Small angle X-ray scattering reveals a compact intermediate in RNA folding

Rick Russell[1], Ian S. Millett[2], Sebastian Doniach[3] and Daniel Herschlag[1,2]

[1]Department of Biochemistry and [2]Department of Chemistry, Stanford University, Stanford, California 94305, USA. [3]Department of Applied Physics, Stanford University, Stanford, California 92343, USA.

We have used small angle X-ray scattering (SAXS) to monitor changes in the overall size and shape of the *Tetrahymena* ribozyme as it folds. The native ribozyme, formed in the presence of Mg^{2+}, is much more compact and globular than the ensemble of unfolded conformations. Time-resolved measurements show that most of the compaction occurs at least 20-fold faster than the overall folding to the native state, suggesting that a compact intermediate or family of intermediates is formed early and then rearranges in the slow steps that limit the overall folding rate. These results lead to a kinetic folding model in which an initial 'electrostatic collapse' of the RNA is followed by slower rearrangements of elements that are initially mispositioned.

The *Tetrahymena* group I ribozyme has served as a powerful system to understand the process by which RNA forms a functional three-dimensional structure (for recent reviews see refs 1,2). The ribozyme contains stretches of nucleotides that form base pairs; these paired regions are referred to as P1–P9 (Fig. 1*a*). Most of these base pairs are expected to form in the absence of Mg^{2+} ions. However, Mg^{2+} or other divalent metal ions are required for the ribozyme to form tertiary contacts, giving it a discrete fold with an interior that is protected from bulk solvent[3,4]. The folded ribozyme catalyzes a reaction (equation 1) in which an oligonucleotide substrate is cleaved by an exogenous guanosine nucleophile.

$$E + CCCUCUA_n + G \rightarrow E + CCCUCU + GA_n \qquad (1)$$

Time-resolved oligonucleotide hybridization and hydroxyl-radical cleavage protection approaches have been used to characterize the kinetic folding process of the ribozyme[5,6]. These studies have revealed a wide range of rate constants for stable formation of parts of the structure (Fig. 1*b*). Residues in the P5abc subdomain form a stable tertiary structure with a rate constant of >100 min^{-1}, followed closely by the rest of the P4–P6 domain[6]. In contrast, minutes are required for portions of the P3-P7-P8 domain in the ribozyme core to form a stable tertiary structure that is protected from solvent[5,6]. These experiments strongly suggest that a folding intermediate accumulates in which the P4–P6 tertiary structure has been formed but the P3-P7-P8 domain is not stably formed. Further, the identification of these disparate folding rates highlights the power of protection approaches to monitor the local environments of individual residues in a time-resolved manner. Small angle X-ray scattering (SAXS) can, in principle, provide information complementary to these techniques that monitor the local environments by reporting, in a model independent fashion, on the overall size and shape of the RNA during folding[7].

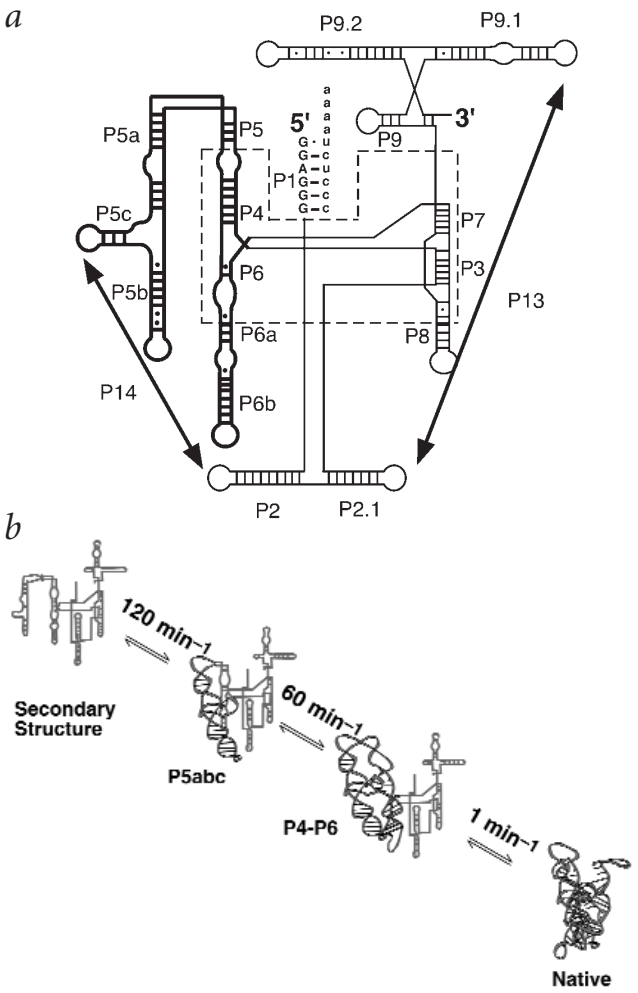

Fig. 1 The *Tetrahymena* ribozyme. ***a***, Paired secondary structure elements P1–P9 are labeled, and the long range pairings P13 and P14 are indicated with arrows[26,33]. The boxed portion is the ribozyme core, the region most highly conserved[33]; the core is largely protected from solvent in the presence of Mg^{2+} (ref. 3). The P4-P6 subdomain, which acquires stable tertiary structure early in the folding pathway subsequent to Mg^{2+} addition is shown in bold (see *b*). ***b***, The ribozyme folding reaction[5,6]. Portions of the ribozyme that have not acquired stable tertiary structure are depicted as two-dimensional secondary structure, while folded elements are shown as ribbons. The P4-P6 subdomain acquires stable tertiary structure with a rate constant of 60 min^{-1} at 42 °C (ref. 6). Overall folding of the ribozyme to its native structure is much slower, with a rate constant of 1 min^{-1}. Peripheral elements P2 and P9 are omitted from the folded structure for clarity.

It has been proposed that the rate-limiting step for overall folding of the *Tetrahymena* ribozyme to the native state involves emergence from a kinetic trap, requiring the disruption of non-native[8] or native[9] structure. However, based on other evidence, a conformational search for long range tertiary contacts has been proposed to be the rate-limiting step of overall folding[10]. We have used SAXS to test and extend these models (Fig. 2). If the slow folding steps involve a search of conformational space by subdomains such as P3-P7-P8 that have not yet formed stable tertiary structure, then a large fraction of the reduction in size would be expected to occur late in folding (Fig. 2*a*). Conversely, if a compact folding intermediate were formed quickly that must rearrange or partially unfold in a slow step, much of the compaction would occur much earlier than completion of overall folding (Fig. 2*b*). It is also possible that a kinetically trapped

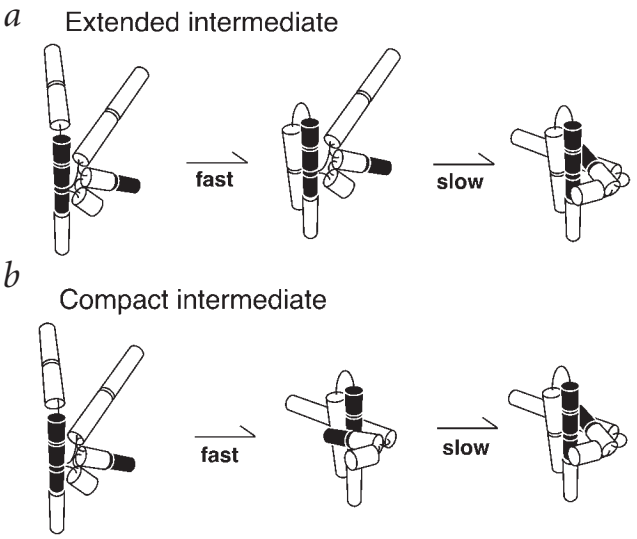

Fig. 2 Limiting models for the rate of compaction of the ribozyme relative to overall folding. ***a***, The slow folding steps involve conformational searching by domains that have not yet formed stable tertiary structure so that folding intermediates are substantially less compact than the final structure. ***b***, The ribozyme compacts rapidly but must then rearrange or partially unfold to reach the final structure.

intermediate remains extended or is only partially compact. Beyond distinguishing these models, SAXS can place unique constraints on our developing molecular understanding of RNA folding by providing a physical description of the intermediate that accumulates.

SAXS has proven to be a powerful approach for monitoring the shape of proteins and RNA[11–13], and more recently for monitoring the rate of protein compaction during folding[14–16]. We first show here that SAXS can distinguish between the folded and the unfolded *Tetrahymena* group I ribozyme. We then demonstrate that compaction of this RNA is much faster than overall folding to the native state.

A compact native state

To determine whether tertiary folding of the ribozyme is detectable by SAXS, measurements were made under conditions that do or do not support formation of tertiary structure. In the absence of Mg^{2+}, conditions under which the ribozyme lacks stable tertiary structure[17], SAXS measurements gave a radius of gyration (R_g) value of 74 Å obtained using the Guinier approximation (Table 1). Addition of 15 mM Mg^{2+} to the ribozyme, with a 30 min, 50 °C incubation, which gives maximal activity[18], resulted in a substantially decreased R_g value of 47 Å (Table 1). Additionally, the shape of the scattering profile revealed that the ribozyme is substantially more globular with Mg^{2+} (Fig. 3*a*).

To address whether the reduction in R_g requires formation of the active structure or whether it is a more general ionic effect, the dependence of R_g on the identity of the metal ion was examined. Addition of Na^+ to concentrations up to 1 M gave only a small reduction in R_g (Table 1). It has been shown that stable tertiary structure is not formed in Na^+ (ref. 17). In contrast, addition of Ca^{2+}, which has been shown to support tertiary structure formation but not activity[17,19,20], gave an R_g value and scattering profile similar to that with Mg^{2+} (Table 1 and data not shown). The correlation between conditions that allow the formation of stable tertiary structure and those that give a substantial reduction in R_g value suggests that this large compaction is dependent on folding of the ribozyme.

To further test whether the reduction in R_g of the ribozyme results from specific tertiary structure formation, SAXS measurements were performed at a range of Mg^{2+} concentrations that spans the concentration required for tertiary structure formation. Formation of native tertiary structure for this and other catalytic RNAs is highly cooperative with Mg^{2+} concentration[17,21–23]. A reduction in R_g was observed with a midpoint of ~0.15 mM Mg^{2+} and a Hill coefficient of ≥3 (Fig. 3*b*). The cooperative nature of the transition and the similarity of the midpoint to that observed for formation of tertiary contacts[17] support the idea that the compaction reflects formation of the native tertiary structure.

To perform its function, the folded ribozyme must bind its two substrates, the oligonucleotide to be cleaved and a guanosine nucleophile. To address whether the conformation that gives an R_g of 47 Å is the completely folded and active form of the ribozyme, both substrates were added at concentrations above their respective dissociation constants. These additions had no detectable effect on the value of R_g or on the X-ray scattering

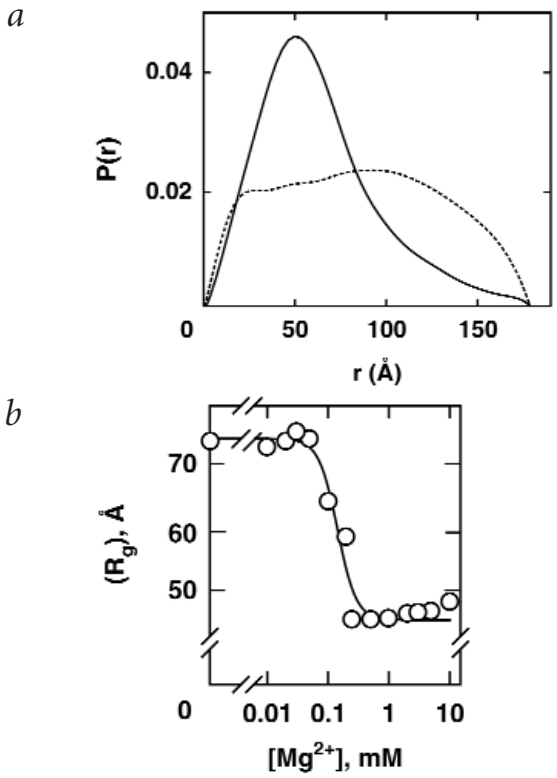

Fig. 3 Tertiary folding of the ribozyme detected by SAXS. ***a,*** P(r), the distance distribution function, for the native (15 mM Mg^{2+}, solid curve) and unfolded (no Mg^{2+}, dashed curve) ribozyme. The peak at low vector length for the native ribozyme indicates a compact, globular shape. ***b,*** Compaction is cooperative with Mg^{2+} concentration. The desired free Mg^{2+} concentration was achieved by repeated dilution and RNA concentration (see Methods). The spacing on the y-axis is scaled as the square of R_g to allow quantitative evaluation. This is necessary because SAXS intensity is proportional to R_g squared[7]. The curve represents a fit to the data around the transition by a cooperative binding equation, giving a value of 0.14 ± 0.02 mM and a Hill coefficient of 2.8 ± 0.9. Because of the small number of points in the transition, the Hill coefficient is best regarded as a minimum. The small but significant increase in R_g observed at the higher concentrations of Mg^{2+} (0.5–10 mM) is likely due to binding of multiple additional Mg^{2+} ions to the folded ribozyme.

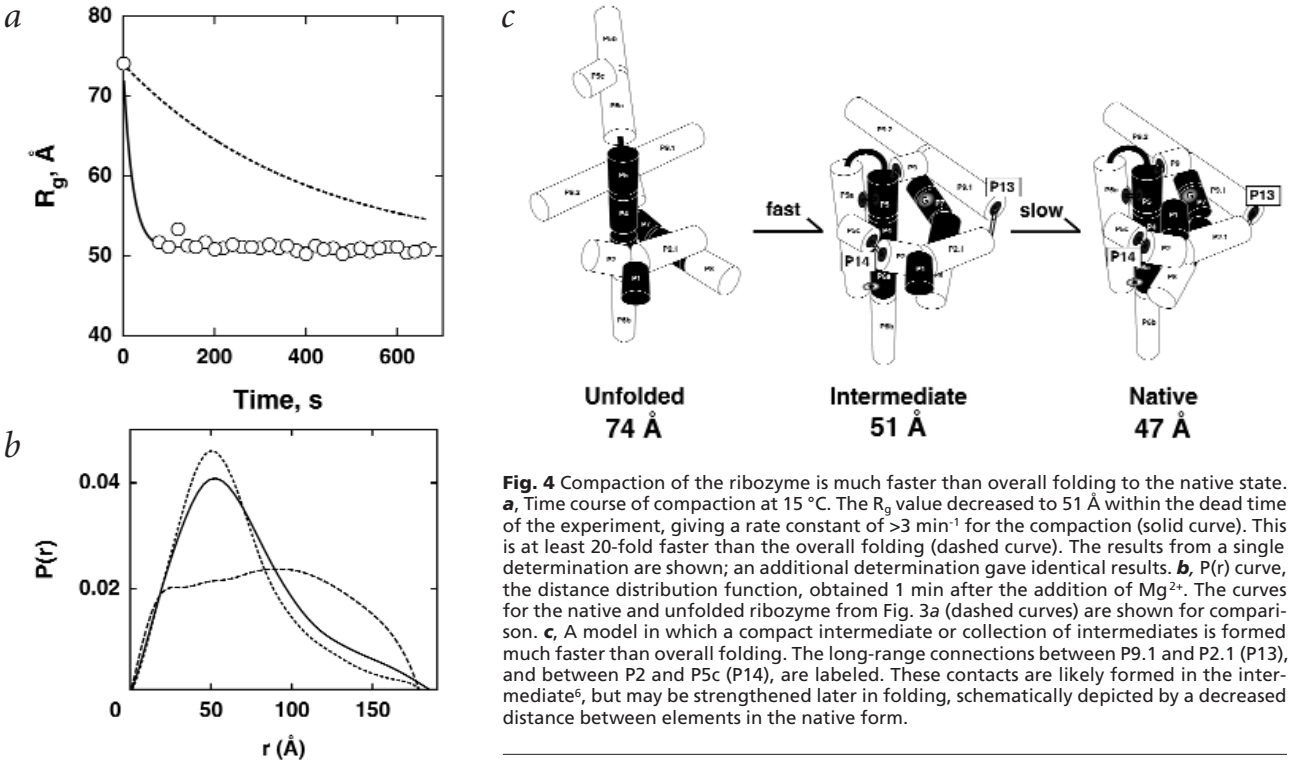

Fig. 4 Compaction of the ribozyme is much faster than overall folding to the native state. ***a***, Time course of compaction at 15 °C. The R_g value decreased to 51 Å within the dead time of the experiment, giving a rate constant of >3 min^{-1} for the compaction (solid curve). This is at least 20-fold faster than the overall folding (dashed curve). The results from a single determination are shown; an additional determination gave identical results. ***b***, P(r) curve, the distance distribution function, obtained 1 min after the addition of Mg^{2+}. The curves for the native and unfolded ribozyme from Fig. 3*a* (dashed curves) are shown for comparison. ***c***, A model in which a compact intermediate or collection of intermediates is formed much faster than overall folding. The long-range connections between P9.1 and P2.1 (P13), and between P2 and P5c (P14), are labeled. These contacts are likely formed in the intermediate[6], but may be strengthened later in folding, schematically depicted by a decreased distance between elements in the native form.

profile (data not shown), suggesting that the ribozyme does not undergo large scale conformational changes upon binding its substrates. This is consistent with the suggestion based on X-ray diffraction data collected at 5 Å resolution that the active site is largely pre-organized for substrate binding and catalysis[4].

The results described above provide information about the global conformations of the native and unfolded ribozyme conformations, and indicate that the native form is much more compact than the unfolded form. Thus, the folding transition can be readily followed by SAXS.

Rapid compaction in folding

Time-resolved SAXS measurements were performed manually at 15 °C, conditions under which the overall rate constant for folding is 0.15 min^{-1} ($t_{1/2}$ ~5 min; ref. 24). Upon initiation of tertiary folding with the addition of Mg^{2+}, the R_g value decreased from 74 Å to 51 Å within the experimental dead time of 1 min (Fig. 4*a*). This gives a minimum rate constant for the compaction of 3 min^{-1}, at least 20-fold faster than the overall rate constant for folding. Analysis of the scattering profile at early times indicates that this 51 Å intermediate is globular in shape, similar to the native ribozyme (Fig. 4*b*).

The value of R_g remained 51 Å for at least 10 min, even though the native state has an R_g value of 47 Å. This is because, under these conditions, 95% of the ribozyme population misfolds to a state that is stable for hours (this misfolded state rapidly converted to the thermodynamically preferred native state at elevated temperature[24]). In a separate experiment, this kinetically stable misfolded state was formed by addition of Mg^{2+} at low temperature[24] and was shown to give an R_g value of 51 Å, identical to that for the folding intermediate. A subsequent 50 °C incubation was required to give the reduced R_g of the native state (Table 1). It is important to note that the pathways that lead to the native and misfolded states diverge from each other late in folding, concomitant with or subsequent to the rate-limiting step (R.R. and D.H., unpublished observations). This indicates that the 51 Å intermediate observed prior to the rate-limiting step (Fig. 4*a*) is populated by molecules that fold correctly as well as incorrectly.

The rapid reduction in R_g relative to overall folding shows that a compact intermediate or family of intermediates is formed early in folding, and the subsequent slow steps to form the native state result in only a small additional compaction. Because the intermediate is already nearly as compact as the folded state (Fig. 4*c*), the slower folding steps are presumably rearrangements of this structure. These rearrangements may involve breaking and exchanging contacts from within the compact intermediate, or transient opening of the intermediate to allow less restricted motion of the core. The results here provide experimental support for the idea that a kinetic trap limits folding of the ribozyme[6,9,23,25], and extend the model by establishing that the kinetically trapped intermediate is nearly as compact as the native state. Combining these and other results, a crude molecular model is developed below for how a kinetic trap can limit folding of the ribozyme.

The time-resolved hydroxyl radical protection studies suggest that two long-range tertiary contacts between loops P13 and P14 are formed significantly faster than overall folding[6] (Fig. 4*c*). The SAXS results here and these previous results would most simply suggest that P13 and P14 constrain the intermediate by forming a nearly continuous ring of peripheral RNA duplexes around the core, as they do in the native state[6,26,27]. We have found, however, that disruption of P13 by mutagenesis[26] has no effect on the rate constant for initial folding; in contrast, this disruption accelerates conversion from the stable misfolded state described above to the native state by 200-fold (data not shown). Thus, formation of P13, and the corresponding ring of peripheral duplex elements, is apparently not responsible for the observed slow rate of initial folding. (Initial folding involves transit through the intermediate but not the stable misfolded form[24].)

Table 1 SAXS characterization of ribozyme folding

Conditions[1]	Ribozyme form[2]	Radius of gyration (R_g; Å)[3]
No Mg^{2+}	Unfolded	74 ± 4
15 mM Mg^{2+}	Native	47.5 ± 0.4
50 mM Mg^{2+}	Native	48.2 ± 0.3
20 mM Ca^{2+}	Native	52 ± 2
100 mM Na^{+}	Unfolded	68 ± 5
1 M Na^{+}	Unfolded	68 ± 5

[1]All measurements were performed at 15 °C in solutions containing 50 mM Na-(3-[N-Morpholino]propanesulfonic acid) (NaMOPS), pH 7.2, in addition to the listed ionic condition. The indicated concentrations of Mg^{2+}, Ca^{2+}, and Na^{+} were added as chloride salts. R_g values were independent of ribozyme concentration (see Methods).
[2]Ribozyme form refers to the presence or absence of stable tertiary structure based on protection of the core from solution radicals[17].
[3]Radius of gyration values were determined using the Guinier approximation. Errors are the standard deviation of the Guinier fit from a single measurement, except at 15 mM Mg^{2+}, in which case data represent the mean and standard deviation from three independent measurements.

Rather than being constrained by outer contacts, the folding intermediate may derive much of its stability from interactions of the improperly folded core elements themselves within a compact structure. P13, though formed in the intermediate, may be only marginally stable until it is strengthened in the misfolded and native states.

How are kinetically trapped species formed early in folding? The SAXS results suggest a general model for RNA folding in which cooperative Mg^{2+} binding to sites along and between the phosphodiester backbones of preformed RNA duplexes results in a rapid compaction of the overall structure. The slow steps in folding are then proposed to be rearrangements of structural elements that have formed incorrectly, including incorrectly paired duplexes and improperly positioned duplexes. Although this 'electrostatic collapse' is conceptually analogous to the hydrophobic collapse that occurs early in the folding of some proteins, there may be larger energetic penalties for rearrangement of misfolded core RNA than the equivalent hydrophobic protein core, leading to slower RNA folding[28]. Extending SAXS to shorter time regimes will provide a critical test of the idea that RNA folding begins with a rapid electrostatic collapse of structure.

Methods

Preparation of ribozyme. The L-21/*Sca*I ribozyme was prepared by *in vitro* transcription and purified using a Qiagen RNeasy column as described[24,29]. To prefold the ribozyme to the native state, Mg^{2+} was added to 15 mM and the ribozyme was incubated at 50 °C for 30 min. This pre-incubation allows the misfolded state that is formed initially to refold to the native state (R.R. and D.H., unpublished observations). The RNA concentration was sufficiently high to bind a substantial fraction of the total Mg^{2+} initially added. Therefore, the concentration of free Mg^{2+} was subsequently adjusted, if necessary, by repeatedly diluting the sample with a solution containing the desired Mg^{2+} concentration followed by concentrating the RNA using a Microcon-50 centrifugation concentrator following the manufacturer's instructions. For SAXS measurements with bound substrates, the oligonucleotides CCCUC(dU)A_5 and GGUCG were added to the prefolded ribozyme (11 μM) to final concentrations of 14 μM each ($K_d^{CCCUC(dU)A_5}$ < 0.02 nM (ref. 30); K_d^{GGUCG} < 2 μM (R.R. and D.H., unpublished observations)). The rate constant for cleavage of CCCUC(dU)A_4 is 0.005 min^{-1} under these conditions (data not shown), so the extent of reaction is insignificant during the 10 min SAXS data collection (see below).

SAXS measurements. Measurements were performed on beamline 4-2 at the Stanford Synchrotron Radiation Laboratory (SSRL). X-ray energy was selected at 8980 eV (Cu edge) using a pair of Mo/B_4C multilayer monochromator crystals[31]. Solution conditions were 50 mM NaMOPS, pH 7.2, and all measurements were made at 15 °C using a static cuvette with a sample volume of 30 μl. The ribozyme concentration was 10–30 μM (1.3–3.8 mg ml^{-1}). Good signal to noise ratios, with scattering at least two fold above the buffer background, were attainable throughout this concentration range. Scattering data for static measurements were routinely collected for 10 min and averaged, but collecting data for only 10 s gave the same values of R_g within error (data not shown). Values of R_g, for both native and unfolded ribozyme samples, were independent of ribozyme concentration over this range, suggesting that the ribozyme was predominantly monomeric, and that interparticle interference was not a large factor at these concentrations of ribozyme (0 and 15 mM Mg^{2+}; data not shown). Time-resolved experiments were performed by rapidly adding Mg^{2+} to a sample of ribozyme previously kept on ice. The sample was placed in a cuvette and the cuvette in the beam path. The experimenter exited the hutch and opened the beam shutters as rapidly as possible. Following an initial 1 min dead time, data were collected in 10 s bins.

Data analysis. Radius of gyration values were calculated using the Guinier approximation[7]. The distribution function of interatomic distances within the RNA, P(r), was estimated from the scattering data using the GNOM algorithm of Semenyuk and Svergum[32]. The longest atom pair distance d_{max} was set by eye so this part of the P(r) curve is subject to error. This uncertainty does not affect the R_g values, which were obtained directly from the scattering data.

Acknowledgments
We thank H. Tsurata for help on beamline 4-2 and K. Hodgson for his support. This research was supported by an NIH grant to D.H. The US Department of Energy and the National Institutes of Health support SSRL. R.R. was supported by an NIH postdoctoral fellowship.

Correspondence should be addressed to D.H. *email: herschla@cmgm.stanford.edu*

Received 20 January, 2000; accepted 27 March, 2000.

1. Treiber, D.K. & Williamson, J.R. *Curr. Opin. Struct. Biol.* **9**, 339–345 (1999).
2. Ferre-D'Amare, A.R. & Doudna, J.A. *Annu. Rev. Biophys. Biomol. Struct.* **28**, 57–73 (1999).
3. Latham, J.A. & Cech, T.R. *Science* **245**, 276–282 (1989).
4. Golden, B.L., Gooding, A.R., Podell, E.R. & Cech, T.R. *Science* **282**, 259–264 (1998).
5. Zarrinkar, P.P. & Williamson, J.R. *Science* **265**, 918–924 (1994).
6. Sclavi, B., Sullivan, M., Chance, M.R., Brenowitz, M. & Woodson, S.A. *Science* **279**, 1940–1943 (1998).
7. Glatter, O. & Kratky, O. *Small angle X-ray scattering*. (Academic Press, London; 1982).
8. Pan, J., Thirumalai, D. & Woodson, S.A. *J. Mol. Biol.* **273**, 7–13 (1997).
9. Treiber, D.K., Rook, M.S., Zarrinkar, P.P. & Williamson, J.R. *Science* **279**, 1943–1946 (1998).
10. Zarrinkar, P.P. & Williamson, J.R. *Nature Struct. Biol.* **3**, 432–438 (1996).
11. Lake, J.A. & Beeman, W.W. *J. Mol. Biol.* **31**, 115–125 (1968).
12. Pilz, I., Kratky, O., Cramer, F., Haar, F.v.d. & Schlimme, E. *Eur. J. Biochem.* **15**, 401–409 (1970).
13. Kratky, O. & Pilz, I. *Q. Rev. Biophys.* **5**, 481–537 (1972).
14. Plaxco, K.W., Millett, I.S., Segel, D.J., Doniach, S. & Baker, D. *Nature Struct. Biol.* **6**, 554–556 (1999).
15. Segel, D.J., *et al. J. Mol. Biol.* **288**, 489–499 (1999).
16. Pollack, L., *et al. Proc. Natl. Acad. Sci. USA* **96**, 10115–10117 (1999).
17. Celander, D.W. & Cech, T.R. *Science* **251**, 401–407 (1991).
18. Herschlag, D. & Cech, T.R. *Biochemistry* **29**, 10159–10171 (1990).
19. Grosshans, C.A. & Cech, T.R. *Biochemistry* **28**, 6888–6894 (1989).
20. McConnell, T.S., Herschlag, D. & Cech, T.R. *Biochemistry* **36**, 8293–8303 (1997).
21. Murphy, F.L. & Cech, T.R. *Biochemistry* **32**, 5291–5300 (1993).
22. Doherty, E.A. & Doudna, J.A. *Biochemistry* **36**, 3159–3169 (1997).
23. Pan, T. & Sosnick, T.R. *Nature Struct. Biol.* **4**, 931–938 (1997).
24. Russell, R. & Herschlag, D. *J. Mol. Biol.* **291**, 1155–1167 (1999).
25. Rook, M.S., Treiber, D.K. & Williamson, J.R. *J. Mol. Biol.* **281**, 609–620 (1998).
26. Lehnert, V., Jaeger, L., Michel, F. & Westhof, E. *Chem. Biol.* **3**, 993–1009 (1996).
27. Pan, J. & Woodson, S.A. *J. Mol. Biol.* **294**, 955–965 (1999).
28. Herschlag, D. *J. Biol. Chem.* **270**, 20871–20874 (1995).
29. Zaug, A.J., Grosshans, C.A. & Cech, T.R. *Biochemistry* **27**, 8924–8931 (1988).
30. Narlikar, G.J., Bartley, L.E., Khosla, M. & Herschlag, D. *Biochemistry* **38**, 14192–14204 (1999).
31. Tsuruta, H., *et al. J. Appl. Crystallogr.* **31**, 672–682 (1998).
32. Semenyuk, A.V. & Svergun, D.I. *J. Appl. Crystallogr.* **24**, 537–540 (1991).
33. Michel, F. & Westhof, E. *J. Mol. Biol.* **216**, 585–610 (1990).