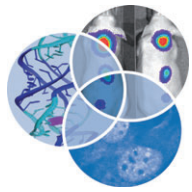
Overview



An introduction to recurrent nucleotide interactions in RNA

Blake A. Sweeney,[1] Poorna Roy[2] and Neocles B. Leontis[2]*

RNA secondary structure diagrams familiar to molecular biologists summarize at a glance the folding of RNA chains to form Watson–Crick paired double helices. However, they can be misleading: First of all, they imply that the nucleotides in loops and linker segments, which can amount to 35% to 50% of a structured RNA, do not significantly interact with other nucleotides. Secondly, they give the impression that RNA molecules are loosely organized in three-dimensional (3D) space. In fact, structured RNAs are compactly folded as a result of numerous long-range, sequence-specific interactions, many of which involve loop or linker nucleotides. Here, we provide an introduction for students and researchers of RNA on the types, prevalence, and sequence variations of inter-nucleotide interactions that structure and stabilize RNA 3D motifs and architectures, using *Escherichia coli* (*E. coli*) 16S ribosomal RNA as a concrete example. The picture that emerges is that almost all nucleotides in structured RNA molecules, including those in nominally single-stranded loop or linker regions, form specific interactions that stabilize functional structures or mediate interactions with other molecules. The small number of noninteracting, 'looped-out' nucleotides make it possible for the RNA chain to form sharp turns. Base-pairing is the most specific interaction in RNA as it involves edge-to-edge hydrogen bonding (H-bonding) of the bases. Non-Watson–Crick base pairs are a significant fraction (30% or more) of base pairs in structured RNAs.

How to cite this article:
WIREs RNA 2015, 6:17–45. doi: 10.1002/wrna.1258

INTRODUCTION

RNA three-dimensional (3D) motifs, folds, and architectures are stabilized by a variety of interactions between individual nucleotides (nts), primarily edge-to-edge base-pairing, face-to-face base-stacking, and base–backbone interactions of various kinds. Base-pairing is the most specific of these interactions, while base-stacking provides much of the stabilization energy for RNA folding.[1] In addition to the well-known Watson–Crick (WC) pairs, structured RNA molecules contain many types of non-WC base pairs.[2] The non-WC pairs constitute a substantial fraction, typically greater than one third, of all base pairs in a structured RNA.[3] Familiarity with these and other recurrent nucleotide-level interactions is fundamental for understanding RNA folding, function, and evolution, because they are so widespread and crucial as building blocks of RNA 3D motifs and complex RNA architectures. Furthermore, analysis of the recurrent geometries of non-WC pairs extends our understanding of the patterns of sequence variation in homologous RNA molecules, beyond the well-known covariation of WC AU and CG base pairs.[4,5]

To alert the reader to the importance of these interactions, we begin by demonstrating their prevalence in the 'loops' of the two-dimensional (2D) representations of structured RNA molecules, using *Escherichia coli* (*E. coli*) 16S ribosomal RNA (rRNA) as an example. Then we provide detailed descriptions of the most important interactions. We introduce the concept of base-pair isostericity and show how

*Correspondence to: leontis@bgsu.edu

[1]Department of Biological Sciences, Bowling Green State University, Bowling Green, OH, USA

[2]Department of Chemistry, Bowling Green State University, Bowling Green, OH, USA

Conflict of interest: The authors have declared no conflicts of interest for this article.
Additional Supporting Information may be found in the online version of this article.

to apply it to understand the sequence variation of both WC and non-WC base pairs in homologous RNA molecules and geometrically similar, recurrent 3D motifs. We describe how to annotate 2D diagrams to capture and communicate the most important interactions of RNA 3D motifs. Finally, we conclude by reviewing online resources available through the Nucleic Acid Database (NDB) web portal (http://ndbserver.rutgers.edu/) and related resources that provide comprehensive lists of interactions and visualizations for all atomic-resolution RNA structures in the Protein Data Bank (PDB), tools for structure search, and compilations of 3D motifs organized by structural similarity.

WHAT ABOUT THE LOOPS IN RNA 2D DIAGRAMS?

We begin with RNA secondary structure or '2D' diagrams, which have a long and remarkable history. The first 2D diagrams were derived in the 1960s for transfer RNA (tRNA). When the first tRNA sequence was obtained, three different 2D structures were proposed,[6] because it was not possible, on the basis of just one sequence, to decide which was correct. One of these was the now familiar cloverleaf structure, with four short helices, three hairpin loops, and one four-way junction. Remarkably, as soon as the second tRNA was sequenced, it became immediately clear that the cloverleaf had to be the correct structure, under the assumption that all tRNAs form a common 2D structure.[7] These two tRNA sequences (yeast alanine and tyrosine tRNAs) happened to differ sufficiently in sequence that the positions of WC base pairs could be deduced by the compensating changes apparent when the two sequences were drawn in the cloverleaf 2D structure. The presence of WC covariation (AU, UA, GC, and CG) at aligned positions of the sequences provided evidence for WC base-pairing, consistent only with this 2D scheme. This early work provided the conceptual basis for the comparative sequence analysis (CSA) method for RNA 2D structure determination.[8,9] The atomic-resolution 3D structures of tRNAs solved by X-ray crystallography in the 1970s[10] confirmed the CSA prediction of the tRNA 2D structure and encouraged scientists to apply the method to much larger RNAs including 16S[11] and 23S rRNA.[12] The first 16S 2D structure was also determined on the basis of just two complete, sufficiently divergent sequences, supplemented in this case by a collection of nuclease fragments from about 100 other sequences.[11] With the addition of more sequences, it was possible to refine the 2D structure, so that by the time the 3D structure of 16S was determined at atomic resolution in 2000, almost 98% of the WC base pairs were correctly predicted.[13] Comparison with X-ray structures revealed not only the strengths but also the weaknesses of the CSA methods. In a small number of cases, 3D modeling errors were corrected by reference to CSA predictions. However, where there is no covariation in sequences, physical interactions cannot be detected by CSA, at least not at the level of pairwise sequence comparisons. Also, to detect non-WC base pairs by CSA new rules of sequence variation were needed, as will be explained below (see section *Base-Pair Isostericity and Sequence Variation*). Pioneering work on the use of non-WC sequence variation to detect tertiary interactions and to use them to build 3D models of RNA was carried out by Westhof and Michel on the group I introns.[14,15]

Strictly speaking, RNA 2D diagrams are minimalist representations of the folding of the RNA chain on itself to form the 'nested' WC base pairs. These representations generally omit all other interactions, although they may display long-range (LR) WC pairs forming pseudoknots (PK), when these can be detected by CSA. In cases with sufficient covariation data, it has been possible to extend secondary structures to include some non-WC pairs.[16] Furthermore, when sequences are analyzed at the level of recurrent motifs, geometries of non-WC pairs formed by interacting nts can also be predicted, as will be explained below.[17–19]

RNA 2D diagrams can be misleading in at least two ways: (1) by implying that loops and linkers are single-stranded and that their nucleotides do not interact with the rest of the RNA in specific, phylogenetically conserved ways; and (2) by implying that RNA 3D structure is extended in 3D space with few or no interactions between different helical elements.

Readers should note that 2D diagrams prepared before 3D structures were solved continue to be widely used. Some of these 2D diagrams contain errors, especially in the pairings of the most conserved bases that show little or no sequence covariation.[20] Some WC pairs observed in the 3D structure are absent in the 2D and some shown in the 2D are not found in the 3D. In addition, some pairs shown as WC in the 2D actually assume non-WC geometries in 3D structures. There are also discrepancies between 3D structures of the same RNA molecules.[3] Some are due to bona fide structural differences related to changes in functional state while others are likely due to differences in 3D modeling related to the quality and interpretation of the underlying experimental electron densities.

Discrepancies between 2D and 3D representations tend to occur either adjacent to or within internal or multi-helix junction (MHJ) loops.[21] There are at

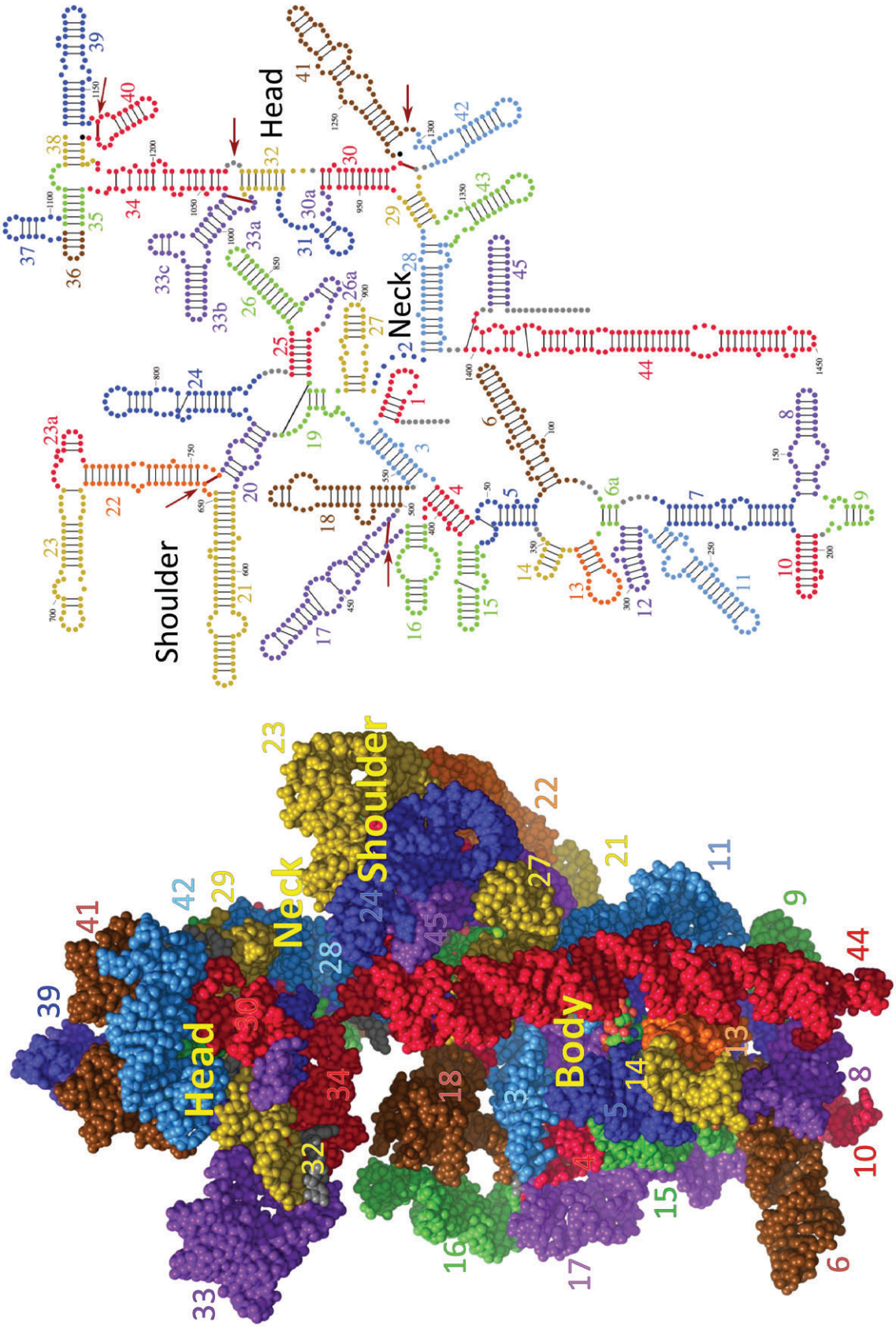


FIGURE 1 | Structural representations of *Escherichia coli* (*E. coli*) 16S rRNA. 2D and 3D representations of 16S rRNA of *E. coli*, with helical elements colored identically. Lines connect WC paired nts. The 3D is PDB file 2AW7[22] whereas the 2D is modified from the study by Petrov et al.[20] The multi-helix junctions that contain embedded WC base pairs are indicated with brick-red arrows and the paired bases with bold lines of the same color.

least five WC pairs in MHJ loops of bacterial 16S rRNA that are not shown in most 2D diagrams, and at least four of these could not be detected by CSA. These pairs are indicated with bold red lines in the 2D diagram of *E. coli* 16S rRNA shown in Figure 1. This 2D representation is based on annotations of nt interactions for NDB file RR0125 (PDB file 2AW7), as posted on the NDB website (http://ndbserver.rutgers.edu/) and verified by visual analysis of the 3D structure.[23] The corresponding helical elements in the 2D and 3D representations of 16S rRNA are shown in Figure 1, highlighted with the same colors.

Types of Loops

The 2D structure partitions the nts of an RNA into disjoint sets constituting helices, hairpin loops (HL), internal loops (IL), MHJ loops, linkers, and 5′- or 3′-terminal sequences. Numbering the nts in a 2D diagram consecutively from the 5′- to 3′-end makes it easy to find the helix, loop, or linker segment to which a given nt belongs. HL occur at the ends of helices and comprise continuous segments of RNA sequence linking the 3′ and 5′ strands at the ends of helices. Therefore, they are also called terminal loops, and many are found at or near the external surfaces of RNA molecules. Certain types of HL, especially those with the consensus sequence, GNRA, interact with other helical elements in the interior. IL occur between two helices and are composed of two strands. MHJ loops occur where three or more helical elements meet. Linkers are single-stranded segments that connect two domains or helical elements.

What Is an RNA Motif?

Motifs are recurrent patterns. For RNA, it is useful to distinguish among motifs in the primary (i.e., sequence), 2D, and 3D structures that the molecule actually forms. 'GNRA' is an example of a sequence-level motif that designates a common pattern of four 'loop' nts seen recurrently in RNA HL, where the first base is almost always G and the last is almost always A, while the other two can vary, with 'N' standing for A, C, G, or U, and 'R' for G or A. Motifs in the 2D structure are distinguished by whether they form HL, IL, or MHJ and by the number of nts they have in their component strands. Thus, '2 × 3' internal loops have two nts in one strand and three in the other and are considered different from 3 × 3 or 2 × 4 IL at the level of 2D motifs. In this article, we are concerned with recurrent 3D motifs, which unlike sequence or 2D motifs, can vary in sequence as well as in the number of nts, as long as they form similar 3D structures. The challenge is to find good ways to measure 3D motif similarity to group them in biologically meaningful ways.

To illustrate the importance of loop nts, the 3D motifs they form and their interactions in structured RNA molecules, in writing this article, we analyzed the atomic-resolution structure of 16S rRNA from *E. coli*, as represented by a high-quality, X-ray crystal structure, PDB file 2AW7.[22] We examined the prevalence of loop nts, their distributions and the number and types of interactions they form compared to interactions of nts in WC-paired helices. The results of this analysis are presented next.

Distribution of Nucleotides and Their Interactions in 16S rRNA

Ribosome scientists partition the 16S rRNA 2D structure into about 50 helical elements, connected to each other by MHJ or by single-stranded linkers. Each helical element comprises one or more helices, connected to each other by IL. Of these, 32 helical elements terminate on one end in HL, as e.g., helices 6 and 8 in Figure 1.[20,24] A further 17 helical elements serve to connect two MHJ (e.g., helices 4, 5, and 7 in Figure 1) or an MHJ and single-stranded linkers (e.g., helix 28). An additional helical element of 16S, helix 2, constitutes a pseudoknot and is not considered part of most 2D representations of 16S. More than half of the helical elements are interrupted by one or more internal loops (IL). For example, helix 44, the longest helical element in 16S and a prominent part of the 30S interface with the 50S subunit, contains nine IL. The large number of IL in *E. coli* 16S (63, see Table 1) accounts for the larger number of helices (~111), defined as uninterrupted WC-paired duplexes, compared with helical elements (~50), and the prevalence of very short 'helices', consisting of less than three base pairs.

Table 1 presents the distribution of nts based on the corrected 2D structure of *E. coli* 16S rRNA shown in Figure 1 and the analysis of annotations of NDB file RR0125 (PDB file 2AW7). This 3D structure has 1530 nts, 12 nts less than the accepted total for *E. coli* 16S (1542), because some of the nts on the 5′- and 3′-ends are not resolved in this structure, which also lacks tRNAs and mRNA.

The most significant result in Table 1 is that fully 43.4% of the nts of *E. coli* 16S belong to the loop or linker regions. The remaining 56.6% of nts form the nested AU, GC, and GU WC base pairs that constitute the helices defining the 2D structure. In addition to the five WC pairs mentioned above that are integral parts of MHJ loops and which we do not include in the 2D, the 3D structure contains twelve more WC base pairs forming LR tertiary (3°) interactions for a

TABLE 1 | Summary of the 2D Structure of *Escherichia coli* 16S rRNA, updated with Base-Pairing Annotations from the NDB Entry for 2AW7 and Manual Analysis of the 3D Structure.

16S rRNA Elements	Number of Elements	Number of Nucleotides	Percentage of Total Nucleotides
Hairpin loops	32	172	11.24
Internal loops	63	256	16.73
Multi-helix junctions	17	197	12.88
Linker segments	8	39	2.55
Total 'loop' nts		664	43.40
Helices	111	866	56.60
Total nts		1530	100.00

total of 17 tertiary WC pairs. The 34 nts forming these tertiary WC pairs are assigned to their respective loop or linker segments rather than to the helices of the 2D structure. Regarding the distribution of nts in different kinds of loops, comparable numbers of nts constitute the HL (11.2%) and MHJ loops (12.9%), while the IL have somewhat more nts (16.7%). The remaining nts belong to linker regions (2.6%). In the category of single-stranded linkers, we include the 5′- and 3′-ends of the molecule. In summary, more than 40% of the nts of 16S rRNA belong to loops or linkers. Next, we examine the size distributions of helices and loops.

Size Distributions of Helices and Loops

Figure 2 shows the distribution, in base pairs, of the lengths of 16S helices. Most helices in 16S are surprisingly short, comprising four or fewer base pairs. This is not atypical of structured RNA molecules. The longest continuous helices in *E. coli* 16S, uninterrupted by internal loops or single-base bulges, are only 12 base-pairs long, and a significant number of 'helices' consist of only one or two WC base pairs. Reasons for including these very short helices in the 2D structure are discussed below.

Figures 3 and 4 show the size distributions of HL and IL in 16S rRNA. Many HL in 16S rRNA comprise 4 nts, and almost all of these are recurrent GNRA or UNCG type loops.[25] The smallest HL is just two nts and is found on the end of helix 6 of *E. coli* 16S rRNA. In this loop, the closing base pair is WC and is not counted as part of the HL. Recurrent GNRA and UNCG HL are closed by non-WC paired nts, which are counted as parts of these loops; therefore, GNRA and UNCG HL also have just two unpaired nts, even though they are commonly called 'tetra-loops' in the literature.

The smallest IL consists of just one 'bulged' nt, flanked on each side by WC base pairs. There are several of these in 16S rRNA, and almost all of them are extremely conserved.[16] The 2D diagram implies that each of these is 'bulged' out. While the 3D structure shows that some are indeed bulged out, others form specific interactions with the adjacent WC base pairs or intercalate between them. Thus, even the simplest IL, drawn identically in 2D diagrams, give rise to different, geometrically distinct 3D motifs. IL having two nts can be symmetric (1×1), with one nt in each strand, or asymmetric (2×0) with both loop nts in one strand. Again, each of these 2D motifs gives rise to different, structurally distinct 3D motifs. To classify loop motifs in functionally meaningful ways, the types of interactions that the loop nts form among themselves must be considered. Using this approach, all HL and IL extracted from unique NDB structures have been classified by structural similarity and organized into an RNA Motif Atlas.[26] The RNA Motif Atlas can be accessed through the redesigned NDB website,[23] and will be discussed in the last section of this review. At this point, we note that computational tools such as FR3D ('Find RNA 3D') that annotate RNA structures and find recurrent 3D RNA motifs,[27,28] as well as compare and cluster them, rely on the analysis of the recurrent nt interactions described in this review.

Interactions of Loop Nucleotides

Using 16S rRNA as an example, it was shown in the previous sections that substantial fractions of nts in structured RNAs belong to 'loops' and these vary in size and structure. Next, we examine the roles that loop nts play in RNA 3D structures. Figures 5 and 6 provide histograms to compare the number of interactions involving loop versus helix nucleotides. Figure 5 shows base-pair interactions, and Figure 6 shows all FR3D annotated inter-nucleotide interactions, including base-pairing, base-stacking, and base–phosphate (BPh) interactions. Each type of interaction will be described in more detail below. Two important points emerge from Figure 5: (1) more than 50% of 'loop' nucleotides form one or more base pairs; and (2) a significant fraction (9%) of helix nts makes one or two additional base pairs, in addition to the defining WC pair each forms. The additional base pairs involve base edges other than the WC edge and are, by definition, non-WC pairs. In no case do bases make more than three base pairs, for reasons explained below.

Figure 6 shows that when we include stacking and BPh interactions, most nts in the 16S structure, whether they are found in loops or helices, form two or more interactions. As expected, most helix nts form

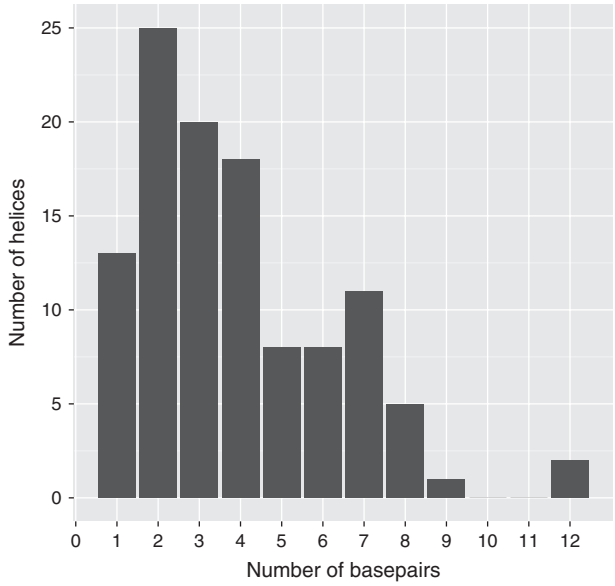


FIGURE 2 | Distribution of helix lengths in 16S rRNA. Histogram of helix lengths (in base pairs) from the 2D representation of 16S rRNA in Figure 1, using definitions of helices explained in the text.

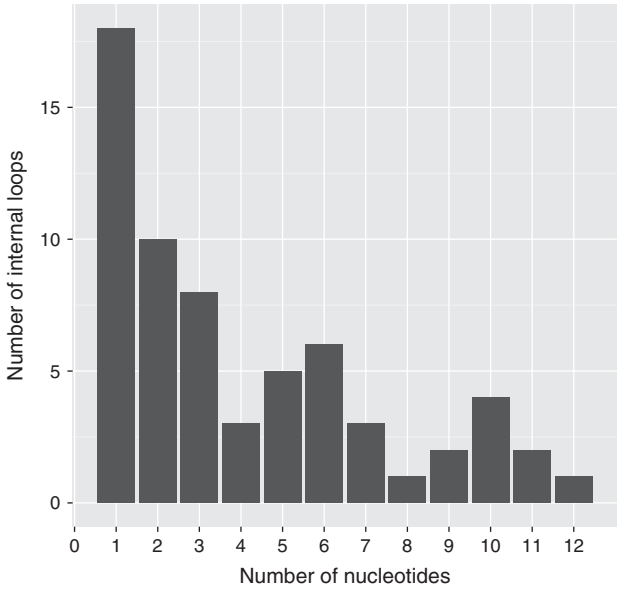


FIGURE 4 | Distribution of internal loop (IL) sizes in 16S rRNA. Histogram of IL sizes (in nts) from the 2D representation of 16S rRNA in Figure 1. Flanking WC basepairs are not included.

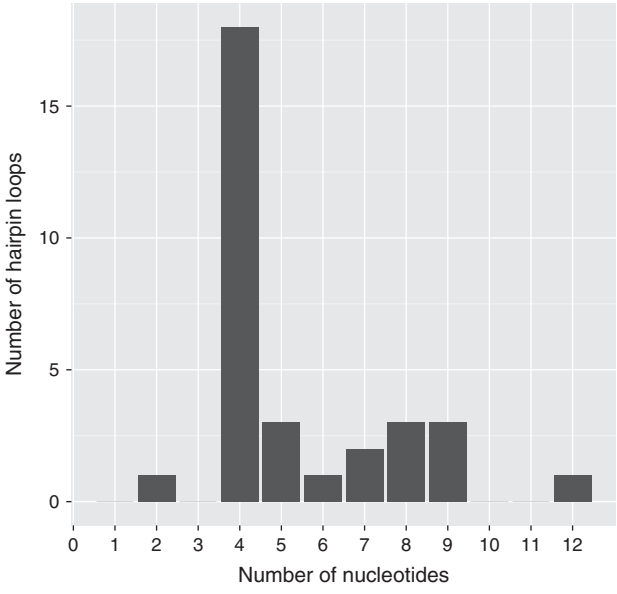


FIGURE 3 | Distribution of hairpin loop (HL) sizes in 16S rRNA. Histogram of HL sizes (in nts) from the 2D representation of 16S rRNA in Figure 1. Flanking WC basepairs are not included.

three or four interactions, corresponding to one base pair and two or three stacking interactions per nt. Even nts on the ends of helices stack on at least one other base. Thus, almost all helical nts have at least two interactions, one pairing and one stacking. With regard to loop nucleotides, Figure 6 shows that almost 80% form two or more interactions with other nts in 16S rRNA—evidence that most loops form specific structures. On average, nucleotides in loops form 2.5 interactions per nt compared with 3.6 for those in helices. The stacking data (not shown) show that 85% of loop nts form one or more stacking interactions, with >60% forming two or more; on average, loop nts form 1.7 stacking interactions per nucleotide, compared with 2.5 for helix nts.

Less than 15% of loop nts form only one interaction and, surprisingly, only ~58 loop bases in 2AW7, or less than 4% of all nts in this 16S rRNA structure, form no classified pairing, stacking, or base–backbone interactions and are candidates for bases that completely 'bulge out' of the structure, as implied by 2D diagrams for all loop nts. As discussed in section *Quaternary Interactions in 16S rRNA*, detailed analysis of 3D structures of ribosomes in different states shows that most of these bases do in fact form some kind of interaction by stacking or H-bonding to tRNA, mRNA, or 23S rRNA, by binding to ribosomal proteins, or by forming as yet unannotated or unclassified base–ribose or perpendicular base–base interactions with other nts (Roy et al., unpublished observations). These interactions are discussed in more detail in the section 'Components of RNA nucleotides'. Classification of base–ribose and perpendicular interactions is an active area of research (Zirbel, private communication); once classes of interactions are agreed on, these interactions will also be annotated in NDB.

Regarding BPh interactions, which are base specific to a considerable extent (see section on *Base–Backbone Interactions*), we find that less than 4% of

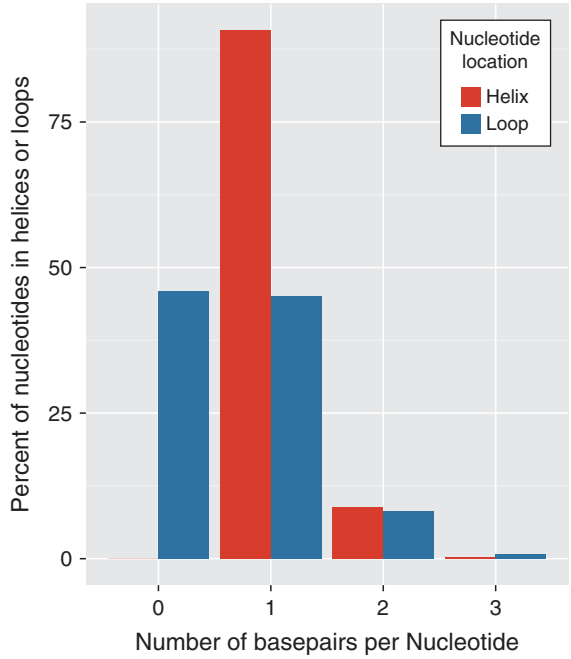


FIGURE 5 | Comparison of base pairs formed by nucleotides in loops versus helices in 16S rRNA. Histogram comparing number of base pairs (non-WC as well as WC) formed by nucleotides in loops (blue) versus helices (red).

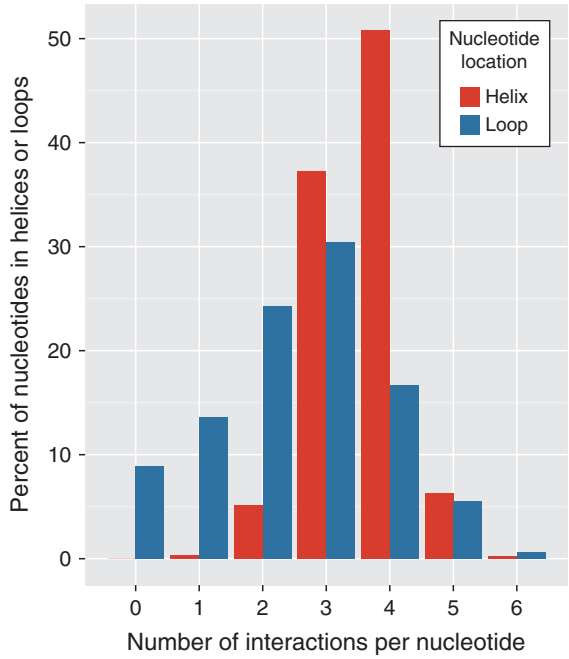


FIGURE 6 | Comparison of interactions formed by nucleotides in loops versus helices in 16S rRNA. Histogram comparing number of annotated interactions (base-pairing, base-stacking, and base–phosphate) formed by nucleotides in loops (blue) versus helices (red).

helical nucleotides (34 of 866) provide the base in such interactions, whereas 18.4% of loop bases do so (122 of 664). In addition, 18% of loop nts provide the phosphate groups for BPh interactions but only 4% of helix nts do so. This shows that loop nts also play a prominent role in mediating BPh interactions, which contribute substantially to stabilizing folded RNA structures.[29,30]

Taken as a whole, these data show that nts in loops and linkers form significant numbers of interactions of all types and, therefore, loop regions are generally well structured and contribute significantly to RNA 3D structure. Next, we analyze these interactions in more detail to identify the locations of interacting nucleotides.

Local versus Long-Range (LR) Interactions in Loops and Helices

To better understand the roles of loop and linker nts in RNA 3D structure, we next address these questions: How many structure-defining interactions occur between loop nts, how many between helical nts, and how many between loop and helical nts? How many of these interactions contribute to local structure and how many to the overall 3D architecture? To answer these questions, it is useful to classify interactions according to whether they are local or LR. LR interactions are of particular interest because they contribute to the overall folding of structured RNA molecules. Before presenting the data, we discuss how local and LR interactions can be distinguished by automated means to facilitate structural analysis, as implemented in the FR3D program suite.[27]

Defining Local and LR Interactions

Local interactions are those between nts belonging to the same helix or loop, or between adjacent elements of the 2D structure, whereas LR interactions are those between nts distant in the 2D. The farther apart two nts are in the 2D structure, the more WC base pairs they 'cross over' when they interact. By definition, all of the WC base pairs that define the 2D structure are 'nested' with respect to each of the others. Two WC pairs (i, k) and (m, n) are nested when either $i < m < n < k$ (as shown in Figure 7(a)) or $m < i < k < n$, where i, k, m, and n are the nt numbers in the linear RNA sequence (given by definition in the 5′ to 3′ direction) and $i < k$, $m < n$. In other words, nested interactions do not cross over each other in the 2D structure. However, if two interactions (i, k) and (m, n) are such that $i < m < k < n$, as shown in Figure 7(b), or alternatively, $m < i < n < k$, then they are said to 'cross' over each other. Once a comprehensive and consistent set of mutually nested WC pairs is identified and assigned to

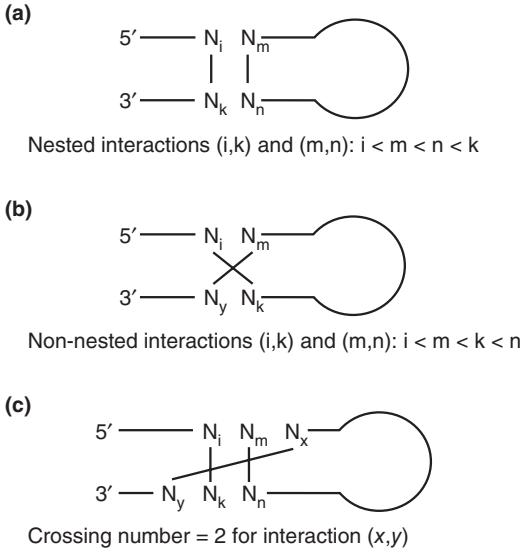


FIGURE 7 | Definition of crossing number for long-range interactions. Base pairs (*i, k*) and (*m, n*) are nested in (a) but non-nested in (b). Interaction (*x, y*) in (C) crosses over two nested base pairs, (*i, k*) and (*m, n*) and has crossing number equal to 2.

the 2D structure, one can measure how 'local' or 'LR' any other interaction is, by counting the number of nested WC pairs it crosses over. This measure is called the 'crossing number' of the interaction, and the larger it is, the more distant the interacting nts are in the 2D structure. By definition, all WC pairs that belong to the 2D structure have crossing number equal to zero. An example of an interaction (*x*, *y*) that crosses two nested WC pairs, (*i*, *j*) and (*m*, *n*), is shown in Figure 7(c). The pair (*x*, *y*) has crossing number equal to two. Whether interactions are labeled local or LR depends on the choice of cutoff for the crossing number. Certainly, the cutoff should be ≥1, so that interactions with crossing number zero are labeled 'local'. The smaller the cutoff, the more interactions will be defined as LR, but if the cutoff is set too low, interactions between nts that are very close to each other in the 2D structure and therefore best considered local, will be labeled LR. If the cutoff is set too high, interactions best classified as LR will be labeled local. We find that a cutoff equal to four provides a good compromise, so we label 'LR' all interactions with crossing number ≥4 (C. Zirbel, private communication).

Local versus LR Interactions in 16S rRNA

Using crossing number ≥4, we find 2307 local and 167 LR interactions in 16S rRNA. The relevant data for *E. coli* 16S rRNA are presented in Table 2, where we further group interactions as 'helix–helix', when the two interacting nts both belong to helices, 'loop–loop' when they are both found in loops, and 'loop–helix' when one is in a loop and other is in a helix. Most local nt interactions are helix–helix (1321 of 2307 or 57%); however, a significant number are loop–loop (538 of 2307 or 23%), while only 19% are loop–helix (448 of 2307). Most local loop–helix interactions are base-stacking interactions (386 of 448 or 86%), indicating that loop nts frequently stack with the WC pairs of adjacent helices. Many of these helices appear in the 2D to be very short, comprising only one or two base pairs (see Figures 1 and 2). Thus, for the most part, these very short helices are not isolated but are stabilized by stacking on bases in the adjacent HL, IL, or MHJ loops, many of which form non-WC base pairs. In fact, 16S rRNA contains 15 local AG *cis* WC (cWW) base pairs, most of which occur at the interfaces between loops and helices.[31] Readers should note that cWW pairs composed of noncanonical base combinations (i.e., other than AU or GC) are considered non-WC pairs. These and other non-WC pairs can be considered to form extensions of the secondary structure.[32] Local stacking interactions also play major roles in loops; they account for 336 of 538 or 62% of local loop–loop interactions. In fact, 20.4% (324 of 1586) of all local stacking interactions occur in loops.

With regard to local base-pairing and BPh interactions, most of these also are loop–loop rather than loop–helix. As shown in Table 2, there are 133 local loop–loop versus 38 local loop–helix base-pairing interactions and 81 local loop–loop versus 24 local loop–helix BPh interactions.

While there are far fewer LR than local interactions within 16S rRNA, the LR contacts are crucial for defining the 3D architecture. As discussed below, the actual folding into the correct 3D structure also requires other components, especially divalent magnesium ions.[33] The absolute number of LR interactions depends, of course, on the value of the crossing number cutoff used to calculate them, as discussed above. In generating the data reported here, we set the cutoff to 4 and found with this value that most LR interactions occur between nts distant in the 2D structure, i.e., nts belonging to different helical elements and even different structural domains. While 77% of local interactions involve helix nts (1769 of 2307 interactions), 93% of LR interactions involve loop nts (156 of 167) of which 52% (87 of 167) are between two loop nts. Only 11 of 167 or 6.5% of LR interactions are between two helix nts, and all of these are BPh or helix packing ('P-') interactions.[34] The latter involve highly conserved GU pairs that interlock on their sugar edges with AU or GC WC pairs and occur exclusively between helices, forming a small number of crucial and highly conserved inter-helical tertiary contacts in structured RNAs.[35,36]

TABLE 2 | Interactions between Helix and Loop nts, Local versus Long-Range

	Local Interactions				Long-Range Interactions				
	Base-Pairing	Base-Stacking	Base–Phosphate	Subtotals	Base-Pairing	Base-Stacking	Base-Phosphate	Helix Packing	Subtotals
Helix–helix	440	876	5	1321	0	0	6	5	11
Helix–loop	38	386	17	448	42	6	6		70
Loop–helix			7				16		
Loop–loop	133	324	81	538	35	33	18		86
Totals	611	1586	110	2307	77	39	46	5	167

Analysis of interactions between nts in helices, in loops, and between loops and helices. Interactions are classified as long-range if they cross four or more nested Watson–Crick (WC) base pairs. All other interactions are local. Base–phosphate interactions are separated, depending on whether the helix or loop nt contributes the base to the interactions. In the row labeled "Helix-loop", the helical nt contributes the base to BPh interactions while in the row "Loop-helix" the loop nt contributes the base.

Table 2 shows that LR loop–loop interactions include comparable numbers of pairing, stacking, and BPh interactions, whereas LR loop–helix interactions are dominated by non-WC base pairing. Most of these involve the sugar edges of the helix nts.

In BPh interactions, one nt contributes the base and the other the phosphate group (see section on Base–Backbone Interactions). The BPh interactions are separated in Table 2 according to whether the helix or loop nt contributes the base to the interaction. The row in Table 2 labeled 'Helix–Loop' includes interactions in which the helical nts contribute the base and vice versa for the row labeled 'Loop–Helix'. These data show that the loop nt usually provides the base in LR loop–helix BPh interactions, whereas this is reversed in local interactions. There are comparable numbers of LR base-pairing and LR BPh interactions between loops and helices as between two loop nts, but far more LR stacking interactions between two loop nts than between loop and helix nts (34 vs. 6).

Base-Stacking Interactions of Loop nts

What kinds of stacking interactions are observed in loops? As in helices, there is stacking between consecutive bases in the same strand, but in loops there are additional types of stacking. In many internal loops, there is extensive 'cross-strand' stacking between bases forming consecutive non-WC base pairs but located in opposite strands. Cross-strand stacking is a characteristic feature of many RNA 3D motifs, including recurrent sarcin–ricin motifs.[37] Extensive stacking is also observed between helical elements that meet at MHJ. As helices at MHJ can stack in different ways, the pattern of inter-helix stacking is a defining feature of each MHJ. Thus, stacking at junctions is a local interaction with important implications for the global RNA architecture.[38]

Among LR interactions, stacking occurs between the apical bases of certain HL, most commonly GNRA loops, and the IL forming 'platform' or loop receptor motifs.[39] LR stacking interactions reinforce tertiary base-pairing and BPh interactions to anchor the docking of these loops to their receptors. LR tertiary (3°) stacking interactions between two 'bulged out' bases also occur to stabilize the compact folding of the RNA.

Finally, intercalation of a bulged base from one loop into a binding site created by another loop results in two or more LR stacking interactions and generally at least one LR pairing or BPh interactions. The interaction between the D- and T-loops of tRNA involves intercalation of bulged bases from the D-loop in the T-loop.[40] All T-loop HL motifs provide sites for intercalation of a bulged base from another loop.[41]

Quaternary Interactions of 16S rRNA

Most, if not all, cellular RNAs interact specifically with one or more protein. *E. coli* 16S rRNA associates with 21 different ribosomal proteins (r-proteins) to form the small (30S) ribosomal subunit or 'SSU', and transiently with several translation factors. Figure 8 shows a histogram of SSU protein–RNA interactions for loop and helix nts in 16S rRNA (PDB file 2AW7). About 60% of nt-amino acid interactions involve loop nts, even though loop nts constitute just 42% of all 16S nts, demonstrating once again the important role of loop nts in the functional interactions of structured RNA.

The 30S particle also interacts with the large (50S) ribosomal subunit, or LSU, to form the functional 70S ribosome. Noncovalent interactions or 'bridges' that form between the subunits, many of which involve RNA, stabilize the assembled 70S ribosome.[24] Most of the 16S nts interacting with the LSU are loop nts or helix nts adjacent to loops.

About 13 of the approximately 58 'bulged-out' nts that do not interact with other nts in *E. coli* 16S rRNA, interact with SSU protein molecules. From structures that contain bound mRNA and tRNA,

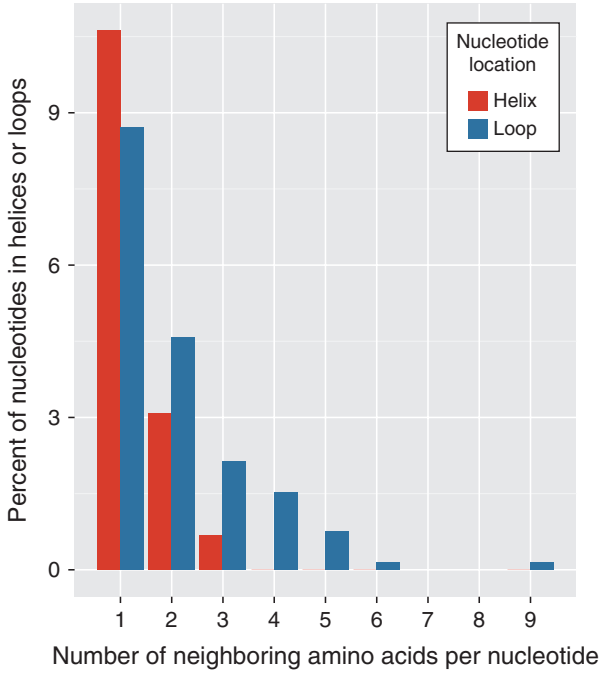


FIGURE 8 | Amino Acid interactions for loop versus helix nucleotides. Histogram of number of amino acids within 4 Å for nucleotides in loops (blue) versus helices (red) in *Escherichia coli* 30S ribosome.

we observe that three others interact with tRNA, three with mRNA, and A532, which is bulged out in 2AW7, forms a bridge to the head domain of 16S when the head clamps down on the mRNAs and the tRNA bound to the 16S A-site. Another bulged base, U723, forms a BPh interaction with the mRNA-Shine-Dalgarno helix. A702, which bulges out of the helix 23 kink-turn IL, interacts with 23S rRNA. Five others of these 58 'looped-out' bases form perpendicular base–base interactions and eight form base–ribose interactions. Thus, we see that only about half of the 58 nts (<30) are truly bulged out and of these ~21 facilitate tight turns in the 16S backbone. Examples of tight turns facilitated by looped-out bases have been known at least since the elucidation by NMR of the 3D structure of recurrent UNCG HL in the early 1990s.[42] In these HL, the second variable nt ('N') is bulged out, allowing a tight 180° turn in the backbone between the conserved U and the C bases. *E. coli* 16S rRNA contains seven UNCG loops. The remaining 16S bulged bases that facilitate tight turns do so in a variety of contexts. They include C280, U485, A702, U723, U1212, U1240, C1397, C1400, and A1503.

The conclusion of this overview of the prevalence and roles of loop nts, exemplified by 16S rRNA, is that almost all nts, whether they belong to 'loops', linkers, or helices in the 2D structure, form some kind of pairing, stacking, or base–backbone interactions. Moreover, interactions of loop nts constitute most of the crucial LR pairing, stacking, and BPh interactions that stabilize domain structures and architectures and mediate most interactions with proteins and other RNAs. Most of the very small number of truly 'looped-out' bases play a role in facilitating formation of sharp ~180° turns in the RNA backbone.

Box 1 entitled *Challenges in RNA Structural Bioinformatics* is provided for readers who are interested in the criteria used to assign nts to helices and loops in a consistent way to enable automated analysis of RNA 3D structures and extraction of loop motifs for detailed comparison and analysis. The description will enable readers to better understand how motifs are extracted from structures for the RNA 3D Motif Atlas,[26] now accessible through the revised NDB website.[23]

BOX 1

CHALLENGES IN RNA STRUCTURAL BIOINFORMATICS

Assigning nts to helices and loops in structured RNAs

While it is usually clear from accurate 2D diagrams, drawn with reference to 3D data, whether a given nt should be assigned to the 2D structure, or to linkers, HL, IL, or MHJ loops, there are ambiguous cases that require clear definitions to obtain consistent assignments. First, the choice of which helices to count as PK and which as part of the 2° structure is to some degree arbitrary and different choices will also change the number of HL, IL, MHJ loops, and linker regions.[43] For example, if helix 2 (formed by interactions between nts in the HL of helix 1 and the linker joining helices 27 and 28) is included in the 16S 2D, then H1 becomes the PK and a new four-way junction is defined, comprising H2, H3, H19, and H27. To build a comprehensive Motif Atlas, that includes all motifs, it may be necessary, in analyzing each 3D structure, to cycle systematically through alternative PK definitions. Generally, the helix formed by nts closest in the primary sequence is given precedence. Following this guideline, helix 1 is favored over helix 2 for inclusion in the 2D.

Second, there may be structural differences between PDB/NDB files for the same RNA molecule. These differences can be due to inconsistencies or errors in the 3D modeling,[3] or to

bona fide structural changes that occur upon binding of substrates or other ligands. Furthermore, some regions of RNA molecules are so dynamic that it is not possible to build unique atomic-resolution models for them. In these cases, the structure must be inferred by combining other data, including CSA, 3D modeling and structural studies of RNA fragments. Even where all 3D structural models agree, further questions arise regarding how to partition structures into helices and loops:

Should single (isolated) Watson–Crick pairs that occur between two loops be assigned to the loops or to the 2D structure?
Not infrequently, a single WC pair separates two loops. In which cases should the WC pair be assigned to one of the adjacent loops and when should it be considered part of the 2D structure? Alternatively, should the loops be merged into a single 3D motif and the WC pair treated as belonging to this larger motif? Examples of such 'isolated' WC base pairs in 16S rRNA occur between IL in H17 and in H44 and between IL and MHJ loops in H5, H19, H22, H23, H33, H40, and H42.

The solution we propose is to treat loops separated by single WC pairs as separate 3D motifs and assign the WC pair to the 2D, unless the nts of this WC pair form extensive interactions with nts of each of the adjacent loops. In the case of the H17 and H44, the interactions of the WC pairs are limited to base stacking with neighboring bases in the sequence. In these cases the sequence of the WC pair is less likely to be conserved and typical WC covariation is usually observed, and the adjacent loops are best treated as distinct motifs with the WC pairs assigned to the 2D structure. However, in the case of IL adjacent to five of the MHJ loops of 16S rRNA there are additional interactions involving the embedded WC pairs. Moreover, these pairs tend to be highly conserved in sequence, due to the additional interactions. In these cases we propose that the WC pair should be assigned to the MHJ loop and not to the 2D structure. This was done in constructing Table 1. All the WC pairs marked with red lines in Figure 1 are of this type.

Should GU 'wobble' pairs always be assigned to the 2D regardless where they are located?
Sequence analysis shows that most cWW GU pairs in an RNA structure covary with AU, UA, GC, and CG in homologous sequences, indicating that at these positions GU pairs are functionally indistinguishable from WC pairs. Therefore we assign cWW GU pairs that are adjacent to HL, IL, or MHJ loop motifs to the 2D structure, just as is done for AU and GC pairs. This is the approach taken in the construction of the 3D Motif Atlas.[26] Nonetheless, readers should note that sequence analysis shows that highly conserved cWW GU base pairs flank some HL, IL, and MHJ loop motifs and perhaps these should be considered parts of these motifs, at least operationally, for example, in designing self-assembling RNA molecules for RNA nanotechnology or in 3D structure prediction. This is an area of current research that requires integration of CSA and biophysical characterization. In fact, in a number of 3D motifs, exemplified by C-loops, the flanking WC pairs actually form base triples with the loop nts, and therefore could be considered part of the motif.[19] Even though the RNA 3D Motif Atlas extracts each loop motif together with its flanking WC pairs to facilitate analysis of their interactions and their sequence variations, the flanking WC pairs are still considered part of the 2D structure.[26]

How should other cis *Watson–Crick pairs (AA, AC, AG, CC, CU, and UU) be treated, especially when these form interfaces between helices and loops?*
AG *cis* Watson–Crick (cWW) pairs are not uncommon in structured RNAs; they tend to occur at the ends of helices flanking loop motifs, especially MHJ motifs.[31] They rarely occur within helices because they are larger than AU or GC WC pairs and therefore not isosteric to them, and when substituted for them, distort the local helical geometry and destabilize adjacent WC base pairs. Because cWW AG pairs almost always occur adjacent to loops and rarely covary with GC or AU, we assign all cWW GA pairs to the loops rather than the adjacent helices, consistent with the way they are treated in the 3D Motif Atlas. For consistency, other non-WC cWW base pairs, including AC, AA, UU, CU, and CC, are also treated this way.

In the next section, we describe the components of RNA nucleotides to provide a basis for understanding nucleotide-level interactions in RNA.

COMPONENTS OF RNA NUCLEOTIDES

The nucleotide is the basic unit of RNA structure. It is also the synthetic unit (or 'synthon') from which

RNA is produced in vivo. Each nucleotide is made of one of the four RNA bases, adenine (A), cytosine (C), guanine (G), or uracil (U), attached to ribose, a five-membered sugar ring, which in turn is linked to one or more phosphate groups.

Two features distinguish RNA from DNA and have important structural consequences: the substitution of -OH at the 2′-carbon of the ribose ring in RNA in place of -H in the deoxy-ribose ring of DNA, and the substitution of uracil (U) in RNA for thymidine (T) in DNA. The 2′-OH facilitates H-bonding along the 'sugar edge' of each RNA nt. U lacks the 5-methyl group of thymidine (T), located on the Hoogsteen edge of T; consequently U can form base pairs along its Hoogsteen edge whereas T cannot. These structural differences make RNA more versatile than DNA in forming interactions that support complex structures.

Bases and Base Edges: Each base is a nitrogen-rich, heterocyclic aromatic ring system that is planar in its equilibrium geometry and quite rigid, due to delocalization of π electrons. RNA bases are of two types, the pyrimidines (U and C), composed of one six-membered aromatic, heterocyclic ring, and the larger purines (A and G), composed of fused five- and six-membered rings. Structures of nts with the numbering of base positions are shown in Figure 9. Base ring atoms are numbered from 1 to 9 in purines and from 1 to 6 in pyrimidines. Exocyclic groups and attached hydrogen atoms are numbered according to ring position. Ribose atoms are numbered with primes, from 1′ to 5′, to distinguish their atoms from those of the base.

The planar RNA bases present three 'edges' studded with H-bonding donor and acceptor groups along which H-bonds can form with complementary groups on the edges of other bases, as well as with phosphate and ribose groups, amino acid side-chains and backbone atoms of proteins and small molecules. These edges are called the Watson–Crick (W), Hoogsteen (H), and sugar (S) edges.[2] Because there are just three edges, it is useful to represent RNA bases, purines as well as pyrimidines, as triangles (or more precisely, oblate triangular prisms) to describe and classify their diverse interactions as observed in RNA 3D structures.[43] The base edges are indicated for each RNA base in Figure 9. Further support for representing bases as triangles comes from the observation that bases can pair edge-to-edge with up to but no more than three other bases at the same time.[44] This explains the observation that no nt in 16S forms more than three base pairs (see Figure 5).

In the triangles representing the RNA bases, two of the vertices coincide with exocyclic H-bonding functional groups. The H and W edges meet at the vertex defined by the N4 or N6 amino groups of C and A, respectively. The corresponding atoms for U and G are the O4 and O6 carbonyl oxygen atoms. The W and S edges meet at the vertex defined by the O2 carbonyl group in C and U, the polarized C2–H2 of A, and the N2 amino group of G. The glycosidic bond connecting each base to the C1′ carbon of the ribose defines the third vertex of the triangle, where the H and S edges meet. The 2′-OH group, unique to RNA, extends the H-bonding capability of the sugar edge of each nucleotide.

The RNA bases are shown in Figure 9 with H-bond donor groups shaded with blue 'clouds' and the H-bond acceptor groups shaded with red 'clouds', representing centers of positive and negative charges, respectively. The ribose 2′-OH group is shaded in purple because it can serve as either H-bond donor or H-bond acceptor.

Ribose Sugar Rings: RNA bases are covalently attached to the 1′-carbon atoms of their respective ribose rings by single C–N bonds, the 'beta-glycosidic' bonds. In the beta configuration, each base is attached by the glycosidic bond to the same side of its ribose ring as the ribose 5′-carbon atom. The ribose ring is a five-carbon aldose sugar present in the furanose form. The flexible beta-glycosidic bonds allow for relatively free rotation of the RNA base relative to the ribose moiety. Rotation of the base gives rise to two distinct conformational classes of the nucleotide, called *anti* and *syn*.[45] In the *anti* configuration, the WC edges of the bases point away from the 5′-phosphate group. In the *syn* configuration, the base is rotated ~180° about the glycosidic bond, so the W edge faces the 5′-phosphate. G is the base that is most often observed in *syn* in 3D structures. In the *syn* configuration, the N2 amino group of G can H-bond to a 5′-phosphate oxygen atom. However, in polynucleotides, the *anti* configuration is much more common. The *anti* configuration is stabilized by weak H-bonding of the purine H8 or pyrimidine H6 atoms on the Hoogsteen edge to the 5′-phosphate group.[45] Whether a base is *syn* or *anti* affects the kinds of base pairs, it can form with nearby bases.[2] One of the most common discrepancies between published RNA 3D structures of the same molecule is the glycosidic bond configuration of certain nts, especially Gs.[3]

Phosphate Groups: Phosphate groups are derived from phosphoric acid (H_3PO_4), a weak acid with three dissociable protons that can form up to three phospho-ester linkages. In RNA and DNA, there is one phosphate per base–sugar unit ('nucleoside') and each phosphate links two nucleosides, by forming two phospho-ester bonds. Each phosphate is arbitrarily assigned to the nt to which it is attached

at the 5′-hydroxyl group. The phosphate groups link adjacent nucleoside units to each other by forming second phospho-ester linkages to the 3′-hydroxyls of the preceding one in the linear chain. To link two nucleosides, each phosphate loses two protons (H+) and forms two electrically neutral phospho-ester linkages. The remaining phosphate proton is acidic and largely dissociates at neutral pH. Therefore, each phosphate group in RNA carries a full negative (−1) electrical charge, which is delocalized over the two nonbridging oxygen atoms of the phosphate group, making each of them strong H-bond acceptors. Consequently, BPh interactions can be as stabilizing as GC base pairs and therefore structurally important.[29,30] Large RNA molecules such as 16S rRNA, therefore, have substantial negative charge that must be neutralized by H-bonding interactions and mobile positive ions in order for the RNA to fold into its functional structure. In vivo, charge neutralization depends on magnesium ions (Mg^{2+}), assisted by basic proteins, monovalent cations (principally K^+), and polyamines.[46]

RNA Backbone Conformations

The backbone of the RNA chain is very flexible and able to assume many different conformations because each nt contributes six covalent single bonds, including the C3′–C4′ and C4′–C5′ bonds of the ribose ring, and two C–O and two O–P bonds comprising the two phospho-ester linkages. The conformation of each nt is defined by the set of values of the dihedral angles of these six bonds together with that of the glycosidic bond.[45] The dihedral angles of each nt are assigned Greek letters, alpha (α) to zeta (ζ), starting with the P–O5′ bond of the 5′-phosphate and continuing consecutively with the O5′–C5′, C5′–C4′, C4′–C3′, C3′–O3′, and ending with the O3′–P bond of the 3′-phosphate. Rotations about the glycosidic bond ('chi' or χ) define the seventh dihedral angle. These seven dihedral angles define a very large, seven-dimensional conformational space for each nucleotide. However, there are extensive correlations between the values of the dihedral angles, so that relatively few regions of this seven-dimensional space are populated to a significant extent in observed or theoretically possible structures.[47] It is most convenient to cluster conformations by parsing the RNA backbone in overlapping 'suites' that extend from one sugar to the next, rather than from one phosphate group to the next.[47] Analysis of atomic-resolution experimental data using this approach enabled researchers to determine that the backbone conformations of structured RNA molecules are 'rotameric'.[47,48] In other words, most observed conformations can be grouped in well-defined clusters of conformations, many of which are characteristic of particular structural motifs. This analysis identified 42 recurrent rotamer clusters in RNA structures, each of which was assigned a two-symbol representation.[49] While the experimental data are still incomplete and we can expect to discover additional energetically accessible conformations or rotamers, these are likely to be rare or similar to ones already reported. Conformational analysis of each suite in each atomic-resolution 3D structure is now a standard reporting function of NDB.[23]

As many conformations have similar energies, the selection of local nt conformations as RNA molecules fold is guided largely by the specific interactions formed by the bases, especially base-pairing. We turn to these next.

NUCLEOTIDE INTERACTIONS

For an RNA chain to fold into a distinct 3D structure, the nucleotides must form specific and energetically favorable noncovalent contacts. RNA nucleotides can interact with each other in many different ways. These can be broadly classified as (1) base with base, (2) base with ribose sugar, (3) base with phosphate, (4) ribose with ribose, (5) ribose with phosphate, and (6) phosphate with phosphate.

We begin with base–base interactions and then consider base–phosphate and base–sugar interactions. Base–base interactions are the most sequence specific, and therefore the most useful for predicting RNA structures or designing RNA sequences to fold in desired ways for RNA nanotechnology applications. BPh and base–ribose interactions are base specific for one of the two interacting nucleotides. Ribose–ribose, ribose–phosphate, and phosphate–phosphate interactions are not sequence-specific and consequently difficult to classify informatively for RNA prediction and design, and will not be discussed further. Nonetheless, all these interactions contribute to the energetics of RNA folding, and we can expect that analysis of their local contexts, recurrent geometries, and relative frequencies will prove valuable in the evaluation of *de novo* modeled and predicted 3D RNA structures. Although the negatively charged phosphate groups repel each other, precluding direct contact, the phosphates can indirectly interact through bridging divalent ions, especially Mg^{2+}, which is maintained at millimolar (mM) concentration in cells. Metal-bridging interactions are very important for the compact folding of complex RNA architectures and have been reviewed elsewhere.[50,51]

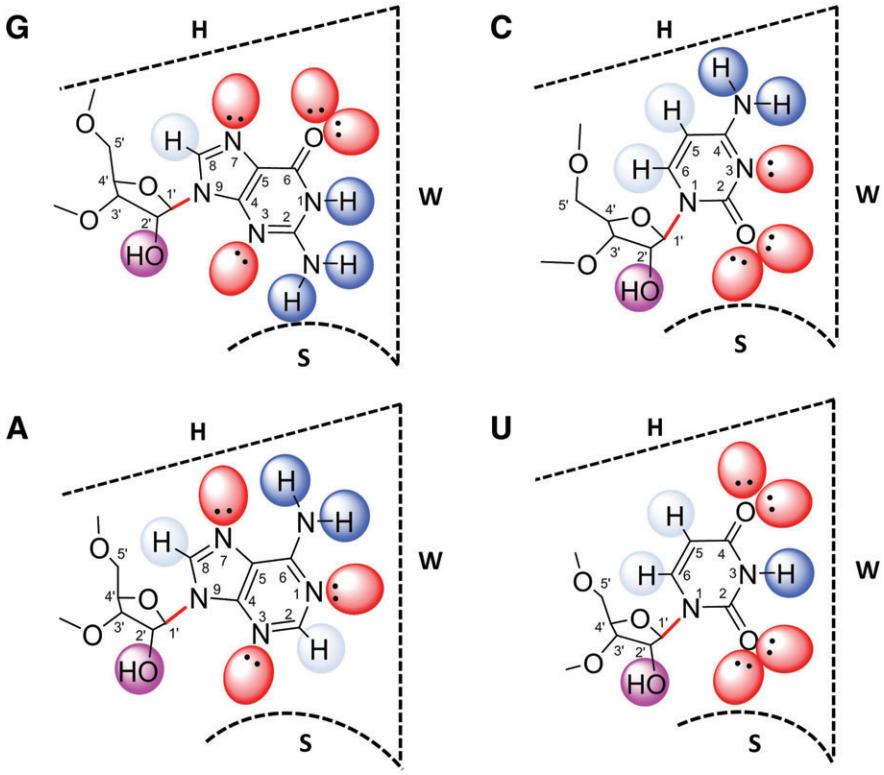


FIGURE 9 | Structures of RNA nucleotides. RNA nts G, C, A, and U each have three edges along which H-bonding takes place. Hoogsteen (H), Watson–Crick (W), and sugar (S) edges are marked with dotted lines. Base ring atoms are numbered from 1 to 9 for purines and 1 to 6 for pyrimidines. Ribose carbons are numbered 1′ to 5′. Exocyclic groups and attached hydrogens are numbered according to ring position. H-bond donors are highlighted with blue and H-bond acceptors with red.

Base–Base Interactions

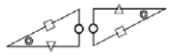
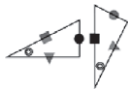
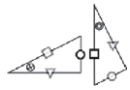
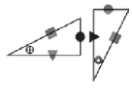
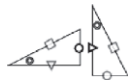
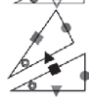
RNA bases interact with each other primarily by edge-to-edge H-bonding ('base-pairing') and by face-to-face van der Waals interactions ('base-stacking'). In principle, bases can also interact edge-to-face ('perpendicular interactions'). In crystals of aromatic hydrocarbons compounds, perpendicular interactions are frequently observed.[52] Perpendicular interactions can be found in RNA 3D structures using the FR3D motif search program.[28] Although relatively rare, they deserve to be studied systematically to determine recurrent geometries and sequence propensities.

Base-Pairing Interactions

Base-pairing results from H-bonding interactions between the edges of two RNA bases, a consequence of the geometric regularities of the RNA bases and the presence of at least two H-bond donor or acceptor groups on each base edge. H-bonds are attractive electrostatic interactions between positively charged H-atoms that are covalently bonded to highly electronegative atoms, primarily O and N in biomolecules, and electronegative atoms bearing unpaired electrons, also O or N for the most part. Individual H-bonds are highly directional but relatively weak noncovalent interactions. Therefore, stable association between the edges of two bases generally requires forming two or more H-bonds. Specificity is achieved because H-bonds are directional and require juxtaposition of complementary donors and acceptors, as exemplified by the H-bonding between the W edges of G and C or A and U to form canonical WC base pairs.

RNA bases can form many types of base pairs, in addition to the well-known WC pairs, because they have three distinct edges available for H-bonding,[2,5] the Watson-Crick edge (W), the Hoogsteen edge (H), and the sugar edge (S) as shown in Figure 9. The base edges can interact with each other in all combinations, W with W, W with H, W with S, H with H, H with S, and S with S, and in each of two orientations, *cis* and *trans*, to create 12 geometrically distinct types of base pairs.[2] These 12 types of base-pairing geometries are summarized schematically in Table 3 using right triangles to represent RNA bases. The Hoogsteen edge corresponds to the hypotenuse of each triangle in Table 3, and the marked vertex indicates the location of the ribose sugar.

TABLE 3 | The 12 Geometric Families of RNA Base Pairs

No.	Glycosidic Bond Orientation	Interacting Edges		Abbreviation	Symbol	Triangle Abstraction
		NT1	NT2			
1	*Cis*	Watson–Crick	Watson–Crick	cWW		
2	*Trans*	Watson–Crick	Watson–Crick	tWW		
3	*Cis*	Watson–Crick	Hoogsteen	cWH		
		Hoogsteen	Watson–Crick	cHW		
4	*Trans*	Watson–Crick	Hoogsteen	tWH		
		Hoogsteen	Watson–Crick	tHW		
5	*Cis*	Watson–Crick	Sugar edge	cWS		
		Sugar edge	Watson–Crick	cSW		
6	*Trans*	Watson–Crick	Sugar edge	tWS		
		Sugar edge	Watson–Crick	tSW		
7	*Cis*	Hoogsteen	Hoogsteen	cHH		
8	*Trans*	Hoogsteen	Hoogsteen	tHH		
9	*Cis*	Hoogsteen	Sugar edge	cHS		
		Sugar edge	Hoogsteen	cSH		
10	*Trans*	Hoogsteen	Sugar edge	tHS		
		Sugar edge	Hoogsteen	tSH		
11	*Cis*	Sugar edge (priority)	Sugar edge	cSs		
		Sugar edge	Sugar edge (priority)	csS		
12	*Trans*	Sugar edge (priority)	Sugar edge	tSs		
		Sugar edge	Sugar edge (priority)	tsS		

Each geometric base-pair family (numbered 1 to 12 in column 1 is defined by the interacting edges of the bases and the relative orientation of the glycosidic bonds (columns 2–4). Abbreviations and symbols for representing each base-pair family in text and secondary structures are shown in columns 5 and 6. To construct symbols for base pairs, circles represent WC edges, squares Hoogsteen edges, and small triangles Sugar edges. Solid symbols represent cis pairs and open symbols trans pairs. Column 7 shows an abstract representation of each base-pair family using right triangles to represent the bases, where the hypotenuse corresponds to the Hoogsteen edge of each base. Each base edge is labeled with its symbol (circle, square, or triangle), and the position of the glycosidic bond is indicated by the interior circle in one vertex.

Table 3 also shows symbols developed to mark base pairs in extended 2D diagrams. In these symbols, circles represent W edges, squares H edges, and triangles S edges. Base pairs involving distinct edges are represented by two symbols connected by a line. For example, a *trans* WC/Hoogsteen base pair (tWH) is represented with a circle linked to a square. The symbol is placed in 2D diagrams, so that the circle faces the letter representing the base bonding with its W edge and the square the base bonding with its H edge. When both bases use the same edge to form the pair, a single symbol suffices. Filled symbols represent *cis* base pairs and open symbols denote *trans* base pairs. Annotating 2D diagrams in this way conveys crucial 3D interaction information to readers, at least for local interactions. Representing all LR interactions observed in the 3D structure is more difficult because the 2D diagram can become very cluttered with symbols if these are not drawn carefully.

Exploring Non-WC Base-Pairing

Within each base-pair family, different base combinations (A with U, G with C, U with U, etc.) can form geometrically similar base pairs, depending

on the arrangement of H-bond donor and acceptor groups on each base edge. To help the reader explore the base-pairing possibilities of RNA bases, we have prepared planar structures of the four nucleotides that can be color xeroxed onto transparencies and cut out individually (see Figures 10 and 11). Each nt is reproduced in two distinct orientations to aid in forming both *cis* and *trans* base pairs. Possible base pairs can be explored by aligning bases in various combinations along their edges to identify potentially stable H-bonding arrangements. To guide the reader in this exercise, lone pair electrons have been included on the exocyclic carbonyl oxygen and ring imine nitrogen atoms; these serve as electron-rich, H-bond acceptors and are shaded red. Readers will note that the lone pairs of exocyclic amino groups (GN2, CN6, and AN6) are not shown in these diagrams, because these electrons are delocalized into the aromatic rings and are generally not available as H-bond acceptors, except in special cases.[30,31]

The H-bond acceptor groups are colored red in Figures 10 and 11, to reflect their overall negative charge. H-bond donor groups, comprising H-atoms covalently bonded to electronegative oxygen or nitrogen atoms, are colored blue, reflecting their overall positive charges. The goal when manipulating the colored transparencies is to obtain stable base pairs by juxtaposing H-bond donors and acceptors, so as to form at least two H-bonds. Each red-colored functional group should partly overlap a blue-colored functional group while avoiding any red-with-red or blue-with-blue juxtapositions. The 2′-OH (hydroxyl) groups are colored purple to indicate that they can serve either as H-bond donors or acceptors. Moreover, a single hydroxyl group can simultaneously interact with an H-bond acceptor and at least one H-bond donor. The –OH single bond can rotate as needed to optimize H-bonding.

To get started, readers can arrange the transparent cutouts of uracil and adenine having W edges facing in opposite directions to form the canonical AU *cis* WC (cWW) base pair. This pair is shown in Figure 12(a) to illustrate the complementary arrangements of color-coded H-bond donating and accepting groups, red opposite blue. The H-atom attached to C2 of adenine (AH2) is colored light blue to show that it is slightly polarized and capable of weak H-bonding with O2 of uracil. Other H-atoms covalently bonded to carbon atoms that are sufficiently polarized to form weak H-bonds are also colored light blue.

To form the *trans* WC (tWW) UA pair, the A and U cutouts printed with W edges facing in the same direction can be used. Bringing together the W edges forms an AU *trans* WC (tWW) base pair, that can be compared with Figure 12(b). This pair differs from the canonical (*cis*) WC AU base pair in the mutual orientation of the glycosidic bonds. In the *cis* pair, the glycosidic bonds are on the same side of the base-pair axis running through the base centers parallel to the H-bonds, while in the *trans* pair, the glycosidic bonds are on opposite sides of this axis.

Next readers can reorient the W edge of the U so that it faces the Hoogsteen edge of the A, juxtaposing complementary H-bond donor and acceptor groups. Results can be compared with the pairs shown in Figure 12(c) and (d). There are two possible results, the *cis* and the *trans* WC/Hoogsteen pairings (cWH or tWH UA). These pairs are stabilized by two strong H-bonds, involving NH donors, and one relatively weak bond involving AH8. These base pairs have the same base combination (UA) as the cWW and tWW pairs, but involve the Hoogsteen edge of A, illustrating that each distinct combination of edges and glycosidic bond orientations produces a different base-pairing geometry and therefore a different pair, even when the base combination remains the same (here, UA).

Readers are encouraged to try other base combinations, to see how many form stable pairs for each geometric family, and to check their base-pairing models against the online RNA Base Pair Catalog, which contains structures of all observed and predicted base pairs organized by geometric family (http://ndbserver.rutgers.edu/ndbmodule/services/BPCatalog/bpCatalog.html). The base combinations that form pairs in each pairing geometry are summarized in Table 4.

The Sugar Edge

A key concept for understanding sugar-edge base-pairing is the role of the ribose 2′-OH group, which can serve as an H-bond donor or acceptor, and often, both simultaneously, on account of the free rotation of the hydroxyl group about the C2′–O2′ single bond and the presence of the nonbonding electron orbitals on the oxygen. Consequently, many different pairs can form along the sugar edges of nts.

Protonation of the WC Edge of A and C

Another important concept is that the imine nitrogen atoms, AN1 and CN3, on the W edges of A and C, which are normally unprotonated and act as H-bond acceptors, can be protonated at a modest energetic cost, when required by the context, to convert these groups to H-bond donors. This allows certain base combinations to form base pairs that cannot form when A or C are in their unprotonated forms. Protonation confers a positive charge to the resulting base

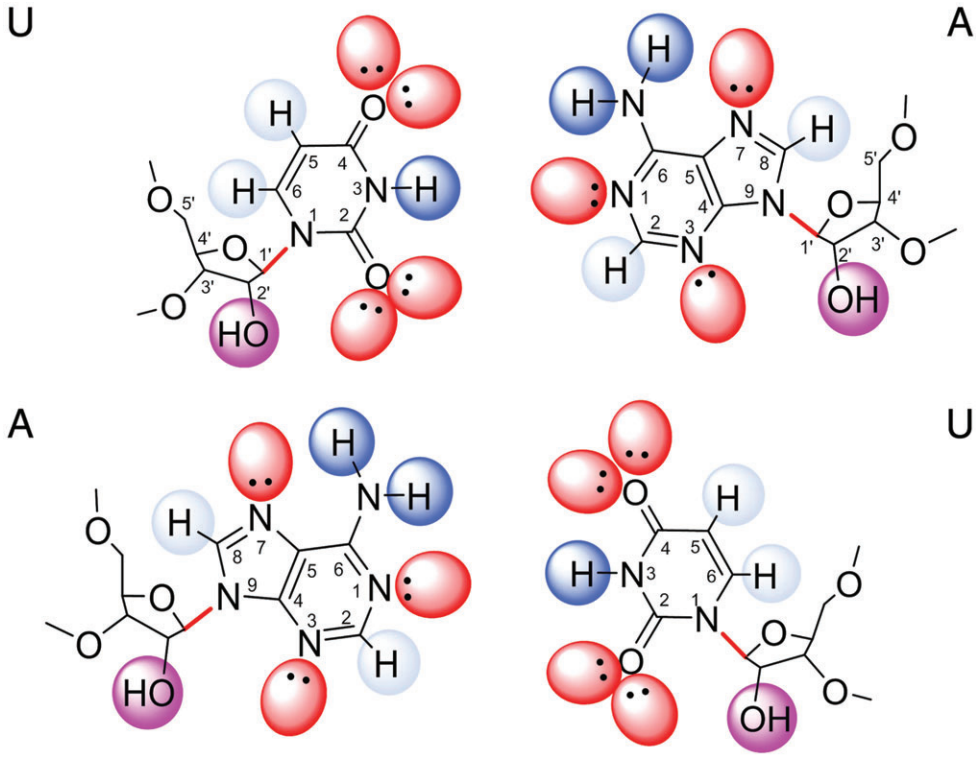


FIGURE 10 | U and A nucleotides to print on transparencies. U and A nts in two orientations to print on transparencies for making base pairs in different geometries by juxtaposing H-bonding donor (blue) and acceptor (red) groups.

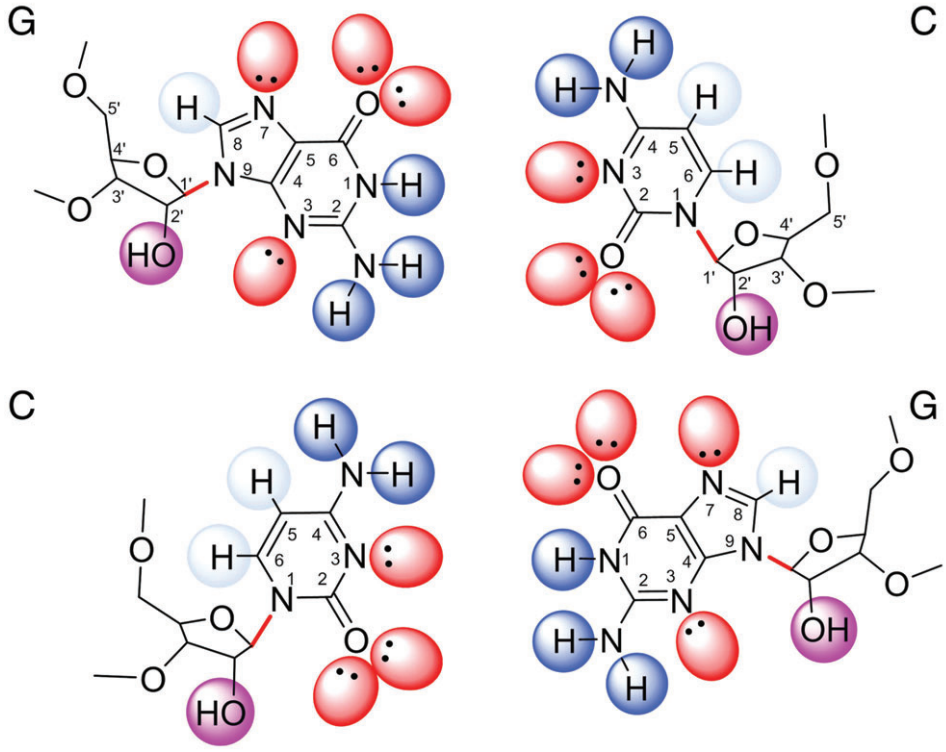


FIGURE 11 | G and C nucleotides to print on transparencies. U and A nts in two orientations to print on transparencies for making base pairs in different geometries by juxtaposing H-bonding donor (blue) and acceptor (red) groups.

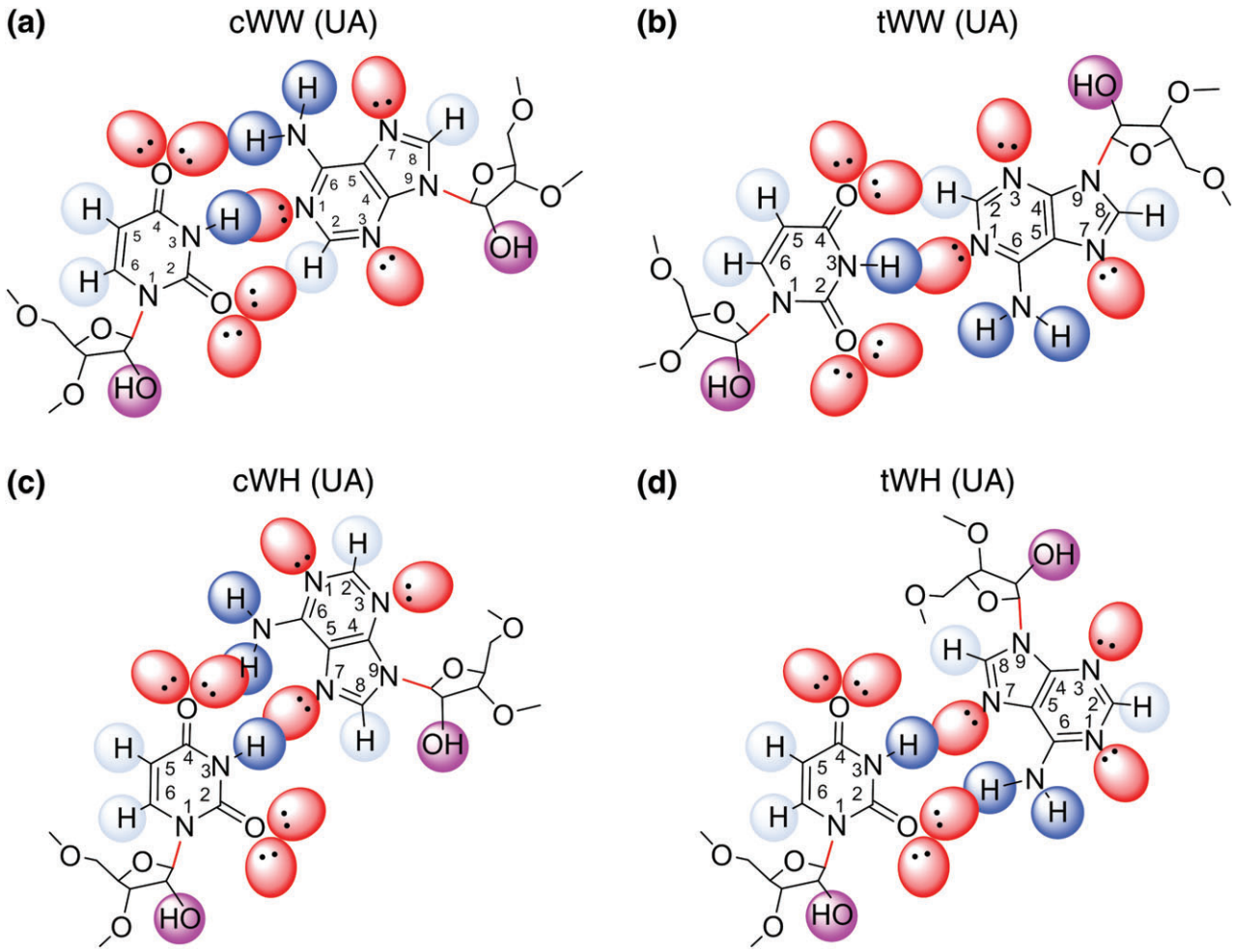


FIGURE 12 | Representative base pairs. UA base combination in four different base-pairing geometries: (a) *cis* Watson–Crick/Watson–Crick (cWW); (b) *trans* Watson–Crick/Watson–Crick (tWW); (c) *cis* Watson–Crick/Hoogsteen (cWH); and (d) *trans* Watson–Crick/Hoogsteen (cWH).

pair, which can help stabilize accumulations of negative charge, as occur, e.g., in the close packing of phosphate groups or during chemical reactions.[53,54] Allowing for H-bonding to 2′-OH and protonation of A and C W edges expands the number of base pairs that one can construct. The RNA Base Pair Catalog includes such pairings.

Specifying non-WC Base Pairs: Base Combinations and Base-Pair Families

As readers experiment with the RNA base cut outs to construct base pairs, they will quickly discover that the same base combination, e.g., UA in the examples above, can form several different types of base pairs. To keep track of the base pairs that form, readers should note the geometric base-pair type or family, in addition to the base combination. By giving the interacting edges of the bases and the relative orientations, *cis* or *trans*, of the glycosidic bonds, one can completely specify the base-pair family. To specify individual pairs, both the base combination *and* the base-pair family are needed, e.g., 'tWH UA', 'cWW UA', or 'cWW CG'.

It is necessary to specify the base combination because in each base-pair family, different base combinations (up to 16) can form and for each base combination up to 12 different base-pair types are available. For example, 15 of the 16 base combinations (all except GG) can form cWW base pairs, while the base combination AG can form 11 of the 12 different types of base pairs, all families except tWW. Referring to non-WC base pairs as 'mismatches', as is frequently done in the literature, is imprecise and misleading because for stable non-WC pairings, as for WC pairs, at least two H-bonds form between the interacting base edges. The only 'mismatches' are those base combinations that cannot form a particular base-pair type. For example, GG is a 'mismatch' for cWW but a stable 'match' for tWW.

TABLE 4 | Base Combinations that Form Base Pairs in Each Geometric Family

	AA	AC	AG	AU	CA	CC	CG	CU	GA	GC	GG	GU	UA	UC	UG	UU
cWW	$I_{1.4}$	$I_{1.2a}$	$I_{1.3}$	$I_{1.1}$	$I_{1.2b}$	$I_{1.6}$	$I_{1.1}$	$I_{1.5}$	$I_{1.3}$	$I_{1.1}$		$I_{1.2a}$	$I_{1.1}$	$I_{1.5}$	$I_{1.2b}$	$I_{1.7}$
tWW	$I_{2.7}$	$I_{2.4}$		$I_{2.2}$	$I_{2.3}$	$I_{2.9}$	$I_{2.6}$	$I_{2.8}$		$I_{2.5}$	$I_{2.7}$	$I_{2.4}$	$I_{2.1}$	$I_{2.8}$	$I_{2.3}$	$I_{2.9}$
cWH			$I_{3.3}$	$(I_{3.3})$		$I_{3.2}$	$I_{3.1}$	$(I_{3.2})$	$I_{3.3}$		$I_{3.4}$		$(I_{3.3})$		$I_{3.1}$	$I_{3.2}$
tWH	$I_{4.3}$		$I_{4.3/4.2}$		$I_{4.2}$	$I_{4.1}$	$I_{4.2}$				$I_{4.5}$	$I_{4.3}$	$I_{4.1}$		$I_{4.4}$	$I_{4.2}$
cWS	$I_{5.1}$	$I_{5.1}$	$I_{5.1}$	$I_{5.1}$	$I_{5.2}$	$I_{5.2}$	$I_{5.2}$	$I_{5.2}$	$I_{5.3}$	$I_{5.3}$	$I_{5.5}$	$I_{5.3}$	$I_{5.4}$	$I_{5.4}$	$I_{5.4}$	$I_{5.4}$
tWS	$I_{6.1}$	$I_{6.2}$	$I_{6.2}$	$I_{6.1}$	$I_{6.2}$	$I_{6.1}$	$I_{6.3}$	$(I_{6.1})$		$I_{6.3}$		$I_{6.3}$	$I_{6.3}$	$I_{6.4}$	$I_{6.4}$	$I_{6.4}$
cHH			$I_{7.2}$				$I_{7.1a}$		$I_{7.3}$	$I_{7.1b}$	$I_{7.1}$					
tHH	$I_{8.1}$	$I_{8.1}$	$I_{8.3}$	$I_{8.3}$	$I_{8.1}$		$I_{8.1}$	$I_{8.3}$	$I_{8.2}$	$I_{8.1}$	$I_{8.4}$		$I_{8.2}$	$I_{8.2}$		
cHS	$I_{9.1}$	$I_{9.1}$	$I_{9.1}$	$I_{9.1}$	$I_{9.1}$	$I_{9.1}$	$I_{9.2}$	$I_{9.1}$	$I_{9.1}$		$I_{9.1}$		$I_{9.3}$	$I_{9.1}$	$I_{9.1}$	$I_{9.1}$
tHS	$I_{10.1}$	$I_{10.1}$	$I_{10.1}$	$I_{10.1}$	$I_{10.1}$	$I_{10.1}$		$I_{10.1}$			$I_{10.2}$		$I_{10.2}$		$I_{10.2}$	
cSS	$I_{11.1}$	$I_{11.1}$	$I_{11.1}$	$I_{11.1}$	$I_{11.1}$	$I_{11.1}$	$I_{11.1}$	$I_{11.1}$	$I_{11.1}$	$I_{11.1}$	$I_{11.1}$	$I_{11.1}$	$I_{11.1}$	$I_{11.1}$	$I_{11.1}$	$(I_{11.1})$
tSS	$I_{12.1}$	$I_{12.1}$	$I_{12.1}$	$I_{12.1}$					$I_{12.2}$	$I_{12.2}$	$I_{12.2}$	$I_{12.2}$				

The base combinations (columns) that form observed or predicted base pairs of the geometric type indicated in each row are marked with '$I_{n.m}$' where n runs from 1 to 12 and refers to the base-pair family (i.e., cWW, cWH, tHS, etc.) and m enumerates the isosteric subgroups within that family. Base combinations that do not form base pairs are indicated by blank cells.

It is better, therefore, to simply speak of the 'allowed' and 'nonallowed' base combinations for each pairing geometry. As noted above, some base combinations require protonation to form. In such cases, the energy penalty for protonation is compensated by the more favorable binding energy.

We summarize all the data found in the Base-Pair Catalog in Table 4, which shows all base combinations that form in each family. The base combinations (columns of Table 4) that form stable pairs in each base-pair geometry or 'family' (rows of Table 4) are indicated with an entry marked '$I_{n.m}$' where n runs from 1 to 12 and refers to the base-pair family (i.e., cWW, cWH, tHS, etc.) and m enumerates the isosteric subgroups within that family. If no stable pair forms for a particular base combination and geometry, the corresponding cell in Table 4 is left empty. Each column shows the base-pair geometries that each base combination can adopt. The 3D structures of base combinations forming pairs in each geometric family can be viewed and downloaded as PDB files from the pages of the online Base-Pair Catalog.

Regarding Table 4, readers should note that for base-pair geometries that involve different edges (e.g., cWH or tHS) the order in which the bases are listed matters. For example, in the tWH AG base pair, the A interacts with its W edge and the G with its Hoogsteen edge. Consequently, 'tWH AG' and 'tWH GA' refer to different base pairs. However, 'tWH AG' and 'tHW GA' refer to the same base pair. However, tHW pairs do not have their own row in Table 4 because they are redundant with tWH pairs.

Base-Pair Isostericity and Sequence Variation

The classification of base pairs into geometric families based on edges provides the framework for making sense of the base substitutions observed in structural alignments of recurrent RNA 3D motifs and sequence alignments of RNA homologs (e.g., 16S rRNAs from different micro-organisms). To illustrate the basic ideas, we first consider the AU, UA, GC, and CG WC (cWW) base pairs that constitute the geometrically regular, antiparallel double helices of RNA. RNA helices are regular precisely because AU and GC pairs can substitute for each other with little or no distortion of the helical geometry. They are said to be isomorphic or 'isosteric' to each other in the sense of occupying the same space between the backbone atoms of the two strands of the helix. The property of being isosteric can be quantified using a measure called the IsoDiscrepancy Index (IDI), which depends on three distinct geometric features of base pairs, as illustrated in Figure 13.[3] The geometric classification of base pairs is useful because only base pairs belonging to the same family are isosteric by qualitative and quantitative criteria (i.e., the IDI). Isosteric pairs substitute for each other without distorting the local RNA 3D structure. The IDI was calibrated by carrying out statistical analysis of IDI values of isosteric and near isosteric AU, GC, and GU base pairs extracted from high-quality structures. Quantitative analysis with IDI values confirms that for two base pairs to be isosteric, they must belong to the same geometric family.[3]

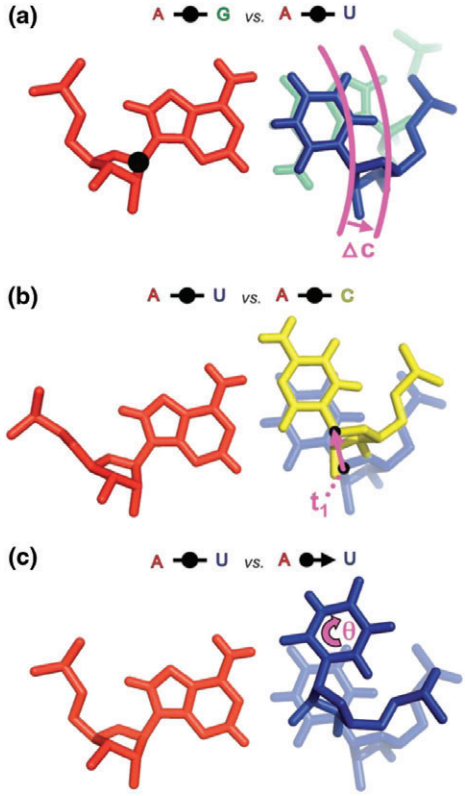


FIGURE 13 | Calculation of IsoDiscrepancy Index (IDI) to compare geometries of base pairs. The IDI is illustrated using nonisosteric base pairs. To calculate the IDI for two base pairs, the bases designated 'first base' in each base pair are superposed (bases on the left in each panel), and then the following three quantities are evaluated, normalized, and summed: (1) The difference, Δc, in the intra-base pair C1′–C1′ distances, illustrated for nonisosteric cWW AG and AU. (2) The inter-base pair C1′–C1′ distance, t1, between the C1′ atoms of the second bases of the base pairs, illustrated for the near isosteric cWW AU and AC base pairs. (3) The angle, theta, about an axis perpendicular to the base pair plane, required to superpose the second bases, illustrated using nonisosteric cWW AU and cWS AU base pairs. Adapted from Stombaugh et al. (3) with permission.

All base pairs from the same geometric family have equal or very similar values of the angle describing the mutual orientations of the glycosidic bonds of the interacting bases; however they can differ by the other geometric measures used to define the IDI (see Figure 13). Therefore, not all base pairs belonging to the same geometric family are necessarily isosteric. For example, purines and pyrimidines differ considerably in size, so substituting a purine by a pyrimidine or vice versa may change the space occupied by the base pair within its structural context, as measured by the distance between C1′ atoms of the interacting nts. For example, substituting A for C in a cWW CG base pair to produce cWW AG significantly increases the C1′–C1′ distance from ~10.4 Å, typical for AU or GC, to ~12.3 Å. Consequently, this substitution distorts the helix and destabilizes neighboring base pairs in the helix. Consistent with this prediction, substitution of base combinations AG or GA for cWW CG or AU pairs within regular helices of homologous RNA molecules is rare, and cWW AG pairs occur almost exclusively on the ends of helices, often adjacent to junction motifs where they provide a large surface for stabilizing inter-helix stacking interactions.[31] An example is the cWW AG pair, found in most tRNAs at the top of the anticodon stem-loop and flanking the tRNAs MHJ.[55] As mentioned above, 16S rRNA contains 15 cWW AG pairs.

Two base pairs from the same family may not be isosteric for a second reason: to align properly the H-bond donor and acceptor groups of the substituted bases, it may be necessary to shift the bases laterally relative to the positions of the original bases. Such a shift is needed when U substitutes for C to form a cWW GU pair. The U shifts laterally toward the major (deep) groove relative to the G, so that the H-bond donor and acceptor groups on the W edges of G and U can align. This geometric change is not as disruptive to adjacent base pairs, as the change in C1′–C1′ distance that occurs with cWW AG pairs. Consequently, GU pairs are observed quite frequently within WC helices in RNA molecules. The cWW GU pair is classified 'near isosteric' to cWW GC and AU. However, the lateral shift breaks the symmetry of the cWW geometry, so cWW GU and UG pairs are geometrically distinct and not isosteric or even near isosteric to each other. In each base-pair family, different sets of base combinations form pairs that are isosteric to each other. For example, the base combinations AG and CU are isosteric in the tHS family although they are not isosteric in the cWW family.[5,56]

These considerations predict that those base substitutions that result in isosteric or near isosteric base pairs should be much more likely to occur at equivalent positions of homologous molecules than those that produce nonisosteric ones. The rRNAs of *E. coli* and *Thermus thermophilus* provide an ideal, large-scale test case for this hypothesis, because these two bacteria are phylogenetically and ecologically divergent. Nonetheless, their 3D rRNA structures are sufficiently conserved that a large percentage of base pairs in the two structures can be structurally aligned and compared.[3] The base pairs of 5S, 16S, and 23S rRNAs from high-quality structures of 70S ribosomes of these species were aligned manually to identify corresponding base pairs. It was found that over 90% of base pairs could be aligned and analyzed using the IDI. This analysis revealed that 72% of corresponding base pairs in the 5S, 16S, and 23S rRNAs of *E. coli* and *T. thermophilus* were isosteric

and unchanged in sequence, 19% were isosteric and different in sequence, 7% were near isosteric, and only 2% were nonisosteric.[3] In total, almost 98% of the alignable base pairs in the two structures were isosteric or near isosteric. This result provides strong support for the generalization that base substitutions in homologous RNA molecules or recurrent structural motifs are constrained to preserve isosteric base pairings, whether the pairings are WC or non-WC.

NDB provides annotations for base pairs and other nt interactions for all atomic-resolution 3D structures.[23] The annotations are accessible from the main page of each structure.

Base Triples

Base triples are common submotifs in RNA 3D motifs. All base triples can be decomposed into combinations of base pairs, in which a central base is paired to each of the other two bases of the triple using a different base edge. All in all, there are 108 different geometric base-triple families of which 68 have been observed in experimental structures.[44] The base-pair families that compose each base triple define the base-triple family to which it belongs. A catalog of observed base triples can be accessed online through NDB. New base triples can be found by symbolic search using FR3D.[28]

Base-Stacking Interactions

The energetically most stabilizing contributions to RNA structure are provided by the hydrophobic van der Waals forces mediating stacking of the faces of RNA bases on each other.[30] Because RNA bases lack rotational symmetry, the two faces are not equivalent. Therefore, two RNA bases can stack face-to-face in four distinct ways, depending on which base faces come into contact. The faces are distinguished by reference to the orientation of each base in the WC helix, in which all bases are in the *anti*-glycosidic conformation; the '5′-face' points toward the 5′-end of the strand and the '3′-face' points toward the 3′-end of the strand.[27,43] A systematic analysis of the sequence propensities of base-stacking in all possible geometries and contexts is still lacking. NDB provides base-stacking annotations for all RNA structures that can be used as a basis for such studies.

Base–Backbone Interactions

BPh Interactions

As each nt bears a full negative electrical charge, RNA molecules need to overcome electrostatic self-repulsion to achieve compact folding. The negative charge of each nucleotide is largely concentrated on the two nonbridging oxygen atoms of the

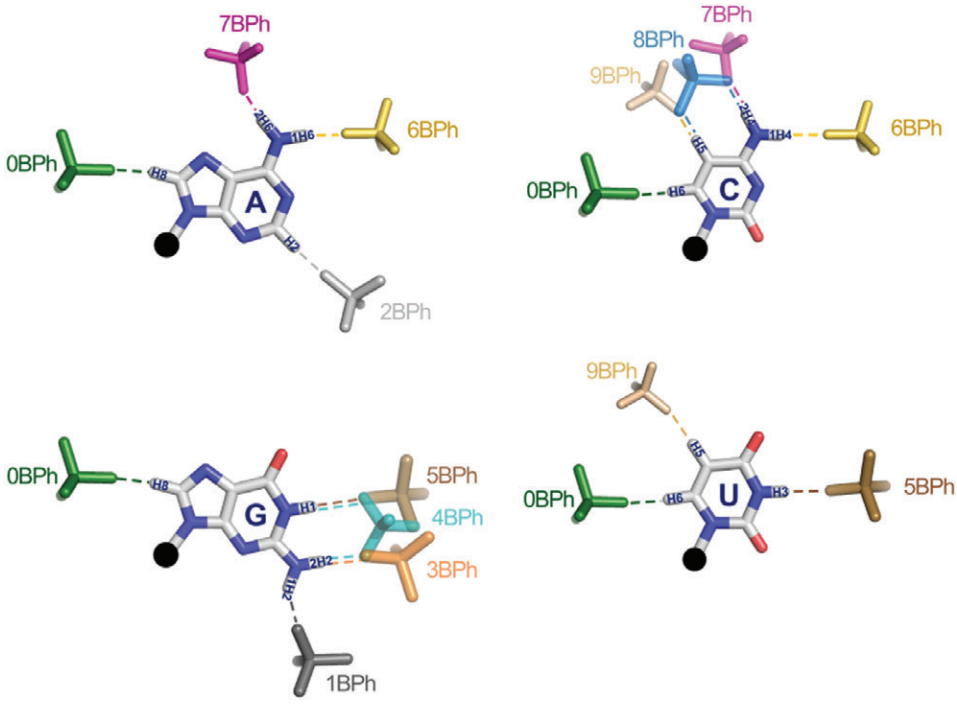


FIGURE 14 | Base–phosphate (BPh) interactions observed in RNA 3D structures for each base. H-bonds are indicated with dashed lines. BPh categories are numbered 0–9, starting at the H6 (pyrimidine) or H8 (purine) base positions. BPh interactions that involve equivalent functional groups on different bases are grouped together: 0 BPh (A, C, G, U), 5BPh (G, U), 6BPh (A, C), 7BPh (A, C) and 9BPh (C, U). Bridging interactions, 8BPh and 4BPh, are especially stable. Adapted from Zirbel et al. (9) with permission.

phosphate groups, resulting in electrostatic repulsion between phosphate groups but enhancement of H-bonding with donors on base edges of the RNA bases. The stabilizing H-bonding interactions between bases and the phosphate backbone moieties, referred to as BPh interactions, help to reduce intra-molecular repulsion between phosphate groups while specifically stabilizing compactly folded structural motifs and architectures. The classification of BPh interactions by base is shown in Figure 14. This classification was produced by quantum mechanical (QM) calculations and molecular dynamics simulations informed by structural bioinformatics analysis.[29,57] The QM calculations revealed the optimal geometries and the intrinsic stabilization energies of these interactions, and showed that the most stable BPh interactions, those on the W edge of G, rival GC WC base pairs in stability. Very stable BPh interactions are also formed by the W edge of U and the H edge of C (see Figure 14 and Ref 29).

Many recurrent 3D motifs contain specific BPh interactions that tend to be highly conserved. For instance, many HL, including the anticodon and T-loops of tRNAs, contain the 'U-turn' submotif, a sharp bend in the backbone stabilized by H-bonding between the W edge of U and the phosphate of the N+3 base of the loop.[40] A conserved BPh interaction involving the conserved 'bulged' G is observed in recurrent sarcin–ricin (S/R) internal loop motifs.[58] Conserved GU wobble base pairs are observed to bind anionic oxygen phosphate atoms in the minor groove to facilitate tight packing of helical elements.[34] It is necessary to consider conserved BPh interactions to fully account for the chemical protections observed in 16S rRNA.[59,60] These examples show that BPh interactions are widespread in RNA structures and play significant roles in RNA folding, as documented in Table 2.

Base–Ribose Interactions

The 2′-OH of ribose participates in S-edge base pairs, acting as both H-bond donor and acceptor and is considered part of the nt sugar edge. H-bonds can also form with the ribose O4′ atom, acting as H-bond donor. Another type of interaction involves stacking of bases on ribose rings as shown in Figure 15. The example in this figure is a key interaction that stabilizes the binding of tRNAs to the P-site of the 30S subunit. The base–ribose stacking occurs between the ribose of the first anticodon nt of tRNA (G34) in the anticodon HL (nts 32–38 of tRNA) and the highly conserved base G966, located in HL 34 of 16S rRNA (an intermolecular loop–loop interaction). Because only the tRNA ribose and not the base is involved in the interaction with 16S, the tRNA anticodon base sequence can vary, allowing different tRNAs to bind to the same site on 16S rRNA. However, G966, the 16S base involved in the interaction, is highly conserved and, moreover, chemically modified.[62] The sequence propensities of base–ribose stacking interactions deserve further study. Annotations of base–ribose interactions are also posted on NDB.

Metal Ion-Mediated Phosphate–Phosphate Interactions

Structured RNA molecules generally require multivalent metal ions to fold into their compact, functional structures.[50,51,63–66] This is because cations such as Mg^{2+} can bridge between two or more negatively charged phosphate groups, allowing compact structures to form in which phosphate groups are in close proximity. Cationic polyamines and basic proteins can also promote RNA folding.[67,68] Mg^{2+} is particularly well suited to facilitate rRNA compaction because it is abundant in cells and has a very high charge density among biologically available ions, owing to its relatively small ionic radius (0.6 Å) and +2 electrical charge. Mg^{2+} associates preferentially with the anionic nonbridging oxygen atoms of the phosphate groups.[64,69–72]

RNA 3D MOTIFS AND NON-WC PAIRS

Recurrent modular 3D motifs generally correspond to individual HL, IL, or MHJ loops. As we have seen, 'loops' play very important roles in structured RNA molecules. While some motifs appear to be unique, many occur over and over in unrelated RNA molecules. Most recurrent 3D motifs are relatively small (less than ~20 nts) and therefore can evolve independently in unrelated molecules. For example, kink-turn and sarcin–ricin motifs, C-loops, and GNRA or UNCG HL are found in many different molecules and even in multiple locations in the same large RNA molecule. In fact, 16S rRNA contains instances of each of these motifs.

Non-WC pairs play two fundamental roles in RNA 3D structure: (1) They are the building blocks of RNA 3D motifs,[73] where they provide the specific local interactions that structure individual HL, IL, and MHJ 3D motifs; and (2) Non-WC pairs mediate most LR (tertiary) interactions, usually in combination with LR WC pairs, base-stacking, and BPh interactions, as discussed above for 16S rRNA. As we have seen, most of these interactions occur between loop nts or loop and helix nts. The different types of non-WC pairs tend to contribute to different extents to local versus LR interactions, as shown in Table 5. For example,

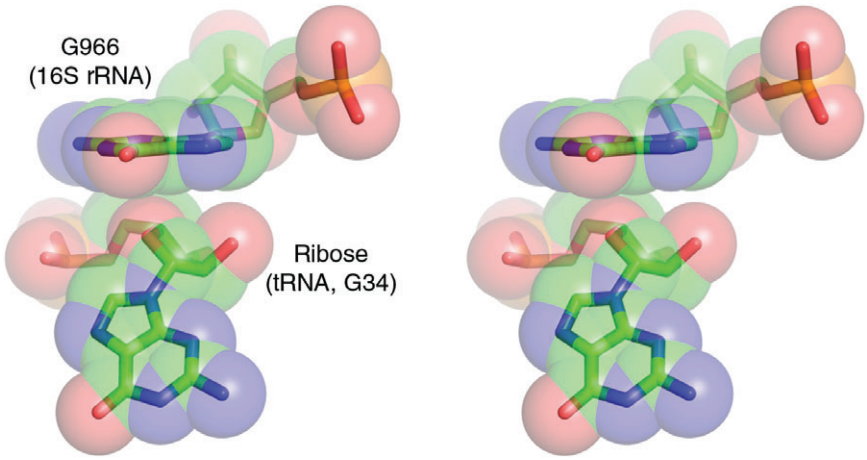


FIGURE 15 | Base–ribose stacking interaction. The conserved base–ribose stacking interaction involving ribose 34 in the anticodon of tRNA bound to the P-site of 16S rRNA and conserved base G966 in 16S. The figure is a stereo-view of the base–ribose interaction. From PDB file 4GD1.[61]

TABLE 5 | Local versus Long-Range Interactions in *Escherichia coli* (*E. coli*) 16S rRNA by Base-Pairing Family

	Local Interactions	Local Interactions	Long-Range Interactions	Long-Range Interactions	All Interactions	All Interactions
Base-Pair Family	Counts	Relative Frequency (%)	Counts	Relative Frequency (%)	Counts	Relative Frequency (%)
cWW	460	75.3	12	15.6	472	68.6
tWW	9	1.5	2	2.6	11	1.6
cWH	8	1.3	5	6.5	13	1.9
tWH	37	6.1	1	1.3	38	5.5
cWS	9	1.5	4	5.2	13	1.9
tWS	7	1.1	5	6.5	12	1.7
cHH	0	0.0	0	0.0	0	0.0
tHH	3	0.5	0	0.0	3	0.4
cHS	11	1.8	0	0.0	11	1.6
tHS	41	6.7	2	2.6	43	6.3
cSS	12	2.0	30	39.0	42	6.1
tSS	14	2.3	16	20.8	30	4.4
Totals	611	100%	77	100%	688	100%

cWW, tWH, and tHS together account for almost 90% of local base pairs, whereas cSS, tSS, cWS, tWS, and cWH account for almost 80% of LR pairs.

Use of the base-pairing symbols shown in Table 3 to annotate local interactions that structure HL, IL, and MHJ captures much of the crucial information found in the 3D structure. An example of a recurrent 3D motif from 16S rRNA is the internal loop of helix 20 which is a variant of bacterial loop E of 5S rRNA.[18] Renditions of the 3D and annotated 2D structures are shown in Figure 16(a) and (b). The *E. coli* and *T. thermophilus* versions of this recurrent motif are superposable in 3D space even though the sequences are different because corresponding bases form isosteric base pairs of the same geometric type, as shown in Figure 16(b) and (c). This and several related 3D motifs were predicted from sequence analysis based on non-WC base variations observed in sequence alignments, before the release of the atomic-resolution 3D structures of the 16S rRNAs.[17,75] This example illustrates the general principle that corresponding motifs in homologous

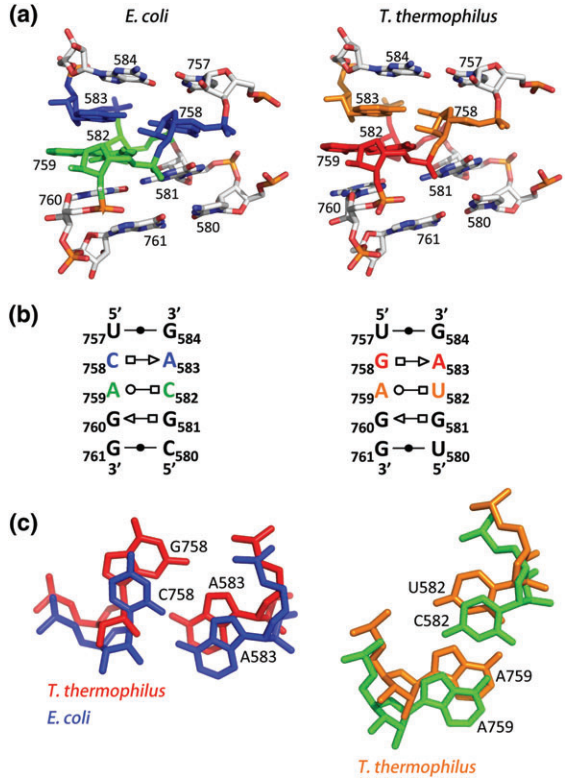


FIGURE 16 | (a) 3D structures of internal loop from helix 20 of 16S rRNA PDB files 2AW7[22] and 1FJG.[74] (b) 2D annotations showing conserved non-WC base pairs. Although they differ in sequence, the two motifs have the same interactions and are assigned to the same motif group. (c) Superposition of isosteric tWH and tSH non-Watson–Crick base pairs from the *Escherichia coli* and *Thermus thermophilus* versions of the helix 20 motif, slightly shifted for visual clarity.

RNA molecules often fold into very similar 3D structures and are stabilized by the same interactions, i.e., the same types of base pairs, stacking of BPh interactions.[21]

Grouping structurally similar RNA motif instances in motif families and aligning corresponding nts among them are crucial to improving bioinformatic methods for RNA 3D structure prediction based on sequence. In addition to variations in sequence, recurrent, structurally similar motifs can also differ in the number of nts that they comprise. A key question is which motifs to group together. Structural analysis based on conserved nt interactions suggests that when the extra nts found in the larger motifs are bulged out so as not to interact with the core nts of the motif and the corresponding core nts form identical interactions, then the motifs should be assigned to the same motif group. For example, some GNRA-type HL have five nts instead of four, but the additional nt is bulged out without affecting the conformations of the other nts.[76] Similarly, there is considerable variation in the number of nts in kink-turn and C-loop IL motifs, but this variation is largely confined to the looped-out nts on one strand.[19]

RESOURCES FOR EXPLORING RNA 3D STRUCTURES

RNA 3D Motif Atlas

These observations provide a framework for classifying recurrent 3D RNA motifs by geometric similarity. They suggest that motifs should be assigned to the same motif groups when corresponding nts form the same interactions, regardless of motif size. This approach was found to be more successful than relying exclusively on root-mean square deviations (RMSD) of atomic coordinates and was implemented to construct the RNA 3D Motif Atlas, a continuously updated online resource.[26] The Motif Atlas features automatic extraction of 3D motifs (HL and IL) from the current nonredundant (NR) dataset of RNA-containing NDB files and their clustering into structurally similar motif families.[26] The Motif Atlas is comprehensive and representative and can be accessed through NDB or directly at http://rna.bgsu.edu/rna3dhub/motifs.

Nucleic Acid Database (NDB)

The NDB (http://ndbserver.rutgers.edu) is a web portal providing access to information about 3D nucleic acid structures and their complexes. In addition to primary data that are archived in PDB, the NDB contains derived geometric data, classifications of structures and motifs, standards for describing nucleic acid features, and tools and software for analyzing DNA and RNA. A variety of search capabilities are available, as are many different types of reports. The NDB was recently redesigned and continues to evolve to meet the changing needs of RNA scientists. NDB provides sophisticated search capabilities at the level of whole structures by a large number of criteria. With the growth of interest in RNA biology and chemistry, the NDB is offering new RNA-derived data and annotations and integrating them into the search capabilities.[23] For example, NDB is developing finer-grained search capabilities at the level of individual motifs and interactions. NDB also provides curated descriptions and links to useful tools and software for RNA scientists (e.g., readers can visit http://ndbserver.rutgers.edu/ndbmodule/services/softwares.html).

Automated Annotation of nt Interactions in RNA 3D Structures

Atomic-resolution 3D structures reveal the architectures of RNA molecules and the stabilizing local

and LR nt interactions. However, RNA 3D structures are complex and difficult for students and non-specialists to comprehend and interpret. Identifying and classifying individual nt interactions is tedious and error-prone work when done manually. It is for this reason that computer programs have been written by several different research groups to automate and standardize this process.[27,28,77–79] The different programs produce annotations that largely agree with each other; discrepancies occur mainly when annotating low-resolution structures that are poorly modeled. Annotations of most of the recurrent nt interactions discussed here are now made available for each new atomic-resolution experimental RNA 3D structure when it appears in NDB.[23] Annotations are accessible under 'Structural Features' in the upper left of each structure summary page of NDB.

Integrating 2D and 3D RNA Structural Representations

RNA 2D diagrams illustrate at a glance the folding of the chain to form the nested WC-paired helices and identify the domain structure of the RNA and the sizes and positions of individual HL, IL, and MHJ loops. They can provide road maps for accessing the 3D structures because it is so easy to find individual nts in the 2D and to identify neighboring nts in the same helical elements and loops. What the 2D lacks, of course, are the local and LR non-WC pairing, stacking, and backbone interactions and the spatial relationships between helical elements. However, interactive computer technology can fill the gap between the over-simplification of 2D diagrams and the complexity of the 3D structures.

The most basic step is the use of color to coordinate the 2D and 3D representations of RNA molecules, as illustrated for 16S in Figure 1. Nts belonging to the same helical element are colored in 3D as shown in the 2D. Coordinated color coding of 2D and 3D structures by helical element facilitates identifying the elements to which interacting nts belong. In the 3D structure, helical elements distant in the 2D can be in close proximity and each helical element can interact with several different elements, not to mention proteins and ligands. To allow readers to visualize *E. coli* 16S rRNA in color on their own computers, we have provided as supplementary files to this article a SwissPDBViewer file (2AW7-16S-helices-colored.pdb) with these color codings and a Pymol script (2AW7-helices-colored-script.pml) to be used with PDB file 2AW7 to generate the display.

The 2D can also be used to rapidly access specific regions of the 3D structure. In this approach, clickable 'hotspots' are embedded at the locations of structural features on the online 2D diagram, e.g., all HL and IL of the molecule. Clicking on the hotspots brings up an interactive window in which the 3D structure of the selected motif is displayed. Links can be provided to the motif family to which the motif belongs to access sequence variants of the motif. One can access 2D diagrams of representative 16S and 23S rRNA molecules that feature this capability at this link: http://rna.bgsu.edu/rna3dhub/motifs/2ds.

While it is possible to represent all local and LR base pairs on a static 2D representation, the resulting 'circuit diagram' is difficult to interpret and comprehend.[80] An alternative approach is to equip interactive 2D diagrams with sets of controls to display selectively the different classes of interaction by type or location. A prototype of such a display for 16S rRNA can be accessed at http://rna.bgsu.edu/rna3dhub/pdb/2AW7/2d.

From 1D to 2D to 3D RNA Structures

It will never be possible to solve the 3D structures of all the RNAs we find in nature at atomic resolution. Fortunately, it may not be necessary to do so as bioinformatic tools have been developed and are under constant improvement to predict 2D and 3D RNA structures starting from individual sequences or alignments of homologous RNA molecules.[81]

The first step is to predict the 2D structure to identify the helical regions and define the nts that belong to HL, IL, MJ, or linker segments. Dynamic programming algorithms that make use of the growing database of experimentally determined, nearest neighbor thermodynamic parameters have been refined to make this step quite reliable, especially when chemical probing or complementary phylogenetic data are available, i.e., sufficiently diverged homologous sequences.[82] The most commonly used resources are listed in the table provided in Supporting Information.

The next step in structure prediction is to identify recurrent HL, IL, or MHJ motifs based on loop sequences identified in the 2D or to carry out *de novo* modeling based on energy parameters. Many groups are developing *de novo* modeling tools.[83–87] 'RNA Puzzles' has been established to provide CASP-like blind tests of these tools, leading to rapid progress in RNA 3D modeling.[88] Accessible through NDB is JAR3D, a new online tool that is closely linked to the 3D Motif Atlas. This tool matches user-provided sequences for HL or IL to the most probable matches of motifs found in the Motif Atlas using observed sequence variations and considerations of isosteric base pairs to score sequences (see http://rna.bgsu.edu/main/webapps/jar3d/).

The next step once the 2D structure is determined and possible 3D structures of HL and IL are proposed is to predict the conformations of MHJs.[89,90] This step is crucial as the MHJ determine the relative orientations of helical elements in 3D space. It is computationally challenging. Improvements in the prediction of MHJ in 2D structures also continue to be made.[91]

The final step is to predict tertiary interactions between loops and helical elements that stabilize the 3D architecture. This step is also very challenging but is greatly facilitated by sequence analysis when sufficiently diverged and properly aligned homologs are available for analysis.[14,92] Structural analysis of the recurrent 'interaction motifs' formed by structurally similar HL, IL, or MHL when they interact with each other or with helices can also be expected to contribute significantly to improvements in modeling tertiary interactions.

CONCLUSION

We have illustrated, using 16S rRNA, the prevalence of loop nts in structured RNAs and their central roles in mediating local and LR interactions that stabilize the 3D architecture and make possible the specific binding of proteins, ligands, and other RNA molecules. We have provided a tutorial to familiarize readers with the most sequence specific of those interactions, the WC and non-WC base pairs, and showed that by grouping them in geometrically similar families and isosteric groups, it is possible to interpret and even predict the sequence variations of recurrent RNA 3D motifs. The distinction between base combination and base-pair geometry was emphasized, because the same combinations of bases (e.g., UA, AG, and CC) can make different types of base pairs. Base substitutions that preserve the base-pairing geometries maintain functional structures and are more likely to occur. Annotating nt interactions is a prerequisite for grouping 3D motifs into structurally similar groups. This is how the RNA Motif Atlas is organized. As new RNA structures are solved experimentally, new motifs are uncovered and the number of sequence variants for known motifs increases. The RNA Motif Atlas makes this information available on an ongoing, continually updated basis to allow researchers to improve their algorithms to predict RNA 3D structure.

Finally, we have provided links to resources to allow readers to deepen their understanding of RNA 3D structure and access specific information about RNAs of interest to them. New ways of integrating 2D and 3D representations of structured RNAs will make it easier for students and scientists to explore and comprehend the structures, functions, and evolution of these amazing, ancient molecules. New bioinformatic tools will make it possible to evermore reliably predict the 2D and 3D structures and possible functions and interactions of new RNA molecules.

ACKNOWLEDGMENTS

We would like to thank Todd Childers for designing the colors used in Figure 1, Craig L. Zirbel for stimulating discussions. This work was supported by grant GM085328 from the National Institutes of Health to N.B.L and C.L.Z.

REFERENCES

1. Sponer J, Sponer JE, Mladek A, Jurecka P, Banas P, Otyepka M. Nature and magnitude of aromatic base stacking in DNA and RNA: quantum chemistry, molecular mechanics, and experiment. *Biopolymers* 2013, 99:978–988.
2. Leontis NB, Westhof E. Geometric nomenclature and classification of RNA base pairs. *RNA* 2001, 7: 499–512.
3. Stombaugh J, Zirbel CL, Westhof E, Leontis NB. Frequency and isostericity of RNA base pairs. *Nucleic Acids Res* 2009, 37:2294–2312.
4. Dutheil JY, Jossinet F, Westhof E. Base pairing constraints drive structural epistasis in ribosomal RNA sequences. *Mol Biol Evol* 2010, 27:1868–1876.
5. Leontis NB, Stombaugh J, Westhof E. The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res* 2002, 30:3497–3531.
6. Holley RW, Apgar J, Everett GA, Madison JT, Marquisee M, Merrill SH, Penswick JR, Zamir A. Structure of a ribonucleic acid. *Science* 1965, 147: 1462–1465.
7. Madison JT, Everett GA, Kung H. Nucleotide sequence of a yeast tyrosine transfer RNA. *Science* 1966, 153:531–534.
8. Woese CR, Gutell R, Gupta R, Noller HF. Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiol Rev* 1983, 47:621–669.
9. Noller HF, Woese CR. Secondary structure of 16S ribosomal RNA. *Science* 1981, 212:403–411.

10. Kim SH, Quigley GJ, Suddath FL, McPherson A, Sneden D, Kim JJ, Weinzierl J, Rich A. Three-dimensional structure of yeast phenylalanine transfer RNA: folding of the polynucleotide chain. *Science* 1973, 179: 285–288.
11. Woese CR, Magrum LJ, Gupta R, Siegel RB, Stahl DA, Kop J, Crawford N, Brosius J, Gutell R, Hogan JJ, et al. Secondary structure model for bacterial 16S ribosomal RNA: phylogenetic, enzymatic and chemical evidence. *Nucleic Acids Res* 1980, 8:2275–2293.
12. Noller HF, Kop J, Wheaton V, Brosius J, Gutell RR, Kopylov AM, Dohme F, Herr W, Stahl DA, Gupta R, et al. Secondary structure model for 23S ribosomal RNA. *Nucleic Acids Res* 1981, 9:6167–6189.
13. Gutell RR, Lee JC, Cannone JJ. The accuracy of ribosomal RNA comparative structure models. *Curr Opin Struct Biol* 2002, 12:301–310.
14. Michel F, Costa M, Massire C, Westhof E. Modeling RNA tertiary structure from patterns of sequence variation. *Methods Enzymol* 2000, 317:491–510.
15. Michel F, Westhof E. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J Mol Biol* 1990, 216:585–610.
16. Gutell RR, Larsen N, Woese CR. Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiol Rev* 1994, 58:10–26.
17. Leontis NB, Westhof E. A common motif organizes the structure of multi-helix loops in 16 S and 23 S ribosomal RNAs. *J Mol Biol* 1998, 283:571–583.
18. Leontis NB, Westhof E. The 5S rRNA loop E: chemical probing and phylogenetic data versus crystal structure. *RNA* 1998, 4:1134–1153.
19. Lescoute A, Leontis NB, Massire C, Westhof E. Recurrent structural RNA motifs, isostericity matrices and sequence alignments. *Nucleic Acids Res* 2005, 33:2395–2409.
20. Petrov AS, Bernier CR, Hershkovits E, Xue Y, Waterbury CC, Hsiao C, Stepanov VG, Gaucher EA, Grover MA, Harvey SC, et al. Secondary structure and domain architecture of the 23S and 5S rRNAs. *Nucleic Acids Res* 2013, 41:7522–7535.
21. Petrov AI, Sweeney BA, Leontis NB. Analyzing, searching, and annotating recurrent RNA three-dimensional motifs. In: Klostermeier D, Hammann C, eds. *RNA Structure and Folding*. Berlin; Boston, MA: De Gruyter; 2013, 363–398.
22. Schuwirth BS, Borovinskaya MA, Hau CW, Zhang W, Vila-Sanjurjo A, Holton JM, Cate JH. Structures of the bacterial ribosome at 3.5 A resolution. *Science* 2005, 310:827–834.
23. Coimbatore Narayanan B, Westbrook J, Ghosh S, Petrov AI, Sweeney B, Zirbel CL, Leontis NB, Berman HM. The nucleic acid database: new features and capabilities. *Nucleic Acids Res* 2014, 42:D114–D122.
24. Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, Cate JH, Noller HF. Crystal structure of the ribosome at 5.5 A resolution. *Science* 2001, 292:883–896.
25. Woese CR, Winker S, Gutell RR. Architecture of ribosomal RNA: constraints on the sequence of "tetra-loops". *Proc Natl Acad Sci U S A* 1990, 87:8467–8471.
26. Petrov AI, Zirbel CL, Leontis NB. Automated classification of RNA 3D motifs and the RNA 3D motif Atlas. *RNA* 2013, 19:1327–1340.
27. Sarver M, Zirbel CL, Stombaugh J, Mokdad A, Leontis NB. FR3D: finding local and composite recurrent structural motifs in RNA 3D structures. *J Math Biol* 2008, 56:215–252.
28. Petrov AI, Zirbel CL, Leontis NB. WebFR3D—a server for finding, aligning and analyzing recurrent RNA 3D motifs. *Nucleic Acids Res* 2011, 39:W50–W55.
29. Zirbel CL, Sponer JE, Sponer J, Stombaugh J, Leontis NB. Classification and energetics of the base-phosphate interactions in RNA. *Nucleic Acids Res* 2009, 37:4898–4918.
30. Sponer J, Sponer JE, Petrov AI, Leontis NB. Quantum chemical studies of nucleic acids: can we construct a bridge to the RNA structural biology and bioinformatics communities? *J Phys Chem B* 2010, 114: 15723–15741.
31. Sponer J, Mokdad A, Sponer JE, Spackova N, Leszczynski J, Leontis NB. Unique tertiary and neighbor interactions determine conservation patterns of Cis Watson-Crick A/G base-pairs. *J Mol Biol* 2003, 330: 967–978.
32. Gutell RR, Woese CR. Higher order structural elements in ribosomal RNAs: pseudo-knots and the use of noncanonical pairs. *Proc Natl Acad Sci U S A* 1990, 87:663–667.
33. Klein BK, Forman P, Shiomi Y, Schlessinger D. Electron microscopy of secondary structure in partially denatured Escherichia coli 16S rRNA and 30S subunits. *Biochemistry* 1984, 23:3927–3933.
34. Mokdad A, Krasovska MV, Sponer J, Leontis NB. Structural and evolutionary classification of G/U wobble basepairs in the ribosome. *Nucleic Acids Res* 2006, 34:1326–1341.
35. Gagnon MG, Mukhopadhyay A, Steinberg SV. Close packing of helices 3 and 12 of 16 S rRNA is required for the normal ribosome function. *J Biol Chem* 2006, 281:39349–39357.
36. Gagnon MG, Steinberg SV. GU receptors of double helices mediate tRNA movement in the ribosome. *RNA* 2002, 8:873–877.
37. Szewczak AA, Moore PB, Chang YL, Wool IG. The conformation of the sarcin/ricin loop from 28S ribosomal RNA. *Proc Natl Acad Sci U S A* 1993, 90: 9581–9585.
38. Lescoute A, Westhof E. Topology of three-way junctions in folded RNAs. *RNA* 2006, 12:83–93.

39. Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Szewczak AA, Kundrot CE, Cech TR, Doudna JA. RNA tertiary structure mediation by adenosine platforms. *Science* 1996, 273:1696–1699.
40. Quigley GJ, Rich A. Structural domains of transfer RNA molecules. *Science* 1976, 194:796–806.
41. Nagaswamy U, Fox GE. Frequent occurrence of the T-loop RNA folding motif in ribosomal RNAs. *RNA* 2002, 8:1112–1119.
42. Varani G, Cheong C, Tinoco I Jr. Structure of an unusually stable RNA hairpin. *Biochemistry* 1991, 30:3280–3289.
43. Hoehndorf R, Batchelor C, Bittner T, Dumontier M, Eilbeck K, Knight R, Mungall CJ, Richardson JS, Stombaugh J, Westhof E, et al. The RNA ontology (RNAO): an ontology for integrating RNA sequence and structure data. *Appl Ontol* 2011, 6:53–89.
44. Abu Almakarem AS, Petrov AI, Stombaugh J, Zirbel CL, Leontis NB. Comprehensive survey and geometric classification of base triples in RNA structures. *Nucleic Acids Res* 2012, 40:1407–1423.
45. Saenger W. *Principles of Nucleic Acid Structure*. New York: Springer Verlag; 1984.
46. Draper DE. RNA folding: thermodynamic and molecular descriptions of the roles of ions. *Biophys J* 2008, 95:5489–5495.
47. Murray LJ, Arendall WB 3rd, Richardson DC, Richardson JS. RNA backbone is rotameric. *Proc Natl Acad Sci U S A* 2003, 100:13904–13909.
48. Murray LJ, Richardson JS, Arendall WB, Richardson DC. RNA backbone rotamers–finding your way in seven dimensions. *Biochem Soc Trans* 2005, 33(Pt 3): 485–487.
49. Richardson JS, Schneider B, Murray LW, Kapral GJ, Immormino RM, Headd JJ, Richardson DC, Ham D, Hershkovits E, Williams LD, et al. RNA backbone: consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *RNA* 2008, 14:465–481.
50. Woodson SA. Metal ions and RNA folding: a highly charged topic with a dynamic future. *Curr Opin Chem Biol* 2005, 9:104–109.
51. Bowman JC, Lenz TK, Hud NV, Williams LD. Cations in charge: magnesium ions in RNA folding and catalysis. *Curr Opin Struct Biol* 2012, 22:262–272.
52. Desiraju GR, Gavezzotti A. Crystal structures of polynuclear aromatic hydrocarbons. Classification, rationalization and prediction from molecular structure. *Acta Crystallogr Sect B Struct Sci* 1989, 45:473–482.
53. Siegfried NA, O'Hare B, Bevilacqua PC. Driving forces for nucleic acid pK(a) shifting in an A(+).C wobble: effects of helix position, temperature, and ionic strength. *Biochemistry* 2010, 49:3225–3236.
54. Veeraraghavan N, Ganguly A, Golden BL, Bevilacqua PC, Hammes-Schiffer S. Mechanistic strategies in the HDV ribozyme: chelated and diffuse metal ion interactions and active site protonation. *J Phys Chem B* 2011, 115:8346–8357.
55. Romby P, Moras D, Bergdoll M, Dumas P, Vlassov VV, Westhof E, Ebel JP, Giege R. Yeast tRNAAsp tertiary structure in solution and areas of interaction of the tRNA with aspartyl-tRNA synthetase. A comparative study of the yeast phenylalanine system by phosphate alkylation experiments with ethylnitrosourea. *J Mol Biol* 1985, 184:455–471.
56. Leontis NB, Westhof E. Conserved geometrical base-pairing patterns in RNA. *Q Rev Biophys* 1998, 31:399–455.
57. Zgarbova M, Jurecka P, Banas P, Otyepka M, Sponer JE, Leontis NB, Zirbel CL, Sponer J. Noncanonical hydrogen bonding in nucleic acids. Benchmark evaluation of key base-phosphate interactions in folded RNA molecules using quantum-chemical calculations and molecular dynamics simulations. *J Phys Chem A* 2011, 115:11277–11292.
58. Correll CC, Munishkin A, Chan YL, Ren Z, Wool IG, Steitz TA. Crystal structure of the ribosomal RNA domain essential for binding elongation factors. *Proc Natl Acad Sci U S A* 1998, 95:13436–13441.
59. Merryman C, Moazed D, McWhirter J, Noller HF. Nucleotides in 16S rRNA protected by the association of 30S and 50S ribosomal subunits. *J Mol Biol* 1999, 285:97–105.
60. Stern S, Moazed D, Noller HF. Structural analysis of RNA using chemical and enzymatic probing monitored by primer extension. *Methods Enzymol* 1988, 164:481–489.
61. Dunkle JA, Wang L, Feldman MB, Pulk A, Chen VB, Kapral GJ, Noeske J, Richardson JS, Blanchard SC, Cate JHD. Structures of the bacterial ribosome in classical and hybrid states of tRNA binding. *Science* 2011, 332:981–984.
62. Burakovsky DE, Prokhorova IV, Sergiev PV, Milon P, Sergeeva OV, Bogdanov AA, Rodnina MV, Dontsova OA. Impact of methylations of m2G966/m5C967 in 16S rRNA on bacterial fitness and translation initiation. *Nucleic Acids Res* 2012, 40:7885–7895.
63. Draper DE, Grilley D, Soto AM. Ions and RNA folding. *Annu Rev Biophys Biomol Struct* 2005, 34:221–243.
64. Draper DE. Folding of RNA tertiary structure: linkages between backbone phosphates, ions, and water. *Biopolymers* 2013, 99:1105–1113.
65. Auffinger P, Grover N, Westhof E. Metal ion binding to RNA. *Met Ions Life Sci* 2011, 9:1–35.
66. Tan ZJ, Chen SJ. Importance of diffuse metal ion binding to RNA. *Met Ions Life Sci* 2011, 9:101–124.
67. Klein DJ, Moore PB, Steitz TA. The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *J Mol Biol* 2004, 340:141–177.
68. Koculi E, Thirumalai D, Woodson SA. Counterion charge density determines the position and plasticity

of RNA folding transition states. *J Mol Biol* 2006, 359:446–454.

69. Draper DE. A guide to ions and RNA structure. *RNA* 2004, 10:335–343.
70. Draper DE. RNA folding: thermodynamic and molecular descriptions of the roles of ions. *Biophys J* 2008, 95:5489–5495.
71. Misra VK, Shiman R, Draper DE. A thermodynamic framework for the magnesium-dependent folding of RNA. *Biopolymers* 2003, 69:118–136.
72. Tan ZJ, Chen SJ. Predicting electrostatic forces in RNA folding. *Methods Enzymol* 2009, 469:465–487.
73. Leontis NB, Lescoute A, Westhof E. The building blocks and motifs of RNA architecture. *Curr Opin Struct Biol* 2006, 16:279–287.
74. Carter AP, Clemons WM, Brodersen DE, Morgan-Warren RJ, Wimberly BT, Ramakrishnan V. Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature* 2000, 407:340–348.
75. Leontis NB, Stombaugh J, Westhof E. Motif prediction in ribosomal RNAs Lessons and prospects for automated motif prediction in homologous RNA molecules. *Biochimie* 2002, 84:961–973.
76. Stombaugh J, Nasalean L, Zirbel C, Leontis NB. RNA 3D structural motifs: definition, identification, annotation, and database searching. In: Walter NG, Woodsen SA, Batey RT, eds. *Non-Protein Coding RNAs*. Heidelberg: Springer; 2008.
77. Yang H, Jossinet F, Leontis N, Chen L, Westbrook J, Berman H, Westhof E. Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res* 2003, 31:3450–3460.
78. Gendron P, Lemieux S, Major F. Quantitative analysis of nucleic acid three-dimensional structures. *J Mol Biol* 2001, 308:919–936.
79. Parisien M, Major F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* 2008, 452:51–55.
80. Lescoute A, Westhof E. The interaction networks of structured RNAs. *Nucleic Acids Res* 2006, 34: 6587–6604.
81. Leontis N, Westhof E. Modeling RNA molecules. In: Leontis N, Westhof E, eds. *RNA 3D Structure Analysis and Prediction*. Heidelberg: Springer; 2012, 5–18.
82. Aigner K, Dressen F, Steger G. Methods for predicting RNA secondary structure. In: Leontis N, Westhof E, eds. *RNA 3D Structure Analysis and Prediction*. Heidelberg: Springer; 2012, 5–18.
83. Rother K, Rother M, Boniecki M, Puton T, Tomala K, Lukasz P, Bujnicki JM. Template-based and template-free modeling of RNA 3D structure. In: Leontis N, Westhof E, eds. *RNA 3D Structure Analysis and Prediction*. Heidelberg: Springer; 2012, 5–18.
84. Sijenyi F, Saro P, Ouyang Z, Damm-Ganamet K, Wood M, Jiang J, Santalucia J. The RNA folding problems: different levels of sRNA structure prediction. In: Leontis N, Westhof E, eds. *RNA 3D Structure Analysis and Prediction*. Heidelberg: Springer; 2012, 5–18.
85. Flores SC, Jonikas MA, Bruns CM, Ku JP, Schmidt J, Altman R. Methods for building and refining 3D models of RNA. In: Leontis N, Westhof E, eds. *RNA 3D Structure Analysis and Prediction*. Heidelberg: Springer; 2012, 5–18.
86. Ding F, Dokholyan NV. Multiscale modeling of RNA structure and dynamics. In: Leontis N, Westhof E, eds. *RNA 3D Structure Analysis and Prediction*. Heidelberg: Springer; 2012, 5–18.
87. Cao S, Chen SJ. Statistical mechanical modeling of RNA folding: from free energy landscape to tertiary structural prediction. In: Leontis N, Westhof E, eds. *RNA 3D Structure Analysis and Prediction*. Heidelberg: Springer; 2012, 5–18.
88. Cruz JA, Blanchet MF, Boniecki M, Bujnicki JM, Chen SJ, Cao S, Das R, Ding F, Dokholyan NV, Flores SC, et al. RNA-puzzles: A CASP-like evaluation of RNA three-dimensional structure prediction. *RNA* 2012, 18:610–625.
89. Lamiable A, Barth D, Denise A, Quessette F, Vial S, Westhof E. Automated prediction of three-way junction topological families in RNA secondary structures. *Comput Biol Chem* 2012, 37:1–5.
90. Laing C, Wen D, Wang JT, Schlick T. Predicting coaxial helical stacking in RNA junctions. *Nucleic Acids Res* 2012, 40:487–498.
91. Liu B, Diamond JM, Mathews DH, Turner DH. Fluorescence competition and optical melting measurements of RNA three-way multibranch loops provide a revised model for thermodynamic parameters. *Biochemistry* 2011, 50:640–653.
92. Westhof E, Masquida B, Jossinet F. Predicting and modeling RNA architecture. *Cold Spring Harb Perspect Biol* 2011, 3. doi:10.1101/cshperspect.a003632.

