

# Kinetic Intermediates in RNA Folding

Patrick P. Zarrinkar and James R. Williamson\*

The folding pathways of large, highly structured RNA molecules are largely unexplored. Insight into both the kinetics of folding and the presence of intermediates was provided in a study of the  $Mg^{2+}$ -induced folding of the *Tetrahymena* ribozyme by hybridization of complementary oligodeoxynucleotide probes. This RNA folds via a complex mechanism involving both  $Mg^{2+}$ -dependent and  $Mg^{2+}$ -independent steps. A hierarchical model for the folding pathway is proposed in which formation of one helical domain (P4-P6) precedes that of a second helical domain (P3-P7). The overall rate-limiting step is formation of P3-P7, and takes place with an observed rate constant of  $0.72 \pm 0.14 \text{ minute}^{-1}$ . The folding mechanism of large RNAs appears similar to that of many multidomain proteins in that formation of independently stable substructures precedes their association into the final conformation.

A complete understanding of RNA folding requires the determination of both the final structures adopted by functional RNAs and the processes of their formation. Given the central role of RNA in cellular functions such as splicing and translation (1), it is important to understand the mechanism by which RNA can form complex three-dimensional structures capable of ligand binding and catalysis. Thus it is necessary to consider how a linear polyanionic chain of nucleotides folds into a compact, stable conformation.

Good models for the secondary and tertiary structure of RNAs of interest can be constructed from a combination of phylogenetic and biochemical data (2-4). Improvements in both crystallography and nuclear magnetic resonance methods continue to provide more detailed structural information (5, 6). In contrast, little is known about either the kinetics of folding or the identity of intermediates on the folding pathways of large RNAs.

We studied the mechanism of RNA folding by taking advantage of sequence-specific oligodeoxynucleotide hybridization to monitor the kinetics of  $Mg^{2+}$ -induced folding of the *Tetrahymena* ribozyme. Formation of the active structure of this complex RNA occurs by way of multiple intermediates in a process involving both  $Mg^{2+}$ -dependent and  $Mg^{2+}$ -independent steps. The rate limiting step, which takes place on a time scale of minutes, is formation of the P3-P7 domain. The data support a model for the folding pathway in which short-range secondary structure formation and subsequent stabilization of the P4-P6 heli-

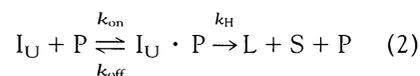
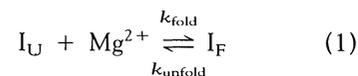
cal domain precede P3-P7 formation. The hierarchical mechanism presented, where folding of domains occurs prior to slow formation of the final structure, may illustrate the general features observed in the folding of large RNAs both in vitro and in vivo.

**RNA folding monitored by oligonucleotide hybridization.** The stability of the native conformation of large, highly structured RNA molecules generally depends on divalent metal ions, usually  $Mg^{2+}$  (7, 8). In the absence of divalent ions, many RNAs exist in a partially denatured state where some local secondary structure elements, such as hairpins, are stable, but many tertiary interactions are not. The addition of  $Mg^{2+}$  can thus be used as a means to initiate formation of the native structure.

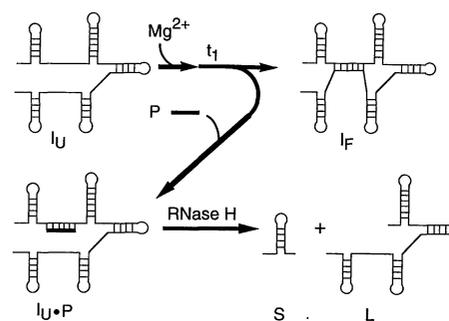
When RNA is subjected to  $Mg^{2+}$ -induced folding, parts of its sequence previously accessible become inaccessible to binding by complementary DNA oligonucleotides. The hybridization of such probes to structured RNAs under various equilibrium conditions has already been examined (9-12). In order to obtain information about the rates at which conformational rearrangements take place, we measured the time dependence of changes in the accessibility to specific oligonucleotide probes.

**Fig. 1.** Schematic representation of the kinetic oligonucleotide hybridization experiment. Folding is induced by addition of  $Mg^{2+}$ , and after time  $t_1$  the fraction of RNA still unfolded ( $I_U$ ) is determined by rapid binding to an oligonucleotide probe (P, shown in bold), followed by RNase H cleavage. S and L are the short and long fragments produced by RNase H cleavage of the probe-RNA complex ( $I_U \cdot P$ ). Folded RNA ( $I_F$ ) is not cleaved. The fraction of observed cleavage corresponds to the fraction  $I_U$  left at time  $t_1$ .

Folding was initiated by the addition of  $Mg^{2+}$  to  $^{32}P$ -labeled RNA equilibrated in tris buffer. Samples from the folding reaction were added to an excess of oligonucleotide probe and ribonuclease (RNase) H, setting up a kinetic competition between continued folding and probe binding (Fig. 1). Under the appropriate conditions, the fraction of the RNA that remained unfolded ( $I_U$ ) was rapidly bound by the probe (P) and cleaved by RNase H at the site of hybridization (resulting in fragments L and S). RNA that had already folded ( $I_F$ ) was thus inaccessible and would remain uncleaved. Electrophoretic separation of cleaved and uncleaved RNA allowed quantitation of the fraction cleaved, and thus the fraction unfolded as a function of time. Such an experiment is described by the kinetic mechanism shown in Eqs. 1 and 2 below, where each  $k$  denotes the first-order rate constant for an individual step.



The most revealing application of oligonucleotide hybridization and subsequent RNase H cleavage was a rapid quench experiment to measure the fraction of total RNA that remains unfolded at any given time after addition of  $Mg^{2+}$ . To obtain such a snapshot of progress toward the folded state, three kinetic conditions must be met. First, the rates of probe hybridization and RNase H cleavage must be rapid compared to the folding rate ( $k_{\text{on}}[P], k_{\text{H}} \gg k_{\text{fold}}$ ). Second, RNase H cleavage must be much faster than dissociation of the RNA-DNA hybrid ( $k_{\text{H}} \gg k_{\text{off}}$ ). Third, unfolding must be slow compared to folding ( $k_{\text{unfold}} < k_{\text{fold}}$ ). Conditions were adjusted so that all three requirements were satisfied for each probe used in kinetic experiments (13). In principle, if oligonucleotides complementary to different regions are used, folding of each part of the RNA can be probed specifically and independently. In practice, only certain sequences are amenable to this analysis (14).



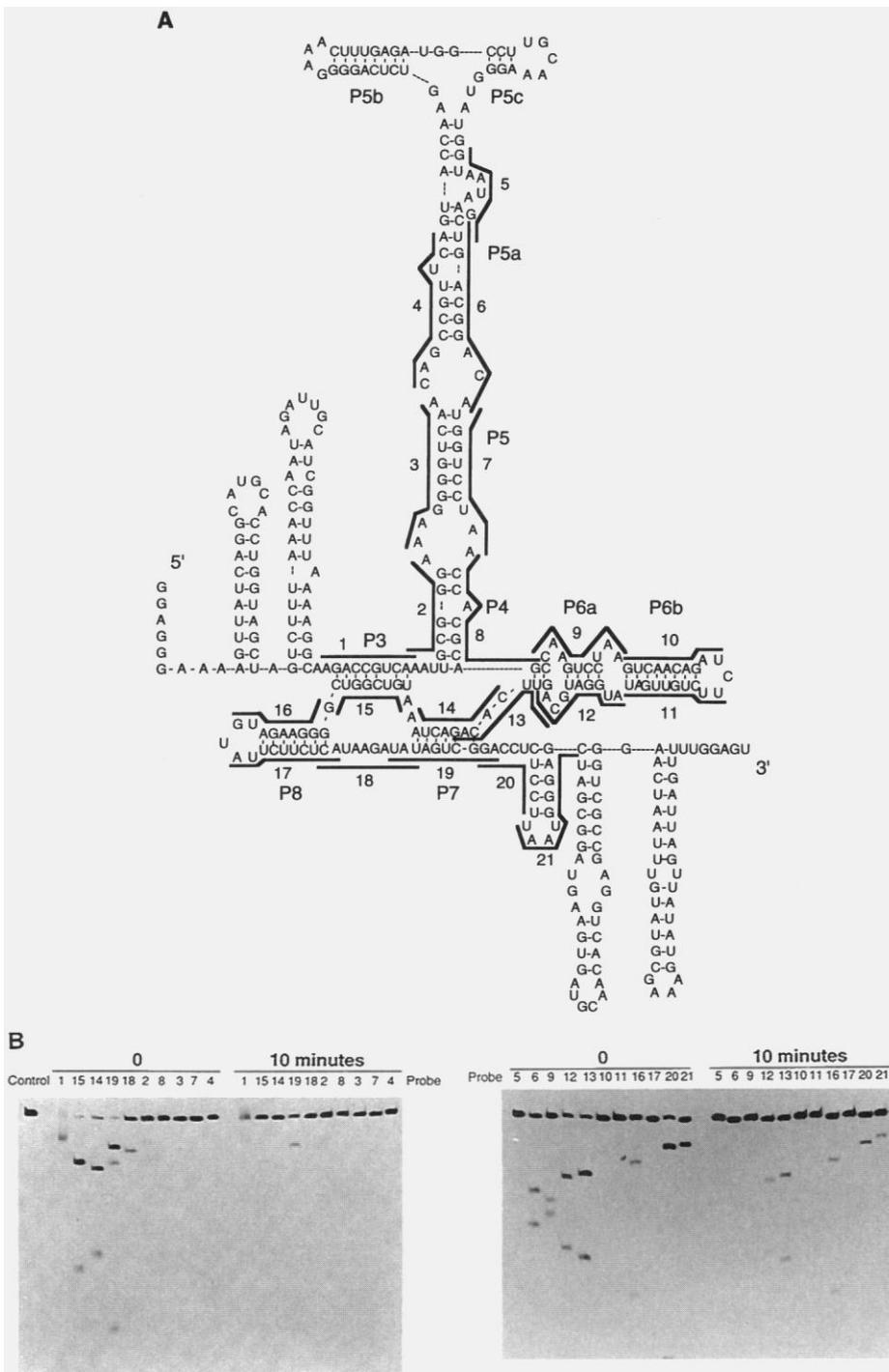
The authors are in the Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

\*To whom correspondence should be addressed.

The *Tetrahymena* L-21 Sca I ribozyme was chosen as a model system for studying RNA folding. This catalytic RNA, a shortened version of a group I intron from *Tetrahymena* pre-ribosomal RNA, acts as an endoribonuclease on small oligoribonucleotide substrates in *trans* with multiple turnover (15). It is one of the most studied RNAs, moderately large, and highly structured (16). Its secondary structure is known, and a good model exists for the three-dimensional structure of the conserved core region (2). Furthermore, it is an RNA enzyme with a well-defined activity (17, 18) and thus allows functional analysis. Both the native structure and function of this RNA require  $Mg^{2+}$ , and it has been observed that in order to obtain fully active ribozyme, a 10-minute equilibration in  $Mg^{2+}$  is required before the reaction is initiated (17). Previous equilibrium studies of  $Mg^{2+}$ -induced folding have suggested the presence of distinct folding domains (19, 20). We have extended these findings by focusing on the kinetics of specific folding events.

First, the regions of the L-21 Sca I RNA suitable for kinetic analysis were determined. We screened 21 oligonucleotides complementary to different sequences within the ribozyme to survey most of the conserved core (Fig. 2A). Each oligonucleotide was tested for its ability to hybridize, in competition with structure formation, when added at high concentration simultaneously with  $Mg^{2+}$ . In a second experiment, we tested each probe for binding after the RNA had been equilibrated in  $Mg^{2+}$  for 10 minutes. These two extremes represent the initial and final points of a folding time course.

The probes can be grouped into three classes (Fig. 2B). The first class gave no observable RNase H cleavage when added with, or after,  $Mg^{2+}$ . Members of this class generally target secondary structures that would be expected to form in the absence of  $Mg^{2+}$ , and the lack of cleavage was therefore not surprising (21). The second class competed successfully with structure formation when added simultaneously with  $Mg^{2+}$ , resulting in almost complete cleavage of the RNA, but gave almost no cleavage after equilibration in  $Mg^{2+}$ . This second class included probes targeting sequences involved in long-range interactions, such as the P3 and P7 helices, that might be expected to require  $Mg^{2+}$  for their stabilization if not their formation. The third class of probes gave incomplete cleavage at zero time, and significant remaining cleavage at 10 minutes. The observation of these three different classes of probe behavior suggests that there are distinguishable folding events that can be addressed by oligonucleotide hybridization.



**Fig. 2.** Initial screening of oligodeoxynucleotide probes complementary to the *Tetrahymena* group I intron. **(A)** Secondary structure of the L-21 Sca I ribozyme. The target sequence for each oligonucleotide is outlined in bold. Probes were numbered according to their target sequence, starting from the 5' end of the RNA. **(B)** Autoradiograph showing the results of the screening. In all experiments, RNA was annealed by heating to 95°C for 1 minute, and then equilibrated at 37°C for 3 minutes. For the zero time point, uniformly  $^{32}P$ -labeled L-21 Sca I RNA (58) in 10  $\mu$ l of buffer containing 1 mM tris (pH 8.1) and 0.1 mM EDTA (0.1  $\times$  tris-EDTA) was annealed and then added directly to an equal volume of 2 $\times$  folding buffer [1 $\times$ :50 mM tris (pH 8.1), 10 mM  $MgCl_2$ , 10 mM NaCl, 0.1 mM EDTA, 1 mM dithiothreitol] containing oligonucleotide probe (final concentration  $\sim$ 300  $\mu$ M) and RNase H (United States Biochemical, final concentration 0.1 U/ $\mu$ l) that had been equilibrated at 37°C. The final volume was 20  $\mu$ l, and the final RNA concentration was 1 nM. After cleavage for 30 seconds at 37°C, the reaction was quenched with 14  $\mu$ l of stop solution containing 90 mM EDTA and gel-running dyes in 82 percent formamide. For the 10-minute time point, L-21 Sca I in 5  $\mu$ l of 0.1  $\times$  tris-EDTA was annealed, then added to an equal volume of 2 $\times$  folding buffer. After the RNA had been allowed to fold at 37°C for 10 minutes, 10  $\mu$ l of 37°C equilibrated, 1 $\times$  folding buffer containing probe and RNase H were added and RNase H cleavage was performed as described. Products were separated on 6 percent denaturing polyacrylamide gels.

**Kinetic measurements of *Tetrahymena* ribozyme folding.** A detailed analysis was undertaken to explore the kinetics of formation of the P3 helix. P3 is part of a pseudo-knot constituting one domain of the core of the intron and is essential for ribozyme activity (22). It is a long-range base-pairing interaction, with its two strands far apart in the linear sequence of the RNA, and we therefore anticipated that it might exhibit interesting folding behavior. In the initial screening, oligonucleotides complementary to both strands of this helix (probes 1 and 15) showed the second class of behavior (Fig. 2B), with accessibility to probe binding being  $Mg^{2+}$ -dependent.

After  $Mg^{2+}$  was added to RNA equilibrated in tris buffer, the accessibility of P3 to probe 15 decreased with time (Fig. 3A). The observed decay is apparently first order (Fig. 3B), is dependent on the presence of  $Mg^{2+}$  (23), and is independent of the concentrations of RNA, oligonucleotide probe,

RNase H (13),  $Na^+$  (24), and the time allowed for RNase H cleavage. The rate of decay therefore appears to reflect the rate limiting step for P3 folding and becoming inaccessible to probe binding or RNase H cleavage (25). Therefore, there is at least one slow step on the folding pathway of the *Tetrahymena* ribozyme which occurs on the minute time scale ( $k_{obs} = 0.72 \pm 0.14 \text{ min}^{-1}$ ) (26).

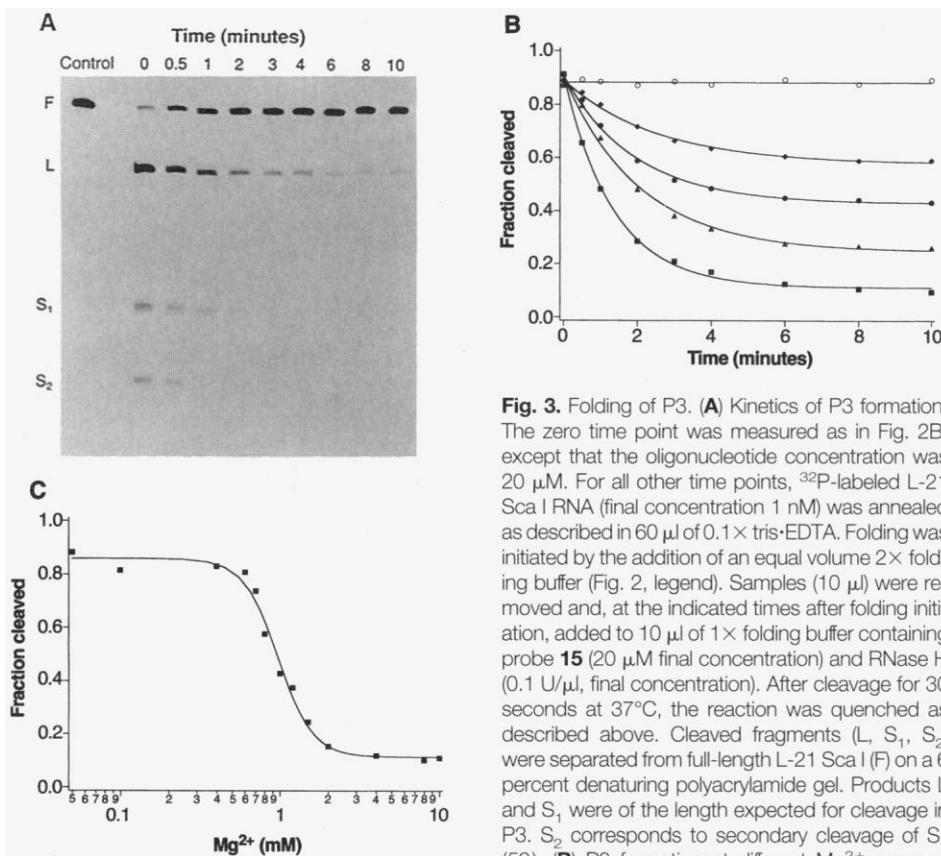
The simplest mechanism that could account for the  $Mg^{2+}$  requirement is a two-state transition involving the binding of  $Mg^{2+}$  upon folding as shown in Eq. 1. If such a one-step mechanism is correct, then the observed rate of folding should be dependent on  $Mg^{2+}$  concentration. To test this possibility, we measured the rate at which P3 becomes inaccessible at different  $Mg^{2+}$  concentrations (Fig. 3B) (27). We observed that at lower concentrations of  $Mg^{2+}$  the extent of cleavage once equilibrium had been reached increased, consistent with a shift in the equilibrium from the

folded to the unfolded conformation. The midpoint of this transition ( $[Mg^{2+}]_{1/2}$ ) occurred when the  $Mg^{2+}$  concentration was 0.97 mM (Fig. 3C). A similar transition was seen when the  $Mg^{2+}$  dependence of ribozyme activity, global structure formation, or active site formation was examined (19, 28), an indication that our observations correlate closely with results obtained by independent methods. The rate of approach to equilibrium, however, was found to be independent of  $Mg^{2+}$  concentration. Apparently, the formation of P3 occurs by way of a more complex mechanism where rapid  $Mg^{2+}$  binding follows a slow,  $Mg^{2+}$ -independent step. Thus, there is a transient intermediate on the folding pathway, and the rate limiting step is  $Mg^{2+}$ -independent.

To compare the rate of formation of P3 with the rate at which the active structure is formed, we used the well-characterized endonuclease activity of this RNA (17). The amount of active ribozyme can be determined from the magnitude of the kinetic burst corresponding to the first turnover in a multiple turnover reaction (17). The relative amount of active ribozyme present was measured as a function of increasing folding times after the addition of  $Mg^{2+}$  (Fig. 4). Full activity was obtained at a rate ( $k_{obs} = 0.61 \text{ min}^{-1}$ ) comparable to that of P3 becoming inaccessible to probe binding as determined by oligonucleotide hybridization (29). We therefore conclude that P3 formation is the rate-limiting step in formation of the active structure.

To obtain information about the folding kinetics of various regions of the RNA, we examined probes targeting other sequences. P7 is the second stem of the central pseudo-knot in the conserved core of group I introns. Together with P3, P7 forms part of one of several stacked helical domains in the three-dimensional structure of the *Tetrahymena* ribozyme (2). In the initial screen, probes complementary to both strands of P7 (probes 14 and 19) fell into the second class (Fig. 2B). In all experiments defining the kinetic mechanism of P3 formation, the probes for P7 gave identical results. P3 and P7 are thus kinetically indistinguishable. Although our experiments cannot yet determine whether they form cooperatively, sequentially, or independently, the P3 and P7 helices clearly become inaccessible to probe binding and RNase H cleavage on the same time scale.

Two further tests of the kinetic mechanism of rapid  $Mg^{2+}$  binding after the rate limiting,  $Mg^{2+}$ -independent step were performed with probes targeting P3 and P7. First, unfolding was monitored after dilution of the  $Mg^{2+}$ -equilibrated sample, and both reverse steps were fast. Second, refolding was initiated after unfolding by addition of  $Mg^{2+}$ , and the refolding occurred at the



**Fig. 3.** Folding of P3. (A) Kinetics of P3 formation. The zero time point was measured as in Fig. 2B, except that the oligonucleotide concentration was 20  $\mu\text{M}$ . For all other time points,  $^{32}\text{P}$ -labeled L-21 Sca I RNA (final concentration 1 nM) was annealed as described in 60  $\mu\text{l}$  of 0.1  $\times$  tris-EDTA. Folding was initiated by the addition of an equal volume 2 $\times$  folding buffer (Fig. 2, legend). Samples (10  $\mu\text{l}$ ) were removed and, at the indicated times after folding initiation, added to 10  $\mu\text{l}$  of 1 $\times$  folding buffer containing probe 15 (20  $\mu\text{M}$  final concentration) and RNase H (0.1 U/ $\mu\text{l}$ , final concentration). After cleavage for 30 seconds at 37 $^{\circ}\text{C}$ , the reaction was quenched as described above. Cleaved fragments (L, S<sub>1</sub>, S<sub>2</sub>) were separated from full-length L-21 Sca I (F) on a 6 percent denaturing polyacrylamide gel. Products L and S<sub>1</sub> were of the length expected for cleavage in P3. S<sub>2</sub> corresponds to secondary cleavage of S<sub>1</sub> (59). (B) P3 formation at different  $Mg^{2+}$  concentrations. (C) Thermodynamic dependence of P3 formation on  $Mg^{2+}$  concentration. Kinetic experiments as in (B) were performed for each  $[Mg^{2+}]$  shown. The fraction cleaved at equilibrium for each concentration was determined by a fit of the kinetic data to a first order exponential. The midpoint of the transition ( $[Mg^{2+}]_{1/2}$ ) was obtained by a fit of these equilibrium values to an expression of simple two state binding of  $x$   $Mg^{2+}$  ions [ $f = 1/([Mg^{2+}]/[Mg^{2+}]_{1/2} + 1)$ ], where  $f$  is the fraction cleaved at equilibrium, and  $x$  the number of  $Mg^{2+}$  bound per RNA molecule). The fit gave values of  $[Mg^{2+}]_{1/2} = 0.97 \text{ mM}$  and  $x = 3.97$ , suggesting that approximately four magnesium ions are involved in the transition.

concentrations ( $[Mg^{2+}]$ ). Experiments were performed as described in (A), except that  $[Mg^{2+}]$  in 1 $\times$  folding buffer was varied: (■) 10 mM, (▲) 1.5 mM, (●) 1.0 mM, (◆) 0.8 mM, (○) 0.05 mM. The final  $[Mg^{2+}]$  during RNase H cleavage was always adjusted to 10 mM. Results were quantitated with a Molecular Dynamics PhosphorImager. (C) Thermodynamic dependence of P3 formation on  $Mg^{2+}$  concentration. Kinetic experiments as in (B) were performed for each  $[Mg^{2+}]$  shown. The fraction cleaved at equilibrium for each concentration was determined by a fit of the kinetic data to a first order exponential. The midpoint of the transition ( $[Mg^{2+}]_{1/2}$ ) was obtained by a fit of these equilibrium values to an expression of simple two state binding of  $x$   $Mg^{2+}$  ions [ $f = 1/([Mg^{2+}]/[Mg^{2+}]_{1/2} + 1)$ ], where  $f$  is the fraction cleaved at equilibrium, and  $x$  the number of  $Mg^{2+}$  bound per RNA molecule). The fit gave values of  $[Mg^{2+}]_{1/2} = 0.97 \text{ mM}$  and  $x = 3.97$ , suggesting that approximately four magnesium ions are involved in the transition.

same rate as  $Mg^{2+}$ -initiated folding. These two observations further support the proposed mechanism of P3-P7 formation.

A second helical domain in the model of the three-dimensional structure of the intron is formed by the P4, P5, and P6 stems (2). This domain is independently stable (30, 31). When oligonucleotides complementary to sequences in P4 (probe 8) and P6a (probe 12) were added to the RNA simultaneously with  $Mg^{2+}$ , no more than 40 to 50 percent cleavage was observed. These probes therefore are members of the third class (Fig. 2B). If the limited cleavage were due to similar rates for folding and probe binding, it should have been possible to increase the extent of cleavage by increasing the probe concentration. At higher concentrations of either probe, however, the extent of cleavage remained unchanged. Alternatively, the population of unfolded molecules may be heterogeneous, with only a subpopulation able to bind oligonucleotides.

When either of the probes for P4 or P6a was added to the RNA before  $Mg^{2+}$ , cleavage increased to almost 80 percent. In a kinetic experiment a slow folding phase appeared to be preceded by a rapid burst (Fig. 5), thus providing evidence that an equilibrium between at least two conformations exists in the absence of  $Mg^{2+}$ . One of these is inaccessible to binding by 8 and 12 or RNase H cleavage and can fold rapidly

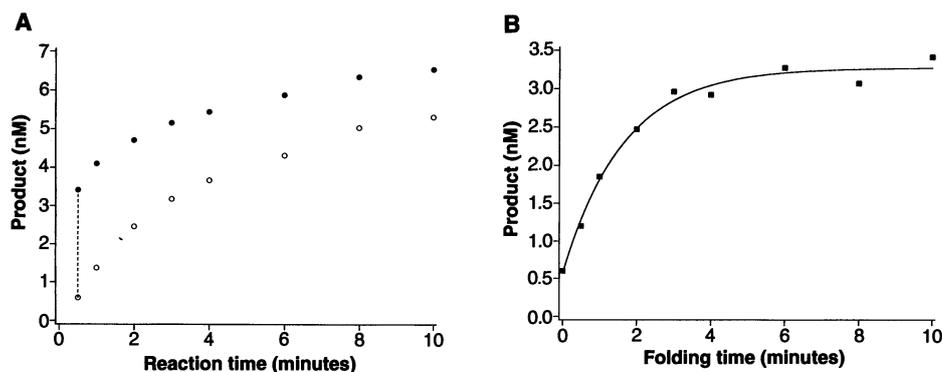
once  $Mg^{2+}$  is added, while the second conformation is accessible to the probes and must go through a slow step in order to fold. In the absence of bound  $Mg^{2+}$ , hybridization of oligonucleotide can drive the equilibrium toward the accessible conformation to give increased cleavage. When we attempted to improve the resolution of the kinetic phases by folding the RNA at a lower temperature, we found that the kinetic behavior of P4 and P6a at 37° and 27°C was indistinguishable by kinetic oligonucleotide hybridization. The apparent temperature independence of P4 and P6a formation contrasts with P3 and P7, whose rates of formation were very sensitive to temperature [activation energy ~20 kcal/mol (32)] (Fig. 5). Thus, these two helical domains of the *Tetrahymena* intron clearly show different kinetic behavior during  $Mg^{2+}$ -induced folding, with the P4-P6 domain apparently forming faster than P3-P7.

**The folding pathway of the *Tetrahymena* ribozyme.** A model for the folding pathway of this RNA includes several intermediates and involves the sequential formation and stabilization of helical domains (Fig. 6). The first intermediate,  $I_U$ , consists of short-range secondary structures, such as hairpins, that form in the absence of  $Mg^{2+}$ . Experiments with probes targeting P4 and P6a show that  $I_U$  is in a slow equilibrium with another conformation ( $I_1$ ). At present, it is not possible to distinguish whether  $I_1$  is

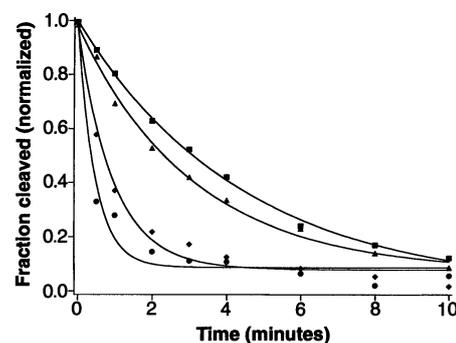
a productive intermediate, or is on a non-productive branch of the folding pathway. Rapid  $Mg^{2+}$  binding by either  $I_U$  or  $I_1$  leads to formation of  $I_2$  and is reflected by the rapid burst observed in the kinetic experiments probing P4 and P6a.  $I_2$  corresponds to an intermediate in which P4 and P6 are inaccessible to the probes, while P3 and P7 are still accessible.

$I_3$  is the transient intermediate inferred from the  $Mg^{2+}$  concentration independence of the rate of P3 and P7 formation. Only the presence of such an intermediate in the mechanism of P3-P7 formation can explain the requirement for  $Mg^{2+}$ , since the rate limiting step itself does not involve metal binding. The species  $I_3$  may be one in which P3 and P7 are formed, but are not yet stable or in the correct orientation relative to P4-P6. Further binding of  $Mg^{2+}$  may allow the two helical stacks to come together in a conformation in which P3 and P7, as well as P4 and P6, are protected ( $I_F$ ). Additional fast rearrangements may lead to the final structure (N), but since the rate at which full activity is gained is the same as the rate at which P3 and P7 become inaccessible, formation of the P3-P7 domain appears to be the overall rate limiting step on the folding pathway.

The above folding mechanism (Fig. 6) has three implications. First,  $Mg^{2+}$  binding to  $I_3$  must in fact be significantly stronger than suggested by the transition midpoint ( $[Mg^{2+}]_{1/2}$ ) observed in the experiments with probes targeting P3 and P7, as well as in previous studies of the activity and structure of the *Tetrahymena* ribozyme (19, 28). In order to overcome the unfavorable equilibrium between  $I_2$  and  $I_3$ , high concentrations of  $Mg^{2+}$  are required, even though  $I_3$  binds  $Mg^{2+}$  tightly. From our experimental data, we can estimate an approximate upper



**Fig. 4.** Ribozyme endonuclease assay for folding. Reaction conditions were kept as close as possible to those used for probe hybridization experiments. L-21 Sca I (final concentration 7 nM) (60) was annealed in 30  $\mu$ l of 0.1  $\times$  tris-EDTA, as described, and an equal volume of 2 $\times$  folding buffer was added. The RNA was allowed to fold at 37°C for the specified amount of time before the reaction was initiated by addition of 5'  $^{32}$ P-labeled CCCUCUAAAAA substrate (final concentration 16 nM) (61) in 60  $\mu$ l of 1 $\times$  folding buffer containing GTP (final concentration 0.5 mM). Samples (10  $\mu$ l) were taken and added to 8  $\mu$ l stop solution. For the zero time point, substrate and GTP were added together with folding buffer immediately after annealing. Products were separated on 20 percent denaturing polyacrylamide gels, and the results were quantitated as above. (A) The time course of substrate cleavage for folding times of 0 minutes and 10 minutes in  $Mg^{2+}$  (●) before the addition of substrate addition. When the RNA was first incubated for 10 minutes in  $Mg^{2+}$ , a rapid kinetic burst was followed by a slower multiple turnover rate. Without prior incubation, the size of the burst was significantly reduced, indicating that initially little active ribozyme was present. The largest difference in the amount of product formed with and without prior incubation of the RNA in  $Mg^{2+}$  before initiation of the reaction was found after 30 seconds of substrate cleavage (indicated by the dashed line). (B) Extent of substrate cleavage after 30 seconds as a function of the time allowed for the ribozyme to fold before initiation of the reaction. The rate of formation of the active structure was determined by a fit of the 30-second time points to a single exponential ( $k_{obs} = 0.61 \text{ min}^{-1}$ ). The same result was obtained when the folding time course was measured with the 1 minute reaction time points.



**Fig. 5.** Folding kinetics of P3, P7, P4, and P6a at 27°C. Reactions were performed as described for Fig. 3A, except that folding was done at 27°C, followed by probe binding and RNase H cleavage at 37°C. The data for each probe was normalized to allow a direct comparison, and fit to a single exponential (62). Probes and concentrations used: (●) 12 (60  $\mu$ M), (◆) 8 (60  $\mu$ M), (▲) 19 (60  $\mu$ M), (■) 15 (20  $\mu$ M).

limit of 100  $\mu\text{M}$  for the midpoint of the transition of  $\text{Mg}^{2+}$  binding to  $I_3$  (33). This value is consistent with the magnitude of dissociation constants measured for high affinity  $\text{Mg}^{2+}$  binding sites in transfer RNA (tRNA) (7, 34). Second, the folding of this RNA involves several discrete steps, reflected in the inferred kinetic intermediates. Third, and most fundamental, the folding pathway includes a slow step, and this rate limiting step itself does not involve binding of  $\text{Mg}^{2+}$ .

These results support and expand upon previous investigations of RNA folding. Studies on tRNA by Crothers and others have demonstrated the existence of folding intermediates (35–37). Lynch and Schimmel, in a fluorescence study (38), have shown that the rate limiting step for  $\text{Mg}^{2+}$ -induced folding of tRNA, as with the *Tetrahymena* intron, does not involve  $\text{Mg}^{2+}$  binding. In contrast to our findings for the ribozyme, however, in tRNA  $\text{Mg}^{2+}$  binding precedes the slow,  $\text{Mg}^{2+}$ -independent, steps. For both group I introns and tRNA, secondary structure generally precedes formation of tertiary interactions, which can be quite slow (36, 39, 40). The rate of helix formation in nucleic acids is known to be extremely rapid ( $k = 10^7$  to  $10^9 \text{ M}^{-1} \text{ min}^{-1}$ ) (41, 42). Short-range helices, and thus the formation of  $I_U$  from the denatured RNA, are expected to form at such fast rates.

The presence of a slow step on the folding pathway of the *Tetrahymena* ribozyme is consistent with the reported observation that prior incubation in  $\text{Mg}^{2+}$  is necessary in order to obtain fully active ribozyme (17). Studies on the thermodynamics of  $\text{Mg}^{2+}$ -induced folding of this RNA (19, 20) have provided evidence for two equilibrium folding intermediates; one in which a subdomain involving P5abc is formed, and one in which only the P4-P6 domain, but not P3-P7, is formed. These observations, together with the demonstration that the P4-P6 domain is stable independently, but only

in the presence of  $\text{Mg}^{2+}$  (30, 31), support the plausibility of the proposed kinetic intermediate  $I_2$ . In these studies it was shown that the P5abc subdomain facilitates the formation of the P4-P6 domain, and that the P4-P6 domain stabilizes the catalytic core, including P3-P7. From this we can infer that the P5abc intermediate should precede  $I_2$  in our mechanism (43). The localization of  $\text{Mg}^{2+}$  interactions to sites between the two helical stacks formed by P4-P6 and P3-P7 (44) may explain the  $\text{Mg}^{2+}$  requirement for association of these domains. The general mechanism of helix formation followed by docking into the correct orientation follows a precedent set by kinetic and thermodynamic studies of an oligonucleotide substrate binding to the intron, where hybridization to the internal guide sequence precedes movement of the resulting P1 helix into the active site (45–47).

This model provides a conceptual framework for studying kinetic RNA folding intermediates. The active structures of large RNA molecules appear to consist of stacks of helices forming domains and to be stabilized by interdomain tertiary contacts. It has been suggested that these structures are formed in a hierarchical process, with individual helices forming first and providing nucleation sites for formation of the stacked domains, which then come together in the final structure (16, 30, 48). We have provided direct kinetic evidence for such a hierarchical folding pathway in the *Tetrahymena* ribozyme. Furthermore, the kinetic intermediates we observe directly parallel previously identified equilibrium folding intermediates. The intermediates we observe suggest that the folding domains generally form in two steps. Base-pairing interactions form first, with subsequent formation of tertiary contacts and stabilization through specific binding of  $\text{Mg}^{2+}$ .

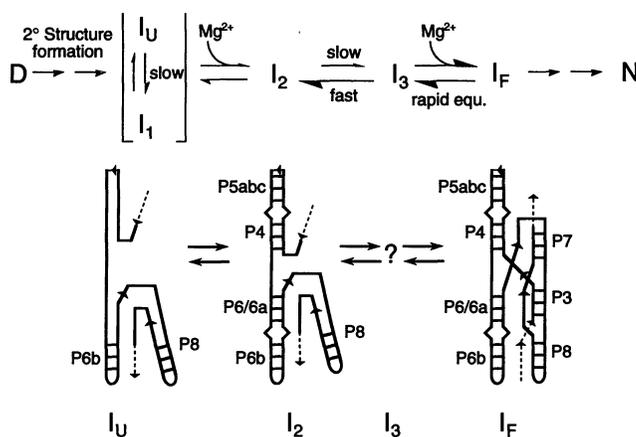
In the presence of denaturants such as urea, proteins are in an unfolded state. Formation of the folded structure is induced by

reducing the denaturant concentration and can be followed either thermodynamically at different levels of denaturation, or kinetically by a “concentration jump” experiment. Analogously, many RNAs are in a denatured state in the absence of  $\text{Mg}^{2+}$ , and the addition of a divalent metal is therefore equivalent to dilution of a denaturant, allowing application of the same experimental approaches used to study protein folding. Conceptually, the correlation of kinetic information with sequence by targeted hybridization most resembles amide  $^1\text{H}$ - $^2\text{H}$  exchange experiments (49). For both, a snapshot of the folding reaction is obtained by a rapid quench experiment. The state of accessibility to external probes is determined in a sequence-specific way, yielding information not only on the rates of folding events, but also on which regions of the molecule are involved in a particular step.

The presence of independently folding domains in RNA has clear parallels in the folding of multidomain proteins, where formation of the individual domains in general precedes their association into the final structure (50). In proteins, most of the structural stability is derived from tertiary interactions and the formation of a hydrophobic core (51). Secondary structures are weak, and are usually only stable in the context of the whole protein. In contrast, RNA is mostly stabilized by helical secondary structures, while tertiary interactions are weak and often require  $\text{Mg}^{2+}$  for stabilization (16, 52). For proteins the smallest independently folding unit generally corresponds to small globular domains stabilized by tertiary interactions. For RNA, the smallest folding units are groups of helical stems stabilized by base-pairing and stacking. Thus, for both proteins and RNA the smallest thermodynamically stable structures also appear to be kinetically independent folding units.

We have presented a model for the folding pathway of the *Tetrahymena* ribozyme. Folding proceeds by way of several intermediates, with the P4-P6 domain being formed before, and potentially required for, formation of the P3-P7 domain. The rate limiting step is formation of P3-P7, and in itself does not involve binding of  $\text{Mg}^{2+}$ . Although only a part of the overall folding process could be examined, it is clear that folding is not a simple two-state transition, but takes place by a complicated mechanism that includes both  $\text{Mg}^{2+}$ -dependent and  $\text{Mg}^{2+}$ -independent steps. Since the ability of proteins to affect the stability of RNA structures is well documented (53–55), it may be that RNA folding in vivo, like that of proteins, is assisted by additional factors. The rapid transcription rates of RNA polymerases (56) would result in the synthesis of an RNA the size of a group I intron in a

**Fig. 6.** Model for the folding pathway of the *Tetrahymena* group I intron.



matter of seconds. This may affect the formation of higher order structure (57), which can be much slower, and probably takes place post-transcriptionally rather than during transcription. Therefore, the intermediates, rates and pathways presented above may be generally relevant to the folding of large RNAs in vivo.

## REFERENCES AND NOTES

- R. F. Gesteland and J. F. Atkins, Eds., *The RNA World* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1993).
- F. Michel and E. Westhof, *ibid.* **216**, 585 (1990).
- S. Stern, B. Weiser, H. F. Noller, *ibid.*, **204**, 447 (1988).
- B. Felden, C. Florentz, R. Giegé, E. Westhof, *ibid.* **235**, 508 (1994).
- J. A. Doudna, C. Grosshans, A. Gooding, C. E. Kundrot, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 7829 (1993).
- E. P. Nikonowicz and A. Pardi, *J. Mol. Biol.* **232**, 1141 (1993).
- T. Pan, D. M. Long, O. C. Uhlenbeck, in (1), pp. 271-302.
- A. M. Pyle, *Science* **261**, 709 (1993).
- O. C. Uhlenbeck, J. Baller, P. Doty, *Nature* **225**, 508 (1970).
- J. W. Weller and W. E. Hill, *Biochemistry* **31**, 2748 (1992).
- S. T. Cload, P. L. Richardson, Y.-H. Huang, A. Scheppartz, *J. Am. Chem. Soc.* **115**, 5005 (1993).
- K. A. LeCuyer and D. M. Crothers, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 3373 (1994).
- (i) To ensure that probe binding and RNase H cleavage were fast, we increased the concentrations of oligonucleotide and RNase H until the extent and the rate of cleavage were maximal. At 20 to 60  $\mu\text{M}$  oligonucleotide more than 90 percent of the RNA was cleaved for all probes targeting P3 and P7. Similar oligonucleotide concentrations for P4 and P6a produced lower, but still saturated, cleavage (see text). When RNase H was 0.1 U/ $\mu\text{l}$  (corresponding to a large stoichiometric excess of enzyme over RNA), the reaction was always complete in less than 30 seconds. Under these conditions, therefore, hybridization and cleavage were rapid compared to the measured folding rates. (ii) To confirm that RNase H cleavage was faster than probe dissociation, we measured as a function of time the decrease in the extent of cleavage after incubation of preformed probe•RNA complex under folding conditions. For probes 10 nucleotides long, cleavage decreased less than 10 percent after 1 minute when P3, P4, or P7 were targeted, and less than 20 percent after 30 seconds when P6a was targeted. This shows that probe dissociation is slower than RNase H cleavage. A probe length of 10 nucleotides was chosen because shorter probes did not form sufficiently stable complexes, while longer probes anneal to target sequences that contribute to more than one structural feature in the RNA. (iii) To determine whether significant unfolding of the RNA occurred under native conditions, fully folded ribozyme was incubated in the presence of excess probe and RNase H. Less than 20 percent cleavage was observed after 2 minutes, indicating that the RNA remains in the folded, inaccessible conformation. These results show that, if probe binding and RNase H cleavage are performed for 30 seconds at high concentrations of oligonucleotide and RNase H, the fraction of RNA cleaved corresponds directly to the fraction that is unfolded.
- Only those sequences can be studied that are accessible to DNA probe hybridization and RNase H cleavage in the absence, but not in the presence, of  $\text{Mg}^{2+}$ , and where binding of the probe can compete successfully with structure formation when probe and  $\text{Mg}^{2+}$  are added at the same time. Therefore, the entire folding process cannot be followed from the denatured (D) to the native (N) state, but only the progression from the  $\text{Mg}^{2+}$ -free form of the RNA ( $I_0$ ) to the first folding intermediate in which binding of the DNA probe or cleavage by RNase H, or both, are blocked ( $I_1$ ). There certainly are steps preceding the addition of  $\text{Mg}^{2+}$ , and there may be further rearrangements after the formation of the first inaccessible intermediate. Furthermore, only those folding events accompanied by a change in accessibility can be detected. Thus, using these methods, only a part of the folding pathway can be explored (Eq. 3).
 
$$D \rightarrow I_0 \rightarrow I_1 \rightarrow N \quad (3)$$
- A. J. Zaugg, C. A. Grosshans, T. R. Cech, *Biochemistry* **27**, 8924 (1988).
- T. R. Cech, in (1), pp. 239-269.
- D. Herschlag and T. R. Cech, *Biochemistry* **29**, 10159 (1990).
- \_\_\_\_\_, *ibid.*, p. 10172.
- D. W. Celander and T. R. Cech, *Science* **251**, 401 (1991).
- B. Laggenbauer, F. L. Murphy, T. R. Cech, *EMBO J.* **13**, 2669 (1994).
- Although the lack of cleavage likely is due to the target site being unavailable for probe binding in the absence of  $\text{Mg}^{2+}$ , it may also be due to  $\text{Mg}^{2+}$ -induced structure formation in these regions being much faster than oligonucleotide hybridization, or to the RNA•DNA complex being sterically inaccessible to RNase H.
- S. Couture *et al.*, *J. Mol. Biol.* **215**, 345 (1990).
- When folding buffer containing  $\text{Na}^+$  only was added, no decrease in cleavage was observed. Folding could also be initiated by adding  $\text{Mg}^{2+}$  to RNA that had already been equilibrated in  $\text{Na}^+$ .
- Between 10 and  $\sim 400$  mM  $\text{Na}^+$  the extent and rate of decrease of cleavage were constant. At higher  $\text{Na}^+$  concentrations, the fraction of RNA competent for cleavage decreased in a gradual way, even in the absence of  $\text{Mg}^{2+}$ . This was not due to the inability of RNase H to function at such high salt concentrations (before cleavage,  $\text{Na}^+$  was diluted to 100 mM, where RNase H still cleaves rapidly). The decreased accessibility is most likely due to  $\text{Na}^+$ -induced stabilization of secondary structure, since Fe(II)-EDTA mapping has shown that the native tertiary structure does not form in the presence of monovalent ions only (19).
- While the most straightforward interpretation is that the observed decrease in cleavage is due to actual helix formation, it is not possible to rule out the possibility that other structural rearrangements result in the target sequence becoming inaccessible.
- We obtained  $k_{\text{obs}}$  by fitting the data to a single exponential. The value shown is derived from 17 independent experiments performed over a period of 10 months with several preparations of each component.
- Probes to both strands of P3 (1 and 15) gave identical results in all experiments, supporting the interpretation that the decrease in cleavage is due directly to formation or stabilization of the P3 helix. The only difference was that a three times higher concentration of probe 1 was required compared to probe 15 to achieve the same extent of cleavage, and fulfill all the kinetic conditions outlined above.
- J.-F. Wang and T. R. Cech, *J. Am. Chem. Soc.* **116**, 4178 (1994).
- A 14 nucleotide long version of probe 15 also acts as an inhibitor of the ribozyme reaction when added to the RNA at the same time as  $\text{Mg}^{2+}$ . The apparent inhibition constant ( $K_i$ ) was equivalent to the apparent dissociation constant ( $K_d$ ) of the oligonucleotide as determined by RNase H cleavage ( $\sim 0.4 \mu\text{M}$ ). This supports the interpretation that lack of RNase H cleavage is due to lack of probe hybridization, not to steric occlusion of RNase H.
- F. L. Murphy and T. R. Cech, *Biochemistry* **32**, 5291 (1993).
- \_\_\_\_\_, *J. Mol. Biol.* **236**, 49 (1994).
- The activation energy was obtained from an Arrhenius plot of the observed rates of formation of P3 and P7 (with probes 15 and 19) at five temperatures between 25° and 37°C.
- With probes targeting P3 or P7, we observe more than 90 percent cleavage if they are added together with  $\text{Mg}^{2+}$ , suggesting that the equilibrium between  $I_2$  and  $I_3$  is at least tenfold in favor of  $I_2$ , and thus that the true  $K_d$  of magnesium is at least ten times lower than the observed midpoint of the transition ( $[\text{Mg}^{2+}]_{1/2}$ ) of 0.97 mM.
- P. R. Schimmel and A. G. Redfield, *Annu. Rev. Biochem. Bioeng.* **9**, 181 (1980).
- P. E. Cole and D. M. Crothers, *Biochemistry* **11**, 4368 (1972).
- D. M. Crothers, P. E. Cole, C. W. Hilbers, R. G. Shulman, *J. Mol. Biol.* **87**, 63 (1974).
- A. Stein and D. M. Crothers, *Biochemistry* **15**, 160 (1976).
- D. C. Lynch and P. R. Schimmel, *ibid.* **13**, 1841 (1974).
- A. R. Banerjee, J. A. Jaeger, D. H. Turner, *ibid.* **32**, 153 (1993).
- L. Jaeger, E. Westhof, F. Michel, *J. Mol. Biol.* **234**, 331 (1993).
- I. E. Scheffler, E. L. Elson, R. L. Baldwin, *ibid.* **36**, 291 (1968).
- D. Pörschke and M. Eigen, *ibid.* **62**, 361 (1971).
- Since only those folding events can be monitored that are accompanied by a change in accessibility to our probes, and formation of the P5abc intermediate does not appear to cause such a change, it cannot definitively be placed on the folding pathway.
- E. L. Christian and M. Yarus, *Biochemistry* **32**, 4475 (1993).
- D. Herschlag, *ibid.* **31**, 1386 (1992).
- J.-F. Wang, W. D. Downs, T. R. Cech, *Science* **260**, 504 (1993).
- P. C. Bevilacqua, R. Kierzek, K. A. Johnson, D. H. Turner, *ibid.* **258**, 1355 (1992).
- E. Westhof and F. Michel, in *Structural Tools for the Analysis of Protein-Nucleic Acid Complexes*, D. J. M. Lilley, H. Heumann, D. Suck, Eds. (Birkhäuser Verlag, Basel, 1992), pp. 255-267.
- C. K. Woodward, *Curr. Opin. Struct. Biol.* **4**, 112 (1994).
- J.-R. Garel, in *Protein Folding*, T. E. Creighton, Ed. (Freeman, New York, NY, 1992), pp. 405-454.
- K. A. Dill, *Biochemistry* **29**, 7133 (1990).
- J. A. Jaeger, M. Zuker, D. H. Turner, *ibid.* **29**, 10147 (1990).
- Q. Guo and A. M. Lambowitz, *Genes Dev.* **6**, 1357 (1992).
- Z. Tsuchihashi, M. Khosla, D. Herschlag, *Science* **262**, 99 (1993).
- C. Reich, G. J. Olsen, B. Pace, N. R. Pace, *ibid.* **239**, 178 (1988).
- P. H. von Hippel, D. G. Bear, W. D. Morgan, J. A. McSwiggen, *Annu. Rev. Biochem.* **53**, 389 (1984).
- V. L. Emerick and S. A. Woodson, *Biochemistry* **32**, 14062 (1993).
- L-21 Sca I RNA was transcribed from plasmid pT7L-21 linearized with Sca I. Reactions (100  $\mu\text{l}$ ) were performed for 3.5 hours at 37°C in 40 mM tris (pH 8.1), 2 mM spermidine, 10 mM dithiothreitol, 25 mM  $\text{MgCl}_2$ , 1 mM each GTP, CTP, UTP, 0.1 mM ATP and 250  $\mu\text{Ci}$  of [ $\alpha$ - $^{32}\text{P}$ ] ATP (NEN) (for unlabeled RNA the concentration of each NTP was 4 mM), with 500 units of T7 RNA polymerase (New England Biolabs) and 15  $\mu\text{g}$  of template. The reaction was quenched with stop solution, and the product was purified on a 6 percent polyacrylamide gel and soaked into buffer containing 10 mM tris (pH 8.1), 1 mM EDTA, and 0.3 M sodium acetate. After ethanol precipitation and resuspension in tris-EDTA the RNA was quantitated by Cerenkov counting.
- To confirm the identity of the cleavage products, we subjected RNA to 5' or 3' end-labeling and subsequent probe binding and RNase H cleavage. The products were compared directly with an RNA sequencing ladder generated by  $I_2$  cleavage of phosphorothioate containing ribozyme [J. Rudinger *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5882 (1992)].  $S_2$  is due to cleavage at sequences in P7 that have partial homology to the target sequence in P3. Since no intermediate resulting from cleavage at the P7 site only was ever detected, it is assumed that  $S_2$  can only derive from  $S_1$  after the initial cleavage event in P3. Therefore, the presence of  $S_2$  does not alter the quantitation of the fraction cleaved, which is obtained from the ratio of  $L/(L + F)$  only.
- The concentration of ribozyme was determined kinetically, as described by Herschlag and Cech (17), by quantitating the initial burst in a multiple turnover experiment in which the RNA was allowed to fold to completion before the reaction was initiated.

61. Near stoichiometric concentrations of ribozyme and substrate were used, instead of more standard single or multiple turnover conditions, in order to be able to follow the time course of gain of full activity. In large ribozyme excess fully active ribozyme accumulates too rapidly to observe changes in initial reaction rate, whereas in large substrate excess the fraction of product formed in the very early stages of the reaction is too small for accurate quantitation.

62. For probes **8** and **12** the initial burst described in the text is reflected by the imperfect fit of the data to a single exponential.

63. We thank J. Puglisi and the members of the Williamson Lab for many discussions, P. Kim for helpful suggestions, M. Fedor for advice on kinetics and critical comments, J. McKnight, D. Fass, and S. Blacklowe for critical reading of the manuscript, J. Essigmann for use of the PhosphorImager, T.

Cech for providing the pT7L-21 plasmid, N. Usman for the ribozyme substrate, and Q. Eastman for help in initiating these studies. Supported by a grant from the Searle Scholar Program of the Chicago Community Trust, and a predoctoral fellowship (P.P.Z.) from the Howard Hughes Medical Institute.

15 April 1994; accepted 27 June 1994