

# Tertiary Motifs Revealed in Analyses of Higher-Order RNA Junctions

Christian Laing<sup>1,2</sup>, Segun Jung<sup>1,3</sup>, Abdul Iqbal<sup>1,2</sup> and Tamar Schlick<sup>1,2,3\*</sup>

<sup>1</sup>Department of Chemistry,  
New York University,  
251 Mercer Street, New York,  
NY 10012, USA

<sup>2</sup>Courant Institute of  
Mathematical Sciences,  
New York University,  
251 Mercer Street, New York,  
NY 10012, USA

<sup>3</sup>Computational Biology  
Program, New York University,  
251 Mercer Street, New York,  
NY 10012, USA

Received 2 June 2009;  
received in revised form  
29 July 2009;  
accepted 29 July 2009  
Available online  
3 August 2009

RNA junctions are secondary-structure elements formed when three or more helices come together. They are present in diverse RNA molecules with various fundamental functions in the cell. To better understand the intricate architecture of three-dimensional (3D) RNAs, we analyze currently solved 3D RNA junctions in terms of base-pair interactions and 3D configurations. First, we study base-pair interaction diagrams for solved RNA junctions with 5 to 10 helices and discuss common features. Second, we compare these higher-order junctions to those containing 3 or 4 helices and identify global motif patterns such as coaxial stacking and parallel and perpendicular helical configurations. These analyses show that higher-order junctions organize their helical components in parallel and helical configurations similar to lower-order junctions. Their sub-junctions also resemble local helical configurations found in three- and four-way junctions and are stabilized by similar long-range interaction preferences such as A-minor interactions. Furthermore, loop regions within junctions are high in adenine but low in cytosine, and in agreement with previous studies, we suggest that coaxial stacking between helices likely forms when the common single-stranded loop is small in size; however, other factors such as stacking interactions involving noncanonical base pairs and proteins can greatly determine or disrupt coaxial stacking. Finally, we introduce the ribo–base interactions: when combined with the along-groove packing motif, these ribo–base interactions form novel motifs involved in perpendicular helix–helix interactions. Overall, these analyses suggest recurrent tertiary motifs that stabilize junction architecture, pack helices, and help form helical configurations that occur as sub-elements of larger junction networks. The frequent occurrence of similar helical motifs suggest nature's finite and perhaps limited repertoire of RNA helical conformation preferences. More generally, studies of RNA junctions and tertiary building blocks can ultimately help in the difficult task of RNA 3D structure prediction.

© 2009 Elsevier Ltd. All rights reserved.

**Keywords:** RNA structure; junction; tertiary motif; ribo–base interaction; non-Watson–Crick base pair

*Edited by A. Pyle*

## Introduction

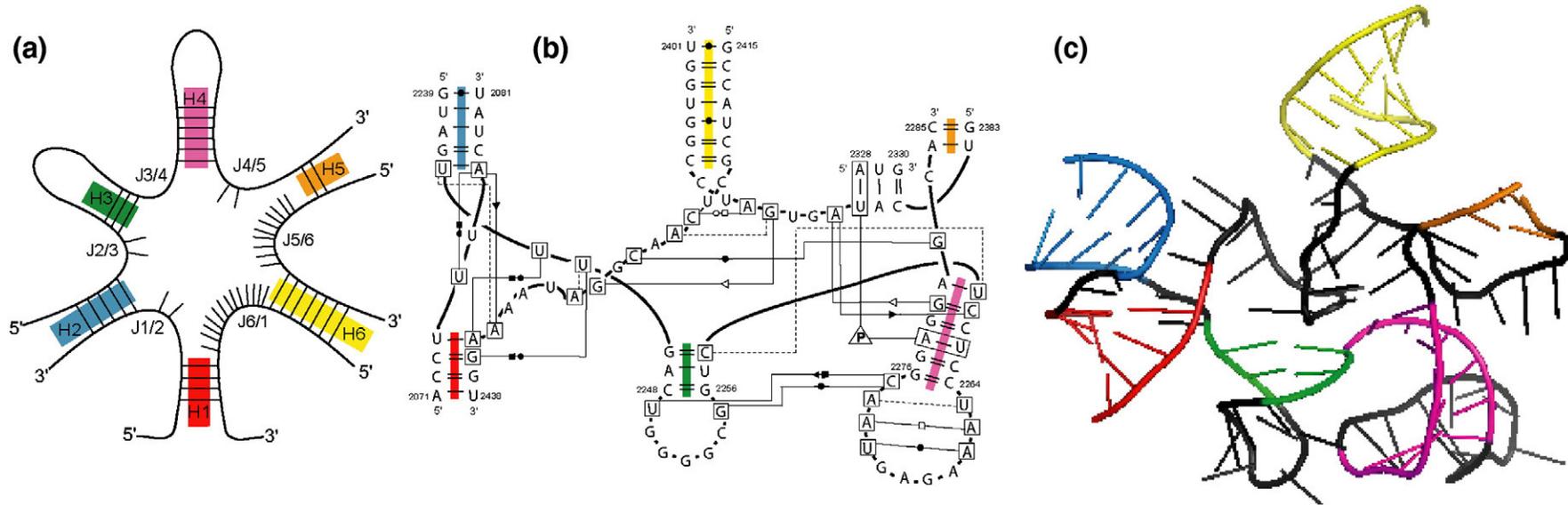
RNA molecules adopt well-defined three-dimensional (3D) structures of high complexity to

perform specific functions in the cell. These complex architectures piece together basic secondary structural elements such as helices, hairpins, internal loops, and junctions, which bind together via tertiary interactions to form compact structures of active RNAs.

An RNA junction can be defined as the point of connection between different helical segments<sup>1,2</sup> (Fig. 1a). This secondary-structure arrangement is present in a wide range of RNA molecules and is involved in a variety of different functional roles, including the self-cleaving catalytic properties of the hammerhead ribozyme,<sup>4</sup> promotion of func-

\*Corresponding author. Department of Chemistry, New York University, 251 Mercer Street, New York, NY 10012, USA. E-mail address: [schlick@nyu.edu](mailto:schlick@nyu.edu).

Abbreviations used: 3D, three-dimensional; WC, Watson–Crick; PDB, Protein Data Bank; AGPM, along-groove packing motif.



**Fig. 1.** Junction architecture for *E. coli* 23S rRNA (2AW4\_2073 from Table 1). (a) Secondary-structure diagram of the six-way junction element composed of six helices labeled  $H_1$  to  $H_6$  (color coded) and six loop regions labeled  $J_{1/2}$  to  $J_{6/1}$  with nucleotide positions marked in black. Helices and loop regions are labeled uniquely according to the 5'-to-3' orientation of the entire RNA structure, by labeling  $H_1$  as the first helix encountered while entering the junction in the 5'-to-3' direction; subsequent helices within the junction are labeled as one moves along the nucleotide chain in the same direction. Lines inside the helices represent the canonical WC base pairs G-C, A-U, and the G-U wobble base pair. (b) Network interaction diagram representing base pairs of the same six-way junction according to the Leontis-Westhof symbology.<sup>3</sup> (c) 3D representation diagram containing a coaxial stacking between helices  $H_1$  (red) and  $H_2$  (blue).

tional folded states of the hairpin ribozyme,<sup>5</sup> recognition of the binding pocket domain by purine riboswitches,<sup>6,7</sup> and translation initiation of the hepatitis C virus at the internal ribosome entry site.<sup>8</sup> Several junctions also occur within ribosomal RNA subunits<sup>9–11</sup> where they play important roles and often bind to ribosomal proteins.<sup>12</sup> While more is known about other secondary-structure elements such as hairpins and internal loops,<sup>13</sup> our current understanding of the more complex junction element, especially higher-order junctions, is limited. Because junctions serve as major architectural features in RNA, it is essential to better understand the structural, energetic, and dynamic aspects of these elements.

RNA crystallography, NMR, and other experimental techniques such as fluorescence resonance energy transfer and small-angle X-ray scattering have offered unprecedented opportunities to analyze RNA tertiary (3D) structure.<sup>11,14–19</sup> Such views have revealed structural properties of junctions such as coaxial stacking of helices and long-range tertiary interactions<sup>20–23</sup> (see Fig. 1b and c). For instance, Lilley *et al.*<sup>24–26</sup> analyzed the conformations of specific examples of three-way and four-way junctions (junctions composed of three and four helical arms, respectively) in DNA and RNA using fluorescence resonance energy transfer and observed transitional changes and flexibility in their helical configuration under Mg<sup>2+</sup> and Na<sup>+</sup> concentration variations. Lescoute and Westhof compiled and analyzed the topology of three-way junctions in folded RNAs, categorizing these junctions into three families and specifying rules to predict coaxial stacking, which occurs when two separate helical regions stack to form coaxial helices as a pseudo-continuous helix (see Fig. 1c).<sup>27</sup> The loops connecting the stacked helices constrain the configuration space that these helical axes can explore. Laing and Schlick analyzed the topology of solved 3D four-way junctions and grouped them into nine families according to coaxial stacking interactions and helical conformation signatures.<sup>28</sup> Tyagi and Mathews predicted coaxial stacking based on free-energy minimization.<sup>29</sup> Finally, Bindewald *et al.* developed RNAJunction, a database that contains information on RNA structural elements including junctions.<sup>30</sup>

These combined analyses of RNA structures have unraveled recurrent structural motifs across a variety of RNA molecules. Our previous work on annotation and analysis of RNA tertiary motifs<sup>23</sup> based on a representative set of high-resolution RNA structures showed that coaxial helices are abundant tertiary motifs that often cooperate with other long-range interactions such as A-minor motifs to stabilize RNA's structure. Building upon existing work on 3-way and 4-way junctions,<sup>27,28</sup> we extend the analysis here to higher-order junctions (5- to 10-way junctions) and combine our findings to describe common motifs, including recurrent helical configurations, occurring across all junctions found in solved structures, regardless of their degree of branching. Our analysis reveals

novel interaction motifs formed between perpendicular alignments of helices as well as common internal base pairs that help form long-range interactions. We also discuss how junctions arrange their helical arms in similar configurations, regardless of their degree of branching. Statistical data showing base pair and base stacking preferences are also reported.

## Results

Network interaction diagrams (see Fig. 1b) indicating base-pair interactions have proven useful in understanding RNA tertiary motifs<sup>31–33</sup> and in investigating the topology of three- and four-way junctions.<sup>27,28</sup> Here, we extend such analyses to higher-order junctions from degrees 5 to 10. We begin with a description of the higher-order junctions using network interaction diagrams. For clarification, we label and color code helices sequentially according to the 5'-to-3' orientation of the entire RNA as shown in Fig. 1a. A helix here is required to contain at least two consecutive Watson–Crick (WC) base pairs G–C, A–U, and G–U. The single-stranded region between each pair of consecutive helices  $H_i$  and  $H_{i+1}$  is labeled by  $J_{i/i+1}$ . Each junction element is labeled by its Protein Data Bank (PDB) code<sup>34</sup> followed by the first residue number of the first helix  $H_1$  in the junction. The point where strands cross over is called the point of strand exchange. We use the Leontis and Westhof notation<sup>35,3</sup> to study base-pair interactions occurring within junctions and to describe common motifs. Our list of 207 junctions contains junctions of degree 3 to 10 (see Table 1 and Tables S1 and S2) and has been assembled by taking all high-resolution RNA structures from the PDB database<sup>34</sup> as of April 2009.

In our previous analysis of four-way RNA junctions,<sup>28</sup> we identified nine broad four-way junction families according to coaxial stacking patterns and helical configurations (Fig. 2). Helices within these junctions stabilize their conformations using common tertiary motifs such as coaxial stacking, loop–helix interaction, and loop–loop interactions. Novel interactions involving A-minor motifs and coaxial stacking were revealed repeatedly at the point of strand exchange in many elements within families *cH*, *cL*, and *cK* (Fig. 2b–d). In our analysis of higher-order junctions, we find more disorder in the organization of their components. Still, similar to three-way and four-way junctions, helices tend to arrange locally in parallel and perpendicular patterns. Similar repeating motifs such as the A-minor interactions and the sarcin/ricin-like motifs are also commonly encountered.

### Higher-order junctions

Due to the small number of examples available for higher-order junctions (Fig. 3), it is not possible to design a classification scheme similar to the families

**Table 1.** List of RNA 3D structures containing 23 five-way junctions, 9 six-way junctions, 4 seven-way junctions, 1 nine-way junction, and 2 ten-way junctions

Name	Degree	RNA type	Coaxial stacks	Helical alignments	Nomenclature	Domain	Helix numbers
1U6B_8	5WJ	GI intron <i>Azoarcus</i>	H <sub>1</sub> H <sub>2</sub> , H <sub>3</sub> H <sub>4</sub>		HS <sub>3</sub> HS <sub>3</sub> 2HS <sub>13</sub> HS <sub>2</sub>		P2-3-8-7.1-10
1Y0Q_43	5WJ	GI intron <i>Twort</i>	H <sub>2</sub> H <sub>3</sub> , H <sub>4</sub> H <sub>5</sub>		HS <sub>4</sub> 2HS <sub>10</sub> HS <sub>3</sub> HS <sub>5</sub>		P3-4-6-7.1-7.2
1S72_238	5WJ	23S rRNA <i>H. marismortui</i>	H <sub>1</sub> H <sub>2</sub> , H <sub>3</sub> H <sub>5</sub>		HS <sub>4</sub> HS <sub>5</sub> HS <sub>6</sub> HS <sub>6</sub> HS <sub>2</sub>	I	H14-15-16-21-22
2J01_267	5WJ	23S rRNA <i>T. thermophilus</i>	H <sub>1</sub> H <sub>2</sub> , H <sub>3</sub> H <sub>5</sub>		HS <sub>3</sub> HS <sub>5</sub> HS <sub>5</sub> HS <sub>6</sub> HS <sub>2</sub>	I	H14-15-16-21-22
2BTE_6	5WJ	Leuyl-tRNA <i>T. thermophilus</i>	H <sub>1</sub> H <sub>5</sub> , H <sub>2</sub> H <sub>3</sub>		HS <sub>2</sub> 3HS <sub>2</sub> H		H1-2-3-4-5
2NR0_6	5WJ	Leucyl-tRNA <i>E. coli</i>	H <sub>1</sub> H <sub>5</sub> , H <sub>2</sub> H <sub>3</sub>		HS <sub>2</sub> HS <sub>4</sub> HS <sub>1</sub> HS <sub>2</sub> H		H1-2-3-4-5
2J01_45	5WJ	23S rRNA <i>T. thermophilus</i>	H <sub>2</sub> H <sub>5</sub> , H <sub>3</sub> H <sub>4</sub>		HS <sub>6</sub> HS <sub>2</sub> HS <sub>1</sub> HS <sub>1</sub>	I	H4A-5-8-9-10
2AW4_46	5WJ	23S rRNA <i>E. coli</i>	H <sub>2</sub> H <sub>5</sub> , H <sub>3</sub> H <sub>4</sub>		HS <sub>6</sub> HS <sub>2</sub> HS <sub>1</sub> HS <sub>1</sub>	I	H4A-5-8-9-10
1NKW_45	5WJ	23S rRNA <i>D. radiodurans</i>	H <sub>2</sub> H <sub>5</sub> , H <sub>3</sub> H <sub>4</sub>		HS <sub>6</sub> HS <sub>2</sub> HS <sub>1</sub> HS <sub>1</sub>	I	H4A-5-8-9-10
3BWP_23	5WJ	GII intron <i>O. theyensis</i>	H <sub>2</sub> H <sub>3</sub>		HS <sub>3</sub> HS <sub>2</sub> HS <sub>2</sub> HS <sub>3</sub> HS <sub>3</sub>		IA-IB-IC-ID1-I(ii)
2J00_35	5WJ	16S rRNA <i>T. thermophilus</i>	H <sub>3</sub> H <sub>4</sub>		HS <sub>2</sub> HS <sub>2</sub> HS <sub>4</sub> HS <sub>6</sub> HS <sub>2</sub>	5'	H3-4-16-17-18
2AVY_35	5WJ	16S rRNA <i>E. coli</i>	H <sub>3</sub> H <sub>4</sub>		HS <sub>2</sub> HS <sub>2</sub> HS <sub>5</sub> HS <sub>7</sub> HS <sub>2</sub>	5'	H3-4-16-17-18
1NKW_592	5WJ	23S rRNA <i>D. radiodurans</i>	H <sub>4</sub> H <sub>5</sub>		HS <sub>4</sub> HS <sub>2</sub> HS <sub>2</sub> 2HS <sub>5</sub>	II	H26-27-32-36-46
1S72_640	5WJ	23S rRNA <i>H. marismortui</i>	H <sub>4</sub> H <sub>5</sub>		HS <sub>4</sub> HS <sub>2</sub> HS <sub>2</sub> HS <sub>4</sub> HS <sub>10</sub>	II	H26-27-32-36-46
2AW4_583	5WJ	23S rRNA <i>E. coli</i>	H <sub>4</sub> H <sub>5</sub>		HS <sub>4</sub> HS <sub>2</sub> HS <sub>2</sub> HS <sub>2</sub> HS <sub>8</sub>	II	H26-27-32-36-46
2J01_583	5WJ	23S rRNA <i>T. thermophilus</i>	H <sub>4</sub> H <sub>5</sub>		HS <sub>4</sub> HS <sub>2</sub> HS <sub>2</sub> HS <sub>2</sub> HS <sub>8</sub>	II	H26-27-32-36-46
1S72_657	5WJ	23S rRNA <i>H. marismortui</i>	H <sub>4</sub> H <sub>5</sub>		HS <sub>2</sub> HS <sub>3</sub> HS <sub>5</sub> HS <sub>8</sub> HS <sub>3</sub>	II	H27-28-29-30-31
2AVY_57	5WJ	16S rRNA <i>E. coli</i>	H <sub>4</sub> H <sub>5</sub>	H <sub>2</sub> H <sub>3</sub>	HS <sub>2</sub> HS <sub>6</sub> HS <sub>1</sub> HS <sub>1</sub> HS <sub>3</sub>	5'	H5-6-7-13-14
2J00_56	5WJ	16S rRNA <i>T. thermophilus</i>	H <sub>4</sub> H <sub>5</sub>	H <sub>2</sub> H <sub>3</sub>	HS <sub>3</sub> HS <sub>6</sub> HS <sub>1</sub> HS <sub>1</sub> HS <sub>4</sub>	5'	H5-6-7-13-14
1NKW_2036	5WJ	23S rRNA <i>D. radiodurans</i>	H <sub>2</sub> H <sub>3</sub>	H <sub>4</sub> H <sub>5</sub>	HS <sub>9</sub> HS <sub>8</sub> HS <sub>11</sub> HS <sub>7</sub> HS <sub>8</sub>	V	H73-74-89-90-93
1S72_2097	5WJ	23S rRNA <i>H. marismortui</i>	H <sub>2</sub> H <sub>3</sub>	H <sub>4</sub> H <sub>5</sub>	HS <sub>6</sub> HS <sub>8</sub> HS <sub>9</sub> HS <sub>4</sub> HS <sub>4</sub>	V	H73-74-89-90-93
2AW4_2056	5WJ	23S rRNA <i>E. coli</i>	H <sub>2</sub> H <sub>3</sub>	H <sub>4</sub> H <sub>5</sub>	HS <sub>4</sub> HS <sub>8</sub> HS <sub>2</sub> HS <sub>4</sub> HS <sub>4</sub>	V	H73-74-89-90-93
2J01_2053	5WJ	23S rRNA <i>T. thermophilus</i>	H <sub>2</sub> H <sub>3</sub>	H <sub>4</sub> H <sub>5</sub>	HS <sub>9</sub> HS <sub>8</sub> HS <sub>11</sub> HS <sub>7</sub> HS <sub>8</sub>	V	H73-34-89-90-93
2A64_20	6WJ	RNase P type B <i>Bacillus stearothermophilus</i>	H <sub>1</sub> H <sub>2</sub> , H <sub>5</sub> H <sub>6</sub>		HS <sub>1</sub> HS <sub>13</sub> HS <sub>3</sub> HS <sub>3</sub> 2HS <sub>8</sub>		P2-3-5-15-15.1-15.2
1NKW_2056	6WJ	23S rRNA <i>D. radiodurans</i>	H <sub>1</sub> H <sub>2</sub>	H <sub>3</sub> H <sub>6</sub>	HS <sub>2</sub> HS <sub>2</sub> 2HS <sub>2</sub> HS <sub>10</sub> HS <sub>13</sub>	V	H74-75-80-81-82-88
2J01_2073	6WJ	23S rRNA <i>T. thermophilus</i>	H <sub>1</sub> H <sub>2</sub>	H <sub>3</sub> H <sub>6</sub>	HS <sub>2</sub> HS <sub>2</sub> 2HS <sub>2</sub> HS <sub>10</sub> HS <sub>13</sub>	V	H74-75-80-81-82-88
1S72_2114	6WJ	23S rRNA <i>H. marismortui</i>	H <sub>1</sub> H <sub>2</sub>	H <sub>3</sub> H <sub>6</sub>	HS <sub>2</sub> HS <sub>3</sub> HS <sub>2</sub> HS <sub>2</sub> HS <sub>10</sub> HS <sub>10</sub>	V	H74-75-80-81-82-88
2AW4_2073	6WJ	23S rRNA <i>E. coli</i>	H <sub>1</sub> H <sub>2</sub>	H <sub>3</sub> H <sub>6</sub>	HS <sub>2</sub> HS <sub>2</sub> 2HS <sub>2</sub> HS <sub>10</sub> HS <sub>13</sub>	V	H74-75-80-81-82-88
1S72_38	6WJ	23S rRNA <i>H. marismortui</i>	H <sub>2</sub> H <sub>3</sub>		HS <sub>2</sub> HS <sub>4</sub> HS <sub>11</sub> HS <sub>3</sub> HS <sub>3</sub> HS <sub>9</sub>	I	H4-4A-11-12-13-14
2J01_43	6WJ	23S rRNA <i>T. thermophilus</i>	H <sub>2</sub> H <sub>3</sub>		2HS <sub>4</sub> HS <sub>11</sub> HS <sub>3</sub> HS <sub>3</sub> HS <sub>9</sub>	I	H4-4A-11-12-13-14
2AW4_44	6WJ	23S rRNA <i>E. coli</i>	H <sub>2</sub> H <sub>3</sub>		2HS <sub>4</sub> HS <sub>11</sub> HS <sub>3</sub> HS <sub>3</sub> HS <sub>9</sub>	I	H4-4A-11-12-13-14
1NKW_43	6WJ	23S rRNA <i>D. radiodurans</i>	H <sub>2</sub> H <sub>3</sub>		2HS <sub>4</sub> HS <sub>11</sub> HS <sub>3</sub> HS <sub>3</sub> HS <sub>9</sub>	I	H4-4A-11-12-13-14
1NKW_829	7WJ	23S rRNA <i>D. radiodurans</i>	H <sub>2</sub> H <sub>3</sub> , H <sub>6</sub> H <sub>7</sub>	H <sub>1</sub> H <sub>4</sub>	HS <sub>4</sub> 2HS <sub>2</sub> HS <sub>4</sub> HS <sub>3</sub> HS <sub>3</sub> HS <sub>7</sub>	II	H36-37-38-39-40-41-45
2AW4_816	7WJ	23S rRNA <i>E. coli</i>	H <sub>2</sub> H <sub>3</sub> , H <sub>6</sub> H <sub>7</sub>	H <sub>1</sub> H <sub>4</sub>	HS <sub>4</sub> HS <sub>2</sub> HS <sub>3</sub> HS <sub>4</sub> HS <sub>3</sub> 2HS <sub>4</sub>	II	H36-37-38-39-40-41-45
2J01_816	7WJ	23S rRNA <i>T. thermophilus</i>	H <sub>2</sub> H <sub>3</sub> , H <sub>6</sub> H <sub>7</sub>	H <sub>1</sub> H <sub>4</sub>	HS <sub>4</sub> 2HS <sub>2</sub> HS <sub>5</sub> HS <sub>3</sub> 2HS <sub>4</sub>	II	H36-37-38-39-40-41-45

Table 1 (continued)

Name	Degree	RNA type	Coaxial stacks	Helical alignments	Nomenclature	Domain	Helix numbers
1S72_909	7WJ	23S rRNA <i>H. marismortui</i>	H <sub>2</sub> H <sub>3</sub> , H <sub>6</sub> H <sub>7</sub>	H <sub>1</sub> H <sub>4</sub>	HS <sub>4</sub> HS <sub>2</sub> HS <sub>3</sub> HS <sub>4</sub> HS <sub>3</sub> 2HS <sub>4</sub>	II	H36–37–38–39–40–41–45
2J01_6	9WJ	23S rRNA <i>T. thermophilus</i>	H <sub>4</sub> H <sub>5</sub>		HS <sub>7</sub> HS <sub>7</sub> HS <sub>18</sub> HS <sub>7</sub> HS <sub>11</sub> H S <sub>2</sub> HS <sub>4</sub> HS <sub>19</sub> HS <sub>5</sub>	I	H1–2–25–26–47– 72–73–94–99
1NKW_7	10WJ	23S rRNA <i>D. radiodurans</i>	H <sub>4</sub> H <sub>5</sub> , H <sub>9</sub> H <sub>10</sub>		HS <sub>8</sub> HS <sub>9</sub> HS <sub>18</sub> HS <sub>7</sub> HS <sub>11</sub> H S <sub>2</sub> HS <sub>2</sub> HS <sub>2</sub> HS <sub>4</sub> HS <sub>4</sub>	I	H1–2–25–26–47– 72–73–94–98–99
2AW4_7	10WJ	23S rRNA <i>E. coli</i>	H <sub>4</sub> H <sub>5</sub> , H <sub>9</sub> H <sub>10</sub>		HS <sub>8</sub> HS <sub>7</sub> HS <sub>18</sub> HS <sub>7</sub> HS <sub>11</sub> H S <sub>2</sub> HS <sub>4</sub> HS <sub>2</sub> HS <sub>5</sub> HS <sub>5</sub>	I	H1–2–25–26–47– 72–73–94–98–99

The name describes the PDB code and the number of the first residue of helix H1 in the junction. The nomenclature is based on Ref. 1, and the helices are numbered according to the scheme in Leffers *et al.*<sup>36</sup> A single line between rows separates junctions with the same number of coaxial helices. A double line between rows separates the junction's degree of branching.

assigned in junctions with small degrees.<sup>27,28</sup> However, a number of recurrent interaction patterns and motifs can be observed, and their helical elements can be organized using coaxial stacking patterns and other helical arrangements as described below.

#### Five-way junctions

Five-way junctions resemble lower-order junctions in terms of their helical arrangements. For instance, Fig. 4a–c shows junction diagrams with two coaxial stacking interactions (seen as aligned colored helices) analogous to families in four-way junctions.<sup>28</sup> Specifically in Fig. 4a, a junction found in the *Azoarcus* intron<sup>37</sup> contains all its helical axes aligned roughly in a coplanar and parallel arrangement and stabilized by long-range interactions, forming a crossing at the point of strand exchange similar to elements in the four-way junction family *cH*. A-minor interactions<sup>38</sup> (denoted by open and filled triangles known as Sugar–Sugar interactions) are the most conserved interactions responsible for such crossings. Similarly, the junction 2BTE\_6 in Fig. 4b corresponds to the transfer RNA, where four helices form the well-known “L” shape while an extra helix bulges out of the “L” shape. Also of interest, both Fig. 4b and c contain junction

examples with a pair of perpendicular coaxial stacking interactions. While the pattern in Fig. 4b is a coaxial stacking produced between consecutive helices, that in Fig. 4c is a coaxial stacking between pairs of nonconsecutive helices (H<sub>2</sub>H<sub>5</sub> and H<sub>3</sub>H<sub>5</sub> for each case). Thus, coaxial stacking interactions are not exclusively formed between neighboring helices.

Figure 4d–f shows junction diagrams with one coaxial stacking perpendicular to at least one helix. Specifically, Fig. 4d illustrates a junction with one coaxial stack and one helical alignment (helices aligned without stacking interactions) arranged in a perpendicular configuration. As observed in three- and four-way junctions, such perpendicular arrangements among helices are stabilized by loop–loop interactions (2BTE\_6 in Fig. 4b and 2AVY\_57 in Fig. 4d), loop–helix interactions (2J01\_45 in Fig. 4c), or helix–helix interactions (1S72\_657 and 2AVY\_35 in Fig. 4f). Loop–loop interactions typically involve Hoogsteen or Sugar–edge interactions but can also involve WC base pairs. Loop–helix interactions primarily involve Sugar–Sugar interactions forming A-minor motifs. Helix–helix interactions involve minor groove interactions and will be discussed below in more detail. Junction diagrams in Fig. 4f resemble family *cK* of four-way junctions, which are composed of one coaxial stacking between two helices, while a third helix aligns perpendicular to

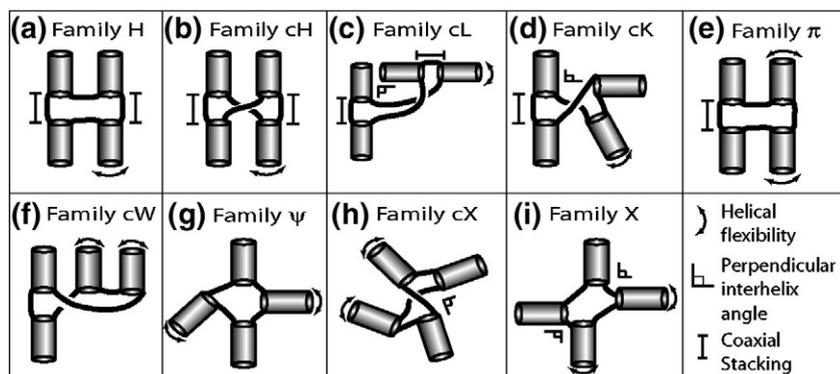
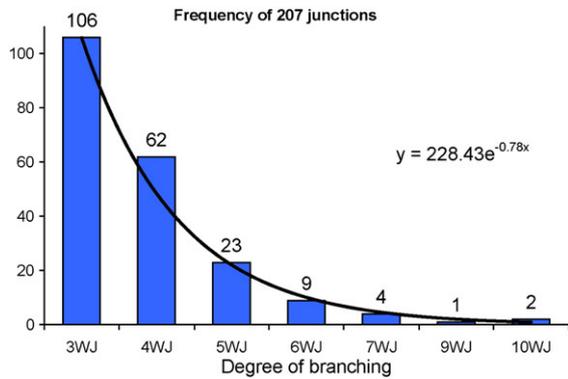


Fig. 2. Classification of RNA four-way junctions into nine families according to their coaxial stacking properties, perpendicular helical configurations, and flexible helical arms (see inset box). (a)–(c) consist of junction families with two coaxial helices sorted by their interhelical angles. (d) and (e) consist of junction families containing one coaxial stacking, while (f)–(i) contain no coaxial stacking but can be characterized by helical alignments and perpendicular configurations stabilized by key tertiary motifs.



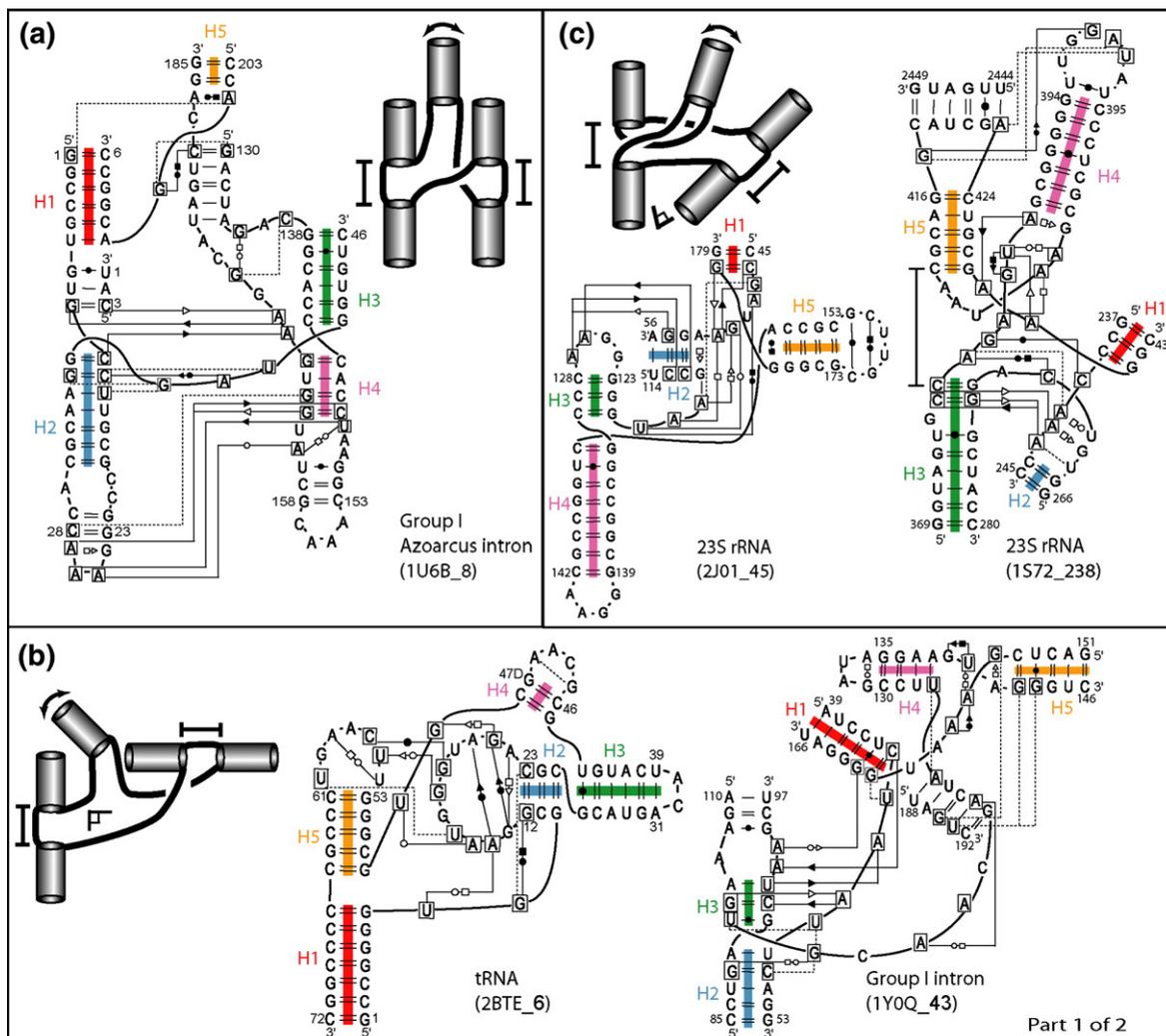
**Fig. 3.** Histogram from a total of 207 RNA junctions sorted by branching degree ranging from 3 (3WJ) to 10 (10WJ).

the coaxial stack. The remaining two helices are arranged based upon the length of their flanking loop elements.

### Six- to 10-way junctions

In contrast to the compact globular shapes that many protein structures have, RNA molecules prefer rather compact prolate ellipsoidal shapes.<sup>14,39</sup> This property reflects the way junctions form by keeping most of their helical axes roughly coplanar. Compared to junctions with a low degree of branching, higher-order junctions are more disordered in the organization of their components; still, the basic helical arrangements such as coaxial stacking (present in every high-order junction), parallel, and perpendicular helical axes are retained, as described next.

Figure 5a shows a six-way junction from the ribonuclease P, forming a coaxial helix  $H_1H_2$  and helices  $H_3$  and  $H_4$  in a plane, with the coaxial helix  $H_5H_6$  leaving this plane. The conformation produced by coaxial helices  $H_1H_2$  and  $H_5H_6$  is similar to the antiparallel conformation found in the four-way junction in the hairpin ribozyme.<sup>40</sup> The diagram in Fig. 5b shows a six-way junction with the



**Fig. 4.** Network interaction diagrams of five-way junctions sorted by coaxial stacking and helical configurations. The network symbology follows the Leontis–Westhof notation<sup>3</sup> (see inset boxes). (a)–(c) show junction diagrams with two coaxial stacks aligned either in parallel (a) or perpendicular to each other (b and c), while (d)–(f) show junction diagrams with one coaxial stack perpendicular to at least one helical arm.

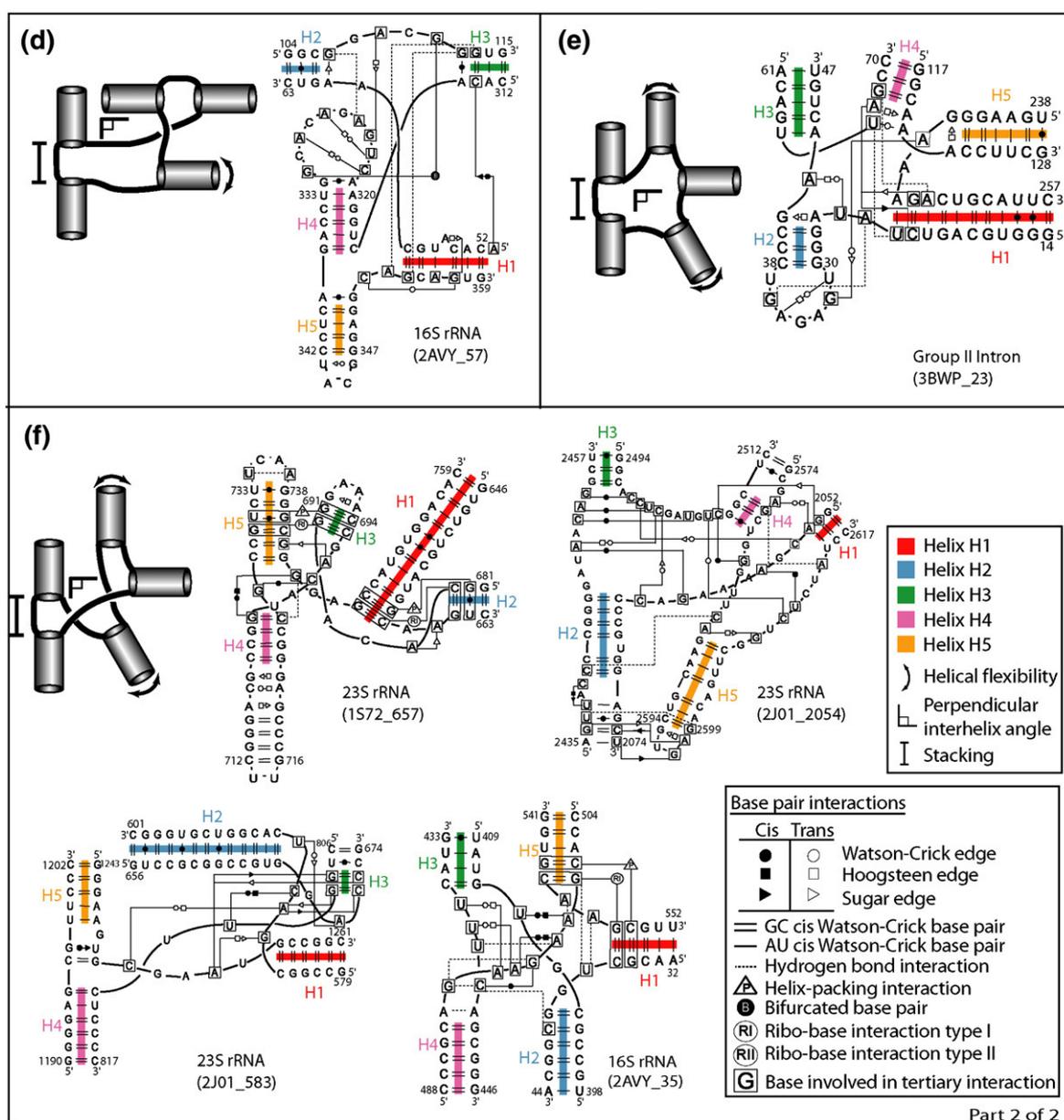


Fig. 4 (legend on previous page)

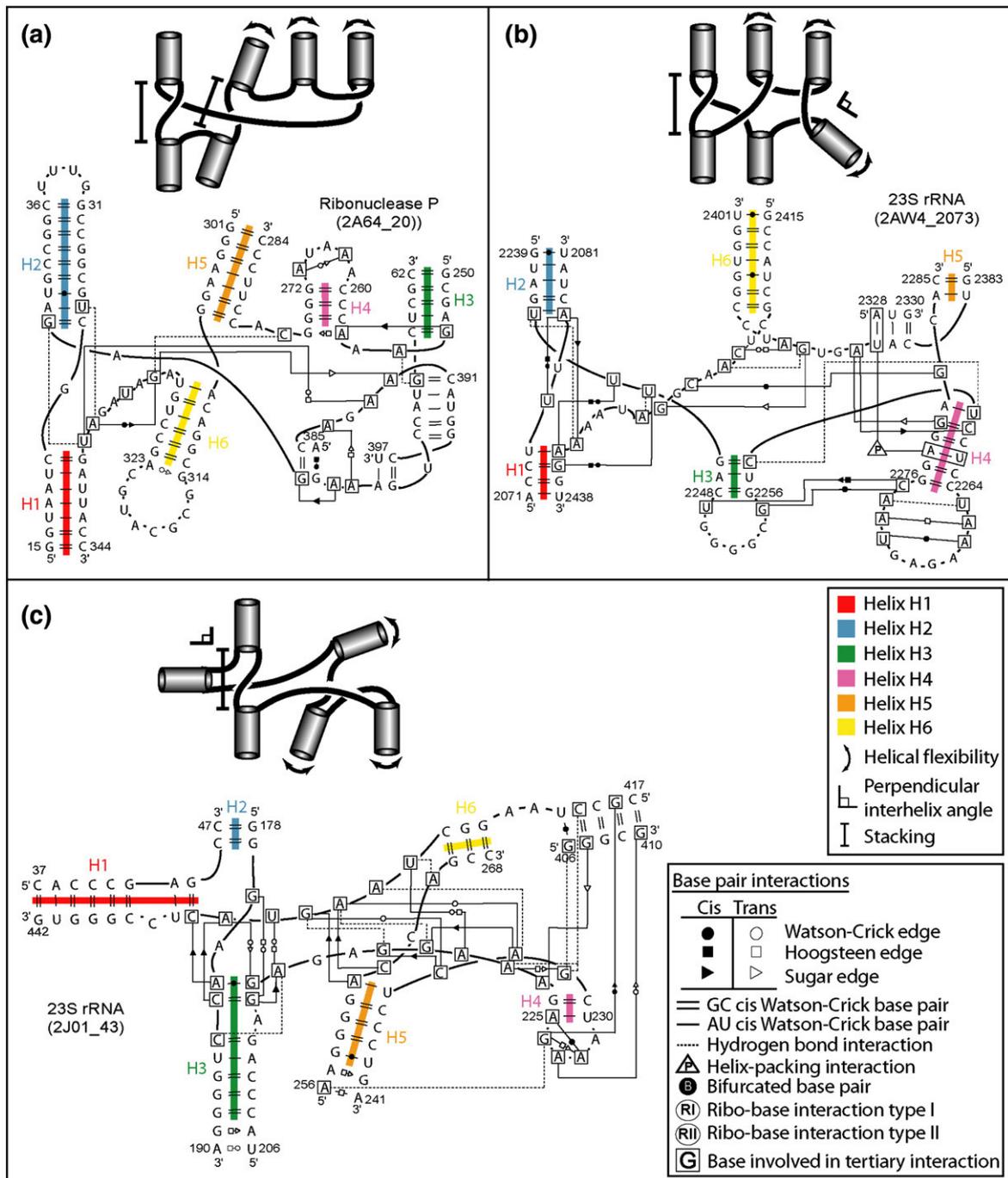
helical axes in a plane. The single-strand  $J_{5/6}$  contains nucleotides 2385–2387 base pairing with a hairpin loop, forming a (pseudoknot) helix perpendicular to  $H_4$ . The homologous six-way junctions found in the *Haloarcula marismortui* and *Thermus thermophilus* (1S72\_2114 and 2J01\_2073 in Table 1) show helices  $H_3$  and  $H_6$  aligned.

In Fig. 5c, the six-way junction 2J01\_43 contains helices  $H_1$ – $H_3$  arranged by forming a coaxial helix  $H_2H_3$ , which aligns perpendicular to  $H_1$ , in a similar conformation to members of family A in three-way junctions such as the M-box riboswitch (2QBZ\_53 in Table S1) and three-way junctions found in the large ribosomal subunit (1S72\_51, 1S72\_1403, and 1S72\_2130 in Table S1).

The seven-way junction in Fig. 6a is formed by three coaxial helices aligning their axes more or

less in a plane. The coaxial stacking between non-neighboring helices  $H_1$  and  $H_4$  is due to a sarcin/ricin motif<sup>41</sup> formed between strands  $J_{1/2}$  and  $J_{7/1}$ . The pair of coaxial helices  $H_2H_3$  and  $H_1H_4$  aligns similar to family *cH* in four-way junctions,<sup>28</sup> where a crossing occurs at the point of strand exchange caused by A-minor interactions. At the same time, the pair of coaxial helices  $H_1H_4$  and  $H_6H_7$  aligns similar to family *H* with its extra helix  $H_5$  in between. Helices  $H_1$  and  $H_3$  arrange perpendicular to each other.

The 9- and 10-way junctions shown in Fig. 6b and c correspond to the central junction connecting all domains in the 23S rRNA. The 10-way junction contains an extra helix presumably formed through evolutionary variation. Note that in both cases, the strand  $J_{3/4}$  forms a “helical region” composed of

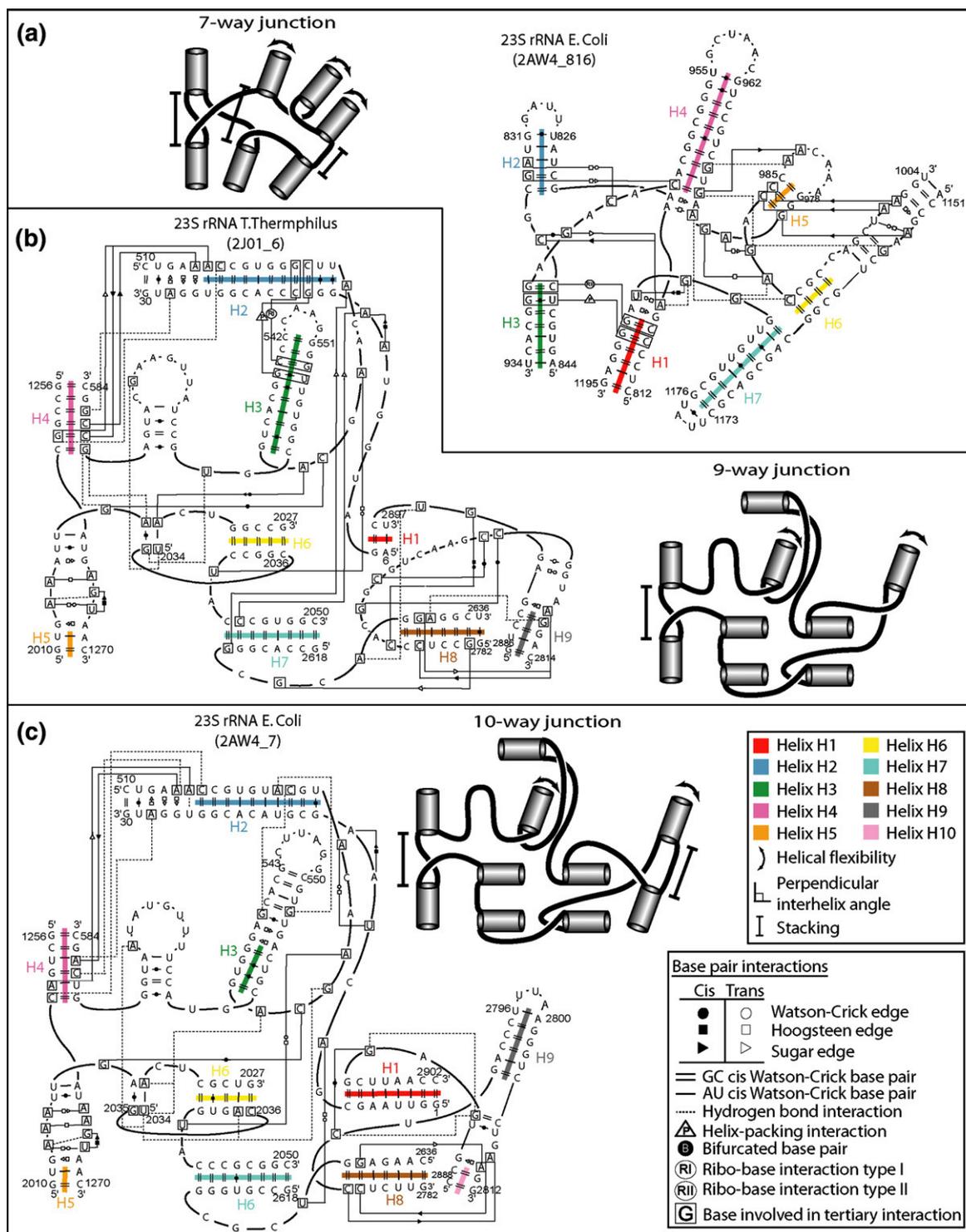


**Fig. 5.** Network interaction diagrams of six-way junctions sorted by coaxial stacking and helical configurations. The network symbology follows the Leontis–Westhof notation<sup>3</sup> (see inset boxes). (a) Junction diagram with two coaxial stacks; (b and c) junction diagrams with one coaxial stack and one perpendicular helical alignment.

alternating WC canonical and non-WC base pairs. Our definition of a helix requires at least two consecutive WC canonical base pairs to be formed; therefore, this region is considered as a strand. Both junctions are nonplanar due to the high degree of branching and form three small globular helical regions. The first region is composed of helices H<sub>1</sub> and H<sub>8</sub>–H<sub>9</sub> (and H<sub>10</sub> for the 10-way junction) arranging similarly to family *cK* in four-way junctions.<sup>28</sup> Helices H<sub>2</sub>, H<sub>3</sub>, H<sub>6</sub>, and H<sub>7</sub> align similar

to family *X* in four-way junctions. The third region is the coaxial helix between H<sub>4</sub> and H<sub>5</sub>.

Another common characteristic of higher-order junctions is that long, single-stranded elements occur to reduce steric clashing caused by junctions with many helical arms, while preserving the preferred prolate and ellipsoid shapes of RNA 3D structures. The single strands connecting two helices often traverse or “jump over” a third helix in between as it occurs in the strand J<sub>3/4</sub> shown in Fig. 5c.



**Fig. 6.** Network interaction diagrams of 7–10 order junctions. (a) 7-way junction, (b) 9-way junction, and (c) 10-way junction. The network symbology follows the Leontis–Westhof notation<sup>3</sup> (see inset boxes).

Moreover, these single strands interact with several junction components while traversing as in the example 2AVY\_35 in Fig. 4f. Here, the strand  $J_{4/5}$  connecting helices  $H_4$  (magenta) and  $H_5$  (orange) interacts with  $J_{3/4}$  and with itself and then interacts with  $J_{2/3}$  and finally with  $J_{5/1}$ . These longer strands between helices allow frequent formation of pseudoknots (Figs. 5a and b and 6b and c). Other

properties of higher-order junctions that are shared by junctions with lower degrees are described in the following sections.

### Statistical features in RNA junctions

From our data set of 207 RNA junctions listed in Table 1 and Tables S1 and S2, more than half are

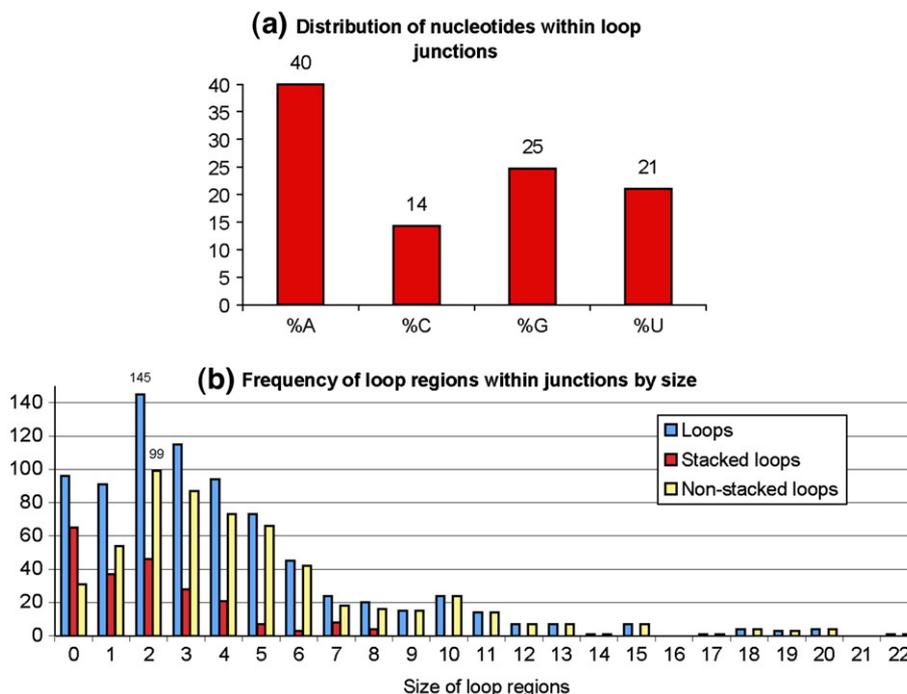
three-way junctions, and the number decreases as the degree of branching increases. Figure 3 shows that the frequency of junctions arranged by degree of branching can be estimated by the exponential function  $y = 228.4e^{-0.78x}$  ( $R^2 = 0.94$ ), but it is not clear how this estimate will change with increased RNA structures. Junctions of higher degree of branching are observed in RNAs of larger size such as the ribonuclease, group II intron, and ribosomal RNA. In contrast, junctions with a small degree of branching occur in a wide range of RNAs, from riboswitches to ribosomal RNAs.

The loop (single-stranded) regions connecting helical elements in junctions are composed by uneven proportions of nucleotide composition as shown in Fig. 7a. While a low percentage of Cs (14%) can be noted, loop regions are strikingly A-rich (40%) for two reasons: A-minor interactions are important in stabilizing helical arms, and adenines offer flexibility to the loop regions. Conversely, the lower concentration of Cs in loop regions corresponds to the smaller number of non-WC base pairs known involving cytosine; however, a reasonable number of these Cs (14%) participate in pseudoknot formation or WC GC base pairs between loops within the junction. In addition, the concentration of WC base pairs near the end of helices (first and second position) produces a high concentration of GC (73%) base pairs, compared to lower AU (20%) and GU wobble (7%) base pairs (data not shown); this might be explained by the high stability (three hydrogen bonds) of GC base pairs.

Figure 7b describes the distribution of the loop size for all loops within helical junctions (blue),

loops between coaxial helices (stacked loop, shown in red), and loops between helices where no coaxial stacking is present (non-stacked loops shown in yellow). In general, a large number of loops range in size from 0 to 6 with a peak at 2, while the less frequent cases are loops of sizes 14 to 22. Figure 7b also shows (in red) that coaxial stacking occurs preferentially in helices adjacent to loops of smaller size, and no stacking is observed for helices between loops of size greater than 8. Coaxial stacking of helices adjacent to loops of size 6 or 7 occurs often due to many noncanonical base-pair interactions, which, in turn, stack with such helices, or also due to the presence of pseudoknots forming at the loop. While a preference for coaxial stacking formation between loops of small size can be noted, there are several cases in which helices with a small loop size do not stack. Particularly, Fig. 7b shows a peak at 2, corresponding to loops between non-stacked helices (99 out of 143). Many reasons could explain the absence of coaxial stacking in these cases, for example, the influence of external forces such as pseudoknot formation, long-range tertiary interactions, and protein binding.

In agreement with work by Elgavish *et al.*, non-canonical base pairs involving AG occur frequently at the end of helices, particularly a trans AG Hoogsteen–Sugar or cis AG Watson–Watson base pairs.<sup>42</sup> These, along with standard WC GC base pairs forming a pseudoknot, are the most frequent interactions observed at the end of helices in junctions. When a noncanonical base-pair AG trans Hoogsteen–Sugar is formed, it often stacks to a trans AU Hoogsteen–Watson base pair. These two base pairs are recurrent



**Fig. 7.** Junction statistics. (a) Proportion of nucleotides at the single-stranded (loop) elements within junctions. (b) Frequency distribution of loop regions within junctions arranged by size. Values for any loop, for loops between coaxial stacking, and for loops between helices forming no coaxial stacking are given in blue, red, and yellow, respectively.

interactions observed in many junctions and become parts of larger 3D motifs such as the sarcin/ricin<sup>32,41</sup> or UA-handle motifs.<sup>43</sup> However, they can also form as independent and stable sub-motifs, often binding to RNA or proteins and assisting in the formation of coaxial stacking between helices.

Other important base-pair interactions found in junctions are the Sugar–Sugar base pairs, which can form A-minor motifs<sup>38</sup> and often combine with coaxial helices forming higher-order motifs<sup>23</sup> (A-minor/coaxial helix). In addition, when long-range interactions occur in junctions, a vast majority of A-minor motifs are formed between loop regions flanking helices (e.g., hairpins and internal loops), while the helical receptors are located near the end of helices.<sup>23</sup> Other base-pair interactions also occur and are composed mostly of purine–purine interactions. Long-range interactions such as A-minor are important elements because they stabilize helical arms in junctions and allow the proper function of RNA molecules.

### Ribo–base interactions stabilize perpendicular helical conformations

One of the most common elements in the ribosome, highlighted by the authors who solved the structures, is the interaction of RNA double helices via minor grooves. Examples of such interactions are A-minor,<sup>38</sup> ribose zipper,<sup>44</sup> G-ribo,<sup>45</sup> and along-groove packing motif (AGPM),<sup>46,47</sup> also known as p-interaction.<sup>48</sup> The interactions presented here describe yet another strategy used for packing minor grooves of rRNA helices against each other.

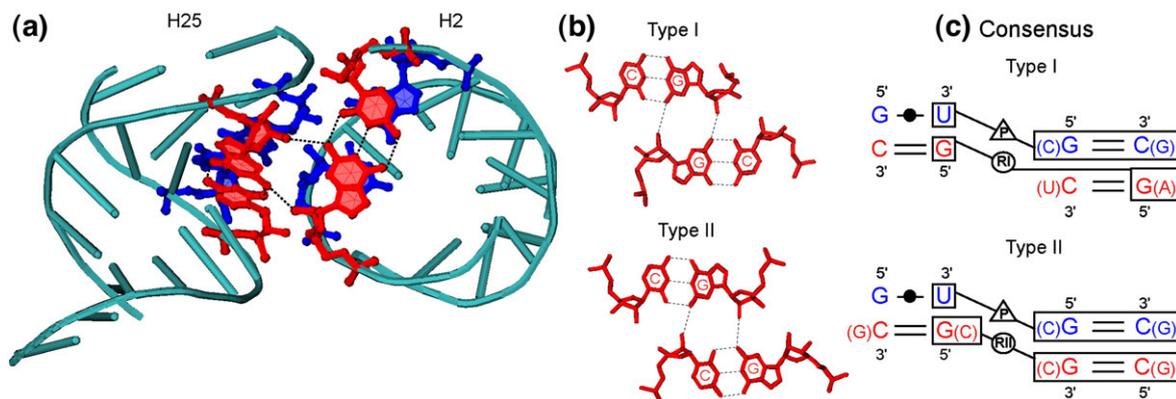
Helices in junctions often align their axes more or less perpendicular to each other via helix–helix interactions along their minor grooves (Fig. 8a). Because the minor groove in A-RNA has a slightly concave shape, the sugar-phosphate backbone of each helix can pack along the minor groove of the other helix. We previously reported perpendicular interactions in four-way junctions where the AGPM

motif is present<sup>28</sup> (GU–WC interaction in blue shown in Fig. 8a). A full analysis based on all junctions allows us to recognize two new interactions that often cooperate with AGPM motifs. The combined interactions are composed of four WC base pairs, forming an angle of approximately 60° between their corresponding base-pair planes and occurring when helices are closely packed. Because these new interactions involve ribose–base interactions, we denote them as ribo–base type I (RI) and ribo–base type II (RII) interactions (see Fig. 8).

The ribo–base type I is characterized by a 2-fold symmetry between two canonical WC base pairs connected by hydrogen bond interactions between the O2' of a G residue of the first base pair and an N2 of a G (or N3 of an A) residue of the second base pair and between O2' of a G (or A) residue of the second base pair and N2 of a G residue of the first base pair (see Fig. 8b). Ribo–base type I occurs between a G of the first base pair and a purine (A or G) of the second base pair. Interestingly, when it appears next to an AGPM motif, a WC CG appears stacked below the WC GU wobble base pair. Indeed, this base-pair signature is even more conserved than the WC GC receptor of the GU wobble in the AGPM motif (Table S3).

The ribo–base type II consists of a roto-reflection symmetry (rotation by 180° followed by a reflection around its axis) where two WC base pairs interact by hydrogen bonds between the O2' of a G residue of the first base pair and an N2 of a G (or N3 of an A) residue of the second base pair and between O2' of a C or U residue of the second base pair and N2 of a G residue of the first base pair (Fig. 8b). When appearing next to the AGPM motif, the CG base pair stacked below the GU base pair can be replaced by a GC base pair, as long as a substitution from CG to GC (or AU) on the receptor base pair of the second stack occurs (see Table S3).

We found 45 instances of ribo–base interactions, mostly located in homologous regions of the ribosomal RNAs considered, and most of them



**Fig. 8.** (a) Perpendicular alignment between helices H2 and H25 in the *T. thermophilus* 23S rRNA structure (PDB code: 2J01). Residues in blue correspond to the AGPM motif (G539–U554 and G17–C523), and residues in red correspond to ribo–base interaction type I (C540–G553 and C18–G522). (b) Ribo–base interaction types I and II. (c) Consensus motif for the perpendicular interaction of helices composed of four stacked base pairs at each helix.

form next to the AGPM motif. While most cases occur between helical elements in junctions, other instances also occur in pseudoknots or near internal loops. Sequence and secondary-structure signature consensus elements for these motifs are shown on Fig. 8c, where the ribo–base interactions appear next to AGPM. There are, however, cases where AGPM motifs with no ribo–base interaction appear or ribo–base interactions in non-AGPM patterns. These cases usually occur when WC base pairs are replaced by other base pairs such as cis Watson–Watson AG or when the GU wobble is replaced by a WC AU base pair (Table S3). Furthermore, crystallographic data from a hammerhead ribozyme (PDB code: 1HMH) and tRNA-Gly (PDB code: 1VAL) show type I and type II interactions forming between a pair of helices, which are possibly tightly packed during the formation of the crystal.

Analogous to AGPM motifs,<sup>46–48</sup> ribo–base interactions bring together helical elements and stabilize RNA molecule for proper function. Another possible role is to act as a mechanism for promoting RNA–protein interactions of neighboring purine nucleotides. Klein *et al.* reported that proteins L18e and L15 in the *H. marismortui* have a high structural homology in the C-terminal domains and both interact with the five-way junction 1S72\_657 (Fig. 3f), forming a near-identical nucleotide and amino acid composition.<sup>12</sup> Both proteins L18e and L15 each interact near ribo–base interactions type I (C658–G747 with C685–G661, and C696–G689 with C741–G730, respectively). A close examination of both cases reveals purine bases that expose their hydrophobic surfaces at the protein–RNA interaction site. In other instances, when pairs of helices are closely packed through AGPM and ribo–base interactions, the AGPM/ribose–base motif appears near the end of helices flanking a trans AG Hoogsteen–Sugar base-pair interactions. This allows these purine bases to expose their hydrophobic surfaces for possible RNA–protein interactions.

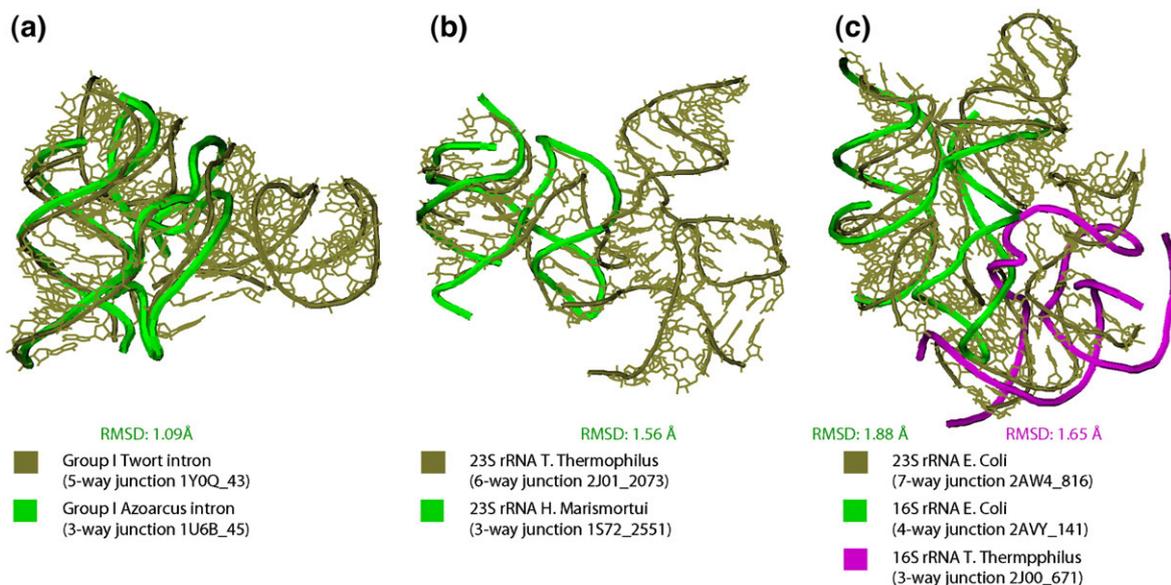
### Folding similarity among junctions with different degrees of branching

With the available 3D structures of large RNA molecules such as ribosomal RNAs,<sup>9–11</sup> group I introns,<sup>37,49,50</sup> and RNase P structures,<sup>51,52</sup> it is now evident that there is a high degree of structural conservation in tertiary structures between homologous RNAs. This fact reflects the similarity among junction architectures despite differences in secondary structure. For instance, Krasilnikov *et al.*<sup>52</sup> reported 3D structural similarities in the S domain of RNase P between an internal loop in RNase P type A and a four-way junction in RNase P type B. Also, most transfer RNA structures are composed of a four-way junction (e.g., 1EFW\_6 in Table S2), but the example shown in Fig. 4b illustrates a tRNA with a five-way junction conformation. Another interesting example is found in the group I introns (see Fig. 8a), where a three-way junction (1U6B\_45 in Table S1) in the *Azoarcus* intron<sup>37</sup> and a five-way

junction (see 1Y0Q\_43 in Table 1 and Fig. 4b) in the *Twort* intron<sup>49</sup> align their corresponding helices P3, P4, and P6 with a high degree of similarity (RMSD=1.09 Å) despite differences in their secondary structure. This structural similarity is in agreement with the observations that group I introns contain conserved core elements formed by junctions, which provide structural stability with the help of conserved peripheral elements by forming long-range contacts.<sup>53</sup>

Moreover, the modular architecture of folded RNAs implies that distances between interacting parts are conserved in functionally homologous molecules;<sup>33</sup> thus, similarities in junctions can be made apparent by observing network interaction diagrams and their 3D motifs. For example, in the large subunit of the ribosomal RNA, a five-way junction in *H. marismortui* (see 1S72\_657 in Table 1 and Fig. 4f) is structurally similar to the four-way junctions found in homologous counterparts in *T. thermophilus*, *Escherichia coli*, and *Deinococcus radiodurans* (2J01\_600, 2AW4\_600, and 1NKW\_608 in Table S2). In all cases, four helices interact in pairs via perpendicular motifs caused by ribo–base interactions with AGPM. Similarly, the core junctions whose diagrams are shown in Fig. 6b and c present a highly conserved structural similarity between the 9-way junction found in the *T. thermophilus* and the 10-way junctions found in both the *E. coli* and *D. radiodurans*. These observations suggest that the extra helices that are “left out” might have formed later in evolution for particular advantages in species.

Strikingly, a structural similarity of junctions with diverse degree of branching was also observed in nonhomologous elements where junctions with a larger degree of branching arrange their helical elements to form “sub-junctions” of smaller degrees. For instance, the six-way junction 2J01\_2073 arranges helices H<sub>1</sub>, H<sub>2</sub>, and H<sub>3</sub> locally, similar to three-way junctions of the C family. Elements in family C consist of one coaxial stacking and a helix aligning parallel with the coaxial helix by allowing the single strand connecting the coaxial helix to the parallel helix to structure like a hairpin using the standard U-turn. The six-way junction also forms a U-turn hairpin within the loop J<sub>6/1</sub> between helices H<sub>1</sub> and H<sub>6</sub>. Figure 9b shows a pairwise structural alignment (RMSD=1.56 Å) between this six-way junction and the three-way junction 1S72\_2551 (Table S1) of family C. Similarly, the U-turn hairpin motif is also found in the four-way junction 2AW4\_1832 (Table S2) within the loop J<sub>3/4</sub>, forming a sub-three-way junction element between helices H<sub>2</sub>, H<sub>3</sub>, and H<sub>4</sub> (helices also labeled 65–67 by Leffers *et al.*<sup>36</sup>). Another example is found in helical elements H<sub>1</sub>–H<sub>4</sub> in the seven-way junction, shown in Fig. 6a, which can be decomposed into a four-way junction of the *cH* family<sup>28</sup> while helices H<sub>5</sub>–H<sub>7</sub> can be associated to a three-way junction of the C family<sup>27</sup> as observed in Fig. 9c. Here, both the four-way junction 2AVY\_141 from Table S2 and the three-way junction 2J00\_671 from Table S1 super-



**Fig. 9.** Structural similarity between (a) homologous and (b and c) nonhomologous junctions. (a) Alignment between the *Azoarcus* intron (olive green) and the *Twort* intron (bright green). (b) A six-way junction (olive green) in the 23S rRNA presenting structural similarity to a three-way junction (bright green) in the 16S rRNA. (c) A seven-way junction (olive green) in the 23S rRNA presenting structural similarity to a three-way (magenta) and a four-way (bright green) junction in the 16S rRNA.

imposed with the seven-way junction 2AW4\_816 (RMSD = 1.88 and 1.65 Å, respectively).

## Summary and Discussion

RNA junctions are important structural elements that serve as major architectural components in RNA. While most junctions found in solved crystal structures are formed by a small number of helical branches, higher-order junctions with as many as 10 helices exist. Junctions organize their helical elements using various common interactions, such as long-range interactions, coaxial stacking, and many 3D motifs.

Our analysis of higher-order junctions using network interaction diagrams is a complementary and compatible approach to the classification of RNA three-way and four-way junctions given by Lescote and Westhof<sup>27</sup> and Laing and Schlick,<sup>28</sup> which organize elements according to their helical configurations. Our work also complements other studies. For instance, the SCOR<sup>54</sup> database lists examples of coaxial helices as elements of tertiary motifs. Similarly, RNA junctions contained in the RNAJunction<sup>28</sup> database have been grouped by standard nomenclature<sup>2</sup> based on the size of each loop region. However, similar junctions from homologous RNAs can differ by single insertions or deletions in the loop regions, leading to different classifications under the standard nomenclature.

In the present analysis, we considered higher-order junctions from 5 to 10 helices and compared coaxial stacking and base-pair configuration properties to those noted in lower-order junctions. We described statistical properties of helices and loop

regions for all these RNA junctions and introduced a new motif composed of ribo-base interactions and the AGPM, which is involved in perpendicular helical arrangements. We noted the folding similarity that exists among junctions with different degrees of branching.

In agreement with previous works,<sup>27–29</sup> the data from Fig. 7b indicate a preference for coaxial stacking formation for helices whose common single-stranded loop is small in size. However, there are several cases where helices with a small loop between them do not stack. The reasons for the absence of coaxial stacking are diverse. Often, elements in the loop regions within junctions form noncanonical base pairs, which, in turn, can help reduce the spatial distance between helical arms and facilitate coaxial stacking. In many cases of the family C of the three-way junctions,<sup>27</sup> a small U-turn motif forms at the end of a helix,<sup>27</sup> possibly preventing a coaxial stacking on the capped helix. In addition, proteins can disrupt coaxial helices when their presence alters helical orientations. The four-way junction 1S72\_1743 (Table S2) found in the *H. marismortui* 23S rRNA contains a pair of helices (labeled 62–63 by Leffers *et al.*<sup>36</sup>) with no single-stranded region between them, but the helices are distorted by the protein L19e, thus preventing the formation of coaxial stacking.

Furthermore, in some cases, even if the size of a loop  $J_{i/i+1}$  is small, the size of neighboring loops  $J_{i-1/i}$  and  $J_{i+1/i+2}$  can be equal or smaller, as observed in elements of four-way junction families *H* and *cH*<sup>28</sup> (Table S2). This can lead to an interconversion of stacking conformers or to a competition for coaxial stacking conformers, which can ultimately be decided by long-range interactions.

Indeed, experiments for the hammerhead ribozyme<sup>55</sup> and hairpin ribozyme<sup>56</sup> have shown that loop–loop interactions act as important elements in the function of these ribozymes, by stabilizing the correct conformation of these junctions. In particular, A-minor motifs occurring within the junction (e.g., Fig. 2a and 1S72\_238 from Fig. 2c) help stabilize the structure and avoid interconversion of different configurations.

Although, in general, due to the conformational flexibility and dynamic character of junctions, a continuum of junction conformations might be possible, our compilation of RNA junction domains based on available structures illustrates nature's strong preferences for the arrangement of RNA helical elements in parallel and perpendicular patterns, while keeping the helical axes coplanar. As recently discussed in an essay,<sup>57</sup> most RNA structure and folding data come from *in vitro* experiments, where high ionic concentrations can compensate for the lack of *in vivo* folding factors such as ligands and RNA chaperones. Differences between *in vitro* and *in vivo* folds of RNA are still being investigated. Long-range interactions that stabilize helical elements are very diverse but often involve Sugar–Sugar interactions in the form of A-minor motifs. Other interactions such as base–ribose and long-range stacking interactions are also observed. One advantage of studying junctions with different numbers of helices is that it allows recognition of important repeating motifs such as the Sugar–edge interactions (A-minor), sarcin/ricin, and trans AG Hoogsteen–Sugar interactions. These sets of noncanonical base pairs play important roles in RNA's structure and therefore function.

Ribo–base interactions are novel helix–helix interactions found in perpendicular helical conformations. They belong to the same family of helix packing interactions such as the G-ribo,<sup>45</sup> A-minor,<sup>38</sup> AGPM,<sup>47</sup> and ribose zipper.<sup>44</sup> Because ribo–base interactions often appear next to the AGPM, both motifs form parts of a larger motif (AGPM/ribo–base), whose main function is to pack together helical elements and stabilize RNA molecule for proper function. Such motifs can also act as RNA–RNA or RNA–protein binding promoters by helping their flanking trans AG Hoogsteen–Sugar base-pair interactions to expose their hydrophobic surfaces for binding.

As more interactions involving RNA base and ribose are discovered, one can foresee the need to extend the current RNA base-pair classification given by Leontis and Westhof to include ribose–base interactions.<sup>3</sup>

We encountered many examples of higher-order junctions that arrange their helical elements similar to lower-order junctions. The junction examples belong to both homologous and nonhomologous RNAs. One can then ask: how are higher-order junctions formed? We propose that some junctions with a high degree of branching are formed from insertions and unions of smaller-order junctions under evolutionary pressure; the optimal junction

sites for insertions and unions likely correspond to regions that would not dramatically change its internal tertiary-structure conformation. Our analysis also suggests that higher-order junctions can be decomposed into smaller “sub-junctions”. Ultimately, a better understanding of junction decompositions can help predict RNA 3D structures and functions.

## Materials and Methods

The data set of our 3D RNA junctions was collected from the Research Collaboratory for Structural Bioinformatics PDB.<sup>34</sup> Based on available structures as of April 2009, 554 high-resolution structures were selected, with repetitions omitted by choosing the more recent structures. Junction elements were searched within these and analyzed for base-pair interactions.

To perform our comprehensive search of  $k$ -junctions ( $3 \leq k \leq 10$ ) in the set of RNA structures above, we first considered the secondary structure associated with every 3D structure defined in terms of its canonical WC base pairs and the single-stranded regions. The search for canonical WC base pairs was performed using the program FR3D.<sup>58</sup> Second, we searched for sets of  $k$  distinct strands connecting in a cyclical way by at least two consecutive canonical WC base pairs (Fig. 1a). For simplicity, pseudoknots were automatically removed during the search but later re-inserted for statistical analysis. Visual inspection was also used to verify the correctness of our procedure. In addition, we compared our search outcome to data available from the RNAJunction database,<sup>30</sup> to ensure the verity of all junctions.

Crystal structures containing at least one junction each were identified, 43 in total. The structures include the two high-resolution crystal structures of the 16S (PDB codes: 2AVY and 2J00) and four 23S rRNA (PDB codes: 1NKW, 1S72, 2AW4, and 2J01). Although the 3D shape of equivalent rRNA molecules is highly conserved among species, differences are informative because they help to understand evolutionary changes that nature allows while keeping their molecular function intact. In total, our data set thus contains 207 RNA junctions as listed in Table 1 and Tables S1 and S2. Additional detailed junction information such as PDB source, sequence, and residue numbers are available in Tables S4–S10.

Noncanonical base pairing with alternate hydrogen-bonding patterns occur often in RNA. A consensus between FR3D and RNAVIEW<sup>59</sup> was considered to classify base pairs. Where discrepancies occur, we employed visual programs such as PyMOL (DeLano Scientific LLC) and Swiss-PdbViewer.<sup>60</sup> Additionally, the junction data were analyzed from different perspectives: sequence signatures, length of loop regions, 3D motifs, and the 3D organization of their helices. Orientation aspects such as in coaxial stacking, helices that form perpendicular inter-helical angles, and helices aligning their axes in parallel without the use of stacking forces were analyzed by inspection. Pairwise structure alignment between junction domains was done using the ARTS web server.<sup>61</sup>

Network interaction diagrams describing base-pair interactions are represented symbolically according to the Leontis and Westhof base-pairing classification.<sup>35,3</sup> The diagrams were created using VMD<sup>62</sup> and S2S,<sup>63</sup> a visual aid program based on RNAVIEW.

## Acknowledgements

This work was supported by the Human Frontier Science Program, by a joint National Science Foundation (NSF)/National Institute of General Medical Sciences initiative in Mathematical Biology (DMS-0201160), and by NSF EMT award # CF-0727001. Partial support by the National Institutes of Health (grant # R01-GM055164 and grant # 1 R01 ES 012692) and NSF (grant # CCF-0727001) is also gratefully acknowledged. A.I. was supported by the New York University Dean's Undergraduate Research Fund FAS Frances and Benjamin Benenson Research Scholar, and S.J. was supported by the Sackler Institute Biomedical Science Training Fellowship and a MacCracken Fellowship.

## Supplementary Data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.jmb.2009.07.089](https://doi.org/10.1016/j.jmb.2009.07.089)

## References

- Lilley, D. M., Clegg, R. M., Diekmann, S., Seeman, N. C., von Kitzing, E. & Hagerman, P. (1995). Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB). A nomenclature of junctions and branchpoints in nucleic acids. Recommendations 1994. *Eur. J. Biochem.* **230**, 1–2.
- Lilley, D. M., Clegg, R. M., Diekmann, S., Seeman, N. C., Von Kitzing, E. & Hagerman, P. J. (1995). A nomenclature of junctions and branchpoints in nucleic acids. *Nucleic Acids Res.* **23**, 3363–3364.
- Leontis, N. B. & Westhof, E. (2001). Geometric nomenclature and classification of RNA base pairs. *RNA*, **7**, 499–512.
- Scott, W. G., Murray, J. B., Arnold, J. R., Stoddard, B. L. & Klug, A. (1996). Capturing the structure of a catalytic RNA intermediate: the hammerhead ribozyme. *Science*, **274**, 2065–2069.
- Wilson, T. J., Nahas, M., Ha, T. & Lilley, D. M. (2005). Folding and catalysis of the hairpin ribozyme. *Biochem. Soc. Trans.* **33**, 461–465.
- Batey, R. T., Gilbert, S. D. & Montange, R. K. (2004). Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature*, **432**, 411–415.
- Serganov, A., Yuan, Y. R., Pikovskaya, O., Polonskaia, A., Malinina, L., Phan, A. T. *et al.* (2004). Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem. Biol.* **11**, 1729–1741.
- Kieft, J. S., Zhou, K., Grech, A., Jubin, R. & Doudna, J. A. (2002). Crystal structure of an RNA tertiary domain essential to HCV IRES-mediated translation initiation. *Nat. Struct. Biol.* **9**, 370–374.
- Cate, J. H., Yusupov, M. M., Yusupova, G. Z., Earnest, T. N. & Noller, H. F. (1999). X-ray crystal structures of 70S ribosome functional complexes. *Science*, **285**, 2095–2104.
- Noller, H. F. (2005). RNA structure: reading the ribosome. *Science*, **309**, 1508–1514.
- Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H. & Noller, H. F. (2001). Crystal structure of the ribosome at 5.5 Å resolution. *Science*, **292**, 883–896.
- Klein, D. J., Moore, P. B. & 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.
- Hendrix, D. K., Brenner, S. E. & Holbrook, S. R. (2005). RNA structural motifs: building blocks of a modular biomolecule. *Q. Rev. Biophys.* **38**, 221–243.
- Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905–920.
- Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E. *et al.* (1996). Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science*, **273**, 1678–1685.
- Lipfert, J., Ouellet, J., Norman, D. G., Doniach, S. & Lilley, D. M. (2008). The complete VS ribozyme in solution studied by small-angle X-ray scattering. *Structure*, **16**, 1357–1367.
- Toor, N., Keating, K. S., Taylor, S. D. & Pyle, A. M. (2008). Crystal structure of a self-spliced group II intron. *Science*, **320**, 77–82.
- Walter, F., Murchie, A. I., Duckett, D. R. & Lilley, D. M. (1998). Global structure of four-way RNA junctions studied using fluorescence resonance energy transfer. *RNA*, **4**, 719–728.
- Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Jr, Morgan-Warren, R. J., Carter, A. P., Vonnrhein, C. *et al.* (2000). Structure of the 30S ribosomal subunit. *Nature*, **407**, 327–339.
- Holbrook, S. R. (2005). RNA structure: the long and the short of it. *Curr. Opin. Struct. Biol.* **15**, 302–308.
- Holbrook, S. R. (2008). Structural principles from large RNAs. *Annu. Rev. Biophys.* **37**, 445–464.
- Kim, S. H., Sussman, J. L., Suddath, F. L., Quigley, G. J., McPherson, A., Wang, A. H. *et al.* (1974). The general structure of transfer RNA molecules. *Proc. Natl Acad. Sci. USA*, **71**, 4970–4974.
- Xin, Y., Laing, C., Leontis, N. B. & Schlick, T. (2008). Annotation of tertiary interactions in RNA structures reveals variations and correlations. *RNA*, **14**, 2465–2477.
- Hohng, S., Wilson, T. J., Tan, E., Clegg, R. M., Lilley, D. M. & Ha, T. (2004). Conformational flexibility of four-way junctions in RNA. *J. Mol. Biol.* **336**, 69–79.
- Lilley, D. M. (1998). Folding of branched RNA species. *Biopolymers*, **48**, 101–112.
- Lilley, D. M. (2000). Structures of helical junctions in nucleic acids. *Q. Rev. Biophys.* **33**, 109–159.
- Lescoute, A. & Westhof, E. (2006). Topology of three-way junctions in folded RNAs. *RNA*, **12**, 83–93.
- Laing, C. & Schlick, T. (2009). Analysis of four-way junctions in RNA structures. *J. Mol. Biol.* **390**, 547–559.
- Tyagi, R. & Mathews, D. H. (2007). Predicting helical coaxial stacking in RNA multibranch loops. *RNA*, **13**, 939–951.
- Bindewald, E., Hayes, R., Yingling, Y. G., Kasprzak, W. & Shapiro, B. A. (2008). RNAjunction: a database of RNA junctions and kissing loops for three-dimensional structural analysis and nanodesign. *Nucleic Acids Res.* **36**, D392–D397.
- Lemieux, S. & Major, F. (2006). Automated extraction and classification of RNA tertiary structure cyclic motifs. *Nucleic Acids Res.* **34**, 2340–2346.
- Leontis, N. B., Stombaugh, J. & Westhof, E. (2002).

- Motif prediction in ribosomal RNAs. Lessons and prospects for automated motif prediction in homologous RNA molecules. *Biochimie*, **84**, 961–973.
33. Lescoute, A. & Westhof, E. (2006). The interaction networks of structured RNAs. *Nucleic Acids Res.* **34**, 6587–6604.
  34. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H. *et al.* (2000). The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242.
  35. Leontis, N. B., Stombaugh, J. & Westhof, E. (2002). The non-Watson–Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.* **30**, 3497–3531.
  36. Leffers, H., Kjems, J., Ostergaard, L., Larsen, N. & Garrett, R. A. (1987). Evolutionary relationships amongst archaeobacteria. A comparative study of 23 S ribosomal RNAs of a sulphur-dependent extreme thermophile, an extreme halophile and a thermophilic methanogen. *J. Mol. Biol.* **195**, 43–61.
  37. Adams, P. L., Stahley, M. R., Kosek, A. B., Wang, J. & Strobel, S. A. (2004). Crystal structure of a self-splicing group I intron with both exons. *Nature*, **430**, 45–50.
  38. Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B. & Steitz, T. A. (2001). RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc. Natl Acad. Sci. USA*, **98**, 4899–4903.
  39. Tirumalai, D. & Hyeon, C. (2009). Theory of RNA folding: from hairpins to ribozymes. *Springer Ser. Biophys.* **13**, 27–47.
  40. Pley, H. W., Flaherty, K. M. & McKay, D. B. (1994). Three-dimensional structure of a hammerhead ribozyme. *Nature*, **372**, 68–74.
  41. Leontis, N. B. & Westhof, E. (1998). A common motif organizes the structure of multi-helix loops in 16 S and 23 S ribosomal RNAs. *J. Mol. Biol.* **283**, 571–583.
  42. Elgavish, T., Cannone, J. J., Lee, J. C., Harvey, S. C. & Gutell, R. R. (2001). AA.AG@helix.ends: A:A and A:G base-pairs at the ends of 16 S and 23 S rRNA helices. *J. Mol. Biol.* **310**, 735–753.
  43. Jaeger, L., Verzemnieks, E. J. & Geary, C. (2009). The UA\_handle: a versatile submotif in stable RNA architectures. *Nucleic Acids Res.* **37**, 215–230.
  44. Tamura, M. & Holbrook, S. R. (2002). Sequence and structural conservation in RNA ribose zippers. *J. Mol. Biol.* **320**, 455–474.
  45. Steinberg, S. V. & Boutorine, Y. I. (2007). G-ribo: a new structural motif in ribosomal RNA. *RNA*, **13**, 549–554.
  46. Gagnon, M. G., Mukhopadhyay, A. & Steinberg, S. V. (2006). Close packing of helices 3 and 12 of 16 S rRNA is required for the normal ribosome function. *J. Biol. Chem.* **281**, 39349–39357.
  47. Gagnon, M. G. & Steinberg, S. V. (2002). GU receptors of double helices mediate tRNA movement in the ribosome. *RNA*, **8**, 873–877.
  48. Mokdad, A., Krasovska, M. V., Sponer, J. & Leontis, N. B. (2006). Structural and evolutionary classification of G/U wobble basepairs in the ribosome. *Nucleic Acids Res.* **34**, 1326–1341.
  49. Golden, B. L., Kim, H. & Chase, E. (2005). Crystal structure of a phage Twort group I ribozyme–product complex. *Nat. Struct. Mol. Biol.* **12**, 82–89.
  50. Guo, F., Gooding, A. R. & Cech, T. R. (2004). Structure of the *Tetrahymena* ribozyme: base triple sandwich and metal ion at the active site. *Mol. Cell*, **16**, 351–362.
  51. Kazantsev, A. V., Krivenko, A. A., Harrington, D. J., Holbrook, S. R., Adams, P. D. & Pace, N. R. (2005). Crystal structure of a bacterial ribonuclease P RNA. *Proc. Natl Acad. Sci. USA*, **102**, 13392–13397.
  52. Krasilnikov, A. S., Xiao, Y., Pan, T. & Mondragon, A. (2004). Basis for structural diversity in homologous RNAs. *Science*, **306**, 104–107.
  53. Lehnert, V., Jaeger, L., Michel, F. & 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.
  54. Klosterman, P. S., Hendrix, D. K., Tamura, M., Holbrook, S. R. & Brenner, S. E. (2004). Three-dimensional motifs from the SCOR, structural classification of RNA database: extruded strands, base triples, tetraloops and U-turns. *Nucleic Acids Res.* **32**, 2342–2352.
  55. Penedo, J. C., Wilson, T. J., Jayasena, S. D., Khvorova, A. & Lilley, D. M. (2004). Folding of the natural hammerhead ribozyme is enhanced by interaction of auxiliary elements. *RNA*, **10**, 880–888.
  56. Walter, F., Murchie, A. I., Thomson, J. B. & Lilley, D. M. (1998). Structure and activity of the hairpin ribozyme in its natural junction conformation: effect of metal ions. *Biochemistry*, **37**, 14195–14203.
  57. Cruz, J. A. & Westhof, E. (2009). The dynamic landscapes of RNA architecture. *Cell*, **136**, 604–609.
  58. Sarver, M., Zirbel, C. L., Stombaugh, J., Mokdad, A. & Leontis, N. B. (2008). FR3D: finding local and composite recurrent structural motifs in RNA 3D structures. *J. Math. Biol.* **56**, 215–252.
  59. Yang, H., Jossinet, F., Leontis, N., Chen, L., Westbrook, J., Berman, H. & Westhof, E. (2003). Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res.* **31**, 3450–3460.
  60. Guex, N. & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modelling. *Electrophoresis*, **18**, 2714–2723.
  61. Dror, O., Nussinov, R. & Wolfson, H. J. (2006). The ARTS web server for aligning RNA tertiary structures. *Nucleic Acids Res.* **34**, W412–W415.
  62. Hsin, J., Arkhipov, A., Yin, Y., Stone, J. E. & Schulten, K. (2008). Using VMD: an introductory tutorial. *Curr. Protoc. Bioinformatics*, Chapter 5, Unit 5.7, 5.7.1–5.7.48.
  63. Jossinet, F. & Westhof, E. (2005). Sequence to Structure (S2S): display, manipulate and interconnect RNA data from sequence to structure. *Bioinformatics*, **21**, 3320–3321.