Acc. Chem. Res. 1992, 25, 201–207

The RNA-Folding Problem

DAVID E. DRAPER

Department of Chemistry, Johns Hopkins University, Baltimore, Maryland 21218

Received November 4, 1991

It is a central tenet of molecular biology that structure determines function, and indeed, spectacular insights into functions of specific proteins and DNA sequences have been based on structural studies of these molecules. The structures of a third class of biological macromolecules, RNAs, have not been studied nearly as intensively, with the consequence that details of their functions are poorly understood. This situation is changing as genetic and biochemical studies are bringing the in vivo roles of RNAs into sharper focus, and systematic methods for approaching RNA structure are developed.

RNAs are intimately involved in the process of gene expression in cells, both as informational molecules and as part of the decoding machinery. By the early 1960s three kinds of RNAs had been discerned: messenger RNAs are DNA copies carrying the sequence specifying a gene; transfer RNAs are the "adaptors" which actually read the genetic code off the messenger and substitute the correct amino acid in the growing peptide chain; and the ribosome is an ~3-MDa (megadalton) complex of RNA and protein which facilitates the interaction of transfer and messenger RNAs while catalyzing peptide bond formation. Of these RNA classes, only the structure of transfer RNA is understood in any detail, based on single-crystal X-ray diffraction results[1,2] and extensive physical studies of its folding.[3] Messenger and ribosomal RNAs may reach thousands of nucleotides (in contrast to 75–80 nucleotides for transfer RNA); this has made determination of their structures a formidable experimental problem and is in part responsible for their neglect. The detailed structures of these larger RNAs probably were not seen as a pressing problem either: messengers were viewed as largely passive molecules, while ribosomal RNAs were thought to serve only as the scaffolding for the 50–80 proteins which were presumed to endow the ribosome with enzymatic properties.

About a decade ago evidence started to accumulate in favor of active roles for messenger and ribosomal RNAs in gene expression. Direct involvement of ribosomal RNAs in several essential ribosome functions was discovered,[4] while specific roles for ribosomal proteins remain elusive. In defiance of the classical formulation that gene expression is regulated by repressor proteins binding specific DNA sequences, many important regulatory pathways were found in which messenger RNA structures modulate translation and bind repressor (or activator) proteins.[5] Finally, the discoveries of RNAs with catalytic activities provided convincing proof that RNAs are functionally sophisticated.[6]

How RNAs fold in three dimensions to achieve their specific biological functions is therefore an important problem and is analogous to the protein-folding problem which has occupied chemists for so long. With both proteins and RNAs, folding can be conceptually divided into the formation of secondary structure, followed by the addition of tertiary interactions. RNA secondary structure is defined as segments of contiguous Watson–Crick (A-U, C-G) or wobble (G-U) base pairs (see illustration of transfer RNA in Figure 1). Tertiary interactions are the additional hydrogen-bonding and base-stacking interactions which stabilize a unique three-dimensional structure of the molecule. Some of the same methodologies used to investigate protein folding have been applied to RNAs: there have been attempts to accumulate enough thermodynamic information to predict RNA secondary structure by computer searches for the free energy minimum,[7] and chemical reagents which distinguish between the solvent accessibilities of paired and unpaired bases have been used.[8] While these methods have had some success, they generally do not unambiguously predict a single structure. The obvious method of choice, X-ray crystallography, has not been practical as RNAs have shown extreme reluctance to give crystals diffracting to high resolution. The current challenge is to devise systematic approaches for deducing RNA folding at a sufficient level of detail to be useful in rationalizing functional properties.

Rather than deal with the structure of an entire messenger or ribosomal RNA, we have first tried to define interesting fragments of these RNAs, much as one might isolate a protein domain or subunit for initial study. For instance, certain limited regions of the ribosomal RNAs are very highly conserved, and some ribosomal proteins also protect specific ribosomal RNA structures from digestion with nucleases. Thus it seemed feasible to isolate ribosomal RNA domains which fold independently of the rest of the RNA and are small enough to allow one to deduce useful structural details from physical studies. This approach also seemed feasible with messenger RNAs, where regulatory

David Draper, born in California in 1950, did his undergraduate work in biochemistry at the University of California, Berkeley, and the University of Sussex, England. He received a Ph.D. degree in chemistry from the University of Oregon in 1977 and was then a postdoctoral fellow in the Molecular, Cellular, and Development Biology Department of the University of Colorado until 1980. He is now a professor in the Department of Chemistry, Johns Hopkins University.

(1) Quigley, G. J.; Wang, A. H. J.; Seeman, N. C.; Suddath, F. L.; Rich, A.; Sussman, J. L.; Kim, S. H. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 4866–4870.
(2) Ladner, J. E.; Jack, A.; Robertus, J. D.; Brown, R. S.; Rhodes, D.; Clark, B. F. C.; Klug, A. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 4414–4418.
(3) Cole, P. E.; Yang, S. K.; Crothers, D. M. *Biochemistry* **1972**, *11*, 4358–4368.
(4) Dahlberg, A. E. *Cell* **1989**, *57*, 525–529.
(5) Ilan, J., Ed. *Translational Regulation of Gene Expression*; Plenum Press: New York, 1987.
(6) Cech, T. R.; Bass, B. L. *Annu. Rev. Biochem.* **1986**, *55*, 599–630.
(7) Jaeger, J. A.; Turner, D. H.; Zuker, M. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 7706–7710.
(8) Ehresmann, C.; Baudin, F.; Mougel, M.; Romby, P.; Ebel, J.-P.; Ehresmann, B. *Nucleic Acids Res.* **1987**, *15*, 9109–9128.

0001-4842/92/0125-0201$03.00/0 © 1992 American Chemical Society

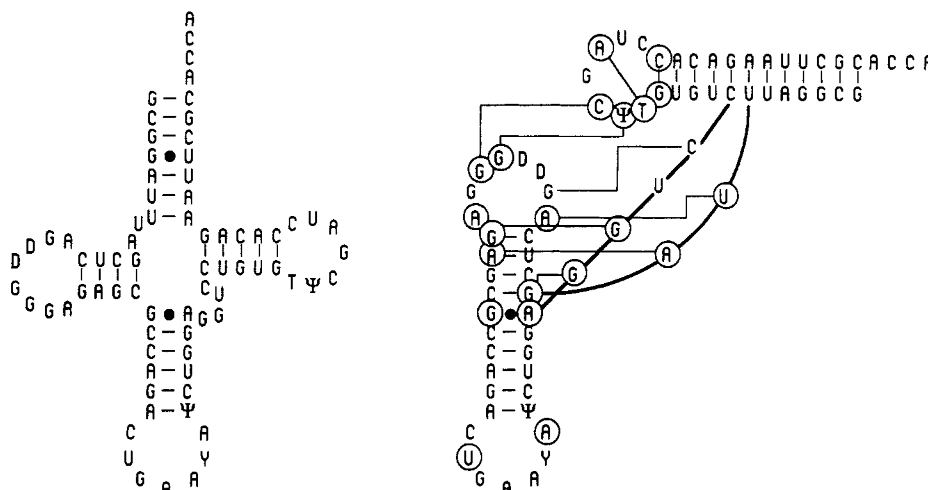


Figure 1. Levels of RNA structure. Schematic representations of the secondary (left) and tertiary (right) structures of yeast tRNA^{phe} are shown. Thin lines indicate hydrogen bonds; thick lines connect adjacent nucleotides. Circled bases are invariant among all tRNA sequences. Besides the modified bases shown (T, D, Y, and Ψ), tRNAs have a number of methylated bases which are not indicated; the modifications are not essential for function and are rarely found in other RNAs.

regions are interspersed with relatively unstructured coding domains. Our strategy has been to look for RNA structures that are specifically recognized by a protein; the protein binding affinity then serves as an assay for native folding of the RNA fragment.

Physical studies of most RNAs have become practical only since the availability of a method for synthesizing defined RNAs in quantity. Phage RNA polymerases of high efficiency and high purity have made it possible to prepare enzymatically tens of milligrams of RNAs from DNA templates.[9,10] All of the RNAs described here were made in this way.

Isolation of Functionally Interesting RNA Fragments

We have been working on two RNA fragments which are of moderate size (60–110 nucleotides) and have distinctive tertiary structures specifically recognized by proteins. One is derived from the *Escherichia coli* α messenger RNA, which displays classical feedback regulation at the translational level: a repressor protein encoded by the mRNA binds to a structure in the vicinity of the ribosome initiation site and prevents translation. In this case, the repressor is a ribosomal protein called S4, which also binds directly to a ribosomal RNA. The S4 binding affinity gave us a simple in vitro assay for native folding of RNA fragments we synthesized, and RNA fragments containing a 114-nucleotide mRNA sequence retained the ability to bind repressor with the same affinity as a much larger piece of RNA (see Figure 2).[11,12] More elaborate in vivo assays revealed S4 repression of messenger RNAs containing these fragment sequences[13,14] and provided convincing evidence that only a limited RNA domain is required for the biological function of translational repression.

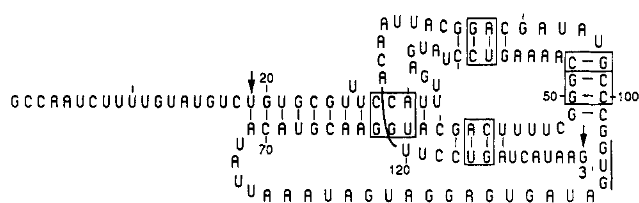


Figure 2. Primary and secondary structure of the *E. coli* α messenger RNA ribosome binding site and regulatory region. Numbering of the bases is from the first nucleotide of the in vivo transcript; tick marks are located every 10 nucleotides. Arrows indicate the most extreme 5′ and 3′ termini of RNA fragments which bind S4 as tightly as much longer transcripts in vitro. Underlining indicates signals recognized by ribosomes in initiating translation; GUG codes for the first amino acid of the protein. Boxes indicate positions where sets of compensatory base changes indicated that base pairing is required for S4 recognition.

The other RNA fragment is derived from the large subunit of the ribosome. GTP is hydrolyzed in response to binding of a protein factor at two points in the ribosome cycle, once upon binding of aminoacyl-tRNA (catalyzed by elongation factor Tu), and once upon translocation to the next codon (catalyzed by elongation factor G). An "uncoupled" GTPase activity can be assayed with the large 50S subunit and one of the two elongation factors. This uncoupled activity is affected both by the ribosomal protein L11 and by the thiostrepton family of peptide antibiotics, and both L11 and the antibiotics were known to bind to a limited region of the 23S ribosomal RNA.[15] We prepared a series of ribosomal RNA fragments and found that relatively small RNAs are able to bind these ligands with the same affinities as intact ribosomal RNA (Figure 3); binding affinity again serves as an assay for correct folding of the RNA structure.[16-18]

Determination of RNA Secondary Structures

The first step in our studies of these RNAs was to find the secondary structure or helical-base-pairing scheme. The first RNA secondary structure ever de-

(9) Milligan, J. F.; Groebe, D. R.; Witherell, G. W.; Uhlenbeck, O. C. *Nucleic Acids Res.* **1987**, *15*, 8783–8798.
(10) Draper, D. E.; White, S. A.; Kean, J. M. *Methods Enzymol.* **1988**, *164*, 221–237.
(11) Deckman, I. C.; Draper, D. E. *Biochemistry* **1985**, *24*, 7860–7865.
(12) Deckman, I. C.; Draper, D. E.; Thomas, M. S. *J. Mol. Biol.* **1987**, *196*, 313–322.
(13) Thomas, M.; Bedwell, D. M.; Nomura, M. *J. Mol. Biol.* **1987**, *196*, 333–345.
(14) Tang, C. K.; Draper, D. E. *Biochemistry* **1990**, *29*, 4434–4439.
(15) Schmidt, F. J.; Thompson, J.; Lee, K.; Dijk, J.; Cundliffe, E. *J. Biol. Chem.* **1981**, *256*, 12301–12305.
(16) Ryan, P. C.; Draper, D. E. *Biochemistry* **1989**, *28*, 9949–9956.
(17) Ryan, P. C.; Lu, M.; Draper, D. E. *J. Mol. Biol.* **1991**, *221*, 1257–1268.
(18) Laing, L.; Wang, Y. X.; Draper, D. Unpublished observations.

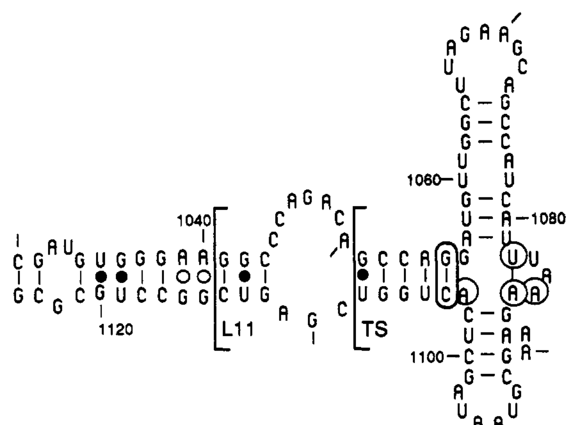


Figure 3. Primary and secondary structure of the "GTPase center" from the large subunit ribosomal RNA. Numbering of the bases is from the 5′ nucleotide of the *E. coli* 23S ribosomal RNA; tick marks are located every 10 nucleotides. Brackets indicate the termini of the smallest RNA fragments found to bind either protein L11 or the antibiotic thiostrepton (TS) with the same affinity as much larger RNA fragments. Positions at which mutations disrupt both L11 and thiostrepton binding are circled. An interaction between positions 1082 and 1086, deduced by compensatory base mutations, is shown.

duced, the cloverleaf structure of tRNAs, was arrived at after only two different tRNA sequences were available.[19] It was known that all tRNAs must have generally the same structure, since they all function in the same ribosomal apparatus; thus it was expected that a base-pairing scheme which could be drawn the same way for all tRNA sequences must be the correct one. The same approach is still the most reliable method for finding the secondary structure of larger RNAs: first phylogenetically related RNAs from a variety of organisms are sequenced to obtain a set of homologous, but not identical, RNA sequences. Then "compensatory changes", pairs of sequence changes substituting one base pair for another in a potential helix, are taken as evidence for the existence of that helix. As hundreds of ribosomal RNA sequences are now available, the secondary structures of even these very large RNAs have been established in considerable detail.[20] This is a rare case in which reliable structural information can be had by "sequence gazing" alone and contrasts with the ambiguous information available from comparisons of homologous protein sequences.

The S4 messenger RNA is not conserved among a sufficiently diverse set of organisms to use this phylogenetic approach, so we had to generate our own sequence variants. We first listed all thermodynamically plausible helices of four or more base pairs; fewer than a third of these 31 helices could coexist at any one time in the RNA. Some simple "structure-mapping" experiments eliminated a number of helices from consideration.[21] In this approach, gel methods are used to rapidly screen for nucleotides sensitive to nucleases with specificity for either single-stranded, unstructured RNA or structured, approximately helical RNA. Twenty-two of the potential helices were inconsistent with the nuclease digestion patterns, though several different mutually exclusive secondary structures could still be drawn with the remaining helices. Finally, we used site-directed mutagenesis to make sets of compensatory base changes in each of the remaining potential helices.[22] First two RNAs were constructed with mutations in either of the two complementary strands of a putative helix. In most cases we found weakened S4 binding affinities. Next, a double mutant was made which had complementary alterations in both strands of the helix and which restored the potential for base pairing. We deduced that the helix must exist if the S4 binding constant returned to full strength. This gave us a methodical way to arrive at the biologically functional secondary structure, with the level of detail limited only by our patience in constructing site-directed mutations.

Figure 2 shows all the positions at which compensatory changes provided positive evidence for secondary structure. The molecule has an unusual "pseudoknot" secondary structure, i.e., sequences in the 3′ half of the RNA pair at several points with the loop of a hairpin in the 5′ half. (If the pairings were full helical turns, the 3′ terminus would be pulled through the hairpin loop; joining of the 5′ and 3′ ends would then create a topological knot. Such a structure has not been observed in RNA, and the term "pseudoknot" was coined to describe loop pairings with less than a full turn of helix.[23]) This structure shows the value of systematically testing the possible secondary structures, since we did not seriously consider this bizarre folding until forced to do so by the data.

There are several caveats in the interpretation of compensatory base change data. First, it is important to understand that the secondary structure deduced by this method is not necessarily the minimum free energy structure, but the functional structure for the particular assay used (protein recognition, in this case). This is an advantage, in that a biologically relevant structure is determined, but the deduced structure may not be found in the free RNA. Second, a protein may recognize a specific sequence in a particular helix, and compensatory changes at these positions will give negative results. A possible instance of this is the potential pair C52-G98: substitution with G52-C98 severely disrupts protein binding, even though it would appear that these bases should pair and stack onto the neighboring confirmed helix. (The two bases could also be involved in some critical tertiary interactions.) Lastly, helices can be involved in additional interactions which may give unusual patterns of compensatory changes. For example, the mutation CU41 → GA has no effect on protein binding even though it compensates for the effect of the AG110 → UC mutation, which does have a substantial effect alone. This asymmetry suggests that there are additional interactions taking place at this position which allow the RNA to tolerate disruption of one strand of the helix but not the other.

With the GTPase center RNA we had the luxury of a substantial data base of sequences from diverse organisms which determined the main features of the secondary structure. (The *E. coli* L11 protein has been shown by others to recognize the large subunit ribosomal RNA from a wide variety of organisms; thus it is

(19) Madison, J. T.; Everett, G. A.; Kung, H. *Science* **1966**, *153*, 531–534.
(20) Gutell, R. R.; Woese, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 663–667.
(21) Deckman, I. C.; Draper, D. E. *J. Mol. Biol.* **1987**, *196*, 323–332.
(22) Tang, C. K.; Draper, D. E. *Cell* **1989**, *57*, 531–536.
(23) Studnicka, G. M.; Rahn, G. M.; Cummings, I. W.; Salser, W. A. *Nucleic Acids Res.* **1978**, *5*, 3365–3387.

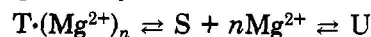
reasonable to assume that all of the related GTPase center sequences have about the same structure.) This region is, however, very highly conserved, and a number of potential base pairs are invariant among the available sequences; therefore the data base provides no evidence for or against their pairing. In addition, several positions in the *E. coli* sequence look like they should extend conserved helical segments but are not conserved as Watson–Crick pairs. This may mean either that noncanonical pairings may sometimes substitute for Watson–Crick pairs or that in some cases an entirely different, nonhelical structure should be considered. An example is nucleotides U1065 and A1073, which look like they should form a Watson–Crick pair stacked onto the adjacent helix. Among other organisms, only C1065-A1073 or A1065-C1073 are found (besides U1065-A1073). Curiously, we found that neither of the single base mutations U1065 → A or A1073 → U had any effect on either L11 protein or thiostrepton binding, but the double mutation A1065-U1073 weakened the binding of both ligands.[17] Perhaps the double mutant does form a standard Watson–Crick pair and alters the structure of the molecule, while the other base juxtapositions hydrogen bond in a noncanonical structure.

Existence of RNA Tertiary Structure

In contrast to the simple rules for forming RNA secondary structures, there is a large variety of possible tertiary interactions. In transfer RNA there are many noncanonical hydrogen bonds between the bases, sugars, and phosphates and a wide range of backbone conformations including unusually sharp bends. There is no way to know whether tRNA represents the gamut of tertiary structures, or whether a yet larger range of hydrogen-bonding patterns and backbone conformations are common in other RNAs. Even some simple four-base RNA hairpin loops have an unexpected set of hydrogen bonds which confer unusual stability on the structure,[24,25] so it is likely that many new kinds of tertiary structures remain to be discovered.

Even if many kinds of tertiary interactions are likely, it does not follow that a stable tertiary folding is a ubiquitous feature of RNAs. One can imagine that the intricate set of tertiary interactions found in transfer RNA is an exception and that the RNA fragments described here assume specific three-dimensional structures only when bound by proteins. In the case of the ribosomal RNA fragment, the function of the protein might be to order an otherwise floppy or unstable structure. The messenger RNA must be unfolded to allow translation of the sequences following G95, and it is possible that the pseudoknot is not a stable structure at all in the absence of the repressor protein. Thus it is important to ask whether these RNAs assume stable secondary and tertiary foldings by themselves or only in association with proteins.

The stability of RNA structures in this size range is easily determined since denaturation of the molecule produces a significant increase in the absorbance. In addition, the effects of Mg^{2+} ions on the RNA stability provide a way to assay for the presence of tertiary structure. Studies of transfer RNA showed that Mg^{2+} ions preferentially stabilize tertiary over secondary interactions.[26] Unfolding of the native tRNA structure proceeds sequentially:



$$T\cdot(Mg^{2+})_n \rightleftharpoons S + nMg^{2+} \rightleftharpoons U$$

where T represents the native RNA with intact tertiary structure, S is a form with secondary structure but no tertiary structure, and U is the completely unfolded form. (The secondary structure may of course denature in several steps; only one structure is shown for simplicity.) As indicated, formation of the tertiary structure is associated with uptake of Mg^{2+} ions. The rationale for this observation is that tertiary folds generally stabilize a more compact RNA structure which brings phosphates into close proximity. Multivalent ions are more effective than monovalent ions at reducing the resulting electrostatic repulsion, especially in the tight quarters of a compact structure. (Since a variety of di- and trivalent ions stabilize the tRNA structure, the effects are probably purely electrostatic and not dependent on any specific coordination of Mg^{2+} with the phosphates.) Charging of a transfer RNA with an amino acid by its cognate synthetase is highly sensitive to the divalent cation concentration, as one might expect if the synthetase recognizes the native tertiary structure of the RNA.[27]

Very similar behavior is seen with our ribosomal RNA fragment. First, binding of either L11 protein or the antibiotics depends strongly on the Mg^{2+} ion concentration (Figure 4A); one or two Mg^{2+} ions must bind the RNA with an affinity of $\sim 3\ mM^{-1}$.[16,17] Mg^{2+} binding to a simple RNA helix under the same conditions would be 5–10-fold weaker, so stabilization of an RNA structure with a higher charge density than a helix is suggested. Increasing Mg^{2+} concentration also preferentially stabilizes a major melting transition (Figure 4B); the concentration range in which this happens is the same as that promoting protein binding. We therefore conclude that the RNA has a Mg^{2+}-stabilized tertiary structure which is recognized by the protein and antibiotic ligands.

Melting experiments with RNA fragments containing all or part of the messenger pseudoknot structure demonstrate that the pseudoknot is in fact a stable structure under physiological conditions, with a melting temperature of $\sim 65\ °C$.[28] Protein binding is not detectable in the absence of Mg^{2+}, and from the dependence of the protein binding affinity on Mg^{2+} ion concentration, we have estimated that approximately five Mg^{2+} ions bind the messenger fragment with affinities of $0.6\ mM^{-1}$.[12] The messenger fragment is therefore similar to the ribosomal fragment in having a Mg^{2+}-stabilized tertiary structure that is recognized by the protein.

Detection of a Specific Tertiary Interaction

In principle, the same kinds of compensatory changes used to establish an RNA secondary structure can also detect tertiary hydrogen bonding between bases. There are of course many possible combinations of mutants to test, and tertiary pairings may not follow Watson–Crick rules, so a way to narrow down the possible interactions is needed. A set of phylogenetically related

(24) Cheong, C.; Varani, G.; Tinoco, I., Jr. *Nature* **1990**, *346*, 680–682.
(25) Heus, H. A.; Pardi, R. *Science* **1991**, *253*, 191–194.
(26) Stein, A.; Crothers, D. M. *Biochemistry* **1976**, *15*, 160–167.
(27) Karpel, R. L.; Miller, N. S.; Lesk, A. M.; Fresco, J. R. *J. Mol. Biol.* **1975**, *97*, 519–532.
(28) Gluick, T. C.; Spedding, G.; Draper, D. E. Manuscript in preparation.

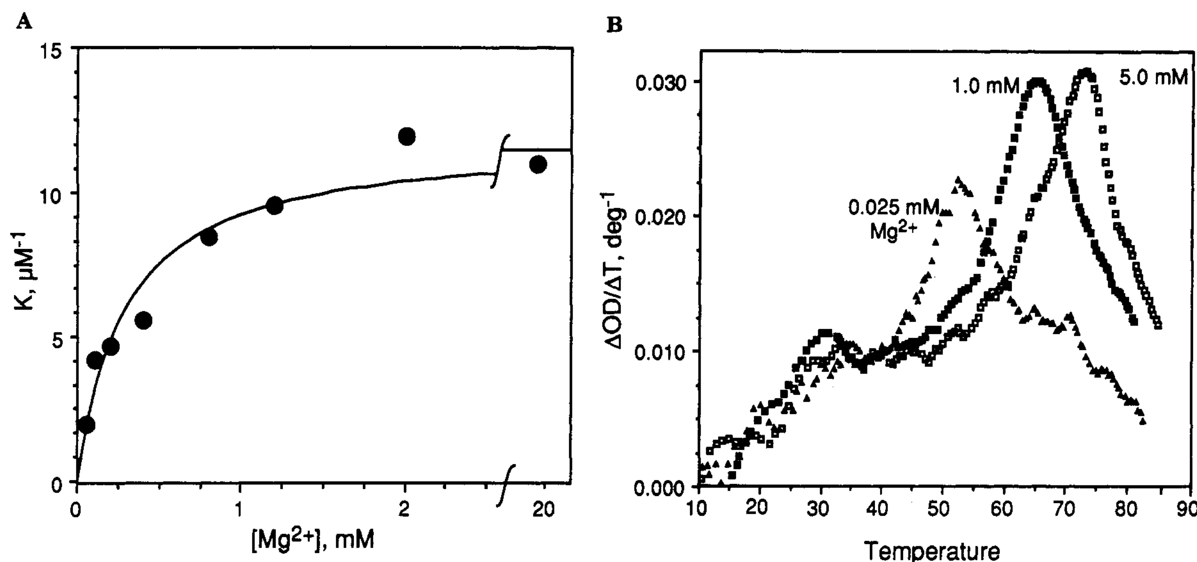


Figure 4. Effects of Mg^{2+} ions on the ribosomal RNA fragment folding. (A) The L11 protein–RNA fragment binding constant increases with Mg^{2+} ion concentration. The curve is calculated with the assumption that one ion binding the RNA with an affinity of 3.5 μM^{-1} is a prerequisite for protein binding. (B) First derivative optical melting curves (260 nm) of the 1029–1122 ribosomal RNA fragment. The solution contained 0.1 M KCl in addition to the indicated Mg^{2+} concentrations.

sequences can suggest unusual compensatory changes, though the data set must be very large because loops generally evolve much more slowly than helical stems. Tertiary interactions in the large ribosomal RNAs have been detected in this way.[20] Systematic sets of mutations can also narrow the search considerably. In the GTPase center RNA, we constructed a set of 44 mutations which, taken together, change every base within the main recognition region (1051–1108) to its Watson–Crick complement. Compensatory mutations were made where necessary to preserve the known secondary structure. All of these sequence variants were tested for both L11 and thiostrepton binding, and only a small number of mutants had a major effect on the binding of *both* ligands (Figure 3). The two ligands can both bind simultaneously to the RNA and, therefore, must contact different sets of RNA features. Mutations which affect both ligands are probably affecting the overall structure, rather than removing an RNA contact with the ligand, and are good candidates for involvement in tertiary interactions.

Most of the positions identified as potential participants in the tertiary structure are conserved among all of the available ribosomal RNA sequences, as one might expect for residues involved in crucial interactions. (Many of the bases in the tRNA tertiary interactions are invariant among all tRNAs; see Figure 1.) Only two of the candidate positions are not invariant: all prokaryotes have U1082 and A1086, while all eukaryotes have C1082 and G1086 and archaebacteria have either the prokaryotic or eukaryotic combination of bases at these positions. This pattern of conservation looks like a compensatory base change, but the very high degree of conservation among kingdoms makes this a shaky conclusion. We therefore tested this by constructing the appropriate single and double mutations, including one double mutation which would substitute A1082-U1086, a potential pairing not detected in nature. The results were unequivocal: both L11 and thiostrepton require complementary bases at the two positions.[29]

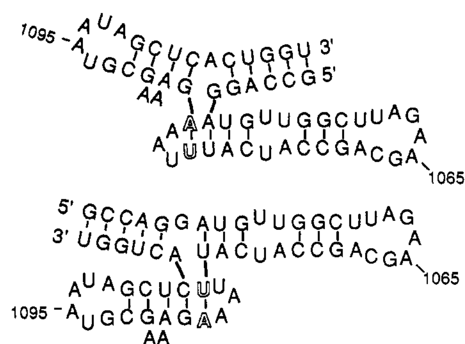


Figure 5. Two possible stacking schemes for the ribosomal RNA fragment. The U1082-A1086 pairing has been stacked onto either A1057-U1081 (upper) or G1087-C1102 (lower). Thick lines connect adjacent nucleotides.

This establishes one tertiary connection which further constrains the three-dimensional folding of the RNA. A selection/amplification experiment, which generates all possible combinations of sequence variants at a set of nucleotides and then enzymatically amplifies only those sequences which are selected by ligand binding,[30] is in progress and may reveal other compensatory changes among the remaining candidates for tertiary interactions.

A caution is again in order here. The pattern of allowed 1082-1086 substitutions follows the Watson–Crick pairing rules, but so do other kinds of hydrogen-bonding schemes. Hoogsteen hydrogen bonding, for instance, takes advantage of the purine N7 positions to also pair A-U and C-G. In transfer RNA, a pairing conserved as A-U or G-C is actually a noncanonical pairing between parallel RNA strands. Although we are confident that bases 1082 and 1086 interact, these precedents force us to remain undecided about the precise bonding scheme.

The presence of the 1082-1086 interaction narrows down the likely tertiary structures to two basic models. When one is building RNA models, a principle to keep in mind is that bases have a very strong tendency to stack (similar to the protein-folding principle that hy-

(29) Ryan, P. C.; Lu, M.; Draper, D. E. *Proc. Natl. Acad. Sci. U.S.A.* 1991, *88*, 6308–6312.

(30) Tuerk, C.; Gold, L. *Science* 1990, *249*, 505–510.

drophobic side chains are buried, though the origin of the force is different). A 1082-1086 pair is therefore almost certainly stacked with some other bases, and the obvious candidates are the adjacent A1057-U1081 and G1087-C1102 pairs. These two stacking possibilities lead to two different arrangements of the three main helical segments of the RNA (Figure 5). Additional tertiary interactions involving the other bases that are candidates for tertiary interactions are possible in each case.

RNA Conformational Equilibria

An additional complication in thinking about tertiary structures is the probability that many RNAs are rather flexible molecules with multiple conformations; thus we may need to determine not one but several related and functionally important foldings for any single RNA molecule. There is evidence in favor of functionally relevant conformational "switches" for both the RNAs discussed here. In the case of the messenger RNA, some mutations which have no effect on repressor binding affinity are still not regulated by the repressor in vivo.[14] The best way to explain this is by supposing that the repressor and ribosome binding sites on the mRNA are distinct and that the protein binding transmits a conformational change to the ribosome binding site. Mutations can disrupt this communication between the two sites and have only small effects on protein affinity while abolishing the translational regulation. Enzymologists will recognize this as a classical allosteric mechanism, with the repressor acting as an effector of ribosome initiation.[31] The proposal predicts that the RNA can switch between active and inactive conformations. Recently we have detected a transition of the pseudoknotted RNA between two conformations, only one of which is able to form translational initiation complexes. Mutations that enhance or depress translation rates in vivo tend to shift the transition toward the active or inactive conformations, respectively.[28] It will be extremely interesting to find out the structural differences between the two conformations.

The ribosomal GTPase domain is also thought to have multiple functionally important conformations. Briefly, the observation is that the different related antibiotics which bind this RNA have very different effects on the uncoupled ribosomal GTPase activity: e.g., thiostrepton inhibits the activity while micrococcin stimulates it. Cundliffe has proposed that the GTPase center RNA normally cycles through several conformations as the ribosome goes through the steps associated with peptide-bond formation and mRNA translocation, and that the antibiotics exert their effects by trapping the RNA in one of these conformations.[32] We have recently detected a substantial rearrangement of the RNA fragment structure taking place at ~30 °C by changes in either UV absorbance or the imino proton NMR spectrum.[18] Whether this is related to conformational changes taking place in the intact ribosome is a matter for further experiment, but the observation at least underscores how frequently RNAs are found to adopt alternate structures. Obviously the tertiary structures of alternate RNA conformations, as well as the thermodynamics of their interconversion, are important questions to be resolved before the functions of these RNAs will be understood.

Concluding Remarks

It is now fairly routine for an RNA secondary structure to be determined using the described phylogenetic and site-directed mutagenesis methods, and initial progress in studying the folding of an RNA is usually rapid. Although the secondary structure for an RNA cannot be as informative as a complete three-dimensional structure, it is extremely useful information which can provide some insight into function as well as direction for further physical studies. For instance, the way the ribosome binding site (~U75-U105 in Figure 2) is incorporated into the unexpected pseudoknot folding of the messenger RNA must be considered in thinking about mechanisms for translational repression. In the ribosomal RNA fragment secondary structure, the clustering of nucleotides important for both protein and antibiotic binding in one section of the secondary structure suggests that further studies of this RNA structure should focus on the conformation of the indicated helix junction.

The rate-limiting step in RNA-folding studies is now the determination of tertiary interactions. The methods that have provided a wealth of detailed information about protein and DNA structure, single-crystal X-ray diffraction and high-field NMR, have been difficult to apply; consequently there is a dearth of high-resolution RNA structures. This situation should improve in the near future. As seen with our ribosomal RNA fragment, compensatory base changes can suggest tertiary interactions, and computer studies show that relatively few such interactions need to be specified before the overall three-dimensional folding is strongly constrained.[33] Compensatory changes together with molecular modeling have already led to a surprisingly detailed proposal for the structure of the active site of the group I intron self-splicing RNA.[34] Multidimensional NMR methods are currently being developed that will probably allow the complete proton spectra of modest-sized (≲40 nucleotide) RNAs to be assigned,[35] and even in molecules the size of transfer RNA, techniques for extracting useful information from the imino proton region of the spectrum have been available for some time.[36] Thus there is every prospect that some aspects of our ribosomal RNA fragment structure will yield to NMR studies. Finally, it is possible that RNA-protein (or RNA-antibiotic) complexes will crystallize more readily than RNA molecules alone, because complexes may have less conformational heterogeneity; the recent spectacular successes in crystallizing transfer RNAs with their cognate synthetases bodes well.[37,38]

At the present time, details of the tertiary foldings of very few RNAs are known; even the number of RNAs demonstrated to have significant tertiary interactions

(31) Monod, J.; Wyman, J.; Changuex, J.-P. *J. Mol. Biol.* **1965**, *12*, 88-118.
(32) Cundliffe, E. In *Structure, Function, and Genetics of Ribosomes*; Hardesty, B., Kramer, G., Eds.; Springer-Verlag: New York, 1986; pp 586-604.

(33) Hubbard, J. M.; Hearst, J. E. *Biochemistry* **1991**, *30*, 5458-5465.
(34) Michel, F.; Westhof, E. *J. Mol. Biol.* **1990**, *216*, 585-610.
(35) Varani, G.; Tinoco, I. *J. Am. Chem. Soc.* **1991**, *113*, 9349-9354.
(36) Hare, D. R.; Ribeiro, N. S.; Wemmer, D. E.; Reid, B. R. *Biochemistry* **1985**, *24*, 4300-4306.
(37) Rould, M. A.; Perona, J. J.; Söll, D.; Steitz, T. A. *Science* **1989**, *246*, 1135-1142.
(38) Ruff, M.; Krishaswamy, S.; Boeglin, M.; Poterszman, A.; Mitschler, A.; Podjarny, A.; Rees, B.; Thierry, J. C.; Moras, D. *Science* **1991**, *252*, 1682-1689.

in the absence of bound proteins is only a handful. By comparison there is a rapidly expanding catalog of biologically interesting RNAs: besides the transfer, messenger, and ribosomal RNAs already mentioned, there are structures able to catalyze hydrolysis or ligation of other RNAs[39-41] as well as a variety of RNA-protein complexes involved in messenger RNA processing.[42] The number and importance of these RNA structures provide a strong impetus for structural studies. There undoubtedly remain many unexpected discoveries to be made in this area, and the next few years should be very exciting as more RNA structures are studied in detail.

A number of past and present students have contributed to the work described here; I thank particularly Drs. Ingrid Deckman and Patricia Ryan, who initiated the work on the messenger and ribosomal RNA fragments, respectively. The work has been supported by NIH Grant GM29048.

(39) Hampel, A.; Tritz, R.; Hicks, M.; Cruz, P. *Nucleic Acids Res.* **1990**, *18*, 299–304.
(40) Perrotta, A. T.; Been, M. D. *Nature* **1991**, *350*, 434–436.
(41) Cech, T. R. *Annu. Rev. Biochem.* **1990**, *59*, 543–568.
(42) Guthrie, C. *Science* **1991**, *253*, 157–163.