4868–4876 *Nucleic Acids Research, 2010, Vol. 38, No. 14* *Published online 11 March 2010*
doi:10.1093/nar/gkq155

The RNA backbone plays a crucial role in mediating the intrinsic stability of the GpU dinucleotide platform and the GpUpA/GpA miniduplex

Xiang-Jun Lu[1,*], Wilma K. Olson[2,3] and Harmen J. Bussemaker[1,4,*]

[1]Department of Biological Sciences, Columbia University, New York, NY 10027, [2]Department of Chemistry and Chemical Biology, [3]BioMaPS Institute for Quantitative Biology, Rutgers – The State University of New Jersey, Piscataway, NJ 08854 and [4]Center for Computational Biology and Bioinformatics, Columbia University, New York, NY 10027, USA

Received December 4, 2009; Revised February 11, 2010; Accepted February 17, 2010

ABSTRACT

The side-by-side interactions of nucleobases contribute to the organization of RNA, forming the planar building blocks of helices and mediating chain folding. Dinucleotide platforms, formed by side-by-side pairing of adjacent bases, frequently anchor helices against loops. Surprisingly, GpU steps account for over half of the dinucleotide platforms observed in RNA-containing structures. Why GpU should stand out from other dinucleotides in this respect is not clear from the single well-characterized H-bond found between the guanine N2 and the uracil O4 groups. Here, we describe how an RNA-specific H-bond between O2′(G) and O2P(U) adds to the stability of the GpU platform. Moreover, we show how this pair of oxygen atoms forms an out-of-plane backbone 'edge' that is specifically recognized by a non-adjacent guanine in over 90% of the cases, leading to the formation of an asymmetric miniduplex consisting of 'complementary' GpUpA and GpA subunits. Together, these five nucleotides constitute the conserved core of the well-known loop-E motif. The backbone-mediated intrinsic stabilities of the GpU dinucleotide platform and the GpUpA/GpA miniduplex plausibly underlie observed evolutionary constraints on base identity. We propose that they may also provide a reason for the extreme conservation of GpU observed at most 5′-splice sites.

INTRODUCTION

Recent years have witnessed a dramatic increase in our appreciation of the crucial role played by RNA in a variety of structural, regulatory and enzymatic processes in the cell (1). Knowing how the base sequence of an RNA molecule determines its 3D structure is crucial for understanding its biological function (2). Hydrogen bonding (H-bonding) and stacking interactions between bases are major driving forces for RNA secondary and tertiary structure formation, and the large number of distinct structural motifs to which such interactions can give rise has been the subject of intense research (3–8). One of the features of folded RNA molecules is the frequent occurrence of higher order structural motifs involving three or more bases, knowledge of which can be valuable for the computational prediction of tertiary structure from sequence (9,10). H-bond interactions involving the ribose sugar of the RNA backbone, the 2′-hydroxyl (O2′) group in particular, can play an important role in defining RNA tertiary structure (5,11).

Many RNA structural motifs involve non-Watson–Crick base pairing (12). A special case of such non-canonical pairing occurs in the so-called dinucleotide platform, defined as two neighboring nucleotides arranged in a side-by-side planar arrangement with an H-bond between the respective bases. The best known examples of such platforms include: the ApA or adenosine platform, first identified by Doudna and coworkers (13) in the P4–P6 domain of a Group I intron; the GpU platform, later found in the crystal structures of the complex of a small fragment of *Escherichia coli* 23S ribosomal RNA (rRNA) and ribosomal protein L11 (14); the sarcin/ricin domain of *E. coli* 23S rRNA (15);

*To whom correspondence should be addressed. Tel: +1 212 854 1527; Fax: +1 212 865 8246; Email: xl2134@columbia.edu
Correspondence may also be addressed to Harmen J. Bussemaker. Tel: +1 212 854 9932; Fax: +1 212 865 8246; Email: hjb2004@columbia.edu

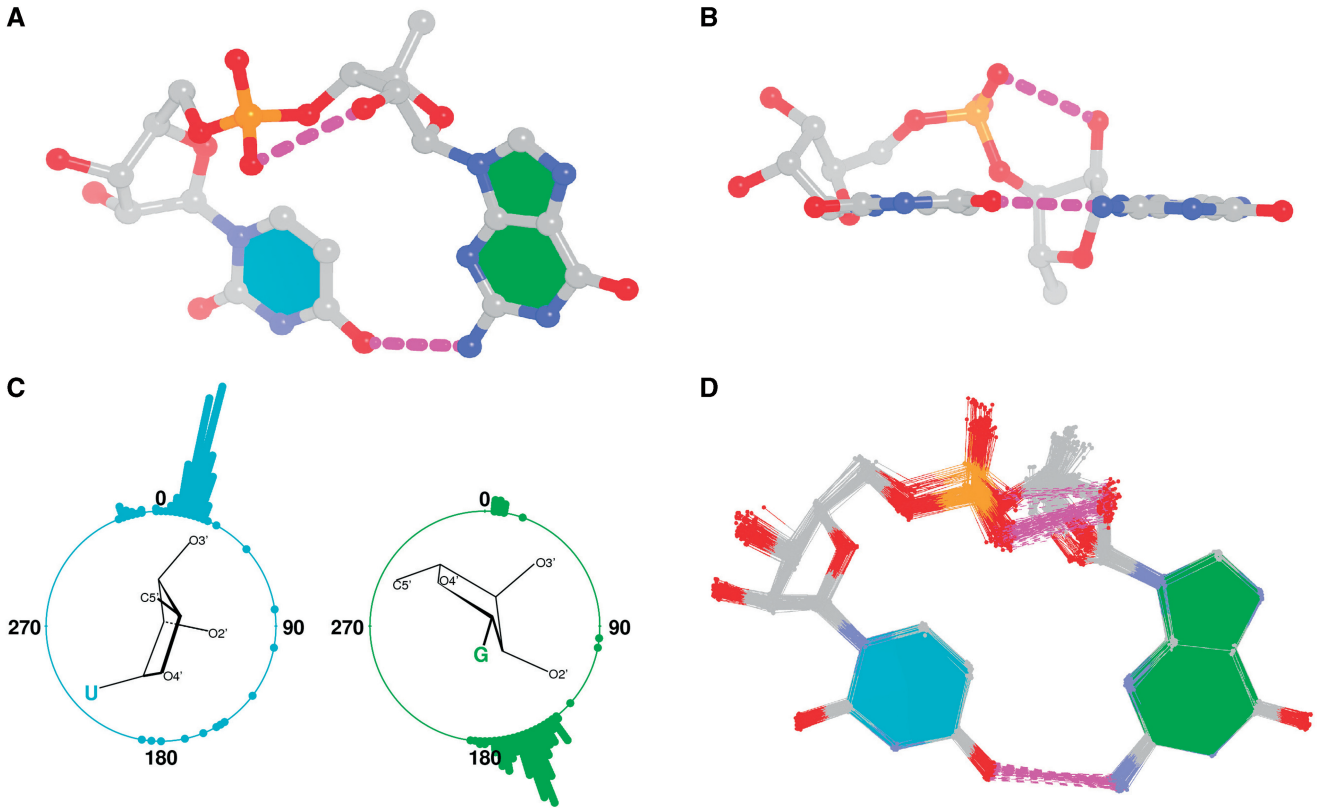

Figure 1. Structural characterization of the G + U platform. (**A**) Representative molecular image depicting the well-characterized N2(G)–O4(U) H-bond between consecutive bases and the heretofore little-noticed O2′(G) – O2P(U) H-bond in the sugar–phosphate backbone. (**B**) Side view highlighting the out-of-plane backbone 'edge' of the platform, with the H-bonded O2′(G) and O2P(U) atoms directed away from the base-pair plane. (**C**) Distribution of the phase angle of pseudorotation (55) *P* of the U and G sugar rings across the 193 G + U platforms detailed in Supplementary Table S1. Atomic models illustrate the dominant C3′-endo ($0° < P < 36°$) and C2′-endo ($144° < P < 180°$) conformations adopted, respectively, by the U and G sugars. (**D**) Superposed images emphasizing the stiffness of the 140 G + U platforms with mixed C2′-endo/C3′-endo (G/U) puckering. Composite images are obtained by superposition of the mean coordinate frame of the two bases in each platform. The G and U are represented, respectively, throughout in green and cyan. Dashed lines in magenta denote H-bonds. The structures shown in A and B depict the $G_{2655}pU_{2656}$ platform from a 27-nt fragment that mimics the sarcin/ricin loop from *E. coli* 23S rRNA [PDB entry 1MSY (6)].

and the purine riboswitch from the *xpt-pbuX* operon of *Bacillus subtilis* (16) (see molecular images in Supplementary Figure S1). The GpU platform is particularly prevalent in complex RNA structures (17), but the reasons for its wide occurrence are not known.

In this report, we demonstrate the crucial importance of an intra-backbone H-bond within the GpU platform, between the O2′ of guanosine and one of the non-bridging oxygens (O2P) of the phosphate that connects the two nucleotides (Figure 1). We show that the backbone H-bond-stabilized GpU platform almost always participates in a highly distinctive structural motif, consisting of a GpUpA trinucleotide interacting with a non-adjacent GpA dinucleotide through an intricate network of H-bonding and base-stacking interactions. The asymmetric GpUpA/GpA miniduplex coincides with the conserved core of the well-known loop-E motif (3), also known as the bulged-G motif (15). The miniduplex occurs as well in the crystal structure of the self-spliced Group IIC intron from *Oceanobacillus iheyensis* (18). In what follows, we show that the backbone 'edge' of the GpU platform and the interactions that it enables provide a structural rationale for the prevalence and evolutionary conservation of this motif.

MATERIALS AND METHODS

Structural data

We downloaded and analyzed all of the structures available in the Nucleic Acid Database (NDB) (19) as of October 2008. Unless otherwise mentioned, we limit our discussion to RNA X-ray crystal structures solved at 2.5 Å or a better resolution.

Identification of base pairs

We used the 3DNA software package (20,21) to characterize the spatial arrangements of interacting bases. We chose the following set of stringent parameters to ensure that the geometry of each identified base pair is nearly planar and supports at least one inter-base H-bond: (i) a vertical distance (stagger) between base planes ≤1.5 Å; (ii) an angle between base normal vectors ≤30°; and (iii) a pair of nitrogen and/or oxygen base atoms at a distance ≤3.3 Å. This purely geometric approach allows for the identification of canonical Watson–Crick as well as non-canonical base pairs, made up of normal or modified bases, regardless of tautomeric or protonation state.

Identification of dinucleotide platforms and higher order interactions

For a base pair between nucleotides i and $i+1$ to be classified as a dinucleotide platform, we required formation of a covalent bond between the O3′(i) and P($i+1$) atoms. To identify higher order associations, we searched in 3D space for nucleotides that have stacking and H-bonding interactions with GpU dinucleotide platforms.

Evolutionary conservation

For phylogenetic analysis of archaeal 23S rRNA, we downloaded the highly refined seed alignment from the comparative RNA web site (http://www.rna.ccbb.utexas.edu/) (22) maintained by the Gutell Laboratory. The secondary structure of *Haloarcula marismortui* 23S rRNA was adapted from the 2D folding pattern generated from the same website.

Supplementary materials

The PDF file 'Lu_supp_info.pdf' contains three tables and five figures. Tables S1 and S3 provide full structural information (PDB id, NDB id, chain id, residue name and number and associated parameters) to verify the results reported in this work.

RESULTS

Predominance of GpU among the dinucleotide platforms

Using the 3DNA (20,21) software package (see 'Materials and Methods' section for details), we analyzed the spatial arrangements of adjacent nucleotides in all nucleic acid structures stored, as of October 2008, in the NDB (19). Among X-ray crystal structures solved at 2.5 Å or better resolution, we identified a total of 312 dinucleotide platforms (Supplementary Table S1). All but 10 of the dimers occur in RNA, with adjacent A, C, G or U bases lying in the same plane and adopting a so-called M + N pairing scheme (20), i.e. with the faces of the two bases, like those of an A + U Hoogsteen pair, pointing in the same direction (see Supplementary Figure S2 and text below). These 302 platforms occur in 48 of the 373 RNA-containing crystal structures that pass the 2.5-Å resolution cutoff.

The frequency of each dinucleotide in the full set of structures, regardless of whether or not it forms a platform, ranges from 5 to 9% for the six dinucleotides adopting the most platform arrangements (Table 1). The G + U platform stands out from the other platforms in two respects: it accounts for most of the M + N platforms (193/302 or 64%) and shows the greatest propensity to adopt a platform conformation in the RNA structures (193/3605 or 5.4%). The frequency of occurrence (45/302 or 15%) of the next most prevalent platform, A + A, is several times less than that of the G + U platform. Moreover, the platform conformation occurs for only 1.3% (45/3586) of all ApA dinucleotides. None of the 14 other possible dinucleotide platforms is significantly over-represented.

The over-representation of the G + U platform persists if we use a more lenient 3.2-Å resolution cutoff or a more stringent 2.0-Å resolution cutoff, at which none of the ribosomal structures is included (Supplementary Table S2A). Structures of *H. marismortui* 23S rRNA are highly over-represented in the NDB (Supplementary Table S1). We, therefore, repeated our analysis by deleting the recurring *H. marismortui* 23S rRNA entries in our dataset of 2.5-Å or better resolution structures, and using two other datasets from the literature: 342 RNA structures of 'reduced redundancy' subject to a 4.0-Å resolution cutoff (23) and 54 non-redundant RNA structures selected with a 3.0-Å cutoff (24). Our results did not change in any substantive way (Supplementary Table S2B). In fact, analysis of a single 23S rRNA structure (9) leads to the same findings: 11 of the 19 platforms detected in the fully refined large subunit (50S) of the *H. marismortui* ribosome (25) [Protein Data Bank (PDB) (26) entry 1JJ2] are G + U platforms ($P = 4.8 \times 10^{-9}$, cumulative binomial distribution). This establishes statistical significance beyond any reasonable doubt. Moreover, even if the over-representation of the GpU platform should turn out to be limited to these specific currently solved RNA structures, a structural explanation is still wanting.

Table 1. RNA dinucleotide platforms found in 373 crystal structures of 2.5 Å or better resolution

	Platforms		Dinucleotides		Platforms per dinucleotide (%)
	Count	Percentage	Count	Percentage	
G + U	193	63.9	3605	6.6	5.4
A + A	45	14.9	3586	6.5	1.3
U + C	19	6.3	2840	5.2	0.7
A + C	14	4.6	3576	6.5	0.4
C + A	14	4.6	3293	6.0	0.4
G + G	14	4.6	4830	8.8	0.3
Others	3	1.0	33 141	60.4	0.0
Total	302	100	54 871	100	–

Platforms are defined as two consecutive nucleotides with coplanar bases and stabilized by a base–base H-bond. In virtually all platforms identified, the faces of the bases point in the same direction [the so-called M + N pairing scheme (20)]. The G + U platform stands out from the other platforms in two respects: it accounts for most (64%) of the platforms and shows the greatest propensity (5.4%) of all dinucleotides to adopt a platform arrangement.

A crucial role for an intra-backbone H-bond within the GpU platform

The extreme over-representation of the G + U association distinguishes it from all other dinucleotide platforms, suggesting an intrinsic structural propensity. The G + U platform is stabilized by a well-characterized N2(G) – O4(U) H-bond in a favorable donor–acceptor arrangement (Figure 1A) (27). However, based on a single base–base H-bond, it is hard to rationalize the predominance of the interaction. Exhaustive identification of all possible H-bonds, made possible with 3DNA (20,21), uncovered a crucial second contribution to the stability of the G + U platform: an H-bond between the O2′ of guanosine on a given residue i, O2′(i), and the O2P of the backbone phosphate group, O2P($i+1$), that connects the two nucleotides (Figure 1A and B). As this H-bond has

received only scant attention in the RNA literature (28), especially in the context of a dinucleotide platform (21), we queried the Cambridge Structure Database (29) for hydroxyl-phosphate H-bonds with similar relative geometry and chemical identity. We found that an H-bond of this type in the phospholipid lysophosphatidyl-ethanolamine (30) plays a critical role in the organization of that molecule. Moreover, the O2′(i) – O2P(i+1) H-bond is not specific to dinucleotide platforms: there are 1186 such pairwise interactions within a distance cutoff of 3.3 Å in the current set of RNA crystal structures.

The two H-bonds within the G + U platform are likely to act cooperatively, thus providing a structural rationale for the high prevalence of the paired arrangement (Figure 1A and B). The O2′(G)–O2P(U) H-bond occurs in 82% (158/193) of all G + U platforms, underscoring its structural importance. The distance between the O2′(G) and O2P(U), 2.68 ± 0.14 Å, is close to optimal (31,32), and the roughly tetrahedral angles formed by the C2′–O2′ bond of G and the O2P on U, 114 ± 3°, and the P–O2P bond on U and the O2′ on G, 79 ± 5°, allow for the formation of reasonable H-bonds. In contrast, the base–base H-bond in the less prevalent A + A platform is suboptimal (longer and less linear) compared to that of the G + U platform (Supplementary Figure S1), and the O2′(i) – O2P(i+1) H-bond occurs in only 31% (14/45) of the coplanar ApA examples.

The conformation of the ribose sugar ring of a nucleotide affects the way in which its 2′-hydroxyl interacts with other groups. We therefore analyzed the puckering of the sugar rings in the G + U platform (Figure 1C) and found that whereas uridine preferentially adopts the C3′-endo form (the conformation of the sugar characteristic of A-form helical RNA), guanosine occurs almost exclusively in the C2′-endo form (the conformation typical of B-form DNA). This supports our hypothesis that the O2′(i) – O2P(i+1) H-bond plays a crucial role. The G + U platform is an exceptionally rigid structural unit (Figure 1D): the root-mean-square deviation (RMSD) of the atoms in all 140 G + U platforms that contain both the O2′(G) – O2P(U) H-bond and the mixed (C2′-endo/C3′-endo) puckering of the guanosine and uridine sugars is only 0.17 ± 0.07 Å (distribution relative to the centroid, i.e. the structure with the smallest average RMSD from all other structures). This makes the G + U platform even more rigid than the Watson–Crick G–C and A–U base pairs in RNA duplexes (33).

The backbone 'edge' of the GpU platform mediates a GpUpA/GpA miniduplex

Strikingly, the backbone-stabilized G + U platform virtually always participates in an asymmetric miniduplex, consisting of a GpUpA trinucleotide (of which it is part) and a 'complementary', non-adjacent GpA dinucleotide. The GpUpA and GpA subunits are held together by an intricate network of H-bonding and base-stacking interactions (Figure 2). The miniduplex consists of two layers: three nucleotides in the lower plane containing the G + U platform plus the A of GpA, and two nucleotides in the upper plane containing the A of GpUpA plus the G of GpA (Figure 2A and B). The O2′(G) and O2P(U) atoms lie 2.22 ± 0.23 and 3.35 ± 0.22 Å, respectively, above the G + U platform plane (Figure 1B). In 96% (152/158) of the cases, this sugar–phosphate feature interacts with a non-adjacent guanine in the upper plane through 2–3 H-bonds (Figure 2C). While these guanine-backbone H-bonds have been previously noted in the context of the sarcin/ricin loop (15), the role of the intra-backbone O2′(G) – O2P(U) H-bond described above has heretofore been largely ignored. Significantly, it naturally subdivides the nine-membered (N1–C6–O6–O2′–C2′–C3′–O3′–P–O2P) ring formed by the guanine-backbone contacts into fused five- and six-membered rings (N1–C6–O6–O2′–O2P and O2′–C2′–C3′–O3′–P–O2P) that plausibly contribute to the specificity, rigidity and stability of the miniduplex interaction. The offset of the O2′ and O2P atoms from the GpU plane further contributes to the formation of well-directed H-bonds with a stacked, but sequentially distant guanine.

An exhaustive search of other base pairs formed by the upper G from GpA reveals a sheared G–A interaction (34–36) with the A of the GpUpA (Figure 2D). The same G forms an intra-strand O2′(G)–O4′(A) H-bond (of near-optimal length, 2.83 ± 0.20 Å) with the A in the lower plane, to which it is covalently attached. The same O2′(G) atom also contributes to the sheared G – A pair, forming an H-bond, 2.96 ± 0.11 Å in length, with the N6 of A. A corresponding search in the lower plane reveals that the A of the GpA forms a reverse Hoogsteen pair (36) with the U of the G + U platform in all cases; the phosphate of the A interacts specifically with the platform G through two additional H-bonds. Thus, the G + U platform is part of a 'complementary' G + U/A base triplet held together by 5–6 H-bonds (Figure 2E). The intra- and inter-strand interactions apparently work in concert to organize the miniduplex as a whole.

Overall, the two-layered, backbone-stabilized GpUpA/GpA miniduplex is held together by ~12 H-bonds, as well as cross-strand purine-stacking interactions between the two adenines and the two guanines in the lower and upper planes (Figure 2A and B). The five-nucleotide structural unit is exceptionally rigid; the RMSD of the 152 GpUpA/GpA examples is only 0.35 ± 0.13 Å. Detailed inspection of the intricate network of interactions shows that the base identities of the five nucleotides are highly specific. For example, mutating the guanine in the upper plane to a pyrimidine would increase the distance of potential proton donor and acceptor atoms from the O2′(G) and O2P(U) of the G + U platform, disallowing the H-bonds observed in Figure 2C; changing the G to an adenine would change the donor/acceptor pattern at the Watson–Crick edge, allowing a single N6(A)–O2′(G) H-bond or possibly an additional N6(A)–O2P(U) H-bond, while eliminating one of the H-bonds to the upper-plane adenine of GpUpA. Furthermore, a systematic search reveals that among the 1186 RNA dinucleotides with an O2′(i)–O2P(i+1) backbone 'edge' (regardless of platform conformation or base identity, see above), there are 237 cases where both O2′(i) and O2P(i+1) are H-bonded to base atoms of another

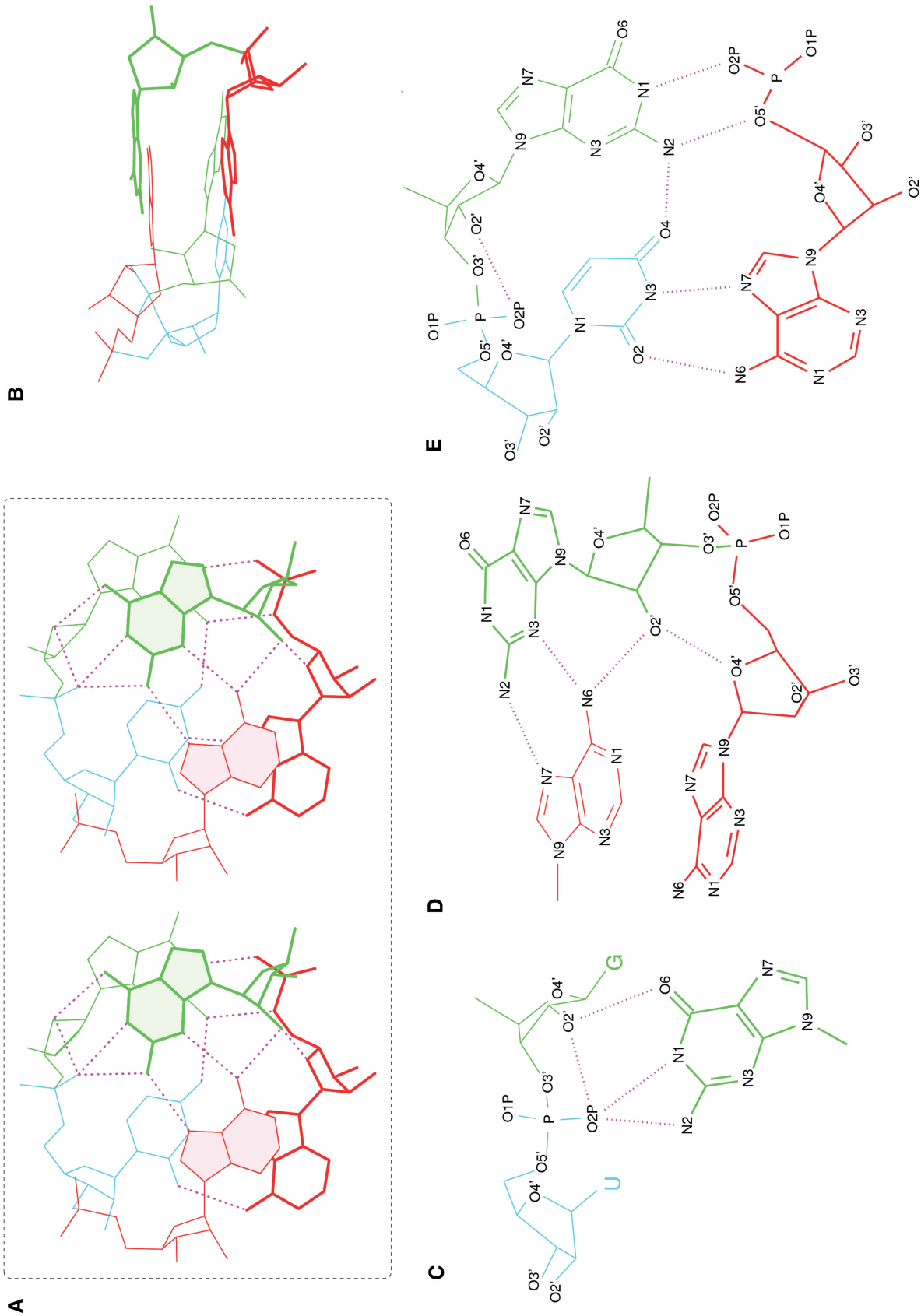

Figure 2. Characterization of the two-layered GpUpA/GpA miniduplex adopted by 152 of the 158 backbone H-bond-stabilized G+U platforms. (**A**) Top view (stereo) showing the intricate network of long-range interactions and the anti-parallel 5′→3′ directions of the GpUpA trinucleotide and the non-adjacent GpA dinucleotide backbones (thin and thick lines, respectively). (**B**) Side view highlighting the two-layered arrangement and the opposing chain directions. (**C**) The multiple H-bonding interactions between the upper-plane G from the GpA dinucleotide and the O2′(G) and O2P(U) atoms of the G+U platform within the GpUpA trinucleotide. (**D**) Additional interactions of the upper-plane G with (i) the upper-plane A from GpUpA, via a sheared G–A pair and an additional O2′(G)—N6(A) sugar–base H-bond and (ii) the lower-plane A to which it is covalently attached via a O2′(G)—O4′(A) inter-sugar H-bond. (**E**) The reverse A–U Hoogsteen pair in the lower plane formed by the A from the GpA dinucleotide with the U of the G+U platform and the two sequence-specific H-bonding interactions between the 5′-phosphate of the same A and the Watson–Crick edge of the platform G. The structural fragment shown here is taken from PDB entry 1MSY (6).

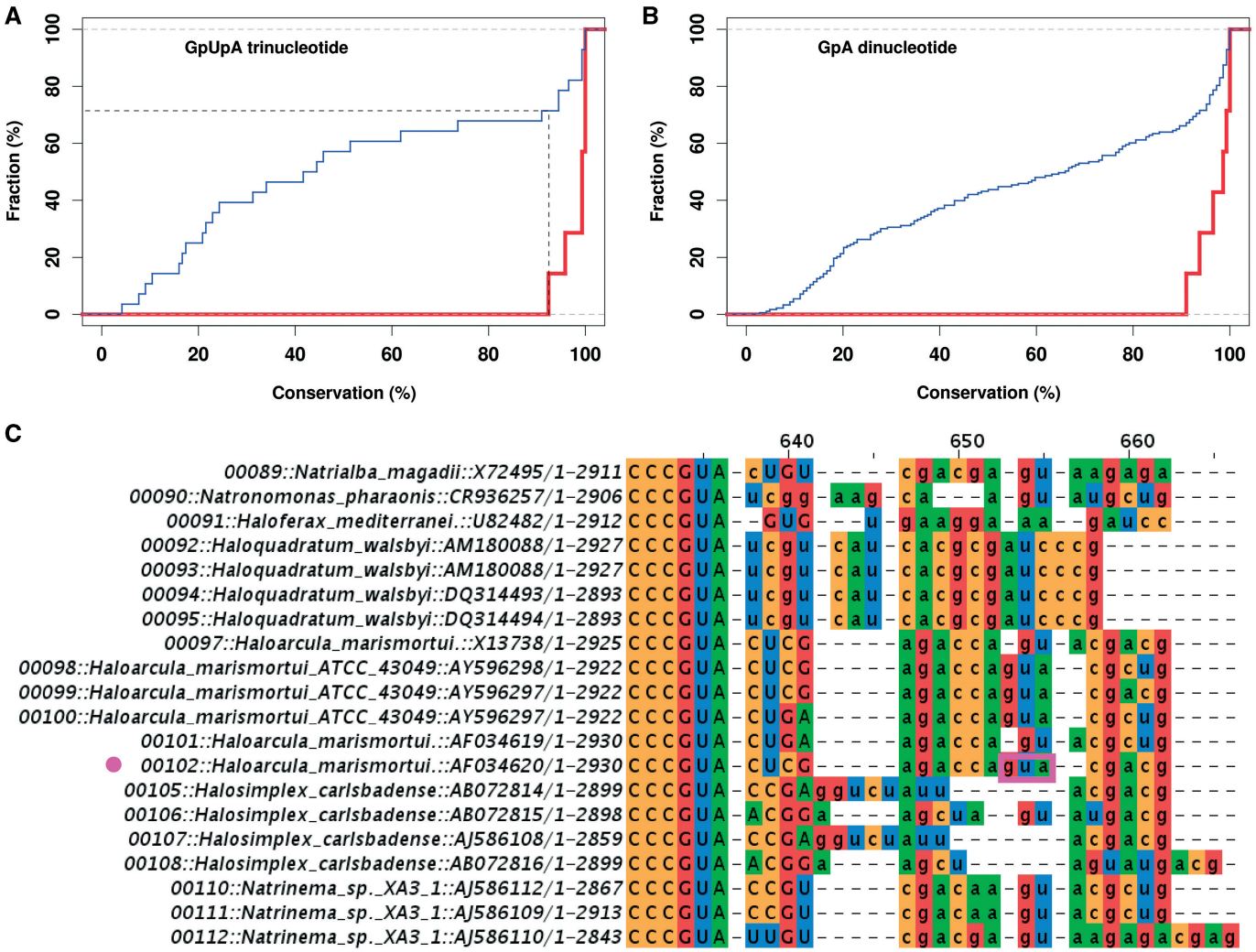

Figure 3. Strong evolutionary conservation of nucleotides involved in the GpUpA/GpA miniduplexes identified in *H. marismortui* 23S rRNA [PDB entry 1JJ2 (25)]. Analysis based on the manually curated multiple alignment of 144 archaeal sequences by Gutell and coworkers (22). (**A**) Effect of structural context on sequence conservation of the GpUpA trinucleotides. Shown is a comparison in high-confidence regions (uppercase base letters) between the cumulative distribution of percent conservation of those seven trinucleotides that participate in a GpUpA/GpA motif (red line) and those that do not (blue line). More than 70% of all non-structured GpUpA trinucleotides are less conserved than the least conserved structured GpUpA (dashed line). (**B**) Idem, for the GpA dinucleotide. (**C**) Suboptimal alignment of 23S rRNA sequences around the only one of the structured GpUpAs (magenta box) in *H. marismortui* (magenta dot), which occurs in a low-confidence region (lower-case base letters). The corresponding GpA dinucleotide of the single unconserved GpUpA/GpA motif (#3 in Supplementary Figure S3) also occurs in a region of low-confidence alignment. Consideration of the structural context of this trimer may improve the alignment in this region.

nucleotide. Strikingly, guanine accounts for 91.6% (217/237) of the interacting nucleotides with at least two H-bonds (details will be reported elsewhere). Such recognition of nucleotide sequence through the sugar–phosphate backbone is unprecedented.

Evolutionary conservation of the miniduplex

The backbone-stabilized GpUpA/GpA motif occurs eight times in the structure of the 23S rRNA of the *H. marismortui* large ribosomal subunit [PDB entry 1JJ2 (25)]. Examination of the interactions in the context of the 23S rRNA secondary structure (22) reveals that all GpUpA/GpA motifs occur in loop regions, either extending a double-helical stem or bringing sequentially distant nucleotides into contact at a multi-armed helical junction (Supplementary Figure S3). We analyzed a manually curated multiple alignment of 144 archaeal 23S rRNA sequences downloaded from the Gutell laboratory website (22) and found that GpUpA and GpA are almost entirely conserved when they occur in regions marked by the Gutell group as high-confidence. Figure 3A and B shows that the conservation of GpUpA and GpA at sites outside the structural context of the GpUpA/GpA miniduplex in *H. marismortui* 23S rRNA is significantly lower than that at the structured sites ($P = 0.002$ for GpUpA and 0.003 for GpA; Wilcoxon–Mann–Whitney test). The GpUpA and GpA comprising the single unconserved structural motif (#3 in Supplementary Figure S3) occur in regions of low-confidence alignment. Figure 3C illustrates the suboptimal alignment of the set

of archaeal sequences around the GpUpA in this region of the *H. marismortui* 23S rRNA, suggesting that the alignment might be improved by taking into account the new information reported here.

Miniduplex recognition

The rigid structure of the GpUpA/GpA miniduplex presents a variety of features for association with other moieties, such as other nucleotides, the backbones or side chains of proteins and metal ions. For example, A_{2010} in the *H. marismortui* 23S rRNA structure [PDB entry 1JJ2 (25)], which lies in the lower (G + U/A) plane of a GpUpA/GpA motif (site #7 in Supplementary Figure S3), interacts with the minor-groove edge of a G – C base pair (21) via an A-minor motif of Type I (37). Together with the G + U dinucleotide platform, these bases form a nearly planar pentaplet (Supplementary Figure S4). Furthermore, two neighboring backbone NH-groups of the zinc-finger protein TFIIIA recognize the O6 and N7 atoms on the major-groove edge of the guanine in the G + U platform found in the crystal complex with a fragment of *Xenopus laevis* 5S rRNA (38) (Supplementary Figure S5A). Finally, a magnesium ion interacts with the non-bridging oxygen (O2P) of the phosphate group immediately preceding the guanosine of one of the G + U platforms in the *H. marismortui* 23S rRNA structure (Supplementary Figure S5B; motif #1 in Figure S3).

DISCUSSION

The backbone 'edge' of the GpU platform

Our unbiased, data-driven structural analysis of the GpU dinucleotide platform reveals two crucial roles for the RNA sugar–phosphate backbone. First, an H-bond formed in most cases between the O2′ of the guanosine ribose sugar and the O2P of the intervening phosphate group provides stability and rigidity to the platform beyond the single N2(G)–O4(U) H-bond between the two bases. Accordingly, physical model building demonstrates that the O2′(G)–O2P(U) H-bond restricts the GpU dinucleotide platform to a virtually inflexible structure. We note, however, that the energetic contribution of this H-bond is likely to depend on both sequence context and environment, and a quantitative assessment of its net value would require carefully designed experiments (39) or high-level quantum chemical calculations (40). Second, the same two backbone atoms constitute a novel out-of-plane 'edge'—distinct from the well-documented in-plane edges of bases (5)—that can be recognized by other moieties (e.g. a nucleotide) through additional H-bonds (to a guanine in over 90% of the cases). Strikingly, the GpU platform, when present in the O2′–O2P backbone-stabilized form, nearly always appears in the context of an extremely rigid miniduplex consisting of 'complementary' GpUpA and GpA subunits. These five nucleotides form the conserved core of the loop-E (3) or bulged-G (15) motif found in a wide variety of functionally important RNA molecules, such as the sarcin/ricin loop (15) and other locations (41,42) in 23S rRNA, loop E region of 5S rRNA (43,44), helix 27 in 16S rRNA (45) and the lysine riboswitch (46). We also observed the GpUpA/GpA miniduplex within domain I of the group IIC intron (18), where it anchors two crucial structural features: the long-range α–α′ kissing loop interaction and the coaxial stacking of stems IA and IB.

The interactions that keep the GpUpA/GpA miniduplex in place are highly cooperative. However, other energetically favorable interactions (e.g. Watson–Crick pairing) may compete for the five constituent bases. The formation of the GpUpA/GpA motif may therefore depend on the structural and environmental context in which the bases occur. Indeed, while the GpU dinucleotide platform conformation and the miniduplex occur in loop E of the *H. marismortui* 5S rRNA structure (44), they are absent in the crystal structure of a loop-E fragment from *E. coli* (47). We note that the short 'extended' fragment at loop E of 5S rRNA has been documented as a loop-E motif (43), even though it lacks the bulged-G conformation of the GpU platform and the GpUpA/GpA miniduplex (48). The capability to distinguish such differences in structure underscores the strength of our geometry-based approach.

A role for the GpU dinucleotide platform and the GpUpA/GpA miniduplex during pre-mRNA splicing?

While it is well known that a GpU dinucleotide demarcates the 5′ end of virtually every intron processed by the major spliceosome (49), there is currently no structural rationale for this extreme evolutionary conservation. If the 5′-splice site (5′-SS) GpU were to adopt a platform conformation, its associated intrinsic rigidity and salient features could serve as a target for recognition by other spliceosomal components. Indeed, the geometry of the G + U platform (Figure 1A) is consistent with the experimental observation that *in vitro* recognition of the 5′-SS GpU by p220 (the human equivalent of the yeast protein Prp8) in the U5 small nuclear ribonucleoprotein (snRNP) is perturbed by substitution of a large methyl or iodo group, but not a small fluoro group, at position C5 of the uracil (50).

It is also tempting to speculate that beyond the backbone-stabilized GpU platform conformation itself, the larger network of interactions that holds together the GpUpA/GpA miniduplex might transiently form during the second step of the messenger RNA (mRNA) splicing reaction in yeast. At the presumed catalytic center of the spliceosome, Watson–Crick base pairing between the underlined flanking bases of the 5′-SS consensus GUAUGU and the conserved hexamer ACAGAG that starts at residue 47 of the yeast U6 snRNA (51) juxtaposes the GpUpA trinucleotide in the 5′-SS with the GpA dinucleotide of U6 across a loop region. Whether the interaction between these residues requires protein is an open question, and the detailed structural mechanism is still unknown (52,53). We propose that the 5′-SS GpUpA may interact with the U6 GpA using the GpUpA/GpA miniduplex conformation. The specific lower-plane pairing of the G + U platform and the non-adjacent

adenosine shown in Figure 2E is consistent with the observation that any mutation of A_{51} (the adenine in the putative GpA fragment) leads to accumulation of a lariat intermediate, but does not block the first step of the splicing reaction (54). Therefore, the GpUpA/GpA miniduplex would have to form between the first and second steps of the splicing reaction. Possible ways of experimentally testing our hypothesis include splicing assays using RNA molecules with targeted chemical modifications, designed to disrupt the O2′(G)–O2P(U) H-bond or other interactions within the miniduplex.

Summary

The detailed structural analysis presented in this paper points to an important role for the RNA backbone in mediating sequence-specific interactions, and provides a rationale for the over-representation and evolutionary conservation of the GpUpA/GpA miniduplex at the core of the loop-E/bulged-G motif. Our structural insights might help to interpret other extant data and guide the design of experiments aimed at elucidating the mechanism of mRNA splicing. The successful outcome of our computational structural analysis provides motivation for further unbiased searches for RNA structural motifs. Finally, algorithms for the prediction of RNA secondary and tertiary structure from sequence might benefit from taking the GpU dinucleotide platform and GpUpA/GpA miniduplex into account.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors are grateful to Daniel Aalberts, Larry Chasin, Dan Herschlag, Magda Konarska, Jim Manley and Ruben Gonzalez for valuable discussions and/or a critical reading of the manuscript. They thank Yurong Xin for valuable discussions in the early stages of this project. They also thank the anonymous reviewers, whose comments helped clarify the presentation of the manuscript.

FUNDING

National Institutes of Health grants (R01HG003008 and U54CA121852 to H.J.B. and R01GM20861 and R01GM034809 to W.K.O.). Funding for open access charge: National Institutes of Health grant R01HG003008.

REFERENCES

1. Gesteland,R.F., Cech,T.R. and Atkins,J.F. (eds) (2006), *The RNA World*, 3rd edn. Cold Spring Harbor Laboratory Press, New York.
2. Noller,H.F. (2005) RNA structure: reading the ribosome. *Science*, **309**, 1508–1514.
3. Moore,P.B. (1999) Structural motifs in RNA. *Annu. Rev. Biochem.*, **68**, 287–300.
4. Hermann,T. and Patel,D.J. (1999) Stitching together RNA tertiary architectures. *J. Mol. Biol.*, **294**, 829–849.
5. Leontis,N.B. and Westhof,E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA*, **7**, 499–512.
6. Correll,C.C. and Swinger,K. (2003) Common and distinctive features of GNRA tetraloops based on a GUAA tetraloop structure at 1.4 Å resolution. *RNA*, **9**, 355–363.
7. Jossinet,F., Ludwig,T.E. and Westhof,E. (2007) RNA structure: bioinformatic analysis. *Curr. Opin. Microbiol.*, **10**, 279–285.
8. Holbrook,S.R. (2008) Structural principles from large RNAs. *Annu. Rev. Biophys.*, **37**, 445–464.
9. Das,R. and Baker,D. (2007) Automated de novo prediction of native-like RNA tertiary structures. *Proc. Natl Acad. Sci. USA*, **104**, 14664–14669.
10. Parisien,M. and Major,F. (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature*, **452**, 51–55.
11. Tamura,M. and Holbrook,S.R. (2002) Sequence and structural conservation in RNA ribose zippers. *J. Mol. Biol.*, **320**, 455–474.
12. Westhof,E. and Fritsch,V. (2000) RNA folding: beyond Watson-Crick pairs. *Structure*, **8**, R55–R65.
13. Cate,J.H., Gooding,A.R., Podell,E., Zhou,K., Golden,B.L., Szewczak,A.A., Kundrot,C.E., Cech,T.R. and Doudna,J.A. (1996) RNA tertiary structure mediation by adenosine platforms. *Science*, **273**, 1696–1699.
14. Wimberly,B.T., Guymon,R., McCutcheon,J.P., White,S.W. and Ramakrishnan,V. (1999) A detailed view of a ribosomal active site: the structure of the L11-RNA complex. *Cell*, **97**, 491–502.
15. Correll,C.C., Beneken,J., Plantinga,M.J., Lubbers,M. and Chan,Y.L. (2003) The common and the distinctive features of the bulged-G motif based on a 1.04 Å resolution RNA structure. *Nucleic Acids Res.*, **31**, 6806–6818.
16. Batey,R.T., Gilbert,S.D. and Montange,R.K. (2004) Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature*, **432**, 411–415.
17. Harrison,A.M., South,D.R., Willett,P. and Artymiuk,P.J. (2003) Representation, searching and discovery of patterns of bases in complex RNA structures. *J. Comput. Aided Mol. Des.*, **17**, 537–549.
18. Toor,N., Keating,K.S., Taylor,S.D. and Pyle,A.M. (2008) Crystal structure of a self-spliced group II intron. *Science*, **320**, 77–82.
19. Berman,H.M., Olson,W.K., Beveridge,D.L., Westbrook,J., Gelbin,A., Demeny,T., Hsieh,S.H., Srinivasan,A.R. and Schneider,B. (1992) The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.*, **63**, 751–759.
20. Lu,X.J. and Olson,W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.
21. Lu,X.J. and Olson,W.K. (2008) 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc.*, **3**, 1213–1227.
22. Cannone,J.J., Subramanian,S., Schnare,M.N., Collett,J.R., D'Souza,L.M., Du,Y., Feng,B., Lin,N., Madabusi,L.V., Muller,K.M. *et al.* (2002) The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics*, **3**, 2.
23. Stombaugh,J., Zirbel,C.L., Westhof,E. and Leontis,N.B. (2009) Frequency and isostericity of RNA base pairs. *Nucleic Acids Res.*, **37**, 2294–2312.
24. Xin,Y., Laing,C., Leontis,N.B. and Schlick,T. (2008) Annotation of tertiary interactions in RNA structures reveals variations and correlations. *RNA*, **14**, 2465–2477.
25. Klein,D.J., Schmeing,T.M., Moore,P.B. and Steitz,T.A. (2001) The kink-turn: a new RNA secondary structure motif. *EMBO J.*, **20**, 4214–4221.
26. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
27. Coulocheri,S.A., Pigis,D.G., Papavassiliou,K.A. and Papavassiliou,A.G. (2007) Hydrogen bonds in protein-DNA

complexes: where geometry meets plasticity. *Biochimie*, **89**, 1291–1303.
28. Richardson,J.S., Schneider,B., Murray,L.W., Kapral,G.J., Immormino,R.M., Headd,J.J., Richardson,D.C., Ham,D., Hershkovits,E., Williams,L.D. *et al.* (2008) RNA backbone: consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *RNA*, **14**, 465–481.
29. Allen,F.H. (2002) The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr. B*, **58**, 380–388.
30. Pascher,I., Sundell,S. and Hauser,H. (1981) Polar group interaction and molecular packing of membrane lipids. The crystal structure of lysophosphatidylethanolamine. *J. Mol. Biol.*, **153**, 807–824.
31. Taylor,R. and Kennard,O. (1984) Hydrogen-bond geometry in organic crystals. *Acc. Chem. Res.*, **17**, 320–326.
32. Jeffrey,G.A., Maluszynska,H. and Mitra,J. (1985) Hydrogen bonding in nucleosides and nucleotides. *Int. J. Biol. Macromol.*, **7**, 336–348.
33. Olson,W.K., Esguerra,M., Xin,Y. and Lu,X.J. (2009) New information content in RNA base pairing deduced from quantitative analysis of high-resolution structures. *Methods*, **47**, 177–186.
34. Heus,H.A. and Pardi,A. (1991) Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science*, **253**, 191–194.
35. Li,Y., Zon,G. and Wilson,W.D. (1991) NMR and molecular modeling evidence for a G·A mismatch base pair in a purine-rich DNA duplex. *Proc. Natl Acad. Sci. USA*, **88**, 26–30.
36. Saenger,W. (1984) *Principles of Nucleic Acid Structure*. Springer, New York.
37. Nissen,P., Ippolito,J.A., Ban,N., Moore,P.B. and Steitz,T.A. (2001) RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc. Natl Acad. Sci. USA*, **98**, 4899–4903.
38. Lu,D., Searles,M.A. and Klug,A. (2003) Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition. *Nature*, **426**, 96–100.
39. SantaLucia,J. Jr, Kierzek,R. and Turner,D.H. (1992) Context dependence of hydrogen bond free energy revealed by substitutions in an RNA hairpin. *Science*, **256**, 217–219.
40. Zirbel,C.L., Sponer,J.E., Sponer,J., Stombaugh,J. and Leontis,N.B. (2009) Classification and energetics of the base-phosphate interactions in RNA. *Nucleic Acids Res.*, **37**, 4898–4918.
41. Yang,X., Gerczei,T., Glover,L.T. and Correll,C.C. (2001) Crystal structures of restrictocin-inhibitor complexes with implications for RNA recognition and base flipping. *Nat. Struct. Biol.*, **8**, 968–973.
42. Leontis,N.B., Stombaugh,J. and Westhof,E. (2002) Motif prediction in ribosomal RNAs: lessons and prospects for automated motif prediction in homologous RNA molecules. *Biochimie*, **84**, 961–973.
43. Leontis,N.B. and Westhof,E. (1998) The 5S rRNA loop E: chemical probing and phylogenetic data versus crystal structure. *RNA*, **4**, 1134–1153.
44. Ban,N., Nissen,P., Hansen,J., Moore,P.B. and Steitz,T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science*, **289**, 905–920.
45. Carter,A.P., Clemons,W.M., Brodersen,D.E., Morgan-Warren,R.J., Wimberly,B.T. and Ramakrishnan,V. (2000) Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature*, **407**, 340–348.
46. Serganov,A., Huang,L. and Patel,D.J. (2008) Structural insights into amino acid binding and gene control by a lysine riboswitch. *Nature*, **455**, 1263–1267.
47. Correll,C.C., Freeborn,B., Moore,P.B. and Steitz,T.A. (1997) Metals, motifs, and recognition in the crystal structure of a 5S rRNA domain. *Cell*, **91**, 705–712.
48. Leontis,N.B. and Westhof,E. (1998) A common motif organizes the structure of multi-helix loops in 16 S and 23 S ribosomal RNAs. *J. Mol. Biol.*, **283**, 571–583.
49. Black,D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev. Biochem.*, **72**, 291–336.
50. Reyes,J.L., Kois,P., Konforti,B.B. and Konarska,M.M. (1996) The canonical GU dinucleotide at the 5′ splice site is recognized by p220 of the U5 snRNP within the spliceosome. *RNA*, **2**, 213–225.
51. Lesser,C.F. and Guthrie,C. (1993) Mutations in U6 snRNA that alter splice site specificity: implications for the active site. *Science*, **262**, 1982–1988.
52. Collins,C.A. and Guthrie,C. (2000) The question remains: is the spliceosome a ribozyme? *Nat. Struct. Biol.*, **7**, 850–854.
53. Smith,D.J., Query,C.C. and Konarska,M.M. (2008) 'Nought may endure but mutability': spliceosome dynamics and the regulation of splicing. *Mol. Cell*, **30**, 657–666.
54. Fabrizio,P. and Abelson,J. (1990) Two domains of yeast U6 small nuclear RNA required for both steps of nuclear precursor messenger RNA splicing. *Science*, **250**, 404–409.
55. Altona,C. and Sundaralingam,M. (1972) Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation. *J. Am. Chem. Soc.*, **94**, 8205–8212.