Methods to program and to probe RNA tertiary structure with nucleic acid origami

by

Molly F. Parsons

B.S., Iowa State University (2016)

Submitted to the Department of Biological Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Biological Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2022

© Massachusetts Institute of Technology 2022. All rights reserved.

Author .....
Department of Biological Engineering
May 13, 2022

Certified by .....
Mark Bathe, Ph.D.
Professor of Biological Engineering, MIT
Thesis Supervisor

Accepted by .....
Katharina Ribbeck, Ph.D.
Hyman Career Development Professor, Graduate Program Committee
Co-Chair, Biological Engineering, MIT

Methods to program and to probe RNA tertiary structure with nucleic acid origami

by

Molly F. Parsons

Submitted to the Department of Biological Engineering
on May 13, 2022, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Biological Engineering

Abstract

Biological structure determination has revolutionized mechanistic understandings, nanotechnology, and drug design. Despite advances in structural determination technologies, from nuclear magnetic resonance to cryo-electron microscopy (cryo-EM), one class of biomolecules has resisted 3D structure characterization. RNA, particularly larger RNAs, often dynamically adopt multiple conformations in a structural ensemble, and this heterogeneity has made 3D structure determination challenging through conventional techniques.

In this thesis, I investigated two avenues for improving RNA 3D structure determination, both leveraging the nanoscale programmability of nucleic acid origami. Nucleic acid origami generally involves folding one long single-stranded nucleic acid, the scaffold, into a target geometry via hybridization with short oligonucleotide "staples." First, we expanded the geometric space accessible to 3D wireframe DNA-scaffolded origami with edges composed of two helix bundles, optimizing folding conditions and crossover design and analyzing the final folded 3D structures, for a new design algorithm. I designed a tetrahedral wireframe DNA origami to capture an engineered tRNA via hybridization at three sites. For this complex, I verified stable, cooperative binding, and characterized the 3D structure with cryo-EM, which confirmed binding at all three sites and yielded a 17-Å resolution reconstruction of the tRNA. I also outlined a high-throughput workflow to probe the unknown tertiary structure of a target RNA with varied designs of DNA origami. Additionally, I studied the design of 3D wireframe RNA-scaffolded origami, characterizing the folded structure for several crossover schemes to evaluate how best to accommodate the A-form helical geometry of RNA for robust designs. The resulting algorithm for designing RNA-scaffolded polyhedra enables precise, covalent anchoring of a target RNA fragment onto a wireframe polyhedra. I tested this anchoring approach to attach a 232-nt HIV-1 RNA fragment to an RNA-scaffolded pentagonal bipyramid as a method to improve cryo-EM characterization. The particles folded into the expected pentagonal bipyramidal geometries, and cryo-EM micrographs suggested anchored target RNA, but the design and data analysis need further refinement to determine a 3D structure

for the anchored RNA fragment. These studies together represent proofs-of-concept for stabilizing RNA structures on nucleic acid origami, enabled by the expansion of origami design.

Thesis Supervisor: Mark Bathe, Ph.D.
Title: Professor of Biological Engineering, MIT

Acknowledgments

Thanks to the Office of Naval Research, the National Defense Science and Engineering Graduate Fellowship program, and the Siebel Scholars program for financial support throughout my PhD candidacy.

My graduate school career, and the work leading up to this thesis, have not been solo efforts, and I am grateful to the community of mentors, scientists and friends that supported me.

I thank my thesis advisor Prof. Mark Bathe for his guidance on project directions, his support in making connections with collaborators, and his career mentorship. He set up a collaborative lab environment that helped me to learn, to grow in confidence and to become an independent researcher. Prof. Nils Walter, a member of my thesis committee, provided invaluable insight into RNA structural biology and pushed my scientific rigor, and I really appreciate his kind support at all committee meetings. My thesis committee chair Prof. Paul Blainey helped me to explore and refine engineering approaches to tackle my biological questions, and also kindly spoke with me about career directions and provided advice at a critical moment.

Thank you to my collaborators for their expertise and assistance on numerous projects. Prof. Silvi Rouskin was an early mentor for me in the RNA space, introducing me to DMS-MaPseq chemical probing and HIV RNA biology. Prof. Wah Chiu provided insight into the application of cryo-EM to our designs, and how we might optimize designs, as well as support for the data collection and analysis itself with Dr. Kaiming Zhang, Dr. Shanshan Li, and Htet Khant in his lab. I appreciate these last three for the great deal of time they put in to carry out cryo-EM structure analysis for many structures. Similarly, Dr. Ed Brignole, assistant director of the Cryo-EM Facilitiy at MIT.nano, dedicated much time to training me in cryo-EM sample preparation and data collection, and I appreciate his numerous late nights of assistance and conversation during cryo-EM imaging.

Within the Bathe BioNanoLab, I owe a great deal of my early development to the mentorship of Dr. Tyson Shepherd. I thank him for teaching me many lab

techniques and experiment design, teaching me to avoid spending "clean thoughts on dirty data," and for his friendship. I also appreciate the valuable lab mentorship of Dr. Sakul Ratanalert, Dr. James Banal, Dr. Eike Wamhoff, and Dr. Floris Engelhardt. Sakul provided crucial friendship and advice for navigating graduate school as well. Thanks to Dr. Xiao Wang for supporting my work with atomic force microscopy data and for jumping in to clip some of my grids for cryo-EM. I thank Hellen Huang for frequent experiment support during a couple of intense years. And many thanks to Matthew Allan for being excited about RNA, teaching me much about RNA biology and statistics, and working with me on the RNA-scaffolded origami project beyond the contents of this thesis. Thank you in general to all the current and former members of the Bathe BioNanoLab with whom I have overlapped for their scientific discussion, friendship, river runs, and trivia wins and losses. I was fortunate to get to mentor several amazing undergraduates: Jiayi Dong, Jaquesta Adams, Shreya Gupta, and Sal Ibarra. I am thankful for the opportunity to work with them and for their inspiring attitudes and dedication.

In the BE community, I thank Sue Jaskela and Dalia Fares at the academic office and Aran Parillo in the BE IT office for frequent administrative and IT support. The BE community has been my favourite part of my PhD experience. I am grateful to have gone through the program with the BE 2016 graduate student cohort, who have all been so supportive of each other. Thank you to Connor Dobson for fun discussions of viral biology, lesson planning and co-teaching SPLASH and HSSP classes for high school students, and thesis proposal and dissertation writing parties. I am likewise very grateful to my many other BE friends, including Shalmalee Pandit, Noreen Waufford, Divya Ramamoorthy, Meelim Lee, Ian Andrews, Patrick Holec, Eric Lenhardt–our cohort dinners and climbing and gaming and ice skating sessions bolstered me at many difficult points. In particular, I'd like to thank my fellow lab and cohort members Rebecca Du and Joseph Berleant, who were both great support in lab, from discussing thermodynamics and teaching me Python to saving late night gels, as well as steadfast friends whose support I treasure.

Outside of MIT, I want to thank Dr. Alina Ott who mentored me during my

undergraduate research in the Schnable Lab at Iowa State University, and taught me a love of research and of carefully organized lab notebooks.

Thank you to my friends Sophie Donta and Ella Brown and to Bill, Ruth, and Geordie Lindemann for reminding me about life oustide of graduate school and supporting me through it.

To my family, Terry ODonnell Parsons, Marshall Parsons, Emily Parsons and Adam Wozniak, Lynn Parsons, Henry Parsons, and Oliver Parsons, thank you for shaping me into the person I am today, showing interest in my work, and helping me to keep laughing. And finally, thank you to my husband Robin Lindemann for being an emotional tether since the beginning of graduate school applications, for your constant love and support, and for holding off on the grad school jokes for two years when you defended before me.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Although of critical relevance to disease and underlying cellular functions, there is currently a dearth of known three-dimensional structures of ribonucleic acids (RNA). A search of the PDB revealed that 3D structures have been reported for only 3% of the HIV-1 RNA genome, for example, and typically for very short fragments. The driving motivation for my thesis work was to develop new strategies for tertiary structure determination of larger RNAs.

1.1 RNA structure

From skeletal movement to cell membranes and RNA riboswitches, the relationship of structure and function is fundamental to biology at every level. For example, RNaseP, one of the first recognized catalytic RNAs, has a highly-conserved flat-surface structure that enables the enzyme to specifically recognize its precursor tRNA substrate and catalyze hydrolysis—in some cases even in the absence of coordinating proteins [1–3]. Even for non-catalytic RNAs, structure affects their interaction with the cellular environment; 35% of mRNA untranslated regions (UTRs) in Escherichia coli have conserved structural motifs important to protein binding [4]. Additionally, the tertiary structure of viral RNA genomes plays a key role in their infection and replication.

Like proteins, RNA has several levels of structure: primary (1^o), the sequence of

nucleotides; secondary (2^o), the pattern of unpaired nucleotides and hybridization of complementary nucleotides; tertiary (3^o), the ultimate three-dimensional RNA structure; and quarternary, higher-order complexation with other molecules. Researchers can use classical structure characterization techniques like crystallography, cryogenic-electron microscopy (cryo-EM), and nuclear magnetic resonance (NMR) spectroscopy to find high-resolution 3D structures of rigid RNA [5]. However, most, if not all, RNAs are dynamic and adopt multiple conformations to fulfill their functions. For such RNAs, it is more accurate and relevant to consider their structure as an ensemble: a distribution of different conformations [6–8]. The heterogeneous populations that comprise the RNA structure ensemble make 3^o structure determination of larger RNAs nearly intractable using existing techniques.

1.1.1 RNA structure-function relationships

The distribution of conformations of an RNA can change dynamically in response to their environment, and the nature of the structural distribution–i.e. which conformations are dominant in the population–in turn affects the function of the RNA [6]. For example, the 5' UTR of the HIV-1 RNA genome adopts two well-documented 2^o structure conformations [9]. In one conformation, a palindromic sequence termed the "Dimerization Initiation Site" (DIS) is buried, hybridized with another region in the 5'UTR, and the ribosome binding site and start codon are exposed, enabling translation of viral proteins. In another conformation, the ribosome binding site and start codon are base-paired and buried, inhibiting translation, while the DIS is exposed, enabling genome dimerization. When the genome dimerizes, it can adopt an extended base-paired conformation and bind to tRNA and to viral proteins that package it into a capsid [10–12].

HIV-1 RNA genome structure is thus clearly important to regulating the viral life cycle. A quinolinium derivative compound inhibits viral packaging by affecting the RNA structure: the compound stabilizes a stem loop in the global packaging signal region, preventing interaction with the Gag protein required to direct packaging of the viral genome [13]. Scientists have hypothesized that the virus manages to package

exclusively unspliced genomes because a portion of the intron participates in a stem that is key to the presentation of a binding site for the Gag protein [10].

Additional predicted secondary structures in HIV-1 have as yet unknown function, but are highly conserved across the virus's subtypes. For example, Qi Wang *et al.* identified a conserved 100-nt three-stem-loop structure in the region coding for the protease in HIV-1 [14]. The codons in this region exhibit a lower-than-normal rate of synonymous mutation across subtypes, suggesting that the specific nucleotide identity is important beyond amino acid encoding, and co-varying nucleotides are base paired in the predicted 2^o structure. Such a highly conserved structure suggests an importance to the survival or reproduction of the virus, but how remains an open question.

Riboswitches are another clear example of the RNA structure-function relationship, with one conformation capable of–and stabilized by–binding a metabolite, suppressing or promoting downstream gene expression [8, 15]. Ganser et al. extensively review this and other relationships between RNA structural ensembles and cellular function [6].

1.1.2 Biochemical strategies to study RNA structure

One common method for investigating RNA structure is through molecular dynamics (MD) simulations and other energy minimizations. Experimental data can inform constraints for these simulations to make them more reliable. Two main chemical probing techniques are popular for informing RNA secondary structure predictions: selective 2'hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) [16, 17] and dimethyl sulfate mutational profiling with sequencing (DMS-MapSeq) [18]. Both work via similar methods: a reagent preferentially modifies single-stranded nucleotides, and these modifications lead to mutations during reverse transcription and can thus be read out after sequencing. The mutational profiles are used as parameters in an energy minimization algorithm that predicts which of the double-stranded nucleotides are likely paired with each other to output possible secondary structures [16–18].

DMS in particular probes the hybridization face of nucleotides, with potential to methylate ring nitrogens, which are hydrogen-bonded and therefore blocked from DMS modification when hybridized in a Watson-Crick-Franklin (WCF) base pair. DMS modifies principally adenines (A's) and cytosines (C's), and to a lesser extent guanines (G's) [19, 20]. SHAPE chemistry, on the other hand, effectively measures single nucleotide flexibility; the greater the local flexibility, the more reactive the 2' hydroxyl group is and more prone to modification in SHAPE. WCF base pairing decreases flexibility and reduces SHAPE reactivity, but non-canonical and non-base-pairing interactions can have the same effect [21]. Unlike DMS, SHAPE reagents modify all four RNA nucleotide varieties [22]. More recent analysis pipelines have improved the ability to directly identify nucleotide pairs [23, 24] and to cluster populations of differing structure for a single RNA sequence [25]. Proximity ligation, another structure determination technique with a sequencing readout, has the potential to identify some tertiary interactions, although the resulting data are noisy and subject to sequence biases [26].

Another source of structural information and constraints for computational simulations, Förster Resonance Energy Transfer (FRET) gives information about the distance between two probes, one labeled with a "donor" dye whose emission spectrum overlaps with the excitation spectrum of the "acceptor" dye on the other probe. After exciting the donor, some of its energy will excite the acceptor dye instead of emitting directly. The efficiency of this energy transfer (FRET efficiency) directly correlates with the proximity of the two dyes. The distances obtained from these experiments serve as constraints in MD simulations, which Stephenson *et al.* used to propose several possible models of the packaging signal region in HIV RNA [27]. However, bulk FRET has limited use for RNA because it finds an average distance between the probes on the heterogeneous conformations of RNA and does not necessarily represent any individual conformation. Single molecule FRET (smFRET), on the other hand, can provide information about the distinct conformational populations. With smFRET, individual molecules are tethered to a surface and FRET is observed with Total Internal Reflection Fluorescence (TIRF) and single molecule

imaging techniques. Several distances might be found between probes on different molecules, from which we can infer a probability distribution of conformations [8, 28, 29].

1.1.3 Structural approaches to study RNA tertiary structure

X-ray crystallography, one of the earliest methods to study molecular structure, provided the structure of many proteins and the double-helical structure of DNA. Unfortunately, due to the heterogeneous conformations that most larger RNAs adopt and the tendency of the molecule to form multiple different crystallites, homogeneous crystal formation of RNAs is limited to short oligonucleotide duplexes, engineered RNAs (e.g. with tetraloops to promote crystal growth), and RNAs bound within larger complexes of proteins (e.g. the Group II intron) [30, 31].

Small angle X-ray scattering (SAXS) is a solution technique, eliminating the need to make the RNA crystallize. Because the molecules are in solution, the isotropic scattering profile is a result of scattering from objects in a multitude of random orientations at once, and provides global structural information about the objects. Although useful as a point of comparison or validation for simulated atomic structures, the molecular envelopes that can be calculated from SAXS are low-resolution [32]. The resolution limitation is especially true for conformationally heterogeneous populations of RNA because of the spatial averaging across molecules. SAXS has been applied to a few small fragments of the HIV RNA genome, such as the TAR/PolyA stem loops, the packaging signal (Ψ), and the Rev Response Element (RRE), the latter with a resolution of 21 Å [33, 34].

Similar to the biochemical strategies described above, nuclear magnetic resonance (NMR) spectroscopy indirectly characterizes tertiary structure, providing distance constraints to an energy minimization model or combining with structure data from other techniques to generate a model. Keane et al, for example, applied NMR to a 95-nt fragment of the HIV-1 packaging signal, replacing some of the especially flexible portions of the RNA with a short tetraloop to reduce the molecule size and conformation space [10]. An advantage of NMR is the potential to detect different

conformations of a heterogeneous population in solution, and with much higher resolution than SAXS, but the method is limited to smaller molecules due to the spectral overlap of nuclear resonances. Unless the RNA is rigid, NMR is further limited to identifying only short-range interactions [35]

1.1.4 Cryo-electron microscopy

Cryogenic-electron microscopy (cryo-EM) is perhaps the most direct characterization technique for 3D structure. For single molecule cryo-EM, samples in solution are blotted onto a carbon mesh grid and flash frozen in liquid ethane with a plunge system, leaving a thin film of vitreous (non-crystalline) ice containing many molecules in a variety of orientations. This frozen grid is then imaged with an electron microscope, each image typically consisting of a couple hundred individual frames taken over the course of a several second exposure. Collecting movies enables improvements in image processing: correction of beam-induced motion over the course of the exposure and weighting data by dose – by the end of the exposure, particles have been exposed to much electron-induced damage and are more degraded [36].

After such image pre-processing, particles present in the images are identified–a process called "particle picking." The collection of selected 2D images of individual particles are clustered in sets to define class averages; i.e. the average electron density of each particular 2D orientation—with better signal-to-noise ratio than an individual particle image alone. These class averages are essentially 2D projections of the molecule from different viewpoints, and these are assembled to reconstruct a 3-dimensional structure of the molecule, with potentially near-atomic resolution [36]. Recently advanced image processing and classification algorithms can handle a modest amount of heterogeneity, potentially sorting particle images into a few 3D classes [37–40]. However, minimizing heterogeneity is still considered a key step in cryo-EM sample preparation; the classification is still limited, and the greater the heterogeneity and number of 3D classes, the more total data is required for successful reconstruction.

Recently, Zhang et al reconstructed a 47-nt structure of the rigid DIS duplex from

the HIV-1 genome with 9 Å resolution [41], a 119-nt SAM-IV riboswitch with 3.7 Å resolution [42], and the 88-nt frameshift element of the SARS-CoV-2 genome with 6.9 Å resolution [43], the highest resolution EM reconstructions of protein-free RNA to-date. Generally with cryo-EM, scientists have achieved low resolution reconstructions of free RNA [5] and high resolution (3-5 Å) reconstructions of RNA in the context of ribonucleoprotein complexes such as the spliceosome and ribosome [44, 45]. The singular RNA conformation enforced by interacting proteins in these complexes enabled such high resolution, but for typical RNAs in solution, the averaging across particles with heterogeneous conformation leads to low-resolution structure maps.

The successful reconstructions of RNA structures mentioned above have all been cases of small, rigid RNAs or RNAs stabilized by binding to proteins that force the RNA to hold a particular conformation, enabling crystallography and cryo-EM of a homogeneous population. These results suggest that if we bound other dynamic RNAs in a way that encourages them to hold a single conformation, we might be able to employ the classical techniques listed above to solve the structure of that conformation with high resolution. To that end, **I aimed to develop nucleic acid origami tools to stably capture or anchor RNA molecules with a single conformation at a time from a heterogeneous ensemble**, and to reconstruct the structure of that conformation with cryo-EM.

1.2 Nucleic acid origami design technologies

Nucleic acid origami involves folding a long, single-stranded nucleic acid (the "scaffold") into a target geometry via Watson-Crick-Franklin (WCF) base-pairing with short nucleic acid oligos (the "staples"). Staples have multiple binding domains, each complementary to distinct sections of the scaffold, forcing the scaffold to fold upon itself to bring these otherwise distant regions together for maximal hybridization with the staples. The target geometry can be two-dimensional or three-dimensional and have edges comprised of a single double helix or a bundle of two or more double helices, with the scaffold routing through each helix. To connect bundles of helices, both

scaffold and staple strands cross from one double helix to an adjacent one, forming a nucleic acid junction like the Holliday junctions observed in DNA recombination and repair [46–49]. Depending on the exact routing design, these "crossovers" can also form mesojunctions, where only a single strand crosses helices and continues in the reverse direction, or antijunctions, where strands cross helices and continue in the same direction [50].

The scaffold and staples, in theory, can be either DNA or RNA or a mixture of the two. Each staple also offers two points of modification one at the 5' and one at the 3' end. With various chemistries, each end can attach to therapeutic moieties, fluorophores, proteins, etc., or the staple ends can be simply extended with a sequence not complementary to anywhere on the scaffold, leaving an overhanging oligo at a particular location on the origami edge [51]. On a wireframe polyhedron, according to the placement of this extension, the overhang or other modification will extend outwards, inwards, or along the face of the origami. These modifications on the nanostructures can be precisely positioned with nanometer resolution, due to the predictable nature of WCF base-pairing. Nucleic acid origami is thus highly programmable at the nanoscale, making it an appealing choice for precisely capturing for anchoring different RNAs of interest.

1.2.1 DNA-scaffolded origami

The history of DNA nanotechnology and DNA origami in particular is reviewed in detail elsewhere [51–53]. Of particular relevance to this thesis work are advances in the automated design of DNA-scaffolded wireframe polyhedra. The Bathe BioNanoLab and others developed software to automate sequence design for wireframe 2D and 3D objects, with either two double helix bundles (2HB or DX) or six double helix bundles (6HB) composing the edges [54–58]. While 6HB edges help to rigidify structures [55], they are more difficult to modify at arbitrary desired locations along the edge, as many potential modification positions (i.e. allowable staple nick positions) are internal to the honeycomb-like bundle, and the routing design is more complex. 2HB-edged polyhedra also require much less scaffold than 6HB for the same nanoscale polyhedron.

However, DNA wireframe polyhedra design with 2HB edges has to date been limited to edge lengths that are multiples of 10.5 bp, i.e. full helical turns, which requires some morphing of input geometries so that all edges meet this requirement, and having multiple different edge lengths in a single object is only feasible for particular ratios of edge lengths.

DNA origami offers utility in a variety of applications [51]. Leveraging the ability to position antigens with precise spacing, copy number, and 2D or 3D patterning, Veneziano et al. used DNA origami to study principles of vaccine design [59]. Various groups are also investigating the use of DNA origami as a therapeutic delivery vehicle [60–63]. In the materials space, Oleg Gang and others use DNA origami to program assembly of metal or biological nanomaterials [64–67], or to template lithographic etching [68–70]. Taking programmable design one step further, dynamic DNA assemblies function as "nanomachines" [71, 72]. In an example more closely related to the goal of structuring biological RNA, Timm et al. tethered enzymes to DNA origami to study the effects of organization and spacing on enzyme activity [73].

1.2.2 RNA origami

When forming double helices, RNA adopts A-form helical geometry, with 11 base pairs (bp) per helical turn and significant base pair pitch, unlike DNA's typical B-form helical geometry that has approximately 10.5 bp per helical turn and base pairs more perpendicular to the helical axis. Even in DNA:RNA hybrids, RNA dominates and forces and A-form helical structure [74]. RNA-scaffolded origami therefore requires different, if potentially analogous, design to DNA origami.

RNA nanotechnology often leverages natural RNA structures to construct objects. For example, Severcan et al. built a nanocube out of engineered tRNAs, with their known and reliable L-shaped 3D structure [75, 76], and Geary et al. incorporated kissing loops to connect helices in an ssRNA origami tile [77].

Demonstrating that the scaffolded folding approach commonly used in DNA origami could be applied to RNA, Wang et al. folded an RNA-scaffolded ribbon (3 double helices), rectangle (8 double helices), and triangle (with 3 double helices per edge)

using DNA staple strands, reporting near-quantitative yields [78]. Their manually-designed routing scheme used 11 bp per helical turn–i.e. crossovers between the same two helices were spaced by multiples of 11 bp–as an A-form accommodation. Endo et al. designed an RNA-scaffolded origami tile and 6HB rod, using either DNA or RNA as staples [79]. Their 6HB rod used the exact same routing scheme as DNA-scaffolded B-form 6HBs, with 10.5 bp per turn and no accommodation for A-form twist, and folded with 44% yield. The tile did account for the 11 bp per helical turn in its design but not for the twist and pitch of RNA A-form geometry. Their tile folded with 58% yield and with a longer axial length than expected for A-form, and the authors postulated that the twist of RNA A-form helices was not amenable to the planar form. These results suggest that, at least for small and simpler origami, RNA tolerates a degree of non-ideality in designs.

Further work in the RNA-scaffolded folding design space explored design principles for creating ssRNA tiles composed of double crossover (DX) motifs, the junction similar to Holliday junctions, where two strands cross between helices at adjacent nucleotides and reverse direction upon crossing over. Geary et al. found that, due to the pitch of A-form helices, adjacent crossovers of different strands (such as a scaffold strand crossover and a neighboring staple strand crossover) should be spaced asymmetrically on the two connected double helices [80]. They implemented these rules to fold a tile using only a single ssRNA strand with no staples [77].

RNA-scaffolded origami has several advantages over DNA-scaffolded origami. First, generating RNA scaffolds is much simpler than generating ssDNA scaffolds, as transcription naturally produces single-stranded RNA, whereas most DNA amplification systems inherently make double-stranded DNA. *In vitro* transcription reliably produces long ssRNA of arbitrary sequence with easier purification and higher yields than asymmetric polymerase chain reaction (aPCR) does for ssDNA of arbitrary sequence [81]. And ribosomal RNA is highly abundant in cells, a potentially cheap source of RNA scaffold for applications in which sequence is unimportant [82]. Second, RNA-RNA and RNA-DNA hybridization is more thermodynamically stable than DNA-DNA hybridization [74], which, combined with the different enzymatic degra-

dation rates of DNA, RNA, and modified RNA, provides the opportunity to tune release kinetics for therapeutic delivery applications.

Several studies have applied functionalized RNA nanotechnology to modulate cellular function [83, 84]. For example, siRNA-presenting nanocubes activated RNA interference pathways and reduced viral RNA load in cells [85], and ssRNA origami tiles presenting thrombin-binding RNA aptamers modulated clotting time better than free aptamers or DNA origami with thrombin-binding DNA aptamers [86]. RNA origami can also precisely organize ribozymes or other enzymes for intracellular or cell-free reactions, as applied to production of hydrogen in bacterial metabolism [87] and proposed for protein-free peptidyl transfer [51].

1.3 Overview of the thesis

My thesis work centers around developing a toolkit for enabling tertiary structure determination of larger RNAs, with the aid of nucleic acid origami. I worked on two approaches to that effect: "baited capture" of RNA via origami hybridization with RNA loops and "anchoring" via incorporation of target RNAs into an RNA scaffold (figure 1-1).

In Chapter 2 of my thesis, I discuss expanding the geometric space accessible to wireframe DNA origami via validation of the new algorithm DAEDALUSv2, an important step towards tailoring origami design to target RNA geometry for optimal stabilization. Chapter 3 details a proof-of-concept application of the baited capture approach, in which I used DNA origami to stably capture an engineered tRNA with three hybridization sites. I also discuss progress towards a high throughput platform for probing unknown RNA tertiary structures based on the baited capture approach.

Chapter 4 covers the design and characterization of RNA-scaffolded 3D wireframe origami, which enabled the anchoring approach to RNA stabilization. Initial experiments prototyping this anchoring approach with the Rev Response Element of HIV-1 are covered in Chapter 5.

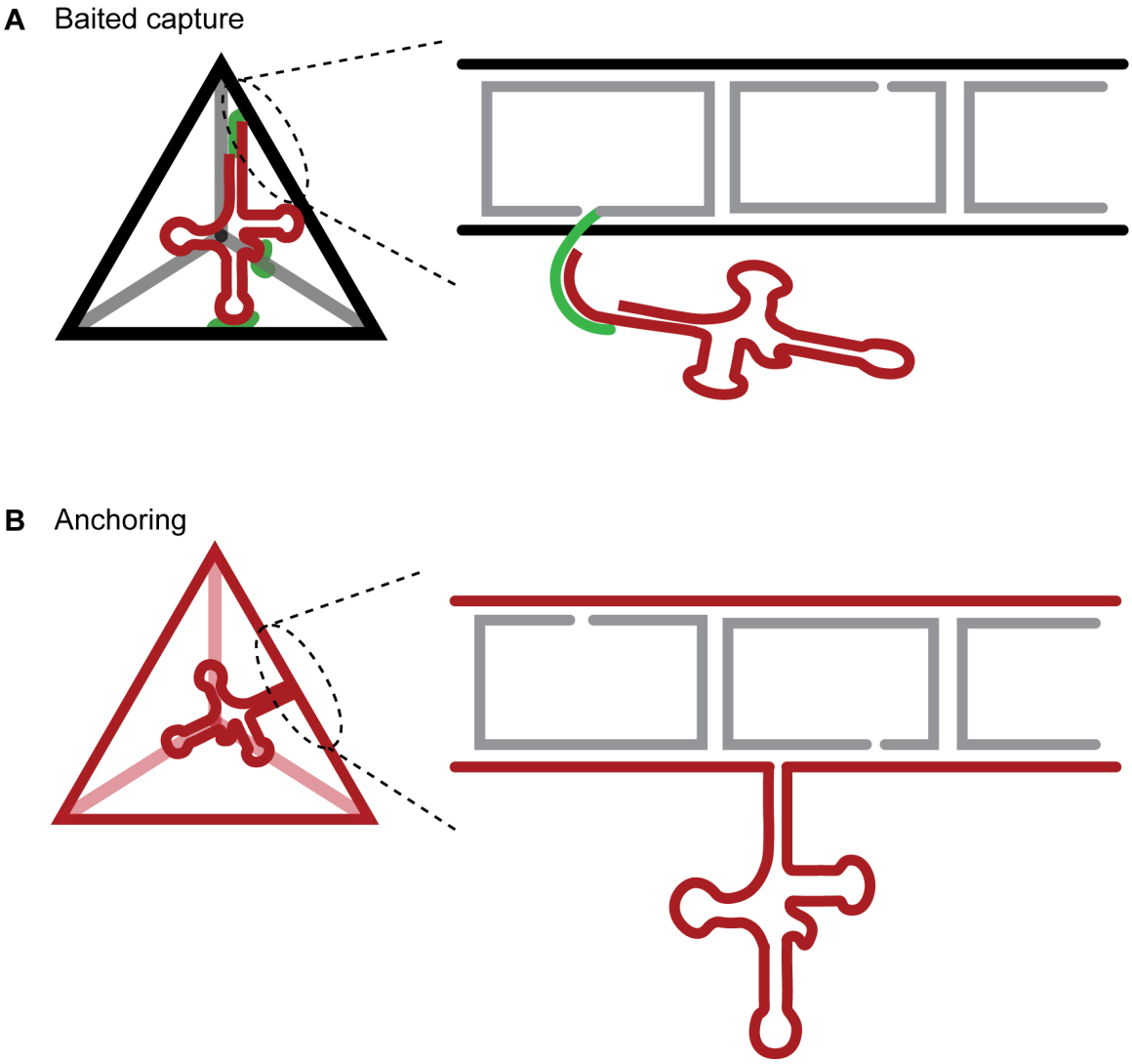

Figure 1-1: **Schematic of two approaches to stabilizing RNA with origami (A)** The baited capture approach, in which overhanging sequences (green) from staples (grey) hybridize with the target RNA (red). **(B)** The anchoring approach, in which extra RNA sequence on either side of the target RNA (red) is used to scaffold an origami to which the target RNA is anchored.

Chapter 2

DAEDALUSv2[1]

2.1 Introduction

Synthetic DNA is endowed with unique properties as a nanoscale engineering material due to its sequence programmability and reliable Watson-Crick base paring, which, combined with the rigidity of the double-helix and immobile double crossovers (DX), offers full programmability of user-defined 2D and 3D objects [47, 48, 88–90]. Scaffolded DNA origami [49] is particularly powerful in synthesizing macromolecular assemblies, using a long DNA scaffold strand that is routed through every base-pair of nearly any desired target shape in 1D [91], 2D [49, 92], or 3D [91–94] and folded via slow thermal annealing in the presence of a molar excess of short oligonucleotide staple strands. Scaffolded wireframe DNA nanoparticles [65, 95–99] (DNA-NPs) are one class of scaffolded DNA origami objects that offer major opportunity for diverse applications, from therapeutics to structural biology to functional materials [51, 59, 100–103]. However, the presence of numerous multiway junctions in wireframe geometries renders the manual scaffold routing and complementary staple design needed for fabrication of these assemblies extremely challenging. The previously published sequence design algorithms vHelix-BSCOR [58] and PERDIX [56] for 2D and vHelix-

[1]This work was performed in conjunction with Dr. Hyungmin Jun, who created the DAEDALUSv2 software. Dr. Shanshan Li and Dr. Kaiming Zhang reconstructed structures from my cryo-EM data while in the lab of Prof. Wah Chiu at Stanford. Dr. Xiao Wang performed AFM. **This work is being written up as a manuscript for publication.**

BSCOR [58] and DAEDALUS [57] for 3D wireframe scaffolded DNA origami have enabled the facile design of 2D and 3D wireframe geometries. vHelix-BSCOR for 2D and 3D employs single duplex DNA edges, whereas PERDIX and DAEDALUS render target 2D and 3D geometries using dual-duplex (DX-based) edges. For greater mechanical stiffness of DNA nanostructures, automatic sequence design of 6HB-based 2D wireframe polygonal objects has also been enabled by METIS [54] and 3D wireframe polyhedral objects by TALOS [55].

DX-based wireframe DNA-NPs from DAEDALUS have two advantages over 6HB-edge based objects. First, DX-based objects require shorter scaffold length compared with the approximately three-fold longer scaffold needed for 6HB edges, which limits the design of 6HB objects to relatively simple polyhedra when using the M13mp18 (7,249-nt) scaffold. Additionally, base-level editing for functionalization is more straightforward for DX-based DNA-NPs due to their relative structural simplicity. However, despite these advantages, DAEDALUS is still limited in its design approach by the fact that edge lengths must have integral numbers of double helical turns of B-form DNA (10.5 bp), and arbitrary vertex angles are not properly handled for high fidelity structural design.

To overcome these limitations, here we introduce a new algorithm that enables fully automated, top-down sequence design of polyhedral DX DNA-NPs to handle arbitrary edge lengths for the precise design of wireframe objects of asymmetric and irregular geometries. This offers the unique ability to organize molecules in nearly arbitrary spatial patterns at the nanometer scale, including asymmetric organizations. Importantly, akin to the 2D implementation of PERDIX, the design motif of the "continuous edge," which enables DNA-NPs to consist of arbitrary edge lengths and vertex angles, is newly integrated within the DAEDALUSv2 algorithm and the Graphical User Interface ATHENA [104]. Unpaired scaffold nucleotides are used at the vertices to accommodate 5'- and 3'-end misalignments that allow arbitrary edge length and vertex angles to be designed. In addition, the new algorithm includes the capability for manual base editing and staple functionalization using the popular computer-aided design software caDNAno [105].

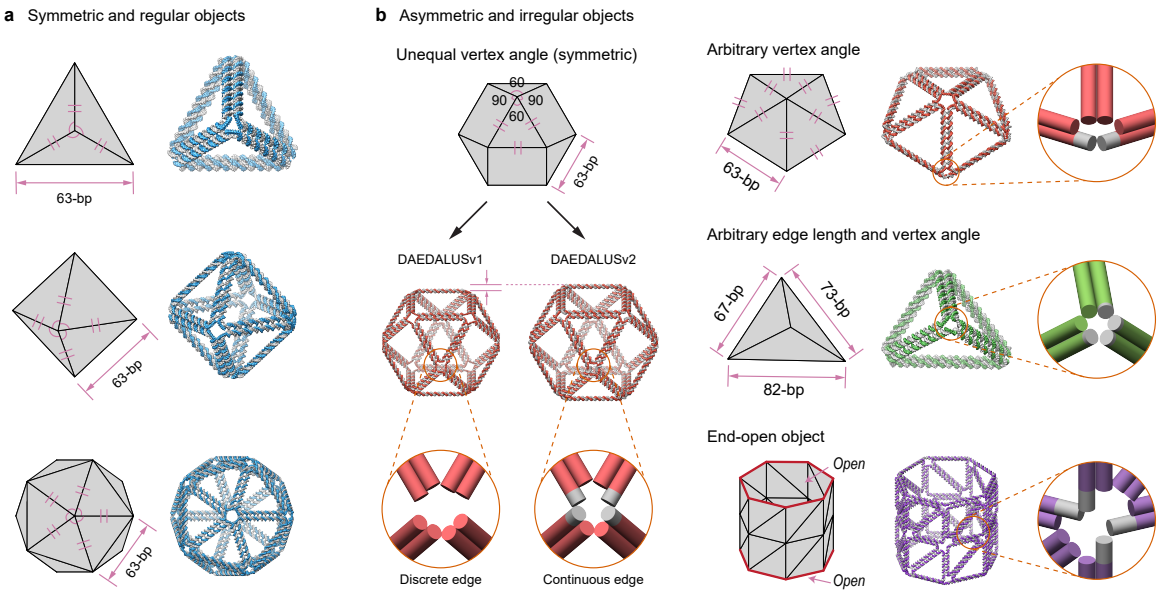

Figure 2-1: **Design of 3D wireframe scaffold DNA origami object with DX-based edges. (a)** Symmetric and regular objects with equal edge lengths and vertex angles. **(b)** Asymmetric and irregular objects with unequal edge lengths or unequal vertex angles.

We validate our new automatic sequence design procedure by applying it to fabricate three objects of uniform edge lengths and vertex angles and demonstrate the utility of the procedure with three new objects of non-uniform edge lengths and/or vertex angles. Proper folding and monodispersity of DNA-NPs are confirmed using gel mobility shift assays, cryo-electron microscopy (cryo-EM) and 3D cryo-EM reconstruction. Results demonstrate high yield of formation of origami objects and high structural fidelity at the nanometer scale. For broad dissemination of our approach, our algorithm is integrated within DAEDALUS in ATHENA [104] (https://github.com/lcbb/athena), which renders automatically the scaffold routing path and staple strand sequences, and is provided as a standalone open-source package (https://github.com/lcbb/daedalus2) for custom wireframe DNA origami design in 3D.

2.2 Results

2.2.1 Automated design of 3D wireframe origami

Automatic sequence design for target 3D geometries (Figures 2-1, A-1 and A-2) is based foundationally on a previous automated 2D design approach, PERDIX [56]. Unlike PERDIX, the new algorithm uses the spanning tree based on the target geometry [57] to determine the scaffold routing, instead of the spanning tree based on the dual graph of the loop-crossover structure [54–56], although these approaches produce equivalent results for the scaffold routing for DX-edges. The target geometry can be defined in two ways: Using a discrete edge-length consisting of multiples of 10.5-bp rounded to the nearest nucleotide or a continuous edge-length with no constraint on length (Figures 2-1 and A-3). Discrete edge length design [57] uses the common assumption of DNA origami design that satisfies the natural helicity of B-form DNA, thereby requiring staple poly-T bulges of length five crossing between edges without any unpaired nucleotides in the scaffold [57]. In contrast, continuous edge design enables objects with continuous, arbitrary edge-lengths and vertex angles, and requires a single duplex to fill the gap in each vertex. Unpaired scaffold nucleotides are used to span the distance between the 3' and 5' end between incoming and outgoing edges, which would otherwise be misaligned due to the native twist of B-form DNA 2-2 [54, 56].

2.2.2 Folding validation

To validate this continuous edge design algorithm for 3D wireframe structures, we tested folding of six objects designed with the new DAEDALUS algorithm, also referred to as DAEDALUSv2. Three regular objects, a pentagonal pyramid with 84-bp edge length (PB84) and two icosahedra with edge lengths of 42- and 52-bp (I42 and I52, respectively), were designed to show equivalency with the earlier algorithm [57] for regular target geometries of discrete edge length. The sequence designs for these objects are illustrated in Figure A-4 and Table A.1. Electrophoresis gel-shift assays

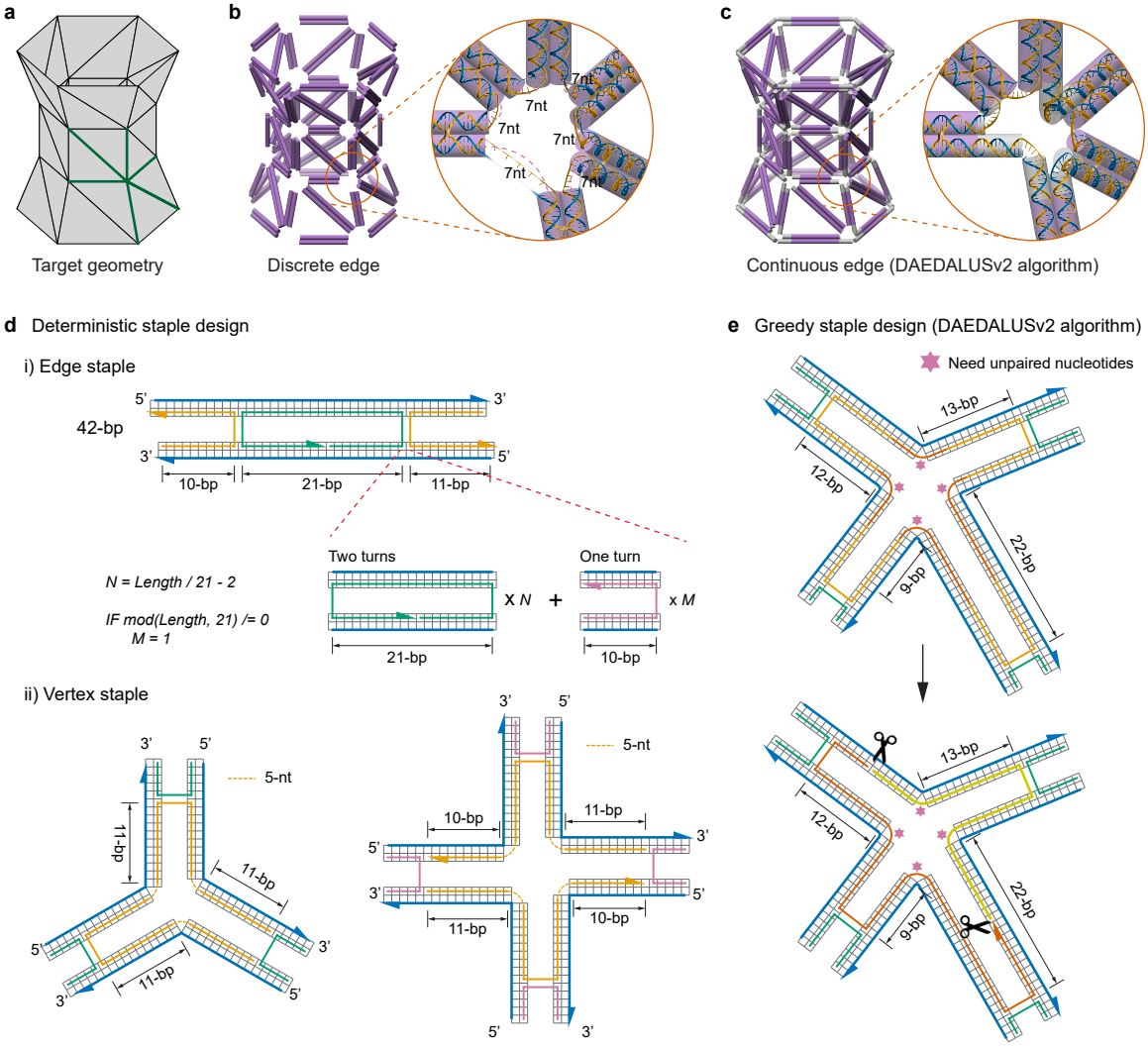

Figure 2-2: **Discrete versus continuous edge lengths for asymmetric and irregular shapes.** **(a)** The target geometry can be defined in two ways: Using a discrete edge-length **(b)** consisting of multiples of 10.5-bp rounded to the nearest nucleotides or a continuous edge-length **(c)** with no constraint on lengths and vertex degrees. **(d)** For the discrete edge design, edge staple is designed based on the 21-bp length and vertex staple on the degree of the vertex. **(e)** For the continuous edge design, staples are designed by cutting the non-circular staple paths with from 20-bp to 60-bp nucleotides.

showed formation of a tight, shifted band after folding, suggesting that these regular particles folded well with near-100% monomer yield (Figure A-5).

Similarly, three irregular objects of continuous edge length and variable vertex angles, including a pentagonal pyramid, asymmetric octahedron, and chiral object, all with base-line 63-bp edge length (PP63, AO63, and CO63, respectively, designs in Figure A-6 and Table A.2), folded as expected according to gel-shift assays (Figure A-7). Although monomer yield for the irregular objects was already high at 70-85%, we used the asymmetric octahedron as a test case for varying folding conditions from the published protocol [57] to find an improved procedure. Gel-shift assays quantified with ImageJ [106] revealed that the asymmetric octahedron could fold in as little as two hours with on average a 14% increase in monomer yield relative to the originally published 12.8h folding ramp (Figure 2-3). Slight increases in monomer yield were also apparent with decreasing concentrations of magnesium (Figure 2-3), possibly due to reduced electrostatic shielding that might inhibit aggregation. Sample atomic force microscopy (AFM) showed the expected portion of dimerization based on gel results (Figures A-8 and A-9).

We also compared the gel-shift assay results for objects designed with the original and new algorithms for the same input geometries (Figure 2-4). All objects showed distinct band shifts relative to the scaffold, suggesting proper folding. For irregular objects, the differences in the degree of band shifting are expected due to differences in the output geometries for the two algorithms, as the original DAEDALUS algorithm slightly distorts the input geometry to enforce full helical turns on all edges. Such distortion is unnecessary for regular objects with defined edge lengths of a multiple of 10.5-bp, so the output geometries for the two algorithms should be equivalent for regular objects—and they did indeed migrate similarly, with the exception of I52, discussed below.

Hypothesizing that double crossover motifs would provide greater stability than single crossovers [50, 107], we initially tested a constraint in DAEDALUSv2 that disallowed single crossovers on edge staples, unlike in the original DAEDALUS algorithm (Figure 2-5). For regular geometries of discrete edge lengths, the constraint would

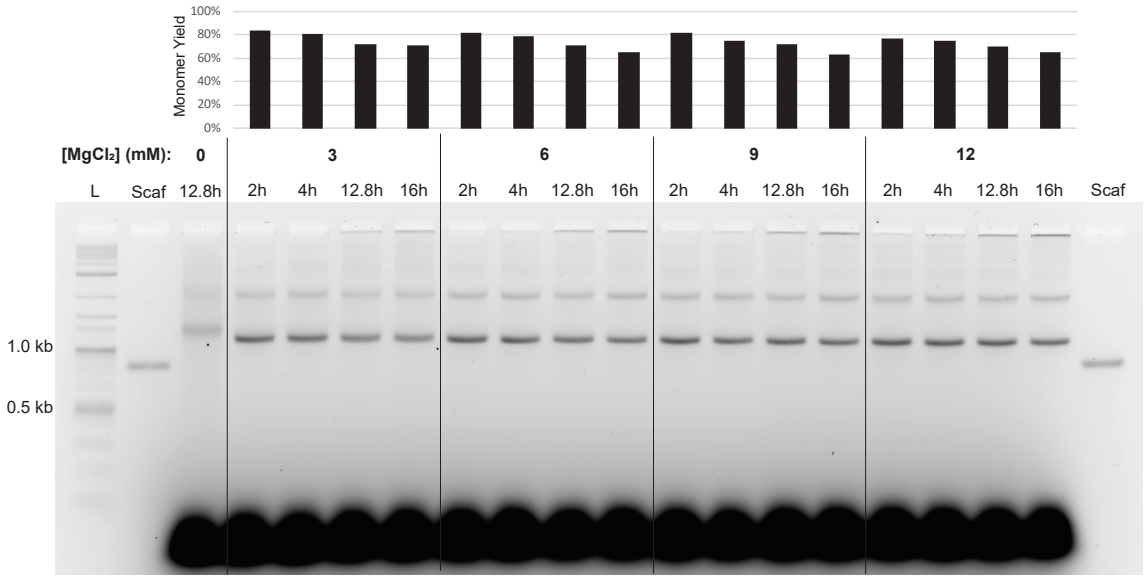

Figure 2-3: **Agarose gel electrophoresis for the asymmetric octahedron of minimum 63-bp edge length with varied magnesium content and length of folding ramp.** Percent monomer yield was calculated from band intensities determined with ImageJ.

only effect a difference for edges with odd numbers of helical turns; edges with even numbers of helical turns are evenly divided into double crossover staple motifs, with double crossovers 10 bases away from the vertices, without single crossover motifs in either algorithm. The I42 and PB84 objects we tested thus have equivalent routing in the original design and DAEDALUSv2 with or without the single crossover constraint. However, when an edge has an odd number of helical turns (e.g., five for 52-bp edge length rather than four for 42-bp edge length), the extra helical turn may be incorporated by either extending the distance between one pair of double crossovers to 21 bases (Figure 2-5a) or by adding a single crossover 10 bases away from a double crossover (Figure 2-5b).

To test our hypothesis, we first used the former design strategy, avoiding single crossovers. We observed a notable shift between the two bands of folded I52 (Figure 2-4), which has 11 single crossovers in the previously published algorithm that are absent in DAEDALUSv2 tested with the crossover constraint. This shift may be due to greater flexibility of the DNA-NP designed sans single crossovers: Increasing flexibility and the presence of curvature for nucleic acids of comparable size have been

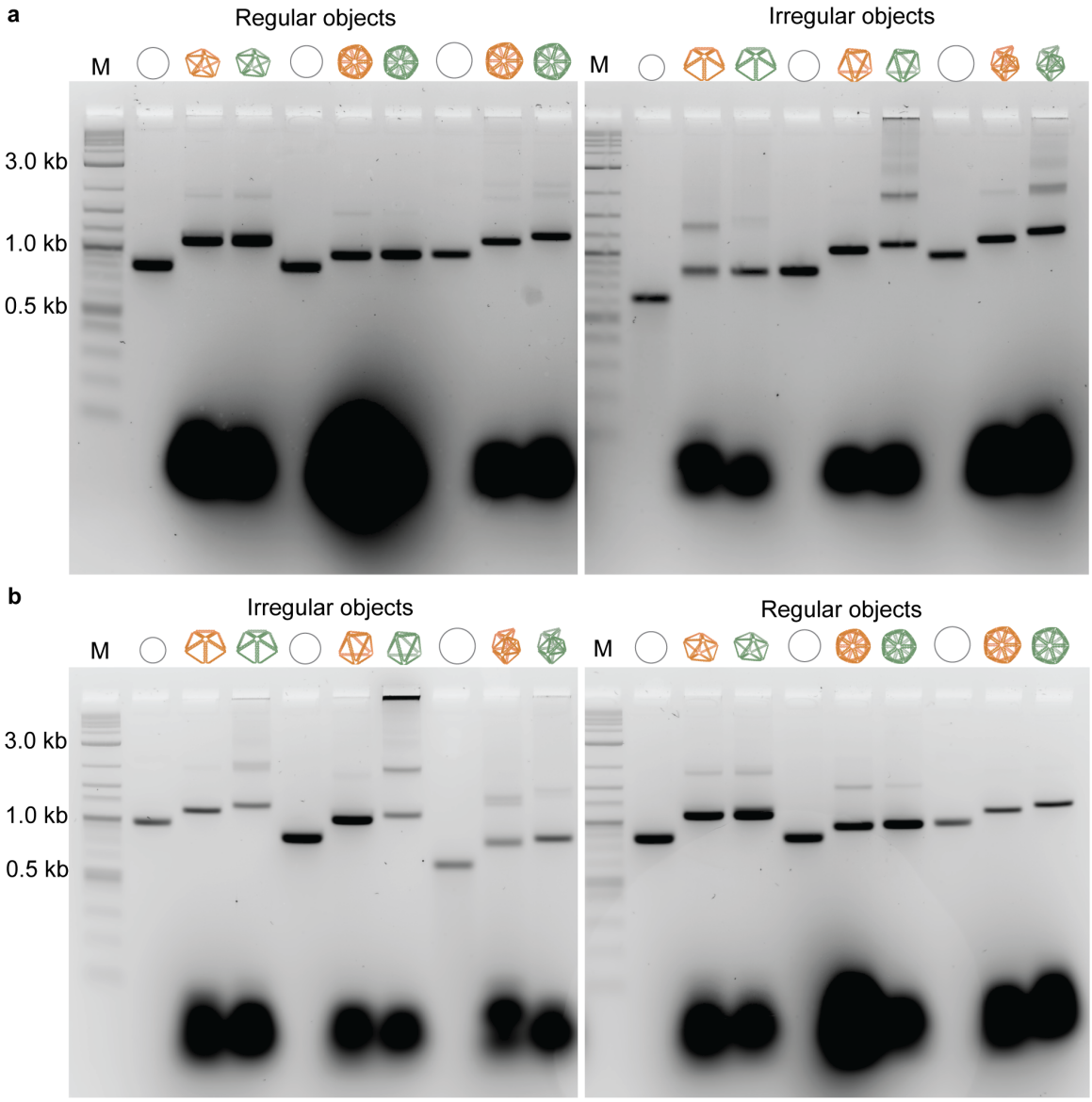

Figure 2-4: **Gel comparison of objects designed with DAEDALUS vs. DAEDALUSv2. (a)** The DAEDALUSv2 objects (green) were folded with the 2h folding ramp, while the DAEDALUS objects (orange) were folded with the 12.5h folding ramp. **(b)** Both DAEDALUS (orange) and DAEDALUSv2 (green) objects were folded with the 12.5h folding ramp.

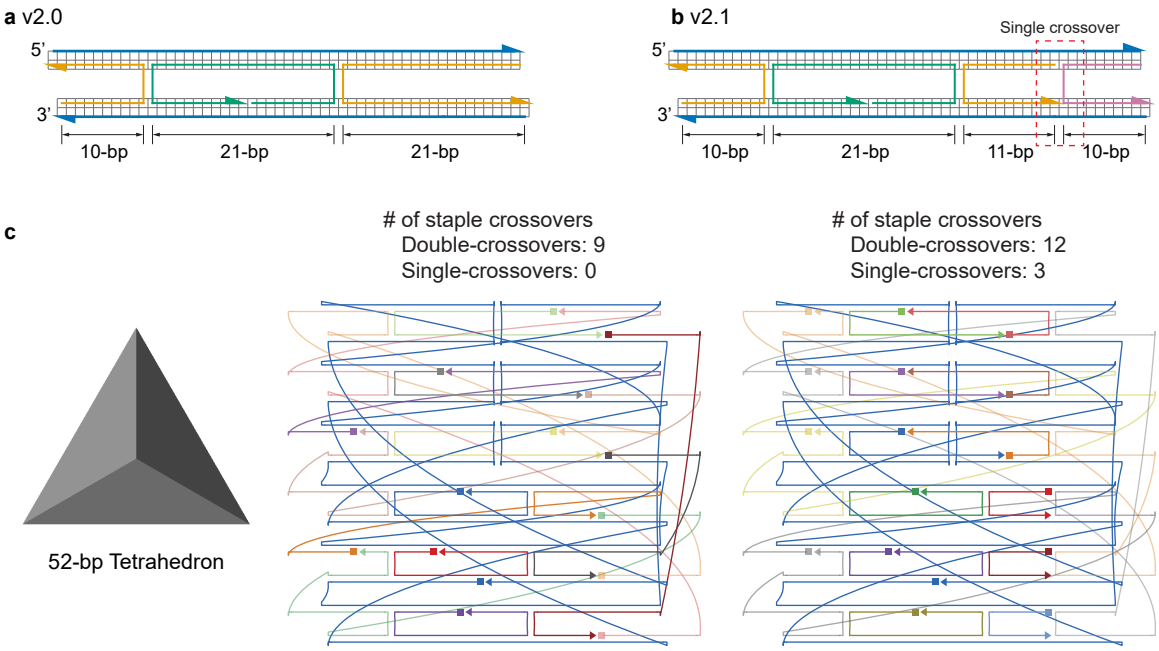

Figure 2-5: **DAEDALUS2 algorithm with a single crossover. (a)** The first iteration of DAEDALUSv2 without single crossovers and **(b)** updated DAEDALUSv2 with a single crossover 10 bases away from a double crossover.

observed to retard movement through a gel [108–111]. Disallowing single crossovers and instead extending the space between a double crossover motif results in one fewer crossover per edge, purportedly leading to this change in flexibility.

A gel mobility shift assay for comparison of the irregular objects designed without (v2.0) and with (v2.1) single crossovers allowed is shown in Figure 2-6. The design of the PP63 did not materially change between these two strategies, as all the edge lengths were even multiples of full helical turns and did not incorporate single crossovers even when allowed, and we accordingly observe no gel shift between design strategies for this object. For the AO63 and the CO63, 3 edges each incorporated a single crossover when allowed—a minor change in flexibility, compared with the 11 additional single crossovers when allowed in I52, and the effect on gel mobility is present but minor. Considering the rigidity that is lost when avoiding single crossovers on edges with an odd number of helical turns, and to allow for 31-bp edge lengths in smaller DNA-NPs, we therefore implemented the use of single crossovers within the final version of DAEDALUSv2, consistent with this choice in the original software

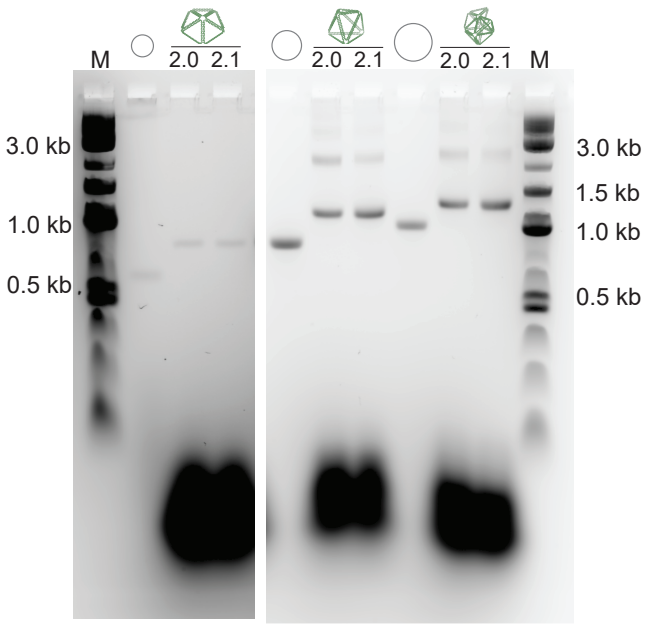

Figure 2-6: **Gel mobility shift assay comparing irregular objects designed without (2.0) or with (2.1) single crossovers.** Note that the pentagonal pyramid, left, does not have single crossovers even when they are allowed by the algorithm.

DAEDALUS [57].

2.2.3 3D cryo-EM reconstruction

Our cryo-EM reconstructions of the wireframe objects validated the 3D structural predictions from DAEDALUSv2. Generally, the reconstructions were correlated more highly with predictions for the regular objects than the irregular objects, due at least in part to the ability to use symmetry for reconstructing regular objects. The PB84 reconstruction achieved a 22.9 Å resolution and fit the algorithm's predicted model with a 0.80 correlation (UCSF Chimera ver 1.14) [112]. The I42 reconstruction achieved a resolution of 18.1 Å with a 0.81 correlation with the predicted model. The I52 reconstruction achieved 17.1 Å resolution and a 0.83 correlation with the predicted model. The reconstructions for these regular objects, along with micrographs showing largely monodisperse well-formed particles, are shown in Figures 2-7 and A-10 to A-12.

Reconstructions and micrographs for the three irregular objects are shown in Fig-

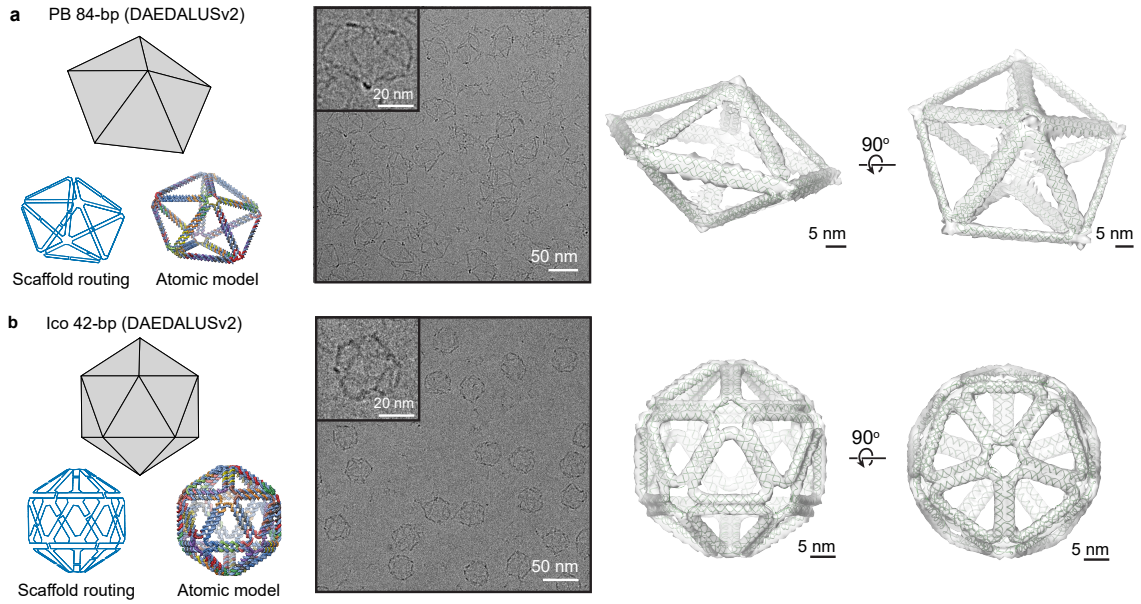

Figure 2-7: **Designing symmetric and regular objects.** Design, cryo-EM micrographs and reconstructions of **(a)** Pentagonal bipyramid (PB) of 84-bp edge-length with DAEDALUSv2 design algorithm. **(b)** Icosahedron (Ico) of 42-bp edge-length with the DAEDALUSv2 algorithm.

ures 2-8 and A-13 to A-15. Both the asymmetric octahedron and the chiral object reconstruction have at least one edge that is ill-defined; due to the lack of symmetry, many more particles are required to properly reconstruct these objects. However, inspection of micrographs and 2D class averages suggests that these particles folded homogenously as designed, although the long edge lengths of the AO63 (some greater than 90 bp) appear to lead to high flexibility. Micrographs for the AO63 and CO63 designed with single crossovers are shown in Figure 2-9.

2.3 Discussion

Taken together, the gel mobility shift assays, AFM, and cryo-EM reconstructions of a variety of objects validate the DAEDALUSv2 designs and their ability to fold with high yields, for both regular and irregular geometries. The pentagonal pyramid (one axis of five-fold rotational symmetry) reconstruction fits well with the predicted model. Although extra webbing-like density at the five-way vertex suggest this geom-

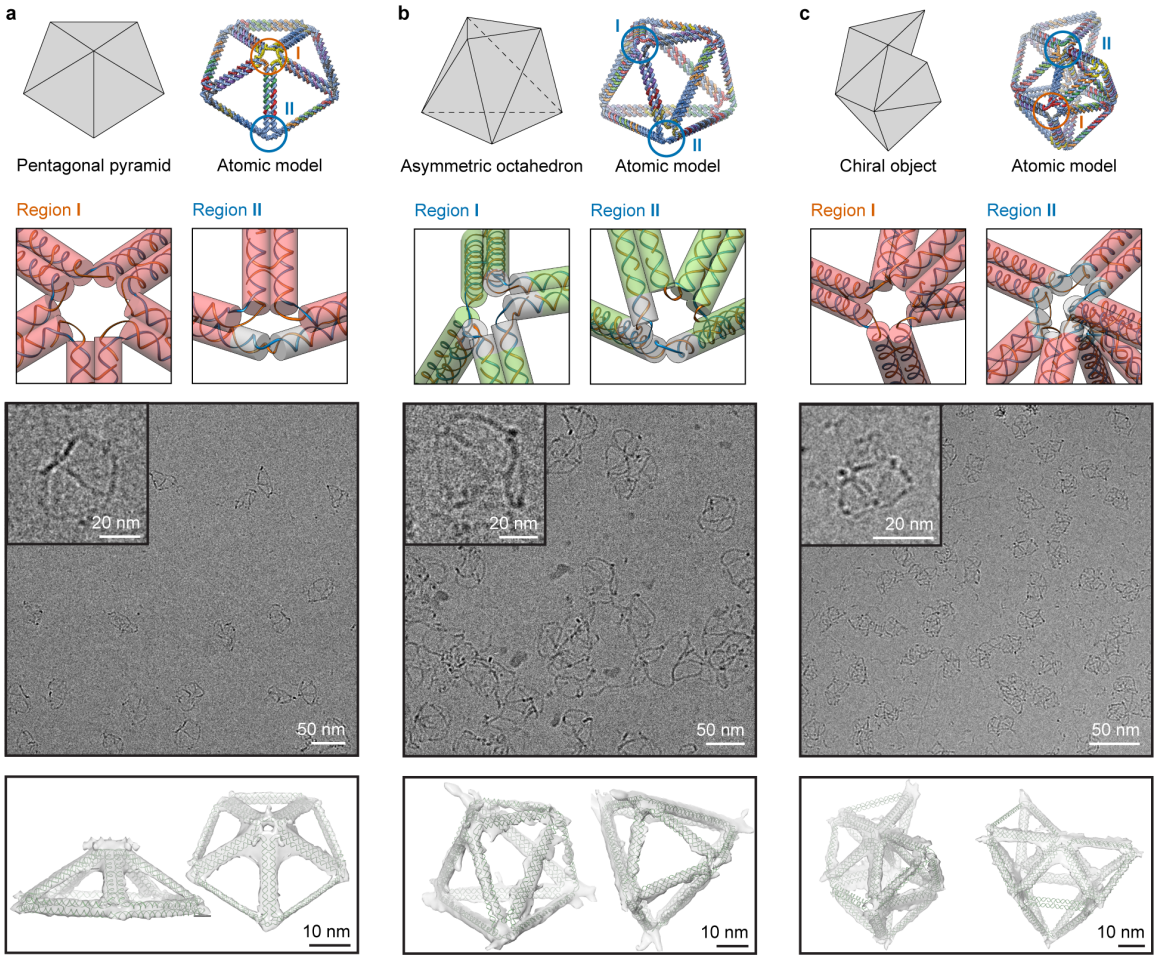

Figure 2-8: **Designing asymmetric and irregular objects with DAEDALUSv2. Design, cryo-EM micrographs and reconstructions of (a)** Pentagonal pyramid of 63-bp edge length, **(b)** asymmetric octahedron of minimum 63-bp edge length, (without single crossovers) and **(c)** chiral object of minimum 63-bp edge length (without single crossovers)

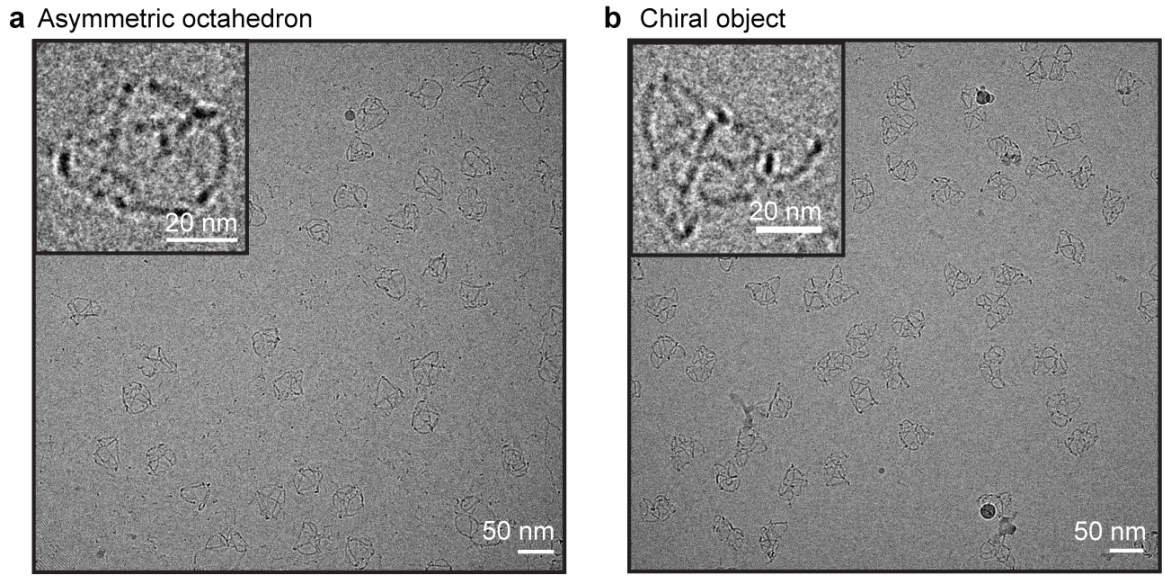

Figure 2-9: **Representative cryo-EM micrographs for irregular objects designed with single crossovers. (a)** asymmetric octahedron with minimum 63-bp edge length, and **(b)** chiral object with minimum 63-bp edge length.

etry has some flexibility in solution, taken with inspection of micrographs, the PP63 appears to have folded as designed.

A potential area for improvement lies in refinements to the folding method to optimize monomer yield across objects. The titration of magnesium chloride and folding times in Figure 2-3 show that moderate improvements are possible with shorter folding ramps and lower amounts of magnesium chloride, and fine-tuning these parameters may increase yield further. Even without such optimization, however, monomer yields for DAEDALUSv2-designed objects are high, with the chiral object achieving 42-52% and all others achieving 73-100% monomer yield (determined from ImageJ analysis of gel images).

DAEDALUSv2 offers an alternative top-down design algorithm for DX-based wireframe origami, with base enumeration and staple routing more similar to the family of other software used for top-down design of 2D DX (PERDIX) and 2D or 3D 6-helix-bundle (TALOS and METIS) origami. By designing edges continuously rather than discretely, the algorithm avoids leaving gaps at vertices when designing objects with multiple edge lengths and vertex angles (Figure 2-1), while staying true to the input

geometry. Unpaired scaffold is allowed at vertices to achieve irregular vertex angles. This concession, as well as following a greedy staple design—ensuring each connecting edge has a minimum hybridization region (6-nt)—then allows the vertex staples to intrude an arbitrary distance into each edge, enabling arbitrary edge lengths beyond just multiples of 10.5 bp (Figure 2-2). DAEDALUSv2 can thus more accurately scale an input geometry than the original DAEDALUS, with the exception of regular geometries. For regular objects with equal edge lengths, multiples of 10.5 bp/edge, the function of the two algorithms is more or less equivalent. In DAEDALUSv2, the staple lengths are kept shorter, making synthesis less costly, although the nick positions are less standardized across the object.

It is worth noting that when dealing with irregular objects, some monomer yield may be sacrificed to prioritize shape fidelity, when comparing objects designed with original DAEDALUS for the same input geometries (Figure 2-4). The original algorithm is thus most useful for applications in which the precise geometry is not crucial but achieving very high monomeric yield before purification is important. DAEDALUSv2 expands to applications that require the design of specific irregular geometries. Figure 2-10 samples the breadth of irregular geometries DAEDALUSv2 can handle. The ability to precisely recreate a desired geometry in DNA could be useful in, for example, templating metals or proteins to create materials with desired mechanical properties, in building a particle around a particular RNA for structural biology or enzymatic purposes [51], or in precisely mimicking viral geometry for vaccines [59]. The top-down design of DAEDALUSv2, with an intuitive user interface in ATHENA [104], should make this precise design of DX DNA-NPs broadly accessible to scientists for creative applications.

2.4 Methods

Top-down sequence design

The new algorithm is provided online for use as a graphical user interface (https://github.com/lcbb/athena) [104] for custom design of 3D wireframe scaffolded

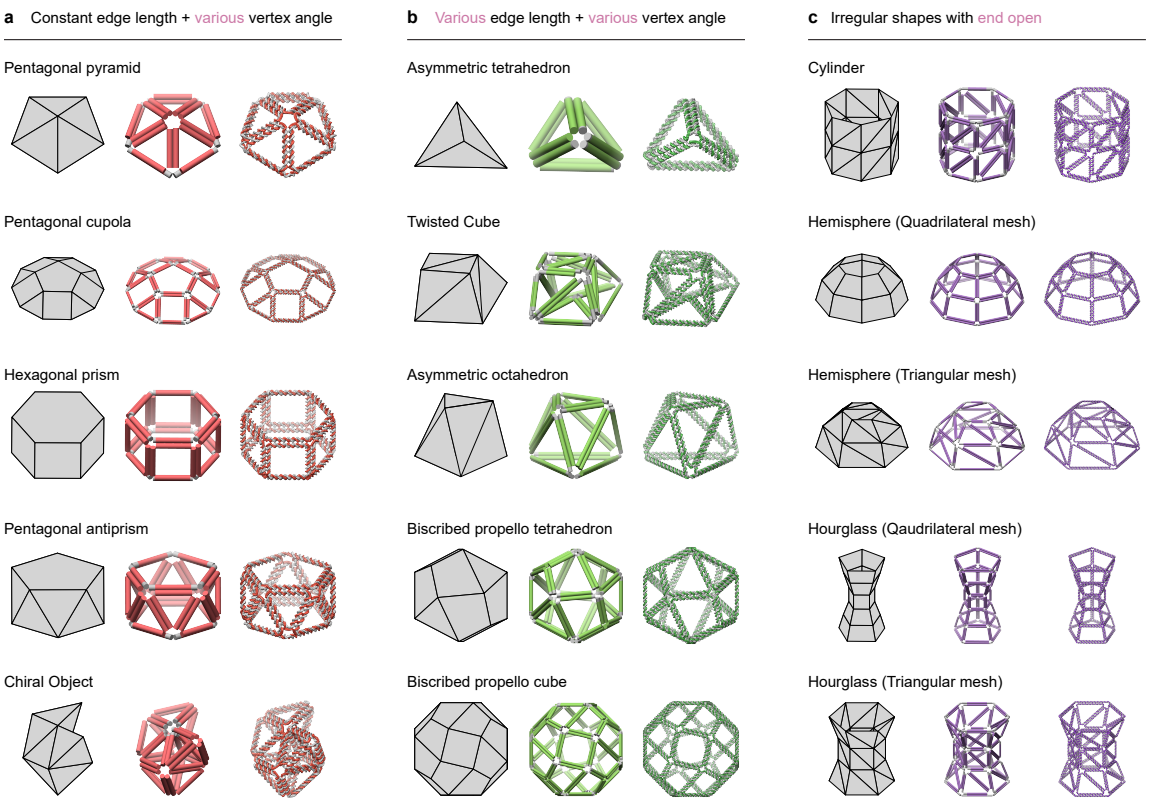

Figure 2-10: **Fully automatic sequence design of 15 diverse scaffolded origami objects with irregular shapes, designed with DAEDALUSv2.** Target geometry, cylindrical model, and atomic model with **(a)** equal edge lengths and unequal vertex angles, **(b)** unequal both edge lengths and vertex angles, and **(c)** with end open.

DNA origami objects. Output files include caDNAno [105] for sequence design editing and oxDNA [113, 114] for coarse-grained simulation of structure and conformational dynamics.

Materials

The three ssDNA scaffolds were produced by using previously described bioproduction methods [115] (sequences in Appendix D). Staple oligonucleotides (see Appendix D for sequences) were purchased from Integrated DNA Technologies (Coralville, IA) in a 96-well plate format at 200 nM concentration in nuclease-free water. Staples were combined into equimolar pools. Tris-acetate-EDTA (TAE, 10×), $MgCl_2$ hexahydrate, and NaCl were purchased from Sigma-Aldrich (St. Louis, MO).

Origami self-assembly

DNA-NPs were folded following the previously published protocol for DX wireframe structures [57], in a solution of 40 nM scaffold, 800 nM staples, 1× TAE, and 12 mM MgCl2, except where other concentrations of $MgCl_2$ are noted (3, 6, or 9 mM). Initially the objects were annealed over the course of 12.8 h on a Bio-Rad T100 thermocycler (Hercules, CA): After a 95°C hold for 5 min, samples were cooled from 80°C to 76°C holding for 5 min at each degree, then from 75°C to 30°C holding for 15 min at each degree, and finally from 29°C to 25°C holding for 10 min at each degree and storing at room temperature. Other folding ramps are listed in Figure S8, where the ramps from 80°C to 25°C are proportionally adjusted at each incubation step so the annealing takes 2h, 4h, or 16h. Folding was initially checked by agarose gel mobility shift assays. 20-22 μl of folded sample was combined with 4 μL of 6× purple loading dye (NEB) and loaded to a 2.5% agarose gel with 1× TAE and 12 mM MgCl2 and 1× SybrSafe (ThermoFisher, Waltham, MA). Each gel was run at 65 V for 3.5 h in 1× TAE with 12 mM $MgCl_2$. Gels were run in an ice-chilled water bath in a cold room and visualized under blue light or using a Typhoon FLA 7000 biomolecular imager. Folded DNA-NPs were purified from excess staples using buffer exchange with Amicon Ultra 0.5ml or 4ml centrifugal spin filters with 100 kDa

MWCO (Sigma-Millipore, St.Louis, MO). After washing the spin filters with water, we exchanged DNA-NPs into buffer composed of 30 mM Tris-HCl and 8 mM $MgCl_2$ by centrifugation at 3200 rpm (approximately 870xg) for 20-45 min at 20°C, diluted approximately 10-fold, and re-concentrated a total of six times.

Cryo-EM data collection and single-particle image

processing Freshly purified and concentrated DNA-NP sample (3 μL) was applied onto the glow-discharged 200-mesh Quantifoil 2/1 grid, blotted for three to five seconds, and rapidly frozen in liquid ethane using a FEI Vitrobot Mark IV (ThermoFisher Scientific). All grids were screened and imaged on an FEI Talos Arctica G2 cryo-electron microscope operated at 200 kV. Micrographs were recorded with a Falcon 3EC direct electron detector in counting mode, where each image is composed of 10 individual frames with an exposure time of 6.5 s and a total dose of 50 electrons per $Å^2$ or with a Gatan K2 Summit direct electron detector in counting mode, where each image is composed of 40 individual frames with an exposure time of 8 s and a total does of 46 electrons per $Å^2$. The nominal pixel size for the PP63 and CO63 was 2.56 Å and for PB84, AO63, I42, and I52 was 1.76 Å. A total of 145 images for the PB84, 1049 images for the PP63, 1012 images for the CO63, 665 images for the AO63, 1098 images for the I42, and 451 images for the I52 were collected with a defocus range of -1 to -4 μm. All the images were motion-corrected using MotionCor2 [116]. Single-particle image processing and 3D reconstruction was performed using the image processing software package EMAN2[117]. All particles were picked manually by e2boxer.py in EMAN2. The initial models generated by the DAEDALUSv2 software were low pass filtered to 60 Å to avoid model bias. The following steps were performed as previously described [55]. A total of 5447 particles for the PB84, 32662 particles for the PP63, 8293 particles for the CO63, 3148 particles for AO63, 24276 particles for the I42, and 17467 particles for the I52 were used for final refinement, applying D5, C5, C1, C1, icosahedral, and icosahedral symmetries, respectively. Resolution for the final maps was estimated using the 0.143 criterion of the Fourier shell correlation (FSC) curve without any mask. A Gaussian low-pass filter was applied to the final 3D maps

displayed in the UCSF Chimera software package [112]. Correlation of each map with its corresponding atomic model is calculated by the UCSF Chimera fitmap function.

Acknowledgments

Funding from the National Science Foundation CCF-1564025, CBET-1729397, and CHE-1839155, the Office of Naval Research N00014-17-1-2609, National Institute of Environmental Health Sciences Center Core grant P30-ES002109, and the National Institute of General Medical Sciences (NIGMS) P41GM103832, R01GM079429, and S10OD021600 are gratefully acknowledged. Research was sponsored by the U.S. Army Research Office and was accomplished under Cooperative Agreement Number W911NF-19-2-0026 for the Institute for Collaborative Biotechnologies.

Chapter 3

Baited capture of an engineered tRNA

As described in Chapter 1, large RNAs often dynamically adopt heterogeneous conformations within a structural ensemble. One approach to stabilize an RNA of heterogeneous structure for cryo-EM imaging and reconstruction is to hybridize segments of the RNA to sequences ("baits") extending from staple ends in a DNA origami structure. We hypothesized that, when hybridized to baits in a sufficient number of positions, a target RNA could be locked into a single conformation, making higher-resolution reconstruction feasible.

3.1 Design considerations

3.1.1 Bait targets

We expected that baits binding to a short stretch of a larger single-stranded region in the target RNA would be less likely to disrupt the RNA 3D architecture than binding to half of a stem. A bait interrupting a native stem structure would release the other half of the stem to potentially find other binding partners in the RNA molecule and thus alter the RNA structure in a cascade of disruptions. Additionally, disrupting an existing long stem would be thermodynamically unfavorable, and the bait might need

to be longer to compensate and achieve stable binding. To identify single-stranded loops to target, we can leverage models of secondary structure, whether from chemical probing experiments like DMS-MaPseq [18] and SHAPE-MaP [16] or derived from models in prior 3D structural studies.

3.1.2 Bait lengths

One concern with such an approach is that binding to baits would in itself disrupt the native RNA structure. We approached our design variables to minimize the likelihood of such a distortion. In considering the length of baits, we reasoned that shorter sequences would reduce the structure disruption by (1) minimizing the RNA sequence directly and locally impacted by the invasion of the bait strand, whether via displacement of an intra-RNA binding partner or via formation of a more-constrained helix at a previously single-stranded loop; and (2) minimizing the enthalpic gain from binding a single bait, thereby reducing the thermodynamic incentive for the RNA to distort from its minimum-free-energy structure to bind to the bait.

However, a minimum bait length must be required for specificity. The actual minimum usable bait length may depend on the length of the target RNA sequence. For example, a section in a particularly short RNA may be specifically targeted with just 3 or 4 nucleotides that are unlikely to recur in that order in such a short sequence, although they might occur many times in a long RNA sequence. A 4-mer has $4^4 = 256$ possible sequences, and in a 100 nt sequence that has 96 4-mers, like a tRNA, a given 4-mer might be expected to occur $96/(4^4) = 0.375$ times (i.e. usually expected to occur at most once in the sequence). Table 3.1 shows this calculation for 4-mers and 8-mers for several different target sequences. For targeting the HIV-1 genome, a 4-nt bait would be insufficient length, with many likely exact matches to the complement throughout the RNA, but an 8-nt bait would likely have no more than one exact complement in the RNA.

Beyond specificity requirements, required bait lengths are also impacted by the need to form a stable complex. The binding affinity of a single bait should be low enough not that the energy released from the RNA binding to it alone is not greater

Table 3.1: Effect of target RNA length on the effective specificity for bait sequences of 4 or 8 nt.

Target sequence	**Length**	**Expected # of matches to a given random 4-mer**	**Expected # of matches to a given random 8-mer**
Engineered tRNA	134	$107/(4^4) = 0.42$	$103/(4^8) = 0.0016$
HIV-1 5'UTR	343	$340/(4^4) = 1.33$	$336/(4^8) = 0.0051$
Full HIV-1 gRNA	9173	$9170/(4^4) = 35.82$	$9166/(4^8) = 0.14$

than the energy involved in displacing intra-RNA interactions, so that distortion of RNA from its native structure remains thermodynamically unfavorable. However, the overall binding affinity of an origami presenting multiple baits with the RNA should be high, so that the complex stably captures the RNA in one particular conformation and remains bound during the sample preparation required for cryo-EM. We can achieve this high overall affinity with a set of low-affinity binding sites through avidity: In multisite binding with connected ligands (such as the different 'prey' regions on an RNA, connected through the rest of the molecule), effective binding affinity for a site increases when a connected ligand binds to a nearby site. In Appendix C, I explain a model for this complexation in more depth.

3.2 Rational design of an origami lock for a tRNA

As a proof of concept for the baited capture approach, we chose to target a Leucine (Leu) transfer RNA (tRNA). tRNA is an attractive initial target because it has well-documented secondary structure with 3 single stranded regions [118, 119]; its approximately 100-nt length is just at the limit of RNA sizes for which conventional techniques have produced tertiary structure models; and it has published 3D models (in the stabilized context of the ribosome or synthetases) [120, 121] we can use to rationally design bait placement and to compare with our 3D structure in the origami complex.

The secondary structure of the tRNA forms a clover-like pattern from several stem-loops: the D-loop, the anticodon loop used to base-pair with mRNA in the

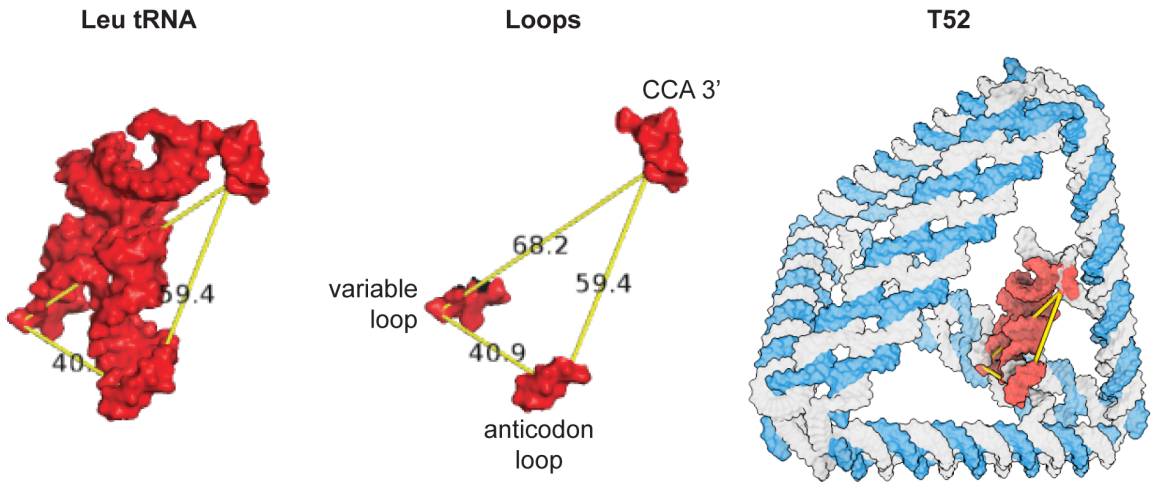

Figure 3-1: **3D spatial matching between a Leu tRNA and a DNA-scaffolded tetrahedron**. Left: the crystal structure of a Leu tRNA (derived from PDB ID: 1WZ2). Middle: the same tRNA model, showing ownly the single stranded loops for targeting. Right: the tRNA model positioned inside the atomic model for a DNA-scaffolded tetrahedron with five helical turns per edge (T52), to align with overhanging baits from staples (grey). The yellow lines mark out the distance between targeted loops.

ribosome, the variable loop just before the $T\Psi C$ loop, and the CCA 3' tail that accepts aminoacylation. The D-loop and the $T\Psi C$ loop interact closely in 3D space to form an overall 'L'-like tertiary structure, in which the variable loop extends from the kink in the 'L' and the CCA 3' tail and the anticodon loop are on opposite ends. Using an atomic model of a Leu tRNA from an x-ray crystallography structure in complex with a Leucyl-tRNA synthetase (PDB ID: 1WZ2 [121]), we aligned the CCA tail, the variable loop, and the anticodon loop with staple nicks on three adjacent edges of the predicted atomic model for a DNA-scaffolded DX tetrahedron with five helical turns per edge (T52), as illustrated in Figure 3-1.

Because the three accessible loops of the Leu tRNA are only 3 to 4 nucleotides long, we engineered the sequence so that each targeted loop could hybridize with at least 8 nucleotides (Figure 3-2). We chose extended loop sequences such that the predicted secondary structure (from Vienna RNAFold [122] and RNAstructure [123] servers) maintained the overall cloverleaf structure, similar to the native secondary structure. We then extended the staple sequences of the T52 so that baits complementary to the

middle 8 nucleotides of the extended CCA 3' tail, variable loop, and anticodon loop were attached at the aligned staple nick positions determined above, with a triple thymine ('TTT') spacer between the origami edge and the start of the bait sequence.

3.3 Binding assays

We used three gel-based binding assays to validate complexation of the *in vitro*-transcribed tRNA and the T52 with three baits: a gel mobility shift assay for a brief binary determination of binding, a depletion assay, and a co-localization assay for quantitative estimates of binding affinity. For all assays, the T52 was folded first on its own and purified with spin filters to remove excess staples, and then annealed with freshly folded RNA using a 1-hour ramp from $42°C$ to $25°C$ followed by a 1 hour hold at $25°C$.

The gel mobility shift assay shown in Figure 3-3A shows an upward shift in the T52 band after incubation with the folded engineered tRNA with extended loops (Ext tRNA). The Ext tRNA is approximately 110 nt, and when associated with the T52 (containing 624 bp = 1248 nt), it corresponds to a 9% increase in total molecular weight, leading to the reduced electrophoretic mobility.

A depletion assay, in which unbound tRNA is able to move into the gel while the origami and any complexed tRNA are stuck in the loading wells, is useful for assaying low-affinity binding. Because it measures unbound RNA, small amounts of binding correspond to high and therefore detectable signal. As visible in Figure 3-3B, increasing T52 concentrations result in less unbound Ext tRNA able to move through the polyacrylamide gel, further confirming complexation.

To quantitatively estimate the binding affinity, we implemented a gel-based co-localization assay. We incubated increasing amounts of Cy5-labeled folded Ext tRNA with 50 nM unlabeled folded T52, and in a gel determined the amount of fluorescent signal co-localized with the SYBR Safe-stained T52 band at each concentration using a fluorometric scanner. The composite image of the co-localization gel is shown in Figure 3-4. Through quantification of the Cy5 band intensities for bound and

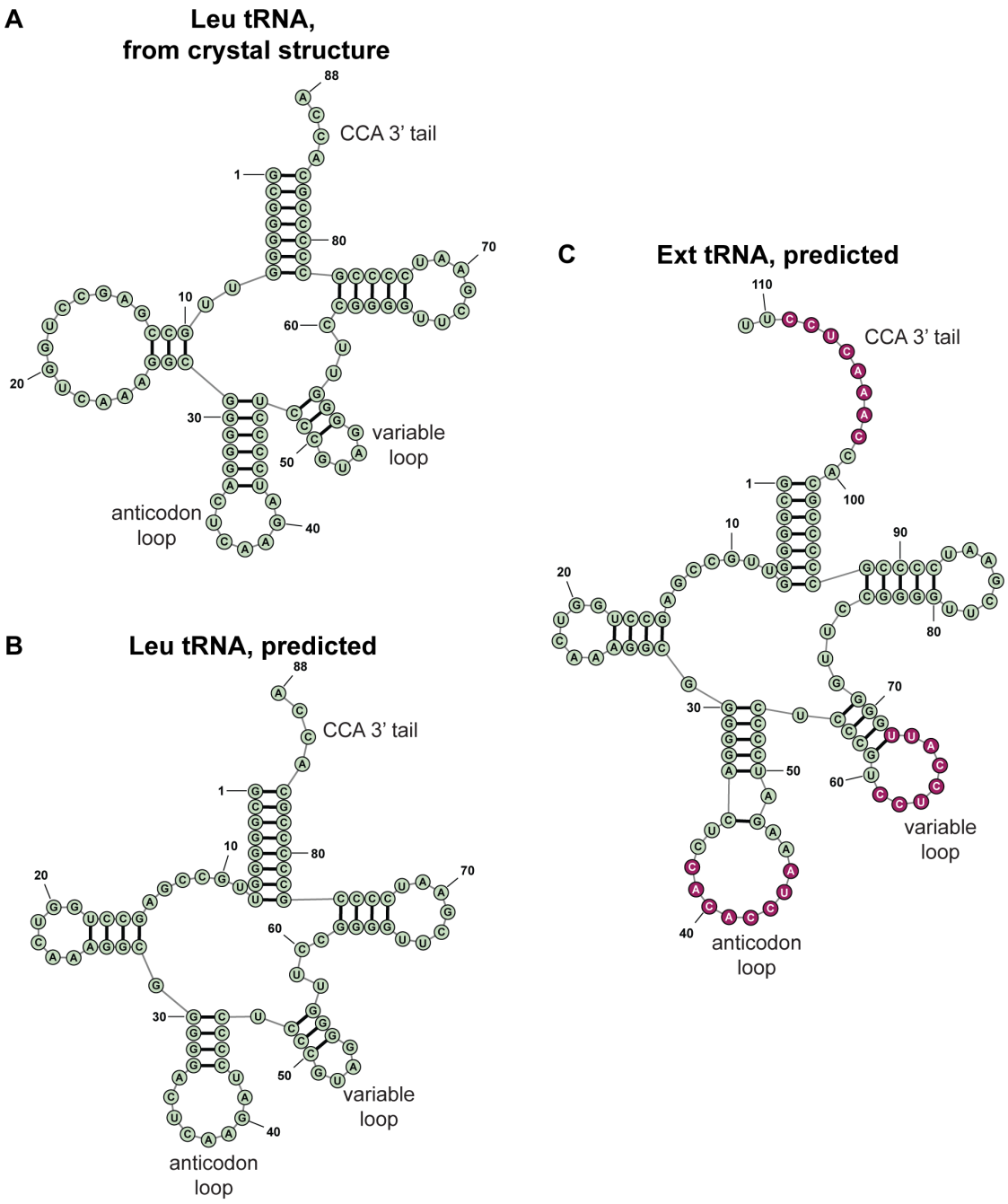

Figure 3-2: **Native vs. engineered Leu tRNA sequence and predicted secondary structure**. Drawings produced with RNAStructure. **(A)** Native secondary structure for a leucine tRNA as determined from a crystal structure of the tRNA bound to a synthetase (PDB: 1WZ2). **(B)** Native secondary structure for a leucine tRNA as predicted by RNAStructure [123]. This structure matches (A) overall with small differences in the start and end points of stems and loops. **(C)** Secondary structure for an engineered tRNA (leucine tRNA sequence modified at the CCA 3' tail, anticodon loop, and variable loop), as predicted by RNAStructure. The purple-shaded nucleotides represent the target sequences of complementary 8-nt baits.

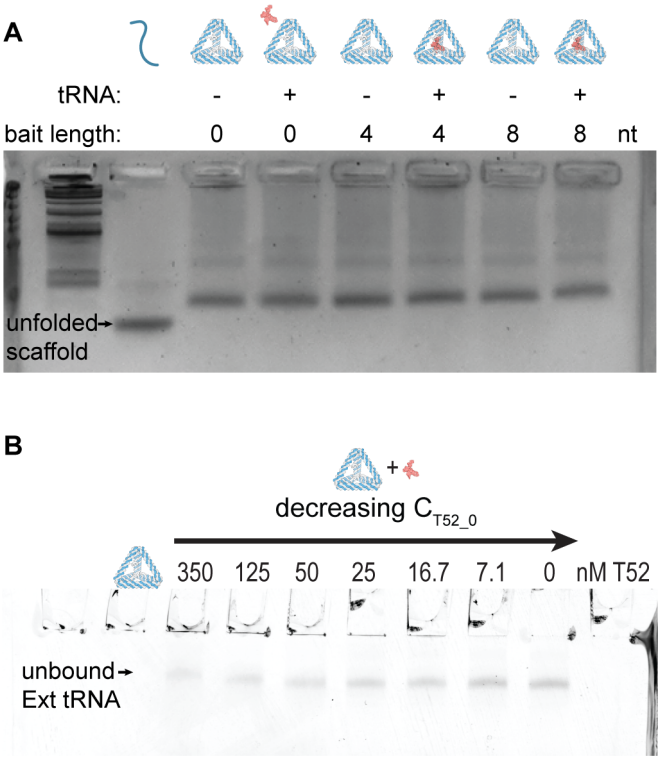

Figure 3-3: **Preliminary gel-based assays for baited capture of tRNA. (A)** Agarose gel mobility shift assay for T52 with no baits (bait length 0 nt), with three baits of length 4 nt incubated with the native Leu tRNA, or with three baits of length 8 nt incubated with the Ext tRNA. **(B)** Gel depletion assay for Cy5-labeled Ext tRNA incubated with T52 with three baits of length 8 nt, run on a polyacrylamide gel. Input T52 concentration decreases from left to right, with a constant tRNA input of 50 nM. Imaged with the Cy5 channel on a Typhoon FLA 7000.

unbound tRNA at each input concentration, and assuming a 1:1 ratio of bound Ext tRNA to bound T52, we plotted the binding curve (Figure 3-4) and fit to the Langmuir equation below by varying K_D (derived from the definition of K_D):

$$f = \frac{C_{tRNA}}{C_{tRNA} + K_D} \tag{3.1}$$

Where f is the fraction of T52 that is bound and C_{tRNA} is the concentration of unbound tRNA. From fitting the quantifiable points (visible but not oversaturated) in the colocalization gel, the apparent dissociation constant is approximately $K_D = 170$ nM ± 40 nM. We note that the fit is relatively poor in this instance, as band intensity quantification is somewhat variable. Future work would benefit from replicates of the co-localization assay, as well as potentially from titrating in folded origami, rather than RNA, for a wider measurable range of data points.

A note for future work

For those looking to implement a different method to measure binding affinities, recent work from the lab of Daniel Herschlag summarizes common pitfalls of thermodynamic characterizations well [124]. For future characterization of RNA/Origami binding thermodynamics, it will be important to establish the time required to equilibrate for accurate equilibrium measurements. In most cases, it is also necessary to assume that the free ligand concentration is effectively the same as the total ligand concentration in order to fit data to the Langmuir equation (3.1) and find apparent K_D, thus requiring the ligand to always be present in great excess to the other binding partner. However, in the gel-based co-localization assay, we can determine both the bound and the unbound tRNA concentration separately, and therefore we do not need to maintain a much lower concentration of origami.

The gel-based co-localization is limited in both throughput and dynamic range, however. Although lower concentrations of RNA can be made more detectable by incorporating a higher percentage of Cy5-modified nucleotides during transcription, this lowers the threshold of RNA for oversaturation.

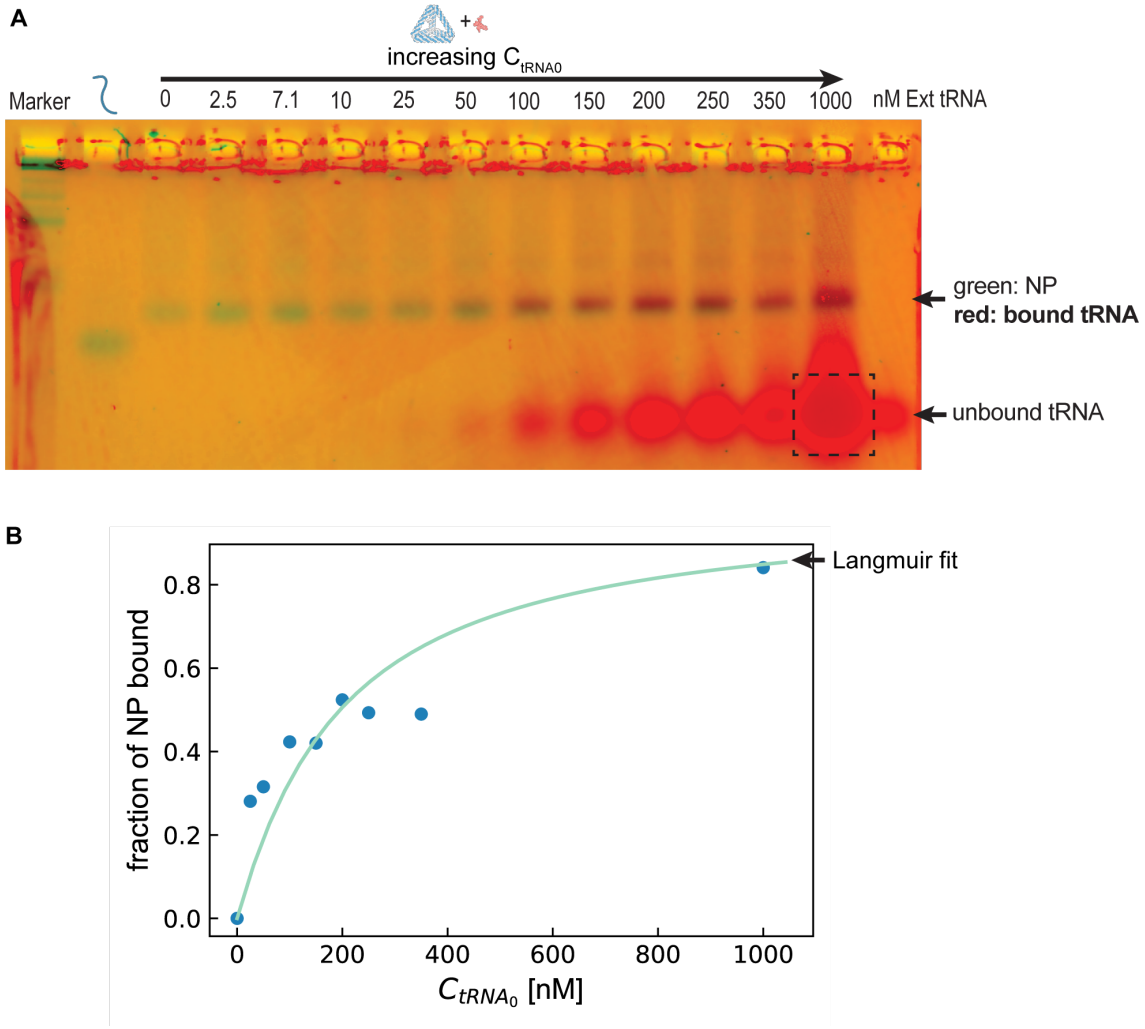

Figure 3-4: **Co-localization gel assay for Ext tRNA and T52 with three baits**, imaged with a Typhoon FLA 7000 biomolecular imager. **(A)** Overlay of the Cy5 channel (in red) on the SYBRsafe imaging channel (in greed) for the co-localization gel. Input Ext tRNA concentration increases from left to right, with a constant T52 input of 50 nM. The dashed box highlights the oversaturated signal from the unbound Ext tRNA band at 1 mM input Ext tRNA. **(B)** Binding curve derived from the intensities in the co-localization image above, for the lanes with 25 nM $\leq C_{tRNA_0} \leq$ 350 nM (detectable but not oversaturated Cy5 signal). The line in green represents the predicted curve from fitting the Langmuir equation (3.1) to the measured data to find K_D.

3.4 Structural characterization with Cryo-EM

Having established tight binding between the Ext tRNA and the T52 with three 8 nt baits, we characterized the structure of the complex using cryo-EM. To prepare the sample for cryo-EM, we annealed purified folded T52 with 5x molar excess folded Ext tRNA, and then washed away excess tRNA and concentrated the complex using an ultracentrifugal spin filter. We separately prepared a sample of concentrated Ext tRNA alone for comparison.

Representative micrographs for the complex and tRNA alone samples are provided in Figure 3-5. In the micrographs for Ext tRNA alone, the individual molecules are too small and too close to background to clearly pick out particles, precluding the possibility of reconstruction. In contrast, the micrograph of the T52-tRNA complex sample shows many monodisperse, easy-to-identify particles. Even absent issues of heterogeneous RNA structure, the approach of capturing RNAs in an origami frame can improve RNA tertiary structure determination through facilitation of particle picking.

Collaborator Dr. Kaiming Zhang in the Chiu Lab at Stanford University processed and reconstructed the cryo-EM data. After classification of 13,346 manually picked particles, 23% contained density inside the T52, 60% were empty, and the remainder were poorly folded or had contaminating surface ice and needed to be discarded. Reconstruction of each class achieved density maps with 17Å resolution.

The interior density, corresponding to the captured Ext tRNA, connects with the wireframe density at three adjacent edges of the tetrahedron, consistent with the placement of the three overhanging baits designed to hybridize with the tRNA. The size of the interior density is similar to that of the model Leu tRNA used for the rational design, although we would not expect the density to exactly match the model due to the engineered sequence with longer loops, the fact that the tRNA is not bound to the synthetase and its lack of post-transcriptional modifications common to tRNAs. Overall, the Ext tRNA and T52 reconstruction shows the expected complexation and provide a tertiary structure for an RNA that could not be determined without

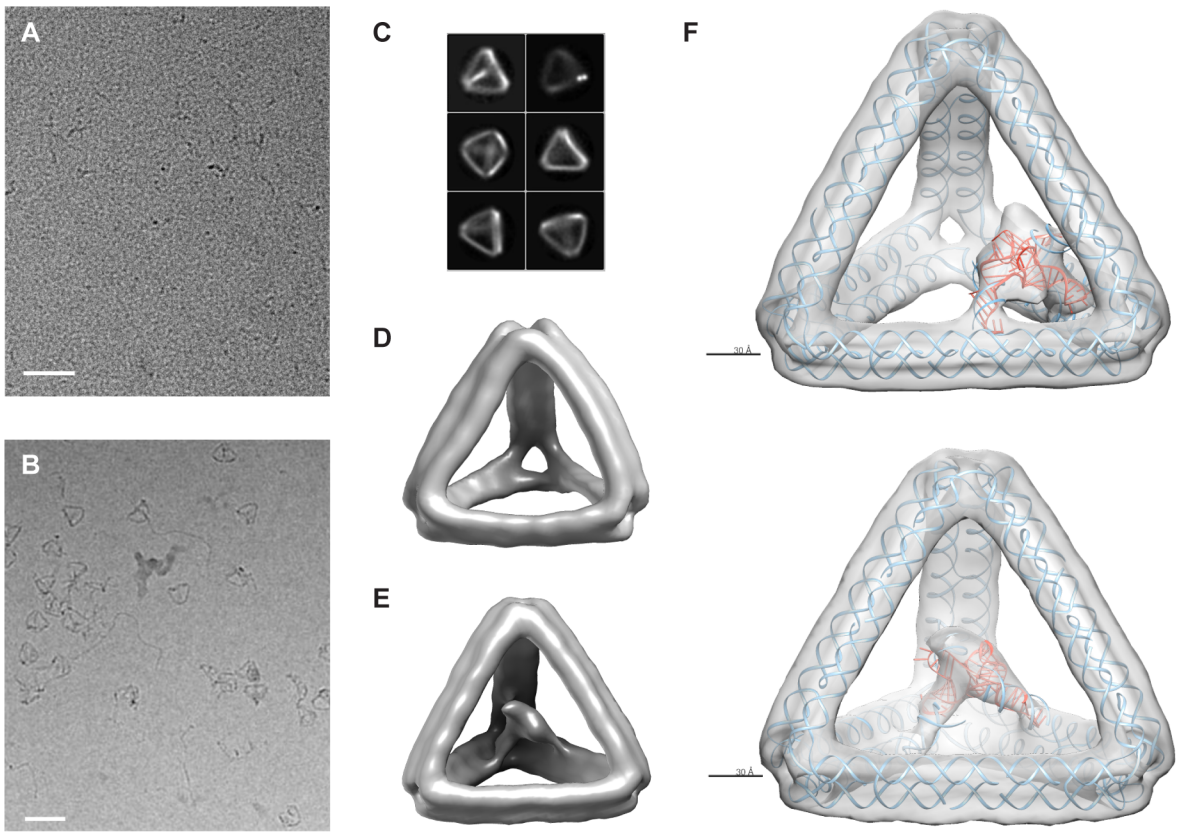

Figure 3-5: **Cryo-EM characterization of an engineered tRNA captured in a DNA tetrahedron.** **(A)** Example micrograph for folded Ext tRNA alone (scale bar: 50 nm). **(B)** Example micrograph for folded Ext tRNA after incubation with T52 (scale bar: 50 nm). **(C)** Example class averages from the T52 + Ext tRNA sample. **(D)** Reconstruction of the particles without tRNA density. **(E)** Reconstruction of the particles with tRNA density. **(F)** Fit of the T52 atomic model and the model of the native Leu tRNA in the density map (scale bars: 3 nm).

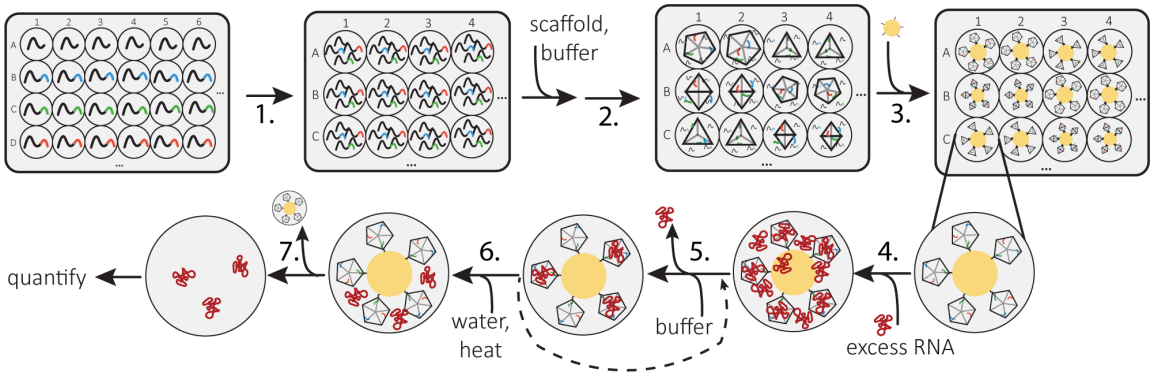

Figure 3-6: **Overview of a high-throughput workflow to identify origami locks with bait arrangements that match the geometry of target RNAs, for stabilization and subsequent tertiary structure characterization.** The overall steps are to 1. form unique staple pools (each including a biotinylated staple), to which to add scaffold; 2. subsequent folding; 3. attach the folded origami to streptavidin-coated magnetic beads and wash away excess staples; 4. to each well, add target RNA in excess; 5. 2 washes to remove unbound RNA; 6. elute remaining bound RNA; 7. extract the eluted RNA for quantification.

complexation (with existing techniques).

This proof-of-concept demonstrates the promise of the baited capture approach for determining otherwise inaccessible RNA tertiary structures.

3.5 A high-throughput platform to capture unknown structure

For the proof-of-concept application described above, we were able to rationally design a geometric match between baits and RNA loops because of an existing 3D structure model. To capture and characterize RNAs of unknown tertiary structure, many arrangements of baits would need to be tested to identify geometric matches. We thus need a method to synthesize, and test for binding, a large library of wireframe origami for each target RNA. I have designed the architecture for such a workflow and validated the individual components, described below and outlined in Figure 3-6.

3.5.1 An algorithm to identify bait attachment positions

With no information about the tertiary structure of a target RNA, the design space to cover to be sure of a geometric match with baits is essentially infinite. To make the design space more tractable, we can use information from computational simulations, Förster Resonance Energy Transfer (FRET) experiments, or other intramolecular distance data to inform bait placement. I wrote an algorithm (further refined by undergraduate mentee Sal Ibarra and graduate student Matthew Allan) that would read in the coarse-grained atomic model output from the Bathe BioNanoLab's top-down origami design software and search for sets of base pair positions whose distances matched user input distances. The structure of the algorithm is shown in Figure 3-7.

The algorithm accepts as inputs the coarse-grained origami model or models output from DAEDALUS or other origami design software (CanDo format), a list of staple sequences (also output form DAEDALUS), a set of point coordinates or a set of distances between points, and the desired number S of output designs. In a first pass, the algorithm determines the set of possible origami nucleotides at which to place an overhang: staple nucleotides positioned towards the inside of the origami wireframe and not near a crossover, to maintain a minimum hybridization region length of 7 nt. Next, one of these allowable nucleotides is randomly selected to attach the first bait. The algorithm then searches through the remaining allowable nucleotides until it identifies one that is a distance d_1 from the first nucleotide to attach the second bait, where $d1$ is the user input distance between point 1 and point 2 on the target RNA, with some tolerance. Again, the algorithm searches through remaining nucleotides to find one the appropriate distance from the first two placed baits to attach the third bait. If no such nucleotide can be found, a different second bait placement is chosen that is still d_1 from the first bait, and the search for third bait placement continues. The nucleotide search process repeats until origami positions on the origami matching all input points are found, and the set of chosen nucleotides for bait attachment from this process is considered one solution.

Solutions are identified until no more exist or until the algorithm has identified S

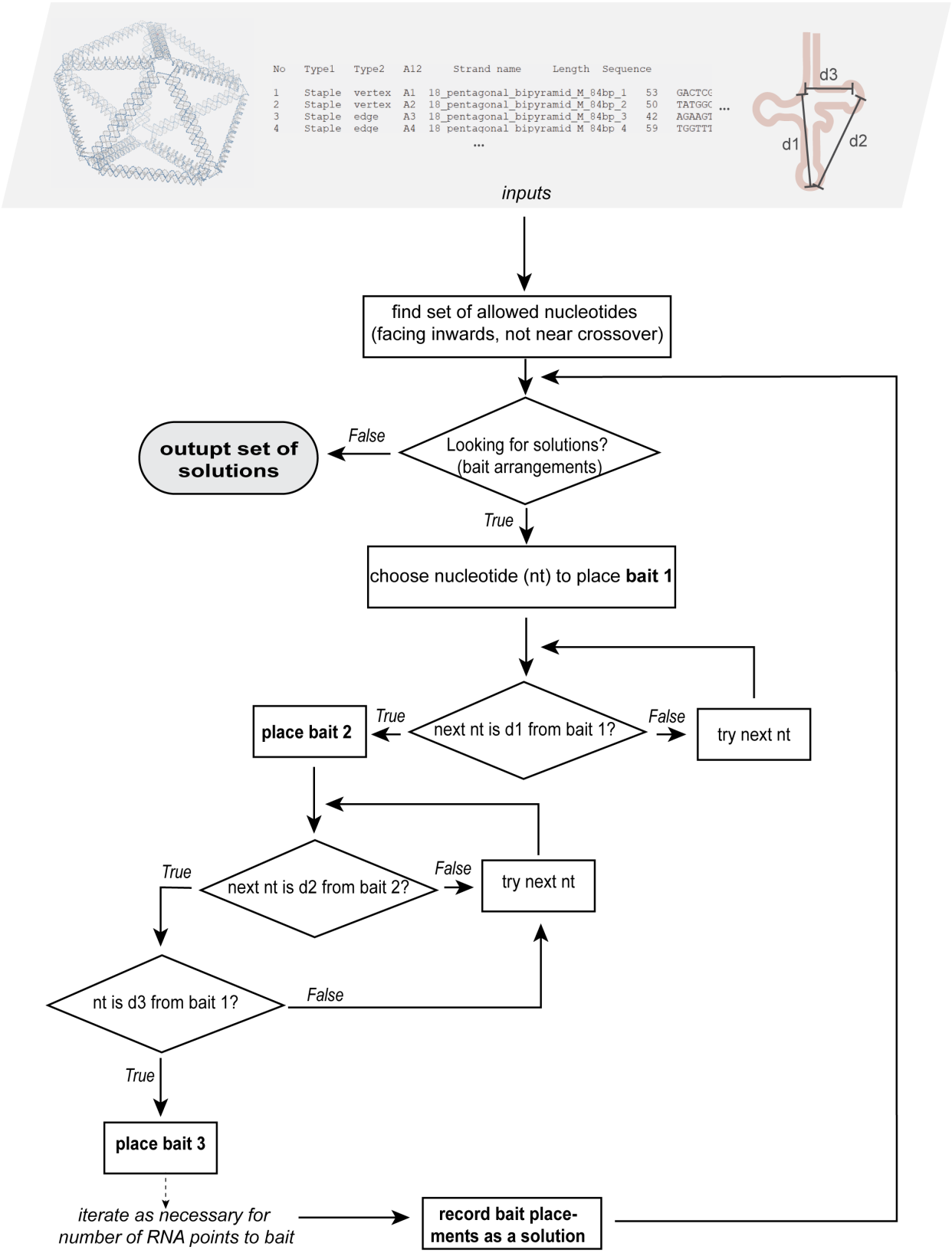

Figure 3-7: **Code architecture for the bait placement algorithm.**

solutions–the user input max number of solutions to find. We included the S input because a particular origami may have many hundreds of solutions and the researcher may only plan to test, for example, one 96-well plate of solutions per origami to start. Outputs from the algorithm include (1) a file with a list of staple sequences to order, with bait sequences appended according to the solutions identified, (2) a file of commands for the LabCyte Echo liquid handler to combine the output staples as ordered in (1) to form S pools for folding origami with varied bait placement, and (3) graphical sketches of the arrangement of baits in each solution.

Figure 3-8 shows the application of this algorithm to generate other capture designs for the Ext tRNA characterized above. Undergraduate mentee Sal Ibarra generated solutions for four different origami designs, using distances from either the Leu tRNA crystal structure or from the bait placement in the rationally designed T52 that captured Ext tRNA as inputs. We tested folding with the output staples for two solutions per origami. By gel mobility shift assay, all objects appeared to fold compactly, forming tight bands with slight upward shifts relative to their respective scaffold. For some origami, staples for one solution seemed to lead to more dimerization than the other solution, but the difference in the PP63 was mitigated when we tested folding with less magnesium–conditions which are more appropriate to native RNA structure as well.

A co-localization assay with one of the solutions produced for capturing the Ext tRNA with a truncated tetrahedron with 4 helical turns per edge (TT42) shows that this particular design does not bind the tRNA with high affinity. The fluorescent tRNA signal only colocalizes with the TT42 band for ??nM Ext tRNA. This is much lower affinity than the rationally designed T52 for the Ext tRNA, and possibly suggests that the distance tolerance used in the algorithm is too high. Broadly speaking, including a degree of tolerance is important, to account for error in experimental data used as input and for the flexibility of the single stranded bait sequences with thymine linkers. However, too large of a tolerance would produce mostly bait arrangements that do not closely match the physical distances in the target RNA and require testing of many more designs to find those that have high avidity and stably bind.

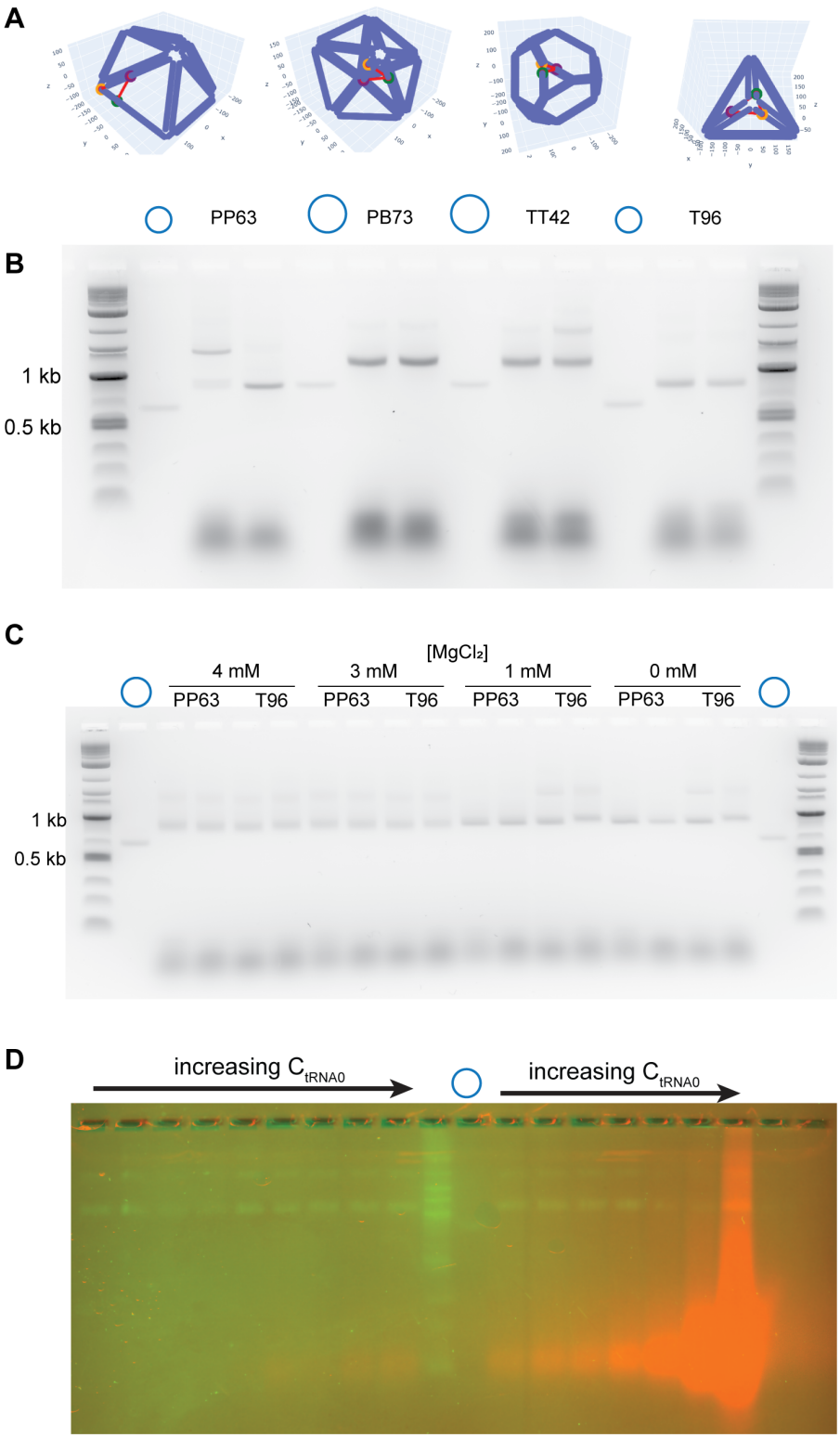
A
PP63
PB73
TT42
T96
B
1 kb
0.5 kb
C
[MgCl₂]
4 mM
3 mM
1 mM
0 mM
PP63
T96
PP63
T96
PP63
T96
PP63
T96
1 kb
0.5 kb
D
increasing C tRNA0
increasing C tRNA0

Figure 3-8: **An example application of the bait placement algorithm to the Ext tRNA. (A)** Graphical sketches of an identified solution for bait placement on four different origami designs. **(B)** Gel mobility shift assay for two solutions each of the four origami shown in (A), to assay folding. Lanes for the respective circular ssDNA scaffolds are indicated with blue circles. **(C)** Gel mobility shift assay with magnesium titration for the folding of the PP63 and Tet96. **(D)** Co-localization assay for one solution of the TruncTet42 incubated with increasing concentrations of Cy5-labeled Ext tRNA, showing the overlay of the gel imaged with the SYBRsafe channel (in green) and with the Cy5 channel (in red). Concentration of input Ext tRNA increases from left to right, with an interruption in the middle for marker and scaffold reference lanes.

3.5.2 Construction of the origami library

To synthesize a library of origami objects with varied bait placements, we need the ability to generate hundreds of staple pools, but manually pooling just one set of staples can take an hour. We tested the use of a LabCyte Echo liquid handler to pool staples for a pentagonal bipyramid with 7 helical turns per edge (PB73), with random arrangements of five overhanging baits complementary to targets in the 5'UTR fragment of the HIV-1 genome, and formed 48 unique staple pools (100 nl per staple) in approximately one hour. To test folding of a subset of these, we subsequently added scaffold and buffer with a multichannel pipette, although a robotic liquid handler capable of handling volumes 10-100 μl would also work well for large libraries in this instance. Validated by gel mobility shift assay (see Figure 3-9), all tested origami folded compactly.

3.5.3 A binding assay on magnetic beads

Although the ideal binding assay would be one that provides quantitative estimates for dissociation constants between the RNA and the origami, such assays are typically not feasible at the scale of a library of hundreds of origami. The existing automated and parallelized binding assays, like biolayer interferometry (BLI) and surface plasmon resonance (SPR), can operate somewhat closer to the appropriate scale given the appropriate instrumentation. However, these techniques operate by affixing the smaller binding partner, in this case the RNA, to a surface, and that poses both more

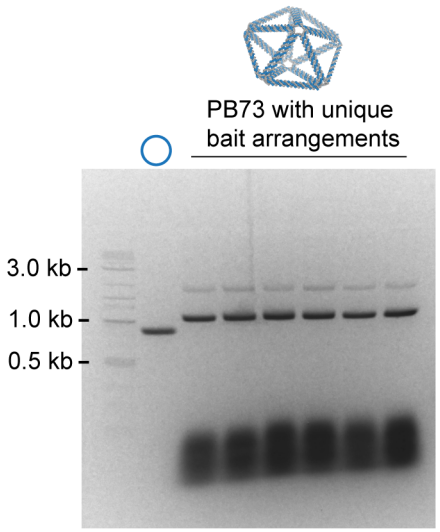

Figure 3-9: **Gel mobility shift assay for origami objects folded using LabCyte Echo-pooled staples**, each with a different arrangement of five overhanging baits that target the HIV-1 5'UTR.

challenges with the RNA and minor concerns about the effect on RNA conformations and binding than attaching the origami nanoparticle.

We thus developed a binding assay in which the origami is affixed, via binding of a biotinylated staple, to streptavidin-coated magnetic beads (Figure 3-6). Once attached to the beads, excess staples can be removed and the origami washed with buffer. RNA is added in 5x molar excess and incubated for 1 hour to allow binding, after which the supernatant is removed and unbound RNA is washed away with buffer. Depending on the strength of the binding interaction and the kinetics of dissociation, some amount of bound RNA will also unbind and be removed during the two wash steps. Finally, any remaining bound RNA is eluted from the origami in water by heating the complex. We quantify the eluted RNA using qPCR, with standards for absolute quantification. The stronger the binding interaction between the target RNA and a particular origami with baits, the more RNA will bind during incubation and remain bound through the washes.

I tested the bead-based binding assay for the T52 that captures Ext tRNA with three baits, and included for comparison the T52 folded with only one or two of these baits. We expected that three-baited design would recover significantly more

RNA, given an increase in affinity derived from more hybridization. Quantification results for nine replicates (three technical replicates in the binding assay and three qPCR replicates each) reveals a mean of 4.3 times greater RNA captured using the three-baited T52 than the one-baited T52 (Figure 3-10). If no cooperativity was in effect, with the low probability of RNA binding at any one site we would expect approximately 3 times greater RNA recovery for the three-baited T52 than the one-baited T52, so the observed increase in captured RNA is consistent with a moderate improvement in binding affinity due to avidity.

I also tested the bead-based binding assay with the five-baited PB73 for capturing the HIV-1 gRNA 5'UTR, using the LabCyte Echo-pooled staples to fold six PB73 with unique arrangements of the five HIV-1-targeting baits. Quantification with a TaqMan qPCR assay, again of three technical replicates in the binding assay and three qPCR replicates each, showed that one presentation of baits ('B4') captured approximately 1.5 times as much HIV-1 5'UTR RNA as the next highest-affinity design, and approximately 6 times as much RNA as the lowest-affinity design tested. We can thus apply the bead-based binding assay to compare a variety of designs and distinguish those that capture the most RNA as 'hits' for further validation and characterization.

3.5.4 Handling an RNA structural ensemble

Because most RNA exists as a structural ensemble, adopting several different conformation and sometimes switching between them, we would expect several significantly different presentations of baits to stably capture the same RNA. Each presentation may align to a different 3D configuration of the RNA in the ensemble. Multiple hits identified from the high-throughput binding assay might thus provide information about the distribution of conformations in the structural ensemble of the target RNA, and cryo-EM characterization of all stable complexes would provide tertiary structure models for each of these conformations. In future applications, we could repeat the binding assay under varied salt and buffer conditions, or including cellular components like RNA binding proteins, and use the degree of RNA recovery by the

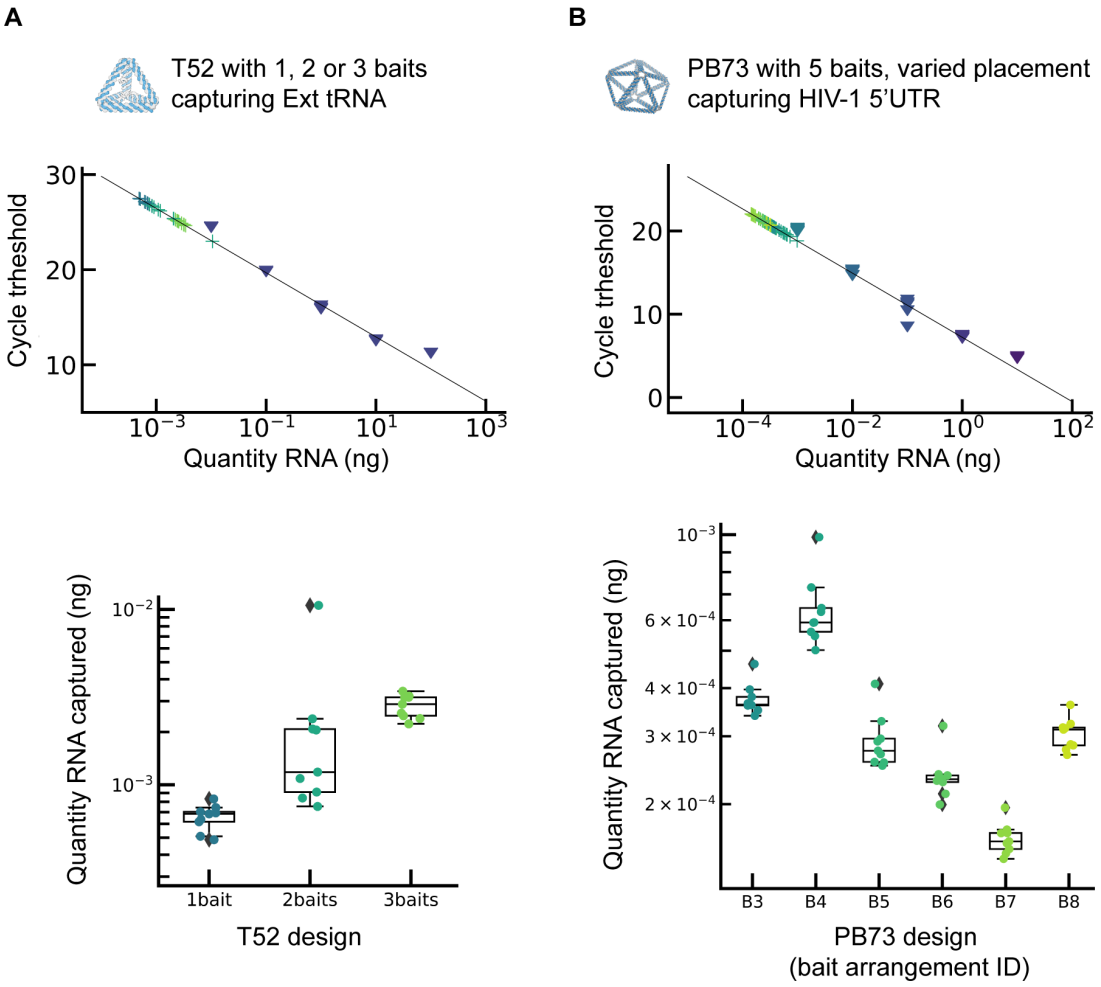

Figure 3-10: **qPCR quantitation of RNA captured with origami in the magnetic bead binding assay**. The top plots show the standard curve (standard points as triangles, data as '+', with nine replicates total for each data and standard type), while the bottom box plots compare the calculated RNA mass. **(A)** Ext tRNA bound and recovered from T52 with 1, 2, or 3 baits. **(B)** HIV-1 5'UTR RNA bound and recovered from PB73 with several different random arrangement of five baits.

different baited origami to evaluate how the distribution of conformations shifts in each condition.

3.6 Methods

Materials

Oligonucleotide staples and primers synthetic DNA sequences were purchased from Integrated DNA Technologies (IDT, Coralville, IA). HIV-1 NL4-3 plasmid DNA, p83-2, was obtained from the NIH Aids Reagent Program (now BEI Resources, NIAID's Microbiology and Infectious Disease Research Reagent Repository). HEPES, KCl, 10X TAE, magnesium chloride, magnesium acetate, high-resolution agarose, and Amicon Ultra 50 kDa and 100 kDa MWCO spin filters were purchased from Sigma-Millipore (MA). HiScribe T7 RNA polymerase kits, Q5 2x HiFidelity PCR mastermixes, and Luna Universal One-Step RT-qPCR kits were purchased from New England Biolabs (NEB, Ipswitch, MA). Quantifoil 200 mesh copper grids with R2/1 spacing were purchased from Electron Microscopy Sciences. Quiaquick PCR purification kits (Qiagen brand) and RNA clean-and-concentrator-5 kits (Zymo Research brand) were purchased from VWR. Dynabeads™ M-270 Streptavidin magnetic beads, SuperScript II Reverse Transcriptase, and the gel stain SYBRsafe (10,000x concentration) were purchased from Thermo Fisher Scientific (MA). Cyanine 5-AA-UTP was purchased from TriLink Biotechnologies (CA). were purchased from Thermo

RNA transcription and folding

Full sequences for the native leucine tRNA, engineered tRNA, and HIV 5'UTR templates and primers are included in Appendix D. For the tRNA templates, due to their short length, we split each sequence into two overlapping long forward and reverse "primers," which we then extended using Q5 2x HiFidelity PCR mastermix (NEB) using only 4 cycles of PCR (72° annealing temperature and 10s extension time), with 0.5 μM of each primer in a total 400 μl reaction. For the HIV 5'UTR DNA template, We amplified DNA from the plasmid p83-2 (NIH Aids Reagent Program) using Q5

2x HiFidelity PCR mastermix (NEB) according to manufacturer instructions, using forward primers containing the T7 promoter sequence. We purified the DNA templates using a Qiaquick PCR purification kit (Qiagen) according to manufacturer instructions.

From each DNA template, we transcribed RNA using a HiScribe T7 RNA polymerase kit (NEB) with overnight incubation at 37°C. For the tRNAs, the suggested modified protocol for short transcripts was used (1.5 μl each of 10X T7 reaction buffer, each 100 mM NTP, and T7 polymerase mix, instead of 2 μl each). For Cy5-labeled transcripts, we used approximately 1 μl 7.7 mM Cy5-AA-UTP (TriLink) in addition to 1.5 μl of 100 mM unlabeled CTP in a 20 μl transcription reaction, such that approximately five Cy5 labels would be incorporated per transcript on average. We then treated the 20 μl finished transcription reactions with 2 μl DNase I for 15 min at 37°C before purifying with Zymo RNA clean-and-concentrator-5 kits (Zymo Research), following manufacturer instructions except for an extra wash step.

To fold RNA, we first prepared and syringe-filtered 3x folding buffer: 30 mM HEPES-KOH (from 2M pH 7.5 stock), 300 mM potassium glutamate, and 36 mM magnesium glutamate. added approximately 40 pmol purified RNA to nuclease free water for a total volume of 12 μl and denatured the RNA at 95°C for 2 min, then placed immediately on ice for at least 2 minutes. To this rapidly-cooled denatured RNA, we added 6 μl of 3x folding buffer and incubated the mixture at 37°C for 20 minutes.

Origami self-assembly

Sequences for the scaffold and staples, including overhanging baits, are included in Appendix D. Staple oligonucleotides (IDT) at a stock concentration of 200 μM were pooled either manually or with the LabCyte Echo 525 Acoustic Liquid Handler. Origami nanoparticles were folded using 20 or 40 nM scaffold and 20x molar excess of each staple (400 or 800 nM respectively) in 1x TAE and 12 mM $MgCl_2$, annealed in a 12.8-h overnight folding ramp as described previously [57]. Folded origami was purified with 0.5-ml Amicon Ultra spin filters with 100 kDa molecular weight cut off

(Sigma-Millipore). After washing the spin filters with water, we exchanged the folded origami nanoparticles into buffer composed of 30 mM Tris-HCl and 8 mM $MgCl_2$ by centrifugation at 3000 rpm (approximately 850 xg) for 20-45 min at 20°C, diluted approximately 10-fold, and re-concentrated a total of six times.

Gel-based binding assays

Gel mobility shift assay: We mixed 250 nM of freshly folded unlabeled RNA with 50 nM purified origami, adding a buffer composed of 50 mM HEPES, 100 mM KCl, and 10 mM $MgCl_2$ as necessary for a total reaction volume of 15 μl. We annealed the samples in a ramp from 37°C to 25°C over 1 hour before loading to a 2% high-resolution agarose gel in 1x TBE, 12 mM $MgCl_2$ and 1x SYBRsafe (Thermo Fisher). We ran the gel at 65V for 150 min in a 4°C cold room and imaged with blue light.

Depletion assay: We mixed purified origami in varying concentrations (350, 125, 50, 25, 16.67, or 7.14 nM) with 50 nM of freshly folded Cy5-labeled tRNA, adding buffer composed of 30 mM Tris pH8 and 12 mM $MgCl_2$ as necessary to reach a total reaction volume of 10 μ. Each sample was annealed in a Bio-Rad thermal cycler with a 1 hour ramp from 42°C to 25°C (-1°C every 3.5 min) and held at 25°C for 1 hour. For all samples, we added 2 μl of 6x DNA loading dye and loaded directly into a 12% polyacrylamide gel in 8 mM $MgCl_2$ and 1x TBE. We measured the Cy5 signal of the tRNA using a Typhoon FLA 7000 imager (GE).

Co-localization assay: We mixed freshly folded cy5-labeled tRNA in varying concentrations (0, 2.5, 7.14, 10, 25, 50, 100, 150, 200, 250, 350, or 1000 nM) with 50 nM purified origami, adding buffer composed of 30 mM Tris pH 8 and 12 mM $MgCl_2$ as necessary to reach a total reaction volume of 10 μl. Each sample was annealed with a 1 hour ramp from 42°C to 25°C (-1°C every 3.5 min) and held at 25°C for 1 hour. We then added 2 μl of 6x DNA loading dye (no SDS) to all samples and loaded them directly into a 2% high-resolution agarose gel in 1x TBE with 8 mM $MgCl_2$. After running the gel at 60V for 3 hours in a 4°C cold room, we imaged the gel with a Typhoon FLA 7000 imager (GE) using the Cy5 channel. We then post-stained the gel in 1x SYBRsafe (Thermo Fisher) and 1x TBE and re-imaged with the SYBRSafe

channel, and repeated the Cy5-channel imaging for alignment. Band intensities were quantified with Fiji (Fiji Is Just Image J) [125], taking the average of 3 area-under-the-curve measurements for each intensity. For the final lane of $C_{tRNA_0} = 1mM$, in which the band for unbound tRNA was oversaturated, we estimated the bound tRNA concentration from the average ratio of known tRNA input concentration to total tRNA signal in other lanes, and then estimated unbound tRNA concentration from $C_{tRNA_{unbound}} = C_{tRNA_0} - C_{tRNA_{bound}}$.

Cryo-electron microscopy

We annealed 175 nM purified Tet52 (with three 8 nt baits) with 870 nM of freshly folded, unlabeled Ext tRNA (5x molar excess RNA) in 30 mM Tris and 8 mM $MgCl_2$ in a total volume of 350 μl using the 42°C to 25°C ramp over 1 h followed by an hour hold at 25°C. We added 150 μl of the 30 mM Tris and 8 mM $MgCl_2$ buffer to the annealed sample and concentrated it in a 0.5-ml Amicon Ultra spin filter with 100 kDa MWCO (Sigma-Millipore) for approximately 1 hour and 20 min at approximately 850xg. The estimated final concentration (measured on a Nanodrop One microvolume UV-vis spectrophotometer) was 780 ng/μl, or approximately 1.2 μM. 3 μl of this concentrated annealed complex was applied immediately to a Quantifoil copper grid and flash frozen with a Vitrobot automated plunger at the SLAC National Accelerator laboratory. Grids were screened and imaged with a Talos Arctica cryo-electron microscope. We collected 600 micrographs with a 1.37 Å nominal pixel size. Micrographs were motion-corrected using MotionCor2 [116], and single-particle image processing and 3D reconstruction was performed using the image processing software package EMAN2[117]. 13,346 particles were picked manually by e2boxer.py in EMAN2.

Magnetic bead-based capture assay

To prepare origami for a high-throughput magnetic bead assay for RNA capture, we assembled origami as described above except that one staple oligonucleotide was biotinylated at the 5' end following a double thymine spacer (synthesized at IDT).

For the DNA tetrahedron with 52 bp per edge, the biotinylated staple was added at only 10x molar excess to the scaffold, where all others were added at 20x molar excess, to reduce competition between folded origami and excess staple for binding streptavidin. Folded origami was diluted to 7.5 nM with buffer (30 mM Tris-HCl, 12 mM $MgCl_2$ before aliquoting into a deep-well plate, without spin filter purification. We aliquoted each type of origami into three wells for technical replicates.

RNA was folded as described above except in a modified buffer: 100 mM HEPES pH 7.5, 10 mM $MgCl_2$, and 100 mM KCl, and with RNA at 37.5 nM (so that it would be at approximately 5x molar excess to the origami based on predicted origami binding to streptavidin beads).

We used the Kingfisher Flex (Thermo Fisher) to perform the bead-based capture assay. First, M-270 streptavidin-coated magnetic beads (Thermo Fisher) were collected onto the instrument's magnetic head from wells containing 40 μl of vortexed beads each. These were then washed in 300 μl of buffer composed of 30 mM Tris-HCl and 12 mM $MgCl_2$ for 30 seconds with medium mixing. The washed beads were recollected and deposited in wells of 100 μl folded origami at 7.5 nM for 15 minutes with slow mixing to bind the biotinylated origami. The beads with bound origami were then recollected and washed again in 300 μl of buffer for 1 min with slow mixing. To capture the folded RNA, the origami-bound beads were deposited in wells of 100 μl of 37.5 nM folded RNA and left to incubate at room temperature, with slow mixing, for approximately 1 hour. Beads were collected from the RNA well and washed in 300 μl of the Tris/$MgCl_2$ buffer for 20 seconds of slow mixing, repeated twice before depositing beads in 50 μl of water to elute any bound RNA for 2 min of slow mixing, pre-heated to 50°C. Beads were collected out of the well, leaving behind the eluant for downstream processing (qPCR)

Quantitative RT-PCR

Quantification of RNA after the bead-based capture assay was performed on a QuantStudio 6 Flex real-time PCR system (Thermo Fisher).

For the Ext tRNA: Standards were prepared of 0.01, 0.1, 1, 10, 100, and 1000

ng/μl of Ext tRNA. For each standard and each Ext tRNA eluant well from the bead-based capture assay, we added 1 μl RNA to a 20 μl reaction of Luna Universal One-Step RT-qPCR (NEB), performed in triplicate and according to manufacturer instructions. On the QuantStudio 6 system, we used the SYBR reporter and ROX as the passive reference.

For the HIV-1 5'UTR: Standards were prepared of 0.005, 0.05, 0.5, 5, 50, and 500 ng/μl of 5'UTR RNA (for 0.001, 0.01, 0.1, 1, 10 and 100 ng of RNA in each final qPCR reaction). For each standard and each HIV-1 5'UTR eluant well from the bead-based capture assay, we mixed 4 μl of RNA with 2 pmol of reverse transcription primer and 10 μmol of each dNTP in a total volume of 12 μl. We heated this mixture to 65°C for 5 min and chilled on ice to denature the RNA, then added 8 μl of reverse transcription mastermix (4 μl 5x FS buffer, 2 μl 0.1M DTT, 1 μl RNase OUT, and 1 μl SuperScript II RTase per reaction) and incubated at 42°C for 50 min to reverse transcribe cDNA, followed by heat inactivation of the enzymes at 70°C for 15 min.

3 μl of each reverse transcription was added to 15 μl 2x TaqMan Fast Advanced MasterMix (Thermo Fisher), 1.5 μl 20x TaqMan Assay Pa03453409_s1 (Thermo Fisher, pre-designed assay targeting the HIV-1 Long Terminal Repeat), and 10.5 μl nuclease-free water. This 30 μl total volume was aliquoted into three wells of 10 μl each in a 384-well plate for 3 replicates of each qPCR reaction. Real-time PCR was carried out and measured with the FAM signal (NFQ-MGB quencher) and ROX as the passive reference on the QuantStudio 6 system.

Chapter 4

RNA-scaffolded 3D wireframe origami[1]

4.1 Introduction

Nucleic acid nanotechnology offers promise in diverse applications, from enzymatic nanoreactors [73] to wafer-scale lithography [70] and therapeutics [126, 127]. The predictability of Watson-Crick-Franklin base-pairing in nucleic acids makes this type of nanotechnology easily programmable. When also employing structural motifs pulled from biology—such as Holliday junction crossovers formed during homologous recombination—nucleic acids can be engineered to form a wide variety of structures. Scaffolded DNA origami [49], in particular, enables the fabrication of nearly arbitrary 2D and 3D dense, bricklike as well as wireframe objects by folding a single-stranded DNA scaffold to user-specified geometries via annealing with short staple strands [54, 55, 57, 58, 91, 97, 128, 129]. For some applications, designs leverage dynamic behaviors in scaffolded origami for controlled cargo release and signaling [127, 130]. The tunable rates of degradation of RNA/DNA hybrid particles by native nucleases

[1]Matthew F. Allan carried out the DMS-MaPseq experiments, and Dr. Shanshan Li and Dr. Kaiming Zhang performed the cryo-EM analysis. Dr. Tyson Shepherd and Dr. Sakul Ratanalert created the initial Python version of DAEDALUS adapted for 11 nucleotides per helical turn and scaffold crossover asymmetry. Dr. Tyson Shepherd also worked on many of the initial experiments to characterize the alternative A-form designs. Hellen Huang performed the biochemical stability assay. **This work is being written up as a manuscript for publication.**

suggest a promising avenue for mediating such dynamic behavior. However, RNA origami has thus far generally been less explored than DNA origami, largely utilizing RNA/RNA interactions [75–77, 79, 131–136], with only several studies probing the ability to fabricate hybrid RNA/DNA origami [78, 137, 138]. And fully automated, top-down sequence design procedures remain sparse [139, 140], although such algorithms have greatly aided scaffold routing and staple sequence design for DNA origami [54, 55, 57, 58, 128]. Full control over both the RNA/DNA composition and the geometry of nucleic acid origami would offer additional application avenues in nanoscale materials synthesis and therapeutics that are not currently offered by either RNA/RNA or DNA/DNA origami alone [83, 137, 141, 142].

RNA shares a similar 4-base code with DNA, including common bases of adenine (A), cytosine (C), and guanine (G), with the exception that RNA uses uracil (U) rather than thymine (T) as a fourth base. Chemically, RNA carries an additional 2'-hydroxyl group on the sugar that forces a C3'-endo form, leading to an A-form double helix (11 base pairs per helical turn, 2.6 Å vertical rise per base pair, 23 Å helical diameter), when hybridized with either DNA or RNA, rather than the B-form of duplex DNA (10.5 base pairs per helical turn, 3.4 Å vertical rise per base pair, 20 Å helical diameter) [143, 144]. RNA is typically single stranded in the cell and therefore adopts complex tertiary folds, often using alternatives to Watson-Crick base pairing with Hoogsteen and sugar edge base interactions, thus rendering reliable de novo structural prediction and programmability more challenging [145, 146]. In spite of this difficulty, knowledge gained from 3D RNA structures has been used to generate RNA nanoparticles by engineering RNA fragments to assemble into programmed higher-order geometries, using, for example, tRNAs and multi-way junctions, to create complex shapes[75, 147, 148]. However, absolute control over programmability of the object is constrained by the required sequence space of the folds.

Beyond this approach, RNA nanotechnology has recently seen significant advances in programmed folding of long single-stranded RNA objects, with predefined tertiary junctions used to assemble complex nanoparticle shapes [76, 77, 133]. While developments in hybrid RNA/DNA origami folding with high yield and purity have been

reported [78, 79, 149], offering the potential of full programmability of geometry, catalytic activity, and chemical stability, the breadth of applications has only partially been explored [85, 150], in part because of a lack of algorithmic methods for staple design and scaffold routing of A-form geometries. No studies have yet been done on the design and fabrication of arbitrary double-duplex (DX) wireframe structures with RNA, having focused principally on brick-like or single-duplex-edged structures. The limited prior work on hybrid RNA/DNA origami suggested that such designs need to account for the extreme pitch and twist of A-form helical geometries to achieve optimal folding yield, and that to accommodate these features, adjacent crossovers of different strands (e.g. scaffold and staple) must be spaced asymmetrically along the helices [80, 136].

The ability to scaffold wireframe origami with RNA, in addition to enabling particular applications as mentioned above, creates an opportunity to study nucleic acid origami folding and stability with nucleotide-level precision. Few methods for nucleic acid origami approach the base-pair level information provided by chemical probing; the nearest may be the use of molecular dynamics simulations combined with detailed cryo-EM characterization [151], and a large-scale Förster Resonance Energy Transfer study that provided hybridization-domain-level information on an origami folding pathways [152]. Hydroxy radical footprinting has been used to investigate stability of DNA 3-way and 4-way junctions46, but to our knowledge chemical probing such as dimethyl sulfate mutational profiling with sequencing (DMS-MaPseq)[18] has not yet been applied to large-scale scaffolded origami.

Here, we have designed and characterized hybrid nucleic acid origami, using in vitro-transcribed RNA scaffold and synthesized DNA staples to create seven wireframe polyhedra with DX edges. Impacts of ionic strength and ionic species on folding were characterized with gel electrophoresis. From cryo-electron microscopy (cryo-EM) reconstruction, we verified the predicted structures. Folding was assessed at single-nucleotide resolution using DMS-MaPseq, which we were further able to use for insights into sources of instability in origami design. We modified the open source software DAEDALUS to implement our A-form DX wireframe design rules in

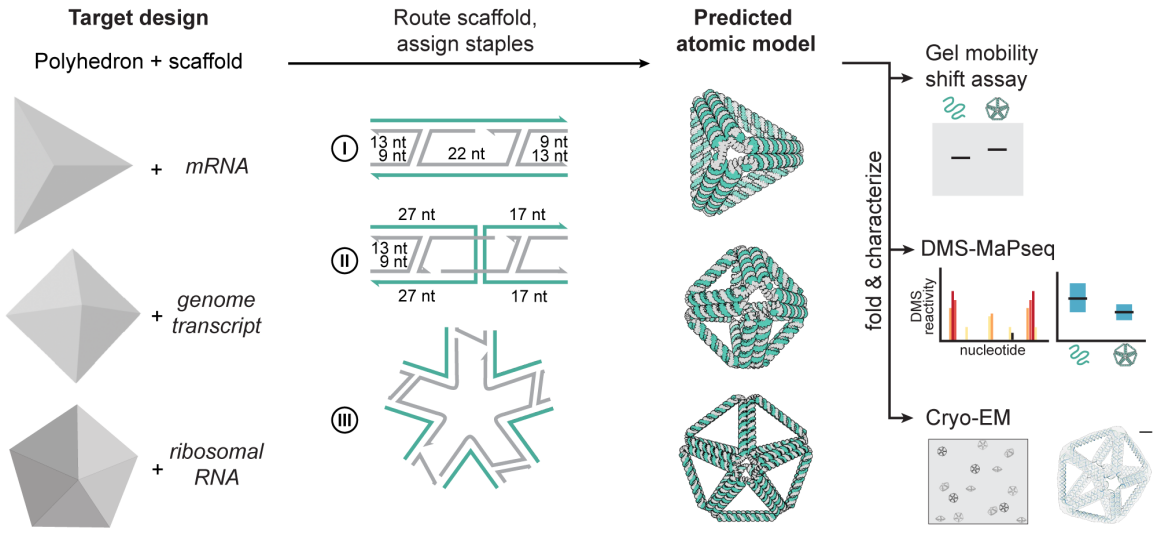

Figure 4-1: **Design overview for A-form DX wireframe origami.** Starting with a target polyhedron and scaffold sequence as inputs, we algorithmically route the scaffold through the polyhedron and route and assign staple for I. edges with no scaffold crossover, II. edges with a scaffold crossover, III. vertices, and predict an atomic model structure. The basic routing scheme for edges with the 4 helical turns is shown, scaffold in green and staples in grey. Using the calculated staple sequences, we can then fold the RNA scaffold into the target structure and characterize with gel mobility shift assays for preliminary folding evaluation, DMS-MaPseq or other RNA chemical probing method to evaluate base-pairing per nucleotide, and cryo-electron microscopy to evaluate overall structure formation.

a top-down staple design algorithm for input target geometries and target scaffold sequences, incorporating the aforementioned asymmetry in the staple crossover calculation (Figure 4-1). The design software has application to diverse alternative 3D wireframe geometries and sizes on the 5-50nm scale.

4.2 Results

4.2.1 Biochemical characterization to establish a folding protocol

Optimal folding conditions for origami typically vary based on type of design; for example, protocols differ between wireframe and bricklike, six- and two-helix-bundle, and single-stranded RNA- vs. DNA-scaffolded origami [55, 57, 140, 153]. We thus

first sought to establish the folding conditions for hybrid two-helix-bundle wireframe origami. Because the designs presented here are most similar to the DX wireframe designs of DAEDALUS, we based our folding protocol on the one reported in that paper [57], albeit with reduced time at higher temperatures to account for RNA instability. For a manual design of an RNA-scaffolded tetrahedron with six helical turns per edge, gel mobility shift assays of KCl and NaCl showed an upward shift of the major band, relative to unpaired scaffold, with the band position and tightness stabilized at 300 mM monovalent salt (Figures 4-2 and B-1). These results suggest the origami particles are fully formed in 300 mM monovalent salt and 10 mM HEPES-KOH pH 7.5 with our 13-hour folding ramp. No major band was observed when attempting to fold the tetrahedron in magnesium, likely due to RNA degradation from the divalent salt at the high temperatures during annealing (Figure B-2). This effect was mitigated by using previously published fast-folding protocols that used magnesium during folding [78] (Figure B-3). Likewise, higher yields of the RNA-scaffolded tetrahedra were achieved in HEPES-KOH pH 7.5 buffer than in Tris-HCl pH 8.1 buffer (Figure B-2), possibly due to the combined effect of higher pH and temperature. Dynamic light scattering (DLS) characterization of the tetrahedral origami folded in 300 mM KCl and 10 mM HEPES showed primarily monomeric populations with 33% polydispersity (Figures 4-2 and B-4). When folding other geometries in this solution, we again saw tight shifted bands in gel mobility shift assays, and a pentagonal bipyramidal geometry had a monodisperse population by DLS (Figures 4-2 and B-4). We thus determined that 300 mM KCl and 10 mM HEPES-KOH pH 7.5 was the optimal buffer for folding our hybrid RNA/DNA A-form DX wireframe origami with a 13-hour annealing ramp, and we used this folding method for all subsequent studies.

For the same initial EGFP mRNA-scaffolded tetrahedron with 66-bp edges, enzymatic degradation confirmed that the structure was scaffolded with RNA hybridized to DNA. Treatment with RNase A for 5 min resulted in no folded origami band on a gel and some apparent aggregation–possibly RNase A bound without cutting (Figure B-15). This interpretation is supported by the lack of release of DNA staple

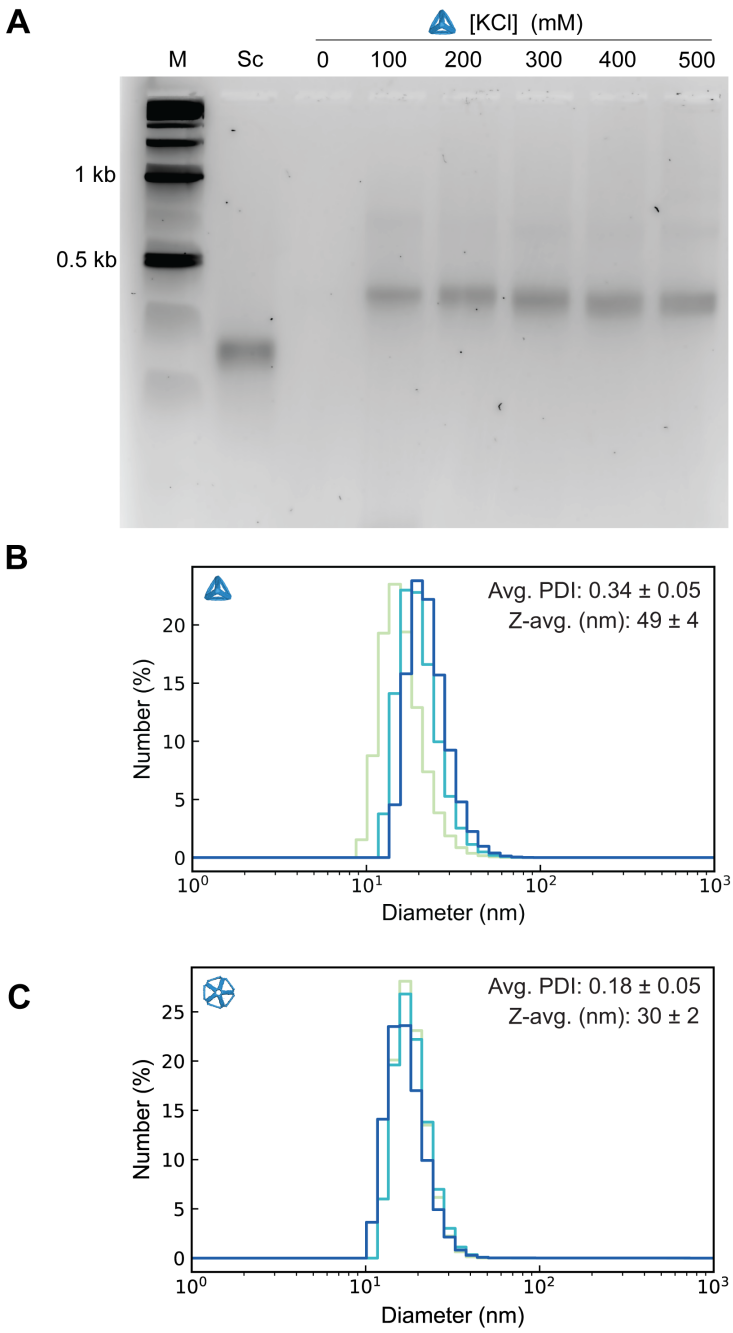

Figure 4-2: **Salt titration and initial characterization of folding. (A)** Gel mobility shift assay for an EGFP mRNA-scaffolded tetrahedron with 6 helical turns per edge (A-form design with asymmetry in scaffold crossover design – see Figure 4-5) (B) Dynamic light scattering histogram for the EGFP mRNA-scaffolded tetrahedron (scaffold crossover asymmetry design), with three measurement replicates shown in shades of blue-green. (C) Dynamic light scattering histogram for a 23s rRNA-scaffolded pentagonal bipyramid (scaffold crossover asymmetry design), with three measurement replicates shown in shades of blue-green.

strands, which should not be digested by RNase A. RNase H, on the other hand, specifically digests DNA-bound RNA, and completely degraded the folded origami and released staples within 5 minutes according to gel analysis. The folded origami band completely disappeared and was not replaced with a band corresponding to scaffold, suggesting full digestion of the scaffold, and bands corresponding to the short DNA staple strands appeared (Figure B-15). Had the folded origami in fact been formed with trace template DNA rather than the transcribed RNA, RNase H would not be able to degrade it. Interestingly, DNase I did not degrade the DNA staples within a 5 minute incubation, as the folded origami band remained intact according to gel electrophoresis (Figure B-15). This result suggests that the nanostructuring with RNA protects the DNA staples from a degree of enzymatic degradation.

4.2.2 Simple origami geometries folded with mRNA, M13 RNA and de Bruijn sequence

We examined the ability of three types of RNA sequences to scaffold A-form DX wireframe origami: mRNA, of interest in therapeutic delivery and vaccine applications; a de Bruijn sequence, designed to have minimal self-complementarity and repetition; and a transcript from the M13 viral genome, which is a sequence frequently used to scaffold DNA origami as a scalable source of single-stranded DNA. For an initial test with a simple polyhedron, we targeted a regular tetrahedral geometry, with six edges of equal length and four three-way vertices. We used in vitro-transcribed 792-nt prokaryotic EGFP mRNA and 660-nt and 924-nt de Bruijn RNA sequences to scaffold tetrahedra with six, five, and seven helical turns per edge, respectively (rT66, rT55, and rT77). To accommodate A-form helical geometry, the DX edges were designed with asymmetry in the staple crossover calculation, and 11 nt per helical turn, as illustrated in Figure 4-1.

The gel mobility shift assays in Figures 4-3A and B-6 shows that the tetrahedra formed discrete bands, shifted up slightly relative to the scaffold bands, suggesting a compactly folded particle. Because these objects were scaffolded with RNA, we

could apply dimethyl sulfate mutational profiling with sequencing (DMS-MaPseq)[18] to further characterize folding biochemically with single base resolution. With the nucleotide-level data from DMS-MaPseq, we could evaluate the degree of base pairing between scaffold and staples. We expected that DMS reactivities would be lower in well-folded regions of origami than in the RNA scaffold without staples, because the scaffold should be hybridized to staples and its nucleotides thereby protected from modification by DMS. Figure 4-3B plots the mean normalized DMS reactivity (see Methods) for each of the 64 double helical segments in the folded origami and each corresponding segment of nucleotides in the EGFP mRNA scaffold alone (Each data point corresponds to a nucleotide in the scaffold sequence. Whiskers indicate the furthest points within a distance 1.5 times that of the interquartile distance from the box limits, with outliers shown as dots). As expected, the median normalized DMS reactivity was 82% lower ($P = 2.7 \times 10^{-10}$, two-sided Wilcoxon signed-rank test) in the folded origami (median = 0.63%) compared to the scaffold folded without staples (median = 3.4%). We verified that reduced DMS reactivity was due to specific staple binding by refolding rT66 without one of its staples (B-5). The median DMS reactivity for the scaffold nucleotides targeted by the staple was 6.2-fold as high when the staple was omitted as when it was included ($P = 1.4 \times 10^{-4}$, two-sided Wilcoxon signed-rank test), while no such change was seen in the off-target nucleotides (fold-change = 1.08, $P = 0.76$), which suggests that the scaffold hybridized specifically to staples. For the rT55 and rT77, gel mobility shift assays also showed tight bands for the folded origami with a slight upward shift relative to the scaffold band, indicating compactly folded particles (Figure B-6). The median DMS reactivity of the double helical segments of each folded tetrahedron with a de Bruijn scaffold sequence was also lower than that of its scaffold folded without staples (79% lower for rT55, 70% lower for rT77), further suggesting that these scaffolds hybridized to staples as intended (Figure B-6).

Having determined biochemically that these RNA-scaffolded tetrahedra hybridized to staples and folded compactly, we next wanted to characterize the tertiary structure with cryo-electron microscopy (cryo-EM). Cryo-EM micrographs for the rT66 (Fig-

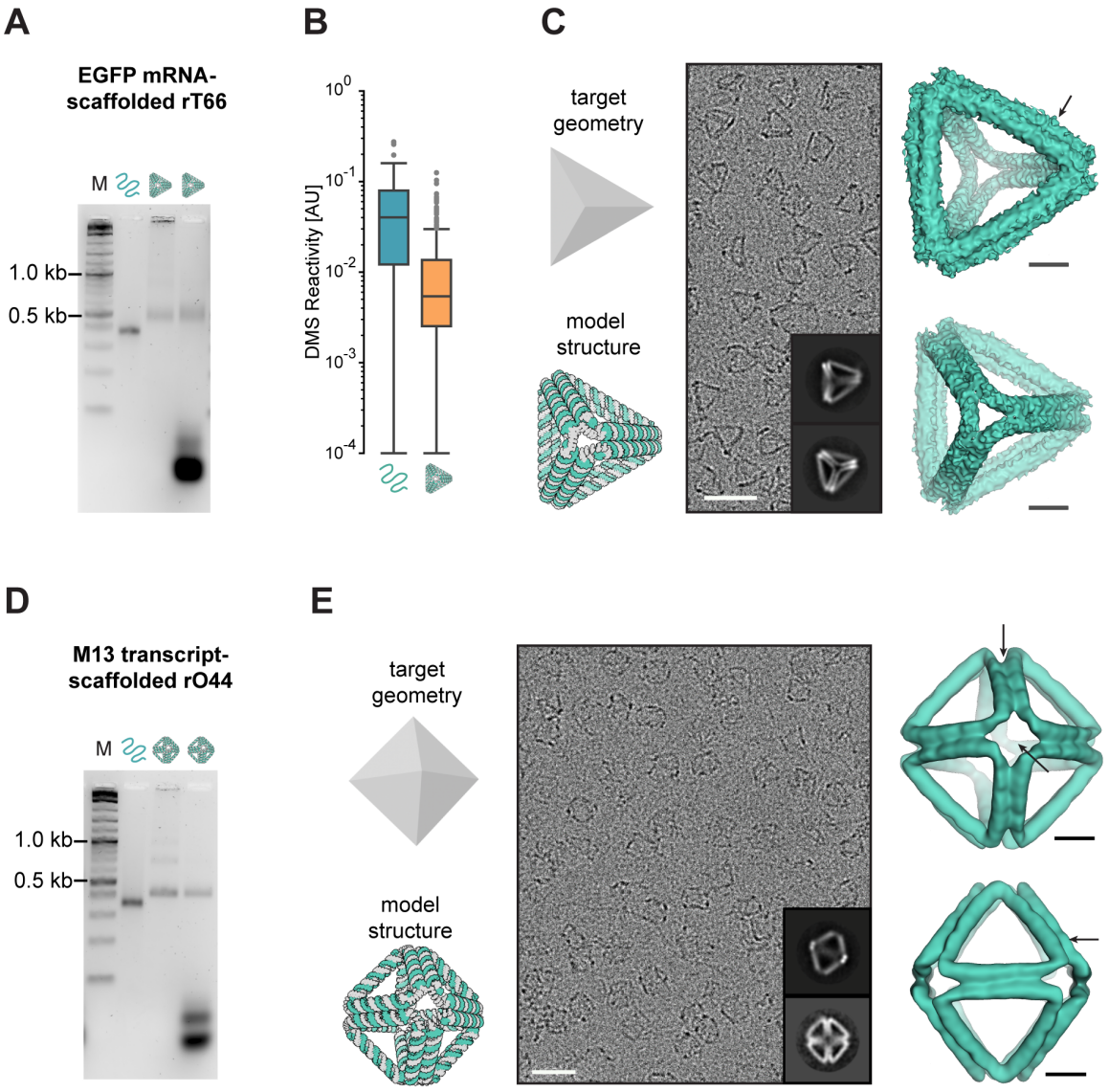

Figure 4-3: **Characterization of EGFP mRNA- and M13 RNA-scaffolded origami. (A)** Gel mobility shift assay, left to right: marker, unfolded EGFP mRNA scaffold, spin filter-purified folded rT66, and unpurified folded rT66. **(B)** Box plots of normalized DMS reactivity of segments, corresponding to how frequently nucleotides are unpaired, for the EGFP mRNA scaffold, folded without (left) or with (right) rT66 staples. **(C)** The input target geometry and predicted DX wireframe atomic model for the rT66, followed by an example micrograph (scale bar: 50 nm), two 2D class averages (insets), and two views of the reconstructed density map (scale bars: 5 nm). Arrow indicates modest edge bowing. **(D)** Gel mobility shift assay, left to right: marker, unfolded M13 transcript scaffold, spin filter-purified folded rO44, and unpurified folded rO44. **(E)** The input target geometry and predicted DX wireframe atomic model for the rO44, followed by an example micrograph (scale bar: 50 nm), two 2D class averages (insets), and two views of the reconstructed density map (scale bars: 5 nm). Arrows indicate evidence of twist between the two helices of each edge.

ure 4-3C) showed monodisperse tetrahedral particles, and the reconstruction at 12 Å resolution had slightly bowed edges, but matched the predicted atomic model with a correlation of 0.76.

Turning to a slightly more complex geometry, with twelve edges of equal length and six four-way vertices, we used a 1056-nt transcript of the M13 phage genome to scaffold a regular octahedron with four helical turns per edge (rO44). The M13 transcript-scaffolded rO44 likewise formed a discrete band with lower electrophoretic mobility than the scaffold in a gel mobility shift assay Figure 4-3D), and cryo-EM screening showed monodisperse octahedral particles (Figures 4-3E, B-7 and B-2). Reconstruction from the cryo-EM data achieved a resolution of 17 Å, and the map had a 0.90 correlation with the predicted atomic model. Unlike the longer-edged rT66, bowed edges were not evident in the reconstructed rO44.

Both the rT66 and the rO44 density maps had edge lengths corresponding to an average helical rise of 0.29 nm/bp (approximately 12.8 nm per edge for the rO44 and 18.9 nm per edge for the rT66, as measured in UCSF ChimeraX [154]), which is 11% larger than the canonical A-form rise of 0.26 nm/bp and 8% larger than the simulated energy-minimized A-form rise of 0.267 nm/bp [155]. Although still generally consistent with expectations for A-form helices, this increase in average rise may indicate that the helices are slightly underwound [155] or that the crossover junction geometry creates space that modestly lengthens the edge. For both the rT66 and rO44, we note that the two duplexes in an edge modestly twisted or sheared relative to one another (Figure 4-3C and E). This distortion may indicate an inability of the A-form twist to relax fully in these structures and is corroborated by molecular dynamics simulations [156].

4.2.3 23s-rRNA-scaffolded origami

Having validated folding of A-form DX wireframe origami with simple geometries and scaffolds, we moved to a larger RNA scaffold known to have significant native structure. Ribosomal RNA (rRNA) is the most abundant, by mass, type of RNA in mammalian cells [82], and its use as a scaffolding material for wireframe origami

holds promise for applications in studying RNA-mediated catalysis [51]. We therefore tested the ability of an in vitro-transcribed 1980-nt fragment of the E. Coli 23s rRNA to scaffold two different A-form DX wireframe origami objects, of varying complexity: a regular octahedron and a pentagonal bipyramid, each with six helical turns per edge (rO66 and rPB66, respectively), as well as a pentagonal bipyramid with five helical turns per edge (rPB55). Unlike the other geometries tested, the rPB66 has multiple vertex types, with both four-way and five-way vertices and corresponding variation in dihedral angles, making it a more complex target geometry in addition to using the longer, inherently structured scaffold. The gel mobility shift assays in Figures 4-4A and B and B-6 show that each folded object formed a discrete band shifted slightly upwards from the scaffold band, suggesting compactly folded objects.

As additional evidence of the proper folding of the 23s-scaffolded origami, the median normalized DMS reactivities among their double helical segments was lower than that of the 23s fragment scaffold folded without staples (72% lower for rO66, 71% lower for rPB66) (Figure 4-4C and D), indicating that the scaffolds hybridized to their staples in the folded origami. In further support of specific staple hybridization, the rO66 design only has staples hybridized to the first 1584 nucleotides of the scaffold, and we determined that the DMS reactivities in the excess scaffold region were well-correlated with those in the same region for the scaffold folded without staples (Pearson Correlation Coefficient = 1.0), rather than being depressed (Fig B-9).

With biochemical evidence of folding, we proceeded to use cryo-EM to characterize the tertiary structures of the rRNA-scaffolded origami. The cryo-EM micrographs for each particle showed well-folded, monodisperse particles (Figures 4-4E and F, B-10 and B-11). The rO66 reconstruction achieved 13 Å resolution and had a correlation of 0.85 with the predicted atomic model. The reconstruction of the rPB66 achieved 19 Å resolution, and the resulting density map fits with the predicted atomic model with a correlation of 0.92. In the cryo-EM reconstructions for both objects, we again observed a slight twisting or shearing of the two duplexes that make up an edge. The rO66 edges showed some outward bowing, which is not apparent in the rPB66 density map. The edges of the rO66 were approximately 17.9 nm long, as measured

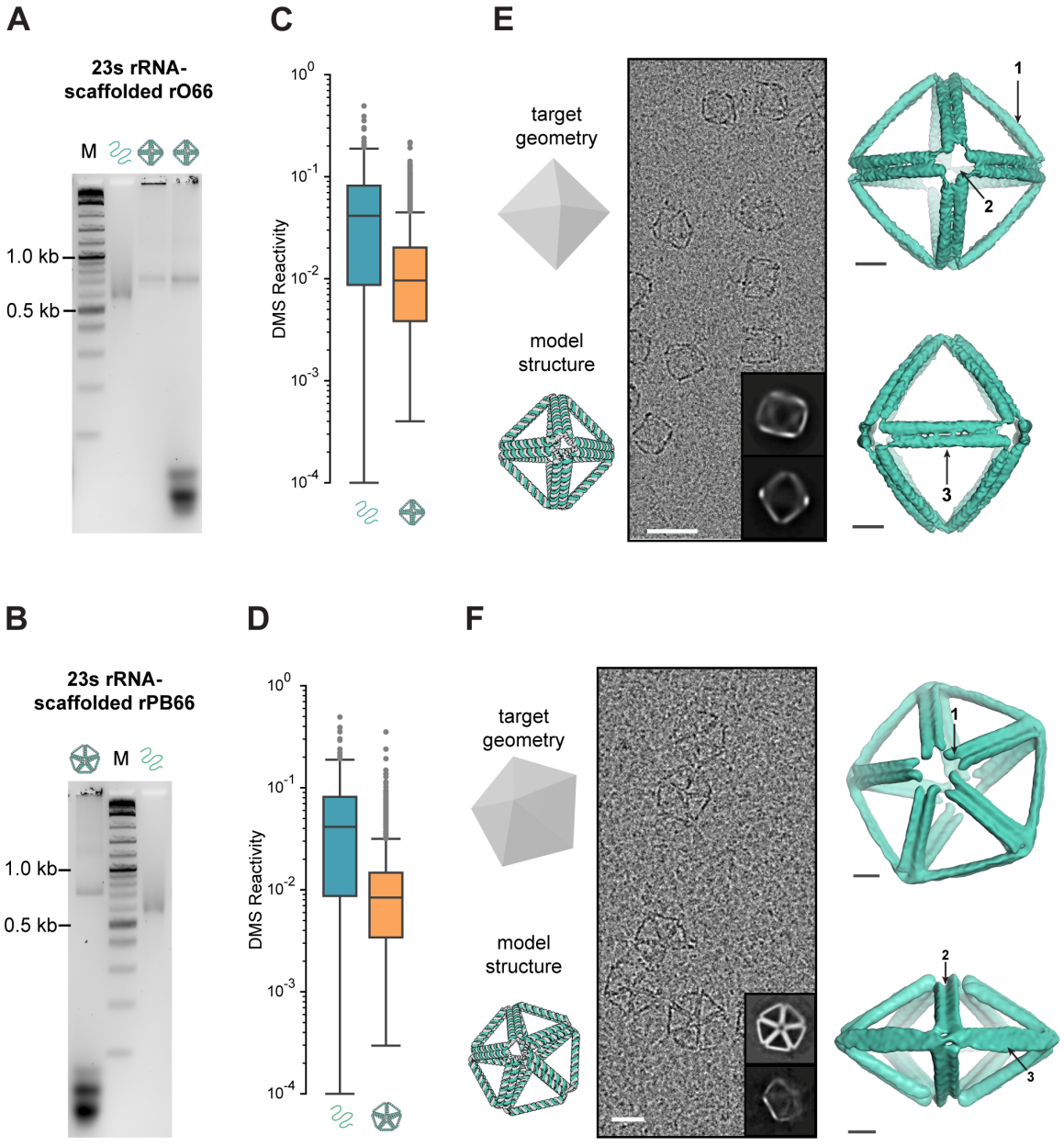
A
23s rRNA-scaffolded rO66
M
1.0 kb
0.5 kb
C
DMS Reactivity
E
target geometry
model structure
1
2
3
B
23s rRNA-scaffolded rPB66
M
1.0 kb
0.5 kb
D
DMS Reactivity
F
target geometry
model structure
1
2
3

Figure 4-4: **Characterization of 23s rRNA-scaffolded origami. (A)** Gel mobility shift assay, left to right: marker (1kb plus DNA ladder, NEB), unfolded 23s rRNA fragment scaffold, spin filter-purified folded rO66, and unpurified folded rO66. **(B)** Gel mobility shift assay, left to right: unpurified folded rPB66, marker (1kb plus DNA ladder, NEB), unfolded 23s rRNA fragment scaffold. **(C)** and **(D)** Box plots of normalized DMS Reactivity, corresponding to how frequently a nucleotide is unpaired, for the 23s rRNA fragment scaffold, folded without (left) or with (right) staples for the origami (rO66 in (C) and rPB66 in (D)). **(E)** and **(F)** The input target geometry and predicted DX wireframe atomic model for the rO66 (E) and rPB66 (F), followed by an example micrograph (scale bar: 50 nm), two 2D class averages (insets), and two views of the reconstructed density map (scale bars: 5 nm). Arrows in (E) indicate 1. Slight bowing in the DX edge, 2. Apparent twist in the vertex due to offset helical ends, and 3. Twist in the DX edge. Arrows in (F) indicate 1. The offset in the ends of helices due to the A-form pitch, 2-3. Indications of twist in the DX edges.

in UCSF ChimeraX [154], corresponding to an average rise of 0.271 nm/bp. The rPB66 has edges averaging approximately 17.8 nm long, corresponding to an average rise of 0.269 nm/bp. These values are consistent with the canonical A-form rise of 0.26 nm/bp and the simulated energy-minimized A-form rise of 0.267 nm/bp [155]. In both the rO66 and rPB66 reconstructions, we observed an offset in the helical ends of a DX edge at vertices, likely corresponding to the pitch of A-form helices and the asymmetry in staple crossover design.

4.2.4 Alternative routing designs tested

Besides the A-form design implemented for the origami described above, with 11 nt per helical turn and asymmetry in the staple crossover positions, we tested additional routing schemes, similar to those applied in prior hybrid RNA-DNA origami contexts [79], for the RNA-scaffolded DX wireframe origami: a B-form design, with no crossover asymmetry and 10.5 bp per helical turn; a “Hybrid"-form design with no crossover asymmetry and 11 nt per helical turn; and an alternative A-form design, with asymmetry incorporated into the scaffold crossover calculation and 11 nt per helical turn (Figure 4-5). The latter design maintains the asymmetrical spacing between adjacent scaffold and staple crossovers on neighboring helices that was suggested from models and implementations of A-form DX routing in RNA-scaffolded designs [80],

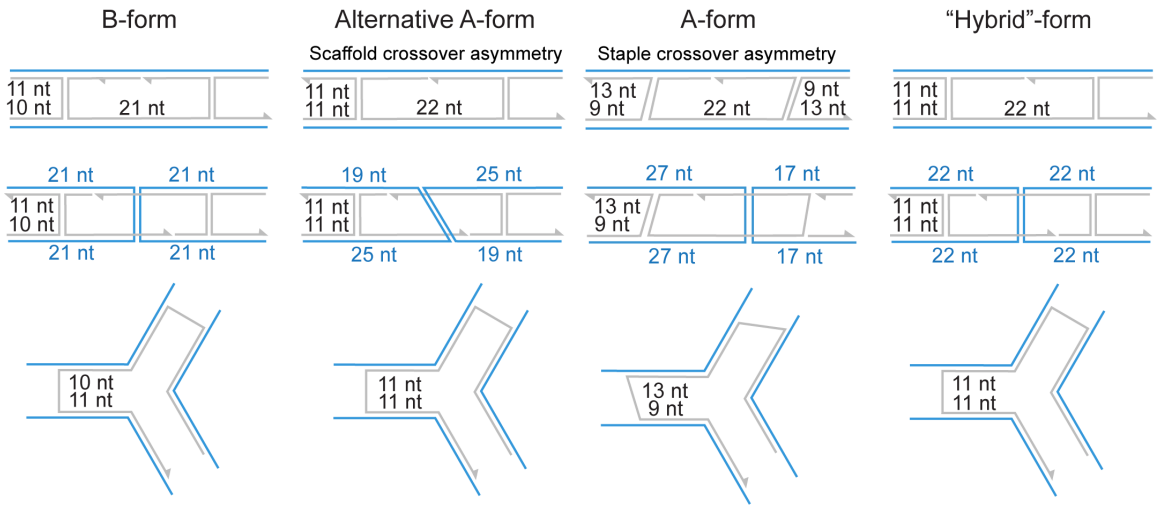

Figure 4-5: **Comparison of B-form (DAEDALUS [57]), A-form, and Hybrid-form scaffold and staple routings on edges and vertices.** Scaffold is in blue and staples are in grey. The edge diagrams represent edges 4 helical turns in length.

but the asymmetry is incorporated in scaffold instead of staple crossover positions.

The EGFP-mRNA-scaffolded Hybrid-form rT66 had high folding yield, with cryo-EM micrographs showing well-formed tetrahedral particles, and a reconstruction yielding a density map with 0.96 alignment correlation with the predicted model (Figure B-12). However, folding the EGFP mRNA scaffold using staples designed for a B-form fold showed a notably higher gel band shift, and cryo-EM micrographs did not show folded tetrahedron-shaped particles (Figure B-13).

The alternative A-form rT66 and the rPB66 designed with the scaffold crossover asymmetry behaved similarly to their staple asymmetry A-form designs above, with the same electrophoretic mobility and median DMS reactivity, and showing well-folded particles in cryo-EM imaging (Figures B-14 and 4-6). However, the alternative A-form rO44 and the rO66 had lower electrophoretic mobility than their staple crossover asymmetry counterparts, and showed no octahedral particles in cryo-EM screens (Figure 4-7). Further, the median DMS reactivity of the alternative A-form rO66 was only approximately 50% lower than the scaffold.

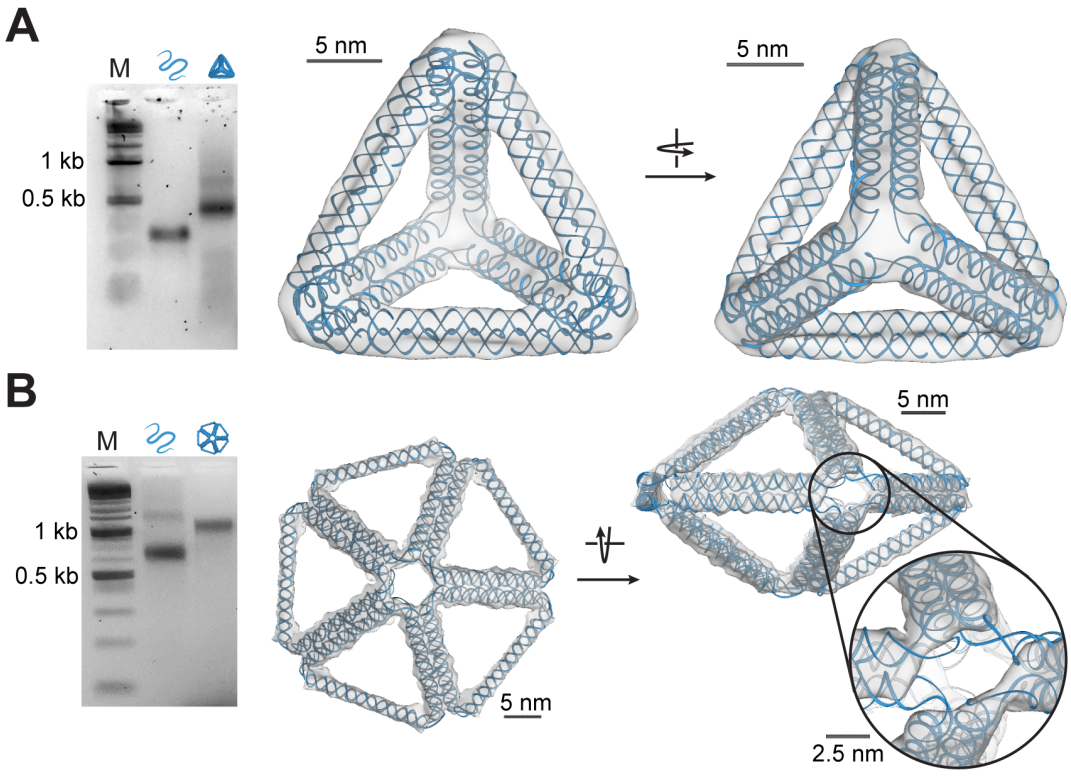

Figure 4-6: **3D structural characterization of RNA-scaffolded wireframe nanoparticles designed with scaffold crossover asymmetry (alternative A-form).** **(A)** A regular EGFP mRNA-scaffolded tetrahedron showing the distinct wireframe structure. **(B)** A regular 23s rRNA-scaffolded pentagonal bipyramid with 66-bp edge lengths showing the wireframe structure. A notable twist is seen along the edge, which disrupts the electron density at the vertices, as shown.

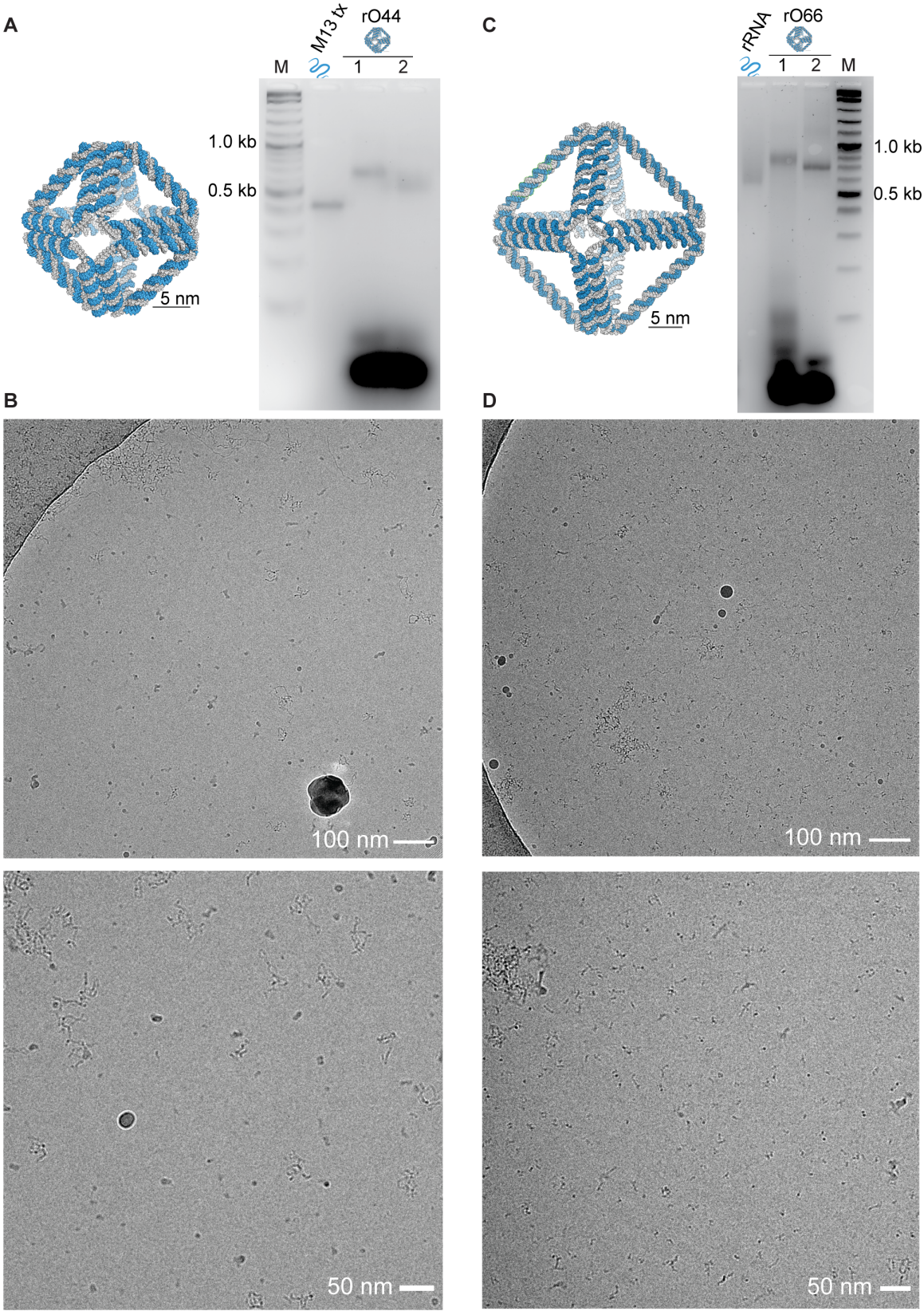
A
M
M13 tx
rO44
1
2
1.0 kb
0.5 kb
5 nm
C
rRNA
rO66
1
2
M
1.0 kb
0.5 kb
5 nm
B
100 nm
50 nm
D
100 nm
50 nm

Figure 4-7: **Characterization of alternative A-form (with scaffold crossover asymmetry) regular octahedra** with **(A)** four helical turns per edge or **(B)** six helical turns per edge, both scaffolded with M13 transcript RNA. The gel mobility shift assays show both the alternative A-form (**1** - scaffold crossover asymmetry) and the A-form (**2** - staple crossover asymmetry) designs, showing the greater shift, corresponding to a less compact object, for the alternative A-form. Representative micrographs show no well-folded octahedral particles for the alternative A-form design.

4.3 Concluding Discussion

To broaden application of nucleic acid origami to nanoscale materials, therapeutics, and structural biology, we introduced here a method for the synthesis of RNA-scaffolded origami using DNA staples. We have demonstrated folding of a variety of RNA scaffolds into several different DX wireframe geometries. Although for some geometries, like the tetrahedron, the RNA-scaffolded structures seem to be able to tolerate a range of different routing designs that use 11 nt per helical turn, for octahedra, a routing scheme that incorporated asymmetry into scaffold rather than staple crossovers resulted in a failure to fold as visualized with cryo-EM imaging. The A-form design with asymmetric staple crossovers thus is robust for a wider variety of geometries than other routing schemes.

In addition to the need to accommodate A-form helical geometry, one concern with using long RNA strands to scaffold origami is its tendency to be highly structured internally. However, the presence of native secondary structure in the target RNA scaffold does not inherently prevent folding; the 23s rRNA fragment used to scaffold two objects has a variety of secondary structural motifs, and is approximately 58% base paired on its own, but after denaturing at high temperatures and re-annealing, the RNA scaffold bound preferentially to the staples and folded the target structure with high yield.

Scaffolding wireframe origami with RNA allowed us to study nucleic acid origami folding and stability with nucleotide-level precision. To our knowledge, the DMS-MaPseq profiling presented in this work is the first application of chemical probing to large-scale scaffolded origami, made possible by the use of RNA scaffolds. We can

apply this protocol to investigate base pairing stability for improved sequence design. The approach also holds promise for kinetic studies of nucleic acid origami folding; probing samples at various stages of the annealing ramp and comparing what sections of the scaffold are bound at each stage might improve mechanistic understanding of scaffolded origami folding.

Beyond studies of nucleic acid origami itself, the use of RNA as scaffold for DX wireframe origami enables a variety of applications, depending on the particular scaffold used. RNA has a number of useful features distinct from standards DNA, including the ability to modify nucleotides for stability and translatability and to introduce riboswitches and aptamers, ribozymes, antisense oligos, and long RNAs. In particular, here we used mRNA as a scaffold for the assembly of a nanoparticle containing the sequence encoding a fluorescent protein. Delivery of such a nanoparticle to a cell could offer important potential for nuclease-specific release of scaffold or staples within the cell, with applications in antisense oligonucleotide (ASO) therapy [157, 158], multiplex automated genome engineering (MAGE) [159], and homologous recombination template [160] delivery. The tunable degradation rates introduced by the combined use of RNA, modified RNA and DNA may also prove useful for material templating and etching. We additionally used ribosomal RNA to scaffold a pentagonal bipyramid that would leave domains V and VI of the 23S rRNA to fold freely. Future engineering will allow for the generation of synthetic nucleic acid assemblies that can coordinate catalytic ribozymes [60, 161], test modification enzyme substrates[162, 163], and develop novel ribosomes and translation systems [164–166].

4.4 Methods

Reagents

Oligonucleotide staples and primers and gBlock synthetic DNA sequences were purchased from Integrated DNA Technologies (IDT, Coralville, IA). HEPES, Trizma base, EDTA, NaCl, KCl, $MgCl_2$, magnesium acetate, and high-resolution agarose were purchased from Sigma-Millipore (MA). HiScribe T7 RNA polymerase kits and

Q5 2x HiFidelity PCR mastermixes were purchased from New England Biolabs (NEB, Ipswitch, MA). Sodium cacodylate solution, pH 7.2, was purchased from VWR.

A-form DX wireframe origami programmed using pyDAEDALUS

To generalize to A-form helical geometries to allow for RNA/RNA or hybrid RNA/DNA DX-based origami, edge lengths were discretized to multiples of 11 rather than rounded multiples of 10.5, and crossover positions were changed to be compatible with A-form helix crossovers. As described previously [136], scaffold crossover edges, which have adjacent crossovers occurring on different strands (scaffold vs. staple) and thus must occur an odd number of half-twists apart, require that the crossovers on the two helices of the edge be spaced asymmetrically to be compatible with A-form helical geometry. We investigated two approaches to implementing this asymmetry, in addition to testing a design with no asymmetry incorporated (Hybrid-form).

In one approach to A-form, the required asymmetry is incorporated into the staple crossover position calculation (Figures 4-1 and 4-5). In this case, staple crossovers are asymmetric across the two helices, with a 4-nt difference between the nucleotide position on the two helices (e.g., the first staple crossover occurs 9 nt from the vertex on the 5' side and 13 nt from the vertex on the 3' side).

An alternative approach incorporates asymmetry into the scaffold crossover position calculation for A-form (Figure 4-3). In this approach, the scaffold crossover had a 5-nt difference between the nucleotide position on the two helices (e.g. the scaffold crossover on a 33-nt edge would occur 14 nt from the vertex on the 5' side and 19 nt from the vertex on the 3' side). Manual modifications from B-form designs to the alternative A-form were initially implemented using Tiamat software66 and subsequently automated. For the tetrahedron and pentagonal bipyramid, a "Hybrid-form" design was also manually generated, which does not have the 5-nt difference in nucleotide position between the two helices, but instead directly crosses over as in the standard DNA design (Figures 4-5 and B-12).

We implemented the A-form design rules with staple crossover asymmetry in a top-down design algorithm that calculates sequences for folding an input target

shape with wireframe DX edges (Figure 1). The code architecture in pyDAEDALUS (http://github.com/lcbb/pyDAEDALUSX) mirrors the format and naming conventions of DAEDALUS [57]. Briefly, the input target geometry file in the Polygon File Format (PLY) is parsed to identify relevant geometric parameters including coordinates of vertices, edge and face connectivities that form the graph of the shape, and edge lengths. Scaffold routing is achieved by calculating the spanning tree of the graph, and staples are added according to either standard geometric rules for B-form DNA or the A-form design rules, depending on user specification. The resulting outputs of the algorithm are plaintext and Comma-Separated Values (CSV) files that store the routing information, staple sequences, and nucleotide spatial coordinates. The positions and orientations of each nucleotide are represented as vectors following the convention from the software 3DNA [167].

While the overall architecture of DAEDALUS is preserved in pyDAEDALUS, several fundamental changes were required. First, the connectivities passed through the functions are stored in pyDAEDALUS as NetworkX Graph objects, rather than sparse matrices as in DAEDALUS. Second, Prim's algorithm [168], which is used to generate the spanning trees required to route the scaffold strand, was used in both algorithms as built-in functions. However, for many structures the Python version generates a spanning tree different from the MATLAB version. Although this will affect the scaffold routing and staple sequences, the fidelity of the final design should not be affected, because each possible spanning tree of an object corresponds to a valid scaffold routing [57]. Third, in order to exploit the object-oriented structure that Python enables, the DNAInfo class was introduced, which packages together the many variables associated with the geometry, routing, and structure generated in intermediate sub-functions of the algorithm.

To render the code more robust and offer a platform for further development by other contributors, additional frameworks were constructed. A style guide was implemented to help readability of the code, and linting, i.e. automatic checking of adherence to the style guide, is also enforced. In addition, unit tests were introduced to ensure that the functionality of the code is preserved as intended by the original

authors.

RNA transcription

The full sequence of each DNA template is listed in Appendix D. For the RNA-scaffolded tetrahedron, the EGFP sequence was generated as a gBlock and cloned with a T7 promoter and Shine-Dalgarno (SD) sequence 5' of the coding sequence into a pUC19 vector using restriction cloning (EcoRI, PstI). RNA was transcribed from a Phusion PCR-generated double-stranded DNA (dsDNA) template containing a 5' T7 promoter, amplified using primers listed in Appendix D, and gel purified. For the pentagonal bipyramid and octahedron with 6 helical turns per edge, primers were chosen flanking Domains I–IV of the rrlB gene encoding the 23S rRNA from the pCW1 plasmid [169]. For the octahedron with 4 helical turns per edge, partial M13 DNA template was amplified from mp18 ssDNA (NEB). For the fragment of human immunodeficiency virus 1 (HIV-1) Rev response element (RRE) used as a control in DMS-MaPseq experiments, the sequence was synthesized as a gBlock (IDT) containing a 5' T7 promoter, then PCR amplified using the Q5 High-Fidelity 2X Master Mix (NEB). For the tetrahedra with 5 and 7 helical turns per edge, we designed a random scaffold sequence with minimal self-complementarity (rsc1218v1) and again obtained the sequence as a synthetic gBlock (IDT), then amplified with Q5 High-Fidelity 2x Master Mix (NEB) to create dsDNA templates for 660 nt and 924 nt scaffolds, each with a 5' T7 promoter.

Using these dsDNA templates, RNA was transcribed using the manufacturers protocol for HiScribe T7 (NEB) for canonical base RNAs or DuraScribe T7 (Lucigen) for 2'-fluoro-modified base RNAs. RNA was treated with DNase I (NEB), then pre-cleaned on a ZymoClean RNA cleanup kit (RNA Clean-and-concentrator 5). Urea polyacrylamide gel (PAGE) was used to validate purity, and PAGE or HPLC was used to purify the RNAs if byproducts were present. With RNA pre-cleaned using the RNA clean-and-concentrator-5 kit (Zymo), and denatured by addition of 1x RNA loading dye (NEB) and 5-10-minute incubation at 70^{o}C, PAGE purification was performed on a 6% gel containing 8M urea. RNA was sliced from the gel after visualization with

SybrSafe (ThermoFisher) and eluted in 300 mM sodium acetate pH 5.2, precipitated in 70% ethanol at -20ºC for >2 hours, and then pelleted at 14,000 RPM for 30 minutes at 4ºC.

For HPLC purification, transcribed and column-purified (with ZymoClean RNA clean-and-concentrator-5 kit) RNA was diluted with nuclease-free water and injected into an XBridge Oligonucleotide BEH C18 column (130 Å, 2.5 μm, 4.6 mm x 50 mm, Waters) under the following gradient, flowing at 0.9 ml/min: increasing from 38-40% solvent B over 1 minute, increasing to 60% buffer B across 15 minutes, increasing to 66% buffer B across 6 minutes, increasing to 70% buffer B across 30 seconds, reaching 100% buffer B across 30 seconds, maintaining 100% buffer B for 1 minute, decreasing to 38% solvent B over 1 minute, where it was finally held for 2 minutes (adapted from previously-published protocol70). Buffer A was a solution containing 0.1 M TEAA, while buffer B included 0.1 M TEAA and 25% (v/v) acetonitrile. All HPLC purification of the RNA scaffold was run at 65ºC to prevent formation of secondary structure. Sodium acetate, pH 5.2, was added to a final concentration of 300 mM in the collected fraction, and the RNA was precipitated in 70% ethanol at -20ºC for >2 hours, then pelleted at 14,000 RPM for 30 minutes at 4ºC.

Hybrid RNA/DNA nanoparticle folding and characterization

Using RNase-free buffers and conditions, 20 nM of purified scaffold was mixed with 400 nM individual staples and buffer and salt and brought to 50 μl aliquots for temperature ramping. 10 mM and 50 mM HEPES-KOH pH 7.5 and 50 mM Tris-HCl pH 8.1 were tested, and salt concentrations were tested in 10 mM HEPES-KOH such that the final concentrations of KCl and NaCl individually were 0, 100, 200, 300, 400, and 500 mM, and for $MgCl_2$: 0, 2, 4, 8, 12, and 16 mM. Folding was performed using a modification of the previously published wireframe origami thermal annealing protocol14 but with reduced incubation time at high temperatures. Briefly, the folding protocol was 90ºC for 45 s; ramp 85ºC to 70ºC at 45s/ºC; ramp 70ºC to 29ºC at 15 m/ºC; ramp 29ºC to 25ºC at 10 m/ºC; 10 m at 37ºC; hold at 4ºC until purification. Folded particles were purified away from excess staples using Amicon Ultra 100 kDa

0.5 ml filter columns and buffer exchanged into the same buffer used for folding.

The size distribution of the origami nanoparticles was measured via DLS using a Zetasizer Nano ZSP (model ZEN5600, Malvern Instruments, UK). Purified nanoparticles were concentrated to 75 nM in 50 μl in 10 mM HEPES-KOH pH 7.5 and 300 mM KCl. The default procedure for DNA was used, only customizing the buffer to include 300 mM KCl. Three serial DLS measurements were performed on the same folded sample at 25^{o}C. The average nanoparticle diameter (nm) and polydispersity index (PdI) were computed using the associated Malvern software (Zetasizer Software v 7.12).

Biochemical stability in the presence of RNases A and H was also tested. 50 nM RNA-scaffolded tetrahedron with 66-bp edge length was incubated for 5 min at 37^{o}C in the presence of buffer alone, 25 units of RNase H or 3.5 units of RNase A. Reactions were quenched at 4^{o}C and run at 65V for 180 minutes on a high-resolution 2.5% agarose gel in TBE with 2.5 mM $Mg(OAc)_2$, maintained at 4^{o}C on ice.

Chemical probing of secondary structure with DMS-MaPseq

The RNA for each scaffold and the control 232-nt HIV-1 RRE fragment was produced as described above, except that they were purified with the following steps:

Polyacrylamide gels with 6 M urea (2.4 ml 5X Tris-Borate-EDTA (TBE), 4.32 g urea, 1.2 ml 40% 19:1 acrylamide:bis-acrylamide, 120 μl ammonium persulfate, 12 μl TEMED, nuclease-free water to 12 ml) were pre-run at 160 volts for 30 min. The RNAs were denatured in 2X RNA Loading Dye (New England BioLabs) at 70°C for 10 min, immediately placed on ice, and run on the gels at 160 volts for 60 min in 1X TBE in a Mini-PROTEAN Tetra Cell (Bio-Rad). The gels were stained in 1X TBE containing 1X SYBR Safe (Thermo Fisher) for 5 min, and each band of expected molecular weight was excised and transferred to a 0.5 ml tube at the bottom of which a hole had been punctured with a needle. Each 0.5 ml tube was placed in a 1.5 ml tube and spun at 16,000×g for 60 sec to extrude the gel slice into the 1.5 ml tube. Each gel slice was covered with 400 ul of gel elution buffer (250 mM sodium acetate pH 5.2, 20 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 0.25% w/v sodium dodecyl

sulfate) and incubated in a thermomixer at 20°C for 11 hr while shaking at 500 rpm. The slurries were decanted into Costar Spin-X Centrifuge Tube Filters (Corning) and spun at 16,000×g for 1 min to remove gel particles. To each filtrate, 1 ml 100% ethanol was added, and the tubes were frozen at -80°C for 1 hr. The tubes were then spun at 12,700×g for 1 hr at 4°C. The pellets were washed with 500 μl 75% ethanol at -20°C and spun for another 10 min. The supernatants were removed and the tubes uncapped and placed on a 37°C heat block to dry the pellets for 10-20 min. The pellets containing the RNA were resuspended in 10 μl nuclease-free water.

The gel-purified RNA scaffolds were used to fold nanoparticles in folding buffer (10 mM HEPES-KOH pH 7.5, 300 mM KCl) using 20 nM scaffold and 400 nM for each staple with the temperature steps described above. Each RNA scaffold was also folded using the same protocol but without adding staples. The 23S scaffold was folded in two tubes each containing 85 μl; rT66 without staple 10 was folded in one tube containing 70 μl; all other nanoparticles and scaffolds were folded in two tubes each containing 60 μl.

Folded nanoparticles/scaffolds were purified by five rounds of filtration through Amicon Ultra 100 kDa 0.5 ml filter columns. One Amicon filter for each of the 16 nanoparticle/scaffold samples was first spun at 2,400×g for 30 min at 4°C with 500 μl of the buffer used to fold the origami. To each pre-spun filter, 350 μl of 300 mM sodium cacodylate pH 7.2 (Electron Microscopy Sciences) was added, followed by 50 – 150 μl of the pooled folding product of one nanoparticle/scaffold. The samples were spun at 850×g for 30 min at 4°C, after which the filtrate was decanted and 450 μl sodium cacodylate added to the filter, and these steps were repeated for a total of five filtrations. The fifth filtration was run for 50 min, after which each filter (containing approximately 50 μl) was inverted into a clean collection tube and spun at 1,500×g for 1 min at 4°C to collect the sample of nanoparticles/scaffold. A 10 μl aliquot of each sample was transferred to a 1.5 ml tube.

As a control for normalization of the DMS reactivities, 1.3 μg of gel-purified RRE RNA was denatured in 8 μl of RNase-free water at 95^{o}C for 60 seconds and immediately placed on ice for 60 seconds. The denatured RRE RNA was mixed

with 612 μl of 300 mM sodium cacodylate pH 7.2 (Electron Microscopy Sciences) and incubated at 37°C for 20 min to refold its structure. A 38.5 μl aliquot of refolded RRE RNA was added into each of 16 tubes containing 10 μl of one nanoparticle/scaffold. To each sample, 1.5 μl of neat dimethyl sulfate (DMS) was added (50 μl total volume, 3% DMS v/v), stirred with a pipette tip, and incubated at 37°C for 5 min in a thermomixer while shaking at 500 rpm. Each reaction was quenched by adding 30 μl neat beta-mercaptoethanol (MilliporeSigma). DMS-modified nucleic acids were purified using a Zymo RNA Clean and Concentrator-5 Kit (Zymo Research) and eluted in 10 μl RNase-free water.

For each RNA sample, 4 μl was reverse transcribed in a 20 μl reaction containing 1 μl pooled reverse primers (10 μM each), 1 μl TGIRT-III enzyme (Ingex), 4 μl 5X First Strand buffer (Invitrogen), 1 μl 10 mM dNTPs (Promega), 1 μl 0.1 M dithiothreitol (Invitrogen), and 1 μl RNaseOUT (Invitrogen). The reactions were incubated at 57°C in a thermocycler with the lid set to 60°C for 90 min. The RNA templates were degraded by adding 1 μl of 4.0 M sodium hydroxide to each reaction and incubating at 95°C for 3 min. Each cDNA was purified using a Zymo Oligo Clean and Concentrator-5 Kit (Zymo Research) and eluted in 10 μl nuclease-free water.

The cDNA from each of the 16 samples was amplified as a set of overlapping amplicons, each 250 – 556 bp (47 amplicons total), plus one amplicon spanning the entire RRE (16 amplicons total). For each amplicon, 1 μl purified cDNA was amplified with an Advantage HF 2 PCR kit (Takara) in a 25 μl reaction containing 0.5 μl forward primer (10 μM, IDT), 0.5 μl reverse primer (10 μM, IDT), 0.5 μl 50x Advantage-HF 2 Polymerase Mix, 2.5 μl 10x Advantage 2 PCR Buffer, 2.5 μl 10x HF dNTP Mix, and 17.5 μl nuclease-free water. The PCR entailed an initial denaturation step at 94^{o}C for 60 sec, followed by 25 cycles of 94^{o}C for 30 sec, 60^{o}C for 30 sec, and 68^{o}C for 60 sec, with a final extension at 68^{o}C for 60 sec. All PCR products were validated using E-Gel EX-Gels with 2% Agarose (Thermo Fisher).

All 47 PCR products from nanoparticles/scaffolds and 5 RRE products were consolidated into 5 pools such that no two amplicons from the same RNA sequence were pooled together. Pools 1 – 4 contained 6 μl each of 10 PCR products; pool 5 con-

tained 5 μl each of 12 PCR products. For each pool, 30 μl was mixed with 6 μl 6X gel loading dye and run on a 50 ml gel containing 2% SeaKem Agarose (Lonza), 1x Tris-Acetate-EDTA (Boston BioProducts), and 5 μl 10,000X SYBR Safe DNA Gel Stain (Thermo Fisher) at 60 volts for 105 min. Bands at the expected sizes were excised and the DNA extracted using a Zymoclean Gel DNA Recovery Kit (Zymo Research) and eluted in 12 μl 10 mM Tris pH 8.0. DNA libraries were generated and sequenced on an Illumina MiSeq using a 300 $\times$ 300 read length at the MIT BioMicroCenter sequencing core.

Statistical analysis of DMS reactivities and structural features

DMS-induced mutation rates ("DMS reactivities") were determined using the Detection of RNA folding Ensembles with Expectation Maximization clustering (DREEM) pipeline [25], using the default parameters except for a 90% coverage threshold for clustering. In order to control for variations in DMS treatment among different samples, the DMS reactivities were normalized using a custom script as follows. The median DMS reactivity among the top 50% (n = 46) of the 91 adenine (A) and cytosine (C) bases in the spiked-in RRE control was computed for each sample. For each other sample, the ratio of the median DMS reactivity of the RRE to the median DMS reactivity of the RRE in the sample of 23S scaffold without staples (the reference sample) was computed, and the DMS reactivity of the sample was divided by this ratio to normalize it.

Cryo-electron microscopy

Three microliters of the folded and purified RNA nanostructure solution (approximately 600 nM) was applied onto the glow-discharged 200-mesh Quantifoil 2/1 grid, blotted for four seconds and rapidly frozen in liquid ethane using a Vitrobot Mark IV (Thermo Fisher Scientific). All grids were screened and imaged on a Talos Arctica cryo-electron microscope (Thermo Fisher Scientific) operated at 200 kV at a magnification of 79,000$\times$ (corresponding to a calibrated sampling of 1.76 Å per pixel). Micrographs were recorded by EPU software (Thermo Fisher Scientific) with a Gatan

Table 4.1: Particle numbers and symmetry numbered used to reconstruct RNA-scaffolded origami.

Object	**Helical geometry design**	**# Particles used in reconstruction**	**Symmetry used**
rT66	alternate A-form (scaffold xover asymmetry)	819	Tetrahedral
rPB66	alternate A-form (scaffold xover asymmetry)	7073	D5
rT66	Hybrid-form	2171	Tetrahedral
rT66	A-form (staple xover asymmetry)	∼42,000	Tetrahedral
rPB66	A-form (staple xover asymmetry)	14,402	D5
rO44	A-form (staple xover asymmetry)	∼22,000	Octahedral
rO66	A-form (staple xover asymmetry)	38,580	Octahedral

K2 Summit direct electron detector in counting mode, where each image is composed of 24 individual frames with an exposure time of 6 s and a total dose 63 electrons per Å^2. We used a defocus range of -1.5 – -3 μm to collect images, which were subsequently motion-corrected using MotionCor2 [116]. Single-particle image processing and 3D reconstruction was performed as previously described14 using the image processing software package EMAN2 [117]. All particles were picked manually by e2boxer.py in EMAN2. The total number of particles used for final refinement are listed in Table 4.1. Resolution for the final maps were estimated using the 0.143 criterion of the Fourier shell correlation (FSC) curve without any mask. A Gaussian low-pass filter was applied to the final 3D maps displayed in the UCSF Chimera software package [112]. Correlation of each map with its corresponding atomic model is calculated by the UCSF Chimera fitmap function, with density simulated from the model at the same resolution as the corresponding reconstruction.

Acknowledgements

This work was supported by the National Science Foundation (NSF-EAGER) CCF-1547999, the Office of Naval Research (N000141612953, N000141210621, N00014161295,

N000141612506), a NDSEG fellowship to MFP, a NSF GRFP to MFA, the National Institute of Environmental Health Sciences (P30-ES002109) and the National Institutes of Health (1-R21-EB026008-01).

We thank Tammy C. T. Lan for assistance with DMS-MaPseq, and Krista Pullen for help with salt titration triplicates and DLS characterization.

Molecular graphics and analyses performed with UCSF Chimera and UCSF ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311, NIH R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases.

Some cryo-EM specimens were prepared and imaged at the Automated Cryogenic Electron Microscopy Facility in MIT.nano on a Talos Arctica microscope, which was a gift from the Arnold and Mabel Beckman Foundation.

Chapter 5

Anchoring RNA fragments on RNA-scaffolded origami

5.1 Introduction

5.1.1 Approach overview

Although of great interest for drug design and nanotechnology [170–173], tertiary structure determination remains an elusive goal for a wide variety of RNAs that do not crystallize and whose size and/or heterogeneity frustrates analysis by cryo-electron microscopy (cryo-EM) or nuclear magnetic resonance. Successful RNA structure determination with cryo-EM most often occurs when the RNA is stabilized through interactions with coordinating proteins, such as in the ribosome [174, 175]. For other RNA structure targets of interest, we hypothesized that anchoring the RNA on nucleic acid origami would provide sufficient stability and improvement in particle selection to enable 3D structural characterization.

Anchoring the RNA on nucleic acid origami could be accomplished via RNA ligation to staple ends, but ligation is often inefficient. Further, the structure could largely fold–and therefore be selected from cryo-EM micrographs–even if the staple does not incorporate (consider the missing staple control discussed in section 4.2.2). We can instead directly incorporate RNA fragment sequence within an RNA scaffold,

i.e. using long tails on either side of the target RNA fragment to hybridize to staples and form a nanostructure (Figure 1-1). This approach is more straightforward to implement in the laboratory, and has the major advantage that every folded particle selected in cryo-EM micrographs will be guaranteed to have the target RNA fragment anchored to it. With this ongoing work, we aimed to elucidate some best practices for designing anchored RNA objects for structural characterization.

5.1.2 Proof-of-concept target: HIV-1 Rev Response Element

We selected the HIV-1 Rev Response Element (RRE) as an initial target RNA fragment to test the anchoring approach. The RRE is a long stem-loop feature in the HIV genome to which the protein Rev binds and oligomerizes [34, 176–178]. This is a key step in the viral life cycle, as Rev binding enables the unspliced viral genome and incompletely spliced viral mRNA to leave the nucleus (where they were transcribed) for translation and packaging in the cytoplasm of the infected host cell [176]. The 232-nt branched head group of the RRE has two secondary structure conformations, as determined by in-gel SHAPE and in-cell DMS-MaPseq, which have been determined to affect functional activity and viral fitness [179, 180]. Small angle X-ray scattering (SAXS) provided a 21Å-resolution envelope for the 3D structure of the RRE headgroup, although the study only fit one of the two documented secondary structures into the envelope to create a tertiary structure model.

As with the tRNA in Chapter 3, it is useful to have existing structure data for our target RNA in proof-of-concept experiments, so we can compare structural observations. In particular, with cellular DMS-MaPseq data, we can see if the secondary structures adopted by an RRE anchored on nucleic acid origami align with the secondary structures in the full genomic and cellular context. With the RRE, there is opportunity to advance existing structural knowledge as well, as the highest resolution 3D structure is 21Å [179] (whereas RNA-scaffolded origami structures have reached 11Å resolution [section 4.2.2]), and does not account for the heterogeneous structural ensemble the RRE has been shown to adopt.

5.2 Design considerations

Several design factors could influence the reconstruction of an anchored RNA on nucleic acid origami. First, the choice of origami to which to anchor the RNA could affect the resolution limit, particularly for origami with highly flexible edges. Shorter edge lengths and more rigid geometries are thus preferable. However, if the anchored RNA protrudes to the interior of the origami, the structure must be large enough to contain the anchored RNA. The type of geometry could further have an impact on the success of reconstruction; an overall symmetric geometry could result in unintentionally averaging out the anchored RNA, when particle images are mistakenly misaligned with respect to the anchored RNA but appear well-aligned and averaged due to the symmetry. Given the ability to uniquely align the tRNA in the symmetric tetrahedron in Chapter 3, however, we hypothesized that asymmetry in the origami may not be crucial to reconstruction. In other words, we expected the anchored RNA would be enough of an asymmetric feature in itself that the attached origami could be otherwise symmetrical.

Another parameter to consider in designing anchored RNA origami is the location and direction of the anchoring. We hypothesized that fragment placement may affect the ability of the object and anchored fragment to fold, as well as our ability to image and reconstruct the fragment structure. For example, inserting the target RNA sequence within scaffold such that it points inwards into the wireframe origami may cause steric hindrance that inhibits the fold. Or if it does fold, the presence of origami edges in front of and behind the RNA fragment might impact the resolution of data we can record with cryo-EM. On the other hand, allowing the fragment to fold pointing outwards from the origami may allow the RNA too many degrees of freedom and too much heterogeneity in conformations, leading to noisy structure averaging.

We could also tune the number of points of attachment between the anchored RNA and the origami. The target RNA could anchor into the scaffold at only its 5' end, only its 3' end, or both. The scaffold anchoring could be further supplemented with hybridization to a staple overhang, as in baited capture, if the anchored RNA

protrudes internally to wireframe origami.

5.3 Results

5.3.1 Anchored RRE designs

Although the above parameters represent a rather large design space, we focused first on a single RNA-scaffolded origami, a pentagonal bipyramid with six helical turns per edge (rPB66) that we know folds reliably and have characterized with cryo-EM [section 4.2.3], and varied the anchoring location and direction. The rPB66 is approximately 21 nm x 36 nm, and the RRE headgroup is approximately 6 nm x 12 nm, so the target RNA should fit within the rPB66 if designed to protrude inward.

To anchor the HIV-1 RRE fragment on the rPB66, we inserted the 232-nt sequence into the 1920-nt 23s rRNA scaffold sequence and used the original rPB66 staple set to fold the 23s rRNA while leaving the RRE fragment un-stapled, free to fold its own structure (Figure 5-1A). The placement of the sequence insertion dictates the location of the RRE fragment on the final folded object, predictable with the atomic model output from our origami design algorithm. This approach avoids folding pitfalls uncovered for the HIV-1 genome, which failed to fold a pentagonal bipyramid and an icosahedron and is adaptable to other RNA targets of interest without needing to drastically redesign the origami for each new target.

The RRE headgroup ends in a stem, such that the 5' and 3' ends of the sequence are base paired, which means both ends can anchor into the scaffold in between two adjacent scaffold nucleotides without disrupting staple hybridization to the scaffold. I created three anchored RNA designs on the rPB66 to test: the RRE protruding outwards from a vertex (rPB66_RREvertex), outwards from an edge (rPB66_RREout), or inwards from an edge (rPB66_RREin) (Figure 5-1).

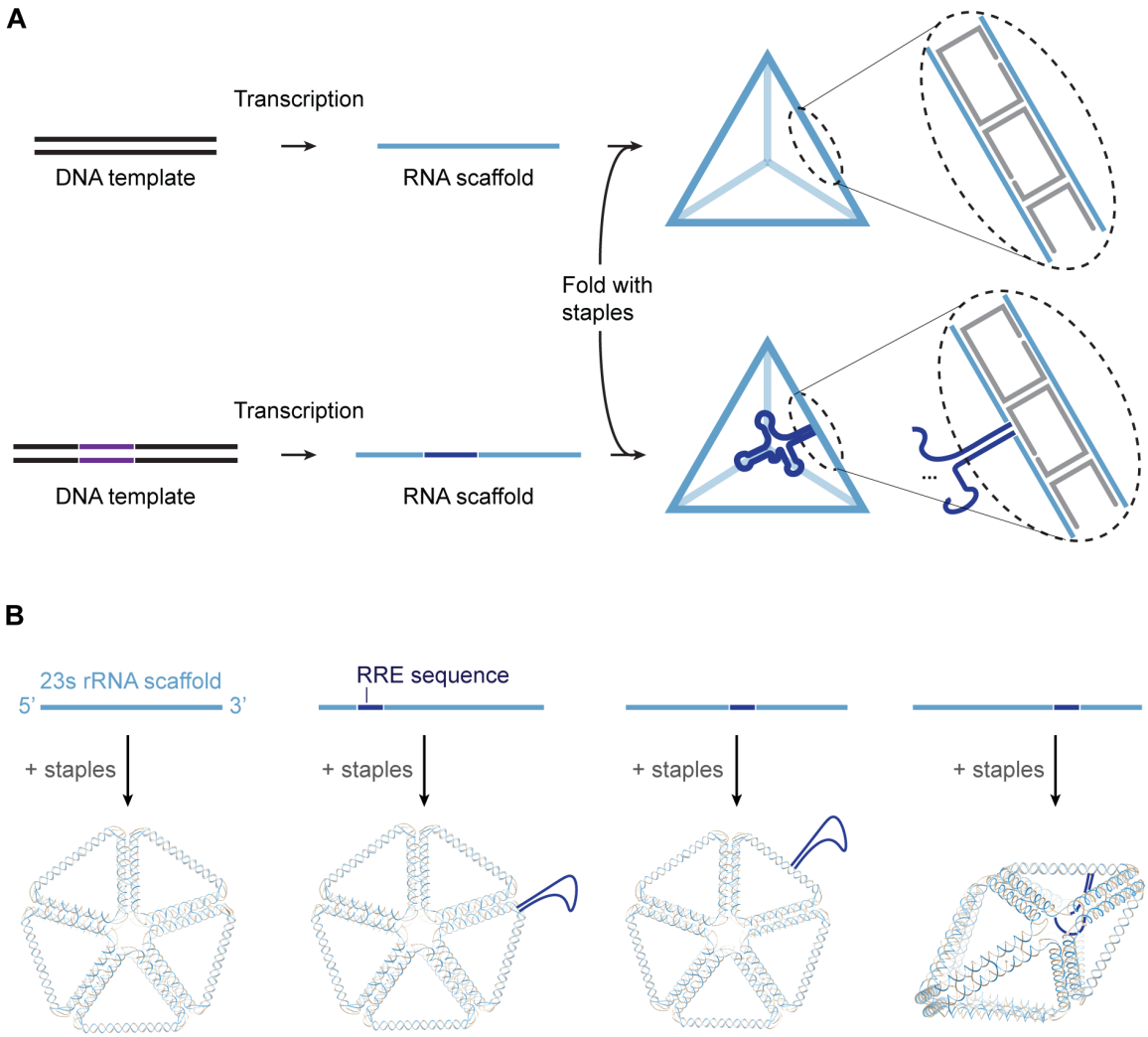

Figure 5-1: **Schematic of (A) anchoring approach and (B) anchored RRE designs.** RNA is in blue, and DNA is in grey or black, with the inserted RRE sequence in dark blue. Left to right in (B): bare rPB66, rPB66_RREvertex, rPB66_RREout, and rPB66_RREin

5.3.2 Folding characterization

Gel mobility shift assay suggest all three anchored RRE designs fold compactly with high monomeric yield (Figure 5-2). The scaffold lanes in Figure 5-2B show smears due to the strong secondary structure of the ribosomal RNA that folds in the 2 mM magnesium acetate buffer (required to run the gel of folded origami), but analysis of the scaffolds in a denaturing polyacrylamide gel shows clean transcripts with the expected molecular weights (5-2A). The folded objects each form a dense single band, collapsed into a single structure from the secondary structure smear of the 23s rRNA scaffolds. As expected, rPB66 with anchored RRE has slightly lower electrophoretic mobility than rPB66 without anchored RRE, due to the increased molecular weight.

Cryo-EM screening provided further evidence of proper folding for the three anchored RNA designs (5-2C, D, and E). Pentagonal bipyramidal objects of the correct dimension are apparent in micrographs for all three designs, although the rPB66_RREin construct appears to have lower yield for fully folded objects .

Cryo-electron microscopy

Reconstruction of rPB66_RREout from a small dataset (13000 manually picked particles) yielded inconclusive results. 2D class averages from this dataset frequently showed blurry edges, indicating heterogeneity averaged together with low signal-to-noise (Figure 5-3A). Initial 3D classification (in Relion) with eight classes likewise included no class with clear anchored RNA density. Although several classes had deformed edge density, these irregularities appeared more like malformed particles than additional RNA density (Figure 5-3B). I refined the 3D model using the boxed class in Figure 5-3B as a reference, achieving a 30Å-resolution final reconstruction The refined structure, although clearly pentagonal bipyramidal, shows minimal anchored RNA density, if any.

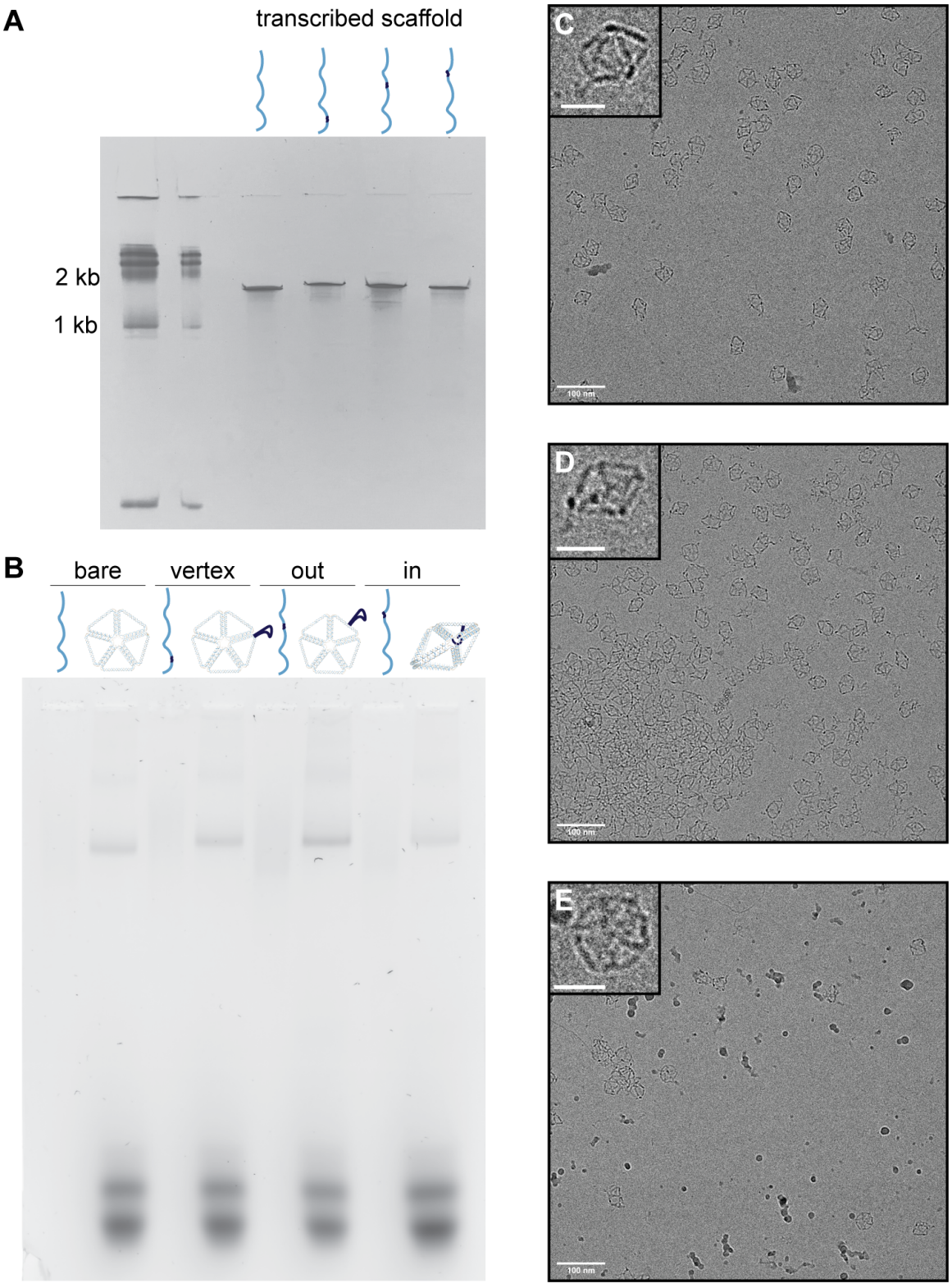
A
transcribed scaffold
2 kb
1 kb
B
bare
vertex
out
in
C
100 nm
D
100 nm
E
100 nm

Figure 5-2: **Folding and characterization of RNA-scaffolded pentagonal bipyramid with anchored HIV RRE fragments. (A)** 4% denaturing PAGE of in vitro-transcription products for four different scaffolds, left to right: 1920-nt 23s rRNA scaffold, same rRNA scaffold with 232-nt RRE sequence inserted to extrude at vertex, with RRE sequence inserted to extrude outwards at an edge, and with RRE sequence inserted to extrude inwards at an edge. **(B)** 2.5% high-resolution agarose gel in 1x TAE and 2 mM magnesium acetate, with 23s rRNA scaffolds and respective folded origami for, left to right: a pentagonal bipyramid with no RRE, with RRE at the vertex, with RRE pointing in at an edge, and with RRE pointing out at an edge. **(C)** Cryo-EM micrographs for a pentagonal bipyramid with RRE fragment anchored at a vertex pointing outwards. **(D)** Cryo-EM micrographs for a pentagonal bipyramid with RRE fragment anchored on an edge pointing outwards.**(E)** Cryo-EM micrograph for a pentagonal bipyramid with RRE fragment anchored on an edge pointing inwards. Scale bars are 100 nm in each micrograph and 25 nm in the insets.

5.4 Conclusions and future work

Despite performing analysis without symmetry, the anchored RNA fragment appears to have largely averaged into the background in the rPB66_RREout reconstruction. A few possibilities present themselves. First, the innate 5-fold rotational symmetry of the rPB66 itself may have led to particle images misaligning such that the rPB66 edges aligned within this symmetry but the anchored RRE protruded from different sides–the anchored fragment did not contribute significantly to alignment scores. In support of this hypothesis, two outer ring edges each show protruding density that appears somewhat symmetrical (in translation and reflection), and could both be the base of the RRE fragment that should have been aligned on a single edge. As a second and related possibility, the anchored RRE could be too flexible relative to the rPB66 with only the base of its stem attached to the structure, and the multiple sampled angles of RRE to the rPB66 edge result in the RRE averaging out. A third possibility is that the RRE structural ensemble remains too widely distributed, and these multiple conformations average out into noise despite being anchored to an origami.

The first two cases might be solved with a much larger data set, in which more particles can be sorted into each class, improving alignments, and finer gradations of heterogeneity can be observed in the 3D class models. Masking and focused recon-

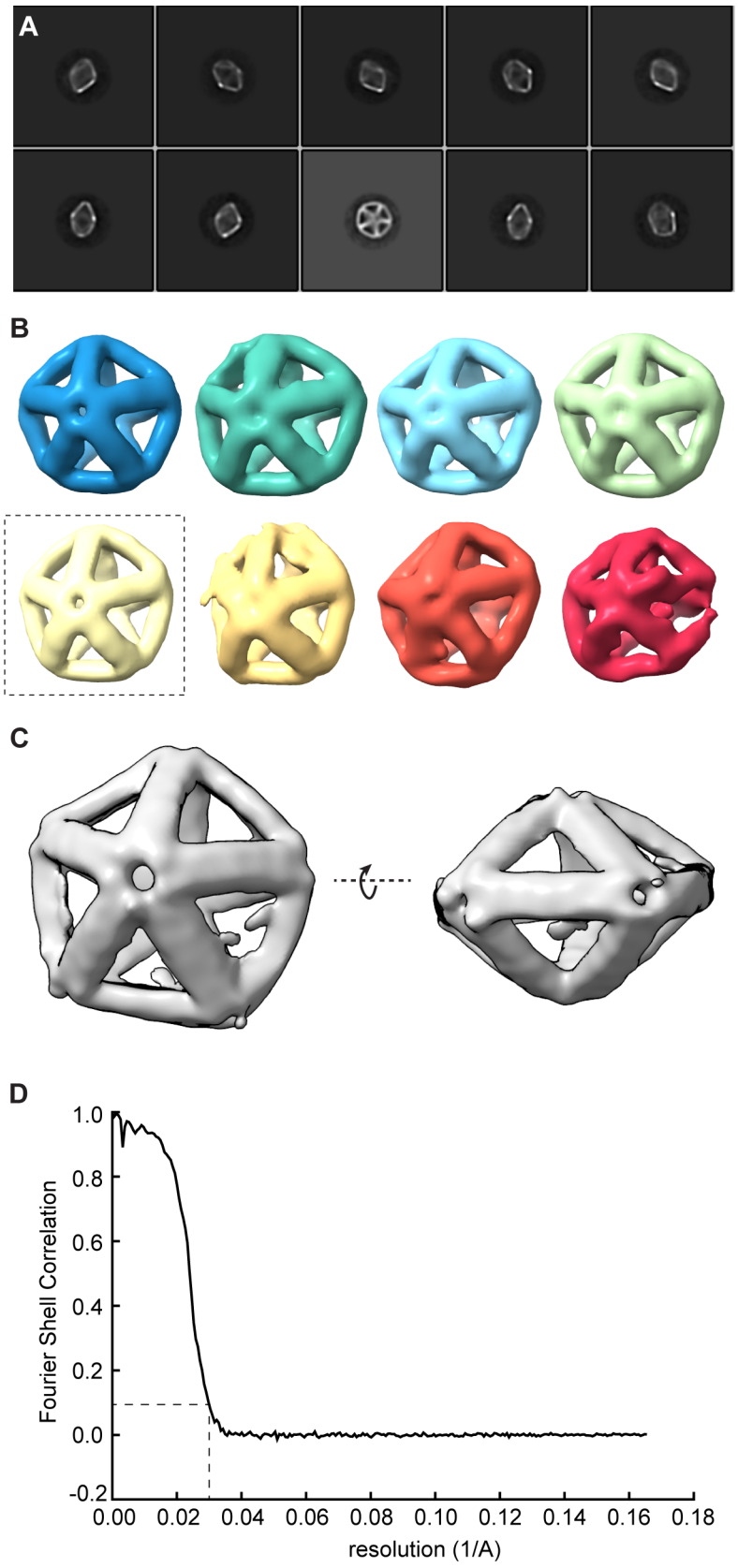
A
B
C
D
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
Fourier Shell Correlation
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16
0.18
resolution (1/A)

Figure 5-3: **Cryo-EM reconstruction of RNA-scaffolded pentagonal bipyramid with an HIV-1 RRE RNA fragment anchored on an edge protruding outwards. (A)** Example 2D class averages. **(B)** Initial 3D classification into eight classes. **(C)** Two views of a refined 3D model **(D)** Fourier shell correlation for the unmasked refined model, showing a resolution of 30 Å for the refined reconstruction.

struction on the anchored RNA density could then optimize the RRE resolution. To address the third possibility, if no anchored RRE density can be observed despite a large data set, we need to add an overhanging bait, complementary to an RRE loop, within the rPB66, although this approach only applies to the rPB66_RREin design. Overall, with the preliminary folding data and micrographs, the anchoring approach still seems promising for attaching RNA fragments at specific locations on origami. If despite larger data sets and further development the method is insufficient for studying RNA tertiary structures with high resolution, it still has promise for organizing RNA on the nanoscale, and could be applied to study or create ribozyme activity in a new context.

5.5 Methods

Materials

Oligonucleotide staples and primers and gBlock synthetic DNA sequences were purchased from Integrated DNA Technologies (IDT, Coralville, IA). HEPES, KCl, 10X TAE, magnesium acetate, high-resolution agarose, and Amicon Ultra 100 kDa MWCO spin filters were purchased from Sigma-Millipore (MA). HiScribe T7 RNA polymerase kits and Q5 2x HiFidelity PCR mastermixes were purchased from New England Biolabs (NEB, Ipswitch, MA). Quantifoil 200 mesh copper grids with R2/1 spacing were purchased from Electron Microscopy Sciences. Quiaquick PCR purification kits (Qiagen brand) and RNA clean-and-concentrator-5 kits (Zymo Research brand) were purchased from VWR. The gel stain SYBRsafe (Invitrogen brand) was purchased at 10,000x from Thermo Fisher Scientific (MA).

RNA transcription

Full sequences of the DNA templates and primers are included in Appendix D. For each anchored RRE design, we amplified DNA template from a gBlock (IDT) using Q5 2x HiFidelity PCR mastermix (NEB) according to manufacturer instructions, using a forward primer containing the T7 promoter sequence. We purified the DNA templates using a Qiaquick PCR purification kit (Qiagen) according to manufacturer instructions.

We transcribed each RNA scaffold using a HiScribe T7 RNA polymerase kit (NEB) with overnight incubation at 37°C. We then treated the 20 μl finished transcription reactions with 2 μl DNase I for 15 min at 37°C before purifying with Zymo RNA clean-and-concentrator-5 kits (Zymo Research), following manufacturer instructions except for an extra wash step. We analyzed purified transcripts with a denaturing 4% polyacrylamide gel with 6M urea, pre-run or 30 min at 130V and then loaded with samples mixed in 1x RNA loading dye (NEB) (denatured at 70°C for 10 min) and run at 160V for 2 hours. Gels were stained with 1x SYBRsafe (Invitrogen) and imaged using a Typhoon FLA 7000 imager (GE).

Origami folding and purification

We folded origami as described in Chapter 4, with 20 nM RNA scaffold and 200 or 400 nM of each DNA staple in 10 mM HEPES-KOH (from 2M pH 7.5 stock) and 300 mM KCl, folded using the 13-h ramp described in Chapter 4. For initial folding analysis, I used 50 μl folding reactions, and for cryo-EM samples, I used 4 ml folding reactions aliquotted into 96-well plates.

Amicon Ultra spin filters with 100 kDa MWCO were used to filter out excess staples and concentrate the folded origami as follows: 4-ml spin filters were cleaned with 4 ml of 'rNP filter buffer' (10 mM HEPES-KOH and 300 mM KCl, filtered with a 0.2-μm syringe filter) and spun at 1500xg for 10 min. We discarded the flowthrough and retentate, added 4 ml of folded origami sample and spun at 1200xg for 3 min, to a volume of approximately 500 μl. We discarded the flowthrough and added rNP

filter bufer to bring the total volume in the spin filter to 4 ml and again spun at 1200xg for 3-4 min, down to a volume of 200-500 μl. We repeated this step 4 times for a total of 5 buffer-exchanging spins. Samples were transferred to cleaned 0.5-ml Amicon Ultra filters (also with 100 kDa MWCO) for final concentration with a 70 min spin at 1200xg, to a volume of approximately 50 μl.

We analyzed folded origami products, purified and unpurified, using a 2.5% high-resolution agarose gel in 1X TAE, 2 mM magnesium acetate, and 1x SYBRSafe run in a 4°C room at 65 V for 2.5-4 hours. Folded origami samples were loaded with 1x DNA loading dye without SDS (NEB), while RNA scaffolds were loaded with 1x RNA loading dye (NEB) after denaturing at 70C for 10 min. Gels were imaged using a Typhoon FLA 7000 imager (GE).

Cryo-electron microscopy

3 μl of folded and purified origami (approximately 400 nM, or 1200-2000 nM in initial screens of aggregated rPB66_RREout and rPB66_RREvertex samples) was added to a Quantifoil R2/1 copper 200 mesh grid and flash frozen in liquid ethane using a Vitrobot Mark IV (Thermo Fisher Scientific), using a blot time of 4-6 s and 95% humidity. All grids were screened and imaged on an FEI Talos Arctica G2 cryo-electron microscope operated at 200 kV. Micrographs were recorded with a Falcon 3EC direct electron detector in counting mode, where each image of the movie is composed of 6 individual frames (final fraction 86 frames) with an exposure time of 5.01 s and a total dose of 56.54 electrons per Å^2. The nominal pixel size was 2.0143 Å. Micrographs were collected for rPB66_RREout with a defocus range of -1.2 to -3.0 μm, and all images were motion-corrected using RELION's implementation of MotionCor2 [116, 181]. Single-particle image processing and 3D reconstruction was performed using the image processing software package RELION [181]. 13,190 individual particles were manually picked using the RELION Manual Picker interface, with 12,779 particles used for the final 3D reconstruction after discarding bad 2D class averages. Resolution for the final map was estimated using the 0.143 criterion of the Fourier shell correlation (FSC) curve without any mask.

Chapter 6

Conclusions and Outlook

6.1 Summary and perspectives

In this thesis, we explored two avenues for stabilizing RNAs to better resolve conformations in their tertiary structure ensembles. Each approach first required the development of nucleic acid origami tools.

In the baited capture approach, an RNA is captured in a single conformation within a wireframe origami via precise placement of short overhangs complementary to single stranded RNA regions. For this approach to be broadly applicable, we first had to expand the accessible geometric space of the capturing origami (Chapter 2). We adapted a design algorithm for DX two-helix-bundle (2HB) wireframe origami to allow unpaired scaffold at vertices and adjust staple routing so that irregular objects with varied vertex angles and edge lengths could be designed faithfully to the input geometry. We found that allowing single crossovers led to more rigid and compactly folded particles than the alternative of an extra long segment between double helical crossovers. We also discovered that shorter folding ramps and lower magnesium content tended to decrease aggregation of this origami. With cryo-EM, we determined that the origami folded into the predicted geometry, although objects with long edges (>90 bp) were highly flexible. The latter result informs origami design for the purpose of RNA stabilization: minimizing edge lengths will likely reduce the inherent flexibility of the system and improve the maximum attainable resolution for

the origami and attached RNA.

I then designed a 52-bp-per-edge DNA-scaffolded tetrahedron with three overhanging baits to capture an engineered tRNA as a proof-of-principle for the baited capture approach (Chapter 3). I confirmed capture with co-localization assays, and cryo-EM revealed the tRNA density connected to the origami at 3 sites, suggesting all three baits bound as designed. With this initial design, we tested steps of a high throughput workflow that would allow application of the baited capture approach to RNAs of unknown structure.

The development of RNA-scaffolded wireframe origami enabled a second approach to stabilizing RNA for better resolution of tertiary structure conformations, in which RNA is directly anchored by using trailing RNA sequence on one or both sides of the target RNA to scaffold the origami. We modified the original DAEDALUS algorithm for DX 2HB wireframe origami to design A-form structures–the geometry adopted by RNA in double helices–and I compared multiple routing schemes for staple and scaffold crossovers to account for pitch and twist in A-form helices (Chapter 4). We analyzed four of our structures, folded with three different RNA scaffolds, with cryo-EM, and while there was some moderate edge bowing and twisting in a couple of the origami, the reconstructions correlated well with the predicted structures.

For an RNA-scaffolded origami structure we verified to fold robustly, I inserted a 232-nt fragment of the HIV-1 genome into the 23s rRNA scaffold sequence to investigate the anchoring approach (Chapter 5). Gel mobility shift assays and cryo-EM screening suggest that the original origami structure, a pentagonal bipyramid, still folds with this inserted sequence anchored on an edge or vertex, whether interior or exterior to the wireframe object. Further study is required to thoroughly compare how the placement of the anchored RNA affects the imaging and reconstruction process, and it is possible that additional stabilization will be required to reconstruct larger RNAs. I expect the anchoring approach will be more useful for structural biology when combined with the recent computational advances in classification of heterogeneity.

6.2 Relative advantages and limitations of baited capture vs. anchoring approach

Although both approaches have promise in uncovering RNA tertiary structure, each method has its own advantages and its own applications in RNA structural biology and beyond. For both approaches, a key challenge is ensuring that any tertiary structures determined are relevant to the native structure. Using minimal bait lengths helps address this challenge in the baited capture approach, and an advantage of the anchoring approach is the even more minimal interference with RNA structure. Chemical probing like DMS-MaPseq is an important validation step in both approaches, comparing secondary structures of RNA on the origami with those in the cell.

The anchoring approach is much more straightforward to implement than the baited capture approach, theoretically requiring only a single or very few origami designs even with no prior knowledge of the RNA structure beyond length. To identify origami that capture a target RNA of unknown structure, without reshaping the target RNA, many thousands of bait arrangements may need to be folded and tested. Additionally, in the baited capture approach, many well-folded origami particles may be empty, requiring classification and discarding of a large fraction of picked particles from cryo-EM images as they cannot be used to reconstruct the target RNA. The anchoring approach, on the other hand, guarantees that every folded origami particle has the target RNA attached, so the full data set can be used for reconstruction. However, for particularly large and heterogeneous RNAs, the anchoring approach may require supplemental overhanging baits, and multiple arrangements of these baits may need testing, for successful classification and reconstruction of the individual conformations.

The baited capture approach is more suited to studying RNA tertiary structure in a cellular context. Whereas in the anchoring approach the target RNA is subject to the folding conditions of the origami, the target RNA and the origami are folded separately and then incubated in the baited capture approach. Baited capture can thus

probe the target RNA structure in more cell-like conditions, perhaps eventually even in cells directly. For example, biotin-labeled origami delivered to cells could capture target RNA in the cytoplasm, and we could purify the resulting bound complex for cryo-EM characterization.

6.3 Future directions

Both the baited capture design and anchoring approach have opportunities for further development, as well as potential non-structural-biology applications.

Improving bait placement design for baited capture

One key area for refinement of the baited capture approach lies in the library design stage. While the bait placement algorithm described in Chapter 3 addresses the large design space, two other computational approaches could further tailor the origami design to target RNAs, whether for structural biology purposes or other applications. The first is to model the structural ensemble of the target RNA with a focus on the loops to be baited. With a secondary structure as input, the RNA can be abstracted to a ball-and-stick model: single-stranded loops are balls on the end of hairpin helices (sticks), and the phosphate backbones between neighboring hairpins serve as hinge points, around which the ball-and-sticks can rotate. With a Monte-Carlo approach applied to the ball-and-stick model, we can compute sets of possible distances between modeled loops. These distance sets then serve as input to automated DNA origami design and bait placement. Such a method would provide information on the potential heterogeneity of the target RNA and enable us to tailor library design to a wider range of conformations. Any available experimental data on conformations, e.g. from smFRET, could also complement the modeling step.

As an alternative to the bait placement algorithm that uses existing origami designs, we could further tailor the origami library design with custom geometries based on the RNA loop distances, as determined with the method described above or from pre-existing models. For example, one could design new origami by creating poly-

hedra with connecting two RNA loop points at a time (work now begun by fellow graduate student Matthew Allan). With the updated DAEDALUSv2 algorithm described in chapter 2, we can actually design DNA origami able to accommodate such customized geometry, regardless of any irregular edge lengths or vertex angles.

Structural biology applications

Once expanded and refined, either approach to RNA tertiary structure determination could be applied to answer open structural biology questions in the field. For example, although some evidence suggests that RNA folds modularly, our inability to obtain full genome structures has prevented us from confirming this theory—we study large RNA tertiary structure modularly, resulting in some inherent bias [182–184]. And in some cases, like with the SARS-CoV-2 frameshift stimulating element, the larger RNA context was shown to be important to the secondary structure of a small fragment [117]. To address this knowlege gap at the tertiary structure level, we could apply our origami-based platform to study tertiary structure of RNA fragments of varying length. For example, the 350-nt 5' untranslated region (5'UTR) of HIV-1 compared with the first 500 nt of the genome, or the first 2000 nt (which would include the frameshift stimulating element as well). Note, however, that particularly large RNA segments may not be sufficiently stabilized in the anchoring method and may not fit in a small rigid origami, and therefore require baited capture in a superstructured origami or in chains of origami, although such an approach becomes highly complicated in the existing workflow. The limits of tractable RNA size with the baited capture and the anchoring approaches still need to be determined.

Other aspects of tertiary structural biology for viral RNAs can be probed with the origami framework in the presence of proteins or other cellular components. For example, how do specific cellular components important to HIV reverse transcription (RT) and packaging relate with the tertiary structure of the HIV 5'UTR? Reverse transcription is of great clinical interest—a majority of the anti-HIV drugs currently target the RT step [185]. Reverse transcription is initially primed by a cellular $tRNA_{Lys}$ annealing to the primer binding site (PBS), a large loop in the 5'UTR [12, 185].

This annealing requires some denaturing of both the tRNA and presumably the viral RNA, and investigating how this event alters the structure of the 5'UTR could be instructive about the early stages of RT, such as binding of the reverse transcriptase. The tRNA is predicted to interact with a few other sites in the viral genome as well, such as an A-rich stem loop near the PBS, and a cryo-EM structure of the complex could confirm which of these predictions are accurate [12]. It is possible a population of viral RNA does not bind to the tRNA, if the PBS is secluded in that conformation, for example, and knowledge of these structures too would inform understanding of the mechanism of RT initiation.

Of course, many other large RNAs of mechanistic and disease relevance have unknown 3D structures, and the frameworks outlined in this thesis could be applied to study tertiary structures of a wide variety of RNA targets, from other viral RNA genomes to long non-coding RNAs and mRNAs.

Applications in RNA detection

For RNAs with known tertiary structure, the workflows described above and in Chapter 3 have translational potential in RNA detection or selection. Most current RNA detection techniques are purely sequence based, with the highly sensitive quantitative reverse transcription and polymerase chain reaction (qRT-PCR) as the gold standard [186, 187]. Not all rely on amplification and thus can provide faster detection results, particularly useful in viral RNA detection [188].

With the baited capture approach, origami selectively binds RNA based in small part on sequence, due to the short baits that need to be complementary to the RNA, and in large part on shape, because binding affinity depends on matching 3D geometric arrangement of baits to target locations on the RNA. The ability to detect RNA shape specifically may be useful in viral detection, particularly when many variants emerge and circulate. Even when related viruses differ in sequence, they experience strong evolutionary pressure to maintain particular RNA structures [189]. Thus, a shape-based RNA selection system has potential as a virus detector, specific to a particular family or to a strain, that may be less likely to experience high false negative

rates when mutations occur.

Several stages are required to develop the baited capture approach into a detection method. First, we need to further characterize the sequence- and shape-specificity of baited capture. Can the origami designed to bind the engineered tRNA or a viral genome fragment preferentially bind the target RNA over other tRNAs or scrambled sequence? Does it select related RNAs with the same predicted structure and target loop sequences, but with internal sequence differences? Second, we need to add a signaling component to the RNA, for example releasing a fluorescent signal upon RNA binding. This signal could potentially be achieved with strand displacement of a fluorescently-labeled oligo bound to the baits, or with CRISPR/Cas-driven cleavage of a signaling moiety, or measurement of a molecular weight shift after binding.

Synthetic biology applications

We could leverage the ability to precisely organize RNAs in three dimensions on the nanoscale to direct catalysis with ribozymes in a cell-free, and potentially protein-free, context. Efforts in this direction could prove useful in synthetic biology, and also in investigating basic biology questions around what aspects of particular ribonucleoprotein complexes contribute to catalysis. Do the proteins in RNase P, the ribosome, or the spliceosome specifically participate in catalysis, or are they purely structural as has been speculated [190]? Can we achieve peptidyl transfer when the peptidyl transferase center of the ribosome is organized by nucleic acid origami instead of proteins, and if we bring two tRNAs in close proximity via bait binding to the 3'-CCA? To answer such questions, and to organize ribozymes, precise knowledge of 3D structure is unnecessary, as complete fixation of the RNAs in one conformation is unnecessary to–and indeed would likely inhibit–their activity. For these applications, anchoring might be preferable to baited capture because the covalent anchoring in scaffold is a much stronger attachment point than hybridization with short overhanging oligos. However, both approaches could be employed or eventually even used in tandem to orchestrate multi-step ribozyme activity.

6.4 Closing thoughts

With my thesis work, I expanded tools for wireframe nucleic acid origami design and folding, and with these tools, investigated two routes to tertiary structure determination for larger, dynamic and previously intractable RNA targets. More 3D RNA structures will contribute to a mechanistic understanding of the roles cellular or viral RNAs play in regulation of the cell or viral life cycle, and could expand drug design to include a wide range of RNA targets in genetic or viral disease.

Appendix A

Supplement to DAEDALUSv2

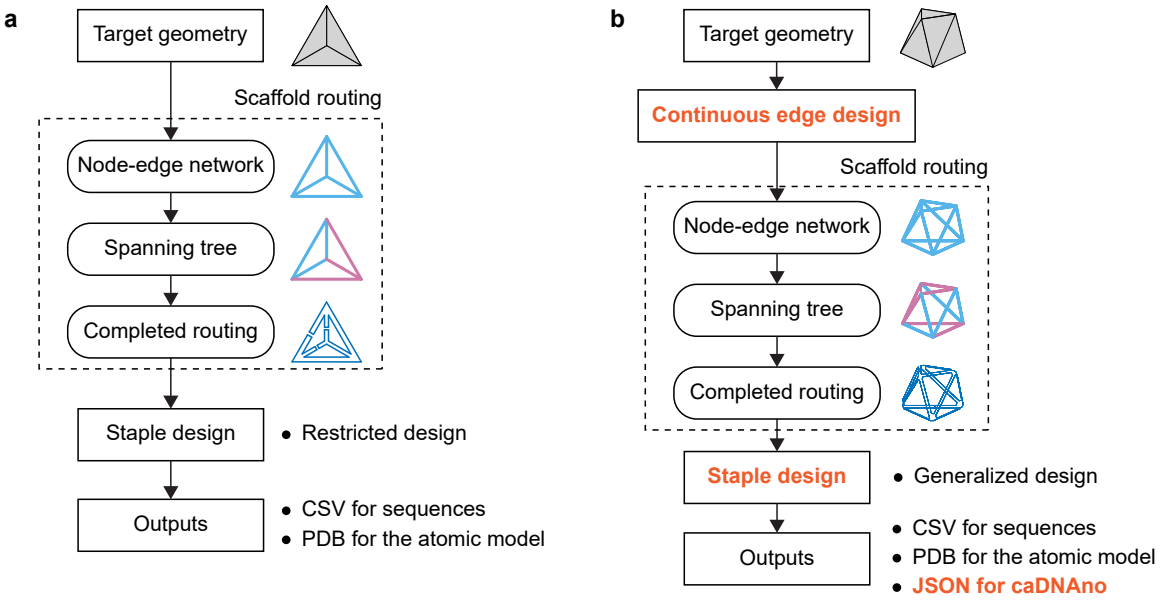

Figure A-1: **Comparison of the procedure between DAEDALUS and DAEDALUSv2.** **(a)** DAEDALUS and **(b)** DAEDALUSv2 design procedure.

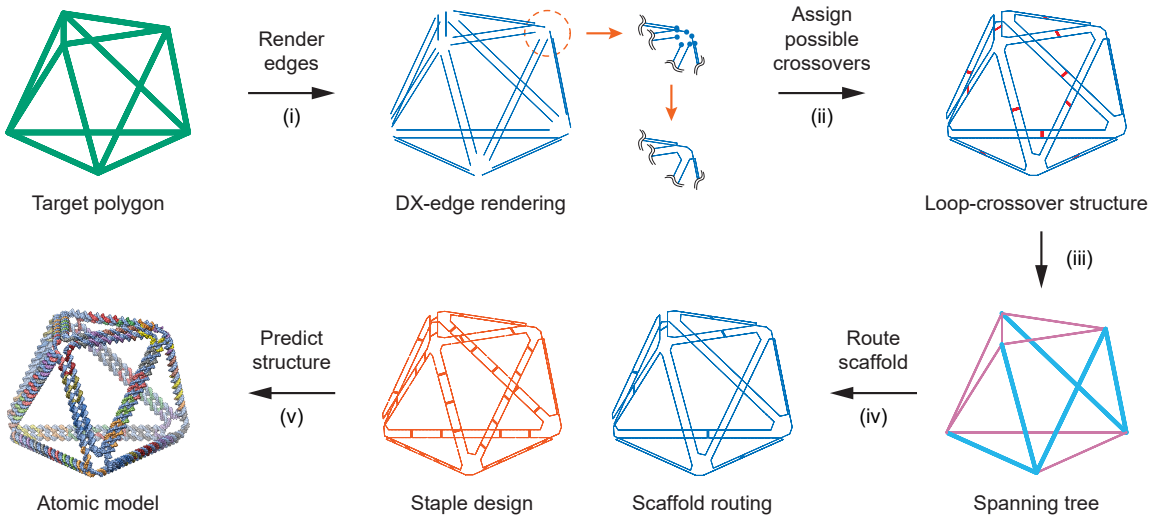

Figure A-2: **Schematic illustrating design algorithm for the 3D wireframe scaffolded DNA origami with asymmetric and irregular shapes.** The arbitrary polyhedral geometry is discretized to line segments (step i) to represent two DNA duplexes per wireframe edge with the endpoints joined (step ii) to form closed loops with geometrically allowable scaffold double-crossovers between them. For the scaffold routing, spanning tree is computed (step iii) based on the target geometry. Each of the edges that is a member of the spanning tree is assigned no scaffold double-crossover, whereas each remaining edge is assigned the scaffold double-crossover (step iv). Staple sequences generated by the algorithm were used with the input scaffold and a 3D atomic-level structural model is generated (step v) assuming canonical B-form DNA geometry.

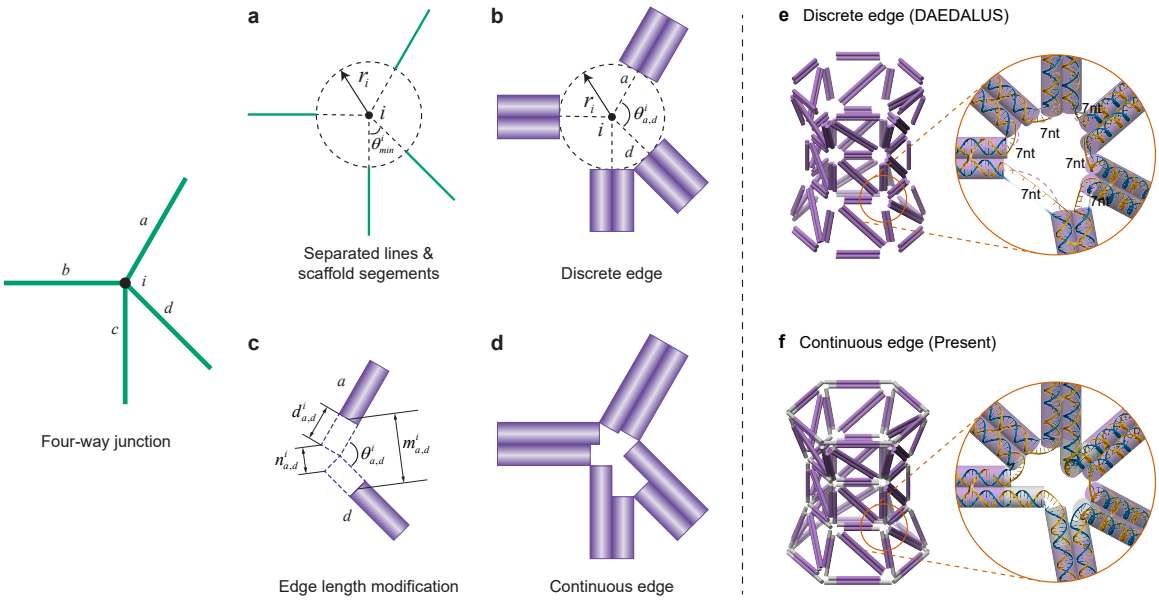

Figure A-3: **Continuous edge design for irregular shapes.** The initial geometry has a four-way junction at an unequal vertex, , having each of four edges denoted by from 'a' to 'd'. **(a)** The initial off-set distance (apothem), , is calculated by the minimum angle, at the th vertex and the number of arms in th vertex. **(b)** Each separated line with initial offset distance, is drawn by the cylinder representing the double strand DNA. **(c)** Two cylinders located in adjacent separated lines that do not contact each other are modified to connect each other at the end. The new off-set distance of two cylinders from the vertex can be calculated by two given distances and and the given angle , in which the subscripts a and d represent the edge identifier. **(d)** The continuous edges incorporating the offset distance are shown. The wireframe DNA origami with the **(e)** discrete edge and **(f)** continuous edge length.

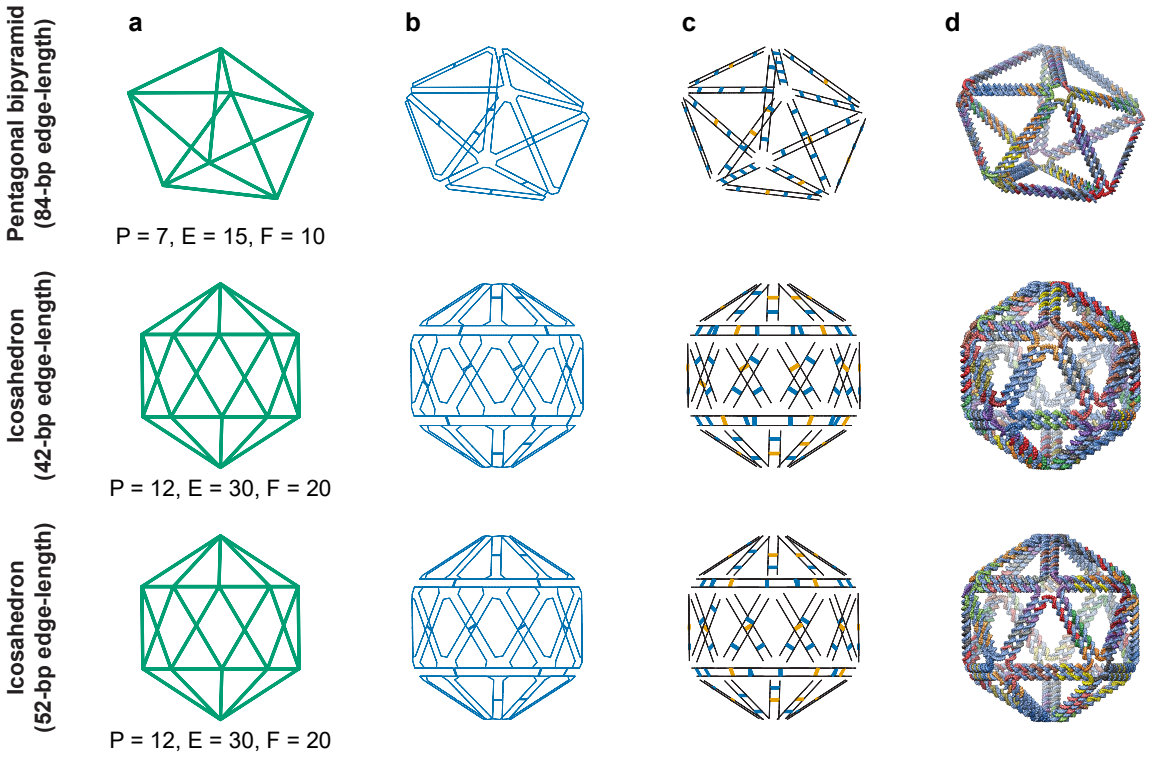

Figure A-4: **Target geometries, scaffold routing paths, crossover models, and atomic models for regular objects.** **(a)** The target geometries. The letters, P, E, and F denote the number of points, edges, and faces, respectively. **(b)** Scaffold routing path. The continuous blue loop is the single-stranded DNA scaffold that routes throughout the entire origami object of arbitrary shape. **(c)** Scaffold double crossovers (blue) determined by the spanning tree algorithm and staple double crossovers (orange). **(d)** Atomic model.

Table A.1: **Design parameters for regular objects.** # indicates the type of scaffolds in Table 3 of this supplement

Target geometry	**Edge Length**	**Scaffold**			**Staples**		
		Required length	# of double crossovers	# of unpaired nucleotides	# of staples	# of double crossovers	# of unpaired nucleotides
Pentagonal bipyramid	84-bp	2,520-nt (#phPB84)	9	0	61	51	155
Icosahedron	42-bp	2,520-nt (#phPB84)	19	0	64	41	300
Icosahedron	52-bp	3,120-nt (#phIco52)	19	0	72	41	300

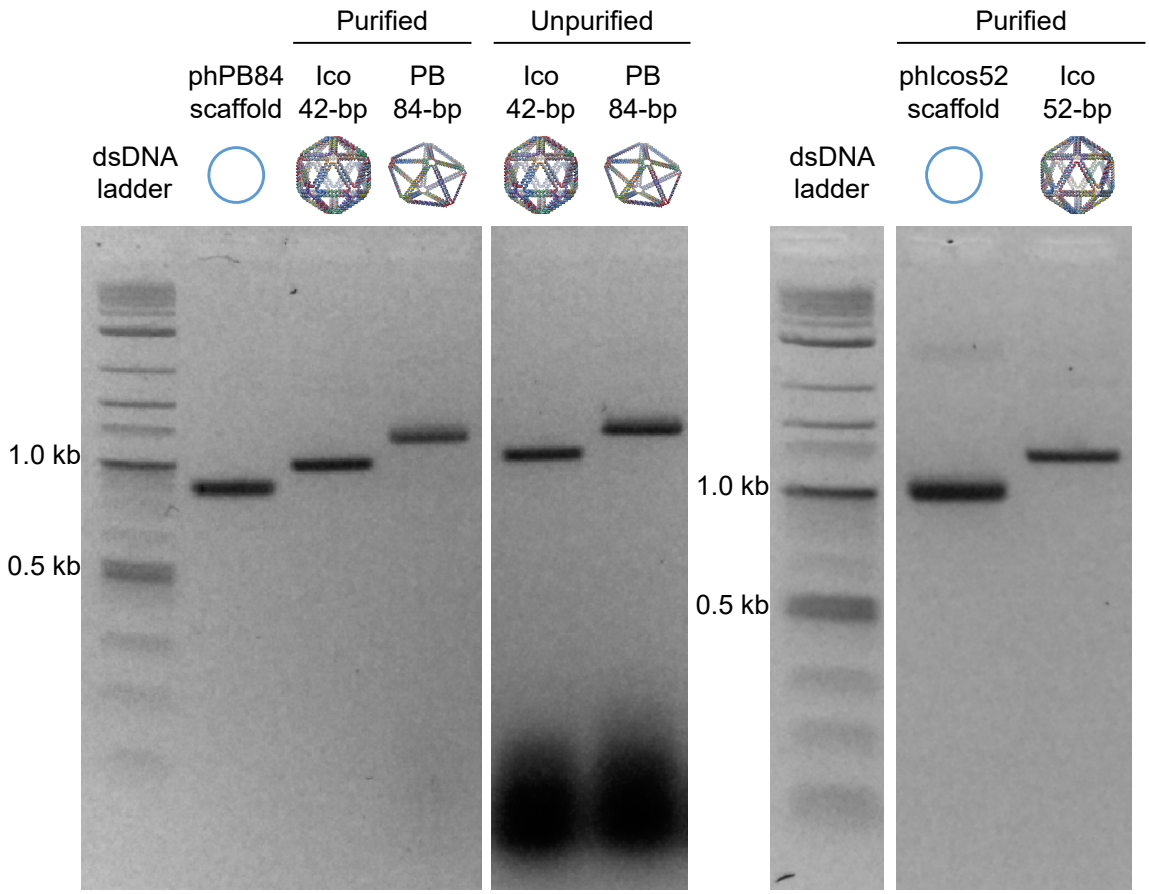

Figure A-5: **Agarose gel electrophoresis for DX-based regular objects.** The Ico 42-bp and PB 84-bp both are folded using the phPB84 scaffold, while the Ico 52-bp is folded using the phIco52 scaffold (Table 3 of this supplement). Ladder is the 1-kb plus ladder from New England Biosciences.

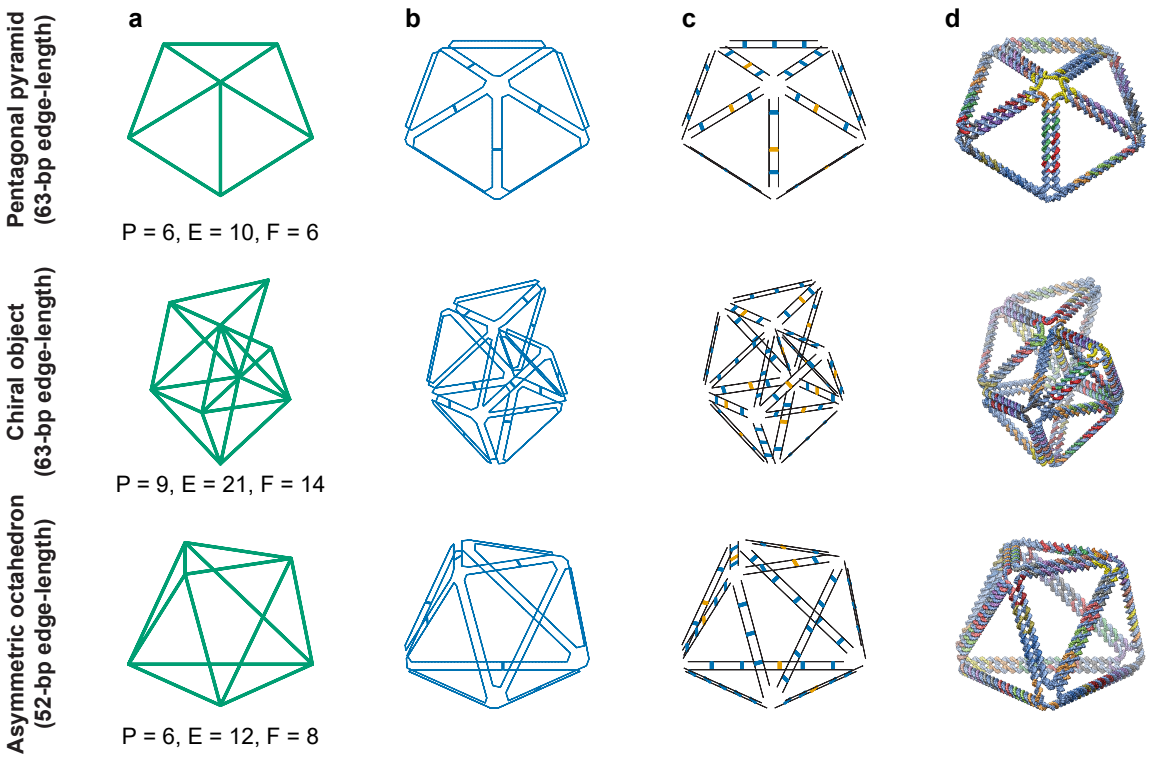

Figure A-6: **Target geometries, scaffold routing paths, crossover models, and atomic models for irregular objects.** **(a)** The target geometries. The letters, P, E, and F denote the number of points, edges, and faces, respectively. **(b)** Scaffold routing path. The continuous blue loop is the single-stranded DNA scaffold that routes throughout the entire origami object of arbitrary shape. **(c)** Scaffold double crossovers (blue) determined by the spanning tree algorithm and staple double crossovers (orange). **(d)** Atomic model.

Table A.2: **Design parameters for irregular objects.** # indicates the type of scaffolds in Table 3 of this supplement.

Target geometry	**Min. Edge Length**	**Scaffold**			**Staples**		
		Required length	# of double crossovers	# of unpaired nucleotides	# of staples	# of double crossovers	# of unpaired nucleotides
Pentagonal pyramid	63-bp	1,376-nt (#pF1a)	5	41	33	25	80
Chiral object	63-bp	2,914-nt (#phIco52)	13	117	69	50	161
Asymmetric octahedron	63-bp	2,256-nt (#phPB84)	7	80	54	41	69

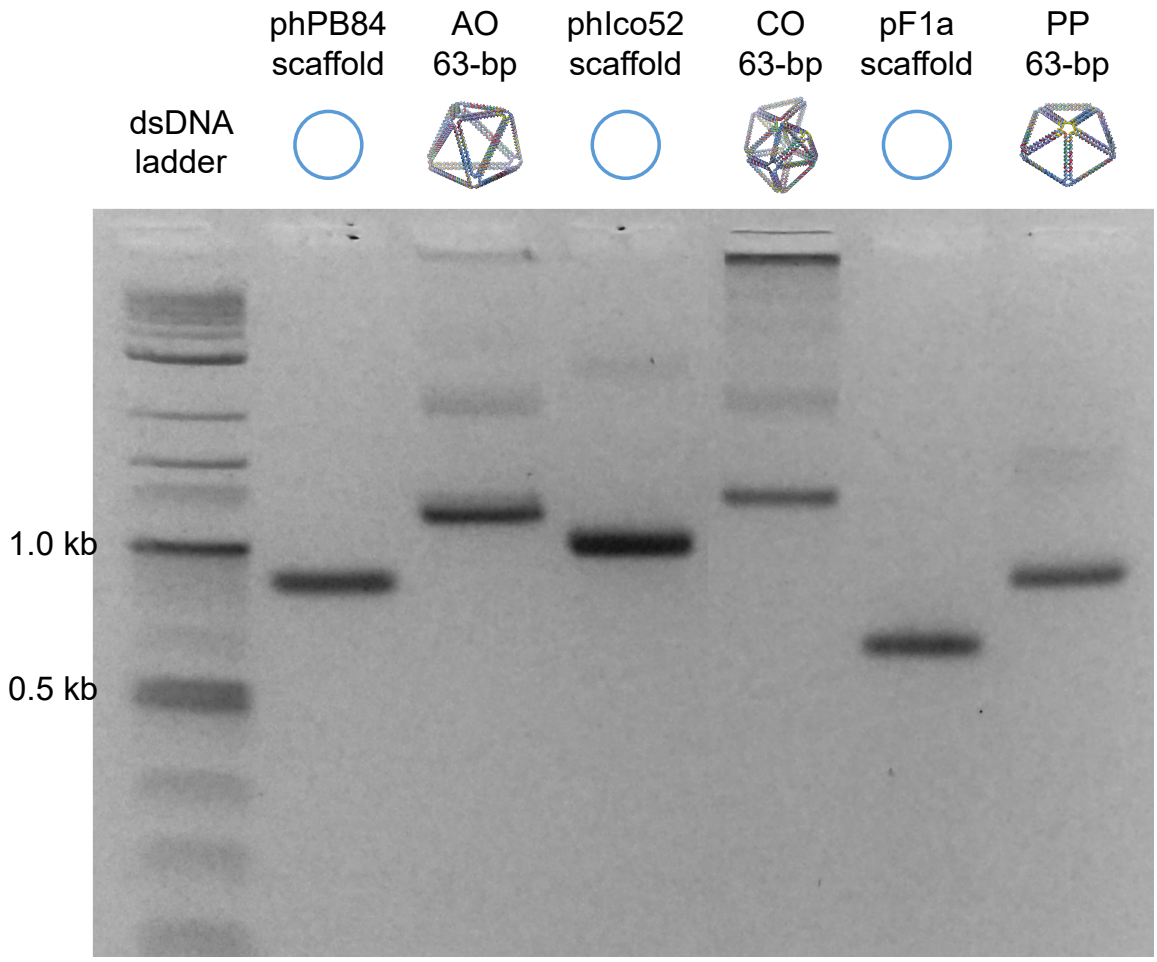

Figure A-7: **Agarose gel electrophoresis for DX-based irregular objects.** AO: asymmetric octahedron, CO: chiral object, and PP: pentagonal pyramid. The AO 63-bp is folded using the phPB84 scaffold, while the CO 63-bp and PP 63-bp are folded using the phIco52 and pF1a scaffold, respectively (Table S3). Ladder is the 1-kb plus ladder from New England Biosciences.

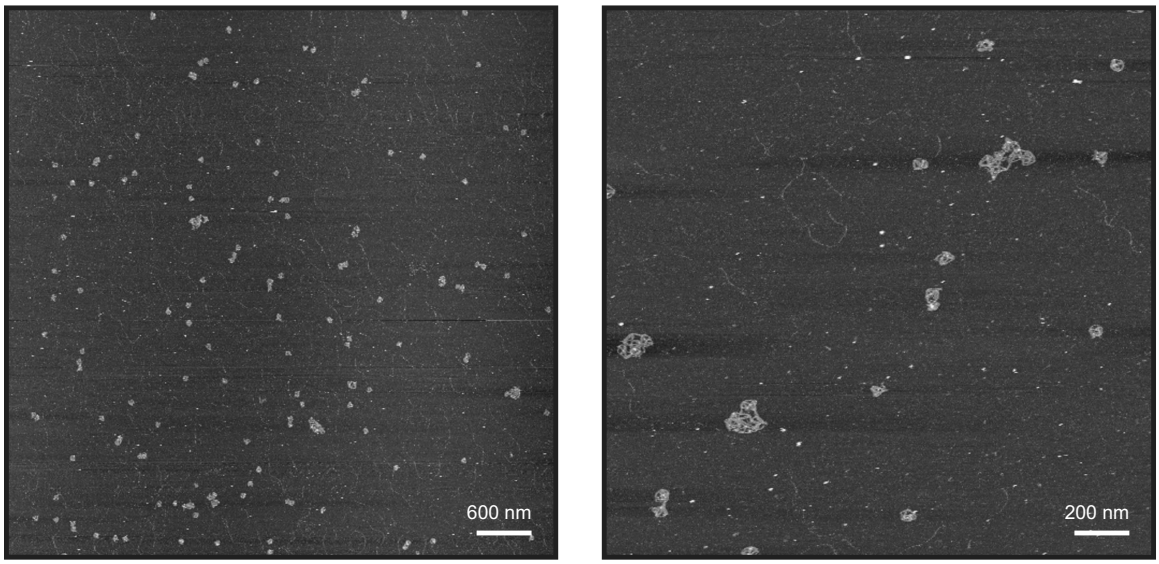

Figure A-8: **AFM imaging of asymmetric 63-bp asymmetric octahedron folded over 12.5 hours.**

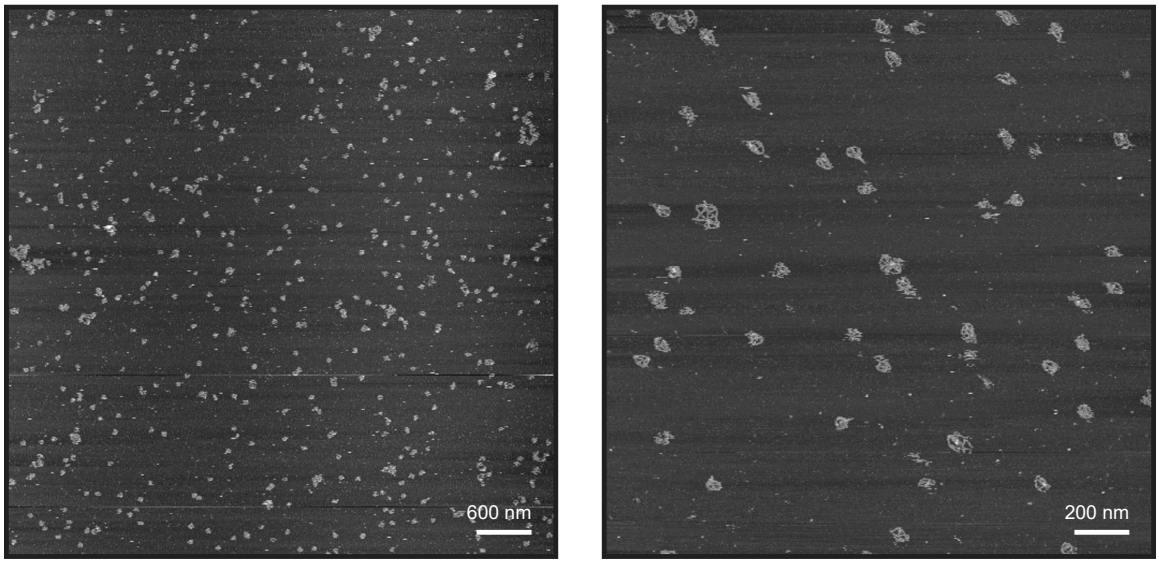

Figure A-9: **AFM imaging of asymmetric 63-bp asymmetric octahedron folded over 2 hours.**

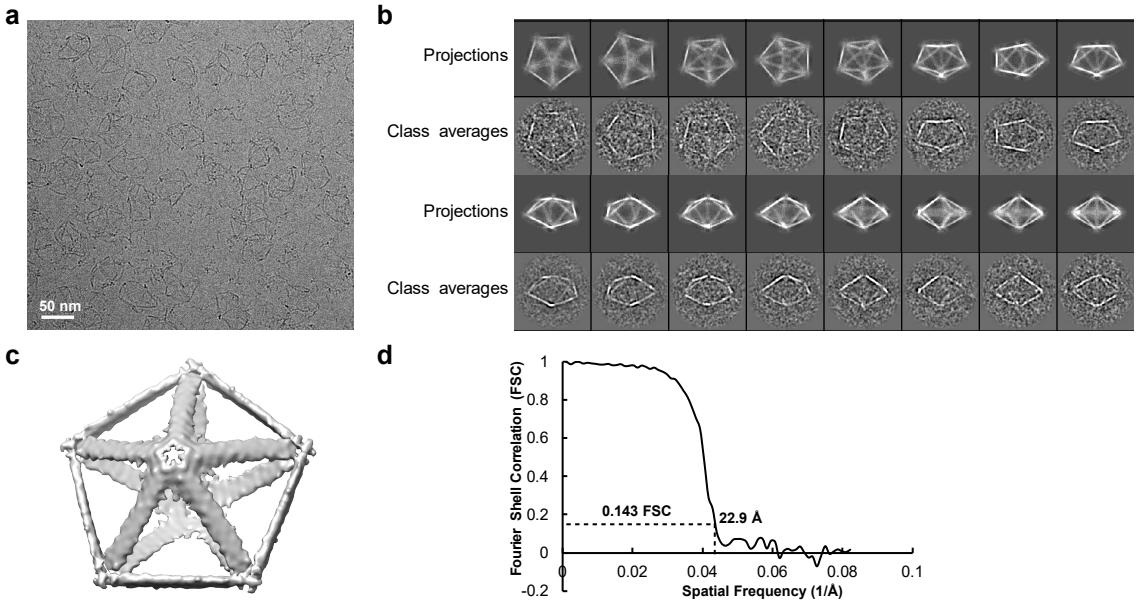

Figure A-10: **Cryo-EM imaging of a DAEDALUS2 pentagonal bipyramid of 84-bp edge length.** **(a)** Representative electron micrograph, **(b)** sample 2D class averages and model projections, **(c)** 3D reconstructed density, **(d)** Fourier Shell Correlation plot showing resolution information.

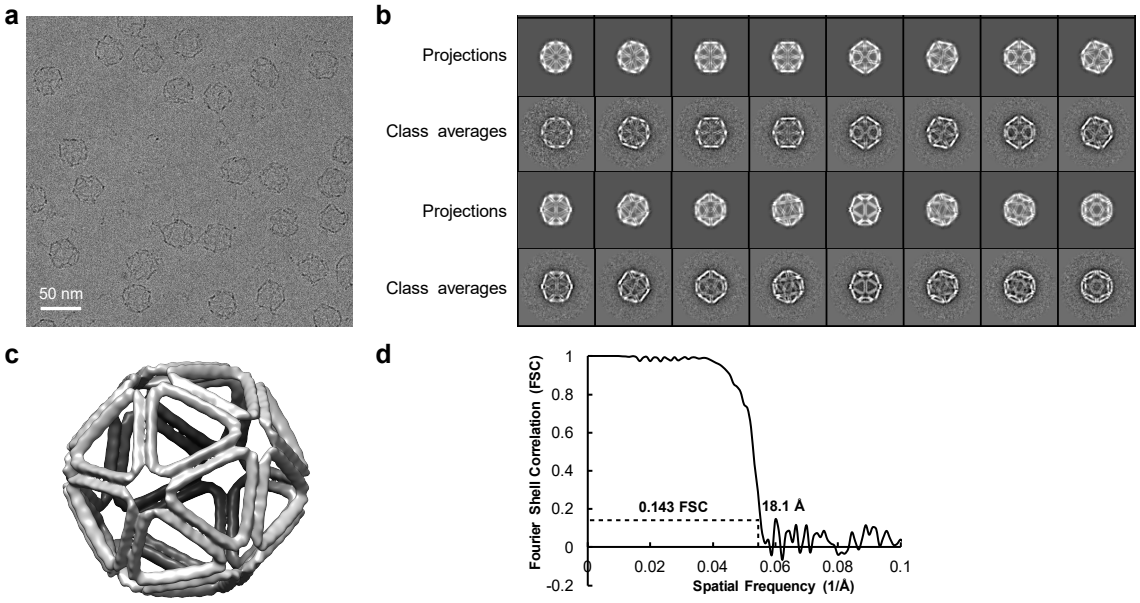

Figure A-11: **Cryo-EM imaging of an icosahedron of 42-bp edge length.** **(a)** Representative electron micrograph, **(b)** sample 2D class averages and model projections, **(c)** 3D reconstructed density, **(d)** Fourier Shell Correlation plot showing resolution information.

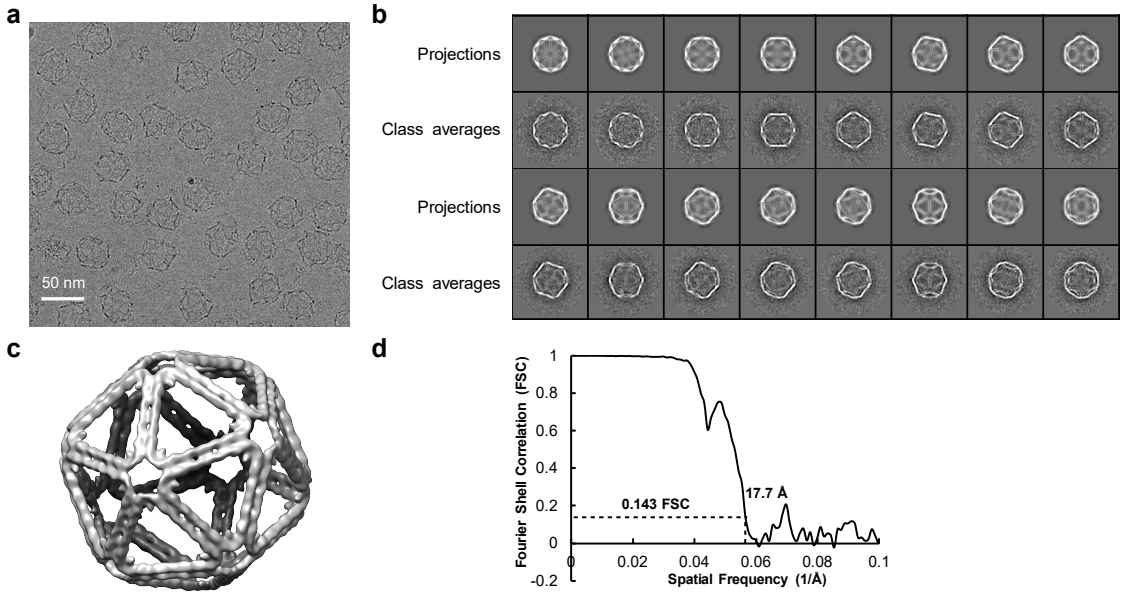

Figure A-12: **Cryo-EM imaging of an icosahedron of 52-bp edge length.** **(a)** Representative electron micrograph, **(b)** sample 2D class averages and model projections, **(c)** 3D reconstructed density, **(d)** Fourier Shell Correlation plot showing resolution information.

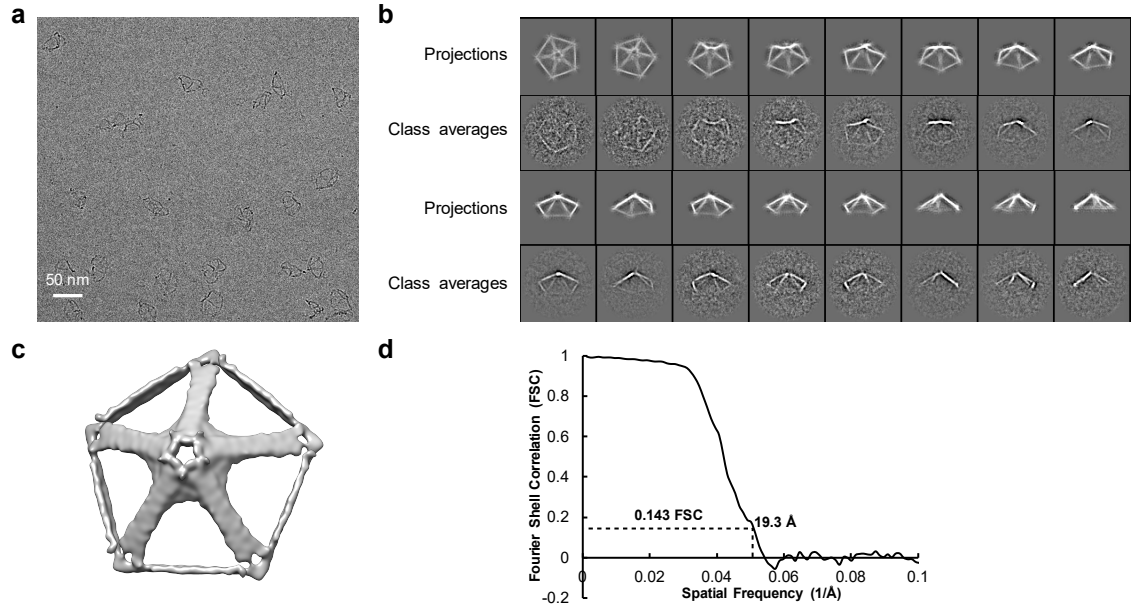

Figure A-13: **Cryo-EM imaging of an pentagonal pyramid of 63-bp edge length.** **(a)** Representative electron micrograph, **(b)** sample 2D class averages and model projections, **(c)** 3D reconstructed density, **(d)** Fourier Shell Correlation plot showing resolution information.

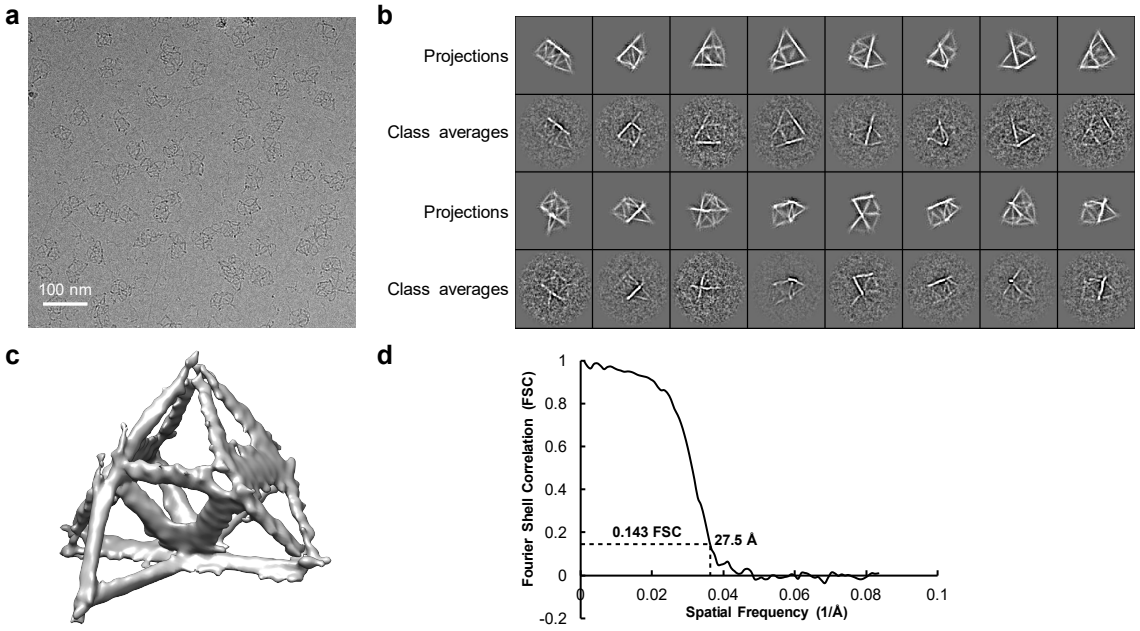

Figure A-14: **Cryo-EM imaging of a chiral object of minimum 63-bp edge length.** **(a)** Representative electron micrograph, **(b)** sample 2D class averages and model projections, **(c)** 3D reconstructed density, **(d)** Fourier Shell Correlation plot showing resolution information.

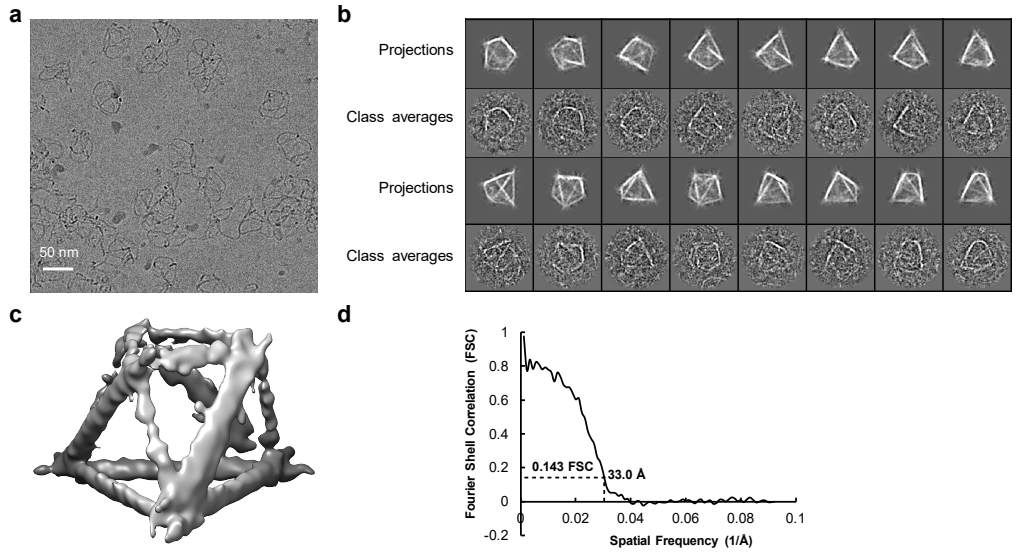

Figure A-15: **Cryo-EM imaging of an asymmetric octahedron of minimum 63-bp edge length.** **(a)** Representative electron micrograph, **(b)** sample 2D class averages and model projections, **(c)** 3D reconstructed density, **(d)** Fourier Shell Correlation plot showing resolution information.

Appendix B

Supplement to RNA-scaffolded 3D wireframe origami

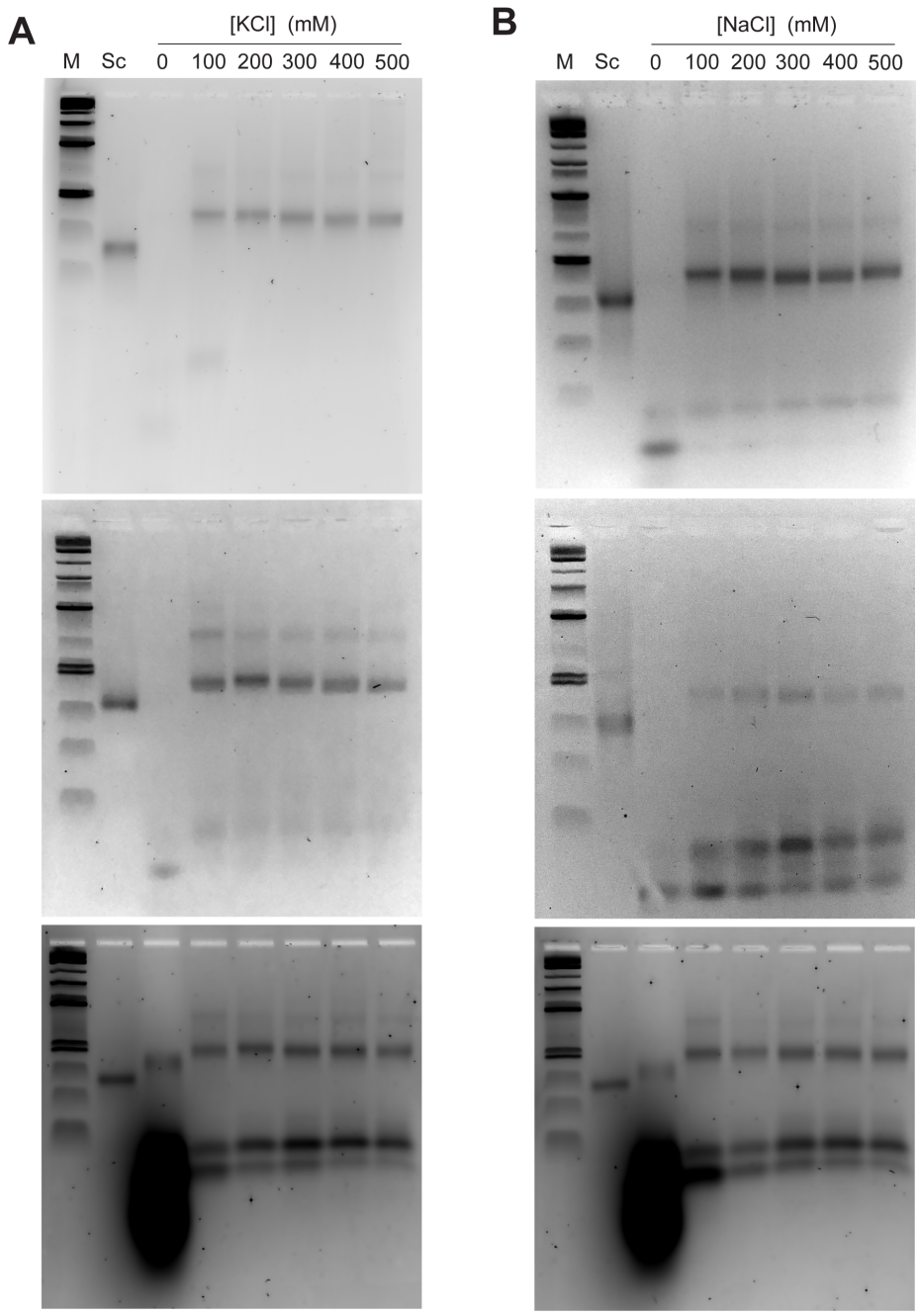

Figure B-1: **Triplicate titration of monovalent salts for a RNA-scaffolded tetrahedron (scaffold crossover asymmetry design) with 66-bp edge lengths.** **(A)** Titration series of KCl. **(B)** Titration series of NaCl. Marker is 1kb plus DNA ladder from New England Biosciences.

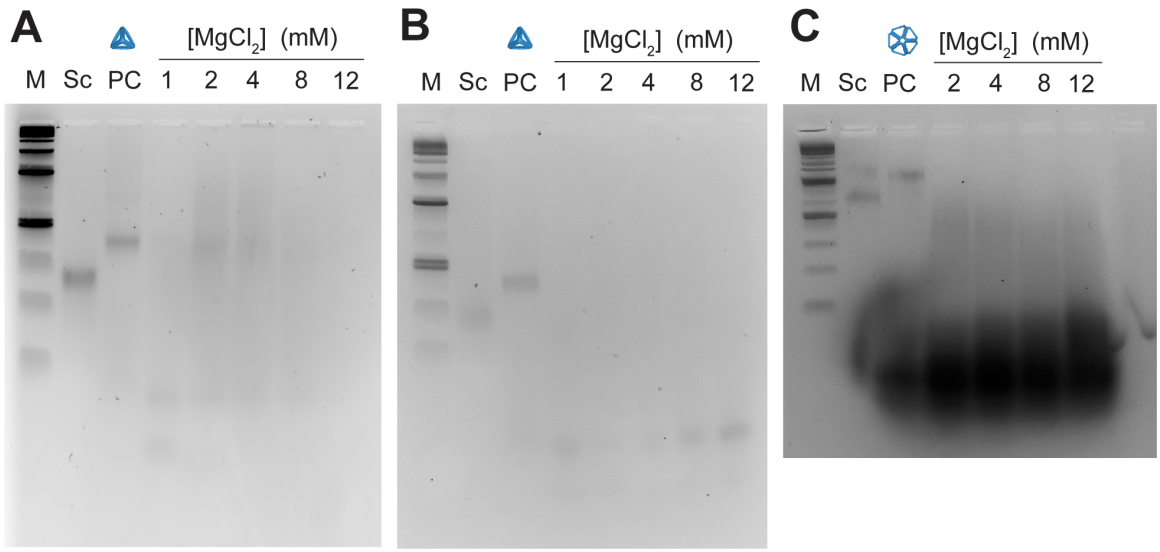

Figure B-2: **(A)** and **(B)** Duplicate titration of MgCl2 on a regular RNA-scaffolded tetrahedron (scaffold crossover asymmetry design) with 66-bp edge lengths. **(C)** A titration of MgCl2 on a regular RNA-scaffolded pentagonal bipyramid (alternative A-form design) with 66-bp edge lengths.

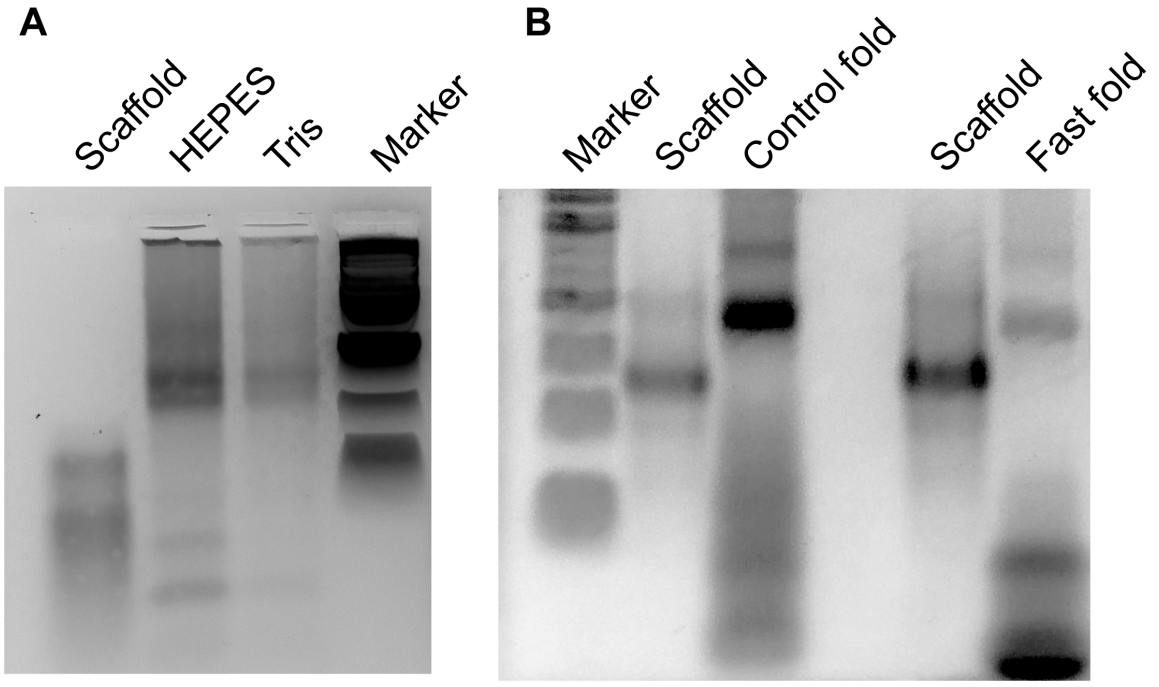

Figure B-3: **Comparison of alternative folding conditions for a tetrahedron with alternative A-form design (scaffold crossover asymmetry). (A)** HEPES was compared to Tris-HCl pH 8.1 for the long folding protocol modified from [57]. **(B)** HEPES with the modified folding protocol ("Control fold") was tested against a previously published fast-folding protocol in magnesium [78], showing near equivalent yields when adjusted for loading.

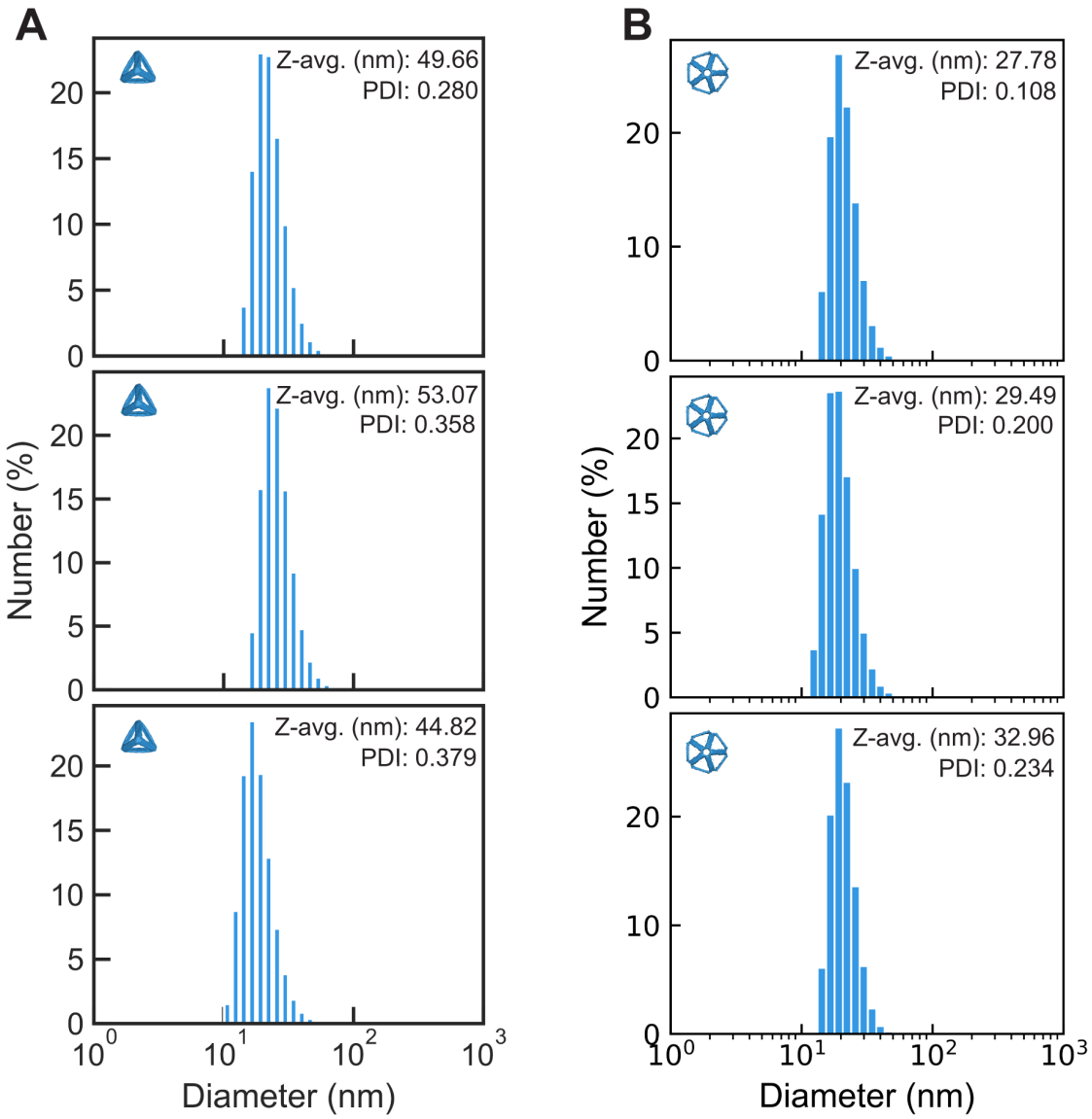

Figure B-4: **Dynamic light scattering showing monodispersity** for **(A)** scaffold asymmetry-designed RNA-scaffolded tetrahedron with 66 bp per edge and **(B)** scaffold asymmetry-designed RNA-scaffolded pentagonal bipyramid with 66 bp per edge.

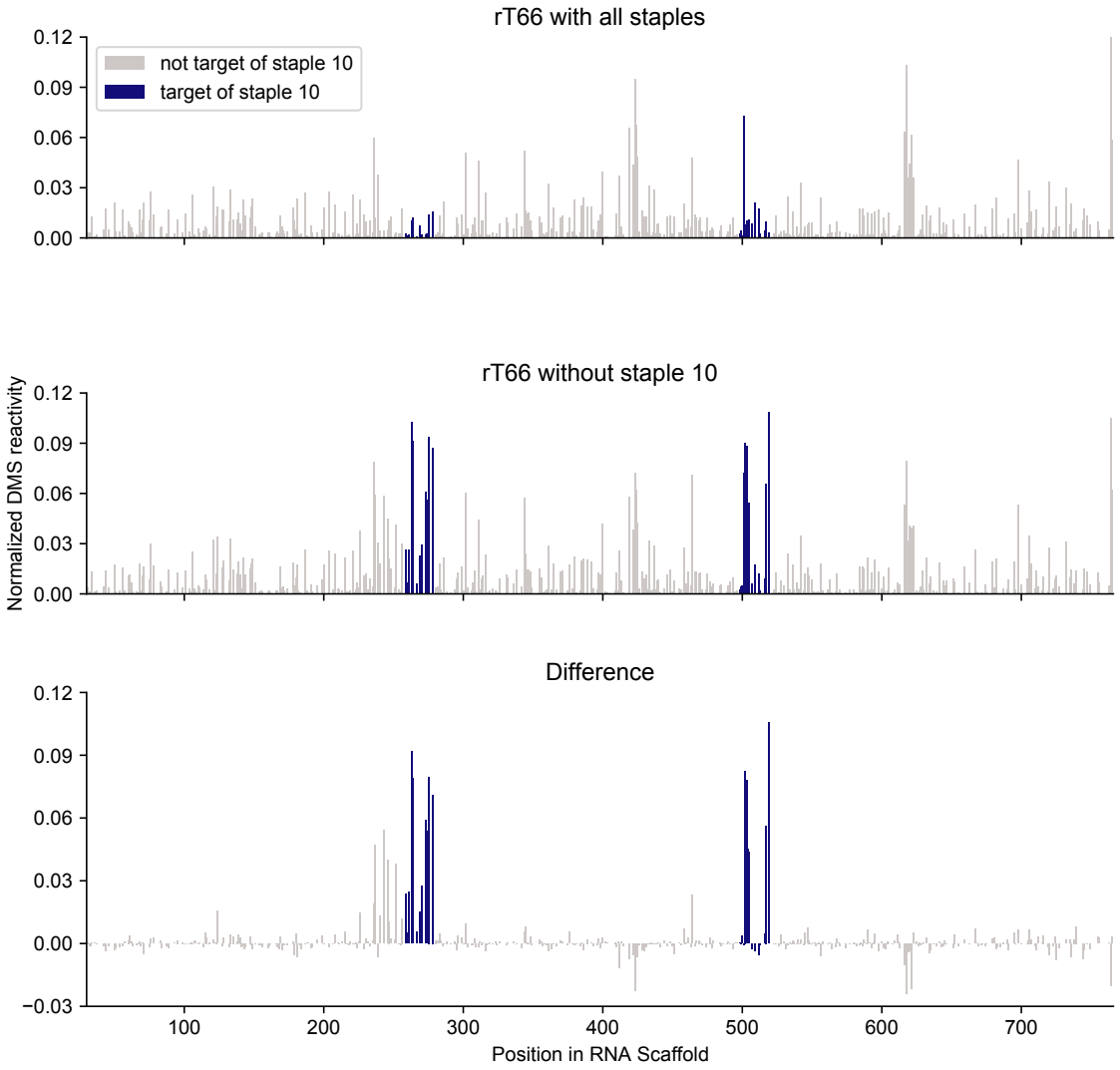

Figure B-5: **DMS-MaPseq mutational profiles for the staple-asymmetry-designed rT66 scaffolded with EGFP mRNA, (A)** with and **(B)** without staple 10, whose targeted nucleotides are shown in blue.

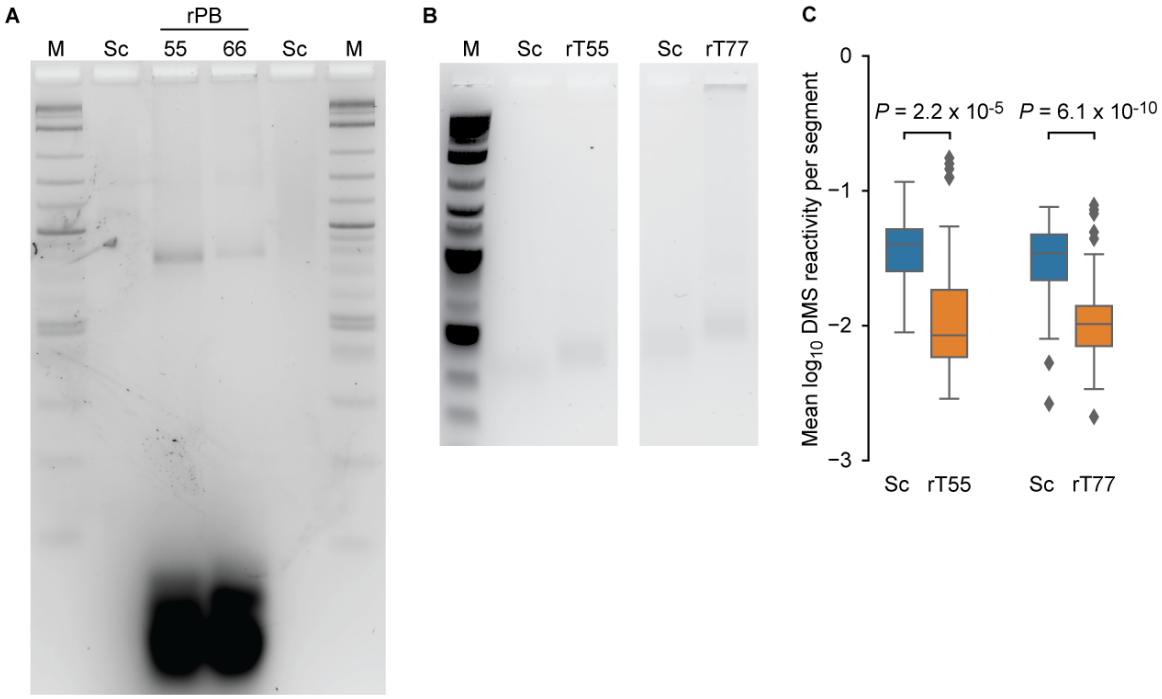

Figure B-6: **Gel mobility shift assays for RNA-scaffolded origami with odd edge lengths**, all with the staple asymmetry design. **(A)** Pentagonal bipyramid with five helical turns per edge (rPB55), with 23s rRNA fragment scaffold. This scaffold often forms a smear in agarose gels with salt, likely due to the formation of secondary structure, but it shows a single band on denaturing PAGE gels. The pentagonal bipyramid with six helical turns per edge (rPB66) folded with the same scaffold is included in the gel for comparison. **(B)** (left) Tetrahedron with five helical turns per edge, with $rsc1218v1\_T55$ synthetic RNA fragment scaffold and (right) tetrahedron with seven helical turns per edge, with $rsc1218v1\_T77$ synthetic RNA fragment scaffold. **(C)** Box plots of normalized DMS reactivity per double-helical segment for the two odd-edge-length tetrahedra and their scaffolds.

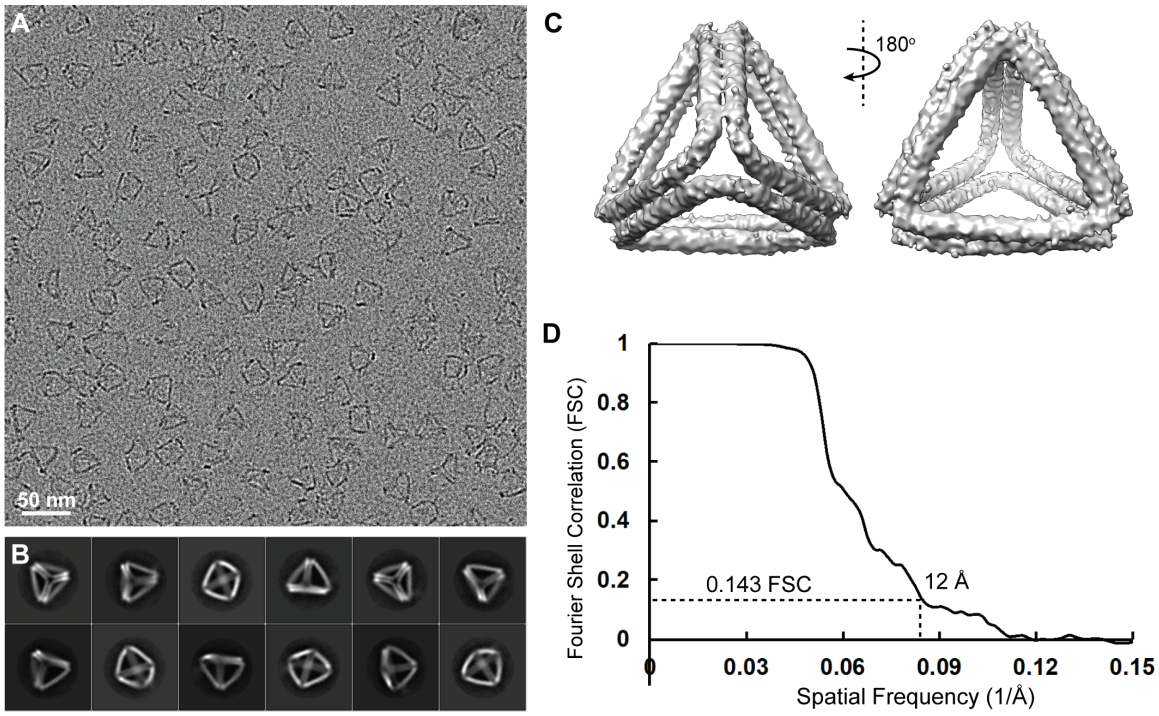

Figure B-7: **Cryo-EM reconstruction for the staple-asymmetry-designed A-form tetrahedron with EGFP mRNA scaffold.** **(A)** Representative micrograph. **(B)** 2D class averages. **(C)** two views of the reconstruction. **(D)** Fourier shell correlation plot; the resolution of the reconstruction is 12Å.

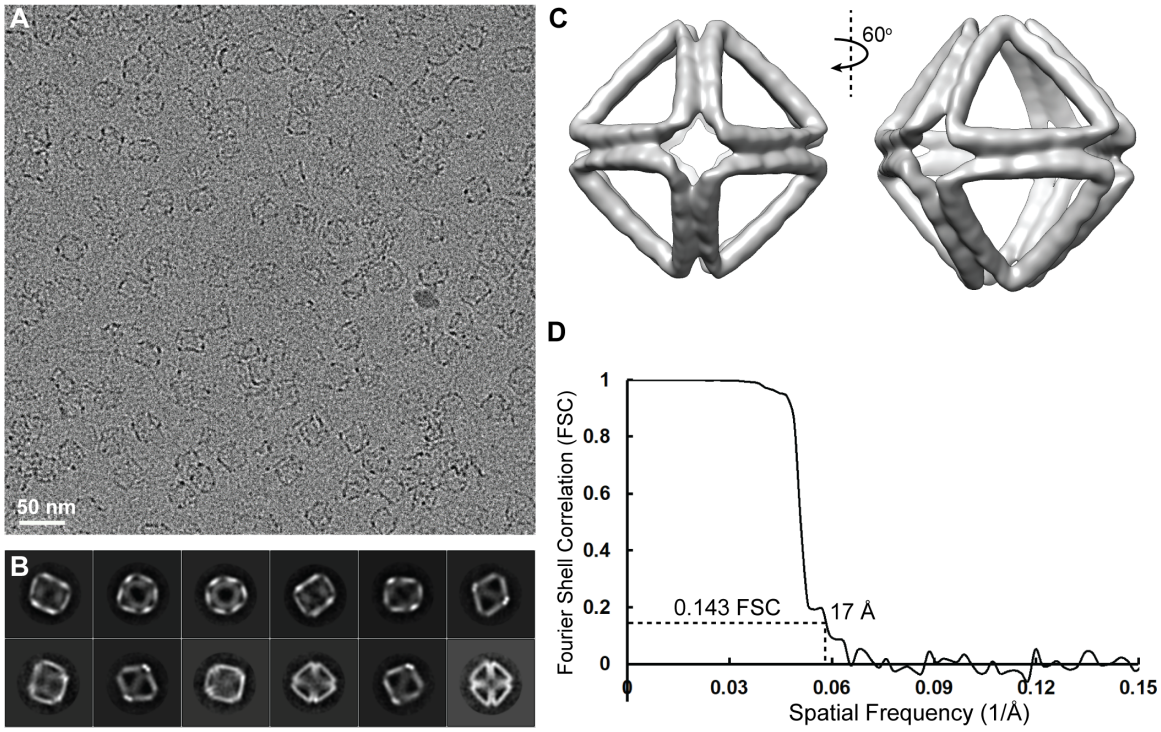

Figure B-8: **Cryo-EM reconstruction for the staple-asymmetry-designed A-form octahedron with four helical turns per edge with M13 transcript scaffold.** **(A)** Representative micrograph. **(B)** 2D class averages. **(C)** two views of the reconstruction. **(D)** Fourier shell correlation plot; the resolution of the reconstruction is 17Å.

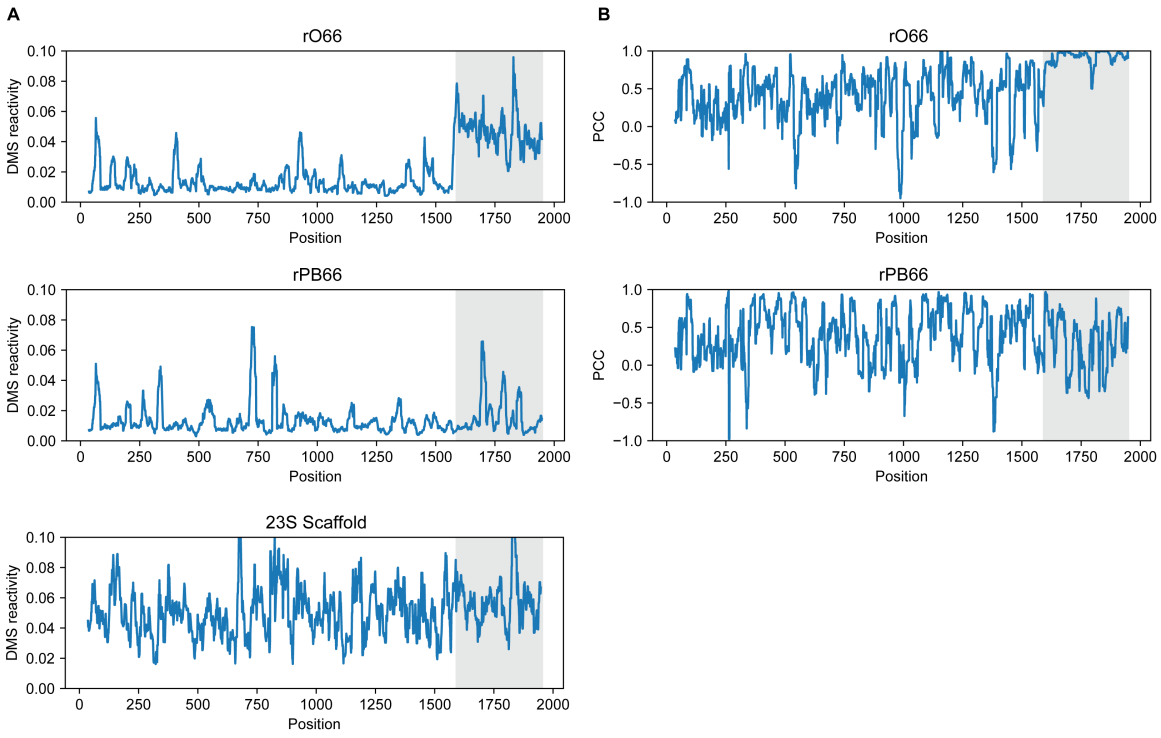

Figure B-9: **DMS-MaPseq reactivity profiles for the staple-asymmetry-designed rO66 and rPB66 scaffolded with 23s rRNA. (A)** DMS reactivity by nucleotide position in the rO66, the rPB66, and the scaffold without staples. **(B)** Pearson correlation of the DMS reactivity in the folded origami with that of the scaffold without staples.

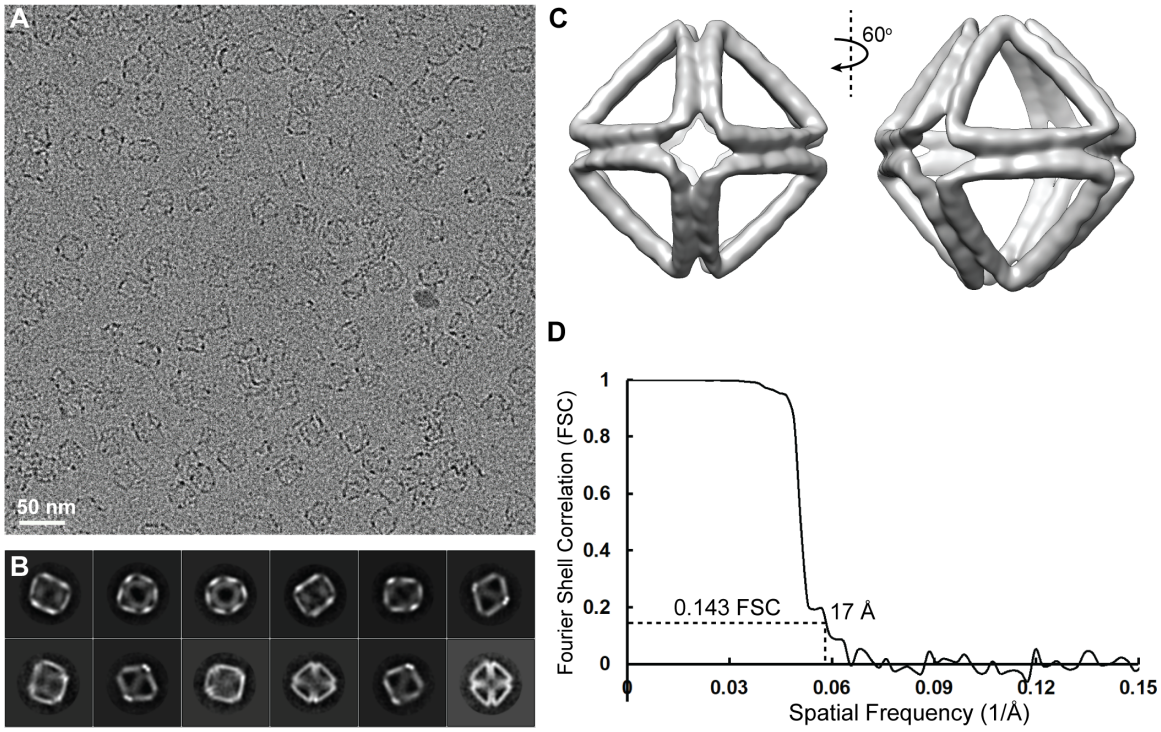

Figure B-10: **Cryo-EM reconstruction for the staple-asymmetry-designed A-form octahedron with six helical turns per edge and 23s rRNA fragment scaffold.** **(A)** Representative micrograph. **(B)** 2D class averages. **(C)** two views of the reconstruction. **(D)** Fourier shell correlation plot; the resolution of the reconstruction is 13Å.

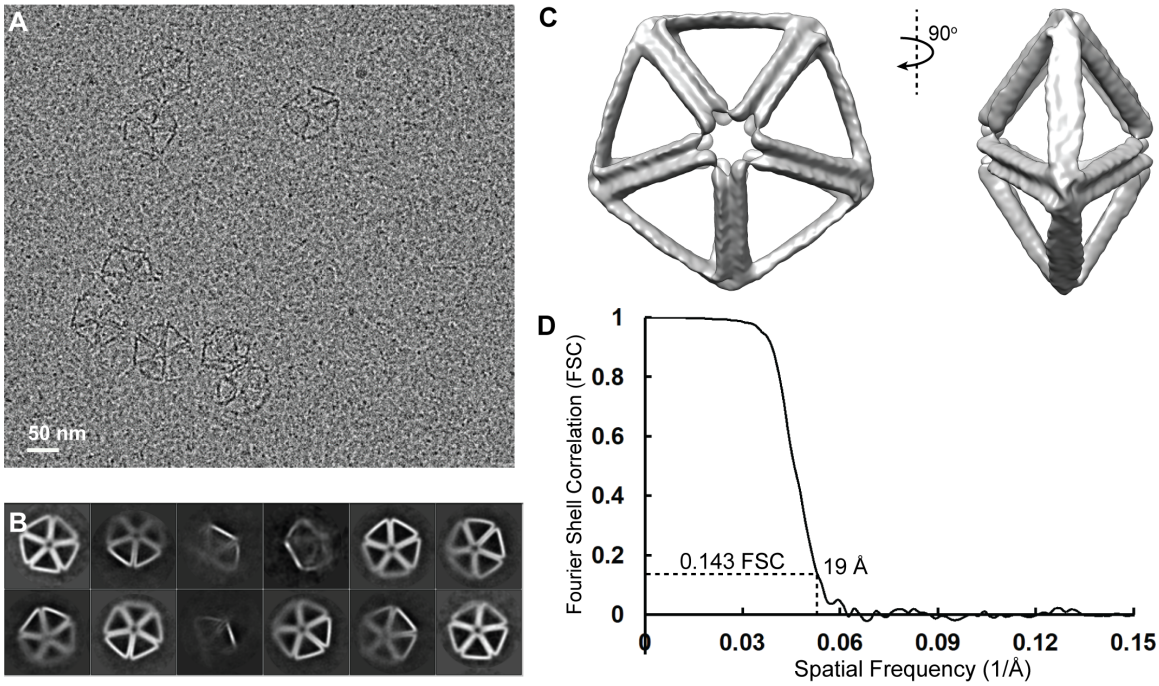

Figure B-11: **Cryo-EM reconstruction for the staple-asymmetry-designed A-form pentagonal bipyramid with six helical turns per edge and 23s rRNA fragment scaffold.** **(A)** Representative micrograph. **(B)** 2D class averages. **(C)** two views of the reconstruction. **(D)** Fourier shell correlation plot; the resolution of the reconstruction is 19Å.

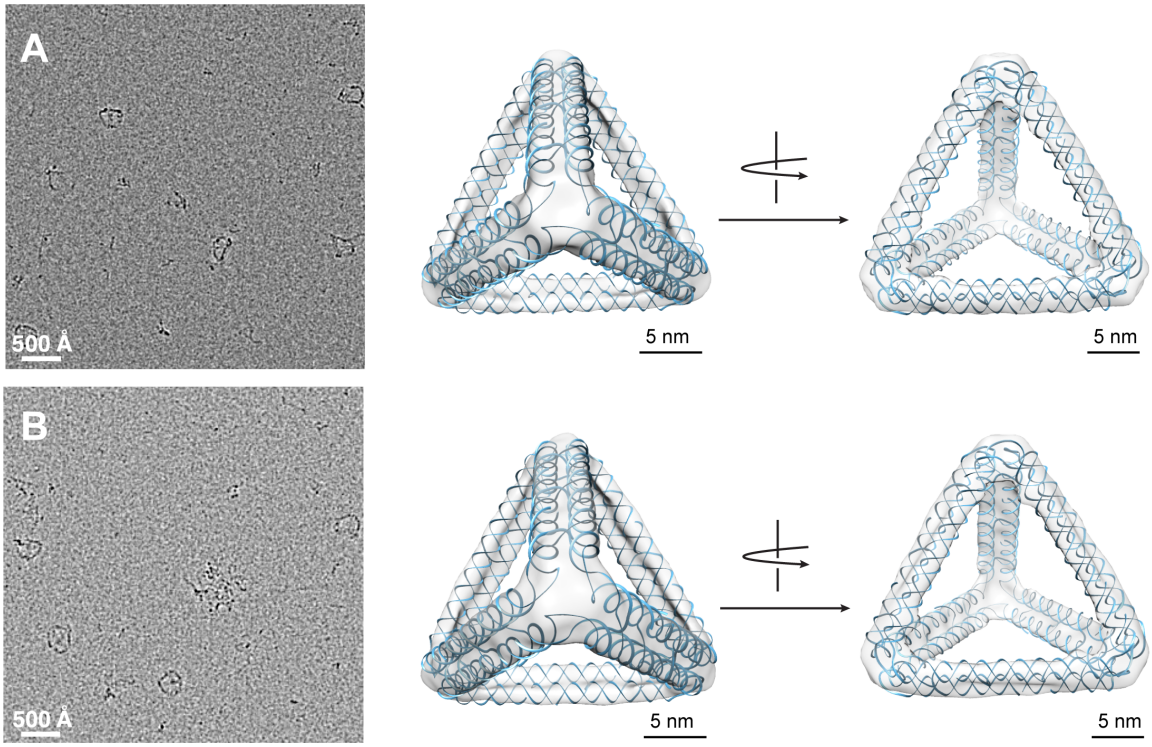

Figure B-12: **Cryo-EM comparison of RNA-scaffolded tetrahedra with a 66-bp edge lengths folded using (A)** alternative A-form geometry staples or **(B)** Hybrid-form geometry staples.

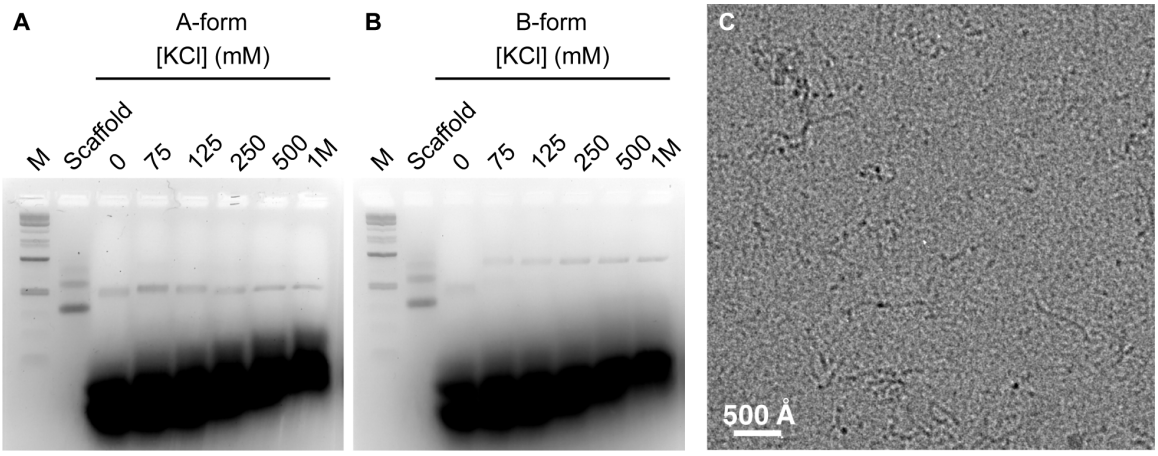

Figure B-13: **Characterization of B-form staple designs from DAEDALUS used to fold RNA-scaffolded origami. (A)** KCl titration of alternative A-form staple designs with RNA scaffolding **(B)** KCl titration of B-form staple designs with RNA scaffolding with the band being noticeably higher. **(C)** Cryo-electron microscopy micrograph showing the unfolded structure of the B-form-stapled origami.

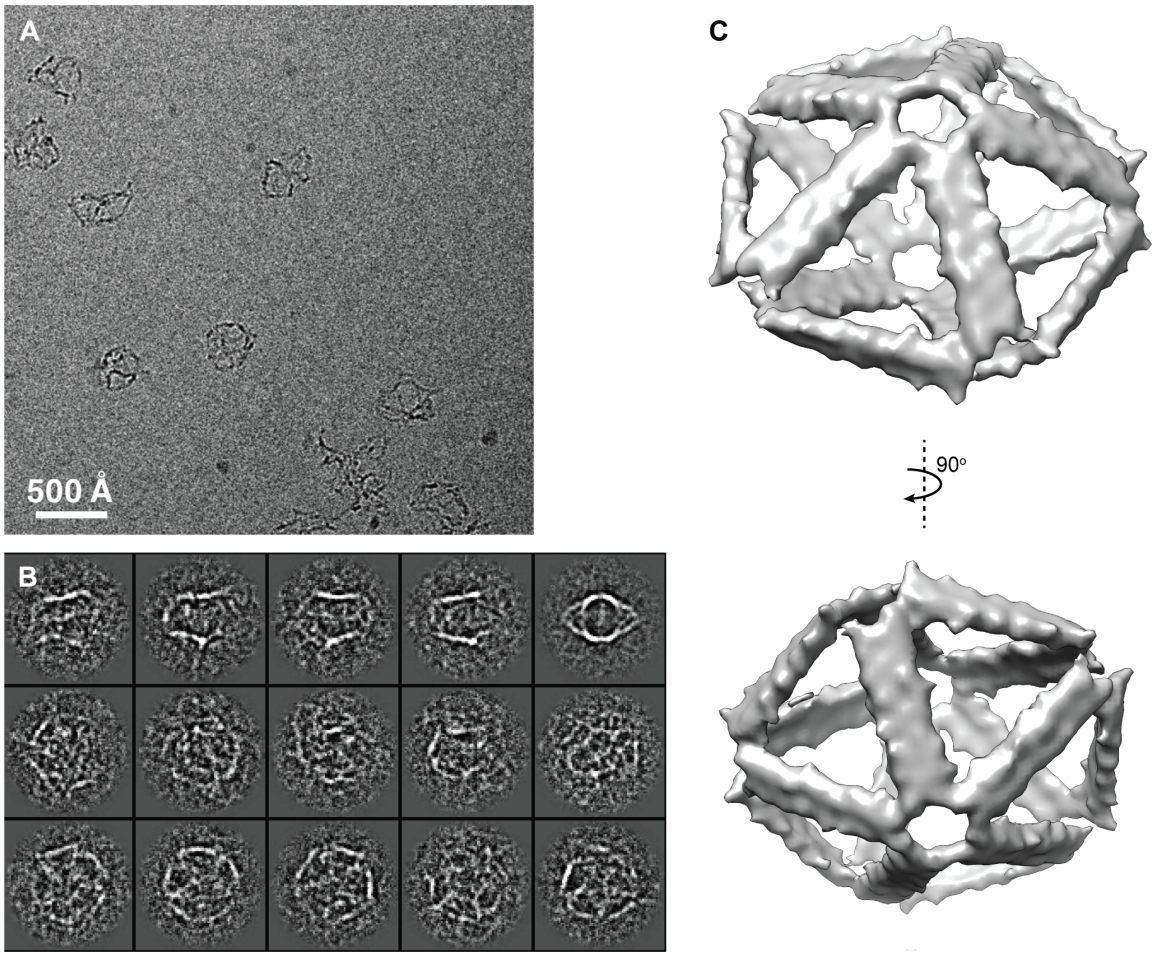

Figure B-14: **Cryo-EM micrograph of RNA-scaffolded pentagonal bipyramid with 66-bp edge lengths folded using scaffold asymmetry design A-form geometry staples.** **(A)** Representative micrograph. **(B)** 2D class averages. **(C)** two views of the 3D reconstruction.

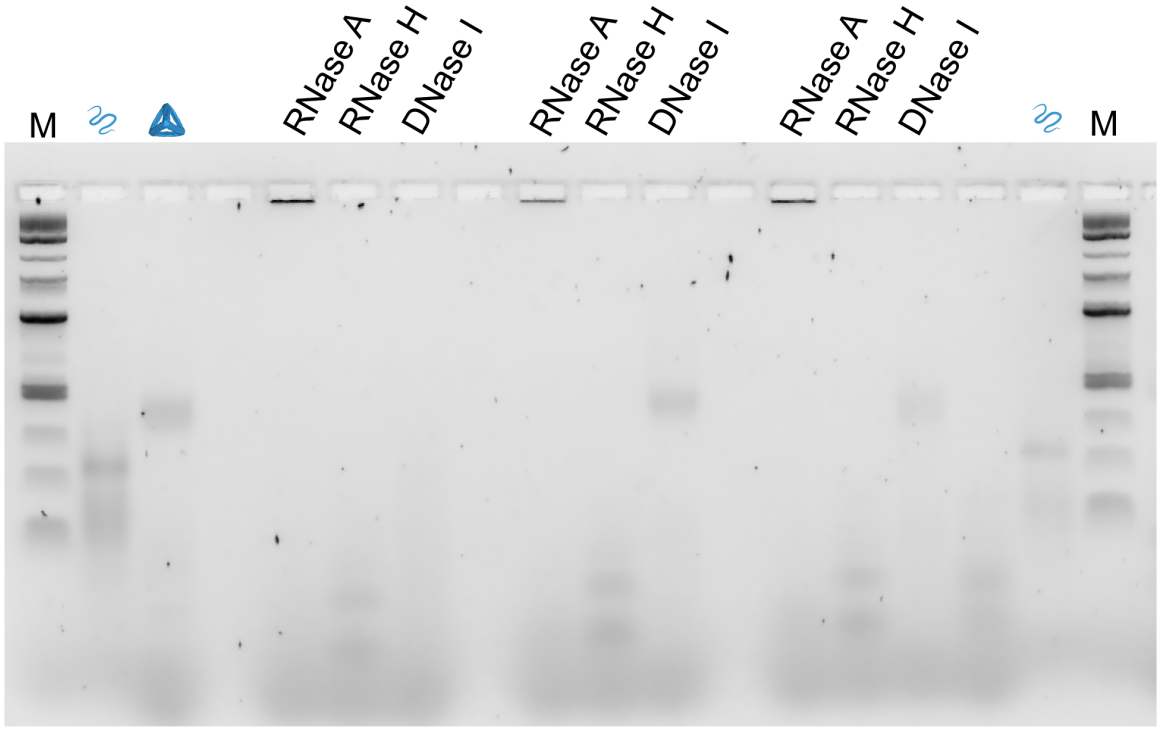

Figure B-15: **Triplicate characterization of biochemical stability of the scaffold-asymmetry-designed EGFP mRNA-scaffolded tetrahedron with 66-bp edge length.** The RNase A-, RNase H-, and DNase I-labeled lanes all represent the folded rT66 treated with the respective nuclease for 5 min at 37^{o}C. Two of the three DNase I replicates show the intact folded origami band.

Appendix C

Thermodynamics

C.1 A brief introduction to statistical mechanics

Employing thermodynamic models can help us predict and interpret interactions between origami with baits and the target RNA.

One approach to thermodynamic modeling is at the microscale with statistical mechanics. The Boltzmann distribution (equations C.1 and C.2) can be applied to describe systems at equilibrium, with the basic principal that the lower the free energy of a particular state, the more favorable it is, and the more likely a molecule will adopt that state [191].

$$p_s = \frac{\exp(-G_s/kT)}{Q} \tag{C.1}$$

$$Q = \sum_s \exp(-G_s/kT) \tag{C.2}$$

Where s is a state of the system, k is Boltzmann's constant, T is the temperature of the system, and G_s is the free energy of state s. Q is referred to as the partition function, representing the sum of energies of all possible states of the system. p_s describes the probability of a molecule adopting state s, and at equilibrium is equal to the proportion of molecules in that state.

C.2 Modeling RNA capture by origami

We describe the system of RNA and origami within the statistical mechanical framework as a set of several possible states:

- bound at all sites
- partially bound at a particular subset of sites (multiple different states fall under this category)
- fully unbound

Most types of experiments, including the co-localization assay employed in Chapter 3, cannot distinguish between fully and partially bound in their measurements. However, we still incorporate the multiple ways to be bound into our model, because it does affect the final distribution between fully unbound and bound.

For a system with n binding sites, each state s is described by the set of microstates s_i for each binding site i, which can be either 0 (unbound) or 1 (bound).

$$s = [s_1, s_2 ... s_n] \tag{C.3}$$

For example, in a system with three binding sites like our tetrahedron designed to capture tRNA, the state in which only the first and third sites are bound to RNA is described as $s = [1, 0, 1]$.

In the simplest model, we use the fully unbound state as a reference point ($G_unbound = 0$), and the energy G_s of each state is only determined by the chemical potentials μ_i of each site i that is bound in that state. For example, in the state where only sites 1 and 2 are bound, $G_s = \mu_1 + \mu_2$. This model assumes that the binding state of one site has no effect on the energy of binding for another site. In other words, this would be a completely non-cooperative system.

However, such an assumption is unrealistic to the baited capture RNA and origami system, which may be either positively or negatively cooperative. With one loop of RNA bound to a single bait, binding to a second site could be much more favorable

due to avidity if the second bait is accessible to the second RNA loop (positive cooperativity). Avidity can be thought about from a macroscale perspective as a locally increased effective concentration of binding partners, because the second bait and RNA loop are already forced in close proximity, and the volume accessible to the RNA loop is much smaller than the solution as a whole. From another, more microscale perspective, much of the entropic cost of binding–going from two molecules to one–is already paid with the first site bound, making the energy of binding a second site lower than binding the single site alone.

If, on the other hand, the second site is at a particularly non ideal distance, either so far away that the RNA cannot possibly bind to the two baits at once or simply positioned such that RNA has to alter conformation somewhat to hybridize to both, the binding would be negatively cooperative. Even though hybridization with the second bait would release enthalpy, enthalpy would be required to break and/or rearrange the internal hybridization in the RNA, thus resulting in a less favorable binding to the second bait than its hybridization energy alone would imply.

C.2.1 Accounting for avidity and deformation

A model for the thermodynamics of RNA baited capture thus needs to account for possible cooperativity. One favoured method involves adjusting individual binding affinities to account for the local effective concentration of additional binding sites when one or more sites is already bound [191]. In statistical mechanics, this local effective concentration adjustment would be incorporated into the chemical potential μ_i for site i, which depends on the activity and therefore concentration of binding partner, as in equation C.4 [192].

$$\mu_i = kT \ln(aK_i) \tag{C.4}$$

where K_i is the individual association constant for site i and a is the activity (related to concentration) of the ligand (i.e. RNA binding loop).

To determine the local effective concentration, we must determine the volume

accessible by the ligand/RNA binding loop and the number of binding sites present within that volume (in our system of heterogeneous binding sites, at most one binding site for that loop will be present in the volume). However, this calculation is complicated for our system. The accessible volume for a second binding loop would probably be best described by a shell centered at the first binding loop, but accurately determining the minimum and maximum radius to delineate the shell is difficult for RNA with unknown structure and unknown dynamic range–the typical target of our baited capture for studying structural biology. We could say the maximum radius accessible to the second loop is the maximum distance between the second loop sequence and the first loop sequence if the RNA were completely stretched out with no internal hybridization, and the minimum radius is 0, assuming sufficient flexibility (sequence distance between loops sufficiently greater than the persistence length of RNA) to bring the second and first loops together. These limits, however, do not likely describe the actually folded RNA we are targeting. And describing accessible volume gets still more complicated for a third loop and beyond, when the RNA is fixed at two or more points but with still some dynamicism and an indeterminate and/or mobile centre point for the shell accessible by the additional loop.

This approach to modeling the system, describing all possible parameters from a physical basis, is useful if the goal of the model is to predict binding affinity and outcomes for the system, and to possibly compare the prediction with experimental values. A different use for modeling is to fit experimental data in order to draw further conclusions about the system. For the baited capture approach to studying RNA tertiary structure, we want to compare bait arrangements to determine the closest geometric match to our target RNA–identifying 'hits' that not only bind stably but do so without deforming the native RNA structure. We can employ thermodynamic modeling to aid us in this comparison, without needing to know a priori a complete description of the system.

Inspired by Borkovec et al. [192], I realized we could incorporate pairwise site 'interaction energies' into a statistical mechanical model of baited capture for more relevant comparisons between bait arrangements on origami. In this model, the energy

of a state s is determined not only by chemical potentials mu_i of individual bound sites, but also by the interaction energies E_{ij} between each pair of sites i and j that are both bound in that state (equation C.5) [192].

$$G_s = \sum_i \mu_i s_i + \sum_{ij} E_{ij} s_i s_j \tag{C.5}$$

Note that an advantage to writing the equation as above is an easier computational implementation [192]; the chemical potential for site i is only included if the site is actually bound and $s_i = 1$, and likewise, the interaction energy E_{ij} is only included if both sites i and j are bound and $s_i = s_j = 1$. If site i is unbound, $s_i = 0$ and thus naturally removes the μ_i and E_{ij} terms from the equation, therefore we do not require a separate energy equation for each state or a different set of energy equations for systems with different numbers of baits.

With this definition for state energy incorporated into the general statistical mechanical model described in section C.1, and combined with equation C.4, we can fit experimental values of the fraction bound for known free RNA concentration and known individual site association constants to estimate interaction energies between each pair of sites. Evaluating interaction energies, in addition to overall affinities, will allow us to compare capture designs that not only have different bait arrangements, but different bait sequences or varying numbers of baits included.

If all values of the interaction energy matrix E are 0, the binding is non-cooperative. An estimated interaction energy $E_{ij} >>> 0$ indicates that binding both sites i and j is highly unfavourable (negative cooperativity), and likely the two sites are either much too far apart or much too close to be bound simultaneously by the RNA. $E_{ij} <<< 0$ indicates positive cooperativity between the two sites, suggesting that sites i and j are spaced at the same distance as their targets on the RNA, such that avidity makes binding both simultaneously much more favorable than the sum of individual site affinities would suggest. An intermediate interaction energy, $E_{ij} < 0$ or $E_{ij} > 0$, suggests only moderate positive or negative cooperativity in binding the two sites–the RNA can only simultaneously bind i and j sites if it contorts away from its native,

minimum-free-energy structure.

We can thus use estimated interaction energies to identify a "good fit" of origami presenting baits to the target RNA.

C.2.2 Algorithm to estimate interaction energies for the baited capture model

The overall steps in the algorithmic implementation of the statistical mechanical model with interaction energies are as follows:

1. Provide the experimental data

    - NP_0, scalar, the total input concentration of origami (nanoparticle) [M]
    - R_0, 1-d array, total input concentrations of RNA [M]
    - $R_{bound,obs}$, 1-d array, the experimentally measured concentrations of bound RNA at equilibrium [M]
    - $R_{unbound,obs}$, 1-d array, the experimentally measured concentrations of unbound RNA at equilibrium [M]. This could alternatively be calculated as $R_0 - R_{bound,obs}$
    - $NP_{bound,obs}$, 1-d array, the experimentally measured concentration of bound origami [M]. From the co-localization experiment, I assumed one-to-one binding of RNA to origami, i.e. $NP_{bound,obs} = R_{bound,obs}$
    - $f_{obs} = NP_{bound,obs}/NP_0$, 1-d array, the fraction of origami bound to RNA [unitless]

2. Define the system and related constants (temperature, number of binding sites, individual binding site association constants, possible system states, and the Boltzmann constant k in appropriate units)

3. Define the model

    - G to calculate the energy of a state s given RNA concentration, individual site binding affinities, and the interaction energy matrix

- Q to calculate the semi-grand canonical ensemble partition function for the system (relies on G)
- p to calculate the probability of a particular state s (relies on G and Q)
- f to determine the predicted fraction of origami bound for the system with a particular RNA concentration (relies on G, Q and p)

4. Solve for the interaction energy matrix that minimizes error between the observed fraction bound and that predicted by the model

Sample code written in Python 3 is presented below, for the system of a tetrahedron with 3 overhanging baits (8 nt each) capturing an engineered tRNA. Individual binding affinities for each bait were estimated using a nearest neighbor model [193]. The experimental data included is from the co-localization assay in Chapter 3.

```
import numpy as np
from scipy.optimize import curve_fit
import itertools as it

#%% Input experimental data

NP_0 = np.array([50*10**-9]*11) # M | total concentration of NP
R_0 = np.array([2.5*10**-9,7.5*10**-9,10*10**-9,25*10**-9,50*10**-9,
    100*10**-9,150*10**-9,200*10**-9,250*10**-9,350*10**-9,
                1000*10**-9]) # M | total concentration of RNA

Rbound_obs = np.array
    ([0,6.527198978*10**-9,2.832178703*10**-9,9.452187896*10**-9,
    15.03971061*10**-9,26.04244856*10**-9,24.59785865*10**-9,
    34.92603804*10**-9,29.96802724*10**-9,25.24992105*10**-9,
                        36.36133607*10**-9]) # M | concentration of
    bound RNA (experimentally observed)
Rub_obs = R_0 - Rbound_obs # M | concentration of free RNA (
    experimentally observed)
```

```
17
18 NPbound_obs = Rbound_obs # M | concentration of bound NP (
       experimentally observed)
19
20 f_obs = NPbound_obs/NP_0 # fraction of NP bound to RNA (
       experimentally observed)
21
22 #%% Define constants
23 nsites = 3 # number of binding sites on NP with complements on the
       RNA
24
25 k = 1.38064852*10**-23 # J/K | Boltzmann constant
26 T = 300 # K | Temperature
27
28 #%% Define system inputs
29
30 # System states (list of all microstates, i.e. possible combinations
       of sites bound)
31 system = list(it.product([0,1], repeat=nsites))
32
33
34 # microscopic association constants for each site
35 dH = [-51.33, -45.93, -54.34] # kcal/mol; estimated from
       Tm_estimator (NN model)
36 dS = [-143.91/1000, -129.3/1000, -156.97/1000] # kcal/mol-K
37 T = 300 # K;
38 R = 1.9872036*(10**(-3)) # (kcal/mol-K) #ideal gas constant
39 Ki = [np.exp(-(dH[site] - T*dS[site])/R/T) for site in range(3)] #
       association constants
40
41 #%% Define model
42
43 # Free energy of a microstate
44 def G(a,Ki,E,s):
45     # a is scalar | activity of RNA
46     # Ki is list of length nsites | microscopic association
```

```
    constants for each site
47      # E is a matrix sized nsites^2 | pairwise site interaction
    energies
48      # s is a list of length nsites | binary binding state (0 or 1)
    of each site
49
50      # format E as a full, symmetric matrix (enforced within the
    function so we can
51      # reduce the fitting space, because there are redundant
    parameters within the matrix)
52
53      # create a square matrix pre-populated with zeros -- interaction
     energies between a
54      # site and itself is always zero (since a site can only be bound
     by one target at a time)
55      # so all the diagonals remain zero
56      E_sym = np.zeros([len(E),len(E)])
57
58      for i,row in enumerate(E):
59          for j,elem in enumerate(row):
60              if j>i: # only retrieving the values to the top right of
     the diagonal
61                  E_sym[i][j] = elem
62                  E_sym[j][i] = elem
63
64      mu = k*T*np.log(a*np.array(Ki)) # array of site chemical
    potentials
65
66      G_si = -sum(mu[i]*state for i,state in enumerate(s)) + 0.5*sum(
    E_sym[i][j]*statei*statej
67                for i,statei in enumerate(s) for j,statej in enumerate
    (s))
68
69      return G_si
70
71  # Semi-grand canonical ensemble partition function
```

```
72 def Q(a,Ki,E,system):
73     Q_system = sum([np.exp(-G(a,Ki,E,s)/k/T) for s in system])
74     return Q_system
75 
76 # Probability of each microstate
77 def p(a,Ki,E,s,system):
78     Q_sys = Q(a,Ki,E,system)
79     p_si = np.exp(-G(a,Ki,E,s)/k/T)/Q_sys
80     return p_si
81 
82 # fraction bound of NP
83 def f(a,Ki,E,system,nsites):
84     s_unbound = np.zeros(nsites) # microstate with nothing bound (
    state = 0) at all sites
85     Q_sys = Q(a,Ki,E,system)
86     G_unbound = G(a,Ki,E,s_unbound)
87     f_pred = (Q_sys - np.exp(-G_unbound/k/T))/Q_sys
88     # "bound" state is all microstates except case where nothing is
    bound at any site
89     return f_pred
90 
91 #%% Solve for Ei that minimize error between predicted and observed
    fraction NP bound
92 
93 E0 = np.array([[0,0,0],
94                [0,0,0],
95                [0,0,0]]) # initial guess for interaction energies
96 
97 def f_fit(a,*E):
98     E = np.array(E).reshape((nsites,nsites))
99     f_pred = [f(conc, Ki, E, system, nsites) for conc in a]
100 
101     return f_pred
102 
103 E_calc = curve_fit(f_fit,Rub_obs,f_obs,p0=E0.flatten())
```

Appendix D

Nucleic acid sequences used in this work

D.1 DAEDALUSv2

Table D.1: Scaffold sequences used for folding: 2,520-nt (#phPB84), 3,120-nt (#phIco52) and 1,676-nt (#pF1a) lengths.

Scaf. name	**Scaffold sequence**
#phPB84	GAGCGCAACGCAATTAATGTGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCG CAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCG CCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCT TTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGAT AGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGA ACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTG GTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTACAACCGGGGTACATATGATTGGGGT CTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTT CACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGT CTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATA GTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTG CAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAG GGCCGAGCGCAtAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAG CTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGT GTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGA TCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGG CCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGA TGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTT GCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATT GGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAAC CCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACA GGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCC TTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATT TAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAA CCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCGAATTCGTCGT CGTCCCCTCAAACTCTTGGGTGGAGAGGCTATTCGTTTAAGGTCACATCGCATGTAATTTACTTATTC TCTGTTGTTGAGCCACCCGGGCGCCAGATTTTGTTTAAAGCTTTGTCTCTTAGTTTGTATAGACAGAT TCAGAGTGCAAGGTTTCGTTCGCTCGTACCTGGTTTTCCCTGGTTCTTCACAGATAGGATTTGACTTT CTACAACACTTATGCGGCTTCCTACCCGTTTGAAGGCCGATACAGGTGCTGCGCAAAATGCGGGCGA ACATAGAGTATCAAAACAACGCCTTCTAATCTAGGAATATAGGGAAGATACGTATTTGCTACCATGCT TTCTTGGGTCATTAACGACCAACCTCTTTTCTTTTAAAGTAGGATTGCACAATGAATGAATACACGTG GTCCGATAACTGACCAAGTAACATGGTTATCACTaGATGTCCGCCAGACGTGTGCAAACCAACCCGGG AGTTACGTCACTAATCCTTCGCTACGTCGTGAAGATATTTACTTGTGAATATCGAGGGTAATAAGAT AATAGACTGTGACTAGTATTGCCAGACTGTCGCTACCTGCAACACATAACTATCCTGAGGTTACTGCA TAGTACTGATTACACCCGAGTCAAAATTTCTAACTTCTAACATGTACCTAGTAACCAGCTCAATAATT ATGTCAGAATATAGCTCTGGGAACCCTCGGACAATTATGATACACGGTATTAATATCTTGCTTGCGTT AGCCACTTCTCATCTTTGGATACCGATTCTATTTTGCATAGCAGTTCCTTTTACACATATAAGAATTT CGCCATAGGTATGCTGCAG

#phIco52

AGAATTTCGCCATAGGTATGCTTAAGGAAGTCGAGATTGCGAACCATTACCGAGACTATGGCTTCAT
GTGGTGATTTCACCCGACCCACCCTTGGCGCCAGCTTTACGCAGCTTCCTGACGATACGTGGTGTAAC
GTTGTGTTTGGCAATGGAAACCGAGATCAACTATTTCTAATGCTGATATAGCAGAGTCTCGCGTCTA
TCATACGCAAGTCGCACGTCATTTTCGAGAGCAGCGTAAGACTCTGAAGGTCATGAGCCCAGATGTT
ATTACCCTCTACCTATAAACATCAAAATTGTAGTCGTTTTACAGTCCATCGTCGCTCCAGAGCGAAGA
TTAAGGTTAGATCTAGATTATCTTTGCACGTGTGGACCGACGCAGCTGGGGCTCTAGCTCCACTACG
GTTACGAAACTGCTGAACGATCTGGTCCACTTCAAGATTCACACATCGTTTCATTCTTTGGACAACCA
ACACTCTCAGTCAGAGTTTCGAGTATAATAATTCTTCCGCGCTAGGGTAAAAAGCAGATATGGGGAG
ACATTCCGGGCTTTTGAGCCGATACACTAAGCACTTGACATACTCACATCAGTAGAGGTTAACATTCA
TGACTATCACGCGCTGCAGGAGCGCAACGCAATTAATGTGCGCCCTGTAGCGGCGCATTAAGCGCGG
CGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGC
TTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTT
TAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACG
TAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGT
GGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGAT
TTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTACAACC
GGGGTACATATGATTGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATG
AGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAG
TATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCT
GTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTA
CCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAAT
AAACCAGCCAGCCGGAAGGGCCGAGCGCATAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCT
ATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCA
TTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACG
ATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATC
GTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTA
CTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAG
TGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAA
CTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTG
AGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGT
TTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGT
TGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGG
ATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGC
CACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCC
TTTCGTCGAATTCGTCGTCGTCCCCTCAAACTCTTGGGTGGAGAGGCTATTCGTTTAAGGTCACATCG
CATGTAATTTACTTATTCTCTGTTGTTGAGCCACCCGGGCGCCAGATTTTGTTTAAAGCTTTGTCTCT
TAGTTTGTATAGACAGATTCAGAGTGCAAGGTTTCGTTCGCTCGTACCTGGTTTTCCCTGGTTCTTCA
CAGATAGGATTTGACTTTCTACAACACTTATGCGGCTTCCTACCCGTTTGAAGGCCGATACAGGTGCT
GCGCAAAATGCGGGCGAACATAGAGTATCAAAACAACGCCTTCTAATCTAGGAATATAGGGAAGATA
CGTATTTGCTACCATGCTTTCTTGGGTCATTAACGACCAACCTCTTTTCTTTTAAAGTAGGATTGCAC
AATGAATGAATACACGTGGTCCGATAACTGACCAAGTAACATGGTTATCACTAGATGTCCGCCAGAC
GTGTGCAAACCAACCCGGGAGTTACGTCACTAATCCTTCGCTACGTCGTGAAGATATTTACTTGTGAA
TATCGAGGGTAATAAGATAATAGACTGTGACTAGTATTGCCAGACTGTCGCTACCTGCAACACATAA
CTATCCTGAGGTTACTGCATAGTACTGATTACACCCGAGTCAAAATTTCTAACTTCTAACATGTACCT
AGTAACCAGCTCAATAATTATGTCAGAATATAGCTCTGGGAACCCTCGGACAATTATGATACACGGT
ATTAATATCTTGCTTGCGTTAGCCACTTCTCATCTTTGGATACCGATTCTATTTTGCATAGCAGTTCC
TTTTACACATATA

#pF1a	GAGCGCAACGCAATTAATGTGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCG CAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCG CCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCT TTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGAT AGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGA ACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTG GTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTACAACCGGGGTACATATGATTGGGGT CTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTT CACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGT CTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATA GTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTG CAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAG GGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAA GCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGG TGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATG ATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTG GCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAG ATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGT TGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCAT TGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAA CCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAAC AGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTC CTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTAT TTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAA ACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTC

Table D.2: Staple sequences used to fold the pentagonal pyramid of 84-bp edge length.

Staple ID	**Length (bp)**	**Staple sequences**
1	59	TCGCCCGCATTTTGCGCCGACGACGAATTCGACCAAGAGTTTGAGGGGAAGCACCTGTA
2	59	CACTGATTAAGCATTGGAATAGACCGAGATAGGCCTTATAAATCAAAAGTAACTGTCAG
3	59	TGGTTTCTTAGACGTCAATTGAGCTGGTTACTAATATTCTGACATAATTGGTGGCACTT
4	59	CGTAGTTATCTACACGATTATTCCCTTTTTTGCCATTTCCGTGTCGCCCCGGGGAGTCA
5	59	GTAACCTCAGGATAGTTCATGGTAGCAAATACGTAATGACCCAAGAAAGATGTGTTGCA
6	59	AACGACGAGCGTGACACCTATTCTCAGAATGACTCGGTCGCCGCATACACACGATGCCT
7	59	GAAGAACGTTTTCCAATAGCCGGTGAGCGTGGGTTGCTGATAAATCTGGGATGAGCACT
8	59	ACATTCAAATATGTATCCTATCAGGGCGATGGCAAGGGCGAAAAACCGTCGCTCATGAG
9	59	AAGCACTAAATCGGAACTATCCAAAGATGAGAATGCAAAATAGAATCGGCCTAAAGGGA
10	54	GACTCGTTTTTGGTGTAATCAGCAATACTAGTCTTTTTTACAGTCTATTATACC
11	54	AAGTTTTTTTTTTTGGGGTCGATTGACGGGGAATTTTTTAGCCGGCGAAATCAG
12	53	GTAAGATTTTTTCCTTGAGAGTGTGGCGCGGTATTTTTTTATCCCGTAACAGA
13	53	AGCCGCTTTTTATAAGTGTTGTTTACTTCTGACTTTTTAACGATCGGATTCAT
14	53	TAGACATTTTTGATCGCTGAGATATACTTTAGATTTTTTTGATTTAAACAGGA
15	52	CCGGGTTTTTTTGGTTTGCACTGGGAACCGGATTTTTGCTGAATGAACGTTG
16	52	TCAATATTTTTATATTGAAAATGCTCACCCAGTTTTTAAACGCTGGTAAATT
17	52	CCACTTTTTTTTATGCGCTCGGCTGGGGCCAGTTTTTATGGTAAGCCCGAAA
18	52	TGTGCATTTTTTATCCTACTTTTTAGAAGGCGTTTTTTTGTTTTGATTAGGA
19	52	AAAGCATTTTTTTCTTACGGATACTCTGAATCTTTTTTGTCTATACACAGCG
20	52	CGCAAATTTTTCTATTAACTGCCCTTAACGTGTTTTTAGTTTTCGTTGGTCA
21	52	TTATAGTTTTTGTTAATGTCACGGAACCCCTATTTTTTTTGTTTATTATGCT

22	51	TATGGCTTTTTGAAATTCTTATTAATACCGTTTTTTGTATCATAATATTTT
23	51	CTCATTTTTTTTTTTTAACCAAGAACAAGAGTTTTTTCCACTATTAATAATC
24	42	GTTCCAGTTTGTAGGCCGAAATCGGCAAAATCGTTGAGTGTT
25	42	AAGGAAGGGAATTCGCGTTAAATTTTTGTTAACGTGGCGAGA
26	42	GGAGCGGGCGCTCATATGTACCCCGGTTGTAAGAAAGCGAAA
27	42	GCAAGTGTAGCCCACTGAGCGTCAGACCCCAATAGGGCGCTG
28	42	CGTGATACGCCTAAACGAATAGCCTCTCCACCGAAAGGGCCT
29	42	AACCATCACCCAGAACGTGGACTCCAACGTCACCACTACGTG
30	42	CAAGCAAGATATATGTGTAAAAGGAACTGCTAGTGGCTAACG
31	42	TACTCTAGCTTTGATAATCTCATGACCAAAATGCGAACTACT
32	42	ATTAATAGACTGATCTAGGTGAAGATCCTTTTCCCGGCAACA
33	42	GGATAAAGTTGACTTCATTTTTAATTTAAAAGGGATGGAGGC
34	42	ACTCACCAGTCTTGACGCCGGGCAAGAGCAACTTGGTTGAGT
35	42	GATATTCATTGCGTTGCGCTCCTGCAGCATCTTATTACCCTC
36	42	ATCTTCACGACGCCGCTACAGGGCGCACATTAACAAGTAAAT
37	42	ATTAGTGACGTCACACCCGCCGCGCTTAATGCGTAGCGAAGG
38	42	GACATCTAGTGCATGTAACTCGCCTTGATCGTACGTCTGGCG
39	42	TACTTGGTCAGTTTTTGCACAACATGGGGGATATAACCATGT
40	42	ACGTGTATTCAGGACCGAAGGAGCTAACCGCTTTATCGGACC
41	42	TCATTGCAGCACCCTTCCGGCTGGCTGGTTTATCTCGCGGTA
42	42	CTTCCTGTTTTAGGAAGAGTATGAGTATTCAAGGCATTTTGC
43	42	ATCCTATCTGTAGTGATAACACTGCGGCCAACAGAAAGTCAA
44	42	GGAAAACCAGGTGCAGTGCTGCCATAACCATGGAAGAACCAG
45	42	CGAAACCTTGCGGCATGACAGTAAGAGAATTATACGAGCGAA
46	42	CAAAGCTTTAATTACATCGAACTGGATCTCAAAACTAAGAGA
47	42	GCGCCCGGGTGTCAGTTGGGTGCACGAGTGGGACAAAATCTG
48	42	AGAGAATAAGTGAAAGTAAAAGATGCTGAAGAGCTCAACAAC
49	42	AGAAGTTAGAATGTCCGAGGGTTCCCAGAGCTGGTACATGTT
50	42	ATATTCCTAGAAAAAGAAAAGAGGTTGGTCGTTATCTTCCCT

51	26	ACATGCTTTTTGATGTGACCTTATTT
52	26	CGCTGCTTTTTGCGTAACCACAACTC
53	25	ACCAAGTTTACTCATATAGGTGCCT
54	25	ACAATAACCCTGATAATTTCTAAAT
55	25	GCCCCCGATTTAGAGCGGTGCCGTA
56	25	GTAGCAATGGCAACAAGCCATACCA
57	25	TTTAAAGTTCTGCTATTTTCGCCCC
58	25	GGCAACTATGGATGAACTCCCGTAT
59	25	GGTAGCGACAGTCTGGTACTATGCA
60	25	TCGGCCTTCAAACGGGACTCTATGT
61	25	TTCGGGGAAATGTGCGTGATAATAA

Table D.3: Staple sequences used to fold the icosahedron of 42-bp edge length.

Staple ID	**Length (bp)**	**Staple sequences**
1	57	GCATCGTAGCGAAGGTTTTTATTAGTGACGGACATCTAGTGTTTTTATAACCATGTT
2	57	TACTATACCTATGGCTTTTTGAAATTCTTAATCGGTATCCATTTTTAAGATGAGAAG
3	54	ACGCTGGTTTTTTGAAAGTAAAATTACATCGAACTTTTTTGGATCTCAAGCCGC
4	54	TCGCTGAGTTTTTATAGGTGCCTGTTTACTCATATTTTTTATACTTTAGGATCT
5	54	AAAGGGCCTTTTTTCGTGATACGTGGTTTCTTAGTTTTTACGTCAGGTGATTTG
6	52	AAGAACCATTTTTGGGAAAACCAGGTACAAAGTTGCAGGACCACCTGGATGG
7	52	TAAGCCTTTTTCTCCCGTATCCCGGTTGTAATTTTTTTCGCGTTAAAAGCAC
8	52	TATTAACTTTTTTGGCGAACTACTTACTAATGTCATGATAATAACCTATTTT
9	52	CGCCGGTTTTTGCAAGAGCAAGAGCTGAATGATTTTTAGCCATACCAGATGG
10	52	CGCTCGTTTTTGCCCTTCCGGGGTGAGCGTGGTTTTTGTCTCGCGGTATGCC
11	52	TACATGTTTTTCGATGTGACCGAGTATTCAACTTTTTATTTCCGTGTGATAG
12	52	AGGTGATTTTTAGATCCTTTTAGTTTTCGTTCTTTTTCACTGAGCGTAGTCA
13	52	TATTACTTTTTCCTCGATATTCTATATTCCTATTTTTGATTAGAAGGTAAAT

14	52	ATAACCATTTTTTGAGTGATAACACTGCTGTGCGCGGAACCCCTGCACTTTT
15	52	GGGTTGGTATAATTGTCCGAGGGTTTTTTTCCCAGAGCTATATTACTAGTCA
16	52	AAAACTTCTTGTGCAATCCTACTTTTTTTTAAAAGAAAAGAGGTAACGGGTA
17	52	TTTATTTTTTTTTTCTAAATAACCCTGATAAATTTTTTGCTTCAATACCAAG
18	52	AATCTGTTTTTGCGCCCGGGTTCGCCCGCATTTTTTTTTGCGCAGCACATAA
19	52	TACAGGTTTTTGCGCACATTAAATTTTGACTCTTTTTGGGTGTAATCGGCAT
20	52	GGGGGATTTTTTCATGTAACTTCTCAGAATGATTTTTCTTGGTTGAGTGGCA
21	51	TTTAATTTTTAGTTCTGCTATTGACGGGGAATTTTTAGCCGGCGAAGAGCG
22	51	CAGTCTATGCGACAGTCTGGCAATCTGACATAATTATTGATTTTTGCTGGT
23	51	AAATCTTTTTAAAAGAATAGGAACAAGAGTCTTTTTCACTATTAAACCGTC
24	51	GGAAGCCGCCTGTATCGGCCTTCATGGTCGTTAATGACCCTTTTTAAGAAA
25	50	CACGCTGCTTTTTGCGTAACCACTTTTCGCCCCGTTTTTAAGAACGTTTT
26	50	ATCACCCTTTTTTAATCAAGTTTTTTTAACCAATTTTTTAGGCCGAAATC
27	49	AGATATTTTTTTAATACCGTGTATCTTGCACACGTCTGGCGTAACTCCC
28	49	GGACCTTTTTACGTGTATTCATTCAATTTTTAATTTAAAAGATTGATTT
29	47	ACTTGTGTTGTTCCAGTTTGACCGAGATAGGGTTGAGGTCAGTTATC
30	47	TGGCGTCAAAGGGCGAAAAAGAACGTGGACTCCAACTAACGCAAGCA
31	47	CCAAAAAAGCATCTTACGGATACTCACCAGTCACAGTGATGAGCACT
32	47	GGCAAAAATCCCTTAACGTGTGATAATCTCATGACCAAATCCCTTAT
33	44	TATAGGTTCTAGCTTCCCGGCAACTTTTTAATTAATAGATTATG
34	44	CGGGGAAAGGCCAACTTACTTCTGTTTTTACAACGATCGAACAT
35	44	AGGCGGATGAGCGAACGAAACCTTTTTTTGCACTCTGAAAACAA
36	42	AGTAACCTCAGCGCCCTTATTCCCTTTTTTGCAGTACTATGC
37	42	GTGACACCACGATCATTGCAGCACTGGGGCCAAACGACGAGC
38	42	AGCCTCTCCACATATTGAAAAAGGAAGAGTATTTAAACGAAT
39	42	ATACTCTATGTGGCTCAACAACAGAGAATAAGCGTTGTTTTG
40	42	TCATATGTACCGTAGTTATCTACACGACGGGGCAGACCCCAA
41	42	AGGTGCCGTAATTTTTGTTAAATCAGCTCATTTTTGGGGTCG
42	42	TATTATCCCGTAGGGAGCCCCCGATTTAGAGCTGTGGCGCGG

43	42	TATCTTCACGAGGTAGCAAATACGTATCTTCCCACAAGTAAA
44	42	GCATACACTATCGCCTTGATCGTTGGGAACCGCTCGGTCGCC
45	42	GCGCTTAATGCCAGCGGTAAGATCCTTGAGAGCACACCCGCC
46	42	AGTTAGAATTGCGTTGCGCTCCTGCAGCAGGTACATGTTAGA
47	34	GGCAACTTTTTTATGGATGAACGAAATGATAAAT
48	34	TGTAGCTTTTTAATGGCAACAACGTTAACCGCTT
49	34	TTTGCCTTTTTTTCCTGTTTTTGCTCCGCTCATG
50	34	AGTTTGTTTTTAGGGGACGACGACGAAAGAGACA
51	34	GTGTTGTTTTTTAGAAAGTCAAATCCGTAACTGT
52	34	TATCAGTTTTTGGCGATGGCCCACTAGGGAAGAA
53	34	GGCGCTTTTTTAGGGCGCTGGCAAGTTGCTATGC
54	34	TGACAGTTTTTTAAGAGAATTATGCATGGGTGCA
55	32	CTGGAGCCCTGGCTGGTTTATTGCTAGACAGA
56	32	TTTTGCACGAGGACCGAAGGAGCTGCGCAAAC
57	32	AGACAATACATTCAAATATGTATCACCCAGAA
58	32	AAGCTTTATCTGTCTATACAAACTATTCGACG
59	32	CAGACCAACACTGATTAAGCATTGTATCTGTG
60	32	AAAATAGATATGTGTAAAAGGAACGTAGCGGT
61	32	CGAGTGGGGATGCTGAAGATCAGTGTGCTGCC
62	32	AGCGAAAGCGTGGCGAGAAAGGAACGTGAACC
63	26	TAAATCTTTTTGGAACCCTAAATTGA
64	26	TTATGTTTTTTGTTGCAGGTATATCT

Table D.4: Staple sequences used to fold the icosahedron of 52-bp edge length (with no single crossovers).

Staple ID	**Length (bp)**	**Staple sequences**
1	59	AACGATGTGTGAATTTTTTCTTGAAGTGTGGAGCTAGAGTTTTTCCCCAGCTGCTAACC
2	59	AACCAGGTACGAGTTTTTCGAACGAAACTAAAAGAAAAGTTTTTAGGTTGGTCGGTATC
3	59	ACTTTTCGGGGAATTTTTATGTGCGCGGCAAATATGTATTTTTTCCGCTCATGATAAGT
4	59	ATAGACAGATCGCTTTTTTTGAGATAGGTCCAAGTTTACTTTTTTCATATATACTCGGCC
5	59	TGAAAGTAAAAGATTTTTTTGCTGAAGATGGATCTCAACATTTTTGCGGTAAGATCAATG
6	59	TAGGCCGAAATCGTTTTTTGCAAAATCCCTGAGTGTTGTTTTTTTTCCAGTTTGGAAACGT
7	58	CTCTGCTATATTTTTTCAGCATTAGAAACGTTACACCTTTTTACGTATCGTCAGGAAG
8	58	CTGCTATGCAATTTTTAATAGAATCGAAGATATTAATTTTTTACCGTGTATCATAATT
9	57	GTCCGAGGGAAAGATAATCTAGATCGTCGGTCCACACGTGCTTCCCAGAGCTATATT
10	57	GGGCCAGATACCGCTTTTTTTGCACAAGGACCGAAGGAGCTAGGTAAGCCCTCCCGTA
11	57	CATCACCCTAATTTTTTCAAGTTTTTCAAAAGCCCGGTTTTTAATGTCTCCCTCTTA
12	57	AAAGGGCCTGGCGCTAGGGCGCTGGGAAAGCGAAAGGAGCGCGTGATACGCCTATTT
13	57	AGCTTGACGAAAGAACGTGGACTCCACAAGAGTCCACTATTGGGAAAGCCGGCGAAC
14	57	CAGACCCCAATTGGTAACTGTCAGAGCCTCACTGATTAAGCATCATATGTACCCCGG
15	57	CTGCGTAAAAGTGGCTAACGCAAGCGTATCCAAAGATGAGAGCTGGCGCCAAGGGTG
16	57	CATGAGTGAGGTATTATCCCGTATTTTCTGCTATGTGGCGCTAACACTGCGGCCAAC
17	57	TGGTTGAGTACTTTTTTCACCAGTCATAACTCCCGGGTTTTTTTGGTTTGCATACTT

18	56	TTACTTCTGATTTTTCAACGATCGGACATGGGGGATTTTTTCATGTAACTCCTCGA
19	56	GTGGCGAGAATTTTTAGGAAGGGAACAAGTGTAGCGTTTTTGTCACGCTGCCCGCT
20	53	CTTCCGTTTTTGCTGGCTGGTGGGTCTCGCGGTTTTTTATCATTGCAGCACTG
21	53	CAAAGGTTTTTGCGAAAAACCAACGTGAGTTTTTTTTTTCGTTCCACTGAGCGT
22	53	AAATTATTTTTCATGCGATGTTTTGAGGGGACTTTTTGACGACGAATTCGACG
23	52	ATGAGCTTTTTACTTTTAAAGGACGCCGGGCATTTTTAGAGCAACTCTTTTT
24	51	TGAGTATGTCACTAAATCGGAACCCTAAAGTTTTTGGAGCCCCCGATTTAG
25	51	ATTTTGACTCGGGTTTTTTGTAATCAGTACTATGCAGTAACCTCTAGTGAT
26	51	TCTTCACGAGCATGACAGTAAGAGAATTATTTTTTGCAGTGCTGCCATAAC
27	51	GCCGCATAAGTGTTTTTTTGTAGAAAGTCAAATCCTATCTGTGCCTCTCCA
28	51	GTTCGCCCGCATTTTTTTTTGCGCAGCACCTGTATCGGCCTTCCGCCGCGC
29	49	AGCCATACCAATTTTTACGACGAGCGTGACACCACGATGCCTGGAGCCG
30	49	TTATTATACTCTTTTTGAAACTCTGACTGAGAGTGTTGGTTCAGCGCGT
31	48	TTATAGGTTATTTTTATGTCATGATAATAATGGTTTCTTATAGACCGA
32	48	TTGTAATTCGTTTTTCGTTAAATTTTTGTTAAATCAGCTCTGGGTTAC
33	48	CTGACATAATTTTTTTATTGAGCTGGTTACTAGGTACATGGCATGGTA
34	48	TCGTAGTTATTTTTTCTACACGACGGGGAGTCAGGCAACTCCCGAAGA
35	48	ACAGGGTTTTTCGCACATTAATGAATGTTAACTTTTTCTCTACTGATG
36	48	TGTTGCAGGTAGCTTTTTGACAGTCTGGCAATACTAGTGTAAAAGGAA
37	48	ATTTTGATGTTTATTTTTTAGGTAGAGGGTAATAACATCGCGGAAGAA
38	48	ACAACGTTGCGCATTTTTAACTATTAACTGGCGAACTAAGACGCGAGA
39	48	CTTCCTTTTTTTAAGCATACCCCTCGATATTCTTTTTACAAGTAAATA
40	46	AAAAGGATCTATTTTTGGTGAAGATCCTTTTTGATACTACGTGAAC
41	46	TTGAAAAAGGATTTTTAGAGTATGAGTATTCAACATCAGAATGACT
42	46	GAGACAAAGCTTTTTTTTAAACAAAATCTGGCGCCCTTCAATAATA
43	45	GGTCGGGTGATTTTTAATCACCACATGAAGCCATAAGCTGAATGA
44	44	TTAATCTTTTTTTCGCTCTGGAGCGACGATGGACTGTTCCATTG
45	44	TGCGGCTTTTTATTTTGCCTTCCTGTTTTTGCTCACTTTTTCTA
46	44	TTCCCTTTTTTATATTCCTAGATTAGAAGGCGTTGTAGTTTCGT

47	42	CAATCCTACTTCTTGCACTCTGAATCTGTCTAATTCATTGTG
48	42	CAAAATCCCTTGTCTATCAGGGCGATGGCCCAATCTCATGAC
49	42	CCCTTATTCCCGGTCGCCGCATACACTATTCTTTCCGTGTCG
50	42	ACAACAGAGAAGACAATAACCCTGATAAATGCGGGTGGCTCA
51	42	GTGTATCGGCTTGGGGTCGAGGTGCCGTAAAGCAAGTGCTTA
52	42	GACCTTCAGAGCATATCTGCTTTTTACCCTAGCTGGGCTCAT
53	42	CTTCCCGGCAAGACGTGCGACTTGCGTATGATCTTACTCTAG
54	42	ACTTATGCGCTTTAGATTGATTTAAAACTTCATTGCAGGACC
55	42	TGGTTCGCAATGCCTTGATCGTTGGGAACCGGGTCTCGGTAA
56	42	TTATTACTATGGCGAAATTCTTATATGTCACAGTCTATTATC
57	42	GATTAGTGACGCAGAAAAGCATCTTACGGATGCGTAGCGAAG
58	41	AATAGATTTTTCTGGATGGAGGCGGATAAAGTTTTTAATTT
59	41	GGTCAGTTTTTTTATCGGACCACGTGTATTCTACAAACTAA
60	37	GATAGGGTTTATAAATCAAAAGAAGACGTCAGGTGGC
61	37	ATCGAACTCAGTTGGGTGCACGAGATTTTTTAACCAA
62	37	ACGTTTTCCCTTGAGAGTTTTCGCATGGATGAACGAA
63	37	AATACATTAACCCCTATTTGTTTACCAGAAACGCTGG
64	37	CCCAAGAGGACCTTAAACGAATAGAAGAACCAGGGAA
65	37	TTAATGCGGCGTAACCACCACACCAAACGGGTAGGAA
66	37	GATAGTCATTGCGTTGCGCTCCTGGTCCAAAGAATGA
67	37	GTGAGCGTTTATTGCTGATAAATCTGTAGCAATGGCA
68	37	AACCATGTCACGTCTGGCGGACATCAGGATAGTTATG
69	37	GCAAATACTTAATGACCCAAGAAATTAGAAGTTAGAA
70	37	AACCGTAGGACCAGATCGTTCAGCTTTGATACTCTAT
71	37	CCAAACACAATAGTTGATCTCGGTTAAAACGACTACA
72	26	CGCTGCTTTTTTCTCGAAAATCAATT

Table D.5: Staple sequences used to fold the pentagonal pyramid of 63-bp edge length

Staple ID	Length (bp)	Staple sequences
1	59	TTTTGGGGTCTTGTGCCGTAAAGCACTAAGGCAACAATTTTTTAGACTGGATGGAATCC
2	59	GCGAACGTGGTTGAAAGGAAGGGAAGAAAGATCATGTAATTTTGCCTTGATCGTTGCCT
3	59	GTGCACGAGTTTTTACATCGAACTGGATCGATGAGCACTTTTTAAAGTTCTGCTGGCAA
4	59	CGTAACCACCTTCCCGCCGCGCTTAATGCAGAAAAGCATTTTTACGGATGGCATGCCAA
5	59	CGTGGACTCCTTGTCAAAGGGCGAAAAACTTCGCGTTAATTTTTTTGTTAAATCAAGAA
6	52	TTGGTATTTTTACTGTCAGACAAAACTTCATTTTTTTTTTAATTTAACGGCT
7	52	GGCTGGTTTTTTTTATTGCTGGGTATCATTGCTTTTTAGCACTGGGGCTACA
8	48	CCCTTATAAATCAAAGCTCATTTTTTAACCAATAGGCCGATACTTTAG
9	48	GAACGTTTTCCAATTCAACAGCGGTAAGATCCTTGAGAGTGAACAAGA
10	48	GTGATAACACTGCGGACAGTAAGAGAATTATGCAGTGCTGCCCGTATC
11	48	GTGCCTCACTGATTAACTATGGATGAACGAAATAGACAGATCCCGTAT
12	48	GCGTGACACCACGATGGGAACCGGAGCTGAATGAAGCCATTGAGCGTG
13	42	ATGACTTATTGCGTTGCGCTCCTGAAGAACACTATTCTCAGA
14	42	TCACCAGTCACGCCGCTACAGGGCGCACATTAGGTTGAGTAC
15	42	AGGAGCTAACCGCGCTGGCAAGTGTAGCGGTCGGAGGACCGA
16	42	ACAACATGGGGGCGAAAGGAGCGGGCGCTAGGGCTTTTTTGC
17	42	ATTAACTGGCGCCCGATTTAGAGCTTGACGGGTGCGCAAACT
18	42	TCTAGCTTCCCATCGGAACCCTAAAGGGAGCCAACTACTTAC
19	42	GTGAAGATCCTCACTTCTGCGCTCGGCCCTTCAAGGATCTAG
20	42	CTCATGACCAAAGGCGGATAAAGTTGCAGGACTTTTGATAAT
21	42	TCAGACCCCAAACTACGTGAACCATCACCCTACCACTGAGCG
22	42	CCCGGTTGTAACGTCTATCAGGGCGATGGCCCTCATATGTAC
23	36	GTAGTTATCCAGATGGTAAGCCCTCCATAACCATGA

24	36	GGTCTCGCATAAATCTGGAGCCGGACCAAACGACGA
25	36	GTCCACTAGTGTTGTTCCAGTTTGTTTCGCCCCGAA
26	36	TGACGCCGATGTGGCGCGGTATTATCGCTGAGATAG
27	36	ATTGATTTCAAGTTTACTCATATAAATCGGCAAAAT
28	30	GAGCAATTTTTGGTCGCCGCATTCAGTTGG
29	30	CTTACTTTTTTTGACAACGATCACGCTGCG
30	30	GTAGCATTTTTGGCAACAACGTGAAAGCCG
31	30	CTTAACTTTTTGAGTTTTCGTTATCAAGTT
32	30	TAGACCTTTTTGATAGGGTTGATTAAAGAA
33	26	CGACGGTTTTTGGAGTCAGGCAAGCA

Table D.6: Staple sequences used to fold the asymmetric octahedron of 63-bp edge length (with no single crossovers).

Staple ID	**Length (bp)**	**Staple sequences**
1	59	CTTGAGAGTTTTCGCCCTTGGGAACCGGAGCTGTAACTCGCCTTGATCGCGAAGAACGT
2	57	AGTTTGGAACAAGTCCACTATTAAAGAACGTCGTAAAGCACTAAATCTTCCTAAAGG
3	57	GCTTCCCGGCAACTTAATAGACTGGATGGAGGTATCATTGCAGCACTTTCAGATGGT
4	57	TCACCAGTCACAGTAGCATCTTACGGATGGCCTGATTAAGCATTGGTTTTCAGACCA
5	53	AAGCCCCCGAAGGATTTTACCGCTTTTTGAGCGTGACATTTTTCGATGCCTGT
6	53	AGTTTAATCTCATGTTTTAAATCCCTTAAAATGTGCGCTTTTTACCCCTATTT
7	53	GAGCCCCTAGGGCGTTTTCAAGTGTAGCGTAGGAAGCCTTTTTTAAGTGTTGT
8	48	TTTTTGCGGCATTTTATGAGTATTCAACATTTCCGTGTCGACCCAAGA
9	48	CGTGTATTCATTCACTAGTGATAACCATGTTACTTGGTCAGATGCTGA
10	48	CTCAACAACAGAGAAGAGACAAAGCTTTAAACAAAATCTGCCAAGAAA
11	48	CATCACCCTAATCAAAAAACCGTCTATCAGGGCGATGGCCTAATTTAA

12	48	GAAGGATTAGTGACCCCTCGATATTCACAAGTAAATATCTACTATTAA
13	48	TGATAAATCTGGAGCTTATGCGCTCGGCCCTTCCGGCTGGGCGAGAAA
14	47	AGAAACTGGGTGCACGTTTGGTTACGCGCGGTATTATTTCGTATTGA
15	47	ACCTTAGGACGACGACTTTTCGACGTCATTTTTTAATTTATAGGCCG
16	46	TGGCGGACATTTGTGCAATCCTATTTTTTAAAAGAAAAGAGGCAAA
17	42	ATCTCAACAGCCTTTTAAAGTTCTGCTATGTGATCGAACTGG
18	42	TACCAAACGACTGCACAACATGGGGGATCATGAATGAAGCCA
19	42	GGTCGAGGTGCGGACTCCAACGTCAAAGGGCGAGTTTTTTGG
20	42	CTTTAGATTGAAGGTGAAGATCCTTTTTGATACTCATATATA
21	42	TAGAAGGATTGCGTTGCGCTCGTCTGGCCTATATTCCTAGAT
22	42	ATACTCTATGTGCCGCTACAGGGCGCACATTACGTTGTTTTG
23	42	TTTGCGCAGCACACACCCGCCGCGCTTAATGCTCGCCCGCAT
24	42	CCTTCAAACGGGGTCACGCTGCGCGTAACCACCCTGTATCGG
25	42	TGTGAAGAACCAGAATAGACCGAGATAGGGTTAAATCCTATC
26	42	AGGTACGAGCGCAAAATCCCTTATAAATCAAAAGGGAAAACC
27	42	GTGATACGCCTTTAAATTTTTGTTAAATCAGCAAAGGGCCTC
28	42	GTTAATGTCATGTACCCCGGTTGTAATTCGCGATTTTTATAG
29	42	GTTTCTTAGACAGCGTCAGACCCCAATCATATGATAATAATG
30	42	ACTTTTCGGGGACGTGAGTTTTCGTTCCACTGGTCAGGTGGC
31	42	ATTCAAATATGATACACTATTCTCAGAATGACTTCTAAATAC
32	42	TGAGACAATAACAAGAGCAACTCGGTCGCCGCTATCCGCTCA
33	42	GCTTGACGGGGAGAAAGCGAAAGGAGCGGGCGCCGATTTAGA
34	42	TGGGTCTCGCGGCGGATAAAGTTGCAGGACCACCGGTGAGCG
35	42	TAGTTATCTACCTTCTGACAACGATCGGAGGATCCCGTATCG
36	42	GTCAGGCAACTGATAACACTGCGGCCAACTTAACGACGGGGA
37	42	GAAATAGACAGAGTGCTGCCATAACCATGAGTATGGATGAAC
38	42	TAGGTGCCTCAATGACAGTAAGAGAATTATGCATCGCTGAGA
39	37	TACGTATTTTCTTCCAATACTAGTCATTTTCTATTAT
40	36	AGATCAGTGCTGGTGAAAGTAAAAGTTATCGGACCA

41	36	CTGGCGAACAACAACGTTGCGCAATCACGACGTAGC
42	36	GTTTGAGGAACGAATAGCCTCTCCCCCTTATTCCCT
43	36	GGAAGGGAAAAGCCGGCGAACGTGCTGGTTTATTGC
44	36	AAGGATCTTTTAAAACTTCATTTTCACTACGTGAAC
45	36	GCATGGTAGTTGGTCGTTAATGACGCGCCCGGGTGG
46	35	AAAGGAAGAGTGCCTTCCTGTTTTTTTTCTCACCC
47	35	ATACAAACTAATAAGTAAATTACTTTTTCGATGTG
48	32	CTTATTAGTAACTCCCGGGTTGGTCACACGTC
49	32	AAATCGGAACGAAACCTTGCACTCATCTGTCT
50	32	CGCCGGGCCCTGATAAATGCTTCAATATTGAA
51	25	TTTCCAATGATGAGCAGGTAAGATC
52	20	AGCAATGGCTACTTACTCTA
53	20	AGAAAGTCGAGTGTTGTTCC
54	20	GTTTATTTTTGGTTGAGTAC

Table D.7: Staple sequences used to fold the asymmetric octahedron of 63-bp edge length (with single crossovers).

Staple ID	**Length (bp)**	**Staple sequences**
1	46	TGGCGGACATTTGTGCAATCCTATTTTTTAAAAGAAAAGAGGCAAA
2	48	CGTGTATTCATTCACTAGTGATAACCATGTTACTTGGTCAGATGCTGA
3	36	AGATCAGTGCTGGTGAAAGTAAAAGTTATCGGACCA
4	47	AGAAACTGGGTGCACGTTTGGTTACGCGCGGTATTATTTCGTATTGA
5	35	AAAGGAAGAGTGCCTTCCTGTTTTTTTTCTCACCC
6	48	TTTTTGCGGCATTTTATGAGTATTCAACATTTCCGTGTCGACCCAAGA
7	36	GTTTGAGGAACGAATAGCCTCTCCCCCTTATTCCCT
8	47	ACCTTAGGACGACGACTTTTCGACGTCATTTTTTAATTTATAGGCCG
9	35	ATACAAACTAATAAGTAAATTACTTTTTCGATGTG

10	48	CTCAACAACAGAGAAGAGACAAAGCTTTAAACAAAATCTGCCAAGAAA
11	36	GCATGGTAGTTGGTCGTTAATGACGCGCCCGGGTGG
12	36	TAAAGAACGTCGTAAAGCACTAAATCTTCCTAAAGG
13	42	GGTCGAGGTGCGGACTCCAACGTCAAAGGGCGAGTTTTTTGG
14	48	CATCACCCTAATCAAAAAACCGTCTATCAGGGCGATGGCCTAATTTAA
15	36	AAGGATCTTTTAAAACTTCATTTTCACTACGTGAAC
16	42	CTTTAGATTGAAGGTGAAGATCCTTTTTGATACTCATATATA
17	24	AGTTTAATCTCATGTTTTAAATCC
18	36	TACGGATGGCCTGATTAAGCATTGGTTTTCAGACCA
19	42	TAGGTGCCTCAATGACAGTAAGAGAATTATGCATCGCTGAGA
20	42	GAAATAGACAGAGTGCTGCCATAACCATGAGTATGGATGAAC
21	42	GTCAGGCAACTGATAACACTGCGGCCAACTTAACGACGGGGA
22	42	TAGTTATCTACCTTCTGACAACGATCGGAGGATCCCGTATCG
23	24	AAGCCCCCGAAGGATTTTACCGCT
24	36	CTGGATGGAGGTATCATTGCAGCACTTTCAGATGGT
25	42	TGGGTCTCGCGGCGGATAAAGTTGCAGGACCACCGGTGAGCG
26	48	TGATAAATCTGGAGCTTATGCGCTCGGCCCTTCCGGCTGGGCGAGAAA
27	36	GGAAGGGAAAAGCCGGCGAACGTGCTGGTTTATTGC
28	42	GCTTGACGGGGAGAAAGCGAAAGGAGCGGGCGCCGATTTAGA
29	24	GAGCCCCTAGGGCGTTTTCAAGTG
30	32	CGCCGGGCCCTGATAAATGCTTCAATATTGAA
31	42	TGAGACAATAACAAGAGCAACTCGGTCGCCGCTATCCGCTCA
32	42	ATTCAAATATGATACACTATTCTCAGAATGACTTCTAAATAC
33	20	TTGTTTATTTTTGGTTGAGT
34	50	CTTAAAATGTGCGCTTTTTACCCCTATACTCACCAGTCACAGTAGCATCT
35	42	ACTTTTCGGGGACGTGAGTTTTCGTTCCACTGGTCAGGTGGC
36	42	GTTTCTTAGACAGCGTCAGACCCCAATCATATGATAATAATG
37	42	GTTAATGTCATGTACCCCGGTTGTAATTCGCGATTTTTATAG
38	42	GTGATACGCCTTTAAATTTTTGTTAAATCAGCAAAGGGCCTC

39	32	AAATCGGAACGAAACCTTGCACTCATCTGTCT
40	42	AGGTACGAGCGCAAAATCCCTTATAAATCAAAAGGGAAAACC
41	42	TGTGAAGAACCAGAATAGACCGAGATAGGGTTAAATCCTATC
42	20	GTAGAAAGTCGAGTGTTGTT
43	50	TAGCGTAGGAAGCCTTTTTTAAGTGTTCCAGTTTGGAACAAGTCCACTAT
44	42	CCTTCAAACGGGGTCACGCTGCGCGTAACCACCCTGTATCGG
45	42	TTTGCGCAGCACACACCCGCCGCGCTTAATGCTCGCCCGCAT
46	42	ATACTCTATGTGCCGCTACAGGGCGCACATTACGTTGTTTTG
47	42	TAGAAGGATTGCGTTGCGCTCGTCTGGCCTATATTCCTAGAT
48	37	TACGTATTTTCTTCCAATACTAGTCATTTTCTATTAT
49	32	CTTATTAGTAACTCCCGGGTTGGTCACACGTC
50	48	GAAGGATTAGTGACCCCTCGATATTCACAAGTAAATATCTACTATTAA
51	36	CTGGCGAACAACAACGTTGCGCAATCACGACGTAGC
52	20	GTAGCAATGGCTACTTACTC
53	50	TTTTGAGCGTGACATTTTTCGATGCCTTAGCTTCCCGGCAACTTAATAGA
54	42	TACCAAACGACTGCACAACATGGGGGATCATGAATGAAGCCA
55	59	CTTGAGAGTTTTCGCCCTTGGGAACCGGAGCTGTAACTCGCCTTGATCGCGAAGAACGT
56	25	TTTCCAATGATGAGCAGGTAAGATC
57	42	ATCTCAACAGCCTTTTAAAGTTCTGCTATGTGATCGAACTGG

Table D.8: Staple sequences used to fold the chiral object of 63-bp edge length (with no single crossovers).

Staple ID	**Length (bp)**	**Staple sequences**
1	59	AGAGCTGATCCTTTTTGATAATTTATGACCAAAATCCCGGTTGTAATTTTTTCGTTAAA
2	58	GGGCGCGCAATGGCAACAACTTTTTTTGCGCAAACTGGCAACAATTATTTGACTGGAT
3	57	GGAGGCGGGACCAAGTTTACTCATTTTACTTTAGATAAAGCGAAAGTTTGGGCGCTA

4	56	AACATTTTTTGCCTTCCTGTTTTTTGCTCACCCAGTTATTTTTCTTTTTTATACAT
5	55	AATTATGTCCAAAGAATGATTTTAACGATGTGTGATCGTTGGGATTTTTCGGAGC
6	54	TGAATCCGCGCTTAATTTTTTCCGCTACAGGGCCTCTACTGATTTTTTTGTGAG
7	54	TCCCGTAACTATGGATGAACGATTTTAGACAGATCGCTGGTTTATTTGATAAAT
8	51	TGTGTTGCTTTTGTAGCGACAGTAAATATCTTTTTTACGTAGCGAAGTGAT
9	50	ATCGGAGTTTTTTACCGAAGGAGCTATAGTGGAGCTAGTTTTTGCCCCAG
10	50	GTCTAGCGAACGAATTTTTGCACTCTGAGTGGCTCAACATTTAGAGAATA
11	50	ACCATCATTTTTTCTAATCAAGTTTTGAACCCTAAAGGGTTTTCCGATTT
12	50	TTCCAATTTTTTTGAGCACTTTTAATTGACGCCGGGCAAGTTTCGGTCGC
13	49	TAGACGTCACACGAGTGGGTTACATTTTTCTGGATCTCAACAGAACGTT
14	48	CCAATCATATGTACCCCTTAACGTGAGTTTTCGTTCCACTGACGGGGA
15	48	AATCTGGCGCCCGGATCTGTCTATACAAACTAAGAGACAAATCTTCCC
16	48	TCGATATTCACAAGTCTGGCAATACTAGTCACAGTCTATTTGGAACAA
17	48	CTTTAAAAGAAAAGTTATCGGACCACGTGTATTCATTCATCCCTTTTT
18	48	CACAGAAAAGCATCCACTATTCTCAGAATGACTTGGTTGATCGCGGTA
19	48	CAACTTACTTCTGATGCAGTGCTGCCATAACCATGAGTGACTTACTCT
20	48	GGCAAAATCCCTTAGTTAAATCAGCTCATTTTTTAACCAAGCGGTATT
21	48	CTCCCCATATCTGCTCAAGTGCTTAGTGTATCGGCTCAAAATTGCCAA
22	48	CTTCAATAATATTGATATGTATCCGCTCATGAGACAATAATAAACGAA
23	48	CTGGCGGACATCTAGGATTAGTGACGTAACTCCCGGGTTGAGTTTTCG
24	48	GTCATGAATGTTAACGCACATTAATTGCGTTGCGCTCCTGGTAAAGCA
25	48	CCACATGAAGCCATGCGTAAAGCTGGCGCCAAGGGTGGGTAGGGCGAT
26	48	GGGTAATAACATCTACTGTAAAACGACTACAATTTTGATGACTGAGAG
27	45	AACCATTTTTTTTACTTGGTCAGAGGTTGGTCGTTTTTTGACCCA
28	44	TGTAGAACGCGAGACTCTGCTTTTCAGCATTAGAAATATTACAC
29	44	TAGGGTTTATTAAAGAACGTGTTTCCAACGTCAAAGGGACGTGA
30	43	AGTAAATTCCACCCAAGAGTTTGATTTGGACGACGACGTTCGC
31	43	CGCATATTACGGATGGCATTTTTTCAGTAAGAGAATTACAACG
32	43	TCTGGATTTTTGCGACGATGGGGGCTCATGATTTTTTTTCAGA

33	42	AGGACCACTTAATTAAGCATTGGTAACTGTCAATAAAGTTGC
34	42	CCTTCCGGCTGGCTGAGATAGGTGCCTCACTGTGCGCTCGGC
35	42	AAGGGCCTCGTAATCTAGATCTAACCTTAATCAATTCGACGA
36	42	TTTTTATAGGTTCGGTCCACACGTGCAAAGATGATACGCCTA
37	42	AGCCGGCGAACATTTAAAAGGATCTAGGTGAATGACGGGGAA
38	42	AGGAAGGGAAGTGATTTAAAACTTCATTTTTAGTGGCGAGAA
39	42	AGCGGTCACGCCGTGACACCACGATGCCTGTATGGCAAGTGT
40	42	CACCACACCCGGAAGCCATACCAAACGACGAGTGCGCGTAAC
41	42	CACGTATCTTTTCAGGAAGCTAGTCTCGGTATTTTTGGTTCG
42	42	TTCGGGGAAATATGCTGAAGATCAGTTGGGTGGGTGGCACTT
43	42	CCCCTATTTGTAAACGCTGGTGAAAGTAAAAGGTGCGCGGAA
44	42	AGAAAGCTAGATTAGAAGGCTTTTTGATACTCTATGTAGTTA
45	42	GTGGACCAGATGGGGATCATGTAACTCGCCTTGAATCTTGAA
46	42	TTTCGTAACCGACCGCTTTTTTGCACAACATGCGTTCAGCAG
47	42	GCGCAGCTATGGCGAAATTCTCTCAGGATTCGCCCGCATTTT
48	42	GCCTTCAAACGCTCGACTTCCTTAAGCATACCACCTGTATCG
49	42	TATCTGTGAAGGTGCGACTTGCGTATGATAGAAGTCAAATCC
50	42	AACCAGGTACGTACGCTGCTCTCGAAAATGACAACCAGGGAA
51	36	GTCAGGCATCGTAGTTATCTACACGAGCGTCAGACC
52	36	TCATTGCACCGGTGAGCGTGGGTCGTACTCACCAGT
53	36	AGCTTCCCATTAACTGGCGAACTATAACACTGCGGC
54	36	CTAAATCGTTGGGGTCGAGGTGCCCAGCGCGTGATA
55	36	GGCCCACTCGAAAAACCGTCTATCCGGGTGAAATCA
56	36	GAGTCCACGAGTGTTGTTCCAGTTATCTTATTACCC
57	36	ATCCCGTAAGTTCTGCTATGTGGCTAGGCCGAAATC
58	36	ACACAACGGTTGATCTCGGTTTCCAGCCCGGAATGT
59	36	TATATTCCATGGTAGCAAATACGTAGCTTTAAACAA
60	36	TGTTGGTTTATACTCGAAACTCTGTTTATAGGTAGA
61	36	CCCCGAAGCGGTAAGATCCTTGAGGTTTGCACACGT

62	36	TAGCCTCTACATGCGATGTGACCTCCCTGATAAATG
63	36	TGCGGCATCCGTGTCGCCCTTATTTGTGCAATCCTA
64	29	CTGGAGGCACTGGGGCCAGTTGTAAGCCC
65	27	CTGCGTAATGTCATGATAATGGTTTCT
66	26	TTTTTTAAATCAAAAGTTTACCGAGA
67	25	TATGTTTTTACCCTTTTTGCGGAAG
68	24	CAATGGTAGGAAGCTTTTAAGTGT
69	24	TCAAAAAAAGGAAGTTGAGTATTC

Table D.9: Staple sequences used to fold the chiral object of 63-bp edge length (with single crossovers).

Staple ID	**Length (bp)**	**Staple sequences**
1	30	TCAACATTTAGAGAAGTTTGATTTGGACGA
2	33	GTCTAGCGAACGAATTTTTGCACTCTGAGTGGC
3	48	AATCTGGCGCCCGGATCTGTCTATACAAACTAAGAGACAAATCTTCCC
4	36	TATATTCCATGGTAGCAAATACGTAGCTTTAAACAA
5	42	AGAAAGCTAGATTAGAAGGCTTTTTGATACTCTATGTAGTTA
6	45	AACCATTTTTTTTACTTGGTCAGAGGTTGGTCGTTTTTTGACCCA
7	48	CTTTAAAAGAAAAGTTATCGGACCACGTGTATTCATTCATCCCTTTTT
8	36	TGCGGCATCCGTGTCGCCCTTATTTGTGCAATCCTA
9	56	AACATTTTTTGCCTTCCTGTTTTTTGCTCACCCAGTTATTTTTCTTTTTTATACAT
10	24	TCAAAAAAAGGAAGTTGAGTATTC
11	48	CTTCAATAATATTGATATGTATCCGCTCATGAGACAATAATAAACGAA
12	36	TAGCCTCTACATGCGATGTGACCTCCCTGATAAATG
13	20	TAAGTAAATTCCACCCAAGA
14	42	CCCCTATTTGTAAACGCTGGTGAAAGTAAAAGGTGCGCGGAA
15	42	TTCGGGGAAATATGCTGAAGATCAGTTGGGTGGGTGGCACTT
16	20	TTAGACGTCACACGAGTGGG

17	32	CATGATAATGGTTTCTTACATTTTTCTGGATC
18	30	GAGCTAGTTTTTGCCCCAGCTGCGTAATGT
19	42	TTTTTATAGGTTCGGTCCACACGTGCAAAGATGATACGCCTA
20	42	AAGGGCCTCGTAATCTAGATCTAACCTTAATCAATTCGACGA
21	53	CGACGTTCGCTCTGGATTTTTGCGACGATGGGGGCTCATGATTTTTTTTCAGA
22	51	TGTGTTGCTTTTGTAGCGACAGTAAATATCTTTTTTACGTAGCGAAGTGAT
23	48	CTGGCGGACATCTAGGATTAGTGACGTAACTCCCGGGTTGAGTTTTCG
24	36	CCCCGAAGCGGTAAGATCCTTGAGGTTTGCACACGT
25	45	TCAACAGAACGTTTTCCAATTTTTTTGAGCACTTTTAATTGACGC
26	42	AACCAGGTACGTACGCTGCTCTCGAAAATGACAACCAGGGAA
27	42	TATCTGTGAAGGTGCGACTTGCGTATGATAGAAGTCAAATCC
28	44	TGTAGAACGCGAGACTCTGCTTTTCAGCATTAGAAATATTACAC
29	24	CAATGGTAGGAAGCTTTTAAGTGT
30	42	GCCTTCAAACGCTCGACTTCCTTAAGCATACCACCTGTATCG
31	42	GCGCAGCTATGGCGAAATTCTCTCAGGATTCGCCCGCATTTT
32	48	GGGTAATAACATCTACTGTAAAACGACTACAATTTTGATGACTGAGAG
33	36	TGTTGGTTTATACTCGAAACTCTGTTTATAGGTAGA
34	55	AATTATGTCCAAAGAATGATTTTAACGATGTGTGATCGTTGGGATTTTTCGGAGC
35	25	TATGTTTTTACCCTTTTTGCGGAAG
36	48	CTCCCCATATCTGCTCAAGTGCTTAGTGTATCGGCTCAAAATTGCCAA
37	36	ACACAACGGTTGATCTCGGTTTCCAGCCCGGAATGT
38	42	GAATTACAACGATCGGAGTTTTTTACCGAAGGAGCTATAGTG
39	42	TTTCGTAACCGACCGCTTTTTTGCACAACATGCGTTCAGCAG
40	42	GTGGACCAGATGGGGATCATGTAACTCGCCTTGAATCTTGAA
41	26	TTTTTTAAATCAAAAGTTTACCGAGA
42	48	GGCAAAATCCCTTAGTTAAATCAGCTCATTTTTTAACCAAGCGGTATT
43	36	ATCCCGTAAGTTCTGCTATGTGGCTAGGCCGAAATC
44	48	TCGATATTCACAAGTCTGGCAATACTAGTCACAGTCTATTTGGAACAA
45	36	GAGTCCACGAGTGTTGTTCCAGTTATCTTATTACCC

46	44	TAGGGTTTATTAAAGAACGTGTTTCCAACGTCAAAGGGACGTGA
47	42	CACGTATCTTTTCAGGAAGCTAGTCTCGGTATTTTTGGTTCG
48	48	CCACATGAAGCCATGCGTAAAGCTGGCGCCAAGGGTGGGTAGGGCGAT
49	36	GGCCCACTCGAAAAACCGTCTATCCGGGTGAAATCA
50	54	TGAATCCGCGCTTAATTTTTTCCGCTACAGGGCCTCTACTGATTTTTTTGTGAG
51	48	GTCATGAATGTTAACGCACATTAATTGCGTTGCGCTCCTGGTAAAGCA
52	36	CTAAATCGTTGGGGTCGAGGTGCCCAGCGCGTGATA
53	50	ACCATCATTTTTTCTAATCAAGTTTTGAACCCTAAAGGGTTTTCCGATTT
54	42	CACCACACCCGGAAGCCATACCAAACGACGAGTGCGCGTAAC
55	42	AGCGGTCACGCCGTGACACCACGATGCCTGTATGGCAAGTGT
56	42	AGGGCGCGCAATGGCAACAACTTTTTTTGCGCAAACTGGCAA
57	53	CAATTATTTGACTGGACTCATTTTACTTTAGATAAAGCGAAAGTTTGGGCGCT
58	42	AGGAAGGGAAGTGATTTAAAACTTCATTTTTAGTGGCGAGAA
59	42	AGCCGGCGAACATTTAAAAGGATCTAGGTGAATGACGGGGAA
60	59	AGAGCTGATCCTTTTTGATAATTTATGACCAAAATCCCGGTTGTAATTTTTTCGTTAAA
61	48	CAACTTACTTCTGATGCAGTGCTGCCATAACCATGAGTGACTTACTCT
62	36	AGCTTCCCATTAACTGGCGAACTATAACACTGCGGC
63	50	CGGGCAAGTTTCGGTCGCCGCATATTACGGATGGCATTTTTTCAGTAAGA
64	48	CACAGAAAAGCATCCACTATTCTCAGAATGACTTGGTTGATCGCGGTA
65	36	TCATTGCACCGGTGAGCGTGGGTCGTACTCACCAGT
66	29	CTGGAGGCACTGGGGCCAGTTGTAAGCCC
67	54	TCCCGTAACTATGGATGAACGATTTTAGACAGATCGCTGGTTTATTTGATAAAT
68	42	CCTTCCGGCTGGCTGAGATAGGTGCCTCACTGTGCGCTCGGC
69	42	AGGACCACTTAATTAAGCATTGGTAACTGTCAATAAAGTTGC
70	20	ATGGAGGCGGGACCAAGTTT
71	48	CCAATCATATGTACCCCTTAACGTGAGTTTTCGTTCCACTGACGGGGA
72	36	GTCAGGCATCGTAGTTATCTACACGAGCGTCAGACC

D.2 Baited capture of an engineered tRNA

Table D.10: RNA sequences and their corresponding primer sequences and DNA templates for tRNA (native and engineered, DNA templates from extension of the two long overlapping primers) and the HIV 5'UTR

Name	**Sequence**
Ext Leu tRNA	GCGGGGGUUGCCGAGCCUGGUCAAAGGCGGGGGACUCCACACCUAAAGAUCCC CUCCCGUCCUCCAUUGGGGUUCCGGGGUUCGAAUCCCCGCCCCCGCACCAAACU CCUU
Ext Leu tRNA for	CTGCAGTAATACGACTCACTATAGCGGGGGTTGCCGAGCCTGGTCAAAGGCGG GGGACTCCACACCTAAAGATCCCC
Ext Leu tRNA rev	AAGGAGTTTGGTGCGGGGGCGGGGATTCGAACCCCGGAACCCCAATGGAGGAC GGGAGGGGATCTTTAGGTGTGGAGT
Nat Leu tRNA	GCGGGGGUUGCCGAGCCUGGUCAAAGGCGGGGGACUCAAGAUCCCCUCCCGUA GGGGUUCCGGGGUUCGAAUCCCCGCCCCCGCACCA
Nat Leu tRNA for	CTGCAGTAATACGACTCACTATAGCGGGGGTTGCCGAGCCTGGTCAAAGGCGG GGGACTCAAGATC
Nat Leu tRNA rev	TGGTGCGGGGGCGGGGATTCGAACCCCGGAACCCCTACGGGAGGGGATCTTGA GTCCCCCGCCTTTG
HIV 5'UTR	GGGUCUCUCUGGUUAGACCAGAUCUGAGCCUGGGAGCUCUCUGGCUAACUAGG GAACCCACUGCUUAAGCCUCAAUAAAGCUUGCCUUGAGUGCUCAAAGUAGUGU GUGCCCGUCUGUUGUGUGACUCUGGUAACUAGAGAUCCCUCAGACCCUUUUAG UCAGUGUGGAAAAUCUCUAGCAGUGGCGCCCGAACAGGGACUUGAAAGCGAAA GUAAAGCCAGAGGAGAUCUCUCGACGCAGGACUCGGCUUGCUGAAGCGCGCAC GGCAAGAGGCGAGGGGCGGCGACUGGUGAGUACGCCAAAAAUUUUGACUAGCG GAGGCUAGAAGGAGAGAGAUGGGUGCG
HIV 5'UTR for	TAATACGACTCACTATAGGGTCTCTCTGGTTAGACCAGATC
HIV 5'UTR rev	CGCACCCATCTCTC
HIV-1 NL4-3 p83-2	ctcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggAGACGGTCACAGCTTGTCTGT AAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGC GGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGT GCACCATGGTGACCTTCGTGCACATGGCCGGAGGAACTGCCATGTCGGAGGTG CAAGCACACCTGCGCATCAGAGTCCTTGGTGTGGAGGGAGGGACCAGCGCAGC TTCCAGCCATCCACCTGATGAACAGAACCTAGGGAAAGCCCCAGTTCTACTTAC ACCAGGAAAGGCTGGAAGGGCTAATTTGGTCCCAAAAAAGACAAGAGATCCTT GATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTGGCAGAACTACACA CCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGCTTCAAGTT...

HIV-1 NL4-3 p83-2 continued	AGTACCAGTTGAACCAGAGCAAGTAGAAGAGGCCAATGAAGGAGAGAACAACA GCTTGTTACACCCTATGAGCCAGCATGGGATGGAGGACCCGGAGGGAGAAGTA TTAGTGTGGAAGTTTGACAGCCTCCTAGCATTTCGTCACATGGCCCGAGAGCTG CATCCGGAGTACTACAAAGACTGCTGACATCGAGCTTTCTACAAGGGACTTTCC GCTGGGGACTTTCCAGGGAGGTGTGGCCTGGGCGGGACTGGGGAGTGGCGAGC CCTCAGATGCTACATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGT TAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTA AGCCTCAATAAAGCTTGCCTTGAGTGCTCAAAGTAGTGTGTGCCCGTCTGTTGT GTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCT CTAGCAGTGGCGCCCGAACAGGGACTTGAAAGCGAAAGTAAAGCCAGAGGAGA TCTCTCGACGCAGGACTCGGCTTGCTGAAGCGCGCACGGCAAGAGGCGAGGGG CGGCGACTGGTGAGTACGCCAAAAATTTTGACTAGCGGAGGCTAGAAGGAGAG AGATGGGTGCGAGAGCGTCGGTATTAAGCGGGGGAGAATTAGATAAATGGGAA AAAATTCGGTTAAGGCCAGGGGGAAAGAAACAATATAAACTAAAACATATAGT ATGGGCAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTTTTAGAGA CATCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAG GATCAGAAGAACTTAGATCATTATATAATACAATAGCAGTCCTCTATTGTGTGC ATCAAAGGATAGATGTAAAAGACACCAAGGAAGCCTTAGATAAGATAGAGGAA GAGCAAAACAAAAGTAAGAAAAAGGCACAGCAAGCAGCAGCTGACACAGGAAA CAACAGCCAGGTCAGCCAAAATTACCCTATAGTGCAGAACCTCCAGGGGCAAAT GGTACATCAGGCCATATCACCTAGAACTTTAAATGCATGGGTAAAAGTAGTAGA AGAGAAGGCTTTCAGCCCAGAAGTAATACCCATGTTTTCAGCATTATCAGAAGG AGCCACCCCACAAGATTTAAATACCATGCTAAACACAGTGGGGGGACATCAAGC AGCCATGCAAATGTTAAAAGAGACCAGAGGAAGCTGCAGAATGGGATAGATTG CATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAGATGAGAGAACCAAGGGG AAGTGACATAGCAGGAACTACTAGTACCCTTCAGGAACAAATAGGATGGATGA CACATAATCCACCTATCCCAGTAGGAGAAATCTATAAAAGATGGATAATCCTGG GATTAAATAAAATAGTAAGAATGTATAGCCCTACCAGCATTCTGGACATAAGAC AAGGACCAAAGGAACCCTTTAGAGACTATGTAGACCGATTCTATAAAACTCTAA GAGCCGAGCAAGCTTCACAAGAGGTAAAAAATTGGATGACAGAAACCTTGTTG GTCCAAAATGCGAACCCAGATTGTAAGACTATTTTAAAAGCATTGGGACCAGGA GCGACACTAGAAGAAATGATGACAGCATGTCAGGGAGTGGGGGGACCCGGCCA TAAAGCAAGAGTTTTGGCTGAAGCAATGAGCCAAGTAACAAATCCAGCTACCAT AATGATACAGAAAGGCAATTTTAGGAACCAAAGAAAGACTGTTAAGTGTTTCAA TTGTGGCAAAGAAGGGCACATAGCCAAAAATTGCAGGGCCCCTAGGAAAAAGG GCTGTTGGAAATGTGGAAAGGAAGGACACCAAATGAAAGATTGTACTGAGAGA CAGGCTAATTTTTTAGGGAAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAA TTTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTTCAGGTT TGGGGAAGAGACAACAACTCCCTCTCAGAAGCAGGAGCCGATAGACAAGGAAC TGTATCCTTTAGCTTCCCTCAGATCACTCTTTGGCAGCGACCCCTCGTCACAAT AAAGATAGGGGGGCAATTAAAGGAAGCTCTATTAGATACAGGAGCAGATGATA CAGTATTAGAAGAAATGAATTTGCCAGGAAGATGGAAACCAAAAATGATAGGG GGAATTGGAGGTTTTATCAAAGTAAGACAGTATGATCAGATACTCATAGAAATC TGCGGACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATA ATTGGAAGAAATCTGTTGACTCAGATTGGCTGCACTTTAAATTTTCCCATTAGT CCTATTGAGACTGTACCAGTAAAATTAAAGCCAGGAATGGATGGCCCAAAAGTT AAACAATGGCCATTGACAGAAGAAAAAATAAAAGCATTAGTAGAAATTTGTACA GAAATGGAAAAGGAAGGAAAAATTTCAAAAATTGGGCCTGAAAATCCATACAA TACTCCAGTATTTGCCATAAAGAAAAAAGACAGTACTAAATGGAGAAAATTAGT AGATTTCAGAGAACTTAATAAGAGAACTCAAGATTTCTGGGAAGTTCAATTAGG AATACCACATCCTGCAGGGTTAAAACAGAAAAAATCAGTAACAGTACTG...

HIV-1 NL4-3 p83-2 continued	GATGTGGGCGATGCATATTTTTCAGTTCCCTTAGATAAAGACTTCAGGAAGTAT ACTGCATTTACCATACCTAGTATAAACAATGAGACACCAGGGATTAGATATCAG TACAATGTGCTTCCACAGGGATGGAAAGGATCACCAGCAATATTCCAGTGTAGC ATGACAAAAATCTTAGAGCCTTTTAGAAAACAAAATCCAGACATAGTCATCTAT CAATACATGGATGATTTGTATGTAGGATCTGACTTAGAAATAGGGCAGCATAG AACAAAAATAGAGGAACTGAGACAACATCTGTTGAGGTGGGGATTTACCACAC CAGACAAAAAACATCAGAAAGAACCTCCATTCCTTTGGATGGGTTATGAACTCC ATCCTGATAAATGGACAGTACAGCCTATAGTGCTGCCAGAAAAGGACAGCTGG ACTGTCAATGACATACAGAAATTAGTGGGAAAATTGAATTGGGCAAGTCAGAT TTATGCAGGGATTAAAGTAAGGCAATTATGTAAACTTCTTAGGGGAACCAAAG CACTAACAGAAGTAGTACCACTAACAGAAGAAGCAGAGCTAGAACTGGCAGAA AACAGGGAGATTCTAAAAGAACCGGTACATGGAGTGTATTATGACCCATCAAAA GACTTAATAGCAGAAATACAGAAGCAGGGGCAAGGCCAATGGACATATCAAAT TTATCAAGAGCCATTTAAAAATCTGAAAACAGGAAAGTATGCAAGAATGAAGG GTGCCCACACTAATGATGTGAAACAATTAACAGAGGCAGTACAAAAAATAGCCA CAGAAAGCATAGTAATATGGGGAAAGACTCCTAAATTTAAATTACCCATACAAA AGGAAACATGGGAAGCATGGTGGACAGAGTATTGGCAAGCCACCTGGATTCCT GAGTGGGAGTTTGTCAATACCCCTCCCTTAGTGAAGTTATGGTACCAGTTAGAG AAAGAACCCATAATAGGAGCAGAAACTTTCTATGTAGATGGGGCAGCCAATAG GGAAACTAAATTAGGAAAAGCAGGATATGTAACTGACAGAGGAAGACAAAAAG TTGTCCCCCTAACGGACACAACAAATCAGAAGACTGAGTTACAAGCAATTCATC TAGCTTTGCAGGATTCGGGATTAGAAGTAAACATAGTGACAGACTCACAATATG CATTGGGAATCATTCAAGCACAACCAGATAAGAGTGAATCAGAGTTAGTCAGTC AAATAATAGAGCAGTTAATAAAAAAGGAAAAAGTCTACCTGGCATGGGTACCA GCACACAAAGGAATTGGAGGAAATGAACAAGTAGATAAATTGGTCAGTGCTGG AATCAGGAAAGTACTATTTTTAGATGGAATAGATAAGGCCCAAGAAGAACATG AGAAATATCACAGTAATTGGAGAGCAATGGCTAGTGATTTTAACCTACCACCTG TAGTAGCAAAAGAAATAGTAGCCAGCTGTGATAAATGTCAGCTAAAAGGGGAA GCCATGCATGGACAAGTAGACTGTAGCCCAGGAATATGGCAGCTAGATTGTAC ACATTTAGAAGGAAAAGTTATCTTGGTAGCAGTTCATGTAGCCAGTGGATATAT AGAAGCAGAAGTAATTCCAGCAGAGACAGGGCAAGAAACAGCATACTTCCTCTT AAAATTAGCAGGAAGATGGCCAGTAAAAACAGTACATACAGACAATGGCAGCA ATTTCACCAGTACTACAGTTAAGGCCGCCTGTTGGTGGGCGGGGATCAAGCAG GAATTTGGCATTCCCTACAATCCCCAAAGTCAAGGAGTAATAGAATCTATGAAT AAAGAATTAAAGAAAATTATAGGACAGGTAAGAGATCAGGCTGAACATCTTAA GACAGCAGTACAAATGGCAGTATTCATCCACAATTTTAAAAGAAAAGGGGGGA TTGGGGGGTACAGTGCAGGGGAAAGAATAGTAGACATAATAGCAACAGACATA CAAACTAAAGAATTACAAAAACAAATTACAAAAATTCAAAATTTTCGGGTTTAT TACAGGGACAGCAGAGATCCAGTTTGGAAAGGACCAGCAAAGCTCCTCTGGAA AGGTGAAGGGGCAGTAGTAATACAAGATAATAGTGACATAAAAGTAGTGCCAA GAAGAAAAGCAAAGATCATCAGGGATTATGGAAAACAGATGGCAGGTGATGAT TGTGTGGCAAGTAGACAGGATGAGGATTAACACATGGAAAAGATTAGTAAAAC ACCATATGTATATTTCAAGGAAAGCTAAGGACTGGTTTTATAGACATCACTATG AAAGTACTAATCCAAAAATAAGTTCAGAAGTACACATCCCACTAGGGGATGCTA AATTAGTAATAACAACATATTGGGGTCTGCATACAGGAGAAAGAGACTGGCAT TTGGGTCAGGGAGTCTCCATAGAATGGAGGAAAAAGAGATATAGCACACAAGT AGACCCTGACCTAGCAGACCAACTAATTCATCTGCACTATTTTGATTGTTTTTC AGAATCTGCTATAAGAAATACCATATTAGGACGTATAGTTAGTCCTAGGTGTGA ATATCAAGCAGGACATAACAAGGTAGGATCTCTACAGTACTTGGCACTAGCAGC ATTAATAAAACCAAAACAGATAAAGCCACCTTTGCCTAGTGTTAGGAAACTGAC AGAGGACAGATGGAACAAGCCCCAGAAGACCAAGGGCCACAGAGGGAGCCATA CAATGAATGGACACTAGAGCTTTTAGAGGAACTTAAGAGTGAAGCTGTTAGAC ATTTTCCTAGGATATGGCTCCATAACTTAGGACAACATATCTATGAAACTTACG GGGATAC...

HIV-1 NL4-3 p83-2 continued	TTGGGCAGGAGTGGAAGCCATAATAAGAATTCGGTCACCATCCTCGCTCACTGA CTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGG CGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGA GCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTT TTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCA GAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAA GCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCG CCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATC TCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCG TTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGG TAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGA GCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGG CTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTT CGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCG GTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAA GAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCA CGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTT TTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGG TCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTA TTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGG GAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCA CCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAG AAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGA AGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGC TACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGG TTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGT TAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATC ACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAG ATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTAT GCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACA TAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACT CTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACC CAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAAC AGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAA TACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTC TCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTC CGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCA TGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGT

Table D.11: Scaffold sequences for DNA-scaffolded wireframe origami used to capture tRNA or HIV 5'UTR.

Name	**Sequence**
T52 scaffold Gv3	GTTTAAGGTCACATCGCATGTTGCACAACTATTGCCTTTGTTTTGGCAGTCGTC TATCTAGACGGCTACTACCGGTTTGCTATCCTGATGACTAATCTTAGATTGAGC TAGCGGCCCTTAATCAATGCCTTGAACTATTACACTAATCTAACTCAGAACAGA CCAAGTGTCCCATTCGTGTACTGAGGAGATATCAGCCTCTGGTTAAGGAAAACG CGGGTCTGGTCCGGCACCCGATATTCTTCTTACAACTCACTGTTAACTGTAAGT ACCAGACTCCCAGATGGCATTGCGGGCTTGCGAATGAAGGTAACACAATATTA GTTTGGTTTAAGCAGATGGGACGTTCTGCTAACTGGATAGCCAGCAGATTGAA AGGCATGTGTATCGCCGCCGTACCTGGTTAGACAGATCACCGCCACTTCGTGGA CCGACGTTAGAGACCTCCGTGAGGGTTAAGAGCACTGTTTCCAAACCCTACCGC AACCGCAATGCCATCAATATGTGTGGGGTAGAATGGATGAGCGAATGCTCGTA TCTATTATGTGGCATATTAGAGTGTACTAGTCGACGCGTGGCCAAGTTGTATAT TCGATCGGGCCTCGGTATAGAGTGAAATCCAAA
PP63 and T96 scaffolds	#pF1A (see Table D.1)
PB73 and TT42 scaffolds	#phPB84 (see Table D.1)

Table D.12: Staple sequences used to fold the DNA-scaffolded tetrahedron with 52-bp edge length (T52). 3 staples have four versions each: with no bait sequence extension, or a 4-, 8-, or 12-nt bait sequence extension following a 'TTT' spacer. Only one version of each staple was used per staple pool, depending on the experiment design. Likewise, some experiments replaced staple 9 with a biotinylated version.

Staple ID	**Length (bp)**	**Staple sequences**
1	33	GATGTGACCTTAAACTTTGGATTACGAATGGGA
3	78	GATTAGTCATCTTTTTAGGATAGCAAAAAACAAAGGCTTTTTAATAGT TGTGGATTAGTGTAATTTTTTAGTTCAAGG
5	33	TAGCCGTCTCGCTCATCCATTCTAATAGATACG
7	31	AAGGGCCTTACAGTTAACAGTGATCTGGGAG
8	78	TGTCTAACCAGTTTTTGTACGGCGGCAGGCTGATATCTTTTTTCCTCA GTACTCACTCTATACTTTTTCGAGGCCCGA
10	20	GCTGGCTATCGCCGGACCAG

11	78	CAGTTAGCAGATTTTTACGTCCCATCATTCGCAAGCCTTTTTCGCAATGCCAGTTGTAAGAAGTTTTTAATATCGGGT
12	31	TTACCTTCTGCTTAAACCAAACTAGGTTTGG
13	31	AAACAGTGTTGCGGTTGCGGTAGATATTGTG
14	42	TTGGCCACGTCCACGAAGTGGCGGTGATCTCGAATATACAAC
15	20	GCGTCGACTATCTAACGTCG
16	78	ACCCCACACATTTTTTATTGATGGCACTCTTAACCCTTTTTTCACGGAGGTCGTACACTCTAATTTTTTATGCCACAT
9	42	TTCAATCTACCCGCGTTTTCCTTAACCAGGATACACATGCCT
9 (biotinylated)	67	/5BioTEG/TTTTCAATCTACCCGCGTTTTCCTTAACCAGGATACACATGCCT
2 (no bait)	29	CACTTGGTCTGTTCTGAGTTACAACATGC
4 (no bait)	29	AGCATTAGATAGACGACTGCCACCGGTAG
6 (no bait)	31	TCTGGTACGCTAGCTCAATCTAACATTGATT
2 (4nt bait)	36	CACTTGGTCTGTTCTGAGTTACAACATGCTTTTTGA
4 (4nt bait)	36	AGCATTAGATAGACGACTGCCACCGGTAGTTTTGGT
6 (4 nt bait)	38	TCTGGTACGCTAGCTCAATCTAACATTGATTTTTTCTA
2 (8 nt bait)	40	CACTTGGTCTGTTCTGAGTTACAACATGCTTTTAGGTGTG
4 (8 nt bait)	40	AGCATTAGATAGACGACTGCCACCGGTAGTTTGGAGTTTG
6 (8 nt bait)	42	TCTGGTACGCTAGCTCAATCTAACATTGATTTTTAATGGAGG
2 (12 nt bait)	44	CACTTGGTCTGTTCTGAGTTACAACATGCTTTTTTAGGTGTGGA
4 (12 nt bait)	44	AGCATTAGATAGACGACTGCCACCGGTAGTTTAAGGAGTTTGGT
6 (12 nt bait)	46	TCTGGTACGCTAGCTCAATCTAACATTGATTTTTCCAATGGAGGAC

Table D.13: Staple sequences used to fold the DNA-scaffolded pentagonal bipyramid with 73-bp edge length (PB73) for capture of HIV 5'UTR RNA. Many staples have two versions each: with no bait sequence extension, or one of five 8-nt bait sequence extensions following a 'TTT' spacer. Only one version of each staple was used per staple pool, depending, for a total of five baited staples in each pool for HIV 5'UTR capture. For staples with a bait sequence extension, the bait ID is given in parentheses next to the Staple ID. Every pool incorporated bait 5 on staple 1, and a biotinylated staple 37.

Staple ID	**Length (bp)**	**Staple sequences**
1 (bait 5)	53	ACGGGGAGTCAGGTTGCGTTGCGCTCGACGTAGCTCTACACGTTTGCAAGCTT
2	42	GGGCGCACATTAACAACTATGGATGAACGAAATACCGCTACA
3	20	GACAGATCGCGCTTAATGCG
4	78	CGTAACCACCATTTTTCACCCGCCGCTGAGATAGGTGTTTTTCCTCACTGATTTAAAACTTCATTTTTTTTTTAATTT
5	20	TCACGCTGCGAAGCGAAAGG
6	32	TCCTACTTCTGGCAAGTGTAGCGGAGCGGGCG
7	32	CTAGGGCGTAAAAGAAAAGAGGTTTTGTGCAA
8	52	AACGTGAGTTTTTTTTTCGTTCCACTTGGCGAGAAAGTTTTTGAAGGGAAGA
9	42	GACCCCAATCATAGACGGGGAAAGCCGGCGAACGGAGCGTCA
10	42	CGATTTAGAGCTTTGTACCCCGGTTGTAATTCGCGGAGCCCC
11	20	GTTAAATTTTAACCCTAAAG
12	52	GCGAAAAACCGTTTTTTCTATCAGGGTGCCGTAAAGCTTTTTACTAAATCGG
13	20	GGGGTCGAGGCGATGGCCCA
14	32	TATTCCTACCTAATCAAGTTTTTTCTACGTGA
15	32	ACCATCACGATTAGAAGGCGTTGTCTTCCCTA
16	20	ACGTCAAAGGCAGTTTGGAA
17	32	ATGATAATAAGAACGTGGACTCCACAAGAGTC

18	32	CACTATTAAATGGTTTCTTAGACGTTAATGTC
19	52	TGTTAAATCAGTTTTTCTCATTTTTTAGATAGGGTTGTTTTTAGTGTTGTTC
20	20	GAATAGACCGAACCAATAGG
21	32	GAAGGAGCCCCTTATAAATCAAAACCGAAATC
22	32	GGCAAAATTAACCGCTTTTTTGCAGGAGGACC
23	20	AAAATCCCTTAAAAGGATCT
24	32	AATGAAGCTGATAATCTCATGACCAGGTGAAG
25	32	ATCCTTTTCATACCAAACGACGAGCGGAGCTG
26	20	TTAGATTGATTAAGCATTGG
27	32	ACCACTTATTTACTCATATATACTTAACTGTC
28	32	AGACCAAGTGCGCTCGGCCCTTCCGTTGCAGG
29	52	AGCCCTCCCGTTTTTTATCGTAGTTAGAAGGATTAGTTTTTTGACGTAACTC
30	42	GGTTACATCGAACCAGCACTGGGGCCAGATGGTACACGAGTG
31	42	CGCGGTATCATTGTGGATCTCAACAGCGGTAAGAGTGGGTCT
32	20	TCCTTGAGAGGCCGGTGAGC
33	52	TTTATTGCTGATTTTTTAAATCTGGATTTTCGCCCCGTTTTTAAGAACGTTT
34	20	GGCTGGCTGGGGCGGATAAA
35	52	ATTGACGCCGGTTTTTGCAAGAGCAAACAATTAATAGTTTTTACTGGATGGA
36	42	CCGCATACACTATACTTACTCTAGCTTCCCGGCACTCGGTCG
37 (biotinylated)	54	TGGTTGAGCAAACTATTAACTGGCGAACTTCTCAGAATGACTTTT/3BioTEG/
38	20	TACTCACCAGCAACGTTGCG
39	52	GATCATGTAACTTTTTTCGCCTTGATCGATGCCTGTATTTTTGCAATGGCAA
40	20	CGTGACACCACGTTGGGAAC
41	20	CAACATGGGGGACAACGATC
42	52	TCACAGAAAAGTTTTTCATCTTACGGCACTGCGGCCATTTTTACTTACTTCT
43	20	TGAGTGATAAATGGCATGAC

44	32	TTTTCTAAAGTGCTGCCATAACCAAGTAAGAG
45	32	AATTATGCATACATTCAAATATGTTTGTTTAT
46	20	ATTATCCCGTTCCAATGATG
47	32	AAGGAAGATGCTATGTGGCGCGGTAGCACTTT
48	32	TAAAGTTCGTATGAGTATTCAACATATTGAAA
49	52	CTATACAAACTTTTTTAAGAGACAAAATGCTGAAGATTTTTTCAGTTG GGTG
50	42	CAAAATCTGGCGCAAACGCTGGTGAAAGTAAAAGGCTTTAAA
51	42	TTTGCTCACCCAGCCGGGTGGCTCAACAACAGAGTTCCTGTT
52	20	AATAAGTAAAGCATTTTGCC
53	78	GAGACAATAACTTTTTCCTGATAAATGCCCTTATTCCTTTTTCTTTTTT GCGTTACATGCGATTTTTTGTGACCTTAA
54	20	TTTCCGTGTCGCTTCAATAA
55	20	ATCCGCTCATGAACCCCTAT
56	52	AAGGGCCTCGTTTTTTGATACGCCTACTTTTCGGGGATTTTTAATGTG CGCG
57	20	TCAGGTGGCATTTTTATAGG
58	20	AATTCGACGAACGAATAGCC
59	32	TCAAACGGTGAGGGGACGACGACGTCTCCACC
60	32	CAAGAGTTGTAGGAAGCCGCATAAATCGGCCT
61	42	GTTTGCACACGTCAACCTTGCACTCTGAATCTGTCCGGGTTG
62	42	TACGAGCGAACGATGGCGGACATCTAGTGATAACAAACCAGG
63	20	CATGTTACTTGAACCAGGGA
64	52	AAGTCAAATCCTTTTTTATCTGTGAAGGTCAGTTATCTTTTTGGACCA CGTG
65	20	GTGTTGTAGACAGCACCTGT
66	52	GACCCAAGAAATTTTTGCATGGTAGCTATGTTCGCCCTTTTTGCATTT TGCG
67	20	TTTGATACTCAAATACGTAT
68	20	GGTCGTTAATTATTCATTCA

Substitutions for arrangement ‘B3’

10 (bait 1)	53	CGATTTAGAGCTTTGTACCCCGGTTGTAATTCGCGGAGCCCCTTTTCAAGTCC
36 (bait 2)	53	CCGCATACACTATACTTACTCTAGCTTCCCGGCACTCGGTCGTTTTACTCACC
68 (bait 3)	31	GGTCGTTAATTATTCATTCATTTCACACAAC
51 (bait 4)	53	TTTGCTCACCCAGCCGGGTGGCTCAACAACAGAGTTCCTGTTTTTGATTTTCC

Substitutions for arrangement 'B4'

47 (bait 1)	43	AAGGAAGATGCTATGTGGCGCGGTAGCACTTTTTTTCAAGTCC
32 (bait 2)	31	TCCTTGAGAGGCCGGTGAGCTTTTACTCACC
62 (bait 3)	53	TACGAGCGAACGATGGCGGACATCTAGTGATAACAAACCAGGTTTCACACAAC
27 (bait 4)	43	ACCACTTATTTACTCATATATACTTAACTGTCTTTGATTTTCC

Substitutions for arrangement 'B5'

68 (bait 1)	31	GGTCGTTAATTATTCATTCATTTTCAAGTCC
5 (bait 2)	31	TCACGCTGCGAAGCGAAAGGTTTTACTCACC
61 (bait 3)	53	GTTTGCACACGTCAACCTTGCACTCTGAATCTGTCCGGGTTGTTTCACACAAC
60 (bait 4)	43	CAAGAGTTGTAGGAAGCCGCATAAATCGGCCTTTTGATTTTCC

Substitutions for arrangement 'B6'

17 (bait 1)	43	ATGATAATAAGAACGTGGACTCCACAAGAGTCTTTTCAAGTCC
54 (bait 2)	31	TTTCCGTGTCGCTTCAATAATTTTACTCACC
57 (bait 3)	31	TCAGGTGGCATTTTTATAGGTTTCACACAAC
23 (bait 4)	31	AAAATCCCTTAAAAGGATCTTTTGATTTTCC

Substitutions for arrangement 'B7'

11 (bait 1)	31	GTTAAATTTTAACCCTAAAGTTTTCAAGTCC
38 (bait2)	31	TACTCACCAGCAACGTTGCGTTTTACTCACC
16 (bait 3)	31	ACGTCAAAGGCAGTTTGGAATTTCACACAAC

63 (bait 4)	31	CATGTTACTTGAACCAGGGATTTGATTTTCC
Substitutions for arrangement 'B8'		
52 (bait 1)	31	AATAAGTAAAGCATTTTGCCTTTTCAAGTCC
2 (bait 2)	53	GGGCGCACATTAACAACTATGGATGAACGAAATACCGCTACATTTTACTCACC
25 (bait 3)	43	ATCCTTTTCATACCAAACGACGAGCGGAGCTGTTTCACACAAC
61 (bait 4)	53	GTTTGCACACGTCAACCTTGCACTCTGAATCTGTCCGGGTTGTTTGATTTTCC

Table D.14: Staple sequences used to fold the tetrahedron with 94-bp edge length (T94) for Ext tRNA capture. Substitutions for staples with 8-nt bait extensions are indicated by solution (sol.) number, corresponding to the solution output from the bait placement algorithm discussed in Chapter 3. Each solution substitutes 3 staples with bait-extended staples.

Staple ID	**Length (bp)**	**Staple sequences**
1	41	CATGTATTTTTACTCGCCTTGATCGTTGGGAGTAGCAATGG
2	31	TGCTGCCATAATTTTTCCATGAGTGAGGGAT
3	42	GCACAACATGGTAACACTGCGGCCAACTTACTCCGCTTTTTT
4	59	TTTGGGGTCGAGGTGCCGGACCGAAGGAGCTAATCTGACAACGATCGGAGTAAAGCACT
5	25	AAATCGGAACCCTAAAATCAAGTTT
6	41	GGCCCATTTTTCTACGTGAACCATCACCCTAGGGAGCCCCC
7	31	CGAGATAGGGTTTTTTTGAGTGTTGTGCGAT
8	42	CGTCTATCAGGTCCAGTTTGGAACAAGAGTCCGGCGAAAAAC
9	59	TACTCATATATACTTTAGACTCCAACGTCAAAGACTATTAAAGAACGTGGATTGATTTA
10	25	AAACTTCATTTTTAATGACCAAGTT
11	41	CACTGATTTTTTTAAGCATTGGTAACTGTCATTAAAAGGAT

12	57	CTAGGTGAAGATTTTTTCCTTTTTGAGGGCCAGATGGTTTTTTTAAGCCCTCCTGCCT
13	42	GCTGAGATAGGCGTATCGTAGTTATCTACACGTAGACAGATC
14	59	GAATGAAGCCATACCAACTATGGATGAACGAAAACGGGGAGTCAGGCAAACGACGAGCG
15	25	TGACACCACGATGCCTACCGGAGCT
16	42	TTGCAGCACTGTAATCTCATGACCAAAATCCCCGCGGTATCA
17	42	AGCGTGGGTCTTTAACGTGAGTTTTCGTTCCAGGAGCCGGTG
18	42	CTGATAAATCTCTGAGCGTCAGACCCCAATCATGGTTTATTG
19	41	TTCTGCTTTTTGCTCGGCCCTTCCGGCTGGCTATGTACCCC
20	57	GGTTGTAATTCTTTTTGCGTTAAATTGGTCACGCTGCTTTTTGCGTAACCACACCAC
21	42	AAAGTTGCAGGCACACCCGCCGCGCTTAATGCGGAGGCGGAT
22	42	ATAGACTGGATGCCGCTACAGGGCGCACATTAGCAACAATTA
23	42	CTTCCCGATTGCGTTGCGCTCGTAAGAGACTACTTACTCTAG
24	46	CAACAACGTTGTTTTTCGCAAACTATTAACTGGCGAAATTATGCAG
25	42	GCAAGTGTAGCTTTGTTAAATCAGCTCATTTTTAGGGCGCTG
26	42	GGAGCGGGCGCTTAACCAATAGGCCGAAATCGGAAAGCGAAA
27	42	AAGGAAGGGAAGCAAAATCCCTTATAAATCAACGTGGCGAGA
28	46	GATTTAGAGCTTTTTTTGACGGGGAAAGCCGGCGAAAAGAATAGAC

Substitutions for different bait arrangement solutions

5 (sol. 0 bait)	35	GTTTAAATCGGAACCCTAAAATCAATTTAGGTGTG
8 (sol. 0 bait)	52	AAAACCGTCTATCAGGTCCAGTTTGGAACAAGAGTCCGGCGATTAATGGAGG
7 (sol. 0 bait)	41	TTTTTTGAGTGTTGTGCGATCGAGATAGGGTTTGGAGTTTG
8 (sol. 1 bait)	52	TCAGGTCCAGTTTGGAACAAGAGTCCGGCGAAAAACCGTCTATTTAGGTGTG
27 (sol. 1 bait)	52	GTGGCGAGAAAGGAAGGGAAGCAAAATCCCTTATAAATCAACTTAATGGAGG
5 (sol. 1 bait)	35	AACCCTAAAATCAAGTTTAAATCGGTTGGAGTTTG
16 (sol. 2 bait)	52	AATCCCCGCGGTATCATTGCAGCACTGTAATCTCATGACCAATTTAGGTGTG
10 (sol. 2 bait)	35	ACCAAGTTAAACTTCATTTTTAATGTTAATGGAGG

13 (sol. 2 bait)	52	TCTACACGTAGACAGATCGCTGAGATAGGCGTATCGTAGTTATTGGAGTTTG
7 (sol. 3 bait)	41	TTTTTTGAGTGTTGTGCGATCGAGATAGGGTTTTAGGTGTG
5 (sol. 3 bait)	35	CAAGTTTAAATCGGAACCCTAAAATTTAATGGAGG
8 (sol. 3 bait)	52	GAAAAACCGTCTATCAGGTCCAGTTTGGAACAAGAGTCCGGCTTGGAGTTTG
25 (sol. 4 bait)	52	TCATTTTTAGGGCGCTGGCAAGTGTAGCTTTGTTAAATCAGCTTTAGGTGTG
21 (sol. 4 bait)	52	CGCTTAATGCGGAGGCGGATAAAGTTGCAGGCACACCCGCCGTTAATGGAGG
18 (sol. 4 bait)	52	TTTATTGCTGATAAATCTCTGAGCGTCAGACCCCAATCATGGTTGGAGTTTG
20 (sol. 10 bait)	67	TAATTCTTTTTGCGTTAAATTGGTCACGCTGCTTTTTGCGTAACCACACCACGGTTGTTTAGGTGTG
18 (sol. 10 bait)	52	TTTATTGCTGATAAATCTCTGAGCGTCAGACCCCAATCATGGTTAATGGAGG
21 (sol. 10 bait)	52	GCTTAATGCGGAGGCGGATAAAGTTGCAGGCACACCCGCCGCTTGGAGTTTG
3 (sol. 11 bait)	52	CAACATGGTAACACTGCGGCCAACTTACTCCGCTTTTTTGCATTTAGGTGTG
2 (sol. 11 bait)	41	GAGTGAGGGATTGCTGCCATAATTTTTCCATTTAATGGAGG
15 (sol. 11 bait)	35	AGCTTGACACCACGATGCCTACCGGTTGGAGTTTG
2 (sol. 12 bait)	41	CCATAATTTTTCCATGAGTGAGGGATTGCTGTTTAGGTGTG
23 (sol. 12 bait)	52	ACTTACTCTAGCTTCCCGATTGCGTTGCGCTCGTAAGAGACTTTAATGGAGG
15 (sol. 12 bait)	35	CGGAGCTTGACACCACGATGCCTACTTGGAGTTTG
12 (sol. 13 bait)	67	TGAAGATTTTTTCCTTTTTGAGGGCCAGATGGTTTTTTAAGCCCTCCTGCCTCTAGGTTTAGGTGTG
10 (sol. 13 bait)	35	ACCAAGTTAAACTTCATTTTTAATGTTAATGGAGG
13 (sol. 13 bait)	52	TATCTACACGTAGACAGATCGCTGAGATAGGCGTATCGTAGTTTGGAGTTTG
25 (sol. 14 bait)	52	TGTTAAATCAGCTCATTTTTAGGGCGCTGGCAAGTGTAGCTTTTTAGGTGTG
20 (sol. 14 bait)	67	ACCACACCACGGTTGTAATTCTTTTTGCGTTAAATTGGTCACGCTGCTTTTTGCGTATTAATGGAGG
18 (sol. 14 bait)	52	TATTGCTGATAAATCTCTGAGCGTCAGACCCCAATCATGGTTTTGGAGTTTG

Table D.15: Staple sequences used to fold the truncated tetrahedron with 42-bp edge length (TT42) for Ext tRNA capture. Substitutions for staples with 8-nt bait extensions are indicated by solution (sol.) number, corresponding to the solution output from the bait placement algorithm discussed in Chapter 3. Each solution substitutes 3 staples with bait-extended staples.

Staple ID	**Length (bp)**	**Staple sequences**
1	55	GGTAGGAAGTTGCATAAGTGTTGTAGAAAGCAGGTACGAGCTTTTTTG AACGAAA
2	50	TGGACTCAAGAACCAGGGAAAACTCAAATCCTATCTGTGCAACGTCAA AG
3	55	TCCCTTATATTTCAAAAGAATAGACCGAGAAGTCCACTATTTTTTTTA AAGAACG
4	50	TTTGCGGTCCAGTTTGGAACAAGTAGGGTTGAGTGTTGTCATTTTGCC TT
5	54	TAACCCTGATAATGCTTCAATAATATTGATGTCGCCCTTATTTTTTTTC CCTTT
6	50	CCTTGCAGTATTCAACATTTCCGAAAAGGAAGAGTATGACTCTGAATC TG
7	47	CGATGTGACCTACTGGCGAACTTTTTTTACTTACTCTAACCCGCCGC
8	42	GCAAACTATTATAAACGAATAGCCTCTCCACCACAACGTTGC
9	45	CGGAGCTGAACGATGCCTGTTTTTTTAGCAATGGCACAAGAGTTT
10	50	AAGTTGCCGACGAGCGTGACACCATGAAGCCATACCAAAAGGACCACT TA
11	54	GGGCGCTAGTCGCTGGCAAGTGTAGCGGTGACTGGATGGATTTTTTG GCGGATA
12	42	AACAATTAATACACGCTGCGCGTAACCACCACGCTTCCCGGC
13	45	GGATCTCAAAAGAACGTTTTTTTTTTCCAATGATGAATGTGCGCG
14	35	TCGCCCCGCAGCGGTAAGATCCTTAATGACTTGGT
15	49	AACCATGAGTCGGTCGCCGCTTTTTTATACACTATTCTCAGGAGAGTT T
16	42	GCAAGAGCAACTGATAACACTGCGGCCAACTTTTGACGCCGG
17	47	TTCTTAGACGTGTGGCGCGGTATTTTTTTTTATCCCGTAACTTCTGAC

18	42	AGTTCTGCTATCAGGTGGCACTTTTCGGGGAAGCACTTTTAA
19	54	TTGGGGTCGTGTGCCGTAAAGCACTAAATTGAGCGTCAGATTTTTTCCCCCAATC
20	42	TTTCGTTCCACCGGAACCCTAAAGGGAGCCCCTAACGTGAGT
21	45	CTGGGGCCAAATCTCATGACTTTTTTCAAAATCCCTCGATTTAGA
22	42	TCCTTTTTGATGATGGTAAGCCCTCCCGTATCTAGGTGAAGA
23	47	AGACCAAGTTTCATTTTTAATTTTTTTTTAAAAGGATCGTAGTTATC
24	50	ATATGTAGATTGATTTAAAACTTACTCATATATACTTTACCCCGGTTGTA
25	41	CCTGTTTTTGCTCTTTCAGAAACGCTGGTGAAATGAGACAA
26	42	GTATCCGCTCAGTAAAAGATGCTGAAGATCAGATTCAAATAT
27	59	GAACCCCTATTTTTTTATTTTTCTAAATACTTGGGTGCACGAGTGGGTTACATCGAACT
28	56	AACGATCGGATTTCCGAAGGAGCTAACCGCATAGGTTAATGTCATGATATAATGGT
29	42	CGCCTATTTTTTTTTTTTGCACAACATGGGGGACCTCGTGATA
30	59	GAGGGGACGATTTCGAATTCGACGAAAGGGTCATGTAACTCGCCTTGTTCGTTGGGAAC
31	56	GCTTAATGCGTTTCTACAGGGCGCACATTAAACAACAGAGAATAAGTTATTACATG
32	42	GTGGCTCATTGCGTTGCGCTCATCGGCCAATCTGGCGCCCGG
33	41	TCTATACAAACTATTTGACAAAGCTTTAAACAATTCAAACG
34	41	TGCGCTCGGCCCTTTTGGCTGGCTGGTTTATTGAAAGGAGC
35	42	GGAAGAAAGCGCTGATAAATCTGGAGCCGGTGAGAAAGGAAG
36	59	GCTTGACGGGTTTAGCCGGCGAACGTGGCGAGCGTGGGTCTCGCGGTTTCATTGCAGCA
37	41	ATTCGCGTTAAATTTTTGTTAAATCAGCTCATTCAAGTTTT
38	42	ATCACCCTAATTTTTAACCAATAGGCCGAAATTACGTGAACC
39	41	GGCGAAAAACCGTTTTTCAGGGCGATGGCCCACCGGCAAAA
40	56	TACACGACGGTTTGTCAGGCAACTATGGATCTCACTGATTAAGCATTTTAACTGTC
41	58	ATAGGTGCGAACGAAATAGACAGAGCATGACAGTAAGAGAATTATGTTGTGCTGCCAT
42	46	TGAGTACTCACCATTTACAGAAAAGCATCTTACGGATGTCGCTGAG

Substitutions for different bait arrangement solutions

24 (sol. 0 bait)	60	TTAAAACTTACTCATATATACTTTACCCCGGTTGTAATATGTAGATTG ATTTTAGGTGTG
20 (sol. 0 bait)	52	CCCTAAAGGGAGCCCCTAACGTGAGTTTTCGTTCCACCGGAATTAATG GAGG
21 (sol. 0 bait)	55	TCTCATGACTTTTTTCAAAATCCCTCGATTTAGACTGGGGCCAAATTG GAGTTTG
20 (sol. 1 bait)	52	TTCCACCGGAACCCTAAAGGGAGCCCCTAACGTGAGTTTTCGTTTAGG TGTG
22 (sol. 1 bait)	52	GAAGATCCTTTTTGATGATGGTAAGCCCTCCCGTATCTAGGTTTAATG GAGG
36 (sol. 1 bait)	69	CAGCTTGACGGGTTTAGCCGGCGAACGTGGCGAGCGTGGGTCTCGCG GTTTCATTGCAGTTGGAGTTTG
6 (sol. 2 bait)	60	AAGGAAGAGTATGACTCTGAATCTGCCTTGCAGTATTCAACATTTCCG AATTTAGGTGTG
4 (sol. 2 bait)	60	TGGAACAAGTAGGGTTGAGTGTTGTCATTTTGCCTTTTTGCGGTCCAG TTTTAATGGAGG
2 (sol. 2 bait)	60	AAACTCAAATCCTATCTGTGCAACGTCAAAGTGGACTCAAGAACCAGG GATTGGAGTTTG
8 (sol. 3 bait)	52	ACTATTATAAACGAATAGCCTCTCCACCACAACGTTGCGCAATTTAGG TGTG
10 (sol. 3 bait)	60	GCCATACCAAAAGGACCACTTAAAGTTGCCGACGAGCGTGACACCATG AATTAATGGAGG
11 (sol. 3 bait)	64	CTGGATGGATTTTTTGGCGGATAGGGCGCTAGTCGCTGGCAAGTGTAG CGGTGATTGGAGTTTG
2 (sol. 4 bait)	60	AGAACCAGGGAAAACTCAAATCCTATCTGTGCAACGTCAAAGTGGACT CATTTAGGTGTG
4 (sol. 4 bait)	60	GTCCAGTTTGGAACAAGTAGGGTTGAGTGTTGTCATTTTGCCTTTTTG CGTTAATGGAGG
6 (sol. 4 bait)	60	ATGACTCTGAATCTGCCTTGCAGTATTCAACATTTCCGAAAAGGAAGA GTTTGGAGTTTG
7 (sol. 10 bait)	57	CGAACTTTTTTTACTTACTCTAACCCGCCGCCGATGTGACCTACTGGT TTAGGTGTG
31 (sol. 10 bait)	66	ATGCGTTTCTACAGGGCGCACATTAAACAACAGAGAATAAGTTATTAC ATGGCTTATTAATGGAGG
12 (sol. 10 bait)	52	AATACACGCTGCGCGTAACCACCACGCTTCCCGGCAACAATTTTGGAG TTTG
24 (sol. 11 bait)	60	TAATATGTAGATTGATTTAAAACTTACTCATATATACTTTACCCCGGT TGTTTAGGTGTG

37 (sol. 11 bait)	51	GTTAAATCAGCTCATTCAAGTTTTATTCGCGTTAAATTTTTTTAATGGAGG
20 (sol. 11 bait)	52	CCCTAAAGGGAGCCCCTAACGTGAGTTTTCGTTCCACCGGAATTGGAGTTTG
8 (sol. 12 bait)	52	CGTTGCGCAAACTATTATAAACGAATAGCCTCTCCACCACAATTTAGGTGTG
30 (sol. 12 bait)	69	GACGATTTCGAATTCGACGAAAGGGTCATGTAACTCGCCTTGTTCGTTGGGAACGAGGGTTAATGGAGG
10 (sol. 12 bait)	60	ACCAAAAGGACCACTTAAAGTTGCCGACGAGCGTGACACCATGAAGCCATTTGGAGTTTG
10 (sol. 13 bait)	60	CCGACGAGCGTGACACCATGAAGCCATACCAAAAGGACCACTTAAAGTTGTTTAGGTGTG
34 (sol. 13 bait)	51	GCCCTTTTGGCTGGCTGGTTTATTGAAAGGAGCTGCGCTCGTTAATGGAGG
12 (sol. 13 bait)	52	GTAACCACCACGCTTCCCGGCAACAATTAATACACGCTGCGCTTGGAGTTTG
6 (sol. 14 bait)	60	CAGTATTCAACATTTCCGAAAAGGAAGAGTATGACTCTGAATCTGCCTTGTTTAGGTGTG
33 (sol. 14 bait)	51	ACTATTTGACAAAGCTTTAAACAATTCAAACGTCTATACAATTAATGGAGG
2 (sol. 14 bait)	60	CTGTGCAACGTCAAAGTGGACTCAAGAACCAGGGAAAACTCAAATCCTATTTGGAGTTTG

Table D.16: Staple sequences used to fold the pentagonal bipyramid with 73-bp edge length (PB73) for Ext tRNA capture. Substitutions for staples with 8-nt bait extensions are indicated by solution (sol.) number, corresponding to the solution output from the bait placement algorithm discussed in Chapter 3. Each solution substitutes 3 staples with bait-extended staples.

Staple ID	**Length (bp)**	**Staple sequences**
1	47	ACTTGTTTTTGTCAGTTATCGGTCGTTAATGTTTTTACCCAAGAAAG
2	48	AAAGAAAAGAGGTTGGACCACGTGTATTCATTCATTGTGCTTTTTGCT
3	36	CACCCAGAGCATTTTGCCTTCCTGAATCCTACTTTA
4	41	GTGTCGTTTTTCCCTTATTCCCTTTTTTGCGAACGCTGGTG

5	26	CTCATGTTTTTAGACAATAACTTTCC
6	48	ATGAGTATTCAACACCTGATAAATGCTTCAATAATATTGACCCAAGAG
7	36	TTTGAGGGACGAATAGCCTCTCCAAAAAGGAAGAGT
8	40	TAAATTTTTTTACATGCGATGTGACCTTAAGACGACGACG
9	58	GTGACACCACGTTTTTATGCCTGTAGTGAATCTGTCTTTTTTTATACAAACTAATAAG
10	48	CTCAACAACAGAGAAGAGACAAAGCTTTAAACAAAATCTGATATTCCT
11	36	AGATTAGAAATACGTATCTTCCCTGCGCCCGGGTGG
12	47	CATGGTAGCAAGGCGTTGTTTTGATACTCTATTTTTTGTTCGCCCGC
13	57	TCCTTGAGAGTTTTTTTTTCGCCCCGATTATCCCGTATTTTTTTGACGCCGGATCCG
14	42	TTCAAATATGTGCAAGAGCAACTCGGTCGCCGTCTAAATACA
15	42	TTGTTTATTTTCATACACTATTCTCAGAATGAGAACCCCTAT
16	40	GGCACTTTTTTTTTCGGGGAAATGTGCGCGCTTGGTTGAG
17	58	GAGTGATAACATTTTTCTGCGGCCAAAACATGGGGGATTTTTTTCATGTAACTCAGGT
18	42	TTCTTAGACGTCGCCTTGATCGTTGGGAACCGTAATAATGGT
19	42	TAATGTCATGAGAGCTGAATGAAGCCATACCATTTTATAGGT
20	47	AATTCGACGAATTTTTAGGGCCTCGTGATACGCCTATAACGACGAGC
21	48	CTATGTGGCGCGGTAAGAACGTTTTCCAATGATGAGCACTTCAGACCA
22	36	AGTTTACTTAAGCATTGGTAACTGTTTAAAGTTCTG
23	40	TCGCTTTTTTGAGATAGGTGCCTCACTGATCATATATACT
24	58	AAATCCCTTAATTTTTCGTGAGTTTTCAGATGGTAAGTTTTTTCCCTCCCGTAACAGA
25	48	GATGAACGAAATAGTCGTAGTTATCTACACGACGGGGAGTGAATTATG
26	36	CAGTGCTGTGGCATGACAGTAAGACAGGCAACTATG
27	47	TACTCACCAGTTTTTTCACAGAAAAGCATCTTACGGACCATAACCAT
28	42	ACCTTGCACTCCAATGGCAACAACGTTGCGCAAGCGAACGAA
29	42	AACCAGGTACGAACTATTAACTGGCGAACTACAACCAGGGAA
30	41	TAGAAATTTTTGTCAAATCCTATCTGTGAAGTTACTCTAGC
31	57	TTAAATTTTTGTTTTTTTAAATCAGCCACGCTGCGCGTTTTTTAACCACCACTGTTG

32	42	AGCCGCATAAGACCCGCCGCGCTTAATGCGCCACGGGTAGGA
33	42	TCGGCCTTCAAGCTACAGGGCGCACATTAATTAGCACCTGTA
34	51	ATTTTGCGCGCGTTGCGCTCGACGTAGCGATTTTTTAGGATTAGTGATGTT
35	48	CCGCTTTTTTGCACCTTACTTCTGACAACGATCGGAGGACGACCACTT
36	36	ATGCGCTCGCGGATAAAGTTGCAGCGAAGGAGCTAA
37	46	TTCCCGGCAACTTTTTAATTAATAGACTGGATGGAGGGCCCTTCCG
38	42	AGCACTGGGGCCGTTCCACTGAGCGTCAGACCGTATCATTGC
39	42	TGGGTCTCGCGCCAATCATATGTACCCCGGTTCCGGTGAGCG
40	46	GCTGGCTGGTTTTTTTTATTGCTGATAAATCTGGAGGTAATTCGCG
41	48	ATCTAGTGATAACCACGTAACTCCCGGGTTGGTTTGCACACCGATTTA
42	36	GAGCTTGAACCCTAAAGGGAGCCCCGTCTGGCGGAC
43	47	CTAAATCGGACGGGGAAAGCCGGCGAACGTGTTTTTGCGAGAAAGGA
44	47	AAGGGTTTTTCGAAAAACCGGGGGTCGAGGTTTTTTGCCGTAAAGCA
45	48	TAATCAAGTTTTTTTCTATCAGGGCGATGGCCCACTACGTCGAACTGG
46	36	ATCTCAACCACGAGTGGGTTACATGAACCATCACCC
47	46	AAAGTAAAAGATTTTTTGCTGAAGATCAGTTGGGTGAGCGGTAAGA
48	42	AGTGTAGCGGTTCATTTTTTAACCAATAGGCCGGCGCTGGCA
49	42	GCGGGCGCTAGGAAATCGGCAAAATCCCTTATAGCGAAAGGA
50	51	AGGGAAGAAAAATCAAAAGAATAGACCGAGTTTTTTATAGGGTTGACGTCA
51	48	AACGTGGACTCCAAGTGTTGTTCCAGTTTGGAACAAGAGTGATCCTTT
52	36	TTGATAATAAAGGATCTAGGTGAACCACTATTAAAG
53	47	TTAGATTGATTTTTTTTAAAACTTCATTTTTAATTTACTCATGACCA

Substitutions for different bait arrangement solutions

29 (sol. 0 bait)	52	ACCAGGGAAAACCAGGTACGAACTATTAACTGGCGAACTACATTTAGGTGTG
31 (sol. 0 bait)	67	CCACCACTGTTGTTAAATTTTTGTTTTTTTAAATCAGCCACGCTGCGCGTTTTTTAATTAATGGAGG
32 (sol. 0 bait)	52	CGGGTAGGAAGCCGCATAAGACCCGCCGCGCTTAATGCGCCATTGGAGTTTG

38 (sol. 1 bait)	52	ATTGCAGCACTGGGGCCGTTCCACTGAGCGTCAGACCGTATCTTTAGGTGTG
25 (sol. 1 bait)	58	GACGGGGAGTGAATTATGGATGAACGAAATAGTCGTAGTTATCTACACTTAATGGAGG
23 (sol. 1 bait)	50	CACTGATCATATATACTTCGCTTTTTTGAGATAGGTGCCTTTGGAGTTTG
43 (sol. 2 bait)	57	AGAAAGGACTAAATCGGACGGGGAAAGCCGGCGAACGTGTTTTTGCGTTTAGGTGTG
42 (sol. 2 bait)	46	AAGGGAGCCCCGTCTGGCGGACGAGCTTGAACCCTATTAATGGAGG
46 (sol. 2 bait)	46	CATCTCAACCACGAGTGGGTTACATGAACCATCACCTTGGAGTTTG
14 (sol. 3 bait)	52	CGCCGTCTAAATACATTCAAATATGTGCAAGAGCAACTCGGTTTTAGGTGTG
6 (sol. 3 bait)	58	ATAAATGCTTCAATAATATTGACCCAAGAGATGAGTATTCAACACCTGTTAATGGAGG
13 (sol. 3 bait)	67	TTCGCCCCGATTATCCCGTATTTTTTTGACGCCGGATCCGTCCTTGAGAGTTTTTTTTTGGAGTTTG
19 (sol. 4 bait)	52	TACCATTTTATAGGTTAATGTCATGAGAGCTGAATGAAGCCATTTAGGTGTG
8 (sol. 4 bait)	50	CGACGTAAATTTTTTTACATGCGATGTGACCTTAAGACGATTAATGGAGG
28 (sol. 4 bait)	52	TGCACTCCAATGGCAACAACGTTGCGCAAGCGAACGAAACCTTTGGAGTTTG
9 (sol. 10 bait)	68	TGACACCACGTTTTTATGCCTGTAGTGAATCTGTCTTTTTTTATACAAACTAATAAGGTTTAGGTGTG
28 (sol. 10 bait)	52	GCAACAACGTTGCGCAAGCGAACGAAACCTTGCACTCCAATGTTAATGGAGG
7 (sol. 10 bait)	46	ACGAATAGCCTCTCCAAAAAGGAAGAGTTTTGAGGGTTGGAGTTTG
12 (sol. 11 bait)	57	TTGTTTTGATACTCTATTTTTTGTTCGCCCGCCATGGTAGCAAGGCGTTTAGGTGTG
1 (sol. 11 bait)	57	CCAAGAAAGACTTGTTTTTGTCAGTTATCGGTCGTTAATGTTTTTACTTAATGGAGG
33 (sol. 11 bait)	52	CTGTATCGGCCTTCAAGCTACAGGGCGCACATTAATTAGCACTTGGAGTTTG
31 (sol. 12 bait)	67	CACGCTGCGCGTTTTTTAACCACCACTGTTGTTAAATTTTTGTTTTTTTAAATCAGCTTTAGGTGTG
48 (sol. 12 bait)	52	TGGCAAGTGTAGCGGTTCATTTTTTAACCAATAGGCCGGCGCTTAATGGAGG

32 (sol. 12 bait)	52	TTAATGCGCCACGGGTAGGAAGCCGCATAAGACCCGCCGCGCTTGGAGTTTG
1 (sol. 13 bait)	57	GTCGTTAATGTTTTTACCCAAGAAAGACTTGTTTTTGTCAGTTATCGTTTAGGTGTG
12 (sol. 13 bait)	57	GTTCGCCCGCCATGGTAGCAAGGCGTTGTTTTGATACTCTATTTTTTTTTAATGGAGG
11 (sol. 13 bait)	46	TATCTTCCCTGCGCCCGGGTGGAGATTAGAAATACGTTGGAGTTTG
46 (sol. 14 bait)	46	CACGAGTGGGTTACATGAACCATCACCCATCTCAACTTTAGGTGTG
47 (sol. 14 bait)	56	GAAGATCAGTTGGGTGAGCGGTAAGAAAAGTAAAAGATTTTTTGCTTTAATGGAGG
21 (sol. 14 bait)	58	AATGATGAGCACTTCAGACCACTATGTGGCGCGGTAAGAACGTTTTCCTTGGAGTTTG

Table D.17: Staple sequences used to fold the pentagonal pyramid with 63-bp edge length (PP63) for Ext tRNA capture. Substitutions for staples with 8-nt bait extensions are indicated by solution (sol.) number, corresponding to the solution output from the bait placement algorithm discussed in Chapter 3. Each solution substitutes 3 staples with bait-extended staples.

Staple ID	**Length (bp)**	**Staple sequences**
1	59	GTGCACGAGTTTTTACATCGAACTGGATCGATGAGCACTTTTTAAAGTTCTGCTGGCAA
2	48	GAACGTTTTCCAATTCAACAGCGGTAAGATCCTTGAGAGTGAACAAGA
3	36	GTCCACTAGTGTTGTTCCAGTTTGTTTCGCCCCGAA
4	30	TAGACCTTTTTGATAGGGTTGATTAAAGAA
5	59	CGTGGACTCCTTGTCAAAGGGCGAAAAACTTCGCGTTAATTTTTTTGTTAAATCAAGAA
6	48	CCCTTATAAATCAAAGCTCATTTTTTAACCAATAGGCCGATACTTTAG
7	36	ATTGATTTCAAGTTTACTCATATAAATCGGCAAAAT
8	52	TTGGTATTTTTACTGTCAGACAAAACTTCATTTTTTTTTTAATTTAACGGCT
9	26	CGACGGTTTTTGGAGTCAGGCAAGCA
10	48	GTGCCTCACTGATTAACTATGGATGAACGAAATAGACAGATCCCGTAT

11	36	TGACGCCGATGTGGCGCGGTATTATCGCTGAGATAG
12	42	CCCGGTTGTAACGTCTATCAGGGCGATGGCCCTCATATGTAC
13	42	TCAGACCCCAAACTACGTGAACCATCACCCTACCACTGAGCG
14	30	CTTAACTTTTTGAGTTTTCGTTATCAAGTT
15	59	TTTTGGGGTCTTGTGCCGTAAAGCACTAAGGCAACAATTTTTTAGACTGGATGGAATCC
16	42	CTCATGACCAAAGGCGGATAAAGTTGCAGGACTTTTGATAAT
17	42	GTGAAGATCCTCACTTCTGCGCTCGGCCCTTCAAGGATCTAG
18	42	TCTAGCTTCCCATCGGAACCCTAAAGGGAGCCAACTACTTAC
19	42	ATTAACTGGCGCCCGATTTAGAGCTTGACGGGTGCGCAAACT
20	30	GTAGCATTTTTGGCAACAACGTGAAAGCCG
21	59	GCGAACGTGGTTGAAAGGAAGGGAAGAAAGATCATGTAATTTTGCCTTGATCGTTGCCT
22	48	GCGTGACACCACGATGGGAACCGGAGCTGAATGAAGCCATTGAGCGTG
23	36	GGTCTCGCATAAATCTGGAGCCGGACCAAACGACGA
24	52	GGCTGGTTTTTTTTATTGCTGGGTATCATTGCTTTTTAGCACTGGGGCTACA
25	42	ACAACATGGGGGCGAAAGGAGCGGGCGCTAGGGCTTTTTTGC
26	42	AGGAGCTAACCGCGCTGGCAAGTGTAGCGGTCGGAGGACCGA
27	30	CTTACTTTTTTTGACAACGATCACGCTGCG
28	59	CGTAACCACCTTCCCGCCGCGCTTAATGCAGAAAAGCATTTTTACGGATGGCATGCCAA
29	48	GTGATAACACTGCGGACAGTAAGAGAATTATGCAGTGCTGCCCGTATC
30	36	GTAGTTATCCAGATGGTAAGCCCTCCATAACCATGA
31	42	TCACCAGTCACGCCGCTACAGGGCGCACATTAGGTTGAGTAC
32	42	ATGACTTATTGCGTTGCGCTCCTGAAGAACACTATTCTCAGA
33	30	GAGCAATTTTTGGTCGCCGCATTCAGTTGG

Substitutions for different bait arrangement solutions

16 (sol. 0 bait)	52	TTTGATAATCTCATGACCAAAGGCGGATAAAGTTGCAGGACTTTTAGGTGTG
8 (sol. 0 bait)	62	TAATTTAACGGCTTTGGTATTTTTACTGTCAGACAAAACTTCATTTTTTTTTTTAATGGAGG

6 (sol. 0 bait)	58	TACTTTAGCCCTTATAAATCAAAGCTCATTTTTTAACCAATAGGCCGATTGGAGTTTG
12 (sol. 1 bait)	52	TGTAACGTCTATCAGGGCGATGGCCCTCATATGTACCCCGGTTTTAGGTGTG
4 (sol. 1 bait)	40	TAGGGTTGATTAAAGAATAGACCTTTTTGATTAATGGAGG
7 (sol. 1 bait)	46	AATATTGATTTCAAGTTTACTCATATAAATCGGCAATTGGAGTTTG
4 (sol. 2 bait)	40	AAAGAATAGACCTTTTTGATAGGGTTGATTTTTAGGTGTG
6 (sol. 2 bait)	58	CCAATAGGCCGATACTTTAGCCCTTATAAATCAAAGCTCATTTTTTAATTAATGGAGG
12 (sol. 2 bait)	52	ATCAGGGCGATGGCCCTCATATGTACCCCGGTTGTAACGTCTTTGGAGTTTG
19 (sol. 3 bait)	52	CGCAAACTATTAACTGGCGCCCGATTTAGAGCTTGACGGGTGTTTAGGTGTG
21 (sol. 3 bait)	69	GTAATTTTGCCTTGATCGTTGCCTGCGAACGTGGTTGAAAGGAAGGGAAGAAAGATCATTTAATGGAGG
22 (sol. 3 bait)	58	ATGAAGCCATTGAGCGTGGCGTGACACCACGATGGGAACCGGAGCTGATTGGAGTTTG
17 (sol. 4 bait)	52	GGCCCTTCAAGGATCTAGGTGAAGATCCTCACTTCTGCGCTCTTTAGGTGTG
7 (sol. 4 bait)	46	AAGTTTACTCATATAAATCGGCAAAATATTGATTTCTTAATGGAGG
24 (sol. 4 bait)	62	ACTGGGGCTACAGGCTGGTTTTTTTTATTGCTGGGTATCATTGCTTTTTAGCTTGGAGTTTG
12 (sol. 10 bait)	52	CTATCAGGGCGATGGCCCTCATATGTACCCCGGTTGTAACGTTTTAGGTGTG
5 (sol. 10 bait)	69	TCAAAGGGCGAAAAACTTCGCGTTAATTTTTTTGTTAAATCAAGAACGTGGACTCCTTGTTAATGGAGG
6 (sol. 10 bait)	58	ACCAATAGGCCGATACTTTAGCCCTTATAAATCAAAGCTCATTTTTTATTGGAGTTTG
5 (sol. 11 bait)	69	AAATCAAGAACGTGGACTCCTTGTCAAAGGGCGAAAAACTTCGCGTTAATTTTTTTGTTTTTAGGTGTG
6 (sol. 11 bait)	58	ACCAATAGGCCGATACTTTAGCCCTTATAAATCAAAGCTCATTTTTTATTAATGGAGG
12 (sol. 11 bait)	52	TGGCCCTCATATGTACCCCGGTTGTAACGTCTATCAGGGCGATTGGAGTTTG
2 (sol. 12 bait)	58	CAGCGGTAAGATCCTTGAGAGTGAACAAGAGAACGTTTTCCAATTCAATTTAGGTGTG
3 (sol. 12 bait)	46	CGAAGTCCACTAGTGTTGTTCCAGTTTGTTTCGCCCTTAATGGAGG

11 (sol. 12 bait)	46	ATTATCGCTGAGATAGTGACGCCGATGTGGCGCGGTTTGGAGTTTG
33 (sol. 13 bait)	40	TCGCCGCATTCAGTTGGGAGCAATTTTTGGTTTAGGTGTG
1 (sol. 13 bait)	69	GAACTGGATCGATGAGCACTTTTTAAAGTTCTGCTGGCAAGTGCACGAGTTTTTACATCTTAATGGAGG
11 (sol. 13 bait)	46	TTATCGCTGAGATAGTGACGCCGATGTGGCGCGGTATTGGAGTTTG
8 (sol. 14 bait)	62	GGTATTTTTACTGTCAGACAAAACTTCATTTTTTTTTTAATTTAACGGCTTTTTTAGGTGTG
24 (sol. 14 bait)	62	CACTGGGGCTACAGGCTGGTTTTTTTTATTGCTGGGTATCATTGCTTTTTAGTTAATGGAGG
17 (sol. 14 bait)	52	GATCTAGGTGAAGATCCTCACTTCTGCGCTCGGCCCTTCAAGTTGGAGTTTG

D.3 RNA-scaffolded 3D wireframe origami[1]

Table D.18: Scaffold sequences used for folded the RNA-scaffolded objects in Chapter 4, as well as the corresponding DNA template sequences.

Name	Sequence
EGFP gBlock	GCCAGTGAATTCTAATACGACTCACTATAGGTAGCTAAGGAGGTAAATAATGGTGAGCAAGG GCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGC CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAA GTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTA CGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGC CATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGAC CCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCG ACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAAC GTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAA CATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACG GCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCA ACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGC ATGGACGAGCTGTACAAGTAACTGCAGGCATGC
EGFP mRNA scaffold	AAGGGCGAGGAGCUGUUCACCGGGGUGGUGCCCAUCCUGGUCGAGCUGGACGGCGACGUAAA CGGCCACAAGUUCAGCGUGUCCGGCGAGGGCGAGGGCGAUGCCACCUACGGCAAGCUGACCC UGAAGUUCAUCUGCACCACCGGCAAGCUGCCCGUGCCCUGGCCCACCCUCGUGACCACCCUG ACCUACGGCGUGCAGUGCUUCAGCCGCUACCCCGACCACAUGAAGCAGCACGACUUCUUCAA GUCCGCCAUGCCCGAAGGCUACGUCCAGGAGCGCACCAUCUUCUUCAAGGACGACGGCAACU ACAAGACCCGCGCCGAGGUGAAGUUCGAGGGCGACACCCUGGUGAACCGCAUCGAGCUGAAG GGCAUCGACUUCAAGGAGGACGGCAACAUCCUGGGGCACAAGCUGGAGUACAACUACAACAG CCACAACGUCUAUAUCAUGGCCGACAAGCAGAAGAACGGCAUCAAGGUGAACUUCAAGAUCC GCCACAACAUCGAGGACGGCAGCGUGCAGCUCGCCGACCACUACCAGCAGAACACCCCCAUC GGCGACGGCCCCGUGCUGCUGCCCGACAACCACUACCUGAGCACCCAGUCCGCCCUGAGCAA AGACCCCAACGAGAAGCGCGAUCACAUGGUCCUGCUGGAGUUCGUGACCGCCGCCGGGAUCA CUCUCGGCAUGGACGAGCUGUACAAGUAACUGCAGGCAUGCAAGCUUGGCGUAAUCAUGGUC AUAGCUGUUUCCUGUGUGAAAUUGUUAUCCGCUCACAAUUCCACACAA
EGFP mRNA sequence	GGUAGCUAAGGAGGUAAAUAAUGGUGAGCAAGGGCGAGGAGCUGUUCACCGGGGUGGUGCC CAUCCUGGUCGAGCUGGACGGCGACGUAAACGGCCACAAGUUCAGCGUGUCCGGCGAGGGCG AGGGCGAUGCCACCUACGGCAAGCUGACCCUGAAGUUCAUCUGCACCACCGGCAAGCUGCCC GUGCCCUGGCCCACCCUCGUGACCACCCUGACCUACGGCGUGCAGUGCUUCAGCCGCUACCC CGACCACAUGAAGCAGCACGACUUCUUCAAGUCCGCCAUGCCCGAAGGCUACGUCCAGGAGC GCACCAUCUUCUUCAAGGACGACGGCAACUACAAGACCCGCGCCGAGGUGAAGUUCGAGGGC GACACCCUGGUGAACCGCAUCGAGCUGAAGGGCAUCGACUUCAAGGAGGACGGCAACAUCCU GGGGCACAAGCUGGAGUACAACUACAACAGCCACAACGUCUAUAUCAUGGCCGACAAGCAGA AGAACGGCAUCAAGGUGAACUUCAAGAUCCGCCACAACAUCGAGGACGGCAGCGUGCAGCUC GCCGACCACUACCAGCAGAACACCCCCAUCGGCGACGGCCCCGUGCUGCUGCCCGACAACCAC UACCUGAGCACCCAGUCCGCCCUGAGCAAAGACCCCAACGAGAAGCGCGAUCACAUGGUCCU GCUGGAGUUCGUGACCGCCGCCGGGAUCACUCUCGGCAUGGACGAGCUGUACAAGUAACUGC AGGCAUGCAAGCUUGGCGUAAUCAUGGUCAUAGCUGUUUCCUGUGUGAAAUUGUUAUCCGCU CACAAUUCCACACAACAUACG

[1]In the scaffold sequence table, "RNA scaffold" refers to the sequence used to design staples, while "RNA sequence" refers to the full sequence of the transcribed RNA molecule, which may be longer than the scaffold sequence alone.

23s rRNA rrlB RNA scaffold	GCCCUGUUUUUGCAGUCAGAGGCGAUGAAGGACGUGCUAAUCUGCGAUAAGCGUCGGUAAGG UGAUAUGAACCGUUAUAACCGGCGAUUUCCGAAUGGGGAAACCCAGUGUGUUUCGACACACU AUCAUUAACUGAAUCCAUAGGUUAAUGAGGCGAACCGGGGGAACUGAAACAUCUAAGUACCC CGAGGAAAAGAAAUCAACCGAGAUUCCCCCAGUAGCGGCGAGCGAACGGGGAGCAGCCCAGA GCCUGAAUCAGUGUGUGUGUUAGUGGAAGCGUCUGGAAAGGCGCGCGAUACAGGGUGACAGC CCCGUACACAAAAAUGCACAUGCUGUGAGCUCGAUGAGUAGGGCGGGACACGUGGUAUCCUG UCUGAAUAUGGGGGGACCAUCCUCCAAGGCUAAAUACUCCUGACUGACCGAUAGUGAACCAG UACCGUGAGGGAAAGGCGAAAAGAACCCCGGCGAGGGGAGUGAAAAAGAACCUGAAACCGUG UACGUACAAGCAGUGGGAGCACGCUUAGGCGUGUGACUGCGUACCUUUUGUAUAAUGGGUCA GCGACUUAUAUUCUGUAGCAAGGUUAACCGAAUAGGGGAGCCGAAGGGAAACCGAGUCUUAA CUGGGCGUUAAGUUGCAGGGUAUAGACCCGAAACCCGGUGAUCUAGCCAUGGGCAGGUUGAA GGUUGGGUAACACUAACUGGAGGACCGAACCGACUAAUGUUGAAAAAUUAGCGGAUGACUUG UGGCUGGGGGUGAAAGGCCAAUCAAACCGGGAGAUAGCUGGUUCUCCCCGAAAGCUAUUUAG GUAGCGCCUCGUGAAUUCAUCUCCGGGGGUAGAGCACUGUUUCGGCAAGGGGGUCAUCCCGA CUUACCAACCCGAUGCAAACUGCGAAUACCGGAGAAUGUUAUCACGGGAGACACACGGCGGG UGCUAACGUCCGUCGUGAAGAGGGAAACAACCCAGACCGCCAGCUAAGGUCCCAAAGUCAUG GUUAAGUGGGAAACGAUGUGGGAAGGCCCAGACAGCCAGGAUGUUGGCUUAGAAGCAGCCAU CAUUUAAAGAAAGCGUAAUAGCUCACUGGUCGAGUCGGCCUGCGCGGAAGAUGUAACGGGGC UAAACCAUGCACCGAAGCUGCGGCAGCGACGCUUAUGCGUUGUUGGGUAGGGGAGCGUUCUG UAAGCCUGCGAAGGUGUGCUGUGAGGCAUGCUGGAGGUAUCAGAAGUGCGAAUGCUGACAUA AGUAACGAUAAAGCGGGUGAAAAGCCCGCUCGCCGGAAGACCAAGGGUUCCUGUCCAACGUU AAUCGGGGCAGGGUGAGUCGACCCCUAAGGCGAGGCCGAAAGGCGUAGUCGAUGGGAAACAG GUUAAUAUUCCUGUACUUGGUGUUACUGCGAAGGGGGGACGGAGAAGGCUAUGUUGGCCGG GCGACGGUUGUCCCGGUUUAAGCGUGUAGGCUGGUUUUCCAGGCAAAUCCGGAAAAUCAAGG CUGAGGCGUGAUGACGAGGCACUACGGUGCUGAAGCAACAAAUGCCCUGCUUCCAGGAAAAG CCUCUAAGCAUCAGGUAACAUCAAAUCGUACCCCAAACCGACACAGGUGGUCAGGUAGAGAA UACCAAGGCGCUUGAGAGAACUCGGGUGAAGGAACUAGGCAAAAUGGUGCCGUAACUUCGGG AGAAGGCACGCUGAUAUGUAGGUGAAGCGACUUGCUCGUGGAGCUGAAAUCAGUCGAAGAUA CCAGCUGGCUGCAACUGUUUAUUAAAAACACAGCACUGUGCAAACACGAAAGUGGACGUAUA CGGUGUGACGCCUGCCCGGUGCCGGAAGGUUAAUUGAUGGGGUUAGCGCAAGCGAAGCUCUU GAUCGAAGCCCCGGUAAACGGCGGCCGUAACUAUAACGGUCCUAAGGUAGCGAAAUUCCUUG UCGGGUAAGUUCCGACCUGCACGAAUGGCGUAAUGAUGGCCAGGCUGUCUCCACCCGAGACU CA

23s rRNA rrlB RNA sequence

GCCCUGUUUUUGCAGUCAGAGGCGAUGAAGGACGUGCUAAUCUGCGAUAAGCGUCGGUAAGG
UGAUAUGAACCGUUAUAACCGGCGAUUUCCGAAUGGGGAAACCCAGUGUGUUUCGACACACU
AUCAUUAACUGAAUCCAUAGGUUAAUGAGGCGAACCGGGGGAACUGAAACAUCUAAGUACCC
CGAGGAAAAGAAAUCAACCGAGAUUCCCCCAGUAGCGGCGAGCGAACGGGGAGCAGCCCAGA
GCCUGAAUCAGUGUGUGUGUUAGUGGAAGCGUCUGGAAAGGCGCGCGAUACAGGGUGACAGC
CCCGUACACAAAAAUGCACAUGCUGUGAGCUCGAUGAGUAGGGCGGGACACGUGGUAUCCUG
UCUGAAUAUGGGGGGACCAUCCUCCAAGGCUAAAUACUCCUGACUGACCGAUAGUGAACCAG
UACCGUGAGGGAAAGGCGAAAAGAACCCCGGCGAGGGGAGUGAAAAAGAACCUGAAACCGUG
UACGUACAAGCAGUGGGAGCACGCUUAGGCGUGUGACUGCGUACCUUUUGUAUAAUGGGUCA
GCGACUUAUAUUCUGUAGCAAGGUUAACCGAAUAGGGGAGCCGAAGGGAAACCGAGUCUUAA
CUGGGCGUUAAGUUGCAGGGUAUAGACCCGAAACCCGGUGAUCUAGCCAUGGGCAGGUUGAA
GGUUGGGUAACACUAACUGGAGGACCGAACCGACUAAUGUUGAAAAAUUAGCGGAUGACUUG
UGGCUGGGGGUGAAAGGCCAAUCAAACCGGGAGAUAGCUGGUUCUCCCCGAAAGCUAUUUAG
GUAGCGCCUCGUGAAUUCAUCUCCGGGGGUAGAGCACUGUUUCGGCAAGGGGGUCAUCCCGA
CUUACCAACCCGAUGCAAACUGCGAAUACCGGAGAAUGUUAUCACGGGAGACACACGGCGGG
UGCUAACGUCCGUCGUGAAGAGGGAAACAACCCAGACCGCCAGCUAAGGUCCCAAAGUCAUG
GUUAAGUGGGAAACGAUGUGGGAAGGCCCAGACAGCCAGGAUGUUGGCUUAGAAGCAGCCAU
CAUUUAAAGAAAGCGUAAUAGCUCACUGGUCGAGUCGGCCUGCGCGGAAGAUGUAACGGGGC
UAAACCAUGCACCGAAGCUGCGGCAGCGACGCUUAUGCGUUGUUGGGUAGGGGAGCGUUCUG
UAAGCCUGCGAAGGUGUGCUGUGAGGCAUGCUGGAGGUAUCAGAAGUGCGAAUGCUGACAUA
AGUAACGAUAAAGCGGGUGAAAAGCCCGCUCGCCGGAAGACCAAGGGUUCCUGUCCAACGUU
AAUCGGGGCAGGGUGAGUCGACCCCUAAGGCGAGGCCGAAAGGCGUAGUCGAUGGGAAACAG
GUUAAUAUUCCUGUACUUGGUGUUACUGCGAAGGGGGGACGGAGAAGGCUAUGUUGGCCGG
GCGACGGUUGUCCCGGUUUAAGCGUGUAGGCUGGUUUUCCAGGCAAAUCCGGAAAAUCAAGG
CUGAGGCGUGAUGACGAGGCACUACGGUGCUGAAGCAACAAAUGCCCUGCUUCCAGGAAAAG
CCUCUAAGCAUCAGGUAACAUCAAAUCGUACCCCAAACCGACACAGGUGGUCAGGUAGAGAA
UACCAAGGCGCUUGAGAGAACUCGGGUGAAGGAACUAGGCAAAAUGGUGCCGUAACUUCGGG
AGAAGGCACGCUGAUAUGUAGGUGAAGCGACUUGCUCGUGGAGCUGAAAUCAGUCGAAGAUA
CCAGCUGGCUGCAACUGUUUAUUAAAAACACAGCACUGUGCAAACACGAAAGUGGACGUAUA
CGGUGUGACGCCUGCCCGGUGCCGGAAGGUUAAUUGAUGGGGUUAGCGCAAGCGAAGCUCUU
GAUCGAAGCCCCGGUAAACGGCGGCCGUAACUAUAACGGUCCUAAGGUAGCGAAAUUCCUUG
UCGGGUAAGUUCCGACCUGCACGAAUGGCGUAAUGAUGGCCAGGCUGUCUCCACCCGAGACU
CA

23s rRNA rrlB DNA template	CTTAAGTAATACGACTCACTATAGCCCTGTTTTTGCAGTCAGAGGCGATGAAGGACGTGCTAA TCTGCGATAAGCGTCGGTAAGGTGATATGAACCGTTATAACCGGCGATTTCCGAATGGGGAA ACCCAGTGTGTTTCGACACACTATCATTAACTGAATCCATAGGTTAATGAGGCGAACCGGGG GAACTGAAACATCTAAGTACCCCGAGGAAAAGAAATCAACCGAGATTCCCCCAGTAGCGGCG AGCGAACGGGGAGCAGCCCAGAGCCTGAATCAGTGTGTGTGTTAGTGGAAGCGTCTGGAAAG GCGCGCGATACAGGGTGACAGCCCCGTACACAAAAATGCACATGCTGTGAGCTCGATGAGTA GGGCGGGACACGTGGTATCCTGTCTGAATATGGGGGGACCATCCTCCAAGGCTAAATACTCC TGACTGACCGATAGTGAACCAGTACCGTGAGGGAAAGGCGAAAAGAACCCCGGCGAGGGGAG TGAAAAAGAACCTGAAACCGTGTACGTACAAGCAGTGGGAGCACGCTTAGGCGTGTGACTGC GTACCTTTTGTATAATGGGTCAGCGACTTATATTCTGTAGCAAGGTTAACCGAATAGGGGAG CCGAAGGGAAACCGAGTCTTAACTGGGCGTTAAGTTGCAGGGTATAGACCCGAAACCCGGTG ATCTAGCCATGGGCAGGTTGAAGGTTGGGTAACACTAACTGGAGGACCGAACCGACTAATGT TGAAAAATTAGCGGATGACTTGTGGCTGGGGGTGAAAGGCCAATCAAACCGGGAGATAGCTG GTTCTCCCCGAAAGCTATTTAGGTAGCGCCTCGTGAATTCATCTCCGGGGGGTAGAGCACTGTT TCGGCAAGGGGGTCATCCCGACTTACCAACCCGATGCAAACTGCGAATACCGGAGAATGTTA TCACGGGAGACACACGGCGGGTGCTAACGTCCGTCGTGAAGAGGGAAACAACCCAGACCGCC AGCTAAGGTCCCAAAGTCATGGTTAAGTGGGAAACGATGTGGGAAGGCCCAGACAGCCAGGA TGTTGGCTTAGAAGCAGCCATCATTTAAAGAAAGCGTAATAGCTCACTGGTCGAGTCGGCCT GCGCGGAAGATGTAACGGGGCTAAACCATGCACCGAAGCTGCGGCAGCGACGCTTATGCGTT GTTGGGTAGGGGAGCGTTCTGTAAGCCTGCGAAGGTGTGCTGTGAGGCATGCTGGAGGTATC AGAAGTGCGAATGCTGACATAAGTAACGATAAAGCGGGTGAAAAGCCCGCTCGCCGGAAGAC CAAGGGTTCCTGTCCAACGTTAATCGGGGCAGGGTGAGTCGACCCCTAAGGCGAGGCCGAAA GGCGTAGTCGATGGGAAACAGGTTAATATTCCTGTACTTGGTGTTACTGCGAAGGGGGGACG GAGAAGGCTATGTTGGCCGGGCGACGGTTGTCCCGGTTTAAGCGTGTAGGCTGGTTTTCCAG GCAAATCCGGAAAATCAAGGCTGAGGCGTGATGACGAGGCACTACGGTGCTGAAGCAACAAA TGCCCTGCTTCCAGGAAAAGCCTCTAAGCATCAGGTAACATCAAATCGTACCCCAAACCGACA CAGGTGGTCAGGTAGAGAATACCAAGGCGCTTGAGAGAACTCGGGTGAAGGAACTAGGCAAA ATGGTGCCGTAACTTCGGGAGAAGGCACGCTGATATGTAGGTGAAGCGACTTGCTCGTGGAG CTGAAATCAGTCGAAGATACCAGCTGGCTGCAACTGTTTATTAAAAACACAGCACTGTGCAAA CACGAAAGTGGACGTATACGGTGTGACGCCTGCCCGGTGCCGGAAGGTTAATTGATGGGGTT AGCGCAAGCGAAGCTCTTGATCGAAGCCCCGGTAAACGGCGGCCGTAACTATAACGGTCCTA AGGTAGCGAAATTCCTTGTCGGGTAAGTTCCGACCTGCACGAATGGCGTAATGATGGCCAGG CTGTCTCCACCCGAGACTCA
M13 transcript RNA scaffold	GGUCUCACUGGUGAAAAGAAAAACCACCCUGGCGCCCAAUACGCAAACCGCCUCUCCCCGCG CGUUGGCCGAUUCAUUAAUGCAGCUGGCACGACAGGUUUCCCGACUGGAAAGCGGGCAGUGA GCGCAACGCAAUUAAUGUGAGUUAGCUCACUCAUUAGGCACCCCAGGCUUUACACUUUAUGC UUCCGGCUCGUAUGUUGUGUGGAAUUGUGAGCGGAUAACAAUUUCACACAGGAAACAGCUAU GACCAUGAUUACGAAUUCGAGCUCGGUACCCGGGGAUCCUCUAGAGUCGACCUGCAGGCAUG CAAGCUUGGCACUGGCCGUCGUUUUACAACGUCGUGACUGGGAAAACCCUGGCGUUACCCAA CUUAAUCGCCUUGCAGCACAUCCCCCUUUCGCCAGCUGGCGUAAUAGCGAAGAGGCCCGCAC CGAUCGCCCUUCCCAACAGUUGCGCAGCCUGAAUGGCGAAUGGCGCUUUGCCUGGUUUCCGG CACCAGAAGCGGUGCCGGAAAGCUGGCUGGAGUGCGAUCUUCCUGAGGCCGAUACUGUCGUC GUCCCCUCAAACUGGCAGAUGCACGGUUACGAUGCGCCCAUCUACACCAACGUGACCUAUCC CAUUACGGUCAAUCCGCCGUUUGUUCCCACGGAGAAUCCGACGGGUUGUUACUCGCUCACAU UUAAUGUUGAUGAAAGCUGGCUACAGGAAGGCCAGACGCGAAUUAUUUUUGAUGGCGUUCCU AUUGGUUAAAAAAUGAGCUGAUUUAACAAAAAUUUAAUGCGAAUUUUAACAAAAUAUUAACG UUUACAAUUUAAAUAUUUGCUUAUACAAUCUUCCUGUUUUUGGGGCUUUUCUGAUUAUCAAC CGGGGUACAUAUGAUUGACAUGCUAGUUUUACGAUUACCGUUCAUCGAUUCUCUUGUUUGCU CCAGACUCUCAGGCAAUGACCUGAUAGCCUUUGUAGAUCUCUCAAAAAUAGCUACCCUCUCC GGCAUUAAUUUAUCAGCUAGAACGGUUGAAUAUCAUAUUGAUGGUGAUUUGACUGUCUCCGG CC

M13 transcript RNA sequence	GGGUCUCACUGGUGAAAAGAAAAACCACCCUGGCGCCCAAUACGCAAACCGCCUCUCCCCGC GCGUUGGCCGAUUCAUUAAUGCAGCUGGCACGACAGGUUUCCCGACUGGAAAGCGGGCAGUG AGCGCAACGCAAUUAAUGUGAGUUAGCUCACUCAUUAGGCACCCCAGGCUUUACACUUUAUG CUUCCGGCUCGUAUGUUGUGUGGAAUUGUGAGCGGAUAACAAUUUCACACAGGAAACAGCUA UGACCAUGAUUACGAAUUCGAGCUCGGUACCCGGGGAUCCUCUAGAGUCGACCUGCAGGCAU GCAAGCUUGGCACUGGCCGUCGUUUUACAACGUCGUGACUGGGAAAACCCUGGCGUUACCCA ACUUAAUCGCCUUGCAGCACAUCCCCCUUUCGCCAGCUGGCGUAAUAGCGAAGAGGCCCGCA CCGAUCGCCCUUCCCAACAGUUGCGCAGCCUGAAUGGCGAAUGGCGCUUUGCCUGGUUUCCG GCACCAGAAGCGGUGCCGGAAAGCUGGCUGGAGUGCGAUCUUCCUGAGGCCGAUACUGUCGU CGUCCCCUCAAACUGGCAGAUGCACGGUUACGAUGCGCCCAUCUACACCAACGUGACCUAUC CCAUUACGGUCAAUCCGCCGUUUGUUCCCACGGAGAAUCCGACGGGUUGUUACUCGCUCACA UUUAAUGUUGAUGAAAGCUGGCUACAGGAAGGCCAGACGCGAAUUAUUUUUGAUGGCGUUCC UAUUGGUUAAAAAAUGAGCUGAUUUAACAAAAAUUUAAUGCGAAUUUUAACAAAAUAUUAAC GUUUACAAUUUAAAUAUUUGCUUAUACAAUCUUCCUGUUUUUGGGGCUUUUCUGAUUAUCAA CCGGGGUACAUAUGAUUGACAUGCUAGUUUUACGAUUACCGUUCAUCGAUUCUCUUGUUUGC UCCAGACUCUCAGGCAAUGACCUGAUAGCCUUUGUAGAUCUCUCAAAAAUAGCUACCCUCUCC GGCAUUAAUUUAUCAGCUAGAACGGUUGAAUAUCAUAUUGAUGGUGAUUUGACUGUCUCCGG CC
M13 transcript DNA template	GAATTCTAATACGACTCACTATAGGGTCTCACTGGTGAAAAGAAAAACCACCCTGGCGCCCAA TACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTC CCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCA CCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAA TTTCACACAGGAAACAGCTATGACCATGATTACGAATTCGAGCTCGGTACCCGGGGATCCTCT AGAGTCGACCTGCAGGCATGCAAGCTTGGCACTGGCCGTCGTTTTACAACGTCGTGACTGGG AAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTA ATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGG CGCTTTGCCTGGTTTCCGGCACCAGAAGCGGTGCCGGAAAGCTGGCTGGAGTGCGATCTTCC TGAGGCCGATACTGTCGTCGTCCCCTCAAACTGGCAGATGCACGGTTACGATGCGCCCATCTA CACCAACGTGACCTATCCCATTACGGTCAATCCGCCGTTTGTTCCCACGGAGAATCCGACGGG TTGTTACTCGCTCACATTTAATGTTGATGAAAGCTGGCTACAGGAAGGCCAGACGCGAATTAT TTTTGATGGCGTTCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAATGCGAATTTT AACAAAATATTAACGTTTACAATTTAAATATTTGCTTATACAATCTTCCTGTTTTTGGGGCTT TTCTGATTATCAACCGGGGTACATATGATTGACATGCTAGTTTTACGATTACCGTTCATCGAT TCTCTTGTTTGCTCCAGACTCTCAGGCAATGACCTGATAGCCTTTGTAGATCTCTCAAAAATA GCTACCCTCTCCGGCATTAATTTATCAGCTAGAACGGTTGAATATCATATTGATGGTGATTTG ACTGTCTCCGGCC
De Bruijn rsc1218v1 gblock	GGGCCTAGTGAGTGCATTAGAGACTCAAGCCATGTATCCATGACCAGAAGAGGGGACCCTAG GCCAAGGACATGAGGGGCTCTATGCACTAATGTCTAGTCAGGCCTTCACCCTAGTCATCTCTA TGTGGAATACAATTGGGATTCATGAGAGAATCATGCAAGCCAAAGTCAGCAGGTAGCCTTAT TTGATGGAGTTAACCAACAGCCAAATTTATTGCCCAGAGCCCATTAACCATGCTAGAGGAGG GTTTCTCCTGATGACTGGCTCCCTGACATCTGTGTCTGGGCCTCAAATTGGTTTGAAGCTGAT GGGCTATGCATAATCCCATCACTAAGACCTTGGTCTTGCTTGCTAGAACTTTTCATTACACCA CTAGCTGAAGTTTGTTATGACTGCCCCTAATGGCAGGCTCCTCAGGGACAGAAGGAACTGGC CAACCCCATATCTAATGTTGTTGGGCTTGCTGTAGGGTACCCATCTATGAGGAGGCTTGTAAC TGTGATGCTCCCTTAGCTACTTTTGATGTATATGCACAGTGTTGGCAGTGGACCTTGCTCCAT GCAATTCATATGGCTATCTTGCAGTCCATGGCCCACCCCAGACTTCATTGCTGCCCTGTGACA CCATTCATTAGAACCTGGTATGGAAGTGTACCACAGAGCTGCAAGGGGTTGTAGACCACTTA TGGTGAAATTTGTAGAATCCAGGGTGGTGAGCTGTAAATGAAACTTTGGATAACTGGCTTTC TAGAGCCTTCCTCCCACCCTCCCCTGTTTGCCATGCCCCAGGGGCAAGCACTGCCCTTTGCAC TCTCACTACAGGTAAAGAGTGTCCCCAGGAGTAAATTTCTAGCCCCATAGAAAAGGAAGGTCT AGAGGGAAATTGGCAATGGGCACCTGTCCCATTATAAGCATCTATTTGAAGATGCCTGGCAC AATAGATCAGTAGGGGAGTTGCTATACCCACTGGTAAGGTTAGATGGTGTTAGTAGGCAGGT CAGCACAGACCTGCACATACAGCATGGGTGGTTACCCAACCATACACAAGACCCATGAAAGTG TGGAAGCTCTTGTACACCCTCTGGCACCAGCATAGTGTTGCATTGTGCACTGAAAGAGCTACC CCAGGCTGTAGCCTAATTTCTTATGGGGATAGGGGCCTTGCAATTTCAAAGCCTTGGACACCT GACCCCACCTCTGAGGGTCTCAGGCCCTC

De Bruijn rsc1218v1 rT55 RNA scaffold	GGGCCUAGUGAGUGCAUUAGAGACUCAAGCCAUGUAUCCAUGACCAGAAGAGGGGACCCUAG GCCAAGGACAUGAGGGGCUCUAUGCACUAAUGUCUAGUCAGGCCUUCACCCUAGUCAUCUCU AUGUGGAAUACAAUUGGGAUUCAUGAGAGAAUCAUGCAAGCCAAAGUCAGCAGGUAGCCUUA UUUGAUGGAGUUAACCAACAGCCAAAUUUAUUGCCCAGAGCCCAUUAACCAUGCUAGAGGAG GGUUUCUCCUGAUGACUGGCUCCCCUGACAUCUGUGUCUGGGCCUCAAAUUGGUUUGAAGCUG AUGGGCUAUGCAUAAUCCCAUCACUAAGACCUUGGUCUUGCUUGCUAGAACUUUUCAUUACA CCACUAGCUGAAGUUUGUUAUGACUGCCCCUAAUGGCAGGCUCCUCAGGGACAGAAGGAACU GGCCAACCCCAUAUCUAAUGUUGUUGGGCUUGCUGUAGGGUACCCAUCUAUGAGGAGGCUUG UAACUGUGAUGCUCCCUUAGCUACUUUUGAUGUAUAUGCACAGUGUUGGCAGUGGACCUUGC UCCAUGCAAUUCAUAUGGCUAUCUUGCAGUCCAUGGCCCACCCCAGACUUCAUUGCUGCCCU GUGACACCAUUCAUUAGAACCUGGUAUGGAAGUGUACCAC
De Bruijn rsc1218v1 rT55 RNA sequence	GGGCCUAGUGAGUGCAUUAGAGACUCAAGCCAUGUAUCCAUGACCAGAAGAGGGGACCCUAG GCCAAGGACAUGAGGGGCUCUAUGCACUAAUGUCUAGUCAGGCCUUCACCCUAGUCAUCUCU AUGUGGAAUACAAUUGGGAUUCAUGAGAGAAUCAUGCAAGCCAAAGUCAGCAGGUAGCCUUA UUUGAUGGAGUUAACCAACAGCCAAAUUUAUUGCCCAGAGCCCAUUAACCAUGCUAGAGGAG GGUUUCUCCUGAUGACUGGCUCCCCUGACAUCUGUGUCUGGGCCUCAAAUUGGUUUGAAGCUG AUGGGCUAUGCAUAAUCCCAUCACUAAGACCUUGGUCUUGCUUGCUAGAACUUUUCAUUACA CCACUAGCUGAAGUUUGUUAUGACUGCCCCUAAUGGCAGGCUCCUCAGGGACAGAAGGAACU GGCCAACCCCAUAUCUAAUGUUGUUGGGCUUGCUGUAGGGUACCCAUCUAUGAGGAGGCUUG UAACUGUGAUGCUCCCUUAGCUACUUUUGAUGUAUAUGCACAGUGUUGGCAGUGGACCUUGC UCCAUGCAAUUCAUAUGGCUAUCUUGCAGUCCAUGGCCCACCCCAGACUUCAUUGCUGCCCU GUGACACCAUUCAUUAGAACCUGGUAUGGAAGUGUACCAC
De Bruijn rsc1218v1 rT77 RNA scaffold	GGGCCUAGUGAGUGCAUUAGAGACUCAAGCCAUGUAUCCAUGACCAGAAGAGGGGACCCUAG GCCAAGGACAUGAGGGGCUCUAUGCACUAAUGUCUAGUCAGGCCUUCACCCUAGUCAUCUCU AUGUGGAAUACAAUUGGGAUUCAUGAGAGAAUCAUGCAAGCCAAAGUCAGCAGGUAGCCUUA UUUGAUGGAGUUAACCAACAGCCAAAUUUAUUGCCCAGAGCCCAUUAACCAUGCUAGAGGAG GGUUUCUCCUGAUGACUGGCUCCCCUGACAUCUGUGUCUGGGCCUCAAAUUGGUUUGAAGCUG AUGGGCUAUGCAUAAUCCCAUCACUAAGACCUUGGUCUUGCUUGCUAGAACUUUUCAUUACA CCACUAGCUGAAGUUUGUUAUGACUGCCCCUAAUGGCAGGCUCCUCAGGGACAGAAGGAACU GGCCAACCCCAUAUCUAAUGUUGUUGGGCUUGCUGUAGGGUACCCAUCUAUGAGGAGGCUUG UAACUGUGAUGCUCCCUUAGCUACUUUUGAUGUAUAUGCACAGUGUUGGCAGUGGACCUUGC UCCAUGCAAUUCAUAUGGCUAUCUUGCAGUCCAUGGCCCACCCCAGACUUCAUUGCUGCCCU GUGACACCAUUCAUUAGAACCUGGUAUGGAAGUGUACCACAGAGCUGCAAGGGGUUGUAGAC CACUUAUGGUGAAAUUUGUAGAAUCCAGGGUGGUGAGCUGUAAAUGAAACUUUGGAUAACUG GCUUUCUAGAGCCUUCCUCCCACCCUCCCCUGUUUGCCAUGCCCCAGGGGCAAGCACUGCCCU UUGCACUCUCACUACAGGUAAAGAGUGUCCCCAGGAGUAAAUUUCUAGCCCCAUAGAAAAGG AAGGUCUAGAGGGAAAUUGGCAAUGGGCACCUGUCCCAUUAUAAGCAUCUAUUUG
De Bruijn rsc1218v1 rT77 RNA sequence	GGGCCUAGUGAGUGCAUUAGAGACUCAAGCCAUGUAUCCAUGACCAGAAGAGGGGACCCUAG GCCAAGGACAUGAGGGGCUCUAUGCACUAAUGUCUAGUCAGGCCUUCACCCUAGUCAUCUCU AUGUGGAAUACAAUUGGGAUUCAUGAGAGAAUCAUGCAAGCCAAAGUCAGCAGGUAGCCUUA UUUGAUGGAGUUAACCAACAGCCAAAUUUAUUGCCCAGAGCCCAUUAACCAUGCUAGAGGAG GGUUUCUCCUGAUGACUGGCUCCCCUGACAUCUGUGUCUGGGCCUCAAAUUGGUUUGAAGCUG AUGGGCUAUGCAUAAUCCCAUCACUAAGACCUUGGUCUUGCUUGCUAGAACUUUUCAUUACA CCACUAGCUGAAGUUUGUUAUGACUGCCCCUAAUGGCAGGCUCCUCAGGGACAGAAGGAACU GGCCAACCCCAUAUCUAAUGUUGUUGGGCUUGCUGUAGGGUACCCAUCUAUGAGGAGGCUUG UAACUGUGAUGCUCCCUUAGCUACUUUUGAUGUAUAUGCACAGUGUUGGCAGUGGACCUUGC UCCAUGCAAUUCAUAUGGCUAUCUUGCAGUCCAUGGCCCACCCCAGACUUCAUUGCUGCCCU GUGACACCAUUCAUUAGAACCUGGUAUGGAAGUGUACCACAGAGCUGCAAGGGGUUGUAGAC CACUUAUGGUGAAAUUUGUAGAAUCCAGGGUGGUGAGCUGUAAAUGAAACUUUGGAUAACUG GCUUUCUAGAGCCUUCCUCCCACCCUCCCCUGUUUGCCAUGCCCCAGGGGCAAGCACUGCCCU UUGCACUCUCACUACAGGUAAAGAGUGUCCCCAGGAGUAAAUUUCUAGCCCCAUAGAAAAGG AAGGUCUAGAGGGAAAUUGGCAAUGGGCACCUGUCCCAUUAUAAGCAUCUAUUUG

HIV RRE DNA template	GGCTTATCGAAATTAATACGACTCACTATAGGAGCTTTGTTCCTTGGGTTCTTGGGAGCAGCA GGAAGCACTATGGGCGCAGCGTCAATGACGCTGACGGTACAGGCCAGACAATTATTGTCTGA TATAGTGCAGCAGCAGAACAATTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAAC TCACAGTCTGGGGCATCAAACAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGATACCTAAAG GATCAACAGCTCC
HIV RRE RNA sequence	GGAGCUUUGUUCCUUGGGUUCUUGGGAGCAGCAGGAAGCACUAUGGGCGCAGCGUCAAUGAC GCUGACGGUACAGGCCAGACAAUUAUUGUCUGAUAUAGUGCAGCAGCAGAACAAUUUGCUGA GGGCUAUUGAGGCGCAACAGCAUCUGUUGCAACUCACAGUCUGGGGCAUCAAACAGCUCCAG GCAAGAAUCCUGGCUGUGGAAAGAUACCUAAAGGAUCAACAGCUCC

Table D.19: Primer sequences to amplify DNA templates for the RNA scaffold sequences in the previous table.

Name	**Sequence**
EGFP for	GAATTCTAATACGACTCACTATAGGTAGCTAAGG
EGFP rev	CGTATGTTGTGTGGAATTGTGAG
23SdomIIV for	CTTAAGTAATACGACTCACTATAGCCCTGGCAGTCAGAGG
23SdomIIV rev	TGAGTCTCGGGTGGAGACAG
M13o44 for	TAATACGACTCACTATAGGGTCTCGCTGGTGAAAAGAAA
M13o44 rev	GGCCGGAGACAGTCAAATC
rsc1218v1 for	GGCTTATCGAAATTAATACGACTCACTATAGGGCCTAGTGAGTGCATTAGAG
rsc1218v1 rT55 rev	GTGGTACACTTCCATACCAGGTTC
rsc1218v1 rT77 rev	CAAATAGATGCTTATAATGGGACAGGTGC
HIV RRE for	GGAGCTTTGTTCCTTGGGTTCTTGG
HIV RRE T7 for	GGCTTATCGAAATTAATACGACTCACTA
HIV RRE rev	GGAGCTGTTGATCCTTTAGGTATCTTTC

D.3.1 Staple sequences for RNA-scaffolded objects with A-form routing

Table D.20: Staple sequences used to fold the **A-form** EGFP mRNA-scaffolded tetrahedron of 66-bp edge length (staple crossover asymmetry design).

Staple ID	**Length (bp)**	**Staple sequences**
1	44	GTGGTCGGGGTACAGCTCCTCGCCCTTTTGTGTTGCTGCTTCAT
2	44	CACTGCACGCCAGGATGGGCACCACCCCGGTGAAGCGGCTGAAG
3	81	TGGTAGTGGTCGGTTTTTCGAGCTGCACATGGCGGACTTGTTTTTAAGAAGTCGGG AATTGTGAGCGTTTTTGATAACAAT
4	44	AACAGCTATGAGTCGCCGATGGGGGTGTTCTGCTTCACACAGGA
5	44	CCAAGCTTGCATGTCGGGCAGCAGCACGGGGCCCCATGATTACG
6	33	TCCAGCTTACGTTGTGGCTGTTGTAGTCTTTGC
7	33	TCAGGGCGATCGCGCTTCTCGTTGGGGTTGTAC
8	22	GTGCCCCAGGACCATGATATAG
9	81	ATGCCGTTCTTCTTTTTTGCTTGTCGGTGTTGCCGTCCTCTTTTTCTTGAAGTCGG CGCGGGTCTTGTTTTTTAGTTGCCG
10	44	TCGATGTTGTGGCTCCTGGACGTAGCCTTCGGGCGCTGCCGTCC
11	44	AGTTCACCTTGTCGTCCTTGAAGAAGATGGTGCGCGGATCTTGA
12	81	CGAACTCCAGCAGTTTTTGACCATGTGGACTGGGTGCTCATTTTTGGTAGTGGTTG CCTGCAGTTACTTTTTTTGTACAGC
13	33	CGCTGAACCATCGCCCTCGCCCTCGCTCCCGGC
14	33	GGCGGTCATCGTCCATGCCGAGAGTGACGGACA
15	22	TTGTGGCCGTTGCCGTAGGTGG
16	81	AGATGAACTTCAGTTTTTGGTCAGCTTTACGTCGCCGTCCTTTTTAGCTCGACCGT AGGTCAGGGTGTTTTTGTCACGAGG
17	33	CGCCCTCGGCTCGATGCGGTTCACCAGCTTGCC
18	33	GGTGGTGCGTGGGCCAGGGCACGGGCAGGGTGT
19	22	AACTTCACCTCGATGCCCTTCA

Table D.21: Staple sequences used to fold the **A-form** de Bruijn RNA-scaffolded tetrahedron of 55-bp edge length (staple crossover asymmetry design).

Staple ID	Length (bp)	Staple sequences
1	44	CTACCTGCTGATGCACTCACTAGGCCCGTGGTAATCAAATAAGG
2	22	CTTTGGCTTGCTGAGTCTCTAA
3	81	AGTTCCTTCTGTCTTTTTCCTGAGGAGAATTTGGCTGTTGTTTTTGTTAACTCCCA CTTCCATACCATTTTTGGTTCTAAT
4	44	CAGGGCAGCAAAACATTAGATATGGGGTTGGCCGAATGGTGTCA
5	22	TGAAGTCTGGGGCAAGCCCAAC
6	33	GATGGGATTAGCAAGCAAGACCAAGGTACAGTT
7	33	ACAAGCCTAAGTAGCTAAGGGAGCATCCTTAGT
8	81	CTAGTGGTGTAATTTTTTGAAAAGTTCTATGCATAGCCCATTTTTTCAGCTTCAAG CCAGTCATCAGTTTTTGAGAAACCC
9	44	GGGCAGTCATAGGTTAATGGGCTCTGGGCAATACCTGCCATTAG
10	22	ACAAACTTCAGTCCTCTAGCAT
11	81	CAACACTGTGCATTTTTTATACATCAACCTCATAGATGGGTTTTTTACCCTACAGTG GGCCATGGACTTTTTTGCAAGATA
12	33	GTCCCCTCGAGCCCCTCATGTCCTTGGGCAAGG
13	33	TCCACTGCGCCATATGAATTGCATGGACCTAGG
14	81	CCTGACTAGACATTTTTTTAGTGCATATTCTGGTCATGGATTTTTTACATGGCTAT GATTCTCTCATTTTTTGAATCCCAA
15	33	TGTCAGGGAACCAATTTGAGGCCCAGAGACTAG
16	33	GGTGAAGGTTGTATTCCACATAGAGATCACAGA

Table D.22: Staple sequences used to fold the **A-form** de Bruijn RNA-scaffolded tetrahedron of 77-bp edge length (staple crossover asymmetry design).

Staple ID	Length (bp)	Staple sequences
1	44	AGCCAGTCATCTGCACTCACTAGGCCCCAAATAAGATGTCAGGG

2	44	TCCTCTAGCATGATACATGGCTTGAGTCTCTAAAGGAGAAACCC
3	22	GGTTAATGGGCTTCTGGTCATG
4	81	TGAAGTCTGGGGTTTTTTGGGCCATGGAACCAATTTGAGGTTTTTCCCAGACACGA TGCTTATAATGTTTTTGGACAGGTG
5	44	TTTCCCTCTAGGAATGGTGTCACAGGGCAGCAACCCATTGCCAA
6	44	CTATGGGGCTACACTTCCATACCAGGTTCTAATACCTTCCTTTT
7	22	GAAATTTACTCGCTCTGTGGTA
8	33	ACCCTACAGTTACAAGCCTCCTCATAGCCCTGG
9	33	ATTCTACATTTCATTTACAGCTCACCAATGGGT
10	22	AATTTCACCATGTTATCCAAAG
11	22	GCAAGCCCAACGGAGCATCACA
12	81	ATATACATCAAAATTTTTGTAGCTAAGAACATTAGATATGTTTTTGGGTTGGCCCT AGTGGTGTAATTTTTTGAAAAGTTC
13	44	GCCATATGAATTATGCATAGCCCATCAGCTTCAACTGCAAGATA
14	44	AGGTCCACTGCACCAAGGTCTTAGTGATGGGATTGCATGGAGCA
15	22	CAACACTGTGCTAGCAAGCAAG
16	81	GGAGGAAGGCTCTTTTTTAGAAAGCCAAAGTGGTCTACAATTTTTCCCCTTGCACT GGGGACACTCTTTTTTTTACCTGTA
17	33	AGTGCATATAGGGTGAAGGCCTGACTAGGGGCA
18	33	TGGCAAACAAGGGCAGTGCTTGCCCCTGACATT
19	22	AGGGGAGGGTGGTGAGAGTGCA
20	22	GAGCCCCTCATATAGAGATGAC
21	81	ATGAATCCCAATTTTTTTGTATTCCACGTCCTTGGCCTAGTTTTTGGTCCCCTCTCT GGGCAATAAATTTTTTTTGGCTGT
22	33	CAGTCATATCCCTGAGGAGCCTGCCATTGACTT
23	33	TGGCTTGCATCAAATAAGGCTACCTGCTAGGGG
24	22	ATGATTCTCTCTGGTTAACTCC
25	22	ACAAACTTCAGAGTTCCTTCTG

Table D.23: Staple sequences used to fold the **A-form** M13 transcript-scaffolded octahedron of 44-bp edge length (staple crossover asymmetry design).

Staple ID	**Length (bp)**	**Staple sequences**
1	54	TGATAAATTAATGTTTTTCCGGAGAGGCAGGTCATTGCCTTTTTTGAGAGTCTG
2	44	ACCGTTCTAGCAATATTTAAATTGTAAACGTTATATGATATTCA
3	54	GAGACAGTCAAATTTTTTCACCATCAAATATTTTGTTAAATTTTTATTCGCATT
4	44	GCCTGGGGTGCTTTTCACCAGTGAGACAGGCCGTAAAGTGTAAA
5	54	TTGGGCGCCAGGGTTTTTTGGTTTTTCCTAATGAGTGAGCTTTTTTAACTCACA
6	44	AAGGCTATGTAGCTATTTTTGAGAGACGTAGCGGTTTGTCTACA
7	54	CCAGCTGGCGAAATTTTTGGGGGATGTGTTTTCCCAGTCATTTTTCGACGTTGT
8	44	TCGCTATTACGTTGTTATCCGCTCACAATTCCATGCGGGCCTCT
9	54	AACTGTTGGGAAGTTTTTGGCGATCGGCACAACATACGAGTTTTTCCGGAAGCA
10	44	AAATCAGCTCACCATTCGCCATTCAGGCTGCGCAAATTTTTGTT
11	54	GTGCCGGAAACCATTTTTGGCAAAGCGTTTTTTAACCAATTTTTTAGGAACGCC
12	44	ACGCCAGGGCTGCAAGGCGATTAAGTTCTGGCACCGCTTGGGTA
13	54	TGAGCGAGTAACATTTTTACCCGTCGGAACAAACGGCGGATTTTTTTGACCGTA
14	44	CCTCAGGAAGACAGCTTTCATCAACATTAAATGGACAGTATCGG
15	54	TCGCGTCTGGCCTTTTTTTCCTGTAGCTCGCACTCCAGCCTTTTTAGCTTTCCG
16	44	AGCCCCAATGTACCCCGGTTGATAATTAATATCAAAAACAGAAA
17	54	TAAAACTAGCATGTTTTTTCAATCATAAAACAGGAAGATTTTTTTGTATAAGCA
18	44	GGTAATCGGAGCAAACAAGAGAATCGTGGGATTCTCCGATGAAC
19	54	ACGTTGGTGTAGATTTTTTGGGCGCATTCTGCCAGTTTGATTTTTGGGGACGAC
20	44	ATTAATGACGGGAAACCTGTCGTGCCGGTCATGGGATAAGCTGC
21	54	CGCTCACTGCCCGTTTTTCTTTCCAGTATCGGCCAACGCGTTTTTCGGGGAGAG
22	44	ATCATGGTCCGGGTACCGAGCTCGAAGTTGTTAATTGCTTCGTA
23	54	GCAGGTCGACTCTTTTTTAGAGGATCCCATAGCTGTTTCCTTTTTTGTGTGAAA
24	44	GCATGCCTAAAACGACGGCCAGTGCCTGCACGTAACCGAAGCTT

Table D.24: Staple sequences used to fold the **A-form** 23s rRNA-scaffolded octahedron of 66-bp edge length (staple crossover asymmetry design).

Staple ID	**Length (bp)**	**Staple sequences**
1	54	GCACCGTAGTGCCTTTTTTCGTCATCACCGGGACAACCGTTTTTTCGCCCGGCC
2	44	GCTTTTCCTGGTTACTTATGTCAGCATTCGCACGATGCTTAGAG
3	44	TTGTTGCTTCAGGGCTTTTCACCCGCTTTATCGAAGCAGGGCAT
4	54	GGGGTACGATTTGTTTTTATGTTACCTTTCTGATACCTCCTTTTTAGCATGCCT
5	44	TTCCAGACGCTCTCTGACTGCCAGGGCCGGTTTATCGCGCGCCT
6	44	ACACACTGATTAGATTAGCACGTCCTTCATCGCTCCACTAACAC
7	54	TCACCTTACCGACTTTTTGCTTATCGCCAGGCTCTGGGCTTTTTTGCTCCCCGT
8	33	ACCAGCCTTGATTTTCCGGATTTGCCTTATAAC
9	33	GGTTCATACCCATTCGGAAATCGCCGGTGGAAA
10	22	ACACGCTTAAACGCCTCAGCCT
11	54	TCGGTTTCCCTTCTTTTTGGCTCCCCTCGCAGTCACACGCTTTTTCTAAGCGTG
12	44	TCTATACCCTGAGCTCACAGCATGTGCATTTTTCGGGTTTCGGG
13	44	CCAGTTAAGACACGTGTCCCGCCCTACTCATCGCAACTTAACGC
14	54	AACCTGCCCATGGTTTTTCTAGATCACGTGTACGGGGCTGTTTTTTCACCCTGT
15	44	TTCGCAGGCTTCAGTTAGTGTTACCCAACCTTCCACAGCACACC
16	44	CCCTACCCAACAACATTAGTCGGTTCGGTCCTCACAGAACGCTC
17	54	ACAAGTCATCCGCTTTTTTAATTTTTCAACGCATAAGCGTTTTTTCGCTGCCGC
18	33	ACCCATTACTTGCTACAGAATATAAGCTTTCAC
19	33	CCCCAGCCATCTCCCGGTTTGATTGGCTCGCTG
20	22	TACAAAAGGTAATTCGGTTAAC
21	54	CATCCTGGCTGTCTTTTTTGGGCCTTCCCTTAGCTGGCGGTTTTTTCTGGGTTG
22	44	CCCCGGAGATGATGATGGCTGCTTCTAAGCCAACAGTGCTCTAC
23	44	CGCTACCTAAATGAGCTATTACGCTTTCTTTAAAATTCACGAGG
24	54	CCGCGCAGGCCGATTTTTCTCGACCAGTAGCTTTCGGGGATTTTTGAACCAGCT
25	33	CCGATTAACCTTAGGGGTCGACTCACGCCCCGT

26	33	TACATCTTAGCTTCGGTGCATGGTTTACCTGCC
27	22	CGTTGGACAGGTTTCGGCCTCG
28	54	CCTGTTTCCCATCTTTTTGACTACGCCAACCCTTGGTCTTTTTTTCCGGCGAGC
29	33	CCAAGTACCTCCGTCCCCCCTTCGCAACCATGA
30	33	CTTTGGGACCACATCGTTTCCCACTTAGTAACA
31	22	AGGAATATTAAAACATAGCCTT
32	54	TGTGTCTCCCGTGTTTTTATAACATTCCGGGATGACCCCCTTTTTTTGCCGAAA
33	33	GGATTCAGCCCCGGTTCGCCTCATTACGTTAGC
34	33	ACCCGCCGTTTCCCTCTTCACGACGGAACCTAT
35	22	TTAATGATAGTTGTTTCAGTTC
36	54	TCTTTTCCTCGGGTTTTTGTACTTAGAGTGTCGAAACACATTTTTCTGGGTTTC
37	33	TAGCCTTGCACTATCGGTCAGTCAGGGAATCTC
38	33	GGTTGATTTCGCTCGCCGCTACTGGGGAGTATT
39	22	GAGGATGGTCCCGGTACTGGTT
40	54	GGTTCTTTTCGCCTTTTTTTTCCCTCACCCCATATTCAGATTTTTCAGGATACC
41	33	TTTTCACTTGTACGTACACGGTTTCAATCGGGT
42	33	TGGTAAGTTCCGGTATTCGCAGTTTGCGGTTCT
43	22	CCCCTCGCCGGCTCCCACTGCT

Table D.25: Staple sequences used to fold the **A-form** 23s rRNA-scaffolded pentagonal bipyramid of 66-bp edge length (staple crossover asymmetry design).

Staple ID	**Length (bp)**	**Staple sequences**
1	54	TTTTCACCCGCTTTTTTTTATCGTTACAACCTTGCTACAGTTTTTAATATAAGT
2	44	GCTTTTCCTGGCTTGGTCTTCCGGCGAGCGGGCGATGCTTAGAG
3	44	TTGTTGCTTCACGATTAACGTTGGACAGGAACCAAGCAGGGCAT
4	44	TTCGCACTTCTCCTTCGGCTCCCCTATTCGGTTTTATGTCAGCA
5	44	CATGCCTCACACGCCCAGTTAAGACTCGGTTTCGATACCTCCAG

6	81	TCAACATTAGTCGTTTTTGTTCGGTCCGGGTCTATACCCTTTTTTGCAACTTAAGC ACACCTTCGCATTTTTGGCTTACAG
7	33	TCACAGCACGTGTCCCGCCCTACTCATCATCCG
8	33	CTAATTTTTTTCACCCCCAGCCACAAGTCGAGC
9	22	TGTGCATTTTTACAGGATACCA
10	54	GGTTCTTTTCGCCTTTTTTTTCCCTCAGTGTCGAAACACATTTTTCTGGGTTTC
11	33	GGCTAGATTTACCCAACCTTCAACCTCACTCCC
12	33	CTCGCCGGACGGTTTCAGGTTCTTTTTGCCCAT
13	22	CACCGGGTTTCTCCAGTTAGTG
14	44	TTCGTGCAGGTACCTGACCACCTGTGTCGGTTTTCATTACGCCA
15	44	CGACAAGGAATCTCAAGCGCCTTGGTATTCTCTCGGAACTTACC
16	54	TCGGGTGGAGACATTTTTGCCTGGCCAGGGGTACGATTTGTTTTTATGTTACCT
17	44	TATACAAAAGGCTCTGACTGCCAGGGCTGAGTCCGCTGACCCAT
18	44	ACGCCTAAGCGAGATTAGCACGTCCTTCATCGCTACGCAGTCAC
19	54	TCACCTTACCGACTTTTTGCTTATCGCTGCTCCCACTGCTTTTTTTGTACGTAC
20	33	AGCTTCGCCCGTTTACCGGGGCTTCGTTATAAC
21	33	GGTTCATACCCATTCGGAAATCGCCGGATCAAG
22	22	TTGCGCTAACCAGTTACGGCCG
23	54	AGGATGGTCCCCCTTTTTCATATTCAGGTGTACGGGGCTGTTTTTTCACCCTGT
24	44	CACTATCGGTCTATGGATTCAGTTAATGATAGTCGGTACTGGTT
25	44	TTTAGCCTTGGCCCCCGGTTCGCCTCATTAACCAGTCAGGAGTA
26	54	TGCCGCAGCTTCGTTTTTGTGCATGGTCTATCTCCCGGTTTTTTTTGATTGGCC
27	54	TAGGGGTCGACTCTTTTTACCCTGCCCGCACCGTAGTGCCTTTTTTCGTCATCA
28	33	ACGCCTTTATATTAACCTGTTTCCCAACGCATA
29	33	AGCGTCGCAACGCTCCCCTACCCAACATCGACT
30	22	CGGCCTCGCCTCAAGTACAGGA
31	33	CCTTCCCAAAGCCAACATCCTGGCTGCAACATA
32	33	GCCTTCTCGGGACAACCGTCGCCCGGCTCTGGG
33	22	CATCGTTTCCCTGGCTGCTTCT
34	54	AACCAGCCTACACTTTTTGCTTAAACCCGTCCCCCCTTCGTTTTTCAGTAACAC

35	81	CTAGTTCCTTCACTTTTTCCGAGTTCTTTCGCTACCTTAGTTTTTGACCGTTATCCA TCAATTAACCTTTTTTTCCGGCAC
36	33	AGTTACGGACATATCAGCGTGCCTTCCGGATTT
37	33	GCCTGGAACGCCTCAGCCTTGATTTTCTCCCGA
38	22	CACCATTTTGCTCGCTTCACCT
39	54	TGTGTCTCCCGTGTTTTTATAACATTCAACAGTGCTCTACTTTTTCCCCGGAGA
40	33	CGGACGTTTCTGGGTTGTTTCCCTCTGCTGGTA
41	33	TCTTCGACTAATAAACAGTTGCAGCCATCACGA
42	22	AGCACCCGCCGTTAGCTGGCGG
43	54	CTATTACGCTTTCTTTTTTTTAAATGAACTTAACCATGACTTTTTTTTGGGACC
44	44	ACATCTTCCGCAATAGCTTTCGGGGAGAACCAGTTAGCCCCGTT
45	44	CGACCAGTGAGTGAATTCACGAGGCGCTACCTAGCAGGCCGACT
46	54	TGTTTGCACAGTGTTTTTCTGTGTTTTTGATTTCAGCTCCTTTTTACGAGCAAG
47	33	CGGTTGATCGCTCGCCGCTACTGGGGATACGTC
48	33	CACTTTCGCGGGCAGGCGTCACACCGTGAATCT
49	22	TTCTTTTCCTCTGCTCCCCGTT
50	54	CACACTGATTCAGTTTTTGCTCTGGGCGGGGTACTTAGATTTTTTGTTTCAGTT
51	33	GGGATGACCAGTTTGCATCGGGTTGGGCTTCCA
52	33	CTAACACAATCGCGCGCCTTTCCAGACTAAGTC
53	22	CCCCTTGCCGATCCGGTATTCG

Table D.26: Staple sequences used to fold the **A-form** 23s rRNA-scaffolded pentagonal bipyramid of 55-bp edge length (staple crossover asymmetry design).

Staple ID	**Length (bp)**	**Staple sequences**
1	54	CTAAGCCAACATCTTTTTCTGGCTGTCCGGTTTCAGGTTCTTTTTTTTTTTCACT
2	44	CCTTGGTCTTCTTTCTTTAAATGATGGCTGCTTTGGACAGGAAC
3	22	CGGCGAGCGGGAGCTATTACGC
4	44	ACATCGTTTCCGCTCCCACTGCTTGTACGTACATGGGCCTTCCC

5	22	CACTTAACCATCGCCTAAGCGT
6	81	CTTCGGCTCCCCTTTTTTATTCGGTTAATACAAAAGGTACTTTTTGCAGTCACAGAC TTTGGGACCTTTTTTTAGCTGGCG
7	33	AGACGCTTGTCACCCTGTATCGCGCGCAAGACT
8	33	CGGTTTCCTGCAACTTAACGCCCAGTTCTTTCC
9	54	CCCATATTCAGACTTTTTAGGATACCAACTGGGTTTCCCCTTTTTATTCGGAAA
10	33	GACCCATTACCTTGCTACAGAATATAATGGAGG
11	33	ATGGTCCCAGTCAGGAGTATTTAGCCTGTCGCT
12	44	TGGTATTCTCTTCGGCCTCGCCTTAGGGGTCGACTCAAGCGCCT
13	22	ACCTGACCACCGACTACGCCTT
14	54	CTAGTTCCTTCACTTTTTCCGAGTTCTCTCACCCTGCCCCTTTTTGATTAACGT
15	44	GGTTCTTTTCGCTCTGACTGCCAGGGCTTTTGCCCCCTCGCCGG
16	22	CCTTTCCCTCATCCTTCATCGC
17	54	ACGCTTATCGCAGTTTTTATTAGCACGCGGTACTGGTTCATTTTTCTATCGGTC
18	33	TTTCCTGGTGATGTTACCTGATGCTTAATATCA
19	33	CCTTACCGTCGCCGGTTATAACGGTTCGAGGCT
20	54	GTGCATTTTTGTGTTTTTTACGGGGCTCCACTAACACACATTTTTCACTGATTC
21	44	CTACTCATCGATAATGATAGTGTGTCGAAACACCGTGTCCCGCC
22	22	GCTCACAGCATATGGATTCAGT
23	54	AGCACCCGCCGTGTTTTTTGTCTCCCGACCGGGTTTCGGGTTTTTTCTATACCC
24	54	GCGCAGGCCGACTTTTTTCGACCAGTGCTTTTCACCCGCTTTTTTTTATCGTTA
25	33	CATCTTCCTTCGGTGCATGGTTTAGCCTCACGA
26	33	CGGACGTTGTCTGGGTTGTTTCCCTCTCCGTTA
27	33	CCGGAGATCCCTTGCCGAAACAGTGCTCCCTAC
28	33	CCAACAACGCAGGCTTACAGAACGCTCCTACCC
29	54	GCATGCCTCACAGTTTTTCACACCTTCGCATAAGCGTCGCTTTTTTGCCGCAGC
30	81	GAATATTAACCTGTTTTTTTTCCCATCTGTGTCGGTTTGGTTTTTGGTACGATTAA GCAGGGCATTTTTTTTGTTGCTTCA
31	33	AAGTACAGTCCGTCCCCCCTTCGCAGTTTCTGA
32	33	TACCTCCACTTATGTCAGCATTCGCACAACACC

33	54	TTGATTGGCCTTTTTTTTCACCCCCAGCGGTTCGGTCCTCTTTTTCAGTTAGTG
34	33	CTCCCGGTATAGCTTTCGGGGAGAACCAACCGG
35	33	GACAACCGAAACCAGCCTACACGCTTAAGCTAT
36	54	GGGTTGGTAAGTCTTTTTGGGATGACCGAATTCACGAGGCTTTTTGCTACCTAA
37	44	TCCGGTATTCGTCAACCTGCCCATGGCTAGATCTGATAACATTC
38	22	CAGTTTGCATCTTACCCAACCT
39	54	TGATTTTCCGGATTTTTTTTGCCTGGATCGCCCGGCCAACTTTTTATAGCCTTC
40	33	TTCAGTTCTCTTTTCCTCGGGGTACTTTCACGC
41	33	CTCAGCCTGCACCGTAGTGCCTCGTCAAGATGT
42	54	ACTGGGGGAATCTTTTTTCGGTTGATTCCCCGGTTCGCCTTTTTTCATTAACCT
43	33	ACATTAGTCCACAAGTCATCCGCTAATGTTCGC
44	33	TCGCCGCTAGGCTCTGGGCTGCTCCCCTTTTCA

D.3.2 Staple sequences for RNA-scaffolded objects with alternative routing schemes

Table D.27: Staple sequences used to fold the **alternative A-form** EGFP mRNA-scaffolded tetrahedron of 66-bp edge length (scaffold crossover asymmetry design).

Staple ID	**Length (bp)**	**Staple sequences**
1	44	GGTCGGGGTAGAGCTCCTCGCCCTTTTGTGTGGCTGCTTCATGT
2	44	CTGCACGCCGTGATGGGCACCACCCCGGTGAACCGGCTGAAGCA
3	44	CAGCTATGACCGTCGCCGATGGGGGTGTTCTGCCACACAGGAAA
4	44	AAGCTTGCATGTGTCGGGCAGCAGCACGGGGCCATGATTACGCC
5	33	CCAGCTTGACGTTGTGGCTGTTGTAGTCTTTGC
6	33	TCAGGGCGTCGCGCTTCTCGTTGGGGTTGTACT
7	22	TGCCCCAGGATCCATGATATAG
8	44	TCGATGTTGTGTCCTGGACGTAGCCTTCGGGCACGCTGCCGTCC
9	44	AGTTCACCTTGGTCCTTGAAGAAGATGGTGCGCGCGGATCTTGA

10	33	CTGAACTTATCGCCCTCGCCCTCGCCGCCGGCG
11	33	GCGGTCACGTCCATGCCGAGAGTGATCGACACG
12	22	GTGGCCGTTTACCGTAGGTGGC
13	33	CCCTCGAACTCGATGCGGTTCACCAGGTTGCCG
14	33	GTGGTGCAGGGCCAGGGCACGGGCAGCGTGTCG
15	22	CTTCACCTCGGATGCCCTTCAG
16	81	TGGTAGTGGTCTTTTTGGCGAGCTGCATGGCGGACTTGTTTTTAAGAAGTCGTGAA TTGTGAGCGTTTTTGATAACAATTT
17	81	ATGCCGTTCTTTTTTTCTGCTTGTCGGGTTGCCGTCCTTTTTTCCTTGAAGTCGCG CGGGTCTTGTTTTTTAGTTGCCGTC
18	81	GAACTCCAGCATTTTTGGACCATGTGAGACTGGGTGCTTTTTTCAGGTAGTGGTCC TGCAGTTACTTTTTTTGTACAGCTC
19	81	GATGAACTTCATTTTTGGGTCAGCTTGCGTCGCCGTCCTTTTTAGCTCGACCAGAG GTCAGGGTGTTTTTGTCACGAGGGT

Table D.28: Staple sequences used to fold the **Hybrid-form** EGFP mRNA-scaffolded tetrahedron of 66-bp edge length (no crossover asymmetry).

Staple ID	**Length (bp)**	**Staple sequences**
1	44	GGTCGGGGTAGAGCTCCTCGCCCTTTTGTGTGGCTGCTTCATGT
2	44	CTGCACGCCGTGATGGGCACCACCCCGGTGAACCGGCTGAAGCA
3	44	CAGCTATGACCGTCGCCGATGGGGGTGTTCTGCCACACAGGAAA
4	44	AAGCTTGCATGTGTCGGGCAGCAGCACGGGGCCATGATTACGCC
5	33	CCAGCTTGACGTTGTGGCTGTTGTAGTCTTTGC
6	33	TCAGGGCGTCGCGCTTCTCGTTGGGGTTGTACT
7	22	TGCCCCAGGATCCATGATATAG
8	44	TCGATGTTGTGTCCTGGACGTAGCCTTCGGGCACGCTGCCGTCC
9	44	AGTTCACCTTGGTCCTTGAAGAAGATGGTGCGCGCGGATCTTGA
10	33	CTGAACTTATCGCCCTCGCCCTCGCCGCCGGCG
11	33	GCGGTCACGTCCATGCCGAGAGTGATCGACACG
12	22	GTGGCCGTTTACCGTAGGTGGC

13	33	CCCTCGAACTCGATGCGGTTCACCAGGTTGCCG
14	33	GTGGTGCAGGGCCAGGGCACGGGCAGCGTGTCG
15	22	CTTCACCTCGGATGCCCTTCAG
16	81	TGGTAGTGGTCTTTTTGGCGAGCTGCATGGCGGACTTGTTTTTAAGAAGTCGTGAA TTGTGAGCGTTTTTGATAACAATTT
17	81	ATGCCGTTCTTTTTTTCTGCTTGTCGGGTTGCCGTCCTTTTTTCCTTGAAGTCGCG CGGGTCTTGTTTTTTAGTTGCCGTC
18	81	GAACTCCAGCATTTTTGGACCATGTGAGACTGGGTGCTTTTTTCAGGTAGTGGTCC TGCAGTTACTTTTTTTGTACAGCTC
19	81	GATGAACTTCATTTTTGGGTCAGCTTGCGTCGCCGTCCTTTTTAGCTCGACCAGAG GTCAGGGTGTTTTTGTCACGAGGGT

Table D.29: Staple sequences used to fold the **alternative A-form** 23s rRNA-scaffolded pentagonal bipyramid of 66-bp edge length (scaffold crossover asymmetry design).

Staple ID	**Length (bp)**	**Staple sequences**
1	44	ACAAAAGGTACCTGACTGCCAGGGCTGAGTCTCTGACCCATTAT
2	44	CCTAAGCGTGCATTAGCACGTCCTTCATCGCCTGCAGTCACACG
3	54	ACCTTACCGACTTTTTGCTTATCGCAGTCCCACTGCTTTTTTTGTACGTACACG
4	33	TTCATATCATTCGGAAATCGCCGGTTAAAGAGC
5	33	TTCGCTTGGTTTACCGGGGCTTCGATCTAACGG
6	54	TTCTTTTCGCCTTTTTTTTCCCTCACGTCGAAACACACTTTTTTGGGTTTCCCC
7	44	CTATCGGTCAGGGATTCAGTTAATGATAGTGTGGTACTGGTTCA
8	44	TAGCCTTGGAGCCGGTTCGCCTCATTAACCTATTCAGGAGTATT
9	54	CACTGATTCAGTTTTTGCTCTGGGCTGGTACTTAGATGTTTTTTTTCAGTTCCC
10	22	TTTTCCTCGGGCTCCCCGTTCG
11	33	TTGATTTCCTCGCCGCTACTGGGGGAACGTCCA
12	33	CTTTCGTGGCAGGCGTCACACCGTATATCTCGG
13	33	AACACACAGCGCGCCTTTCCAGACGCTGTCGGG
14	33	ATGACCCCGTTTGCATCGGGTTGGTAATCCACT

15	54	GATGGTCCCCCTTTTTCATATTCAGACTACGGGGCTGTTTTTTCACCCTGTATC
16	22	GCATTTTTGTGAGGATACCACG
17	33	CAGCATGTTGTCCCGCCCTACTCATCGTCCGCT
18	33	AATTTTTCCACCCCCAGCCACAAGTCAAGCTCA
19	33	CGCCGGGGGTTTCAGGTTCTTTTTCACCATGGC
20	33	TAGATCACACCCAACCTTCAACCTGCCTCCCCT
21	54	TTTCACCCGCTTTTTTTTATCGTTACTCTTGCTACAGATTTTTATATAAGTCGC
22	44	TCGCACTTCTGTCGGCTCCCCTATTCGGTTAACTATGTCAGCAT
23	44	ATGCCTCACAGCCAGTTAAGACTCGGTTTCCCTATACCTCCAGC
24	81	AACATTAGTCGTTTTTGTTCGGTCCTCTCTATACCCTGTTTTTCAACTTAACGCCAC ACCTTCGCTTTTTAGGCTTACAGA
25	22	CGGGTTTCGGGCAGTTAGTGTT
26	54	TGCCGCAGCTTTTTTTCGGTGCATGGTTCTCCCGGTTTTTTTTGATTGGCCTTT
27	44	ACATCTTCCGCAGCTTTCGGGGAGAACCAGCTATTAGCCCCGTT
28	44	CGACCAGTGAGATTCACGAGGCGCTACCTAAATGCAGGCCGACT
29	54	TGTCTCCCGTGTTTTTATAACATTCTCAGTGCTCTACCTTTTTCCCGGAGATGA
30	22	CTTGCCGAAACCGGTATTCGCA
31	22	CACCCGCCGTGTAGCTGGCGGT
32	33	GACGTTAGCTGGGTTGTTTCCCTCTTCTGGTAT
33	33	CTTCGACTATAAACAGTTGCAGCCAGCACGACG
34	54	CTATTACGCTTTTTTTTCTTTAAATGACTTAACCATGATTTTTCTTTGGGACCT
35	22	ATCGTTTCCCATGGCTGCTTCT
36	33	CTTCCCACAAGCCAACATCCTGGCTGTAACATA
37	33	GCCTTCTCGGACAACCGTCGCCCGGCCCTGGGC
38	33	AGCGTCGCACGCTCCCCTACCCAACAACGACTA
39	33	CGCCTTTCATATTAACCTGTTTCCCATCGCATA
40	44	TTTTCCTGGAATTGGTCTTCCGGCGAGCGGGCTTGCTTAGAGGC
41	44	GTTGCTTCAGCGATTAACGTTGGACAGGAACCCGCAGGGCATTT
42	54	AGGGGTCGACTTTTTTCACCCTGCCCCACCGTAGTGCCTTTTTTCGTCATCACG
43	22	GGCCTCGCCTTCAAGTACAGGA

44	54	ACCAGCCTACATTTTTCGCTTAAACCGCGTCCCCCCTTTTTTTCGCAGTAACAC
45	33	CCTGGAAACCTCAGCCTTGATTTTCCGCCGAAG
46	33	TTACGGCACATATCAGCGTGCCTTCTCGATTTG
47	54	GGGTGGAGACATTTTTGCCTGGCCATCGGTACGATTTGTTTTTATGTTACCTGA
48	44	CGTGCAGGTCGCTGACCACCTGTGTCGGTTTGGATTACGCCATT
49	44	ACAAGGAATTTCAAGCGCCTTGGTATTCTCTACGAACTTACCCG
50	81	AGTTCCTTCACTTTTTCCGAGTTCTCTCGCTACCTTAGTTTTTGACCGTTATAGTCA ATTAACCTTTTTTTCCGGCACCGG
51	22	CCATTTTGCCTCGCTTCACCTA
52	54	TTTGCACAGTGTTTTTCTGTGTTTTTAGATTTCAGCTCTTTTTCACGAGCAAGT
53	22	CGCTAACCCCATTACGGCCGCC

Table D.30: Staple sequences used to fold the **alternative A-form** M13 transcript-scaffolded octahedron of 44-bp edge length (scaffold crossover asymmetry design).

Staple ID	**Length (bp)**	**Staple sequences**
1	54	CTGTGTGAAATTTTTTTGTTATCCGCTCAAATCACCATTTTTTCAATATGATAT
2	44	CACACAACATACGAAGTGAGACCGGCCGGAGACAGTCACAATTC
3	54	GCCGGAAGCATTTTTTAAAGTGTAAAGTTAATTGCGTTTTTTTTGCGCTCACTGC
4	54	CTGCATTAATGTTTTTAATCGGCCAACCAGGGTGGTTTTTTTTTTTCTTTTCACC
5	54	CTTTCATCAACTTTTTATTAAATGTGAGGAAGGGCGATTTTTTCGGTGCGGGCC
6	54	GCCTCAGGAAGTTTTTATCGCACTCCAAACCAGGCAAATTTTTGCGCCATTCGC
7	54	CTCCGTGGGAATTTTTCAAACGGCGGAATGGGCGCATCTTTTTGTAACCGTGCA
8	44	TGCGCAACTGTTGGCGAGTAACAACCCGTCGGATTCATTCAGGC
9	44	GCCAGCTGGTAATCATGGTCATAGCTGTTTCTCTTCGCTATTAC
10	54	GCTTGCATGCCTTTTTTGCAGGTCGACCCGGGTACCGATTTTTGCTCGAATTCG
11	54	CGAAAGGGGGATTTTTTGTGCTGCAAGGGTAACGCCAGTTTTTGGTTTTCCCAG
12	44	TGATAAATCGTCTGGCCTTCCTGTAGCCAGTCAACCGTTCTAGC

13	54	TTTTGTTAAATTTTTTCAGCTCATTTTGAACGCCATCATTTTTAAAATAATTCG
14	54	TAATGCCGGAGTTTTTAGGGTAGCTATTACAAAGGCTATTTTTTCAGGTCATTG
15	44	CGTGCCAGCCGCTTTCCAGTCGGGTAGCATTAAAACAAACCTGT
16	54	GAGAATCGATGTTTTTAACGGTAATCGGTCAATCATATTTTTTGTACCCCGGTT
17	54	AGCCCCAAAAATTTTTCAGGAAGATTGTTTAAATTGTATTTTTAACGTTAATAT
18	44	TTGTAAAACAGAAAGATAATCGACGGCCAGTGCCAATCACGACG
19	44	CAGTATCGTCTGCCAGTTTGAGGGCAAATATATAAGGACGACGA
20	44	CGCATTAAATCCTGAGAGTCTGGAGCAAACAATTTGTTAAAATT
21	44	GTGCCGGAGCCAGCTTTCCGGCACAAGTTGGCGATTCGCTTCTG
22	44	TTGGGCGCGCGCGGGGAGAGGCGGGAGATCTTTTGATTTGCGTA
23	44	TGGTGTAGTTGACCGTAATGGGATCAATAGTTAACAGGTCACGT
24	44	AACTCACACCTGGGGTGCCTAATGGGATCCTCTAGAAGTGAGCT

Table D.31: Staple sequences used to fold the **alternative A-form** 23s rRNA-scaffolded octahedron of 66-bp edge length (scaffold crossover asymmetry design).

Staple ID	**Length (bp)**	**Staple sequences**
1	54	GGTGCCTCGTCTTTTTGTCGCGCCTCGGCCGTCGCCCGTTTTTGCCGGCGTGGC
2	44	CTGGGGGCGGGCGGCGTTCGCGCTTCTGGTGCCGGGGGCTTTTC
3	44	TTCGGCGCCGTCCCGCTTTGTCGTTGCTTGTGTGCGTTTGTTGC
4	54	GGTTTGGTGTTTTTTTGCCTGGTGCTTTCCGGCGTGCCTTTTTTCGCGGCGCGC
5	44	GCTTCCGCTGGTGCGGGGGCGGGGCTGGGGTGCCCTTTCCGGGC
6	44	GTTCGGGCTCTCGCGTCCTTCGTCGCCTCTGGCCGCGCGCGCTG
7	54	GCCGGCGCTTGTTTTTTCGCGGGTTGGGGGCTGCTCCCTTTTTCGTTCGCTCGC
8	33	GCGCGCTTCCGGGTTTGCCTGGGGGGCGTTCGT
9	33	GTCGCCTTGGGGTCGCCGGTTGTGGCGCGGCCT
10	22	GGGCCGGGGCGGCCTTGGTTTT
11	54	TTCCCTTCGGCTTTTTTCCCCTGTTCGCGCGCGCCTGGTTTTTGCGTGCTCCCG
12	44	GCCCTGCGGCTGCGTGTGCGTTTTTGTGTGCGGTTCGGGTCTGT

13	44	TGGGGCTCGGTCGCCCTGCTCGTCGGGCTCGCGTGGCGCCCGGT
14	54	GCCCGTGGCTGTTTTTGGTCGCCGGGTGGCTGTCGCCCTTTTTTGTGTCGCGCG
15	44	TGCGGGGCGCTGGTGTTGCCCGGCCTTCGGCCTCTTCGCGGGCT
16	44	CGGCGCGTGGGTGGTCGGTTCGGTCCTCCGGTTCCCCTGCCCGG
17	54	TCGTCCGCTGGTTTTTTTTTTCGGCGTCGTCGCTGCCGTTTTTCGGCTTCGGTG
18	33	TGTGCGGGTGCGGGGTGTGGGTCGCTGCCCCCG
19	33	GCCGCGGGCGGTTTGGTTGGCCTTTCGGCCCGT
20	22	GGGTGCGCGGTGTTGGCCTTGC
21	54	TGTCTGGGCCTTTTTTTCCCGCGTCGTGGCGGTCTGGGTTTTTTTGTTTCCCTC
22	44	GGGTGGGTTCGGCTTCTGGGCCGGCGTCCTGGCTCTGCCCCCGG
23	44	CTGGGTGGCTTGCGCTTTCTTTGGGTGGTGGCTCGGGGCGCTGC
24	54	CCGGCTCGGCCTTTTTGGTGGGCTGTTTCGGGGGGGGCTTTTTCGGCTGTCTCC
25	33	TTGGGCGGTCGGCTCGCCCTGCCCCGGGTCTTC
26	33	CGCGCGGGCGTGGTTTGGCCCCGTTGCTTGGCG
27	22	GGGCCCTTGGTCGCCTTGGGGG
28	54	CGTCGGCTGCGTTTTTCCTTTCGGCCTCTTCCGGCGGGTTTTTCGGGCTTTTCG
29	33	GGGGTGTTCCCCTTCGCGGTGGCGCCGTTGGGG
30	33	CCTTGGCTTTCCCGCTTGGCCGTGGCTGGTGCG
31	22	GGCCTGTTTCCCTTCTCCGTCC
32	54	CCCGTGGTGGCTTTTTGTTCTCCGGTGGGCCCCCTTGCTTTTTCGGGGCGGTGC
33	33	TTGGTGGTTCGCCTCGTTGGCCTGTGGCCCGCC
34	33	GTGTGTCTTTCGCGGCGGGCGTTGGCGGTTCGG
35	22	GGTGTGTCGGGGTTCCCCCGGT
36	54	CTCGGGGTGCTTTTTTTGGGTGTTTCGGCGCGCTGGGTTTTTTTTCCCCGTTCG
37	33	GGGGGTGGGGTCGGTCGGGGGTGTTTGGTTGGT
38	33	TTCTTTTCCGCTGCTGGGGGGGTCTCGGCCTTG
39	22	TCCCCCCGTGTGGTTCGCTGTC
40	54	TTCGCCTTTCCTTTTTCTCGCGGTGCTTCGGGCGGGGTTTTTTGCCGCGTGTCC
41	33	TCCCCTCGTGCGCGGTTTCGGGTTCTTTGGTGG

42	33	GTCGGGGTTTCGCGGTTTGCGTCGGGTTTTCGC
43	22	CCGGGGTTCTTCTGCTTGTGCG

D.4 Anchoring RNA fragments on RNA-scaffolded origami

The staples used to fold all pentagonal bipyramids with 66 bp edge length (rPB66, rPB66_RREvertex, rPB66_RREout, and rPB66_RREin) are the same as those used to fold the 23s rRNA-scaffolded rPB66 in Chapter 4 (see Table D.25). The 23s rRNA scaffold without inserted RRE is also given Table D.18.

Table D.32: Scaffold sequences for 23s rRNA-scaffolded pentagonal bipyramids with anchored RRE (inserted RRE sequence indicated in bold).

Name	**Sequence**
23s RRE Vertex Out scaffold	GCCCUGGCAGUCAGAGGCGAUGAAGGACGUGCUAAUCUGCGAUAAGC**GGAGC UUUGUUCCUUGGGUUCUUGGGAGCAGCAGGAAGCACUAUGGGCGCA GCGUCAAUGACGCUGACGGUACAGGCCAGACAAUUAUUGUCUGAUA UAGUGCAGCAGCAGAACAAUUUGCUGAGGGCUAUUGAGGCGCAACA GCAUCUGUUGCAACUCACAGUCUGGGGCAUCAAACAGCUCCAGGCA AGAAUCCUGGCUGUGGAAAGAUACCUAAAGGAUCAACAGCUCC**GUCG GUAAGGUGAUAUGAACCGUUAUAACCGGCGAUUUCCGAAUGGGGAAACCCAGU GUGUUUCGACACACUAUCAUUAACUGAAUCCAUAGGUUAAUGAGGCGAACCGG GGGAACUGAAACAUCUAAGUACCCCGAGGAAAAGAAAUCAACCGAGAUUCCCC CAGUAGCGGCGAGCGAACGGGGAGCAGCCCAGAGCCUGAAUCAGUGUGUGUGU UAGUGGAAGCGUCUGGAAAGGCGCGCGAUACAGGGUGACAGCCCCGUACACAA AAAUGCACAUGCUGUGAGCUCGAUGAGUAGGGCGGGACACGUGGUAUCCUGUC UGAAUAUGGGGGGACCAUCCUCCAAGGCUAAAUACUCCUGACUGACCGAUAGU GAACCAGUACCGUGAGGGAAAGGCGAAAAGAACCCCGGCGAGGGGAGUGAAAA AGAACCUGAAACCGUGUACGUACAAGCAGUGGGAGCACGCUUAGGCGUGUGAC UGCGUACCUUUUGUAUAAUGGGUCAGCGACUUAUAUUCUGUAGCAAGGUUAAC CGAAUAGGGGAGCCGAAGGGAAACCGAGUCUUAACUGGGCGUUAAGUUGCAGG GUAUAGACCCGAAACCCGGUGAUCUAGCCAUGGGCAGGUUGAAGGUUGGGUAA CACUAACUGGAGGACCGAACCGACUAAUGUUGAAAAAUUAGCGGAUGACUUGU GGCUGGGGGUGAAAGGCCAAUCAAACCGGGAGAUAGCUGGUUCUCCCCGAAAG CUAUUUAGGUAGCGCCUCGUGAAUUCAUCUCCGGGGGUAGAGCACUGUUUCGG CAAGGGGGUCAUCCCGACUUACCAACCCGAUGCAAACUGCGAAUACCGGAGAA UGUUAUCACGGGAGACACACGGCGGGUGCUAACGUCCGUCGUGAAGAGGGAAA CAACCCAGACCGCCAGCUAAGGUCCCAAAGUCAUGGUUAAGUGGGAAACGAUG UGGGAAGGCCCAGACAGCCAGGAUGUUGGCUUAGAAGCAGCCAUCAUUUAAAG AAAGCGUAAUAGCUCACUGGUCGAGUCGGCCUGCGCGGAAGAUGUAACGGGGC UAAACCAUGCACCGAAGCUGCGGCAGCGACGCUUAUGCGUUGUUGGGUAGGGG AGCGUUCUGUAAGCCUGCGAAGGUGUGCUGUGAGGCAUGCUGGAGGUAUCAGA AGUGCGAAUGCUGACAUAAGUAACGAUAAAGCGGGUGAAAAGCCCGCUCGCCG GAAGACCAAGGGUUCCUGUCCAACGUUAAUCGGGGCAGGGUGAGUCGACCCCU AAGGCGAGGCCGAAAGGCGUAGUCGAUGGGAAACAGGUUAAUAUUCCUGUACU UGGUGUUACUGCGAAGGGGGGACGGAGAAGGCUAUGUUGGCCGGGCGACGGUU GUCCCGGUUUAAGCGUGUAGGCUGGUUUUCCAGGCAAAUCCGGAAAAUCAAGG CUGAGGCGUGAUGACGAGGCACUACGGUGCUGAAGCAACAAAUGCCCUGCUUC CAGGAAAAGCCUCUAAGCAUCAGGUAACAUCAAAUCGUACCCCAAACCGACACA GGUGGUCAGGUAGAGAAUACCAAGGCGCUUGAGAGAACUCGGGUGAAGGAACU AGGCAAAAUGGUGCCGUAACUUCGGGAGAAGGCACGCUGAUAUGUAGGUGAAG CGACUUGCUCGUGGAGCUGAAAUCAGUCGAAGAUACCAGCUGGCUGCAACUGU UUAUUAAAAACACAGCACUGUGCAAACACGAAAGUGGACGUAUACGGUGUGAC GCCUGCCCGGUGCCGGAAGGUUAAUUGAUGGGGUUAGCGCAAGCGAAGCUCUU GAUCGAAGCCCCGGUAAACGGCGGCCGUAACUAUAACGGUCCUAAGGUAGCGA AAUUCCUUGUCGGGUAAGUUCCGACCUGCACGAAUGGCGUAAUGAUGGCCAGG CUGUCUCCACCCGAGACUCA

23s RRE Edge Out scaffold	GCCCUGGCAGUCAGAGGCGAUGAAGGACGUGCUAAUCUGCGAUAAGCGUCGGU AAGGUGAUAUGAACCGUUAUAACCGGCGAUUUCCGAAUGGGGAAACCCAGUGU GUUUCGACACACUAUCAUUAACUGAAUCCAUAGGUUAAUGAGGCGAACCGGGG GAACUGAAACAUCUAAGUACCCCGAGGAAAAGAAAUCAACCGAGAUUCCCCCA GUAGCGGCGAGCGAACGGGGAGCAGCCCAGAGCCUGAAUCAGUGUGUGUGUUA GUGGAAGCGUCUGGAAAGGCGCGCGAUACAGGGUGACAGCCCCGUACACAAAA AUGCACAUGCUGUGAGCUCGAUGAGUAGGGCGGGACACGUGGUAUCCUGUCUG AAUAUGGGGGGACCAUCCUCCAAGGCUAAAUACUCCUGACUGACCGAUAGUGA ACCAGUACCGUGAGGGAAAGGCGAAAAGAACCCCGGCGAGGGGAGUGAAAAAG AACCUGAAACCGUGUACGUACAAGCAGUGGGAGCACGCUUAGGCGUGUGACUG CGUACCUUUUGUAUAAUGGGUCAGCGACUUAUAUUCUGUAGCAAGGUUAACCG AAUAGGGGAGCCGAAGGGAAACCGAGUCUUAACUGGGCGUUAAGUUGCAGGGU AUAGACCCGAAACCCGGUGAUCUAGCCAUGGGCAGGUUGAAGGUUGGGUAACA CUAACUGGAGGACCGAACCGACUAAUGUUGAAAAAUUAGCGGAUGACUUGUGG CUGGGGGUGAAAGGCCAAUCAAACCGGGAGAUAGCUGGUUCUCCCCGAAAGCU AUUUAGGUAGCGCCUCGUGAAUUCAUCUCCGGGGGUAGAGCACUGUUUCGGCA AGGGGGUCAUCCCGACUUACCAACCCGAUGCAAACUGCGAAUACCGGAGAAUG UUAUCACGGGAGACACACGGCGGGUGCUAACGUCCGUCGUGAAGAGGGAAACA ACCCAGACCGCCAGCUAAGGUCCCAAAGUCAUGGUUAAGUGGGAAACGAUGUG GGAAGGCCCAGACAGCCAGGAUGUUGGCUUAGAAGCAGCCAUCAUUUAAAGAA AGCGUAAUAGCUCACUGGUCGAGUCGGCCUGCGCGGAAGAUGUAACGGGGCUA AACCAUGCACCGAAGCUGCGGCAGCGACGCUUAUGCGUUGUUGGGUAGGGGAG CGUUCUGUAAGCCUGCGAAGGUGUGCUGUGAGGCAUGCUGGAGGUAUCAGAAG UGCGAAUGCUGACAUAAGUAACGAUAAAGCGGGUGAAAAGCCCGCUCGCC**GGA** **GCUUUGUUCCUUGGGUUCUUGGGAGCAGCAGGAAGCACUAUGGGCG** **CAGCGUCAAUGACGCUGACGGUACAGGCCAGACAAUUAUUGUCUGA** **UAUAGUGCAGCAGCAGAACAAUUUGCUGAGGGCUAUUGAGGCGCAA** **CAGCAUCUGUUGCAACUCACAGUCUGGGGCAUCAAACAGCUCCAGG** **CAAGAAUCCUGGCUGUGGAAAGAUACCUAAAGGAUCAACAGCUCC**GG AAGACCAAGGGUUCCUGUCCAACGUUAAUCGGGGCAGGGUGAGUCGACCCCUA AGGCGAGGCCGAAAGGCGUAGUCGAUGGGAAACAGGUUAAUAUUCCUGUACUU GGUGUUACUGCGAAGGGGGGACGGAGAAGGCUAUGUUGGCCGGGCGACGGUU GUCCCGGUUUAAGCGUGUAGGCUGGUUUUCCAGGCAAAUCCGGAAAAUCAAGG CUGAGGCGUGAUGACGAGGCACUACGGUGCUGAAGCAACAAAUGCCCUGCUUC CAGGAAAAGCCUCUAAGCAUCAGGUAACAUCAAAUCGUACCCCAAACCGACACA GGUGGUCAGGUAGAGAAUACCAAGGCGCUUGAGAGAACUCGGGUGAAGGAACU AGGCAAAAUGGUGCCGUAACUUCGGGAGAAGGCACGCUGAUAUGUAGGUGAAG CGACUUGCUCGUGGAGCUGAAAUCAGUCGAAGAUACCAGCUGGCUGCAACUGU UUAUUAAAAACACAGCACUGUGCAAACACGAAAGUGGACGUAUACGGUGUGAC GCCUGCCCGGUGCCGGAAGGUUAAUUGAUGGGGUUAGCGCAAGCGAAGCUCUU GAUCGAAGCCCCGGUAAACGGCGGCCGUAACUAUAACGGUCCUAAGGUAGCGA AAUUCCUUGUCGGGUAAGUUCCGACCUGCACGAAUGGCGUAAUGAUGGCCAGG CUGUCUCCACCCGAGACUCA

23s RRE Edge In scaffold	GCCCUGGCAGUCAGAGGCGAUGAAGGACGUGCUAAUCUGCGAUAAGCGUCGGU AAGGUGAUAUGAACCGUUAUAACCGGCGAUUUCCGAAUGGGGAAACCCAGUGU GUUUCGACACACUAUCAUUAACUGAAUCCAUAGGUUAAUGAGGCGAACCGGGG GAACUGAAACAUCUAAGUACCCCGAGGAAAAGAAAUCAACCGAGAUUCCCCCA GUAGCGGCGAGCGAACGGGGAGCAGCCCAGAGCCUGAAUCAGUGUGUGUGUUA GUGGAAGCGUCUGGAAAGGCGCGCGAUACAGGGUGACAGCCCCGUACACAAAA AUGCACAUGCUGUGAGCUCGAUGAGUAGGGCGGGACACGUGGUAUCCUGUCUG AAUAUGGGGGGACCAUCCUCCAAGGCUAAAUACUCCUGACUGACCGAUAGUGA ACCAGUACCGUGAGGGAAAGGCGAAAAGAACCCCGGCGAGGGGAGUGAAAAAG AACCUGAAACCGUGUACGUACAAGCAGUGGGAGCACGCUUAGGCGUGUGACUG CGUACCUUUUGUAUAAUGGGUCAGCGACUUAUAUUCUGUAGCAAGGUUAACCG AAUAGGGGAGCCGAAGGGAAACCGAGUCUUAACUGGGCGUUAAGUUGCAGGGU AUAGACCCGAAACCCGGUGAUCUAGCCAUGGGCAGGUUGAAGGUUGGGUAACA CUAACUGGAGGACCGAACCGACUAAUGUUGAAAAAUUAGCGGAUGACUUGUGG CUGGGGGUGAAAGGCCAAUCAAACCGGGAGAUAGCUGGUUCUCCCCGAAAGCU AUUUAGGUAGCGCCUCGUGAAUUCAUCUCCGGGGGUAGAGCACUGUUUCGGCA AGGGGGUCAUCCCGACUUACCAACCCGAUGCAAACUGCGAAUACCGGAGAAUG UUAUCACGGGAGACACACGGCGGGUGCUAACGUCCGUCGUGAAGAGGGAAACA ACCCAGACCGCCAGCUAAGGUCCCAAAGUCAUGGUUAAGUGGGAAACGAUGUG GGAAGGCCCAGACAGCCAGGAUGUUGGCUUAGAAGCAGCCAUCAUUUAAAGAA AGCGUAAUAGCUCACUGGUCGAGUCGGCCUGCGCGGAAGAUGUAACGGGGCUA AACCAUGCACCGAAGCUGCGGCAGCGACGCUUAUGCGUUGUUGGGUAGGGGAG CGUUCUGUAAGCCUGCGAAGGUGUGCUGUGAGGCAUGCUGGAGGUAUCAGAAG UGCGAAUGCUGACAUAAGUAACGAUAAAGCGGGUGAAAAGCCCGCUCGCCGGA AGACCAAGGGUUCCUGUCCAACGUUAAUCGGGGCAGGGUGAGUCGACCCCUAA GGCGAGGCCGAAAGGCGUAGUCGAUGGGAAACAGGUUAAUAUUCCUGUACUUG GUGUUACUGCGAAGGGGGGACGGAGAAGGCUAUGUUGGCCGGGCGACGGUUGU CCCGGUUUAAGCGUGUAGGCUGGUUUUCCAGGCAAAUCCGGAAAAUCAAGGCU GAGGCGUGAUGACGAGGCACUACGGUGCUGAAGCAACAAAUGCCCUGCUUCCA GGAAAAGCCUCUAAGCAUCAGGUAACAUCAAAUCGUACCCCAAACCG**GGAGCU** **UUGUUCCUUGGGUUCUUGGGAGCAGCAGGAAGCACUAUGGGCGCAG** **CGUCAAUGACGCUGACGGUACAGGCCAGACAAUUAUUGUCUGAUAU** **AGUGCAGCAGCAGAACAAUUUGCUGAGGGCUAUUGAGGCGCAACAG** **CAUCUGUUGCAACUCACAGUCUGGGGCAUCAAACAGCUCC**AGGCAAG AAUCCUGGCUGUGGAAAGAUACCUAAAGGAUCAACAGCUCCACACAGGUGGUC AGGUAGAGAAUACCAAGGCGCUUGAGAGAACUCGGGUGAAGGAACUAGGCAAA AUGGUGCCGUAACUUCGGGAGAAGGCACGCUGAUAUGUAGGUGAAGCGACUUG CUCGUGGAGCUGAAAUCAGUCGAAGAUACCAGCUGGCUGCAACUGUUUAUUAA AAACACAGCACUGUGCAAACACGAAAGUGGACGUAUACGGUGUGACGCCUGCC CGGUGCCGGAAGGUUAAUUGAUGGGGUUAGCGCAAGCGAAGCUCUUGAUCGAA GCCCCGGUAAACGGCGGCCGUAACUAUAACGGUCCUAAGGUAGCGAAAUUCCU UGUCGGGUAAGUUCCGACCUGCACGAAUGGCGUAAUGAUGGCCAGGCUGUCUC CACCCGAGACUCA

Table D.33: DNA template sequences and corresponding primers (same primers for all templates) for the 23s rRNA scaffolds with RRE inserts listed in Table D.32.

Name	**Sequence**
23SdomIIV for	CTTAAGTAATACGACTCACTATAGCCCTGGCAGTCAGAGG
23SdomIIV rev	TGAGTCTCGGGTGGAGACAG

23s gblock	RREvertex	GCCCTGGCAGTCAGAGGCGATGAAGGACGTGCTAATCTGCGATAAGCGGAGCT TTGTTCCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAAT GACGCTGACGGTACAGGCCAGACAATTATTGTCTGATATAGTGCAGCAGCAGA ACAATTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCT GGGGCATCAAACAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGATACCTAAAG GATCAACAGCTCCGTCGGTAAGGTGATATGAACCGTTATAACCGGCGATTTCCG AATGGGGAAACCCAGTGTGTTTCGACACACTATCATTAACTGAATCCATAGGTT AATGAGGCGAACCGGGGGAACTGAAACATCTAAGTACCCCGAGGAAAAGAAAT CAACCGAGATTCCCCCAGTAGCGGCGAGCGAACGGGGAGCAGCCCAGAGCCTG AATCAGTGTGTGTGTTAGTGGAAGCGTCTGGAAAGGCGCGCGATACAGGGTGA CAGCCCCGTACACAAAAATGCACATGCTGTGAGCTCGATGAGTAGGGCGGGAC ACGTGGTATCCTGTCTGAATATGGGGGGACCATCCTCCAAGGCTAAATACTCCT GACTGACCGATAGTGAACCAGTACCGTGAGGGAAAGGCGAAAAGAACCCCGGC GAGGGGAGTGAAAAAGAACCTGAAACCGTGTACGTACAAGCAGTGGGAGCACG CTTAGGCGTGTGACTGCGTACCTTTTGTATAATGGGTCAGCGACTTATATTCTG TAGCAAGGTTAACCGAATAGGGGAGCCGAAGGGAAACCGAGTCTTAACTGGGC GTTAAGTTGCAGGGTATAGACCCGAAACCCGGTGATCTAGCCATGGGCAGGTT GAAGGTTGGGTAACACTAACTGGAGGACCGAACCGACTAATGTTGAAAAATTA GCGGATGACTTGTGGCTGGGGGTGAAAGGCCAATCAAACCGGGAGATAGCTGG TTCTCCCCGAAAGCTATTTAGGTAGCGCCTCGTGAATTCATCTCCGGGGGTAGA GCACTGTTTCGGCAAGGGGGTCATCCCGACTTACCAACCCGATGCAAACTGCGA ATACCGGAGAATGTTATCACGGGAGACACACGGCGGGTGCTAACGTCCGTCGT GAAGAGGGAAACAACCCAGACCGCCAGCTAAGGTCCCAAAGTCATGGTTAAGT GGGAAACGATGTGGGAAGGCCCAGACAGCCAGGATGTTGGCTTAGAAGCAGCC ATCATTTAAAGAAAGCGTAATAGCTCACTGGTCGAGTCGGCCTGCGCGGAAGA TGTAACGGGGCTAAACCATGCACCGAAGCTGCGGCAGCGACGCTTATGCGTTG TTGGGTAGGGGAGCGTTCTGTAAGCCTGCGAAGGTGTGCTGTGAGGCATGCTG GAGGTATCAGAAGTGCGAATGCTGACATAAGTAACGATAAAGCGGGTGAAAAG CCCGCTCGCCGGAAGACCAAGGGTTCCTGTCCAACGTTAATCGGGGCAGGGTG AGTCGACCCCTAAGGCGAGGCCGAAAGGCGTAGTCGATGGGAAACAGGTTAAT ATTCCTGTACTTGGTGTTACTGCGAAGGGGGGACGGAGAAGGCTATGTTGGCC GGGCGACGGTTGTCCCGGTTTAAGCGTGTAGGCTGGTTTTCCAGGCAAATCCG GAAAATCAAGGCTGAGGCGTGATGACGAGGCACTACGGTGCTGAAGCAACAAA TGCCCTGCTTCCAGGAAAAGCCTCTAAGCATCAGGTAACATCAAATCGTACCCC AAACCGACACAGGTGGTCAGGTAGAGAATACCAAGGCGCTTGAGAGAACTCGG GTGAAGGAACTAGGCAAAATGGTGCCGTAACTTCGGGAGAAGGCACGCTGATA TGTAGGTGAAGCGACTTGCTCGTGGAGCTGAAATCAGTCGAAGATACCAGCTG GCTGCAACTGTTTATTAAAAACACAGCACTGTGCAAACACGAAAGTGGACGTAT ACGGTGTGACGCCTGCCCGGTGCCGGAAGGTTAATTGATGGGGTTAGCGCAAG CGAAGCTCTTGATCGAAGCCCCGGTAAACGGCGGCCGTAACTATAACGGTCCTA AGGTAGCGAAATTCCTTGTCGGGTAAGTTCCGACCTGCACGAATGGCGTAATG ATGGCCAGGCTGTCTCCACCCGAGACTCA

23s RREout gblock	GCCCTGGCAGTCAGAGGCGATGAAGGACGTGCTAATCTGCGATAAGCGTCGGT AAGGTGATATGAACCGTTATAACCGGCGATTTCCGAATGGGGAAACCCAGTGT GTTTCGACACACTATCATTAACTGAATCCATAGGTTAATGAGGCGAACCGGGGG AACTGAAACATCTAAGTACCCCGAGGAAAAGAAATCAACCGAGATTCCCCCAGT AGCGGCGAGCGAACGGGGAGCAGCCCAGAGCCTGAATCAGTGTGTGTGTTAGT GGAAGCGTCTGGAAAGGCGCGCGATACAGGGTGACAGCCCCGTACACAAAAAT GCACATGCTGTGAGCTCGATGAGTAGGGCGGGACACGTGGTATCCTGTCTGAA TATGGGGGGACCATCCTCCAAGGCTAAATACTCCTGACTGACCGATAGTGAACC AGTACCGTGAGGGAAAGGCGAAAAGAACCCCGGCGAGGGGAGTGAAAAAGAAC CTGAAACCGTGTACGTACAAGCAGTGGGAGCACGCTTAGGCGTGTGACTGCGT ACCTTTTGTATAATGGGTCAGCGACTTATATTCTGTAGCAAGGTTAACCGAATA GGGGAGCCGAAGGGAAACCGAGTCTTAACTGGGCGTTAAGTTGCAGGGTATAG ACCCGAAACCCGGTGATCTAGCCATGGGCAGGTTGAAGGTTGGGTAACACTAA CTGGAGGACCGAACCGACTAATGTTGAAAAATTAGCGGATGACTTGTGGCTGG GGGTGAAAGGCCAATCAAACCGGGAGATAGCTGGTTCTCCCCGAAAGCTATTT AGGTAGCGCCTCGTGAATTCATCTCCGGGGGTAGAGCACTGTTTCGGCAAGGG GGTCATCCCGACTTACCAACCCGATGCAAACTGCGAATACCGGAGAATGTTATC ACGGGAGACACACGGCGGGTGCTAACGTCCGTCGTGAAGAGGGAAACAACCCA GACCGCCAGCTAAGGTCCCAAAGTCATGGTTAAGTGGGAAACGATGTGGGAAG GCCCAGACAGCCAGGATGTTGGCTTAGAAGCAGCCATCATTTAAAGAAAGCGT AATAGCTCACTGGTCGAGTCGGCCTGCGCGGAAGATGTAACGGGGCTAAACCA TGCACCGAAGCTGCGGCAGCGACGCTTATGCGTTGTTGGGTAGGGGAGCGTTC TGTAAGCCTGCGAAGGTGTGCTGTGAGGCATGCTGGAGGTATCAGAAGTGCGA ATGCTGACATAAGTAACGATAAAGCGGGTGAAAAGCCCGCTCGCCGGAGCTTT GTTCCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAATGA CGCTGACGGTACAGGCCAGACAATTATTGTCTGATATAGTGCAGCAGCAGAAC AATTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGG GGCATCAAACAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGATACCTAAAGGA TCAACAGCTCCGGAAGACCAAGGGTTCCTGTCCAACGTTAATCGGGGCAGGGT GAGTCGACCCCTAAGGCGAGGCCGAAAGGCGTAGTCGATGGGAAACAGGTTAA TATTCCTGTACTTGGTGTTACTGCGAAGGGGGGACGGAGAAGGCTATGTTGGC CGGGCGACGGTTGTCCCGGTTTAAGCGTGTAGGCTGGTTTTCCAGGCAAATCC GGAAAATCAAGGCTGAGGCGTGATGACGAGGCACTACGGTGCTGAAGCAACAA ATGCCCTGCTTCCAGGAAAAGCCTCTAAGCATCAGGTAACATCAAATCGTACCC CAAACCGACACAGGTGGTCAGGTAGAGAATACCAAGGCGCTTGAGAGAACTCG GGTGAAGGAACTAGGCAAAATGGTGCCGTAACTTCGGGAGAAGGCACGCTGAT ATGTAGGTGAAGCGACTTGCTCGTGGAGCTGAAATCAGTCGAAGATACCAGCT GGCTGCAACTGTTTATTAAAAACACAGCACTGTGCAAACACGAAAGTGGACGTA TACGGTGTGACGCCTGCCCGGTGCCGGAAGGTTAATTGATGGGGTTAGCGCAA GCGAAGCTCTTGATCGAAGCCCCGGTAAACGGCGGCCGTAACTATAACGGTCCT AAGGTAGCGAAATTCCTTGTCGGGTAAGTTCCGACCTGCACGAATGGCGTAAT GATGGCCAGGCTGTCTCCACCCGAGACTCA

23s RREin gblock	GCCCTGGCAGTCAGAGGCGATGAAGGACGTGCTAATCTGCGATAAGCGTCGGT AAGGTGATATGAACCGTTATAACCGGCGATTTCCGAATGGGGAAACCCAGTGT GTTTCGACACACTATCATTAACTGAATCCATAGGTTAATGAGGCGAACCGGGGG AACTGAAACATCTAAGTACCCCGAGGAAAAGAAATCAACCGAGATTCCCCCAGT AGCGGCGAGCGAACGGGGAGCAGCCCAGAGCCTGAATCAGTGTGTGTGTTAGT GGAAGCGTCTGGAAAGGCGCGCGATACAGGGTGACAGCCCCGTACACAAAAAT GCACATGCTGTGAGCTCGATGAGTAGGGCGGGACACGTGGTATCCTGTCTGAA TATGGGGGGACCATCCTCCAAGGCTAAATACTCCTGACTGACCGATAGTGAACC AGTACCGTGAGGGAAAGGCGAAAAGAACCCCGGCGAGGGGAGTGAAAAAGAAC CTGAAACCGTGTACGTACAAGCAGTGGGAGCACGCTTAGGCGTGTGACTGCGT ACCTTTTGTATAATGGGTCAGCGACTTATATTCTGTAGCAAGGTTAACCGAATA GGGGAGCCGAAGGGAAACCGAGTCTTAACTGGGCGTTAAGTTGCAGGGTATAG ACCCGAAACCCGGTGATCTAGCCATGGGCAGGTTGAAGGTTGGGTAACACTAA CTGGAGGACCGAACCGACTAATGTTGAAAAATTAGCGGATGACTTGTGGCTGG GGGTGAAAGGCCAATCAAACCGGGAGATAGCTGGTTCTCCCCGAAAGCTATTT AGGTAGCGCCTCGTGAATTCATCTCCGGGGGTAGAGCACTGTTTCGGCAAGGG GGTCATCCCGACTTACCAACCCGATGCAAACTGCGAATACCGGAGAATGTTATC ACGGGAGACACACGGCGGGTGCTAACGTCCGTCGTGAAGAGGGAAACAACCCA GACCGCCAGCTAAGGTCCCAAAGTCATGGTTAAGTGGGAAACGATGTGGGAAG GCCCAGACAGCCAGGATGTTGGCTTAGAAGCAGCCATCATTTAAAGAAAGCGT AATAGCTCACTGGTCGAGTCGGCCTGCGCGGAAGATGTAACGGGGCTAAACCA TGCACCGAAGCTGCGGCAGCGACGCTTATGCGTTGTTGGGTAGGGGAGCGTTC TGTAAGCCTGCGAAGGTGTGCTGTGAGGCATGCTGGAGGTATCAGAAGTGCGA ATGCTGACATAAGTAACGATAAAGCGGGTGAAAAGCCCGCTCGCCGGAAGACC AAGGGTTCCTGTCCAACGTTAATCGGGGCAGGGTGAGTCGACCCCTAAGGCGA GGCCGAAAGGCGTAGTCGATGGGAAACAGGTTAATATTCCTGTACTTGGTGTT ACTGCGAAGGGGGGACGGAGAAGGCTATGTTGGCCGGGCGACGGTTGTCCCGG TTTAAGCGTGTAGGCTGGTTTTCCAGGCAAATCCGGAAAATCAAGGCTGAGGC GTGATGACGAGGCACTACGGTGCTGAAGCAACAAATGCCCTGCTTCCAGGAAA AGCCTCTAAGCATCAGGTAACATCAAATCGTACCCCAAACCGGGAGCTTTGTTC CTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAATGACGCT GACGGTACAGGCCAGACAATTATTGTCTGATATAGTGCAGCAGCAGAACAATTT GCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCA TCAAACAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGATACCTAAAGGATCAAC AGCTCCACACAGGTGGTCAGGTAGAGAATACCAAGGCGCTTGAGAGAACTCGG GTGAAGGAACTAGGCAAAATGGTGCCGTAACTTCGGGAGAAGGCACGCTGATA TGTAGGTGAAGCGACTTGCTCGTGGAGCTGAAATCAGTCGAAGATACCAGCTG GCTGCAACTGTTTATTAAAAACACAGCACTGTGCAAACACGAAAGTGGACGTAT ACGGTGTGACGCCTGCCCGGTGCCGGAAGGTTAATTGATGGGGTTAGCGCAAG CGAAGCTCTTGATCGAAGCCCCGGTAAACGGCGGCCGTAACTATAACGGTCCTA AGGTAGCGAAATTCCTTGTCGGGTAAGTTCCGACCTGCACGAATGGCGTAATG ATGGCCAGGCTGTCTCCACCCGAGACTCA

Bibliography

1. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35,** 849–857 (1983).
2. Kazantsev, A. V. *et al.* Crystal structure of a bacterial ribonuclease P RNA (2005).
3. Marquez, S. M., Chen, J. L., Evans, D. & Pace, N. R. Structure and Function of Eukaryotic Ribonuclease P RNA. *Molecular Cell* **24,** 445–456 (2006).
4. Mustoe, A. M. *et al.* Pervasive Regulatory Functions of mRNA Structure Revealed by High-Resolution SHAPE Probing. *Cell* **173,** 181–195 (2018).
5. Baird, N. J. *et al.* Discrete Structure of an RNA Folding Intermediate Revealed by Cryo-electron Microscopy. *Journal of the American Chemical Society2* **132,** 16352–16353 (2010).
6. Ganser, L. R., Kelly, M. L., Herschlag, D. & Al-Hashimi, H. M. The roles of structural dynamics in the cellular functions of RNAs. *Nature Reviews Molecular Cell Biology* **20,** 474–489 (June 2019).
7. Chauvier, A., Cabello-Villegas, J. & Walter, N. G. Probing RNA structure and interaction dynamics at the single molecule level. *Methods* **162-163,** 3–11 (June 2019).
8. Ray, S., Chauvier, A. & Walter, N. G. Kinetics coming into focus: single-molecule microscopy of riboswitch dynamics. *RNA Biology* **16,** 1077–1085 (Sept. 2019).
9. Boeras, I. *et al.* The basal translation rate of authentic HIV-1 RNA is regulated by 5'UTR nt-pairings at junction of R and U5. *Scientific Reports* **7,** 1–10 (2017).
10. Keane, S. C. *et al.* Structure of the HIV - 1 RNA packaging signal. *Science* **348,** 917–922 (2015).
11. Kuzembayeva, M., Dilley, K., Sardo, L. & Hu, W. S. Life of psi: How full-length HIV-1 RNAs become packaged genomes in the viral particles. *Virology* **454-455,** 362–370 (2014).
12. Kleiman, L. tRNALys3: The primer tRNA for reverse transcription in HIV-1. *IUBMB Life* **53,** 107–114 (2002).

13. Ingemarsdotter, C. K., Zeng, J., Long, Z., Lever, A. M. & Kenyon, J. C. An RNA-binding compound that stabilizes the HIV-1 gRNA packaging signal structure and specifically blocks HIV-1 RNA encapsidation. *Retrovirology* **15,** 1–14 (2018).

14. Wang, Q., Barr, I., Guo, F. & Lee, C. Evidence of a novel RNA secondary structurein the coding region of HIV-1 pol gene. *RNA* **14,** 2478–2488 (Oct. 2008).

15. Serganov, A. & Patel, D. J. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat Rev Genet* **8,** 776–790 (2007).

16. Siegfried, N. A., Busan, S., Rice, G. M., Nelson, J. A. & Weeks, K. M. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nature methods* (2014).

17. Smola, M. J., Rice, G. M., Busan, S., Siegfried, N. A. & Weeks, K. M. Selective 2 ' -hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct , versatile and accurate RNA structure analysis. *Nature Protocols* **10,** 1643–1669 (2015).

18. Zubradt, M. *et al.* DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nature Methods* **14,** 75–82 (2017).

19. Peattie, D. A. & Gilbert, W. Chemical probes for higher-order structure in RNA. *Proceedings of the National Academy of Sciences* **77,** 4679–4682 (Aug. 1980).

20. Cordero, P., Kladwang, W., VanLang, C. C. & Das, R. Quantitative Dimethyl Sulfate Mapping for Automated RNA Secondary Structure Inference. *Biochemistry* **51,** 7037–7039 (Aug. 2012).

21. Weeks, K. M. Advances in RNA structure analysis by chemical probing. *Current Opinion in Structural Biology* **20,** 295–304 (June 2010).

22. Wilkinson, K. A. *et al.* Influence of nucleotide identity on ribose 2'-hydroxyl reactivity in RNA. *RNA* **15,** 1314–1321 (May 2009).

23. Homan, P. J. *et al.* Single-molecule correlated chemical probing of RNA. *Proceedings of the National Academy of Sciences of the United States of America* (2014).

24. Krokhotin, A., Mustoe, A. M., Weeks, K. M. & Dokholyan, N. V. Direct identification of base-paired RNA nucleotides by correlated chemical probing. *RNA* **23,** 6–13 (Jan. 2017).

25. Tomezsko, P., Swaminathan, H. & Rouskin, S. *Viral RNA structure analysis using DMS-MaPseq* 2020.

26. Ramani, V., Qiu, R. & Shendure, J. High-throughput determination of RNA structure by proximity ligation. *Nature Biotechnology* **33,** 980–984 (2015).

27. Stephenson, J. D. *et al.* Three-Dimensional RNA Structure of the Major HIV-1 Packaging Signal Region. *Structure* **21,** 951–962 (2013).

28. Rinaldi, A. J., Suddala, K. C. & Walter, N. G. *RNA-RNA Interactions* 63–95 (2015).

29. Kim, H. *et al.* Protein-guided RNA dynamics during early ribosome assembly. *Nature* **506,** 334–338 (2014).

30. Lebars, I. *et al.* Exploring TAR-RNA aptamer loop-loop interaction by X-ray crystallography, UV spectroscopy and surface plasmon resonance. *Nucleic Acids Research* **36,** 7146–7156 (2008).

31. Ferré-D'Amaré, A. R., Zhou, K. & Doudna, J. A. A general module for RNA crystallization. *Journal of Molecular Biology* **279,** 621–631 (1998).

32. Chen, Y. & Pollack, L. SAXS studies of RNA: structures, dynamics, and interactions with partners. *Wiley Interdisciplinary Reviews: RNA* **7,** 512–526 (2016).

33. Jones, C. P., Cantara, W. A., Olson, E. D. & Musier-Forsyth, K. Small-angle X-ray scattering-derived structure of the HIV-1 5' UTR reveals 3D tRNA mimicry. *Proceedings of the National Academy of Sciences* **111,** 3395–3400 (2014).

34. Fang, X. *et al.* An unusual topological structure of the HIV-1 rev response element. *Cell* **155,** 594–605 (2013).

35. Keane, S. C. & Summers, M. F. NMR studies of the structure and function of the HIV-1 5'-leader. *Viruses* **8,** 17–19 (2016).

36. Cheng, Y., Grigorieff, N., Penczek, P. A. & Walz, T. A primer to single-particle cryo-electron microscopy. *Cell* **161,** 439–449 (2015).

37. Zhong, E. D., Bepler, T., Berger, B. & Davis, J. H. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nature Methods* **18,** 176–185 (Feb. 2021).

38. Punjani, A. & Fleet, D. J. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *Journal of Structural Biology* **213,** 107702 (June 2021).

39. Scheres, S. in *Methods in Enzymology* (ed Crowther, R.) 125–157 (2016).

40. Serna, M. Hands on Methods for High Resolution Cryo-Electron Microscopy Structures of Heterogeneous Macromolecular Complexes. *Frontiers in Molecular Biosciences* **6** (May 2019).

41. Zhang, K. *et al.* Structure of the 30 kDa HIV-1 RNA Dimerization Signal by a Hybrid Cryo-EM, NMR, and Molecular Dynamics Approach. *Structure* **26,** 490–498 (Mar. 2018).

42. Zhang, K. *et al.* Cryo-EM structure of a 40 kDa SAM-IV riboswitch RNA at 3.7 Å resolution. *Nature Communications* **10,** 1–6 (Dec. 2019).

43. Zhang, K. *et al.* Cryo-EM and antisense targeting of the 28-kDa frameshift stimulation element from the SARS-CoV-2 RNA genome. *Nature Structural & Molecular Biology* **28,** 747–754 (Aug. 2021).

44. Nikolay, R. *et al.* Structural Visualization of the Formation and Activation of the 50S Ribosomal Subunit during In Vitro Reconstitution. *Molecular Cell* **70,** 881–893 (2018).

45. Bai, R., Wan, R., Yan, C., Lei, J. & Shi, Y. Structures of the fully assembled Saccharomyces cerevisiae spliceosome before activation. *Science* **360,** 1423–1429 (2018).

46. Holliday, R. A mechanism for gene conversion in fungi. *Genetics Research* **5,** 282–304 (1964).

47. Winfree, E., Liu, F., Wenzler, L. A. & Seeman, N. C. Design and self-assembly of two-dimensional DNA crystals. *Nature* **394,** 539–544 (Aug. 1998).

48. Yan, H., Park, S. H., Finkelstein, G., Reif, J. H. & LaBean, T. H. DNA-Templated Self-Assembly of Protein Arrays and Highly Conductive Nanowires. *Science* **301,** 1882–1884 (Sept. 2003).

49. Rothemund, P. W. Folding DNA to create nanoscale shapes and patterns. *Nature* **440,** 297–302 (2006).

50. Du, S. M., Zhang, S. & Seeman, N. C. DNA junctions, antijunctions, and mesojunctions. *Biochemistry* **31,** 10955–10963 (Nov. 1992).

51. Wamhoff, E.-C. *et al.* Programming Structured DNA Assemblies to Probe Biophysical Processes. *Annual Review of Biophysics* (2019).

52. Dey, S. *et al.* DNA origami. *Nature Reviews Methods Primers* **1,** 13 (Jan. 2021).

53. Castro, C. E. *et al.* A primer to scaffolded DNA origami. *Nature Methods* **8,** 221–229 (Mar. 2011).

54. Jun, H., Wang, X., Bricker, W. P. & Bathe, M. Automated sequence design of 2D wireframe DNA origami with honeycomb edges. *Nature Communications* **10,** 5419 (Nov. 2019).

55. Jun, H. *et al.* Automated Sequence Design of 3D Polyhedral Wireframe DNA Origami with Honeycomb Edges. *ACS Nano,* doi: 10.1021/acsnano.8b08671 (2019).

56. Jun, H. *et al.* Autonomously designed free-form 2D DNA origami. *Sci Adv* **5,** eaav0655 (2019).

57. Veneziano, R. *et al.* Designer nanoscale DNA assemblies programmed from the top down. *Science* **352,** 1534 (2016).

58. Benson, E. *et al.* Computer-Aided Production of Scaffolded DNA Nanostructures from Flat Sheet Meshes. *Angew Chem Int Ed Engl* **55,** 8869–8872 (2016).

59. Veneziano, R. *et al.* Role of nanoscale antigen organization on B-cell activation probed using DNA origami. *Nature Nanotechnology* **15,** 716–723 (Aug. 2020).

60. Sakai, Y. *et al.* DNA Aptamers for the Functionalisation of DNA Origami Nanostructures. *Genes (Basel)* **9** (2018).

61. Zhang, Q. *et al.* DNA Origami as an In Vivo Drug Delivery Vehicle for Cancer Therapy. *ACS Nano* **8,** 6633–6643 (2014).

62. Wang, Q., Ma, Y., Lu, Z., Yu, H. & Li, Z. Co-delivery of Chemotherapeutic Drugs and Immune Adjuvants by Nanoscale DNA Tetrahedrons for Synergistic Cancer Therapy. *ACS Applied Nano Materials* **5,** 101–106 (Jan. 2022).

63. Weiden, J. & Bastings, M. M. DNA origami nanostructures for controlled therapeutic drug delivery. *Current Opinion in Colloid & Interface Science* **52,** 101411 (Apr. 2021).

64. Kahn, J. S. & Gang, O. Designer Nanomaterials through Programmable Assembly. *Angewandte Chemie International Edition* **61** (June 2021).

65. Tian, Y. *et al.* Lattice engineering through nanoparticle–DNA frameworks. *Nature Materials* **15,** 654–661 (June 2016).

66. Wang, S.-T. *et al.* Designed and biologically active protein lattices. *Nature Communications* **12,** 3702 (June 2021).

67. Tian, Y. *et al.* Prescribed nanoparticle cluster architectures and low-dimensional arrays built using octahedral DNA origami frames. *Nature Nanotechnology* **10,** 637–644 (July 2015).

68. Shen, B. *et al.* Plasmonic nanostructures through DNA-assisted lithography. *Science Advances* **4** (Feb. 2018).

69. Gopinath, A., Miyazono, E., Faraon, A. & Rothemund, P. W. K. Engineering and mapping nanocavity emission via precision placement of DNA origami. *Nature* **535,** 401–405 (July 2016).

70. Gallego, I. *et al.* DNA-Origami-Driven Lithography for Patterning on Gold Surfaces with Sub-10 nm Resolution. *Adv Mater* **29** (2017).

71. Yang, P. *et al.* Stimuli-Responsive Three-Dimensional DNA Nanomachines Engineered by Controlling Dynamic Interactions at Biomolecule-Nanoparticle Interfaces. *ACS Nano* **15,** 16870–16877 (Oct. 2021).

72. Endo, M. & Sugiyama, H. DNA Origami Nanomachines. *Molecules* **23,** 1766 (July 2018).

73. Timm, C. & Niemeyer, C. M. Assembly and purification of enzyme-functionalized DNA origami structures. *Angewandte Chemie (International ed. in English)* **54** (June 2015).

74. Kankia, B. I. & Marky, L. A. DNA, RNA, and DNA/RNA Oligomer Duplexes: A Comparative Study of Their Stability, Heat, Hydration, and Mg ²⁺ Binding Properties. *The Journal of Physical Chemistry B* **103,** 8759–8767 (1999).

75. Severcan, I. *et al.* A polyhedron made of tRNAs. *Nature Chemistry* **2,** 772–779 (2010).

76. Severcan, I., Geary, C., Verzemnieks, E., Chworos, A. & Jaeger, L. Square-shaped RNA particles from different RNA folds. *Nano Lett* **9,** 1270–1277 (2009).

77. Geary, C., Rothemund, P. W. & Andersen, E. S. A single-stranded architecture for cotranscriptional folding of RNA nanostructures. *Science* (2014).

78. Wang, P., Ko, S. H., Tian, C., Hao, C. & Mao, C. RNA-DNA hybrid origami: folding of a long RNA single strand into complex nanostructures using short DNA helper strands. *Chem Commun (Camb)* **49,** 5462–5464 (2013).

79. Endo, M., Takeuchi, Y., Emura, T., Hidaka, K. & Sugiyama, H. Preparation of chemically modified RNA origami nanostructures. *Chemistry* **20,** 15330–15333 (2014).

80. Geary, C. W. & Andersen, E. S. *Design Principles for Single-Stranded RNA Origami Structures* in *DNA Computing and Molecular Programming: International Conference, DNA 20, Kyoto, Japan, September 22-26, 2014.* (eds Murata, S. & Kobayashi, S.) (Springer US, 2014), 1–19.

81. Veneziano, R., Shepherd, T. R., Ratanalert, S., Bellou, L. & Tao, C. In vitro synthesis of gene-length single-stranded DNA, 1–7 (2018).

82. Palazzo, A. F. & Lee, E. S. Non-coding RNA: what is functional and what is junk? *Frontiers in Genetics* **6** (Jan. 2015).

83. Ohno, H., Akamine, S. & Saito, H. RNA nanostructures and scaffolds for biotechnology applications. *Curr Opin Biotechnol* **58,** 53–61 (2018).

84. Shibata, T. *et al.* Protein-driven RNA nanostructured devices that function in vitro and control mammalian cell fate. *Nature Communications* **8,** 540 (Sept. 2017).

85. Afonin, K. A. *et al.* Triggering of RNA Interference with RNA–RNA, RNA–DNA, and DNA–RNA Nanoparticles. *ACS Nano* **9** (Jan. 2015).

86. Krissanaprasit, A. *et al.* Multivalent Aptamer-Functionalized Single-Strand RNA Origami as Effective, Target-Specific Anticoagulants with Corresponding Reversal Agents. *Advanced Healthcare Materials* **10,** e2001826 (June 2021).

87. Delebecque, C. J., Lindner, A. B., Silver, P. A. & Aldaye, F. A. Organization of Intracellular Reactions with Rationally Designed RNA Assemblies. *Science* **333,** 470–474 (July 2011).

88. Chen, J. & Seeman, N. C. Synthesis from DNA of a molecule with the connectivity of a cube. *Nature* **350,** 631–633 (Apr. 1991).

89. Mao, C., Sun, W. & Seeman, N. C. Designed Two-Dimensional DNA Holliday Junction Arrays Visualized by Atomic Force Microscopy. *Journal of the American Chemical Society* **121,** 5437–5443 (June 1999).

90. Seeman, N. C. DNA in a material world. *Nature* **421,** 427–431 (Jan. 2003).

91. Dietz, H., Douglas, S. M. & Shih, W. M. Folding DNA into twisted and curved nanoscale shapes. *Science* **325,** 725–730 (2009).

92. Han, D. *et al.* DNA Origami with Complex Curvatures in Three-Dimensional Space. *Science* **332,** 342–346 (Apr. 2011).

93. Andersen, E. S. *et al.* Self-assembly of a nanoscale DNA box with a controllable lid. *Nature* **459,** 73–76 (May 2009).

94. Douglas, S. M. *et al.* Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459,** 414–418 (2009).

95. Han, D. *et al.* DNA Gridiron Nanostructures Based on Four-Arm Junctions. *Science* **339,** 1412–1415 (Mar. 2013).

96. Iinuma, R. *et al.* Polyhedra Self-Assembled from DNA Tripods and Characterized with 3D DNA-PAINT. *Science* **344,** 65–69 (Apr. 2014).

97. Zhang, F. *et al.* Complex wireframe DNA origami nanostructures with multi-arm junction vertices. *Nat Nanotechnol* **10,** 779–784 (2015).

98. Liu, W. *et al.* Diamond family of nanoparticle superlattices. *Science* **351,** 582–586 (Feb. 2016).

99. Matthies, M., Agarwal, N. P. & Schmidt, T. L. Design and Synthesis of Triangulated DNA Origami Trusses. *Nano Letters* **16,** 2108–2113 (Mar. 2016).

100. Seeman, N. C. & Sleiman, H. F. DNA nanotechnology. *Nature Reviews Materials* **3,** 17068 (Jan. 2018).

101. Ke, Y., Castro, C. & Choi, J. H. Structural DNA Nanotechnology: Artificial Nanostructures for Biomedical Research. *Annual Review of Biomedical Engineering* **20,** 375–401 (June 2018).

102. Hu, Q., Li, H., Wang, L., Gu, H. & Fan, C. DNA Nanotechnology-Enabled Drug Delivery Systems. *Chemical Reviews* **119,** 6459–6506 (May 2019).

103. Bathe, M. & Rothemund, P. W. DNA Nanotechnology: A foundation for Programmable Nanoscale Materials. *MRS Bulletin* **42,** 882–888 (Dec. 2017).

104. Jun, H. *et al.* Rapid prototyping of arbitrary 2D and 3D wireframe DNA origami. *Nucleic Acids Research* **49,** 10265–10274 (Oct. 2021).

105. Douglas, S. M. *et al.* Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nucleic Acids Res* **37,** 5001–5006 (2009).

106. Abràmoff, M. D., Magalhães, P. J. & Ram, S. J. Image Processing with ImageJ. *Biophotonics International* **11,** 36–42 (2004).

107. Wang, H. & Seeman, N. C. Structural Domains of DNA Mesojunctions. *Biochemistry* **34,** 920–929 (Jan. 1995).

108. Harrington, R. E. Studies of DNA bending and flexibility using gel electrophoresis. *Electrophoresis* **14,** 732–746 (1993).

109. Hagerman, P. J. Sequence-directed curvature of DNA. *Annual Review of Biochemistry* **59,** 755–781 (June 1990).

110. Livshits, M. A., Amosova, O. A. & Lyubchenko, Y. L. Flexibility Difference Between Double-Stranded RNA and DNA as Revealed by Gel Electrophoresis. *Journal of Biomolecular Structure and Dynamics* **7,** 1237–1249 (June 1990).

111. Calladine, C., Collis, C., Drew, H. & Mott, M. A study of electrophoretic mobility of DNA in agarose and polyacrylamide gels. *Journal of Molecular Biology* **221,** 981–1005 (Oct. 1991).

112. Pettersen, E. F. *et al.* UCSF Chimera–a visualization system for exploratory research and analysis. *J Comput Chem* **25,** 1605–1612 (2004).

113. Ouldridge, T. E., Louis, A. A. & Doye, J. P. K. Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *The Journal of Chemical Physics* **134,** 085101 (Feb. 2011).

114. Snodin, B. E. K. *et al.* Introducing improved structural properties and salt dependence into a coarse-grained model of DNA. *The Journal of Chemical Physics* **142,** 234901 (June 2015).

115. Shepherd, T. R., Du, R., Huang, H., Wamhoff, E.-C. & Bathe, M. Bioproduction of pure, kilobase-scale single-stranded DNA. *Sci Rep* **9,** 6121 (2019).

116. Zheng, S. Q. *et al. MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy* Feb. 2017.

117. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol* **157,** 38–46 (2007).

118. Tukalo, M., Yaremchuk, A., Fukunaga, R., Yokoyama, S. & Cusack, S. The crystal structure of leucyl-tRNA synthetase complexed with tRNALeu in the post-transfer–editing conformation. *Nature Structural & Molecular Biology* **12,** 923–930 (Sept. 2005).

119. Agirrezabala, X. & Valle, M. Structural Insights into tRNA Dynamics on the Ribosome. *International Journal of Molecular Sciences* **16,** 9866–9895 (Apr. 2015).

120. Korostelev, A., Trakhanov, S., Laurberg, M. & Noller, H. F. Crystal Structure of a 70S Ribosome-tRNA Complex Reveals Functional Interactions and Rearrangements. *Cell* **126,** 1065–1077 (Sept. 2006).

121. Fukunaga, R., Yokoyama, S. & RIKEN Structural Genomics/Proteomics Initiative (RSGI). *The crystal structure of Leucyl-tRNA synthetase and tRNA(leucine) complex* 2005.

122. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms for Molecular Biology* **6,** 26 (2011).

123. Reuter, J. S. & Mathews, D. H. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11,** 129 (Mar. 2010).

124. Jarmoskaite, I., AlSadhan, I., Vaidyanathan, P. P. & Herschlag, D. How to measure and evaluate binding affinities. *eLife* **9** (Aug. 2020).

125. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9,** 676–682 (June 2012).

126. Douglas, S. M., Bachelet, I. & Church, G. M. A logic-gated nanorobot for targeted transport of molecular payloads. *Science* **335,** 831–834 (2012).

127. Li, S. *et al.* A DNA nanorobot functions as a cancer therapeutic in response to a molecular trigger in vivo. *Nat Biotechnol* **36,** 258–264 (2018).

128. Benson, E. *et al.* DNA rendering of polyhedral meshes at the nanoscale. *Nature* **523,** 441–444 (2015).

129. He, Y. *et al.* Hierarchical self-assembly of DNA into symmetric supramolecular polyhedra. *Nature* **452,** 198–201 (2008).

130. Nummelin, S. *et al.* Robotic DNA Nanostructures. *ACS synthetic biology* **9,** 1923–1940 (Aug. 2020).

131. Afonin, K. A. *et al.* In vitro assembly of cubic RNA-based scaffolds designed in silico. *Nat Nanotechnol* **5,** 676–682 (2010).

132. Afonin, K. A. *et al.* Co-transcriptional assembly of chemically modified RNA nanoparticles functionalized with siRNAs. *Nano Lett* **12,** 5192–5195 (2012).

133. Han, D. *et al.* Single-stranded DNA and RNA origami. *Science* **358** (2017).

134. Hoiberg, H. C., Sparvath, S. M., Andersen, V. L., Kjems, J. & Andersen, E. S. An RNA Origami Octahedron with Intrinsic siRNAs for Potent Gene Knockdown. *Biotechnol J* **14,** e1700634 (2019).

135. Kozyra, J. *et al.* Designing Uniquely Addressable Bio-orthogonal Synthetic Scaffolds for DNA and RNA Origami. *ACS Synth Biol* **6,** 1140–1149 (2017).

136. Sparvath, S. L., Geary, C. W. & Andersen, E. S. in *Methods in Molecular Biology* (2017).

137. Ke, W. *et al.* RNA-DNA fibers and polygons with controlled immunorecognition activate RNAi, FRET and transcriptional regulation of NF-kappaB in human cells. *Nucleic Acids Res* **47,** 1350–1361 (2019).

138. Monferrer, A., Zhang, D., Lushnikov, A. J. & Hermann, T. Versatile kit of robust nanoshapes self-assembling from RNA and DNA modules. *Nat Commun* **10,** 608 (2019).

139. Mohammed, A., Orponen, P., Pai, S., Stepney, S. & Verlan, S. *Algorithmic Design of Cotranscriptionally Folding 2D RNA Origami Structures* in (Springer International Publishing, Cham, 2018), 159–172.

140. Geary, C., Grossi, G., McRae, E. K. S., Rothemund, P. W. K. & Andersen, E. S. RNA origami design tools enable cotranscriptional folding of kilobase-sized nanoscaffolds. *Nature Chemistry* **13,** 549–558 (May 2021).

141. Hong, E. *et al.* Structure and Composition Define Immunorecognition of Nucleic Acid Nanoparticles. *Nano Lett* **18,** 4309–4321 (2018).

142. Johnson, M. B. *et al.* Programmable Nucleic Acid Based Polygons with Controlled Neuroimmunomodulatory Properties for Predictive QSAR Modeling. *Small* **13** (2017).

143. Rich, A. Discovery of the hybrid helix and the first DNA-RNA hybridization. *J Biol Chem* **281,** 7693–7696 (2006).

144. Rich, A. A Hybrid Helix Containing Both Deoxyribose and Ribose Polynucleotides and Its Relation to the Transfer of Information between the Nucleic Acids. *Proc Natl Acad Sci U S A* **46,** 1044–1053 (1960).

145. Miao, Z. & Westhof, E. RNA Structure: Advances and Assessment of 3D Structure Prediction. *Annual review of biophysics* **46** (May 2017).

146. Shapiro, B. A. & Le Grice, S. F. J. Advances in RNA structure determination. *Methods* **103** (July 2016).

147. Shu, D., Shu, Y., Haque, F., Abdelmawla, S. & Guo, P. Thermodynamically stable RNA three-way junction for constructing multifunctional nanoparticles for delivery of therapeutics. *Nature Nanotechnology* **6,** 658–667 (Sept. 2011).

148. Jasinski, D., Haque, F., Binzel, D. W. & Guo, P. Advancement of the Emerging Field of RNA Nanotechnology. *ACS Nano* **11,** 1142–1164 (2017).

149. Ko, S. H. *et al.* Synergistic self-assembly of RNA and DNA molecules. *Nature Chemistry* **2,** 1050–1055 (Dec. 2010).

150. Afonin, K. A. *et al.* Multifunctional RNA nanoparticles. *Nano letters* **14,** 5662–5671 (Oct. 2014).

151. Kube, M. *et al.* Revealing the structures of megadalton-scale DNA complexes with nucleotide resolution. *Nature Communications* **11,** 6229 (Dec. 2020).

152. Schneider, F., Möritz, N. & Dietz, H. The sequence of events during folding of a DNA origami. *Science Advances* **5** (May 2019).

153. Gerling, T., Kube, M., Kick, B. & Dietz, H. Sequence-programmable covalent bonding of designed DNA assemblies. *Science Advances* **4** (Aug. 2018).

154. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science* **30,** 70–82 (Jan. 2021).

155. Liebl, K., Drsata, T., Lankas, F., Lipfert, J. & Zacharias, M. Explaining the striking difference in twist-stretch coupling between DNA and RNA: A comparative molecular dynamics analysis. *Nucleic Acids Research* **43,** 10143–10156 (Dec. 2015).

156. Naskar, S., Joshi, H., Chakraborty, B., Seeman, N. C. & Maiti, P. K. Atomic structures of RNA nanotubes and their comparison with DNA nanotubes. *Nanoscale* **11,** 14863–14878 (2019).

157. Morcos, P. A. Achieving targeted and quantifiable alteration of mRNA splicing with Morpholino oligos. *Biochem Biophys Res Commun* **358,** 521–527 (2007).

158. Rinaldi, C. & Wood, M. J. A. Antisense oligonucleotides: the next frontier for treatment of neurological disorders. *Nat Rev Neurol* **14,** 9–21 (2018).

159. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460,** 894–898 (2009).

160. Song, F. & Stieger, K. Optimizing the DNA Donor Template for Homology-Directed Repair of Double-Strand Breaks. *Mol Ther Nucleic Acids* **7,** 53–60 (2017).

161. Walter, H. K. *et al.* "DNA Origami Traffic Lights" with a Split Aptamer Sensor for a Bicolor Fluorescence Readout. *Nano Lett* **17,** 2467–2472 (2017).

162. Punekar, A. S., Liljeruhm, J., Shepherd, T. R., Forster, A. C. & Selmer, M. Structural and functional insights into the molecular mechanism of rRNA m6A methyltransferase RlmJ. *Nucleic Acids Res* **41,** 9537–9548 (2013).

163. Punekar, A. S., Shepherd, T. R., Liljeruhm, J., Forster, A. C. & Selmer, M. Crystal structure of RlmM, the 2'O-ribose methyltransferase for C2498 of Escherichia coli 23S rRNA. *Nucleic Acids Res* **40,** 10507–10520 (2012).

164. Neumann, H., Wang, K., Davis, L., Garcia-Alai, M. & Chin, J. W. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **464,** 441–444 (2010).

165. Orelle, C. *et al.* Protein synthesis by ribosomes with tethered subunits. *Nature* **524,** 119–124 (2015).

166. Shepherd, T. R. *et al.* De novo design and synthesis of a 30-cistron translation-factor module. *Nucleic Acids Res* **45,** 10895–10905 (2017).

167. Lu, X. J. & Olson, W. K. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat Protoc* **3,** 1213–1227 (2008).

168. Prim, R. C. Shortest connection networks and some generalizations. *Bell System Technical Journal* **36,** 1389–1401 (1957).

169. Weitzmann, C. J., Cunningham, P. R. & Ofengand, J. Cloning, in vitro transcription, and biological activity of Escherichia coli 23S ribosomal RNA. *Nucleic Acids Res* **18,** 3515–3520 (1990).

170. Lundquist, K. P., Panchal, V., Gotfredsen, C. H., Brenk, R. & Clausen, M. H. Fragment-Based Drug Discovery for RNA Targets. *ChemMedChem* **16,** 2588–2603 (June 2021).

171. Suresh, B. M. *et al.* A general fragment-based approach to identify and optimize bioactive ligands targeting RNA. *Proceedings of the National Academy of Sciences* **117,** 33197–33203 (Dec. 2020).

172. Kozlovskii, I. & Popov, P. Structure-based deep learning for binding site detection in nucleic acid macromolecules. *NAR Genomics and Bioinformatics* **3** (Nov. 2021).

173. Warner, K. D., Hajdin, C. E. & Weeks, K. M. Principles for targeting RNA with drug-like small molecules. *Nature Reviews Drug Discovery* **17,** 547–558 (July 2018).

174. Razi, A., Britton, R. A. & Ortega, J. The impact of recent improvements in cryo-electron microscopy technology on the understanding of bacterial ribosome assembly. *Nucleic Acids Research* **45,** 1027–1040 (2017).

175. Brown, A. & Shao, S. Ribosomes and cryo-EM: a duet. *Current Opinion in Structural Biology* **52,** 1–7 (2018).

176. Fernandes, J., Jayaraman, B. & Frankel, A. The HIV-1 Rev response element. *RNA Biology* **9,** 6–11 (Jan. 2012).

177. Bai, Y., Tambe, A., Zhou, K. & Doudna, J. A. RNA-guided assembly of Rev-RRE nuclear export complexes. *eLife* **3,** e03656 (Aug. 2014).

178. Daugherty, M. D., Booth, D. S., Jayaraman, B., Cheng, Y. & Frankel, A. D. HIV Rev response element (RRE) directs assembly of the Rev homooligomer into discrete asymmetric complexes. *Proceedings of the National Academy of Sciences* **107,** 12481–12486 (July 2010).

179. Sherpa, C., Rausch, J. W., Le Grice, S. F., Hammarskjold, M. L. & Rekosh, D. The HIV-1 Rev response element (RRE) adopts alternative conformations that promote different rates of virus replication. *Nucleic Acids Research* **43,** 4676–4686 (Mar. 2015).

180. Tomezsko, P. J. *et al.* Determination of RNA structural diversity and its role in HIV-1 RNA splicing. *Nature* (2020).

181. Scheres, S. H. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology* **180,** 519–530 (Dec. 2012).

182. Denny, S. K. *et al.* High-Throughput Investigation of Diverse Junction Elements in RNA Tertiary Folding. *Cell* **174,** 377–390 (2018).

183. Gracia, B. *et al.* RNA Structural Modules Control the Rate and Pathway of RNA Folding and Assembly. *Journal of Molecular Biology* **428,** 3972–3985 (2016).

184. Gracia, B. *et al.* Hidden Structural Modules in a Cooperative RNA Folding Transition. *Cell Reports* **22,** 3240–3250 (2018).

185. Hu, W. S. & Hughes, S. H. HIV-1 reverse transcription. *Cold Spring Harb Perspect Med* **2** (2012).

186. Liolios, L. *et al.* Comparison of a Multiplex Reverse Transcription-PCR-Enzyme Hybridization Assay with Conventional Viral Culture and Immunofluorescence Techniques for the Detection of Seven Viral Respiratory Pathogens. *Journal of Clinical Microbiology* **39,** 2779–2783 (Aug. 2001).

187. Medioni, C. & Besse, F. in *RNA Detection* (ed Gaspar, I.) 1–28 (Humana Press, New York, NY, 2018).

188. Shinoda, H. *et al.* Amplification-free RNA detection with CRISPR–Cas13. *Communications Biology* **4,** 476 (Apr. 2021).

189. Smyth, R. P., Negroni, M., Lever, A. M., Mak, J. & Kenyon, J. C. RNA Structure—A Neglected Puppet Master for the Evolution of Virus and Host Immunity. *Frontiers in Immunology* **9** (Sept. 2018).

190. Walter, N. G. & Engelke, D. R. Ribozymes: catalytic RNAs that cut things, make things, and do odd and useful jobs. *Biologist (London, England)* **49,** 199–203 (Oct. 2002).

191. Dill, K. A., Bromberg, S. & Stigter, D. *Molecular Driving Forces: Statistical Thermodynamics in Biology, Chemistry, Physics and Nanoscience* 2nd (Garland Science, New York, Oct. 2010).

192. Borkovec, M., Hamáček, J. & Piguet, C. Statistical mechanical approach to competitive binding of metal ions to multi-center receptors. *Dalton Transactions,* 4096–4105 (Nov. 2004).

193. Wu, P., Nakano, S.-i. & Sugimoto, N. Temperature dependence of thermodynamic properties for DNA/DNA and RNA/DNA duplex formation. *European Journal of Biochemistry* **269,** 2821–2830 (June 2002).