



# Structural Insights Into Tautomeric Dynamics in Nucleic Acids and in Antiviral Nucleoside Analogs

Bogdan I. Fedeles<sup>1</sup>, Deyu Li<sup>2</sup> and Vipender Singh<sup>3\*</sup>

<sup>1</sup>Departments of Chemistry and Biological Engineering and Center for Environmental Health Sciences, Massachusetts Institute of Technology, Cambridge, MA, United States, <sup>2</sup>Department of Pharmaceutical Sciences, College of Pharmacy, University of Rhode Island, Kingston, RI, United States, <sup>3</sup>Department of Biochemistry and Biophysics, Novartis Institute of Biomedical Research, Cambridge, MA, United States

## OPEN ACCESS

### Edited by:

Annalisa Pastore,  
King's College London,  
United Kingdom

### Reviewed by:

Pablo D. Dans,  
Institut Pasteur de Montevideo,  
Uruguay

Pierre Barraud,  
UMR8261 Expression Génétique  
Microbienne, France

### \*Correspondence:

Vipender Singh  
vipender02118@gmail.com

### Specialty section:

This article was submitted to  
Structural Biology,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 27 November 2021

**Accepted:** 30 December 2021

**Published:** 25 January 2022

### Citation:

Fedeles BI, Li D and Singh V (2022)  
Structural Insights Into Tautomeric  
Dynamics in Nucleic Acids and in  
Antiviral Nucleoside Analogs.  
Front. Mol. Biosci. 8:823253.  
doi: 10.3389/fmolb.2021.823253

DNA (2'-deoxyribonucleic acid) and RNA (ribonucleic acid) play diverse functional roles in biology and disease. Despite being comprised primarily of only four cognate nucleobases, nucleic acids can adopt complex three-dimensional structures, and RNA in particular, can catalyze biochemical reactions to regulate a wide variety of biological processes. Such chemical versatility is due in part to the phenomenon of nucleobase tautomerism, whereby the bases can adopt multiple, yet distinct isomeric forms, known as tautomers. For nucleobases, tautomers refer to structural isomers that differ from one another by the position of protons. By altering the position of protons on nucleobases, many of which play critical roles for hydrogen bonding and base pairing interactions, tautomerism has profound effects on the biochemical processes involving nucleic acids. For example, the transient formation of minor tautomers during replication could generate spontaneous mutations. These mutations could arise from the stabilization of mismatches, in the active site of polymerases, in conformations involving minor tautomers that are indistinguishable from canonical base pairs. In this review, we discuss the evidence for tautomerism in DNA, and its consequences to the fidelity of DNA replication. Also reviewed are RNA systems, such as the riboswitches and self-cleaving ribozymes, in which tautomerism plays a functional role in ligand recognition and catalysis, respectively. We also discuss tautomeric nucleoside analogs that are efficacious as antiviral drug candidates such as molnupiravir for coronaviruses and KP1212 for HIV. The antiviral efficacy of these analogs is due, in part, to their ability to exist in multiple tautomeric forms and induce mutations in the replicating viral genomes. From a technical standpoint, minor tautomers of nucleobases are challenging to identify directly because they are rare and interconvert on a fast, millisecond to nanosecond, time scale. Nevertheless, many approaches including biochemical, structural, computational and spectroscopic methods have been developed to study tautomeric dynamics in RNA and DNA systems, and in antiviral nucleoside analogs. An overview of these methods and their applications is included here.

**Keywords:** Tautomerism, ribozymes, riboswitches, mutagenesis, COVID-19, antivirals, nucleoside analogs therapy, spontaneous mutations

## INTRODUCTION

Nucleic acid bases exhibit keto-enol and amino-imino prototropic tautomerism due to the presence of multiple solvent-exchangeable protons (**Figure 1**) (Watson and Crick, 1953; Topal and Fresco, 1976; Brown et al., 1989; Colominas et al., 1996; Mons et al., 2002). The formation of minor tautomers can increase the overall structural and chemical diversity of nucleic acids, which enables their diverse functions in biology (Topal and Fresco, 1976; Cochrane and Strobel, 2008a; Singh et al., 2015). For example, many self-cleaving ribozymes (RNA enzymes) and some riboswitches (RNA aptamers) are proposed to utilize tautomerism to perform their biological function (**Figure 2, Figure 3**) (Singh et al., 2015; Singh et al., 2014). Formation of minor tautomers in DNA, at low frequency, is proposed to contribute to the phenomenon of 'spontaneous mutagenesis', which denotes the background level of mutations that appear during the replication of undamaged DNA (Watson and Crick, 1953; Topal and Fresco, 1976; Wang et al., 2011; Rangadurai et al., 2020). These mutations are thought to arise due, in part, to the altered base pairing properties of minor tautomers (**Figure 4**) (Watson and Crick, 1953; Topal and Fresco, 1976; Wang et al., 2011).

Therapeutics based on tautomeric nucleoside analogs have proven effective as antiviral agents against a range of retro- and ribo-viruses, including influenza (Delang et al., 2018), hepatitis-C-virus (HCV) (Crotty et al., 2001), human immunodeficiency virus (HIV) (Li et al., 2014) and coronaviruses, including COVID-19 (**Figure 5**) (Shannon et al., 2020; Kabinger et al., 2021). The antiviral efficacy of these analogs stems from their ability to exist in multiple tautomeric or rotameric states, which help mutagenize the viral genomes to error catastrophe and even to viral population extinction. The formation of minor tautomers in nucleic acids is a rare event, and their dynamics of interconversion is fast, on a millisecond to nanosecond time scale (Peng et al., 2013; Rangadurai et al., 2019). Recent developments in methods have allowed direct identification and quantification of minor tautomers in nucleic acid bases and in nucleoside/nucleotide analogs (Peng et al., 2011; Li et al., 2014; Singh et al., 2014; Peng et al., 2015; Rangadurai et al., 2019).

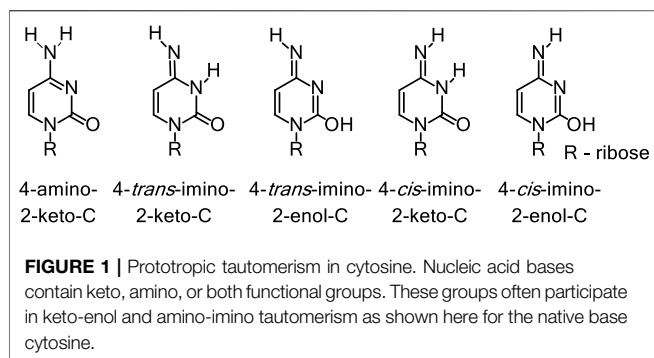
Small self-cleaving ribozymes and many riboswitches are proposed to utilize minor tautomeric forms of catalytic residues and ligands to perform their biological functions (**Figure 2, Figure 3**) (Cochrane et al., 2007; Cochrane and Strobel, 2008a; Cochrane and Strobel, 2008b; Thore et al., 2008; Gilbert et al., 2009; Wilcox and Bevilacqua, 2013; Singh et al., 2014; Singh et al., 2015). Among riboswitches, crystallographic evidence shows that the purine and the thiamine pyrophosphate riboswitches recognize the minor tautomeric forms of their non-natural ligands xanthine and oxythiamine pyrophosphate (OxyTPP), respectively (**Figure 2**) (Thore et al., 2008; Gilbert et al., 2009; Singh et al., 2015; Singh et al., 2014). Self-cleaving ribozymes such as hammerhead, hairpin, glmS, varkud satellite (VS), hepatitis delta virus (HDV), and twister catalyze nucleolytic intramolecular self-scission reactions (**Figure 3**) (Rupert and Ferré-D'Amaré,

2001; Martick and Scott, 2006; Cochrane et al., 2007; Cochrane and Strobel, 2008a; Suslov et al., 2015). In these ribozymes, it is proposed that the tautomeric forms of their catalytic guanosines act as a general base to activate the 2'-hydroxyl nucleophile to initiate the cleavage reactions (Klein et al., 2007; Cochrane and Strobel, 2008a).

During DNA replication, formation of minor tautomeric forms of nucleobases could lead to mispairing that generates mutations in the absence of any DNA lesion or other form of DNA damage (Watson and Crick, 1953; Löwdin, 1963; Topal and Fresco, 1976; Wang et al., 2011; Bebenek et al., 2011; Rangadurai et al., 2019; Rangadurai et al., 2020). This phenomenon is thought to explain the appearance of spontaneous mutations during the replication of undamaged DNA. The transient generation of minor tautomers in DNA allows stabilization of mismatches, in the polymerase active site, in conformations that are indistinguishable from the structures of canonical base pairs (**Figure 4**) (Watson and Crick, 1953; Topal and Fresco, 1976; Wang et al., 2011; Rangadurai et al., 2020).

While assembling their double-helix model of DNA, James Watson and Francis Crick stumbled over the phenomenon of base tautomerism; in fact, their model required that the bases adopt specific tautomeric forms in order to base-pair, with the suggestion that alternate tautomers would lead to mispairing and spontaneous mutations. (Watson and Crick, 1953). These ideas were later formalized by Michael Topal and Jacques Fresco, who described in 1976 the specific pairings between the minor tautomeric forms of canonical nucleic acid bases during replication that are likely to lead to mutations (Topal and Fresco, 1976). Since then, structural studies have shown that the formation of minor tautomers in DNA could allow wobble mismatches, such as the A-C and G-T base pairs, to adopt a geometry similar to that of canonical base pairs in Watson and Crick (W-C) conformations (**Figure 4**) (Wang et al., 2011; Szymanski et al., 2017). These studies provided structural evidence in support of the rare tautomer hypothesis for spontaneous mutations (Wang et al., 2011). Furthermore, tautomeric dynamics occur on a time scale that is similar to that of DNA replication, suggesting that the rate of rare tautomer formation could determine the probability of base misincorporation during replication (Rangadurai et al., 2019; Rangadurai et al., 2020).

In this review, we will discuss mechanisms involving tautomerism in RNA biology and in the generation of spontaneous mutations during DNA replication. We will also discuss the role of tautomerism in the development of nucleoside/nucleotide analogs based antiviral therapeutics. Specifically, we will focus on tautomerism in the catalysis of small self-cleaving ribozymes and in ligands recognition by riboswitches. We will discuss evidence of tautomerism in DNA and the associated mechanisms that contribute to the generation of spontaneous mutations. We will also review the role of tautomerism in the mechanism by which antiviral drug candidates based on nucleoside analogs, such as molnupiravir (isopropyl ester prodrug of the ribonucleoside analog  $\beta$ -D-N<sup>4</sup>-hydroxycytidine) and KP1212 (5-aza-5,6-dihydro-2'-deoxycytidine), induce lethal mutagenesis in the genomes of SARS-CoV-2 and HIV,



respectively. Minor tautomers are rare and their dynamics of interconversion is fast, which makes them challenging to identify in complex macromolecular systems, such as DNA and RNA. In the last section, we will summarize the current, state-of-the-art methods for studying tautomerism in nucleic acids and in antiviral nucleoside analogs, and discuss the future directions of the field.

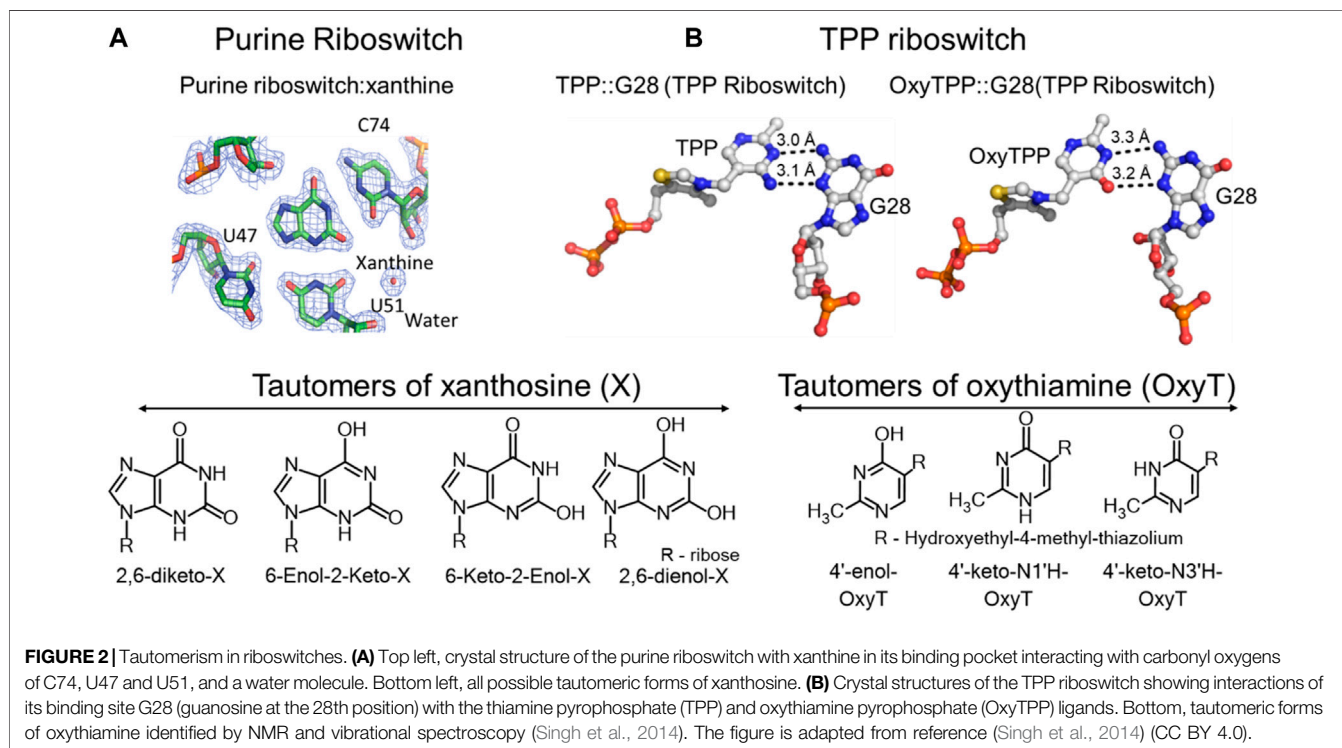
## Tautomerism in Nucleic Acids

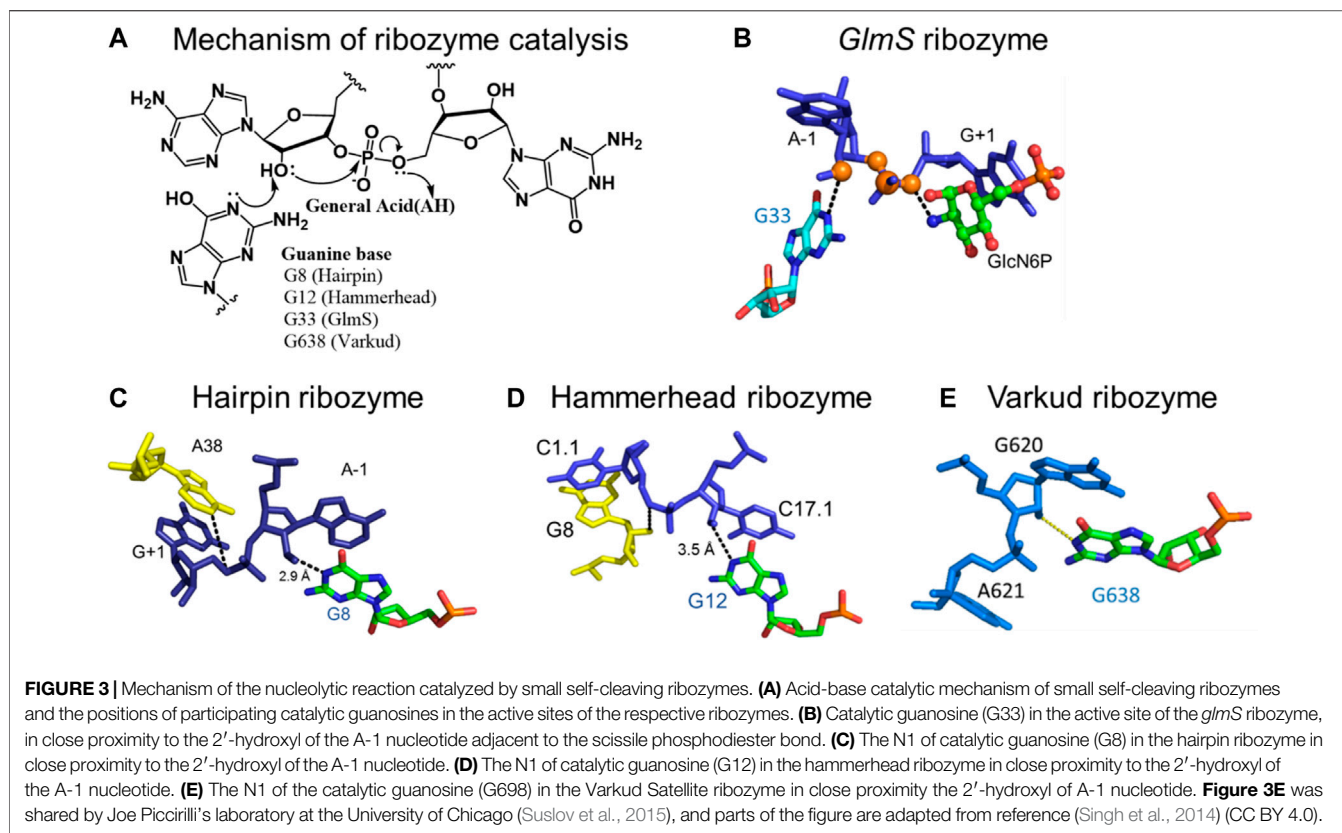
Nucleic acids display keto-enol and amino-imino types of tautomerism (Singh et al., 2015), as shown in **Figure 1** using cytosine as an example. This type of tautomerism, called prototropic tautomerism (Laar, 1886; Antonov, 2013) involves the movement of protons in the form of deprotonation at one site and protonation at another site, which is accompanied by a rearrangement of the double bonds in between the two sites. The base pairing complementarity in nucleic acids depends on

the position of protons on the Watson-Crick (W-C) interface (Watson and Crick, 1953; Wang et al., 2011). As the positions of protons change between different tautomers (**Figure 1**) (Singh et al., 2015), the distribution of hydrogen donor and acceptor sites is altered on the W-C face of each base. Therefore, certain minor tautomeric forms of canonical nucleic acid bases can engage in alternative base pairings (Watson and Crick, 1953; Topal and Fresco, 1976; Wang et al., 2011; Singh et al., 2015; Rangadurai et al., 2019; Rangadurai et al., 2020). Since base pairing interactions are predominantly responsible for the biological functions performed by nucleic acids, prototropic tautomerism is expected to influence the biology of both DNA and RNA (Singh et al., 2015). Below we discuss a few notable examples where nucleobase tautomerism is a key contributor or otherwise modulates the biological functions performed by nucleic acids.

## Tautomerism in RNA Riboswitches

The ligand recognition by the purine and thiamine pyrophosphate riboswitches involves binding to minor tautomeric forms of their non-natural ligands (**Figure 2**) (Thore et al., 2008; Gilbert et al., 2009; Singh et al., 2014). The purine riboswitch regulates expression of genes involved in the metabolism of adenine and guanine. The regulation is mediated by the conformational change in response to binding to purine ligands (Gilbert et al., 2009; Gilbert et al., 2006). In addition to recognizing adenine and guanine, the purine riboswitch can also bind to metabolites of purines such as xanthine, hypoxanthine and 2,6-diaminopurine (Gilbert et al., 2009; Gilbert et al., 2006). Crystal structures of these ligands bound to the purine riboswitch show that the cytosine at the 74 position (C74) is critical for determining substrate specificity (**Figure 2A**) (Gilbert et al.,

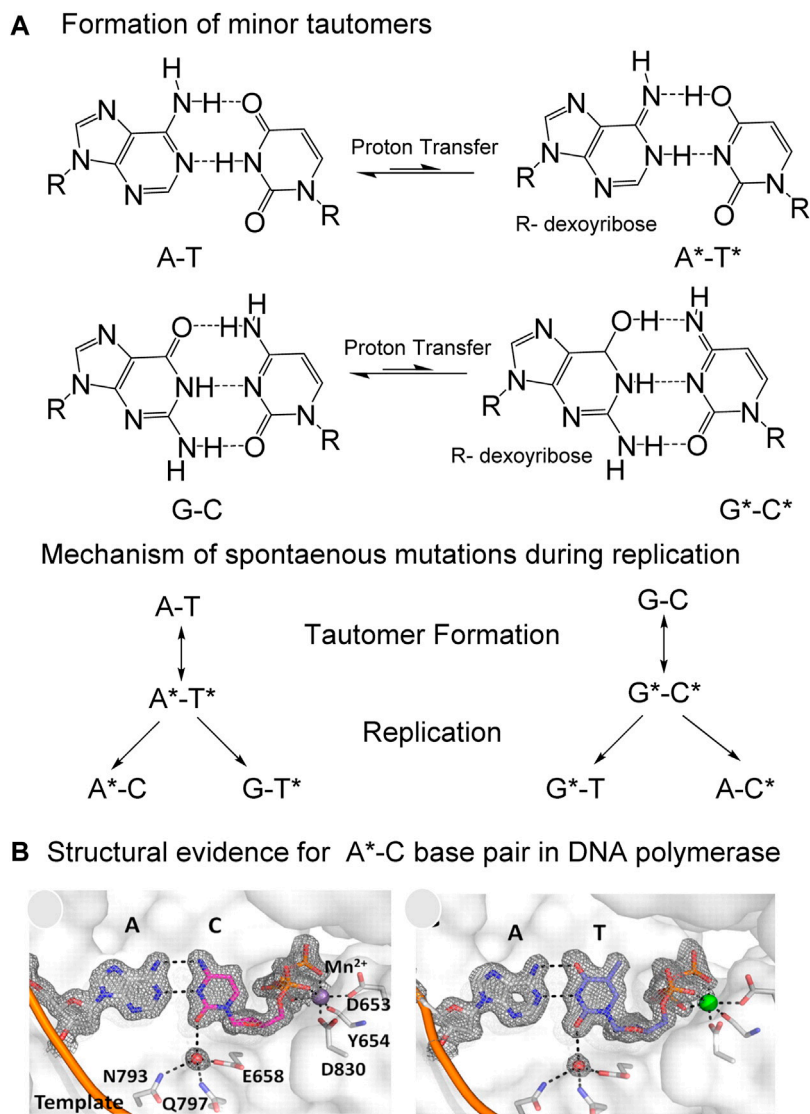




2009). The carbonyl oxygen (O2) of C74 and U51 form hydrogen bonds with the 2-amino functional group of the native ligand guanine (Gilbert et al., 2009), which explain the riboswitch preference for ligands that have a hydrogen bond donor at the 2-position of purines such as guanine and 2,6-diaminopurine (Gilbert et al., 2009; Gilbert et al., 2006). However, the riboswitch can also bind to xanthine, which has a carbonyl group at the 2-position, and hypoxanthine, which lacks any functional group at that position, albeit with a weaker affinity (Figure 2A) (Gilbert et al., 2009). Xanthine binding has a dissociation constant ( $K_D$ ) of 32  $\mu$ M, three orders of magnitude higher than the nanomolar  $K_D$ s for guanine and 2,6-diaminopurine (Gilbert et al., 2009). Hypoxanthine has 200-fold less affinity compared to guanine (Gilbert et al., 2009). However, the crystal structures of the riboswitch bound to xanthine or guanine ligands reveal no significant structural differences (Gilbert et al., 2009). To rationalize their identical structures, it was proposed that the 2-enol form of xanthine would alleviate the repulsive interactions that exist due to the presence of three negatively charged oxygens in close proximity (Figure 2A) (Singh et al., 2015). This proposal was also consistent with the biochemical observation that xanthine binds the riboswitch at a pH of 6 ( $K_D = 33 \mu$ M), but no binding is detected at a higher pH of 8.5 (Gilbert et al., 2009). The enol tautomers are expected to be more stable at lower pH (Gilbert et al., 2009). Despite the observed crystallographic and biochemical evidence in support of the minor 2-enol tautomer, the direct existence of 2-enol xanthine has not been established. This is due, in part, to the lack of sensitive methods for

distinguishing keto and enol forms of xanthine in the background overlapping signals from polymeric RNA (Gilbert et al., 2009; Singh et al., 2014).

Tautomerism is also suggested to influence oxythiamine pyrophosphate (OxyTPP) binding to the thiamine pyrophosphate (TPP) riboswitch (Thore et al., 2006; Thore et al., 2008). The TPP riboswitch binds to its natural ligand, TPP, to negatively regulate the expression of genes involved in the biosynthesis and transport of thiamine (Thore et al., 2006). 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 (Figure 2B) (Thore et al., 2006; Thore et al., 2008). The OxyTPP has a carbonyl group at the 4'-position. Yet, under crystalline conditions, its hydrogen bonding interactions to G28 are almost identical to those of TPP (Figure 2B) (Thore et al., 2008). The 4'-position of OxyTPP can only act as a hydrogen bond donor in its enol form, leading to the hypothesis that OxyTPP binds to the riboswitch as an enol tautomer (Thore et al., 2008). The Oxythiamine (Oxy) portion of OxyTPP exists in three tautomeric forms including the proposed 4'-enol tautomer (Singh et al., 2014). Biochemical binding isotope effect (BIE) experiments combined with density functional theory (DFT) calculations performed using O-18 labelled OxyTPP could not unambiguously identify the tautomer of OxyTPP in the binding pocket of the TPP riboswitch (Singh et al., 2014). More sensitive methods, such as NMR relaxation dispersion,



**FIGURE 4** | Proposed mechanism by which “spontaneous” mutations are introduced by the inter-strand movement of protons in Watson-Crick (W-C) base pairs during replication. **(A)** Spontaneous transfer of two protons from one strand to another in A-T and G-C base pairs generate minor tautomers, which can form mismatches during replication to cause mutations. **(B)** Structural evidence for the stabilization of A-C base pair in W-C conformation, almost indistinguishable from the A-T base pair in active site of a high fidelity DNA polymerase. Figure is adapted from reference (Wang et al., 2011; Slocombe et al., 2021).

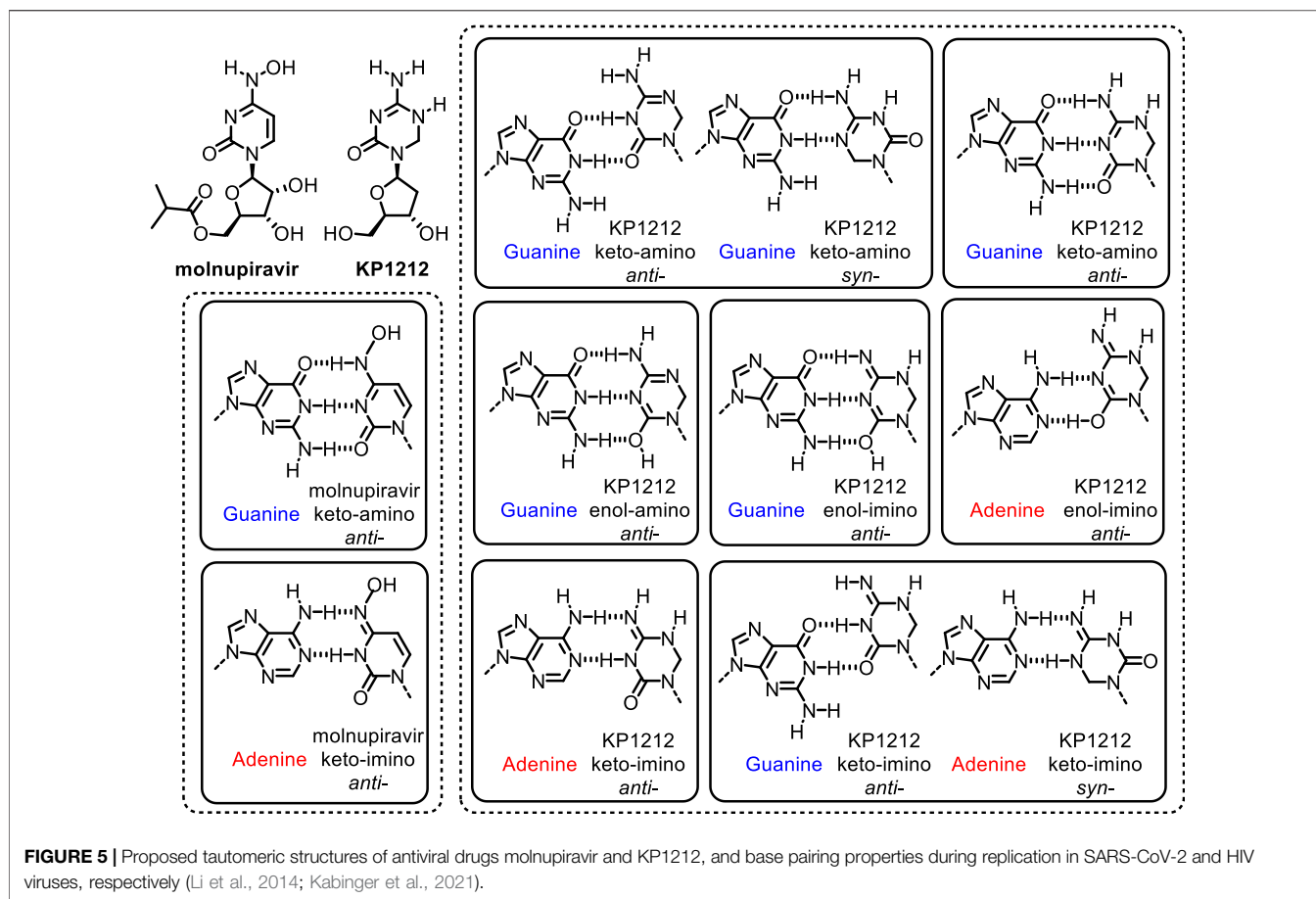
low temperature NMR, FTIR (Fourier Transform Infrared Spectroscopy) and 2D-IR (2-Dimensional Infrared Spectroscopy) (discussed in the Methods overview) are needed to unambiguously characterize the tautomeric form of OxyTPP or its binding partner G28 on the TPP riboswitch.

### Tautomerism in Self-Cleaving Ribozymes

Small self-cleaving ribozymes are proposed to utilize the minor tautomeric form(s) of catalytic guanosines to execute the first step of the self-cleavage reactions (**Figure 3**) (Cochrane and Strobel, 2008a). As mentioned above, the hammerhead, hairpin, *glmS*, VS, HDV, and the twister ribozymes are examples of small self-cleaving ribozymes that catalyze the nucleolytic intramolecular self-

scission reactions (**Figure 3**) (reviewed in (Cochrane and Strobel, 2008a)). Their reaction mechanism are very similar, and they typically involve base catalyzed activation of the site-specific 2'-hydroxyl that acts as a nucleophile to attack the adjacent scissile 3'-phosphate. This reaction yields two RNA products: one containing the 2',3'-cyclic phosphate and the other the 5'-hydroxyl functional group (**Figure 3A**). The only exception to this mechanism is the *glmS* ribozyme, which requires an external cofactor in the self-cleavage reaction (Cochrane et al., 2007).

The self-scission reactions catalyzed by these ribozymes are proposed to utilize the minor tautomeric form of catalytic guanosines, in which the N1 is not protonated, to act as a general base in the 2'-hydroxyl activation step of the reaction



(Singh et al., 2015) (**Figure 3A**). Structural studies of hammerhead (Martick and Scott, 2006), hairpin (Rupert and Ferré-D'Amaré, 2001), VS (Suslov et al., 2015) and the *glmS* (Cochrane et al., 2007; Klein and Ferré-D'Amaré, 2006) ribozymes identified the N1 of catalytic guanines in close proximity, within hydrogen bonding distance, to the 2'-hydroxyl nucleophile. These studies established that the N1 of G33 in *glmS* (**Figure 3B**) (Cochrane et al., 2007; Klein and Ferré-D'Amaré, 2006), G8 in hairpin (**Figure 3C**) (Fedor, 2000; Pinard et al., 2001; Rupert and Ferré-D'Amaré, 2001; Kuzmin et al., 2004; Bevilacqua and Yajima, 2006), G12 in hammerhead (**Figure 3D**) (McKay, 1996; Han and Burke, 2005; Martick and Scott, 2006; Thomas and Perrin, 2008), and G638 in VS (**Figure 3E**) (Lafontaine et al., 2001; Hiley et al., 2002; Sood and Collins, 2002; Suslov et al., 2015) play the role of a general base in the self-cleavage reactions catalyzed by these ribozymes.

However, the N1 of guanosine has a  $pK_a$  of  $\sim 10$  and is protonated at the physiological pH of  $\sim 7.4$  (Singh et al., 2015). The protonated N1 of guanosine is a poor base to abstract a proton from the 2'-hydroxyl, which has a  $pK_a$  of  $\sim 13$  (Velikyan et al., 2001). Therefore, tautomeric or ionic forms of the catalytic guanines in which the N1 is not protonated are expected to form transiently during the catalysis. These minor tautomers are likely more

nucleophilic at N1, and thus could extract the proton from the 2'-hydroxyl groups (Singh et al., 2015). Generation of N1 unprotonated guanosine would require perturbation of its  $pK_a$  towards neutrality. Significant perturbations in  $pK_a$ , by as much as four units, have been reported in RNA systems, including ribozymes and riboswitches (Legault and Pardi, 1997; Wilcox and Bevilacqua, 2013). Perturbation of  $pK_a$  towards neutrality would facilitate deprotonation at the N1 site of catalytic guanines because prototropic tautomerism involves the deprotonation and the protonation steps, and is optimal in functional groups whose  $pK_a$ 's are close to neutral (Singh et al., 2015).

Despite significant structural and biochemical evidence in support for the presence minor tautomeric forms of catalytic guanines in small self-cleaving ribozymes, their direct identification has proven challenging, owing to the lack of sensitive methods (Singh et al., 2015).

### Tautomerism in DNA and its Role in Replication Fidelity

The genetic integrity of genomic DNA relies on adenine, guanine, cytosine and thymine existing predominantly in their keto and amino tautomeric forms during replication and transcription (Watson and Crick, 1953; Topal and Fresco, 1976; Wang et al., 2011; Rangadurai et al., 2020).

Therefore, replication fidelity is expected to be influenced by the formation of minor tautomers (Watson and Crick, 1953; Topal and Fresco, 1976; Wang et al., 2011; Rangadurai et al., 2019; Rangadurai et al., 2020). In their work on the structure of DNA, Watson and Crick did appreciate that the formation of minor tautomeric forms would alter the base pairing properties of nucleic acid bases, potentially with mutagenic consequences (Watson and Crick, 1953). These minor tautomeric forms could arise from inter-helical transfer of protons in a DNA duplex (**Figure 4**) (Löwdin, 1963; Sevilla et al., 1995). Transient formation of minor tautomeric forms of DNA bases, and their stabilization in the active site of DNA polymerases during replication, could lead to incorporation of mismatched base pairs (Topal and Fresco, 1976; Watson and Crick, 1953; Wang et al., 2011). This phenomenon is plausible considering that the kinetics of minor tautomer formation and their lifetime in the active site is comparable with the kinetics of nucleotide incorporation by the polymerase. Therefore, the probability of base misincorporation during DNA replication may be correlated with the probability of rare tautomer formation (**Figure 4A**) (Topal and Fresco, 1976; Wang et al., 2011; Peng et al., 2015; Rangadurai et al., 2019).

Structural evidence for the rare tautomer hypothesis for spontaneous mutation came from the high-resolution crystal structure of a DNA polymerase that catalyzes replication in crystals (**Figure 4B**) (Wang et al., 2011). It was observed that a C•A mismatch mimics the shape of the cognate C•G base pair in the crystal (Wang et al., 2011). The movement of protons in the mismatched bases alter the hydrogen-bonding pattern such that the base pairs involving the minor tautomeric forms adopt an overall shape that is virtually indistinguishable from the canonical W-C base pair in DNA (Wang et al., 2011). This “shape mimicry” allows the mismatch to evade error detection mechanisms of human polymerases (**Figure 4B**). These observations provided structural support for the rare tautomer hypothesis of spontaneous mutagenesis.

NMR dispersion methods have allowed us to gain better understanding of tautomeric dynamics in DNA, and in RNA duplexes (Kimsey et al., 2018; Rangadurai et al., 2019; Rangadurai et al., 2020). Using W-C mismatches such as G•T or G•U, these studies established that G•T/U wobble mismatches exist in dynamic equilibrium between three distinct W-C mismatched base pairs within the DNA and RNA duplexes (Kimsey et al., 2018; Rangadurai et al., 2019; Rangadurai et al., 2020). The three distinct W-C mismatches include two tautomeric and one anionic species (Kimsey et al., 2018; Rangadurai et al., 2019; Rangadurai et al., 2020). The tautomeric forms were established using the chemical shifts of guanine N1 and thymidine/uridine N3. The chemical shifts, although consistent with G<sup>enol</sup>•T/U base pair (minor enol tautomer of G paired with the dominant keto tautomeric form of T or U) were partially skewed toward G•T<sup>enol</sup>/U<sup>enol</sup> pairs (keto-amino tautomer of G paired with enol tautomer of T or U) (Kimsey et al., 2018). This skewness was interpreted as evidence for a rapid equilibrium between the major G<sup>enol</sup>•T/U and the minor G•T<sup>enol</sup>/U<sup>enol</sup> tautomeric base pairs (Kimsey et al., 2018). The two rapidly exchanging tautomeric species (G<sup>enol</sup>•T/

U ⇌ G•T<sup>enol</sup>/U<sup>enol</sup>) were also quantitated and found to be around 0.4% of the total population at neutral pH. Increased understanding of base pairing mismatches like G•T, G•U, and A•C that nevertheless adopt W-C like geometry through either tautomerization or ionization allowed us to appreciate that these mismatches appear to be more common (Kimsey et al., 2018; Rangadurai et al., 2019; Rangadurai et al., 2020). Stabilization of W-C mismatches through tautomerization, either in the template strand or in the incoming nucleotide, could allow the incorporation of mismatches during replication (Topal and Fresco, 1976; Wang et al., 2011; Rangadurai et al., 2020), and remains a compelling mechanistic explanation for spontaneous mutagenesis.

## Therapeutic Implications of Tautomeric Nucleosides

Nucleoside/nucleotide analogs enriched in minor tautomers are effective as antiviral agents against many viruses (Crotty et al., 2001; Baranovich et al., 2013; Li et al., 2014; Singh et al., 2015; Delang et al., 2018; Kabinger et al., 2021). The antiviral property of these analogs stem from their ability to induce mutations in viral genomes (Li et al., 2014; Delang et al., 2018; Gordon et al., 2021; Kabinger et al., 2021). Such nucleoside analogs are mutagenic, in part, because they form significant amounts of minor tautomeric forms, which enables them to engage in ambiguous base pairing, i.e., pairing with more than one base (Li et al., 2014; Delang et al., 2018; Kabinger et al., 2021). The concept of specifically increasing the mutation rates of pathogens to drive their population collapse is particularly effective against retroviruses and riboviruses (Crotty et al., 2001; Anderson et al., 2004; Domingo et al., 2008; Manrubia et al., 2010; Domingo et al., 2012). These viruses have high mutation rates and their population lives close to the mutational threshold called the error catastrophe limit (ECL), a theoretical mutational rate above which producing viable viral progeny becomes impossible and leads to viral population collapse (Eigen, 2002). This phenomenon is called lethal mutagenesis. Additionally, unlike the human replicative polymerases that feature both high selectivity and high fidelity, the replication machinery of retroviruses is both promiscuous and error prone and thus, it can readily incorporate mutagenic nucleoside analogs (Anderson et al., 2004; Li et al., 2014; Kabinger et al., 2021). Therefore, these nucleoside analogs selectively mutagenize the viral genomes, while leaving the host (human) genome essentially untouched (Crotty et al., 2001; Anderson et al., 2004; Delang et al., 2018; Kabinger et al., 2021).

Nucleoside analogs that induce lethal mutagenesis are effective as antiviral drugs against many retroviruses and riboviruses, such as HIV, HCV, influenza virus and coronaviruses (Anderson et al., 2004; Baranovich et al., 2013; Li et al., 2014; Delang et al., 2018; Gordon et al., 2021; Kabinger et al., 2021). These viruses have high mutation rates and their replication machinery is error-prone (Watson and Crick, 1953; Crotty et al., 2001; Wang et al., 2011; Li et al., 2014; Rangadurai et al., 2020; Kabinger et al., 2021). The error prone nature of some viral polymerases allows incorporation of modified nucleosides (Crotty et al., 2001; Li

et al., 2014; Peng et al., 2015; Kabinger et al., 2021). The 5-aza-5,6-dihydro-2'-deoxycytidine (KP1212) is an experimental drug candidate that targets HIV by increasing the mutation rate of the virus (Figure 5) (Harris et al., 2005; Mullins et al., 2011; Li et al., 2014). The antiviral efficacy of KP1212 stems from its ability to exist in multiple tautomeric states (Li et al., 2014; Peng et al., 2015). These tautomers can engage in ambiguous base pairings to induce mutagenesis in HIV (Li et al., 2014; Peng et al., 2015). Other antiviral drugs such as ribavirin for HCV (Crotty et al., 2001; Li et al., 2014), and favipiravir for influenza (Baranovich et al., 2013; Delang et al., 2018) are also efficacious due, in part, to their ability to exist in multiple tautomeric or rotameric forms and induce mutagenesis in viral genomes. Lethal mutagenesis has also been implicated as the main mechanism of action for the recently developed small-molecule therapeutics, such as molnupiravir from Merck, for severe acute respiratory syndrome coronavirus (SARS-CoV-2, the virus that causes Covid-19) (Kabinger et al., 2021; Zhou, 2021).

The mechanism by which KP1212 induces lethal mutagenesis in HIV has been studied extensively (Li et al., 2014; Peng et al., 2015). KP1212 is mutagenic, inducing G to A and A to G transition mutations (Figure 5) (Harris et al., 2005; Li et al., 2014). The mutagenicity of KP1212 is due, in part, to its ability to exist in multiple tautomeric or rotameric forms (Peng et al., 2013; Li et al., 2014; Peng et al., 2015). While KP1212 is considered a dC (deoxycytidine) analog, because it features the functional groups and the W-C face of dC, it has a near neutral pK<sub>a</sub> of ~7, compared to pK<sub>a</sub> of ~4 for dC. As a consequence, KP1212 in solution exists in multiple tautomeric states (Li et al., 2014; Peng et al., 2015). These include keto-amino, keto-imino and its rotameric form, enol-amino and enol-imino and its rotameric form (Li et al., 2014). The enol-amino and enol-imino are the dominant tautomeric forms of KP1212, in contrast to the dominant keto-amino form observed for dC (Li et al., 2014; Peng et al., 2015). The alternative tautomeric forms of KP1212 are expected to have perturbed base pairing properties (Li et al., 2014), and thus contribute to the mutagenic properties of the base. KP1212 was shown to induce G to A and A to G mutations in the HIV genomes in cellular models, in preclinical rodents studies and in HIV patients who participated in the clinical trials for KP1212 (Harris et al., 2005; Mullins et al., 2011; Li et al., 2014). However, these studies also revealed that KP1212 is not mutagenic to human cells. Lack of mutagenicity in humans is likely due to the higher fidelity and selectivity of human DNA replicative polymerases (Mullins et al., 2011). Unlike dC, KP1212 has a saturated carbon center at the 6-position, which causes the base ring to be puckered (Li et al., 2014). This geometric distortion is significant enough to be selected against by the human polymerases. KP1212, therefore, exemplifies the features of an ideal lethal mutagen because it specifically induces mutations in HIV, driving viral population collapse, while sparing the human genome.

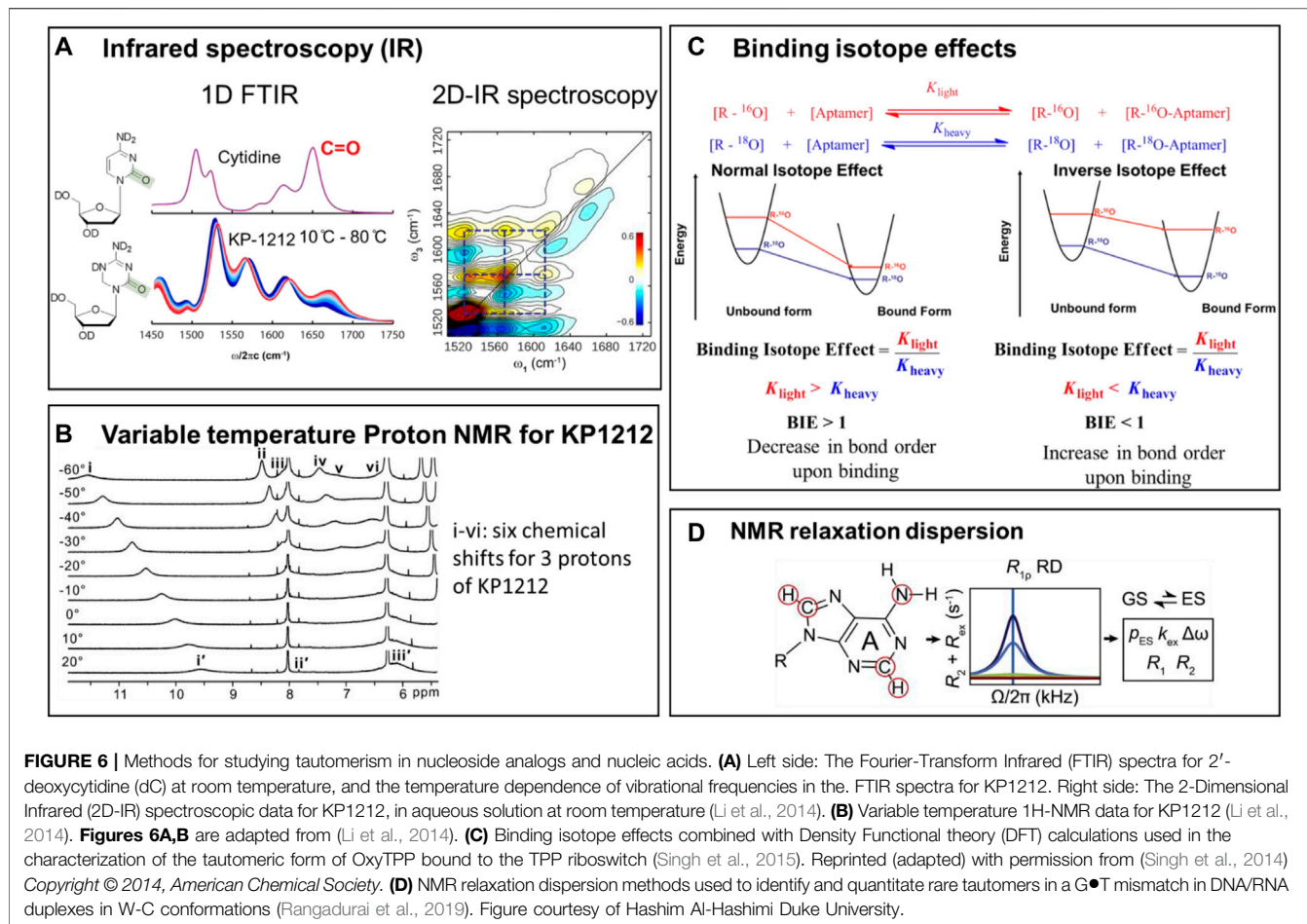
Molnupiravir is a broad-spectrum nucleoside analog drug that is efficacious against many viruses including the proofreading-intact SARS-CoV-2 coronavirus with a high genetic barrier to resistance (Gordon et al., 2021; Agostini et al., 2019). Its efficacy stems from its ability to induce lethal

mutagenesis in SARS-CoV-2, during the early stages of Covid-19 (Figure 5) (Agostini et al., 2019; Kabinger et al., 2021; Gordon et al., 2021; Menéndez-Arias, 2021). Molnupiravir is an isopropyl ester prodrug of the ribonucleoside analog β-D-N<sup>4</sup>-hydroxycytidine (NHC) (Kabinger et al., 2021). It is currently in the phase-III clinical trial for Covid-19 (Kabinger et al., 2021). Biochemical studies using SARS-CoV-2 RNA-dependent-RNA polymerase (SARS-CoV-2 RdRp) have shown that the triphosphate form of NHC can be incorporated into RNA, albeit at a lower frequency compared to canonical nucleosides (Gordon et al., 2021; Menéndez-Arias, 2021). Once incorporated into the growing strand, it can be extended at its 3'-end. When present in the template strand it can form base pairs with G (NHC:G) or A (NHC:A) to induce G to A or A to G mutations (Figure 5) (Gordon et al., 2021). The NHC:A base pairs is more efficiently extended compared to NHC:G (Kabinger et al., 2021). Cellular studies have shown that NHC is 100-fold more active than ribavirin and favipiravir against SARS-CoV-2 (Zhou, 2021). The higher activity molnupiravir also induces a higher mutation frequency in the viral RNA (Menéndez-Arias, 2021; Zhou, 2021). Structural analysis of RdRp-RNA complexes containing mutagenesis products revealed that the NHC moiety can form stable base pairs with either G or A in the active site of RdRp (Kabinger et al., 2021). The formation of stable base mispairs with G or A, in W-C conformations, allow molnupiravir to escape the proofreading mechanism of the SARS-CoV-2 RdRp polymerase (Agostini et al., 2019; Gordon et al., 2021; Kabinger et al., 2021; Menéndez-Arias, 2021). The accumulation of mutations with each round of viral multiplication would lead to weakening of the virus (Kabinger et al., 2021; Menéndez-Arias, 2021). Similar to molnupiravir against SARS-CoV-2, ribavirin and favipiravir target HCV, and influenza, respectively, in part by the lethal mutagenesis mechanism (Crotty et al., 2001; Baranovich et al., 2013). Taken together, all the above studies underscore the importance of better understanding the mechanisms by which nucleoside analogs are incorporated and lead to mispaired bases in viral genomes.

## Methods for Studying Tautomerism in Nucleic Acids and in Nucleoside Analogs

Minor tautomers of canonical nucleic acid bases and their analogs are rare, transient, and interconvert on a fast time scale (Topal and Fresco, 1976; Rangadurai et al., 2019; Peng et al., 2015). Recent developments of spectroscopic methods have allowed their direct identification in nucleic acids and in nucleoside analogs (Figure 6) (Wang et al., 2011; Peng and Tokmakoff, 2012; Peng et al., 2013; Singh et al., 2014; Li et al., 2014; Peng et al., 2015; Szymanski et al., 2017; Rangadurai et al., 2019; Rangadurai et al., 2020; Kabinger et al., 2021). Prototropic tautomers differ from one another based on the position of protons, and the position of double bonds. Therefore, methods that are capable of detecting and distinguishing protons in different chemical





environment (for example, NMR) are useful for the direct identification of tautomeric forms (Singh et al., 2014; Li et al., 2014; Peng et al., 2013) in both nucleic acids and nucleoside analogs (Figures 6B,C) (Li et al., 2014; Peng et al., 2015; Rangadurai et al., 2019; Rangadurai et al., 2020). Similarly, methods that provide information on the bond order and strength of a specific chemical bond (such as IR and Raman spectroscopy) are useful for distinguishing between keto-enol or amino-imino tautomeric forms, because the functional groups in each tautomeric pair have very different and characteristic vibrational properties. In addition, biochemical binding isotope effects and their interpretation using computational DFT have been used for characterizing tautomeric forms of a ligand in the binding pocket of an RNA aptamer (Figure 6C) (Singh et al., 2014).

NMR measures the chemical shift of individual protons, a property that depends on the chemical environment of the proton. Since the protons in tautomeric pairs are attached to different heteroatoms, their NMR chemical shifts are very different and thus distinguishable. Often, because the tautomeric equilibria are faster than the NMR acquisition time scale, broader peaks are observed, that span areas of multiple tautomeric forms. However, by substantially lowering the temperature of the sample (a method called variable

temperature NMR), the tautomeric equilibria can be slowed down sufficiently to resolve the individual tautomeric forms.

Infrared spectroscopy (IR) measures the vibrational properties of the chemical bonds. Depending on their chemical makeup and bond order, different bonds absorb infrared radiation at different wavelengths. IR spectra provide direct information on the types of bonds present in a molecule, and therefore can easily distinguish between keto-enol or amino-imino tautomeric forms. Two-dimensional IR (2D-IR), an advanced version of IR, uses the same principle but provides, in addition, information on bond connectivity. Bonds in the same molecule vibrate together, and thus give a characteristic cross peak in the spectrum, which allows the distinction between the individual tautomeric forms present at equilibrium.

In addition to the direct detection of rare tautomers in nucleic acids, which is very challenging, indirect approaches, based on binding isotope effects (BIEs), exist that allow inference of minor tautomer formation. Theoretically, BIEs report on changes in vibrational frequencies of a ligand upon binding to its target. Experimentally, they are estimated using the binding affinities of ligands carrying either light or heavy isotopes of atoms directly involved in the binding interaction. The experimental BIEs are then compared to BIEs calculated, using the Density Functional Theory (DFT), from frequencies of optimized tautomeric forms.

The exact characterization is completed by identifying a tautomeric structure whose calculated BIEs closely match the experimental BIEs. A good example of this approach is the identification of tautomeric form of OxyTPP recognized by the TPP riboswitch (**Figure 6C**), where spectroscopic observations indicated that each tautomer has a unique vibrational frequency pattern. This approach was used to study OxyTPP bound to the TPP riboswitch (**Figure 6C**), using  $^{18}\text{O}$  and  $^{32}\text{P}/^{33}\text{P}$  labeled OxyTPP (Singh et al., 2014; Singh et al., 2015). Although the calculated BIEs are different for the keto and enol forms of OxyTPP, the magnitude of the difference between the BIEs corresponding to various tautomers of OxyTPP was too small to yield a definite conclusion. More precise measurements of BIEs are nevertheless expected to provide clarity on the tautomeric form of OxyTPP bound to the TPP riboswitch (Sood and Collins, 2002; Singh et al., 2014). These studies also underscore the potential of indirect measurements like BIEs and kinetic isotope effects (KIEs) to increase our understanding of tautomerism in nucleic acids and base analogs.

The variable temperature NMR, FT-IR (Fourier Transform Infrared) and 2D-IR (2-dimensional Infrared) methods were used to identify the tautomers of KP1212 and oxythiamine (Singh et al., 2014; Li et al., 2014; Peng et al., 2015). Tautomeric dynamics of these nucleoside analogs are fast, typically on the nanosecond timescale (Peng et al., 2013; Peng et al., 2015). Therefore, attempts to distinguish between the multiple tautomers of KP1212 at room temperature using NMR were not successful. Lowering the temperature to  $-20^\circ\text{C}$  allowed identification of individual tautomers of KP1212 and oxythiamine by 1-dimensional proton NMR ( $^1\text{H}$ -NMR) in the dimethylformamide (DMF) solvent (**Figure 6B**) (Singh et al., 2014; Li et al., 2014; Peng et al., 2015). While the characterization and distribution of tautomers of KP1212 and oxythiamine by  $^1\text{H}$ -NMR utilized non-physiological conditions with low temperature and an aprotic solvent (DMF) (Li et al., 2014), these experiments clearly outlined the chemical plausibility of the tautomeric forms of these compounds. For detecting multiple tautomers under physiologically relevant aqueous conditions at room temperature, IR-based methods have proven to be more effective (Singh et al., 2014; Li et al., 2014; Peng et al., 2015). The temperature dependence of change in amplitude of vibrational frequencies in the 1D-FTIR spectrum confirmed the presence of multiple tautomers for KP1212 and oxythiamine (**Figure 6C**) (Li et al., 2014). To directly identify minor tautomers based on altered vibrational frequency of minor tautomers, 2D-IR data was combined with DFT calculations (**Figure 6A**) (Peng et al., 2015). The 2D IR spectrum has a greater spatial and temporal resolution than FTIR, and thus allows direct identification of multiple fast interconverting tautomers (Peng et al., 2013; Li et al., 2014; Peng et al., 2015). These studies established that KP1212 exists in seven different tautomeric or rotameric forms and oxythiamine exists in three different tautomeric forms (Li et al., 2014; Singh et al., 2014). Furthermore, the biochemical observations that KP1212 is 10% mutagenic when replicated *in vitro* and in living cells, and induces G to A mutations, can be rationalized by the tautomeric

distribution identified from these spectroscopic studies (Li et al., 2014; Peng et al., 2015).

By contrast with the nucleoside analogs described above, the minor tautomers of the canonical DNA bases are significantly less abundant, shorter lived and more challenging to detect (Peng et al., 2013; Rangadurai et al., 2019; Peng et al., 2015). NMR relaxation dispersion (RD) methods allow identification of low-abundance short-lived conformational states in biomolecules (**Figure 6D**) (Mulder et al., 2001; Rangadurai et al., 2019; Rangadurai et al., 2020). NMR relaxation dispersion (RD) has been used to characterize rare tautomers in Hoogsteen base pairs versus Watson-Crick base pairs in different types of DNA/RNA, such as A or B type of DNA/RNA (Nikolova et al., 2011; Alvey et al., 2014; Zhou et al., 2016; Rangadurai et al., 2018). These results offer understandings into differences between A-RNA and B-DNA duplexes and provide possible explanations for how they respond to damage and modifications.

Using NMR RD, it has been shown that the G•T/U mismatches exist in dynamic equilibrium between tautomeric and anionic W-C conformations within the DNA and RNA duplexes. These studies also provided insights into the kinetics of rare tautomer formation in nucleic acids, and the probability of base misincorporation due to the formation of minor tautomers (Rangadurai et al., 2019; Rangadurai et al., 2020). These methods continue to enhance our understanding of the mechanisms by which the formation of minor tautomers contribute to the generation of spontaneous mutations.

Structural methods based on X-ray crystallography and Cryo-EM (Cryogenic electron microscopy) have proven effective in studying mismatched base pairs that potentially involve minor tautomers. While protons are very difficult to visualize directly by X-ray crystallography and Cryo-EM, the overall geometry of base pairs and the inter-heteroatom distances can be accurately measured, and thus allow to infer the presence of protons and hydrogen bonding interactions. Specifically, crystallographic studies have shown that mismatches in high fidelity DNA polymerases can adopt structures that closely mimic the W-C base pairing geometry (Wang et al., 2011; Kabinger et al., 2021). The existence of mismatches in conformations that are identical to W-C shapes may allow them to evade the polymerase proof reading mechanisms and lead to mutagenesis (**Figure 4B**, **Figure 5A**) (Bebenek et al., 2011; Wang et al., 2011; Fedeles et al., 2015; Kabinger et al., 2021). In sum, these studies have provided convincing structural evidence in support of rare tautomer hypothesis for the generation of spontaneous mutations that arise during nucleic acid replication.

## FUTURE DIRECTIONS

The last decade has seen significant progress in our understanding of tautomerism in DNA and RNA. Despite the early realization that tautomerism could influence nucleic acid biology, directly identifying rare tautomers of nucleic acid bases proved very challenging. The minor tautomers of DNA bases are present at equilibrium in minute amounts ( $< 0.1\%$ ) and their dynamics of interconversion is generally very fast. Until recently,

little progress was made in directly visualizing minor tautomers of DNA bases. The NMR RD allowed direct detection of rare tautomers and characterized the dynamics of their formation in short DNA sequences. Structurally establishing that mismatches can be stabilized in W-C conformation by high fidelity DNA polymerases, provided crystallographic support for the rare tautomer hypothesis of spontaneous mutations. Advancement in methods based on BIEs, high-resolution crystallography, NMR and IR significantly enhanced our mechanistic understanding of mutagenesis from nucleoside analogs based antiviral drugs. While all these advancements are notable, they only reflect special cases of tautomerism in a limited set of experimental conditions. Much work remains to develop more broadly applicable methods that enable the direct visualization and identification of minor tautomers of canonical bases in complex biological systems, such as genomic DNA, and in the binding pocket of polymerases, under physiological conditions. Sensitivity improvements in all the methods summarized above are all expected to improve our understanding of tautomerism. In the case of BIEs, more precise measurements are expected to help identify minor tautomers in the context of RNA systems and may provide further insights into the role of tautomerism in RNA biology, including catalytic processes and ribozymes. More work is also needed to bridge the knowledge gap between the in-solution tautomeric equilibria and dynamics of nucleobases and analogs, and their biological consequences. In the case of antiviral nucleoside analogs, such as KP1212 and molnupiravir, there is great interest to understand how tautomeric equilibria translate into mutagenic potential

inside the viral polymerases. These nucleoside analogs have already proven very effective as antiviral drugs. However, our understanding of the physico-chemical properties that enable these compounds to adopt multiple tautomeric forms is still limited. Further insight into tautomeric equilibria and dynamics should allow us to develop more potent and safer antiviral therapeutics.

## AUTHOR CONTRIBUTIONS

All the authors made substantial, direct, and intellectual contribution to the manuscript and approved it for publication.

## FUNDING

BF is supported by NIH grants R01-CA080024, P30-ES002109, P42-ES027707 and T32-ES007020. DL is supported by R01-ES028865. VS is supported by Novartis Institute of Biomedical Research.

## ACKNOWLEDGMENTS

The authors thank Michael Romanowski of Novartis Institute of Biomedical Research (NIBR); Hashim Al-Hashimi of Duke University for the NMR dispersion figure (**Figure 6D**), and Joe Piccirilli of University of Chicago for providing the Varkud ribozyme figure (**Figure 3E**).

## REFERENCES

- Agostini, M. L., Pruijssers, A. J., Chappell, J. D., Gribble, J., Lu, X., Andres, E. L., et al. (2019). Small-Molecule Antiviral  $\beta$ -d-N 4-Hydroxycytidine Inhibits a Proofreading-Intact Coronavirus with a High Genetic Barrier to Resistance. *J. Virol.* 93 (24), e01348–19. doi:10.1128/JVI.01348-19
- Alvey, H. S., Gottardo, F. L., Nikolova, E. N., and Al-Hashimi, H. M. (2014). Widespread Transient Hoogsteen Base Pairs in Canonical Duplex DNA with Variable Energetics. *Nat. Commun.* 5 (1), 4786. doi:10.1038/ncomms5786
- Anderson, J. P., Daifuku, R., and Loeb, L. A. (2004). Viral Error Catastrophe by Mutagenic Nucleosides. *Annu. Rev. Microbiol.* 58, 183–205. doi:10.1146/annurev.micro.58.030603.123649
- Antonov, L. (2013). *Tautomerism: Methods and Theories*. Sofia, Bulgaria: John Wiley & Sons.
- Baranovich, T., Wong, S.-S., Armstrong, J., Marjuki, H., Webby, R. J., Webster, R. G., et al. (2013). T-705 (Favipiravir) Induces Lethal Mutagenesis in Influenza A H1N1 Viruses *In Vitro*. *J. Virol.* 87 (7), 3741–3751. doi:10.1128/jvi.02346-12
- Bebenek, K., Pedersen, L. C., and Kunkel, T. A. (2011). Replication Infidelity via a Mismatch with Watson-Crick Geometry. *Proc. Natl. Acad. Sci. USA* 108 (5), 1862–1867. doi:10.1073/pnas.1012825108
- Bevilacqua, P., and Yajima, R. (2006). Nucleobase Catalysis in Ribozyme Mechanism. *Curr. Opin. Chem. Biol.* 10 (5), 455–464. doi:10.1016/j.cbpa.2006.08.014
- Brown, R. D., Godfrey, P. D., McNaughton, D., and Pierlot, A. P. (1989). Tautomers of Cytosine by Microwave Spectroscopy. *J. Am. Chem. Soc.* 111 (6), 2308–2310. doi:10.1021/ja00188a058
- Cochrane, J. C., Lipchock, S. V., and Strobel, S. A. (2007). Structural Investigation of the GlmS Ribozyme Bound to its Catalytic Cofactor. *Chem. Biol.* 14 (1), 97–105. doi:10.1016/j.chembiol.2006.12.005
- Cochrane, J. C., and Strobel, S. A. (2008). Catalytic Strategies of Self-Cleaving Ribozymes. *Acc. Chem. Res.* 41 (8), 1027–1035. doi:10.1021/ar800050c
- Cochrane, J. C., and Strobel, S. A. (2008). Riboswitch Effectors as Protein Enzyme Cofactors. *Rna* 14 (6), 993–1002. doi:10.1261/rna.908408
- Colominas, C., Luque, F. J., and Orozco, M. (1996). Tautomerism and Protonation of Guanine and Cytosine. Implications in the Formation of Hydrogen-Bonded Complexes. *J. Am. Chem. Soc.* 118 (29), 6811–6821. doi:10.1021/ja954293l
- Crotty, S., Cameron, C. E., and Andino, R. (2001). RNA Virus Error Catastrophe: Direct Molecular Test by Using Ribavirin. *Proc. Natl. Acad. Sci.* 98 (12), 6895–6900. doi:10.1073/pnas.111085598
- Delang, L., Abdelnabi, R., and Neyts, J. (2018). Favipiravir as a Potential Countermeasure against Neglected and Emerging RNA Viruses. *Antivir. Res.* 153, 85–94. doi:10.1016/j.antiviral.2018.03.003
- Domingo, E., Grande-Pérez, A., and Martín, V. (2008). Future Prospects for the Treatment of Rapidly Evolving Viral Pathogens: Insights from Evolutionary Biology. *Expert Opin. Biol. Ther.* 8 (10), 1455–1460. doi:10.1517/14712598.8.10.1455
- Domingo, E., Sheldon, J., and Perales, C. (2012). Viral Quasispecies Evolution. *Microbiol. Mol. Biol. Rev.* 76 (2), 159–216. doi:10.1128/mmbr.05023-11
- Eigen, M. (2002). Error Catastrophe and Antiviral Strategy. *Proc. Natl. Acad. Sci.* 99 (21), 13374–13376. doi:10.1073/pnas.212514799
- Fedele, B. I., Freudenthal, B. D., Yau, E., Singh, V., Chang, S.-c., Li, D., et al. (2015). Intrinsic Mutagenic Properties of 5-chlorocytosine: A Mechanistic Connection between Chronic Inflammation and Cancer. *Proc. Natl. Acad. Sci. USA* 112 (33), E4571–E4580. doi:10.1073/pnas.1507709112
- Fedor, M. J. (2000). Structure and Function of the Hairpin Ribozyme. *J. Mol. Biol.* 297 (2), 269–291. doi:10.1006/jmbi.2000.3560

- Gilbert, S. D., Mediatore, S. J., and Batey, R. T. (2006). Modified Pyrimidines Specifically Bind the Purine Riboswitch. *J. Am. Chem. Soc.* 128 (44), 14214–14215. doi:10.1021/ja063645t
- Gilbert, S. D., Reyes, F. E., Edwards, A. L., and Batey, R. T. (2009). Adaptive Ligand Binding by the Purine Riboswitch in the Recognition of Guanine and Adenine Analogs. *Structure* 17 (6), 857–868. doi:10.1016/j.str.2009.04.009
- Gordon, C. J., Tchesnokov, E. P., Schinazi, R. F., and Götte, M. (2021). Molnupiravir Promotes SARS-CoV-2 Mutagenesis via the RNA Template. *J. Biol. Chem.* 297, 100770. doi:10.1016/j.jbc.2021.100770
- Han, J., and Burke, J. M. (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 (21), 7864–7870. doi:10.1021/bi047941z
- Harris, K., Brabant, W., Styrchak, S., Gall, A., and Daifuku, R. (2005). KP-1212/1461, a Nucleoside Designed for the Treatment of HIV by Viral Mutagenesis. *Antivir. Res.* 67 (1), 1–9. doi:10.1016/j.antiviral.2005.03.004
- Hiley, S. L., Sood, V. D., and Fan, J. (2002). 4-thio-U Cross-Linking Identifies the Active Site of the VS Ribozyme. *Embo J* 21 (17), 4691–4698. doi:10.1093/emboj/cdf462
- Kabinger, F., Stiller, C., Schmitzová, J., Dienemann, C., Kocic, G., Hillen, H. S., et al. (2021). Mechanism of Molnupiravir-Induced SARS-CoV-2 Mutagenesis. *Nat. Struct. Mol. Biol.* 28 (9), 740–746. doi:10.1038/s41594-021-00651-0
- Kimsey, I. J., Szymanski, E. S., Zahurancik, W. J., Shakya, A., Xue, Y., Chu, C.-C., et al. (2018). Dynamic Basis for dGdT Misincorporation via Tautomerization and Ionization. *Nature* 554 (7691), 195–201. doi:10.1038/nature25487
- Klein, D. J., Been, M. D., and Ferré-D'Amaré, A. R. (2007). Essential Role of an Active-Site Guanine in glmS Ribozyme Catalysis. *J. Am. Chem. Soc.* 129 (48), 14858–14859. doi:10.1021/ja0768441
- Klein, D. J., and Ferré-D'Amaré, A. R. (2006). Structural Basis of glmS Ribozyme Activation by Glucosamine-6-Phosphate. *Science* 313 (5794), 1752–1756. doi:10.1126/science.1129666
- Kuzmin, Y. I., Da Costa, C. P., and Fedor, M. J. (2004). Role of an Active Site Guanine in Hairpin Ribozyme Catalysis Probed by Exogenous Nucleobase rescue. *J. Mol. Biol.* 340 (2), 233–251. doi:10.1016/j.jmb.2004.04.067
- Laar, C. (1886). Ueber die Hypothese der wechselnden Bindung. *Ber. Dtsch. Chem. Ges.* 19 (1), 730–741. doi:10.1002/cber.188601901165
- Lafontaine, D. A., Wilson, T. J., Norman, D. G., and Lilley, D. M. J. (2001). The A730 Loop Is an Important Component of the Active Site of the VS Ribozyme 1 Edited by J. Karn. *J. Mol. Biol.* 312 (4), 663–674. doi:10.1006/jmbi.2001.4996
- Legault, P., and Pardi, A. (1997). Unusual Dynamics and pKa Shift at the Active Site of a Lead-Dependent Ribozyme. *J. Am. Chem. Soc.* 119 (28), 6621–6628. doi:10.1021/ja9640051
- Li, D., Fedeles, B. I., Singh, V., Peng, C. S., Silvestre, K. J., Simi, A. K., et al. (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 (32), E3252–E3259. doi:10.1073/pnas.1405635111
- Löwdin, P.-O. (1963). Proton Tunneling in DNA and its Biological Implications. *Rev. Mod. Phys.* 35 (3), 724–732. doi:10.1103/revmodphys.35.724
- Manrubia, S. C., Domingo, E., and Lázaro, E. (2010). Pathways to Extinction: beyond the Error Threshold. *Phil. Trans. R. Soc. B* 365 (1548), 1943–1952. doi:10.1098/rstb.2010.0076
- Martick, M., and Scott, W. G. (2006). Tertiary Contacts Distant from the Active Site Prime a Ribozyme for Catalysis. *Cell* 126 (2), 309–320. doi:10.1016/j.cell.2006.06.036
- McKay, D. B. (1996). Structure and Function of the Hammerhead Ribozyme: an Unfinished story. *Rna* 2 (5), 395–403.
- Menéndez-Arias, L. (2021). Decoding Molnupiravir-Induced Mutagenesis in SARS-CoV-2. *J. Biol. Chem.* 297 (1), 100867. doi:10.1016/j.jbc.2021.100867
- Mons, M., Dimicoli, I., Pizzuti, F., Tardivel, B., and 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 (20), 5088–5094. doi:10.1021/jp0139742
- Mulder, F. A. A., Mittermaier, A., Hon, B., Dahlquist, F. W., and Kay, L. E. (2001). Studying Excited States of Proteins by NMR Spectroscopy. *Nat. Struct. Biol.* 8 (11), 932–935. doi:10.1038/nsb1101-932
- Mullins, J. I., Heath, L., Hughes, J. P., Kicha, J., Styrchak, S., Wong, K. G., et al. (2011). Mutation of HIV-1 Genomes in a Clinical Population Treated with the Mutagenic Nucleoside KP1461. *PLoS one* 6 (1), e15135. doi:10.1371/journal.pone.0015135
- Nikolova, E. N., Kim, E., Wise, A. A., O'Brien, P. J., Andricioaei, I., and Al-Hashimi, H. M. (2011). Transient Hoogsteen Base Pairs in Canonical Duplex DNA. *Nature* 470 (7335), 498–502. doi:10.1038/nature09775
- Peng, C. S., Baiz, C. R., and Tokmakoff, A. (2013). Direct Observation of Ground-State Lactam-Lactim Tautomerization Using Temperature-Jump Transient 2D IR Spectroscopy. *Proc. Natl. Acad. Sci. USA* 110 (23), 9243–9248. doi:10.1073/pnas.1303235110
- Peng, C. S., Fedeles, B. I., Singh, V., Li, D., Amariuta, T., Essigmann, J. M., et al. (2015). Two-dimensional IR Spectroscopy of the Anti-HIV Agent KP1212 Reveals Protonated and Neutral Tautomers that Influence pH-dependent Mutagenicity. *Proc. Natl. Acad. Sci. USA* 112 (11), 3229–3234. doi:10.1073/pnas.1415974112
- Peng, C. S., Jones, K. C., and Tokmakoff, A. (2011). Anharmonic Vibrational Modes of Nucleic Acid Bases Revealed by 2D IR Spectroscopy. *J. Am. Chem. Soc.* 133 (39), 15650–15660. doi:10.1021/ja205636h
- Peng, C. S., and Tokmakoff, A. (2012). Identification of Lactam-Lactim Tautomers of Aromatic Heterocycles in Aqueous Solution Using 2D IR Spectroscopy. *J. Phys. Chem. Lett.* 3 (22), 3302–3306. doi:10.1021/jz301706a
- Pinard, R., Hampel, K. J., and Heckman, J. E. (2001). Functional Involvement of G8 in the Hairpin Ribozyme Cleavage Mechanism. *Embo j* 20 (22), 6434–6442. doi:10.1093/emboj/20.22.6434
- Rangadurai, A., Szymanski, E. S., Kimsey, I., Shi, H., and Al-Hashimi, H. M. (2020). Probing Conformational Transitions towards Mutagenic Watson-crick-like G-T Mismatches Using Off-Resonance Sugar Carbon R1ρ Relaxation Dispersion. *J. Biomol. NMR* 74 (8-9), 457–471. doi:10.1007/s10858-020-00337-7
- Rangadurai, A., Zhou, H., Merriman, D. K., Meiser, N., Liu, B., Shi, H., et al. (2018). Why Are Hoogsteen Base Pairs Energetically Disfavored in A-RNA Compared to B-DNA. *Nucleic Acids Res.* 46 (20), 11099–11114. doi:10.1093/nar/gky885
- Rangadurai, A., Szymanski, E. S., Kimsey, I. J., Shi, H., and Al-Hashimi, H. M. (2019). Characterizing Micro-to-millisecond Chemical Exchange in Nucleic Acids Using Off-Resonance R1ρ Relaxation Dispersion. *Prog. Nucl. Magn. Reson. Spectrosc.* 112–113, 55–102. doi:10.1016/j.pnmrs.2019.05.002
- Rupert, P. B., and Ferré-D'Amaré, A. R. (2001). Crystal Structure of a Hairpin Ribozyme-Inhibitor Complex with Implications for Catalysis. *Nature* 410 (6830), 780–786. doi:10.1038/35071009
- Sevilla, M. D., Besler, B., and Colson, A.-O. (1995). Ab Initio molecular Orbital Calculations of DNA Radical Ions. 5. Scaling of Calculated Electron Affinities and Ionization Potentials to Experimental Values. *J. Phys. Chem.* 99 (3), 1060–1063. doi:10.1021/j100003a032
- Shannon, A., Selisko, B., Le, N.-T., Huchting, J., Touret, F., Piorkowski, G., et al. (2020). Rapid Incorporation of Favipiravir by the Fast and Permissive Viral RNA Polymerase Complex Results in SARS-CoV-2 Lethal Mutagenesis. *Nat. Commun.* 11 (1), 4682. doi:10.1038/s41467-020-18463-z
- Singh, V., Fedeles, B. I., and Essigmann, J. M. (2015). Role of Tautomerism in RNA Biochemistry. *Rna* 21 (1), 1–13. doi:10.1261/rna.048371.114
- Singh, V., Peng, C. S., Li, D., Mitra, K., Silvestre, K. J., Tokmakoff, A., et al. (2014). Direct Observation of Multiple Tautomers of Oxythiamine and Their Recognition by the Thiamine Pyrophosphate Riboswitch. *ACS Chem. Biol.* 9 (1), 227–236. doi:10.1021/cb400581f
- Slocombe, L., Al-Khalili, J. S., and Sacchi, M. (2021). Quantum And Classical Effects In DNA Point Mutations: Watson-Crick Tautomerism In AT And GC Base Pairs. *Phy. Chemi.* 23 (7), 4141–4150. doi:10.1039/D0CP05781A
- Sood, V. D., and Collins, R. A. (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 (3), 443–454. doi:10.1016/s0022-2836(02)00521-1
- Suslov, N. B., DasGupta, S., Huang, H., Fuller, J. R., Lilley, D. M. J., Rice, P. A., et al. (2015). Crystal Structure of the Varkud Satellite Ribozyme. *Nat. Chem. Biol.* 11 (11), 840–846. doi:10.1038/nchembio.1929
- Szymanski, E. S., Kimsey, I. J., and Al-Hashimi, H. M. (2017). Direct NMR Evidence that Transient Tautomeric and Anionic States in dG-dT Form

- Watson-crick-like Base Pairs. *J. Am. Chem. Soc.* 139 (12), 4326–4329. doi:10.1021/jacs.7b01156
- Thomas, J. M., and Perrin, D. M. (2008). Probing General Base Catalysis in the Hammerhead Ribozyme. *J. Am. Chem. Soc.* 130 (46), 15467–15475. doi:10.1021/ja804496z
- Thore, S., Frick, C., and Ban, N. (2008). Structural Basis of Thiamine Pyrophosphate Analogues Binding to the Eukaryotic Riboswitch. *J. Am. Chem. Soc.* 130 (26), 8116–8117. doi:10.1021/ja801708e
- Thore, S., Leibundgut, M., and Ban, N. (2006). Structure of the Eukaryotic Thiamine Pyrophosphate Riboswitch with its Regulatory Ligand. *Science* 312 (5777), 1208–1211. doi:10.1126/science.1128451
- Topal, M. D., and Fresco, J. R. (1976). Complementary Base Pairing and the Origin of Substitution Mutations. *Nature* 263 (5575), 285–289. doi:10.1038/263285a0
- Veliky, I., Acharya, S., Trifonova, A., Földesi, A., and Chattopadhyaya, J. (2001). The pKa's of 2'-Hydroxyl Group in Nucleosides and Nucleotides. *J. Am. Chem. Soc.* 123 (12), 2893–2894. doi:10.1021/ja0036312
- Wang, W., Hellinga, H. W., and Beese, L. S. (2011). Structural Evidence for the Rare Tautomer Hypothesis of Spontaneous Mutagenesis. *Proc. Natl. Acad. Sci.* 108 (43), 17644–17648. doi:10.1073/pnas.1114496108
- Watson, J. D., and Crick, F. H. C. (1953). Genetical Implications of the Structure of Deoxyribonucleic Acid. *Nature* 171 (4361), 964–967. doi:10.1038/171964b0
- Wilcox, J. L., and Bevilacqua, P. C. (2013). A Simple Fluorescence Method for pKa Determination in RNA and DNA Reveals Highly Shifted pKa's. *J. Am. Chem. Soc.* 135 (20), 7390–7393. doi:10.1021/ja3125299
- Zhou, H., Kimsey, I. J., Nikolova, E. N., Sathyamoorthy, B., Grazioli, G., McSally, J., et al. (2016). m1A and m1G Disrupt A-RNA Structure through the Intrinsic Instability of Hoogsteen Base Pairs. *Nat. Struct. Mol. Biol.* 23 (9), 803–810. doi:10.1038/nsmb.3270
- Zhou, S., Hill, C. S., Sarkar, S., Tse, L. V., Woodburn, B. M. D., Schinazi, R. F., et al. (2021).  $\beta$ -DN 4-hydroxycytidine (NHC) Inhibits SARS-CoV-2 through Lethal Mutagenesis but Is Also Mutagenic to Mammalian Cells. *J. Infect. Dis.* 224 (3), 415–419. doi:10.1093/infdis/jiab247

**Conflict of Interest:** Declaration of interest DL is an author on the patent US9283242B2 (assigned to Massachusetts Institute of Technology), BF and VS are authors on the patent US9714265B2 (assigned to Massachusetts Institute of Technology). These patents describe two classes of mutagenic nucleoside analogs that can adopt multiple tautomeric forms and their potential uses as anti-viral therapeutics.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Fedeles, Li and Singh. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.