Three-dimensional motifs from the SCOR, structural classification of RNA database: extruded strands, base triples, tetraloops and U-turns

Peter S. Klosterman[1], Donna K. Hendrix[1], Makio Tamura[2], Stephen R. Holbrook[2] and Steven E. Brenner[1,2],*

[1]Department of Plant and Microbial Biology, University of California at Berkeley, 111 Koshland Hall, Berkeley, CA 94720-3102, USA and [2]Physical Biosciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA

ABSTRACT

Release 2.0.1 of the Structural Classification of RNA (SCOR) database, http://scor.lbl.gov, contains a classification of the internal and hairpin loops in a comprehensive collection of 497 NMR and X-ray RNA structures. This report discusses findings of the classification that have not been reported previously. The SCOR database contains multiple examples of a newly described RNA motif, the *extruded helical single strand*. Internal loop base triples are classified in SCOR according to their three-dimensional context. These internal loop triples contain several examples of a frequently found motif, the minor groove AGC triple. SCOR also presents the predominant and alternate conformations of hairpin loops, as shown in the most well represented tetraloops, with consensus sequences GNRA, UNCG and ANYA. The ubiquity of the GNRA hairpin turn motif is illustrated by its presence in complex internal loops.

INTRODUCTION

The structural and evolutionary classifications of proteins, publicly available in the SCOP (1), CATH (2) and FSSP (3) databases, have been key to our understanding of protein structure. Although the number of three-dimensional RNA structures currently available is small when compared with the number of protein structures, it is still substantial, and the complexity of RNA structures is comparable with that of proteins. The biological significance of RNA is reflected in the large number of systems in which RNA plays a key role, in particular the numerous examples of RNA with catalytic activity (4,5). The three-dimensional structure of the 50S ribosomal subunit (6,7) has shown that in protein synthesis, RNA, rather than protein, catalyzes peptide bond formation.

The solution of high-resolution structures of the 30S (8) and 50S (9) ribosomal subunits has provided a greatly expanded view of RNA structure. Analysis of these structures has deepened understanding of well-known motifs such as the T-loop U-turn and identified previously uncharacterized motifs (10). Analysis focusing on base pairs has produced a comprehensive compendium of non-Watson–Crick base pairs (11,12), with an associated standard for annotations of RNA motifs (13). This analysis has recently been summarized in a base-pair centered review of RNA motifs (14).

The Structural Classification of RNA (SCOR) database (15,16) has been created to aid in the understanding of RNA structures, using data deposited in the Protein Data Bank (PDB) (17) and Nucleic Acid Database (18). Release 2.0.1 of this database contains a comprehensive classification of internal and hairpin loops contained in 497 NMR and X-ray PDB entries, released through May 15, 2003. Here we describe findings of the classification that have not been reported previously, based on analysis of structures classified in releases 1.2 and 2.0.1 of the SCOR database. These findings include the extruded strand motif, which is frequently involved in tertiary interactions, a classification of internal loop triples that reflects the local topology, and the conservation and variability of the structures of hairpin tetraloops.

THE EXTRUDED HELICAL SINGLE STRAND

The great majority of entries in the loop classification are simple, well-recognized motifs, such as *stacked duplexes with one non-Watson–Crick pair*, most commonly a single G·U wobble base pair, and the single, looped-out A base. Our classification identified the *extruded helical single strand*, another simple, yet previously unrecognized, RNA motif, found in both internal and hairpin loops and frequently associated with tertiary interactions. This motif is unusual in being characterized by bases not involved in local base pairs; however, the extruded bases are frequently paired in a tertiary interaction. The extruded strand consists of two or more stacked residues from a single strand, forming a mini-helix,

*To whom correspondence should be addressed. Tel: +1 510 643 9131; Fax: +1 208 279 8978; Email: brenner@compbio.berkeley.edu
Present address:
Peter S. Klosterman, Department of Pediatrics, University of California, San Francisco, 3333 California Street, Box 1245, San Francisco, CA 94118-1245, USA
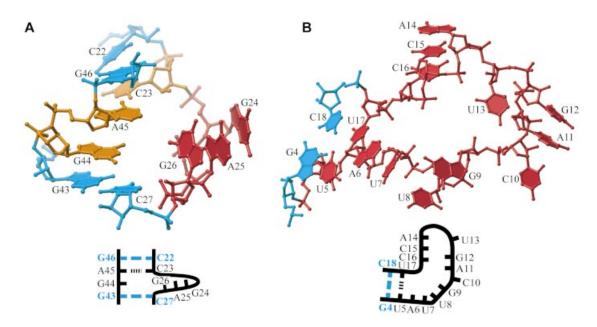
Figure 1. Loops with an *extruded helical strand*. In the three-dimensional Ribbons (89) figures that follow, the residues involved in Watson–Crick base pairs are drawn in cyan. (**A**) Three-base *extruded helical strand*, consisting of residues G24, A25 and G26, from the lead-dependent ribozyme, entry 429d (30). The extruded residues are in red and the remaining loop residues, in the main stack, are in orange. (**B**) Thirteen base hairpin loop with several independent stacks, from the spliceosomal U2B"–U2A′ protein complex bound to a fragment of U2 small nuclear RNA, entry 1a9n (33). No protein residues are shown, to emphasize the RNA structure, although the protein structure plays an important role in determining the structure of this loop. The stacks consist of the closing duplex stack, containing residues G4 through U7 as well as U17 and C18, two two-base stacks, one including U8 and G9, the second A11 and G12, and one three-base stack, containing A14 through C16; the last three stacks are extruded strands. The conventions used for the sketches are as follows. Watson–Crick base pairs are drawn as solid lines in cyan, non-Watson–Crick pairs as dashed lines. Vertical and curved lines denote the sugar-phosphate backbone. Base stacking is denoted by a series of equally spaced parallel line segments, such as bases G24 through G26 in (A); looped-out bases are drawn facing out of the backbone, e.g. bases C10 and U13 in (B). Bases involved in tertiary interactions, interacting with the bases in the loop but arising from outside the loop sequence, are drawn in gray, e.g. base A15 in Figure 5.

which is extruded from the surrounding double-stranded helix (Fig. 1A). We call this motif an interaction element because it appears to mediate RNA–RNA and RNA–protein interactions. A list of examples from various structures is found in Table 1. The loop naming convention follows SCOP; hairpin loops are identified by the structure's PDB id, the chain id if present, followed by the beginning and ending residue ids. Internal loop identifiers contain PDB id, chain id and beginning and ending residue ids of both strands. Bulge loops, internal loops in which residues from only one strand are not Watson–Crick paired, are identified by a series of residues followed by a comma and zero, to distinguish them from hairpin loops.

This motif has clear functional importance in many of the examples listed. Three types of functionality have been identified: RNA–RNA recognition, RNA–protein recognition and participation in RNA catalysis. A particularly well-studied example of an extruded strand involved in RNA–RNA recognition is the strand consisting of residues A1492 and A1493 in the 30S ribosomal subunit, which interacts directly with the tRNA anticodon–mRNA helix and plays a critical role in mRNA decoding. A series of elegant structural studies focusing on these residues have provided substantial insight into the mechanism underlying the fidelity of message decoding and the loss of fidelity on binding by aminoglycoside antibiotics such as paromomycin; representative entries include 1pbr (19), 1ibm (20), 1n32 (21) and 1j5a (22). Higher resolution structures containing a fragment of the 16S rRNA complexed with aminoglycoside antibiotics have

provided detailed structural understanding of the activity of these antibiotics, in entries 1j7t (23), 1lc4 (24), 1mwl (25) and 1o9m (26). The three-residue extruded strand in the group I intron A-rich bulge, entry 1gid (27), provides another example of RNA–RNA recognition. Here the motif nucleates folding of a substructure containing the bulge and a three-helix junction.

Extruded strands are also involved in RNA–protein recognition. Examples include the three-residue UCU strand, which is a central part of the Tat binding site of HIV-1 TAR RNA, entry 397d (28), and the three-residue extruded strand in the universal core of the signal recognition particle (SRP), entry 1dul (29), where residue A140 of this strand contacts both invariant arginine residues in a highly conserved amino acid sequence of the 4.5S RNA M domain. We have identified two examples of extruded strands involved in RNA catalysis. In the lead-dependent ribozyme (leadzyme), entry 429d (30), the cleavage site is residue G24, the 5′ residue of a three-residue extruded strand. The structure of domains 5 and 6 of the group II self-splicing intron, entry 1kxk (94), contains a two-residue extruded strand; the 5′ residue of this strand, A880, is the branch point nucleotide, whose 2′ hydroxyl attacks the phosphodiester bond at the 5′ splice site.

Analysis of ribosomal structures has identified unusually long extruded strands, such as the loop 1jj2:9:549–552,607, which contains a four base extruded strand. The 5′ residue of this strand lies on the surface of the ribosome and is highly exposed, so that it may have functional importance in interaction with another molecule. Other noteworthy extruded

Table 1. Examples of the *extruded helical strand*

Loop id	Structure description	Sequence	Reference
1pbr:7,20–21	16S ribosomal RNA A site with paromomycin	A,**AA**	(19)
1ibm:a:1408,a:1492–1493	30S ribosomal subunit with mRNA, anticodon stem–loop	A,**AA**	(20)
1gid:a:135,a:183–188	Group I self-splicing intron	U,**AAUAAG**	(27)
397d:a:23–25,0	HIV TAR RNA	**UCU**,0	(28)
429d:a:23–26,b:44–45	Lead-dependent ribozyme	**CGAG**,GA	(30)
1dul:b:140–143,b:168	Signal recognition particle, ribonucleoprotein core	**ACCA**,A	(29)
1kxk:a:8–10,a:62	Group II self-splicing intron	UA**U**,G	(92)
1jj2:9:549–552,607	*H.marismortui* 50S ribosomal subunit	**ACAA**,G	(9)
1jj2:0:1102–1107,0	*H.marismortui* 50S ribosomal subunit	**CCCCAA**,0	(9)
1j5e:a:505–510,0	*T.thermophilus* 30S ribosomal subunit	**GGCCAA**,0	(8)
1mzp:b:7–15,0	*S.acidocaldarius* protein L1 with 55mer 23S rRNA fragment	**CGUAGGA**U**A**,0	(31)
1aud:b:24,b:39–45	U1A-UTR RNA	A,A**UUGCAC**	(35)
1a9n:q:5–17	Spliceosomal U2B"–U2A′ protein, fragment of U2 small nuclear RNA	UA**UUGCAGUACCU**	(33)
1ec6:d:5–16	Nova-2 Kh3 K-homology RNA-binding domain, 20mer RNA hairpin	ACCU**AGAU**CACC	(37)
1ehz:z:54–60	Yeast phenylalanine tRNA	5MU_PSU_C_G_1MA_**U**_**C**	(93)
1c0a:b:632–638	Aspartyl-tRNA synthetase: tRNA asp	C_U_QUO_**U**_**C**_A_C	(38)

Residues contained in an extruded strand are indicated in boldface.

strands observed in ribosomal structures include the loops 1jj2:0:1102–1107,0 and 1j5e:a:505-510,0; both contain two independent extruded strands. In the 1jj2 loop, one extruded strand consists of residues C1102, C1103, C1104 and C1105, with C1102, C1103 and C1104 engaged in a tertiary interaction through Watson–Crick base pairs. The second extruded strand in this loop consists of A1106 and A1107. The 1j5e loop similarly contains one extruded strand with sequence GGCC, followed by a second with sequence AA. As with 1jj2 bases 1102–1104, the first three residues of the 1j5e loop, sequence GGC, are remotely Watson–Crick paired. The longest single extruded strand was observed in the recently published structure of the *Sulfolobus acidocaldarius* L1 protein complexed with a 55-nt fragment of the *Thermus thermophilus* 23S rRNA, entry 1mzp (31). This strand consists of residues G8, U9, A10, G11, A13; four of the five residues are stabilized by tertiary non-Watson–Crick base pairs, the fifth by a tertiary Watson–Crick base pair.

The structure of these loops emphasizes the importance of stacking interactions in determining RNA structure (32). The extruded strand is also seen in several of the larger, more complicated hairpin loops. The longest of these is the 13 residue hairpin loop in entry 1a9n (33), from a fragment of U2 small nuclear RNA (Fig. 1B). This loop contains two extruded strands of two residues and one of three residues. Similar loops are found in structures of the U1A spliceosomal protein complexed with an RNA hairpin, entry 1urn (34), the U1A polyadenylation element, entries 1aud (35) and 1dz5 (36), and the Nova KH domain aptamer, entry 1ec6 (37). Note that these six structures are all of RNA–protein complexes, in which the loops are tightly bound to the protein and the extruded strands are contained within protein pockets.

Among the hairpin loops containing an extruded strand are the well-known tRNA T-loops, which contain a two-residue extruded strand that forms part of the interface with the tRNA D-loop. Other examples of this motif are found in tRNA anticodon loops in complexes of *Escherichia coli* tRNA (Asp) with synthetase, entries 1c0a (38) and 1efw (39); in these cases the extruded strand lies in a protein pocket and one of the bases stacks with a phenylalanine residue. Additional examples in tRNA structures include two-residue extruded

strands in the D-loops of entries 1qu2 (40) and 2fmt (41). This element is also found in shorter hairpin loops, including the tetraloops in structures 1biv (42), 1mnb (43) and 1rht (44), the pentaloop in entry 1a4d (45), and the heptaloop in entry 1bzu (46).

We observe that an extruded strand is involved in crystal contacts in entries 1duh (47) and 429d. RNA crystal growth is the result of the production of an extended RNA–RNA complex, so the involvement of this element in crystal contacts seems similar to the role of the A-rich bulge in nucleating the folding of the group I intron. Sequence-specific recognition of another nucleic acid is frequently involved in the function of non-coding RNAs (48). Such recognition may be an important role for this motif, which presents extrahelical structure and is more ordered than a single looped-out base. It is more flexible than the U-turn or a base triple, thus is compatible with an induced fit mechanism of complex formation. The binding of the 50S ribosomal subunit to messenger RNA and cognate transfer RNA in the A site is a well-studied case of induced fit binding. In free rRNA, A1492 and A1493 are looped in, with a stack interruption between G1491 and A1492 and continuous stacking among A1492, A1493 and G1494. On binding of either paromomycin or mRNA with tRNA, A1492 and A1493 loop out, forming the extruded strand, and the nearby residue G530 rotates from the *syn* to the *anti* conformation (21).

As is evident in Figure 1A, the first or 5′ residue of an extruded strand is the most exposed, so that this residue is functionally highly important. Biological evidence for this importance includes the invariance of the 5′ residue in group IC1-2 introns (49) (http://www.rna.icmb.utexas.edu/), the interaction of Tat with the first uridine residue of the UCU bulge, and the observations that the 5′ residue is the cleavage site of the leadzyme and that the 5′ residue of the extruded strand in the SRP RNA–protein complex contacts the invariant arginines.

Classically, RNA motifs have been defined by their sequence; these motifs have been described as forming a distinctive conformation independent of their context (50). We observe a tendency for the extruded strands to contain either all purines or all pyrimidines, particularly in extruded strands

with an apparent functional role. Among the short internal loops characterized by this element, there are three all-purine sequences (entry 1gid: AAG, 429d: GAG, 1byj: AA), one all-pyrimidine sequence, UCU, in entry 1anr (95) and two with mixed sequence, in the core of the signal recognition particle, entries 1cq5 (96), 1duh and 1dul (1cq5: UAC, 1duh: ACA, 1dul: ACC), and the group II intron (1kxk: AU). Two of the strands with all-purine sequences are involved in RNA–RNA interactions: the group I intron, entry 1gid and the leadzyme, entry 429d; while in the remaining case, entry 1byj (97), the all-purine strand is part of a complex with an aminoglycoside antibiotic. The all-pyrimidine extruded strand is involved in an RNA–protein complex, the Tat binding site of HIV-1 TAR RNA. The three structures of the signal recognition particle RNA containing this motif all contain a three-residue mixed purine–pyrimidine *extruded helical strand* and an adjacent looped-out base. Remarkably, the structures differ in which residues form the extruded strand. The biological significance of this structural variability is unclear; it illustrates the flexibility of this motif. This flexibility may play a role in the assembly or function of the ribonucleoprotein complex. In tRNA structures, there is a tendency for the two-residue extruded strand found in the T-loop to contain two pyrimidines; this is seen in five of the eleven distinct (by species and tRNA function) kinds of tRNAs with structures available. Of the remaining kinds of tRNAs, five contain T-loop extruded strands with sequence RU and one with sequence AA. Strands with mixed purine–pyrimidine content prevail in the long loops, which break up into several independent, small strands in complex with proteins, in entries 1aud, 1a9n, 1ec6 and 1urn. These make up six of the nine distinct extruded strands in the four structures.

The position of the extruded strand relative to the main helical stack varies from structure to structure. We compared the A880, U881 branch point extruded strand in the group II self-splicing intron, entry 1kxk, with the U59, C60 T-loop strand in tRNA(Phe), entry 1ehz, which have the same secondary structural context. Both are two-residue strands connected to a continuously base-paired and stacked helix. The residues of the main helix superpose well: the C1′ atoms of seven main stack residues adjacent to the extruded strands superpose with an r.m.s. fit of 0.81 Å. The orientation of the extruded strand is similar in the two structures, with the bases perpendicular to the axis of the main stack, but the location of the extruded strand is significantly different in the superposed structures; the average distance separating the glycosidic nitrogens (N1 of pyrimidines, N9 of purines) of corresponding bases is 5.8 Å.

A recently published review of RNA motifs (14) states that RNA motifs are arrays of non-Watson–Crick base pairs. This useful definition is shown here to be too narrow for describing extruded strands, which characteristically contain unpaired bases forming an independent stack. In accord with the dynamic role of extruded strands, they frequently take part in tertiary base pairing; nevertheless, the inclusion of tertiary interactions does not overcome the limitation of the base-pair based definition and fails to recognize the fundamental commonality amongst the extruded strands. As pointed out by Doherty *et al*. (51), the 5′ A39 base of the extruded strand in the signal recognition particle structure (entry 1dul) is not base paired, but rather lies in a protein pocket;

interestingly, the hydrogen bonding between this base and its protein pocket is very similar to that seen in Type I base triples.

INTERNAL LOOP TRIPLES

Base triples and quadruples are well-recognized elements of RNA structure, triples having been identified in the early tRNA crystal structures (52,53). Our loop classification describes the three-dimensional context of base triples contained within internal loops. This terminology is based on the well-known dinucleotide platform terminology (54) and describes the relation of the triple residues to the adjoining helical stack. The internal loop base triple containing categories in our classification include *loops with a dinucleotide platform in a triple, loops forming base triples with base pair N+1*, triples with base pairs N+2, N+3, N–1 and N–2, as well as *loops with multiple triples*. This base-pair oriented classification counts the distance from the least-paired base in the triple (described below) to the base pair with the other two bases. Looped-out bases do not contribute to the count in this nomenclature, since, e.g. a triple with base pair N–1 which contains a looped-out base is structurally more similar to a triple with base pair N–1 lacking a looped-out base than to a triple with base pair N–2. This decision is guided by the observation that U-turns are frequently interrupted by looped-out bases, which do not alter the essential structure of the U-turn, e.g. the hairpin loops in entries 1a4t (55), 1qfq (56) and 2tob (57). These internal loop triples are distinguished from base triples that result from tertiary interactions. We summarize here 47 non-redundant triple-containing internal loops (Table 2).

The most commonly found triple-containing internal loops, comprising 20 of the 47 loops, are *loops with a dinucleotide platform in a triple*. Dinucleotide platforms consist of bases that are adjacent in sequence and side-by-side, coplanar, and non-Watson–Crick base paired in structure. This well-recognized structural motif is an extension of the adenosine platform (58). There is no strong sequence pattern in the structures surveyed. We have distinguished between simple dinucleotide platforms, which lack a base triple (Fig. 2A), and dinucleotide platforms in triples, in which the platform is part of a base triple (Fig. 2B). Simple dinucleotide platforms are relatively uncommon; only two clear examples, in loops 1gid:a:218–220,a:253 and 1j5e:a:641–642,0, were identified in these structures.

The three-dimensional classification of internal loop triples is based on the relationship between the sequence numbers of the triple residues and those of the surrounding helical stack. This description begins with the identification of the least-paired base in the triple (N). The procedure for identifying the least-paired base is as follows. In the great majority of the triples analyzed, one of the three potential base pairs involves either two or three hydrogen bonds and both bases contact along their Watson–Crick edges (13). This occurs most often because the two bases are Watson–Crick paired (Fig. 3A). The least-paired base is the remaining base in the triple. In most triples that lack this Watson–Crick/Watson–Crick pairing, two bases are paired by two hydrogen bonds and the remaining, least-paired, base has only one base–hydrogen bond with either of the other two bases (Fig. 3B). Hydrogen bonds

Table 2. Internal loop triples

Dinucleotide platforms in triples: S-turn/loop E motif	1j5e:a:888–894,a:905–909 1jj2:0:159–163,0:172–177
	1jj2:0:210–215,0:225–229 1jj2:0:291–295,0:356–361
	1jj2:0:568–571,0:586–590 1jj2:0:2690–2694,0:2701–2704
	1jj2:9:76–80,9:102–105 430d:a:7–12,a:19–23
	483d:a:2653–2657,a:2664–2667
Other dinucleotide platforms in triples	1eht:5–7,27–29 1esy:a:5,a:14–15 1gid:a:224–227,a:247–249
	1j5e:a:748,0 1j5e:a:1227,0 1j5e:a:1402–1403,a:1498–1500
	1jj2:0:40,0:441–442 1jj2:0:831–832,0:850 1jj2:0:1010,0
	1mms:c:1088–1090,c:1101 1qa6:c:138–139,0
Loops forming base triples with base pair N+1	1cq5:a:14–18,a:29–33 1fmn:8–13,24–28
	1j5e:a:595,0 1jj2:0:2492–2495,0:2525–2529
Loops forming base triples with base pair N–1	1exy:a:21–22,0 1f1t:a:30–32,0 1j5e:a:55,0 1jj2:0:2474–2475,0
Loops forming base triples with base pair N+2	1akx:a:23–24,0 1biv:a:10,0 1d0t:a:7,0
	1jj2:0:1894–1895,0 1jj2:0:1898,0:1939–1941
	1jj2:0:2096,0:2648–2649 1ull:a:23–24,0
Loops forming base triples with base pair N–2	1jj2:0:2291,0
Loops forming base triples with base pair N+3	1j5e:a:64–65,0 1jj2:0:2856,0:2901–2902
Loops with multiple triples	1cx0:b:140–143,b:162 1f1t:a:9,a:24–27
	1j5e:a:372–374,0 1j5e:a:410–416,a:427–432
	1j5e:a:1052–1055,a:1205–1206 1j5e:a:1058,a:1199–1202
	1jj2:0:959–963,0:1005–1007 1jj2:0:1426–1428,0:1438
	1jj2:0:2557,0:2576–2578

These were identified in the structures in the PDB entries whose coordinates were released before October 4, 2000, plus the 2.4 Å structure of the 50S ribosomal subunit from *H.marismortui*, entry 1jj2, and the 3.05 Å structure of the 30S ribosomal subunit from *T.thermophilus,* entry 1j5e. This list excludes a small number of poorly formed or borderline loops that are difficult to classify. Loops from distinct PDB entries are considered redundant if they have the same structure and arise from the same species.

between a base atom and a 2′ hydroxyl group are also included in this analysis. If the least-paired base is in the major groove of the pair between the other two bases, this is termed a major groove triple, including the base pair N+1, N+2 and N+3 triples, with the least-paired base contacting a 3′, subsequent base pair. Conversely, a minor groove triple, either base pair N–1 or N–2, is formed when the least-paired base lies in the minor groove and the least-paired base contacts a 5′ or previous base pair. We have identified two exceptional, symmetric triples in which it is difficult to identify a least-paired base, both from the 30S ribosomal subunit structure (1j5e) and both occurring in *loops with multiple triples*. An example (Fig. 3C) is the A411·A414·A430 triple in loop 1j5e:a:410–416,a:427–432. Here both the A411·A430 pair and the A430·A414 pair are two-hydrogen bond *trans* Watson–Crick/Hoogsteen pairs.

A total of 18 of the 20 dinucleotide platforms in triples are major groove triples, in which the least-paired base is the 5′ base of the platform and forms a triple with the N+1 pair, containing the 3′ base of the platform. Among the 47 triple-containing loops, we have identified two minor groove dinucleotide platforms in triples, in loops 1jj2:0:1010,0 and 1j5e:a:1227,0. Here the least-paired base is the 3′ base of the platform and forms a triple with the N–1 pair, which contains the 5′ base of the platform and is Watson–Crick in both cases.

Following *loops with a dinucleotide platform in a triple* in frequency are *loops with multiple triples*, found in nine loops. This category reflects the tendency for base triples to be adjacent in sequence and consecutively stacked in three-dimensional structure (Fig. 2C). Of the nine examples in this list, two, 1jj2:0:2557,0:2576–2578 and 1cx0:b:140–143,b:162 (98), contain three adjacent base triples; the others all contain two adjacent triples.

Also common are *loops forming base triples with base pair N+2*, seen in seven instances. We also observed four each of *loops forming base triples with base pair N+1* and *loops forming*

triples with base pair N–1. These are similar to dinucleotide platforms in triples in that the least-paired base contacts the adjacent base pair in the stack. These loops, however, lack dinucleotide platforms, either because of an intervening, looped-out base, as is the case for all four N–1 triples, or because the two bases adjacent in sequence are not paired. Notable examples with unusual structure include one loop forming a base triple with base pair N–2 and two remarkable loops forming base triples with base pair N+3 (Fig. 2D).

Because of *loops with multiple triples*, the 47 loops contain 55 triples, including 39 major groove, 14 minor groove and two symmetric triples. Although the triples contain a wide variety of base pairs, two base pair configurations were found with particularly high frequency. The most common is the S-turn GUA triple (Fig. 3B), so named because it is found in the well-documented S-turn or eukaryotic loop E motif (50,59,60). This triple contains both a two-hydrogen bond *trans* Watson–Crick/Hoogsteen U·A pair (13), also known as a reversed Hoogsteen pair (61), and a one-hydrogen bond *cis* Hoogsteen/Sugar Edge U·G pair. The A and G bases are not directly paired. Ten examples of this triple were observed, including the nine examples of the loop E motif listed in Table 2, plus loop 1fmn:8–13,24–28, which is similar to a loop E motif, found in the FMN–RNA aptamer complex (62).

We also recognize the common minor groove AGC triple (Fig. 3A), in which the A base lies in the minor groove of the G–C pair. Here there is a G–C Watson–Crick pair and a *trans* Hoogsteen/Sugar Edge A·G pair, and the A and C bases are unpaired. Five examples of this triple were identified, two in the 50S subunit, one in the 30S subunit, one in the malachite green aptamer entry 1flt (72), and one in the HTLV-1 Rex peptide aptamer, entry 1exy (73) (loops 1j5e:a:55,0, 1f1t:a:30–32,0, 1exy:a:21–22,0, 1jj2:0:959–963,0:1005–1007, and 1jj2:0:2291,0).

Thirteen of the fourteen N–1, N–2, N+2 and N+3 triples contain Watson–Crick base pairs. The exception among these
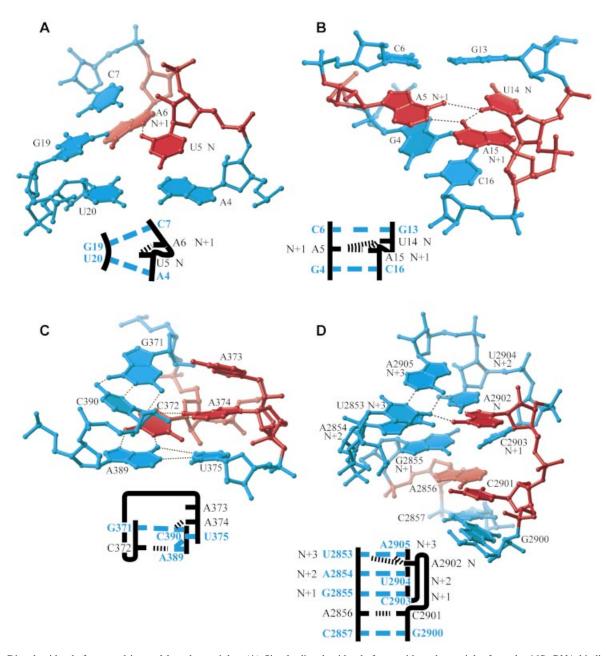
Figure 2. Dinucleotide platforms and internal loop base triples. (**A**) Simple dinucleotide platform, with no base triple, from the 16S rRNA binding site for protein S8, entry 1bgz (72). Bases U5 and A6 form the dinucleotide platform; neither base is paired with the opposite strand. (**B**) Dinucleotide platform in a triple from stem–loop SL2 of the HIV-1 psi RNA packaging signal, entry 1esy (80). The triple includes bases U14 (base N), A15 and A5, the subsequent, N+1 base pair. U14 is the least-paired, or N base in this triple because A4 and A15 pair along their Watson–Crick edges. (**C**) Loop with multiple triples from the *T.thermophilus* 30S ribosomal subunit, entry 1j5e. Bases in the triple G371·C390·A374 stack over consecutive bases in the triple C372·A389·U375. Base A373 is stacked and unpaired. (**D**) Triple with base pair N+3, from the *H.marismortui* 50S ribosomal subunit, entry 1jj2. Bases U2853, A2905 and A2902 form a triple, with A2902 the least-paired or N base since U2853 and A2905 pair along their Watson–Crick edges. This is a triple with base pair N+3 because two base pairs, G2855–C2903 and A2854–U2904, lie between base N and the other two bases in the triple.

triples, the N+3 triple in loop 1j5e:a:64–65,0, contains a similar two-hydrogen bond *cis* Watson–Crick edge/Watson–Crick edge pair, A101·G68.

Of the 55 triples, 39 are found in ribosomal subunit structures, including both of the base pair N+3 triples and the single base pair N–2 triple. Dinucleotide platforms in triples are also highly represented in ribosomal subunit structures. These make up 17 of the 20 dinucleotide platforms in triples,

and include nine from the *Haloarcula marismortui* 50S subunit (1jj2), four from structures of 23S or 28S rRNA from other species, and four from the *T.thermophilus* 30S subunit (1j5e).

A recently published structure illustrating the functional importance of these internal loop triples is the specificity domain of RNase P, entry 1nbs (63), which contains an S-turn or loop E motif, in loop 1nbs:b:139–142,b:166–170, with the
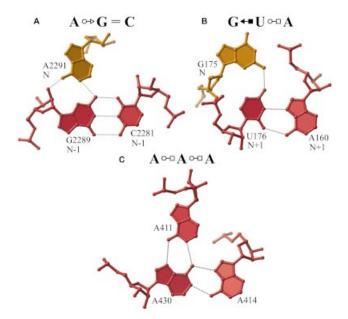
Figure 3. Base triples, annotated using the symbols of Leontis and Westhof (13). (**A**) Minor groove AGC triple, from entry 1jj2. A2291 and G2289 form a *trans* Watson–Crick/Sugar Edge pair; G2289 and C2281 are Watson–Crick paired. A2291 is the least-paired base, base N, because of the Watson–Crick pair between G2289 and C2281. The G2289–C2281 pair is base pair N–1 because base A2291 lies in the minor groove. (**B**) S-turn GUA triple, from entry 1jj2. U176 and G175 form a *cis* Hoogsteen/Sugar Edge pair; U176 and A160 form a *trans* Watson–Crick/Hoogsteen pair. G175 is the least-paired base, base N, because there is only one hydrogen bond in the G175·U176 pair, while there are two hydrogen bonds in the U176·A160 pair. The U176·A160 pair is base pair N+1 because base G175 lies in the major groove. (**C**) Symmetric triple from entry 1j5e. A411 and A430 form a *trans* Watson–Crick/Hoogsteen pair, as do A430 and A414. Here it is difficult to identify a least-paired base, since the A411·A430 and A430·A414 pairs both involve two hydrogen bonds and no two bases form a Watson–Crick/Watson–Crick pair.

associated dinucleotide platform in a triple, an S-turn GUA triple, as well as a *loop forming a base triple with base pair N+2*, 1nbs:b:182,b:228–230. The looped-out, unpaired A230 residue in this loop directly interacts with the 2′-OH of nucleotide 62 in the TψC stem of the substrate pre-tRNA.

TETRALOOPS

The structures classified in SCOR and analyzed here contain a diverse collection of hairpin loop tetraloops, of which the well-known conserved ribosomal tetraloops (64), with sequences GNRA, UNCG and CUUG, are well represented. The CUUG hairpin loop is an interesting special case. It was identified by Woese *et al.* (64) as a tetraloop because it is structurally homologous to many sequences, such as GNRA, where the first and last bases cannot Watson–Crick pair. Structurally, however, the CUUG loop is a diloop: the C and G bases are Watson–Crick paired, as shown in entry 1rng (50,65).

The GNRA U-turn structure (66) (Fig. 4A) is highly conserved; 42 of the 47 (89%) non-redundant GNRA tetraloops in the structures here surveyed (described in the caption for Table 2) have the standard conformation, with one base in the 5′ stack, three in the 3′ stack, frequently with an

interruption in the stacking between the 3′ A base and the subsequent Watson–Crick paired base, and a basal A·G *trans* Hoogsteen/Sugar Edge pair. In the most common alternate GNRA loop conformation, found in entries 1cn8 (67), 1etf (68) and 1f1t (69), the second, 'N' base, which caps the turn, is looped out or disordered (Fig. 4B); here the basal A·G pair is in the standard *trans* Hoogstein/Sugar Edge conformation. One of the GNRA tetraloops in the 50S ribosomal subunit, 1jj2:0:2738–2741, has an unusual conformation, in which all four bases lie in the 3′ stack (Fig. 4C); here the first two bases, G2738 and A2739, are Watson–Crick paired in tertiary interactions. We emphasize the variability of these loops to illustrate that these are not wholly invariant structural units.

The variability in the structures of UNCG tetraloops also lies in the location of the capping base, here the third, C base. Fifteen of the twenty (75%) non-redundant examples of UNCG loops in this sample have the standard structure, with the U and C bases in the 5′ stack, the G base in the 3′ stack, the second, and the 'N' base looped out (Fig. 4D). Alternate structures include two examples, 1raw (70) and 1ebr (71), in which the C base stacks over both the U and the G bases, one example, 1ebq (71), in which the C base stacks over the G base, and three examples, 1bgz (72), 1d6k (73) and 1tlr (74), in which both the 'N' and C bases are looped out. All of the UNCG tetraloops contain a basal *trans* Watson–Crick/Sugar Edge G·U pair.

There are many PDB entries containing phage coat protein-binding tetraloops, with consensus sequence RNYA (75) or ANYA (76); complexes of RNA containing this tetraloop and MS2 or R17 phage coat proteins crystallize readily. In the most common conformation of the RNYA tetraloop, with seven non-redundant examples, of which five contain the complex with protein and two the free RNA, two bases are in the 5′ stack and two are looped out, interacting with protein, and no bases paired (Fig. 4E). Our sample included three RNYA tetraloops with non-standard structures, all of free RNA: one, in entry 1tfn (77), with the first base in the 5′ stack, the fourth base in the 3′ stack, and the second and third bases looped out, and two, from entries 1d0t and 1d0u (78), in which the first two bases were in the 5′ stack, the fourth in the 3′ stack, and the third looped out; all three contain basal *cis* Watson–Crick/Sugar Edge pairs.

A motif which is similar in structure and function, with consensus sequence GGNG, is found in the HIV-1 psi-RNA recognition element, in both stem–loops 2 and 3. This is also a viral packaging motif, forming a specific complex with the HIV-1 nucleocapsid protein. In the two structures containing this motif in complex with protein, entries 1a1t (79) and 1esy (80), the first base terminates the 5′ stack and the remaining bases are looped out and interact with protein. We call this combination of a well-ordered 5′ stack followed by looped-out bases a protein-binding tetraloop. As is evident in Figure 4, the UNCG and RNYA loops share a more extended 5′ stack, with two bases not Watson–Crick paired, while in the GNRA loops the 3′ stack is more extended, containing three bases not Watson–Crick paired.

Of the 90 non-redundant tetraloops in the structures here surveyed, all but nine contain the sequences GNRA, UNCG, RNYA or GGNG. Remarkably, three of these, two from 30S ribosomal subunit and one from 50S subunit, exhibit the GNRA U-turn stacking even though their sequences, UCAC,
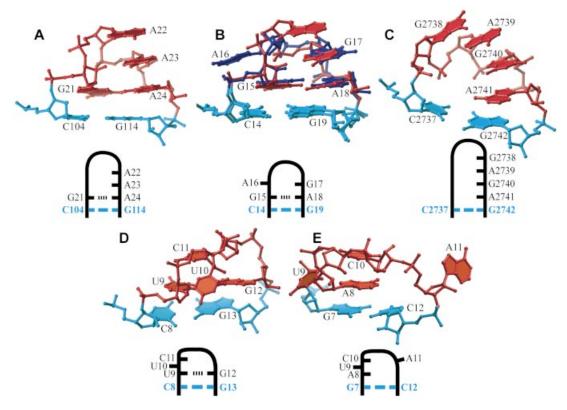
Figure 4. Tetraloops. (**A**) GNRA tetraloop in the standard conformation, a U-turn, from the hammerhead ribozyme, entry 1hmh (66). All bases lie in a single double helical stack; base G21 ends the 5′ stack, bases A22, A23 and A24 begin the 3′ stack. (**B**) Non-canonical GNRA tetraloop, not a U-turn, from the malachite green aptamer, entry 1f1t (69), superposed with the canonical GNRA tetraloop from entry 1hmh. Tetraloop residues from 1f1t are drawn in blue and labeled, those from 1hmh in red, and Watson–Crick paired closing residues of both structures in cyan. Residues C14, G15, G17, G18 and G19 of 1f1t superpose well with the corresponding residues in the 1hmh loop, A16 is looped out. The looping out of A16 allows stacking between G17 and a base from symmetry-related RNA molecule (not shown here). (**C**) GNRA tetraloop with an unusual structure, from the 50S ribosomal subunit, entry 1jj2. Here base C2737, part of the closing Watson–Crick pair, ends the 5′ stack; G2738, A2739, G2740 and A2741 all lie in the 3′ stack. (**D**) UNCG tetraloop in the standard conformation, from the *T.thermophilus* ribosomal S15–rRNA complex, entry 1dk1 (90). Bases U9 and C11 lie in the 5′ stack, G12 in the 3′ stack, and U10 (the 'N' base) is looped out. (**E**) RNYA tetraloop in the standard conformation, from the RNA aptamer–MS2 phage coat protein complex, entry 5msf (91). Bases A8 and C10 lie in the 5′ stack, while U9 and A11 are looped out.

AGCC and UAAC, are very different from the GNRA consensus.

U-TURNS IN INTERNAL LOOPS

The well-known U-turn motif (81–85) is widespread among hairpin loops. The most common U-turn sequences are UNR and GNRA. The U-turn caps a helical stack; in its pure form, all bases are well stacked. In addition to the loops in which all bases are stacked, such as GNRA tetraloops and tRNA anticodon and T-loops, there are many examples where the U-turn is decorated with looped out bases, which do not alter the conformation of the stacked bases. This motif is also present in the large, highly asymmetric internal loop found in structures of an RNA aptamer in complex with the AMP ligand, entries 1am0 (70) and 1raw (86). Here residues G8, A9 and A10 and the AMP ligand form a GNRA tetraloop, with the AMP replacing the fourth residue, here in a *trans* AMP·G Watson–Crick/Hoogsteen pair, the stack capped by the U-turn is extended by a non-Watson–Crick *trans* Watson–Crick/Hoogsteen G7·G11 pair, and the closing base pair is G6·C35. A similar example is found in loop 1mzp:b:7–15, from the recently published structure of the L1 protein complexed with

a 55-nt 23S rRNA fragment (Fig. 5). Here the GNRA tetraloop consists of residues G39, A40, A41 and A15, there is a *trans* Watson–Crick/Sugar Edge A15·G39 pair, and the stack capped by the U-turn is extended by a non-Watson–Crick *trans* Watson–Crick/Hoogsteen U38·A42 pair. We note that these U-turns exhibit both of the canonical base-backbone interactions described by Gutell *et al.* (85); we have also identified several hairpin loops with the U-turn base stacking and backbone turn structure that lack these base-backbone interactions.

CONCLUSIONS

The classification of the internal and hairpin loops in 497 PDB entries, publicly available at the SCOR web site, http://scor.lbl.gov, provides a comprehensive database of information about RNA structural motifs. Our findings emphasize the importance of base stacking, as illustrated in the identification of a previously unrecognized structural element, the *extruded helical single strand*, most commonly consisting of two or three residues in an independent stack. This motif plays an important functional role in the short internal loops that it characterizes.
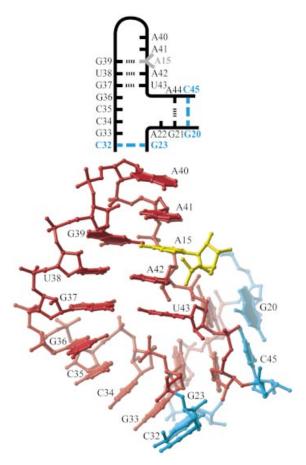
Figure 5. GNRA-like U-turn in a large, highly asymmetric internal loop, from the *S.acidocaldarius* protein L1 with a 55mer 23S rRNA fragment, entry 1mzp (31). G39, A40, A41 and A15 have the classic stacking pattern of a GNRA U-turn; this motif is extended by the non-Watson–Crick U38·A42 and G37·U43 pairs. This remarkable loop is also involved in three adjacent base triples, not shown here.

We have identified multiple examples of several categories of internal loop base triples, using a classification which defines the relationship of the bases in the triple to the surrounding helical stack. These include the dinucleotide platform in a triple and triples with base pair N+1, N+2, N+3, N–1 and N–2. Here the plus and minus sign refers to major groove and minor groove triples, respectively. We describe a procedure for distinguishing major groove from minor triples, through the identification of the least-paired base, and describe two unusual, symmetric triples, both in *loops with multiple triples*, which cannot be readily described as either major or minor groove triples. We also describe a commonly found triple motif, the minor groove AGC triple.

An overriding theme which emerged from the classification is that tetraloops adopt a limited repertoire of primary conformations, with detailed variations. The importance of the U-turn motif is emphasized by its presence in two distinct extended, highly asymmetric internal loops. GNRA, UNCG and RNYA loops dominate the well-populated tetraloop class. In these loops, although the majority of examples share a common structure, there are several instances in which apical residues are found in alternate conformations.

The analysis of the internal and hairpin loops classified in the SCOR database demonstrates the power of such classification to recognize new structural elements and canonical patterns of structure and to describe RNA structural variation.

ACKNOWLEDGEMENTS

REFERENCES

1. Murzin,A.G., Brenner,S.E., Hubbard,T. and Chothia,C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, **247**, 536–540.
2. Orengo,C.A., Michie,A.D., Jones,S., Jones,D.T., Swindells,M.B. and Thornton,J.M. (1997) CATH—a hierarchic classification of protein domain structures. *Structure*, **5**, 1093–1108.
3. Holm,L. and Sander,C. (1996) Mapping the protein universe. *Science*, **273**, 595–603.
4. Kruger,K., Grabowski,P.J., Zaug,A.J., Sands,J., Gottschling,D.E. and Cech,T.R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell*, **31**, 147–157.
5. Guerrier-Takada,C., Gardiner,K., Marsh,T., Pace,N. and Altman,S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, **35**, 849–857.
6. Nissen,P., Hansen,J., Ban,N., Moore,P.B. and Steitz,T.A. (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science*, **289**, 920–930.
7. Schmeing,T.M., Seila,A.C., Hansen,J.L., Freeborn,B., Soukup,J.K., Scaringe,S.A., Strobel,S.A., Moore,P.B. and Steitz,T.A. (2002) A pre-translocational intermediate in protein synthesis observed in crystals of enzymatically active 50S subunits. *Nature Struct. Biol.*, **9**, 225–230.
8. Wimberly,B.T., Brodersen,D.E., Clemons,W.M.,Jr, Morgan-Warren,R.J., Carter,A.P., Vonrhein,C., Hartsch,T. and Ramakrishnan,V. (2000) Structure of the 30S ribosomal subunit. *Nature*, **407**, 327–339.
9. 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.
10. 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.
11. Nagaswamy,U., Larios-Sanz,M., Hury,J., Collins,S., Zhang,Z., Zhao,Q. and Fox,G.E. (2002) NCIR: a database of non-canonical interactions in known RNA structures. *Nucleic Acids Res.*, **30**, 395–397.
12. Leontis,N.B., Stombaugh,J. and Westhof,E. (2002) The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.*, **30**, 3497–3531.
13. Leontis,N.B. and Westhof,E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA*, **7**, 499–512.
14. Leontis,N.B. and Westhof,E. (2003) Analysis of RNA motifs. *Curr. Opin. Struct. Biol.*, **13**, 300–308.
15. Klosterman,P.S., Tamura,M., Holbrook,S.R. and Brenner,S.E. (2002) SCOR: a structural classification of RNA database. *Nucleic Acids Res.*, **30**, 392–394.
16. Tamura,M., Hendrix,D.K., Klosterman,P.S., Schimmelman,N.R., Brenner,S.E. and Holbrook,S.R. (2004) SCOR: Structural Classification of RNA, version 2.0. *Nucleic Acids Res.*, **32**, D182–D184.

17. 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.

18. 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.

19. Fourmy,D., Recht,M.I., Blanchard,S.C. and Puglisi,J.D. (1996) Structure of the A site of *Escherichia coli* 16S ribosomal RNA complexed with an aminoglycoside antibiotic. *Science*, **274**, 1367–1371.

20. Ogle,J.M., Brodersen,D.E., Clemons,W.M.,Jr, Tarry,M.J., Carter,A.P. and Ramakrishnan,V. (2001) Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science*, **292**, 897–902.

21. Ogle,J.M., Murphy,F.V., Tarry,M.J. and Ramakrishnan,V. (2002) Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell*, **111**, 721–732.

22. Schlunzen,F., Zarivach,R., Harms,J., Bashan,A., Tocilj,A., Albrecht,R., Yonath,A. and Franceschi,F. (2001) Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. *Nature*, **413**, 814–821.

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

24. Vicens,Q. and Westhof,E. (2002) Crystal structure of a complex between the aminoglycoside tobramycin and an oligonucleotide containing the ribosomal decoding a site. *Chem. Biol.*, **9**, 747–755.

25. Vicens,Q. and Westhof,E. (2003) Crystal structure of geneticin bound to a bacterial 16S ribosomal RNA A site oligonucleotide. *J. Mol. Biol.*, **326**, 1175–1188.

26. Russell,R.J., Murray,J.B., Lentzen,G., Haddad,J. and Mobashery,S. (2003) The complex of a designer antibiotic with a model aminoacyl site of the 30S ribosomal subunit revealed by X-ray crystallography. *J. Am. Chem. Soc.*, **125**, 3410–3411.

27. Cate,J.H., Gooding,A.R., Podell,E., Zhou,K., Golden,B.L., Kundrot,C.E., Cech,T.R. and Doudna,J.A. (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science*, **273**, 1678–1685.

28. Ippolito,J.A. and Steitz,T.A. (1998) A 1.3-Å resolution crystal structure of the HIV-1 *trans*-activation response region RNA stem reveals a metal ion-dependent bulge conformation. *Proc. Natl Acad. Sci. USA*, **95**, 9819–9824.

29. Batey,R.T., Rambo,R.P., Lucast,L., Rha,B. and Doudna,J.A. (2000) Crystal structure of the ribonucleoprotein core of the signal recognition particle. *Science*, **287**, 1232–1239.

30. Wedekind,J.E. and McKay,D.B. (1999) Crystal structure of a lead-dependent ribozyme revealing metal binding sites relevant to catalysis. *Nature Struct. Biol.*, **6**, 261–268.

31. Nikulin,A., Eliseikina,I., Tishchenko,S., Nevskaya,N., Davydova,N., Platonova,O., Piendl,W., Selmer,M., Liljas,A., Drygin,D., Zimmermann,R., Garber,M. and Nikonov,S. (2003) Structure of the L1 protuberance in the ribosome. *Nature Struct. Biol.*, **10**, 104–108.

32. Saenger,W. (1984) *Principles of Nucleic Acid Structure*. Springer-Verlag, New York, 116.

33. Price,S.R., Evans,P.R. and Nagai,K. (1998) Crystal structure of the spliceosomal U2B"-U2A′ protein complex bound to a fragment of U2 small nuclear RNA. *Nature*, **394**, 645–650.

34. Oubridge,C., Ito,N., Evans,P.R., Teo,C.H. and Nagai,K. (1994) Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, **372**, 432–438.

35. Allain,F.H., Howe,P.W., Neuhaus,D. and Varani,G. (1997) Structural basis of the RNA-binding specificity of human U1A protein. *EMBO J.*, **16**, 5764–5772.

36. Varani,L., Gunderson,S.I., Mattaj,I.W., Kay,L.E., Neuhaus,D. and Varani,G. (2000) The NMR structure of the 38 kDa U1A protein–PIE RNA complex reveals the basis of cooperativity in regulation of polyadenylation by human U1A protein. *Nature Struct. Biol.*, **7**, 329–335.

37. Lewis,H.A., Musunuru,K., Jensen,K.B., Edo,C., Chen,H., Darnell,R.B. and Burley,S.K. (2000) Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome. *Cell*, **100**, 323–332.

38. Eiler,S., Dock-Bregeon,A., Moulinier,L., Thierry,J.C. and Moras,D. (1999) Synthesis of aspartyl-tRNA(Asp) in *Escherichia coli*—a snapshot of the second step. *EMBO J.*, **18**, 6532–6541.

39. Briand,C., Poterszman,A., Eiler,S., Webster,G., Thierry,J. and Moras,D. (2000) An intermediate step in the recognition of tRNA(Asp) by aspartyl-tRNA synthetase. *J. Mol. Biol.*, **299**, 1051–1060.

40. Silvian,L.F., Wang,J. and Steitz,T.A. (1999) Insights into editing from an ile-tRNA synthetase structure with tRNA(Ile) and mupirocin. *Science*, **285**, 1074–1077.

41. Schmitt,E., Panvert,M., Blanquet,S. and Mechulam,Y. (1998) Crystal structure of methionyl-tRNA(fMet) transformylase complexed with the initiator formyl-methionyl-tRNA(fMet). *EMBO J.*, **17**, 6819–6826.

42. Ye,X., Kumar,R.A. and Patel,D.J. (1995) Molecular recognition in the bovine immunodeficiency virus Tat peptide-TAR RNA complex. *Chem. Biol.*, **2**, 827–840.

43. Puglisi,J.D., Chen,L., Blanchard,S. and Frankel,A.D. (1995) Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex. *Science*, **270**, 1200–1203.

44. Borer,P.N., Lin,Y., Wang,S., Roggenbuck,M.W., Gott,J.M., Uhlenbeck,O.C. and Pelczer,I. (1995) Proton NMR and structural features of a 24-nucleotide RNA hairpin. *Biochemistry*, **34**, 6488–6503.

45. Dallas,A. and Moore,P.B. (1997) The loop E-loop D region of *Escherichia coli* 5S rRNA: the solution structure reveals an unusual loop that may be important for binding ribosomal proteins. *Structure*, **5**, 1639–1653.

46. Durant,P.C. and Davis,D.R. (1999) Stabilization of the anticodon stem-loop of tRNA(Lys,3) by an A+-C base-pair and by pseudouridine. *J. Mol. Biol.*, **285**, 115–131.

47. Jovine,L., Hainzl,T., Oubridge,C., Scott,W.G., Li,J., Sixma,T.K., Wonacott,A., Skarzynski,T. and Nagai,K. (2000) Crystal structure of the ffh and EF-G binding sites in the conserved domain IV of *Escherichia coli* 4.5S RNA. *Structure Fold. Des.*, **8**, 527–540.

48. Eddy,S.R. (2001) Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.*, **2**, 919–929.

49. 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., Pande,N., Shang,Z., Yu,N. and Gutell,R.R. (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.

50. Moore,P.B. (1999) Structural motifs in RNA. *Annu. Rev. Biochem.*, **68**, 287–300.

51. Doherty,E.A., Batey,R.T., Masquida,B. and Doudna,J.A. (2001) A universal mode of helix packing in RNA. *Nature Struct. Biol.*, **8**, 339–343.

52. Robertus,J.D., Ladner,J.E., Finch,J.T., Rhodes,D., Brown,R.S., Clark,B.F. and Klug,A. (1974) Structure of yeast phenylalanine tRNA at 3 Å resolution. *Nature*, **250**, 546–551.

53. Kim,S.H., Suddath,F.L., Quigley,G.J., McPherson,A., Sussman,J.L., Wang,A.H., Seeman,N.C. and Rich,A. (1974) Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. *Science*, **185**, 435–440.

54. 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.

55. Cai,Z., Gorin,A., Frederick,R., Ye,X., Hu,W., Majumdar,A., Kettani,A. and Patel,D.J. (1998) Solution structure of P22 transcriptional antitermination N peptide-boxB RNA complex. *Nature Struct. Biol.*, **5**, 203–212.

56. Scharpf,M., Sticht,H., Schweimer,K., Boehm,M., Hoffmann,S. and Rosch,P. (2000) Antitermination in bacteriophage lambda. The structure of the N36 peptide-boxB RNA complex. *Eur. J. Biochem.*, **267**, 2397–2408.

57. Jiang,L. and Patel,D.J. (1998) Solution structure of the tobramycin-RNA aptamer complex. *Nature Struct. Biol.*, **5**, 769–774.

58. 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.

59. Brunel,C., Romby,P., Westhof,E., Ehresmann,C. and Ehresmann,B. (1991) Three-dimensional model of *Escherichia coli* ribosomal 5 S RNA as deduced from structure probing in solution and computer modeling. *J. Mol. Biol.*, **221**, 293–308.

60. 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.

61. Saenger,W. (1984) *Principles of Nucleic Acid Structure*. Springer-Verlag, New York.

62. Fan,P., Suri,A.K., Fiala,R., Live,D. and Patel,D.J. (1996) Molecular recognition in the FMN-RNA aptamer complex. *J. Mol. Biol.*, **258**, 480–500.
63. Krasilnikov,A.S., Yang,X., Pan,T. and Mondragon,A. (2003) Crystal structure of the specificity domain of ribonuclease P. *Nature*, **421**, 760–764.
64. Woese,C.R., Winker,S. and Gutell,R.R. (1990) Architecture of ribosomal RNA: constraints on the sequence of 'tetra-loops'. *Proc. Natl Acad. Sci. USA*, **87**, 8467–8471.
65. Jucker,F.M. and Pardi,A. (1995) Solution structure of the CUUG hairpin loop: a novel RNA tetraloop motif. *Biochemistry*, **34**, 14416–14427.
66. Pley,H.W., Flaherty,K.M. and McKay,D.B. (1994) Three-dimensional structure of a hammerhead ribozyme. *Nature*, **372**, 68–74.
67. Mao,H., White,S.A. and Williamson,J.R. (1999) A novel loop-loop recognition motif in the yeast ribosomal protein L30 autoregulatory RNA complex. *Nature Struct. Biol.*, **6**, 1139–1147.
68. Battiste,J.L., Mao,H., Rao,N.S., Tan,R., Muhandiram,D.R., Kay,L.E., Frankel,A.D. and Williamson,J.R. (1996) Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science*, **273**, 1547–1551.
69. Baugh,C., Grate,D. and Wilson,C. (2000) 2.8 Å crystal structure of the malachite green aptamer. *J. Mol. Biol.*, **301**, 117–128.
70. Jiang,F., Kumar,R.A., Jones,R.A. and Patel,D.J. (1996) Structural basis of RNA folding and recognition in an AMP-RNA aptamer complex. *Nature*, **382**, 183–186.
71. Peterson,R.D. and Feigon,J. (1996) Structural change in Rev responsive element RNA of HIV-1 on binding Rev peptide. *J. Mol. Biol.*, **264**, 863–877.
72. Kalurachchi,K. and Nikonowicz,E.P. (1998) NMR structure determination of the binding site for ribosomal protein S8 from *Escherichia coli* 16 S rRNA. *J. Mol. Biol.*, **280**, 639–654.
73. Stoldt,M., Wohnert,J., Ohlenschlager,O., Gorlach,M. and Brown,L.R. (1999) The NMR structure of the 5S rRNA E-domain-protein L25 complex shows preformed and induced recognition. *EMBO J.*, **18**, 6508–6521.
74. Butcher,S.E., Dieckmann,T. and Feigon,J. (1997) Solution structure of a GAAA tetraloop receptor RNA. *EMBO J.*, **16**, 7490–7499.
75. Witherell,G.W., Gott,J.M. and Uhlenbeck,O.C. (1991) Specific interaction between RNA phage coat proteins and RNA. *Prog. Nucleic Acid Res. Mol. Biol.*, **40**, 185–220.
76. Convery,M.A., Rowsell,S., Stonehouse,N.J., Ellington,A.D., Hirao,I., Murray,J.B., Peabody,D.S., Phillips,S.E. and Stockley,P.G. (1998) Crystal structure of an RNA aptamer-protein complex at 2.8 Å resolution. *Nature Struct. Biol.*, **5**, 133–139.
77. Kerwood,D.J. and Borer,P.N. (1996) Structure refinement for a 24-nucleotide RNA hairpin. *Magn. Resonance Chem.*, **34**, S136–S146.
78. Smith,J.S. and Nikonowicz,E.P. (2000) Phosphorothioate substitution can substantially alter RNA conformation. *Biochemistry*, **39**, 5642–5652.
79. De Guzman,R.N., Wu,Z.R., Stalling,C.C., Pappalardo,L., Borer,P.N. and Summers,M.F. (1998) Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science*, **279**, 384–388.
80. Amarasinghe,G.K., De Guzman,R.N., Turner,R.B. and Summers,M.F. (2000) NMR structure of stem-loop SL2 of the HIV-1 psi RNA packaging signal reveals a novel A-U-A base-triple platform. *J. Mol. Biol.*, **299**, 145–156.
81. Hingerty,B., Brown,R.S. and Jack,A. (1978) Further refinement of the structure of yeast tRNA(Phe). *J. Mol. Biol.*, **124**, 523–534.
82. Holbrook,S.R., Sussman,J.L., Warrant,R.W. and Kim,S.H. (1978) Crystal structure of yeast phenylalanine transfer RNA. II. Structural features and functional implications. *J. Mol. Biol.*, **123**, 631–660.
83. Westhof,E. and Sundaralingam,M. (1986) Restrained refinement of the monoclinic form of yeast phenylalanine transfer RNA. Temperature factors and dynamics, coordinated waters and base-pair propeller twist angles. *Biochemistry*, **25**, 4868–4878.
84. Westhof,E., Dumas,P. and Moras,D. (1988) Restrained refinement of two crystalline forms of yeast aspartic acid and phenylalanine transfer RNA crystals. *Acta Crystallogr. A*, **44** (Pt 2), 112–123.
85. Gutell,R.R., Cannone,J.J., Konings,D. and Gautheret,D. (2000) Predicting U-turns in ribosomal RNA with comparative sequence analysis. *J. Mol. Biol.*, **300**, 791–803.
86. Dieckmann,T., Suzuki,E., Nakamura,G.K. and Feigon,J. (1996) Solution structure of an ATP-binding RNA aptamer reveals a novel fold. *RNA*, **2**, 628–640.
87. Jones,T.A., Zou,J.Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A*, **47** (Pt 2), 110–119.
88. Guex,N. and Peitsch,M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, **18**, 2714–2723.
89. Carson,M. (1997) Ribbons. *Methods Enzymol.*, **277**, 493–505.
90. Nikulin,A., Serganov,A., Ennifar,E., Tishchenko,S., Nevskaya,N., Shepard,W., Portier,C., Garber,M., Ehresmann,B., Ehresmann,C., Nikonov,S. and Dumas,P. (2000) Crystal structure of the S15-rRNA complex. *Nature Struct. Biol.*, **7**, 273–277.
91. Rowsell,S., Stonehouse,N.J., Convery,M.A., Adams,C.J., Ellington,A.D., Hirao,I., Peabody,D.S., Stockley,P.G. and Phillips,S.E. (1998) Crystal structures of a series of RNA aptamers complexed to the same protein target. *Nature Struct. Biol.*, **5**, 970–975.
92. Zhang,L. and Doudna,J.A. (2002) Structural insights into group II intron catalysis and branch-site selection. *Science*, **295**, 2084–2088.
93. Shi,H. and Moore,P.B. (2000) The crystal structure of yeast phenylalanine tRNA at 1.93 Å resolution: a classic structure revisited. *RNA*, **6**, 1091–1105.
94. Zhang,L. and Doudna,J.A. (2002) Structural insights into group II intron catalysis and branch-site selection. *Science*, **295**, 2084–2088.
95. Aboul-ela,F., Karn,J. and Varani,G. (1996) Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. *Nucleic Acids Res.*, **24**, 3974–3981.
96. Schmitz,U., Behrens,S., Freymann,D.M., Keenan,R.J., Lukavsky,P., Walter,P. and James,T.L. (1999) Structure of the phylogenetically most conserved domain of SRP RNA. *RNA*, **5**, 1419–1429.
97. Yoshizawa,S., Fourmy,D. and Puglisi,J.D. (1998) Structural origins of gentamicin antibiotic action. *EMBO J.*, **17**, 6437–6448.
98. Ferre-D'Amare,A.R., Zhou,K. and Doudna,J.A. (1998) Crystal structure of a hepatitis delta virus ribozyme. *Nature*, **395**, 567–574.