nature nanotechnology

ARTICLES

https://doi.org/10.1038/s41565-019-0517-8

Computational design of three-dimensional RNA structure and function

Joseph D. Yesselman[1], Daniel Eiler[2], Erik D. Carlson[3,4,5], Michael R. Gotrik[1], Anne E. d'Aquino[3,5,6], Alexandra N. Ooms[7], Wipapat Kladwang[1], Paul D. Carlson[8], Xuesong Shi[1], David A. Costantino[2], Daniel Herschlag[1,9,10], Julius B. Lucks[3,4,5,6], Michael C. Jewett[3,4,5,6], Jeffrey S. Kieft[2] and Rhiju Das[1,11]*

RNA nanotechnology seeks to create nanoscale machines by repurposing natural RNA modules. The field is slowed by the current need for human intuition during three-dimensional structural design. Here, we demonstrate that three distinct problems in RNA nanotechnology can be reduced to a pathfinding problem and automatically solved through an algorithm called RNAMake. First, RNAMake discovers highly stable single-chain solutions to the classic problem of aligning a tetraloop and its sequence-distal receptor, with experimental validation from chemical mapping, gel electrophoresis, solution X-ray scattering and crystallography with 2.55 Å resolution. Second, RNAMake automatically generates structured tethers that integrate 16S and 23S ribosomal RNAs into single-chain ribosomal RNAs that remain uncleaved by ribonucleases and assemble onto messenger RNA. Third, RNAMake enables the automated stabilization of small-molecule binding RNAs, with designed tertiary contacts that improve the binding affinity of the ATP aptamer and improve the fluorescence and stability of the Spinach RNA in cell extracts and in living *Escherichia coli* cells.

RNA-based nanotechnology is an emerging field that harnesses RNA's unique structural properties to create new nanostructures and machines[1,2]. Perhaps more so than for other biomolecules, the RNA tertiary structure is composed of discrete and recurring components known as tertiary 'motifs'[3]. Along with the helices that they interconnect, many of these structural motifs appear highly modular; that is, each motif folds into a well-defined three-dimensional (3D) structure in a broad range of contexts[2,4–6]. By exploiting symmetry, motif repetition, expert modelling and computational tools for visualization and modelling flexibility, these motifs have been assembled into new polyhedra, sheets and cargo-carrying nanoparticles for biomedical use[7–10]. Despite these advances, current methods still rely on human intuition and the field cannot yet generate RNAs as sophisticated as natural RNA machines, which are asymmetric, too large to be modelled by 3D RNA structure prediction methods and composed of vast repertoires of distinct interacting motifs, most of which are not yet well characterized[11–13].

We present here a new approach to 3D RNA design based on the recognition that numerous recurring problems in the field can be cast into the same 'pathfinding' problem (Fig. 1). First, a founding problem of RNA nanotechnology involves designing a compact nanostructure that aligns the two parts of the tetraloop/tetraloop receptor (TTR) so that they can form a tertiary contact upon RNA chain folding (Fig. 1a). This task requires finding RNA sequences that interconnect the 5′ and 3′ ends of the tetraloop (orange) to the 3′ and 5′ ends of the tetraloop receptor, respectively (blue in Fig. 1a). The problem has previously been solved through a combination of expert manual modelling and the symmetric assembly of multiple chains[5,14]. In all cases, an important guiding principle—sometimes called RNA architectonics[4]—is to design the intermediate RNA chains so that they form RNA modules previously seen in nature, including both canonical double-stranded helices and non-canonical RNA motifs that twist and translate between two desired helical end-points at the tetraloop and the receptor. We call this design task the 'RNA motif pathfinding problem'. The general complexity of this pathfinding task has prevented the design of asymmetric, single-chain solutions to the TTR stabilization problem.

A second problem is highly analogous to the TTR stabilization problem but is more difficult. Efforts to select engineered ribosomes with messenger RNA decoding, polypeptide synthesis and protein excretion functions optimized for new substrates might be dramatically accelerated through the design of integrated ribosomes. An important step towards this goal involves tethering the two 23S and 16S ribosomal RNAs (rRNAs) of the ribosome into a single RNA strand that supports *Escherichia coli* growth[15–18]. 3D designs for the tether require solving the RNA motif pathfinding problem over >100 Å distances and avoiding steric collisions with the ribosome's RNA and protein components (blue and orange strands in Fig. 1b). Even after the identification of appropriate helix end-points, this difficult design challenge previously took more than a year to solve using in vivo assays based on trial-and-error refinement[16,17] or the ad hoc combination of non-canonical motifs without explicit 3D modelling[15,18].

[1]Department of Biochemistry, Stanford University School of Medicine, Stanford, CA, USA. [2]Department of Biochemistry and Molecular Genetics, University of Colorado Denver School of Medicine, Aurora, CO, USA. [3]Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL, USA. [4]Chemistry of Life Processes Institute, Northwestern University, Evanston, IL, USA. [5]Center for Synthetic Biology, Northwestern University, Evanston, IL, USA. [6]Interdisciplinary Biological Sciences Graduate Program, Northwestern University, Evanston, IL, USA. [7]Department of Cancer Genetics & Genomics, Stanford University School of Medicine, Stanford, CA, USA. [8]Robert F. Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY, USA. [9]Department of Chemistry, Stanford University School of Medicine, Stanford, CA, USA. [10]Stanford ChEM-H (Chemistry, Engineering, and Medicine for Human Health), Stanford University, Stanford, CA, USA. [11]Department of Physics, Stanford University, Stanford, CA, USA. *e-mail: rhiju@stanford.edu

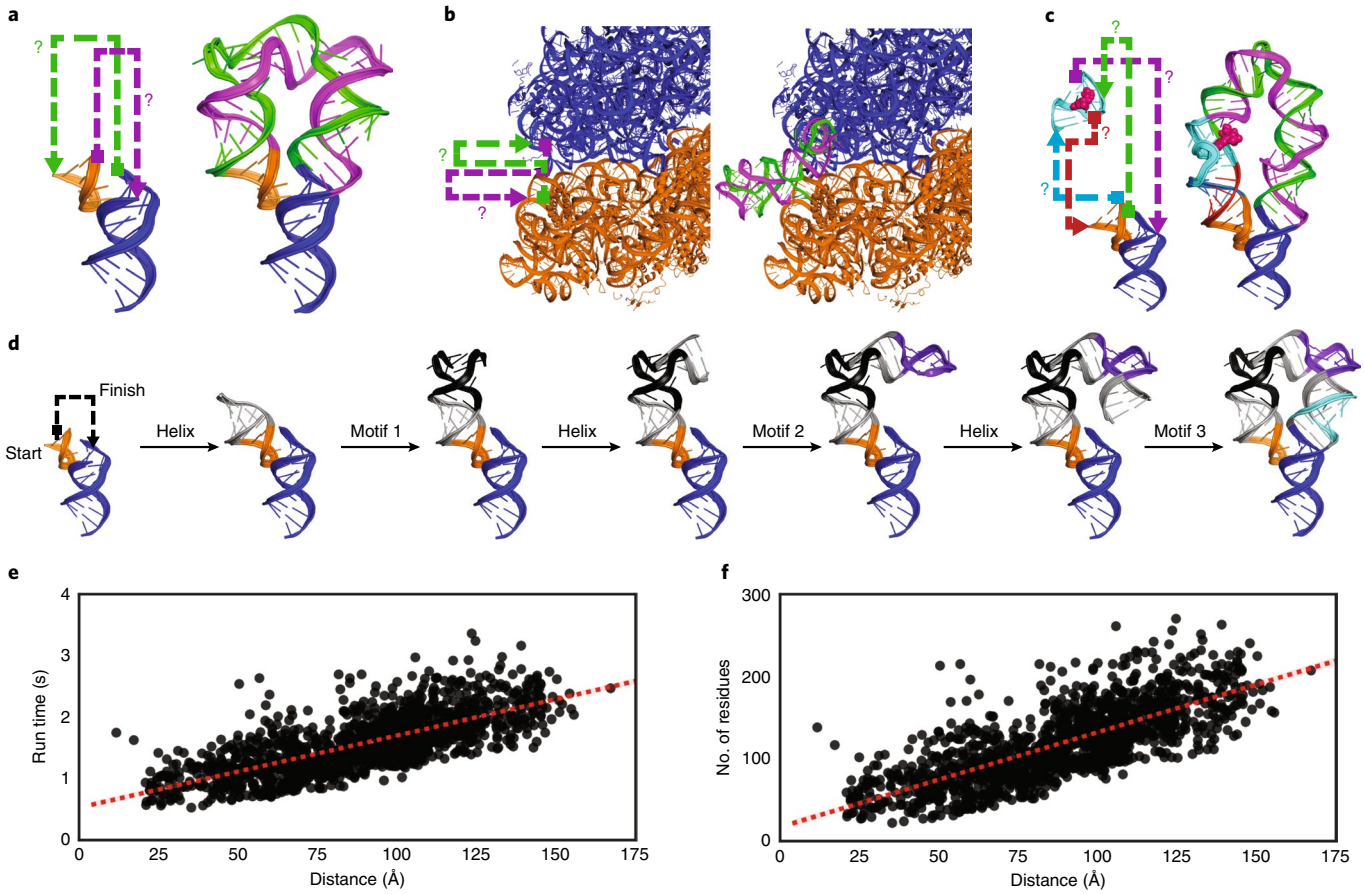

Fig. 1 | Problems in RNA nanotechnology reduced to RNA motif pathfinding problems and solved by RNAMake. a, 'MiniTTRs' require two strands (green and purple) between a tetraloop (orange) and tetraloop receptor (blue). **b**, Tethered ribosomes require two strands (green and purple) to link the small subunit (orange) to the large subunit (blue). **c**, 'Locking' a small-molecule binding aptamer (cyan) by designing four strands (green, purple, teal and red) to a peripheral tertiary contact (orange and blue). Red spheres, ATP molecules. **d**, Demonstration of the RNAMake design algorithm, which builds an RNA path via the successive addition of motifs and helices from a starting base pair to the ending base pair. **e**,**f**, Computational efficiency for RNAMake to design connections between each pair of hairpins on the 50S *E. coli* ribosome. The run time scales linearly with problem size, as measured by the translational distance between helical end-points (**e**) or the number of residues required for segments (**f**) (higher order junctions are utilized in Supplementary Fig. 14).

A third problem involves a more complex instance of two RNA motif pathfinding problems (green, purple, red and teal lines in Fig. 1c). A ubiquitous task in RNA nanotechnology is the selection of 'aptamer' RNAs that sense or carry target small molecules, such as adenosine 5′-triphosphate or fluorophores[19]. Despite recent progress[20,21], improving aptamers requires numerous rounds of tedious selections, with few design tools available to guide consistent improvements. The desired stabilizations might be achieved by peripheral tertiary contacts that extend out of either end of the aptamers and encircle them, bracing them into their functional 3D arrangements (Fig. 1c), analogous to the tertiary contacts that 'lock' natural riboswitch aptamers[22]. However, such rational design has not been carried out due to the difficulty of finding the required four strands that interconnect a given aptamer structure and a tertiary contact.

Here, we present a 3D RNA design algorithm, RNAMake, that solves all three cases of the RNA motif pathfinding problem described above. Gauntlets of structural and functional measurements test that these computationally designed nanostructures, ribosomes and ATP and fluorescent RNA aptamers achieve their design goals, without the need for any further rounds of trial and error.

The RNAMake algorithm and motif library

RNAMake uses a 3D motif library drawn from all unique, publicly deposited crystallographic RNA structures and an efficient algorithm to discover combinations of these motifs and helices that solve the RNA motif pathfinding problem (Methods and Supplementary Table 1). The final set of non-canonical motifs contained 461 unique two-way junctions, 61 higher-order junctions, 290 variable-length hairpins and 89 tertiary contacts. The pathfinding algorithm assembles canonical helical segments that range from 1 to 22 base pairs with these non-canonical structural motifs, step-by-step in a depth-first search (Fig. 1d and Methods). The canonical helical segments are idealized and sequence invariant[23]; after completion of the 3D structural designs, they are filled in with sequences that best match the target secondary structure and minimize alternative secondary structures[24]. Owing to its efficient algorithmic implementation, RNAMake is able to find solutions rapidly; the run time scales linearly with the problem size, and the discovery of exceptionally long double-stranded RNA paths that snake around the entire ribosome takes less than 3 s (run on a Macbook Pro 2016, 2.9 GHz Intel Core i7) (Fig. 1e,f).

RNAMake TTR designs achieve high stability

The problem of creating a well-folded RNA nanostructure was first solved two decades ago by repurposing the well-characterized TTR tertiary contact to bring together two separate RNA chains[5], analogous to the P4–P6 domain of the *Tetrahymena* group I self-splicing intron and other natural functional RNAs. Although later

RNA nanotechnology studies used the TTR module and other structural motifs to design different nanostructures, the original and later designs are all multichain assemblies[25–29]. We chose to test RNAMake on the TTR problem because of the prospect of achieving de novo single-chain solutions to this fundamental problem, which we hypothesized might also help crystallization. We generated 16 diverse single-chain solutions with RNAMake, which we called 'miniTTR' designs.

Standard biochemical and biophysical assays for the RNA structure confirmed folding for the majority of the miniTTR designs. We tested the miniTTR RNAs for the correct secondary structure and tertiary contact formation with single nucleotide resolution chemical mapping (SHAPE (selective 2′-hydroxyl acylation analysed by primer extension) and DMS (dimethyl sulfate) (Supplementary Fig. 1b); Fig. 2a gives DMS at the tetraloop and receptor nucleotides), for compact folds through native gel electrophoresis and mutational analysis (Fig. 2b and Supplementary Fig. 1c) and for tertiary stability through Mg^{2+} binding curves (Fig. 2c,d and Supplementary Table 2). Overall, 11 of the 16 designs passed these experimental screens (details given in Supplementary results and Supplementary Table 3). Several miniTTR constructs required less than 1 mM Mg^{2+} to fold stably, similarly to or better than reported midpoints for natural TTR-containing RNA nanostructures. Indeed, miniTTR 2 and miniTTR 6 exhibited folding stabilities better than that of the P4–P6 RNA in side-by-side assays (Fig. 2c). Furthermore, miniTTR 6 had a much sharper Mg^{2+} dependence than P4–P6 with an apparent Hill coefficient of over ten (Fig. 2c). The stability of the RNAMake designs was particularly notable given that P4–P6 and other natural TTR-containing RNAs are larger than the miniTTR designs and have additional stabilizing tertiary contacts[30–32] and that other attempts to make artificial minimized TTR constructs gave significantly worse stabilities[33].

After the gel-based and chemical mapping tests, we tested whether the RNAMake designs might allow crystallization and thereby enable high-resolution characterization of the structural accuracy of the designs. After small-angle X-ray scattering (SAXS) measurements confirmed a monomeric structure even at high RNA concentrations (>1 μM) (Fig. 2e,f, Supplementary results and Supplementary Fig. 2a,b), we were able to obtain crystals of miniTTR 6 that diffracted at 2.55 Å resolution ($1/\sigma$ of 1.0) (Fig. 2g and Supplementary Table 10). The crystal structure and the RNAMake model agreed with an all-heavy-atom root mean squared deviation (r.m.s.d.) of 4.2 Å, better than the nanometre-scale accuracy typically sought in RNA nanotechnology. The primary discrepancy between our modelled 3D structure and the crystal structure was a single motif, a triple mismatch drawn from the large ribosomal subunit (Fig. 2h, right). This motif formed multiple consecutive non-canonical base pairs with high B factors in our miniTTR 6 crystal instead of the conformation found in the ribosomal structure, which involved flipped-out adenosines (residues O2360–O2363 and O2424–O2426 (Protein Data Bank (PDB) 1S72)) (Fig. 2h, left). Other motifs in the design achieved near-atomic accuracy, including the TTR tertiary contact (r.m.s.d. 0.45 Å (Fig. 2i)), a kink-turn variant drawn from the archaeal 50S ribosomal subunit (r.m.s.d. 2.0 Å (Fig. 2j))[34] and a 'right-angle turn' drawn from a viral internal ribosomal entry site domain (r.m.s.d. 1.28 Å (Fig. 2k))[25].

Automated 3D design of tethered ribosomal subunits

After testing RNAMake's performance in designing compact RNA nanostructures, we evaluated whether it might solve a practical problem involving nanostructures that must traverse long distances (compare Fig. 1a,b). The ribosome is a ribonucleoprotein machine dominated by two extensive RNA subunits, the 16S and 23S rRNAs. In previous work, we constructed a tethered ribosome called Ribo-T in which the large and small subunit rRNAs were connected by an RNA tether to form a single subunit ribosome[17]. In that work, the major bottleneck involved more than a year of numerous trial-and-error iterations to identify RNA tethers that were not cleaved by ribonucleases in vivo when wild-type ribosomes were replaced in the Squires strain of *E. coli*[17]. The Squires strain cells lack genetic rRNA alleles, surviving off plasmids that can be exchanged using positive and negative selections. Early failure rounds that involved ribosomes from our and other studies are shown in Fig. 3a,b and the success with Ribo-T in Fig. 3c. Nevertheless, the current tethers in Ribo-T are unstructured and unlikely to remain stable if other modules are incorporated (Fig. 3c). We hypothesized that an automated design by RNAMake might give structured, chemically stable tethers for this design problem.

RNAMake generated 100 designs (RM-Tethers), which contained either four or five non-canonical structural motifs each (Methods gives the motif selections), to tether the H101 helix on a circularly permuted 23S rRNA to the h44 helix on the 16S rRNA (Fig. 1b and Supplementary Fig. 3b). Of the nine diverse solutions we tested (RM-Tether 1–9), DNA templates for seven could be synthesized, and the transformation of these DNA templates into the Squires strain allowed us to assay whether the RNAMake designs could replace wild-type ribosomes deleted from growing bacteria. One of these seven constructs, RM-Tether 4, led to the viable growth of bacterial colonies. DNA sequencing confirmed that these colonies harboured the correct RM-Tether 4 plasmid; and RNA electrophoresis confirmed the presence of a single dominant RNA species with the same length as Ribo-T, with no detectable products that corresponded to separate 16S or 23S rRNA lengths or other cleavage products (Fig. 3d). Although the growth rate of this strain was low (Supplementary Fig. 4d), we were able to confirm independently that the ribosomes loaded on messenger RNA in vitro, using integrated synthesis, assembly, and translation (iSAT) in ribosome-free S150 extracts[35,36]. Similar to Ribo-T[16], we detected 70S/monosome[37] and polysomes (and no 30S or 50S subunits) by separation of the iSAT-prepared RM-Tether 4 ribosomes on a sucrose gradient (Fig. 3e and Methods). Electrophoresis of the polysome fraction confirmed that it contained an uncleaved rRNA the same size as Ribo-T (Fig. 3f). In addition, SHAPE-Seq mapping on this rRNA confirmed that the RM-Tether 4 can be reverse transcribed from one ribosomal subunit to the other across both strands of the tether and highlights a chemical reactivity profile consistent with the design, with one region of flexibility around the middle junction (Supplementary Fig. 5). Taken together, these data demonstrate that RNAMake-designed ribosomes with structured, chemically stable tethers can replace wild-type ribosomes in vivo and more than one such ribosome can be loaded onto a single message in vitro. RNAMake obviates the repeated rounds of trial and error that were previously required to achieve these design goals.

RNAMake stabilizes small-molecule binding aptamers

As a final series of tests, we evaluated whether RNAMake could solve 3D design problems whose complexity precluded prior progress even with trial and error or large-scale library selections. Small molecules can be bound and sensed by artificially selected RNA aptamers. Unfortunately, these molecules often exhibit weakened binding affinities or instability in biological environments and additional rounds of selection to improve aptamers typically give diminishing returns[38–40]. By expanding RNAMake to allow the design of interconnections between multiple pairs of helices (Fig. 1c), we tested the hypothesis that the computational design of peripheral tertiary contacts might 'lock' these artificial aptamers into their bound conformation even in the absence of a ligand. By reducing the number of alternative structures available in the unbound state, such locking contacts could selectively increase the free energy of the unbound state and thereby improve the free-energy difference to the bound state, and so lead to a better affinity to small molecule targets.

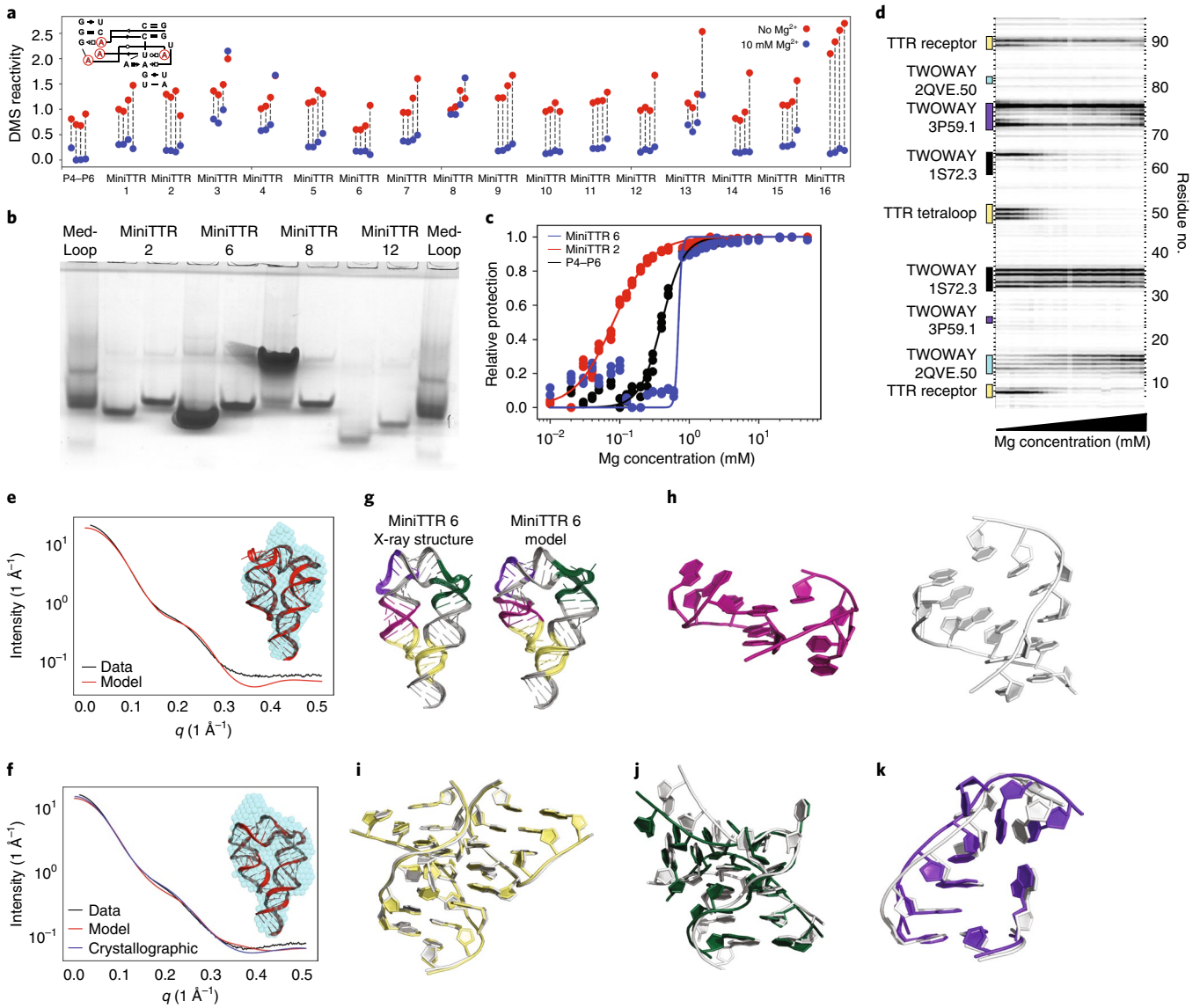

Fig. 2 | Solving the TTR design problem. a, Quantification of DMS reactivity in the absence of Mg^{2+} and with 10 mM Mg^{2+} for RNAMake-designed miniTTR constructs and the P4–P6 domain of the *Tetrahymena* ribozyme as a large natural RNA comparison. Inset: the four adenosines in the TTR (red circles) that undergo protection on Mg^{2+}-dependent folding. **b**, Native gel assays testing whether mutation of the GAAA tetraloop (left lane) to the UUCG mutant (right lane) disrupts the miniTTR tertiary fold and slows its mobility. 'MedLoop' lanes are a control RNA with similar size. Supplementary Fig. 1c gives other constructs. **c**, Quantification of the miniTTR folding stability based on Mg^{2+} binding curves read out by DMS mapping for miniTTR 6, miniTTR 2 and P4–P6. **d**, Raw data from the Mg^{2+} titration (concentrations: 0.00, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 1.00, 1.25, 1.50, 1.75, 2.00, 2.50, 3.00, 3.50, 4.00, 5.00, 7.50, 10.00 mM) of miniTTR 2 highlight the change in DMS reactivity in the TTR and the motifs used in the design. **e**,**f**, SAXS analysis: experimental intensity versus scattering amplitude and low-resolution reconstruction derived from experimental scattering profiles (blue beads, inset) overlaid on the RNAMake-designed model (cartoon, inset) for miniTTR 2 (**e**) and miniTTR 6 (**f**). In **f**, the SAXS prediction from the miniTTR crystal structure is also shown (blue line). **g**–**l**, X-ray crystal structure of miniTTR 6 tests the accuracy of the RNAMake model at atomic resolution: overall RNAMake and X-ray structure (**g**), magnified views of a triple mismatch motif (right) from the ribosome (left) (**h**), TTR (**i**), kink-turn motif (**j**) and right-angle turn (**k**). In **h**–**k**, the crystal structures are white.

First, we sought to stabilize a classic aptamer for adenosine 5′-triphosphate and adenosine 5′-monophosphate (ATP and AMP, respectively), which is in wide use in RNA nanotechnology but whose binding has not been appreciably improved since its discovery in 1993[41–45]. In total, we tested ten ATP aptamers embedded by RNAMake into scaffolds with tetraloop–receptor contacts, which we called ATP-TTR designs (Fig. 4a and Supplementary Fig. 6; Methods describes the modelling of helix flexibility used for these designs). Chemical mapping confirmed that four of these RNAs formed the TTR and also retained their ability to bind to ATP, as assessed by the DMS protection of aptamer nucleotides A13 and A14 (Supplementary Table 4 and Supplementary Fig. 7). Titrations of ATP read out through chemical mapping (Fig. 4d and Supplementary Table 4) showed that three designs achieved better ATP dissociation constants (K_d of 1.5, 4.1 and 1.4 μM) than that of the isolated ATP aptamer under the same conditions (K_d = 16.2 μM), improvements by up to an order of magnitude. Three of the ATP-TTRs gave ligand-free DMS reactivity profiles in the aptamer regions similar to that of the ligand-bound aptamer, which suggests that they preform the structure needed for ATP binding rather than

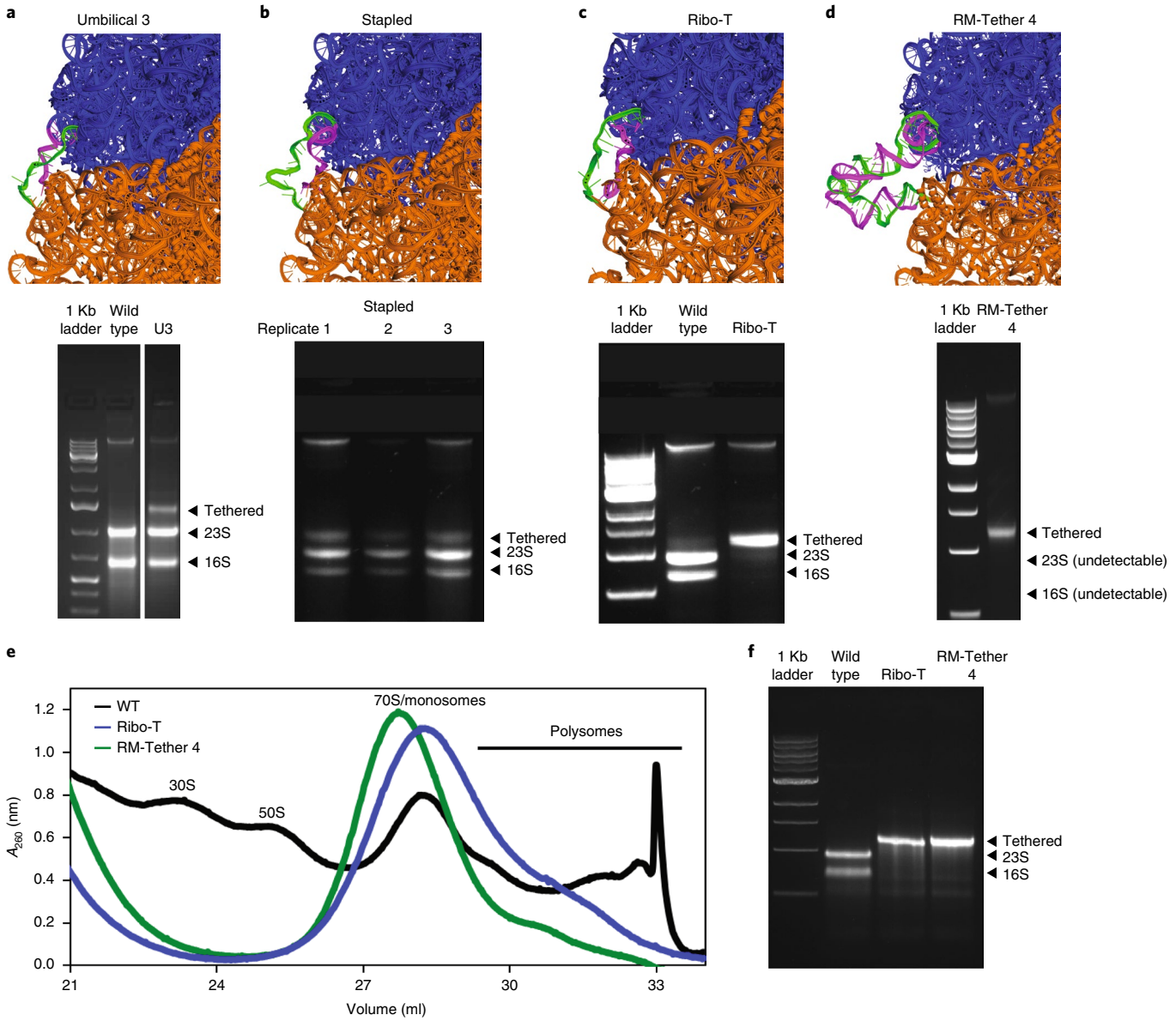

Fig. 3 | The ribosome tethering problem. a–d, Modelling (top panels) of tethers (green and magenta strands) that connect 16S and 23S rRNA into a single rRNA, and agarose gel electrophoresis (bottom panels) of RNA extracted from *E. coli* (Squires strain) in which wild-type ribosomes were completely replaced with the designed molecules (top): an early U3 tether (Umbilical 3) designed by the Jewett lab, cleaved into two bands in vivo (**a**), a stapled ribosomes developed by a separate group[15] (three replicates shown), also cleaved in this plasmid context and strain in vivo (**b**), the successful Ribo-T design developed after a year of manual trial and error to withstand cleavage in vivo (**c**) and RM-Tether 4, a design automatically generated by RNAMake, which also presents as a single band in vivo (**d**). **e**, Sucrose gradient fractionation prepared from in vitro iSAT reactions that express wild-type ribosomes, Ribo-T version 1.0 and RM-Tether 4. Peaks correspond to small subunits (30S), large subunits (50S), monosomes/70S and polysomes (standard assignments of the peaks are given in, for example, Underwood et al.[37]). **f**, Agarose electrophoresis analysis confirms that the polysome fraction of **e** is composed of tethered ribosomes. Full gels are given in Supplementary Figs. 5 and 6.

require conformational rearrangements observed in the isolated ATP aptamer (Fig. 4b,c and Supplementary Table 4)[41]. These results demonstrate that the TTR peripheral contact efficiently couples to enhance the binding of ATP in the aptameric region, as desired. As a further test of this coupling, we confirmed that the Mg^{2+} requirements to form the TTR were reduced in the presence compared to the absence of the small-molecule ligand in these constructs (Supplementary results and Supplementary Fig. 7b).

As a second test of aptamer stabilization, we assessed whether RNAMake could stabilize the Spinach RNA, which binds an analogue of the green fluorescent protein chromophore (*Z*)-4-(3,5-difluoro-4-hydroxybenzylidene)-1,2-dimethyl-1H-imidazol-5(4H)-one (DFHBI) within a G-quadruplex. Binding to Spinach enhances the fluorescence of DFHBI by ~1,000-fold relative to that of unbound ligand, which makes this RNA useful for biological interrogations[39,46], although its binding affinity, brightness, folding efficiency and biological stability remain poor even after extensive efforts to discover improvements, such as the minimized Spinach and Broccoli aptamers[47–50]. We characterized 16 'Spinach-TTR molecules' designed by RNAMake to embed the Spinach aptamer into scaffolds with tetraloop–receptor contacts (Fig. 4e and Supplementary Fig. 8). SHAPE chemical mapping confirmed that

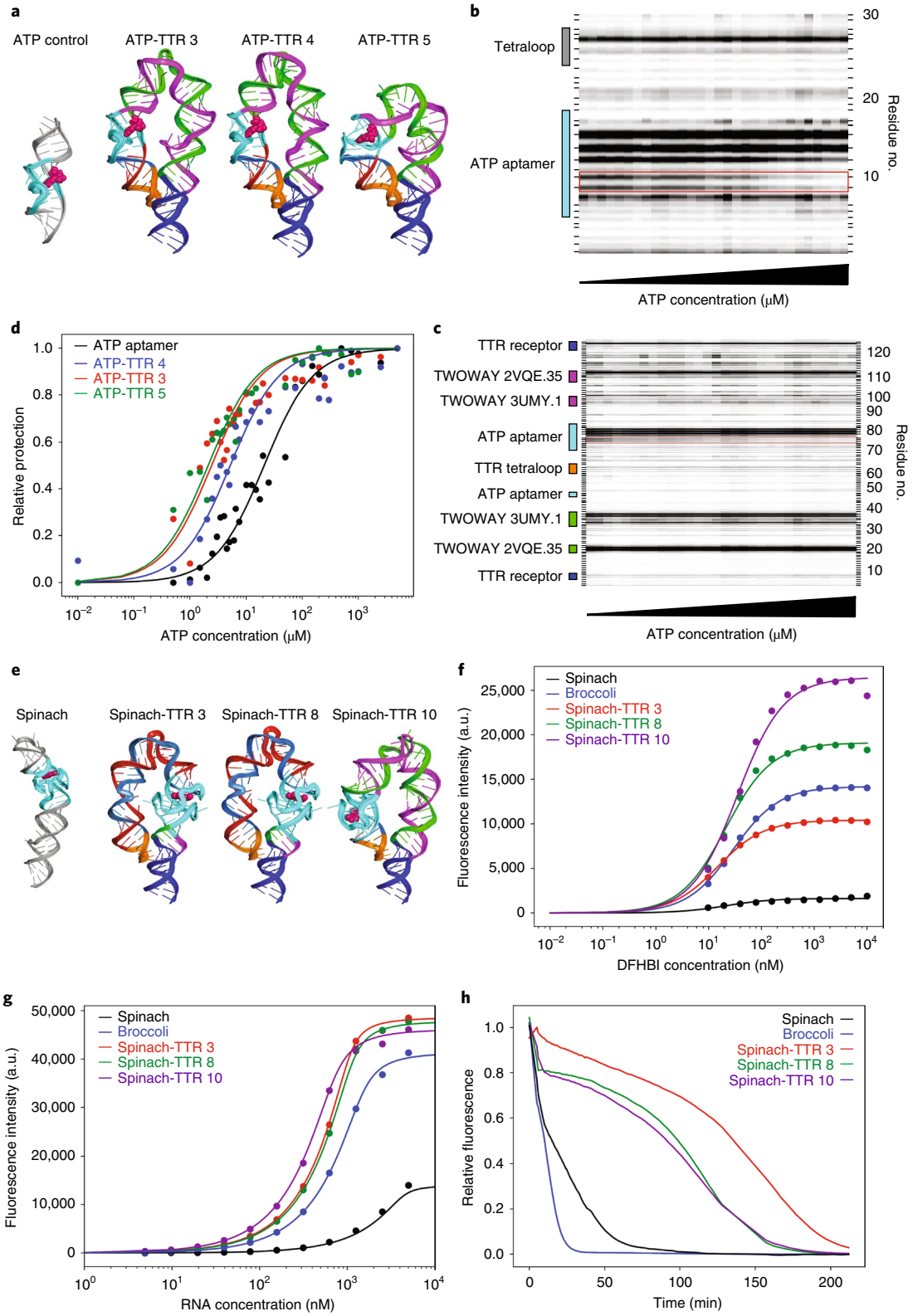

Fig. 4 | Stabilizing aptamers for ATP and light-up fluorophores through designer tertiary contacts. a, 3D models of the ATP aptamer alone and of ATP-TTR 3, 4 and 5. **b**, DMS probing of ATP titration of the ATP aptamer; red box denotes the two adenines that become protected on addition of ATP. **c**) DMS probing of ATP titration of ATP-TTR 4; the red box denotes the two adenines that become protected on the addition of ATP. **d**, Quantified DMS protection as a function of ATP concentration for the ATP aptamer compared to those of ATP-TTR 3, 4 and 5. **e**, The crystal structure of the Spinach aptamer bound to DFHBI (left) and 3D models of Spinach-TTR 3, 8 and 10. **f**, Fluorescence measurements of a DFHBI titration at constant RNA concentrations (200 nM) of Spinach, Broccoli and Spinach-TTR 3, 8 and 10. **g**, Fluorescence measurements of an RNA titration with constant DFHBI concentrations (400 nM) of Spinach, Broccoli and Spinach-TTR 3, 8 and 10. **h**, Fluorescence in 20% *X. laevis* egg extract as compared to the construct in buffer for Spinach, Broccoli and Spinach-TTR 3, 8 and 10. a.u., arbitrary units.

Table 1 | RNAMake designs and component motifs generated for all four design problems

Design problem	Designs	Successful designs	Motifs used	Experimentally validated motifs
MiniTTR	16	11	23	20
Ribosome tether	9	1	14	3
ATP-TTR	10	3	16	6
Spinach-TTR	16	13	26	20
All	51	28	56[a]	41[a]

[a]Some motifs were used in multiple design problems.

these molecules form tetraloop–receptor contacts in 13 of the 15 cases that could be tested (Supplementary Fig. 9 and Supplementary Table 5). By carrying out fluorescence assays that titrated both RNA and DFHBI concentrations, we evaluated the dissociation constants, brightness and folding efficiency of these designs. Seven of the 16 Spinach-TTR designs exhibited a twofold brighter fluorescence than that of the original Spinach as well as a brighter Broccoli aptamer (Fig. 4f, Supplementary Fig. 10 and Supplementary Tables 5 and 6). Two of these constructs, Spinach-TTR 3 and 8, were not only brighter but also gave a higher affinity and improved folding efficiency relative to that of Broccoli and a minimized Spinach construct, Spinach-min (Fig. 4g and Supplementary Tables 5 and 6).

We hypothesized that these improvements to in vitro stability measures might also lead to an improved stability in a harsh biological environment. When the DFHBI-bound RNAs were challenged with 20% whole cell lysate extracted from the eggs of *Xenopus laevis*, six of the seven Spinach-TTR constructs exhibited fluorescence for a longer time than the control Spinach and Broccoli sequences (Methods). Spinach-TTR 3 exhibited a particularly high stability (Fig. 4h), with a time to half fluorescence of 131 minutes, compared to <20 minutes for Spinach, Spinach-min and Broccoli (Supplementary Table 6 and Supplementary Fig. 11). This same robust fluorescence of the Spinach-TTRs was observed in 20% *E. coli*. lysate, which suggests a general stabilization in biological environments (Supplementary Fig. 12). We finally sought to assess the ability of the Spinach-TTR constructs to activate fluorescence in cells, using *E. coli* as a test bed. Six Spinach-TTR designs were cloned into a plasmid for T7 RNA polymerase-driven expression (Methods). Each Spinach-TTR variant was able to significantly activate expression above the background, and several designs exceeded the fluorescence observed for both Spinach and Broccoli in vivo (Supplementary Fig. 13).

Conclusions

As RNA nanotechnology seeks to create artificial molecules closer in sophistication to natural RNA molecules, the design of tertiary structures that are as complex, asymmetric and diverse as natural RNAs becomes an important goal. Here, we hypothesized that several distinct tasks in designing complex RNA tertiary structures might be reduced to instances of a single RNA motif pathfinding problem and developed the algorithm RNAMake to solve the pathfinding task (Table 1 and Supplementary results). For the miniTTR nanostructure design problem, 11 of 16 molecules exhibited the correct tertiary fold in nucleotide-resolution chemical mapping and electrophoresis assays, and we achieved a crystal structure of one design that confirmed its accuracy at a high resolution. For the problem of tethering *E. coli* 16S and 23S rRNAs into a single RNA molecule, one of nine RNAMake-designed molecules replaced ribosomes in vivo and was confirmed to translate in polysomes in cell-free translation reactions. For the problem of stabilizing aptamers through locking tertiary contacts, 3 of 10 RNAMake-designed ATP-TTR molecules achieved an improved affinity to ATP compared to that of the starting aptamer, and seven of 16 Spinach-TTR designs maintained their binding affinity for the DFHBI fluorophore while achieving improvements in fluorescence and folding efficiency in vitro, and in stability in extracto and in vivo. In each task, RNAMake achieved its design objectives in a single round of tests that involved the parallel synthesis of 8–16 constructs, without further trial-and-error iteration.

As RNAMake is applied to more problems, we expect its success rate to improve further. Accumulating knowledge as to which structural motifs recur in successful versus failing designs may allow an empirical scoring for the modularity of each motif; inferences for some motifs, such as A–A mismatches, are already possible (Supplementary discussion). Second, the incorporation of motifs that are known to sample at least two conformations (for example, the triple mismatch in miniTTR6 herein or kink turns) may allow an improved design of such machines as the ribosome, and improved cryogenic electron microscopy methods may provide more detailed feedback on such distinct states[18,51]. Third, natural structured RNAs often contain multiple tertiary contacts and multibranched junctions and we have extended RNAMake's pathfinding method to design such motifs (Supplementary Fig. 14). Finally, we expect RNAMake's computational design approach to be complementary to library selection and high-throughput screening methods[52], especially for larger problems that require numerous non-canonical motifs. By distributing RNAMake as a source code and a server, we hope to encourage these applications and extensions of computational RNA design.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41565-019-0517-8.

Received: 26 July 2017; Accepted: 24 June 2019;
Published online: 19 August 2019

References

1. Guo, P. The emerging field of RNA nanotechnology. *Nat. Nanotechnol.* **5**, 833–842 (2010).
2. Grabow, W. W. & Jaeger, L. RNA self-assembly and RNA nanotechnology. *Acc. Chem. Res.* **47**, 1871–1880 (2014).
3. Leontis, N. B., Lescoute, A. & Westhof, E. The building blocks and motifs of RNA architecture. *Curr. Opin. Struct. Biol.* **16**, 279–287 (2006).
4. Jaeger, L. & Chworos, A. The architectonics of programmable RNA and DNA nanostructures. *Curr. Opin. Struct. Biol.* **16**, 531–543 (2006).
5. Jaeger, L. & Leontis, N. B. Tecto-RNA: one-dimensional self-assembly through tertiary interactions. *Angew. Chem. Int. Ed.* **39**, 2521–2524 (2000).
6. Zhang, H. et al. Crystal structure of 3WJ core revealing divalent ion-promoted thermostability and assembly of the Phi29 hexameric motor pRNA. *RNA* **19**, 1226–1237 (2013).
7. Weizmann, Y. & Andersen, E. S. RNA nanotechnology—the knots and folds of RNA nanoparticle engineering. *MRS Bull.* **42**, 930–935 (2017).
8. Jasinski, D., Haque, F., Binzel, D. W. & Guo, P. Advancement of the emerging field of RNA nanotechnology. *ACS Nano* **11**, 1142–1164 (2017).
9. Afonin, K. A. et al. Computational and experimental characterization of RNA cubic nanoscaffolds. *Methods* **67**, 256–265 (2014).
10. Jossinet, F., Ludwig, T. E. & Westhof, E. Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. *Bioinformatics* **26**, 2057–2059 (2010).
11. Wimberly, B. T. et al. Structure of the 30S ribosomal subunit. *Nature* **407**, 327–339 (2000).
12. Nguyen, T. H. D. et al. The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature* **523**, 47–52 (2015).
13. Miao, Z. et al. RNA-Puzzles Round III: 3D RNA structure prediction of five riboswitches and one ribozyme. *RNA* **23**, 655–672 (2017).
14. Nasalean, L., Baudrey, S., Leontis, N. B. & Jaeger, L. Controlling RNA self-assembly to form filaments. *Nucleic Acids Res.* **34**, 1381–1392 (2006).
15. Fried, S. D., Schmied, W. H., Uttamapinant, C. & Chin, J. W. Ribosome subunit stapling for orthogonal translation in *E. coli*. *Angew. Chem. Int. Ed.* **54**, 12791–12794 (2015).

16. Orelle, C. et al. Protein synthesis by ribosomes with tethered subunits. *Nature* **524**, 119–124 (2015).
17. Carlson, E. D. Creating Ribo-T: (design, build, test)^n. *ACS Synth. Biol.* **4**, 1173–1175 (2015).
18. Schmied, W. H. et al. Controlling orthogonal ribosome subunit interactions enables evolution of new function. *Nature* **564**, 444–448 (2018).
19. Famulok, M. Oligonucleotide aptamers that recognize small molecules. *Curr. Opin. Struct. Biol.* **9**, 324–329 (1999).
20. Porter, E. B., Polaski, J. T., Morck, M. M. & Batey, R. T. Recurrent RNA motifs as scaffolds for genetically encodable small-molecule biosensors. *Nat. Chem. Biol.* **13**, 295–301 (2017).
21. Gotrik, M. et al. Direct selection of fluorescence-enhancing RNA aptamers. *J. Am. Chem. Soc.* **140**, 3583–3591 (2018).
22. Montange, R. K. & Batey, R. T. Riboswitches: emerging themes in RNA structure and function. *Annu. Rev. Biophys.* **37**, 117–133 (2008).
23. Macke, T. J. & Case, D. A. in *Molecular Modeling of Nucleic Acids* (eds Leontis, N. B. & SantaLucia, J.) 379–393 (American Chemical Society, 1997).
24. Lee, J. et al. RNA design rules from a massive open laboratory. *Proc. Natl Acad. Sci. USA* **111**, 2122–2127 (2014).
25. Dibrov, S. M., McLean, J., Parsons, J. & Hermann, T. Self-assembling RNA square. *Proc. Natl Acad. Sci. USA* **108**, 6405–6408 (2011).
26. Afonin, K. A. et al. Multifunctional RNA nanoparticles. *Nano Lett.* **14**, 5662–5671 (2014).
27. Khisamutdinov, E. F. et al. Fabrication of RNA 3D nanoprisms for loading and protection of small RNAs and model drugs. *Adv. Mater.* **28**, 10079–10087 (2016).
28. Bindewald, E., Grunewald, C., Boyle, B., O'Connor, M. & Shapiro, B. A. Computational strategies for the automated design of RNA nanoscale structures from building blocks using NanoTiler. *J. Mol. Graph Model* **27**, 299–308 (2008).
29. Huang, L. & Lilley, D. M. J. A quasi-cyclic RNA nano-scale molecular object constructed using kink turns. *Nanoscale* **8**, 15189–15195 (2016).
30. Wu, L., Chai, D., Fraser, M. E. & Zimmerly, S. Structural variation and uniformity among tetraloop–receptor interactions and other loop–helix interactions in RNA crystal structures. *PLoS ONE* **7**, e49225 (2012).
31. Frederiksen, J. K., Li, N.-S., Das, R., Herschlag, D. & Piccirilli, J. A. Metal-ion rescue revisited: biochemical detection of site-bound metal ions important for RNA folding. *RNA* **18**, 1123–1141 (2012).
32. Rangan, P., Masquida, B., Westhof, E. & Woodson, S. A. Assembly of core helices and rapid tertiary folding of a small bacterial group I ribozyme. *Proc. Natl Acad. Sci. USA* **100**, 1574–1579 (2003).
33. Fiore, J. L. & Nesbitt, D. J., An RNA folding motif: GNRA tetraloop–receptor interactions. *Q Rev. Biophys.* **46**, 223–264 (2013).
34. 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).
35. Jewett, M. C., Fritz, B. R., Timmerman, L. E. & Church, G. M. In vitro integration of ribosomal RNA synthesis, ribosome assembly, and translation. *Mol. Syst. Biol.* **9**, 678 (2013).
36. Fritz, B. R., Jamil, O. K. & Jewett, M. C. Implications of macromolecular crowding and reducing conditions for in vitro ribosome construction. *Nucleic Acids Res.* **43**, 4774–4784 (2015).
37. Underwood, K. A., Swartz, J. R. & Puglisi, J. D. Quantitative polysome analysis identifies limitations in bacterial cell-free protein synthesis. *Biotechnol. Bioeng.* **91**, 425–435 (2005).
38. Carothers, J. M., Oestreich, S. C. & Szostak, J. W. Aptamers selected for higher-affinity binding are not more specific for the target ligand. *J. Am. Chem. Soc.* **128**, 7929–7937 (2006).
39. Paige, J. S., Wu, K. Y. & Jaffrey, S. R. RNA mimics of green fluorescent protein. *Science* **333**, 642–646 (2011).
40. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).
41. Jiang, F., Kumar, R. A., Jones, R. A. & Patel, D. J. Structural basis of RNA folding and recognition in an AMP–RNA aptamer complex. *Nature* **382**, 183–186 (1996).
42. Huang, Z. & Szostak, J. W. Evolution of aptamers with a new specificity and new secondary structures from an ATP aptamer. *RNA* **9**, 1456–1463 (2003).
43. Sassanfar, M. & Szostak, J. W. An RNA motif that binds ATP. *Nature* **364**, 550–553 (1993).
44. Sazani, P. L., Larralde, R. & Szostak, J. W. A small aptamer with strong and specific recognition of the triphosphate of ATP. *J. Am. Chem. Soc.* **126**, 8370–8371 (2004).
45. Geary, C., Chworos, A., Verzemnieks, E., Voss, N. R. & Jaeger, L. Composing RNA nanostructures from a syntax of RNA structural modules. *Nano Lett.* **17**, 7095–7101 (2017).
46. Kellenberger, C. A., Chen, C., Whiteley, A. T., Portnoy, D. A. & Hammond, M. C. RNA-based fluorescent biosensors for live cell imaging of second messenger cyclic di-AMP. *J. Am. Chem. Soc.* **137**, 6432–6435 (2015).
47. Strack, R. L., Disney, M. D. & Jaffrey, S. R. A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat-containing RNA. *Nat. Methods* **10**, 1219–1224 (2013).
48. Filonov, G. S., Moon, J. D., Svensen, N. & Jaffrey, S. R. Broccoli: rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *J. Am. Chem. Soc.* **136**, 16299–16308 (2014).
49. Ketterer, S., Fuchs, D., Weber, W. & Meier, M. Systematic reconstruction of binding and stability landscapes of the fluorogenic aptamer Spinach. *Nucleic Acids Res.* **43**, 9564–9572 (2015).
50. Song, W., Strack, R. L., Svensen, N. & Jaffrey, S. R. Plug-and-play fluorophores extend the spectral properties of Spinach. *J. Am. Chem. Soc.* **136**, 1198–1201 (2014).
51. Shi, X., Huang, L., Lilley, D. M. J., Harbury, P. B. & Herschlag, D. The solution structural ensembles of RNA kink-turn motifs and their protein complexes. *Nat. Chem. Biol.* **12**, 146–152 (2016).
52. Buenrostro, J. D. et al. Quantitative analysis of RNA–protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. *Nat. Biotechnol.* **32**, 562–568 (2014).

Acknowledgements

We thank S. Bonilla for assistance in performing the native gel assays and A. Watkins for discussions about ribosome tether design. We thank the Straight lab for graciously providing *X. laevis* whole cell lysate. This work was supported by the National Institutes of Health, through NIGMS MIRA R35 GM122579 (R.D.), R01 GM121487 (R.D.), New Innovator Award 1DP2GM110838 (J.B.L.), Ruth L. Kirschstein National Research Service Award Postdoctoral Fellowships GM112294 (J.D.Y.) and GM100953 (D.E.), P01 GM066275 (D.H.) and R35 GM118070 (J.S.K.), a Stanford School of Medicine Discovery Innovation Award (R.D.), Army Research Office W911NF-16-1-0372 (M.C.J.), the National Science Foundation through MCB-1716766 (M.C.J.), Career Award 1452441 (J.B.L.) and Graduate Research Fellowship DGE-1324585 (A.E.D.), the David and Lucile Packard Foundation (M.C.J.) and the Camille Dreyfus Teacher-Scholar Program (M.C.J.).

Author contributions

R.D. and J.D.Y. conceived the study. J.D.Y. developed RNAMake and generated the models and sequences used throughout the study. A.N.O., W.K. and J.D.Y. performed the chemical mapping, titrations and native gel assays for miniTTR constructs. X.S. and D.H. designed and performed the SAXS on miniTTR 2 and 6. D.E. solved the miniTTR crystal structure assisted by D.A.C. and J.S.K. in preparing the RNA and analysis. E.D.C., A.E.D. and M.C.J. made and tested the RNAMake-designed ribosomes P.D.C. and J.B.L. carried out the SHAPE-seq and in vivo tests of Spinach-TTRs. M.R.G. performed fluorescence and lysate experiments on Spinach-TTRs. J.D.Y. and R.D. wrote the paper, with participation by all the authors.

Competing interests

Stanford University is filing a patent on aptamer stabilization with RNAMake.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41565-019-0517-8.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to R.D.

Peer review information: *Nature Nanotechnology* thanks Peixuan Guo, Nils (G) Walter and other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

Methods

Software availability. All the software and source code used in this work are freely available for non-commercial use. RNAMake software and documentation are at http://rnamake.stanford.edu. An RNAMake server to perform scaffolding and aptamer stabilization is available at http://rnamake.stanford.edu. EteRNAbot secondary structure design is available at https://software.eternagame.org.

Sequences and primers. All the sequences and primers used in this study are given in Supplementary sequences and Supplementary primers, respectively.

Building the motif library. To build a curated motif library of all the RNA structural components, we obtained the set of non-redundant RNA crystal structures managed by the Leontis and Zirbel groups[53] (version 1.45, http://rna.bgsu.edu/rna3dhub/nrlist/release/1.45). This set specifically removes redundant RNA structures that are identical to previously solved structures, such as ribosomes crystallized with different antibiotics. We processed each RNA structure to extract every motif with Dissecting the Spatial Structure of RNA (DSSR)[54] with the following command:

```
x3dna-dssr -i file.pdb -o file_dssr.out
```

We manually checked each extracted motif to confirm that it was the correct type, as DSSR sometimes classifies tertiary contacts as higher-order junctions and vice versa. For each motif collected from DSSR, we ran the X3DNA `find_pair` and `analyze` programs to determine the reference frame for the first and last base pair of each motif to allow for the alignment between motifs:

```
find_pair file.pdb 2> /dev/null stdout | analyze stdin
>& /dev/null
```

The naming convention for each motif involves the motif classification, the originating PDB accession code and a unique number to distinguish from other motifs of the same type, all separated by periods. For example, TWOWAY.1GID.2, is a two-way junction from the PDB 1GID and is the third two-way junction to be found in this structure. All the motifs retain their original residue numbering, chain identifications and relative position compared to their originating structure.

In addition to the motifs derived from the PDB, we also utilized the make-na web server (http://casegroup.rutgers.edu/nab.html) to generate idealized helices of between 2 and 22 base pairs in length[23]. All the motifs in these generated libraries are bundled with RNAMake and are grouped together by type (junctions, hairpins and so on) in sqlite3 databases in the RNAMake directory RNAMake/resources/motif_libraries_new.

Automatically building new RNA segments. RNAMake seeks a path for RNA helices and non-canonical motifs that can connect two base pairs separated by a target translation and rotation. We developed a depth-first search algorithm to discover such RNA paths. The algorithm is guided by a heuristic cost function f inspired by prior manual design efforts[2,25] and is composed of two terms:

$$f(\text{path}) = h(\text{path}) + g(\text{path}) \tag{1}$$

The first term, h(path), describes how close the last base pair in the path is to the target base pair; h(path) = 0 corresponds to a perfect overlap in translation and rotation. The functional form for h(path) depends on the spatial position of each base pair's centroid d and an orthonormal coordinate frame R that defines the rotational orientation of each base pair[46]:

$$h(\text{path}) = |\vec{d_1} - \vec{d_2}| + W\left(|\vec{d_1} - \vec{d_2}|\right) \sum_{i}^{3} \sum_{j}^{3} \text{abs}\left(R_{1ij} - R_{2ij}\right) \tag{2}$$

Here $W(d)$ is:

$$W(d) = \begin{cases} 0, & \text{if } d > 150 \\ \log \frac{150}{d}, & \text{if } 1.5 < d < 150 \\ 2, & \text{if } 1.5 > d \end{cases} \tag{3}$$

where d is measured in ångströms. The weight $W(d)$ reduces the importance of the current base pair and the target base pair with a similar alignment when they are spatially far apart. This term conveys the intuition that aligning the two coordinate frames becomes important only as the path of the motif and helices approaches the target base pair. RNAMake readily allows for the exploration of alternative forms of the cost function terms (2) and (3), which include more standard rotationally invariant metrics to define rotation matrix differences[55] or base-pair-to-base-pair r.m.s.d. values based on quaternions[56], but these were not tested in the current study.

The second term in the cost function (1) is g(path), which parameterizes the properties of the non-canonical RNA motifs and helices that comprise the path at each stage of the calculation:

$$g(\text{path}) = \frac{S_{ss}(\text{path})}{2} + 2N_{\text{motifs}} \tag{4}$$

where S_{ss} is a secondary structure score for all the motifs and helices in the path. This S_{ss} term favours longer canonical helices as well as motifs with frequently recurring base pairs, as follows. All the base pairs found in the RNA motif are scored based on their relative occurrences in all the high-resolution crystal structures; all the unpaired residues receive a penalty and Watson–Crick base pairs receive an additional bonus score (Supplementary Table 9). Values were derived based on logarithms of the frequencies of these elements in the crystallographic database, that is the inverse Boltzmann approximation[57], so that that the frequency of the elements in RNAMake designs was similar to that seen in natural RNA tertiary structures. In addition to the secondary structure score, N_{motifs} penalizes the total number of motifs in the path, here taken as the number of non-canonical motifs plus the number of helices (independent of the helix length).

The search adds motifs and helices to the path in a depth-first manner, and as the total cost function f(path) decreases, the back-tracking of f(path) increases. Any solutions with h(path) less than five, that is, overlap at approximately nucleotide resolution between the path's last base pair and the target base pair, are accepted into a list of final designs. The balance between g(path) and h(path) allows RNAMake to reduce the number of motif combinations considered and find most solutions in a few seconds. For each solution, we then used EteRNAbot, a secondary structure optimization algorithm that has undergone extensive empirical tests[24], to fill in the helix sequences.

Proteins that are included in the coordinates supplied to RNAMake are represented as steric beads centred at the Cα atom of each amino acid. This representation allows RNAMake to avoid steric clashes with proteins, particularly for the ribosome tethering problems.

Design, synthesis and experimental testing of miniTTR constructs. RNAMake designs of miniTTR constructs, in vitro synthesis and experimental testing are given in the Supplementary methods.

Design, construction and experimental testing of ribosome tether constructs. RNAMake designs of ribosome tether constructs, cloning and replacing wild-type ribosome and experimental are given in the Supplementary methods.

Design, synthesis and experimental testing of aptamer-stabilizing constructs. RNAMake designs of ATP- and Spinach-stabilized constructs and experimental testing by RNA structure probing, fluorescence measurements, Spinach-TTR stability assay in *Xenopus* egg extract and in vivo Spinach aptamer testing are given in the Supplementary methods.

Data availability

The data that support the plots within this paper and other findings of this study are available from the corresponding author upon reasonable request. Furthermore, all of our chemical mapping data are available on https://rmdb.stanford.edu, and a detailed table of the accession identifications is given in the Supplementary Information.

References

53. Petrov, A. I., Zirbel, C. L. & Leontis, N. B. Automated classification of RNA 3D motifs and the RNA 3D Motif Atlas. *RNA* **19**, 1327–1340 (2013).
54. 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).
55. Huynh, D. Q. Metrics for 3D rotations: comparison and analysis. *J. Math. Imaging Vis.* **35**, 155–164 (2009).
56. Karney, C. F. F. Quaternions in molecular modeling. *J. Mol. Graph Model* **25**, 595–604 (2007).
57. Finkelstein, A. V., Badretdinov, A. Ya & Gutin, A. M. Why do protein architectures have Boltzmann-like statistics? *Proteins* **23**, 142–150 (1995).