ARTICLES
https://doi.org/10.1038/s41592-022-01455-w

nature methods

Sub-3-Å cryo-EM structure of RNA enabled by engineered homomeric self-assembly

Di Liu[1,2,6], François A. Thélot[3,6], Joseph A. Piccirilli[4,5], Maofu Liao[3]✉ and Peng Yin[1,2]✉

High-resolution structural studies are essential for understanding the folding and function of diverse RNAs. Herein, we present a nanoarchitectural engineering strategy for efficient structural determination of RNA-only structures using single-particle cryogenic electron microscopy (cryo-EM). This strategy—ROCK (RNA oligomerization-enabled cryo-EM via installing kissing loops)—involves installing kissing-loop sequences onto the functionally nonessential stems of RNAs for homomeric self-assembly into closed rings with multiplied molecular weights and mitigated structural flexibility. ROCK enables cryo-EM reconstruction of the *Tetrahymena* group I intron at 2.98-Å resolution overall (2.85 Å for the core), allowing de novo model building of the complete RNA, including the previously unknown peripheral domains. ROCK is further applied to two smaller RNAs—the *Azoarcus* group I intron and the FMN riboswitch, revealing the conformational change of the former and the bound ligand in the latter. ROCK holds promise to greatly facilitate the use of cryo-EM in RNA structural studies.

Recent decades have witnessed the growing appreciation for the versatile and far-reaching roles of RNAs: beyond guiding protein biosynthesis, they regulate gene expression and modulate other important biological processes by various mechanisms[1,2], such as binding proteins, recognizing metabolites and catalyzing chemical transformations. More than 85% of the human genome is transcribed, but less than 3% of genome encodes proteins[3], indicating that a large portion of transcribed RNAs have unknown functions and underlying structures. Furthermore, the functional capacity of RNAs has been expanded by in vitro selection and evolution[4–6]. High-resolution structural information is essential for understanding the function of such RNAs. However, experimentally solved three-dimensional (3D) RNA structures remain scarce: among the total of ~180,000 structures presently deposited in the Protein Data Bank (PDB), only 0.9% are RNA structures. This reflects the difficulties associated with the structural acquisition of RNAs, especially for large RNAs (>~100 nucleotides (nt); small RNAs of up to ~100 nt can be studied by nuclear magnetic resonance (NMR)), which are studied primarily by X-ray crystallography[7]. First, the intrinsic properties of RNA, such as (1) a poorly differentiated anionic surface, (2) an irregular and elongated shape, and (3) conformational heterogeneity and structural flexibility, present challenges for obtaining crystals that diffract well. Second, phase determination of RNA crystals is also difficult owing to the lack of convenient and general strategies, such as selenomethionine substitution[8] used in protein crystallography.

Without the need for procuring crystals and solving phase problems, cryo-EM is gaining increasing popularity in structural determination of protein-containing systems, and the resolution it provides is beginning to rival that of crystallography, thanks to the ongoing advances in instruments and computational techniques[9]. Nevertheless, cryo-EM has not been well-explored for RNAs. So far, there are only two reported examples of cryo-EM-determined RNA-only structures[10–12] that have achieved a resolution of 4.5 Å or better. The first is the 4.5-Å structure of the *Lactococcus lactis* group IIA intron[10], a large RNA containing >600 nt. The other more recent case is the smaller 119-nt *Mycobacterium* sp. MCS SAM-IV riboswitch[12]: 3.7- and 4.1-Å cryo-EM maps were reported for the unliganded and ligand-bound states, respectively, but large datasets (~2 million initial particles) were required for cryo-EM reconstruction. Even for the best-resolved map at 3.7 Å, the intensities for nucleobases are barely separate and backbone features are not well delineated, making the model-building process heavily dependent on computer modeling[13].

Single-particle cryo-EM reconstruction is capable of identifying different conformations by computationally classifying the particles on the basis of their 3D shapes to yield structures of discrete conformations, and, consequently, higher-resolution structures can be typically obtained from the refinement of the most populated class of particles than from the whole set of particles. Thus, in principle, cryo-EM holds an additional advantage in studying RNA structures whose misfolding propensity and structural heterogeneity are well-known to complicate their structural studies. However, structural flexibility of RNA—likely arising from the less compact shapes of folded RNAs[14] compared to proteins and fewer extensive long-range tertiary interactions to stabilize the overall 3D architectures[7]—can interfere with the cryo-EM-enabled conformational classification. In addition, many structured RNAs are relatively small (<100 kDa, or <300 nt), making it challenging to accurately align single-particle images owing to the low signal-to-noise ratio and limited spatial information.

To address the challenges in RNA cryo-EM, we herein present ROCK, a nanoarchitectural engineering strategy derived from nucleic-acid nanotechnology[15–19]. Kissing-loop sequences are installed onto the peripheral stems of the target RNA to mediate its self-assembly into a closed homooligomeric ring. The assembled structure has a multiplied molecular weight and each constituent monomeric unit is expected to have mitigated flexibility, thus simultaneously addressing both challenges of small molecular weight and structural flexibility for RNA cryo-EM. Harnessing the capabilities

[1]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, USA. [2]Department of Systems Biology, Harvard Medical School, Boston, MA, USA. [3]Department of Cell Biology, Blavatnik Institute, Harvard Medical School, Boston, MA, USA. [4]Department of Chemistry, the University of Chicago, Chicago, IL, USA. [5]Department of Biochemistry and Molecular Biology, the University of Chicago, Chicago, IL, USA. [6]These authors contributed equally: Di Liu, François A. Thélot. ✉e-mail: maofu_liao@hms.harvard.edu; py@hms.harvard.edu

of ROCK, we determined the structure of the complete *Tetrahymena* group I intron[20] (TetGI) at 2.98 Å resolution (2.85 Å for the core). The cryo-EM map presents clear features of base-to-base separation and sugar-phosphate backbone that are characteristic for RNA and thereby enables de novo model building of the complete TetGI, including the previously unknown peripheral domains. We also demonstrated that different regions of the TetGI can be selectively stabilized by configuring it into two different oligomers (dimer or trimer) via engineering different pairs of peripheral stems, and that structures corresponding to the states of the intron with and without bound substrate can be reconstructed. Lastly, ROCK's general utility is further validated by its application to two smaller RNAs—the *Azoarcus* group I intron (AzoGI; 206 nt) and the FMN riboswitch (112 nt), of which modest-resolution cryo-EM maps are readily obtainable from datasets of reasonable sizes and by use of more accessible instruments. These maps allow us to capture the conformational change of the AzoGI from closed to open conformations after splicing, and to delineate the ligand-binding environment of the FMN riboswitch, demonstrating the potential of cryo-EM for studying RNA dynamics and RNA-binding molecules.

Results

RNA construct engineering for cryo-EM. Construct engineering is a common strategy in RNA crystallography[7,21] and many methods of general applicability have been developed, including the installment of the tetraloop–tetraloop receptor interaction[22] and the introduction of binding proteins, such as U1A protein[23,24] or antibody fragments[25,26]. The main purpose of construct engineering in crystallography is to create preferred intermolecular crystal-packing interactions for mediating crystal growth and, at the same time, dampening structural flexibility by locking the molecules in ordered lattice. For crystallography, because the effects of construct engineering are difficult to predict, and diffraction quality assessment and data collection can be performed very fast thanks to the technical advances at synchrotron beamlines, various constructs of a target RNA are routinely prepared and screened in a high-throughput fashion. In contrast, cryo-EM remains a low-throughput and expensive technique, calling for a high success rate if construct engineering is to be deployed.

Taking the advantage of the programmability of RNA and the advances in nucleic-acid nanotechnology[15–19], ROCK involves rationally designing RNA constructs to self-assemble into a homomeric closed ring via installing kissing-loop sequences onto the peripheral helices without perturbing the functional core of the RNA (Fig. 1a; see Supplementary Note 1 for the considerations in choosing the 7-bp kissing-loop motif[27] throughout this work). Compared with a stand-alone RNA, the assembled ring structure is more amenable to cryo-EM structure determination (Fig. 1b and Extended Data Fig. 1a,b). First, the assembled structure has a multiplied molecular weight and a characteristic, symmetric shape, substantially streamlining cryo-EM experiments and reconstruction. Second, each constituent monomeric unit is expected to have mitigated flexibility due to the geometric restraints imposed by the kissing loops-mediated ring closure—this is analogous to quaternary structures that are often observed in proteins[28] but are rare in known natural RNA structures[29,30]. Third, the self-assembled system also facilitates the experimental procedures of folding optimization and native purification (Extended Data Fig. 1b). This helps eliminate the misfolding and conformational heterogeneity that are well-known to complicate the functional and structural studies of RNAs.

The lengths of the stems onto which the kissing-loop sequences are installed need to be optimized to ensure formation of a ring (a closed structure) instead of other linear or spiral assemblies (open structures; Extended Data Fig. 1c). Promising constructs can be readily designed in silico with software such as NanoTiler[31] and experimentally validated by native polyacrylamide gel electrophoresis (PAGE), which also helps determine the optimal folding/assembly condition and enables the native purification of the desired oligomer containing correctly folded RNA subunits (Fig. 1c).

Engineering the TetGI for homomeric self-assembly. To demonstrate the utility of ROCK, we chose the TetGI for an in-depth case study because it is a representative and challenging RNA structure: it is the first discovered and most iconic catalytic RNA[20] and, as one of the most interrogated folded RNA molecules, serves as a rewarding model for research of RNA biochemistry and structural biology[32,33]; however, its complete structure was unknown in the roughly four decades following its discovery. Figure 2a presents the secondary structure (shown as the pre-2S state, that is the state before the second step of splicing) of the TetGI derived from previous prediction[34,35] and amended by this work. Besides the catalytic core[34] (containing P4–P6, P3–P9 and P1–P10 domains) conserved for all group I introns, the TetGI, as a subgroup IC1 intron, also possesses peripheral P2–P2.1 and P9.1–P9.2 domains (or domains 2 and 9, respectively; Fig. 2b). Though a number of partial structures of the TetGI have been determined by crystallography (including the P4–P6 domain[36,37] and the core[38,39]; see Supplementary Table 1 for representative solved group I intron structures), its complete structure has not been determined at high resolution. Nonetheless, the complete TetGI has been modeled computationally based on phylogeny, biochemistry data and a number of distance constraints that were derived from long-range tertiary interactions[34,35] before crystal structures were available.

The TetGI catalyzes two consecutive phosphotransesterification reactions, and we chose two of its reaction states for our construct designs. Figure 2c presents a state initially designed as the post-2S (the state after the second step of splicing) complex. Two deoxy substitutions were introduced at the splice junction (u–1 and u+1) of the ligated exon mimic (TetLEM) to prevent the reverse reaction of the second step of splicing[40]. In attempting to prepare this construct, we observed that the intron RNA was cleaved co-transcriptionally between U20 and U21 (Extended Data Fig. 2a,b), likely owing to the formation of a hairpin at the 5′ sequence (Extended Data Fig. 2c). Therefore, the truncated intron RNA, when hybridized to TetLEM (Fig. 2d), can also be regarded as the *trans*-acting ribozyme (for endonucleolytic reaction[41]) in complex with its oligonucleotide substrate. Figure 2d shows the other reaction state, termed pre-2SΔ5′ex, corresponding to the pre-2S complex[40] but without the 5′ exon. This construct is formed by two fragments: an in vitro-transcribed (IVT) RNA corresponding to the 5′ fragment of the TetGI through nucleotide A386 and a chemically synthesized 37-nt chimeric oligonucleotide (dTetCIRC) corresponding to the intron's 3′ fragment and 3′ exon. We introduced mutations to the 5′ sequence of the IVT RNA (along with the compensatory mutations to the 3′ exon sequence of dTetCIRC to maintain P10) to prevent the possible formation of the aforementioned hairpin presumably responsible for the 5′ cleavage (Extended Data Fig. 2b,c). Because the TetGI is susceptible to hydrolysis at the 3′ splice site[42] in the absence of the 5′ exon (Extended Data Fig. 2d–f), we introduced two deoxy substitutions in dTetCIRC at G414 and a+1 (the u+1a mutation introduces an additional base pair at the base of P10, potentially improving the construct's rigidity[40]) to inhibit this hydrolysis reaction (Extended Data Fig. 2g–j). Compared to the post-2S complex, the pre-2SΔ5′ex complex has P9.0 and P10 covalently linked, which is anticipated to help rigidify the architecture[40] and therefore is beneficial for obtaining a higher-resolution structure.

According to the knowledge gained from the previous structural and functional studies, P6b, P8 and P9.2 extend away from the catalytic core and do not participate in tertiary interactions; therefore, we chose these three stems for ROCK engineering. In this work, three TetGI constructs are designed and studied by cryo-EM: two dimeric constructs (Fig. 2e), TetGI-D and TetGI-DS, designed as

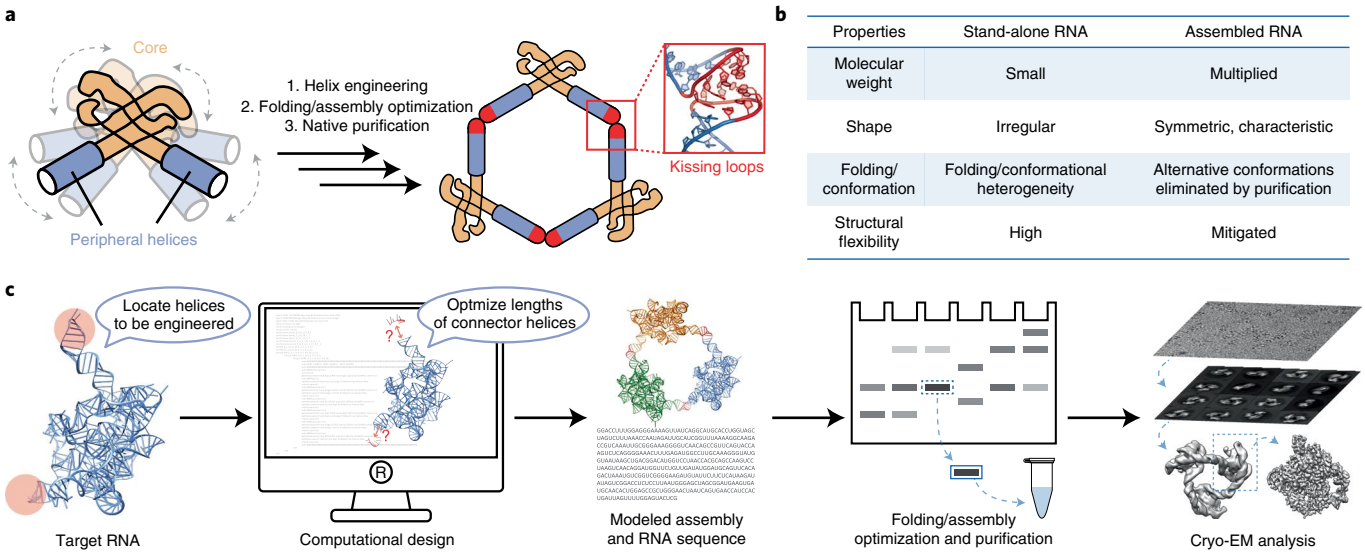


Fig. 1 | Conceptualization and workflow of ROCK. a, Engineering a target RNA for the self-assembly of a closed homomeric ring through installing kissing-loop sequences onto the functionally nonessential peripheral helices. Throughout this work, we use a 7-bp kissing-loop motif that adopts an included angle of ~120° based on its NMR structure[27] (PDB: 2BJ2). **b**, Tabulated comparison of the properties of stand-alone RNA and assembled RNA shows that the latter utilized in our method is more amenable to cryo-EM structural determination. **c**, Workflow of ROCK. The peripheral helices of the target RNA onto which kissing-loop sequences are installed (highlighted by light red shadows) are located. Computational design involves length optimization of the connector helices between RNA core and the kissing loops (to ensure the formation of a ring structure), computer modeling of the homomeric ring and generation of RNA sequence. Native polyacrylamide gel electrophoresis (PAGE) is used to screen the optimal folding/assembly conditions and to purify the desired homooligomer containing the correctly folded RNA. The gel-purified sample is subjected to structural determination by cryo-EM.

the post-2S and pre-2SΔ5′ex states, respectively; and a post-2S trimeric construct (Fig. 2f), TetGI-T. Another dimeric construct designed by engineering P6b and P8 failed to give a high yield of the desired dimer and was not further studied (Supplementary Fig. 1). We note that, during the preparation of this paper, a 6.8-Å cryo-EM structure of the TetGI in the *trans*-acting ribozyme form was reported[13], and its structural information, although not used in the present study, can guide construct engineering.

Sub-3-Å cryo-EM structure of the TetGI-DS construct. The best-resolved cryo-EM map of the TetGI was obtained from the TetGI-DS construct (Fig. 2d,e). The characteristic shape of the TetGI-DS homodimer, clearly visible from the raw micrographs (Extended Data Fig. 3a) and two-dimensional (2D) class averages (Fig. 3a and Extended Data Fig. 3b), facilitates particle-picking and initial alignment. Reconstruction of the dimer with C_2 symmetry delivered a cryo-EM map of a modest 3.92-Å resolution (see Fig. 3b for the whole dimer, and Fig. 3c for the symmetrized monomer, or the C_2 monomer). Symmetry-expansion (SE) allows finer 3D classification of the monomers, and subsequent refinement yielded a map of a substantially improved resolution (Fig. 3d and Extended Data Fig. 3c–e): as estimated by the Fourier shell correlation (FSC) curves (Fig. 3e), the overall resolution for the SE monomer is 2.98 Å at FSC = 0.143, with the core arriving at 2.85-Å resolution, superior to the best-resolved group I intron to date, the 3.10-Å AzoGI crystal structure[40]. This represents the first sub-3-Å cryo-EM map obtained for an RNA-only structure and enables the de novo model building of the complete TetGI (Fig. 3f,g). In this map, fine details of structural features pertaining to the sugar-phosphate backbone and nucleobases are resolved (Fig. 3h; see Supplementary Fig. 2 for comparing the features of different X-ray and EM maps). To demonstrate the excellent map quality, we show that the intensities for individual bases are well separated without breaking the backbone continuity at a wide range of contour levels (Fig. 3i), that the distinct geometries of different types of base pairs can be readily recognized (Fig. 3j,k), and, notably, that features corresponding to the exocyclic amino groups are visible (blue arrows in Fig. 3i–k). Additionally, we can visualize the strong intensities of ordered native Mg^{2+} ions (Fig. 3l; see Supplementary Fig. 3 for the Mg^{2+} ions at the A-rich bulge of P4–P6 domain), which have important roles in RNA folding and sometimes serve as ligands in RNA catalysis[43].

Assembly, activity and cryo-EM analyses of TetGI-D and -T. Both TetGI-D and TetGI-T are constructs designed in the post-2S state, enabling their parallel comparison. Because the cation contents (species and concentrations) can substantially influence RNA folding and assembly[44], we tested the assembly of TetGI-D and TetGI-T (along with the monomer control, TetGI-M) in different annealing buffers. We then analyzed the assemblies by native PAGE as shown in Fig. 4a. Indeed, different annealing buffers resulted in different assembly patterns and the maximum yields of the desired dimer for TetGI-D (lanes 4 and 5) and the desired trimer for TetGI-T (lanes 10 and 11) were obtained in a buffer containing no Na^+ and 1 or 3 mM Mg^{2+}. We note that the optimal Mg^{2+} concentrations for yielding the desired oligomers are close to the previously determined optimal concentration (2 mM) for TetGI folding[45], implying that the homomeric self-assembly system can serve as a read-out platform for optimizing RNA-folding conditions. Consequently, we assembled the TetGI-D dimer and TetGI-T trimer in large scale with 3 mM Mg^{2+} and purified them by preparative native PAGE for the subsequent activity assays and cryo-EM analyses.

To ensure the catalytic activity of the engineered, assembled constructs, we assayed their *trans*-acting endonucleolytic activity[41] (Fig. 4b). The reaction kinetics for TetGI-M, TetGI-D and TetGI-T are almost identical within the first 15 minutes of reaction when less than 50% of substrate is cleaved, indicating that the activities of the engineered TetGI subunits within the homomeric assemblies are not notably affected. Interestingly, TetGI-D and TetGI-T are slightly

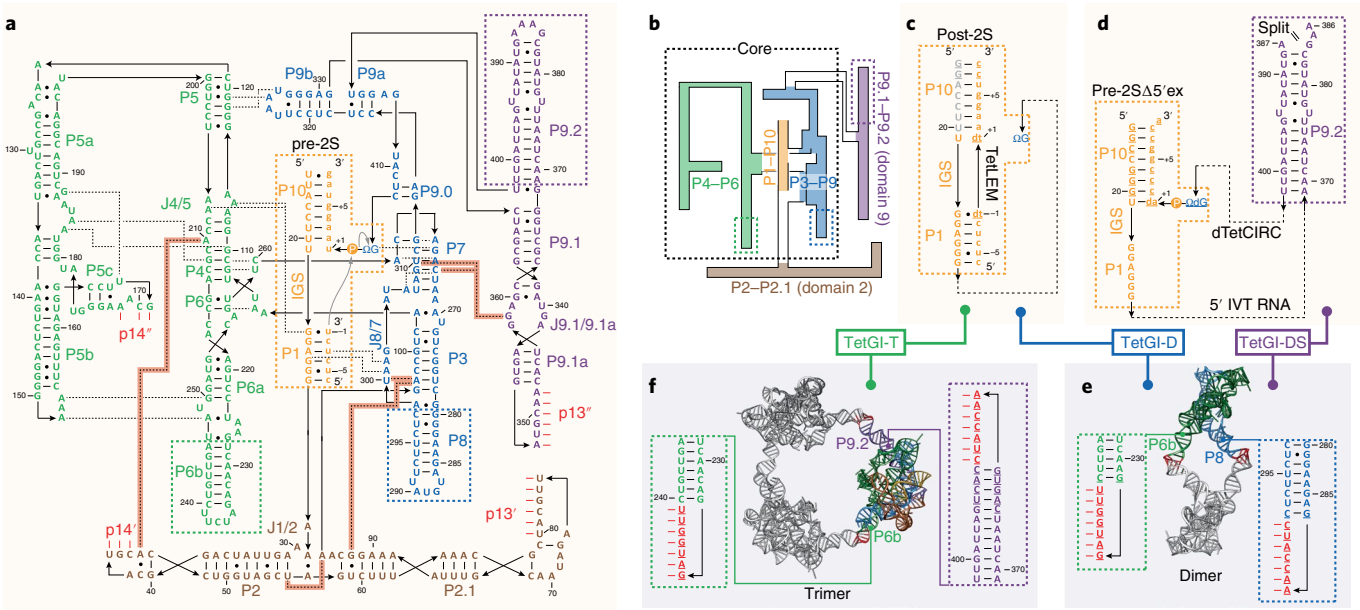


Fig. 2 | Construct engineering of the *Tetrahymena* group I intron (TetGI) as a case study for ROCK. a, Secondary structure of the wild-type TetGI in the pre-2S state. Black arrows indicate direct covalent connections (from 5′ to 3′) in the primary sequence. Key tertiary contacts are shown with black dashed lines (tertiary interactions newly discovered in this work are highlighted by red shadows). Short lines (–) designate canonical Watson–Crick (WC) base pairs, and dots (•) designate non-WC base pairs. "P" and "J" refer to "pairing" and "joining" regions, respectively, in the secondary structure. Nucleotides involved in P*n* (*n* = 13 or 14) tertiary base-pairings are indicated by short red sticks, and p*n*′ and p*n*″ correspond to the 5′ and 3′ strands of P*n*. Lower-case nucleotides are from exons. ΩG, the intron's 3′-terminal guanosine. IGS, the internal guide sequence. Dashed boxes mark the locations to be engineered. The catalyzed reaction is shown by gray arrows. **b**, Schematic of the core (black dashed box; including the P4–P6, P1–P10 and P3–P9 domains), and the peripheral P2–P2.1 and P9.1–P9.2 domains (or domains 2 and 9, respectively) of the TetGI. **c**,**d**, Engineered constructs corresponding to the post-2S (**c**) and the pre-2SΔ5′ex (pre-2S without 5′ exon bound; **d**) states. Underlined nucleotides are mutated from the wild type. In **c**, gray nucleotides are absent owing to the cleavage after U20 (Extended Data Fig. 2b) during the preparation by in vitro transcription (IVT); the ligated exon mimic (TetLEM; chemically synthesized) has two deoxy mutations. In **d**, the construct, with a strand split at the apical loop of P9.2 (between A386 and A387), is formed by an IVT-synthesized 5′ fragment and a chemically synthesized 3′ fragment with two deoxy substitutions (dTetCIRC). **e**,**f**, The dimeric (**e**) and trimeric (**f**) assemblies are designed by installing kissing-loop sequences (red) onto P6b and P8, and P6b and P9.2, respectively. The design of dimer is based on the crystal structure[39] of the TetGI core (PDB: 1X8W), and the design of trimer is based on a computer model[35] of the complete intron. In this work, three TetGI constructs are designed and studied by cryo-EM: TetGI-D (dimeric post-2S construct), TetGI-T (trimeric post-2S construct) and TetGI-DS (dimeric pre-2SΔ5′ex construct).

slower in cleaving the remaining substrate than is TetGI-M, probably owing to the tighter binding of the product to the homooligomers (that is, via the avidity effect), which would limit the reaction rate[46] more significantly in the later stages of the assayed reaction.

Figure 4c shows the cryo-EM map of the C_2 dimer of TetGI-D. SE of the dimer allows finer 3D classification (Extended Data Fig. 4a–e), revealing two conformations that are different in P1 (Fig. 3d,e, Extended Data Fig. 4c and Supplementary Video 1): Fig. 4d shows the conformation with double-stranded P1 docked between the P4–P6 and P3–P9 domains[47–49] (Extended Data Fig. 4c); Fig. 4e shows the conformation without TetLEM bound, and the internal guide sequence (IGS) remains single-stranded and interacts with the minor groove of P2.1. Except for P1, these two conformations are almost identical, so the final refinement was conducted by combining these two classes, resulting in a final map (Fig. 4f and Extended Data Fig. 4c) with a better resolved core (core resolution is 3.68 Å overall resolution is 3.78 Å) than either class. The cryo-EM reconstruction of TetGI-T (Fig. 4g,h and Extended Data Fig. 5a–e) was similarly performed, and its resolution (overall 4.17 Å, core 4.10 Å) is lower than that of the TetGI-D (Fig. 4i). The overall architectures of the cryo-EM structures of TetGI-D and TetGI-T are mostly consistent with each other, and only subtle differences are observed in the helical directions of P6b and P9.2 owing to the different geometric restraints applied (Fig. 4j). Because TetGI-D has a higher overall resolution than TetGI-T, the map quality of TetGI-D is better than TetGI-T in most parts of the structure as indicated by the atom-averaged local resolution (Fig. 4k) and map intensity (Fig. 4l) calculated for each residue. However, for the peripheral P2.1 and P9.1–P9.2 regions, the map quality of TetGI-T is comparable to or even better than TetGI-D, reflecting the reduced conformational dynamics of these regions in TetGI-T owing to P9.2 being geometrically restrained in the assembled trimer. This result demonstrates the effectiveness of using cyclic oligomerization to mitigate structural flexibility and suggests that different regions can be preferentially stabilized by being configured within different oligomeric constructs.

Structural insights into the TetGI's peripheral domains. Our cryo-EM structures of the TetGI support the configurations of peripheral (P2–P2.1 and P9.1–P9.2) domains predicted by the decades-old computationally predicted model[35] and corroborated by a recent modest-resolution (6.8-Å) cryo-EM structure[13]. As expected for a subgroup IC1 intron, P2–P2.1 connects P4–P6 and P9.1–P9.2 via the tertiary base-pairings of P14 and P13, respectively, constituting a pseudo-continuous belt enclosing the core (Fig. 5a). Besides the overall structural organization, the present high-resolution cryo-EM structures provide a clearer view of the structural elements involving the peripheral domains (five important structural elements are highlighted by red dashed circles; see the descriptions for Fig. 5b,c below, and Extended Data Fig. 6a–c

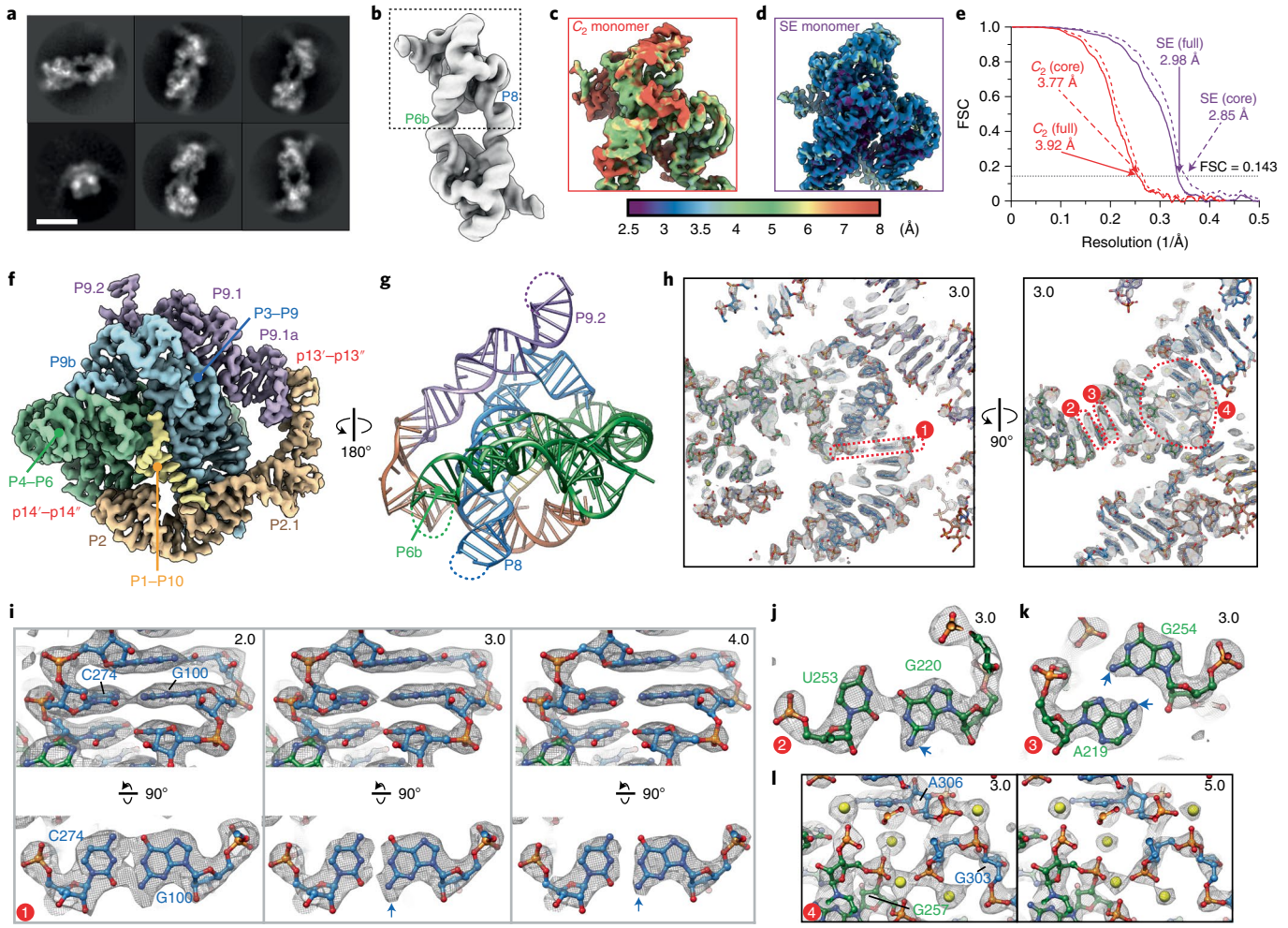


Fig. 3 | Sub-3-Å-resolution cryo-EM map of TetGI-DS, a dimeric pre-2SΔ5′ex construct of the TetGI, enabled by ROCK. a, Representative 2D class averages of TetGI-DS dimer. Scale bar, 10 nm. **b**, Cryo-EM map of TetGI-DS dimer refined with C_2 symmetry. **c**–**e**, Reconstruction with symmetry expansion (SE) improves the resolution as indicated by the local resolution maps of the monomer from the C_2 dimer (C_2 monomer, **c**) and the monomer after SE (SE monomer, **d**), and by the Fourier shell correlation (FSC) curves (**e**) calculated for the full maps (solid lines) and the core (dashed lines) of the C_2 monomer (red) and SE monomer (purple). **f**,**g**, Cryo-EM map (**f**) and the derived atomic model (**g**) colored according to the secondary structure in Fig. 2a,b. **h**, Two clipped views of the map (gray mesh) with the atomic model (sticks for RNA and yellow spheres for Mg^{2+} ions) illustrating the quality of the high-resolution cryo-EM map. Numbers in the corners of the figure panels (throughout all figures in this work) indicate the root mean square deviation (RMSD, analogous to σ for X-ray crystallography maps) contour level of the raw EM map (without any zoning or carving applied unless specified otherwise). Numbered (1 through 4) red dashed boxes mark the regions in **i**–**l**. **i**–**l**, Detailed views of an example WC base pair (G100-C274, **i**), a wobble base pair (G220-U253, **j**), a noncanonical base pair (A219-G254, **k**) and a region enclosing five Mg^{2+} ions (**l**). Blue arrows point to the map features corresponding to exocyclic amino groups. In **i**, two views related by a rotation of 90° are shown with maps rendered at three different contour levels to show the map features of base-to-base separation and sugar-phosphate backbone. In **l**, the same area is shown with maps rendered at two different contour levels to show the strong map intensities of Mg^{2+} ions.

for the tertiary interaction of P13, the four-way junction joining P9.1–P9.2 and P9a–P9b, and the junction connecting P2–P2.1 with P1 and P3–P8). The tertiary interaction P14 (Fig. 5b) consists of U43, G44 and C45 (p14′) pairing with A172, C170 and G169 (p14″), respectively, and an unexpected feature of P14 is the unpaired A171, which does not pair with U43 as previously predicted[35]. Also notable is that the bulged A210 from P4—which has been suggested to destabilize the folding of the isolated P4–P6 domain[37] and was either eliminated (ΔA210 or ΔC209, refs. [25,37]) or mutated (A210G, ref. [39]) in previous structural studies—participates in a long-range base-triple with the noncanonical C41-A46 pair (inset of Fig. 5b), possibly serving to stabilize P14. On the side of domain 9, we discovered a new long-range tertiary contact involving the interaction of G358 from J9.1/9.1a and the minor groove of P7 (Fig. 5c). This contact has likely functional implications because (1) a similar contact involving the peripheral P7.2 and the P7 also occurs in the Twort group I intron[50] (TwoGI, a subgroup IA2 intron; Extended Data Fig. 6d–f) and (2) the purine-rich internal loop of J9.1/9.1a is conserved in different subgroup IC1 and IE introns possessing the peripheral pseudo-continuous helix of P2.1–P13–P9.1a (Extended Data Fig. 6g), implying the presence of this new visualized contact in these introns. Considering the reaction site is located on the major-groove side of P7, this newly visualized tertiary contact is likely to buttress the reaction site from the other side. Preceding biochemical and chemical probing studies[51,52] indicate that domain 9 functions to stabilize the P3–P7 region, and its removal affects some of the intron's reactions involving the 3′ exon. However, previously, the major focus of the function of domain 9 has been directed

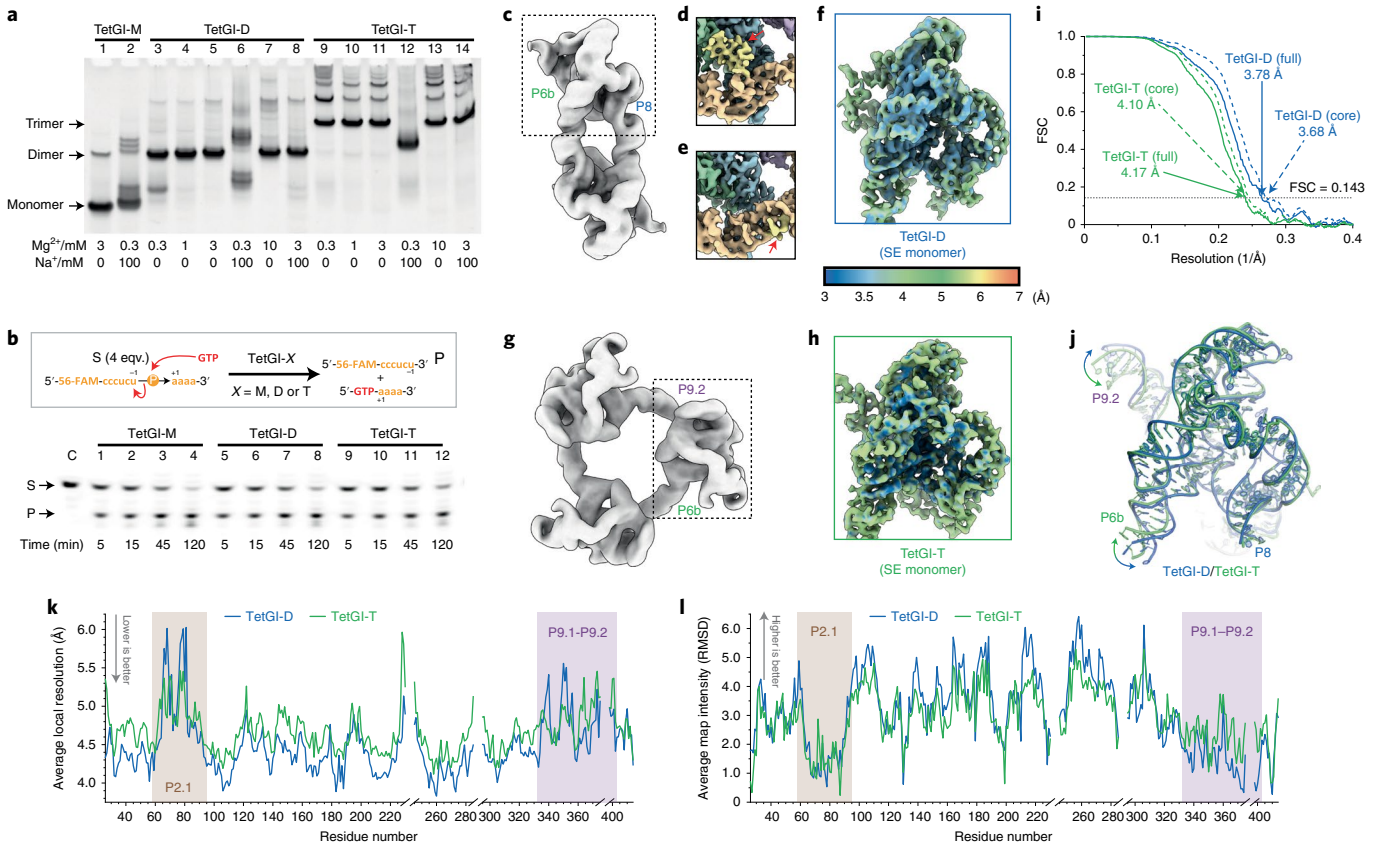


Fig. 4 | Assembly, activity and cryo-EM analyses of TetGI-D and TetGI-T. a, Native PAGE analyses of the assembly products under different annealing buffers. The bands for the target dimer and trimer, along with the monomer control (TetGI-M), are marked on the left of the gel image. **b**, TetGI-D and TetGI-T are catalytically active as determined by the *trans*-acting endonucleolytic activity. In the presence of each intron construct (0.5 μM of monomer unit), the substrate (S, 2 μM; 4 equivalents (eqv.) of intron monomer unit), a 10-nt single-stranded RNA 5′ labeled with 6-fluorescein amidite (56-FAM), is cleaved by exogenic GTP (1 mM) at 37 °C. **c**, Cryo-EM map of TetGI-D dimer refined with C_2 symmetry. **d**,**e**, 3D classification of the SE monomers of TetGI-D results in two distinct conformations with (**d**) or without (**e**) TetLEM bound. Red arrows indicate the docked double-stranded P1 (**d**) and the single-stranded IGS (**e**). **f**, Local resolution map of the SE monomer of TetGI-D dimer refined by combining the two classes in **d** and **e** (see Extended Data Fig. 4c for details of the reconstruction workflow). **g**, Cryo-EM map of TetGI-T trimer refined with C_3 symmetry. **h**, Local resolution map of the SE monomer of TetGI-T trimer. The shown map is refined by combining the two classes corresponding to the conformations with double-stranded P1 and single-stranded IGS (Extended Data Fig. 5c). **i**, FSC curves calculated for the full maps (solid lines) and the core (dashed lines) of the TetGI-D (blue) and TetGI-T (green) SE monomers. **j**, Overlay of the cryo-EM structures of TetGI-D (blue) and TetGI-T (green). Double-headed arrows mark the structural divergence of P6b and P9.2 of the two structures. **k**,**l**, Plots of the atom-averaged local resolution (**k**) and map intensity (**l**) calculated for each residue of TetGI-D (blue) and TetGI-T (green). Though TetGI-T has an overall lower map quality (marked by higher values in **k** and lower values in **l**) than TetGI-D, it has comparable or better map quality in the regions of P2.1 and P9.1–P9.2 (highlighted by brown and purple shades, respectively) because these regions are geometrically restrained in TetGI-T.

to the apical loop of P9.1a (that is, p13″) as its participation in the P13 interaction is readily predicted by sequence complementarity[35]. Here, our ROCK-enabled high-resolution structures provide an additional structural basis for the role of domain 9.

J8/7 and active-site magnesium ions of TetGI. In TetGI-DS (a pre-2SΔ5′ex construct), the IGS is left single-stranded, prompting us to compare J8/7 (functioning as a docking site for P1) of this construct with that of the AzoGI crystal structure[40] (with a double-stranded P1–P2; the AzoGI's P1–P2 is equivalent to the TetGI's P1) and of the TetGI core structure[39] (without P1 or the IGS). As shown in Fig. 5d, the configurations of J8/7 in TetGI-DS and the AzoGI are nearly identical. Some discrepancies lie in the bound metal ions. For instance, we localize a Mg^{2+} ion, which is not present in the AzoGI structure[40], interacting with the phosphate oxygen atoms of A301 and A302 (corresponding to the AzoGI's A167 and A168), and this Mg^{2+} ion has been predicted to stabilize the stack-exchange junction at P3–P8 (ref. [53]). As our cryo-EM model and the previous X-ray model[39] of the TetGI differ considerably in J8/7 (Fig. 5e, left), we attempted to use our cryo-EM structure to rebuild and re-refine the X-ray model. Upon doing so, we observed substantial improvement of the X-ray map quality and refinement statistics (Supplementary Fig. 4), and the structure differences in this region are largely eliminated (Fig. 5e, right). These results indicate that J8/7 is pre-organized for P1 docking, and also demonstrate the utility of cryo-EM structures in improving the RNA model building that has been difficult for low-resolution X-ray data.

In the TetGI-DS structure, we observed an active-site Mg^{2+} (Fig. 5f) corresponding to M1 in the AzoGI ribo-ΩG pre-2S crystal structure[54]. M1 functions as an activator of the nucleophile and the scissile phosphate in the second step of splicing[55–58], and its presence in the 5′ exon-free pre-2SΔ5′ex intron suggests its possible role of activating a water molecule as the nucleophile of the 3′ splice site hydrolysis reaction[42]. However, the other metal M2 (stabilizing the leaving group in the second step of splicing[59]) in the AzoGI structure is not observed in TetGI-DS, likely owing to

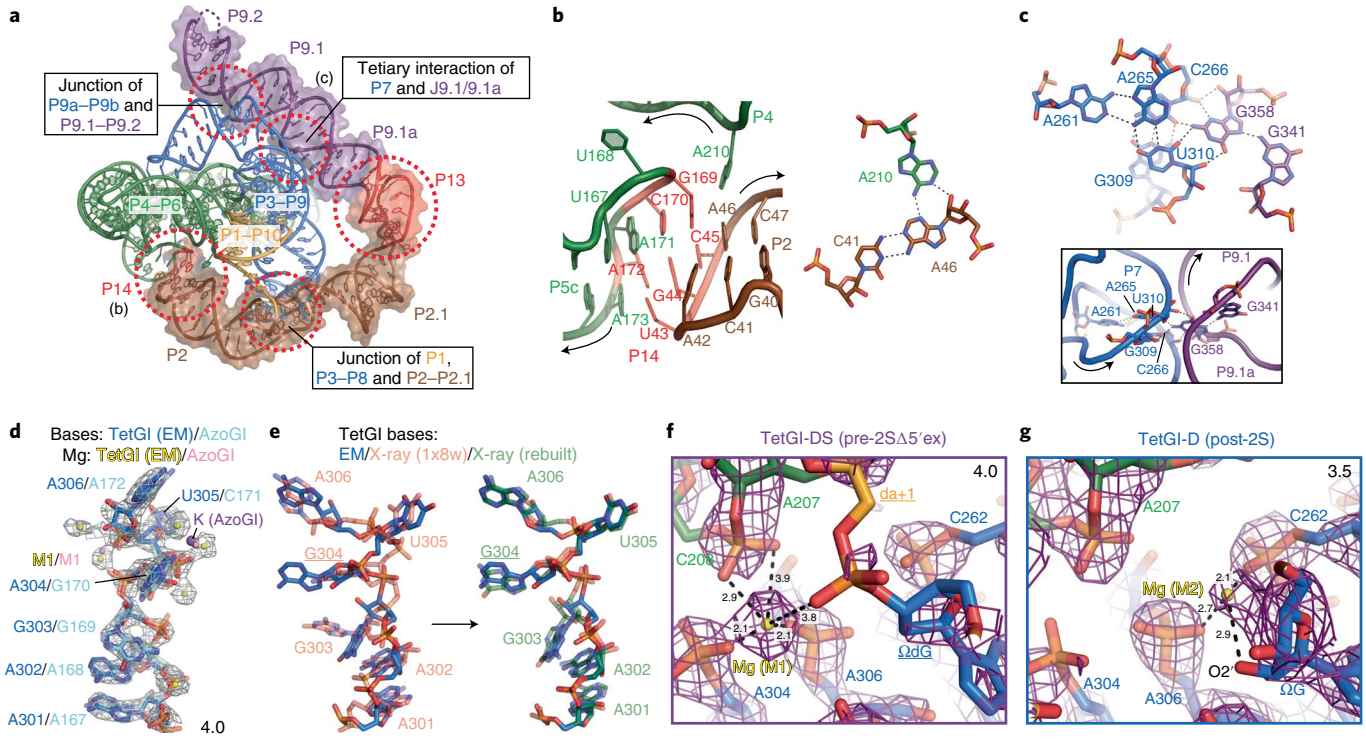


Fig. 5 | Structural insights gained from the cryo-EM structures of the TetGI. a, An overview of tertiary interactions and junction structures involving the peripheral domains (highlighted by surface rendering) of the TetGI. Unless specified otherwise, the TetGI structural elements are from the atomic model of the construct TetGI-DS. **b**, Structural details of P14. The right panel shows the base-triple formed by the bulged-out A210 from P4 and the noncanonical C41-A46 pair from P2. **c**, A newly visualized tertiary contact formed by the docking of G358 of J9.1/9.1a into the minor groove of P7. **d**, A comparison of J8/7 region of the TetGI-DS cryo-EM structure and the *Azoarcus* group I intron (AzoGI) crystal structure[40] (PDB: 1U6B) reveals close structural resemblance. Gray mesh is the EM map of TetGI-DS contoured at 4.0 RMSD level and carved within 2.0 Å of the displayed atoms. **e**, The high-resolution cryo-EM structure enables the rebuilding and re-refinement of J8/7 of the previous crystal structure of the TetGI core[39] (PDB: 1X8W). The structures before (light red) and after (light green) rebuilding are overlayed with the TetGI cryo-EM structure (blue). Underlined nucleotides indicate the mutations introduced for the crystallographic study. **f**,**g**, Active-site Mg^{2+} ions, M1 and M2, are observed in the cryo-EM maps (magenta meshes) of TetGI-DS (**f**) and TetGI-D (**g**), respectively. In the second step of splicing, M1 activates the attacking nucleophile and M2 stabilizes the leaving group. Dashed lines mark the distances (Å) between the Mg^{2+} ions and RNA heteroatoms.

the deoxy substitution of ΩG (G414 for the TetGI), which lacks the 2′-OH coordinating to M2 (ref. [56]) (M2 is present in the deoxy-ΩG pre-2S AzoGI crystal structure[40] as a monovalent metal, but in our cryo-EM study, the only metal ion present is Mg^{2+}). The incapability of recruiting a divalent cation at M2 site thus is likely the cause for the suppression of the 3′ splice site hydrolysis reaction by deoxy-ΩG mutation (Extended Data Fig. 2i). Nonetheless, M2 could be spotted in the cryo-EM map of TetGI-D, which has a native ribo-ΩG (Fig. 5g; M1 could not be confidently built in TetGI-D owing to the limited resolution). In our structures, we could not unambiguously localize the possible third active-site metal ion implicated in some functional studies[56].

Extending the application of ROCK to smaller RNA targets. To demonstrate the general utility of ROCK, we next set out to assess it for smaller RNA structures below 100 kDa and with more affordable, lower-performance instruments (Polara microscope with K2 camera, rather than Krios microscope with K3 camera for the TetGI structures; Methods and Supplementary Table 3). The AzoGI is a subgroup IC3 intron of 206 nt (~70 kDa), and compared with the TetGI, it lacks the extensive peripheral domains that help stabilize the overall structure. Both the smaller molecular weight and the presumed higher flexibility make the AzoGI a more challenging target for cryo-EM, although its smaller size and simpler fold make it more attractive for crystallographic studies[40,54,60]. On the basis of its crystal structures, we designed a post-2S trimeric construct of the AzoGI, AzoGI-T, by engineering P5a and P8a (Fig. 6a,b; see Supplementary Fig. 5 for a dimeric construct designed by engineering P6a and P8a, which was not further pursued owing to its similarity to TetGI-D). Similar to the TetGI-D and TetGI-T (which are also post-2S constructs, as shown in Fig. 2c), two deoxy substitutions are introduced in the in-*trans* added ligated exon mimic (AzoLEM). Using a workflow (Extended Data Figs. 7a–d and 8a–e) similar to that for the TetGI constructs, we obtained a 4.9-Å structure (Fig. 6c,d) corresponding to a conformation with P1–P10 tightly docked (Fig. 6e), which is from the most populated class from 3D classification (Extended Data Fig. 8c) and is similar to the post-2S AzoGI crystal structure[60]. We also reconstructed the cryo-EM map from another class of particles corresponding to an alternative open conformation (Fig. 6f). Although this map is of a substantially lower resolution (~8 Å), the relative movement of P2–P1–P10 and P4–P5 could be clearly discerned (Fig. 6f and Supplementary Video 2). Such a large conformational change has not been observed among different constructs in previous crystallographic studies[40,54,60] of the AzoGI, likely owing to the constraint applied by the similar crystal-packing interfaces.

Lastly, we used ROCK to solve an even smaller target—the *Fusobacterium nucleatum* FMN riboswitch of 112 nt (~35 kDa).

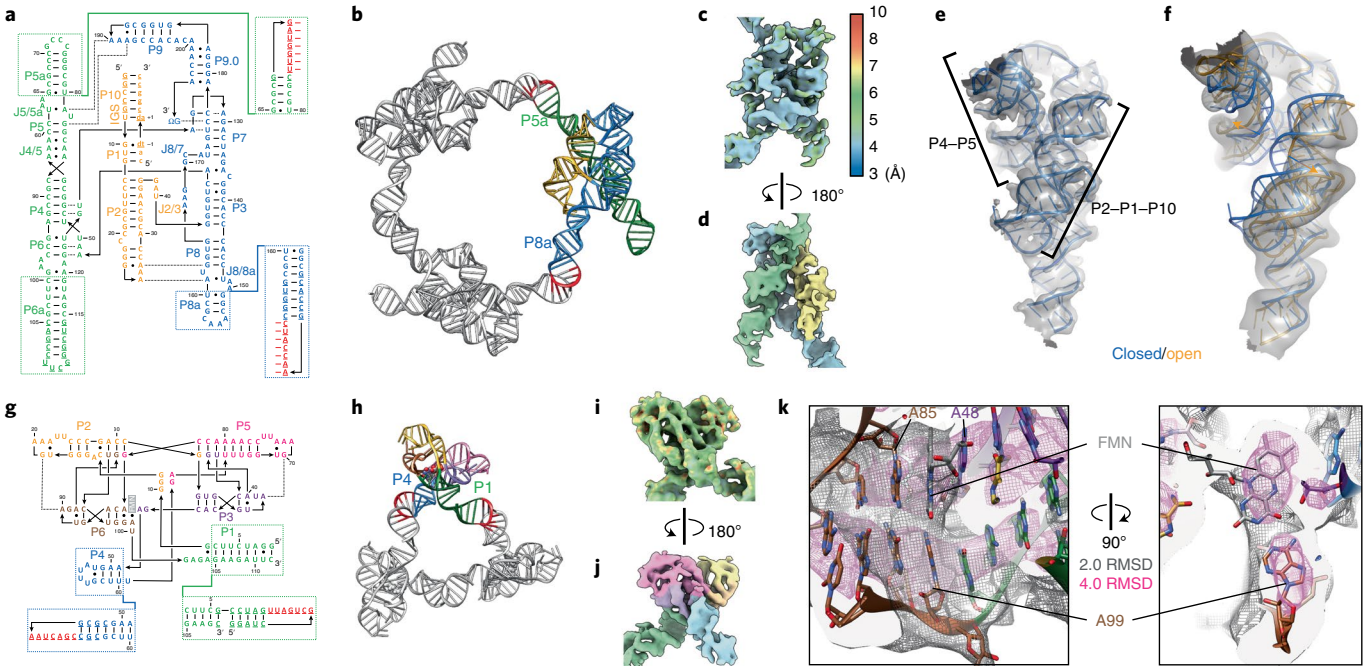


Fig. 6 | The homomeric self-assembly strategy applied to two smaller structured RNAs. a, Sequence and secondary structure of the post-2S AzoGI. Dashed boxes mark the engineering of P5a and P8a in the construct for trimeric self-assembly, AzoGI-T. **b**, Computer model of the assembled homotrimer of AzoGI-T. **c**,**d**, Cryo-EM maps rendered by local resolution (**c**) and colored on the basis of the secondary structure in **a** (**d**). **e**,**f**, A comparison of the refined cryo-EM maps of the best-resolved (4.9 Å same as the map in **c** and **d**) class of particles corresponding to the closed conformation (**e**) from 3D classification of the SE monomer and another less-resolved (~8.0 Å) one corresponding to the open conformation. In **f**, the fitted atomic model of the open conformation (orange) is overlayed with that of the close conformation (blue), and the arrows mark the structural movement of P2–P1–P10 and P4–P5. **g**, Sequence and secondary structure of the *Fusobacterium nucleatum* FMN riboswitch. Dashed boxes mark the engineering of P1 and P4 in the construct for trimeric self-assembly, FMNrbsw-T. **h**, Computer model of the assembled homotrimer of FMNrbsw-T. **i**,**j**, Cryo-EM maps rendered by local resolution (**i**) and colored on the basis of the secondary structure in **g** (**j**). The color key for local resolution in **i** is the same as in **c**. **k**, Clipped views showing the cryo-EM map of the FMN ligand and its binding environment. The cryo-EM map is rendered at two different contour levels (gray mesh at 2.0 RMSD and magenta mesh at 4.0 RMSD).

On the basis of its crystal structure[61], we installed the kissing-loop sequences onto P1 and P4 for a homotrimeric assembly (Fig. 6g,h; referred to as FMNrbsw-T). Interestingly, the RNA of FMNrbsw-T assembles into two different homooligomers, dimer or trimer, when annealed in different buffer conditions (Extended Data Fig. 9a–c). Ligand-binding assays (Extended Data Fig. 9d,e) show that the RNA within the expected trimer is in its functionally relevant conformation, whereas the dimer is less competent in ligand binding and is likely a kinetic product or as a result of an alternate conformation of the apo riboswitch[62]. The trimer was natively purified and mixed with the FMN ligand, and then subjected to cryo-EM analysis (Extended Data Fig. 10a–e). Due to the small size, the final refinement was conducted on the whole trimer particles with one monomer focused (Extended Data Fig. 10c), resulting in a 5.9-Å resolution cryo-EM map of the focused monomer (Fig. 6i,j). The bound ligand, along with its binding environment, can be visualized in this modest-resolution cryo-EM map owing to the strong map intensities of the ligand and its vicinity (Fig. 6k), probably as a result of the more restricted motion of the ligand-binding site relative to the peripheries (Supplementary Video 3). While we attribute the limited resolutions for AzoGI-T and FMNrbsw-T to their smaller sizes and increased structural flexibility compared with the TetGI constructs, or to the less sophisticated instruments, we could not exclude the possibility that other experimental variables, such as ice thickness, may also affect the achievable resolution.

Technical requirements of ROCK. Here we share important technical requirements of ROCK. First, the workflow of ROCK starts with construct engineering that requires the RNA having at least two functionally nonessential helices for installing kissing-loop sequences. For natural RNAs, nonessential helices can typically be identified by phylogenetic analysis of homologs from different organisms[63], and two nonessential helices can be readily identified within a homolog possessing more extensive peripheral elements. Second, construct engineering can be substantially facilitated by the knowledge of the relative spatial positions of the nonessential helices to be engineered, which can be obtained from solved homologs or partial structures (for example by crystallography or NMR), computer modeling[35], atomic force microscopy[64], small-angle X-ray scattering[65], comparative gel electrophoresis[66], and preliminary low-resolution cryo-EM models[13]. It is worth noting that a known structure is not a prerequisite for construct engineering: for instance, for designing TetGI-T, we used the computationally predicted model by Westhof and coworkers in 1996 (ref. [35]), which preceded the appearance of the crystal structure of any complete group I intron. Lastly, for all structural-biology studies, one must ensure the biological meaningfulness of the obtained structures. Specifically for ROCK, to exclude the possibility that de novo-designed oligomerization alters the native structures and functions, we have assayed the activities of the engineered, assembled RNAs.

Discussion

In this work, we have determined the cryo-EM structures of three RNAs with sizes ranging from ~35 to ~140 kDa and belonging to two important categories of functional RNAs—ribozymes and riboswitches—demonstrating the general utility of ROCK. The core of

ROCK is construct engineering for homooligomeric self-assembly, which facilitates the whole pipeline of RNA cryo-EM determination from RNA-folding optimization to cryo-EM reconstruction. Previous RNA cryo-EM studies[11–13] aiming for high-resolution structures typically entail collecting large datasets, which makes the workflow experimentally and computationally demanding, and do not identify different conformational states. (After the initial submission of our work (preprint accessible at https://doi.org/10.1101/2021.08.11.455951v1), Su et al. independently reported cryo-EM structures of the TetGI from stand-alone RNA constructs by collecting large datasets[67]. Their best-resolved structure at 3.06-Å resolution (~3.0 Å for the core) is of the pre-2S state, which is close to our pre-2SΔ5′ex (TetGI-DS) construct at 2.98-Å resolution (2.85 Å for the core). It is worth noting that Su et al. used 2.8 times as many particles as ours (SE monomers; Extended Data Fig. 3c) in the final refinement.) In contrast, our strategy directly deals with RNA's intrinsic molecular features that fundamentally limit the obtainable resolution: nonfunctional structural dynamics are mitigated with geometric restraints, and sample homogeneity is improved by natively purifying the target homooligomers. Furthermore, the resulting symmetric assemblies are also preferred targets for cryo-EM reconstruction: (1) the characteristic shapes are more convenient for initial alignment of the particles; and (2) special data-processing procedures, such as SE and individual subunit-focused classification and refinement, can be utilized to achieve higher resolution and obtain structures of different conformations.

ROCK substantially improves the utility and performance of cryo-EM in RNA structural studies, making it a complementary approach[68] to the currently more prevalent X-ray crystallography with respect to construct choice, conformational dynamics, and accessible structural information (see Supplementary Note 2 for detailed discussions). For example, we demonstrate that native Mg^{2+} ions can be localized in the cryo-EM maps of different TetGI constructs with resolution ranging from 2.98 to 4.17 Å. This cryo-EM approach to metal ion localization is complementary to X-ray crystallography, where, at similar or worse resolutions typically obtained for large RNAs, heavy metals are often introduced as mimics of small native metals by crystal soaking to provide anomalous scattering signals. Future development of ROCK may potentially generate RNA cryo-EM structures of higher resolution via the introduction of additional structure-stabilizing RNA-RNA interactions or RNA-binding proteins. ROCK will also benefit from the advances of computational tools for RNA cryo-EM map interpretation and model building to address the challenges that have not been encountered in protein cryo-EM or RNA X-ray crystallography. Taken together, we believe that ROCK will help further unleash the largely unexplored potential of cryo-EM in RNA structural studies, opening new opportunities for elucidating the mechanisms of functional RNAs and potentially facilitating structure-based approaches to RNA-targeting therapeutics[69].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41592-022-01455-w.

Received: 16 August 2021; Accepted: 9 March 2022;
Published online: 2 May 2022

References

1. Mortimer, S. A., Kidwell, M. A. & Doudna, J. A. Insights into RNA structure and function from genome-wide studies. *Nat. Rev. Genet.* **15**, 469–479 (2014).
2. Serganov, A. & Patel, D. J. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat. Rev. Genet.* **8**, 776–790 (2007).
3. Hangauer, M. J., Vaughn, I. W. & McManus, M. T. Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. *PLoS Genet.* **9**, e1003569 (2013).
4. Robertson, D. L. & Joyce, G. F. Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature* **344**, 467–468 (1990).
5. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).
6. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510 (1990).
7. Zhang, J. & Ferre-D'Amare, A. R. New molecular engineering approaches for crystallographic studies of large RNAs. *Curr. Opin. Struct. Biol.* **26**, 9–15 (2014).
8. Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* **9**, 1665–1672 (1990).
9. Nogales, E. The development of cryo-EM into a mainstream structural biology technique. *Nat. Methods* **13**, 24–27 (2016).
10. Qu, G. et al. Structure of a group II intron in complex with its reverse transcriptase. *Nat. Struct. Mol. Biol.* **23**, 549–557 (2016).
11. Li, S. et al. Structural basis of amino acid surveillance by higher-order tRNA-mRNA interactions. *Nat. Struct. Mol. Biol.* **26**, 1094–1105 (2019).
12. Zhang, K. et al. Cryo-EM structure of a 40 kDa SAM-IV riboswitch RNA at 3.7 A resolution. *Nat. Commun.* **10**, 5511 (2019).
13. Kappel, K. et al. Accelerated cryo-EM-guided determination of three-dimensional RNA-only structures. *Nat. Methods* **17**, 699–707 (2020).
14. Hyeon, C., Dima, R. I. & Thirumalai, D. Size, shape, and flexibility of RNA structures. *J. Chem. Phys.* **125**, 194905 (2006).
15. Seeman, N. C. Nanomaterials based on DNA. *Annu. Rev. Biochem.* **79**, 65–87 (2010).
16. Seeman, N. C. *Structural DNA Nanotechnology* (Cambridge University Press, 2016).
17. Seeman, N. C. & Sleiman, H. F. DNA nanotechnology. *Nat. Rev. Mater.* **3**, 17068 (2017).
18. Guo, P. The emerging field of RNA nanotechnology. *Nat. Nanotechnol.* **5**, 833–842 (2010).
19. Grabow, W. W. & Jaeger, L. RNA self-assembly and RNA nanotechnology. *Acc. Chem. Res.* **47**, 1871–1880 (2014).
20. Kruger, K. et al. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell* **31**, 147–157 (1982).
21. Pujari, N. et al. Engineering crystal packing in RNA structures I: past and future strategies for engineering RNA packing in crystals. *Cryst.* **11**, 8 (2021).
22. Ferre-D'Amare, A. R., Zhou, K. & Doudna, J. A. A general module for RNA crystallization. *J. Mol. Biol.* **279**, 621–631 (1998).
23. Ferre-D'Amare, A. R. & Doudna, J. A. Crystallization and structure determination of a hepatitis delta virus ribozyme: use of the RNA-binding protein U1A as a crystallization module. *J. Mol. Biol.* **295**, 541–556 (2000).
24. Ferre-D'Amare, A. R. Use of the spliceosomal protein U1A to facilitate crystallization and structure determination of complex RNAs. *Methods* **52**, 159–167 (2010).
25. Ye, J. D. et al. Synthetic antibodies for specific recognition and crystallization of structured RNA. *Proc. Natl Acad. Sci. USA* **105**, 82–87 (2008).
26. Koldobskaya, Y. et al. A portable RNA sequence whose recognition by a synthetic antibody facilitates structural determination. *Nat. Struct. Mol. Biol.* **18**, 100–106 (2011).
27. Lee, A. J. & Crothers, D. M. The solution structure of an RNA loop–loop complex: the ColE1 inverted loop sequence. *Structure* **6**, 993–1007 (1998).
28. Goodsell, D. S. & Olson, A. J. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 105–153 (2000).
29. Jones, C. P. & Ferre-D'Amare, A. R. RNA quaternary structure and global symmetry. *Trends Biochem. Sci.* **40**, 211–220 (2015).
30. Bou-Nader, C. & Zhang, J. Structural Insights into RNA dimerization: motifs, interfaces and functions. *Molecules* **25**, 12 (2020).
31. Bindewald, E., Grunewald, C., Boyle, B., O'Connor, M. & Shapiro, B. A. Computational strategies for the automated design of RNA nanoscale structures from building blocks using NanoTiler. *J. Mol. Graph Model* **27**, 299–308 (2008).
32. Hougland, J. L., Piccirilli, J. A., Forconi, M., Lee, J. & Herschlag, D. in *RNA World* 3rd edn (eds Gesteland, R. F., Atkins, J. F. & Cech, T. R.) 133–205 (Cold Spring Harbor Laboratory Press, 2006).
33. Golden, B. L. in *Ribozymes and RNA Catalysis* 178–200 (The Royal Society of Chemistry, 2007).
34. Michel, F. & Westhof, E. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* **216**, 585–610 (1990).

35. Lehnert, V., Jaeger, L., Michel, F. & Westhof, E. New loop-loop tertiary interactions in self-splicing introns of subgroup IC and ID: a complete 3D model of the Tetrahymena thermophila ribozyme. *Chem. Biol.* **3**, 993–1009 (1996).
36. Cate, J. H. et al. Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **273**, 1678–1685 (1996).
37. Juneau, K., Podell, E., Harrington, D. J. & Cech, T. R. Structural basis of the enhanced stability of a mutant ribozyme domain and a detailed view of RNA–solvent interactions. *Structure* **9**, 221–231 (2001).
38. Golden, B. L., Gooding, A. R., Podell, E. R. & Cech, T. R. A preorganized active site in the crystal structure of the *Tetrahymena* ribozyme. *Science* **282**, 259–264 (1998).
39. Guo, F., Gooding, A. R. & Cech, T. R. Structure of the *Tetrahymena* ribozyme: base triple sandwich and metal ion at the active site. *Mol. Cell* **16**, 351–362 (2004).
40. Adams, P. L., Stahley, M. R., Kosek, A. B., Wang, J. & Strobel, S. A. Crystal structure of a self-splicing group I intron with both exons. *Nature* **430**, 45–50 (2004).
41. Zaug, A. J., Been, M. D. & Cech, T. R. The Tetrahymena ribozyme acts like an RNA restriction endonuclease. *Nature* **324**, 429–433 (1986).
42. Inoue, T., Sullivan, F. X. & Cech, T. R. New reactions of the ribosomal RNA precursor of *Tetrahymena* and the mechanism of self-splicing. *J. Mol. Biol.* **189**, 143–165 (1986).
43. Woodson, S. A. Metal ions and RNA folding: a highly charged topic with a dynamic future. *Curr. Opin. Chem. Biol.* **9**, 104–109 (2005).
44. Liu, D. et al. Branched kissing loops for the construction of diverse RNA homooligomeric nanostructures. *Nat. Chem.* **12**, 249–259 (2020).
45. Rook, M. S., Treiber, D. K. & Williamson, J. R. An optimal Mg^{2+} concentration for kinetic folding of the *Tetrahymena* ribozyme. *Proc. Natl Acad. Sci. U. S. A.* **96**, 12471–12476 (1999).
46. Herschlag, D. & Cech, T. R. Catalysis of RNA cleavage by the *Tetrahymena thermophila* ribozyme. 2. Kinetic description of the reaction of an RNA substrate that forms a mismatch at the active site. *Biochemistry* **29**, 10172–10180 (1990).
47. Pyle, A. M. & Cech, T. R. Ribozyme recognition of RNA by tertiary interactions with specific ribose 2′-OH groups. *Nature* **350**, 628–631 (1991).
48. Herschlag, D., Eckstein, F. & Cech, T. R. Contributions of 2′-hydroxyl groups of the RNA substrate to binding and catalysis by the Tetrahymena ribozyme. An energetic picture of an active site composed of RNA. *Biochemistry* **32**, 8299–8311 (1993).
49. Strobel, S. A. & Cech, T. R. Tertiary interactions with the internal guide sequence mediate docking of the P1 helix into the catalytic core of the *Tetrahymena* ribozyme. *Biochemistry* **32**, 13593–13604 (1993).
50. Golden, B. L., Kim, H. & Chase, E. Crystal structure of a phage Twort group I ribozyme-product complex. *Nat. Struct. Mol. Biol.* **12**, 82–89 (2005).
51. Barfod, E. T. & Cech, T. R. Deletion of nonconserved helices near the 3′ end of the rRNA intron of *Tetrahymena thermophila* alters self-splicing but not core catalytic activity. *Genes Dev.* **2**, 652–663 (1988).
52. Laggerbauer, B., Murphy, F. L. & Cech, T. R. Two major tertiary folding transitions of the *Tetrahymena* catalytic RNA. *EMBO J.* **13**, 2669–2676 (1994).
53. Denesyuk, N. A. & Thirumalai, D. How do metal ions direct ribozyme folding? *Nat. Chem.* **7**, 793–801 (2015).
54. Stahley, M. R. & Strobel, S. A. Structural evidence for a two-metal-ion mechanism of group I intron splicing. *Science* **309**, 1587–1590 (2005).
55. Piccirilli, J. A., Vyle, J. S., Caruthers, M. H. & Cech, T. R. Metal ion catalysis in the *Tetrahymena* ribozyme reaction. *Nature* **361**, 85–88 (1993).
56. Shan, S., Yoshida, A., Sun, S., Piccirilli, J. A. & Herschlag, D. Three metal ions at the active site of the *Tetrahymena* group I ribozyme. *Proc. Natl Acad. Sci. USA* **96**, 12299–12304 (1999).
57. Yoshida, A., Sun, S. & Piccirilli, J. A. A new metal ion interaction in the *Tetrahymena* ribozyme reaction revealed by double sulfur substitution. *Nat. Struct. Biol.* **6**, 318–321 (1999).
58. Kuo, L. Y. & Piccirilli, J. A. Leaving group stabilization by metal ion coordination and hydrogen bond donation is an evolutionarily conserved feature of group I introns. *Biochim. Biophys. Acta* **1522**, 158–166 (2001).
59. Weinstein, L. B., Jones, B. C., Cosstick, R. & Cech, T. R. A second catalytic metal ion in group I ribozyme. *Nature* **388**, 805–808 (1997).
60. Lipchock, S. V. & Strobel, S. A. A relaxed active site after exon ligation by the group I intron. *Proc. Natl Acad. Sci. USA* **105**, 5699–5704 (2008).
61. Serganov, A., Huang, L. & Patel, D. J. Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch. *Nature* **458**, 233–237 (2009).
62. Wilt, H. M., Yu, P., Tan, K., Wang, Y. X. & Stagno, J. R. FMN riboswitch aptamer symmetry facilitates conformational switching through mutually exclusive coaxial stacking configurations. *J. Struct. Biol. X* **4**, 100035 (2020).
63. Michel, F. & Costa, M. Inferring RNA structure by phylogenetic and genetic analyses. *Cold Spring Harb. Monogr. Ser.* **35**, 175–202 (1998). %@ 0270-1847.
64. Schön, P. Imaging and force probing RNA by atomic force microscopy. *Methods* **103**, 25–33 (2016).
65. Chen, Y. & Pollack, L. SAXS studies of RNA: structures, dynamics, and interactions with partners. *Wiley Interdiscip. Rev. RNA* **7**, 512–526 (2016).
66. Lilley, D. M. Analysis of branched nucleic acid structure using comparative gel electrophoresis. *Q. Rev. Biophys.* **41**, 1–39 (2008).
67. Su, Z. et al. Cryo-EM structures of full-length *Tetrahymena* ribozyme at 3.1 A resolution. *Nature* **596**, 603–607 2021.
68. Wang, H. W. & Wang, J. W. How cryo-electron microscopy and X-ray crystallography complement each other. *Protein Sci.* **26**, 32–39 (2017).
69. Warner, K. D., Hajdin, C. E. & Weeks, K. M. Principles for targeting RNA with drug-like small molecules. *Nat. Rev. Drug Discov.* **17**, 547–558 (2018).

Methods

RNA construct design. The software NanoTiler[31], developed by B. Shapiro's group (website: https://fscnpl-bnkly01p.ncifcrf.gov/software.html), was used to design the RNA constructs with kissing loops installed for homomerization. Throughout this work, we chose the 7-bp kissing-loop motif which is adapted from the RNA I–RNA II complex of the *Escherichia coli* ColE1 plasmid[70] and has been determined by NMR[27] (PDB: 2BJ2) to have an included angle of ~120°. For TetGI-D and TetGI-DS, the design was based on crystal structure[39] of the core domains of the TetGI (PDB: 1X8W). Before being imported into NanoTiler as structural modules, the PDB files of 1X8W (the TetGI core) and 2BJ2 (the 7-bp kissing-loop motif) were first edited so that the helices (P6b and P8 for the TetGI core, and the two stems of the 7-bp kissing-loop motif) to be connected in the final construct were trimmed to 3 bp. Then NanoTiler's command 'clone' was used to generate another copy of these structural models; commands 'genhelixconstraint' and 'opthelices' were used to generate ideal A-form RNA helices of certain lengths for connecting the TetGI core and the 7-bp KL; the 'start_score' in the output of the 'opthelices' for placing the last connecting helix indicates how good the ring closure is, with a lower value indicating more favorable ring closure. The processes can be automated with the 'NanoScript' of NanoTiler in which 'foreach' loops can be incorporated to iterate through connecting helices of different lengths to perform the screening of the optimal lengths for ring closure. Similarly, the trimeric construct TetGI-T was designed based on the computational model[35] of the complete TetGI kindly provided by E. Westhof; AzoGI-T based on the crystal structure[40] of the AzoGI (PDB: 1U6B); and FMNrbsw-T based on the crystal structure[61] of the FMN riboswitch bound to FMN (PDB: 3F2Q). After installing the kissing-loop sequence and connector helices onto the peripheral stems of each target RNA, the secondary structure of each newly formed stem-loop was further checked by Mfold[71] to ensure that the new stem-loop folds correctly.

RNA preparation and nanostructure assembly. All RNA molecules were synthesized by in vitro transcription (IVT) using the HiScribe T7 High Yield RNA Synthesis Kit from the New England Biolabs (NEB). The corresponding DNA templates for in vitro transcription were the PCR products of gene fragments (sequences shown in Supplementary Table 2) ordered from WuXi Qinglan Biotech. The PCR experiments were conducted using the Q5 Hot Start High-Fidelity DNA Polymerase (NEB) following the recommended protocol provided by manufacturer. Primers and other modified oligonucleotides were ordered from the Integrated DNA Technologies (IDT). All IVT RNA molecules were purified by denaturing PAGE (containing 7 M urea), then ethanol-precipitated and suspended in pure water. The RNA concentration was determined by measuring OD_{260}.

Protocols for RNA nanostructure assembly were adapted from a previous publication[44]. RNAs were first denatured at 85 °C for 1 min and snap-cooled on ice. Then, the annealing buffers containing 20 mM of Tris-acetate (pH 8.0) and varied concentrations of Mg^{2+} (using 100 mM Mg^{2+} stock solution containing 110 mM of $MgCl_2$ and 10 mM of EDTA) and Na^+ (using 1 M Na^+ stock solution containing 1 M of NaCl) ions were added to the denatured RNAs (to a final RNA concentration of ~800 nM). The mixtures were then annealed from 70 °C to 4 °C with the following protocol: 70 °C to 50 °C over 6 min, 50 °C to 37 °C over 20 min, and 37 °C to 4 °C over 2 hr. The annealed RNAs were analyzed by native PAGE in 0.5 × TBE buffer supplemented by 3 mM of $MgCl_2$ (so that the final concentration of free Mg^{2+} is 2 mM because 1 mM of EDTA is included in 0.5 × TBE). To prepare the samples for cryo-EM experiments, the cation contents in the annealing buffers were chosen based on the native PAGE results: for TetGI-DS, TetGI-D, TetGI-T and AzoGI-T, the annealing buffer contains 1 mM of Mg^{2+} and 1.2 equivalents (eqv.) of associated synthetic oligonucleotides (dTetCIRC for TetGI-DS; TetLEM for TetGI-D and TetGI-T; and AzoLEM for AzoGI-T); for FMNrbsw-T, the annealing buffer contains 0.3 mM of Mg^{2+} and 10 mM of Na^+ supplemented by 1 mM of FMN. The assembled nanostructures were purified by preparative native PAGE, eluted to the annealing buffer of 3 mM Mg^{2+}, and concentrated to ~1 μg/μL using Amicon Ultra centrifugal filters (MWCO 30 kDa).

Cryo-EM sample preparation and data acquisition. Before grid preparation, Mg^{2+} and other components (such as short oligonucleotides or ligand) were added to the purified and concentrated RNA assemblies. For the four ribozyme constructs (TetGI-DS, TetGI-D, TetGI-T and AzoGI-T), Mg^{2+} was added to a final concentration of 30 mM to ensure complete folding[53]. Additional synthetic oligonucleotide was added to each of the ribozyme construct in case of the dissociation at equilibrium or during the purification processes: for TetGI-DS, 1.2 equivalents of dTetCIRC was added; for TetGI-D and TetGI-T, 3 equivalents of TetLEM was added; and for AzoGI-T, 3 equivalents of AzoLEM was added. For FMNrbsw-T, Mg^{2+} was added to a final concentration of 10 mM, and the ligand FMN to a final concentration of 200 μM. After the addition of Mg^{2+} and other components, the samples were left at room temperature for at least 15 minutes before grid preparation.

Each RNA sample was applied to a glow-discharged Quantifoil (R1.2/1.3, 400 mesh) holey carbon grid. The three TetGI samples were vitrified in liquid ethane with a Thermo Fisher Vitrobot Mark IV (blot for ~7 s, force +12) and imaged on a Titan Krios microscope equipped with a Gatan K3 camera. The Krios movie stacks were acquired in counting mode, with a physical pixel size of 0.825 Å, total exposure of 47 e/Å^2 and at a defocus ranging from −0.8 to −2.0 μm. The AzoGI-T and FMNrbsw-T samples were frozen using a Gatan Cryoplunge 3 (blot ~3 s) and imaged on an FEI T30 Polara microscope equipped with a K2 summit detector. The Polara micrographs were acquired in super-resolution mode, with 0.62-Å super-resolution pixel size and 52 e/Å^2 exposure, with a defocus ranging from −0.8 to −2.0 μm. SerialEM[72] was used for collection of all datasets. More information about cryo-EM data collection can be found in Supplementary Table 3.

Cryo-EM data processing. The overall processing pipeline was similar for each dataset. Briefly, movie stacks were binned 2× if collected in super-resolution mode, then motion-corrected with MotionCor2 (ref. [73]). The contrast transfer function was subsequently computed with CTFFIND4[74]. Particle picking and 2D particle curation were performed with Simplified Application Managing Utilities of EM Labs (https://liao.hms.harvard.edu/samuel), a set of protocols built on the SPIDER[75] image processing system. About 2,000 particles were manually picked to generate 2D initial models, which were used to auto-pick 10% of the micrographs and generate refined 2D templates. The refined templates were then used to auto-pick from all micrographs. The particles were curated with 'samtree2dv3.py', which runs iterative principal component analysis (PCA), *k*-means clustering and multireference alignment. Selected particles were then imported to RELION 3.0 (ref. [76]), which was used for all subsequent processing steps. Briefly, the particles were downsampled and sorted with 3D classification, before unbinning and 3D refinement with enforced symmetry. As the constructs include monomers of uneven integrity, we used symmetry expansion[77] to address the pseudo-symmetry of the particles and separate the most stable monomers. Monomers were then classified through 3D classification without particle alignment, and selected classes were refined using local angle search refinement. Monomer resolution was calculated with the FSC = 0.143 criterion on half maps from independent halves of the dataset, while local resolution was determined with ResMap[78]. All maps were density modified with 'phenix.resolve_cryo_em'[79] that is recently incorporated in PHENIX[80], and no model information was provided to avoid any possible bias. More information about cryo-EM data processing can be found in Extended Data Figs. 3–5, 8 and 10.

Structure modeling. To build the structural model of the TetGI, we began with the core domains by rigid-body fitting the crystallographic model of the core (PDB: 1X8W) into the TetGI-DS map using UCSF Chimera[81] and performing several rounds of real-space refinements[82] in PHENIX[80]. Though the crystallographic model mostly agreed with our cryo-EM map, two regions of the core show the most significant discrepancy—the region of P9.0 and G414, which were not present in the crystallographic model, and the region near J8/7, which we believe was built incorrectly in the crystallographic model (Supplementary Fig. 4). These two regions were manually built in COOT[83]. To build the peripheral domains of the TetGI, ideal A-form helices were generated and fitted to the TetGI-DS map using UCSF Chimera[81] based on the known secondary structure. The rest of structure was manually built in COOT[83]. The most difficult region is P2.1 to P13, where the map density is relatively weaker and the secondary structure assignment was ambiguous. For instance, any residue from A87 through A90 could be possibly unpaired, and the exact number of base pairs in kissing loops P13 was not known. We manually tested different possibilities of secondary structure assignments for this region to ensure its successful joining to the helices on both sides. The cryo-EM map of TetGI-T was referenced to build the region of P2.1 to P13 because the map quality is better for this region than TetGI-DS. The complete TetGI model was finally real-space refined against the maps of the three TetGI constructs and manually checked and adjusted in COOT. The statistics of model refinement and validation were tabulated in Supplementary Table 3.

For AzoGI-T, we used the AzoGI crystal structure (PDB: 1U6B) as the initial model, which was rigid-body fitted to the AzoGI-T map using Chimera, and then the atomic model was iteratively improved by real-space refinements in PHENIX and manual adjustments in COOT. Because the final cryo-EM map for FMNrbsw-T is the trimer instead of SE monomer, the designed NanoTiler model was used as the initial model to fit the map by MDFF[84], and then the atomic model was iteratively improved by real-space refinements in PHENIX and manual adjustments in COOT. All structure figures were prepared with UCSF Chimera or PyMOL (DeLano Scientific).

Ribozyme activity assays. To measure the *trans*-acting endonucleolytic activity of the TetGI constructs (TetGI-M, TetGI-D and TetGI-T), reactions were performed with 2 μM 56-FAM-labeled 10-nt RNA substrate, 0.5 μM ribozyme monomeric units and 1 mM GTP cofactor in 1 × ribozyme reaction buffer (10 mM sodium cacodylate at pH 6.8, 30 mM $MgCl_2$) at 37 °C. For the AzoGI constructs, the activity assayed was based on the second step of splicing because the 3′ exon of the G4-a + 7 AzoGI constructs (AzoGI*-M and AzoGI*-T, see Extended Data Fig. 7) was not completely cleaved during IVT. Reactions were performed with 2 μM 56-FAM-labeled 5-nt RNA substrate (mimicking 5′ exon) and 1 μM ribozyme monomeric units in 1 × ribozyme reaction buffer at 25 °C. Aliquots of the reaction mixtures were removed at various times and quenched by adding an equal volume of 2 × TBE–urea sample buffer (Bio-Rad) supplemented with extra 50 mM EDTA. Reactions were analyzed by denaturing PAGE, and the gels were visualized using a Sapphire Biomolecular Imager (Azure Biosystems).

Reverse transcription and polyadenylation assay. Reverse transcription (RT) was used to study the 5′ sequence of the TetGI constructs. The reactions were

performed with 0.8 μM RNA, 1.2 μM 56-FAM labeled RT primer (complementary to nucleotides 29 to 46 of the TetGI), 0.5 mM of each dNTP and 1 U/μL AMV reverse transcriptase (NEB) in 1 × AMV Reverse Transcriptase Reaction Buffer (supplied by manufacturer with the enzyme) for 1 hour at 42 °C. The TetGI 5′-mimic RNA (Extended Data Fig. 2a) was prepared by IVT and analyzed by dideoxynucleotide sequencing[85] to provide length markers for determining the 5′ cleavage site. For these sequencing reactions, 1.25 mM of a single ddNTP (dideoxynucleoside triphosphate; GE Healthcare) was added to each RT reaction for the 5′-mimic RNA. A polyadenylation assay[86] was used to determine whether the RNA has a 3′-hydroxyl group or a 2′,3′-cyclic phosphate group, because the former can be polyadenylated while the latter cannot. Polyadenylation reactions were performed with 200 nM RNA, 0.5 mM ATP and 24 U/μL yeast poly(A) polymerase (Thermo Scientific) in 1 × Poly(A) Polymerase Reaction Buffer (supplied by manufacturer) for 20 min at 37 °C.

Ligand-binding assays for the FMN riboswitch assemblies. The binding affinities of the FMN riboswitch assemblies to FMN ligand were measured on the basis of the fluorescence quenching of FMN upon specific binding to the riboswitch[61]. Varying concentrations (ranging from 0.1 nM to 100 nM of monomeric units) of the FMN riboswitch assemblies (monomer control, FMNrbsw-T dimer and trimer) were mixed with 60 nM FMN in 50 mM Tris-HCl (pH 7.4), 100 mM KCl and 2 mM $MgCl_2$ for at least 30 min at room temperature for equilibrium before data collection. The fluorescence intensity was measured at 530 nm emission with 450 nm excitation at room temperature using a Synergy H1 Hybrid multi-mode microplate reader (BioTek). Each data point was measured from three independent experiments. Data were fitted using a two-parameter (K_d and f_c) quadratic equation (4) derived in Extended Data Fig. 9d, implying 1:1 stoichiometry of ligand to monomeric unit. Data plotting and curve fitting were performed using OriginPro 2018 software.

Statistics and reproducibility. The following experiments were repeated three times to ensure reproducibility: ribozyme activity assays for the TetGI constructs (Fig. 4b) and AzoGI constructs (Extended Data Fig. 7d). The following experiments were repeated twice to ensure reproducibility: assembly assays for the TetGI constructs (Fig. 4b, Extended Data Fig. 2j and Supplementary Fig. 1b), the AzoGI constructs (Extended Data Fig. 7c and Supplementary Fig. 5b) and the FMN riboswitch constructs (Extended Data Fig. 9a–c); assays for determining the 5′ or 3′ ends of the TetGI constructs (Extended Data Fig. 2b,f,i) and the AzoGI constructs (Extended Data Fig. 7b). For cryo-EM studies of TetGI-DS, TetGI-D, TetGI-T, AzoGI-T and FMNrbsw-T, 4,170, 4,068, 4,284, 4,884 and 3,465 raw micrographs were respectively collected, and the representative raw micrographs are shown respectively in Extended Data Figs. 3a, 4a, 5a, 8a and 10a.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data supporting the findings of this study are principally within the figures and the associated Supplementary Information. Atomic coordinates and cryo-EM maps have been deposited with the Protein Data Bank and the Electron Microscopy Data Bank under the accession codes: 7R6L and EMD-24281 for TetGI-DS, 7R6M and EMD-24282 for TetGI-D, 7R6N and EMD-24283 for TetGI-T, EMD-24284 for AzoGI-T and EMD-24285 for FMNrbsw-T. In this study, the following structures from the PDB were utilized: 2BJ2, 1X8W, 1U6B and 3F2Q.

References

70. Tomizawa, J.-i Control of ColE1 plasmid replication: the process of binding of RNA I to the primer transcript. *Cell* **38**, 861–870 (1984).
71. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
72. Schorb, M., Haberbosch, I., Hagen, W. J. H., Schwab, Y. & Mastronarde, D. N. Software tools for automated transmission electron microscopy. *Nat. Methods* **16**, 471–477 (2019).
73. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
74. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
75. Shaikh, T. R. et al. SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nat. Protoc.* **3**, 1941–1974 (2008).
76. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
77. Scheres, S. H. Processing of structurally heterogeneous cryo-EM data in RELION. *Methods Enzymol.* **579**, 125–157 (2016).
78. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
79. Terwilliger, T. C., Ludtke, S. J., Read, R. J., Adams, P. D. & Afonine, P. V. Improvement of cryo-EM maps by density modification. *Nat. Methods* **17**, 923–927 (2020).
80. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D. Biol. Crystallogr.* **66**, 213–221 (2010).
81. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
82. Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D. Struct. Biol.* **74**, 531–544 (2018).
83. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D. Biol. Crystallogr.* **60**, 2126–2132 (2004).
84. McGreevy, R., Teo, I., Singharoy, A. & Schulten, K. Advances in the molecular dynamics flexible fitting method for cryo-EM modeling. *Methods* **100**, 50–60 (2016).
85. Hahn, C. S., Strauss, E. G. & Strauss, J. H. Dideoxy sequencing of RNA using reverse transcriptase. *Methods Enzymol.* **180**, 121–130 (1989).
86. Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E. & Bartel, D. P. RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* **292**, 1319–1325 (2001).
87. Zaug, A. J., Kent, J. R. & Cech, T. R. A labile phosphodiester bond at the ligation junction in a circular intervening sequence RNA. *Science* **224**, 574–578 (1984).
88. Ennifar, E., Walter, P., Ehresmann, B., Ehresmann, C. & Dumas, P. Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nat. Struct. Biol.* **8**, 1064–1068 (2001).
89. Lebars, I. et al. Exploring TAR-RNA aptamer loop-loop interaction by X-ray crystallography, UV spectroscopy and surface plasmon resonance. *Nucleic Acids Res.* **36**, 7146–7156 (2008).
90. Lilley, D. M. Structures of helical junctions in nucleic acids. *Q. Rev. Biophys.* **33**, 109–159 (2000).
91. Szewczak, A. A. et al. An important base triple anchors the substrate helix recognition surface within the *Tetrahymena* ribozyme active site. *Proc. Natl Acad. Sci. USA* **96**, 11183–11188 (1999).
92. Suh, S. O., Jones, K. G. & Blackwell, M. A group I intron in the nuclear small subunit rRNA gene of *Cryptendoxyla hypophloia*, an ascomycetous fungus: evidence for a new major class of group I introns. *J. Mol. Evol.* **48**, 493–500 (1999).
93. Leontis, N. B. & Westhof, E. A common motif organizes the structure of multi-helix loops in 16S and 23S ribosomal RNAs. *J. Mol. Biol.* **283**, 571–583 (1998).

Acknowledgements

We thank E. Westhof for providing the computer model of the complete TetGI and information on the peripheral domains of group I introns, T. Cech and F. Guo for information on the X-ray model of the TetGI core, M. Dai and Y. Shao for helpful discussions, S. Stoilova-McPhie for the help in pilot EM experiments, and M. Ziegler for proofreading. D.L. is a Merck Fellow of the Life Sciences Research Foundation. This work was supported by NSF grants (CMMI-1333215, CMMI-1344915 and CBET-1729397), AFOSR grant (MURI FATE, FA9550-15-1-0514), NIH grant (5DP1GM133052) and Molecular Robotics Initiative fund from the Wyss Institute to P.Y., a NIH grant (R01GM122797) to M.L. and a NIH grant (R01GM102489) to J.A.P.

Author contributions

D.L. conceived and designed the study, designed and prepared the constructs, performed the functional assays, conducted pilot EM experiments, built and refined the atomic models and drafted the manuscript. F.A.T. collected the EM data, performed the reconstruction and generated the maps, assisted in building atomic models and drafted the manuscript. J.A.P. advised the experiments and drafted the manuscript. M.L. conceived, designed and supervised the study, performed the reconstruction and generated the maps. P.Y. conceived, designed and supervised the study and drafted the manuscript. All authors analyzed the data and commented on the manuscript.

Competing interests

A provisional patent related to this work has been filed with D.L., F.A.T., M.L. and P.Y. listed as coinventors. J.A.P. declares no competing interests.

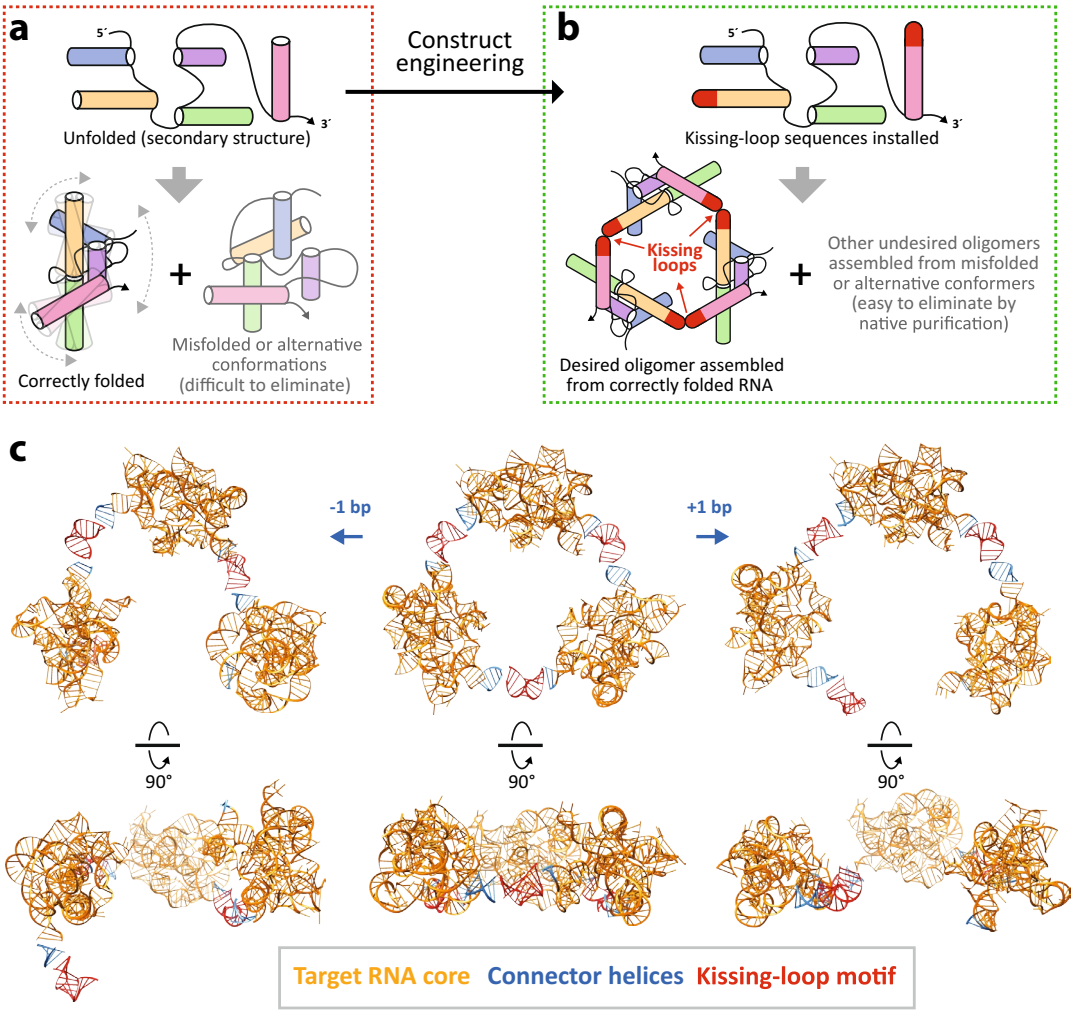
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41592-022-01455-w.

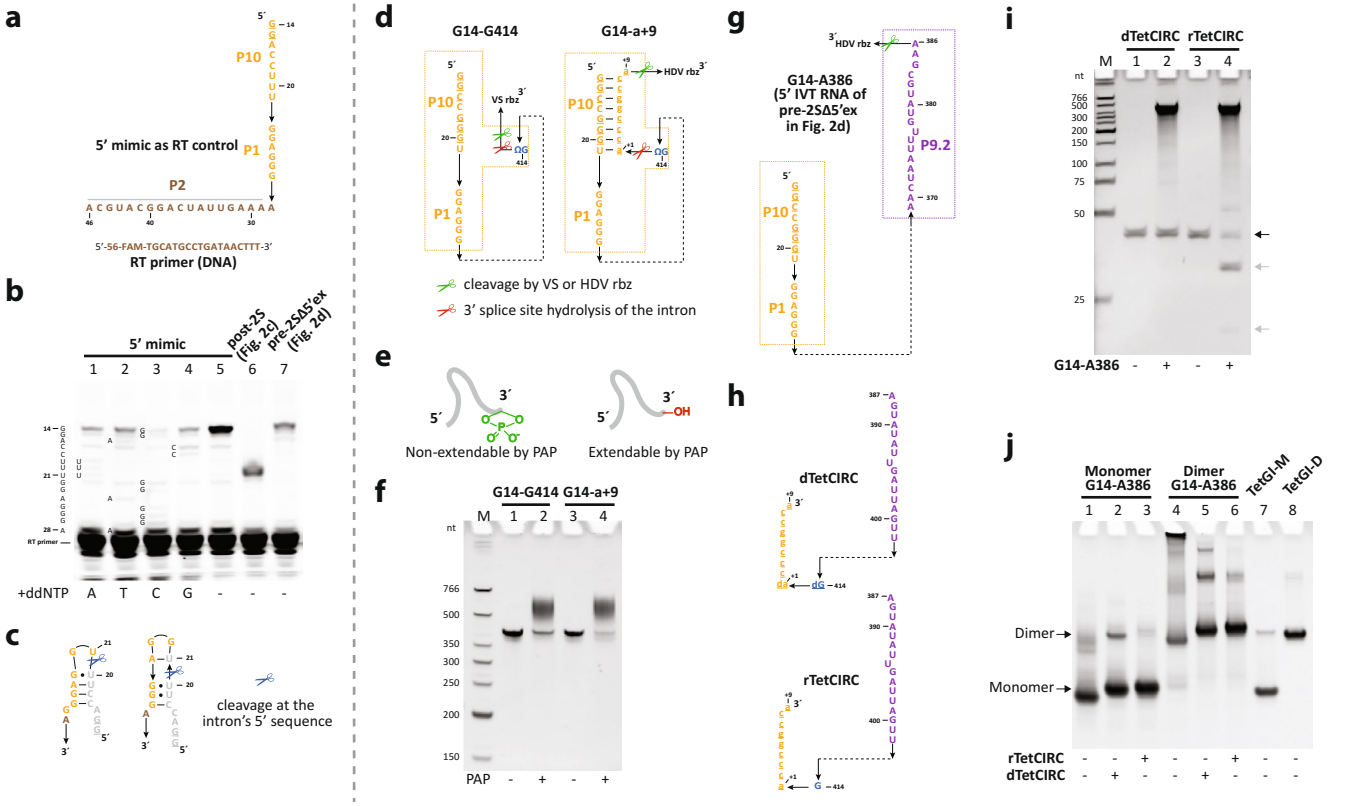
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41592-022-01455-w.

Correspondence and requests for materials should be addressed to Maofu Liao or Peng Yin.

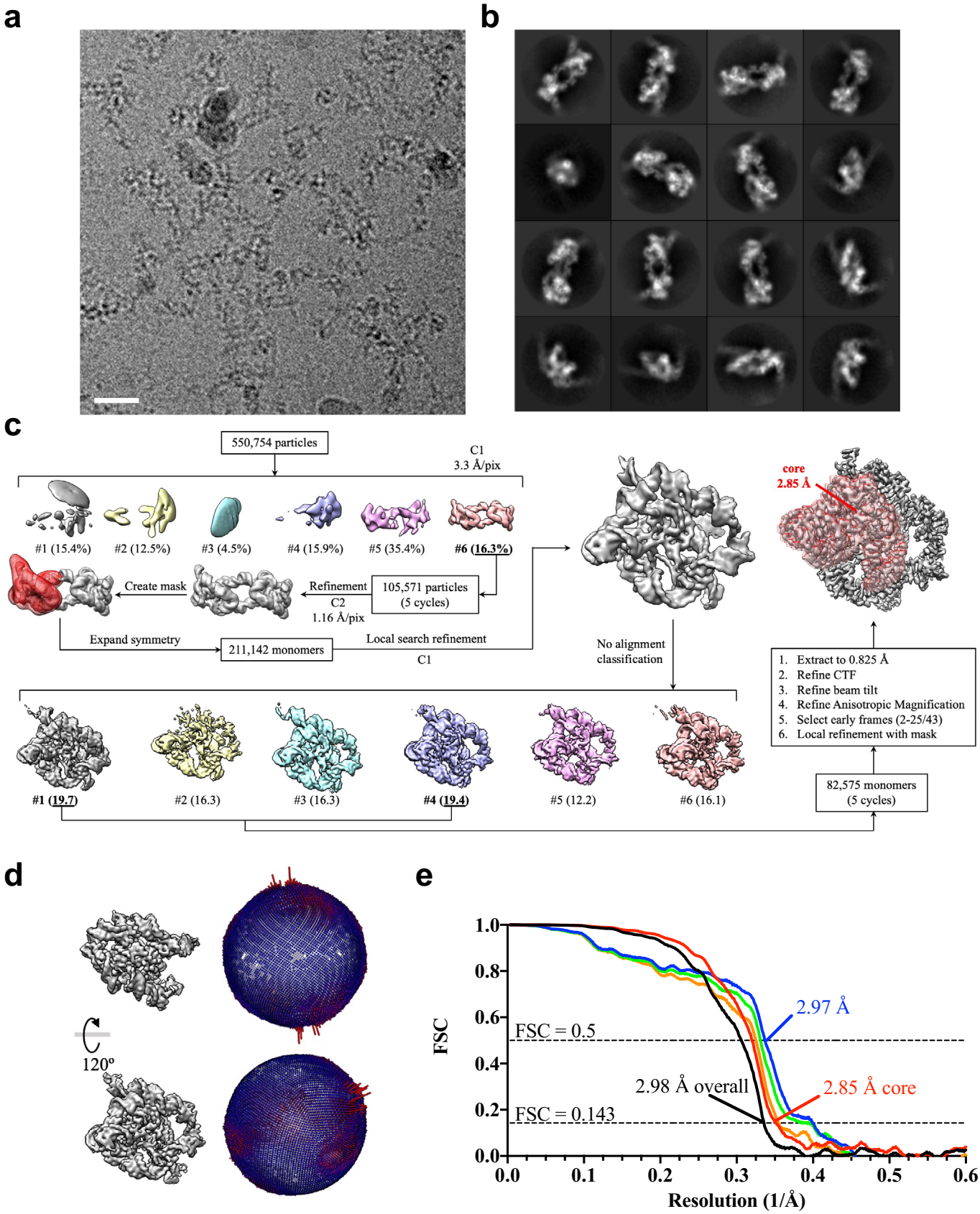
Peer review information *Nature Methods* thanks the anonymous reviewers for their contribution to the peer review of this work. Rita Strack was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.



Extended Data Fig. 1 | The benefits of RNA oligomerization-enabled cryo-EM via installing kissing-loops (ROCK) and the process of screening the lengths of connector helices for ring closure. a, The folding wild-type monomeric RNA leads to correctly folded RNA and species of misfolded or alternative conformations: the former is a difficult subject for cryo-EM structural determination due to small size and structural flexibility and the latter are difficult to eliminate. **b**, The engineered RNA construct with kissing-loop sequences installed folds and assembles into the desired oligomer from the correctly folded RNA that has a larger size and mitigated structural flexibility, while the species of misfolded or alternative conformations assembled into other assemblies of undesired oligomeric states that can be readily eliminated by native purification methods such as native PAGE. Therefore, besides offering RNA constructs more amenable to high-resolution cryo-EM reconstruction, the self-assembled system also facilitates the experimental procedures of RNA folding optimization and native RNA purification. This helps eliminate the misfolding and conformational heterogeneity that are well known to complicate the functional and structural studies of RNA. **c**, The lengths of the connector helices between the RNA core and kissing-loop motif need to be optimized for ring closure. The computer model with optimized lengths of connector helices is shown in the middle, and the removal (left) or addition (right) of one base-pair (bp) would lead to the formation of open spiral structures. The screening process can be performed by NanoTiler as described in Methods.

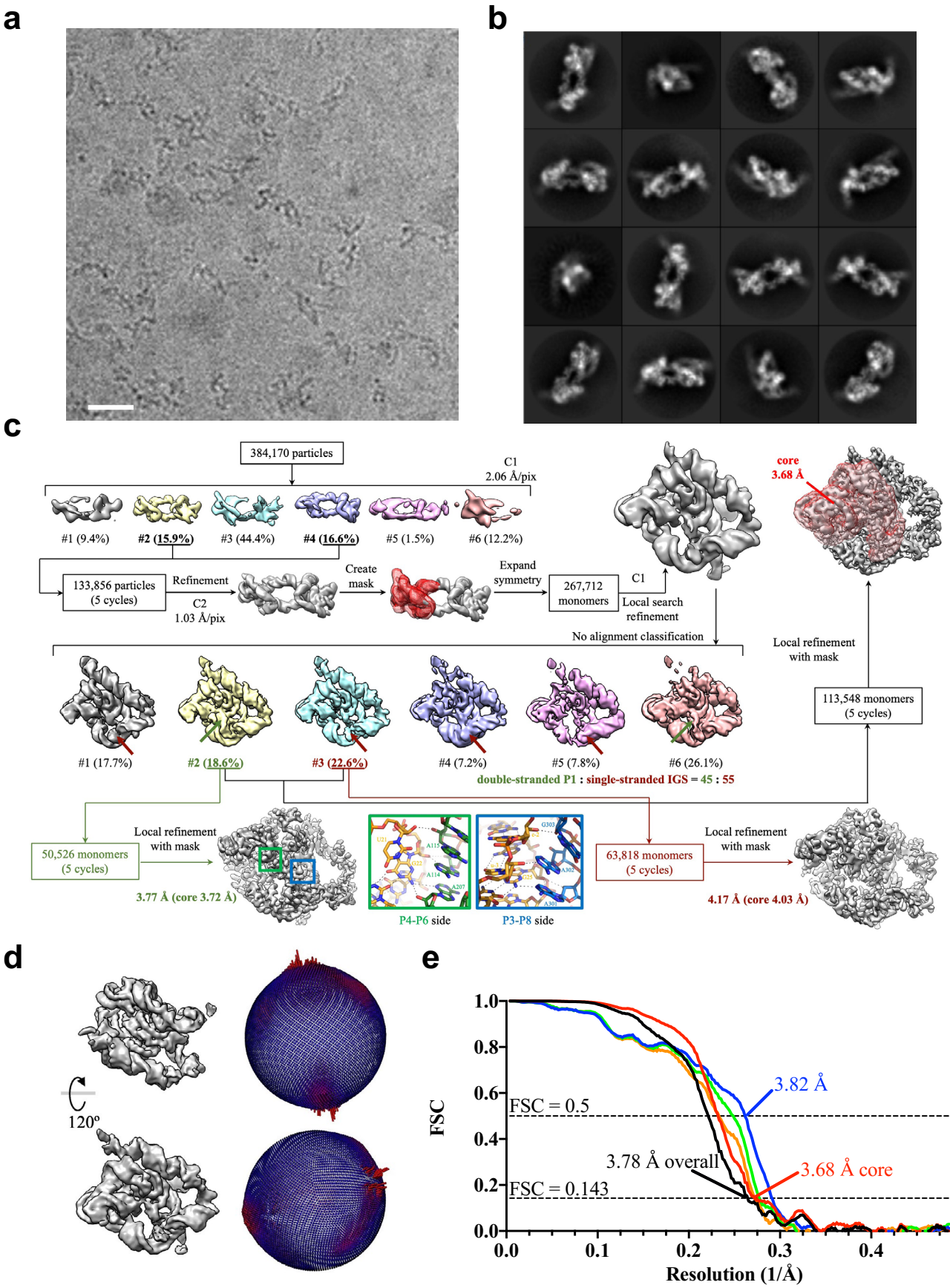


Extended Data Fig. 2 | Designing the 5′ and 3′ fragments for the pre-2SΔ5′ex TetGI complex. a, Sequences of the RNA mimicking the 5′ fragment of the TetGI (5′ mimic) as the reverse transcription (RT) control and the 56-FAM-labelled RT primer. Underlined nucleotides in 5′ mimic were mutated to facilitate the synthesis by IVT; the nucleotides marked by grey line above are the primer-binding region. **b**, Analysis of the 5′ sequences of the TetGI by RT. Lanes 1 to 5 are the dideoxynucleotide sequencing results of 5′ mimic with a single dideoxynucleoside triphosphate (ddNTP; indicated under the gel) added (lanes 1 to 4) or no ddNTP added (lane 5); lanes 6 and 7 are the results of the constructs with wild-type 5′ sequence (same as the post-2S complex in Fig. 2c) or mutated 5′ sequence (same as the pre-2SΔ5′ex complex in Fig. 2d). While the RNA used in the post-2S complex is cleaved after U20, the RNA in the pre-2SΔ5′ex complex is not cleaved at this site. **c**, The likely formation of hairpin structures (two possibilities shown) of the wild-type 5′ sequence accounting for the self-cleavage. We note that the reaction of the wild-type leads to a L-19 product that is cleaved after U19 (ref. [87]) and we attribute the difference of cleavage products to the different 5′ sequences. **d**, Two RNA constructs transcribed with a self-cleaving ribozyme (rbz; VS or HDV) for producing a homogeneous 3′-end. The G14-a+9 RNA was originally designed to have 3′-exon, which was almost completely cleaved in the absence of the 5′-exon. Green and red scissors mark the cleavage sites for the appended ribozyme and the group I intron itself, respectively. **e**, Structures of the 3′-ends produced by the cleavage of the appended ribozyme or the self-cleavage of the intron. While the latter can be extended by yeast poly(A) polymerase (PAP), the former cannot. **f**, PAP extension assay demonstrates that the majority products of both G14-A386 (lanes 1 and 2) and G14-a+9 (lanes 3 and 4) RNA preparation are extended by PAP (lanes 2 and 4), indicating that they are mostly the products from the intron's self-cleavage. **g, h**, Sequences of the 5′ IVT RNA (**g**; G14-A386, from nucleotides 14 to 386) and the 3′ fragments (**h**; dTetCIRC with two deoxynucleotides flanking the 3′ splice site, and rTetCIRC with all-RNA nucleotides) for the pre-2SΔ5′ex TetGI complex. **i**, TetGI-catalyzed hydrolysis at the 3′ splice site is inhibited by the modification of DNA bases flanking the scissile phosphate of the 3′ splice site. Black arrow points to the band of the 3′ fragments of (dTetCIRC or rTetCIRC); grey arrows point to the bands of the cleaved 3′ fragments. **j**, Assembly assays for the monomeric (lanes 1 to 3) and dimeric (lanes 4 to 6) constructs of 5′ IVT G14-A386 RNA without 3′ fragment (lanes 1 and 4), with dTetCIRC (lanes 2 and 5) or with rTetCIRC (lanes 3 and 6). The annealing buffer contains 3 mM Mg^{2+}, which is chosen from the assembly assay of the post-2S constructs (TetGI-M, -D, and -T; see Fig. 3a). Lanes 7 and 8 are control assemblies of TetGI-M and -D. Without 3′ fragment (lanes 1 and 4), there are trailing smears for the 5′ IVT RNA, indicating incorrect folding of the intron if P9.2, P9a and P9.0 are not formed. In the presence of dTetCIRC (lanes 2 and 5) or rTetCIRC (lanes 3 and 7), especially for dTetCIRC, some slower-migrating bands emerged, probably due to domain swapping, that is a 3′ fragment simultaneously binds to a 5′ IVT RNA molecule to form P9.2 and another 5′ IVT RNA molecule to form P10.



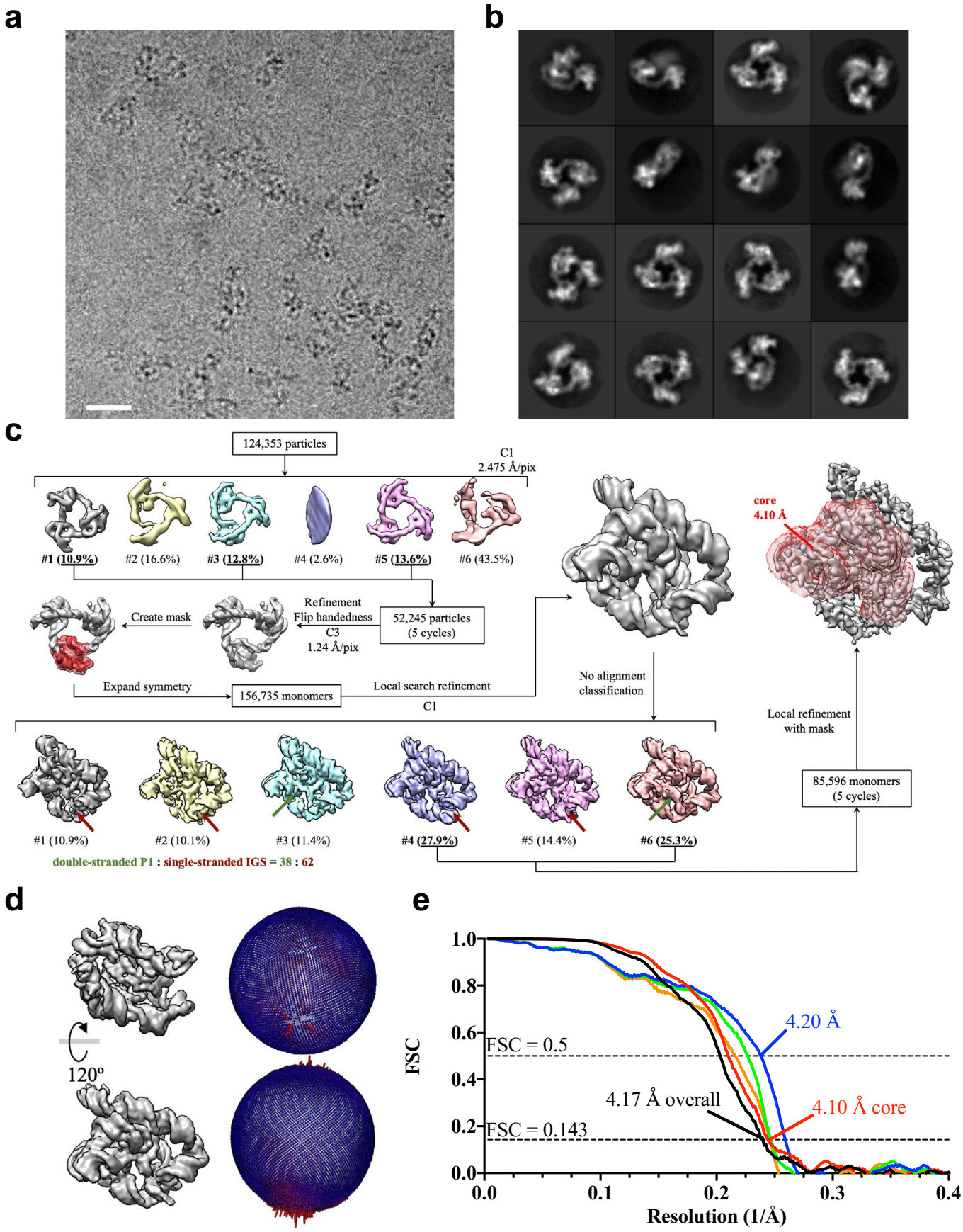
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Cryo-EM imaging, processing and validation for TetGI-DS. a, Representative cryo-EM image of TetGI-DS. The scale bar represents 20 nm. **b**, 2D class averages of TetGI-DS. Box size is 264 Å. **c**, Processing flowchart for the TetGI-DS dataset. **d**, Angle distribution for the particles included in the final 3D reconstruction. **e**, Fourier Shell Correlation (FSC) curves of the final TetGI-DS reconstruction. Half map #1 vs. half map #2 for the entire monomer is shown in black. The remaining FSC curves were calculated for the core domains only: half map #1 vs. half map #2 (red), model vs. refined map (blue), model refined in half map #1 vs. half map #1 (green), and model refined in half map #1 vs. half map #2 (orange).



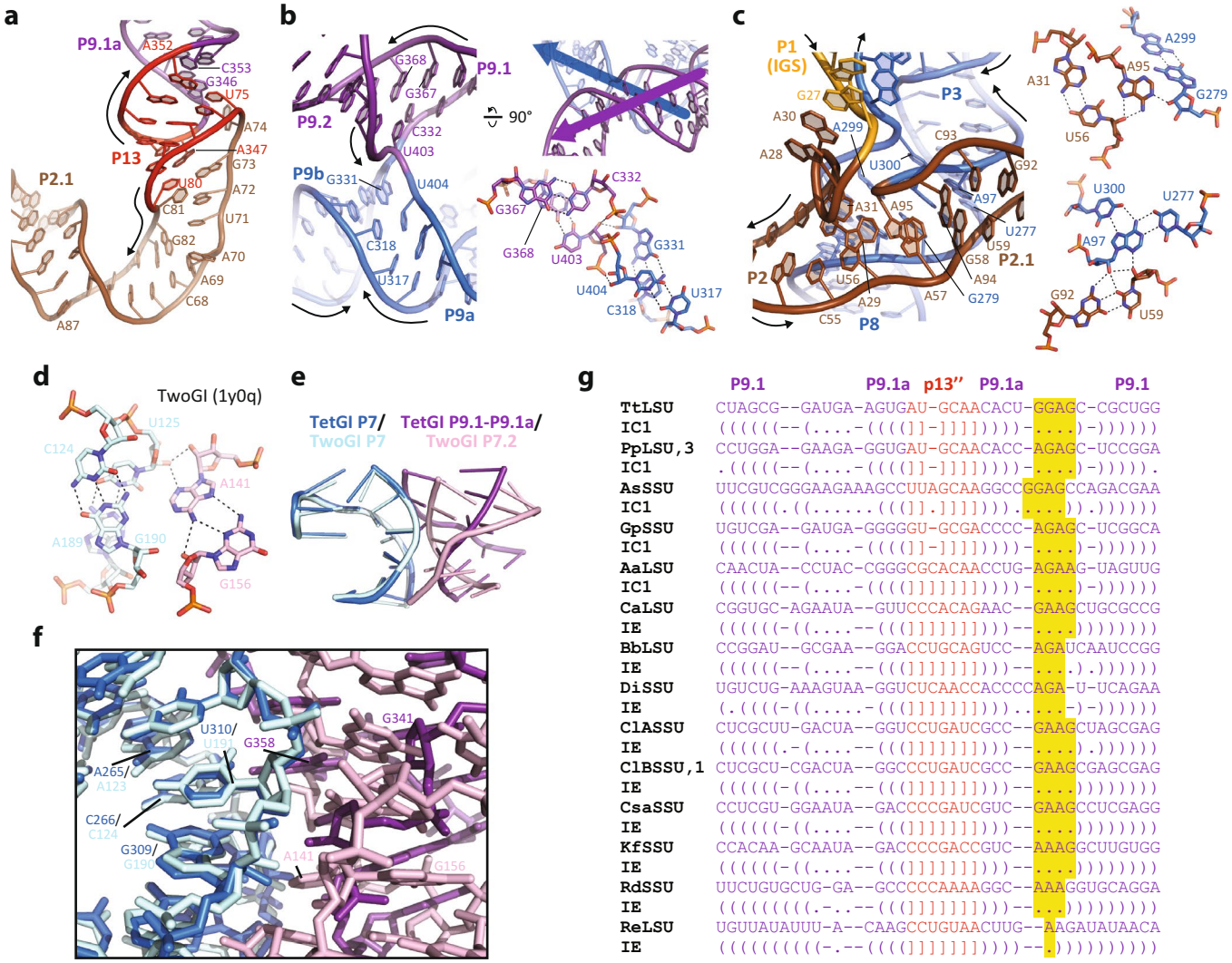
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Cryo-EM imaging, processing and validation for TetGI-D. a, Representative cryo-EM image of TetGI-D. The scale bar represents 20 nm. **b**, 2D class averages of TetGI-D. Box size is 264 Å. **c**, Processing flowchart for the TetGI-D dataset. 3D classification of the symmetry-expanded monomers results in classes according to the conformations with double-stranded P1 (green arrows) or with single-stranded IGS (red arrows). The ratio of the two conformations is calculated and shown. For the conformation with double-stranded P1, green and blue boxes show the details of tertiary contacts of P1 on the P4-P6 side and P3-P8 side, respectively. The contacts agree well with the previous biochemical studies[48–50] and the crystal structure of AzoGI[40]. The final cryo-EM map for TetGI-D was refined from two classes, and exhibits a stronger map intensity of double-stranded P1 than single-stranded IGS; therefore, the atomic model was built with double-stranded P1. **d**, Angle distribution for the particles included in the final 3D reconstruction. **e**, Fourier Shell Correlation (FSC) curves of the final TetGI-D reconstruction. Half map #1 vs. half map #2 for the entire monomer is shown in black. The remaining FSC curves were calculated for the core domains only: half map #1 vs. half map #2 (red), model vs. refined map (blue), model refined in half map #1 vs. half map #1 (green), and model refined in half map #1 vs. half map #2 (orange).

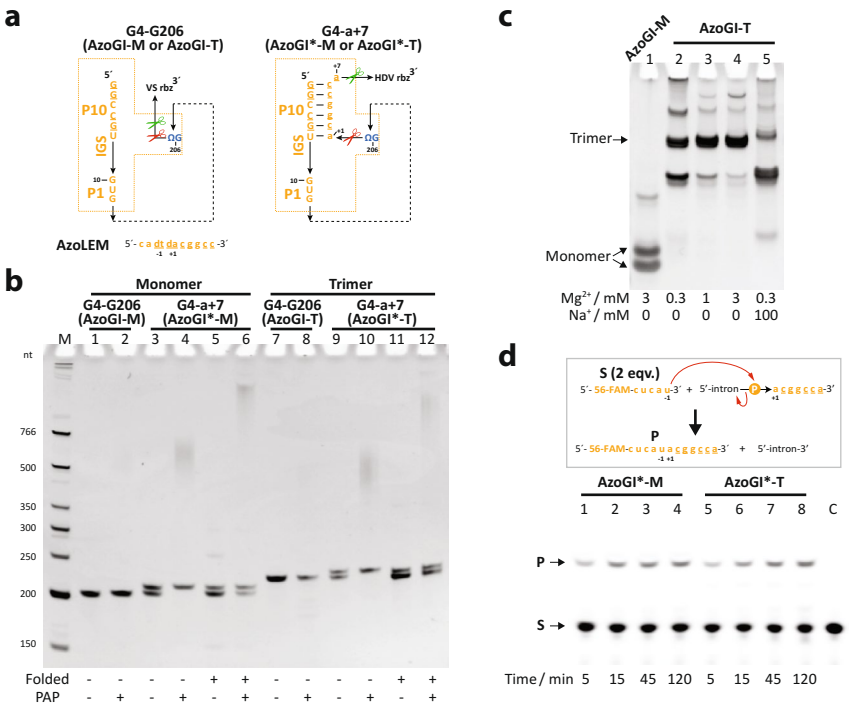


Extended Data Fig. 5 | See next page for caption.

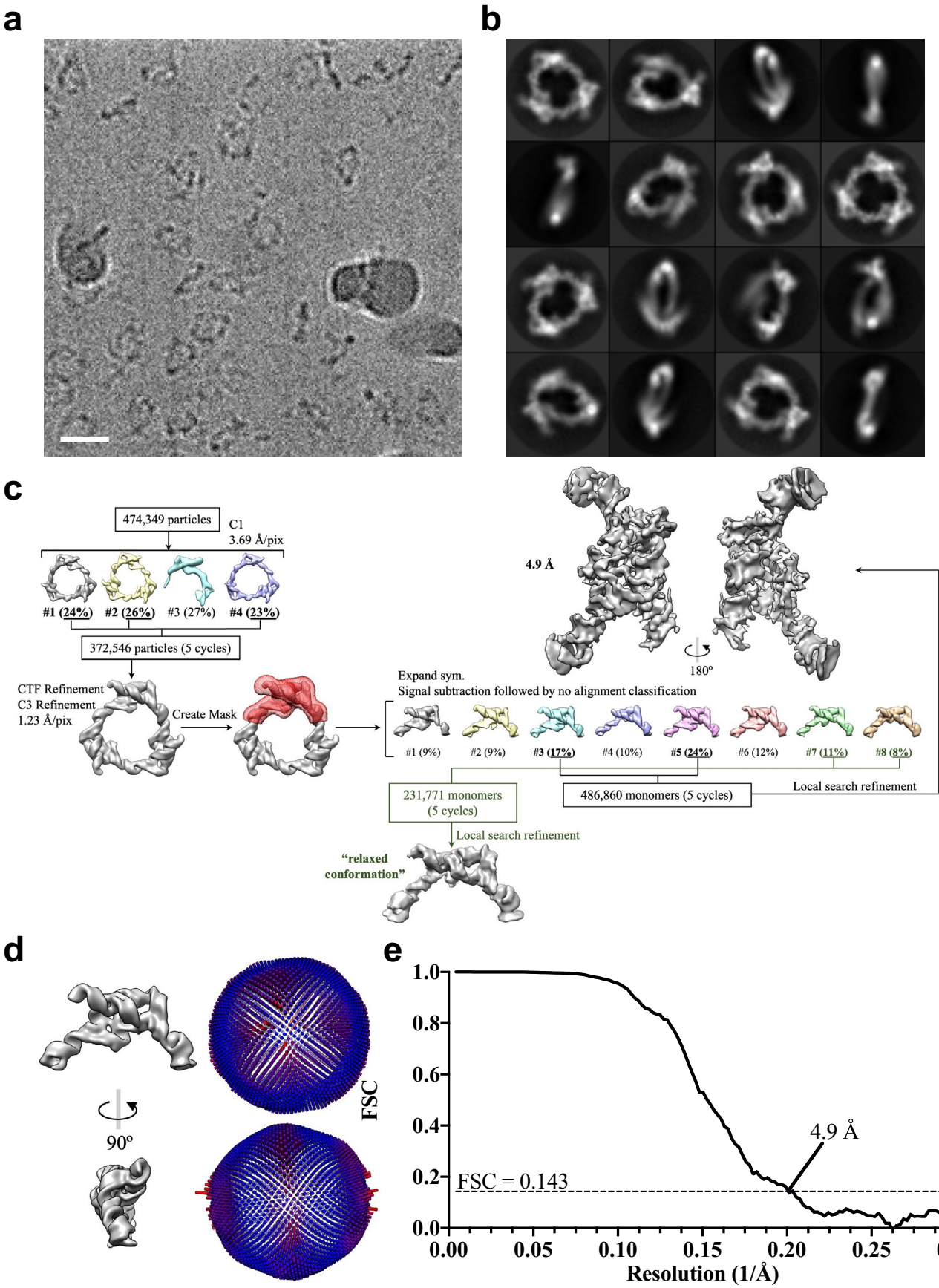
Extended Data Fig. 5 | Cryo-EM imaging, processing and validation for TetGI-T. a, Representative cryo-EM image of TetGI-T. The scale bar represents 20 nm. **b**, 2D class averages of TetGI-T. Box size is 317 Å. **c**, Processing flowchart for the TetGI-T dataset. Similar to the case of TetGI-D, 3D classification of the symmetry-expanded monomers of TetGI-T also results in classes according to the conformations with double-stranded P1 (green arrows) or with single-stranded IGS (red arrows). The ratio of the two conformations is calculated and shown, which is close to the dimeric construct TetGI-D shown in Extended Data Fig. 4c. The final cryo-EM map for TetGI-T was refined from two classes, and exhibits a stronger map of single-stranded IGS than double-stranded P1, probably due to the stabilization of the peripheral domains by the trimer construct; therefore, the atomic model was built with the single-stranded IGS. **d**, Angle distribution for the particles included in the final 3D reconstruction. **e**, Fourier Shell Correlation (FSC) curves of the final TetGI-T reconstruction. Half map #1 vs. half map #2 for the entire monomer is shown in black. The remaining FSC curves were calculated for the core domains only: half map #1 vs. half map #2 (red), model vs. refined map (blue), model refined in half map #1 vs. half map #1 (green), and model refined in half map #1 vs. half map #2 (orange).



Extended Data Fig. 6 | Newly visualized structural elements involving the peripheral domains of the TetGI. a, Structural details of the tertiary interaction P13. P13 is a 6-bp duplex formed by the base-pairings of U75 through U80 (p13′) and A352 through A347 (p13′), and stacks coaxially between the G73:C81 pair of P2.1 and the G346:C353 pair of P9.1a, bearing a conformational resemblance to some other 6-bp kissing-loop complexes[88,89]. A 4-nt bulge consisting of A69 through A72 is present near the tip of P2.1, allowing for the bent shape at the junction of P2.1 and P13. **b**, Structure of the four-way junction (4WJ) at P9a-P9b and P9.1-P9.2. Top-right inset shows the strand directions of the continuous strands of the 4WJ, indicating the flanking helices stacked in a left-handed parallel configuration[90]. Bottom-right inset shows interaction details at the crossover site. Though the overall connectivity and other long-range tertiary interactions may be the major determinant for the configuration of this 4WJ, the sugar-phosphate interactions of the nucleotides from the exchanging strands at the crossover site may contribute to the configuration of this 4WJ. **c**, Structure of the complex multiway junction at P1, P3-P8 and P2-P2.1. Insets show the details of two tertiary contacts centered by A95 (top-right) and A97 (bottom-right), respectively, stabilizing the juxtaposition of the pseudo-continuous helices of P2-P2.1 and P3-P8. are two tertiary interactions between the P2-P2.1 and P3-P9 domains. Within the A97-centered tertiary interaction (right), U300 forms a base-triple with A97:U277 pair, corroborating the previous biochemical evidence[91]. **d**, A similar contact is observed in the crystal structure of the Twort group I intron[51] (TwoGI; PDB code: 1y0q) formed between P7 and the internal loop in P7.2. **e**, **f**, Comparing the tertiary interactions observed in TetGI and TwoGI (overlayed by P7). **g**, Sequence alignment of different class IC1 and class IE group I introns reveals a conserved purine-rich loop (highlighted in yellow) at J9.1/9.1a. The extracts of the alignments of 14 sequences are from Lehnert et al. [35], where all the sequences were regarded as subgroup IC1, but later some of them were categorized as subgroup IE[92]. TtLSU (intron in the large ribosomal RNA precursor of *Tetrahymena thermophila*) is the TetGI studied in this work. We note that the internal loop of J9.1/9.1a of the TetGI, which has the sequence of 5′-AUGA-3′/5′-GGAG-3′, is reminiscent of but not exactly the loop E motif[93], which is normally 5′-AGUA-3′/5′-GAA-3′, though the sequence of the abovementioned internal loop in P7.2 of the TwoGI is the same as loop E motif.

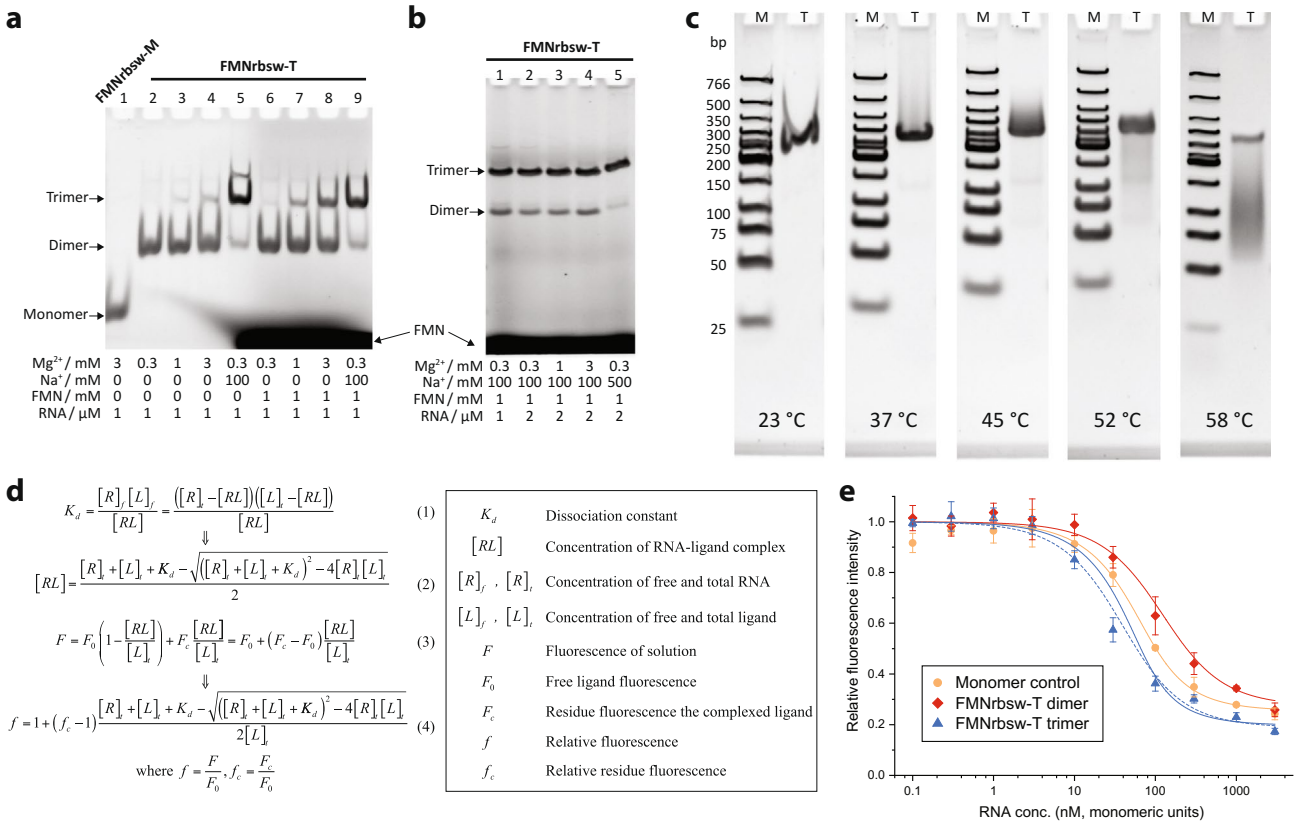


Extended Data Fig. 7 | AzoGI-T sequence design, assembly and activity. a, Two RNA designs, G4-G206 and G4-a+7, transcribed with a self-cleaving ribozyme (rbz) for producing a homogeneous 3′ end. Green and red scissors mark the cleavage site for the appended ribozyme or the intron itself. AzoLEM is the ligated exon mimic for the AzoGI, and its complexing with G4-G206 RNA forms the post-2S state of the intron. **b**, The PAP extension assay of various constructs for the AzoGI. The products from the preparation of G4-G206 RNA constructs cannot be extended by PAP (lanes 2 and 8), indicting a 2′,3′-cyclic phosphate generated by rbz at the 3′ end (this is different from the G14-G414 RNA of TetGI, the majority of which can be extended by PAP; see lane 2 of Extended Data Fig. 2f). For G4-a+7 constructs, the products generated by the cleavage of rbz and intron itself can be directly distinguished based on the different electrophoretic mobilities in the presented analytical gel because the former is longer due to the appendage of a 7-nt 3′ exon (the small size difference is not noticeable in the preparative gel for RNA purification). As shown in lanes 3 and 9, about half of G4-a+7 RNA of either construct is cleaved at the 3′ splice site (whereas, in the case of the TetGI G14-a+9 RNA, more than 90% is cleaved at the 3′ splice site, indicating a substantially higher activity of the TetGI; see lane 4 of Extended Data Fig. 2f). Only the shorter products, which is produced by intron cleavage, can be extended by PAP. After folding, the intron-cleaved shorter products increase to ~70% (lanes 5 and 11), and some portion of these shorter products cannot be extended by PAP, probably due to the lower accessibility of the 3′ end after RNA folding. **c**, Assembly assay of the trimeric construct AzoGI-T (lanes 2 to 5) and the monomer control AzoGI-M (lane 1). Similar to the TetGI constructs, the optimal condition for folding/assembly was determined to be 3 mM Mg^{2+}. Interestingly, monomer control AzoGI-M runs into two major bands, indicating conformational heterogeneity that is likely due to the open and closed conformations of the post-2S construct. **d**, Activity assay to test the activity of the second step of splicing. The reactions were conducted with 1 μM monomer units of either intron construct and 2 μM substrate (S, reacting as the 5′ exon) at 25 °C. The assay takes the advantage of the fact that there is still about 30% of AzoGI*-M and AzoGI*-T constructs containing 5′ exon after IVT preparation, purification and folding due to presumably lower splicing activity of the AzoGI.

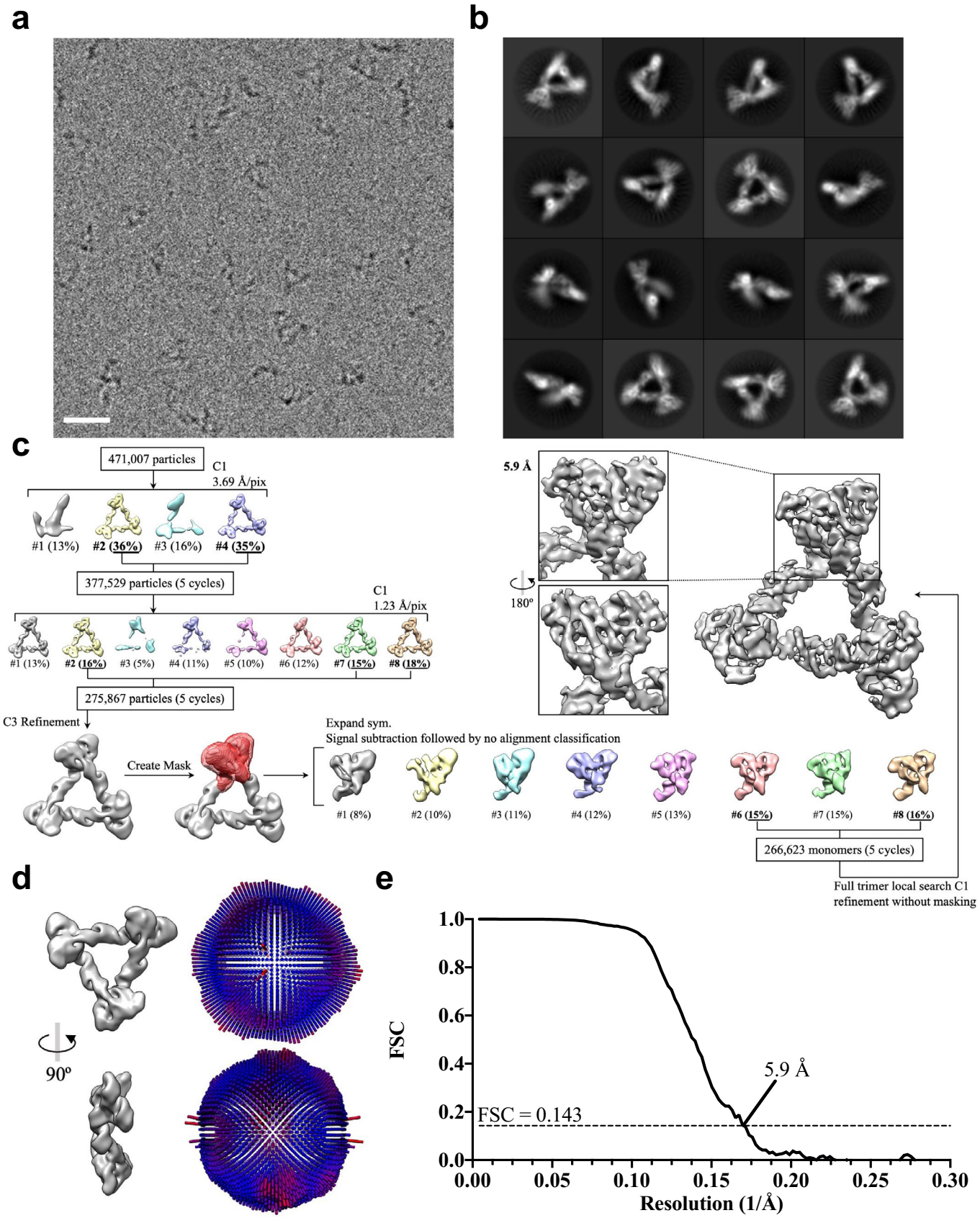


Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Cryo-EM imaging, processing and validation for AzoGI-T. a, Representative cryo-EM image of AzoGI-T. The scale bar represents 20 nm. **b**, 2D class averages of AzoGI-T. Box size is 236 Å. **c**, Processing flowchart for the AzoGI-T dataset. **d**, Angle distribution for the particles included in the final 3D reconstruction. **e**, Gold-standard Fourier Shell Correlation (FSC) for the final AzoGI-T reconstruction.



Extended Data Fig. 9 | FMNrbsw-T assembly and activity. a, **b**, Assembly assay. FMNrbsw-M is the monomer control. Three factors to increase the trimer yield: the presence of FMN ligand; high Na^+; and increasing RNA concentration. **c**, Native PAGE (6%, in the presence of 2 mM free Mg^{2+}) analysis of preassembled FMNrbsw-T trimer (lane T) under different temperatures ranging from 23 °C to 58 °C. Disassembly of the trimer starts to occur at 45 °C. The high thermal stability of the kissing-loop motif may contribute to the formation of the kinetic assembly of dimer because the kissing-loop formation may precede the complete folding of the riboswitch during annealing as suggested by its weaker ligand binding as shown in **e**. Lane M contains the low-molecular weight DNA ladder (NEB). **d**, Derivation of the equation to analyze the ligand binding (1:1 stoichiometry) based on fluorescence quenching. Data were fitted using a two-parameter (K_d and f_c) quadratic Equation (4). **e**, Fluorescent binding assay of FMN (60 nM) with the riboswitch constructs conducted in 100 mM KCl and 2 mM $MgCl_2$. Each data point is represented as mean ± s.d from three independent measurements. FMNrbsw-T dimer is the alternate dimeric assembly of the FMNrbsw-T RNA. The calculated K_d (nM, mean ± s.d.), f_c (mean ± s.d.), and R^2 are: monomer control, 30.1 ± 2.2, 0.254 ± 0.006, 0.9961; FMNrbsw-T dimer, 91.6 ± 21.9, 0.273 ± 0.017, 0.9944; FMNrbsw-T trimer, 17.4 ± 7.6, 0.197 ± 0.020, 0.9881. The dashed blue line for is fitted FMNrbsw-T trimer taking $[L]_t$ as 20 nM, which assumes that each monomeric subunit in the trimer can locally sense only one third of the ligand, and calculated values of K_d (nM, mean ± s.d.), f_c (mean ± s.d.), and R^2 are: 28.8 ± 6.3, 0.188 ± 0.015, 0.9937. This adjustment results in an improved fitting as reflected by an improved R^2.



Extended Data Fig. 10 | Cryo-EM imaging, processing and validation for FMNrbsw-T. a, Representative cryo-EM image of FMNrbsw-T. The scale bar represents 20 nm. **b**, 2D class averages of FMNrbsw-T. Box size is 236 Å. **c**, Processing flowchart for the FMNrbsw-T dataset **d**, Angle distribution for the particles included in the final 3D reconstruction. **e**, Gold-standard Fourier Shell Correlation (FSC) for the final FMNrbsw-T reconstruction.

nature research

Corresponding author(s): Maofu Liao, Peng Yin

Last updated by author(s): Jan 5, 2022

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	SerialEM (v 3.7)
Data analysis	PHENIX (v1.18.2-3874); Coot (v 0.9.3); Pymol (v 2.4.1); UCSF Chimera (v 1.14); SAMUEL (v 21.01); MotionCor2 (v 1.1.0); CTFFIND4 (v 4.1.5); RELION 3.0; ResMap (v 1.1.4); NanoTiler (v 0.11.1); Mfold (v 3.6); OriginPro 2018 (v 9.5); VMD (v 1.9.3) with MDFF (the version included within VMD) and NAMD (v 2.13).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data supporting the findings of this study are principally within the figures and the associated Supplementary Information. Atomic coordinates and cryo-EM maps have been deposited with the Protein Data Bank and the Electron Microscopy Data Bank under the accession codes: 7R6L and EMD-24281 for TetGI-DS, 7R6M and EMD-24282 for TetGI-D, 7R6N and EMD-24283 for TetGI-T, EMD-24284 for AzoGI-T, and EMD-24285 for FMNrsw-T. In this study, the following structures from the PDB were utilized: 2bj2, 1x8w, 1u6b and 3f2q.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample size. The sample size was determined based on sufficient EM data that can achieve adequate single-particle EM analysis and 3D cryo-EM reconstructions.
Data exclusions	No data were excluded from analyses.
Replication	Biochemical and functional experiments were repeated two or three times in independent experiments and findings were reliably reproduced. For cryo-EM data, data were collected by acquiring sufficiently large number of micrographs as specified in Methods from at least four independently prepared cryo-EM grids. The cryo-EM data processing would be reproducible following the procedures in Methods.
Randomization	This is not relevant to our study, because no grouping was needed.
Blinding	Investigators were not blinded to group allocation, because no grouping was needed for this study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging