Structural inference of native and partially folded RNA by high-throughput contact mapping

Rhiju Das*[†], Madhuri Kudaravalli*, Magdalena Jonikas[‡], Alain Laederach[§], Robert Fong[¶], Jason P. Schwans*, David Baker[‖], Joseph A. Piccirilli[¶], Russ B. Altman[‡§], and Daniel Herschlag*[,**]

Departments of *Biochemistry, [‡]Bioengineering, and [§]Genetics, Stanford University, Stanford, CA 94305; [¶]Department of Chemistry, University of Chicago, Chicago, IL 60637; and [‖]Department of Biochemistry, University of Washington, Seattle, WA 98195

Edited by Alan M. Lambowitz, University of Texas, Austin, TX, and approved December 18, 2007 (received for review September 24, 2007)

The biological behaviors of ribozymes, riboswitches, and numerous other functional RNA molecules are critically dependent on their tertiary folding and their ability to sample multiple functional states. The conformational heterogeneity and partially folded nature of most of these states has rendered their characterization by high-resolution structural approaches difficult or even intractable. Here we introduce a method to rapidly infer the tertiary helical arrangements of large RNA molecules in their native and non-native solution states. Multiplexed hydroxyl radical (·OH) cleavage analysis (MOHCA) enables the high-throughput detection of numerous pairs of contacting residues via random incorporation of radical cleavage agents followed by two-dimensional gel electrophoresis. We validated this technology by recapitulating the unfolded and native states of a well studied model RNA, the P4–P6 domain of the *Tetrahymena* ribozyme, at subhelical resolution. We then applied MOHCA to a recently discovered third state of the P4–P6 RNA that is stabilized by high concentrations of monovalent salt and whose partial order precludes conventional techniques for structure determination. The three-dimensional portrait of a compact, non-native RNA state reveals a well ordered subset of native tertiary contacts, in contrast to the dynamic but otherwise similar molten globule states of proteins. With its applicability to nearly any solution state, we expect MOHCA to be a powerful tool for illuminating the many functional structures of large RNA molecules and RNA/protein complexes.

hydroxyl radical | molten globule | *Tetrahymena* ribozyme | two-dimensional gel

The discoveries of catalytic RNAs, silencing RNAs, riboswitches, and a panoply of functional RNA molecules have sweeping implications for our views of evolution from an early "RNA World" and for the potential of structured RNAs to act in roles beyond the simple transmission of information laid out in the Central Dogma of Molecular Biology (1). The functions of these RNAs in primitive and modern life are being elucidated at an explosive pace. Nevertheless, a deep understanding of these fundamental biopolymers and their biological roles requires structural portraits of their functional states, and, in this respect, progress has been slow.

Our understanding of RNA structure has greatly lagged behind that of protein structure: compared with nearly 40,000 protein structures in the Protein Data Bank, there are currently <1,000 experimentally determined RNA structures, most of which are small fragments (2). High-resolution approaches using NMR spectroscopy (NMR) and x-ray crystallography have the potential to describe RNA structure at the atomic level, but have been considerably hampered by numerous factors, including limited chemical shift dispersion, the large sizes of structured RNAs, and the poor behavior of RNA at high concentrations.

Further enriching and complicating the modeling of RNA behavior is the seemingly pervasive tendency of RNA to form alternative secondary and tertiary structures. Indeed, cycling between partially structured folding intermediates or alternative conformations appears to be crucial to the versatility and function of RNAs in their cellular milieu (3–7). Furthermore, many RNAs and RNA/protein complexes function through structures in which only a portion of the RNA molecule adopts a discrete three-dimensional fold (4, 8, 9). Overall, understanding the fundamental properties, folding pathways, and energetics of functional RNA molecules requires information about intermediates in the folding process and not just a single final folded state (10). The partially folded nature of these intermediates renders their characterization by x-ray crystallography and NMR particularly difficult and possibly intractable; structural information for these states, even at a modest spatial resolution, would be a boon for the understanding of RNA behavior.

The difficulty of inferring RNA structure is underscored by continuing investigations of the first large globular RNA whose native structure was determined by crystallography (11, 12), the 158-nucleotide P4–P6 domain of the *Tetrahymena thermophila* group I intron. Despite its status as one of the best-studied model systems in RNA folding (see, e.g., refs. 13–17), the structure of a partially folded state of this RNA, discovered in studies at high monovalent ion concentrations (17), remains unknown. This state, attained under unusual solution conditions, mimics a kinetic and thermodynamic folding intermediate observed in folding pathway of the P4–P6 RNA and its mutants under physiological conditions (17). As illustrated by the insightful studies of molten globule states in protein folding (see, e.g., ref. 18), a complete picture of RNA structure and energetics will be greatly aided by understanding such folding intermediates and their analogs.

We present a high-throughput contact-mapping technique that permits the rapid inference of the helical arrangement of RNA conformations and apply it to three states of the P4–P6 RNA: the unfolded state, the folded state, and the hitherto unsolved partially folded state. Although currently at modest resolution, these structural portraits clearly illustrate physical distinctions between the three states and illuminate informative similarities and differences relative to protein folding intermediates. High-throughput contact mapping holds immediate promise for revealing previously unavailable structural information for the many folded and partially folded states of large RNAs.

Results

Contact Mapping by Multiplexed Hydroxyl Radical Cleavage Analysis. Without crystallographic diffraction data, pairwise distance constraints are the most valuable type of data for structural inference,

Author contributions: R.D., A.L., D.B., R.B.A., and D.H. designed research; R.D., M.K., and M.J. performed research; R.D., R.F., J.P.S., and J.A.P. contributed new reagents/analytic tools; R.D. analyzed data; and R.D. and D.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[†]Present address: Department of Biochemistry, University of Washington, Seattle, WA 98195.

**To whom correspondence should be addressed. E-mail: herschla@stanford.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0709032105/DC1.

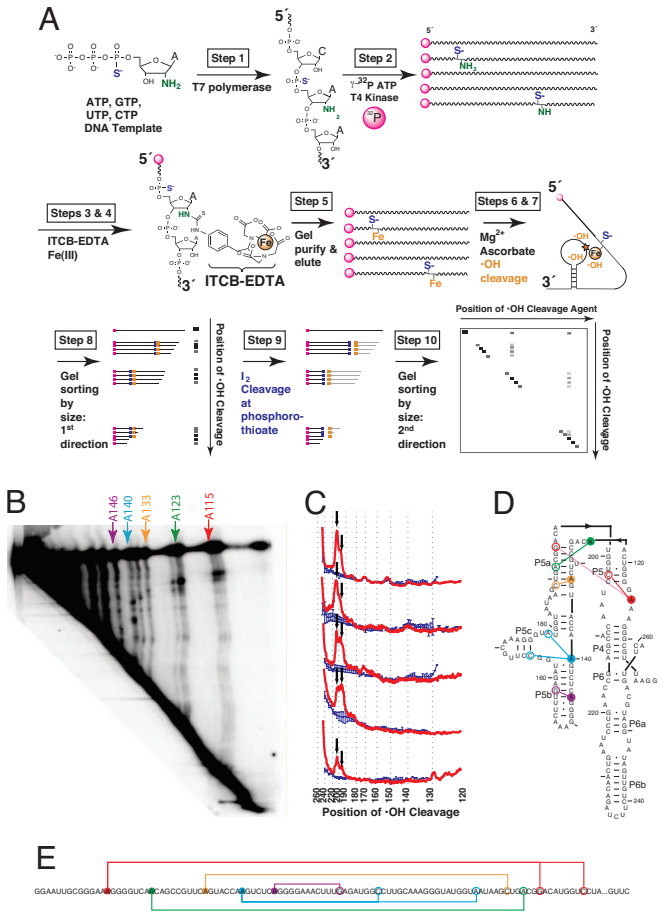

Fig. 1. Overview of high-throughput RNA contact mapping by Multiplexed ·OH Cleavage Analysis (MOHCA). (*A*) Schematic of the preparation of RNA containing randomly incorporated phosphorothioate-tagged radical sources followed by cleavage pattern readout via two-dimensional gel electrophoresis. (*B*) Sample MOHCA gel image for the P4–P6 domain of the *Tetrahymena* ribozyme in an unfolded state containing secondary structure but not tertiary structure, with cleavage agents tethered to adenosine residues and radiolabels at the 5′ end of the RNA. Cleavage patterns due to numerous radical sources are visible, with five cleavage agents highlighted. A white triangle has been placed over the bottom left of the image over radioactive signals from a different experiment that was included in the PhosphorImager. (*C*) Cleavage profile (red) due to Fe(II) tethered at A115 for the gel in *B* (top) and four replicate gels with independently prepared samples. Background estimates (see *SI Text*) are shown as blue error bars, and reproducible ''hits'' are marked with black arrows. (*D* and *E*) Constraints inferred from the gel in *B* and supported by replicates, overlaid on the native secondary structure (*D*) and the sequence (*E*) of P4–P6; for clarity, only five cleavage agents are marked (filled circles) connected to representative cleaved residues (open circles); data from additional agents (e.g., A113 and A114) are visible in *B* and tabulated in SI Table 1.

underlying both NMR structure determination (19) and lower resolution manual modeling approaches using phylogenetic covariance (20, 21). In the absence of assigned NMR spectra or phylogenetic constraints, several laboratories have demonstrated that accurate residue–residue distance constraints can be derived from hydroxyl radical (·OH) cleavage patterns generated by Fe(II) atoms tethered to specific locations in the nucleic acid backbone (see, e.g., refs. 22, 23). However, the labor and expense required to tether cleavage agents to each of the hundreds of residues in a large RNA have prevented broad and intense investigation of RNA structures by cleavage mapping.

We present a protocol for incorporating cleavage agents randomly throughout an RNA followed by a two-dimensional gel electrophoresis procedure that reads out the position of both the cleavage agent and the residues cleaved by that particular agent. Combining the resulting data with new computational tools then generates three-dimensional structures of folding states. The technique, which we term Multiplexed ·OH Cleavage Analysis (MOHCA), is briefly summarized below and in Fig. 1*A*. A complete step-by-step protocol is given in the supporting information (SI) *Text*.

Incorporation of the cleavage agent into a large RNA is accomplished by *in vitro* transcription of the RNA in the presence of a modified nucleotide triphosphate (step 1, Fig. 1*A*). The nucleotide contains a 2′-amino modification, which allows attachment of the hydroxyl radical cleavage agent, and a phosphorothioate group, which allows specific cleavage at a later stage (24); this "tagging" strategy has been widely used for high-throughput mutational analysis (25). After random incorporation of the modified nucleotide at a frequency of approximately one modification per RNA (step 1), the RNA is end-radiolabeled (step 2) and the cleavage agent, an Fe-EDTA chelate, is tethered to the 2′-amino groups that are present throughout the RNA molecules (steps 3 and 4). This pool of RNA molecules is purified (step 5) and folded to the desired state (step 6), after which localized radical generation is initiated by reduction of Fe(III) to Fe(II) and then quenched after several minutes (step 7). The cleavage products are first separated by length via polyacrylamide gel electrophoresis (step 8). This pattern indicates where each RNA was cleaved but not the location of the responsible cleavage agent. However, this information is not lost, as the modified nucleotide that harbored each cleavage agent remains physically a part of one of the two fragments resulting from each cleavage event. We take advantage of the ability of iodine to induce backbone scission at the phosphorothioate (step 9), allowing the position of the cleavage agent to be read out by subsequent separation of the initial lane of RNA molecules in a second, orthogonal dimension (step 10). Thus, the first gel dimension identifies the cleavage position and the second gel dimension identifies the tether position of the agent that caused each cleavage, revealing a pair of residues whose sugar moieties were brought within ≈25 Å of each other by the RNA's three-dimensional structure.

High-Throughput Contact Mapping for the P4–P6 RNA Model System. To validate this method, we first investigated the unfolded state of the P4–P6 RNA, populated in a moderate concentration of monovalent ions (50 mM Na-MOPS, pH 7.0). Under these solution conditions, the RNA attains much or all of its secondary structure (17, 26) in the form of Watson–Crick base pairs, but the resulting double helices do not dock together into a compact tertiary structure (13, 14, 17). MOHCA analysis should therefore reveal signals corresponding to the RNA's secondary structure but not its tertiary structure. A representative MOHCA gel image for this unfolded state is shown in Fig. 1*B*, with cleavage agents incorporated at adenosine positions.

The strongest feature on the gel image, along the diagonal, is composed of RNA fragments that were cleaved by hydroxyl radicals but did not undergo phosphorothioate cleavage upon treatment with iodine.[††] Above this diagonal, punctate cleavage features, or "hits," from at least 10 different adenosine-tethered cleavage agents are visible, with patterns for five agents marked on the figure. These MOHCA cleavage products are less well-resolved than typical patterns from solution hydroxyl radical cleavage, potentially due to heterogeneity in the RNA fragments after the cleavage agent excises itself from the RNA (R.D. and D.H., unpublished data). Nevertheless, this method's weak signal-to-noise ratio can be ad-

[††]The diagonal stripe includes cleavage products by cleavage agents 3′ to the cleaved residue, as well as fragments that undergo desulfurization at phosphorothioates (85–90%) rather than strand scission (10–15%) upon iodine treatment (ref. 24 and R.D., M.K., and D.H., unpublished data).

BIOPHYSICS

dressed in standard ways, through systematic image annotation procedures, the establishment of statistical criteria for assessing the significance of each hit, and the performance of multiple replicates.

Accurate annotation of the cleavage agents associated with each feature is straightforward, based on matching the horizontal spacing of the patterns to the known sequence of adenosines (A113, A114, A115, A122, etc.) in the RNA. To accurately annotate the vertical dimension of the gel, corresponding to the positions of radical-cleaved residues, separate experiments were carried out using an enzymatically digested sample that produces a two-dimensional grid of fiducial markers (SI Figs. 5 and 6), analogous to reference ladders in traditional one-dimensional RNA footprinting. Imperfections in RNA migration across the gel and the intrinsic spread of the cleavage products can complicate analysis; we thus extended a previously described one-dimensional computational approach (SAFA; ref. 27) to align two-dimensional gels and permit rapid, semiautomated annotation (SI Figs. 5–7).

After this alignment and annotation, cleavage profiles were calculated by integrating image intensity down through each pattern corresponding to cleavage by a single cleavage agent (Fig. 1*C* and SI Fig. 7), as in one-dimensional footprinting analysis (27, 28), and background estimates were obtained from neighboring profiles (SI Figs. 8–10). All experiments have been carried out minimally in triplicate (with an average of six replicates; Fig. 1*C*), and straightforward, well defined criteria have been established to assess the statistical significance of each hit (*SI Text*). For the gel in Fig. 1*B*, observed MOHCA features agree well with contacts expected in the secondary structure of the P4–P6 RNA between residues that are quite far apart in the RNA's primary structure (Fig. 1 *D* and *E*), illustrating how a single MOHCA experiment (with replicates) can produce a wealth of accurate structural data.

Using these procedures for gel annotation, signal identification, and statistical appraisal, we obtained complete MOHCA data sets for the P4–P6 RNA in unfolded conditions, native conditions (10 mM $MgCl_2$), and non-native high monovalent ion conditions (2 M NaCl). To maximize the number of distance constraints, separate gels have been carried out with cleavage agents introduced at adenosine, cytidine, and uridine positions[‡‡] and with the radiolabel at 5′ and 3′ ends. Each two-dimensional gel gave resolvable and reproducible cleavage patterns for cleavage agents extending 120 residues into this 158-nt RNA from the radiolabeled end. A total of 60, 63, and 77 hits were identified for the unfolded, high NaCl, and native states, respectively. All observed signals and their statistical significances are given in SI Fig. 11 and SI Table 1.

The confident cleavage signals observed in these experiments fell into three major classes. The majority of hits were present in all three states (see, e.g., Fig. 2, first and second columns). Some hits were observed for the native state and for the high monovalent state but were not detected in the unfolded state (Fig. 2, third column). A final major class of signals was uniquely observed in the native state (Fig. 2, third and fourth column). These signals define an experimentally inferred contact map for each state (Fig. 3 *A*–*C*).

Internal Consistency and Accuracy of High-Throughput Contact Mapping. Due to the large number of data points available, high-throughput contact mapping allows excellent tests of internal consistency. Most regions of the experimentally inferred contact map are given support from the cleavage patterns of different radical sources (overlap of filled symbols in Fig. 3 *A*–*C*). Further, many contacts are supported by symmetry-related pairs of hits in which the cleavage agent and the cleaved residue are interchanged (overlap of open and filled symbols in Fig. 3 *A*–*C*).

Beyond internal consistency, the data for the previously well characterized unfolded and native states provide excellent checks

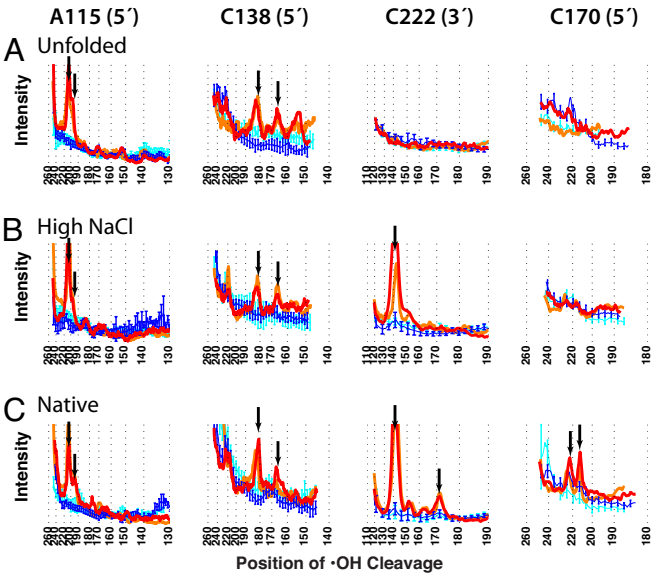

Fig. 2. Representative MOHCA cleavage profiles illustrating different classes of "hits" observed in the unfolded state (*A*), the compact non-native state in high monovalent ion concentrations (*B*), and the native state (*C*) of the P4–P6 RNA. Cleavage agents depicted are, from left to right, A115 (5′ radiolabeled RNA), C138 (5′), C222 (3′), and C170 (5′). Two replicates are shown (red and orange) with corresponding background estimates as error bars in blue and cyan, respectively. Statistically significant hits, supported by additional replicates (SI Table 1), are marked with black arrows. Some smaller peaks are also visible but do not pass the stringent statistical criteria used in this work (*SI Text*).

on the accuracy of distance constraints inferred from MOHCA. First, all hits observed in the unfolded state correspond to residue pairs brought close together by the native secondary structure but not to pairs brought together by tertiary contacts, as expected (Fig. 3 *A* and *D*) (29; §§). Second, the residue pairs inferred to be neighboring in the native state based on the MOHCA data are also in contact in the crystal structure of P4–P6 (Fig. 3 *C* and *F*). As expected from prior studies on tethered radical cleavage agents (22, 23, 30), the majority of the MOHCA hits correspond to distances closer than 25 Å in the crystal structure (SI Fig. 12). Taking this distance as a cutoff for a "correct" hit, the rate of false positives is estimated to be quite small, 7% for the unfolded state and 4% for the native state (*SI Text*). These extra hits are reproducible and pass stringent criteria for statistical significance. They may reflect alternative structures induced by a small subset of incorporated cleavage sources or may signal intrinsic RNA solution dynamics not captured in the reference crystal structure. The latter effects have been suggested by chemical cross-linking measurements (31, 32) and are supported by localized discrepancies of footprinting data for the RNA in solution with footprinting data for the RNA in crystallized form (33) and with the solvent-accessible surface area computed for the P4–P6 crystal structure (SI Fig. 13).

Assuming the same distance cutoff of 25 Å, the true positive rates for reconstituting the expected neighboring pairs in unfolded state and the native state are 29% and 26% (*SI Text*). This rate of contact recovery is much higher than previous, low-throughput approaches that yield only a handful of constraints (22, 34–36) but does not reach 100% due to the experimental inaccessibility of peripheral regions of the contact map (24% of long-range hits fall outside the

[‡‡]Experiments with cleavage agents at guanosines could not be carried out due to the scarcity and difficulty of synthesis of 2′-NH_2-guanosine α-thio-triphosphate.

[§§]At its present resolution, the MOHCA data cannot resolve the single-nucleotide shift in helix "register" expected for the P5abc loop based on NMR studies (29). However, the data rule out dramatic differences in secondary structure between the unfolded and native state, consistent with prior chemical protection studies.

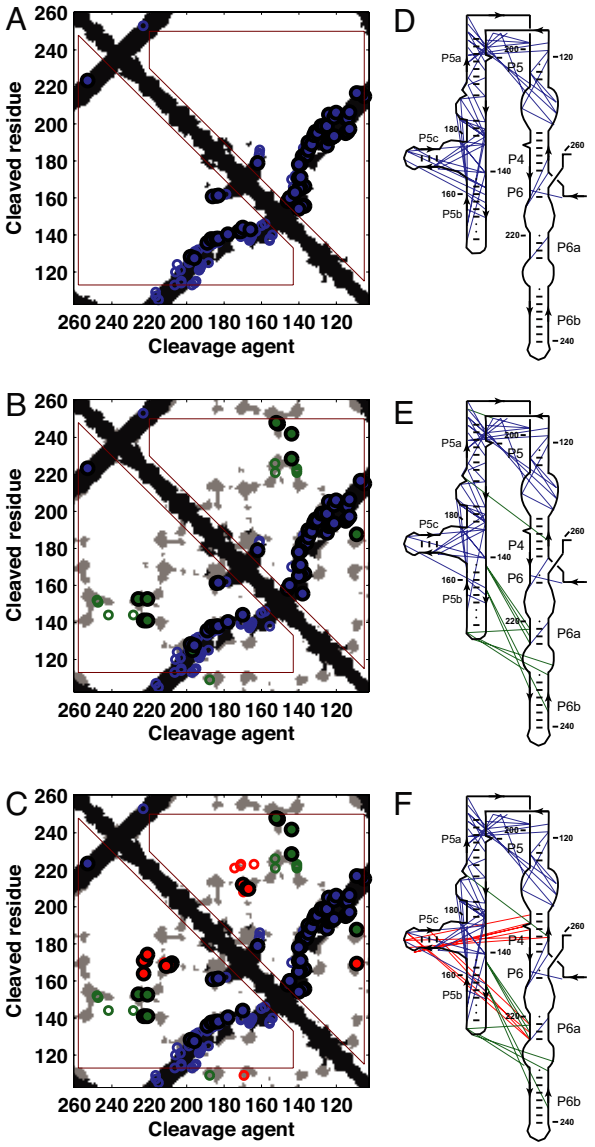

Fig. 3. MOHCA distance constraints inferred for three states of the P4–P6 RNA. (*A–C*) Contact maps derived from experimental MOHCA hits (filled symbols) for the unfolded state, the state in high monovalent ion concentrations, and the native state, respectively. Hits are color-coded based on whether they are present in all three states (blue), in only the high monovalent and native state (green), or solely in the native state (red). To facilitate visual comparison of symmetry-related hits, the data are also plotted with the cleavage agent and cleaved residue interchanged (open symbols). For comparison to the crystal structure, the hits are overlaid on a grayscale map of residue pairs brought within 25 Å by secondary structure (black) and by tertiary structure (gray) (see *SI Text* and SI Figs. 11 and 12). Brown trapezoids approximately demarcate regions of the contact map in which hits are readily distinguished on MOHCA gels. Note that, on the horizontal axis, the residue numbering increases from right to left to aid comparison to Fig. 1*B*. (*D–F*) The same contacts as *A–C*, respectively, as lines connecting residues in the P4–P6 RNA secondary structure.

boundaries in Fig. 3 *A–C*) and to limitations in signal-to-noise. In addition, a significant fraction (≈25%; see SI Fig. 13) of P4–P6 residues are protected from cleavage by radicals in solution and are thus presumably sterically inaccessible. The lower incorporation of 2′-amino modified nucleotides at certain residues may further explain the absence of some hits (data not shown). These effects, along with the variation in observed MOHCA signal strength between replicates (see, e.g., Figs. 1*C* and 2), currently preclude a quantitative correlation between increased signal strength and decreased distance (SI Fig. 12). Despite this limitation, conservatively assuming that each observed MOHCA hit defines a soft distance constraint (≈25 Å) between a pair of residues allows modeling of tertiary structure at modest resolution, as described next.

Models of RNA Tertiary Structure. To visualize and interpret the information from MOHCA, structural reconstructions have been carried out with a recently developed automated *de novo* fold prediction method based on Fragment Assembly of RNA (FARNA) (37) constrained by the MOHCA data; similar results have been obtained with an independently developed structure simulation method, the Nucleic Acid Simulation Tool (NAST; *SI Text* and SI Fig. 14). These tertiary structure models assume no crystallographic information except the native secondary structure (11, 29), as this information is often available from phylogenetic data; models generated with the quite accurate secondary structure available before the crystal structure (13) gave similar results (data not shown). Structural models satisfying the MOHCA constraints were generated and clustered, with the largest cluster presented as the model ensemble (*SI Text*). The P4–P6 RNA data under the three assayed conditions gave clearly distinguishable models (Fig. 4).

As a calibration of the structural inference, the MOHCA-constrained ensemble for the native state was compared with the crystal structure (Fig. 4 *A* and *B*). The x-ray structure and MOHCA model agree well. The C4′ rmsd of the cluster center and the mean cluster coordinates from the crystal structure are 13.3 and 11.1 Å, respectively. This accuracy is consistent with a modeling precision estimate of 13.1 Å, obtained from the mean RMSD of the cluster center to other members of the same cluster (Fig. 4*C*). Although these values appear poor in comparison to the 4–6 Å resolution reached by medium-resolution models of protein structures (38), RNA residues are nearly three times the molecular weight of protein residues and RNA double helices are more than twice as wide as protein alpha helices. Thus, these models at ≈13-Å resolution have positioned the secondary structural elements within approximately half the diameter of an RNA double helix. The MOHCA data are critical for obtaining this subhelical resolution; the probability of reaching this accuracy by chance is estimated to be $<10^{-4}$, based on FARNA reconstructions assuming secondary structure constraints but no MOHCA data (SI Fig. 15).

Although the model of the unfolded state is quite extended (Fig. 4 *F* and *G*), the shape of the compact, non-native state under high monovalent ion concentrations is similar to the native state ensemble at first glance (Fig. 4 *D* and *E*). However, the backbone RMSD of the structure is 17 Å from the crystal structure. Most strikingly, the position of one helix (P5c, shown in green and marked with arrow) is no longer inside the core of the molecule; the tip of the helix (residue 170) moves ≈40 Å, well beyond the modeling error. The result is consistent with a working model of this non-native state in which a central metal ion core involving P5c is absent (17); further, the ensemble's shape is in excellent agreement with experimental solution x-ray scattering data (see SI Fig. 16). Unexpectedly, the modeling converged to structures with helix P5c favoring an orientation at the "front" of the molecules rather than forming a completely disordered cloud. This predicted orientation, observed in the majority of the FARNA models for this non-native state (SI Fig. 17), may be physically enforced by rigid helical connections to the well ordered tetraloop tertiary contact (cyan/orange in Fig. 4 *D* and *E*) and is potentially testable by extensive molecular ruler measurements (39, 40).

Discussion

A Molten-Globule-Like State of RNA? In recent years, the discoveries of collapsed, non-native states for several large RNA molecules

BIOPHYSICS

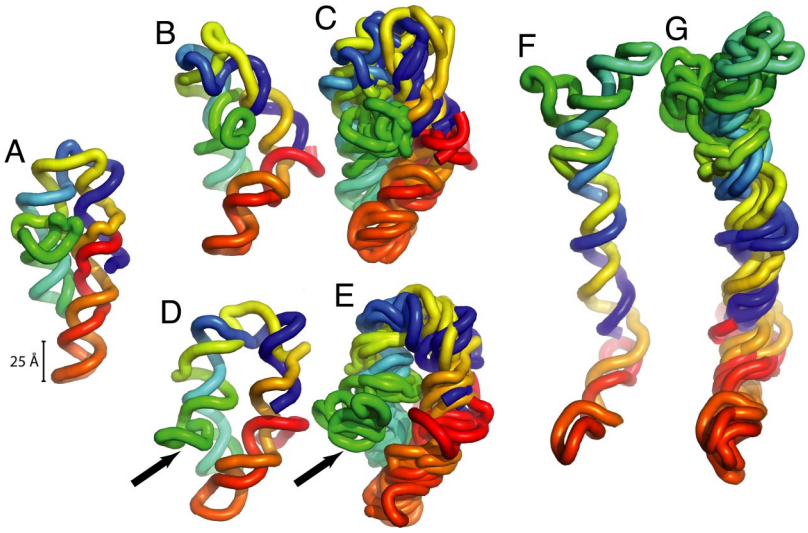

Fig. 4. Structural models of three states of the P4–P6 RNA using Fragment Assembly of RNA (FARNA) constrained by MOHCA. The model for the native state [(*B*); rainbow coloring from blue (5′) to red (3′)] agrees with the crystal structure (*A*) within the present resolution (≈13 Å) of the method, as estimated from the mean pair-wise RMSD of the ensemble (*C*). Models for the P4–P6 RNA at high monovalent ion concentrations (*D* and *E*) are native-like except for the conformation of the P5c helix (green; marked with arrow). Models for the unfolded state at low monovalent ion concentrations (F and G) gave highly extended configurations. Models in *B*, *D*, and *F* are cluster centers from FARNA ensembles; *C*, *E*, and *G* show five randomly chosen examples of 50 ensemble members (see SI Fig. 14 for full ensembles).

(41–44) have prompted tentative analogies to the molten globule states of proteins (45), whose roles in thermodynamic and kinetic folding pathways have been intensively studied (18). The medium-resolution structural portrait presented here of the P4–P6 RNA at high monovalent ion concentrations allows for a more precise comparison and contrast of a compact, non-native RNA state to its potential protein counterpart. Two of the defining characteristics of protein molten globules are the possession of significant secondary structure and near-native compactness; these properties are clearly attained by the P4–P6 state at high monovalent ion concentrations (Fig. 4 *D* and *E* and ref. 17). Indeed, the long-range base pairs defining the double helices of this compact state are indistinguishable from the native state at the resolution of MOHCA (blue lines in Fig. 3 *E* and *F*). Furthermore, the compactness of protein molten globules partly stems from the presence of some long-range native-like tertiary interactions. For example, the N and C terminus α-helical regions are associated with each other in the alpha lactalbumin native state and molten globule (18). Analogously, a long-range tertiary interaction between an RNA tetraloop and its receptor (cyan/orange in Fig. 4 *D* and *E*) appears to be responsible for the compactness of the nonnative state of the P4–P6 RNA (17).

Despite these similarities, the state of the P4–P6 RNA at high monovalent ion concentrations lacks one defining characteristic of protein molten globules, the disorder (and concomitant solvent accessibility) present throughout the entire molecule. First, the compact non-native RNA state appears to have its complete secondary structure stably formed (Fig. 3*E* and refs. 17 and 26), in contrast to protein molten globules in which only alpha helical and not beta strand secondary structure is typically detected (18). Further, based on the extent of protection from hydroxyl radicals (17) and the observation of well-defined MOHCA constraints (Fig. 3), the tetraloop/receptor tertiary contact in the P4–P6 RNA appears just as specifically ordered in its high monovalent state as it does in the native state (compare Fig. 4 *C* and *E*) in contrast to the dynamic side-chain interactions seen throughout protein molten globules (18). Finally, the MOHCA-derived model suggests that this well ordered tertiary contact may induce a preferred non-native conformation for other, solvent-accessible regions of the RNA molecule (green helix, Fig. 4*E*).

These results support a picture in which RNA self-assembly is more modular than protein self-assembly (see also refs. 28 and 46), with RNA thermodynamic intermediates favoring the complete formation of a fraction of the tertiary native contacts and possessing partial order elsewhere. This picture differs strikingly from the traditional view of protein molten globules as highly dynamic throughout the entire molecule, suggesting fundamental differences in the folding intermediates and the basic physical properties of proteins and RNA. More examples will be needed to test the generality of this comparison. It will thus be of interest to obtain structural models of the non-native states identified in other RNAs in high monovalent ion concentrations and other solution conditions (42–44).

Toward High-Throughput RNA Structure Modeling. Multiplexed ·OH Cleavage Analysis (MOHCA) appears capable of providing rapid and accurate three-dimensional models of the multiple solution states of a large RNA molecule. The resolution of this structural inference method does not currently reach the distance scale of nucleotides (≈6 Å), much less the atomic resolution offered, in favorable cases, by crystallography or NMR. Nevertheless, the ≈13 Å accuracy and precision of the models is similar to the best resolution achieved for large RNAs by manual modeling methods with local and long range phylogenetic constraints (e.g., for the *Tetrahymena* group I ribozyme; refs. 21 and 47). Cryoelectron microscopy can also reach this resolution for large complexes, although the resulting density maps require separately solved structures for assigning helix identities (7). Even at this modest resolution, such models have been useful guides in the design and interpretation of experiments to understand ribozyme catalysis (see, e.g., refs. 48 and 49) and have led to powerful biological insights into the multiple functional states of the ribosome (7). The MOHCA work presented herein further demonstrates the value of a medium-resolution model for a compact, non-native state of the P4–P6 RNA; the structure permits an approximate analogy with protein molten globule intermediates to be delineated and then qualified.

The structures of metabolite-sensing riboswitches with and without their ligands (4) and the possibly topologically tangled conformations of misfolded ribozymes (50) have been refractory to traditional structural approaches and are excellent future targets for MOHCA. Further, given that the majority of residue-residue contacts determined by MOHCA are induced by Watson–Crick base pairs (Fig. 3), the technology has much to offer in characterizing functionally important switches in RNA secondary structure, e.g., within the dynamic and complex spliceosome. The step-by-step experimental protocol given in the Supporting Information, the widespread availability of the required gel electrophoresis equipment, and free analysis software and structural inference software should aid the application of the MOHCA technology to an even wider array of targets by the experimental RNA community.

Beyond its application to new RNAs, an exciting future challenge for the MOHCA method is to improve its structural resolution to

the length scale of individual residues. Advances in the MOHCA experimental procedure, such as coloading of fluorescently labeled internal controls, are expected to enhance the accuracy of two-dimensional gel annotation, giving constraints at the nucleotide resolution routinely achieved with traditional one-dimensional gels. Immediately, MOHCA can be used as a high-throughput screen; follow-up experiments with a subset of tethered cleavage agents introduced singly by ligation can confirm these hits and more precisely read out their cleavage profiles. Furthermore, high-throughput data on the conformational flexibility and the solvent accessibility of the RNA backbone and bases are readily available from chemical footprinting at nucleotide resolution. The generation of models that automatically fit these types of "one-dimensional" data are computationally challenging, and we thus invite structural modelers to download the MOHCA and solution radical footprinting data for the P4–P6 states probed in this work (SI Tables 2–4) to develop and calibrate higher resolution structural inference algorithms. With continuing advances in both its experimental and computational aspects, we hope that the MOHCA technology will become widely useful for revealing native folds of new RNAs and alternative topologies of well known RNAs, both critical stages in understanding the fundamental behavior of this ubiquitous biopolymer.

Methods

Multiplexed Hydroxyl Radical Cleavage Analysis. The new contact-mapping approach (MOHCA; Fig. 1*A*) makes use of widely available molecular biology tools. A full description of sample preparation and two-dimensional gel electrophoresis, adapted from an in-house protocol, is given in *SI Text*.

Gel Alignment and Annotation. Alignment of MOHCA two-dimensional gel images, extraction of cleavage profiles, and estimation of background and statistical significance of observed hits are described in Supporting Method Sections 2–4. The MATLAB software developed for this analysis, along with a short tutorial with data presented in this manuscript, is available at http://faculty.washington.edu/rhiju/MOHCA/.

Structural Inference. Structural models were prepared by using Fragment Assembly of RNA (FARNA) or the Nucleic Acid simulation Tool, guided by MOHCA constraints, as is described in *SI Text*. FARNA is part of the Rosetta package, freely available to academic users (http://rosettacommons.org); a tutorial is available at http://faculty.washington.edu/rhiju/MOHCA/. NAST is freely available as part of SimTK (https://simtk.org/home/nast).

ACKNOWLEDGMENTS. We thank M. Greenfeld for assistance in compiling the MOHCA experimental protocol; M. Dreyfus (Centre National de la Recherche Scientifique, Paris) and O. Uhlenbeck (Northern University, Evanston, IL) for the gift of the P226L T7 polymerase; J. Lipfert, M. Tyka, and members of the Herschlag group for comments on the manuscript; and the users of Rosetta@home for enabling rapid structural inference of three states of P4–P6. This work was funded by National Institutes of Health Program Project Grant P01-GM-066275 (to D.H.), SimBios National Center for Biomedical Computing Grant U54 GM072970 (to R.B.A.), an Abbott Laboratories Stanford Graduate Fellowship (to R.D.), a Jane Coffin Childs Fellowship (to R.D.), National Library of Medicine Training Grant LM-07033 (to M.J.), and a Damon Runyon Fellowship (to A.L.).

1. Gesteland RF, Cech TR, Atkins JF, eds (2006) *The RNA World: The Nature of Modern RNA Suggests a Prebiotic RNA World* (Cold Spring Harbor Lab Press, Plainview, NY).
2. Berman H, Henrick K, Nakamura H, Markley JL (2007) The worldwide Protein Data Bank (wwPDB): Ensuring a single, uniform archive of PDB data. *Nucleic Acids Res* 35:D301–D303.
3. Stroynowski I, Kuroda M, Yanofsky C (1983) Transcription termination *in vitro* at the tryptophan operon attenuator is controlled by secondary structures in the leader transcript. *Proc Natl Acad Sci USA* 80:2206–2210.
4. Winkler WC, Breaker RR (2003) Genetic control by metabolite-binding riboswitches. *Chembiochem* 4:1024–1032.
5. Noller HF (2005) RNA structure: Reading the ribosome. *Science* 309:1508–1514.
6. Staley JP, Guthrie C (1998) Mechanical devices of the spliceosome: Motors, clocks, springs, and things. *Cell* 92:315–326.
7. Mitra K, Frank J (2006) Ribosome dynamics: Insights from atomic structure modeling into cryo-electron microscopy maps. *Annu Rev Biophys Biomol Struct* 35:299–317.
8. Hentze MW, Kuhn LC (1996) Molecular control of vertebrate iron metabolism: mRNA-based regulatory circuits operated by iron, nitric oxide, and oxidative stress. *Proc Natl Acad Sci USA* 93:8175–8182.
9. Kieft JS, Zhou K, Jubin R, Doudna JA (2001) Mechanism of ribosome recruitment by hepatitis C IRES RNA. *Rna* 7:194–206.
10. Baird NJ, Westhof E, Qin H, Pan T, Sosnick TR (2005) Structure of a folding intermediate reveals the interplay between core and peripheral elements in RNA folding. *J Mol Biol* 352:712–722.
11. Cate JH, *et al.* (1996) Crystal structure of a group I ribozyme domain: Principles of RNA packing. *Science* 273:1678–1685.
12. Cate JH, Hanna RL, Doudna JA (1997) A magnesium ion core at the heart of a ribozyme domain. *Nat Struct Biol* 4:553–558.
13. Murphy FL, Cech TR (1993) An independently folding domain of RNA tertiary structure within the Tetrahymena ribozyme. *Biochemistry* 32:5291–5300.
14. Murphy FL, Wang YH, Griffith JD, Cech TR (1994) Coaxially stacked RNA helices in the catalytic center of the Tetrahymena ribozyme. *Science* 265:1709–1712.
15. Deras ML, Brenowitz M, Ralston CY, Chance MR, Woodson SA (2000) Folding mechanism of the Tetrahymena ribozyme P4–P6 domain. *Biochemistry* 39:10975–10985.
16. Silverman SK, Deras ML, Woodson SA, Scaringe SA, Cech TR (2000) Multiple folding pathways for the P4–P6 RNA domain. *Biochemistry* 39:12465–12475.
17. Takamoto K, *et al.* (2004) Principles of RNA compaction: Insights from the equilibrium folding pathway of the P4–P6 RNA domain in monovalent cations. *J Mol Biol* 343:1195–1206.
18. Arai M, Kuwajima K (2000) Role of the molten globule state in protein folding. *Adv Protein Chem* 53:209–282.
19. Varani G, Tinoco I, Jr (1991) RNA structure and NMR spectroscopy. *Q Rev Biophys* 24:479–532.
20. Levitt M (1969) Detailed molecular model for transfer ribonucleic acid. *Nature* 224:759–763.
21. Lehnert V, Jaeger L, Michel F, Westhof E (1996) New loop-loop tertiary interactions in self-splicing introns of subgroup IC, ID: A complete 3D model of the Tetrahymena thermophila ribozyme. *Chem Biol* 3:993–1009.
22. Han H, Dervan PB (1994) Visualization of RNA tertiary structure by RNA-EDTA.Fe(II) autocleavage: Analysis of tRNA(Phe) with uridine-EDTA.Fe(II) at position 47. *Proc Natl Acad Sci USA* 91:4955–4959.
23. Joseph S, Whirl ML, Kondo D, Noller HF, Altman RB (2000) Calculation of the relative geometry of tRNAs in the ribosome from directed hydroxyl-radical probing data. *RNA* 6:220–232.
24. Gish G, Eckstein F (1988) DNA and RNA sequence determination based on phosphorothioate chemistry. *Science* 240:1520–1522.
25. Strobel SA (1999) A chemogenetic approach to RNA function/structure analysis. *Curr Opin Struct Biol* 9:346–352.
26. Banerjee AR, Jaeger JA, Turner DH (1993) Thermal unfolding of a group I ribozyme: The low-temperature transition is primarily disruption of tertiary structure. *Biochemistry* 32:153–163.
27. Das R, Laederach A, Pearlman SM, Herschlag D, Altman RB (2005) SAFA: Semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments. *RNA* 11:344–354.
28. Laggerbauer B, Murphy FL, Cech TR (1994) Two major tertiary folding transitions of the Tetrahymena catalytic RNA. *EMBO J* 13:2669–2676.
29. Wu M, Tinoco I, Jr (1998) RNA folding causes secondary structure rearrangement. *Proc Natl Acad Sci USA* 95:11555–11560.
30. Whirl-Carrillo M, Gabashvili IS, Bada M, Banatao DR, Altman RB (2002) Mining biochemical information: lessons taught by the ribosome. *RNA* 8:279–289.
31. Cohen SB, Cech TR (1997) Dynamics of thermal motions within a large catalytic RNA investigated by cross-linking with thiol-disulfide interchange. *J Am Chem Soc* 119:6259–6268.
32. Gabashvili IS, Whirl-Carrillo M, Bada M, Banatao DR, Altman RB (2003) Ribosomal dynamics inferred from variations in experimental measurements. *RNA* 9:1301–1307.
33. Vicens Q, Gooding AR, Laederach A, Cech TR (2007) Local RNA structural changes induced by crystallization are revealed by SHAPE. *RNA* 13:536–548.
34. Badorrek CS, Gherghe CM, Weeks KM (2006) Structure of an RNA switch that enforces stringent retroviral genomic RNA dimerization. *Proc Natl Acad Sci USA* 103:13640–13645.
35. Tsai HY, Masquida B, Biswas R, Westhof E, Gopalan V (2003) Molecular modeling of the three-dimensional structure of the bacterial RNase P holoenzyme. *J Mol Biol* 325:661–675.
36. Spanggord RJ, Siu F, Ke A, Doudna JA (2005) RNA-mediated interaction between the peptide-binding and GTPase domains of the signal recognition particle. *Nat Struct Mol Biol* 12:1116–1122.
37. Das R, Baker D (2007) Automated *de novo* prediction of native-like RNA tertiary structures. *Proc Natl Acad Sci USA* 104:14664–14669.
38. Kryshtafovych A, Venclovas C, Fidelis K, Moult J (2005) Progress over the first decade of CASP experiments. *Proteins 61 Suppl* 7:225–236.
39. Lafontaine DA, Norman DG, Lilley DM (2002) The global structure of the VS ribozyme. *EMBO J* 21:2461–2471.
40. Capel MS, Kjeldgaard M, Engelman DM, Moore PB (1988) Positions of S2, S13, S16, S17, S19 and S21 in the 30 S ribosomal subunit of Escherichia coli. *J Mol Biol* 200:65–87.
41. Russell R, Millett IS, Doniach S, Herschlag D (2000) Small angle X-ray scattering reveals a compact intermediate in RNA folding. *Nat Struct Biol* 7:367–370.
42. Russell R, *et al.* (2002) Rapid compaction during RNA folding. *Proc Natl Acad Sci USA* 99:4266–4271.
43. Rangan P, Woodson SA (2003) Structural requirement for Mg^{2+} binding in the group I intron core. *J Mol Biol* 329:229–238.
44. Buchmueller KL, Webb AE, Richardson DA, Weeks KM (2000) A collapsed non-native RNA folding state. *Nat Struct Biol* 7:362–366.
45. Woodson SA (2000) Compact but disordered states of RNA. *Nat Struct Biol* 7:349–352.
46. Brion P, Westhof E (1997) Hierarchy and dynamics of RNA folding. *Annu Rev Biophys Biomol Struct* 26:113–137.
47. Michel F, Westhof E (1990) Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J Mol Biol* 216:585–610.
48. Michel F, Hanna M, Green R, Bartel DP, Szostak JW (1989) The guanosine binding site of the Tetrahymena ribozyme. *Nature* 342:391–395.
49. Hougland JL, Piccirilli J, Forconi M, Lee J, Herschlag D (2005) in *The RNA World*, eds Gesteland RF, Cech TR, Atkins JF (Cold Spring Harbor Lab Press, Plainview, NY), pp 133–205.
50. Russell R, *et al.* (2006) The paradoxical behavior of a highly structured misfolded intermediate in RNA folding. *J Mol Biol* 363:531–544.