

Naturally ornate RNA-only complexes revealed by cryo-EM

<https://doi.org/10.1038/s41586-025-09073-0>

Received: 8 December 2024

Accepted: 25 April 2025

Published online: 6 May 2025

Open access

 Check for updates

Rachael C. Kretsch¹, Yuan Wu², Svetlana A. Shabalina³, Hyunbin Lee⁴, Grace Nye⁵, Eugene V. Koonin³, Alex Gao^{1,4}, Wah Chiu^{1,5,6,7}✉ & Rhiju Das^{1,2,4}✉

The structures of natural RNAs remain poorly characterized and may hold numerous surprises^{1–4}. Here we report three-dimensional structures of three large ornate bacterial RNAs using cryo-electron microscopy (cryo-EM). GOLLD (Giant, Ornate, Lake- and Lactobacillales-Derived), ROOL (Rumen-Originating, Ornate, Large) and OLE (Ornate Large Extremophilic) RNAs form homo-oligomeric complexes whose stoichiometries are retained at lower concentrations than measured in cells. OLE RNA forms a dimeric complex with long co-axial pipes spanning two monomers. Both GOLLD and ROOL form distinct RNA-only multimeric nanocages with diameters larger than the ribosome, each empty except for a disordered loop. Extensive intramolecular and intermolecular A-minor interactions, kissing loops, an unusual A–A helix and other interactions stabilize the three complexes. Sequence covariation analysis of these large RNAs reveals evolutionary conservation of intermolecular interactions, supporting the biological importance of large, ornate RNA quaternary structures that can assemble without any involvement of proteins.

The importance of non-coding RNAs (ncRNAs) across biology is increasingly recognized, but only a small number have been functionally characterized, with studies revealing sophisticated catalytic and sensory functions in some cases^{5–9}. Bacteria, archaea and their viruses are thought to possess numerous diverse and complex ncRNAs, but most of these have not been thoroughly studied^{1–4}. Furthermore, there is a conspicuous shortage of data on the 3D structures of RNA molecules. Out of more than 4,000 RNA classes in the RNA Families (RFAM) database 15.0, only 143 have experimentally resolved tertiary structures¹⁰. For many of the remaining cases, it appears likely that structural characterization will depend on reconstitution of the RNA with small molecule, protein or nucleic acid partners, which are unknown in most cases.

The Breaker laboratory and collaborators have previously described three classes of bacterial and phage RNAs for which covariance analysis of genomic and metagenomic sequences revealed secondary structures that were so extensive and elaborate that ‘ornate’, ‘giant’ or ‘large’ were included in their names: GOLLD RNA³, ROOL RNA^{1,11} (concomitantly reported in ref. 12) and OLE RNA². The functions of these three classes of large RNAs remain poorly understood.

Here, using cryo-EM, we show that OLE, ROOL and GOLLD all form atomically ordered 3D structures. Unexpectedly, the three structures are stabilized not by proteins but by other copies of the same RNA molecule in ornate quaternary assemblies with many intermolecular bridges, a phenomenon that has not previously been observed for natural RNA molecules¹³.

OLE forms an RNA-only dimer

OLE is a class of large RNAs with an ornate secondary structure that is conserved throughout evolution². OLE is found mainly in extremophilic bacteria, and experimental characterization in *Halalkalibacterium halodurans* has demonstrated its involvement in integrating energy availability, metal ion homeostasis and drug treatment to mediate cellular adaptation, although the underlying molecular mechanisms remain unknown^{2,14–16}. Cellular localization to the membrane, binding to at least six protein partners^{15,17–22} and evidence of alternative secondary structures¹⁷ suggested that OLE was unlikely to form a well-defined RNA-only 3D structure. However, our study showed that the 577-nucleotide (nt) OLE RNA from *Clostridium acetobutylicum*^{2,23} formed distinct, compact particles that were clearly visible in cryo-EM images (Fig. 1a). Furthermore, a 2.9 Å resolution 3D map of a dimeric OLE RNA could be reconstructed with two-fold imposed symmetry (Extended Data Fig. 1). A model of the each chain has been built for 308 nt in the OLE 5′ region, with Q-scores²⁴ exceeding the expected score at this resolution (Fig. 1b, Extended Data Fig. 1e and Supplementary Video 1).

Our OLE dimer map shows that it is organized as a series of parallel A-form helices, resembling a bundle of pipes. The exterior ends of these pipes from each chain are interconnected into a five-way junction, with a secondary structure that agrees with the previously proposed one for the observed domain with stems P3 to P9.3 (refs. 2,15) (Fig. 1c; hereafter, paired stems, hairpin loops and joining linkers are designated ‘P’, ‘L’ and ‘J’, respectively, following conventional RNA nomenclature).

¹Biophysics Program, Stanford University, Stanford, CA, USA. ²Howard Hughes Medical Institute, Stanford University, Stanford, CA, USA. ³Computational Biology Branch, Division of Intramural Research, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA. ⁴Department of Biochemistry, Stanford University School of Medicine, Stanford, CA, USA. ⁵Division of CryoEM and Bioimaging, SSRL–SLAC National Accelerator Laboratory, Menlo Park, CA, USA. ⁶Department of Bioengineering and James Clark Center, Stanford University, Stanford, CA, USA. ⁷Department of Microbiology and Immunology, Stanford University, Stanford, CA, USA. ✉e-mail: wahc@stanford.edu; rhiju@stanford.edu

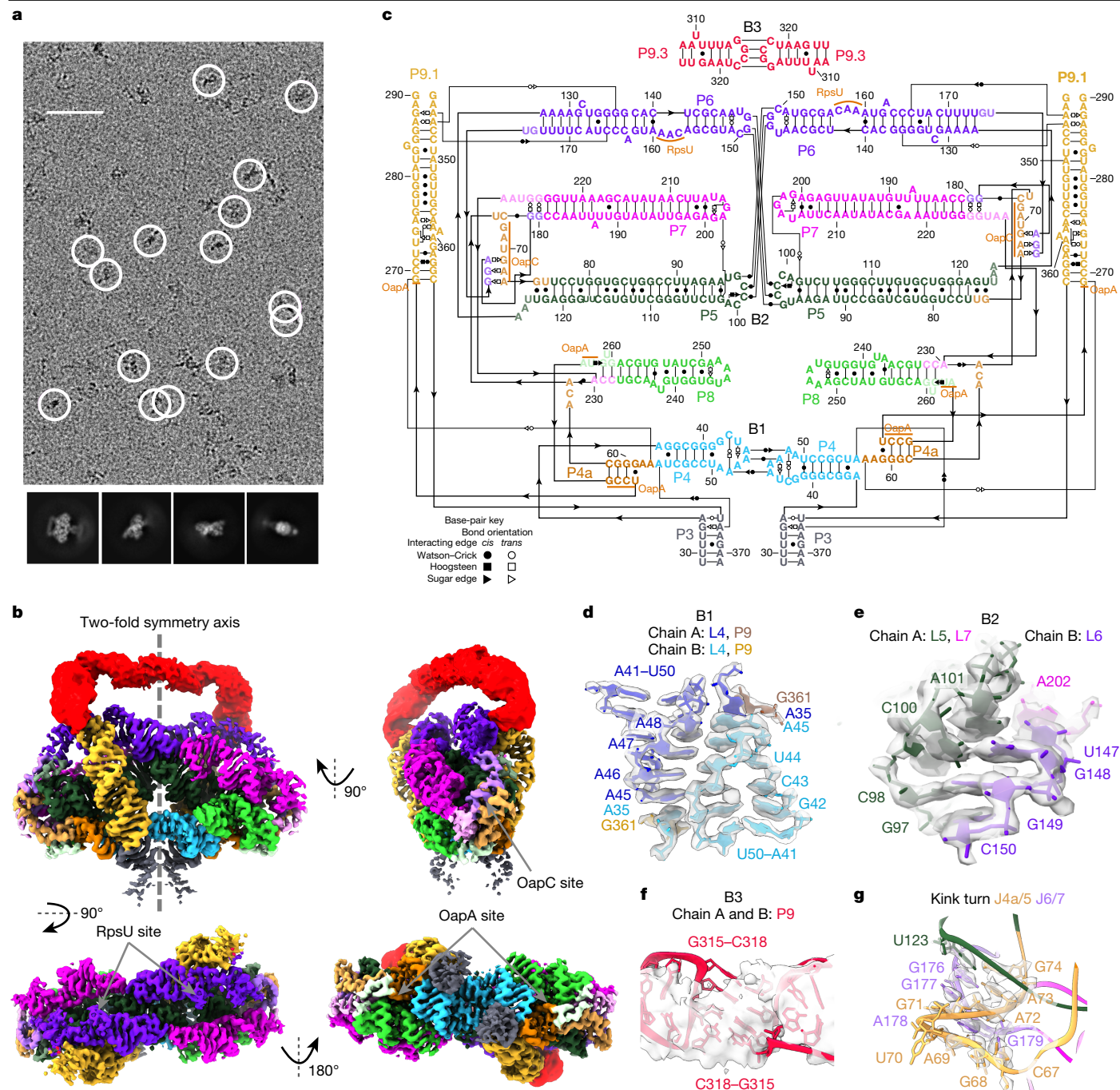


Fig. 1 | Structure of OLE homodimer. **a**, Top, representative micrograph (6,752 micrographs total). Particles selected for reconstruction are circled in white. Bottom, 2D class averages. Scale bar, 50 nm. **b**, The cryo-EM reconstruction of the OLE dimer. The top left image depicts the separation of the two chains: one chain on the right and the other on the left. Each domain of both chains is coloured. To aid visualization, the flexible P9.3 domain (red) is displayed with the unsharpened map at 10 σ contour. The proposed binding

sites of previously described proteins (RpsU, OapC and OapA) are labelled. **c**, Secondary structure of OLE dimer. The domains are coloured as in **b**. **d-f**, The intermolecular bridge interactions B1 (**d**), B2 (**e**) and B3 (**f**), coloured by domain. In **d**, the domain colouring is darker for chain A to differentiate the chains. **g**, The kink-turn motif that may bind the OapC protein, identical for each monomer. The sharpened cryo-EM map is displayed at the following contours: 7 σ (**b**), 15 σ (**d**), 12 σ (**f**) and 10 σ (**g**).

An unusual but highly conserved symmetric interaction comprised of four A-A base pairs between two chains (L4, Fig. 1d), intermolecular base pairing and stacking interactions connecting L5, L6 and L7 (Fig. 1e), and a kissing loop (L9.3, Fig. 1f) ‘weld’ the pipes together in the middle of the complex. We denote these intermolecular interactions ‘bridges’ B1–B3, as used in ribosome nomenclature²⁵. An elaborative list of intramolecular motifs and intermolecular interactions is presented in Supplementary Tables 1 and 2, respectively. Beyond the 5’ region, other

conserved parts of OLE were not resolved in the structure, suggesting flexibility.

Surprisingly, regions of OLE that were previously thought to adopt alternative structures upon protein binding are clearly resolved and solvent-accessible, suggesting that proteins may bind the OLE dimer in the pre-formed RNA conformation that we observed here. Our cryo-EM data show that these proteins are not required for the folding of the 5’ domain of OLE, and the RNA structure itself may have a crucial role in

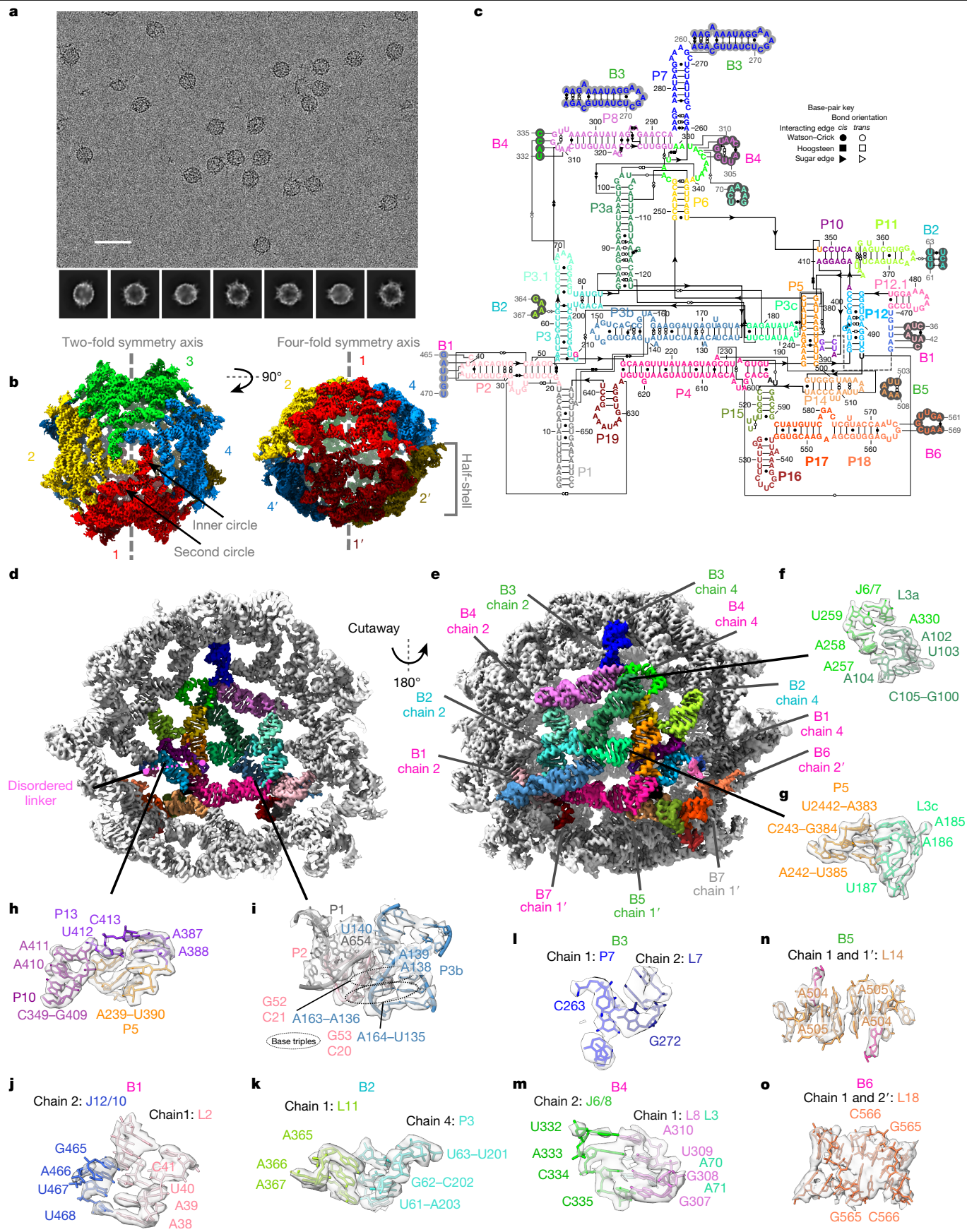


Fig. 2 | See next page for caption.

Fig. 2 | Atomically ordered structure of ROOL homo-octamer.

a, Representative micrograph (top; 4,462 micrographs total) with 2D class averages (bottom). Scale bar, 50 nm. **b**, The 3.1 Å cryo-EM reconstruction of the ROOL complex with D_4 symmetry. The map is coloured by 8 labelled chains. In the top view (left), the inner and second circle are labelled. **c**, Secondary structure of ROOL coloured by domain. Only one chain is shown, in full. Nucleotides involved in intermolecular interactions have been circled in light or dark grey. **d, e**, Chain 1 is coloured by domain with all other chains in grey. **d**, Cutaway view, showing chain 1 from the interior of the nanocage with the

disordered linker labelled in pink (nucleotides 414–464). **e**, Intermolecular interactions or bridges of chain 1 are labelled, with kissing loops labelled in magenta, A-minor interactions in cyan and other interactions in lime. One interaction, B7 (grey), is not ordered in this cryo-EM map, but residues come in sufficiently close contact that interactions could form. The same interactions, but between different pairs of chains, share the same number. **f–i**, Selected intramolecular interactions as labelled in **d, e, j–o**. Intermolecular interactions as labelled in **e**. The sharpened cryo-EM map is displayed at the following contours: 6σ (**b, d, e**), 8σ (**o**), 16σ (**g, i–n**) and 20σ (**f, h**).

organizing these proteins. In particular, the protein OapC was previously hypothesized to bind a kink turn between J4a/5 and J5/6, and binding of OapC was thought to alter secondary structure, in particular increasing protection of J6/7 to in-line hydrolysis¹⁷. Our OLE dimer structure supports formation of a kink turn^{26,27} in J4a/5 at the base of P5, but this kink turn is formed with J6/7, not J5/6 (Fig. 1g). The previously observed protection of J6/7 may therefore be explained by direct binding to the protein, and not by a rearrangement of secondary structure. In addition, whereas the internal loop of the P6 stem is different from the previously proposed one, it exposes residues I63–I65, which were proposed to bind the protein RpsU¹⁵. A163 is flipped out of the helix and docks into a pocket created by P5, P6, P7 and dimer interface. This OLE dimer pocket is reminiscent of the pocket RpsU occupies in the ribosome, supporting the previous hypothesis that OLE could sequester RpsU¹⁵.

ROOL assembles into an ordered nanocage

ROOL is a class of RNAs that is encoded in a wide variety of bacterial prophages and phages, often near tRNA islands^{1,11,12}. The predicted secondary structure is highly complex with multiple pseudoknots, but no protein binding partners have been identified, leading to the hypothesis that ROOL may function as an RNA-only complex¹. Although no function has been described for ROOL, it has been shown to be as abundant as 16S ribosomal RNA, but non-essential, in at least one strain of *Ligilactobacillus salivarius*¹².

The 659-nt ROOL env-120, discovered in cow rumen^{1,28}, produces visually clear, symmetric particles in cryo-EM micrographs (Fig. 2a and Extended Data Fig. 2). The 3.1 Å reconstructed map reveals a closed, hollow nanocage structure that comprises 8 chains with dihedral symmetry and a diameter of approximately 280 Å, larger than the maximal dimension of a bacterial ribosome (approximately 250 Å) (Fig. 2b and Supplementary Video 2). Each chain has a secondary structure that is consistent with the stems P1 to P19 proposed previously by covariation analysis¹², including the pseudoknot P10 (Fig. 2c). Atomic models for each chain can be built with a good match to the map density as shown by the Q-scores²⁴ (Extended Data Fig. 2e). Our model shows intramolecular tertiary interactions (Fig. 2d–i), which scaffold the flat monomer structure (Fig. 2f–i), including a set of non-canonical base pairs and stacking interactions that connect loops L3a and J6/7 (Fig. 2f), an A-minor interaction between L3c and P5 (Fig. 2g), an additional pseudoknot P13 (adjacent to P5 and P10, Fig. 2h), a complex set of non-canonical pairs between nucleotides that are already in stems P1, P2 and P3b (Fig. 2i), and other motifs (Supplementary Table 1).

The ROOL quaternary complex is an octameric nanocage, with a top and bottom half-shells, each formed by 4 chains, hereafter labelled chains 1–4 and 1'–4'. Within a half-shell, each chain forms 8 bridges with its neighbours, 4 on each side, labelled B1–B4 (Fig. 2j–o). Starting from the top, the loop of stem P7 forms an isolated base pair with a bulged out base in stem 7 of the next chain (B3, Fig. 2l). This 'daisy chain' of interacting stem-loops forms an inner circle on the top of the half-shell approximately 36 Å in diameter (Fig. 2b). The P7 stem is not always conserved, but a second circle of RNA (Fig. 2b), involving a quaternary kissing loop (B4, Fig. 2m), is highly conserved in evolution, and was

identified as tertiary interaction by previous covariation analysis¹. An A-minor interaction (B2, Fig. 2k) and a novel quaternary kissing loop (B1, Fig. 2j) further glue together the chains in the half-shell. Between a novel intramolecular tertiary interaction (Fig. 2g) and the intermolecular kissing loop B1 (Fig. 2j), we identified a disordered region that appears to be located inside the nanocage, based on the position of flanking regions (Fig. 2d and Extended Data Fig. 2d,e). This region was previously identified as a linker with little to no sequence or structural similarity across homologues¹.

As opposed to a simple dimer such as the OLE interface, where each chain interacts with a single partner, in the ROOL complex, each chain reaches over and interacts with two chains in the other half-shell. These interactions favour the full cage assembly, as opposed to isolated dimers. B5 and B6 are quaternary interactions in which the same sequences from different chains interact via adenosine stacking and Watson–Crick–Franklin base pairing, respectively (Fig. 2n,o). An additional interaction between the internal loop J17/18 of chain 1, previously proposed to form a pseudoknot with the flank of the linker region, and P19 of chain 1' seems plausible given their proximity, but that region was not well-resolved in our structure.

GOLLD assembles into a distinct nanocage

GOLLD RNAs are the largest among the three RNA classes analysed here, with many members exceeding 800 nucleotides in length^{3,11,29}. GOLLD, similar to ROOL, is a molecule of unknown function encoded in bacterial prophages and phages, often near tRNA islands, but with sequences and secondary structures that are distinct from those of ROOL^{3,11,29}. GOLLD expression has been shown to increase during the lysis of bacterial cells infected by phage¹¹. Unlike ROOL, the predicted secondary structures of GOLLD RNAs consist of a universally conserved 3' region and a less conserved 5' region³.

The GOLLD env-38 RNA, first identified in a marine metagenomic sample downstream of Met-tRNA^{3,30}, produces visually striking flower-like particles in cryo-EM micrographs (Fig. 3a and Extended Data Fig. 3). The 3D reconstruction at 3.0 Å resolution shows that GOLLD forms a nanocage, similar to the one formed by ROOL, but larger. The GOLLD structure is a closed 14-mer with D_7 quaternary symmetry, with a diameter of 380 Å and a completely empty interior except for a disordered loop (Fig. 3b,c and Supplementary Video 3). Models for each of the 14 chains were built with Q-score²⁴ in accordance with the map resolution (Extended Data Fig. 3g). As with ROOL, kissing loops and A-minor interactions underlie the tertiary and quaternary structure of GOLLD in addition to other motifs (Fig. 3d–t and Supplementary Tables 1 and 2) but the specific interactions are distinct. Beyond confirming the accuracy of the previously predicted secondary structure with stems P1–P27 (Fig. 3d), the tertiary structure of GOLLD reveals prominent interactions, including A-minor interactions involving adenosines at the P3–P4–P5 junction (Fig. 3g), an A-minor interaction between adenosines in L26 and stem P14 (Fig. 3h) and a loop L22 that forms a pseudoknot with the nearby linker J17/22, in addition to an A-minor interaction with that pseudoknot (Fig. 3i). Furthermore, loop L27 brings together seven regions by forming base pairs with stem P23 and linker J24/26 as well as base-backbone interactions with

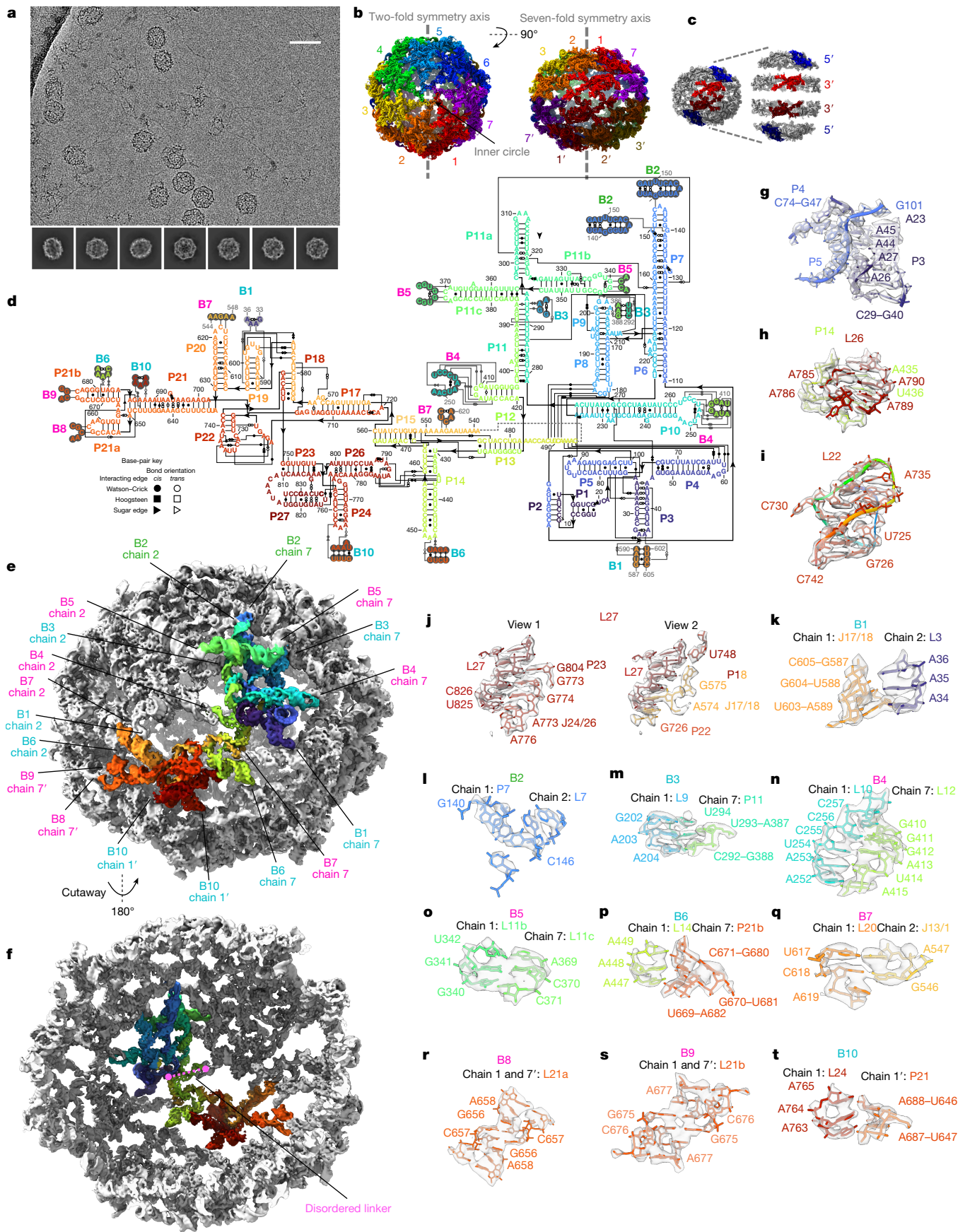


Fig. 3 | See next page for caption.

Fig. 3 | Atomically ordered structure of GOLLD homo-14-mer.

a, Representative micrograph (23,281 micrographs total) and 2D class averages. Scale bar, 50 nm. **b**, The 3.0 Å cryo-EM reconstruction of GOLLD with D_7 symmetry, coloured by chain. In the top view (left), the inner circle is labelled. **c**, The 5' (blue, residues 1–420) and 3' (red, residues 421–833) regions of GOLLD organize into the cap and a ring of the half-shell of the nanocage, respectively. To demonstrate the separation of domains, the four regions are artificially moved apart. **d**, The secondary structure of GOLLD. Only one chain is displayed in full, nucleotides participating in intermolecular interactions that are from other chains are circled in light or dark grey. **e**, One chain of GOLLD is depicted

two additional stems, P18 and P22, and linker J17/18 (Fig. 3j). Similar to ROOL, the variable linker within each chain is not resolved, but the positions of immediate flanking sequences in the 5' and 3' regions indicate that the linker resides in the interior of the cage (Fig. 3f and Extended Data Fig. 3f,g). Globally, the cryo-EM structure shows that the 5' region and the 3' region form separate domains in the 3D structure (Fig. 3c). This separation could explain why the 3' and 5' domains are divergent in GOLLD, whereas, in ROOL, the 5' and 3' regions are intertwined and hence have to co-evolve to maintain the tertiary and quaternary structure.

The 5' domains of GOLLD form the cap of each half-shell of the nanocage. Within the cap of each half-shell, each of 7 monomers forms 8 quaternary bridges to other chains—4 on each side, including kissing loops, A-minor interactions and other interactions (B2–B5, Fig. 3e,l–o). B2 (Fig. 3l) closely resembles the daisy chain of interacting stem-loops from ROOL, except that the distance between the pairs of interacting residues is reduced from 9 nt to 4 nt. This compensates for the increased number of chains in GOLLD, resulting in an inner circle of roughly the same diameter as of ROOL. In GOLLD, the only non-interacting loop with a conserved sequence, L11a (sequence GAAA), points towards this inner circle. The 3' regions complete the half-shell below this 5' cap through two interactions: an A-minor interaction (B6, Fig. 3p) and a kissing loop between L20 and J13/15, which was previously identified by covariation analysis² and here shown to be an intermolecular bridge (B7, Fig. 3q). Only a single intermolecular A-minor interaction, B1, glues the 3' and 5' regions from different chains together (Fig. 3k).

Finally, similar to the ROOL nanocage, the two half-shells come together with each chain in the top half-shell interacting with two chains in the bottom half-shell. In the GOLLD nanocage, these interactions consist of two self-interacting kissing-loop interactions (B8 and B9, Fig. 3r,s) and an A-minor interaction (B10, Fig. 3t) involving 3' regions from different chains.

Biological relevance of homo-multimers

Symmetric multimers are common among proteins and rationally designed RNA molecules^{26,31–33}, but observations of natural RNA multimers are rare. When observed, natural RNA homomeric interactions typically involve a single contact¹³. Further, with notable exceptions of viruses, such as HIV and other retroviruses³⁴ and the Φ 29 bacteriophage^{35–37}, the biological relevance of RNA homomeric complexes has not been demonstrated, leaving the possibility that they form only at high RNA concentrations and extreme ionic conditions or in the context of the specific constructs chosen for in vitro structural characterization. By contrast, several lines of evidence support GOLLD, ROOL and OLE forming multimers in their biological contexts.

First, concomitant with the same set of cryo-EM studies presented above, we resolved a 2.9 Å resolution map of another large RNA molecule, the *raiA* motif from *C. acetobutylicum*^{23,38}, as a pure monomer (Extended Data Figs. 4 and 5 and Supplementary Text 1), refuting the possibility that any large RNA would form a multimer in our experimental conditions. Independent studies have also resolved the *raiA* motif as a monomer^{39,40}. Additionally, we characterized the 343-nt HNH

in rainbow, intermolecular interactions or bridges are labelled, with kissing loops labelled in magenta, A-minor interactions in cyan and other interactions in lime. The same interaction, but between different pairs of chains, share the same number. Each chain interacts with four other chains. **f**, Same as **e**, but rotated and is cut away to show the rainbow-labelled chain from the interior of the nanocage with the disordered linker labelled in pink (nucleotides 497–538). **g–j**, Select intramolecular interactions. **k–t**, Intermolecular interactions as labelled in **e**. The sharpened cryo-EM map is displayed at the following contours: 12 σ (**e,f,i**), 14 σ (**c,o**), 16 σ (**b,n**), 18 σ (**s,q**), 20 σ (**k–m,p**), 22 σ (**t**), 25 σ (**h,i**) and 30 σ (**g,j**).

endonuclease-associated RNA and open reading frame (HEARO)³⁹ from *Limnospira maxima*, which is known to form a defined RNA structure that is involved in DNA nickase activity when bound to the protein IsrB⁴¹. Unlike the 5' region of OLE, which is also known to bind proteins, the HEARO RNA was disordered in the absence of the protein (Extended Data Fig. 6), suggesting that multimer formation of protein-binding RNAs is not an artefact of cryo-EM experimental conditions.

Second, mass photometry, which gives high precision estimates of molecular weight but requires molecular binding to surfaces, confirms the stoichiometry of GOLLD, ROOL and OLE to be 14, 8 and 2, respectively, at RNA concentrations as low as 12.5 nM (Extended Data Fig. 7). This concentration is three orders of magnitude lower than the concentrations in our cryo-EM experiments and corresponds to a population of only around ten RNA molecules in a bacterial cell, substantially lower than what is expected from observed expression levels¹.

Third, using dynamic light scattering (DLS; Extended Data Fig. 7), we confirmed that both ROOL and GOLLD primarily form thermostable multimers, with no detectable fraction of monomers, at temperatures up to 55 °C and concentrations as low as 110 nM.

Fourth, for all three structures each chain contains five or more conserved inter-subunit contacts, indicative of intricate arrangements that suggest selection pressure during the evolution of these RNAs.

Fifth, using comparative analysis of both sequences and secondary structures, we detected evolutionary conservation of structural elements and, in particular, the sites of intermolecular interactions supporting RNA homo-oligomerization (Supplementary Files 1–3 and Supplementary Table 3). Comparative analysis of OLE, ROOL and GOLLD showed that, although the sequences of these RNAs are not highly conserved, all intramolecular stems exhibit extensive base pairing supported by covariation analysis, including stems whose loops are involved in intermolecular bridges (Extended Data Fig. 8a–c, Supplementary Text 2 and Supplementary Table 3). The A positions in the OLE non-canonical A–A base-pair stem bridge B1 and the GOLLD A-minor interaction bridge B6 are highly conserved (Figs. 1d and 3p and Extended Data Fig. 8d,e). The intermolecular base pairs between ROOL J6/8 and L8 (bridge B4, Fig. 2m) were detected as a prominent, conserved quaternary interaction in prior covariation analysis⁴⁰ (Extended Data Fig. 8b and Supplementary Table 3). In other bridges, we observed intermolecular symmetric kissing loops that had base pairs between the same loop from two different chains: nucleotides 315–318 in chain A and B of OLE (B3) and nucleotides 656–657 from chain 1 and 7' of GOLLD (B8). Apparent covariance at immediately adjacent nucleotides in these loop sequences supports intermolecular base pairs because base pairing of adjacent nucleotides within the same chain is stereochemically precluded (Extended Data Fig. 8f,g). OLE L9.3 and GOLLD L21a were each found to covary in this manner, switching an internal tetranucleotide between palindromes GGCC to GAUC or AGCU and an internal dinucleotide between GC and CG, supporting bridges OLE B3 and GOLLD B8, respectively (Extended Data Fig. 8f,g). The other symmetric kissing loops in our structures, GOLLD L21b (bridge B9) and ROOL L18 (bridge B6), were highly conserved across the variants for which the loops could be confidently aligned (Extended Data Fig. 8g,h), precluding similar

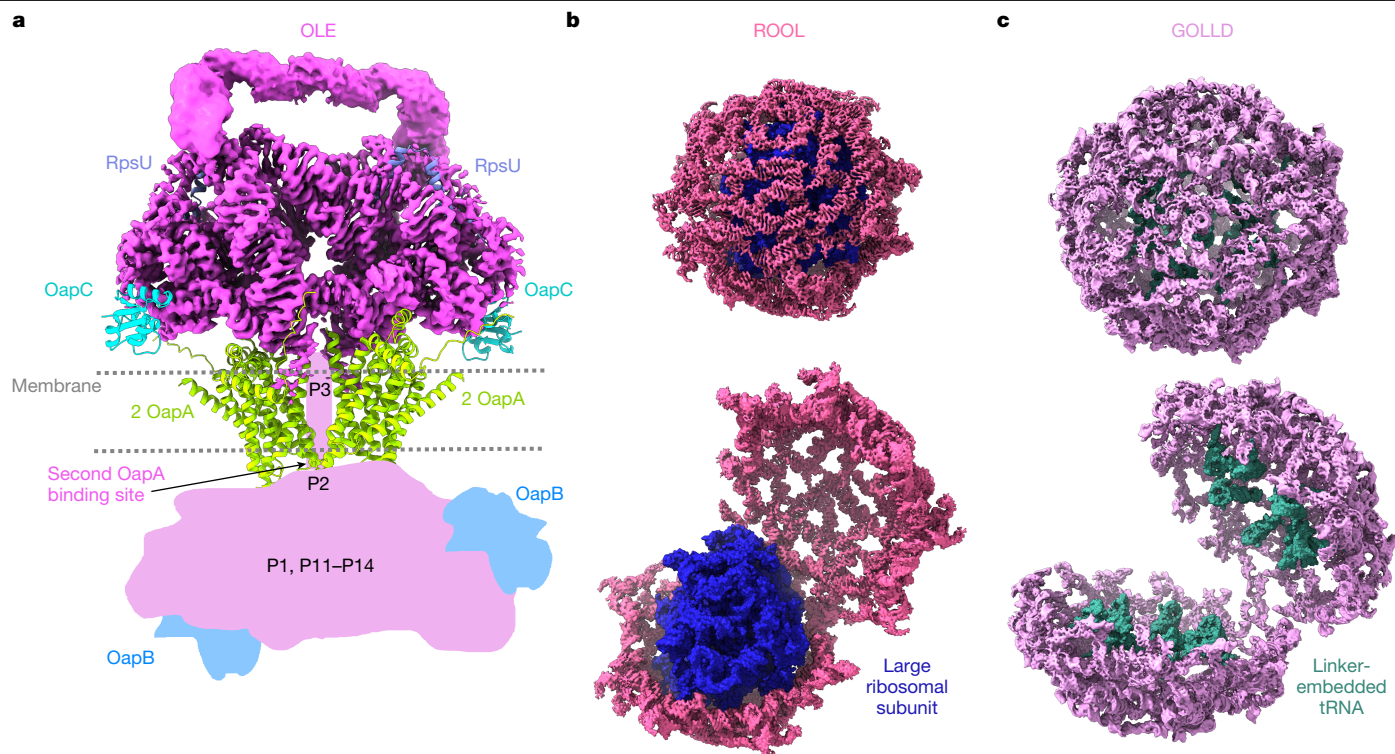


Fig. 4 | Structure-guided hypotheses for homo-oligomeric RNAs. **a**, OLE dimer displayed with AlphaFold 3⁴⁷ models of OapC, OapA dimer and RpsU proteins at their proposed binding sites. **b,c**, The two half-shells of the RNA nanocages are held together by only a few interactions and hence could open up to encapsulate other biomolecules, such as protein, metabolites,

nucleic acids or RNA–protein complexes. **b**, The RNA nanocage formed by ROOL is shown encapsulating a ribosomal large subunit. **c**, The RNA nanocage formed by GOLLD is shown exposing the covalently linked tRNAs when open. The sharpened cryo-EM maps are displayed at the following contours 7σ (**a**), 6σ (**b**) and 12σ (**c**).

covariance analysis but consistent with the importance of the observed intermolecular interactions.

Discussion

Together, our cryo-EM data, biophysical experiments and evolutionary analyses show that GOLLD, ROOL and OLE each form not only ornate secondary structures but also symmetric quaternary assemblies stabilized by many intermolecular contacts. While this Article was being revised, a publication appeared reporting similar cryo-EM structures, supporting the reproducibility of cryo-EM⁴². These structures and their complex network of RNA structure motifs offer a rich source of data for RNA structure prediction and design efforts. OLE forms a dimer shaped like a bundle of pipes and exposes structured binding pockets for protein partners such as the membrane-associated OLE-associated protein A (OapA). After superimposing an OapA dimer to each P4a site (OapA is known to bind OLE in a 2:1 ratio^{41–43}), we note that the RNA could induce the formation of an OapA tetramer. OapA is a membrane protein, and the tetramer is reminiscent of the double-stranded RNA transporter SID-1^{43–45}, suggesting that it may be able to accommodate RNA elements, such as the 3' region of OLE, which was not resolved here (Fig. 4a). In contrast to OLE, and despite unrelated sequences and distinct secondary and tertiary structures, GOLLD and ROOL both form nanocages, suggesting that their function might involve encapsulating their internal disordered linkers and/or other molecules, analogous to proteinaceous microcompartments that are common in bacteria and archaea⁴⁶. Although not large enough to enclose entire DNA genomes of their parent phages, these cages might contain macromolecules of significant size (Fig. 4b,c), such as phage-encoded tRNAs, which are sometimes present in the GOLLD linker region, bacterial ribosomes, which have been shown to bind GOLLD in pull-down assays⁴⁵, metabolites, or stress

response proteins. It remains to be determined whether nanocage formation is a common feature among large natural RNAs.

Online content

Any methods, additional references, Nature Portfolio 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/s41586-025-09073-0>.

- Weinberg, Z. et al. Detection of 224 candidate structured RNAs by comparative analysis of specific subsets of intergenic regions. *Nucleic Acids Res.* **45**, 10811–10823 (2017).
- Puerta-Fernandez, E., Barrick, J. E., Roth, A. & Breaker, R. R. Identification of a large noncoding RNA in extremophilic eubacteria. *Proc. Natl Acad. Sci. USA* **103**, 19490–19495 (2006).
- Weinberg, Z., Perreault, J., Meyer, M. M. & Breaker, R. R. Exceptional structured noncoding RNAs revealed by bacterial metagenome analysis. *Nature* **462**, 656–659 (2009).
- Narunsky, A. et al. The discovery of novel noncoding RNAs in 50 bacterial genomes. *Nucleic Acids Res.* **52**, 5152–5165 (2024).
- ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
- Nemeth, K., Bayraktar, R., Ferracin, M. & Calin, G. A. Non-coding RNAs in disease: from mechanisms to therapeutics. *Nat. Rev. Genet.* **25**, 211–232 (2024).
- Eddy, S. R. Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.* **2**, 919–929 (2001).
- Stav, S. et al. Genome-wide discovery of structured noncoding RNAs in bacteria. *BMC Microbiol.* **19**, 66 (2019).
- Chen, Y. et al. Hovlinc is a recently evolved class of ribozyme found in human lincRNA. *Nat. Chem. Biol.* **17**, 601–607 (2021).
- Ontiveros-Palacios, N. et al. Rfam 15: RNA families database in 2025. *Nucleic Acids Res.* **53**, D258–D267 (2025).
- Chen, A. G. *Functional Investigation of Ribozymes and Ribozyme Candidates in Viruses, Bacteria and Eukaryotes*. PhD dissertation, Yale University (2014).
- Cousin, F. J. et al. A long and abundant non-coding RNA in *Lactobacillus salivarius*. *Microb. Genomics* **3**, e000126 (2017).

13. Jones, C. P. & Ferré-D'Amaré, A. R. RNA quaternary structure and global symmetry. *Trends Biochem. Sci.* **40**, 211–220 (2015).
14. Lyon, S. E., Wencker, F. D. R., Fernando, C. M., Harris, K. A. & Breaker, R. R. Disruption of the bacterial OLE RNP complex impairs growth on alternative carbon sources. *PNAS Nexus* **3**, gae075 (2024).
15. Breaker, R. R., Harris, K. A., Lyon, S. E., Wencker, F. D. R. & Fernando, C. M. Evidence that OLE RNA is a component of a major stress-responsive ribonucleoprotein particle in extremophilic bacteria. *Mol. Microbiol.* **120**, 324–340 (2023).
16. Wallace, J. G., Zhou, Z. & Breaker, R. R. OLE RNA protects extremophilic bacteria from alcohol toxicity. *Nucleic Acids Res.* **40**, 6898–6907 (2012).
17. Lyon, S. E., Harris, K. A., Odzer, N. B., Wilkins, S. G. & Breaker, R. R. Ornate, large, extremophilic (OLE) RNA forms a kink turn necessary for OapC protein recognition and RNA function. *J. Biol. Chem.* **298**, 102674 (2022).
18. Widner, D. L., Harris, K. A., Corey, L. & Breaker, R. R. *Bacillus halodurans* OapB forms a high-affinity complex with the P13 region of the noncoding RNA OLE. *J. Biol. Chem.* **295**, 9326–9334 (2020).
19. Yang, Y., Harris, K. A., Widner, D. L. & Breaker, R. R. Structure of a bacterial OapB protein with its OLE RNA target gives insights into the architecture of the OLE ribonucleoprotein complex. *Proc. Natl Acad. Sci. USA* **118**, e2020393118 (2021).
20. Fernando, C. M. & Breaker, R. R. Bioinformatic prediction of proteins relevant to functions of the bacterial OLE ribonucleoprotein complex. *mSphere* **9**, e0015924 (2024).
21. Harris, K. A., Zhou, Z., Peters, M. L., Wilkins, S. G. & Breaker, R. R. A second RNA-binding protein is essential for ethanol tolerance provided by the bacterial OLE ribonucleoprotein complex. *Proc. Natl Acad. Sci. USA* **115**, E6319–E6328 (2018).
22. Block, K. F., Puerta-Fernandez, E., Wallace, J. G. & Breaker, R. R. Association of OLE RNA with bacterial membranes via an RNA-protein interaction. *Mol. Microbiol.* **79**, 21–34 (2011).
23. Nölling, J. et al. Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J. Bacteriol.* **183**, 4823–4838 (2001).
24. Pintilie, G. et al. Measurement of atom resolvability in cryo-EM maps with Q-scores. *Nat. Methods* **17**, 328–334 (2020).
25. Frank, J. et al. A model of the translational apparatus based on a three-dimensional reconstruction of the *Escherichia coli* ribosome. *Biochem. Cell Biol.* **73**, 757–765 (1995).
26. Huang, L. & Lilley, D. M. J. A quasi-cyclic RNA nano-scale molecular object constructed using kink turns. *Nanoscale* **8**, 15189–15195 (2016).
27. Klein, D. J., Schmeing, T. M., Moore, P. B. & Steitz, T. A. The kink-turn: a new RNA secondary structure motif. *EMBO J.* **20**, 4214–4221 (2001).
28. Hess, M. et al. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* **331**, 463–467 (2011).
29. Morgado, S., Antunes, D., Caffarena, E. & Vicente, A. C. The rare IncRNA GOLLDD is widespread and structurally conserved among *Mycobacterium* tRNA arrays. *RNA Biol.* **17**, 1001–1008 (2020).
30. Yooseph, S. et al. The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. *PLoS Biol.* **5**, e16 (2007).
31. Hao, C. et al. Construction of RNA nanocages by re-engineering the packaging RNA of Phi29 bacteriophage. *Nat. Commun.* **5**, 3890 (2014).
32. Liu, D., Thélot, F. A., Piccirilli, J. A., Liao, M. & Yin, P. Sub-3-Å cryo-EM structure of RNA enabled by engineered homomeric self-assembly. *Nat. Methods* **19**, 576–585 (2022).
33. Jaeger, L. & Chworos, A. The architectonics of programmable RNA and DNA nanostructures. *Curr. Opin. Struct. Biol.* **16**, 531–543 (2006).
34. Dubois, N., Marquet, R., Paillart, J.-C. & Bernacchi, S. Retroviral RNA dimerization: From structure to functions. *Front. Microbiol.* **9**, 527 (2018).
35. Ding, F. et al. Structure and assembly of the essential RNA ring component of a viral DNA packaging motor. *Proc. Natl Acad. Sci. USA* **108**, 7357–7362 (2011).
36. Simpson, A. A. et al. Structure of the bacteriophage phi29 DNA packaging motor. *Nature* **408**, 745–750 (2000).
37. Xu, J., Wang, D., Gui, M. & Xiang, Y. Structural assembly of the tailed bacteriophage phi29. *Nat. Commun.* **10**, 2366 (2019).
38. Soares, L. W., King, C. G., Fernando, C. M., Roth, A. & Breaker, R. R. Genetic disruption of the bacterial *raiA* motif noncoding RNA causes defects in sporulation and aggregation. *Proc. Natl Acad. Sci. USA* **121**, e2318008121 (2024).
39. Badepally, N. G., de Moura, T. R., Purta, E., Baulin, E. F. & Bujnicki, J. M. Cryo-EM structure of *raiA* ncRNA from *Clostridium* reveals a new RNA 3D fold. *J. Mol. Biol.* **436**, 168833 (2024).
40. Haack, D. B. et al. Scaffold-enabled high-resolution cryo-EM structure determination of RNA. *Nat. Commun.* **16**, 880 (2025).
41. Hirano, S. et al. Structure of the OMEGA nickase IsrB in complex with omegaRNA and target DNA. *Nature* **610**, 575–581 (2022).
42. Wang, L. et al. Cryo-EM reveals mechanisms of natural RNA multivalency. *Science* **388**, 545–550 (2025).
43. Zheng, L. et al. Cryo-EM structures of human SID-1 transmembrane family proteins and implications for their low-pH-dependent RNA transport activity. *Cell Res.* **34**, 80–83 (2024).
44. Hirano, Y. et al. Cryo-EM analysis reveals human SID-1 transmembrane family member 1 dynamics underlying lipid hydrolytic activity. *Commun. Biol.* **7**, 664 (2024).
45. Qian, D. et al. Structural insight into the human SID1 transmembrane family member 2 reveals its lipid hydrolytic activity. *Nat. Commun.* **14**, 3568 (2023).
46. Greening, C. & Lithgow, T. Formation and function of bacterial organelles. *Nat. Rev. Microbiol.* **18**, 677–689 (2020).
47. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025

Methods

In vitro RNA synthesis

DNA templates containing the RNA sequence of interest prepended with the T7 promoter (see Supplementary Table 4 for sequences) were ordered as gBlocks from IDT. Primers designed to amplify these sequences (see Supplementary Table 4 for sequences) were also from IDT. PCR amplification was carried out with NEBNext Ultra II Q5 Master Mix (NEB M0544S) using 10 ng of template per reaction. The thermocycler settings were: 98 °C for 30 s, then 35 cycles of 98 °C for 10 s, 55 °C for 30 s, then 72 °C for 30 s, and a final step of 72 °C for 5 min. The PCR products were then column purified using the QIAquick PCR Purification Kit (Qiagen 28104) and run on a 2% E-Gel agarose gel (Thermo Scientific A42135) to check DNA quality. DNA concentration was measured using a NanoDrop. Purified DNA smaller than 515 bp were in vitro transcribed using TranscriptAid T7 High Yield Transcription Kit (Thermo Scientific K0441) with 6 µl of DNA template per reaction. Purified DNA longer than 515 bp were in vitro transcribed using MEGAscript T7 Transcription Kit (Thermo Scientific AM1334) with 6–8 µl DNA template per reaction. These in vitro transcription reactions were incubated at 37 °C for 6 h, then held at 4 °C before DNase treatment. The RNA was then purified using the RNA Clean & Concentrator-25 Kit (Zymo Research R1017) and eluted in 30 µl water. The concentration of purified RNA was measured using a NanoDrop, and the quality was checked using the Agilent 2100 Bioanalyzer (Nano RNA Assay, run by the Stanford PAN Facility; Bioanalyzer 2100 Expert B.02.11.S1824), as shown in Extended Data Fig. 7a.

RNA folding

For all subsequent experiments, RNA was re-folded using the same basic protocol. RNA concentrations used and any other modifications to this standard protocol are mentioned in each section. RNA was denatured (90 °C for 3 min, room temperature for 10 min) in 50 mM Na-HEPES pH 8.0. RNA was then folded with 10 mM MgCl₂ at 50 °C for 20 min, and cooled to room temperature for at least 10 min before taking measurements.

Mass photometry

Mass photometry data were collected using the Refeyn TwoMP, using AcquireMP version 2024-R1.1 and DiscoverMP version 2024-R1 to obtain histogram data. For the OLE data, coated glass slides from the MP Sample Preparation Pack (MP-CON-21014) were used, for the ROOL and GOLLD data Mass Glass UC slides (MP-CON-41001) were used after coating with poly-L-lysine. The instrument was focused using droplet-dilution. Data were collected for 1 min using the large image size. The contrast data were calibrated to nucleotide length using the Millennium RNA Markers (Thermo Scientific AM7150). Gaussians were fitted by the automatic analysis in DiscoverMP. The resulting data and plotting code can be found in the accompanying GitHub repository.

Mass photometry data were not reliable for the *raiA* motif. *raiA* motif RNA was folded at 1 µM following the standard procedure above. On the stage two dilutions were attempted, 15 µl buffer:2 µl sample for final concentration of 118 nM and 18 µl buffer:2 µl 10x diluted sample for a final concentration of 10 nM. There is a known issue with nucleic acid samples, whereby there are noisy low-mass peaks⁴⁸ (communication with the company Refeyn). These are not present in the buffer alone. For this reason, *raiA* motif (205 nt) is below the recommended minimal size for mass photometry, and indeed when we attempted to collect data on the *raiA* motif we observed noise peaks, containing the size of *raiA* motif monomer but smaller than any multimer, in both binding and unbinding regimes, indicating unreliable results (data not shown).

OLE was folded at 0.25 µM and was diluted to 12.5 nM on the stage. OLE folded in various buffers, all including 50 mM Na-HEPES pH 8.0, with other components added at the time when MgCl₂ is added in the standard protocol: (1) nothing added; (2) 1 mM MgCl₂; (3) 10 mM MgCl₂, standard; (4) 100 mM MgCl₂; (5) 10 mM MgCl₂ and 1% ethanol; (6) 10 mM MgCl₂

and 5% ethanol; (7) 0 mM MgCl₂ and 200 mM KCl; (8) 10 mM MgCl₂ and 200 mM KCl; (9) 0 mM MgCl₂ and 200 mM NaCl; and (10) 10 mM MgCl₂ and 200 mM NaCl. Buffers with MnCl₂ were attempted but the manganese saturated the detector.

ROOL and GOLLD were folded at 1 µM. The samples were diluted 10× prior to taking data. On the stage the samples were further diluted (10 µl buffer:10 µl sample) for a final concentration of 50 nM.

DLS of RNA nanocages

RNA was folded at 30 ng µl⁻¹ using the standard folding protocol. DLS traces were collected using the Prometheus Panta. Two replicates (2 capillaries of 10 µl volume, NanoTemper PR-CO02) for each RNA were obtained. DLS data of 10× 5 s acquisitions per capillary with laser power 100% were obtained using PR.PantaControl v.1.8.0. The auto-correlation function was calculated and size distribution was fitted using default parameters in PR.PantaAnalysis v.1.8.0. The resulting size distribution tables and plotting code can be found in the accompanying GitHub repository.

Cryo-electron microscopy grid preparation

For all samples, the RNA was frozen using a VitroBot Mark IV, using no. 542 filter paper and Quantifoil 1.2/1.3 200 mesh copper grids which were glow discharged for 30 s at 15 mA. GOLLD was folded at 8 µM, using the standard folding conditions except, after the 50 °C incubation, the temperature was lowered to 37 °C at a rate of 0.1 °C s⁻¹, held at 37 °C for 2 min, and then reduced to room temperature at a rate of 0.1 °C s⁻¹. To increase concentration of GOLLD in the ice, 4 cycles of applying 2 µl of sample and blotting for 3 s were performed before plunging. ROOL was folded at 9.1 µM with the standard folding protocol. The grid was coated with 2 µl of 100 mM NaCl which was blotted for 3 s. Then, 2 µl sample was immediately applied to the grid and blotted for 3 s before plunging into liquid ethane. OLE and *raiA* motif RNA were frozen with the standard folding protocol at 20 µM and 15 µM respectively; 2 µl of sample was applied to the grid, followed by 3 s blot and plunge into liquid ethane.

Cryo-electron microscopy data collection

All datasets were collected on Titan Krios G3 microscopes using a 50 µm C2 aperture and 100 µm objective aperture and EPU software (v.3.5). The OLE dataset was collected using a Falcon 4 camera with a 10 eV slit on a Selectris energy filter, while the other datasets were collected using a K3 camera with a 20 eV slit on a Bio Quantum energy filter and EPU software. Additional information on dose, magnification, and data collected for each RNA can be found in Extended Data Table 1.

Cryo-electron microscopy data processing

Data were processed live using CryoSparc (v.4.5.3)⁴⁹ and then further refined, including non-uniform refinement⁵⁰. For OLE and *raiA* motif per particle motion correction was performed⁵¹. For all datasets, symmetry was not applied until final refinement stages. For OLE, C₂ symmetry was applied. For ROOL and GOLLD, D₄ and D₇ symmetry were applied, respectively, followed by symmetry expansion of the particles and local refinement for one asymmetric unit. Finally, for GOLLD and ROOL subdomains of one asymmetric unit were locally refined and composite maps, and half-maps were created for one asymmetric unit and then composited to the full symmetry using phenix.combine_focused_maps (v.1.21)⁵². Local resolution was estimated using CryoSparc. See Extended Data Figs. 1–5 for more details on processing pipelines.

Modelling

Maps were sharpened using phenix.auto_sharpen with half-maps (v.1.21). Initial models for a monomer were obtained from ModelAngelo (Relion-5.0)⁵³; because current versions of ModelAngelo cannot be run on a pure RNA structure, EMDB-17659 was added to the corner of the map, the corresponding protein sequence (Protein Data Bank (PDB): 8PHE) was provided, and protein residues were subsequently

Article

deleted from the model. The RNA modelled chains were manually combined tracing the RNA sequences, adding and mutating residues when necessary (in particular, C to U mutations were commonly required). Manual model correction and refinement was accomplished in Coot (version 0.9.8)⁵⁴. Manual refinement of the monomer was performed using Isolde and Coot⁵⁴. Symmetry was applied to the model, from henceforth refinement was done asymmetrically due to limitations in refinement programs. Intermolecular interactions were analysed by hand and corrected using Isolde⁵⁵ and Coot⁵⁴. DRRAFTER⁵⁶ (Rosetta 3.10 (2020.42)) was used to fill in low resolution areas. For symmetric kissing loops, these models were selected and fit into the map and refined more symmetrically by hand using Isolde⁵⁵. Final refinement was first run through phenix.real_space_refine⁵⁷ followed by piecewise corrections using ERRASER2⁵⁸ (Rosetta 3.10 (2020.42)), followed by manual refinement in Coot⁵⁴ and Isolde⁵⁵ when necessary. A protocol was created to enable refinement on the large ROOL and GOLLD complexes. First, the monomer was refined in -30 sections splitting the model and map prior to using ERRASER2. These were then stitched together and regions encompassing the stitch sites were further refined. Finally, problematic regions of the monomer were refined further. Symmetry was applied to the monomer and the interaction sites were refined in parallel until interactions were sufficiently realistic with only minor clashes. Throughout, split points were manually edited if they caused minimization errors or to include interaction residues. The following ERRASER2 command was used, repeating if not yet converged:

```
$ERRASER -s $PDB -edensity:mapfile $MAP -edensity::mapreso $RESOLUTION -score:weights stepwise/rna/rna_res_level_energy7beta.wts -set_weights elec_dens_fast 40 cart_bonded 5.0 linear_chainbreak 10.0 chainbreak 10.0 fa_rep 1.5 fa_intra_rep 0.5 rna_torsion 10 suite-ness_bonus 5 rna_sugar_close 10 -rmsd_screen 3.0 -mute core.scoring.CartesianBondedEnergy core.scoring.electron_density.xray_scattering -rounds 3 -fasta $FASTA -cryoem_scatterers -rna:eraser:fixed_res $FIXED.
```

Validation metrics were calculated using Phenix, including phenix.rna_validate^{59–61}. ChimeraX (version 1.8)⁶² was used to calculate Q-score⁶¹ and for all visuals.

Base pairing and base stacking were identified using Rosetta rna_motif⁶³. Kink turns and ribose zippers were identified using DSSR with the “-k-turns” flag (v.1.9.9)⁶⁴. Z-anchors were manually labelled by aligning every 5-nt range of each structure to a representative Z-anchor (4E8Q residues 108–111) and manually inspected each region that had root mean squared deviation (rmsd) < 4 Å. Secondary structure was drawn using RiboDraw with manual manipulation⁶³. For visualizing a hypothetical OLE RNA–protein complex, AlphaFold 3 (server version)⁴⁷ was used to predict: (1) a OapA dimer with a OapC monomer; and (2) RpsU using the sequences in (Supplementary Table 4). The OapA dimer was fitted into the proposed RNA site manually. The OapC was close to its presumed binding site, but clashed with RNA and therefore its position was manually adjusted. RpsU was also placed manually in its proposed binding site. C₂ symmetry was then applied to visualize the full complex.

Bioinformatic analysis

Bacterial genomes were downloaded from National Center for Biotechnology Information (NCBI) Genome database in February 2024 (<https://ftp.ncbi.nlm.nih.gov/genomes/genbank/bacteria/>). GenBank records for phage genomes were downloaded in March 2024 (<https://millardlab.org/bacteriophage-genomics/phage-genomes-march-2024/>). Sequence profiles of GOLLD, ROOL and OLE were downloaded from the Rfam database (<ftp.ebi.ac.uk/pub/databases/Rfam/>) on March 2024. A custom sequence profile of raiA motif was built using the reported alignment⁶³. To retrieve ncRNAs from genome sequences, cmsearch was conducted using sequence profiles with a cutoff value of 10⁻⁵ (Infernal 1.15)⁶⁵. The overall procedure yielded the following numbers of nonredundant ncRNA sequences: 806, GOLLD; 1,596, ROOL; 8,585, OLE; 4,875, raiA motif.

The Infernal software⁶⁵ (v.1.1.2) was used to compare candidate RNA structures against Rfam models (cmScan), build and calibrate new covariance models (cmbuild, cmcalibrate) for separate clusters of RNAs, and perform structure-informed homology searches (cmsearch). Comparative analysis and multiple alignments for isolated RNA candidates were conducted using cmalign⁶⁵ and MUSCLE (v.5)⁶⁶ with pairwise comparisons refined using the OWEN program⁶⁷. Evolutionary history was inferred via the Maximum Likelihood method with different models in MEGA X⁶⁸. Evolutionary analysis of compensatory substitutions in isolated clusters was performed by DecipherSSC⁶⁹.

RNAalifold (from ViennaRNA 2.7.0)⁷⁰ applied to computationally fold multiple RNA alignments, and AFold/Hybrid^{71,72} were used to predict locally folded secondary structures or hybrid duplex elements within clusters. Covariation analysis was performed with R-scape (v.1.2.3)⁷³, which annotates multiple structural alignments of RNAs using statistically significant covariations (E-value < 0.05) as base-pairing constraints.

Reporting summary

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

Data availability

The cryo-EM micrographs and particles, cryo-EM maps and model coordinates have been made available on Electron Microscopy Public Image Archive (EMPIAR), Electron Microscopy Data Bank (EMDB) and Protein Data Bank (PDB), respectively (raiA motif: EMPIAR-12706, EMD-48162 and 9ELY; OLE: EMPIAR-12707, EMD-48163 and 9MCW; ROOL: EMPIAR-12708, EMD-48179 and 9MDS; GOLLD: EMPIAR-12709, EMD-48214 and 9MEE). Bioanalyzer, DLS and mass photometry data are presented in Supplementary Data 2.

Code availability

Custom scripts can be found at https://github.com/DasLab/RNA_multimer_2024.

- Li, Y., Struwe, W. B. & Kukura, P. Single molecule mass photometry of nucleic acids. *Nucleic Acids Res.* **48**, e97 (2020).
- Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
- Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* **17**, 1214–1221 (2020).
- Rubinstein, J. L. & Brubaker, M. A. Alignment of cryo-EM movies of individual particles by optimization of image translations. *J. Struct. Biol.* **192**, 188–195 (2015).
- Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D* **75**, 861–877 (2019).
- Jamali, K. et al. Automated model building and protein identification in cryo-EM maps. *Nature* **628**, 450–457 (2024).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
- Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr. D* **74**, 519–530 (2018).
- Kappel, K. et al. De novo computational RNA modeling into cryo-EM maps of large ribonucleoprotein complexes. *Nat. Methods* **15**, 947–954 (2018).
- Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D* **74**, 531–544 (2018).
- Chou, F.-C., Sripakdeevong, P., Dibrov, S. M., Hermann, T. & Das, R. Correcting pervasive errors in RNA crystallography through enumerative structure prediction. *Nat. Methods* **10**, 74–76 (2013).
- Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
- Afonine, P. V. et al. New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr. D* **74**, 814–840 (2018).
- Richardson, J. S. et al. RNA backbone: consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *RNA* **14**, 465–481 (2008).
- Meng, E. C. et al. UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci.* **32**, e4792 (2023).
- Das, R. & Watkins, A. M. RiboDraw: semiautomated two-dimensional drawing of RNA tertiary structure diagrams. *NAR Genomics Bioinform.* **3**, lqab091 (2021).
- Lu, X.-J., Bussemaker, H. J. & Olson, W. K. DSSR: an integrated software tool for dissecting the spatial structure of RNA. *Nucleic Acids Res.* **43**, e142 (2015).

65. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).
66. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004).
67. Ogurtsov, A. Y., Roytberg, M. A., Shabalina, S. A. & Kondrashov, A. S. OWEN: aligning long collinear regions of genomes. *Bioinformatics* **18**, 1703–1704 (2002).
68. Kumar, S., Stecher, G., Li, M., Niyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
69. Spiridonov, A. N. & Shabalina, S. A. Deciphering structural selective constraints: a comparative evolutionary analysis of RNA hairpin structures. In *Computational Methods in Systems Biology: 22nd International Conference Proceedings* (eds Gori, R. et al.) 196–208 (Springer, 2024).
70. Bernhart, S. H., Hofacker, I. L., Will, S., Gruber, A. R. & Stadler, P. F. RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics* **9**, 474 (2008).
71. Ogurtsov, A. Y., Shabalina, S. A., Kondrashov, A. S. & Roytberg, M. A. Analysis of internal loops within the RNA secondary structure in almost quadratic time. *Bioinformatics* **22**, 1317–1324 (2006).
72. Kondrashov, A. S. & Shabalina, S. A. Classification of common conserved sequences in mammalian intergenic regions. *Hum. Mol. Genet.* **11**, 669–674 (2002).
73. Rivas, E., Clements, J. & Eddy, S. R. A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat. Methods* **14**, 45–48 (2017).
74. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
75. Young, G. et al. Quantitative mass imaging of single biological macromolecules. *Science* **360**, 423–427 (2018).

Acknowledgements The authors thank Stanford Research Computing Center, SLAC Shared Scientific Data Facility, the Stanford-SLAC Cryo-EM Center, the Stanford Chem-H Macromolecular Structure Knowledge Center and the Stanford Biochemistry administrative

staff for resources and support that contributed to this research. This work was supported by Stanford Bio-X (Bowes Graduate Student Fellowship to R.C.K.), the National Institute for Health (R35 GM122579 to R.D. and Common Fund Transformative High-Resolution Cryo-Electron Microscopy program U24 GM129541 to W.C.), Howard Hughes Medical Institute (HHMI) (to R.D.), the National Science Foundation (Grant No. 2330652 to R.D. and W.C.), and the G. Harold & Leila Y. Mathers Foundation (MF-2303-04116 to A.G.). S.A.S. and E.V.K. report funding from the Intramural Research Program of the National Institutes of Health of the United States of America (National Library of Medicine). This article is subject to HHMI's Open Access to Publications policy. HHMI lab heads have previously granted a nonexclusive CC BY 4.0 license to the public and a sublicensable license to HHMI in their research articles. Pursuant to those licenses, the author-accepted manuscript of this article can be made freely available under a CC BY 4.0 license immediately upon publication.

Author contributions R.C.K., W.C., and R.D. conceptualized and designed the study. R.C.K. and G.N. selected sequences for the study. Y.W. and R.C.K. performed in vitro RNA transcription and collected DLS and mass photometry data. R.C.K. froze cryo-EM grids, collected and processed cryo-EM data, and modelled the cryo-EM maps with advice from W.C. and R.D. G.N. screened cryo-EM grids. H.L. and A.G. generated sequence alignments and S.A.S., R.C.K., R.D. and E.V.K. analysed these sequences for covariation. R.C.K., S.A.S. and R.D. prepared the manuscript with input from all authors.

Competing interests The authors declare no competing interests.

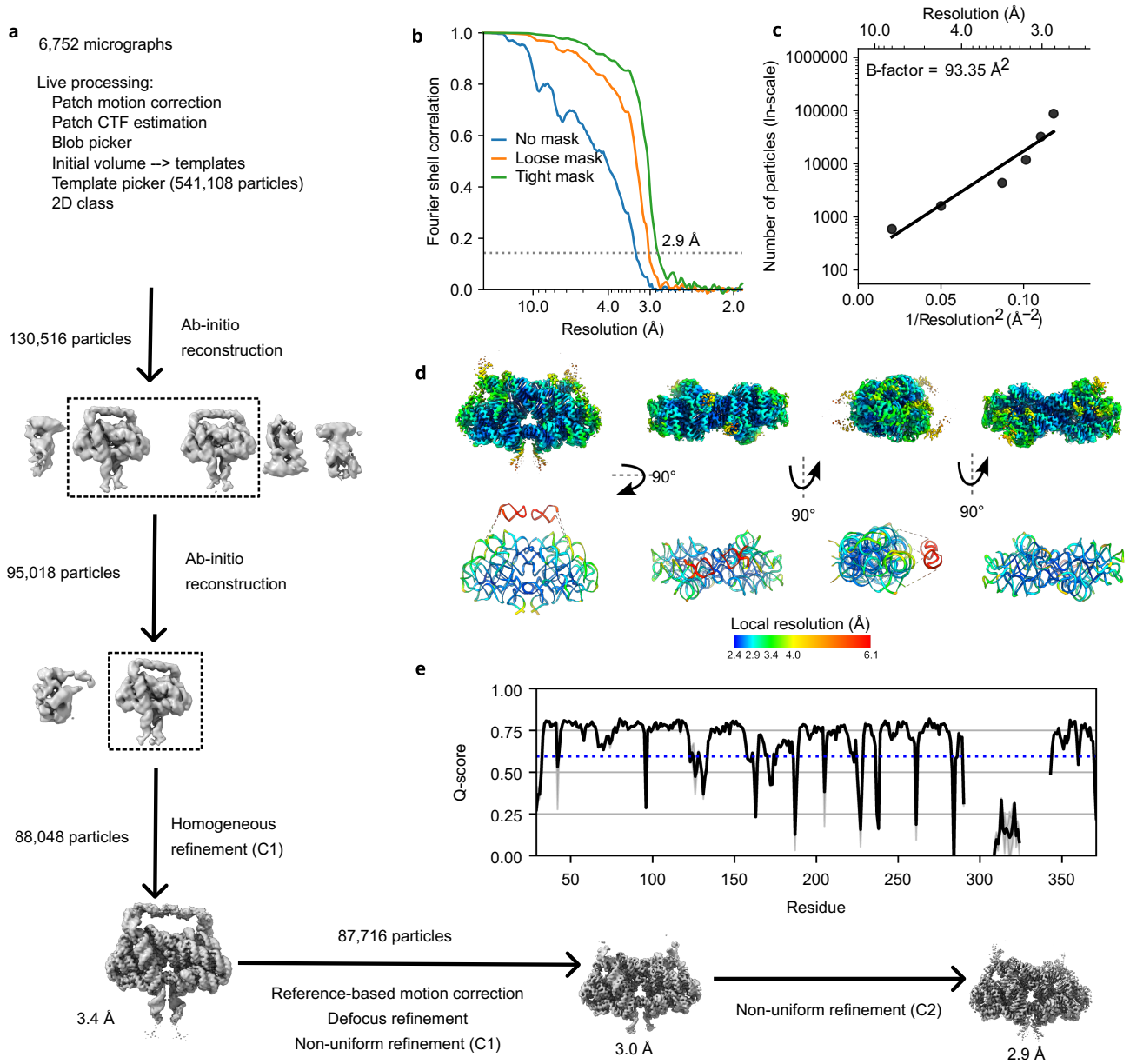
Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-025-09073-0>.

Correspondence and requests for materials should be addressed to Wah Chiu or Rhiju Das.

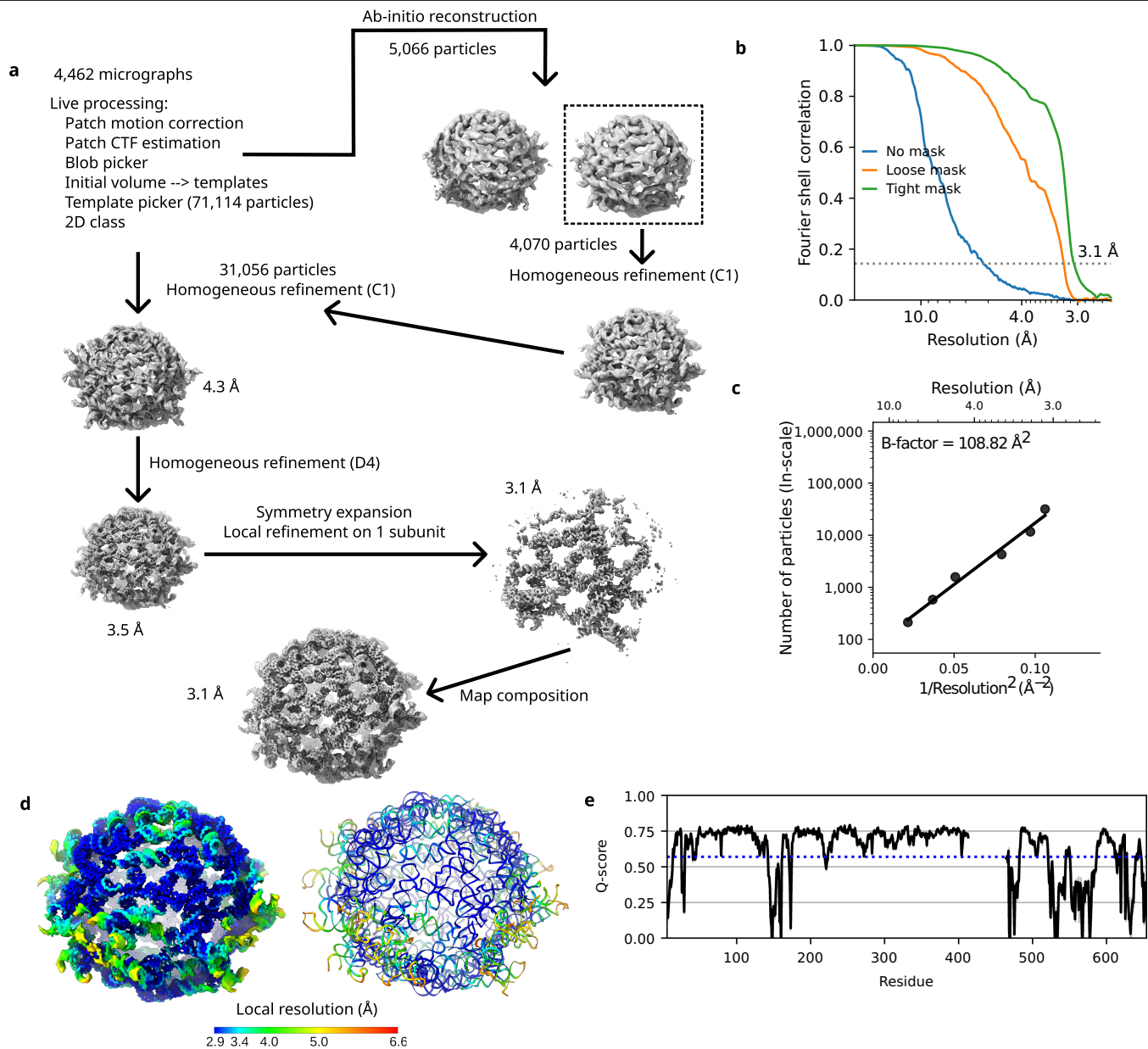
Peer review information *Nature* thanks Anna Pyle, Jane Richardson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer review reports are available.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



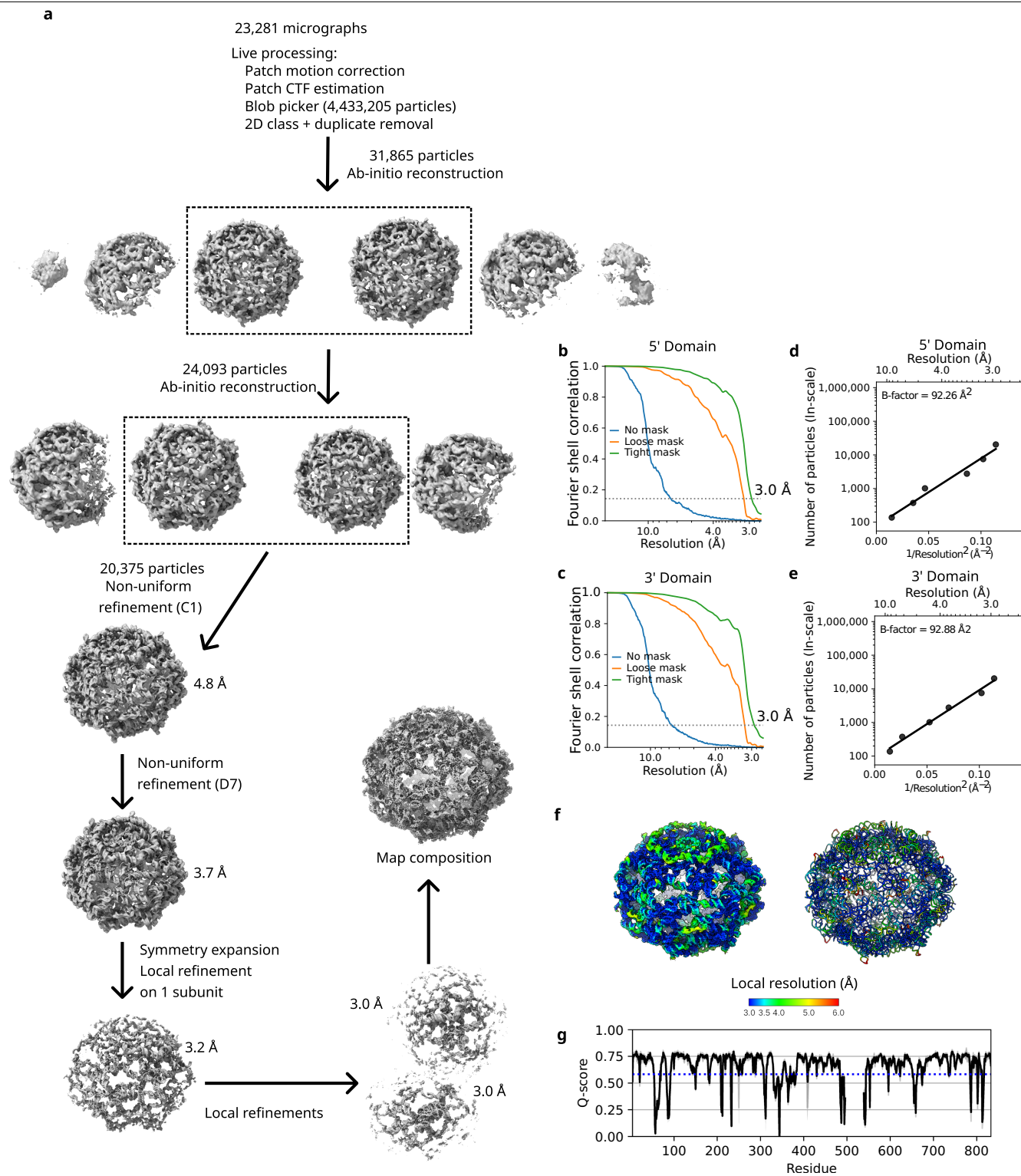
Extended Data Fig. 1 | Cryo-EM data processing workflow for OLE dimer. (a-e) OLE resolves into a high resolution dimer, even in the absence of protein. (a) Data processing flowchart for the OLE dimer. (b) Fourier shell correlation (FSC) plot for final refinement of OLE dimer. (c) Plot of particle number against the reciprocal squared resolution for OLE dimer. The B-factor was calculated as twice the linearly fitted slope⁷⁴. (d) Local resolution of the OLE dimer on the

cryo-EM map (top) and the molecular model (bottom). (e) Resolvability of the built model of the OLE dimer as measured by Q-score. The black line is the mean across all chains, with the maximum and minimum values depicted in light grey (N = 2 chains). The expected Q-score at this resolution⁷² is labeled with a blue dotted line.



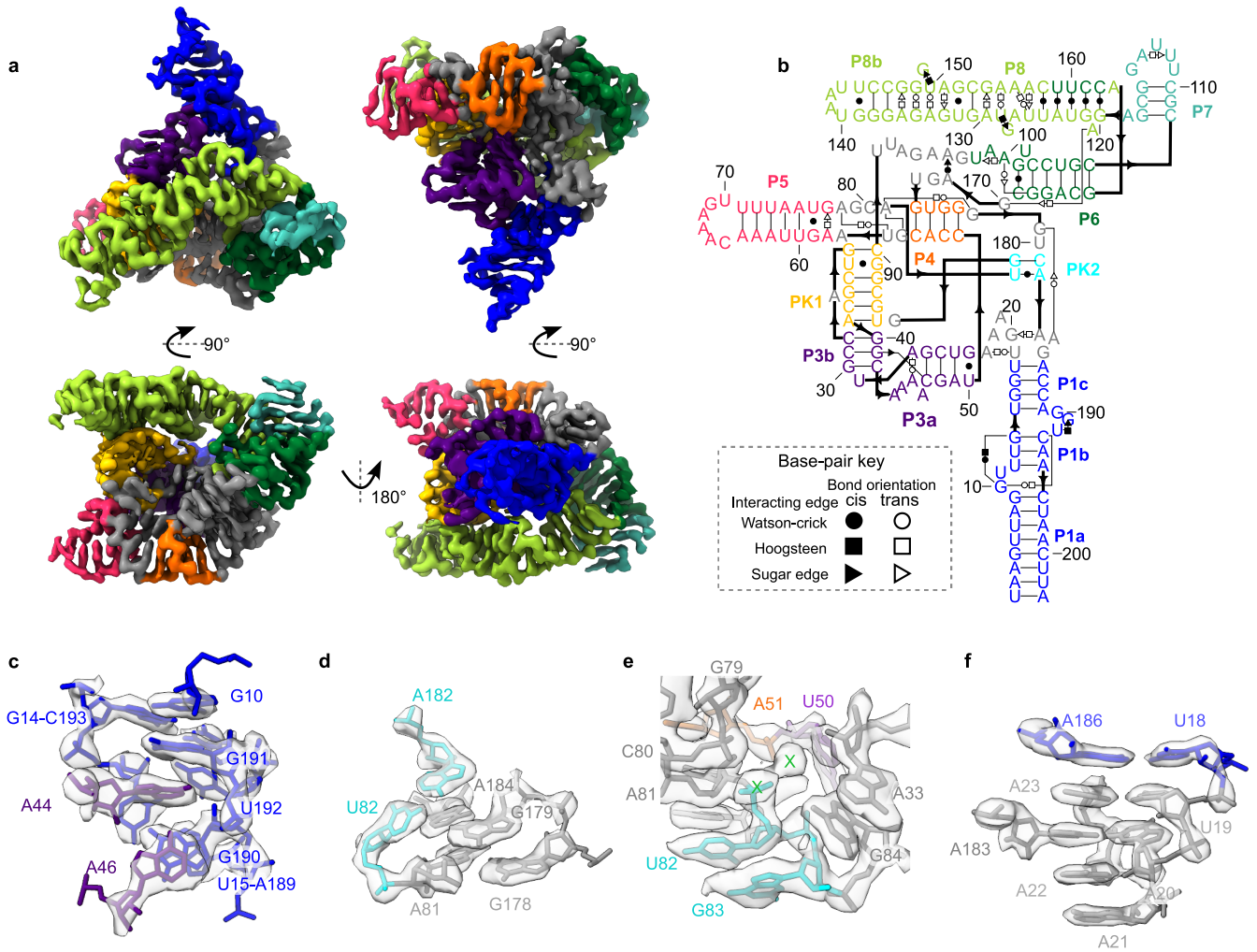
Extended Data Fig. 2 | Cryo-EM data processing workflow for ROOL nanocage complex. (a) Data processing flowchart. (b) Fourier shell correlation (FSC) plots of the single subunit local refinement. (c) Plot of particle number against the reciprocal squared resolution for the single subunit local refinement. The B-factor was calculated as twice the linearly fitted slope⁷⁴. (d) Local resolution

on the cryo-EM map (right) and the molecular model (left). (e) Resolvability of the built model as measured by Q-score. The black line is the mean across all chains, with the maximum and minimum values depicted in light grey (N = 8 chains). The expected Q-score at this resolution⁷² is labeled with a blue dotted line.



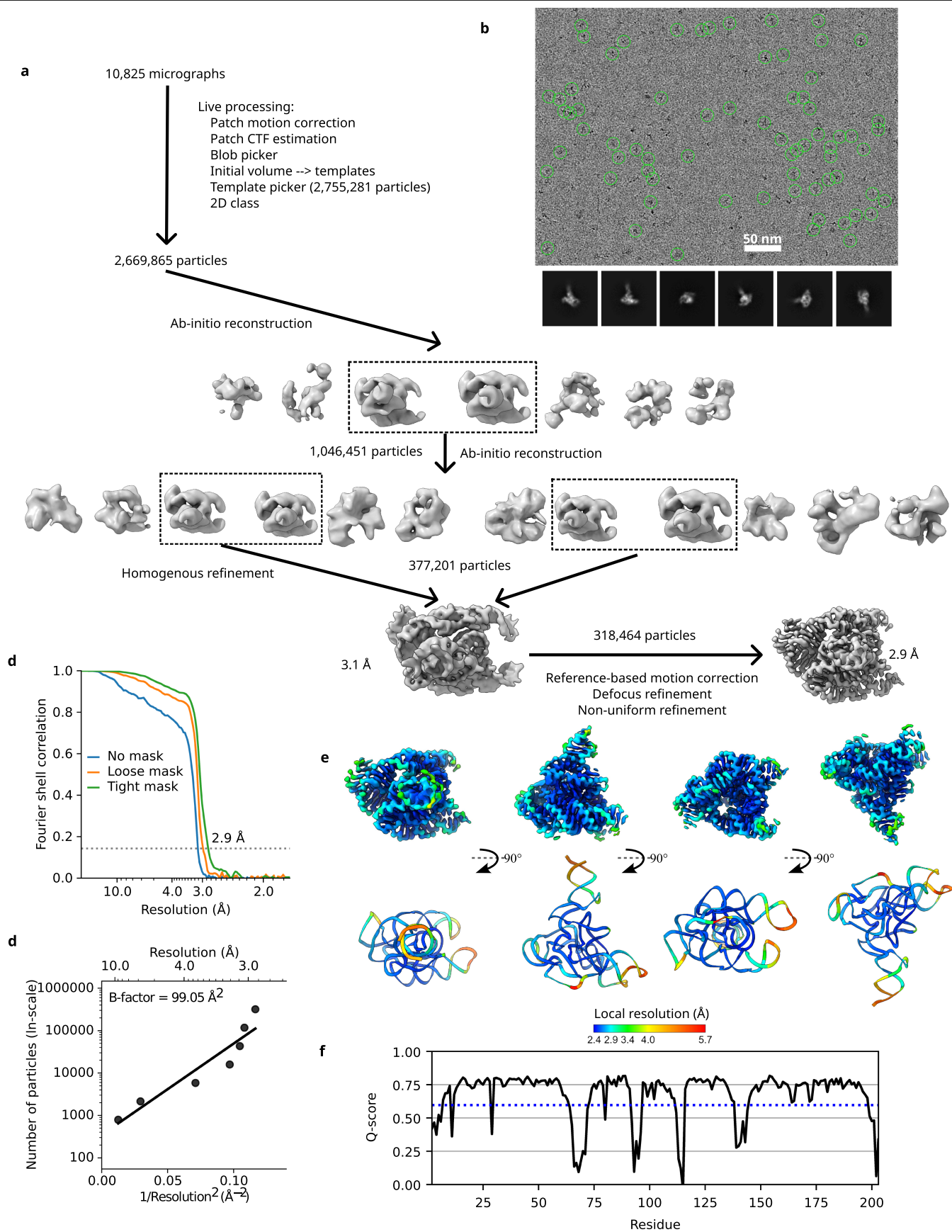
Extended Data Fig. 3 | Cryo-EM data processing workflow for GOLLN nanocage complex. (a) Data processing flowchart. (b-c) Fourier shell correlation (FSC) plots for the local refinement of the 5' and 3' domains respectively. (d-e) Plots of particle number against the reciprocal squared resolution for the local refinement of the 5' and 3' domains respectively. The B-factor was calculated as twice the linearly fitted slope⁷⁴. (f) Local

resolution on the cryo-EM map (left) and the molecular model (right). (g) Resolvability of the built model as measured by Q-score. The black line is the mean across all chains, with the maximum and minimum values depicted in light grey (N = 14 chains). The expected Q-score at this resolution⁷² is labeled with a blue dotted line.



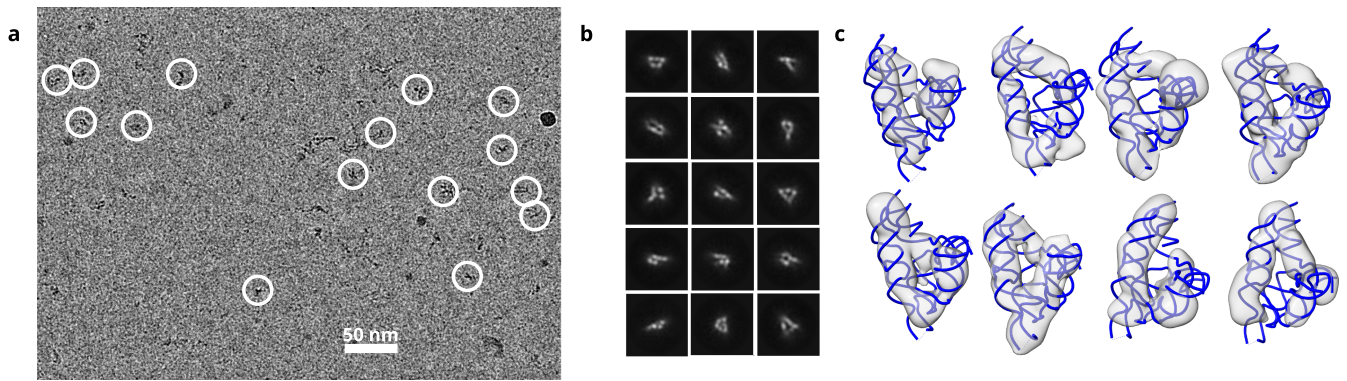
Extended Data Fig. 4 | Tertiary structure of *raiA* RNA motif. (a) Global view of tertiary structure of *raiA* motif and 2.9 Å cryo-EM map coloured by as labeled in the secondary structure, (b). (c-f) Select tertiary interactions. Description

can be found in Supplemental Text 1. The sharpened cryo-EM map is displayed at the following contours (a): 8 σ , (c,e,f): 16 σ , (d): 20 σ .



Extended Data Fig. 5 | Cryo-EM data processing workflow for *raiA* motif. (a) Data processing flowchart. (b) Representative micrograph (10,825 micrographs total) and 2D class averages. (c) Fourier shell correlation (FSC) plot. (d) Plot of particle number against the reciprocal squared resolution.

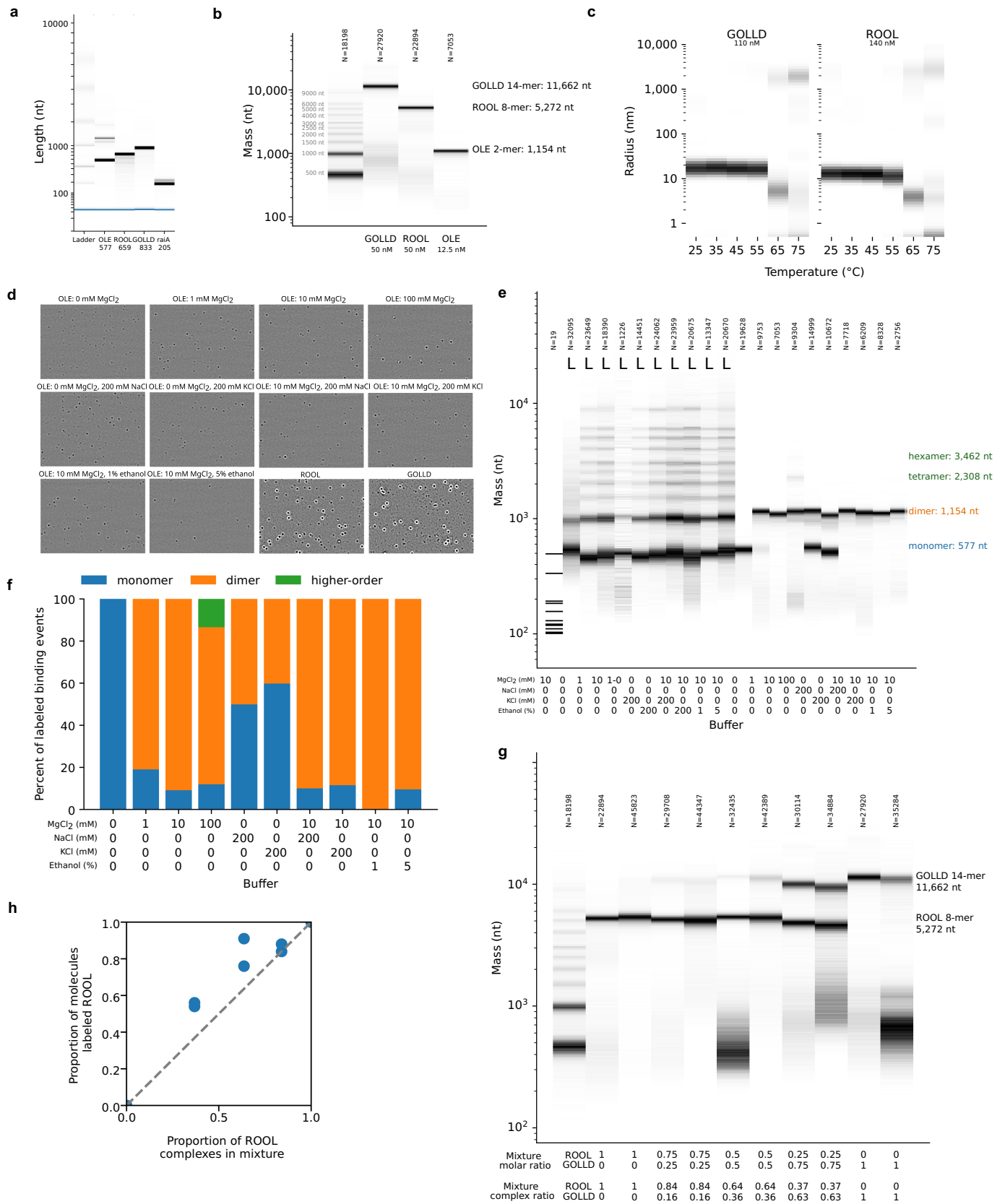
The B-factor was calculated as twice the linearly fitted slope⁷⁴. (e) Local resolution on the cryo-EM map (top) and the molecular model (bottom). (f) Resolvability of the built model as measured by Q-score. The expected Q-score of a RNA model at this resolution is labeled with a blue dotted line.



Extended Data Fig. 6 | Cryo-EM data of HEARO RNA without protein shows disorder. HEARO did not resolve into a high resolution structure, despite similar amount and quality of data as OLE-dimer. (a) The representative micrograph (8,294 micrographs total) of HEARO shows clear particles. (b) Select 2D class averages show that HEARO is forming RNA helices, but they have diverse orientations and are blurred, suggesting high flexibility. (c) 3D reconstructions

of HEARO, overlaid with the known structure of this RNA in the OMEGA nickase complex bound to protein IsrB (PDB: 8DMB⁴¹), show RNA of a similar fold to the complexed RNA. Multiple conformations are reconstructed, but with poorly resolved features, suggesting that HEARO may not form an atomically ordered structure when not in complex with its partner proteins.

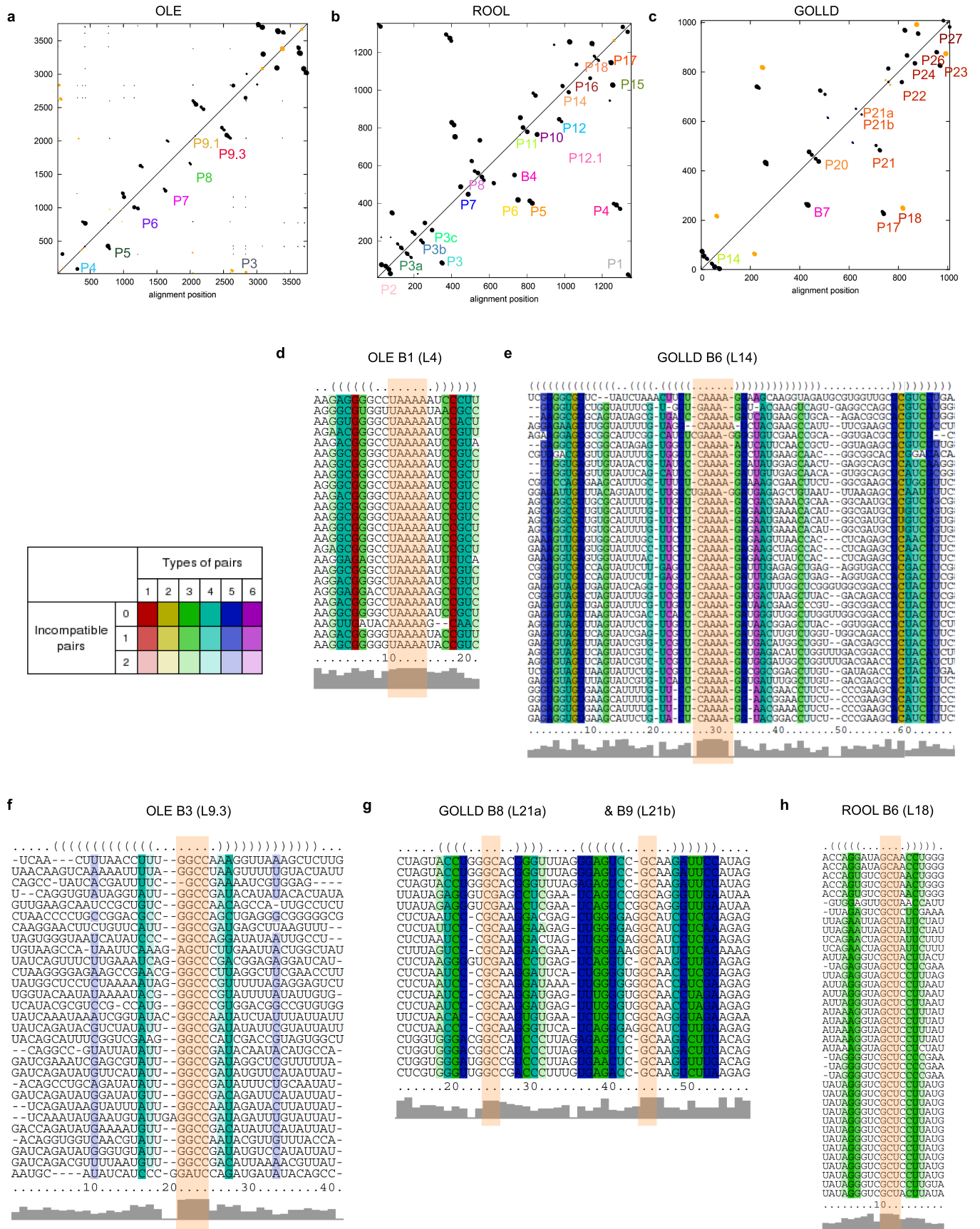
Article



Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Evidence of multimer formation of GOLLD, ROOL, and OLE in biologically relevant concentrations. (a) Agilent Bioanalyzer traces demonstrate the purity of the samples. The second peak for OLE is a common artifact of poor denaturation of sample in Bioanalyzer traces. The pure monomeric reading in mass photometry, (b), shows that this peak is likely not a covalently linked dimer. (b) Mass of GOLLD, ROOL, and OLE complexes as obtained from mass photometry at 50 nM, 50 nM, and 12.5 nM respectively. The data is a histogram of particle count density, normalized per sample, where dark is many counts, white is none. Total particle counts are shown above the graph. (c) Hydrodynamic radius of GOLLD and ROOL complexes as derived from dynamic light scattering at 110 nM and 140 nM respectively. The data are plotted as relative population density, normalized by density per sample, with dark representing highly populated radius values. The temperature of the sample was raised from 25 °C to 75 °C and dynamic light scattering traces were

obtained every 10 °C, showing complex melting into monomers at 65 °C and aggregation at high temperatures. (d) Representative ratiometric image for all mass photometry data (1 frame from a 60 s collection at 331 Hz). (e) Mass photometry data of OLE in different buffer conditions demonstrates OLE can dimerize at low RNA concentration, low magnesium concentration, and in the absence of magnesium with sufficient monovalent cations. (f) The mass photometric data is summarized by counting the amount of hits in the monomer, dimer, and high stoichiometry peaks. The absolute ratio of monomer:dimer is accurate as assessed in (g-h). (g) Mass photometry traces of mixtures of ROOL and GOLLD, ratiometric image examples can be found in. (h) Summary of the mixture results, with the known complex ratio plotted against the ratio reported by mass photometry. There is agreement, but with slight bias towards higher counts for the smaller species, ROOL, opposite of the previously observed trend⁵.



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

Extended Data Fig. 8 | Comparative and covariation sequence analysis of homo-oligomer forming RNAs. (a-c) Distributions of covariation scores in multiple sequence alignments of (a) OLE, (b) ROOL, and (c) GOLLD sequences with select stems labeled. Dot size is proportional to the covariation score. In blue the consensus base pairs are depicted; in green, the consensus base pairs that show significant covariation are shown; in orange, other pairs that have significant covariation were depicted, they are not part of the consensus secondary structure but are compatible with it; in black, other significant pairs are depicted. Positions are relative to the original input alignment (before any gapped column is removed). (d-h) Examples of multiple alignments and profiles of sequence identity of selected stable hairpins with highly conserved loops which are involved in the intermolecular interactions are shown. Nucleotides

involved in intermolecular interactions are labeled as in main Figs. 1, 2 and 3 for the RNAs OLE (B1, B3), ROOL (B6), and GOLLD (B6, B8) respectively, and highlighted with an orange box. A coloring scheme for highlighting the mutational pattern with respect to the secondary structure (folding) was used and can be found next to (d). If one predicted base-pair is formed by several different combinations of nucleotides, consistent or compensatory mutations have taken place. This is indicated by different colors. Pale colors indicate that a base pair cannot be formed in some sequences of the alignment. The sequence variants for the examples were selected from the closest branches of the evolutionary trees built based on the multiple sequence alignments used for the covariation analysis.

Article

Extended Data Table 1 | Cryo-EM experiments on four large non-coding RNAs

	OLE EMPIAR- 12707 EMDB-48163 PDB-9MCW	ROOL EMPIAR- 12708 EMDB-48179 PDB-9MDS	GOLLD EMPIAR- 12709 EMDB-48214 PDB-9MEE	raiA motif EMPIAR- 12706 EMDB-48162 PDB-9ELY
Sample preparation				
RNA length (nt)	577	659	833	205
Molecular weight (kDa)	187	213	269	67
Stoichiometry	2	8	14	1
Complex molecular weight (kDa)	373	1,702	3,768	67
Concentration (μM)	15	9.1	8.0	20
Concentration (mg/mL)	2.8	1.9	2.1	1.3
Image acquisition				
Nominal magnification	75,000	105,000	105,000	105,000
Voltage (kV)	300	300	300	300
Electron exposure ($e^-/\text{\AA}^2$)	50	50	50	60
Defocus range (μm)	-1.0 to -2.6	-0.6 to -2.0	-1.0 to -2.4	-0.6 to -2.0
Pixel size (\AA)	0.954	0.86	0.86	0.86
Dose per frame ($e^-/\text{\AA}^2/\text{frame}$)	1.25	1.25	1.25	1.5
Exposure time (s)	4.74	1.42	2.00	1.65
Micrographs acquired	6,752	4,462	23,281	10,825
Data processing				
Symmetry imposed	C2	D4	D7	C1
Particle box size* (pix) (extraction size / fourier cropped size)	360 / 240	600 / 400	800 / 500	320 / 320
Initial particles images	542,670	71,114	147,754	2,755,281
Final particle images	87,716	31,656	20,357	318,464
Map resolution (\AA) FSC threshold 0.143	2.9	3.1	3.0	2.9
Map resolution range (\AA)	2.4-6.1	2.9-6.6	3.0-6.0	2.4-5.7
B-factor (\AA^2)	93	109	93	99
Modeling				
Non-hydrogen atoms	13,206	102,240	235,872	4,318
RNA nucleotides	616	4,784	11,032	202
Average Q-score	0.66	0.62	0.66	0.68
Model resolution (\AA) FSC threshold 0.5	3.04	3.38	3.18	3.11
Molprobity score	1.73	1.69	1.80	1.71
R.m.s.d. bond lengths (\AA)	0.015	0.003	0.003	0.003
R.m.s.d. bond angles ($^\circ$)	1.583	0.620	0.620	0.703
RNA sugar pucker outliers	9 (1%)	0 (0%)	0 (0%)	0 (0%)
RNA suite outliers	41 (7%)	24 (4%)	654 (6%)	16 (8%)
RNA suiteness	0.684	0.804	0.787	0.735
Clashscore	0.20	0.10	0.42	0.15

*Additional box size used during processing.

Supplementary information

**Naturally ornate RNA-only complexes
revealed by cryo-EM**

In the format provided by the
authors and unedited

Supplementary information for:

Naturally ornate RNA-only complexes revealed by cryo-EM

Rachael C. Kretsch¹, Yuan Wu², Svetlana A. Shabalina³, Hyunbin Lee⁴, Grace Nye⁵, Eugene V. Koonin³, Alex Gao^{1,4}, Wah Chiu^{1,5,6,7,*}, Rhiju Das^{1,2,4,*}

¹ Biophysics Program, Stanford University, Stanford, CA USA

² Howard Hughes Medical Institute, Stanford University, Stanford, CA USA

³ Computational Biology Branch, Division of Intramural Research, National Library of Medicine, National Institutes of Health, Bethesda, MD USA

⁴ Department of Biochemistry, Stanford University School of Medicine, Stanford, CA USA

⁵ Division of CryoEM and Bioimaging, SSRL, SLAC National Accelerator Laboratory, Menlo Park, CA USA

⁶ Department of Bioengineering and James Clark Center, Stanford University, Stanford, CA USA

⁷ Department of Microbiology and Immunology, Stanford University, Stanford, CA USA

*Correspondence to Wah Chiu (wahc@stanford.edu) and Rhiju Das (rhiju@stanford.edu).

Page 3-4	Supplemental Text 1: Description of <i>raiA</i> RNA motif tertiary structure.
Page 4	Supplemental Text 2: Comparative and covariation analysis of ornate large RNAs.
Page 5-7	Supplemental Table 1: Intramolecular motifs identified.
Page 8-9	Supplemental Table 2: Intermolecular interactions, base pairs, base stacks, and motifs.
Provided as additional file	Supplemental Table 3: Summary of the nucleotide-nucleotide covariations identified in the OLE, ROOL, and GOLLD alignments.
Page 10	Supplemental Table 4: Sequences used in this study.
Provided as additional file	Supplemental File 1: The multiple sequence alignment of OLE in Stockholm format.
Provided as additional file	Supplemental File 2: The multiple sequence alignment of ROOL in Stockholm format.
Provided as additional file	Supplemental File 3: The multiple sequence alignment of GOLLD in Stockholm format.
Provided as additional file	Supplemental Video 1: Structure of OLE dimer. The overall topology of the OLE dimer is displayed with the regions in Fig. 1 highlighted.
Provided as additional file	Supplemental Video 2: Structure of ROOL nanocage. The overall topology of the ROOL nanocage is displayed with the regions in Fig. 2 highlighted.
Provided as additional file	Supplemental Video 3: Structure of GOLLD nanocage. The overall topology of the GOLLD nanocage is displayed with the regions in Fig. 3 highlighted.
Provided as additional file	Source Data for Extended Data Figure 7: The Bioanalyzer tables for OLE, ROOL, GOLLD, and <i>raiA</i> contain the intensity for the ladder used and the sample. The band location, in nucleotide units was calculated by linearly fitting the inverse migration time of the 7 highest intensity peaks to the known length of the ladder components (25, 200, 500, 1000, 2000, 4000, and 6000 nt), as is standard. The dynamic light scattering (DLS) table show the average intensity across 10 acquisitions and 2 replicates. Each column represents the DLS trace at a given temperature: data was collected at 25°C and temperature was increased by 10°C until 75°C. The mass photometry table lists every event recorded for each sample. The size of the RNA was calibrated using Millenium RNA Ladder (Ambion AM7150).

Supplemental Text 1: Description of *raiA* RNA motif tertiary structure.

The structure of another family of long RNAs, the *raiA* motif, from *Clostridium acetobutylicum* ATCC 824, was determined. Under similar concentrations and conditions to the other long RNAs, the *raiA* motif was found to be a well-ordered monomer (**Extended Data Fig. 4a**). The secondary structure of the *raiA* RNA is very similar to the structure proposed by Soares et al¹ based on covariance analysis and inline probing experiments (**Extended Data Fig. 4b**). The few differences are trivial, four previously proposed base pairs were not found in our structure: U19-G185, U29-A43, U50-G178, G55-U173, and an additional four base pairs were identified not previously annotated, mostly within the region for which there was no prior data from inline probing assays: G24-U50, G101-C169, C103-G168, G120-C162. Notably, although U29-A43 was proposed to be a base-pair, both bases were sites of spontaneous scission supporting our model with no base pair.

The linkers J1a/1b, J1b/1c and J3a/3b all interact in a complex tertiary interaction rigidifying P1 and P3 (**Extended Data Figure 4c**). A44 and A46 in this interaction, are members of a Z-anchor, one of many motifs found throughout the *raiA* motif structure (**Supplemental Table 1**). Around PK2, U82-A182 and G83-C181, our structure shows an interesting interaction where the strand loops around itself to form non-canonical interaction A184-G179 which facilitates the sharp turn of the backbone necessary to form this pseudoknot (**Extended Data Fig. 4d**). The backbone of the 3' strand of PK2 is part of an interesting pocket. The pocket contains two cryo-EM densities of high signal indicating ions (**Extended Data Fig. 4e**). These are 6 Å apart. The pocket is formed by backbone atoms as well as G84, which is observed pointing into the pocket.

There are two major secondary structure variants in the *raiA* motif family; both are consistent with our 3D structure. In ~30% of *raiA* motif sequences, P7 and P8 are not present¹. These stems appear peripheral, wrapping around the core of the RNA. There is weak density in the cryo-EM map indicating the P8-loop may interact with the P5-loop; however this region has poor local resolution, suggesting the interaction is transient or flexible and hence does not play an important stabilizing role for the atomically ordered core. Second, the loop in J1c/3a (19-23) forms a T-loop, into which A183 intercalates (**Extended Data Fig. 4f**). ~40% of *raiA* motifs have an additional stem, P2, between P1c and P3a¹, which would not support a T-loop. We hypothesize that the T-loop interaction is replaced by an A-minor interaction between the P2 minor groove and A183 and/or A184 (>97% sequence conserved¹). Hence, we hypothesize that these secondary structure variants would likely have similar folds to the structure shown herein.

The long stretch of nucleotides, Jpk1/6, was previously identified as having many conserved nucleotides but being unpaired according to covariation and scission experiments¹. We observe low local resolution in this region, supporting the lack of structure and hence flexibility of this region. Interestingly, this region is looped out into solution as opposed to being tucked into the core as might be expected for regions with high sequence conservation. Hence, this sequence does not seem to play a vital role in the structure of this RNA, and the sequence may be an important recognition signal for other nucleic acid or protein partners. Alternatively, it may play an

important structural role in an alternative biologically important conformation of this molecule that may be visited at different stages of this RNA's function, which remains unknown.

Supplemental Text 2: Comparative and covariation analysis of ornate large RNAs.

To gain insight into the biological significance and divergence of protein-independent RNA quaternary ensembles, we analyzed the evolutionary conservation of their intermolecular interactions across distinct families of large RNA molecules. This included comparative and covariance analyses of the intermolecular contacts identified in this study and the structural elements in their neighborhoods. The analysis of OLE, ROOL, and GOLLD showed that, despite limited sequence identity, they all exhibit extensive folding and some interactions that are long range in sequence, as confirmed by covariation statistics (**Extended Data Fig. 8a-c** and **Supplemental Files 1-3**). Covariance analysis across the three large RNA families revealed similar patterns in the distribution of covariance scores (**Extended Data Fig. 8a-c** and **Supplemental Table 3**), with significant enrichment of covariance scores at sites close to identified intermolecular interactions. Notably, most of the consensus positions lacked significant covariation scores due to the requirement of some variability to be able to observe sufficient mutations to detect covariation. Instead, pairs with significant covariation scores ($E < 0.05$) tended to be located in variable regions of hairpin stems, while their loops remained highly conserved (**Supplemental Table 3**). This pattern underscores the role of covariation in maintaining structural stability despite sequence variability.

We observed that some hairpins with variable stems maintained highly conserved loops, particularly those containing A-runs, for example, B1 interaction in OLE (P4/L4) and B6 interaction (L14/P14) in GOLLD (see **Fig. 1d, 3p** and **Extended Data Fig. 8d-e**). In these examples, the four or three highly conserved A positions (**Extended Data Fig. 8d-e**), flanked by variable stems with a number of significant covariations (**Supplemental Table 3**), are indeed involved in A-minor interactions in our cryo-EM structures (B1 in **Fig. 1d**, B6 in **Fig. 3p**, respectively).

Supplemental Table 1: Intramolecular motifs identified.

RNA	Motif type	Residues involved
<i>raiA</i>	Z-anchor	A43, A44, A45, A46
	Kink-turn	116-124, 159-171, 106-110
	Ribose zipper	A57, A58, C90, U91
		U85, G86, G154, A155
		A99, A100, C103, C104
	U-Turn	G20, A21, A22
		A66, A67, A68
		U98, A99, A100
		U140, A141, A142
	UA-Handle	U18, U19, A23, A186
		G177, G178, A81, C51
	T-Loop	U18, U19, G20, A21, A22, A23, A186
	Intercalated T-Loop	U18, U19, G20, A21, A22, A23, A186, A183
	GNRA-Tetraloop	U140, A141, A142, U143
	A-minor	A57, G34, C90
		A100, G168, C103
Platform	G128, U129, A130, G154	
Loop-E-Submotif	C153, G154, A155, U129, A130, G131	
	U132, G133, A134, U150, A151, G152	
Bulged-G	C153, G154, A155, G128, U129, A130, G131	
	U132, G133, A134, G149, U150, A151, G152	
OLE	Z-anchor	A359, A360, G361, A362
	Kink-turn	67-74, 123-127, 173-179
	Ribose zipper	C165, C166, A345, A346
	U-Turn	U44, A45, A46
		G201, A202, G203
		U245, A246, A247
	UA-Handle	U272, U273, A358, A362
	GNRA-Tetraloop	G201, A202, G203, A204
	A-minor	A57, G269, C365
		A345, C166, G135
	Platform	C99, A101, G102, A95
	Tandem-GA-sheared	G71, A72, G177, A178

RNA	Motif type	Residues involved		
ROOL	Z-anchor	U209, G210, U211, G212		
	Kink-turn	127-135, 165-172		
	Ribose zipper		C20, C21, A163, A164	
			A257, A258, C288, C289	
			U61, G62, A366, A367	
	U-Turn		U37, A38, A39	
			G101, A102, U103	
			U306, G307, G308	
			G364, A365, A366	
			U563, U564, G565	
	UA-Handle		G24, U25, A18, C49	
			C373, U374, A356, G358	
			G500, U501, A510, C513	
	A-minor		U507, U508, A503, A228	
			A163, C21, G52	
			A257, C289, G327	
			A185, U244, A383	
			A341, U107, A98	
			A366, G62, C202	
			A404, G406, C352	
		Platform		G397, U398, A399, G489
				G116, U117, A118, G88
		Loop-E-Submotif		G87, G88, A89, U117, A118, C119
			G84, A85, A86, A120, A121, U122	
			U488, G489, A490, U398, A399, G400	
	Bulged-G		G87, G88, A89, G116, U117, A118, C119	
			U488, G489, A490, G397, U398, A399, G400	
Tandem-GA Sheared		G261, A262, G284, A285		
GA-minor		A378, G379, C248, G249, C250		

RNA	Motif type	Residues involved
GOLLD	Ribose zipper	U17, C18, A22, A23
		A44, A45, G76, U77
		G107, C108, A278, A279
		A218, A219, C393, U394
		C292, U293, A203, A204
	U-Turn	G33, A34, A35
		G201, G202, A203
		U251, A252, A253
		U339, G340, G341
		G664, A665, A666
		U725, G726, G727
	UA-Handle	C580, U581, A572, G743
		U754, U755, A803, A805
	A-minor	A26, G97, U77
		A219, G287, C393
		A551, G428, C475
		A665, G686, C648
		A736, C732, G722
		A785, G460, U434
		A792, C461, G433
	Platform	U791, A792, A793, G784
		G694, U695, A696, G639
		G132, U133, A134, G156
		A295, C297, A298, U323
	Loop-E-Submotif	A155, G156, A157, U133, A134, U135
		U137, G138, A139, U151, A152, G153
		C638, G639, A640, U695, A696, A697
Bulged-G	A155, G156, A157, G132, U133, A134, U135	
	C638, G639, A640, G694, U695, A696, A697	

Supplemental Table 2: Intermolecular interactions, base pairs, base stacks, and motifs.

RNA	Intermolecular Interaction	Motif, pair or stack type	Residues involved
OLE	B1	Base pair: W-S cis	A:A45 B:A48
		Base pair: W-W cis	A:A47 B:A47
		Stack: Parallel	A:A35 B:A45
	B2	Base pair: W-W cis	A:G97 B:C150
		Base pair: W-W cis	A:C98 B:G149
		Base pair: W-W cis	A:C100 B:G148
		Stack: Antiparallel	A:G97 B:A151
		Stack: Antiparallel	A:A101 B:G148
		Stack: Antiparallel	A:U147 B:A202
	B3	Base pair: W-W cis	A:G315 B:C318
		Base pair: W-W cis	A:G316 B:C317
		Base pair: W-W cis	A:C317 B:G316
Base pair: W-W cis		A:C318 B:G315	
ROOL	B1	Base pair: W-W cis	1:A38 2:U468
		Base pair: W-W cis	1:A39 2:U467
		Base pair: W-W cis	1:U40 2:A466
		Base pair: W-W cis	1:C41 2:G465
		Stack: Antiparallel	1:A38 2:G485
		Stack: Antiparallel	1:C41 2:A466
	B2	Base pair: S-W cis	1:G62 2:A366
	B3	Base pair: W-W cis	1:C263 2:G272
	B4	Base pair: W-S trans	1:A71 2:C334
		Base pair: W-W cis	1:G307 2:C335
		Base pair: W-W cis	1:G308 2:C334
		Base pair: W-W cis	1:U309 2:A333
		Base pair: W-W cis	1:A310 2:U332
		Stack: Antiparallel	1:G307 2:A336
		Stack: Antiparallel	1:G308 2:C335
		Stack: Antiparallel	1:A310 2:A333
		A-minor	1:A71, 2:C334, 1:G308
	B5	Stack: Antiparallel	1:A505 1':A505
B6	Base pair: W-W cis	1:G565 2':C566	
	Stack: Antiparallel	1:U564 2':A568	
	Stack: Antiparallel	1:C566 2':C566	

RNA	Intermolecular Interaction	Motif, pair or stack type	Residues involved
GOLLD	B1	Base pair: S-W cis	1:A589 2:A34
		Base pair: S-W trans	1:G604 2:A35
		A-minor	2:A34, 1:A589, 1:U603
		A-minor	2:A35, 1:G604, 1:U588
	B2	Base pair: W-W cis	1:G140 2:C146
		Base pair: S-H cis	1:U150 2:C146
		Stack: Antiparallel	1:G141 2:C146
	B3	Base pair: S-W cis	1:U293 2:A203
		Base pair: S-W trans	1:G388 2:A204
		A-minor	2:A203, 1:U293, 1:A387
	B4	Base pair: W-S cis	1:A409 2:U251
		Base pair: S-W cis	1:A415 2:A252
		Base pair: W-W cis	1:U414 2:A253
		Base pair: W-W cis	1:A413 2:U254
		Base pair: W-W cis	1:G412 2:C255
		Base pair: W-W cis	1:G411 2:C256
		Base pair: W-W cis	1:G410 2:C257
		Stack: Antiparallel	1:A409 2:C250
		Stack: Antiparallel	1:A413 2:C255
		Stack: Antiparallel	1:G412 2:C256
		Stack: Antiparallel	1:G411 2:C257
		Stack: Antiparallel	1:G410 2:U258
		A-minor	2:A252, 1:A415, 1:U408
	B5	Base pair: W-W cis	1:A372 2:G340
		Base pair: W-W cis	1:C371 2:G341
		Base pair: W-W cis	1:A369 2:U342
		Stack: Antiparallel	1:A372 2:G341
	B6	Base pair: S-W cis	1:A682 2:A447
		Base pair: S-W trans	1:G670 2:A448
		A-minor	2:A447, 1:A682, 1:U669
		A-minor	2:A448, 1:G670, 1:U681
	B7	Base pair: W-W cis	1:C618 2:G546
		Base pair: W-W cis	1:U617 2:A547
		Stack: Antiparallel	1:C618 2:A547
B8	Base pair: W-W cis	1:G656 7':C657	
B9	Base pair: W-W cis	1:G675 7':C676	
	Stack: Antiparallel	1:G675 7':A677	
B10	Base pair: S-W cis	1:U647 1':A763	
	Base pair: W-S trans	1:A764 1':A688	
	A-minor	1:A763, 1':U647, 1':A687	
	A-minor	1:A764, 1':A688, 1':U646	
	GA-minor	2:A145, 2:C146, 1:G140, 1:G141, 1:G142	

Supplemental Table 4: Sequences used in this study.

Genes (<i>T7 promoter bolded, priming site for amplification underlined, added or mutated nucleotides highlighted</i>)	
raiA (<i>Clostridium acetobutylicum</i> ATCC 824 AE001437.1/ 2985433-2985231 ^{1,2})	<u>TTCTAATACGACTCACTATA</u> GGTTAAAGTTAGGTTTGTGGTTGAAAGTCGATGCCAGTCGACAGGCAAAACG ATCCACGTAAGTTAAACAAGTTTTAATGAGCATGGTGGGCTTAGAAGTAAGTCCGCCGCTTAGGCG AGAGTATTAGTAGTGAGAGGTAATCCGGGTAGCGAACTCCAGCAGGCGAGTGTGGGTCAAAGACC <u>AGGTCACCTAACTTA</u>
OLE (<i>Clostridium acetobutylicum</i> ATCC 824 AE001437.1/ 2182120-2181516 "env-38" ^{2,3}) *	<u>TTCTAATACGACTCACTATA</u> GGTGCAGTATTCTAGTCAGGGAATGCTTTTTGAAAGCGGGGCTAAAAAT CCGCTAAAGGGCACATCGATGAAGTTCCTGGTGTGGCTTAGAATGCCAGCTTGGGCTTGTGCTGGG AGTTAAAAAGCTGGGGCACTCGCAATGGCATGCGACAAATGACCCTACTTTTGTGGAGGCCAATTTATTG TATATTGAGAGGATATTTCAATATACGAAATTTGGGGTAAACCTGCAATGTGGTGTAAAAGCTATGTGCAG TGTAGCCTGCCTTGTGGTATGGGAGAGGAGATAAACAAGTCAAAAAATTTAGGCCCTAAGTTTTTGTGA CTATTGAACCTCGAAACCTATGTTGCAAAAGAGGCTAAGAAAGCATCTAACTTTGAGGAAACCTCTAG ACTGTTTTGGTAAATGAGGATTGCAAGTCCGGACTTAGTGGCAATTCAGTCTGAAAGTGGCAACACTTC AGCTCGGATATTAAAGGAAACCGCTATATGGCGACGTATAGTATTCTGTGGGAAAGCCTACTGAACCT ATGCCGTAAGATTACTTATTTTGTACCACATTG GC
ROOL (RUMENNODE4169944_96001/22320-21654 "env-120" ^{6,7}) †	<u>TTCTAATACGACTCACTATA</u> GAAGTGTATAGACATAGCCTTGTGTATGACTGTCTAATCAACAGTGCA AGGAATTAGTTGTGCTCTCAAAAGGAGTTATGTGAAGGAAAGATTAAATGGATACATTTAATTAAGTACA ATTACTACAAAATCTATAATGACTGGGTAGTCAACCCGTAAGAAGGATGAGTTAGTAGGATATAAAT ATCTTACAGTCAACTGATGTGCAAGTTTATAAGTAGACGTAGTGTGAAAGACTTATCGCTAACGCAATA GAGCTTATCTCGAAAGGATAAAGAAACCAAGAAGATATACAAATTTGGTAACTTGTATAAGCCCTTGGTA ATACAAAATAAGTTAGTTCCTCATGATGTG CGTGA AACATGACTATAAGATAAGTTGAATTCGATAGTAG CCCAAGAGGAATCAAATGAGTAAACAAGTATTATAGGTAATATTATGTTTTCTATCTGAATGATGTT GCCTTGAAAAGGTGGTGGTCTGTGGTAAAAATTAATCCATCGTGGTTTTGGATTTTCTTCGGAA AATTTGGTCAAGAAGCGTGGAGTTGCTAACCATTTGCTCAGACTTGTATCGCCAGTAGCACGATAACGATA TTTATGAATGTTGAAGGTGAATAAAGCCTATGATTTGTGAACATTCC
GOLLD (JCVI SCAF 1096627009515/1239-410 env-38 ^{6,7})	<u>TTCTAATACGACTCACTATA</u> GGATGGCCGCTATGGTCGTCAAATAAATACGAAAGTAGGATAAAGTGG AATAGTAAGGTTTAGCTATTCTGCGGTTTCATCTTGTAAAGTGGAGCTTGAGTAGGCAAGTGAGATATC AATTTTTCTTAGTATGTGAGGGTAACACTTTAGCAGAGGAGAAAAGTTGACTGGGCTTTGACAGAGATTC AGTTGAGTTCGGAAGGCAATAAGAATAACTCATAGAATTCATGCGAGAGGTAGGGAACATATCCCTCC CTAATAATCGCGGTATTCAATATGATAGTTTACTTAACATCAAATTTGGGAAACCCCTTTAAGATAAGATAG TGTACGGGTGGTCCGTTATTATCCTTTGTATAGAGTGTACCAGCACCTATCGATGAAGTAACTTGGAAAT ATGGTGGTAGGGATACCACATCGGGTAGTTAAGTATTCTGTGCTTCAAAGGCGCGGAGCTTGTGATAG ACCGCTACCTGAAACCACTGCAAAACCAACTTATATTGCAATTTAGGATTGCAAAAATCATAAAAATCAA AATAAGAAAAAGTGTCTATCAGTTTTGACCAGAAAGGTGTCTACATAGTAATGAGTTGTTCAATGCCACAA ACACTCTCAAGGTGAATGTGATTTCTTCGAAAGTTTCTAACCCGCAAGTGTGAATCTGCTCGCCAGGG TAGAAGAAAATAAGTAAGAAGAGAGTAGTTAAAACTCAAAGAGTGGTCTACTTAATAACCCGACATTGG TTGTTACATCTCAAAAGGATGTGGAAACAAAGGGAATAGATAATCCTTTTAAAGACAAC TA CAATATGGT GTATTCTAGCCTI
Primers	
Forward for all	TTCTAATACGACTCACTATAGG
raiA_R	TAAGTTAGTTGACCTGGCTTTGA
OLE_R	GGCAATGTGGTAACAAAAAAGTAAAT
ROOL_R	GGAAATGTTCACAATCATAGGCTT
GOLLD_R	AGGCTGAGAATACACCATATTTGTA
Sequences For AlphaFold 3 Prediction	
OapA	MKSLGKSNWTLAIAVITLVLAIFSIIVTFMLSGVSWAVGMIVLLIVLVGILFDITGVAATAADEKPF HAMASERLKGAKQAVAITRNDRFANFCNDVIGDISGIVSGTAATYVVIQLALQLGYGENSGQFALS FTSVVAALTVGGKAIGKTLAIEHSTAIYQVGVFYFLEEKMNINFLPNGKKKKEKNSRK
OapC	MSYEKVSQATDLVIGTKQLKALEQEEVLEVVIAKDAEPRVNVKVEAMSVKQIPIIYVDSMKKLKAKG IDVGAATVALKK
RpsU	MAETRVKRNESIDAALRRFKRSLKEGTLAEVRKRKHYEKPSVRRKKKSEARKRKF

* Due to DNA synthesis problems, OLE P1 was trimmed to a 3 base-pair stem

WT P1 : AUUAUUAUGUGGU--ACCACAUUGAUUUUAU

(((((...((((((((((--))))))))))))))...)))

P1 of sequence used herein: GGU--ACCACAUUGCC

(((--)))

† Due to DNA synthesis limitations, a base pair was mutated from U-A to G-C

References

1. Soares, L. W., King, C. G., Fernando, C. M., Roth, A. & Breaker, R. R. Genetic disruption of the bacterial *raiA* motif noncoding RNA causes defects in sporulation and aggregation. *Proc. Natl. Acad. Sci. U. S. A.* **121**, e2318008121 (2024).
2. Nölling, J. *et al.* Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J. Bacteriol.* **183**, 4823–4838 (2001).
3. Puerta-Fernandez, E., Barrick, J. E., Roth, A. & Breaker, R. R. Identification of a large noncoding RNA in extremophilic eubacteria. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 19490–19495 (2006).
4. Weinberg, Z. *et al.* Detection of 224 candidate structured RNAs by comparative analysis of specific subsets of intergenic regions. *Nucleic Acids Res.* **45**, 10811–10823 (2017).
5. Hess, M. *et al.* Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* **331**, 463–467 (2011).
6. Weinberg, Z., Perreault, J., Meyer, M. M. & Breaker, R. R. Exceptional structured noncoding RNAs revealed by bacterial metagenome analysis. *Nature* **462**, 656–659 (2009).
7. Yooseph, S. *et al.* The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. *PLoS Biol.* **5**, e16 (2007).