1-12-2023

Structural and Energetic Features of Base–Base Stacking Contacts in RNA

Zakir Ali
Computational Biochemistry Laboratory, Department of Chemistry and Centre for Advanced Studies in Chemistry, Panjab University, Chandigarh

Ambika Goyal
Computational Biochemistry Laboratory, Department of Chemistry and Centre for Advanced Studies in Chemistry, Panjab University, Chandigarh

Ayush Jhunjhunwala
Center for Computational Natural Sciences and Bioinformatics, International Institute of Information Technology, Hyderabad, Gachibowli, Hyderabad, Telangana

Abhijit Mitra
Center for Computational Natural Sciences and Bioinformatics, International Institute of Information Technology, Hyderabad, Gachibowli, Hyderabad, Telangana

John F. Trant
Department of Chemistry and Biochemistry, University of Windsor, 401 Sunset Avenue, Windsor, Ontario N9B 3P4, Canada Binary Star Research Services, LaSalle, Ontario N9J 3X8, Canada

Structural and energetic features of base-base stacking contacts in RNA

Zakir Ali,[1] Ambika Goyal,[1] Ayush Jhunjhunwala,[2] Abhijit Mitra,[2] John F. Trant,[3,4,†]* and Purshotam Sharma[1,3,]*

[1]Computational Biochemistry Laboratory, Department of Chemistry and Centre for Advanced Studies in Chemistry, Panjab University, Chandigarh, 160014, India.

[2]Center for Computational Natural Sciences and Bioinformatics, International Institute of Information Technology, Gachibowli, Hyderabad, Telangana, 500032, India.

[3]Department of Chemistry and Biochemistry, University of Windsor, 401 Sunset Ave. Windsor ON, N9B 3P4, Canada

[4]Binary Star Research Services, LaSalle ON N9J 3X8, Canada

[†]Dr. John F. Trant, a co-corresponding author on this work, is also affiliated to Binary Star Research Services as the Chief Executive Officer. However, Binary Star had no input into the paper and the research is not associated with Binary Star activities.

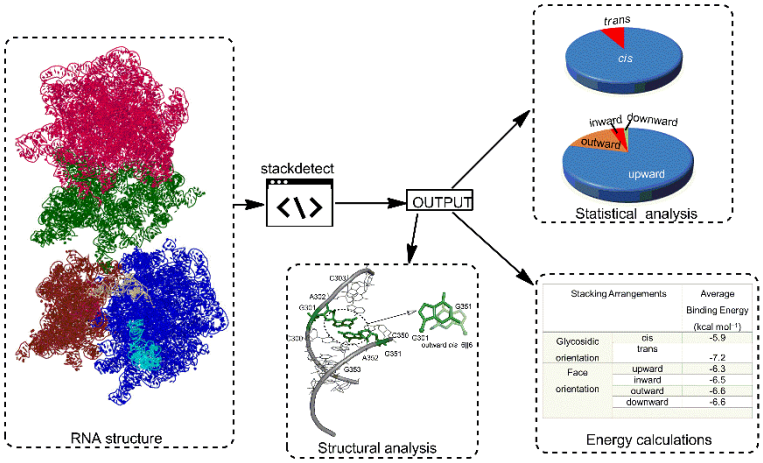
*Email: j.trant@uwindsor.ca, psharma@pu.ac.in

ABSTRACT

Nucleobase π–π stacking is one of the crucial organizing interactions within 3D RNA architectures. Characterizing the structural variability of these contacts in RNA crystal structures will help delineate their subtleties and their role in determining function. This analysis of different stacking geometries found in RNA X-ray crystal structures is the largest such survey to date; coupled with quantum-mechanical calculations on typical representatives of each possible stacking arrangement, we determined the distribution of stacking interaction energies. A total of 1,735,481 stacking contacts, spanning 359 of the 384 theoretically possible distinct stacking geometries, were identified. Our analysis reveals preferential occurrences of specific consecutive stacking arrangements in certain regions of RNA architectures. Quantum chemical calculations suggest that 88 of the 359 contacts possess intrinsically stable stacking

geometries, whereas remaining stacks require the RNA backbone or surrounding macromolecular environment to force their formation and maintain their stability. Our systematic analysis of π–π stacks in RNA highlights trends in the occurrence and localization of these noncovalent interactions and may help better understand the structural intricacies of functional RNA-based molecular architectures.

GRAPHICAL ABSTRACT



INTRODUCTION

Improved structure elucidation techniques have led to an ever-accelerating availability of high-resolution crystal structures of RNA; analysis has likewise evolved from only considering simple 2D base-paired helices joined by single strands, to highly-complex 3D structures supported by a variety of noncovalent interactions. These stems, loops and knots owe their existence to a range of noncovalent interactions.[1] A comprehensive index of these motifs, and how they emerge from specific combinations of inter-base physicochemical forces will not only describe the relationship between RNA structure and function but will also allow for the rational design of functional RNA for non-translational applications.

Higher-order RNA structures emerge from the combination of base-base,[2] base-sugar,[3] base-phosphate,[4] base-ion,[5] sugar-phosphate [6] and RNA-solvent interactions.[7] Base-base contacts are often the most important. Advances in identification,[8] classification,[2b, 4] annotation,[2b] statistics,[2f] and molecular level features, including associated energetics,[2a, b, e, 8b, 9] have ensured that the mechanics of RNA base pairing are well established. However, this information alone is insufficient to explain RNA structure. In addition to traditional inter-strand base pairing, conformationally diverse π–π stacking interaction between bases, often adjacent in the linear sequence, is another driver for RNA self-organization into preferred 3D shapes. Single stranded RNA is stabilized by π–π stacking, but these aromatic interactions also help stabilize higher-order topologies: the four stems of tRNA engage in end-to-end coaxial stacking, facilitating the formation of two long double-stranded helices that define the distinctive L-shaped tertiary structure (Figure 1A).[10] Stacking is also important in inter-loop (kissing-loop) interactions (e.g., stacking of A2513 and G2564 in the 23S RNA of *Haloarcula marismortui*, and of A74 and A291 in bacterial ribonuclease P RNA, Figures 1B and C respectively). Similarly, adenine stacking plays a crucial role in stabilizing the tertiary GC pairings in the kissing complex of hairpins with a (cognate) GACG tetraloop,[11] and guanine-guanine stacks stabilize G-quadraplexes.[12]

Although the physicochemical features of isolated stacked nucleobase dimers have been extensively studied,[13] as have the role of stacking in terms of DNA–protein recognition,[14] RNA–protein recognition,[15] base-backbone lone pair–π interactions,[3] base-water lone pair–π or OH–π bonding,[16] and base-wedged and base-intercalated contacts in RNA,[17] the diversity and occurrence of stacking interactions in RNA structures is not well known.

Summary of the nomenclature

To better organize and classify the RNA stacks, we proposed a systematic base stacking nomenclature and identification scheme[18] built on integrating and extending the existing methodologies.[8a, 19] This unambiguous scheme classifies the base stacks at four distinct levels. At the first level, the stacks are classified into one of the 10 possible distinct nucleobase binary sequence combinations (A||A, G||G, A||G, A||C, A||U, G||C, G||U, C||C, C||U and U||U). The second level classifies each of these combinations according to their relative facial orientation. Specifically, each of the six heterodimeric sequence combinations (A||G, A||C, A||U, G||C, G||U and C||U) involve four possible relative facial orientations (upward, downward, inward, or outward, Figure S1). However, degeneracy between the upward and downward arrangements leaves only three distinct facial arrangements (inward, outward, and upward) for each of the four homodimeric sequence combinations (A||A, C||C, G||G and U||U, Figure S1C). At the third level, each arrangement at the second level is further classified into one of two types based on the relative orientation of the glycosidic bonds of the interacting bases (*cis* and *trans*). At the final level, all stacks are classified according to the interacting ring(s) of the purines (i.e., 5-membered ring, 6-membered ring or both rings, Figure 2 and 3). For a further discussion of how this organizes all possible structures, please see the discussion in the SI. In total there are 384 possible binary stacks that can be further classified as either consecutive (if they involve adjacent base pairs) or non-consecutive (otherwise) to fully describe every possible geometric interaction.

Despite enhanced awareness of the general relevance of π–π contacts in RNA, and the development of computer tools for locating base stacks in RNA 3D structures,[8a, 9h, 19-20] there is no comprehensive categorization of these interactions. In this work, we approach this challenge by applying our unambiguous classification scheme[18] to characterize the π–π interactions in all available high quality X-ray crystal structures containing RNA. Our detailed statistical analysis on 1,735,481 identified stacking contacts enables us to quantify the relative frequencies of different stacks; this analysis is then extended to determine their relative abundance in specific RNA backbone topologies. We then conducted quantum chemical geometry optimizations and stacking interaction energy calculations on a "typical" selected representative example from each stacking geometry. This provides the intrinsic stacking interaction energies and the associated geometric parameters — vertical separation ($\vec{d}_{ab}$), tilt (θ_{ab}), horizontal shift (τ_a or τ_b), and relative glycosidic angle (σ_{ab}), for each possible base stack (Figure 4). The statistical and correlational information, stacking interaction energy values and

optimized stacking geometric parameters, may aid in better understanding of the intricacies of complex RNA structures.

MATERIALS AND METHODS

Occurrence and detection of base stacks

The dataset comprised RNA X-ray crystal structures published on the PDB through 12 December 2020 (https://www.rcsb.org/). The search was conducted by setting the "Polymer Entity Type" to "RNA," providing 5746 hits. This was filtered by setting the "Method" to "X-ray" (3410 hits), and then by restricting the resolution to ≤ 3.5 Å (2999 hits). This, admittedly arbitrary, resolution cut-off was selected to be in agreement with that chosen by Wilson *et al.* in a similar effort,[15b] as well as other studies based on RNA crystal structures.[9a, 21] These hits were then filtered using the CD-HIT suite,[22] to retain only a single representative from clusters of similar structures with >80% similarity, generating a non-degenerate dataset of 2114 structures (Table S1). The data is inherently biased by what is selected and deposited into the database, but this filtering removes effective duplicates that might further sway the data away from a population estimate. Further, to test the robustness of the derived results, the analysis was repeated on structures selected from the 2114 structures after restricting the refinement resolution to ≤ 2.5 Å (i.e., 743 structures, Table S2).

We processed these PDB files using our 'StackDetect' tool to detect and classify stacks, available on GitHub (https://github.com/PSCPU/Stack_detect),[18] using the search criteria adapted from Gaab and coworkers[23] ($\vec{d}_{ab}$ < 4.5 Å, θ_{ab} < 23°, and τ_a or τ_b < 40°). Although our search for stacking interactions was restricted to these well-established criteria, the search space can be easily extended by increasing the values of $\vec{d}_{ab}$, θ_{ab} and τ_a or τ_b to include additional significantly weaker dispersive stacking forces that originate from the inherently flat nature of the stacking potential energy surface.[13] For further analysis, the output files obtained from running the program for each separate pdb file were consolidated into a single file using in-house python code, accessible through GitHub (https://github.com /PSCPU/Stack_detect).

Stacking analysis in RNA crystal structures

To complement the statistical analysis, we analyzed of the structural role of base stacks present in select functionally-significant RNA crystal structures. These include the structures of 67 riboswitches, spanning most of the riboswitch categories described by Sherwood and Henkin,[24] also examined in our previous study (Table S3).[25] We also analysed specific stacks in the crystal structure of the 16S rRNA of *Thermus thermophilus* (PDB code: *1n33*).[26] We further analyzed stacking interactions of bases involved in structurally-important A-minor motifs and

higher order structures (base triplets, quartets and pentets) identified in the large ribosomal subunit of *Haloarcula marismortui* (PDB code: *1jj2*),[27] using the DSSR program[28], as well as the functionally-important inter-RNA stacking contacts present in the crystal structure of the bacterial ribosome containing rRNAs, tRNAs and a portion of messenger RNA (PDB *4w2i*)[29]. Structures were visualized using Pymol.[30]

QM calculations

For the QM calculations, one example was initially chosen for each stacking geometry present in the dataset. In synchrony with previous studies,[4, 9e, 31] the models for geometry optimization were generated by replacing the ribose moiety with a hydrogen atom. An alternative to be considered would be the replacement of the C1' of the ribose sugar with a methyl group. The decision to use hydrogen was based on precedent for alignment with other work in the field. Any substitution has an impact on the specifics of a molecular interaction, but as the focus of our calculations is to properly evaluate of the strength of the base:base stacking interactions, we prioritized the comparison and itnegration of extant data sets. Geometry optimizations were carried out using ωB97X-D/6-31G(d,p), by employing the default D2 version of the disperison correction available in Gaussian 16 Rev. C.01.[32]

For stacked dimers that did not retain the face-on ring–ring geometry on full optimization, the models were prepared by replacing the crystal-structure coordinates of each base (monomer) with their ωB97X-D/6-31G(d,p) optimized geometries, while retaining the relative orientation of the bases within the stack. These non-optimized crystallographic examples were chosen by first sorting the stacks of a particular arrangement with respect to the vertical separation ($\vec{d}_{ab}$ > 3 Å) between the bases (Figure 4). The final representative example was chosen from the resulting top twenty structures, by sorting them with respect to the sum of θ_{ab}, τ_a and τ_b (Figure 4).

The intrinsic stability of each base stack was calculated in terms of stacking interaction energy, defined as the stabilization acquired by the two bases when they form a stack. The stacking interaction energies were calculated at the ωB97X-D/6-311+G(2df,p) level. The stacking interaction energies were corrected for basis set superposition error (BSSE) using the standard counterpoise procedure.[33] QM calculations were carried out using the Gaussian 16 (Rev. C.01) suite of programs.[32]

RESULTS

1. Statistical analysis of stacking interactions in RNA crystal structures

Analysis of the dataset of 2114 RNA crystal structures reveals 1,735,481 stacks. These stacks were grouped, based on their base combinational, geometrical, and positional characteristics, and were statistically analysed to discover correlations within and across these groupings (Tables S4 – S73).

Stacking geometries differ in consecutive, non-consecutive and inter RNA stacks

Consecutive stacks. Approximately three-quarters of the total stacks are consecutive (Table 1), and almost all (99.7%) consecutive stacks adopt a *cis* orientation (Table S74), mainly since most of these stacks occur in RNA motifs that adopt a uniform backbone arrangement, (*e.g.*, stems or helical regions) which restricts them to a *cis* orientation (Figure 5).[18] However, the 0.3% *trans* oriented stacks (Table S74) occur where the strand is abruptly bent or folded (Figure 6). In terms of the face orientation, 97.7% stacks are upward, which is the orientation adopted in regular strands (*vide supra*, Figure S2). Ring arrangement does affect the likelihood of a consecutive stack, with 6||5 and 6||6 stacks contributing the most (26.5% each), followed by 5||6,6|||6 (18.5%) and 5||5,6||56 (14.6%, Supplemental Table S75).

'Upward *cis*' is the most frequent orientation in a regular strand (Figure 5); 'inward *cis*' and 'inward *trans*' are usually present in regions that involve a sharp 'V' type turn in RNA (Figure 6). In contrast, other stacking classes, including 'outward *cis*', 'outward *trans*', 'upward *trans*' and 'downward *cis*' occur where slight irregularities exist in the strands, including the translation of the flanking bases on one or both sides of the stack from their regular positions, and in certain cases, straightening of the local helix (Figures 6 and S3).

Non-consecutive stacks. Only 66.4% of the non-consecutive stacks are *cis* (Figure S2 and Table S76). The remarkably enhanced presence of the *trans* orientation in these stacks is because they do not necessarily occur in regular double-stranded regions, and therefore evade the strict requirements that demand *cis* orientation. The outward face orientation, usually occurring between bases present on opposite strands in stem regions, is dominant (62.4%) in these stacks (Figure S2). Furthermore, because the constituent bases are predominantly located on the opposite sides of base-paired stems, the 6||6 arrangement has the highest occurrence (51.0%) in non-consecutive stacks (Table S75).

Inter-RNA stacks. Such stacks are rare, comprising only 0.4% of the total stacks (Table S74). They require multiple RNA chains to be present, and only occur in complex RNA architectures, like the ribosome. However, when present, these stacks play an important role in folding different RNA chains in a single structure, like how key amide hydrogen-bond interactions in proteins define the quaternary structure. Most such stacks are *cis* and around half (50.3%) are

outward oriented (Figure S2 and Table S76). 6||6 arrangement is the most prominent orientation in these stacks (48.1%, Table S75). Due to the involvement of two RNA chains, the backbone of the interacting nucleotides exhibits variable geometrical characteristics (Figure S5), which may depend on the identities of the interacting RNA. Therefore, like the non-consecutive stacks, no straightforward relationship between stacking and local RNA backbone is observed within the most prominent (i.e., 'outward' and '*cis*') face and glycosidic orientations respectively (Tables S74 and S78).

Overall, *cis* orientation is the overwhelmingly dominant glycosidic orientation for all stacks (91.6%, Figure 7). A majority (83.1%) of the *cis* stacks occur in the upward face orientation. The dominance of the upward, *cis* orientation is unsurprising as it is the only possible intrastrand stacking contact in regular stems, the most common RNA structural motif (Figure 5). Further, the proportions of *cis* and 'upward' stacks decrease with increase in purine content within a stack (Figure 8B and Tables S79 – S82). Since both '*cis*' and 'upward' orientations are characteristic of stacks present stem regions of RNA (Figure 5),[18] this suggests that purine stacking might be more important in irregular (i.e., nonhelical) RNA regions.

Most stacks generally involve purines, and guanine in particular; significant occurrence bias exists in favour of G and C over A and U respectively

Purine||purine stacks, which account for 48.2% of the total, mainly involve guanine (80% Figure 8 .and Tables S79 – S83) Further, guanine (66.2%) is again overrepresented in purine||pyrimidine stacks, which themselves account for 39.8% of the examples (Figure 7). Guanine is generally overrepresented: 66.2% of all stacks containing guanine (Figure 8 and Table S84). Thus, 64.5% stacks contain at least one guanine, although the total percentage of guanines in crystal structures is only 34%. In contrast, pyrimidine||pyrimidine stacks make up only 12.0% stacks (Figure 8 and Tables S79 – S82). Cytosine is present in 86% of these stacks (Table S85), although the percentage of cytosine in crystal structures is only 26%. Overall, although the dominance of purine||purine stacks points towards a crude relationship between the size of the nucleobase ring skeleton and its stacking ability, the dominance of guanine over adenine and of cytosine over uracil in base stacks correlates with the high dipole moments of guanine and cytosine, which may favourably align with diploe moment of the other base of the stack to form strong contacts.[34]

Occurrence of ring–ring stacking geometries do not necessarily correlate with the extent of stacking and have variable face characteristics

Purine||purine stacks can be divided into up to 15 distinct geometries, while purine||pyrimidine stacks involve 3 unique geometries (Figures 2, 3 and Table S86).[18] Within the purine||purine

stacks, the 5||5,6||56 arrangement, which possesses a substantial overlap of the five-membered and six-membered rings of both purines (Figure 3), and which shows a significant average stacking interaction energy (–8.8 kcal mol^{-1}), is dominant (24.6%, Tables S87 and S88). However, the 5||6,6||5 arrangement, which involves stacking between both 5-membered and 6-membered purine rings, has a negligible (0.1%, Table S88) contribution, which correlates with its comparably low average stacking interaction energy (–2.9 kcal mol^{-1}, Table S87). Similarly, among the three purine||pyrimidine arrangements, the 5||6,6||6 arrangement, that involves the greatest stacking and strongest average interaction energy (–10.3 kcal mol^{-1}, Table S87), is the most prominent (43.5%, Table S89). However, surprisingly, despite relatively small stacking extent and consequent moderate average stacking interaction energy (–4.6 kcal mol^{-1} and –5.6 kcal mol^{-1}, Table S87), 6||5 (24.0%) and 6||6 (17.8%, Table S88) arrangements possess the second and third most significant occurrences within the purine||purine stacks. This unexpected high population of weakly interacting arrangements in purine||purine stacks means that the stacking interaction energy does not necessarily correlate directly with occupancy of the arrangement. As expected, a combination of entropic factors and local RNA backbone forces too are likely responsible for positioning the bases relative to one another.

Statistical distribution of stacks remains unchanged on further refinement of the structural dataset

To further solidify our statistical analysis, we analysed the effect of the resolution cut-off on the derived results by selecting X-ray structures with only up 2.5 Å resolution (Table S2). The resulting dataset of 743 high resolution structures excluded a considerable number of structures containing ribosomal subunits, since most of them were crystallized at a relatively lower resolution. However, the statistical analysis on the refined dataset produced very similar results to the complete dataset (Tables S90 – S96), with only slightly increased preferences towards the *cis* glycosidic orientation (by about 0.4%, Table S90), consecutive arrangement (by about 0.8%, Table S91), and upward face orientation (by about 0.5%, Table S92). Further, the preferred ring arrangement in purine||purine stacks (5||5,6||56) and purine||pyrimidine stacks (5||6,6||6) also remains same in both datasets (Tables S93 and S94). Furthermore, the relative preference of bases involved in stacks remains similar in both datasets (Tables S95 and S96). As taking different subsets of the dataset returns the same results, this suggests that these ratios of the different stacks are likely representative of their frequency in the population.

2. Analysis of stacking trends in specific functionally-important RNA classes

Aptamer domains of the riboswitches

Riboswitches are structured RNA elements usually located at the 5' end of untranslated mRNA. These play a vital role in gene expression regulation by sensing and binding specific small metabolites.[35] Formed from two distinct regions (the 'ligand-binding' aptamer domain and the expression platform), riboswitches regulate the gene expression through binding of a specific ligand to the aptamer domain, which triggers structural changes in the expression platform that regulates the gene expression. Owing to their importance, several crystal structures of the aptamer domains of riboswitches have been published in literature over the last two decades, which provides wealth of data to understand the specific stacking features in this specific class of structured RNA. An analysis of stacking interactions in the aptamer domains of riboswitches may aid in better understanding about the associated structural intricacies.

A total of 5056 stacks were detected in the 67 PDB entries containing riboswitch aptamer domains. 80.7% of these stacks are consecutive (Table S2), and most stacks are *cis* oriented (95%) and upward oriented (82%, Figure S6 and Table S97). Thus, compared to the total dataset, an increased prominence of consecutive, *cis*, and upward stacks is observed in riboswitches. This in turn points toward the similarities in stacking patterns within the structures of the aptamer domains of different riboswitches.

To structurally understand riboswitches' deviation from the mean stacking population, we examined the TPP binding (Thi Box) riboswitch of *E. coli* (PDB *2hom*, Figure 9). The most-common 'upward *cis'* stacks occur in consecutive bases of a single strand (Figure 9). "Upward *cis*" also commonly occurs between consecutive bases in a regular strand (A9||C10 and A80||G81 stacks, Figure 9). However, stacks between bases belonging to opposing strands of helical regions are commonly "outward *cis*" (G16||G51 and G65||A75, Figure 9). In contrast, the less–frequent stacking classes such as "outward *trans"* (A61||C63, Figure 9), "inward *cis/trans*" (U20||A44 and G60||61, Figure 9) and "downward *cis/trans*" (U52||A53 and G21||A43, Figure 9) occur at the folds and at places where the loops meet. Riboswitches are far richer in these motifs than other RNA; an increased appearance of minor stacks suggests architectural complexity, and likely specialized function.

16S rRNA of the 30S ribosomal subunit of Thermus thermophilus

We also analyzed the stacking interactions in the crystal structure of the 30S ribosomal subunit of *Thermus thermophilus* (Figure 10). As in the complete dataset, a majority of the stacks are *cis* (92.9%), consecutive (74.9%), upward (79.0%), and a plurality are 6||6 (34.4%, Figure S7 and Table S98). These statistical trends are in line with the overall dataset: unlike riboswitches, ribosomes do not differ statistically from the conformational preferences of RNA as a whole. The resemblence of the statistical results of the analyzed 30S ribosomal subunit with that of

total dataset is due the dominance of ribosomal structures in the dataset (i.e., 96.6% of stacking interactions occur in different ribosomal structures, Table S99). Select stacking arrangements in the 16S rRNA structure are illustrated in Figure 10 to aid in visual understanding of how stacking contacts shape the 3D RNA structures.

Stacking around A-minor motifs and base multiplets

Although the analysis was developed with binary pairs in mind, the strategy, and the software, can be readily applied to higher order stacks. To demonstrate this potential, and to explore potentially more important structures, we investigated A-minor motifs, base-triplets, quartets and pentets.

A-minor motifs are structurally and functionally important hydrogen-bonded RNA tertiary contacts, first discovered in the large ribosomal subunit of the archeon *Haloarcula marismortui*.[36] To understand the stacking characteristics of the bases constituting the A-minor motifs, we carried out a detailed analysis of a high (2.4 Å) resolution crystal structure of the large ribosomal subunit of *Haloarcula marismortui* (PDB code: *1jj2*). The stacking interactions were analyzed for a total of 156 A-minor motifs (49 of Type 0, 71 of Type I and 36 of Type II, Figure 11)

Similar to the complete dataset, all three bases (minor groove adenine, and the Watson-Crick base pair) constituting each type of A-minor motif (Type 0, I and II) predominantly adopt a *cis* orientation (93.2%, (Figure S8A)), although the minor-groove adenine exhibits increased preference for the *trans* glycosidic orientation (25%, 19% and 15.2% for Type 0, I and II A-minor motifs respectively), compared to the two bases constituting the Watson-Crick pair (up to 7.3%, Figure 11 and Table S100). Furthermore, the minor-groove adenine exhibits reduced preference towards the upward face orientation, compared to the Watson-Crick base pair with the A-minor motif (Table S101). This correlates with the fact that the minor-groove adenine generally participates in tertiary contacts, and unlike the Watson-Crick pair, is not a part of a regular stem region and consequently need not observe strict backbone conformation requirements. Moreover, although all three bases constituting the A-minor motifs are generally involved in stacking with the consecutive bases, the minor-groove adenine in Type 0 and Type I motifs has a relatively greater involvement of non-consecutive stacking (Table S102). Nevertheless, like the total dataset, 6||6 is the most prominent stacking arrangement for the bases involved in A-minor motifs (30.4%, Table S103).

We further extended our analysis to study the stacking interactions of the bases involved in other higher-order hydrogen-bonded structures (*i.e.*, base-triplets, quartets and pentets) within the large ribosomal subunit (PDB code: *1jj2*). Specifically, we analysed stacking

interactions of bases involved in a total of 252 such higher-order structures, which includes 180 base triplets, 66 quartets and 6 pentets. The stacks involving in these higher-order motifs are mostly *cis*-oriented (86.5%, Figure S8B), although slightly reduced *cis* preference is observed compared to the complete dataset. However, the preference for the *cis* orientation slightly increases with increase in the number of bases within the multiplet – 84.8% for triplets, 89.2% for quartets and 92.5% for pentets (Table 104). In terms of the face orientation, stacks involving the bases of quartets contribute most to the upward face orientation (76.3%), compared to those of triplets (71.8%) and pentets (65.0%, Table S105). Further, the stacks involving quartets are more prominently consecutive (71.6%), compared to those in triplets (67.4%) and pentets (60%, Table S106). This points to the dependence of the stacking pattern on the number of hydrogen-bonded bases within a multiplet. Regardless, like the distribution the complete dataset, as well as the A-minor motifs, 6||6 is most dominant ring arrangement in the multiplets (30.6%) (Table S107).

Analysis of Inter-RNA interactions illustrates how different base stacks play a role in the translation process

As mentioned earlier, inter-RNA stacks contribute to only a small fraction (0.4%) of the total stacking interactions in the whole dataset (Table S74). However, our analysis reveals that such stacks are vital for the 3D arrangement of different RNA chains in large RNA structures such as the ribosomal subunits (Figure S9). Interestingly almost half of these inter-RNA stackings are located in the ribosomal structures (Table S108). For example, analysis of the structure of *Escherichia coli* ribosomal complex (PDB code: *4w2i*) containing rRNAs, tRNAs and mRNA reveals the identity of base stacks in the interaction between mRNA and tRNA positioned on the ribosome (Figure 12). Specifically, an intrinsically strong (–9.3 kcal mol^{-1}, Table S109) "outward *trans* 5||5,6||56" pattern of stacking occurs between A14 of mRNA and G34 of the anticodon loop of tRNA present at the ribosomal A site, which may help in binding of tRNA present on the ribosomal A site with mRNA, and therefore, may help in the mRNA decoding process. Further, a strong (–5.7 kcal mol^{-1}, Table S109) upward *trans* 6||56 stacking occurs between the 76^{th} base of the acceptor stem of tRNA at A site with the 2434^{th} base of 23S rRNA (Figure 12), that may aid in establishing the optimal tRNA:rRNA contact for facilitating the translation process.

In addition, analysis of the rRNA:mRNA:tRNA complexes involving the P site tRNA reveals two important mRNA:tRNA stacking contacts. The first contact occurs between A16 of mRNA and A37 of the anticodon loop of tRNA in an intrinsically stable (–9.7 kcal mol^{-1})

outward *cis* 6||6 orientation (Table S109). However, the second contact involves A18 of mRNA and G35 of the anticodon loop of tRNA in an intrinsically unfavourable (0.3 kcal mol^{-1}) outward d *cis* 6||6 orientation (Table S109). Interestingly, no stacking occurs between the bases of mRNA and tRNA at the ribosomal E site. Altogether, these examples illustrate how diverse stacking geometries participate in RNA–mediated translation.

3. Quantum chemical calculations on different stacking orientations

We carried out quantum chemical calculations to evaluate the strength of stacking interactions. As expected, the 88 full-optimized stacking arrangements have better (more negative) interaction energy (–20.8 kcal mol^{-1} to –6.6 kcal mol^{-1}) than the 271 models drawn directly from the crystal structures (–10.8 kcal mol^{-1} to 3.6 kcal mol^{-1}, Tables S109 – S111 and Figures S10 – S22). The fact that those 271 stacks optimize to either one of the 88, or change to a planar base-pairing arrangement, demonstrates that these are not intrinsically stable forms, but rather are prevented from adopting a preferred orientation due to positioning induced by the local RNA environment. This effect is less pronounced for the stronger interactions which already dominate: stacks with participation of both rings of the purine skeleton generally retain the ring arrangement on optimization and exhibit stronger interaction than the other stacks (Tables S109 – S111 and Figure 13).

The average stacking interaction energies of purine||purine stacks (–6.3 kcal mol^{-1}, Table S95 and Figures S10 – S17) are like those of purine||pyrimidine stacks (–7.4 kcal mol^{-1}, Table S96 and Figures S18 – S21). However, due to a relatively smaller possibility of overlap, the pyrimidine||pyrimidine stacks bind relatively weakly compared to purine||purine and purine||pyrimidine stacks (–4.6 kcal mol^{-1}, Table S111). Alternatively, when averaged over the glycosidic orientations of all stacks, *trans* stacks are slightly more stable than their *cis* counterparts (by 1.5 kcal mol^{-1}, Table S112). This contrasts with the observed dominance of *cis* oriented stacks in the crystal structures (Figure 7) and suggests that although less favourable than *trans* stacks in isolation, the regular helical backbone topology preferentially induces *cis* stacks in RNA structures. Similar results were observed on averaging the energies over face orientations. Specifically, the upward face-oriented stacks which make about 79% of the total stacks (Table 1) are comparable in average stability to the downward stacks (within 0.3 kcal mol^{-1}, Table S112), which contribute only 1.4% (Table 2). However, averaging the interaction energies over different ring arrangements reveal that with the exception to two arrangements — 5||6,6||5 and 5||5,6||6 that are not possible in certain glycosidic and face combinations — the stacking interaction energies directly correlate with the extent of stacking — highest in case of

5||56,6||56 (–9.0 kcal mol^{-1}) and lowest for 5||5 (–4.3 kcal mol^{-1}, Table S112). In all cases, the archetypal stacking is favourable; highlighting the unusual unfavourable interaction we found in the ribosome.

DISCUSSION AND CONCLUSIONS

Base-base stacking is one of the forces responsible for the folding of complex 3D architectures of RNA. To analyse the structural significance of these contacts, we performed a detailed statistical, energetic, and structural context analysis of base stacks in the library of higher resolution RNA crystal structures available in the PDB, using our recently developed classification scheme for RNA stacking[18]. Our study reveals a correlation between the type of stacking interaction in a given region of the RNA and the local RNA fold. For example, consecutive "upward *cis*" stacks mostly occur in regular strands, whereas "outward *cis*" arrangements occur when the stacked bases are present on opposite strands in a helix; sharp "V" or "Z" type folds involve the presence of an "inward *cis"* or "inward *trans*" stack. Based on these correlations between the stacking pattern and the folding pattern of a particular region of the RNA structure, significant predictions can be made regarding the structural complexity of a specific RNA region. For example, the extent of stacking between two opposite running strands can be predicted from the dominance of "outward" face-oriented stacks, In a similar way, a "consecutive inward or outward" stack in the structure will aid in precisely locating a sharp turn or fold in a regular running strand, and vice versa. Thus, the valuable relationships between backbone topologies and base stacking characteristics derived from our analysis can be used to digitize RNA folding and may help improve RNA structure prediction and refinement algorithms.

We further highlight the dependence of geometric characteristics of stacks on the relative position of the interacting bases in the sequence space. Consecutive stacks are mostly upward and *cis* oriented, but non-consecutive stacks generally adopt outward and *trans* orientations. More importantly, our analysis reveals that the rare inter-RNA stacks are overwhelmingly *cis* oriented, with half of them oriented in the outward conformation. Altogether, these conclusions provide important insights for understanding the complexities of RNA structures. Furthermore, the structural principles revealed from our analysis can be used to assign correct geometries to the base stacks predicted using the available RNA structure prediction methods and may thus help test and improve the RNA structure prediction and folding algorithms.

Statistical analysis of base stacks at the level of base sequence combinations reveals the dominance of guanine over adenine and of cytosine over uracil in base stacks, which correlates with the high dipole moments of guanine and cytosine.[34] A relative "upward face" and *cis* glycosidic orientation is the most common form, which correlates with their dominant occurrence in the most common RNA motifs (i.e., helical regions or stems). Furthermore, the dominant "*cis*" and "upward" orientations become less common with increasing purine content, pointing to their abundance in irregular RNA regions. However, the higher-than-expected occurrence of weakly-interacting arrangements in purine||purine stacks suggests that RNA backbone forces might be responsible for stacking in arrangements that involve such weak π–π contacts.

We further examined the prevalence of stacks in selected functionally important RNA structures such as riboswitches and rRNA. Detailed analysis of the distribution of stacking contacts within the aptamer domains of riboswitches may aid in understanding their structural intricacies, and thereby help design better synthetic aptamers with improved ligand-binding characteristics. The analysis of the stacking contacts at the tRNA:mRNA:rRNA interface highlights the role that diverse stacking geometries play in facilitating the appropriate architecture required for RNA–mediated translation. More importantly, analysis of stacking around the structurally important A-minor motifs reveals differences in the stacking pattern around both the minor-groove adenine, and the Watson-Crick pair, that constitute these motifs. Similarly, analysis of stacking contacts in base multiplets points to the dependence of the stacking pattern on the number of hydrogen-bonded bases comprising the multiplets. Altogether, these observations on stacking contacts within specific RNA structures and recurring motifs may help design improved RNA structure prediction methods. They certainly need to be included in any tool that seeks to predict structure.

QM calculations were carried out on all possible stacking arrangements to augment our analysis. These calculations reveal the optimal geometrical parameters for the 88 full optimized stacking geometries, which can be useful for generating stable stacking contacts in RNA 3D structures, and may, therefore, help generate high quality models of specific RNA motifs and noncanonical RNA oligonucleotides. The QM optimized structures and interaction energies of different stacking geometries may help in choosing sequences for forming structures stabilized through stacking interactions, and may thus be useful for improving the stability of synthetic mRNA constructs for potential therapeutic applications.

Overall, this work characterizes the intrinsic structural aspects of base-base stacking interactions. When coupled with the recent advances in understanding other noncovalent forces within RNA, including base pairing,[2f] base-phosphate interactions[4] and other RNA interactions involving the π-clouds of nucleobases,[3, 16-17] our study may help in improving our understanding of the intricacies of RNA structures. In future, our classification and analysis scheme can be extended to detect and categorize diverse stacking interactions between nucleobases and amino acids in RNA:protein complexes, and between nucleobases and ligands in RNA:ligand complexes.

FUNDING

The research project was supported by Department of Science and Technology (DST), University Grants Commission (UGC). P.S. 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. JFT thanks the Natural Sciences and Engineering Research Council of Canada (2018-06338). Z.A. thanks UGC for a Senior Research Fellowship.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org

AUTHOR INFORMATION

Corresponding Authors

Purshotam Sharma – *Computational Biochemistry Laboratory, Department of Chemistry and Centre for Advanced Studies in Chemistry, Panjab University, Chandigarh, 160014, India*; orcid.org/0000-0002-2561-7117; Email: psharma@pu.ac.in

John F. Trant – University *of Windsor, 401 Sunset Avenue Windsor, Ontario, Canada N9B 3P4.* Email: j.trant@uwindsor.ca

Authors

Zakir Ali – *Computational Biochemistry Laboratory, Department of Chemistry and Centre for Advanced Studies in Chemistry, Panjab University, Chandigarh, 160014, India;* orcid.org/0000-0001-9597-4737

Ambika Goyal – *Computational Biochemistry Laboratory, Department of Chemistry and Centre for Advanced Studies in Chemistry, Panjab University, Chandigarh, 160014, India.* Email: ambikagoyal1511@gmail.com

Ayush Jhunjhunwala – *Center for Computational Natural Sciences and Bioinformatics, International Institute of Information Technology, Gachibowli, Hyderabad, Telangana, 500032; India;* orcid.org/0000-0002-3507-7193, Email: ayushjjwala94@gmail.com

Abhijit Mitra – *Center for Computational Natural Sciences and Bioinformatics, International Institute of Information Technology, Gachibowli, Hyderabad, Telangana, 500032, India*; orcid.org/ 0000-0002-6876-4838 Email: abi_chem@iiit.ac.in

REFERENCES

1. (a) Saenger, W., Principles of nucleic acid structure. **1984**; (b) Pleij, C. W., Pseudoknots: a new motif in the RNA game. *Trends in biochemical sciences* **1990,** *15* (4), 143-147.

2. (a) Halder, S.; Bhattacharyya, D., RNA structure and dynamics: a base pairing perspective. *Progress in Biophysics and Molecular Biology* **2013,** *113* (2), 264-283; (b) Leontis, N. B.; Westhof, E., Geometric nomenclature and classification of RNA base pairs. *RNA* **2001,** *7* (4), 499-512; (c) Mladek, A.; Sharma, P.; Mitra, A.; Bhattacharyya, D.; Šponer, J. í.; Šponer, J. E., Trans Hoogsteen/sugar edge base pairing in RNA. Structures, energies, and stabilities from quantum chemical calculations. *The Journal of Physical Chemistry B* **2009,** *113* (6), 1743-1755; (d) Šponer, J.; Leszczyński, J.; Hobza, P., Nature of nucleic acid− base stacking: nonempirical ab initio and empirical potential characterization of 10 stacked base dimers. Comparison of stacked and H-bonded base pairs. *The Journal of Physical Chemistry* **1996,** *100* (13), 5590-5596; (e) Šponer, J. E.; Špačková, N. a.; Leszczynski, J.; Šponer, J., Principles of RNA base pairing: structures and energies of the trans Watson− Crick/sugar edge base pairs. *The Journal of Physical Chemistry B* **2005,** *109* (22), 11399-11410; (f) Stombaugh, J.; Zirbel, C. L.; Westhof, E.; Leontis, N. B., Frequency and isostericity of RNA base pairs. *Nucleic Acids Research* **2009,** *37* (7), 2294-2312.

3. Chawla, M.; Chermak, E.; Zhang, Q.; Bujnicki, J. M.; Oliva, R.; Cavallo, L., Occurrence and stability of lone pair–π stacking interactions between ribose and nucleobases in functional RNAs. *Nucleic Acids Research* **2017**.

4. Zirbel, C. L.; Šponer, J. E.; Šponer, J.; Stombaugh, J.; Leontis, N. B., Classification and energetics of the base-phosphate interactions in RNA. *Nucleic Acids Research* **2009,** *37* (15), 4898-4918.

5. (a) Draper, D. E., A guide to ions and RNA structure. *RNA* **2004,** *10* (3), 335-343; (b) Kolev, S. K.; Petkov, P. S.; Rangelov, M. A.; Trifonov, D. V.; Milenov, T. I.; Vayssilov, G. N., Interaction of Na+, K+, Mg 2+ and Ca 2+ counter cations with RNA. *Metallomics* **2018,** *10* (5), 659-678; (c) Misra, V. K.; Draper, D. E., On the role of magnesium ions in RNA stability. *Biopolymers: Original Research on Biomolecules* **1998,** *48* (2-3), 113-135.

6. (a) Ulyanov, N. B.; James, T. L., RNA structural motifs that entail hydrogen bonds involving sugar–phosphate backbone atoms of RNA. *New Journal of Chemistry* **2010,** *34* (5), 910-917; (b) Šponer, J.; Mládek, A.; Šponer, J. E.; Svozil, D.; Zgarbová, M.; Banáš, P.; Jurečka, P.; Otyepka, M., The DNA and RNA sugar–phosphate backbone emerges as the key player. An overview of quantum-chemical, structural biology and simulation studies. *Physical Chemistry Chemical Physics* **2012,** *14* (44), 15257-15277.

7. (a) Fingerhut, B. P., The mutual interactions of RNA, counterions and water–quantifying the electrostatics at the phosphate–water interface. *Chemical Communications* **2021,** *57* (96), 12880-12897; (b) Likhtenshtein, G. I., Nucleic Acids Hydration. In *Biological Water*, Springer: 2021; pp 371-405.

8. (a) Sarver, M.; Zirbel, C. L.; Stombaugh, J.; Mokdad, A.; Leontis, N. B., FR3D: finding local and composite recurrent structural motifs in RNA 3D structures. *Journal of mathematical biology* **2008,** *56* (1-2), 215-252; (b) Das, J.; Mukherjee, S.; Mitra, A.; Bhattacharyya, D., Non-canonical base pairs and higher order structures in nucleic acids: crystal structure database analysis. *Journal of Biomolecular Structure and Dynamics* **2006,** *24* (2), 149-161.

9. (a) Chawla, M.; Oliva, R.; Bujnicki, J. M.; Cavallo, L., An atlas of RNA base pairs involving modified nucleobases with optimal geometries and accurate energies. *Nucleic Acids Research* **2015,** *43* (14), 6714-6729; (b) Seelam, P. P.; Sharma, P.; Mitra, A., Structural landscape of base pairs containing post-transcriptional modifications in RNA. *RNA* **2017,** *23* (6), 847-859; (c) Šponer, J. E.; Leszczynski, J.; Sychrovský, V.; Šponer, J., Sugar edge/sugar edge base pairs in RNA: stabilities and structures from quantum chemical calculations. *The Journal of Physical Chemistry B* **2005,** *109* (39), 18680-18689; (d) Lemieux, S.; Major, F., RNA canonical and non-canonical base pairing types: a recognition method and complete repertoire. *Nucleic Acids Research* **2002,** *30* (19), 4250-4263; (e) Sharma, P.; Mitra, A.; Sharma, S.; Singh, H.; Bhattacharyya, D., Quantum chemical studies of structures and binding in noncanonical RNA base pairs: the trans Watson-Crick: Watson-Crick family. *Journal of Biomolecular Structure and Dynamics* **2008,** *25* (6), 709-732; (f) Šponer, J. E.; Špačková, N. a.; Kulhánek, P.; Leszczynski, J.; Šponer, J., Non-Watson− Crick base pairing in RNA. quantum chemical analysis of the cis Watson− Crick/sugar edge base pair family. *The Journal of Physical Chemistry A* **2005,** *109* (10),

2292-2301; (g) Sharma, P.; Chawla, M.; Sharma, S.; Mitra, A., On the role of Hoogsteen: Hoogsteen interactions in RNA: ab initio investigations of structures and energies. *RNA* **2010,** *16* (5), 942-957; (h) Sharma, P.; Sponer, J. E.; Sponer, J.; Sharma, S.; Bhattacharyya, D.; Mitra, A., On the Role of the cis Hoogsteen: Sugar-Edge Family of Base Pairs in Platforms and Triplets · Quantum Chemical Insights into RNA Structural Biology. *The Journal of Physical Chemistry B* **2010,** *114* (9), 3307-3320.

10. (a) Kim, S.; Suddath, F.; Quigley, G.; McPherson, A.; Sussman, J.; Wang, A.; Seeman, N.; Rich, A., Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. *Science* **1974,** *185* (4149), 435-440; (b) Byron, K.; Wang, J. T.; Wen, D. In *Genome-wide search for coaxial helical stacking motifs*, 2012 IEEE 12th International Conference on Bioinformatics & Bioengineering (BIBE), IEEE: 2012; pp 260-265.

11. Stephenson, W.; Asare-Okai, P. N.; Chen, A. A.; Keller, S.; Santiago, R.; Tenenbaum, S. A.; Garcia, A. E.; Fabris, D.; Li, P. T., The essential role of stacking adenines in a two-base-pair RNA kissing complex. *Journal of the American Chemical Society* **2013,** *135* (15), 5602-5611.

12. Lech, C. J.; Heddi, B.; Phan, A. T. n., Guanine base stacking in G-quadruplex nucleic acids. *Nucleic Acids Research* **2013,** *41* (3), 2034-2046.

13. Šponer, J.; Riley, K. E.; Hobza, P., Nature and magnitude of aromatic stacking of nucleic acid bases. *Physical Chemistry Chemical Physics* **2008,** *10* (19), 2595-2610.

14. Wilson, K. A.; Kellie, J. L.; Wetmore, S. D., DNA–protein π-interactions in nature: abundance, structure, composition and strength of contacts between aromatic amino acids and DNA nucleobases or deoxyribose sugar. *Nucleic Acids Research* **2014,** *42* (10), 6726-6741.

15. (a) Wilson, K. A.; Holland, D. J.; Wetmore, S. D., Topology of RNA–protein nucleobase–amino acid π–π interactions and comparison to analogous DNA–protein π–π contacts. *RNA* **2016,** *22* (5), 696-708; (b) Wilson, K. A.; Kung, R. W.; D'souza, S.; Wetmore, S. D., Anatomy of noncovalent interactions between the nucleobases or ribose and π-containing amino acids in RNA–protein complexes. *Nucleic Acids Research* **2021,** *49* (4), 2213-2225.

16. Kalra, K.; Gorle, S.; Cavallo, L.; Oliva, R.; Chawla, M., Occurrence and stability of lone pair-π and OH–π interactions between water and nucleobases in functional RNAs. *Nucleic Acids Research* **2020,** *48* (11), 5825-5838.

17. Baulin, E.; Metelev, V.; Bogdanov, A., Base-intercalated and base-wedged stacking elements in 3D-structure of RNA and RNA–protein complexes. *Nucleic acids research* **2020,** *48* (15), 8675-8685.

18. Jhunjhunwala, A.; Ali, Z.; Bhattacharya, S.; Halder, A.; Mitra, A.; Sharma, P., On the Nature of Nucleobase Stacking in RNA: A Comprehensive Survey of Its Structural Variability and a Systematic Classification of Associated Interactions. *Journal of Chemical Information and Modeling* **2021,** *61* (3), 1470-1480.

19. (a) Lu, X.-J.; Olson, W. K., 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nature Protocols* **2008,** *3* (7), 1213; (b) Gendron, P.; Lemieux, S.; Major, F., Quantitative analysis of nucleic acid three-dimensional structures. *Journal of molecular biology* **2001,** *308* (5), 919-936; (c) Waleń, T.; Chojnowski, G.; Gierski, P.; Bujnicki, J. M., ClaRNA: a classifier of contacts in RNA 3D structures based on a comparative analysis of various classification schemes. *Nucleic Acids Research* **2014,** *42* (19), e151-e151.

20. (a) Major, F.; Thibault, P., RNA tertiary structure prediction. *Bioinformatics: from genomes to therepies* **2007**; (b) Lengauer, T., Bioinformatics—from genomes to therapies. *Bioinformatics-From Genomes to Therapies* **2007**, 1-24.

21. Chawla, M.; Abdel-Azeim, S.; Oliva, R.; Cavallo, L., Higher order structural effects stabilizing the reverse Watson–Crick Guanine-Cytosine base pair in functional RNAs. *Nucleic acids research* **2014,** *42* (2), 714-726.

22. Huang, Y.; Niu, B.; Gao, Y.; Fu, L.; Li, W., CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* **2010,** *26* (5), 680-682.

23. Gabb, H.; Sanghani, S.; Robert, C.; Prevost, C., Finding and visualizing nucleic acid base stacking. *Journal of Molecular Graphics* **1996,** *14* (1), 6-11.

24. Sherwood, A. V.; Henkin, T. M., Riboswitch-mediated gene regulation: novel RNA architectures dictate gene expression responses. *Annual review of microbiology* **2016,** *70*, 361-374.
25. Seelam, P. P.; Mitra, A.; Sharma, P., Pairing interactions between nucleobases and ligands in aptamer: Ligand complexes of riboswitches: Crystal structure analysis, classification, optimal structures, and accurate interaction energies. *RNA* **2019,** *25* (10), 1274-1290.
26. Ogle, J.; Murphy, F.; Tarry, M.; Ramakrishnan, V., Structure of the Thermus thermophilus 30s ribosomal subunit bound to codon and near-cognate transfer RNA anticodon stem-loop mismatched at the second codon position at the a site with paromomycin. *Cell* **2002,** *111*, 721-732.
27. Klein, D. J.; Schmeing, T. M.; Moore, P. B.; Steitz, T. A., The kink-turn: a new RNA secondary structure motif. *The EMBO journal* **2001,** *20* (15), 4214-4221.
28. Lu, X.-J.; Bussemaker, H. J.; Olson, W. K., DSSR: an integrated software tool for dissecting the spatial structure of RNA. *Nucleic acids research* **2015,** *43* (21), e142-e142.
29. Polikanov, Y. S.; Szal, T.; Jiang, F.; Gupta, P.; Matsuda, R.; Shiozuka, M.; Steitz, T. A.; Vázquez-Laslop, N.; Mankin, A. S., Negamycin interferes with decoding and translocation by simultaneous interaction with rRNA and tRNA. *Molecular cell* **2014,** *56* (4), 541-550.
30. DeLano, W. L., The PyMOL molecular graphics system. *http://www. pymol. org* **2002**.
31. (a) Copeland, K. L.; Anderson, J. A.; Farley, A. R.; Cox, J. R.; Tschumper, G. S., Probing phenylalanine/adenine π-stacking interactions in protein complexes with explicitly correlated and CCSD (T) computations. *The Journal of Physical Chemistry B* **2008,** *112* (45), 14291-14295; (b) Rutledge, L. R.; Campbell-Verduyn, L. S.; Wetmore, S. D., Characterization of the stacking interactions between DNA or RNA nucleobases and the aromatic amino acids. *Chemical Physics Letters* **2007,** *444* (1-3), 167-175; (c) Rutledge, L. R.; Durst, H. F.; Wetmore, S. D., Evidence for stabilization of DNA/RNA− protein complexes arising from nucleobase− amino acid stacking and T-shaped interactions. *Journal of Chemical Theory and Computation* **2009,** *5* (5), 1400-1410.
32. Frisch, M. e.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Petersson, G.; Nakatsuji, H., Gaussian 16. Gaussian, Inc. Wallingford, CT: 2016.
33. Boys, S. F.; Bernardi, F., The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Molecular Physics* **1970,** *19* (4), 553-566.
34. (a) Sharma, P.; Lait, L. A.; Wetmore, S. D., yDNA versus yyDNA pyrimidines: computational analysis of the effects of unidirectional ring expansion on the preferred sugar–base orientation, hydrogen-bonding interactions and stacking abilities. *Physical Chemistry Chemical Physics* **2013,** *15* (7), 2435-2448; (b) McConnell, T. L.; Wetmore, S. D., How do size-expanded DNA nucleobases enhance duplex stability? Computational analysis of the hydrogen-bonding and stacking ability of xDNA bases. *The Journal of Physical Chemistry B* **2007,** *111* (11), 2999-3009; (c) Wheaton, C. A.; Dobrowolski, S. L.; Millen, A. L.; Wetmore, S. D., Nitrosubstituted aromatic molecules as universal nucleobases: Computational analysis of stacking interactions. *Chemical physics letters* **2006,** *428* (1-3), 157-166.
35. (a) Mandal, M.; Breaker, R. R., Gene regulation by riboswitches. *Nature reviews Molecular cell biology* **2004,** *5* (6), 451-463; (b) Tucker, B. J.; Breaker, R. R., Riboswitches as versatile gene control elements. *Current opinion in structural biology* **2005,** *15* (3), 342-348.
36. Nissen, P.; Ippolito, J. A.; Ban, N.; Moore, P. B.; Steitz, T. A., RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proceedings of the National Academy of Sciences* **2001,** *98* (9), 4899-4903.

Figures and Tables

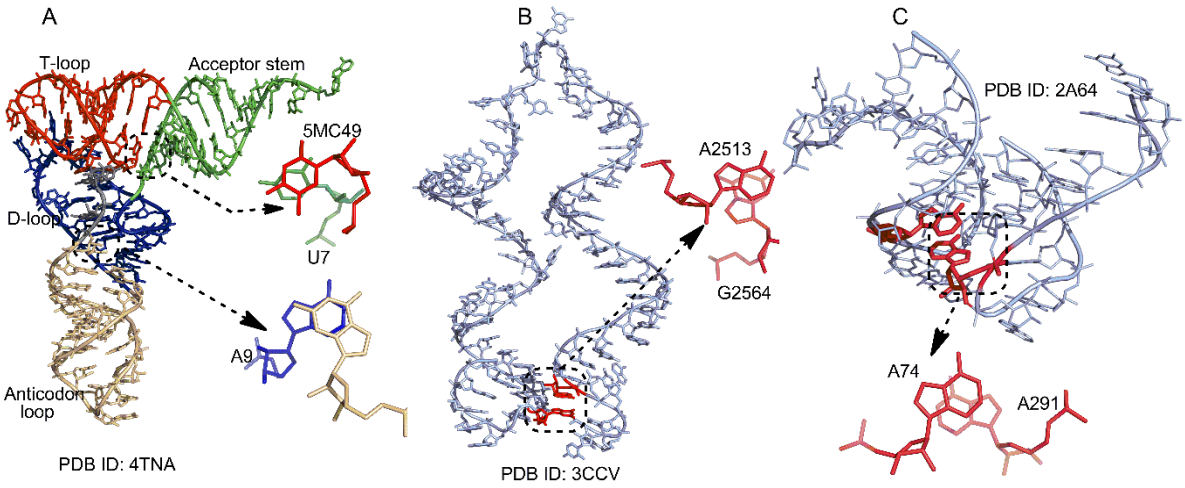


Figure 1. Examples of stacking interactions in RNA structures. (A) Coaxial stacking between anticodon loop (ivory) and D-loop (blue), and between T-loop (red) and acceptor stem (green). (B) Inter-loop stacking between A2513 and G2564 of 23S RNA of *Haloarcula marismortui*. (C) Kissing loop stacking interaction between A74 and A291 of the bacterial ribonuclease P.

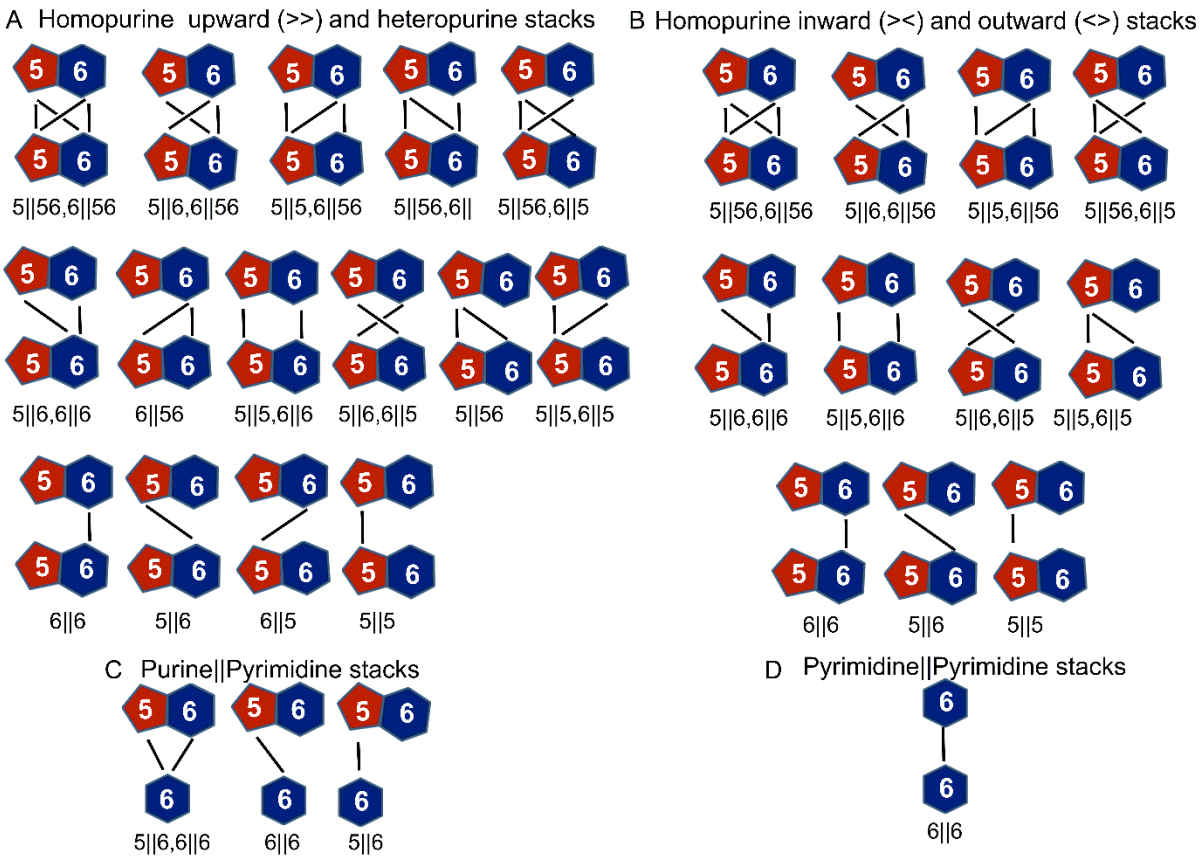


Figure 2. Possible stacking arrangements in (A) homopurine (upward (>>)) and heteropurine stacks, (B) homopurine stacks (inward (><) and outward (<>) face orientations), (C) purine||pyrimidine stacks and (D) pyrimidine||pyrimidine stacks.

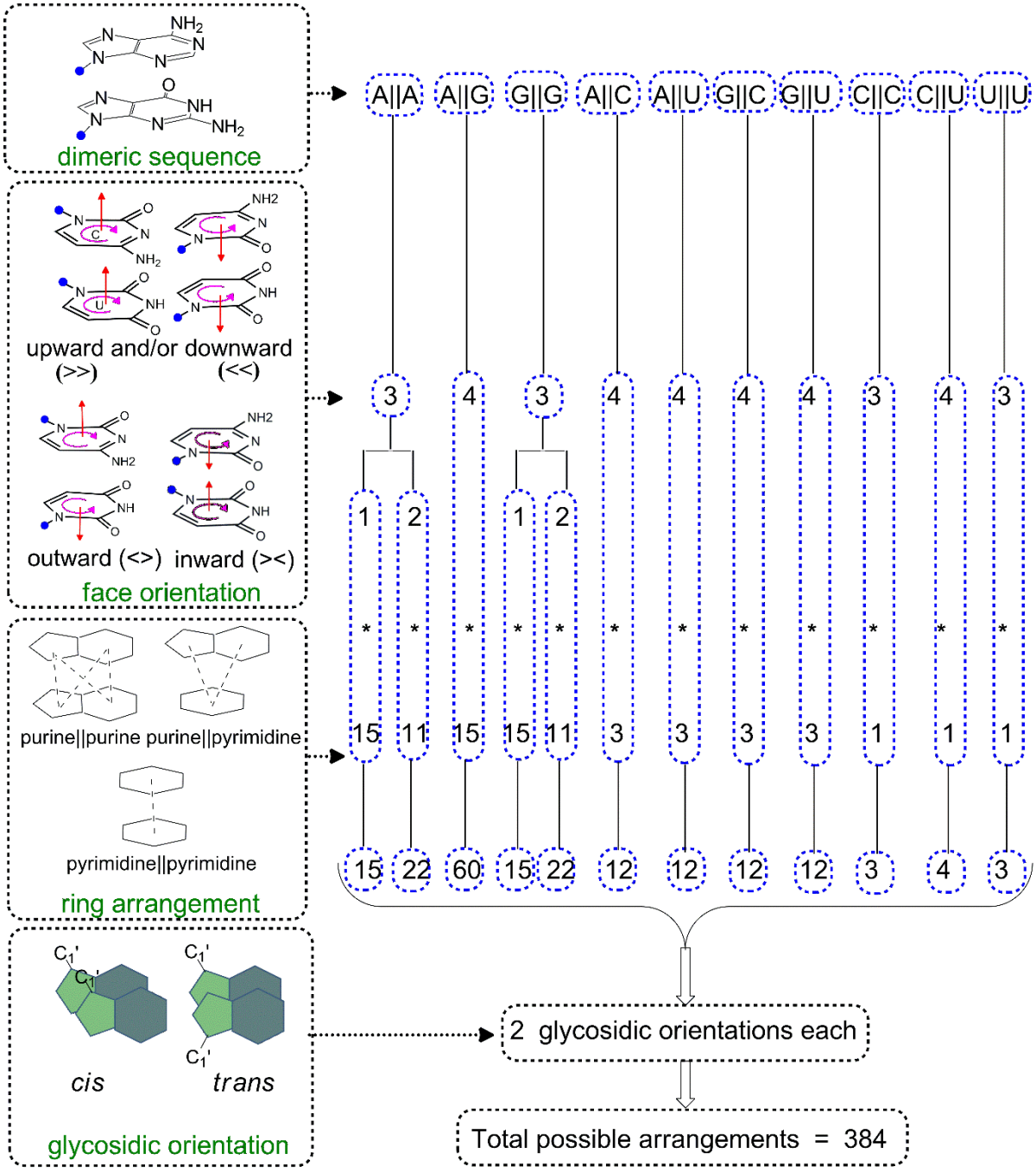


Figure 3 Possible stacking arrangements of the bases, classified with respect to the base sequence, face orientation, ring arrangement and glycosidic orientation.[18]

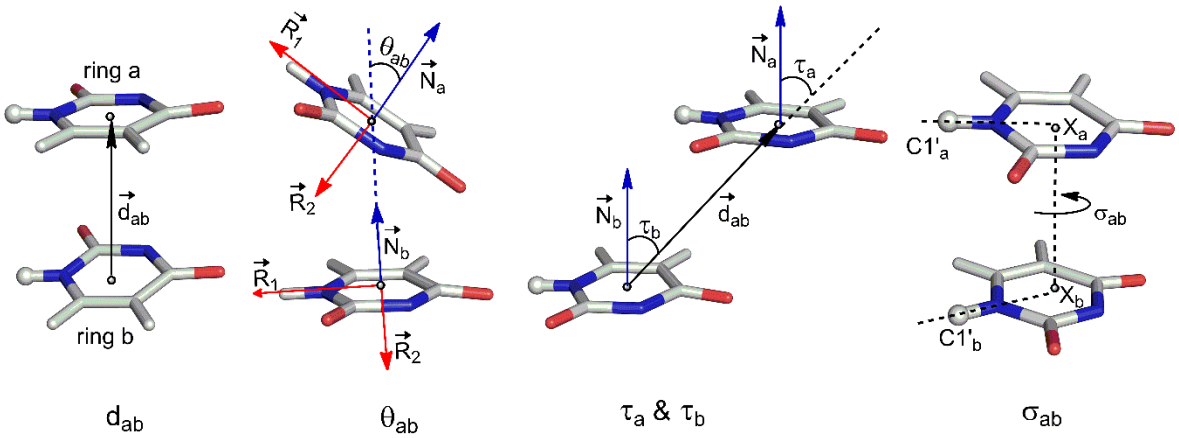


Figure 4. Pictorial representation of the geometrical parameters used to identify stacking contacts – the vertical distance between the stacked rings (d_{ab}), tilt angle (θ_{ab}), horizontal shift angles (τ_a and τ_b) and mutual glycosidic orientation of the bases within a stack (σ_{ab}).

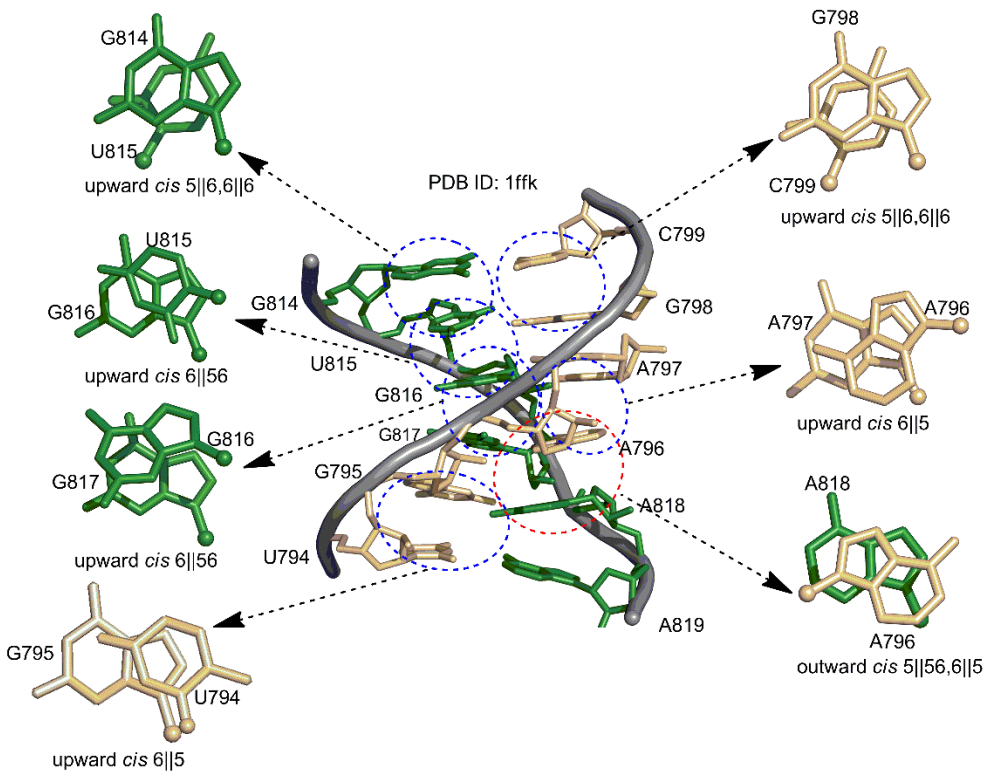


Figure 5. Examples of interstrand and intrastrand stacking interactions among the bases in an RNA helical region.

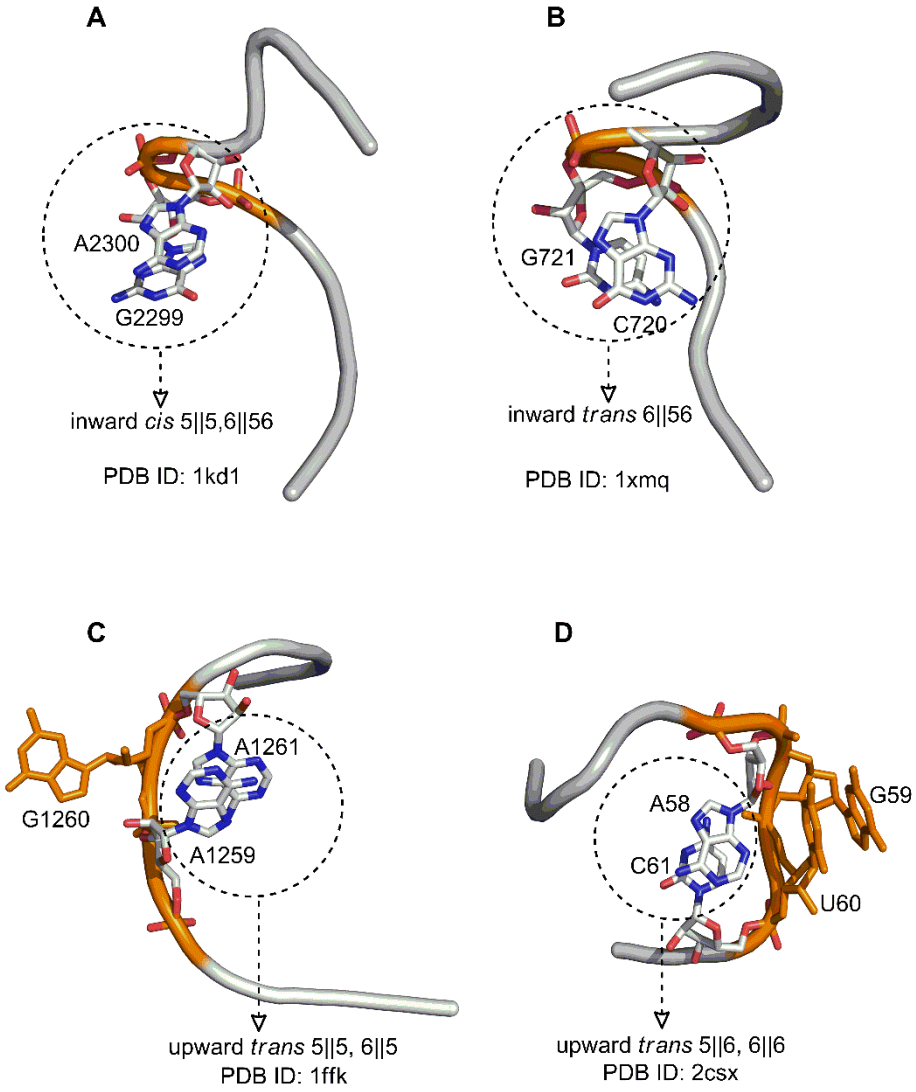


Figure 6. (A and B) Examples of "V" shaped bends in an RNA strand. (C) A non-consecutive stack with a single flipped-out base in between. (D) A non-consecutive stack with two flipped-out bases.

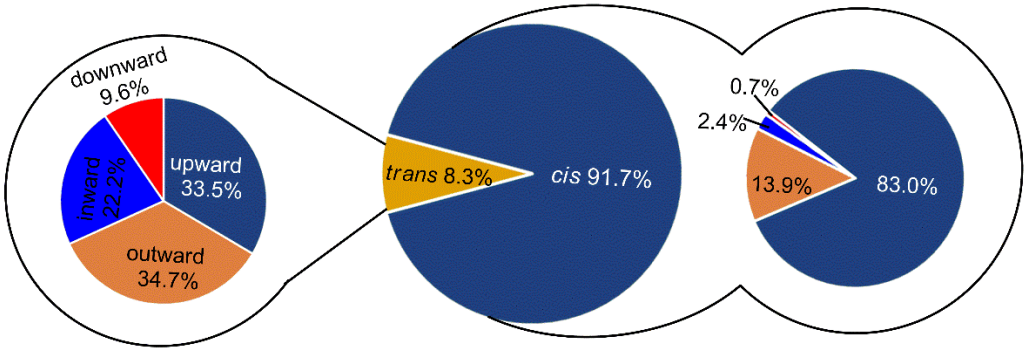


Figure 7. Distribution of stacks based on relative glycosidic orientation of the interacting bases, and further distribution in terms of nucleobase face orientations.

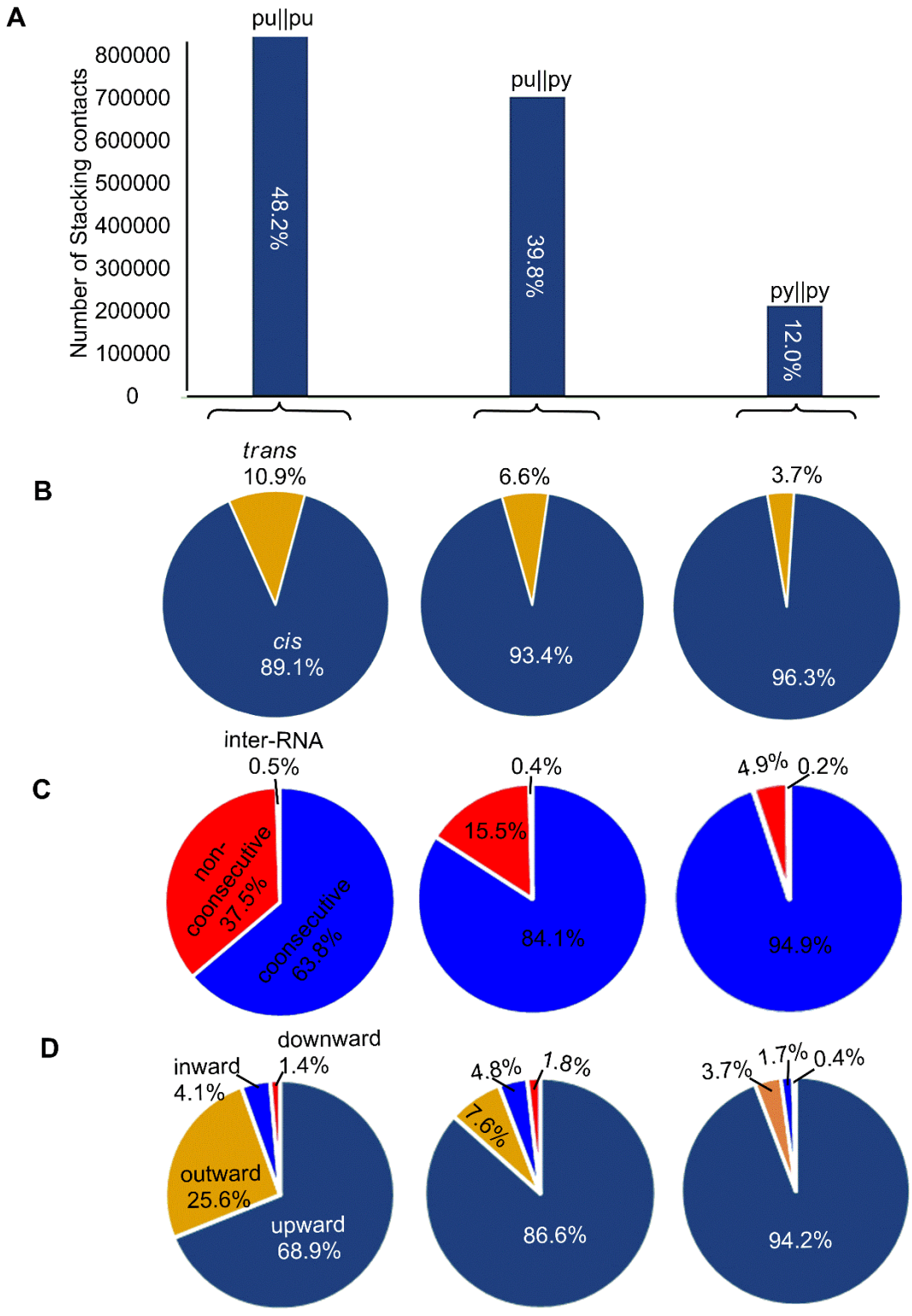


Figure 8. Distribution of purine||purine (pu||pu), purine||pyrimidine (pu||py) and pyrimidine||pyrimidine (py||py) stacks in terms of (A), their further distribution into *cis* and *trans* orientations (B), consecutive, non-consecutive and inter-RNA types (C) and nucleobase face orientations and (D).

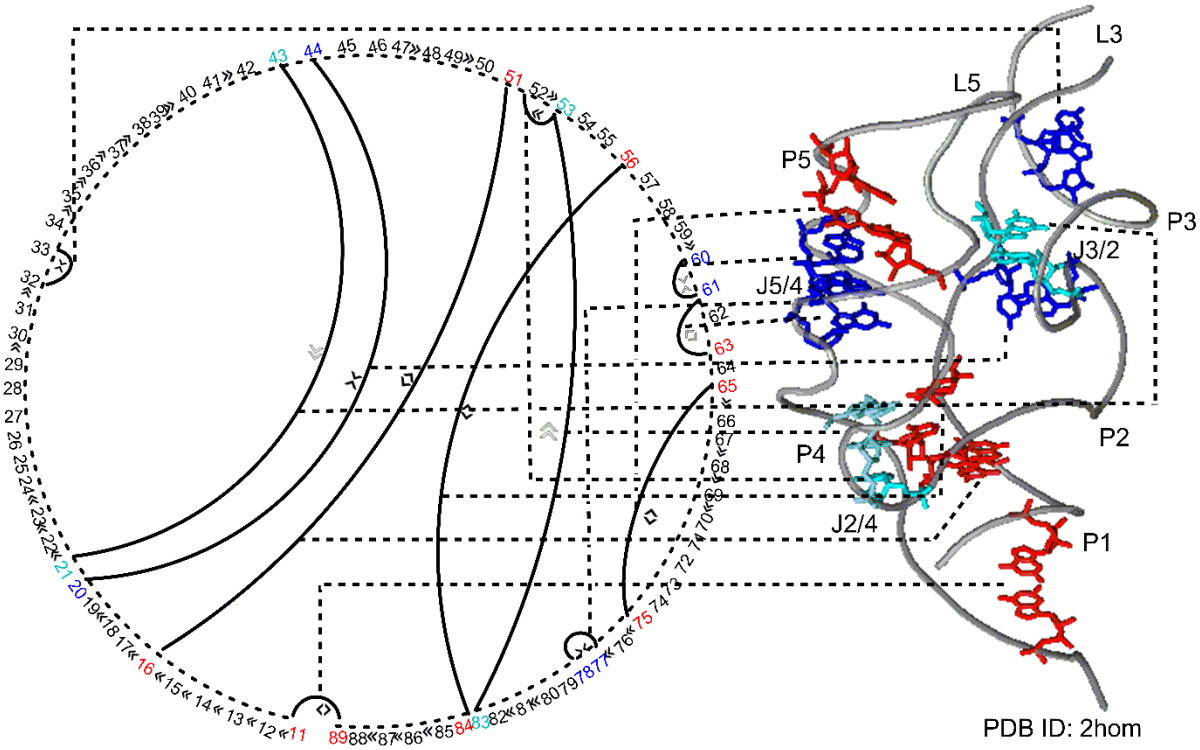


Figure 9. 2D representation of the crystal structure of Thi–Box riboswitch (PDB code:2hom) to illustrate the stacking contacts (left). 3D structure showing stacking arrangements that aid in folding of the structure (right). Colors represent different type of stacks (red = outward, blue = inward and cyan = downward).

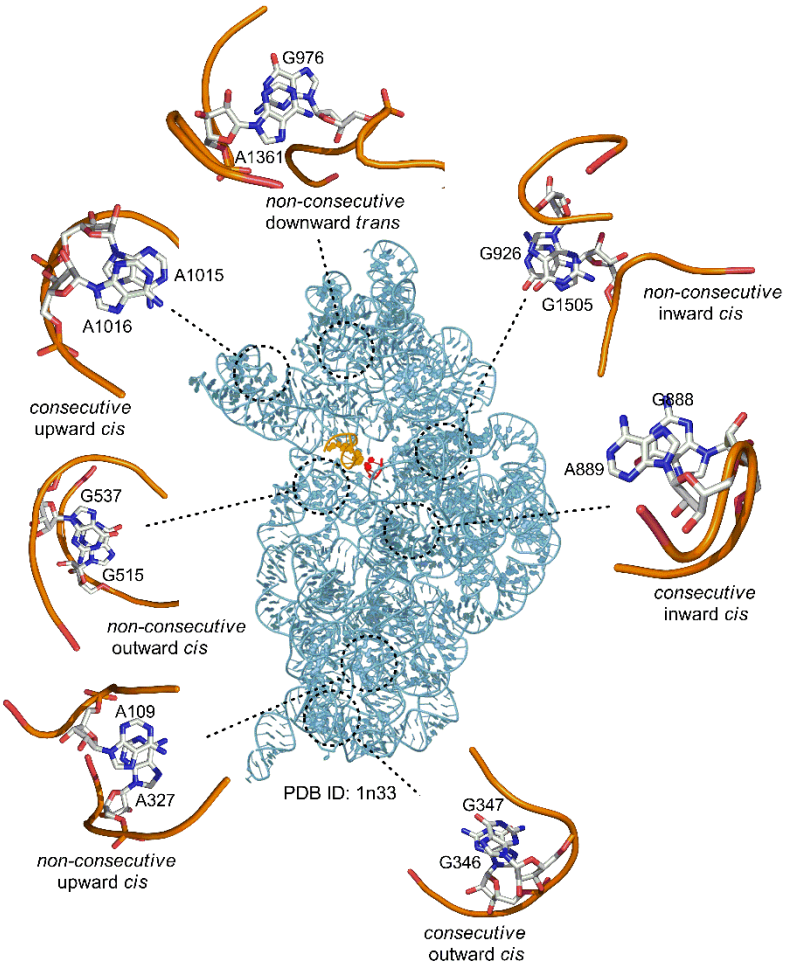


Figure 10. Examples of base stacks in 16S rRNA of the 30S ribosomal subunit of *Thermus thermophilus*.

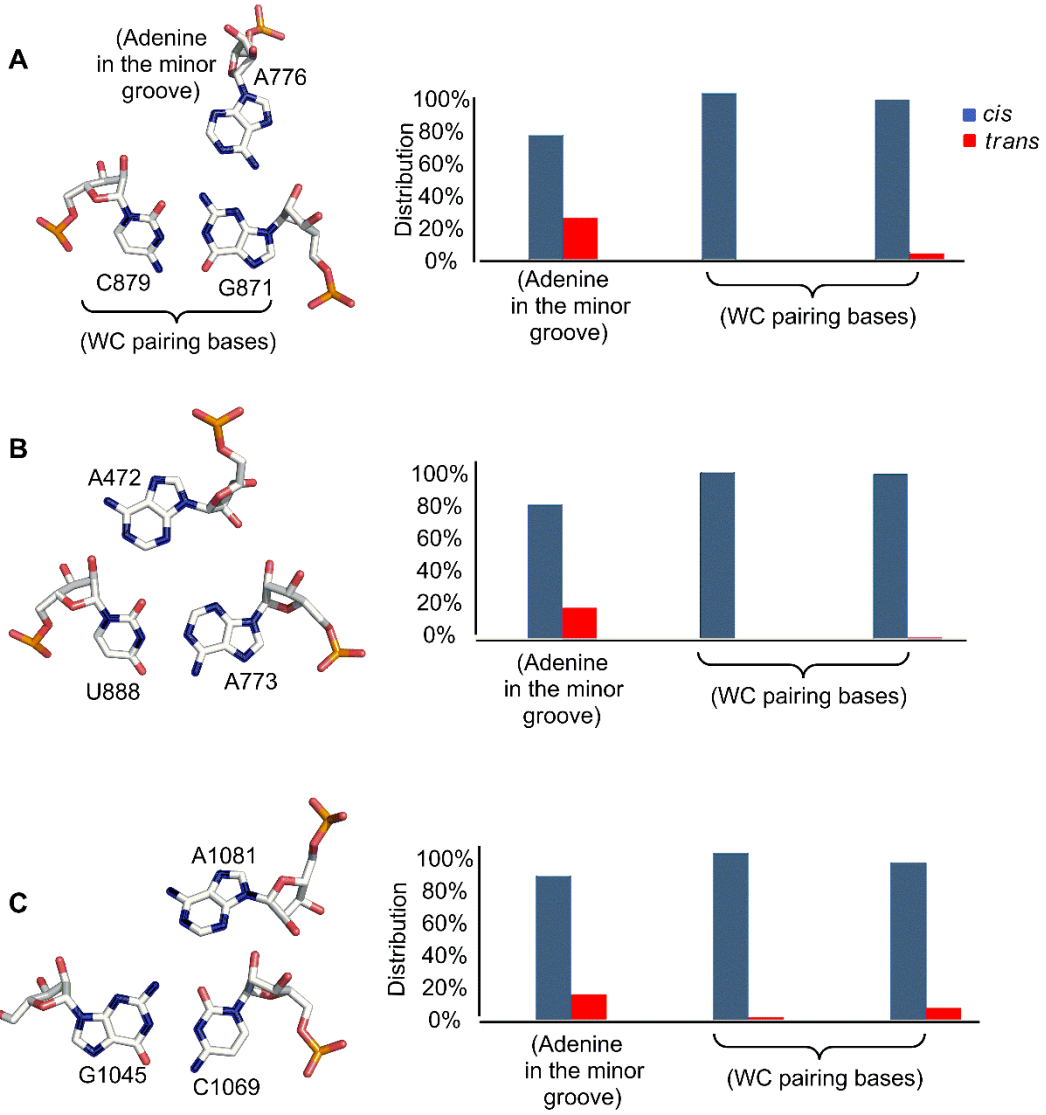


Figure 11. Example of A-minor motif and the distribution of stacks of their bases among different glycosidic orientation. (A) type 0 A-minor motif, (B) type I A-minor motif and (C) type II A-minor motif

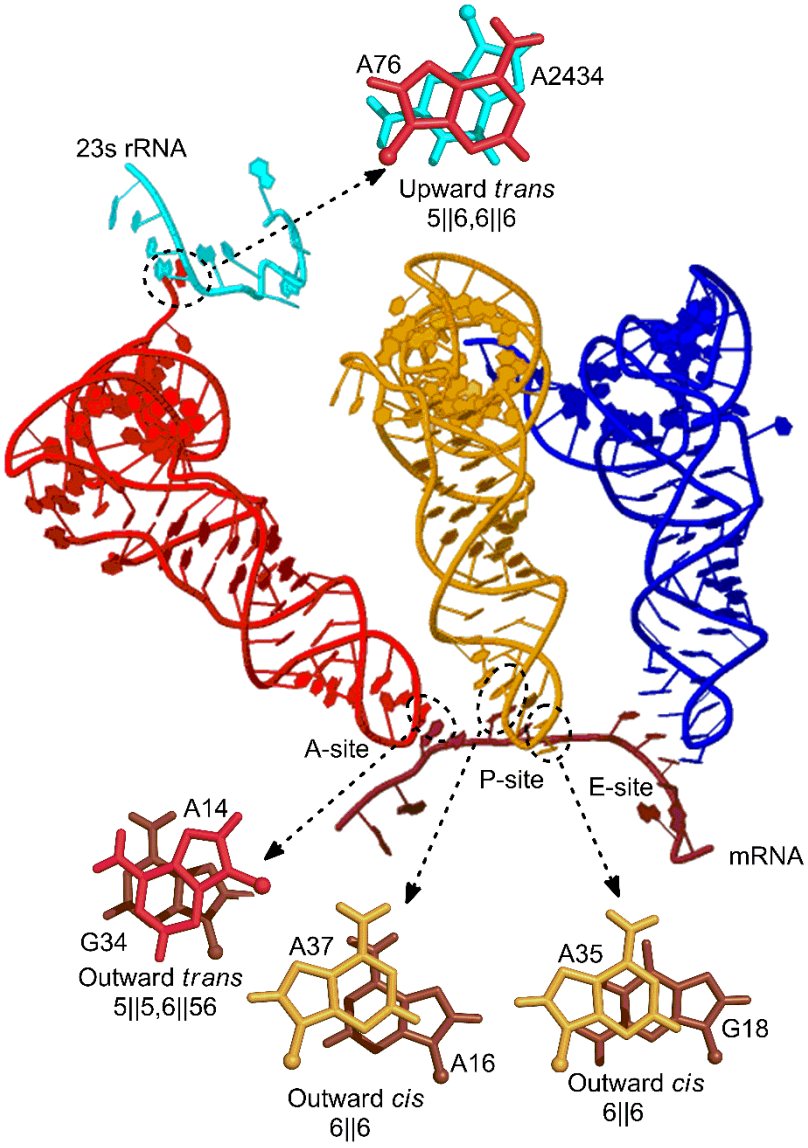


Figure 12. Stacking interactions between the bases of tRNA, mRNA and rRNA of tRNA in a translation complex.

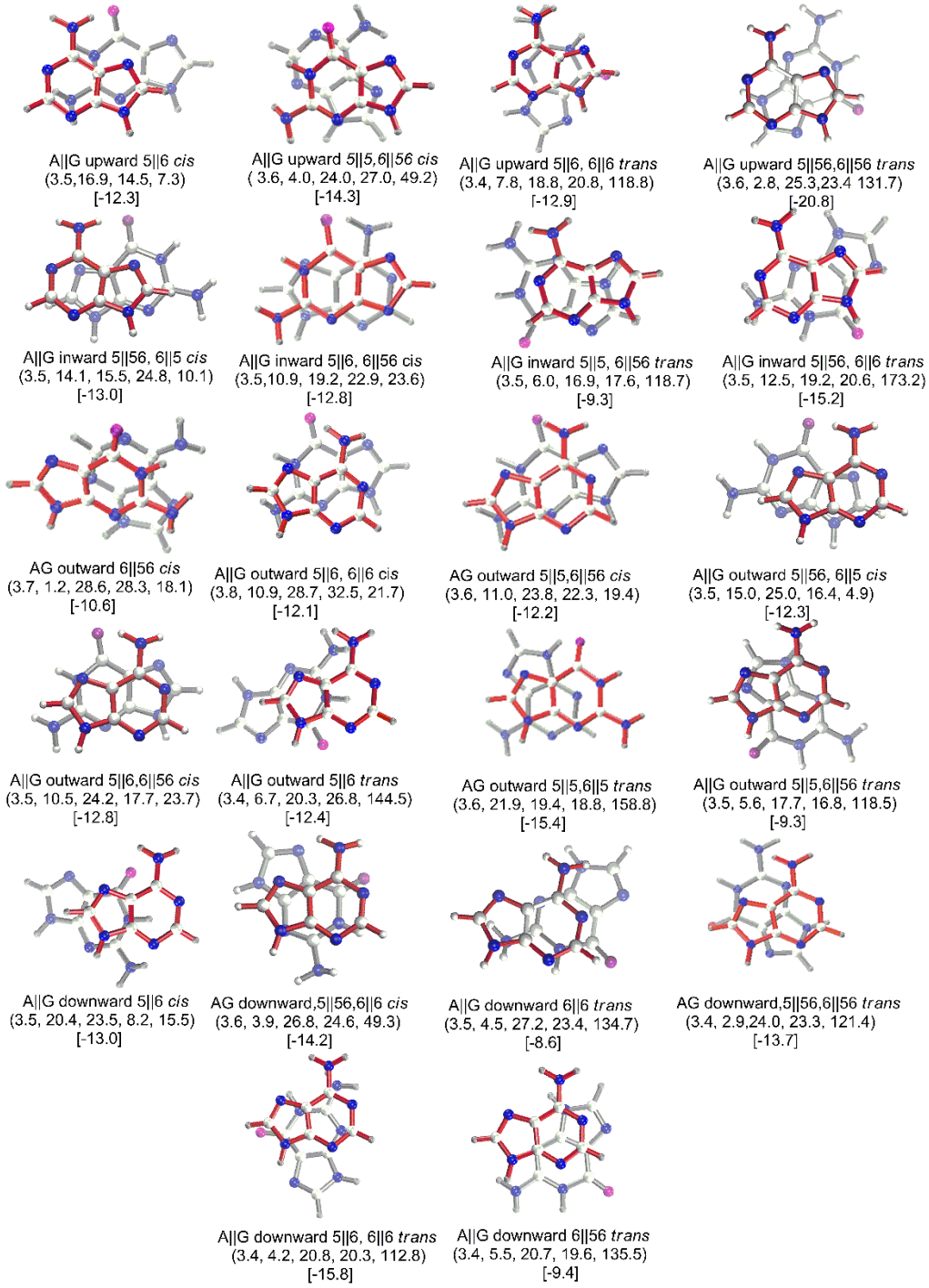


Figure 13. Quantum mechanically optimized AG stacks. Structural parameters of each arrangement (*i.e.*, $\vec{d}_{ab}$ (Å), θ_{ab} (deg.), τ_a (deg.), τ_b (deg.) and σ_{ab} (deg.)) are provided in parentheses in the same order. Stacking interaction energies (kcal mol^{-1}) are provided in square brackets.

Table 1. Distribution of the base stacks based on their relative position in the RNA structures.

Consecutive	**Non consecutive**	**Inter-RNA**	**Total**
1312017 (75.6%)	416354 (24.0%)	7110 (0.4%)	**1735481**

Table 2. Distribution of the base stacks based on relative face orientation of the interacting bases.

upward	downward	inward	outward	Total
1370087 (78.9%)	24757(1.4%)	69964 (4.0%)	270673(15.6%)	**1735481**