Review TRENDS in Biochemical Sciences Vol.31 No.1 January 2006

Atomic level architecture of group I introns revealed

Quentin Vicens and Thomas R. Cech

Howard Hughes Medical Institute and Department of Chemistry and Biochemistry, University of Colorado, UCB 215, Boulder, CO 80309-0215, USA

Twenty-two years after their discovery as ribozymes, the self-splicing group I introns are finally disclosing their architecture at the atomic level. The crystal structures of three group I introns solved at moderately high resolution (3.1–3.8 Å) reveal a remarkably conserved catalytic core bound to the metal ions required for activity. The structure of the core is stabilized by an intron-specific set of long-range interactions that involves peripheral elements. Group I intron structures thus provide much awaited and extremely valuable snapshots of how these ribozymes coordinate substrate binding and catalysis.

Group I introns in the spotlight

Since the early 1980s, RNA molecules such as ribonuclease P [1] and the self-splicing introns [2–4] have been known to catalyze the cleavage and formation of covalent bonds. Such molecules were called 'ribozymes' to specify that they are enzymes with a catalytic RNA component [2]. Since then, several new classes of ribozyme have been identified in nature – for example, the hammerhead, the hairpin and the hepatitis delta virus ribozymes (reviewed in Refs [5,6]), and the more recently discovered glmS ribozyme [7] – or have been evolved *in vitro* from pools of random RNA sequences (e.g. see Refs [8,9]). Even the ribosome has been shown to be a ribozyme [10].

Self-splicing introns occur naturally in many organisms, including algae, lichens and fungi, as well as in some bacteria, but few have been found in animals [11]. They are not essential for cell viability unlike, for example, the peptidyl transferase center of rRNA [12] and the ribonuclease P responsible for tRNA maturation [13]. Instead, self-splicing introns seem to be selfish genetic elements that have a successful strategy for survival: they paste themselves in and out of various genes and, because they self-splice at the RNA level, they are not deleterious. Self-splicing introns are nonetheless highly instructive models for RNA folding and catalysis. In addition, understanding the function of these molecules helps to address the 'RNA world' hypothesis [14]: the discovery that RNA molecules, like proteins, could possess catalytic properties as well as being able to store genetic information (e.g. as in HIV and the tobacco mosaic, hepatitis C and SARS viruses), implied that an era based solely on RNA could have preceded DNA- and protein-based life on this planet [14,15]. Furthermore, self-splicing introns can be engineered as new tools for gene regulation, discovery and analysis [16]. Modified group I introns can, for example, replace portions of mRNA sequences [17,18], a strategy that could help to repair mutations causing diseases or to reprogram genetic expression.

The most abundant self-splicing introns are the group I introns (>2000 sequences have been found so far [19]), which catalyze a two-step transesterification reaction using a guanosine molecule as a cofactor (Figure 1). They are thus distinguished from the group II introns, which generally use an internal adenosine to initiate self-splicing [5]. The other important feature that distinguishes group I from group II introns and other RNA molecules is their structure. Group I introns are formed by a specific arrangement of about ten Watson–Crick paired elements (indicated by P) or helices that are capped by loops (L) and connected by junctions (J).

Helices P1 to P10 (and the intervening junctions and loops) assemble to form the catalytic core, from which different peripheral elements branch out [20–23]. Helix P1 contains the 5′ substrate strand or '5′ exon' and ends with a conserved G•U pair that contributes to 5′ splice site recognition. Helix P10, which forms after the first step of splicing, involves base pairs between the intron and the 3′ substrate strand or '3′ exon' (Figure 1). Recognition of the 3′ splice site is achieved, in part, by the conserved guanine, termed omega G (ΩG), at the 3′-terminal position of the intron. The active site is located in the vicinity of the P3–P7 pseudoknot and the joining regions J4/5 and J8/7 [24,25]. Even though the core structure is highly conserved, there is very poor sequence conservation apart from a few crucial nucleotides located at the active site [26]. However, patterns of sequence similarity observed at the core of group I introns, as well as common peripheral structures, have facilitated a further classification into 13 subgroups [21,26,27].

To understand how the P1–P10 elements fold in three dimensions so that the ribozyme can perform its precise set of cleavage and ligation reactions, efforts began in the early 1990s to solve the structures of group I introns by X-ray crystallography. It is thus remarkable that, after a decade of attempts, the crystal structures of three different introns were solved to moderately high resolution by independent groups and published separately within a 6-month time frame between July 2004 and January 2005. Group I introns from the purple bacterium

Corresponding author: Cech, T.R. (president@hhmi.org).
Available online 13 December 2005

 0968-0004/$ - see front matter doi:10.1016/j.tibs.2005.11.008

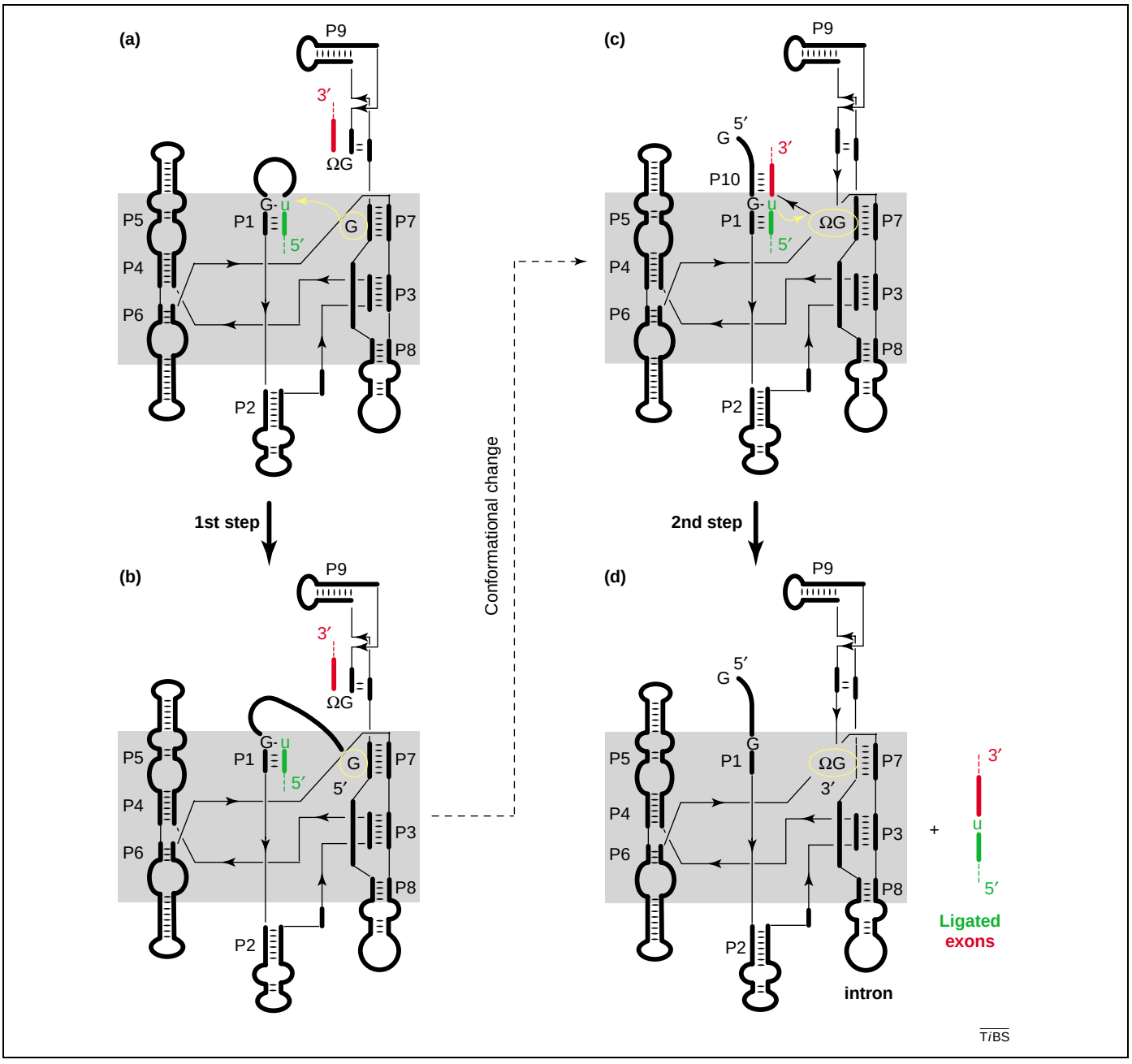

Figure 1. Group I intron self-splicing occurs by a two-step transesterification mechanism. **(a)** The 5′ splice site (marked by a conserved G•U pair, green) undergoes nucleophilic attack (yellow arrow) by the 3′-OH group of a guanosine (or GMP or GTP) cofactor bound to the intron at the G-binding site (ringed in yellow). Lower- and upper-case characters stand for exon and intron sequences, respectively. **(b)** After the reaction, this guanosine is covalently linked to the 5′ end of the intron. **(c)** During a conformational change, the guanosine is displaced from the G-binding site by the 3′-terminal omega G (ΩG) that marks the 3′ splice site (red). The 3′-OH group of the terminal residue of the 5′ exon attacks the 3′ splice site in a reaction that is chemically equivalent to the reverse of step 1. **(d)** The 5′ and 3′ exons are ligated and the intron is released. The group I intron is shown adopting its conserved secondary structure in black; the shaded box delimits its catalytic core (see text and Figure 2b for details).

Azoarcus (*Azo*) [28,29], from the ciliate *Tetrahymena thermophila* (*Tet*) [30] and from the *Staphylococcus aureus* bacteriophage Twort (*Two*) [31] were solved to resolutions of 3.1, 3.8 and 3.6 Å, respectively (Table 1). These recent structures build on the previous *Tet* structures of the non-catalytic P4–P6 domain (160 nt) [32,33] and the P4–P6 domain linked to the catalytic core (247 nt) solved to 5 Å resolution [34], bringing the understanding of self-splicing activity to the atomic level. They capture different snapshots along the splicing pathway that rationalize more than 20 years of biochemical data.

Here, we offer an integrative view of group I intron organization based on comparative structural analysis. Superimposition of the three structures reveals a defined set of structural domains that assemble into a conserved core, which organizes the catalytic site. The structure of the catalytic site is extraordinarily well preserved, as inferred by comparative sequence analysis, but reveals tertiary contacts that could not have been predicted. Moreover, the structures show three distinct ways in which intron-specific peripheral elements establish long-range tertiary interactions that help to stabilize the

Table 1. General information on the *Azo, Tet* and *Two* sequences and structures

	Azoarcus[a]	*Tetrahymena*	Twort
General information			
Origin (gene location)	Purple bacterium ($tRNA^{Ile}$)	*T. thermophila* (26S rRNA)	*S. aureus* bacteriophage (ORF 142)
Subgroup	IC3	IC1	IA2
Length of wild-type intron (nt)	206	413	252
Length of construct (nt)	219 (intron) +3 (5′ exon) +6 (3′ exon)	247	242 (intron) +4 (5′ exon)
Self-splicing step mimicked	Before second step of splicing	Product form with exons released or before docking of substrate	Product form with 5′ exon still bound
G+C content (%)	60	44	37
Refinement parameters			
Resolution (Å)	3.1 (deoxy), 3.4 (ribo)	3.8	3.6
R/R_{free} (%)	24.6/27.9 (deoxy), 26.9/30.7 (ribo)	26.3/32.0	27.9/31.5
Number of molecules per asymmetric unit	1	4	1
Number of nt observed per number of nt in asymmetric unit	222/222	968/988	233/246
PDB ID	1U6B (deoxy), 1ZZN (ribo)	1X8W	1Y0Q

[a]The *Azo* structures containing a deoxy [29] or ribo [35] ΩG are referred to as deoxy or ribo respectively, where information differs between the two. Abbreviations: nt, nucleotide; ORF, open reading frame.

conserved architecture of the core. Finally, the structures allow the long-anticipated catalytic metal ions to be observed in the active site.

Different roads to a common destination

Two significant observations can be made when looking at the *Azo*, *Tet* and *Two* introns and the corresponding constructs that were crystallized. First, the three introns represent 3 of the 13 different group I intron subgroups defined by Michel and Westhof [26]: IC3 (*Azo*), IC1 (*Tet*) and IA2 (*Two*). As such, they offer three different examples of how the global architecture of a group I intron supports a common core structure. This aspect will be emphasized in the penultimate section of the review. Second, the three structures reflect three different solutions to the problem of avoiding RNA self-splicing during transcription and crystallization, which might fortuitously mimic three different steps along the splicing pathway:

(i) The *Azo* construct was designed to represent an intermediate state just before the second step of the self-splicing reaction (Figure 1). To achieve this, four deoxy mutations were introduced at residues in the active site (the last two nucleotides of the intron, the first one of the 3′ exon and the last one of the 5′ exon) to retain both exons in the final construct. The deoxy ΩG, however, favored binding of a monovalent cation in the catalytic site, which is not thought to be biologically relevant [29]. The latest structure by the same laboratory therefore retains only the deoxy mutation at the 3′ end of the 5′ exon, which slows rather than abolishes self-splicing activity [35]. In this structure, an Mg^{2+} ion is now bound to the wild-type ribo ΩG and the intron is fully active in the crystal.

(ii) The *Tet* construct lacks both exons and several helices including helix P1, and the structure shows ΩG bound to the G-binding site [30]. In addition, splicing has been shown to occur on addition of the substrate *in trans* [36]. This structure could therefore reflect either a product formed after exon release (Figure 1) or the intron before docking of its substrate.

(iii) The *Two* construct contains the intron bound to a mimic of the 5′ exon, which was added *in trans* to form the P1 helix. In this structure, ΩG is also bound to the G-binding site and the intron is active when embedded in the 5′ and 3′ exons [31]. Consequently, the *Two* structure might be analogous to the product formed before exon release.

The three structures were solved to moderately high resolution (Table 1), providing electron density for bases, sugar rings, phosphate groups and metal ions. The higher resolution obtained for *Azo* made the refinement of water molecules also possible. All nucleotides are observed in *Azo*, whereas a few residues are missing in the final models of *Tet* and *Two* because they were disordered in the crystal (Table 1). One molecule of the *Azo* or *Two* intron is observed in the asymmetric unit of the crystallographic unit cell, whereas four copies (named A to D) of the truncated but still slightly flexible *Tet* construct are present in the asymmetric unit of the unit cell. (The asymmetric unit corresponds to the smallest unit that can be rotated and translated using only the crystallographic symmetry operators to generate one unit cell.) This made the refinement of *Tet* very difficult, but in the end offered four different views or conformational states of the same intron.

In addition, the highest temperature *B* factors (which indicate a higher uncertainty in atomic positions) are observed for residues adjacent to the disordered regions (such as the P5a helix in *Two*), for nucleotides involved in loose crystal packing contacts (such as the P6b helix bound to the U1A protein in *Azo*, and the P8 and P5a helices in *Two*) and for flexible domains (such as the P9 helix in *Tet*, which adopts different orientations in

molecules B and C; see Supplementary Figure S1). The three structures are well-ordered at the active site, however, where the lowest *B* factors are found (Supplementary Figure S1).

Because comparative structural analyses involve superimpositions of 3D structures, it is worth recognizing that at resolutions worse than 3.0 Å precise atomic positions, sugar puckers and contacts between nucleotides are not unambiguously determined. Such limitations do not hamper comparisons of global and local architectures (because overall helical geometries and base-pairing schemes are well established) but preclude definitive answers to some atomic level questions about the self-splicing mechanism.

Introns share a conserved catalytic core

Global fold and the G-binding site

The overall folds of the core portion of the three structures are very similar [37] (Figure 2 and Supplementary Movie; see Supplementary Figure S2 for the detailed sequence) and globally confirm previous 3D models [23,26,38] and crystal structures [32–34]. The structure of the group I intron forms by assembly of three main domains: first, P4–P5–P6 (also named P4–P6), which in *Tet* is extended and forms a hairpin clamped by tertiary interactions; second, P3–P7–P8–P9 (also named P3–P9), which is roughly perpendicular to P4–P6 and wraps around it on one side; and third, P1–P2, which lies side by side with P3–P9 (Figure 2a–c and Supplementary Movie). In addition to the universally conserved G•U pair that marks the 5′ splice site and the 3′-terminal ΩG, the sequence conservation between *Azo*, *Tet* and *Two* is localized mostly in the P4, P6 and P7 helices, matching observations from large-scale sequence alignments (Figure 2 and Supplementary Figure S2) [26]. The 3D structures confirm that these conserved nucleotides are clustered in and around the active site (Figure 2d).

The *Azo*, *Tet* and *Two* structures show the G-binding site occupied by the 3′-terminal ΩG. Remarkably, even though the three structures are proposed to represent different snapshots along the self-splicing pathway, the structure of the G-binding site is globally conserved (the root mean square deviation of the sugar-phosphate atoms is <1 Å; Figure 3a). The ΩG is involved in coplanar hydrogen-bonding interactions to the deep groove of a universally conserved G–C pair of helix P7, as predicted previously [39,40]. What had been largely unforeseen by sequence analyses and biochemical data was that this base-triple interaction is sandwiched by three other base-triple interactions comprising residues in the P7 helix and the J6/7 region, so that the guanine base is stacked on adjacent purines (only one of these additional triples had been predicted and modeled [41]). This mode of interaction provides specific recognition of the guanine base, while enabling the sugar moiety of the guanosine to remain accessible for catalysis. It can be contrasted with the situation in the recently solved structure of a riboswitch bound to the purine base hypoxanthine [42], where the base is involved in coplanar hydrogen bonds, as it is in the introns, but no base–base stacking interactions are formed and the residue is buried deeply in its binding site (Figure 3b).

Joining regions between helices

The junctions of 3–7 nt that connect the various helices of the introns are not floppy, but instead form specific structures that are essential to proper folding and to substrate recognition. The region located around the bases of helices P4 and P6 that involves the junctions J3/4 and J6/7 forms the interface between domains P4–P6 and P3–P9. It is stabilized similarly in the three structures by a succession of base-triple interactions that line up along the same side of the P4–P6 junction (Figure 4a). This triple helical scaffold was predicted early on from extensive mutagenesis studies [43–45]. A side-by-side comparison of the five layers of triples shows that most of the pairing partners had been predicted accurately, with the exception of the triple involving the first nucleotide of J3/4 (Supplementary Figure S3). But although the hydrogen-bond patterns observed in the three crystal structures are highly similar, they differ from the modeled patterns (Supplementary Figure S3). This is not unexpected, however, because the goal of 3D modeling based on comparative sequence analysis and mutational studies is to predict the molecular architecture and not necessarily the atomic details.

Because the J8/7 and J4/5 junctions are directly involved in RNA substrate recognition, it is interesting to contrast the *Azo* and *Two* structures, which contain the P1 substrate, with the *Tet* structure, which lacks this substrate (reviewed in [46]). The length and sequence of the J8/7 junction, located in the vicinity of the P3 pseudoknot, are highly conserved, corresponding to seven nucleotides in the IC1 introns (such as *Tet*) and six nucleotides in the other introns, with a particular order of conserved purines and pyrimidines. The latter six nucleotides are involved in bringing together the P3–P7 region that contains the G-binding site and the P1 helix, as shown by the *Azo* and *Two* structures (Figure 4b and Supplementary Figure S4). The *Tet* structure, which lacks the P1 helix, has a considerably different J8/7 structure (Figure 4b). Furthermore, J8/7 seems to be flexible in the absence of P1, as indicated by the different structures that it adopts in the four *Tet* molecules present in the asymmetric unit (Supplementary Figure S4b). The *Tet* structure also shows that the extra uridine at the 5′ end of J8/7 (U300 is not present in the other introns) interacts in the deep groove of the P3 helix (Figure 4b). Also notable is that the additional G–C pair on the 5′ side of the P3 helix of *Tet* superimposes over the G–C pair on the 'top' of helix P8 of *Azo* and *Two* (see orientation in Supplementary Figure S2).

In the structures of *Azo* and *Two*, the conserved adenines in the J4/5 joining region form a tandem of sheared A◦A pairs, in which one adenine of each strand comes into direct contact with the G•U pair on the 5′ splice site (Supplementary Figures S4a and S5). The accessible N2 of the guanine is contacted by both the N3 atom and the 2′-OH group of the 3′ adenine on the 3′ side of J4/5 (A87 in *Azo*, A83 in *Two*). The N3 atom and the 2′-OH group of the

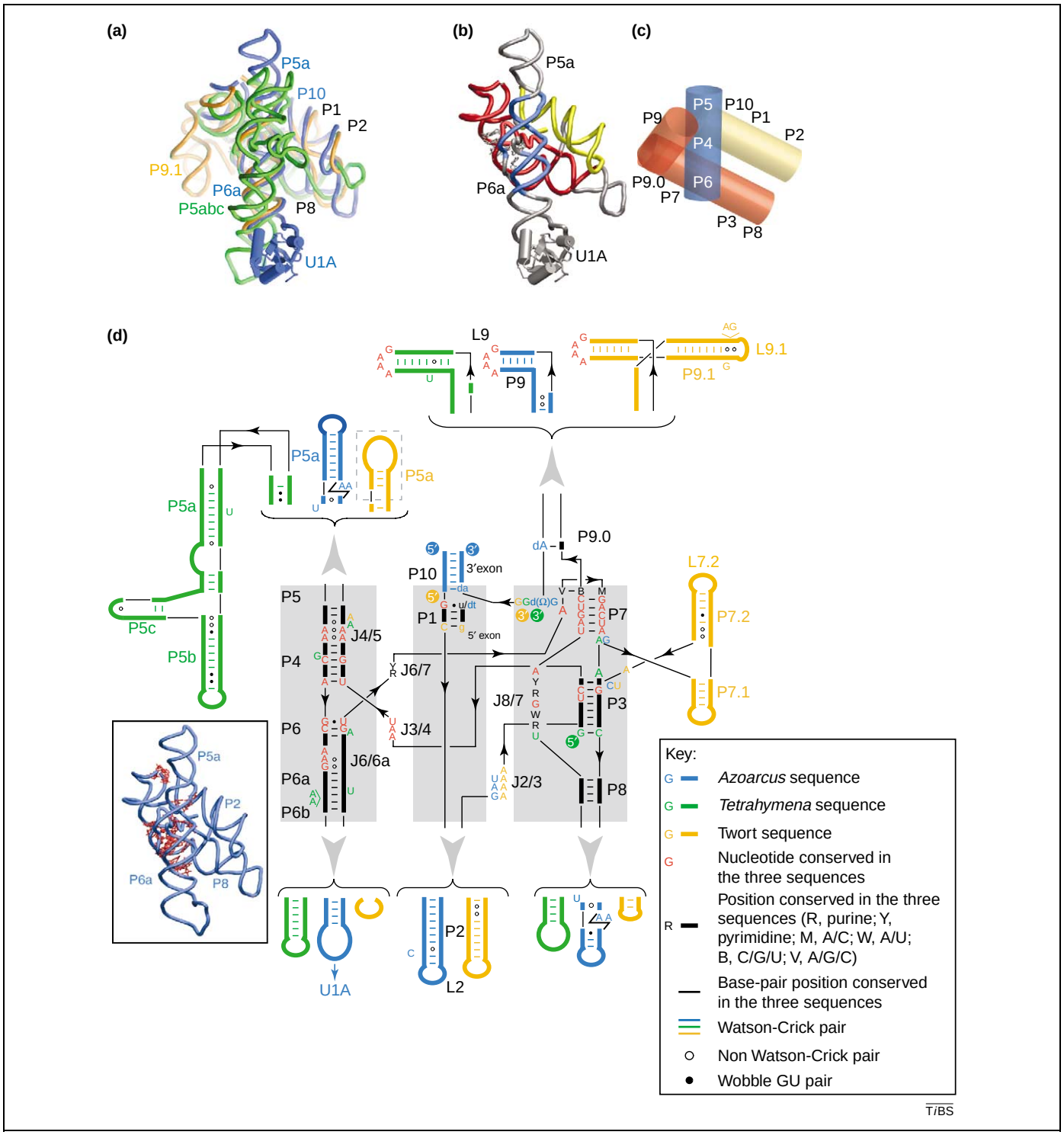

Figure 2. Introns from *Azoarcus, Tetrahymena* and Twort share a common core. **(a)** Superimposition of the 3D structures (based on the P7 helix) of *Azo* (blue), *Tet* (green) and *Two* (orange). Of the four structures in the asymmetric unit, the structure of molecule B was used for *Tet* in this and subsequent figures unless otherwise stated. The U1A protein is shown here but is removed in subsequent figures. **(b)** The P4–P6 (blue), P3–P9 (red) and P1–P2 (yellow) domains highlighted on the *Azo* structure (the P7 region is shown in ball-and-stick notation). **(c)** The relative positioning of the domains shown in (a,b). **(d)** Overlay of secondary structure diagrams emphasizing conserved structural elements (black) inside a conserved core (formed by the three shaded boxes that define, from left to right, the P4–P6, P1–P2 and P3–P9 domains). P1–P10 elements, important junctions and loops, and the intron-specific peripheral domains are indicated. Colored 5′ and 3′ symbols locate the different ends of the three constructs. The broken box around the P5a element of *Two* specifies that this region is disordered in the final model. Lower- and upper-case characters stand for exon and intron sequences, respectively. Nucleotides conserved in the three sequences are shown in red: these residues on the 3D structure of *Azo* are shown in the inset. Three-dimensional structures in this and all figures except Figure 4 were drawn with VMD v.1.8.2 [82] and rendered with Povray v.3.6 (http://www.povray.org/).

3′ adenine on the 5′ side of J4/5 (A58 in *Azo*, A56 in *Two*) form hydrogen bonds to the 2′-OH group of the guanine (Supplementary Figure S4a). These observations rationalize previous experimental and modeling results obtained with the full-length *Tet* RNA [47,48]. Remarkably, the geometry of the A◦A tandem is maintained in molecules A, C and D of *Tet*, even though the relevant substrate is missing (Supplementary Figure S4b).

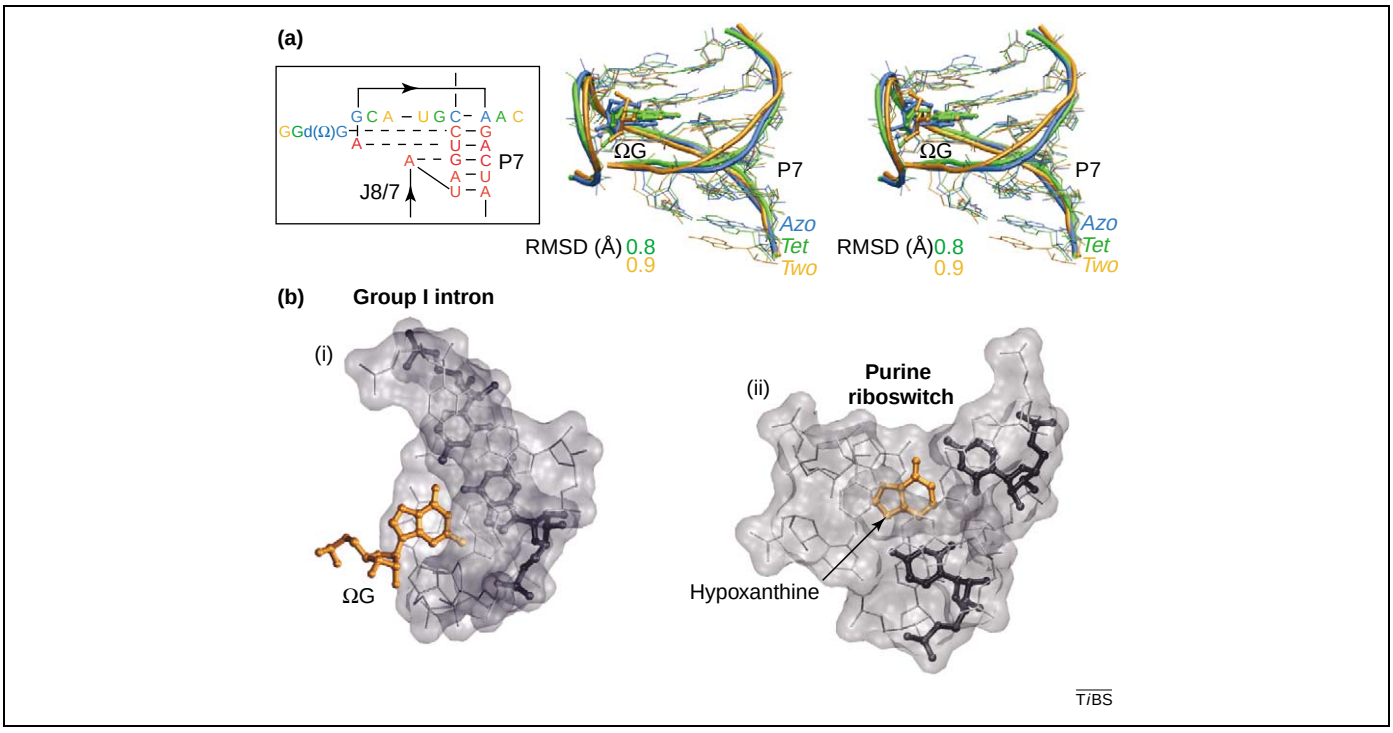

Figure 3. The conserved G-binding site. **(a)** Stereo view showing superimposition of the G-binding site in the *Azo* (blue), *Tet* (green) and *Two* (orange) structures. Inset shows the base-pairing scheme. The ΩG is in bold. Values in green and orange indicate the root mean square deviations (RMSDs) for the *Tet*/*Azo* and *Two*/*Azo* comparisons, respectively. Unless otherwise indicated, the RMSDs were calculated in this and subsequent figures for the sugar-phosphate atoms of the depicted nucleotides, with the software Lsqman [83]. **(b)** Two modes of interaction involving a purine base: (i) the G-binding site in group I introns, and (ii) the binding site for purines in the guanine riboswitch. The ligand is in orange, partners involved in direct hydrogen bonds are in black, and the binding site is shown as a gray surface.

Tetraloop docking interaction

In most introns the P9 stem is capped by a stable GNRA tetraloop (L9), where N is any nucleotide and R is purine [26]. Sequence alignments and experimental data have identified two modes of specific interactions involving the L9 tetraloop and a region in P5: in some introns, L9 binds to consecutive Watson–Crick pairs [26,49], whereas in others it contacts a conserved 11-nt receptor motif [50].

The *Azo*, *Tet* and *Two* constructs used for crystallization possess a GAAA L9 loop (although the wild-type *Tet* sequence does not). The L9 loop of the *Tet* RNA interacts with two adjacent base pairs in P5 (Figure 4c), similar to the interaction predicted for other group I introns [51]. This interaction is analogous to that observed in a structure of the hammerhead ribozyme [52]. The part of P5 observed in the *Two* structure and the loop L9 superimpose on these regions of the *Tet* RNA, a good indication that they are involved in similar interactions. The *Azo* structure shows L9 in contact with the conserved 11-nt receptor motif located in P5 (Figure 4c). This interaction is similar to one observed on another side of the *Azo* intron between loop L2 and stem P8 [28] (Supplementary Figure S5) and to one described previously in the *Tet* P4–P6 domain [32]. As emphasized by this structural comparison, the two types of interaction (involving either adjacent Watson–Crick pairs or the 11-nt receptor) have a similar role in clamping the P5 and P9 regions to one another (Supplementary Figure S5).

Pseudoknot belts

The *Azo* structure showed a novel feature – a 'pseudoknot belt', which involves a single stretch of ~25 nucleotides that spans three helices (including the pseudoknot P3) and two junctions from the 3′ end of P2 to the 5′ end of P4 [28]. This stretch encircles the intron at its midpoint (Figure 5). This structure is compatible with previous kinetic experiments that indicated that P3 is the last helix to form during folding of the *Tet* intron [53–55].

Highlighting the corresponding nucleotides in the *Tet* and *Two* structures reveals that these introns also have a pseudoknot belt, which wraps around the circumference in a similar manner (Figure 5). Notably, the belt remains on the outside of the molecule, despite differences in the set of peripheral elements possessed by each intron: the loop L2 binds to P8 in *Azo* and *Two*, whereas it forms a pseudoknot with a region in the P5abc extension in *Tet*, as shown in a 3D model of the full-length *Tet* sequence by Lehnert *et al.* [23] (Figure 5).

The architecture of the core is supported by diverse peripheral elements

The classification of group I introns into 13 structural subgroups is based not only on different sets of conserved sequences in the core, but also on subgroup-specific peripheral elements [23,26]. These elements typically branch out from the core in the P2, P5, P6, P8 and P9 regions. Some of them are longer than 500 nt and contain open reading frames encoding endonucleases that promote intron mobility [11]. The main role of the peripheral

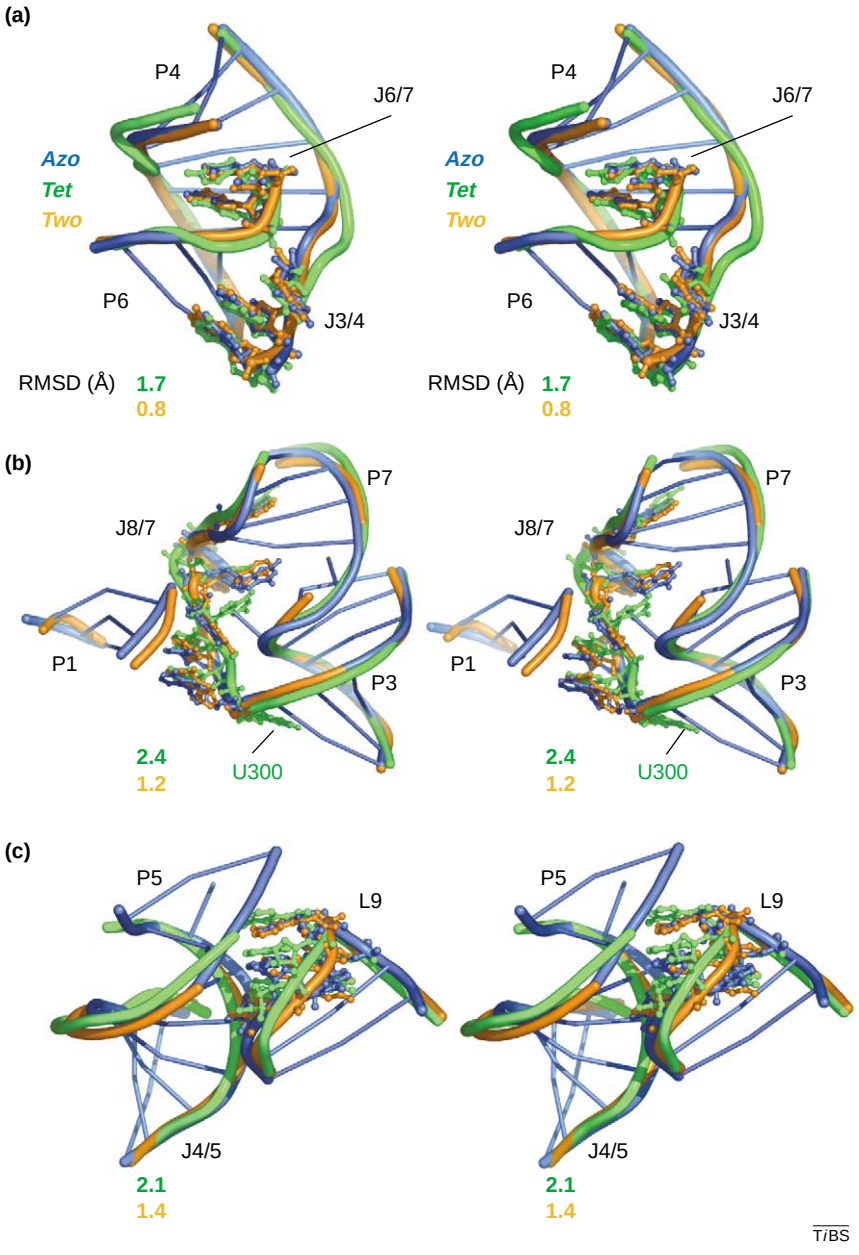

Figure 4. Stereo views showing superimposition of various conserved structural domains. **(a)** Interactions at the junction of the P4 and P6 helices. **(b)** The J8/7 junction brings together helices P1, P3 and P7 (helices P1 and P7 were not used in the RMSD calculations). **(c)** Tetraloop docking interactions that clamp domains P5 and P9 to one another (RMSD calculated for residues in L9 and J4/5). Nucleotides belonging to junctions J3/4 and J6/7 (a), to junction J8/7 (b) and to loop L9 (c) are shown in ball-and-stick notation. Base pairs of the *Azo* structure only are indicated by rods. Figure was drawn with PyMOL v.0.95 (http://pymol.sourceforge.net/) and nuccyl v.1.5 (http://www.biosci.ki.se/groups/ljo/software/nuccyl.html).

elements is to stabilize the structure of the core by establishing long-range tertiary interactions [56,57]; for example, when progressively deprived of its peripheral elements, the intron loses activity [58], although very low levels persist in the complete absence of some peripheral elements [24]. Artificial peripheral elements can even be engineered around a minimal catalytic module to enhance activity [59]. Because the *Azo*, *Tet* and *Two* introns belong to three different subgroups, a comparative analysis of their structures offers an exciting opportunity to understand how different sets of elements support a common architecture of the core.

In the three structures, the peripheral elements poke out in different directions to form intron-specific sets of long-range interactions. Remarkably, peripheral elements form buttresses that can wrap around the conserved active site either halfway, as in the *Two* intron, or totally, as in the *Tet* intron (Figure 6). For example, the P5abc extension, which is typical of IC1 introns such as *Tet*, protrudes on the side opposite to that of the P7.1–P7.2 extension, which is characteristic of IA introns such as *Two* (Figure 2a and Supplementary Movie). Consequently, the P5abc extension establishes several long-range interactions within the P4–P6 domain and with loop L9, and it

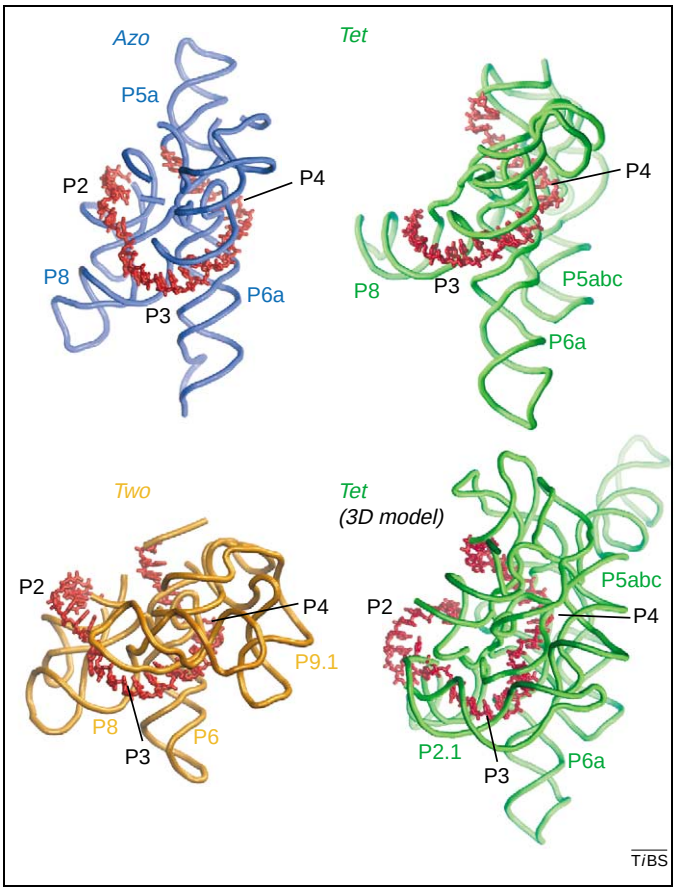

Figure 5. Pseudoknot belts on the circumference of the introns. Adjacent nucleotides involved in formation of the belt are shown in red in each of the crystal structures [residues 34–59 in *Azo* (blue), residues 94–116 in *Tet* (green) and residues 31–58 in *Two* (orange)] and in a 3D model (green) of the full-length *Tet* sequence (residues 47–56 and 94–116) [23]. Introns are shown in a similar orientation.

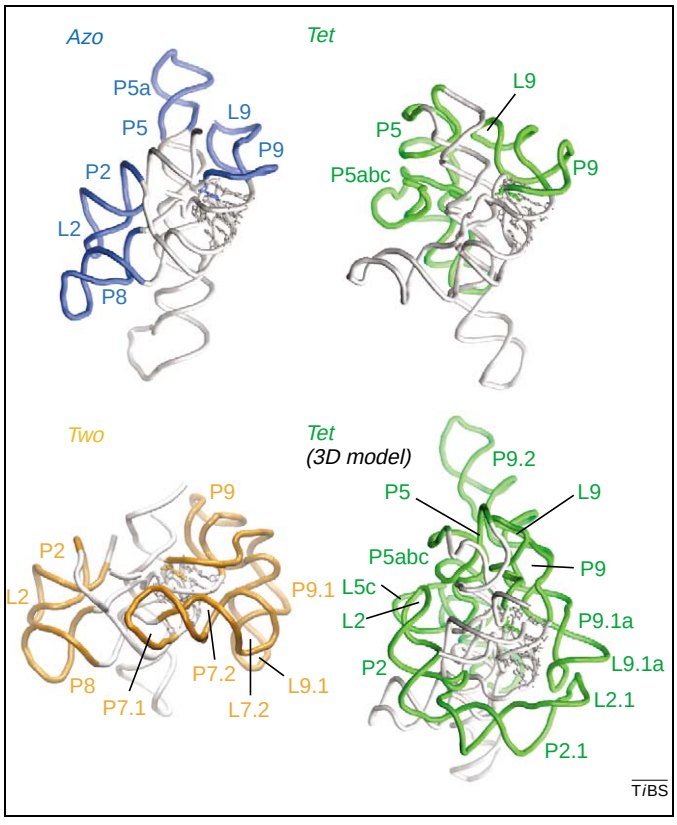

Figure 6. Buttresses around the active site. Intron-specific peripheral elements are highlighted in color in each of the structures of *Azo* (blue), *Tet* (green) and *Two* (orange) as well as in the full-length model of *Tet* (green) [23]. The rest of the intron is represented in gray. The catalytic site is shown in gray ball-and-stick notation, and the ΩG is shown in blue, green or orange.

facilitates pseudoknot formation between L2 and L5c, and between L2.1 and L9.1a (Figure 6 and Supplementary Figure S5) [23]. The P7.1–P7.2 extension of *Two* forms a bow-tie-like protuberance that enables loop L7.2 to interact with L9.1, which is another extension that is typical of IA introns (Figure 6 and Supplementary Figure S5) [26].

The *Two* structure reveals tetraloop–base-pair interactions between L2 and L9 and the helices P8 and P5 (the residues in most of the P5 extension are, however, absent from the final model; Supplementary Figure S5). Notably, these interactions are replaced by tetraloop–receptor interactions in the *Azo* structure (Figure 6 and Supplementary Figure S5). Together with the high G+C content of the intron (Table 1), these stronger interactions, established on the outermost ends of the structure, probably account for the activity of this intron at high temperatures [60]. Of note, no other long-range tertiary interactions involving peripheral domains are observed in *Azo*.

Some group I introns require a protein for activity; for example, the *Neurospora crassa* ND1 intron is stabilized by CYT-18 [61,62], and in *Saccharomyces cerevisiae* the bI3 intron is stabilized by the bI3 and Mrs1 proteins [63], and the bI5 intron by CBP2 [64]. In these introns, the protein does not directly participate in the catalysis, but seems to be involved instead in indirectly stabilizing the catalytic core either by reinforcing long-range interactions between peripheral elements (e.g. CYT-18 stabilizes the L9–P5 interaction) or by binding to a peripheral element (e.g. the bI3 protein binds to the P5abc extension). In addition, extensive biochemical data obtained for group IE introns have recently shown the crucial role played by a part of stem P2.1 (a marker for group IE introns) in forming a triple helical interaction with P3 and P6 that is required for catalysis [65].

How many metal ions in the active site?

The folding of group I introns brings many phosphate groups into close proximity within the active site. The resulting high electronegative potential is compensated by the specific binding of several metal ions [66]. Among these metal ions, some have been found to be directly involved in catalysis [67,68].

Computational investigations [69], comparisons with protein enzyme active sites [70], and biochemical and modeling experiments [71] have led several groups to suggest mechanisms based on two metal ions for the self-splicing reaction [71–73]. Indeed, *in vitro* experiments uncovered a first (M_1) [68] and then a second (M_2) [74] metal ion with specific roles in catalysis: (i) in the first step one metal ion serves as a nucleophile activator (M_2 bound to the 3′-OH of the guanosine cofactor) and the other serves as a leaving group stabilizer (M_1 bound to

the 3′ oxygen atom of the uridine of the substrate G•U pair); (ii) in the second step, M_2 (now bound to the 3′ oxygen atom of ΩG) and M_1 (still bound to the same uridine) exchange their roles [72,73]. In addition, both M_1 and M_2 coordinate the scissile phosphate at the 5′ splice site during the first step, and then the scissile phosphate at the 3′ splice site during the second step. The 2′-OH group of the guanosine cofactor has also been shown to be contacted by a metal ion during the first step of splicing [72]. A third metal ion was subsequently suggested to have this role [75,76] and to coordinate residue 262 (*Tet* numbering), which lies in the catalytic site [77].

The resolution at which the structures of *Azo*, *Tet* and *Two* were solved has enabled metal ions to be placed in the active site (Figure 7): two metals in the *Azo* active site (one Mg^{2+} and one K^+ ion in the first structure containing the deoxy ΩG [29], two Mg^{2+} ions in the most recent ribo ΩG structure [35]), one metal in the *Tet* active site (a Eu^{3+} or Ir^{3+} ion soaked in to identify sites normally occupied by Mg^{2+}) and one metal in the *Two* active site (Mg^{2+}). The binding mode of the two metal ions identified in the *Azo* active site of the most recent structure supports a two-metal ion model; first, two metal ions could be sufficient for catalysis to occur in the second step of splicing: M_1 as a nucleophile activator and M_2 as a leaving group stabilizer (Figure 7); second, the hydroxyl and phosphate oxygen groups that are directly coordinated by the two metal ions (Figure 7) are the biochemically predicted ligands; and last, the metal–metal distance is 3.9 Å, in agreement with the model of a two-metal-ion catalytic center [73].

In all four *Tet* structures, the metal ion observed contacts the 3′ oxygen atom of the ΩG and the O2P from residue 306 (*Tet* numbering), similar to metal M_2 in the *Azo* structure (Figure 7 and Supplementary Figure S6). In the *Two* structure, a metal ion contacts the 3′-OH group of the ΩG and the phosphate oxygen atom O2P of residue 306, suggesting that this metal ion could be M_2. The distances are longer, however, and potential

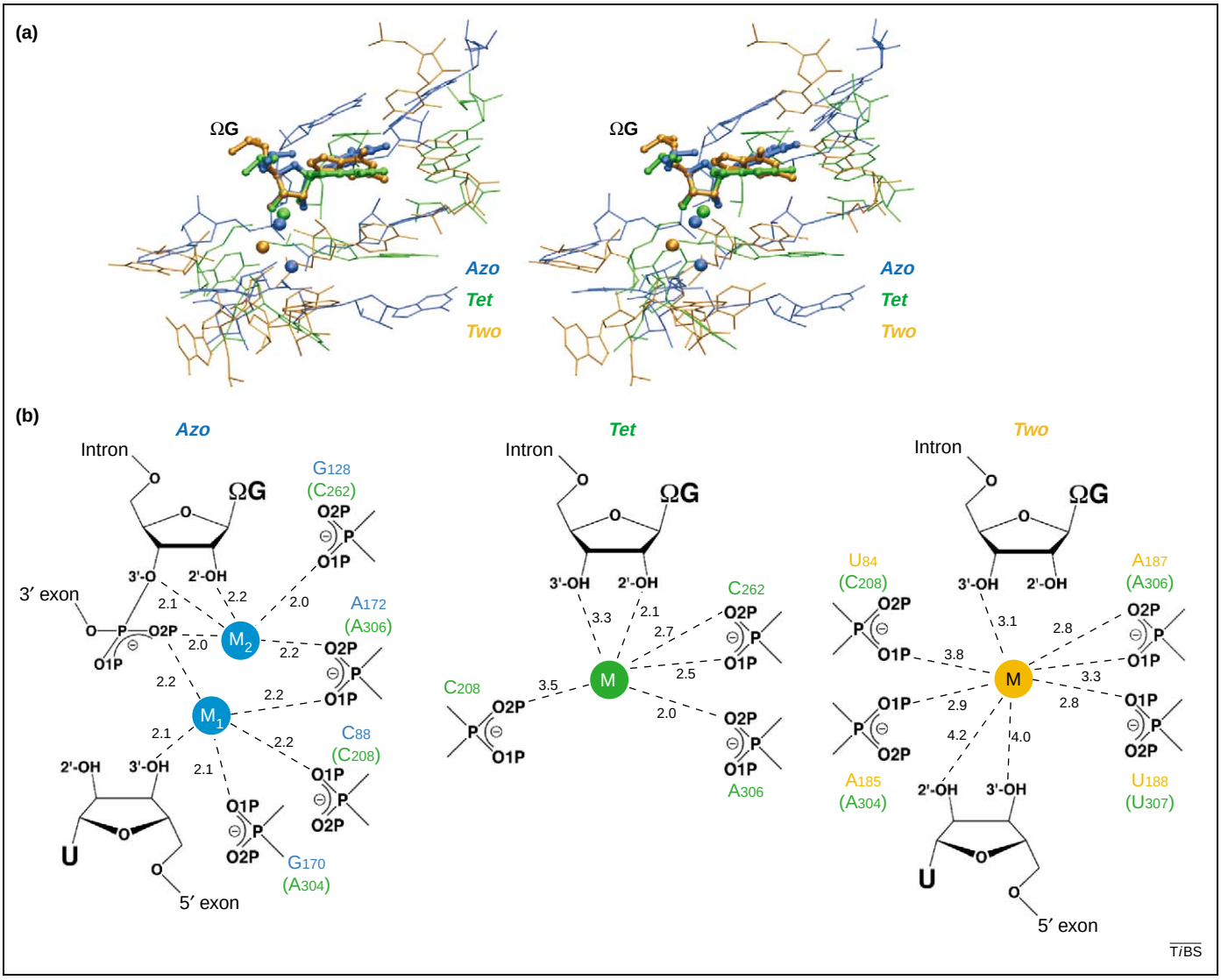

Figure 7. Metal ions at the active site. **(a)** Superimposition of the sugar ring atoms and the 2′-OH and 3′-OH groups of the ΩG from *Azo* (structure from Ref. [35] in blue), *Tet* (molecule C in green) and *Two* (orange). Residues bound to the ΩG and to the metal ions are shown. **(b)** Representation of the active site in the three introns. Potential metal–ligand bonds are represented by broken lines and distances (Å) are indicated. For each intron, the *Tet* numbering is indicated in green in parentheses as a reference.

interactions (although they would involve outer-sphere coordination of the metal ion) could be formed to residue 304 and the uridine from the substrate G•U pair, which are ligands of metal M_1 in the *Azo* structure (Figure 7).

The *Azo* structure shows how the scissile phosphate at the 3′ splice site bridges the two metal ions. In absence of this phosphate group, as observed in *Tet* and *Two*, the catalytic site might contain only one well-ordered metal ion. Subsequent rearrangement could then occur to facilitate positioning of a second metal ion for catalysis.

Notably, the present structures do not rule out the presence of a third metal ion at the active site. A third metal ion could be (i) disordered enough in the crystal structures that it is not observed at these resolutions, (ii) displaced by the high concentrations of monovalent salt used during crystallization in the case of *Two* [31], (iii) involved in steps or transitory stages of the self-splicing reaction other than those captured by crystallography, (iv) necessary in other biochemical contexts (e.g. under different splicing conditions or when analyzing an amino- or sulfur-substituted intron such as the one used in recent biochemical studies [75]), or (v) even possibly involved in the catalysis of some group I introns (such as *Tet*) but not others (such as *Azo*).

Concluding remarks

The crystal structures of self-splicing introns from *Azo*, *Tet* and *Two* represent a major breakthrough in the field of group I intron and ribozyme research. Twenty-two years after the discovery of the first ribozyme, they have revealed at the atomic level a conserved catalytic core built around a G-binding site and a precise set of structural domains that help to stabilize the binding of substrates. Metal ions are identified in the catalytic site at positions predicted by models based on solution data and by comparisons with metalloprotein enzymes.

Group I introns are subdivided into 13 subgroups that are distinguished by particular conserved sequences at the core and diverse sets of peripheral elements. Structural comparisons show how the conserved architecture of the catalytic core is supported by long-range tertiary interactions formed by these intron-specific peripheral elements. Investigation of additional introns is needed to address whether the sequence identities observed in the same subgroup are functionally linked to the presence of a specific set of peripheral elements, or whether they are simply a fossil of the evolution of these introns. Such studies would emphasize the importance of the peripheral elements in stabilizing a catalytic core – a function that is also known to be crucial in other ribozymes such as the hammerhead ribozyme [78] and the catalytic RNA of ribonuclease P [79–81].

Acknowledgements

We gratefully acknowledge Scott Strobel for sharing data before publication, as well as Alan Lambowitz, Paul Paukstelis and Kevin Weeks for providing coordinates of 3D models beneficial to the preparation of this manuscript. We thank Robert Batey, Eric Westhof, David Zappulla and the referees of this review for helpful suggestions on the manuscript.

Supplementary data

Supplementary data associated with this article can be found at doi:10.1016/j.tibs.2005.11.008

References

1 Guerrier-Takada, C. *et al*. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* 35, 849–857
2 Kruger, K. *et al*. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* 31, 147–157
3 Peebles, C.L. *et al*. (1986) A self-splicing RNA excises an intron lariat. *Cell* 44, 213–223
4 van der Veen, R. *et al*. (1986) Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing *in vitro*. *Cell* 44, 225–234
5 Doudna, J.A. and Cech, T.R. (2002) The chemical repertoire of natural ribozymes. *Nature* 418, 222–228
6 Fedor, M.J. and Williamson, J.R. (2005) The catalytic diversity of RNAs. *Nat. Rev. Mol. Cell Biol.* 6, 399–412
7 Winkler, W.C. *et al*. (2004) Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* 428, 281–286
8 Bartel, D.P. and Szostak, J.W. (1993) Isolation of new ribozymes from a large pool of random sequences. *Science* 261, 1411–1418
9 Tarasow, T.M. *et al*. (1997) RNA-catalysed carbon–carbon bond formation. *Nature* 389, 54–57
10 Steitz, T.A. and Moore, P.B. (2003) RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends Biochem. Sci.* 28, 411–418
11 Haugen, P. *et al*. (2005) The natural history of group I introns. *Trends Genet.* 21, 111–119
12 Bashan, A. *et al*. (2003) Structural basis of the ribosomal machinery for peptide bond formation, translocation, and nascent chain progression. *Mol. Cell* 11, 91–102
13 Kirsebom, L.A. (2002) RNase P RNA-mediated catalysis. *Biochem. Soc. Trans.* 30, 1153–1158
14 Joyce, G.F. (2002) The antiquity of RNA-based evolution. *Nature* 418, 214–221
15 Gilbert, W. (1986) Origin of life: the RNA world. *Nature* 319, 618–618
16 Akashi, H. *et al*. (2005) Gene discovery by ribozyme and siRNA libraries. *Nat. Rev. Mol. Cell Biol.* 6, 413–422
17 Alexander, R.C. *et al*. (2005) 5′ Transcript replacement *in vitro* catalyzed by a group I intron-derived ribozyme. *Biochemistry* 44, 7796–7804
18 Sullenger, B.A. and Cech, T.R. (1994) Ribozyme-mediated repair of defective mRNA by targeted, *trans*-splicing. *Nature* 371, 619–622
19 Cannone, J.J. *et al*. (2002) The Comparative RNA Web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* 3, 2
20 Davies, R.W. *et al*. (1982) Making ends meet: a model for RNA splicing in fungal mitochondria. *Nature* 300, 719–724
21 Michel, F. *et al*. (1982) Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. *Biochimie* 64, 867–881
22 Waring, R.B. and Davies, R.W. (1984) Assessment of a model for intron RNA secondary structure relevant to RNA self-splicing – a review. *Gene* 28, 277–291
23 Lehnert, V. *et al*. (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
24 Ikawa, Y. *et al*. (2000) Minimal catalytic domain of a group I self-splicing intron RNA. *Nat. Struct. Biol.* 7, 1032–1035
25 Strobel, S.A. and Ortoleva-Donnelly, L. (1999) A hydrogen-bonding triad stabilizes the chemical transition state of a group I ribozyme. *Chem. Biol.* 6, 153–165
26 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
27 Suh, S.O. *et al*. (1999) A group I intron in the nuclear small subunit rRNA gene of *Cryptendoxyla hypophloia*, an ascomycetous fungus: evidence for a new major class of group I introns. *J. Mol. Evol.* 48, 493–500

28 Adams, P.L. *et al*. (2004) Crystal structure of a group I intron splicing intermediate. *RNA* 10, 1867–1887
29 Adams, P.L. *et al*. (2004) Crystal structure of a self-splicing group I intron with both exons. *Nature* 430, 45–50
30 Guo, F. *et al*. (2004) Structure of the *Tetrahymena* ribozyme: base triple sandwich and metal ion at the active site. *Mol. Cell* 16, 351–362
31 Golden, B.L. *et al*. (2005) Crystal structure of a phage Twort group I ribozyme–product complex. *Nat. Struct. Mol. Biol.* 12, 82–89
32 Cate, J.H. *et al*. (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* 273, 1678–1684
33 Juneau, K. *et al*. (2001) Structural basis of the enhanced stability of a mutant ribozyme domain and a detailed view of RNA-solvent interactions. *Structure* 9, 221–231
34 Golden, B.L. *et al*. (1998) A preorganized active site in the crystal structure of the *Tetrahymena* ribozyme. *Science* 282, 259–264
35 Stahley, M.R. and Strobel, S.A. (2005) Structural evidence for a two-metal-ion mechanism of group I intron splicing. *Science* 309, 1587–1590
36 Golden, B.L. *et al*. (1997) Crystals by design: a strategy for crystallization of a ribozyme derived from the *Tetrahymena* group I intron. *J. Mol. Biol.* 270, 711–723
37 Dror, O. *et al*. (2005) ARTS: alignment of RNA tertiary structures. *Bioinformatics* 21(Suppl 2), ii47–ii53
38 Rangan, P. *et al*. (2004) Architecture and folding mechanism of the *Azoarcus* group I Pre-tRNA. *J. Mol. Biol.* 339, 41–51
39 Been, M.D. and Perrotta, A.T. (1991) Group I intron self-splicing with adenosine: evidence for a single nucleoside-binding site. *Science* 252, 434–437
40 Michel, F. *et al*. (1989) The guanosine binding site of the *Tetrahymena* ribozyme. *Nature* 342, 391–395
41 Yarus, M. *et al*. (1991) An axial binding site in the *Tetrahymena* precursor RNA. *J. Mol. Biol.* 222, 995–1012
42 Batey, R.T. *et al*. (2004) Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature* 432, 411–415
43 Flor, P.J. *et al*. (1989) A conserved base pair within helix P4 of the *Tetrahymena* ribozyme helps to form the tertiary structure required for self-splicing. *EMBO J.* 8, 3391–3399
44 Green, R. *et al*. (1990) *In vitro* genetic analysis of the *Tetrahymena* self-splicing intron. *Nature* 347, 406–408
45 Michel, F. *et al*. (1990) Phylogenetic and genetic evidence for base-triples in the catalytic domain of group I introns. *Nature* 347, 578–580
46 Woodson, S.A. (2005) Structure and assembly of group I introns. *Curr. Opin. Struct. Biol.* 15, 324–330
47 Strobel, S.A. and Cech, T.R. (1995) Minor groove recognition of the conserved G.U pair at the *Tetrahymena* ribozyme reaction site. *Science* 267, 675–679
48 Strobel, S.A. *et al*. (1998) Complementary sets of non-canonical base pairs mediate RNA helix packing in the group I intron active site. *Nat. Struct. Biol.* 5, 60–65
49 Jaeger, L. *et al*. (1994) Involvement of a GNRA tetraloop in long-range RNA tertiary interactions. *J. Mol. Biol.* 236, 1271–1276
50 Costa, M. and Michel, F. (1995) Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.* 14, 1276–1285
51 Waldsich, C. *et al*. (2002) Monitoring intermediate folding states of the *td* group I intron *in vivo*. *EMBO J.* 21, 5281–5291
52 Pley, H.W. *et al*. (1994) Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature* 372, 111–113
53 Rangan, P. *et al*. (2003) Assembly of core helices and rapid tertiary folding of a small bacterial group I ribozyme. *Proc. Natl. Acad. Sci. U. S. A.* 100, 1574–1579
54 Pan, J. and Woodson, S.A. (1998) Folding intermediates of a self-splicing RNA: mispairing of the catalytic core. *J. Mol. Biol.* 280, 597–609
55 Zarrinkar, P.P. and Williamson, J.R. (1994) Kinetic intermediates in RNA folding. *Science* 265, 918–924
56 Jaeger, L. *et al*. (1996) The structure of group I ribozymes. In *Nucleic Acids and Molecular Biology* (Eckstein, F. and Lilley, D.M.J., eds), pp. 33–51, Springer-Verlag
57 Naito, Y. *et al*. (1998) P5abc of the *Tetrahymena* ribozyme consists of three functionally independent elements. *RNA* 4, 837–846
58 Cech, T.R. (1990) Self-splicing of group I introns. *Annu. Rev. Biochem.* 59, 543–568
59 Ohuchi, S.J. *et al*. (2002) Modular engineering of a group I intron ribozyme. *Nucleic Acids Res.* 30, 3473–3480
60 Tanner, M. and Cech, T. (1996) Activity and thermostability of the small self-splicing group I intron in the pre-tRNA^{Ile} of the purple bacterium *Azoarcus*. *RNA* 2, 74–83
61 Caprara, M.G. *et al*. (1996) A tyrosyl-tRNA synthetase recognizes a conserved tRNA-like structural motif in the group I intron catalytic core. *Cell* 87, 1135–1145
62 Paukstelis, P.J. *et al*. (2005) A tyrosyl-tRNA synthetase adapted to function in group I intron splicing by acquiring a new RNA binding surface. *Mol. Cell* 17, 417–428
63 Longo, A. *et al*. (2005) Evolution from DNA to RNA recognition by the bI3 LAGLIDADG maturase. *Nat. Struct. Mol. Biol.* 12, 779–787
64 Webb, A.E. *et al*. (2001) Protein-dependent transition states for ribonucleoprotein assembly. *J. Mol. Biol.* 309, 1087–1100
65 Xiao, M. *et al*. (2005) A peripheral element assembles the compact core structure essential for group I intron self-splicing. *Nucleic Acids Res.* 33, 4602–4611
66 Woodson, S.A. (2005) Metal ions and RNA folding: a highly charged topic with a dynamic future. *Curr. Opin. Chem. Biol.* 9, 104–109
67 Cech, T.R. *et al*. (1992) RNA catalysis by a group I ribozyme. Developing a model for transition state stabilization. *J. Biol. Chem.* 267, 17479–17482
68 Piccirilli, J.A. *et al*. (1993) Metal ion catalysis in the *Tetrahymena* ribozyme reaction. *Nature* 361, 85–88
69 Uchimaru, T. *et al*. (1993) Theoretical analyses on the role of Mg^{2+} ions in ribozyme reactions. *FASEB J.* 7, 137–142
70 Beese, L.S. and Steitz, T.A. (1991) Structural basis for the 3′–5′ exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J.* 10, 25–33
71 Streicher, B. *et al*. (1996) The environment of two metal ions surrounding the splice site of a group I intron. *EMBO J.* 15, 2556–2564
72 Sjogren, A.S. *et al*. (1997) Metal ion interaction with cosubstrate in self-splicing of group I introns. *Nucleic Acids Res.* 25, 648–653
73 Steitz, T.A. and Steitz, J.A. (1993) A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. U. S. A.* 90, 6498–6502
74 Weinstein, L.B. *et al*. (1997) A second catalytic metal ion in group I ribozyme. *Nature* 388, 805–808
75 Shan, S. *et al*. (1999) Three metal ions at the active site of the *Tetrahymena* group I ribozyme. *Proc. Natl. Acad. Sci. U. S. A.* 96, 12299–12304
76 Shan, S.O. and Herschlag, D. (1999) Probing the role of metal ions in RNA catalysis: kinetic and thermodynamic characterization of a metal ion interaction with the 2′-moiety of the guanosine nucleophile in the *Tetrahymena* group I ribozyme. *Biochemistry* 38, 10958–10975
77 Hougland, J.L. *et al*. (2005) Functional identification of catalytic metal ion binding sites within RNA. *PLoS Biol.* 3, e277
78 Khvorova, A. *et al*. (2003) Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. *Nat. Struct. Biol.* 10, 708–712
79 Krasilnikov, A.S. *et al*. (2004) Basis for structural diversity in homologous RNAs. *Science* 306, 104–107
80 Baird, N.J. *et al*. (2005) Structure of a folding intermediate reveals the interplay between core and peripheral elements in RNA folding. *J. Mol. Biol.* 352, 712–722
81 Kazantsev, A.V. *et al*. (2005) Crystal structure of a bacterial ribonuclease P RNA. *Proc. Natl. Acad. Sci. U. S. A.* 102, 13392–13397
82 Humphrey, W., *et al*. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33–38, 27–38
83 Sierk, M.L. and Kleywegt, G.J. (2004) *Deja vu* all over again: finding and analyzing protein structure similarities. *Structure* 12, 2103–2111