et al. 2012). At typical resolutions, however, it is challenging to unambiguously assign the positions of protons to distinguish between various tautomeric forms. Also, the position of a tautomeric equilibrium may be different inside a crystal than in aqueous solution.

Binding isotope effects (BIEs) combined with DFT calculations are useful for studying tautomerism of a ligand bound to an RNA aptamer (Singh et al. 2014). BIEs are suited for studying tautomerism because, like vibrational spectroscopy, they are also influenced by alterations in vibrational frequencies between the bound and the unbound states of a ligand (Schramm 2007). This approach has been used for characterizing the tautomeric forms of OxyTPP bound to the TPP riboswitch (Fig. 6C; Singh et al. 2014). BIEs can potentially be used for studying both amino–imino and keto–enol tautomerism of any ligand bound to its target by using an appropriately labeled ligand; ^{15}N BIEs are useful for studying amino–imino tautomerism, while ^{18}O BIEs can be used for characterizing keto–enol tautomeric forms (Singh et al. 2014). Using the methods described in Singh et al. (2014) to study the tautomeric forms of OxyTPP bound to TPP riboswitch, one can envision an experimental setup to study the tautomeric forms of catalytic guanosines in ribozymes using BIEs. Specifically, an abasic site is introduced in the ribozyme in place of the catalytic guanosine and a mixture of labeled guanines containing heavy (^{15}N or ^{18}O) and light (^{14}N or ^{16}O) isotopes are added separately. By quantitating the ratio of light (^{14}N or ^{16}O) to heavy (^{15}N or ^{18}O) isotopes in the bound and the unbound forms of the guanine mixture, BIEs can be obtained. The ^{18}O or the ^{15}N BIEs measured experimentally can then be correlated with BIEs obtained computationally using the ISOEFF98 software, from all the $3N - 6$ vibrational modes obtained from DFT calculations performed using Gaussian software, for all possible tautomeric forms of guanines.

RNA systems are complex and given the challenges associated with studying tautomerism, currently no one technique can comprehensively address the question of tautomerism in every system. As it was shown recently (Singh et al. 2014), a combination of methods based on 2D IR, FTIR, variable temperature NMR, BIEs, and DFT calculations were needed to provide full characterization of the tautomeric forms of OxyT in solution and OxyTPP bound to the TPP riboswitch.

These methods would be even more powerful if combined with other techniques such as high-resolution X-ray crystallography and other kinds of spectroscopies to address the question of tautomerism in RNA enzymes and aptamers.

CONCLUSION AND PERSPECTIVE

Significant structural and biochemical evidence exists to suggest that tautomerism plays an important role in the biochemistry of many self-cleaving ribozymes and riboswitches. Given that tautomerism is an intrinsic property of ribonucleotides, one can imagine that other biological processes involving RNA could also be influenced by the structural diversity generated by tautomeric equilibria. Additionally, tautomerism is important in other areas of nucleic acid biochemistry. Spontaneous formation of minor tautomeric forms of nucleic acid bases, either in the nucleotide pool or in the template strand, has been proposed to generate mutations, owing to suspected noncanonical base-pairing properties of minor tautomers. Tautomeric nucleoside analogs also have therapeutic applications as antiviral drugs because of their ability to induce lethal mutagenesis (Li et al. 2014), increasing the viral mutation rates above the error catastrophe limit of a virus. Furthermore, because small self-cleaving ribozymes and riboswitches are considered as evolutionary remnants that have persisted in the modern genomes (Ferre-D'Amare and Scott 2010), biochemical versatility of RNA enhanced by tautomerism may have also played a crucial role in the ancestral RNA World (Gilbert 1986). Tautomerism of nucleic acid bases is perhaps also relevant to the origin of life question, and pertinent for understanding why Nature selected A, T, G, C, and U as nucleic acid bases for encoding the genetic information of all life forms.

ACKNOWLEDGMENTS

We thank Prof. P. Bevilacqua from Penn State University, Prof. R. Batey from U Colorado, Prof. U. RajBhandary from MIT, and Dr. P. McCown and Dr. N. Sudarsan from Yale University for critical reading of the manuscript and insightful comments and suggestions. This work was supported by grants from National Institutes of Health P01 CA26731, R37 CA080024, and P30 ES002109. V.S. was supported by NIH Training Grant T32 ES007020.

REFERENCES

- Alkorta I, Elguero J, Liebman JF. 2006. The annular tautomerism of imidazoles and pyrazoles: the possible existence of nonaromatic forms. *Struct Chem* 17: 439–444.
- Antonov L. 2014. *Tautomerism—methods and theories*. Wiley, New York.
- Baker JW, Ingold CK, Thorpe JF. 1924. XXXVI.—Ring-chain tautomerism. Part IX. The mutarotation of the sugars. *J Chem Soc Trans* 125: 268–291.
- Bakker JM, Compagnon I, Meijer G, von Helden G, Kabeláč M, Hobza P, de Vries MS. 2004. The mid-IR absorption spectrum of gas-phase clusters of the nucleobases guanine and cytosine. *Phys Chem Chem Phys* 6: 2810–2815.

- Batey RT, Gilbert SD, Montange RK. 2004. Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature* **432**: 411–415.
- Bebenek K, Pedersen LC, Kunkel TA. 2011. Replication infidelity via a mismatch with Watson–Crick geometry. *Proc Natl Acad Sci* **108**: 1862–1867.
- Bensaude O, Dreyfus M, Dodin G, Dubois JE. 1977. Intramolecular nondissociative proton transfer in aqueous solutions of tautomeric heterocycles: a temperature-jump kinetic study. *J Am Chem Soc* **99**: 4438–4446.
- Bevilacqua PC, Yajima R. 2006. Nucleobase catalysis in ribozyme mechanism. *Curr Opin Chem Biol* **10**: 455–464.
- Bevilacqua PC, Brown TS, Nakano S, Yajima R. 2004. Catalytic roles for proton transfer and protonation in ribozymes. *Biopolymers* **73**: 90–109.
- Björk GR, Durand JMB, Hagervall TG, Leipuviene R, Lundgren HK, Nilsson K, Chen P, Qian Q, Urbonavičius J. 1999. Transfer RNA modification: influence on translational frameshifting and metabolism. *FEBS Lett* **452**: 47–51.
- Breaker RR. 2011. Prospects for riboswitch discovery and analysis. *Mol Cell* **43**: 867–879.
- Brown RD, Godfrey PD, McNaughton D, Pierlot AP. 1989. Tautomers of cytosine by microwave spectroscopy. *J Am Chem Soc* **111**: 2308–2310.
- Chi YI, Martick M, Lares M, Kim R, Scott WG, Kim SH. 2008. Capturing hammerhead ribozyme structures in action by modulating general base catalysis. *PLoS Biol* **6**: e234.
- Cochrane JC, Strobel SA. 2008a. Catalytic strategies of self-cleaving ribozymes. *Acc Chem Res* **41**: 1027–1035.
- Cochrane JC, Strobel SA. 2008b. Riboswitch effectors as protein enzyme cofactors. *RNA* **14**: 993–1002.
- Cochrane JC, Lipchock SV, Strobel SA. 2007. Structural investigation of the GlmS ribozyme bound to its catalytic cofactor. *Chem Biol* **14**: 97–105.
- Collins JA, Irnov I, Baker S, Winkler WC. 2007. Mechanism of mRNA destabilization by the *glmS* ribozyme. *Genes Dev* **21**: 3356–3368.
- Colominas C, Luque FJ, Orozco M. 1996. Tautomerism and protonation of guanine and cytosine. Implications in the formation of hydrogen-bonded complexes. *J Am Chem Soc* **118**: 6811–6821.
- Czerwonec A, Dunin-Horkawicz S, Purta E, Kaminska KH, Kasprzak JM, Bujnicki JM, Grosjean H, Rother K. 2009. MODOMICS: a database of RNA modification pathways. 2008 update. *Nucleic Acids Res* **37**: D118–D121.
- Das SR, Piccirilli JA. 2005. General acid catalysis by the hepatitis δ virus ribozyme. *Nat Chem Biol* **1**: 45–52.
- Davis JH, Dunican BF, Strobel SA. 2011. *glmS* Riboswitch binding to the glucosamine-6-phosphate α -anomer shifts the pK_a toward neutrality. *Biochemistry* **50**: 7236–7242.
- Demeshkina N, Jenner L, Westhof E, Yusupov M, Yusupova G. 2012. A new understanding of the decoding principle on the ribosome. *Nature* **484**: 256–259.
- Donghi D, Schnabl J. 2011. Multiple roles of metal ions in large ribozymes. *Met Ions Life Sci* **9**: 197–234.
- Fedor MJ. 2000. Structure and function of the hairpin ribozyme. *J Mol Biol* **297**: 269–291.
- Ferre-D'Amare AR, Scott WG. 2010. Small self-cleaving ribozymes. *Cold Spring Harb Perspect Biol* **2**: a003574.
- Gilbert W. 1986. Origin of life: the RNA world. *Nature* **319**: 618.
- Gilbert SD, Mediator SJ, Batey RT. 2006a. Modified pyrimidines specifically bind the purine riboswitch. *J Am Chem Soc* **128**: 14214–14215.
- Gilbert SD, Stoddard CD, Wise SJ, Batey RT. 2006b. Thermodynamic and kinetic characterization of ligand binding to the purine riboswitch aptamer domain. *J Mol Biol* **359**: 754–768.
- Gilbert SD, Reyes FE, Edwards AL, Batey RT. 2009. Adaptive ligand binding by the purine riboswitch in the recognition of guanine and adenine analogs. *Structure* **17**: 857–868.
- Gong B, Chen JH, Chase E, Chadalavada DM, Yajima R, Golden BL, Bevilacqua PC, Carey PR. 2007. Direct measurement of a pK_a near neutrality for the catalytic cytosine in the genomic HDV ribozyme using Raman crystallography. *J Am Chem Soc* **129**: 13335–13342.
- Gong B, Chen JH, Yajima R, Chen Y, Chase E, Chadalavada DM, Golden BL, Carey PR, Bevilacqua PC. 2009. Raman crystallography of RNA. *Methods* **49**: 101–111.
- Gong B, Klein DJ, Ferré-D'Amaré AR, Carey PR. 2011. The *glmS* ribozyme tunes the catalytically critical pK_a of its coenzyme glucosamine-6-phosphate. *J Am Chem Soc* **133**: 14188–14191.
- Griffiths AJ. 1995. Natural plasmids of filamentous fungi. *Microbiol Rev* **59**: 673–685.
- Hamm P, Zanni M. 2011. *Concepts and methods of 2D infrared spectroscopy*, 1st ed. Cambridge University Press, Cambridge, New York, UK, New York.
- Hammann C, Luptak A, Perreault J, de la Peña M. 2012. The ubiquitous hammerhead ribozyme. *RNA* **18**: 871–885.
- Han J, Burke JM. 2005. Model for general acid–base catalysis by the hammerhead ribozyme: pH-activity relationships of G8 and G12 variants at the putative active site. *Biochemistry* **44**: 7864–7870.
- Hiley SL, Sood VD, Fan J, Collins RA. 2002. 4-Thio-U cross-linking identifies the active site of the VS ribozyme. *EMBO J* **21**: 4691–4698.
- Hiller DA, Strobel SA. 2011. The chemical versatility of RNA. *Philos Trans R Soc Lond B Biol Sci* **366**: 2929–2935.
- Huang F, Bugg CW, Yarus M. 2000. RNA-catalyzed CoA, NAD, and FAD synthesis from phosphopantetheine, NMN, and FMN. *Biochemistry* **39**: 15548–15555.
- Jaworski A. 1990. Infrared spectra and tautomerism of 5-fluorocytosine, 5-bromocytosine and 5-iodocytosine. Matrix isolation and theoretical AB initio studies. *J Mol Struct* **223**: 63–92.
- Jones PR. 1963. Ring-chain tautomerism. *Chem Rev* **63**: 461–487.
- Jones FD, Strobel SA. 2003. Ionization of a critical adenosine residue in the *Neurospora* Varkud Satellite ribozyme active site. *Biochemistry* **42**: 4265–4276.
- Klein DJ, Ferré-D'Amaré AR. 2006. Structural basis of *glmS* ribozyme activation by glucosamine-6-phosphate. *Science* **313**: 1752–1756.
- Klein DJ, Been MD, Ferré-D'Amaré AR. 2007. Essential role of an active-site guanine in *glmS* ribozyme catalysis. *J Am Chem Soc* **129**: 14858–14859.
- Kühne RO, Schaffhauser T, Wokaun A, Ernst RR. 1979. Study of transient chemical reactions by NMR. Fast stopped-flow Fourier transform experiments. *J Magn Reson* **1969** **35**: 39–67.
- Kuzmin YI, Da Costa CP, Fedor MJ. 2004. Role of an active site guanine in hairpin ribozyme catalysis probed by exogenous nucleobase rescue. *J Mol Biol* **340**: 233–251.
- Laar C. 1886. Ueber die Hypothese der wechselnden Bindung. *Ber Dtsch Chem Ges* **19**: 730–741.
- Lafontaine DA, Wilson TJ, Norman DG, Lilley DM. 2001. The A730 loop is an important component of the active site of the VS ribozyme. *J Mol Biol* **312**: 663–674.
- Lafontaine DA, Wilson TJ, Zhao ZY, Lilley DMJ. 2002. Functional group requirements in the probable active site of the VS ribozyme. *J Mol Biol* **323**: 23–34.
- Legault P, Pardi A. 1997. Unusual dynamics and pK_a shift at the active site of a lead-dependent ribozyme. *J Am Chem Soc* **119**: 6621–6628.
- Li D, Fedele BI, Singh V, Peng CS, Silvestre KJ, Simi A, Simpson J, Tokmakoff A, Essigmann JM. 2014. Tautomerism provides a molecular explanation for the mutagenic properties of the anti-HIV nucleoside 5-aza-5,6-dihydro-2'-deoxycytidine. *Proc Natl Acad Sci* **111**: E3252–E3259.
- Lilley DMJ. 2004. The Varkud satellite ribozyme. *RNA* **10**: 151–158.
- Lippert B, Gupta D. 2009. Promotion of rare nucleobase tautomers by metal binding. *Dalton Trans* 4619–4634.
- Mandal M, Breaker RR. 2004. Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat Struct Mol Biol* **11**: 29–35.
- Mandal M, Boese B, Barrick JE, Winkler WC, Breaker RR. 2003. Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. *Cell* **113**: 577–586.
- Martick M, Scott WG. 2006. Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell* **126**: 309–320.

- Mata S, Cortijo V, Caminati W, Alonso JL, Sanz ME, López JC, Blanco S. 2010. Tautomerism and microsolvation in 2-hydroxypyridine/2-pyridone. *J Phys Chem A* **114**: 11393–11398.
- McCarthy TJ, Plog MA, Floy SA, Jansen JA, Soukup JK, Soukup GA. 2005. Ligand requirements for glmS ribozyme self-cleavage. *Chem Biol* **12**: 1221–1226.
- McCown PJ, Winkler WC, Breaker RR. 2012. Mechanism and distribution of glmS ribozymes. *Methods Mol Biol* **848**: 113–129.
- McKay DB. 1996. Structure and function of the hammerhead ribozyme: an unfinished story. *RNA* **2**: 395–403.
- Milewski S. 2002. Glucosamine-6-phosphate synthase—the multi-facets enzyme. *Biochim Biophys Acta* **1597**: 173–192.
- Minkin VI, Garnovskii AD, Elguero J, Katritzky AR, Denisko OV. 2000. The tautomerism of heterocycles: five-membered rings with two or more heteroatoms. *Adv Heterocycl Chem* **76**: 157–323.
- Mons M, Dimicoli I, Piuze F, Tardivel B, Elhanine M. 2002. Tautomerism of the DNA base guanine and its methylated derivatives as studied by gas-phase infrared and ultraviolet spectroscopy. *J Phys Chem A* **106**: 5088–5094.
- Moreno M, Miller WH. 1990. On the tautomerization reaction 2-pyridone \rightleftharpoons 2-hydroxypyridine: an ab initio study. *Chem Phys Lett* **171**: 475–479.
- Nakano S, Chadalavada DM, Bevilacqua PC. 2000. General acid-base catalysis in the mechanism of a hepatitis δ virus ribozyme. *Science* **287**: 1493–1497.
- Nir E, Plützer C, Kleinermanns K, de Vries M. 2002. Properties of isolated DNA bases, base pairs and nucleosides examined by laser spectroscopy. *Eur Phys J D* **20**: 317–329.
- Nissen P, Hansen J, Ban N, Moore PB, Steitz TA. 2000. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**: 920–930.
- Peng CS, Tokmakoff A. 2012. Identification of lactam-lactim tautomers of aromatic heterocycles in aqueous solution using 2D IR spectroscopy. *J Phys Chem Lett* **3**: 3302–3306.
- Peng CS, Jones KC, Tokmakoff A. 2011. Anharmonic vibrational modes of nucleic acid bases revealed by 2D IR spectroscopy. *J Am Chem Soc* **133**: 15650–15660.
- Peng CS, Baiz CR, Tokmakoff A. 2013. Direct observation of ground-state lactam-lactim tautomerization using temperature-jump transient 2D IR spectroscopy. *Proc Natl Acad Sci* **110**: 9243–9248.
- Perrotta AT, Wadkins TS, Been MD. 2006. Chemical rescue, multiple ionizable groups, and general acid–base catalysis in the HDV genomic ribozyme. *RNA* **12**: 1282–1291.
- Pinard R, Hampel KJ, Heckman JE, Lambert D, Chan PA, Major F, Burke JM. 2001. Functional involvement of G8 in the hairpin ribozyme cleavage mechanism. *EMBO J* **20**: 6434–6442.
- Prudent JR, Uno T, Schultz PG. 1994. Expanding the scope of RNA catalysis. *Science* **264**: 1924–1927.
- Ravindranathan S, Butcher SE, Feigon J. 2000. Adenine protonation in domain B of the hairpin ribozyme. *Biochemistry* **39**: 16026–16032.
- Roth A, Breaker RR. 2009. The structural and functional diversity of metabolite-binding riboswitches. *Annu Rev Biochem* **78**: 305–334.
- Ruffner DE, Stormo GD, Uhlenbeck OC. 1990. Sequence requirements of the hammerhead RNA self-cleavage reaction. *Biochemistry* **29**: 10695–10702.
- Rupert PB, Ferré-D'Amaré AR. 2001. Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. *Nature* **410**: 780–786.
- Sato O, Cui A, Matsuda R, Tao J, Hayami S. 2007. Photo-induced valence tautomerism in Co complexes. *Acc Chem Res* **40**: 361–369.
- Saville BJ, Collins RA. 1990. A site-specific self-cleavage reaction performed by a novel RNA in *Neurospora* mitochondria. *Cell* **61**: 685–696.
- Sayle RA. 2010. So you think you understand tautomerism? *J Comput Aided Mol Des* **24**: 485–496.
- Schramm VL. 2007. Binding isotope effects: boon and bane. *Curr Opin Chem Biol* **11**: 529–536.
- Seelig B, Jäschke A. 1999. A small catalytic RNA motif with Diels-Alderase activity. *Chem Biol* **6**: 167–176.
- Serganov A, Yuan YR, Pikovskaya O, Polonskaia A, Malinina L, Phan AT, Hobartner C, Micura R, Breaker RR, Patel DJ. 2004. Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem Biol* **11**: 1729–1741.
- Shugar D, Kierdaszuk B. 1985. New light on tautomerism of purines and pyrimidines and its biological and genetic implications. *J Biosci* **8**: 657–668.
- Singh V, Peng CS, Li D, Mitra K, Silvestre KJ, Tokmakoff A, Essigmann JM. 2014. Direct observation of multiple tautomers of oxythiamine and their recognition by the thiamine pyrophosphate riboswitch. *ACS Chem Biol* **9**: 227–236.
- Sood VD, Collins RA. 2002. Identification of the catalytic subdomain of the VS ribozyme and evidence for remarkable sequence tolerance in the active site loop. *J Mol Biol* **320**: 443–454.
- Su D, Chan CTY, Gu C, Lim KS, Chionh YH, McBee ME, Russell BS, Babu IR, Begley TJ, Dedon PC. 2014. Quantitative analysis of ribonucleoside modifications in tRNA by HPLC-coupled mass spectrometry. *Nat Protoc* **9**: 828–841.
- Sudarsan N, Wickiser JK, Nakamura S, Ebert MS, Breaker RR. 2003. An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes Dev* **17**: 2688–2697.
- Suydam IT, Strobel SA. 2008. Fluorine substituted adenosines as probes of nucleobase protonation in functional RNAs. *J Am Chem Soc* **130**: 13639–13648.
- Thomas JM, Perrin DM. 2008. Probing general base catalysis in the hammerhead ribozyme. *J Am Chem Soc* **130**: 15467–15475.
- Thore S, Leibundgut M, Ban N. 2006. Structure of the eukaryotic thiamine pyrophosphate riboswitch with its regulatory ligand. *Science* **312**: 1208–1211.
- Thore S, Frick C, Ban N. 2008. Structural basis of thiamine pyrophosphate analogues binding to the eukaryotic riboswitch. *J Am Chem Soc* **130**: 8116–8117.
- Topal MD, Fresco JR. 1976. Complementary base pairing and the origin of substitution mutations. *Nature* **263**: 285–289.
- Tsukiji S, Pattnaik SB, Suga H. 2003. An alcohol dehydrogenase ribozyme. *Nat Struct Mol Biol* **10**: 713–717.
- Velikyan I, Acharya S, Trifonova A, Földesi A, Chattopadhyaya J. 2001. The pK_a's of 2'-hydroxyl group in nucleosides and nucleotides. *J Am Chem Soc* **123**: 2893–2894.
- Viladoms J, Fedor MJ. 2012. The glmS ribozyme cofactor is a general acid–base catalyst. *J Am Chem Soc* **134**: 19043–19049.
- Viladoms J, Scott LG, Fedor MJ. 2011. An active-site guanine participates in glmS ribozyme catalysis in its protonated state. *J Am Chem Soc* **133**: 18388–18396.
- Wang W, Hellinga HW, Beese LS. 2011. Structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis. *Proc Natl Acad Sci* **108**: 17644–17648.
- Watson JD, Crick FH. 1953. Genetical implications of the structure of deoxyribonucleic acid. *Nature* **171**: 964–967.
- Wilcox JL, Bevilacqua PC. 2013. A simple fluorescence method for pK_a determination in RNA and DNA reveals highly shifted pK_a's. *J Am Chem Soc* **135**: 7390–7393.
- Wilcox JL, Ahluwalia AK, Bevilacqua PC. 2011. Charged nucleobases and their potential for RNA catalysis. *Acc Chem Res* **44**: 1270–1279.
- Winkler WC, Cohen-Chalamish S, Breaker RR. 2002a. An mRNA structure that controls gene expression by binding FMN. *Proc Natl Acad Sci* **99**: 15908–15913.
- Winkler W, Nahvi A, Breaker RR. 2002b. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**: 952–956.
- Winkler WC, Nahvi A, Sudarsan N, Barrick JE, Breaker RR. 2003. An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nat Struct Biol* **10**: 701–707.
- Winkler WC, Nahvi A, Roth A, Collins JA, Breaker RR. 2004. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428**: 281–286.



Role of tautomerism in RNA biochemistry

Vipender Singh, Bogdan I. Fedeles and John M. Essigmann

RNA 2015 21: 1-13

References

This article cites 102 articles, 20 of which can be accessed free at:
<http://rnajournal.cshlp.org/content/21/1/1.full.html#ref-list-1>

Open Access

Freely available online through the *RNA* Open Access option.

Creative Commons License

This article, published in *RNA*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *RNA* go to:
<http://rnajournal.cshlp.org/subscriptions>
