

# Recurrent structural RNA motifs, Isostericity Matrices and sequence alignments

Aurélie Lescoute, Neocles B. Leontis<sup>1</sup>, Christian Massire<sup>2</sup> and Eric Westhof\*

Institut de Biologie Moléculaire et Cellulaire du CNRS, UPR 9002, Université Louis Pasteur, 15 rue René Descartes, F-67084 Strasbourg Cedex, France, <sup>1</sup>Department of Chemistry and Center for Biomolecular Sciences, Bowling Green State University, Bowling Green, OH 43403, USA and <sup>2</sup>Ibis Therapeutics, Carlsbad Research Center, 2292 Faraday Avenue, Carlsbad, CA 92008, USA

Received February 28, 2005; Revised and Accepted April 6, 2005

## ABSTRACT

**The occurrences of two recurrent motifs in ribosomal RNA sequences, the Kink-turn and the C-loop, are examined in crystal structures and systematically compared with sequence alignments of rRNAs from the three kingdoms of life in order to identify the range of the structural and sequence variations. Isostericity Matrices are used to analyze structurally the sequence variations of the characteristic non-Watson–Crick base pairs for each motif. We show that Isostericity Matrices for non-Watson–Crick base pairs provide important tools for deriving the sequence signatures of recurrent motifs, for scoring and refining sequence alignments, and for determining whether motifs are conserved throughout evolution. The systematic use of Isostericity Matrices identifies the positions of the insertion or deletion of one or more nucleotides relative to the structurally characterized examples of motifs and, most importantly, specifies whether these changes result in new motifs. Thus, comparative analysis coupled with Isostericity Matrices allows one to produce and refine structural sequence alignments. The analysis, based on both sequence and structure, permits therefore the evaluation of the conservation of motifs across phylogeny and the derivation of rules of equivalence between structural motifs. The conservations observed in Isostericity Matrices form a predictive basis for identifying motifs in sequences.**

## INTRODUCTION

Folded RNA molecules exhibit complex architectures in which a large fraction of the bases engage in non-Watson–Crick

(non-WC) base pairs and form motifs that mediate long-range RNA–RNA interactions and create binding sites for proteins and small molecule ligands. For homologous RNA molecules, the architectures are highly conserved, more than the underlying secondary structures, as different secondary structures can produce basically similar three-dimensional (3D) folds (1,2). Because of the prevalence and central roles of non-WC base pairs in RNA tertiary interactions, we have sought in previous works to extend the isostericity concept to include them so as to develop rules for identifying base substitutions in RNA sequences that conserve the 3D structures of motifs (3,4). Moreover, elucidating these substitution rules is the key for refining and eventually automatizing sequence alignment, and is crucial for deeper understanding of the coupling between RNA evolution and RNA architecture. Our aim is to derive sets of evolutionary and structural rules for constructing pertinent architectures for RNA molecules on the basis of an analysis of sequences. Ultimately, the derived sets of rules should help predict RNA folds from sequences alone. Thus, the aim of the present analysis is not to classify RNA motifs or to search for motifs in folded X-ray structures of RNA (5,6). We wish to relate a previously identified RNA motif to the sets of constraints between nucleotides which allow the formation of the motif. The starting point is first crystal structures and, then, an alignment of RNA sequences functionally homologous to the one crystallized. Therefore, the present analysis cannot offer explanations for the kinetics of folding or its underlying physico-chemical requirements (e.g. presence of magnesium ions, protein cofactors or small ligands), since it assumes the final folded state. Experimental and theoretical studies have emphasized the dynamical flexibility and the magnesium or protein requirements of K-turns (7–10). Furthermore, while motifs with recurrent conformers of the sugar-phosphate backbone but without apparent sequence conservation are very useful to our understanding (5,6), they cannot at this stage be used for predicting RNA

\*To whom correspondence should be addressed. Tel: +33 0 3 88 417046; Fax: +33 0 3 88 417066; Email: e.westhof@ibmc.u-strasbg.fr  
Correspondence may also be addressed to N. B. Leontis. Tel: +1 419 372-2031; Fax: +1 419 372 9809; Email: leontis@bgnet.bgsu.edu  
Correspondence may also be addressed to C. Massire. Tel: +1 760 931 9200; Fax: +1 760 431 2768; Email: cmassire@isisph.com

structure from sequence alone and are, thus, not considered here. It is clear that an RNA motif is the result of a combination of base–base interactions and of backbone preferences; for example, the double helix has defined backbone preferences (11) but its identification from sequences alone relies on base–base WC covariations. Likewise, our present analysis aims at deriving the covariation rules for more complex RNA motifs with non-WC base pairs.

We define RNA motifs as ordered arrays of non-WC base pairs under constraints, which may be interspersed with bulged out or inserted bases. Motifs are often embedded within regular helical regions forming internal loops, but may also comprise hairpin or junction loops. The concept of isostericity provides the basis for extracting the constraints acting on RNA motifs. These constraints can then be applied (i) for structure-based alignment of homologous sequences; (ii) for analyzing and classifying RNA motifs; and (iii) for the prediction of 3D structure based on sequence variation in homologous RNA molecules. In the present contribution, we develop and apply Isostericity Matrix analysis to determine sequence signatures for recurrent Kink-turn and C-loop motifs and to evaluate their conservation at corresponding homologous positions in families of ribosomal RNA molecules. We show that this process simultaneously leads to a realignment of homologous sequences. Therefore, Isostericity Matrices are not only useful for the analysis, classification and prediction of the 3D structure of RNA motifs based on sequence signatures in homologous RNA molecules, but also and importantly for the structure-based alignment of homologous sequences and the analysis of the evolution of RNA motifs.

## MATERIALS AND METHODS

### Definitions

*Non-Watson–Crick base pairs.* Just as WC base pairs combine to form regular A-form RNA helices, non-WC base pairs combine to form more complex ‘local RNA motifs’, which we define operationally as ‘ordered arrays of non-WC base pairs’ (12). Local motifs are composed of nucleotides that are ‘bracketed’ by secondary structure and thus include the hairpin, internal bulge and junction (or multi-helix) ‘loops’ that punctuate and organize RNA secondary structures. We distinguish ‘local’ motifs from ‘composite’ motifs that are formed by non-WC interactions involving distant regions of the RNA, as defined by the secondary structure (see below for examples). Such ordered arrays of non-WC base pairs are clearly under structural constraints dictated by the formation of the covalent linkages characteristic of the polynucleotide, the stabilization provided by stacking interactions, as well as the stereochemical preferences of the backbone torsion angles. Furthermore, motifs can often be interrupted by variable numbers of extruded bases. Here, we focus on the constraints on the base identity which are observable in homologous sequences.

*Base pair classification.* To systematically describe and annotate non-WC base pairs in a manner that can elucidate the rules for base substitutions in motifs, we classified them according to simple geometric criteria: the base edges involved in H-bonding interactions (Watson–Crick, Hoogsteen/CH or Sugar-Edges) and the relative orientations of the glycosidic

bonds (*cis* or *trans*) of the interacting nucleotides (3). For each of the twelve distinct geometric families of base pairs that result from this classification, we introduced symbols to annotate secondary structure drawings so as to display the decomposition of motifs into non-WC base pairs.

*Isosteric base pairs and Isostericity Matrices.* To identify isosteric base pairs, we begin with base pairs that belong to the same geometric family because they share the same ‘relative orientations’ of the glycosidic bonds of their interacting bases and so meet the first criterion for isostericity. However, owing to the size difference between purine versus pyrimidine bases, not all base pairs in the same geometric family have the same C1′–C1′ distances separating the interacting bases. The C1′–C1′ distance is therefore the second criterion for isostericity. Thus, base pairs that are in the same geometric family have roughly equal C1′–C1′ distances, and are hydrogen-bonded between equivalent atomic positions and comprise ‘isosteric subgroups’ of base pairs. As optimal hydrogen-bonding can result in shifts of one base relative to another, base pairs that differ only in lateral shifting of the pairing partners belong to almost isosteric subgroups. These base pairs, while not exactly isosteric, may also substitute for each other without significant distortion of motifs, as for example the wobble and the WC pairs. We have presented the isosteric subgroups in  $4 \times 4$  Isostericity Matrices, one for each geometric family (4). Isostericity Matrices therefore indicate which base pairs in each geometric family can substitute for each other without distorting the 3D structure of the motif to which they belong.

*Sequence signatures of motifs.* We define the ‘sequence signature’ of a motif as the set of nucleotide sequences that fold to form the same 3D motif. The sequence signature is different from the consensus sequence because it takes into account coordinated changes needed to maintain non-WC base-pairing and other base-specific interactions. Theoretical sequence signatures for recurrent motifs can be generated combinatorially from all possible isosteric base pairs for each non-WC pair in the motif. Generally, it appears that other constraints—e.g. base stacking and tertiary or quaternary RNA–RNA or RNA–protein interactions—limit the possibilities to a smaller set of sequences than predicted combinatorially.

Identical or nearly identical features that occur independently at non-equivalent positions are recurrent. Thus, ‘recurrent motifs’ are motifs that occur independently in RNA molecules, whether the molecules are homologous or not. Recurrent motifs usually have very similar 3D structures and, consequently, similar, if not identical, sequence signatures, depending on other constraints. Most recurrent motifs are relatively small, but no upper limit on their size has been determined. Kink-turns and C-loops represent examples of recurrent motifs and are the focus of this study. Other examples of recurrent motifs include GNRA and UNCG ‘hairpin’ loops or the sarcin/ricin and loop E ‘internal’ loops.

*Corresponding motifs.* We say two motifs ‘correspond’ to each other when they occur at ‘equivalent positions’ in the 3D structures of a family of RNA molecules. For example, the Kink-turns that occur at equivalent positions in Helix 7 of different 23S rRNAs are corresponding motifs. Corresponding motifs themselves are considered equivalent when their 3D

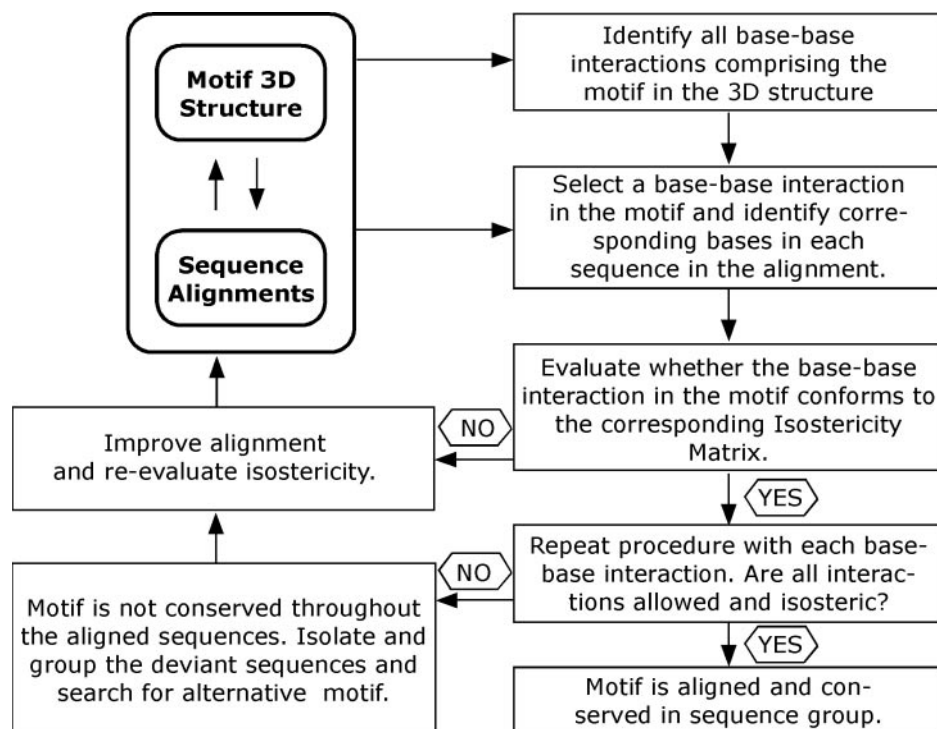
structures are essentially the same. The loop E motifs of homologous 5S ribosomal RNAs occur at 'equivalent positions' and interact with the same regions of their respective 23S rRNAs. The loop E motifs of eukaryal and archaeal 5S rRNAs are also 'equivalent motifs' as they have the same 3D structures. However, the loop E motifs of bacterial and eukaryal 5S rRNAs have different 3D structures and thus are examples of corresponding motifs that are not equivalent. By definition, corresponding motifs that are not equivalent cannot be aligned nucleotide-by-nucleotide.

### Methodological implementations

We manually aligned a set of ribosomal 16S and 23S rRNA sequences extracted from fully sequenced genomes against the available crystal structures of ribosomal RNAs. For Archaea (A) and Bacteria (B) sequences, the atomic resolution crystal structures of *Haloarcula marismortui* (*H.m.*) 23S rRNA and *Thermus thermophilus* (*T.th.*) 16S rRNA were used to define secondary structure masks for the sequence alignments (13,14). In the text below, by default the nucleotide numbering refers to those two crystal structures. To align the conserved primary sequences, we added genomic RNA sequences in a progressive fashion using clustalX (15). Sequences with unidentified nucleotides within a given motif region were not included in the statistics of that motif. For Archaea, 24 sequences of 16S and 23S rRNAs were used; for Bacteria, 800 16S rRNA sequences and 805 23S rRNA sequences were aligned; for Eukarya (E), 5190 18S rRNA sequences and 133 28S rRNA sequences were considered (16). The sequence alignments were analyzed using COSEQ, BioEdit (<http://www.mbio.ncsu.edu/BioEdit/>) and RNAMLview to determine covariations and substitutions at sequence positions

corresponding to non-WC pairs observed in crystal structures and to compare these to the corresponding Isostericity Matrices (17).

Our procedure for analyzing an RNA motif family begins with a detailed structural comparison of all occurrences of the motif in crystal structures to identify characteristic and variable structural features of the motif family. This step provides crucial information regarding variations in base-pairing as well as positions where insertions or deletions occur. The procedure we devised so as to (i) analyze the conservation of RNA motifs and (ii) refine the structure-based alignments of families of RNA molecules is shown in the Flow Chart in Figure 1. The inputs, framed by rounded boxes, are (i) a set of preliminary sequence alignments and (ii) the 3D structure of one of the sequences in the alignment. For each motif, the bases in the 3D structure forming non-WC base pairs are identified (first square box, upper right). For each base pair in the 3D structure, the corresponding nucleotides in each sequence in the alignment are identified (second box on right) and a covariation matrix is constructed using the current alignment and then evaluated by comparison with the corresponding Isostericity Matrix (third box on right). The procedure is repeated for each base pair in the motif (fourth box on right). If the base substitutions for all base pairs are isosteric and insertions or deletions occur at reasonable positions, the motif is considered aligned and conserved in the sequence group (last box, lower right). If the covariation matrix for one or more base pairs in the motif does not conform to the corresponding Isostericity Matrix, improvement of the alignment is attempted, and the isostericities are checked at each step, in an iterative manner (square box, middle left). When further improvement is not possible, the sequences are grouped according to those that conform to the Isostericity Matrices



**Figure 1.** Flow chart showing iterative process relating structure-based RNA motifs and accuracy of sequence alignments.

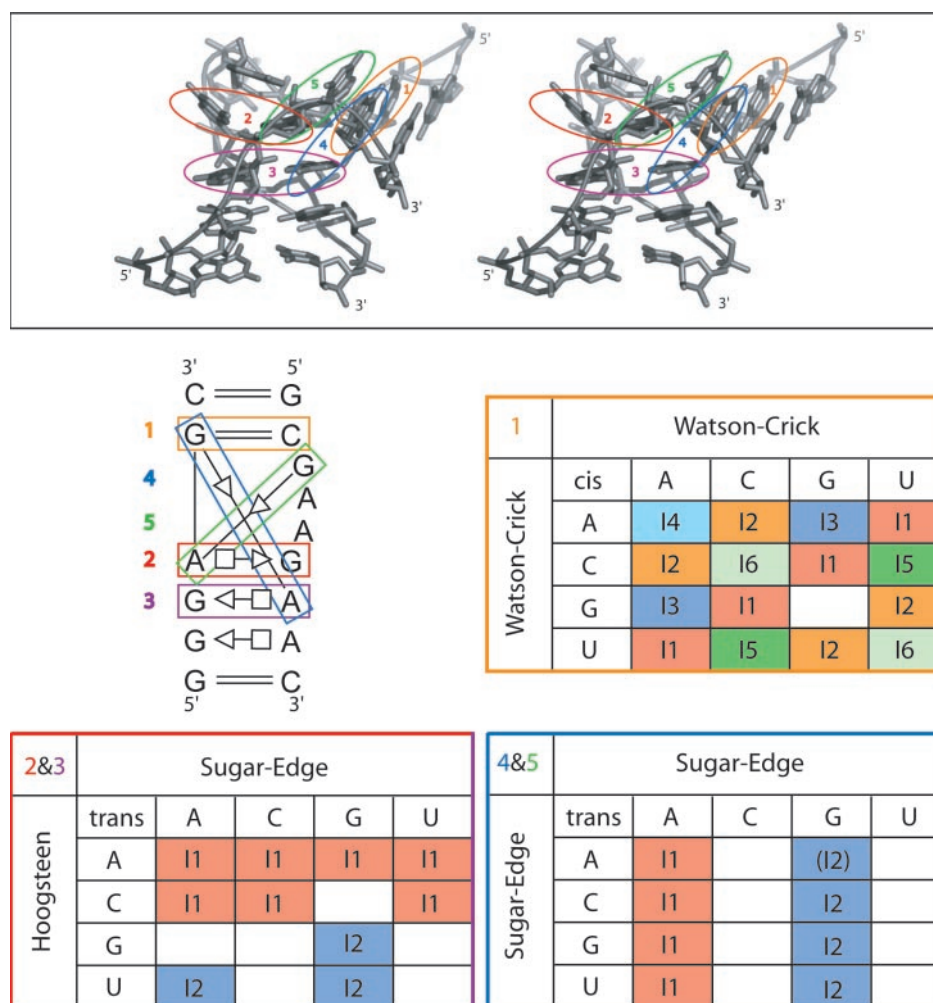
for each base pair and those that do not, i.e. those in which the motif is present and those in which it is not. For the latter, alternative motifs are considered and evaluated in the same manner (square box, lower left). Two examples, one for each motif, of the matrices before and after analysis are shown on Figures S1 and S2 in the Supplementary Material. The resulting sequence alignments can be obtained from the authors upon request and on the web site (<http://www-ibmc.u-strasbg.fr/upr9002/westhof/>). The secondary structures, with their attached Isostericity Matrices, for all motifs analyzed and discussed below are presented in the Supplementary Material (Figures S3 and S4).

## RESULTS

### Kink-turns

*Characteristic features.* Kink-turn motifs are recurrent internal loops that produce sharp bends (kinks) in RNA helices (18).

The bend occurs on the shallow/minor groove side and, consequently, brings together the shallow/minor grooves of the two supporting helices. One helix, the so-called ‘C-stem’ (for ‘canonical-stem’), comprises only WC base pairs, while the other so-called ‘NC-stem’ (for non-canonical-stem) is composed of non-WC base pairs, usually two tandem sheared (*trans*-H/SE) A/G base pairs that present the Watson–Crick and Sugar-Edges of conserved Adenosines for interaction with the shallow/minor groove of the C-stem. The interaction involves tandem *trans*-SE base pairs, as also observed for certain long-range RNA tertiary interactions, for example, the interaction of the tandem ‘sheared’ G/A motif in Helix 101 (H101) of *H.m.* 23S rRNA (G2874/A2883 and A2875/G2882) with the minor groove of H63 (C1786/G1806 and C1787/G1805) (12). With the inclusion of the flanking base pairs and the bases forming the non-WC base pairs, the K-turns comprise about 13 nt. Five base pairs characterize the motif and are found in almost all K-turns. These are numbered to facilitate the discussion as follows (see also Figure 2).



**Figure 2.** Stereographic view of a crystallographic structure of a typical Kink-turn (14,28) with its annotated secondary structure following the nomenclature for non-WC pairs (3). Each characteristic base pair is circled in the 3D diagram with colors corresponding to those in the 2D diagram. The same color code is used to frame the Isostericity Matrix attached to each base pair. Base pair 1 (BP1), colored orange, is *cis*-WC/WC; Base pair 2 (BP2), in red, is *trans*-H/SE; Base pair 3 (BP3), in purple, is *trans*-H/SE; Base pair 4 (BP4), in blue, is *trans*-SE/SE; Base pair 5, in green, is *trans*-SE/SE. In each Isostericity Matrix, the families of isosteric pairs (I1, I2, etc.) have an identical colored background. Parentheses indicate modeled interactions for the isosteric relationships not yet observed in high-resolution X-ray structures (4).

- Base pair 1 (shown in orange in Figure 2) is the last canonical *cis*-WC/WC base pair of the C-stem, flanking the internal loop. It is usually G=C.
- Base pairs 2 and 3 (in red and purple, respectively) are usually tandem *trans*-Hoogsteen/Sugar-Edge (*trans*-H/SE) and occur at the end of the NC-stem.
- Base pair 4 (in blue) is *trans*-Sugar-Edge/Sugar-Edge (*trans*-SE/SE) (type I A-minor interaction) and involves the adenine of base pair 3 (purple) and the first nucleotide of base pair 1 (orange).
- Base pair 5 (in green) is also *trans*-SE/SE and comprises the adenine of base pair 2 in the NC-stem interacting with the 5'-most unpaired base of the longer strand of the asymmetric loop.

The corresponding Isostericity Matrices are also shown in Figure 2. In the 2D schematic figures representing the motifs, colored frames are placed around each non-WC base pair to link them visually to the corresponding covariation tables and Isostericity Matrices. Thus, red and purple frames indicate *trans*-H/SE base pairs in the NC-stem and blue and green frames indicate the *trans*-SE/SE tertiary base pairs whereby the NC- and C-stems interact on their shallow/minor grooves. The boxes of each matrix are colored to indicate which base pairs belong to the same isosteric subgroups. Isosteric subgroups that are related by lateral shifts (as are the *cis*-WC and wobble pairs) are colored with similar colors.

A list of K-turns, together with composite K-turn motifs identified in the ribosome and other RNAs is given in Table 1. The K-turn motifs are shown schematically in Figure 3. Each K-turn in the ribosome is indicated by the molecule (16S or

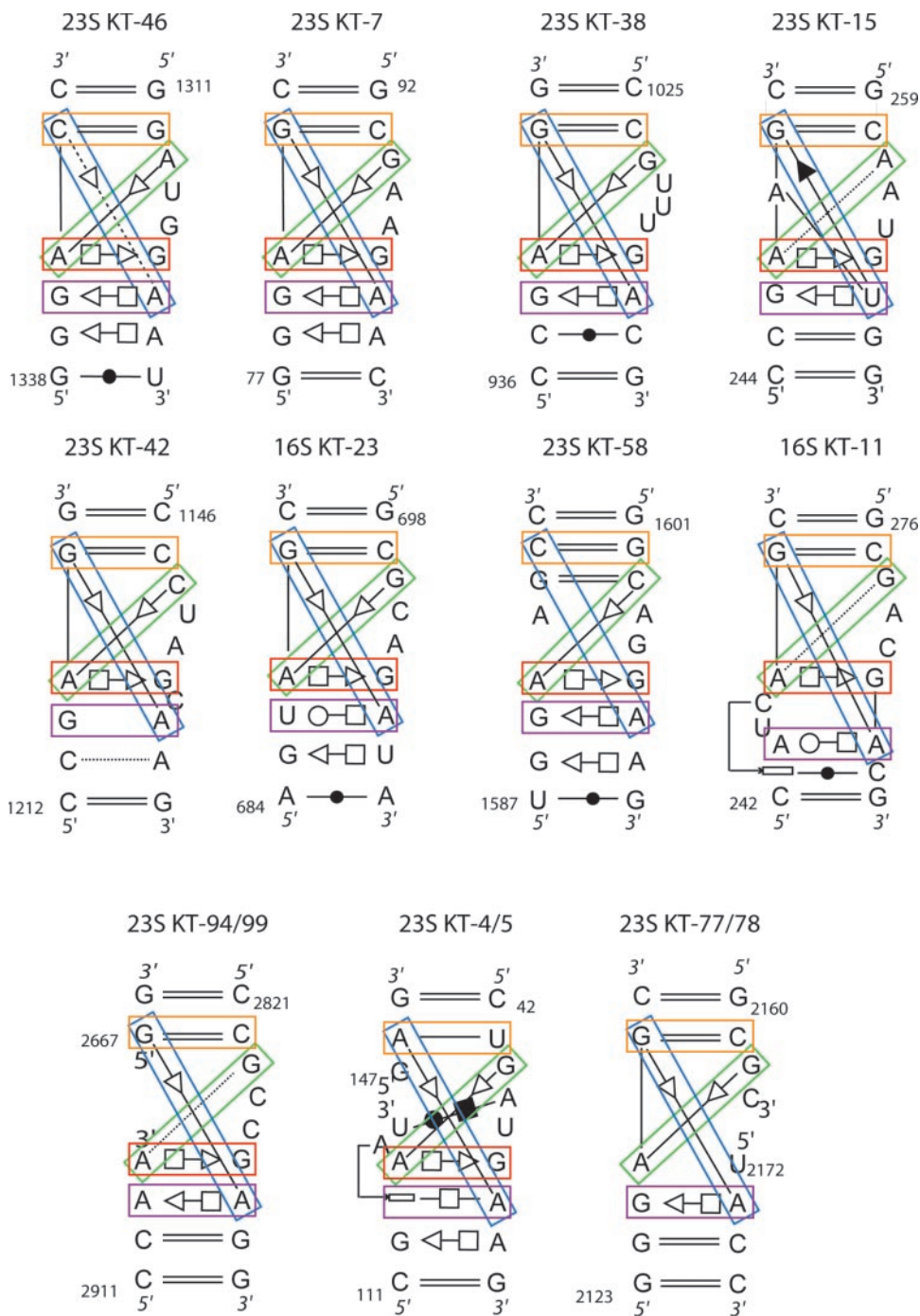
23S) and the helix (or nearest helices) in which it occurs. For example, 23S Kt-7 refers to the K-turn in Helix 7 of 23S rRNA and 16S Kt-11 refers to the K-turn in Helix 11 of 16S rRNA. For completeness, we list in Table 1 K-turns identified in other protein-RNA complexes besides the ribosomal subunits: the U4 snRNP (19), the box H/ACA pseudouridine synthases (20), the box C/D methylases (21) and the L1 mRNA autoregulatory element (22). These are not discussed further. We exclude also from this analysis the 'reverse K-turn', which kinks toward the deep/major groove instead of the shallow/minor groove (23). Although several examples of those reverse K-turns exist, they present, among the four non-WC pairs characteristic of K-turns, only a single invariant *trans*-H/SE G/A pair (equivalent to base pair 3 on Figure 2). Thus, the reverse K-turns constitute an independent RNA motif with specific constraints different from those of the standard K-turns.

*Structural variants.* As noted above, the first non-WC basepair of the NC-stem (the 'red' base pair) is present in almost every structurally known K-turn and is conserved with regard to geometry and even sequence (Figure 3). The one exception is the 'composite' K-turn 23S Kt-77/78, in which the conserved A2167 at this position is not paired. While the geometry of base pair 1 (red) is very conserved, the second (purple) base pair can be replaced by other, related base pair types, *trans*-WC/H (16S Kt-23 and 16S Kt-11) or *trans*-H (23S Kt-4/5). All of these base pair geometries have one feature in common: they present the conserved A in the shallow/minor groove to interact with the last base pair in the C-stem, forming Base

**Table 1.** List of the K-turns and C-motifs considered in the analysis and alignments

	Molecule	Nucleotides	PDB file	Features	Conservation
<b>Kink-turns</b>					
U4 Kt	U4	26–35/41–47	1E7K		—
16S Kt-11	<i>T.th.</i> 16S	242–247/277–284	1J5E	A243 inserted between C245 and A246 A243/A282 form <i>trans</i> WC/H pair	A, B, E
16S Kt-23	<i>T.th.</i> 16S	683–688/699–707	1J5E	U/A <i>trans</i> WC/H base pair	A, B, E
23S Kt-7	<i>H.m.</i> 23S	77–81/93–100	1S72		A, B
23S Kt-7	<i>D.r.</i> 23S	80–84/96–103	1LNR		
23S Kt-15	<i>H.m.</i> 23S	245–249/259–265	1S72	U265/G249 form P-interaction	A
23S Kt-38	<i>H.m.</i> 23S	936–940/1026–1034	1S72	Inserted base in long strand	A
23S Kt-42	<i>H.m.</i> 23S	1147–1155/1212–1216	1S72	Inserted C1153 pairs with G2786	A, B, E
23S Kt-42	<i>D.r.</i> 23S	1054–1062/1119–1123	1LNR	Inserted C1060 pairs with G2731	
23S Kt-46	<i>H.m.</i> 23S	1312–1319/1338–1342	1S72	3rd NC base is sheared A/G	A, B, E
23S Kt-46	<i>D.r.</i> 23S	1221–1228/1247–1251	1LNR	3rd NC base is sheared A/G	
23S Kt-58	<i>H.m.</i> 23S	1587–1592/1602–1608	1S72	Two inserted bases in short strand	A
23S Kt-4/5	<i>H.m.</i> 23S	111–115/147–149 42–53	1S72	Composite motif—interrupted in short strand	A, B, E
23S Kt-4/5	<i>D.r.</i> 23S	114–118/154–156 45–53	1LNR	Composite motif—interrupted in short strand	
23S Kt-77/78	<i>T.th.</i> 23S	2124–2127/2161–2163 2172–2828	1MZP	Composite motif—interrupted in long strand	A, B, E
23S Kt-94/99	<i>H.m.</i> 23S	2912–2914/2667–2668 2821–2828	1S72	Composite motif—interrupted in short strand	A
<b>C-loops</b>					
D2 domain	<i>E.coli</i> TRS mRNA	74–77/95–100	1KOG	Canonical	
23S C38	<i>H.m.</i> 23S	958–964/1004/1009	1S72	Extruded AA interact with Helix 81	A, B, E
23S C38	<i>D.r.</i> 23S	877–881/921–924	1LNR	Extruded AA interact with Helix 81	A, B, E
23S C50	<i>H.m.</i> 23S	1425–1428/1437–1439	1S72		A, B, E
23S C50	<i>D.r.</i> 23S	1332–1336/1344–1346	1LNR		A, B, E
23S C96	<i>H.m.</i> 23S	2717–2721/2761–2763	1S72		A, B, E
23S C96	<i>D.r.</i> 23S	2659–2663/2704–2706	1LNR		A, B, E
16S C15	<i>T.th.</i> 16S	371–376/387–390	1J5E		A, B, E

Their positions in the molecule is indicated with the PDB number corresponding to the crystal structure and a short description of the structural features.



**Figure 3.** Annotated secondary structures of Kink-turn motifs from crystal structures, comparing structural variants to the typical Kink-turn, exemplified by KT-7 from archaeal 23S rRNA. Each characteristic base pair is framed in a different color: The last base pair of the C-stem in orange (base pair 1), the two *trans*-H/SE base pairs of the NC-stem, base pair 2 in red and base pair 3 in purple, and the two *trans*-SE/SE, base pair 4 in blue and base pair 5 in green. Each tertiary interaction is represented by a unique symbol indicating the interacting edges of the bases and whether the pair is *cis* or *trans* (3).

pair 4 (blue). Even more variation is displayed by the third base pair of the NC-stem, which does not participate directly in K-turn interactions. It is therefore not discussed any further.

Besides variations in base-pairing geometry, K-turn structures exhibit variations in the number of nucleotides in the internal loop. Thus, *H.m.* 23S Kt-38 has one extra nucleotide in the longer strand, which bulges out just above the NC-stem.

This is the most common point of insertion. Bases inserted here have very little effect on the 3D structure of the motif. Insertions also occur in the shorter strand, e.g. 23S Kt-15 (A248), 23S Kt-58 (A1591 and G1592) and 16S Kt-11 (U244 and C245). Unlike the insertions in the longer strand, these usually participate in additional base pairs, as shown in Figure 3.

**Composite K-turns.** Just as for sarcin/ricin loops (24), K-turns can exist as composite motifs. Composite motifs have the same (or very similar) base-pairing and base-stacking interactions but more complex strand topologies. Three composite K-turns have been identified, all of them in 23S rRNA. One occurs in the complex junction in Domain I, between Helix 4 and Helix 5. This K-turn has a discontinuity in the shorter strand of the internal loop of the canonical motif. The identical motif is observed in the *H.m.* and *D.r.* 23S crystal structures and the break in the shorter strand occurs at equivalent positions, between the nucleotides of the NC- and C-stems (after U115 in the *H.m.* or U118 in the *D.r.*). The composite K-turn in the L1 binding region of Domain V, occurs between Helices 77 and 78. Unlike the other composite K-turns, a discontinuity occurs in the longer strand. This region, which is not well resolved in the complete 50S structures, has been crystallized as an RNA fragment corresponding to the *T.th.* 23S sequence, and includes all of Helix 77 and parts of Helices 76 and 78 as well as the loop regions forming the Kink-turn (22). The discontinuity occurs between C2163 and U2172 in the *T.th.* sequence. The third composite motif occurs in Domain VI of *H.m.* 23S between Helices 94 and 99. Here, the discontinuity occurs in the shorter strand of the asymmetric loop, between A2914 and G2667. Interestingly, G2667 is located 3' to A2914 whereas in all other K-turns the corresponding nucleotides are oriented with the opposite polarity.

**Isostericity Matrix analysis.** Covariation tables were calculated for the characteristic base pairs of each K-turn from the current alignments and the covariation tables were compared to the corresponding Isostericity Matrices for the base pair geometry identified in the crystal structures (Figure 2). Figure 4 shows the annotated secondary structures for a typical K-turn, Kt-46, for representative sequences from Archaea (*H.m.*), Bacteria (*D.r.*) and Eukarya (*Saccharomyces Cerevisiae*, *S.c.*). Crystal structures, available for the first two, show that the K-turn is present in both. The secondary structure and Isostericity Matrix for the *S.c.* motif indicates the K-turn is also conserved in this sequence. The sequence variations for each of the five characteristic base pairs of the K-turn are shown in the tables in Figure 4. To facilitate the comparison, the observed values are presented in the tables in the same orientation as the published Isostericity Matrices, shown in Figure 2 (4). The three numbers given in each box indicate the percentage of archaeal, bacterial and eukaryal sequences in the alignments having the corresponding base pair. These covariation tables are the result of iterations carried out as shown in Figure 1.

The sequence variations for the first base pair (orange) are typical of WC base pairs. All four WC base pairs occur in one or more phylogenetic domain. The matrices for the *trans* H/SE base pairs, the 'red' and 'purple' base pairs, show almost exclusively a subset of isosteric group I<sub>1</sub> from the Isostericity Matrix for this base-pairing geometry. The exceptions occur in Eukarya and include a small fraction of sequences with disallowed base pairs (C/G) for which it appears that the K-turn motif is not present. The small number of eukaryal sequences for the 28S rRNA precludes firm conclusions on those instances. Interestingly, for a significant percentage of Eukarya, the 'purple' base pair is C/A, which is isosteric

to A/G. The *trans*-SE/SE base pairs 4 and 5 (green and blue) involve the Hoogsteen base of the red and purple, which is usually A and more rarely C, interacting with bases in the C-stem, which can be A, C, G or U, as shown in the covariation matrices. Again only a subset of the sequence variations permitted by isostericity considerations is observed for each base pair, when these are considered in isolation.

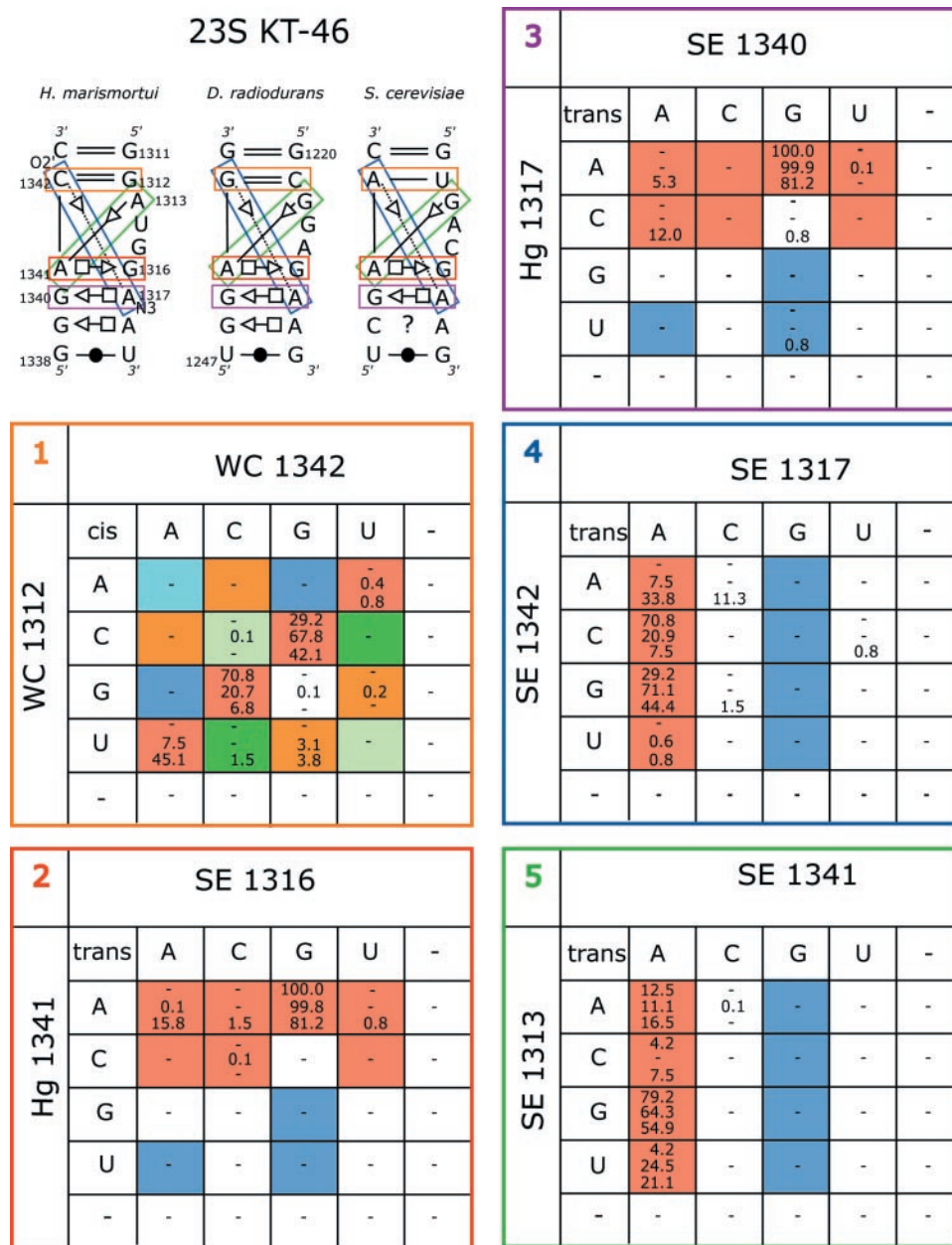
Thus, one observes that unsuspected concerted base changes maintaining the order and the geometrical arrangements of the base pairs constituting the motif occur. The observed sequence variations show that the pairings follow the corresponding Isostericity Matrix and, preferably, the same isosteric subgroup. For some types of interactions only a sub-class of pairings can occur. For example, in Figure 2 the matrix of *trans*-SE/SE pairing, Bp 5 and Bp 6, show that only A and G can serve as receptor of 2'-OH of the second nucleotide implicated in the interaction.

**Sequence signatures.** We have applied the procedure of Figure 1 to each K-turn in the ribosome (see also Supplementary Material S3). These covariation matrices are the final ones calculated at the conclusion of the realignment process. The matrices for individual K-turn motifs are combined in Figure 5 to provide aggregated statistics for K-turns. To construct the matrices of Figure 5, counts were accumulated only from sequences for which it was concluded that the corresponding base pair is present. The resulting matrices define the sequence signature for K-turns at the level of base pairs. It is apparent that for each non-WC base pair, the sequence signature includes a subset of the possible isosteric base pairs expected from the Isostericity Matrices. This indicates the presence of other constraints, including stacking interactions and tertiary interactions involving one or both bases of a pair (see Discussion).

### C-loop motifs

**Characteristic features.** Like the K-turn, the C-loop motif is an asymmetric internal loop (Figure 6). Two bases in the longer strand form non-WC base pairs with bases in the shorter strand. These bases belong to the flanking canonical base pairs. A third base in the longer strand, usually an A, can and often does interact with a third base pair, one removed from the internal loop (Figure 7). Indeed, in two cases this base forms a *cis*-WC/SE interaction. There are three C motifs in 23SrRNA (14), one in 16SrRNA (13,25) and one in the mRNA of threonine synthetase (26).

All C-loops in the ribosome are found in hairpin stem-loop structures that engage in tertiary interactions involving the hairpin loop. The C-loop itself increases the helical twist of the stem between the two WC base pairs flanking the motif. This probably optimizes the geometry of the interaction mediated by the hairpin loop distal to the C-loop. However, the large helical twist between the flanking base pairs (~90°) greatly diminishes the stacking between them. Three nucleotides in the longer strand of the C-loop interact with the base pairs flanking the motif by stacking and forming non-WC base pairs. The first base in the loop is usually a C (and hence 'C-loop'), and it stacks below the preceding flanking base pair where it forms a *trans*-WC/H interaction in the major/deep groove with the second base pair flanking the motif, thus reinforcing the large helical twist between the flanking base



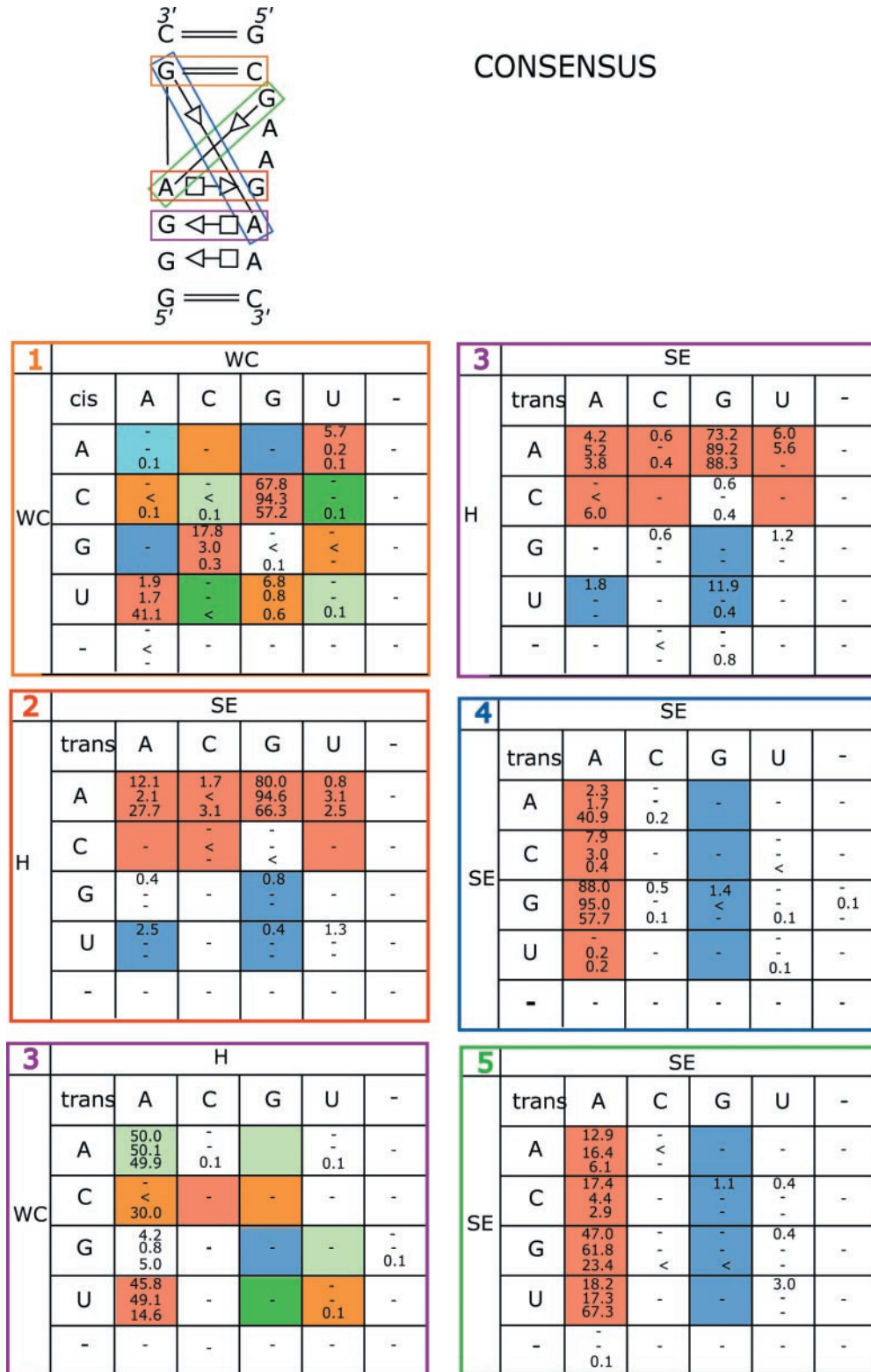
**Figure 4.** Upper panel: Annotated secondary structures of the conserved 23S rRNA Kt-46 Kink-turn motif. Lower panel: Isostericity Matrix analysis of characteristic base pairs of Kt-46. In each box, the percentage of observed sequences for each base pair is given for Archaea (A), Bacteria (B) and Eukarya (E) (upper, middle and lower).

pairs. The third base of the longer strand stack above the second flanking base pair and forms *cis* WC/SE base pairs with the base pair above it. The shorter strand usually has one or two unpaired nucleotides that are invariably extruded. In all C-loops in the ribosome, the extruded bases engage in tertiary interactions, some of which are long range and involve RNA elements directly implicated in ribosome function. A schematic figure of a typical C-loop motif is shown in Figure 6 with the four characteristic base pairs of the motif framed in colored boxes to visually link them to the corresponding Isostericity Matrices and covariation tables. Schematic diagrams of C-loops observed in crystal structures are

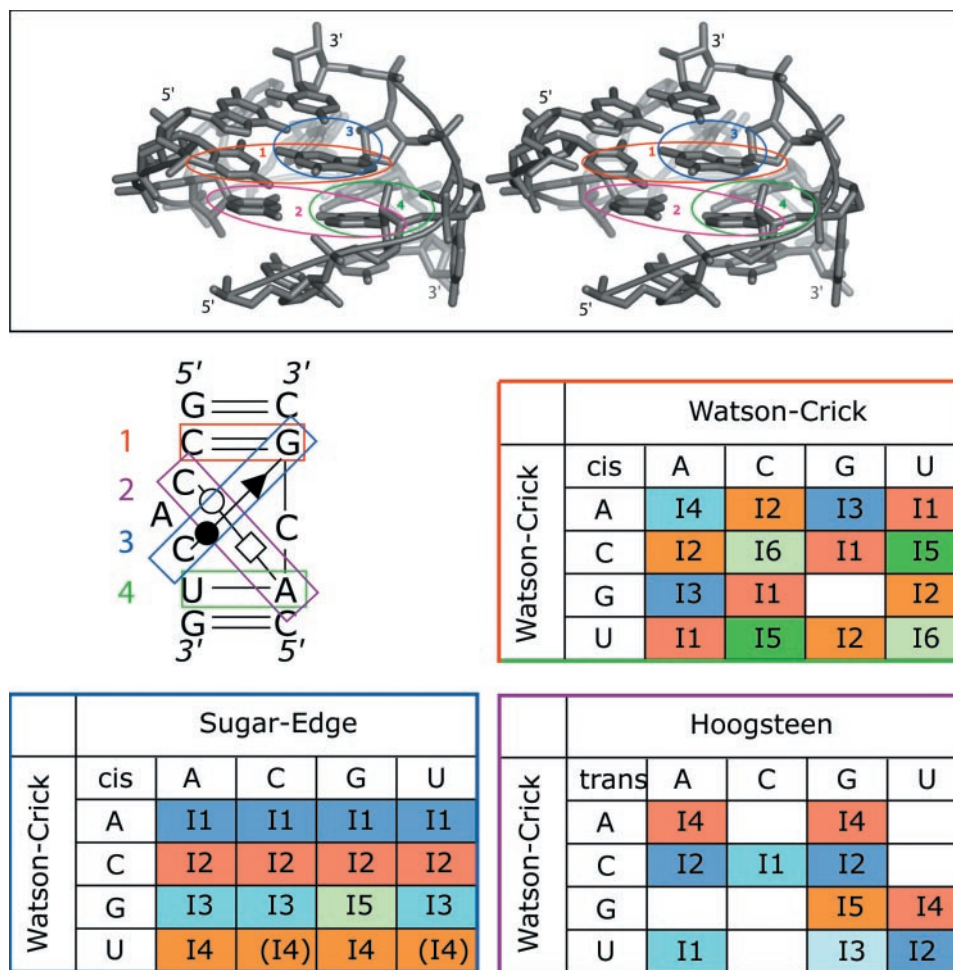
shown in Figure 7. These diagrams show that C-loops can differ in the number of extruded bases in the shorter strand as well as the number of bases in the longer strand. Thus, 23S C38 has two extruded bases in the shorter strand, C50 has one, and C96, one or two, while in 16S C15 the extruded base occurs after the flanking base pair.

*Isostericity Matrix analysis.* For each of the C-loop motifs identified in the ribosome, covariation matrices for each of the base pairs were determined from the current alignments and compared to the corresponding Isostericity Matrices (see Supplementary Material S4). The analysis for a typical





**Figure 5.** Aggregated tables of sequence variations for each of the characteristic base pairs of conserved Kink-turns. The data for the conserved Kink-turns are given. These serve to define the sequence signatures for Kink-turn motifs. For pair 1 (orange *cis*-WC/WC), the number of sequences were A = 264, B = 5625, E = 10 846. For Pair 2 (red *trans*-H/SE), A = 240, B = 4820, E = 10 713; KT-77/78 was excluded, since pair 2 is absent. For Pair 3 (purple *trans*-H/SE), A = 168, B = 2415 and E = 266. The following K-turns were excluded: KT-4/5 where *trans*-H/SE is replaced by *trans*-H/H base pair; KT-42 where *trans*-H/SE is absent in X-ray structure although bases are present; KT-23 where *trans*-H/SE is replaced by *trans*-WC/H base pair (see below); KT-11 where *trans*-H/SE is replaced by *trans*-WC/H base pair (see below). For Pair 3 (purple *trans*-WC/H), A = 48, B = 1600 and E = 10 377. KT-11 and KT-23 were taken into account. For Pair 4 (blue *trans*-SE/SE), A = 216, B = 5625 and E = 10 846. The following were excluded: KT-58 where pair 4 is absent in the structure; KT-15 where *trans*-SE/SE is replaced by a *cis*-SE/SE base pair. For Pair 5 (green *trans*-SE/SE), A = 264, B = 5625 and E = 10 846.

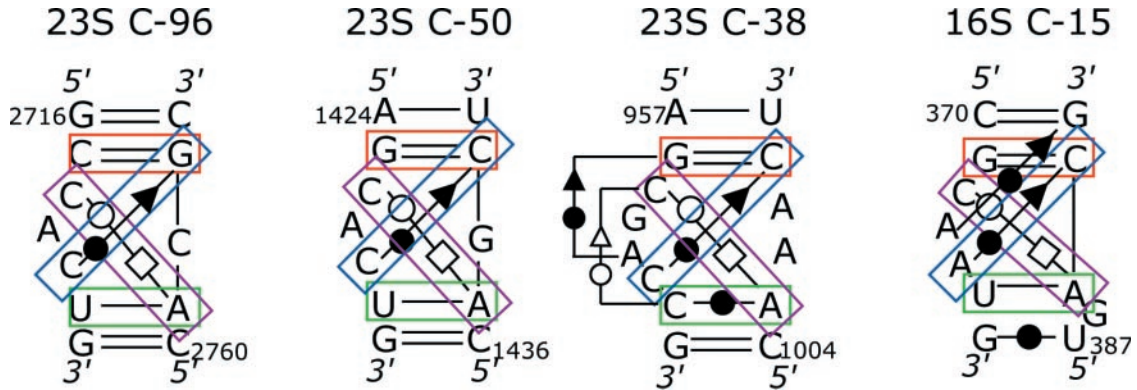


**Figure 6.** Stereographic view of a crystallographic structure of a typical C-loop (14,28) with its annotated secondary structure following the nomenclature for non-WC pairs (3). Each characteristic base pair is circled in the 3D diagram with colors corresponding to those in the 2D diagram. The same color code is used to frame the Isostericity Matrix attached to each base pair. Base pair 1 (BP1) is colored red: *cis*-WC/WC; Base pair 2 (BP2) in purple: *trans*-WC/H; Base pair 3 (BP3) in blue: *cis*-WC/SE; Base pair 4 (BP4) in green: *cis*-WC/WC. In each Isostericity Matrix, the families of isosteric pairs (I1, I2, etc.) have an identical colored background. Parentheses indicate modeled interactions for the isosteric relationships not yet observed in high-resolution X-ray structures (4).

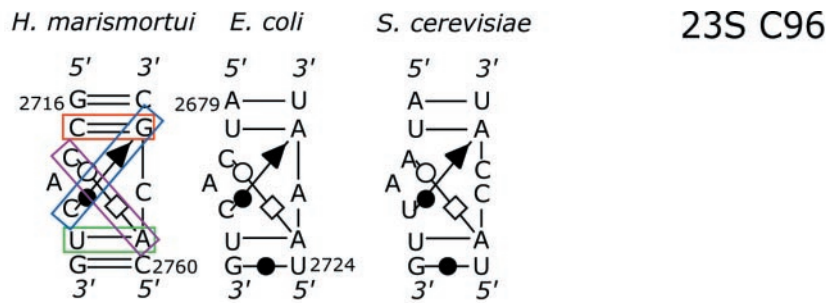
C-loop motif, 23S C96, is shown in Figure 8. The first *cis*-WC/WC pair (red box) of the motif, G2763=C2717 in the *H.m.* sequence, is almost exclusively G=C in Archaea and Bacteria, but in Eukarya A-U and A-C are also represented. The second *cis*-WC/WC base pair (green box) of the motif is almost exclusively A/U in all three phylogenetic groups. They otherwise follow the diagonal repartition typical of WC pairs. Base pair 2 (purple), the *trans*-WC/H pair, is almost exclusively C/A as in the crystal structure, but some isosteric C/G occur in Archaea, Bacteria and Eukarya (isosteric subgroup 1) while A/A which has a distinctly longer C1'-C1' distance and is not isosteric to the other base pairs (isosteric subgroup 2) also occurs in Eukarya. No crystal structure of a C-loop with the *trans*-WC/H A/A pair is known at this time. Regarding the *cis*-WC/SE pair of C96 (BP 3, blue), only C/G is observed for Archaea while the isosteric C/A base pair also occurs in Bacteria. U/A occurs in most eukaryal sequences but U/G and C/G are also represented too. C/C is observed for BP3 in the crystal structures C38 and C50, while C/G is observed in C96 and A/C is observed in 16S C15 in the crystal structure of the *T.th.* 16S rRNA. The *cis*-WC/SE A/N pairs form an isosteric

subgroup with a longer C1'-C1' distance than the C/N pairs. The presence of A/C therefore modifies the geometry of the 16S C15 motif noticeably, compared to the other C-loops, but not so radically as to disrupt the other base pairs of the motif (see below).

About half the bacterial sequences for C50 have G at the position corresponding to C1426, implying G1426/C1437 (*H.m.* numbering) for the *trans*-WC/H base pair, a pairing that cannot exist. In these sequences, the second flanking WC base pair of the motif (BP4) is invariably G1429/C1437 (*H.m.* numbering) and the *cis*-WC/SE (BP3) is always A1428/C1439. While the *D.r.* 23S sequence provides an example of this variant and the crystal structure shows that the G corresponding to G1426 is in the deep/major groove, the structure is not resolved well enough to determine the nature of the interactions. Surprising variations occur in Eukarya for C50 and C96, which cannot be properly assessed owing to the lack of structures and the limited number of sequences. Most Eukarya (65%) have G/G for the *trans*-WC/H base pair of C50 (corresponding to C1426/A1437 in *H.m.*), which is not observed in other C-loops, but is nearly isosteric to the



**Figure 7.** Annotated secondary structures for C-loop motifs from crystal structures, comparing structural variants with a typical C-loop, exemplified by C15 from helix 15 of 16S rRNA. The four characteristic interactions of this motif are shown: canonical basepairs 1 in red and 4 in green, *trans*-WC/H base pair 2 in purple and *cis*-WC/SE base pair 3 in blue.



1		WC 2717					
WC 2763	cis	A	C	G	U	-	
	A	-	20.4	-	17.4	41.6	
	C	-	-	0.4	-	-	
	G	-	100.0	82.2	38.1	-	
	U	-	-	-	-	-	
	-	-	-	-	-	-	

2		H 2761					
WC 2718	trans	A	C	G	U	-	
	A	-	-	-	-	-	
	C	78.8	-	4.2	0.3	1.8	
	G	95.8	-	0.9	-	-	
	U	99.8	-	-	-	-	
	-	17.7	-	-	-	-	

3		SE 2763					
WC 2720	cis	A	C	G	U	-	
	A	-	-	2.4	3.5	-	
	C	17.4	0.4	79.9	10.6	-	
	G	2.7	-	-	-	-	
	U	0.9	-	0.9	-	-	
	-	58.4	-	23.0	-	-	

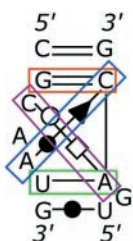
  

4		WC 2721					
WC 2761	cis	A	C	G	U	-	
	A	-	0.2	-	95.8	99.5	
	C	0.9	-	-	96.5	-	
	G	-	4.2	0.2	2.7	-	
	U	-	-	-	-	-	
	-	-	-	-	-	-	

**Figure 8.** Upper panel: Annotated secondary structures of 16S C96, for representative archaeal, bacterial and eukaryal sequences. Lower panel: Isostericity Matrix analysis of characteristic base pairs of 16S C96. In each box, the percentage of observed sequences for each base pair is given for Archaea, Bacteria and Eukarya (upper, middle and lower).

subgroup that includes A/A, A/G and G/U. Interestingly, some Eukarya have A/G (2.3%) or G/U (0.8%). In 23S C96, most Eukarya have A/A for the *trans*-WC/H while a small percentage show G/G (0.9%). These sequences have U/A or U/G for the *cis*-WC/SE pair. U is not observed widely in BP3 of other C-loops, so this combination may signal a further variant of the motif.

**Sequence signatures.** The aggregated covariation matrices for C-loops are shown in Figure 9. Except for base pair 2, the Isostericity Matrices are obeyed. Since base pair 2, the *trans*-WC/H pair, is central to the definition of a C-motif, those related motifs are designated 'C-like motifs'. Some C-motifs show additional base pairs (e.g. C15 and C38). The additional base pairs behave differently in their corresponding Isostericity Matrices (Figure 10). The additional base pair in C15, a *cis*-WC/SE pair, follows the Isostericity Matrix. On the contrary, the additional base pair in C38, a *trans*-WC/SE pair, occupies three subgroups in the matrix, all compatible, but not isosteric. The latter pair is, thus, rather variable and opportunistic, and forms in response to the environment.



**C-like loops.** The 2D diagrams of two C-like motifs are shown in Figure 11. In those C-like motifs, the high twist angle between the flanking base pairs is maintained but base pair 2 is absent. The usual C residue is replaced by other nucleotides and, in the 3D structures, it points away from the loop towards the exterior, sometimes forming an additional pair. Base pair 3, the *cis*-WC/SE, is present and, interestingly, a supplementary pair appears, also *cis*-WC/SE, equivalent to the one which occasionally, but in a constrained fashion, appears in the C-motifs (e.g. 16S C15).

## DISCUSSION AND CONCLUSIONS

The systematic comparisons between structures, sequences and Isostericity Matrices have shown that deviations of covariation matrices from the corresponding Isostericity Matrices indicate either incorrect sequence alignment or change of motifs between species or phylogenetic groups. Through the iterative application of the procedures in Figure 1, it was possible to realign some sequences to correct several

## Consensus

1		WC				
	cis	A	C	G	U	-
WC	A	<	0.9	0.1	6.8 2.7	-
	C	<	<	63.5 52.6 2.7	<	-
	G	<	27.0 26.7 92.7	-	0.1	-
	U	9.4 13.8 0.3	0.1	<	<	-
	-	-	-	-	-	-

2		H				
	trans	A	C	G	U	-
WC	A	< 1.7	-	-	-	-
	C	94.8 75.5 94.7	1.0 1.0 -	2.1 9.8 0.7	-	-
	G	-	10.0 0.4	< 1.6	<	-
	U	1.0 0.2	-	0.3	1.0 3.7	<
	-	-	-	-	-	-

3		SE					
	cis	A	C	G	U	-	
WC	A	-	36.5 27.4 2.7	2.1 6.7 69.3	6.3 13.8 0.3	-	
	C	-	4.4 0.3	27.0 25.2 0.1	25.0 20.0 22.0	3.1 <	-
	G	<	<	<	0.1	-	-
	U	-	-	-	-	-	-
	-	-	1.2	-	1.4	-	-

4		WC					
	cis	A	C	G	U	-	
WC	A	-	6.3 0.1 0.1	1.0	88.5 75.4 96.8	-	
	C	-	-	1.0 11.0 0.4	-	-	
	G	-	2.0 9.8 2.4	-	-	0.3	-
	U	1.0 3.7	-	<	-	-	-
	-	-	-	<	-	-	-

Figure 9. Aggregated tables of sequence variations for each of the characteristic base pairs of conserved C-loops.

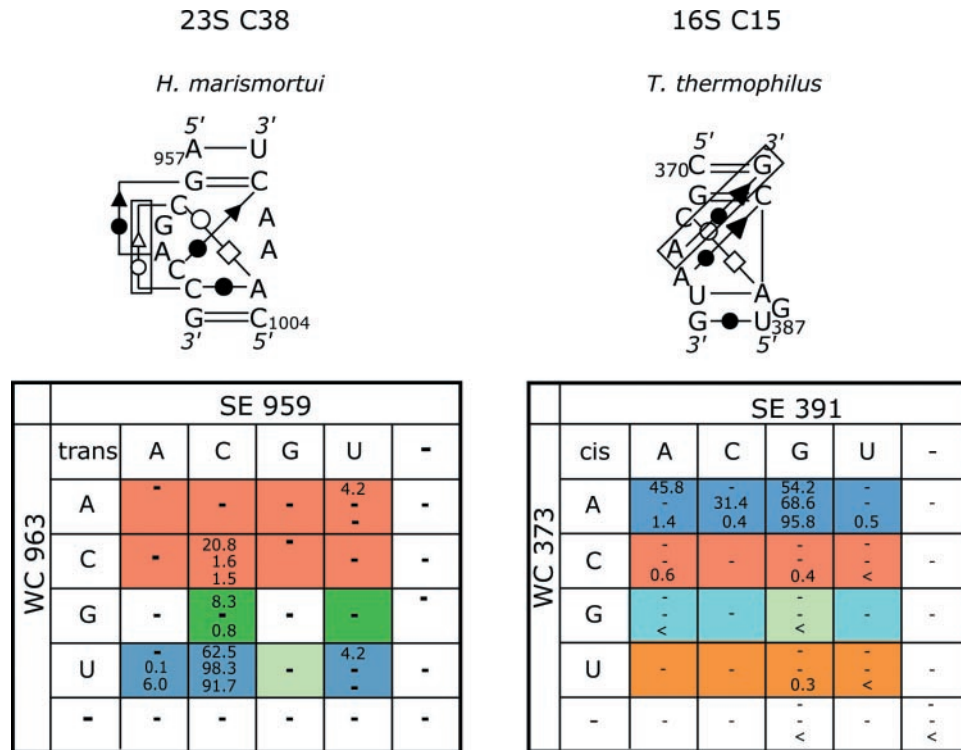


Figure 10. Covariation tables for two additional pairs present in the C-loop motifs 23S C38 and 16S C15.

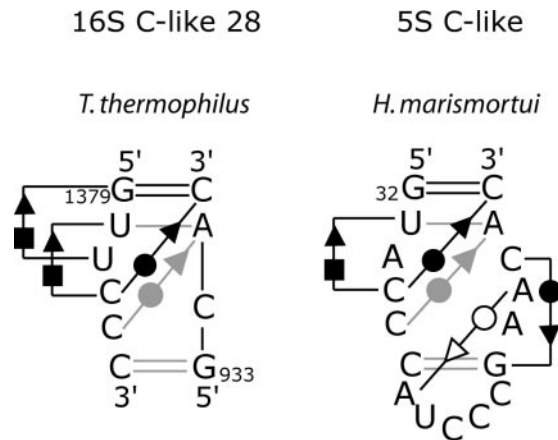


Figure 11. Annotated secondary structures for two C-like motifs, the C-like 28 from the 16S rRNA and loop C in the 5S rRNA. Although the overall 3D fold is maintained, the distinctive C of the *trans*-WC/H pair is absent in those motifs. An additional pair is present (black). In grey are shown the usual pairs of the C-motifs.

discrepancies between the covariation and Isostericity Matrices. When realignment did not improve the agreement between observed covariations and Isostericity Matrices, the particular motif was identified either as a variant of the structurally characterized parent motif or as absent and replaced by a structurally distinct motif in that sequence or phylogenetic group of sequences. Figures 5 and 9 provide data for constructing sequence signatures for K-turn and C-loop motifs. In order to best take into account the covariations observed for some non-WC base pairs, isosteric subgroups should be grouped into

structurally similar groups that are related by lateral shifts. This is exactly what is done for wobble and WC pairs, which are not identical, but which can co-vary without disrupting or distorting regular A-type helices. The comparison between the data in Figures 5 and 9 with the respective Isostericity Matrices shows that generally a subset of possible base pairs identified in the Isostericity Matrices actually occur in the sequence signatures. This observation shows the central importance of geometrical conservation in the rules of transformation during RNA evolution.

Thus, the variations in RNA motifs are constrained at different levels: (i) The base changes follow the Isostericity Matrices corresponding to the nature of the non-WC pair. There is a strong tendency to vary within the same isosteric subgroup within a given matrix. (ii) For some non-WC pairs, a closely related pair, not strictly isosteric, can occur. (iii) Within the motif, the length of some unpaired segment of nucleotides can vary, usually at identical insertion sites. (iv) The loss of one characteristic non-WC pair leads to structural variants of the motifs.

Superimposed on those general constraints, other structural constraints specific to the motif type or due to the particular environment come into play. For example, each *trans*-H/SE base pair in the NC-stem of the K-turn exhibits A as the Hoogsteen base and almost never C. In other motifs where *trans*-H/SE base pairs occur, C is often observed to substitute for A. The exclusive occurrence of A in the *trans*-H/SE base pairs of K-turn motifs can be understood by the added constraint due to the observation that these A residues mediate the shallow/minor-groove interactions with the C-stem, crucial for the folding of K-turn motifs, and for which adenines are the best suited (18,27).

Our analysis identifies sequences or groups of sequences in which K-turns are not conserved. These are given at the level of phylogenetic groups in Table 1. Some K-turn or C-loop motifs do not appear to be conserved in some sequences even though they appear to be present in most other sequences in their phylogenetic groups. K-turns are not the only motifs producing sharp bends on the shallow/minor groove sides of helices. Motifs that produce similar effects occur in Helix 68 of *H.m.* 23S rRNA, Helix 41 of *T.th.* 16S rRNA, and between P5 and P5a in certain Group I introns. Moreover, K-turns can occur as complex composite motifs that are difficult to identify in the absence of 3D structure. Thus, it is not surprising that for some sequences and even whole phylogenetic domains, K-turns are not conserved. Our analysis, summarized in Table 1, shows that Kink-turns are conserved across the three major phylogenetic domains for Kt-42 and Kt-46 in 23S rRNA and for Kt-11 and Kt-23 in 16S rRNA. The Kt-7 motif is conserved in Archaea and Bacteria, while the Kt-15, K-38 and Kt-58 motifs are only conserved in Archaea. Similarly, except for C50, the C-loop motifs in 16S and 23S rRNA appear to be conserved in all three phylogenetic domains. The composite Kt-4/5 and Kt-77/78 motifs appear to be conserved in all three domains, but Kt-94/99 does not appear to be conserved. These examples mirror the composite sarcin/ricin motifs (24) and emphasize the role of structural constraints underlying RNA motifs beyond the conservation of the nature of the base pairs, like stacking, hydrogen bonding, and 3'-5' covalent linkages.

In summary, we have shown that Isostericity Matrices provide tools for productive iteration of multiple sequence alignment of RNA sequences. For this process to be successful, several goals must be met simultaneously: (i) corresponding motifs must be identified in homologous sequences and it must be determined which are structurally conserved. (ii) Insertions and deletions must be correctly positioned and bases forming structurally conserved base pairs characteristic of the motif must be aligned. (iii) Sequences forming alternative motifs must be identified, segregated from the other sequences, and aligned separately. We have demonstrated the use of Isostericity Matrix to iteratively refine rRNA sequence alignments for the biologically significant and recurrent K-turn and C-loop motifs, and identified sequences or phylogenetic groups for which the motifs were not conserved. We have constructed sequence signatures that can be used to search for these motifs in sequences. This body of data contributes to a programmed process of automatic identification, annotation and prediction of RNA motifs in sequence alignments.

The key conclusion that emerges from this work is that the structural content, quality and pertinence of RNA sequence alignments depends crucially on a careful analysis of RNA motifs. Together with the use of Isostericity Matrices of non-WC base pairs and the definition of RNA motifs as ordered arrays of non-WC base pairs under constraints, the rich 3D information hidden within sequences can be extracted. As a corollary, attempts to align structurally non-equivalent motifs at the nucleotide level fail and give nonsensical results.

## SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

## ACKNOWLEDGEMENTS

N.B.L. is supported by NIH grant 3R15-GM55898. E.W. thanks the Institut Universitaire de France for support. Funding to pay the Open Access publication charges for this article was provided by the Centre National de la Recherche Scientifique (France).

*Conflict of interest statement.* None declared.

## REFERENCES

- Westhof,E. and Massire,C. (2004) Structural biology. Evolution of RNA architecture. *Science*, **306**, 62–63.
- Krasilnikov,A.S., Xiao,Y., Pan,T. and Mondragon,A. (2004) Basis for structural diversity in homologous RNAs. *Science*, **306**, 104–107.
- Leontis,N.B. and Westhof,E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA*, **7**, 499–512.
- Leontis,N.B., Stombaugh,J. and Westhof,E. (2002) The non-Watson–Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.*, **30**, 3497–3531.
- Wadley,L.M. and Pyle,A.M. (2004) The identification of novel RNA structural motifs using COMPADRES: an automated approach to structural discovery. *Nucleic Acids Res.*, **32**, 6650–6659.
- Hershkovitz,E., Tannenbaum,E., Howerton,S.B., Sheth,A., Tannenbaum,A. and Williams,L.D. (2003) Automated identification of RNA conformational motifs: theory and application to the HM LSU 23S rRNA. *Nucleic Acids Res.*, **31**, 6249–6257.
- Goody,T.A., Melcher,S.E., Norman,D.G. and Lilley,D.M. (2004) The kink-turn motif in RNA is dimorphic, and metal ion-dependent. *RNA*, **10**, 254–264.
- Matsumura,S., Ikawa,Y. and Inoue,T. (2003) Biochemical characterization of the kink-turn RNA motif. *Nucleic Acids Res.*, **31**, 5544–5551.
- Cojocaru,V., Nottrott,S., Klement,R. and Jovin,T.M. (2005) The snRNP 15.5K protein folds its cognate K-turn RNA: a combined theoretical and biochemical study. *RNA*, **11**, 197–209.
- Razga,F., Spackova,N., Reblova,K., Koca,J., Leontis,N.B. and Sponer,J. (2004) Ribosomal RNA kink-turn motif—a flexible molecular hinge. *J. Biomol. Struct. Dyn.*, **22**, 183–194.
- Sundaralingam,M. and Arora,S.K. (1969) Stereochemistry of nucleic acids and their constituents. IX. The conformation of the antibiotic puromycin dihydrochloride pentahydrate. *Proc. Natl Acad. Sci. USA*, **64**, 1021–1026.
- Leontis,N.B. and Westhof,E. (2003) Analysis of RNA motifs. *Curr. Opin. Struct. Biol.*, **13**, 300–308.
- Clemons,W.M.Jr, Brodersen,D.E., McCutcheon,J.P., May,J.L., Carter,A.P., Morgan-Warren,R.J., Wimberly,B.T. and Ramakrishnan,V. (2001) Crystal structure of the 30 S ribosomal subunit from *Thermus thermophilus*: purification, crystallization and structure determination. *J. Mol. Biol.*, **310**, 827–843.
- Ban,N., Nissen,P., Hansen,J., Moore,P.B. and Steitz,T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905–920.
- Thompson,J.D., Gibson,T.J., Plewniak,F., Jeanmougin,F. and Higgins,D.G. (1997) The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, **25**, 4876–4882.
- Wuyts,J., Perriere,G. and Van De Peer,Y. (2004) The European ribosomal RNA database. *Nucleic Acids Res.*, **32**, D101–D103.
- Yang,H., Jossinet,F., Leontis,N., Chen,L., Westbrook,J., Berman,H. and Westhof,E. (2003) Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res.*, **31**, 3450–3460.
- Klein,D.J., Schmeing,T.M., Moore,P.B. and Steitz,T.A. (2001) The kink-turn: a new RNA secondary structure motif. *EMBO J.*, **20**, 4214–4221.
- Vidovic,I., Nottrott,S., Hartmuth,K., Luhrmann,R. and Ficner,R. (2000) Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol. Cell*, **6**, 1331–1342.
- Rozhdvestvensky,T.S., Tang,T.H., Tchirkova,I.V., Brosius,J., Bachelierie,J.P. and Huttenhofer,A. (2003) Binding of L7Ae protein to

- the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. *Nucleic Acids Res.*, **31**, 869–877.
21. Kuhn, J.F., Tran, E.J. and Maxwell, E.S. (2002) Archaeal ribosomal protein L7 is a functional homolog of the eukaryotic 15.5kD/Snu13p snoRNP core protein. *Nucleic Acids Res.*, **30**, 931–941.
  22. Nevskaya, N., Tishchenko, S., Gabdoulkhakov, A., Nikonova, E., Nikonov, O., Nikulin, A., Platonova, O., Garber, M., Nikonov, S. and Piendl, W. (2005) Ribosomal protein L1 recognizes the same specific structural motif in its target sites on the autoregulatory mRNA and 23S rRNA. *Nucleic Acids Res.*, **33**, 478–485.
  23. Strobel, S.A., Adams, P.L., Stahley, M.R. and Wang, J. (2004) RNA kink turns to the left and to the right. *RNA*, **10**, 1852–1854.
  24. Leontis, N.B., Stombaugh, J. and Westhof, E. (2002) Motif prediction in ribosomal RNAs—lessons and prospects for automated motif prediction in homologous RNA molecules. *Biochimie.*, **84**, 961–973.
  25. Wimberly, B.T., Brodersen, D.E., Clemons, W.M.Jr, Morgan-Warren, R.J., Carter, A.P., Vornrhein, C., Hartsch, T. and Ramakrishnan, V. (2000) *Nature*, **407**, 327–339.
  26. Torres-Larios, A., Dock-Bregeon, A.C., Romby, P., Rees, B., Sankaranarayanan, R., Caillet, J., Springer, M., Ehresmann, C., Ehresmann, B. and Moras, D. (2002) Structural basis of translational control by *Escherichia coli* threonyl tRNA synthetase. *Nature Struct. Biol.*, **9**, 343–347.
  27. Nissen, P., Ippolito, J.A., Ban, N., Moore, P.B. and Steitz, T.A. (2001) RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc. Natl Acad. Sci. USA*, **98**, 4899–4903.
  28. Klein, D.J., Moore, P.B. and Steitz, T.A. (2004) The contribution of metal ions to the structural stability of the large ribosomal subunit. *RNA*, **10**, 1366–1379.