STRUCTURAL BIOLOGY

Cryo-EM reveals mechanisms of natural RNA multivalency

Liu Wang[1,2]†, Jiahao Xie[3]†, Tao Gong[1]†, Hao Wu[4,5]†, Yifan Tu[6]†, Xin Peng[6]†, Sitong Shang[6]†, Xinyu Jia[1]†, Haiyun Ma[1]†, Jian Zou[1]†, Sheng Xu[4,5]†, Xin Zheng[1,2]†, Dong Zhang[6]†, Yang Liu[7]†, Chong Zhang[1], Yongbo Luo[1], Zirui Huang[1], Bin Shao[1], Binwu Ying[7], Yu Cheng[8], Yingqiang Guo[9], Ying Lai[1], Dingming Huang[1,2], Jianquan Liu[6], Yuquan Wei[1], Siqi Sun[4,5]*, Xuedong Zhou[1,2]*, Zhaoming Su[1]*

Homo-oligomerization of biological macromolecules leads to functional assemblies that are critical to understanding various cellular processes. However, RNA quaternary structures have rarely been reported. Comparative genomics analysis has identified RNA families containing hundreds of sequences that adopt conserved secondary structures and likely fold into complex three-dimensional structures. In this study, we used cryo–electron microscopy (cryo-EM) to determine structures from four RNA families, including ARRPOF and OLE forming dimers and ROOL and GOLLD forming hexameric, octameric, and dodecameric nanostructures, at 2.6- to 4.6-angstrom resolutions. These homo-oligomeric assemblies reveal a plethora of structural motifs that contribute to RNA multivalency, including kissing-loop, palindromic base-pairing, A-stacking, metal ion coordination, pseudoknot, and minor-groove interactions. These results provide the molecular basis of intermolecular interactions driving RNA multivalency with potential functional relevance.

Symmetric assembly in macromolecular complexes frequently underlies the core of biological processes and functions (*1*, *2*). Homo-oligomerization of multiple identical subunits through a network of intermolecular interactions forms higher-order quaternary structures that acquire distinctive features such as increased stability, allosteric regulation, and multivalent and cooperative interactions (*2*). These homomers have been predicted to occur in 50% of proteins (*2*, *3*), forming structures including chaperonins and viral capsids in dihedral and cubic symmetry with a large central cavity as well as cytoskeleton filaments in helical symmetry (*2*). Impairment of homo-oligomerization is implicated in disease (*4*, *5*).

RNAs fold into intricate tertiary structures to play critical regulatory roles in numerous biological processes (*6*). However, although protein homomers are ubiquitous across all life domains, the oligomeric quaternary structures of natural RNA sequences have not been well characterized (*7*, *8*). The predominantly observed RNA multimers are homodimers, enabled primarily by kissing-loop (KL) and complementary (including palindromic) base-pairing interactions and occasionally by tertiary interactions such as minor-groove, A-stacking, and pseudoknot (PK) interactions (*8*). A recent study has also reported the dimerization of large group II introns through such interactions (*9*). Although these structural motifs have been widely used for higher-order assembly in RNA nanotechnology (*10*), the only natural RNA homomer structure consisting of more than two protomers is the φ29 prohead RNA (*11*), which was shown to assemble into a four-membered ring in crystalline conditions (*12*) and a five-membered ring in the context of the φ29 phage packaging motor (*13*). The leading factor that limits our understanding of RNA quaternary interactions is the challenge posed to RNA three-dimensional structure determination, highlighted by orders of magnitude fewer depositions of RNA structures in the Protein Data Bank (PDB) compared with proteins (*14*). This paucity largely prompts the insufficient accuracy of RNA structure predictions in contrast to that of proteins (*15*, *16*).

The RNA families database (Rfam) has classified thousands of RNA families on the basis of multiple sequence alignment (MSA) and covariance models of conserved secondary structures (*17*). Aside from families with well-characterized functions and structures, more RNA families are constantly being discovered by comparative genomic analysis, yet their structures and functions remain mostly unexplored (*18*, *19*). Here, we focus on four RNA families with uncharacterized structures: ARRPOF (area required for replication in a plasmid of *Fusobacterium*), OLE (ornate, large, and extremophilic), ROOL (rumen-originated, ornate, large), and GOLLD (giant, ornate, lake- and Lactobacillales-derived). ARRPOF is predicted to perform functions related to plasmid replication and regulation (*19*), OLE participates in bacterial cell growth under extreme conditions such as cold or alcohol stress (*20*), and ROOL and GOLLD are frequently associated with prophages and reside close to tRNAs either in the chromosomes or plasmids (*21*).

Single-particle cryo–electron microscopy (cryo-EM) has recently emerged as a powerful method to study protein-free RNA structures with heterogeneous conformations and compositions (*22*). We selected sequences from bacteria with characterized functions in ARRPOF, OLE, ROOL, and GOLLD RNA families, which are predicted to form complex tertiary structures, whose functions could be further explored. Cryo-EM structures determined at 2.6- to 4.6-Å resolutions reveal quaternary assemblies of all these RNAs through conserved intermolecular interfaces and elucidate molecular mechanisms of RNA multivalency.

ARRPOF forms a homodimer and adopts two conformations by swapping the 3′ domain

ARRPOF has been shown by comparative genomics analysis to adopt a conserved secondary structure (*19*). The sequence is present in the pKH9 plasmid from *Fusobacterium nucleatum* (*Fnu*) identified in an earlier study, which showed that the truncated plasmids, while still retaining the ARRPOF sequence but lacking a predicted *rep* (replication) gene, could still replicate (*23*), suggesting that ARRPOF might play a role in plasmid replication (*19*). We carried out in vitro transcription, denatured gel purification, and subsequent refolding to prepare RNAs for cryo-EM structure determination (*22*). Cryo-EM reconstructions revealed that the 255-nucleotide (255-nt) *Fnu* ARRPOF assembles into a homodimer and adopts two alternative conformations with swapped 3′ domains, determined at 3.9- and 4.0-Å resolutions (Fig. 1, A and B; fig. S1; and table S1).

ARRPOF adopts eight consecutive paired stems (P1 to P8) from 5′ to 3′ end that form a three-way junction (3WJ) and a four-way junction (4WJ) (Fig. 1C and fig. S2). In conformation 1, the 5′ end starts with two intermolecular complementary base pairings, P9 and P10, with L6 (L denotes loop) and the 5′ leader of the other protomer (Fig. 1, D and E). P10 is followed by the 3WJ and two intramolecular PKs

[1]The State Key Laboratory of Biotherapy, National Clinical Research Center for Geriatrics, West China Hospital; The State Key Laboratory of Oral Diseases, National Clinical Research Center for Oral Diseases, National Center for Stomatology, West China Hospital of Stomatology, Sichuan University, Chengdu, China. [2]Department of Cariology and Endodontics, West China Hospital of Stomatology, Sichuan University, Chengdu, China. [3]Mingle Scope (Chengdu), Chengdu, China. [4]Research Institute of Intelligent Complex Systems, Fudan University, Shanghai, China. [5]Shanghai Artificial Intelligence Laboratory, Shanghai, China. [6]The Key Laboratory for Bio-resources and Eco-environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu, China. [7]Department of Laboratory Medicine, West China Hospital, Sichuan University, Chengdu, China. [8]Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong, China. [9]Cardiovascular Surgery Research Laboratory, Department of Cardiovascular Surgery, West China Hospital, Sichuan University, Chengdu, China.
***Corresponding author. Email: zsu@wchscu.cn (Z.M.S.); zhouxd@scu.edu.cn (X.D.Z.); siqisun@fudan.edu.cn (S.Q.S.)** †These authors contributed equally to this work.

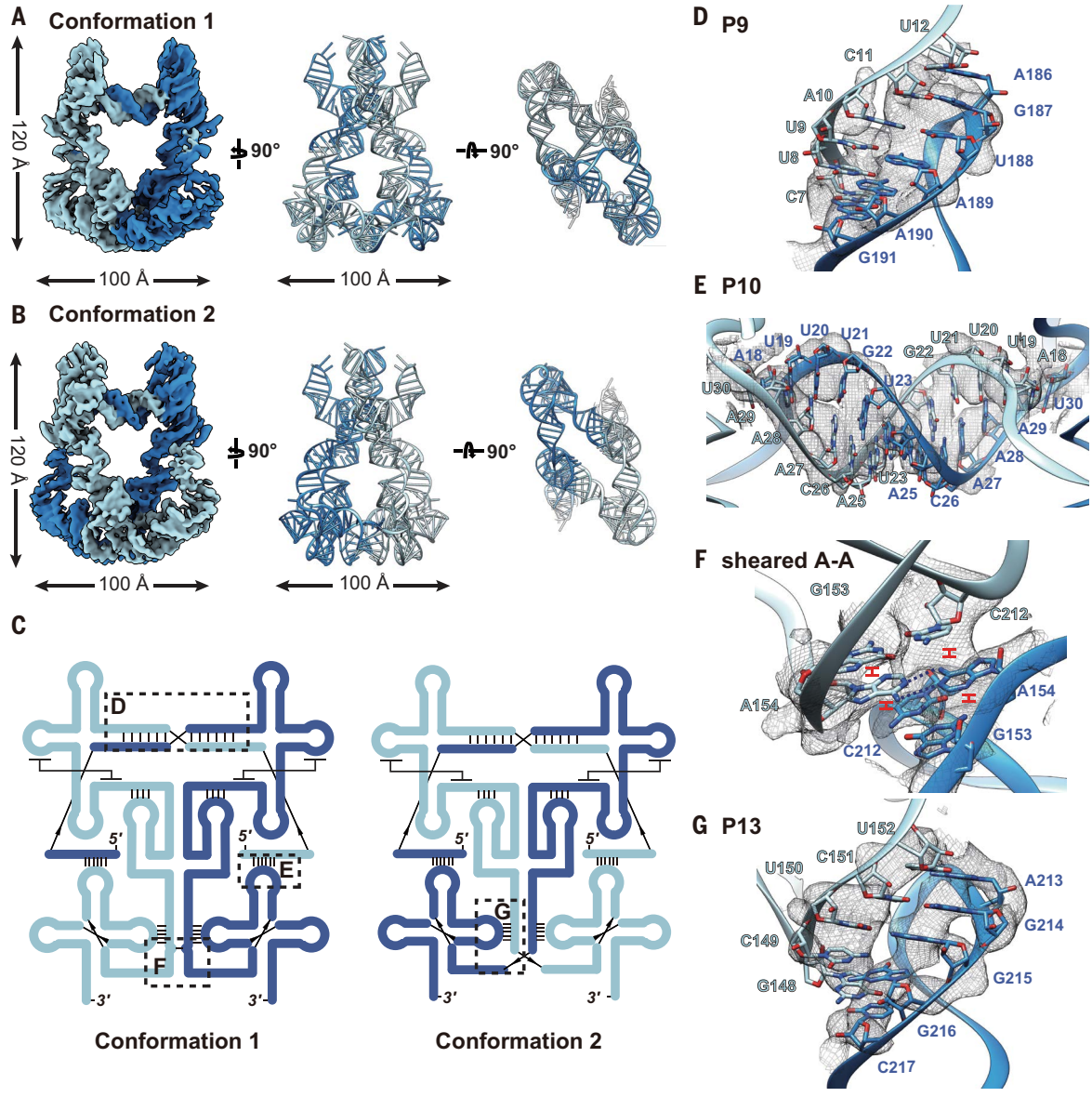

Fig. 1. Cryo-EM structures reveal alternative conformations and intermolecular interfaces of ARRPOF dimer. (**A** and **B**) Overall cryo-EM density maps and models at 5.0σ threshold of conformation 1 (A) and conformation 2 (B) with the 3′ domain P5-P8 swapped to the other side. (**C**) Projected secondary structures in cartoon of conformations 1 and 2 with designated tertiary interactions. (**D** to **G**) Cryo-EM density maps and models of intermolecular interfaces of P9 in both conformations (D), P10 (E), and sheared A-A base pair stacked by P13 from both protomers (F) in conformation 1 and junction-loop interaction P13 in conformation 2 (G), all at 2.0σ threshold.

formed between J3/4 (J denotes a junction between paired regions) and L2 and L4, namely P11 and P12, before reaching P4 (fig. S2, A to C). Following P4 is the intramolecular PK P13 between J4/5 and L7 that stacks on an intermolecular sheared A-A pair (Fig. 1F), after which the remaining 3′ domain consisting of the 4WJ swaps to the opposite side in conformation 2 (movie S1). All tertiary interactions have been found by comparative genomics analysis to be conserved (*19*), including the intermolecular palindromic P10 observed in the cryo-EM structures (fig. S2A). The sheared A-A pair is not conserved, implying that those sequences lacking this A-A pair and adjacent stacking interactions may only adopt conformation 2 (Fig. 1C), in which P9 is intramolecular and P13 is intermolecular (Fig. 1G).

OLE assembles into a homodimer for protein interactions

A recent surge of notable structural and functional studies has demonstrated that OLE RNA binds to proteins and forms ribonucleoprotein (RNP) complexes to participate in a wide range of fundamental processes in bacteria such as cell growth, stress response, and replication (*20*). OLE colocalizes with OapA (OLE-associated protein A) at the cell membrane, whose coding region is located in tandem of *ole* in the gene cluster (*24*). OapB and OapC also directly bind to OLE (*25*, *26*), and impaired OLE RNP function makes bacterial cells more sensitive to unconventional conditions such as cold temperatures, alcohols (*27*), and magnesium ions (*28*).

We determined the 2.6-Å-resolution cryo-EM structure of the OLE (578-nt) dimer from *Clostridium botulinum* (*Cbo*) (Fig. 2A, fig. S3, and table S1), a pathogen known to produce a lethal neurotoxin that causes botulism and muscle paralysis, which has been used in cosmetic dermatology to induce facial muscle paralysis (*29*). The *Cbo* OLE contains 15 conserved paired stems (P1 to P15) (*24*), but the cryo-EM density only accounts for P3 to P9.3, including the preformed protein binding sites for OapA (*24*), OapC (*26*), and ribosomal protein RpsU (*20*) (Fig. 2A).

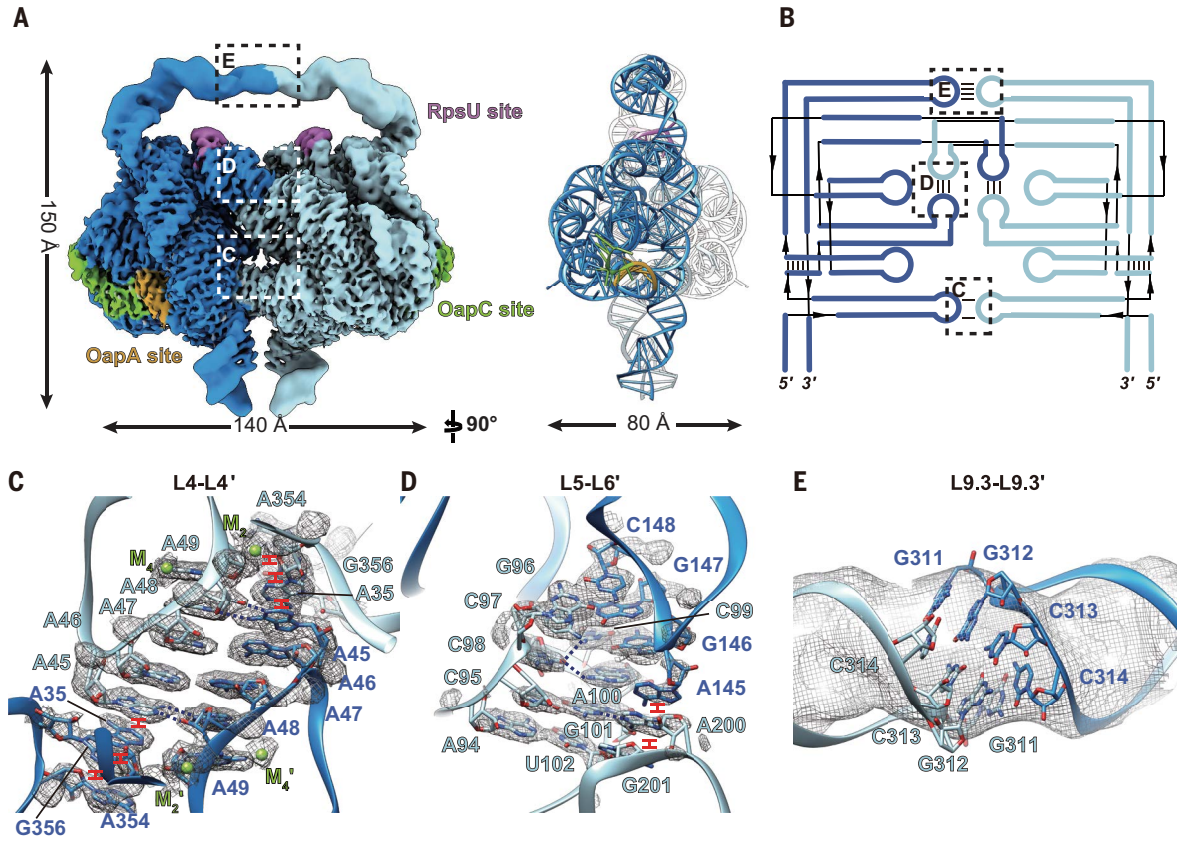

Fig. 2. Cryo-EM structures reveal intermolecular A-stacking and KL interactions of OLE dimer. (**A**) Cryo-EM density map and model of OLE dimer at 3.0σ threshold. (**B**) Projected secondary structure in cartoon with designated intermolecular interactions. (**C** to **E**) Cryo-EM density maps and models of intermolecular interfaces of A-stacking interactions between L4 and L4′ at 1.0σ threshold (C), KL interaction between L5 and L6′ (D), and KL interaction between L9.3 and L9.3′ at 2.0σ threshold (E). Blue dashed lines indicate hydrogen bonds of noncanonical interactions, and red labels indicate stacking.

The P3–P9.3 domain forms an intramolecular PK between J4/5 and J8/9 (fig. S2, D and E) and assembles into a homodimer through three intermolecular interactions (Fig. 2B and movie S2), including the continuous A-stacking and sheared A-A pairs between L4 in each protomer (Fig. 2C) and two KL interactions between L5 and L6 (Fig. 2D) and L9.3 in each protomer (Fig. 2E). Comparative genomics analysis indicates that all intermolecular interactions are highly conserved in primary sequence, suggesting that the dimeric form is likely critical for OLE function. The remaining stems were not resolved owing to high flexibility, which could potentially be stabilized upon binding to OapB as previously reported (*25*).

Cryo-EM reveals multimerization interfaces and different assembly mechanisms of ROOL RNAs

ROOL RNAs are usually associated with bacteriophages and are located close to tRNA genes (*21*), which have been primarily identified in *Lactobacillus* and *Enterococcus*. Recent studies have reported that *Enterococcus faecalis* (*Efa*) strains carrying ROOL in their chromosomes have been found in red tilapia and the international space station (*30*, *31*), whereas megaplasmid-encoded ROOL in *Lactobacillus salivarius* (*Lsa*) had unusually abundant expression levels that exceeded 16*S* ribosomal RNA (rRNA) in the late stationary phase in some strains (*32*).

We observed cage-like particles under cryo-EM directly after in vitro transcription of both *Efa* and *Lsa* ROOL RNAs (figs. S4 and S5), indicating that these RNAs could form quaternary assemblies under native cotranscriptional folding conditions (*22*). Cryo-EM reconstructions of *Efa* ROOL (580-nt) yielded monomer, homotetramer, and dihedral homooctamer at 3.1-, 4.7-, and 3.8-Å resolutions, respectively, whereas the cryo-EM structure of *Lsa* ROOL (526-nt) showed a dihedral homohexamer at 3.1-Å resolution (Fig. 3, A and B; figs. S4 and S5; and table S1). The density of *Efa* ROOL monomer allowed modeling of roughly half of the RNA sequence, suggesting that the remaining unresolved regions are highly dynamic. These dynamic regions are all present in the tetramer and octamer, stabilized by intermolecular interactions in the quaternary structures, allowing modeling of the entire ROOL RNA (fig. S6A). The *Lsa* ROOL monomer from the homohexamer adopts almost identical architecture (fig. S6B) and was modeled by substitution of the *Efa* ROOL sequence.

Both Efa and Lsa ROOL RNAs adopt conserved secondary structures containing 16 paired stems (P1 to P16), five multiway junctions, and four intramolecular tertiary interactions (fig. S6, A and B). The four intramolecular interactions include a tetraloop/tetraloop receptor (TL/TLR) interaction between P2 and L5 (fig. S6C), an A-stacking interaction between L6 and P8 (fig. S6D), and two PKs, between J9/12 and L13 and L13 and P15 (fig. S6, E and F). Five additional intermolecular interactions are observed in both ROOL assemblies (fig. S6, G to K, and movies S3 and S4). Three interfaces are within the same plane of the dihedral assemblies, including two 3-bp (base pair) loop-bulge interactions, between L2 and L13′ and J11/9 and L11′ (the prime symbol denotes the structural motif from the adjacent protomer in the same plane), and one minor-groove interaction, between L12 and P3′ (Fig. 3, E to G). The dihedral complexes are assembled through two interplanar interactions, a minor-groove A-stacking interaction between L14 and P7″ (the double prime symbol denotes the structural motif from the protomer in the opposite plane) and a 4-bp KL between L15 and L15″ (Fig. 3, H and I). Whereas all intermolecular base pairs are conserved, as shown by covariation analysis, interplanar interfaces

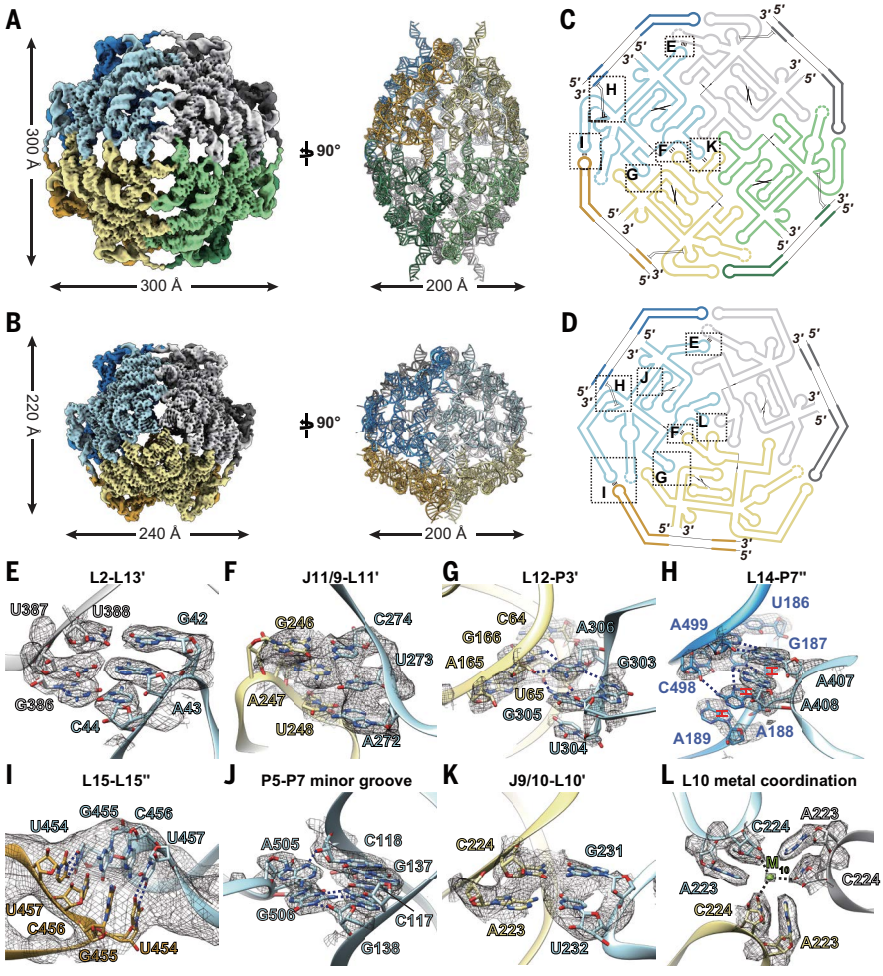

Fig. 3. Cryo-EM structures and intermolecular interfaces of ROOL hexamer and octamer. (**A** and **B**) Overall cryo-EM density maps and architectures of *Efa* ROOL octamer (A) and *Lsa* ROOL hexamer (B) at 4.0σ threshold. (**C** and **D**) Projected secondary structures with designated intermolecular interfaces of *Efa* ROOL octamer (C) and *Lsa* ROOL hexamer (D). (**E** to **I**) Intermolecular interactions in *Lsa* ROOL hexamer, including KL interactions between L2 and L13′ (E) and J11/9 and L11′ at 4.0σ threshold (F), TL/TLR interaction between L12 and P3′ at 1.0σ threshold (G), A-stacking interaction between L14 and P7″ (H), and interplanar KL interaction between L15 and L15″ at 2.0σ threshold (I). (**J**) Intramolecular minor-groove interaction between P5 and P7 observed only in *Lsa* ROOL at 2.0σ threshold. (**K**) KL interaction between J9/10 and L10′ at 2.0σ threshold reveals intraplanar assembly mechanism of *Efa* ROOL. (**L**) Metal ion coordination of three cytidines (C224) in L10s of different protomers at 0.8σ threshold reveals intraplanar assembly mechanism of *Lsa* ROOL. Blue dashed lines indicate hydrogen bonds of noncanonical interactions, and black dashed line indicates metal coordination.

in *Efa* ROOL contain fewer interactions compared with *Lsa* ROOL (fig. S6, J and K), which may result in dissociation of the upper and lower halves of the octamer into tetramers under cryo-EM conditions (fig. S4). Notably, *Lsa* ROOL adopts an extra purine minor-groove interaction between P5 and P7 that is not present in the *Efa* ROOL (Fig. 3J), resulting in a more-compact P2 and P15 that affects assembly stoichiometry (fig. S7, A to D). Moreover, P10 adopts different structures in *Efa* and *Lsa* ROOL RNAs, leading to drastically different intraplanar assembly mechanisms (Fig. 3, C and D, and fig. S7E). The *Efa* ROOL uses four 2-bp KL interactions between J9/10 and L10′ at the apex along the main symmetry axis to assemble the intraplanar tetramer (Fig. 3K), whereas the *Lsa* ROOL presents an interaction right on the main symmetry axis, enabled by metal ion coordination with C224 bases in L10 from all three protomers for homotrimeric assembly (Fig. 3L and table S2). Similar divalent metal ion coordination of cytosines has been previously reported in crystal packing (*33*), and metal ion coordination could also facilitate protein self-assembly (*34*).

Cryo-EM structures of GOLLD RNAs reveal dynamic multimerization

Akin to ROOL, GOLLD RNAs are frequently found near tRNA genes, and their expression levels correlate with bacteriophage packaging in certain bacterial strains (*35*). GOLLD RNAs consist of variable 5′ domains and highly conserved 3′ domains. We prepared two RNA constructs of the full-length (FL, 700-nt) GOLLD and 3′ domain (375-nt) of GOLLD from *Streptococcus agalactiae* (*Sag*) by in vitro transcription, after which we also observed cotranscriptionally folded cage-like particles under cryo-EM. Reconstruction of the GOLLD 3′ domain showed both dihedral homodecamer and dodecamer at 3.6- and 6.1-Å resolutions, with the homodecamer as the major class that contains 12 times more particles than the dodecamer (fig. S8 and table S1). In contrast, the FL GOLLD was reconstructed only as the dihedral homododecamer at 4.6-Å resolution (Fig. 4A and fig. S9).

The FL GOLLD is the largest and most complex RNA structure resolved in this work, containing 25 paired stems (P2–P8 and P10–P27, numbered according to the conserved secondary structure) (*35*), two 4WJs, three 3WJs, and five intramolecular tertiary interactions (fig. S10A). These include two KLs, between L8b and L11 and J24/26 and L27 (P25) (fig. S10, B and C), and three minor-groove interactions, between L5 and P3, P7 and P10, and L27 and P14 (fig. S10, D to F). Additional intermolecular interactions facilitate RNA multimerization (Fig. 4B and movie S5), including four minor-groove interactions, between L3 and P19′, P8 and P11′, L8a and P8a′, and L14 and P21′ (Fig. 4, C to F), and two KL interactions, between L10 and L12′ and L15 and L20′ (intermolecular P16) (Fig. 4, G and H). Together these six intermolecular interfaces assemble the hexameric plane. Another two interplanar interactions, a TL/TLR interaction between L24 and P21″ and a KL between L21 and L21″ (Fig. 4, I and J), complete the dihedral dodecameric nanocage. Intermolecular interfaces involving base pairings show either covariation or sequence conservation (fig. S10A). While the GOLLD 3′ domain multimerization dynamics is likely enabled by the 4WJ formed by P19 to P22 (fig. S11A), the presence of 5′ domain in the FL GOLLD ensures homogeneous assembly through intraplanar interfaces and shape complementarity (fig. S11, B and C), which also allows various assemblies and stoichiometries facilitated by the variable 5′ domains of different GOLLD RNAs (*36*).

Discussion

RNA multivalency has been increasingly recognized to play important roles in biological processes and diseases, such as triggering innate immune responses and driving the formation of pathological condensates and aggregations (*37*). However, the molecular basis behind RNA multimerization remains largely unknown (*8*). Almost all experimentally determined RNA multimers are homodimers with total intermolecular interface areas ranging from 134 to 3143 Å^2 (Fig. 5A). Those with larger interfaces usually contain long complementary base-pairing or

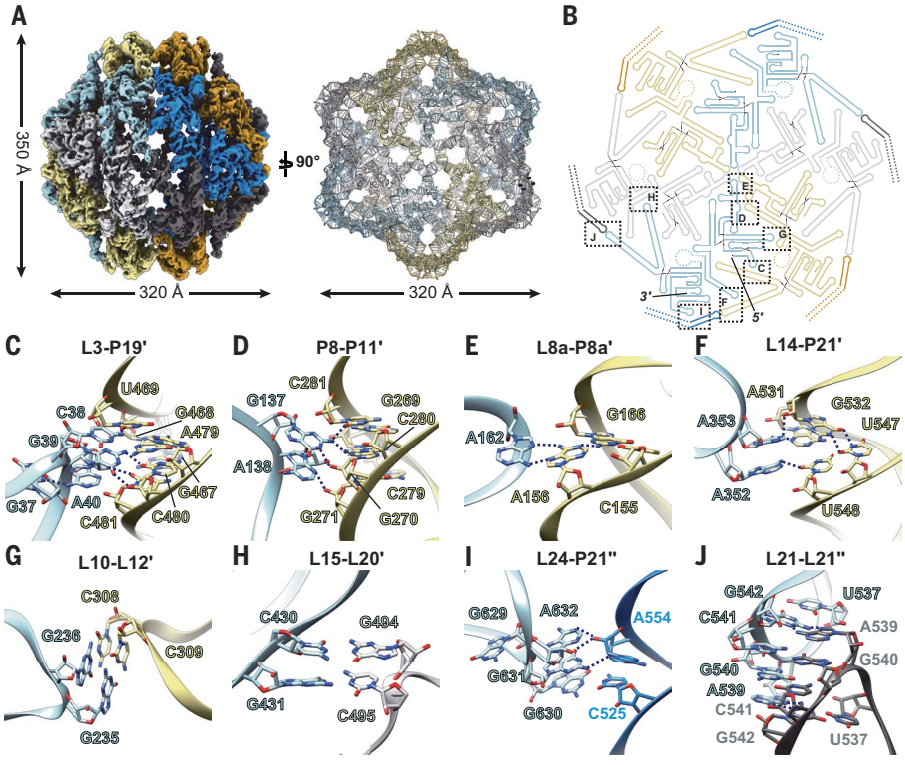

Fig. 4. Cryo-EM structures and intermolecular interfaces of FL GOLLD dodecamer. (**A**) Cryo-EM density map and model of GOLLD dodecamer at 2.0σ threshold. (**B**) Projected secondary structure in cartoon with designated intermolecular interactions. (**C** to **F**) Minor-groove interactions with the adjacent protomer between L3 and P19′ (C), P8 and P11′ (D), L8a and P8a′ (E), and L14 and P21′ (F). (**G** and **H**) KL interactions between L10 and L12′ (G) and L15 and L20′ (intermolecular P16) (H). (**I**) Interplanar TL/TLR interaction between L24 and P21″ in the protomer from the different plane. (**J**) Interplanar KL interaction between L21 and L21″. Blue dashed lines indicate hydrogen bonds of noncanonical interactions.

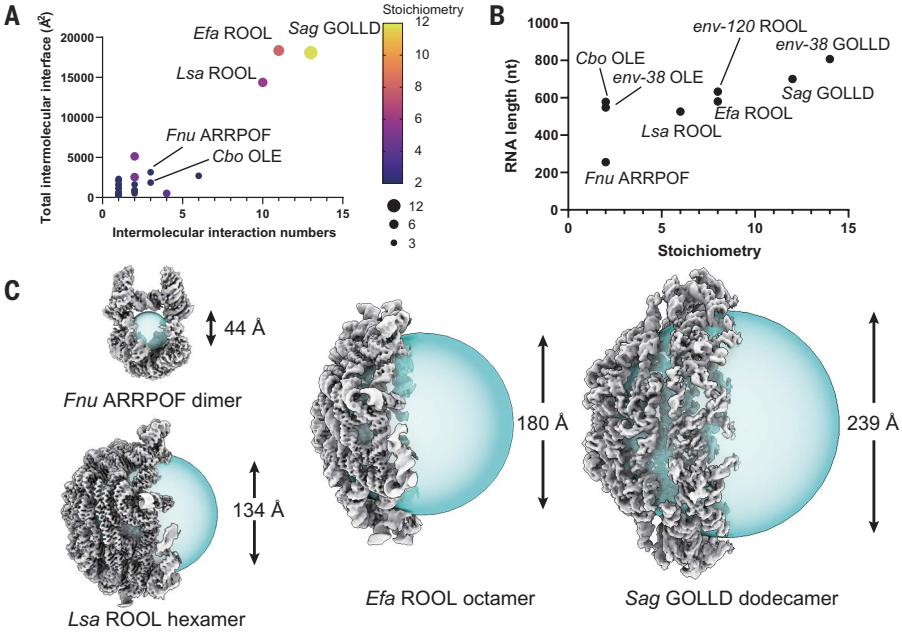

Fig. 5. Correlations between RNA multivalency and numbers of intermolecular interfaces, total interface areas, and RNA lengths. (**A**) Correlation between numbers of intermolecular interfaces, interface areas, and multimer stoichiometry. (**B**) Correlation between RNA length and multimer stoichiometry. (**C**) Central cavities of ARRPOF, ROOL, and GOLLD assemblies.

A-stacking interactions, such as ARRPOF (3143 Å^2) and OLE (1855 Å^2). Proteins also primarily form homodimers, because dimers normally contain fewer intermolecular interfaces and likely emerged earlier in evolution, whereas higher-order multimers require accumulation of mutations for more intermolecular interactions and thus evolved later (Fig. 5A) (*1*). Homomers with dihedral symmetry (D_n) are more prevalent than those with cyclic symmetry (C_n), as observed in this work and a recent independent study of the same RNA families (*36*), because there are potentially multiple evolutionary paths to D_n symmetry, in contrast to the single pathway toward C_n symmetry (*1*).

Within the ROOL and GOLLD families, the dihedral symmetry of the RNA assemblies varies with different sequences and correlates with RNA length; longer RNAs tend to assemble into multimers with higher stoichiometry (Fig. 5B). The 526-nt *Lsa* ROOL assembles into a dihedral homohexamer, whereas the *Efa* and *env-120* ROOL of 580- and 659-nt, respectively, form homooctamers (*36*). The 700-nt *Sag* GOLLD forms a 12-membered dodecamer, and the 833-nt *env-38* GOLLD forms a 14-membered tetradecamer (*36*). The more conserved *Sag* GOLLD 3′ domain of 375 nt could form decameric and dodecameric assemblies and even more heterogeneous states, as observed in cryo-EM two-dimensional averages (fig. S8). Variations in critical structural motifs could lead to multimerization dynamics and drastically different assembly mechanisms within the same RNA family. Superposition of the *Efa* and *Lsa* ROOL monomers reveals that the additional P5-P7 minor-groove interaction only observed in the *Lsa* ROOL induces more-compact P2 and P15 (fig. S7, A to D), whereas the missing P10 facilitates the intraplanar metal ion coordination of three *Lsa* ROOL protomers (Fig. 3L and fig. S7E), which together affect the stoichiometry and symmetry of RNA multivalent assembly. For GOLLD, the 4WJ consisting of P19 to P22 in the 3′ domain allow heterogeneous assemblies (figs. S8 and S11A), whereas its 5′ domain in the FL GOLLD promotes homogeneous quaternary assembly through intermolecular interactions and shape complementarity (fig. S11, B and C).

ROOL and GOLLD assemblies have demonstrated that RNA sequences selected from different bacterial strains can adopt different multimerization states, which may also affect their functions. Although comparative genomics analysis has provided a collection of sequences that adopt conserved secondary structures, selecting proper sequences for structural and functional analyses remains difficult for RNA families such as GOLLD, whose variable 5′ domain can adopt alternative secondary structures in different bacteria (*35*). Alignment of the *Sag* and *env-38* GOLLD sequences to the previously reported conserved secondary structure revealed that *env-38*

GOLLD matched perfectly with the presence of all structural motifs, whereas *Sag* GOLLD contained extra P8a and P8b but lacked part of P7 and P10 as well as the entirety of P9, P11a to P11c, and P21b. Nonetheless, *Sag* GOLLD could still form the quaternary assembly, albeit with different stoichiometry (fig. S12, A to C). Another GOLLD sequence from *Flavobacteriales bacterium* ALC-1 (*Fba*), also retrieved from the Rfam and previous covariation analysis (*17*, *35*), failed to form a rigid tertiary structure that could be experimentally determined, despite exhaustive RNA and cryo-EM sample optimizations (*22*). Comparison between *Fba* and *Sag* GOLLD showed that *Fba* GOLLD is missing L8b-L11, P22, and part of P27 (fig. S12, D and E). While these differences in secondary structures are not apparently related to tertiary structure formations, we explored whether a pretrained contrastive learning language model, fine-tuned by sequences from structured RNA families (CRAFTS), could extract structural features from RNA sequences and retrieve their likelihood of forming tertiary structures within their respective families (fig. S13 and supplementary text, note 1). Although the ARRPOF sequences were clustered on the basis of variations in P4 and P6 and extra nucleotides in P10 (fig. S14, A to C), which may result in steric clash of extended P4 and disruption of P9 with truncated P6 (fig. S14D), the remaining OLE, ROOL, and GOLLD families showed inconclusive sequence consensus variations that require further validations (fig. S15). These observations suggest that general application of such a model across RNA families remains challenging and is primarily limited by the scarcity of RNA structures available in the PDB.

Our cryo-EM structures of ARRPOF, ROOL, and GOLLD RNAs revealed cage-like nanostructures with various central cavities ranging from 44 to 239 Å in diameter (Fig. 5C), which could form under extremely low concentrations (*36*). This might imply that natural RNA homomers can encapsulate macromolecules such as double-stranded DNA during transcription, tRNAs, and even molecules as large as rRNAs. These homomers may function as delivery vehicles or chaperones, possibly representing relics of "the RNA world" (*38*, *39*). In modern organisms, the roles of these homomers have largely been taken over by proteins. Preliminary functional results indicate that ARRPOF dimerization may play critical roles in transcription regulation of the fructose-associated pathway that modulates *Fnu* autoaggregation (supplementary text, note 2, and fig. S16) (*40*). The OLE RNP complex has been recently proposed to participate in multiple pathways essential in cellular processes, equivalent to TOR complexes in eukaryotes (*20*). ROOL and GOLLD are bacterial noncoding RNAs larger than 500 nt that frequently reside in tRNA clusters, whereas other structured RNAs of similar sizes have been characterized to participate in catalytic reactions (*21*, *41*, *42*). Although their possible catalytic functions as large ribozymes await to be tested, the highly symmetrical assembly may enhance RNA stability and potentially allow their movements within cells and across organisms and species as mobile elements (*43*, *44*).

REFERENCES AND NOTES

1. E. D. Levy, E. Boeri Erba, C. V. Robinson, S. A. Teichmann, *Nature* **453**, 1262–1265 (2008).
2. D. S. Goodsell, A. J. Olson, *Annu. Rev. Biophys. Biomol. Struct.* **29**, 105–153 (2000).
3. H. Schweke *et al.*, *Cell* **187**, 999–1010.e15 (2024).
4. C. M. Dobson, *Nature* **426**, 884–890 (2003).
5. T. A. Thibaudeau, R. T. Anderson, D. M. Smith, *Nat. Commun.* **9**, 1097 (2018).
6. L. Statello, C.-J. Guo, L.-L. Chen, M. Huarte, *Nat. Rev. Mol. Cell Biol.* **22**, 96–118 (2021).
7. C. P. Jones, A. R. Ferré-D'Amaré, *Trends Biochem. Sci.* **40**, 211–220 (2015).
8. C. Bou-Nader, J. Zhang, *Molecules* **25**, 2881 (2020).
9. Z.-X. Liu *et al.*, *Science* **383**, eadh4859 (2024).
10. H. Ohno, S. Akamine, H. Saito, *Curr. Opin. Biotechnol.* **58**, 53–61 (2019).
11. P. X. Guo, S. Erickson, D. Anderson, *Science* **236**, 690–694 (1987).
12. F. Ding *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 7357–7362 (2011).
13. H. Mao *et al.*, *Cell Rep.* **14**, 2017–2029 (2016).
14. H. Ma, X. Jia, K. Zhang, Z. Su, *Signal Transduct. Target. Ther.* **7**, 58 (2022).
15. R. C. Kretsch *et al.*, *Proteins* **91**, 1600–1615 (2023).
16. F. Bu *et al.*, *Nat. Methods* **22**, 399–411 (2025).
17. N. Ontiveros-Palacios *et al.*, *Nucleic Acids Res.* **53**, D258–D267 (2025).
18. S. Stav *et al.*, *BMC Microbiol.* **19**, 66 (2019).
19. Z. Weinberg *et al.*, *Nucleic Acids Res.* **45**, 10811–10823 (2017).
20. R. R. Breaker, K. A. Harris, S. E. Lyon, F. D. R. Wencker, C. M. Fernando, *Mol. Microbiol.* **120**, 324–340 (2023).
21. K. A. Harris, R. R. Breaker, *Microbiol. Spectr.* **6**, RWR-0005-2017 (2018).
22. X. Chen *et al.*, *Nat. Protoc.* 10.1038/s41596-024-01072-1 (2024).
23. G. Bachrach *et al.*, *Appl. Environ. Microbiol.* **70**, 6957–6962 (2004).
24. E. Puerta-Fernandez, J. E. Barrick, A. Roth, R. R. Breaker, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 19490–19495 (2006).
25. Y. Yang, K. A. Harris, D. L. Widner, R. R. Breaker, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2020393118 (2021).
26. S. E. Lyon, K. A. Harris, N. B. Odzer, S. G. Wilkins, R. R. Breaker, *J. Biol. Chem.* **298**, 102674 (2022).
27. J. G. Wallace, Z. Zhou, R. R. Breaker, *Nucleic Acids Res.* **40**, 6898–6907 (2012).
28. K. A. Harris, N. B. Odzer, R. R. Breaker, *Mol. Microbiol.* **112**, 1552–1563 (2019).
29. B. M. Lund, M. W. Peck, in *Guide to Foodborne Pathogens*, R. G. Labbé, S. García, Eds. (Wiley, 2013), pp. 91–111.
30. R. Rizkiantino *et al.*, *J. Surv. Fish. Sci.* **7**, 27–42 (2020).
31. N. C. Bryan *et al.*, *Front. Microbiol.* **11**, 515319 (2021).
32. F. J. Cousin *et al.*, *Microb. Genom.* **3**, e000126 (2017).
33. N. Marino, D. Armentano, C. Zanchini, G. De Munno, *CrystEngComm* **16**, 8286–8296 (2014).
34. E. N. Salgado, R. J. Radford, F. A. Tezcan, *Acc. Chem. Res.* **43**, 661–672 (2010).
35. Z. Weinberg, J. Perreault, M. M. Meyer, R. R. Breaker, *Nature* **462**, 656–659 (2009).
36. R. C. Kretsch *et al.*, bioRxiv 2024.12.08.627333 [Preprint] (2024); https://doi.org/10.1101/2024.12.08.627333.
37. G. M. Wadsworth *et al.*, *Mol. Cell* **84**, 3692–3705 (2024).
38. W. Gilbert, *Nature* **319**, 618 (1986).
39. T. R. Cech, *Cold Spring Harb. Perspect. Biol.* **4**, a006742 (2012).
40. X. Zheng *et al.*, *Sci. Adv.* **10**, eado0016 (2024).
41. L. Wang *et al.*, *Nat. Struct. Mol. Biol.* 10.1038/s41594-025-01484-x (2025).
42. A. M. Pyle, *Annu. Rev. Biophys.* **45**, 183–205 (2016).
43. X. Chen, O. Rechavi, *Nat. Rev. Mol. Cell Biol.* **23**, 185–203 (2022).
44. A. M. Jose, *Genesis* **53**, 395–416 (2015).
45. H. Wu *et al.*, Code for CRAFTS model, Zenodo (2025); https://doi.org/10.5281/zenodo.14966520.

ACKNOWLEDGMENTS

We thank members from J. Bujnicki's and S. Glatt's groups for constructive discussions under the 1000 RNAs initiative and S. Y. Huang for RNA modeling suggestions. Cryo-EM data were collected on Can Cong at SKLB West China Cryo-EM Center and processed at SKLB Duyu High Performance Computing Center at West China Hospital. This work was partially supported by the Shanghai Artificial Intelligence Laboratory (S.Q.S.). **Funding:** National Key Research and Development Program of China grant 2022YFC2303700 (Z.M.S.); Natural Science Foundation of China grant 32222040 (Z.M.S.); Natural Science Foundation of China grant 32270120 (X.Z.); and the 1.3.5 Project for Disciplines Excellence of West China Hospital grant ZYYC21006 (Z.M.S.).
Author contributions: Conceptualization: Z.M.S.; RNA and cryo-EM grids preparation: L.W., Y.F.T., X.P., X.Y.J., D.Z., Y.Liu, Z.R.H.; Cryo-EM data collection and processing: L.W., Y.F.T., X.P., X.Y.J., H.Y.M., D.Z., Y.Liu, Y.B.L., Z.R.H.; Generating RNA atomic coordinates: L.W., J.H.X., X.P., S.T.S., X.Y.J., J.Z., Z.M.S.; Bacteria assays: T.G., X.Z., X.D.Z.; Language models: H.W., S.X., Y.C., S.Q.S.; Methodology: L.W., J.H.X., T.G., H.W., C.Z., S.X., Y.Lai, X.Z., B.S., Y.C., S.Q.S., X.D.Z., Z.M.S.; Investigation: L.W., J.H.X., T.G., H.W., X.Z., J.Q.L., S.Q.S., X.D.Z., Z.M.S.; Visualization: L.W., T.G., H.W., Y.F.T., X.P., X.Y.J., H.Y.M., J.Z., S.X.; Funding acquisition: Z.M.S., X.D.Z., S.Q.S.; Project administration: Z.M.S.; Supervision: Z.M.S., X.D.Z., S.Q.S., Y.Q.W., D.M.H., J.Q.L., Y.Q.G., B.W.Y.; Writing – original draft: Z.M.S., X.D.Z., S.Q.S., L.W., T.G., H.W.; Writing – review & editing: L.W., J.H.X., TG.,. H.W., Y.F.T., X.P., S.T.S., X.Y.J., H.Y.M., J.Z., S.X., X.Z., D.Z., Y.Liu, C.Z., Y.B.L., Z.R.H., B.S., B.W.Y., Y.C., Y.Q.G., Y.Lai, D.M.H., J.Q.L., Y.Q.W., S.Q.S., X.D.Z., Z.M.S. All authors contributed to the preparation of the manuscript. **Competing interests:** The authors declare that they have no competing interests.
Data and materials availability: The cryo-EM maps and associated atomic coordinate models have been deposited in the wwPDB OneDep System under the following EMD accession codes and PDB IDs: EMD-62489 and 9KPO for ARRPOF dimer conformation 1, EMD-62487 and 9KPH for ARRPOF dimer conformation 2, EMD-62991 and 9LCR for OLE dimer, EMD-60850 and 9ISV for *Efa* ROOL monomer, EMD-61123 and 9J3R for *Efa* ROOL tetramer, EMD-61125 and 9J3T for *Efa* ROOL octamer, EMD-61189 and 9J6Y for *Lsa* ROOL hexamer, EMD-62721 and 9L0R for FL GOLLD dodecamer, EMD-63218 and 9LMF for GOLLD 3′ domain decamer, and EMD-62724 for GOLLD 3′ domain dodecamer. Code for CRAFTS model and benchmark data are available in Zenodo (*45*). Data for transcriptomics analysis of *Fnu* with ARRPOF overexpression have been deposited in the Gene Expression Omnibus (GEO) of NCBI (BioProject accession number PRJNA1207299). All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adv3451
Materials and Methods; Supplementary Text; Figs. S1 to S16; Tables S1 to S5; References (*46*–*80*); MDAR Reproducibility Checklist; Movies S1 to S5; Data S1 to S3

Submitted 17 December 2024; accepted 4 March 2025; published online 13 March 2025
10.1126/science.adv3451