J|A|C|S COMMUNICATIONS

Published on Web 11/01/2010

Discrete Structure of an RNA Folding Intermediate Revealed by Cryo-electron Microscopy

Nathan J. Baird,[†] Steven J. Ludtke,[§] Htet Khant,[§] Wah Chiu,[§] Tao Pan,*[,†,‡] and Tobin R. Sosnick*[,†,‡]

Department of Biochemistry and Molecular Biology and Institute for Biophysical Dynamics, University of Chicago, Chicago, Illinois 60637, United States, and National Center for Macromolecular Imaging, Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas 77030, United States

Received August 19, 2010; E-mail: trsosnic@uchicago.edu

Abstract: RNA folding occurs via a series of transitions between metastable intermediate states. It is unknown whether folding intermediates are discrete structures folding along defined pathways or heterogeneous ensembles folding along broad landscapes. We use cryo-electron microscopy and single-particle image reconstruction to determine the structure of the major folding intermediate of the specificity domain of a ribonuclease P ribozyme. Our results support the existence of a discrete conformation for this folding intermediate.

A major challenge in RNA folding is to define the role of folding intermediates. Folding to the native state occurs in a series of transitions. Populated intermediate states have been characterized by a variety of biophysical and biochemical methods.[1−8] Folding intermediates contain a majority of the native secondary structure and some tertiary structure. These intermediates may act as structural checkpoints, serving to minimize the propensity of RNA to misfold.

Whether intermediate states represent rapidly interconverting heterogeneous ensembles or discrete structures remains an open issue. Compared to native states, intermediate states are less compact, have more solvent exposed residues, and contain fewer tertiary interactions. The majority of techniques used in RNA folding studies are ensemble methods which generate an average description of structures, pathways, and landscapes. Single-molecule studies can track properties of individual RNAs, for example, by following either the dynamics of FRET pairs or the force-versus-extension traces for two points on the RNA. However, these techniques do not provide structural information away from the probed sites and have limited time resolution, leaving the issue of structural heterogeneity unresolved. Cryo-electron microscopy (cryo-EM), by virtue of aligning individual molecules and averaging structurally identical molecules in the reconstruction process, is well poised to address this issue.

We determined the structure of a folding intermediate of the specificity domain of *Bacillus subtilis* RNase P ribozyme (S-domain) using cryo-EM. The native structure of this 154-residue RNA was solved by X-ray crystallography,[9] and its folding behavior in solution has been studied extensively.[8,10,11] In equilibrium folding, the S-domain folds with well-separated unfolded-to-intermediate (U-to-I) and intermediate-to-native (I-to-N) transitions according to circular dichroism (CD) spectroscopy and small-angle X-ray scattering (SAXS).[10,11] Under the experimental condition for cryo-EM (∼19 μM RNA, 1 mM $MgCl_2$, 20 mM TrisHCl, pH 8.0, 22 °C; red data point in Figure 1A), the intermediate state comprised over 95% of the total population. The [Mg^{2+}] used here to populate the intermediate state was significantly higher than in previous CD

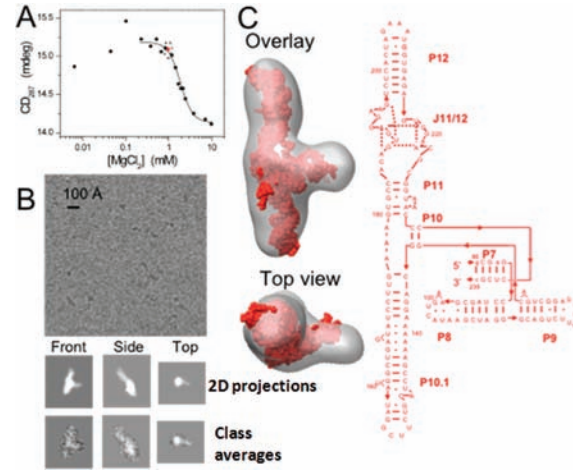

Figure 1. Cryo-EM and SPR of the S-domain intermediate structure. (A) Folding monitored by CD spectroscopy under conditions similar to those of cryo-EM studies. (B) Raw CCD image of RNA molecules (spots) and selected class averages of the intermediate compared with 2D projections from the 3D cryo-EM reconstruction. (C) SPR of the intermediate overlaid to an atomic model and secondary structure. The final 3D map was generated using 60 class averages from ∼11 600 particles.

(0.3 μM RNA, 0.1 mM Mg^{2+}) and SAXS (6.3 μM RNA, 0.4 mM Mg^{2+}) studies because of the higher RNA concentration required in the cryo-EM experiment (19 μM RNA, 1 mM Mg^{2+}). This solution was deposited onto a cryo-EM grid and rapidly frozen into vitreous ice by plunging into liquid ethane.[12] As the molecules freely moved about in solution prior to freezing, snapshots of the RNA solution structure were captured.

During the freezing process, molecules in the intermediate state were unlikely to convert to either the unfolded state or the native state. The S-domain becomes more stable at lower temperature, which precludes the intermediate reverting to the unfolded state. Forward folding to the native state was precluded as the folding time constant is over 10 s at 22 °C.[8] This was much longer than the freezing time, estimated to be less than 100 ms.

After cryo-EM visualization at 60000× magnification, single-particle reconstruction (SPR) was performed in three stages, including particle selection, particle image classification, and iterative 3D refinement using EMAN software.[13] Prior to 3D reconstruction, the particles were subjected to reference-free 2D classification and averaging (Figure 1B, where the black spots are individual RNA molecules). These class averages exhibited characteristic shapes similar to the expected structure (see below), implying that we successfully selected RNA particles from the raw images. The resulting class averages contained less noise than individual particles, and their orientations were determined using a cross-common-lines approach. An initial 3D model was generated from

[†] Department of Biochemistry and Molecular Biology, University of Chicago.
[‡] Institute for Biophysical Dynamics, University of Chicago.
[§] Baylor College of Medicine.

10.1021/ja107492b

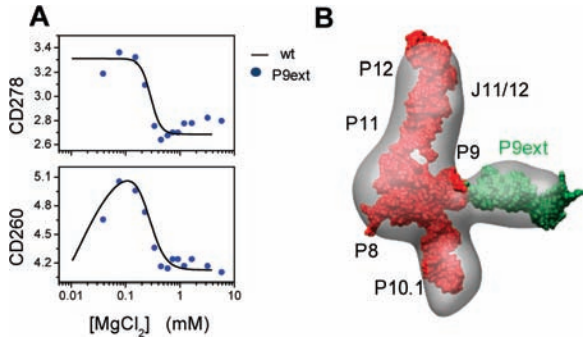

Figure 2. Cryo-EM of S-domain intermediate with the helical extension on helix P9 (P9ext, green). (A) P9ext (blue circles) has the same folding behavior as the wild-type S-domain (black line). (B) Cryo-EM reconstruction of the P9ext intermediate. The final 3D map was generated using 60 class averages from ~19 800 particles.

the class averages and iteratively refined using standard methods of projection matching to produce a final 3D reconstruction (Figure 1C).

The cryo-EM density map of the intermediate is elongated, with a distinct protrusion near the center. The map contains many features found in a model of this intermediate previously generated using solution data[10,11] (Figure 1C). The salient feature of this model is the collinear arrangement of the P12 and P10.1 helices in the intermediate structure; these regions have a side-by-side arrangement in the native fold. The cryo-EM density map of the intermediate recapitulates this principal feature along with the absolute dimensions of the original model.

The major difference between our previously generated model and the cryo-EM density map is at the tip of the P12 region (Figure S1A, Supporting Information). Because we lacked chemical mapping data for this region, we had left it in its native arrangement in the original, SAXS-derived model. The new cryo-EM density map contains additional density above the P12 helix and therefore warrants an adjustment of the orientation of the P12 helix. Using the program Assemble,[14] we rotate P12 into a more vertical, extended conformation to fit the EM density map (Figure S1B). The resulting model also results in better agreement with previous SAXS experiments[10,11] (data not shown).

The size of the S-domain is below the commonly pursued limit of SPR methods.[15] The size, however, is within the range predicted as being possible to align for SPR,[16] and it is further enhanced by the relatively higher scattering of RNA as compared to protein. Because the S-domain is close to the limit of theoretical feasibility, we conducted additional analyses and experiments to verify our findings. To test the robustness of the final reconstruction, we initiated a 3D reconstruction starting from a simplified Gaussian ellipsoid (Figure S1C). Although this reconstruction was devoid of any initial 2D class information, the result agreed well with the cryo-EM density map generated using the 2D class information (Figure S1D).

To further validate our cryo-EM results, we constructed an S-domain variant with a 16-base-pair extension added to the P9 helix, as was done previously for EM studies of a group I intron ribozyme.[17] The P9 hairpin does not participate in any nonlocal interactions in either the native or the intermediate state. Lengthening the P9 stem by 1.5 helical turns (~5 nm) did not alter the folding behavior of the S-domain (Figure 2A). In cryo-EM measurements, the P9 extension significantly helped particle selection and 2D classification because of the additional mass and more identifiable shape. These data were processed similarly to the wild-type data and without any reference to the original intermediate model. The cryo-EM density map of the P9ext-intermediate (Figure 2B) closely resembled that of the wild-type intermediate. Importantly, it clearly exhibited the expected density of the P9 extension (shown in green), allowing unambiguous assignment of the P9 helical arm in the intermediate structure.

The presence of identifiable features in the cryo-EM density map indicated that the intermediate was dominated by conformations having limited global variability. The power of cryo-EM lies in its ability to classify and align thousands of individual snapshots, and thereby reconstruct a 3D map. When interpreting the density map, three possible sample compositions were considered: (i) a significant population of native and unfolded molecules, (ii) a mixture of structures with widely varying angles between the P10.1 and P12 helices, and (iii) a dominant discrete structure. The first two possibilities would result in only the central core intensity being seen in the 2D classification and 3D reconstruction, because the protruding helical arms of the different structures would not be repetitively aligned. Hence, the convergence of both the wild-type and P9ext reconstructions to the observed structural envelopes demonstrates that the intermediate has one dominant conformation.

This study indicates that RNA folding pathways can proceed through discrete and specific global architectures. A sequential folding process may aid in efficient folding by reducing conformational variability and the natural propensity of RNA to misfold. The presence of a distinct energy well for the intermediate would also increase the degree of folding cooperativity because the I-to-N transition occurs between two well-defined structural states. In the case of the S-domain, the discrete structure of the intermediate is a result of both native and non-native interactions.[11]

The application of cryo-EM to determine structural features of biomolecules has been limited to ~200 kDa.[15] The present results demonstrate that cryo-EM methods can generate density maps of metastable RNA structures as small as ~150 residues or ~50 kDa, which is within the theoretical expectation.[16] The feasibility of visualizing such a "small" molecule in our study can be explained partly by the high scattering of the RNA phosphate backbone and the extended helical features that assist in particle alignment during data processing. The ability to obtain molecular envelopes of "small" RNA opens the possibility of structural determination for metastable RNAs, assuming they maintain a single dominant conformation.

Acknowledgment. This work was supported by NIH grants (GM57880 to T.P. and T.R.S., P41RR002250 to S.L. and W.C.). Accession codes. The three-dimensional density maps of the wild type and P9 extension S-domain intermediates have been deposited into the EMDB with accession codes EMD-5242 and EMD-5243, respectively.

Supporting Information Available: Cryo-EM methods and a figure describing modeling of the intermediate structure. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Brion, P.; Westhof, E. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 113–137.
(2) Tinoco, I., Jr.; Bustamante, C. *J. Mol. Biol.* **1999**, *293*, 271–281.
(3) Treiber, D. K.; Williamson, J. R. *Curr. Opin. Struct. Biol.* **2001**, *11*, 309–314.
(4) Thirumalai, D.; et al. *Annu. Rev. Phys. Chem.* **2001**, *52*, 751–762.
(5) Woodson, S. A. *Biochem. Soc. Trans.* **2002**, *30*, 1166–1169.
(6) Russell, R.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 155–160.
(7) Su, L. J.; Waldsich, C.; Pyle, A. M. *Nucleic Acids Res.* **2005**, *33*, 6674–6687.
(8) Baird, N. J.; et al. *Q. Rev. Biophys.* **2007**, *40*, 113–161.
(9) Krasilnikov, A. S.; et al. *Nature* **2003**, *421*, 760–764.
(10) Baird, N. J.; et al. *J. Mol. Biol.* **2005**, *352*, 712–722.
(11) Baird, N. J.; et al. *J. Mol. Biol.* **2010**, *397*, 1298–1306.
(12) Dubochet, J.; et al. *Q. Rev. Biophys.* **1988**, *21*, 129–228.
(13) Ludtke, S. J.; Baldwin, P. R.; Chiu, W. *J. Struct. Biol.* **1999**, *128*, 82–97.
(14) Jossinet, F.; Ludwig, T. E.; Westhof, E. *Bioinformatics* **2010**, *26*, 2057–2059.
(15) Chiu, W.; et al. *Structure* **2005**, *13*, 363–372.
(16) Henderson, R. *Q. Rev. Biophys.* **1995**, *28*, 171–193.
(17) Nakamura, T. M.; et al. *Embo J.* **1995**, *14*, 4849–4859.

JA107492B