Published in final edited form as:
Curr Opin Struct Biol. 2011 June ; 21(3): 319–326. doi:10.1016/j.sbi.2011.03.003.

Emerging structural themes in large RNA molecules

Nicholas J. Reiter, **Clarence W. Chan**, and **Alfonso Mondragón**‡
Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208 USA

Summary

Extensive networks of tertiary interactions give rise to unique, highly-organized domain architectures that characterize the three-dimensional structure of large RNA molecules. Formed by stacked layers ofa near-planar arrangement of contiguous coaxial helices, large RNA molecules are relatively flat in overall shape. The functional core of these molecules is stabilized by a diverse set of tertiary interaction motifs that often bring together distant regions of conserved nucleotides. Although homologous RNAs from different organisms can be structurally diverse, they adopt a structurally conserved functional core that includes preassembled active and/or substrate binding sites. These findings broaden our understanding of RNA folding and tertiary structure stabilization, illustrating how large, complex RNAs assemble into unique structures to perform recognition and catalysis.

Introduction

RNAs assume diverse roles in numerous cellular processes, including serving as molecules to store or transfer genetic information, acting as catalysts (ribozymes) to accelerate specific chemical reactions, and functioning as gene-regulatory elements (riboswitches) to direct gene expression. Comparable to protein molecules, many RNAs adopt complex three-dimensional folds in order to function. Recent structural determinations of several large RNA molecules in complex with their natural substrates or ligand cofactors have provided important insights into the mechanisms of RNA folding, RNA-based molecular recognition, and RNA-based enzymatic catalysis.

Early insights into the principles underlying RNA tertiary architecture were derived from the three-dimensional structures of $tRNA^{Phe}$ [1,2]; small nucleolytic ribozymes, such as the hammerhead [3,4], hepatitis delta virus [5], and hairpin ribozymes [6]; and domains of significantly more complex molecules, such as the group I P4-P6 domain [7] and the RNase P specificity domain [8,9]. Collectively, these structures revealed an architecture dominated by groups of single and coaxial helical stems that are interconnected by a now familiar set of secondary and tertiary structural motifs serving to stabilize their overall fold, such as coaxial helical stacking, tetraloop-tetraloop receptor interactions, pseudoknots, ribose zippers, and dinucleotide platforms (reviewed in [10]). However, subsequent studies have shown that these motifs also act as essential components in the formation of larger, multi-domain RNA-based molecules, which prompts the questions: Are there architectural themes that are unique to larger RNAs? Is there a recurrence of structural features that enable these RNAs to recognize their natural targets? Specifically, do large catalytic and regulatory RNAs, like most protein enzymes, contain preassembled active sites and utilize shape complementarity to perform molecular recognition?

‡Corresponding author Tel: (847)491-7726 FAX: (847)467-6489 a-mondragon@northwestern.edu.

In this review, structural studies of large RNA molecules (> 100 nucleotides) in complex with their natural targets or ligand cofactors will be examined with a focus on ribonuclease (RNase) P [11] and self-splicing (group I and group II) introns [12,13]. The architectural themes emerging from these molecules are further illustrated by features present in various riboswitch structures [14-17] (reviewed in [18-21]). We confine our analysis to large RNA molecules that are stable without or with minimal assistance from protein molecules, and hence very large ribonucleoprotein (RNP) complexes, such as the spliceosome, telomerase, and the ribosome, are beyond the scope of this review.

Parallelisms in the tertiary folds

Consistent with observations that small nucleolytic RNAs and domains of large RNAs assume a relatively planar tertiary conformation (reviewed in [10,22], the recent crystal structures of a bacterial RNase P holoenzyme and various full-length group I and II introns in complex with their respective molecular targets demonstrate that even large RNAs possess a relatively flat overall fold (Figure 1). The dominant tertiary motif that helps to preserve their relatively planar conformation and to form their core scaffold appears to be coaxial helical stacking. Several stems – typically separated by internal loops and non-canonical structures – pack tightly in a coplanar fashion to form large quasi-two dimensional layers that can stack to produce a final structure that is generally only 1-2 helical layers deep in thickness, conferring their characteristic shape. Within these RNA molecules, distinct domains, such as the P4-P6 domain in group I introns [7] and the specificity and catalytic domains in bacterial RNase P [8,9,23] can be present with each domain exhibiting a quasi-planar, stem-rich fold. This type of domain architecture is prevalent in nearly all large RNA molecules characterized to date, including the RNA of the 30S subunit of the bacterial ribosome. Notable exceptions to this pattern include the large subunits of the bacterial and eukaryotic ribosomes, which contain multiple embedded proteins, forming isotropically arranged multi-domain RNP complexes [24,25].

Conservation of the functional core

Comparative sequence analysis of related RNA molecules from different organisms has shown that it is possible to distinguish conserved structural elements despite significant divergence in the corresponding sequences [26-28]. In many instances, covariant base-pair changes in helical stems are easily accommodated at the structural level. However, strongly conserved sequences have been noted in many RNA molecules and are associated with structurally and functionally important regions. In the case of bacterial RNase P, the sequence variability is large enough to warrant a classification scheme based on sequence-level similarity alone [29]. In fact, until recently it was unclear whether RNA molecules of identical function but with significant sequence differences would resemble each other structurally. Initial structural analysis of the specificity domains of two different RNase P molecules [8], followed by comparisons of various different group I introns [30] and of the full-length RNA component of RNase P from two bacteria [23], have shown that the core of these molecules is structurally conserved despite variable peripheral elements. These analyses demonstrate that different combinations of peripheral elements can stabilize a particular conserved core scaffold (Figure 2) and suggests, as previously proposed [8,30], that the evolutionary pressure to preserve the core RNA fold is strong.

Although the generality of this observation is unknown due to the paucity of determined structures of homologous RNA molecules showing significant sequence-level differences, it is possible to anticipate that the core of homologous molecules will likely be conserved even when the sequence variability is high. For example, determining whether different group II introns resemble each other at the catalytic core and share many of the essential structural

features will be an interesting test of this hypothesis, particularly because their sequence variability is even higher [31] than that observed in both group I introns and RNase P.

Preassembled functional cores

The structures of RNase P and of the group I and II introns all suggest predominantly preassembled and organized active sites that require metals for catalysis [11-13,32-34] (Figures 2 and 3). Whereas metal-assisted catalysis is not a universal feature of ribozymes, as there are both small [6] and large [35] ribozymes that operate without direct involvement of metals, preassembled active sites appear to be a common feature. A preassembled active site does not preclude local conformational changes or movements of important functional groups, but instead implies that the main components are already in place and that no large-scale conformational rearrangements are needed to bring them together for activity. This appears to hold true for both small and large RNA molecules with the caveat that defining the correct boundaries of small functional RNAs existing within the context of a larger RNA molecule is critical [3].

The active site environment is typically located within loop regions or at the junction of coaxial helical stacks. These elements are stabilized by RNA tertiary motifs, metal ions, and sometimes accessory protein subunits. Conserved non-canonical base pairs and/or extrahelical nucleobases are also common features at the active site. Whereas many small endonucleolytic ribozymes utilize the 2′-hydroxyl of an internal ribose for catalysis [36], both RNase P and group I introns use an external metal-activated nucleophile. These are not the only possibilities as group II introns utilize both the 2′-hydroxyl of an internal adenosine and at least two metal ions to catalyze the splicing reaction. Current data suggest that any significant conformational changes associated with catalytic activity are confined to local regions within the active site [37]. To provide a clearer understanding of their catalytic mechanism and to address active site plasticity in these large ribozymes, additional structures at atomic-resolution and at different states of their catalytic cycle are needed.

Molecular recognition

Comparative analysis of large RNA structures highlights the common theme of shape complementarity and specific interactions in target recognition (Figure 4). Whereas self-cleaving RNA molecules recognize their target cleavage site in *cis* to perform single-turnover reactions and primarily use specific base-pair interactions to facilitate recognition [36], multiple-turnover ribozymes, like RNase P and the ribosome, recognize their substrate in *trans*. In the case of bacterial RNase P, substrate recognition primarily occurs via specific base-stacking interactions between bases in highly conserved regions in the specificity domain and the tRNA, in addition to the formation of an A-minor motif [11]. Base-pairing involving the 3′-CCA tRNA sequence also contributes to recognition, but this is not a universal feature as some mature tRNAs acquire the 3′-CCA residues through a modification following 3′ and 5′ end-processing [38]. Structural evidence shows that group I introns utilize extensive tertiary motifs to create precisely-shaped binding scaffolds [12]. Hence, it appears that shape complementarity is fundamental to substrate recognition in large catalytic RNAs.

Although riboswitches recognize small metabolites, they also utilize both specific base-contacts and precise shape complementarity to distinguish exquisitely between closely related molecules. For instance, the lysine and purine riboswitches show at least a 5,000-fold level of discrimination between lysine and ornithine and a 10,000-fold level of discrimination between adenine and guanine, respectively [39]. One of the best examples of a selective and preassembled binding site is the large *glmS* ribozyme riboswitch [16,40], where the RNA structure does not change upon binding its target glucosamine-6-phosphate

molecule. Instead, ligand binding serves to function as the catalytic cofactor in the reaction. Taken together, these large ribozyme and riboswitch structures support the notion that RNA conforms to the same principles that are common to all macromolecular recognition processes: shape complementarity and specific atomic interactions. Not surprisingly, base pairing and base stacking are widespread in RNA recognition.

The role of adenosines

Of the four nucleobases, adenine appears to be preferred in functional cores and junctions. In RNase P, a constellation of conserved adenosines surrounds the active site scaffold and help to maintain the integrity of the RNase P active site environment. Interestingly, adenosine is the nucleotide that is universally conserved with greatest frequency across all RNase P/MRP family members [41]. Other large RNA molecules, such as the group I introns and the glycine riboswitch [12,15] also contain critical adenosines within loops and junctions. Conserved A-rich regions (J4/5 and J8/7) in group I introns are important for substrate recognition and splice-site selection, serving as a platform during exon ligation, helping to orient catalytic metals, and promoting the formation of stabilizing A-minor and ribose zipper tertiary interactions [30,42]. The glycine riboswitch, which utilizes two RNA domains to bind glycine cooperatively, contains both a cluster of purine nucleotides around its ligand-binding interface and a highly conserved adenosine that extrudes from the binding pocket to intercalate with an A-rich junction, facilitating interdomain tertiary contacts and ligand cooperativity [15]. Finally, approximately 43% of all 23S and 16S rRNA loop and interhelical bridging regions are composed of adenosines, corresponding to a sequence preference 18-28% greater than that of C, G, or U nucleotides [10]. Whether this purine base is indeed preferred at loop or interhelical regions in the assembly of higher order, multi-domain RNA enzymes remains to be confirmed, and will require the structural elucidation of additional large RNA and RNP complexes.

The role of proteins

Proteins greatly extend the functional versatility of RNA molecules by stabilizing a specific three-dimensional fold, by making an unstructured RNA region more accessible for its biological function, or by guiding molecular recognition. Group I and II introns and RNase P all contain important accessory protein subunits that assist in the folding and activation of the RNA catalyst. For example, the *Neurospora crassa* CYT-18 protein, a tyrosyl-tRNA synthetase, has adapted to function in splicing by binding and stabilizing some phage group I introns at a similar location as the peripheral group I intron P5abc domain [43]. Thus it appears that CYT-18 functionally replaces parts of the peripheral fold in some group I introns, consistent with the theme that proteins help to extend the functionality of a ribozyme and were likely later additions in evolutionary time. Similarly, DEAD-box proteins CYT-19 and/or Mss116 help to facilitate RNA folding and specific nucleic acid rearrangements *in vivo* in the group II splicing reaction [44,45]. In addition, it is likely that other maturase proteins form part of the *in vivo* group II RNP enzyme by binding and stabilizing core RNA domains and by guiding the flexible domain VI within the active site [46]. In contrast, the protein of bacterial RNase P is not known to affect significantly the folding or the final conformation of the RNA, but rather, docks on to the RNA surface of the ribozyme and serves to position the pre-cleaved tRNA 5′-leader region [11].

The RNase P and group I and II intron protein subunits are not embedded in the RNA structures, but are instead positioned at the surface or periphery, helping to promote a stable and functional RNP complex *in vivo*. This is not the case in more complex ribonucleoprotein complexes. For instance, essential RNAs in the spliceosome are surrounded and stabilized by multiple proteins [47]. Thus, the role of proteins spans a broad spectrum, with bacterial

RNase P and group I and II introns whose protein components are not essential *in vitro* near one end, and larger assemblies with multiple essential protein components, like the ribosome and the splicesome, near the other.

Conclusions and outlook

The advent of several large RNA crystal structures, both alone and in complex with their substrates and associated proteins, has permitted analysis of the general properties important for RNA architecture and the fundamental relations between structure and function. Large RNAs are composed mainly of packed stacked stems that are linked together by a variety of tertiary interactions, often forming extensive layers that aggregate into anisotropic molecules. While there is variability in the structure of homologous RNAs, which is mostly confined to peripheral elements, functional RNA molecules have well-defined structures with preassembled active sites or binding regions. As in all other biological molecules, recognition involves specific atomic interactions and shape complementarity, and in the case of RNA, base-pairing and base-stacking play a central role.

To date, most of the known structures correspond to the first functional RNA molecules identified. As more structures become available it is likely that some of these observations will need to be revisited or revised. Additional structures of more non-coding RNA molecules spanning a wider range of functions are needed, along with more examples of homologous RNA structures. This is an important task for the future if we want to achieve a complete understanding of RNA function and structure. Indeed, many surprises await as we discover even more about RNA and its numerous roles in biology.

Acknowledgments

Work was supported by a National Institutes of Health grant (GM58443) to AM, a NRSA post-doctoral fellowship to NJR, and an NIH Molecular Biophysics Training grant (T32GM008382) fellowship to CWC.

References and recommended reading

1. Kim SH, Suddath FL, Quigley GJ, McPherson A, Sussman JL, Wang AH, Seeman NC, Rich A. Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. Science. 1974; 185:435–440. [PubMed: 4601792]
2. Robertus JD, Ladner JE, Finch JT, Rhodes D, Brown RS, Clark BF, Klug A. Structure of yeast phenylalanine tRNA at 3 A resolution. Nature. 1974; 250:546–551. [PubMed: 4602655]
3. Martick M, Scott WG. Tertiary contacts distant from the active site prime a ribozyme for catalysis. Cell. 2006; 126:309–320. [PubMed: 16859740]
4. Pley HW, Flaherty KM, McKay DB. Three-dimensional structure of a hammerhead ribozyme. Nature. 1994; 372:68–74. [PubMed: 7969422]
5. Ferre-d'Amare AR, Zhou K, Doudna JA. Crystal structure of a hepatitis delta virus ribozyme. Nature. 1998; 395:567–574. [PubMed: 9783582]
6. Rupert PB, Ferre-D'Amare AR. Crystal structure of a hairpin ribozyme-inhibitor complex with implications for catalysis. Nature. 2001; 410:780–786. [PubMed: 11298439]
7. Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, Doudna JA. Crystal structure of a group I ribozyme domain: principles of RNA packing. Science. 1996; 273:1678–1685. [PubMed: 8781224]
8. Krasilnikov AS, Xiao Y, Pan T, Mondragón A. Basis for structural diversity in homologous RNAs. Science. 2004; 306:104–107. [PubMed: 15459389]
9. Krasilnikov AS, Yang X, Pan T, Mondragón A. Crystal structure of the specificity domain of ribonuclease P. Nature. 2003; 421:760–764. [PubMed: 12610630]
10. Holbrook SR. Structural principles from large RNAs. Annu Rev Biophys. 2008; 37:445–464. [PubMed: 18573090]

**11. Reiter NJ, Osterman A, Torres-Larios A, Swinger KK, Pan T, Mondragon A. Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. Nature. 2010; 468:784–789. [PubMed: 21076397] This study illustrates how RNase P is preassembled to engage in shape selective recognition of its tRNA target. The two structures support earlier models of substrate recognition and catalysis, identify the location of the proposed active site, and show how the protein component increases RNase P functionality.

12. Adams PL, Stahley MR, Kosek AB, Wang J, Strobel SA. Crystal structure of a self-splicing group I intron with both exons. Nature. 2004; 430:45–50. [PubMed: 15175762]

**13. Toor N, Keating KS, Taylor SD, Pyle AM. Crystal structure of a self-spliced group II intron. Science. 2008; 320:77–82. [PubMed: 18388288] One of the most elaborate RNA-only structures determined to date. This work reveals how multiple RNA helices come together to stabilize the catalytic core of this ribozyme. This study provides a crucial starting point for future structural investigations of group II introns and the spliceosome.

14. Dann CE 3rd, Wakeman CA, Sieling CL, Baker SC, Irnov I, Winkler WC. Structure and mechanism of a metal-sensing regulatory RNA. Cell. 2007; 130:878–892. [PubMed: 17803910]

**15. Huang L, Serganov A, Patel DJ. Structural insights into ligand recognition by a sensing domain of the cooperative glycine riboswitch. Mol Cell. 2010; 40:774–786. [PubMed: 21145485] Structures of free and ligand-bound states of one domain of the glycine riboswitch show how a unique RNA junction and metal ions create a highly selectively binding pocket for the smallest amino acid. Structural analysis and footprinting studies provide insight into how the entire, two domain riboswitch forms specific interdomain interactions, thereby facilitating cooperative ligand binding.

16. Klein DJ, Ferre-D'Amare AR. Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. Science. 2006; 313:1752–1756. [PubMed: 16990543]

**17. Serganov A, Huang L, Patel DJ. Structural insights into amino acid binding and gene control by a lysine riboswitch. Nature. 2008; 455:1263–1267. [PubMed: 18784651] Structures of free and ligand-bound states of the riboswitch are presented, not only illustrating how long-range tertiary interactions help to stabilize a multi-junction RNA core, but also detailing how a specific metal and hydrogen bonding network within the binding pocket give rise to shape complementarity. The ligand binding capabilities of the riboswitch are also explored though multiple structures containing lysine-related small molecules.

18. Baird NJ, Kulshina N, Ferre-D'Amare AR. Riboswitch function: Flipping the switch or tuning the dimmer? RNA Biology. 2010; 7:328–332. [PubMed: 20458165]

19. Garst AD, Batey RT. A switch in time: detailing the life of a riboswitch. Biochim Biophys Acta. 2009; 1789:584–591. [PubMed: 19595806]

20. Serganov A. Determination of riboswitch structures: light at the end of the tunnel? RNA Biol. 2010; 7:98–103. [PubMed: 20061809]

*21. Zhang J, Lau MW, Ferre-D'Amare AR. Ribozymes and Riboswitches: Modulation of RNA Function by Small Molecules. Biochemistry. 2010; 49:9123–9131. [PubMed: 20931966] An insightful review describing how an RNA molecule can recognize a small molecule and act as a genetic switch.

22. Cruz JA, Westhof E. The dynamic landscapes of RNA architecture. Cell. 2009; 136:604–609. [PubMed: 19239882]

23. Torres-Larios A, Swinger KK, Pan T, Mondragon A. Structure of ribonuclease P--a universal ribozyme. Curr Opin Struct Biol. 2006; 16:327–335. [PubMed: 16650980]

24. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science. 2000; 289:905–920. [PubMed: 10937989]

25. Ben-Shem A, Jenner L, Yusupova G, Yusupov M. Crystal structure of the eukaryotic ribosome. Science. 2010; 330:1203–1209. [PubMed: 21109664]

26. Haas ES, Brown JW. Evolutionary variation in bacterial RNase P RNAs. Nucleic Acids Res. 1998; 26:4093–4099. [PubMed: 9722626]

27. Michel F, Westhof E. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. J Mol Biol. 1990; 216:585–610. [PubMed: 2258934]

28. Zimmerly S, Hausner G, Wu X. Phylogenetic relationships among group II intron ORFs. Nucleic Acids Res. 2001; 29:1238–1250. [PubMed: 11222775]
29. Brown JW, Haas ES, James BD, Hunt DA, Liu JS, Pace NR. Phylogenetic analysis and evolution of RNase P RNA in proteobacteria. J. Bacteriol. 1991; 173:3855–3863. [PubMed: 1711030]
30. Vicens Q, Cech TR. Atomic level architecture of group I introns revealed. Trends Biochem. Sci. 2005
31. Michel F, Umesono K, Ozeki H. Comparative and functional anatomy of group II catalytic introns--a review. Gene. 1989; 82:5–30. [PubMed: 2684776]
32. Kazantsev AV, Krivenko AA, Harrington DJ, Holbrook SR, Adams PD, Pace NR. Crystal structure of a bacterial ribonuclease P RNA. Proc. Natl. Acad. Sci. U. S. A. 2005; 102:13392–13397. [PubMed: 16157868]
33. Toor N, Rajashankar K, Keating KS, Pyle AM. Structural basis for exon recognition by a group II intron. Nat Struct Mol Biol. 2008; 15:1221–1222. [PubMed: 18953333]
34. Torres-Larios A, Swinger KK, Krasilnikov AS, Pan T, Mondragon A. Crystal structure of the RNA component of bacterial ribonuclease P. Nature. 2005; 437:584–587. [PubMed: 16113684]
35. Schmeing TM, Huang KS, Kitchen DE, Strobel SA, Steitz TA. Structural insights into the roles of water and the 2′ hydroxyl of the P site tRNA in the peptidyl transferase reaction. Mol Cell. 2005; 20:437–448. [PubMed: 16285925]
36. Ferre-D'Amare AR, Scott WG. Small Self-cleaving Ribozymes. Cold Spring Harb. Perspect. in Biol. 2010; 2
**37. Lipchock SV, Strobel SA. A relaxed active site after exon ligation by the group I intron. Proc Natl Acad Sci U S A. 2008; 105:5699–5704. [PubMed: 18408159] Structures of three conformational states related to the goup I exon ligation step of the reaction are presented, detailing the role of metal ions in catalysis and providing mechanistic insight into exon binding and product release.
38. Hartmann RK, Gossringer M, Spath B, Fischer S, Marchfelder A. The making of tRNAs and more - RNase P and tRNase Z. Prog Mol Biol Transl Sci. 2009; 85:319–368. [PubMed: 19215776]
39. Garst AD, Edwards AL, Batey RT. Riboswitches: Structures and Mechanisms. Cold Spring Harb Perspect Biol. 2010
40. Link KH, Guo L, Breaker RR. Examination of the structural and functional versatility of glmS ribozymes by using in vitro selection. Nucleic Acids Res. 2006; 34:4968–4975. [PubMed: 16982640]
41. Walker SC, Engelke DR. Ribonuclease P: the evolution of an ancient RNA enzyme. Crit Rev Biochem Mol Biol. 2006; 41:77–102. [PubMed: 16595295]
42. Stahley MR, Strobel SA. Structural evidence for a two-metal-ion mechanism of group I intron splicing. Science. 2005; 309:1587–1590. [PubMed: 16141079]
**43. Paukstelis PJ, Chen JH, Chase E, Lambowitz AM, Golden BL. Structure of a tyrosyl-tRNA synthetase splicing factor bound to a group I intron RNA. Nature. 2008; 451:94–97. [PubMed: 18172503] The co-crystal structure of the Twort group I ribozyme in complex with the *N. crassa* tyrosyl-tRNA synthetase CYT-18 shows elegantly how proteins could have incrementally replaced RNA structures during the transition from an ancient RNA-based world to an RNP world.
44. Karunatilaka KS, Solem A, Pyle AM, Rueda D. Single-molecule analysis of Mss116-mediated group II intron folding. Nature. 2010; 467:935–939. [PubMed: 20944626]
45. Zingler N, Solem A, Pyle AM. Dual roles for the Mss116 cofactor during splicing of the ai5gamma group II intron. Nucleic Acids Res. 2010; 38:6602–6609. [PubMed: 20554854]
46. Toor N, Keating KS, Fedorova O, Rajashankar K, Wang J, Pyle AM. Tertiary architecture of the Oceanobacillus iheyensis group II intron. RNA. 2010; 16:57–69. [PubMed: 19952115]
47. Krummel, DA Pomeranz; Oubridge, C.; Leung, AK.; Li, J.; Nagai, K. Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. Nature. 2009; 458:475–480. [PubMed: 19325628]
48. Wimberly BT, Brodersen DE, Clemons WM Jr, Morgan-Warren RJ, Carter AP, Vonrhein C, Hartsch T, Ramakrishnan V. Structure of the 30S ribosomal subunit. Nature. 2000; 407:327–339. [PubMed: 11014182]

49. Guo F, Gooding AR, Cech TR. Structure of the Tetrahymena ribozyme: base triple sandwich and metal ion at the active site. Mol. Cell. 2004; 16:351–362. [PubMed: 15525509]

50. Golden BL, Kim H, Chase E. Crystal structure of a phage Twort group I ribozyme-product complex. Nat. Struct. Mol. Biol. 2005; 12:82–89. [PubMed: 15580277]

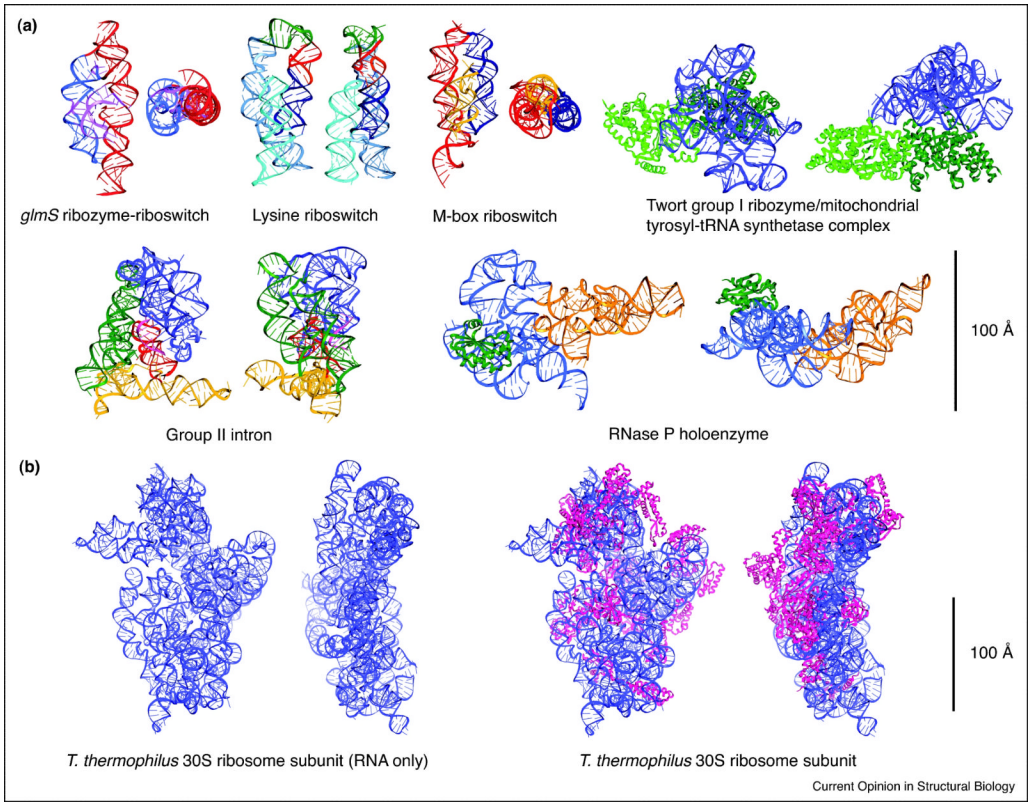

Figure 1. Architecture of RNA molecules

A) Diagrams of various RNA molecules (> 100 nucleotides) illustrating some of their general properties. Two perpendicular views of each structure are shown. Large RNAs are formed by stacked stems that are stabilized by a variety of tertiary interactions. The *Thermoanaerobacter tengcongensis glmS* riboswitch ribozyme structure illustrates the packing of two long stems that form a single, extended layer. The *Thermotoga maritima* lysine riboswitch illustrates an elongated molecule formed by three packed stems, forming a two layer molecule. The two major stems interact with each other through a kissing loop (red/orange). The *Bacillus subtilis* M-box riboswitch also forms parallel stacked stems. More complex molecules have a more globular, but still quasi-planar structure. In some cases, the addition of protein increases their functionality. In the case of the phage Twort group I ribozyme, binding to the homodimeric *N. crassa* CYT-18 protein promotes the splicing of many introns. Note that the protein binds on the surface of the RNA. The *Oceanobacillus iheyensis* group II intron structure represents a remarkable example of an RNA molecule with a complex fold. This large RNA forms a tight, compact structure around the active site region (red). Similar to proteins, large RNA molecules can consist of functional domains. The structure of the RNase P holoenzyme is an excellent example of a ribozyme composed of two functional domains, a catalytic (blue) and a specificity (yellow) domain. In this case, the protein also binds on the surface and does not alter the RNA structure. **B)** Diagrams of the structure of the *Thermus thermophilus* 30S ribosome subunit with and without the protein subunits. The RNA moiety (left) is also quasi-planar, despite its large size. In the presence of the proteins (right) the overall shape remains, with the protein molecules mostly found on the surface. The structures (on a clockwise fashion starting from the top left corner) correspond to PDBs: 2Z75, 3DIL, 2QBZ, 2RKJ, 3IGI, and 3OK7 [11,13,14,16,17,43,48]. The structures on each panel are drawn at the same scale, the bar on the right corresponds to 100 Å.

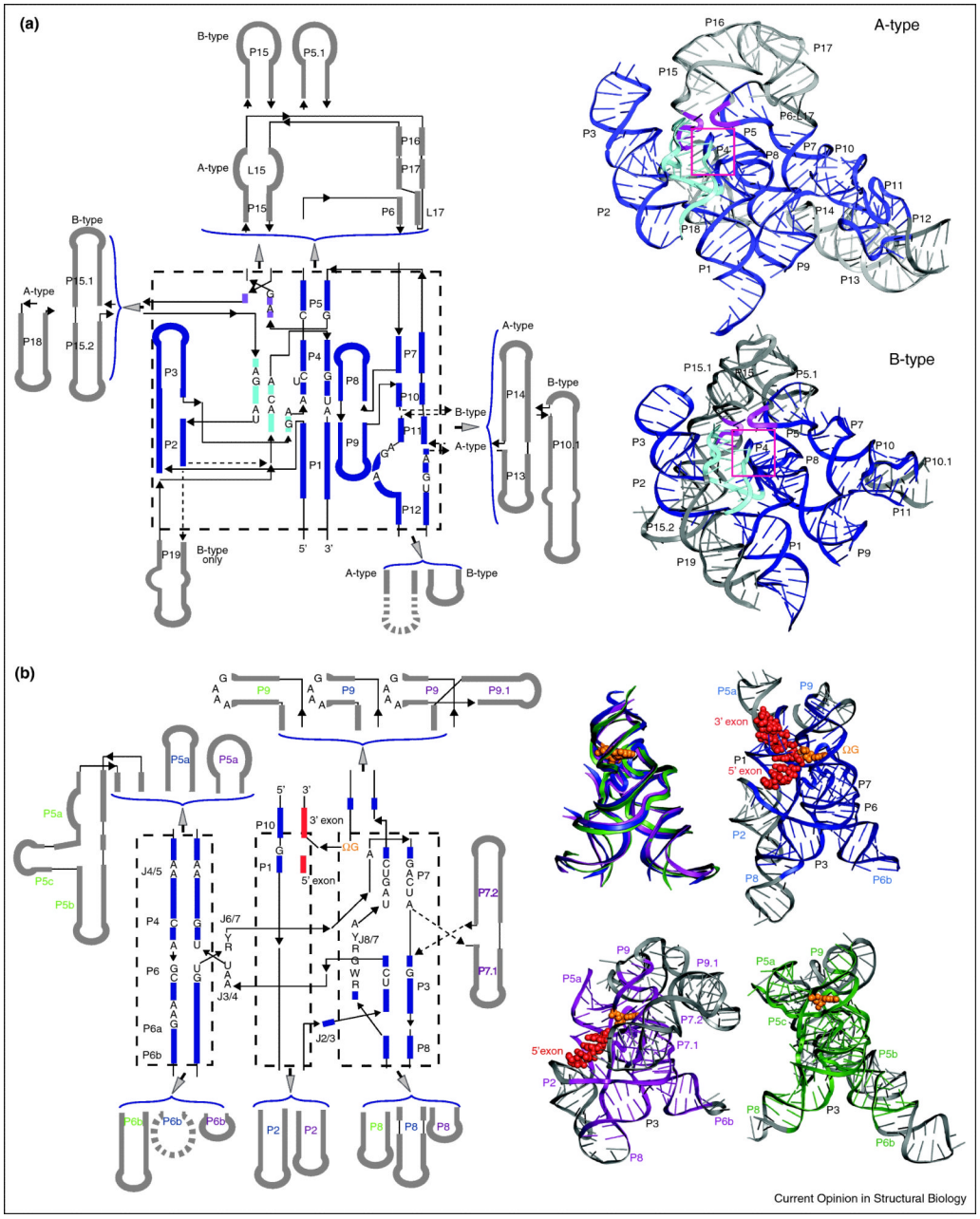

Figure 2. Large ribozymes have defined core architectures that are similar across different organisms

A) Secondary structure schematic diagram emphasizing the shared functional core between the known ancestral (A-type) and bacillus (B-type) P RNA structures of RNase P [11,32]. The dashed box encompasses elements involved in core formation, whereas grey denotes regions that are variable between the structures. Regions that form junctions and that contain universal (cyan) or bacterially (purple) conserved nucleotides are also shown. All conserved nucleotides are labeled. Crystal structures of the *T. maritima and Bacillus stearothermophilus* P RNAs are shown for comparison [11,32] (PDB 3OK7 and 2A64). **B)** Schematic diagram of the shared functional core amongst three group I intron structures from different organisms. Colors and labels are as defined in **A)** with the green, blue, and purple labels in the variable grey helical region corresponding to the different intron

structures from *Tetrahymena thermophila*, *Azoarcus sp. BH72*, and phage Twort, respectively [12,49,50] (PDB 1X8W, 1U6B, 1Y0Q) [12,49,50]. Despite the presence of different peripheral domains, all three group I structures exhibit the same core domain architecture. Grey dashed regions represent engineered RNA motifs added to promote crystal formation. The diagram in **B)** was adapted from [30].

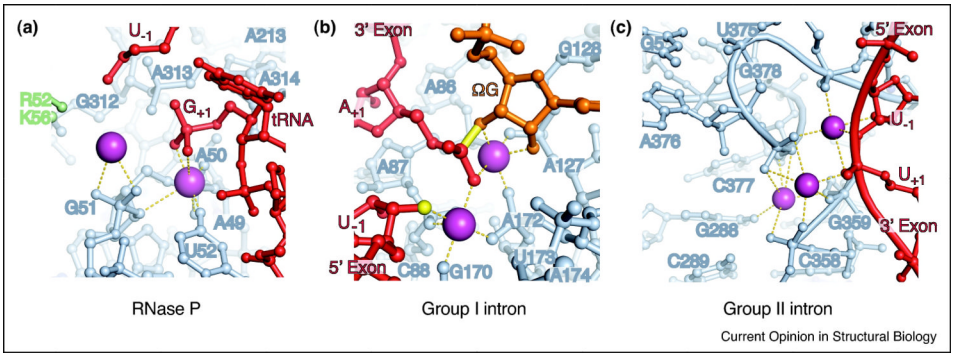

Figure 3. Proposed active sites based on RNase P, group I, and group II intron structures

Whereas these catalytic RNAs require at least two metal ions, putatively magnesium (pink spheres), not all ribozymes contain metal cofactors within their active sites. In these large RNA molecules the overall active site structure does not change upon substrate binding and is largely preassembled, although correct substrate positioning likely enhances the rigidity of the active site and binding of catalytic metal ions. **A)** Structure of the *T. maritima* RNase P holoenzyme in complex with tRNA [11] (PDB 3OKB) shows components that comprise the active site scaffold. **B)** The *Azoarcus* group I intron (ribo-ΩG) with bound exons [42] (PDB 1ZZN). The group I active site (light blue) contains two A-rich regions and structurally conserved base triples that help to position an external nucleotide (ΩG, orange) and exons (red) during cleavage and ligation. The nucleophile and scissile bond for the ligation step are shown in yellow. **C)** Structure of the *O. iheyensis* group II intron shows three metal ions in the region where exon ligation occurs [33,46] (PDB 3EOG). Yellow dashed lines indicate metal-ligand interactions that are ≤ 3.0 Å for the RNase P and group II structures, and within ≤ 2.2 Å for the *Azoarcus* group I structure.

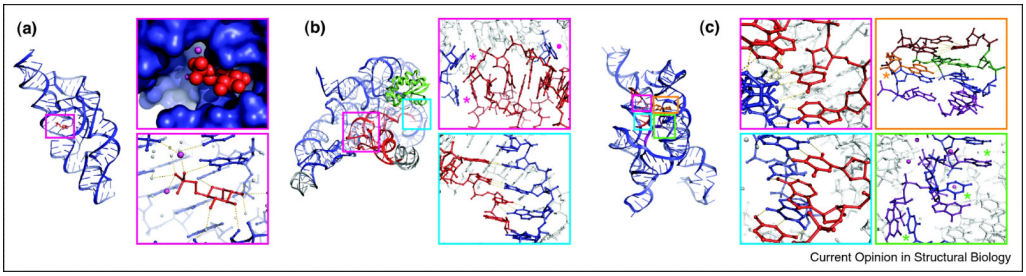

Figure 4. Recognition of molecular targets by large RNA molecules is based on shape complementarity and specific interactions

A) *T. tengcongensis glmS* ribozyme riboswitch [16] (PDB 2Z75) in complex with its ligand cofactor, glucosamine-6-phosphate (red), reveals the importance of specific stabilizing interactions that orient and position the substrate within the cofactor binding pocket. **B)** *T. maritima* RNase P holoenzyme recognition of $tRNA^{Phe}$ (red) [11] (PDB 3OKB) is primarily achieved by precise positioning of two base-stacks (marked by an *) and an A-minor motif (•) in the specificity domain with the unique, tertiary fold of the tRNA TΨC- and D-loops (pink panel). The complex is further stabilized by base-pairing interactions between the RNase P catalytic domain and the tRNA 3′-CCA sequence (blue panel). **C)** *Azoarcus* group I intron [42] (PDB 1ZZN) with both exons (red) positioned by base-pairing and base-stacking interactions in a catalytic state prior to the exon ligation step (pink and blue panels). Three base-stacks (*, green panel) stabilize an important junction (dark purple) in close proximity to the active site. The 3′-terminal guanosine (ΩG,*) nucleophile is both positioned by and embedded within a conserved cluster of stacked base triples (orange panel). In all panels, purple and white spheres represent magnesium atoms and water molecules, respectively.