doi:10.1016/j.jmb.2005.07.010 J. Mol. Biol. (2005) 352, 712–722

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Structure of a Folding Intermediate Reveals the Interplay Between Core and Peripheral Elements in RNA Folding

Nathan J. Baird[1,2], Eric Westhof[3], Hong Qin[4], Tao Pan[4]* and Tobin R. Sosnick[2,4]*

[1]*Department of Chemistry University of Chicago, Chicago IL 60637, USA*

[2]*Institute for Biophysical Dynamics, University of Chicago, Chicago, IL 60637 USA*

[3]*Institut de biologie moléculaire et cellulaire, UPR9002, CNFRS Université Louis Pasteur Strasbourg Cedex, F-67084 France*

[4]*Department of Biochemistry and Molecular Biology, University of Chicago, Chicago IL 60637, USA*

**Corresponding authors*

Though the molecular architecture of many native RNA structures has been characterized, the structures of folding intermediates are poorly defined. Here, we present a nucleotide-level model of a highly structured equilibrium folding intermediate of the specificity domain of the *Bacillus subtilis* RNase P RNA, obtained using chemical and nuclease mapping, circular dichroism spectroscopy, small-angle X-ray scattering and molecular modeling. The crystal structure indicates that the 154 nucleotide specificity domain is composed of several secondary and tertiary structural modules. The structure of the intermediate contains modules composed of secondary structures and short-range tertiary interactions, implying a sequential order of tertiary structure formation during folding. The intermediate lacks the native core and several long-range interactions among peripheral regions, such as a GAAA tetraloop and its receptor. Folding to the native structure requires the local rearrangement of a T-loop in the core in concert with the formation of the GAAA tetraloop–receptor interaction. The interplay of core and peripheral structure formation rationalizes the high degree of cooperativity observed in the folding transition leading to the native structure.

Keywords: modeling; P RNA; S-domain

Introduction

As a result of the hierarchical nature of RNA structure, intermediates populate at significant levels in RNA folding.[1,2] The folding transition to the native structure often occurs in a highly cooperative manner, which can be identified by monitoring structural or spectroscopic features that change as a function of Mg^{2+} concentration.[3,4] A general perception is that the folding intermediates are rich in secondary structure (e.g. Watson–Crick helices) but lack extensive amounts of tertiary contacts. In this view, the transition to the native structure would involve only the organization of the pre-formed helical domains into a compact architecture. This notion of "secondary structure first, tertiary structure later" is applicable to the Mg^{2+}-dependent, equilibrium folding of small tertiary RNAs such as yeast $tRNA^{Phe}$,[5] or the hammerhead and hairpin ribozymes.[6,7]

The folding of larger RNAs is likely to be more complex. Structural information of key folding intermediates at the nucleotide level of resolution is required to investigate the interplay of secondary and tertiary structure formation in these large systems. Such structural detail, however, is sparse, as compared to the tremendous amount of information available on native RNA structures. Because partially folded structures are dynamic and unlikely to be crystallized, other strategies must be employed to obtain the detailed structural information for better understanding of the forces mediating the assembly and stability of large RNAs.

Here, we investigate the equilibrium folding of the specificity domain, or S-domain, of the *Bacillus subtilis* RNase P RNA. The crystal structure contains four tertiary structural modules: a rigid core (core); a four-way junction; a GAAA tetraloop (TL)–receptor; and an unusual motif involving two tertiary interacting loops (J11/12

Abbreviations used: DEPC, diethylpyrocarbonate; DMS, dimethyl sulfate; kethoxal or KE, α-keto-β-ethoxybutyraldehyde; SAXS, small-angle X-ray scattering; TL, tetraloop; U, unfolded; I_{eq}, equilibrium intermediate; N, native.

E-mail addresses of the corresponding authors: taopan@uchicago.edu; trsosnic@uchicago.edu

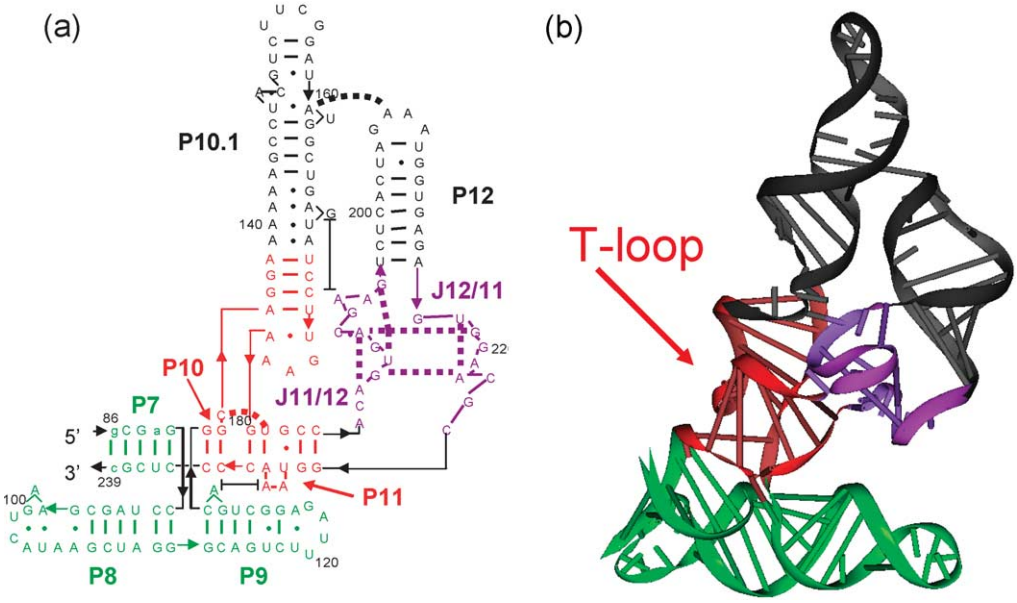

Figure 1. The crystal structure of the specificity domain of RNase P RNA from *Bacillus subtilis*.[8] (a) Secondary structure representation. Color coding is according to the native structure modules, four-way junction (green), core (red), J11/12 junction (purple), TL-receptor (black). Broken lines represent tertiary interactions. Continuous lines between A130/A230 and G168/A194 indicate stacking interactions in the native structure. (b) Three-dimensional structure (PDB 1NBS). The T-loop in the core is composed of five nucleotides, $U_{175}GAAA_{179}$, and is of particular importance to the folding intermediate.

and J12/11) (Figure 1).[8] In the thermodynamic folding pathway, an intermediate, termed I_{eq}, populates. *A priori*, it is unclear which of the native secondary and tertiary modules are present in the folding intermediate.

We derive a nucleotide-level model of the intermediate using chemical and nuclease mapping, circular dichroism (CD) spectroscopy, small-angle X-ray scattering (SAXS) and molecular modeling. The intermediate contains two of the four native tertiary modules. It lacks the core, which results in many of the structural elements extending outwards. Folding from this intermediate to the native structure involves the formation of the core in concert with large-scale motions that bring two peripheral helices into proximity to form the tetraloop–receptor interaction. The coupling of core and peripheral structure formation rationalizes the high degree of cooperativity observed in the folding of this RNA.

Results

The S-domain architecture

The core of the S-domain is composed of the U-shaped, T-loop motif[8,9] located at the end of the P10.1 helix, two coaxially stacked short helices (P10 and P11) and two bulged adenosine bases: A229 intercalates between G133 and C134, while A230 stacks on A130 of P9 (Figure 1). The four-way junction is composed of helices P7–P10. The J11/12 module is composed of two large internal loops, termed J11/12 and J12/11, located between helices P11 and P12. Each of the two loops forms a T-loop motif, and the nucleotide bases within these loops interact extensively through non-Watson–Crick base-pairs. The TL–receptor module is composed of a GAAA tetraloop (P12) interacting with a canonical tetraloop receptor embedded in P10.1.

Mg^{2+}-dependent folding transition

Equilibrium folding is monitored as a function of increasing concentration of Mg^{2+} using CD at 260 nm and 275 nm (Figure 2). The signal at 260 nm first increases and then decreases. This behavior is indicative of a sequential, three-state folding transition (unfolded-to-intermediate-to-nativ, U-to-I_{eq}-to-N), similar to the folding behavior of other tertiary RNAs.[10] At 275 nm, only the I_{eq}-to-N transition is observed. This transition corresponds to the folding to the native structure, as confirmed by structural mapping (see below). The midpoint of the I_{eq}-to-N transition, K_{Mg}, is 0.23($\pm$ 0.03) mM with a Hill constant of 4.8 ± 0.3 (Figure 2(b)).

The thermodynamic I_{eq} state represents the ensemble of RNA structures present at the beginning of the I_{eq}-to-N transition, under the following conditions; 20 mM Tris–HCl (pH 8), 0.1 mM $MgCl_2$ and <0.5 μM S-domain RNA at 37 °C. This intermediate has properties distinct from those of both the unfolded and the native populations. The CD signal at 260 nm for the intermediate under the above conditions is 4.0 $cm^2\ mmol^{-1}$, whereas the CD signals of both the unfolded and native

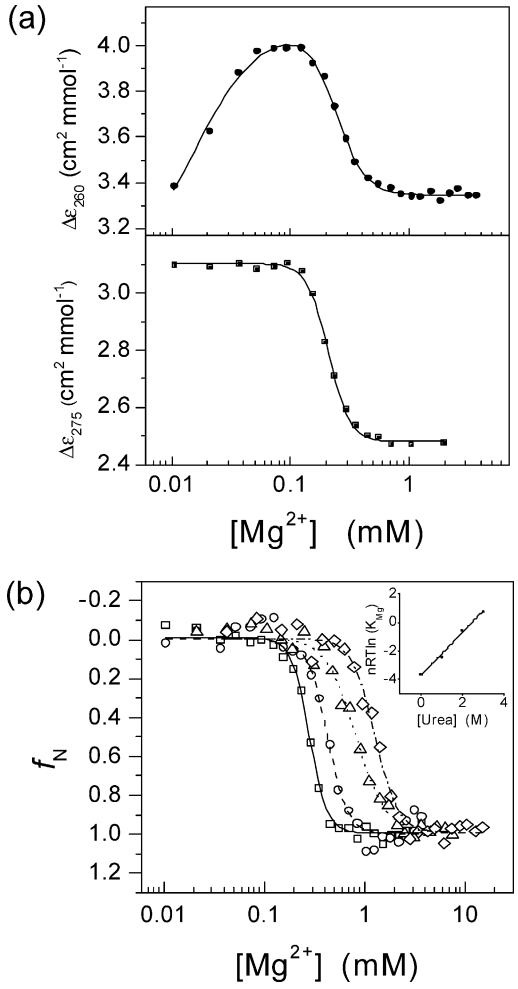

Figure 2. Folding transitions monitored by CD spectroscopy. (a) Mg^{2+} titration monitored by CD at 260 nm (top). The data are fit to a Hill-type equation (equation (1)), incorporating three populated states: U, I_{eq}, and N. The Mg^{2+} titration monitored by CD at 275 nm (bottom), where the signal sensitive only to the $I_{eq} \leftrightarrow N$ transition is fit to equation (2). (b) Mg^{2+} titrations in the presence of urea: 0 M (□), 1 M (○), 2 M (Δ), 3 M (◇). The inset is a plot of nRT ln(K_{Mg}) *versus* [urea]. The slope is m = 1.5(±0.1) kcal mol^{-1} M^{-1}.

states are lower, 2.9 cm^2 mmol^{-1}and 3.3 cm^2 mmol^{-1}, respectively. No combination of unfolded and native signals can produce the signal of the intermediate. Therefore, the thermodynamic intermediate is not a combination of the unfolded and native states.

To determine the surface area buried in the I_{eq}-to-N folding transition, Mg^{2+} titrations are carried out in the presence of 0–3 M urea. The accompanying decrease in stability (Figure 2(b)) has a linear dependence on the concentration of urea:

$$\Delta G([\text{urea}]) = \Delta G(0) + m[\text{urea}]$$

where the slope (m-value, the dependence of the free energy on urea) is proportional to the amount of surface area buried in the transition.[4] K_{Mg} and n were used to calculate the free energy of the N relative to I_{eq} according to:[11]

$$\Delta G_{IN}([Mg^{2+}]) = -nRT \ln([Mg^{2+}]/K_{Mg})$$

The m-value for this transition is 1.5(±0.1) kcal mol^{-1} M^{-1} (Figure 2(b)), which corresponds to the amount of surface area buried in a 22 base-pair RNA duplex.[4]

Characterization of I_{eq}

The extent of solvent exposure of nucleotide bases and ribosephosphate backbone as a function of [Mg^{2+}] is determined using three chemicals and two nucleases to obtain structural information at the nucleotide level (Figure 3).[12,13] Dimethyl sulfate (DMS) modifies N1 of adenosine and N3 of cytidine; kethoxal (KE) modifies N1 and N2 of guanosine; and diethyl pyrocarbonate (DEPC) modifies N7 of adenosine.[14] Nuclease T_1 cleaves the backbone 3′ to guanosine and nuclease V_1 cleaves the backbone of stacked nucleotides, particularly those in RNA helices.[12]

The extent of reactivity for selected residues monitored by structural mapping agrees well with the overall folding transition monitored by CD (Figure 4(a)). Some nucleotides change structure only in the U-to-I_{eq} transition, while other residues change structure only in the I_{eq}-to-N transition. Additional residues exhibit changes in protection in both transitions (Figure 3).

The overall protection pattern indicates that two structural modules are formed largely in the I_{eq} intermediate: the helices involved in the four-way junction (P7–P10) and the tertiary interactions of the structure involving J11/12 and J12/11 loops (residues 185–197 and 216–225). The helices in the four-way junction are formed largely in the U-to-I_{eq} transition, although P9 undergoes a slight rearrangement in the native transition. The extent of nuclease protection for the J11/12 and J12/11 loops reveals that this non-canonical module is formed in the I_{eq} intermediate (Figure 4(b)). The core of the S-domain, however, does not form until the I_{eq}-to-N transition. The formation of the basal T-loop of P10.1 (U175–A179) in the core is particularly striking in the I_{eq}-to-N transition (Figure 4(c)). Lastly, while the P12 and P10.1 helices are present in I_{eq}, the TL–receptor interaction between them is not made until the I_{eq}-to-N transition (Figure 4(d)).

The size and shape of the native and I_{eq} intermediate structures are determined by SAXS.[15] The size is characterized by the radius of gyration (R_g), whereas the global shape is derived from the pair-distance distribution function, $P(r)$, which is the sum of all mass pair distances in the molecule. Both the R_g and the $P(r)$ functions for the native structure obtained from SAXS are in good agreement with those calculated from the crystal structure (Figure 4(e), inset). R_g for the I_{eq} intermediate is 41.1(±1.1) Å, which is 8.8(±1.2) Å

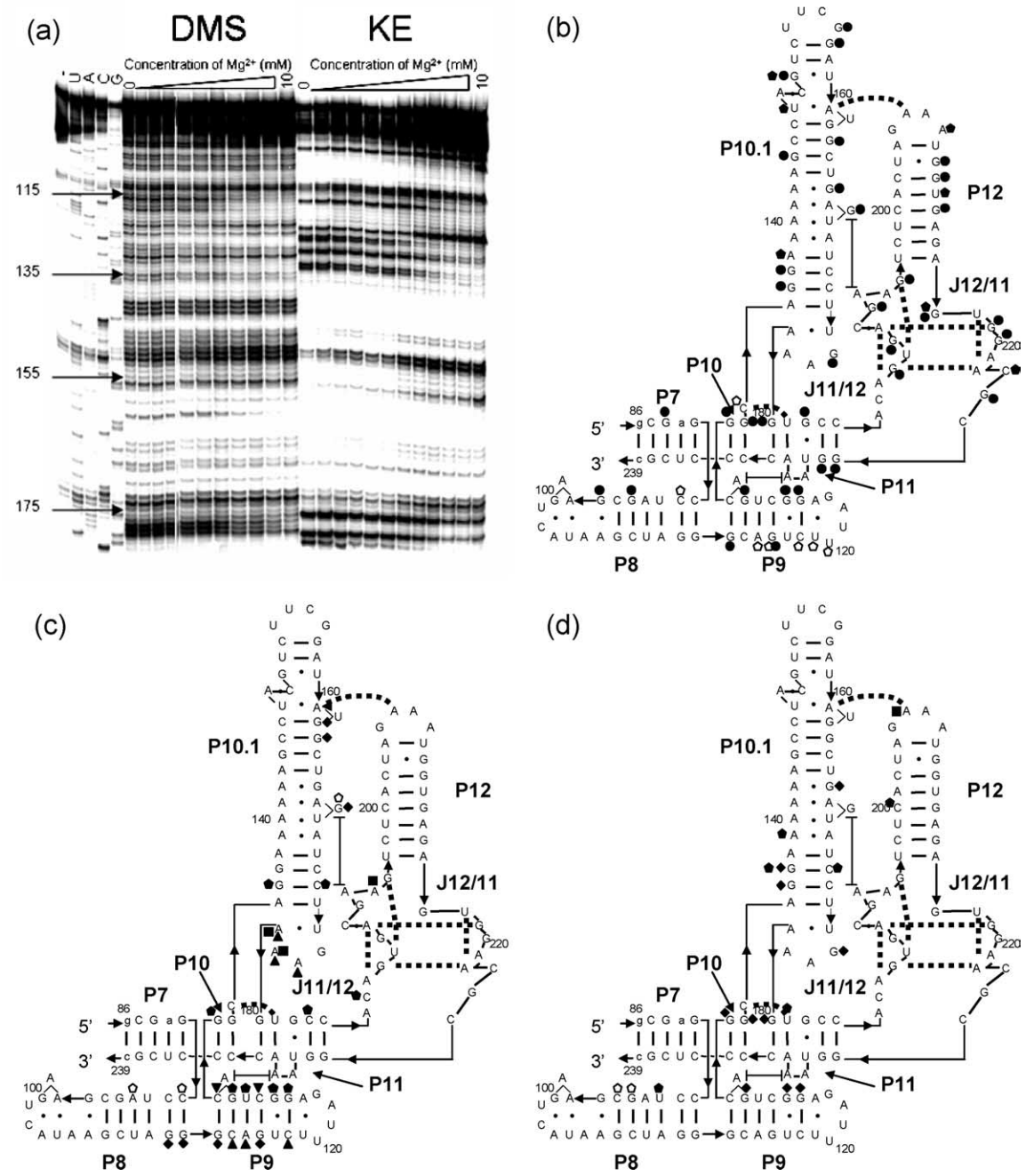

Figure 3. Chemical and nuclease mapping of the S-domain. (a) Chemical modification by DMS (left) or KE (right) over a range of concentrations of Mg^{2+} from 0–10 mM. (b) Nucleotides protected (⬟ V_1; ⬢ T_1) and exposed (⬠ V_1) only in the U-to-I_{eq} transition. (c) Nucleotides protected (⬟ V_1; ♦KE; ▲DMS; ■DEPC) and exposed (⬠ V_1) only in the I_{eq}-to-N transition. (d) Nucleotides protected and exposed (⬟, ⬠ V1; ♦KE; ■DEPC) in both transitions.

larger than the R_g of the experimentally measured native structure. $P(r)$ analysis indicates that I_{eq} has a more extended shape, with a maximum dimension of 135 Å, *versus* 110 Å for the native structure (Figure 4(e), continuous line). As expected, the U state is much more extended than either of the other two states (Figure 4(e)).

Molecular modeling

The site-resolved structural mapping patterns along with the size and shape information were applied as constraints to generate molecular models of the I_{eq} intermediate (Figure 5) using the modeling software MANIP.[16] The modeling process included both global and local arrangements of modules and specific nucleotides. The crystal structure of the S-domain serves as a starting point for developing a model of the intermediate.[8]

In the first stage, the structural mapping data are used to identify the modules that may be rearranged and those that should be left largely

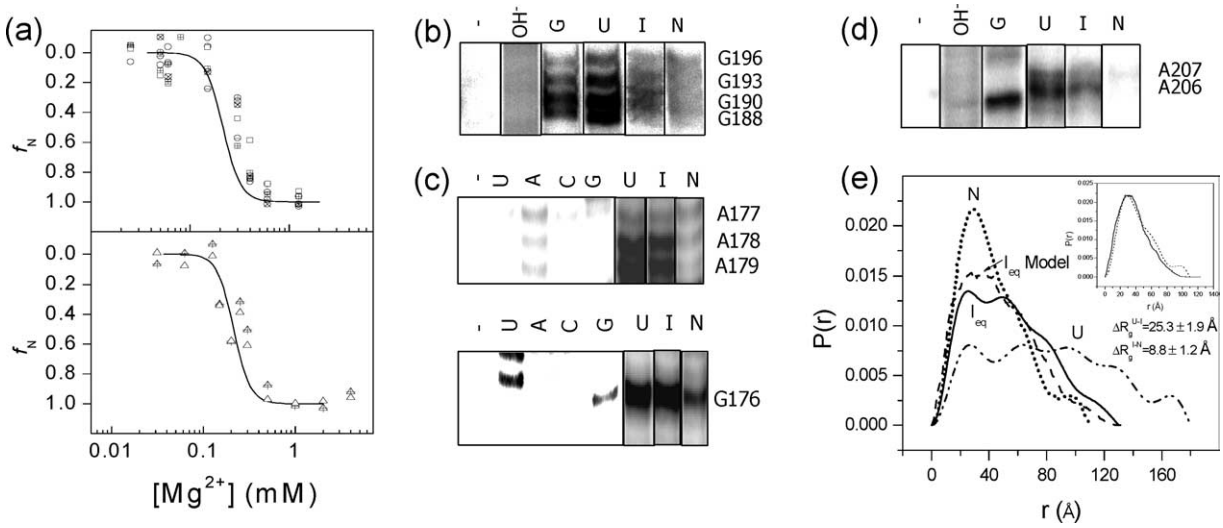

Figure 4. Structural information for I_{eq}. (a) I_{eq}-to-N transition monitored by KE modification at nucleotide 111 (○), 112 (⊕), 113 (⊗), or 116 (⊖) and by DMS modification at nucleotide 114 (□) or 117 (⊞) (upper panel). Transition monitored by digestion with nuclease V_1 (lower panel) at nucleotide 132 (△) or 137 (⏃). The overlay of the CD fit at 275 nm (—) in both panels demonstrates the agreement of the spectroscopic and biochemical probes. (b) T_1 cleavage of the J11/12 region in the U, I_{eq}, and N states shows protection in the U-to-I_{eq} transition. (c) DMS (upper) and KE (lower) modification of the T-loop in the core reveal strong protections only in N compared to U and I. (d) DEPC modification of the GAAA tetraloop shows dramatic protection in the native state. Reference lanes for (b)–(d) are: —, no nuclease; OH^-, partial alkaline hydrolysis; U, A, C, G, sequencing ladder. [$MgCl_2$] for U, I, and N in (b)–(d): U (0 mM), I (0.5×K_{Mg}), N (>5×K_{Mg}). (e) Pair-distribution functions determined by SAXS (⋯, N; —,I_{eq}; • • – • •, U) reveal compaction of I_{eq} state. $P(r)$ calculated from the I_{eq} model (- - -) is shown. The inset shows the overlay of native state $P(r)$ functions from SAXS (⋯) and from the crystal structure (—).

intact. Accordingly, the J11/12 module is kept native-like along with most of the helices in the four-way junction, except P9, which undergoes a local rearrangement in the I_{eq}-to-N transition. The helices involved in the TL–receptor interaction are formed largely in I_{eq}, although the TL–receptor interaction itself is absent.

In the second stage, the native structure is divided into modular pieces near the core to allow for movement and adjustment of the individual helices using the modeling program. In order to generate a structure with an increased R_g value (ΔR_g) that approaches the experimentally determined ΔR_g value between the native and the I_{eq} structure (8.8(±1.2) Å or 7.6–10.0 Å), selected tertiary interactions are removed:

(1) The TL–receptor interaction between helices P10.1 and P12 is broken; R_g increases by 1.6 Å (Figures 5 and 6(a)).
(2) P10.1 is further rotated while leaving the rest of the molecule in its native conformation; R_g increases by a maximum of 4.7 Å (Figures 5 and 6(a)).
(3) To obtain an additional increase in R_g, the four-way junction and the arm containing P11-J11/12-P12 (P12 arm) are adjusted. In the four-way junction, P9 is sterically unhindered but

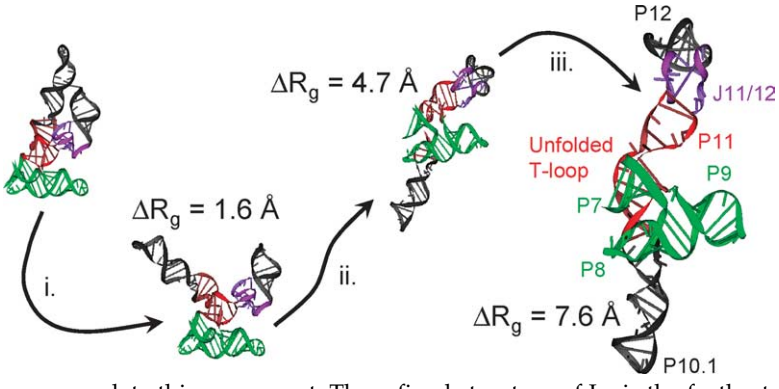

Figure 5. Sequential I_{eq} modeling using (R_g as a constraint. (i) The TL–receptor interaction was disrupted and P10.1 was rotated away from its native position. This rotation achieved an increase in R_g of 1.6 Å, as compared to the native structure. (ii) P10.1 pivoted to a larger angle increased ΔR_g to 4.7 Å. (iii) The P12 arm was adjusted to a ΔR_g value similar to that obtained *via* SAXS experiments for the I_{eq}-to-N transition. The four-way junction was rotated commensurately to accommodate this movement. The refined structure of I_{eq} is the furthest on the right and has ΔR_g=7.6 Å as compared to the native structure. Prominent features include the extended P10.1 and P12 helices, the folded non-canonical J11/12 module, and the expanded nucleotides U175–A179 in the core.

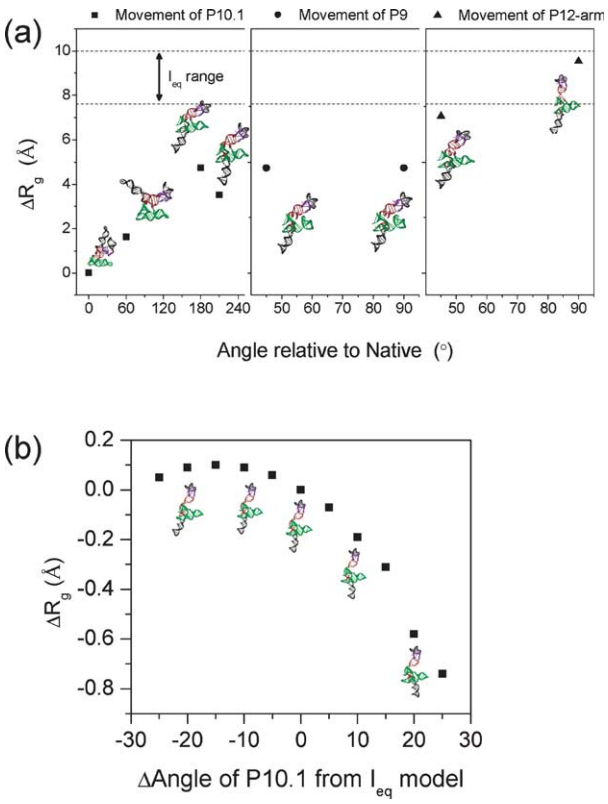

Figure 6. Detailed, sequential manipulation during modeling. (a) Left panel: ΔR_g obtained by varying the position of P10.1. The maximum increase in ΔR_g was 4.7 Å, occurring at a 180° angle from its native position. Middle panel: The orientation of P9 was manipulated following P10.1 adjustment. Extensive movement of P9 did not increase the value of R_g. P9 was therefore left in a native-like orientation relative to P8 in subsequent modeling. Right panel: Movement of the P12 arm following P10.1 adjustment alone achieved $\Delta R_g = 9.5$ Å, similar to that determined experimentally. (b) The value of R_g changed by less than 1 Å within a 50° range in the angle of P10.1 relative to its orientation in I_{eq}.

its rotation leaves R_g unchanged (Figure 6(a)). Therefore, further modeling does not include any adjustments of P9. The position of P10.1 alongside P7 and P8 of the four-way junction precludes the movement of most of the junction. The P12 arm, however, is adjustable, and its movement increases ΔR_g to 9.5 Å, within the experimental range for ΔR_g (8.8($\pm$1.2) Å).

In summary, in order to achieve a ΔR_g value that approaches the experimental value, modular rearrangements in accordance with the mapping data require that both P10.1 and the P12 arm extend in opposite directions. Following adjustment of the backbone dihedral angles to values observed in crystal structures, the structure has $\Delta R_g = 7.6$ Å (Figure 5), which is at the lower end of the experimental values.

The possible alternative structures with this value of ΔR_g are very limited. The only plausible option involves increasing the amount of single-stranded regions at the expense of helical structures, for example, P10.1. Nuclease protection patterns argued against this possibility. Hence, the global arrangement with P10.1 and P12 splayed out in opposite directions is the major structural event consistent with the SAXS and structural mapping data.

The reorientation of the P10.1 and J11/12-P12 regions from their native positions results in rearrangement of other areas. Various linking sequences are adjusted manually and oriented appropriately in accordance with structural mapping data. Particular emphasis is devoted to nucleotides U175–A179 at the base of P10.1, and A231–C234 mediating the P10/P11 stack.

On the basis of the arrangement of these linking regions and the proximity of various nucleotide bases, two non-native base-pairs are introduced into the model, G176-C134 and A135-U175, consistent with structural mapping data. Additionally, the movement of P11 disrupts the A130/A230 stacking interaction and the intercalation of A229 between G133 and C134, which occur in the native structure. Due to the structural arrangement in the I_{eq}, we model A229 as stacking inside the P11 helix between A231 and G182. Minor adjustments are made in the orientation of the four-way junction to accommodate the rearrangement of linking regions with the core in order to relieve steric constraint.

Validation of the model

The I_{eq} model is examined in detail for further consistency against the global and site-specific experimental information including shape (SAXS), surface area burial (urea-dependence), and site-resolved information (nuclease and chemical mapping). The extended nature of both P10.1 and P12 is necessary to obtain the experimental value of ΔR_g (Figures 5 and 6(a)). Comparing $P(r)$ functions of the I_{eq} calculated from the model and from SAXS (Figure 4(e)) indicates that the shape of the model approximates that measured by experiment.

The slight dearth of long vectors for the model suggests that the model could be improved by adjustments, e.g. straightening P12 out even further. This adjustment would increase the R_g by $\sim$1 Å. Perturbations around the present orientation of P10.1 result in only a slight change in R_g as the angle is varied (Figure 6(b)). Although the I_{eq} model is quite extended, various regions may be flexible enough to allow for further increase in R_g. Such regions may include the loop E motif in the middle of P10.1 (residues 139–142 and 166–170), which has a strong dependence on Mg^{2+} for stabilization,[1,17] and the area surrounding the bulged nucleotides in P11. In both of these regions, only a minimal amount of structural mapping information could be resolved. Hence, our model may be best viewed as a lower bound on the extended nature of the intermediate.

Another global validation of the model is the urea *m*-value, which correlates with the change in

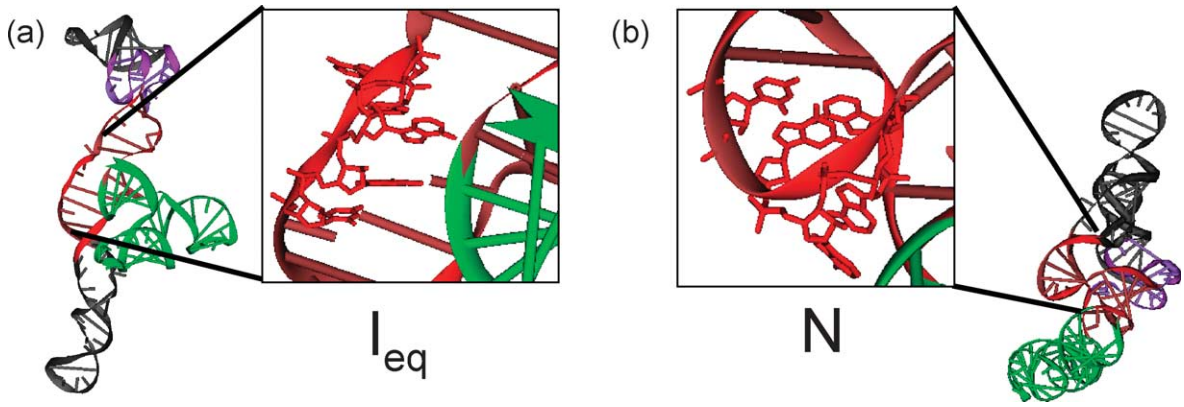

Figure 7. Highlights of the five nucleotide linker U175–A179 in the core. In the I_{eq} structure (a) this region is extended whereas in (b), the native structure, it forms a compact T-loop. In this loop, U175 and A179 forms a reversed Watson–Crick pair. A177 and A178 flip out of the loop forming an A-minor motif that docks in the P7/P10 stack (green and red junction).

accessible surface area in the I_{eq}-to-N transition. The total amount of surface area buried is roughly proportional to the number of nucleotides in the RNA,[4] giving a total *m*-value for the S-domain of 5.4 kcal mol^{-1} M^{-1}. The I_{eq} model has ~70% of secondary and tertiary contacts present in the native structure. Hence, 30% of the surface area burial occurs in the I_{eq}-to-N transition, giving a predicted *m*-value (1.6 kcal mol^{-1} M^{-1}) similar to that determined experimentally (1.5 kcal mol^{-1} M^{-1}).

Site-specific protection patterns also support the structure of the intermediate. The T-loop (U175–A179) at the base of P10.1 is extended in the I_{eq} structure (Figure 7(a)). This open conformation allows for increased modification by chemical probes as compared to the compact, H-bonded native structure. A177 and A178 are largely solvent-exposed in the I_{eq} structure but in the native conformation, they interact with base-pairs G90-C235 and G132-C234 *via* an A-minor motif. Consistently, DMS and KE modify this region strongly in both the U and I states but not in the native structure (Figure 4(c)).

Discussion

This work describes the identification of the extensive tertiary interactions in the major thermodynamic folding intermediate. The majority of the S-domain molecules adopt this structure at a solution condition located between two distinct folding transitions, U-to-I_{eq} and I_{eq}-to-N. The same folding transitions were observed when monitored by chemical and nuclease mapping, CD spectroscopy and SAXS.

Features of the I_{eq} structure and the I_{eq}-to-N transition

A general expectation is that RNA folding intermediates largely lack tertiary structures. However, we find that the J11/12-J12/11 junction is formed in I_{eq} despite the fact that this structure is composed entirely of tertiary interactions. This region yielded strong protection factors when probed with nucleases T_1 and V_1 in the U-to-I_{eq} transition, but protection is largely absent from the I_{eq}-to-N transition. Despite the presence of extensive tertiary interactions, however, this loop–loop interaction can still be considered to be local because the loops are the exit strands of the same hairpin. This result suggests that the topology and connectivity of the tertiary modules may dictate the order of structure formation in the equilibrium folding of an RNA.

A striking implication of the model is that the P10.1 helix must significantly reorient upon folding to the native structure. This reorientation occurs upon the formation of the U-shaped T-loop, which is fully extended in the intermediate (Figure 7(a) and (b)). This T-loop, formed by U175–A179, is present at the junction connecting two tertiary structure modules, the central core and the TL–receptor. The I_{eq}-to-N transition involves the rearrangement of this extended five nucleotide region to a U-shaped turn (Figure 7). Formation of the turn swings the P10.1 helix into a native-like position.

The T-loop region plays a pivotal role in the folding transition through the formation of the U175·A179 reversed Watson–Crick base-pair. The formation of this base-pair enables A177 and A178 to form an A-minor motif with the P7/P10 stack in the core. This docking event, in concert with the formation of the U175·A179 pair, orients the P10.1 helix properly for its TL–receptor (P12-P10.1) interaction. Additionally, the bulged G168 in the P10.1 helix is positioned for stacking on A194 of J11/12. The TL–receptor and the interaction of these bulged, stacked bases serve as latches holding this upper module in the native conformation.

Our model also implies structural changes in the

four-way junction and the core occur in the I_{eq}-to-N transition. In the I_{eq} structure, the P8/P9 stack is slightly tilted away from the junction where P7–P10 converge. This tilt orients P8 closer to the inverted P10.1, in agreement with the increase in chemical and nuclease protection in these two helices in I_{eq} (Figure 3: U93, G137, A138, G168). The final transition requires that these stacked helices undergo a slight adjustment with respect to the core.

The formation of the core in the I_{eq}-to-N transition depends on the folding of the T-loop at the base of P10.1 and the return of the P12 arm to its native arrangement, with P11 stacking on the P10 helix. As the core and four-way junction adjust toward their native conformations, A130 and A230 are positioned to stack on top of each other, serving as a tertiary latch connecting the core to the four-way junction. In the I_{eq} structure, these two bases are separated by about ~12 Å, four times greater than their separation in the native structure. The completion of the core module involves the docking of the T-loop into the P7/P10 junction. Additionally, A229, which stacks within P11 in I_{eq}, will intercalate between G133 and C134 of P10. The C134·(U181-A231) triplet interaction in the core also is formed in the transition to the N state.

Comparisons to other RNAs

The role of the T-loop in the core of the S-domain is similar to the J5/J5a loop of the P4-P6 domain of *Tetrahymena* group I intron. Mutational analysis has shown that the J5/J5a internal loop serves as a hinge for the folding of the P4-P6 domain.[18] The intermediate in P4-P6 folding is thought to be an extended conformation of two, largely helical, modules. Similarly, in the S-domain folding intermediate and in the P4-P6 domain, the GAAA tetraloop and its receptor are in an extended conformation. In S-domain folding, the single five nucleotide linker region serves two purposes; it provides flexibility to permit movement of P10.1 and stability (tertiary contacts with P7-P10 core). In contrast, the internal loop of J5/J5a of the P4-P6 domain does not interact with other regions of the RNA, and appears to be a passive contributor to the tertiary fold, serving only as a flexible hinge rather than stabilizing the native fold directly.[18]

Other elements of S-domain folding are seen in the two folding transitions of the hammerhead ribozyme.[7,19] At low [Mg^{2+}], an intermediate is populated in the hammerhead folding pathway having domains II and III stacked on top of each other, while domain I is positioned at a small angle relative to domain III. The transition to the N state incorporates a rearrangement of the single-stranded region that connects domain I to domain II. This adjustment changes the orientation of domain I dramatically and completes the formation of the core of the hammerhead ribozyme. Similarly, in S-domain folding, the compaction of U175–A179 changes the orientation of P10.1 dramatically and permits the formation of the core.

Metal ion binding and equilibrium folding

In the crystal structure of the S-domain, a number of metal ions are bound with inner sphere coordination to the RNA.[8] In the I_{eq} structure, metal ions are likely to remain in the native-like regions such as J11/12 junctions, P10.1 and P8 helical elements, but they are unlikely to remain in the unfolded regions (Figure 8). A subset of these metal ions that are absent from the intermediate are likely to play an important role in the Mg^{2+}-dependent folding transition to the native structure.

Three scenarios are plausible for how metal ions could facilitate the I_{eq}-to-N transition. First, metal ions may mediate long-range tertiary interactions directly through explicit RNA–metal–RNA contacts at an interface between structural elements. Second, metal ions may indirectly stabilize RNA structures involved in the tertiary interaction.[20] For example, a metal-assisted interaction is present in the interaction of the A-rich bulge with the P4 helix and P5abc junction in the P4-P6 domain from group I intron.[21]

We propose a third modality, involving the folding of a binding site or pocket in concert with the formation of long-range interactions in distal regions of the molecule. For example, the metal-induced folding of the centrally located T-loop region occurs along with the TL–receptor docking and the base-stacking between P10.1 and J11/12. Metal ion binding in the S-domain core stabilizes the formation of the TL–receptor by increasing the local concentration of its two constituents. Metal ion binding "acting at a distance", linking local and

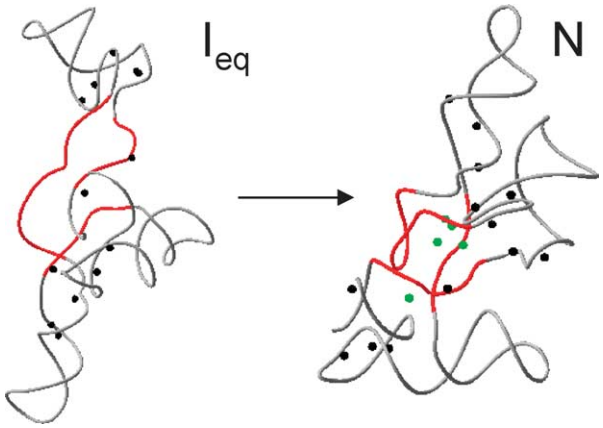

Figure 8. Depiction of metal ions bound within the core in the transition to the native structure. At the resolution of the crystal structure (3.1 Å), five metal-binding sites (green) that are present in the core module (red) are likely to be absent from the unfolded regions of I_{eq}. These metal sites likely are important in facilitating the I_{eq}-to-N folding transition by stabilizing the T-loop, which in turn increases the local concentration of the tetraloop and its receptor, enhancing their association.

non-local events, is likely to be a major part of the origin of the folding cooperativity observed for tertiary RNAs.

The action of metal ion binding in a core region and stabilizing a structural adjustment of other regions appears to be general, being observed in the folding of the hammerhead ribozyme and the four-way junction of the hairpin ribozymes.[7,22] In each of these ribozymes, a metal ion binds near the core junction of the helices, thus promoting the correct orientation of the peripheral elements so that their apical loops are positioned for enzymatic activity.

The detailed structural model of the S-domain can serve as a foundation for investigating other questions of the I_{eq}-to-N transition. Folding kinetic studies from U to N indicate the appearance of the same I_{eq} state, followed by two additional intermediates (N.B., T.P. & T.R.S., unpublished results). The most conspicuous structural elements that are absent from I_{eq} are the peripheral TL–receptor and the T-loop in the core. Modeling both kinetic intermediates that follow I_{eq} will provide structural insights into the movements involved in the rate-limiting step. One possibility is that the formation of the core precedes, and is held in place by the TL–receptor interaction.

Concluding remarks

The equilibrium folding of the S-domain from the intermediate to the native state is a multifaceted process that involves movements of the four-way junction, the core, the TL–receptor, and base-stacking interactions. The structure of the intermediate hints at the impending intricate network of interactions to be formed in the folding transition to the native structure. While full atomic-level simulations of RNA folding processes for large RNAs are currently unfeasible, a more reasonable aim is to understand the thermodynamic and kinetic properties of the folding transition that leads to the native structure.[23] However, to date, the necessary high-resolution models of intermediate states have not been available to achieve these goals.[24] Our nucleotide-level model provides detailed structural information required for future studies, such as calculating the stability of the native structure using methods such as non-linear Poisson–Boltzmann formalism.[23,24]

Materials and Methods

Sample preparation

The S-domain was transcribed by *in vitro* transcription using phage T7 RNA polymerase,[25] purified on denaturing polyacrylamide gels, and stored in water at −20 °C. The unfolded (U) state was prepared by heating in 20 mM Tris–HCl (pH 8.1) at 85 °C for 2 min followed by 5 min at room temperature.

Mg^{2+}-titration monitored by circular dichroism

To RNA (0.3 μM) in the U state, Mg^{2+} solutions were added using a Hamilton titrator connected to a Jasco J715 spectrometer.[11] The waiting time was up to 60 min due to slow folding kinetics (unpublished results), and the data acquisition time was 30 s.

The two Mg^{2+}-dependent transitions observed at CD_{260} were described according to a cooperative binding model:

$$U \overset{K_{UI},n_{UI}}{\leftrightarrow} I_{eq} \overset{K_{IN},n_{IN}}{\leftrightarrow} N$$

where n and K are the Hill coefficient and Mg^{2+}-midpoint of the transitions. The signal, S, is fit for simplicity with K_{IN} and n_{IN} denoted as K_{Mg} and n, according to:

$$S = \frac{S_u + S_{I_{eq}}([Mg^{2+}]/K_{UI})^{n_{UI}} + S_N([Mg^{2+}]/K_{UI})^{n_{UI}}([Mg^{2+}]/K_{Mg})^n}{1 + ([Mg^{2+}]/K_{UI})^{n_{UI}} + ([Mg^{2+}]/K_{UI})^{n_{UI}}([Mg^{2+}]/K_{Mg})^n} \quad (1)$$

The CD_{275} trace has one transition and was fit to:

$$S = \frac{S_{I_{eq}} + S_N([Mg^{2+}]/K_{Mg})^n}{1 + ([Mg^{2+}]/K_{Mg})^n} \quad (2)$$

Mg^{2+}-titration monitored by partial digestion by nucleases T_1 and V_1

A sample of 0.3 μM (5′ ^{32}P)-labeled S-domain in the U state was mixed with $MgCl_2$ in individual tubes, each with a different [Mg^{2+}]. The mixture was incubated at 50 °C for 5 min to allow a rapid approach to equilibrium. Samples were then returned to room temperature for ~5 min before the addition of nuclease. Nuclease reaction was performed at 37 °C for 5 min at 0.1 unit/μl of T_1 and 0.32 munit/μl of V_1. The solutions were mixed with an equal volume of 9 M urea, 100 mM EDTA and loaded immediately onto denaturing polyacrylamide gels. After electrophoresis, the gels were quantified using a phosphorimager. For nuclease V_1, increased cleavage is interpreted as an increase in structure formation. Decreases in cleavages by V_1 could be a result of either structural disorganizations or inaccessibility of the nuclease to the nucleotides due to structural sterics.

Mg^{2+}-titration monitored by chemical modification

A sample of 0.3 μM RNA in the U state was mixed with $MgCl_2$ in individual tubes, each at a different [Mg^{2+}]. The mixture was incubated at 37 °C for 10 min. Reactions were initiated by adding 1 μl of DMS (from 400× concentrated stock) or 1 μl of kethoxal (KE) (from 400× concentrated stock) and performed at 37 °C for 5 min. Reactions were quenched by adding 40 μl of stopping mixture (for DMS: 62.5 mM β-mercaptoethanol, 1 μg of *Escherichia coli* tRNA, 45 mM potassium acetate, 180 mM KCl; for KE: 75 mM sodium acetate (pH 6.0), 6.25 mM boric acid, 1 μg of *E. coli* tRNA, 45 mM potassium acetate, 180 mM KCl). RNA was then precipitated twice with ethanol, and stored in water. Modification was detected by reverse transcription using AMV reverse transcriptase and a (5′-^{32}P)-labeled primer complementary to nucleotides 221–239 of the S-domain (5′ GCGAGGGGTTTACCGCGTT). Modification by DEPC was carried out as described.[26]

Protection factors for chemical and nuclease mapping experiments were calculated from data obtained from the

phosphorimager. Based on CD titration data, certain lanes were chosen where populations of U, I_{eq}, and N states would dominate. Protected regions are defined as having a protection factor of >1.5 among U and I_{eq} or I_{eq} and N. Exposures are defined as regions having a protection factor of <0.67.

SAXS

Experiments were performed at the BioCAT beamline at the Advanced Photon Source at the Argonne National Laboratories.[27] Samples (6.3 μM, 0.3 mg/ml) at various [Mg^{2+}] (20 mM Tris–HCl, pH 8.1) were incubated at 50 °C for 10 min, followed by incubation at room temperature for 5 min. The sample was loaded into a capillary tube controlled by a Hamilton titrator thermostatically controlled at 37 °C. Data collection was performed under continuous-flow conditions to avoid radiation damage.

The equilibrium SAXS measurements covered a range of [Mg^{2+}] from 0–10 mM. The 0.4 mM $MgCl_2$ condition best correlated with the I_{eq} state, as indicated by nuclease mapping at this [RNA] (data not shown). This condition has [Mg^{2+}] about fourfold higher than that obtained from CD or structural mapping, due to the higher [RNA] used in SAXS (6.3 μM *versus* 0.3 μM). The observed shift, 0.3 mM Mg^{2+} at 6.3 μM S-domain, was within the expected range (0.23–0.46 mM, assuming every two to four ribosephosphate molecules bind an extra magnesium ion[27–29]).

Molecular modeling

To model I_{eq}, secondary structure modules as well as individual nucleotide bases were manipulated using MANIP,[16] and changes were refined with the restraint refinement program NUCLIN-NUCLSQ.[30] Adjustment of the backbone dihedral angles to values observed in crystal structures is done automatically by NUCLIN-NUCLSQ. The process was followed as described.[16,31]

Calculations of $P(r)$ functions from crystal structures and predicted models were conducted using the program XTL.[32]

Acknowledgements

This work was supported by a NIH grant (GM57880). We thank the reviewers for their insightful comments. We thank Dr Jaby Jacob for dedicated assistance in collection of the SAXS data. Use of the Advanced Photon Source was supported by the US Department of Energy, Basic Energy Sciences, Office of Science, under contract no. W-31-109-ENG-38. BioCAT is a National Institutes of Health-supported Research Center RR-08630.

References

1. Tinoco, I., Jr & Bustamante, C. (1999). How RNA folds. *J. Mol. Biol.* **293**, 271–281.
2. Brion, P. & Westhof, E. (1997). Hierarchy and dynamics of RNA folding. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 113–137.
3. Sosnick, T. R., Fang, X. & Shelton, V. M. (2000). Application of circular dichroism to study RNA folding transitions. *Methods Enzymol.* **317**, 393–409.
4. Shelton, V. M., Sosnick, T. R. & Pan, T. (1999). Applicability of urea in the thermodynamic analysis of secondary and tertiary RNA folding. *Biochemistry*, **38**, 16831–16839.
5. Shelton, V. M., Sosnick, T. R. & Pan, T. (2001). Altering the intermediate in the equilibrium folding of unmodified yeast tRNA(Phe) with monovalent and divalent cations. *Biochemistry*, **40**, 3629–3638.
6. Butcher, S. E. & Burke, J. M. (1994). Structure-mapping of the hairpin ribozyme. Magnesium-dependent folding and evidence for tertiary interactions within the ribozyme-substrate complex. *J. Mol. Biol.* **244**, 52–63.
7. Bassi, G. S., Murchie, A. I., Walter, F., Clegg, R. M. & Lilley, D. M. (1997). Ion-induced folding of the hammerhead ribozyme: a fluorescence resonance energy transfer study. *EMBO J.* **16**, 7481–7489.
8. Krasilnikov, A. S., Yang, X., Pan, T. & Mondragón, A. (2003). Crystal structure of the specificity domain of ribonuclease P. *Nature*, **421**, 760–764.
9. Krasilnikov, A. S. & Mondragon, A. (2003). On the occurrence of the T-loop RNA folding motif in large RNA molecules. *RNA*, **9**, 640–643.
10. Fang, X. W., Golden, B. L., Littrell, K., Shelton, V., Thiyagarajan, P., Pan, T. & Sosnick, T. R. (2001). The thermodynamic origin of the stability of a thermophilic ribozyme. *Proc. Natl Acad. Sci. USA*, **98**, 4355–4360.
11. Fang, X., Pan, T. & Sosnick, T. R. (1999). A thermodynamic framework and cooperativity in the tertiary folding of a Mg^{2+}-dependent ribozyme. *Biochemistry*, **38**, 16840–16846.
12. Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J. P. & Ehresmann, B. (1987). Probing the structure of RNAs in solution. *Nucl. Acids Res.* **15**, 9109–9128.
13. Krol, A. & Carbon, P. (1989). A guide for probing native small nuclear RNA and ribonucleoprotein structures. *Methods Enzymol.* **180**, 212–227.
14. Jaeger, J. A., SantaLucia, J., Jr & Tinoco, I., Jr (1993). Determination of RNA structure and thermodynamics. *Annu. Rev. Biochem.* **62**, 255–287.
15. Cantor, C. & Schimmel, P. (1980). *Biophysical Chemistry: Part II*, W.H. Freeman, New York.
16. Massire, C. & Westhof, E. (1998). MANIP: an interactive tool for modelling RNA. *J. Mol. Graph. Model.* **16**, 197–205. 255–257.
17. Correll, C. C., Wool, I. G. & Munishkin, A. (1999). The two faces of the *Escherichia coli* 23 S rRNA sarcin/ricin domain: the structure at 1.11 Å resolution. *J. Mol. Biol.* **292**, 275–287.
18. Szewczak, A. A. & Cech, T. R. (1997). An RNA internal loop acts as a hinge to facilitate ribozyme folding and catalysis. *RNA*, **3**, 838–849.
19. Hammann, C., Norman, D. G. & Lilley, D. M. (2001). Dissection of the ion-induced folding of the hammerhead ribozyme using 19F NMR. *Proc. Natl Acad. Sci. USA*, **98**, 5503–5508.
20. Klein, D. J., Moore, P. B. & Steitz, T. A. (2004). The contribution of metal ions to the structural stability of the large ribosomal subunit. *RNA*, **10**, 1366–1379.
21. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E. *et al.* (1996). Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science*, **273**, 1678–1685.

22. Walter, F., Murchie, A. I. H. & Lilley, D. M. J. (1998). Folding of the four-way RNA junction of the hairpin ribozyme. *Biochemistry*, **37**, 17629–17636.
23. Sosnick, T. R. & Pan, T. (2003). RNA folding: models and perspectives. *Curr. Opin. Struct. Biol.* **13**, 309–316.
24. Misra, V. K. & Draper, D. E. (2002). The linkage between magnesium binding and RNA folding. *J. Mol. Biol.* **317**, 507–521.
25. Loria, A. & Pan, T. (1996). Domain structure of the ribozyme from eubacterial ribonuclease P. *RNA*, **2**, 551–563.
26. Qin, H., Sosnick, T. R. & Pan, T. (2001). Modular construction of a tertiary RNA structure: the specificity domain of the *Bacillus subtilis* RNase P RNA. *Biochemistry*, **40**, 11202–11210.
27. Fang, X., Littrell, K., Yang, X., Henderson, S. J., Siefert, S., Thiyagarajan, P. *et al*. (2000). Mg(2+)-dependent compaction and folding of yeast tRNA(Phe) and the catalytic domain of the *B. subtilis* RNase P RNA determined by small-angle X-ray scattering. *Biochemistry*, **39**, 11107–11113.
28. Stein, A. & Crothers, D. M. (1976). Equilibrium binding of magnesium(II) by *Escherichia coli* tRNAfMet. *Biochemistry*, **15**, 157–160.
29. Bina-Stein, M. & Stein, A. (1976). Allosteric interpretation of Mg2+ binding to the denaturable *Escherichia coli* tRNAGlu2+. *Biochemistry*, **15**, 3912–3917.
30. Westhof, E., Dumas, P. & Moras, D. (1985). Crystallographic refinement of yeast aspartic acid transfer RNA. *J. Mol. Biol.* **184**, 119–145.
31. Masquida, B. & Westhof, E. (2005). Modelling the architecture of structured RNAs within a modular and hierarchical framework. In *Handbook of RNA Biochemistry* (Hartmann, R. K., Bindereif, A., Schön, A. & Westhof, E., eds), pp. 536–544, Wiley, Weinheim.
32. Thiyagarajan, P., Henderson, S. J. & Joachimiak, A. (1996). Solution structures of GroEL and its complex with rhodanese from small-angle neutron scattering. *Structure*, **4**, 79–88.

Edited by J. Doudna

(Received 17 May 2005; received in revised form 27 June 2005; accepted 5 July 2005)
Available online 25 July 2005