Structural landscape of base pairs containing post-transcriptional modifications in RNA

PREETHI P. SEELAM,[1] PURSHOTAM SHARMA,[2] and ABHIJIT MITRA[1]

[1]Center for Computational Natural Sciences and Bioinformatics, International Institute of Information Technology Hyderabad (IIIT-H), Gachibowli, Hyderabad, Telangana 500032, India
[2]Computational Biochemistry Laboratory, Department of Chemistry and Centre for Advanced Studies in Chemistry, Panjab University, Chandigarh 160014, India

ABSTRACT

Base pairs involving post-transcriptionally modified nucleobases are believed to play important roles in a wide variety of functional RNAs. Here we present our attempts toward understanding the structural and functional role of naturally occurring modified base pairs using a combination of X-ray crystal structure database analysis, sequence analysis, and advanced quantum chemical methods. Our bioinformatics analysis reveals that despite their presence in all major secondary structural elements, modified base pairs are most prevalent in tRNA crystal structures and most commonly involve guanine or uridine modifications. Further, analysis of tRNA sequences reveals additional examples of modified base pairs at structurally conserved tRNA regions and highlights the conservation patterns of these base pairs in three domains of life. Comparison of structures and binding energies of modified base pairs with their unmodified counterparts, using quantum chemical methods, allowed us to classify the base modifications in terms of the nature of their electronic structure effects on base-pairing. Analysis of specific structural contexts of modified base pairs in RNA crystal structures revealed several interesting scenarios, including those at the tRNA:rRNA interface, antibiotic-binding sites on the ribosome, and the three-way junctions within tRNA. These scenarios, when analyzed in the context of available experimental data, allowed us to correlate the occurrence and strength of modified base pairs with their specific functional roles. Overall, our study highlights the structural importance of modified base pairs in RNA and points toward the need for greater appreciation of the role of modified bases and their interactions, in the context of many biological processes involving RNA.

Keywords: modified base pairs; X-ray crystal structures; interaction energies; base-pair parameters; post-transcriptional modifications

INTRODUCTION

Recent structural and mechanistic studies on RNA molecules illustrate that although tremendous progress has been achieved toward understanding their versatile role in modern biology, there is a need to deepen our understanding of the principles governing the structure, dynamics, and functions of these fascinating biomacromolecules. For example, similar to proteins, where the post-translational modifications are associated with catalysis, initiation, and termination of signal cascades, and integration of information at many metabolic intersections (Walsh et al. 2005), post-transcriptionally modified nucleobases may also be associated with a variety of RNA functionalities. A detailed understanding of chemical modifications of RNA nucleobases, and resulting changes in associated noncovalent interactions, is therefore one of the necessary requirements for investigating the functional diversity of RNA molecules.

Post-transcriptional modifications in RNA range from the addition of simple functional groups (e.g., base/ribose methylation) to that of complex side chains (e.g., hypermodifications) (Denmon et al. 2011). In addition, such modifications may also include substitution (e.g., conversion of uridine to 4-thiouridine, s4U), isomerization (e.g., conversion of uridine to pseudouridine, Ψ), and reduction (e.g., conversion of uridine to dihydrouridine, D) (Fig. 1; Mueller et al. 1998). Available literature suggests that modifications serve as important evolutionary tools for tuning up the RNA structure to perform its biological functions with greater fidelity (Ofengand and Bakin 1997; Ofengand 2002; Emmerechts et al. 2008). Thus, it is not surprising that modifications are

Corresponding authors: psharma@pu.ac.in, abi_chem@iiit.ac.in
Article is online at http://www.rnajournal.org/cgi/doi/10.1261/rna.060749.117.
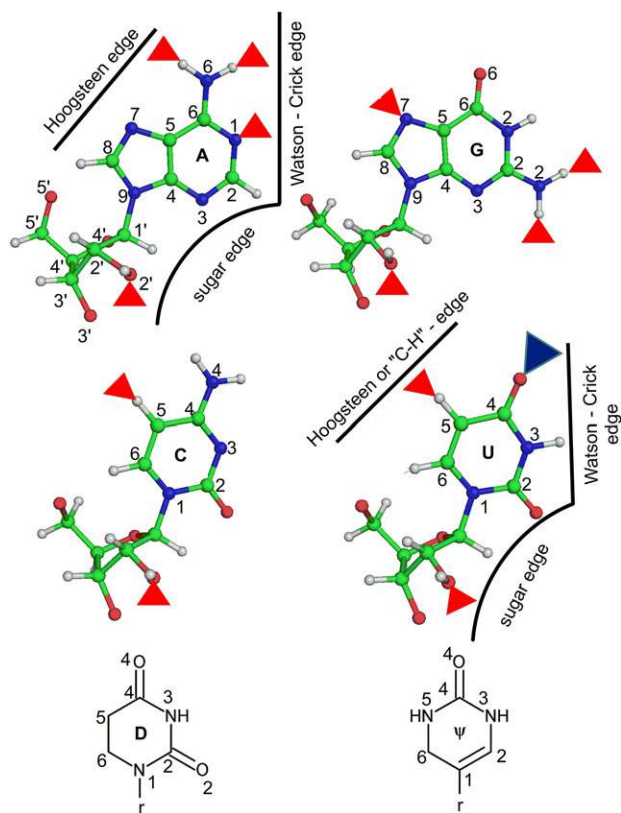
FIGURE 1. Schematic representation of modified base pairs showing their interacting edges. Red triangles represent modification involving methyl group substitution, whereas the blue triangle represents substitution of oxygen with sulfur atom. The ribose sugar is represented by r in the structures of dihydrouridine (D) and pseudouridine (ψ).

present in RNA of organisms belonging to all three (i.e., archaea, bacteria, and eukarya) domains of life (Decatur and Fournier 2002), and the percentage of chemical modifications in an RNA sequence is roughly proportional to the complexity of the organism (Chow et al. 2007). In this context, repositories of modified RNA bases available in databases such as MODOMICS (Dunin-Horkawicz et al. 2006), Transfer RNA database (tRNAdb, Jühling et al. 2009), RNA Modification Database (RNAMDB, Cantara et al. 2011), 3D ribosomal modification maps database (3Dmodmap, Piekna-Przybylska et al. 2008), and Small Subunit rRNA Modification Database (SSUmods, McCloskey and Rozenski 2005) provide a comprehensive listing of post-transcriptionally modified nucleosides in RNA, which are useful in understanding RNA nucleoside modification pathways.

In terms of mechanistic understanding, one of the ways through which modified nucleobases provide stability to the RNA 3D structures is by inducing tailor-made alterations to the conformational preferences of the nucleotides. For example, methylation at the 2′-OH group of ribose shifts the equilibrium toward C3′-endo sugar pucker, thus favoring the A-form RNA helices (Motorin and Helm 2010). Further, contrary to the *anti*-conformation adopted by natu-

rally occurring nucleotides, pseudouridine prefers the *syn* conformation at the glycosidic bond. Given the low-energy requirement for the *anti/syn* transition, pseudouridine can shift between the two conformations with relatively greater ease and can function as a conformational switch in RNA (Charette and Gray 2000). Dihydrouridine, on the other hand, significantly destabilizes the C3′-*endo* sugar conformation, which is associated with base stacked, ordered, A-type helical RNA. Thus, it is not surprising that dihydrouridine is found in higher percentages in RNA of organisms that grow in low temperatures, where it provides extra flexibility (Dalluge et al. 1996).

Apart from changing the conformational preferences, modifications can also affect the noncovalent interactions in RNA (Davis 1995). Base-pairing and base stacking constitute the major noncovalent interactions through which RNA nucleotides interact with each other. Although base stacking helps in RNA folding, it is less specific and weaker compared to base-pairing (Leontis and Westhof 2001). In contrast, base-pairing provides directionality and specificity (Leontis and Westhof 2001; Leontis et al. 2002) and plays a crucial role in scripting the structural variety and functional dynamics of RNA molecules. In this context, the base-pairing classification (Fig. 2) efforts by Leontis and Westhof (2001) (Leontis et al. 2002; Stombaugh et al. 2009) and the quantum chemical revelation of physicochemical principles of RNA base-pairing (Šponer et al. 2005a,b,c, 2010; Sharma et al. 2008, 2010b; Mládek et al. 2009; Chawla et al. 2011; Halder et al. 2014, 2015; Bhattacharya et al. 2015), coupled with increasing availability of X-ray crystal structures of functional RNA, have significantly enhanced our understanding of base-pairing involving canonical nucleosides in RNA. However, the effect of nucleoside modifications on intrinsic properties of RNA base pairs has been considered only in a few quantum chemical or structural studies (Oliva et al. 2006, 2007; Chawla et al. 2015).

Previous studies on tRNA post-transcriptional modifications observed that modifications that introduce positive
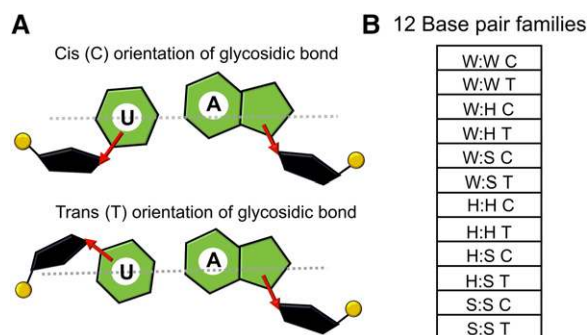

FIGURE 2. (*A*) Schematic representation of *cis* (C) or *trans* (T) orientation of the glycosidic bond. (*B*) List of 12 RNA base-pairing families. W, H, and S represent Watson–Crick, Hoogsteen, and sugar edges, respectively.

charges strongly stabilize the geometry of the corresponding base pairs. An example in context is the stabilization of a reverse Watson–Crick geometry of G15:C48 tertiary interaction in RNA on positively charged archaeosine modification of guanine (Oliva et al. 2007). More recently, analysis of available RNA crystal structures (Chawla et al. 2015) revealed that 11 types of base modifications participate in base pair formation, forming 27 distinct base pair combinations. Further quantum chemical studies revealed that whereas methyl modifications either impart steric clashes or introduce positive charge, other modifications such as Ψ and D affect the stability and flexibility of the structure (Chawla et al. 2015).

Be that as it may, there remains a significant gap in understanding of the structural principles involving modified base pairs in RNA. A number of factors need to be considered in order to address it. First, since the structural diversity of modifications varies across different groups of RNAs (Cantara et al. 2011), the relative abundance of modified base pairs with respect to different RNA classes needs to be considered. Further, due to the prevalence of sugar modifications in RNA, and given the fact that ribose plays an important role in RNA base-pairing (Šponer et al. 2005a; Sharma et al. 2008; Mládek et al. 2009), the effect of ribose modifications on geometries and stabilities of RNA base pairs needs to be analyzed. In addition, the geometric characteristics of crystal occurrences of modified base pairs need to be analyzed in detail, in order to quantify the effect of base modifications on conformational flexibilities of base pairs in crystal contexts. Finally, the structural context of occurrence of modified base pairs in RNA structures needs to be considered, in order to evolve a deeper understanding of the functional roles of such base pairs.

In the present work, we attempt to fill this void in literature by probing into the geometrical features and the intrinsic sta-bilities of base pairs containing modified RNA bases, in terms of their molecular-level interactions, as well as their macromolecular context of occurrence in RNA structures. For this, we have chosen a multipronged approach that uses a combination of crystal structure database analysis using tools of structural bioinformatics, sequence analysis and state-of-the-art quantum chemical methods. Overall, our study provides a comprehensive analysis of modified base pairs in RNA, which may inspire future studies on the specific functional context of individual base modifications in RNA.

RESULTS AND DISCUSSION

Sequence–structure–energetic context of occurrences of modified base pairs in functional RNAs

Statistical overview of modified base pairs in RNA 3D structures

(i) tRNA structures show a remarkably high occurrence of modified base pairs: Fifteen different types of naturally occurring modified RNA nucleosides were searched to analyze their propensities to form base pairs (Table 1). Eleven of them involved modification(s) of the nucleobase moiety, and were previously found to participate in base-pairing in RNA structures (Chawla et al. 2015). Additionally, since methylation of the 2′-OH group is also known to affect base-pairing involving sugar edge interactions (Leontis and Westhof 2001), this modification was also considered for all four nucleosides (A, C, G, and U). A set of 207 high-resolution RNA crystal structures, containing at least one modified base (Supplemental Table S1), and determined according to specific search criteria (see Materials and Methods), was selected for analysis. More than half of these modified bases participate in base-pairing, whereas the unpaired modified bases were present

TABLE 1. Naturally occurring modified bases that participate in base-pairing in RNA crystal structures

Modified bases[a]	Observed base pairs
N1-methyl adenine (m^1A)	m^1A:A W:HT, m^1A:m^5U W:HT, m^1A:m^5Um W:HT
N6,6-dimethyl adenine (m^6_2A)	m^6_2A:G S:ST
2′-O-methyl adenine (Am)	Am:G S:ST
N2-methyl guanine (m^2G)	m^2G:C W:WC, m^2G: C W:W(+)C, m^2G:U W:WC
N2,2-dimethyl guanine (m^2_2G)	m^2_2G:A W:WC
N7-methyl guanine (m^7G)	m^7G:G W:HT, m^7G:C W:WC, m^7G:A S:ST, m^7G:A S:WT
2′-O-methyl guanine (Gm)	Gm:A W:WC, Gm:G S:SC, Gm:C W:WC,
C5-methyl cytosine (m^5C)	m^5C:G W:WC, m^5C:G W:WT
2′-O-methyl cytosine (Cm)	Cm:Gm W:WC
Pseudouridine (Ψ)	Ψ:A W:WC, Ψ:A W:HT, Ψ:G W:WC, Ψ:C S:WC
5,6-dihidrouridine (D)	D:G W:ST, D:G H:SC, D:U W:WT, D:U S:ST
C5-methyl uracil (m^5U)	m^5U:A W:HT, m^5U:G W:WC, m^5U:G W:HT
2′,5-dimethyl uracil (m^5Um)	m^1A: m^5Um H:WT
2′-O-methyl uracil (Um)	Um:A W:WC, Um:A W:HT, Um:G H:SC
4-thiouridine (s^4U)	s^4U:A W:HT, s^4U:A W:ST, s^4U:A S:SC

[a]Methylated bases are represented as m^XN, where "m" represents methyl group, the superscript "X" represents the position of the methyl group, and "N" represents the nucleobase. m^X_YN is used to represent more than one methyl substituent, where the subscript "Y" represents the number of methyl substituents. 2′-O-ribose methylation is represented as Nm.

in other variable structural contexts (Supplemental Tables S2–S4). While 80% of the crystal structures belong to four major RNA classes (tRNA [25%], 16S rRNA [24%], 23S rRNA [23%], and RNA-binding proteins [11%], Fig. 3A), the occurrence of modified bases in the rest of the RNA classes was rather marginal (each below 10%).

Base pairs involving at least one modified base were detected in 65% (135) of the total (207) crystal structures (Supplemental Table S2). Approximately one-third of such structures belongs to tRNA (36%), another one-third to 23S rRNA (34%), and one-sixth belongs to 16S rRNA (18%, Fig. 3). Overall, a total of 453 modified base pairs were detected, half of which belong to tRNA (Supplemental Fig. S1). This agrees with earlier reports regarding relatively greater occurrence of modified bases in tRNA compared to that in other RNA classes (Limbach et al. 1994; Machnicka et al. 2014). Given that most of the nucleobases in tRNA are involved either in base-pairing or in tertiary interactions (Oliva et al. 2006), it is not surprising that most of the modified bases present in tRNA also participate in base-pairing.

(ii) Modified base pairs are observed in all major RNA structural elements and span diverse RNA base-pairing geometries: Distribution of modified base pairs with respect to their contextual occurrence in RNA crystal structures reveals that approximately half (49%) of them are present in stem (helical) regions, 14% in loop regions, and the rest (37%) are involved in tertiary interactions (Supplemental Fig. S1). Overall, the results are in line with a previous crystal structure analysis (Chawla et al. 2015), where 41% of the total modified base pairs were found to be involved in tertiary interactions (Supplemental Tables S5, S6).

Categorization of the modified base pairs, in terms of the portion of the nucleoside that interacts with the partner nucleoside, reveals that 80% of them involve base–base (B–B) interactions, 12% involve base–nucleoside (B–S) interactions, and 8% involve nucleoside–nucleoside (S–S) interactions (Supplemental Tables S7–S9). Further categorization of base pairs in terms of the interacting edge (Watson–Crick [W], Hoogsteen [H] or Sugar [S]) and the glycosidic bond (*cis* [C] or *trans* [T], Figs. 1, 2) orientation reveals

that the B–B interactions involving modified bases (80%) span four of the six associated base-pairing families—W:WC (49%), W:HT (27%), W:WT (3%), and W:HC (1%). Notably, no examples of modified base pairs are observed among H:HC and H:HT families of base pairs, possibly since, due to their unique backbone topology requirements, these base pair families are themselves known to occur rarely in RNA structures (Sharma et al. 2010a). On the other hand, B–S (12%) and S–S (8%) interactions span all six possible base pair geometries (W:SC [5%], W:ST [4%], H:ST [2%], H:SC [1%], S:SC [6%], and S:ST [2%]). Overall, the total 453 base pairs identified in RNA crystal structures belong to 36 unique base-pairing combinations, 24 of which involve B–B interactions, six involve B–S interactions, and six involve S–S interactions (Table 2).

(iii) Methylation is the preferred chemical modification in RNA base pairs: Distribution of modified base pairs with respect to the type of modification reveals that more than half of them contain at least one methylated base (60% total, 35% in tRNA, and 22% in 16S rRNA). Further, depending on the number of potential methylation sites available in the parent nucleobase, substantial diversity is observed in methylated base pairs. For example, one-third of the total methylated base pairs contain m^5C, which is followed by three varieties of methylated G (26%), namely m^7G (13%), m^2G (9%), and m^2_2G (4%, Supplemental Fig. S1). However, in contrast to the abundance of base pairs containing methyl modification at the nucleobase moiety (86%), only 14% contain sugar methylation at 2′-OH (Supplemental Fig. S1). Greater abundance of methylated base pairs in RNA structures can be correlated to the wide variety of structural roles played by these bases that include enhancement of base stacking, enhanced nucleobase polarizability, tendency to favor C3′ endo-conformation, block sugar–edge interactions and enhancement of stability against hydrolysis by methylation at 2′-OH of sugar (Helm 2006).

(iv) Base pairs containing modified uridine or guanosine are relatively more abundant: Crystal structure analysis reveals that 72% of the modified base pairs contained either modified uridine (37%) or modified guanosine (35%, Supplemental Fig. S1). The greater proportion of base pairs containing uridine modifications can be attributed to the natural occurrence of a rich variety in uridine modifications (base methylation and/or sugar methylation, thiolation, pseudouridylation, or reduction), each of which has the propensity to form base pairs. In fact, six of the 15 modified nucleosides that form base pairs (Table 1) contain modification of uridine. On the other hand, greater abundance of base pairs containing guanosine modifications can be correlated to occurrence of a variety of methylation sites at guanosines (e.g.,
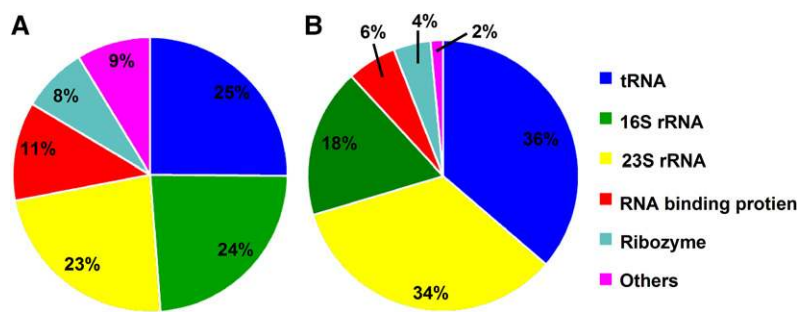

FIGURE 3. (*A*) Percent distribution of total 207 crystal structures in the data set as a function of RNA type. (*B*) Percent distribution of those 135 crystal structures as a function of RNA type that contain at least one modified base.

TABLE 2. Occurrence frequency and the type of RNA in which the 36 unique modified base pairs were identified in the data set

Base pair	Type[a]	F[b]	RNA type
m^5C:G W:WC	B–B	87	tRNA, 16S rRNA
m^2G:C W:WC	B–B	38	tRNA, 16S rRNA
Ψ:A W:WC	B–B	24	tRNA, mRNA:tRNA
m^7G:C W:WC	B–B	22	16S rRNA
m^2_2G:A W:WC	B–B	16	tRNA
Gm:C W:WC	B–B	12	23S rRNA
Ψ:G W:WC	B–B	10	tRNA
Um:A W:WC	B–B	4	23S rRNA
m^5U:G W:WC	B–B	3	snRNA
m^2G: C W:W(+)C	B–B	3	tRNA
m^2G:U W:WC	B–B	1	tRNA
Gm:A W:WC	B–B	1	23S rRNA
Cm:Gm W:WC	B–B	1	23S rRNA
m^5U:A W:HT	B–B	34	tRNA
m^7G:G W:HT	B–B	27	tRNA
Ψ:A W:HT	B–B	22	16S rRNA
m^1A: m^5U H:WT	B–B	18	tRNA
s^4U:A W:HT	B–B	15	tRNA
m^5U:G W:HT	B–B	2	tRNA
m^1A:A H:WT	B–B	1	tRNA
m^1A:m^5Um H:WT	B–B	1	tRNA
Um:A W:HT	B–B	1	23S rRNA
D:U W:WT	B–B	12	tRNA
m^5C:G W:WT	B–B	3	tRNA
Ψ:C S:WC	B–S	22	16S rRNA
m^7G:A S:WT	B–S	10	16S rRNA
s^4U:A W:ST	B–S	7	tRNA
D:G W:ST	B–S	1	tRNA
D:G H:SC	B–S	1	tRNA
Um:G H:SC	B–S	1	23S rRNA
Gm:G S:SC	S–S	30	23S rRNA
s^4U:A S:SC	S–S	1	tRNA
m^6_2A:G S:ST	S–S	3	23S rRNA:tRNA
Am:G S:ST	S–S	2	Ribozyme
m^7G:A S:ST	S–S	1	16S rRNA
D:U S:ST	S–S	1	tRNA

[a]B–B, B–S, and S–S stand for base–base, base–sugar, and sugar–sugar interactions, respectively.
[b]Occurrence frequency.

N1, N2, N7, or 2′-OH), as well as the propensity of guanosine to form singly and doubly methylated structures at N2, all of which participate in base-pairing.

Statistical analysis of modified base pairs in the tRNA sequence database

As mentioned above, the highest fraction of modified base pairs are observed in tRNA crystal structures. Owing to the availability of a greater number of tRNA sequences compared to 3D structures, sequence analysis can provide more detailed information about conservation patterns of modified base pairs. However, the usual method of sequence annotation in the sequence database available at the National Center for Biotechnology Information (Pruitt et al. 2007) does not include information on the presence of modified bases

in nucleic acid sequences. This excludes the possibility of use of sequence alignment algorithms, such as BLAST (Altschul et al. 1990), etc., for analysis of modified base pairs in the sequence space.

To overcome this difficulty, we have used the tRNA sequence database (Jühling et al. 2009), which is a repository of sequences that provides information on the presence of modified bases at different tRNA positions (see Materials and Methods). Within these sequences, base pair combinations present at 10 different positions in tRNA structures were searched and graded according to their occurrence frequency in all the sequences (Fig. 4; Supplemental Table S10). Analysis revealed additional examples of modified base pairs at these 10 selected positions. For example, although the m^2_2G:A combination in W:WC geometry, observed most frequently at the 26:44 position in tRNA crystal structures, is also observed most frequently within the tRNA sequences, our sequence analysis reveals three new examples of modified base pair combinations (m^2_2G:U, m^2_2G:Um, and m^2G:A) at this position. Similarly, at the position 54:58 of TΨC-loop in tRNA, although the m^5U:A W:HT is the most frequently observed combination, and covaries with m^5U:m^1A, m^5Um:m^1A, and A:m^1A pairs in tRNA crystal structures, sequence analysis reveals five new examples of modified base pair combinations (m^1Ψ:A, U:m^1A, Ψ:m^1A, Ψ:A, and m^5Um:A) at this position. On similar lines, tRNA sequence analysis identified additional modified base pair combinations at other positions (Fig. 4).

Our analysis further reveals that occurrence of some modified base pairs are restricted to certain domains of life, and are completely absent in others. For example, although modified base pairs m^2G:C, m^2_2G:A, and m^5C:G at positions 10:25, 26:44, and 49:65, respectively, are observed in tRNA sequences of archaea and eukarya, they are absent in bacteria (Fig. 4). Similarly, the m^5U:m^1A base pair present at the 54:58 position and the Ψ:G base pair at the 13:22 position were found only in eukaryotic tRNA, and were absent in lower domains (bacteria and archaea). While these examples suggest the absence of some modified base pair combinations in lower organisms, certain modified base pairs are only observed in lower organisms, and have not reached higher domains of life. For example, s^4U:A at position 8:14 is observed in bacteria and archaea, but not in eukarya. Nevertheless, four modified base pairs G:m^7G, A:m^5U, G:Ψ, and A:Ψ, observed at tRNA positions 22:46, 54:58, 30:40, and 31:39, respectively, are present in all three domains of life. Overall, our results point toward phylogeny-dependent distribution of modified base pairs in tRNA, which may stem from domain-specific strategies of RNA maturation (Machnicka et al. 2014).

Geometric and energetic characterization of modified base pairs

Of the 36 unique modified base pair combinations identified, 23 are present in multiple instances in RNA crystal structures
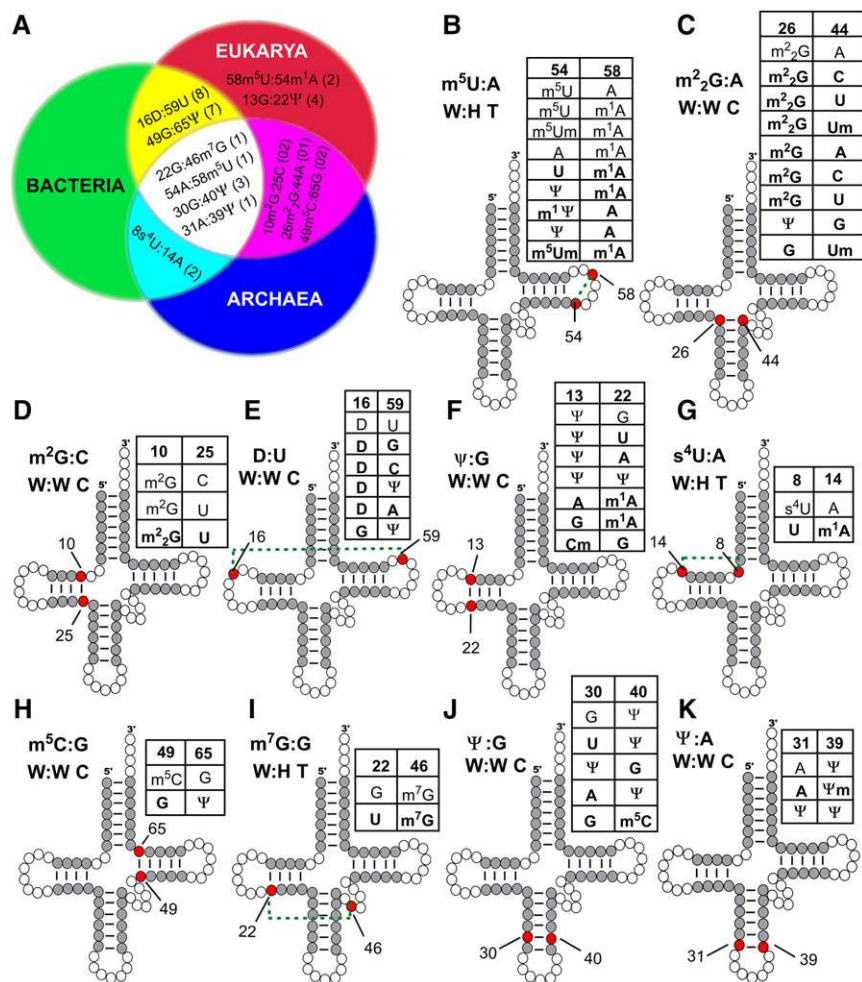
FIGURE 4. Schematic representation of most commonly observed modified base pairs in tRNA sequences. (*A*) Distribution of modified base pairs in tRNA sequences divided according to the domains of life. (*B–K*) Presence of modified base pairs in 10 major base pair positions (represented by red circles) in tRNA structures. The newly identified modified base pair combinations observed from sequence analysis are shown in bold in the corresponding tables.

(Table 2; Supplemental Table S11), where geometrical variations were observed within different occurrences of each base pair. Such variations arise due to both, difference in macromolecular context as well as on the identity of the base pairs. We used eight different structural parameters, viz., root mean square deviation (rmsd), buckle (κ), propeller twist (π), open angle (σ), stagger (s_x), shear (s_y), stretch (s_z) (Mukherjee et al. 2006), and E-value (Das et al. 2006) to quantify the variation in geometries and hydrogen bonding characteristics of modified base pairs in their crystal contexts (Supplemental Section S1; Supplemental Tables S12–S16; Supplemental Figs. S2–S5). Analysis of average and standard deviation in these parameters reveals that most of the base pairs involving B–B interactions exhibit relatively smaller deviation among crystal occurrences. However, significant variation is observed in base pairs involving B–S and S–S interactions, which can be mainly attributed to the flexibility of ribose sugar and the associated glycosidic torsional freedom.

Geometry optimization of a suitably chosen representative crystal occurrence of each modified base pair, using quantum chemical methods, allowed us to locate the minimum energy structures of isolated base pairs. These optimized isolated base pair structures represent the ideal base geometries that would be obtained in the absence of macromolecular crystal structure effects, and are useful to quantify the role of interbase hydrogen bonding in determining the structure of the pair. Thus, the comparison of geometries of the base pairs observed in its isolated form, with those observed in RNA structural context, can provide useful insights into the interplay of the forces within the crystal environment (Fig. 5). Our results reiterate that the variations in geometrical parameters and E-values, between the crystal and the optimized geometry of each base pair, respectively, depend on the combination of geometry of the base pair, the type of interaction (B–B vs. B–S vs. S–S), and the identity of the interacting bases. For example, the high rmsd$_{av}$ (0.8 Å) of the optimized structure of the Um:A W:WC pair, compared to its crystal occurrences, may be understood in terms of significant relaxation of buckle and propeller parameters on optimization. Similarly, the high rmsd$_{av}$ (1.2 Å) of the optimized structure of m$^2_2$G:A W:WC, from its crystal occurrences, can be explained in terms of optimization of hydrogen bond distances (and consequent large deviation in E-val-

ues) on optimization. Detailed comparison of structural parameters in crystal geometries and energy minimized (optimized) geometries of the base pairs are provided in Supplemental Section S1. It may be noted that due to limited resolution of crystal structures within the data set, many of the observed base pair deformations in crystal geometries may also reflect refinement errors of the crystal structure data. This may, in turn, affect the distribution of the geometrical parameters of crystal occurrences of base pairs. Hence appropriate caution needs to be exerted while drawing inferences, regarding the effects of a macromolecular environment on base-pairing, from the comparison of structural properties of crystal and optimized geometries of base pairs.

The refinement inaccuracy related uncertainties notwithstanding, the comparison of optimized geometries, and corresponding interaction energies, of modified base pairs with their respective unmodified counterparts, can reveal important clues regarding the effect of base modification on the
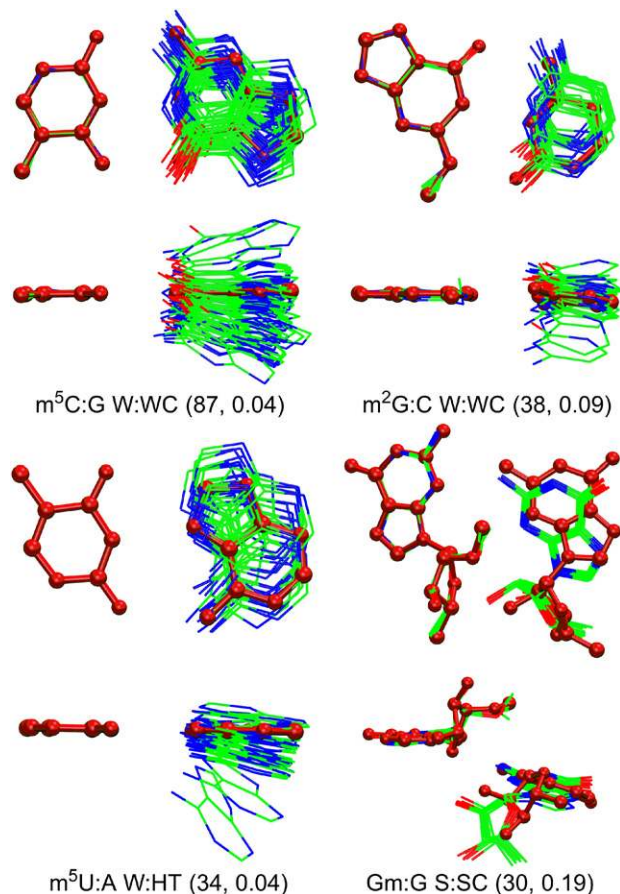
FIGURE 5. Structural alignment of crystal occurrences of modified base pairs (with occurrence frequency ≥30) with their corresponding optimized structures. Occurrence frequency and average RMSD (in Å) with respect to the optimized structure (ball and stick, red) is given in the parentheses.

base pair geometries. This holds true particularly for those related to the structure and strength of hydrogen bonding interactions between the paired bases (Supplemental Table S13). In this context, geometrical deviations and interaction energy differences between each modified base pair and its unmodified counterpart were measured and analyzed (Supplemental Section S1; Supplemental Tables S9–S12; Supplemental Figs. S2–S5). Based on our analysis, the effects of base modification on base-pairing can be divided into two broad categories:

(i) Base pairs where modification induces electronic effects: These include 17 base pairs that involve significant (>2 kcal/mol) change in interaction energy on base modification. Such base pairs can further be grouped into five subcategories:

(a) Base pairs involving alteration of charge on modification: These include seven base pairs, four of which (m^{7}G:G, m^{5}U: m^{1}A, m^{5}Um:m^{1}A, and A:m^{1}A) belong to the W:HT family and one each to W:WC (m^{7}G:C), S:WT (m^{7}G:A), and S:ST (m^{7}G:A) families. All of these acquire positive charge on modification. The resulting enhanced electrostatic compo-

nent of interaction energy significantly increases the overall base-pairing energy (by up to 15 kcal/mol) on modification. Additionally, the alteration of formal charge may affect the pKa values of those titrable groups of nucleobases that are not sequestered in interbase hydrogen bonding. For example, methylation at N7 of G can significantly lower the pKa value of the imino (N1–H) group of the modified G within the m^{7}G:G W:H T pair.

(b) Base pairs involving alteration of hydrogen bonding pattern on modification: These include two base pairs, viz., Ψ:C S:WC and Gm:G S:SC. The former pair disrupts one of the interbase hydrogen bondings, resulting in a decrease in binding energy by 6.3 kcal/mol, whereas the latter alters the H-bonding interactions without affecting the stability of the base pair.

(c) Base pairs involving change in position of electronegative atoms on modification: These include three base pairs involving Ψ. Since Ψ differs from U in terms of the direction of glycosidic bond (e.g., a *trans* base pair involving U will be equivalent to a corresponding *cis* pair involving Ψ), replacement of U with Ψ changes the location of glycosidic nitrogen with respect to the partner base. This results in a change in binding energy of the base pair (Supplemental Table S14).

(d) Base pairs involving replacement of the highly electronegative element (O) with a less electronegative element (S) present on the interacting edge: This category includes the s^{4}U:A W:HT base pair, where the O4 atom present on the WC edge is replaced by the S atom. Since S4 or O4 is not involved in interbase hydrogen bonding in the modified or the unmodified base pair, the interaction energy of s^{4}U:A W:HT is similar (0.4 kcal/mol) to the unmodified pair.

(e) Base pairs involving change in aromaticity of the nucleobase ring on modification: These include four dihydrouridine-containing base pairs, viz., D:U W:WT, D:G W:ST, D:G H: SC, and D:U S:ST, where interaction energy changes up to 2.8 kcal/mol. Further, since the loss of aromaticity increases the pucker of the six-membered nucleobase ring, dihydrouridine-containing base pairs adopt different geometries compared to their unmodified counterparts.

(ii) Base pairs where modification may result in alteration of the surrounding steric environment: This category includes 19 base pairs that involve negligible (<2 kcal/mol) change in interaction energy on base modification, indicating that modification does not significantly change the electronic structure of the base pairs. It is of course possible that base modification in such base pairs is important for providing appropriate steric alterations of the local environment. Such alterations may include blocking the hydrogen bonding capability of the base pair with other surrounding nucleosides, or change the conformational space available to other ligands/proteins present at the interface. Depending on the site of modification, such base pairs can further be grouped into two classes:

(a) Base pairs involving change in steric environment on the minor groove side: These include 11 base pairs, which involve modification of the amino group of guanine (m^{2}G and

m^2_2G) or the 2′-OH (Am, Cm, Gm, and Um). Such modifications may alter the accessibility of the minor groove of the base pair, resulting in potential disruption of associated RNA motifs.

(b) Base pairs involving change in steric environment on the major groove side: These include eight base pairs that involve m^5C, m^5U, m^6_2A, or s^4U bases, where modification occurs at the major groove side of the base pair. Through introduction of the hydrophobic (methyl) groups on nucleobases, such modifications affect the conformational space available for other molecules such as proteins, ligands, other RNA, and antibiotics (Demirci et al. 2014) to interact with RNA.

Functional roles of modified base pairs

Investigating structure–function correlations for some frequently occurring modified base pairs

The above analysis expands the scope for annotation involving their geometric and energetic features by providing useful insights into the occurrence frequencies of modified base pairs within different RNA classes. Additional analysis of macromolecular structural context of occurrence of modified base pairs, as well as of their associated functional roles, is expected to provide an understanding of "why" base modifications occur in the context of RNA. Based on several clues from experimental structures available in literature, and adequately supported by our own structural analysis, here we attempt to provide structural and energetic explanations on why nature may have invoked the modification of bases in functional RNA, in some of the cases.

(i) Presence of methylated base pairs at the hinge regions of tRNA facilitates molecular flexibility: During tRNA transitions at the ribosome, two regions (i.e., the interface of the D-stem/anticodon stem and the TΨC-stem/acceptor-stem) have been proposed to act as hinges for providing flexibility to tRNA (Frank et al. 2005; Sanbonmatsu 2006). Our analysis reveals a significant occurrence frequency of methylated base pairs at three important base-pairing positions (i.e., 10:25, 26:44, and 49:65) within both these hinge regions of tRNA (Supplemental Section S3). Analysis of base-pairing geometries at these positions of tRNA suggests that methylation may help in preventing unwanted hydrogen bonding interactions of these base pairs with their surrounding bases, without compromising the base pair stability (Supplemental Fig. S7). This may, in turn, provide conformational flexibility to hinge regions of tRNA, and points toward the potential role, of the methyl substituents of the nucleobases, in providing flexibility to the tRNA structures, and in facilitating its dynamics.

(ii) Putative role of sugar methylation in tRNA accommodation on the ribosome platform: The crystal structure of the tRNA:rRNA complex of *H. marismortui* (Nissen et al. 2000) reveals that during the accommodation of tRNA on the ribosome, the C75 base of the 3′-CCA end of tRNA forms a W:WC pair with Gm2588 of the Gm2588:G2617:C2542

triplet of 23rRNA. The interaction of C75 with this triplet stabilizes the free 3′-CCA end of tRNA during the translation process. Surprisingly, although this triplet–quartet–triplet association is retained in the analogous crystal structure of *E. coli* (Dunkle et al. 2011), in the similar sequence context, the Gm at 2588 is replaced by unmodified G (Fig. 6). Our analysis of the crystal structure of the ribosome of *E. coli* reveals that the interaction of the 2′-OH group of G2588, with the G2617, reduces the planarity and decreases the hydrogen bond strength within the C75:G2588:G2617:C2542 quartet (PDB: 4V9D, Fig. 6). On the other hand, although the presence of Gm at the 2588 position in *H. marismortui* (PDB: 3CME, Fig. 6) alters the H-bonding interactions involving its 2′-OH group with G2617, without affecting the base pair stability, significant optimization of almost all the base pair parameters is observed on sugar methylation (Supplemental Tables S14, S15). Thus, it appears that the ribose methylation of G2588 helps in maintaining the interaction of tRNA with a
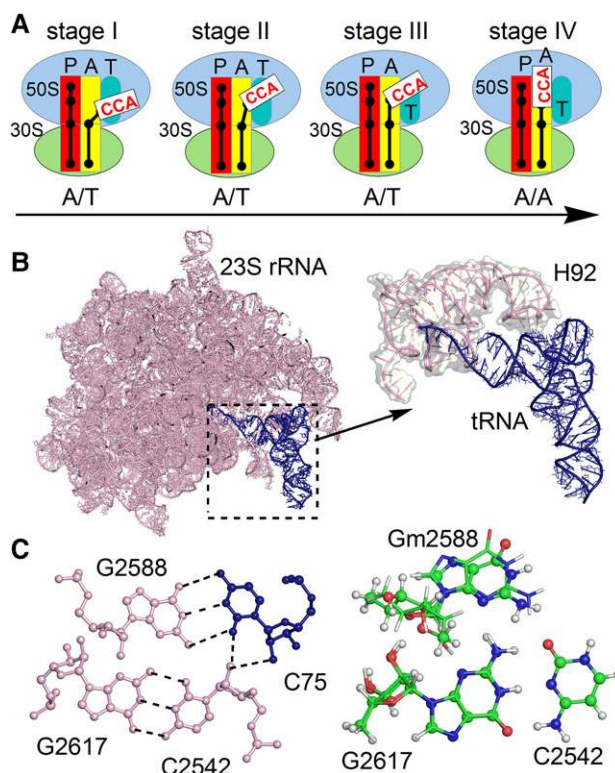

FIGURE 6. (*A*) Flexible 3′-CCA end (white box) of tRNA during various stages of tRNA accommodation at the A-site (yellow box) of the 70S ribosome. The neighboring P-site of rRNA is shown as a red box. (*B*) Interaction of 3′-CCA containing amino acceptor arm of tRNA blue) of tRNA (blue) with the A-loop (H92) of 23S rRNA (pink). (*C*) Structure of base quartet formed from the interaction of the preformed G-minor base triplet (C2542:G2617:Gm2588) present at H92 of rRNA and the C75 of the 3′-CCA of tRNA. Alignment of the preformed rRNA triplet containing the 2′-methylated G2588 present in the crystal structure of the tRNA:rRNA complex of *H. marismortui* (PDB: 3cme), with the corresponding triplet containing the unmodified G2588 present in one of the crystal structures of the tRNA:rRNA complex of *E. coli* (PDB: 4v9d).

large ribosomal subunit and helps in the smooth transition of tRNA from A/T phase to A/A phase. The presence of a modified guanine nucleotide (ribose methylated, Gm2588) in evolutionarily advanced archaea (*H. marismortui*) appears to provide it with a structural advantage over bacteria (*E. coli*) for optimizing the tRNA–rRNA interactions during protein synthesis. This, in turn, underscores the potential role of modified bases, and corresponding base pairs, in facilitating the RNA–RNA interactions in nature.

(iii) Putative role of modified base pairs in T-loop motif of tRNA: The T-loop of tRNA includes three unpaired residues and a single loop closing the 54:58 W:H T base pair at the 54:58 position (Supplemental Fig. S8). Due to backbone flexibility, the 54:58 position can accommodate both purine–pyrimidine as well as purine–purine residues with various degrees of modifications. Our quantum chemical analysis suggests that all five modified base pairs observed at the 54:58 position in tRNA crystal structures possess enhanced interaction energy (up to 7.4 kcal/mol) compared to their unmodified counterparts. This indicates that the presence of modified base pairs within the T-loop may provide additional stabilization to the motif. Further structural analysis reveals that the modified pairs at 54:58 positions may provide additional stabilization to the T-loop by enhancing its associated tertiary interactions with the D loop, which may in turn help in maintaining the functional conformation of tRNA (Supplemental Section S3).

(iv) Possible role of methylated base pairs in the antibiotics binding regions of the bacterial ribosome: Previous experimen-tal studies suggest that N7-methylation at G527 of the G527: C522 W:WC pair of helix 18 in *T. thermophilus* 16S rRNA leads to streptomycin resistance in bacteria (Demirci et al. 2014). Our analysis indicates that N7-methylation imparts a positive charge to G, which enhances the intrinsic stability of the G:C W:WC base pair by ∼9 kcal/mol, while maintaining the geometry similar to that of the canonical G:C base pair. This suggests that methylation at N7 can change the hydrophobic environment in the antibiotic-binding pocket without destabilizing the G527:C522 base pair. The change in the hydrophobic environment may influence the position and orientation of the hydrophobic side chain of the amino acid residues of the S12 protein present in the binding pocket (Fig. 7), which may in turn affect the backbone conformation and size of the binding pocket, thus leading to streptomycin resistance.

It is known from the literature that C5-methylation of C1407 of the C1407:G1494 W:WC pair of 23S rRNA causes resistance toward paramomycin binding to the bacterial ribosome (Demirci et al. 2014). Our analysis indicates C5-methylation does not affect the geometry and the intrinsic stability of the C:G W:WC base pair. Since paromomycin interacts with the major groove of base pair C1407:G1494 W:WC only when C1407 is nonmethylated (Vicens and Westhof 2001), it is possible that the role of methylation lies in providing a steric hindrance for antibiotic binding, without affecting the electronic structure within the binding pocket (Fig. 7). This illustrates the role of the steric effect of modified bases in determining the RNA–ligand interactions.
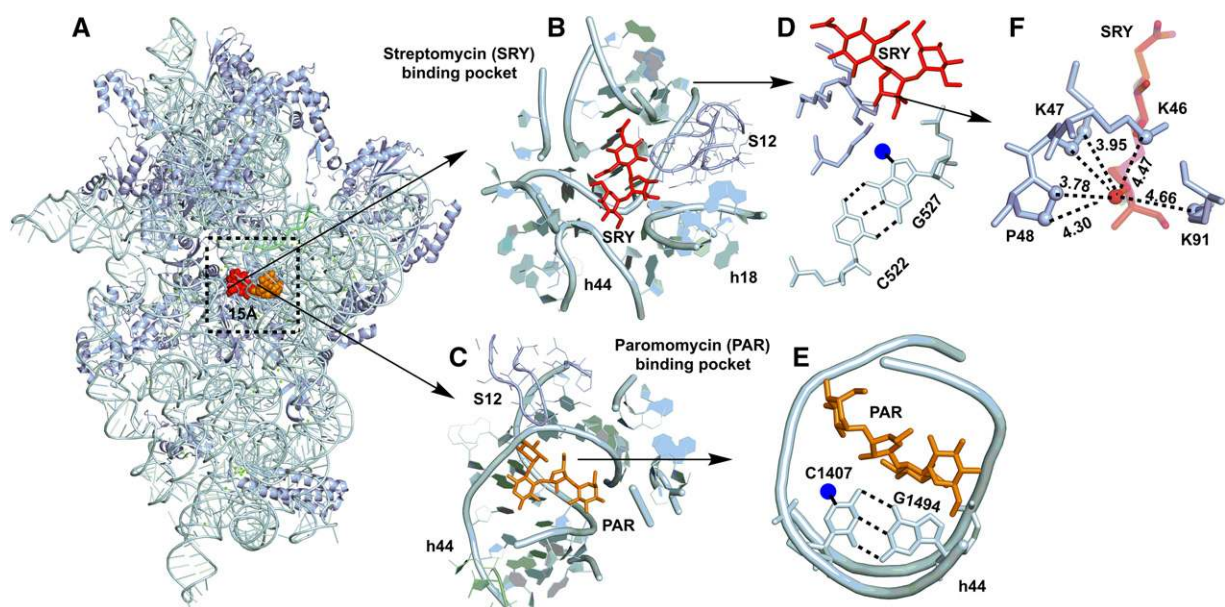

FIGURE 7. Presence of modified base pairs at the binding site of antibiotics streptomycin and paromomycin. (*A*) Structure of 16S rRNA bound to streptomycin (red) and paromomycin (orange). (*B,C*) Antibiotic-binding pocket with surrounding proteins (S12). (*D,E*) Interaction of base pairs C522:527 and C1407:G1494 present in the binding pocket with the antiobiotics streptomycin and paromomycin, respectively. (*F*) Hydrophobic cloud created by surrounding amino acid residues around the methyl group attached to ring II of streptomycin (red). Methyl modification of G527 or C1407 at the nucleobase sites represented by blue circles result in resistance to antibiotic binding.

(v) Potential involvement of modified base pairs in higher order structures and their putative functional roles: Several modified base pairs are involved in formation of higher order structures such as base triples and quadruples. Specifically, we observed 11 distinct triples and two quadruples spanning nine modified bases in our data set (Supplemental Table S17). Such motifs are present at important positions in RNA, including the mRNA:tRNA:rRNA interface, the D loop–V loop interface, the acceptor/D-stem junction of tRNA, the 5′-splice site of group-I intron, and the sarcin-ricin domain in the large ribosomal subunit of *E. coli* (see Supplemental Section S3). Figure 8 shows the geometrical arrangement of some representative higher-order structures that have potential functional roles.

Conclusions

We carried out detailed statistical, geometrical, energetic, and contextual analysis of 36 naturally occurring post-transcriptionally modified base pairs present in RNA macromolecules. Such base pairs span diverse structures and include base–base, base–sugar, and sugar–sugar interactions. Our results reveal that overall, a greater proportion of modified base pairs occurs in tRNA. Further analysis of available tRNA sequences reveals 28 additional examples of modified base pairs, at 10 selected positions in the tRNA sequences that are not observed in reported RNA crystal structures. This adds to the available list of modified base pairs and underscores the importance of sequence analysis in understanding of conservation patterns of RNA motifs.

In general, methylated base pairs are found to be more abundant compared to base pairs containing other modifications in RNA, which can be correlated with the variety of functional roles that methylated bases play in functional RNA, including alteration of base conformation and affecting base stacking. Further, base pairs containing uracil and guanine modifications are more abundant compared to those containing modifications of cytosine or adenine, which can be explained on the basis of the substantial variety in types of uracil modifications and guanine methylation. Detailed analysis of local RNA topology, around the location of modified base pairs, reveals that such base pairs are present in almost all major RNA motifs and points to the diverse structural roles that modified base pairs may play in RNA.

We used advanced quantum chemical methods [MP2/aug-cc-pVDZ//B3LYP/6-31G (d,p)] to analyze the optimal geometries, strengths of interbase interactions, and effects of base modifications on the geometries and interaction energies of RNA base pairs. On the basis of change in strength of interaction on base modification, we classified the effects of base pair modification into steric and electronic perturbations on the unmodified base-pairing geometry. Further, analysis of the surrounding macromolecular environment, as well as of the local RNA structural topology around the modified base pairs, revealed certain important structural and functional contexts. These include contexts involving unique modified base pairs in tRNA, as well as sugar-modified base pairs in rRNA, which suggest that some of them may be playing important roles in maintaining the structure, dynamics, and functions of RNA molecules. Overall, our studies highlight the need for, and provide a comprehensive approach toward, further studies investigating the role of modified bases and their interactions, in the context of many biological processes involving RNA.
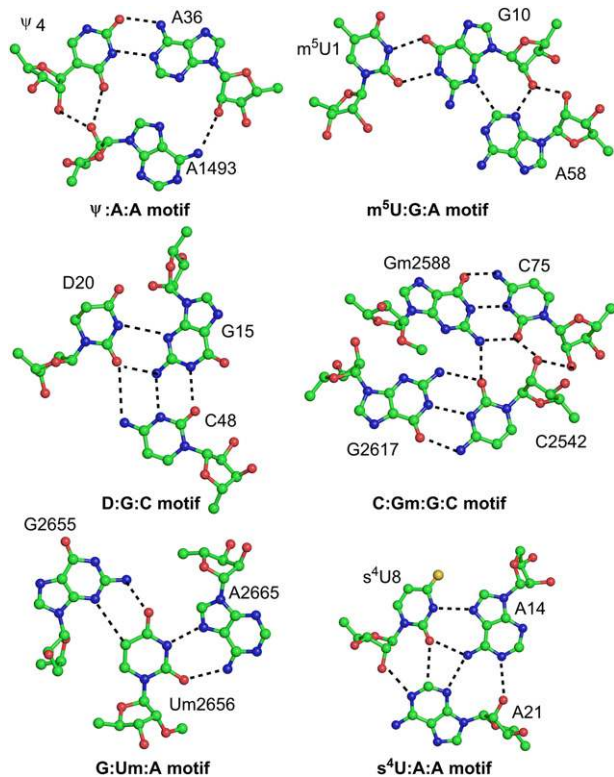
MATERIALS AND METHODS

Data set of RNA crystal structures

To identify base pairs containing modified bases, the occurrence of modified bases was first searched in RNA crystal structures. For this, the PDB*sum* (Laskowski 2009) database, which summarizes information on X-ray crystal structures deposited in the protein databank (PDB), was used. Specifically, using the "Het Groups" option of PDB*sum*, a unique three-letter code corresponding to each of the 15 modified residues (Table 1) was used to retrieve the relevant list of PDB entries submitted until July 18, 2016. The retrieved crystal structures were further filtered according to their resolution, and in synchrony with previous crystal structure study, structures with resolution better than 3.5 Å were selected for further analysis. The data set is intentionally kept redundant with respect to sequence, since the previous study has shown that possible modified base

FIGURE 8. Modified base pairs involved in higher-order interaction motifs.

pair types and base conformations may differ within crystal structures of the same RNA (Chawla et al. 2015). BPFind software (Das et al. 2006) was used to analyze the occurrence, location, and type of modified base pairs with at least two hydrogen bonds in the selected RNA crystal structures.

tRNA sequence analysis

We analyzed all the 474 tRNA cytoplasmic sequences belonging to 73 organisms (i.e., prokaryotes [19], archaea [9], eukaryotes [41], and viruses [4]) from the tRNAdb database (Jühling et al. 2009). For each sequence, we recorded which bases are present at positions where modified base pairs occur in the analyzed crystal structures of tRNA. Thus, at each of the positions, the relative occurrence frequency of the modified base pair was recorded, and ranked within all available combinations. Once the tRNA sequences that contained modified base pairs at specific positions were identified, the sequences that contained the modified pair were further classified according to the type of corresponding aminoacyl tRNA.

Quantum mechanical energy minimization and interaction energies

Among 36 distinct base pair combinations studied, 24 combinations contained base pairs with only base–base interactions, six combinations contained base–nucleoside interactions, and six base pairs contained sugar–sugar interactions. For geometry optimization (energy minimization) of the base pairs that do not involve interaction of ribose sugar with the pairing base, the C1′ atoms of both the participating nucleosides were replaced with hydrogen atoms. For the base pairs involving base–nucleoside interactions, depending on whether one or both the sugars are involved in base-pairing, the respective ribose sugars were retained during energy minimization. In these cases, the 5′-OH group of the interacting ribose sugar was replaced by hydrogen atom, whereas the 3′-OH group was retained during calculations.

   Geometry optimization of the base pairs was carried out at the B3LYP/6-31G(d,p) (Lee et al. 1988; Becke 1993) level using Gaussian 09 (Frisch et al. 2009), which was selected in synchrony with previous studies on RNA base pairs (Šponer et al. 2004, 2005c; Sharma et al. 2008, 2010b). The strength of hydrogen bonds between two bases of the modified base pair was calculated in terms of binding energy or interaction energy, which is defined as the extra stabilization acquired by two bases when they form the base pair. Thus, the interaction energy (ΔE^{AB}) of a base pair AB composed of two bases A and B is defined as

$$\Delta E = E^{AB} - (E^A + E^B),$$

where E^{AB} is energy of the base pair, and E^A and E^B are the energies of bases A and B, respectively. Interaction energies were calculated at the RIMP2/aug-cc-pVDZ level (Ahlrichs et al. 1998; Weigend et al. 1998), using the Turbomole v6.2 (http://www.turbomole.com) suite of quantum chemical programs. Although the interaction energies were corrected for basis set superposition error (Boys and Bernardi 1970), the monomer deformation energies were not included in these calculations, since inclusion of deformation energies may bias the strength of base pairs containing flexible ribose sugar (Šponer et al. 2005a,b,c; Mládek et al. 2009; Sharma et al. 2010b).

We note that the interaction energies of base pairs calculated using quantum chemical methods describe the component of intermolecular interaction that originate from interaction of the electronic structure of the constituent bases. However, since the interaction energies do not include other context-dependent interaction components (e.g., solvent effects and entropic contributions), these values may not be directly comparable with the free energies of binding of base pairs. The relation between intrinsic interaction energies and thermodynamic stabilities of base pairs may be especially complex, when the base modification introduces a formal charge (e.g., in base pairs involving N7-methylation of guanine), which may inevitably affect the (long range) electrostatic component of the interaction energy. Nevertheless, the calculated interaction provides the basic stability order for the base pairs, and helps us in understanding the modification-introduced perturbation to the electronic structure of base pairs.

Comparison of macromolecular crystal and optimized geometries of base pairs

RMSD

To understand the difference in base pair geometry in optimized form and in crystal occurrences, the root mean square deviations (RMSDs) of crystal-constrained geometries of each modified base pair were calculated from their corresponding optimized geometries. In addition, to analyze the variation in base pair geometry within crystal occurrences, the RMSD of each crystal occurrence of the base pair was calculated with respect to the average structure among all crystal occurrences. These calculations were done using VMD v1.9 software (Humphrey et al. 1996).

Base-pair parameters

Change in the geometries of base pairs upon optimization and variation in structures of crystal occurrences of base pairs were quantitatively evaluated by comparing the base pair parameters (buckle, propeller, open angle, shear, stretch, and stagger) of the crystal occurrences with the optimized geometry of each base pair, as well as among the different crystal occurrences of the base pair. These calculations were done using an upgraded version of NUPARM software (Bansal et al. 1995; Mukherjee et al. 2006), which uses the edge-specific system for calculation of base pair parameters, which is specific to RNA base pairs.

E-values of hydrogen bonds

To evaluate the relative goodness of hydrogen bonds within base pairs in their crystal occurrences as well as in optimized geometries, we have calculated a parameter called E-value, which is defined as

$$E = \sum_i (d_i - 3.0)^2 + \frac{1}{2}\sum_j (\theta_j - \pi)^2.$$

Here d is the heavy atom distance for each hydrogen bond between two bases under consideration and θ is a pseudo angle subtended by precursor atoms of both the bases (Das et al. 2006). This parameter was used, since the RNA crystal structures from which base pairs were extracted did not contain hydrogen atoms. The E-value parameter assesses the quality of hydrogen bonds in the absence of

hydrogen atom coordinates, and is useful in analyzing hydrogen bonds within the crystal occurrences of base pairs.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

REFERENCES

Ahlrichs R, Elliott SD, Huniar U. 1998. Quantum chemistry: large molecules–small computers. *Ber Bunsen-Ges Phys Chem* **102:** 795–804.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* **215:** 403–410.

Bansal M, Bhattacharyya D, Ravi B. 1995. NUPARM and NUCGEN: software for analysis and generation of sequence dependent nucleic acid structures. *Comput Appl Biosci* **11:** 281–287.

Becke AD. 1993. Density-functional thermochemistry. III. The role of exact exchange. *J Chem Phys* **98:** 5648–5652.

Bhattacharya S, Mittal S, Panigrahi S, Sharma P, Seelam PP, Paul R, Halder S, Halder A, Bhattacharyya D, Mitra A. 2015. RNABP COGEST: a resource for investigating functional RNAs. *Database* **2015:** bav011.

Boys SF, Bernardi F. 1970. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol Phys* **19:** 553–566.

Cantara WA, Crain PF, Rozenski J, McCloskey JA, Harris KA, Zhang X, Vendeix FAP, Fabris D, Agris PF. 2011. The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Res* **39:** D195–D201.

Charette M, Gray MW. 2000. Pseudouridine in RNA: what, where, how, and why. *IUBMB Life* **49:** 341–351.

Chawla M, Sharma P, Halder S, Bhattacharyya D, Mitra A. 2011. Protonation of base pairs in RNA: context analysis and quantum chemical investigations of their geometries and stabilities. *J Phys Chem B* **115:** 1469–1484.

Chawla M, Oliva R, Bujnicki JM, Cavallo L. 2015. An atlas of RNA base pairs involving modified nucleobases with optimal geometries and accurate energies. *Nucleic Acids Res* **43:** 9573.

Chow CS, Lamichhane TN, Mahto SK. 2007. Expanding the nucleotide repertoire of the ribosome with post-transcriptional modifications. *ACS Chem Biol* **2:** 610–619.

Dalluge JJ, Hashizume T, Sopchik AE, McCloskey JA, Davis DR. 1996. Conformational flexibility in RNA: the role of dihydrouridine. *Nucleic Acids Res* **24:** 1073–1079.

Das J, Mukherjee S, Mitra A, Bhattacharyya D. 2006. Non-canonical base pairs and higher order structures in nucleic acids: crystal structure database analysis. *J Biomol Struct Dyn* **24:** 149–161.

Davis DR. 1995. Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Res* **23:** 5020–5026.

Decatur WA, Fournier MJ. 2002. rRNA modifications and ribosome function. *Trends Biochem Sci* **27:** 344–351.

Demirci H, Murphy FV, Murphy EL, Connetti JL, Dahlberg AE, Jogl G, Gregory ST. 2014. Structural analysis of base substitutions in *Thermus thermophilus* 16S rRNA conferring streptomycin resistance. *Antimicrob Agents Chemother* **58:** 4308–4317.

Denmon AP, Wang J, Nikonowicz EP. 2011. Conformation effects of base modification on the anticodon stem–loop of *Bacillus subtilis* tRNA^Tyr. *J Mol Biol* **412:** 285–303.

Dunin-Horkawicz S, Czerwoniec A, Gajda MJ, Feder M, Grosjean H, Bujnicki JM. 2006. MODOMICS: a database of RNA modification pathways. *Nucleic Acids Res* **34:** D145–D149.

Dunkle JA, Wang L, Feldman MB, Pulk A, Chen VB, Kapral GJ, Noeske J, Richardson JS, Blanchard SC, Cate JHD. 2011. Structures of the bacterial ribosome in classical and hybrid states of tRNA binding. *Science* **332:** 981.

Emmerechts G, Maes L, Herdewijn P, Anné J, Rozenski J. 2008. Characterization of the posttranscriptional modifications in *Legionella pneumophila* small-subunit ribosomal RNA. *Chem Biodiv* **5:** 2640–2653.

Frank J, Sengupta J, Gao H, Li W, Valle M, Zavialov A, Ehrenberg M. 2005. The role of tRNA as a molecular spring in decoding, accommodation, and peptidyl transfer. *FEBS Lett* **579:** 959–962.

Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, et al. 2009. *Gaussian 09*. Gaussian, Inc., Wallingford, CT.

Halder A, Halder S, Bhattacharyya D, Mitra A. 2014. Feasibility of occurrence of different types of protonated base pairs in RNA: a quantum chemical study. *Phys Chem Chem Phys* **16:** 18383–18396.

Halder A, Bhattacharya S, Datta A, Bhattacharyya D, Mitra A. 2015. The role of N7 protonation of guanine in determining the structure, stability and function of RNA base pairs. *Phys Chem Chem Phys* **17:** 26249–26263.

Helm M. 2006. Post-transcriptional nucleotide modification and alternative folding of RNA. *Nucleic Acids Res* **34:** 721–733.

Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. *J Mol Graph Model* **14:** 33–38.

Jühling F, Mörl M, Hartmann RK, Sprinzl M, Stadler PF, Pütz J. 2009. tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res* **37:** D159–D162.

Laskowski RA. 2009. PDBsum new things. *Nucleic Acids Res* **37:** D355–D359.

Lee C, Yang W, Parr RG. 1988. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys Rev B* **37:** 785–789.

Leontis NB, Westhof E. 2001. Geometric nomenclature and classification of RNA base pairs. *RNA* **7:** 499–512.

Leontis NB, Stombaugh J, Westhof E. 2002. The non-Watson–Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res* **30:** 3497–3531.

Limbach PA, Crain PF, McCloskey JA. 1994. Summary: the modified nucleosides of RNA. *Nucleic Acids Res* **22:** 2183–2196.

Machnicka MA, Olchowik A, Grosjean H, Bujnicki JM. 2014. Distribution and frequencies of post-transcriptional modifications in tRNAs. *RNA Biol* **11:** 1619–1629.

McCloskey JA, Rozenski J. 2005. The small subunit rRNA modification database. *Nucleic Acids Res* **33:** D135–D138.

Mládek A, Sharma P, Mitra A, Bhattacharyya D, Šponer J, Šponer JE. 2009. *Trans* Hoogsteen/sugar edge base pairing in RNA. Structures, energies, and stabilities from quantum chemical calculations. *J Phys Chem B* **113:** 1743–1755.

Motorin Y, Helm M. 2010. tRNA stabilization by modified nucleotides. *Biochemistry* **49:** 4934–4944.

Mueller EG, Buck CJ, Palenchar PM, Barnhart LE, Paulson JL. 1998. Identification of a gene involved in the generation of 4-thiouridine in tRNA. *Nucleic Acids Res* **26:** 2606–2610.

Mukherjee S, Bansal M, Bhattacharyya D. 2006. Conformational specificity of non-canonical base pairs and higher order structures in

nucleic acids: crystal structure database analysis. *J Comput Aided Mol Des* **20:** 629–645.

Nissen P, Hansen J, Ban N, Moore PB, Steitz TA. 2000. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289:** 920.

Ofengand J. 2002. Ribosomal RNA pseudouridines and pseudouridine synthases. *FEBS Lett* **514:** 17–25.

Ofengand J, Bakin A. 1997. Mapping to nucleotide resolution of pseu-douridine residues in large subunit ribosomal RNAs from representative eukaryotes, prokaryotes, archaebacteria, mitochondria and chloroplasts. *J Mol Biol* **266:** 246–268.

Oliva R, Cavallo L, Tramontano A. 2006. Accurate energies of hydrogen bonded nucleic acid base pairs and triplets in tRNA tertiary interactions. *Nucleic Acids Res* **34:** 865–879.

Oliva R, Tramontano A, Cavallo L. 2007. Mg^{2+} binding and archaeosine modification stabilize the G15–C48 Levitt base pair in tRNAs. *RNA* **13:** 1427–1436.

Piekna-Przybylska D, Decatur WA, Fournier MJ. 2008. The 3D rRNA modification maps database: with interactive tools for ribosome analysis. *Nucleic Acids Res* **36:** D178–D183.

Pruitt KD, Tatusova T, Maglott DR. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **35:** D61–D65.

Sanbonmatsu KY. 2006. Alignment/misalignment hypothesis for tRNA selection by the ribosome. *Biochimie* **88:** 1075–1089.

Sharma P, Mitra A, Sharma S, Singh H, Bhattacharyya D. 2008. Quantum chemical studies of structures and binding in noncanonical RNA base pairs: the *trans* Watson-Crick:Watson-Crick family. *J Biomol Struct Dyn* **25:** 709–732.

Sharma P, Chawla M, Sharma S, Mitra A. 2010a. On the role of Hoogsteen: Hoogsteen interactions in RNA: ab initio investigations of structures and energies. *RNA* **16:** 942–957.

Sharma P, Šponer JE, Šponer J, Sharma S, Bhattacharyya D, Mitra A. 2010b. On the role of *cis* Hoogsteen:sugar edge family of base pairs in platforms and triplets—quantum chemical insights into RNA structural biology. *J Phys Chem B* **114:** 10234.

Šponer J, Jurečka P, Hobza P. 2004. Accurate interaction energies of hy-drogen-bonded nucleic acid base pairs. *J Am Chem Soc* **126:** 10142–10151.

Šponer JE, Leszczynski J, Sychrovský V, Šponer J. 2005a. Sugar edge/sug-ar edge base pairs in RNA: stabilities and structures from quantum chemical calculations. *J Phys Chem B* **109:** 18680–18689.

Šponer JE, Špačková Na, Leszczynski J, Šponer J. 2005b. Non-Watson-Crick base pairing in RNA. Quantum chemical analysis of the *cis* Watson-Crick/sugar edge base pair family. *J Phys Chem A* **109:** 2292–2301.

Šponer JE, Špačková Na, Leszczynski J, Šponer J. 2005c. Principles of RNA base pairing: structures and energies of the *trans* Watson-Crick/sugar edge base pairs. *J Phys Chem B* **109:** 11399–11410.

Šponer J, Šponer JE, Petrov AI, Leontis NB. 2010. Quantum chemical studies of nucleic acids: can we construct a bridge to the RNA structural biology and bioinformatics communities? *J Phys Chem B* **114:** 15723–15741.

Stombaugh J, Zirbel CL, Westhof E, Leontis NB. 2009. Frequency and isostericity of RNA base pairs. *Nucleic Acids Res* **37:** 2294–2312.

Vicens Q, Westhof E. 2001. Crystal structure of paromomycin docked into the eubacterial ribosomal decoding A site. *Structure* **9:** 647–658.

Walsh CT, Garneau-Tsodikova S, Gatto GJ. 2005. Protein posttransla-tional modifications: the chemistry of proteome diversifications. *Angew Chem Int Ed* **44:** 7342–7372.

Weigend F, Häser M, Patzelt H, Ahlrichs R. 1998. RI-MP2: optimized auxiliary basis sets and demonstration of efficiency. *Chem Phys Lett* **294:** 143–152.