Emerging themes in RNA folding

Jennifer A Doudna and Elizabeth A Doherty

RNAs, like proteins, readily form specific structures adapted for ligand binding and catalysis. Since they are composed of completely different chemical building blocks, however, RNAs and proteins necessarily use distinct strategies to assemble complex architectures. While burial of hydrophobic residues drives protein folding, the hydrophobic effect in RNA contributes primarily to the formation of secondary structure. To form tertiary structure, RNA must overcome electrostatic repulsions from the phosphate backbone. How do negatively charged double helices pack together to produce catalytic centers and ligand binding surfaces? Here, we review our understanding of the principles that underlie RNA folding based on the structural information currently available.

Address: Department of Molecular Biophysics and Biochemistry and Howard Hughes Medical Institute, Yale University, New Haven, CT 06520-8114, USA.

Correspondence: Jennifer A Doudna
E-mail: doudna@csb.yale.edu

Folding & Design 01 October 1997, **2**:R65–R70
http://biomednet.com/elecref/13590278002R0065

© Current Biology Ltd ISSN 1359-0278

Introduction

Despite its limited functional group diversity, RNA forms a remarkable variety of complex structures capable of specific ligand recognition and catalysis. How does ribonucleic acid achieve a level of structural complexity once thought confined to proteins? While protein secondary structure positions amino acid sidechains on the surface of α-helices and β-sheets, optimally located for tertiary contacts, the opposite is true in RNA. Here, the double helical secondary structure positions the unique chemical groups of the nucleotides at the interior of base-paired duplexes in an environment largely inaccessible for tertiary structural interaction. As a result, the molecular surfaces of canonical RNA helices are relatively smooth and featureless, with a high density of negative charge from the phosphate backbone. The A-form duplex of RNA is thus a less-than-ideal building block for the construction of three-dimensional shapes.

Unlike its DNA sibling, however, RNA rarely exists as a long uninterrupted double helix in nature. Instead, structured RNAs such as ribozymes and ribosomal subunits contain a series of short helical regions interspersed with noncanonical base pairs and unpaired bases. These nonstandard segments are critical for RNA tertiary structure formation, enabling A-form helices to pack together to produce unique architectures. In this review, we discuss our current understanding of the molecular basis for RNA folding and contrast the emerging themes in this rapidly growing field to those known for protein folding. Excellent recent reviews have described aspects of RNA folding defined by studies of transfer RNA (tRNA) and ribosomal RNAs [1–3]. Here, we focus primarily on biochemical and biophysical studies of the *Tetrahymena* group I intron, an autocatalytic RNA capable of excising itself from a primary transcript without the need for protein cofactors. The *Tetrahymena* ribozyme is one of only three classes of multi-helical RNAs for which high-resolution structural information is available [4–7]. The crystal structure of the 160-nucleotide P4–P6 domain of this ribozyme [7] has provided important clues about the rules for helical packing in RNA.

The folding problem revisited

How does an extended biopolymer fold into a specific shape adapted for catalysis or molecular recognition? This fascinating biophysical question has traditionally been addressed for proteins, but it is now clear that RNA molecules are similarly characterized by unique folding pathways and structural motifs. In both proteins and RNAs, the hydrophobic effect, hydrogen bonding, metal ion coordination and van der Waals' forces all contribute to formation of compact molecular structures.

The hydrophobic effect plays a central role in driving protein folding [8], as shown by a considerable number of experiments ranging from mutagenesis [9–11] through calorimetry [12,13] to simulations of folding [14,15]. Polar interactions such as hydrogen bonds, though abundant, appear to play a less significant energetic role in protein structures since they occur both within folded proteins and between residues and solvent in the denatured state [8,16]. As a result, short α-helices and β-sheets are usually stable only in the context of an entire protein or independent domain. In contrast, RNA double helices have essentially the same structures in the context of a complex tertiary structure as alone in solution. Hydrophobic effects in RNA occur mainly at the level of secondary structure, making a contribution to the vertical stacking of purine and pyrimidine bases [17]. Thus, RNA necessarily requires a different strategy than protein to drive tertiary structure assembly.

In both protein and RNA, folding is opposed by a large configurational entropy due to the reduction in local backbone rotations and the compactness of native states [8]. RNA folding is additionally opposed by electrostatic

repulsion from the negatively charged phosphate backbone. Nevertheless, to achieve the level of compactness required for the function of many ribozymes, helices must pack together. Tertiary folding mechanisms in RNA thus involve both alterations in backbone geometry to prevent close association of phosphate groups, and direct and indirect metal ion coordination, as discussed below.

Loops and bulges in RNA helix packing

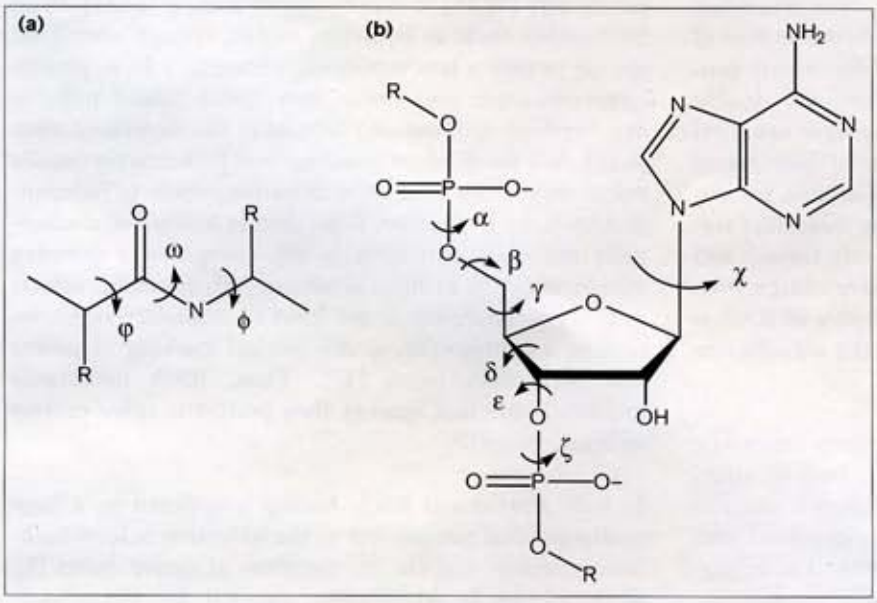
Polynucleotides and proteins feature fundamentally different backbone and sidechain torsional constraints (Figure 1). The peptide backbone is fairly rigid with little allowed rotation about two of the angles and almost none about the third. Greater rotational freedom exists for sidechains [18]. In contrast, RNA contains six torsion angles for the backbone and one for the base. While the positions of these angles are partially interdependent [19–21], they easily allow non-Watson–Crick-paired residues to twist out of their normal helical perch and adopt many noncanonical orientations useful for tertiary structure formation or protein recognition.

Nucleotides that do not form standard Watson–Crick base pairs are common and often phylogenetically conserved at junctions between helical stems, in bulges and loops within helices and in terminal loops. Although experiments on short duplex RNAs suggested that loops and bulges are energetically destabilizing [22], small RNA oligonucleotide structures have shown that unpaired or loop nucleotides are often well ordered. These residues form either nonstandard base pairs or multi-base contacts, or they flip out of the helix to interact with adjacent molecules in a crystal lattice [23–26].

Structures of larger RNA molecules show that these noncanonical regions are key to higher-order RNA folding. In crystal structures of tRNA, nonstandard base pairs and base triples occur near the hinge and outer corner of the L where the two helical domains are joined [27–30] (Figure 2a). Changes in sugar pucker and backbone torsion angles facilitate turns and kinks in the RNA backbone, allowing cross-strand base pairing or base intercalation [21,31]. Likewise, the three helical arms of the wishbone-shaped hammerhead ribozyme are organized by a core of non-Watson–Crick base pairs (Figure 2b). Two of the arms stack coaxially, while the third, adjacent to the cleavage site, juts out from the core at an angle dictated by a sharp turn in the phosphodiester backbone of the core at a highly conserved CUGA sequence, the U-turn [5,6].

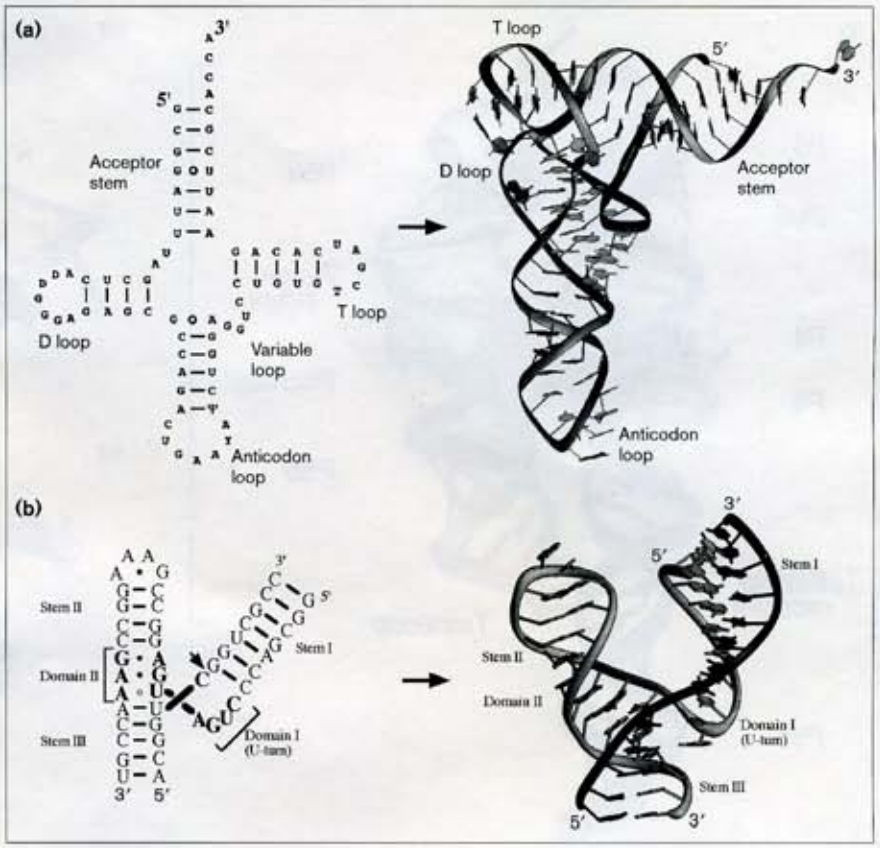
Larger RNAs containing hundreds of nucleotides, such as self-splicing introns and ribosomal RNA, form compact structures that necessarily involve extensive helical packing not seen in tRNA or the hammerhead ribozyme. The crystal structure of the 160-nucleotide P4–P6 domain of the *Tetrahymena* group I self-splicing intron revealed for the first time how RNA uses noncanonical loop and bulge motifs to mediate intramolecular helical packing. The domain is shaped like a candycane in which three stacked helices of the conserved catalytic core dock against a three-helix junction called P5abc [7] (Figure 3a). Two specific long-range interactions clamp the two halves of the domain together. First, a metal-stabilized adenosine-rich corkscrew binds in the minor groove of the P4 helix (Figure 3b; see below). Second, a GAAA tetraloop binds to a conserved 11-nucleotide internal loop called the tetraloop receptor (Figure 3c). The tetraloop–tetraloop receptor interaction

Figure 1



Comparison of peptide and RNA torsional constraints. **(a)** Backbone torsion angles in peptides. Amino acid sidechains, which contain up to five additional torsion angles, are not shown but are denoted by R. **(b)** Torsion angles in ribonucleotides include six backbone angles, α, β, γ, δ, ε, and ζ, and one base angle, χ.

Figure 2



Secondary and tertiary structure diagrams of **(a)** yeast $tRNA^{Phe}$ and **(b)** a hammerhead ribozyme. Dashes between bases in the secondary structure diagrams indicate Watson–Crick pairing between bases, while open and closed circles between bases indicate wobble or other types of base pairs. The cleavage site on the hammerhead substrate strand is marked with an arrow. The CUGA U-turn motif in $tRNA^{Phe}$ similar to the one in the hammerhead ribozyme is located in the anticodon loop.

creates a network of cross-helix base stacking and hydrogen bonding facilitated by a widened minor groove in the receptor. Phylogenetic analyses have indicated that this interaction occurs frequently in several classes of RNA [32]. At both sites of backbone contact, ribose 2′ hydroxyls form hydrogen bonds that stitch together the two helical stacks.

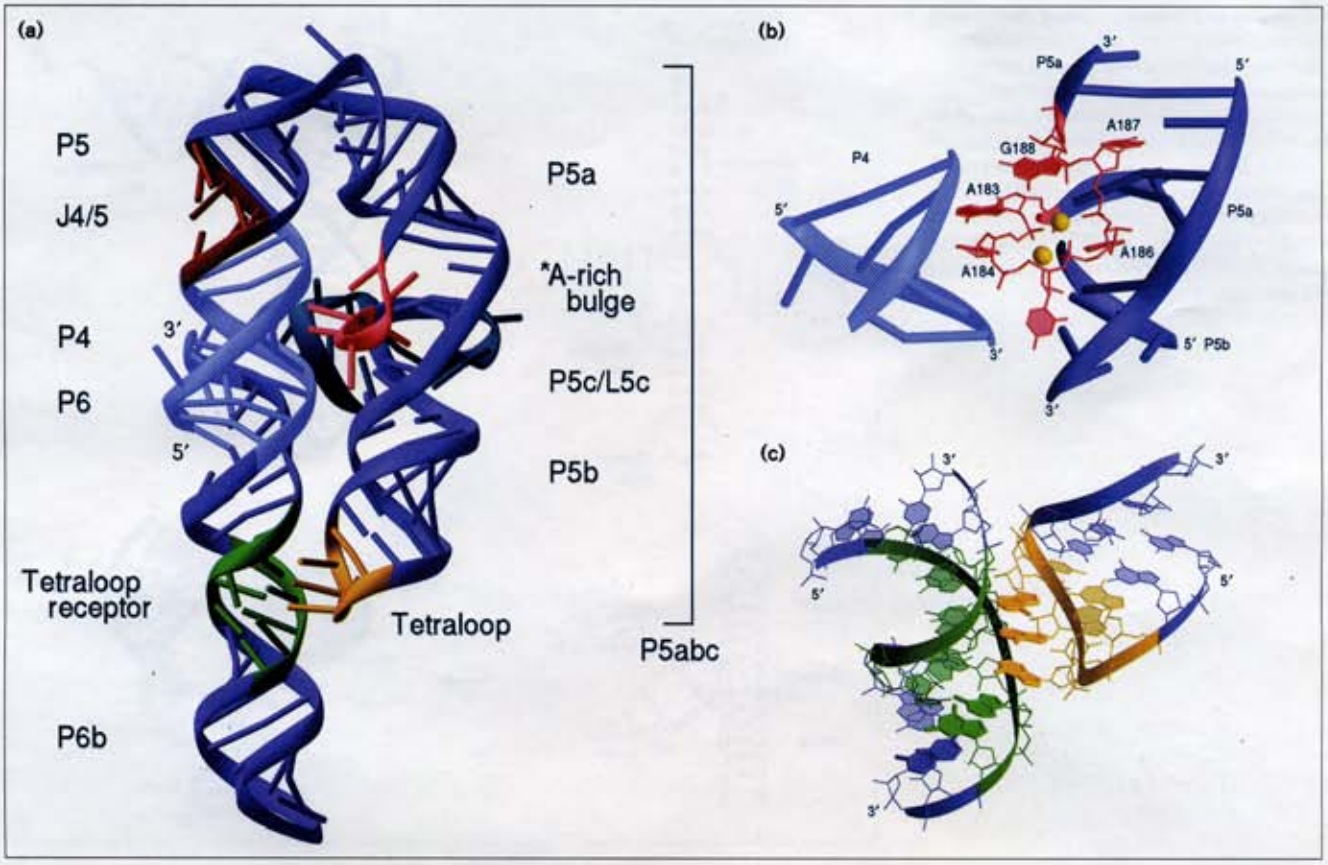
Strikingly, common structural motifs have emerged within the limited number of RNA structures currently available. The U-turn motif of the hammerhead ribozyme (Figure 2b) is virtually identical to the U-turn first observed in the anticodon and pseudouridine loops of phenylalanine tRNA [4–6]. The conformation of the GAAA tetraloop motif, common in many RNA phylogenies [32,33], is the same in a small RNA hairpin, the hammerhead ribozyme and the P4–P6 domain despite differences in molecular environments [5–7,34]. Other examples include an AA platform motif that occurs at three places in the crystal structure of the *Tetrahymena* ribozyme P4–P6 domain [35] (Figure 4) and sheared purine base pairs (A–A and G–A) that occur in the P4–P6 domain [7] and in the hammerhead ribozyme [5,6] (Figure 5). These observations imply that the library of RNA structural motifs is relatively small, and it may be possible to learn the rules and apply them to the rational design of RNA molecules [2].

Magnesium-dependent helix packing in the *Tetrahymena* ribozyme

RNA folding pathways involve multiple transitions and stable intermediates [36–39]. Multi-helical RNAs assemble hierarchically in a metal ion dependent manner, both kinetically [40,41] and at equilibrium [42–44], with secondary structure formation preceding a buildup of higher-order assemblies. Divalent cations, which provide excellent screening of charge from the phosphate backbone, stabilize the tertiary structures of tRNA and hammerhead ribozymes at relatively low concentrations [45,46]. High concentrations of monovalent cations also enable these RNAs to adopt their active structures [5,36].

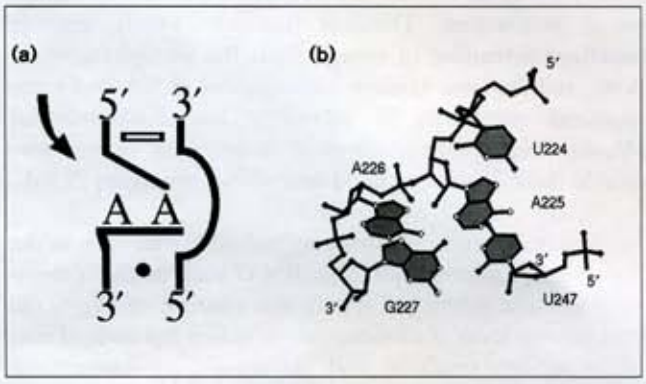
In contrast, magnesium ions play indispensable roles in the higher-order assembly of large RNAs such as the *Tetrahymena* group I intron [47,48]. In the absence of Mg^{2+}, the intron forms most of its secondary structure but little, if any, of its tertiary structure [49]. At least three cooperative folding transitions occur as a function of magnesium ion concentration on the intron folding pathway, as monitored by the appearance of solvent-inaccessible backbone positions and internal cross-links [38,41,50]. The P5abc subdomain folds at the lowest concentration of magnesium [50]. Next, the P5abc subdomain docks against the core helices

Figure 3



(a) Tertiary structure of the *Tetrahymena* P4–P6 domain with base-paired regions (P) and one internal loop (J) marked. Elements of the conserved catalytic core are in light blue and red. The adenosine-rich bulge is in pink. The GAAA tetraloop and tetraloop receptor are in yellow and green. The P5c helix (blue-gray) is behind stems P5a and P5b. **(b)** Close-up of the adenosine-rich bulge and the P5a, P5b, P5c (P5abc) triple junction. Helix P5c is removed for clarity. Magnesium ions are shown as gold spheres. **(c)** Close-up of the GAAA tetraloop–tetraloop receptor interaction.

Figure 4

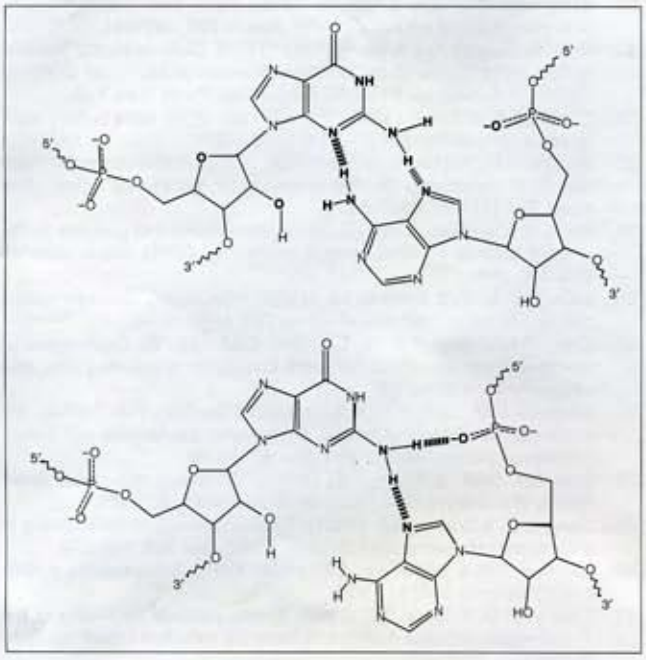


The AA platform stacking motif found at three locations in the *Tetrahymena* P4–P6 domain. **(a)** The arrow demonstrates how side-by-side positioning of two adenosine residues alters the structure of an internal loop. **(b)** Stacking pattern of the AA platform in the tetraloop receptor of P4–P6.

in the P4–P6 domain at a slightly higher magnesium ion concentration. Finally, the remaining helices and joining regions assemble on the P4–P6 scaffold to form the active structure [41,44,51]. The results of kinetic experiments support this folding pathway and suggest that P5abc and P4–P6 structure formation are fast relative to formation of the final active conformation [40,41,52].

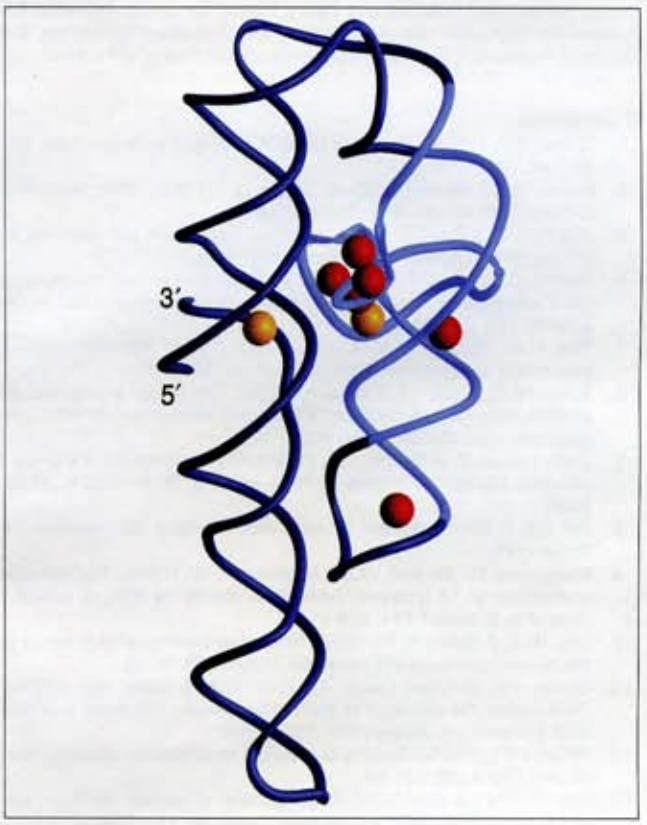
Crystallographic and biochemical experiments on the P4–P6 domain show that a group of magnesium ions positioned in the adenosine-rich corkscrew and the P5abc triple-helix junction (Figure 3b) help orchestrate the first two tertiary folding transitions of the ribozyme (Figure 6). In the P4–P6 domain crystal structure, these ions are coordinated directly or indirectly through water to nonbridging phosphate oxygens or directly to oxygens on guanosine bases [53]. Sulfur substitution at any of the four directly coordinated phosphate oxygens prevents magnesium binding, completely destroying folding of both the P5abc subdomain and the P4–P6 domain [53]. The A-rich

Figure 5



Conformation of sheared G–A base pairs. Two examples of hydrogen-bonding schemes are shown. Within each scheme, bold atoms are sometimes or always involved in hydrogen bonds. Thick dashed lines indicate hydrogen bonds that are always present.

Figure 6



Ribbon diagram of the P4–P6 domain showing the positions of magnesium ion binding sites determined crystallographically by replacement with manganese. The P5abc subdomain is shown in light blue. Red spheres indicate sites occupied by metal ions in both molecules of the crystallographic asymmetric unit. Yellow spheres are sites occupied in one of the two molecules.

bulge/triple-helix junction region forms an exceedingly complex structure involving dramatic backbone kinks and twists and internalization of several phosphate groups [7]. Two of the magnesium ions cross-link noncontiguous phosphate oxygens in the adenosine-rich bulge, holding the corkscrew in place so that adenosine bases in the bulge can flip out, making contacts with surrounding helices (Figure 3b). All of the magnesium ions allow burial of phosphate oxygens from solvent, forming a highly charged interior core not unlike a hydrated salt [53].

This metal ion core is somewhat analogous to the hydrophobic core of proteins in that both allow burial of parts of the chain, providing a scaffold for the construction of specific exterior surfaces. The need to neutralize charge and the absolute dependence on divalent metals for stability suggests that a metal ion core may be a general feature of large RNA molecules. It remains to be determined, however, whether magnesium coordination provides the primary energetic contribution to P4–P6 domain tertiary structure. The interior core of the domain also features extensive base-stacking and hydrogen-bonding interactions which contribute to the formation of specific metal ion binding sites. The P4–P6 domain provides an excellent model system for examining how metal ion coordination, base stacking and other interactions stabilize complex RNAs. An understanding of the balance of these forces will be essential to the study of ribozyme function and RNA–protein interactions.

Conclusions

The large number of available solution and X-ray crystal structures has allowed extensive investigation of the factors that stabilize proteins. While RNA molecules also form complex architectures involving the internalization of the backbone from solvent, only three classes of RNA structures are known in atomic detail. The *Tetrahymena* ribozyme P4–P6 domain demonstrates extensive use of magnesium ions to facilitate packing of noncanonical loops into the back of a distant helical stack. Will other RNAs pack duplexes together using an analogous metal ion core, or will other types of interactions primarily stabilize helix packing? With the limited number of atomic resolution structures currently available, we are just scratching the surface of this new territory. Hopefully, additional structure determinations will soon help us to dig deeper into this vast macromolecular folding problem.

Acknowledgements

We thank Jamie Cate and Don Engelman for helpful discussions, and Dan Battle, Adrian Ferré-D'Amaré and Raven Hanna for critical review of the manuscript. This work was supported by the Donaghue Foundation, the Beckman Foundation, the Lucille Markey Charitable Trust and the NIH.

References

1. Draper, D.E. (1996). Strategies for RNA folding. *Trends Biochem. Sci.* **21**, 145-149.
2. Westhof, E., Masquida, B. & Jaeger, L. (1996). RNA tectonics: towards RNA design. *Fold. Des.* **1**, R78-R88.
3. Westhof, E. & Patel, D. (1997). Nucleic acids: from self-assembly to induced fit recognition. *Curr. Opin. Struct. Biol.* **7**, 305-309.
4. Kim, S.H., *et al.*, & Rich, A. (1973). Three-dimensional structure of yeast phenylalanine transfer RNA: folding of the polynucleotide chain. *Science* **179**, 285-288.
5. Pley, H.W., Flaherty, K.M. & McKay, D.B. (1994). Three-dimensional structure of a hammerhead ribozyme. *Nature* **372**, 68-74.
6. Scott, W.G., Finch, J.T. & Klug, A. (1995). The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage. *Cell* **81**, 995-1002.
7. Cate, J.H., *et al.*, & Doudna, J.A. (1996). Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **273**, 1678-1685.
8. Dill, K.A. (1990). Dominant forces in protein folding. *Biochemistry* **29**, 7133-7155.
9. Matsumura, M., Becktel, W.J. & Matthews, B.W. (1988). Hydrophobic stabilization in T4 lysozyme determined directly by multiple substitutions of Ile 3. *Nature* **334**, 406-410.
10. Lim, W.A. & Sauer, R.T. (1989). Alternative packing arrangements in the hydrophobic core of λ repressor. *Nature* **339**, 31-36.
11. Bowie, J.U., Reidhaar-Olson, J.F., Lim, W.A. & Sauer, R.T. (1990). Deciphering the message in protein sequences: tolerance to amino acid substitutions. *Science* **247**, 1306-1310.
12. Privalov, P.L. (1979). Stability of proteins: small globular proteins. *Adv. Protein Chem.* **33**, 167-241.
13. Privalov, P.L. & Gill, S.J. (1988). Stability of protein structure and hydrophobic interaction. *Adv. Protein Chem.* **39**, 191-234.
14. Dill, K.A., Alonso, D.O. & Hutchinson, K. (1989). Thermal stabilities of globular proteins. *Biochemistry* **28**, 5439-5449.
15. Dill, K.A., *et al.*, & Chan, H.S. (1995). Principles of protein folding – a perspective from simple exact models. *Protein Sci.* **4**, 561-602.
16. Kauzman, W. (1959). Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* **14**, 1-63.
17. Friedman, R.A. & Honig, B. (1995). A free energy analysis of nucleic acid base stacking in aqueous solution. *Biophys. J.* **69**, 1528-1535.
18. Ramachandran, G.N. & Saisekharan, V. (1968). Conformation of polypeptides and proteins. *Adv. Protein Chem.* **23**, 283-437.
19. Jack, A., Ladner, J.E. & Klug, A. (1976). 'Non-rigid' nucleotides in tRNA: a new correlation in the conformation of a ribose. *Nature* **261**, 250-251.
20. Holbrook, S.R., Sussman, J.L., Warrant, R.W. & Kim, S.H. (1978). Crystal structure of yeast phenylalanine transfer RNA. II. Structural features and functional implications. *J. Mol. Biol.* **123**, 631-660.
21. Saenger, W. (1984). *Principles of Nucleic Acid Structure.* Springer-Verlag, New York.
22. Freier, S.M., *et al.*, & Turner, D.H. (1986). Improved free energy parameters for predictions of RNA duplex stability. *Proc. Natl Acad. Sci. USA* **83**, 9373-9377.
23. Holbrook, S.R., Cheong, C. & Tinoco, I. Jr. (1991). Crystal structure of an RNA double helix incorporating a track of non-Watson–Crick base pairs. *Nature* **353**, 579-581.
24. Betzel, C., *et al.*, & Erdmann, V.A. (1994). Crystal structure of domain A of *Thermus flavis* 5s rRNA and the contribution of water molecules to its structure. *FEBS Lett.* **351**, 159-164.
25. Cruse, W.B., Saludjian, P., Biala, E., Strazewski, P., Prange, T. & Kennard, O. (1994). Structure of a mispaired RNA double helix at 1.6-Å resolution and implications for the prediction of RNA secondary structure. *Proc. Natl Acad. Sci. USA* **91**, 4160-4164.
26. Egli, M., Portmann, S. & Usman, N. (1996). RNA hydration: a detailed look. *Biochemistry* **35**, 8489-8994.
27. Robertus, J.D., *et al.*, & Klug, A. (1974). Structure of yeast phenylalanine tRNA at 3 Å resolution. *Nature* **250**, 546-551.
28. Schevitz, R.W., Podjarny, A.D., Krishnamachari, N., Hughes, J.J. & Sigler, P.B. (1979). Crystal structure of a eukaryotic initiator tRNA. *Nature* **278**, 188-190.
29. Moras, D., *et al.*, & Giegé, R. (1980). Crystal structure of yeast $tRNA^{Asp}$. *Nature* **288**, 669-674.
30. Woo, N.H., Roe, B.A. & Rich, A. (1980). Three-dimensional structure of *Escherichia coli* initiator $tRNA^{fMet}$. *Nature* **286**, 346-351.
31. Rich, A., Quigley, G.J. & Wang, A.H.J. (1979). Conformational flexibility of the polynucleotide chain. In *Stereodynamics of Molecular Systems*, (Sarma, R.H., ed.), pp. 315-330, Pergammon Press, New York.
32. Costa, M. & Michel, F. (1995). Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.* **14**, 1276-1285.
33. Woese, C.R., Winker, S. & Gutell, R.R. (1990). Architecture of ribosomal RNA: constraints on the sequence of 'tetra-loops'. *Proc. Natl Acad. Sci. USA* **87**, 8467-8471.
34. Heus, H.A. & Pardi, A. (1991). Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science* **253**, 191-194.
35. Cate, J.H., *et al.*, & Doudna, J.A. (1996). RNA tertiary structure mediation by adenosine platforms. *Science* **273**, 1696-1699.
36. Cole, P.E., Yang, S.K. & Crothers, D.M. (1972). Conformational changes of transfer ribonucleic acid. Equilibrium phase diagrams. *Biochemistry* **11**, 4358-4368.
37. Crothers, D.M., Cole, P.E., Hilbers, C.W. & Shulman, R.G. (1974). The molecular mechanism of thermal unfolding of *Escherichia coli* formylmethionine transfer RNA. *J. Mol. Biol.* **87**, 63-88.
38. Celander, D.W. & Cech, T.R. (1991). Visualizing the higher order folding of a catalytic RNA molecule. *Science* **251**, 401-407.
39. Liang, L.G. & Draper, D.E. (1994). Thermodynamics of RNA folding in a conserved ribosomal RNA domain. *J. Mol. Biol.* **237**, 560-576.
40. Zarrinkar, P.P. & Williamson, J.R. (1994). Kinetic intermediates in RNA folding. *Science* **265**, 918-924.
41. Downs, W.D. & Cech, T.R. (1996). Kinetic pathway for folding of the *Tetrahymena* ribozyme revealed by three UV-inducible crosslinks. *RNA* **2**, 718-732.
42. Banerjee, A.R., Jaeger, J.A. & Turner, D.H. (1993). Thermal unfolding of a group I ribozyme: the low-temperature transition is primarily disruption of tertiary structure. *Biochemistry* **32**, 153-163.
43. Laggerbauer, B., Murphy, F.L. & Cech, T.R. (1994). Two major tertiary folding transitions of the *Tetrahymena* catalytic RNA. *EMBO J.* **13**, 2669-2676.
44. Doherty, E.A. &. Doudna, J.A. (1997). The P4-P6 domain directs higher order folding of the *Tetrahymena* ribozyme core. *Biochemistry* **36**, 3159-3169.
45. Stein, A. & Crothers, D.M. (1976). Conformational changes of transfer RNA. The role of magnesium(II). *Biochemistry* **15**, 160-168.
46. Bassi, G.S., Mollegaard, N.-E., Murchie, A.I.H., von Kitzing, E. & Lilley, D.M.J. (1995). Ionic interactions and the global conformations of the hammerhead ribozyme. *Nat. Struct. Biol.* **2**, 45-55.
47. Christian, E.L. & Yarus, M. (1992). Analysis of the role of phosphate oxygens in the group I intron from *Tetrahymena*. *J. Mol. Biol.* **228**, 743-758.
48. Pyle, A.M. (1993). Ribozymes: a distinct class of metalloenzymes. *Science* **261**, 709-714.
49. Latham, J.A. & Cech, T.R. (1989). Defining the inside and outside of a catalytic RNA molecule. *Science* **245**, 276-282.
50. Murphy, F.L. & Cech, T.R. (1993). An independently folding domain of RNA tertiary structure within the *Tetrahymena* ribozyme. *Biochemistry* **32**, 5291-5300.
51. Wang, J.-F. & Cech, T.R. (1994). Metal ion dependence of active-site structure of the *Tetrahymena* ribozyme revealed by site-specific photocross-linking. *J. Am. Chem. Soc.* **116**, 4178-4182.
52. Sclavi, B., Woodson, S., Sullivan, M., Chance, M.R. & Brenowitz, M. (1997). Time-resolved synchrotron X-ray "footprinting", a new approach to the study of nucleic acid structure and function: application to protein–DNA interactions and RNA folding. *J. Mol. Biol.* **266**, 144-159.
53. Cate, J.H., Hanna, R.L. & Doudna, J.A. (1997). A magnesium ion core at the heart of a ribozyme domain. *Nat. Struct. Biol.* **4**, 553-558.