Published in final edited form as:
Nat Chem. 2021 June 01; 13(6): 549–558. doi:10.1038/s41557-021-00679-1.

RNA origami design tools enable cotranscriptional folding of kilobase-sized nanoscaffolds

Cody Geary[1,2], **Guido Grossi**[1], **Ewan K. S. McRae**[1], **Paul W. K. Rothemund**[2,*], **Ebbe S. Andersen**[1,*]

[1]Interdisciplinary Nanoscience Center and Department of Molecular Biology and Genetics, Aarhus University, Aarhus C, 8000, Denmark

[2]Bioengineering, Computing + Mathematical Sciences, and Computation & Neural Systems, California Institute of Technology, Pasadena, CA 91125, USA

Abstract

RNA origami is a framework for the modular design of nanoscaffolds that can be folded from a single strand of RNA, and used to organize molecular components with nanoscale precision. Design of genetically expressible RNA origami, which must cotranscriptionally fold, requires modeling and design tools that simultaneously consider thermodynamics, folding pathway, sequence constraints, and pseudoknot optimization. Here, we describe RNA Origami Automated Design software (ROAD), which builds origami models from a library of structural modules, identifies potential folding barriers, and designs optimized sequences. Using ROAD, we extend the scale and functional diversity of RNA scaffolds, creating 32 designs of up to 2360 nucleotides, five that scaffold two proteins, and seven that scaffold two small molecules at precise distances. Micrographic and chromatographic comparison of optimized and nonoptimized structures validates that our principles for strand routing and sequence design substantially improve yield. By providing efficient design of RNA origami, ROAD may simplify construction of custom RNA scaffolds for nanomedicine and synthetic biology.

The field of RNA nanotechnology began by extracting RNA structural modules from natural RNA molecules and connecting them to create engineered constructs[1, 2, 3]. This approach was enabled by the structural determination of biological RNA molecules, such as the ribosomal subunits[4, 5], which provided a large library of RNA modules to build from. With these modules, architectures ranging from multi-stranded tiles to single-stranded origami have been explored. Of particular recent interest are RNA structures designed to fold cotranscriptionally, during their synthesis by RNA polymerase, which have the benefit that they can be genetically expressed and folded within cells. Previously, we introduced the

*correspondence to: pwkr@dna.caltech.edu and esa@inano.au.dk.

Author Contributions

C.G., P.W.K.R and E.S.A conceived the project. C.G., G.G., and E.K.S.M. performed the research. P.W.K.R and E.S.A supervised the project. C.G., P.W.K.R and E.S.A wrote the manuscript. All authors analyzed the data and commented on the manuscript.

Competing Interests

The authors declare no competing interests.

Europe PMC Funders Author Manuscripts

RNA origami method[6]—a highly regular architecture that arrange RNA helices into parallel arrays held together by crossovers and kissing loops (KLs)—which is compatible with cotranscriptional folding, but several bottlenecks in computational design methods limited the size (450 nt) and folding yield. Later studies constructed somewhat larger (715 nt) wireframe single-stranded cotranscriptional shapes by composing complex tertiary motifs *in vitro* [7] and *in vivo* [8]. The largest currently achieved structures (6000 nts) require long (~18-hour) thermal anneals[9] making them incompatible with cotranscriptional folding in cells.

RNA nanostructures can serve as functional scaffolds by directly incorporating RNA-protein binding domains[10, 11], small molecule aptamers[12, 13], biosensors[14], ribozymes[15], siRNAs[16], or combinations of such modifications to create multifunctional nanoparticles[17, 18]. RNA nanostructures that fold cotranscriptionally[6, 7] have been expressed in cells[8, 12] where they have the potential to be used as biosensors, scaffolds or regulators for synthetic biology applications[19]—for example, to control product formation from colocalized enzymes[20, 21] and perform gene regulation via recruitment of transcription factors[22]. To verify that two proteins are located on the same scaffold, split fluorescent proteins[23] or Förster resonance energy transfer (FRET) between fluorescent proteins[24] are often used. Similarly, fluorescent RNA aptamers (split-Spinach[25] and apta-FRET[12]) have been used to verify scaffolding effects. RNA origami structures may incorporate tertiary motifs such as the IRES[13] or bKL[26] motifs to produce ~90° bends that allow out-of-plane functionalization. 2'-fluoro-modified RNA origami scaffolds carrying the thrombin aptamer have been used to produce a potent therapeutic anticoagulant[11].

Computational methods have played a central role in developing RNA nanotechnology by facilitating core tasks[27, 28]. Dedicated software has been developed to ease the construction of RNA nanostructures from 3D structural motifs[29, 30, 31]. However, no software exists for the interactive 3D modeling of large and regular RNA scaffolds such as the RNA origami architecture. Algorithms simulating RNA cotranscriptional folding have been developed for predicting folding pathways,[32, 33] which for small structures enables designers to verify that their sequences will avoid kinetic traps; it has not been possible to do this for RNA origami. RNA sequence design algorithms were originally developed based on secondary structure thermodynamic folding algorithms[34], but these lack the ability to efficiently predict pseudoknots (such as KLs). RNA origami, which are stabilized by numerous KL interactions along their strand path, necessarily contain numerous pseudoknots and are therefore not easy to design. Another important element for RNA sequence design is the ability to incorporate numerous sequence constraints to allow RNA sequence and structural motifs to be added, but current design pipelines lack the ability to simultaneously incorporate the multiple constraints necessary for the design of RNA origami structures[6, 23, 35].

Here we introduce the RNA Origami Automated Design software (ROAD)—a computer-aided design software to automate the 3D modeling of structures, analyze folding paths, and design sequences and KLs that fold into the designated structures—allowing us to greatly extend the scale and diversity of RNA origami scaffolds. ROAD allows us to rapidly prototype multiple distinct scaffolds, and investigate the effects of different design parameters with a short design cycle. We study the effect of curvature and cross-over placement within RNA origami structures by atomic force microscopy (AFM) allowing us to

greatly increase the scale of the structures. To study the effect on yield, we then constructed a set of non-optimal designs and analyzed yields by size-exclusion chromatography (SEC) and negative stain transmission electron microscopy (TEM). Finally, we tested the ability of ROAD to design RNA origami scaffolds embedded with aptamers for binding fluorescent proteins and small molecule fluorophores, and used FRET between the fluorescent molecules as a distance indicator to validate the precision of scaffolding.

Results

Design tools for creating RNA origami scaffolds

We have developed the ROAD software package (see code availability) to automate the main design steps for RNA origami: model building, folding path analysis, and sequence design. ROAD is based on a library of compatible structural modules used to construct RNA origami structures. Core modules such as helices, junctions, and 180KLs (KLs that interact at an angle of 180°)[36] are used to build the central scaffold (Fig. 1a) and peripheral modules such as tetraloops[37], 120KL-connectors (KLs that interact at an angle of 120°)[38], light-up aptamers[39, 40] and protein binding aptamers[41, 42] are used to add functionality (Fig. 1b). Schematic representations of the core modules can be used like Lego bricks to compose a large diversity of different designs (Fig. 1c) that directly translate to atomic coordinates (Fig. 1d). Closely spaced crossovers between three helices result in ‘dovetail’ (DT) junctions[6] (Fig. 1e), which is an important design parameter for RNA origami, since the DT length (in bps) changes the dihedral angle Φ between connected helices (Fig. 1f). DTs can only have certain lengths to avoid helices to sterically clash, and are named sDT, where the spacing s can have values from -5 to +2 bps[43] (Supplementary Fig. 4).

The ROAD software package consists of three main algorithms: RNAbuild, RNApath and Revolvr, that take a user-specified ‘RNA blueprint’ as input. RNA blueprints are text-based diagrams that encode all Watson-Crick base pairs (bp), sequence constraints, pseudoknots, base stacking at junctions, and 5’ to 3’ strand orientations (Fig. 1g). RNAbuild uses a module library to build atomic models according to specifications in the blueprint (Fig. 1h). The automated atomic modeling helps the user to design curvature and avoid steric clashes within larger RNA structures that are otherwise unapparent in the RNA blueprint. RNApath analyzes the folding path for potential topological barriers that may arise during the cotranscriptional folding process (Fig. 1i). Topological barriers can arise if a KL interaction (Fig.1g blue arrow) forms before the formation of helices in the loop region (Fig. 1g, pink and orange arrows) since the formation of a double helix may be sterically hindered by the closed loop region. RNApath determines topological barriers based on the relative rates of KL and helix formation, as well as the speed of synthesis, and generates plots and 3D folding animations (Supplementary Videos 1–6) to help the user avoid topology-based misfolding. Revolvr is a sequence design algorithm that uses a multi-stage sequence optimization procedure involving positive design by minimum free energy (MFE)[44] prediction, negative design by sequence symmetry minimization (SSM)[45], and KL orthogonalization, to develop a sequence that folds into the target structure (Fig. 1j). The ROAD package and the analysis scripts are described in the Methods section, a tutorial is

provided as Supplementary Note 1, and a webserver has been established to make the software easily accessible (see code availability).

Design of multivalent interfaces for RNA origami tiles

The ROAD software was used to design three-helix (3H) RNA origami tiles with edge interactions to form fibers or rings, which make them easier to observe by AFM imaging. To make the interactions stronger the 3H tiles were connected by two 120KL interactions[38]; their relative in-plane positioning defines the tile-tile interaction angle θ (Fig. 2a) that can deviate from 120° since the KL motif is flexible enough to accommodate a range of angles[46]. Using RNAbuild, we designed three trapezoidal 3H tiles with different tile-tile interaction angles (θ = 120°, 135°, 108°) that form closed polygonal objects (blue and white models in Fig. 2c–e, Supplementary Fig. 6). The different θ angles were made possible by changing the tile geometry by different DT spacings (named 3H*s*DT, where s = -2, -3 and -4 bps corresponding to Φ = 155°, 122°, and 89°, respectively). For characterization of the designs, we introduce a new near native sample preparation protocol for AFM imaging to capture structures formed in the transcription reaction on the mica surface (Fig. 2b). AFM experiments (Fig. 2f–h, Supplementary Fig. 5) showed that of polygons observed for 3H-2DT (n = 27), 59% were hexagons, 30% were pentagons, and 11% were heptagons; for 3H-3DT only a few octagons were observed, but most tiles participated in open structures that we interpret as helical fibers; for 3H-4DT (n = 72), 69% were pentagons, 26% were hexagons, and 4% were heptagons or quadrilaterals (Supplementary Fig. 6). The data show that 120KLs can be used to create multivalent binding interfaces with θ from 108° to 135°. The folding yield of the individual RNA origami tiles was estimated to be 72-89% by counting of well-formed versus broken structures in AFM images (Supplementary Fig. 12–14, Supplementary Table 3).

Expanding the size of RNA origami structures

Motivated by a desire to make scaffolds large enough for organizing multiple proteins, we explored geometric details and design approaches important for scaling up RNA origami. The modular combination of smaller, already validated RNA motifs is a common and successful approach to the design of larger structures[47, 48]. Here, starting with domains from tile 3H-2DT (Fig. 2c,f), we hierarchically applied duplication and fusion (Supplementary Fig. 8) to design sets of taller and wider scaffolds (Fig. 3a). Extension of RNA origami in the x-direction required no geometric innovation, but extension in the y-direction required consideration of Φ-based curvature when adding multiple rows of helices. DTs that alternates between 0 and -2 bp DTs result in minimum curvature of the RNA origami, but unfortunately 0 bp DTs introduce a potential weakness into an RNA origami, since each 0 bp DT is effectively a 6-arm junction with at least three sterically plausible alternative stacking conformations (*cf.* the two stacking isomers observed in 4-arm junctions[49]). To better stabilize and specify desired folds, we used larger ‘offset DTs’—DTs displaced by a helical turn of RNA, which maintain the same dihedral angle Φ as shorter DTs (e.g. -11 bp and +11 bp DTs rather than 0 bp DTs).

To reach five helices tall, two copies of 3H-2DT were merged via +11 or -11 bp offset DTs to create tiles ZigZag-A-1X and ZigZag-B-1X, respectively (Fig. 3b,c, Supplementary Fig.

8). Three 120KLs added to the edges of these tiles were programmed to join the tiles in a *trans* configuration, resulting in zigzag-shaped filaments (Fig. 3b,c, blue and grey models) in which alternating tiles face up and down, a corrugated configuration which balances tile curvature (cf. previous polygons in Fig. 2c–e in which all tiles face in the same direction). Samples were imaged by AFM and analyzed to show a folding yield of 77-85% similar to the 3H-2DT tiles (Fig. 3b,c, Supplementary Fig. 15–16). A few alternative 6H tiles that contained isolated 0 bp DTs were shown to fold well (Supplementary Fig. 7 combining +11, 0, -2, -11 and -13 bp DTs). To create still taller tiles, two copies of ZigZag-A-1X were merged, via -11 bp DTs, to create the core of a nine-helix tile (Supplementary Fig. 8). Addition of 120KLs resulted in *trans* connections and filaments of alternating up-down orientations for ZigZag-B-9H tiles (Fig. 3d); addition of 180KLs resulted in *cis* connections and filaments of consistent orientation for Ribbon-9H tiles (Fig. 3e). The 9H tiles showed more partial structures and had a reduced folding yield estimated to be 51-62% (Supplementary Fig. 17–18), which could be caused by topological folding barriers (marked in red and orange in Supplementary Video 1) as suggested by RNApath analysis.

We used lateral duplication and fusion (Supplementary Fig. 8) of ZigZag-B-1X to create tiles with two repeats (ZigZag-B-2X in Fig. 3f and Supplementary Fig. 10; Supplementary Fig. 11 shows unexpected edge interactions) or four repeats (ZigZag-B-4X in Fig. 3g and Supplementary Fig. 10). The 2X duplication did not seem to affect yield (estimated to be 78%), whereas the 4X duplication had a reduced yield of 58% (Supplementary Fig. 19–20). The reduction in yield of the large 12 nm x 48 nm ZigZag-B-4X could not be explained by RNApath analysis (Supplementary Video 2), but is more likely caused by the misfolding and aggregation of its long transient 5’ single stranded end. Tiles with alternating -2 bp and +/-11 bp DTs will be flat but have steeply sloped sides. To obtain a more rectangular tile we replaced each -2 bp DTs with +9 bp DTs (-2 bp offset by +11 bp), so that every repeat unit has a counterbalanced set of +9 bp and -11 bp DTs. As an example of this architecture we designed the three-repeat Ribbon-5H-3X with 180KLs connectors resulting in straight linear chains as observed by AFM with a yield of 42% (Fig. 3h, Supplementary Fig. 21). As a second example the tile was extended to 9 helices tall and designed without intermolecular connections, as a standalone scaffold, reaching a length of 2360 nts and a size of 20 nm x 36 nm (Fig. 3a, Rectangle-9H-3X), however, the expansion resulted in only few examples of rectangular shapes that all had folding defects (Fig.3i). Finally, we designed RNA origamis with shorter or longer double crossover spacing: ZigZag-B-2X-Mini with two-turns between crossovers (Supplementary Figs. 7) and Ribbon-5H-3X-bumps with four-turns between crossovers (Supplementary Figs. 7 and 10). The latter was designed with six out-of-plane dumbbells placed in the middle of the 4-turn stem regions, however, the complexity of the design resulted in a low observed yield of 30% (Supplementary Fig. 22), and the three-dimensionality of the design resulted in poor imaging by AFM.

The manual evaluation of folding yields from the AFM images is summarized in Supplementary Table 3. The folding yields negatively correlates with increasing length of the RNA origami structures tested, and with RNApath-predicted topologically blocked positions. This observation is supported by an apparent correlation between the number of observed misfolded structures and RNApath-predicted topologically blocked positions. The data indicates that folding topology is important and that increasing the height of the tile

results in the increased occurrence of predicted topological barriers, that arise because of the longer delay between the synthesis of KL partners. The large Rectangle-9H-3X was predicted to have several topological barriers, this correlated with the larger folding defects observed (Fig.3i, Supplementary Fig. 23). Another example is a merged version of ZigZag-A-1X and ZigZag-B-1X that is 10 helices tall, where we observe partly-formed tiles that again have large defects that seem to correspond to the regions with predicted topological barriers (Supplementary Fig. 23).

Effects of design parameters on folding yield

To support our AFM yield analysis, we performed negative stain TEM imaging of SEC-purified monomer RNA origami structures. A monomeric 5-helix scaffold (5HS, Fig. 4a), based on one of our best performing RNA tiles ZigZag-B-1X (85% yield by AFM) resulted in 86% yield of monomer as determined by SEC analysis (Supplementary Fig. 24) and TEM images of the monomer sample revealed homogeneous and monodisperse particles with class averages displaying highly resolved details of tight helix packing (Fig. 4c and Supplementary Fig. 24). The TEM analysis revealed a clear preference for observing either front or back face views of the 5HS structure (Fig. 4c, Supplementary Fig. 27 for plot of orientation distribution) even though a few edge views were observed as well (Supplementary Fig. 24). Although we were not able to obtain an *ab initio* model, a 3D reconstruction could be made by using the theoretical model as input search volume (Fig. 4b and Supplementary Fig. 24).

To investigate the robustness of the RNA origami method in relation to core design parameters we generated a challenging monomeric design with 5-helices and 2-KL columns, with an unconventional meandering strand path and generated two different versions with different 5’ start sites (Path1 shown in Fig. 4d and Path2 shown in Fig. 4e). The two strand paths are equivalent in 3D structure, but the different positioning of the 5’ start sites (Fig. 4d,e, blue circles) has a large effect on folding topology as predicted by RNApath. During transcription, Path1 has a long transient 5’ single strand but no predicted topological barriers (Fig. 4f, Supplementary Video 3), whereas Path2 has no transient 5’ single strand but has substantial topological barriers predicted (Fig. 4g, orange and red regions, Supplementary Video 4). Previously, we have avoided designs with a long 5’ transient single stranded region because transient single strands are expected to increase aggregation during cotranscriptional folding[7]. To investigate the effect of sequence design optimization, a third design was created based on Path1 satisfying the MFE structure (stage 1-4 of Revolvr, see Methods and Supplementary Fig. 1–2) but lacking the final KL optimization (stage 5 of Revolvr). As expected these designs display a substantial amount of aggregation that resulted in relatively low yield of monomers of 26-44% as determined by SEC analysis (Fig. 4k,l), which can be compared to 5HS which displays 86% monomer yield by SEC analysis (Fig. 4l and Supplementary Fig. 24). The monomers were observed to be stable post SEC purification (Fig. 4k) indicating that aggregation is happening during the cotranscriptional folding process and is not the result of a subsequent equilibration process.

TEM imaging was performed on the purified monomer and aggregate peaks and the monomers were observed to be monodisperse (Supplementary Fig. 25). To be able to

address the folding yield, the TEM grids were prepared from the same concentration of purified RNA samples and quantified from the same number of acquired images. Unbiased blob picking was used to identify particles and 2D class averages showed that the number of face views of the RNA origami structure were very different between the three samples and that several alternative particle views and shapes were observed (Fig. 4h–j and Supplementary Fig. 26). Since each design has the same predicted 3D structure, they should have the same angular distribution on the foils. We used the number of, easily recognizable, face views observed in the 2D class averages as an estimate of the cotranscriptional folding yield of the three samples (Fig. 4m, see Methods for description of folding yield calculation). Path1 with an optimized sequence had a 30% folding yield. Path2 with optimized sequence had 25% folding yield, but of these only 3/4 adopted the designed structure, whereas 1/4 displayed a “purse-handle” phenotype (Fig. 4i, blue arrow) that we suggest to correspond to distortions in the long topologically blocked helix (Fig. 4e) due to partial inhibition of Watson-Crick base pairing (Fig. 4g, blue arrow). Path1 with non-optimized KL sequences had a reduced folding yield of 6% (Fig. 4m) and a large fraction of alternative shapes (Fig. 4j right, Supplementary Fig. 26).

From the limited, but equivalent, datasets acquired for each design only the particles picked from the Path1 data produced a reasonable *ab initio* reconstruction (Fig. 4n). As observed previously in the TEM analysis of the 5HS, the Path1 structure had preferential face adsorption to the carbon foil, but in this case one face was strongly preferred (see Supplementary Fig. 27 for plot of orientation distribution), which indicates that the larger monomer structure has an asymmetric shape in solution that affects the adsorption to the carbon. Although the tested designs can all fold into the correct 3D structure, the choice of strand path and sequence optimization have large effects on both the yield and structural homogeneity of the origami particles.

Scaffolding of proteins and small molecules

To test the ability of RNA origami to scaffold proteins, we used the high-yield 5HS scaffold (Fig. 4a) containing 10 hairpin sites that can be used for functionalization (Fig. 5a). RNAbuild was used to design a series of five scaffolds that positioned two different protein-binding aptamers at increasing distances of approximately 2.5, 5, 7.5, 10, and 22 nm using scaffolds named M*x*P*y*, where *x* refers to the position of the MS2 aptamer[41] and *y* to the position of the PP7 aptamer[42] (Fig. 5b,c). All scaffolds were designed by Revolvr to have unrelated sequences, except for the fixed sequence of the aptamers. Similar to a previous scaffolding study[24], we fused mTurquoise2[50] (a Cyan Fluorescent Protein, CFP) and YPet[51] (a Yellow Fluorescent Protein, YFP) with viral coat proteins MS2 Coat Protein (MCP)[41] and PP7 Coat Protein (PCP)[42], respectively (Fig. 5b,c, sequences in Supplementary Table 8). When the M5P3 scaffold was transcribed in the presense of excess fluorescent proteins it resulted in a FRET signal that reached saturation after 20 min (Supplementary Fig. 28), showing that the scaffold cotranscriptionally folds and brings the two proteins together within FRET distance. To compare several RNA scaffolds, we normalized concentrations of co-transcriptionally folded RNA products and incubated them with excess amounts of fluorescent proteins. The FRET signal was observed to generally decrease with increasing distance between aptamers (Fig. 5d, full spectra in Supplementary Fig. 30), however, some

constructs with a spacing differing by ~2.5 nm were not significantly different in FRET signal (M5P4 ≈ M5P3; M5P2 ≈ M5P1, $p \geq 0.05$ in Student's t-test), and the control constructs, M5P10 with a nominal distance beyond the Förster radius and 5HS with no aptamers, showed measurable levels of FRET. These non-ideal effects may be explained by the large size of the fusion proteins with long linkers used as well as the documented tendency of the fluorescent proteins to form dimers in a colocalized context[52]. In general, the results may also be affected by scaffold flexibility and sequence-specific conformations of particular constructs.

RNAbuild was used to design two series of scaffolds that positioned the fluorescent aptamers Spinach[53, 54] and Mango[40] in various structural contexts (Fig. 5e–h, Supplementary Fig. 29). The first series was based on a two-helix scaffold S2T (short, 2 turns) with short stems to position Spinach and Mango aptamers and 2 helical turns between crossovers (Fig. 5e), which was previously shown to produce a strong FRET signal between the fluorophores DFHBI-1T and YO3-biotin[12]. Two variations of the S2T scaffold were produced: S3T (short, 3 turns) with wider crossover spacing (Fig. 5f), and L3T (long, 3 turns) with longer stems for positioning fluorescent aptamers and wider crossover spacing (Fig. 5g). The S2T scaffold transcribed in the presence of fluorophores shows slowly increasing fluorescence and FRET signals over at least 90 min (Supplementary Fig. 28), which is likely caused by the slow folding of the fluorescent aptamers. To compare several RNA scaffolds, we normalized the RNA concentrations before incubation with excess amount of fluorophores. Fluorescence measurements show ~35% FRET for S2T, ~30% FRET for S3T and ~5% FRET for L3T scaffolds (Fig. 5i, full spectra in Supplementary Fig. 30). While RNAbuild models predict that all three scaffolds have the same distance and orientation between donor and acceptor fluorophores (Fig. 5e–g), the large decrease in FRET signal with increasing construct size suggests that scaffold flexibility (due to longer stems and to a lesser extent larger crossover spacing) strongly influences the FRET signal. The second series was based on the three-helix scaffold from Fig. 2 with fluorescent aptamers placed on top and bottom helices and 2 turns between crossovers (L2T*s*DT in Fig. 5h, Supplementary Fig. 29). This scaffold is able to tune fluorophore spacing (from 1.3 to 3.2 nm in increments of 0.6 nm) by changing DT length (from s = -5 bp to -2 bp in increments of 32.7°), respectively. Fluorescence measurements for the L2T*s*DT scaffolds show a decrease in FRET signal as the predicted distance between the fluorophores is increased (Fig. 5i) (statistically significant $p<0.05$ in Student's t-test except for L2T-3DT). Within this series, care was taken to maintain the relative orientation of the Spinach and Mango aptamers to avoid the possible effects of oriented dipoles on FRET[55] (Supplementary Fig. 29). Comparing between series, we attribute the low FRET signal of the sterically overlapped construct L2T-5DT relative to construct S2T which shares a similar crossover spacing, primarily to the flexibility contributed from a longer aptamer bearing arm.

Discussion

The design and synthesis of cotranscriptional RNA structures in high yield is very challenging, and in our previous work[6] we were only able to achieve cotransctiptional folds of 440 nts in length with yields so low that only a few correctly formed objects could be identified. In the current work we have improved the RNA origami method to greatly expand

both the size and functional complexity of RNA nanostructure designs, as well as dramatically improving the yields of correct products that are able to be produced by cotranscriptional folding. We have rapidly prototyped 32 different RNA origami designs in this work, which allowed us to explore the effect of multiple RNA origami design parameters: DT geometry, multivalent interfaces, taller and wider structures, different strand routing strategies, as well as designs incorporating aptamers for scaffolding proteins and small molecules. The achievements were enabled by the development of the ROAD software package, comprising the programs RNAbuild, RNApath and Revolvr that work together to facilitate the design of large and complex RNA structures and were all found to be crucial for obtaining high-yield RNA scaffolds.

RNAbuild automates the rough 3D modeling of RNA origami structures, which were previously constructed by hand, allowing us to design much larger and more sophisticated designs than before. In this work we demonstrated that the DT seam can be used to adjust the curvature of RNA origami structures to tune tile-tile interaction angle to form rings of defined size (Fig. 2) and to tune the distance between attached fluorescent aptamers (Fig. 5h,i). RNAbuild further allowed us to expand the RNA origami architecture by domain duplication and fusion (Supplementary Fig. 8) reaching sizes of approximately 2000 nts albeit with decreasing yields as estimated from AFM images (Fig. 3; Supplementary Table 3). Interestingly, TEM analysis revealed preferred landing of larger RNA origami structures on the carbon film (Fig. 4h; Supplementary Fig. 27), which indicates that larger RNA origami structures may have a curved structure in solution. Even though *ab initio* reconstruction from TEM images showed that the RNA origami structures are flat this may be an artifact of the deposition on the carbon film. RNAbuild can in the future be improved by extending the library of functional motifs and by supporting for alternative architectures such as parallel crossover RNA origami[9] and wireframe RNA origami[8] as well as allowing physical simulation of the structures to address strain-induced distortions using e.g. oxRNA[46]. The recently developed program RNAmake—which specializes in grafting and stabilizing tertiary motifs onto an input model[31]—could complement and extend RNAbuild.

RNApath makes a simple folding path analysis based on the RNA blueprint (while not taking account of the designed sequence) to predict possible topological barriers for the cotranscriptional folding process. Comparing the number of predicted topological barriers to the folding yield estimation from AFM images revealed a strong correlation, where the most severe cases did not result in any correctly folded objects (Supplementary Table 3). However, the effect of size and number of topological barriers could not easily be separated in this evaluation, since topological barriers arise when designs get larger (and especially taller). The effect of folding path choice was investigated further by designing an RNA origami structure with two alternative folding paths. TEM analysis revealed that there was an approximately 30% decrease in folding yield for the path with topological barriers (Fig. 4m) and that misfolds could be observed with severe distortions of the topologically trapped helix (forming a “purse handle”) (Fig. 4i). The observation that only the structure without topological barriers resulted in a reasonable 3D reconstruction further underscores the importance of taking this design parameter into account (Fig. 4n). The kinetic folding analysis of RNApath may be improved by using thermodynamic kinetic folding algorithms like Kinefold[32] or by using coarse-grained molecular simulations like oxRNA[46].

Revolvr designs sequences for RNA blueprints with a high content of pseudoknots - a task that has not been approached by any other RNA design program. Revolvr solves this task by using a multi-stage sequence optimization procedure involving MFE-based positive design, SSM-based negative design, and KL orthogonalization, that makes it very efficient in the use of computational time (Supplementary Fig. 31). The sequence design by Revolvr has a high success rate, since most of the structures presented in this study worked on first try. The high success rate prompted us to design new sequences for each new RNA origami for scaffolding of proteins and small molecules with the assumption that the geometry of the RNA origami and not the precise sequence was important, which was verified to some extent by the overall ability to control distances on the scaffolds. The effect of sequence design was investigated by designing a non-optimal sequence, where only the last stage of the 5-step design procedure (KL orthogonalization) was omitted. Here the folding yield was observed to decrease from 30% to 6%, showing that this sequence design step has a substantial effect. Future improvements to Revolvr sequence design could be to include an RNA secondary structure partition function in the design optimization, like NUPACK[35]. The partition function optimization may become especially important when designing more challenging RNA structures with smaller stem regions. A great future challenge would also be to include pseudoknot-prediction and kinetic folding simulation directly in the sequence design algorithm.

While RNA nanostructures have previously been used to scaffold protein binding aptamers[10, 11] and small molecule aptamers[12, 13], here we demonstrate distance control by changing the position of aptamers on parallel helix ends and by tuning the DT length to gradually change the distance between helices. Our FRET studies highlight a potential size-flexibility tradeoff in RNA scaffold design: based on our current architecture, larger structures enable complex spatial arrangements of proteins to be constructed, but smaller, more rigid structures are required if more precise distances are desired. The elaboration of RNA origami to multilayer 3D structures may obviate this tradeoff by achieving simultaneously large and rigid structures, as has been achieved for DNA origami design[56]. Rigidity and precision of arrangement will also be improved by exchanging large, flexible, dimeric linkers such as MS2 and PP7 aptamer-protein constructs with smaller, monomeric RNA-binding proteins or peptides such as L7Ae[10] or BIV-Tat[57]. With improved protein scaffolding methods, the RNA origami scaffolds may be used to control product formation from colocalized enzymes[20, 21] and perform gene regulation via recruitment of transcription factors[22].

In this study we have improved the RNA origami method to allow the design of cotranscriptionally folding RNA nanostructures approaching the size of ribosomal RNAs. However, the structural complexity and strategies for cotranscriptional folding of RNA origami and the ribosome are very different. The ribosome is constructed from a high percentage of tertiary structural motifs, with almost 50% non-Watson-Crick base pairs. In contrast, the RNA origami architecture is mainly constructed from Watson-Crick base pairs formed by secondary structure elements and pseudoknots. The cotranscriptional folding of the ribosome involves transiently stable helices, protein chaperones and structural switches to guide the strand into a final native state that doesn’t correspond to the MFE. By contrast, RNA origami takes advantage of a very different, very unnatural design construction, in

which every helix of the design is able to rapidly find its MFE structure during the kinetic folding process. Thus, we are engineering very smoothed-out folding landscapes, with strand paths designed to minimize the possibility for the strand to misfold during the process. A recent computational study[58] suggests a general method for choosing strand-paths that minimizes the risk of topological barriers, and finds that KLs arranged into columns connected by a single common helix will result in the fewest topological barriers. While many of our designs have this property (minimized topological barriers), this work deserves to be further explored quantitatively; and determining whether or not RNA origami contain minor misfolded elements may require the adaptation of SHAPE-seq[59] or other techniques to very large RNA structures, or perhaps high-resolution cryo-EM. Future challenges for the cotranscriptional RNA origami method will be to increase the structural complexity by 3D architectures and tertiary motifs. Likewise, the ability to program RNA origami scaffolds with functional, dynamic features and molecular computing elements (like strand displacement logic gates), as have been achieved with DNA origami structures, to create biosensor devices and nanorobots is still to be explored.

Methods

RNA Origami Automated Design (ROAD) package

To automate key processes of RNA origami design, a set of algorithms were made that together constitutes a design pipeline: RNAbuild – Builds a PDB structure from an RNA blueprint. RNApath – Analyses the folding path of a given RNA blueprint and highlights topological barriers. Animations can be generated in the form of a series of keyframes and a Chimera command file for automatic generation of a movie. Revolvr – Designs RNA sequences that fold into target structures (requires installation of the Vienna RNA package https://www.tbi.univie.ac.at/RNA/). Trace – Takes an RNA blueprint without sequence assigned to it, and a candidate sequence, and creates a new blueprint with the candidate sequence threaded onto the blueprint. Trace_pattern – Converts RNA blueprint diagrams into dot-paren notation and candidate sequence for Revolvr. Trace_analysis – Analyzes an input blueprint and annotates it with features such as duplicated sequences, unintended complementary sequences, GC content, restriction sites, etc. Flip_trace – Flips a blueprint horizontally, vertically, and in both directions, which aids in the design of more complex patterns using domain duplication and fusion. The analysis package is available for download at GitHub (https://github.com/esa-lab/ROAD/) and has been made available as a webserver with accompanying tutorials (https://bion.au.dk/software/rnao-design/).

RNAbuild

The RNAbuild algorithm automates structural modeling of RNA origami structures. As input the algorithm takes an RNA blueprint and parses the RNA structure from 5’ to 3’ end to identify a set of predefined 2D motifs. The 3D atomic model is constructed from 5’ to 3’ end by serial addition of 3D structural modules from a library that matches the 2D motifs and 3D structural modules. When each structural module is added to a structure, it is rotated and translated to the correct position using a single reference base or base pair having A-form parameters (Supplementary Table 1 and Supplementary Fig. 3); these reference bases are added to modules when they are put in the library (Supplementary Fig. 1). For building

RNA helices, there are four different trivial nucleotide modules, simply the PDB coordinates for each RNA nucleotide. Modules for KLs, terminal tetraloops, fluorogenic aptamers, and the RNA-protein binding domains are all based on known crystal structures. The helix axes of the crossover module are modeled as parallel, rather than at the 60°-70° angle found in crystal structures (see PDB-ID: 1HP6)[60], under the assumption that the coupling of adjacent crossovers forces them flat. The spacing of crossovers between three or more parallel helices defines the geometry of the DT junction and RNAbuild helps model these.

RNApath

The RNApath algorithm analyses the folding path of RNA origami designs to identify possible topological barriers during the cotranscriptional folding process. The algorithm takes an RNA blueprint as input and analyses the RNA structure as it is being extended from 5’ to 3’ end. For each subsequence of length k, RNApath takes the fold computed for the k-1 subsequence, and decides what new base pairs can be added to the fold. RNApath adds a secondary structure (e.g. a particular hairpin) to the fold for the subsequence having the smallest k possible, which models the situation that secondary structure folds immediately, as soon as the necessary sequence is transcribed. By default, a particular KL is added to the fold of a subsequence k only when k is at least 150 nt longer than the smallest subsequence that contains both halves of the KL. This feature roughly captures the KL formation time, modeling the situation that KL formation is delayed by ~0.7 s (assuming a transcription speed of 4.3 ms/nt)[61] after the KL sequence has been transcribed. Where the folding of KLs might topologically clash with secondary structure formation, RNApath labels barrier loops by ‘~’ and topologically blocked nts by ‘X’ in an analysis blueprint output. It additionally outputs a list of substructures in dot-paren where transient single strands are shown as ‘,’ and crossovers are shown as ‘^’. The delay is adjustable from 0 nt (for which almost all KLs cause clashes), to N nt (for which no clashes will occur). In addition, RNApath can output a series of PDB models which can be rendered to create a movie in UCSF Chimera v1.10 where pseudoknot loops are colored orange and topologically blocked nts are colored red (Supplementary Videos 1–6 and Note 1). Alternatively, the program trace_analysis provides a fast summary of any patterns in the sequence as well as positions of wobbles within the design, and lastly the strand path analysis.

Revolvr

The Revolvr algorithm designs sequences for target structures by using a five-stage variant of stochastic gradient descent where each stage has an increasingly restrictive cost function (see algorithm flowchart in Supplementary Fig. 1 and 2). The cost function is a score that combines the minimum free energy (MFE) folding prediction and different measures of sequence symmetry, and for each successive round of design becomes stricter. The input file defines the target secondary structure, pseudoknots and sequence constraints, and is initially seeded with a random sequence that satisfies the constraints (or with user inputted sequence). The first stage optimizes the MFE structure over five rounds of positive design to stabilize helix ends and multi-junctions at a cost of raising the GC content. The current sequence’s MFE structure, as computed by the ViennaRNA package[44] is used to calculate the Hamming distance of the current structure to the target structure, which is used as the cost function score for this round of design. The second stage applies alternating positive

and neutral design for a variable number of rounds, until the target structure is achieved: mutations are either targeted to regions that misfolded in the previous round (two out of every three rounds) or spread randomly throughout the sequence to enable neutral drift (every third round). To increase the speed of design, we scale the rate of mutation per design based on the success/failure of each iteration. A design round is considered successful when the cost function remains the same or decreases. The third stage uses negative design to decrease the probability of misfolding, prevent the inclusion of particular sequences (e.g. restriction sites) where undesired, and make the DNA template from which RNA origami are transcribed easier to synthesize and PCR amplify. Sequence symmetry minimization[62] (SSM) is applied to remove all repeated sequences or regions of undesired complementarity above a threshold length (default setting = 10 nt). Similarly, removal of long homopolymer stretches, and a known transcriptional pause site[63], may help to reduce the frequency of unwanted transcriptional termination. The fraction of GC bp is reduced to below 55%, to encourage correct folding at 37°C. Finally, GU wobble pairs are introduced to simultaneously preserve the helix within the desired RNA structure, and weaken it within the corresponding DNA templates. All of the above constraints are applied through successive rounds of targeted mutation, until they and the MFE fold are simultaneously satisfied. The fourth stage eliminates repetition from the set of KLs. Repeated (and thus also palindromic) KL sequences are targeted for mutation in successive rounds until all KL interactions are unique. The fifth stage optimizes the sequences of the KLs to have uniform binding energy and greater specificity. Energies for all possible KL interactions are estimated with the Duplex function of ViennaRNA[44] and KLs are targeted for mutation until all desired KLs have energies between -10.7 and -7.2 kcal/mol, and all undesired KL interactions have energies greater than -6.0 kcal/mol. Revolvr enables potentially conflicting requirements for positive vs negative design, sequence vs secondary structure constraints, and pseudoknotted vs non-pseudoknotted structure to be balanced and satisfied. User-specified sequence constraints supersedes user-specified secondary structure, which supersedes all other constraints. Sequences explicitly specified in a blueprint (such as aptamers) are left unmutated, even if the secondary structure specified for them cannot be achieved in an MFE structure, or if they violate a sequence symmetry constraint. Upon termination, Revolvr outputs an analysis of the designed sequence which includes a blueprint populated with the sequence, KL energies, and the position of potential topological clashes, violations of sequence symmetry constraints, and GU wobbles.

Synthesis of RNA origami structures

DNA templates were commercially synthesized (Integrated DNA Technologies) as double-stranded gBlocks. DNA gBlocks were PCR-amplified using 19-20 nt primers ($T_m \approx 56°C$) complementary to the ends of the gBlock, using standard Taq DNA polymerase, and purified using a Qiagen PCR purification kit. RNAs were transcribed and cotranscriptionally folded in a one-pot reaction containing: template DNA (~4 ng/μl final of PCR amplicon), 6 mM $Mg(OAc)_2$, 40 mM Na OAc, 40 mM KCl, 50 mM Tris-OAc (pH 7.8), rNTPs (0.5 mM each) and 1 mM DTT. Reactions were initiated by adding T7 RNA polymerase (~0.2 U/50 μl). Transcription reactions were carried out in 50 μl volumes at 37.0°C for 45 minutes to 2 hrs, depending on the sequence length. Larger designs required longer synthesis time (1-2

hours), whereas smaller designs (e.g. 2AE) required just a few minutes to reveal multimeric products by AFM.

AFM sample preparation and imaging

1-5 μl of transcription product was mixed with 40 μl AFM dilution buffer (12.5 mM $Mg(OAc)_2$, 40mM KCl, 40mM NaCl, Tris-Borate pH 7.8) directly on the surface of a freshly-cleaved mica puck. Mixing is performed by vigorously pumping a 200 μl pipette tip ten times, before removing and discarding the fluid. The mica was washed with a solution of 60 mM $NiCl_2$. Most AFM images were collected using a Multimode AFM (Digital Instruments) with a Nanoscope IIIA controller and a J-scanner. Olympus TR400PSA silicon nitride probes with a spring constant of ~.08 N/m were used for imaging, with a drive frequency of ~6-9 kHz. AFM in Supplementary Figs. 10 and 11 were collected with a Bruker Fastscan Bio AFM (Bruker) under buffer using FastScan-D probes (Bruker).

Purification of RNA Origami

RNA origamis were transcribed from linearized pUC19 plasmid for large-scale synthesis and purification. Briefly, 25 μg of linearized plasmid was used as template in a 0.5 mL reaction containing 40 mM Tris-Cl pH 8.1, 1 mM spermidine, 0.001% Triton X-100, 100 mM DTT, 12 mM $MgCl_2$, 0.1 mM $CaCl_2$, 0.5X ribolock (Thermo Fisher Scientific) and in house prepared T7 polymerase. Transcription was carried out at 37°C for 3 hours before the addition of 40 μL of DNase I (NEB). After 30 minutes of DNA digestion the reaction was centrifuged at 17,000 RCF (x g) for 10 minutes to pellet precipitated pyrophosphate. The supernatant was loaded onto a Superose 6 column (GE) equilibrated with 25 mM Hepes pH 7.5, 50 mM KCl and 5 mM $MgCl_2$.

Negative Stain TEM

CF400 Au grids (Electron Microscopy Sciences) were glow discharged for 45 seconds at 25 mA before application of 3 μL of sample and then thrice blotted with 3 μL of 0.5% uranyl formate. Peak 2 from the 5HS purification was diluted to 25ng/μL prior to blotting and Peak 2 from the Path1-optimized, Path2-optimized and Path1-nonoptimized purifications were diluted to 50ng/μL. TEMImages were obtained on a 120 kV Tecnai Spirit TEM equipped with a 4K TVIPS CMOS camera at 67k magnification. The images were contrast inverted and converted from tif to mrc using Eman2[64] before being imported into CryoSparc V2.0[65]. CTF correction was applied with CTFFIND4[66]. For 3D reconstructions, ~300 particles were manually picked in Cryosparc and used to generate templates for the first round of templated particle picking. These particles were sorted into 50 2D classes, the best of which were used for 3D reconstructions. *Ab initio* 3D reconstruction of the Path1-optimized design produced a volume with real space slices similar to what would be expected from our design and was further refined by a non-uniform refinement. *Ab initio* 3D reconstruction repeatedly failed for the 5HS design and so a homogeneous refinement of the 5HS structure was performed using a 40 Å masked volume of our predicted RNA origami structure as an initial volume. For the comparison of the Path1-optimized, Path2-optimized and Path1-nonoptimized datasets blob picking was performed on 88 images of each design with the default settings in CryoSparc V2.0, followed by a single round of 2D class averaging into 50 classes. Structural deformities observed were measured with Eman2.

Europe PMC Funders Author Manuscripts

Folding yield calculation

The SEC yield was calculated as the monomer to aggregate fraction based on the average peak heights of the SEC chromatogram (UV 255 nm) for 2-3 transcription reactions. The TEM folding yield was estimated by counting the number of face views of the RNA origami particles based on the assumption that correctly folded particles would have a similar preference of adsorbing to the carbon film in this orientation. The number of face views were adjusted to the amount of ng of RNA loaded on the grids, the number of images obtained, and the molar mass. The folding yield of the 5HS was assumed to be 95% based on analysis of TEM and AFM images (data not shown) and was used to calculate the relative folding yield of the other samples. The transcription folding yield was calculated by multiplying the SEC monomer yield with the TEM folding yield. An alternative fold, named the “purse-handle”, was identified by measurement of a helix gap of minimum 1.5 nm in the TEM class average images of the Path2-optimized sample accounting for 25.3% of the face views.

Protein design, expression, and purification

MCP-ΔFG/V29I[67] and PCP-ΔFG[68] were codon-optimized for expression in *E. coli*. Expression plasmids pJ431 encoding for His(6)-tagged mTurquoise2-MS2 Coat Protein (mTq-MCP) or His(6)-tagged YPet-PP7 Coat Protein (YPet-PCP) under the control of an IPTG-inducible T7 RNA polymerase promoter were ordered from ATUM (USA). Proteins were expressed in E. coli BL21 Star (DE3). Cells were inoculated from a single colony in Luria Bertani (LB) medium with Kanamycin and grown at 37°C overnight, under shaking. The next day media was refreshed and cells were grown at 37°C under shaking to a density of 0.3 OD/ml. Cells were then induced with 1 mM IPTG and incubated at 29°C under shaking for 5 hours. Induced cells were harvested and sonicated with a Q125 microtip sonicator (Qsonica Sonicators, USA). Both His(6)-tagged proteins were purified by gravity-flow chromatography with TALON® Metal Affinity Resin (Takara Bio Inc, Japan). After purification, the proteins were dialyzed overnight at 4°C using a Spectra-Por® Float-A-Lyzer® G2 dialysis device (8-10 kDa MWCO; Spectrum Labs Inc, USA) in Protein Storage Buffer (25 mM Tris/HCl (pH 8.0) + 0.3 M NaCl) and stored at 4°C.

Fluorescent proteins experiments

For protein scaffolding experiments fluorescence measurements were performed on a VarioskanFlash 4 (Thermofisher). Excitation of mTurquoise2-MCP and YPet-PCP were performed at 434 nm and 505 nm, respectively. Emissions of mTurquoise2-MCP and YPet-PCP were recorded at 474 nm and 525 nm, respectively. Excitation bandwidths were set to 5 nm, and measurement time was 0.1 s. Measurement during transcription was done for a 58 µl transcription reaction containing transcription mix, 100 ng of DNA template, and protein concentrations of 500 nM each. Transcription was started by adding 2 µl NTPs (25 mM each) and measured every 5 min for 50 min (Supplementary Fig. 28). To compare several constructs, RNA was transcribed (NEB T7 RNApol protocol, incubated overnight at 37°C) and the reaction was stopped by adding DNase I (NEB) at 10 U/100 µl and incubated for 45 min at 37°C. Produced RNA was quantified on denaturing PAGE using a Typhoon Laser Scanner (Amersham). Florescence measurement was performed on 60 µl samples with 330

nM RNA and protein concentration of 500 nM, each incubated for 5 min to allow proteins to bind the scaffolds.

Fluorescent aptamer experiments

RNA was prepared as described in the previous section with the addition of 100 mM KCl to the transcription reactions to facilitate folding of G quadruplex aptamers. DFHBI-1T was purchased from Lucerna Technologies (USA) and YO3-biotin was purchased as custom synthesis from Apigenex (Czech Republic). Fluorescence measurements were performed on a FluoroMax 4 (Horiba, Jobin Yvon) by excitation of DFHBI-1T and YO3-biotin at 450 nm and 580 nm, respectively. Emissions of DFHBI-1T and YO3-biotin were recorded at 503 nm and 620 nm, respectively. Monochromator slits were set to 5 nm, and integration time was 0.2 s. Measurement during transcription was done for a 58 μl sample of transcription mix containing 100 mM KCl, 30 ng of DNA template, 2 μM DFHBI-1T and 10 μM YO3-biotin. Transcription was started by adding 2 μl NTPs (25 mM each) and measurements were performed every 5 min for 90 min (Supplementary Fig. 28). To compare several constructs, the produced RNA was quantified on denaturing PAGE as described above. Florescence measurement was performed on 60 μl samples with 150 nM RNA incubated at RT for 20 min with 2 μM DFHBI-1T and 10 μM YO3-biotin.

Ensemble FRET calculations

The emission intensity arising from the donor tail at acceptor wavelength was calculated to obtain the leak of the donor emission using the equation: $D_{leak}=I_D(ex_D,em_A)/I_D(ex_D,em_D)$, where $I_D(ex_D,em_A)$ is the emission at acceptor wavelength after donor excitation with only donor present, and $I_D(ex_D,em_D)$ is the emission at donor wavelength after donor excitation with only donor present. Relative FRET values were calculated using the equation:

$$FRET\ Output = \frac{I_{DA}(ex_D, em_A) - D_{leak} * I_{DA}(ex_D, em_D)}{I_{DA}(ex_D, em_A) - D_{leak} * I_{DA}(ex_D, em_D) + I_{DA}(ex_D, em_D)}$$

where $I_{DA}(ex_D,em_A)$ is the emission at acceptor wavelength after donor excitation, and $I_{DA}(ex_D, em_D)$ is the emission at donor wavelength after donor excitation with both donor and acceptor present.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Lulu Qian and Erik Winfree for the use of their atomic force microscopes. Grigory Tikhomirov for help with AFM and Mette Jepsen for help with FRET. We acknowledge the EteRNA community for conducting an experiment which suggested that kissing loop sequences are less constrained than previously assumed; this inspired us to add *de novo* design of KLs to Revolvr. C.G. acknowledges a fellowship from the Carlsberg Research Foundation. E.K.S.M. acknowledges the Natural Sciences and Engineering Research Council of Canada for his post doctoral fellowship. P.W.K.R. acknowledges funding by NSF grants (CCF-1317694 and CMMI-1636364) and ONR grants (N00014-16-1-2159, N00014-17-1-2610, and N00014-18-1-2649). E.S.A. acknowledges funding by the ERC Consolidator Grant (RNA ORIGAMI - RNA-protein nanostructures for synthetic biology, 683305) that supported the work of C.G., G.G. and E.K.S.M. and the Independent Research Fund Denmark (9040-00425B) that supported the work of E.K.S.M.

Europe PMC Funders Author Manuscripts

Data availability

The data supporting the findings of this study are further documented in the associated Supplementary Information. All raw data and analysis files used in the study are available upon request from the authors.

Code availability

The code used to generate RNA origami designs in this study is included in the associated Supplementary Information. Future updates to the code will be made available on Github (https://github.com/esa-lab/ROAD) and on a dedicated web server with accompanying tutorials (https://bion.au.dk/software/rnao-design/). The code is licensed under the MIT License.

References

1. Grabow WW, Jaeger L. RNA self-assembly and RNA nanotechnology. Acc Chem Res. 2014; 47:1871–1880. [PubMed: 24856178]
2. Jasinski D, Haque F, Binzel DW, Guo P. Advancement of the Emerging Field of RNA Nanotechnology. ACS Nano. 2017; 11:1142–1164. [PubMed: 28045501]
3. Weizmann Y, Andersen ES. RNA nanotechnology—The knots and folds of RNA nanoparticle engineering. MRS Bull. 2017; 42:930–935.
4. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science. 2000; 289:905–920. [PubMed: 10937989]
5. Wimberly BT, et al. Structure of the 30S ribosomal subunit. Nature. 2000; 407:327–339. [PubMed: 11014182]
6. Geary C, Rothemund PWK, Andersen ES. A single-stranded architecture for cotranscriptional folding of RNA nanostructures. Science. 2014; 345:799–804. [PubMed: 25124436]
7. Geary C, Chworos A, Verzemnieks E, Voss NR, Jaeger L. Composing RNA Nanostructures from a Syntax of RNA Structural Modules. Nano Lett. 2017; 17:7095–7101. [PubMed: 29039189]
8. Li M, et al. In vivo production of RNA nanostructures via programmed folding of single-stranded RNAs. Nat Commun. 2018; 9:2196. [PubMed: 29875441]
9. Han D, et al. Single-stranded DNA and RNA origami. Science. 2017; 358
10. Shibata T, et al. Protein-driven RNA nanostructured devices that function in vitro and control mammalian cell fate. Nat Commun. 2017; 8:540. [PubMed: 28912471]
11. Krissanaprasit A, et al. Genetically Encoded, Functional Single-Strand RNA Origami: Anticoagulant. Adv Mater. 2019; 31 e1808262 [PubMed: 30972819]
12. Jepsen MDE, et al. Development of a genetically encodable FRET system using fluorescent RNA aptamers. Nat Commun. 2018; 9:18–18. [PubMed: 29295996]
13. Chopra A, Sagredo S, Grossi G, Andersen ES, Simmel FC. Out-of-Plane Aptamer Functionalization of RNA Three-Helix Tiles. Nanomaterials (Basel). 2019; 9
14. Porter EB, Polaski JT, Morck MM, Batey RT. Recurrent RNA motifs as scaffolds for genetically encodable small-molecule biosensors. Nat Chem Biol. 2017; 13:295–301. [PubMed: 28092358]
15. Oi H, et al. Programmable formation of catalytic RNA triangles and squares by assembling modular RNA enzymes. J Biochem. 2017; 161:451–462. [PubMed: 28096453]
16. Afonin KA, et al. Design and self-assembly of siRNA-functionalized RNA nanoparticles for use in automated nanomedicine. Nat Protoc. 2011; 6:2022–2034. [PubMed: 22134126]
17. Cui D, et al. Regression of Gastric Cancer by Systemic Injection of RNA Nanoparticles Carrying both Ligand and siRNA. Sci Rep. 2015; 5 10726 [PubMed: 26137913]
18. Afonin KA, et al. Co-transcriptional assembly of chemically modified RNA nanoparticles functionalized with siRNAs. Nano Lett. 2012; 12:5192–5195. [PubMed: 23016824]

19. Myhrvold C, Silver PA. Using synthetic RNAs as scaffolds and regulators. Nat Struct Mol Biol. 2015; 22:8–10. [PubMed: 25565027]
20. Delebecque CJ, Lindner AB, Silver PA, Aldaye FA. Organization of Intracellular Reactions with Rationally Designed RNA Assemblies. Science. 2011; 333:470–474. [PubMed: 21700839]
21. Sachdeva G, Garg A, Godding D, Way JC, Silver PA. In vivo co-localization of enzymes on RNA scaffolds increases metabolic production in a geometrically dependent manner. Nucleic Acids Res. 2014; 42:9493–9503. [PubMed: 25034694]
22. Zalatan JG, et al. Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. Cell. 2015; 160:339–350. [PubMed: 25533786]
23. Delebecque CJ, Silver PA, Lindner AB. Designing and using RNA scaffolds to assemble proteins in vivo. Nat Protoc. 2012; 7:1797–1807. [PubMed: 22955695]
24. Schwarz-Schilling M, et al. Optimized Assembly of a Multifunctional RNA-Protein Nanostructure in a Cell-Free Gene Expression System. Nano Lett. 2018; 18:2650–2657. [PubMed: 29564885]
25. Rogers TA, Andrews GE, Jaeger L, Grabow WW. Fluorescent monitoring of RNA assembly and processing using the split-spinach aptamer. ACS Synth Biol. 2015; 4:162–166. [PubMed: 24932527]
26. Liu D, et al. Branched kissing loops for the construction of diverse RNA homooligomeric nanostructures. Nat Chem. 2020; 12:249–259. [PubMed: 31959958]
27. Andersen ES. Prediction and design of DNA and RNA structures. N Biotechnol. 2010; 27:184–193. [PubMed: 20193785]
28. Jabbari H, Aminpour M, Montemagno C. Computational Approaches to Nucleic Acid Origami. ACS Comb Sci. 2015; 17:535–547. [PubMed: 26348196]
29. Bindewald E, Grunewald C, Boyle B, O'Connor M, Shapiro BA. Computational strategies for the automated design of RNA nanoscale structures from building blocks using NanoTiler. J Mol Graph Model. 2008; 27:299–308. [PubMed: 18838281]
30. Jossinet F, Ludwig TE, Westhof E. Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. Bioinformatics. 2010; 26:2057–2059. [PubMed: 20562414]
31. Yesselman JD, et al. Computational design of three-dimensional RNA structure and function. Nat Nanotechnol. 2019; 14:866–873. [PubMed: 31427748]
32. Xayaphoummine A, Bucher T, Isambert H. Kinefold web server for RNA/DNA folding path and structure prediction including pseudoknots and knots. Nucleic Acids Res. 2005; 33:W605–610. [PubMed: 15980546]
33. Zhao P, Zhang W, Chen SJ. Cotranscriptional folding kinetics of ribonucleic acid secondary structures. J Chem Phys. 2011; 135 245101 [PubMed: 22225186]
34. Churkin A, et al. Design of RNAs: comparing programs for inverse RNA folding. Brief Bioinform. 2018; 19:350–358. [PubMed: 28049135]
35. Zadeh JN, et al. NUPACK: Analysis and design of nucleic acid systems. J Comput Chem. 2011; 32:170–173. [PubMed: 20645303]
36. Ennifar E, Walter P, Ehresmann B, Ehresmann C, Dumas P. Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. Nat Struct Biol. 2001; 8:1064–1064. [PubMed: 11702070]
37. Fiore JL, Nesbitt DJ. An RNA folding motif: GNRA tetraloop–receptor interactions. Q Rev Biophys. 2013; 46:223–264. [PubMed: 23915736]
38. Lee AJ, Crothers DM. The solution structure of an RNA loop-loop complex: the ColE1 inverted loop sequence. Structure. 1998; 6:993–1005. [PubMed: 9739090]
39. Paige JS, Wu KY, Jaffrey SR. RNA Mimics of Green Fluorescent Protein. Science. 2011; 333:642–646. [PubMed: 21798953]
40. Dolgosheina EV, et al. RNA mango aptamer-fluorophore: a bright, high-affinity complex for RNA labeling and tracking. ACS Chem Biol. 2014; 9:2412–2420. [PubMed: 25101481]
41. Convery MA, et al. Crystal structure of an RNA aptamer-protein complex at 2.8 A resolution. Nat Struct Biol. 1998; 5:133–139. [PubMed: 9461079]

42. Lim F, Peabody DS. RNA recognition site of PP7 coat protein. Nucleic Acids Res. 2002; 30:4138–4144. [PubMed: 12364592]
43. Geary, CW, Andersen, ES. Design Principles for Single-Stranded RNA Origami StructuresDNA Computing and Molecular Programming 2014. Murata, S, Kobayashi, S, editors. Springer International Publishing; Cham: 2014. 1–19.
44. Lorenz R, et al. ViennaRNA Package 2.0. Algorithms Mol Biol. 2011; 6:26. [PubMed: 22115189]
45. Seeman NC. De novo design of sequences for nucleic acid structural engineering. J BioMol Struct Dyn. 1990; 8:573–581. [PubMed: 2100519]
46. Sulc P, Romano F, Ouldridge TE, Doye JP, Louis AA. A nucleotide-level coarse-grained model of RNA. J Chem Phys. 2014; 140 235102 [PubMed: 24952569]
47. Chworos A, et al. Building Programmable Jigsaw Puzzles with RNA. Science. 2004; 306:2068–2072. [PubMed: 15604402]
48. Jaeger L, Chworos A. The architectonics of programmable RNA and DNA nanostructures. Curr OpiN Struct Biol. 2006; 16:531–543. [PubMed: 16843653]
49. Grainger RJ, Murchie AI, Lilley DM. Exchange between stacking conformers in a four-Way DNA junction. Biochemistry. 1998; 37:23–32. [PubMed: 9425022]
50. Goedhart J, et al. Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. Nat Commun. 2012; 3:751. [PubMed: 22434194]
51. Nguyen AW, Daugherty PS. Evolutionary optimization of fluorescent proteins for intracellular FRET. Nat Biotechnol. 2005; 23:355–360. [PubMed: 15696158]
52. Ohashi T, Galiacy SD, Briscoe G, Erickson HP. An experimental study of GFP-based FRET, with application to intrinsically unstructured proteins. Protein Sci. 2007; 16:1429–1438. [PubMed: 17586775]
53. Autour A, Westhof E, Ryckelynck M. iSpinach: a fluorogenic RNA aptamer optimized for in vitro applications. Nucleic Acids Res. 2016; 44:2491–2500. [PubMed: 26932363]
54. Fernandez-Millan P, Autour A, Ennifar E, Westhof E, Ryckelynck M. Crystal structure and fluorescence properties of the iSpinach aptamer in complex with DFHBI. RNA. 2017; 23:1788–1795. [PubMed: 28939697]
55. Jeng SS, et al. Fluorogenic aptamers resolve the flexibility of RNA junctions using orientation-dependent FRET. RNA. 2020
56. Douglas SM, et al. Self-assembly of DNA into nanoscale three-dimensional shapes. Nature. 2009; 459:414–418. [PubMed: 19458720]
57. Bayer TS, Booth LN, Knudsen SM, Ellington AD. Arginine-rich motifs present multiple interfaces for specific binding by RNA. RNA. 2005; 11:1848–1857. [PubMed: 16314457]
58. Mohammed, A, Orponen, P, Pai, S. Algorithmic Design of Cotranscriptionally Folding 2D RNA Origami StructuresUnconventional Computation and Natural Computation 2018. Stepney, S, Verlan, S, editors. Cham: Springer International Publishing; 2018. 159–172.
59. Lucks JB, et al. Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). Proc Natl Acad Sci U S A. 2011; 108:11063–11068. [PubMed: 21642531]
60. Rupert PB, Massey AP, Sigurdsson ST, Ferre-D'Amare AR. Transition state stabilization by a catalytic RNA. Science. 2002; 298:1421–1424. [PubMed: 12376595]
61. Golomb M, Chamberlin M. Characterization of T7-specific ribonucleic acid polymerase. IV. Resolution of the major in vitro transcripts by gel electrophoresis. J Biol Chem. 1974; 249:2858–2863. [PubMed: 4828324]
62. Seeman NC. Nucleic-Acid Junctions and Lattices. J Theor Biol. 1982; 99:237–247. [PubMed: 6188926]
63. Lyakhov DL, et al. Pausing and termination by bacteriophage T7 RNA polymerase. J Mol Biol. 1998; 280:201–213. [PubMed: 9654445]
64. Tang G, et al. EMAN2: an extensible image processing suite for electron microscopy. J Struct Biol. 2007; 157:38–46. [PubMed: 16859925]
65. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods. 2017; 14:290–296. [PubMed: 28165473]

66. Rohou A, Grigorieff N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. J Struct Biol. 2015; 192:216–221. [PubMed: 26278980]

67. Lim F, Peabody DS. Mutations that increase the affinity of a translational repressor for RNA. Nucleic Acids Res. 1994; 22:3748–3752. [PubMed: 7937087]

68. Chao JA, Patskovsky Y, Almo SC, Singer RH. Structural basis for the coevolution of a viral RNA-protein complex. Nat Struct Mol Biol. 2008; 15:103–105. [PubMed: 18066080]

Europe PMC Funders Author Manuscripts

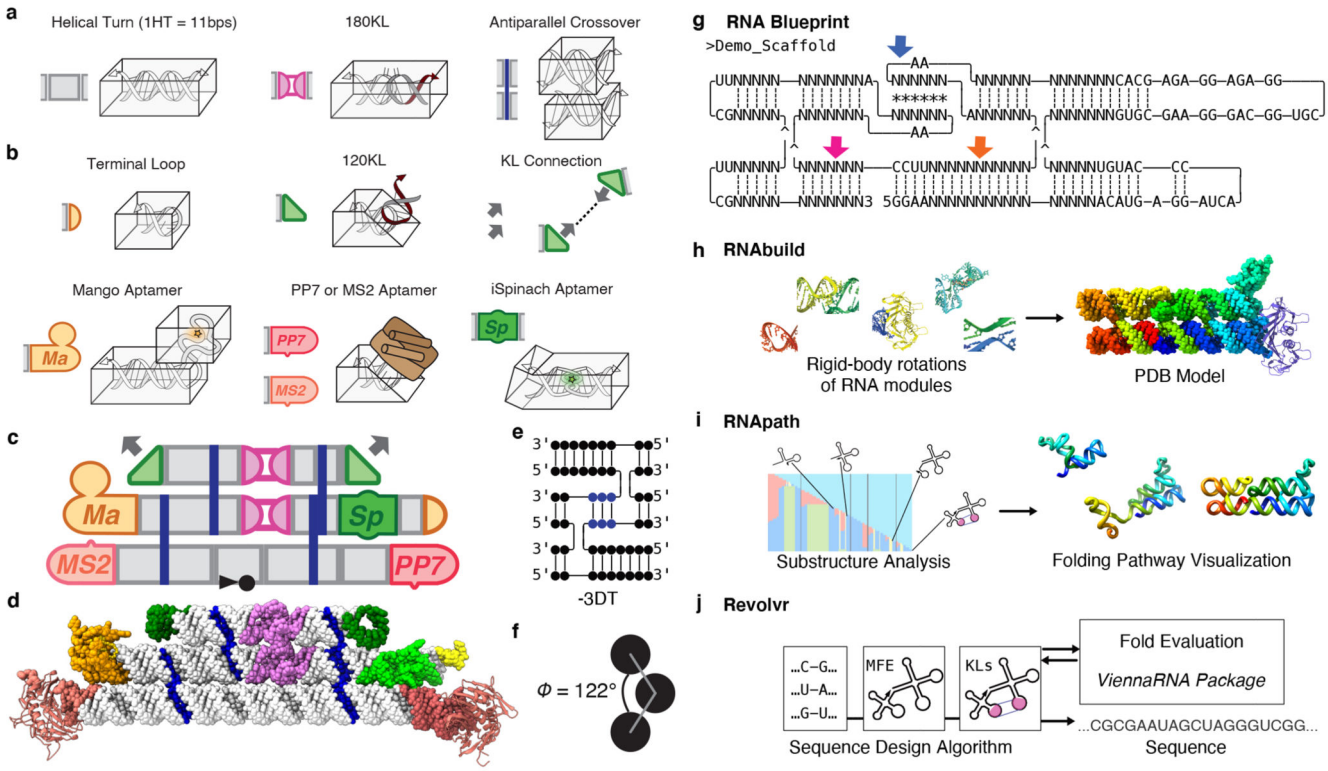

Figure 1. Modular design of RNA origami.

a, Schematics and models of core modules that compose the bulk of an RNA origami. **b,** Edge modules that compose the periphery of an RNA origami. **c,** Modular schematic of a three-helix tall origami. Modules are arranged so that they form a single continuous strand, with 5’ and 3’ ends indicated by the black circle and arrow, respectively. **d,** Atomic model with module coloring corresponding to panel c. **e,** Dovetail junction between three parallel helices with strand directions and central stem in blue. **f,** Side view of dovetail junction helices with indication of dihedral angle *Φ*. **g,** Example RNA blueprint, a text-based input format for all components of our software suite. Asterisks: base pairs within KL pseudoknots. Dashed lines: all other base pairs. Blue arrow denotes KL interaction; pink and orange arrows denote helices in the loop region. **h,** RNAbuild parses blueprints into modules and produces molecular models in PDB format. **i,** RNApath analyzes blueprints and produces a visualization of the order in which helical domains and KLs form, flagging regions which might be susceptible to misfolding. **j,** Revolvr takes sequence constraints and secondary structure from blueprints and uses random mutation to generate sequences that simultaneously satisfy the sequence constraints and are predicted to fold into the desired structure.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

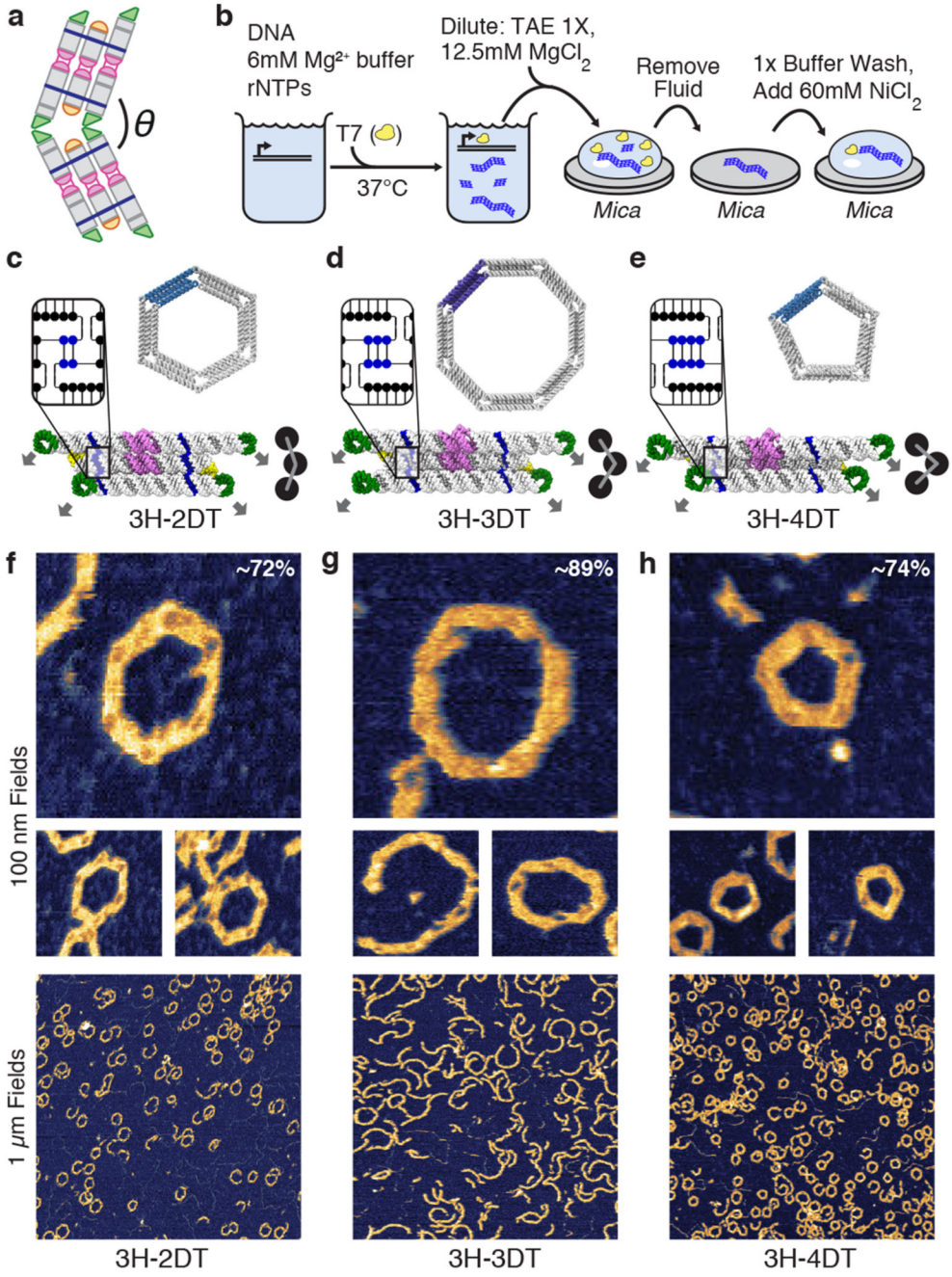

Figure 2. Design of RNA origami curvature and tile-tile interfaces.

a, Schematic model tile-tile interaction and angle *θ*. **b,** Protocol comprising isothermal synthesis at 37°C by T7 polymerase, cotranscriptional folding, and sample preparation for AFM. $NiCl_2$ binds RNA strongly to mica, improving image quality. **c-e,** RNAbuild models of three-helix designs having DTs of different length (named 3H-*s*DT, where *s* indicates the length of the DT); each 1 bp increase in DT length decreases the dihedral angle between the three helices (black cross sections) by 32.7°. Insets show secondary structure with DT marked in blue. Most prevalent closed polygons are shown: hexamer for 3H-2DT, octamer

for 3H-3DT, pentamer for 3H-4DT. **f-h,** AFM images of RNAs corresponding to 100-nm and 1-μm fields. Three-helix designs 3H-2DT (**f**), 3H-3DT (**g**) and 3H-4DT (**h**) are shown. Thin filaments in the background are DNA templates from which the RNA structures are transcribed. Yield of tiles shown on top of AFM image are from Supplementary Table 3.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

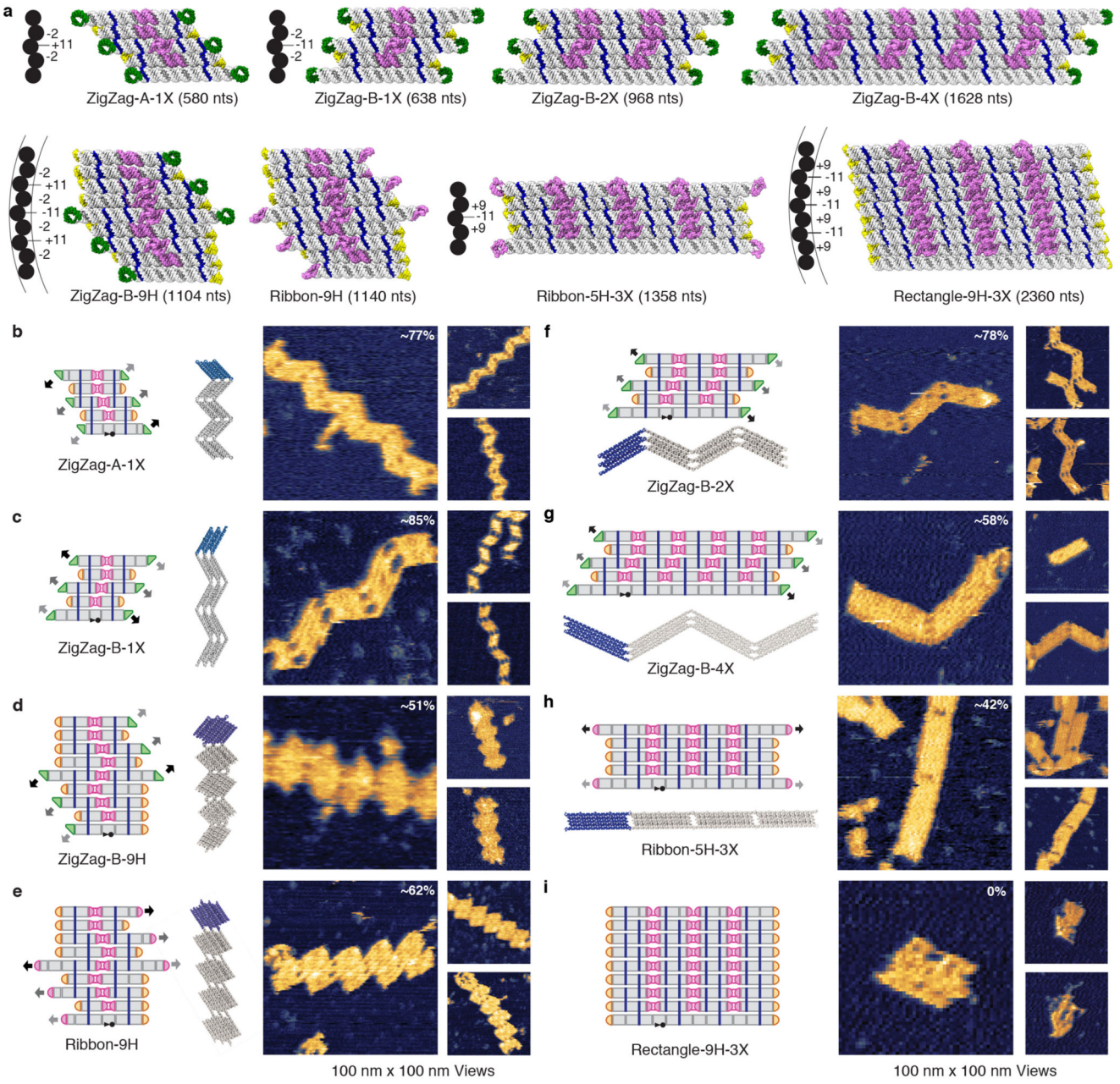

Figure 3. Building taller and wider RNA origami scaffolds.

a, RNAbuild models of RNA origami scaffolds with increasing size colored as in Fig. 1d. Schematic end views show pleating based on alternating DT lengths and scaffold curvature. **b-i,** Schematic models as in Fig. 1c with indication of inter-tile KL connectivity. Blue and gray models (middle) show RNAbuild models connected into fibers by intermolecular KLs. AFM (right) shows 100 nm fields; resolution is high enough that fine features such as gaps between tile connections can be observed. **b,** ZigZag-A-1X, a 5-helix tall tile with -11DT offsets and 120KL connectors, produces zig-zag chains. **c,** ZigZag-B-1X, a 5-helix tall tile similar to that in (b) but with +11DT offsets. **d-e,** Two versions of a 9-helix tall tile with both

-11DT and +11DT offsets: using 120KLs as connectors (ZigZag-B-9H) results in zig-zag chains; using 180KLs (Ribbon-9H) yields straight ribbons. **f-g,** ZigZag-B-2X and ZigZag-B-4X derive from ZigZag-B-1X and share a (-2-bp, -11-bp, -2-bp) series of DTs. **h-i,** Ribbon-5H-3X assembles with 180KL connectors to form a straight ribbon; it features a (+9-bp, -11-bp, +9-bp) series of DTs. Monomer-9H-3X expands this to 9-helices tall with no KL connectors. Yield of tiles shown on main AFM image from Supplementary Table 3.

Europe PMC Funders Author Manuscripts

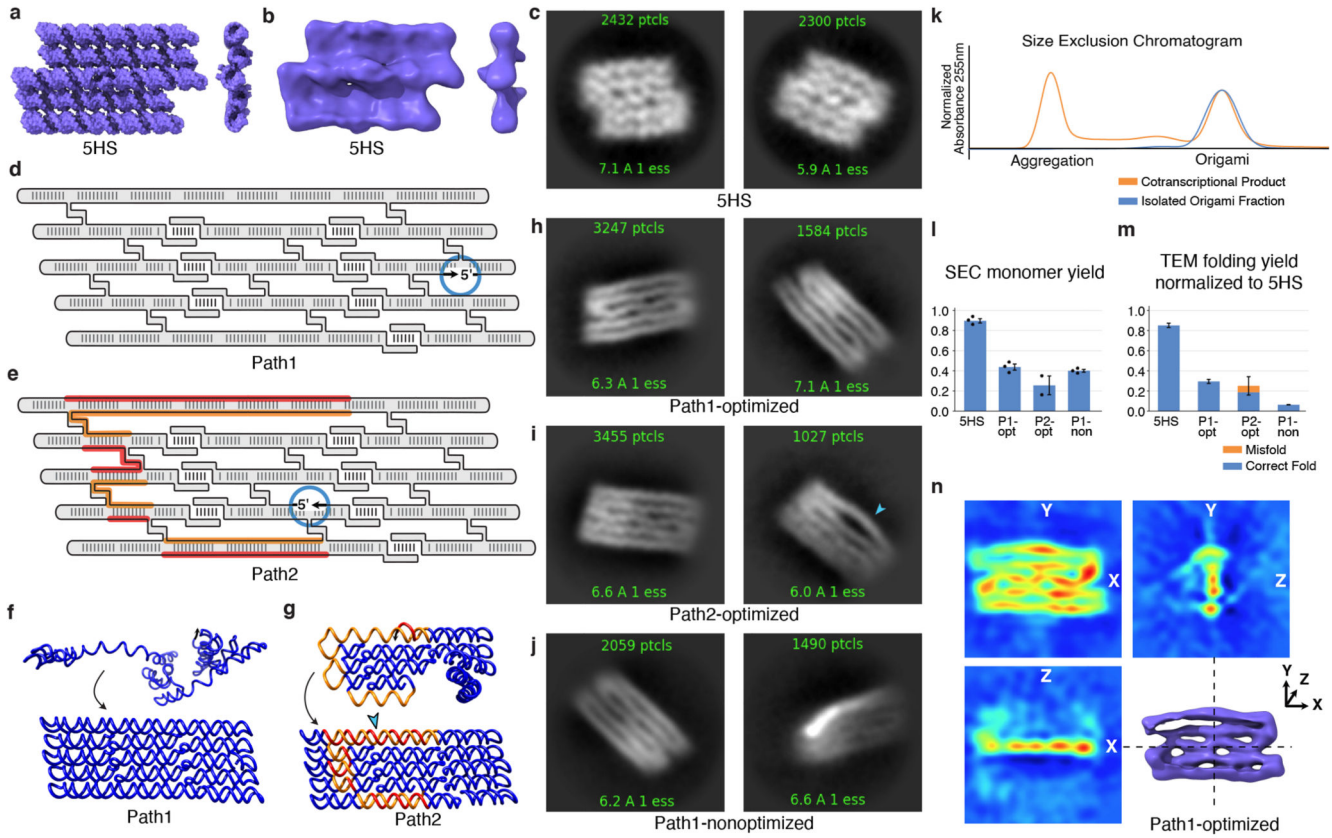

Figure 4. TEM analysis of RNA origami structures with optimal and suboptimal designs.

a, RNAbuild model of the 5HS structure shown in face and edge view. **b,** Model-seeded TEM reconstruction shown in face and edge view. **c,** Class averages of 5HS showing the two dominant views. **d,** Path1 with a long transient 5’ single strand, but no predicted topological barriers. **e,** Path2 with no transient 5’ single stranded region, but substantial predicted topological problems. KL interactions that fold rapidly generate looped out regions (orange) that later must be wrapped around by the nascent chain (red), constituting potential topological barriers to folding. Blue circles (**d, e**) indicate 5’ start sites. **f,** Intermediate (top) and final frame (bottom) of folding movie for Path1 (taken from Supplementary Video 3). **g,** Intermediate (top) and final frame (bottom) of folding movie for Path2 (taken from Supplementary Video 4). Orange and red colors show topological barriers as described in e. Cyan arrow point to position of bend helix in panel i. **h-j,** Representative TEM class averages of the particles. Identifiable structural defects predicted by design software are indicated by a cyan arrow. **k,** Size-exclusion liquid chromatography (SEC) trace at UV 255 nm of transcription reaction showing aggregate and monomer peaks, and monomer peak rerun after purification. **l,** Quantification of FPLC purified aggregation peaks and monomer origami peaks. Error bars show standard error for n=2-3. **m,** Relative folding yield determined from number of face views observed in TEM images adjusted for monomer peak yield. Orange bar shows misfolded state. **n,** *Ab initio* reconstruction of Path1 with sequence optimization.

Europe PMC Funders Author Manuscripts

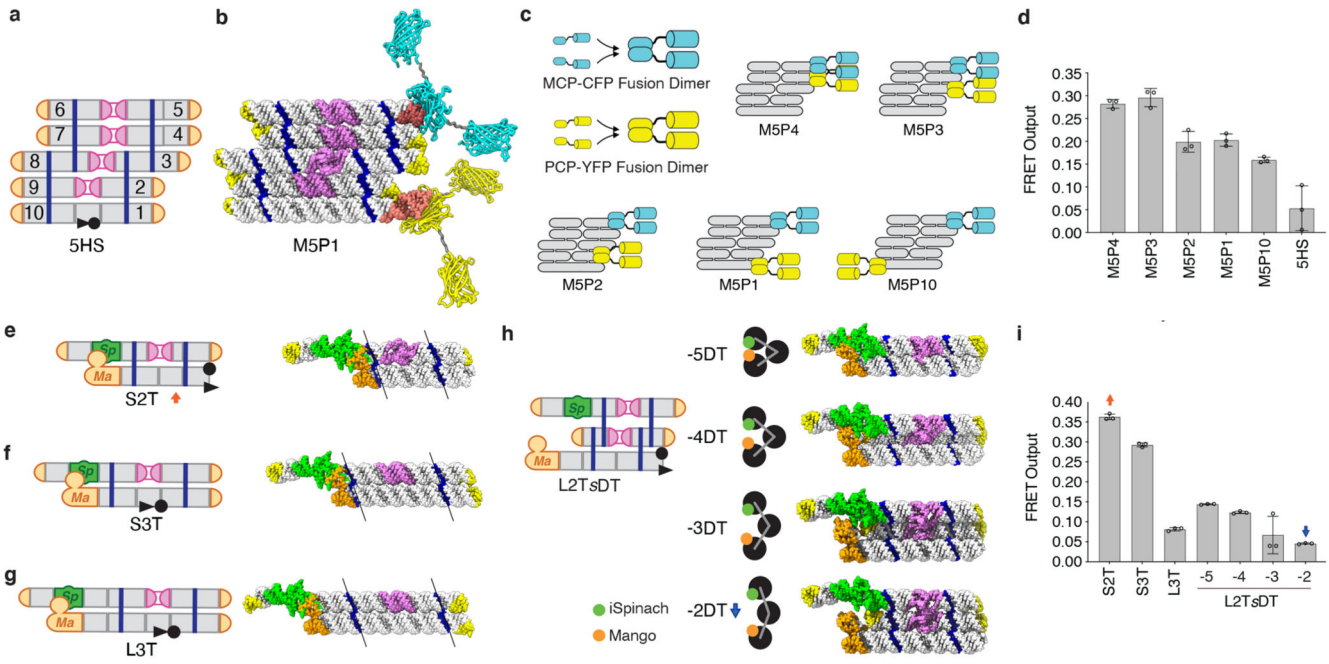

Figure 5. Scaffolding fluorescent proteins and small molecules on RNA origami.

a, Schematic shows ten positions at which aptamers can be added. **b,** Atomic model shows the tile with proteins attached on position 5 and 10. **c,** Schematics for RNA aptamer-binding domains (MCP or PCP) fused to fluorescent proteins (CFP or YFP). Both MCP-CFP (cyan) and PCP-YFP (yellow) self-dimerize via their aptamer binding domains[60]. Schematics of variants positioning MCP and PCP fusions with five different separations named. **d,** FRET measurements for the scaffold variants binding to MCP-CFP and PCP-YFP. Error bars show standard deviation for n=3. **e,** S2T, the variant with highest FRET marked by red up-arrow. **f,** S3T differs from S2T by the addition of one turn between crossovers. **g,** L3T differs from S3T by the addition of one turn to the aptamer-bearing arms. RNAbuild predicts the same distance and steric clash between aptamers for all three variants. **h,** For L2T*s*DT three-helix variants having different DT spacing (from s = -5 to -2 bps) the dihedral angle (black cross-sections) increases by 32.7° for each 1 bp decrease. L2T-2DT, which exhibits the lowest FRET marked by blue down-arrow, places iSpinach and Mango ~4 nm apart. **i,** FRET measurements for the seven scaffold variants. Error bars show standard deviation for n=3. Red and blue arrows denote S2T and L2T-2DT, respectively.

Europe PMC Funders Author Manuscripts