Analysis of RNA motifs

Neocles B Leontis* and Eric Westhof†

RNA motifs are directed and ordered stacked arrays of non-Watson–Crick base pairs forming distinctive foldings of the phosphodiester backbones of the interacting RNA strands. They correspond to the 'loops' — hairpin, internal and junction — that intersperse the Watson–Crick two-dimensional helices as seen in two-dimensional representations of RNA structure. RNA motifs mediate the specific interactions that induce the compact folding of complex RNAs. RNA motifs also constitute specific protein or ligand binding sites. A given motif is characterized by all the sequences that fold into essentially identical three-dimensional structures with the same ordered array of isosteric non-Watson–Crick base pairs. It is therefore crucial, when analyzing a three-dimensional RNA structure in order to identify and compare motifs, to first classify its non-Watson–Crick base pairs geometrically.

Addresses
*Department of Chemistry and Center for Biomolecular Sciences, Bowling Green State University, Bowling Green, Ohio 43403, USA
e-mail: leontis@bgnet.bgsu.edu
†Institut de Biologie Moléculaire et Cellulaire du CNRS, Modélisation et Simulations des Acides Nucléiques, UPR 9002, 15 rue René Descartes, F-67084 Strasbourg Cedex, France
e-mail: e.westhof@ibmc.u-strasbg.fr

Current Opinion in Structural Biology 2003, **13**:300–308

This review comes from a themed issue on
Nucleic acids
Edited by Carl C Correll and David MJ Lilley

0959-440X/03/$ – see front matter

DOI 10.1016/S0959-440X(03)00076-9

Abbreviations

H	Hoogsteen
NOE	nuclear Overhauser effect
SE	sugar-edge
WC	Watson–Crick

Introduction

The architecture of folded RNA molecules is shaped by specific tertiary interactions that involve edge-to-edge base pairing as well as face-to-face stacking interactions. Tertiary interactions generally are mediated by 'RNA motifs', which can be defined as directed and ordered arrays of non-WC base pairs (where WC stands for Watson–Crick) forming distinctive foldings of the phosphodiester backbones of the interacting RNA strands. To identify, classify and analyze RNA motifs, they are first deconstructed into their most elemental components — the individual non-WC base pairs. These, when combined, can form base triples, quadruples and stacks comprising hairpin loops, symmetric and asymmetric internal loops, and junction loops. Additional contacts (hydrogen bonds between sugar–phosphate backbones, between backbone and bases, or forming tertiary base pairs, which may be either WC or non-WC) are formed when motifs interact with each other in space. A given RNA motif comprises all the sequences that form the same directed and ordered array of isosteric non-WC base pairs and fold into similar, if not essentially identical, three-dimensional structures. Therefore, when analyzing new three-dimensional RNA structures to identify and compare motifs, it is crucial to classify their non-WC base pairs geometrically. Edge-to-edge base pairs (as distinct from bifurcated pairs) fall naturally into one of 12 geometric families, defined by the base edges that interact — WC, Hoogsteen (H) or sugar-edge (SE) — and the relative orientation (*cis* or *trans*) of the base–base interaction [1•,2••]. Isosteric base pairs necessarily belong to the same geometric family and can substitute for each other without distorting fundamentally the three-dimensional structure of a motif. All base pairs belonging to the same geometric family display very similar relative orientations of the glycosidic bonds, but may belong to distinct isosteric subfamilies. Base pairs belonging to the same family but to different isosteric subfamilies differ with regard to the C1′–C1′distance. At the level of the individual non-WC base pairs that comprise a motif, the members of each subfamily of isosteric base pairs define the sequence signature of the motif. Thus, the geometric base pair classification provides the framework for the present review. Here, we will first discuss the various views that have been taken on the concept of 'motif'. Afterwards, we will describe, following a base-centered view of RNA motifs, the motifs that have been identified and characterized. The names of several motifs have not yet 'crystallized' and we will attempt to use the most global name that is in accordance with the history of discovery (for a pertinent discussion on some of those issues, see [3]). This classification has been adopted by the Nucleic Acid Database for annotation of nucleic acid structures (http://beta-ndb.rutgers.edu/) and has been incorporated in RNAML, the markup language for RNA bioinformatics [4•].

RNA motifs: the view from the phosphodiester backbone

In 1969, Sundaralingam introduced the 'rigid nucleotide' concept on the basis of analysis of experimentally determined torsion angles in nucleotides and polynucleotides [5]. In 1980, Olson pointed out that representing the nucleic acid backbone by two virtual bonds defined by the C4′ and P atoms of each nucleotide was useful for

the statistical analysis of polynucleotide conformations [6]. The virtual bonds proposed by Olson naturally divide each nucleotide into two heminucleotide blocks (C4′–C5′–O5′–P and C4′–C3′–O3′–P) defining two virtual torsional angles, which were labeled ω_v and ω_v'. Olson and Srinivasan introduced conformation wheels for displaying and correlating polynucleotide conformations in complex molecules such as tRNA [7]. In a series of articles, Yathindra and co-workers showed that ω_v and ω_v' provide a useful way to reduce the description of polynucleotide conformations to two-dimensional Ramachandran-like plots [8–10]. They also applied this analysis to tRNAs, the only complex RNAs for which three-dimensional structures were then known at atomic resolution. Recently, Duarte and Pyle rediscovered this virtual bond description and applied it to the expanding RNA structural database to identify and classify motifs [11]. Their virtual torsions η and θ correspond to ω_v and ω_v' of the older nomenclature, respectively. The two-dimensional plot of η versus θ corresponds formally to the Ramachandran plot of ϕ versus ψ used for analyzing protein conformation and, in fact, clusters of residues that share conformational properties often belong to the same type of structural submotif and appear in discrete regions of the η versus θ plots [11].

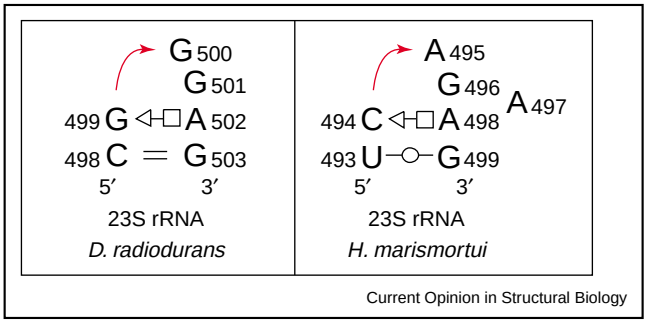
RNA motifs: the view from the bases and their pairing interactions

Given the flexibility of the RNA backbone, an exclusive focus on backbone conformation does not guarantee identification of related, recurrent motifs. This is compounded by the composite nature of certain occurrences of well-known motifs, such as the sarcin/ricin loop motif [12••,13•]. A complementary approach focuses on the specific interactions between the bases comprising motifs, especially the edge-to-edge pairing interactions. The nature and geometrical constraints of these interactions largely determine the trajectory and the mobility of the backbone. We recently presented a catalogue of base pairs grouped into the 12 geometric families (plus additional families comprising bifurcated base pairs [14]) and, within each family, classified the base pairs into isosteric subfamilies using C1′–C1′ distances as the criterion [1•]. The set of observed sequence variants for a putative base pair in a properly aligned set of sequences should be a subset of the isosteric subfamily. Using this criterion, comparative sequence analysis may be productively applied to identify recurrent motifs of known three-dimensional structure, as we demonstrated with regard to the sarcin/ricin and loop E motifs in the 16S and 23S rRNAs [13•,15].

Hairpin loop motifs

A systematic cataloguing of all hairpin loops in the rRNA structures, many of which are unique and quite complex, has not yet been presented. Attention has been focused rather on identifying recurrent motifs. As predicted by sequence analysis, a significant number of GNRA and UNCG tetraloops, as well as variations thereof, are found in the ribosomal structures, most of which involve inserted unpaired nucleotides. For example, the $C_{494}AGAA_{498}$ pentaloop in *Haloarcula marismortui* 23S rRNA [16] is essentially a tetraloop with an inserted bulged base, A497, and is essentially identical to the $G_{499}GGA_{502}$ tetraloop that is observed in the corresponding position of the 23S rRNA of *Deinococcus radiodurans* [17], as shown schematically in Figure 1. These two hairpins are closed by isosteric *trans* H/SE ('sheared') base pairs, A_{498}/C_{494} in *H. marismortui* and A_{502}/G_{499} in *D. radiodurans*. Although A/G and A/A sheared pairs have been known for some time, A/C, A/U, C/A and C/Y sheared base pairs were first predicted from analysis of sequence variations in loop E motifs [18] and were subsequently observed in various contexts, including hairpins.

Figure 1



Comparison of GNRA-like loops at equivalent positions in the 23S rRNAs of *H. marismortui* and *D. radiodurans* — one with a bulge that does not change the structure. The loops are closed by isosteric closing base pairs (sheared G/A and C/A). See Table 1 for a key to the symbols denoting the different families of base pairs.

Table 1

Symbols representing the 12 families of base pairs.

Glycosidic bond orientation	Interacting edges	Symbol
Cis	WC/WC	
Trans	WC/WC	
Cis	WC/H	
Trans	WC/H	
Cis	WC/SE	
Trans	WC/SE	
Cis	H/H	
Trans	H/H	
Cis	H/SE	
Trans	H/SE	
Cis	SE/SE	
Trans	SE/SE	

The 12 families of base pairs as characterized by their interacting edges with, at the right, the symbols used (adapted from [2••]).

Nagaswamy and Fox [19•] recently analyzed the structures and interactions of motifs related to the conserved T-loop motif of tRNA [20,21] in the large [16] and small subunits of the ribosome. They identified conserved features shared by these motifs and the interactions they form. Most importantly, all are closed by *trans* WC/H base pairs, which are usually U/A but are sometimes isosteric C/A. The ability of C/A to substitute for U/A *trans* WC/H base pairs was also proposed on isosteric grounds on the basis of analysis of sequence variants of loop E motifs [18].

Gutell and co-workers recently proposed grouping together 24 RNA hairpin loops from the 23S and 16S rRNAs under the name 'lone-pair triloops' [22]. These hairpins share the property of comprising a loop of three unpaired nucleotides closed by a single base pair that does not belong to a contiguous helix. However, the closing base pairs of these loops belong to several different geometric families. Moreover, the strand orientations of the nucleotides comprising the closing base pairs are parallel in some of the motifs but antiparallel in others. Thus, the loops display no common conformational features, which calls into question the usefulness of grouping them together. Applying motif analysis based on the geometric classification of base pairs, as outlined above, the 24 triloops presented by Lee *et al.* [22] comprise several distinct groups, representatives of which are shown in Figure 2 using annotated diagrams.

Some of the examples of 'lone-pair triloops' are actually portions of T-loop motifs. It seems to us more useful to consider the T-loop *per se* as a distinct motif, as presented by other authors [19•]. Nonetheless, we have noted that certain motifs combine in a hierarchically nested fashion to form more complex motifs [13•,23]. Molecular dynamics and phylogenetic studies on the anticodon loop of tRNAs [24] showed an extended molecular signature of such loops with a non-WC closing pair between residues 32 and 38, and additional sugar–base contacts between residues 33 and 35.

Asymmetric internal loops

'A-minor' motifs

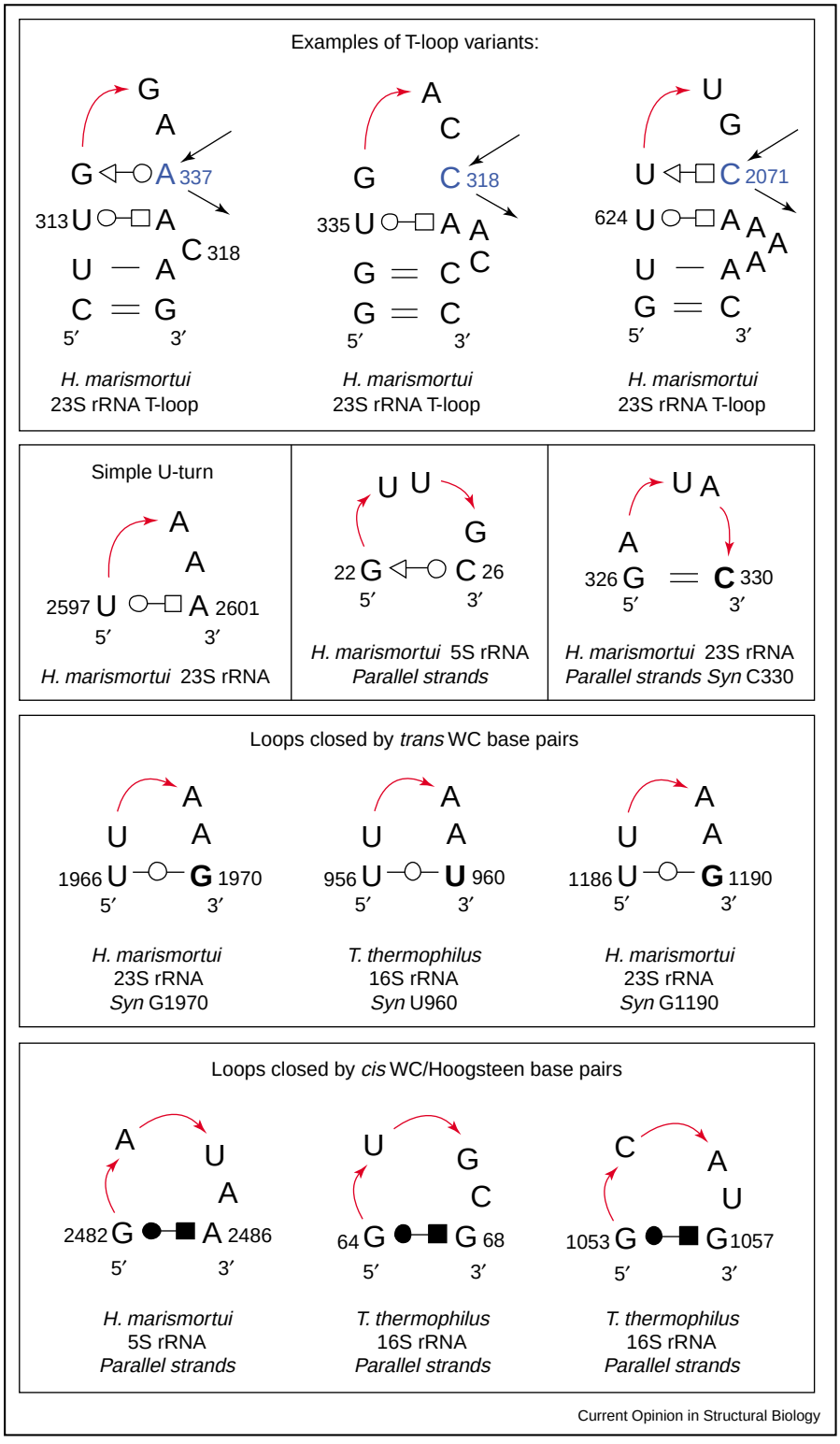
The importance to RNA folding of tertiary interactions involving formally unpaired adenosines with the shallow (minor) groove of canonical WC helices was first suggested on the basis of three-dimensional modeling and comparative sequence analysis of group I introns in the early 1990s [25]. Experimental support for such interactions was obtained by modular replacement of loop–helix interactions by tertiary WC base pairs [26,27]. Subsequently, sequence analysis, base substitution and *in vitro* selection experiments showed that the 11-nucleotide RNA motif CCUAAG...UAUGG interacts specifically with GAAA tetraloops [28,29]. Soon after, X-ray crystallography revealed these interactions in detail in the structure of the hammerhead ribozyme [30,31] and the P4–P6 domain of group I introns [32]. In the hammerhead structure, the GNRA–helix interactions were intermolecular, whereas the GAAA–11-nucleotide receptor interactions in the P4–P6 domain are intramolecular.

Recently, the *trans* and *cis* SE/SE base pairs involving adenosines were given the alternative designations 'type I and type II A-minor motifs' [33•]. This derives from the fact that, in these interactions, one of the bases is an adenosine that interacts in the minor groove of a (usually) canonically paired base pair. The adenosine interacts primarily with one of the two bases, but may also form one or more hydrogen bonds to the other. The 'A-minor' designation can be misleading, however, because other nucleotides besides adenosines can participate in these interactions [1•]. Moreover, base triples involving canonical base pairs in helices are not necessarily implied. A further problem is the confusion resulting from the arbitrariness of the designations type I, II or II′. Thus, the geometric designation is preferred for RNA bioinformatics applications. The 'A-minor' motifs (*cis* and *trans* SE/SE base pairs), as well as the related *cis* and *trans* WC/SE base pairs, being single non-WC base pairs, must generally mediate interactions in combinations. Several different combinations have been observed in a variety of different structures and have been reviewed [13•,34].

The combination of *cis* and *trans* SE/SE is especially prevalent and occurs in different contexts, including GNRA hairpin loops and asymmetric internal loops. For example, an asymmetric loop in the penultimate helix of 16S comprises two highly conserved, unpaired adenosines that swing out to form stacked *cis* and *trans* SE/SE base pairs when cognate mRNA codon and tRNA anticodon nucleotides pair, thus proof-reading the decoding interaction between tRNA and mRNA [35••,36••]. Unless the correct tRNA is bound, these non-WC pairs do not form and the unpaired adenosines contract into the asymmetric loop. However, the binding of certain drugs can lock the motif into the extended conformation even when the cognate codon–anticodon interaction is not present [37]. Interestingly, the same interaction is seen in crystals of the aminoglycoside antibiotics paromomycin and tobramycin bound to an oligonucleotide model of the 30S decoding site [38•,39•].

Although in the ribosome the conserved, unpaired adenosines can interact productively with any of the canonical WC base pairs formed by different cognate codon–anticodon interactions to form the *cis* and *trans* SE/SE decoding motif, in other contexts, such as tertiary interactions in the group I RNAs, a preference for certain WC base pairs over others has been measured [40•]. In any case, WC pairs are preferred over non-WC pairs. The preference for unpaired adenines to mediate these interactions was also quantified in the group I model system [41•].

Figure 2



Hairpin loops classified as 'lone-pair triloops' [22] reclassified according to their closing base pairs. See Table 1 for a key to the symbols denoting the different families of base pairs.

K-turn motifs

A-minor motifs also play crucial roles in *local* tertiary interactions that create bends or kinks in helical stems. The so-called K-turn was first recognized as a motif in the analysis of the complex between the human 15.5 kDa protein and a U4 snRNA fragment [42]. The recurrence of the motif was later clearly established following the resolution of the large and small subunits of the ribosome [12••]. K-turns are composite motifs built on a core comprising two cross-strand stacked adenosines that stabilize the 'kink' in the strands of the double helix by interacting in the shallow (minor) groove of the second helix adjoining the motif. These adenosines form tandem sheared (*trans* H/SE) base pairs or, in one case, a sheared pair adjacent to a *trans* WC/H pair, as in loop E and sarcin-like motifs. An example of a tandem sheared A/G motif from *H. marismortui* 23S rRNA that interacts with a distant helix in the same way is compared with some examples of K-turns in Figure 3.

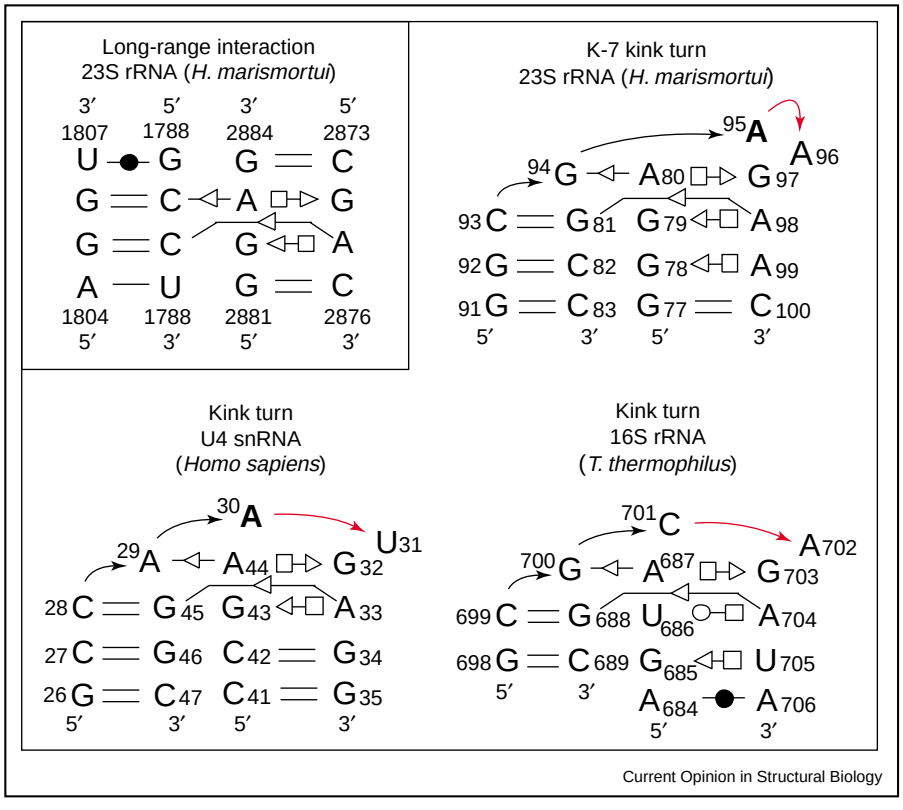
Sarcin-like motif

This motif was identified by NMR almost simultaneously in the highly conserved factor-binding site of 23S-like rRNAs that is subject to attack by the endotoxins sarcin and ricin [43], and in eucaryal 5S loop E [44]. Based on the NMR structure, comparative sequence analysis at the level of motifs was employed to predict its occurrence at several positions in the 16S and 23S rRNAs, as well as in group I and group II introns [15]. The appearance of crystal structures for the rRNAs provided the opportunity to assess the predictive power, as well as the pitfalls, of motif analysis [13•]. One conclusion of this work is that it is essential to determine the surrounding secondary structure accurately before attempting to analyze motifs. One motif was incorrectly identified as a sarcin-like motif because of an error in the secondary structure. A second conclusion was that recurrent motifs can appear in variable regions that are structured in a different way in distantly related organisms or are altogether absent.

The C-motif

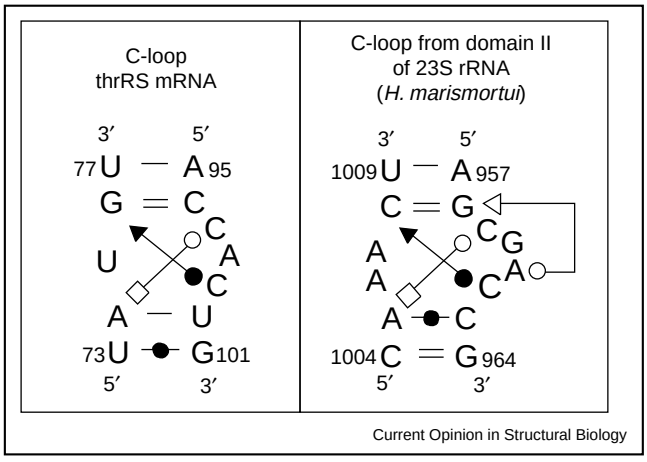
The C-motif, an asymmetric internal loop, is observed to interact with proteins in the 16S rRNA [45,46], the 23S rRNA [16] and the complex between the messenger for the threonine synthetase and the synthetase itself [47]. Two cytosine residues in the longer strand of the asymmetric loop form non-WC pairs with residues on the other strand participating in the WC pairs enclosing the loop. The 5′-end cytosine forms a *trans* WC/H pair with an adenine on the other strand. The 3′-end cytosine forms

Figure 3



Examples of kink turns and comparison to a distant tertiary motif. Upper right: archetypal K-turn in helix 7 (K-7) of *H. marismortui* 23S rRNA [12••]. Upper left: tandem sheared G/A motif forming identical interaction with a distant helix in the 23S rRNA of *H. marismortui*. Lower left: K-turn from U4 snRNA. Lower right: K-turn from 16S rRNA. See Table 1 for a key to the symbols denoting the different families of base pairs.

Figure 4



Examples of C-motifs. Left panel: the canonical C-motif, with three residues on one strand and one residue on the other, from the *E. coli* threonyl tRNA synthetase mRNA. Right panel: variant C-motif, with four residues on one strand and two residues on the other, from domain II of 23S rRNA of *H. marismortui*. The presence of bulged G960 facilitates *trans* WC/H pairing of A961 with G958, further stabilizing the motif without distorting it. The bulged A1006 and A1007 form a long-range A-minor motif with a hairpin stem in domain V of 23S rRNA. See Table 1 for a key to the symbols denoting the different families of base pairs.

a *cis* WC/SE pair with the other closing pair but also on the other strand. The canonical C-loop is an asymmetric loop (Figure 4, left panel) with three residues including the two cytosines on one strand and one bulged base in the other strand. Variant C-loops comprise additional bulged bases. An example is the C-loop in domain II of the 23S rRNA (Figure 4, right panel), which has four residues on one strand and two on the other. As in the GNRA hairpin loops, the additional bulged bases do not affect the geometry of the motif or formation of the non-WC base pairs. In fact, it appears that, in the domain II C-loop, the presence of bulged G960 helps form additional stabilizing interactions (Figure 4).

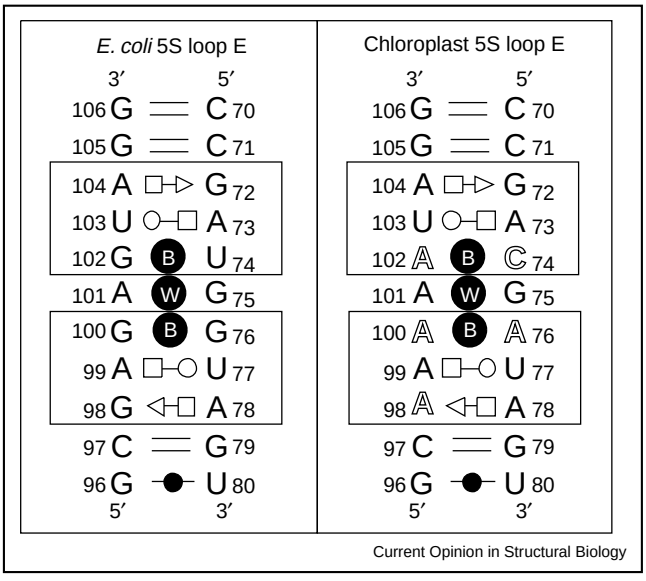
In the sarcin-like motif, residues that occupy equivalent positions interact alternatively with RNA or protein [13•]. With the C-motif, the tendency is to bind solely protein. Protein atoms interact mainly with the sugar–phosphate backbone on either side of the asymmetric loop. The shallow groove of the middle adenine in the long internal bulge interacts with protein atoms in a couple cases of the motif. The C-loop in domain II of the 23S rRNA is an exception in that the two bulged-out adenines, A1006 and A1007, form A-minor-type interactions with domain V. In most C-motifs, there is only one bulged base at this position. The domain II C-loop is therefore an RNA–RNA motif similar to the GNRA tetraloops and their receptors [25,28] and related motifs [13•,33•]. It thus resembles the asymmetric loop found in the 7SL RNA of the signal recognition particle [48••], in which the long segment of the asymmetric loop helps to organize the motif so that two adenines on the other segment can form A-minor motif interactions with the shallow groove of another helix.

Symmetric internal loops

Chloroplast 5S rRNA loop E

The isostericity matrix is a 4 × 4 matrix (where row and column are A, G, C, U) that gives, for each of the 12 families of base–base pairing, the pairs that are isosteric (i.e. that give a good superposition of the attached C1′ carbon atoms). A striking example of the usefulness of the isostericity matrix approach to modeling RNA motifs concerns chloroplast 5S rRNA loop E. Plant chloroplasts are derived from photosynthetic bacteria, yet the loop E region of their 5S rRNA differs considerably from the consensus bacterial loop E sequence (Figure 5), whereas the sequence of *Escherichia coli* loop E is almost identical to the consensus sequence. The crystal structure of *E. coli* 5S loop E was published in 1997 and revealed that loop E comprises a unique tract of seven non-WC pairs, including bifurcated and water-inserted pairs never before seen [49,50]. The structure of chloroplast 5S loop E was modeled in 1998 by homology to the structure of *E. coli* loop E [34], on the basis of reasoning that the structures should be essentially identical as a result of the evolutionary relationship between chloroplasts and bacteria. The base pairing observed in the crystal structure was analyzed and used to construct sequence alignments for

Figure 5



Loop E of *E. coli* and *Spinachia oleracea* (spinach) chloroplast 5S rRNAs compared to bacterial consensus sequence. Bases differing from consensus sequence are in outline font. B indicates *cis* WC bifurcated base pairs and *cis* WC water-inserted pairs. The structure of chloroplast 5S loop E was recently solved by NMR [52•], in agreement with the prediction of its structure using isostericity matrices [34]. W, water. See Table 1 for a key to the symbols denoting the different families of base pairs.

all bacterial and chloroplast loop E sequences. From the alignments, isostericity matrices were derived for each base pair type and used to construct plausible models for the base pairing in the chloroplast sequence [34]. The recently published NMR structure, derived using a variety of constraints including nuclear Overhauser effect (NOE), J-coupling, chemical shift and residual dipolar couplings [51,52•], confirmed the predictions of the geometric family to which each base pair belongs.

Junction loops

As for hairpin loops, a thorough analysis of junction loops in complex RNA structures has not yet been attempted. It is clear, however, that many junctions are composite, comprising nested or adjacent submotifs. Thus, sarcin/ricin or loop E motifs participate in the assembly of complex RNA by forming elements of certain junctions [15]. Another example is the hook-turn motif.

Hook-turn motif

The hook-turn motif was identified while attempting to characterize the aberrant 5S loop E of *Corynebacterium minutissimum* with complementary oligonucleotides [53]. The backbone of one of the strands folds at an angle of almost 180° and interacts in the shallow groove of the helix from which it originates, in a manner seen in certain multihelix junctions in the large rRNAs. As in the junctions, non-WC base pairs form where the strands begin to diverge.

NMR structures versus X-ray structures

Although it has been partially eclipsed by the rapid progress in X-ray determinations of RNA structure, NMR continues to be applied to RNA, with a shift in focus toward identifying differences between crystal and solution structures, and better characterizing solution dynamics and the effects of protein or RNA binding interactions ('induced fit') upon RNA conformation [54]. New methods have been applied to overcome the difficulties of RNA structure determination by NMR, which are caused by the low density of protons in RNA compared to proteins. Most notably, it has been demonstrated that the information provided by residual dipolar couplings, which complements local structural information from NOE and J-couplings, improves the quality of the RNA structures that can be obtained by NMR [51,52•,55].

In addition to the chloroplast 5S loop E, we discuss two examples of recently reported NMR structures. The solution structure of hairpin UUAAGU from the S2 ribosomal protein binding site of 16S was determined by NMR [56]. As in the crystal structure, the closing base pair is a wobble U/U and the loop itself comprises a U-turn with the G bulged.

An NMR study on the effect of chemical modification (2′-O-methylation) of U2552 (*E. coli* numbering) in the A-loop of 23S rRNA, on the other hand, reported that the solution and crystal structures are dramatically different [57]. The A-loop directly interacts with aminoacyl-site tRNA. In the crystal structure, the closing U/C base pair is *trans* SE/H, whereas in the NMR structure it is *cis* WC with the U and C directly hydrogen bonded at the imino positions. When the U is modified, the U and C are unpaired in three of the four structures. In crystal structures, *cis* WC U/C pairs invariably have an inserted water molecule, perhaps on account of the U(O2)–C(O2) repulsions. Residual dipolar couplings were not employed for the refinement of this structure. It is not clear whether NMR techniques relying on NOE and J-coupling are adequate to determine all non-WC pairs.

Conclusions

This short survey of RNA motifs shows that the field is far from being settled. The crystallographic work on the ribosome has revolutionized our perceptions of RNA structures and several groups are struggling to understand and organize the wealth of new data available [1•,58,59]. Even the definition of motif is still open to discussion. How different must two structures be to be recognized as two distinct motifs? How different must two base pairings be if they are not to be considered 'isosteric'? How should we most meaningfully classify structural entities? Should we classify them according to the number of unpaired bases in internal or hairpin loops? Very often, one or more additional residues occur without significantly affecting the fold or backbone trajectory, as, for example, in the GNRA tetraloops [60] and the looped-out bases in T-loops [19•] or in K-turns [12••] or the variants of the C-loop. The most prevalent RNA–RNA long-range contacts implicate adenines interacting with the shallow grooves of base pairs (A-minor motifs). However, there are many ways of presenting bulging adenines for the formation of A-minor contacts. It is now appreciated that complex RNA motifs are assembled from smaller ones; where is the limit? The most astonishing case is the formation of a given motif by the association of more than two strands, as observed, for example, with composite sarcin-like motifs in complex junctions. A clearer understanding of these issues will ultimately help us unveil the underlying laws of Darwinian evolution at the molecular level.

Acknowledgements

NBL was supported by National Institutes of Health grant 2R15-GM55898.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. • Leontis NB, Stombaugh J, Westhof E: **The non-Watson–Crick base pairs and their associated isostericity matrices**. *Nucleic Acids Res* 2002, **30**:3497-3531.
Compendium of observed and modeled non-WC base pairs, organized into geometric families and isosteric subfamilies.

2. •• Leontis NB, Westhof E: **Geometric nomenclature and classification of RNA base pairs**. *RNA* 2001, **7**:499-512.
Explains the geometric classification of RNA base pairs into 12 families and proposes annotations for RNA motifs in two-dimensional diagrams.

3. Moore PB: **Structural motifs in RNA**. *Annu Rev Biochem* 1999, **68**:287-300.

4. • Waugh A, Gendron P, Altman R, Brown JW, Case D, Gautheret D, Harvey SC, Leontis N, Westbrook J, Westhof E *et al.*: **RNAML: a standard syntax for exchanging RNA information**. *RNA* 2002, **8**:707-717.
Describes the RNAML markup language for RNA structural bioinformatics.

5. Sundaralingam M: **Stereochemistry of nucleic acids and their constituents**. *Biopolymers* 1969, **7**:821-838.

6. Olson WK: **Configuration statistics of polynucleotide chains**. *Macromolecules* 1980, **13**:721-728.

7. Srinivasan AR, Olson WK: **Yeast tRNAPhe conformation wheels: a novel probe of the monoclinic and orthorhombic models**. *Nucleic Acids Res* 1980, **8**:2307-2329.

8. Malathi R, Yathindra N: **Secondary and tertiary structural foldings in tRNA. A diagonal plot analysis using the blocked nucleotide scheme**. *Biochem J* 1982, **205**:457-460.

9. Malathi R, Yathindra N: **The heminucleotide scheme: an effective probe in the analysis and description of ordered polynucleotide structures**. *Biopolymers* 1983, **22**:2961-2976.

10. Malathi R, Yathindra N: **Backbone conformation in nucleic acids: an analysis of local helicity through heminucleotide scheme and a proposal for a unified conformational plot**. *J Biomol Struct Dyn* 1985, **3**:127-144.

11. Duarte CM, Pyle AM: **Stepping through an RNA structure: a novel approach to conformational analysis**. *J Mol Biol* 1998, **284**:1465-1478.

12. •• Klein DJ, Schmeing TM, Moore PB, Steitz TA: **The kink-turn: a new RNA secondary structure motif**. *EMBO J* 2001, **20**:4214-4221.
Analysis of an important recurrent motif that induces sharp turns in RNA stems.

13. • Leontis NB, Stombaugh J, Westhof E: **Motif prediction in ribosomal RNAs: lessons and prospects for automated motif prediction in homologous RNA molecules**. *Biochimie* 2002, **84**:961-973.
Evaluates the isostericity matrix approach to prediction of recurrent motifs using comparative sequence analysis.

14. Auffinger P, Westhof E: **Singly and bifurcated hydrogen-bonded base-pairs in tRNA anticodon hairpins and ribozymes**. *J Mol Biol* 1999, **292**:467-483.

15. Leontis NB, Westhof E: **A common motif organizes the structure of multi-helix loops in 16S and 23S ribosomal RNAs**. *J Mol Biol* 1998, **283**:571-583.

16. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA: **The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution**. *Science* 2000, **289**:905-920.

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

18. Leontis NB, Westhof E: **The 5S rRNA loop E: chemical probing and phylogenetic data versus crystal structure**. *RNA* 1998, **4**:1134-1153.

19. • Nagaswamy U, Fox GE: **Frequent occurrence of the T-loop RNA folding motif in ribosomal RNAs**. *RNA* 2002, **8**:1112-1119.
Analysis of hairpin loops with features common to tRNA T-loops.

20. Quigley GJ, Rich A: **Structural domains of transfer RNA molecules**. *Science* 1976, **194**:796-806.

21. Stout CD, Mizuno H, Rubin J, Brennan T, Rao ST, Sundaralingam M: **Atomic coordinates and molecular conformation of yeast phenylalanyl tRNA. An independent investigation**. *Nucleic Acids Res* 1976, **3**:1111-1123.

22. Lee J, Cannone J, Gutell R: **The lonepair triloop: a new motif in RNA structure**. *J Mol Biol* 2003, **325**:65-83.

23. Westhof E, Fritsch V: **RNA folding: beyond Watson–Crick pairs**. *Structure Fold Des* 2000, **8**:R55-R65.

24. Auffinger P, Westhof E: **An extended structural signature for the tRNA anticodon loop**. *RNA* 2001, **7**:334-341.

25. Michel F, Westhof E: **Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis**. *J Mol Biol* 1990, **216**:585-610.

26. Jaeger L, Michel F, Westhof E: **Involvement of a GNRA tetraloop in long-range RNA tertiary interactions**. *J Mol Biol* 1994, **236**:1271-1276.

27. Murphy FL, Cech TR: **GAAA tetraloop and conserved bulge stabilize tertiary structure of a group I intron domain**. *J Mol Biol* 1994, **236**:49-63.

28. Costa M, Michel F: **Frequent use of the same tertiary motif by self-folding RNAs**. *EMBO J* 1995, **14**:1276-1285.

29. Costa M, Michel F: **Rules for RNA recognition of GNRA tetraloops deduced by *in vitro* selection: comparison with *in vivo* evolution**. *EMBO J* 1997, **16**:3289-3302.

30. Pley HW, Flaherty KM, McKay DB: **Three-dimensional structure of a hammerhead ribozyme**. *Nature* 1994, **372**:68-74.

31. Pley HW, Flaherty KM, McKay DB: **Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix**. *Nature* 1994, **372**:111-113.

32. Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, Doudna JA: **Crystal structure of a group I ribozyme domain: principles of RNA packing**. *Science* 1996, **273**:1678-1685.

33. • Nissen P, Ippolito JA, Ban N, Moore PB, Steitz TA: **RNA tertiary interactions in the large ribosomal subunit: the A-minor motif**. *Proc Natl Acad Sci USA* 2001, **98**:4899-4903.
Identifies recurrent motifs involving interactions of adenosines in the shallow (minor) grooves of RNA helices.

34. Leontis NB, Westhof E: **Conserved geometrical base-pairing patterns in RNA**. *Q Rev Biophys* 1998, **31**:399-455.

35. •• Ogle JM, Brodersen DE, Clemons WM Jr, Tarry MJ, Carter AP, Ramakrishnan V: **Recognition of cognate transfer RNA by the 30S ribosomal subunit**. *Science* 2001, **292**:897-902.
See annotation to [36••].

36. •• Ogle JM, Murphy FV, Tarry MJ, Ramakrishnan V: **Selection of tRNA by the ribosome requires a transition from an open to a closed form**. *Cell* 2002, **111**:721-732.
These two papers [35••,36••] describe the conformational changes in 16S rRNA that allow discrimination between the binding of cognate and near-cognate tRNA to the A-site.

37. Carter AP, Clemons WM, Brodersen DE, Morgan-Warren RJ, Wimberly BT, Ramakrishnan V: **Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics**. *Nature* 2000, **407**:340-348.

38. • Vicens Q, Westhof E: **Crystal structure of paromomycin docked into the eubacterial ribosomal decoding A site**. *Structure (Camb)* 2001, **9**:647-658.
See annotation to [39•].

39. • Vicens Q, Westhof E: **Crystal structure of a complex between the aminoglycoside tobramycin and an oligonucleotide containing the ribosomal decoding A site**. *Chem Biol* 2002, **9**:747-755.
These two papers [38•,39•] describe conformational changes induced by drug binding to oligonucleotide models of the 16S decoding site.

40. • Battle DJ, Doudna JA: **Specificity of RNA–RNA helix recognition**. *Proc Natl Acad Sci USA* 2002, **99**:11676-11681.
See annotation to [41•].

41. • Doherty EA, Batey RT, Masquida B, Doudna JA: **A universal mode of helix packing in RNA**. *Nat Struct Biol* 2001, **8**:339-343.
These two papers [40•,41•] analyze the base specificities of non-WC base pairs in the shallow (minor) groove of RNA helices.

42. Vidovic I, Nottrott S, Hartmuth K, Luhrmann R, Ficner R: **Crystal structure of the spliceosomal 15.5 kD protein bound to a U4 snRNA fragment**. *Mol Cell* 2000, **6**:1331-1342.

43. Szewczak AA, Moore PB, Chang YL, Wool IG: **The conformation of the sarcin/ricin loop from 28S ribosomal RNA**. *Proc Natl Acad Sci USA* 1993, **90**:9581-9585.

44. Wimberly B, Varani G, Tinoco I Jr: **The conformation of loop E of eukaryotic 5S ribosomal RNA**. *Biochemistry* 1993, **32**:1078-1087.

45. Clemons WM Jr, Brodersen DE, McCutcheon JP, May JL, Carter AP, Morgan-Warren RJ, Wimberly BT, Ramakrishnan V: **Crystal structure of the 30S ribosomal subunit from *Thermus thermophilus*: purification, crystallization and structure determination**. *J Mol Biol* 2001, **310**:827-843.

46. Wimberly BT, Brodersen DE, Clemons WM Jr, Morgan-Warren RJ, Carter AP, Vonrhein C, Hartsch T, Ramakrishnan V: **Structure of the 30S ribosomal subunit**. *Nature* 2000, **407**:327-339.

47. Torres-Larios A, Dock-Bregeon AC, Romby P, Rees B, Sankaranarayanan R, Caillet J, Springer M, Ehresmann C, Ehresmann B, Moras D: **Structural basis of translational control by *Escherichia coli* threonyl tRNA synthetase**. *Nat Struct Biol* 2002, **9**:343-347.

48. •• Kuglstatter A, Oubridge C, Nagai K: **Induced structural changes of 7SL RNA during the assembly of human signal recognition particle**. *Nat Struct Biol* 2002, **9**:740-744.
Two adenosines in an asymmetric loop bulge out to form A-minor interactions with another helix identical to those formed by A1492 and A1493 in the decoding site of 30S ribosomes.

49. Correll CC, Freeborn B, Moore PB, Steitz TA: **Metals, motifs, and recognition in the crystal structure of a 5S rRNA domain**. *Cell* 1997, **91**:705-712.

50. Correll CC, Wool IG, Munishkin A: **The two faces of the *Escherichia coli* 23 S rRNA sarcin/ricin domain: the structure at 1.11 Å resolution**. *J Mol Biol* 1999, **292**:275-287.

51. Warren JJ, Moore PB: **Application of dipolar coupling data to the refinement of the solution structure of the sarcin–ricin loop RNA**. *J Biomol NMR* 2001, **20**:311-323.

52. • Vallurupalli P, Moore PB: **The solution structure of the loop E region of the 5S rRNA from spinach chloroplasts**. *J Mol Biol* 2003, **325**:843-856.
An NMR study of chloroplast 5S rRNA loop E providing evidence for conservation of the motif.

53. Szep S, Wang J, Moore PB: **The crystal structure of a 26-nucleotide RNA containing a hook-turn**. *RNA* 2003, **9**:44-51.

54. Leulliot N, Varani G: **Current topics in RNA-protein recognition: control of specificity and biological function through induced fit and conformational capture**. *Biochemistry* 2001, **40**:7947-7956.

55. Warren JJ, Moore PB: **A maximum likelihood method for determining D(a)(PQ) and R for sets of dipolar coupling data**. *J Magn Reson* 2001, **149**:271-275.

56. Zhang H, Fountain MA, Krugh TR: **Structural characterization of a six-nucleotide RNA hairpin loop found in *Escherichia coli*, r(UUAAGU)**. *Biochemistry* 2001, **40**:9879-9886.

57. Blanchard SC, Puglisi JD: **Solution structure of the A loop of 23S ribosomal RNA**. *Proc Natl Acad Sci USA* 2001, **98**:3720-3725.

58. Klosterman PS, Tamura M, Holbrook SR, Brenner SE: **SCOR: a structural classification of RNA database**. *Nucleic Acids Res* 2002, **30**:392-394.

59. Nagaswamy U, Larios-Sanz M, Hury J, Collins S, Zhang Z, Zhao Q, Fox GE: **NCIR: a database of non-canonical interactions in known RNA structures**. *Nucleic Acids Res* 2002, **30**:395-397.

60. Abramovitz DL, Pyle AM: **Remarkable morphological variability of a common RNA folding motif: the GNRA tetraloop–receptor interaction**. *J Mol Biol* 1997, **266**:493-506.