Quarterly Reviews of Biophysics **38**, 3 (2005), pp. 221–243. © 2006 Cambridge University Press
doi:10.1017/S0033583506004215 Printed in the United Kingdom
First published online 3 July 2006

RNA structural motifs: building blocks of a modular biomolecule

Donna K. Hendrix[1], Steven E. Brenner[1,2] and Stephen R. Holbrook[2]*

[1] Department of Plant & Microbial Biology, University of California, Berkeley, CA, USA
[2] Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Abstract. RNAs are modular biomolecules, composed largely of conserved structural subunits, or motifs. These structural motifs comprise the secondary structure of RNA and are knit together via tertiary interactions into a compact, functional, three-dimensional structure and are to be distinguished from motifs defined by sequence or function. A relatively small number of structural motifs are found repeatedly in RNA hairpin and internal loops, and are observed to be composed of a limited number of common 'structural elements'. In addition to secondary and tertiary structure motifs, there are functional motifs specific for certain biological roles and binding motifs that serve to complex metals or other ligands. Research is continuing into the identification and classification of RNA structural motifs and is being initiated to predict motifs from sequence, to trace their phylogenetic relationships and to use them as building blocks in RNA engineering.

* Author for correspondence: Dr S. R. Holbrook, Department of Structural Biology, Physical Biosciences Division, Lawrence Berkeley National Laboratory, MS 64R0121, Berkeley, CA 94720-8118, USA.
Tel.: 510-486-4304; Fax: 510-486-6798; E-mail: srholbrook@lbl.gov

1. Introduction

Our rapidly expanding knowledge of the forms and functions of RNAs in biological systems clearly illustrates that, like proteins, RNA assumes complex three-dimensional (3D) structures to perform specific roles. Unlike proteins however, RNA forms more locally stable structures, or structural motifs, that are combinatorially linked and constrained by tertiary interactions to create a 3D structure. The number of these identifiable motifs, although each customizable, appears to be relatively small. In this review, we enumerate and describe some currently known motifs, relate them to structural and functional roles and discuss current and future studies and applications in this area.

2. What is an RNA motif?

2.1 Sequence *vs.* structural motifs

Much of the confusion regarding 'RNA motifs' is due to their definition and description at different levels of detail. Traditionally, RNA structure, like protein structure, is described at the sequence or primary structure level, the secondary, tertiary and quaternary levels.

Initially, RNA motifs were identified and described at the sequence level as commonly occurring short sequences in functional RNAs, such as transfer RNA (tRNA) or ribosomal RNA (rRNA) (Woese *et al.* 1990). The variation in these sequences, within or between RNAs, is often represented by an expanded alphabet describing a consensus sequence (e.g. GNRA where N is any nucleotide and R is a purine). These conserved sequences are often given in the context of the predicted base-pairing structure (e.g. tetraloops) surrounding them.

The next level of RNA structure is the base-pairing, or secondary structure, which identifies both the canonically base-paired regions (helices) and non-paired regions (loops). This structure can sometimes be predicted by various computational methods (Zuker, 1989; Gutell, 1995; Mathews *et al.* 1998; Rivas & Eddy, 1999; Hofacker, 2003) that may or may not include pseudoknots, but generally does not include non-canonical pairing or 3D information such as details of base stacking, backbone hydrogen bonding and tertiary interactions. When multiple sequences of an RNA family are available, analysis of patterns of sequence variation can be used to infer secondary (and tertiary) structure. Such covariation analysis is possible because base pairing can be maintained with compensatory pairings. This approach has been shown to be highly accurate when compared to crystallographic results (Gutell *et al.* 2002). Conserved secondary structures identified by covariation have often been proposed as 'motifs' (Gast, 2003). RNA secondary structure has been catalogued in various databases including the comprehensive database of RNA families, Rfam (Griffiths-Jones *et al.* 2003) and RNA specific databases for ribosomal (Wuyts *et al.* 2004), ribonucleaseP (Brown, 1999), tmRNA (Zweib *et al.* 2003), SRP (Rosenblad *et al.* 2003) and others. A different type of database containing graphical representations of RNA secondary structure, RAG (Fera *et al.* 2004), has recently been introduced. Various biochemical and biophysical experimental methods can also be used to infer secondary structure (Ehresmann *et al.* 1987) or in the case of NMR (Furtig *et al.* 2003) and X-ray crystallographic methods (Holbrook & Kim, 1997) describe secondary and tertiary structure in detail.

RNA tertiary structure describes the overall 3D conformation of a single molecule as determined by crystallography, NMR or modeling methods. The 3D structure is determined by

long-range intramolecular interactions such as pseudoknots, ribose zippers, kissing hairpin loops, tetraloop–tetraloop receptor interactions, coaxial helices and other yet uncharacterized modes of interaction, and can be mediated by intermolecular interactions with ligands including metals and small molecules or other macromolecules (DNA, RNA or protein). Some of these interactions can be identified computationally (e.g. covariation analysis) or by biochemical/biophysical experiments, but in general, 3D structural information from crystallography or NMR is necessary for a complete description. This coordinate information is provided in both the Nucleic Acid Database (NDB; Berman *et al.* 1992) and the RCSB Protein Data Bank (PDB; Berman *et al.* 2000).

2.2 RNA structural motifs

RNA motifs have been defined variously as 'directed and ordered stacked arrays of non-Watson–Crick base pairs forming distinctive foldings of the phosphodiester backbones of the interacting RNA strands' (Leontis & Westhof, 2003) and 'a discrete sequence or combination of base juxtapositions found in naturally occurring RNAs in unexpectedly high abundance' (Moore, 1999). A comprehensive definition of an RNA structural motif should be based on and consist of not only base-pairing or secondary structure constraints, but a complete 3D description, including backbone conformation, all hydrogen-bonding and base-stacking interactions, and sequence preferences. In addition, a RNA structural motif may have co-factors such as bound waters, metals or other ions, to support its conformation; it may have a specific functional role or a primary role in tertiary structure formation; and it may be subject to evolutionary constraints. In order to determine these characteristics, it is necessary that the motif be frequently observed among known RNA structures. The overall 3D structure of such a recurrent motif is largely independent of the context in which it is found. Moreover, RNA structural motifs are truly structural, and there may be several sequences, seemingly unrelated, that obtain the same 3D structure. Examples include the tetraloop with sequence UMAC (M is A or C) which forms a structure almost identical to that of the GNRA tetraloop (Leontis & Westhof, 2002; Klosterman *et al.* 2004), and a tetraloop of sequence GUUA which has a fold like UNCG (Ihle *et al.* 2005). We denote these as the GNRA fold and the UNCG fold. Motifs characterized by sequence or predicted secondary structure alone are not discussed here, although they may ultimately be structurally characterized as further RNA structures are determined.

2.3 RNA structural elements *vs.* motifs

RNA structural motifs typically (but not always) comprise an entire hairpin loop or internal loop (and occasionally multiple loops) to perform their structural, recognition, interaction, or enzymatic role. Motifs can also usually be characterized or identified by sequences compatible with their structure and function. Examples of these motifs include the various tetraloops, the sarcin–ricin loop, the kink-turn, and the T-loop. Smaller, conserved RNA structures are often observed as components of these motifs or in isolation. We term these 'structural elements' or 'structural attributes'. These structural elements may be observed in several types of motifs and may lack a characteristic sequence. Examples of these structural elements include the dinucleotide platform, non-canonical base pairs, the A-minor interaction, and backbone rotamers. The A-minor interaction, for instance, is observed both in the secondary structure motif – the

Table 1. *Characteristics of RNA structural elements and motifs*

Characteristic	Element	Loop motif	Tertiary interaction motif
Size	Small, local	May span entire loop	Multiple loops or stems involved
Sequence conservation	Little or none	Often have sequence preferences/isosteric	Interaction sites
Structural conservation	By definition	Often conserved	Evolutionarily conserved
Features (i.e. pairing, stacking, turn)	Usually single feature	Multiple features/elements	Multiple in each interacting motif
Occurrence	Found within various motifs	Not nested; may occur in tertiary interaction motifs	May include multiple elements and motifs

kink-turn and the tertiary interaction motif – the ribose zipper. Most of these elements are characterized in terms of orientation of the nucleotide bases; however, recent studies (Hershkovitz *et al.* 2003; Murray *et al.* 2003; Schneider *et al.* 2004) have shown that RNA backbone conformational freedom is far more limited than suggested by the six variable torsion angles (plus glycosyl torsion) and that, in fact, only a relatively small number of conformations are observed in biological RNAs. Although further analysis is needed, these favored conformations may comprise a set of RNA 'conformational elements' as opposed to the 'base-orientation elements' that have been identified previously. Characteristics of RNA structural elements *vs.* motifs are summarized in Table 1. Specific RNA structural elements are listed and described in Table 2. In some cases the distinction between a structural element and motif is unclear, for example the U-turn that may be classified as either an element based on occurrence in multiple motifs and lack of sequence conservation or a motif based on size and multiple features. Figure 1 shows an example of a complex RNA motif, the sarcin–ricin loop (Correll *et al.* 1999) and the structural elements from which it is composed. Another viewpoint is given in Fig. 2, which shows how an element, the U-turn, can occur in multiple motifs.

2.4 Specific recognition motifs

In addition and in contrast to RNA structural motifs that occur ubiquitously in RNA biomolecules and perform general structural or functional roles, there is a large class of RNA recognition motifs that are the binding sites for specific proteins and other molecules. These specific recognition motifs may have well defined, stable and conserved 3D structures and sequence signatures, but differ necessarily from the general structural motifs in that they are narrowly distributed within an organism to perform their singular recognition function, which would not be possible if they were common motifs. Thus, these sequences may be broadly distributed phylogenetically and may even have several variants within a genome, but they do not serve general architectural roles in RNA folding and tertiary structure stabilization and do not interact with many different types of ligands. Examples of specific recognition motifs are TAR (Richter *et al.* 2002) and RBE (RRE) (Hung *et al.* 2000; Lesnik *et al.* 2002) motifs of HIV-1 and the IRE motifs found in mRNAs related to iron metabolism (Theil, 2000; Pantopoulos, 2004). Many riboregulators, riboswitches, and aptamers form potentially highly structured regions for protein or small molecule interaction, and mutations in these structures may result in a loss of regulation

Table 2. *Some RNA structural elements*

Element name(s)	Description	Found in	Reference
U turn/Uridine turn/Pi turn	A sharp bend in the phosphate-sugar backbone between the first and second nucleotides, followed by characteristic stacking of the second and third nucleotides. Original descriptions include a stabilizing hydrogen bond between the first and third residues	Hairpin loops (e.g. GNRA, TΨC loop) and internal loops	Kim & Sussman, 1976; Quigley & Rich, 1976; Holbrook *et al.* 1978; Klosterman *et al.* 2004
A-minor interaction	The insertion of minor groove edges of an adenine into the minor groove of neighboring helices. Four types have been identified	Ribose zipper, kink-turn	Nissen *et al.* 2001
S-turn	Two consecutive bends in the phosphate-sugar backbone characterized by backbone distortions and inverted sugar puckers, resulting in an 'S' shape	Loop E motif, Sarcin–ricin loop	Szewczak *et al.* 1993; Wimberly *et al.* 1993; Correll *et al.* 1999
Dinucleotide platform	Two adjacent, covalently linked, co-planar residues that form a non-Watson–Crick pairing	Internal loops, often involved in a base triple	Cate *et al.* 1996b; Klosterman *et al.* 2004
Base triples	Three hydrogen-bonded, co-planar bases with two of the bases sometimes forming a Watson–Crick pair or dinucleotide platform	Loop E motif, Sarcin–ricin loop	Nagaswamy *et al.* 2002; Klosterman *et al.* 2004
Cross-strand stack	A base on one strand stacks with a base on the opposing strand, rather than stacking with the adjacent bases on its own strand	Internal loops, e.g. bacterial loop E motif	Correll *et al.* 1997
Non-canonical base pairs	Two bases of any type interacting in a generally planar arrangement can form hydrogen bonds in characteristic patterns	Double helices	Leontis & Westhof, 2001; Nagaswamy *et al.* 2002
Extruded helical single strand	Two or three unpaired bases extruded from the main double helical stack forming an independent stack	Internal and hairpin loops	Klosterman *et al.* 2004
Backbone rotamers	Commonly occurring RNA backbone conformations	Double helices, hairpin, internal and junction loops	Duarte *et al.* 2003; Hershkovitz *et al.* 2003; Murray *et al.* 2003; Schneider *et al.* 2004

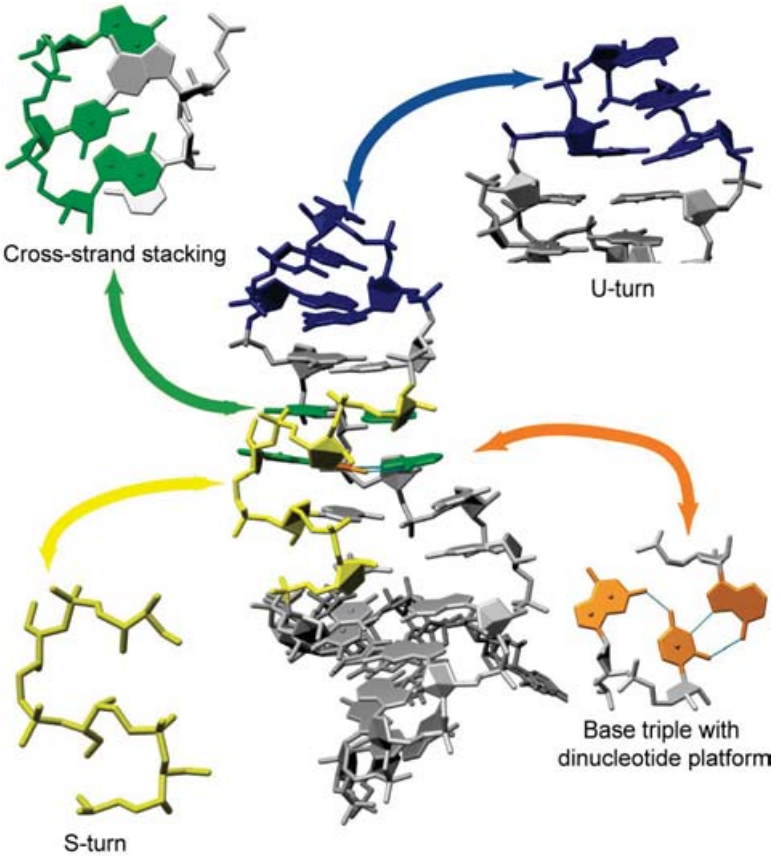

Fig. 1. The sarcin–ricin loop, depicted by elements from the rat 28S ribosomal RNA (PDB identifier 483d). The sarcin–ricin loop is an assembly of two secondary structure motifs: an internal loop (bulged G), made up of residues a:2653-2657 and a:2664-2667, and a hairpin loop (GNRA tetraloop) of residues a:2659-2662. It can also be considered to be composed of several elements. These elements include cross-strand stacking (in green and gray) with a:2655 stacking on a:2664 and a:2657 on a:2665; a U-turn (in blue) in the GNRA tetraloop; a base triple with a dinucleotide platform (orange); an S-turn (yellow). Not shown are several non-canonical base pairs and the backbone-base hydrogen bonds. Note that the base triple includes nucleotides that are also involved in cross-strand stacking.

and a corresponding disease state (Sobczak & Krzyzosiak, 2002; Wong *et al.* 2005). In the remainder of this review we will focus on general RNA structural motifs, their identification, classification, types, and roles.

2.5 Tools for identifying and classifying elements and motifs

In recent years a variety of computational, database and graphics tools have been developed that are extremely valuable in identifying, describing, characterizing and classifying RNA motifs and elements. Descriptive base-pairing nomenclatures (Leontis & Westhof, 2001; Lee & Gutell, 2004) and databases (Nagaswamy *et al.* 2002) have been developed to annotate and classify the wide variety of non-canonical base pairs observed in RNA structures. These have been incorporated into an automated assignment and graphics program (Yang *et al.* 2003) for

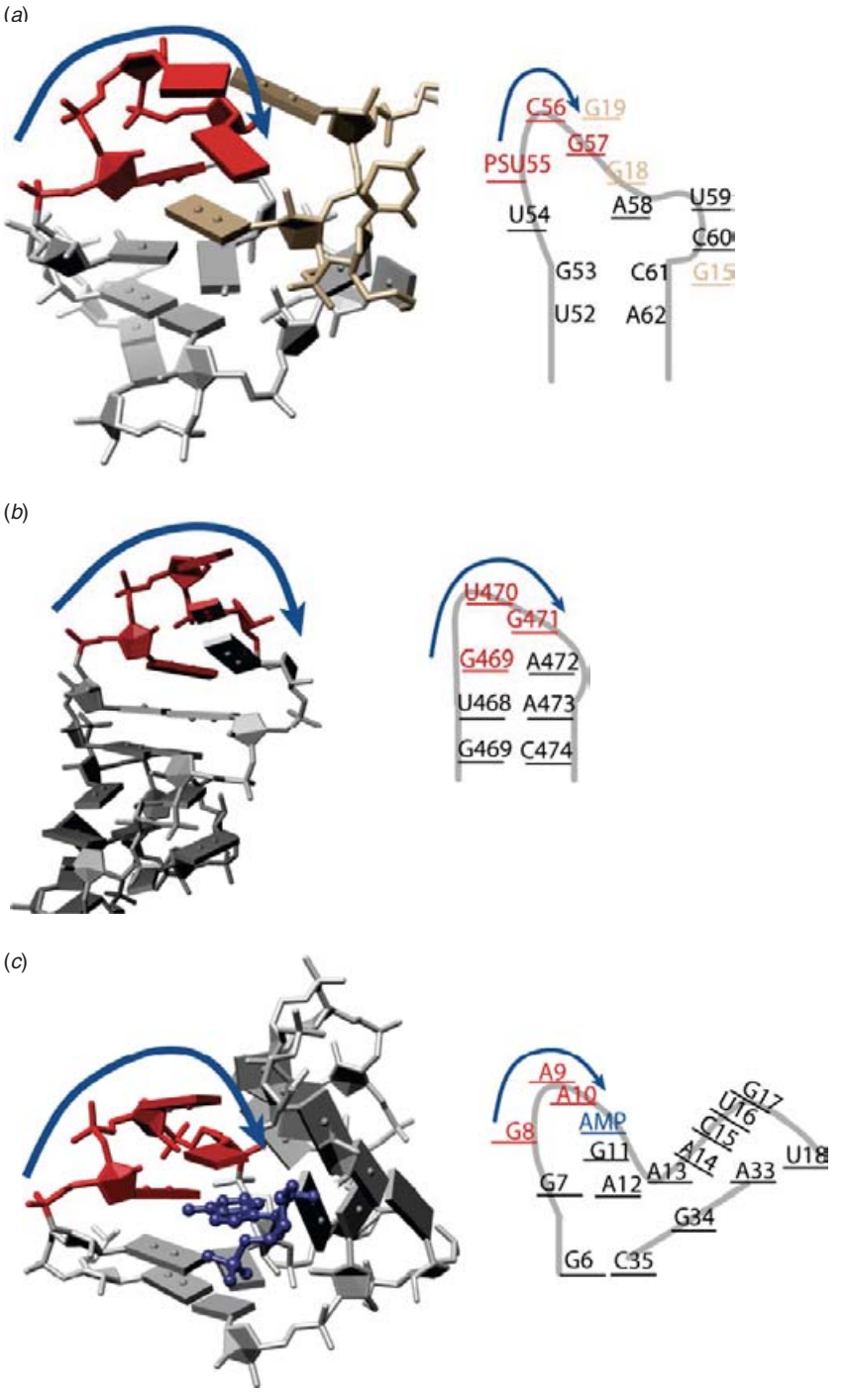

Fig. 2. The U-turn element in multiple loops. 3D representations are given on the left and corresponding 2D diagrams on the right. U-turns are colored red with a blue arrow indicating chain direction. (*a*) The U-turn in the T-loop of yeast $tRNA^{Phe}$ (PDB identifier 6tna). (*b*) The U-turn in the GNRA loop from the 23S ribosomal RNA of *Haloarcula marismortui* (PDB identifier 1jj2). (*c*) The U-turn in an internal loop, from an AMP-aptamer complex (PDB identifier 1am0).

visualization of base-pairing and stacking patterns. A web server providing a sophisticated annotation of secondary structures and some tertiary interactions has been presented by Major and co-workers (Gendron *et al.* 2001). Conformational analysis of the RNA backbone using reduced representations has also been used to characterize known motifs (Duarte & Pyle, 1998; Duarte *et al.* 2003; Hershkovitz *et al.* 2003; Murray *et al.* 2003) and even identify new motifs (Wadley & Pyle, 2004). Cluster analysis of root-mean-square deviations between pairs of RNA fragments has been used as the basis for cluster analysis of RNA loop structures to group tetraloops into various families (Huang *et al.* 2005). A graph theoretic approach has been used to search for substructure patterns in RNA structures using a vectorial approach (Harrison *et al.* 2003). Finally, the Structural Classification of RNA (SCOR), database has been developed to organize and classify RNA structural motifs (Klosterman *et al.* 2002; Tamura *et al.* 2004). SCOR currently relies on manual classification based on features recorded in the literature, computed, or observed by inspection.

3. Types of RNA structural motifs

Examination of RNA 3D structures shows that they can be considered to be composed of a combination of recurring structural motifs joined by a small number of types of tertiary interaction motifs. An example of the secondary structure and tertiary interaction motifs in the P4–P6 domain of the group I intron is shown in Fig. 3. These motifs are commonly subgrouped by secondary structure type: double helices, hairpin loops, internal loops, or junction loops. Moreover, the interaction of RNA with metal ions, small molecules, proteins and other RNAs can also be characterized by a set of recurring motifs. These various types of RNA motifs are discussed in detail below.

3.1 Helices

Although RNA is composed primarily of Watson–Crick base-paired A-form double helices, other helical forms have been observed. Although RNA double helices are generally thought to not have a sequence-structure relationship (Holbrook *et al.* 1981), a comprehensive analysis with the large dataset currently available has not been completed. Recently, the structures of a high salt left-handed RNA duplex (Popenda *et al.* 2004) and a mirror image or L-configuration (Vallazza *et al.* 2004) Spiegelmer RNA duplex were determined. While the left-handed RNA duplex of repeating (CG) units differed significantly from its Z-DNA counterpart, the Spiegelmer RNA had a very similar structure to that of its right-handed, mirror-image structure even though crystallization conditions were quite different. Structures of RNA quadruplexes have also been determined (Liu *et al.* 2002; Pan *et al.* 2003a, b). These structures include both guanine and adenine tetrads as well as bulged and looped out residues and generally differ from their DNA homologs.

3.2 Hairpin loops

Hairpin loops link the 3′- and 5′-ends of a double helix. Within the SCOR database classification, structurally characterized hairpin or external loops must close with a Watson–Crick pairing and vary in length from 2 to 14 nucleotides. The most common and well studied of these are the tetraloops. Of these tetraloops, there are at least four that are characterized by their

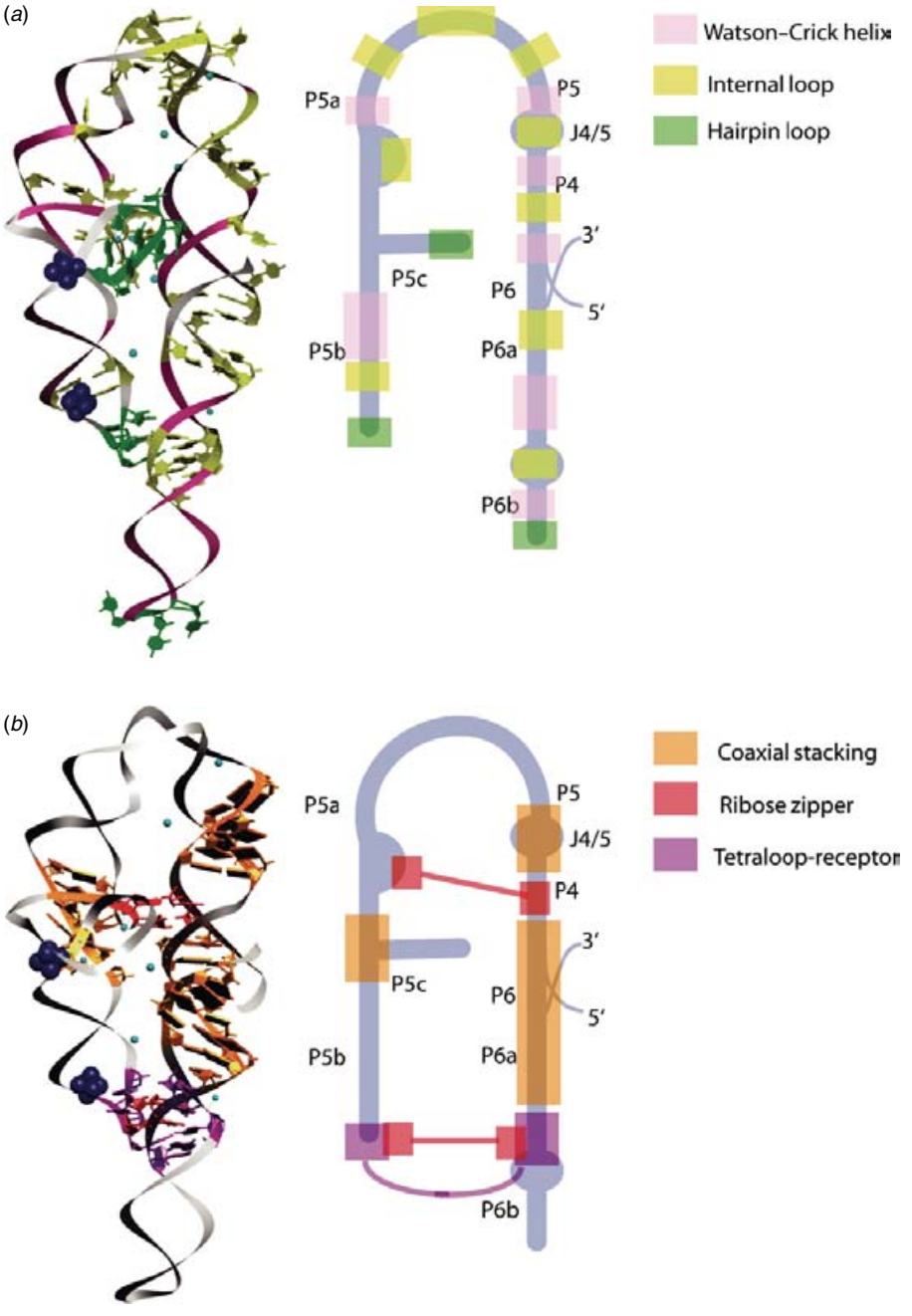

Fig. 3. Motifs in the P4–P6 domain of the group I intron structure. Motifs are truly the structural units, or building blocks, of the *Thermus thermophilus* group I intron structure (PDB identifier 1gid). (*a*) Secondary structure motifs by color: fully-paired Watson–Crick helices in pink, internal loops in yellow, and hairpin loops in green. What remains (in gray) are junction loops, the 3′ and 5′ overhang ends, and a few residues that did not meet the strict definition of Watson- Crick pairings. (*b*) Tertiary interaction motifs by color: coaxial regions of helices in orange, ribose zippers in red, and the tetraloop–tetraloop receptor interaction in purple. Note that these tertiary interactions have significant overlap. Also shown in both panels (*a*) and (*b*) are magnesium ions (cyan) and cobalt hexammine (dark blue). These metal ions are localized within critical tertiary interactions.

sequence and conserved structures: the GNRA type, the UNCG type, the ANYA type and the (U/A)GNN type. In ribosomal RNAs about 70% of tetraloops belong to either the GNRA or the UNCG families and are unusually stable thermodynamically compared to other tetraloop sequences (Antao *et al.* 1991). The well-known GNRA tetraloop (Heus & Pardi, 1991; Jucker & Pardi, 1995a; Leontis & Westhof, 2002; Correll & Swinger, 2003) is the most frequently observed tetraloop in currently available RNA structures. The GNRA tetraloop is frequently used as part of a tertiary interaction motif in the formation of tetraloop–tetraloop receptor interactions.

Other hairpin loop motifs include the T-loop (Nagaswamy & Fox, 2002) and D-loop motifs of tRNA (Quigley & Rich, 1976), the lonepair triloop (Lee *et al.* 2003), and the sarcin-ricin loop (Szewczak & Moore, 1995). These and other RNA hairpin loop motifs are summarized in Table 3. Depending on definition, the U-turn (Jucker & Pardi, 1995a), and reversed U-turn (Kolk *et al.* 1997), may be considered structural elements that are present in several motifs including the GNRA tetraloop and the T-loop. Examples of the T-loop hairpin motif and its U-turn subunit are shown in Fig. 4.

3.3 Internal loops

An internal loop separates double helical RNA into two segments by inclusion of residues that are not Watson–Crick paired in at least one strand of the duplex. Sometimes insertions on only one strand are defined as 'bulge loops', and we include this as a special case of internal loops. Two types of internal loops can be distinguished: symmetric, with the same number of nucleotides inserted on both strands; and asymmetric, with a different number of nucleotides inserted on the opposing strands. Non-canonical base pairing is common in internal loops. A frequently observed motif involves extension of double helical structure through continuous formation of non-Watson–Crick pairs in a symmetric internal loop. This double helical structure is distorted by unwinding, unstacking, and kinking formed by the non-canonical pairs. Fully paired and stacked internal loops of up to eight non-canonical pairs have been structurally observed (Vallurupalli & Moore, 2003) (loop E). Fully paired internal loops can be well described using a standardized base-pairing nomenclature such as that developed by Leontis & Westhof (2001). Such nomenclature classification and isostericity relationships are useful for prediction of these types of motifs from sequence and secondary structure (Leontis *et al.* 2002; Lescoute *et al.* 2005).

Certain asymmetric internal loop motifs have been identified and characterized as resulting in sharp turns important for tertiary structure formation. These include the kink-turn (K-turn) (Klein *et al.* 2001), reverse kink-turn (Strobel *et al.* 2004) and hook-turn (Szep *et al.* 2003). Some RNA internal loop motifs are summarized in Table 4.

3.4 Junction loops/multiloops

Junction loops are formed by the intersection of three or more double helices. These double helices are separated by single-strand sequences of zero or more residues. There are N linker (joining) sequences for N helices in a junction loop, although some of the linker sequences may be of zero length. Although junction loops have not been as systematically or extensively studied as the simpler hairpin and internal loops, some generalizations have been made for the more common three-way and four-way junctions (Lilley, 1998, 2000). Common examples of

Table 3. *Some RNA hairpin loop motifs*

Motif name(s)	Description	Reference
Lonepair triloop	Identified by covariation analysis of 16S rRNA sequences and T-loop of tRNA. Characterized by a single base pair, either Watson–Crick or non-canonical, capped by a hairpin loop containing three nucleotides. Bases immediately 5′ or 3′ to the motif are NOT base paired to one another. Consensus sequences: Type R1: UGNRA; Type R2: UUYRA; Type R3: NRWAN-; Type R4: NRYAN-; Type R5: NCNUN-	Lee *et al.* 2003
GNRA tetraloop	A common tetraloop found in ribosomal RNA, Group I intron, and Hammerhead ribozyme. The GNRA loop sequence is often closed by a C-G Watson–Crick pair. Commonly folds into the 'GNRA fold' of one base on the 5′ stack and three in the 3′ stack, and contains a U-turn. This fold is also formed by the tetraloop family of sequence UMAC	Heus & Pardi, 1991; Jucker & Pardi, 1995a; Correll & Swinger, 2003
UNCG tetraloop	A stable tetraloop found in ribosomal and other functional RNAs. Commonly forms the 'UNCG fold', with the U and C bases in the 5′ stack, the G in the 3′ stack and the N looped out. Also observed in a GUUA tetraloop	Cheong *et al.* 1990
ANYA tetraloop	Identified in aptamers binding to the MS2 coat protein. It has 2 common folds: one in a bound form, the other in the apo-form. The bound form has the 1st and 3rd bases in the 5′ stack, and the 2nd and 4th bases looped out, interacting with protein. The apo- form has the 1st base in the 5′ stack and the 4th base in the 3′ stack, forming a Watson–Crick /Sugar Edge base pair	Convery *et al.* 1998; Rowsell *et al.* 1998; Klosterman *et al.* 2004
(U/A)GNN tetraloop	Seen in RNase III endoribonucleases and the 18S rRNA of yeast, this tetraloop has first and second bases in the 5′ stack and third and fourth bases in the 3′ stack	Butcher *et al.* 1997b
CUYG tetraloop	One of the tetraloop sequences common in ribosomal RNAs, this sequence has been seen in two different forms: as a di-loopin a solution structure, with the C and G Watson–Crick-paired, and as a hairpin loop in *D. radiodurans* 23S rRNA and *T. theromphilus* 16S rRNA	Woese *et al.* 1990; Jucker & Pardi, 1995b
D-loop	In tRNA, the D-loop contains the modified base dihydrouracil. Is composed of 7–11 bases, and inserts bases into the T-loop to form the tRNA T-loop:D-loop tertiary interaction	Quigley & Rich, 1976; Holbrook *et al.* 1978
T-loop	The T-loop, originally characterized in tRNA is a 5-base hairpin closed by a trans-Watson–Crick/Hoogsteen base-pair interaction between bases N and N + 4. Contains a classic U-turn between in the first three bases. Consensus sequence is U(G/U)NR(A/U)	Nagaswamy & Fox, 2002; Krasilnikov & Mondragon, 2003

junction loops are those in tRNA and the hammerhead ribozyme. Coaxial stacking of the helices is a key feature of junction loops as observed in these and many other examples. It has been proposed that coaxial (continuous) stacking in three helix junctions occurs opposite the longest junction strand (Lescoute & Westhof, 2005). The tendency for pairwise coaxial stacking of helical arms, the importance of metal ion interactions in the induction of tertiary folding, and the importance of hairpin or internal loop–loop interactions in stabilization of the tertiary structure (Penedo *et al.* 2004) are prominent features of junction loop architecture.

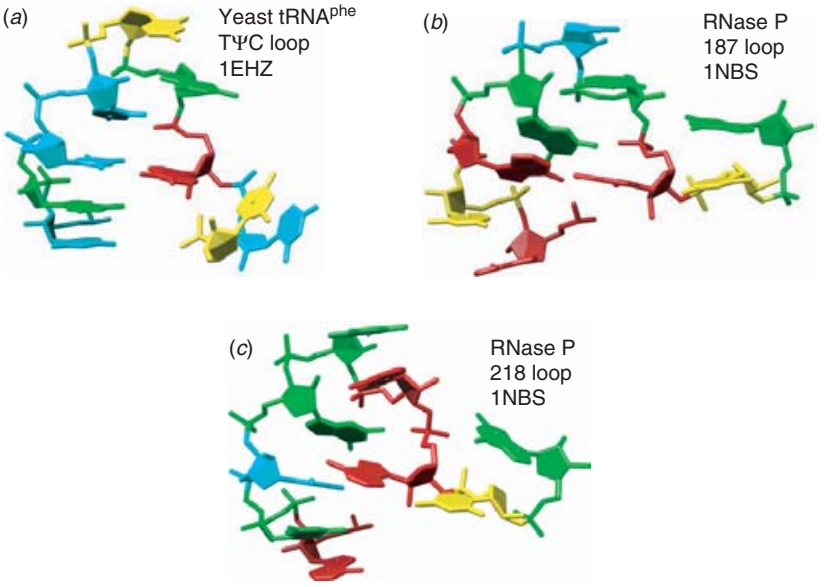

Fig. 4. Common fold of the T-loop motif. (*a*) The TΨC loop of yeast phenylalanine transfer RNA (PDB ID: 1EHZ). (*b*). A T-loop motif found in the crystal structure of the specificity domain of Ribonuclease P RNA (PDB ID 1NBS) at residue 187. (*c*). A T-loop motif found in 1NBS at residue 218. A three-residue U-turn is seen at the apex of the loop bounded by a non-Watson–Crick pair closing the T-loop. Cross-loop hydrogen bonding is observed between base and backbone. Sharp turns are often observed for residues at the 3′-end of the loop. Adenine nucleotides are colored red, guanine green, uracil light blue and cytosine yellow.

3.5 Binding motifs

A primary function of RNA is to bind ligands, either for structural stabilization, as co-factors, substrates or signals. Ligand binding is critical for ribozyme, riboswitch and splicing functions, as well as in mediating RNA–protein and RNA–RNA intermolecular and tertiary interactions. RNA ligand-binding sites often demonstrate high selectivity and specificity, although there may be more than one motif capable of tightly binding a certain type of ligand.

3.5.1 Metal binding

Two types of metal ion interactions with RNA have been described: diffuse ions that accumulate near RNA due to the electrostatic field while retaining their hydration sphere, and chelated ions that are in direct contact with RNA at a specific location, and may have some waters of hydration displaced by coordination with polar RNA atoms (Draper, 2004). Numerous examples of site-specific metal ion binding to RNA have been observed in RNA structures. One of the first of these was the extensively studied binding of metals to tRNA (Holbrook *et al.* 1977; Quigley *et al.* 1978; Shi & Moore, 2000). Although magnesium is considered the biologically relevant metal, other divalent cations, lanthanides and metal hexammines have been observed as strongly bound to tRNA. A database of metal ion-binding sites in RNA structures (MeRNA), has been established to organize the metal-binding sites and identify RNA metal-binding motifs

Table 4. *Some RNA internal loop motifs*

Motif name	Description	Consensus sequence 5′-3′	Reference
Bulged-G	Found in the sarcin–ricin loop and eukaryotic loop E of 5S ribosomal RNA, the loop is characterized by the backbone element 'S-turn'	5′-GA-AY-3′ 3′-AUGAY-5′	Correll *et al.* 2003; Szewczak & Moore, 1995; Wimberly *et al.* 1993
Bacterial loop E	A symmetric internal loop in bacterial 5S ribosomal RNA, this motif is the binding site of the L25 ribosomal protein, and is characterized by three cross-strand purine stacks and several non-canonical base pairings	5′-GAGAGUA-3′ 3′-AUGGUAG-5′	Correll *et al.* 1997
Kink turn/K-turn	Formed by two strands in a helix–internal loop–helix arrangement, this turn is named for the sharp bend, or kink, that is formed in the phosphodiester backbone of the strand, bringing the minor groove side of the two surrounding helices together	5′-GCRNNGANG-3′ 3′-CG—AGNC-5′	Klein *et al.* 2001; Lescoute *et al.* 2005; Vidovic *et al.* 2000
Reverse Kink turn	Like the kink turn, also a helix–internal loop–helix arrangement, but bending in the opposite direction, and thus toward the major grooves of the surrounding helices	5′-ACACAAACC-3′ 3′-UG—AGGG-5′	Strobel *et al.* 2004
Hook turn	A sharp bend in a strand that is helical, A form-like, on its 5′ side, with a $\sim 180^{\circ}$ turn in backbone direction on the 3′ side that occurs between two residues, usually a sheared A-G base pair		Szep *et al.* 2003
C-loop	This internal loop is made of two asymmetric strands and contains 2 (or more) base triples. The longer strand usually begins with the nucleotide C which forms a base triple with the flanking helix at its 3′ side. The nucleotide at the 3′ end of the longer strand forms a base triple with the flanking helix at the 5′ end. The base(s) on the shorter strand are often looped out	5′-C-CAC-U-3′ 3′-G–C–A-5′	Ban *et al.* 2000; Lescoute *et al.* 2005; Torres-Larios *et al.* 2002; Wimberly *et al.* 2000
Tetraloop receptor	This conserved motif is made up of two C-G pairs and an internal loop (including a G-U pair). The internal loop contains an adenosine platform and a looped-out U	5′-CC-UAAG-3′ 3′-GGUA–U-5′	Cate *et al.* 1996a

(Stefan *et al.* 2005). As of August 2005, 9764 metal sites have been identified in 256 PDB entries and eight RNA metal ion-binding motifs identified.

As with proteins, binding of metals or other ligands may either be to preformed sites or motifs, or through induction or selection of a specific binding pocket. Large structural rearrangements have been observed on metal binding (Wu & Tinoco, 1998; Penedo *et al.* 2004) indicating that secondary structure is not always fixed prior to formation of tertiary interactions.

Monovalent cations, particularly sodium and potassium have also been observed to bind specifically to adenosine platforms (Cate *et al.* 1996b; Klein *et al.* 2004) and G-quadruplexes (Pan *et al.* 2003a) among other elements. An especially striking example is a buried potassium ion surrounded by several phosphates that contributes a large binding free energy and allows a complex tertiary fold to be formed (Conn *et al.* 2002).

Metal binding to RNA can serve to stabilize a specific 3D structure (for example see Klein *et al.* 2004), but also may perform a catalytic role (Wedekind & McKay, 2003). In these cases, it is a combination of the 3D fold induced by metal-ion binding and the chemical nature of the metal ion itself that is responsible for catalysis.

3.5.2 Natural and selected aptamers

At least three classes of RNA molecules have been identified that are capable of tight and specific binding to small organic ligands: the *in vitro* selected (SELEX) aptamers, the riboswitches found in the untranslated regions of mRNA, and certain functional RNAs such as the self-splicing group I introns that bind to guanine as a co-factor. 3D structures exist for some members of each of these groups that illustrate the binding motifs (Patel & Suri, 2000; Adams *et al.* 2004; Batey *et al.* 2004; Guo *et al.* 2004; Serganov *et al.* 2004). Guanine binding by group I intron and riboswitch have been shown to be mediated by a base triple sandwich (stacking) motif (Guo *et al.* 2004; Serganov *et al.* 2004).

3.6 Tertiary interactions

As with secondary structure, the tertiary structure of RNA biomolecules is dominated by a limited number of recurring types of interactions or motifs (Batey *et al.* 1999). As enumerated in the SCOR database, these are: coaxial helices, kissing hairpin loops, the tetraloop–tetraloop receptor, the A-minor motif/patch, the tRNA D-loop:T-loop interaction, pseudoknots, and ribose zippers. Descriptions and references for each of these types of tertiary interaction are summarized in Table 5. As shown in Figs 5 and 6, coaxial helices and kissing hairpin loops both involve continuous base stacking, the first between helices either adjacent or connected by single base inserts or internal loops, and the second between two hairpin loops forming base pairs. Several of these modes of tertiary interaction such as the ribose zippers and the A-minor motif can be further subdivided into more specific sub-motifs (Nissen *et al.* 2001; Tamura & Holbrook, 2002), in these cases 11 and four subtypes respectively. Several ribose zippers identified in the bacterial ribonuclease P RNA structure (Kazantsev *et al.* 2005), including two subtypes are shown in Fig. 7. Finally, it is clear that conformational changes can be induced on formation of tertiary interactions. Figure 8 shows the dramatic change in conformation of an isolated tetraloop receptor as compared to a tetraloop–tetraloop receptor interaction.

Table 5. *Some RNA tertiary interaction motifs*

Motif name(s)	Description	Secondary structures	Sequence preference	Reference
Ribose Zipper	Formed by hydrogen bonding between consecutive backbone ribose 2′ hydroxyls from two distant regions of the chain, interacting in an anti-parallel manner. Classified as canonical and 6 other types	Double helix: Hairpin or internal loop	Antiparallel 5′-CC-3′(Stem) 3′-AA-5′(Loop) O2′–O2′ and base triples (e.g. A-minor)	Cate *et al.* 1996a; Tamura & Holbrook, 2002
A-Minor Motif/ A-patch	A clustering of A-minor interactions, often decreasing in type and order going from the 5′ to the 3′ direction	Internal loop, Hairpin loop	Adenosines	Nissen *et al.* 2001
D-Loop:T-Loop	Complex interaction between two conserved hairpins in tRNA, includes interdigitated bases	Hairpin loop: Hairpin loop	Conserved sequences in D-loop and T-loop motifs	Holbrook *et al.* 1978; Holbrook & Kim, 1979
Tetraloop:Tetraloop receptor	Conserved in Group I and II introns occuring between a GNRA tetraloop (GNRA fold) and the receptor; an internal loop plus two C-G pairs. It is characterized by a specific hydrogen bond pattern between the first A of the tetraloop and the U·A of the receptor to form an A·U·A triple; between the second A of the tetraloop and the backbone of the receptor C and U; between the third A of the tetraloop and the C:G pair of the receptor	Hairpin loop: Internal loop	5′-CC-UAAG-3′ 3′-GGUA–U-5′	Pley *et al.* 1994; Cate *et al.* 1996a; Butcher *et al.* 1997a
Kissing Hairpin Loop	The kissing hairpin complex is formed by base pairing between single-stranded residues of two hairpin loops with complementary sequences	Hairpin loop: Hairpin loop	Self-complementary often six nucleotides	Chang & Tinoco, 1994; Ennifar *et al.* 2001
Coaxial Helices Interhelical stacking	Nucleotide bases from two separate helices stack and align axes to form a pseudo-continuous, coaxial helix. Coaxial helices are highly stabilizing and are dominant in several large RNA structures. Interhelical stacking may occur via a single base or base pair bridge between helices, resulting in continuous helical stacking spanning multiple helices	Double helices often across an internal or junction loop	'Bridging' nucleotide between helices has a preference for adenine	Kim *et al.* 1974; Cate *et al.* 1996a
Pseudoknot	When bases pair between nucleotide loops (hairpin or internal) and bases outside the enclosing loop, they form a pseudoknot. This structure often contains coaxial helices. Can be very stable	Hairpin loop: Single strand	Complementary	Shen & Tinoco, 1995; van Batenburg *et al.* 2001

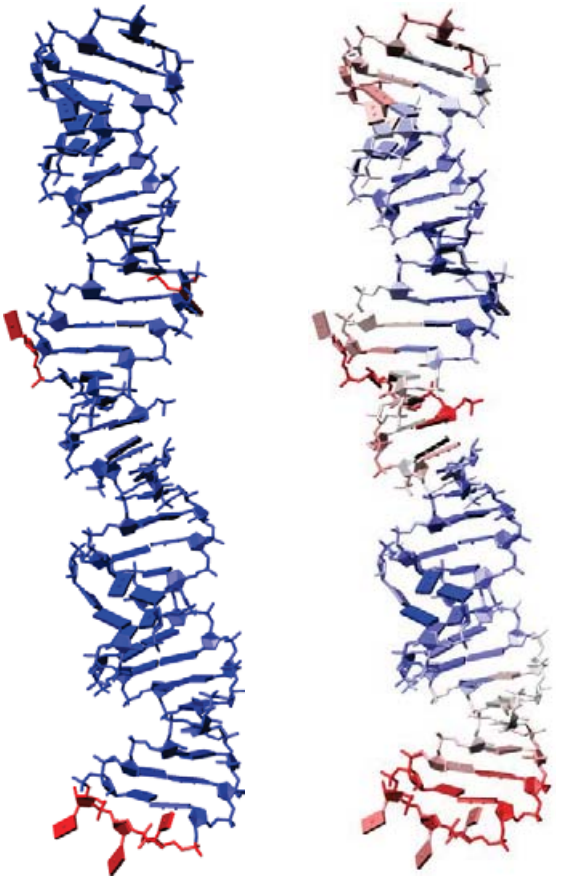

Fig. 5. Coaxial interhelical stacking in the P4–P6 domain of the group I intron. (*a*). Coaxial stack between helices bulges and internal loops with non-Watson–Crick pairing as seen in the P4–P6 domain of the group I intron (PDB ID 1GID). Non-stacked residues are shown in red. (*b*). Crystallographic mobility parameters (B-factors) as observed in 1GID. The correspondence between stacking and reduced mobility is apparent.

Clearly, other motifs remain to be discovered and described, such as novel hairpin loop–loop interactions and base triple and quadruple interactions, but it is becoming more apparent with each new structure that most major classes of RNA tertiary interaction motifs have been identified. A major challenge and opportunity to be undertaken is the prediction of RNA tertiary interactions from sequence alone. Detailed analysis of the known tertiary interaction motifs, together with an accurate secondary structure prediction and comparative sequence analysis should provide strong indications of the presence of specific tertiary interactions in RNA sequences.

4. Future directions

A comprehensive identification, classification and characterization of RNA structural elements, secondary structure motifs, tertiary interaction motifs, and binding motifs is critical for

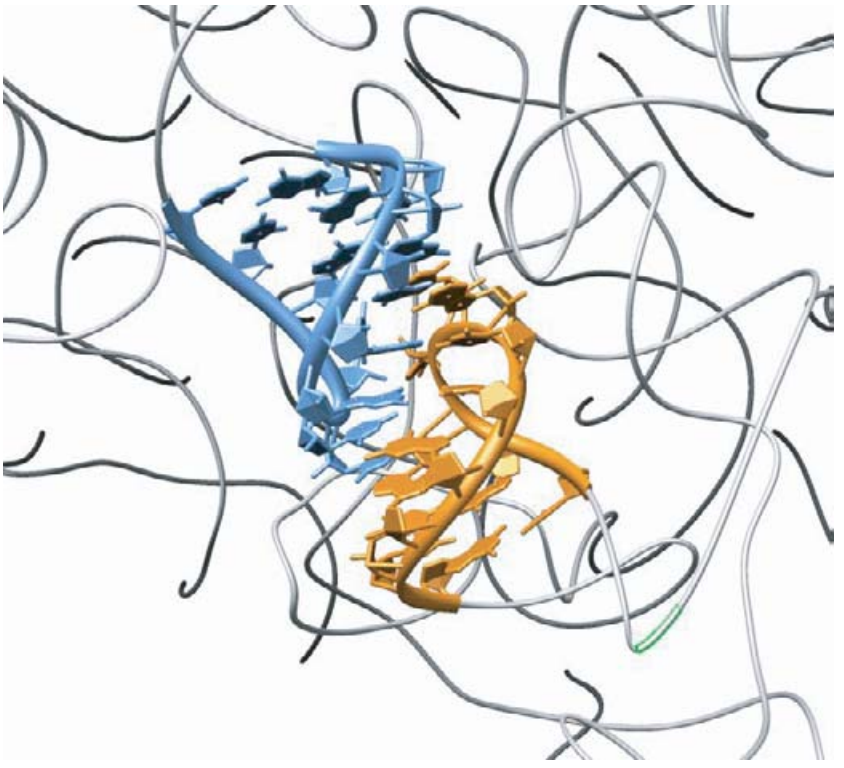

Fig. 6. A kissing hairpin loop found in ribosomal RNA. Two hairpin loops forming a 'kissing' interaction. A coaxial stack is formed by base pairing between the hairpins and their Watson–Crick stems (note the 'minor groove' formed by the backbones of the hairpins). The blue hairpin is formed by residues 414–426 and the orange hairpin by residues 2440–2449 of 23S ribosomal RNA (1JJ2). The surrounding backbone of the ribosomal RNA is shown in gray.

understanding RNA structure–function relationships, folding, evolution, engineering and design. Advances in RNA structure prediction and the identification of RNA genes and genetic control elements in genomic DNA are dependent on our understanding of the modular nature of RNA and the subunits from which it is composed.

In order to obtain such an understanding, not only do we require additional examples of a wide variety of RNA structures, but also development of new and improved computational approaches for the identification of structural motifs and the elements from which they are composed, for classification leading to analysis of sequence and structural variation of these motifs, and for prediction of these motifs from sequences alone. In addition, evolutionary analysis of RNA elements and motifs, and studies relating motif structure to function are needed to place these structural features into a biological context.

In the last few years a strong trend has been established away from structural analysis of model compounds and molecular fragments, toward the structure determination of large, intact, biological RNA molecules (Holbrook, 2005). Analysis of these large RNAs shows a combination of well-known structural motifs and the presence of potential novel motifs (Holbrook, 2005). A future challenge is the confirmation and characterization of these new motifs and forming an understanding of how these motifs are linked and interact to form functional biological

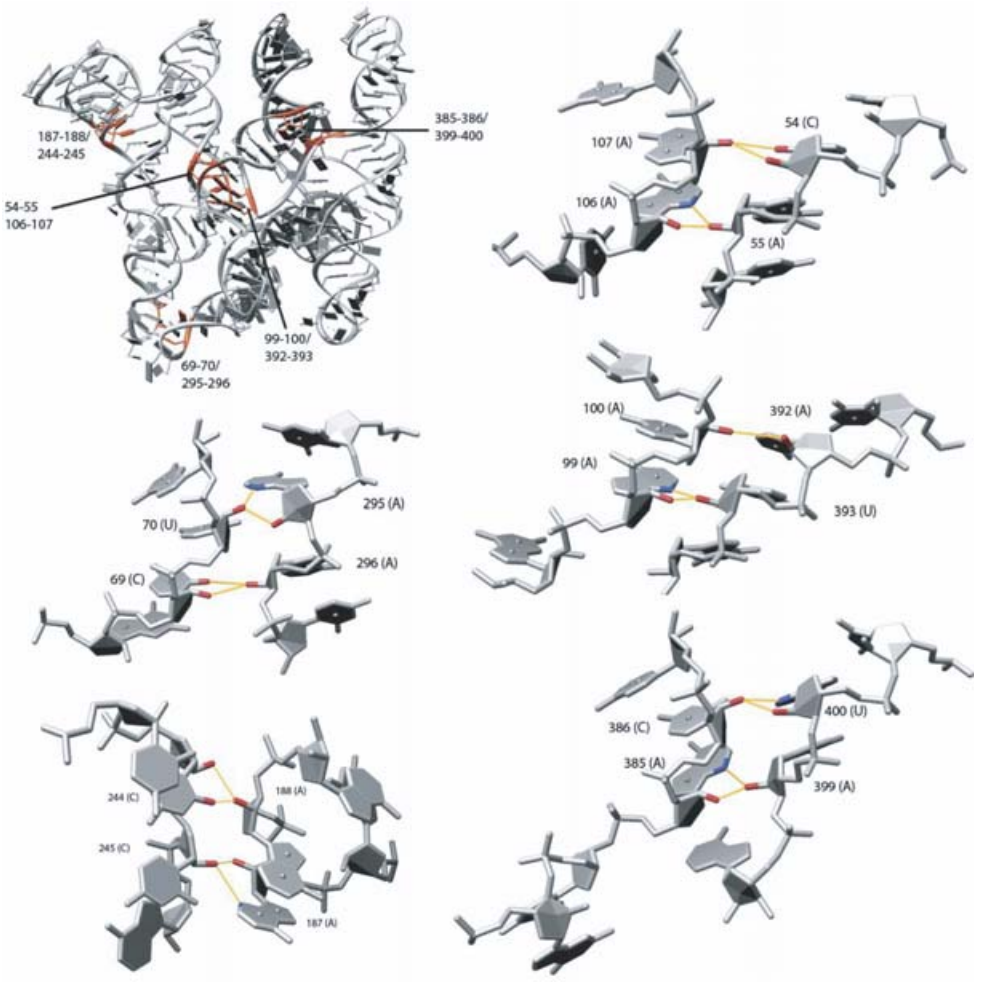

Fig. 7. Ribose zipper motifs observed in Ribonuclease P RNA (PDB identifier 2a64). Locations of ribose zipper tertiary interactions in the crystal structure of bacterial Ribonuclease P RNA are shown in the upper left. The following figures show close-up views of the ribose zippers with O2′-O2′ and O2′-N3 (base) hydrogen bonds indicated. All ribose zippers are of the canonical type except for 99–100;392–393 which is classified as a single ribose zipper due to the single ribose-base hydrogen bond.

RNAs. The existing approaches and tools for identification and analysis of RNA motifs have been described above, as have recent advances in the standardization of nomenclature and definitions. A working group, the RNA Ontology Consortium, has been formally assembled and assigned the task of establishing and improving standards in this area (see Leontis et al. 2006).

Several areas of potentially great significance are currently emerging from the study of RNA structure and RNA structural motifs. These include RNA engineering and design as an approach to nanotechnology (Chworos *et al.* 2004; Guo, 2005); the association of RNA structure with human disease (Michlewski & Krzyzosiak, 2004; Barciszewska *et al.* 2005; Darnell *et al.* 2005) and its treatment through drug design (Tu *et al.* 2005), siRNA, antisense, and ribozyme technology (Scanlon, 2004); and finally the evolution of RNA and RNA motifs, their mutation patterns, and

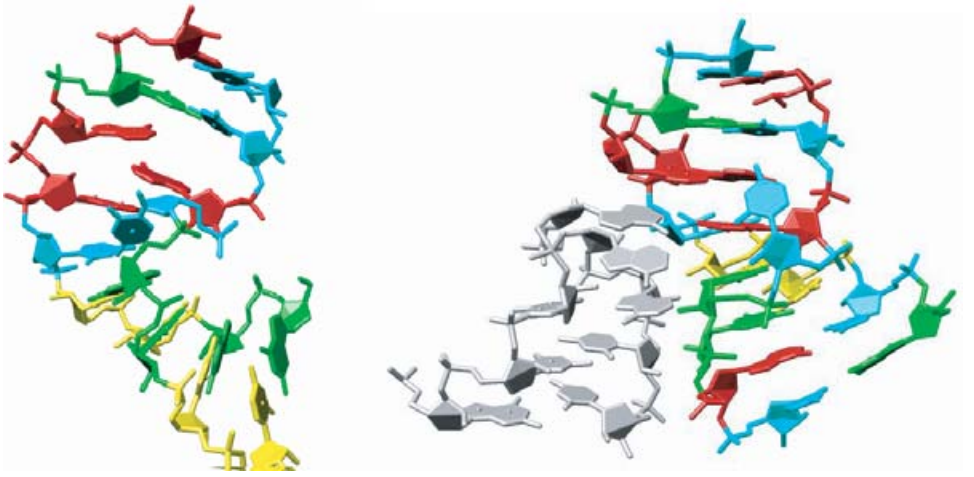

Fig. 8. Conformational change in the tetraloop receptor on binding a GNRA tetraloop. Crystal structures of a 'free' tetraloop receptor motif (1TLR) and the tertiary interaction formed between a GNRA tetraloop and receptor in the P4–P6 domain of the group I intron (1GID). Shown are residues 1–9 and 16–23 of the free tetraloop and 220–228, 148–155 and 246–253 of the complex. The tetraloop receptor sequence is variable and different sequences are shown in the figure: (left GGCCUAAGA/UUAUGGCC, right GUCCUAAGU/AUAUGGAU). Bases are colored as red (A), cyan (U), green (G) and yellow (C) for the tetraloop receptor, while the tetraloop is in gray. Note the conformational change occurring in the receptor on tetraloop binding.

the conservation and identification of functional non-coding RNAs (ncRNAs) in genomic sequences.

5. Acknowledgments

The authors acknowledge Dr Elizabeth Holbrook, Dr Peter Klosterman (U. C. Berkeley), and Dr Liliana Stefan (Lawrence Berkeley Laboratory) for assistance in preparation of this manuscript. S.R.H. was supported by National Institutes of Health grants 1R01GM66199 and 1R01HG002665. S.E.B. was supported by NHGRI of the NIH grant K22 HG00056 and IR01GM66199. D.K.H. was supported by National Institutes of Health grant 1R01GM66199.

6. References

Adams, P. L., Stahley, M. R., Gill, M. L., Kosek, A. B., Wang, J. & Strobel, S. A. (2004). Crystal structure of a group I intron splicing intermediate. *RNA* **10**, 1867–1887.

Antao, V. P., Lai, S. Y. & Tinoco, I. J. (1991). A thermodynamic study of unusually stable RNA and DNA hairpins. *Nucleic Acids Research* **19**, 5901–5905.

Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2·4 Å resolution. *Science* **289**, 905–920.

Barciszewska, M. Z., Szymanski, M., Wyszko, E., Pas, J., Rychlewski, L. & Barciszewska, J. (2005). Lead toxicity through the leadzyme. *Mutation Research* **589**, 103–110.

Batey, R. T., Gilbert, S. D. & Montange, R. K. (2004). Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature* **432**, 411–415.

Batey, R. T., Rambo, R. P. & Doudna, J. A. (1999). Tertiary motifs in RNA structure and folding. *Angewandte Chemie International Edition* **38**, 2326–2343.

Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A. R. & Schneider, B. (1992). The nucleic acid database: a comprehensive relational database of three-dimensional structures of nucleic acids. *Biophysical Journal* **63**, 751–759.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research* **28**, 235–242.

Brown, J. W. (1999). The ribonuclease P database. *Nucleic Acids Research* **27**, 314.

Butcher, S. E., Dieckmann, T. & Feigon, J. (1997a). Solution structure of a GAAA tetraloop receptor RNA. *EMBO Journal* **16**, 7490–7499.

Butcher, S. E., Dieckmann, T. & Feigon, J. (1997b). Solution structure of the conserved 16 S-like ribosomal RNA UGAA tetraloop. *Journal of Molecular Biology* **268**, 348–358.

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

Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Szewczak, A. A., Kundrot, C. E., Cech, T. R. & Doudna, J. A. (1996b). RNA tertiary structure mediation by adenosine platforms. *Science* **273**, 1696–1699.

Chang, K.-Y. & Tinoco Jr., I. (1994). Characterization of a 'kissing' hairpin complex derived from the human immunodeficiency virus genome. *Proceedings of the National Academy of Sciences USA* **91**, 8705–8709.

Cheong, C., Varani, G. & Tinoco Jr., I. (1990). Solution structure of an unusually stable RNA hairpin, 5′GGAC(UUCG)GUCC. Nature **346**, 680–682.

Chworos, A., Severcan, I., Koyfman, A. Y., Weinkam, P., Oroudjev, E., Hansma, H. G. & Jaeger, L. (2004). Building programmable jigsaw puzzles with RNA. *Science* **306**, 2068–2072.

Conn, G. L., Gittis, A. G., Lattman, E. E., Misra, V. K. & Draper, D. E. (2002). A compact RNA tertiary structure contains a buried backbone-K^+ complex. *Journal of Molecular Biology* **318**, 963–973.

Convery, M. A., Rowsell, S., Stonehouse, N. J., Ellington, A. D., Hirao, I., Murray, J. B., Peabody, D. S., Phillips, S. E. & Stockley, P. G. (1998). Crystal structure of an RNA aptamer-protein complex at 2·8 Å resolution. *Nature Structural Biology* **5**, 133–139.

Correll, C. C., Beneken, J., Plantinga, M. J., Lubbers, M. & 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 Research* **31**, 6806–6818.

Correll, C. C., Freeborn, B., Moore, P. B. & Steitz, T. A. (1997). Metals, motifs, and recognition in the crystal structure of a 5S rRNA domain. *Cell* **91**, 705–712.

Correll, C. C. & 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.

Correll, C. C., Wool, I. G. & Munishkin, A. (1999). The two faces of the *Escherichia coli* 23S rRNA sarcin/ricin domain: the structure at 1·11 Å resolution. *Journal of Molecular Biology* **292**, 275–287.

Darnell, J. C., Fraser, C. E., Mostovetsky, O., Stefani, G., Jones, T. A., Eddy, S. R. & Darnell, R. B. (2005). Kissing complex RNAs mediate interaction between the Fragile-X mental retardation protein KH2 domain and brain polyribosomes. *Genes and Development* **19**, 903–918.

Draper, D. E. (2004). A guide to ions and RNA structure. *RNA* **10**, 335–343.

Duarte, C. M. & Pyle, A. M. (1998). Stepping through an RNA structure: a novel approach to conformational analysis. *Journal of Molecular Biology* **284**, 1465–1478.

Duarte, C. M., Wadley, L. M. & Pyle, A. M. (2003). RNA structure comparison, motif search and discovery using a reduced representation of RNA conformational space. *Nucleic Acids Research* **31**, 4755–4761.

Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J. P. & Ehresmann, B. (1987). Probing the structure of RNAs in solution. *Nucleic Acids Research* **15**, 9109–9128.

Ennifar, E., Walter, P., Ehresmann, B., Ehresmann, C. & Dumas, P. (2001). Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nature Structural Biology* **12**, 1064–1068.

Fera, D., Kim, N., Shiffeldrim, N., Zorn, J., Laserson, U., Gan, H. H. & Schlick, T. (2004). RAG: RNA-As-Graphs web resource. *BMC Bioinformatics* **5**, 88.

Furtig, B., Richter, C., Wohnert, J. & Schwalbe, H. (2003). NMR spectroscopy of RNA. *ChemBioChem* **4**, 936–962.

Gast, F.-U. (2003). A new structural motif in the left terminal domain of large viroids identified by covariation analysis. *Virus Genes* **26**, 19–23.

Gendron, P., Lemieux, S. & Major, F. (2001). Quantitative analysis of nucleic acid three-dimensional structures. *Journal of Molecular Biology* **308**, 919–936.

Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A. & Eddy, S. R. (2003). Rfam: an RNA family database. *Nucleic Acids Research* **31**, 439–441.

Guo, F., Gooding, A. R. & Cech, T. R. (2004). Structure of the Tetrahymena ribozyme: base triple sandwich and metal ion at the active site. *Molecular Cell* **16** 351–362.

Guo, P. (2005). Bacterial virus phi29 DNA-packaging motor and its potential applications in gene therapy and nanotechnology. *Methods in Molecular Biology* **300**, 285–324.

Gutell, R. R. (1995). Comparative sequence analysis and the structure of 16S and 23S rRNA. In *Ribosomal RNA: Structure, Evolution, Processing, and Function in Protein Biosynthesis* (eds. R. A. Zimmerman & A. E. Dahlberg), pp. 109–126. Boca Raton, Florida: CRC Press.

Gutell, R. R., Lee, J. C. & Cannone, J. J. (2002). The accuracy of ribosomal RNA comparative structural models. *Current Opinion in Structural Biology* **12**, 301–310.

Harrison, A. M., South, D. R., Willett, P. & Artymiuk, P. J. (2003). Representation, searching and discovery of patterns of bases in complex RNA structures. *Journal of Computer-Aided Molecular Design* **17**, 537–549.

Hershkovitz, E., Tannenbaum, E., Howerton, S. B., Sheth, A., Tannenbaum, A. & Williams, L. D. (2003). Automated identification of RNA conformational motifs: theory and application to the HM LSU 23S rRNA. *Nucleic Acids Research* **31**, 6249–6257.

Heus, H. A. & Pardi, A. (1991). Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science* **253**, 191–194.

Hofacker, I. L. (2003). Vienna RNA secondary structure server. *Nucleic Acids Research* **31**, 3429–3431.

Holbrook, S. R. (2005). RNA structure: the long and short of it. *Current Opinion in Structural Biology* **15**, 302–308.

Holbrook, S. R. & Kim, S.-H. (1997). RNA crystallography. *Biopolymers* **44**, 3–21.

Holbrook, S. R., Kim, S.-H. (1979). Intercalation conformations in single- and double-stranded nucleic acids. *International Journal of Biological Macromolecules* **1**, 233–240.

Holbrook, S. R., Sussman, J. L. & Kim, S. H. (1981). Absence of correlation between base-pair sequence and RNA conformation. *Science* **212**, 1275–1277.

Holbrook, S. R., Sussman, J. L., Warrant, R. W., Church, G. M. & Kim, S.-H. (1977). RNA-ligand interactions. (I) Magnesium binding sites in yeast tRNA(Phe). *Nucleic Acids Research* **4**, 2811–2820.

Holbrook, S. R., Sussman, J. L., Warrant, R. W. & Kim, S. H. (1978). Crystal structure of yeast phenylalanine transfer RNA. II. Structural features and functional implications. *Journal of Molecular Biology* **123**, 631–660.

Huang, H.-C., Nagaswamy, U. & Fox, G. E. (2005). The application of cluster analysis in the intercomparison of loop structures in RNA. *RNA* **11**, 412–423.

Hung, L. W., Holbrook, E. L. & Holbrook, S. R. (2000). The crystal structure of the Rev binding element of HIV-1 reveals novel base pairing and conformational variability. *Proceedings of the National Academy of Sciences USA* **97**, 5107–5112.

Ihle, Y., Ohlenschlager, O., Hafner, S., Duchardt, E., Zacharias, M., Seitz, S., Zell, R., Ramachandran, R. & Gorlach, M. (2005). A novel cGUUAg tetraloop structure with a conserved yYNMGg-type backbone conformation from cloverleaf 1 of bovine enterovirus 1 RNA. *Nucleic Acids Research* **33**, 2003–2011.

Jucker, F. M. & Pardi, A. (1995a). GNRA tetraloops make a U-turn. *RNA* **1**, 219–222.

Jucker, F. M. & Pardi, A. (1995b). Solution structure of the CUUG hairpin loop: a novel RNA tetraloop motif. *Biochemistry* **34**, 14416–14427.

Kazantsev, A. V., Krivenko, A. A., Harrington, D. J., Holbrook, S. R., Adams, P. D. & Pace, N. R. (2005). Crystal structure of a bacterial ribonuclease P RNA. *Proceedings of the National Academy of Sciences USA* **102**, 13392–13397.

Kim, S.-H. & Sussman, J. L. (1976). Pi turn is a conformational pattern in RNA loops and bends. *Nature* **260**, 645–646.

Kim, S. H., Sussman, J. L., Suddath, F. L., Quigley, G. J., McPherson, A., Wang, A. H., Seeman, N. C. & Rich, A. (1974). The general structure of transfer RNA molecules. *Proceedings of the National Academy of Sciences USA* **71**, 4970–4974.

Klein, D. J., Moore, P. B. & Steitz, T. A. (2004). The contribution of metal ions to the structural stability of the large ribosomal subunit. *RNA* **10**, 1366–1379.

Klein, D. J., Schmeing, T. M., Moore, P. B. & Steitz, T. A. (2001). The kink-turn: a new RNA secondary structure motif. *EMBO Journal* **20**, 4214–4221.

Klosterman, P. S., Hendrix, D. K., Tamura, M., Holbrook, S. R. & Brenner, S. E. (2004). Three-dimensional motifs from the SCOR: structural classification of RNA database – extruded strands, base triples, tetraloops, and u-turns. *Nucleic Acids Research* **32**, 2342–2352.

Klosterman, P. S., Tamura, M., Holbrook, S. R. & Brenner, S. E. (2002). SCOR: a Structural Classification of RNA database. *Nucleic Acids Research* **30**, 392–394.

Kolk, M. H., Heus, H. A. & Hilbers, C. W. (1997). The structure of the isolated central hairpin of the HDV antigenomic ribozyme: novel structural features and similarity of the loop in the ribozyme and free in solution. *EMBO Journal* **16**, 3685–3692.

Krasilnikov, A. S. & Mondragon, A. (2003). On the occurrence of the T-loop RNA folding motif in large RNA molecules. *RNA* **9**, 640–643.

Lee, J. C., Cannone, J. J. & Gutell, R. R. (2003). The lonepair triloop: a new motif in RNA structure. *Journal of Molecular Biology* **325**, 65–83.

Lee, J. C. & Gutell, R. R. (2004). Diversity of base-pair conformations and their occurence in rRNA structure and RNA structural motifs. *Journal of Molecular Biology* **344**, 1225–1249.

Leontis, N. B., Altman, R. B., Berman, H. M., Brenner, S. E., Brown, J. W., Engelke, D. R., Harvey, S. C., Holbrook, S. R., Jossinet, F., Lewis, S. E., Major, F., Matthews, D. H., Richardson, J. S., Williamson, J. R. & Westhof, E. (2006). The RNA Ontology Consortium: an open invitation to the RNA community. *RNA* **12**, 533–541.

Leontis, N. B., Stombaugh, J. & Westhof, E. (2002). Motif prediction in ribosomal RNAs. Lessons and prospects for automated motif prediction in homologous RNA molecules. *Biochemie* **84**, 961–973.

Leontis, N. B. & Westhof, E. (2001). Geometric nomenclature and classification of RNA base pairs. *RNA* **7**, 499–512.

Leontis, N. B. & Westhof, E. (2002). The annotation of RNA motifs. *Comparative and Functional Genomics* **3**, 518–524.

Leontis, N. B. & Westhof, E. (2003). Analysis of RNA motifs. *Current Opinion in Structural Biology* **13**, 300–308.

Lescoute, A., Leontis, N. B., Massire, C. & Westhof, E. (2005). Recurrent structural motifs, isostericity matrices and sequence alignments. *Nucleic Acids Research* **33**, 2395–2409.

Lescoute, A. & Westhof, E. (2005). Riboswitch structures: purine ligands replace tertiary contacts. *Chemistry and Biology* **12**, 10–13.

Lesnik, E. A., Sampath, R. & Ecker, D. J. (2002). Rev response elements (RRE) in lentiviruses: an RNAMotif algorithm-based strategy for RRE prediction. *Medicinal Research Reviews* **22**, 617–636.

Lilley, D. M. (2000). Structures of helical junctions in nucleic acids. *Quarterly Reviews of Biophysics* **33**, 109–159.

Lilley, D. M. J. (1998). Folding of branched RNA species. *Biopolymers*, **48**, 101–112.

Liu, H., Matsugami, A., Katahira, M. & Uesugi, S. (2002). A dimeric RNA quadruplex architecture comprised of two G:G(A):G:G(A) hexads, G:G:G:G tetrads and UUUU loops. *Journal of Molecular Biology* **322**, 955–970.

Mathews, D. H., Andre, T. C., Kim, J., Turner, D. H. & Zuker, M. (1998). An updated recursive algorithm for RNA secondary structure prediction with improved free energy parameters. *American Chemical Society Symposium Series* **682**, 246–257.

Michlewski, G. & Krzyzosiak, W. J. (2004). Molecular architecture of CAG repeats in human disease related transcripts. *Journal of Molecular Biology* **340**, 665–679.

Moore, P. B. (1999). Structural motifs in RNA. *Annual Reviews of Biochemistry* **68**, 287–300.

Murray, L. J., Arendall, W. B. R., Richardson, D. C. & Richardson, J. S. (2003). RNA backbone is rotameric. *Proceedings of the National Academy of Sciences USA* **100**, 13904–13909.

Nagaswamy, U. & Fox, G. E. (2002). Frequent occurrence of the T-loop RNA folding motif in ribosomal RNAs. *RNA* **8**, 1112–1119.

Nagaswamy, U., Larios-Sanz, M., Hury, J., Collins, S., Zhang, Z., Zhao, Q. & Fox, G. E. (2002). NCIR: a database of non-canonical interactions in known RNA structures. *Nucleic Acids Research* **30**, 395–397.

Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B. & Steitz, T. A. (2001). RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proceedings of the National Academy of Sciences USA* **98**, 4899–4903.

Pan, B., Xiong, Y., Shi, K., Deng, J. & Sundaralingam, M. (2003a). Crystal structure of an RNA purine-rich tetraplex containing adenine tetrads. Implications for specific binding in RNA tetraplexes. *Structure* **11**, 815–823.

Pan, B., Xiong, Y., Shi, K. & Sundaralingam, M. (2003b). Crystal structure of a bulged RNA tetraplex at 1·1 Å resolution: implications for a novel binding site in RNA tetraplex. *Structure* **11**, 1423–1430.

Pantopoulos, K. (2004). Iron metabolism and the IRE/IRP regulatory system: an update. *Annals of the New York Academy of Sciences* **1012**, 1–13.

Patel, D. J. & Suri, A. K. (2000). Structure, recognition and discrimination in RNA aptamer complexes with cofactors, amino acids, drugs and aminoglycoside antibiotics. *Journal of Biotechnology* **74**, 39–60.

Penedo, J. C., Wilson, T. J., Jayasena, S. D., Khvorova, A. & Lilley, D. M. (2004). Folding of the natural hammerhead ribozyme is enhanced by interaction of auxiliary elements. *RNA* **10**, 880–888.

Pley, H. W., Flaherty, K. M. & McKay, D. B. (1994). Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature* **372**, 111–113.

Popenda, M., Milecki, J. & Adamiak, R. W. (2004). High salt solution structure of a left-handed RNA double helix. *Nucleic Acids Research* **32**, 4044–4054.

Quigley, G. J. & Rich, A. (1976). Structural domains of transfer RNA molecules. *Science* **194**, 796–806.

Quigley, G. J., Teeter, M. M. & Rich, A. (1978). Structural analysis of spermine and magnesium ion binding to yeast phenylalanine transfer RNA. *Proceedings of the National Academy of Sciences USA* **75**, 64–68.

Richter, S., Cao, H. & Rana, T. M. (2002). Specific HIV-1 TAR RNA loop sequence and functional groups are required for human cyclin T1-Tat-TAR ternary complex formation. *Biochemistry* **41**, 6391–6397.

Rivas, E. & Eddy, S. R. (1999). A dynamic programming algortihm for RNA structure prediction including pseudoknots. *Journal of Molecular Biology* **285**, 2053–2068.

Rosenblad, M. A., Gorodkin, J., Knudsen, B., Zweib, C. & Samuelsson, T. (2003). SRPDB: signal recognition particle database. *Nucleic Acids Research* **31**, 363–364.

Rowsell, S., Stonehouse, N. J., Convery, M. A., Adams, C. J., Ellington, A. D., Hirao, I., Peabody, D. S., Stockley, P. G. & Phillips, S. E. (1998). Crystal structures of a series of RNA aptamers complexed to the same protein target. *Nature Structural Biology* **5**, 970–975.

Scanlon, K. J. (2004). Anti-genes: siRNA, ribozymes and antisense. *Current Pharmaceutical Biotechnology* **5**, 415–420.

Schneider, B., Moravek, Z. & Berman, H. M. (2004). RNA conformational classes. *Nucleic Acids Research* **32**, 1666–1677.

Serganov, A., Yuan, Y. R., Pikovskaya, O., Polonskaia, A., Malinina, L., Phan, A. T., Hobartner, C., Micura, R., Breaker, R. R. & Patel, D. J. (2004). Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chemistry and Biology* **11**, 1729–1741.

Shen, L. X. & Tinoco Jr., I. (1995). The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus. *Journal of Molecular Biology* **247**, 963–978.

Shi, H. & Moore, P. B. (2000). The crystal structure of yeast phenylalanine tRNA at 1·93 Å resolution: a classic structure revisited. *RNA* **6**, 1091–1105.

Sobczak, K. & Krzyzosiak, W. G. (2002). Structural determinants of BRCA1 translational regulation. *Journal of Biological Chemistry* **277**, 17349–17358.

Stefan, L. R., Zhang, R., Levitan, A. G., Hendrix, D. K., Brenner, S. E. & Holbrook, S. R. (2005). MeRNA: a database of metal ion binding sites in RNA structures. *Nucleic Acids Research* **34**, D131–D134.

Strobel, S. A., Adams, P. L., Stahley, M. R. & Wang, J. (2004). RNA kink turns to the left and to the right. *RNA* **10**, 1852–1854.

Szep, S., Wang, J. & Moore, P. B. (2003). The crystal structure of a 26-nucleotide RNA containing a hook-turn. *RNA* **9**, 44–51.

Szewczak, A. A. & Moore, P. B. (1995). The sarcin/ricin loop, a modular RNA. *Journal of Molecular Biology* **247**, 81–98.

Szewczak, A. A., Moore, P. B., Chan, Y.-L. & Wool, I. G. (1993). The conformation of the sarcin/ricin loop from 28S ribosomal RNA. *Proceedings of the National Academy of Sciences USA* **90**, 9581–9585.

Tamura, M., Hendrix, D. K., Klosterman, P. S., Schimmelman, N. R., Brenner, S. E. & Holbrook, S. R. (2004). SCOR: Structural Classification of RNA, version 2.0. *Nucleic Acids Research* **32**, D182–D184.

Tamura, M. & Holbrook, S. R. (2002). Sequence and structural conservation in RNA ribose zippers. *Journal of Molecular Biology* **320**, 455–474.

Theil, E. C. (2000). Targeting mRNA to regulate iron and oxygen metabolism. *Biochemical Pharmacology* **59**, 87–93.

Torres-Larios, A., Dock-Bregeon, A. C., Romby, P., Rees, B., Sankaranarayanan, R., Caillet, J., Springer, M., Ehresmann, C., Ehresmann, B. & Moras, D. (2002). Structural basis of translational control by *Escherichia coli* threonyl tRNA synthetase. *Nature Structural Biology* **9**, 343–347.

Tu, D., Blaha, G., Moore, P. B. & Steitz, T. A. (2005). Structures of MLSBK antibiotics bound to mutated large ribosomal subunits provide a structural explanation for resistance. *Cell* **121**, 257–270.

Vallazza, M., Perbandt, M., Klussmann, S., Rypniewski, W., Einspahr, H. M., Erdmann, V. A. & Betzel, C. (2004). First look at RNA in L-configuration. *Acta Crystallographica. Section D: Biological Crystallography* **60**, 1–7.

Vallurupalli, P. & Moore, P. B. (2003). The solution structure of the loop E region of the 5S rRNA from spinach chloroplasts. *Journal of Molecular Biology* **325**, 843–856.

van Batenburg, F. H., Gultyaev, A. P. & Pleij, C. W. (2001). Pseudobase: structural information on RNA pseudoknots. *Nucleic Acids Research* **29**, 194–195.

Vidovic, I., Nottrott, S., Hartmuth, K., Luhrmann, R. & Ficner, R. (2000). Crystal structure of the spliceosomal 15·5 kD protein bound to a U4 snRNA fragment. *Molecular Cell* **6**, 1331–1342.

Wadley, L. M. & Pyle, A. M. (2004). The identification of novel RNA structural motifs using COMPADRES: an automated approach to structural discovery. *Nucleic Acids Research* **32**, 6650–6659.

Wedekind, J. E. & Mckay, D. B. (2003). Crystal structure of the leadzyme at 1·8 Å resolution: metal ion binding and the implications for catalytic mechanism and allo site ion regulation. *Biochemistry* **42**, 9554–9563.

Wimberly, B., Varani, G. & Tinoco, I. J. (1993). The conformation of loop E of eukaryotic 5S ribosomal RNA. *Biochemistry* **32**, 1078–1087.

Wimberly, B. T., Brodersen, D. E., Clemons, W. M. J., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T. & Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. *Nature* **407**, 327–339.

Woese, C. R., Winker, S. & Gutell, R. R. (1990). Architecture of ribosomal RNA: constraints on the sequence of 'tetra-loops'. *Proceedings of the National Academy of Sciences USA* **87**, 8467–8471.

Wong, K., Barbin, Y., Chakrabarti, S. & Adams, P. (2005). A point mutation in the iron-responsive element of the L-ferritin in a family with hereditary hyperferritinemia cataract syndrome. *Canadian Journal of Gastroenterology* **19**, 253–255.

Wu, M. & Tinoco, I. J. (1998). RNA folding causes secondary structure rearrangement. *Proceedings of the National Academy of Sciences USA* **95**, 11555–11560.

Wuyts, J., Perriere, G. & Van De Peer, Y. (2004). The European ribosomal RNA database. *Nucleic Acids Research* **32**, D101–103.

Yang, H., Jossinet, F., Leontis, N., Chen, L., Westbrook, J., Berman, H. M. & Westhof, E. (2003). Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Research* **31**, 3450–3460.

Zuker, M. (1989). Computer prediction of RNA structure. *Methods in Enzymology* **180**, 262–288.

Zweib, C., Gorodkin, J., Knudsen, B., Burks, J. & Wower, J. (2003). tmRDB (tmRNA database). *Nucleic Acids Research* **31**, 446–447.