

How RNA Folds

Ignacio Tinoco Jr^{1*} and Carlos Bustamante^{2*}

¹Department of Chemistry
University of California
Berkeley, Berkeley, CA, 94720-
1460, USA

²Departments of Physics, and
Cellular and Molecular Biology
University of California
Berkeley, Berkeley, CA 94720-
1700, USA

We describe the RNA folding problem and contrast it with the much more difficult protein folding problem. RNA has four similar monomer units, whereas proteins have 20 very different residues. The folding of RNA is hierarchical in that secondary structure is much more stable than tertiary folding. In RNA the two levels of folding (secondary and tertiary) can be experimentally separated by the presence or absence of Mg²⁺. Secondary structure can be predicted successfully from experimental thermodynamic data on secondary structure elements: helices, loops, and bulges. Tertiary interactions can then be added without much distortion of the secondary structure. These observations suggest a folding algorithm to predict the structure of an RNA from its sequence. However, to solve the RNA folding problem one needs thermodynamic data on tertiary structure interactions, and identification and characterization of metal-ion binding sites. These data, together with force *versus* extension measurements on single RNA molecules, should provide the information necessary to test and refine the proposed algorithm.

© 1999 Academic Press

Keywords: RNA folding; RNA secondary structure; RNA tertiary structure; atomic force microscopy

*Corresponding authors

Introduction

The known biological functions of RNA continue to grow in number and to expand in scope. RNA has been transformed from a molecule with a minor role in protein synthesis, to an important player in all of molecular biology. Scarcely ten years ago, messenger RNA was thought to be a passive carrier of genetic information from DNA; its conformation random or irrelevant. Ribosomal RNA was believed to be a scaffold for the essential ribosomal proteins, and the transfer RNAs simply dull adapters that placed each amino acid in the location specified by the genetic code. Now we know that the ribosomal proteins mainly provide the scaffold that allows the ribosomal RNAs to catalyze peptide bond formation (Noller *et al.*, 1992), and to perform their other functions; that transfer RNAs and their mimics are involved in replication, intron splicing, and translational regulation (Giegé *et al.*, 1998); that the conformations of messenger RNAs, particularly at the 3'- and 5'-untranslated regions, determine the lifetimes of the RNAs and control the efficiency of translation (see, for example Spicher *et al.*, 1998). Furthermore, it has been shown that stem-loops in the mRNA can

bind to the protein product of the message to regulate its synthesis (McCarthy & Gualerzi, 1990), and that pseudoknots in retroviral mRNAs cause programmed frameshifts that produce the correct ratios of proteins required for viral propagation (Chamorro *et al.*, 1992).

The importance of RNA has increased most rapidly since the discovery of catalytic RNA. The realization that RNA could be an enzyme has led to a multitude of studies of naturally occurring, and artificially evolved and selected ribozymes. It has become increasingly obvious that RNA-RNA, RNA-DNA and RNA-protein interaction are important in many different biological contexts. For example, there is recent evidence for the role of RNAs in memory at nerve synapses (for a review, see Lisman & Fallon, 1999).

A systematic search for non-coding RNAs in the human and other genomes is just beginning (Lowe & Eddy, 1999). There will certainly be many more unexpected RNA functions found.

RNA folding

The previous paragraphs have been written from an admittedly RNA-centric point of view, but they were intended to justify the effort to understand how RNA folds. We want to be able to interpret an RNA sequence in terms of its folded, three-dimensional functional form. Functional is emphasized

E-mail address of the corresponding author:
intinoco@lbl.gov
carlos@alice.berkeley.edu

because we do not necessarily need, or want, to know the coordinate of every atom in the structure to 1 Å resolution. The functional RNA may have regions where the conformations are not relevant. Some sequences may fold into a structure that is rigid and compact; other sequences may be designed to be flexible to maximize interaction with a particular ligand. We want to know the structure well enough to understand the function. This may mean that we need to know one part of the structure at atomic resolution, another part in terms of Watson-Crick base-pairs formed, and the rest of the molecule only approximately. One criterion for the importance of structure can be inferred from its constancy deduced from phylogenetic comparisons (Michel & Westhof, 1990). If base paired regions in an RNA molecule are conserved over many species, it is assumed that they are important in function. If only a single sequence is available for the RNA, secondary structure calculations can predict the base-paired regions. In viral RNAs the most stable calculated helices correlate with regions that are known to have important biological roles, such as internal ribosome entry sites (Palmenberg & Sgro, 1997).

Here we will try to systematize and generalize some of what is known about how RNA folds, and to suggest an algorithm to predict the three-dimensional structure of RNA from a knowledge of its sequence. For a review that describes the earlier work in this field, see Brion & Westhof (1997). Our goal is to encourage others to enter this relatively unpopulated field. If 10% of protein fold researcher switched to RNA, the problem could be solved in one or two years. If the lack of competition in the field is not sufficiently enticing, we hope that by making sweeping statements about RNA folding we can at least goad people into proving us wrong.

RNA structure is hierarchical and its folding is sequential

The protein folding problem, the search for rules to predict the three-dimensional structures of proteins from their primary sequences, has attracted scientists for almost five decades. Structural prediction in protein folding is difficult for two main reasons. First, because it takes 20 different amino acid residues to build a protein molecule. Thus, the number of distinct interactions among these residues is large, depending not only on the nature of the residues (hydrophobic, hydrophilic, polar, etc.), but also on each residue's detailed structure. Second, because the existence of the various secondary structural elements (α -helices, parallel and anti-parallel β -sheets, β -turns, random coils, etc.) is contextual, i.e. these elements form and are stable in the context of the rest of the protein, but may not form when they are isolated in solution. The energies that stabilize secondary structural elements in proteins are comparable to the energies involved in tertiary interactions. Thus, the formation of secondary structure depends on the nature of the tertiary folding contacts, and *vice versa*. This implies that the

contributions of secondary and tertiary interactions to the energetic stability of a protein are, in principle, not separable and, therefore, no simple rules may exist to predict the three-dimensional structure of the protein from its sequence.

By comparison, RNA folding is simpler. Only four nucleotides, each made of a base, a ribose, and a phosphate, are used as building blocks of the structure. The four bases are very much alike; there are two purines and two pyrimidines that differ only in the placement of carbonyl and amino groups and their interactions are either through hydrogen bonding or base stacking. Each ribose is a five-membered ring with only two main conformations. The phosphate groups with their long-range electrostatic interactions might be considered as providing theoretical difficulties not present in protein folding, but they actually simplify the RNA folding problem both theoretically and experimentally. Electrostatic interactions are well understood theoretically, and experimentally the effective charge on the phosphate groups can be controlled by means of the ionic strength of the solvent.

There are only four basic secondary structure elements in RNA (helices, loops, bulges, and junctions). The helices are A-form Watson-Crick duplexes; the loops, bulges and junctions are all non-Watson-Crick regions terminated by one or more helices. Finally, because the energies involved in the formation of secondary structure are larger than those involved in tertiary interactions, secondary structural elements can exist and be stable by themselves. Thus, the energetic contributions of the secondary and tertiary structural elements are separable, and it is possible to treat the energy of tertiary interactions as a perturbation on the energy stabilizing the secondary structure.

The different challenges posed by the protein and RNA folding problems can be traced to the distinct ways in which the information contained in the sequence of these two molecular species controls the secondary and tertiary structures. We can think of the information in the amino acid sequence of a protein as branching into secondary and tertiary structural elements and also flowing between these elements in both directions. In an RNA, in contrast, the information in the sequence flows linearly, and largely in one direction, first to the secondary and then to the tertiary structure. An RNA molecule can thus be thought of as possessing a hierarchical structure in which the primary sequence determines the secondary structure which, in turn, determines its tertiary folding, whose formation alters only minimally the secondary structure. A folding algorithm then is simply a description of the information flow from the primary to the tertiary structure.

The postulated hierarchical nature of the RNA structure is undoubtedly an oversimplification, but it furnishes the basis on which a predictive folding algorithm can be generated. First, the primary structure can be used to establish the secondary structure through the use of simple, robust rules of secondary

folding. Next, the secondary structure can be used to predict the tertiary contacts in the structure, following again simple tertiary folding rules.

A useful algorithm to predict RNA folding need not tell us anything about the folding pathway of an RNA molecule; the hierarchical nature of the algorithm depends on the relative stabilities of the various structures, not the kinetics of folding. However, we think that the thermodynamics and kinetics of folding are directly linked in RNA. Consider an unfolded single strand of RNA at a temperature and ionic strength in which no base-pairs are stable. The single-stranded RNA is a flexible polynucleotide that has many different conformations. The temperature is now lowered slightly, or the ionic strength is raised slightly, so that base-pair formation is barely possible. Loops and bulges decrease the entropy of the single strand, so they form only if the free energy decrease of base pair formation more than balances the cost of loop closure. We thus expect to find some small amount of secondary structure involving the most stable G + C-rich helices. One or more hairpin loops have formed with the smallest, least destabilizing loops (tetraloops) favored. Larger hairpin loops can form if the loss of configurational entropy of closing the loop is compensated by increasing numbers of successive base-pairs. As the temperature is further lowered, or the ionic strength is raised further, increasing amounts of secondary structure form, including the entire range of loops, bulges and junctions. The structures formed at the higher temperatures do not necessarily survive at lower temperatures and some rearrangements can occur. Eventually, at low temperature or high ionic strength, the least stable structures involving tertiary interactions form.

We have described the hierarchical folding of the RNA as the conditions for folding are improved. Now let us consider the kinetics of folding when a single-stranded RNA is placed in a folding environment. Complementary base-pairs will collide randomly, but a single base-pair is never stable in aqueous solution. Hydrogen bonding provides at most 1 kcal mol^{-1} of stabilization; this is not enough to counteract the entropy loss of even a hairpin loop of four nucleotides ($\Delta G_{37^\circ} = +4.5 \text{ kcal mol}^{-1}$). This means that consecutive base-pairs must form to close a hairpin loop; this is the first step in folding. The rate of hairpin loop formation thus depends on the frequency of collision between complementary bases that initiate a sequence of base-pairs. This rate depends on the effective concentration of one base relative to another, i.e. on the effective relative volume that is accessible to the two bases. The effective relative volume depends, in turn, on the number of nucleotides separating the two bases, so small hairpin loops will form at the shortest times. Later the stems will grow forming internal loops and bulges, and eventually forming junctions of several stems. Tertiary interactions can occur among two, or more, secondary structural elements, but will almost always follow the for-

mation of these elements. The explanation is that the kinetics of folding favors formation of small hairpin loops for the same reasons that make these loops thermodynamically stable. Small loops have their ends close together; this means that the ends collide more often and do not lose much entropy on closing. The main conclusion is that the folding of RNA is sequential as well as hierarchical. We use sequential here to describe the kinetics of formation of secondary structure before tertiary structure; we are not referring to the folding of RNA during its biological synthesis.

We recognize that folding does not always occur on the path towards the lowest free energy structure. Clearly, during folding, secondary contacts may transiently form that must be undone before other interactions occur that lead the molecule in the direction of the final global free energy minimum. Of course, some species can remain kinetically trapped in non-equilibrium states. The enthalpy necessary to break the non-equilibrium base-pairs determines the activation energy of the kinetic trap.

Formation of secondary structure

There is still substantial disagreement in the literature as to what constitutes secondary and what constitutes tertiary structures in RNA (Chastain & Tinoco, 1991; Moore, 1999). The definitions we use here, and the logic for their choice, will become clear from the discussion.

Let us look in detail at the folding of a model RNA to form a hairpin tetraloop, the most abundant of the hairpin loops found in RNA (Woese *et al.*, 1990). As an example, we show the formation of a tetraloop as a function of temperature and Na^+ concentration in Figure 1. The melting

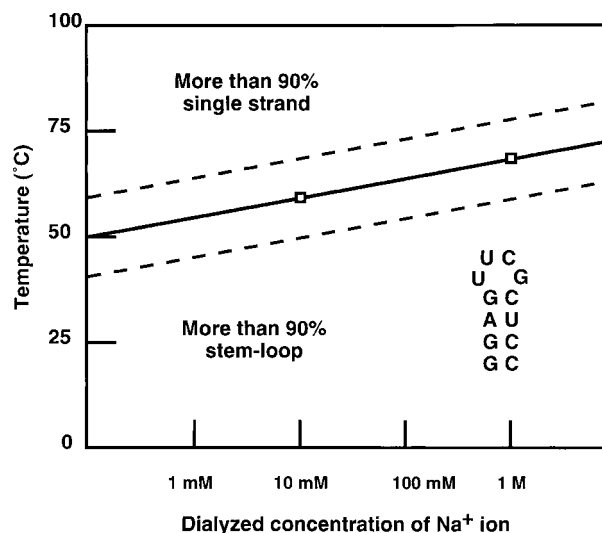


Figure 1. The melting of a stem-loop dependent on temperature and Na^+ ion concentration. Note that there is a 20°C temperature range around the melting temperature (the midpoint of the melting transition) where the RNA is partially folded.

temperature was measured (Antao *et al.*, 1991; Antao & Tinoco, 1992) at pH 7, 0.1 mM EDTA in 1 M NaCl, 10 mM phosphate buffer ($T_m = 67.7^\circ\text{C}$), and in 10 mM phosphate buffer ($T_m = 60.1^\circ\text{C}$), and extrapolated to the other salt concentrations. The broken lines indicate the region where the hairpin is partially formed (between 10% and 90%). The Figure illustrates several facts about RNA folding. The melting temperature is expected to vary logarithmically with salt concentration. The slope depends theoretically on the change in the number of ions bound on melting the structure (Cantor & Schimmel, 1980). Here we find a change in melting temperature of about 3.8°C per tenfold change in Na^+ concentration. For long RNA strands, and depending on the base composition of the duplex, values of from 8 to 20°C per tenfold change in Na^+ have been found (Steger *et al.*, 1980). The transition region is broad, here about 20°C ; the width depends on the enthalpy of melting, the more stable the stem the sharper the melting. The ion concentration is specified in terms of a dialyzed concentration to emphasize the fact that the total ion concentration depends on the RNA concentration. Clearly, an RNA strand of ten nucleotides at 1.0 mM concentration will have a total concentration of 10 mM monovalent positive ions in the absence of any added salt.

Figure 1 shows the folding behavior of a particularly simple RNA molecule, but it illustrates the more general case. As the temperature of an RNA is decreased slowly, the single strand begins to form stem loops. At the higher temperatures (and depending on the salt concentration) there will be multiple species: unfolded, partly folded, and fully folded. Some stems will grow longer, but they are eventually interrupted by internal loops and

bulges. There is competition among the growing stems; a kinetically favored loop may be replaced by a thermodynamically more favored one. Of course conformations formed first need not be replaced by more stable ones, the less stable form can be kinetically trapped if the cooling is too fast. Even if the solution is at thermodynamic equilibrium, the region between the broken lines in Figure 1 reminds us that more than one species may be present.

Figure 2(a) shows the P5abc region of the *Tetrahymena thermophila* group I self-splicing intron. The base-pairing (secondary structure) shown is based on NMR studies of the A-rich bulge (Luebke *et al.*, 1997), the P5b stem loop (Kieft & Tinoco, 1997), and a truncated version of P5abc (Wu & Tinoco, 1998). NMR is particularly appropriate for learning secondary structure because the imino proton resonances are well separated from the rest of the spectrum, and one peak is seen for each A·U and each G·C pair. Each G·U pair shows two imino resonances. We expect that the P5abc RNA starts folding at the two GNRA (N is any nucleotide; R is a purine) tetraloops near nucleotides 25 and 45 in the Figure. These hairpin loops should form as the temperature is lowered to allow the first base-pairs to be stable; they should also form first kinetically when the temperature is dropped below the melting range. The tetraloop stems then grow and form a three-stem junction (nucleotides 10-56 in Figure 2(a)). The 5' and 3' ends have now been brought close together; they form the closing stem that produces the A-rich bulge and a single U-base bulge. A bulge is distinguished from an internal loop by continuous pairing of the bases flanking the bulge. It is possible, but very unlikely, that the 5' and 3' ends would pair first as the temperature

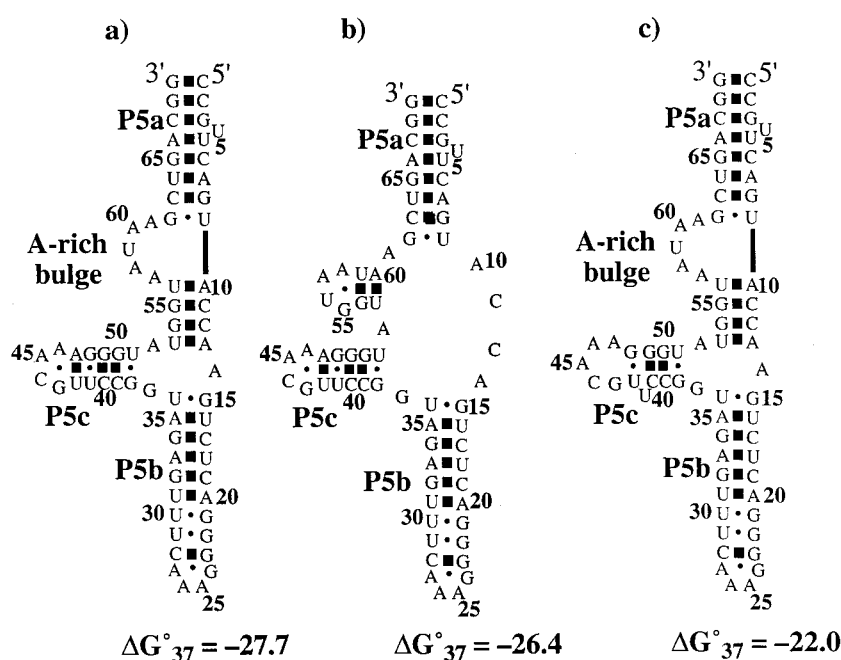


Figure 2. Predicted secondary structures of P5abc from a group I intron ribozyme using the Zuker algorithm available at <http://mfold1.wustl.edu/~mfold/rna/form1.cgi>. The calculated lowest free energy structure is found experimentally in 10 mM Na^+ . Tertiary interactions in Mg^{2+} change the base-pairing in the P5c stem, and at the closing base-pair of the A-rich bulge (see Figure 3).

is lowered. In the single strand the ends are far apart (which slows the kinetics of loop formation), and formation of the base-pair would greatly decrease the entropy of the RNA. Formation of the base-pair between the ends is thus kinetically and thermodynamically unfavored until most of the other pairs have already formed.

It may seem unnecessary to discuss the kinetics of folding if we are trying to predict only the final folded state. If the folded RNA is at equilibrium with respect to its conformation the folding path is irrelevant. However, knowledge of the folding path will make it easier to understand, or to predict, the final result. If the folded RNA is not at equilibrium, then the actual conformation can depend crucially on the path. We expect that most RNAs exist naturally in their thermodynamically most stable conformations; the free energy is a minimum. We can prepare RNA in their most stable conformation by slow cooling. Of course, we can also trap RNA in unstable forms by quick cooling, or by changing solvent at low temperature. Interactions with proteins, other RNAs, or any ligand can also change the conformation by stabilizing a higher free energy species, or catalyzing the refolding of a kinetically trapped species.

Prediction of RNA secondary structure from thermodynamics

The program *mfold* (Zuker, 1989) calculates secondary structures for RNAs in order of increasing free energy; the calculations take a few minutes for RNAs of a few hundred nucleotides. This program can be accessed through the internet at <http://mfold1.wustl.edu/~mfold/rna/form1.cgi>. The free energies of stems and loops used by the program come from experimentally measured values on oligonucleotides obtained by Turner and co-workers (reviewed by Burkard *et al.*, 1999). The main assumption is that the free energies of stems depend on nearest-neighbor sequences of base-pairs only, and that free energies of stems and loops are additive. The minimum free energy secondary structure for the P5abc sequence in Figure 2(a) was obtained in seconds by *mfold*. The calculated structure at 37°C, 1 M NaCl is identical with the NMR-deduced one; it has a calculated free energy, ΔG° , of $-27.7 \text{ kcal mol}^{-1}$. The RNA studied by NMR (56 nt) was a truncated version of P5abc with a shorter P5b stem and a shorter P5a terminal stem. However, the measured and calculated structures both have the same two tetraloops and the same five base-pair P5c stem. The A-rich bulge with its flanking stems is the same in both structures. The junctions are the same in both structures, except that the imino proton of A13·U53 is not seen in the NMR. This is not unusual for fraying base-pairs at the ends of stems; the terminal 5'G·C3' pair is rarely seen in an NMR imino spectrum. Good agreement is usually found between thermodynamically based calculated and

experimental secondary structures (Mathews *et al.*, 1999).

Calculated structures with higher free energies are also available from the program *mfold*. They would correspond to less stable secondary structures if the thermodynamic prediction was correct. One can also think of them as alternate secondary structures that are just as likely to be correct as the minimum free energy one, because of errors in the parameters and assumptions in the calculation. The fifth most stable calculated structure ($\Delta G^\circ = -26.4 \text{ kcal mol}^{-1}$) has another tetraloop instead of the A-rich bulge, and the three-stem junction has been changed to a four-stem junction (see Figure 2(b)). The calculated free energy is within 5% of the minimum, well within errors in the calculation. The 19th most stable calculated structure ($\Delta G^\circ = -22.0 \text{ kcal mol}^{-1}$) has the P5c stem loop changed. The tetraloop has been changed to a loop of five nucleotides, and the stem has only four base-pairs with a one-base bulge. The calculated free energy differs by 25% from the minimum. If this secondary structure for P5abc is correct, the thermodynamic calculation is incorrect; K_{eq} for the less stable form relative to the most stable is 10^{-4} . The 20 calculated lowest free energy conformations are minor variations of the three shown. The P5a and P5b stems are stable and do not vary, the less stable A-rich bulge and P5c region can vary significantly.

Many other RNA secondary structures have been determined (for a review, see Conn & Draper, 1998). The main structural aspects of the secondary folding not included in Figure 2 are the wide variety of non-Watson-Crick base-pairs (base-base mismatches) that have been found. Each base can form hydrogen bonds with every other base in many different orientations (Gesteland *et al.*, 1999); many have been seen. Only the very common sheared G·A base-pairs are shown in Figure 2.

Formation of tertiary structure

When P5abc is part of the entire ribozyme, or part of the P4-P6 domain of the ribozyme, the secondary structure seen is closest to the one with the least stable free energy ($\Delta G^\circ = -22.0 \text{ kcal mol}^{-1}$) shown in Figure 2(c). The reason for this change is the formation of tertiary structure. The structure of P4-P6 was determined by X-ray diffraction in a crystal (Cate *et al.*, 1996a); a schematic of the P5abc domain is shown in Figure 3. The five nucleotide A-rich bulge has been changed to two bulges, a four nucleotide bulge and a single guanine bulge. The guanine base is linked to the P5c stem by a bound Mg^{2+} . The P5c stem loop is changed from the tetraloop to a pentaloop closed by only three base-pairs. Two new G·A pairs bound to Mg^{2+} are formed. Essentially the same base-pairing scheme seen in Figure 3 was found by phylogenetic analysis (Michel & Westhof, 1990) of many group I introns. Covariation in sequences with the same function, of bases that are Watson Crick comp-

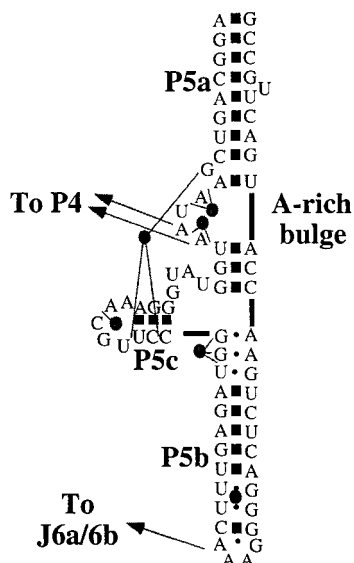


Figure 3. The secondary structure of P5abc when it is crystallized as part of an independently folding domain, P4-P5-P6, of the group I intron ribozyme. The circles show the location of bound Mg^{2+} , and arrows indicate tertiary interactions to the rest of the RNA.

lements indicates base-pairing. The secondary structure thus found corresponds to that present in the functional form i.e. the secondary structure in the presence of tertiary structure.

Comparison of Figure 3 and 2 shows one example of how the formation of tertiary structure can change the secondary structure. For P5abc the presence of Mg^{2+} and the presence of the rest of the ribozyme causes some of the less stable parts of the secondary structure to rearrange. Two adenine bases in the A-rich bulge interact with a loop in P4; the P5b tetraloop docks with its receptor in J4/J5; the P5c stem loop is connected to the surrounding RNA by Mg^{2+} . These tertiary interactions not only change the three-dimensional arrangement of the RNA; they also change the base-pairing. There are many other examples where the base-pairing is not changed. In tRNA the base-pairing in the cloverleaf is not changed when the tRNA is in its folded, functional form. Tertiary folding in tRNA involves formation of base triples and loop-loop base-pairing, but the stems and loops in the secondary structure are not changed.

Role of metal ions

Positively charged ions have a large general effect on RNA folding because of the negative charge on each phosphate group; their effect is like that of ionic strength on polyelectrolytes. However, divalent ions affect the tertiary folding much more than the secondary structure; specific metal ion binding occurs. (For recent reviews with references

to the earlier literature, see Draper & Misra, 1998, Misra & Draper, 1999.)

Metal ions in secondary structure

Divalent ion binding sites in RNA secondary structures seem to be pre-formed sites. This means that the site exists in the absence of the metal ion, and the structure of the site does not change significantly on binding the metal. Examples of these include G·U base-pairs in stem regions. A G·U base-pair has only hydrogen bond acceptors in the major groove: N7 and the carbonyl oxygen of G, and the 4-carbonyl oxygen of U. The A·U and G·C base-pairs have amino hydrogen bond donors from either cytosine or adenine bases in the major groove. Therefore G·U base-pairs, in particular runs of G·U pairs, are expected to be sites for binding solvated magnesium ions, $\text{Mg}(\text{H}_2\text{O})_6^{2+}$.

Cobalt hexammine and osmium hexammine ions in the crystal structure of the P4-P6 domain of the *T. thermophila* group I intron (Cate & Doudna, 1996; Cate *et al.*, 1997) first identified major groove binding sites for fully hydrated Mg^{2+} at tandem G·U base-pairs. Ion binding was studied in solution using $\text{Mn}(\text{H}_2\text{O})_6^{2+}$ (Allain & Varani, 1995), and $\text{Co}(\text{NH}_3)_6^{3+}$ (Colmenarejo & Tinoco, 1999; Kieft & Tinoco, 1997). The Mn^{2+} is paramagnetic and broadens the NMR resonances of nearby nuclei. The cobalt hexammine has 18 protons to provide intermolecular nuclear Overhauser effect (NOE) cross-peaks to precisely locate the RNA binding site. Addition of Mg^{2+} to the RNA causes changes in chemical shifts of the guanine and uracil imino protons at the binding site. These shifts not only confirm the location of the ion binding site, but they can be used to measure an equilibrium constant for the Mg^{2+} binding. Values of dissociation constants, K_d , in the millimolar range have been found (Colmenarejo & Tinoco, 1999; Kieft, 1997) for tandem G·U base-pairs.

Guanine-uracil binding sites have been characterized in solution for the P1 helix (Allain & Varani, 1995), the P5 helix (Colmenarejo & Tinoco, 1999), and the P5b stem loop (Kieft & Tinoco, 1997) of the *Tetrahymena* group I intron ribozyme. These sites are pre-formed in the double helices of the secondary structure, but the sites are all near tertiary interactions in the ribozyme. They are secondary structure ion binding sites that facilitate tertiary structure formation. The G·U pair in the P1 helix is right next to the splice site; it is at the center of the catalytic reaction. The tandem G·U pairs in the P5 helix are part of the docking site for the P9 loop (Golden *et al.*, 1998), and the tandem G·U pairs in the P5b stem loop are near the GAAA tetraloop that docks with its receptor in J6a/6b (Cate *et al.*, 1996b). Presumably the bound metal ions reduce the electrostatic repulsion of the phosphate groups from the different secondary structural elements that must come together in the tertiary structure. In fact, the adenine platform in the J6a/6b tetraloop receptor (a secondary structur-

al element) has a specific binding site for a potassium ion (Basu *et al.*, 1998) that further reduces the net negative potential at the tertiary interaction.

A crystal structure of a 5 S ribosomal RNA domain (Correll *et al.*, 1997) found magnesium ions bound in the major groove of an internal loop with three non-Watson-Crick base-pairs. Hexahydrated $\text{Mg}(\text{H}_2\text{O})_6^{2+}$ was found, but pentahydrated magnesium ions were also seen, with the remaining metal coordination to phosphate oxygen atoms, or N7, or to carbonyl oxygen atoms.

We conclude that after the secondary structure forms, metal ions bind to the pre-formed sites; the metal ion is not required for the formation of the site. However, the presence of the ion in the site favors the formation of tertiary structure by reducing electrostatic repulsion of the phosphate ions.

Metal ions in tertiary structure

Some metal ion binding sites are formed only after, or during, the formation of tertiary structure. For example, when a single strand at the end of a stem loop folds back to form a pseudoknot, a metal ion binding site can form. Such a site has been seen in the mouse mammary tumor virus pseudoknot that promotes the programmed frameshift vital to the propagation of this retrovirus (Gonzalez & Tinoco, 1999). The metal ion, either $\text{Mg}(\text{H}_2\text{O})_6^{2+}$ or $\text{Co}(\text{NH}_3)_6^{3+}$, binds in a pocket between the major groove of the stem and the short loop that crosses the stem. The binding site is formed as the tertiary structure, the pseudoknot, is formed. Divalent ions are known to stabilize pseudoknots relative to their constituent hairpins (Qiu *et al.*, 1996; Wyatt *et al.*, 1990). This does not mean that a pseudoknot forms only in the presence of divalent ions; high concentrations of univalent ions can replace the divalent ions. An analogous binding site can form when two hairpin loops pair to form kissing hairpins. The ion binding site has not been characterized yet, but Mg^{2+} is bound when the kissing complex is formed (Gregorian & Crothers, 1995). Complete base-pairing of the two loops requires tight turns in the loops that brings several phosphates very close together (Chang & Tinoco, 1997). They may produce the preferential divalent ion binding sites.

The crystal structure of the P4-P6 domain of the group I intron (Cate & Doudna, 1996; Cate *et al.*, 1997) have revealed a wide range of tertiary metal ion binding sites. Magnesium ions knit the tertiary structure of the P4-P6 domain together, and in so doing change the secondary structure. The A-rich bulge of P5abc is turned inside out when the tertiary structure forms. The isolated A-rich bulge (Luebke *et al.*, 1997), and the A-rich bulge in P5abc without Mg^{2+} (Wu & Tinoco, 1998) has five nucleotides in the bulge (AAUAA) with the adenine bases stacked inside the bulge. In the crystal structure there are two Mg^{2+} bound to phosphate groups inside the bulge; the bases are on the outside with two adenine bases interacting with the

P4 stem. The bulge now contains only four nucleotides (AAUA) because the terminal adenine base has replaced a G·U pair by an A·U and produced a single-base G-bulge. This guanine base is bonded by Mg^{2+} to the P5c stem. Two G·A base mismatches bound to Mg^{2+} occur in the three-stem junction (see Figure 3). The metal ions shown in the figure are called a magnesium ion core whose formation is necessary for the tertiary folding of the P4-P6 domain (Cate *et al.*, 1997). Again, the binding sites only appear when the tertiary structure forms.

We thus distinguish two types of specifically-bound metal ions: (1) the metal ions that bind to pre-formed sites in secondary structure and do not change the structure of the site significantly; and (2) the metal ions that are part of the tertiary structure site; without the metal ion neither the site nor the tertiary structure exists.

A folding algorithm for RNA

The hierarchical and sequential folding of RNA greatly simplifies the prediction of the RNA folded state, or states, because it divides the problem in two. First, the possible base-pairing schemes of the molecule are calculated from its sequence to obtain the secondary structure. Next, the possible interactions between these base-paired structures are evaluated to obtain the functional, tertiary structure. This formal separation of the RNA folding problem into two parts has an experimental counterpart. Magnesium or other divalent ions preferentially stabilize tertiary structures; thus, the secondary structure of RNA can be characterized at low ionic strength in Na^+ , and then Mg^{2+} can be added to form the tertiary structure. The first part of the RNA folding problem is nearly solved; the second part is just beginning to be attacked.

Prediction of RNA secondary structure

As we have discussed in detail earlier, secondary structure can be predicted well using programs such as that by Zuker (1989). A range of secondary structures (helices, hairpin loops, internal loops, bulges, and junctions) is calculated as the first step of the folding algorithm. We favor the Zuker program 'mfold', but any method that gives many possible secondary structures consistent with the sequence is good. The results from the Zuker program depend on experimental thermodynamic data mainly from the Turner group (Mathews *et al.*, 1999); as more data are obtained the predictions should improve. In whatever way we obtained possible secondary structures, there are extensive thermodynamic data that tell us how changing a secondary structure changes the free energy. We know how breaking base-pairs, or forming new ones, affects the thermodynamic stability of the molecule. We can thus assess which secondary structure elements can be changed without much cost in free energy. The kinetics of unfolding and

refolding may also be important; the time scales can be estimated from the enthalpy changes involved.

Prediction of RNA tertiary structure

In the second step of the algorithm, possible tertiary folds are formed from the various calculated secondary structures. We minimize the destabilization of secondary structure in forming the tertiary structure. Of course, we should also choose the most thermodynamically stable tertiary interactions, but thermodynamic data for these interactions are still very sparse. We will summarize the little that is known about tertiary interactions.

Figure 4 shows thermodynamic data for the formation of a secondary structure hairpin (the equilibrium involves both possible hairpins, but only one is shown in the Figure), followed by a tertiary structure pseudoknot in the presence and absence of Mg^{2+} . In 200 mM NaCl, 10 mM Na phosphate, 100 μ M EDTA (pH 6.4), the pseudoknot melts at 27.8°C, and the hairpins melt at 63.2°C (Gonzalez & Tinoco, 1999). Thus the pseudoknot is unstable at 37°C, whereas the hairpins are still present above 60°C. Adding micromolar amounts of Mg^{2+} stabilize both transitions, but the pseudoknot is preferentially stabilized, causing the two melting transitions to overlap. Even at a concentration of Mg^{2+} as low as 50 μ M the hairpins already melt at over 90°C so their thermodynamics cannot be analyzed, and the pseudoknot melts at 68°C. At a

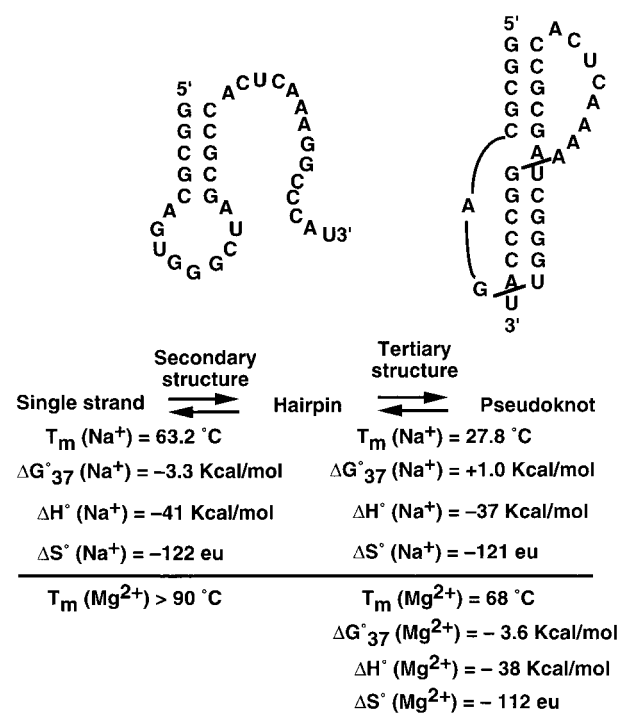


Figure 4. Comparison of thermodynamics for secondary (a hairpin stem loop) and tertiary structure (a pseudoknot) formation. The data in Na^+ refer to 200 mM NaCl, 10 mM Na phosphate, pH 6.4; the data in Mg^{2+} refer to the addition of 50 μ M Mg^{2+} (Gonzalez & Tinoco, 1999).

Mg^{2+} concentration of 5 to 10 mM, typical of biochemical experiments, both transitions are at temperatures too high to measure. The figure gives the thermodynamic parameters for the transitions, and illustrates the small changes in ΔH° and ΔS° that can combine to produce changes in sign of ΔG_{37}° . Of course, to predict tertiary structure we need the thermodynamic data not for one pseudoknot, but for pseudoknot formation as a function of stem sequence and length, and as a function of loop lengths.

Kissing hairpins, or hairpin loop-hairpin loop complexes, were studied by Gregorian & Crothers (1995) in 50 mM Na^+ , 5 mM Mg^{2+} . Four complexes with hairpin loops of complementary sequences of seven nucleotides were melted. Although the four complexes could all form three G-C base-pairs and four A-U base-pairs in the loop-loop helix, the melting temperatures varied from 32°C to 59°C. Kissing complexes with hairpin loops of five to seven nucleotides allow the maximum number of base pairs to form because of the 11 base-pairs per turn in an A-form helix. We thus need to characterize the sequence dependence of this interaction.

Non-Watson-Crick tertiary interactions include base-base mismatches, base-base triples formed in the major groove of the RNA, and base-nucleoside triples formed in the minor groove through hydrogen bonding to the 2'-hydroxyl of ribose-ribose zippers. Other loop-loop, loop-helix, and helix-helix interactions can occur. Many of these interactions have been seen in the group I intron ribozyme (Cate *et al.*, 1996a; Golden *et al.*, 1998), but thermodynamic data that can be assigned to specific interactions is not extensive. Replacement of ribose with its 2'-hydroxyl group by deoxyribose indicates a ΔG_{37}° of about -1 kcal mol^{-1} per ribose-base interaction in the minor groove (Bevilacqua & Turner, 1991; Narlikar *et al.*, 1997). The knowledge about secondary structure thermodynamics and the lack of knowledge about tertiary structure is illustrated in Table 1.

The ability to predict tertiary folding awaits more thermodynamic data, but metal ion binding sites in a secondary structure can suggest the locations of tertiary interactions. Knowledge of the role of Mg^{2+} in tertiary structure formation will be very valuable.

Atomic coordinates for RNA structures

At present, atomic coordinates can be obtained only from experimental measurements; we do not attempt to predict them. X-ray diffraction from crystals gives the highest accuracy, but the structure obtained is not always the desired structure (Holbrook & Kim, 1997). Tetraloops in solution tend to crystallize as double strands with internal loops (Baeyens *et al.*, 1996; Holbrook *et al.*, 1991). A DNA enzyme with its RNA substrate crystallized as a four-stranded cruciform (Nowakowski *et al.*, 1999). NMR provides atomic coordinates in solution for rigid parts of the RNA, and leads to

Table 1. Thermodynamics needed for prediction of RNA folding

A. Secondary structure ^a			
Structural element	ΔG_{37}° (kcal mol ⁻¹)	ΔH° (kcal mol ⁻¹)	ΔS° (cal mol ⁻¹ K ⁻¹)
Hairpin loop closure ($n = 3$ to 10)	+4.5 to +5.9	0	-14.5 to -19.0
Internal loop closure ($n = 2$ to 10)	+4.1 to +6.3	0	-13.2 to -20.3
Bulge loop closure ($n = 1$ to 10)	+3.9 to +5.5	0	-12.6 to -17.7
Watson-Crick base-pair addition per base-pair	-0.9 to -3.4 Average = -1.9	-5.7 to -14.2 Average = -9.6	-15.5 to -34.9 Average = -24.9
B. Tertiary structure in Mg ²⁺			
Structural element	ΔG_{37}° (kcal mol ⁻¹)	ΔH°	ΔS°
Tertiary loop closure	+4 to +6	-	-
Watson-Crick base-pair formation per base-pair (pseudoknots, kissing hairpins)	-1 to -3	-	-
Major groove base-base triples per base-pair	~ -1	-	-
Minor groove base-nucleoside triples (ribose zipper) per base-pair	~ -1	-	-
Tetraloop - receptor	-	-	-
Helix-helix per interaction	-	-	-

^a Data from Turner *et al.* (1988) and Jaeger *et al.* (1989).

ranges of coordinates for flexible parts. The uncertainty in coordinates may be closer to the truth in representing a flexible molecule in solution, than the coordinates obtained from a crystal. However, NMR is limited in the size of the RNA that can be analyzed at atomic resolution. The accessible size is continually increasing; at present the upper limit is about 100 nt. Bound metal ions and their hydration shells, as well as bound water molecules, can be seen in crystal structures. Ion binding sites can be identified in solution by NMR, and the thermodynamics of ion binding can be measured.

Prediction of correct atomic coordinates for RNA from sequence alone, i.e. without experimental data input, is very difficult (for a review, see Auffinger & Westhof, 1998). Molecular mechanics or molecular dynamics calculations need experimental data to guide the calculation toward the the correct free energy minimum. Calculations that sort through possible local conformations (Gautheret *et al.*, 1993) also need experimental guidance to select the right structure. Once an atomic resolution structure is obtained, electrostatic potentials can be calculated to suggest likely ion binding sites and to provide an estimate of the electrostatic contributions to the free energy as a function of conformation (Hermann & Westhof, 1998). The free energies required to dehydrate metal ions can be used to predict whether direct coordination, or outer shell coordination, of the metal ion will occur (Draper & Misra, 1998).

Conclusion

The logic for prediction of tertiary structure follows from the hierarchy of RNA folding. First predict, or determine experimentally, secondary structure, including metal ion binding sites. Second, examine possible tertiary base-pairs that form base triples, pseudoknots, and kissing hairpin

loops or other loop-loop interactions. Finally, consider metal ion-mediated tertiary interactions. Tertiary interactions will change only the weakest of the secondary structures in producing the final structure, leaving the rest largely unaffected. Thus, our ability to correctly predict the folded structure will depend on obtaining thermodynamic data for the different interactions. At present the thermodynamics of secondary structure is in good hands with continuing improvements in the parameters needed for prediction (Mathews *et al.*, 1999). The thermodynamics of tertiary interactions is almost non-existent. Some thermodynamic data for pseudoknots and kissing hairpins have been obtained (Gonzalez & Tinoco, 1999; Qiu *et al.*, 1996; Gregorian & Crothers, 1995; Wyatt *et al.*, 1990), but thermodynamics of other tertiary interactions, particularly involving metal ions, is sorely lacking. Data on ΔH° , ΔS° , and ΔG° of tertiary structure formation in the presence of Mg²⁺ will be the most helpful in achieving the goal of predicting functional structures from sequence alone.

In addition to the standard spectroscopic methods, including NMR, for measuring thermodynamics, force vs. extension methods applied to a single RNA molecule seem very promising (Essevaz-Roulet *et al.*, 1997; Rief *et al.*, 1999). By attaching one end of a tethered RNA molecule to an atomic force microscope cantilever, the force necessary to break base pairs, or tertiary interactions can be measured. The reversible work (omitting PV work) done on a system at constant temperature and pressure is equal to the free energy change of the system. Thus at equilibrium the force times extension of an RNA molecule can be integrated to obtain the free energy of unfolding. The experiment can be done at any temperature in the presence and absence of Mg²⁺ to distinguish secondary structure unfolding from tertiary structure interactions. The ability to unfold a

molecule at 5 °C, or 25 °C, or 37 °C in our choice of solvent is a tremendous advantage. In the past the variables we could control for unfolding an RNA (or protein) were temperature and solvent. Now we can apply a force to the ends of the RNA, or to any two residues where we can attach handles, to study the unfolding in any environment, in the presence of any ligands, including bound proteins. Obtaining and completing this thermodynamic library of tertiary interactions is essential for RNA structural prediction.

We want to predict the folded form of an RNA well enough to determine and understand, its function. The algorithm described here to do this will surely encounter many difficulties. For example, thermodynamics of tertiary interactions will have to be known with an accuracy better than the energy difference between alternative structures, and there may be more than one conformation present at equilibrium. Furthermore, the function may involve a flexible RNA that folds into a single conformation only on binding the correct ligand, or the function may require motion of the RNA; the function may involve an RNA molecule that behaves as a molecular motor, capable of converting chemical into mechanical energy and of generating force. In some instances, only part of the structure may be crucial for function, and we need not even try to predict the rest.

These complications notwithstanding, it is clear that if more experimentalists and theoreticians begin to address the issues discussed here to prove the proposed algorithm wrong (or to improve it), the goal of predicting and understanding the relation between structure and function in RNA will be significantly facilitated.

Acknowledgments

We thank Mr Ruben Gonzalez for providing the thermodynamic data used in Figure 4 and Dr Michael Schmitz for helpful comments. This research was supported in part by National Institute of Health Grant GM 10840 (I.T.), GM 32543 (C.B.) and NSF grant DBI9732140 (C.B.), and by the Department of Energy Grant DE-FG03-86ER60406 (I.T.).

References

- Allain, F. H.-T. & Varani, G. (1995). Divalent metal ion binding to a conserved wobble pair defining the upstream site of cleavage of group I self-splicing introns. *Nucl. Acids Res.* **23**, 341-350.
- Antao, V. P. & Tinoco, I., Jr (1992). Thermodynamic parameters for loop formation in RNA and DNA hairpin tetraloops. *Nucl. Acids Res.* **20**, 819-824.
- Antao, V. P., Lai, S. Y. & Tinoco, I., Jr (1991). A thermodynamic study of unusually stable RNA and DNA hairpins. *Nucl. Acids Res.* **19**, 5901-5905.
- Auffinger, P. & Westhof, E. (1998). Simulations of the molecular dynamics of nucleic acids. *Curr. Opin. Struct. Biol.* **8**, 227-236.
- Baeyens, K. J., De Bondt, H. L., Pardi, A. & Holbrook, S. R. (1996). A curved RNA helix incorporating an internal loop with G.A and A.A non-Watson-Crick base pairing. *Proc. Natl Acad. Sci. USA*, **93**, 12851-12855.
- Basu, S., Rambo, R. P., Strauss-Soukup, J., Cate, J. H., Ferré-D'Amaré, A. R., Strobel, S. A. & Doudna, J. A. (1998). A specific monovalent metal ion integral to the AA platform of the tetraloop receptor. *Nature Struct. Biol.* **11**, 986-992.
- Bervilacqua, P. C. & Turner, D. H. (1991). Comparison of binding of mixed ribose-deoxyribose analogues of CUCU to a ribozyme and to GGAGAA by equilibrium dialysis: evidence for ribozyme specific interactions with 2' OH groups. *Biochemistry*, **30**, 10632-10640.
- Brion, P. & Westhof, E. (1997). Hierarchy and dynamics of RNA folding. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 113-137.
- Burkard, M. E., Turner, D. H. & Tinoco, I., Jr (1999). The interactions that shape RNA structure. In *RNA World* (Gesteland, R. F., Cech, T. R. & Atkins, J. F., eds), 2nd edit., pp. 233-264, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Cantor, C. R. & Schimmel, P. R. (1980). *Biophysical Chemistry. The Behavior of Biological Macromolecules*, vol. 3, pp. 1139-1142, W. H. Freeman, San Francisco.
- Cate, J. H. & Doudna, J. A. (1996). Metal-binding sites in the major groove of a large ribozyme domain. *Structure*, **4**, 1221-1230.
- Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R. & Doudna, J. A. (1996a). Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science*, **273**, 1678-1685.
- Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Szwczak, A. A., Kundrot, C. E., Cech, T. R. & Doudna, J. A. (1996b). RNA tertiary structure mediation by adenosine platforms. *Science*, **273**, 1696-1699.
- Cate, J. H., Hanna, R. L. & Doudna, J. A. (1997). A magnesium ion core at the heart of a ribozyme domain. *Nature Struct. Biol.* **4**, 553-558.
- Chamorro, M., Parkin, N. & Varmus, H. E. (1992). An RNA pseudoknot and an optimal heptameric shift site are required for highly efficient ribosomal frameshifting on a retroviral messenger RNA. *Proc. Natl Acad. Sci. USA*, **89**, 713-717.
- Chang, K.-Y. & Tinoco, I., Jr (1997). The structure of an RNA "kissing" hairpin complex of the HIV TAR hairpin loop and its complement. *J. Mol. Biol.* **269**, 52-66.
- Chastain, M. & Tinoco, I., Jr (1991). Structural elements in RNA. *Prog. Nucl. Acid Res. Mol. Biol.* **41**, 131-177.
- Colmenarejo, G. & Tinoco, I., Jr (1999). Structure and thermodynamics of metal binding in the P5 helix of a group I intron ribozyme. *J. Mol. Biol.* **290**, 119-135.
- Conn, G. L. & Draper, D. E. (1998). RNA structure. *Curr. Opin. Struct. Biol.* **8**, 278-285.
- Correll, C. C., Freeborn, B., Moore, P. B. & Steitz, T. A. (1997). Metals, motifs, and recognition in the crystal structure of a 5S rRNA domain. *Cell*, **91**, 705-712.
- Draper, D. E. & Misra, V. K. (1998). RNA Shows its Metal. *Nature Struct. Biol.* **11**, 927-930.
- Essevaz-Roulet, B., Bockelmann, U. & Heslot, F. (1997). Mechanical separation of the complementary strands of DNA. *Proc. Natl Acad. Sci. USA*, **94**, 11935-11940.

- Gautheret, D., Major, F. & Cedergren, R. (1993). Modeling the three-dimensional structure of RNA using discrete nucleotide conformational sets. *J. Mol. Biol.* **229**, 1049-1064.
- Gesteland, R. F., Cech, T. R. & Atkins, J. F. (1999). Editors of *The RNA World*, 2nd edit., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Giegé, R., Frugier, M. & Rudinger, J. (1998). tRNA mimics. *Curr. Opin. Struct. Biol.* **8**, 286-293.
- Golden, B. L., Gooding, A. R., Podrell, E. R. & Cech, T. R. (1998). A preorganized active site in the crystal structure of the *Tetrahymena* ribozyme. *Science*, **282**, 259-264.
- Gonzalez, R. L., Jr & Tinoco, I., Jr (1999). Solution structure and thermodynamics of a divalent metal ion binding site in an RNA pseudoknot. *J. Mol. Biol.* **289**, 1267-1282.
- Gregorian, R. S., Jr & Crothers, D. M. (1995). Determinants of RNA hairpin loop-loop complex stability. *J. Mol. Biol.* **248**, 968-984.
- Hermann, T. & Westhof, E. (1998). Exploration of metal ion binding sites in RNA folds by Brownian dynamics simulations. *Structure*, **6**, 1303-1314.
- Holbrook, S. R. & Kim, S. H. (1997). RNA crystallography. *Biopolymers*, **44**, 3-21.
- Holbrook, S. R., Cheong, C., Tinoco, I. J. & Kim, S.-H. (1991). Crystal structure of an RNA double helix incorporating a track of non-Watson-Crick base pairs. *Nature*, **353**, 579-581.
- Jaeger, J. A., Turner, D. H. & Zuker, M. (1989). Improved predictions of secondary structures for RNA. *Proc. Natl Acad. Sci. USA*, **86**, 7706-7710.
- Kieft, J. S. (1997). Structure and thermodynamics of a metal binding site in the RNA major groove: cobalt (III) hexammine as a probe. PhD thesis, University of California, Berkeley.
- Kieft, J. S. & Tinoco, I., Jr (1997). Solution structure of a metal-binding site in the major groove of RNA complexed with cobalt (III) hexammine. *Structure*, **5**, 713-721.
- Lisman, J. E. & Fallon, J. R. (1999). What maintains memories? *Science*, **283**, 339-340.
- Lowe, T. M. & Eddy, S. R. (1999). A computational screen for methylation guide snoRNAs in yeast. *Science*, **283**, 1168-1171.
- Luecke, K. J., Landry, S. M. & Tinoco, I., Jr (1997). Solution conformation of a five-nucleotide RNA bulge loop from a group I intron. *Biochemistry*, **36**, 10246-10255.
- Mathews, D. H., Sabina, J., Zuker, M. & Turner, D. H. (1999). Expanded sequence dependence of thermodynamic parameters provides improved prediction of RNA secondary structure. *J. Mol. Biol.* **288**, 911-940.
- McCarthy, J. E. G. & Gualerzi, C. (1990). Translational control. *Trends Genet.* **6**, 78-85.
- Michel, F. & Westhof, E. (1990). Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* **216**, 585-610.
- Misra, V. K. & Draper, D. E. (1999). On the role of magnesium ions in RNA stability. *Biopolymers*, **48**, 113-135.
- Moore, P. B. (1999). The RNA Folding Problem. In *The RNA World* (Gesteland, R. F., Cech, T. R. & Atkins, J. F., eds), 2nd edit., pp. 381-401, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Narlikar, G. J., Khosla, M., Usman, N. & Herschlag, D. (1997). Quantitating tertiary binding energies of 2' OH groups on the P1 duplex of the *Tetrahymena* ribozyme: intrinsic binding energy in an RNA enzyme. *Biochemistry*, **36**, 2465-2477.
- Noller, H. F., Hoffarth, V. & Zimniak, L. (1992). Unusual resistance of peptidyl transferase to protein extraction procedures. *Science*, **256**, 1416-1419.
- Nowakowski, J., Shim, P. J., Prasad, G. S., Stout, C. D. & Joyce, G. F. (1999). Crystal structure of an 82-nucleotide RNA-DNA complex formed by the 10-23 DNA enzyme. *Nature Struct. Biol.* **6**, 151-156.
- Palmenberg, A. C. & Sgro, J.-V. (1997). Topological organization of picornaviral genomes: statistical prediction of RNA structural signals. *Semin. Virol.* **8**, 231-247.
- Qiu, H., Kaluarachchi, K., Du, Z., Hoffman, D. W. & Giedroc, D. P. (1996). Thermodynamics of folding of the RNA pseudoknot of the T4 gene 32 autoregulatory messenger RNA. *Biochemistry*, **35**, 4176-4186.
- Rief, M., Clausen-Schaumann, H. & Gaub, H. E. (1999). Sequence-dependent mechanics of single DNA molecules. *Nature Struct. Biol.* **6**, 346-349.
- Spicher, A., Guicheret, O. M., Duret, L., Aslanian, A., Sanjines, E. M., Denko, N. C., Giacca, A. J. & Blau, H. M. (1998). Highly conserved RNA sequences that are sensors of environmental stress. *Mol. Cell. Biol.* **18**, 7371-7382.
- Steger, G., Müller, H. & Riesner, D. (1980). Helix-coil transitions in double-stranded viral RNA. Fine resolution melting and ionic strength dependence. *Biochim. Biophys. Acta*, **606**, 274-284.
- Turner, D. H. & Sugimoto, N. (1988). RNA structure prediction. *Annu Rev. Biophys. Biophys. Chem.* **17**, 167-192.
- 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.
- Wu, M. & Tinoco, I., Jr (1998). RNA folding causes secondary structure rearrangement. *Proc. Natl Acad. Sci. USA*, **95**, 11555-11560.
- Wyatt, J. R., Puglisi, J. D. & Tinoco, I., Jr (1990). RNA pseudoknots: stability and loop size requirements. *J. Mol. Biol.* **214**, 455-470.
- Zuker, M. (1989). On finding all suboptimal foldings of an RNA molecule. *Science*, **244**, 48-52.