



Cite this: *RSC Adv.*, 2017, 7, 35694

# Higher order structures involving post transcriptionally modified nucleobases in RNA†

Preethi S. P.,<sup>a</sup> Purshotam Sharma <sup>\*b</sup> and Abhijit Mitra <sup>\*a</sup>

Post transcriptionally modified nucleotides play an important structural role in RNA by providing additional chemical diversity to the architecture of the constituent nucleotides. To understand the contribution of these modified nucleotides in tertiary interactions that help in RNA folding, we present a systematic investigation of higher order hydrogen bonded structures involving modified nucleotides using bioinformatics based crystal structural database analysis and quantum chemical calculations. Our analysis reveals that 29% of the modified bases that participate in hydrogen-bonding interactions form higher order motifs, which points towards their importance as RNA building blocks. Although greater geometric variations in such motifs are observed in tRNA, they occur most commonly in rRNA, where they may play a role in RNA folding. Characterization of the optimum geometries and binding strengths of the modified motifs reveals that structures that involve either a positively charged modified nucleobase ( $m^7G^+$ ), or a protonated nucleobase ( $A^{H^+}$ ) paired with a modified base ( $\Psi$ ), possess stronger binding compared to uncharged (neutral) motifs. Thus, they seem to play an important role in stabilizing complex RNA structures, possibly through enhanced electrostatic interactions. Overall, our combined statistical, geometric, energetic and contextual analysis of these modified motifs reveals their importance as stable RNA building blocks, and highlights the need for further investigations related to their functional roles.

Received 10th May 2017  
Accepted 12th July 2017

DOI: 10.1039/c7ra05284g

rsc.li/rsc-advances

## Introduction

The role of posttranscriptional modifications in facilitating the functional diversity of RNA molecules has drawn considerable attention in contemporary biochemical literature.<sup>1–6</sup> From a wider perspective, the characterization of such modified nucleotides as intrinsic building blocks of RNA has been a step forward towards addressing the long standing conundrum on how RNA performs a variety of functional roles, despite its limited constitutional diversity in terms of the presence of only four different nucleobases (adenine (A), cytosine (C), guanine (G) and uracil (U)). However, adequate understanding of the role of modified nucleotides in shaping the structural diversity of RNA macromolecules is yet to be achieved.

The evolutionarily conserved 3D structures of folded RNA macromolecules are stabilized through a variety of non-covalent interactions between the constituent nucleotides. The most prominent among these are the stacking interactions between aromatic nucleobase rings,<sup>7</sup> and the inter-nucleotide hydrogen

bonding (IHB) interactions that constitute the base–base, base–sugar, and base–phosphate interactions.<sup>8–12</sup> Further, in addition to the canonical (A:U and G:C) base–base interactions, ‘non canonical’ IHB interactions play unique structural roles in RNA.<sup>13,14</sup> Such interactions can further be extended to form base triples, quadruples and other higher order structures that contribute towards RNA folding.<sup>15,16</sup>

The structures of IHB motifs have been characterized in previous studies, mainly on the basis of the general assumption that each of these motifs is an intrinsically stable recurrent building block of RNA. Specifically, two complementary techniques: structural bioinformatics based crystal structure database analysis<sup>8,14,16–19</sup> and quantum mechanical (QM) calculations,<sup>11,20–25</sup> have been used to characterize such motifs. In addition to analysing the occurrence frequency and the stability of IHB motifs, such studies have highlighted the differences between the “intrinsic structure” of the motif, and the structure it adopts within RNA.<sup>23,26</sup> These structural differences point towards the role of RNA macromolecular context in shaping the geometries of IHB motifs.

There are a significant number of studies on IHB interactions involving natural nucleotides. However, though modified nucleotides play important roles in determining the structure, folding and function of RNA,<sup>16,17,19,27–30</sup> there are only a few recent studies reporting the occurrence, structures and stabilities of IHB interactions involving modified nucleosides in RNA.<sup>31,32</sup> Oliva *et al.* analysed four modified base pair and two

<sup>a</sup>Center for Computational Natural Sciences and Bioinformatics, International Institute of Information Technology Hyderabad (IIIT-H), Gachibowli, Hyderabad, Telangana, 500032, India. E-mail: abi\_chem@iiit.ac.in

<sup>b</sup>Computational Biochemistry Laboratory, Department of Chemistry and Centre for Advanced Studies in Chemistry, Panjab University, Chandigarh, 160014, India. E-mail: psharma@pu.ac.in

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c7ra05284g



modified triple motifs constituting the conserved tertiary interactions in tRNA using quantum chemical methods.<sup>26</sup> Subsequent studies on one of the conserved (*i.e.* G15:C48) base pair motif in tRNA, established the role of magnesium binding, archaeosine modification,<sup>33</sup> or triple formation with dihydrouridine modification present at position 20,<sup>20</sup> in stabilization of this motif. More recently, studies have characterized the occurrence context, structures and stabilities of naturally occurring IHB pairs that involve modified nucleotides.<sup>31,32</sup> However, a systematic analysis of structural features of higher order interactions involving modified bases in RNA is currently absent in the literature.

Addressing this void, we have carried out a comprehensive analysis of the crystal structures of different classes of RNA, to identify the full range of modified motifs and enumerate their diversity.<sup>31</sup> Based on this search, we identified 15 different types of modified base triples and 3 different types of base quadruple modified motifs. Further, a representative structure corresponding to each modified motif was extracted from the best resolution crystal structure, which was subjected to the quantum chemical analysis based approach of energy minimization and binding energy calculations.<sup>22,34</sup> A detailed comparison of the “experimental” (*i.e.* within RNA macromolecular crystal structure) and “energy minimized” structures, respectively, of each of these motifs was then carried out using standard nucleic acid motif characterization parameters.<sup>35,36</sup> Finally, an analysis of macromolecular occurrence context of these motifs was carried out to provide clues about the associated structural and functional roles of these motifs. Overall, our detailed analysis significantly adds to the understanding of role played by higher order motifs involving modified nucleotides, in RNA structures.

## Materials and methods

To identify modified motifs, the occurrence of such bases was searched in a dataset of RNA crystal structures, that was chosen according to specific search criteria (details in ESI†), and the relevant protein databank (PDB) entries, submitted till 18 July 2016 with resolution better than 3.5 Å,<sup>31</sup> were included. BPFIND software<sup>36</sup> was used to analyse the occurrence, location and type of higher order associations containing modified bases. Geometry optimizations of the modified motifs retrieved from the RNA crystal structures, as well as their unmodified counterparts retrieved either from crystal structures or modelled from modified pairs, was carried out using the B3LYP<sup>37,38</sup> density functional and the 6-31G(d,p) basis set using Gaussian 09.<sup>39</sup> The geometries, optimized using this method, have been shown to compare very well with the reference RIMP2/cc-pVTZ geometries, and are sufficient for interaction energy calculations.<sup>40</sup> Basis set superposition error<sup>41</sup> corrected binding energies of these base–base associations were calculated at the RIMP2 level<sup>42</sup> using the aug-cc-pVDZ basis set. The binding energies calculated at this level compare well with the reference MP2/CBS(T) values, where an underestimation of only up to 1.11 kcal mol<sup>-1</sup> is observed in a previous reference study on RNA base pairs.<sup>43</sup> Further, these methods were selected to

maintain conformity with previous studies on natural and modified IHB motifs involving RNA.<sup>24,25,31,34,44</sup> Additional details regarding model building and the methodology followed for binding energy calculations are provided in ESI.†

Variation in geometries of the modified motifs within the crystal occurrences, between the crystal and the optimized geometries, as well as between the modified and unmodified triples or quadruples, was quantified in terms of their root mean square deviation (rmsd) values using the VMD v1.9 software.<sup>45</sup> Such geometric variations were further analysed by comparing the base pair parameters (*i.e.* buckle, propeller, open angle, shear, stretch and stagger) of the motifs using upgraded version of NUPARM software.<sup>35</sup> In addition, the relative goodness of hydrogen bonds between two different geometries of the same higher order interaction were empirically estimated using an empirical *E*-value parameter (detailed in ESI†) using the BPFIND software.<sup>36</sup>

## Results and discussion

### Statistical overview of higher order interactions of modified bases in RNA 3D structures

Fifteen different naturally occurring post-transcriptionally modified nucleosides that participate in RNA base pairing,<sup>31</sup> were searched in a dataset of 207 high resolution RNA crystal structures identified in our previous study,<sup>31</sup> to analyse their propensity to form modified motifs (Table 1). Of these, nine modified nucleosides participate in motifs (Fig. 1 and Table 1), four of which involve base-methylation (m<sup>6</sup><sub>2</sub>A, m<sup>7</sup>G<sup>+</sup>, m<sup>5</sup>C and m<sup>5</sup>U), two involve 2'-O-methylation at ribose (Gm and Um) and three involve other modifications (s<sup>4</sup>U, Ψ and D, Fig. S1–S3†). More importantly, a sizable proportion (*i.e.* 30%) of the base pairing occurrences involving these nine modified nucleosides occur as part of modified motifs (Fig. 2 and Table 1), which points toward their structural and functional importance in RNA.

Our dataset of RNA crystal structures spanning all major RNA classes (Tables S1 and S2†) shows that modified triples and quadruples (Fig. S4†) are most commonly observed in rRNA (71%), followed by tRNA (26%) and group I intron (3%, Fig. 2). It is noteworthy that despite their high occurrence in rRNA, higher diversity of modified motifs is observed in tRNA, that spans five (*i.e.* m<sup>6</sup><sub>2</sub>A, m<sup>7</sup>G<sup>+</sup>, s<sup>4</sup>U, Ψ and D) of the nine modified nucleotides that form motifs.

Analysis, of occurrence context of modified motifs, reveals that they are observed in all major RNA structural elements. Specifically, more than half of these interactions participate in stem–loop tertiary interactions (57%), followed by loop–loop (17%), junction–loop (14%) and RNA–RNA interface interactions (12%, Fig. 2). Additionally, ten unique modified triple associations participate in the stabilization of A/G-minor motifs<sup>46</sup> or in dinucleotide platforms stabilization (Table 2). Overall, when classified according to the type of interacting edges (*i.e.* Watson–Crick edge (W), Hoogsteen edge (H) and sugar edge (S)<sup>9</sup>), our analysis reveals fifteen distinct modified triples that span nine base triple geometric families,<sup>16</sup> and three distinct modified quadruples. Two of these modified triple



**Table 1** Distinct types of modified bases observed in the RNA crystal structure dataset. Total number of interactions are given in the parenthesis in the first row

Modified bases	Unpaired bases (631)	Base pairs (453)	Base triples (120)	Base quadruples (15)
N6,6-Dimethyl adenine ( $m^6_2A$ )	—	4	1	—
N7-Methyl guanine ( $m^7G^+$ )	14	55	31	—
2'-O-Methyl guanine (Gm)	23	44	44	12
C5-Methyl cytosine ( $m^5C$ )	61	90	2	—
Pseudouridine ( $\Psi$ )	231	79	26	—
5,6-Dihydrouridine (D)	101	15	2	—
C5-Methyl uracil ( $m^5U$ )	15	39	0	3
2'-O-Methyl uracil (Um)	59	9	1	—
4-Thiouridine ( $s^4U$ )	1	23	13	—
N1-Methyl adenine ( $m^1A$ )	53	21	—	—
2'-O-Methyl adenine (Am)	12	12	—	—
N2-Methyl guanine ( $m^2G$ )	1	42	—	—
N2,2-Dimethyl guanine ( $m^2_2G$ )	25	16	—	—
2'-O-Methyl cytosine (Cm)	35	3	—	—
2',5-Dimethyl uracil ( $m^5Um$ )	0	1	—	—

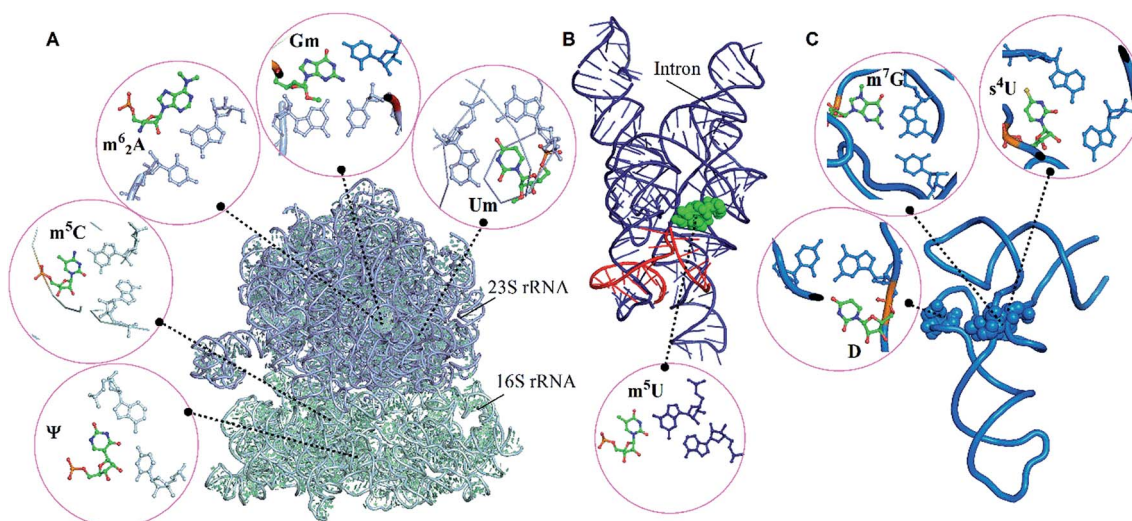
motifs involve the positively charged modified base ( $m^7G^+$ ), whereas one involves a protonated adenine ( $A^{H^+}$ ) interacting with pseudouridine ( $\Psi$ , Table 2).

### Structural characteristics of higher order motifs involving modified bases

(i) **Crystal structure occurrences.** Based on their location within the motif, the higher order associations involving

modified bases can be divided into two classes – (i) structures where the modified base is located at the central position and interacts with all other bases within the motif, and (ii) structures where modified base interacts with only one of the other bases within the motif (Fig. 3). The first category involves two A-minor interactions ( $C:m^7G^+:A$  and  $A:Gm:G$ ), a G-minor interaction ( $C:Gm:G$ ), and a platform interaction ( $G:Um:A$ ). In addition, this category includes three of the motifs ( $A:s^4U:Aw$ ,  $A:s^4U:As$ ,  $A:\Psi:C$ ) involve an uracil modification (*i.e.*  $s^4U$  or  $\Psi$ ), where the modified base interacts through its H-edge with one base and S-edge with the other (Fig. 4). On the other hand, the second category includes three A-minor motifs ( $m^6_2A:G:U$ ,  $m^5C:G:A$  and  $m^5U:G:A$ ), where the modified nucleoside interacts through its S- and W-edges ( $m^6_2A:G:U$ ) or the W-edge ( $m^5C:G:A$  and  $m^5U:G:A$ ), a G-minor motif ( $Gm:G:C$ ) where the modified nucleoside interacts through its S-edge, and a platform interaction ( $Dh:G:C$ ) involving association of D through its H-edge. In remaining three motifs ( $m^7G^+:G:C$ ,  $\Psi:G:A^{H^+}$  and  $Dw:G:C$ ), the modified base interacts with one of the triple bases through its W-edge (Fig. 5). Finally, three modified quadruple motifs are observed ( $C:Gm:G:C$ ,  $A:Gm:G:C$  and  $m^5U:G:A:A$ ).

Of the fifteen modified triple motifs, nine associations involve multiple occurrences (Table S3<sup>†</sup>). Geometrical variations were observed within these multiple occurrences, the magnitude of which depends on the identity of the constituent nucleosides and their respective macromolecular structural context (Fig. S5<sup>†</sup>). For example, although 31 occurrences of Gm2588:G2617:C2542 triple within the crystal structures of 23S rRNA show negligible average rmsd (0.15 Å), the 13 occurrences of A14: $s^4U$ :A21 motif in tRNA show a significant rmsd (1.07 Å). This variability may be attributed to the location of the Gm2588:G2617:C2542 triple in the stable stem-loop interaction region of 23S rRNA, which contrasts the flexible junction loop location of the A14: $s^4U$ :A21 in tRNA structures. Overall, the average rmsd ranges between 0 Å and 1 Å within different crystal occurrences of modified motifs (Table S4<sup>†</sup>). Further, analysis of



**Fig. 1** Modified motifs found in the (A) ribosomal RNA (B) group I intron/exon complex and (C and D) tRNA. The modified base in each of the motif is represented in green color.



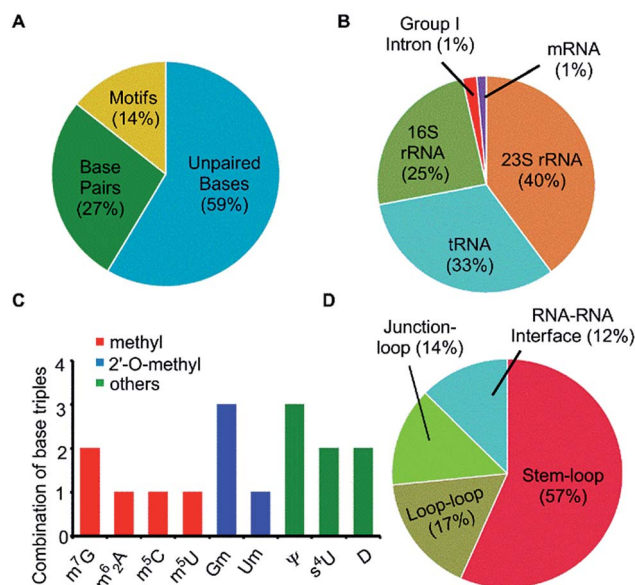


Fig. 2 (A) Distribution of modified bases as function of unpaired bases, base pairs and higher order interactions. (B) Distribution of modified motifs as function of type of RNA. (C) Distribution of nine modified bases that form modified motifs, in terms of observed number of triple geometries. (D) Distribution of modified motifs in different RNA structural elements.

base pair parameters of modified motifs reveals that associations involving only base–base (B–B) interactions show little deviation in the base pair parameters, and exhibits up to 12° of average of buckle, propeller and open values. Within the motifs involving base–sugar (B–S) and/or sugar–sugar (S–S) interactions, largest buckle is observed in  $m^6A:G:U$ , mainly due to deviation in pairing geometry on methylation of A. Similarly,  $Dw:G:C$  motif exhibits largest propeller twist, due to inherent puckering of the dihydrouridine ring. However, owing to the

constraints arising out of satisfying planarity requirements, triples that stabilize the platform interactions (*i.e.*  $G:Um:A$  and  $Dh:G:C$ ) display small deviations in the structural parameters. Overall, when the distribution of base pair parameters with respect to the identity of the modified base is considered, motifs involving D or  $m^6A$  exhibit the most nonplanar geometries with large (up to 40°) deviation in the buckle and propeller parameters. It may however be noted that, due to limited resolution of RNA crystal structures, many of the observed geometrical deformations, in crystal structure occurrences of modified triples or quadruples, may be artefacts arising out of crystal data refinement errors. Thus, to characterize optimum geometries of these motifs in the absence of crystal environment, we carried out quantum chemical energy minimizations.

(ii) **QM optimized geometries.** Geometry optimization (energy minimization) of the crystal occurrences of modified IHB motifs in isolation, is expected to reveal the inherent structural features of these motifs. This may, in turn, provide insights into the contribution of base–base hydrogen bonding in determining the geometries of these motifs. Overlays of the structures of each motif, extracted from the corresponding best resolution RNA crystal structure, on its quantum chemical energy minimized structure, reveal that the structures involving sugar–edge interactions generally show higher rmsd and greater deviations in base pair parameters, compared to those involving other two edges (Tables S5 and S6†). This can be explained on the basis of flexibility associated with the sugar moiety, which relaxes to the nearest local minimum structure on minimization, and increases the rmsd of the overall structure. Nevertheless, no change in hydrogen bonding pattern is observed on optimization of any of the studied motifs, although the quality of interbase hydrogen bonds measured in terms of *E*-values and the geometrical characteristics measured in terms of base pair parameters, improve on optimization (Table S8†). Overall, our analysis indicates that modified triple or quadruple motifs are

Table 2 Distribution of modified motifs as per type of RNA, structural element and tertiary interactions formed

Modified motifs	Geometry	Interaction type	Type of RNA	Position	$f^a$	Structural element
$C:m^7G^+:A$	W:WC/S:WT	B–B/S–B	16S rRNA	522:527:535	11	Stem–loop (A–minor)
$C:Gm:G$	W:WC/S:SC	B–B/S–S	tRNA:rRNA	75:2588:2617	12	RNA–RNA interface (G–minor)
$A:Gm:G$	W:WC/S:SC	B–B/S–S	tRNA:rRNA	76:2588:2617	1	RNA–RNA interface (A–minor)
$A:s^4U:Aw$	H:WT/S:WT	B–B/S–B	tRNA	14:8:21	12	Junction–loop
$A:s^4U:As$	H:WT/S:SC	B–B/S–S	tRNA	14:8:46	1	Junction–loop
$A:\Psi:C$	H:WT/S:WC	B–B/S–B	16S rRNA	533:516:519	22	Loop–loop
$G:Um:A$	S:HC/W:HT	S–B/B–B	23S rRNA	2655:2656:2665	1	Internal loop (platform)
$m^6A:G:U$	W:ST/W:WC	B–B/S–B	tRNA:23S rRNA	76:2618:2541	1	Junction–loop (A–minor)
$m^7G^+:G:C$	W:HT/W:WC	B–B/B–B	tRNA	46:22:13	20	Stem–loop
$Gm:G:C$	S:SC/W:WC	S–S/B–B	23S rRNA	2588:2617:2542	30	Stem–loop (G–minor)
$m^5C:G:A$	W:WC/S:WT	B–B/S–B	16S rRNA	1404:1497:1518	2	Stem–loop (A–minor)
$\Psi:G:A^{H^+}$	W:WC/H:WT	B–B/B–B	tRNA	13:22:46	4	Stem–loop
$m^5U:G:A$	W:WC/S:SC	B–B/S–S	Group I intron	1:10:58	3	RNA–RNA interface (A–minor)
$Dw:G:C$	W:ST/W:WT	B–S/B–B	tRNA	20:15:48	1	Loop–loop
$Dh:G:C$	H:SC/W:WC	B–S/B–B	tRNA	20:19:56	1	Loop–loop (platform)
$C:Gm:G:C$	W:WC/S:SC/W:WC	B–B/S–S/B–B	tRNA:rRNA	75:2588:2617:2542	12	RNA–RNA interface
$A:Gm:G:C$	W:WC/S:SC/W:WC	B–B/S–S/B–B	tRNA:rRNA	76:2588:2617:2542	1	RNA–RNA interface
$m^5U:G:A:A$	W:WC/S:SC/H:ST	B–B/S–S/B–S	Group I intron	1:10:58:84	1	RNA–RNA interface

<sup>a</sup> *f* stands for occurrence frequency of modified base triples in the dataset.



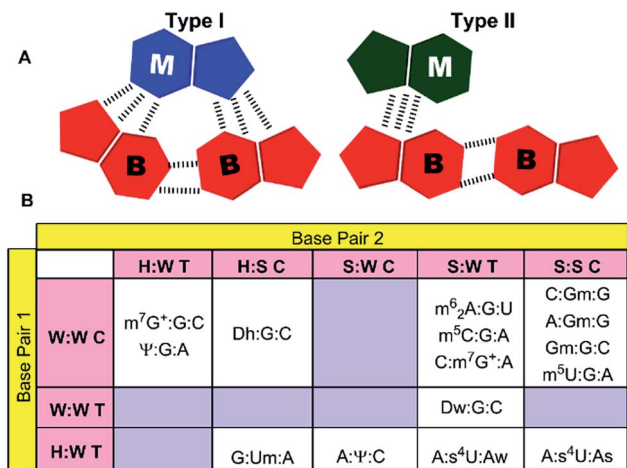


Fig. 3 (A) Graphic representation of two categories of modified motifs. M represents the modified base, whereas B represents the natural bases. (B) Matrix representation of different base triple geometries observed in the modified motifs.

intrinsically stable well-defined building blocks of RNA structures (Table S9<sup>†</sup>). This further highlights that hydrogen bonding is the main stabilizing force in these motifs, and surrounding crystal environment plays a less significant role in determining their geometries and stabilities.

(iii) **Comparison with unmodified motifs.** Comparison of the QM optimized structures of the motifs containing modified bases, with those with the corresponding unmodified motifs, is expected to reveal the effect of modification on the geometry and stability of higher order interactions in RNA (Fig. 6). Out of the four combinations that involve methylation at 2'-OH of the interacting ribose, two (*i.e.* A:Gm:G and Gm:G:C) result in loss of one inter base hydrogen bond on modification, which leads to a reduction in binding energy by up to 8 kcal mol<sup>-1</sup>. Further, although methylation reduces the buckle within the A:Gm pair of A:Gm:G motif by 26°, the buckle within the Gm:G pair of the Gm:G:C motif increases by 15° on methylation (Tables S7 and S8<sup>†</sup>). For the remaining two motifs involving modification at 2'-OH (*i.e.* C:Gm:G and G:Um:A), although the hydrogen bonding pattern remains unaffected, the binding energy of C:Gm:G

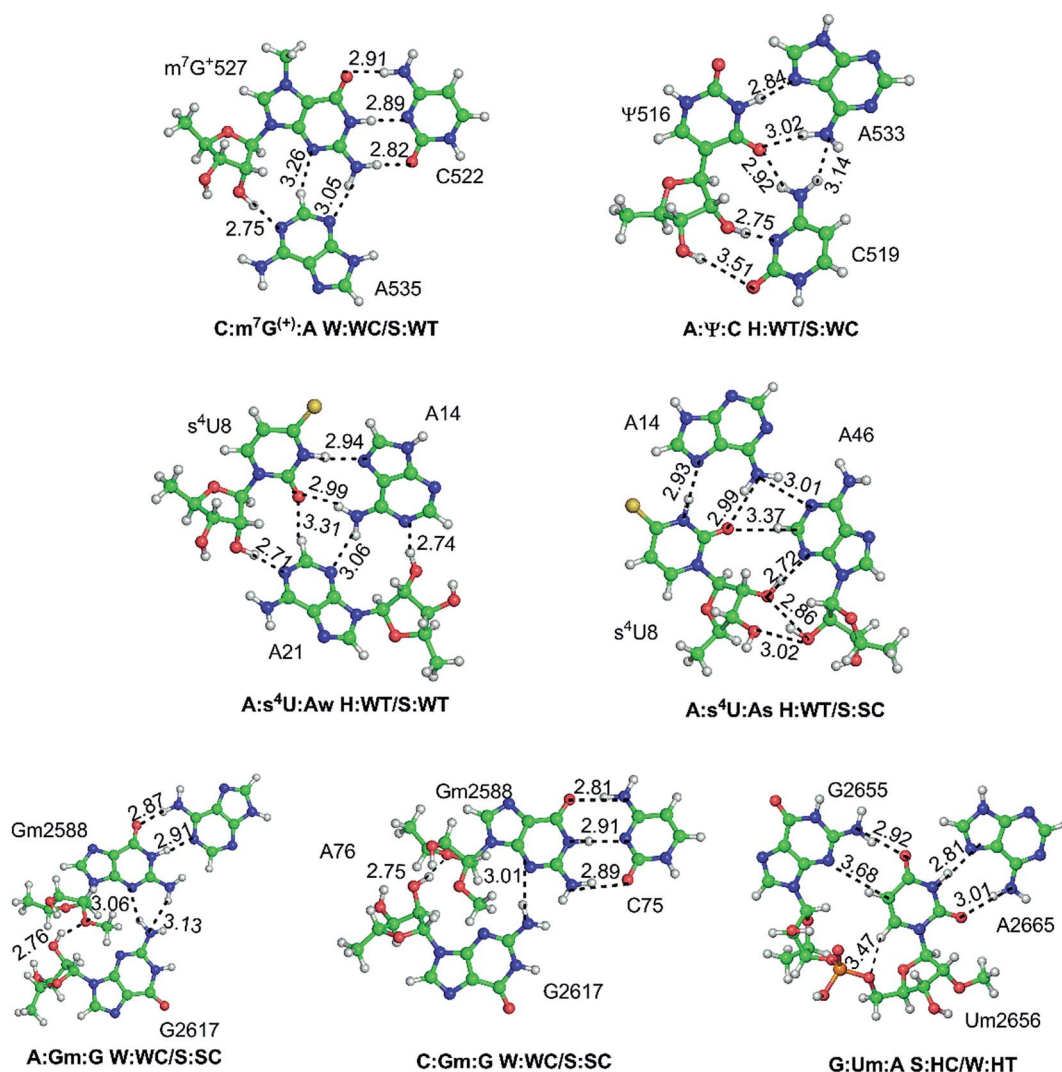


Fig. 4 Schematic representation of optimized geometries of type I modified motifs.



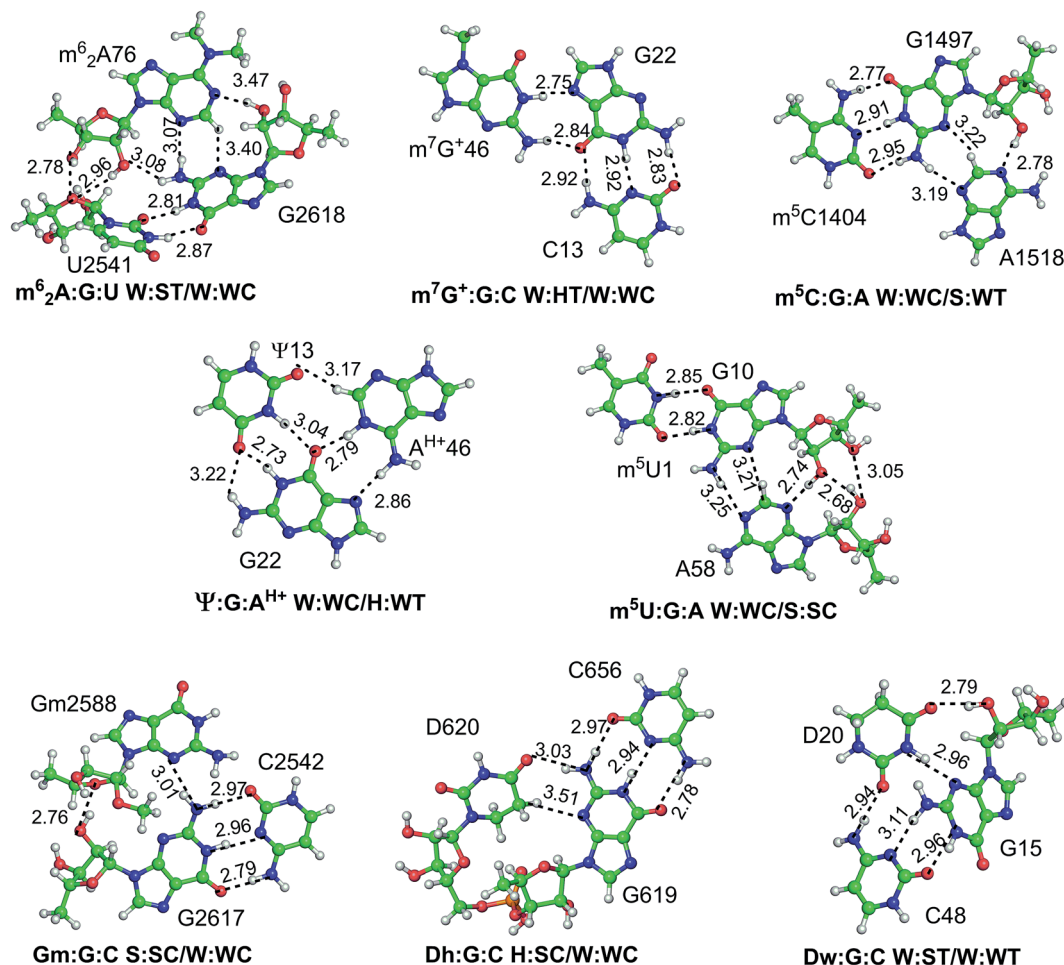


Fig. 5 Schematic representation of optimized geometries of type II modified motifs.

increases by 3 kcal mol<sup>-1</sup> on modification (Table 3), probably due to enhanced dispersion interaction originating from addition of bulky nonpolar methyl group. However, since the ribose sugar of Um of G:Um:A motif is positioned away from the site of internucleotide interactions, the binding energy does not significantly change (*i.e.* remains within 1 kcal mol<sup>-1</sup>) on modification (Table 3).

Within the fifteen motifs that undergo modification at the nucleobase portion of the nucleotide, two motifs (C:m<sup>7</sup>G<sup>+</sup>A and m<sup>7</sup>G<sup>+</sup>:GC) exhibit significant enhancement (by 14 and 28 kcal mol<sup>-1</sup> respectively) in binding energy due to introduction of positive change on the nucleobase architecture. However, rest of the twelve motifs undergo little (*i.e.* within 2 kcal mol<sup>-1</sup>) change in binding energy on modification (Table 3).

### Structurally and functionally important modified base triples and quadruples in RNA

Our detailed analysis of the occurrence contexts of modified triples and quadruples revealed the presence of such motifs within unique structural contexts in different RNA classes, which indicate their putative functional roles. Here we outline the RNA structural environment around each of these motifs,

and project the conclusions derived from our quantum chemical studies on isolated motifs onto the wider biochemical context of their occurrence in RNA structures.

(i) **Modified motifs in tRNA.** Previous structural studies have shown that modified motifs commonly occur at three conserved triple positions (*i.e.* 8:14:21, 13:22:46 and 19:20:56) that constitute a network of tertiary interactions through which the D-arm interacts with other regions of tRNA.<sup>26</sup> The remarkable occurrence of modified motifs, at these positions, point towards their functional importance. Our crystal structure analysis reveals that C:G:m<sup>7</sup>G<sup>+</sup> and Ψ:G:A<sup>H+</sup> are the most commonly detected modified triples at 13:22:46 positions of tRNA in *S. cerevisiae*. This, coupled with the results from our quantum chemical analysis that both these motifs provide the greatest stabilization (up to 13 kcal mol<sup>-1</sup>) compared to the unmodified motifs, indicate that the presence of modified bases, within such tertiary interactions, impart significant structural stabilization to tRNA. It is possible that such local structural stabilization strategies involving base modifications might be common to different RNA structures.

In addition to structure stabilization, our analysis reveals that the presence of modified bases in tRNA tertiary interactions



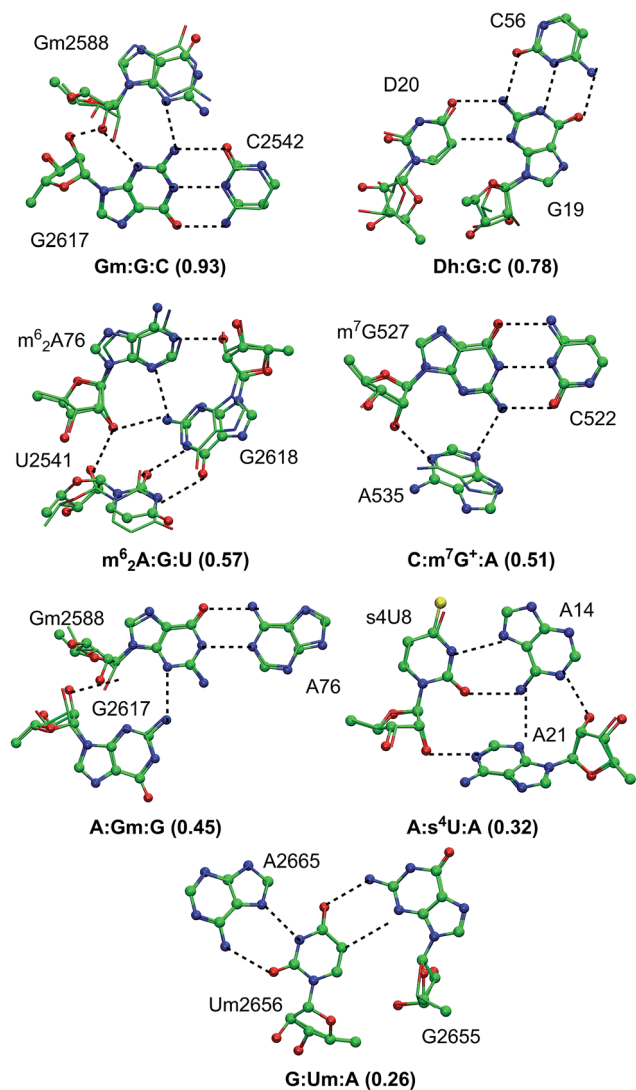


Fig. 6 Alignment of optimized structures of modified and unmodified base triples. RMSD (in Å) for each motif is given in parentheses. Optimized structure of modified motifs is represented as ball and stick, where the unmodified structures as sticks. Motifs with RMSD > 2 Å are shown in the picture. Hydrogen atoms within the motifs are removed for clarity.

may also play a role in stabilization of alternate conformations, which may in turn enhance tRNA flexibility. Specifically, depending on the sequence, the 8:14 pair can form two different triple motifs. The first motif observed in *E. coli* involves the  $s^4U8:A14:A21$  triple that stacks with the 13:22:46 triple (Fig. 7). However, in the second motif present in *T. aquaticus*, A21 is flipped out, and the resulting  $s^4U8:A14:A46$  triple stacks over only the 13:22 pair. The associated loss of hydrogen bonding between 13:22 and 46 in the second motif, contrasting the scenario in the first motif, is partially compensated by enhanced binding (by 7 kcal mol<sup>-1</sup>) within the  $s^4U8:A14:A46$  triple compared to the  $s^4U8:A14:A21$  motif. This intricate trade-off, between competing tertiary interactions, may help in stabilization of the alternate local tRNA conformations on the one hand, and provide local flexibility on the other.

Table 3 Interaction energies (kcal mol<sup>-1</sup>) of the modified base with rest of the bases within the modified motifs

Higher order interaction	Modified	Unmodified
C:m <sup>7</sup> G <sup>+</sup> :A	-59.4	-45.9
C:Gm:G	-45.7	-42.8
A:Gm:G	-34.1	-39.4
A:s <sup>4</sup> U:Aw	-28.1	-29.5
A:s <sup>4</sup> U:As	-35.2	-34.9
A:Ψ:C	-34.2	-35.6
G:Um:A	-25.6	-26.6
m <sup>6</sup> <sub>2</sub> A:G:U	-30.5	-31.4
m <sup>7</sup> G <sup>+</sup> :G:C	-38.0	-20.2
Gm:G:C	-19.3	-25.9
m <sup>5</sup> C:G:A	-29.7	-29.1
Ψ:G:A <sup>H+</sup>	-24.1	-23.1
m <sup>5</sup> U:G:A	-17.2	-16.8
Dw:G:C	-20.3	-21.1
Dh:G:C	-7.7	-10.4
m <sup>5</sup> U:G:A:A	-17.1	-16.7
C:Gm:G:C	-43.1	-38.4

Another example of involvement of modified bases, in stabilization of alternate tertiary interaction motifs, is observed on comparison of tRNA crystal structures of *E. coli* and *T. thermophilus*. Whereas the D20 residue, with its flexible  $\chi$ -torsion interacts with the conserved G15:C48 pair in *T. thermophilus* to form a D:G:C modified motif that connects the D-loop and the V-arm, it interacts with the G19:C56 pair in *E. coli* to form a D20:G19:C56 triple containing a dinucleotide platform,<sup>20</sup> that connects the D-loop with the TΨC loop (Fig. 7). In 20:15:48 modified motif, the binding of D20 to the intrinsically unstable

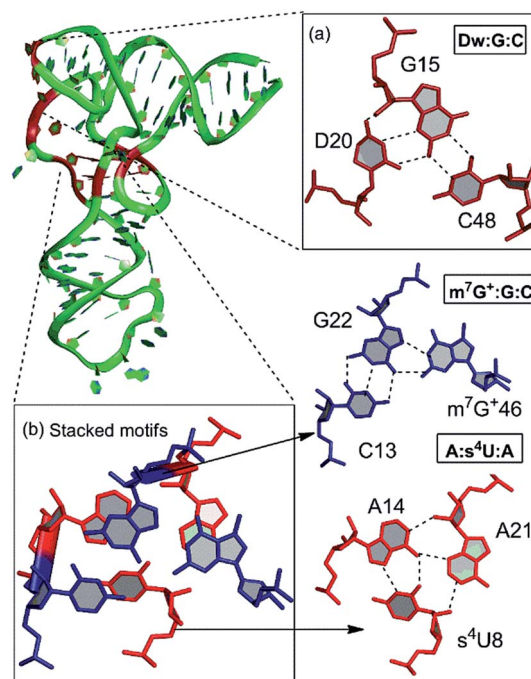


Fig. 7 Modified motifs associated with tertiary interactions in tRNA.



G15:C48 Levitt base pair<sup>20</sup> provides an extra stabilization of 21 kcal mol<sup>-1</sup> to the motif. On the other hand, binding of D20 to the G19:C56 W:W *cis* pair results in the formation of a stable platform, which provides an additional 20 kcal mol<sup>-1</sup> extra stabilization. Overall, our analysis demonstrates how the presence of modified motifs provides enhanced stabilization, and conformational variability to tRNA structures.

**(ii) Quadruple:triple stack at the tRNA:rRNA interface helps in tRNA accommodation on the ribosome.** The crystal structure of tRNA:rRNA complex of *H. marismortui* (an archaea) and *E. coli* (a bacteria) reveal the presence of a triple motif at the tRNA-interface region of rRNA, which stabilizes the 3'-CCA region of tRNA during its accommodation on the ribosome. Previous studies have revealed that G at one of the positions in the *E. coli* triple is replaced by Gm in case of *H. marismortui* (Fig. 8). Although the presence of Gm, in *H. marismortui*, results in loss of hydrogen bonding interactions involving the 2'-OH group, the associated modified triple is more planar compared to the corresponding unmodified triple present in *E. coli*. Thus, the presence of Gm within the triple might help in greater optimization of the triple motif, which may in turn, help in better accommodation of tRNA on the ribosome in the evolutionarily advanced archaea (*H. marismortui*) compared to that in bacteria (*E. coli*).

**(iv) Modified motifs near the codon-anticodon-ribosome interface.** It is known from literature that the bases 1492 and 1493 of helix 44, and 530 of helix 18, of 30S rRNA secure the cognate anti-codon stem loop (ASL) of tRNA with the decoding region of rRNA *via* hydrogen bonding and van der Waals contacts.<sup>47</sup> It has further been proposed that structural changes in these regions can affect the codon-anticodon-ribosome interaction, which might affect the specificity of tRNA binding. In this context, we observed three conserved modified motifs, *i.e.* C522:m<sup>7</sup>G<sup>+</sup>527:A535 and C519:Ψ516:A533, within the helix 18, and m<sup>5</sup>C1404:G1497:A1518 within helix 44. These motifs are near the codon-anticodon-ribosome complex (Fig. 9). Of these, m<sup>5</sup>C:G:A and C:m<sup>7</sup>G<sup>+</sup>:A motifs are supported by A-minor associations, where the base A533 or A535 interacts from the minor groove of m<sup>5</sup>C:G or m<sup>7</sup>G<sup>+</sup>:C base pairs respectively. This involves sugar-sugar interactions and provides additional stabilization to the motif by up to 21 kcal mol<sup>-1</sup>. The presence of these modified

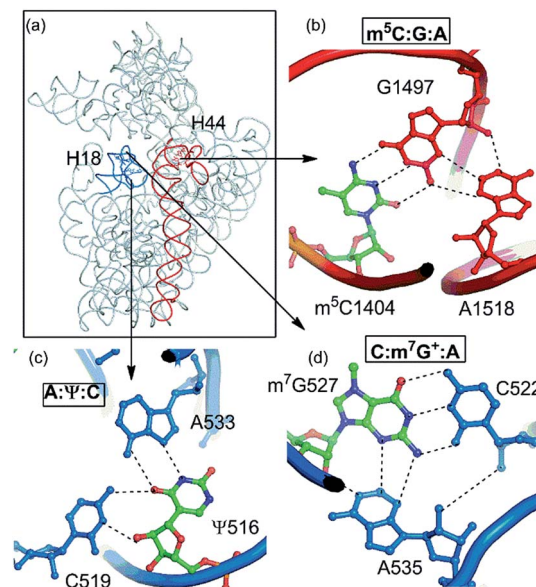


Fig. 9 Modified motifs present in helix 44 and helix 18 respectively of the decoding region of 16S rRNA.

motifs within the functionally important decoding region of rRNA suggests that they may have a role in maintaining the structural and energetic characteristics of the region.

**(v) Modified quadruple helps in the recognition of exon by the group I intron/exon complex.** The m<sup>5</sup>U:G:A:A motif illustrates the role of higher order interactions in the process of recognition of the exon in splice sites. In *Azoarcus* sp. group-I intron/exon complex (PDB 1u6b, Fig. 10), the m<sup>5</sup>U1 of the exon forms a wobble pair with G10 of the internal guide sequence of intron. This pair is specifically recognized by the A-rich wobble receptor domain of intron, where A-minor association of the A58:A84 H:ST pair with the wobble pair results in the m<sup>5</sup>U1:G10:A58:A84 quadruple within the splice site.<sup>48</sup> Although the m<sup>5</sup>U:G W:W C modified pair has a binding energy of -17 kcal mol<sup>-1</sup>, 24 kcal mol<sup>-1</sup> of extra stabilization is achieved through its interaction with the A58:A84 pair. This clearly indicates the role of higher order structures in imparting extra stabilization to the modified base pair, which further stabilizes the exon:intron interaction during splicing.

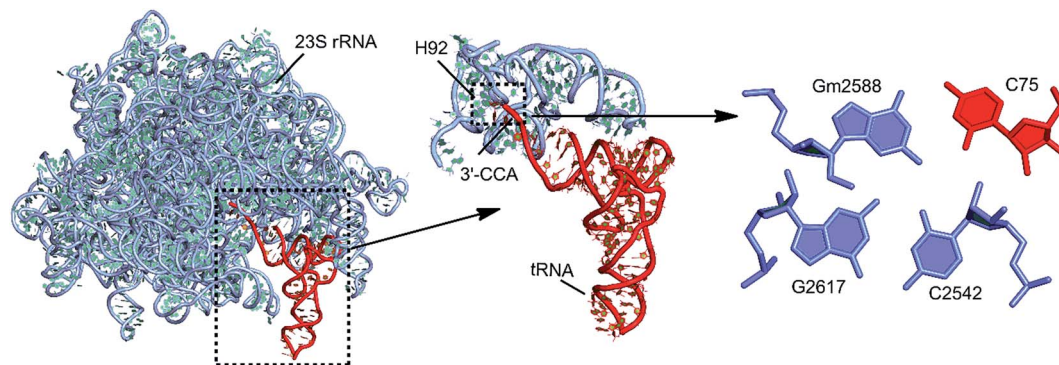


Fig. 8 Modified motifs at the rRNA-tRNA interface, where 3'-CCA end of tRNA interacts with bases near H92 of 23S rRNA (*E. coli*). In *H. marismortui*, 2'-OH methyl modification is observed at G2588.





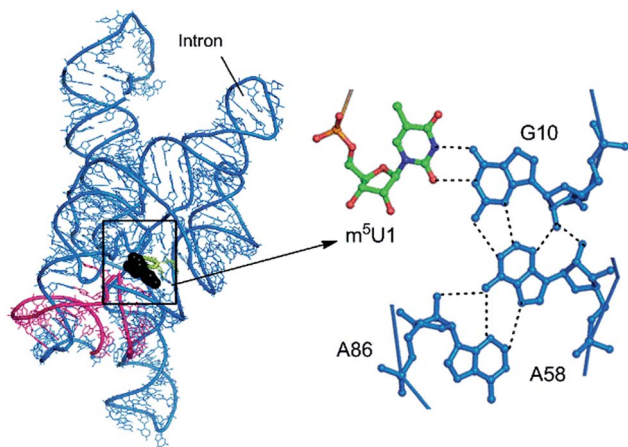


Fig. 10 Modified motif present at the splice site of the intron–exon complex.

## Conclusions

Our crystal structure analysis reveals the importance of higher order associations involving modified base pairs, where ~30% of such base pairs are present as constituents of triples. The significant frequency exhibited by these higher order interactions point towards their important structural and functional roles in RNA macromolecules. Although our previous study reveals that, tRNA contains greater percentage of modified bases and base pairs compared to other RNA classes, the proportion of higher order associations containing modified bases is comparatively greater in 23S rRNA. Further, the fact that the hydrogen bonding pattern within the crystal occurrences of such motifs does not change on optimization, indicates that such motifs are intrinsically stable well-defined building blocks of RNA.

Analysis of stability of higher order modified motifs reveals that those modifications, which introduce positive charge within the motifs, impart greater stabilization compared to other modifications, mainly due to enhanced electrostatic interactions (Table S10<sup>†</sup>). Further analysis of detailed structural environment, around these motifs, also indicates that these motifs are present in structurally and functionally important regions within various RNA structures. This includes the tRNA, tRNA:rRNA interface, rRNA and the intron–exon complex. These findings underscore the need for further investigations towards unravelling the hitherto unknown functional roles of such motifs, in the context of many biological processes involving RNA.

## Conflict of interest

The authors declare no competing financial interest.

## Acknowledgements

PS thanks the Department of Science and Technology (DST) and University Grants Commission (UGC), New Delhi for financial

support through the DST INSPIRE (IFA14-CH162) and the UGC FRP (F.4-5(176-FRP/2015(BSR))) programs, respectively. AM thanks DBT, Government of India project BT/PR-14715/PBD/16/903/2010 for partial financial support.

## References

- 1 P. F. Agris, *EMBO Rep.*, 2008, **9**, 629–635.
- 2 C. S. Chow, T. N. Lamichhane and S. K. Mahto, *ACS Chem. Biol.*, 2007, **2**, 610–619.
- 3 B. El Yacoubi, M. Bailly and V. de Crécy-Lagard, *Annu. Rev. Genet.*, 2012, **46**, 69–95.
- 4 M. Helm, *Nucleic Acids Res.*, 2006, **34**, 721–733.
- 5 Y. Motorin and M. Helm, *Wiley Interdiscip. Rev.: RNA*, 2011, **2**, 611–631.
- 6 S. Raychaudhuri, J. Conrad, B. G. Hall and J. Ofengand, *RNA*, 1998, **4**, 1407–1417.
- 7 J. Černý and P. Hobza, *Phys. Chem. Chem. Phys.*, 2007, **9**, 5291–5303.
- 8 N. B. Leontis, J. Stombaugh and E. Westhof, *Nucleic Acids Res.*, 2002, **30**, 3497–3531.
- 9 N. B. Leontis and E. Westhof, *RNA*, 2001, **7**, 499–512.
- 10 J. Stombaugh, C. L. Zirbel, E. Westhof and N. B. Leontis, *Nucleic Acids Res.*, 2009, **37**, 2294–2312.
- 11 M. Zgarbová, P. Jurecka, P. Banáš, M. Otyepka, J. E. Spöner, N. B. Leontis, C. L. Zirbel and J. I. Šponer, *J. Phys. Chem. A*, 2011, **115**, 11277–11292.
- 12 C. L. Zirbel, J. E. Šponer, J. Šponer, J. Stombaugh and N. B. Leontis, *Nucleic Acids Res.*, 2009, **37**, 4898–4918.
- 13 T. Hermann and E. Westhof, *Chem. Biol.*, 1999, **6**, R335–R343.
- 14 E. Westhof and V. Fritsch, *Struct.*, 2000, **8**, R55–R65.
- 15 P. B. Moore, *Annu. Rev. Biochem.*, 1999, **68**, 287–300.
- 16 A. S. Abu Almakarem, A. I. Petrov, J. Stombaugh, C. L. Zirbel and N. B. Leontis, *Nucleic Acids Res.*, 2012, **40**, 1407–1423.
- 17 S. Bhattacharya, S. Mittal, S. Panigrahi, P. Sharma, S. P. Preethi, R. Paul, S. Halder, A. Halder, D. Bhattacharya and A. Mitra, *Database*, 2015, bav011.
- 18 J. Spöner, J. E. Šponer, A. I. Petrov and N. B. Leontis, *J. Phys. Chem. B*, 2010, **114**, 15723–15741.
- 19 M. Sarver, C. L. Zirbel, J. Stombaugh, A. Mokdad and N. B. Leontis, *J. Math. Biol.*, 2008, **56**, 215–252.
- 20 M. Chawla, S. Abdel-Azeim, R. Oliva and L. Cavallo, *Nucleic Acids Res.*, 2014, **42**, 714–726.
- 21 M. Chawla, P. Sharma, S. Halder, D. Bhattacharya and A. Mitra, *J. Phys. Chem. B*, 2011, **115**, 1469–1484.
- 22 P. Sharma, M. Chawla, S. Sharma and A. Mitra, *RNA*, 2010, **16**, 942–957.
- 23 P. Sharma, A. Mitra, S. Sharma, H. Singh and D. Bhattacharya, *J. Biomol. Struct. Dyn.*, 2008, **25**, 709–732.
- 24 J. Šponer, P. Jurečka and P. Hobza, *J. Am. Chem. Soc.*, 2004, **126**, 10142–10151.
- 25 J. E. Šponer, N. a. Špačková, P. Kulhánek, J. Leszczynski and J. Šponer, *J. Phys. Chem. A*, 2005, **109**, 2292–2301.
- 26 R. Oliva, L. Cavallo and A. Tramontano, *Nucleic Acids Res.*, 2006, **34**, 865–879.



- 27 W. A. Cantara, P. F. Crain, J. Rozenski, J. A. McCloskey, K. A. Harris, X. Zhang, F. A. P. Vendeix, D. Fabris and P. F. Agris, *Nucleic Acids Res.*, 2011, **39**, D195–D201.
- 28 S. Dunin-Horkawicz, A. Czerwoniec, M. J. Gajda, M. Feder, H. Grosjean and J. M. Bujnicki, *Nucleic Acids Res.*, 2006, **34**, D145–D149.
- 29 F. Jühling, M. Mörl, R. K. Hartmann, M. Sprinzl, P. F. Stadler and J. Pütz, *Nucleic Acids Res.*, 2009, **37**, D159–D162.
- 30 D. Piekna-Przybylska, W. A. Decatur and M. J. Fournier, *Nucleic Acids Res.*, 2008, **36**, D178–D183.
- 31 P. P. Seelam, P. Sharma and A. Mitra, *RNA*, 2017, **23**, 847–859.
- 32 M. Chawla, R. Oliva, J. M. Bujnicki and L. Cavallo, *Nucleic Acids Res.*, 2015, **43**, 9573.
- 33 R. Oliva, A. Tramontano and L. Cavallo, *RNA*, 2007, **13**, 1427–1436.
- 34 P. Sharma, J. E. Šponer, J. Šponer, S. Sharma, D. Bhattacharyya and A. Mitra, *J. Phys. Chem. B*, 2010, **114**, 10234.
- 35 S. Mukherjee, M. Bansal and D. Bhattacharyya, *J. Comput.-Aided Mol. Des.*, 2006, **20**, 629–645.
- 36 J. Das, S. Mukherjee, A. Mitra and D. Bhattacharyya, *J. Biomol. Struct. Dyn.*, 2006, **24**, 149–161.
- 37 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
- 38 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785–789.
- 39 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr, J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, Gaussian, Inc., Wallingford, CT, USA, 2009.
- 40 J. Šponer, P. Jurečka and P. Hobza, *J. Am. Chem. Soc.*, 2004, **126**, 10142–10151.
- 41 S. F. Boys and F. Bernardi, *Mol. Phys.*, 1970, **19**, 553–566.
- 42 F. Weigend, M. Häser, H. Patzelt and R. Ahlrichs, *Chem. Phys. Lett.*, 1998, **294**, 143–152.
- 43 A. Mládek, P. Sharma, A. Mitra, D. Bhattacharyya, J. Šponer and J. E. Šponer, *J. Phys. Chem. B*, 2009, **113**, 1743–1755.
- 44 J. Šponer, M. Zgarbová, P. Jurečka, K. E. Riley, J. E. Šponer and P. Hobza, *J. Chem. Theory Comput.*, 2009, **5**, 1166–1179.
- 45 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.
- 46 P. Nissen, J. A. Ippolito, N. Ban, P. B. Moore and T. A. Steitz, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 4899–4903.
- 47 K. Y. Sanbonmatsu, *Biochimie*, 2006, **88**, 1075–1089.
- 48 P. L. Adams, M. R. Stahley, A. B. Kosek, J. Wang and S. A. Strobel, *Nature*, 2004, **430**, 45–50.

