
Topology of three-way junctions in folded RNAs

AURÉLIE LESCOUTE and ERIC WESTHOF

Institut de Biologie Moléculaire et Cellulaire du CNRS, Bioinformatique, modélisation et simulations des acides nucléiques, UPR 9002, Architecture et Réactivité de l'ARN, Université Louis Pasteur, 67084 Strasbourg Cedex, France

ABSTRACT

The three-way junctions contained in X-ray structures of folded RNAs have been compiled and analyzed. Three-way junctions with two helices approximately coaxially stacked can be divided into three main families depending on the relative lengths of the segments linking the three Watson-Crick helices. Each family has topological characteristics with some conservation in the non-Watson-Crick pairs within the linking segments as well as in the types of contacts between the segments and the helices. The most populated family presents tertiary interactions between two helices as well as extensive shallow/minor groove contacts between a linking segment and the third helix. On the basis of the lengths of the linking segments, some guidelines could be deduced for choosing a topology for a three-way junction on the basis of a secondary structure. Examples and prediction based on those rules are discussed.

Keywords: RNA motif; folding; tertiary contact; junction; RNA topology

INTRODUCTION

RNA architecture can be reasonably visualized as the hierarchical assembly of preformed double-stranded helices defined by Watson-Crick base pairs and RNA motifs maintained by non-Watson-Crick base pairs (Michel and Westhof 1990; Brion and Westhof 1997; Batey et al. 1999; Moore 1999; Tinoco and Bustamante 1999; Westhof and Fritsch 2000). The secondary structure is a representation of the helical domains, with the hairpin and internal loops, as well as the junction regions, represented open and unpaired. It is now well appreciated, however, that such regions form, in structured RNAs, compact and sometimes helical-like regions with their bases engaged in non-Watson-Crick pairs, as beautifully demonstrated in the recent ribosomal RNA structures (Ban et al. 2000; Wimberly et al. 2000; Leontis and Westhof 2001; Leontis et al. 2002). The preformed helical domains associate into bundles of helices by end-to-end stacking or parallel packing and form the core of the compact tertiary structure that is further maintained by tertiary interactions between RNA-RNA self-assembly motifs (Cate et al. 1996a,b; Westhof et al. 1996).

In ribozymes, despite invariance in each core, there is a variety in the overall architectures of the catalytic RNAs that promote the stabilization of the helical stems building the core and the correct positioning of the helical substrates (e.g., in group I introns; see Lehnert et al. 1996; Guo et al. 2004; Golden et al. 2005; Woodson 2005). This is achieved by the properties of the RNA anchoring motifs, which allow for the formation of different and often mutually exclusive long-range contacts between nonhomologous peripheral elements. However, in the assembly between domains, one observes the recurrent and systematic use of essentially two main types of long-range RNA-RNA anchors: GNRA tetraloops with their receptors (Costa and Michel 1995, 1997) and loop-loop Watson-Crick or non-Watson-Crick base pairings (Lehnert et al. 1996; Costa et al. 1997, 2000). To promote such long-range contacts, subdomains, which are usually subtended by complex and diverse sets of molecular interactions, have to be assembled. Three-way junctions constitute frequent and critical structured subdomains necessary to promote further long-range RNA-RNA contacts. For example, the three-way junction that forms the catalytic core of the hammerhead ribozyme is constrained by tertiary interactions between peripheral elements. These constraints accelerate the folding of the ribozyme, which is more than 100 times efficient than minimal ribozyme (Khvorova et al. 2003; Canny et al. 2004; Penedo et al. 2004).

Here, we investigate the structures of three-way junctions present in published crystal structures of folded RNAs with the aim of finding sequence signatures, characteristic

Reprint requests to: Eric Westhof, Institut de Biologie Moléculaire et Cellulaire du CNRS, Bioinformatique, modélisation et simulations des acides nucléiques, UPR 9002 Architecture et Réactivité de l'ARN, Université Louis Pasteur, 15 rue René Descartes, 67084 Strasbourg Cedex, France; e-mail: E.Westhof@ibmc.u-strasbg.fr; fax: +33-3-88-60-22-18.

Article and publication are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.2208106>.

of defined structures around three-way junctions and which could be of use for folding three-dimensional structures on the basis of sequence comparisons.

The starting point of the present analysis stems from past experience with the modeling of large RNAs. Over the years, several structured RNAs have been assembled on the basis of sequence analysis, experimental footprinting data, and previously identified RNA–RNA contacts. Crystal structures, published after modeling, are now available for comparison. RMS values ranging between 3.7 and 8.5 Å are obtained depending on the system being compared (Masquida and Westhof 2006). In order to go beyond global agreement between predicted and observed RNA architectures, more information is needed about non-Watson-Crick pairs and their roles in the folding of some critical subdomains; among those, three-way junctions are critical.

One initial hypothesis is that RNA architecture results from the compaction of separate, mostly preformed and stable substructures or modules (Westhof et al. 1996). Although local rearrangements are susceptible to occurring at the interfaces of those building blocks or modules during the process, they are considered minor in comparison to the gross topological features of the final assembly (Wu and Tinoco 1998). In three-way junctions, three Watson-Crick paired helices, linked by at most three single-stranded segments, converge. The number of nucleotides in the single-stranded segments is generally distributed in an unsymmetrical fashion. Further, one observes in three-way junctions first, that two of the helices leading into the junction form an almost contiguous and coaxial stack, with the third helix at an angle to the stack, and, second, that the segment linking the two stacked helices does not contain any nucleotide or only a small number. The objectives of the present work are to find (1) whether three-way junctions can be divided into classes depending on the lengths of the junctions and (2) whether the presence of non-Watson-Crick pairs within nucleotides are characteristic of each class. For example, in the hammerhead ribozyme (Pley et al. 1994; Scott et al. 1995), nucleotides in the junction segments form a stack of three non-Watson-Crick pairs, which allows the coaxial stacking of helices II and III (see Figure 1). We assume that the same coaxial stacks of helices will dominate the fold and, thus, that one should be able to predict the fold of a similar three-way junction in a complex RNA (Westhof et al. 1996). It is well documented that cofactors like ions or proteins contribute to the association process and to the stability of the final fold. Here, we search only for the intrinsic properties and the underlying relationships present in the RNA sequence that promote or allow for the adoption of a particular three-way junction topology.

METHODS AND RESULTS

The three-way junctions present in crystal structures of RNA molecules were searched and collected (Table 1). The coaxially

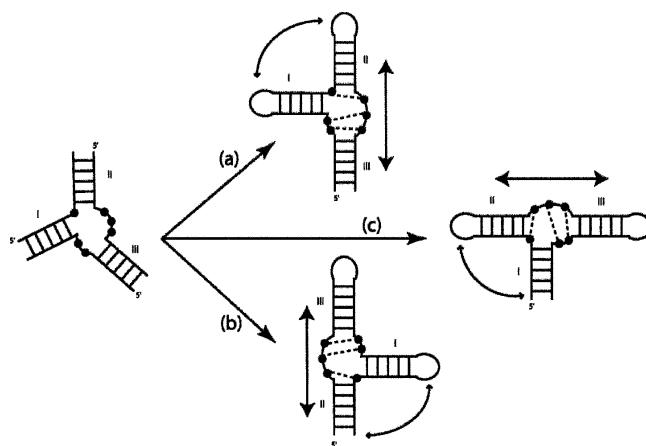


FIGURE 1. Example of a hypothetical three-way junction. The Watson-Crick paired helices are indicated by the ladders. Depending on the structure of the internal junctions, two of the three helices may stack coaxially. In the example shown, helices II and III stack coaxially. Arbitrary non-Watson-Crick contacts are shown as dotted lines. This coaxial stack is maintained whatever the precise secondary structure (whether two loops or only one or none are capped, or whether the closing loops are internal loops in longer helices). In case (a), potential tertiary contacts would occur between loops I and II; in case (b), the contacts would occur potentially between loop I and helix II (or an internal loop within helix II); and in case (c) the contacts would occur potentially between loop I and helix I (or an internal loop within helix I).

stacked helices were visually identified and all three-way junctions represented in a similar fashion: The stacked helices are designated P1 and P2 with P3 on the left side (Fig. 2). In some rare cases, the axes of the three helices are not exactly coplanar. The deviations are small and here will be neglected. The junctions are such that J12 is between P1 and P2, J23 is between P2 and P3, and J31 is between P3 and P1. In all drawings, the green strand is common to P1 and P2; the red strand, to P2 and P3; and the blue strand, to P3 and P1. The number of nucleotides in the three single-stranded junctions (i.e., those not involved in standard helical Watson-Crick pairings) are given in Table 1. A standard helix was considered to exist when there were at least two consecutive canonical Watson-Crick pairs (Vaughn et al. 2002). Three groups (or families) of three-way junctions could be isolated depending on the lengths of the junction strands (Fig. 2). In family A, the number of nucleotides in J31 is smaller than that in J23; in family B, the number of nucleotides in J31 and J23 is the same; in family C, the number of nucleotides in J31 is greater than that in J23. The differences in the number of nucleotides in families A and C is ~ 4 nt. Interestingly, up to now, family A and family B are found only in 16S and 23S rRNAs, while family C is found in various RNAs. Junction J12 contains the smallest number of nucleotides in families A and C (mean is 2 nt), while it contains, on average, the same number of nucleotides (4 nt) as in the other two junctions in family B.

In RNA, helices are right-handed, and single-stranded segments tend also to leave or enter helices in a right-handed fashion. Therefore, at the interface between two

TABLE 1. List of three-way junctions with the number of nucleotides in each segment separating the helices

	J31 Blue	J23 Red	[Δ (J31-J23)]	J12 Green	PDB ID	Domain	Proteins	
Family A								
16S H20-21-22	1	3	2	3	1J5E	Central	<u>S15</u> ;S8; <u>S17</u>	
16S H22-23-23a	1	7	6	0		Central	<u>S6</u> ;S11;S18	
16S H25-25-26a	2	5	3	3		Central	<u>S2</u> ;S8	
16S H34-35-38	4	5	1	5	1S72	3'	S3;S5; <u>S9</u> ;S10;S14	
23S H3-4-23	4	15	11	2		I	<u>L4E</u> ;L24P; <u>L37E</u> ;L39E	
23S H5H6H7	3	5	2	0		I	<u>L29P</u>	
23S H48-X-60	2	4	2	2		III	L19E	
23S H49-59.1-X	1	5	4	2		III	L37E;L39E	
23S H75-76-79	2	5	3	3		V	L15E	
23S H99-100-101	3	4	1	0		V	<u>L3P</u> ;L22P;L31E	
Mean	2	6	4	2				
Sigma (σ)	1	3	3	2				
Family B								
16S H28-29-43	6	5	1	2	1J5E	3'	S7;S9	
16S H32-33-34	6	4	2	2		3'	S14	
16S H33-33a-33b	5	6	1	4		3'	—	
23S H33-34-35	1	2	1	4	1S72	II	L2P	
23S H49-50-51	3	2	1	6		III	L23P; <u>L37E</u> ;L39E	
23S H83-84-85	3	4	1	5		V	L18P;Po	
Mean	4	4	1	4				
Sigma (σ)	2	2	0	2				
Family C								
16S H4-5-15	7	5	2	1	1J5E	5'	S12;S16	
16S H30-31-32	6	3	3	10		3'	<u>S13</u> ;S14;S19	
16S H35-36-37	2	2	0	0		3'	S2;S5	
16S H38-39-40	8	5	3	1	1S72	3'	<u>S14</u>	
23S H2-3-24	14	10	4	0		I	<u>L22P</u> ;L24P	
23S H18-19-20	8	4	4	4		I	<u>L4E</u> ;L24P	
23S H32-33-35	6	3	3	5		II	L2P;L37E	
23S H90-91-92	6	2	4	2		V	<u>L3P</u> ;L14P	
L11 rRNA	5	1	4	1		1HC8	II	L11P
5S	4	3	1	3		1S72		L18P; <u>L21E</u>
Alu domain	4	0	4	0		1e8o		<u>SRP9</u> ;SRP14
S domain	6	0	6	0		1MFQ		
HH	7	1	6	3		1MME		
G-riboswitch	8	3	5	2	1U8D			
P4P6	4	2	2	2	1GID			
Twort Intron	9	3	6	2	1YOQ			
S-dom RNaseP B-type	5	1	4	0	1NBS			
Mean	6	3	4	2				
Sigma (σ)	3	2	2	3				

The references for the various RNAs are the following: 16S rRNA (Wimberly et al. 2000); 23S and 5S rRNA (Klein et al. 2004); L11 rRNA (Conn et al. 1999); Alu domain (Weichenrieder et al. 2000); SRP S domain (Kuglstatter et al. 2002); Hammerhead (Pley et al. 1994); G-riboswitch (Batey et al. 2004); P4P6 (Cate et al. 1996a); Twort intron (Golden et al. 2005); S-domain RNase P type B (Krasilnikov et al. 2003). The PDB ID numbers of the X-ray structures from which the junctions have been extracted are also indicated. For the ribosomal junctions, the domain of the rRNA they are part of is also indicated. In the last column, the proteins interacting with each junction are indicated and the contacts classified as to both RNA backbone and base, to the RNA backbone only (underlined), by stacking only (*italics*). The contacts were assigned visually and only the contacts to the nucleotide junctions and to the next two Watson–Crick base pairs of the helices were considered. For each family, the mean value and sigma value for the lengths of the strand junctions are given.

coaxially and right-handedly stacked helices, the 3'-end strand leaving one helix (P2) will face the deep/major groove of the other helix (P1), while the 5'-end entering strand of helix P1 will face the shallow/minor groove of helix P2 (Fig. 3). The prototypical example is the tRNA structure, where the strand leaving the anti-codon hairpin faces the deep/major groove of the contiguously stacked dihydrouridine helix (Quigley and Rich 1976).

In three-way junctions, with helices P1 and P2 stacked, the right-handedness tendency of the single strands will present J23 toward the deep/major groove of P1, and J31 would normally face the shallow/minor groove of P2. Depending on the relative lengths of the single-stranded segments, actual contacts within the grooves can be made (Fig. 4). The dependency between the relative lengths of the single-stranded and the overall fold of the three-way junctions is due partly to

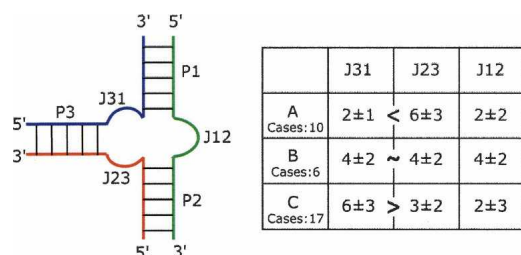


FIGURE 2. The nomenclature used for the three-way junctions with the average numbers of nucleotides in each junction in the table at the right. The number of instances in each of the three families is also indicated.

those geometrical considerations; i.e., that RNA helices are bulky and asymmetric objects with strands of opposite polarity disposed in such a way that when viewing the shallow/minor groove the 5' to 3' strand is at the right.

Organization of the junctions

Are there systematic contacts occurring that are typical of each family? From the preceding paragraph, one can derive the following: For all of the three-way junctions, drawings of the three-dimensional structures together with a representation of the secondary and tertiary base pairs are shown for each family in Figures 5, 7, and 8 (below). Each family contains about the same number of examples stemming from ribosomal RNAs, but families A and B contain only examples from the rRNAs.

Family A

In family A ($J31 < J23$), the junction contacts are the least extensive, and helix P3 is very roughly perpendicular to the P1/P2 coaxial stack (Figs. 4, 5). In most cases, helix P1 ends sharply with a Watson-Crick pair at the junction. In contrast, helices P2 and P3 present often non-Watson-Crick pairs stacked on their last helical base pair. Generally, P2 is organized in a more complex way. There is one motif (or slight variants thereof) found a couple of times and based on a *trans* Hoogsteen–Sugar-Edge AoG base pair. When such a pair is present, the G is always on the 5' end strand.

In almost all cases, J23, the longest strand, interacts in the shallow/minor groove of helix P2 (and rarely with P1), which implies that J23 folds back on itself (Figs. 5, 6). The interactions made by J23 with P2 are diverse, with expected Sugar-Edge–Sugar-Edge pairs comprising A-minor motifs but also Watson-Crick–Watson-Crick or Watson-Crick–Hoogsteen.

Family B

In family B ($J31 \cong J23$), the least populated family, helix P3 bends toward helix P2, and J23 faces the deep/major groove of P1 but does not make contact to it. On the other hand, J31 faces naturally the shallow/minor groove of P2, and contacts

between J31 and the first nucleotides of helix P3 occur. Some elements of family A are present in family B, but no clear trends could be extracted from the available sample (Fig. 7).

Family C

Family C is the most fascinating one (Fig. 8). Family C contains 17 examples, compared to 10 and 6, respectively, in families A and B. Junctions of family C appear in various structured RNAs (from Alu domain to G-riboswitch). In family C ($J31 > J23$), helix P3 bends toward helix P1, and J31 interacts in the shallow/minor groove of helix P2 extensively. J31 is generally structured like a hairpin (using the standard U-turn motif) and often closed by at least 1 bp. The type of base pair forming that pseudo-loop is variable. In several instances, two adenines of the pseudo-loop form A-minor motifs (Nissen et al. 2001) with two consecutive base pairs of helix P2. A clear consensus appears when the pseudo-hairpin contains 3 nt in the loop, two of which make shallow/minor groove contacts with helix P2 (Fig. 9). In such a case, the 3' base of the closing pair is generally in the *syn* conformation (Dock-Bregeon et al. 1989). This type of tri-loop has been described recently (Lee et al. 2003). Four of these are implicated in the core of three-way junctions belonging to family C. The L11 rRNA has been described as a four-way helical junction (Wimberly et al. 1999). According to the definition of a minimal helix as two stacked Watson-Crick base pairs (Waugh et al. 2002), helix 1082 (*Escherichia coli* numbering) should not be considered as a helix. It is rather a tri-loop closed by a non-Watson-Crick base pair (Lee et al. 2003). Other structural characteristics (see Fig. 8) support the classification of L11 rRNA as a three-way junction.

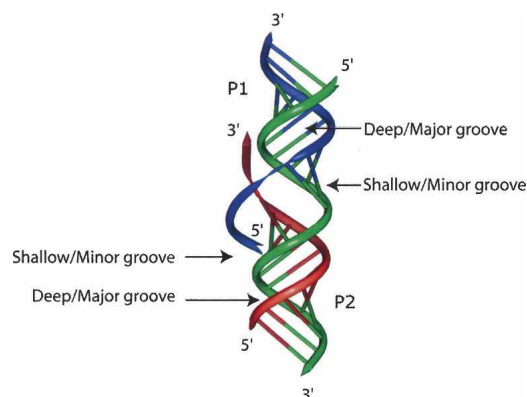


FIGURE 3. Coaxial stack of two helices with single strands entering or leaving the helices. The deep/major and shallow/minor grooves are indicated. The 3'-end strand leaving helix P2 faces the deep/major groove of helix P1. The 5'-end strand entering helix P1 faces the shallow/minor groove of helix P2. One example can be seen in the structure of tRNA, in which the strand leaving the anti-codon hairpin (equivalent to P2) faces the deep/major groove of the dihydrouridine helix (equivalent to P1). Group I introns contain the case depicted at the junction between P4 and P6 (Michel et al. 1990; Adams et al. 2004; Guo et al. 2004; Golden et al. 2005; Woodson 2005).

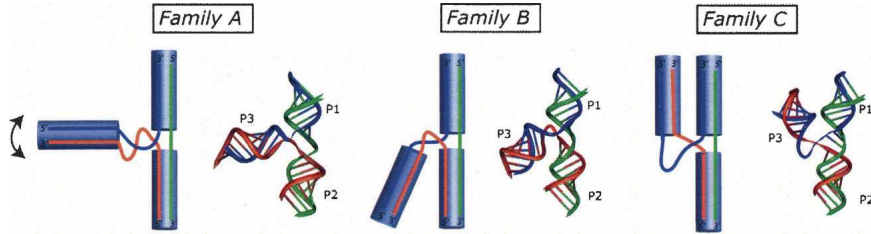


FIGURE 4. Schematic drawings of the three observed families, A, B, and C, in the analyzed three-way junctions. The drawings at the *right* are based on real structures. In family A, the third helix can adopt various angles with respect to the coaxially stacked helices.

Helices P1 and P3 point in the same region of space and, thus, when they are capped by hairpin loops, multiple interactions can occur between these apical loops. The diversity of the interactions between the apical loops defies any attempt at generalization. The extensive contacts on either side of the coaxial interface might explain the fre-

quent occurrence of family C folds in three-way junctions. For example, the 23S H90–H91–H92 three-way junction in the large subunit of the ribosome is key in the organization of the multiple junction around which the peptidyl-ation reaction occurs (Ban et al. 2000; Klein et al. 2004). Similarly, the junction H18–H19–H20 in domain I of the 23S rRNA is critical for the 50S assembly (Klein et al. 2004). Another example is the P5abc three-way junction of the *Tetrahymena* group I intron, which plays a key role in the folding and activity of this ribozyme (van der Horst et al. 1991; Lehnert et al. 1996; Engelhardt et al. 2000). It was shown recently that in absence of P5abc, the RNA of *Tetrahymena* folds in alternative forms that are as stable as the native form, whereas the presence of P5abc induces a very

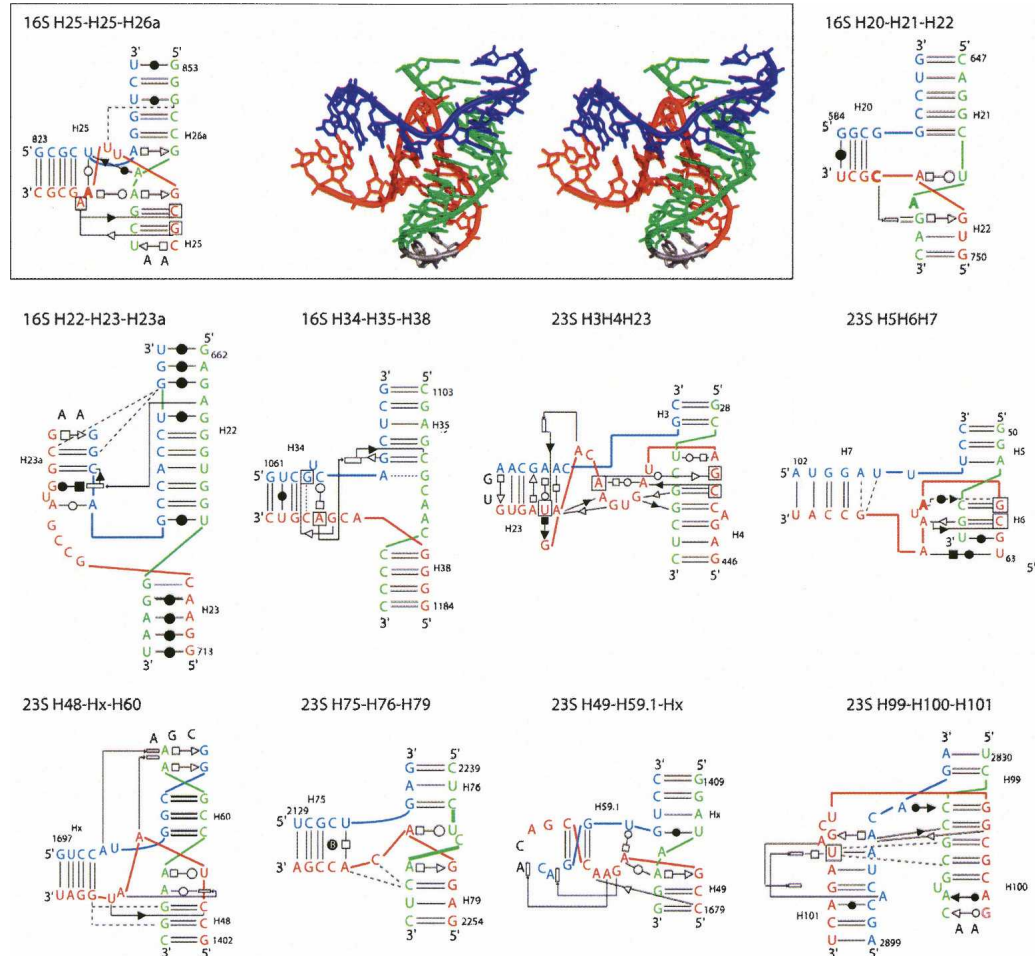


FIGURE 5. The three-way junctions belonging to family A. Ten three-way rRNAs junctions belong to family A. The name of the junction depends on the RNA to which it belongs and on the numbering of the three helices that are anchored to the junction. For a typical junction, the three-dimensional stereo view (DeLano Scientific, <http://www.pymol.org>) is shown on the *right* and the secondary structure with secondary and tertiary interactions is represented on the *left* (Yang et al. 2003). For the other junctions, only the secondary structure diagrams with the symbols for the tertiary contacts are shown. The symbols used are according to the Leontis and Westhof (2001) nomenclature.

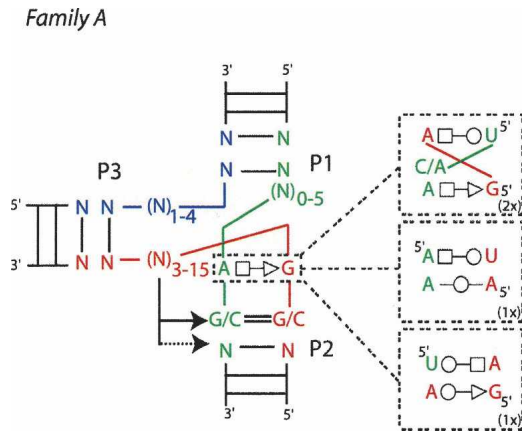


FIGURE 6. A consensus for the family A three-way junction. In family A, the junction J12 includes 0–5 nt; J23, 3–15 nt; and J31, 1–4 nt. Nucleotides at position 3, 4, or 5 of J23 make tertiary interactions with the shallow groove of the first one or the first two Watson-Crick base pairs of P2. The closing base pair of P2 is always in the *trans* orientation.

important stabilization of the native conformation compared to alternative conformations (Johnson et al. 2005). The compactness achieved by this junction promotes the folding of P5abc and ensures the specific contacts with the rest of the intron.

Could one predict the topology adopted by a three-way junction?

As discussed above, the main objective of such an analysis is to deduce rules allowing the prediction of (1) the topology

adopted by a given three-way junction and (2) the potential RNA–RNA contacts linking elements subtended by helices of the three-way junctions. On the basis of a secondary structure displaying a three-way junction, one can suggest the following folding rules:

1. For the continuous strand of the coaxial stack, choose the uninterrupted strand so that J12 contains 0 nt (if there is such a junction) (Kim and Cech 1987). There is no contradictory example. If there are two such strands, pick the coaxial stack so that Watson–Crick pairs are present at the interface. However, there are examples where J12 does not contain the least number of nucleotides (see below).
2. If there is no linking segment longer than the others, family B is the best choice. But, the choice of the coaxially stacked helices is not automatically determined.
3. The choice between families A and C is not straightforward. For family A, check for the possibility of forming a *trans* Hoogsteen–Sugar–Edge at the interface side of P2. For family C, check whether J31 can form a pseudo-hairpin with two to three residues in the loop and one or two adenines for contacting the shallow/minor groove of P2.

When applying the rules above, an important starting point is the available secondary structure, and the outcome of the preceding rules will strongly depend on the accuracy of the secondary structure. For example, it does occur that a terminal base pair of a helix, apparently Watson–Crick, is not formed in the three-way junction and is, instead, engaged in a tertiary contact within the junction. The three-way junction H20–H21–H22 of the 16S rRNA,

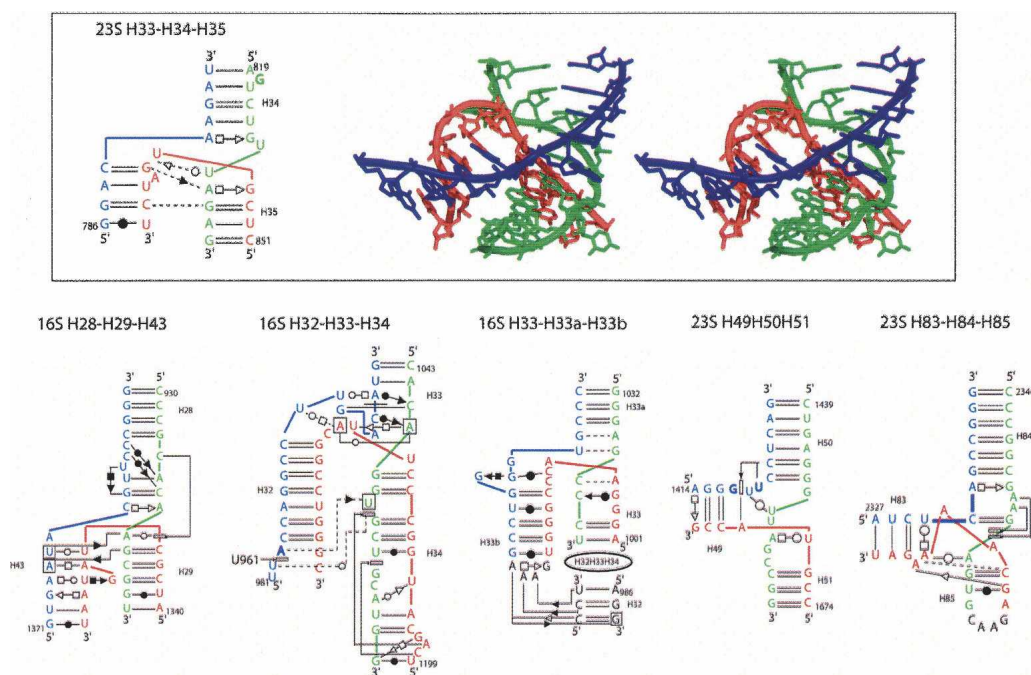


FIGURE 7. The three-way junctions belonging to family B. Same legend as for Figure 5.

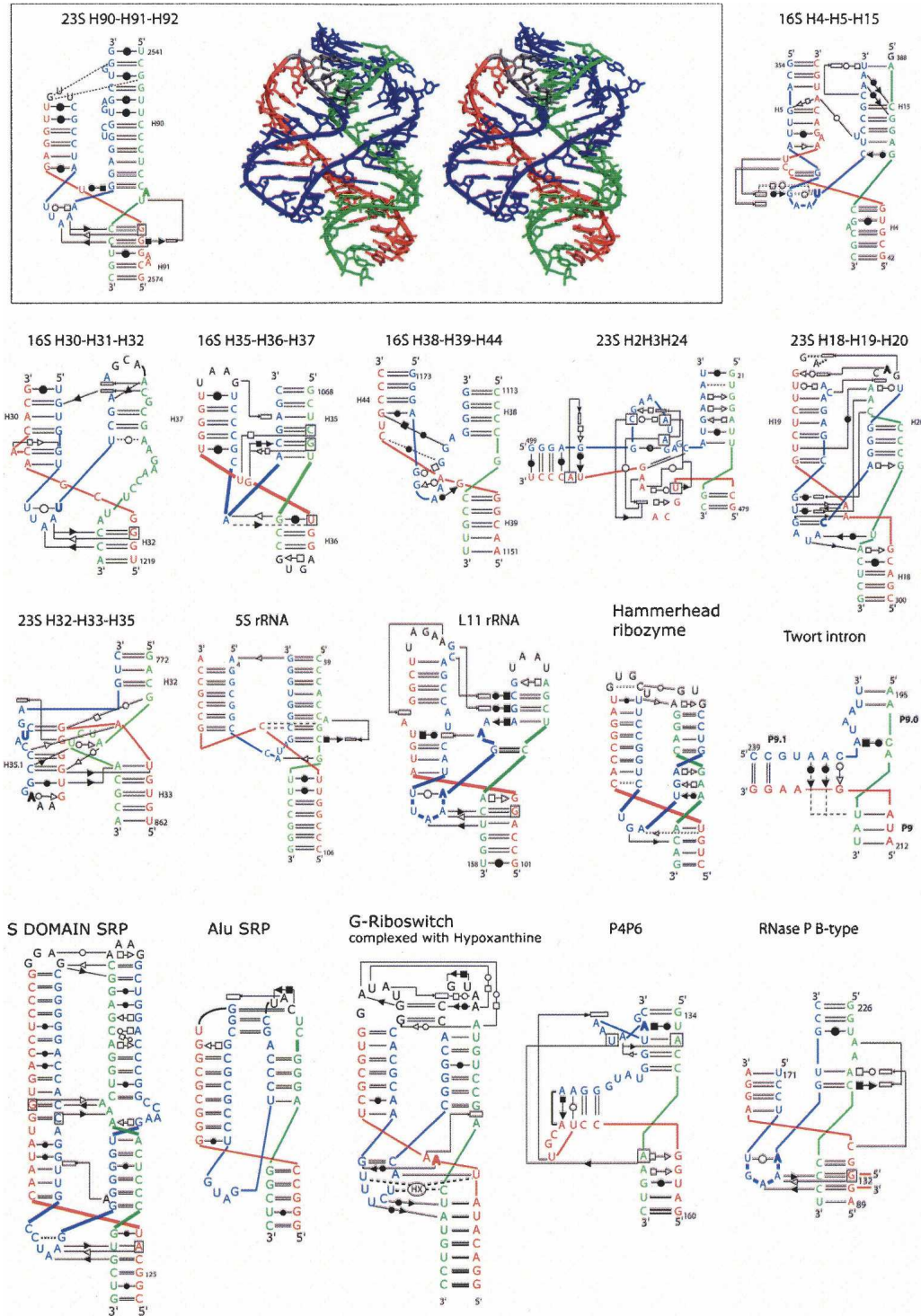


FIGURE 8. The three-way junctions belonging to family C. Ten cases come from the rRNA structures, and the other seven are from different RNA structures. Same legend as for Figure 5.

involved in binding the primary protein S15, is a clear example (see Fig. 4): H20 ends with G587 and C754; these nucleotides are highly conserved (C754, 100%; G587, 96%) leading to assuming erroneously the presence of a G=C pair despite the lack of covariation evidence (Serganov et al.

1996). With such an assumption, following rule 1, a coaxial stack of H20 and H21 would be deduced.

Interestingly, the S15 complex is characterized by a profound conformational change of the three-way junction from the free RNA to the native form of the complexed

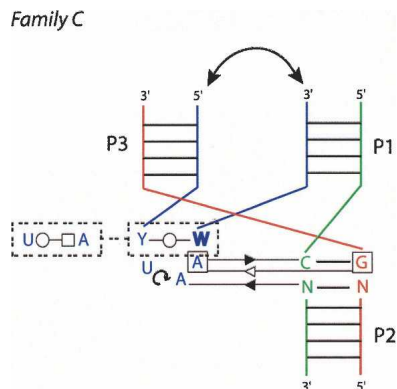


FIGURE 9. A consensus for the family C three-way junction. In some cases, junction strand J31 folds into a tri-loop closed by a *trans* base pair, often a *trans* Watson-Crick. The nucleotide in 3' is in the *syn* conformation (bold) and the free loop nucleotides at positions 2 and 3, often adenines, make Sugar-Edge-Sugar-Edge interactions with nucleotides of base pairs at positions 1 and 2 of helix P2. In almost all the observed cases, there are tertiary interactions between helices P1 and P3 (double arrow).

RNA (Agalarov et al. 2000; Nikulin et al. 2000; Williamson 2000). Experimental evidence has, however, demonstrated that Mg^{2+} ions and protein S15 stabilize the same conformation of H20–H21–H22 (Orr et al. 1998; Agalarov et al. 2000). Furthermore, a detailed analysis has also shown that the folding of the junction is determined by RNA elements rather than by protein binding (Batey and Williamson 1998). Thus, in that particular example, there is in principle enough information in the RNA sequence to deduce the conformation of the three-way junction in the folded RNA. But, as discussed cogently (Williamson 2000), induced-fit could imply that there is not enough information content in the sequence of each partner to predict the fold within the functional complex. The observation that all the compiled examples found in families A and B belong to the ribosomal particles strengthens further that possibility. We surveyed the protein contacts (see Table 1) around the three-way junctions and classified them as backbone only, stacking only, and base with backbone contacts. The trends and the statistics, although weak, show that the least amount of bound proteins is found in family B and the largest in family A, with family C giving an intermediate situation. In each family, there are at least twice as many backbone-only contacts than contacts implying nucleotide bases with or without the backbone. Backbone contacts do occur with the conserved regions of families A (the 3' strand of P2 in Fig. 6) and C (the tri-loop region in Fig. 9). Interestingly, when the junction J12 contains a large number of nucleotides, these nucleotides form extensive contacts with bound proteins. These observations are in agreement with the conclusions reached on the S15 system (Batey and Williamson 1998; Orr et al. 1998; Williamson 2000), namely, that intrinsic RNA elements govern the choice of the folding of the junction but that ions and cofactors may be required for stabilization of the architecture.

Some predictive applications

In the following examples, we will apply the preceding rules for attempting to deduce a possible topology for a three-way junction (Fig. 10). The Varkud satellite (VS) ribozyme contains two important three-way junctions (Beattie et al. 1995). They were both studied in detail (Lafontaine et al. 2001, 2002; Lilley 2004). Although rather precise folds could be proposed, no crystal structure exists yet. Concerning junction II–III–VI, following point 1, helices III and VI should coaxially stack. This choice leads to J23 > J31 and thus to a family A type of junction, as previously shown (Lafontaine et al. 2002). The case of junction III–IV–V is not as straightforward. The junction between helices III and V is longer than the other two: The choice is between family A and C. None of the consensus type of contacts can be made in family A. However, in family C, the J23 junction, UGAUU, could form a pseudo-hairpin capped by a UoU pair and the middle A making a shallow groove contact with helix IV. This choice was also proposed on the basis of experimental data (Lafontaine et al. 2002). Interestingly, for the III–IV–V junction of the VS ribozyme, the deletion of the single free uridine between helices IV and V changes the folding by modifying the coaxial stacking. Indeed, in presence of the single U nucleotide, helices IV and III are stacked, whereas without the U nucleotide in the J45 strand,

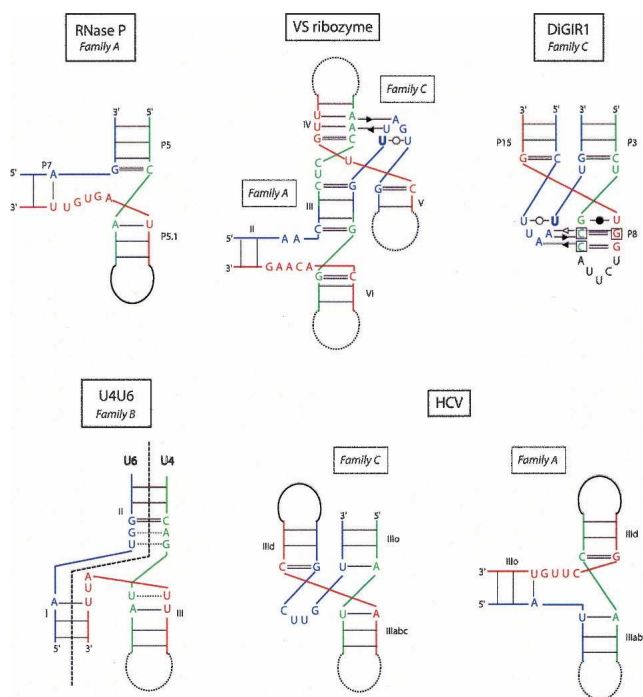


FIGURE 10. Some applications to noncrystallized three-way junctions. The secondary structures of the unknown RNAs have been represented in the fold corresponding to the proposed family. In the case of HCV, as two junction strands have no free nucleotides, two possibilities of stacking are allowed so the junction could belong to either family C or family A.

helices IV and V would be stacked (in agreement with rule 1, above).

Another example of continuing interest is the three-way junction formed between U4 and U6 RNAs in the spliceosome. The number of nucleotides in the junctions is about the same. So one could suggest a family B type. The similarity with the junction of 16S H33–H33a–H33b has been incorporated in the proposed structure shown in Figure 10.

RNase P contains a striking example in one of the two P families (the B-type): the junction between P5, P5.1, and P7. P5 and P5.1 are always contiguous and, thus, a coaxial stacking of P5 and P5.1 is expected. The junction between P5.1 and P7 (J23) is the longest, making this junction a family A type. The segment J23 is highly conserved (AGUGW) and could adopt a complex fold. It was modeled, some years ago, as a right-handed stretch (Massire et al. 1998). The crystal structure of a B-type P RNA (Kazantsev et al. 2005) was published during review of this publication and it does show stacking of P5 and P5.1 with helix P7 at a wide angle.

The group I-like ribozyme (DiGIR1), found in the eukaryotic microorganism *Didymium*, is characterized by a pseudo-knot P15 between P3 and P8 forming a three-way junction (Einvik et al. 1998). In that case, P3 and P8 are coaxially stacked (Michel and Westhof 1990) and the junction belongs clearly to family C like the P10/P10.1/P11 junction of the B-type P RNAs (Krasilnikov et al. 2003). The conserved J31 segment, –UUAU–, can form a pseudo-hairpin closed by a UoU and the two As forming shallow/minor groove contacts with helix P8 (Fig. 10).

The IRES of hepatitis C virus (HCV) contains a structured RNA with one three-way junction (Honda et al. 1999). If the secondary structure is locally correct, two choices are possible involving coaxial stacks: Either helices IIIo and IIIabc stack, giving a family C junction, or helices IIId and IIIabc stack, giving a family A junction. The longest segment, J31 or J23, respectively, does not contain any A (–CUUG–). This fact would tend to favor a family A type for the three-way junction instead of a family C (Fig. 10). With the latter choice, helices IIIabc and IIId would be coaxial and roughly perpendicular to IIIo, leading to a severe reorientation of IIIabc with respect to the other elements (Spahn et al. 2001).

DISCUSSION AND CONCLUSIONS

For RNA motifs, like K-turns (Klein et al. 2001) and C-motifs (Leontis and Westhof 2003), we have shown previously, under the hypothesis that homologous sequences fold into similar three-dimensional structures, that a systematic analysis of sequences in the light of X-ray structures allows us to derive covariation rules for non-Watson-Crick base pairs using a geometric classification of non-Watson-Crick pairs (Lescoute et al. 2005). Such rules, based on isostericity matrices that for a given pair give the structural equivalences observed to substitute in

sequences, allow us to identify with great confidence RNA motifs. In that previous work (Lescoute et al. 2005), local RNA motifs were defined operationally as ordered arrays of non-Watson-Crick base pairs (Leontis and Westhof 2003).

In the present analysis, we considered a more complex assembly, a multiple junction of three helices. Three-way junctions are common in RNA secondary structure. They frequently fulfill essential architectural function as, for example, in the hammerhead (Khvorova et al. 2003; Canny et al. 2004; Penedo et al. 2004) or Varkud (Lafontaine et al. 2001; Lilley 2004) ribozymes and in the ribosomal RNAs (Brodersen et al. 2002; Klein et al. 2004). It is therefore important to derive folding rules for three-way junctions. However, three-way junctions cannot be defined, even a posteriori, as an ordered organization of non-Watson-Crick base pairs.

General trends could be observed like (1) the clear preference for interactions in the shallow/minor groove compared to the deep/major groove and (2) the preference for right-handedness of stacks and sugar-phosphate backbone pathways. Although very strict rules could not be extracted, there are some definite preferences for some non-Watson-Crick pairs at key positions in families A and C. But the variety of the interactions illustrates how versatile they are and how a given type of contact can be replaced by another or others depending on the precise local sequence. Except for those instances following the consensus in families A and C, it is not straightforward to distinguish between the core contacts, key for a given fold, and those contacts that are opportunistic because they depend on the fold and the local sequence environments. Despite this molecular adaptability, some intrinsic properties and underlying relationships present in the RNA sequence that promote or allow for the adoption of a particular three-way junction topology could be extracted. Furthermore, by comparing the various structures in each family, some predictive guidelines could be deduced and applied to noncrystallized RNA systems.

ACKNOWLEDGMENTS

E.W. thanks the Institut universitaire de France for support over the last 10 years.

Received August 25, 2005; accepted October 19, 2005.

REFERENCES

- Adams, P.L., Stahley, M.R., Gill, M.L., Kosek, A.B., Wang, J., and Strobel, S.A. 2004. Crystal structure of a group I intron splicing intermediate. *RNA* **10**: 1867–1887.
- Agalarov, S.C., Sridhar Prasad, G., Funke, P.M., Stout, C.D., and Williamson, J.R. 2000. Structure of the S15, S6, S18-rRNA complex: Assembly of the 30S ribosome central domain. *Science* **288**: 107–113.

- 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.
- Batey, R.T. and Williamson, J.R. 1998. Effects of polyvalent cations on the folding of an rRNA three-way junction and binding of ribosomal protein S15. *RNA* **4**: 984–997.
- Batey, R.T., Rambo, R.P., and Doudna, J.A. 1999. Tertiary motifs in RNA structure and folding. *Angew Chem. Int. Ed. Engl.* **38**: 2326–2343.
- Batey, R.T., Gilbert, S.D., and Montange, R.K. 2004. Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature* **432**: 411–415.
- Beattie, T.L., Olive, J.E., and Collins, R.A. 1995. A secondary-structure model for the self-cleaving region of *Neurospora* VS RNA. *Proc. Natl. Acad. Sci.* **92**: 4686–4690.
- Brion, P. and Westhof, E. 1997. Hierarchy and dynamics of RNA folding. *Annu. Rev. Biophys. Biomol. Struct.* **26**: 113–137.
- Brodersen, D.E., Clemons Jr., W.M., Carter, A.P., Wimberly, B.T., and Ramakrishnan, V. 2002. Crystal structure of the 30 S ribosomal subunit from *Thermus thermophilus*: Structure of the proteins and their interactions with 16 S RNA. *J. Mol. Biol.* **316**: 725–768.
- Canny, M.D., Jucker, F.M., Kellogg, E., Khvorova, A., Jayasena, S.D., and Pardi, A. 2004. Fast cleavage kinetics of a natural hammerhead ribozyme. *J. Am. Chem. Soc.* **126**: 10848–10849.
- Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R., and Doudna, J.A. 1996a. Crystal structure of a group I ribozyme domain: Principles of RNA packing. *Science* **273**: 1678–1685.
- Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Szewczak, A.A., Kundrot, C.E., Cech, T.R., and Doudna, J.A. 1996b. RNA tertiary structure mediation by adenosine platforms. *Science* **273**: 1696–1699.
- Conn, G.L., Draper, D.E., Lattman, E.E., and Gittis, A.G. 1999. Crystal structure of a conserved ribosomal protein-RNA complex. *Science* **284**: 1171–1174.
- Costa, M. and Michel, F. 1995. Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.* **14**: 1276–1285.
- . 1997. Rules for RNA recognition of GNRA tetraloops deduced by in vitro selection: Comparison with in vivo evolution. *EMBO J.* **16**: 3289–3302.
- Costa, M., Deme, E., Jacquier, A., and Michel, F. 1997. Multiple tertiary interactions involving domain II of group II self-splicing introns. *J. Mol. Biol.* **267**: 520–536.
- Costa, M., Michel, F., and Westhof, E. 2000. A three-dimensional perspective on exon binding by a group II self-splicing intron. *EMBO J.* **19**: 5007–5018.
- Dock-Bregeon, A.C., Westhof, E., Giege, R., and Moras, D. 1989. Solution structure of a tRNA with a large variable region: Yeast tRNA^{Ser}. *J. Mol. Biol.* **206**: 707–722.
- Einvik, C., Nielsen, H., Westhof, E., Michel, F., and Johansen, S. 1998. Group I-like ribozymes with a novel core organization perform obligate sequential hydrolytic cleavages at two processing sites. *RNA* **4**: 530–541.
- Engelhardt, M.A., Doherty, E.A., Knitt, D.S., Doudna, J.A., and Herschlag, D. 2000. The P5abc peripheral element facilitates pre-organization of the *Tetrahymena* group I ribozyme for catalysis. *Biochemistry* **39**: 2639–2651.
- Golden, B.L., Kim, H., and Chase, E. 2005. Crystal structure of a phage Twort group I ribozyme-product complex. *Nat. Struct. Mol. Biol.* **12**: 82–89.
- Guo, F., Gooding, A.R., and Cech, T.R. 2004. Structure of the *Tetrahymena* ribozyme: Base triple sandwich and met I ion at the active site. *Mol. Cell* **16**: 351–362.
- Honda, M., Beard, M.R., Ping, L.H., and Lemon, S.M. 1999. A phylogenetically conserved stem-loop structure at the 5′ border of the internal ribosome entry site of hepatitis C virus is required for cap-independent viral translation. *J. Virol.* **73**: 1165–1174.
- Johnson, T.H., Tijerina, P., Chadee, A.B., Herschlag, D., and Russell, R. 2005. Structural specificity conferred by a group I RNA peripheral element. *Proc. Natl. Acad. Sci.* **102**: 10176–10181.
- Kazantsev, A.V., Krivenko, A.A., Harrington, D.J., Holbrook, S.R., Adams, P.D., and Pace, N.R. 2005. Crystal structure of a bacterial ribonuclease P RNA. *Proc. Natl. Acad. Sci.* **102**: 13392–13397.
- Khvorova, A., Lescoute, A., Westhof, E., and Jayasena, S.D. 2003. Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. *Nat. Struct. Biol.* **10**: 708–712.
- Kim, S.H. and Cech, T.R. 1987. Three-dimensional model of the active site of the self-splicing rRNA precursor of *Tetrahymena*. *Proc. Natl. Acad. Sci.* **84**: 8788–8792.
- 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.
- Klein, D.J., Moore, P.B., and Steitz, T.A. 2004. The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *J. Mol. Biol.* **340**: 141–177.
- Krasilnikov, A.S., Yang, X., Pan, T., and Mondragon, A. 2003. Crystal structure of the specificity domain of ribonuclease P. *Nature* **421**: 760–764.
- Kuglstatler, A., Oubridge, C., and Nagai, K. 2002. Induced structural changes of 7SL RNA during the assembly of human signal recognition particle. *Nat. Struct. Biol.* **9**: 740–744.
- Lafontaine, D.A., Norman, D.G., and Lilley, D.M. 2001. Structure, folding and activity of the VS ribozyme: Importance of the 2–3–6 helical junction. *EMBO J.* **20**: 1415–1424.
- . 2002. The global structure of the VS ribozyme. *EMBO J.* **21**: 2461–2471.
- Lee, J.C., Cannone, J.J., and Gutell, R.R. 2003. The lonepair triloop: A new motif in RNA structure. *J. Mol. Biol.* **325**: 65–83.
- Lehnert, V., Jaeger, L., Michel, F., and Westhof, E. 1996. New loop-loop tertiary interactions in self-splicing introns of subgroup IC and ID: A complete 3D model of the *Tetrahymena thermophila* ribozyme. *Chem. Biol.* **3**: 993–1009.
- Leontis, N.B. and Westhof, E. 2001. Geometric nomenclature and classification of RNA base pairs. *RNA* **7**: 499–512.
- . 2002. Analysis of RNA motifs. *Curr. Opin. Struct. Biol.* **13**: 300–308.
- 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.
- Lescoute, A., Leontis, N.B., Massire, C., and Westhof, E. 2005. Recurrent structural RNA motifs, isostericity matrices and sequence alignments. *Nucleic Acids Res.* **33**: 2395–2409.
- Lilley, D.M. 2004. The Varkud satellite ribozyme. *RNA* **10**: 151–158.
- Masquida, B. and Westhof, E. 2006. A modular and hierarchical approach for all-atom modeling. In *The RNA world* (eds. R.F. Gesteland et al.), pp. 659–681. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Massire, C., Jaeger, L., and Westhof, E. 1998. Derivation of the three-dimensional architecture of bacterial ribonuclease P RNAs from comparative sequence analysis. *J. Mol. Biol.* **279**: 773–793.
- Michel, F. and Westhof, E. 1990. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* **216**: 585–610.
- Michel, F., Ellington, A.D., Couture, S., and Szostak, J.W. 1990. Phylogenetic and genetic evidence for base-triples in the catalytic domain of group I introns. *Nature* **347**: 578–580.
- Moore, P.B. 1999. Structural motifs in RNA. *Annu. Rev. Biochem.* **67**: 287–300.
- Nikulin, A., Serganov, A., Ennifar, E., Tishchenko, S., Nevskaya, N., Shepard, W., Portier, C., Garber, M., Ehresmann, B., Ehresmann, C., et al. 2000. Crystal structure of the S15-rRNA complex. *Nat. Struct. Biol.* **7**: 273–277.
- 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.* **98**: 4899–4903.
- Orr, J.W., Hagerman, P.J., and Williamson, J.R. 1998. Protein and Mg²⁺-induced conformational changes in the S15 binding site of 16 S ribosomal RNA. *J. Mol. Biol.* **275**: 453–464.

- Penedo, J.C., Wilson, T.J., Jayasena, S.D., Khvorova, A., and Lilley, D.M. 2004. Folding of the natural hammerhead ribozyme is enhanced by interaction of auxiliary elements. *RNA* **10**: 880–888.
- Pley, H.W., Flaherty, K.M., and McKay, D.B. 1994. Three-dimensional structure of a hammerhead ribozyme. *Nature* **372**: 68–74.
- Quigley, G.J. and Rich, A. 1976. Structural domains of transfer RNA molecules. *Science* **194**: 796–806.
- Scott, W.G., Finch, J.T., and Klug, A. 1995. The crystal structure of an all-RNA hammerhead ribozyme: A proposed mechanism for RNA catalytic cleavage. *Cell* **81**: 991–1002.
- Serganov, A.A., Masquida, B., Westhof, E., Cachia, C., Portier, C., Garber, M., Ehresmann, B., and Ehresmann, C. 1996. The 16S rRNA binding site of *Thermus thermophilus* ribosomal protein S15: Comparison with *Escherichia coli* S15, minimum site and structure. *RNA* **2**: 1124–1138.
- Spahn, C.M., Kieft, J.S., Grassucci, R.A., Penczek, P.A., Zhou, K., Doudna, J.A., and Frank, J. 2001. Hepatitis C virus IRES RNA-induced changes in the conformation of the 40S ribosomal subunit. *Science* **291**: 1959–1962.
- Tinoco Jr., I. and Bustamante, C. 1999. How RNA folds. *J. Mol. Biol.* **293**: 271–281.
- van der Horst, G., Christian, A., and Inoue, T. 1991. Reconstitution of a group I intron self-splicing reaction with an activator RNA. *Proc. Natl. Acad. Sci.* **88**: 184–188.
- Waugh, A., Gendron, P., Altman, R., Brown, J.W., Case, D., Gautheret, D., Harvey, S.C., Leontis, N., Westbrook, J., Westhof, E., et al. 2002. RNAML: A standard syntax for exchanging RNA information. *RNA* **8**: 707–717.
- Weichenrieder, O., Wild, K., Strub, K., and Cusack, S. 2000. Structure and assembly of the Alu domain of the mammalian signal recognition particle. *Nature*. **408**: 167–173.
- Westhof, E. and Fritsch, V. 2000. RNA folding: Beyond Watson-Crick pairs. *Structure Fold. Des.* **8**: R55–R65.
- Westhof, E., Masquida, B., and Jaeger, L. 1996. RNA tectonics: Towards RNA design. *Fold. Des.* **1**: R78–R88.
- Williamson, J.R. 2000. Induced fit in RNA-protein recognition. *Nat. Struct. Biol.* **7**: 834–837.
- Wimberly, B.T., Guymon, R., McCutcheon, J.P., White, S.W., and Ramakrishnan, V. 1999. A detailed view of a ribosomal active site: The structure of the L11-RNA complex. *Cell* **97**: 491–502.
- Wimberly, B.T., Brodersen, D.E., Clemons Jr., W.M., Morgan-Warren, R.J., Carter, A.P., Vonnrhein, C., Hartsch, T., and Ramakrishnan, V. 2000. Structure of the 30S ribosomal subunit. *Nature* **407**: 327–339.
- Woodson, S.A. 2005. Structure and assembly of group I introns. *Curr. Opin. Struct. Biol.* **15**: 324–330.
- Wu, M. and Tinoco Jr., I. 1998. RNA folding causes secondary structure rearrangement. *Proc. Natl. Acad. Sci.* **95**: 11555–11560.
- 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.