Title

A generalizable scaffold-based approach for structure determination of RNAs by cryo-EM.

Journal

Nucleic Acids Research (NAR), 51(20)

Authors

Langeberg, Conner
Kieft, Jeffrey

Publication Date

2023-11-10

DOI

10.1093/nar/gkad784

Peer reviewed

Nucleic Acids Research, 2023, **51**, e100
https://doi.org/10.1093/nar/gkad784
Advance access publication date: 4 October 2023
Methods

A generalizable scaffold-based approach for structure determination of RNAs by cryo-EM

Conner J. Langeberg [1] and Jeffrey S. Kieft [1,2,3,*]

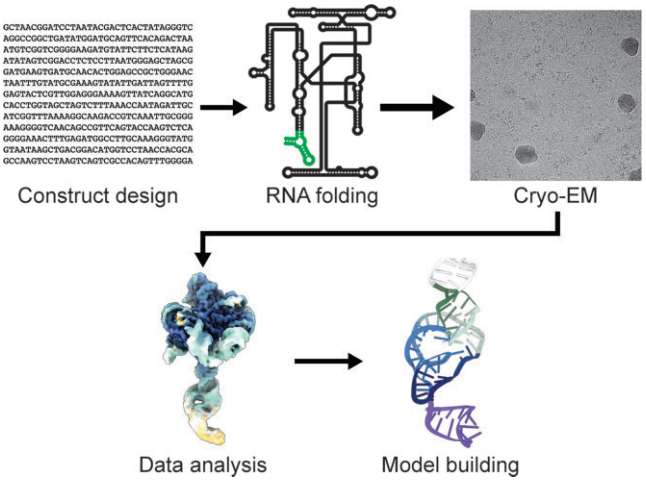
[1]Department of Biochemistry and Molecular Genetics, Aurora, CO 80045, USA
[2]RNA BioScience Initiative, University of Colorado Denver School of Medicine, Aurora, CO 80045, USA
[3]New York Structural Biology Center, New York, NY 10027, USA

[*]To whom correspondence should be addressed. Tel: +1 212 939 0660 (Ext 112); Email: jkieft@nysbc.org

Abstract

Single-particle cryo-electron microscopy (cryo-EM) can reveal the structures of large and often dynamic molecules, but smaller biomolecules (≤50 kDa) remain challenging targets due to their intrinsic low signal to noise ratio. Methods to help resolve small proteins have been applied but development of similar approaches to aid in structural determination of small, structured RNA elements have lagged. Here, we present a scaffold-based approach that we used to recover maps of sub-25 kDa RNA domains to 4.5–5.0 Å. While lacking the detail of true high-resolution maps, these maps are suitable for model building and preliminary structure determination. We demonstrate this method helped faithfully recover the structure of several RNA elements of known structure, and that it promises to be generalized to other RNAs without disturbing their native fold. This approach may streamline the sample preparation process and reduce the optimization required for data collection. This first-generation scaffold approach provides a robust system to aid in RNA structure determination by cryo-EM and lays the groundwork for further scaffold optimization to achieve higher resolution.

Graphical abstract



Introduction

RNAs often require complex tertiary or quaternary folds to affect a phenotypic response. Understanding the structure-function relationship of an RNA molecule therefore requires a detailed understanding of its three-dimensional shape, at an atomic level if possible. Historically this has been achieved through the use of X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. Although powerful, both methods present limitations when working with RNA: X-ray crystallography requires crystallization, and NMR has a practical upper size limit beyond which spectral crowding makes structure determination challenging. Lower-resolution methods such as RNA chemical probing and small angle X-ray scattering (SAXS) supplement our structural understanding of RNA, but do not provide sufficiently high resolution to reliably and routinely provide unambiguous three-dimensional RNA structural information (1–4).

Single-particle cryo-electron microscopy (cryo-EM) now allows determination of atomic resolution structures of many biomolecules and biomolecular complexes (5–7), including those with intrinsic conformational dynamics (8,9). However, this has largely been limited to protein and larger protein-nucleic acid complexes, with relatively few RNA-only structures determined by cryo-EM (10). This can, in part, be

Received: March 20, 2023. **Revised:** August 31, 2023. **Editorial Decision:** September 12, 2023. **Accepted:** September 24, 2023

explained by conformational dynamics in many RNAs, which complicate structural studies (11–13), and many RNA targets of interest are <50 kDa in mass (14–16). The small size of such RNAs results in a low signal to noise ratio in the micrographs, making high-resolution structure determination by cryo-EM difficult or intractable (17,18). Indeed, while improved cryo-EM hardware and sample preparation makes it increasingly possible to achieve higher resolutions for smaller targets (19), a theoretical lower limit on single-particle cryo-EM as applied to proteins was proposed as 38 kDa (20). A number of properties of RNA theoretically may make it more amenable to single particle cryo-EM, such as the presence of phosphorus in the backbone resulting in greater contrast with the solvent compared to protein. Additionally, RNA's lower density compared to protein results in larger 2D projections for the same mass. However, these gains in signal to noise may be offset by RNA's dynamic nature, in part a result of the sparse tertiary interactions that stabilize the overall fold of many RNA elements. Hence, the theoretical lower limit of RNA is likely not dissimilar to that of proteins. When it is possible to recover a cryo-EM map of very low molecular weight RNA molecules, model building is complicated by low resolution and difficulty in determining the base register and stem identity, often requiring alternative structure modeling approaches (21–24).

To circumvent the current size limitation of cryo-EM, various approaches have been developed for protein cryo-EM. These include the use of fragment antigen-binding regions (FAbs) to increase the effective size of a biomolecule (17,25–27) and scaffold-based systems that often use highly symmetric protein complexes to display a cargo of interest (25,28,29). However, similar approaches for RNA-only cryo-EM are sparse (30), often relying on higher-order homo-oligomeric assemblies (31). The fact that RNA elements of interest are often smaller than the theoretical lower limit (14–16) suggests that additional methods would be useful to the field.

We have developed a scaffold-based approach to assist in the determination of small RNA structures by single-particle cryo-EM. The approach involves appending an RNA sequence of interest engineered into a well-ordered larger RNA structure to increase the overall size of the structural target. We first determined that a circularly permuted version of the Tetrahymena ribozyme is an appropriate scaffold, with a rigidity ideal to 'display' smaller RNAs. As a proof of principle, we tested this scaffold approach using three RNAs of known structure: (i) a 23 kDa element from Zika Virus, (ii) a 17 kDa Fluoride riboswitch and (iii) a 17 kDa element from Tamana Bat Virus. The resultant maps of the appended RNAs had resolutions of 5.04, 4.95 and 4.46 Å, respectively, sufficient for *ab initio* structural modeling. While the maps did not allow *de novo* assignment of bound ligands such as magnesium ions, their presence was visible in the maps. Our results illustrate the potential generalizability of this approach as a tool to help determine RNA structures which are significantly smaller than the generally accepted lower cryo-EM size limit. Even when high-resolution maps are not obtained, information about the architecture and behavior of folded RNAs can be assessed.

Materials and methods

in vitro RNA transcription and purification

DNA templates were ordered as gBlock DNA fragments (IDT) and cloned into pUC19 (see Supplementary Table 2). 200 μl PCR reactions using primers containing an upstream T7 promoter were used to generate dsDNA templates for transcription. Typical PCR conditions: 100 ng plasmid DNA, 0.5 μM forward and reverse DNA primers, 500 μM dNTPs, 25 mM TAPS–HCl (pH 9.3), 50 mM KCl, 2 mM $MgCl_2$, 1 mM β-mercaptoethanol and Phusion DNA polymerase (New England BioLabs). dsDNA amplification was confirmed by 1.5% agarose gel electrophoresis. Transcriptions were performed in 1 ml volume using 200 μl of PCR product (∼0.1 μM template DNA) and 10 mM NTPs, 75 mM $MgCl_2$, 30 mM Tris–HCl (pH 8.0), 10 mM DTT, 0.1% spermidine, 0.1% Triton X-100, and T7 RNA polymerase (generated recombinantly in lab). Reactions were incubated at 37°C overnight. After transcription, insoluble inorganic pyrophosphate was removed by centrifugation at 5000×g for 5 min, then the RNA-containing supernatant was precipitated with 3 volumes of 100% ethanol at –80°C for a minimum of 1 h and then centrifuged at 21 000×g for 30 min at 4°C to pellet the RNA, and the ethanolic fraction was decanted. The RNA was resuspended in 9 M urea loading buffer then purified by denaturing 10% PAGE. Bands were visualized by UV shadowing then excised. Bands were then crush-soaked in diethylpyrocarbonate-treated (DEPC) milli-Q water at 4°C overnight. The RNA-containing supernatant was concentrated using spin concentrators (Amicon) to the appropriate concentration in DEPC-treated water. RNAs were stored at 80°C with working stocks stored at 20°C.

RNA refolding

To the RNA solution (usually 18 μl of 3 mg/mL, or approximately 25 μM), 1/20th the volume of 1 M Tris pH 7.5 was added. The solution was heated to 90°C for 3 min then allowed to cool to room temperature for 10 min. Subsequently, 1/20th the volume of 200 mM $MgCl_2$ was added. The solution was heated to 50°C for 30 min, then allowed to cool to room temperature for 10 min. Once cooled, the folded RNA solution was stored on ice until use. For the Fluoride riboswitch construct, 1/20th the volume of 20 mM NaF was added at the time the $MgCl_2$ was added, and all other steps remained the same.

Native gel analysis

To qualitatively assay the folding of each RNA, samples were subject to native gel analysis. Briefly, 1 μg of RNA was refolded as described above. To this, an equal volume of 2× native loading buffer (100 mM Tris pH 7.5, 20 mM $MgCl_2$, 20% sucrose) was added. The samples were then loaded on a 5% native acrylamide gel (5% 37.5:1 acrylamide:bisacrylamide, 0.5× TBE buffer, 10 mM $MgCl_2$, 5% sucrose) and run at 100 V at 4°C for 2 h. RNA was visualized by methylene blue staining of the gel and imaged.

Cryo-EM sample preparation and data acquisition

Each RNA sample was applied to a glow-discharged C-flat (1.2/1.3, 400 mesh) holey carbon grid. Samples were vitrified in liquid ethane using a Vitrobot Mark IV (5 s wait time, 5.5 s blot, −5 blot force, 100% humidity, 4°C). Samples were imaged on a Titan Krios equipped with a Gatan K3 camera and a Bioquantum energy filter. Movies were collected in fringe-free mode, with a physical pixel size of 0.8464 Å, total dose of 50 $e^-/Å^2$, and a defocus range of −0.8 to −2.0 μm. SerialEM was used for all data collection. Additional information on data collection can be found in Supplemental Table 1.

Cryo-EM data processing

The general data processing workflow was approximately equivalent for all datasets using cryoSPARC 4.0. Briefly, all data were patch motion corrected and the contrast transfer function was locally fit. An initial set of particles was picked using blob picker from the first 1000 micrographs. Particles were extracted using a box size of 512 pix and Fourier cropped to 256 pix and 2D classified. Classes which displayed RNA-like features were manually chosen for template picking. Template-based picking was performed using template picker on the entire dataset. Particles were then extracted and subjected to several rounds of 2D classification to sort out junk particles (i.e. particles which did not correspond to the consensus structure). Good 2D classes, which demonstrated features of RNA such as discernable A-form helices, and the associated particles were manually curated after each round. The pruned particles were then used for initial map generation using the *ab initio* reconstruction job requesting 2 volumes. The resulting 2 volumes were used for heterogenous refinement, where one class was the 'good' class and the other a 'junk' class to remove poor quality particles. Good classes were those which demonstrated a map in line with the expected scaffold domain, whereas bad classes did not demonstrate RNA-like features. Several rounds of heterogenous refinement were performed, discarding the 'junk' particles (i.e. those which did not match the consensus model) after each round and using the 'good' particles and resulting maps as inputs for the next round until the map quality stopped improving. The maps shown in Extended Data Figures 1, 5, 8 and 9 are those obtained after the final rounds of heterogeneous refinement. Subsequently, the 'good' volume particles were re-extracted at 512 pix and were subjected to non-uniform refinement. Local motion correction, local CTF refinement, and global CTF refinement were then carried out on the data and the particles were subjected to further rounds of nonuniform refinement to yield the final maps. Additional information on data processing for each map can be found in Extended Data Figures 1, 5, 8 and 9.

Structure modeling

To build each structure, the cryo-EM structure of the Tetrahymena ribozyme was placed into the maps using rigid-body fitting (PDB:7EZ2) followed by manual building of the appended structure. In some regions the solved crystal structure of either the TABV xrRNA (PDB:7K16) or the Fluoride riboswitch (PDB:4ENC) was used to supplement manual building as well as using RNA FRABASE 2.0 (32). The structures were then manually manipulated and joined in COOT (v0.8.9.1). The final model was refined against the map with several rounds of real-space refinement in PHENIX (v1.19.2) and manual adjustments in COOT. Phenix cryo-EM density modification was applied to the Tet_{P6b} map without the atomic model to prevent biases, as the resolution was sufficiently high to benefit the map quality. The statistics of model refinement and validation are listed in supplemental table 1.

Results

Motivation and design considerations for a scaffold-based approach

We reasoned that we could take advantage of the modularity of RNA structure (33), which is largely composed of A-form helices connected by linker regions and stabilized by tertiary interactions (34), to devise a new strategy to aid structural determination of smaller RNAs. It is generally possible to alter distal regions of RNA stem-loops extended away from a core central fold without compromising the overall folded structure or function of the RNA. Indeed, this is a proven method to obtain RNA crystals (35). Appending a small RNA of interest onto an available stem-loop of a larger, well-ordered, recognizable, and robustly folded scaffold would increase the overall effective size of the molecule, likely making it more amenable to cryo-EM (20) (Figure 1A). Also, this strategy would ensure a 1:1 stoichiometry, perhaps help enforce proper folding, and prevent the need to optimize binding conditions necessary with methods that use intermolecular interactions (17,26,27,31). In addition, if the RNA domain is engineered onto the scaffold molecule, the helical orientation and the register of the RNA are known, simplifying model building and removing ambiguity. Other methods to address this often require collecting data on several constructs and determining helical orientation by process of elimination (23). Finally, some RNAs are difficult to characterize structurally due to intrinsic dynamics and heterogeneity (13,23). By applying this approach to RNAs with the propensity to misfold and aggregate, we might mitigate both to yield usable cryo-EM samples.

Tet_{P6b} is a suitable scaffold for high-resolution structure determination

Decades of work have characterized the Tetrahymena group I intron RNA's structure and conformational dynamics (36–38), revealing features that suggest it is a good scaffold for our approach. To test the amenability of this RNA, we used a circularly permuted version (hereafter referred to as Tet_{P6b}). The circularly permuted version of the Tetrahymena ribozyme has the normal 5′ and 3′ ends connected, giving rise to 5′ and 3′ ends in a new position in the structure. In this case, the 5′ and 3′ ends were moved to the P6b helix loop. The resulting map of the Tet_{P6b} yielded a nominal global resolution of 2.44 Å (Figure 1C, D). To date, this is one of two published RNA-only cryo-EM maps <2.5 Å resolution (39). In most regions of the map, the sugar-phosphate chain is clearly resolved, allowing for unambiguous tracing of the entire RNA backbone with sufficient detail to allow *de novo* building of the structure (Figure 1F, G & Extended Data Figure 2). In some parts of the map (mainly around the core domains P7–P3–P8 and most of P4–P6), nucleobases are clearly resolved with features such as exocyclic amines and carbonyl oxygens visible (Extended Data Figure 2). This resolution allowed unambiguous placement of non-Watson–Crick base interactions and native magnesium ions, whose identity was previously known (Extended Data Figure 2). We assessed the map and model by calculating a per-nucleotide *Q*-score to determine the local quality of the map (40) and a map-to-model cross-correlation (CC) score to ensure the built model was in agreement with the recovered map (41) (Figure 1H,I). The Q-score remained within a range expected for a map with a resolution range of 2–4 Å, 0.6–0.9 (Figure 1H), and the CC score was comparably high, greater than 0.9 (Figure 1I). Dips in both the *Q*-score and the CC score matched the lower quality regions of the map such as peripheral helices and loops, as well as regions known to be more dynamic or structurally heterogenous such as the kissing

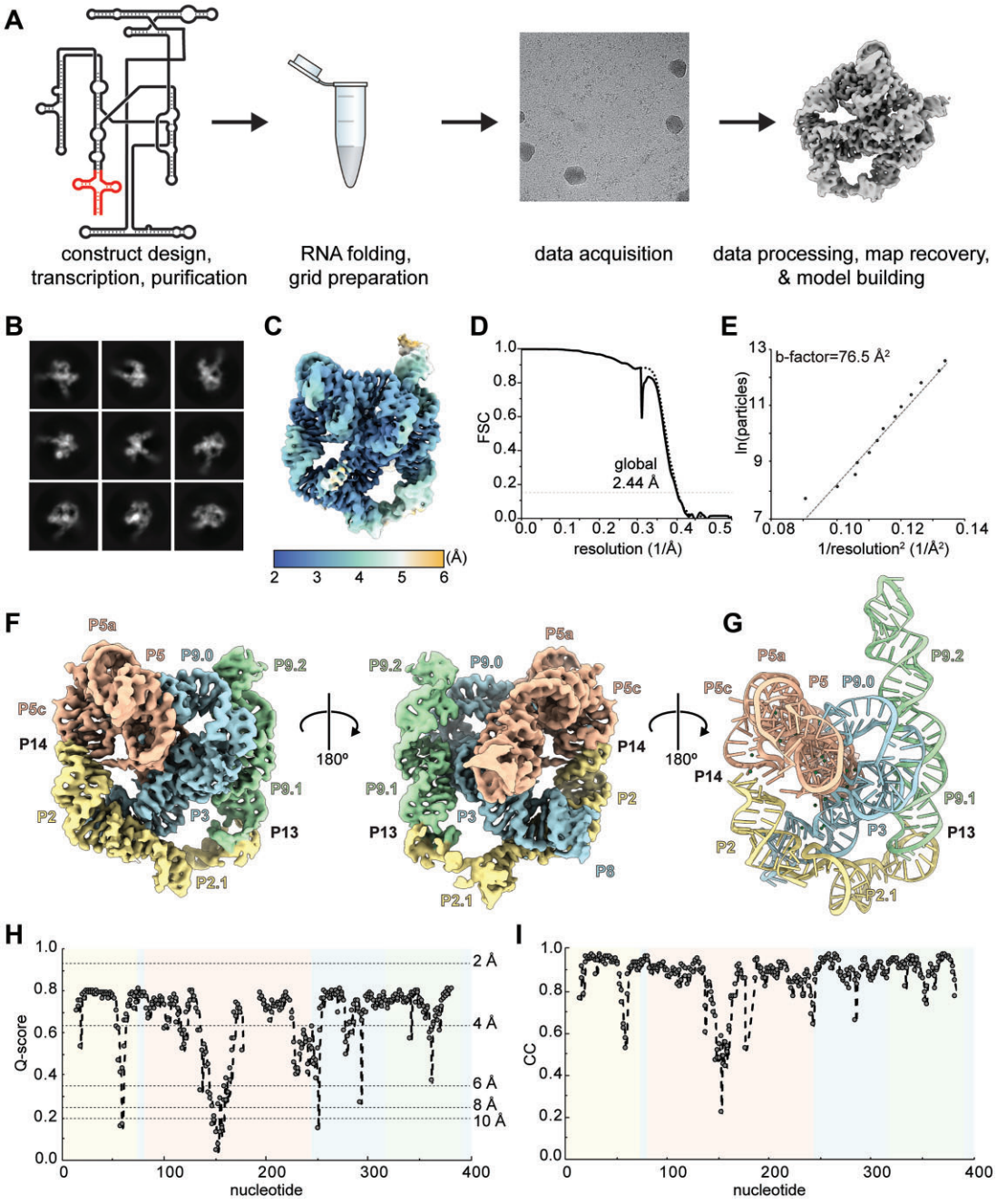


Figure 1. Workflow of the scaffold-based approach and identification of a suitable scaffold. (**A**) The construct has the RNA of interest (red) appended to the larger group I intron RNA. After transcription and purification, the pure RNA is subjected to a standard RNA folding protocol, applied to cryo-EM grids and vitrified. Data collection and processing yields a map for model building and refinement. (**B**) Representative 2D class averages of the Tet_{P6b} construct. (**C**) Local resolution map of the recovered Tet_{P6b} construct. (**D**) Fourier Shell correlation graph of the Tet_{P6b} construct with tight mask (dashed line) and noise subtracted (solid line). (**E**) Henderson-Rosenthal plot of the Tet_{P6b} construct. (**F**) Recovered map of the Tet_{P6b} construct. (**G**) Resulting atomic model of the Tet_{P6b} construct. (**H**) Per-nucleotide *Q*-score graph of the Tet_{P6b} construct. (**I**) Map-to-model cross-correlation score graph of the Tet_{P6b} construct.

loop P13. These trends were also observed in prior cryo-EM studies of the Tetrahymena ribozyme (31,42,43).

Because our method is designed to aid recovery of high-resolution maps of biomolecules, ideally the use of additional particles would result in an even higher-resolution map. Indeed, a Rosenthal–Henderson plot shows a high correlation between particle number and resolution (Figure 1E). Thus, additional particles should contribute to increased resolution, and resolution is therefore not limited by the intrinsic dynamics of the RNA nor by technical limitations. We did not choose to pursue higher resolution of this RNA as our goal was to identify a suitable scaffold for displaying smaller RNAs and this analysis achieved that.

This is the first example of a cryo-EM structure of a circularly permuted version of the Tetrahymena group I intron. In this precatalytic state, the ΩG is poised for the first transesterification step in splicing, surrounded by several well-ordered magnesium ions (Extended Data Figure 2). Unsurprisingly,

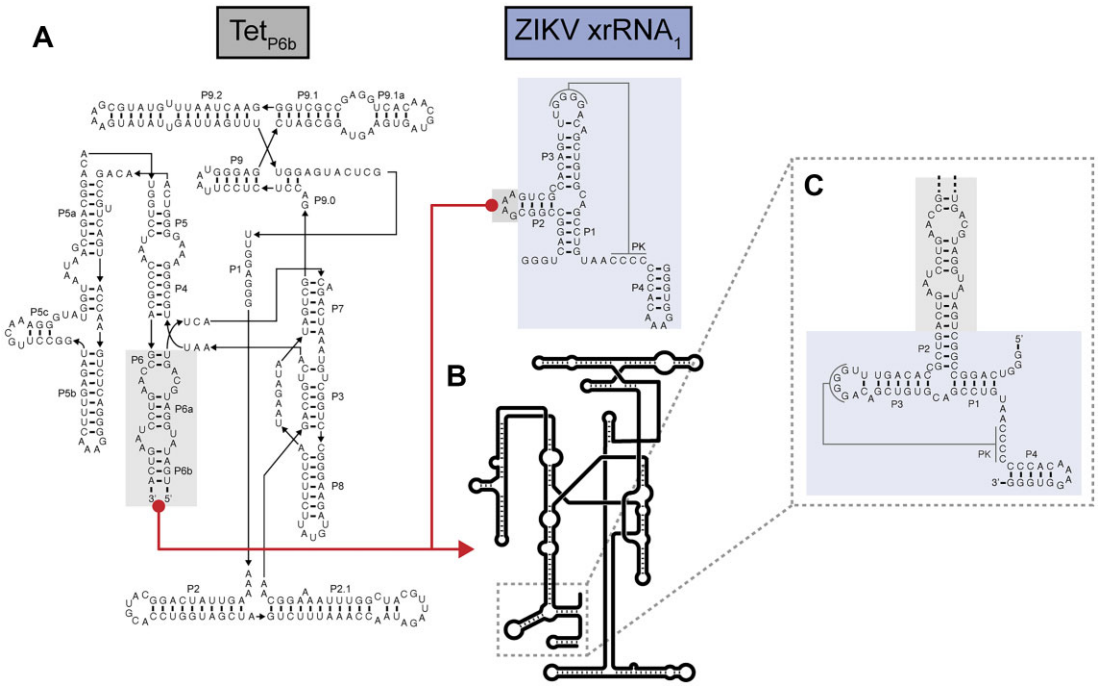


Figure 2. Engineering of the Zika Virus xrRNA-containing test RNA. (**A**) Secondary structures of the Tet$_{P6b}$ RNA and the Zika Virus xrRNA (shaded blue) used in this study. The two RNAs were appended at the locations boxed in grey and designated with red lines. In the xrRNA, P2 can be altered without affecting xrRNA function, therefore the apical loop of this stem was removed and became the point of attachment. (**B**) Cartoon secondary structure depictions of the chimeric RNA, with the xrRNA in blue. (**C**) Close-up of the engineered region.

the active site of the circularly permuted structure remains largely the same as in the solved structure of the wild type RNA (Extended Data Figure 3). However, the regions corresponding to P1 and the 3′ end of P9.0 that comprise the linker are largely unstructured in the circularly permuted version of the ribozyme, making it difficult to trace the RNA here (Extended Data Figure 3). Importantly for the development of our method, this analysis shows Tet$_{P6b}$ is a good candidate for a scaffold-based approach. Notably, circularly permuting and appending domains onto stem-loops other than P6b did not yield usable data, hence we focused on Tet$_{P6b}$ as the scaffold candidate.

The structure of a 23 kDa RNA element from Zika Virus recovered by cryo-EM at 5.04 Å resolution

To benchmark the usefulness of this scaffold-based method to increase the effective size of small RNAs, we tested it with three RNAs of known structure solved by X-ray crystallography. The first was a compact exoribonuclease-resistant RNA (xrRNA) from Zika Virus[40] (Figure 2). We generated a construct with the P2 stem of this 23 kDa RNA fused to the P6b stem of Tet$_{P6b}$. Importantly, the xrRNA's P2 stem is not directly involved in the functional motif of the xrRNA and thus was 'available' to be appended to the scaffold (44). Following transcription of this construct, we analyzed the sample using nondenaturing (native) gels; these were consistent with the RNA folding into a single state and suggested the sample was amenable to single particle cryo-EM analysis (Extended Data Figure 4).

After data collection, particle picking, and initial processes, we examined the resulting 2D classes of xrRNA + Tet$_{P6b}$. In the averages, the appended Zika xrRNA was clearly observable as an additional feature on the Tet$_{P6b}$ (Figure 3A). Initial map generation yielded a class where the appended xrRNA was clear (Figure 3B), but map quality was poor in the distal regions of the xrRNA's P4 stem. To achieve higher resolution of the xrRNA's distal regions, we parsed this pruned dataset with three-dimensional (3D) classification approaches, revealing some conformational heterogeneity in the distal P4 stem, likely reflecting local inherent motion (Extended Data Figure 6). The best resolved 3D class of 68 847 particles was used for final map refinement, yielding a nominal resolution of the total RNA of 3.40 Å (Figure 3C, D), with the highest resolution in the Tet$_{P6b}$ scaffold. Local resolution assessment showed a nominal resolution of 5.04 Å for the xrRNA domain, with decreasing resolution in the distal regions, largely in P4 (Figure 3C, D).

Clear structural features were apparent within the part of the map that contained the appended xrRNA, such as unambiguous major and minor grooves in A-form regions (Figure 3E). Additionally, while individual bases were not directly resolved, clear bumps in the maps for the phosphates of the sugar-phosphate backbone in regions proximal to the scaffold allowed for unambiguous tracing of the backbone. Remarkably, the distinctive ring-like architecture, a defining feature of xrRNA structures determined to date, was clearly visible, and the map showed features consistent with a magnesium ion at the core of the fold that was seen in two crystal structures. The limited resolution in this region was not sufficient to place this ion *de novo*, but was sufficient to confirm the presence of the ion, lend support for its structural importance, and confirm it was not an artifact of crystallization. Hence, globally, the map matches the known structure and confirms many important features that were observed by crystallography.

Using the scaffold to ensure the correct register, we built an *ab initio* atomic model of the Zika xrRNA, relying on the known secondary structure of the xrRNA (Figure 3F). We

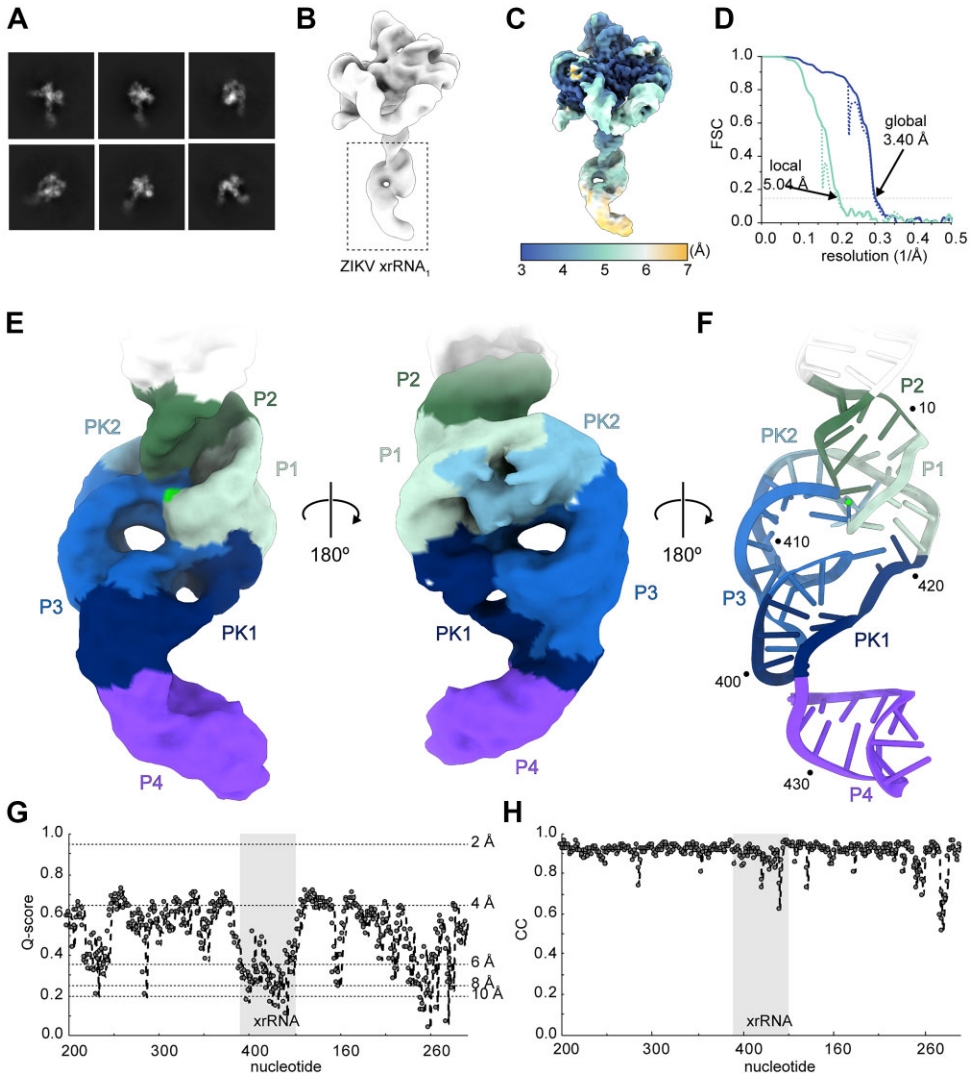


Figure 3. Cryo-EM map of the Zika Virus xrRNA appended onto a scaffold. (**A**) Representative 2D class averages of the Zika Virus xrRNA + Tet_{P6b}. (**B**) Cryo-EM map indicating the location of the Zika Virus xrRNA in the overall initial map. (**C**) Local resolution map of the chimeric Zika Virus xrRNA construct. (**D**) Fourier Shell correlation graph of the Zika Virus xrRNA construct with global resolution (blue) and the local xrRNA resolution (teal) using tight mask (dashed line) and noise subtracted (solid line). (**E**) Recovered map of the chimeric Zika Virus xrRNA construct. (**F**) Resulting atomic model of the chimeric Zika Virus xrRNA construct. A visible magnesium ion is shown as a bright green sphere. (**G**) Per-nucleotide *Q*-score graph of the Tet_{P6b} construct. (**H**) Map-to-model cross-correlation score graph of the Tet_{P6b} construct.

then evaluated the map and model with the *Q*-score (40) and map-to-model CC score (41) (Figure 3G, H). The map of the xrRNA + Tet_{P6b} RNA had *Q*-scores in line with the observed 3.4 Å resolution through much of Tet_{P6b} and consistently high CC scores, both validating the model. Within the xrRNA regions of the map, *Q*-values ranged from 0.5 to 0.2, in agreement with the observed local resolution in this part of the map while suggesting the model was not overfit.

Overall, the Zika Virus xrRNA structural model obtained from cryo-EM aided by our scaffold approach agreed well with the high-resolution crystal structure of the Zika xrRNA (44), with an average backbone RMSD of 3.0 Å across the entire structure and 2.1 Å when P4 is excluded (Figure 4A, B). This is largely because the angle between the P4 stem relative to the rest of the structure was shifted 14° compared to its position in the crystal structure (Figure 4A). Interestingly, in the crystal structure of another xrRNA (from Murray Valley Encephalitis Virus), the P4 stem is nearly perpendicular to its position in the Zika xrRNA structure (45) (Extended Data Figure 7). The cryo-EM map data now support the idea that the P4 stem is mobile compared to the rest of the molecule, and different positions were captured in the two crystal structures. The functional significance of this motion is outside the scope of this study. However, as cryo-EM allows visualization of molecular motions often precluded by crystallization, this observation illustrates that the use of this scaffold-based approach to create RNA samples for use in single-particle cryo-EM may be useful for detecting or confirming putative dynamic regions that can then be studied with other methods.

In summary, while we did not achieve true atomic resolution in the xrRNA, the quality of the map of xrRNA + Tet_{P6b} was sufficient to reveal key architectural features, potential conformational dynamics, and for building and refining a biologically informative structural model. Thus, this scaffold approach is a viable method to 'display' RNAs smaller than the traditional lower size limit of cryo-EM within a

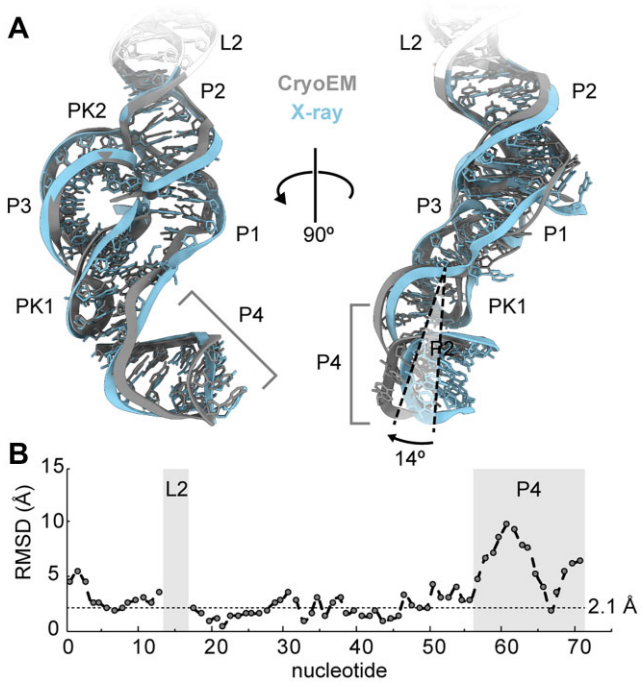


Figure 4. Comparison of cryo-EM-derived model with the X-ray crystallography structure of the Zika Virus xrRNA. (**A**) Overlay of the cryo-EM-based model of the Zika Virus xrRNA (grey) and the crystallography-based model of the Zika Virus xrRNA (cyan). (**B**) Per-nucleotide backbone RMSD graph of both Zika Virus xrRNA structures. The average backbone RMSD across the entire RNA was calculated at 3.0 Å. When P4 is excluded this decreases to 2.1 Å. The RMSD of the L2 loop is omitted as the analogous nucleotides are not present in the cryo-EM model.

larger molecule that can then be used for mid-resolution map generation. When combined with orthogonal data such as covariation analysis and chemical probing, the approach may provide sufficient structural information to build high-confidence models of many RNAs difficult to resolve by other high-resolution structural techniques. In addition, given that this structure was obtained with only ∼68K particles, addition of more data may lead to maps of increased resolution.

Testing the method on two additional small (<20 kDa) RNAs

We tested the general applicability of the use of this scaffold by applying it to two additional examples: (i) a ∼17 kDa xrRNA from Tamana Bat Virus (TABV) (46), distinct in sequence from the Zika Virus xrRNA and (ii) a ∼17 kDa Fluoride riboswitch from *Thermotoga petrophila* (47) (Figure 5). Similar to the scaffolded Zika xrRNA, we used native gels to qualitatively determine that these RNAs were amenable to single particle cryo-EM (Extended Data Figure 4). Then, using the same basic workflow as for the Zika xrRNA construct, we obtained the expected architecture of the scaffold (Extended Data Figures 8, 9). Both resulting maps clearly contain the appended domains with visible structural features such as major and minor grooves.

Within the map of the TABV xrRNA at 4.95 Å, the backbone of the RNA is traceable through the entirety of the structure, with clear phosphate bumps visible in some parts of the map, specifically in regions proximal to the scaffold domain (Figure 5A-C). Similar to the map of the Zika xrRNA, while specific base interactions could not be built *de novo* through the entirety of the structure, the map provides sufficient structural information such that when combined with orthogonal RNA structural techniques such as chemical probing and sequence alignments (16,46), we could build much of the structure without relying on the solved X-ray crystallography structure (Figure 5D). The resulting model demonstrated a backbone RMSD of 1.5 Å compared to the crystal structure throughout the structure, again validating this approach as a viable method to model RNA structures.

The map of the Fluoride riboswitch at 4.46 Å showed similar structural features as the TABV xrRNA, though the quality of the resulting map made data analysis and structure modeling more complicated (Figure 5E–G). However, structural features distinct to this RNA element were clearly present, including density corresponding to a pseudoknot that does not lie in the core of the riboswitch (Figure 5H). Similar to the TABV example, the backbone RMSD compared to the crystal structure throughout the appended domain was calculated at 1.9 Å.

These two additional examples suggest that this scaffold-based method of increasing overall RNA size is generalizable and provide a benchmark for its use. Although high-resolution maps of smaller RNA elements are unlikely to be routinely obtained, the ease of appending different RNA elements of interest to this scaffold may assist in rapid determination of intermediate-to-high-resolution maps of RNAs for model building, likely without requiring exhaustive optimization of RNA folding and grid preparation.

Discussion

Single-particle cryo-EM is now a powerful method for determining macromolecular structure. However, the intrinsic signal to noise limit of the data still makes analysis of small biomolecules or complexes technically challenging (17,18). Whereas methods to circumvent this have been established for protein (17,25–27), similar approaches have not been readily available for RNA, which is particularly limiting as many classes of functional RNAs are too small for traditional single-particle cryo-EM (14–16). Here, we demonstrate a method that can help address this limitation, leveraging RNA's intrinsic modular nature to append diverse RNA elements onto a large scaffold. Alongside other emerging methods and combined with continued advances in other aspects of cryo-EM, this approach for engineering the RNA of interest may assist in more routine structural studies of small RNA elements.

This 'first generation' of our scaffolding approach has several desirable features. First, the samples consistently yielded interpretable maps of the combined Tet_{P6b} scaffold and the appended RNA structures (Zika xrRNA at 5.04 Å, TABV xrRNA at 4.95 Å, and a Fluoride riboswitch at 4.46 Å as assessed using currently accepted map resolution determination methods). Thus, relatively small RNA elements can be examined at resolutions sufficient to gain insight. Intermediate-resolution structures such as these are likely to allow tracing of the backbone of many RNA molecules, and when combined with results from other techniques, biologically relevant and testable tertiary structure models can result. In addition, in some cases the resolution of the map may allow complete or partial structures to be built *de novo*. In other instances, inherent dynamics in the RNA may give rise to uninterpretable maps, but even in those cases information

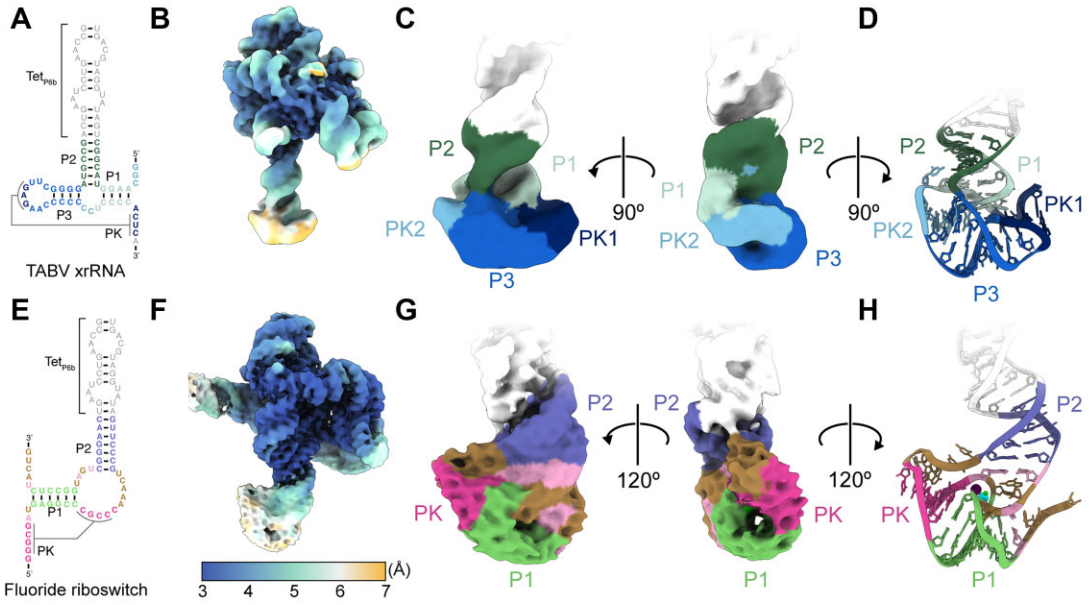


Figure 5. Additional examples of two ∼17 kDa RNA structures recovered using cryo-EM. (**A**) Secondary structure depiction of the TABV xrRNA appended onto Tet$_{P6b}$. (**B**) Local resolution map of the recovered chimeric TABV xrRNA construct. (**C**) Recovered map of the chimeric TABV xrRNA construct. (**D**) Resulting atomic model of the chimeric Zika Virus xrRNA construct. (**E**) Secondary structure depiction of the chimeric *Thermotoga petrophila* Fluoride riboswitch appended onto Tet$_{P6b}$. (**F**) Local resolution map of the recovered Fluoride riboswitch construct. (**G**) Recovered map of the chimeric Fluoride riboswitch construct. (**H**) Resulting atomic model of the chimeric Fluoride riboswitch construct.

regarding the dynamic behavior of the RNA could guide future studies. A similar scaffolding approach has previously been described where a small tetraloop-tetraloop receptor was appended to a small viral frameshifting pseudoknot (48), however this method yielded a relatively low-resolution map that required extensive computational modeling and has not been tested with other targets (49).

The scaffold-based method described here simplifies sample preparation and data analysis, as the same large scaffold is used for every new RNA examined. As most appended RNAs are likely much smaller than the scaffold, we predict similar folding procedures, sample concentration, and grid preparation parameters can be initially used for each new RNA construct, with little or no optimization needed (Figure 1A). Initial data analysis is simplified as the Tet$_{P6b}$ RNA particles are readily recognized and picked using an established template; we observed no preferred orientation, and the initial maps of the scaffold are relatively easy to recover with a standard workflow adapted by the user to each dataset. However, depending on the nature of the appended RNA and the sample, additional computational analyses will likely be unique to each sample.

We chose the Tetrahymena group I intron as the basis for the scaffold as we and others have found it highly amenable to structure determination by cryo-EM (31,42,43). Indeed, here we recovered a map of the circularly permuted Tetrahymena ribozyme at 2.44 Å, similar to that obtained for other versions of the ribozyme (31,42,43). Remarkably, there is still room for improvement as this scaffold is still in a regime where additional particles could yield higher resolution in both the maps of the scaffold and the appended domains. This bodes well for future applications of this or similar scaffolds. Similarly, which stem of the Tetrahymena ribozyme scaffold to append the RNA of interest to may also be optimized on a construct-by-construct basis. Here, P6b was the best stem to minimize motion of the appended domain while not causing misfolding, however the geometry of other RNA elements may necessitate optimization of both the appended stem and linker helix length to prevent steric clashes. Generally, keeping the linking helix as short as possible without causing misfolding is ideal as this minimizes motion of these distal regions which may decrease the local resolution. In the case of the constructs used here, the full-length native helix was used for all three appended constructs while the P6b was shortened to 3 bp.

Importantly, this method does not preclude the careful biochemical, biophysical, and functional studies that precede successful structural analysis. At the very least, the secondary structure of the RNA of interest must be known and validated experimentally to identify stem-loops in the structure. Then, the importance of these stem-loops to the function and fold of the RNA must be determined to identify which stem-loops are suitable for appending to the scaffold without disrupting the native structure. In some cases, and when possible, it may be advantageous to test if the function of the small RNA is retained when appended to the scaffold.

Although this scaffold-based approach described here can be applied immediately with many benefits, there are still features that can be optimized or improved. First, as expected, we observed that within the maps of the appended RNAs, the regions of the maps more proximal to the scaffold showed clearer features. This suggests that it is best to use the shortest possible helix to attach the appended RNA, as long as steric clash does not interfere with the fold. This could be empirically optimized for many RNAs with a set of 2–3 constructs. Related, if an RNA of interest has several available stem-loops that can be altered without loss of function, more than one construct could be examined and the information

from multiple maps combined. However, one important limitation of the approach is that RNA elements that lack an available stem-loop element (e.g. many frameshifting pseudoknots and similar) cannot be appended to the scaffold. Finally, while the scaffold we chose has several highly desirable features, other versions of this scaffold or other larger well-folded RNAs could be tested in 'second-generation' versions.

In recent years, interest in RNA 3D structure determination has increased. The nature of RNA structure and the fact that far fewer RNA structures have been solved relative to protein means that prediction of RNA 3D structures is not yet robust. Machine learning and similar methods hold promise, but require many more RNA structures to examine than currently are available, mandating ways to accelerate experimental RNA structure determination. The application of cryo-EM to RNA structure determination is promising, but thus far has been limited for many technical reasons. The method we present here has the potential to contribute to solving intermediate- and high-resolution structural information on a wide variety of structured RNAs, providing biological insight, testable models, and a richer understanding and database of RNA structure.

Data availability

The data underlying this article are available in the Protein Data Bank at https://www.rcsb.org/, and can be accessed under 8TJX, 8TJQ, 8TJU, and 8TJV. The final refined maps for all structures in this work have been deposited in the Electron Microscopy Data Bank at https://www.ebi.ac.uk/emdb/ and can be accessed under 41313, 41308, 41311 and 41312.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

The authors thank current members of the Kieft Lab for critical reading of the manuscript and insightful comments. Theo Humphries (PNCC) assisted with microscope operations.

Author contributions: C.J.L. and J.S.K. conceptualized and designed the research. C.J.L. prepared the RNA samples for cryo-EM. C.J.L. analyzed the cryo-EM data. C.J.L. performed model building for all RNAs herein. C.J.L. and J.S.K. wrote the manuscript.

Funding

NIH [R35GM118070, R01AI133348 to J.S.K.]; NIH [U24GM129547]; PNCC at OHSU and accessed through EMSL [grid.436923.9]; DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research. Funding for open access charge: Academic Enrichment Funds from CU.

Conflict of interest statement

None declared.

References

1. He,W., Henning-Knechtel,A. and Kirmizialtin,S. (2022) Visualizing RNA structures by SAXS-driven MD simulations. *Front. Bioinform.*, **2**, 781949.
2. Chen,Y. and Pollack,L. (2016) SAXS studies of RNA: structures, dynamics, and interactions with partners. *Wiley Interdiscip. Rev. RNA*, **7**, 512–526.
3. Christy,T.W., Giannetti,C.A., Houlihan,G., Smola,M.J., Rice,G.M., Wang,J., Dokholyan,N.V., Laederach,A., Holliger,P. and Weeks,K.M. (2021) Direct mapping of higher-order RNA interactions by SHAPE-JuMP. *Biochemistry*, **60**, 1971–1982.
4. Yang,S., Parisien,M., Major,F. and Roux,B. (2010) RNA structure determination using SAXS data. *J. Phys. Chem. B*, **114**, 10039–10048.
5. Nakane,T., Kotecha,A., Sente,A., McMullan,G., Masiulis,S., Brown,P., Grigoras,I.T., Malinauskaite,L., Malinauskas,T., Miehling,J., *et al.* (2020) Single-particle cryo-EM at atomic resolution. *Nature*, **587**, 152–156.
6. Yip,K.M., Fischer,N., Paknia,E., Chari,A. and Stark,H. (2020) Atomic-resolution protein structure determination by cryo-EM. *Nature*, **587**, 157–161.
7. Zhang,K., Pintilie,G.D., Li,S., Schmid,M.F. and Chiu,W. (2020) Resolving individual atoms of protein complex by cryo-electron microscopy. *Cell Res.*, **30**, 1136–1139.
8. Murata,K. and Wolf,M. (2018) Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. *Biochim. Biophys. Acta Gen. Subj.*, **1862**, 324–334.
9. Serna,M. (2019) Hands on methods for high resolution cryo-electron microscopy structures of heterogeneous macromolecular complexes. *Front. Mol. Biosci.*, **6**, 33.
10. Ma,H., Jia,X., Zhang,K. and Su,Z. (2022) Cryo-EM advances in RNA structure determination. *Signal Transduct. Target Ther.*, **7**, 58.
11. Al-Hashimi,H.M. and Walter,N.G. (2008) RNA dynamics: it is about time. *Curr. Opin. Struct. Biol.*, **18**, 321–329.
12. Ganser,L.R., Kelly,M.L., Herschlag,D. and Al-Hashimi,H.M. (2019) The roles of structural dynamics in the cellular functions of RNAs. *Nat. Rev. Mol. Cell Biol.*, **20**, 474–489.
13. Bonilla,S.L. and Kieft,J.S. (2022) The promise of cryo-EM to explore RNA structural dynamics. *J. Mol. Biol.*, **434**, 167802.
14. Sherlock,M.E. and Breaker,R.R. (2017) Biochemical validation of a third guanidine riboswitch class in bacteria. *Biochemistry*, **56**, 359–363.
15. Ferre-D'Amare,A.R. and Scott,W.G. (2010) Small self-cleaving ribozymes. *Cold Spring Harb. Perspect. Biol.*, **2**, a003574.
16. Szucs,M.J., Nichols,P.J., Jones,R.A., Vicens,Q. and Kieft,J.S. (2020) A new subclass of exoribonuclease-resistant RNA found in multiple genera of Flaviviridae. *mBio*, **11**, e02352-20.
17. Wu,X. and Rapoport,T.A. (2021) Cryo-EM structure determination of small proteins by nanobody-binding scaffolds (Legobodies). *Proc. Natl. Acad. Sci. U.S.A.*, **118**, e2115001118.
18. Nygaard,R., Kim,J. and Mancia,F. (2020) Cryo-electron microscopy analysis of small membrane proteins. *Curr. Opin. Struct. Biol.*, **64**, 26–33.
19. Herzik,M.A. Jr, Wu,M. and Lander,G.C. (2019) High-resolution structure determination of sub-100 kDa complexes using conventional cryo-EM. *Nat. Commun.*, **10**, 1032.
20. Henderson,R. (1995) The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q. Rev. Biophys.*, **28**, 171–193.
21. Kappel,K., Liu,S., Larsen,K.P., Skiniotis,G., Puglisi,E.V., Puglisi,J.D., Zhou,Z.H., Zhao,R. and Das,R. (2018) De novo computational RNA modeling into cryo-EM maps of large ribonucleoprotein complexes. *Nat. Methods*, **15**, 947–954.
22. Kappel,K., Zhang,K., Su,Z., Watkins,A.M., Kladwang,W., Li,S., Pintilie,G., Topkar,V.V., Rangan,R., Zheludev,I.N., *et al.* (2020) Accelerated cryo-EM-guided determination of three-dimensional RNA-only structures. *Nat. Methods*, **17**, 699–707.

23. Bonilla,S.L., Sherlock,M.E., MacFadden,A. and Kieft,J.S. (2021) A viral RNA hijacks host machinery using dynamic conformational changes of a tRNA-like structure. *Science*, **374**, 955–960.
24. Ma,H., Pham,P., Luo,B., Rangan,R., Kappel,K., Su,Z. and Das,R. (2023) Auto-DRRAFTER: automated RNA modeling based on cryo-EM density. *Methods Mol. Biol.*, **2568**, 193–211.
25. Liu,Y., Huynh,D.T. and Yeates,T.O. (2019) A 3.8 A resolution cryo-EM structure of a small protein bound to an imaging scaffold. *Nat. Commun.*, **10**, 1864.
26. Coscia,F., Estrozi,L.F., Hans,F., Malet,H., Noirclerc-Savoye,M., Schoehn,G. and Petosa,C. (2016) Fusion to a homo-oligomeric scaffold allows cryo-EM analysis of a small protein. *Sci. Rep.*, **6**, 30909.
27. Liu,Y., Gonen,S., Gonen,T. and Yeates,T.O. (2018) Near-atomic cryo-EM imaging of a small protein displayed on a designed scaffolding system. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, 3362–3367.
28. Zhang,K., Horikoshi,N., Li,S., Powers,A.S., Hameedi,M.A., Pintilie,G.D., Chae,H.D., Khan,Y.A., Suomivuori,C.M., Dror,R.O., *et al.* (2022) Cryo-EM, protein engineering, and simulation enable the development of peptide therapeutics against acute myeloid leukemia. *ACS Cent. Sci.*, **8**, 214–222.
29. Zhang,C., Cantara,W., Jeon,Y., Musier-Forsyth,K., Grigorieff,N. and Lyumkis,D. (2019) Analysis of discrete local variability and structural covariance in macromolecular assemblies using Cryo-EM and focused classification. *Ultramicroscopy*, **203**, 170–180.
30. Sampedro Vallina,N., McRae,E.K.S., Hansen,B.K., Boussebayle,A. and Andersen,E.S. (2023) RNA origami scaffolds facilitate cryo-EM characterization of a Broccoli-Pepper aptamer FRET pair. *Nucleic Acids Res.*, **51**, 4613–4624.
31. Liu,D., Thelot,F.A., Piccirilli,J.A., Liao,M. and Yin,P. (2022) Sub-3-A cryo-EM structure of RNA enabled by engineered homomeric self-assembly. *Nat. Methods*, **19**, 576–585.
32. Popenda,M., Szachniuk,M., Blazewicz,M., Wasik,S., Burke,E.K., Blazewicz,J. and Adamiak,R.W. (2010) RNA FRABASE 2.0: an advanced web-accessible database with the capacity to search the three-dimensional fragments within RNA structures. *BMC Bioinf.*, **11**, 231.
33. Grabow,W. and Jaeger,L. (2013) RNA modularity for synthetic biology. *F1000Prime Rep.*, **5**, 46.
34. Leontis,N.B., Lescoute,A. and Westhof,E. (2006) The building blocks and motifs of RNA architecture. *Curr. Opin. Struct. Biol.*, **16**, 279–287.
35. Lu,C., Cai,R., Grigg,J.C. and Ke,A. (2021) Using tRNA scaffold to assist RNA crystallization. *Methods Mol. Biol.*, **2323**, 39–47.
36. Kruger,K., Grabowski,P.J., Zaug,A.J., Sands,J., Gottschling,D.E. and Cech,T.R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell*, **31**, 147–157.
37. Woodson,S.A. (2002) Folding mechanisms of group I ribozymes: role of stability and contact order. *Biochem. Soc. Trans.*, **30**, 1166–1169.
38. Cech,T.R. (2002) Ribozymes, the first 20 years. *Biochem. Soc. Trans.*, **30**, 1162–1166.
39. Li,S., Palo,M.Z., Zhang,X., Pintilie,G. and Zhang,K. (2023) Snapshots of the second-step self-splicing of Tetrahymena ribozyme revealed by cryo-EM. *Nat. Commun.*, **14**, 1294.
40. Pintilie,G., Zhang,K., Su,Z., Li,S., Schmid,M.F. and Chiu,W. (2020) Measurement of atom resolvability in cryo-EM maps with Q-scores. *Nat. Methods*, **17**, 328–334.
41. Afonine,P.V., Klaholz,B.P., Moriarty,N.W., Poon,B.K., Sobolev,O.V., Terwilliger,T.C., Adams,P.D. and Urzhumtsev,A. (2018) New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr. D Struct. Biol.*, **74**, 814–840.
42. Su,Z., Zhang,K., Kappel,K., Li,S., Palo,M.Z., Pintilie,G.D., Rangan,R., Luo,B., Wei,Y., Das,R., *et al.* (2021) Cryo-EM structures of full-length Tetrahymena ribozyme at 3.1 A resolution. *Nature*, **596**, 603–607.
43. Bonilla,S.L., Vicens,Q. and Kieft,J.S. (2022) Cryo-EM reveals an entangled kinetic trap in the folding of a catalytic RNA. *Sci. Adv.*, **8**, eabq4144.
44. Akiyama,B.M., Laurence,H.M., Massey,A.R., Costantino,D.A., Xie,X.P., Yang,Y.J., Shi,P.Y., Nix,J.C., Beckham,J.D. and Kieft,J.S. (2016) Zika virus produces noncoding RNAs using a multi-pseudoknot structure that confounds a cellular exonuclease. *Science*, **354**, 1148–1152.
45. Chapman,E.G., Costantino,D.A., Rabe,J.L., Moon,S.L., Wilusz,J., Nix,J.C. and Kieft,J.S. (2014) The structural basis of pathogenic subgenomic flavivirus RNA (sfRNA) production. *Science*, **344**, 307–310.
46. Jones,R.A., Steckelberg,A.L., Vicens,Q., Szucs,M.J., Akiyama,B.M. and Kieft,J.S. (2021) Different tertiary interactions create the same important 3D features in a distinct flavivirus xrRNA. *RNA*, **27**, 54–65.
47. Ren,A., Rajashankar,K.R. and Patel,D.J. (2012) Fluoride ion encapsulation by Mg2+ ions and phosphates in a fluoride riboswitch. *Nature*, **486**, 85–89.
48. Zhang,K., Zheludev,I.N., Hagey,R.J., Haslecker,R., Hou,Y.J., Kretsch,R., Pintilie,G.D., Rangan,R., Kladwang,W., Li,S., *et al.* (2021) Cryo-EM and antisense targeting of the 28-kDa frameshift stimulation element from the SARS-CoV-2 RNA genome. *Nat. Struct. Mol. Biol.*, **28**, 747–754.
49. Roman,C., Lewicka,A., Koirala,D., Li,N.S. and Piccirilli,J.A. (2021) The SARS-CoV-2 programmed -1 ribosomal frameshifting element crystal structure solved to 2.09 A using chaperone-assisted RNA crystallography. *ACS Chem. Biol.*, **16**, 1469–1481.

Received: March 20, 2023. **Revised:** August 31, 2023. **Editorial Decision:** September 12, 2023. **Accepted:** September 24, 2023