

OVERVIEW OF THE QUESTIONS

The questions to be answered in the study of RNA folding are simple. What is the final structure? How does the molecule get there? How long does it take? If only the answers were equally simple. In cells, RNA is constantly being synthesized and folded while it undergoes continuous conformational switching and degradation. A vast time domain is covered, from picoseconds and nanoseconds for motions in individual nucleotides, to microseconds for forming small hairpin helices, to milliseconds for forming relatively compact tertiary structures such as in tRNA, to minutes for folding more complex molecules such as introns. The process of undoing incorrect folds, or renaturation, can take hours, days, or much longer. The final folded state is usually well defined, but the starting state and transient intermediates are likely to be very polymorphic. The number of possible folding pathways is huge for large molecules, and a correspondingly complex time spectrum can be expected. Different regions of the molecule may have quite distinct time constants for reaching their final form, so it is essential that the experiments be capable of resolution in both time and space. As complex as this problem may seem, there is some consolation in the fact that it is probably simpler than the protein folding problem.

HOW RNA AND PROTEINS DIFFER IN THE FOLDING PROBLEM

Secondary structure is a stronger organizing principle in RNA than in proteins. For example, imagine that an $\alpha\beta$ -dimer-forming protein is mixed with all the proteins in the cell, and the solution is subjected to thermal or chemical denaturation, causing separation of the subunits. Then it is slowly annealed back to native conditions. There is only a small probability that native $\alpha\beta$ dimer will emerge from this messy process. Yet, if the same experiment were done with a double-helical RNA in the presence of all the RNA in the cell, the chances of recovering the duplex with native pairing are quite good, as verified by the widespread use of hybridization to identify DNA and RNA sequences in complex nucleic acid mixtures. (In such hybridizations, of course, an excess of strand α is added in order to drive the kinetics of formation of the $\alpha\beta$ duplex.)

Double-helical nucleic acids are stabilized primarily by the free energy of stacking the bases together. However, this interaction is relatively nonspecific, since there are only modest variations in the enthalpy of base pairing, depending on the nature of the dinucleotide step (Breslauer et al., 1986; Xia et al., 1998; UNIT 11.2). Specificity arises from the highly polar base-pairing interactions. The result of these strong, directional forces is that the secondary structure elements of RNA, i.e., local hairpin helices, are thermodynamically stable in the absence of other interactions. This is in contrast to proteins, in which α helices usually have only marginal stability by themselves.

RNAs are capable of forming stable hairpin helices rapidly, followed by slower formation of the noncanonical interactions in the tertiary structure (Brion and Westhof, 1997; Tinoco and Bustamante, 1999). In these simple cases, the folding pathway through the energy “landscape” is well defined, although even apparently simple cases can involve some rearrangement of secondary structure upon tertiary structure formation (Wu and Tinoco, 1998). In more extreme cases, the high stability of improper double helices can cause RNAs to take long detours through metastable states in deep local energy minima that are characterized as “denatured” by their distinctive mobility on electrophoresis gels. The comparatively large per-residue enthalpy of double helix formation means that elevated temperature is frequently needed to overcome the activation barrier for melting the incorrect structure to produce the native form.

HOW WE KNOW THE FINAL STRUCTURE

In only a few cases does crystallography provide knowledge of the structure of RNA molecules large enough to encode most of the complexity of the folding problem. The classic example is tRNA, augmented by more recent work on ribozymes such as the P4-P6 domain of the *Tetrahymena* ribozyme (Cate et al., 1996). For molecules of this size, NMR spectroscopy is not yet a useful tool, nor is any other spectroscopic method able to provide a general solution of the structural problem.

The primary source of definitive information about the secondary structure and some tertiary structural elements in naturally occur-

ring RNAs is phylogenetic analysis of the conservation of pairing interactions. The basic idea is that base pairs in a structure have been conserved over evolution, even when particular sequences were not. Thus, if mutation of A to G at one site is strongly correlated with mutation of T to C at another, it is likely that the two positions are base paired, probably by Watson-Crick pairing. This logic enabled, for example, the work that defined the secondary structure of 23S ribosomal RNA (Noller et al., 1981), the tertiary structure of RNase P (Harris and Pace, 1995), and the secondary structures of Y RNAs (Farris et al., 1999) and telomerase RNA (Chen et al., 2000).

An alternative approach is to compare the predicted thermodynamic stability of different possible secondary structures, and select the most stable of these. The empirical parameters and software necessary for this analysis have recently been updated by Turner and Zucker and their collaborators (UNIT 11.2; Xia et al., 1998; Matthews et al., 1999). This method has the disadvantage of not accounting for tertiary structure that can influence the relative free energy of structures that differ in secondary structure. The current methods predict ~73% of the known base pairs (Matthews et al., 1999). However, knowledge of relative secondary structure free energies can be vitally important when assessing the possible importance of alternative transient structures in the folding pathway.

Finally, there is the set of low-resolution structural methods that come under the heading of footprinting. These take advantage of the variation in chemical and enzymatic reactivity depending on the structural state of individual nucleotides. Of particular interest for the folding problem are those that can be carried out rapidly enough to follow the folding reaction in real time (see Resolution in Space and Time).

FOLDING VERSUS UNFOLDING

The earliest studies of the RNA folding problem involved observation of thermal melting curves, e.g., for tRNAs (Riesner et al., 1969, 1970, 1973; Cole and Crothers, 1972; Cole et al., 1972). It might be asked how one can learn about folding by studying unfolding, but it should be recognized that relaxation kinetic measurements yield both forward and reverse reaction rate constants, so that information about the folding steps is inherent in the kinetic data. Direct measurement of folding rates for the tertiary structure in yeast tRNA^{Phe} by

stopped-flow mixing with Mg²⁺ to induce folding verifies the convergence of the two approaches (Maglott and Glick, 1997). A perhaps more serious reservation has to do with the series of equilibrium states examined in relaxation kinetics. Small temperature jumps clearly do not correspond to the intracellular process, nor to the way folding is triggered in nature. But neither is a sudden jump in Mg²⁺ concentration to induce folding a physiologically relevant mechanism. The pathway of RNA folding in the cell is a largely unexplored experimental problem.

Understanding of the thermodynamics of RNA secondary structure is now sufficiently advanced that individual components of multiphasic melting curves for small model RNAs can generally be assigned to melting of specific helices (UNIT 11.3; Draper and Gluick, 1995). For somewhat more complicated molecules, the disagreement between predicted and observed melting curves can confidently be assigned to tertiary structural interactions, although often of unknown nature (Draper and Gluick, 1995; UNIT 11.3).

RESOLUTION IN SPACE AND TIME

The general objective of studies of RNA folding pathways is to chart the state of individual nucleotides, helices, or tertiary structural elements as a function of time after a perturbation alters the conformational state. The early work on the dynamics of tRNA tertiary structure, referred to above, exploited the instability of the tertiary structure in absence of Mg²⁺, thus making the tertiary structure unfolding transition the first component of the thermal melting curve (UNIT 11.3). At 100 to 200 mM monovalent counterion concentration, the time constants for tRNA tertiary structure formation are typically in the range of 10 to 100 msec. Faster signals at higher temperature correspond to melting of hairpin helices, which can be identified by a combination of NMR and relaxation kinetics (Crothers et al., 1974; Hilbers et al., 1976). Imino proton exchange of hairpin helices results in line broadening of the corresponding resonances when the helix lifetime is less than ~5 msec (Crothers et al., 1974). The measured temperature dependence of the helix dissociation rate from relaxation kinetics can be extrapolated to lower temperatures, to match a 5 msec lifetime to a temperature zone for broadening of a group of resonances corresponding to a specific helix. In this way, the structural origin of the relaxation signals can

be established (Crothers et al., 1974; Hilbers et al., 1976).

Another important clue for determining the structural origin of relaxation signals is derived from the fortuitous presence of 4-thiouridine (S⁴U) bases in natural tRNAs. For example, the S⁴U residue at position 8 in tRNA^{fMet}, between the acceptor stem and DHU helices, shows a large hyperchromism in the tertiary melting transition (Cole and Crothers, 1972; Cole et al., 1972). Modern synthetic methods, which permit substitution of modified bases at any position, should allow examination of the dynamics of specific sites in any given RNA using absorbance or fluorescence detection.

Two important methods for resolving RNA folding in space and time have been applied to the Group I ribozyme from *Tetrahymena thermophila*. This RNA has played a pivotal role in the field of RNA catalysis, starting with the report by Cech and co-workers (Kruger et al., 1982) that *Tetrahymena* pre-rRNA is capable of excising its own intron. The ribozyme can be split into two domains, one of which, P4-P6, can fold independently, while the other, P3-P9, requires P4-P6 for formation of its native tertiary structure (Doherety and Doudna, 1997). Both the intrinsic interest of the system and the potential for in-depth analysis of the kinetic measurements has been greatly increased by solution of the crystal structure of the P4-P6 subunit of the ribozyme (Cate et al., 1996). Given the size and complex folding pattern of this RNA, it was necessary to develop new methods in order to analyze the kinetic properties at nucleotide resolution.

Sclavi et al. (1998) described the application to RNA folding of a rapid nucleic acid footprinting method that uses a synchrotron source to generate a burst of hydroxyl radicals by radiolysis of water. The X-ray source is combined with a quenched-flow device that allows folding to be triggered by mixing with buffer containing Mg²⁺ ions, followed by a variable time delay before application of the X-ray pulse. Local kinetic parameters can be determined for any nucleotide whose hydroxyl radical reactivity changes upon Mg²⁺ addition on a time scale substantially slower than the fastest resolvable time, i.e., ~10 msec.

Another important technical innovation, employed by Williamson and co-workers, is the use of deoxyoligonucleotide hybridization probes and RNase H cleavage to monitor accessibility of regions of the RNA sequence that are blocked by tertiary structure formation (Zarrinkar and Williamson, 1994; Treiber et al.,

1998). In this method the reaction is initiated by Mg²⁺ addition, followed after a variable time delay by addition of the deoxyoligonucleotide probe and RNase H. The kinetics of processes that occur on a time scale of ~1 min or longer can be resolved, allowing examination of tertiary structure formation. A general conclusion of these and other folding studies on the *Tetrahymena* ribozyme is that formation of the final conformation is slowed by the necessity for disrupting a competing structure.

CONFORMATIONAL SWITCHING

It is often proposed that RNA switches between two or more stable conformations in the course of its biological function. Study of such processes probably has more biological significance than in vitro examination of the rate of folding an intact transcript, which starts from very different boundary conditions than found in the cell, where folding events accompany RNA synthesis. In addition, cellular proteins are likely to affect the pathway and rate of folding reactions (Weeks and Cech, 1995).

Only a few systems of this kind have been intensively studied. For example, the kinetics of a conformational switch that involves two competing secondary structures in a spliced leader RNA was studied by LeCuyer and Crothers (1994). Using stopped-flow and T-jump methods, the authors evaluated the rate constants for the switch, which has a time constant on the order of 100 msec. The estimated activation energy was ~25 kcal/mol, considerably less than would be expected if the mechanism involved dissociation of the 13- to 14-bp of structure 1 before structure 2 begins to form. To explain the results, a branch migration mechanism was proposed, in which the alternate helix is nucleated and grows while the competing structure shrinks.

An example of an even slower conformational switch, whose detailed structural basis is not obvious but probably involves tertiary interactions, comes from the work of Gluick et al. (1997). They observed two electrophoretically distinguishable conformations of a 112-nt fragment containing the first ribosome initiation site of *E. coli* α mRNA. The corresponding relaxation time for the conformational switch is ~2000 sec (~30 min), or ~2 × 10⁴ fold slower than the spliced leader RNA switch.

The RNA conformational switch most intensively studied to date is the reorganization of the P1, or substrate (because it contains the splice site), helix, and its docking as part of the tertiary structure of the *Tetrahymena* ribozyme.

The early stopped-flow experiments of Turner and associates (Bevilacqua et al., 1992) used fluorescently labeled substrate oligonucleotides to form the P1 helix by hybridization with the IGS in *trans*. Two reaction steps could be resolved kinetically, which were interpreted as formation of the P1 helix followed by the slower docking reaction.

Woodson and colleagues adopted the strategy of providing the substrate in *cis*, forming pre-rRNA, and examined the ability of alternate substrate secondary structures to affect the rate of folding (Emerick and Woodson, 1994; Pan and Woodson, 1999). A fraction of the RNA usually renatures rapidly to the form active in splicing, but a variable portion of primary transcripts renatured at physiological temperatures is found in inactive states. The active form of pre-rRNA has the highest electrophoretic mobility, corresponding to a compact structure. The inactive form is found to have an alternate structure, called alt P3 instead of P3, in the catalytic core of the ribozyme. The presence of this non-native structure is coupled with formation of an alternate hairpin helix, P(-1), in the 5' exon (Pan and Woodson, 1999). The switch to the P1-P3 form from the P(-1)-alt P3 structure limits the rate of folding of the P3-P9 domain. The P(-1) structure is modeled as a 12-bp dumbbell helix, whereas P(1) is a 9-bp hairpin (Pan and Woodson, 1999). The activation energy for the folding reaction, ~12 kcal/mol, is much smaller than would be expected if the mechanism involved full melting of incorrect structure before initiating formation of the native form. In this case, as in the spliced leader RNA switch (LeCuyer and Crothers, 1994), it is likely that some form of branch migration pathway is involved, allowing coexistence of portions of each structure in the transition state.

An impressive degree of mechanistic detail at the molecular level has been achieved for *Tetrahymena* pre-rRNA and its corresponding ribozyme. However, it is also striking that a very significant effort by several laboratories was necessary to reach this level of understanding. The methods that have been developed will aid in attacking other RNA folding/switching problems, but present indications are that these systems are idiosyncratic, making each problem a new one and posing a major challenge to formulation of simplifying generalizations.

LITERATURE CITED

- Bevilacqua, P.C., Kierzek, R., Johnson, K.A., and Turner, D.H. 1992. Dynamics of ribozyme binding of substrate revealed by fluorescence-detected stopped-flow methods. *Science* 258:1355-1358.
- Breslauer, K.J., Frank R., Blocker, H., and Marky, L.A. 1986. Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad. Sci. U.S.A.* 83:3746-3750.
- Brion, P. and Westhof, E. 1997. Hierarchy and dynamics of RNA folding. *Annu. Rev. Biophys. Biomol. Struct.* 26:113-137.
- Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R., and Doudna, J.A. 1996. Crystal structure of a group I ribozyme domain: Principles of RNA packing. *Science* 273:1678-1685.
- Chen, J.L., Blasco, M.A., and Greider, C.W. 2000. Secondary structure of vertebrate telomerase RNA. *Cell* 100:503-514.
- Cole, P.E. and Crothers, D.M. 1972. Conformational changes of transfer ribonucleic acid. Relaxation kinetics of the early melting transition of methionine transfer ribonucleic acid (*Escherichia coli*). *Biochemistry* 11:4368-4374.
- Cole, P.E., Yang, S.K., and Crothers, D.M. 1972. Conformational changes of transfer ribonucleic acid. Equilibrium phase diagrams. *Biochemistry* 11:4358-4368.
- Crothers, D.M., Cole, P.E., Hilbers, C.W., and Shulman, R.G. 1974. The molecular mechanism of thermal unfolding of transfer RNA. *J. Mol. Biol.* 87:63-88.
- Doherty, E.A. and Doudna, J.A. 1997. The P4-P6 domain directs higher order folding of the *Tetrahymena* ribozyme core. *Biochemistry* 36:3159-3169.
- Draper, D.E. and Gluick, T.C. 1995. Melting studies of RNA unfolding and RNA-ligand interactions. *Methods Enzymol.* 259:281-305.
- Emerick, V.L. and Woodson, S.A. 1994. Fingerprinting the folding of a group I precursor RNA. *Proc. Natl. Acad. Sci. U.S.A.* 91:9675-9679.
- Farris, A.D., Koelsch, G., Pruijn, G.J., van Venrooij, W.J., and Harley, J.B. 1999. Conserved features of Y RNAs revealed by automated phylogenetic secondary structure analysis. *Nucl. Acids Res.* 27:1070-1078.
- Gluick, T.C., Gerstner, R.B., and Draper, D.E. 1997. Effects of Mg²⁺, K⁺, and H⁺ on an equilibrium between alternative conformations of an RNA pseudoknot. *J. Mol. Biol.* 270:451-463.
- Harris, M.E. and Pace, N.R. 1995-1996. Analysis of the tertiary structure of bacterial RNase P RNA. *Mol. Biol. Rep.* 122:115-123.
- Hilbers, C.W., Robillard, G.T., Shulman, R.G., Blake, R.D., Webb, P.K., Fresco, R., and Riesner, D. 1976. Thermal unfolding of yeast glycine transfer RNA. *Biochemistry* 15:1874-1882.

- Kruger, K., Grabowski, P.J., Zaug, A.J., Sands J., Gottschling, D.E., and Cech, T.R. 1982. Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* 32:147-157.
- LeCuyer, K.A. and Crothers, D.M. 1994. Kinetics of an RNA conformational switch. *Proc. Natl. Acad. Sci. U.S.A.* 91:3373-3377.
- Maglott, E.J. and Glick, G.D. 1997. A new method to monitor the rate of conformational transitions in RNA. *Nucl. Acids Res.* 25:3297-3301.
- Matthews, D.H., Sabina, J., Zucker, M., and Turner, D.H. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* 288:911-940.
- Noller, H.F., Kop, J., Wheaton, V., Brosius, J., Gutell, R.R., Kopylov, A.M., Dohme, F., Herr, W., Stahl, D.A., Gupta, R., and Waese, C.R. 1981. Secondary structure model for 23S ribosomal RNA. *Nucl. Acids Res.* 9:6167-6189.
- Pan, J. and Woodson, S.A. 1999. The effect of long-range loop-loop interactions on folding of the *Tetrahymena* self-splicing RNA. *J. Mol. Biol.* 294:955-965.
- Riesner, D., Romer, R., and Maass, G. 1969. Thermodynamic properties of the three conformational transitions of alanine specific transfer RNA from yeast. *Biochem. Biophys. Res. Commun.* 35:369-376.
- Riesner, D., Romer, R., and Maass, G. 1970. Kinetic study of the three conformational transitions of alanine specific transfer RNA from yeast. *Eur. J. Biochem.* 15:85-91.
- Riesner, D., Maass, G., Thiebe, R., Philippsen, P., and Zachau, H.G. 1973. The conformational transitions in yeast tRNA^{Phe} as studied with tRNA^{Phe} fragments. *Eur. J. Biochem.* 36:76-88.
- Sclavi, B., Sullivan, M., Chance, M.R., Brenowitz, M., and Woodson, S.A. 1998. RNA folding at millisecond intervals by synchrotron hydroxyl radical footprinting. *Science* 279:1940-1943.
- Tinoco, I., Jr. and Bustamante, C. 1999. How RNA folds. *J. Mol. Biol.* 293:271-281.
- Treiber, D.K., Rook, M.S., Zarrinkar, P.P., and Williamson, J.R. 1998. Kinetic intermediates trapped by native interactions in RNA folding. *Science* 279:1943-1946.
- Weeks, K.M. and Cech, T.R. 1995. Protein facilitation of group I intron splicing by assembly of the catalytic core and the 5' splice site domain. *Cell* 82:221-230.
- Wu, M. and Tinoco, I., Jr. 1998. RNA folding causes secondary structure rearrangement. *Proc. Natl. Acad. Sci. U.S.A.* 95:11555-11560.
- Xia, T., SantaLucia, J., Jr., Burkard, M.E., Kierzek, R., Schroeder, S.J., Jiao, X., Cox, C., and Turner, D.H. 1998. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry* 37:14719-14735.
- Zarrinkar, P.P. and Williamson, J.R. 1994. Kinetic intermediates in RNA folding. *Science* 265:918-924.

Contributed by Donald M. Crothers
Yale University
New Haven, Connecticut