Nucleic Acid Nanostructures as Tags for Cryo-Electron Microscopy

Emma Silvester
St. Hugh's College
University of Oxford



Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Oxford

Trinity Term, 2019

Abstract

Cryo-electron microscopy (cryo-EM) is an indispensable tool for structural biologists, allowing the reconstruction of biomolecules in their native, hydrated state to nanometre, or sub-nanometre, resolution. Unfortunately *in situ* cryo-EM characterisation is currently limited by our ability to interpret micrographs and tomograms, as a low signal-to-noise ratio means different structures are usually difficult to distinguish. To overcome this limitation, microscopists often use cryo-EM tags: electron-dense, nano-scale objects with specific targeting properties. This thesis presents two related projects involving the use of nucleic acid nanostructures as cryo-EM tags.

In Project 1, I introduce functionalised DNA-origami nanostructures as tags. I begin with the cryo-EM characterisation of various nanostructures in cell lysate to assess their imaging properties in a biological context. I find that solid lattice-based nanostructures provide high contrast in this environment. By cryo-EM, the outline of nanostructures is often difficult to visualise, whereas the internal arrangement of helices results in a distinctive 'stripy' effect. As a result, honeycomb-lattice origami nanostructures with different shapes can be difficult to distinguish, whereas nanostructures of different lattice types are easily identifiable. Based on this preliminary work, I present a signpost origami nanostructures for tagging, comprising a high-contrast, asymmetric 'sign' for identification, and a six-helix bundle 'post' for targeting. To functionalise nanostructures for targeting, I have characterised an RNA aptamer which binds green fluorescent protein (GFP) and quantified its binding affinity to GFP analogues. The aptamer is attached to a signpost nanostructure by hybridisation and *in situ* binding is demonstrated with a murine leukaemia virus test system. High-contrast, high-efficiency labelling of viral glycoproteins is observed. Electroporation of polyamine-stabilised nanostructures into *E. coli* is also shown for intracellular labelling applications.

In Project 2, I present methods for designing 3D RNA nanostructures for co-transcriptional folding. This provides the groundwork for the development of genetically-encoded cryo-EM tags. I firstly introduce a grid-like architecture, where an RNA nanostructure consists of a set of interacting units arranged on a square lattice. Each unit is a branched hairpin structure, with units connected by a path of single-stranded linkers and through kissing interactions between complementary hairpin arms. For this design, six different nanostructures are synthesised. Supermolecular assembly of nanostructures into finite super-structures is demonstrated, as is the incorporation of an aptamer sequence for GFP binding. Finally, I investigate an origami-like route toward RNA nanostructure design, by introducing two new assembly motifs for incorporation into an already-established 2D RNA origami architecture in an effort to build in 3D. My first origami-like design uses paranemic crossovers formed between bulge loops situated on adjacent sheets. I show that these nanostructures do not fold into the intended 3D structure and suggest explanations for this observation. In my second origami-like design, I introduce the combined bulge-loop/kissing-loop motif. Two-sheet nanostructures, held together through these motifs, have been designed. Although a more promising prospect than the previous design, the yield of well-folded structures is low. Based on these observations, the grid-like architecture will provide a more promising launching pad for the development of genetic tags.

Acknowledgements

First and foremost, thank you to my supervisor, Andrew Turberfield, for offering me the opportunity to undertake a DPhil in his group in the first place, for his continued guidance during my time here, and for his wealth of knowledge and new ideas. Andrew has allowed me a good measure of independence over the past three years, but is always available for input when it is needed. Equal thanks go to Jon Bath, who has been a great source of literature knowledge and laboratory expertise. Jon has an unmatched ability in troubleshooting experiments and suggesting creative solutions to problems, and has provided a great deal of insight and practical advice throughout my DPhil.

Thank you to other members of the Turberfield group, past and present: Alma Chapet–Batlle, Antonio Garcia Guerra, Behnam Najafi, Céline Journot, Erik Benson, Ibon Santiago, Joel Spratt, Joakim Bohlin, Juan Jin, Katie Young, Natalie Haley, Rafael Carrascosa Marzo, Robbie Oppenheimer, Sam Tusk, Seham Helmi and Wenjing Meng have all helped maintain a friendly atmosphere in the group. A special thanks to Sanna Piippo-Henderson keeps the lab well-stocked and running smoothly.

I have also had the pleasure of collaborating with Lindsay Baker. Without Lindsay, and her exhaustive cryo-EM experience, much of this work would have been impossible. Thanks also to other members of the Grünewald group, in particular Ben Vollmer, Cathy Whittle, Emily Machala, Vojta Pražák and Kay himself for their practical contributions to my projects, and for making STRUBI such an enjoyable place to work.

I have spent many hours in the Dunn School basement, where Errin Johnson taught me to use the transmission electron microscope and has provided so much indispensable troubleshooting and advice. Likewise, thanks to David Staunton who runs the biophysics suite in Oxford, and taught me to use the ITC and SPR facilities.

Finally, thanks to my family and my partner, Ryan, for their support over the last few years, and especially over the stressful final months.

Contents

1 Introduction **1**
- 1.1 Scope of Thesis . . . 1
- 1.2 Cryo-Electron Microscopy . . . 2
  - 1.2.1 Why Cryo-EM? . . . 2
  - 1.2.2 Cryo-Electron Tomography . . . 3
  - 1.2.3 Labelling in Cryo-EM . . . 4
- 1.3 Structural DNA Nanotechnology . . . 9
  - 1.3.1 DNA . . . 9
  - 1.3.2 Early DNA Nanostructures . . . 11
  - 1.3.3 DNA Origami . . . 12
- 1.4 Structural RNA Nanotechnology . . . 16
  - 1.4.1 RNA . . . 16
  - 1.4.2 RNA folding: a biological perspective . . . 18
  - 1.4.3 Structural RNA Nanotechnology . . . 19
- 1.5 Summary and Thesis Structure . . . 25

2 DNA Origami Nanostructures for Cryo-electron Microscopy **27**
- 2.1 Author Contributions . . . 27
- 2.2 Project 1 Aim: A DNA Origami Tag for CET . . . 27
- 2.3 Pilot Work with an Origami Wedge . . . 28
  - 2.3.1 Design . . . 28
  - 2.3.2 Synthesis and Characterisation . . . 29
  - 2.3.3 Purifying Origami for Cryo-EM . . . 33
  - 2.3.4 Cryo-EM in Cell Lysate . . . 35
- 2.4 Designing Nanostructures for Multiplexing . . . 36
  - 2.4.1 Hollow Origami Nanostructures . . . 36
  - 2.4.2 Multiplexing Honeycomb Origami Nanostructures . . . 40
- 2.5 Signpost Origami: Design, Synthesis and Characterisation . . . 46
  - 2.5.1 Design . . . 46
  - 2.5.2 Synthesis and Characterisation . . . 48
- 2.6 Signpost Origami: Optimisation . . . 49
  - 2.6.1 Heavy-Metal Pre-Soaking . . . 49
  - 2.6.2 Comparing Lattice Types . . . 52
- 2.7 Summary and Future Directions . . . 56

3 Targeting and Delivering Origami Tags **61**
3.1 Author Contributions . . . 61
3.2 A GFP Aptamer as a Targeting Module . . . 61
3.2.1 GFP Aptamer . . . 61
3.2.2 Conjugation to Origami . . . 63
3.2.3 Quantifying Binding to New GFP Analogues . . . 65
3.3 Moloney murine leukaemia virus: A test system for tagging . . . 68
3.3.1 Murine Leukaemia Virus . . . 68
3.3.2 Aptamer Binding to Env-YFP . . . 68
3.3.3 Negative Stain TEM of Virus Binding . . . 70
3.3.4 Cryo-ET of Virus Binding . . . 70
3.3.5 Cryo-EM of Cell Binding . . . 74
3.4 Electroporation of DNA Origami into Cells . . . 74
3.4.1 Motivation and Background . . . 74
3.4.2 Stabilising Nanostructures with Polyamines . . . 76
3.4.3 Fluorescence Microscopy of Electroporated Cells . . . 78
3.4.4 Discussion . . . 78
3.5 Summary and Future Directions . . . 80

4 RNA Nanostructures: Grid-based Approach **83**
4.1 Author Contributions . . . 83
4.2 Project 2 Aim: 3D RNA Nanostructures by Co-transcriptional Folding . . . 83
4.3 Preliminary Design: A Pentomino Nanostructure . . . 84
4.3.1 Design . . . 84
4.3.2 Synthesis and Characterisation . . . 91
4.3.3 Discussion . . . 95
4.4 Modified Design for Assembly . . . 98
4.4.1 Design . . . 98
4.4.2 Transcription and Characterisation . . . 104
4.4.3 Discussion . . . 110
4.5 Cross-Shaped Assembly . . . 110
4.5.1 Design . . . 110
4.5.2 Synthesis and Characterisation . . . 111
4.5.3 Discussion . . . 121
4.6 Summary . . . 123

5 RNA Nanostructures: Origami-based Approach **125**
5.1 Author Contributions . . . 125
5.2 Extending 2D Origami into the Third Dimension . . . 125
5.2.1 2D Origami . . . 125
5.2.2 Options for building in 3D . . . 127
5.3 3D Origami: Building with PX motifs . . . 133
5.3.1 Design . . . 133
5.3.2 Synthesis and Characterisation . . . 139
5.3.3 Discussion . . . 142

5.4 3D Origami: Building with BLKL motifs . . . 143
5.4.1 Concept . . . 143
5.4.2 2-Sheet Dimer . . . 144
5.4.3 2-sheet Nanostructure . . . 150
5.5 Summary . . . 155

6 Conclusion **163**

A Sequences **167**
A.1 gBlock sequences . . . 167
A.1.1 Grid-based RNA Nanostructures . . . 167
A.1.2 Origami-based RNA Nanostructures . . . 173
A.2 Oligonucleotide Transcription Templates . . . 177
A.2.1 Aptamer Sequences . . . 177
A.2.2 PX_{major} . . . 178
A.3 Origami Sequences . . . 178
A.3.1 Wedge Origami . . . 179
A.3.2 Pointing Hand Origami . . . 187
A.3.3 Honeycomb Lattice Sign-post Origami . . . 194
A.3.4 Square Lattice Sign-post Origami . . . 203

B RNA Methods **211**
B.1 Template Preparation . . . 211
B.1.1 Short Templates . . . 211
B.1.2 Cloning . . . 211
B.1.3 PCR . . . 213
B.2 Transcription Reactions . . . 216
B.2.1 Short RNA Transcription . . . 216
B.2.2 Nanostructure Transcription . . . 216
B.3 Purification . . . 217
B.3.1 Column Purification . . . 217
B.3.2 LiCl Precipitation . . . 217
B.3.3 Ethanol Precipitation . . . 217

C Origami Methods **219**
C.1 Annealing . . . 219
C.1.1 Honeycomb Lattice Origami . . . 219
C.1.2 Square Lattice Origami . . . 220
C.2 Purification . . . 221
C.2.1 MWCO Filtration . . . 221
C.2.2 PEG Precipitation . . . 221
C.2.3 Ethanol Precipitation . . . 222
C.2.4 Column Purificiation . . . 222

D Microscopy **223**
D.1 Transmission Electron Microscopy . . . 223
D.1.1 Grid Preparation: Negative Stain . . . 223
D.1.2 Grid Preparation: Positive Stain . . . 223
D.1.3 Imaging . . . 224
D.2 Atomic Force Microscopy . . . 224
D.3 Cryo-Electron Microscopy . . . 224
D.3.1 Cell Lysate Preparation . . . 224
D.3.2 Grid Preparation . . . 225
D.3.3 Imaging . . . 226
D.4 Fluorescence Microscopy . . . 226

E Gel Electrophoresis **227**
E.1 Agarose Gel Electrophoresis . . . 227
E.2 Polyacrylamide Gel Electrophoresis . . . 227
E.3 Visualising Gels . . . 228

F Matlab Scripts **229**
F.1 Worm-like Chain . . . 229
F.2 Ligate . . . 230
F.3 Gel Peak Analysis . . . 231
F.4 Profile Analysis . . . 232

Chapter 1

Introduction

In this chapter, I will describe the scope of this thesis before introducing the relevant background literature. I will present cryo-electron microscopy as a technique for structural biologists, with a particular emphasis on methods of tagging structures for more effective localisation. I will also discuss advances in structural DNA nanotechnology, focussing on DNA origami which is my architecture of choice for the work presented in Chapters 2 and 3. Finally, I will introduce structural RNA nanotechnology, highlighting the complexities of RNA folding compared with DNA. I will present RNA nanostructure methods, focussing on those relevant to *in vivo* folding.

1.1 Scope of Thesis

In this thesis, I present new strategies for labelling in cryo electron microscopy (cryo-EM) based on nucleic acid nanostructures with inherent targeting properties. Cryo-EM has been a revolutionary technique for structural biologists, permitting the reconstruction of macromolecules in their native, hydrated, state at resolutions that are otherwise unattainable. One of the key limitations of cryo-EM is our ability to interpret data. Contrast is typically low, so structures can be challenging to identify in a noisy biological environment. Data would be far easier to analyse with the addition of a 'tag', an electron-dense nano-scale object which binds to a target molecule that we wish to locate, with a distinctive shape that can be easily recognised in a busy micrograph/tomogram.

DNA and RNA are well-suited as building materials for this application: repeating phosphate groups mean they are more electron-dense than most cellular components, and DNA/RNA strands can be programmably assembled into nanostructures with well-defined 3D architectures. There are also a number of different routes for endowing nucleic acid nanostructures with targeting properties. These include the incorporation of aptamers, the use of natural DNA/RNA binding proteins as protein fusion tags, and hybridisation to chemically modified oligonucleotides for coupling to peptide tags.

This thesis is divided into two projects. For Project 1, I will show the design, synthesis and implementation of tags built from DNA origami and functionalised with RNA aptamers.

These are used to tag protein targets located on viral and cell surfaces. DNA-origami tags have the potential to revolutionise *in situ* cryo-EM of virions and cell membrane complexes (or other systems with external protein domains such as isolated nuclei) as they provide high resolution and high contrast labelling with an unlimited capacity for multiplexing. For Project 2, I will show the design and synthesis of 3D RNA nanostructures by co-transcriptional folding. This project provides the groundwork for generating RNA-based cryo-EM tags *in vivo*, by transcription from a genetically encoded DNA template. Achieving this goal would be a significant milestone for cellular cryo-EM, with RNA-based tags providing an analogous role to fluorescent proteins in light microscopy.

1.2 Cryo-Electron Microscopy

1.2.1 Why Cryo-EM?

The structural characterisation of biological macromolecules is key to an understanding of cellular function. Elucidation of increasingly complex structures has provided many fascinating insights into the behaviour of biological systems, but limitations still exist. The structure of macromolecules and their complexes can be intimately dependent on biological context. Prevalent techniques such as X-ray crystallography and nuclear magnetic resonance require isolation of structures, and therefore may not tell the full story. Electron microscopy, however, allows structures to be imaged without extraction from their immediate biological environment.

Electron microscopy (EM) of biological specimens has made great strides since its inception in the early 20th century [1, 2]. One of the main shortfalls of traditional EM studies is the sample preparation procedure, whereby samples are dehydrated or exposed to harsh fixation conditions which can introduce structural artefacts (even resulting in long-standing misconceptions of cellular ultra-structure [3]). This is necessitated by the high vacuum created in the beam path (required to prevent scattering of the electron beam by molecules in the air), which would compromise aqueous samples. One technique which gained traction in the 1960s, and remains popular today, is negative staining. In this technique, the water surrounding a sample is replaced by a heavy-metal salt solution, whilst the sample itself remains unstained and therefore visible over the dark background [4].

Imaging biological samples in a more native, hydrated, state presented a major challenge, but in the 1980s Dubochet [5, 6, 7] (building on prior work by Unwin [8, 9] and Taylor and Glaeser [10, 11]) showed how perfectly preserved samples could be imaged by freezing them in vitreous ice. This was the birth of cryo-EM, but it took a number of technological developments, most notably direct-electron detectors and improved image processing, for the technique to become widespread [12]. These advances have improved the (now Ångstom-scale) resolution attainable by cryo-EM, often coined the ‘resolution revolution’.

1.2.2 Cryo-Electron Tomography

Electron tomography is a method for obtaining 3D structural information of a sample from electron micrographs. The framework for this technique was developed by De Rosier and Klug in 1968 [13], who described how 3D structures can be reconstructed from a 'tilt series': a set of images of the same sample taken from different angles, appropriately aligned. Their original implementation saw the 3D reconstruction of the T4 bacteriophage tail from a single view, made possible by the high rotational symmetry of this particular structure [13]. With the necessary hardware, however, the technique is general and can be used to create 3D maps of arbitrary structures.

When discussing electron tomography techniques, it is important to distinguish between single-particle averaging and 'true' tomography. The two methods are based on the same mathematical framework, in which 2D projections of a sample are used to build a 3D map. Single-particle averaging is the most commonly-used modality of cryo-EM. Samples for single-particle cryo-EM are solutions of purified molecules/molecular complexes. These molecules are, ideally, randomly orientated in solution, so different projections of the structure are derived by imaging many particles. It is important that the sample is homogenous, since conformational variability leads to uncertainty in the 3D reconstruction (although large conformational changes can be addressed with 3D classification algorithms [12]). Single-particle cryo-EM now rivals X-ray crystallography in terms of resolution, and can determine structures of molecules that are refractory to crystallisation [14].

Of course, not all structural biology questions can be addressed by reconstructing purified molecules in solution. Membrane-dwelling macromolecules must be treated with surfactants for solubilisation, thus any derived structure may not represent the molecule's native configuration. Additionally, the conformation of molecules, and interactions between molecules, can be intimately dependent on their environment. To elucidate these dependencies, it is helpful to image molecules within their biological context. This can even allow us to take a 'snapshot' of biological processes in action. For this, we look towards 'true' tomography. Samples for cryo-electron tomography (CET) include cell-derived vesicles and synthetic lipid structures, pleomorphic viruses, isolated cellular components and thin, or sectioned, cells. For 'true' tomography, the sample to be reconstructed is unique, so different 3D projections are obtained by physically tilting the sample.

The first implementations of CET were in the 1990s [15], but the development of automated tilt series acquisition was a critical step in making the methodology practical [16, 17]. Alignment of images within a tilt series is achieved by cross-correlation of successive images, thus the minimum usable dose is limited by the contrast required for successful alignment. The main concern for cryo-EM methods is limiting exposure to tolerable levels, as high electron doses compromise biological samples and can result in ice de-vitrification. These dose restrictions are far more severe than for single-particle averaging, in which the dose is distributed across many particles. CET alignment procedures use areas outside the region of interest to correct for positioning and focussing errors, which allows all of the electron dose within the region of interest to be put towards data acquisition. Alignment can be further facilitated by the addition of high contrast fiducial markers, such as gold nano-

particles, if the system allows. Tomographic reconstructions of thin (< 500 nm) specimens typically offer resolutions of a few nanometres, but if multiples copies of a particular complex can be identified, sub-tomogram averaging (in which single-particle averaging techniques are applied to CET data) can yield sub-nanometre resolution reconstructions [18].

1.2.3 Labelling in Cryo-EM

Localising Structures by Cryo-EM

Because samples in cryo-EM are unstained, imaging contrast results from natural variation in the electron-refractive index of biological samples and is therefore inherently poor. Cellular tomograms provide data with unprecedented resolution but it is often challenging to unambiguously identify structures of interest, which are typically of low contrast over a noisy background. The task of interpreting data from cellular tomograms can be arduous or even impossible. Larger complexes such as ribosomes [19] or proteasomes [20] can be located by correlation-based searching of tomograms with a template structure (based on a pre-existing reconstruction), but this is unlikely to be useful for detecting smaller structures. Thus, there is a need for a general localisation/tagging method for CET.

A localisation procedure should fulfil two functions: identifying a region of interest within a wide field of view for image acquisition; and interpreting the resulting tomogram. For example, consider using CET to study the structure of complexes formed by a specific membrane-anchored protein. Firstly, a localisation method is used to identify regions in which the target protein resides. Tomograms are then constructed within these regions, and the localisation method allows the identity of the target molecule within the 3D map to be confirmed, even its structure within this context is previously unknown.

One possible route towards tagging is fluorescence microscopy (FM). FM permits the localisation of fluorescent molecules to within a few hundred nanometres for conventional microscopy, or tens of nanometres using super-resolution techniques [21, 22]. Fluorescent proteins such as GFP (green fluorescent protein) can be used as a 'fusion tag', genetically expressed as part of the target protein. This method has generated many important insights into cellular biology [23]. In isolation, FM allows visualisation of the fluorescent tag, but not the context in which it exists. Correlative light and electron microscopy (CLEM) approaches combine the advantages of FM and EM to image macromolecules in a cellular context: FM is used to locate targets in a large field of view, and EM reveals the underlying cellular ultra-structure in the region of interest.

Combining FM with cryo-EM is no easy feat. One strategy is to perform FM prior to vitrification [24]. Whilst this permits more standard and higher resolution FM techniques, it only works for structures that are unperturbed by mild fixation or are stable over the downtime required between imaging and freezing. Performing both imaging steps post-freezing is a more promising strategy, and samples have been successfully imaged in this way [25, 26]. However, the technique is still in its infancy and widespread adoption hinges on the development of a more automated workflow. Additionally, most super-resolution methods require high laser intensities which can cause ice de-vitrification, so a compromise between

FM resolution and sample integrity must be made [27]. For the most part, CLEM is useful for identifying regions of interest, but not identifying specific molecular components within that region.

A more common localisation approach, inspired more abstractly by fluorescence microscopy, entails introducing a high-contrast tag which fuses to or binds a target. A cryo-EM/ET tag must fulfil certain requirements: it must be a suitable size and provide adequate contrast to be distinguished over a noisy cellular environment; it must have a high affinity and specificity for its target; and, for cellular experiments it should ideally be capable of assembling *in vivo*. It is often informative to detect multiple different targets within a tomogram, so a tagging mechanism should be extendible to multiple visibly distinguishable objects with orthogonal targeting properties for multiplexing experiments.

Immunogold Labelling

Currently, the prevailing technique for tagging in CET is immunogold labelling, developed originally for stain EM [28]. Here the tags are gold nano-particles or clusters conjugated to antibodies, soluble proteins or ligands. Whilst any heavy metal could provide contrast, gold is most commonly used as it is relatively inert, forms nano-particles easily and its functionalisation chemistries are well-understood and controlled in aqueous solution [29].

The permeabilisation methods used to label intracellular proteins can result in significant changes to cellular structure, which negates the near-native advantage of cryo-EM/CET over alternative techniques [30]. Nonetheless, immunogold labelling performs well for isolated cellular components. This was illustrated in 2007 with the reconstruction of nuclear pore complexes from CET imaging of intact nuclei [31]. Gold particles, conjugated to fusion proteins carrying a nuclear localisation signal, were used to track the routes taken by cargo through the pore. Likewise, immunogold labelling performs well with viral glycoproteins, or cell surface proteins, where no invasive delivery protocol is required, as demonstrated by Yi and colleagues in 2015[32] by the labelling of viruses and virus-transfected/infected cells. Examples of tomogram slices from this publication are shown in Figure 1.1.

One downside to immunogold, even for these applications, is a limited capacity for multiplexing. Although different sizes of nano-particles can be synthesised, the size requirements for cryo-EM imaging, and the typical size distribution range for a batch of nano-particles, means orthogonal labelling of more than two targets in a single tomogram is challenging. The resolution of tagging (typically $\approx$40 nm [33] due to the size of the nano-particle and targeting moiety) is also limiting for many applications, and stoichiometry of labelling is a further concern [29]. Despite these limitations, immunogold remains a popular tagging strategy with structural biologists, and performs adequately in many circumstances. For cases where higher resolutions are required, or for intra-cellular labelling, we must look towards new and different options.

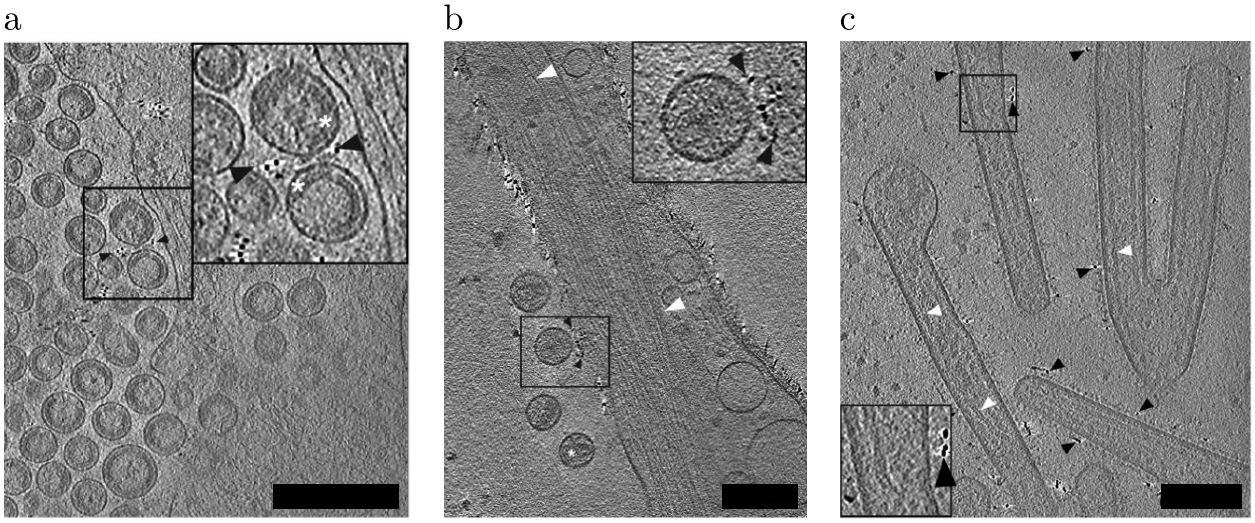


Figure 1.1: Examples of native immunogold labelling with CET, images and caption adapted from Yi et al. (2006)[32]. In all images, black arrowheads indicate immunolabelled proteins, and scale bars are 200 nm. (a) Slice through the middle of a cryo-electron tomogram of immunolabelled-tetherin and HIV-I virus-like particles attached to an HT1080 cell. White asterisks indicate an immature HIV-I Gag lattice. (b) Slice through of a cryo-electron tomogram of immunolabelled tetherin and HIV-I virions attached to an HT1080 cell microspike. Immuno-gold labelled tetherin located between HIV-I virions and the plasma membrane is observed. Preserved cytoskeletal elements are indicated by white arrowheads. (c) Slice through a cryo-electron tomogram of hRSV-infected HeLa cells with the hRSV F glycoprotein immunolabelled. Ribonucleoprotein are visible in some of the viral filaments (white arrowheads).

Metal-binding Proteins as Fusion Tags

Much as fluorescent proteins can be used as fusion tags for light microscopy, attempts have been made to construct an electron-dense clonable tag for CET. Metal-binding proteins are a convenient option; the capture of heavy metal ions provides a region of high local contrast allowing localisation of fused target proteins. Metallothioneins (MT) are low-molecular-weight cysteine-rich proteins which sequester a number of heavy metals at a variety of ionic states and stoichiometries [34, 35]. In 2007, Mercogliano and DeRosier first proposed using MT as an EM tag, demonstrating the technique with functional chimeras of gold-loaded MT fused to maltose-binding protein, which were easily visualised by cryo-EM [36].

Using this strategy, Diestra et al. reported CET imaging of MT-fused proteins (with known distribution patterns) in intact *E. coli* cells treated with gold (Figure 1.2a)[1][37], whilst later work by Bouchet-Marquis et al. used MT as a label for purified kinesin-decorated microtubules and desmin intermediate filaments (Figure 1.2b) [38]. Cd^{2+}-bound MT has been used for the same purpose, although the contrast is much lower than for gold ions [39]. Despite possessing a number of appealing properties, namely a small size (which permits nanometre precision) and clonability (which vastly simplifies intra-cellular use), there are a number of limitations with MT labelling which have prevented it from becoming widespread. Application of this system *in situ* requires an unnaturally high level of MT expression which can result in difficulties in growing cells, likely due to the protein's metal-scavenging nature [38]. Additionally, the high metal concentrations required for suitable contrast can be toxic to cells [38, 40] and can induce endogenous MT expression complicating data interpretation [37].

Ferritin particles have been likewise introduced as clonable CET labels (Figure 1.2c) [40]. This expression system is less toxic (although higher-than-physiological iron concentrations are still required), but has other disadvantages. Most notable are the large (>12 nm) particle size which limits labelling precision and the multivalent nature of ferritin particles which can affect labelled structures through clustering [40]. Metal-binding proteins in general also have a limited capacity for multiplexing and, depending on the target protein, fusion of the metal-binding protein may change the structure or function of its fusion partner.

Nucleic Acid Nanostructure Tags

Nucleic acid nanotechnology offers a potential, unexplored alternative for the future of CET labelling. DNA and RNA exhibit a relatively high contrast in cryo-tomograms and it has been shown extensively that nucleic acid sequences can be designed to assemble into a multitude of 2D and 3D shapes (Section 1.3 and 1.4), allowing many distinguishable labels to be generated and thereby opening up the possibility of multiplexing. Although nanostructure sizes are typically large ($\approx$ 10-100 nm), they can be designed with well-defined, asymmetrical shapes which allows high spatial resolution. Incorporation of aptamers[2] would offer a means of

[1] Cells were embedded in an acrylic resin, i.e. not native cryo-EM.

[2] Nucleic acid constructs optimised by *in vitro* selection experiments for high binding affinity to selected ligands.

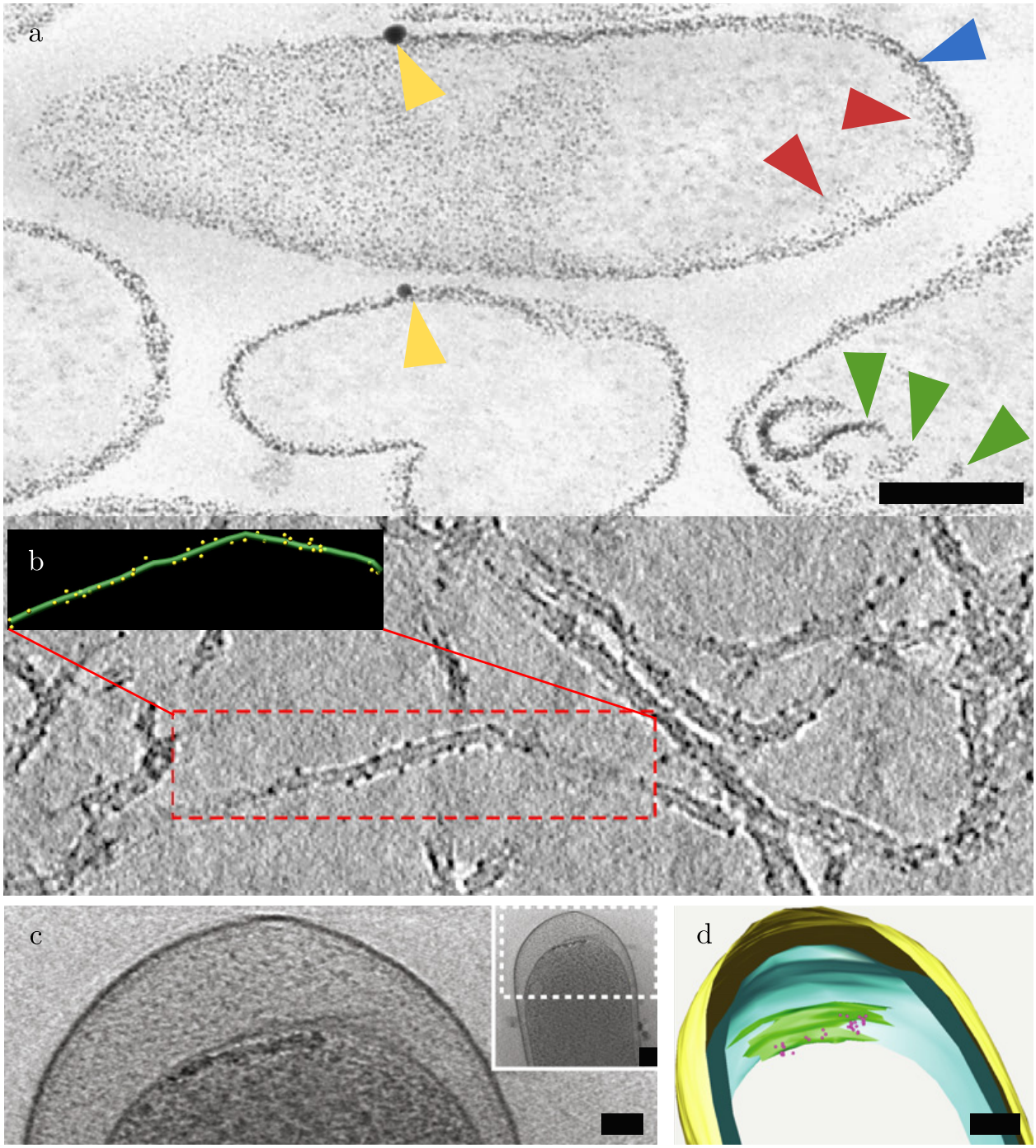


Figure 1.2: Overview of metal-binding proteins as tags for CET, scale bars 100 nm. Images and captions adapted from Diestra et al. (2009) [37] (a), Bouchet-Marquis et al. (2012) [38] (b) and Wang et al. (2011) [40] (c) and (d). (a) Localization patterns of maltose binding protein-MT showing a group of E. coli cells. In the largest bacterium, a cross-section of the cell reveals both internal (red arrowheads) and more peripheral (blue arrowhead) particles. In another bacterium particles form clusters and delineate structures of unknown identity (green arrowheads). Large electron-dense aggregates (yellow arrowheads) are occasionally seen. (b) Tomographic slice of desmin-MT intermediate filaments after incubation in Au(I)Cl. The gold clusters are visible along the outer surface of the filament. The inset shows a surface rendered representation of the indicated area. (c) Slice through a tomogram of an E. coli cell expressing CheY-ferritin. Inset shows the same cell at a lower magnification. (d) 3D segmentation of the cell in (c), showing the inner and outer membranes (cyan and yellow), CheY-ferritin (pink), and putative chemoreceptor clusters (green).

specific targeting.

Clonable structures are desirable for endogenous tagging; this could be achieved, in principle, through co-transcriptional folding of an RNA tag from a clonable DNA template, although *in vivo* assembly of RNA nanostructures of the required size has yet to be demonstrated.

1.3 Structural DNA Nanotechnology

1.3.1 DNA

DNA is a bio-polymer, widely known for its biological role: the faithful storage and replication of genetic information. This information is encoded by a sequence of monomers or nucleotides, each comprising a 5-carbon deoxyribose sugar linked to a phosphate group at the 5′ carbon and a nitrogenous base at the 1′ carbon. Within a DNA strand, the phosphate group of one nucleotide is linked to the following nucleotide ribose via the 3′ carbon, resulting in a repeating sugar/phosphate motif, dubbed the DNA 'backbone'. The informational storage capacity of DNA stems from the four distinct bases, often referred to by single-letter acronyms: adenine (A), cytosine (C), guanine (G) and thymine (T). A and G are both purines, composed of a six-membered ring fused to a five-membered ring, whilst C and T are pyrimidines, composed of a single six-membered ring.

DNA may exist in single-stranded or double-stranded forms, but in a biological context is most commonly found in the double-stranded B-form helix configuration, whose structure was finally elucidated in 1953 following the combined efforts of Franklin, Watson and Crick [41, 42, 43]. Each strand is wound in a right-handed helix, and the two strands are held together in an anti-parallel conformation by hydrophobic interactions and by hydrogen bonding between bases on the two strands. The most remarkable property of DNA is the stringency of these base-pairing interactions: A partners with T forming two hydrogen bonds, and G partners with C forming three hydrogen bonds. In both cases the pairing of a bulky purine with a smaller pyrimidine means the diameter of the double helix remains constant along its length. The double helix is further stabilised by 'aromatic stacking' between adjacent bases, which in reality is a complex interaction resulting from several non-covalent forces [44]. Spaces between the two strands are usually referred to as 'grooves'. As the two strands constituting the helix do not lie exactly opposite one another, there are two non-equivalent grooves: a wide, deep 'major groove' and a narrow, shallow 'minor groove'. A more detailed view of the chemical and structural properties of the B-form DNA double helix is provided in Figure 1.3.

In biology the base-pairing rules permit replication of the genetic code, but these rules also render DNA a convenient material for nano-scale rational design, on the premise that complementary sequences will hybridise to form a duplex. This was first noted by Nadrian Seeman in 1982, where he writes 'It is possible to generate sequences of oligomeric nucleic acids which will preferentially associate to form migrationally immobile junctions rather than linear duplexes, as they usually do' [45]. Beyond the obvious advantage of programmability,

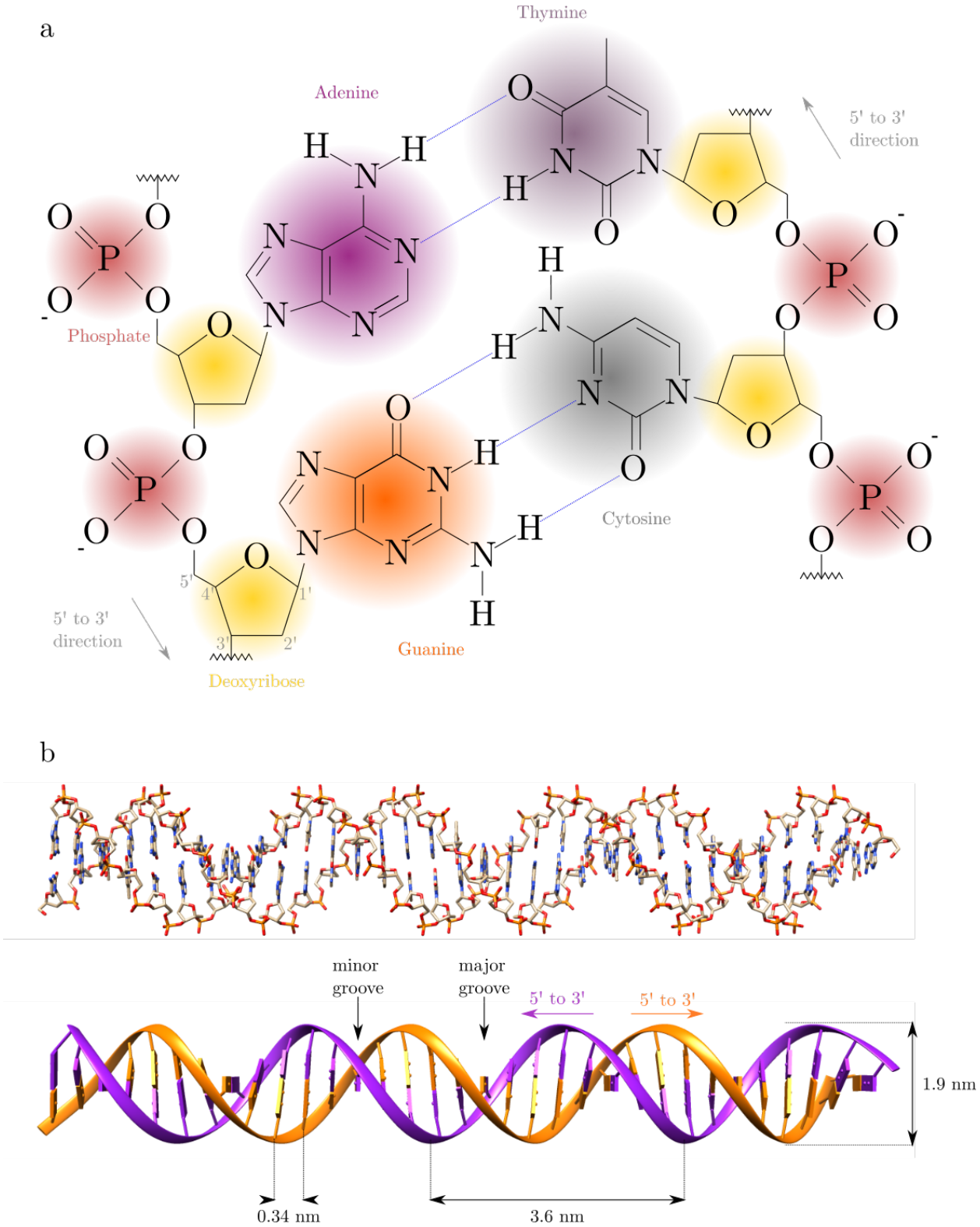


Figure 1.3: The structure of DNA: (a) Chemical structure of a DNA molecule, showing each of the four bases. Hydrogen bonds between bases are shown in blue. (b) Atomistic (above) and schematic (below) representations of the DNA B-Helix.

DNA possesses a number of characteristics that are desirable in a building material. Double-stranded DNA is a stiff polymer (possessing a persistence length of ≈50 nm [46]), allowing placement of structural elements with nanometre precision, but it also exists in a single-stranded form (possessing a persistence length of ≈0.75 nm [47]) which allows the inclusion of flexible regions. DNA strands can also be synthesised cheaply and with a wide range of modifications, and can be acted on by an arsenal of enzymes including ligases, restriction endonucleases, kinases, phosphatases and exonucleases; thus there is a pre-existing tool-kit (of both biological and synthetic origin) for manipulating DNA.

1.3.2 Early DNA Nanostructures

A discussion of the, now diverse, field of structural DNA nanotechnology must begin with Seeman's seminal work in 1982 [45]. Here he proposed using DNA strands to form immobile junctions, which are linked to one another via complementary 'sticky ends' to form a lattice. A 2D realisation of these crystals was achieved experimentally in 1998 [48], but Seeman's initial designs have also inspired a plethora of approaches beyond his original conception.

Many of the earliest reported DNA nanostructures were 2D arrays of repeated units [48, 49, 50] but, in parallel to this, methods of self-assembling discrete geometrical objects were also developed. Consider, for example, the assembly of a DNA tetrahedron from four oligonucleotides (Figure 1.4) [51]. Each edge of the tetrahedron is made from one of six 'sub-sequences' hybridised to its complement, and each oligonucleotide is made from three of these 'sub-sequences', or their complements, delineated by 2-nucleotide linkers which act as flexible hinges. Component oligonucleotides are mixed stoichiometrically and annealed (heated to a high temperature and cooled quasi-statically to reach a free energy minimum) in a buffer containing Mg^{2+}. Similar structures have been assembled using sequential, multi-step strategies [52, 53]. 'Nicks' exist in these structures where the 5′ and 3′ ends of adjacent oligonucleotides are held together spatially by Watson-Crick base-pairing, but without a covalent linkage through the DNA backbone. These nicks can be ligated enzymatically to stabilise the final nanostructures [52, 53, 54].

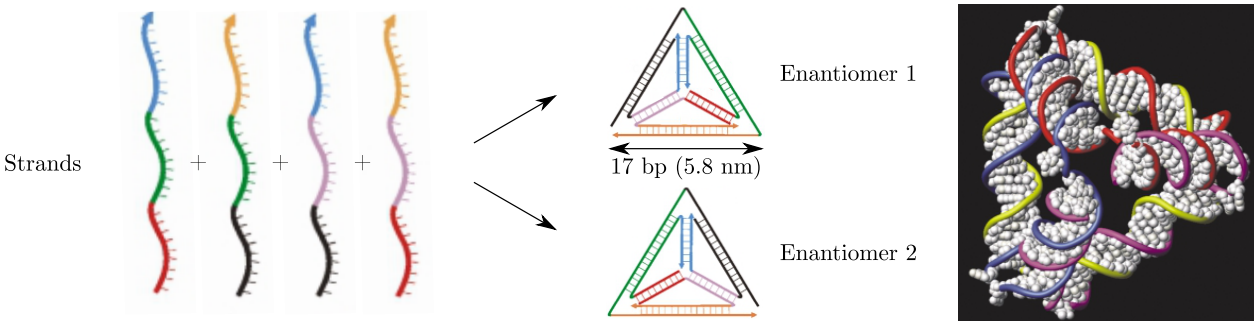


Figure 1.4: One pot assembly of a DNA tetrahedron (figure and caption adapted from Goodman et al. 2004 [51]). (a) Synthetic scheme for the DNA tetrahedron, with Schlegel representations of the final product illustrating both possible enantiomers. Each edge is represented in a different colour. (b) Space filling representation of a DNA tetrahedron with 17 bp edges and 2 bp hinges. The backbone of each strand is coloured uniquely.

A limitation of many of these early nanostructure designs is the formation of unwanted by-products, resulting in low yield of the target structure. Precise stoichiometry is generally required, and successful synthesis often hinges on multiple reaction and purification steps. Yield can be improved by introducing a hierarchy of interactions, whereby sub-units are pre-formed through strong interactions, and link together through weaker interactions to form lattices [48, 49, 55] or finite super-structures [54, 56]. An ingenious alternative to this is scaffolded nanostructures, where a long, clonable, scaffold strand is folded into a target structure by oligonucleotides. The beauty of this idea is that the oligonucleotides can be added in excess to the scaffold strand without causing any misfolding effects. Post-folding, the target structure can be easily purified from the excess of unbound oligonucleotides due to the large difference in molecular weight. The first implementation of this methodology was the synthesis of a DNA octahedron from a ≈1700 nucleotide single-stranded DNA molecule and five 40-mer oligonucleotides [57]. In this design, the overwhelming majority of base-pairs in the final structure are formed between nucleotides within the scaffold strand. Extending this technique to new geometries would therefore require a complete re-design of the scaffold strand, and the library of designable structures is limited by the requirement for a single-stranded folding pathway.

1.3.3 DNA Origami

2D DNA Origami

In 2006, Paul Rothemund introduced a novel technique for scaffolded DNA nanostructure design, DNA origami [58]. DNA origami nanostructures are generated from a long, scaffold strand (usually circular and of biological origin) which is folded into a designed shape by annealing with many synthetic staple strands (Figure 1.5). All the base-pairs in the target structure are formed between a scaffold nucleotide, and a staple nucleotide. The precise sequence of the scaffold is usually unimportant (providing it is known), and entirely different structures can be generated from the same scaffold sequence via hybridisation to different staple sets. This technique allows the construction of 2D and 3D nanostructures of arbitrary shape with high yield and relative ease and has become the architecture of choice within the DNA nanotechnology field.

In its original inception, this technique was used to generate 2D (single-layer) objects from sheets of parallel helices. The scaffold strand is designed to raster-fill a target shape, comprising one strand of each constituent helix. Locations where the scaffold strand switches from one helix in the array to an adjacent helix are 'scaffold crossovers', and the positions of these crossovers are chosen to be consistent with the B-helix geometry. The B-helix possesses an average helical repeat of 10.44 base pairs [59], thus a stretch of 16 base pairs corresponds to approximately 1.5 turns. A scaffold crossover between the nth and $n+1$th helices, must be separated from a crossover between the $n+1$th and $n+2$th helices by an odd number of half-turns. Likewise the distance between two distinct crossovers between the nth and $n+1$th helices must be an even number of half-turns. Rothemund positioned scaffold crossovers based on these requirements to yield flat structures.

The scaffold is pinned into the designed shape by the staple strands, with each staple strand binding two or more different regions on the origami (ergo the staple strand sequences are fully determined by the scaffold sequence). Locations where a staple switches from one helix in the array to an adjacent helix are 'staple crossovers', and these must fulfil the same geometric requirements as the scaffold crossovers. The geometry of the B-helix is such that a scaffold crossover between the nth and $n+1$th helices can occur at the same base pair position (or separated by an even number of half turns) relative to a staple crossover between the $n+1$th and $n+2$th and $n-1$th and nth helices. The link between crossover positions and global geometry has been exploited to generate curved nanostructures for both single-layer [60] and multi-layer [61] origami designs.

Lattice-based 3D DNA Origami

3 years after the origami method was first published [58], the design architecture was extended to multilayer structures, in which parallel helices are packed onto honeycomb [62, 63] or square [64] lattices. As with the 2D designs, scaffold and staple crossover positions respect the natural geometry of the B-helix, minimising global twist and strain in the final structure. More recently this strategy was extended to hexagonal and hybrid lattices [65].

Small modifications to the lattice-based origami method reveal new designable geometries. The inclusion of flexible hinges permits the assembly of wire-frame structures with rigid multi-helix edges [63]. This technique was demonstrated elegantly by Tian *et al.* in 2015 [66] where octahedral origami nanostructures were used to assemble well-defined 3D gold nanoparticle clusters. Multiple origami nanostructures can assemble into homo- or hetero-superstructures through complementary staple extensions [63] or carefully designed stacking interfaces [67]. This latter strategy was recently used to produce fully-specified gigadalton-scale structures, comparable in size and complexity to a viral capsid [68].

Designing 3D origami nanostructures by hand is possible but cumbersome. The design process has been greatly facilitated by the open-source software caDNAno [62], which offers live 2D and 3D representations of a structure throughout its design, and calculates possible staple and scaffold crossover positions for the user-defined geometry. The computational tool CanDo [69] generates 3D models, at both a schematic and atomistic level, from caDNAno input data.

Alternative Origami Architectures

The success and popularity of the origami technique has yielded many origami-based strategies for nano-scale construction. Some of the earliest realisations of 3D origami nanostructures were achieved by folding of 2D Rothemund-style origami nets [60, 70, 71], such as the hollow origami tetrahedron or cube shown in Figure 1.6b and c.

In 2013, the Yan group reported an alternative design strategy, in which four-arm junctions are used as vertices in a network of helical regions to generate both 2D (Figure 1.6g) and 3D (Figure 1.6h) 'gridiron' origami objects [72]. They subsequently published a

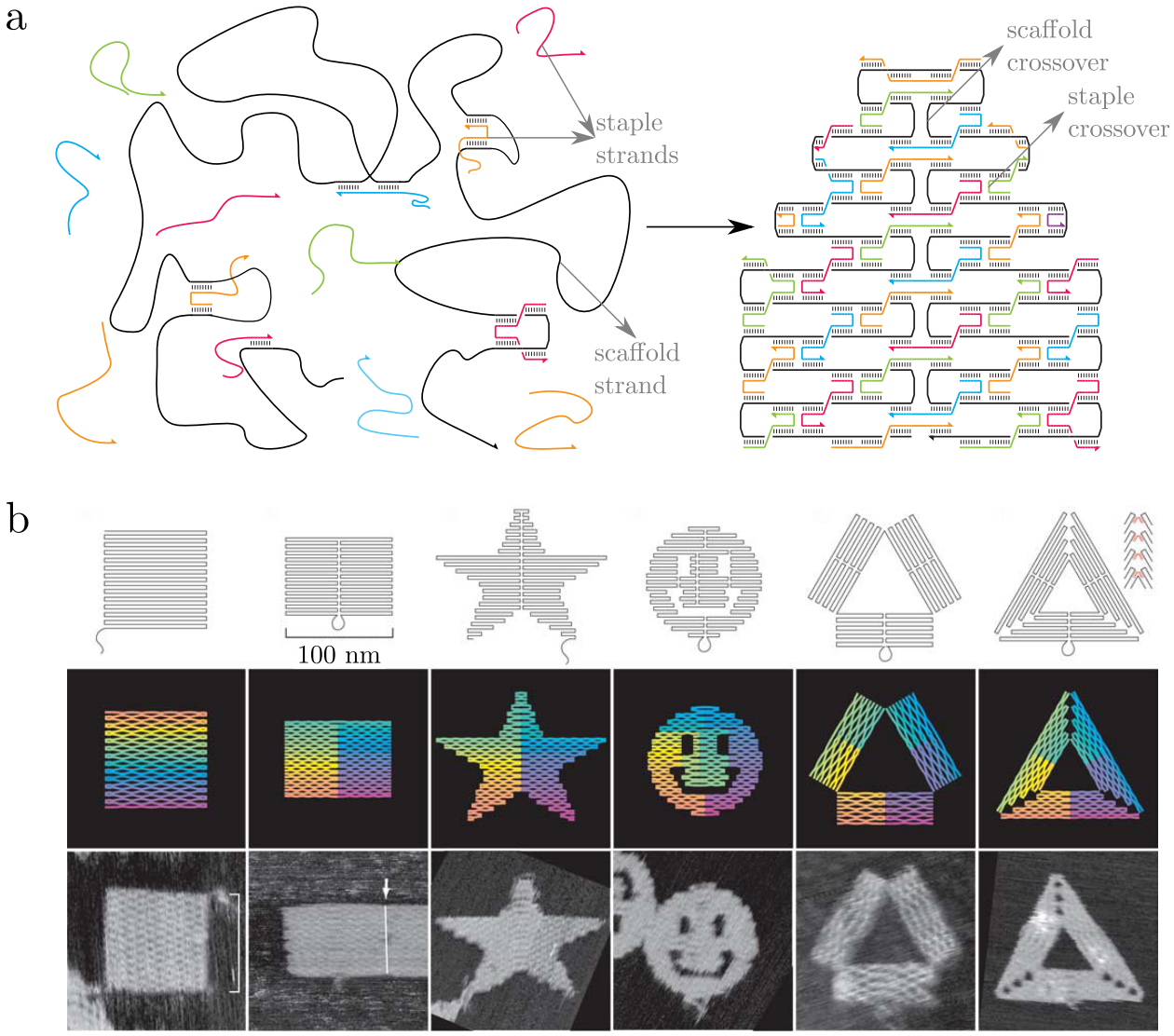


Figure 1.5: Overview of Rothemund's 2D origami method (figure and caption adapted from Rothemund, 2006 [58]). (a) Schematic view of origami folding: the long black scaffold strand is folded into a designed shape by annealing with many coloured staple strands. (b) Top row, folding paths for various shapes. Dangling curves and loops represent unfolded sequence. Second row from top, diagrams showing the bend of helices at crossovers (where helices touch) and away from crossovers (where helices bend apart). Colour indicates the base-pair index along the folding path; red is the 1st base, purple the 7,00th. Bottom row, AFM images. All images and panels are 165 nm × 165 nm.

related method of creating more complex wire-frame objects with a variety of multi-armed structures, such as the flower-and-bird pattern shown in Figure 1.6i [73]. This 'wire-frame' architecture has also been used to assemble a myriad of 3D mesh structures (Figure 1.6j). Automation of the design process makes this wire-frame method comparable to lattice-based structures in terms of versatility and accessibility [74, 75].

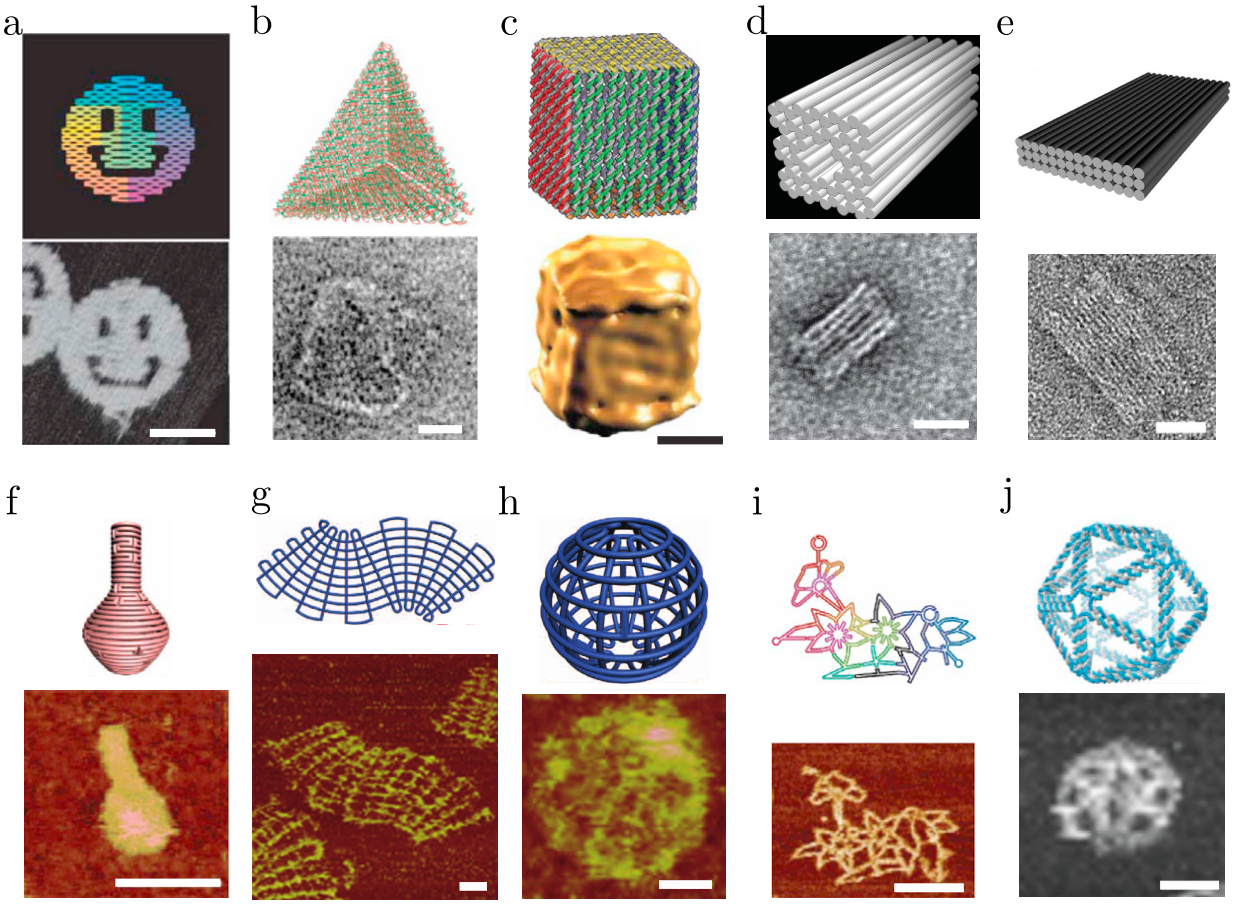


Figure 1.6: Overview of DNA origami design strategies. (a) DNA origami 'smiley' from Rothemund's 2D origami architecture [58]. (b) and (c) A hollow tetrahedron [70] and cube [71] from folding flat origami sheets. (d)'Square nut' origami structure formed from helices packed on a honeycomb lattice [63]. (e) 3 layer origami structure formed from helices packed on a square lattice [64]. (f) A nanoflask built from concentric rings of DNA helices [60]. (g) and (h) 2D and 3D gridiron structures based on DNA 4-arm junctions [72]. (i) A wireframe flower-and-bird pattern [73]. (j) A wire-frame icosahedron [74].Scale bars, 20 nm (b, c, d, e, g, h), 50 nm (a, j), 75 nm (f), 100 nm (i),

1.4 Structural RNA Nanotechnology

1.4.1 RNA

RNA shares many properties with its more famous sister molecule, but possesses two key chemical differences which have important implications for both biology and nanotechnology. The simpler of these differences is the nitrogenous base uracil (U) which replaces thymine in DNA. In biology this difference explains the superior informational integrity DNA possesses relative to RNA. Spontaneous deamination converts cytosine to uracil which is excised from DNA through enzymatic repair mechanisms [76], but in RNA the mutation is not detectable, since uracil is a natural RNA base. Uracil is also commonly seen to form non-canonical Watson-Crick base pairs with guanine; these G·U pairs are significantly more stable than G·T pairs in DNA [77]. In biology, G·U pairings are partly responsible for the degeneracy of the genetic code (as first posited by Crick in 1966 [78]) but are also observed in abundance within natural RNA complexes such as the ribosome [79], and can likewise be incorporated in synthetic RNA nanostructures.

The second key difference is the ribose sugar of the RNA backbone, which has an additional hydroxyl group on the 2′ carbon. This increases the number of possible geometries between nucleotides that give rise to base pairs. In 2001, Neocles Leontis and Eric Westhof offered a convenient classification of these geometries based on the observation that hydrogen bonding interactions between RNA bases exist between 'edges' of the two contributing bases: the Watson-Crick edge, the Hoogsteen edge or the sugar edge (which includes the 2′ OH) [80]. Bases can also interact in one of two orientations relative to the glycosidic bonds giving rise to 12 possible interaction geometries which result in at least 2 hydrogen bonds [80]. Only six of these orientations are even theoretically possible with DNA, and in almost all biological contexts DNA exists as a double-stranded B-helix with only *Cis* Watson-Crick/Watson-Crick base pairs. Conversely, RNA strands adopt complex and varied 3D architectures, and high resolution biological examples of 11 of the 12 possible base-pair geometries are currently available [80]. This diversity in local geometries means predicting the 3D structure of an RNA strand from its sequence is a significant challenge.

The 2′ hydroxyl group also makes RNA susceptible to base-catalysed hydrolysis, which is 100-fold faster for RNA relative to DNA under neutral conditions [81]. RNA therefore has a much reduced chemical stability compared to DNA, an additional reason why DNA is the preferred molecule for storing genetic information.

In addition to single-stranded RNAs, double-stranded RNA (formed of two antiparallel complementary strands) also exists in biology, most notably as the genetic material of some viruses, in the form of small interfering RNA (siRNA), or as intramolecular helical domains within larger structures. The ribose 2′ hydroxyl group would result in a steric clash within a B-form helix, so RNA instead adopts an A-form helix [82] (shown in Figure 1.7). An A-helix is also adopted by DNA under dehydrating conditions [42]. This geometrical difference is an important consideration when designing origami-like RNA nanostructures, as the helix structure affects placement of crossovers.

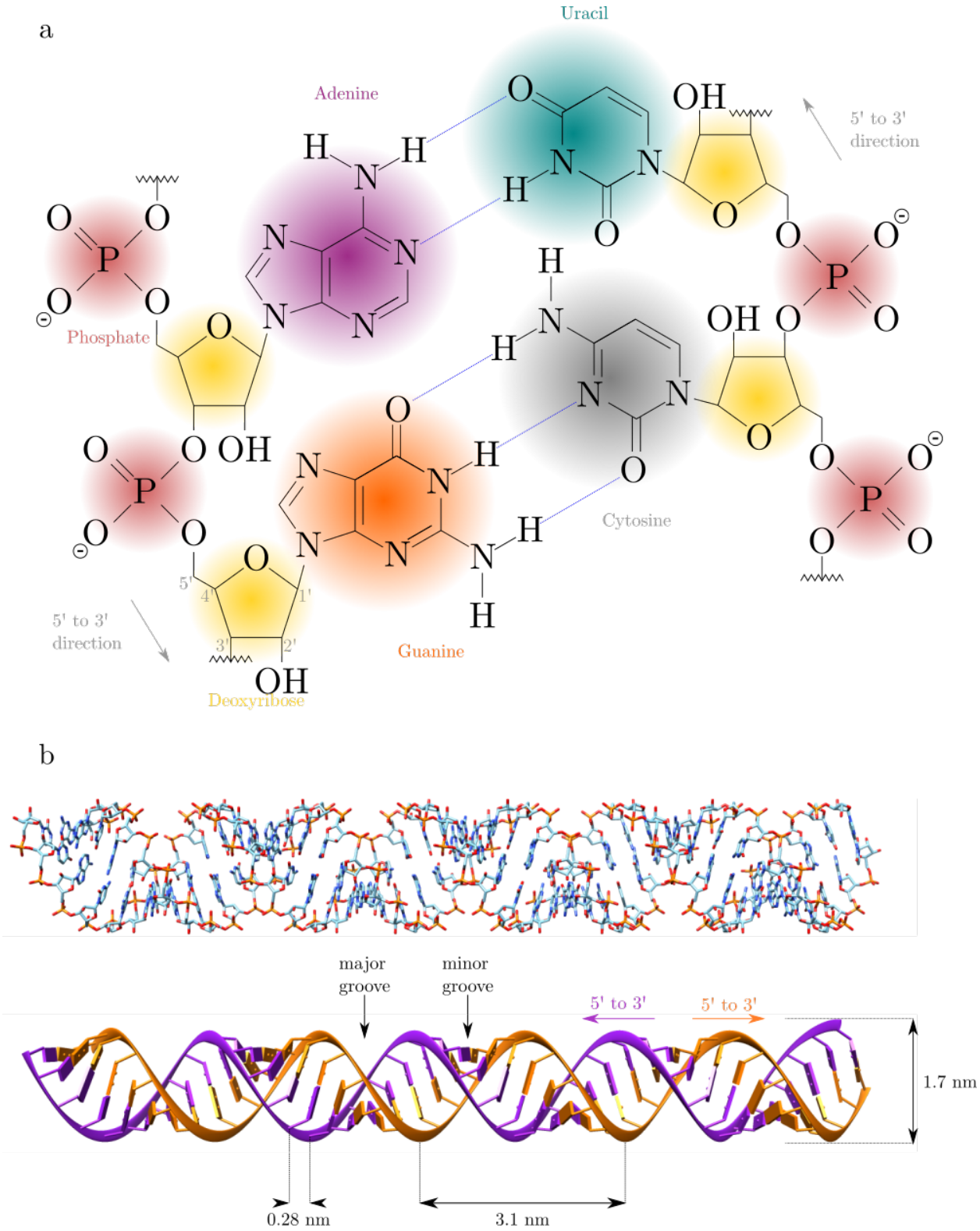


Figure 1.7: The structure of the RNA double helix: (a) Chemical structure of a double-stranded RNA molecule, showing each of the four bases. Hydrogen bonds between bases are shown in blue. (b) Atomistic (above) and schematic (below) representations of the RNA A-Helix.

1.4.2 RNA folding: a biological perspective

Once thought to be the mere middle man of the central dogma, RNA molecules are now known to perform a myriad of roles within the cell [83]. Many cellular processes are controlled by ribonucleoprotein (RNP) machines: complexes of structured RNA molecules and proteins. The most famous example is the ribosome, which performs protein synthesis according to a messenger RNA (mRNA) sequence [84]. Further examples include the splicesome [85], Ribonuclease P [86], telomerase [87] and the signal recognition particle [88]. In all cases, the primary catalytic activity is carried out by RNA components. Certain small RNAs can also form catalytic structures without any associated proteins, known as ribozymes [89]. Even mRNA molecules are replete with functional secondary structures at their non-coding termini, which contribute to a complex system of gene regulation [90].

A central premise of structural biology is that structure dictates function; thus in order to understand RNA's functional diversity it is important to understand how RNA assumes 3D structures in the cell. A wide repertoire of shapes are adopted by RNA molecules, but there exist certain rules which apply to most structured RNAs. Usually a set of short (< 10 base pair) A-form helices comprise the bulk of nucleotides. The base pairs which form this set are dubbed the secondary structure, with the primary structure referring to the $5' \rightarrow 3'$ nucleotide sequence. The secondary structure folds into the final 3D structure through the formation of tertiary contacts [91, 92]. RNA folding is generally hierarchical, in that the secondary structure is stable in the absence of tertiary structure, and does not change when tertiary interactions form [93] although exceptions to this rule exist [94].

The requirement for RNA molecules to faithfully fold into fully-specified 3D conformations is 'the RNA folding problem' which, as described by Daniel Herschlag in 1995,[95] can be sub-divided into a thermodynamic problem and a kinetic one. Firstly, for an RNA strand to adopt a single, native structure at equilibrium, that structure must be strongly energetically favoured over all competing structures, including those with identical secondary structures. This is the thermodynamic folding problem. Herschlag theorised that this problem could be resolved by specific RNA-binding proteins which stabilise a single fold, contributing to its final thermodynamic stability.[95] Continuous RNA-protein interactions are required to maintain this effect. Although this behaviour is not ubiquitous, it does occur in some systems. One example is the CYT-18 protein, which suppresses mutations in the bacteriophage T4 *td* intron through specific binding to the catalytically active (self-splicing) structure of the intron core.[96]

The kinetic problem is that, in order to function in cells, a strand of RNA must fold into a single target structure within an appropriate time-scale. RNA folding energy landscapes are rugged, interspersed with a number of metastable mis-folded states which can act as kinetic traps, slowing the folding process significantly [97]. *In vitro* these kinetic barriers are quite prominent [97], whereas *in vivo* RNA molecules are known to fold efficiently. The solution proposed by Hershlag in his seminal paper [95] is RNA chaperones: transient, non-specific RNA binding proteins which aid RNA folding by preventing mis-folding, or by resolving mis-folded steps. RNA chaperones are now known to be a critical class of proteins, common to all forms of life. Some well-characterised examples are the nucleocapsid proteins of HIV-1

[98], the *E. coli* protein StpA [99] and the prokaryotic ribosomal proteins L1 [100], S1 [101] and S12 [102].

In vivo, nascent RNA molecules fold co-transcriptionally which affects the folding kinetics [103]. As the RNA strand is synthesised from the 5′ end, local interactions can form quickly. These short range structural elements can then rearrange to form larger domains over a longer time scale. Domains of a certain size will become trapped kinetically when the free energy barrier for structural rearrangement becomes too large, and as a result the conformation of a co-transcriptionally folded RNA strand may not be the minimum free energy structure. Instead, longer RNA strands contain domains of low free energy, linked by longer range interactions which fit with the short range domains [104]. This provides a new perspective on the thermodynamic folding problem: the functional conformer of an RNA molecule may not be favoured energetically at all, but trapped in the conformation it adopted whilst being synthesised. Transcription speed also affects folding, and both decreasing or increasing the transcription speed can lead to inactive transcripts [105, 106].

1.4.3 Structural RNA Nanotechnology

In vitro RNA nanostructures by annealing

The field of structural RNA nanotechnology has progressed in parallel with DNA nanotechnology. The greater flexibility of RNA-RNA interactions opens up a world of new design possibilities which are not accessible with DNA, but this is a mixed blessing as it also means RNA structures are more difficult to predict. Coupled with reduced chemical and nuclease stability, and increased synthesis costs, RNA has yet to become a widespread staple material for nano-scale design.

Luc Jaeger pioneered the use of RNA as a synthetic building block, taking advantage of natural folding motifs which facilitate assembly *in vivo*. For example, the 'kissing loop' complex (formed by base-pairing between the ends of two hairpin loops) which facilitates RNA dimerisation in HIV, inspired the design of 'tectosquares', as depicted in Figure 1.8 [107]. In this architecture, four constituent RNA molecules (folded by annealing) assemble through kissing loops interactions to generate one tectosquare module. These tectosquare modules are also designed with sticky ends (lengths of single-stranded RNA, complementary to an equivalent strand on a different tectosquare) which permits assembly of these modules into well-defined multimers and lattices. In a clever extension to 3D, this design strategy was used with modified tRNA units, bringing together naturally occurring 3D motifs to generate rationally designed polyhedra, shown in Figure 1.9 [108].

This modular approach to design has been the driving principle behind most RNA nanostructure research; 3D modules (elucidated from biological RNAs) are connected by designable helical domains, thereby combining top-down and bottom-up approaches. The database RNAjunction [109] contains PDB files of RNA structures (kissing loops, bulge loops and junctions) searchable by geometry and assists the design process considerably. In addition to structural elements, modules may bestow functionality, e.g. through aptamers (nucleic acid sequences which bind a target with high affinity and specificity) or siRNAs

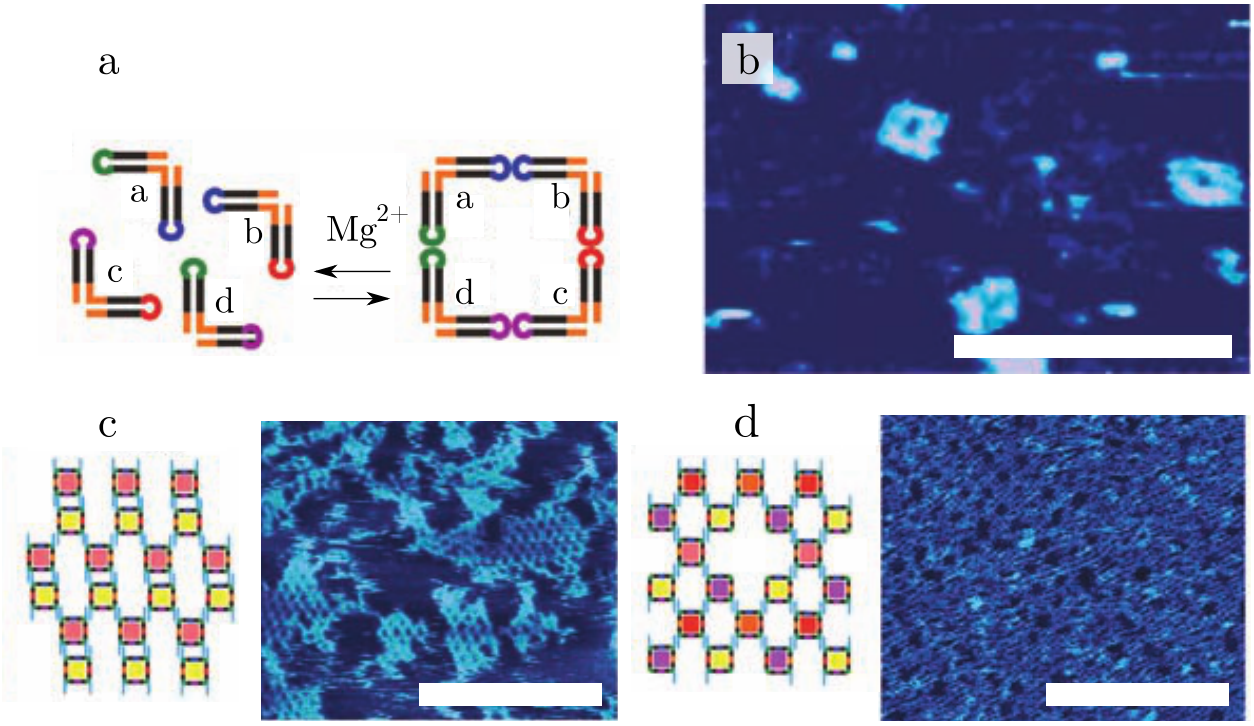


Figure 1.8: Tectosquare RNA nanostructures. (a) Tectosquare assembly scheme. (b) AFM visualisation of distinct tectosquares (scale bar 100 nm). (c) and (d) Schematic (left) and AFM (right) visualisations of two lattices generated from tectosquare modules (scale bars 500 nm). Images and captions adapted from [107].

(short interfering RNAs, involved in genetic regulation).

Following an entirely bottom-up strategy (as is typical in DNA nanotechnology), Han *et al.* assembled remarkable origami-like DNA and RNA nanostructures by annealing from single strands in 2017 (Figure 1.10) [110]. Similar to Rothemund's DNA origami technique [58], this strategy enables construction of arbitrary user-prescribed 2D shapes, raster-filled with parallel DNA/RNA helices. Compared to scaffolded origami, single-stranded nanostructures offer greater potential for clonability, and therefore large-scale production with future possibilities of directed evolution, but there is a key challenge in achieving structural complexity whilst maintaining topology simplicity (to avoid the kinetic traps imposed by knots). The innovation with this architecture is the combination of partially complemented DNA/RNA (helical domains) with paranemic crossover cohesion (locking domains), yielding an unknotted structure. A detailed overview of the design principles alongside an example structure is depicted in Figure 1.10.

In vitro RNA nanostructures by co-transcriptional folding

Thus far, the RNA nanostructures introduced are assembled by annealing one or more RNA strands to achieve the minimum energy conformation, but one of the principle goals behind structural RNA nanotechnology is the assembly of nanostructures in cells by transcription from a DNA template. There are currently very few literature examples of *in vivo* assembled RNA nanostructures (described below), but in recent years there has been an increased focus

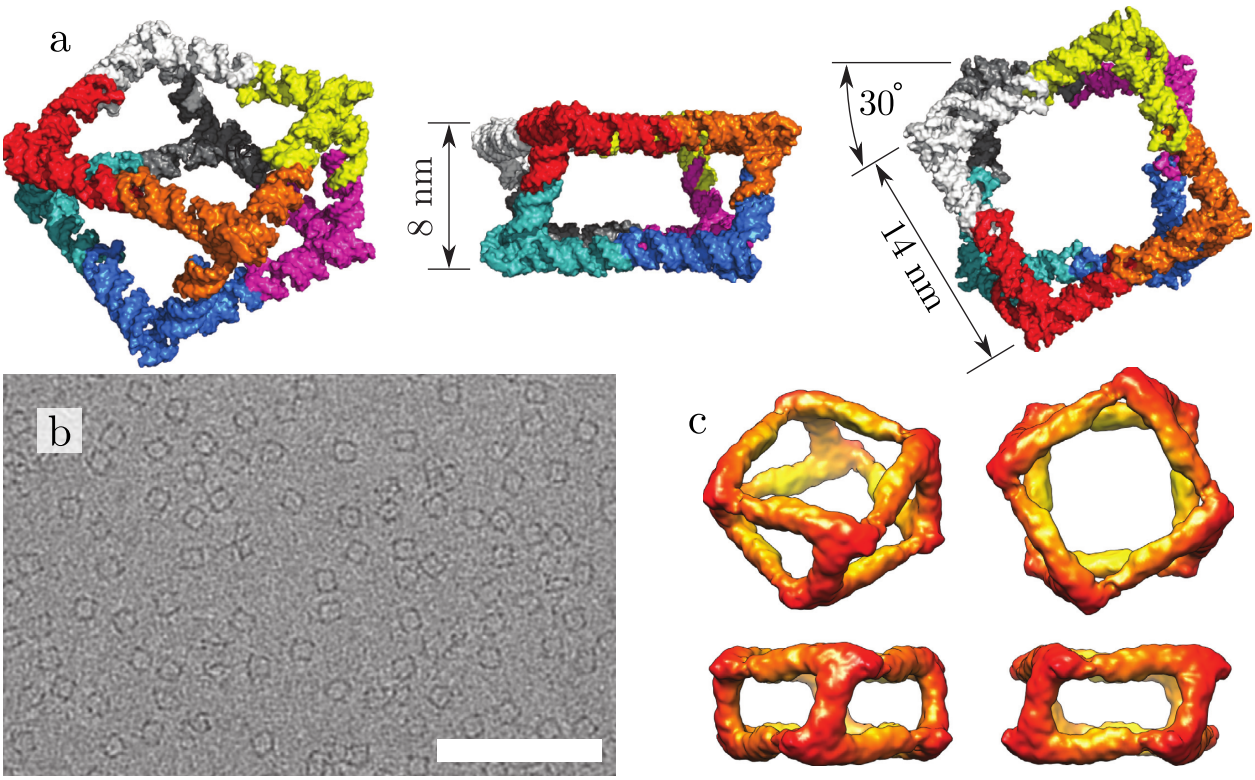


Figure 1.9: RNA polyhedra. (a) 3D model of non-uniform square antiprism formed by six distinct modified tRNA units (represented in different colours). Three different views are shown. (b) Typical cryo-EM image of antiprism particles. More than 85% of objects seen are well-firmed antiprisms, scale bar 100 nm. (c) Reconstructed three-dimensional model of the non-uniform square antiprism at a resolution of 24.5 Å. Images and captions adapted from Severcan et al. (2010) [108]

.

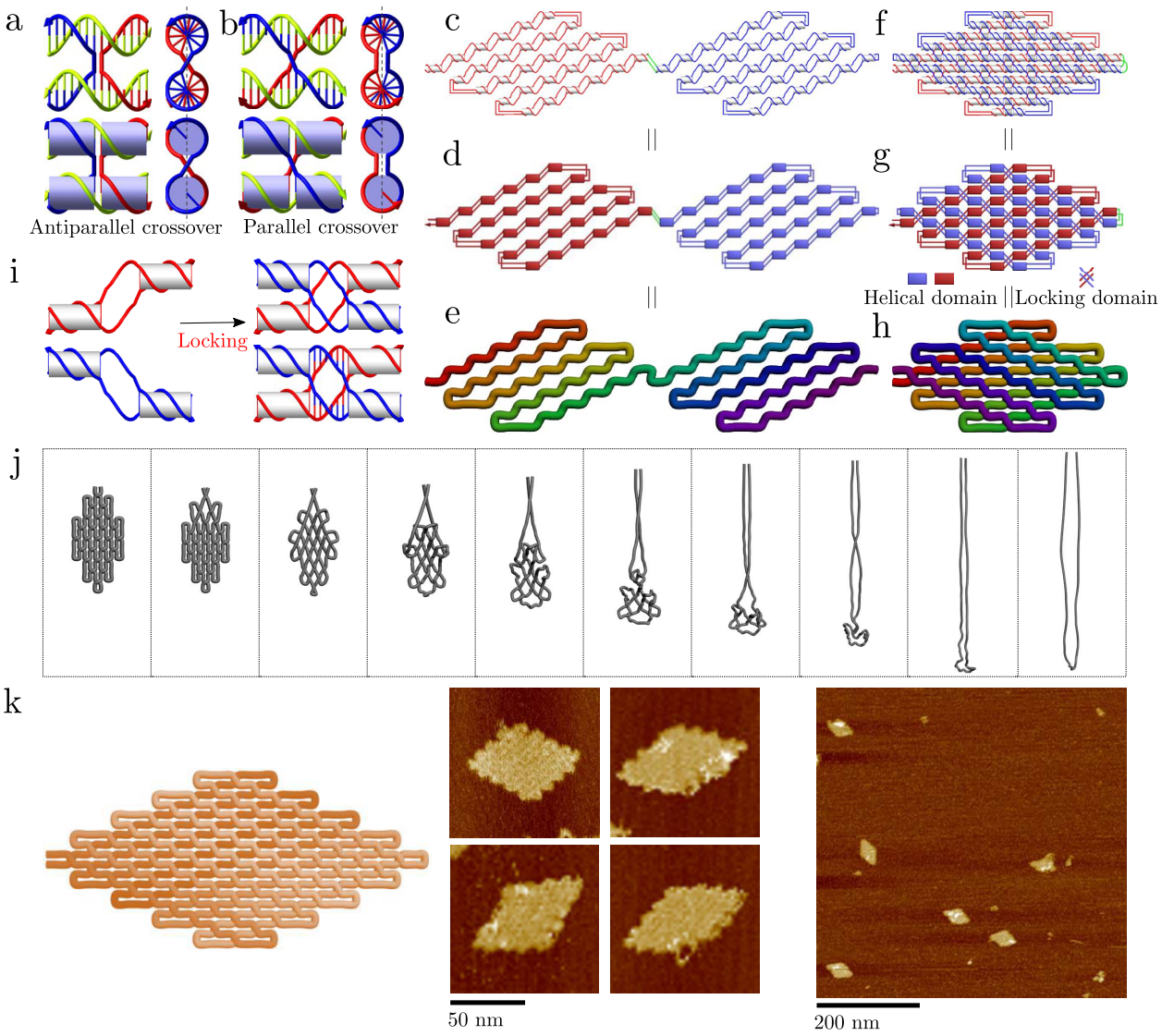


Figure 1.10: Design and assembly of single-stranded origami. (a) and (b) Schematics of antiparallel crossover designs (a) and parallel crossover designs (b) illustrating their local topology. The dashed lines denote the plane containing the helical axes. (c) to (e) Models of a typical illustrative double-stranded intermediate for single-stranded origami. (f) to (h) Models of a fully-formed single-stranded origami structure. (c) and (f) Double helical models with cylinders depicting the uninterrupted base pairs. (d) and (g) Cartoon models highlighting the helical domains (rectangles) corresponding to the white cylinders in (c) and (f) and locking domains (crosses) corresponding to the grey cylinders in (f). Bottom-layer strand sections and rectangles are coloured in red, and the top-layer ones are coloured in blue, whereas the green lines denote the connection between the red layer and the blue layer in (c), (d), (f), and (g). (e) and (h) Pipeline style models representing the folding track of the double-stranded intermediate. The model is coloured with a gradient starting from the 5′ and 3′ ends (red) to the middle of the strand (purple). (i) Schematic depicting the formation of a locking domain. (j) Dynamic relaxation model showing a time-lapse snapshot of the relaxation process of a single-stranded origami pipeline model (h) under simulated gravity, with both of its ends fixed. Schematic (left) and AFM images (centre and right) of a 6337 nucleotide 9× 9 rhombus shape RNA single-stranded origami.

on co-transcriptional assembly of RNA nanostructures *in vitro*. This permits the design of RNA nanostructures which are potentially suitable for *in vivo* applications, but with a more controllable environment for characterisation.

In 2010, Afonin *et al.* demonstrated the co-transcriptional assembly of RNA nano-cubes from 6 or 10 strands which assemble through simple inter-molecular Watson-Crick base-pairing [111]. This architecture is equivalent to that of early DNA structures, such as the tetrahedron depicted in Figure 1.4 [51], and the yield of complete structures ($\approx 70\%$ for co-transcriptional assembly) is likely limiting for extensions to larger and more complex objects. Smaller nanostructures such as these are nonetheless interesting for therapeutic applications, and in 2012 the same group reported the co-transcriptional assembly of siRNA-functionalised nano-rings, shown to induce gene knock-down in *ex vivo* pilot experiments [112]. For more versatile assembly options, structural elements outside of basic helical domains can be incorporated, e.g. 120° kissing loop domains are included in the RNA nano-ring design [112]. In 2017 Cody Geary and co-workers developed a structural syntax of RNA modules for programming strand assembly, and with these modules synthesised a varied library of discrete and repeating RNA nano-structures [113].

As previously mentioned, single-stranded nanostructures offer the advantage of clonability and thus large-scale production. For *in vivo*-folded nanostructures additional advantages exist: structures can be transcribed from a single gene (and therefore only one terminator domain per structure is required); multiple strands do not need to find each other in a crowded cellular environment; and strand stoichiometry is not a concern. In 2014, Geary *et al.* demonstrated the co-transcriptional folding of nanostructures from long, single strands of RNA [114, 115]. These structures comprise anti-parallel RNA helices, precisely organised by intramolecular kissing loop interactions and a newly introduced cross-over pattern termed the 'dovetail seam'. The design architecture was used to generate 2D tiles, capable of assembly into hexagonal lattices via intermolecular kissing interactions. An overview of one of these tiles is presented in Figure 1.11 [114]. Importantly, the dovetail seam and kissing loops motifs are paranemic interactions, meaning they can form without unwinding of the contributing helices, thus circumventing topological traps. Nanostructures of up to 660 nucleotides were reported, and extensions of the design protocol to arbitrarily large 2D shapes have been proposed [116]. Inspired by this work, a wireframe strategy for co-transcriptional assembly of RNA polyhedra has also been introduced, with some promising preliminary results [117].

In vivo RNA Nanostructures

To date, examples of *in vivo* folding of RNA nanostructures have been few in number, but some recent examples demonstrate the potential utility of RNA nano-structures in cells.

In 2011 Delebecque et al. designed and assembled 1D and 2D RNA scaffolds with integrated protein-docking aptamers to spatially organise the hydrogen-producing pathway, allowing control of hydrogen production through scaffold architecture [118]. In 2014 this strategy was used to program different cell behaviours, including the co-localisation of split GFP fragments to generate cell fluorescence and an increase in the metabolic output of

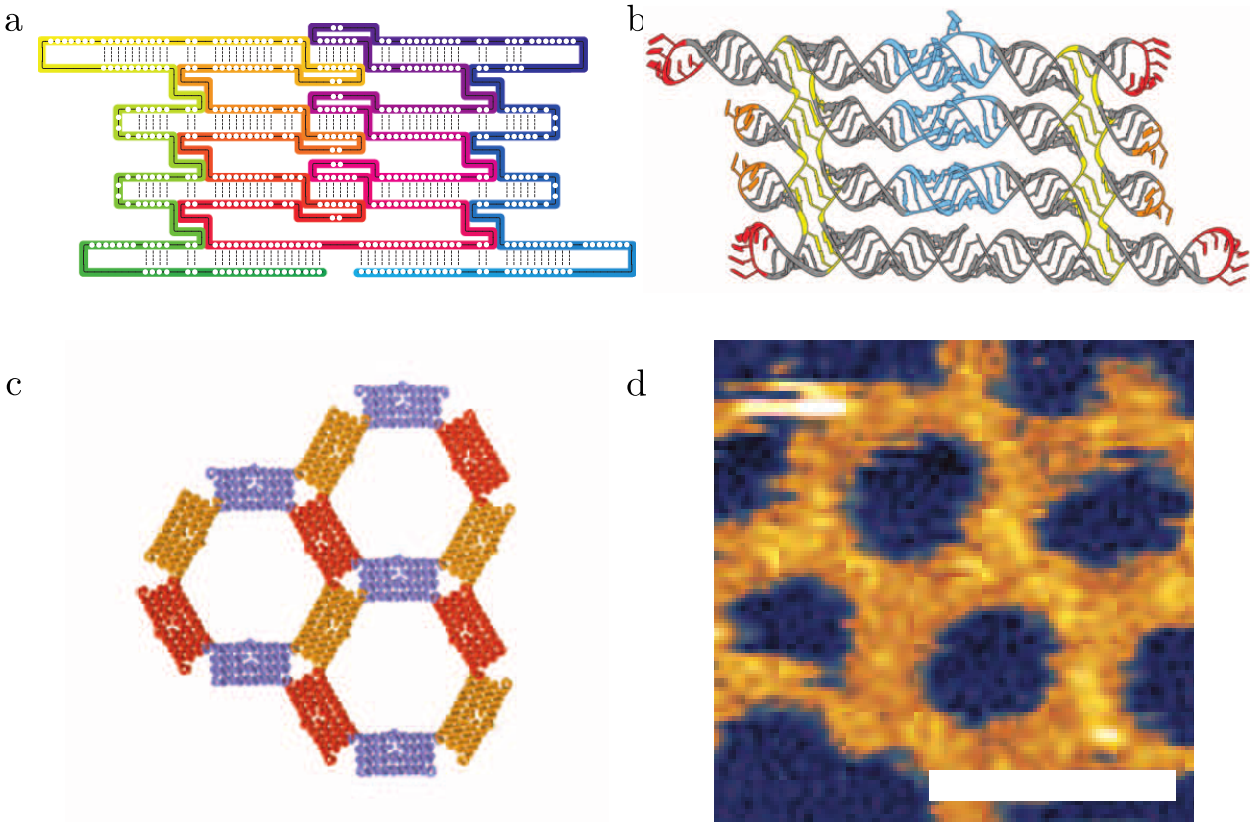


Figure 1.11: Overview of 4H-AE tile. (a) Strand path diagram showing the order of synthesis, 5′ to 3′ from blue to green. (b) Rendering of atomic model. Helices are shown in grey, 180° kissing loop in blue, one half of 120° kissing loop in red, dovetail seam in yellow. (b) Schematic for tile assembly. (d) AFM image of tile assembly, scale bar 50 nm. Images and captions adapted from Geary et al., 2014 [114].

enzyme pathways with up to four constituent enzymes [119].

In 2018, two groups reported the *in vivo* production of discrete single-stranded RNA nano-structures. Li *et al.* introduced a strategy for designing wire-frame structures and demonstrated *in vivo* assembly of a 339 nucleotide double-square and, at much lower efficiency, a 632 nucleotide tetra-square [120]. Jepsen *et al.* made use of the RNA origami method introduced in 2014[114] to develop an RNA aptamer-based FRET system in which the presence or absence of the small molecule S-adenosylmethionine (SAM) alters the conformation of an internal SAM riboswitch [121]. This conformational change affects the relative position of internal spinach [122] and mango [123] aptamers, resulting in a SAM-dependent FRET signal. A system responsive to the presence of a specific RNA sequence was also reported.

1.5 Summary and Thesis Structure

This chapter introduced cryo-EM and CET as techniques for structural biologists, highlighting the challenge of unambiguously identifying specific structures *in situ*. I have introduced currently-used methods of labelling molecules for CET and outlined the limitations of these techniques, illustrating the need for a new tagging method. I propose using DNA and RNA nanostructures with targeting properties to achieve this goal. I have highlighted recent advances in DNA and RNA nanotechnology, focussing on DNA origami and co-transcriptionally folded RNA nanostructures which will be the basis for my tagging methods.

The rest of this thesis will be divided into two projects. Chapters 2 and 3 will be concerned with Project 1: the design and functionalisation of DNA origami nanostructures as tags for cryo-EM. In Chapter 2, I will demonstrate the cryo-EM imaging of a number of different DNA origami nanostructures in a biological context and assess their suitability as tags. Here I develop a purification method for producing extremely clean, high-concentration samples for cryo-EM. I will also investigate origami lattice type and heavy-metal binding as methods of improving contrast. This chapter culminates in the design and synthesis a 'signpost' origami nanostructure for use as a cryo-EM tag.

In Chapter 3, I will address the necessary targeting properties for cryo-EM tags. I will show that an RNA aptamer targeting GFP can be incorporated into DNA origami nanostructures, and retains its targeting properties. I will use these aptamer-conjugated nanostructures to demonstrate *in situ* binding to a murine leukaemia virus test system as well as to cells decorated with the murine leukaemia virus glycoprotein. I will also show the electroporation of DNA nanostructures into cells, with future intra-cellular labelling applications in mind.

Project 2 will be covered in Chapters 4 and 5, where I will introduce novel strategies for building 3D RNA nanostructures co-transcriptionally, for future implementation as a genetically encoded tag for cryo-EM. In Chapter 4, I will introduce a grid-based approach, where nanostructures comprise a set of mutually-interacting branched hairpin

units, connected by single-stranded linkers and arranged on a square lattice. I will show the design, synthesis and characterisation of a number of different nanostructures with this architecture, including promising TEM results that suggest a majority population of well-folded structures. I will also show that these nanostructures can be used to build supermolecular assemblies.

In Chapter 5 I will consider an alternative route towards RNA design. Here I will investigate different ways of extending an already-established 2D RNA origami architecture into 3D. I will introduce and characterise two motifs for achieving this goal: the PX_{major} motif and the BLKL motif. I will explain the design of 3D nanostructures built with these motifs, and characterise the resulting structures.

Finally, Chapter 6 is a conclusive chapter that summarises the full body of work presented in this thesis.

Chapter 2

DNA Origami Nanostructures for Cryo-electron Microscopy

In this chapter, I will present a number of different DNA origami nanostructures and assess their suitability as tags for cryo-EM. After introducing my project aims, I will present pilot data with a wedge origami nanostructure which I used to optimise the purification procedure for cryo-EM, and to show that solid, lattice-based origami structures can be visualised with high contrast in a biological context. I will then present two new nanostructures and establish constraints for reliably visualising structures, and for distinguishing multiple different structures. I will conclude by presenting the design and synthesis of a 'signpost' origami nanostructure, optimised for visualisation by cryo-EM.

2.1 Author Contributions

Céline Journot designed the tetrahedron origami nanostructure in Section 2.4.1 and determined synthesis conditions. AFM was performed by Emma Silvester with guidance from Natalie Haley. The preliminary cryo-EM (Figure 2.6) was performed by Susan Lea and the remaining cryo-EM/CET in this chapter was performed by, or under the supervision of, Lindsay Baker. Lindsay also built the tomogram shown in Figure 2.17. The rest of the work presented in this chapter (design of the remaining origami nanostructures, negative-stain TEM characterisation of nanostructures and image analysis, development of the origami purification protocol and heavy-metal binding experiments) were performed by Emma Silvester.

2.2 Project 1 Aim: A DNA Origami Tag for CET

As discussed in Section 1.2.3, high-electron-density structures with tagging properties are commonly used for CET, but currently there is no method that preserves both contrast and resolution without affecting the target system. DNA origami is a versatile and robust technique for generating user-defined structures with a high degree of uniformity. The rigidity

of DNA origami nanostructures means targeting modules (such as aptamers, or chemically-modified oligonucleotides) can be placed with nanometre precision, and the comparatively high electron-density of DNA (due to repeating phosphate groups) means high contrast can also be achieved. I therefore propose using DNA origami nanostructures with targeting properties as tags for CET.

There are many possible target systems that could be addressed with origami tags. Labelling samples with external targets, such as viral glycoproteins or cell surface proteins is the simplest option, and will be my primary focus. For these systems, origami tags can be introduced without any disruption to the target structures. For downstream applications in intracellular labelling there will be additional challenges, including stability, toxicity and efficient delivery of nanostructures. These points will be addressed in more detail in Section 3.4.1.

In Project 1 (Chapters 2 and 3) I will investigate the use of DNA origami tags as a general tool for CET, focussing on viral and extracellular membrane applications but with future intracellular applications in mind. Chapter 2 will focus on the structural design of origami tags with desirable imaging properties and Chapter 3 will focus on functionalising tags for targeting, and delivering tags to cells.

2.3 Pilot Work with an Origami Wedge

2.3.1 Design

Cryo-EM and CET of DNA origami have become more commonplace over recent years. However, since CET is generally used to characterise origami, as opposed to origami behaving as a tool for CET as in my implementation, imaging is typically performed with purified nanostructures in buffer. Since origami tags will be used in the context of a biological system, the level of background will be much higher than for purified nanostructures, especially for the long-term goal of intracellular labelling. It is therefore necessary to image origami nanostructures in a biological environment before determining if they are suitable as tags for biological systems. I will perform cryo-EM imaging of origami samples mixed with cell lysate as a proxy for an intracellular environment, and assess the contrast of different nanostructures in this context.

As I mentioned in Section 1.2.3 there are requirements that a tag must fulfil: it must be a suitable size and provide adequate contrast to be distinguished in a noisy biological context; it must possess an asymmetry such that different views of the structure are distinguishable and thus the position of an attached targeting region may be deduced; and it must bind a target of interest with high affinity and specificity. This latter point will be addressed in Chapter 3. To address the first two requirements, I designed a wedge-shaped DNA origami nanostructure with the following characteristics:

- The structure comprises 52 helices arranged in a 4×13 honeycomb lattice, in which each internal helix has three adjacent partners. This densely-packed architecture should

offer high contrast from all viewing angles.

- The top view of the nanostructure is asymmetric, allowing orientation to be determined from a transmission image.
- The principle dimensions of the nanostructure range from ≈ 10 to ≈ 100 nm to assess the nanostructure size required for high-contrast labelling.

Scaffold routing and staple sequences were designed using caDNAno software [62], and 3D models of the structure were generated using Cando[1] [69]. Depictions of alternate views of the designed structure are shown in Figure 2.1. Early in the development of DNA origami, experiments revealed that circular scaffold strands result in a higher yield of well-folded structure than linear scaffolds [58], thus I have used the sequence of the circular, phage-derived M13mp18 scaffold in generating this design. A caDNAno overview and full set of staple sequences are provided in Appendix A.3.1.

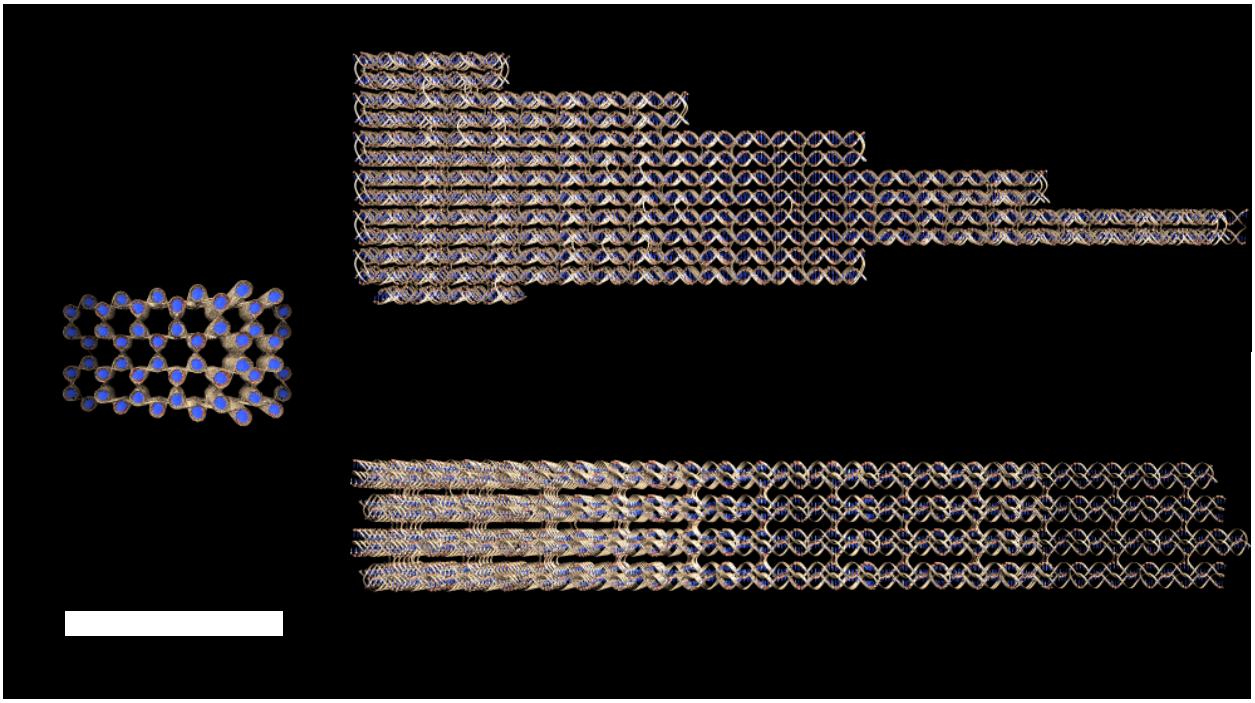


Figure 2.1: Orthogonal views of designed wedge-shaped origami nanostructure as generated by Cando and rendered in UCSF Chimera [124] (scale bar 20 nm). End-, top- and side-views correspond to the left, top right and bottom right panels shown.

2.3.2 Synthesis and Characterisation

Optimising Folding Conditions

DNA origami nanostructures require stabilisation by cations, usually Mg^{2+}. To determine the Mg^{2+} concentration required for folding, M13mp18 scaffold[2] at a final concentration of

[1] Staple sequences and caDNAno design provided in Appendix A.3.1.
[2] Scaffold type p7249, purchased at 400 nM from Tilibit nanosystems.

10 nM was mixed with a 10-fold excess of staple strands and folded over a 16 hour annealing ramp (Appendix C.1.1) in a buffer containing 1×TE[3] and $MgCl_2$ at a range of concentrations. Annealed samples were evaluated by agarose gel electrophoresis (AGE) (Appendix E.1) as shown in Figure 2.2. A shift in mobility relative to the single-stranded scaffold band indicates origami folding, and this can be seen over a wide range of $MgCl_2$ concentrations (9-29 mM). At low (5 mM) $MgCl_2$ concentrations, a low mobility band is observed, which is indicative of poor/incomplete folding, whereas at high (15-29 mM) concentrations, an increasing amount of aggregation occurs. Based on these results, a final $MgCl_2$ concentration of 12.5 mM was used for annealing.

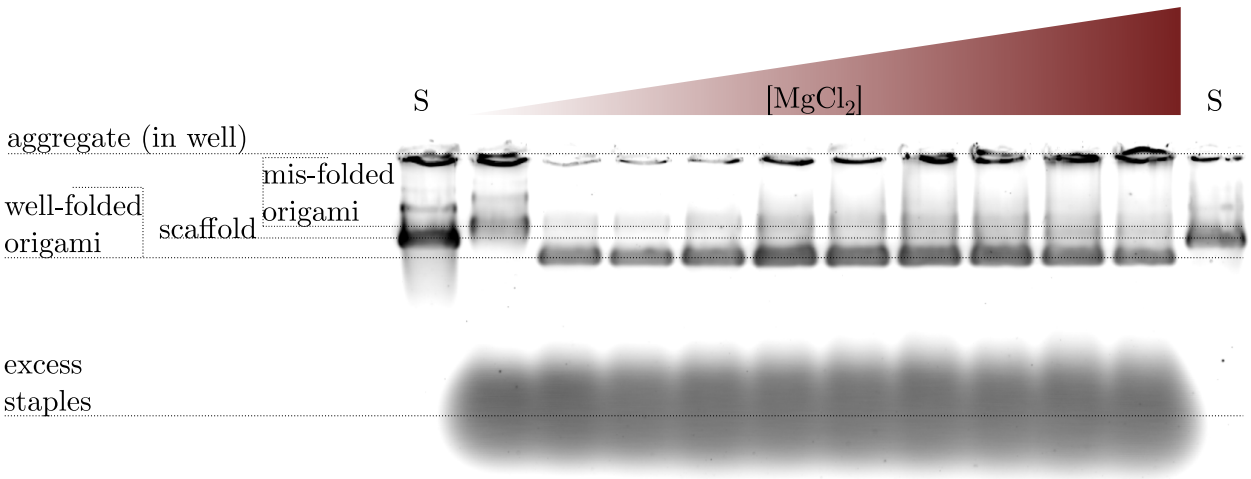


Figure 2.2: Agarose gel electrophoresis of wedge origami folded in different $MgCl_2$ concentrations. M13mp18 (single-stranded) scaffold is shown in lanes 1 and 11 (labelled S). Lanes 2 to 10 show wedge origami folded in TE buffer + 5, 9, 11, 13, 15, 17, 19, 21, 25, 29 mM $MgCl_2$

Characterisation by cryo-EM typically requires high (μM) particle concentrations. Although annealing at these concentrations is not feasible due to component stock concentrations, increasing folding concentration is desirable to reduce the scale of post-purification concentrating required. I have screened scaffold concentration for origami folding between 5-40 nM (Figure 2.3). With the exception of a small increase in the fraction of origami dimer observed for concentrations above 5 nM, there does not appear to be any change in folding behaviour over this range, so a final scaffold concentration of 40 nM was selected for annealing.

AFM

As a first attempt to visually characterise the wedge-origami, nanostructures were imaged by atomic force microscopy (AFM). AFM is commonly used to image 2D DNA origami objects, but the technique can also be used to image 3D origami at the expense of reduced image quality. Wedge-origami were folded using the conditions outlined above (scaffold concentration 40 nM, $MgCl_2$ concentration 12.5 mM), and purified by spin column filtration[4]

[3] 10 mM EDTA-Na_2·$2H_20$, 100 mM Tris base, pH 8.0, purchased at 100× from ITW reagents.
[4] 3× 15 minute spin with 100 kDA 0.5 mL Amicon filters.

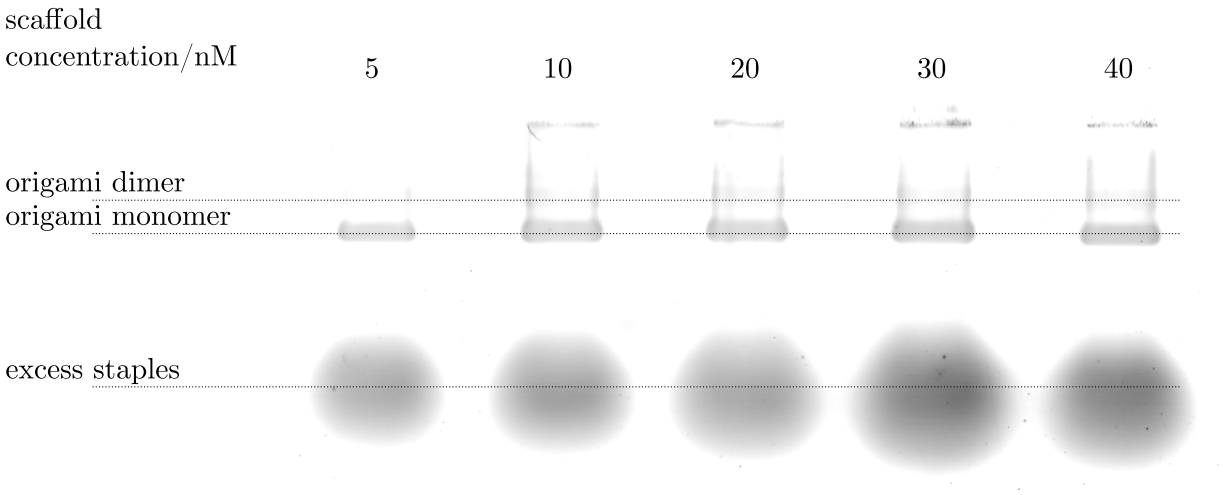


Figure 2.3: Agarose gel electrophoresis of wedge origami folded at different scaffold concentrations (5, 10, 20, 30 and 40 nM).

and imaged in fluid[5] by AFM in tapping mode (Appendix D.2). Results are shown in Figure 2.4. The sample is mono-disperse and structures appear to be well-folded. Front-, rear- and side-views can often be distinguished but high-resolution details of the structure are not readily visible.

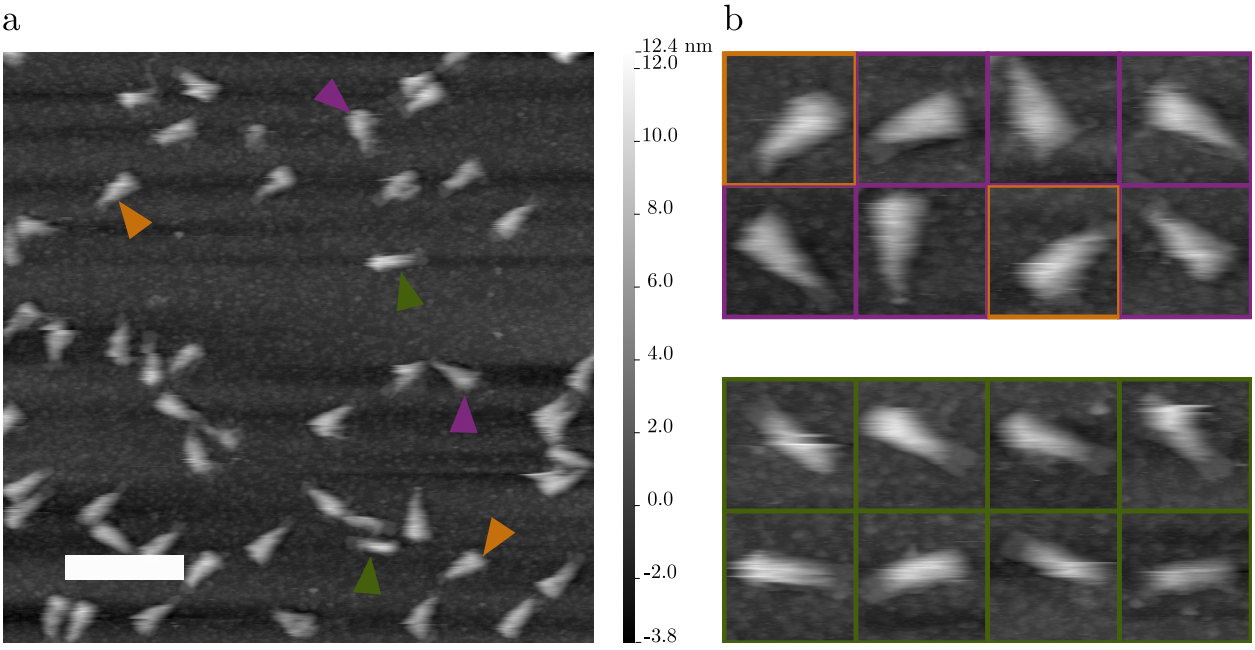


Figure 2.4: Height contrast AFM images of wedge origami nanostructures. (a) Characteristic wide-field view of wedge-shaped origami nanostructures imaged by AFM (scale bar 200 nm). Front-, rear- and side-views are indicated by purple, orange, and greem arrowheads respectively. (b) Single particle images of front-/rear-views (top) and side-views (bottom) of the nanostructures. Squares are 115 nm× 115 nm.

[5]Origami folding buffer: 1×TE, 12.5 mM MgCl2, pH 8.0 (no added Nickel).

Negative-Stain TEM

Negative-stain transmission electron microscopy (TEM) was also used to image wedge origami nanostructures. Samples were deposited onto carbon-coated copper grids (Agar scientific) and stained in 2% uranyl acetate (Appendix D.1.1). The staining process can generate regions of positive and negative staining within the same grid, but this protocol preferentially results in negative staining. Compared to positive staining, negatively stained regions typically provide higher resolution views and allow visualisation of the internal origami structure. Typical micrographs are shown in Figure 2.5.

As expected, TEM provides a much higher resolution view of the nanostructures than AFM. Different views can be reliably distinguished and higher resolution details of the structure are visible. These images confirm that the nanostructures appear as designed and are well-folded at close to 100% yield.

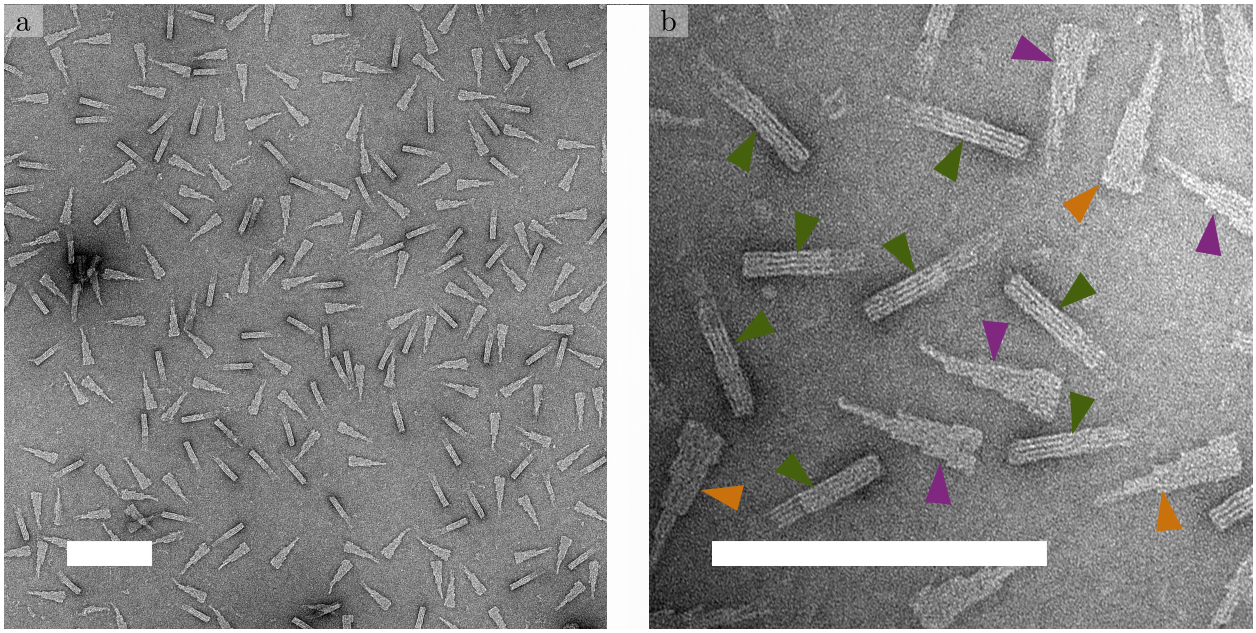


Figure 2.5: Characteristic wide-field (a) and high-magnification (b) views of wedge-origami imaged by negative stain TEM. In (b) front-, rear- and side-views are indicated by purple, orange, and green arrowheads respectively. Scale bars 200 nm.

Preliminary Cryo-EM

Preliminary cryo-EM of wedge origami nanostructures was performed by Susan Lea (Figure 2.6). A 2 mL volume of origami was purified by spin column filtration, and condensed to a small ($\approx$ 20 μL)volume. The sample concentration was measured by A_{260} absorption as $\approx$ 1 μM, but the cryo-EM results show relatively low particle number and a large background of dark spots corresponding to excess staples. This indicates that the measured concentration results primarily from staple excess. Spin column filtration is therefore not an effective purification/concentration method for high concentration nanostructures, so it was necessary to evaluate new methods.

Despite these purification problems, nanostructures can be distinguished at high resolution. Unlike by negative stain TEM or AFM, where origami appear as solid objects, by cryo-EM the internal structure is visible, with the helical grain much easier to distinguish than the origami outline.

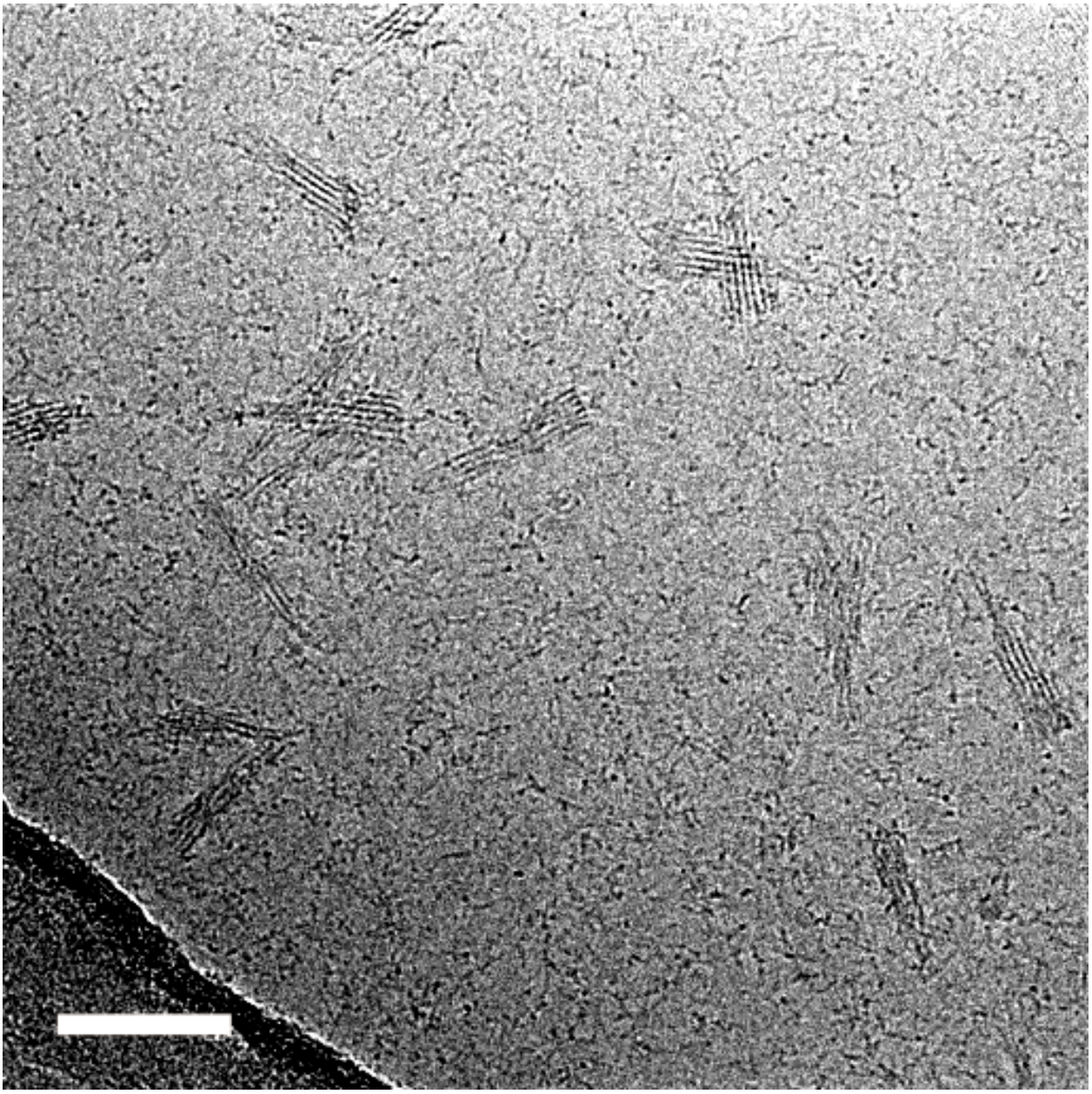


Figure 2.6: Cryo-EM image of wedge-shaped origami nanostructures in buffer. Dark background spots are excess staples (scale bar 100 nm).

2.3.3 Purifying Origami for Cryo-EM

High particle concentrations are usually preferred for CET. This will be particularly important for tags as they will be used to supplement a target system, and must be present concentrations comparable to, or greater than, the target for high efficiency labelling. The

purity of the sample is also paramount as some impurities (e.g. ethanol) can affect freezing, and others (e.g. excess staples) can increase background, compromising the signal-to-noise ratio.

Precipitation by poly(ethylene glycol) (PEG) crowding has been used for decades in the purification of DNA plasmids and phage, and has been more recently optimised for purifying DNA origami [125]. This method gives excellent separation of intact origami from non-integrated staple strands, which is important for minimising background, and will be even more essential when strands with targeting properties are incorporated (since excess targeting strands will block binding sites). PEG precipitation also allows sample concentrations up to the limit of solubility.

DNA origami samples purified by PEG precipitation were prepared at concentrations of ≈ 1 μM for cryo-EM (Appendix C.2.2). Imaging problems were encountered with these samples, attributed to excess PEG, which is a cryo-protectant. Non-vitreous 'stringy' ice is commonly observed, and ice is found to melt at low electron doses (<10 $e^-/Å^2$). An example wide-field micrograph is shown in Figure 2.7. Based on these results, it was determined that PEG should be thoroughly removed from samples before imaging by cryo-EM.

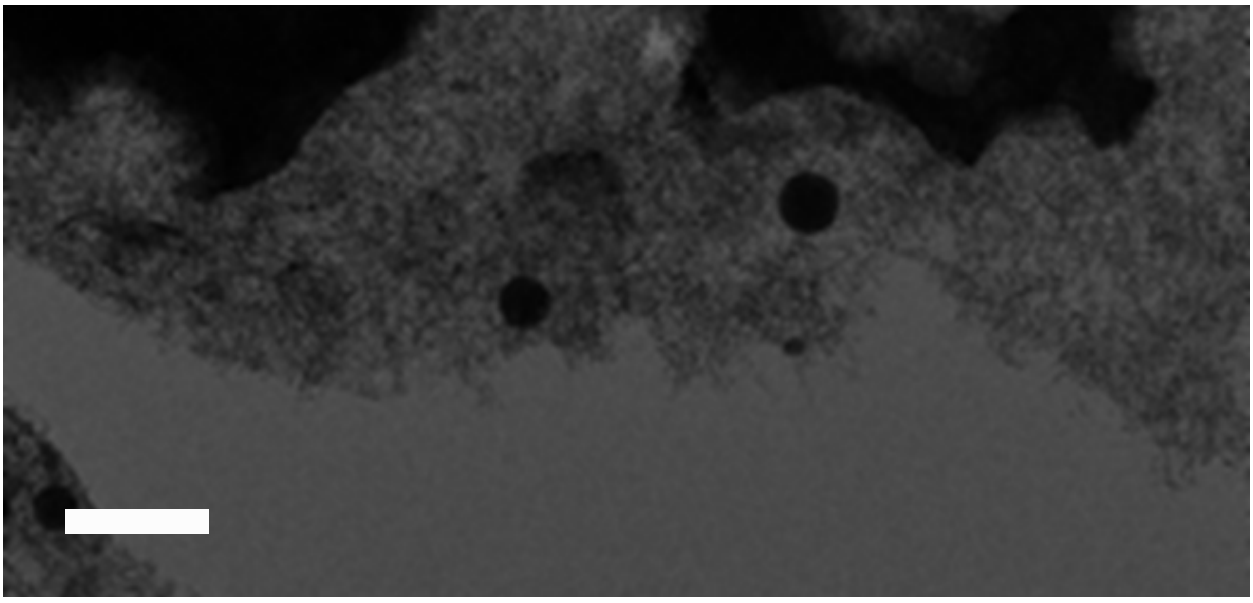


Figure 2.7: Example of cryo-EM imaging problems attributed to excess PEG. At the top, unusual 'stringy' ice is seen. The empty section below is where ice has melted at very low dose. Scale bar 1 μm.

A common method for removing excess PEG from samples is ethanol precipitation. Ethanol precipitation of DNA is very standard in cloning, but is not normally used with DNA nanostructures. Ethanol precipitation was performed on a sample of unpurified 40 nM DNA origami (Appendix C.2.3), and structures were evaluated by AGE (Figure 2.8a) and negative-stain TEM (Figure 2.8b and c). Perhaps surprisingly (as there is no Mg^{2+} in the ethanol precipitation mix), the structural integrity of the origami is preserved. Staple strands are also precipitated with the origami, so this is not a complete purification technique, but can be used to concentrate origami and remove PEG contamination.

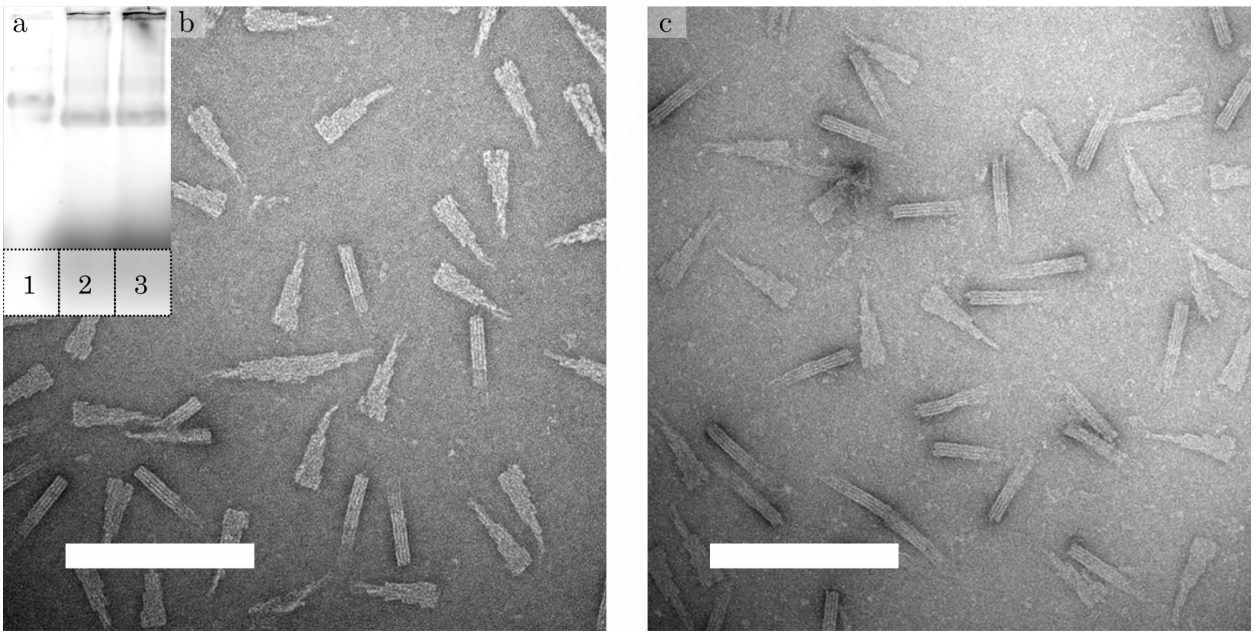


Figure 2.8: Overview of DNA origami ethanol precipitation. (a) Agarose gel electrophoresis of M13mp18 single-stranded scaffold for reference (lane 1), unpurified origami (lane 2), and origami after EtOH precipitation (lane 3). (b) and (c) Negative stain TEM of wedge origami before (b) and after (c) ethanol precipitation. Nanostructures remain well-folded. Scale bars 200 nm.

Origami prepared by PEG precipitation followed by ethanol precipitation was no easier to image by cryo-EM than samples prepared by PEG precipitation only, likely because ethanol is also a cryo-protectant. Purification methods which use ethanol precipitation as an intermediate step may, however, be suitable. Qiagen PCR purification kits were used for this purpose, and it was seen by agarose gel electrophoresis (Figure 2.9) that origami remained well-folded through the purification process providing buffers were supplemented with 5 mM $MgCl_2$ (higher concentrations also work, but compromise yield). Origami concentrations of up to 2-3 μM can be recovered with this method, and samples are clean enough for imaging by cryo-EM.

In experiments described below the purification method used is a 2-step protocol: PEG precipitation followed by column purification (Appendix C.2.2 and C.2.4).

2.3.4 Cryo-EM in Cell Lysate

DNA origami were purified for cryo-EM, and the concentration was measured by A_{260} absorption as 2.1 μM. Lindsay Baker prepared and imaged cryo-EM grids of origami samples mixed 1:1 with bacterial cell lysate. Cryo-EM methods are presented in Appendix D.3. Example micrographs are shown in Figure 2.10.

Origami can be seen with a similar visibility to membrane structures, which are typically the highest contrast objects in cellular micrographs. The 'stripes' of the helical grain do not resemble any components of cellular debris, making the nanostructures distinctive in

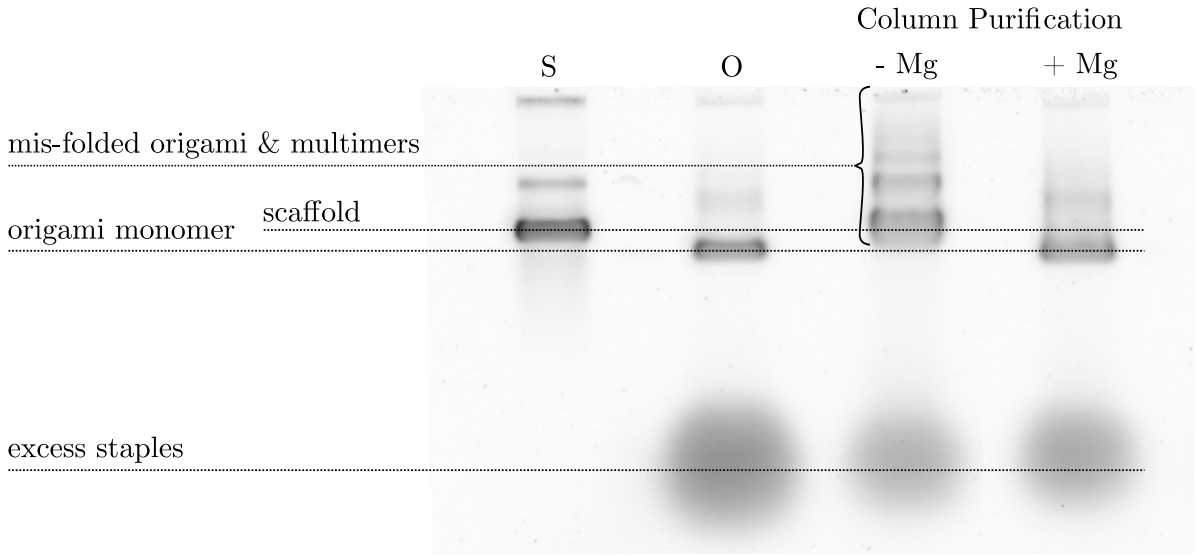


Figure 2.9: Agarose gel electropheris of DNA origami clean-up. For lanes 1-4: (1) scaffold (S), (2) wedge origami, unpurified (O), (3) origami after passing through unmodified Qiagen PCR purification kit, and (4) origami after passing through Qiagen PCR purification kit with all buffers supplemented with 5 mM $MgCl_2$. Purification removes $\approx 60\%$ of free staples.

this environment. In solution, nanostructures may assume any 3D orientation, but most views appear similar to front-/rear-/side-views observed in negative stain TEM. Examples of front- and rear-views are shown in the inset in Figure 2.10, and the degree of asymmetry is usually sufficient to distinguish the two views (although a more asymmetrical face would make analysis easier). End-views of the structure, where present, are the highest contrast components of the image.

Based on this data, I conclude that honeycomb-lattice origami nanostructures should be visible in a busy, biological context, and therefore provide an appropriate architecture for CET tags.

2.4 Designing Nanostructures for Multiplexing

2.4.1 Hollow Origami Nanostructures

Synthesis and Characterisation

One of the leading merits of DNA origami nanostructures as tags compared to other available systems is the possibility of multiplexing, i.e. simultaneous labelling of multiple targets with distinguishable tags. As described in Section 1.3.3, there are many different architectures that can be built with DNA origami, so new origami types, as well as different honeycomb lattice origami designs, can be investigated as options for multiplexing.

Separately to my work, Céline Journot (a previous lab member) designed a hollow

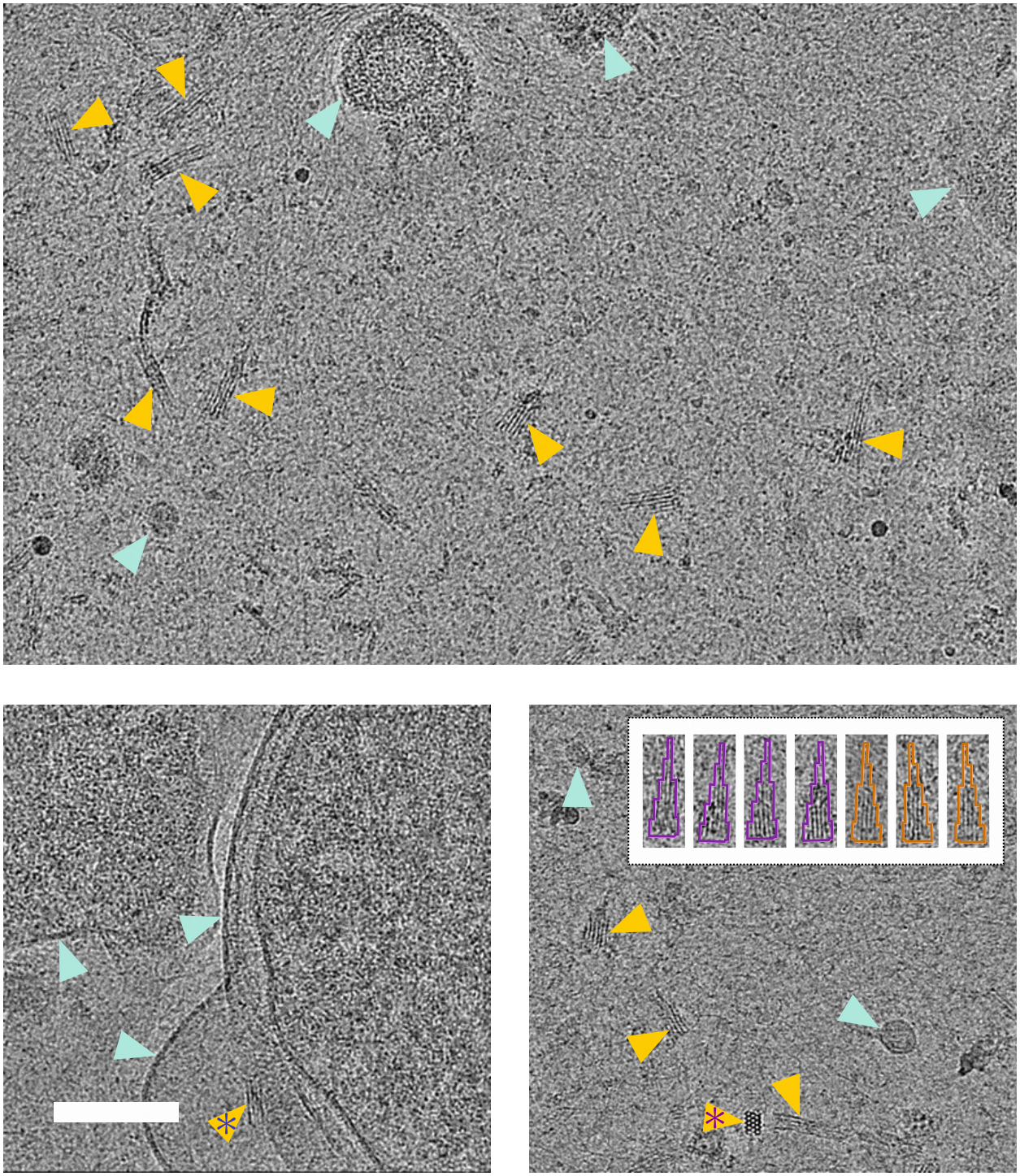


Figure 2.10: Cryo-EM projection images of wedge-shaped origami nanostructures mixed 1:1 with bacterial cell lysate (scale bar 100 nm). Yellow arrowheads point to different views of nanostructures, while cyan arrowheads point to vesicles or other membrane structures. Most origami views are visible with similar contrast to membrane structures, and can be visualised even through membranes (blue asterisk). End-views (purple asterisk) are the highest contrast objects in the micrographs. The inset in the bottom right view shows example front- (purple outline) and rear-views (orange outline) of the nanostructure.

tetrahedron nanostructure for cryo-EM where each arm of the tetrahedron is a 6-helix bundle. An overview of the design is given in Figure 2.11 [126]. The symmetry of a tetrahedron is a problem for tagging, since the position of a tagging module would not be uniquely deducible, but an assessment of the design architecture with regards to multiplexing is useful nonetheless. This open (wireframe-like) architecture is very different to the solid design presented in the previous section, and should be easily distinguishable from the wedge for cryo-EM, providing the contrast is sufficient.

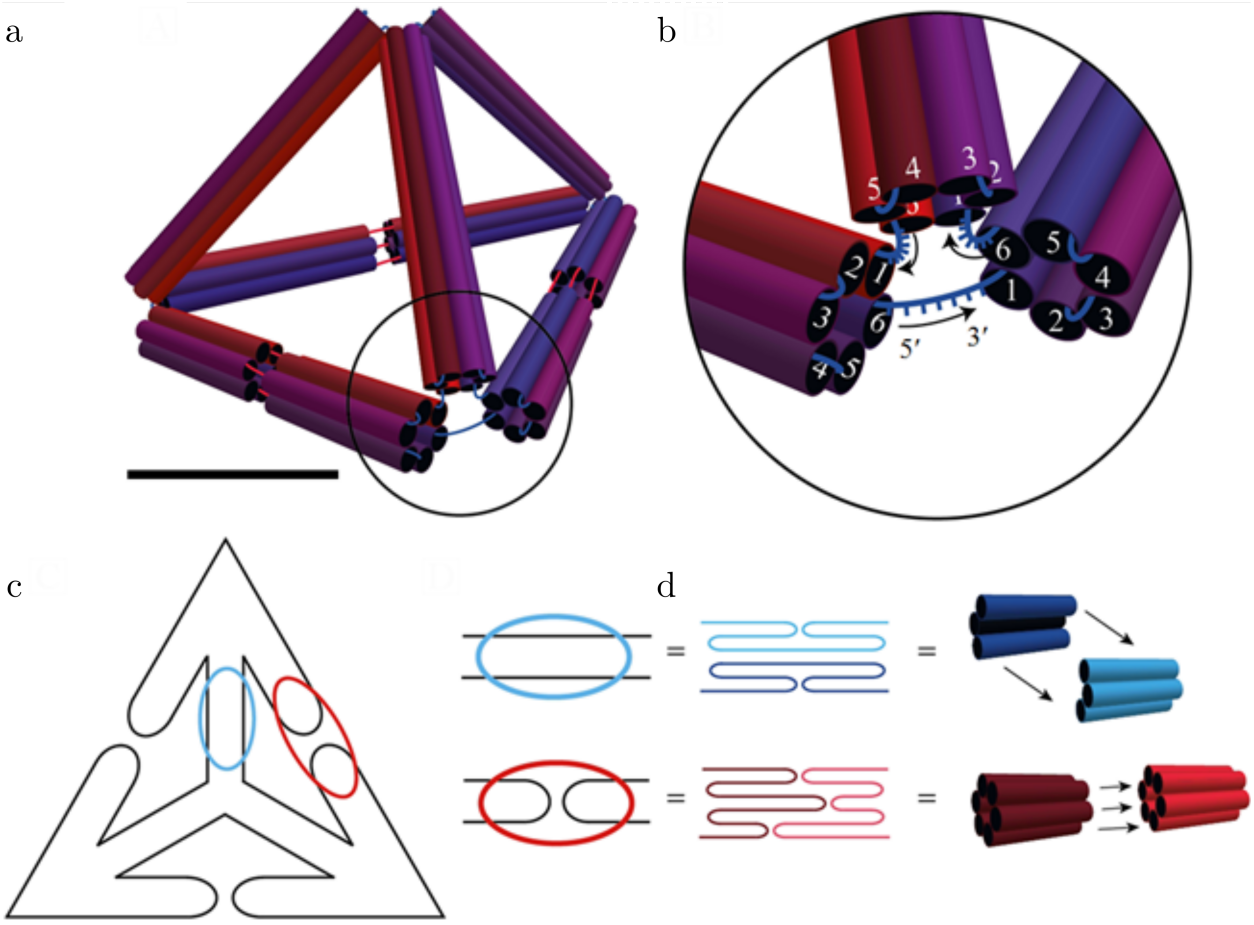


Figure 2.11: Image and caption adapted from Journot, 2017 [126]. Schematics of DNA tetrahedron. (a) 3D view of the entire structure. Each DNA double helix is shown as a tube, and the edges are 6-helix bundles. Red lines represent staples binding two helix-bundles together and blue scaffold strands. Scale bar approximately 20 nm. (b) Details of the summit vertex corresponding to the black circle in (a). The scaffold routing and unpaired bases are showed in blue. A black arrow indicates the direction of the DNA; the 3'-end located at the head. The staples are not drawn for clarity. (c) Schematics of the general routing used for the construction of the tetrahedron. (d) detail of (top) lengthwise and (bottom) crosswise type of edges. Each line of the schemes on the right corresponds to the scaffold routing, which is maintained by the hybridisation of staples (not shown for simplicity).

This nanostructure was designed using the plasmid pUC19 as a scaffold strand. The pUC19 was supplied in double-stranded form[6] and treated with restriction enzymes to generate the single-stranded scaffold. The tetrahedron origami was annealed in a buffer

[6]pUC19 supplied by New England Biolabs at 1,000 µg/ml.

of 1× TE with 10 mM $MgCl_2$ and a 10-fold excess of stables (conditions determined by Celine Journot [126]). The origami was folded at a scaffold concentration of 40 nM which was found to result in a sizeable fraction of multimers. Based on Celine's previous work, it is known that these multimers are of well-defined geometries which, although not useful for a tag, are interesting to image by cryo-EM and can provide a good assessment of whether different geometries of wireframe-like origami objects may be used for multiplexing.

Correct folding was verified by AGE (Figure 2.12a) and negative-stain TEM (Figure 2.12b, c). A mobility shift relative to single-stranded scaffold was seen by gel, as well as a ladder of higher molecular weight bands corresponding to multimers. By TEM structures are easily visualised, and well-folded monomers and multimers are observed.

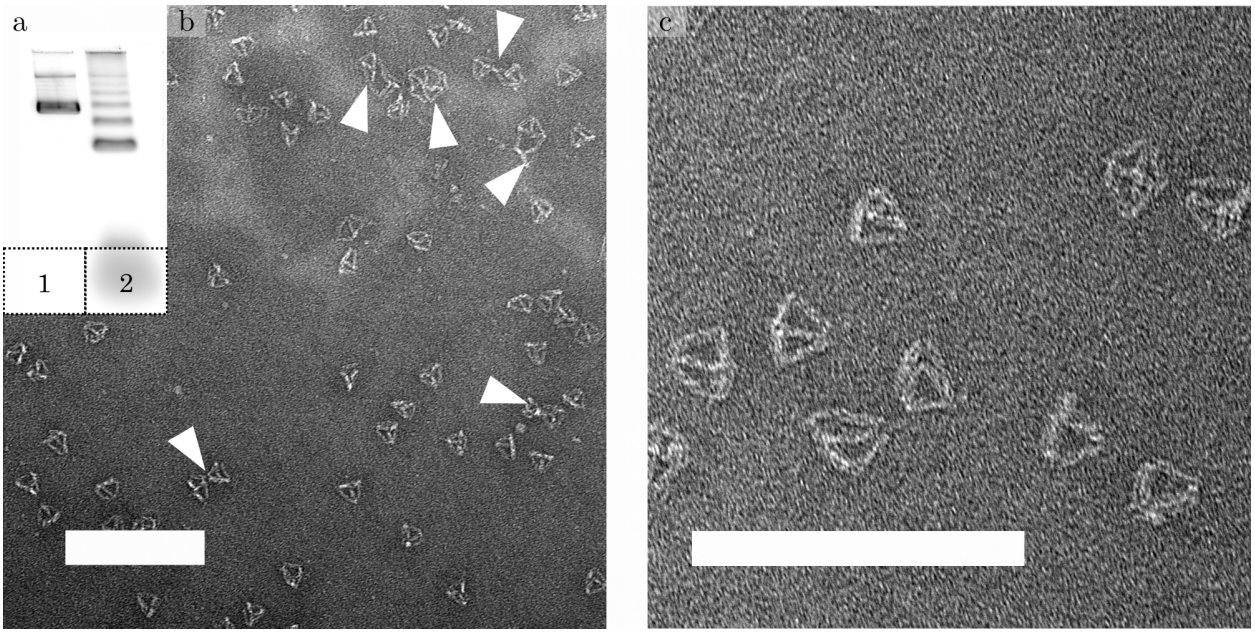


Figure 2.12: (a) Gel electrophoresis of single-stranded puc19 scaffold strand (lane 1) and folded tetrahedron origami, unpurified (lane 2). A mobility shift is observed, indicative of folding. A ladder of higher molecular weight bands is also oberved, corresponding to multimers. (b) Characteristic wide-field and (c) high-magnification views of tetrahedron origami nanostructures imaged by negative stain TEM. Examples of multimers are indicated by white arrowheads, scale bars 200 nm.

Cryo-EM in Cell Lysate

Tetrahedron origami were purified for cryo-EM, and the concentration was measured as 1.9 μM from A_{260} absorption. Nanostructures were mixed 1:1 with mammalian cell lysate and cryo-EM was performed with the help of Lindsay Baker. The mammalian cell lysate is significantly denser than the bacterial cell lysate used for imaging in Figure 2.10, and provides a greater challenge for identifying structures. Results from typical micrographs are depicted in Figure 2.13.

Images were successfully taken of these structures, but the contrast is extremely poor, and structures were only visible in thin ice sections. Even in these regions an electron dose

of >30 $e^-/Å^2$ is required for visualisation. High dose can chemically damage samples, and results in surface charge accumulation which, in turn, deflects incoming electrons and results in imaging artefacts (lower panel of Figure 2.13). The reduced imaging contrast of these structures compared to the honeycomb-lattice wedge origami can be explained by two key structural differences: density projection and regularity.

The wedge nanostructure has a solid architecture and the lowest density projection corresponds to the front/rear view of the structure, where the electron beam passes through 4-helix-thick path. The tetrahedron, however, has an open architecture. Except in special cases where edges of the structure align in a given view, the contrast of the structure depends only on the contrast of the six-helix-bundle components, which have a thickness of only two helices. Likewise, the helices are ≈27 nm long (compared to the wedge length of ≈ 80 nm), so projections with end-on views of the tetrahedron edges will display much lower contrast than the wedge end-views.

The regularity of lattice-based structures means that, from a given projection, there are regions where high contrast elements (helices) align, and regions where low contrast elements (gaps between helices) align. This provides contrast enhancement by way of the 'stripy' effect seen in Figures 2.6 and 2.10. The tetrahedron architecture does not have this feature and, as a result, structures are more difficult to pick out in a noisy micrograph.

Based on these results, it is not recommended that this architecture be used for CET tags.

2.4.2 Multiplexing Honeycomb Origami Nanostructures

Design of a 'Pointing-Hand' Origami

In light of the difficulties encountered imaging the hollow tetrahedron nanostructure, different solid designs were also tested for multiplexing. To test the discernibility of honeycomb lattice origami nanostructures with different global architectures, I designed a 'pointing-hand' origami for comparison with the wedge presented in the previous section. This design comprises 70 helices arranged in a honeycomb lattice, with a highly asymmetric top view, and approximately rectangular side-/end-views which resemble those of the wedge lattice (but with different dimensions). Scaffold routing and staple sequences were designed using caDNAno software for an M13mp19 scaffold [62], and 3D models of the structure were generated using CanDo [69]. Depictions of alternate views of the designed structure are shown in Figure 2.14, and the layout of crossovers and staple sequences are provided in Appendix A.3.2.

Synthesis and Characterisation

Since this origami is structurally similar to the wedge presented previously, I have re-used the folding conditions instead of re-optimising salt and scaffold concentrations. Correct folding was verified by agarose gel electrophoresis (Figure 2.15a) and negative stain TEM (Figure

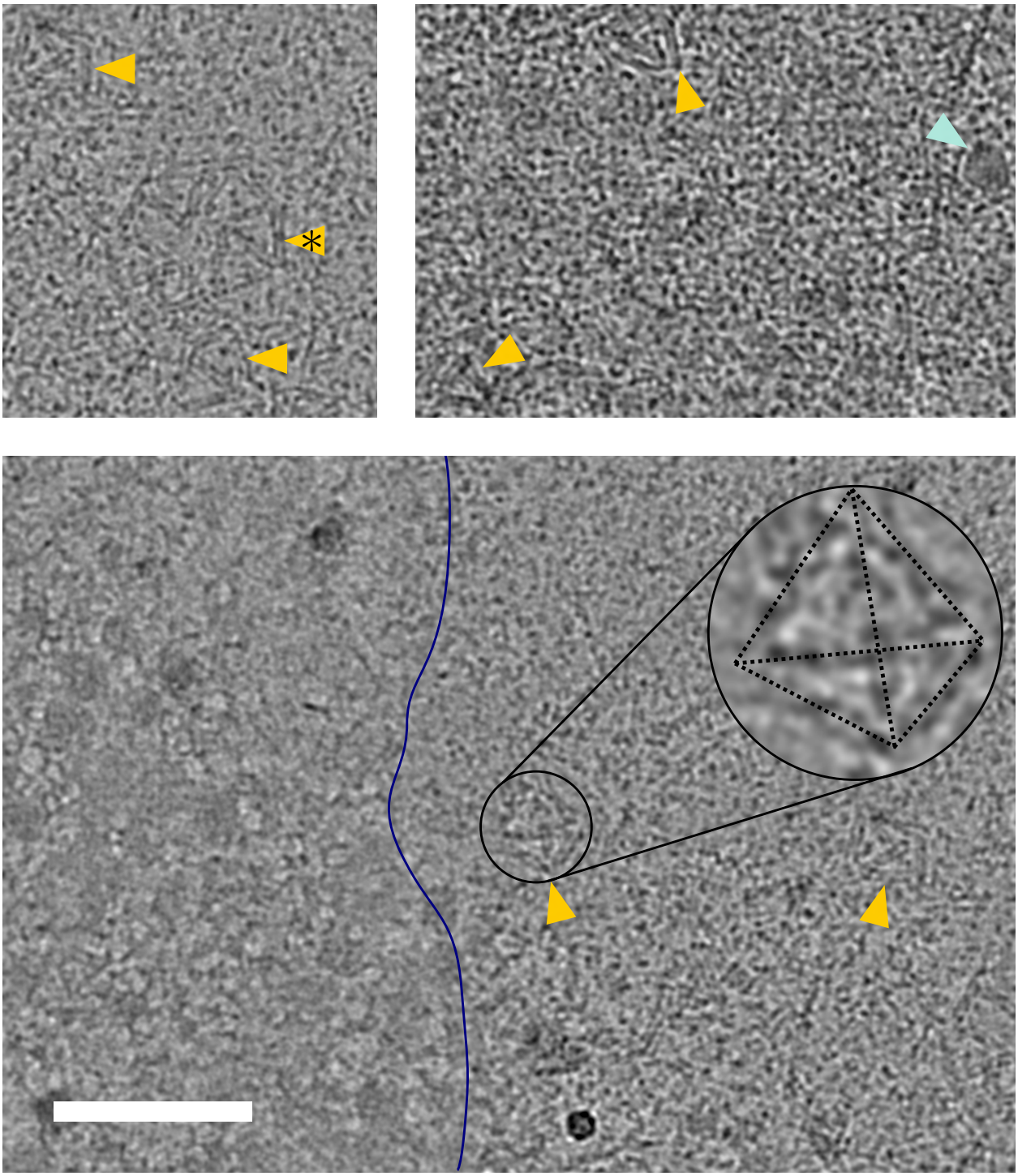


Figure 2.13: Cryo-EM projection images of tetrahedron origami nanostructures mixed with mammalian cell lysate imaged by cryo-EM, scale bar 100 nm. Nanostructures are indicated by yellow arrowheads, including one multimer (black asterisk) and a membrane structure is indicated by a cyan arrowhead. The inset in the lower panel shows a magnified version of a single tetrahedron, overlayed by an outline of the expected structure. High doses result in imaging artefacts, as seen to the left of the solid blue line .

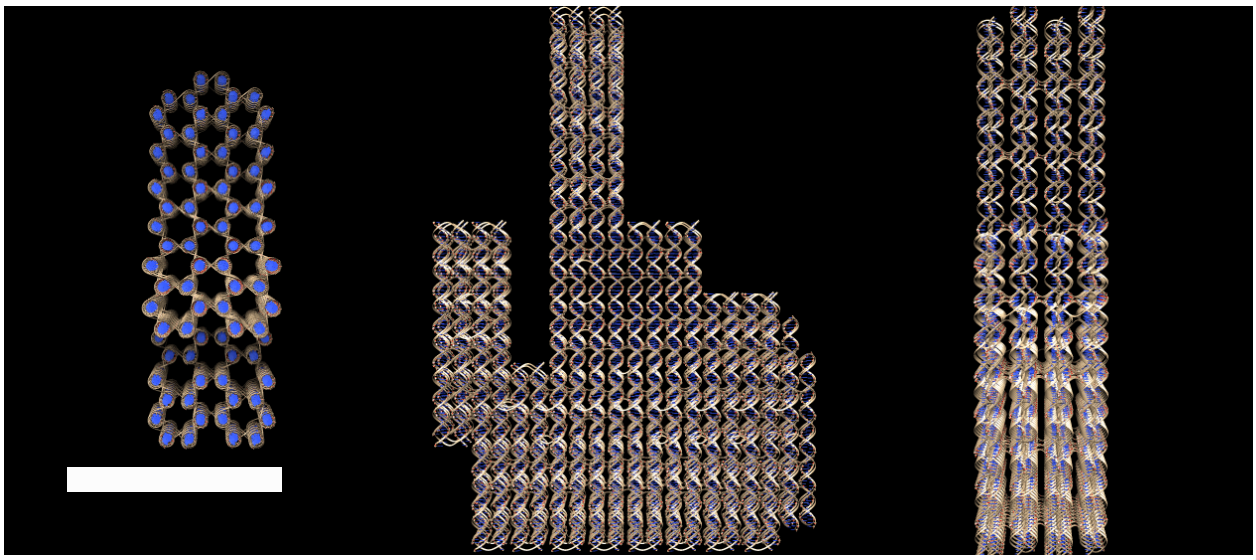


Figure 2.14: Orthogonal views of designed pointing hand origami nanostructure as generated by Cando (scale bar 20 nm). End-, top- and side-views correspond to the left, centre and right panels shown.

2.15b, c). A mobility shift relative to single-stranded scaffold was seen by gel. By TEM, structures are easily visualised and are well folded at close to 100% yield.

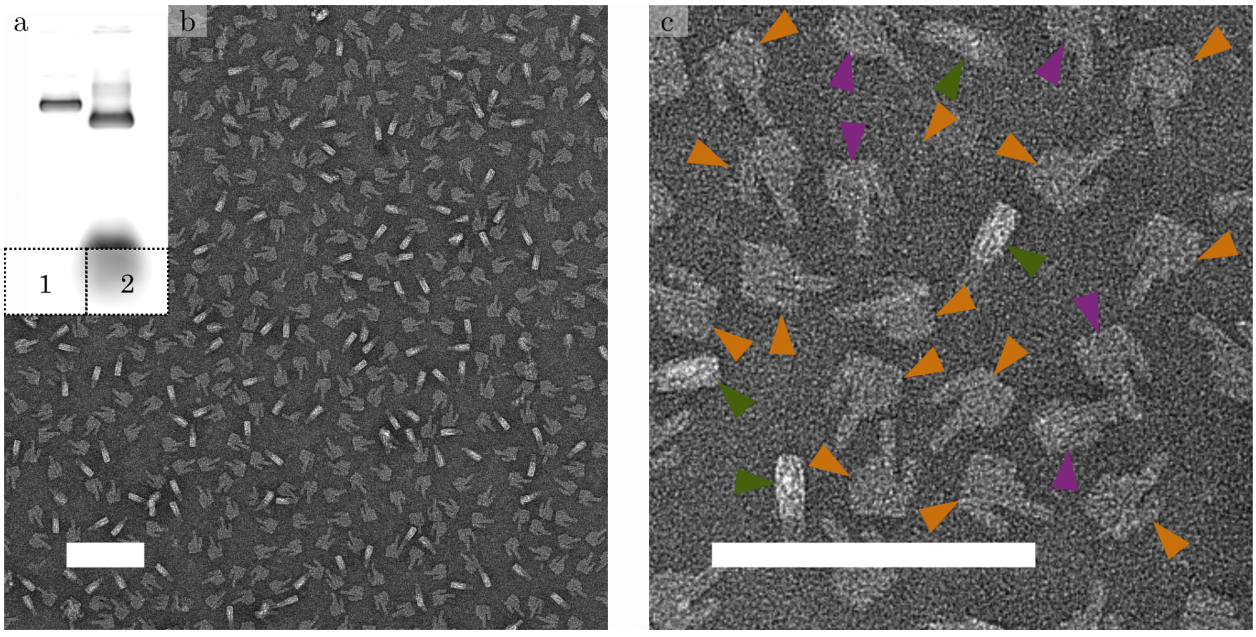


Figure 2.15: (a) Gel electrophoresis of single-stranded M13mp18 scaffold strand (lane 1) and folded pointing-hand origami, unpurified (lane 2). A mobility shift is observed, indicative of folding. (b) Characteristic wide-field and (c) high-magnification views of pointing-hand origami nanostructures imaged by negative stain TEM. In (c) front-, rear- and side-views are indicated by purple, orange, and green arrowheads respectively. Scale bars 200 nm.

Cryo-EM in Cell Lysate

To compare the wedge and pointing hand, both structures were purified for cryo-EM, and concentrations adjusted to 1 μM. Wedge origami, pointing-hand origami and mammalian cell lysate were mixed in a 1:1:2 ratio, and cryo-EM was performed by Lindsay Baker. A typical micrograph is shown in Figure 2.16. Once again, the mammalian cell lysate is thicker than the bacterial cell lysate used to image the wedge origami.

In many cases, the two nanostructures can be distinguished. Projections similar to side- or end-views, although fundamentally alike for the two designs, are generally easier to distinguish than front/rear views. Wedge side-views are longer and lower contrast than those of the pointing-hand, whereas wedge end-views are shorter and higher contrast than pointing-hand end-views. Projections similar to front/rear-views are visible with the lowest contrast, since this corresponds to the electron beam passing through the thinnest part of the structure. In this environment these views are close to the limit of visibility by eye, with the full face of the structure usually not distinguishable.

This data shows that, in principle, multiple honeycomb lattice origami nanostructures can be used for multiplexing, but also points towards some design constraints:

- A thickness of four helices should be an absolute minimum limit for the dimensions of a tag in any one direction. Highest contrast will likely be achieved for nanostructures with similar extent in x, y and z dimensions.
- Most views are similar to the orthogonal end-/front-/side-views seen by negative stain TEM. Multiplexing nanostructures should be distinguishable from any projection for distinction in micrographs.
- The internal structure of the nanostructures is visible, so multiplexing would be greatly facilitated by using different lattice types or more complex arrangements of helices.

Cryo-electron Tomography

The intended use for these nanostructures is as tags for CET, so tomography of the wedge/hand/lysate mix was carried out with the help of Lindsay Baker (who also reconstructed the tomogram) to verify that the structures can still be distinguished in 3D. The sample thickness is ≈150 nm, which is a reasonable approximation to the thickness of a cell edge or a FIB (focus ion beam) milled section.

Example tomogram slices are shown in Figure 2.17. The sample is very busy, which makes it more difficult to see which origami are present, but the contrast is generally good: higher than for ribosomes, but a little lower than for membrane structures. From individual tomogram slices it is sometimes difficult to distinguish origami types, partly because contrast variation (which is helpful in distinguishing side-views) is no longer a factor. Structures can be reliably distinguished using the full 3D tomogram data, but identification from individual slices would be improved with the design constraints outlined previously.

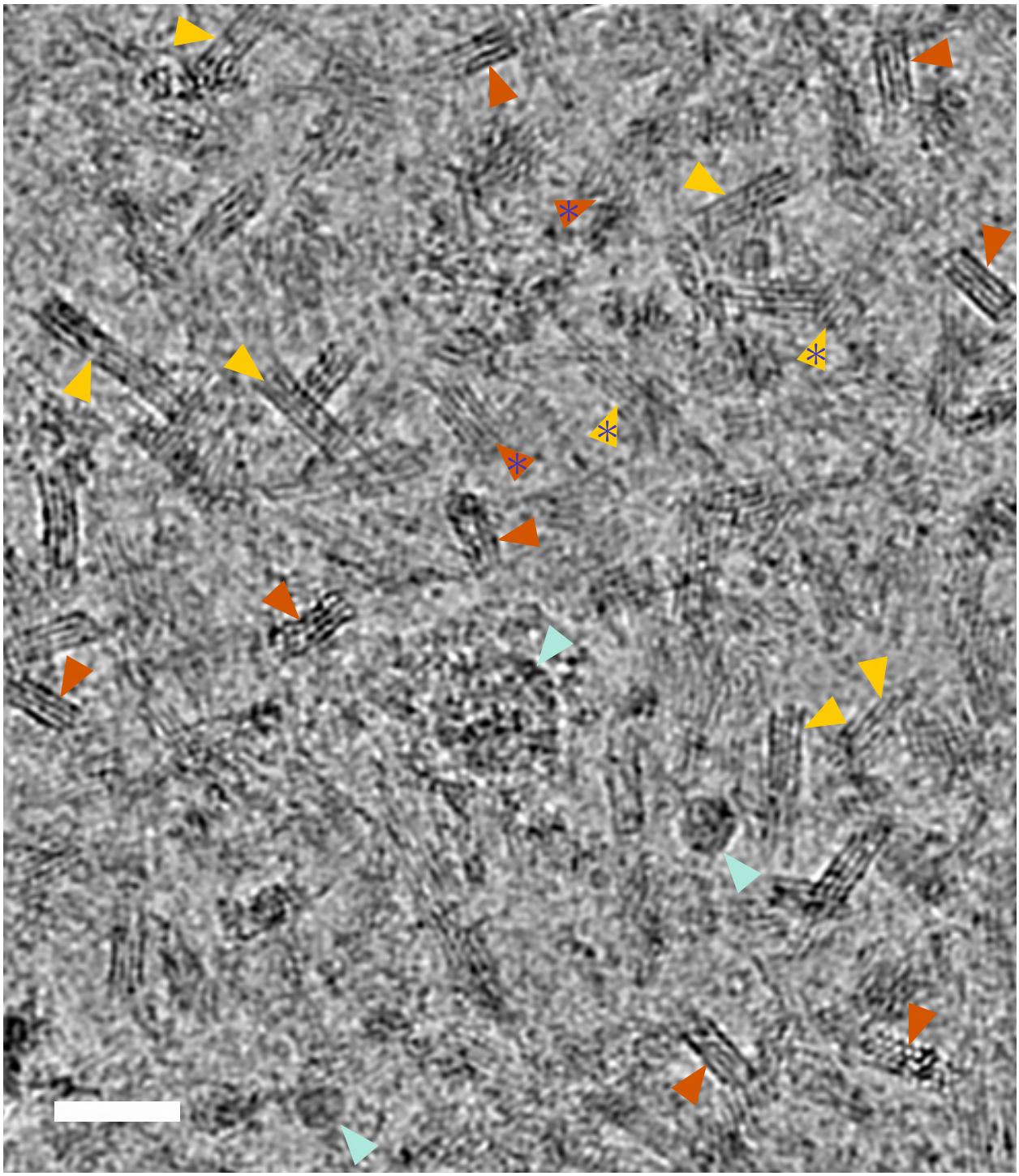


Figure 2.16: Characteristic micrograph of wedge (yellow arrowheads) and hand (orange arrowheads) in mammalian cell lysate. Membrane bound structures are indicated with cyan arrowheads. Views which correspond approximately to front-views (blue asterisks) are close to the visibility limit by eye. Scale bar 50 nm.

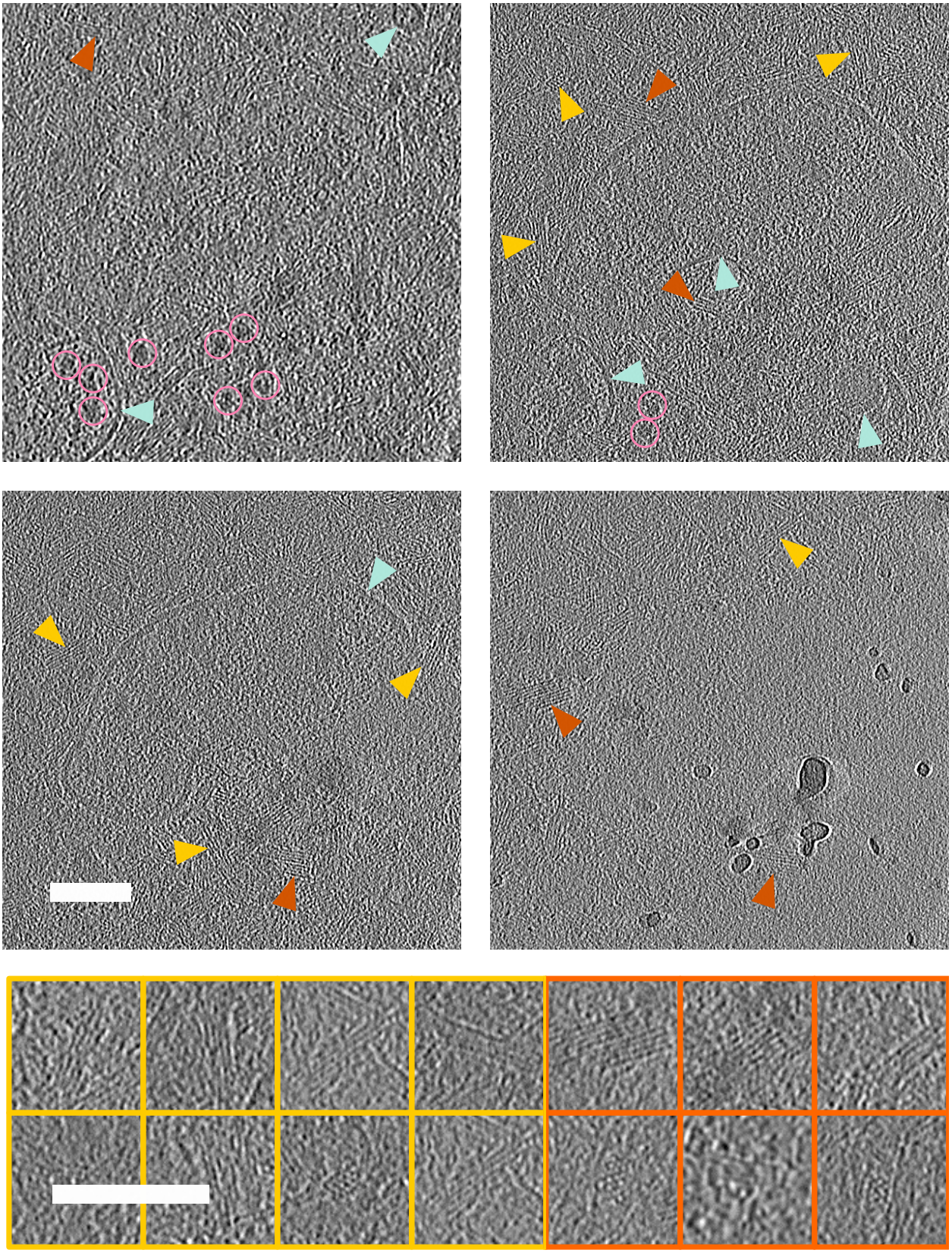


Figure 2.17: Cryo-electron tomography of wedge and pointing-hand origami mixed with mammalian cell lysate. Above: slices from a single tomogram. Orange and yellow arrowheads point towards examples of hand and wedge nanostructures respectively. Membrane structures are shown with turquoise arrowheads, and examples of ribosomes are indicated by pink circles. Below: nanostructure views of wedge (yellow outline) and hand (orange outline) nanostructures. Scale bars 100 nm.

2.5 Signpost Origami: Design, Synthesis and Characterisation

2.5.1 Design

Preliminary cryo-EM results have shown that the solid lattice-based origami nanostructures are visible with high contrast in cell lysate, whereas wireframe-like nanostructures are not. Design constraints have been identified for visualising nanostructures, but for tagging applications further constraints must be considered. Firstly, the protein targets will be significantly smaller than the overall origami size, of order 100-fold in terms of molecular weight. Bulky origami architectures are therefore at risk of concealing the target. Secondly, depending on the target system features (in particular the separation between target molecules), binding efficiency of tags may be limited by nanostructure crowding. To address these constraints, whilst maintaining the high contrast observed for the wedge and pointing-hand nanostructures, I introduce a 'signpost' origami design.

A schematic overview of this design is shown in Figure 2.18a. The structure consists of a bulky, asymmetric 'sign', used for identification purposes, and a six-helix bundle 'post'. A nucleic acid targeting module, which may be a chemically modified oligonucleotide, an aptamer, or the binding partner of a biological DNA/RNA-binding protein, can be attached by hybridisation at the end of the post. The spatial separation of the sign from the target molecule is intended to minimise the risk of concealing the target and to reduce crowding effects (as illustrated in Figure 2.18b).

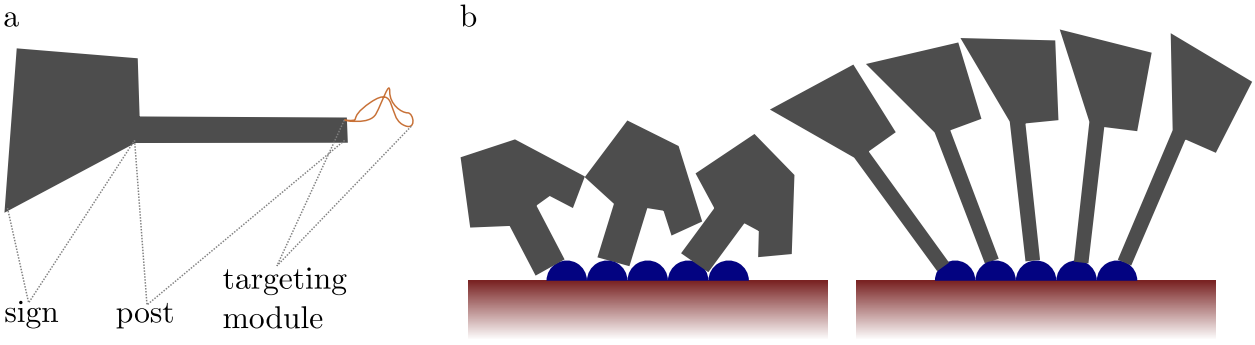


Figure 2.18: Concept design overview for signpost DNA origami nanostructures. (a) Schematic of signpost nanostructure design. The nanostructures consists of a bulky sign and six-helix bundle post, conjugated to a targeting module. (b) Illustration of signpost crowding behaviour for signpost structure (right) compared to pointing hand (left). The separation of the origami bulk from the target allows higher labelling efficiency for the signpost structure.

Cryo-EM imaging of the hollow tetrahedron nanostructure revealed that 6-helix bundles are difficult to visualise in cell lysate. It is therefore important that the position of the targeting module can be deduced from imaging the sign portion of the origami, i.e. flexibility of the post must not introduce a significant uncertainty to the aptamer position. The persistence length of a 6-helix bundle has been measured as approximately 2.4 μm [127]. Note that this is likely an underestimate, since this value was determined from negative-

stain TEM measurements, in which nanostructures are dried under high salt concentrations, and the persistence length of DNA is known to decrease with increasing ionic strength [128].

Using a worm-like chain (WLC) model I have simulated trajectories of 6-helix bundles with a persistence length of 2.4 μm, from which I measured the mean distance $\langle \Delta r \rangle$ between the end position of the simulated bundles and a completely stiff bundle of the same length. $\langle \Delta r \rangle$ values were determined as a function of the number of base pairs per helix. The results of these simulations are shown in Figure 2.19.[7] Typically the distance $\langle \Delta r \rangle$, should be $\leq 3-4$ nm, which is a realistic lower limit for the size of the targeting module.

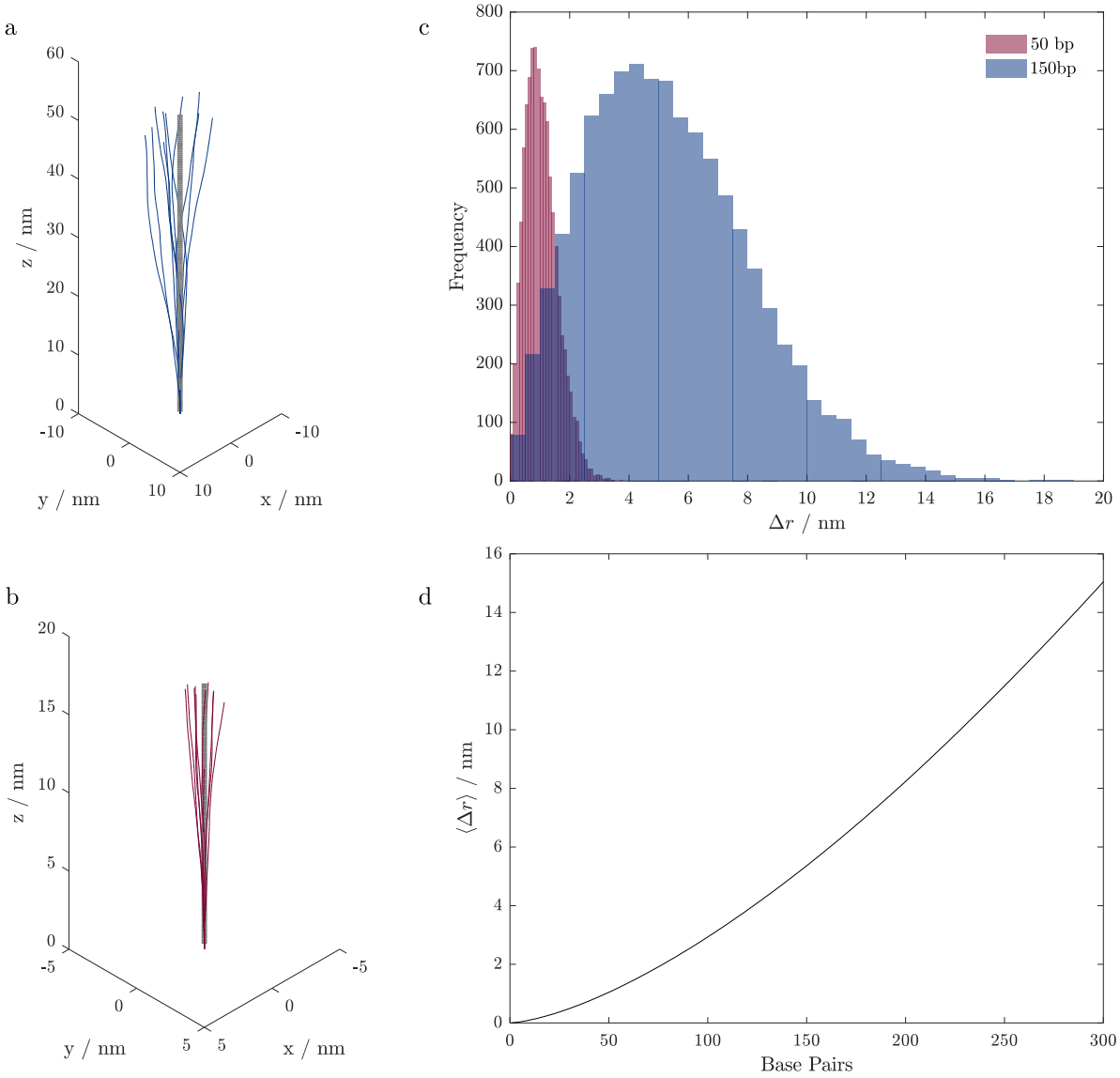


Figure 2.19: WLC model for 6-helix bundle. (a)/(b) 10 simulated traces of 150 bp/50 bp 6-helix bundles (blue)/(red) compared to a perfectly stiff 6-helix bundle of the same length (grey). (c) Histogram of the end position error Δr introduced by bending for 10,000 150 bp and 50 bp six-helix bundles. (d) Plot of the mean end position error ⟨Δr⟩ as a function of number of base pairs.

[7]The MATLAB script used for these simulations is provided in Appendix F.1.

Using this post length requirement, and considering the design constraints identified previously, I designed a signpost DNA origami for application as a CET tag. A post length 93 bp, corresponding to a $\langle \Delta r \rangle$ of 2.6 nm, has been used (for future tagging applications it may be advantageous to optimise this number depending on the system, e.g. more crowded systems will favour longer posts). The sign portion of the origami comprises 96 helices arranged in a honeycomb lattice. This larger cross-section, and correspondingly thicker side-views, compared to the wedge/pointing hand structures should reduce the cryo-EM contrast discrepancy between the end- and side-views. The sign is also designed to be highly asymmetric from all side-views to make micrographs easier to interpret. Scaffold routing and staple sequences were designed using caDNAno software [62] with for an M13mp18 scaffold, and 3D models of the structure were generated using CanDo [69]. Depictions of alternate views of the designed structure are shown in Figure 2.20, and the layout of crossovers and staple sequences are provided in Appendix A.3.3.

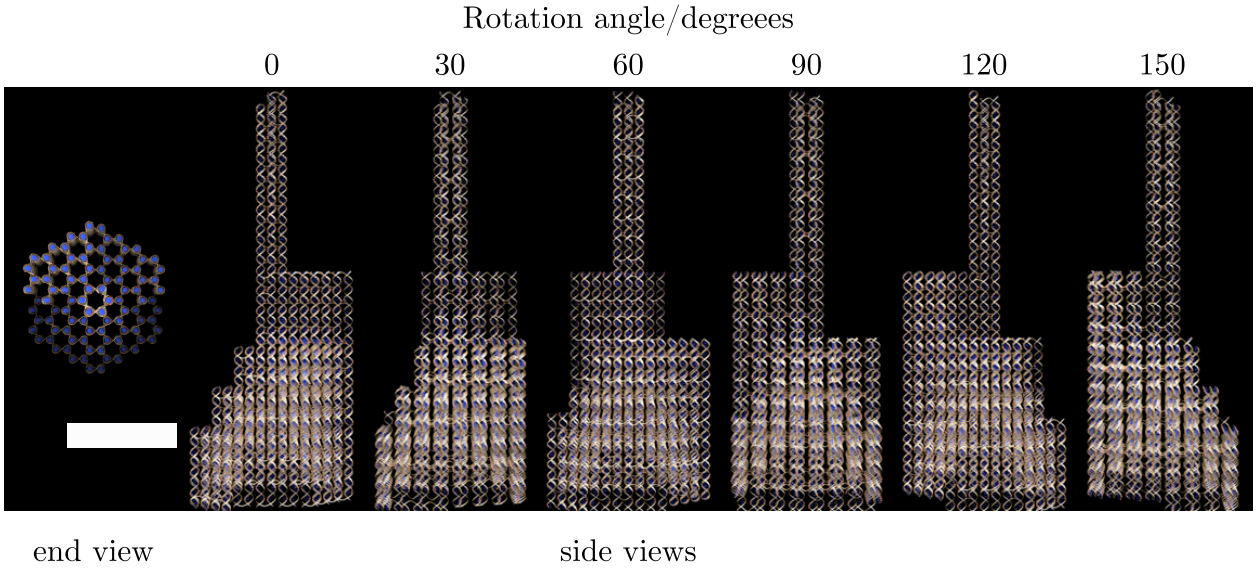


Figure 2.20: Views of designed honeycomb lattice signpost origami nanostructure as generated by Cando (scale bar 20 nm). An end-view is shown on the left, the other views are side-views at different rotations about the post axis.

2.5.2 Synthesis and Characterisation

Signpost origami were prepared at 10 nM concentration (unlike the wedge and hand nanostructures, higher concentration annealing was found to adversely affect folding), using the same 16 hour annealing ramp, buffer conditions and staple concentrations as the nanostructures presented thus far. Folding was verified by AGE (Figure 2.21a), where only a small mobility shift is observed relative to the single-stranded scaffold.

The signpost nanostructures were characterised by negative-stain TEM (Figure 2.21b, c). In some cases the posts appear to be bent relative to the signs, however this is most likely a consequence of the origami landing on the EM grid surface, and not representative of the origami conformation in solution. On the grid, side-views of the origami are in one

of 6 orientations (0 - 300° in 60° intervals, with angles defined as in Figure 2.20). These different views are distinguishable, and structures are annotated according to their angle in Figure 2.21c.

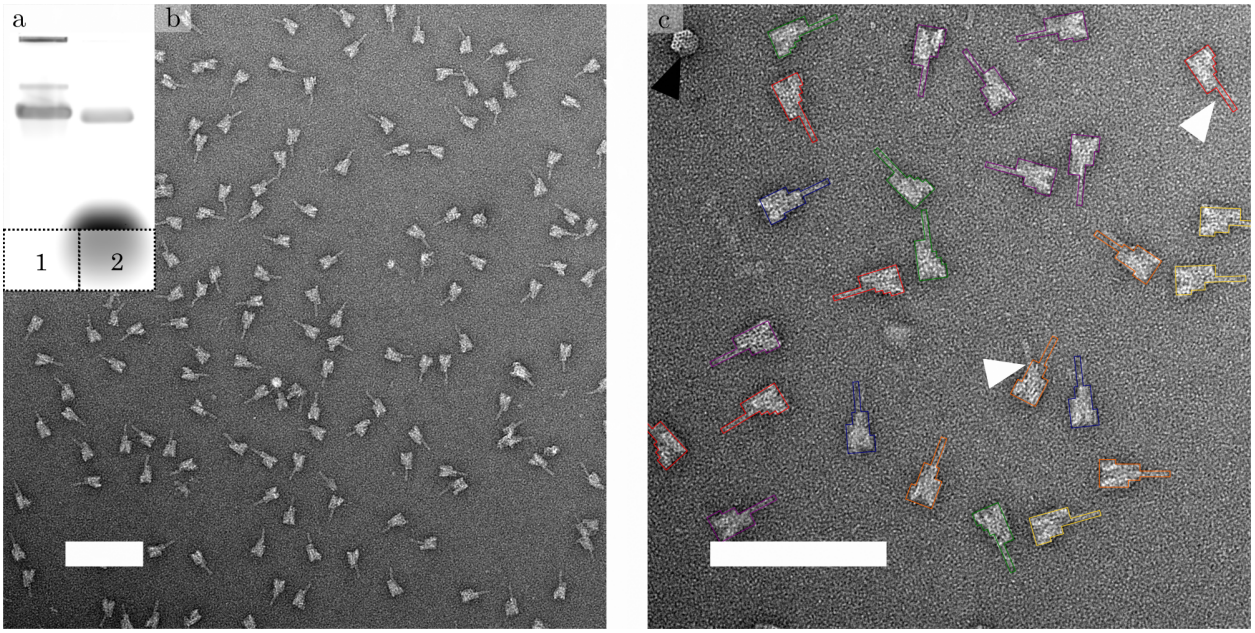


Figure 2.21: (a) Gel electrophoresis of scaffold strand (lane 1) and folded signpost origami, unpurified (lane 2). A small mobility shift is observed. Characteristic wide-field (b) and high-magnification (c) views of honeycomb-lattice signpost origami nanostructures imaged by negative stain TEM (scale bars 200 nm). Examples of bent posts are indicated by white arrowheads. Outlines (from atomic models) are given to side-views in (c) depending on the angle (as defined in Figure 2.20). Red/orange/yellow/green/blue/purple outlines correspond to angles of 0/60/120/180/240/300°. One end-view is also seen (black arrowhead).

Signpost nanostructures were also prepared for cryo-EM in cell lysate, as described previously. Examples micrographs are shown in Figure 2.22. As expected, the sign portion of nanostructures is visible with high contrast, whilst the post portion cannot be visualised for many nanostructure orientations. In cases where the post is visible, there are no obvious examples of post-bending as seen by negative stain TEM. Nanostructures are visible with similar contrast to membrane structures, and views of different nanostructure orientations have comparable contrast.

2.6 Signpost Origami: Optimisation

2.6.1 Heavy-Metal Pre-Soaking

In classical TEM studies, contrast is introduced to the system by heavy-metal soaking. Although naked nanostructures provide good contrast for cryo-EM, achieving contrast enhancement would still be advantageous, particularly for applications to busier and more complex target systems. Since origami nanostructures require cations to maintain structural

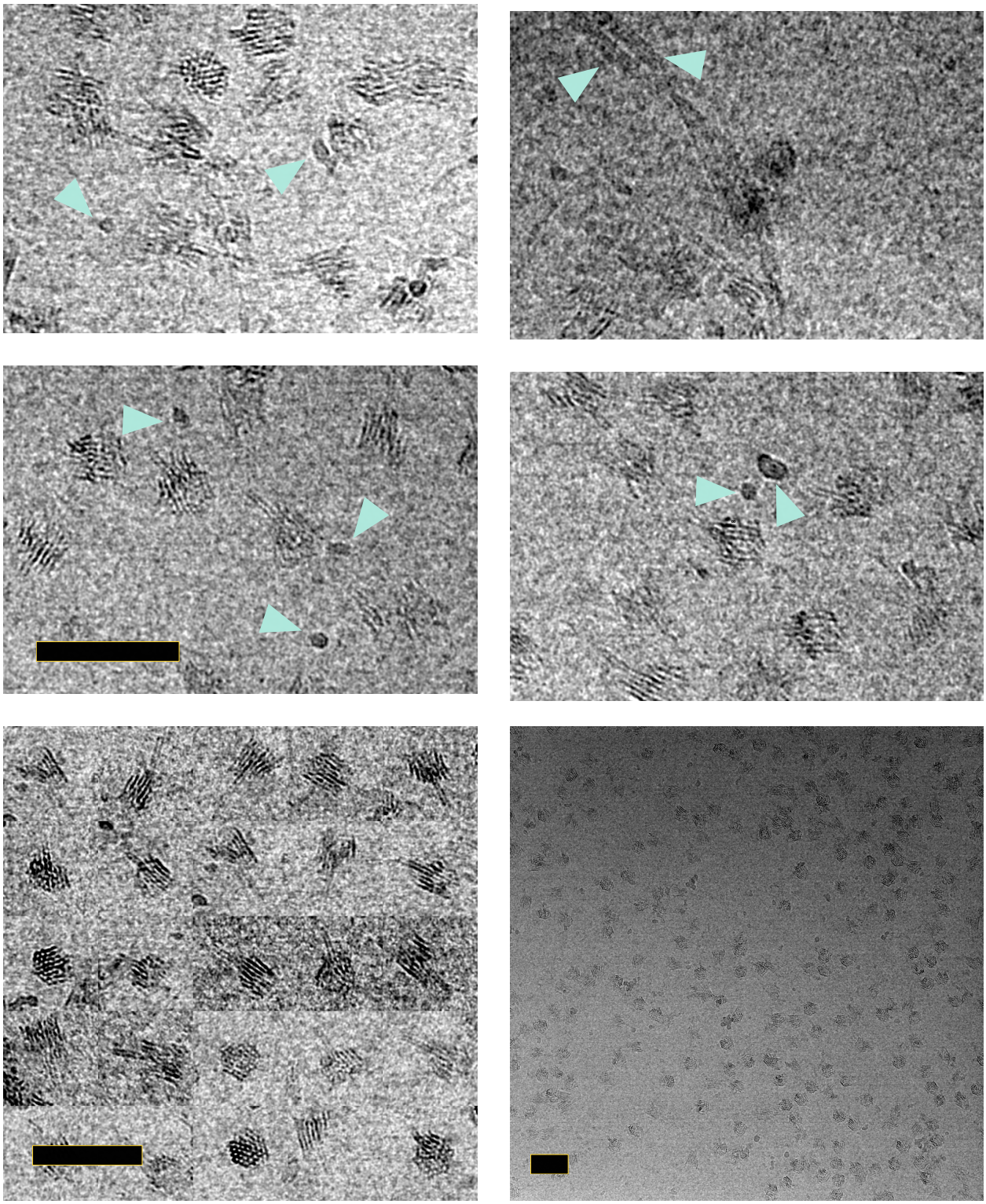


Figure 2.22: Cryo-EM results of signpost origami nanostructures imaged in bacterial cell lysate, scalebars 100 nm. The top four panels shown different high-magnification views, with membrane-bound cell debris indicated by cyan arrowheads. The bottom left panel shows a montage of 25 randomly-selected nanostructure views, and the bottom right panel is a wide-field overview.

stability, I opted to replace the traditionally-used Mg^{2+} ions with heavy-metal ions. The proposition is that these ions will remain bound to the origami when it is introduced to the tagging system, and serve to increase contrast of the nanostructure (analogous to staining in classical TEM) without impacting the target system. I have named this process 'pre-soaking'.

It has been demonstrated that, for cations of charge $\leq 2+$, the total concentration of charge is the relevant factor for origami stability, whereas for cations of charges $> 2+$, only the stoichiometric ratio of cation:phosphate group is important [129]. This suggests a general rule that cations of charges $> 2+$ are tightly bound to the origami, whereas cations of lower charges are being constantly exchanged with ions in solution. I therefore decided that an ion of charge 3+ (or more) would be more suitable for providing higher contrast, and opted for Yb^{3+} (from $YbCl_3$ salt), which has a good balance of high molecular weight, affordability and non-toxicity.

To avoid re-optimising the folding and purification protocols, I folded origami under standard conditions in a buffer containing $MgCl_2$ and transferred into solutions of $YbCl_3$ (Alfa Aesar) post-purification. Origami samples were concentrated to $\approx$ 750 nM[8], and diluted 1/100 in different $YbCl_3$ solutions. Agarose gel results are shown in Figure 2.23.

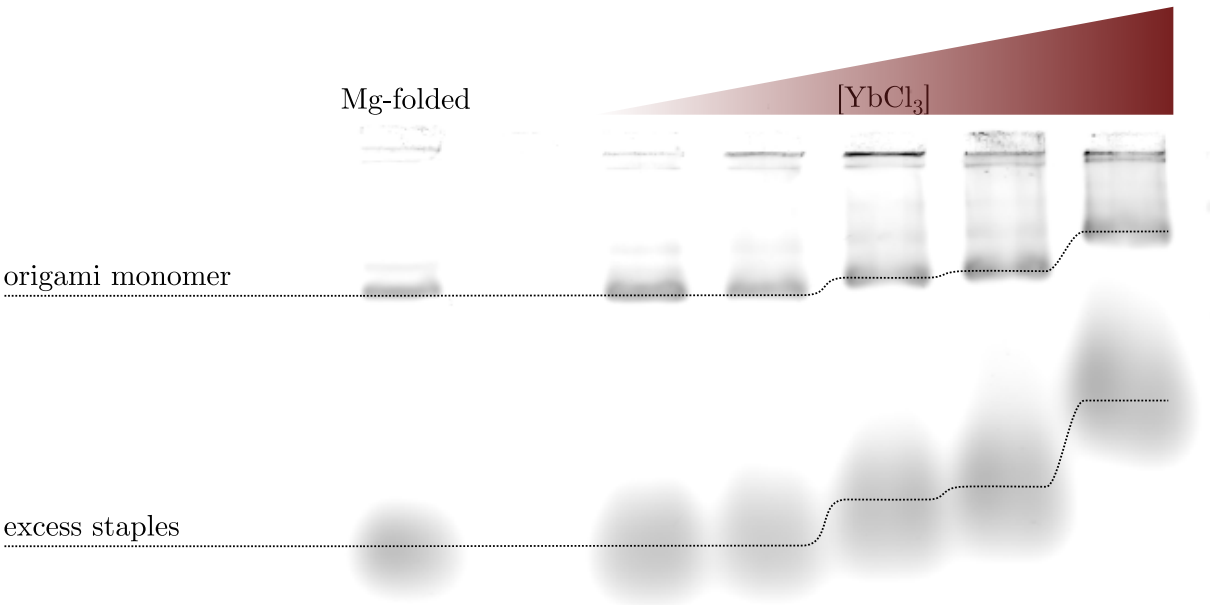


Figure 2.23: AGE of honeycomb signpost origami transferred into solutions of $YbCl_3$. The first lane is the origami sample before buffer exchange, in the $MgCl_2$ buffer. Lanes 2-6 are origami resuspended in 0.1, 0.5, 1, 10, 50 mM $YbCl_3$.

At low ($\leq$ 0.5 mM) concentrations of $YbCl_3$, the origami runs to the same position as the sample in Mg-based buffer, indicating that the origami remains well folded at these low ion concentrations. At higher ion concentrations, both the origami and staple band have lower mobility. Often lower gel mobility indicates nanostructure mis-folding, but this does not explain the staple band shift. The most likely explanation for this is that, at higher concentrations, the Yb^{3+} changes the net charge of both the origami and staples (additional positive charges means a less negative net charge and therefore slower movement through

[8] 500 µL of 10 nM unpurified origami are spun for 30 minutes at 14,000×g in a 100 kDa Amicon filter.

the gel). This effect has been observed previously by Agarwal and colleagues with oligolysine stabilisation of origami structures [130]. It is therefore likely that origami remains well-folded even at high $YbCl_3$ concentrations (since no aggregation is observed, which would be a more likely consequence of excess positive charge).

Attempts to image Yb^{3+}-folded nanostructures by negative stain TEM, resulted in grids coated in large salt crystals which were impossible to image effectively, so Yb-folded origami were imaged directly by cryo-EM. A signpost origami sample was prepared for cryo-EM as described previously, but 5 mM $MgCl_2$ (no buffering) was used in the elution step of column purification. After purification the concentration was measured as ≈400 nM. The sample was dived into three, and aliquots diluted 1 in 2 with 2, 20 and 200 mM $YbCl_3$ solutions. Example cryo-EM micrographs are shown in Figure 2.24

At low (1 mM) $YbCl_3$ concentrations, there is no observable contrast improvement relative to the $MgCl_2$-folded origami. At higher (10 mM and 100 mM) concentrations we see dramatic aggregation of nanostructures. This was not directly observed by AGE, but origami gel samples are run at a lower concentration which would reduce aggregation, and gel electrophoresis is a non-equilibrium technique so aggregates may fall apart in the gel. Transient interactions between structures, which may be responsible for aggregates, are an alternative explanation for the observed mobility changes in Figure 2.23. Additionally, blotting for EM increases the nanostructure concentration, and may induce aggregation. This is the most likely explanation for difficulties in imaging by negative stain TEM: blotting induces aggregation, and large aggregates draw in uranyl acetate resulting in large salt crystals covering the grid.

Altogether, heavy-metal pre-soaking is not recommended as a method for improving nanostructure contrast for cryo-EM, but in the conclusion to this chapter I will suggest alternative options.

2.6.2 Comparing Lattice Types

For building multilayer DNA origami objects, it is commonplace to design helices packed onto square [64] or honeycomb lattices [62, 63]. So far, I have only worked with honeycomb lattice designs, but I also designed a square lattice nanostructure for comparison for two reasons:

Contrast: The square-lattice helix-packing density is higher than for a honeycomb lattice[9] This may result in enhanced imaging contrast for imaging by cryo-EM.

Multiplexing: In cryo-EM micrographs of origami nanostructures it is notable that, unlike for negative stain TEM, the helical grain is more visible than the overall nanostructure shape. This suggests that different lattice types (or more complex arrangements of helices, such as Dietz's 'stacked cross'[63]) might provide better options for multiplexing than different honeycomb-lattice designs.

[9] Factor of $3\sqrt{3}/4 \approx 1.3$.

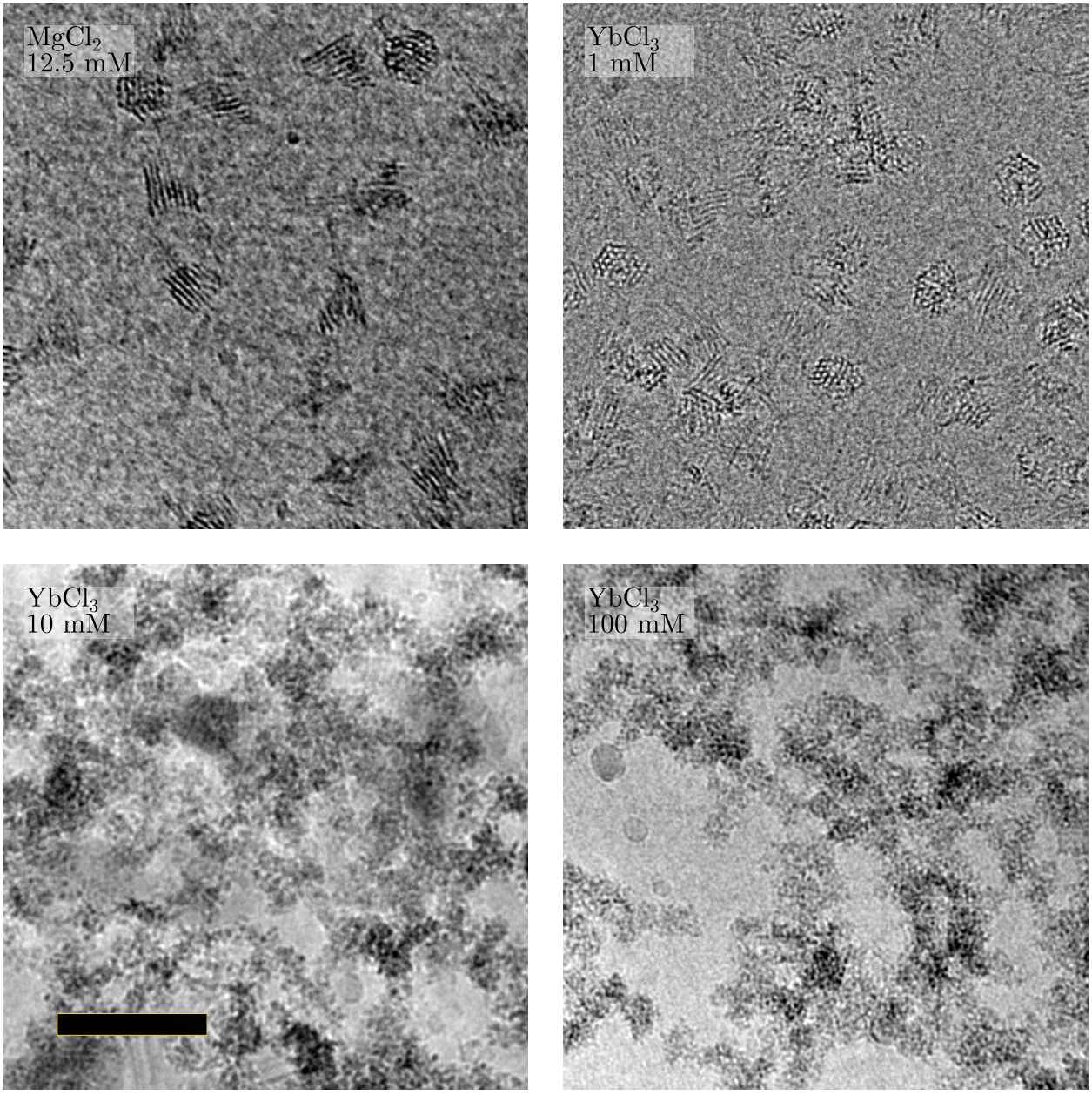


Figure 2.24: Typical cryo-EM micrographs of origami folded in 12.5 mM $MgCl_2$ (top left) and 1 mM (top right), 10 mM (bottom left) and 100 mM (bottom right) $YbCl_3$, scalebar 100 nM.

For a direct comparison of lattice types, I designed a signpost origami with a similar overall architecture based on a square lattice. The 'sign' portion again consists of 96 helices, now arranged in a 8×12 grid, and is asymmetric from all angles. The 'post' is a 3×2 square lattice six-helix bundle which is expected to display similar bending behaviour to a honeycomb six-helix bundle. Scaffold routing and staple sequences were designed using caDNAno software (with advice from William Shih) and 3D models of the structure were generated using Cando.

For square lattice designs, crossover positions are initially placed assuming a helical repeat of 10.66 bp, i.e. successive crossovers between the same pair of helices are separated by 32 bp (3 full turns), and crossovers between one helix and successive neighbours are separated by 8 bp (3/4 turns). The natural helical repeat of the B-helix is closer to 10.44 bp [59], and this small offset can accumulate over the structure resulting in a global twist. To counter this, I have introduced one deletion every 48 bp (i.e. 47 bp/4.5 turns = 10.44 bp/turn), located at the same positions for each helix to minimise any length strain. The caDNAno design and full set of staple sequences are provided in Appendix A.3.4, and depictions of alternate views of the designed structure are shown in Figure 2.25.

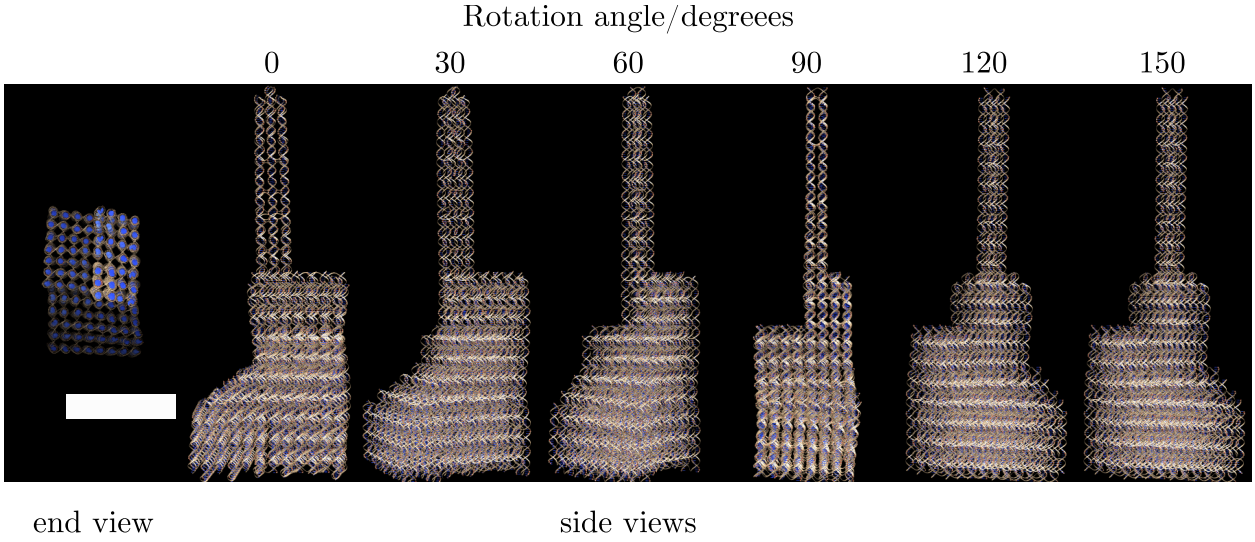


Figure 2.25: Views of designed square lattice signpost origami nanostructure as generated by Cando (scale bar 20 nm). An end-view is shown on the left, the other views are side-views at different rotations about the post axis.

To assemble this nanostructure, I tested four different annealing protocols, each with 20 nM of scaffold and a 10-fold excess of staples in buffers of 1×TE + 6, 8, 10, 12, 14 and 16 mM $MgCl_2$ as shown in Figure 2.26. The annealing protocol 5040-18h (cooling from 50°C to 40° over 18 hours, Appendix C.1) with 16 mM origami produced the best result, with a high yield of folded origami and little aggregation or higher molecular weight species. These conditions were used for all future folding of this origami.

The square lattice signpost nanostructures were imaged by negative stain TEM (Figure 2.27), and are well-folded at high yield. Side-views of the origami can be seen in 4 different orientations (0-270° in 90° intervals with angles defined in Figure 2.25). These different views are distinguishable.

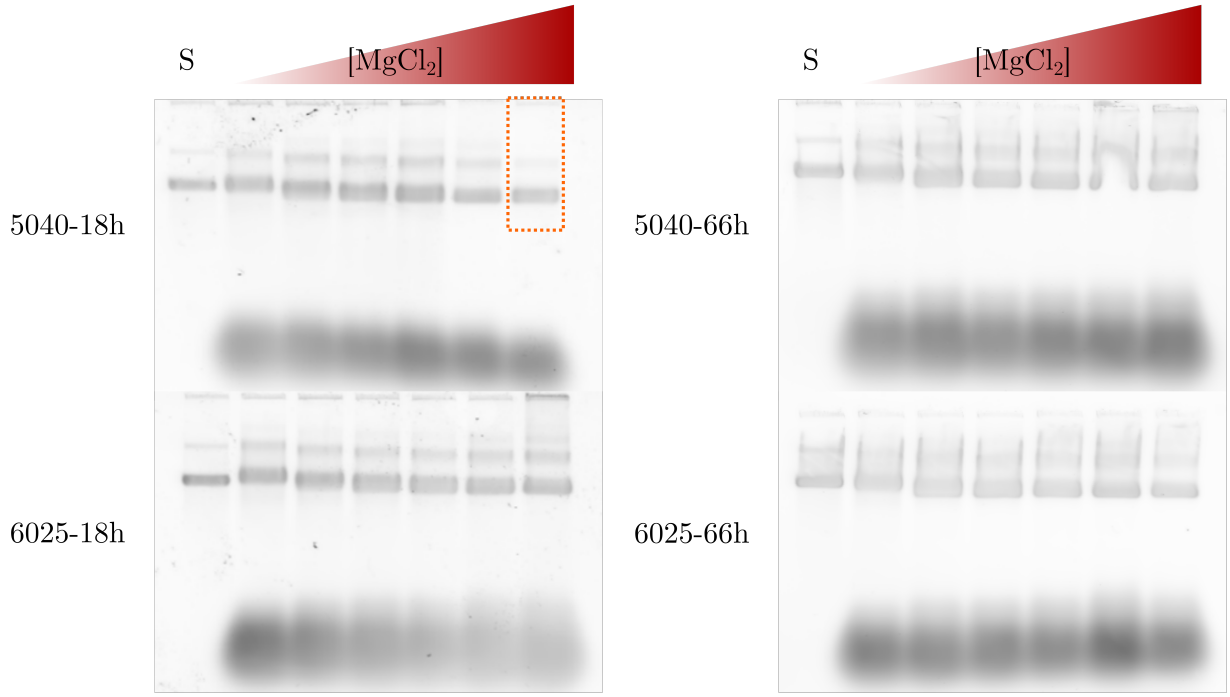


Figure 2.26: Overview of annealing protocol screening for square lattice signpost origami. The four gels correspond to four different annealing protocols. Each gel shows the scaffold in lane 1, and then origami annealed in 1× TE + 6, 8, 12, 14, 16 mM $MgCl_2$. A mobility shift relative to the scaffold indicates binding. The lane of the chosen protocol is indicated in orange.

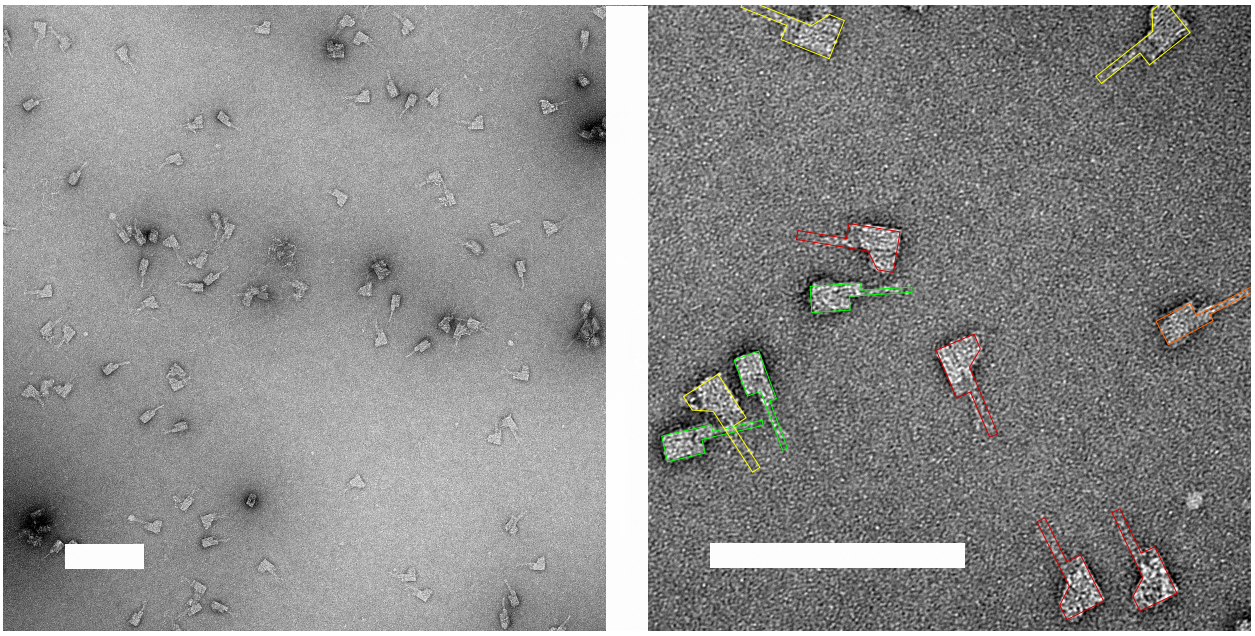


Figure 2.27: Characteristic wide-field (left) and high-magnification (right) views of square-lattice signpost origami nanostructures imaged by negative stain TEM (scale bars 200 nm). Outlines (from atomic models) are given to side-views in (b) depending on the angle (as defined in Figure 2.20). Red/orange/yellow/green outlines correspond to angles of 0/90/180/270°.

To compare helix types (from a point of view of contrast comparison and multiplexing viability within cryo-EM) a mix of the square-lattice and honeycomb-lattice DNA origami nanostructures were purified for cryo-EM and imaged in mammalian cell lysate. Typical wide-field micrographs and high-magnification views of the two structures are depicted in Figure 2.28. Examples of square-lattice and honeycomb-lattice origami nanostructures are indicated by blue and red arrowheads respectively. By eye there is no notable difference in contrast, although the two nanostructures can, in general, be distinguished.

I also made a quantitative comparison of nanostructure imaging properties. The analysis method is illustrated in Figure 2.29a. End-views of the structures, which are very high contrast for both lattice types and easily distinguished from one another, were neglected in this analysis. 56 side-views were randomly selected and the following steps performed:

1. Micrographs were opened in the Fiji distribution of ImageJ [131]. Line profiles across side-views are plotted.
2. Locations and values of the line-profile peaks and troughs are determined using the 'findpeaks' function in MATLAB and the peak-to-peak separation is calculated. The MATLAB script is provided in Appendix F.4.
3. A crude measure of contrast is made by subtracting the mean trough value from the mean peak value, and normalising such that the largest value is one.

A histogram of mean peak-to-peak separation values is shown in Figure 2.29b. The distribution separates into two distinct populations corresponding to the more closely-packed square lattice, and the less closely-packed honeycomb lattice. These two structures, despite their similar global architectures, are therefore suitable as orthogonal tags for multiplexing.

On the basis of this distribution, line profiles were designated an identity according to their mean peak-to-peak separation: if >3 nm the line profile belongs to a honeycomb lattice, and if <3 nm the line profile is of a square lattice. Histograms of contrast values (as defined above) for the two profile populations are shown in Figure 2.29c. Contrast measurements for the square/honeycomb lattice profiles are 0.55 ± 0.15 and 0.60 ± 0.17 (given as mean $\pm$ standard deviation). There is an inherent bias in this analysis, since profiles of particularly low contrast may not be visible and so not selected for analysis, but this comparison supports the visual assessment that there is no significant difference in contrast between the two architectures, and so they are equally suitable as tags.

2.7 Summary and Future Directions

In this chapter, I have shown the design and synthesis of a number of origami nanostructures, assessing each for their suitability as a CET tag. I firstly showed the design and synthesis of a wedge-origami nanostructure, which I used to develop a purification technique for cryo-EM samples and to demonstrate the high contrast provided by honeycomb origami nanostructures within a noisy biological context (cell lysate). Contrast is comparable

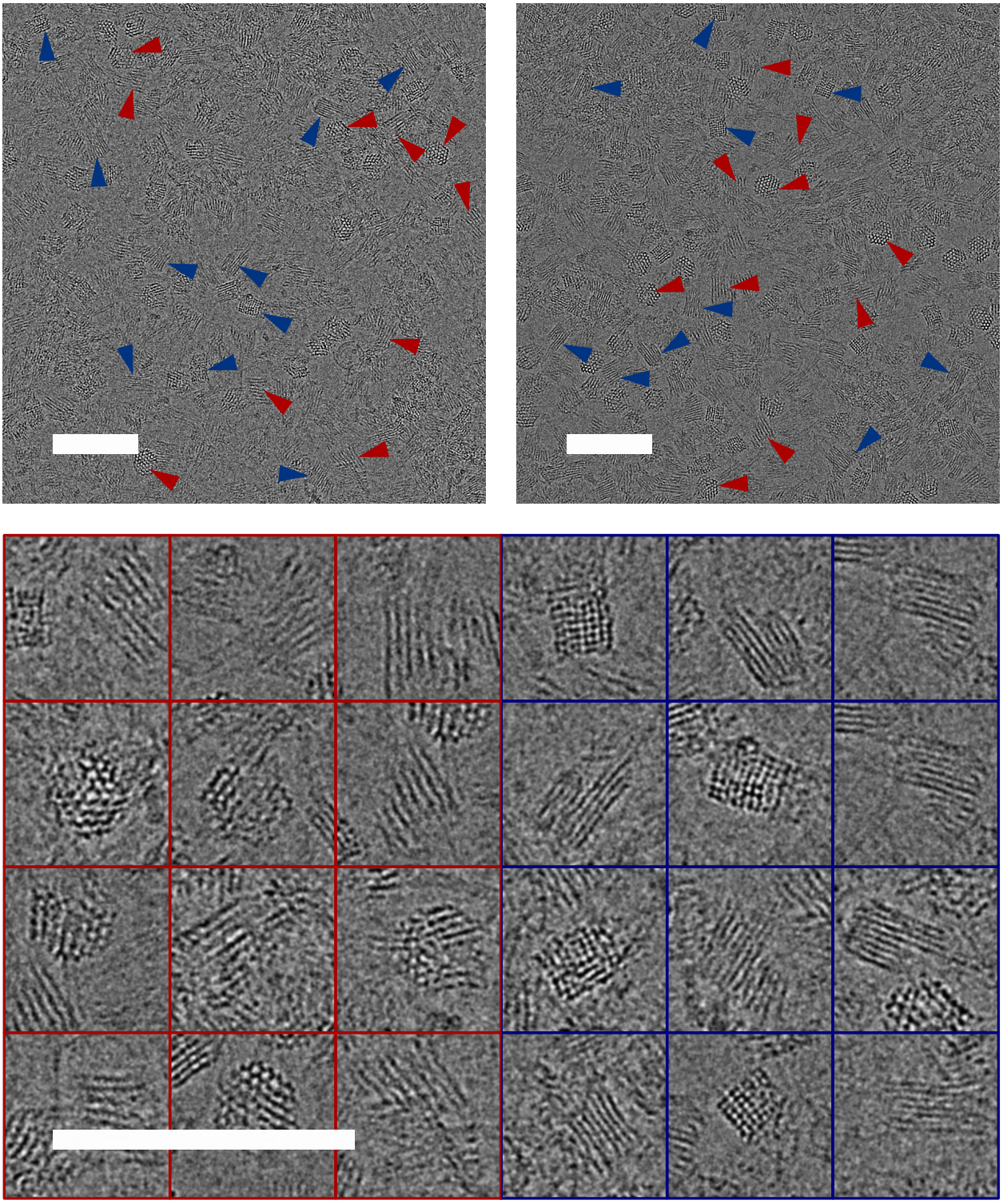


Figure 2.28: Above: two typical wide-field micrographs of mixed square lattice (examples indicated by blue arrowheads) and honeycomb lattice (red arrowheads) signpost origami, imaged in cell lysate. Below: high-magnification views of square lattice (blue outline) and honeycomb lattice (red outline) signpost origami. Scale bars 100 nm.

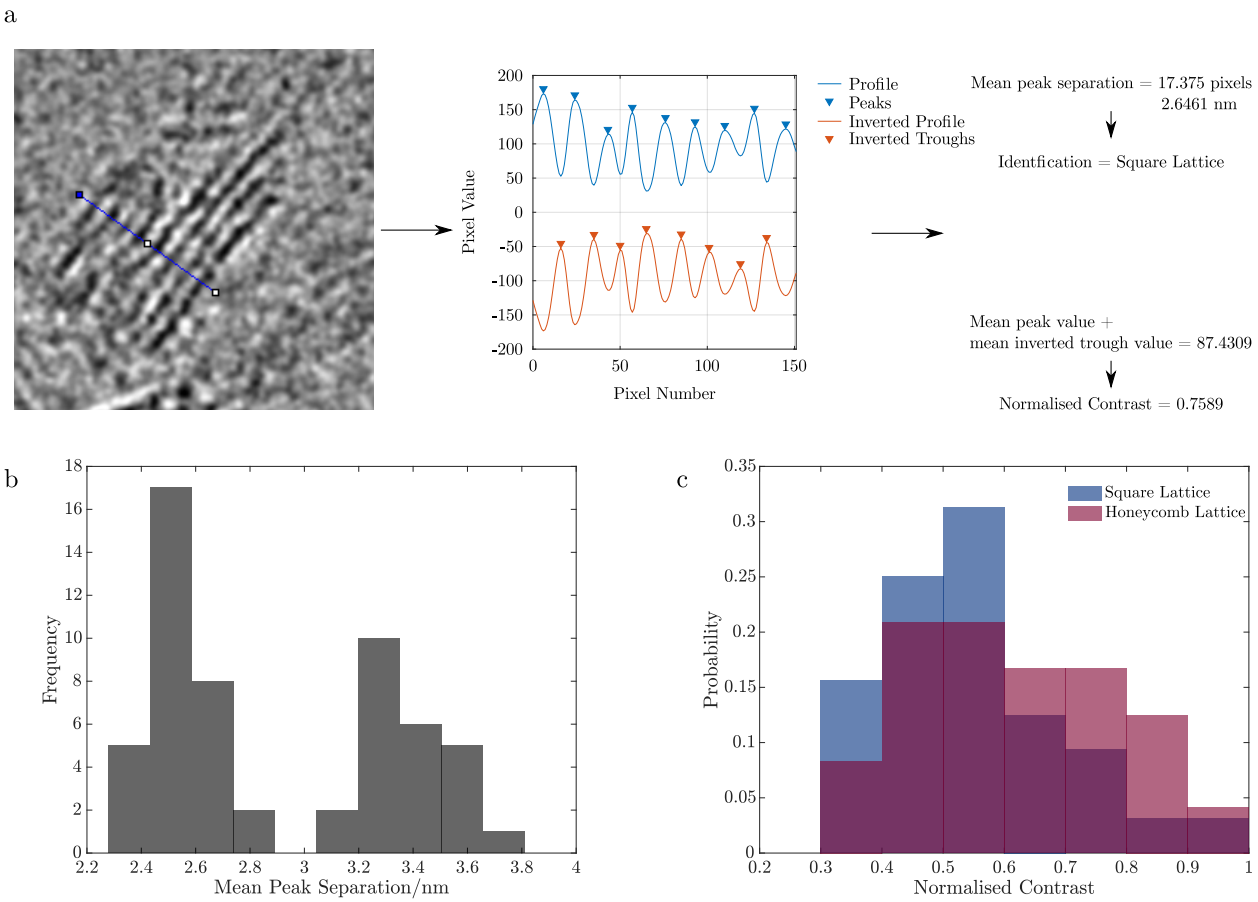


Figure 2.29: (a) Overview of profile analysis for a single nanostructure view: profile is drawn across nanostructure side view in ImageJ, peaks and troughs of profile are identified with MATLAB, peak separation and contrast are calculated for profile. (b) Histogram of mean peak-peak separations measured from 56 selected side profiles. The distribution separates into two distinct populations, corresponding to the more closely-packed square lattice, and the less-closely packed honeycomb lattice. (c) Contrast probability distributions of measured profiles from square (blue) and honeycomb (red) lattice origami. Profiles are designated as square/honeycomb if the mean peak separation is less than/more than 3 nm. Contrast is defined crudely as mean-peak minus mean-trough pixel values across each profile.

to membranes structures which are typically the highest contrast components in cellular micrographs/tomograms. I have also shown that hollow nanostructures, formed by arrangements of linked 6-helix-bundle domains, do not provide sufficient contrast for my application.

For lattice-based nanostructures in lysate, the outline of the structure is not always clear, but the internal arrangement of helices gives rise to a distinctive 'stripy' effect. Consequently, different honeycomb-lattice nanostructures can be difficult to distinguish in crowded micrographs, whereas structures of different lattice types (even with similar global architectures) can be readily distinguished. This chapter culminated in the design and synthesis of a signpost origami nanostructure which possesses suitable characteristics for use as a tag for CET. The nanostructure comprises a high-contrast, asymmetric 'sign' section used for identification, and a six-helix bundle 'post' section, intended to spatially separate the sign from the targeting module to reduce crowding and prevent concealment of the target molecule.

For *in situ* multiplexing, it would be advisable to use origami with different global architectures and different lattice arrangements, with examples of designed signpost variants shown in Figure 2.30. New lattice options may be achieved through rearrangements/modifications to standard honeycomb/square lattices (e.g. structure 3 in Figure 2.30), but hexagonal/hybrid lattices will also be useful options [65]. Variation in post-length will also be an important attribute for multiplexing.

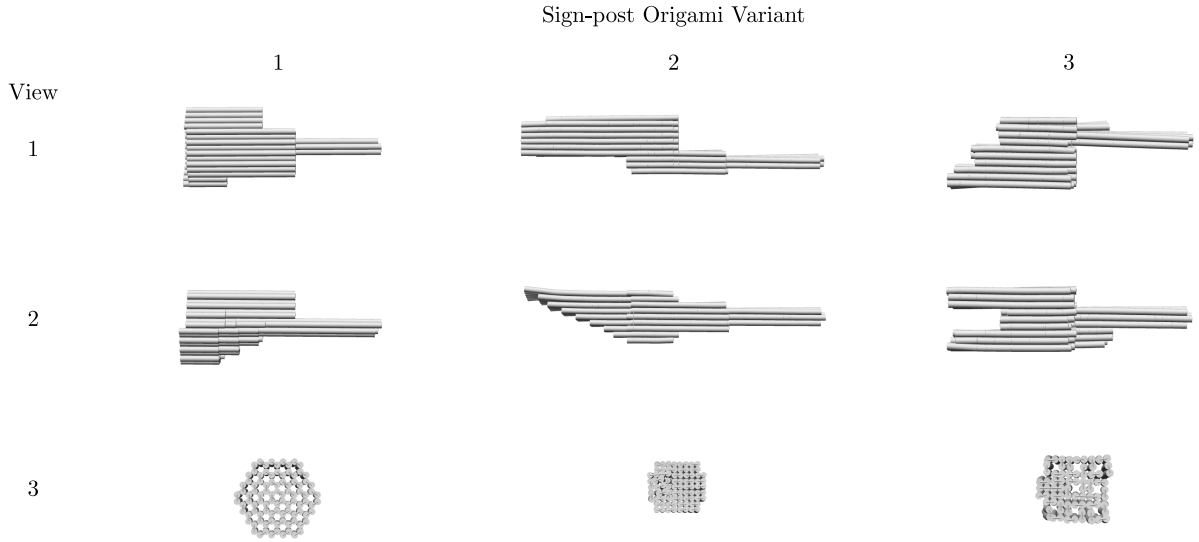


Figure 2.30: Examples of designed signpost origami variants for multiplexing. Variant 1 is the honeycomb lattice origami nanostructure presented previously. Variants 2 and 3 are structures designed for multiplexing, each with different lattice arrangements and global architectures. 3D orthogonal views are shown for each structure. Images are from schematic 3D models, generated in Cando [69] and rendered in Chimera [124].

The contrast of unmodified nanostructures is sufficient for our purposes, but there are options for improving this further. We have seen that heavy-metal pre-soaking at low

concentrations does not give a marked enhancement of contrast, and at high concentrations induces nanostructure aggregation, but the addition of metallo-intercalators such as cis-platin is another possibility. These compounds affect the helical repeat of the DNA B helix [132, 133] which would need to be considered for crossover placements during origami design. Other effects, especially toxicity, would also need to be addressed. Alternatively, chemically modified oligonucleotides may be used to improve contrast, e.g. nano-gold modified oligonucleotides [134] could be used to decorate the origami surface. The high level of control with this approach also provides a new way of multiplexing structures, potentially circumventing the need for complete nanostructure redesign for new tags.

For CET tagging applications, nanostructures must be functionalised for targeting, and for intracellular tagging, delivering of nanostructures into cells must also be considered. These will be the topics of Chapter 3.

Chapter 3

Targeting and Delivering Origami Tags

In this chapter, I will present methods for using DNA nanostructures as tags. Firstly, I will introduce a GFP aptamer and demonstrate its binding properties to GFP analogues with and without conjugation to an origami nanostructure. Next, I will show virus- and cell-binding of origami tags to a murine leukaemia virus test system to demonstrate the viability of this technique. Finally I will explore electroporation of origami into cells with a view to intracellular tagging experiments.

3.1 Author Contributions

The fluorescent proteins mEGFP, mVenus, sfGFP and mCherry were provided by Lindsay Baker and Cathy Whittle. Samples of murine leukaemia virus (MLV) and the MLV glycoprotein Env were provided by Vojta Pražák. Cryo-EM/CET in this chapter was performed by, or under the supervision of, Lindsay Baker or Vojta Pražák. Vojta Pražák also preformed the cell-binding experiments in Section 3.3.5 and built the tomograms in Figure 3.9 and 3.10. Fluorescence microscopy was performed by Emma Silvester at the Division of Structural Biology (STRUBI) with guidance from Lindsay Baker. The rest of the work presented in this chapter (aptamer transcription, characterisation and hybridisation to origami, negative-stain TEM, polyamine stabilisation of origami and electroporation tests) was performed by Emma Silvester.

3.2 A GFP Aptamer as a Targeting Module

3.2.1 GFP Aptamer

In Chapter 2, I showed the design, synthesis and characterisation of a signpost DNA origami nanostructure which has appropriate structural characteristics for a CET tag. A tag must also bind to the protein of interest.

There are a number of different options for functionalising DNA origami nanostructures for specific binding to protein targets [135]. For example, biotinylated or NTA-modified oligonucleotides are available commercially. Incorporation of these modified strands into DNA nanostructures permits targeting of streptavidin-fused and His-tagged proteins respectively. Alternatively, a DNA-/RNA-binding protein/peptide could be introduced into the target protein as a fusion tag, and their nucleic acid binding partner attached to the DNA nanostructure. The targeting method I will focus on is nucleic acid aptamers: DNA/RNA strands which fold into a well-defined 3D structure and display high affinity and specificity for their target. These can easily be incorporated into nanostructures through DNA/DNA or DNA/RNA hybridisation. RNA aptamers are of special interest, since these could also be incorporated into RNA nanostructures for future applications in intracellular labelling.

Although aptamers can be developed for any target system, the selection process is time consuming, so it is helpful to develop a nanostructure which targets a common protein tag. The GFP protein, and its family of analogues, can tolerate N- and C-terminal fusion to wide variety of proteins. This property, in conjunction with their fluorescent behaviour, has made GFP and its variants common choices as fusion tags for fluorescence microscopy [136]. There is therefore an abundance of existing GFP-fusion proteins, making a tag targeting GFP an immediately useful tool for microscopists. Tagging fluorescent proteins also has the added advantage of applicability in correlative light/electron microscopy.

In 2011, a group of RNA aptamers which target GFP and related proteins were developed by Shui et al. [137]. The tightest binding aptamer (AP3-1) possesses a measured K_D of 5.1 nM, which is suitable for most biological targets, so I have selected this aptamer for use with my origami tags. I will characterise binding of the aptamer when free, and when conjugated to the signpost origami nanostructure.

For free-aptamer binding experiments I made a small addition to the AP3-1 sequence to maximise the yield of aptamer obtained by transcription by T7 RNA polymerase from a DNA template. The sequence *GGG* was introduced at the $5'$ end and the reverse complement *CCC* was introduced at the $3'$ end. The first 6 nucleotides of the modified sequence are *GGGAGC*, which has a relative promoter strength of 0.97 compared to the consensus starter sequence *GGGAGA* [138]. This addition is not predicted to affect the overall secondary structure, and therefore binding behaviour, of the aptamer. The predicted secondary structure of the unmodified and modified aptamer sequences (at 37°C, with 1 M NaCl) is shown in Figure 3.1a.

The RNA sequence of the modified aptamer (hereafter referred to as GA) was converted to a DNA sequence, the T7 RNA polymerase promoter sequence added at the $5'$ end, and both duplex strands were ordered from IDT (sequences provided in Appendix A.2.1)[1]. The aptamer was transcribed from the annealed duplex using the NEB HiScribe kit (Appendix B.1.1 and B.2.1), and purified by ethanol precipitation (Appendix B.3.3). The transcript quality was verified by native polyacrylamide gel electrophoresis (PAGE) (Appendix E.2), as shown in Figure 3.1b. A single band is seen, confirming that the transcription product is intact and has one consistent fold.

[1]Template sequence and reverse complement were ordered as 4 nmole Ultramers.

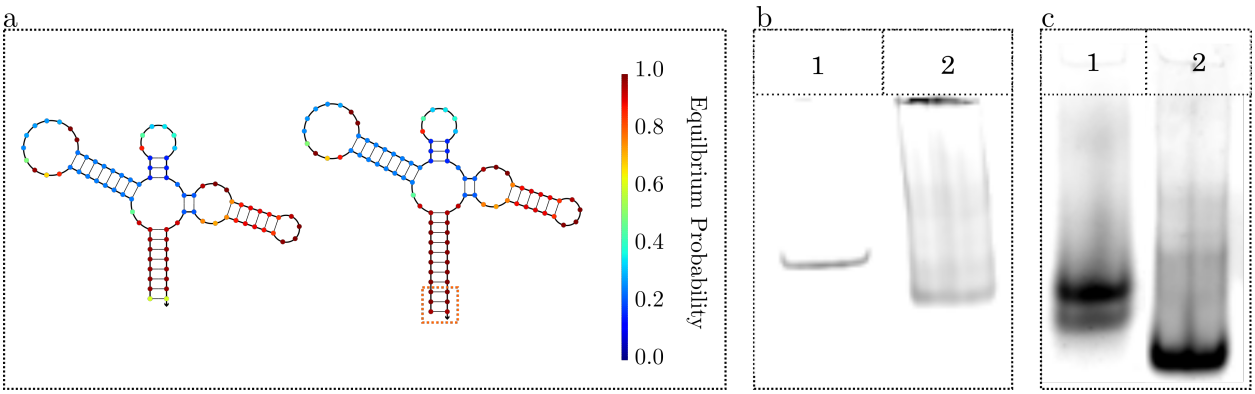


Figure 3.1: Overview of aptamer binding. (a) MFE secondary structures of the published (left) and modified (right) aptamers predicted by NUPACK [167]. Additional nucleotides are indicated by the orange dashed box. (b) PAGE of duplex template (lane 1) and transcription product (lane 2) for modified aptamer. (c) PAGE of EYFP (lane 1) and EYFP in an excess of GA (lane 2). Native fluorescence of EYFP is used (no staining). A gel shift indicates successful binding.

To check GA possesses the same binding properties as its parent AP3-1, samples of 250 nM enhanced yellow fluorescent protein (EYFP)[2] were prepared with and without a ten-fold molar-excess of GA, and incubated for 30 minutes in phosphate-buffered saline (PBS)[3] on ice. These samples were analysed by PAGE (Figure 3.1c), using the native EYFP fluorescence (Appendix E.3). A gel shift indicates that the EYFP is, as expected, fully bound, and thus GA possesses the same qualitative binding behaviour as AP3-1.

3.2.2 Conjugation to Origami

The honeycomb-lattice signpost origami nanostructure introduced in Chapter 2 is used for experiments in this chapter, unless otherwise specified. For hybridisation to origami, I designed a variation of GA with a 12 nucleotide single-stranded 3′ tail, referred to as GA-O. The 3′ tail is complementary to a 5′ extension for a staple whose 5′ terminus otherwise lies at the end of the origami post (see Figure 3.2d). Introducing the aptamer tail at the 3′ end means that any early-termination transcripts (which could result in non-binding RNAs) will not hybridise to the origami as only full transcripts will possess the required sequence.

GA-O was also prepared by transcription from a DNA template, and the transcript quality verified by PAGE (Figure 3.2a). GA-O runs higher than the duplex template, which was initially surprising, however NUPACK analysis revealed that GA-O is able to dimerise through it's 3′ tail, explaining this lower mobility. Binding of EYFP to the GA-O dimer was also confirmed by PAGE (Figure 3.2b).

For hybridisation of GA-O to the origami nanostructure there are two options: addition of GA-O after annealing the rest of the origami structure (including the complementary staple), or inclusion of GA-O in the staple mix during annealing. Efficiency of the former option

[2]EYFP purchase from Cambridge Bioscience.
[3]137 mM NaCl, 2.7 mM KCl, 10 mM phosphate buffer solution (pH 7.4 at 25°C), Sigma.

may be restricted by GA-O dimerisation slowing the kinetics of staple binding, so I opted for the latter option. To verify aptamer integrity after annealing, GA-O was mixed with the complementary staple in origami folding buffer, annealed over 16 hours, and analysed by native PAGE (Figure 3.2c). There does not appear to be any significant degradation or reduction in binding efficiency of GA-O following the origami annealing protocol, indicating that the aptamer is stable over the thermal annealing ramp. Aptamer-conjugated origami (apt-ori) is prepared by annealing M13mp18 scaffold with a 10-fold excess of staples and a 60-fold excess of GA-O over 16 hours. To verify binding of apt-ori with EYFP, 100 nM EYFP (in water) was mixed 1:1 with origami folding buffer, or 400 nM apt-ori and analysed by AGE (Figure 3.2c). Once again, binding was observed, indicating that hybridisation of the aptamer to a large origami nanostructure, and transfer into the origami binding buffer does not disrupt the aptamer/GFP interaction.

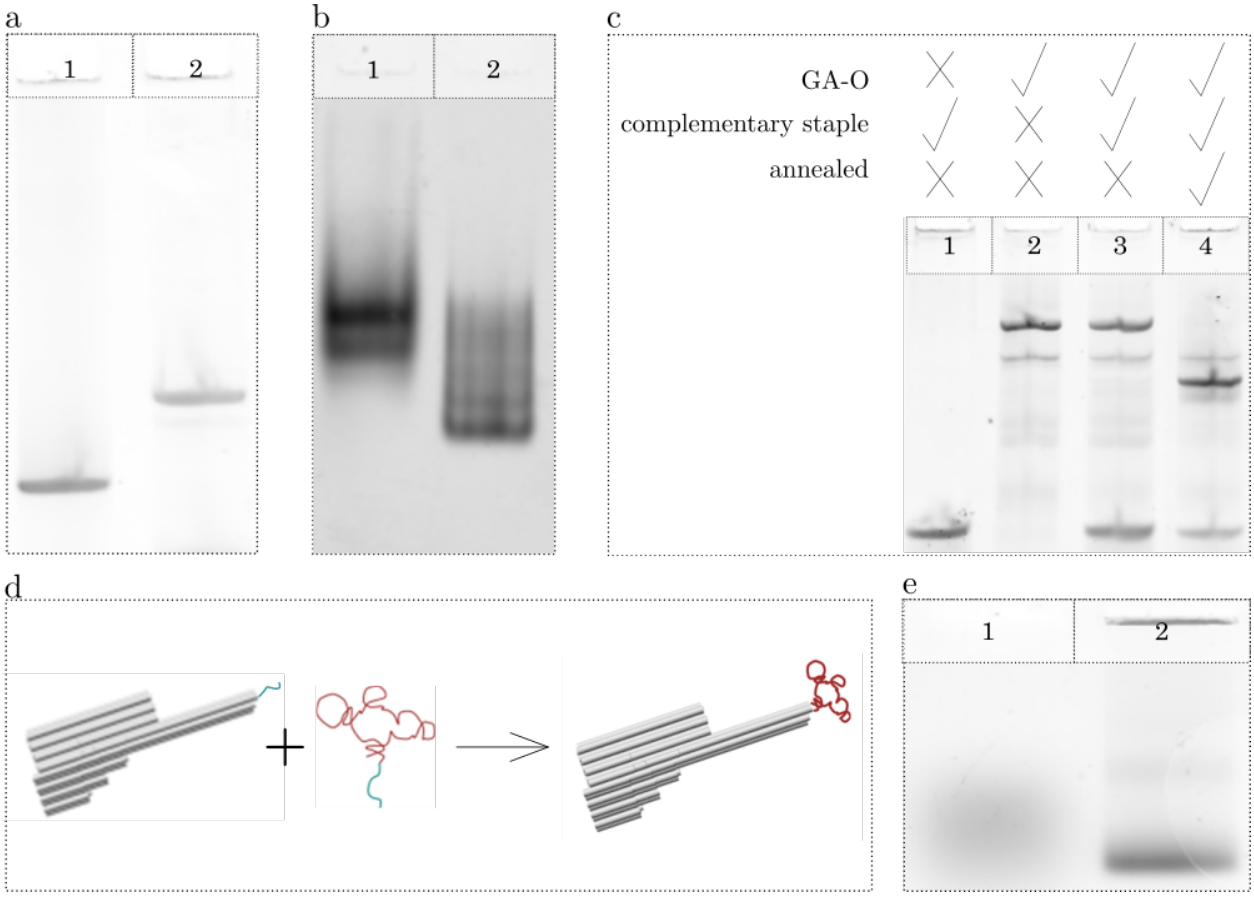


Figure 3.2: Overview of aptamer-origami conjugation and EYFP binding. (a) PAGE of duplex template (lane 1) and transcription product (lane 2) of GA-O. (b) PAGE of EYFP (lane 1) and EYFP in an excess of GA-O. (c) PAGE overview of GA-O binding to its complementary staple. Lane 1 is the complementary staple, lane 2 is GA-O, and lanes 3 and 4 are mixes of the two incubated at RT (lane 3) or annealed over 16 hours (lane 4). (d) Schematic of aptamer-origami conjugation: GA-O (comprising GA, in red and a 3′ tail, in green) is attached to the 'post' of the sign-post origami through hybridisation to a complementary staple extension (also shown in green). (e) AGE of EYFP without (lane 1) and with (lane 2) an excess of aptamer-conjugated origami. The gel shift indicates successful binding.

3.2.3 Quantifying Binding to New GFP Analogues

In their 2011 paper [137], Shui et al. quantified binding of AP3-1 to GFP, YFP and CFP. YFP and CFP are GFP analogues, with different fluorescent properties to their parent protein but very similar polypeptide sequences and crystal structures. For tagging applications there are a number of different GFP analogues which are more relevant/commonly-used, in particular:

mEGFP: GFP and many of its analogues are weak dimers ($K_D \approx 0.1$ mM). For many applications this is not important, but when using GFP as fusion tags for proteins which polymerise or otherwise co-localise, this interaction causes problems. For such systems, monomeric EGFP (mEGFP) [139] or other monomeric GFP variants are preferred tags.

mVenus: Monomeric Venus (mVenus) [140] is a monomeric GFP variant which is the preferred fluorescent protein fusion tag for membrane proteins [141].

sfGFP: Superfolder GFP (sfGFP) [142] is a robustly-folded version of GFP which folds well even when fused to poorly folded polypeptides. sfGFP is the preferred fusion tag for targets that are otherwise difficult to tag.

mCherry: Monomeric Cherry (mCherry) [143] is an RFP-derived fluorescent protein, with a similar crystal structure to GFP but low sequence homology. It is therefore a reasonable test of aptamer specificity (it is expected GA will not bind mCherry), and could be useful as an orthogonal tag for multiplexing.

These four fluorescent proteins, along with YFP (as a positive control) were provided by Lindsay Baker and Cathy Whittle. Binding behaviour was first determined qualitatively by native PAGE (Figure 3.3). 250 nM protein samples were prepared with and without a 10-fold molar excess of GA. All GFP analogues were seen to bind, whilst mCherry was not.

To determine the dissociation constant, K_D, for the four binding proteins, isothermal titration calorimetry (ITC) was used. ITC is a high-accuracy quantitative technique for measuring thermodynamic parameters of binding interactions in solution. Apparatus consists of two cells: a reference cell containing water, and a sample cell containing one of the binding molecules (in my case, aptamer GA). A feedback loop maintains cells at the same temperature. The binding partner (fluorescent protein in my case) is loaded into a syringe and added to the sample cell in small injections. For each injection the heat absorbed or released is measured by the microcalorimeter, and these measurements over the course of the experiment, as binding saturates, are used to compute a binding curve, from which thermodynamic parameters can be extracted.

Samples of GA and fluorescent proteins were dialysed overnight into a common buffer[4], as mismatched buffers can result in high heats of dilution which makes data difficult to analyse.

[4] 100-500 µL 10K Slide-A-Lyzer MINI devices were used for dialysis (ThermoFisher Scientific). 500 µL samples were dialysed in 1 L of PBS + 5 mM $MgCl_2$ at 4°C.

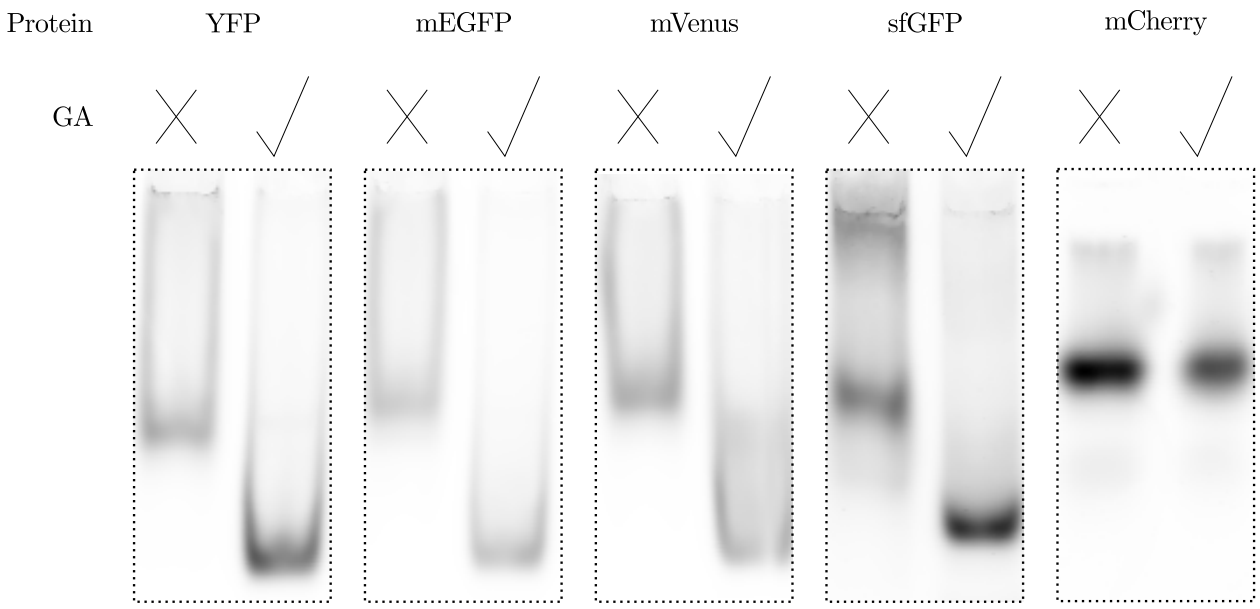


Figure 3.3: PAGE of different fluorescent proteins with and without GA. A binding shift is shown for all proteins except for mCherry.

GA concentration was adjusted to a measured concentration of 10 μM (according to A_{260} measurements) and protein concentration adjusted to ≈ 30 μM (protein concentrations were measured accurately from extinction coefficients after the experiments were performed). Experiments were performed at 25°C, and 4 μL injections were used. For fitting the binding curves, the binding stoichiometry was specified as 1:1 and the protein concentration was also specified. The aptamer concentration was used as a fitting parameter, as an unknown fraction of aptamer is likely to be misfolded or degraded and some fraction of free NTPs may also remain, which would affect the measured concentration.

The raw data and fitted binding curves for the four binding proteins are shown in Figure 3.4 and an overview of the thermodynamic parameters is presented in Table 3.1. For mEGFP, a larger than expected heat of dilution was observed, most likely resulting from a manual error, e.g. incomplete drying of the syringe after washing could result in a buffer mis-match. All four proteins show low nano-molar K_D values with no significant difference between the four proteins. These four proteins are therefore suitable targets for apt-ori, whilst the lack of observed binding to mCherry suggests that this could be used as an orthogonal target (hinging on the development of an mCherry aptamer).

	K_D/nM	ΔG/kJmol^{-1}
YFP	4 ± 2	-48.0
mEGFP	4 ± 6	-47.5
mVenus	4 ± 2	-47.8
sfGFP	13 ± 8	-45.1

Table 3.1: Thermodynamic parameters for GA binding to different GFP analogues.

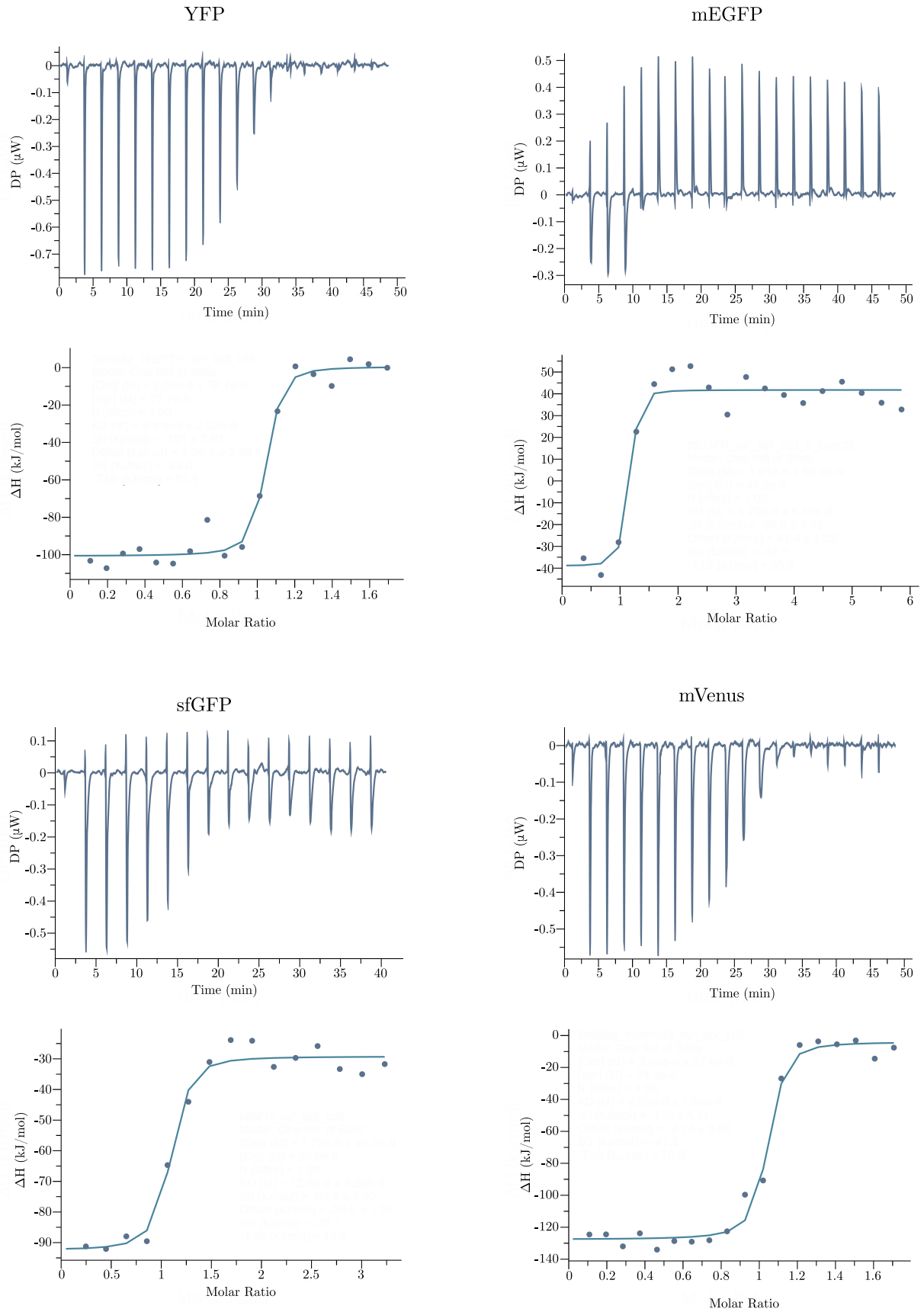


Figure 3.4: ITC results for YFP, mEGFP, sfGFP and mVenus. For each protein, the top panel is raw data after baseline correction, and the bottom panel is integrated data corrected for the heat of dilution, fitted to a one-site binding model. Analysis was performed using MicroCal PEAQ-ITC analysis software.

3.3 Moloney murine leukaemia virus: A test system for tagging

3.3.1 Murine Leukaemia Virus

Once origami binding to GFP and its analogues had been observed, the next step was to demonstrate *in situ* binding to an appropriate GFP-tagged test system. Murine leukaemia viruses, commonly abbreviated to MLVs, are retroviruses named for their ability to induce cancer in murine hosts. Moloney murine leukaemia virus (MoMLV) is a well-studied strain of MLV owing to potential gene therapy applications [144]. In 2005, Förster et al. used cryo-electron tomography to reconstruct MoMLV virions at high resolution and, in conjunction with single-particle averaging techniques, reconstruct envelope-glycoprotein (Env) complexes to 2.7 nm (Figure 3.5) [145].

In this section, a strain of MoMLV with YFP-tagged envelope-glycoproteins (Env-YFP) is presented as a test target-system for DNA-origami tags. The system offers a number of desirable properties:

- Correct binding is easily identifiably as MoMLV has a simple, single-glycoprotein surface.
- Viruses can be imaged in buffer, so no origami delivery protocols are required, and the stability of tags in an intracellular environment is not a concern.
- Binding of Env-YFP to its designated receptor, murine cationic amino acid transporter-1 (mCAT-1), provides an option for cell-tagging experiments using the same system [146].

3.3.2 Aptamer Binding to Env-YFP

A preliminary check of aptamer binding was first performed with solubilised Env-YFP, to verify that fusion of YFP with a target protein does not prohibit binding. Env-YFP with and without an excess of free aptamer and apt-ori was analysed by PAGE and AGE respectively (results shown in Figure 3.6). Binding shifts are seen in both cases.

Interestingly, the agarose gel shows two bands. A reasonable explanation for this is domain shedding. Retrovirus envelope proteins consist of two domains: the surface protein (SU), and the transmembrane protein (TM). They are able to shed the SU domain [147, 148] which, in this case, includes the YFP fusion tag. It is therefore possible these two bands correspond to apt-ori bound to the full Env-YFP protein, and apt-ori bound to the YFP-tagged SU domain. Although the difference in size should not result in such a significant mobility shift, a difference in solubility can also affect gel migration behaviour, and explain these two bands. This same effect is not evident in the polyacrylamide gel, but since less soluble species can run as a smear rather than a distinct band (as evidenced by the left-hand

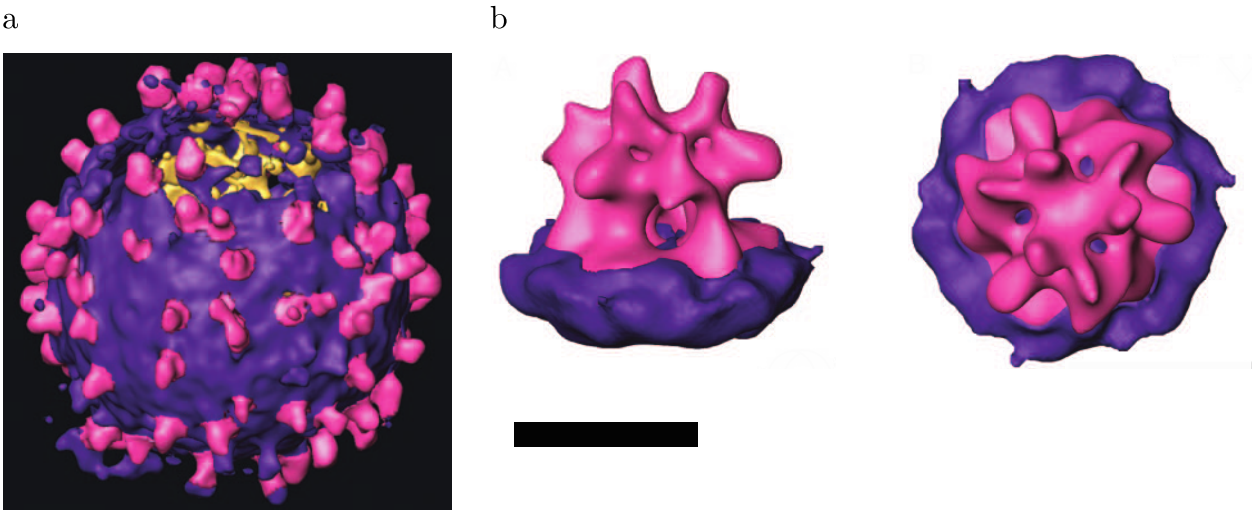


Figure 3.5: Figure and caption adapted from Förster et al. [145] (a) Surface rendered tomogram of an individual MoMLV virion that possesses particularly many spikes. Env complexes (magenta) are located at the lipid surface of the virus (purple). (b) Final Env structure. Isosurface representation of Env as seen from the side (left) and the top (right) (scale bar 10 nm).

Env-YFP lane), two species may still be present in the second lane (the band being the more soluble aptamer-SU domain complex, and the smear corresponding to aptamer-Env-YFP).

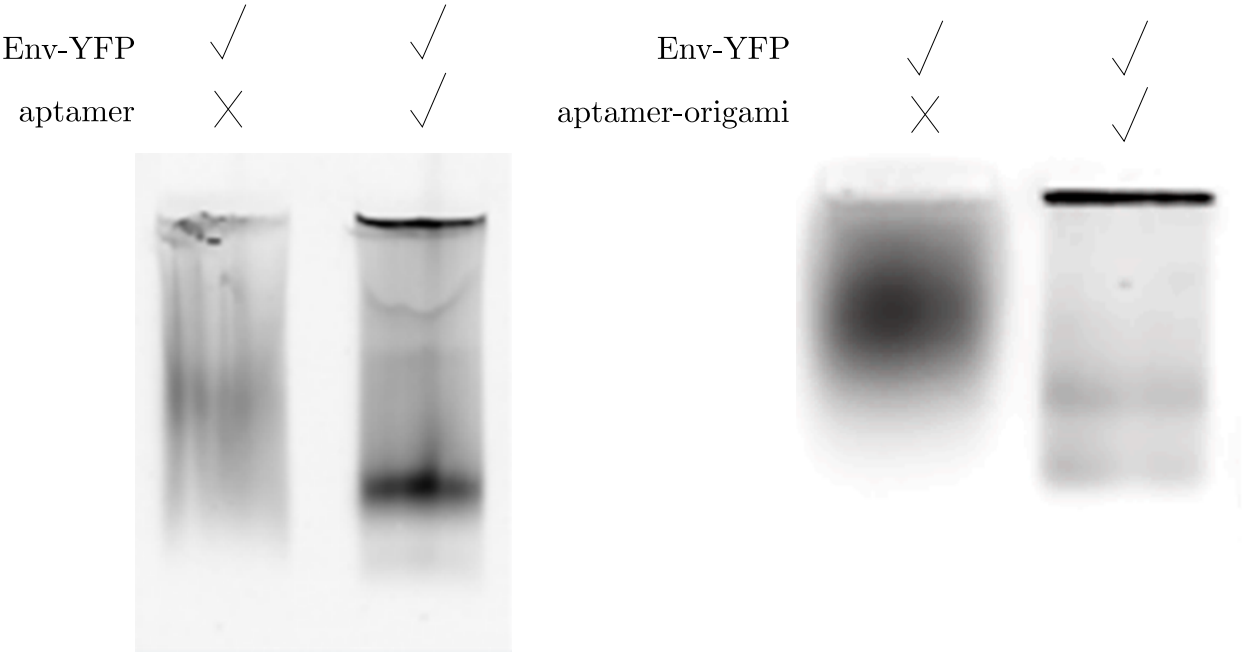


Figure 3.6: Gel electrophoresis of GFP aptamer binding to Env-YFP. Left: PAGE of Env-YFP with and without an excess of aptamer. Right: AGE of Env-YFP with and without an excess of aptamer-conjugated origami. Both gels are unstained and scanned for YFP fluorescence.

3.3.3 Negative Stain TEM of Virus Binding

Negative stain TEM was used as a preliminary check of apt-ori binding to the full MoMLV virus, to verify that the EYFP domain is still accessible when incorporated into the virus. Samples of apt-ori mixed with intact MoMLV virions were prepared for TEM. 1 μL of 20 nM apt-ori was mixed with 1 μL of virus prep, and incubated on ice for 15 minutes. The mix was diluted 1 in 20 immediately before application to the TEM grid. Grids were stained with 2% Uranyl Acetate as usual. Events of origami bound to the viruses are observed, as depicted in Figure 3.7, although these binding events are rare and a high background of unbound apt-ori is also observed (Figure 3.8a). One possible explanation for this is the domain shedding mentioned in Section 3.3.2, i.e. apt-ori may be bound to the shed SU domain complex, and therefore blocked from binding the virus. There are examples of what appear to be apt-ori bound to protein free in solution shown in Figure 3.8b, although the resolution is not sufficient to make a definitive statement about this. If this is the case, it will be important to purify the virus particles immediately before cryo-EM.

3.3.4 Cryo-ET of Virus Binding

Following these promising but inconclusive stain TEM results, apt-ori was prepared for cryo-EM, and concentration was measured as 500 nM. Origami were mixed 1:1 with a virus sample and imaged by CET. In the first set of tomograms, although likely binding events were observed, a high background of unbound origami was present, consistent with results from negative stain TEM (shown in Figure 3.9). CET was repeated by Vojta Pražák, with viruses purified by Capto Core columns immediately before grid preparation. This produced much cleaner tomograms, supporting the theory that high background results from virus shedding rather than low binding efficiency. Slices from an example tomogram are shown in Figure 3.10.

In this tomogram, the virus is almost completely surrounded by DNA origami nanostructures, which consistently point inwards towards the virus. In most cases, the glycoprotein bound to each apt-ori can also be identified. The surrounding origami 'shell' is toroidal in shape as the full ice thickness is only slightly greater than the virion diameter, thus there is not space for isotropic apt-ori binding. Some bending of the apt-ori 'posts' can be seen at the top and bottom of the ice (e.g. left highlighted origami in panel 1 of Figure 3.10) because of this space restriction. Origami crowding also seems to be an issue, so for samples with such a high density of targets, optimisation of origami tags may be required. In particular a longer post, and/or smaller sign, would likely improve labelling efficiency (signpost variant 2 in Figure 2.30 would be a good option). These results are promising for the application of origami tags to viral target systems. The fact that glycoproteins are visible, means that these tags would permit sub-tomogram averaging of the proteins. This would be beneficial for viruses with multiple glycoproteins which, without tagging, would be indistinguishable and thus impossible to independently reconstruct in this way.

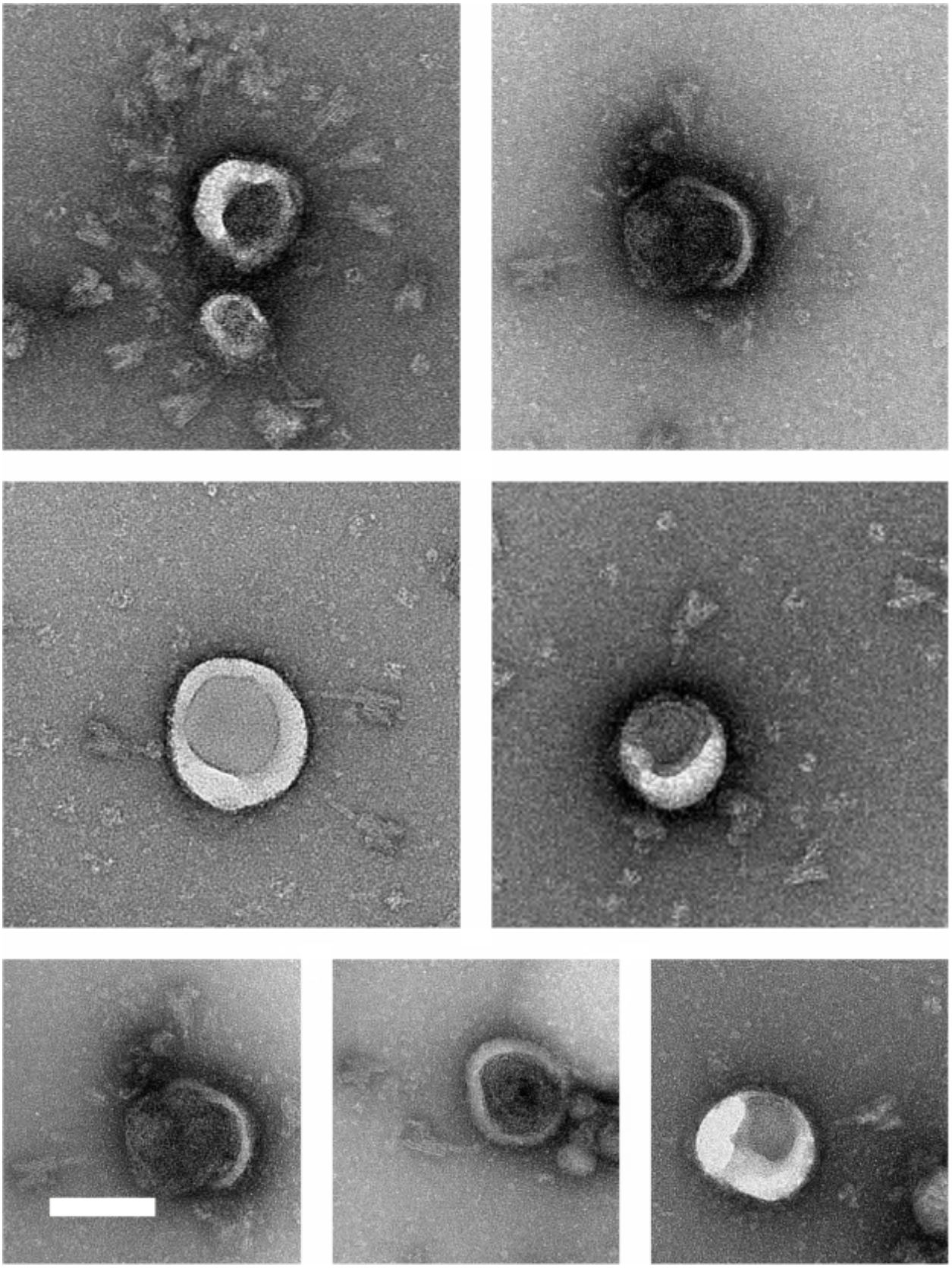


Figure 3.7: TEM views of selected instances of apt-ori binding to MoMLV (scale bar 100 nm).

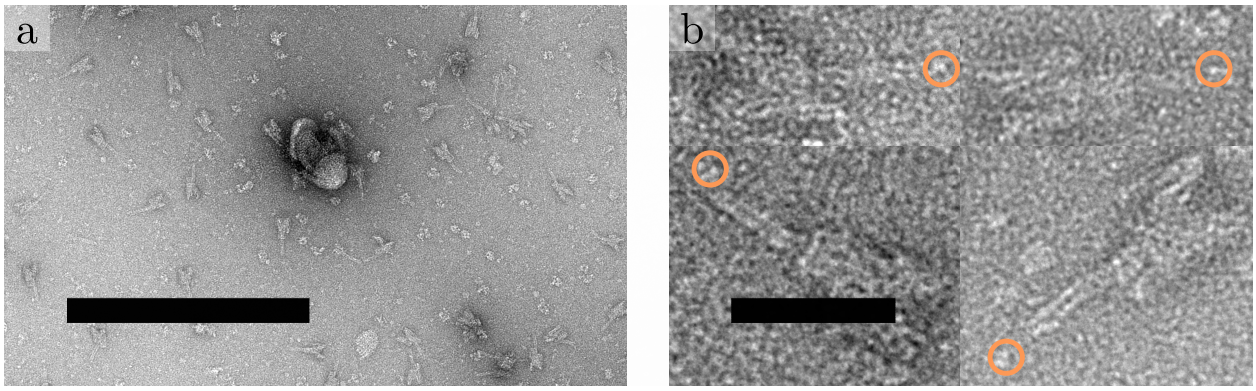


Figure 3.8: (a) Wild-field overview of apt-ori and MoMLV (scale bar 500 nM). There is a high background of free origami. (b) Examples of what appear to be apt-ori structures bound to shed protein domains, outlined in orange (scale bar 50 nM).

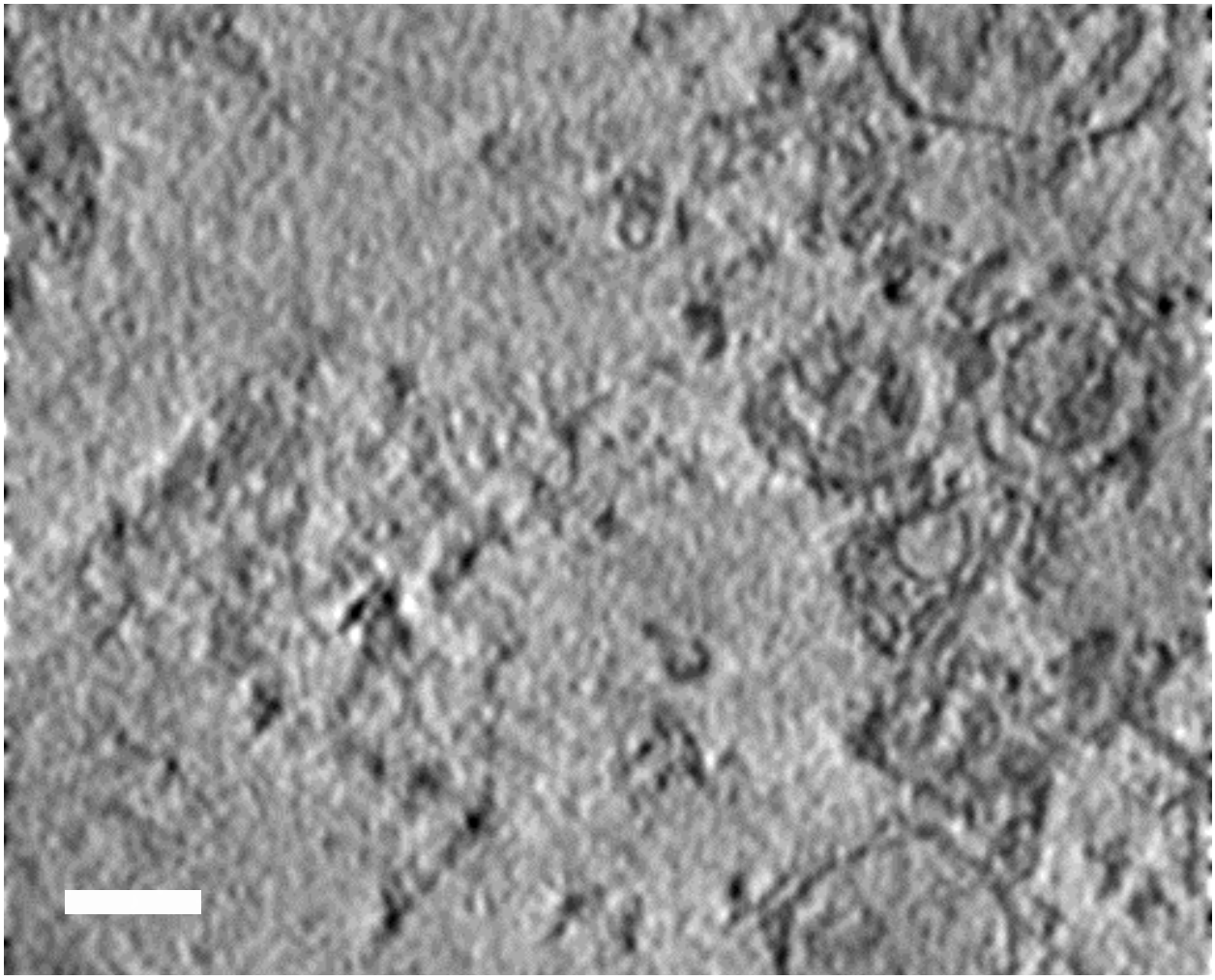


Figure 3.9: Preliminary CET of virus binding apt-ori. A crop from a busy region of a single tomogram slice is shown. A high background of unbound origami is observed (scale bar 100 nm).

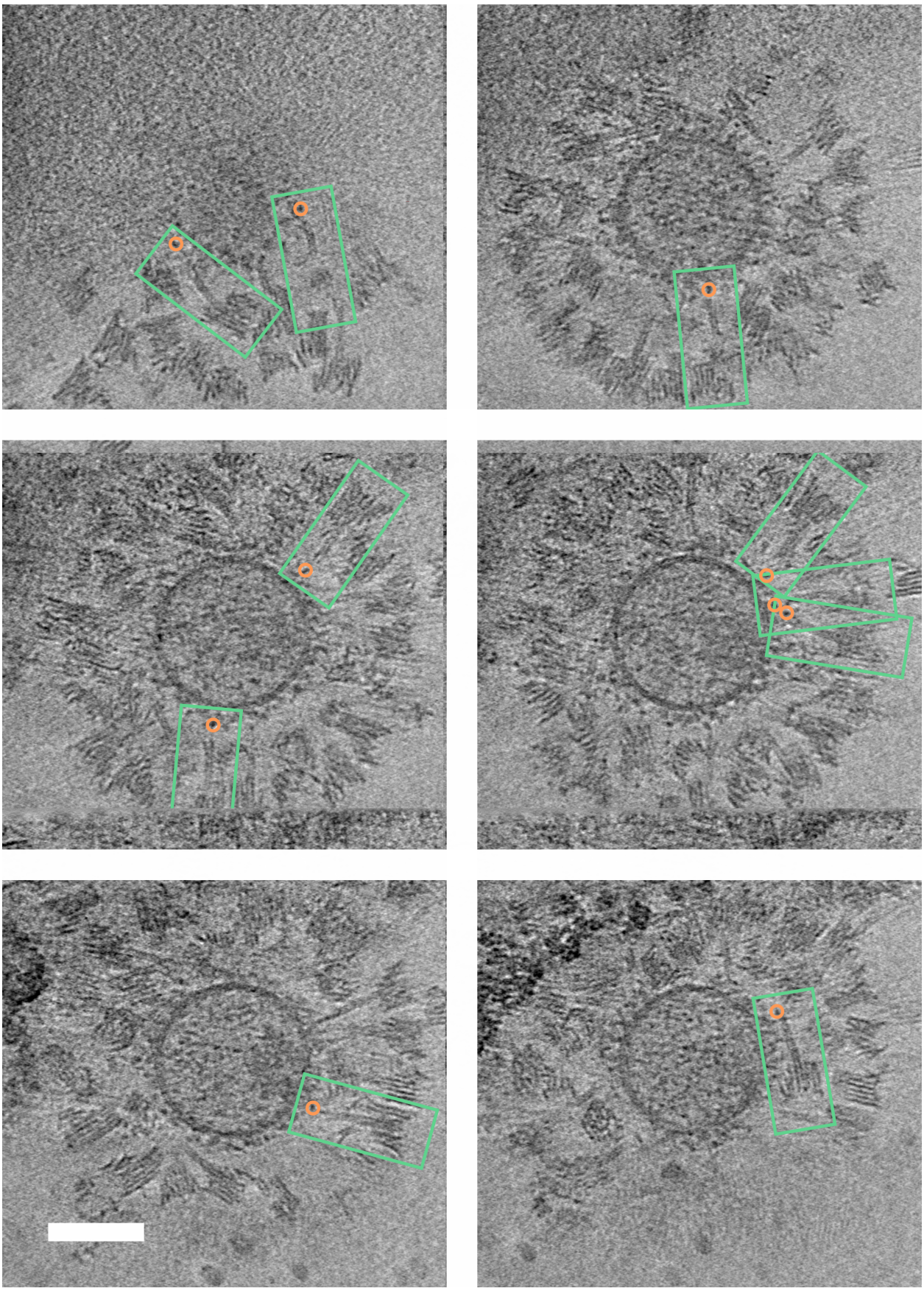


Figure 3.10: Tomogram slices of aptamer-conjugated origami binding to MoMLV. Examples of bound nanostructures are outlined in green, partnered bound Env-YFP are outlined in orange (scale bar 100 nm).

3.3.5 Cryo-EM of Cell Binding

The MoMLV system was also used to demonstrate cell-binding *in situ*. The MoMLV env protein interacts with the host cell through the receptor mCat-1 [146, 149] (Figure 3.11), an integral membrane protein with 14 transmembrane domains. This Env/mCat-1 interaction can be exploited to use apt-ori to tag the cell surface: cells expressing mCat-1 can be exposed to purified Env, which will bind to the receptors. Apt-ori will, in turn, bind to the Env protein, resulting in an indirect apt-ori/cell-surface interaction.

Apt-ori were prepared at 1 μM concentration by PEG precipitation. Cell washing prior to freezing means the final purification step is not necessary as excess PEG will be flushed out with other impurities. Cell-binding experiments were performed by Lindsay Baker and Vojta Pražák. XC rat sarcoma cells, stably expressing the viral receptor mCat-1 were grown on grids and mixed with Env-YFP and apt-ori. A representative micrograph is shown in Figure 3.12. Origami are easily visualised, suggesting that this tagging system works well for cell surface proteins. This tagging system also demonstrates the viability of indirect tagging mechanisms, since here the apt-ori/cell-surface interaction has two parts: the aptamer/YFP interaction and the Env/mCat-1 interaction.

3.4 Electroporation of DNA Origami into Cells

3.4.1 Motivation and Background

DNA origami nanostructure tags have been shown labelling viral and cell-surface proteins and could likewise be used to tag the external surface of isolated cell structures, but many new targets would be made available with nanostructure delivery to cells. There are a number of new challenges associated with intracellular labelling. The first potential concern is the stability of nanostructures within cells. A study testing degradation of DNA origami constructs exposed to multiple endo- and exonucleases found high stability of origami nanostructures compared to duplex DNA [69]. This is supported by a 2011 study by Mei and colleagues, which concluded that DNA origami structures, with a range of shapes and sizes, remain well-folded in cell lysate for at least 12 hours at room temperatures (whilst foreign biological single- or double-stranded DNA is degraded within an hour) [151]. It is therefore anticipated that nanostructures will remain intact within cells for the duration of tagging experiments, although targeting module stability may need to be considered independently.

The second concern is nanostructure toxicity, since a tag should cause minimal disturbance to the target system. Studies of origami delivery and the corresponding cell response have overwhelmingly focussed on mammalian cells due to the potential application of nanostructures as drug delivery vehicles. It is therefore not possible to make a general statement on cell toxicity, although numerous studies of non-loaded DNA origami nanostructures delivered to mammalian cell lines have shown no measurable decrease in cell viability [152, 153, 154, 155]. In 2011, Schuüller and colleagues investigated the potential of FDA origami tubes decorated with CpG sequences as immunostimulants [152]. Even in

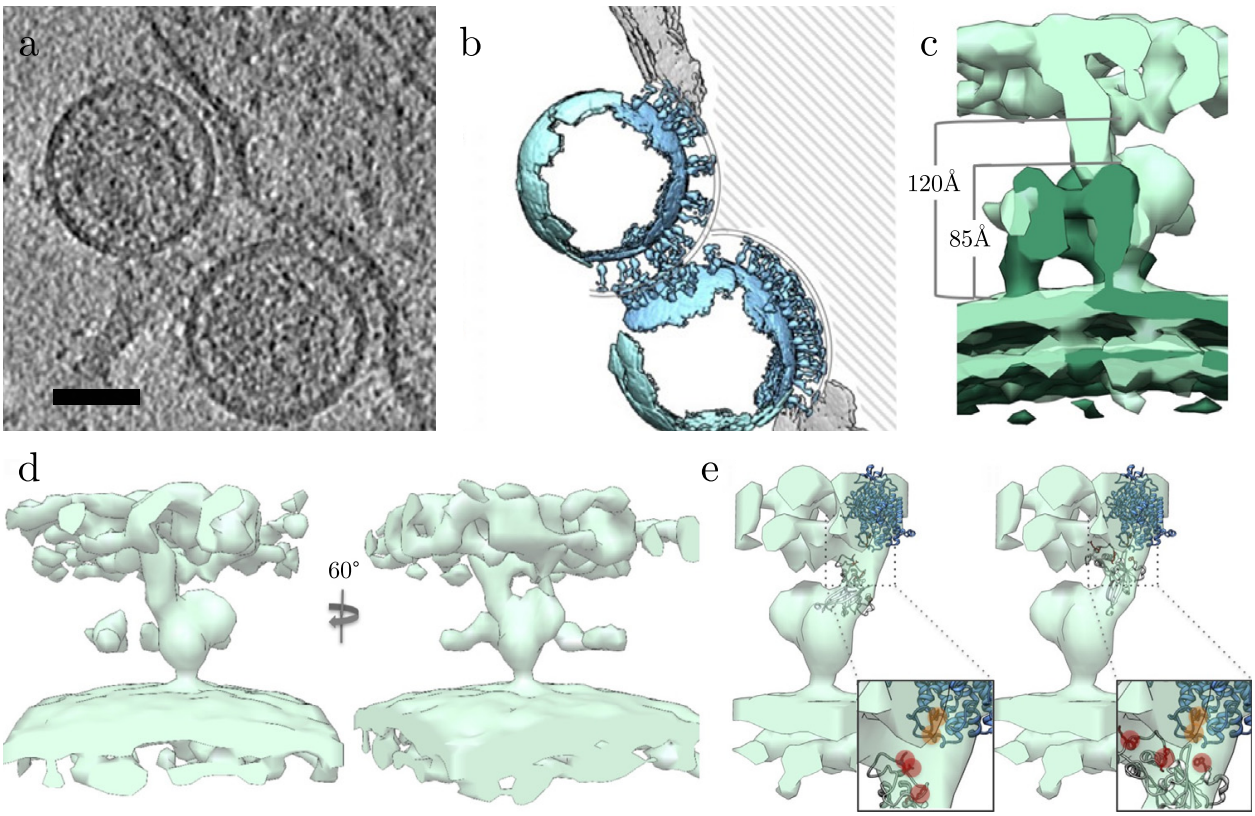


Figure 3.11: Env-YFP interactions with the host cell, figure and caption adapted from Riedel et al., 2017 [149]. (a) Slice through a tomogram showing two viruses in contact with the cell. Scale bar represnts 50 nm. (b) Backplot of final sub-volume averages onto (a) using the positions and orientations found for individual particles as part of the averaging. Interacting Env-YFP is depicted in light blue, non-interacting virus membrane in cyan and non-interacting cell membranes in grey. (c) Surface representation of the EM density map of interacting Env-YFP filtered to 50Å. (d) Two different putative localisations of the receptor binding domain (RBD) in the interacting Env-YFP EM density map are shown . RBD receptor interacting residues are colored red. Further, a putative structure of mCat-1 (generated using the Phyre server [150] has been added with the Env interacting residues shown in orange).

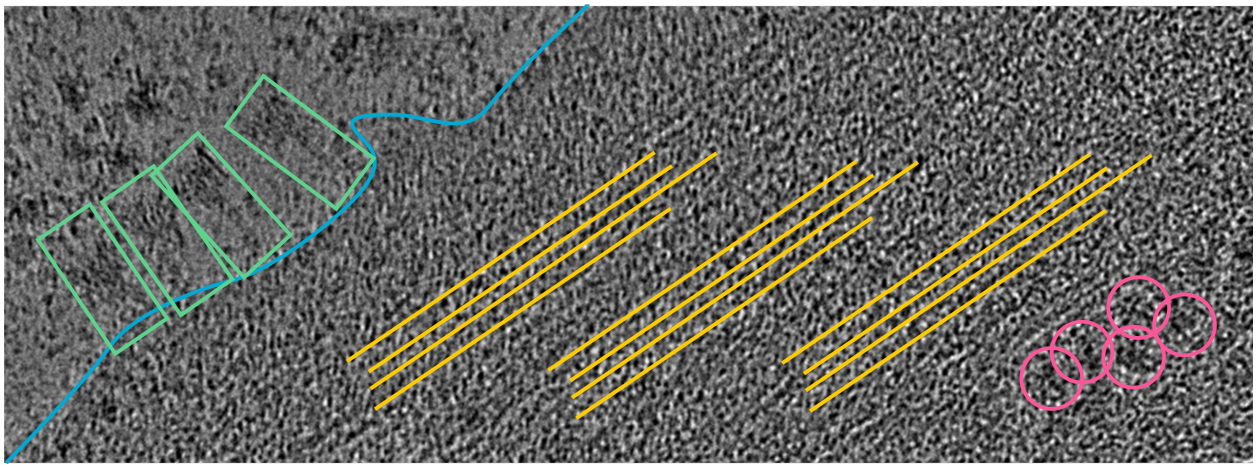


Figure 3.12: Cryo-EM projection image of thin ice at the edge of a cell protrusion. Origami tags are shown in green boxes, the blue line indicates the cell boundary, yellow lines indicate actin bundles and pink circles show ribosomes. Scale bar 50 nm.

the absence of CpG strands, nanostructures can activate immunity via non-TLR9-mediated pathways. Such responses would need to be analysed and addressed for intracellular tagging.

The third, and perhaps most significant, hurdle is efficient delivery of nanostructures. For tagging applications, it will be necessary to deliver many nanostructures per cell. A potential solution is electroporation, where cells are subjected to a brief, high-voltage pulse which increases membrane permeability and allows DNA (or other molecules) to enter the cell. This a common strategy for cloning, and has been demonstrated for large (>100 kb) DNA molecules [156]. For DNA origami delivery, electroporation of nanostructures stored in standard buffers results in disassembly, likely due to the electric field stripping away Mg^{2+} ions which are crucial for maintaining structural integrity [129]. Origami can be stabilised by higher valency (3+ or above) polyamines [129, 130] in Mg-free buffers, which has been exploited to electroporate origami nanostructures into mammalian cells [129].

Targeting within eukaryotic cells is complicated significantly by internal cellular compartments, so initial experiments will focus on nanostructure deliver to prokaryotic cells.

3.4.2 Stabilising Nanostructures with Polyamines

The experiments presented in this Section were performed with the wedge origami nanostructure presented in Section 2.3 (application to different structures may require new conditions). In 2016, Chopra and colleagues reported electroporation of DNA origami nanostructures, stabilised by spermidine ions, into mammalian cells [129]. For my intended application of electroporation into prokaryotic cells, I have investigated spermine (sp^{4+}) and spermidine (spd^{3+}), both polyamines, as stabilising agents.

To avoid re-optimising folding and purification conditions, origami were annealed as usual in 1×TE + 12.5 mM $MgCl_2$ and transferred into polyamine solutions post-folding. After

annealing and purification, origami were concentrated[5], and then diluted into solutions of sp^{4+}/ spd^{3+} to a final origami concentration of 10 nM and a final $MgCl_2$ concentration of 0.55 mM (which is not sufficient to stabilise origami without the polyamines). Different concentrations of sp^{4+} and spd^{3+} were investigated, both in water and in TE buffer. Results are shown in Figure 3.13a, b, d and e. Origami remains well-folded in 50 μM - 10 mM of spd^{3+} in TE or 10 μM - 1 mM of spd^{3+} in water, and likewise remains well-folded in 10 μM - 1 mM of sp^{4+} in TE or 1 μM - 1 mM of Sp^{4+} in water. For Yb^{3+}-stabilised nanostructures presented in Section 2.6.1, a difference in nanostructure mobility was observed, attributed to a change in nanostructure charge from bound counter-ions. Here this is not seen, presumably due to the lower charge density of the polyamines (the same results were observed by Chopra et al. for spermidine stabilisation [129]).

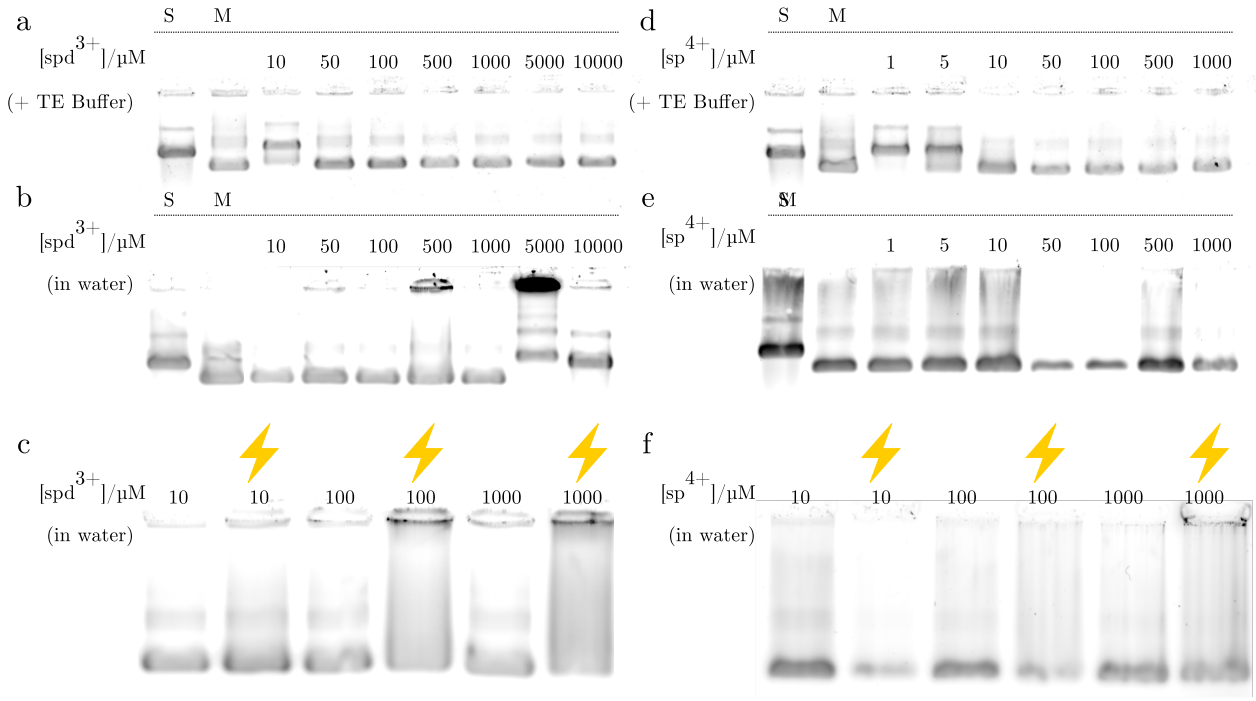


Figure 3.13: Overview of origami stabilisation by polyamines. (a), (b), (d), (e) AGE of origami nanostructures prepared in 1× TE + 12.5 mM $MgCl_2$ (lane 2), diluted into solutions of spermine and spermidine with and without TE buffer (as labelled). M13mp18 is shown in lane 1 for reference. (c)/(f) Origami in solutions of spermidine/spermine (at labelled concentrations) before and after a 2.5 kV electric field pulse.

Unbuffered solutions were used to minimise the presence of unnecessary ions (which can cause arcing), and the stability of samples when exposed to electric field pulses was investigated by gel electrophoresis. 100 μL samples were pipetted into 0.2 cm cuvettes and exposed to an electric field pulse suitable for electroporation of *E. coli*.[6] Samples before and after the pulse were analysed by gel electrophoresis (Figure 3.13c, f). Origami in low (10 μM) concentrations of spd^{3+} remained well-folded, but at higher spd^{3+} concentrations misfolding/aggregation was observed, whilst origami remained folded for a range of sp^{4+} concentrations. Based on these results, a 10 μM sp^{4+} concentration was used

[5] 30 minute spin with 100 kDa Amicon filters.
[6] Micropulser Electroporation Apparatus, 2.5 kV exponential decay pulse.

for electroporation.

3.4.3 Fluorescence Microscopy of Electroporated Cells

Wedge-origami nanostructures were annealed with Cy3-conjugated strands, which were incorporated by hybridisation to staple extensions (see Appendix A.3.1 for details of staple mixes).[7] 10 Cy3-labelled strands were conjugated to each origami nanostructure. We expect this to be sufficient to visualise single origami structures (Vojta Pražák, personal correspondence).

Origami nanostructures were purified by spin filtration (Appendix C.2.1), and prepared in a 10 μM solution of sp^{4+}. Electrocompetent *E. coli* cells were prepared according to the protocol laid out by Gonzales et al. [157]. Two strains were investigated: DH5α and wild-type MG1655. Electroporation was performed with the same pulse conditions outlined above, with final origami concentrations between 2 and 10 nM and a final sp^{4+} concentration of 10 μM. Cells were pelleted (to remove free origami structures), stained with the membrane dye DiO, and analysed by fluorescence microscopy (Appendix D.4). Examples micrographs are shown in Figure 3.14.

3.4.4 Discussion

A number of possible electroporation events can be identified, but further work would be required to confirm that we are legitimately detecting single origami nanostructures within cells.

From these images, it is unclear if the fluorescent signal emanates from the inside or outside of cells; it is possible the nanostructures are adhered to the cell surface. For both strains the fluorescent signals localise preferentially to the poles. This could indicate that the nanostructures have entered the cell and are compartmentalised, or that they attach outside of the cell with a particular curvature preference.

Assuming the fluorescent signal is internal, it also does not indicate that the nanostructures remain well-folded (although this assumption is supported by AGE results in Figure 3.13), and could be alternatively explained by uptake of residual unbound fluorescent oligonucleotides. In the following discussion, I will assume that fluorescent signals indicate internalised, well-folded origami nanostructures, but these outstanding uncertainties would need to be addressed in any further work.

We cannot know that a fluorescent signal emanates from a single nanostructure (since the nanostructure size is smaller than the imaging resolution), but based on the number of observed fluorescence events, I propose that, in most cases, a single nanostructure will be the source of fluorescence. Similar uptake efficiencies were observed for the two strains,

[7] 10 staples with 5′ extensions were included at 10× excess over the scaffold. The Cy3-labelled oligonucleotide was included at 200× excess over the scaffold (i.e. 2× excess over complementary staple domains.)

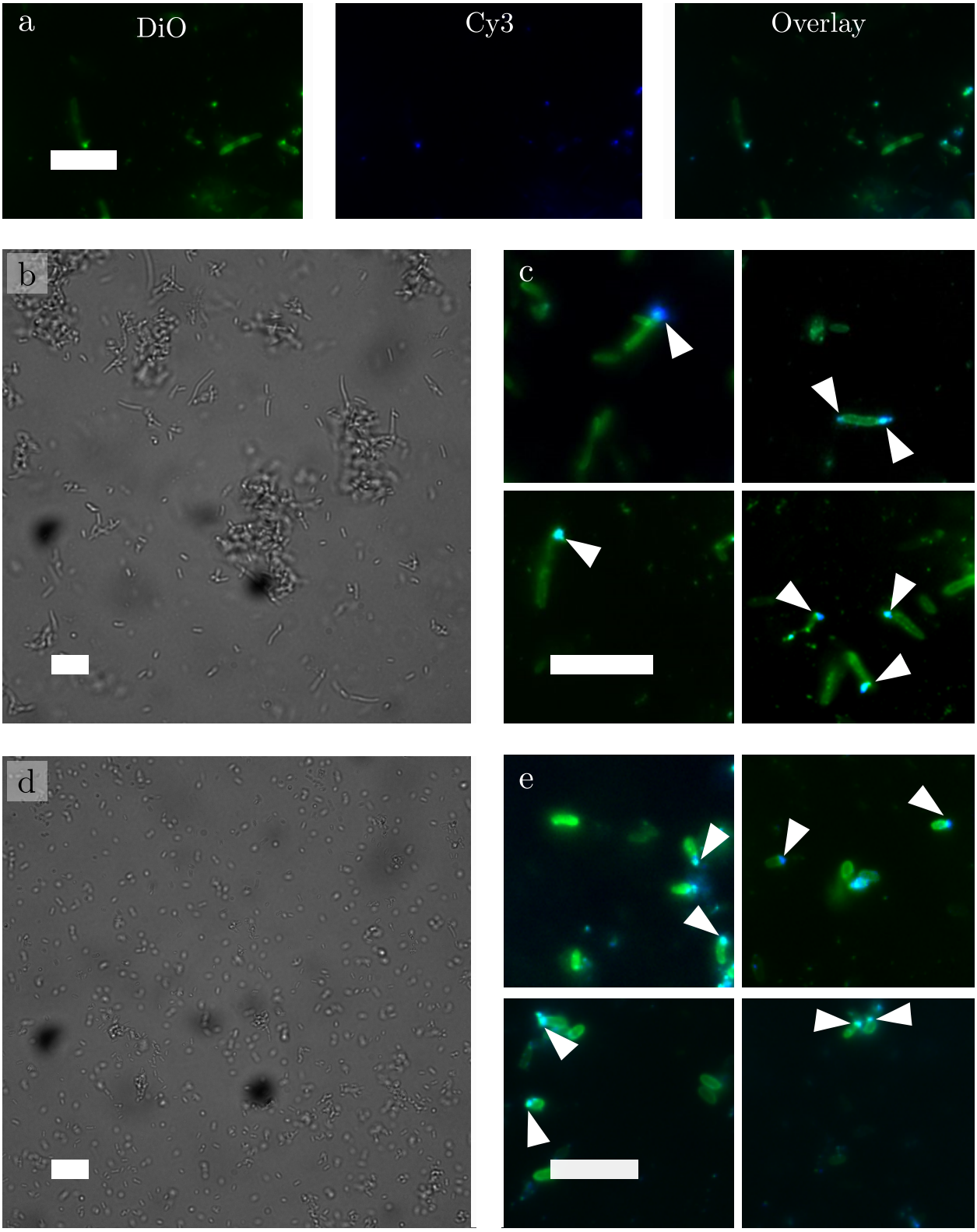


Figure 3.14: Cy3-labelled origami nanostructures were electroporatied into E coli.. Scale bar 5 μM. (a) DiO (membrane dye) and Cy3 channels are shown in green and blue respectively. (b)/(d) Confocal image of electroporated DH5a/MG1655 E coli. cells. (c)/(e) Characteristic fluorescent microcroscopy images of DH5a/MG1655 E coli. cells. Proposed electroporation events ar indicated by white arrowheads.

although with DH5α, elongated cells were observed post-electroporation. This is usually symptomatic of a stress response, thus MG1655 cells are more suitable recipients. Proposed electroporation events are usually one per cell, with occasional observations of 2, and this was not improved by increasing origami concentrations from 2-10 nM. This efficiency would not be suitable for tagging. Copy numbers may be improved with optimisation of strains and electroporation conditions, although a drastic increase in efficiency would be required.

There are a plethora of alternative options that could be explored in the future for delivery of nucleic acid nanostructures [158], although most methods would be expected to have more severe efficiency constraints. For cryo-EM applications a microfluidic delivery platform [159] may be a suitable option. Cells are passed through constrictions smaller than their diameter and the resulting compressional/shear forces result in transient pores which allows uptake of the surrounding medium. Efficiencies 10-fold higher than electroporation have been observed with this technique, and optimisation of cell speed and size and number of constrictions for origami delivery may give sufficient copy numbers for intracellular tagging experiments.

A more elegant solution to intracellular tagging is RNA nanotechnology. RNA nanostructures (with tagging properties), capable of self-assembling isothermally as they are transcribed *in vivo*, could be genetically encoded as tags (and strains expressing the structure can therefore be selected for). This would bypass the need for complex delivery mechanisms, and allow copy number to be controlled through promoter strength. The development of co-transcriptionally folded 3D RNA nanostructures, with this tagging end-goal in mind, will be the topic of the next two chapters of this thesis.

3.5 Summary and Future Directions

In this chapter, I have verified the binding behaviour of an RNA aptamer to EYFP, and quantified its binding affinity to new GFP analogues by isothermal titration calorimetry. I have shown that hybridisation of this aptamer to a signpost origami nanostructure results in a functionalised tag (apt-ori), suitable for CET.

An MoMLV test system with YFP-tagged Env glycoproteins was used to demonstrate the targeting behaviour of the apt-ori constructs. Tomography results of apt-ori bound to the virus demonstrate high-efficiency and high-contrast labelling, illustrating the usefulness of the apt-ori tag for viral target systems. Apt-ori/cell binding has also been demonstrated with cells expressing the mCat-1 receptor protein, showing that tags can be used for cell surface labelling and via an indirect tagging interaction (since this interaction consists of two parts: the aptamer/YFP and the Env/mCat-1 interfaces).

There are many interesting systems that can be studied via this method. We are currently working on using these tags to characterise the FLRT2:Unc5:Latrophilin complex involved in cell adhesion and repulsion in neurones [160]. This involves using the established GFP aptamer with mVenus-tagged proteins, in conjunction with a covalent approach to tagging based on enzymatic coupling. In this new approach, origami nanostructures are functionalised with chemically-modified oligonucleotides

(benzylguanine/benzylcytosine/HaloTag-Ligand) which form covalent linkages with peptide tags (SNAP/CLIP/Halo), which can be introduced into the ectodomain of the target proteins. Colleagues have also adopted this technique to investigate the glycoprotein distribution on the HSV-1 surface, with focus on characterising the Herpes entry complex [161].

Barriers remain for intracellular labelling due to low delivery efficiency by electroporation (Section 3.4.3) and remaining uncertainties with interpreting fluorescent micrographs. In Section 3.4.4, I have suggested alternative routes toward tag delivery, but a more elegant solution is to design RNA nanostructure tags which can be transcribed *in vivo* from a cloneable DNA template. The first steps toward this goal will be the topic of the next two chapters.

Chapter 4

RNA Nanostructures: Grid-based Approach

In this chapter, I will present a novel method for building co-transcriptionally-folded RNA nanostructures in 3D. Firstly I will introduce the 'grid-like' design architecture, and demonstrate the technique by designing and building a pentomino nanostructure. Based on these preliminary results, I will show a modified version of the design architecture which is used to construct a number of nanostructures which interact to form programmable supermolecular assemblies with up to five components.

4.1 Author Contributions

Lindsay Baker performed the cryo-EM (Figure 4.12) in this chapter. The rest of the work in this chapter (design and synthesis of nanostructures, AGE and negative-stain TEM characterisation of nanostructures) was performed by Emma Silvester.

4.2 Project 2 Aim: 3D RNA Nanostructures by Co-transcriptional Folding

In Chapters 2 and 3, I have demonstrated the design and functionalisation of DNA origami nanostructures as tags for cryo-EM. These nanostructures have great potential as exogenous tags, e.g. for viral or outer membrane targeting, but they are not ideally suited for endogenous tagging. With some optimisation, electroporation (Section 3.4), 'cell-squeezing' [159] or other delivery methods may offer a suitable workaround, but a more comprehensive and elegant solution is to use RNA nanostructures which fold co-transcriptionally *in vivo* causing minimal disruption to the cell. Targeting properties could be introduced to RNA nanostructures through incorporation of RNA aptamers such as the GFP aptamer used in Chapter 3. A genetically-encoded tag would be easy to incorporate into biological systems, offering breadth of utility for cryo-EM applications, comparable to the use of fluorescent

protein fusion tags for light microscopy.

Unlike for DNA nanotechnology, there is not a wealth of well-established and widely-used architectures for RNA nanostructures. A genetic tag must be single-stranded (or transcribed from a small number of genes) and able to fold co-transcriptionally/isothermally, which presents additional challenges to the design process. I have shown in Section 2.4.2 that to image DNA nanostructures in mammalian cell lysate (a proxy for an intracellular environment), a thickness of at least 4 helices is required to provide adequate contrast. It will therefore be necessary to design structures which fold into compact three-dimensional structures of at least these dimensions.

In this project (Chapters 4 and 5) I will present novel design architectures for three-dimensional co-transcriptionally-folded RNA nanostructures, with downstream applications as endogenous cryo-EM tags in mind. In this chapter, I will introduce a grid-based approach to designing nanostructures inspired by supermolecular assembly, and in Chapter 5 I will investigate ways of extending an established 'RNA origami' architecture into 3D.

4.3 Preliminary Design: A Pentomino Nanostructure

4.3.1 Design

Concept

Much of the progress in the structural RNA nanotechnology field has been been based on supermolecular assembly of structures through tertiary motifs extracted from resolved structures of biological RNAs. Examples include the RNA tectosquares [107] and polyhedra [108] discussed in Section 1.4.3. Inspired by this, but wishing to retain a single-stranded approach, I have developed a 'pseudo-supermolecular' design methodology for grid-like RNA nanostructures, which are comprised of interacting units arranged on a square lattice and connected by flexible, single-stranded linkers.

A schematic depiction of a unit in this design is shown in Figure 4.1a. Each unit consists of a series of stacked three-way junctions (3WJs), where two contributing helices (H1 and H2) are coaxially stacked, and thus parallel, and the third helix (H3) lies at approximately 90° to the coaxial stack. H2 of one junction becomes the H1 domain for the next. H3 domains are capped by hairpin loops. The result is a long, contiguous hairpin with further hairpin domains branching out at right angles. The length of the helical stretch between two adjacent H3 helices (i.e. H2 for the previous junction and H1 for the following junction) is chosen such that successive H3 helices lie at 90° from one another in the plane perpendicular to the H1/H2 coaxial stack.

Neighbouring units can associate by kissing loop interactions between complementary H3 hairpins. In this way one unit can theoretically interact with four neighbours (four compass directions in the plane perpendicular to the H1/H2 axis). Each pair of interacting units has at least two kissing interactions, to increase binding strength between units and to keep them

parallel. Units are also linked through the nanostructure by the scaffold routing: the 3′ end of one unit will link to the 5′ end of the next unit by a flexible single-stranded RNA linker. Thus the structure is designed in a 'pixel' format, where desired pixels (i.e. unit positions on a square lattice) are identified, a valid scaffold routing through the pixels is drawn out, and orthogonal kissing loops sequences are specified for consistency with the design. An overview of this design concept is shown in Figure 4.1.

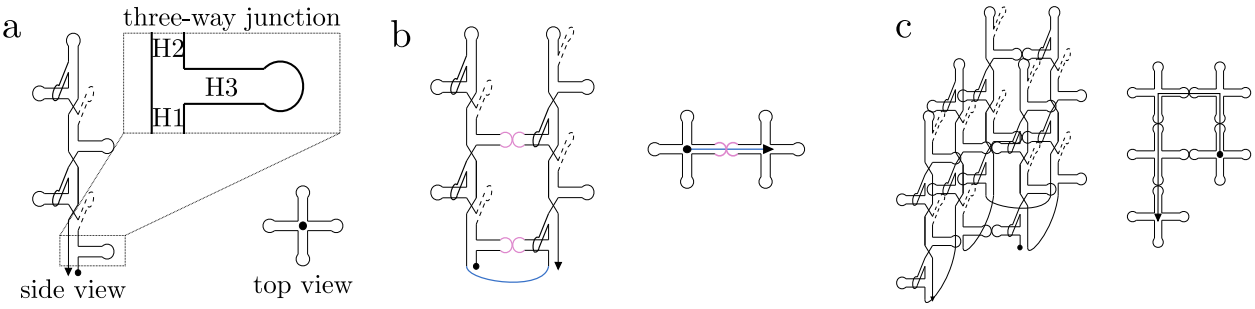


Figure 4.1: Schematic overview of grid-like RNA nanostructure design. (a) Side-/top-view of a single unit. Inset shows a three-way junction, with contributing helices H1, H2 and H3 labelled. (b) Neighbouring units are linked by the strand routing (blue) and kissing interactions between complementary hairpins (pink). (c) This concept can be extended to more units to generate nanostructures.

This modular architecture is designed with co-transcriptional folding in mind. An individual unit has a simple folding pathway, and should find its minimum-free-energy (MFE) secondary-structure quickly as it is transcribed. The kissing loops will then form between complete units on a longer time scale, not interfering with secondary structure formation.

Overview of Motifs and Atomistic Model

The key motifs of this structure (outside RNA A-helix regions) are 3WJs, kissing loops, and single-stranded linkers. The 3WJ I have used in this first design is extracted from the PDB file of the Thermus thermophilus 70S ribosome (PDB ID: 2J02) [162], and possesses almost the desired geometry (which would be contiguous H1 and H2, with H3 at 90° to the H1/H2 axis). The main difference is that H1 and H2 are not quite contiguous, although they are parallel. By matching[1] longer helical domains[2] to H1/H2/H3 regions on the junction PDB file, and positioning successive 3WJs, it was determined that a 12 base pair separation (i.e. H1/H2 length) between adjacent junctions produces arms (H3 duplexes) at right angles from one another in the plane perpendicular to the unit axis. The secondary structure and 3D structure of the 3WJ is shown in Figure 4.2.

After matching 3WJs with 12 base pair H1/H2 domains, to give a 3D unit structure, it was determined that a 120° kissing loop has the correct geometry to link neighbouring parallel units. This discrepancy, compared to the expected 180° angle in the original concept design

[1] Alignment of PDB files was performed in Chimera [124]. The method of matching motif PDBs in this way is described in the first part of the RNA nanostructure design procotol by Sparvath and colleagues [115].

[2] PDB files of RNA A-helices obtained from the make-na server: http://structure.usc.edu/make-na/server.html.

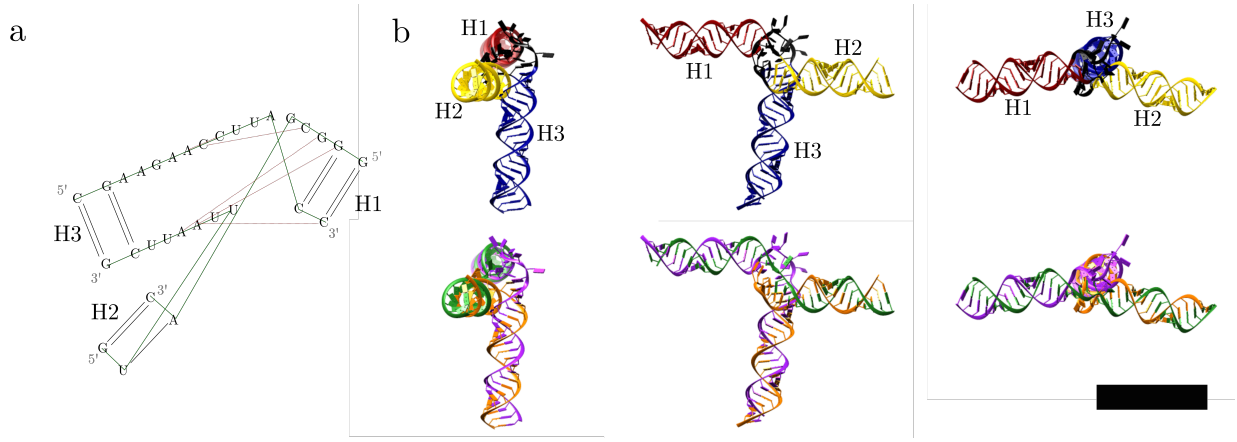


Figure 4.2: (a) Sequence and structure of wild-type 3WJ. Backbone, secondary structure and non-Watson-Crick tertiary contacts depicted in green, black and red respectively. (b) and (c) Rendering of 3WJ with extended arms generated in UCSF Chimera [124]. In (b) the three contributing helical domains are shown in red, blue and yellow, and junction nucleotides are shown in black. In (c) The three contributing strands are shown in purple, orange and green (scale bar 5 nm).

(Figure 4.1), is primarily because H1 and H2 of the 3WJ are not co-axial, so the unit axis lies at an angle ($\approx 30°$) to the helical axes. This results in a favoured 120° kissing geometry between units. The angles for one junction are depicted in Figure 4.3. For simplicity, schematic depictions of nanostructures are still drawn with contiguous H1/H2 domains and 180° kissing loops.

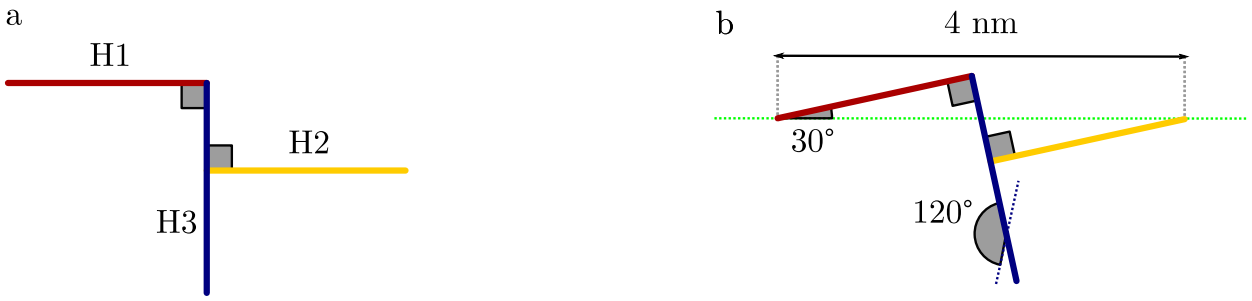


Figure 4.3: (a) 2D schematic depiction of 3WJ. A unit comprises a set of stacked junctions, i.e. another copy of H1 follows from H2. The combined H1/H2 helical domain is 12 base pairs. (b) Overview of angles. Because the H1 and H2 axes are not contiguous, the unit axis (green dotted line) lies at an angle relative to the helical axes ($\approx 30°$). As a result, the a 120° kissing loop is preferred for unit/unit interactions.

The 120° RNAI/II kissing complex (PDB ID: 1BJ2) [163] has been used previously by Luc Jäger in nanostructure design, who has developed many orthogonal sequences with the same interaction geometry [112, 164], which simplifies sequence design considerably. Depictions of the wild-type 120° kissing loop are shown in Figure 4.4. Matching of junctions and kissing loops to create multiple aligned units suggests that the best alignment results from having two distinct unit types: type A and B, where the hairpin contributing to the kissing loop is located at different positions on the H3 helices/branches; H3 lengths are 8 and 18 base pairs for units A and B respectively. Within a nanostructure, A and B units will be positioned in a

alternating grid, such that A units interact only with B units and vice versa. Atomic models of both units were generated using a strategy based on Sparvath's method of designing RNA tiles [115]. In brief:

1. Motifs (junctions, kissing loops and 3WJs) were positioned into the unit structures by matching helical regions in UCSF Chimera [124].
2. The matched motifs were loaded in SwissPdb Viewer [165]. Duplicate nucleotides were removed, and all remaining nucleotides were renamed using the 'Rename Current Layer' tool. Nucleotides are numbered from 5′ to 3′ and labelled as belonging to the same chain.
3. The numbered PDB file of matched motifs was ligated into a single-strand using a MATLAB script (see Appendix F.2), which is more general than the Perl script used in Sparvath's design methodology.

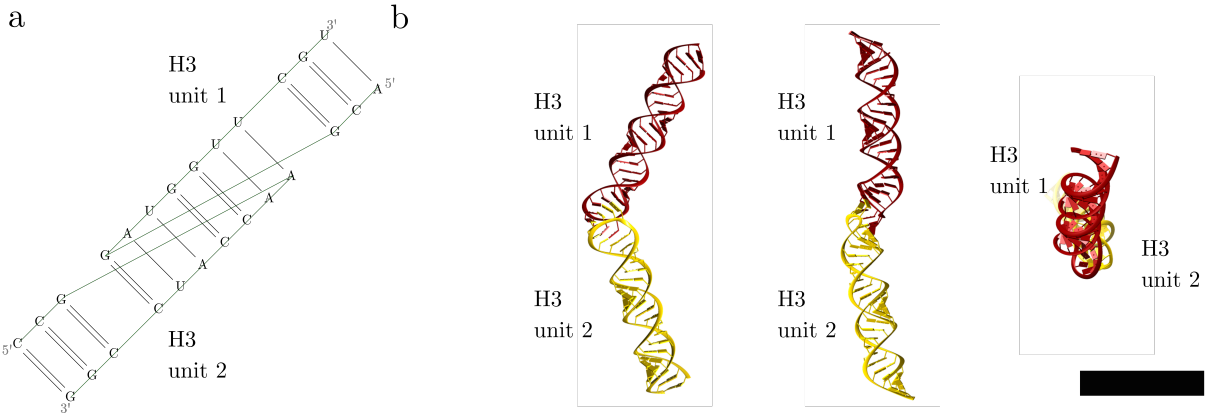


Figure 4.4: (a) Sequence and structure of wild-type 120° kissing loop. Backbone, and base pairs depicted in green and black respectively. (b) 3 orthogonal views of kissing loop with extended arms, rendered in UCSF Chimera [124]. The two contributing helices are shown in red and yellow (scale bar 5 nm).

Renders of units type A and B are shown in Figure 4.5a and b. The first nanostructure I have designed is a 'pentomino' of 5 units, arranged into an asymmetric 'P' shape. A top-view of 5 units arranged in this way is also shown in Figure 4.5c.

The final motifs that need to be incorporated into nanostructures are single-stranded linkers. Measurements of single-stranded RNA lengths at different ionic concentrations [166] give a persistance length (l_p) of 16.4 ± 0.4 Å at a Mg^{2+} concentration of 6 mM (typical for *in vitro* transcription reactions). This work also provided a contour length of 225.6 Å for rU_{40}, i.e. 5.64 Å per nucleotide (a quantity named l_0 below). Therefore, using a worm-like chain model, for a sequence of N unpaired nucleotides the mean-squared end-to-end distance $\langle R^2 \rangle$ is given by:

$$\langle R^2 \rangle = 2 l_p N l_0 = 185 N \text{Å}^2 \tag{4.1}$$

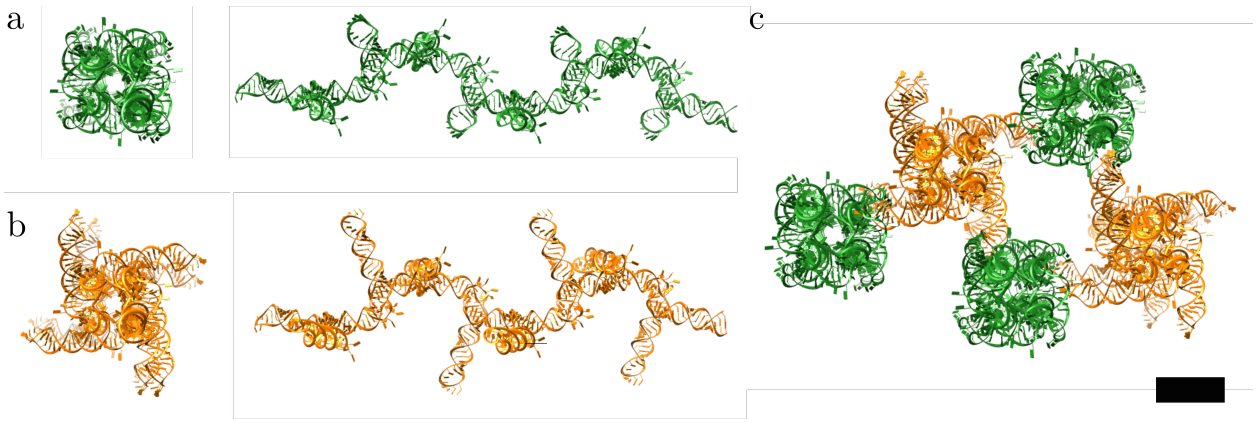


Figure 4.5: Overview of unit structures. (a)/(b) Two orthogonal views of atomic model of unit types A/B. (c) View of 5 interacting units. Models generated in UCSF Chimera [124], scale bar 5 nm.

The separation between units in the atomic model is ≈ 85 Å. Requiring $\sqrt{\langle R^2 \rangle} = 85\mathring{A}$, we find:

$$N = \frac{\langle R^2 \rangle}{2 l_p l_0} = \frac{85^2}{185} = 39 \tag{4.2}$$

A linker length of 39 nucleotides has been used in this design, such that the mean end-to-end linker length will be equal to the required unit-to-unit separation in the transcription buffer. Intramolecular kissing loops should be able to form without any significant stretching or compression of linkers.

Sequence Design

Sequences are designed in a modular fashion, with a view to co-transcriptional folding. It is expected that individual units will fold as they are transcribed, and thus can be designed independently for local free energy minima (rather than designing the whole structure at once). Units are expected to be trapped kinetically in this fold, and so not interfere with the secondary structure folding of other units. Sequence design for individual units is, once again, inspired heavily by Sparvath's design protocol [115]. In brief:

1. 3D structures are converted to secondary structures.
2. Unit sequences are designed in NUPACK [167] with appropriate sequence constraints.
3. NUPACK sequences are evaluated by mfold [171] and optimised by hand.
4. Flanking sequences for transcription and cloning are introduced for the gBlock template.

The secondary structure of both unit types were derived in dot/bracket notation from PDB files via the RNApdbee web server [168, 169] and checked against the expected structure.

Careful sequence specification of kissing loops is important, since this determines the global architecture. The arrangement of units comprising the nanostructure is shown in Figure 4.6a and b. Type B units are rotated 180° around the unit axis direction relative to Type A units such that the H3 hairpins line up with their partners along this axis. There are 5 unit/unit interaction interfaces (1-2, 1-4, 2-3, 3-4, 4-5), each of which comprises 2 kissing loops. I opted to use the same sequence twice for these interfaces to minimise the number of orthogonal sequences needed (mismatches should be prevented sterically).

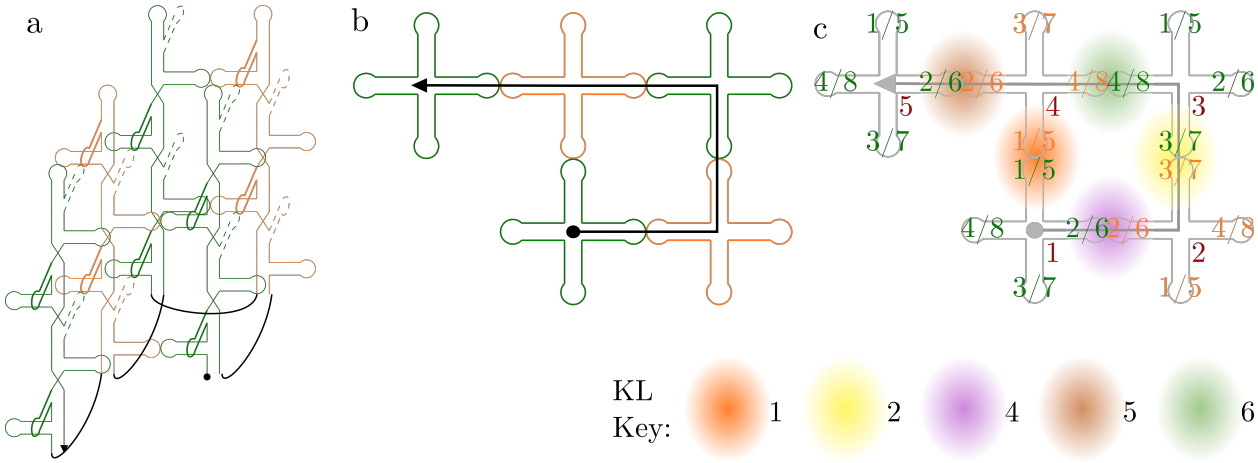


Figure 4.6: Schematic view of pentomino nanostructure. (a)/(b) Side/top view of pentomino with type A units shown in green and type B units shown in orange. The arrow points from 5′ to 3′. (c) Schematic with kissing loops indicated by coloured ovals. Units are labelled in red from 1-5. For each unit, arms are labelled from 1-8 in green/orange for type A/B units.

A list of kissing loop sequences (with the required 120° geometry) used by Jäger [112, 164], with their measured dissociation constants is shown in the Table 4.1. Kissing loop sequences from these examples are assigned to H3 arms as specified in Table 4.2. Half of the H3 helices are external to the structure, these are 8 base pairs regardless of unit type and are capped with stable *UUCG* tetraloops (PDB ID: 1F7Y) [170]. The start and end duplex regions (i.e. the first H1 domain and final H3 domain) are also limited to 8 base pairs, and the final H3 domain is also capped with a *UUCG* tetraloop.

Individual units were initially designed with the NUPACK web server 'design' module [167]. The secondary structure of the unit is specified (with any tertiary contacts instead specified as single-stranded, since the NUPACK web server does not permit pseudo-knotted designs), and the sequence is largely unspecified, with the exception of kissing loops sequences, *UUCG* tetra loops (capping external H3 helices, and the final H2 domain), sequences for the 3WJ motifs, and the *GGGAGA* T7 polymerase starter sequence (which is included at the start of unit 1 only).

Sequences with low normalised ensemble defects (NEDs) and reasonable (< 65%) GC content were selected for further modification. Secondary structures of possible folds for

Kissing Loop	Loop A sequence	Loop B sequence	K_D/nM
1	*GUCCACU*	*AGUGGAC*	6
2	*GCAGGCU*	*AGCCUGC*	18
3	*GCGUUCU*	*AGAACGC*	51
4	*GAGACGU*	*ACGUCUC*	15
5	*ACCACGA*	*UCGUGGU*	52
6	*GAUGGUU*	*AACCAUC*	42
7	*GCAGGCA*	*UGCCUGC*	13
8	*GCGAAGU*	*ACUUCGC*	82
9	*GGUGACU*	*AGUCACC*	79

Table 4.1: Table of 120° kissing loop sequences and their measured affinities. Loop A and B correspond to the two hairpin loops comprising the motif.

	Unit Number				
Loop Number	1	2	3	4	5
1/5	1b	-	-	1a	-
2/6	4b	4a	-	6a	6b
3/7	-	2a	2b	-	-
4/8	-	-	5b	5a	-

Table 4.2: Table of kissing loop identities of arms of all units in pentomino nanostructure. '1a' means loop A of kissing loop 1. Loops of the same colour are complementary, and correspond to the colours of labelled kissing loops in Figure 4.6.

NUPACK sequences were predicted using the mfold web server [171]. Sequences where the correctly-folded structure is the most stable structure predicted (lowest free energy ΔG, of folding), and with a large free energy difference (ΔΔG) between the correctly-folded structure and most stable misfolded structure were selected for further analysis. Changes were made manually to reduce the thermodynamic stability of misfolded structures relative to the designed structure as described by Sparvath [115].

This can be done efficiently by analysing the most stable misfold (lowest ΔG of folding) and identifying the longest duplex present which does not exist in the correctly folded structure. Contributing nucleotides are then changed by hand, such that base pairs are maintained in the designed fold, but mismatches are introduced in the misfold (destabilising the misfolded structure relative to the folded structure). Something as simple as tactically changing a $G{\cdot}C$ pair to a $C{\cdot}G$ pair can have a substantial effect. After making changes, the difference in free energy between the designed structure and most stable misfolded structure (ΔΔG) is calculated. If the difference has increased, the sequence change is retained. This process is iterated until only one fold is predicted, or until no further improvement is seen. During this process, $G{\cdot}U$ 'wobble' base pairs were also introduced at least every 6 base pairs to reduce secondary structure in the DNA template (these were not incorporated in the NUPACK design phase, as it typically resulted in failure of the design algorithm). Mfold secondary structure predictions of the five units are shown in the left part of Figure 4.7.

Sequences for the single-stranded linkers were also designed in NUPACK. The four

linkers were designed simultaneously to prevent interaction between strands. Subsequently structures of each of the five units, with their flanking linkers, were predicted by mfold [171] to ensure linkers do not interfere with unit structures. After selecting suitable linker sequences, the secondary structure of the entire strand was analysed by mfold. The folding prediction (top right of Figure 4.7) showed 3×2-base-pair interactions between linkers in the MFE structure. To determine the stability of these interactions, the secondary structure was also predicted with a sequence constraint which prevents the linker-nucleotides from base-pairing (bottom right of Figure 4.7). This structure was actually found to be more thermodynamically stable (lower ΔG for formation of secondary structure: -1320 kcal/mol compared to -1314 kcal/mol)[3], thus it is not expected that linkers will interact within the nanostructure. Since the full nanostructure is predicted to adopt the intended fold in equilibrium, this architecture should also be suitable to generate nanostructures by annealing.

Once the full nanostructure sequence was designed, and structure verified with NUPACK, the RNA sequence was converted to DNA, the consensus T7 RNA polymerase promoter $CTAATACGACTCACTATA$ is included at the 5′ end of the sequence, and restriction sites introduced for cloning into pUC19. The DNA template was ordered as a double-stranded gBlock.[4] The full gBlock sequence is given in Appendix A.1.1.

4.3.2 Synthesis and Characterisation

Cloning

The pentomino template includes flanking restriction sequences: the EcoRI recognition sequence at the 5′ end, and the HindIII sequence at the 3′ end. Each of these restriction sites appears once in pUC19, so this is an appropriate pair of restriction endonucleases for cloning into this vector. The gBlock duplex and pUC19 plasmid were both digested simultaneously with EcoRI[5] and HindIII, and Antarctic phosphatase was used to dephosphorylate the 5′ terminus of the linearised vector to prevent re-circularisation. The vector and insert were ligated using T4 DNA ligase in a 6:1 molar ratio.

5 μL of ligated DNA was transformed into 50 μL competent DH5α *E. coli* (New England Biolabs) by heat shock at 42°C for 30 seconds, and 4×100 μL sequential 10-fold dilutions were applied to ampicillin agar plates and grown overnight. Colonies were selected, and minipreps were prepared using the QIAprep Spin Miniprep Kit from Qiagen. Recombinant plasmids were analysed by AGE, and those of the expected size were sequenced[6]. The template for the complete pentomino nanostructure was prepared from the miniprep by digestion with HindIII and purified.[7] The full cloning protocol is given in detail in Appendix B.1.2.

[3] I believe this discrepancy to be an artefact of the 2-step process used for mfold predictions [171]
[4] Ordered from Integrated DNA Technologies as a gBlock gene fragment (1000 ng).
[5] All enzymes were purchase from New England Biolabs.
[6] Sequencing performed by Source BioScience.
[7] Plasmid purification using the QIAquick PCR purification kit.

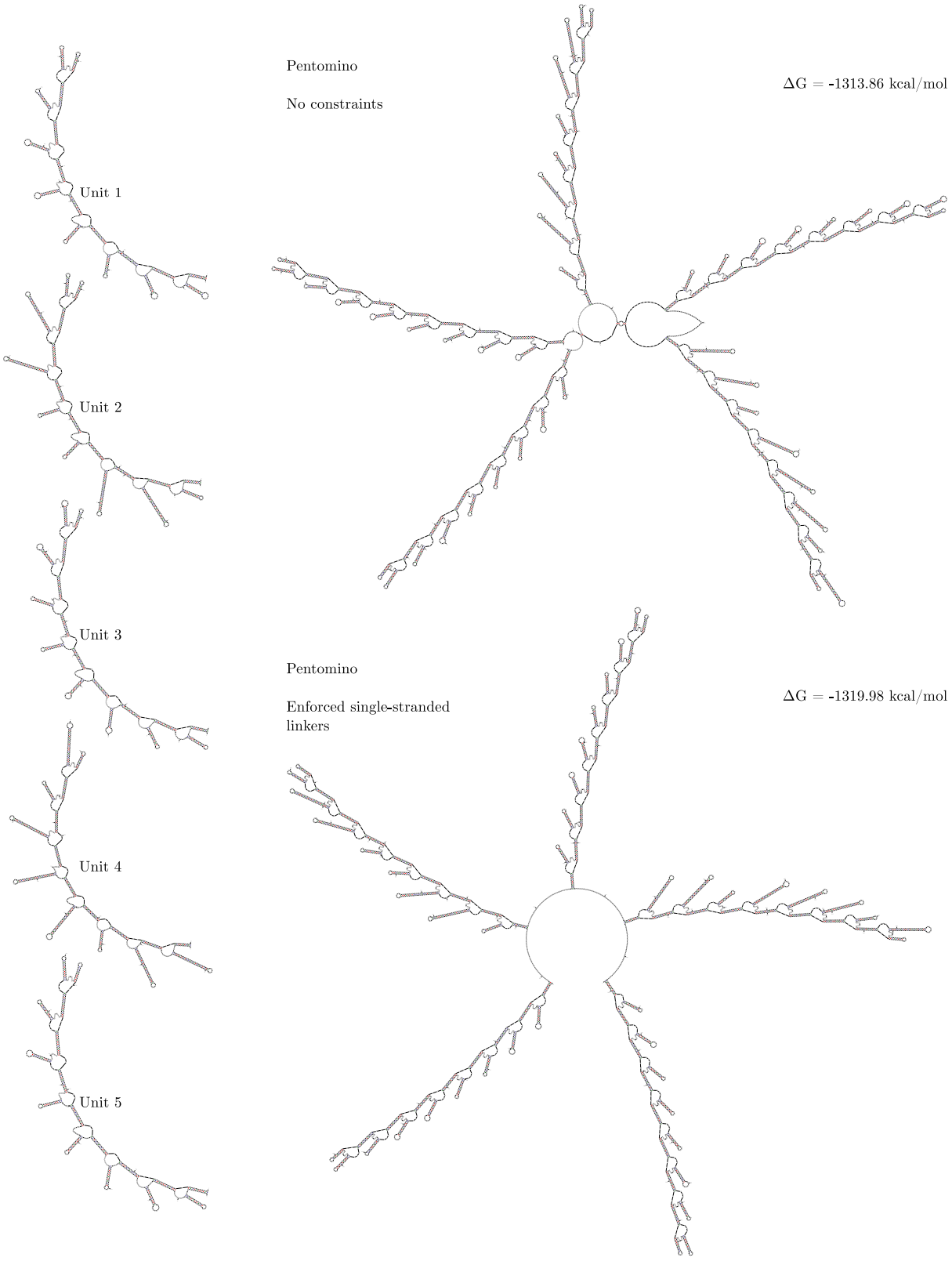


Figure 4.7: Mfold [171] predictions of unit 1-5 secondary structures (left), and the full pentomino nanostructure without (top right) and with (bottom right) enforcement of single-stranded linkers.

Transcription and Gel Characterisation

Pentomino nanostructures were transcribed from linearised plasmid template with the NEB HiScribe kit using 100 ng of template in a 20 µL reaction, incubated at 37°C for 2 hours. This was followed by template digestion using DNase I (protocols provided in Appendix B.2.2). The transcription product was analysed by AGE (Appendix E.1), next to a 1 kb ladder for size reference (Figure 4.8a). Relatively clean bands can be seen, rather than a smear (which would be diagnostic of unfolded RNA). The primary band runs to the roughly same position as the 1000 base-pair DNA duplex of the 1 kb ladder (NEB).

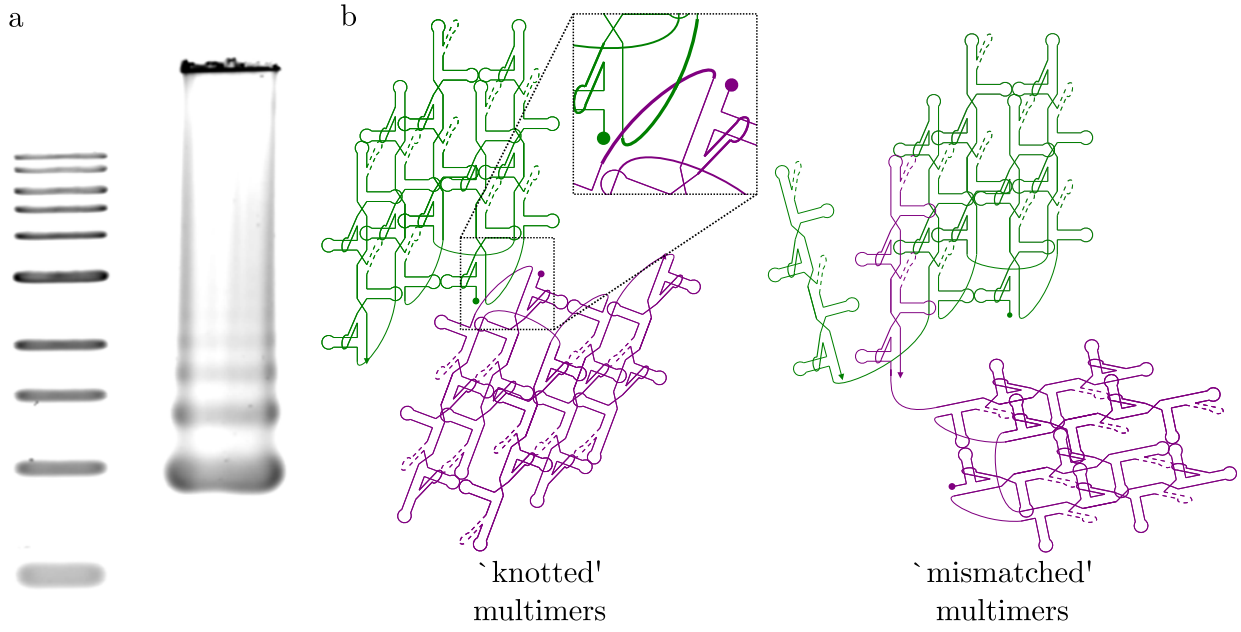


Figure 4.8: (a) AGE of pentomino transcript (lane 2) next to a 1 kb DNA ladder (lane 1). A relatively clean, band (which runs to the same point as a 1000 base-pair duplex) is seen, along with a ladder of higher molecular weight bands. (b) Depictions of possibilities for observed multimers: 'knotted' multimers (left) correspond to nanostructures threaded through one another via single stranded linkers and 'mismatched multimers (right) correspond to nanostructures with intermolecular kissing loop interactions.

In addition to the primary band, a ladder of higher molecular weight bands is present, presumably corresponding to multimers. These may be topological/'knotted' multimers, where structures are threaded through each other via single-stranded linkers, or 'mismatched' multimers in which intermolecular kissing loops form (instead of the designed intramolecular interaction). Multimer types are depicted in Figure 4.8b. In both cases the secondary structure is well-formed. Multimers could also result from mis-formed secondary structures, but this would more likely result in aggregation (which can also be observed). The fraction of multimers can be dramatically reduced by decreasing the template concentration/transcription time.

Because the nanostructure is designed for modular assembly, it is expected that subsets of the full structure should also have well-defined folds. To investigate this, I prepared templates for subsets of the full pentomino structure: units 1, 1-2, 1-3 and 1-4 by PCR (Appendix

B.1.3). Templates were purified using the QIAquick PCR purification kit by Qiagen, and template size and quality verified by AGE. Transcription reactions were prepared (as above, for the full structure) and analysed by AGE, as shown in Figure 4.9.

Similar results to the complete structure (clean band + ladder of multimer bands) are seen for units 1-2, 1-3 and 1-4. For these subsets, a small amount of aggregation is also observed. The fraction of aggregation increases with increasing nanostructure size. For unit 1 only there are no higher molecular weight bands. This strongly supports the theory that multimers are 'knotted' or 'mismatched' rather than misfolded, since there is no way for these super-structures to form from a single unit, whereas multimers from a misfolded secondary structure would still be present.

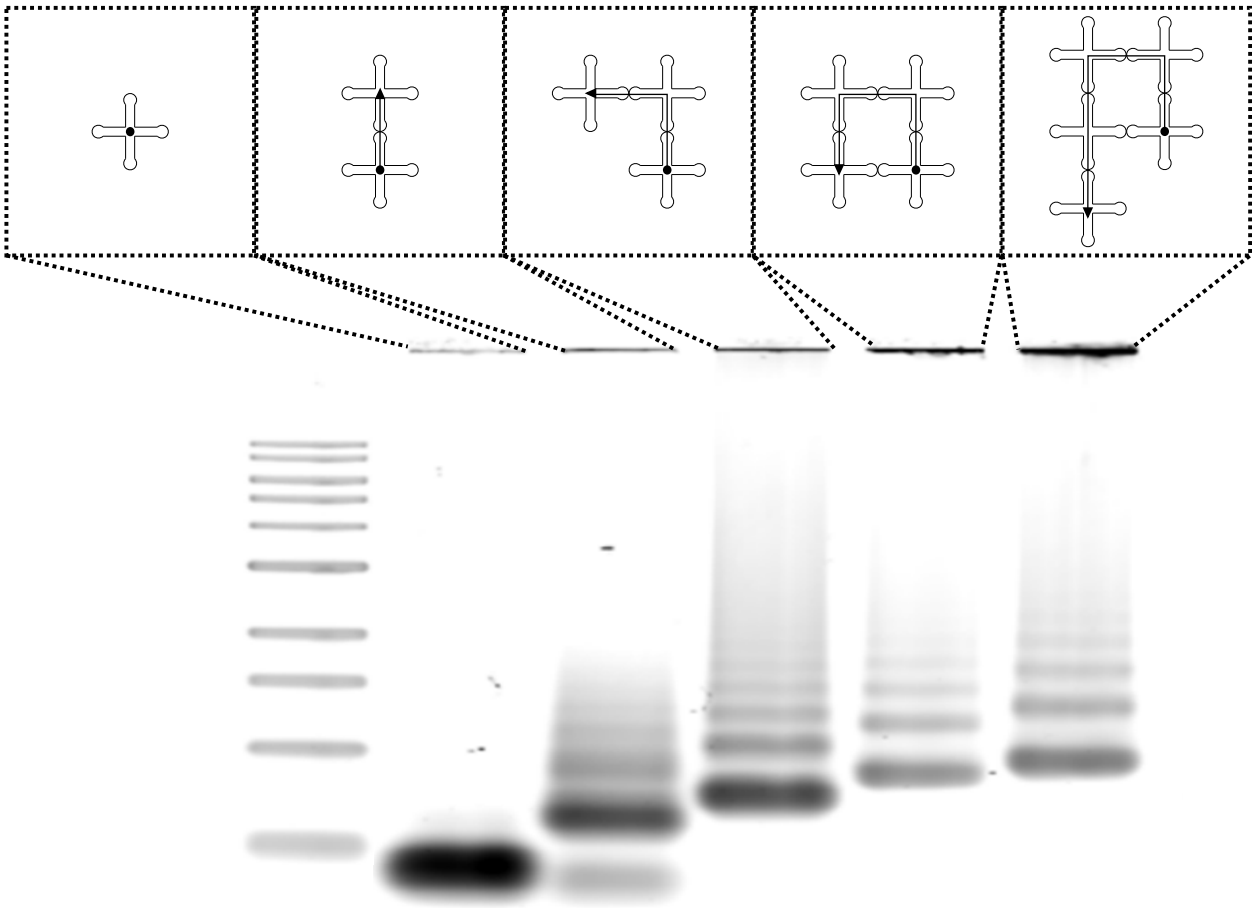


Figure 4.9: AGE of pentomino subsets. Lane 1 is a 1 kb DNA ladder (NEB). Lanes 2-6 are units 1, 1-2, 1-3, 1-4 and 1-5 of the pentomino nanostructure.

Negative Stain TEM

Samples of the pentomino nanostructure were prepared by transcription (Appendix B.2.2) and prepared for TEM by column purification (Appendix B.3.1). Quality of the purified RNA was verified by AGE, and concentration was estimated by A_{260} absorption, and adjusted to 5 nM. Grids were prepared for negative-stain TEM (Appendix D.1.1), and characteristic micrographs are shown in Figure 4.10a and b. Results are encouraging but not conclusive.

Fairly monodisperse structures of the correct size can be seen, in addition to larger structures resembling multimers. Particle sizes were analysed crudely by outlining objects in 49k× micrographs using the freehand selection tool in ImageJ, measuring the area and normalising to nm^2. A histogram of these measured sizes are shown in Figure 4.10c. There is a primary distribution of particles likely corresponding to monomers (with mean ± standard deviation of 350 ± 50 nm^2), what appear to be two distinct higher molecular weight distributions (attributed to dimers/trimers) and occasional larger objects (corresponding to higher order multimers). This is consistent with gel results. Unfortunately the 3D folding of the structures is not visible from these micrographs, although monomers do appear fairly consistent.

Subsets of the pentomino structure were likewise analysed by negative-stain TEM. Example micrographs are shown in Figure 4.11. Results are as expected: increasingly large structures are observed for larger subsets of the full nanostructure. Once again, 3D folding is not visible but particles are largely monodisperse in all cases.

Cryo-EM

To evaluate the usefulness of this architecture for downstream tagging applications nanostructures were prepared for, and analysed by, cryo-EM. Nanostructures were purified by column purification (Appendix B.3.1). The concentration was measured as 0.8 μM by A_{260} absorption. Samples were imaged by cryo-EM in buffer (examples from micrographs shown in Figure 4.12). As with the negative stain TEM, likely nanostructures can be identified, but the 3D folding in not clear. In some cases (e.g. bottom right of top left panel in Figure 4.12), a 'P'-shape, characteristic of the designed nanostructure top view, can be seen, but in most cases structures appear as rings. Objects also appear somewhat smaller by cryo-EM compared to negative stain TEM, most likely because only high-contrast parts of the structure can be visualised by cryo-EM.

4.3.3 Discussion

Preliminary results for this grid-like design architecture are encouraging. Gel analysis shows clean bands and well-defined multimers which is indicative of folding. Compact structures of the expected size can be seen by negative-stain TEM and cryo-EM. For my intended tagging application, however, potential flaws were identified in the sequence design, AGE analysis and cryo-EM characterisation stages.

The DNA template for the RNA nanostructure was difficult to synthesise for the supplier, IDT. This resulted in a synthesis delay and meant many colonies needed to be screened during template cloning to identify one with the correct sequence. This should be limited to save time in future designs. Synthesis difficulties were due primarily to the highly repetitive sequence resulting from the 3WJ which occurs 40 (8×5) times within the nanostructure. For subsequent iterations of this design architecture, a non sequence-specified junction should be used instead. The longer hairpin domains present in the structure, the 18 base pair internal H3 domains of type B units, also complicate synthesis. These 18 base pair domains can be replaced by 7 base pair domains without otherwise affecting the sequence design. This will

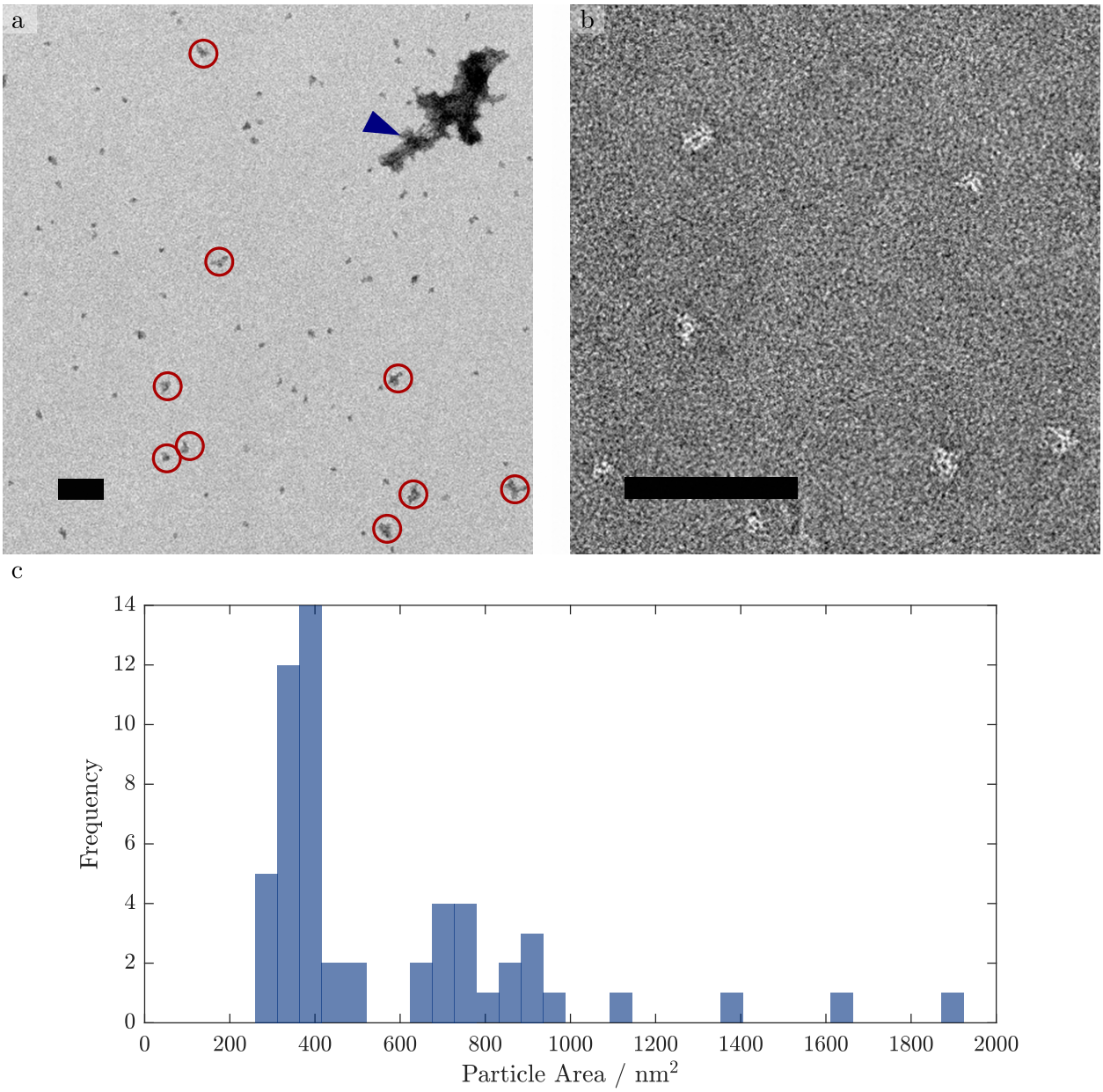


Figure 4.10: Characteristic (a) wide field and (b) high magnification views of pentomino nanostructure. Wide field view is from a positively stained region of the grid, high magnification view is from a negatively stained region. A large RNA aggregate is indicated with a blue arrowhead. Red circles indicate suspected multimeric structures. Scale bars 200 nm. (c) Histogram of particle areas measured from 49k× micrographs.

also give a denser structure which should improve cryo-EM contrast. Furthermore, even at the low concentrations used for negative-stain TEM, a significant population of multimeric structures are observed, attributed to multiple RNA molecules bound through mismatched kissing loops or connected topologically. Reducing the arm length (and correspondingly reducing the linker length) will increase effective concentrations of the intermolecular kissing-loop partner for internal hairpin arms, reducing the concentration of mismatched multimers, and will reduce concentration of topological multimers by reducing the area RNA strands can thread through.

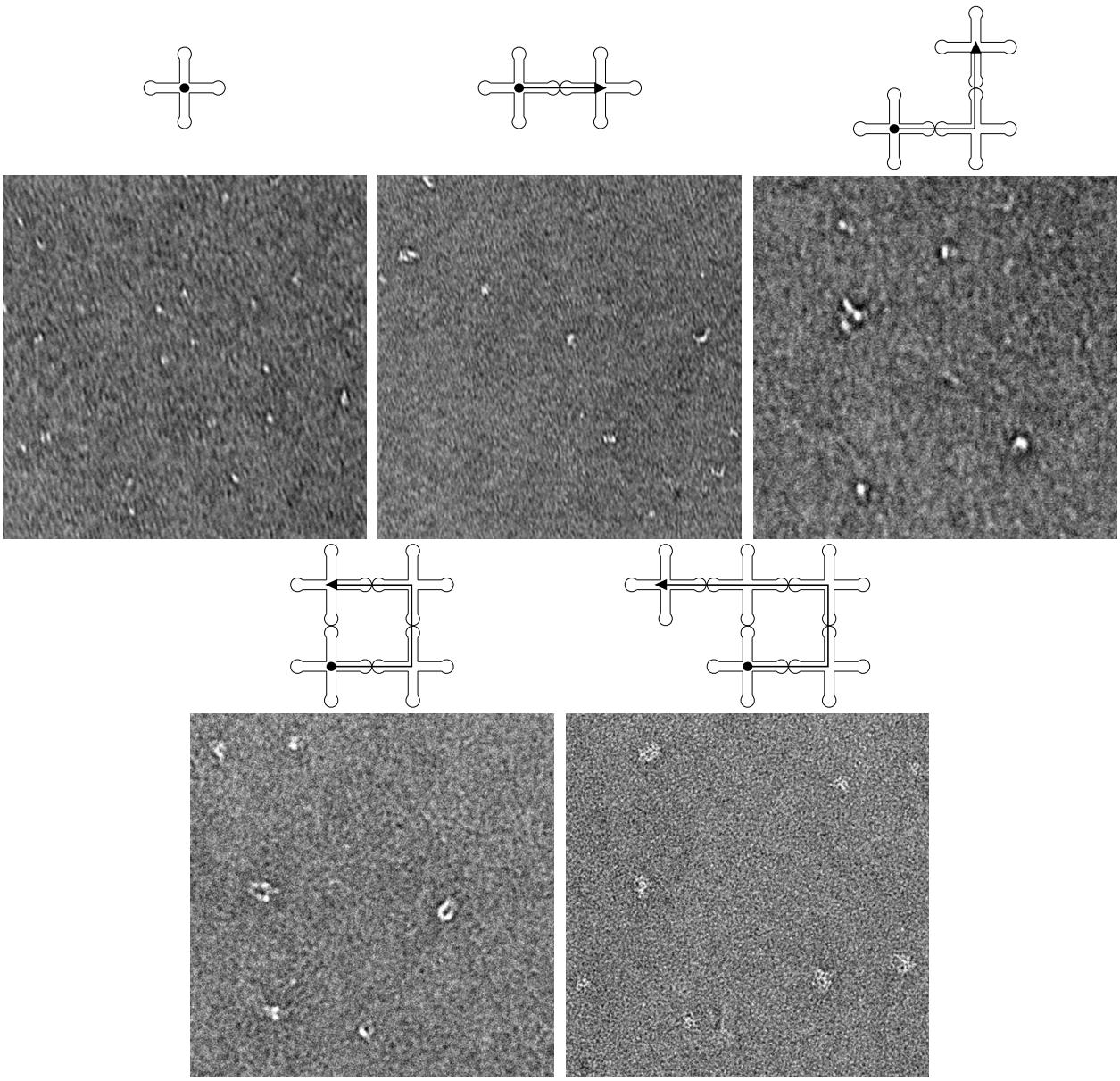


Figure 4.11: Negative stain overview of subsets of pentomino structure (units 1, 1-2, 1-3, 1-4 and 1-5 from top left to bottom right). All panels are 600 nm × 600 nm.

Cryo-EM characterisation of pentomino structures allowed visualisation of compact nano-objects, but the 3D structure was not generally possible to identify. For tagging applications, larger designs with a higher degree of asymmetry will be needed for identification (especially in a noisy biological context). This could be approached by simply extending the design to more units, although nanostructure folding will likely become increasingly difficult with increasing size (as evidenced by the increase in aggregation observed for increasingly large subsets of this five unit structure in Figure 4.9). An alternative is supermolecular assembly of multiple nanostructures. For intermolecular interactions of a suitably affinity, this should be possible even in a biological context.

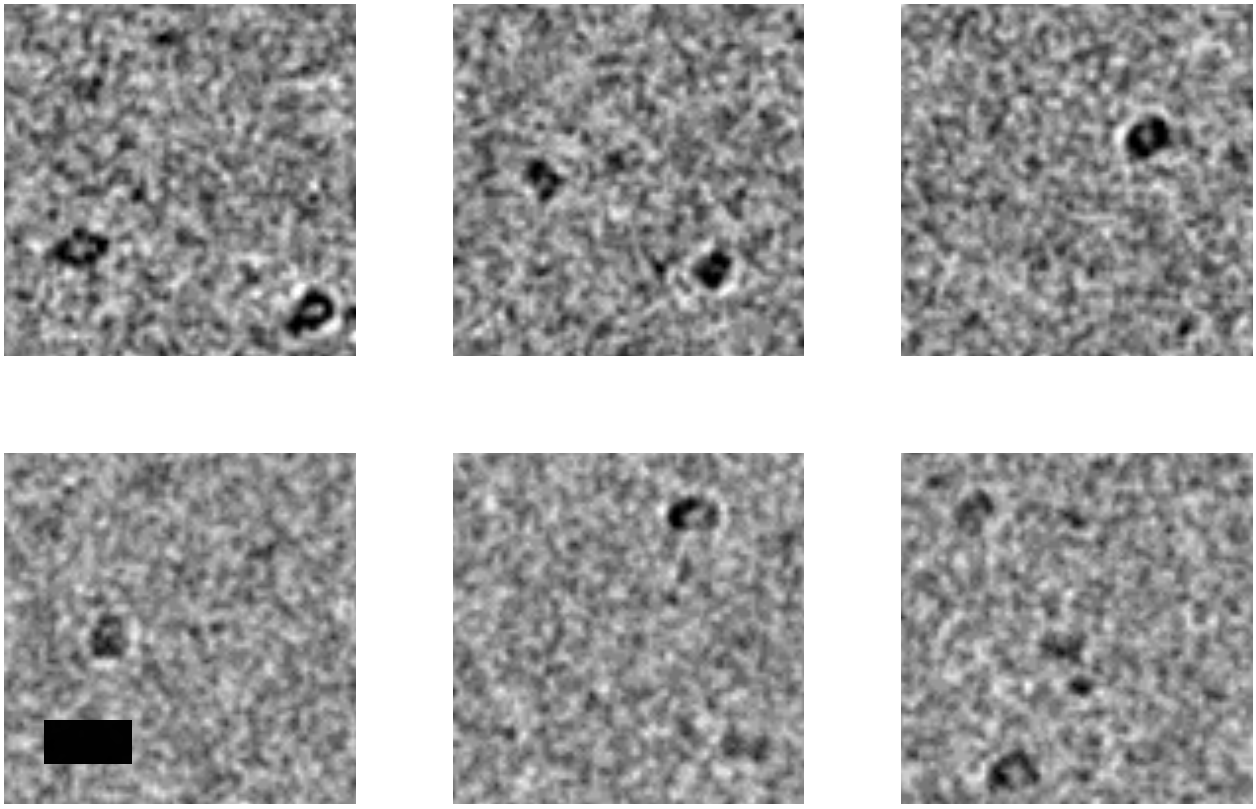


Figure 4.12: Examples of what appear to be pentomino nanostructures image by cryo-EM. Scale bar 20 nm.

4.4 Modified Design for Assembly

4.4.1 Design

Concept

Based on the conclusions discussed in Section 4.3.3, I introduced some design modifications to the grid-like architecture. These include modifications to individual units (to improve template synthesis time/fidelity) and design of nanostructures for isothermal supermolecular assembly. I have designed two simple assembly systems to test this modified architecture.

Modified Units

As mentioned, one of the modifications that should be made to unit design is the replacement of sequence-specified 3WJs with non-specified ones. In 2006, Aurélie Lescoute and Eric Westhof compiled and analysed 36 resolved 3WJs from X-ray structures of folded RNAs [172]. 3WJs were characterised into three families depending on the arrangement of the three contributing helices. The relative length of segments linking the three duplexes is somewhat conserved within each family. Results from this paper are shown in Figure 4.13, but I have changed the nomenclature to match that used in my earlier discussion of junctions and unit design. The family of junctions relevant to this design is Family A, in which helix H3

is roughly perpendicular to the H1/H2 coaxial stack. The contacts made by the junctions regions and the termination of contributing helical domains are varied, but the junction strand lengths (J31, J23, J12[8]) are fairly consistent across the family, so it is proposed here that choosing junction regions of these lengths should result in a junction of this geometry type. I have used J12 and J23 lengths of 2 nucleotides, and a J31 length of 5 nucleotides. Family A represents the most flexible junctions, but the angle of H3 relative to the H1/H2 stack should, in my case, be fixed by the formation of a kissing loop motif. These junction strands are included as single-stranded domains in the sequence design phase.

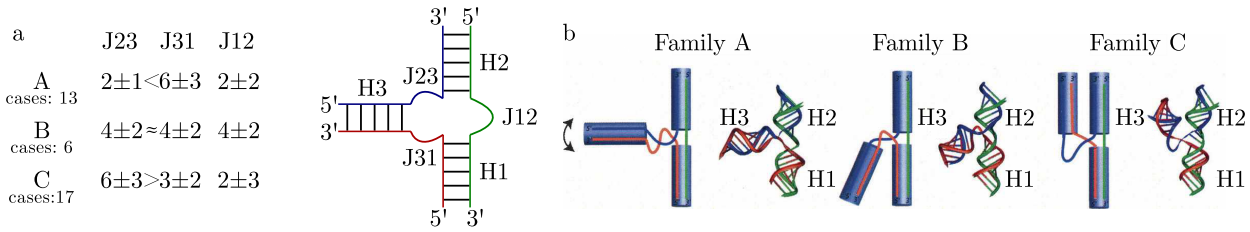


Figure 4.13: Image and caption adapted from Lescoute, 2006 [172] (a) The nomenclature used for the three-way junctions with the average numbers of nucleotides in each junction in the table at the left. The number of instances in each of the three families is also indicated. (b) Schematic drawings of the three observed families, A, B and C, in the analysed three-way junctions. The drawings at the right are based on real structures. In family A, the H3 helix can adopt various angles with respect to the coaxially stacked helices.

Since no crystal structure exists for this more general junction architecture, it is not possible to generate an atomic model for the modified unit design. I have assumed that the H1/H2 helices are stacked precisely, and thus the H1/H2 length between adjacent junctions should be an odd number of helix quarter turns, such that successive units point in 90° from one another in the plane perpendicular to the H1/H2 axis. To this end, I have used an average H1/H2 length of 8.25 base pairs, i.e. every fourth domain is 9 base pairs, and the remaining domains are 8 base pairs. The average twist between H3 arms should therefore be 3/4 of a helical turn (i.e 90° in a left-handed sense). I have also replaced the 18 base pair H3 domains of type B units by 7 base pair domains, as mentioned earlier, to reduce template synthesis problems, to increase density of the resulting structures and to minimise the fraction of multimeric structures.

To verify that the modified units fold as expected, individual units were first designed (as described in the previous section) for hetero-dimerisation, as shown in Figure 4.14a. A type A unit (unit α) and a type B unit (unit β) were designed to interact via 2× kissing loop A (Table 4.1). DNA templates for the designed units were ordered as gBlocks from IDT, and templates prepared by PCR (Appendix B.1.3). The gBlock template sequences for these units are provided in Appendix A.1.1. Units were prepared by transcription and DNAseI template digestion (Appendix B.2.2), then mixed and analysed by AGE. Results are shown in Figure 4.14. The two units interact completely to form a dimer, as designed. There is no evidence of higher molecular weight multimers, which would be expected if only one kissing loop was allowed to form by the unit geometry, so this is a good indication that the units have the expected structure.

[8]Jmn refers to linker strand between helices m and n.

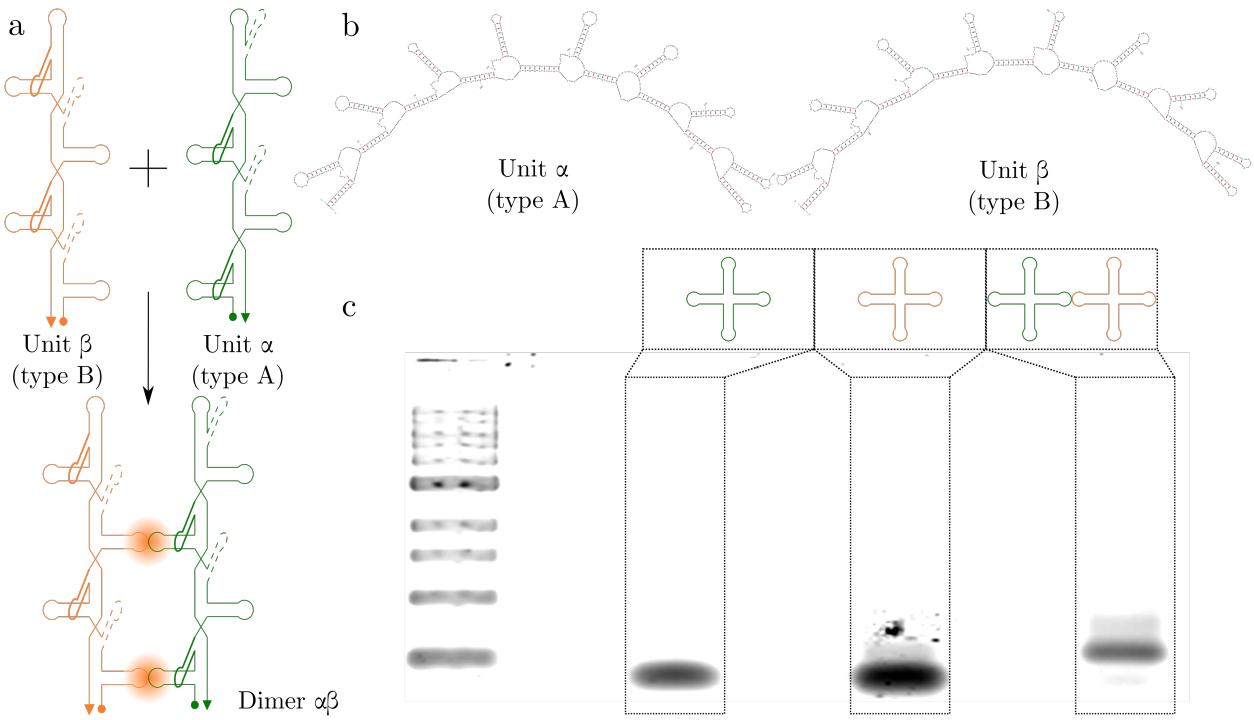


Figure 4.14: (a) Overview of designed dimerisation between units. (b) Mfold predictions of secondary structures for units α and β. (c) Agarose gel electrophoresis of unit dimerisation. Lane 1: 1 kb hyperladder (bioline), fastest two bands correspond to 200 and 400 base pair duplexes. Lane 2: unit α Lane 3: unit β Lane 4: 1:1 mix of units α and β.

Intermolecular Assembly

In addition to these unit modifications, I have opted to increase the total RNA structure size through intermolecular assembly of nanostructures. As in the case of intramolecular assembly this can be achieved through kissing loop interactions, with two key differences.

1. Intermolecular unit-unit interactions can be defined by two kissing loops, but it is important that the two loops are unique and orthogonal, to ensure the nanostructures interact in the correct orientation (since, unlike intramolecular interactions, this will not be sterically enforced).

2. Interacting nanostructures are inverted relative to one another, to help counteract any geometrical strain/distortion which may accumulate with increasingly large structures. The kissing loop angle should therefore be 180°.

The 180° HIV-1 kissing complex (PDB ID: 2FCX) [173] has been used extensively by Luc Jäger and Ebbe Andersen in nanostructure design, and many orthogonal sequences have been developed [107, 114]. Depictions of the wild-type sequence and structure are shown in Figure 4.15. Previously used sequences and their measured affinities are shown in Table 4.3 (1' through 5'), not including the wild-type sequence, which is self-complementary. There is, notably, some overlap between these sequences and the 120° kissing loop sequences shown in Figure 4.1: 180° kissing loops 3', 4' and 5' closely resemble 120° kissing loops 1, 2/7

and 8 and, as such, should not be incorporated into the same structure due to increased opportunities for mis-folding of the tertiary structure.

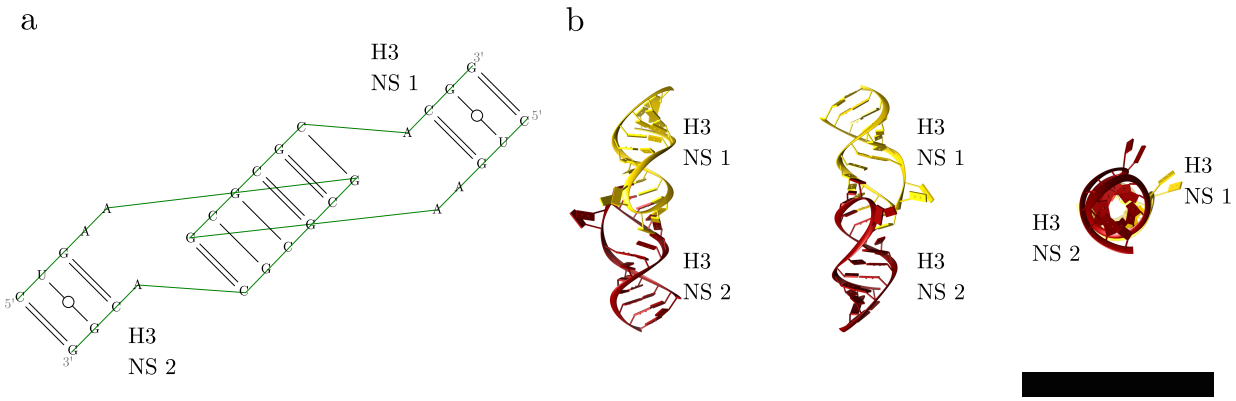


Figure 4.15: (a) Sequence and structure of wild-type 180° kissing loop. Backbone, and base pairs depicted in green and black respectively. (b) 3 orthogonal views of kissing loop with extended arms, rendered in UCSF Chimera [124]. The two contributing helices are shown in red and yellow (scale bar 5 nm).

Kissing Loop	Loop A sequence	Loop B sequence	K_D/nM
1’	*aaGGAGGCa*	*aaGCCUCCa*	5.6
2’	*aaGCGAGCa*	*aaGCUCGCa*	12.6
3’	*aaGUCCACa*	*aaGUGGACa*	15.8
4’	*aaGCAGGCa*	*aaGCCUGCa*	11.1
5’	*aaCUUCGCa*	*aaGCGAAGa*	152
6’	*aaGUGCUGa*	*aaCAGCACa*	?

Table 4.3: Table of 180° kissing loop sequences and their measured affinities. Loop A and B correspond to the two hairpin loops comprising the motif. Lower case ‘a’ nucleotides exist in the loop but are not involved in pairing.

For *in vitro* one-pot assembly or *in vivo* assembly, it is useful to introduce a hierarchy of interaction strengths to promote correct folding. This is a common principle in DNA nanotechnology [48, 49, 55, 54, 56, 68]. The base principle is that strong, intramolecular interactions will form first, and weaker intermolecular interactions will form later between intact nanostructures, not interfering with nanostructure folding as nanostructures are kinetically trapped in their co-transcriptional fold. Loop 5’ possesses an affinity roughly 10-fold lower than the 120° kissing loops, as well as the other 180° kissing loops, so the $aaGN_4Ga/aaCN_4Ca$ sequence structure is used to design a new, orthogonal, intermolecular loop sequence. Kissing loop sequence 6’ (Table 4.3) was introduced. Although the affinity of this kissing interaction was not measured, it is expected to be a similar to loop 5’. Checks were made for the suitability of this loop.

1. The structure of the two contributing hairpins was verified in mfold, to ensure there is no additional structure within the loop which could inhibit binding.

2. The analysis module of the NUPACK web server was used to check for orthogonality with other 180° and the 120° kissing loops (treating the loop sequence as a single strand).

3. The knotted structure of the kissing loop was verified through the 'predict' feature of the Nanofolder web server [174].

Each intermolecular unit/unit interaction interface comprises two kissing loops. I propose using one strong ($aaGN_4Ca/aaGN_4Ca$ type) kissing loop to ensure the nanostructures find each other in solution, and one weak ($aaGN_4Ga/aaCN_4Ca$ type) kissing loop to keep the intermolecular interaction interface weaker than the equivalent intramolecular interface for hierarchical assembly. The length of the H3 domains capped by intermolecular kissing loops is not important because of the 180° geometry. I have used 9 base pair domains, but this can be increased/decreased as necessary.

A schematic of the designed assembly options is shown in Figure 4.16. For initial experiments I designed simple super-structures as a test of supermolecular assembly. Three nanostructures (A, B and C) are used:

Nanostructure A is a 4-unit nanostructure, where the first two units include external loops for intermolecular assembly, arranged into one interaction 'face'. The remaining 3 'faces' are decorated only with non-interactive $UUCG$ tetraloops.

Nanostructure B is a 2-unit nanostructure, where each unit now consists of 12 (rather than 8) stacked 3WJs. Nanostructure B has one interaction face (comprising the four loops complementary to the four loops which make up the interaction face in nanostructure A). Nanostructures A and B are therefore expected to bind in a 1:1 ratio, resulting in super-structure AB.

Nanostructure C is also a 2-unit nanostructure, identical in architecture to nanostructure B, but with two interaction faces (each equivalent to the nanostructure B's interaction face). Nanostructure C should bring together two copies of nanostructure A resulting the super-structure A_2C.

The inter- and intramolecular kissing loops sequences assigned are shown in Table 4.4.1, and depicted spatially in Figure 4.17. Note that intramolecular kissing loops exist between arms of the same number (i.e. at the same height), but this is not the case for intermolecular interactions as interacting nanostructures are inverted relative to one another.

Sequence Design

Unit sequences were designed as described in Section 4.3. Only loop sequences (120° kissing loops, 180° kissing loops and $UUCG$ tetraloops) and the T7 starter sequence (for unit 1 of all nanostructures) were specified. Sequences of nanostructures B and C differ only in the intermolecular kissing loops (as specified in Figure 4.4.1), so we can be certain that any

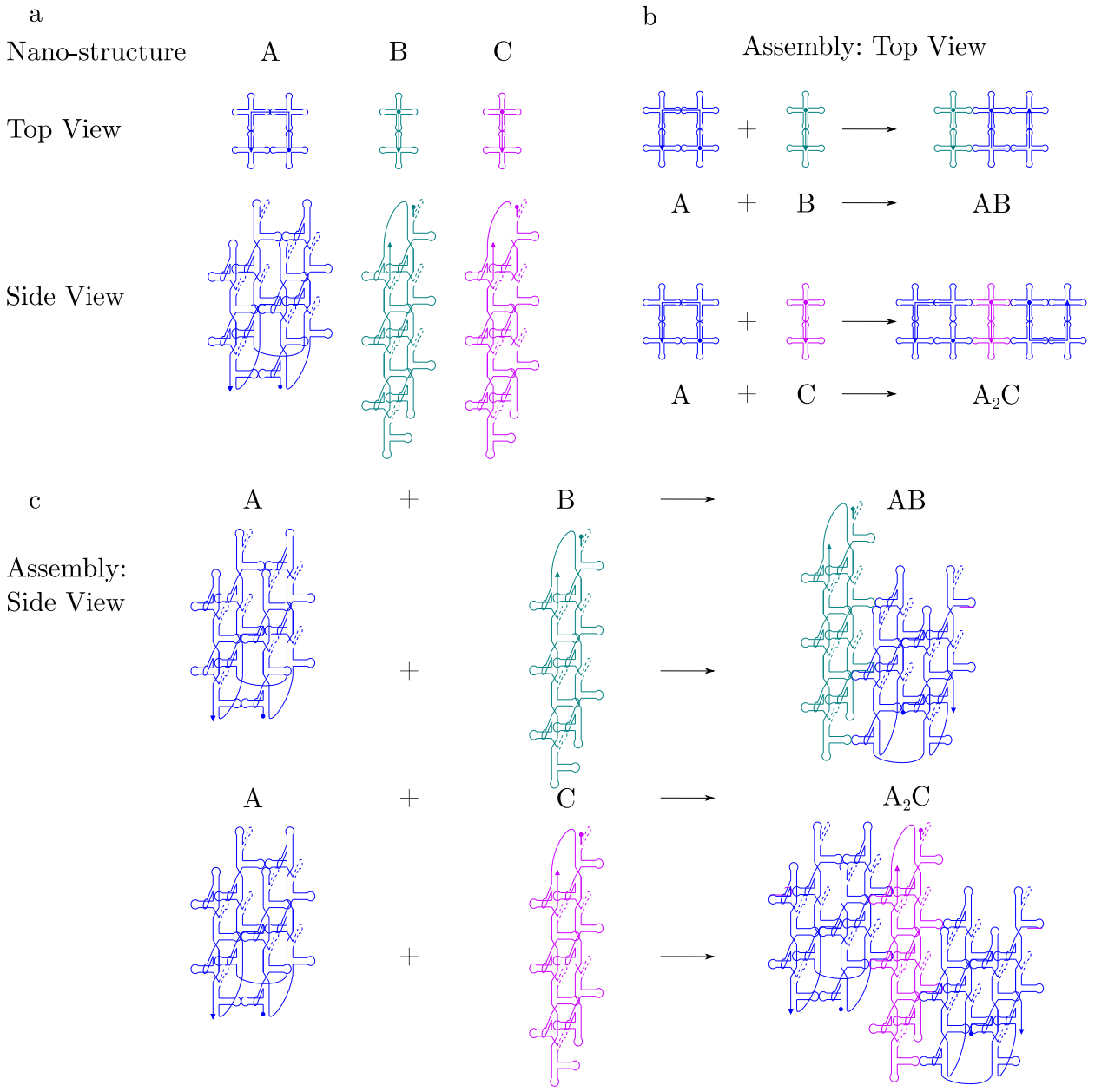


Figure 4.16: Schematic overview of nanostructures for testing supermolecular assembly. (a) Top and side view of individual nanostructures. Top (b) and side (c) views of designed super-structures, assembled by mixing component nanostructures.

differences observed in their interactions with nanostructure A results from the designed kissing interactions.

Because of the shorter H3 domains in type B units compared to the pentomino nanostructure, I have used 30 nucleotide single-stranded linkers. I designed the 4 linkers simultaneously, assigning 4 linkers to nanostructure A, and the remaining linker to nanostructure B/C since it is important that the linkers on different nanostructures also do not interact with one another.

The sequences for the units and linkers were concatenated into a single strand for each of the three nanostructures, and the secondary structure for the strands were predicted by

	NS A Unit #				NS B Unit #		NS C Unit #	
Loop #	1	2	3	4	1	2	1	2
1	1a	5'a	-	1b	-	-	-	-
2	-	-	4a	4b	-	-	-	8'b
3	1'a	3b	3a	-	2a	2b	2a	2b
4	2a	2b	-	-	-	-	2'b	-
5	1a	2'a	-	1b	-	-	-	-
6	-	-	4a	4b	8'b	-	8'b	1'b
7	8'a	3b	3a	-	2a	2b	2a	2b
8	2a	2b	-	-	-	2'b	5'b	2'b
9					-	-	-	-
10					1'b	-	1'b	-
11					2a	2b	2a	2b
12					-	5'b	-	5'b

Table 4.4: Table of kissing loop identities in arms of all units in nanostructures A, B and C. '1a' means loop A of kissing loop 1. Loops of the same colour are complementary, and correspond to the colours of labelled kissing loops in Figure 4.17.

mfold. The predicted folding for the full structures, as well as the constituent units, are shown in Figure 4.18. As for the pentomino nanostructure, the predicted MFE equilibrium structure for the full strand is consistent with the designed structure.

Following sequence design, the consensus T7 RNA polymerase promoter $CTAATACGACTCACTATA$ is included at the 5′ end of the sequence, and restriction sites introduced for cloning into pUC19. The DNA templates were ordered as double-stranded gBlocks. Sequences are provided in Appendix A.1.1.

4.4.2 Transcription and Characterisation

Cloning

Templates for nanostructures B and C were cloned into a pUC19 vector as for the pentomino nanostructure (Appendix B.1.2). For nanostructure A, I inadvertently introduced additional EcoRI sites into the template sequence during design. This template was therefore cloned into pUC19 using one blunt end and one sticky end. The gBlock was digested with HindIII followed by 5′ phosphorylation with T4 PNK and pUC19 was digested simultaneously with HindIII and HincII followed by dephosphorylation by antarctic phosphatase.[9] Following ligation, cloning and sequencing, and template plasmids were linearised for all three nanostructures by HindIII digestion.

[9]All enzymes were purchase from New England Biolabs.

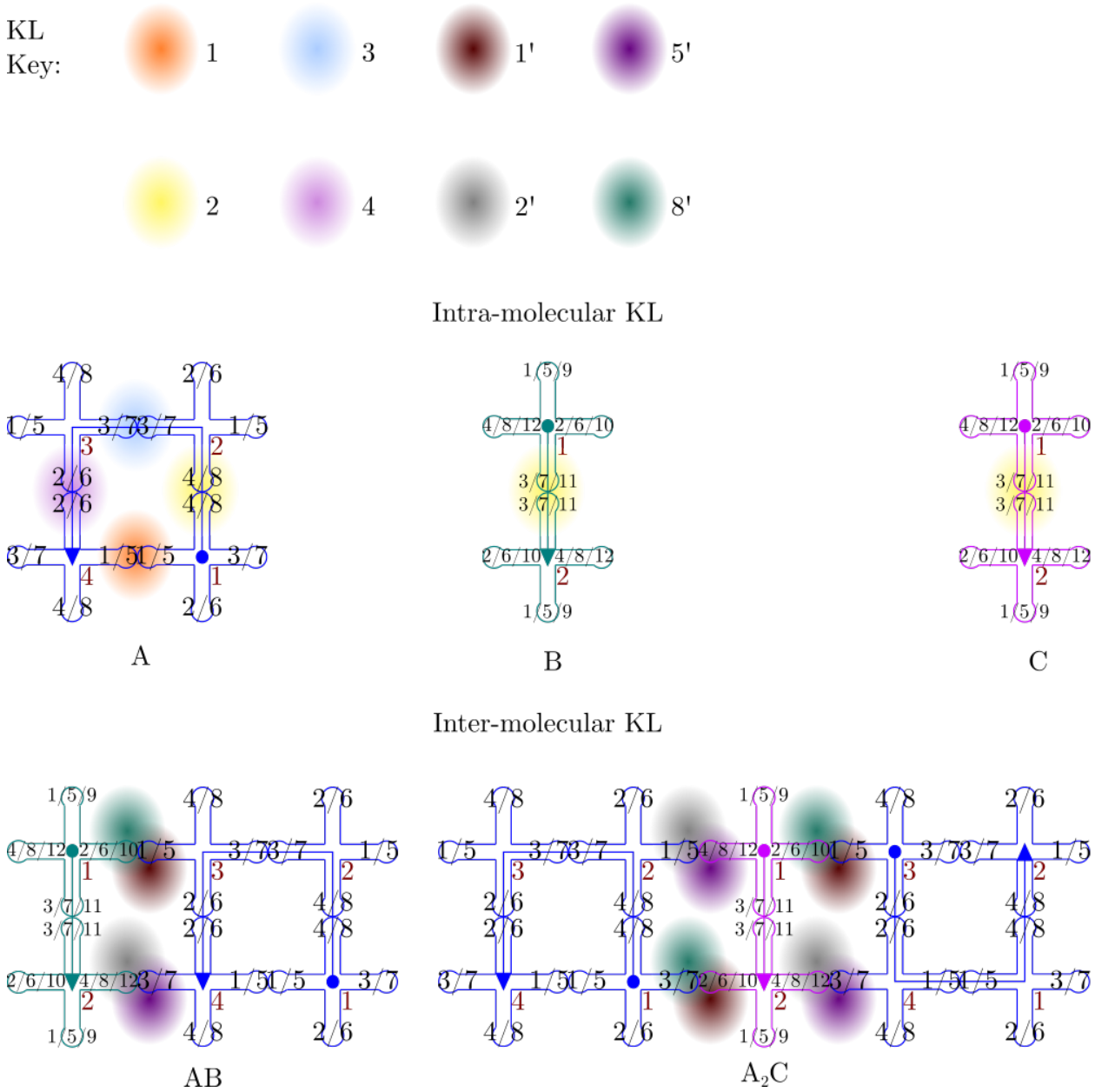


Figure 4.17: Overview of kissing loops in nanostructures A, B and C, including 120° intramolecular kissing loops (above) and the 180° intermolecular kissing loops (below). Arrows point from 5′ to 3′. Units are labelled in red in order of transcription, and for each unit, arm numbers are labelled.

Transcription and Characterisation of Nanostructures

The three nanostructures were each prepared by transcription and analysed by gel electrophoresis and negative stain TEM. AGE results are shown in Figure 4.19. Relatively clean bands are seen for all three nanostructures with little multimerisation/smearing. The decrease in proportion of observed multimers is most likely attributed to the reduced linker/arm lengths. Nanostructure mobilities fall between the 0.5 and 1.0 kb bands of the ladder in lane 1.

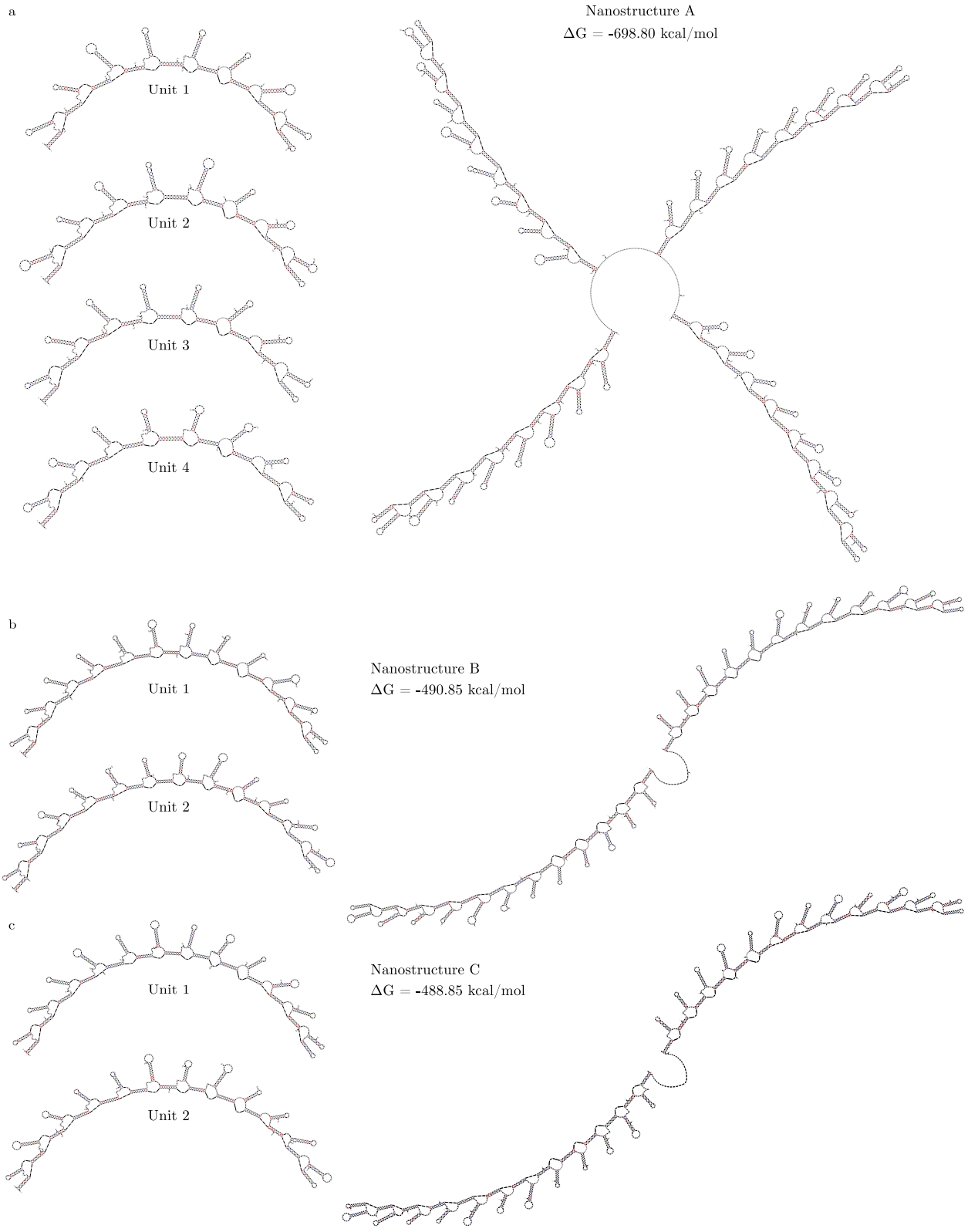


Figure 4.18: Mfold predictions of component units (left) and full nanostructures (right) for nanostructures A, B and C.

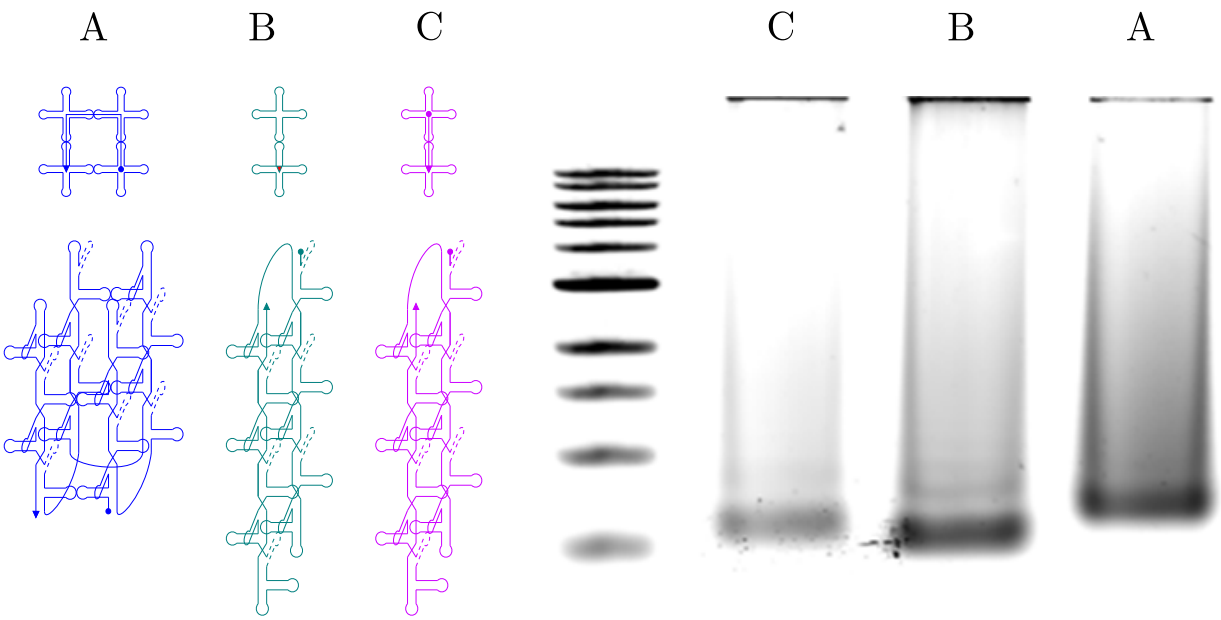


Figure 4.19: AGE of nanostructures A, B and C (with schematic for reference). The NEB Quick-Load 1 kB ladder is shown in lane 1.

For negative-stain TEM, structures were purified by LiCl precipitation (Appendix B.3.2) and resuspended in PBS-Mg buffer[10] to a final concentration of 5 nM. Grids were prepared for negative-stain TEM (Appendix D.1.1) and characteristic wide-field and high-magnification views are shown in Figure 4.20. For all three nanostructures, a majority of mono-disperse structures are observed, likely corresponding to monomers. A small population of larger structures (corresponding to multimers/aggregates) is also seen. Particle size distributions were analysed from 68 k× micrographs by the method described in Section 4.3.2. Histograms of measured areas are shown in Figure 4.21. In all cases, a primary distribution of monomers is seen, with a long tail of larger multimers (unlike for the pentomino nanostructure, these are not present in large enough numbers to identify distinct populations). Nanostructures B and C have extremely similar size distributions, which is expected. Using a monomer cut-off of 600 nm^2 for nanostructure A, and 400 nm^2 for nanostructures B and C, monomer particle sizes (given as mean ± standard deviation are 400 ± 40, 230 ±50 and 230 ± 60 nm^2 for nanostructures A, B and C respectively. Although there is no atomic model for comparison, these sizes are comparable to that of the pentomino structure, and therefore reasonable.

Characterisation of Assemblies

Assembly behaviour of these nanostructures was assessed by AGE. As gel electrophoresis is not an equilibrium technique, a quantitative assessment of binding cannot be made, but a qualitative assessment of the species present can be. Nanostructures were transcribed and relative concentrations estimated by gel band intensity. Nanostructure A was mixed roughly 1:1 with nanostructures B and C,[11] and incubated on ice for 15 minutes. These mixes were ran on a gel alongside the individual nanostructures for reference, results are shown in Figure

[10] 20 mM phosphate buffer, 2.7 mM KCl, 137 mM NaCl, 5 mM $MgCl_2$, pH 7.4.

[11] Relative concentrations estimated by previous measurements of gel band brightness for the three samples.

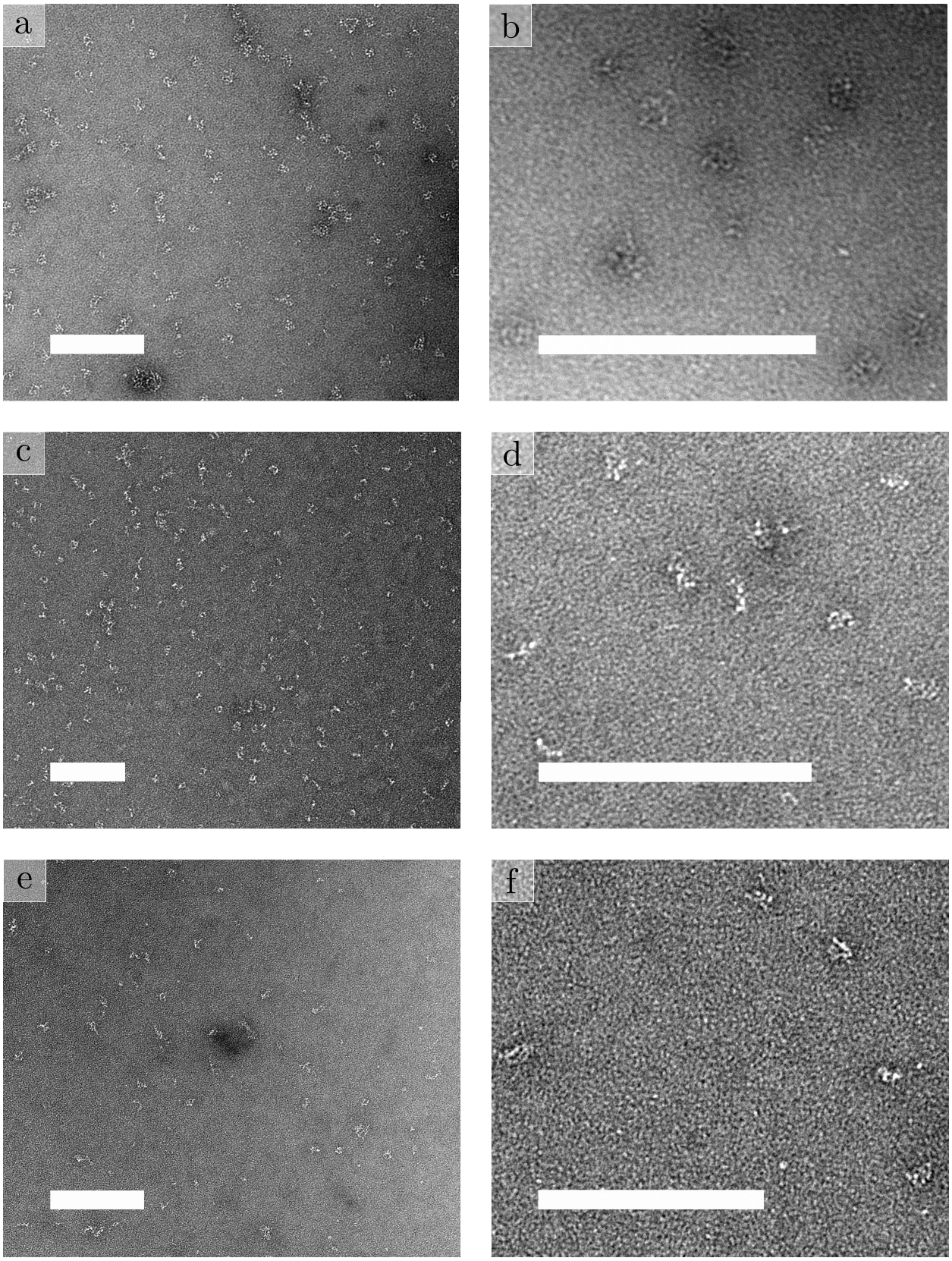


Figure 4.20: TEM overview of nanostructures A, B and C for assembly. (a)/(b) Wide-field/high magnification views of Nanostructure A. (c)/(d) Wide-field/high magnfication views of Nanostructure B. (e)/(f) Wide-field/high magnification views of Nanostructure C. All scale bars are 200 nm.

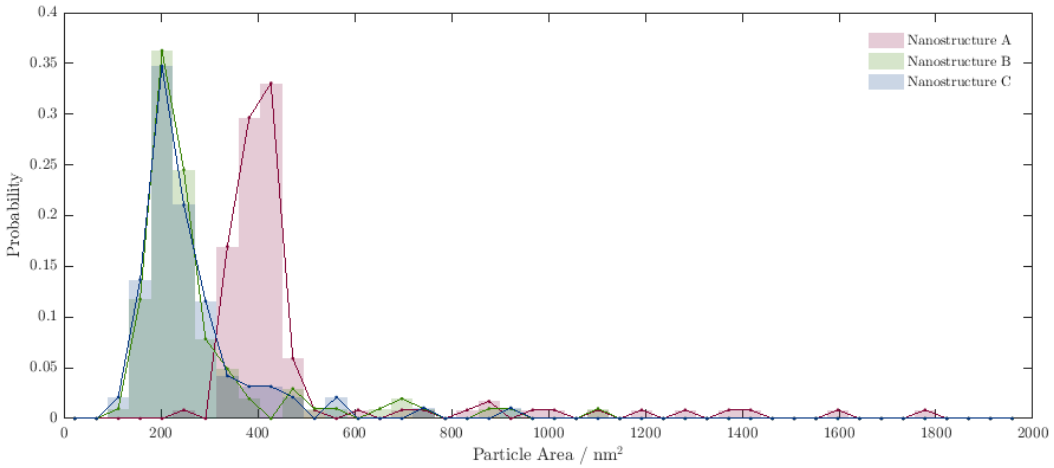


Figure 4.21: Histogram of measured particle areas for nanostructures A, B and C. Histograms are normalised to account for differences in particle number, and a line plot overlays the histogram bars to help distinguish the three distributions.

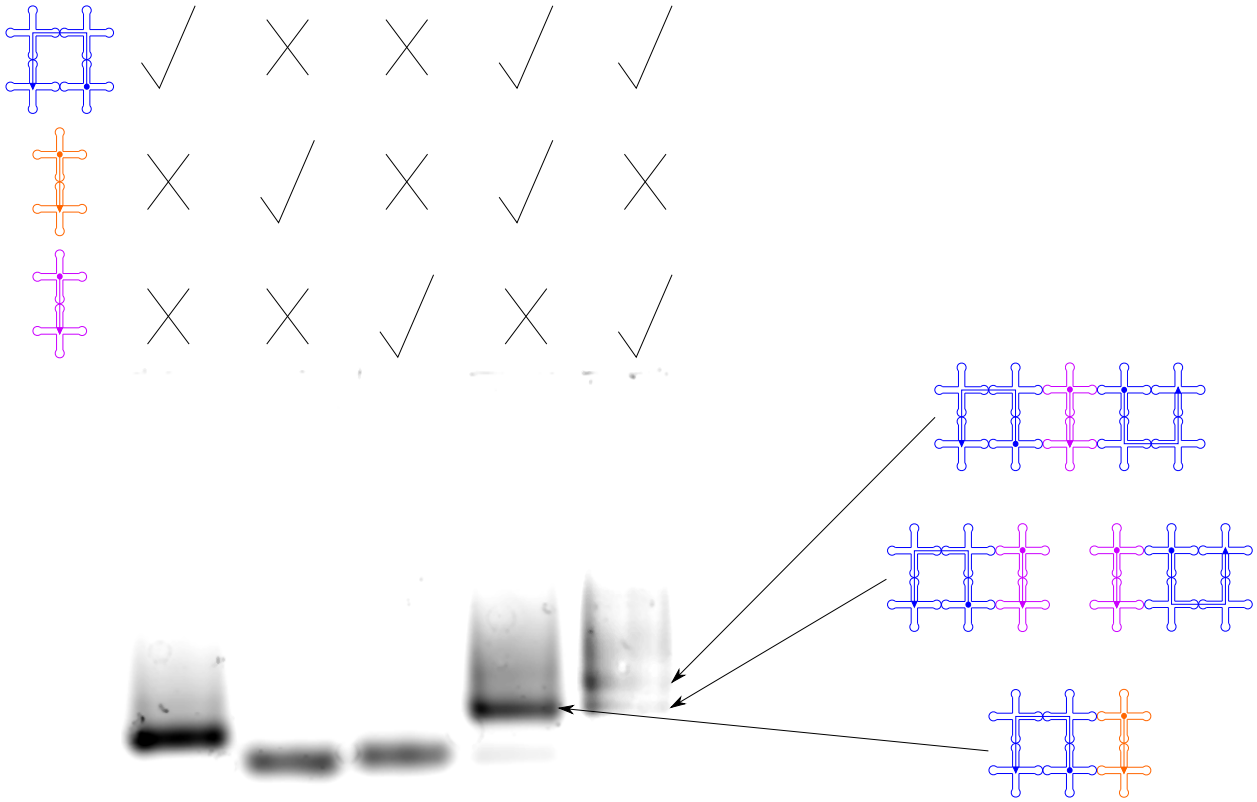


Figure 4.22: AGE of AB and AC/A_2C assemblies. Lanes 1-3 contain nanostructures A, B and C. Lanes 4 and 5 contain roughly 1:1 mixes of nanostructure A with nanostructure B (lane 4) and C (lane 5).

4.22.

Nanostructures A and B interact in a 1:1 ratio to form a hetero-dimer, as designed, with no evidence of higher order interactions. In the A and C mix two bands are seen. The higher mobility band runs to approximately the same place as the AB dimer and is likely an AC

hetero-dimer, with the lower mobility band corresponding to the full A_2C complex (both bands are expected since nanostructure A is not in a 2-fold excess of C).

4.4.3 Discussion

Results for nanostructures A-C, using the modified grid architecture, are promising for design of cryo-EM tags. Individual nanostructures have been characterised by AGE and negative stain TEM, and appear to have a consistent mono-disperse folding, with a low fraction misfolded material. AGE results of mixtures also show that these structures can be designed for programmable supermolecular assembly, which allows for extensions of this architecture to larger sizes. As nanostructures B and C differ only in their external loop sequences, we can be fairly certain that the observed difference in interaction stoichiometry is a consequence of the designed set of intermolecular interactions and not, for example, topological/mismatched multimers like those seen for the pentomino design. For cryo-EM tagging applications, larger assemblies will be advantageous, so in the next section I will describe the design of new nanostructures (D and E) which are intended to interact with nanostructure A for larger cross-shaped assemblies.

4.5 Cross-Shaped Assembly

4.5.1 Design

Based on encouraging preliminary results from simple assembly options, I designed larger cross-shaped super-structures which would be a more appropriate size for tagging applications. These super-structures again consist of only two component nanostructures, which are now designed to interact in a 4:1 ratio to form a cross-shaped pentamer. Nanostructure A is re-purposed as the satellite structure (which appears four times in each super-structure), and two new nanostructures are also designed for the central position in the assembly:

Nanostructure D is a structure composed of a single 8-armed unit, where each 'side' of the unit contains two hairpins which interact with 1/2 of the interaction face of nanostructure A. The H3 domains are extended by 11 base pairs to prevent a steric clash between bound copies of nanostructure A.

Nanostructure E is a structure composed of 4×12-armed units, where each 'side' of the structure contains four hairpins which interact with the full interaction face of nanostructure A. This structure more closely resembles nanostructures B and C from the previous section.

The 4:1 ratio means that the central structure can be functionalised for targeting (e.g. by incorporation of a GFP aptamer), and the targeting region can be uniquely determined within the super-structure since there is only one copy.

A schematic for these nanostructures and their intended assemblies with nanostructure A is depicted in Figure 4.23. It is expected that nanostructure D will have a simpler and more robust folding pathway than nanostructure E, but the higher complexity of nanostructure E may provide advantages. The internal kissing loops should give a more rigid structure, and the four kissing loop interaction interface will be stronger than the two loop interface (although the two loop interface should be sufficient in most contexts).

The inter- and intramolecular kissing loops sequences were assigned are shown in Table 4.5, and depicted spatially in Figure 4.24.

	NS D	NS E			
	Unit #	Unit #			
Loop #	1	1	2	3	4
1	8'b	-	-	4a	4b
2	8'b	-	3b	3a	-
3	8'b	2a	2b	-	8'b
4	8'b	1a	-	8'b	1b
5	1'b	2'b	8'b	4a	4b
6	1'b	8'b	3b	3a	2'b
7	1'b	2a	2b	2'b	1'b
8	1'b	1a	2'b	1'b	1b
9		5'b	1'b	4a	4b
10		1'b	3b	3a	5'b
11		2a	2b	5'b	-
12		1a	5'b	-	1b

Table 4.5: Table of kissing loop identities of arms of all units for nanostructures D and E. '1a' means loop A of kissing loop 1. Loops of the same colour are complementary, and correspond to the colours of labelled kissing loops in Figure 4.24

.

Nanostructure sequences were designed as described previously. Single-stranded linkers for nanostructure E were designed in the presence of the nanostructure A linkers to prevent interactions between these domains. Mfold predictions of unit and full-strand secondary structures are shown in Figure 4.25.

Sequences were converted to DNA, terminal sequencers (including promoter and restriction enzyme sequences) were added, and templates were ordered as gBlocks from IDT.

4.5.2 Synthesis and Characterisation

Synthesis and Characterisation of Nanostructures

Templates for nanostructures D and E were prepared by PCR (Appendix B.1.3) and cloning followed by restriction enzyme digest (Appendix B.1.2) respectively.

Nanostructures A, D and E were prepared by transcription followed by template digestion

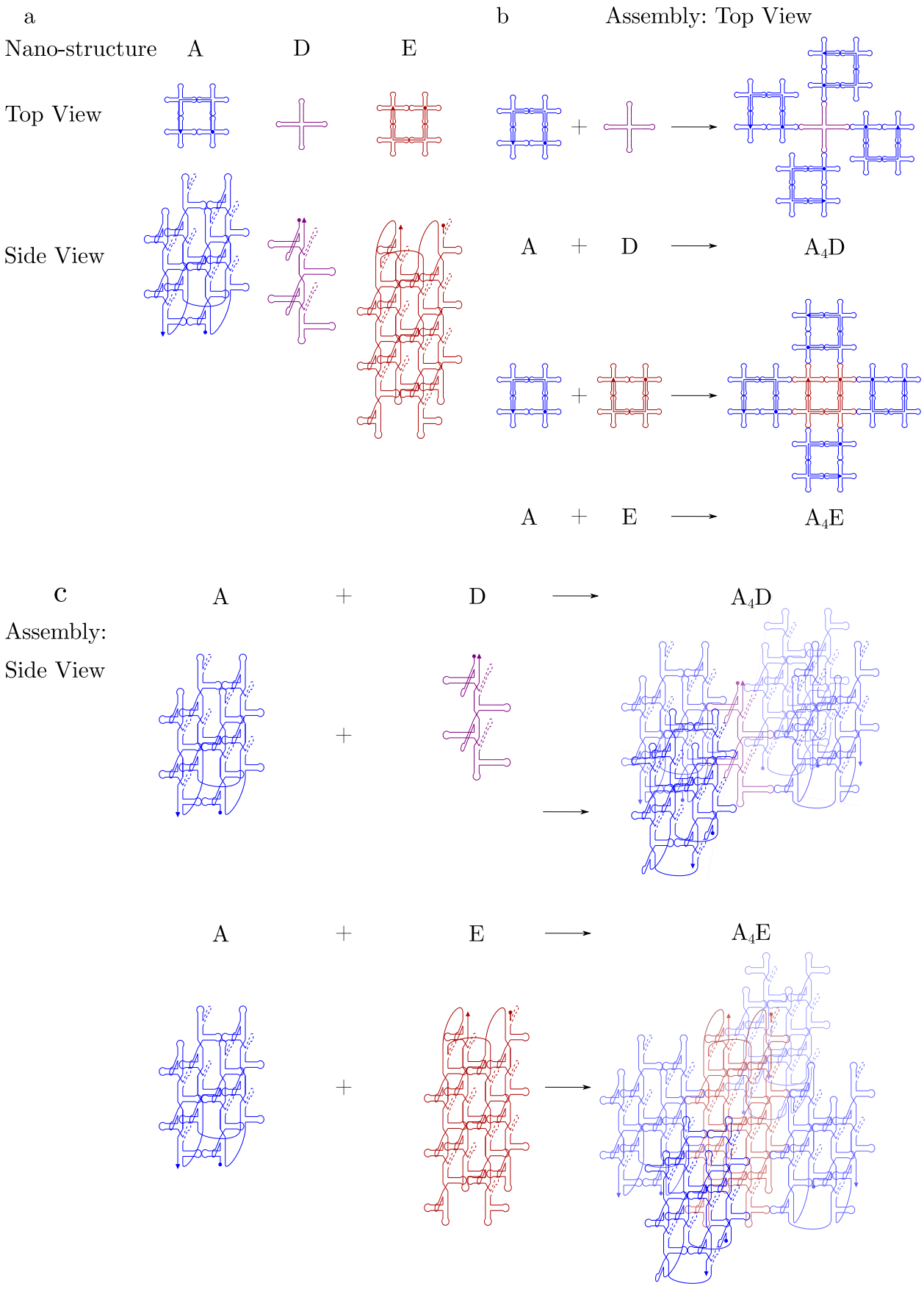


Figure 4.23: Schematic overview of nanostructures for a cross-shaped supermolecular assembly. (a) Top and side view of individual nanostructures. Top (b) and side (c) views of designed super-structures, assembled by mixing component nanostructures.

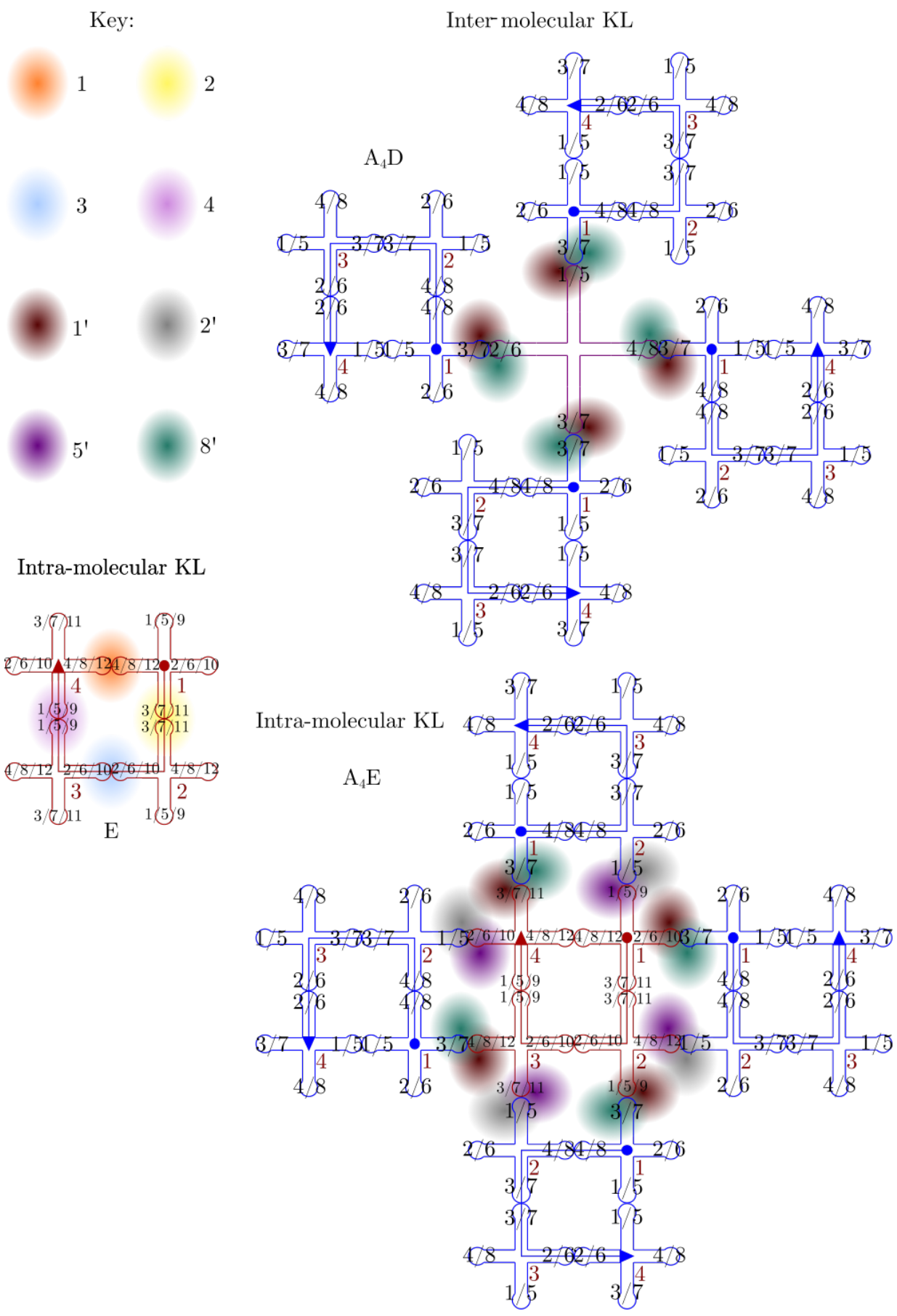


Figure 4.24: Overview of kissing loops for cross-shaped assembly, including 120° intramolecular kissing loops for nanostructure E (left) and the 180° intermolecular kissing loops (right). Arrows point from 5′ to 3′. Units are labelled in red in order of transcription, and for each unit, arm numbers are labelled.

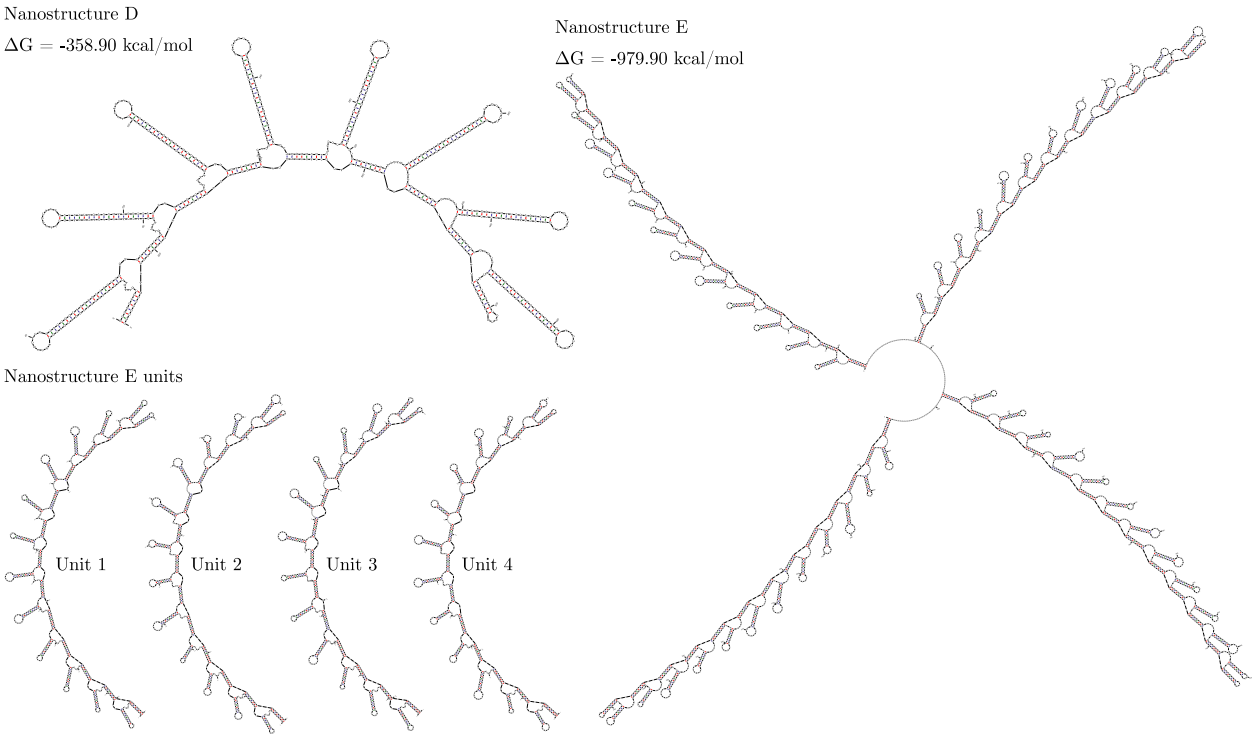


Figure 4.25: Mfold secondary structure predictions for nanostructure D (top left) and nanostructure E (component units bottom left and full nanostructure on the right).

with DNAseI I and analysed by gel electrophoresis. Results are shown in Figure 4.26 and three structures gave clean bands, indicative of a single consistent folding.

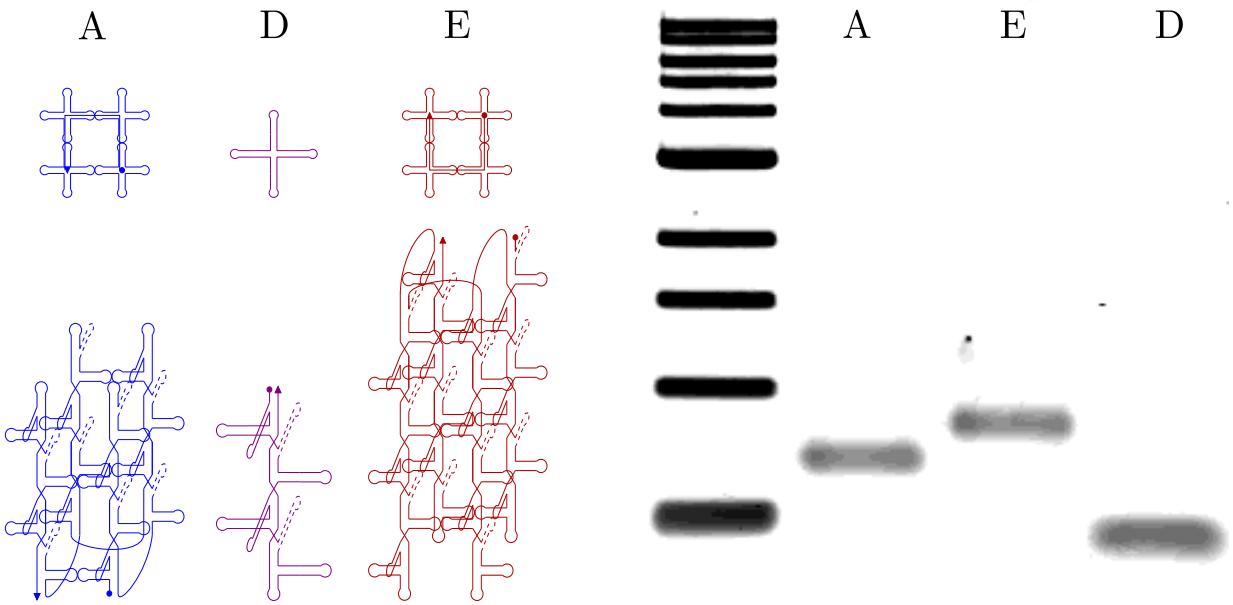


Figure 4.26: AGE of nanostructures A, D and E (with schematic for reference). The NEB Quick-Load 1 kB ladder is shown in lane 1.

Nanostructures D and E were also prepared for negative stain TEM. Samples were purified by LiCl precipitation and resuspended in PBS-Mg buffer. Characteristic wide-field and high magnification views are shown in Figure 4.27a-d. Nanostructure D is difficult to

image due to its small size, but nanostructure E can be imaged reasonably well. A majority of mono-disperse structures can be seen, with what appears to be a relatively consistent fold. As described in Section 4.4.2, particle areas were analysed for Nanostructure E and results are shown in Figure 4.27e. Using a monomer cut-off of 900 nm^2, monomer particle size (given as mean ± standard deviation) is 550 ± 90, slightly larger than nanostructure A, as expected.

Characterisation of Assemblies

Assembly formation was analysed by AGE. Samples of nanostructures A, D and E were prepared by transcription and DNAseI digestion, followed by LiCl precipitation and resuspension into PBS-Mg buffer. Sample concentrations were estimated by A_{260} absorption and adjusted to 100 nM.

Results for the nanostructure A/D assembly is shown in Figure 4.28b. Mixes were prepared at A:D volume ratios of 8, 4, 2, 1 and 0.5:1 (lanes 3-7). The most likely identity of different bands is labelled in Figure 4.28a, and possible AD, A_2D, A_3D and A_4D assemblies are seen, as expected. In lane 1, where there remains an excess of nanostructure A, some of the smaller assemblies are still seen, with the brightest assembly band corresponding to A_3D, which would not be expected if the assemblies form as designed. This observation points towards nanostructure D being too flexible to facilitate defined stoichiometric assemblies, e.g, if an incorrect pair of kissing loop interactions form between nanostructures A and D, then this could block the formation of remaining pairs, resulting in sub-stoichiometric assemblies. Nanostructure E may be a more promising prospect, as the internal structure of kissing loop interactions should result in a more rigid structure and more better-defined interaction interface.

AGE results for the nanostructure A/E assay are shown in Figure 4.28b. Mixes were prepared at A:E volume ratios if 8, 4, 2, 1, 0.5:1 (lanes 3-7). For this mixture, results are much more difficult to interpret; it is clear the two nanostructures are binding, but stoichiometries cannot be determined. Distinct bands for multimers are not observed, instead we see an upward smear in intensity. Possible explanations are the assemblies falling apart (unlikely since the A/E interaction interface is stronger than for A/B) or that multimer bands overlap so cannot be resolved. By eye, it does appear that in lanes with a 4-8 excess of nanostructure A (lanes 3-4) there are no small multimers (close in mobility to the A/E bands), so binding efficiency appears to be improved for this complex.

Because the gels are difficult to interpret fully by eye, I have used a simple method for analysing the gels which permits a more direct comparison. Firstly, the gel image is rotated 90° and a profile along each lane is plotted using the 'Plot Profile' command in ImageJ. The position and relative intensity of bands is found using the 'findpeaks' function in MATLAB, and peaks for each lane are plotted. For each band, its position is normalised relative to the position of the nanostructure A band, giving a relative band mobility. This allows comparison of bands between gels. The MATLAB script is provided in Appendix F.3 and results for each gel are shown in Figure 4.29.

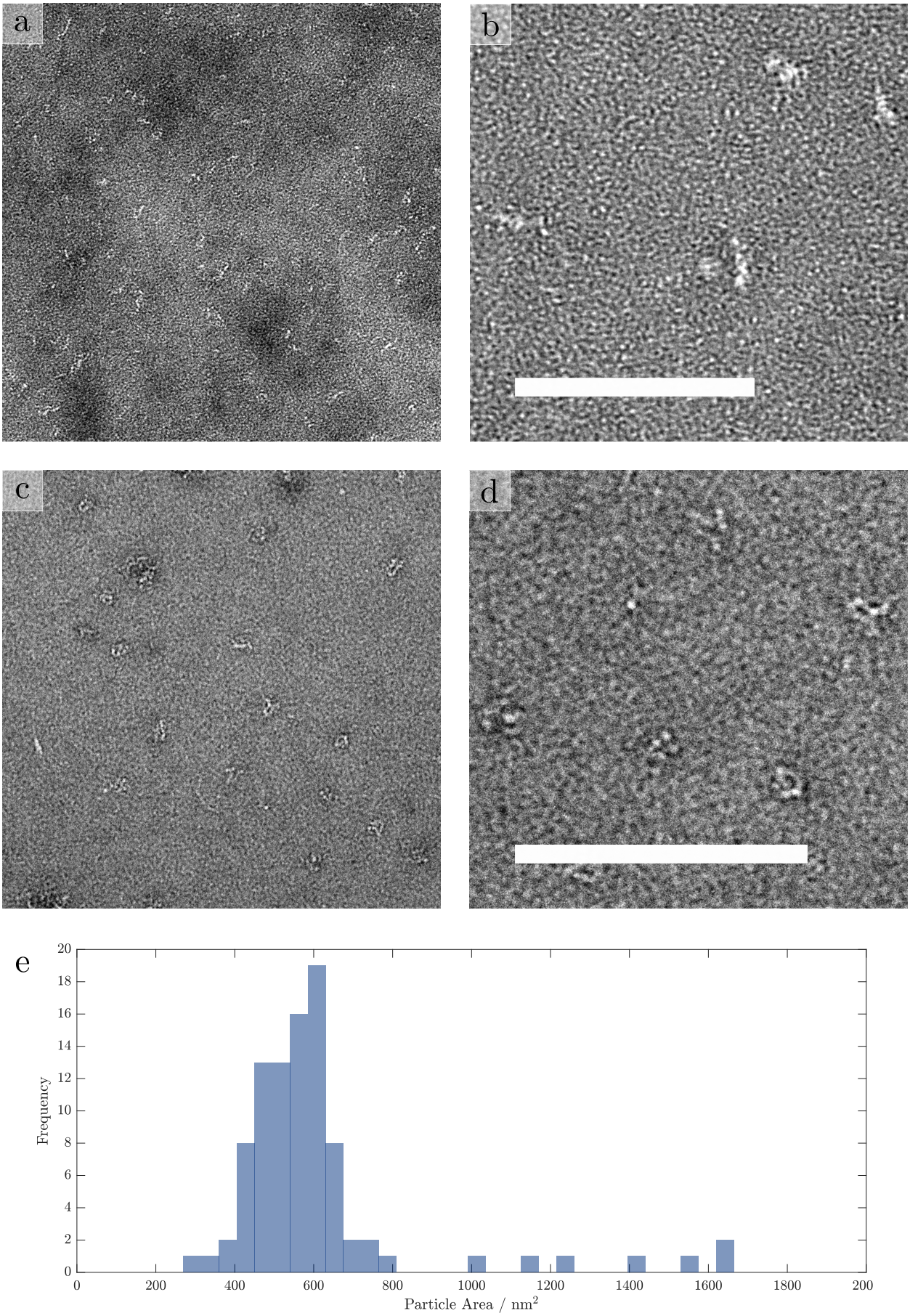


Figure 4.27: TEM overview of nanostructures D and . (a)/(b) Wide-field/high magnification views of Nano-structure D, scale bars 100 nm. (c)/(d) Wide-field/high magnfication views of Nano-structure E, scale bars 200 nm. (e) Histogram of measured particle cross-sectional areas for nano-structure E

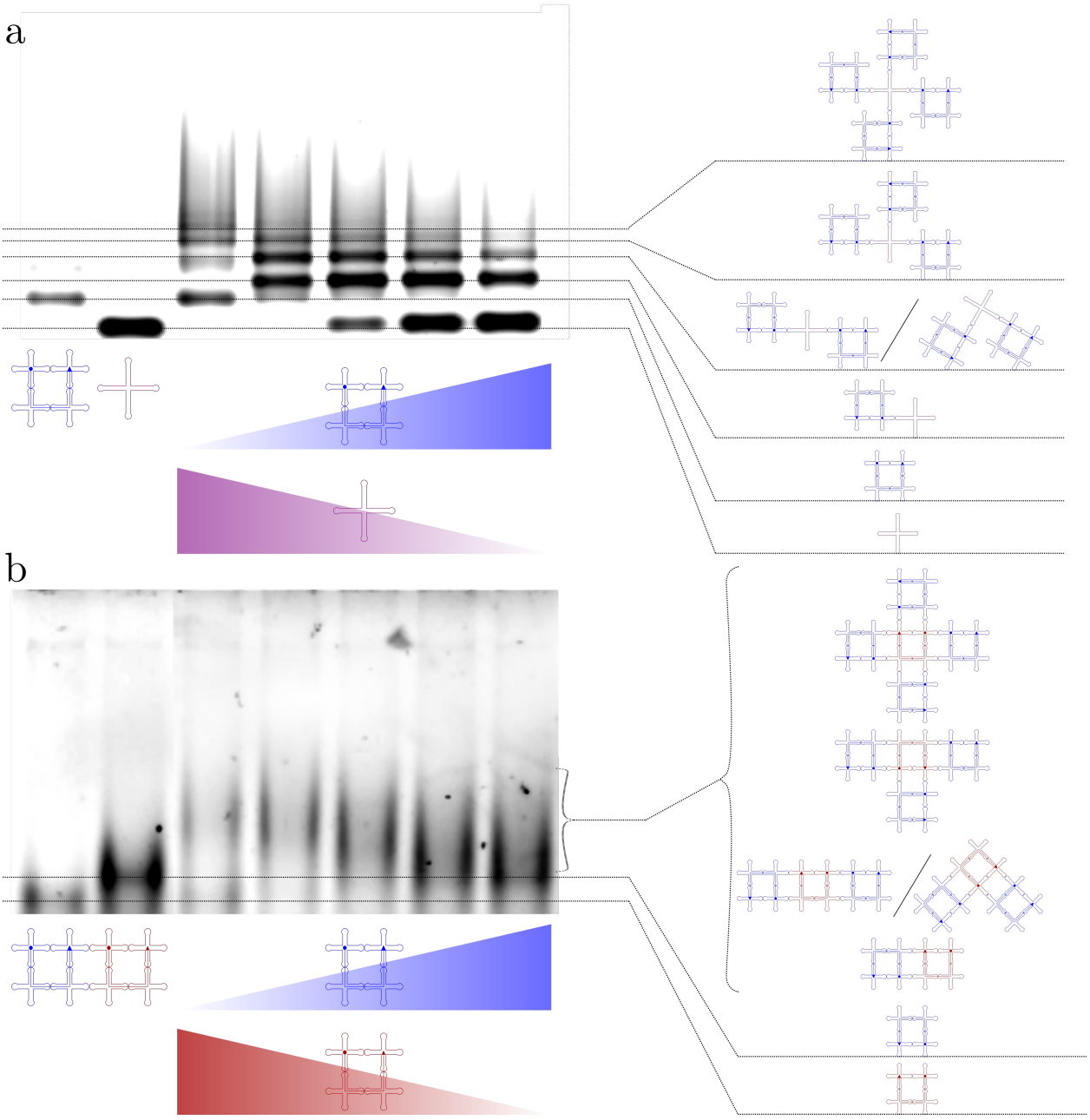


Figure 4.28: AGE assays of pentamer supermolecular assemblies. (a) AGE assay of nanostructure A binding nanostructure D. Lanes 1 and 2 contain nanostructures A and B. Lanes 3-7 contain mixes of nanostructures A and D, with increasing amounts of A and decreasing amounts of D in each successive lane. The most likely assemblies corresponding to different gel bands are indicated on the right. (b) AGE assay of nanostructure A binding nanostructure E (see details for (a) above, replacing D with E). Distinct bands for higher order assemblies are not seen, instead we observe a high molecular weight smear.

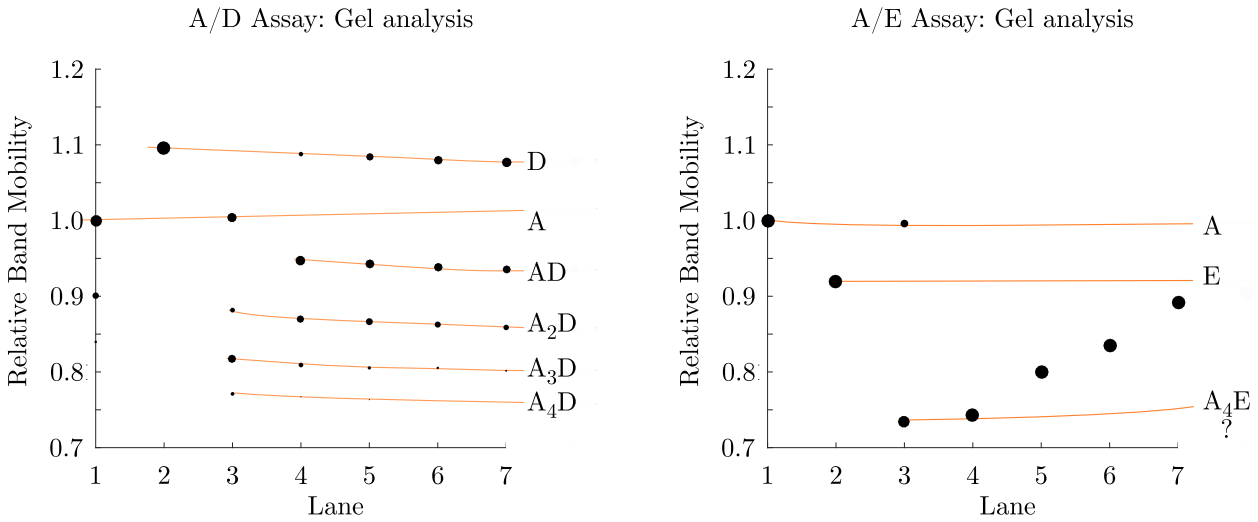


Figure 4.29: AGE analysis of assays of nanostructures D (left) and E (right) against nanostructure A. Data point sizes are scaled according to peak 'prominence' (a measure of band intensity) and corresponding complexes, where known, are labelled.

For the nanostructure A/D assay, multiple peaks are seen in each lane corresponding to the different A/D multimers. For the nanostructure A/E assay, most lanes contain a single peak, but the position of this peak shifts to gradually lower band mobility as the E:A ratio is increased, indicating the formation of increasingly large assemblies. For lane 3, (E:A = 8:1), we see two peaks. The higher mobility peak lines up with nanostructure A, and the lower peak has a relative band mobility lower than the A_4D complex, suggesting that this peak corresponds to the A_4E complex. Thus, despite our inability to resolved different complexes by gel, the pentamer A_4E complex does seem to form as designed, and at higher yield than the A_4D complex.

Options for Fluorescent Labelling

We have seen that gels with multiple nanostructure components can be difficult to analyse, partly because it is not possible to determine the exact source of different bands. It is useful to label structures with fluorophores, such that the presence of a specific structure in a band will be indicated by that structures assigned fluorescent signal. I have explored three different avenues for fluorescent labelling: using fluorescent nucleotides; using GFP recruitment via an incorporated aptamer; and using hybridisation to fluorophore-modified oligonucleotides. These methods have been tested with nanostructure E.

GFP Rectruitment My previous work with an RNA aptamer which targets GFP (Chapter 3) provides an interesting option for fluorescent labelling. The aptamer sequence can be incorporated into the nanostructure, permitting binding to GFP proteins, and so indirect labelling of the nanostructure by GFP fluorescence. This method is interesting for applications to intracellular tagging for cryo-EM which would require a combination of aptamer-functionalised nanostructures and GFP-fused target proteins.

There are many locations within nanostructure E where a GFP aptamer could be

incorporated without affecting the overall structure. I chose to replace the terminal hairpin in unit 4 with the aptamer, but it could likewise be introduced at any of the four terminal hairpins, or any of the eight external branch hairpins which are not involved in supermolecular assembly.

The aptamer sequence was introduced to the plasmid template by site-directed mutagenesis. The modified unit 4 sequence was generated by replacing the *UUCG* tetra loop on the unit four terminal hairpin with the GA sequence (E-apt sequence provided in Appendix A.1.1). Sequences for primers were designed using the NEBaseChanger tool, and the sequence modification is introduced using the NEB Q5 Site-Directed Mutagenesis Kit (using the supplied protocol) and the modified plasmid was sequenced[12].

AGE of aptamer-modified nanostructure E (E-apt) bound to EYFP, with and without a 2-fold excess of nanostructure A is shown in Figure 4.30a, with free EYFP as a reference in lane 1.[13] In lane 2 we see a relatively clean, sharp band, demonstrating that the aptamer can fold well within a large nanostructure and retains its binding behaviour. Nanostructure E-apt is therefore a reasonable starting point for a genetically-encoded RNA tag with GFP-binding behaviour. In lane 3, we can see that nanostructure E-apt is completely bound to nanostructure A since none of the original band remains and we see a decrease in band mobility. The resulting band is diffuse (as seen previously), presumably corresponding to overlapping bands of different assemblies (expected for a 2-fold excess).

Fluorescent Nucleotides Fluorescently-labelled RNA molecules are commonly required for fluorescence *in situ* hybridization (FISH) applications, and commercial kits exist to synthesise fluorescently-modified RNA strands for this purpose. I have synthesised Nanostructure E using a commercial kit,[14] with Alexa Fluor 488 modifications.

Fluorescently modified nanostructure E (E-FM) appears as a smear when analysed by AGE (Figure 4.30 b). The most likely explanations are that the transcription protocol results in a higher amount of aborted transcripts with poorly defined folds, or that the heating step required results in RNA degradation or mis-folding. Any changes to the protocol in an effort to resolve these issues resulted in undetectable levels of transcription product. Despite its poor folding, nanostructure E-FM does interact with nanostructure A (no unbound E-FM remains when mixed with an excess of A), implying that the RNA present is at least partially folded.

Based on these results it is not recommended that commercial transcription kits for fluorescently-modified RNA molecules be used for the synthesis of RNA nanostructures of this size/complexity.

Hybridisation to Fluorescent Oligonucleotides The simplest way of fluorescently labelling an RNA nanostructure is through hybridisation to a synthetic fluorescently-modified DNA oligonucleotide. For this grid-like architecture, tails can be introduced

[12]Sequencing performed by Source Bioscience.
[13]Final concentrations are 50 nM EYFP, 50 nM E-apt and 100 nM A.
[14]FISH Tag RNA Multicolor Kit, Alexa Fluor dye combination, ThermoFisher Scientific.

at the 5′ or 3′ ends of nanostructures without changing the overall strand folding. I introduced a single-stranded extension of 12 nucleotides at the 3′ end of nanostructure E by PCR (Appendix B.1.3) for hybridisation to a complementary 12 nucleotide strand with a FAM-modification (IDT) (sequence provided in Appendix A.1.1). The template for the modified structure was synthesised by PCR. AGE of FAM-labelled nanostructure E (E-FAM) with and without a large excess of nanostructure A is shown in Figure 4.30c. A single band corresponding to an A/E-FAM complex is observed, as is a substantial amount of aggregation. Analysing this gel as in the previous section, the bound assembly has a relative peak mobility of 0.77 compared to nanostructure E-FAM, which would be about 0.71 compared to nanostructure F. This is consistent with the proposed A_4E band in Figure 4.29, supporting the conclusion that nanostructure E is fully bound (i.e. each face is bound to a copy of nanostructure A). This band is somewhat diffuse, which may be a gel artefact, but may also signal that the A_4E complex can adopt a range of geometries with different mobilities. The aggregation is also a concern. This is most likely due to multiple versions of nanostructure A binding to a single nanostructure E interaction interface (i.e. through fewer than 4 kissing loops) leading to higher order assemblies with unrestricted polymerisation. This is unlikely to be a problem at low nanostructure concentrations, but could be limited by reducing the strength of the kissing loop interactions.

For comparison, I also investigated the nanostructure D/A interaction using FAM-labelled nanostructure D (template also generated by PCR). This is shown on the far right of Figure 4.30c. Here we see three very distinct bands in the second lane. Relative mobilities of the three bands are ≈ 0.68, 0.72 and 0.77 relative to nanostructure D, and an estimated ≈ 0.74, 0.79, 0.85 relative to nanostructure A. This does not line up precisely with bands shown in Figure 4.29, but most likely correspond to A_2D, A_3D and A_4D respectively. This means that even in a large excess of A, D is not fully bound, supporting the earlier suggestion that the flexibility of nanostructure D results in sub-stoichiometric assemblies.

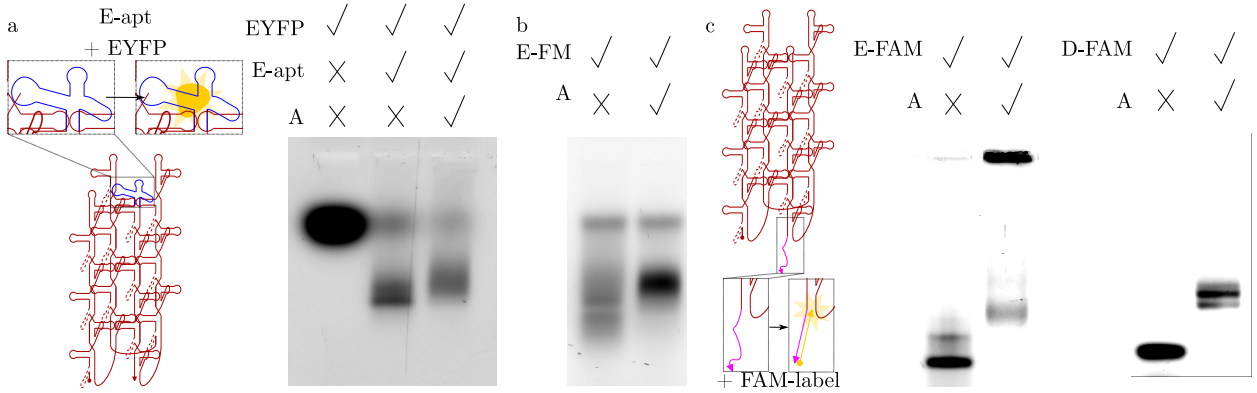


Figure 4.30: AGE overview of fluorescent labelling options for nanostructure E. (a) Fluorescent labelling by GFP recruitment via incorporated aptamer sequence. (b) Fluorescently-modified rNTPs. (c) Hybridisation to a fluorescently-modified DNA oligonucleotides (shown for nanostructures D and E). See text for details.

Negative-Stain TEM

A/D and A/E assemblies were also analysed by negative-stain TEM. Structures were prepared by transcription (Appendix B.2.2), purified by LiCl precipitation (Appendix B.3.2), and resuspended in PBS-Mg buffer. Mixed samples were prepared of A/D and A/E with a final nanostructure A concentration of 10 nM and a final D/E concentration of 2.5 nM. A large fraction of free structures are observed for both mixes (Figure 4.31a and b), which may indicate that the interaction interface is too weak to ensure assembly at the low concentrations required for negative-stain TEM. Of course, the assembly strength is easily tunable through the strength of the contributing kissing loops. Assemblies are sometimes difficult to identify, but some of the clearer examples of A/D and A/E assemblies are shown in Figure 4.31c and d. The main difference that we see between the two assembly types is for the trimer A_2D and A_2E assemblies. For A_2E we see two geometries of assembly: an L shape assembly where the two nanostructure A copies occupy adjacent sites (top three examples) and a straight line assembly where they occupy opposite sites (bottom two examples). For A_2D only the straight line assemblies are seen, which is likely related to the greater flexibility of nanostructure D. Given more freedom of movement, the satellite structures will prefer this positioning as it minimises repulsion. Other than this discrepancy, the assemblies for the different nanostructures look qualitatively similar.

4.5.3 Discussion

In this section, I have shown the design and synthesis of two new nanostructures: D and E, each designed to bind nanostructure A in a 4:1 ratio to yield a superstructure with 5 components. AGE and negative stain TEM results indicate well-folded structures, but gel analysis of assembly formation leads to a few conclusions which may inform future developments of this architecture:

- Nanostructure D, comprising a single unit, is too flexible to enforce stoichiometric assemblies. For well-defined geometries only nanostructures with an internal structure of kissing loops should be used.

- Nanostructure E, comprising four units, shows improved binding efficiency relative to nanostructure D, but it is not clear whether the assemblies have a consistent structure. The correct geometry could be enforced by making all kissing interactions unique and/or by replacing the single-stranded linkers between units with structured linkers.

- For *in vivo* assembly, it would advisable to tune the strength of intermolecular kissing loops, to ensure an interface requires all four loops to form stably. This would further help minimise the fraction of incorrectly-formed assemblies.

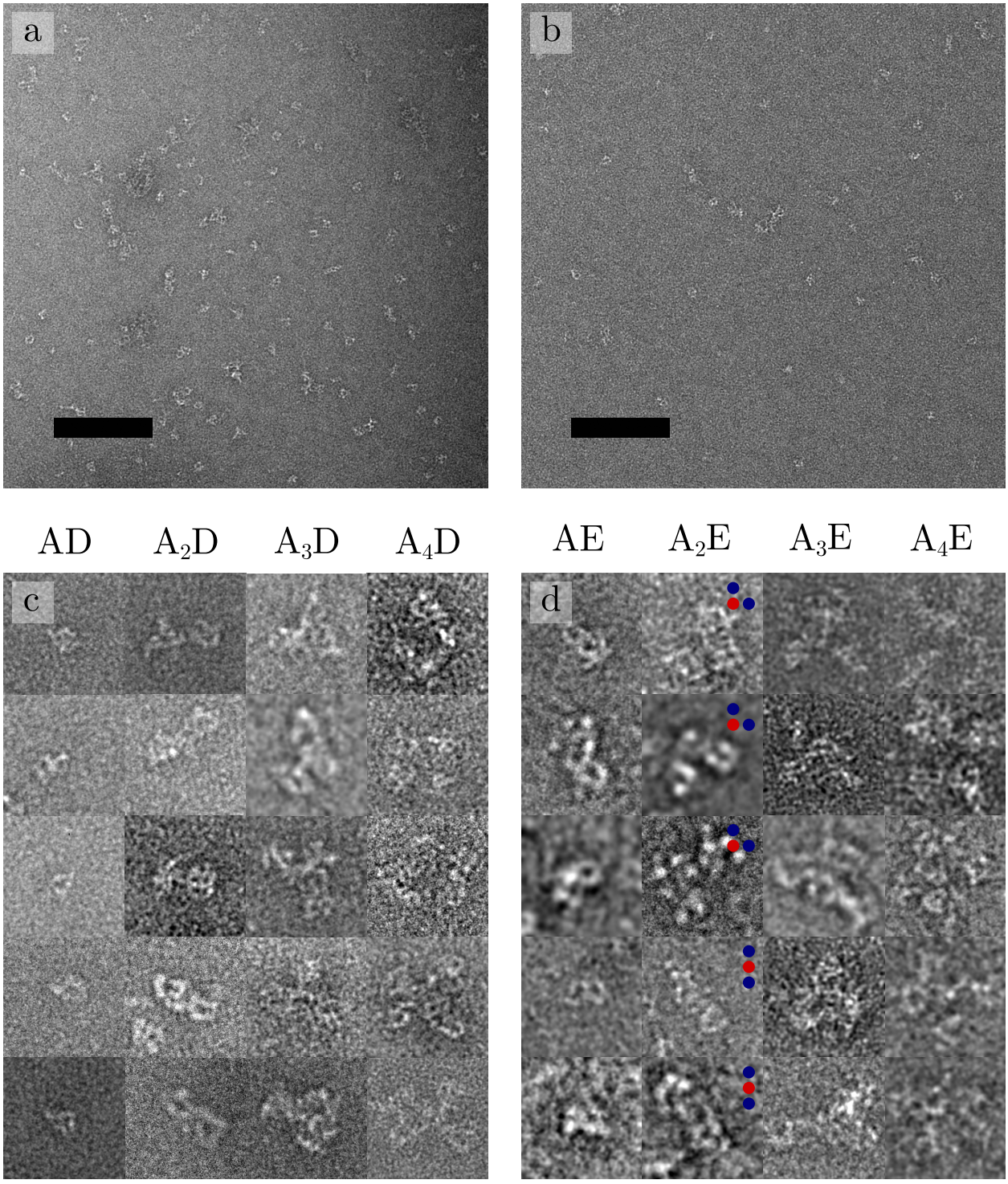


Figure 4.31: TEM overview of cross-shaped assemblies. (a) and (b) are representative views of the A/D and A/E mixes (scale bar 200 nm). (c) and (d) are examples of different supermolecular assemblies (labelled by column). All views are 75 nm × 75 nm.

4.6 Summary

In this chapter I have established a novel grid-like architecture for building 3D RNA nanostructures by co-transcriptional folding. In this architecture, nanostructures comprise a set of mutually interacting units, arranged as pixels on a square lattice. Each unit is a branched hairpin structure, and neighbouring units are connected by a pathway of single-stranded linkers and by kissing-loop interactions between complementary hairpins.

I firstly showed the design and synthesis of a 5-unit (pentomino) nanostructure using this method. The co-transcriptional folding pathway is considered during the design: individual units are designed in parallel, with the expectation that these domains will fold as they are transcribed and become kinetically trapped in the lowest energy conformation. Following the design of all sequences and linkers, the equilibrium fold of the full nanostructure is also analysed. The predicted secondary structure is consistent with the design, indicating that this design architecture can also be applied for nanostructures folded by annealing. Analysis by gel electrophoresis, negative-stain TEM and cryo-EM suggest a majority of well-folded structures and a significant population of multimers.

Based on results from this preliminary design, in particular challenges with template synthesis, nanostructure visibility and the presence of multimers, I then introduced a modified version of the architecture. This modified architecture has slightly different unit structures and a capacity for supermolecular assembly. This new architecture is used to assemble 5 new nanostructures (named A-E). AGE and negative-stain TEM results suggest a large majority of well-folded structures in all cases. These nanostructures are also used to demonstrate programmable supermolecular assembly of dimers (AB), trimers (A_2C) and pentamers (A_4D and A_4E) with mixed success. In particular, comparing results for nanostructures D and E, we can see that internal structure (intramolecular kissing loops) are key to providing the structural rigidity required for stoichiometric assemblies. I have also demonstrated the incorporation of an aptamer into a nanostructure (and subsequently nanostructure assembly), and shown that the aptamer retains its binding properties.

This nanostructure architecture provides one viable route toward designing genetically-encoded tags for cryo-EM. For *in vivo* supermolecular assembly, it may be necessary to tune the intermolecular kissing loop strength, to promote the formation of the intended assemblies whilst preventing sub- and super-stoichiometric assemblies. Structured linkers between units may also improve folding, and may help prevent cleavage in an *in vivo* environment.

Chapter 5

RNA Nanostructures: Origami-based Approach

In this chapter, I will present methods for extending a well-established 2D RNA origami architecture [114, 115] into 3 dimensions. 3D designs can be generated from a stack of interacting 'sheets', with each sheet resembling a 2D origami tile. The final designed architecture resembles that of a square-lattice DNA origami nanostructure. In this chapter, I consider two possible options for inter-sheet interactions: paranemic crossovers which form by hybridisation of complementary bulge loops on adjacent sheets (named PX_{major} motifs), and a combined bulge-loop/kissing-loop (BLKL) motif which form by base-pairing between hairpins on adjacent sheets. For the PX_{major} motif, I will show that the designed motif can form in isolation from two constituent strands, and I will present the design and synthesis of a 4-sheet nanostructure held together with these motifs. I will show that these structures do not fold as intended, and suggest possible explanations. For the BLKL motif, I will show the design and synthesis of a 2-sheet hetero-dimer, as well as single-stranded nanostructures with the same designed architecture.

5.1 Author Contributions

All the work in this chapter (design, synthesis and characterisation of nanostructures and motifs) was performed by Emma Silvester.

5.2 Extending 2D Origami into the Third Dimension

5.2.1 2D Origami

In parallel to the work presented in Chapter 4, I explored a different avenue towards RNA nanostructure design. At the beginning of my DPhil, the only published examples of large-scale single-stranded RNA nanostructures folded co-transcriptionally were Cody Geary's tiles [114], ergo this seemed a reasonable starting point for RNA design. These tile structures

were introduced Section 1.4.3, but I will discuss these nanostructures in more detail here, with emphasis on the motifs involved and their relevance to the co-transcriptional folding pathway. Each tile comprises an array of up to 6 anti-parallel helices, precisely organised by a 'dovetail seam' crossover pattern and a central seam of kissing-loop motifs.

Crossover Motifs

In general, double-crossover motifs can form between anti-parallel (A) or parallel strands (P), and the spacing between crossovers may be an odd (O) or even (E) number of half-turns. Geary employed a 'P-stick' model to calculate optimal crossover positions for these motif types, in which the position of the phosphorous atoms are used as a proxy for nucleotide position and crossovers positions are specified so as to minimise distance between contributing phosphorous atoms/nucleotides [116]. Tiles held together with AO and AE type crossovers were both achieved experimentally [114]. Strand routing diagrams for 4-helix tiles of both types are depicted in Figure 5.1.

In Geary's designs, for structures comprising at least three helices, there exist helical domains separating a double crossover linking the nth and $n + 1^{th}$ helices and a double crossover which links the $n + 1^{th}$ and $n + 2^{th}$ helices (these domains are highlighted in orange in Figure 5.1). These base pairs form later in the folding pathway than the helical domains either side (which are simple hairpin stems which can form immediately as they are transcribed). To avoid topological folding traps, it is therefore preferable that these domains be 'paranemic'. In general, this means that the interaction between helices (in this case the dovetail base pairs between crossovers) can be fully separated without unwinding of the contributing helices (i.e. without breaking any other base pairs). To fulfil this requirement, the 'dovetail seam' domains must be shorter than a helical turn. Spacings of 3 nucleotides and 2 nucleotides were used for AO and AE tiles respectively. The collective crossover pattern and intervening nucleotides are termed the 'dovetail seam' (outlined in grey and indicated by arrows in Figure 5.1).

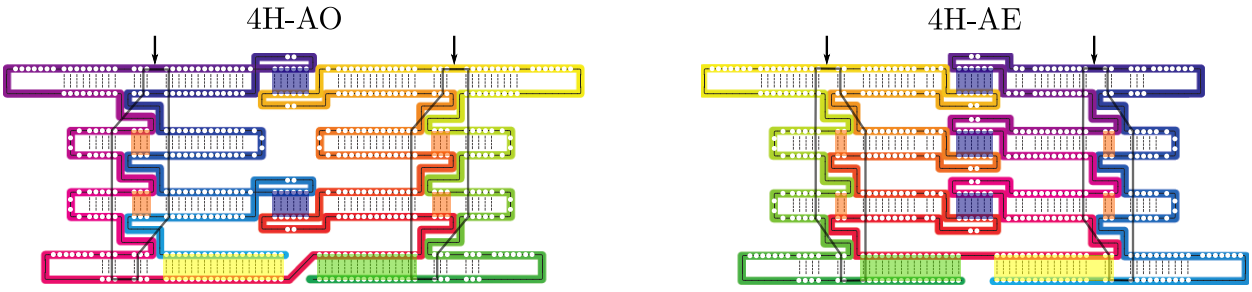


Figure 5.1: Figure and caption adapted for Geary et al. [114]. Strand-path diagrams for 4H-AO and 4H-AE tiles, showing order of synthesis, 5′ to 3′, from blue to purple to red to orange to yelow to green. Nucleotides contributing to the dovetail seams are outlines in grey, base pairs formed between double crossovers are highlighted in orange. The 5′ and 3′ terminal helical domains are highlighted yellow and green respectively.

Kissing Loop Motifs

The central seam of an RNA tile comprises a set of 180° kissing loop motifs (the same motifs used for intermolecular assembly in Chapter 4, highlighted blue in Figure 5.1). These kissing loop motifs are another example of paranemic structures, generated by the formation of six base-pairs between two fully-formed hairpin loops, without breaking of the hairpin stem. By combining sets of dovetail seams and kissing loop seams, arbitrary 2D origami designs inspired by space-filling curves have been designed [116], but currently only tiles (the simplest reduction of this arhictecture) have been demonstrated experimentally. Both 180° and 120° kissing loops were also used for super-molecular assembly of tiles into lattices, but this is not relevant for my intended application of discrete 3D nanostructures.

Terminal Helical Domains

The terminal helical domains (highlighted in yellow and green in Figure 5.1), like the dovetail seam and kissing loops, are long-range interactions. There is a significant delay between the transcription of the first and second strand in these helical domains (unlike the rest of the structure, which consists of hairpins). From a topological point of view, the 5′ helical domain (yellow in Figure 5.1) should form without restriction, since this can fold before the kissing loop partners on the second half of the structure are transcribed. The 3′ helical domain (green in Figure 5.1), however, may introduce topological folding problems. If the internal kissing loops were to form prior to the synthesis of the 3′ end of the structure, then the formation of this helical domain would be blocked. Succesful co-transcriptional folding of these structures suggests that, in fact, the kissing loops fold on a longer time scale than the secondary structure, and that this is not a practical concern. However this issue may need to be revisited as structures become increasingly large. Likewise, a reduction in transcription speed (e.g. due to rNTP concentration or polymerase source), may result in mis-folding if the transcription speed becomes slower than kissing loop formation.

5.2.2 Options for building in 3D

In order to provide adequate contrast for cryo-EM, it is necessary to extend this framework into the third dimension. Inspired by lattice-based DNA origami I proposed stacking ‘sheets’, each resembling a single tile, into a 3D structure, held together through tertiary interactions. Here I discuss three possible mechanisms for achieving this goal, listed below.

1. Extending the ‘dovetail’ seam to 3D.

2. Connecting 2D sheets by paranemic crossovers.

3. Connecting 3D sheets via combined bulge-loop/kissing-loop motifs.

Dovetail Seam: 2D to 3D

The simplest option, conceptually, for building RNA origami in 3D would be to extend the dovetail seam itself into the third dimension. This entails positioning additional double crossovers between adjacent helices on subsequent sheets, and establishing a consistent strand routing which circumnavigates the structure. The RNA A helix has a different geometry to the DNA B helix, and does not lend itself as well to the traditional square and honeycomb lattice arrangements of helices used for DNA origami. The RNA helix geometry does, however, fit neatly with a 'skewed' square lattice, in which each helix has four neighbours, as in a square lattice, but where the 'lattice angle' (see Figure 5.2b) is not 90°. The pattern of crossovers indicated in Figure 5.2 gives a lattice angle of 98°, which a close approximation to a square lattice.

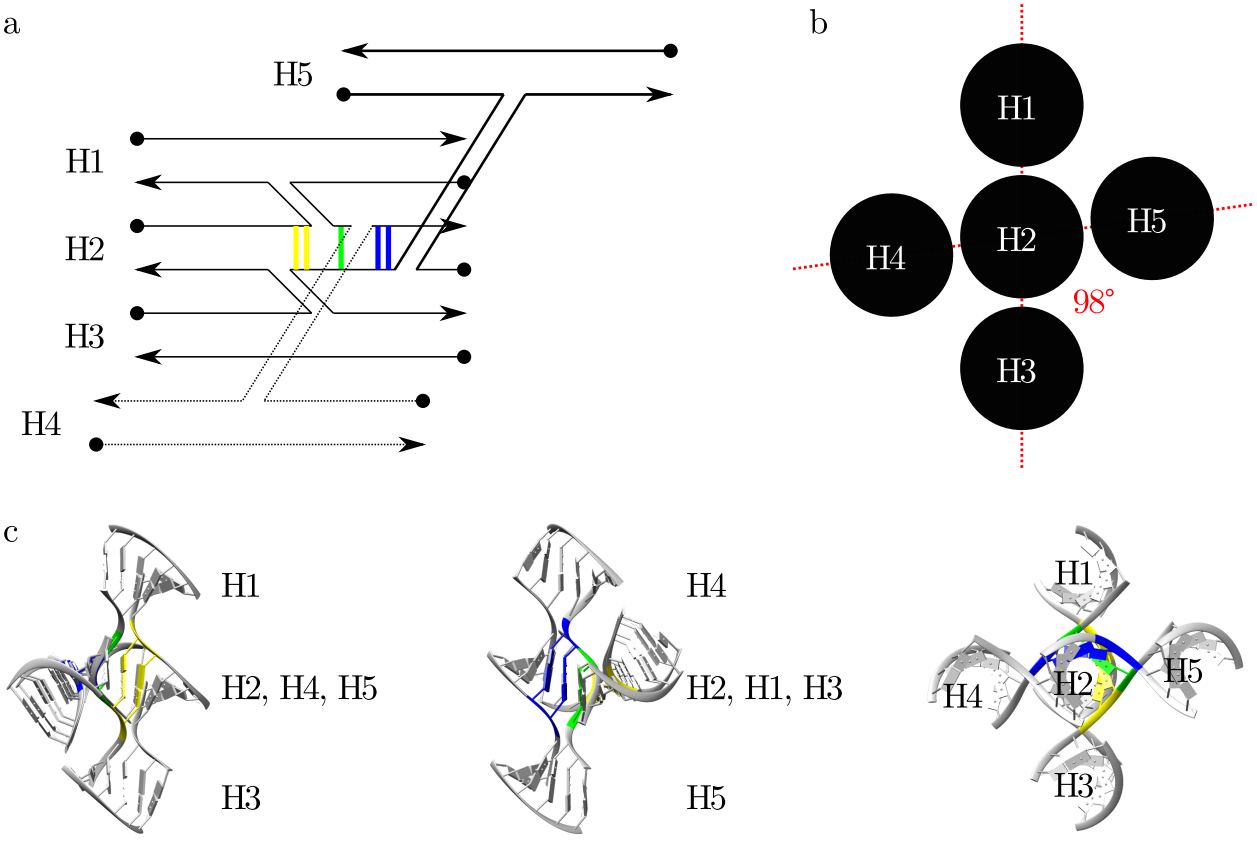


Figure 5.2: Crossover positioning for a skewed lattice arrangement of helices. (a) Expanded view of helices, showing base pairs between crossovers on H2, linking between all its neighbours (H1 and H3 within the sheet and H4 and H5 as neighbouring helices on sheets above and below). (b) End view of helices on skewed lattice. (c) Orthogonal views of 5-way junction between helices H1, H2, H3, H4 and H5, with nucleotides coloured as in (a). Atomic model rendered in Chimera.

Although a strand routing for 3D structures can be generated using this pattern of crossovers, the folding pathway is complicated, and the structure is held together in 3D by a network of extremely weak 1 and 2 base pair interactions. Note that Figure 5.2 shows only the crossovers of H2 with its four neighbours, but each of these will have their further crossovers with other neighbours, e.g. crossovers between H5 and an in-sheet neighbour

would exist 1 nucleotide prior to the H2/H5 crossover shown. For an $m \times n$ lattice, a network of $2mn - (m + n)$ double crossovers, each separated by 1 or 2 base pairs, would be the motif locking the structure together. Geometrically, crossovers could be shifted by $11 \times n$ base pairs (where n is an integer), but this would introduce significant topological problems (see discussion of crossover motifs). This strategy does not appear to offer a folding pathway conducive to co-transcriptional folding. It would be instead preferable for sheets to fold independently (through the 2D origami folding pathway), with paranemic inter-sheet interactions forming between folded sheets.

Paranemic Crossovers

A paranemic assembly motif, once formed, comprises helices which have parallel crossovers at every possible point along the length their length. The simplest reduction of this motif is the paranemic crossover which consists of two parallel double crossovers between helices separated by a helical half turn. Physically these motifs are generated by base-pairing between bulge loop motifs on two helical domains, occasionally referred to as 'bubble cohesion'.

At the time of generating these designs, paranemic crossovers had not been incorporated into RNA nanostructures, although these have since been used extensively in the construction of single-stranded origami-like nanostructures [110]. In this paper [110], as well as in studies of paranemic crossovers in isolation [175], the difference in helicity between the DNA B helix and the RNA A helix was considered in design, but another key geometric difference: the major/minor groove separation, was not addressed. For incorporation into a 3D nanostructure it will be important to optimise the crossover geometry.

Two different PX geometries can be achieved with RNA helices: a $\mathrm{PX_{minor}}$ motif which spans the minor groove, and a $\mathrm{PX_{major}}$ motif which spans the major groove. Optimal $\mathrm{PX_{minor}}$ and $\mathrm{PX_{major}}$ crossover separations were determined using the steps outlined below.

1. PDB files of two RNA A helices were loaded into Chimera.[1]

2. Helices were rotated and aligned so that one parallel crossover is perfectly aligned. The steps for alignment are described by Geary et al. [115].

3. Separation between pairs of nucleotides (using the locations of the nucleotide phosphorous atoms) on the two contributing helices were determined using the 'getcrd' command in Chimera.

4. Nucleotide positions for the second crossover were designated according to the closest nucleotide pairs identified in step 3. The crossover separation was found to be 3 base pairs for $\mathrm{PX_{minor}}$ and 8 base pairs for $\mathrm{PX_{major}}$.

[1]PDB files of RNA A-helices obtained from the make-na server: http://structure.usc. edu/make-na/server.html

PX_{major} was chosen for further investigation, since a 3 base pair interaction will be transient even at high local concentration and an 8 base pair design offers more freedom for designing many orthogonal sequences.

An atomistic model for PX_{major} was generated from the aligned helices, using the technique outlined in Section 4.3.1. In the PDB file, both helices were capped with *UUCG* tetraloops, to give two contributing strands which are renumbered and labelled (SwissPdb Viewer [165]) then ligated independently (MATLAB, Appendix F.2), and then recombined into a single PDB file (Chimera [124]). The contributing bulge loop strands and fully-formed paranemic crossover are shown in Figure 5.3.

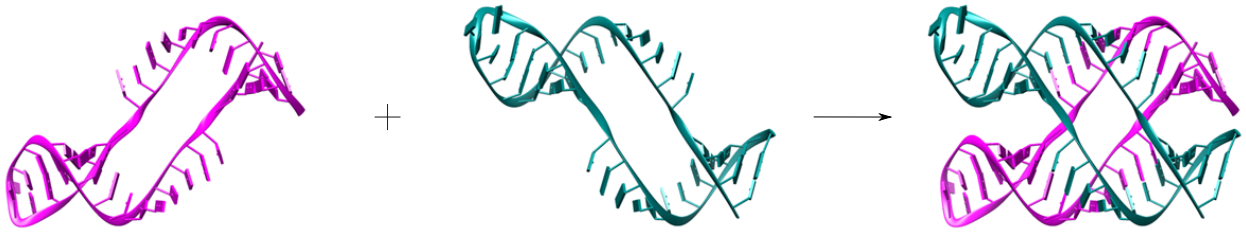


Figure 5.3: Atomistic model of PX_{major} formation, rendered in Chimera. The two contributing strands are shown in pink and teal.

Since this geometry has not been studied previously, I designed sequences for PX_{major} formation to test assembly. Sequence design methodology was based on the protocol used by Afonin et al. [175].

1. In dot/bracket notation, the structures for the two contributing strands are, when unbound:
   (((((((.....(((((...)))))).....)))))))+(((((((.....(((((....)))))).....)))))))
   and, when bound:
   ((((((([[[[[[[[(((((...)))))[[[[[[[[)))))))+(((((((]]]]]]]](((((....)))))]]]]]]])))))))

2. Eight-nucleotide sequences 'A' were generated randomly in MATLAB. For each sequence A (blue in nucleotides in dot-bracket notation above) a sequence 'B' (red notation) is also generated randomly with some constraints. B is designated as the bulge loop partner to A, and as such should not interact with A. To manually reduce interaction possibilities, for A(n) = U or C, then B(9-n) = U or C. Likewise if A(n) = A or G then B(9-n) = A or G. This serves to reduce base-pairing between bulge loop domains in both contributing strands (including $G \cdot U$ pairs). From these sequences the complementary sequences A* and B* are also generated (yellow and green notation respectively).

3. Only sequences with GC content of at least 62.5% were retained, so that PX formation is more energetically favourable.

4. The 'designer' function of RNAsoft[2] [176] is used to design sequences for the first contributing strand (with A/B bulge loop sequences). Sets of sequences are only

[2] www.rnasoft.ca

retained where the correct structure can be generated (distance to desired structure = 0).

5. Of remaining sequences, the same check was performed for the second contributing strand (with B*/A* bulge loop sequences) and only those with the correct structure are retained.

6. Using full sequences generated in RNAsoft, the tertiary structure of the two contributing strands were verified through the 'predict' feature of the Nanofolder web server [174]. It was found that only careful specification of the capping nucleotides, those either side of the bulge loop domains, is important (to prevent branch migration).

7. Of the remaining sets of sequences, four were selected and checked for orthogonality. The selected sequences are shown in Table 5.1

PX #	A	B	A*	B*
1	UCCCCGAU	CAGCUUCC	AUCGGGGA	GGAAGCUG
2	GGUGGGUA	AGGGGGUG	UACCCACC	CACCCCCU
3	CUCCACCG	ACCACUCC	CGGUGGAG	GGAGUGGU
4	GUCCUCCU	CCCCUCUG	AGGAGGAC	CAGAGGGG

Table 5.1: Bulge loop sequences for PX motifs used in RNA nanostructures.

The PX1 sequence was used to investigate PX formation experimentally. The NUPACK-designed sequences for each strand were used as a starting point, but appopriate T7 RNA polymerase starter sequences were included at the 5′ end of both strands. To prevent non-PX dimerisation between the two strands, the consensus sequence was not used, but the starter sequences *GUCAGA* and *GGGGUC* were used for strands 1 and 2 respectively, both of which are expected to give >80% transcription yield relative to the consensus *GGGAGA* starter sequence [138]. The sequences of the contributing strands are converted to DNA, the T7 RNA polymerase promoter region added at the 5′ end, and both duplex strands were ordered from IDT[3] (sequences provided in Appendix A.2.2). Transcripts of the PX strands were prepared by transcription (Appendix B.2.1), purified by ethanol precipitation (Appendix B.3.3), resuspended in 1×TBE + 6 mM $MgCl_2$ and adjusted to the same concentration according to the A_{260} measurement. A 1:1 mix of the two PX strands was then prepared, and incubated for 1 hour at room temperature. Samples were analysed by native PAGE (Appendix E.2). The mechanism and results from this experiment are shown in Figure 5.4. By gel, we can see that the two strands interact completely to form a hetero-dimer, shown in lane 9. This is presumed to be the designed PX motif, since no dimerisation is predicted in NUPACK (which does not allow pseudo-knotted structures).

Based on these preliminary results, a PX_{major} motif, consisting of two double crossovers between parallel helices spanning a major groove of the RNA A helix, appears to be an appropriate motif to incorporate into RNA nanostructures.

[3]Strands ordered as 25 nmole DNA oligo, standard desalting.

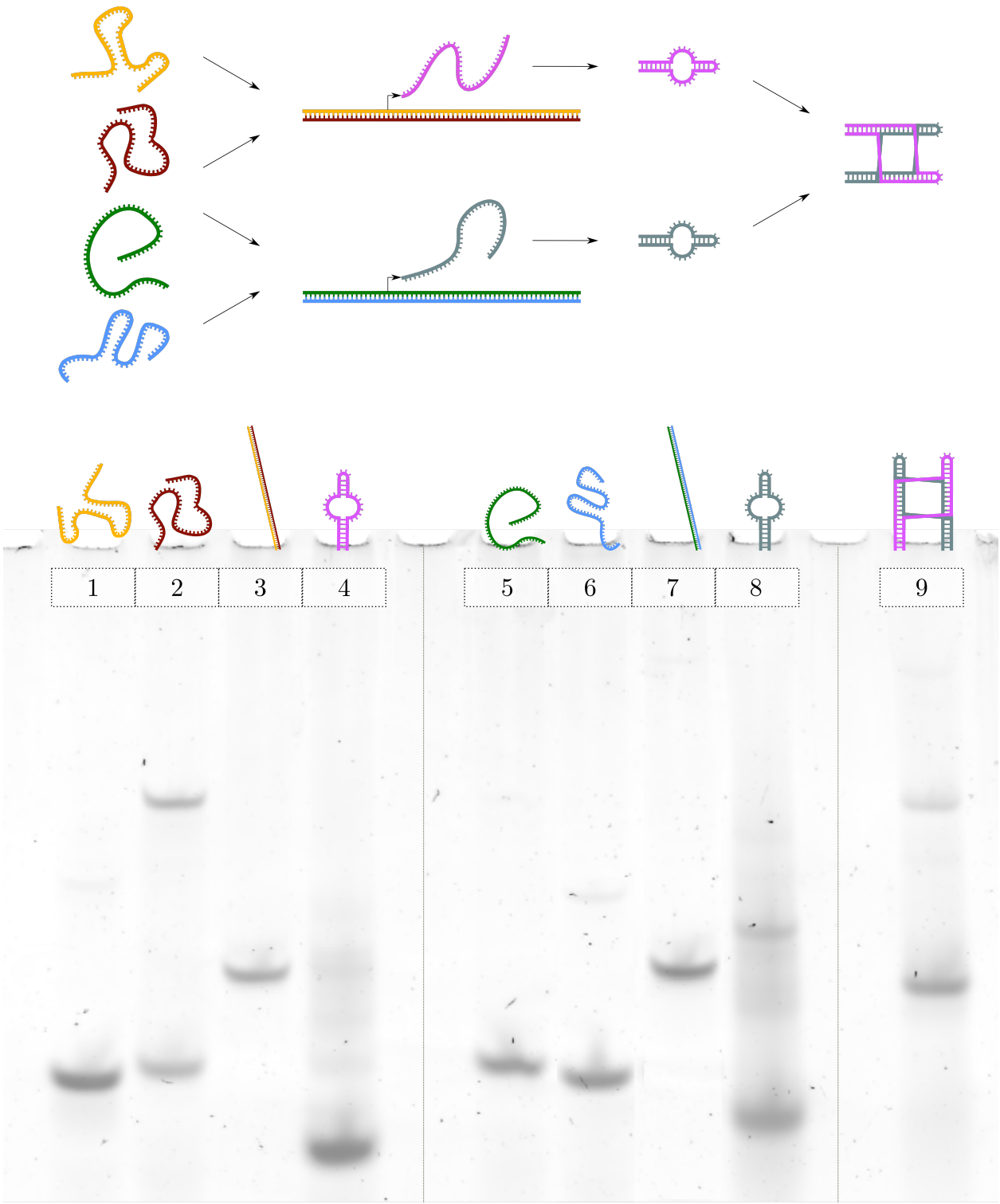


Figure 5.4: Top: Mechanism of PX motif formation experiment: DNA strands are annealed to form a template duplex, from which the two contributing PX motifs are transcribed. Below: PAGE of PX motif formation: lanes 1, 2, 5 and 6 are the DNA template strands, lanes 3 and 7 are the annealed template duplexes, lanes 4 and 8 are the transcribed bulge loops involved in the PX motif (relatively clean, single bands for both). Lane 9 contains a 1:1 mix of the two PX components.

Combined Bulge-Loop/Kissing-Loop Motif

Kissing loop motifs are some of the most widespread structural motifs in RNA synthetic biology, most notably the 180° kissing loop motif of the HIV-1 DIS, and the 120° loop-loop complex of the RNA i/ii inverse loop, both utilised in the grid-like architecture presented in Chapter 4. Given their widespread applications in RNA nanostructures and paranemic nature, it is appealing to utilise kissing loops to extend the tile framework into 3D.

Using a kissing loop interaction to form 3D connections between non-coaxial parallel helices requires a kissing loop angle of 0° (angle as defined by RNAjunction [109]). This geometry has not been identified in biological systems to my knowledge, but could be achieved by combining a kissing loop motif with bulge loops motifs on one or both of the two parallel helices to yield the correct collective geometry. The simplest solution is to introduce a 1° bulge loop in just one of the helices, and using the 180° HIV-1 kissing loop motif (actually 179° despite how it is usually referred to), as shown in Figure 5.5c to connect helices. The 1° bulge loop sequence and structure (shown in Figure 5.5a) was derived from a crystal structure of the 50s ribosomal subunit from Deinococcus radiodurans (PDB ID: 1NJP) [177]. The sequence and structure of the 180° kissing loop is provided again in Figure 5.5 for reference. This combined geometry also gives an inter-sheet helical separation of $\approx$ 22 Å, which is equal to the intra-sheet helical separation [115] which may not be necessary depending on how sheets are connected through the scaffold, but the dense packing enforced by this rules is nonetheless useful for providing structural rigidity and high cryo-EM contrast.

By aligning crystal structures of the various motifs in chimera, and ligating the PDB file in MATLAB, I have generated a PDB structure for the combined bulge-loop/kissing-loop motif, (Figure 5.5c) referred to as BLKL, which can then be incorporated into a 3D nanostructure model as an alternative to the PX motif. The use of the 180° HIV-1 kissing loop motif is serendipitous, as numerous orthogonal sequence variants for the loop have already been identified (see Section 4.4).

5.3 3D Origami: Building with PX motifs

5.3.1 Design

Concept and Atomistic Model

As a first attempt to extend the RNA origami architecture into three dimensions, I designed a nanostructure comprising 4 sheets, where each sheet resembles a 2D RNA origami tile with constituent helices elongated/shortened to provide an asymmetric cross-section. Sheets are designed to interact through paranemic crossover (PX) motifs which form by base-pairing between complementary 8 nucleotide bulge loops on equivalent helices in adjacent sheets. Sheets are also linked through the scaffold routing, resulting in parallel double crossovers between sheets (equivalent to the crossovers that appear twice in each PX motif). This results

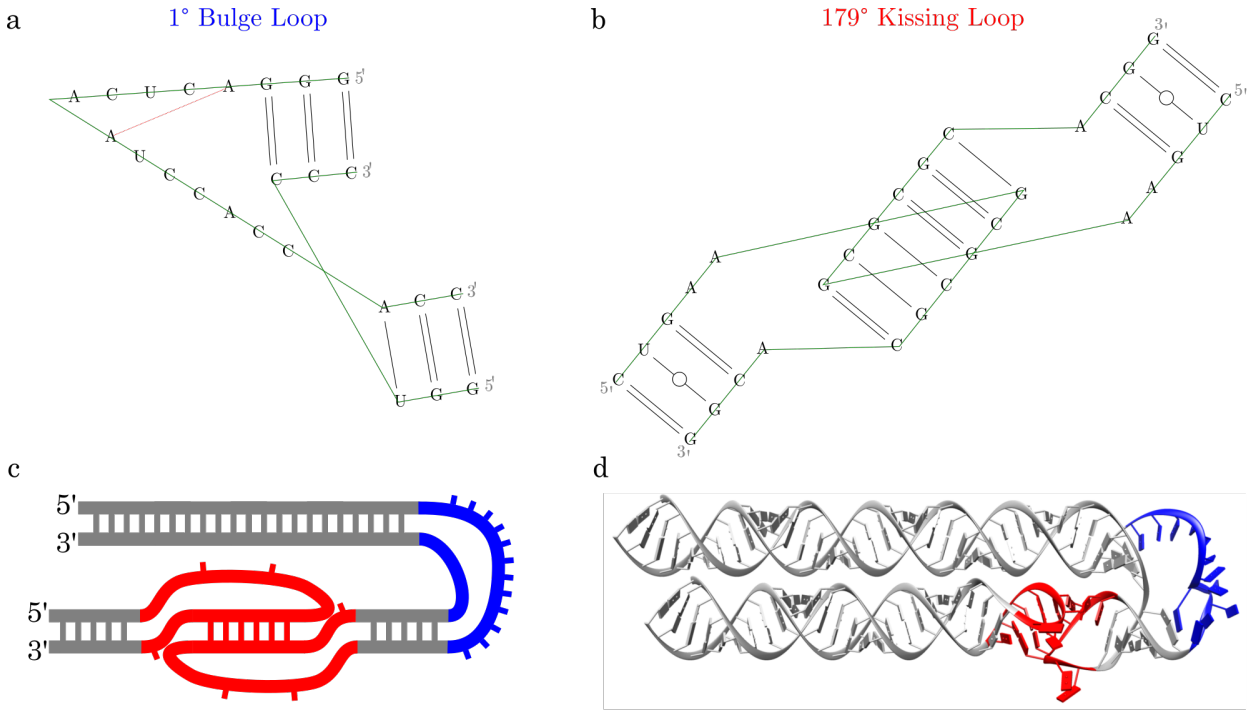


Figure 5.5: (a) Sequence and structure of the wild-type 1° *bulge loop. Backbone, secondary structure and non-Watson-Crick tertiary contacts are depicted in green, black and red respectively. (b) Sequence and structure of the wild-type* 179° *kissing loop, same colour coding as in (a). (c) Schematic depiction of combined bulge-loop/kissing-loop (BLKL) interaction interface. Helical domains are depicted in grey, the kissing loop domain is shown in red, and bulge loop nucleotides are depicted in blue. (d) 3D atomic model of BLKL interaction interface, generated in chimera, same colour coding as in (c).*

in a 3′ tail (3′T) in the co-transcriptional folding pathway, which pairs with short domains between crossovers which remain single-stranded until the final transcription steps. This 3′T motif is paranemic. A schematic of the designed structure, before and after formation of tertiary contacts in shown in Figure 5.6a, with domains coloured according to their folding speed. The only slow-folding, non-paranemic motifs are the terminal helical domains for each sheet, which were not seen to block folding of 2D RNA origami (Section 5.2.1), so are not expected to cause folding problems in this context. A second representation of this schematic is shown in Figure 5.6b, where the scaffold backbone is coloured by sheet number. The sheet numbering is defined according to transcription order rather than position in the folded structure. Atomistic models of the various tertiary motifs are also shown.

An atomistic model of the structure was generated in Chimera using the general method described in Section 4.3.1, and described more specifically below.

1. An atomic model of a single ‘sheet’ is generated according to the protocol described by Sparvath et al. [115]. As with the pentomino unit, ligation of the strand was performed by a MATLAB script (Appendix F.2).

2. Four sheets are loaded into chimera [124], and aligned to give correctly positioned inter-sheet crossovers. Rules for crossover positions are shown in Figure 5.7. Each sheet is

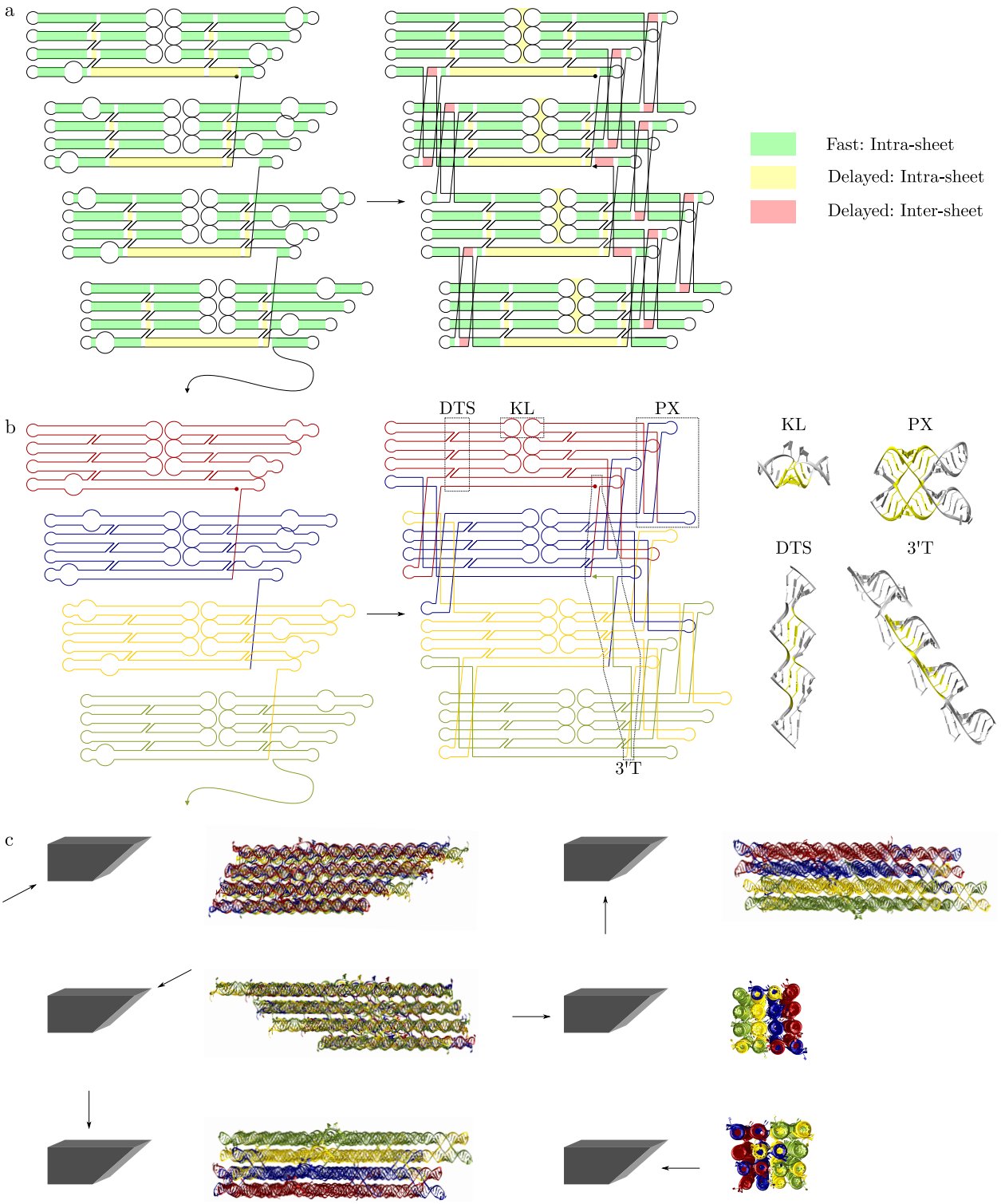


Figure 5.6: Overview of origami$_{PX}$ schematic design and atomic model. (a) Schematic of origami structure before and after formation of tertiary contacts. Base pairs are coloured according to when they form: quick-forming ones (i.e hairpin base pairs) are shown in green, base pairs which form at a delay but within a sheet (kissing loops, dovetail seam and long helical domain) are shown in yellow, and intra-sheet base pairs are shown in red (PX motifs and 3′ tail). (b) Schematic of origami$_{PX}$, as in (a), with the four sheets indicated in four different colours. The kissing loop (KL), paranemic crossover (PX), dovetail seam (DTS) and 3′ tail (3′T) motifs are indicated on the folded schematic, and atomic models of these motifs are shown to the right. (c) 6 orthogonal views of the origami$_{PX}$ atomic model. Sheet colours are matched to the colours in (c).

shifted vertically by 22 Å, and then horizontally by 14 Å (equivalent to 5 nucleotides, the separation between crossovers to sheets above and below on the same strand).

3. A PDB file is generated of the matched set of sheets, and crossovers are introduced by renumbering the strand routing in SwissPdb Viewer [165], and re-ligating the strand in MATLAB.

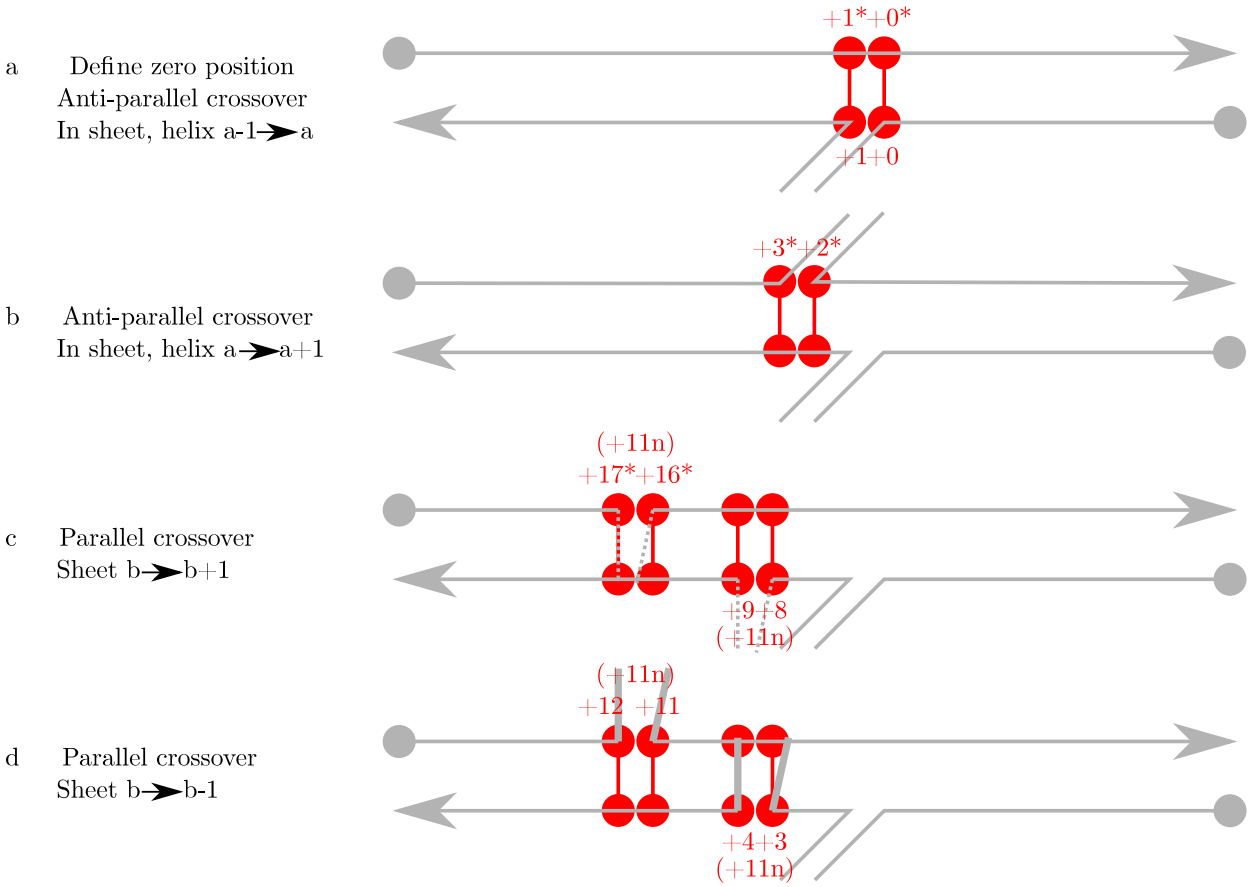


Figure 5.7: Overview of rules for crossover positions. (a) For each helix, the origin of the numbering system is at the in-sheet crossover which links the helix to the previous helix. Crossover positions are specified relative to this crossover (or the theoretical position of this crossover for the first helix of each sheet, where the strand routing begins and ends). Nucleotides on the strand including the crossover are numbered with increasing values in the 5′ direction (and negative values prior to the crossover). The complementary strand is labelled according to its base pairing with the first strand, with an asterisk differentiating the labelling of the two strands. (b) Nucleotide positions for crossovers between a helix and the subsequent helix in the same sheet. (c) Nucleotide positions for crossovers between a helix and the equivalent helix in the next sheet. (d) Nucleotide positions for crossovers between a helix and the equivalent helix in the previous sheet. n indicates any positive or negative integer (11n is a whole turn). Crossover positions may be shifted for the specified number of nucleotides. Positions in (c) and (d) apply for both paranemic crossovers and strand routing. For paranemic crossovers there exist two double crossovers 8 nucleotides apart, with no other crossovers between them.

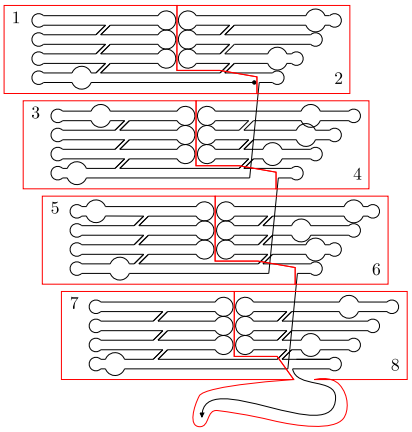
Six orthogonal views of the full atomic model are shown in Figure 5.6c, with sheets

coloured in the same way as the schematic in Figure 5.6b. The overall architecture is reminiscent of a square-lattice DNA origami.

Sequence Design

The sequence design steps are similar to those for the pentomino nanostructure. In this case the structure is sub-divided into 8 half-sheets which are designed independently. A brief overview of steps, followed by a more detailed discussion is outlined below.

1. The 3D structure is converted to a secondary structure using the RNApdbee web server [168, 169] and subdivided into 8 half-sheets (see right). Half sheets are designed according to the secondary structure (i.e. before the formation of PX motifs), and the 3′ tail is included in the design of the 8th half-sheet.
2. Half-sheet sequences are designed in NUPACK [167] with appropriate sequence constraints. Sequences are evaluated by mfold and optimised by hand.
3. Flanking sequences for transcription and cloning are introduced for the gblock template.



Sequences for PX and KL motifs must be specified in the sequence design stage. Four different 180° kissing loops sequences were used in this design, shown in Table 5.2. Kissing loop 7' has been used previously by Severcan et al. [108], although the affinity of the pair has not been measured.

Kissing Loop	Loop A sequence	Loop B sequence	K_D/nM
1'	*aaGGAGGCa*	*aaGCCUCCa*	5.6
2'	*aaGCGAGCa*	*aaGCUCGCa*	12.6
3'	*aaGUCCACa*	*aaGUGGACa*	15.8
7'	*aaAGGCUCa*	*aaGAGCCUa*	?

Table 5.2: Table of 180° kissing loop sequences used in origami$_{PX}$ and their measured affinities where available. Loop A and B correspond to the two hairpin loops comprising the motif. Lower case 'a' nucleotides exist in the loop but are not involved in pairing.

PX motif sequences are also assigned (sequences provided in Tabel 5.1). For both the KL and PX motifs, the number of unique sequences used is lower than the total number of motifs. The positions of different sequences are assigned in an attempt to minimise the formation of unintended matches. Assigned sequences for PX and KL motifs are shown in Figure 5.5a and b respectively.

In addition to the PX and KL sequences, the T7 starter sequence is included at the 5′ end of the first half-sheet, nucleotides flanking crossovers (intra-sheet antiparallel and inter-sheet parallel) are specified to prevent branch migration, restriction enzyme sites were

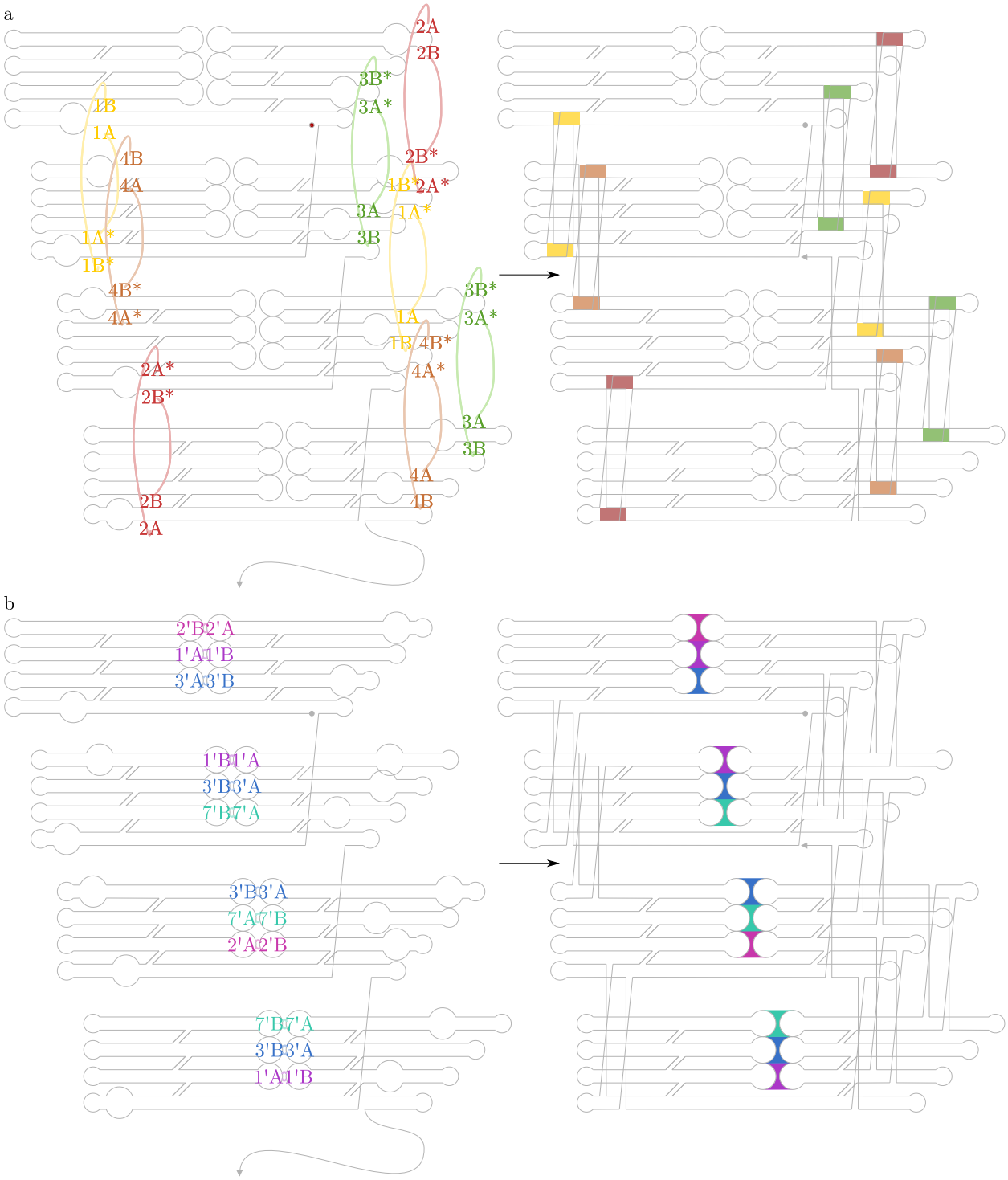


Figure 5.8: Schematic of origami$_{PX}$, with the identities of (a) paranemic crossover motifs and (b) 180° kissing loops labelled. Motifs of the same colour are equivalent. In (a) The left schematic shows the structure before formation of paranemic crossovers or the 3′ tail, wheras the right schematic shows the full folded structure. In (b) the fully-folded schematic is shown on both the left and right.

introduced at the 3′ ends of each sheet, and for sheet 8 (which was designed last) the 3′T sequence is specified according to the designed single-stranded domains in half-sheets 4 and 6. As described in Section 4.3.1 preliminary sequences are designed in NUPACK [167], and

sequences are modified iteratively through mfold [171] analysis, during which $G \cdot U$ wobble pairs are introduced. Mfold predictions for the MFE secondary structures for each of the 8 consitutent half-sheets are shown in Figure 5.9.

For these predicted secondary structures, it is worth noting that, although we can design half-sheets with no unintended base-pairs, some elements of the secondary structure remain single-stranded. These are the two-base-pair domains along the dovetail seam and a short hairpin domain which should be present in each of the odd-numbered half-sheets. The two-base-pair domains are not anticipated to cause problems, since these domains are present in Geary's tiles [114] which are known to fold as designed. It is expected that the formation of tertiary structure (internal kissing loops) will promote the 'clicking together' of the dovetail seam which may otherwise not be energetically favourable. The same is true for the short hairpin domains since formation of the kissing loop interaction will bring the two hairpin strands in closer proximity.

For subsets of the structure larger than a half-sheet, mfold analysis reveals that there are often a number of possible folds with similar energy, with the most stable structure not always corresponding exactly to the designed one. However, since the structure is designed for co-transcriptional folding, this is not necessarily a problem, since individual half-sheets are at a thermodynamic minimum in their designed fold and should become kinetically trapped in this fold as they are transcribed. Additionally, tertiary contacts between half-sheets may start to form during transcription which should stabilise the correct folding.

Following sequence design, the RNA sequence was converted to DNA, the consensus T7 RNA polymerase promoter included at the 5′ end of the sequence, and restriction sites introduced for cloning into pUC19. The DNA template was ordered as a double-stranded gBlock (sequence provided in Appendix A.1.2).[4]

5.3.2 Synthesis and Characterisation

Cloning

The origami$_{PX}$ gblock template was digested with EcoRI, followed by 5′ phosphorylation with T4 PNK and pUC19 was digested simultaneously with EcoRI and HincII in 1×NEBuffer 2.1 followed by dephosphorylation by antarctic phosphatase.[5] Ligation and cloning were performed as described in Appendix B.1.2. After screening a number of colonies of the correct size, there were found to be three consistent point mutations. These were corrected through three successive rounds of site-directed mutagenesis (Appendix B.1.3)[6], giving a plasmid template of the correct sequence.

Linear transcription templates for subsets of the nanostructure were prepared by restriction enzyme digests. Restriction sites were included at the 3′ terminus of each sheet so structural subsets of 1, 2, 3 and 4 sheets were prepared by NdeI, NheI, BglII and EcoRI

[4]Ordered from Integrated DNA Technologies as a gblock gene fragment (1000 ng).
[5]All enzymes were purchase from New England Biolabs.
[6]Mutagenesis performed with the Q5 site-directed mutagenesis kit from NEB.

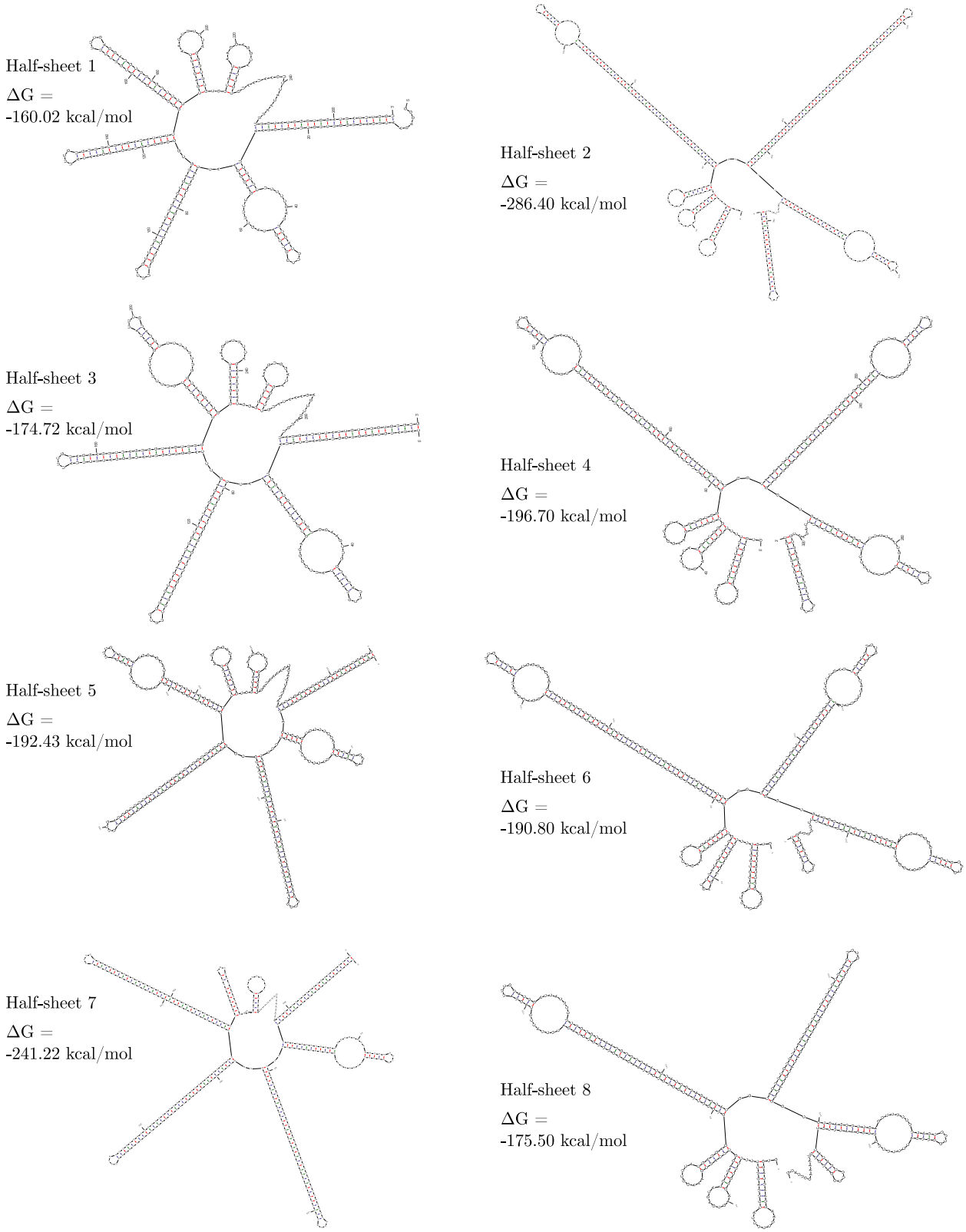


Figure 5.9: Mfold predictions of secondary structures for half-sheets 1-8 of the origami$_{PX}$ nanostructure.

digestion respectively. Linear templates were purified prior to transcription.[7]

Transcription and Gel Characterisation

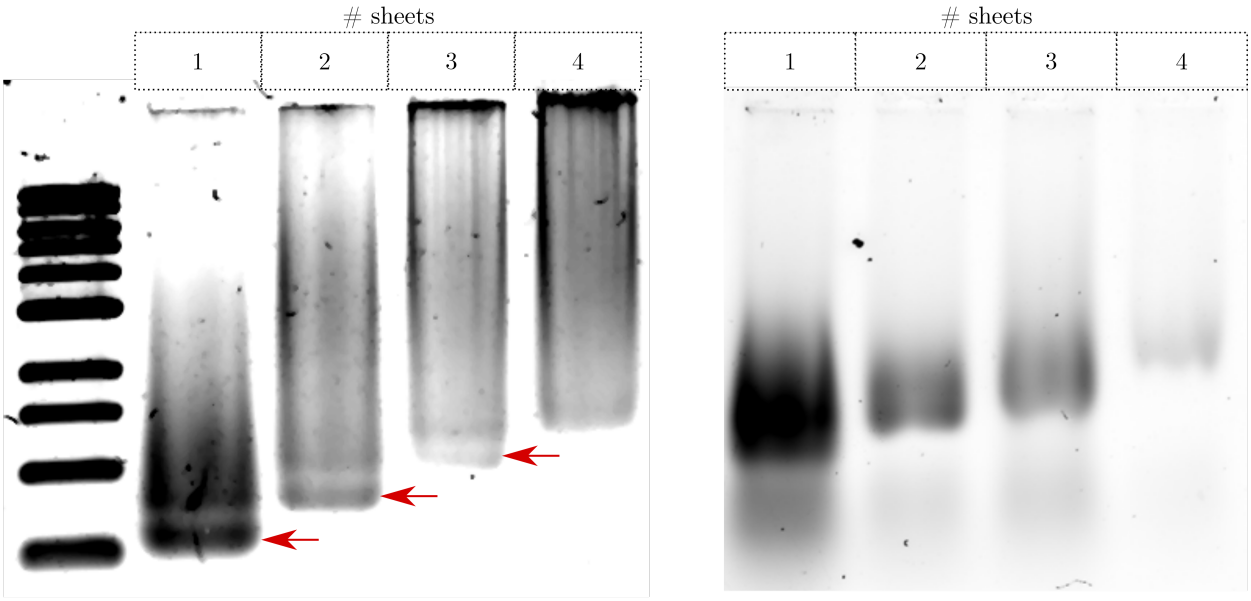
Subsets of the structure were prepared by transcription from templates (Appendix B.2.2), and analysed by AGE (Appendix E.1). Results are shown in Figure 5.10a. A large amount of smearing is seen for all subsets, which is usually indicative of unfolded, mutually-interacting RNA molecules. For smaller subsets of the nanostructure there appears to be a high mobility band which likely corresponds to the well-folded nanostructure/nanostructure-subset (indicated by red arrows in Figure 5.10b) and a smear of lower-mobility material(which likely corresponds to mis-folded strands). Relative amounts of folded and unfolded material can be estimated crudely from gel contrast.

1. Bands are selected and quantified using the 'measure' tool in imagej.
2. A 'background' region of the gel is selected and quantified in the same way.
3. The amount of RNA in each band is quantified by subtracting the mean pixel value from the background and multiplying by the band area.
4. Entire lanes are selected and quantified in the same way.
5. The ratio of the final value found for a band and the full lane gives an estimate of the fraction of folded RNA.

By this method, the fraction of folded RNA was estimated as ≈14, 9, 6 and 0% for 1, 2, 3 and 4 sheet structures respectively. The estimated folded fraction of even the 1 sheet structure is very low, which implies that folding problems are not solely due to difficulties in forming PX motifs.

After the report of 2D single-stranded RNA origami folded by thermal annealing [110], I investigated the effect of thermal annealing on the origami$_{PX}$ structure. The full nanostructure, as well as the 1, 2 and 3-sheet subsets of the structure were prepared by transcription, and purified using Amicon spin filters.[8] During purification, samples were also buffer-exchanged into 1×TAE + 12.5 mM $MgCl_2$, as used by Han et al. for thermal annealing [110]. Samples were then annealed using both short (2 hour) and long (18 hour) annealing ramps [110]. AGE results from the short annealing ramp are shown in Figure 5.10b. The high mobility smear, and in-well aggregation, are absent in annealed samples, but instead of clean bands a short, diffuse smear is seen, suggesting that structures do not fold well by thermal annealing. A high mobility smear is also seen, which appears to be due to RNA hydrolysis. For the long-annealing samples, only this high mobility smear is present (i.e. complete hydrolysis). The lack of well-folded structures by thermal annealing is expected, since mfold analysis shows that the designed secondary structure (in the absence of tertiary contacts) is not the thermodynamic minimum for the full structure.

[7] Template purification using the QIAquick PCR purification kit.
[8] 3×10 minutes spinds, with 100 kDa Amicon filters.



Co-transcriptional Folding Thermal Annealing

Figure 5.10: Agarose gel electrophoresis of subsets of origami$_{PX}$ structure folded (a) co-transcriptionally or (b) by thermal annealing. Red arrows indicate high mobility bands which may contain well-folded structures.

Negative Stain TEM

Origami$_{PX}$ folding was also investigated by negative stain TEM. Samples of the origami$_{PX}$ nanostructure were prepared by transcription and prepared for TEM by column purification (Appendix C.2.4). Grids were prepared for TEM (Appendix D.1.1), and characteristic micrographs are shown in Figure 5.11a and b. Results are consistent with AGE analysis: structures are evidently not folded into the designed 3D structure (no convincingly well-folded structures were seen across the whole data set). Short helical domains are seen, but these are not arranged into a well-ordered or consistent structure.

5.3.3 Discussion

Preliminary results for the PX-based origami architecture demonstrate that these structures are not capable of forming by co-transcriptional folding or thermal annealing. There are a number of possible factors which may contribute to the observed misfolding:

1. The PX and KL motifs are degenerate and may form incorrect interaction pairs.
2. Formation of PX motifs, although paranemic, requires helices to wrap around one another, which may be prevented by the surrounding structure.
3. The secondary structure (prior to formation of PX and KL contacts) is not a strong thermodynamic minimum and initial mis-folding of the secondary structure may block the formation of tertiary interactions.

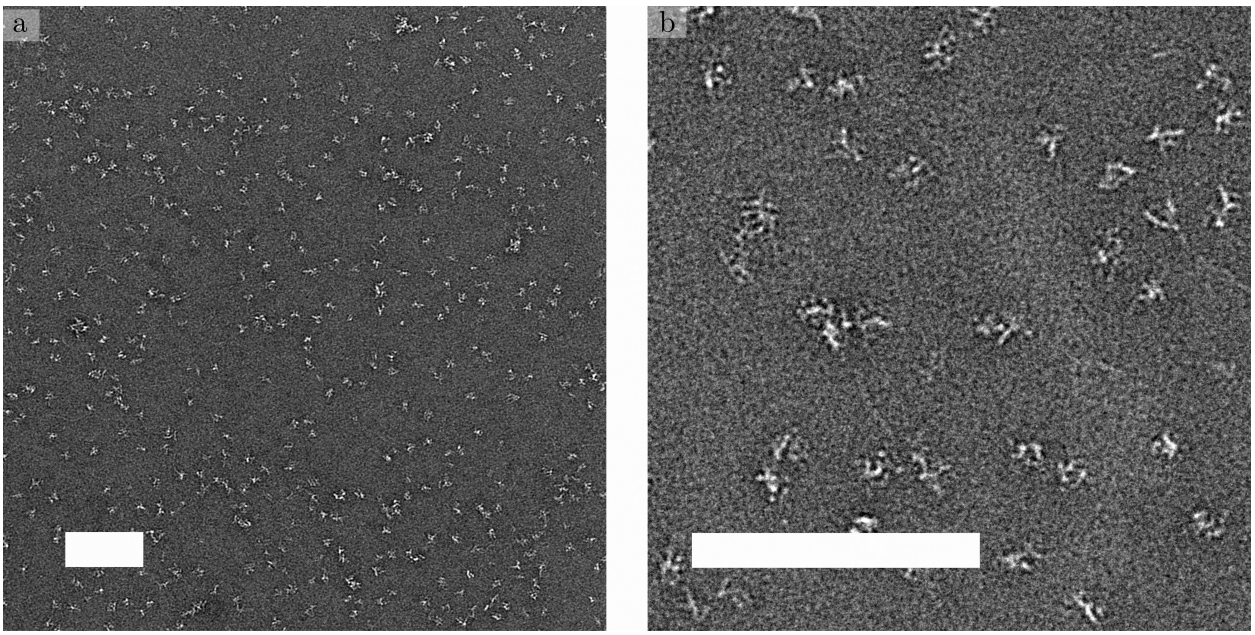


Figure 5.11: Characteristic (a) wide-field and (b) high-magnification views of origami$_{PX}$ strands analyse by negative stain TEM. Scale bar 200 nm

4. The 8 nucleotide bulge loops may sterically block the intending helix stacking behaviour.

It is notable that a subset of the structure comprising only the first sheet does not fold at high yield (Figure 5.10). This mis-folding cannot be explained by points 1 and 2 above, suggesting that one or both of points 3 and 4 (both direct consequences of the 8 nucleotide bulge loops) must contribute to mis-folding. There is no clear way to overcome either of these potential problems, so based on these results I decided to investigate BLKL motifs as an alternative to PX motifs for building origami in 3D.

5.4 3D Origami: Building with BLKL motifs

5.4.1 Concept

The concept of my proposed origami$_{\mathrm{BLKL}}$ architecture is once again based on a set of mutually-interacting sheets, with each sheet resembling a 2D RNA tile. In this case, pairs of equivalent terminal helices on successive sheets are replaced with BLKL motifs, which is intended to force sheets into a compact, parallel conformation. Like the PX$_{\mathrm{major}}$ motif, the BLKL motif is paranemic, but unlike the PX$_{\mathrm{major}}$ motif, the BLKL motif does not require contributing helices to wrap around one another. It is anticipated that two sheets can slot together with minimal deformation, i.e. the two parts of the motif should possess broadly the same geometry before and after formation of the motif. This also means that the half-formed motif should not affect the stacking behaviour of the other helices within the sheet.

5.4.2 2-Sheet Dimer

Design

To investigate BLKL motifs as route to building in 3D, I first designed a simple 2-sheet hetero-dimer to test the proposed interaction interface. The BLKL motif consists of two parts: the 180° bulge loop capped with a hairpin loop, and the complementary hairpin. These two parts will be referred to as $\text{BLKL}_{\text{turn+loop}}$ and $\text{BLKL}_{\text{loop}}$ respectively. By replacing hairpin domains in one Geary tile with a $\text{BLKL}_{\text{turn+loop}}$ motif, and with a $\text{BLKL}_{\text{turn}}$ in compatible positions on a second tile, sheets can be designed to come together. For the two-sheet hetero-dimer, four BLKL motifs are used to hold sheet 1 (S1) and sheet 2 (S2) together. If all four interactions form as intended, this should result in a rigid interaction interface, whilst there remain four free hairpins on each sheet for modification which could be used to build larger structures. Both the conceptual design, and the atomic models (which I will address next) are shown in Figure 5.12.

An atomistic model of the structure was generate in Chimera [124] using the method described previously, outlined more specifically below.

1. Two-tile PDB files are loaded into Chimera, and aligned as described previously. For this two-strand hetero-dimer there will be no scaffold-linkages between sheets, but arranging sheets to permit inter-sheet crossovers will provide more options for single-stranded designs.

2. The aligned tiles are combined into a single PDB file using the 'combine' tool in Chimera.

3. Four BLKL motif PDB files (Section 5.2.2) are loaded into Chimera. BLKL motifs are positioned onto the 2-tile PDB by matching positions of phosphate atoms on the two helical domains of the BLKL motifs with phosphate atoms on the two corresponding helical domains on the two tiles. This 'matching' is described by Geary et al. [115]. The optimal position was found by minimising the RMSD between matched atom pairs. The mismatched is minimised when the $\text{BLKL}_{\text{loop}}$ motif is on S1 and the $\text{BLKL}_{\text{turn+loop}}$ on S2 for the 5′ half-sheet, and vice versa in the 3′ half-sheet.

4. The BLKL positioning rules are shown in Figure 5.13

Sequences of S1 and S2 were designed from the secondary structure. Only certain KL sequences can be used within the BLKL motifs, as many KL sequences interact with the BL sequence resulting in mis-folded $\text{BLKL}_{\text{turn+loop}}$ domains. The 'design' module of NUPACK [167] was used to assess which sequences can be placed within the $\text{BLKL}_{\text{turn+loop}}$ domain.

1. The secondary structure for a $\text{BLKL}_{\text{turn+loop}}$ domain within the $\text{dimer}_{\text{BLKL}}$ framework was isolated. In dot/bracket notation this is:
((((((((((((((((.....)))))).....))))))))))))

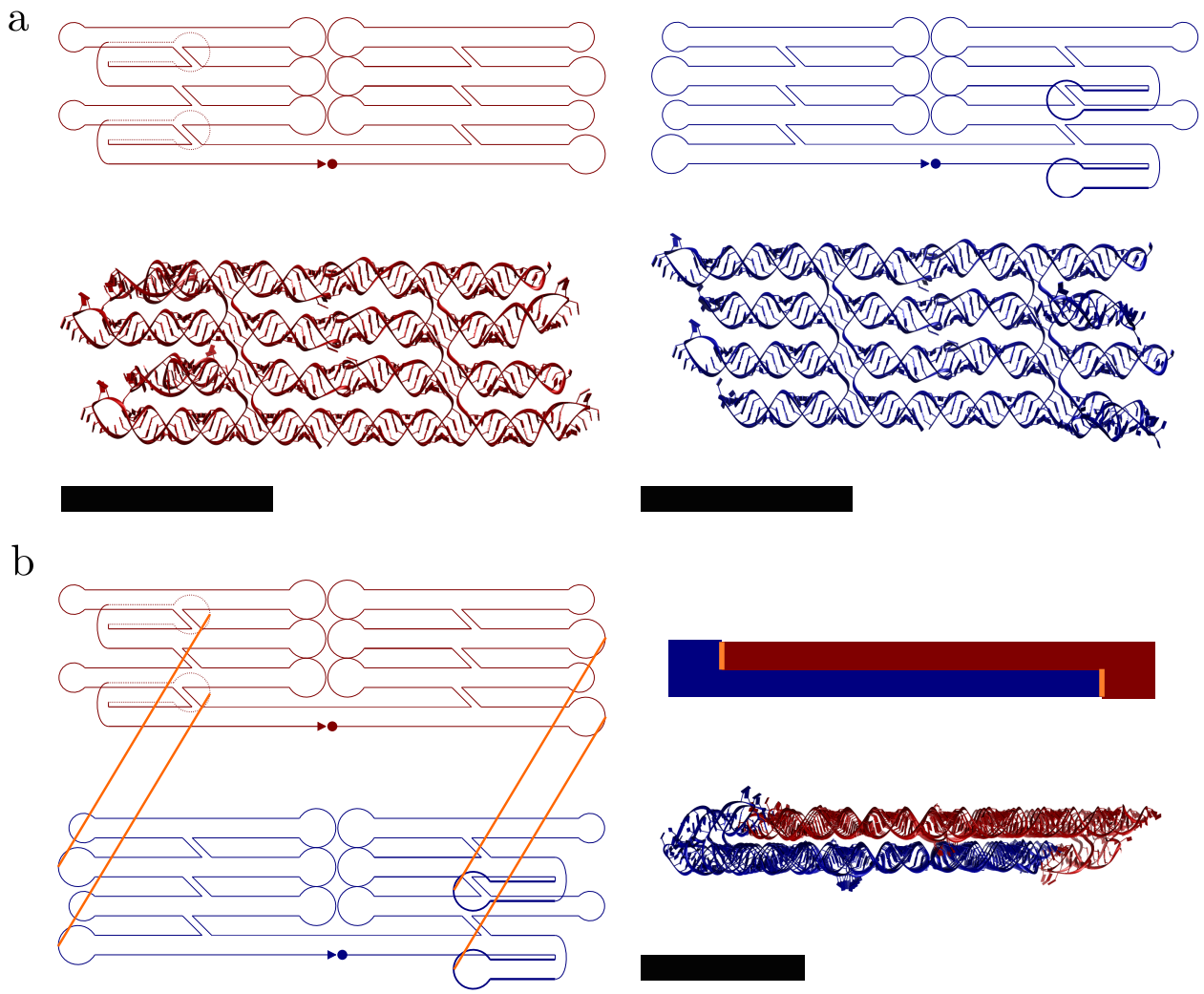


Figure 5.12: Overview of design for an origami$_{BLKL}$ hetero-dimer. (a) Schematic (above) and atomic model (below) for sheets 1 (left, depicted in red) and 2 (right, depicted in blue) for dimerisation. (b) Schematic (left) and atomic model/block view (right) of origami$_{BLKL}$ hetero-dimer. The kissing loop portion of the BLKL motif is shown in orange. Scalebar 10 nm.

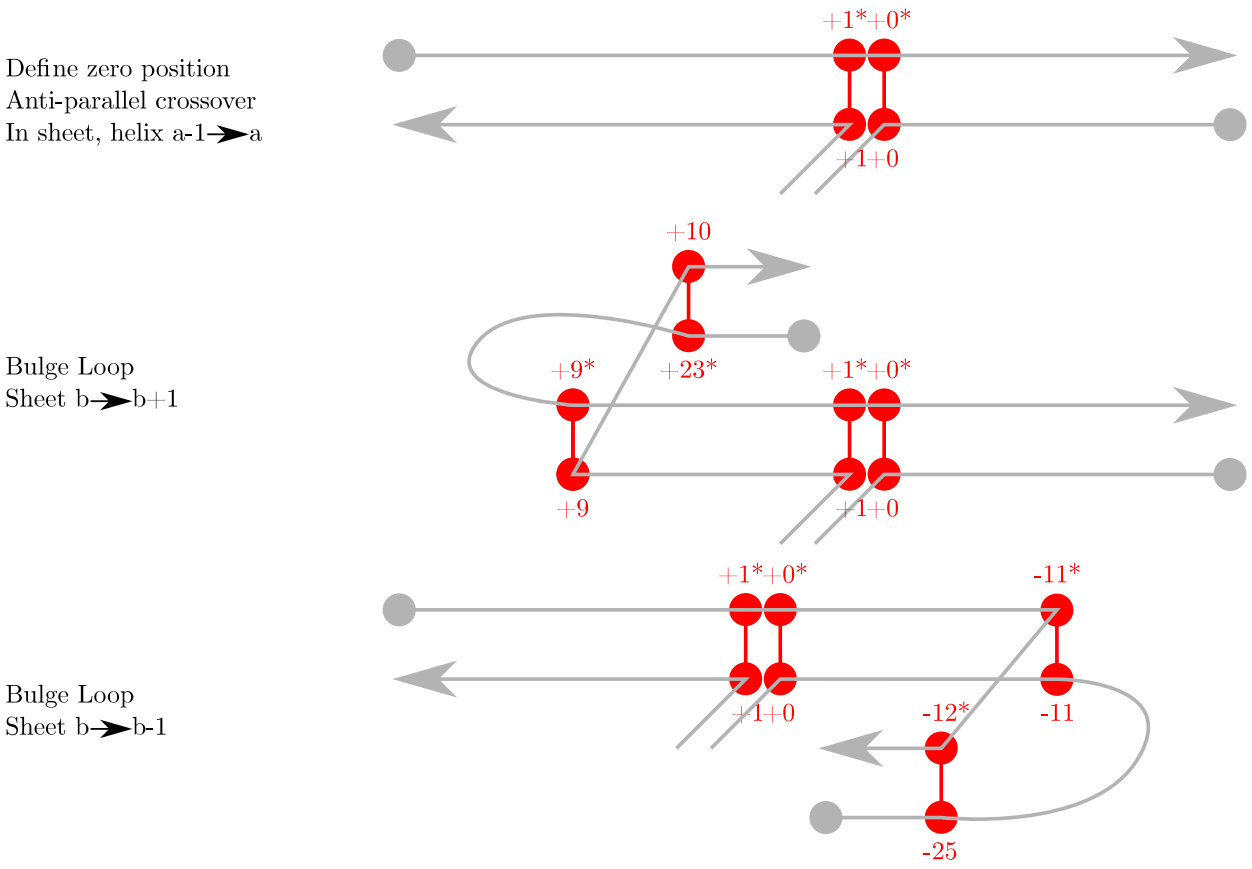


Figure 5.13: Overview of rules for BLKL positions. (a) Origin specification (see Figure 5.7a). (b) Nucleotide positions for bulge loop of BLKL motif linking to the equivalent helix in the subsequent sheet. (c) Nucleotide positions for bulge loop of BLKL motif linking to the equivalent helix in the previous sheet.

2. Sequences are specified for the bulge loop (the same in each case) and for the kissing hairpin (sequences shown in Table 5.4.2). The 'design' module of NUPACK [167] was then used to generate a sequence with the specified secondary structure and sequence constraints. 10 design trials were performed. The loop sequences analysed include the 6 pairs given in Table 4.15, as well as additional pairs (sequences 7' and 8' have been used by previous authors [108, 114] and sequences 9'-13' were designed according to the method outlined in Section 4.4.1,

3. The best (lowest) normalised ensemble defect (NED)[9] was recorded for each sequence. These are shown in Table 5.4.2 below. Only sequences with a 'best NED' below 3% have been use within $\text{BLKL}_{\text{turn+loop}}$ domains.

Loop Name	Loop Sequence	Best NED % (10 trials)	Loop Name	Loop Sequence	Best NED %) (10 trials)
1'A	*aaGGAGGCa*	10.4	7'B	*aaGAGCCUa*	5.4
1'B	*aaGCCUCCa*	1.3	8'A	*aaCGUGGUa*	8.8
2'A	*aaGCGAGCa*	8.2	8'B	*aaACCACGa*	2.7
2'B	*aaGCUCGCa*	4.0	9'A	*aaCGAAACa*	0.7
3'A	*aaGUCCACa*	0.6	9'B	*aaGUUUCGa*	20.0
3'B	*aaGUGGACa*	17.8	10'A	*aaCACUACa*	0.6
4'A	*aaGCAGGCa*	4.0	10'B	*aaGUAGUGa*	9.9
4'B	*aaGCCUGCa*	7.3	11'A	*aaGAUAGGa*	10.0
5'A	*aaCUUCGCa*	4.4	11'B	*aaCCUAUCa*	0.6
5'B	*aaGCGAAGa*	5.6	12'A	*aaGAACGCa*	9.9
6'A	*aaGUGCUGa*	12.0	12'B	*aaGCGUUCa*	4.7
6'B	*aaCAGCACa*	2.2	13'A	*aaGUCACCa*	0.8
7'A	*aaAGGCUCa*	2.7	13'B	*aaGGUGACa*	5.4

Table 5.3: Table of the best NED for 10 design trials of $BLKL_{turn+loop}$ domain for different hairpin loop sequences (see text above for full explanation).

A consistent set of kissing loop sequences were assigned to the different loops on S1 and S2, as indicated in Figure 5.14. Sequences were designed in NUPACK [167] as described previously, with appropriate sequence constraints: the dovetail seam, the T7 starter sequence, kissing loop sequences and bulge loop sequences were all specified. For the $\text{origami}_{\text{PX}}$ nanostructure, the design was split into half-sheet domains with a view to co-transcriptional folding. As a result, full sheets in the nanostructure did not have a strong thermodynamic minimum for the designed fold, which may be a contributing factor to the nanostructure mis-folding. For this design, the full sheets were designed in one go rather than splitting into smaller domains.

As described previously, NUPACK sequences were selected and optimised iteratively in Mfold [171]. Mfold predictions of the secondary structures for the two sheets are shown in Figure 5.15. The structures are predicted to form as designed with the exception of the dovetail seam, as discussed for $\text{origami}_{\text{PX}}$ in Section 5.3.1.

[9]NED: the average percentage of bases that are incorrectly paired at equilibrium

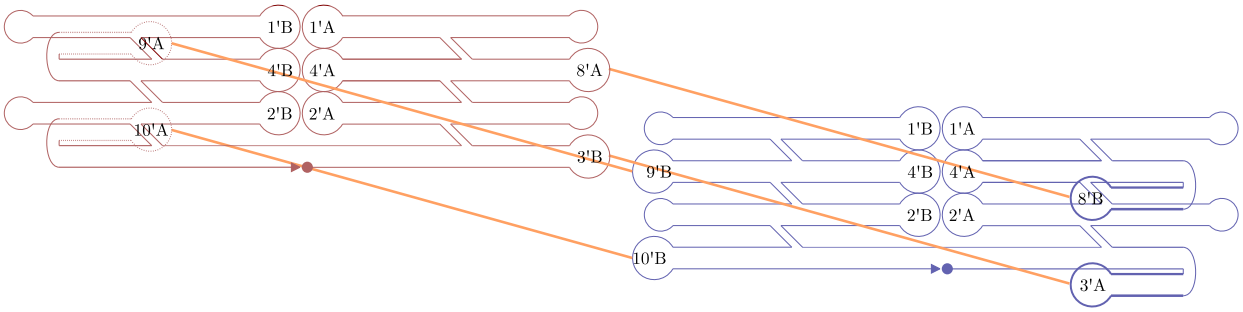


Figure 5.14: Kissing loop identities for internal and external ($BLKL_{turn+loop}$) loops for $origami_{BLKL}$ sheets 1 and 2.

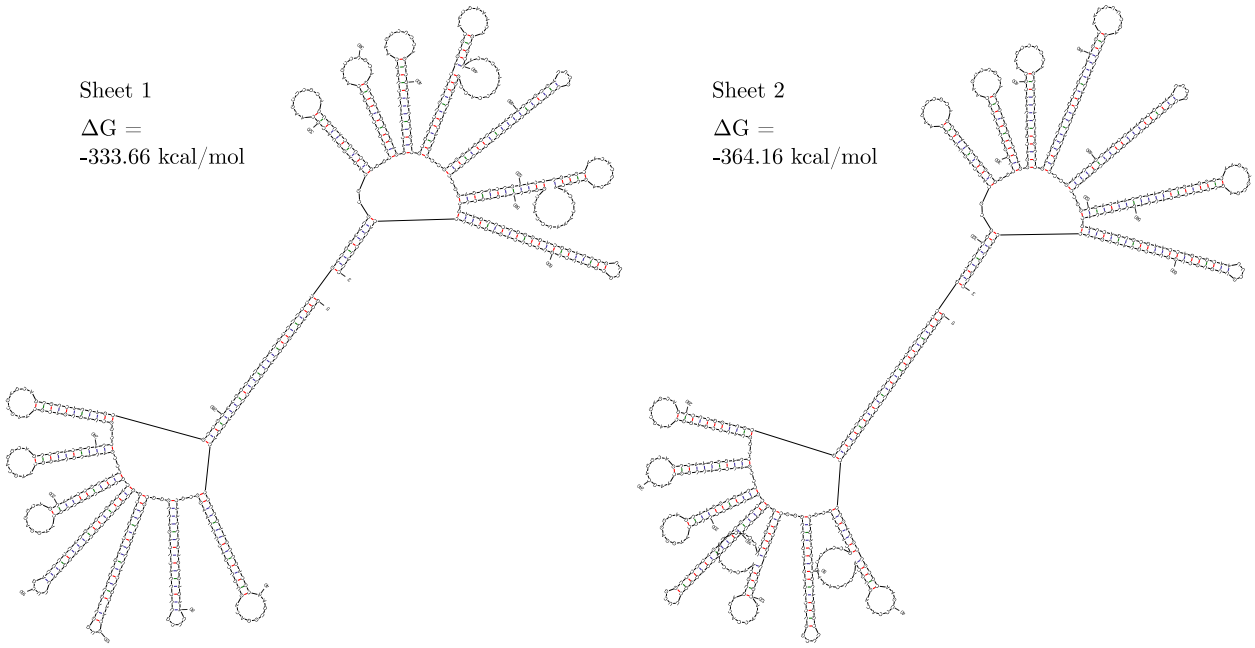


Figure 5.15: Mfold predictions of secondary structure for sheets 1 and 2 of $origami_{BLKL}$ design.

Synthesis and Characterisation

Template duplexes for the two sheets were generated by PCR from a single gblock template (gblock $origami_{BLKL}$-ssl sequence provided in Appendix A.1.2 and PCR details provided in Appendix B.1.3). Nanostructures were prepared (Appendix B.2.2) and analysed by AGE (Figure 5.16a). Each sheet is visible as a single, distinct band which is usually indicative of a consistent fold. The mobilities of the sheets are between that of the 400 and 500 bp DNA duplex bands. Comparing these structures to a sheet 1 of the $origami_{PX}$ nanostructure (Figure 5.10, lane 1 in each gel), a significant improvement of folding is seen, suggesting that the BLKL motif is less disruptive to co-transcriptional folding than the PX_{major} motif.

The nanostructures were also imaged by TEM. Nanostructures were prepared by transcription, purified by LiCl precipitation (Appendix B.3.2), and resuspended in PBS-Mg to a final concentration of 10 nM. Single-helix-thick structures are notoriously difficult to visualise by negative-stain TEM, so a positive staining protocol was used for these samples

(Appendix D.1.2). Example micrographs are shown in Figure 5.16b and c for sheets 1 and 2 respectively. Rectangular structures with dimensions of $\approx$ 10 nm $\times$ 20 nm can be observed, consistent with the designed sheets.

Finally the interaction between the sheets was analysed by AGE. Both sheets were transcribed, and a 1:1 mix of the two samples was prepared. This is shown in Figure 5.16d. An upward smear of lower-mobility RNA (starting at the S1/S2 band position) suggests and interaction between the two sheets, but the lack of a distinct band may indicate that the interaction interface is too weak, or not well defined.

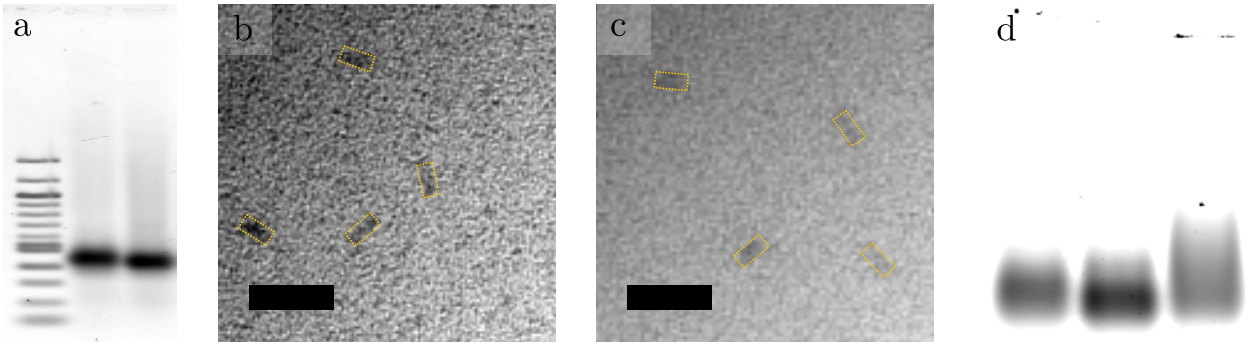


Figure 5.16: Characterisation of S1 and S2 of origami$_{BLKL}$ dimer. (a) AGE of S1 and S2 next to a 100 bp ladder (NEB). Both sheets have a mobility in between the 400 bp and 500 bp references. (b)/(c) Positive stain TEM of S1/S2 nanostructures. Four nanostructures can be seen in each field of view, indicated by yellow outlines. Scale bars 50 nm. (d) AGE of sheets 1 and 2 (lanes 1 and 2) and a 1:1 mix of the sheets (lane 3). An upward smear in mobility suggests interaction between sheets 1 and 2, but the lack of a distinct band may indicate that the interaction interface is not well defined.

Discussion

Analysis by AGE and positive-stain TEM suggest that the dimer$_{\mathrm{BLKL}}$ sheets fold into the designed structures at much higher yields than origami$_{\mathrm{PX}}$ sheets. AGE results from mixing the two sheets show that the two sheets do interact, but do not seem to form rigid/stable dimers. Based on this mixed success, I proceeded with designing single-stranded dimer nanostructures.

By introducing a parallel double-crossover between the two sheets, we can design scaffold-linked dimers. The scaffold link will reduce the range of geometries between sheets, and increase the local concentration of BLKL$_{\mathrm{turn+loop}}$ and BLKL$_{\mathrm{loop}}$ pairs, which may promote the formation of the intended interaction interface.

5.4.3 2-sheet Nanostructure

Design

Three variants of scaffold-linked origami$_{\mathrm{BLKL}}$ can be designed, depending on where the 5′/3′ nick is positioned relative to the S1/S2 double crossover. The rule for positioning inter-sheet crossovers is shown in Figure 5.7. The 5′/3′ nick can be positioned on S1 (5′ tail) on S2 (3′ tail) or at the crossover itself (no tail). Schematic and atomistic views of these three variants are shown in Figure 5.17. Here 'tail' refers to the domain between the S1/S2 double crossover and the 5′/3′ nick. Note that in the 2-sheet nanostructure, S2 (as defined previously) is actually transcribed before S1.

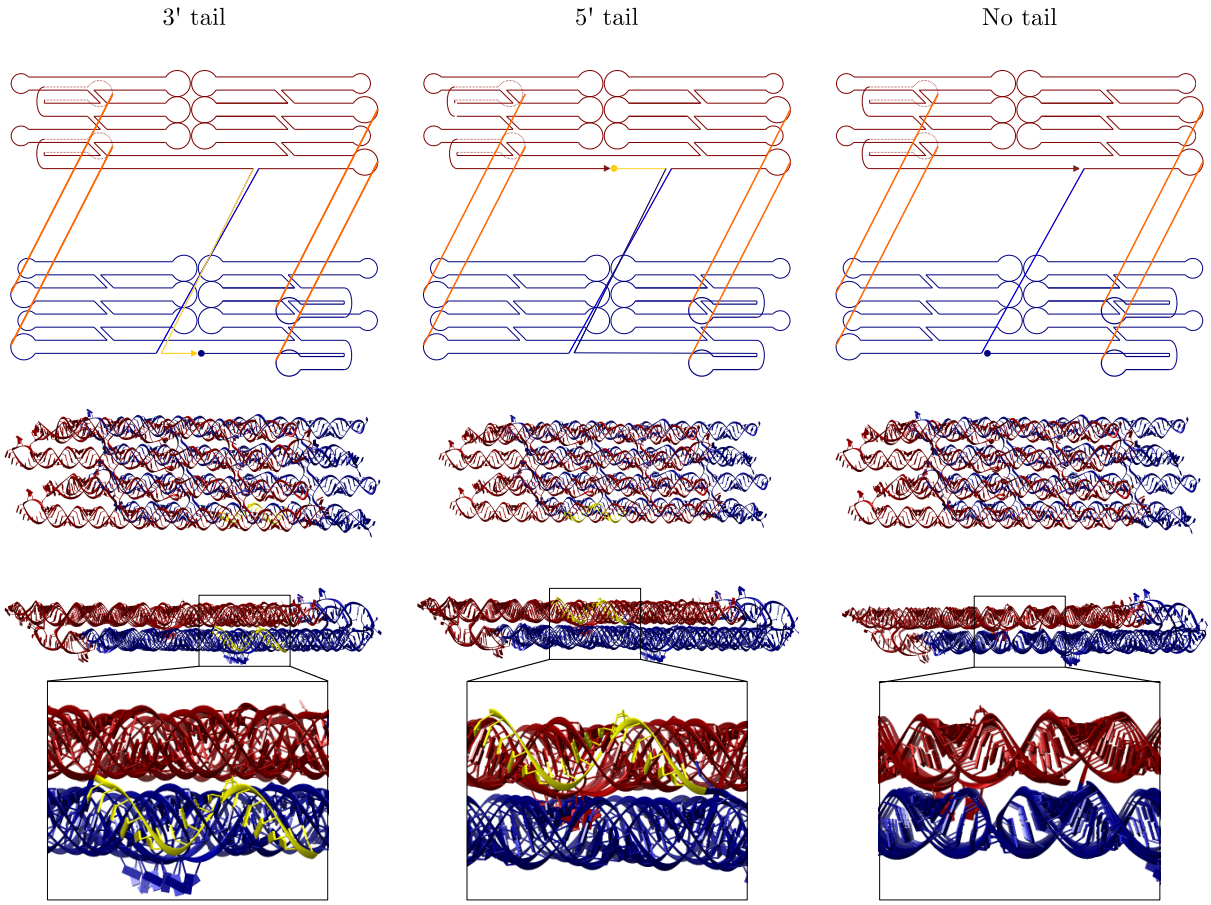


Figure 5.17: Schematic (above) and atomistic (below) views of the three designed scaffold-linked dimers. The 'tail' domain between the S1/S2 double crossover and the 5′/3′ nick is shown in yellow. Insets show magnified views of the tail.

Each variant bears potential advantages and disadvantages. At first glance, the 3′ tail would seem to be the most problematic, since the duplex between the 3′ tail and its complementary sequence on S2 forms last during transcription. The base-pairing will require sheets to fold around one another, which is sterically difficult. This domain is comparable to the 3′ helical domain in the original 2D tile architecture (see discussion in Section 5.2.1), so may not prohibit folding. For the 5′ tail design, the tail is transcribed first. The complementary sequence on S1 is transcribed later during the folding process and

the duplex may form as the complement is transcribed. Because the 5′ tail remains present and single-stranded throughout the transcription process it may transiently interact with the RNA strand as it is transcribed, but it is not clear whether this would help or hinder the folding process. For both the 3′-tail and 5′-tail variants, the 5′/3′ nick is positioned in an appropriate location to introduce a parallel crossover to a third sheet, and so the architectures could be extended to larger nanostructures. Finally if we introduce the 5′/3′ nick exactly at the crossover, we can design a dimer which has no tail. This should fold more easily, but has a significant downside, as these scaffold connections cannot be extended to more than 2 sheets.

I have proceeded with designing and transcribing all three dimer types (named origami$_{\text{BLKL}}$-3′, origami$_{\text{BLKL}}$-5′ and origami$_{\text{BLKL}}$-N for the 3′-tail, 5′-tail and no-tail variants respectively). Atomic models were generated from the 2-sheet atomic model, by identifying the appropriate 5′/3′ nick and parallel crossover positions, renumbering the strand routing accordingly (SwissPdb Viewer [165]) and ligating the strand. Secondary structures were derived from the atomic model and checked against the expected structure.

For the 2-sheet design, the same internal kissing loops were used on each sheet. For single-stranded folding this could be problematic (possibility of mismatches), so a different set of kissing loops were assigned (Figure 5.18). Instead of designing new sequences from scratch, the S1 and S2 sequences, matched onto the new scaffold routings, were used as a starting point. Loop sequences were changed, the parallel double crossover sequences were specified to promote the correct stacking confirmation, and the *GGGAGA* starter sequence and its complement were included. Following these changes, sequences were optimised by iterative mFold analysis of the full 2-sheet structures. Mfold secondary structure predictions are shown in Figure 4.7 and gblock sequences are provided in Appendix A.1.2.

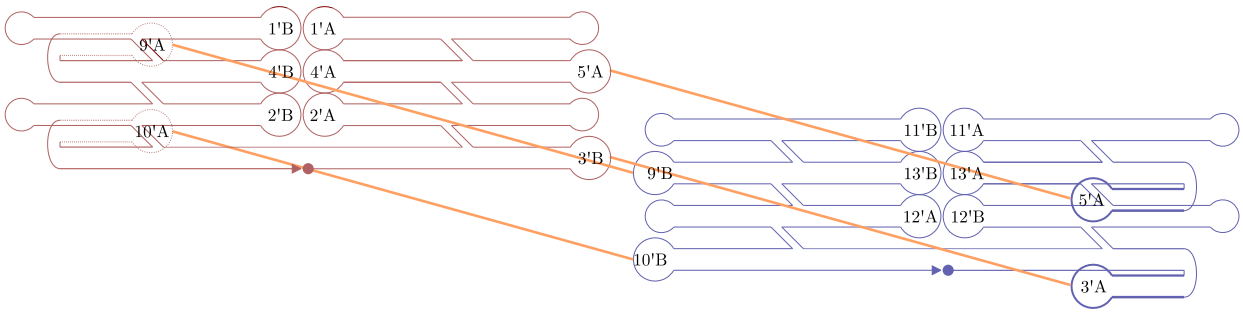


Figure 5.18: Kissing loop identities for internal and external ($BLKL_{turn+loop}$) loops for origami$_{BLKL}$-3′, -5′ and -N variants.

Synthesis and Characterisation

Templates for all three variants were generated by PCR (B.1.3), and nanostructures were prepared by transcription (Appendix B.2.2), purified by LiCl precipitation (Appendix B.3.2) and resuspended in PBS-Mg buffer to a final concentration of 100 nM. A subset of the samples were also prepared by thermal annealing[10], and the nanostructures with and without this

[10]Samples are heated to 95°C and subsequently cooled to 5°C at a rate of 1°C/s

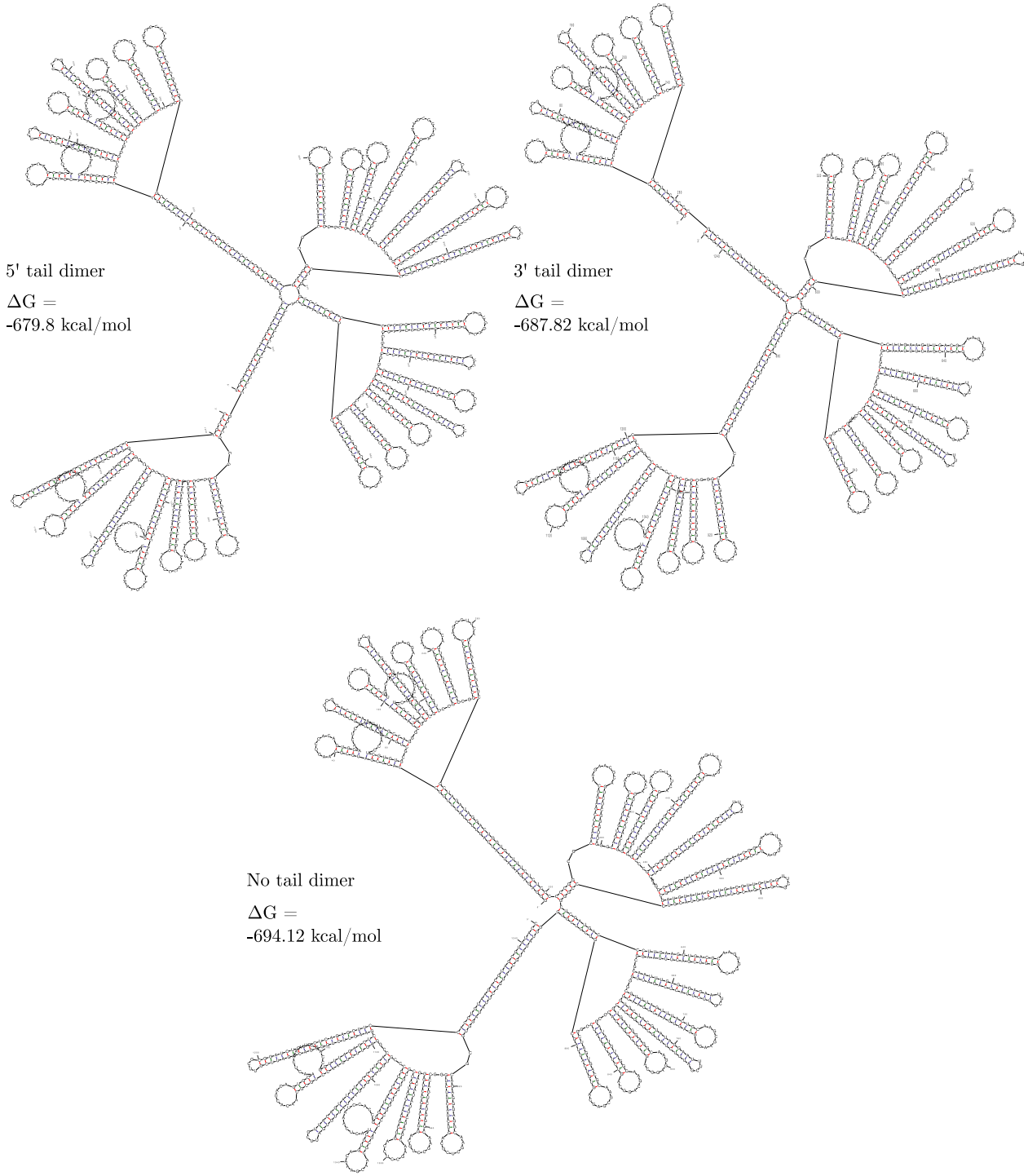


Figure 5.19: Mfold predictions of secondary structure for dimer types (5′ tail, 3′ tail and no tail) of origami$_{BLKL}$ design.

annealing step were analysed by AGE (Figure 5.20).

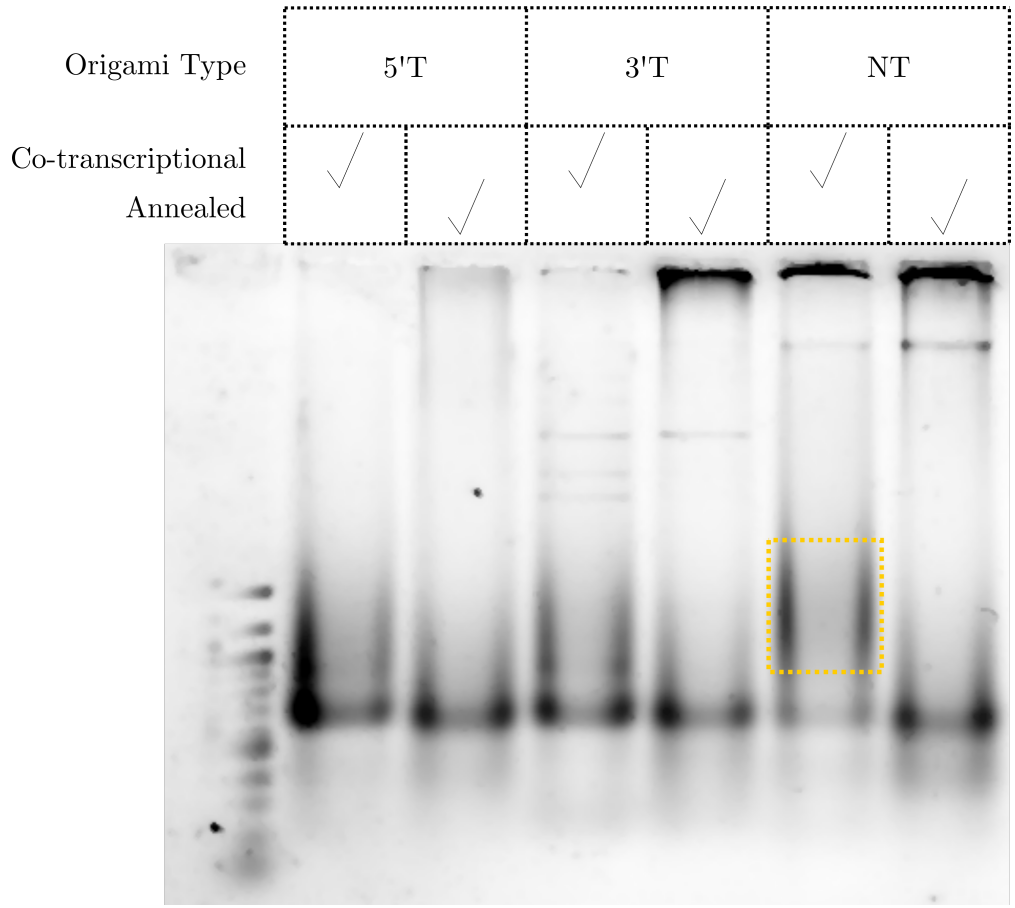


Figure 5.20: Agarose gel electrophoresis of origami$_{BLKL}$ variants generated by co-transcriptional folding and thermal annealing (95-5°C at 1°C/s). The far left lane contains a 100 bp ladder (NEB). The yellow box indicates a diffuse band, suggested to contain multimers of 'open' structures for the NT co-transcriptional variant.

All samples appear reasonably well-folded with a relatively clean primary band. Comparing the various samples we can make some general comments.

- Despite my previous assertion that origami$_{\text{BLKL}}$-N should have the simplest folding pathway, it appears to fold the least well co-transcriptionally (lane 6). The primary band (mobility around 700 bp lane) comprises only ≈15% of material, a diffuse band centred between the 1,200-1,500 bp (yellow box) comprises about 55% of the material and the remaining 30% is present as aggregate in the well. One reasonable explanations for this observation is that the 5′/3′ nick at the double parallel crossover means that the four-way junction linking the two sheets can adopt many more geometries. This may result in an increase in 'open' structures, i.e. strands which adopt the correct secondary structure, but tertiary contacts are not made. Intermolecular kissing loop interactions can then form more freely, giving rise to multimers of various sizes, which could explain both these observations.

- The origami$_{\text{BLKL}}$-5′ and origami$_{\text{BLKL}}$-3′ co-transcriptional variants (lanes 2 and 4) look

similar, with a strong primary band, lining up with the compact band for origami$_{\text{BLKL}}$-N.

- The annealed versions of each sample (lanes 3, 5 and 7) look very similar, although the 3′T and NT variants show a substantial amount of aggregation. The bands corresponding to the annealed nanostructure also seem to run very slightly further (≈1%) than their co-transcriptional counterparts. This may be an indication of a slightly different folding confirmation before and after annealing. Since the annealed samples run faster, these are likely to be more compact, and most likely correspond to the correct folding conformation.

Negative-stain TEM

All three nanostructure variants were prepared by transcription (Appendix B.2.2), purified by LiCl precipitation (Appendix B.3.2) and resuspended in PBS-Mg buffer to a final concentration of 10 nM. A subset of each sample were prepared by thermal annealing, and TEM grids were prepared (Appendix D.1.1) for both the co-transcriptional and annealed samples. Imaging results for the six samples are shown in Figures 5.21-5.26. An overview of these results is given below.

- For all samples a population of apparently well-folded structures is seen along with populations of mis-folded structures and larger aggregates. This is consistent with gel results.
- For all three variants, a much lower grid coverage is seen for co-transcriptionally-folded structures compared to structures generated by thermal annealing. This is most likely a result of sequestration of RNA strands into large aggregates. By AGE, little-to-no aggregation is seen for co-transcriptionally-folded structures, leading to two possible explanations: transient aggregates exist in solution and fall apart in the gel (another explanation for the slightly lower mobility of co-transcriptionally folded structures); or co-transcriptionally-folded nanostructures are monomeric in solution, but in an open conformation (not all intended tertiary contacts form), and aggregation is induced by blotting. In both these explanations, these aggregates are most likely held together by tertiary contacts rather than secondary structure. By contrast, the more stable aggregates observed by AGE of annealed structures may be formed through intermolecular secondary structure formation.
- A quantitative estimate of folding yield was made based on the micrographs for all six grids. Each individual RNA object observed is specified as well-folded or misifolded. Misfolded objects include misshapen monomers, as well as multimers and aggregates, so the fraction of well-folded objects will be significantly greater than the fraction of well-folded RNA strands. Results are shown in the Table 5.4. For the annealed nanostructures, the same results are seen for all three variants (22-23% well-folded), whereas for the co-transcriptional variants the fraction of well-folded structures is greatest for the 5′T variant (32%) followed by the 3′T variant (23%) followed by the NT variant (18%).

	% well-folded objects	
	Co-transcriptional	Annealed
origami$_{\text{BLKL}}$-5′T	32 (219)*	22 (312)
origami$_{\text{BLKL}}$-3′T	23 (191)	22 (250)
origami$_{\text{BLKL}}$-NT	18 (254)	23 (235)

*Number in brackets is total number of objects analysed.

Table 5.4: Fraction of well-folded objects for origami$_{BLKL}$ variants.

Discussion

In this section I have designed nanostructures based on Geary's RNA tile [114, 115], where a BLKL motif is incorporated to build in three dimensions. Unlike the origami$_{\text{PX}}$ architecture, individual sheets appear to fold well, and interact with one another. Three different variants of a 2-sheet nanostructure have been designed, and analysed by negative-stain TEM. For all variants, a significant minority of apparently well-folded structures are observed which, once again, is an improvement on the origami$_{\text{PX}}$ architectures. In all cases, however, the yield of well-folded structures is restrictively low for my intended tagging application. A cryo-EM tag will need to be larger than these designs (with a thickness of at least four sheets) to provide sufficient contrast, and the more complex folding pathway for larger nanostructures would most likely result in even lower yields of folded structures.

5.5 Summary

In this chapter I have considered different methods for extending an established 2D RNA origami architecture [114, 115] into three dimensions. I firstly introduced the RNA origami architecture, highlighting key features of its design, and suggesting different methods for building in 3D. To do this, I envision a multi-layer nanostructure resembling a DNA origami design, where each layer/sheet is a tile based on Geary's 2D RNA tile [114]. For the goal of linking sheets, I have introduced two new motifs: PX$_{\text{major}}$ and BLKL. The PX$_{\text{major}}$ motif is a paranemic crossover between parallel helices, comprising a pair of parallel double crossovers separated by a major groove. The BLKL motif combines a 1° bulge loop and a 179° kissing loop.

My first attempt at designing 3D RNA origami used the PX$_{\text{major}}$ motifs. I have shown that a PX$_{\text{major}}$ motif can form by pairing of two bulge loops. To incorporate this motif into a nanostructure, external hairpins on RNA sheets were selectively replaced with symmetric, 8-nucleotide bulge loops, with sequences specified for the formation of PX$_{\text{major}}$ motifs between adjacent sheets. The design of a full 4-sheet nanostructure is shown. AGE and negative-stain TEM results of the transcribed RNA strands show that the structure does not fold as designed. Even transcription of a single sheet from the nanostructure results in mis-folding, suggesting that the bulge loop motifs themselves prevent folding, possibly due to incorrect formation of the secondary structure, or by blocking the intended helical stacking behaviour. Despite challenges in incorporating PX$_{\text{major}}$ motifs into this design architecture, the motif

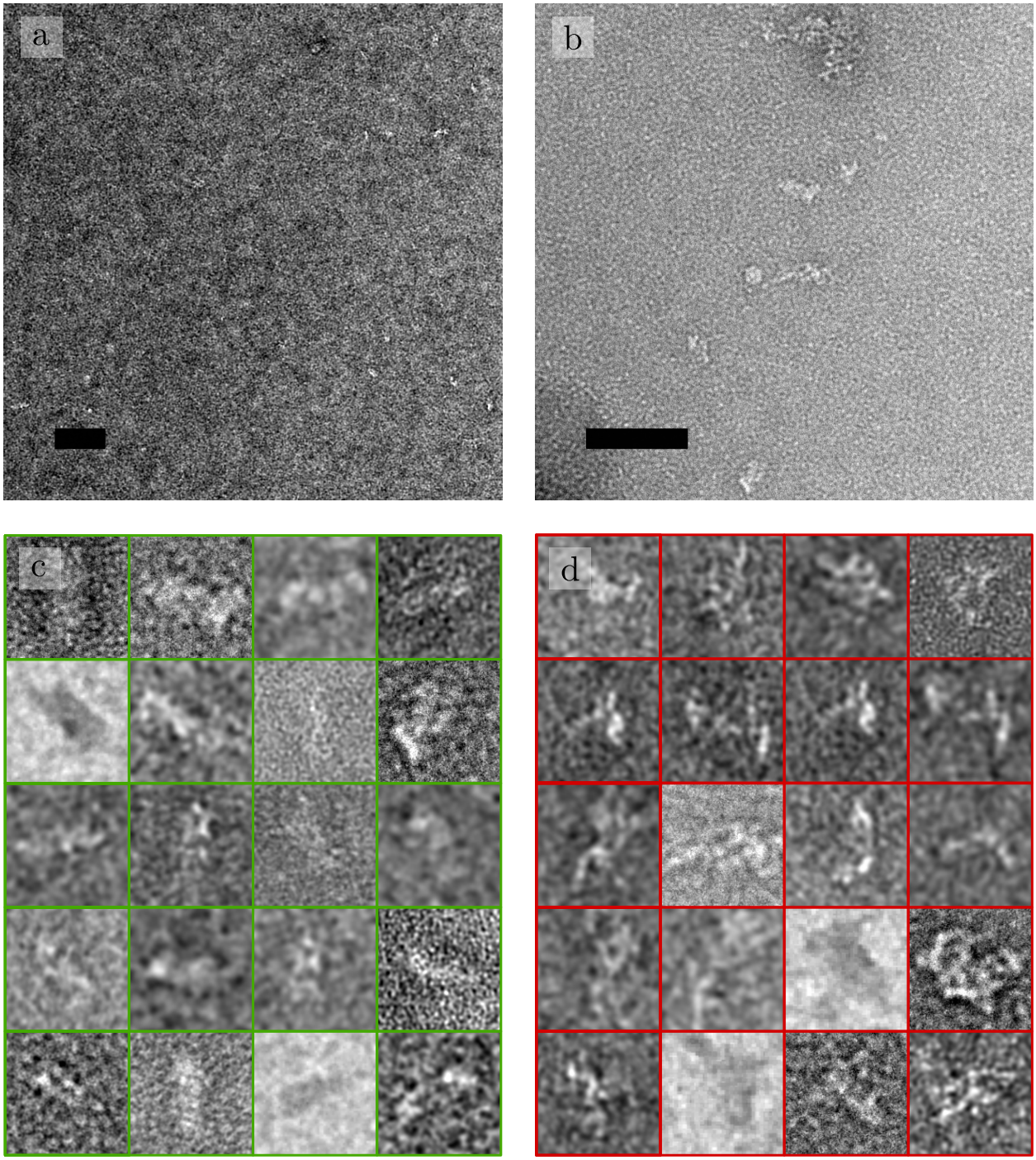


Figure 5.21: TEM overview of origami$_{BLKL}$-5′ generated by co-transcriptional folding. (a) and (b) are example wide-field and high-magnification TEM views, scale bars 100 nm. (c) Examples of what appear to be well-folded structures, panels are 50 nm×50 nm. (d) Examples of what appear to be poorly-folded structures, panels are 50 nm×50 nm.

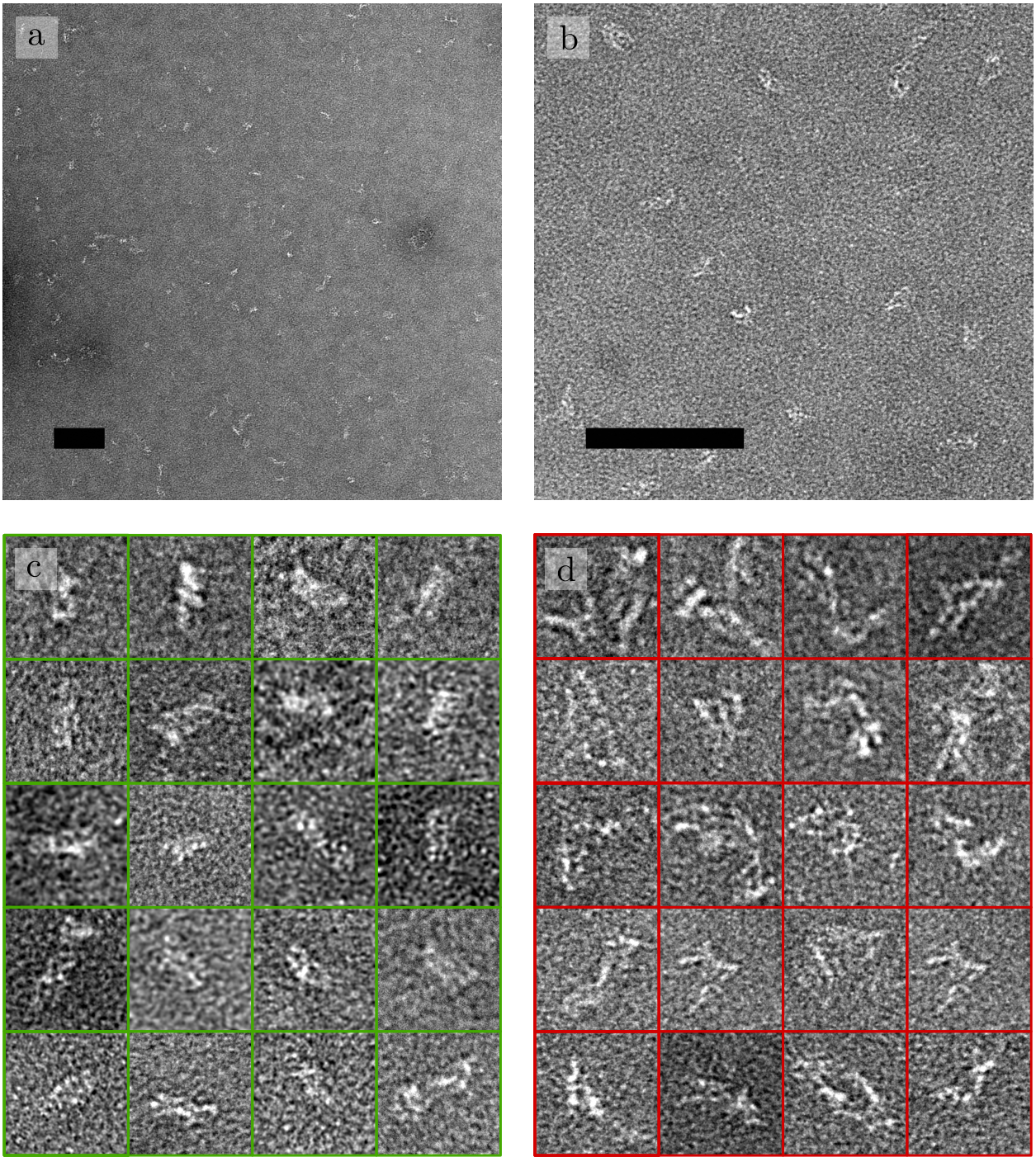


Figure 5.22: TEM overview of origami$_{BLKL}$-5′ generated by thermal annealing. (a) and (b) are example wide-field and high-magnification TEM views, scale bars 100 nm. (c) Examples of what appear to be well-folded structures, panels are 50 nm×50 nm. (d) Examples of what appear to be poorly-folded structures, panels are 50 nm×50 nm.

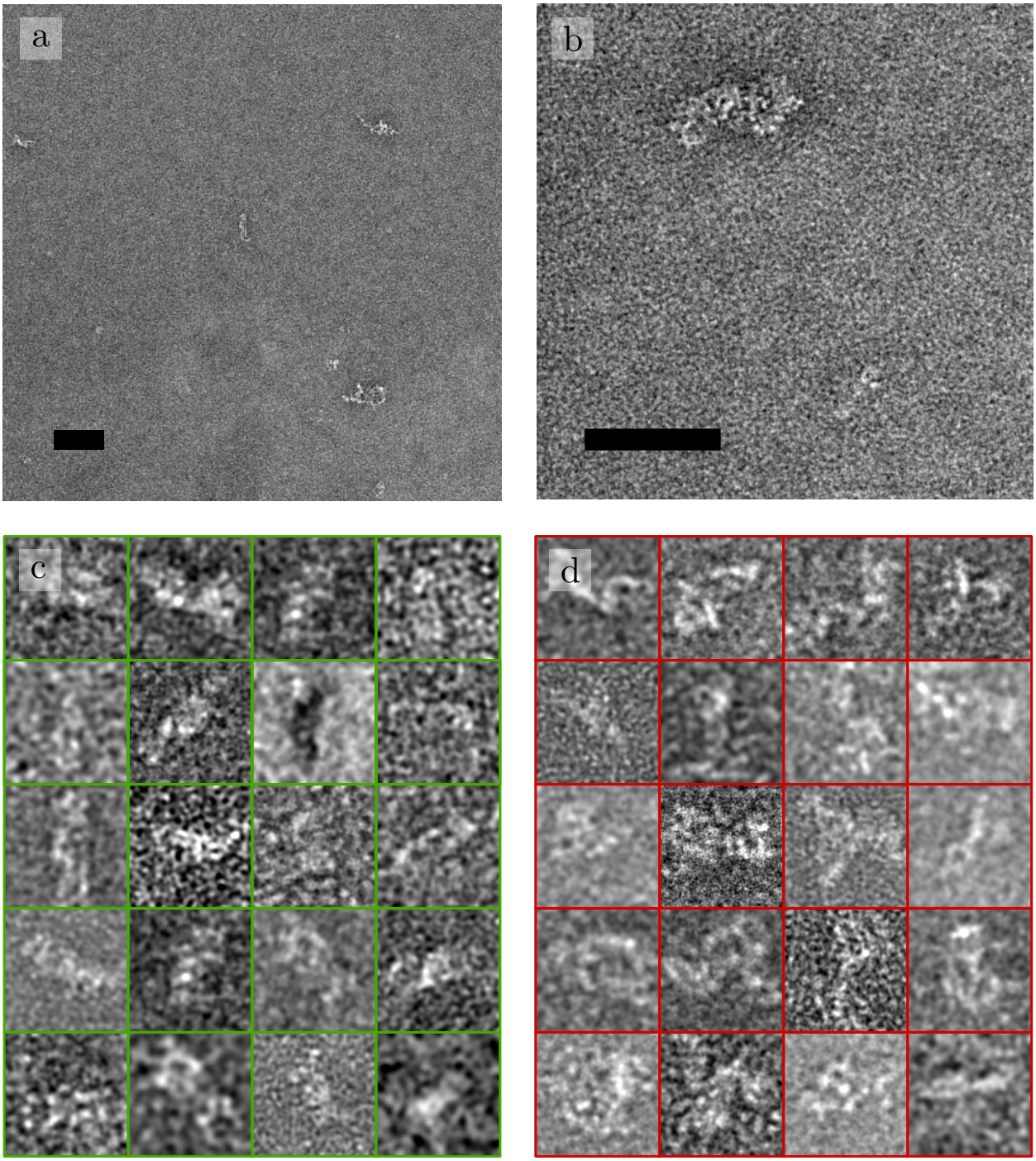


Figure 5.23: TEM overview of origami$_{BLKL}$-3′ generated by co-transcriptional folding. (a) and (b) are example wide-field and high-magnification TEM views, scale bars 100 nm. (c) Examples of what appear to be well-folded structures, panels are 50 nm×50 nm. (d) Examples of what appear to be poorly-folded structures, panels are 50 nm×50 nm.

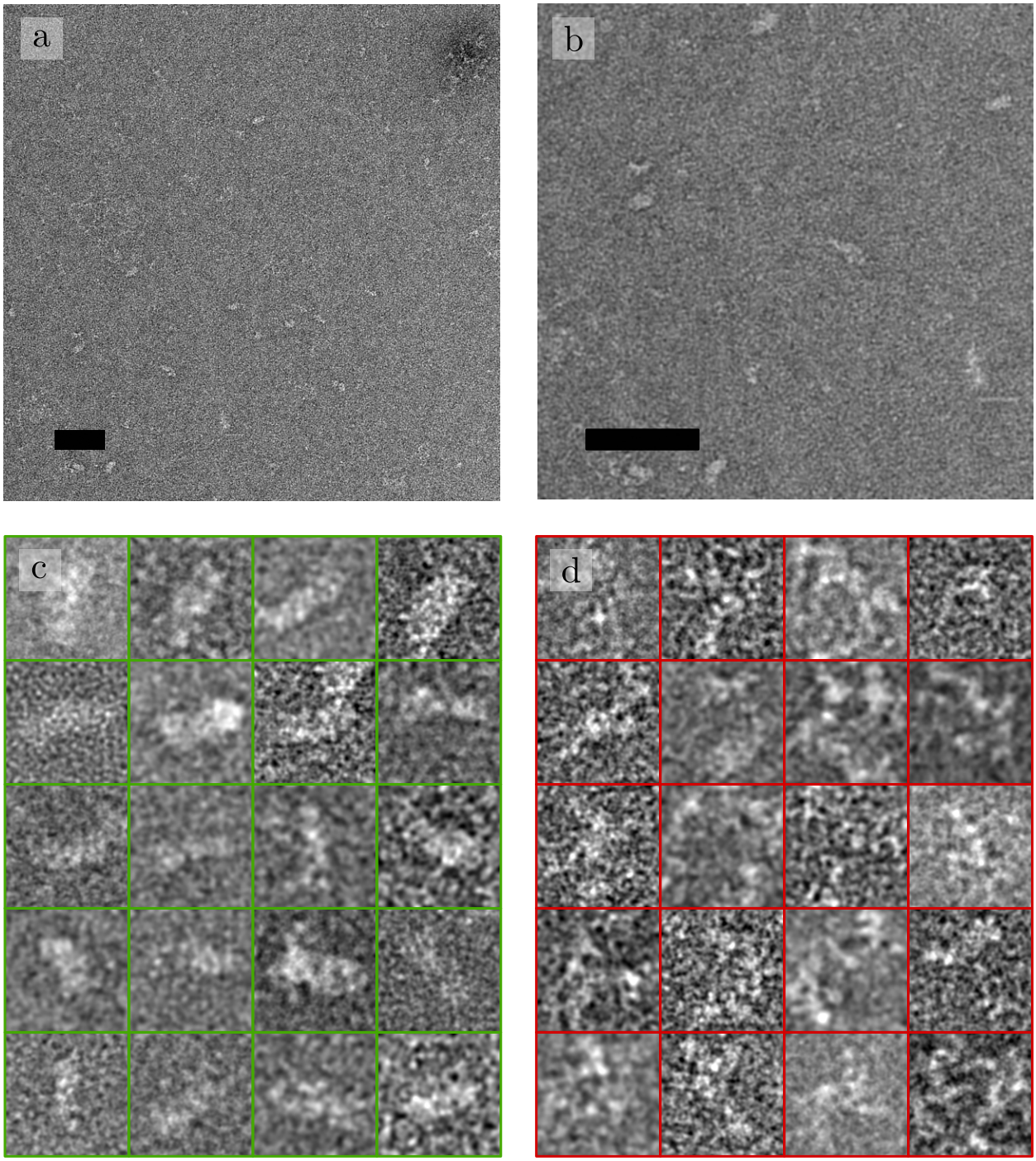


Figure 5.24: TEM overview of origami$_{BLKL}$-3′ generated by thermal annealing. (a) and (b) are example wide-field and high-magnification TEM views, scale bars 100 nm. (c) Examples of what appear to be well-folded structures, panels are 50 nm×50 nm. (d) Examples of what appear to be poorly-folded structures, panels are 50 nm×50 nm.

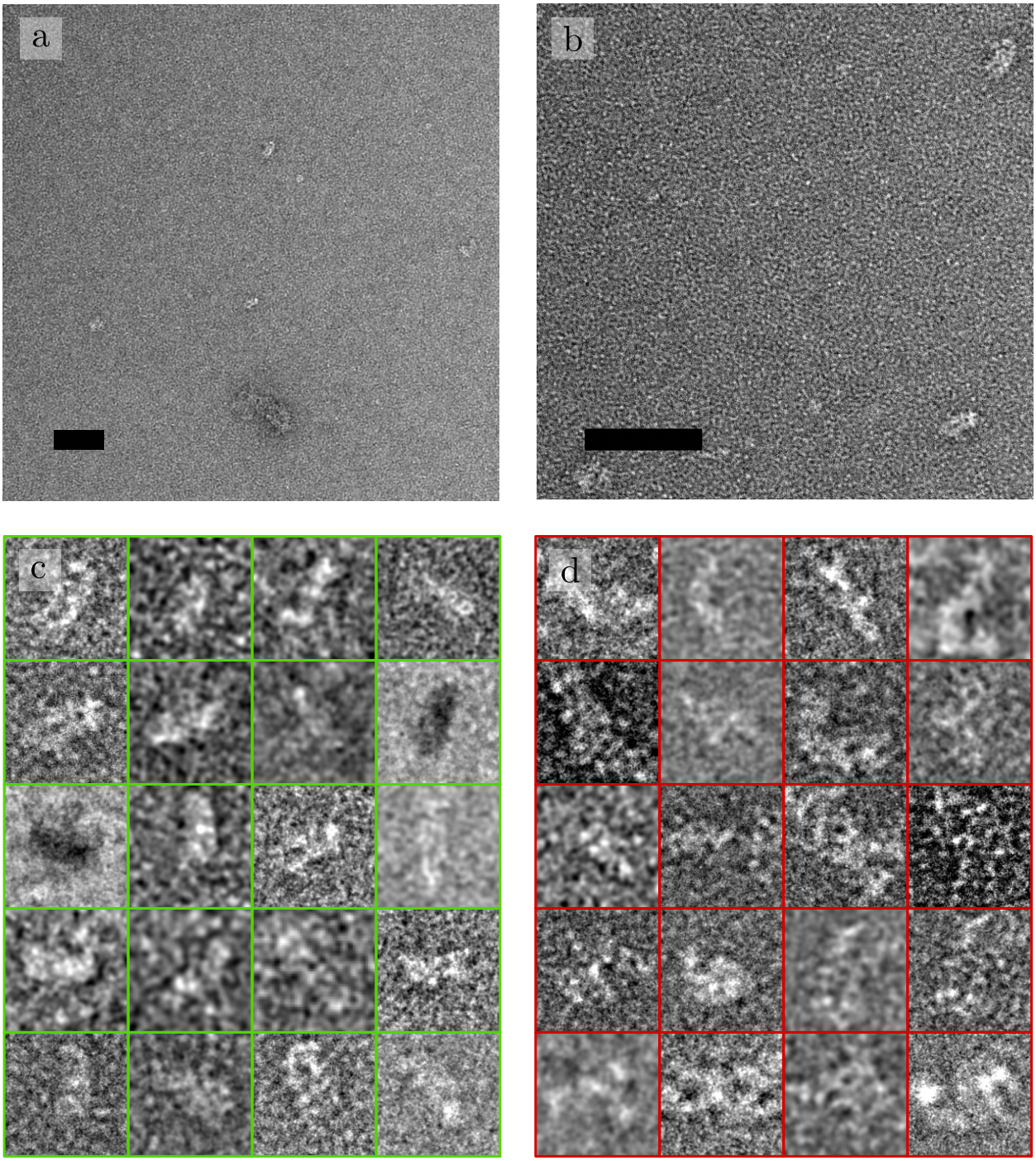


Figure 5.25: TEM overview of origami$_{BLKL}$-N generated by co-transcriptional folding. (a) and (b) are example wide-field and high-magnification TEM views, scale bars 100 nm. (c) Examples of what appear to be well-folded structures, panels are 50 nm×50 nm. (d) Examples of what appear to be poorly-folded structures, panels are 50 nm×50 nm.

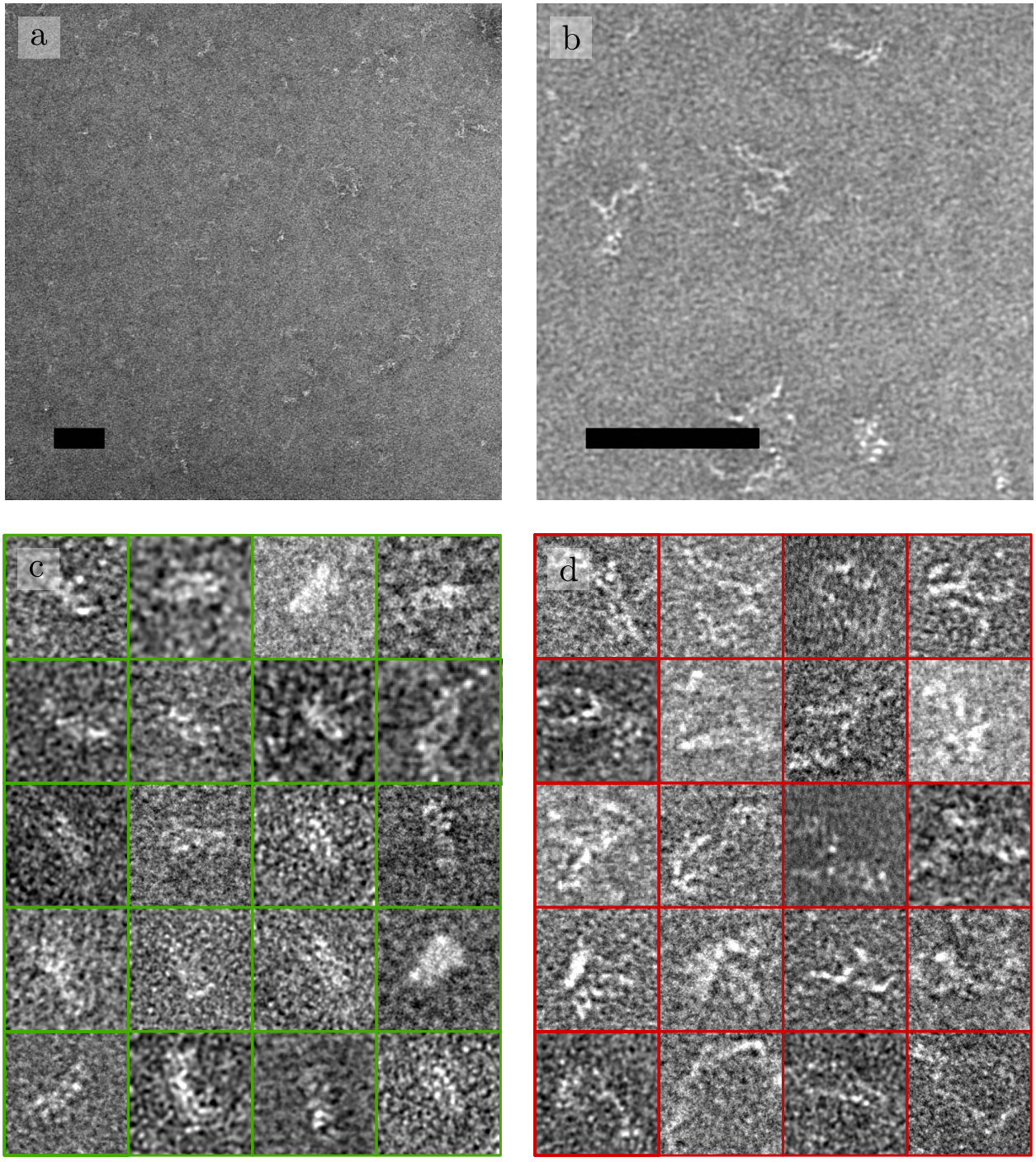


Figure 5.26: TEM overview of origami$_{BLKL}$-N generated by thermal annealing. (a) and (b) are example wide-field and high-magnification TEM views, scale bars 100 nm. (c) Examples of what appear to be well-folded structures, panels are 50 nm×50 nm. (d) Examples of what appear to be poorly-folded structures, panels are 50 nm×50 nm.

itself may still be useful for incorporation into more open designs with fewer geometric restrictions imposed by the surrounding structure.

My second attempt at designing 3D RNA origami used the BLKL motifs. Similar to the previous design, external hairpins on RNA sheets were selectively replaced with BLKL motifs. Firstly, two complementary sheets were designed for hetero-dimer formation. Individual sheets were analysed by AGE and positive-stain TEM, and appear to be well-folded. A 1:1 mix of sheets was also analysed by AGE, revealing that although sheets do appear to interact, they do not do so in a clean 1:1 ratio, most likely due to geometric limitations. Despite this concern, three single-stranded variants of these structures were designed with slightly different scaffold routing. The nanostructures were transcribed and characterised by negative-stain TEM. For each variant, we observe a significant fraction of seemingly well-folded structures, which is an improvement on the origami$_{\mathrm{PX}}$ architecture. This suggests that the BLKL motif is more suitable than the PX$_{\mathrm{major}}$ motif for incorporation into tightly-packed nanostructures. In addition to the well-folded structures, a majority of mis-folded structures and large aggregates are also seen. For the development of genetic tags for cryo-EM, a higher yield of well-folded structures will be required, since mis-folded structures and aggregates will complicate data interpretation, especially if the targeting component is still functional. Therefore, for the development of genetic tags for cryo-EM, the grid-like architecture introduced in Chapter 4 will be a more suitable starting point.

Chapter 6

Conclusion

In this thesis I have presented new methods for using nucleic acid nanostructures as tags for cryo-electron microscopy. This work has been sub-divided into two projects: in Project 1 I have shown the design, functionalisation and application of DNA origami nanostructures as tags for cryo-EM and in Project 2, I have shown the design, synthesis and characterisation of novel 3D RNA nanostructures generated by co-transcriptional folding.

In Chapter 1, I introduced cryo-EM as a technique for structural biologists, highlighting both its strengths (high resolution, minimal disruption to sample) and weaknesses (low contrast, difficulties in locating specific structures) for *in situ* characterisation of biological macromolecules. I discussed methods for labelling target structures in cryo-EM and explained the limitations of currently-used techniques, illustrating the need for a new tagging method. I suggested using functionalised DNA and RNA nanostructures for this goal. I then introduced DNA and RNA nanotechnology, focussing on DNA origami and co-transcriptionally-folded RNA nanostructures, which are the structural basis for my tagging methods.

Chapters 2 and 3 are concerned with Project 1, focussing respectively on the structural design, and functionalisation/delivery of nanostructure tags. In Chapter 2, I showed the design and construction of various DNA origami nanostructures, and assessed their suitability as tags for cryo-EM. Firstly, I showed the design and synthesis of a wedge-shaped origami nanostructure, which I used to develop a purification protocol for cryo-EM. Micrographs of these nanostructures in cell lysate have shown that, even in a noisy biological environment, honeycomb-lattice origami nanostructures can provide high contrast, comparable to membranous cellular structures. I then introduced two new nanostructures to inform choices on mutliplex labelling. A tetrahedron origami nanostructure, where each arm is a six-helix bundle, was imaged in cell lysate and was found to provide insufficient contrast for effective tagging. A pointing-hand honeycomb-lattice origami nanostructure was also evaluated. The wedge and hand structures offered similar contrast, but from some views the two designs are challenging to distinguish. Based on these microscopy results, I proposed a set of design constraints which I used to design a signpost origami nanostructure, comprising a high-contrast, asymmetric 'sign' attached to a six-helix bundle post. Both square- and honeycomb-lattice variations of this design were constructed. There

is no significant difference in contrast for the two types, but they can be easily distinguished despite their similar global architectures. Designs for sign-post variants with different global architectures and different lattice types are also proposed for multiplex labelling.

In Chapter 3, I addressed the aims of targeting and delivering tags. The targeting moiety introduced in this chapter is an RNA aptamer which binds to GFP, developed originally by Shui and colleagues.[137] I used gel electrophoresis to demonstrate EYFP-binding properties of the free aptamer, as well as the aptamer conjugated to a signpost origami nanostructure. Isothermal titration calorimetry was used to quantify the binding of the free aptamer to new GFP analogues: mEGFP, mVenus and sfGFP. I then introduced the Moloney murine leukaemia virus (MoMLV) with YFP-tagged envelope glycoproteins as a test system for tagging, culminating with *in situ* cryo-electron tomography of aptamer-conjugated signpost origami nanostructures targeting the intact virus. High-contrast, high-efficiency labelling was observed, illustrating the usefulness of this tagging method for virus targets. Finally I show the electroporation of spermine-stabilised nanostructures into prokaryotic cells. Currently the delivery efficiency is too low for intracellular labelling applications, but this may be improved by optimising electroporation conditions or nanostructure design.

Chapters 4 and 5 are concerned with Project 2, with the two chapters demonstrating two different avenues toward RNA nanostructure design. In Chapter 4 I introduced a novel, grid-like architecture, where nanostructures comprise units, arranged as pixels on a square lattice. Each unit is a branched hairpin structure, with a contiguous core of stacked helical domains, and hairpin 'arms' branching out. Neighbouring units can associate by 120° kissing loop interactions between arms and are further connected through a path of single-stranded linkers. I have presented a strategy for designing sequences for this architecture. I have shown by agarose gel electrophoresis that individual units interact as designed, and can be incorporated into nanostructures. TEM and gel electrophoresis results suggest a well-folded majority population of nanostructures. To generate structures of a more suitable size for cryo-EM tagging, I have demonstrated the programmable supermolecular assembly of multiple nanostructures into finite complexes, of up to 5 components, through patterns of 180° kissing loops. I have also shown the incorporation of an aptamer into the nanostructure to allow targeting. This design strategy offers one viable route toward building genetically-encoded nanostructure tags for intracellular targets.

In Chapter 5, I presented options for extending an established 2D RNA origami architecture[114, 115] into 3D. I have introduced two new motifs to bring together parallel helices: the PX_{major} motif, comprising two parallel crossovers separated by a major-groove half-turn; and the BLKL motif, which combines a 1° bulge loop and a 179° kissing loop. I have shown the design, atomic model generation, transcription and characterisation of a 3D RNA origami held together with PX_{major} motifs. This nanostructure is not capable of folding as designed, either co-transcriptionally or by thermal annealing. This is most likely due to the formation of the 8 nucleotide bulge loops which are the precursor to the full formed PX_{major} motifs, either resulting in misfolding of the secondary structure, or sterically blocking the required helical stacking behaviour. Despite difficulties with incorporating this motif into the origami architecture, it may prove useful for incorporation into new and different nanostructure architectures. I then presented the design of two-sheet nanostructures

held together by BLKL motifs. Although a more promising prospect, the yield of well-folded structures is low, so the architecture is likely not suitable for designing larger-scale nanostructures as required for cryo-EM tagging. Altogether, the grid-like architecture of Chapter 3 provides a more reasonable starting point for the design of genetically-encoded cryo-EM tags.

I anticipate that the body of work presented in this thesis will be a valuable resource for structural biologists. The DNA origami tagging method introduced in Chapters 2 and 3 can be utilised for the tagging of extracellular domains of cell-membrane proteins and viral glycoproteins with little further development. This will allow the *in situ* identification and characterisation of target proteins which is often impossible with current methods. Membrane proteins are, in fact, some of the most important targets for *in situ* characterisation, since they can adopt different conformations and behaviours in native membranes compared to the more commonly-used purified and reconstituted systems.[178] This tagging method has already been adopted for ongoing work in the characterisation of the FLRT2:Unc5:Latrophillin complex involved in cell adhesion and repulsion in neurones[160] as well as characterisation of the HSV-1 entry machinery[161], and we foresee a multitude of interesting future applications for this technique.

The development of a genetically-encoded RNA-nanostructure tag for cellular cryo-EM is a longer term goal, but one I have made significant progress towards. In particular, the grid-like architecture presented in Chapter 4 offers a viable route towards building RNA nanostructure *in vivo* by virtue of its simple, modular folding pathway. For *in vivo* nanostructure assembly, it will likely be necessary to optimise aspects of the design, e.g. by carefully tuning kissing interactions, or introducing structured linkers between units. Nonetheless, this architecture does provide a suitable starting-point for the design of genetically-encoded tags, which would open up a plethora of new applications.

Appendix A

Sequences

A.1 gBlock sequences

Promoter sequences are shown in blue, nanostructure template sequences are shown in black and non-coding sequences (at 5′ and 3′ termini) are shown in grey.

A.1.1 Grid-based RNA Nanostructures

Pentomino Nanostructure

ACGTCAGAATTCTAATACGACTCACTATAGGGAGACTTTAATTCGGTCGGTAGT
GGACATCGGCCGAAGAACCTTACCTGGCTTAGTTTAATTTATTGTCTACGTCT
CGGACAGTAAAGAACCTTAAATGAAGTAGTTAAATTCGGGTATTTTCGGATGC
TCGAAGAGCCATACCCTGAGTCGTTTAATTCTCGTCGGTTCGCCGATGAGAAG
AACCTTACCGTCGTGAGTTTAATTCGTGCTAGAGTGGACCTAGCGCGAAGAGC
CTTACCGACATCTCTTTAATTCGTGGAGCACGTCTCGTTCTACGAAGAGCCTTA
CCGGAAGCTGTTTAATTGTGCGTATTTCGATACGTACAAGAACCTTATTAGCA
AAGGTTAAATTGCGGACGGTTCGCCGTCTGCAAGAACCTTACCCAGCTTTTCG
AGGTTGGGGCGACTTTTGTTAAGCGACAGTTTTCGGGCGAGAGGTGTTGGGCG
ACTTATGACGGGCGACGATTTAGGGGCGACTATTTTATTGCGACTAGGCTAGG
GCGAGTCTCTCTCCACACACCAAACAACAAACACAAAACACACTACATAAGTCA
TAGTTTAATTCGTTGGCTTTCGAGTCAGCGAAGAACCTTACCTGGTTGGGTTT
AATTCGTGAGTAAGGGATCGGAGAGACGTTTCGGTCTCTTGCTTACGAAGAGC
CTTACCGACCTGAGTTTAATTCGATGAATAATTGGCTGGGCAGGCTCCGGTCA
ATTGTTCGTCGAAGAACCTTACCGTGGTTTGTTTAATTCGTGGAGGTTCGCCT
TTACGAAGAACCTTACCTTAAGCCGTTAAATTCGGACCTGTTCGTAGGTTCGA
AGAGCCTTACCGTTGCCGGTTTAATTCGCGGCACTGGGATATCGGAGACGTCG
GTATCTCAGTGTTGCGAAGAACCTTACCACCGGGCGTTAAATTCGTCTCGTGT
GAGAGGTTGCAGGCTAACCTTTCACATGAGGCGAAGAACCTTACCGTTTTCGG
TTTAATTCGGTGTTGTTCGTAACATCGAAGAACCTTACCCGTCTTTTCGAGGA
TGGGGCGACTGAAGATGGGCGACGTCTGGTGGGCGACTGGTAACGGGCGACG

GTTTGAGGGCGACAGACTACGGGCGACTTAGGTTGGGCGACTCAATCGGGGCG
ACTGTGATCAAACAAAGAAACAAACAAAACCCCCAACAACAAACAAAGAGCGT
GTTCAATTCGGAGTTCTTCGGAGCTTCGAAGAGCCTTACCTCTGGGAGTTAAA
TTCGTGTCTGTTCGCGGACGCGAAGAACCTTACCAGGACGGGTTTAATTCGAG
TTACAGCCTGCGTAATTCGAAGAACCTTACCATCGTGTGTTTAATTCGTACGG
CTCGTGGTGTCGTGCGAAGAACCTTACCTGTACGAGTTTAATTCGGTGACGTT
CGTGTCATCGAAGAACCTTACCAGTTCCGGTTTAATTCGTCGGGCTTCGGCCT
GGCGAAGAGCCTTACCGAGACTTGTTTAATTCGCTCGTTAGCCTGCAGCGGGC
GAAGAACCTTACCGTAACGTGTTTAATTCGGTTTAGTCGTGGTCTAGATCGAA
GAACCTTACCGGTAGTTTCGATTATCGGGCGACATGTTATGGGCGACAGGTTT
CGGGCGACTGGAATTGGGCGACTTGTATAGGGCGACATATGATGGGCGACTTG
TCTTGGGCGACTCTCGGAGGGCGACACGTTTACCAAGCAAAACACACAAACAA
AAACAAACAACAACCCATGACGCGTTTAATTCGTGTCATGTTCTGTCTGGTCC
ACTCAGATAGGATATGACGCGAAGAACCTTACCATGAGTCGTTTAATTCGTTA
GTCATGTAGTAGGGATGGTTCCTGCTACGTGACTGGCGAAGAGCCTTACCACA
GACTGTTAAATTCGGACCTCTTCGGGGGTTCGAAGAGCCTTACCAGATGTGGT
TAAATTCGGGGTATGCGAGGATCGACCACGATGGTCCTTGCGTACTCCGAAGA
ACCTTACCCGTCTTGGTTTAATTCGTGGATAGGGAGTACTCGTCCACTGAGTG
CTTCCTGTCCGCGAAGAACCTTACCGATGGAAGTTAAATTCGGGCGGAGTTGG
GTAGGGATGGTTCCTATCCGACTTCGTCCGAAGAGCCTTACCAGACCGTGTTT
AATTCGTTGACTTTCGGGTTAGCGAAGAACCTTACCAGTCGTGGTTTAATTCG
ATTTCGAGGTGCTGCTACCACGAAGTAGTATCTCGAGATCGAAGAACCTTACC
ATTCACTTCGGTGAGTGGGCGACTATGATTGGGCGACGTGGTTTGGGCGATTT
CTATTGGGCGACCGAGATGGGGCGACTACATTTGGGCGACAGTTTGTGGGCGA
CGGCTTATGGGCGACGTGTTAAGAAAAGAAAAAGGAAGGAGAAGGAAGAAGA
AAAAGAGATGGCCTGTTGAATTTGGCCGTCTTCGGATGGTCGAAGATCCTTAC
CTGTCGCGGTTTAATTCGGGTCCTAACCATCGGGACTCGAAGATCCTTACCAA
AGTTGGTTTAATTCGCTGGTGTTCGCATCGGCGAAGACCCTTAGGGGAATTAG
TTAAATTCGCTGACCTTCGGGTCGGCGAAGAACCTTACCTGTGTTCGTTTAAT
TCGTTGCGTTTCGGCGCGACGAAGATCCTTACCGGATGCAGTTTAATTCGGGT
ATGAACCATCCATATTCGAAGAACCTTACCGGTCCGGGTTTAATTCGGCATCT
TTCGGGATGTCGAAGAACCTTACCTGGCGATGTTAAATTCGCTGTGTTTCGAT
ACGGCGAAGAACCTTACCCGTATTTTCGAATGCGGGGCGACATTGCTAGGGCG
ACCTGGATCGGGCGACTGTATCTGGGCGACGAATACGGGGCGACTAGTTCTCC
GCGACCAATTTTGGGCGACTGCGATAGGGCGACGGGCTAAGCTTCCCCTTGGG
GCCTCTAAACGGGTCTTGAGGGGTTTTTTG

Unit α

GCATACGTCAGAATTCTAATACGACTCACTATAGGGAGACGAACCCACTTGAGC
AGTGGACGCTTAAGTACGCGCGGGACAAGAGCGATCGTACGTCTCACGGTCG
CAGGCTTTGGATTTCTGCGTGCCATTCGTGGTACGCTTGGATGCGACCCTACC
TCGAGCTTCGGCTTGAGGCCGACCGTGTGAAAGACTCATTACAGTGGACGTAG

TGAGAGTCGTGGTCATTCCGGAGGAGTACGTCTCACTTCTCCTTGACTGGGCA
GCGGTCTGTGTATTCGTACGCAGAAATCAGGAGCACTACGCACTCGATTCGTC
GGGTGCACGCTCGCGCTTCGGCGTGAGCCAGCTTCTGAAAGCCTAGTCCCGAC
TACGAAACACATGGTCACTCGTATCCTCTCCGAAGCAATCCTGCGCAACGTCTC
CCAAGCTTCCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCATATG
CGTGACCCGGACGAACCACAGA

Unit β

TTTCAGCAAGATGTTTACGTCAGAATTCTAATACGACTCACTATAGGGAGACCAC
ACATCTTTGCTTCGGCAGAGACAGGCTTAACGAAAGGCTGGTCGAGACGTGAC
TAGCAAGCGCTAGCAAAGAGGAGCTCGCAGGCTGAGTTCCAAGGCGGTGGCCT
CCGTATACATTCGTGTGTACCTTCCTTGTCCACTTACACGGCTTTCGAGCTGTG
CCGATCTATCAAGGAGTCTGTCGAGACGTGACGGACAAGCAGTAGGACCAAGG
CGCTCGCAGGCTGAGTGCCAAGTAGTGTCGAGAACTCTGACTTCGGTCGGAGA
AGGTGGCCGTTCGCGGTCACCAAGACGCTACAACCTGCTGCAAGATGGATCAA
GGACGAGGATCCCATCGCCGAGCTGGCGCAGGTTGAGCCAAGGTCTCCCAAGC
TTCCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCATATGCGTGAC
TTACAAAGTACTGGTTCTTT

Nanostructure A

TGCATGATCTACGTGCGTCACATGCAGTACGAATT CTAATACGACTCACTATA
GGGAGACGAACCCACTTGAACGTCCACTGTTTAAGTACGCGCGGGACAAGAG
CGATCATTTCGATGGTCGCAGGCTTTGGATTTCTGCGTGCCATAAGGAGGCAA
TGGTACGCTTGGATGCGACCCTACCTCGAGAGCAGGCTTCTTGAGGCCGACCG
TGTGAAAGACTTATTACGTCCACTGTAGTAAGAGTCGTGGTCATTCCGGAGGA
GTTTCGACTTCTCCTTGACTGGGCAGCGGTCTGTGTAGAAGTGCTGACTACGC
AGAAATCAGGAGCACTACGCACTCTAGCAGGCTTAGGGTGCACGCTCGTTCTT
CGGAATGAGCCAGCTTCTGAAAGCCTAGTCCCGACTACGAAACACATGGTCAC
TCGTATCCTCTCCGAAGCAATCCTGCGCAACGTCTCCCAACAAAACCCAAAACA
ACAACAAACACCAAGGGAGACCACACATCTATGTCAAACTTCGCATGACGTAG
ACAGGCTTAACGAAAGGCTGGAATTCGTTCTAGCAAGCGCTAGCAAAGAGTAG
CAGAGAACGCCTGTTACAAGGCGGTGGCCTCCGTATACAAGCCTGCTGTGTAC
CTTCCTTGTCCACTTACACTGTCTAAAGCGAGCATAGATAGTGCCGATCTATCA
AGGAGTATGTCTTCGGACGTACAAGCAGTAGGACCAAGAGGATCAGAACGCGA
TTCTCAAGTAGTGTCGAGAACTCTAACAGCCTGCGTTGGAGAAGGTGGAATTT
CGATTTCACCAAGACGCTACAACCTGCTGCAAGATGGATCAAGGACGAGGATC
CCATCGCCGAGCTGGCGCAGGTTGAGCCAAGGTCTCCCGAAAACACAAACAAA
ACAAACACACAAAAAGGGAGATGAACAAACTGGGAATTCGTTCTCAGTCACCG
ATGTCTATAACTCTTAAGGAGACGTCTTGAGAGCCGTTCGGGCACAAAGTAGT
AGAGCGTTCTTCTGCTACAAGCAGTCGGGAGAAGATCTATGTTCGCATGGATC

AGTATATGACCGTAAAGTTGGACATTCGTGTTCAACGAGCCTGCCCAACAAGT
AATGTCGAGACGTGACGTTACAAGGACTGACGAGGGGCATGTTCGCGTTCTGA
ACGTGCAAGAAGTCGCAAGGAGTGTTACATTCGTGTGACACGAGGTTTAGATT
CGTCTGAACCAAGCGGCTTCGGGTCGGTCCCAGGGTAGGCGAGGTCGTATAAA
CCGGCTGCCAGCCTGAACCCGACGTCGGAACATCTCCCAAAACAAAACAACAA
ACAACAAAAACACAACCCAGACCAACACGTTTATCAGTGGACGATGAACCCGC
AGTGACACAACTACGCTAACGTCTCTAGTGTACAGGGATGGAGGAAGGTCGTC
TTTCGAGATGACGACGTGGGCCAGAAGGAGTATGTTCGCATGCTCGGCTGCTC
AGCACACCGAGTTTCAGTGGACGAAGCTCACGTCAGAGTCTTCCTCGGATCAC
GTCTCGATTCGATCGACATAGCACCCAGATGGTCTTCGGACTATCACGCTAGC
CCAAAGAGCATAAGTTCGCTTGTGCGAGCATGACTTTCGAGTTATGCAGGGGT
TAGCAAGCTGTGTCTCACTTTGACAAGCTGGGCAGAGGGCTCACGAATCCATC
CCAAGTCGCTGCAAGGTCTGGGAAGCTTTAGTAATGCAGACACTTGCGGTCCA

Nanostructure B

CTGGGCTGAATTCTAATACGACTCACTATAGGGAGAGGAACCCACTTGAGCTTC
GGCTTAAGTACGCGCGGGACAAGAGACATCATTTCGATGGTGTCAGGCTTTG
GATTTTCAGATGTACGCAGGCTGTATATCTTTGATTGCGACCCTACATCGAGT
TTCGACTTGATGCCGAACGTGTGAAAGACTCATTAATTCGTTAGTGAGAGTCG
TGGTCATTCCGGATGATCAAACAGCACATGATTATCCTTGACGGGGCAGCGGT
GTTTGTCGCAGGCTGACGAACAAATCAAGATCACTACGTACTTAATTCGTTAG
GTACACTAGATGCACGTGAGGTTGGATGTTCGCATTCAACGATCCTGGCCAAC
AAGTAATATCAAAGCCTCCATGATGTTACAAGGACTGACGAAAGGAATCTACG
CAGGCTGTAGGTTCAAGAAGTCGCAAGGAGTGTTACATTCGTGTGACACGAG
GTTTATATTCGTATGAACCAAGCGGCTTCAGGTCGGTCCCAGGCTAGGAGAGT
GCGTCTACAGATTTTGAAAGCCTCGTCCCGACTACGAAACACATGTTCACTCG
CAATCTCTCCGAAGCAATCCTGCGCAACCTCTCCCCAACCACCAAACAAACAA
CCAAACCAAACACGCTGACCAACACGTTGAGCTTCGGCTTAACCCGCAGTGAC
ACAAATCCGGTATTCGTACTGGACAGAGATGGAGGAAGGTCGTCTAGCCTGC
AGATGACGACGTGGGCCAGAAGGATGATGTTCGCATTATCGGCTGCTCAGCA
CACCGAGATGATTCGTCGTCTCACGTCAGAGTCTTCCTCGGATCTTCGGATTC
GATCGAGATAGCAAACACAGGTTCAGCCTGCGAATCTGACGTTAGCCCAAAGA
GAACTAGATAAGCTCGCAATCTGGTTCGATCAGTGTCCACTTACACAGTCTTC
GGATTGTGCCGATCTATCAAGGAGTCTGTCTTCGGACGGACAAGCAGTAGGC
CCAACATGCATAGCCTGCATGTATGAAGTAGTGTCGAGAACTCATTAGTAAGC
GAAGAACTAGTGAGAAGATGGAAGTTCGCTTTCATCAAGACGCTACAACCTGC
TGCAAGATGGATCAAGGACGCTGAAAGGGTTAACAAGCTGTCTCTCACTTTGA
CAAGCTGGGCAGAGGGCTCACGAATCCGTCTCAAGTCGCTGCAAGGTCAGCG
AAGCTTTGCATGATCTACGTGCGTCACATGCAGTAC

Nanostructure C

GGCTGAATTCTAATACGACTCACTATAGGGAGAGGAACCCACTTGAGCTTCGGC
TTAAGTACGCGCGGGACAAGAGACATCATTTCGATGGTGTCAGGCTTTGGATT
TTCAGATGTACGCAGGCTGTATATCTTTGATTGCGACCCTACATCGAGTAAAG
CTCGCATACTTGATGAAGAACGTGTGAAAGACTCATTAATTCGTTAGTGAGAG
TCGTGGTCATTCCGGATGATTTAACAGCACAAAATTATCCTTGACGGGCCAGC
GGTGTTTGTCGCAGGCTGACGAACAAATCAAGATCACTACGTACTTTAAAGCG
AAGATAAGGTACACTAGATGCACGTGAGGTTGGATGTTCGCATTCAACGATCC
TGGCCAACAAGTAATATCAAAGCCTCCATGATGTTACAACGACTGACGAAAGG
AATCTACGCAGGCTGTAGGTTCAAGAAGTCGCAAGGAGTGTTACATTCGTGTG
ACACGAGGTTTATATTCGTATGAACCAAGCGGCTTCAGGTCGGTCGCAGGCTA
GGAGAGTGCGTCTACCGATTTTGAAAGGCTCGTCCCGACTACGAAACACATGT
TCACTCGCAATCTCTCCGAAGCAATCCTGCGCAACCTCTCCCCAACCACCAAAC
AAACAACCAAACCAAACACTCAGACCAACACGTTGAGCTTCGGCTTAACCCGC
AGTGACACAAATCCGGTACAAACAGCACATGTACTGGACAGAGATGGAGGAAG
GTCGTCTAGCCTGCAGATGACGACGTGGGCCAGAAGGATGATGTTCGCATTAT
CGGCTGCTCAGCACACCGAGATGATTCGTCGTCTCACGTCAGAGTCTTCCTCG
GATCACAAGCCTCCAGTGATTCGATCGAGATAGCAAACACAGGTTCAGCCTGC
GAATCTGACGTTAGCCCAAAGAGAACTAGATAAGCTCGCAATCTGGTTCGATC
AGTGTCCACTTACACAGTCTTCGGATTGTGCCGATCTATCAAGGAGTCTGTCT
TCGGACGGACAAGCAGTAGGCCCAACATGCATAGCCTGCATGTATGAAGTAGT
GTCGAGAACTCATTAGTAAGCGAAGAACTAGTGAGAAGATGGAAGTTCGCTTT
CATCAAGACGCTACAACCTGCTGCAAGATGGATCAAGGACGCTGAAAGGGTTA
ACAAGCTGTCTCTCACTTTGACAAGCTGGGCAGAGGGCTCACGAATCCGTCTC
AAGTCGCTGCAAGGTCTGAGAAGCTTAAG

Nanostructure D

TGCATGATCTACGTGCGTCACATGCAGTACGAATT
CTAATACGACTCACTATAGGGAGACCGAAGACAGCTAGCTTCTGACGATGCAAC
AGCACAGTATTGTCAGGAGCTGGCTGGAGCAGTCTGGATAAGGAATCATCGAG
ATGGATGCAACAGCACAGTATTCATCTTGATGGTTCCCAGACGGTGCGTTAAG
GACGGCTCGGCAGTGTGGCAACAGCACAGTCATACTGCTGAGCTGTCCAGGGA
CGCAGAACATGAGTTAGCTGCATGTGTCGCAACAGCACAGTGATACATGTAGC
TGACTCGATAGGGTGaCTATTCGACTTTACTTATAGGTACGCAAGCCTCCAGTG
TGCCTATGAGTAGAGTCTCGTCTTGGAATAAGGCACGTAAGGCTAACTGTGCA
AGCCTCCAGTACGGTTAGTCTTATGTGCGAGTCTGCTCTATCCGCTCGGGACG
ACTTCGATGCAAGCCTCCAGTATTGAAGTTGTCCTGAGCTCGAGTGTTCGATC
ATCACGAGCATGTCACTACGCAAGCCTCCAGTGTGGtGACGTGCTTGtGACAGT
CGTACCTTCGGGTGCGACCAGAATACTCACGAGTAGACGGTCCGAGACACGTC
ATCCTACACTGTGTCCGAGCATCGTCAACAGGCTGCaAGGTCTCTCAAGCTTCA
CTAGCTCAGATTCAGTAGACCGCTGTTG

Nanostructure E

TGCATGATCTACGTGCGTCACATGCAGTACGAATTCTAATACGACTCACTATA
GGGAGACGAACCCACTTGAACTTCGGTTTAAGTACGCACGAGACAAGAGATAT
CATTTCGATGGTATCAGGCTTTGGATTTCTGAATGCCAGCAGGCTTGGTATTC
TTGGATGCGACCCTACCTCGAGAGTCCACTTCTTGAGGCCGACCGTGTGAAAG
ACTTATTACAAAGCTCGCATGTAGTAAGAGTCGTGGTCATTCCGGAGGAGTAA
ACAGCACATACTTCTCCTTGACTGGGCAGAGATCTGTGTAGCAGGCTTACGCA
GAAATCAGGAGCACTACGCAGTATAGTCCACTTATGCTGCACTATATGCACGT
GAGGTTGGATGTAAGCGAAGAACATTCAACGATCCTGGCCAACAAGTAATATC
AAAGCCTCCATGATGTTACAAGGACTGACGAAAGGAATCTACGCAGGCTGTAG
GTTCAAGAAGTCGCAAGGAGTGTTACAGTCCACTTGTGACACGAGGTTTATAT
TCGTATGAACCAAGCGGCTTCAGGTCGGTCCCAGGCTAGGAGAGTGCGTATAC
AGCTTCTGAAAGCCTAGTCCCGACTACGAAACACATGGTCACTCGTATCCTCT
CCGAAGCAATCTTGTGCAACGTCTCCCAACAAAACCCAAAACAACAACAAACA
CCAAGGGAGACCACACATCTATGTTTCGACGTAGACAGGCTTAACGAAAGGCT
GGAAAGAACGCTTCTAGCAAGCGCTAGCAAAGAGTAGCAGAGCCTGCCTGTTA
CAAGGCGGTGGCCTCCGTATACATTCGTGTGTACCTTCCTTGTCCACTTACACT
GTCTAAACAGCACATAGATAGTGCCGATCTATCAAGGAGTATGACAGAACGCG
TCGTACAAGCAGTAGGACCAAGAGGATCAGCCTGCGATTCTCAAGTTGTGTGG
AAAACTCTAACATAAGCTCGCAATGTTGGAGAAGGTGGAATAACTTACACAGT
CTAAAGCCTCCATAGATTGTGCCGATGTATCAAGGAGTCTGTGAGAACGCCAC
GGACAAGCAGTAAGCCCAACATGTTTAGCCTGCAAATATGAAGTAGTGTCGAG
AACTCATTAGTAAGCGAAGAACTAGTGAGAAGATGGAAGTTCGCTTTCATCAA
GACGCTACACCTTGCTGCAAGATGCATCAATATTTCACCAACACGCAACAACC
TGCTGCAAGATGGATCAAGGACGAGGATCCCATCGCCGAGCTGGCGCAGGTTG
AGCCAAGGTCTCCCGAAAACACAAACAAAACAAACACACAAAAAGGGAGATGA
ACAAACTGGGAAGAGACGTTTCTCAGTCACCGATGTCTATAACTCTTAAGGCG
TTCTCTTGAGAGCCGTTCGGGCACAAAGTAGTAGATTCGTCTGCTACAAGCAG
TCGGGAGAAGATCTATGTAACAGCACAACATGGATCAGTATATGACCGTAAAG
TTGGACAGAGACGTTGTTCAACGAGCCTGCCCAACAAGTAATGTCGCGTTCTG
ACGTTACAAGGACTGACGAGGGGCATGTTCAAAGCTCGCATGAACGTGCAAGA
AGTCGCAAAGGAGTGTGATAGAAGCCTCCACTATTACACGAGGGAGAGGAACC
CACTTGAGCGAGACGTGCTTAAGTACGCTCGTGACAAGAGACATCATGCGTTC
TATGGTGTCAGGCTTTGGATTTTCAGATGTAGAAAGCGAAGATCTATATCTTT
GATTGCGACCCTACATCGAGTTTCGACTTGATGCCGAACGTGTTTCGACATGT
TCACTCGCAATCTCTCCGAAGCAATCATGAGCAACCTCTTCCAATGCGGCTTCG
GGTCGGTCCCAGGGTAGGCGAGGTCGTATAAACCGGCTGCCAGCCTGAACCCG
ACGTCGGAACATCTCCCAAAACAAAACAACAAACAACAAAAACACAACCCAGA
CCAACACGTTTATCACGTCTCGATGAACAAGCAGTGACAAAACTGATGACTTC
GGTCGTCACAGGGATGGAGGAAGGTCGTCTCTAACAGCACAAGAGATGACGAC
GTGGGCCAGAAGGAGTATGAGTGGACCATGCTCGGCTGCTCAGCACACCGAGT

TTCACGTCTCGAAGCTCACGTCAGAGTCTTCCTCGGATCAAAAGCTCGCATTG
ATTCGATCGACATAGCACCCAGATGGTCAAAAGCCTCCATTGACTATCACGCT
AGCCCAAAGAGCATAAGAGTGGACCTTGTGCGACTCAGACCAACACAGTTGAG
GACGTCTCCCTTAACCCGCAGTGACACAAATCCGGTACAAAGCGAAGATGTAC
TGGACAGAGATGGAGGAAGGTTGTCTTTCGAGATAACGACGTGGGACAGAAG
GATGGTCAGTGGACGACTATCGGCTGATCAGTTCGCTGGTCAGAGGTCTCACG
AATCCGTCTCAAGTCGCTGCAATGGTTTGAGAGGGGTTAGCAAGCTGTGTCTC
ACTTTGACAAGCTGGGCAGAGGGCTCACGAATCCATCCCAAGTCGCTGCAAGG
TCTGGGAAGCTTCACTAGCTCAGATTCAGTAGACCGCTGTTG

Variants of the nanostructure E sequence are given below:

Nanostructure E-apt

For nanostructure E-apt, the green sequence above is replaced by the following sequence:

AGCTTCTGGACTGCGATGGGAGCACGAAACGTCGTGGCGCAATTGGGTGGGGA
AAGTCCTTAAAAGAGGGCCACCACAGAAGCT

Nanostructure E-FAM

For nanostructure E-FAM, the following sequence is introduced at the end of the template (i.e. immediately after the end of the black sequence, immediately before the final grey sequence):

AGCAGACGAAGC

The nanostructure is then incubated with a FAM-labelled DNA oligonucleotide with a complementary sequence:

GCTTCGTCTGCT/36-FAM/

A.1.2 Origami-based RNA Nanostructures

Origami$_{PX}$

AAGCTTCTAATACGACTCACTATAGGGAGAACTGAGTGGTCTGCTCGTGTAAT
CCCTCTCCCCGATTCGACTTCGGTCGACAGCTTCCGAGGGAACCATTGCAAGG
TATGAGGCTTCGGCCTTATACTTTGCGATGCGGGGGTGTCGAGGAACATTCGT
GTTTCTCGATACCCCGGGGATGAAGCGTCCTTTCGAGGATGCTTCGTCCCCCG
AGACCAAGCTCGCAGGTCTCGCGGCGACAAGGAGGCAGTCGCGTAATAAGTCC
ACAATTTTGCACGGGCGGACTACTCGGTAGCACCACGTGAAGTGGACACGCGT
GGTGGCGCAGGTGAAGCCTCCACGCCTGCCCTATTGAAGCGAGCACGATAGGG
TATTGTGACTGTGATATACGTAGCTGGGTTACCGGACGGGTGGGTAGTTCGTT
CGCGAACAGGGGGTGCGTCCGGTGACCCGGCTATGTATGTCACGGTCATAATA
CGCGCAGGCCTTCGGAGCGTTCCGATGTGCGTCGCGCCGCGTCGAGACGCTTC
GGCGTTTCGATGCGGTGCGATGCACGTCGGGACGCTCCGAGGGTTTGCCTTGG
GCCTGCCAGACTGTGGAGTGGTACATCTTCGGATGTCGGTGGAGATAGTCTGG

CGGGCCCATCTCCCGAGGAGCGAGTGGTCACATTCGTGTGGCCACTCGCTTCT
CCATATGCGGGTGGTGGATCGTATGGTGGTAACGGGGAAGCTGGTTTATTCGT
AAACATCGGGGACCGTTACCATCAACCTCCTCACGGTAGTCTCGGTTGCTTCG
GTAGCCGAGATTGCCGTGGGGAGCCCAGGCTGTCATTGGACTCTTATTCGTAG
GAGTCTAGTGGCAGCTTGCGTCCTGCGTCCTCCTGACTATTCGTAGTCCCCCT
CTGGCAGGACGGTGAGTCAAGCCTCCAGACTCACGGGGCCCAAGTGGACAGGG
CCGTACAAAGAGCCTATGTTATGATCCACCGCCTGCGTGTGGCGAGGTGGTGA
AAGGCTCACGCCACCTCCCCCCGTTGAAGTCCACACGGCGGGCCGCATGAAGG
AGGCACGTGCGGGGATGTTTCCGGACCCTTGGAACTTGTCTCTCCCACCCCCT
TAGTCTTCGGACTATACCCACCGGAGAGGCAGGTTCTAGGGGTCTGGAAACGT
CCGGCAAGATCACAGCTTGTTCCGGCGGACGTGGAAGCTGGTCTATTCGTAGA
CATCGGGGAATGTTCGCTGGAGCAAGTTGTGGTCTTGGCGACAGGCCGGACTC
CACCGCAGTATTCGTACTGACCACTCCTCTGGCTTGTCCATAACGCTGGGCTT
TTATTCGTAAGAGCTCAGCCGCTAGCTCGAAGGGGTCGATTAGCGGGCTCCCC
CTGGTGTTTCGATACCTACCCACCCCCGCTTCCCGGGTAAGAGCTCTCTGTAAC
ACGAGTGTTCGCACTTGTGTTGCAGGGAGCTTTTGCTCGGCGGCTCAGCATAG
GGGTGTAGAAGACCATTTCGATGGTTTTCTATACCTCTATGTTGGGCGTTCCT
AGGACTGGCAGAGGGGTCGCATTCGTGTGAAGGAGGACCCAGTCCTGGGAACC
CTGATCAAGTGGACAGATCAGGCGGCCACAAAGGCTCAGTGGCGATAAAAGCG
AGCATTAAATCGGCCTCTTCGGGCTGGTGGCGACCGCCTGAAGCTCGCACGGG
CGGTCCCCCACGTGAAGAGCCTACGCGTGGCCGGGTGAAGTCCACACGCCCGG
GGGATTGATGGTCTAAATCAATAGGGAGGGGAGTGGTACGGAGTGGTACATC
TTCGGATGTCGGTGGAGGTACCGCTCTCCTCTCTATTGGTTTGGACCATTAAT
CCCGGCATAGTTCTGTCATAGGTGGTGTCCCCGATCTACGTTCGCGTAGCAGC
TTCCTACCACTTATGATAGAATTATGGCGATGTAGTGTTTGGGTCCAGAGGGG
AAGCGTTCGCGCTTAGGAGGACGGCCCAGACGTTACATCCTTTTAGCAGTATT
CGTACTGCCAGATCTTTGCTTAGAGTGGTATGGTATACGCTTGGTGGGTACGC
TCTTCGGAGCGAGGGGGAGGAGCGTGTACCAACCCGGTGTCGTTAGCTGGCTG
GTCTTATCTCGAGATTTCGATTTCGAGGTGGGATCGGTCGGCTAATGATACCG
GCGGACTTTCGTGGCGCTAGCAGAGTTGATCTGGATTTCGATCTAGATCGATT
CTGTTAGTGCCGTGAAGGTCGGTAGTGTTCGGGCTGACTTGTGTCTACGATGA
TTCGTCGTCGTGGATATAGGTCGGCTTGAATACTATCCGGGGCCAAGAGCCTA
GGCCCCGCGGATCCAAGCTCGCAGGATCGTAACAAGGAGGCAGTTTACCATTC
TAGGCGGAGATCTGGCGGACTAATGAAGCCTCCACGTTAGTCCCCCCTCGTGA
AGCGAGCACGCGAGGCTCATCGAAAGGCTCACGATGAGGTGGTCTACGGTGTG
TTAAGACGTTTGTGCAGACTCCACCGCAGTATTCGTACTGACCACTCCTCTGC
GCAAATGTTTTAACGCACTGTAGATCACGGCAGTTCGTGTCAGTGCTTAGCTT
GCGGAGTTCGCTCTGTAAGCTAGGTATTGACACGGATTGGCGCGAGTTCCGAG
TCCTCCTTCAGCTTCGGTTGACCCCTCTGTCGGAACTCGCCACATCTTCGGATG
TGTAAAAGGTTATGAATTC

origami$_{BLKL}$-ssl

TGCATGATCTACGTGCGTCACATGCAGTACGAATTCTAATACGACTCACTATA
GGGAGAGCGTATGTACTTTATGACCCTTGAAGTTGACGCAAGTGGACAGTGTC
GACTTTAAGGCGCAATTTGAGTGAGAATTCGTTCTTACTCAGATTGGGCGTTT
AGTTTACTGCAACGTGGTAGTAGTGAACTGAACGGTGCTACGTATAGGAATTT
CGATTTCTATGCGTGGCACGTGATGGTTGCAAGGAGGCAGTAACCGTCACCCG
ATATGTAGGCAAGCAGGCAGTCTACGTATCCGGCTAGTGACGCAAGCGAGCAG
TGTCATTAGCGTCGTAAGGTATATATGCTTTCCCGATGGAATGCGCAGAGGTA
CGCAAGCTCGCAGTGTACTTCTGGGCTCGTCTGGGCAAGCCTGCAGTCCAGGC
GAGGACATATTCTGGCAAGCCTCCAGTCGGAATGTGTCGCGTGCTGTCCCAGC
GCAACGAAACAGTGCTACTCAAATCCACCGGGATAGCATGCCCGTGGAGGTGA
ATATTGTGATTCGTTACAGTATTTACCTTCACCGGAGTTCTGAACCTGCGCAA
CACTACAGTGCAACTCAAATCCACCGGTTTAGAGCTCGAAGTTCTGGGGACGG
ATTCGTTCGCGGATCTGTCCTCAGAGCTTCCATTCTATCTCTTACACTTCATCA
GGGAGACGGGAGATTCAGGGTCACCTGATGCTGCGCAAGTCCACAGTGCAACT
CAAATCCACCGCATCAGCGCAGACGTCGTAGGGCTTCGGTCCTGCGACGTTTG
GGCAGCGGATGCGCAAACCACGAGTGCAACTCAAATCCACCTCCGCTGGGATT
AGTTTTGAGGGCTTCGGTCCTCAGAACTGATCCGATCGCCTTGCAAGGAGGCA
GTAAGGTGATCCCGATCTATTGGCAAGCAGGCAGTCAATGGATCCGGTATCCG
AGGCAAGCGAGCAGTCTCGGATGCGTGACTCTGAGTCTTCCGTTTCCCGACTT
ATGGCGCACCAGTTCGCAAGCTCGCAGTGAATTGGTGGGCTGGGTCTCGCAAG
CCTGCAGTGAGGCCCAGGGACGGATAACGCAAGCCTCCAGTGTTATTCGTCCG
ACAATGATGCTGTTAAGGCAAGTTTCGAGTCTTAATAGCATTATTGTCCCGAT
TCGTAGCTAAGCGCAATTCGTTGTGCTTAGCTATGAATCCGGTGACAATTAAT
GTGAGGGCAAGTAGTGAGTCCTCGCATTAGTTGTTACGATCAGTGCGAACCAG
CGCAATTCGTTGTGCTGGTTCGTATTGATCCCATGAGTCAAGCTTCACTAGCT
CAGATTCAGTAGACCGCTGTTG

origami$_{BLKL}$-5′

TGCATGATCTACGTGCGTCACATGCAGTACCATATGCTAATACGACTCACTAT
AGGGAGACGTAGCGTATGTTTAGCGGGAGATCATCAGTTCTATTTGGCCACGC
GATGCGCAAGTCCACAGTGCAACTCAAATCCACCTCGTGTGCGCTGGTTAGTT
GTGTCTTCGGACGCAATTAGCCAGGGCCAGTGAACGGCAACTTCGCAGTCGTA
CTCAAATCCACCTCGCTGGGCTGGATCGACTGTATCTTCGGATGCAGTTGATC
TAGCGATAGGATGGCAAGATAGGAGTCATTCTGTCCCGTTTGTTCCGCAAGTC
ACCAGTGGAGCAGACCGGCTCTGTCTGCAAGCGTTCAGTAGACGGAGCGCTAG
ATAGGACTGGTGATTTCCTGCTAGCAGGCGCAACTCGAGGCAAGAACGCAGTC
TCGGGTTGGGCTGTGGTGTGCAAGGTGACAGTACACTACAGGGCATTTCATGG
CAACCTATCAGTCATGGAATGTCGTTAATATTGCTCTGCTGGCAAGTTTCGAG
TCAGTAGAGTAATGTTGACCCGACGGTCTATTCACTGTATTTCGATACGGTGA
GTAGATCGTCCGGCATGTTGGAATCGAACGGCAAGTAGTGAGTCGTTTGATTT

CAACGTGCGATGTATACATGACGTTGTCGTTCGCGACGACGTTATGTGTACGT
CCCTGCGTTTATGACCCTTGAAGTTGACGCAAGTGGACAGTGTCGACTTTAAG
GCGCAATTTGAGTGAGAATTCGTTCTTACTCAGATTGGGCGTTTAGTTTACTG
CAAGCGAAGAGTAGTGAACTGAACGGTGCTACGTATAGGAATTTCGATTTCTA
TGCGTGGCACGTGATGGAGGCAAGGAGGCAGTCTCCGTCACCCGATATGTAGG
CAAGCAGGCAGTCTACGTATCCGGCTAGTGACGCAAGCGAGCAGTGTCATTAG
CGTCGTAAGCAATATATGCTATGTCTTCCGGGCGCGCAGAGGTAGGCAAGCTC
GCAGTCTACTTCTGGGCCAGTCTGGGCAAGCCTGCAGTCCAGGCTGGGACATA
TTCTGGCAAGCCTCCAGTCGGAATGTGTCGCGTGCTGTCCCAGCGCAACGAAA
CAGTGCTACTCAAATCCACCGGGATAGCATGCCCGTGGAGGTGAATATTGTGA
TTCGTTACAGTATTTACCTTCACCGGAGTTCTGAACCTGCGCAACACTACAGT
GCAACTCAAATCCACCGGTTTAGAGCTCGAAGTTCTGGGGACGGATTCGTTCG
CGGATCTGTCCTCAGAGCTTCCGCCCTTTTAGCAGACGAAGCAAGCTTCACTA
GCTCAGATTCAGTAGACCGCTGTTG

origami$_{\mathbf{BLKL}}$-3′

TGCATGATCTACGTGCGTCACATGCAGTACCATATGCTAATACGACTCACTAT
AGGGAGAGGCCACGCGATGCGCAAGTCCACAGTGCAACTCAAATCCACCTCGT
GTGCGCTGGTTAGTTGTGTCTTCGGACGCAATTAGCCAGGGCCAGTGAACGGC
AACTTCGCAGTCGTACTCAAATCCACCTCGCTGGGCTGGATCGACTGTATCTT
CGGATGCAGTTGATCTAGCGATAGGATGGCAAGATAGGAGTCATTCTGTCCCG
TTTGTTCCGCAAGTCACCAGTGGAGCAGACCGGCTCTGTCTGCAAGCGTTCAG
TAGACGGAGCGCTTCTCCCGACTGGTGATTTCCTGCTGGCAGGCGCAACTCGA
GGCAAGAACGCAGTCTCGGGTTGGGCTGTGGTGTGCAAGGTGACAGTACACTA
CAGGGCATTTCATGGCAACCTATCAGTCATGGAATGTCGTTAATATTGCTCTG
CTGGCAAGTTTCGAGTCAGTAGAGTAATGTTGACCCGACGGTCTATTCACTGT
ATTTCGATACGGTGAGTAGATCGTCCGGCATGTTGGAATCGAACGGCAAGTAG
TGAGTCGTTTGATTTCAACGTGCGATGTATACATGACGTTGTCGTTCGCGACG
ACGTTATGTGTACGTCCCTGCGTTTATGACCCTTGAAGTTGACGCAAGTGGAC
AGTGTCGACTTTAAGGCGCAATTTGAGTGAGAATTCGTTCTTACTCAGATTGG
GCGTTTAGTTTACTGCAAGCGAAGAGTAGTGAACTGAACGGTGCTACGTATAG
GAATTTCGATTTCTATGCGTGGCACGTGATGGAGGCAAGGAGGCAGTCTCCGT
CACCCGATATGTAGGCAAGCAGGCAGTCTACGTATCCGGCTAGTGACGCAAGC
GAGCAGTGTCATTAGCGTCGTAAGCCCTATATGCTATGTCTTATAGGTGCGCA
GAGGTAGGCAAGCTCGCAGTCTACTTCTGGGCCAGTCTGGGCAAGCCTGCAGT
CCAGGCTGGGACATATTCTGGCAAGCCTCCAGTCGGAATGTGTCGCGTGCTGT
CCCAGCGCAACGAAACAGTGCTACTCAAATCCACCGGGATAGCATGCCCGTGG
AGGTGAATATTGTGATTCGTTACAGTATTTACCTTCACCGGAGTTCTGAACCT
GCGCAACACTACAGTGCAACTCAAATCCACCGGTTTAGAGCTCGAAGTTCTGG
GGACGGATTCGTTCGCGGATCTGTCCTCAGAGCTTCCACCTGTAAGACGTAGC
GTATGGGCAGCGGGAGATCATCAGTCTTTTAGCAGACGAAGCAAGCTTCACTA
GCTCAGATTCAGTAGACCGCTGTTG

origami$_{BLKL}$-N

TGCATGATCTACGTGCGTCACATGCAGTACCATATGCTAATACGACTCACTAT
AGGGAGAGAGATCATCAGTTCTATTTGGCCACGCGATGCGCAAGTCCACAGTG
CAACTCAAATCCACCTCGTGTGCGCTGGTTAGTTGTGTCTTCGGACGCAATTA
GCCAGGGCCAGTGAACGGCAACTTCGCAGTCGTACTCAAATCCACCTCGCTGG
GCTGGATCGACTGTATCTTCGGATGCAGTTGATCTAGCGATAGGATGGCAAGA
TAGGAGTCATTCTGTCCCGTTTGTTCCGCAAGTCACCAGTGGAGCAGACCGGC
TCTGTCTGCAAGCGTTCAGTAGACGGAGCGCTAGATAGGACTGGTGATTTCTC
TCCCCCCGGCGCAACTCGAGGCAAGAACGCAGTCTCGGGTTGGGCAGTGGTGT
GCAAGGTGACAGTACACTACTGGGCATTTCATGGCAACCTATCAGTCATGGAA
TGTCGTTAATATTGCTCTGCTGGCAAGTTTCGAGTCAGTAGAGTAATGTTGAC
CCGACGGTCTATTCACTGTATTTCGATACGGTGAGTAGATCGTCCGGCATGTT
GGAATCGAACGGCAAGTAGTGAGTCGTTTGATTTCAACGTGCGATGTATACAT
GACGTTGTCGTTCGCGACGACGTTATGTGTACGTCCCGGGCTCGATGACCCTT
GAAGTTGACGCAAGTGGACAGTGTCGACTTTAAGGCGCAATTTGAGTGAGAAT
TCGTTCTTACTCAGATTGGGCGTTTAGTTTACTGCAAGCGAAGAGTAGTGAAC
TGAACGGTGCTACGTATAGGAATTTCGATTTCTATGCGTGGCACGTGATGGAG
GCAAGGAGGCAGTCTCCGTCACCCGATATGTAGGCAAGCAGGCAGTCTACGTA
TCCGGCTAGTGACGCAAGCGAGCAGTGTCATTAGCGTCGTCGGGGGTATATGC
TATGTCTTATAGGTGCGCAGAGGTAGGCAAGCTCGCAGTCTACTTCTGGGCCA
GTCTGGGCAAGCCTGCAGTCCAGGCTGGGACATATTCTGGCAAGCCTCCAGTC
GGAATGTGTCGCGTGCTGTCCCAGCGCAACGAAACAGTGCTACTCAAATCCAC
CGGGATAGCATGCCCGTGGAGGTGAATATTGTGATTCGTTACAGTATTTACCT
TCACCGGAGTTCTGAACCTGCGCAACACTACAGTGCAACTCAAATCCACCGGT
TTAGAGCTCGAAGTTCTGGGGACGGATTCGTTCGCGGATCTGTCCTCAGAGCT
TCCACCTGTAAGACGTAGCGTATACCAAAAAGCAGACGAAGCAAGCTTCACTA
GCTCAGATTCAGTAGACCGCTGTTG

A.2 Oligonucleotide Transcription Templates

A.2.1 Aptamer Sequences

GA

Strand 1: CTAATACGACTCACTATAGGGAGCTTCTGGACTGCGATGGGAGCACG
AAACGTCGTGGCGCAATTGGGTGGGGAAAGTCCTTAAAAGAGGGCCACCACAG
AAGCTCCC

Strand 2: GGGAGCTTCTGTGGTGGCCCTCTTTTAAGGACTTTCCCCACCCAATT
GCGCCACGACGTTTCGTGCTCCCATCGCAGTCCAGAAGCTCCCTATAGTGAGT
CGTATTAG

GA-O

Strand 1: CTAATACGACTCACTATAGGGAGAAGCTTCTGGACTGCGATGGGAGC
ACGAAACGTCGTGGCGCAATTGGGTGGGGAAAGTCCTTAAAAGAGGGCCACCA
CAGAAGCTTCTCCCGCCGCCGGCCGG

Strand 2: CCGGCCGGCGGCGGGAGAAGCTTCTGTGGTGGCCCTCTTTTAAGGAC
TTTCCCCACCCAATTGCGCCACGACGTTTCGTGCTCCCATCGCAGTCCAGAAGC
TTCTCCCTATAGTGAGTCGTATTAG

A.2.2 PX_{major}

PX1a

Strand 1: CTAATACGACTCACTATAGTCAGAGTCCCCGATTCGCTTCGGCGACA
GCTTCCCTCTGAC

Strand 2: GTCAGAGGGAAGCTGTCGCCGAAGCGAATCGGGGACTCTGACTATAG
TGAGTCGTATTAG

PX1b

Strand 1: CTAATACGACTCACTATAGGGGTCCGGAAGCTGGTTATTCGTAACAT
CGGGGAGGACCCC

Strand 2: GGGGTCCTCCCCGATGTTACGAATAACCAGCTTCCGGACCCCTATAG
TGAGTCGTATTAG

A.3 Origami Sequences

For origami staples, both the staple sequences (Tables A.1, A.2, A.3 and A.4), staple mixes and the layout of staples with respect to the scaffold (Figures A.1, A.2, A.3, A.4) are provided

for each nanostructure.

A.3.1 Wedge Origami

Staple name	Sequence
6[229]28[210]	TTTTGTTGATATACCTCATAGAATTAATACAAAACAGTATAA AGCCAA
6[229]28[210]	TTTTGTTGATATACCTCATAGAATTAATACAAAACAGTATAA AGCCAA
10[124]32[105]	TTTTCCGCCACCATCTGAATAATTATCCAATCAAGCAAAAGA AGATGA
50[75]48[84]	TTTTGGTCGAGGTACGGGGAAAGCCGGCGAACGTGGCGAGAA TAGGGT
23[154]49[173]	TACAATTCATGTAAAAGTCCTTTAATGCCGCGTAACCACCAC ACTTTT
30[229]47[237]	TTTTTCCCTTAGAAATGGTTTCGGCCTAGTGAGGCCACCGAG TAAAAG
40[120]17[104]	TTTTTCAATAGATAATACATTTGATCATCAAATATAACGGAA AC
44[225]21[209]	TTTTCGTGGCACAGACAATATTTAAATCGTAAAACATTCTTA CC
34[124]43[125]	TTTTATGATGGCAATATACAAAATCAAACCGCCTGCAACAG
28[281]23[288]	TTTTCATAATTACTAGTTTTTGTTTAACGTCAAAATTTT
22[288]5[281]	TTTTATGAAAATAGCAGCCTTTAAGTAAATGAATTTTTT
4[281]5[258]	TTTTTTCTGTATGGGATTTTGCTAAACAACTCTTTCC
20[225]6[203]	TTTTAAGAAAAGTAAGCAGATAGGGCGGATCCCGGAA
16[120]10[98]	TTTTACCATCGATAGCAGCACCGCCTCAGACACCAGA
29[210]31[229]	TAAATTTATCCTTGCGCTATTAATTAATTTTTTT
13[49]15[68]	TTATTAGATCGGCATAAAATGTTACTGTAGTTTT
29[259]47[278]	TAAGAATTTAGTAAAACCGTTGTAGCAATATTTT
25[56]1[75]	CAAGGATCCATTAAGCAAAAGAATTTCATGTTTT
33[105]35[124]	CAGATGAATTCATCTATTCCTGATTATCAGTTTT
8[176]8[147]	TTTTTAATAAGTTTTAACGGGGTCAGTGCC
19[105]17[120]	TTTGTCAACCGTCAGTCACCAATGAATTTT
2[176]2[147]	TTTTTTGCAGGGAGTTAAAGGCCGCTTTTG
12[71]11[55]	TTTTTTTCATAATCAAAATCACCCCACCAC
23[210]21[225]	TGCCAGTCTGAACAGAAGCCCTTTTTTTTT
18[173]9[176]	TTTTCCGATTGAGGGAGTGTACTGGTTTT
38[68]37[71]	TTTTAAATCCTTTTATTAATTTTAATTTT
24[173]3[176]	TTTTGGCTTATCCGGTCGCTGAGGCTTTT
43[106]41[120]	ATTAACCAGTTGAAGATTAGAGCCGTTTT
42[173]33[176]	TTTTCCTTGCTGACCTGATTGCTTTTTTT
46[278]29[281]	TTTTCTTCTTTGAAAACACCGGAATTTTT
32[176]19[173]	TTTTGAATACCAAGTTGACATTCAATTTT
10[111]11[124]	GAACCACGCCACCACCCTCAGAGTTTT

Continued on next page...

Table A.1 – continued from previous page

Staple name	Sequence
50[62]51[75]	GCCGTAAAAATCAAGTTTTTTGGTTTT
6[216]7[229]	AGTATAGAAGTGCCGTCGAGAGGTTTT
42[160]43[173]	ACCTCAAAAAATCTAAAGCATCATTTT
39[42]39[68]	TCGCCATTCAGGCTGCTCGTATTTTTT
46[216]45[225]	TGCTGGTAGCGTAAGAATATTTT
25[49]50[26]	TTCAACGAATGTGTATCACCCAGCACTAAATCGGAACCCTTCTATCATTTT
43[19]40[42]	TTTTCATGCCTGCAGGTCGACTCTAGAGGATAAAACGCGAAAGGGGGAAGG
1[26]5[48]	TTTTGAAGGCACCAACCTAAAACTCTTTGATACAACGGAACCGAGTGTACA
30[41]4[22]	TTAAAATGAAGCAACAAATGGATGAACGACTGACCAACTTTGAATTTT
28[41]2[22]	ATCAGGTTCATTTGAAAAACAAACAAAGCCCCCAGCGATTATACTTTT
37[47]38[19]	AGCCAACGTGCCCGAGCTTTCCGGCACCGCTTCTGGTGCCGGTTTT
49[19]47[48]	TTTTAGGGCGAAAAACCGAAAGGGAGCCCCTTAAAGAGCTGGTT
48[41]0[26]	ACGTGGAAGACAGTGATTCAAGGAGAAGTAATGCCACTACTTTT
51[26]27[48]	TTTTATGGCCCACTACGTGAACCAGGTAAACAAATCA
18[41]8[22]	CAGGTCTTAATGCAGAGATTTAGGAATACCTTTT
9[42]10[22]	GATACATGGAATTACGAGGCATAGTAAGAGTTTT
39[19]40[35]	TTTTAAACCAGGCAAAGCGCCATGCGATCG
13[22]14[19]	TTTTAAAAACCAAGGCTTTTGCAAATTTT
31[22]44[19]	TTTTATTTTTGTTGGGGTGCCTAATTTTT
21[19]30[22]	TTTTCGAACCAGACCGTCGCATTAATTTT
19[19]32[22]	TTTTTTATAGTCAGAAGAACAAACGTTTT
7[22]20[19]	TTTTTAATTTCAACGAGCTTCAAAGTTTT
9[22]18[19]	TTTTACATTCAACTTACCCTGACTATTTT
33[22]42[19]	TTTTGCGGATTGAGTGCCAAGCTTGTTTT
11[22]16[19]	TTTTCAACACTATGCGGAATCGTCATTTT
17[19]34[22]	TTTTTAAATATTCATTCGTTGGTGTTTTT
27[22]48[19]	TTTTAAAGGCCGGCTCCAACGTCAATTTT
29[22]46[19]	TTTTAGAGTCTGGGGCCCTGAGAGATTTT
5[22]22[19]	TTTTAGAGGACAGTCAATAACCTGTTTTT
23[19]28[22]	TTTTTTAGCTATATTTCATTGCCTGTTTT
35[22]40[19]	TTTTAGATGGGCGGTGCGGGCCTCTTTTT
12[48]12[22]	GGAACCGTTTACCAGACGACGATTTTT
37[22]36[35]	TTTTGCCTCAGGAAGATCGCGACGACG
6[48]6[22]	AGTAAATTGGGCTTGAGATGGTTTTTT
45[19]45[41]	TTTTGAGTGAGCTAACTCACATT
16[97]43[105]	TAATCAGGCAAGGCTCCTGATTTCAGGTTGAGGAAGTTTCCTGTGTCAGT
25[76]20[84]	ACCCTCACCGCACTCCAAAATTTAGCATTCCCAATTTTTGATAT

Continued on next page...

Table A.1 – continued from previous page

Staple name	Sequence
	ACATAC
47[104]25[125]	TTATAAAACGTGCTTTCCTCGTTTTGACGCTAATGCGGCTGTCT CATCGT
18[62]45[55]	GAATGACCATAAATATCAAAAAACCCGTTTTAACCGAAGCATTT GCGCT
8[55]29[62]	AAGATTCAATTACCTTCAAATCAGGTCAACGTTAATCGATGAAC GGTAA
28[97]33[90]	GTAAAGTAATTCTACATATGTCAGGAAGGTCTGGCCCAGCTTGT AGATT
4[153]26[133]	AAAGGAGCCTTTAATTGTATCCCATCGCGTTTTAGATTACCGAC CAATC
21[196]47[216]	ATAGCTAAGCGATATTCATCTAGAACAACTGAGAAGTGTTTTTA TAATC
6[181]28[161]	GAGGTTTCACCAGTCGCTAATACCAACGCGCCATATTTAACAAC GCCAA
4[69]9[76]	CGCAGACGCTGGCTACGAGAAATTTTAATTATTACATTCACAAA CAAAT
6[90]11[97]	CAGTGAATACCAGTAAACGAACATTAAACCAGCATCTCAGAACC GCCAC
46[153]33[146]	TACCTACGTCACACGCCATTACAAATGAATATCAAACCCTCAGG AGAAA
30[132]43[139]	TACCTTTCAATTTCTTTCAATTACCTTTTACATCGATCAATACT GAGAG
18[139]45[146]	GGAAATTGGTTTACTTATTCAATTTGAACGAACGATGGCAGATT CACCA
46[146]3[153]	ATTTTGAATCCAATAGAGGCAATTTTGCGCGAGGCCCACGCATA ACCGA
33[50]6[56]	TTATTTGCACGTAGAGTAACAGATTAAGAAAGACTTATGCGACA CCAG
43[73]22[77]	GTAATCATGGTCTCACAAAATTCGCATTGTATTGCTCCTCTGCG AA
17[72]40[64]	AGTAGTTCTGAAAATAAAGCGCCAGGGTTTTCCGGCGATTCAAT TC
5[70]48[67]	GACCTTCCGAGTAGTCTACTAAGAGGGTGTTCTAGTGTTCCAGT T
19[42]14[39]	GGATTGCCAAAAATCCTCAAATAGCGTCGTAATAGTTTTCGGTC C
5[238]29[238]	CTAAAGTAGGGAAGATTTATCATGCGTTATACAAACCGACCGT
5[161]25[160]	ACACTGAAAAAAAAGGCTCCATATATTCGGTATTCCAAATCA
30[181]47[195]	GCTGAGATTCAAATATTGCAACAGGAACGGTACGCCAGAATC
9[147]31[160]	CTTTTGATGATACAGGAGGGAAAAGGGCACAAAATGGAAACA
4[195]20[196]	AATAATTACAACGCATCACCGTTTGCTCAGTACCACCGAACA
28[48]32[35]	AAAGGCTAAGAGAATATTTTGCTCATTTCGGATTCTCCGTGG
32[146]6[133]	AGGCGAACAGCGCCAAGAACTGTATTAATCAGAACCGCCACC

Continued on next page...

Table A.1 – continued from previous page

Staple name	Sequence
21[147]47[160]	CAACATTGAGTTAAATTTATCCAAAGAAATGGAAAGCCGATT
7[98]30[112]	TTGGGAAGACTATTTCGGAACTACGCAGCTTAGAGCCGGCTT
49[126]46[112]	GGCGCTGTGGTTGCTAGAATCTCGTCTGAAATGGGGAGAGGC
24[125]3[111]	CCCGACTGACAATGCGAAAGACAGCATCGGAACGACCGATAG
44[125]20[126]	GCAGAAGCATTTAATTAACCTCTTAATTGCTGAATTTAAGAC
5[126]9[139]	GGGATAGCTCAGAGTCTGAAACCGTATATCCAGTAAGCGTCA
6[202]28[189]	TAGGTGTCTGTAGCCAAAGTCTCCAGAGCAGTAGGGCTTAAT
22[62]24[49]	TTGACCACGCATAGGGTCAATCATAAGGGAGATTTAAATCGG
48[139]47[153]	CGTACTAGCAAGTGTAGCGGTCACGCTGGCCGCTAACAGGAG
22[251]5[237]	TAAAAACTTTGTCGTTTCAACAGTTTCAGCGGAGTTAACGAT
48[66]43[72]	TGGAAAAATCCTGAGACGCGCTTTCCAACATAGCTCGAATTC
46[111]44[84]	GGTTTGCAATCGGCAGAGGTGAGGCGGTGAAATTGTTATCCG
6[55]20[42]	AACGAGTGACCAGGTTAGATACATTTCGACTCCAAATCGCGT
45[196]23[209]	CCTTCTGACCTGAAAATATCCTCTGACCCGCTCAACCTAATT
32[83]6[70]	TCATCAAACGCAAAATAAAGGCTCATTATAAGGCTTGCCCTG
44[174]22[161]	AACTGATAGCCCTAGTACATAATAGTGAGCCCAATAGATAAC
10[97]20[98]	GCCGCCGGCCAGAAGAGCCATCGGAATAAGTTTATGCAAACG
22[195]6[182]	AGAGGGTAATAGCAAAGTTACCAGAAGGGCGGGGTTACTCAG
45[127]33[125]	TTACATACCACCATGCCACGTCTGGTCAGTTGGCGTAACAG
29[239]46[252]	GTGATATGAGTAGAGTCTGTCCATCACGCAAATTTAACATC
21[91]47[103]	AGAGGTCCCAAAAAACCCCGGGCGCCAGTCGGCAAAATCCC
18[83]20[72]	TTAGAGCAAATCCTCTAACGGAACAACAGAACTGGTGGCA
28[83]2[72]	ATGCCGGATAGTAGAAGCAATTGTCGAACTTTGAGGACTA
23[115]48[98]	AAGCAATATAATGTTCAGAGCACGTATATCAAAAGAATA
35[70]40[87]	CAAAGAATTACAAAAAGTTGGGTTATCTTTAGGAGAGAA
26[73]30[63]	TAAATGCAATGTTCAACCAGCTATTTCGTAAATATTTAA
24[97]26[74]	GAATTAGATCGAGAGGGTATTAAACCAAGTATATATTT
43[53]21[62]	CGGGTACCGACGAGCCGAATAGGAATTGTAAGGATTAG
18[125]8[102]	AAAGGTGTTACCGTAACAGTTAATGCCCCCTGCAAAAT
30[188]4[179]	TTAAGACACAATGAAATTGAGACAAACTTTTTCACGTT
8[142]30[133]	GTAACAGTGCCATGAAAGGCATGAATAATGCGAGAGAC
21[70]47[79]	CTTTAATAAGCAAAACTAGCATCACCAGTGTTTGATGG
2[71]0[49]	AAGACTTTACACTAAAACACTCAGAAAGAGACGGGTA
23[231]4[223]	CCATATTCGCATTAGACGGGAGTTAGCGGAGAATAG
46[237]23[230]	AAGAACTCAAACTATGAAATATTCTTACTAAACAG
21[98]4[91]	ATTTTTGTATAACAAACCGGATTCTTAAACCTCCA
4[118]24[98]	CTTTCGATTGCGCCTGCGGGAGGAGGCAAGGCAAA
46[195]31[188]	TATTACCATAGAACGCTATTAGTCTTTAAGTGAAT
0[48]23[55]	AAATACGCCTTTATTTGTACCGGGCGCGAGCTGAA
46[188]23[195]	GCCAGCCATATTTTTGAGAATCTAACGAGCGTCTT
5[259]28[252]	AGACGTTCAGAGAGAAACGATAAAAAGCCTGTTTA
2[97]5[90]	AACGGCTACAGAGGATCCGCGTGTTACTGTAATCT
26[160]29[153]	CTAATTTACGAGCAAATAGATTTTAGGCCGCAAGA
29[189]44[196]	AGTTAATGCTTAGAAACCTTGCTTCTGTTTGAATG

Continued on next page...

Table A.1 – continued from previous page

Staple name	Sequence
5[133]28[126]	CAAGCCCGCAACTAAGTTGCTTTTTCGAGCCAGTA
28[62]3[69]	TTTGAGAAAGGTGGTAAAGCTGTATCATCGCCTGA
19[35]8[42]	GCAAAGCTTTAATTCTTTAATCATTGTGATCAGTT
28[55]49[62]	GATCTACCCATCAATCCACTACGATTTAGAGCTTG
25[112]28[105]	TTTATTTTTTCCTTATCATTCAACAACAAGTACCG
41[91]32[98]	CACTAACAACTAATAAGGAATTTAACGTTGAAACA
30[153]5[146]	AAAATCATAGGTCTTGTAGCTTAAATATAATAGGA
23[161]20[168]	TTATCCTGAATCTTATCAGAGAATAAGAGGAAACG
47[49]26[56]	TGCCCCAGCAGGCGACAAGAGTATGATACCTGAGT
2[146]4[126]	CGGGATCGTCACCCTCAGCAGACAACAAGGTTTAT
25[126]22[119]	AGGAATCCGAACCTCAAGATTAAGTACGGTGTCTG
24[76]21[69]	AAAGCCTCAGAGCACATCAATATTTAGTAGAGTAC
32[76]35[69]	CATTAAATGTGAGCAAACAGATAATGGATGCGGAA
16[69]34[50]	CTTTAGCGTCAGTAGACTGGATGCTTTAGAACCT
6[101]23[97]	AGCTGCTCATTTGACAAGGTTGATTAACATCC
47[137]26[140]	CGGGAGCTAACAGGGCGTATCAACTGTAGAA
19[84]16[81]	GACACCATTGGGAAACCATTATAGCGACAGA
40[86]17[83]	GTATTAGACTACCACCAATTATACCACCATT
31[112]45[126]	TTAATTAATAAAACCAACGCGCGGATTAT
25[140]23[153]	CGCCCAATAGCAAGTAAGAACACCCAGC
29[161]31[174]	CGCGAGAAAACTTTAGAGTCAAATCAAT
44[160]46[161]	AAACATCGACCAGTAATAAAAAACGCTC
20[167]6[154]	CAATAATTCAAGAGCCACCCTCAGAACC
43[140]31[153]	CCAGCAGAAAATACTTACCTTTTTTAAT
22[118]20[105]	GAAGTTTCATTCCACGGATGGTATGTTA
7[168]21[181]	AAGGATTAGGATTAAAACCGAGCAAGAA
45[70]29[83]	CAGTCGGGAAACCTTTTCTTTTGTCAAT
27[98]25[111]	GACAATACAAGAACACAAGCAAGCCGTT
32[97]31[111]	AATGTAGCTTCCCATCAAGAAAACAAAA
35[49]17[62]	ATCTGTTATCATTTAGGGTTAAACAGTT
20[125]6[112]	TCCTTATCTATTATCCACCACCCAAATC
47[161]45[174]	AAAGGGATTTTAGACAGGAAAGGGACAT
45[42]31[41]	AATTGCGAAAGTGTAAAGCCTAAATCAG
15[35]11[48]	AGAGGGGCAATACTCATAACCCTCAGAG
31[175]45[188]	ATATGTGATGCGCGTCTGGCCAACAGAG
43[84]33[83]	CATAGCTGGTTATCTAAAAAAAAATTGC
22[209]4[196]	CCCTGAAATTCCACAAGGAATTGCGAAT
5[147]7[160]	ACCCATGGCCACCCGAGGCTGAGACTCC
34[90]40[91]	TGTTTGGGAAGGAGCGGAATTAGGATTT
28[251]22[252]	GTATCATCCAATCCAAATAAGAATAACA
29[112]23[125]	ATATAACTATATGTATAAGAGCTTAAAT
32[125]18[126]	TACCTGATAGAAAATTCATATATTCATT
5[77]3[76]	ATCAAGATAGCCGGAACGAGGTAAATTG
27[92]49[104]	GACGACGCCCGAGAAGGAAGGGAAGAA

Continued on next page...

Table A.1 – continued from previous page

Staple name	Sequence
10[68]16[70]	GCAGGTGCCTCCCTCAGAGCAGTTTGC
47[80]27[91]	TGGTTGAGTGTCTGATAAATTGTCCA
16[80]10[69]	ATCACGCCACCTGACAGGAGGTTGAG
35[35]37[46]	CATCGTAACCGTGCTTGAGGGACTCC
47[84]45[94]	TCCGAAAGGTGGTTGTCGTGCCAGC
18[153]8[143]	AGGTAAATATTGACTACATGGTTGA
8[101]20[91]	CTACGTTAATACAGGACGTAGAAAA
41[56]43[52]	GCTGCAACAGTCACGACGTTGTCCC
4[178]6[168]	GAAAATCTCCAGTTTCGTAGTACCG
29[126]47[136]	AAATGCTGATGCAACGCTCAAAGAG
28[104]23[114]	ACAAAAGAATAAATCATACTTTTG
4[111]6[102]	GGTGAATTATTCATAACGTAACAA
18[76]17[71]	CAGCAAAATAACGACAGAACACC
20[71]19[76]	ACATAGCCCGAGGAATAAAAGAA
29[103]30[98]	AATCGGGTTAGGTTAGAAAAGCC
45[95]29[102]	TGCATTAATGGTATTGGTTGAT
34[49]17[41]	ACCATAGATAGGTCAGAATCCC
40[63]41[55]	GACAACGCAACTGTTGGGATGT
42[41]33[49]	ACGGCCACCGTAATGGTCAAAA
32[153]33[160]	CGCGCAGCAATAACGGATTCG
46[251]29[258]	ACTTGCCAATAAGGCGTTAAA
20[160]19[153]	AACGGAATACCCAAAAAGACA
31[63]44[70]	ACGCCATCAAAAATTTCCACA
45[56]46[49]	CACTGCCGGCAACAGCTGATT
3[91]2[98]	ACCTGAGCTTGATAGGGTAGC
5[112]4[119]	TACCCTCATTTTCACAGCTTG
18[104]9[111]	CCGACTTTGGAAAGCGCAGTC
9[49]8[56]	AACGCGGCCTTGATAGGTAGA
46[48]29[41]	GCCCTTCACCGCCTAGCAAAC
49[105]49[125]	AGCGAAAGGAGCGGGCGCTAG
26[132]27[139]	AATAATCAGAACGCGCCTGTT
11[56]10[49]	CGGAACCCAGACGATTCAAAA
5[154]22[147]	TACCGTACCACAAGAATGTTT
4[222]5[216]	AAAGGAACAACTAAGACAGC
14[38]13[48]	GAGAAATAGATAGCCCCC
14[68]13[71]	TTTTCGCGTTTTCCGTTTGCCATCTTTTT
26[176]25[173]	TTTTTAATATCCCATCGATATAGAATTTT
48[173]27[176]	TTTTCCGCCGCGCGAACAAGAAAAATTTT
0[75]25[75]	TTTTAGGAAGTTTAAAAATTTTTAGA
36[71]36[49]	TTTTAAGTTTGAGTAACACCAGT
25[19]26[22]	TTTTAATACTTTTGCGAAGGGTGAGTTTT
3[22]24[19]	TTTTCAAGCGCGATTATGACCCTGTTTTT
15[19]36[22]	TTTTAGAAGTTTTGCCACAGTATCGTTTT
47[19]47[41]	TTTTGTTGCAGCAAGCGGTCCAC

Continued on next page...

Table A.1 – continued from previous page

Staple name	Sequence
41[19]41[41]	TTTTTCGCTATTACGCCAGCTGG
15[3]36[22]_FL	TGGTGGGTTGTTGGTTGTGAAGAAGTTTTGCCACAGTATCGTTTT
36[87]36[49]_FL	TGGTGGGTTGTTGGTTGTGAAAGTTTGAGTAACACCAGT
14[84]13[71]_FL	TGGTGGGTTGTTGGTTGTGACGCGTTTTCCGTTTGCCATCTTTTT
47[3]47[41]_FL	TGGTGGGTTGTTGGTTGTGAGTTGCAGCAAGCGGTCCAC
3[6]24[19]_FL	TGGTGGGTTGTTGGTTGTGACAAGCGCGATTATGACCCTGTTTTT
41[3]41[41]_FL	TGGTGGGTTGTTGGTTGTGATCGCTATTACGCCAGCTGG
0[91]25[75]_FL	TGGTGGGTTGTTGGTTGTGAAGGAAGTTTAAAAATTTTTAGA
26[192]25[173]_FL	TGGTGGGTTGTTGGTTGTGATAATATCCCATCGATATAGAATTTT
48[189]27[176]_FL	TGGTGGGTTGTTGGTTGTGACCGCCGCGCGAACAAGAAAAATTTT
25[3]26[22]_FL	TGGTGGGTTGTTGGTTGTGAAATACTTTTGCGAAGGGTGAGTTTT
cy3_label	/5Cy3/TCACAACCAACAACCCACCA

Table A.1: Staple sequences for wedge origami nanostructure. Colours relate to staple mixes (see below).

Staple mixes for origami variants (staples in 10-fold excess of scaffold):

Basic Mix: Black staples + red staples, for unmodified origami.

Cy3 Mix: Black staples + blue staples (with cy3_label in 20-fold excess over other staples, for cy3-labelled origami (used for electroporation experiments).

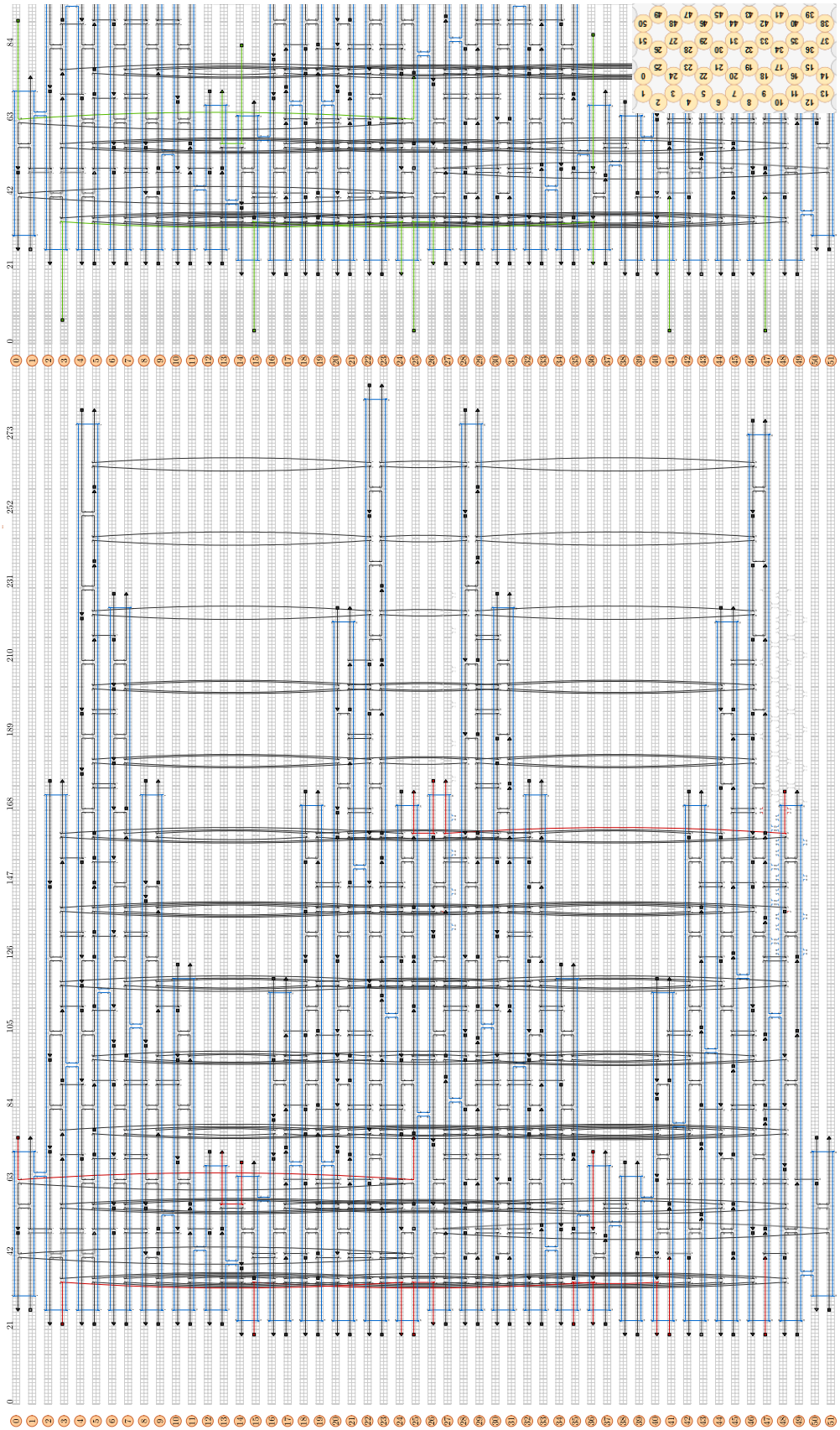


Figure A.1: Wedge origami layout from caDNAno. Main view (below) shows staples for basic nanostructure, and reduced view (above) shows nanostructure with extensions for hybridisation to fluorescent oligonucleotides. Inset shows arrangement of helices.

A.3.2 Pointing Hand Origami

Staple name	Sequence
6[229]28[210]	TTTTGTTGATATACCTCATAGAATTAATACAAAACAGTATAA AGCCAA
10[124]32[105]	TTTTCCGCCACCATCTGAATAATTATCCAATCAAGCAAAAGA AGATGA
50[75]48[84]	TTTTGGTCGAGGTACGGGGAAAGCCGGCGAACGTGGCGAGAA TAGGGT
23[154]49[173]	TACAATTCATGTAAAAGTCCTTTAATGCCGCGTAACCACCAC ACTTTT
30[229]47[237]	TTTTTCCCTTAGAAATGGTTTCGGCCTAGTGAGGCCACCGAG TAAAAG
40[120]17[104]	TTTTTCAATAGATAATACATTTGATCATCAAATATAACGGAA AC
44[225]21[209]	TTTTCGTGGCACAGACAATATTTAAATCGTAAAACATTCTTA CC
34[124]43[125]	TTTTATGATGGCAATATACAAAATCAAACCGCCTGCAACAG
28[281]23[288]	TTTTCATAATTACTAGTTTTTGTTTAACGTCAAAATTTT
22[288]5[281]	TTTTATGAAAATAGCAGCCTTTAAGTAAATGAATTTTTT
4[281]5[258]	TTTTTTCTGTATGGGATTTTGCTAAACAACTCTTTCC
20[225]6[203]	TTTTAAGAAAAGTAAGCAGATAGGGCGGATCCCGGAA
16[120]10[98]	TTTTACCATCGATAGCAGCACCGCCTCAGACACCAGA
29[210]31[229]	TAAATTTATCCTTGCGCTATTAATTAATTTTTTT
13[49]15[68]	TATTAGATCGGCATAAAATGTTACTGTAGTTTT
29[259]47[278]	TAAGAATTTAGTAAAACCGTTGTAGCAATATTTT
25[56]1[75]	CAAGGATCCATTAAGCAAAAGAATTTCATGTTTT
33[105]35[124]	CAGATGAATTCATCTATTCCTGATTATCAGTTTT
8[176]8[147]	TTTTTAATAAGTTTTAACGGGGTCAGTGCC
19[105]17[120]	TTTGTCAACCGTCAGTCACCAATGAATTTT
2[176]2[147]	TTTTTTGCAGGGAGTTAAAGGCCGCTTTTG
12[71]11[55]	TTTTTTTCATAATCAAAATCACCCCACCAC
23[210]21[225]	TGCCAGTCTGAACAGAAGCCCTTTTTTTTT
18[173]9[176]	TTTTCCGATTGAGGGAGTGTACTGGTTTT
38[68]37[71]	TTTTAAATCCTTTTATTAATTTTAATTTT
24[173]3[176]	TTTTGGCTTATCCGGTCGCTGAGGCTTTT
43[106]41[120]	ATTAACCAGTTGAAGATTAGAGCCGTTTT
42[173]33[176]	TTTTCCTTGCTGACCTGATTGCTTTTTTT
46[278]29[281]	TTTTCTTCTTTGAAAACACCGGAATTTTT
32[176]19[173]	TTTTGAATACCAAGTTGACATTCAATTTT
10[111]11[124]	GAACCACGCCACCACCCTCAGAGTTTT
50[62]51[75]	GCCGTAAAAATCAAGTTTTTTGGTTTT
6[216]7[229]	AGTATAGAAGTGCCGTCGAGAGGTTTT
42[160]43[173]	ACCTCAAAAAATCTAAAGCATCATTTT
39[42]39[68]	TCGCCATTCAGGCTGCTCGTATTTTTT
46[216]45[225]	TGCTGGTAGCGTAAGAATATTTT

Continued on next page...

Table A.2 – continued from previous page

Staple name	Sequence
25[49]50[26]	TTCAACGAATGTGTATCACCCAGCACTAAATCGGAACCCTTC TATCATTTT
43[19]40[42]	TTTTCATGCCTGCAGGTCGACTCTAGAGGATAAAACGCGAAA GGGGGAAGG
1[26]5[48]	TTTTGAAGGCACCAACCTAAAACTCTTTGATACAACGGAACC GAGTGTACA
30[41]4[22]	TTAAAATGAAGCAACAAATGGATGAACGACTGACCAACTTTG AATTTT
28[41]2[22]	ATCAGGTTCATTTGAAAAACAAACAAAGCCCCCAGCGATTAT ACTTTT
37[47]38[19]	AGCCAACGTGCCCGAGCTTTCCGGCACCGCTTCTGGTGCCGG TTTT
49[19]47[48]	TTTTAGGGCGAAAAACCGAAAGGGAGCCCCTTAAAGAGCTGG TT
48[41]0[26]	ACGTGGAAGACAGTGATTCAAGGAGAAGTAATGCCACTACTT TT
51[26]27[48]	TTTTATGGCCCACTACGTGAACCAGGTAAACAAATCA
18[41]8[22]	CAGGTCTTAATGCAGAGATTTAGGAATACCTTTT
9[42]10[22]	GATACATGGAATTACGAGGCATAGTAAGAGTTTT
39[19]40[35]	TTTTAAACCAGGCAAAGCGCCATGCGATCG
13[22]14[19]	TTTTAAAAACCAAGGCTTTTGCAAATTTT
31[22]44[19]	TTTTATTTTTGTTGGGGTGCCTAATTTTT
21[19]30[22]	TTTTCGAACCAGACCGTCGCATTAATTTT
19[19]32[22]	TTTTTTATAGTCAGAAGAACAAACGTTTT
7[22]20[19]	TTTTTAATTTCAACGAGCTTCAAAGTTTT
9[22]18[19]	TTTTACATTCAACTTACCCTGACTATTTT
33[22]42[19]	TTTTGCGGATTGAGTGCCAAGCTTGTTTT
11[22]16[19]	TTTTCAACACTATGCGGAATCGTCATTTT
17[19]34[22]	TTTTTAAATATTCATTCGTTGGTGTTTTT
27[22]48[19]	TTTTAAAGGCCGGCTCCAACGTCAATTTT
29[22]46[19]	TTTTAGAGTCTGGGGCCCTGAGAGATTTT
5[22]22[19]	TTTTAGAGGACAGTCAATAACCTGTTTTT
23[19]28[22]	TTTTTTAGCTATATTTCATTGCCTGTTTT
35[22]40[19]	TTTTAGATGGGCGGTGCGGGCCTCTTTTT
12[48]12[22]	GGAACCGTTTACCAGACGACGATTTTT
37[22]36[35]	TTTTGCCTCAGGAAGATCGCGACGACG
6[48]6[22]	AGTAAATTGGGCTTGAGATGGTTTTTT
45[19]45[41]	TTTTGAGTGAGCTAACTCACATT
16[97]43[105]	TAATCAGGCAAGGCTCCTGATTTCAGGTTGAGGAAGTTTCCT GTGTCAGT
25[76]20[84]	ACCCTCACCGCACTCCAAAATTTAGCATTCCCAATTTTTGAT ATACATAC
47[104]25[125]	TTATAAAACGTGCTTTCCTCGTTTTGACGCTAATGCGGCTGT CTCATCGT

Continued on next page...

Table A.2 – continued from previous page

Staple name	Sequence
18[62]45[55]	GAATGACCATAAATATCAAAAAACCCGTTTTAACCGAAGCATTTGCGCT
8[55]29[62]	AAGATTCAATTACCTTCAAATCAGGTCAACGTTAATCGATGAACGGTAA
28[97]33[90]	GTAAAGTAATTCTACATATGTCAGGAAGGTCTGGCCCAGCTTGTAGATT
4[153]26[133]	AAAGGAGCCTTTAATTGTATCCCATCGCGTTTTAGATTACCGACCAATC
21[196]47[216]	ATAGCTAAGCGATATTCATCTAGAACAACTGAGAAGTGTTTTTATAATC
6[181]28[161]	GAGGTTTCACCAGTCGCTAATACCAACGCGCCATATTTAACAACGCCAA
4[69]9[76]	CGCAGACGCTGGCTACGAGAAATTTTAATTATTACATTCACAAACAAAT
6[90]11[97]	CAGTGAATACCAGTAAACGAACATTAAACCAGCATCTCAGAACCGCCAC
46[153]33[146]	TACCTACGTCACACGCCATTACAAATGAATATCAAACCCTCAGGAGAAA
30[132]43[139]	TACCTTTCAATTTCTTTCAATTACCTTTTACATCGATCAATACTGAGAG
18[139]45[146]	GGAAATTGGTTTACTTATTCAATTTGAACGAACGATGGCAGATTCACCA
46[146]3[153]	ATTTTGAATCCAATAGAGGCAATTTTGCGCGAGGCCCACGCATAACCGA
33[50]6[56]	TTATTTGCACGTAGAGTAACAGATTAAGAAAGACTTATGCGACACCAG
43[73]22[77]	GTAATCATGGTCTCACAAAATTCGCATTGTATTGCTCCTCTGCGAA
17[72]40[64]	AGTAGTTCTGAAAATAAAGCGCCAGGGTTTTCCGGCGATTCAATTC
5[70]48[67]	GACCTTCCGAGTAGTCTACTAAGAGGGTGTTCTAGTGTTCCAGTT
19[42]14[39]	GGATTGCCAAAAATCCTCAAATAGCGTCGTAATAGTTTTCGGTCC
5[238]29[238]	CTAAAGTAGGGAAGATTTATCATGCGTTATACAAACCGACCGT
5[161]25[160]	ACACTGAAAAAAAAGGCTCCATATATTCGGTATTCCAAATCA
30[181]47[195]	GCTGAGATTCAAATATTGCAACAGGAACGGTACGCCAGAATC
9[147]31[160]	CTTTTGATGATACAGGAGGGAAAAGGGCACAAAATGGAAACA
4[195]20[196]	AATAATTACAACGCATCACCGTTTGCTCAGTACCACCGAACA
28[48]32[35]	AAAGGCTAAGAGAATATTTTGCTCATTTCGGATTCTCCGTGG
32[146]6[133]	AGGCGAACAGCGCCAAGAACTGTATTAATCAGAACCGCCACC
21[147]47[160]	CAACATTGAGTTAAATTTATCCAAAGAAATGGAAAGCCGATT
7[98]30[112]	TTGGGAAGACTATTTCGGAACTACGCAGCTTAGAGCCGGCTT
49[126]46[112]	GGCGCTGTGGTTGCTAGAATCTCGTCTGAAATGGGGAGAGGC

Continued on next page...

Table A.2 – continued from previous page

Staple name	Sequence
24[125]3[111]	CCCGACTGACAATGCGAAAGACAGCATCGGAACGACCGATAG
44[125]20[126]	GCAGAAGCATTTAATTAACCTCTTAATTGCTGAATTTAAGAC
5[126]9[139]	GGGATAGCTCAGAGTCTGAAACCGTATATCCAGTAAGCGTCA
6[202]28[189]	TAGGTGTCTGTAGCCAAAGTCTCCAGAGCAGTAGGGCTTAAT
22[62]24[49]	TTGACCACGCATAGGGTCAATCATAAGGGAGATTTAAATCGG
48[139]47[153]	CGTACTAGCAAGTGTAGCGGTCACGCTGGCCGCTAACAGGAG
22[251]5[237]	TAAAAACTTTGTCGTTTCAACAGTTTCAGCGGAGTTAACGAT
48[66]43[72]	TGGAAAAATCCTGAGACGCGCTTTCCAACATAGCTCGAATTC
46[111]44[84]	GGTTTGCAATCGGCAGAGGTGAGGCGGTGAAATTGTTATCCG
6[55]20[42]	AACGAGTGACCAGGTTAGATACATTTCGACTCCAAATCGCGT
45[196]23[209]	CCTTCTGACCTGAAAATATCCTCTGACCCGCTCAACCTAATT
32[83]6[70]	TCATCAAACGCAAAATAAAGGCTCATTATAAGGCTTGCCCTG
44[174]22[161]	AACTGATAGCCCTAGTACATAATAGTGAGCCCAATAGATAAC
10[97]20[98]	GCCGCCGGCCAGAAGAGCCATCGGAATAAGTTTATGCAAACG
22[195]6[182]	AGAGGGTAATAGCAAAGTTACCAGAAGGGCGGGGTTACTCAG
45[127]33[125]	TTACATACCACCATGCCACGTCTGGTCAGTTGGCGTAACAG
29[239]46[252]	GTGATATGAGTAGAGTCTGTCCATCACGCAAATTTAACATC
21[91]47[103]	AGAGGTCCCAAAAAACCCCGGGCGCCAGTCGGCAAAATCCC
18[83]20[72]	TTAGAGCAAATCCTCTAACGGAACAACAGAACTGGTGGCA
28[83]2[72]	ATGCCGGATAGTAGAAGCAATTGTCGAACTTTGAGGACTA
23[115]48[98]	AAGCAATATAATGTTCAGAGCACGTATATCAAAAGAATA
35[70]40[87]	CAAAGAATTACAAAAAGTTGGGTTATCTTTAGGAGAGAA
26[73]30[63]	TAAATGCAATGTTCAACCAGCTATTTCGTAAATATTTAA
24[97]26[74]	GAATTAGATCGAGAGGGTATTAAACCAAGTATATATTT
43[53]21[62]	CGGGTACCGACGAGCCGAATAGGAATTGTAAGGATTAG
18[125]8[102]	AAAGGTGTTACCGTAACAGTTAATGCCCCCTGCAAAAT
30[188]4[179]	TTAAGACACAATGAAATTGAGACAAACTTTTTCACGTT
8[142]30[133]	GTAACAGTGCCATGAAAGGCATGAATAATGCGAGAGAC
21[70]47[79]	CTTTAATAAGCAAAACTAGCATCACCAGTGTTTGATGG
2[71]0[49]	AAGACTTTACACTAAAACACTCAGAAAGAGACGGGTA
23[231]4[223]	CCATATTCGCATTAGACGGGAGTTAGCGGAGAATAG
46[237]23[230]	AAGAACTCAAACTATGAAATATTCTTACTAAACAG
21[98]4[91]	ATTTTTGTATAACAAACCGGATTCTTAAACCTCCA
4[118]24[98]	CTTTCGATTGCGCCTGCGGGAGGAGGCAAGGCAAA
46[195]31[188]	TATTACCATAGAACGCTATTAGTCTTTAAGTGAAT
0[48]23[55]	AAATACGCCTTTATTTGTACCGGGCGCGAGCTGAA
46[188]23[195]	GCCAGCCATATTTTTGAGAATCTAACGAGCGTCTT
5[259]28[252]	AGACGTTCAGAGAGAAACGATAAAAAGCCTGTTTA
2[97]5[90]	AACGGCTACAGAGGATCCGCGTGTTACTGTAATCT
26[160]29[153]	CTAATTTACGAGCAAATAGATTTTAGGCCGCAAGA
29[189]44[196]	AGTTAATGCTTAGAAACCTTGCTTCTGTTTGAATG
5[133]28[126]	CAAGCCCGCAACTAAGTTGCTTTTTCGAGCCAGTA
28[62]3[69]	TTTGAGAAAGGTGGTAAAGCTGTATCATCGCCTGA
19[35]8[42]	GCAAAGCTTTAATTCTTTAATCATTGTGATCAGTT

Continued on next page...

Table A.2 – continued from previous page

Staple name	Sequence
28[55]49[62]	GATCTACCCATCAATCCACTACGATTTAGAGCTTG
25[112]28[105]	TTTATTTTTTCCTTATCATTCAACAACAAGTACCG
41[91]32[98]	CACTAACAACTAATAAGGAATTTAACGTTGAAACA
30[153]5[146]	AAAATCATAGGTCTTGTAGCTTAAATATAATAGGA
23[161]20[168]	TTATCCTGAATCTTATCAGAGAATAAGAGGAAACG
47[49]26[56]	TGCCCCAGCAGGCGACAAGAGTATGATACCTGAGT
2[146]4[126]	CGGGATCGTCACCCTCAGCAGACAACAAGGTTTAT
25[126]22[119]	AGGAATCCGAACCTCAAGATTAAGTACGGTGTCTG
24[76]21[69]	AAAGCCTCAGAGCACATCAATATTTAGTAGAGTAC
32[76]35[69]	CATTAAATGTGAGCAAACAGATAATGGATGCGGAA
16[69]34[50]	CTTTAGCGTCAGTAGACTGGATGCTTTAGAACCT
6[101]23[97]	AGCTGCTCATTTGACAAGGTTGATTAACATCC
47[137]26[140]	CGGGAGCTAACAGGGCGTATCAACTGTAGAA
19[84]16[81]	GACACCATTGGGAAACCATTATAGCGACAGA
40[86]17[83]	GTATTAGACTACCACCAATTATACCACCATT
31[112]45[126]	TTAATTAATAAAACCAACGCGCGGATTAT
25[140]23[153]	CGCCCAATAGCAAGTAAGAACACCCAGC
29[161]31[174]	CGCGAGAAAACTTTAGAGTCAAATCAAT
44[160]46[161]	AAACATCGACCAGTAATAAAAAACGCTC
20[167]6[154]	CAATAATTCAAGAGCCACCCTCAGAACC
43[140]31[153]	CCAGCAGAAAATACTTACCTTTTTTAAT
22[118]20[105]	GAAGTTTCATTCCACGGATGGTATGTTA
7[168]21[181]	AAGGATTAGGATTAAAACCGAGCAAGAA
45[70]29[83]	CAGTCGGGAAACCTTTTCTTTTGTCAAT
27[98]25[111]	GACAATACAAGAACACAAGCAAGCCGTT
32[97]31[111]	AATGTAGCTTCCCATCAAGAAAACAAAA
35[49]17[62]	ATCTGTTATCATTTAGGGTTAAACAGTT
20[125]6[112]	TCCTTATCTATTATCCACCACCCAAATC
47[161]45[174]	AAAGGGATTTTAGACAGGAAAGGGACAT
45[42]31[41]	AATTGCGAAAGTGTAAAGCCTAAATCAG
15[35]11[48]	AGAGGGGCAATACTCATAACCCTCAGAG
31[175]45[188]	ATATGTGATGCGCGTCTGGCCAACAGAG
43[84]33[83]	CATAGCTGGTTATCTAAAAAAAAATTGC
22[209]4[196]	CCCTGAAATTCCACAAGGAATTGCGAAT
5[147]7[160]	ACCCATGGCCACCCGAGGCTGAGACTCC
34[90]40[91]	TGTTTGGGAAGGAGCGGAATTAGGATTT
28[251]22[252]	GTATCATCCAATCCAAATAAGAATAACA
29[112]23[125]	ATATAACTATATGTATAAGAGCTTAAAT
32[125]18[126]	TACCTGATAGAAAATTCATATATTCATT
5[77]3[76]	ATCAAGATAGCCGGAACGAGGTAAATTG
27[92]49[104]	GACGACGCCCGAGAAGGAAGGGAAGAA
10[68]16[70]	GCAGGTGCCTCCCTCAGAGCAGTTTGC
47[80]27[91]	TGGTTGAGTGTCTGATAAATTGTCCA
16[80]10[69]	ATCACGCCACCTGACAGGAGGTTGAG

Continued on next page...

Table A.2 – continued from previous page

Staple name	Sequence
35[35]37[46]	CATCGTAACCGTGCTTGAGGGACTCC
47[84]45[94]	TCCGAAAGGTGGTTGTCGTGCCAGC
18[153]8[143]	AGGTAAATATTGACTACATGGTTGA
8[101]20[91]	CTACGTTAATACAGGACGTAGAAAA
41[56]43[52]	GCTGCAACAGTCACGACGTTGTCCC
4[178]6[168]	GAAAATCTCCAGTTTCGTAGTACCG
29[126]47[136]	AAATGCTGATGCAACGCTCAAAGAG
28[104]23[114]	ACAAAAGAATAAATCATACTTTTG
4[111]6[102]	GGTGAATTATTCATAACGTAACAA
18[76]17[71]	CAGCAAAATAACGACAGAACACC
20[71]19[76]	ACATAGCCCGAGGAATAAAAGAA
29[103]30[98]	AATCGGGTTAGGTTAGAAAAGCC
45[95]29[102]	TGCATTAATGGTATTGGTTGAT
34[49]17[41]	ACCATAGATAGGTCAGAATCCC
40[63]41[55]	GACAACGCAACTGTTGGGATGT
42[41]33[49]	ACGGCCACCGTAATGGTCAAAA
32[153]33[160]	CGCGCAGCAATAACGGATTCG
46[251]29[258]	ACTTGCCAATAAGGCGTTAAA
20[160]19[153]	AACGGAATACCCAAAAAGACA
31[63]44[70]	ACGCCATCAAAAATTTCCACA
45[56]46[49]	CACTGCCGGCAACAGCTGATT
3[91]2[98]	ACCTGAGCTTGATAGGGTAGC
5[112]4[119]	TACCCTCATTTTCACAGCTTG
18[104]9[111]	CCGACTTTGGAAAGCGCAGTC
9[49]8[56]	AACGCGGCCTTGATAGGTAGA
46[48]29[41]	GCCCTTCACCGCCTAGCAAAC
49[105]49[125]	AGCGAAAGGAGCGGGCGCTAG
26[132]27[139]	AATAATCAGAACGCGCCTGTT
11[56]10[49]	CGGAACCCAGACGATTCAAAA
5[154]22[147]	TACCGTACCACAAGAATGTTT
4[222]5[216]	AAAGGAACAACTAAGACAGC
14[38]13[48]	GAGAAATAGATAGCCCCC
14[68]13[71]	TTTTCGCGTTTTCCGTTTGCCATCTTTTT
26[176]25[173]	TTTTTAATATCCCATCGATATAGAATTTT
48[173]27[176]	TTTTCCGCCGCGCGAACAAGAAAAATTTT
0[75]25[75]	TTTTAGGAAGTTTAAAAATTTTTAGA
36[71]36[49]	TTTTAAGTTTGAGTAACACCAGT
25[19]26[22]	TTTTAATACTTTTGCGAAGGGTGAGTTTT
3[22]24[19]	TTTTCAAGCGCGATTATGACCCTGTTTTT
15[19]36[22]	TTTTAGAAGTTTTGCCACAGTATCGTTTT
47[19]47[41]	TTTTGTTGCAGCAAGCGGTCCAC
41[19]41[41]	TTTTTCGCTATTACGCCAGCTGG

Table A.2: Staple sequences for pointing-hand origami nanostructure.

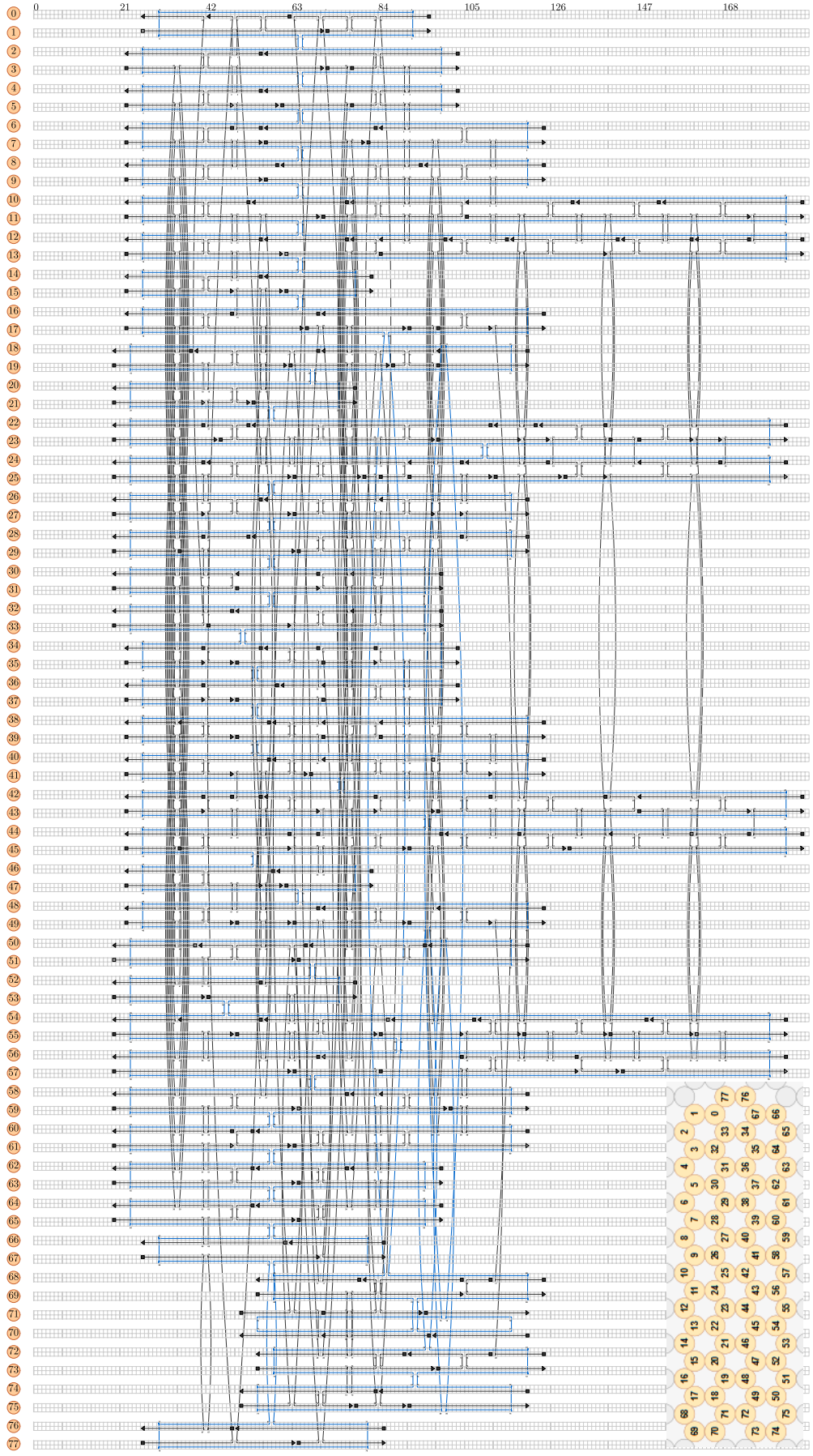


Figure A.2: Hand origami layout from caDNAno. Inset shows arrangement of helices

A.3.3 Honeycomb Lattice Sign-post Origami

Staple name	Sequence
0[111]16[102]	TCGCCCAGCAGGGAGACTTTTCAACATTGAAAGATTCA
0[195]4[182]	TTGTATCGGATTTTCATAGTTAGCGTAAACCCTCATAGGTGT
1[116]67[121]	AAGCATCTTTAATCACCGCCTTAACAATTTTTCAATATATATCT TTAG
1[140]5[149]	CCCTCAGGATACCGAGTAACAGTG
1[163]3[174]	ACGTTGGTGAATAGTTAATGCCCCCTTACTCAGCCCTCAG
1[203]3[216]	AACTTTCAAAGGAGCATGAAAGTATTAAAGAGGGTGGGATAG
1[85]13[90]	TTTTTCGGTCGCTTAGTAAATTGGGCTGAATTAC
10[106]66[92]	CCAGCGGATATTCATTGTCATTTCTTAGAGTGAGTAACCGAAC
10[136]66[130]	CCACGCCTTTAGCCATATGGAGCGGAGACTTTACA
10[166]31[166]	TTTTCCTCAGAACTAAAAACAGTTTT
11[105]6[98]	ACATCAATCATACAGAACGAGGAGGCTTCGCATAAGTGTCGACT GCTCC
11[126]67[139]	GAGCCACCACCGGACGACTTGCAACGCTCCTCAAAACTAATA
11[64]67[90]	TTTTGCTCATTCAGATTCCCGCAACTATCGACTCAGTGCCAAGC TTGCATAAAGG
12[134]5[125]	TTTTATTAGCGTTTGCGCCGCTTAATGACATTTAACG
13[64]69[95]	TTTTCAACTTTAAATGGTCATTTGACCCCATTAACGAACCACCA GCAGAAGATAAAACAG
13[91]13[111]	CTTATGCGATTTTAAGGGCAT
14[111]51[101]	TCATCAACTGGCTCAACGGAATCATGAGGCCACTAAACGGAGTA GCACCCACCGACCAGT
14[127]13[134]	TTTTTGTAGCGCGTTTTTTCGGTCATAGCCCCCTTTTTT
14[90]40[81]	CAGTCAGGAATTAGCAATAAAGCC
15[68]15[90]	TTTTAAAAATCTACGTTAATAAA
16[101]17[127]	TCAGGAATTACGAGGCAGAATCAAGTTTGCTTTT
16[127]15[127]	TTTTCTTTAGCAGGTAATTACGTCAGACTTTT
16[90]42[81]	TTAGGAAGAAGCCTGCAAGGATAA
18[127]20[98]	TTTTTAATCAGTAGCGACATAGTAAGAGCAAGGCTTTGTTTTGC
19[68]20[77]	TTTTACGACGATAAAAGTAAAAT
2[132]21[131]	ACAGAGGCTTTGAGTAAAACGACCAAGCGGCCGGAAATTTT
2[143]10[137]	TAGCATACACTTCTTTGACCCCCAGTGCCTTGATAGTTGGCAGG TCCCA
2[167]1[162]	GTCGTCTTTGCATCAGACACCAG
2[209]0[196]	CCACAGAATTTTCATGATATAAGTATAGTCTGAAACCTTTAA
20[131]19[127]	TTTTCGTCACCAATGAAACCATAAGAATGCAACGATAGCAGCAC CGTTTT
21[84]45[107]	AGCAAAAGGTAATAACCAAAATAAAGTACAGCTAATGCAGAACG CGCTTTTT
22[131]51[131]	TTTTCATTAAAGGCAGTAGGGCTTTT
23[93]7[111]	TGAGGATGTTACCTTTGAAAGAGGAGTAAGCGT
23[98]95[107]	GAGGGAAAAGGGCGTTAAATATACCGACTGCCGTAAAG
24[131]53[131]	TTTTTGGTTTACCAATCATAATTTTT

Continued on next page...

Table A.3 – continued from previous page

Staple name	Sequence
24[92]53[83]	CAACTGCTTTAAACAGAAGATTG
25[112]59[127]	CAATAGAAAAGAAAACATACAGAATACCGATGATGAAACAAATC AATATCGATAGCTT
25[54]51[69]	AATCAAAAATTCGCGTTAAATCAGCTCATTACGTTAAGCCCCAA TGTACCCAGAGAAT
25[86]26[90]	TAGTCAGAAGTTGGAATAAG
26[104]55[107]	ACACCACTATTTTGCAAATCCTCAAATATAT
26[131]25[131]	TTTTCAACATATAAAATTCATATTTT
26[89]6[81]	CAAAGCTGACCAATTAGCCGGAA
27[50]57[65]	TTTTCGAAAGACTGGGAACAATCAACATTA
27[77]62[91]	TTAATTTACGCAGTATGTTAGTGATTAATCAATTAATCGCGC
28[142]62[136]	AACCGAGGAAAAAACAAAATA
28[166]61[166]	TTTTGAACAAAGTAACAATTTCTTTT
28[85]61[76]	CTTATCGAGCTTCAAATCGCACT
29[147]2[144]	AAGTAAGAATCCTCTGGCCTTACAGCTTCAGCGAAGGAACGAGGG
29[54]29[83]	TTTTGGAAGCAAACTCCATTGAGTTAAGCC
29[84]9[93]	CAATAATTGTACAGGAGTAATCTT
3[162]4[156]	CCGCCAGAGGTTTAGTAC
3[175]5[188]	AACCGCCCGATCTAGAATTTTCTGTATGGGTTTATTCGGAAC
3[78]43[107]	TTTTGGGTAAAATCATAACCCATAACGAATAGATACCAATCAA TAATCGGCTGTTTTT
30[104]57[108]	GAGCGCTAGAAACAACTGGCACAAACGTAATAGTGCTTTTTAA CCT
30[153]11[166]	TTAGACGAATAACACGCCACCCCCTCAGAGCCGCCACTTTT
30[166]63[166]	TTTTGGAAGCGCAACAGTAACATTTT
31[54]65[76]	TTTTGAGAGTACCTTTAATTGCTATAATGCATGTGCT
31[84]11[104]	ATAAGAGACCCAAAAGGCTTGCCCTGACGAGAAAC
31[98]10[107]	TTGCGTTTTGTTTACACCA
32[125]27[131]	CCCAATCAAATGAAGAACAAAGCTATCTAATAACGTAAAGGTG GTTTT
32[166]65[166]	TTTTATTTGCCAGTCCTGATTATTTT
33[105]35[115]	CCAGCTAATCAAGATACCGCGCCCA
33[54]34[54]	TTTTTGTTTTAAATATAATTCTGCGAACTTTT
34[153]35[166]	TTTAGCGCCGGTATTCTAAGAACTTTT
34[166]33[166]	TTTTGCGAGGCGTCAGAGCCTATTTT
34[99]8[77]	TGCTTTCATTCCATATAACAGTTGTGAATATCAACGTTCATCA AACCAGGC
35[116]7[131]	ATAGCAAGCATTTTGAAGGAACCAGCCGCCAGACAGGAGCAGT CTGGCTTTTGATTTT
35[133]8[143]	TATAGAAGGCTTATAACCTCCACCGCCTCTCAGAGAGACGATA TTA
35[54]36[54]	TTTTGAGTAGATTTAGATAACCTGTTTATTTT
35[83]34[100]	TACTTTTCATCGTAGGAATCATTTAGT
36[104]36[81]	TTTTTGCAAGCCGTTTTTAATTTC

Continued on next page...

Table A.3 – continued from previous page

Staple name	Sequence
37[54]76[63]	TTTTGCTATATTTTCATCTACTATTGCGCTTTTCCAG
37[83]38[81]	CTGCTCATCCGCAAAAAG
38[107]37[104]	TTTTTTTAAACCAAGTACGAGAACAATTTTT
39[47]78[63]	TTTTTTAACATCCAATAAATCATAGCTAAACGGGGAGCGTATTG
4[111]1[115]	AAAGTACCGAAGGCACCAACCGACTAAAGTTA
4[155]3[161]	CGCCACTCAAAAACACCCTCAGAA
4[181]2[168]	ATCACCGGCCTATTCAGCTTGCTTTCGAAGTAAATAAGTTTT
4[222]2[210]	CCGTCGGAGGCTGAAAAAAAGGCTCCAAACAGTTTAGCATT
40[107]39[107]	TTTTTTCCAAGATTAAGCAAAAACGGGTATTTTT
40[76]71[92]	AGCATAAACAGGCACATCAATTTTGGGGGAATGGCTAATGCGC GAACTGATAGCTTTT
41[47]41[76]	TTTTAAAACATTATGACCCTGTAATACTTT
42[107]41[107]	TTTTTTCTTTCCTCAACTTATTTTATCATTTTTT
42[72]80[66]	AGAACCCTCAGATTGCCTGGC
42[80]47[102]	AAATTGTAGAAAAGTCCTACATGTTCCGACAAACAGTCAAATC AATTTTCGAGCCAGT
43[47]44[47]	TTTTGCAATGCCCATCATATCCTGAGTATTTT
44[107]14[91]	TTTTTCTGTTTATCAACCCAAAAGTTGAGATACGAACTATTATAC
44[69]81[85]	AAAAATACTAATTTCCTACATACATTGGCAGATTCACCAGTTTT
45[47]46[47]	TTTTATGTGTACAGACCTGTCGGTAAAGTTTT
46[107]18[95]	TTTTTAATAAGAGAATATAGCGAGACA
46[72]84[63]	TAAAGTAATTGACGACAGGCCTTG
47[47]48[50]	TTTTATTCAAAAGGGTAGCTGATAATTTT
48[100]89[89]	TTTTTCAGAGGCCCATCAATATGAGTTTTTATATTTT
49[50]50[50]	TTTTATTAATGCCGAGAGTCTGTTTT
49[77]49[100]	TTTTTGAGAGATATTTAGGTTTTT
5[114]66[109]	TAAGTACAACCACAGCATTCCAGAACACGTCAACAAATAAGGA ACAAAACTCGTATT
5[126]1[139]	GGGTCAGCGATTATAAAGAGGCAAAAGAAACGGCTATCGTCA
5[150]9[166]	CCCGTATAAACTTCTTAAGATATTCACTTTT
5[189]1[202]	CTATTATCCCGGAAGAGCCACCACCCTCCAGCCCTGCTAAAC
50[131]6[115]	TTTTTTAATTGAGAATCGCCCGCTCAATGAATTAATTGACGAG TG
50[69]85[85]	ATTGCCTGGAGAGGCCGTTCTGAGAAAGTAACCGTCACTTGCC TGAGTAGAAGATTTT
50[73]51[92]	GGTCCGATGAACGGTAATCGTAAAATT
50[97]91[89]	CAACATGTACTACAAAGGCTAAGGAGGCCGTTTT
51[102]94[91]	ATATCATATGGCTGCGCACGGGGAAAGCCGG
51[50]52[50]	TTTTGAGCAAACACGGTTGATATTTT
51[93]22[84]	CTTACTTGAGCCATTTGGG
52[131]93[131]	TTTTTACTAGAAACGCCGCGCTTTTT
52[92]94[77]	TACAAACTAGCATGTCGCTGGCAGGCGAGA
53[50]95[65]	TTTTATCAGAAAATATTTTGCTACGTGAAC
53[84]24[93]	TATAAAATAAGGCGACATT

Continued on next page...

Table A.3 – continued from previous page

Staple name	Sequence
54[114]60[105]	TGGGTTAATTTGTAAATGCTTAGGCGCTGAGAACATAGATGTG AGTGAATAA
54[131]95[131]	TTTTGACCTAAATAACCCTAAATTTT
54[92]95[83]	GATAGCAAATATTTAATCAAGTT
55[108]50[98]	TTTATTTGAAAAGAATAATTTAGTAAAGCCAAATATTTAACAA CGC
55[40]56[40]	TTTTTTAAATTTTTGTCTGGCCTTCCTGTTTT
55[87]59[97]	AGAAAACTTTTAATCGCAAGACTACAATTTATTTTTCCC
56[131]55[131]	TTTTATAACTATATCATCTTCTTTTT
56[85]57[76]	AAGAGGAACGCCATCATGAGCGA
57[109]4[112]	CCGGCTGATGTCACAATAAGACAAGGTAAATTCACCGTATTAC CATTAGCAAGCGAAAC
57[40]58[40]	TTTTTAGCCAGCTTTCAACGGCGGATTGTTTT
57[66]24[50]	AATGAAAATAATCAGGTCAACGAGAATGACCATATTTT
57[77]56[86]	GTAACAGGTCTGAGAGACA
58[131]57[131]	TTTTAGATTAAGATTGGGTTATTTTT
58[85]59[76]	TCATAACCCGTCGGATGTAGATG
59[40]60[43]	TTTTACCGTAATGGGAAGGGGACGATTTT
59[66]26[50]	TGGTTCTCCGTTCAAATAAGATTAAGAGGAAGCCTTTT
59[77]58[86]	GGCGCATTAATTAACAAAA
6[114]5[113]	TACTGGCGACAATCCGTAA
6[131]23[131]	TTTTTGATACAGGGAAATTATTTTTT
6[80]23[92]	CGAGATCCCCCTCAAACGAT
60[104]26[105]	CCTTGCTTTAGAATCCTTGAAAAGAGTCAGAAAATCGCAAAG
60[135]1[132]	AGTACATAAACATCAAGACGCAATTACCGAAGGAAAGCGGTTG AGCGCCGACTTGCGGG
60[166]28[147]	TTTTATTTGAATTACCTTTTTACATTTTACCAGA
60[69]28[54]	CATCTGCAGGAAGAGCGAACCAGACCTTTT
60[97]60[70]	TCTGTAAATCGTCGCTATCGTAACCGTG
61[105]8[98]	AAAAGAACAAAAGAATGAAATTACCGTTCCACAGA
61[43]62[43]	TTTTCGACAGTATCGGAAACCATTTT
61[77]28[86]	CCAGCTTATTCATTGACTC
62[114]30[105]	GAATACCAAGAGAAATAGGTAATT
62[135]10[126]	ACGGATTCGCATTTTCAACACCCTAATAGCACCTCAGA
62[166]63[153]	TTTTGTACCTTTTACATCGGTGAATAT
62[90]63[69]	AGAGGCGAACAGCTTTCCGGCACCGCTTCCATTCA
63[43]64[43]	TTTTGGCAAAGCGTCTTCGCTATTTT
63[70]30[84]	GGCTGCGCAACTTCAAAATTAGAGAGAT
64[108]32[98]	ATAAATTTTGCGAAACGATTGATGG
64[139]35[132]	ATAATCCTGATTGTACCAGAATATTTATATCTTACCGGGAGGA ATCAGA
64[166]33[153]	TTTTTCAGATGATGGCAATTATCATATTTACAAAGTCTTTC
64[62]59[65]	GCGGGCCCCATTCGCTGGTGCCGGCCTCCAGTTTGTAGGTCAC GT

Continued on next page...

Table A.3 – continued from previous page

Staple name	Sequence
64[69]32[54]	GATCGGTAAGGGGGTGTAGCTCAACATTTT
65[43]66[43]	TTTTTTACGCCAGTTTCCCAGTTTTT
65[86]33[104]	AAGTTCTTAATTGTGTCTGGAAGTATTTTGCAC
66[108]69[115]	AAATATCTAAACTGGTCAGGTCAGTATTA
66[129]69[136]	AACAACTAACATATCAAACAACAGTGCCA
66[166]67[150]	TTTTACATTTGAGGATTTAGGCCG
66[76]67[69]	TAAGTTGGGTAACGGACGGCC
66[91]65[85]	GTTATTAATTGCGATGCAAGTTAA
67[122]62[115]	GAGCATTCGACAGAAACCTTGGATTTGCGTAGCTGATTGCTTT
67[140]60[136]	GATTAGAAAGTATTAATTATCCATCAATCGTCAGAGAGAAACA ATTAATTTAATGGAAAC
67[151]69[166]	TCAATCACCTCAGCAGCAAATGAAAATTTT
67[43]65[62]	TTTTCACGACGTTGTAAAACCCAGGGTCTGGCGA
67[91]61[104]	AATTGAGGAAGGTTCCTTTGCCATTATCTGGAAGGGTAAAACT TACAAACCTGAGC
68[166]67[166]	TTTTATCTAAAGCATAGATAATTTTT
68[69]68[43]	TAGAGGATCCCCGGGTACCGAGCTTTT
68[81]30[54]	GCAGGAAGTACGGCTGAATCCTTTTGAACCCACAAGAAACAGG TCAGGATTATTTT
69[116]33[125]	ACACCGCCTGCCCTCAAATCCTGA
69[137]28[143]	CGCTGAGAGCTGCTGAAAACGAGCATAAACACAGAGAGGGAGA ATTTAAGAAAGGA
69[43]69[69]	TTTTTCGAATTCGTAATCATGGCCGAA
69[96]68[82]	AGGTGAGGCGTTGGCAAATCAACAGTTGGCCT
7[112]64[109]	CATACATCTGAATTAGCAATAGTCAGAGAAGAAATATACTTCTGA
7[77]27[76]	ACCGAACGGATTGCATCAAAATCGCGTT
70[69]70[43]	AAATATCATAGCTGTTTCCTGTGTTTT
70[92]35[82]	TTTTCCTAAAACATCGATTAGA
71[43]71[69]	TTTTTGAAATTGTTATCCGCTCGTCTT
72[69]72[43]	TATTAACAATTCCACACAACATATTTT
72[92]37[82]	TTTTGACAATATTTTTCGCGAG
73[43]74[56]	TTTTCGAGCCGGAAGCATAAAGGAAAGGACCTTGTAAAGCC
74[85]73[92]	TTTTTAGAACCCTTCTCGTAAGAATACGTGGCACATTTT
75[36]76[36]	TTTTTGAGCTAACGTGCCAGCTTTTT
75[49]74[36]	TCACATTTGGGGTGCCTAATGAGTTTT
75[56]38[47]	AATTGCGATAGTAGTAGCATTTT
76[62]40[47]	TCGGGAACAACGCGTCGGTTGTACCATTTT
76[85]75[85]	TTTTCTGGCCACCCGCCACTGACAGAGATTTT
77[36]76[49]	TTTTGCATTAATGAATCGGCACCTGTC
78[62]42[47]	GGCGCCAAACAGCTTATATTTTAAATTTTT
78[85]77[85]	TTTTATAAAAGGTTTGAGGCGGGACATTTTTT
79[36]80[36]	TTTTCACCAGTGAAAGCGGTCCTTTT
79[49]78[36]	GACGGGCGGGTGGTTTTTCTTTTTTTT
8[142]63[139]	AAGCCAGAATGCCCTTTTAACTGAGGTTTAA

Continued on next page...

Table A.3 – continued from previous page

Staple name	Sequence
8[166]29[166]	TTTTAAACAAATACAGATAGCCTTTT
8[97]64[70]	TGAACGGAAGAGCAAATATCATTTGCACGTTAGAACCTACCAT AGTTGGGAAGGGC
80[65]84[56]	CCTGAGAGAGATTATTTTTTGACGCGCCAGCCTGGTAA
80[85]79[85]	TTTTTCACACGCCGCCCTTCAACCAGTATTTT
81[36]80[49]	TTTTACGCTGGCTGAAATGGTTGCAGC
82[85]42[73]	TTTTCTCATGGAAATAACGAGCATTTT
83[36]82[36]	TTTTCAGCAGGCAATATTACCTCAATCGTTTTGCCCTTTT
83[63]83[85]	CATTGCAACAGGAAAAACGTTTT
84[85]44[70]	TTTTACTCAAACTATCATAAACAGAACAAG
85[36]84[36]	TTTTCCTGTTTTGATTAGTATATCCAGAACGAAAATTTTT
85[56]88[40]	ATAACATTGTAGCAAATCCCTAAGAATAGCCCGAGATAGGTTTT
86[85]46[73]	TTTTCATCACGCAAATGCCGGAGAAGG
87[36]86[36]	TTTTGTTCCGAAATCGGCAAATACTTCTTGATGGTGTTTT
87[63]87[85]	TATAATAAAAGAGTCTGTCTTTT
88[89]88[63]	TTTTATCAGTGAGGCCACCGAGATCAA
89[40]89[62]	TTTTGTTGAGTGTTGTTCCGCCA
9[94]9[111]	GACAAGAACAGCCGCCGC
90[62]90[40]	CGGTACAGTTTGGAACAAGTTTT
90[89]89[69]	TTTTATTAAAGGGATTTTAGACAGGAAGAATCCT
91[40]91[62]	TTTTAGTCCACTATTAAAATCAG
92[100]52[93]	TGGTTGCTTTCGGTCACCGTTA
92[131]24[112]	TTTTTAATGCGCCGCTACAGCCACACCAAGCCTGACACCGGAG CGCCA
92[62]92[40]	TTAGAGAACGTGGACTCCATTTT
92[76]25[85]	ACGTGCTCTAGGGCAATCATAAAACAGGTTCAGAATTTACCCT GACTATTA
93[40]93[62]	TTTTACGTCAAAGGGCGAAAGGA
94[131]54[115]	TTTTGGGAGCCCCCGATTTAAAATCGGTTAA
94[62]94[40]	AGCGAAAAACCGTCTATCATTTT
94[76]55[86]	AAGGAAGACCCAAAATTGTAATTTTAACCAATAACGCG
94[90]92[77]	CGAACGTAGTGTAGGACGAGCACGTATA
95[108]92[101]	CACTGAGCTTGGTAACCAGGCGCGTACTA
95[40]54[40]	TTTTGGCGATGGCCCATTAAAATTCGCATTTT
95[66]50[74]	CATCGGAAGAAGCGGGCGTTCCTCGAGCGGGAGCTAAACTCA
95[84]54[93]	TTTTGGGGTCGAGGCGTGT
5[231]3[253]	TCAAGAGCCAGGCGACCCATGTACCGTAACACTTTTT
3[217]5[230]	CAAGCCCACGCCTGTCAGCGGAGTGAGAATCTCCAAGACTCC
0[260]5[257]	TTTTCGAATAATAATTAGGATTAGCTTTT
17[68]18[68]	TTTTCTAATGCAGATACTCGTTTACCAGTTTT
18[94]2[78]	CTATACGTAATGAAGTTTCCATTAAACTTTT
20[97]4[82]	CAGAGGGTCACCAGATTTGTATCATCTTTT
21[61]22[61]	TTTTTAGCGTCCAATACTGAGCCAATTAGCGGAATCGTCATAT TTT

Continued on next page...

Table A.3 – continued from previous page

Staple name	Sequence
23[61]6[61]	TTTTAATATTCATTGAGCGCAGACGGTCTTTT
36[80]12[64]	GCAATCATTGTTGAGATGGTTTAATTTTTTT
38[80]14[68]	GTGGAGGCAAAGACGTTGGGAAGTTTT
40[80]16[68]	TCAGTGCGGGATACCACATTCAATTTT
5[82]0[85]	TTTTGCCTGATAAATTCCGATATATTTTT
7[61]8[64]	TTTTAATCATAAGGGAGCATAGGCTTTTT
89[70]20[61]	GAGAAGTTATTCAAGTAGCTAGTTTAGACTGGATTTT
9[64]10[64]	TTTTGGCTGACCTAACAAAGCTTTTT
17[72]18[52]_FL	TTTTCTAATGCAGATACTCGTTTACCAGACCACCTCCTTACTACTCCC
18[94]2[62]_FL	CTATACGTAATGAAGTTTCCATTAAACACCACCTCCTTACTACTCCC
20[97]4[66]_FL	CAGAGGGTCACCAGATTTGTATCATCACCACCTCCTTACTACTCCC
21[61]22[45]_FL	TTTTTAGCGTCCAATACTGAGCCAATTAGCGGAATCGTCATAACCACCTCCTTACTACTCCC
23[61]6[45]_FL	TTTTAATATTCATTGAGCGCAGACGGTCACCACCTCCTTACTACTCCC
36[80]12[48]_FL	GCAATCATTGTTGAGATGGTTTAATTTACCACCTCCTTACTACTCCC
38[80]14[52]_FL	GTGGAGGCAAAGACGTTGGGAAGACCACCTCCTTACTACTCCC
40[80]16[52]_FL	TCAGTGCGGGATACCACATTCAAACCACCTCCTTACTACTCCC
5[82]0[69]_FL	TTTTGCCTGATAAATTCCGATATATACCACCTCCTTACTACTCCC
7[61]8[48]_FL	TTTTAATCATAAGGGAGCATAGGCTACCACCTCCTTACTACTCCC
89[70]20[45]_FL	GAGAAGTTATTCAAGTAGCTAGTTTAGACTGGAACCACCTCCTTACTACTCCC
9[64]10[48]_FL	TTTTGGCTGACCTAACAAAGCTACCACCTCCTTACTACTCCC
cy5_label	/5Cy5/GGGAGTAGTAAGGAGGTGGT
B-4[257]0[237]	TTTTGGGGTTTTGCTCAGTAAAGGATTTTTTCACG
B-0[236]4[223]	TTGAAAATAGAAACAGTACAAACTACAAATAGGAGATAAGTG
B-2[253]1[260]	TTTTGAGTTTCGTCACGGAACAACTAAAGGAATTGTTTT
3′-0[244]1[237]	TTTTCACGTTGAAAATAGAAA
3′-2[253]4[223]	TTTTGAGTTTCGTCACCAGTACAAACTACAAATAGGAGATAAGTG
3′-4[257]5[244]	TTTTGGGGTTTTGCTCAGTAAAGGATT
3'add	GGAACAACTAAAGGAATTGCCGGCCGGCGGC
3′H'	GGAACAACTAAAGGAATTGCCCGCCAAAAAAGCATTGCTTATCAATTTGTTGCACCggg
H'*	cccGGTGCAACAAATTGATAAGCAATGCTTTTTTGGCGGG
BG_tag	TTTTCACGTTGAAAATAGAAAGGAACAACTAAAGGAATTG/3BG/

Table A.3: Staple sequences for honeycomb-lattice signpost origami nanostructure. Colours relate to staple mixes (see below).

Staple mixes for origami variants (staples in 10-fold excess of scaffold):

Basic Mix: Black + red + green staples, for unmodified origami.

Cy3 Mix: Black + blue staples + green staples with cy5_label in 20-fold excess over other staples, for cy5-labelled origami.

IHF Mix : Black + red + grey + purple + pink + orange staples, for IHF duplex modified origami.

BG Mix : Black + red + purple + cyan staples, for benzylguanine modified origami.

GAO Mix : Black + red + grey + purple + magenta staples + GAO (transcription product) in 12-fold excess, for GFP aptamer modified origami.

Fluorescent/tagged mixes are also used, e.g. cy3-labelled benzylguanine-tagged origami are synthesised with a mix of black, blue, purple and cyan staples.

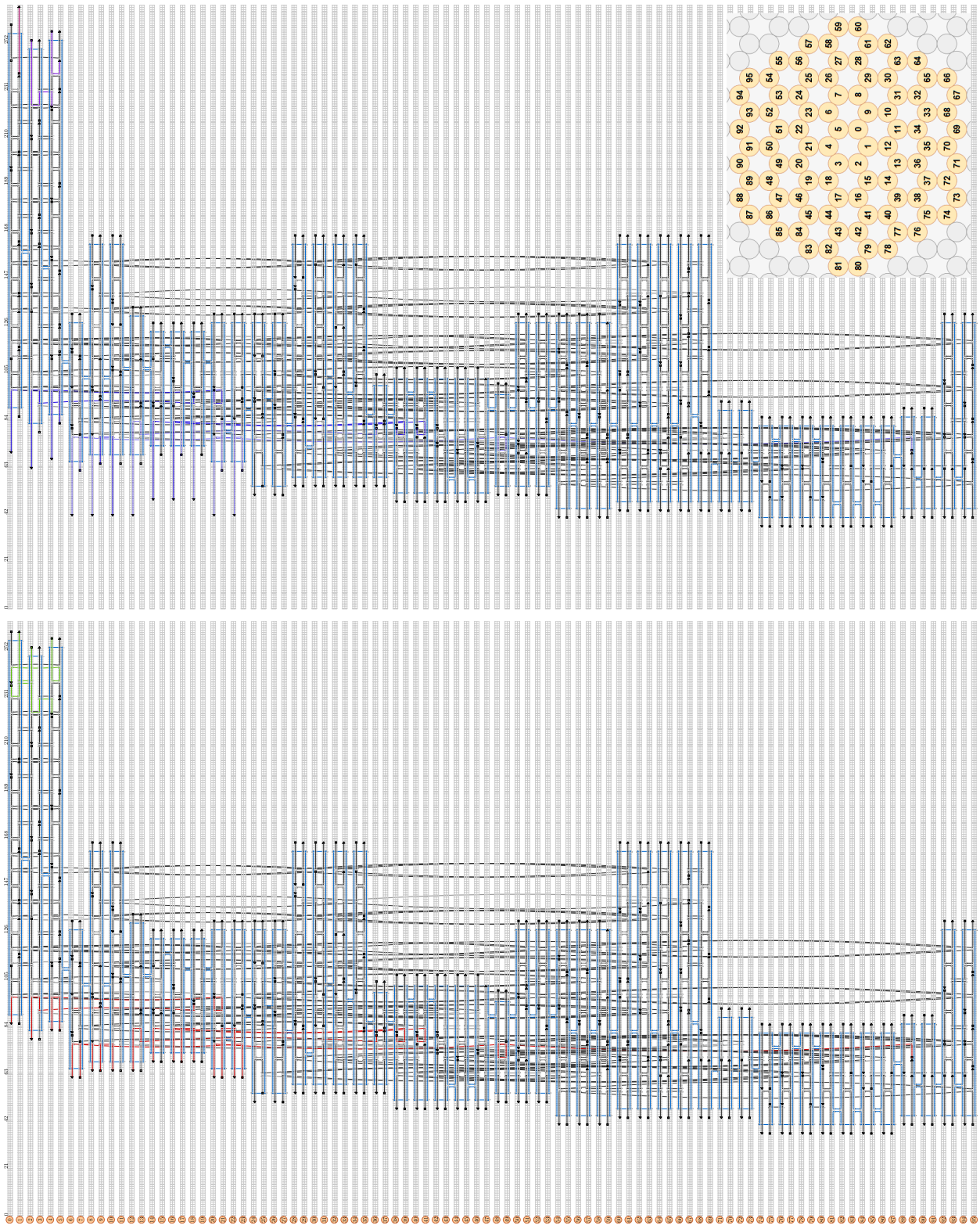


Figure A.3: Honeycomb-lattice signpost origami layout from caDNAno. Two views show staples for the basic nanostructure (bottom), and with extensions for hybridisation to fluorescent oligonucleotides and GA-O (above). Inset shows arrangement of helices.

A.3.4 Square Lattice Sign-post Origami

Staple name	Sequence
54[86]64[68]	TTTTTTTTTTGCCCCAAAAACAGGAAGATGTCAAATCCGAACT AAAACGCAAGACAT
92[87]82[88]	TCATAGGTGAAAACTTTTTCAAATAATTTAATTTACATTTGCG GGAGC
79[7]94[7]	TTTTTTTTTTTCGACAACTCGTAAGTTACAAAATCGTTTTTTT TTT
51[9]55[28]	TTTTTTTTTTCGTCTCATTCATGACAAGAACCGGATATTCAT
88[115]95[99]	TTTTTTTTTTCAGTAGGGCTTAATTAAAGCCAAATCAGAAAC
85[7]85[31]	TTTTTTTTTTGGATTATAGTTTCGTCACCAG
86[63]3[71]	CGTCTTTCTTCCACAGACAGCCCTATAATCGGGCAAAAGA
3[48]45[70]	CCTAAAATAGCGGTTGTAGCACAGACGTATTTTCTGTATGGGA TTTTTTTTTTTTT
11[9]12[9]	TTTTTTTTTTTTTGCCATCTTCACCGGAACCGTTTTTTTTTT
72[86]71[86]	TTTTTTTTTTGACTTCAAATACTTTAATTGCTTTTTTTTTTT
74[31]59[31]	AGCACTAAACTAATGCAGATTCAAAAAGATTCATCAGTTGTAT GCGAT
58[59]50[56]	AGCTATTAATTAGTATTTGAGAGATCGCATGCCT
77[12]71[30]	TTTTTTTTTTGGATTTAGAAGTACTGACCTGAAAATTGCTGAA ATCGTCAAAACAGT
92[118]79[91]	TTTTTTTTTTAAGAGTCAATAGTGAATTTATCAAAAGCGATAG CCATAAATCAATA
5[136]2[153]	GTGCCAGCTGCATTAATGAATCGTCCAGTT
1[4]0[4]	TTTTTTTTTTCCAAAAGGAGCCTTTACTCCAAAAAAAAGGCTT TTTTTTTTT
94[120]91[115]	TTTTTTTTTTCTTTTTTAATGGAAGTTAATTTTTTTTTTT
49[4]47[16]	TTTTTTTTTTCTGCAAGGCGATTAAGAGTATCGGCCTATGTGG GGGGCAGG
78[34]78[7]	ACCCTTTTAGACTTTACAAACAATTTTTTTTTTT
82[43]91[45]	GCCAAAACCACCTCAATTACCTGAGCAATTCGCCTGAATGCT
36[64]19[54]	TTTTTTTTTTTTAGTTGCTATTAGGAAACGATAGCCCATTTTT TTTTTT
44[31]44[9]	GAATAGAAAGGAACCACCATTTTTTTTTT
9[9]10[9]	TTTTTTTTTTCAGACTGTAGCCTTATTAGCGTTTTTTTTTT
66[43]70[52]	CATTCATTTCGCAGCGAACCCAAACTCCAACAGGTATGG
84[31]5[39]	TCAATATAGTAATGCCACTACGAATTTTTCATAGCCGGAA
40[47]47[47]	GAGGCGTTACCGTACTCCAGCTTTCCGGCACC
78[120]93[118]	TTTTTTTTTTTGAAATGGATTATAAGACGCTGAGTTTTTTTTTT
81[84]86[88]	AAAGTAAACAGGCGTACTATCCATATTTCGACGACAATAAACAA TATCAACA
20[54]33[43]	TTTTTTTTTTGGAATACCCAAAATACATAGAACAAAGTAAGAAA ATATTTATCCTTT
44[39]84[32]	CGGAGTGATACAAACTACAACGCCCACGCTGCGCGATTCA
94[36]63[30]	GCTTTGAATACCATTAAATCCCGTTATTAGCCGTCAAGCCTGCA

Continued on next page...

Table A.4 – continued from previous page

Staple name	Sequence
	AGTAGGTAAAGATACATGCAACAC
38[70]37[64]	TTTTTTTTTTGTTTTGAAGCCTTACCGTAGTGAATCAAGATTTTTTTTTTT
91[46]95[59]	GATGCAAATCCAATCGCAAATAAGGTGAT
34[23]32[18]	ACGATTTTTTGACTCCCTGAGTTTAACGTCAAGAACAC
80[40]78[35]	AATACCGAATTTGCTCCAGAACAATCAACAGAGATAGA
77[64]58[60]	AATATTTTGAGGTCATTTTTGCGGCAGGATTAATTCCCAACAATAACCACCCTGTAAAGCCTCAGGGT
20[27]6[9]	GATTAACCACCAGGAAACGTCACCAATGAAATTTTTTTTTT
85[88]2[100]	CCTGAACAAGAAAAATCCTAATCCCCCAGCAGC
2[219]1[183]	TTTTTTTTTTTTCGGCAAAATCCCTTATAAATCAAAAGAATAGCCCGAGAACTCCAAC
5[56]47[71]	CCTGGGGTCAGAGGCTAAATTTTTCTGTTGGGGTGCCGGAAACCAGGC
66[63]54[52]	ATTCTACTAATAGTAAGCAATAATACTTTTTATACCCATTATTTCAATCACAAA
27[9]9[48]	TTTTTTTTTCTGAATACATGGGGTCAGAGACAAAAATTTTGTCCATAGCCCCGCGTTTTCATCGGCACAGTATTTTTTTT
3[9]84[7]	TTTTTTTTTTGGGTAAAATACATCCTGATTGTTTTTTTTTTTTTT
87[76]76[64]	CAGAAACAACGCCAACATGTATCATATGTAAAGGGACAAACATCAGAAGAACTGCAACAGAGAAGATA
63[31]5[55]	TATCATAACCCTCGGGGAGAAGCCATGCCTGGAAAGGCAGGTTATTGTAAAG
74[115]81[120]	TTTTTTTTTTACCCTCAACAAATTAACCGTTTTTTTTTTTT
13[9]13[48]	TTTTTTTTTTCCTCCCTCCGACTTGAGCCATTTGGGAATTAGTTTTTTTTTT
22[59]37[39]	TTTTTTTTTTCGAGGAAACGCAGCCAGCATTGACAGTACCA
43[9]26[9]	TTTTTTTTTTTCAGAGCAACTAAAGGAATAATTAACTAAATGAAAGTTTTAACGGCTTGCAGTCTTTTTTTTTT
91[64]77[59]	AAGACAAATAACCTTGTCCCTTAGAGTAATAAAAGGGACAGCAC
89[25]90[38]	TTGCACGAATCATAATTACTAGATAAGAATAAA
44[70]43[70]	TTTTTTTTTTTGCTAAACAACTAAACCAAGTATTTTTTTTTT
92[63]90[52]	TTAACCTCATTAATTTCTTCTGTAAATCGTCAAGAAGACGTT
79[48]55[59]	ATTACCGCCAGCCATTCAAACTAAGCCAGCAATTGAGGACGGAGACATGTATAAGTATG
15[26]11[48]	AGGGAAAGAGCCACTTCATAATCAAAATCAGCAAATTTTTTTTTT
49[42]56[30]	TTCCCAGACTCTAGGAGCGAGTTGAACGGTAATCGAATCA
70[86]94[84]	TTTTTTTTTTCCTTTTGATAATGAATGGCTATTAGTCCACCAGTCCTCATGGAAATA
64[86]63[86]	TTTTTTTTTTTTGTACCAAAAGATAAAAATTTTTTTTTTTTT
68[47]78[48]	AACGAGTACTGTAGCTAGAGCTTAGCGTAAGAATACGTGTTCTGGC
1[104]83[120]	ACAACCATACCCTCAGCGATTATATCCATGTTCACGTATAACGTGTTTTTTTTTT
12[48]28[36]	TTTTTTTTTTAGCCACCGGAACCGGTAAATATTGACGGACAATGA

Continued on next page...

Table A.4 – continued from previous page

Staple name	Sequence
	AAGTAA
65[4]65[26]	TTTTTTTTTTCTTTTGCAAAAGAAGGCAA
18[54]33[35]	TTTTTTTTTTAAAGGTGGCAATAGAAAATGCCCTTTTGGTAATAA ATAG
2[127]85[120]	CTTTTGCGGGATCGTCCGCCCACGTGACGGGGAAATCCCATAATA GCCGGCGATTTTTTTTTT
56[86]55[86]	TTTTTTTTTTTTGCCTGAGAGTAATCAGAAAATTTTTTTTTT
53[9]2[9]	TTTTTTTTTTAATTGTTATCCTTCCATTAAACTTTTTTTTTT
83[7]87[25]	TTTTTTTTTTATATTCCTGATTATACCATATCTGAATAATG
78[95]81[83]	GGCAGATTTTTAATGCAAAAATACCGAACGAAAAGCATCACATCA CTTAGTA
4[83]87[75]	CAGAAAACACTCGAAAGGAGCGGGCGCTTGTC
81[7]82[7]	TTTTTTTTTTGAGTAACATTATCCGGAATTATCATCTTTTTTTTTT
25[9]19[27]	TTTTTTTTTTGAAAGCTTGATGATACAGGAGAAACGCGTTA
47[17]46[4]	AAGATCGACGCCAGCTGGCGAAATTTTTTTTTT
95[60]4[48]	GAAATTTTAGACTTTTTATAAACCGAACTGACCAA
81[52]87[63]	AGTGAGGAACGCCCGCCGCAGAGGCACCGACAAAAGGTAA
95[100]77[115]	AATTTTTGATTCATTTTGCGCCATTGCGAACTGATAGCCCTTTTT TTTTT
66[91]67[91]	TTTTTTTTTTGCGCGAGCTGAAAAGGTATATTTTCATTTGGGTTT TTTTTTT
0[27]42[9]	AAATATTGTATCTTTCATCGTAGGACCACCCTTTTTTTTTT
58[91]59[91]	TTTTTTTTTTTAGCTGATAAATTAATAGAAAAATCTACGTTCTTT TTTTTTT
64[39]68[9]	CAGACGACGATAAAAACCAAAATAGCGAGAGGACGAGGCATGGTT TTTTTTTT
40[70]39[70]	TTTTTTTTTTCCGGTATTCTAACTTGCGGGAGTTTTTTTTTT
6[48]21[54]	TTTTTTTTTTCATTAGCAAGGCCCCAGAGCCGCCATAATAACTTT TTTTTTT
5[4]52[34]	TTTTTTTTTTATAGGCTGGCTGACCTAACATACGGAGGAAGTGCT CACAAAATATTTTAGCTTT
60[59]52[52]	ACAAAGTCAGGAAAACAAGAGAATCGAAACAACCGGCC
65[27]58[9]	GGCAAAATTTCAACTTTAATCATTTTTTTTTTTT
37[4]35[23]	TTTTTTTTTTGATTAGGATTAGCGGGTGAATCTTACCAGAAGTCA AGAACGCTAACGA
21[4]20[4]	TTTTTTTTTTTCAGAGCCGCCACCAGAAGACTCCTTACACCCTTT TTTTTTT
46[31]40[9]	TCGCTATTCACTCCAGCAGGAGGTTTAGTCCCTCATTTTTTTTTT
90[115]78[96]	TTTTTTTTTTTCATCTTCTGACCTAATATTTTAACAGTATTAGAT TTTACATT
0[55]31[64]	CCTTATCCGGGTATTTTCAACGAGATAACCCACTTTTTTTTTT
90[51]0[56]	AAAAAAAGCCTGTTTAGTAATTTAGGCGCTTAATGGCAAGTGCGA AAGAGCTGTCTTT
19[4]16[20]	TTTTTTTTTTGCCACTTACGCAGTATAAAGACACCACTCAGACAC

Continued on next page...

Table A.4 – continued from previous page

Staple name	Sequence
	CCGGAATAAGTTTGGGC
68[23]62[4]	GTTTAGACTAGTAAGAAACGCCAAAAGGAATTTTTTTTTTTT
0[151]5[135]	GAACCCTAAAGGGAGCCCCCGATGTACAACGAAAGGCCGAACCTGT C
41[9]20[28]	TTTTTTTTTTGAACCGATCATTACCGCGCAATAAGAAGCGTCTTTC CTCATTACTTGATATGCAT
50[55]35[64]	GCAGGTCGTCACGACGTGCATCCTCCCGAGAACGCCAGTTACAAAA TTTTTTTTTTT
69[68]73[63]	AACTAACAGTTGGAGAGTACTCGCGTTTTAATTCGATCAAAAAG
0[79]87[71]	ACCAATCACATAGTTAGCGTAAGTATAGATAACGATCTAAAGTTTT GTAGTAATTC
93[4]92[4]	TTTTTTTTTTCAGTAACAGTACCTTTACGTCAGATGAATATATTTT TTTTTT
79[92]4[84]	CCTAAGTAATAACCTTGCTGCTGGTCAGTTGGCAAAAACTCACAGG CG
1[168]4[163]	GAACGTGGTAGGGTTGTGGTTTGCCCCTTCACCGCCT
54[51]52[68]	TATTCGCATTTTGAGGATTCGAGGTGAATTTCTTAAACAGCTGCGC AAGTTAAATCAAAA
45[38]86[16]	CAATAAATGATAGTAGGAACCCATGTACCGTAACACTGACTT
30[64]29[59]	TTTTTTTTTTAAGAATTGAGTTATAAGAGTTTTTTTTTT
86[115]83[99]	TTTTTTTTTTCAGAACGCGCCTGTTCATGTTCGGAAGGGTGA
84[95]1[103]	AAGAAAGCATCTTTGATTACGAGCGCGCCGACAATGACA
73[40]68[48]	GCAAAGCGGATTGCAGCTTCAAAAATGGTTTCTGCG
56[29]64[40]	ACGTAACTGACGAGAAACACCAGAACGAGTGGGCTTGACTGGCTCA TGCTTTAC
46[86]5[71]	TTTTTTTTTTCGCCATTCAGGCTTGATACAGGGTAGCAACGGCTAG CCTAATG
57[3]52[9]	TTTTTTTTTTGTGAATAAGGCTTGCCCAAAGCTGAATCATGGTTCC TGTGTGATTTTTTTTTT
52[47]74[32]	CTGTAGCCGTTAAAATTTAAATTGGCATAAAGCTAAAATATCTTTA GG
52[33]50[9]	CATCAGCTGTTCATAACATTAAATGTAGGATCCCCGGGTACCGAGC TCGAATTTTTTTTTTT
92[36]93[42]	ATAACCAGGTTTATACATCGGGAGAAACAATA
48[71]47[86]	ATGGGCGCTCTCCGTGGGAACAAACGGCGGATAATAGGAAATTTTT TTTTTTT
91[12]94[37]	TTTTTTTTTTAAATTGCGTAGATTTTTATATGTAATT
80[120]95[120]	TTTTTTTTTTGTAGCAATACTTCTCATTTGAATTACTTTTTTTTTT
62[86]93[87]	TTTTTTTTTTTTAGAACCCTCAAAATCTACCACCAGCGAAAAACGA CACGACCAATCCTTGAAAACATA
16[19]15[25]	GACATTCACCGCTCAGAAACCGATTGAGGG
54[39]54[4]	TAAACGTTTTCCACACTCATCAAGAGTAATCTTTTTTTTTTT
8[48]39[35]	TTTTTTTTTTGCACCGCGACAGAACGTAGAAAAGAACTGTCACAAA CTTACCAGATTGCACCCAGCTACAAGGTG

Continued on next page...

Table A.4 – continued from previous page

Staple name	Sequence
16[54]27[47]	TTTTTTTTTTTATGGTTTACCAGCGCCAATGCCTTGATAGCAATA
81[88]0[80]	AGTCTGTCCGGAACGATTAATTGCAGACAGCATCGGAACGCGATAG TTATGTAGAA
71[31]73[39]	TCAGGGAAGAGACCAAAACGAGAATACTATTATAGTCAGAA
3[120]3[143]	GAAACAAATTAGAGCTCATAACCGATATATTCGGTCGCTGAGGGAG TTGAGATTTG
32[17]30[4]	CCTGAAGAGGTATTAACAAAGTCAGATTCTGAAACATGAAAGTTTT TTTTTT
94[83]83[95]	TATGAAAATTAAGGTTTGAAAAATTCTTACCAGTATGAGAATCGGG TTGCTT
2[167]5[183]	AGTGTTGTGCCAACGCGCGGGGAGAGGCGGTT
64[67]73[86]	TATGTGTTTAGCTGGCATCAATTAAGAGGAAGCCCGAAATTTTTTT TTT
77[60]81[51]	AGACAAACAGAGGTGAGGCCGCTGAGTCGGCCTAGA
52[67]65[86]	TAATCCGGTTGATCTGGAGCCGTTGGGAGCCGGAGAGAGCATAAAG CTAAATCGGTTTTTTTTTT
3[72]90[64]	ATACACTACGGTCAATACAGGGCGAGGCCGATCGTTATACATACCG ACCGTGTGAT
0[214]1[214]	TTTTTTTTTTGTGAACCATCACCCAAATCAAGTTTTGTCAAAGGGC GAAAAACCGTCTATCATTTTTTTTTT
73[4]72[4]	TTTTTTTTTTTCAGGTCTTTACCCTGGACCATAAATCAAAAATTTT TTTTTT
3[40]80[41]	GGCACCAACTTTGAAAGAGGACAGGAACAAAGGAATCCTGTGCTGGT
69[9]80[7]	TTTTTTTTTTATAGCGTCCAAACATTTGAAGATTAGAATTTTAAAA GTTTTTTTTTTTTT
84[120]87[115]	TTTTTTTTTTACGTGGCGAGAAAAGCTAATGTTTTTTTTTT
34[64]17[54]	TTTTTTTTTTAAACAGCCATATGTAAGCATACCGAATCATTTTTTT TTT
55[60]49[86]	TACCTCGCGTCTCGTCGGATATCGTAACCGTTGTAAAACGACGGCC AATAGGTCATTTTTTTTTT
0[183]3[167]	TTGGGGTCGAGGTGCCGTAAAGCAGTCCACGC
59[32]66[44]	TTTAAGAAGATGGTTTAGAATTAGCAAAATTAGTAG
95[7]89[24]	TTTTTTTTTTCGCAGAGGCGAATAAATAAAGCAAAATTAT
83[100]75[115]	CGAGACTTAGCCATCACGTCAATATAACCTCAAATATCAATTTTTT TTTT
84[71]81[79]	AGGGCGCTGCGCCGCTCATAAGGGATCAGTGAGGCCACCG
50[91]50[68]	TTTTTTTTTTTGACCGTAATGGGGTGCCAA
2[152]1[167]	TGGATTGCAGGCACAAGAGTCCACTATTAAA
42[70]41[70]	TTTTTTTTTTCCGCACTCATCAGAAGGCTTATTTTTTTTTTT
71[4]70[4]	TTTTTTTTTTCCCCCTCAAATGCTTTTAAATATTCATTGAATTTTT TTTTTT
61[9]74[12]	TTTTTTTTTTTACCACATTCACAACTAATTTTTTTTTTT
33[36]2[40]	CAGCCCAATCCACCAATAGCAAGCAAATATCGGTGCAGCTTGCTCT AAAGAC

Continued on next page...

Table A.4 – continued from previous page

Staple name	Sequence
48[86]53[91]	TTTTTTTTTTCGTTGGTGTAGAAAGCGCCCGCCATCAAGCTCATTTTTTAACCTTTTTTTTTT
60[91]61[91]	TTTTTTTTTTTCAACCGTTAATAAAAACCATCAATATGATATTTTTTTTTTT
33[44]0[28]	ACAGAGCAAGCCATAAAGAAATTCCATTTTTTCACGTTGA
4[143]5[127]	CCTGATAAATTGTGTCGAAATCCGAGTCGGGA
76[115]79[120]	TTTTTTTTTTTTAAAACATACGCTCAATCGTCTTTTTTTTTT
29[9]14[4]	TTTTTTTTTTGTTCCAGTAAGAAACAGTTATCACCGTCACAGAGCCGCCACCCTTTTTTTTTT
55[29]75[43]	TACCCATAAAACTAGCATGACAGGTAGAAGGGTGAAGTAATGTCAGT
75[44]68[24]	GCCAGGTCAGTATTAACACCTAGATAATTACTGCGGATATAATGGATTTAGTTTGAAAAT
2[99]3[119]	GAAGTTGCGCTCACTGCCCGCTTTCCCGACCTGCCCAAGCGC
28[35]44[40]	CAGTGAGCGCTAATATCAGAAGTTTCAG
4[162]0[152]	GGCCCTGAGAGAGTATCGTATCTGCAGCAAGCGCTAAATCG
39[9]22[9]	TTTTTTTTTTCCGCCAACCGCCCGGAATATTTTATCCGTTTTGCTCAGGAGGTTGAGGCAGGTCTTTTTTTTTT
67[33]66[9]	TTAGATAAACATCCAATAAATCATACAGTTTTGCTTTTTTTTTT
50[67]57[86]	GCTTTACAAAGGCTATCAGGTCATTTTTTTTTT
37[40]49[41]	GGCGGATACGAGAGGGTTGATATACGCCAGGGTT
87[26]88[38]	GAAGGGTAAGAGAATATAAAGTATTTTCGAGC
70[51]60[60]	CTTCAACATGTTTTAAATTAAATGCATTTATTTCCGGA
32[64]15[54]	TTTTTTTTTTAAAAACAGGGAAGCTATCTCAAGAAAAATTTTTTTTTTT
14[54]45[37]	TTTTTTTTTTTATTCATTAAAGGTGAATTAATGCCCCCTGCAATAAAGCCCCTATTTCGGACAAGCC
10[48]35[47]	TTTTTTTTTTATCACTTTTCGGTACAATCAACATATAAAAGTGTACTAAATAAATCCAGAGCCTAATTTGC
32[47]46[32]	GCGCATTAGACGGGAGTGCGAATAGTTTTTATGGTTTATCGGGCCTCT
45[9]28[9]	TTTTTTTTTTCCCTCATTTTCAGGGATAGACCTATTAGGGTAATTGCCCGTATCGTCATTTACCTTTTTTTTTT
5[72]69[67]	AGTGAGCTTCAACAGTTGAAAGGAGCAAATGAATATATTTATGC
90[37]4[4]	CACCGGTAAAACAGTATTCATTAGAAGGAGATTTTGCGATGAACGGTGTACAGACCAGGCGCTTTTTTTTTT
80[79]92[64]	GCCTGAGTAAGAAAACTGAGTGAAGAACGCGACTGAGAGACTACCTTT
39[36]38[9]	TATCTTAGCGAACTGCCAGTTTGAGGGGACGACGACTTGGGTAAAGTATAGCCACCCTCAGAATTTTTTTTTT
52[51]33[64]	TTCGCTTCTGAAGGGCGCAGATATGAGAACAAGAGAATAACATTTTTTTTTTT
93[43]92[37]	ACGGAGCTATTACGGCTTAGGTTGGGTTAT
68[91]69[91]	TTTTTTTTTTAAGTTTCATTCCATATAAAGTACGGTGTCTGGTTTT

Continued on next page...

Table A.4 – continued from previous page

Staple name	Sequence
	TTTTTT
82[120]89[115]	TTTTTTTTTTCTTTCCTCGTTAGACGCTCAATTTTTTTTTT
7[9]8[9]	TTTTTTTTTTCCATCGATAGCTGCCTTTAGCGTTTTTTTTTT
59[9]60[9]	TTTTTTTTTTGTGAATTACCTAGATTTAGGAATTTTTTTTTT
4[214]5[214]	TTTTTTTTTTTCTTTTCACCAGTGAGACGGGCAACATGCGTATTGG GCGCCAGGGTGGTTTTTTTTTTTTTT
88[37]82[44]	CAGTAATTAGAACCTCAGATGATGGCATAACCACCACAGTAC
23[9]24[9]	TTTTTTTTTTAGACGATTGGCAAGCCAGAATGTTTTTTTTTT
67[9]67[32]	TTTTTTTTTTCAGAGGGGGTAATAGTACCA
19[28]7[48]	GCAAATCAAGTTAGCACCGTAATCAGTAATTACTTTTTTTTTT
4[183]3[219]	GCTGATTGCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTCCGAA ATTTTTTTTTT

Table A.4: Staple sequences for square-lattice signpost origami nanostructure.

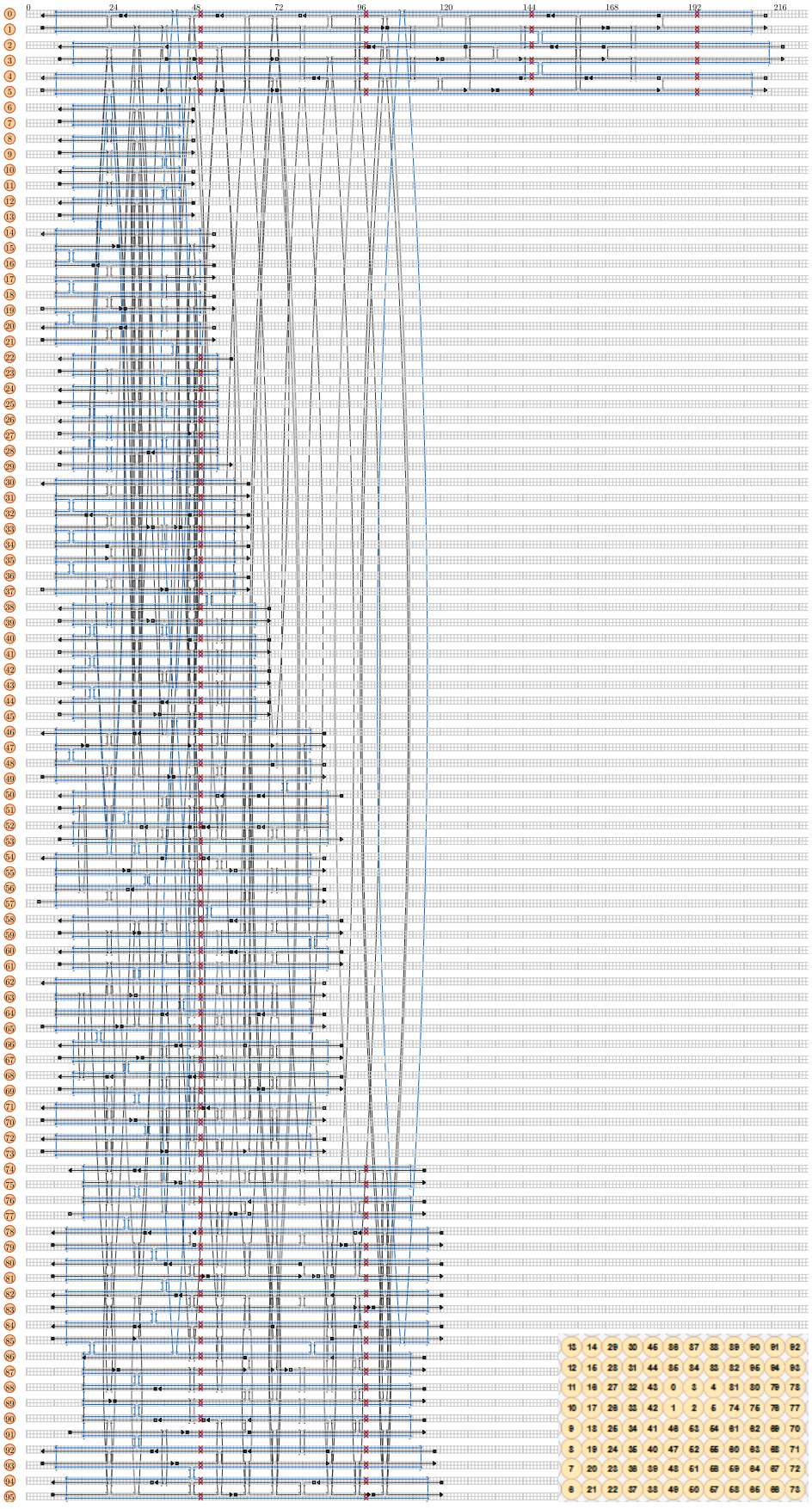


Figure A.4: Square-lattice signpost origami layout from caDNAno. In staple sequences the 4 nucleotide poly-T extensions shown are replaced by 10 nucleotide poly-T extensions. Inset shows arrangement of helices.

Appendix B

RNA Methods

B.1 Template Preparation

B.1.1 Short Templates

Templates for aptamer GA and GA-O were ordered as 2×4 nmole ultramer strands from IDT (template strand and reverse complement). Templates for the PX motif folding experiment were ordered as 2× oligonucleotides (template strand and reverse complement), 25 nmole scale and standard desalting purification. To prepare duplex template:

1. Suspend DNA strands in 1×TE at 200 μM concentration
2. Mix 4 μL of each strand is with 2 μL 5× duplex annealing buffer[1] to give a final duplex concentration of 80 μM.
3. Heat mixture to 94°C for 2 minutes, then cool to 5°C at 1°C/min.

B.1.2 Cloning

Most longer templates were cloned into plasmids, as the presence of shorter side products compromised PCRs with the gBlock as template. 0.3 μg of gblock template or 1 pmole of pUC19 vector is used in a 50 μL restriction enzyme digestion reaction. Cloning is performed as follows:

Step 1: Template Digestion/Phosphorylation

1. In a 200 μL Eppendorf tube at room temperature, mix: 10 μL of each restriction enzyme[2] (this may be 1 or 2 enzymes, depending on the presence of usable restriction sites at the 3′ and 5′ termini of the gblock), 5 μL of 10× NEBuffer (buffer choice

[1] 1× duplex annealing buffer: 100 mM Potassium Acetate, 30 mM HEPES, pH 7.5
[2] All enzymes purchased from New England Biolabs

determined using the NEB double digest tool) and 0.3 µg gblock template made up to 50 µL total volume with Milli-Q water.

2. Incubate for 4 hours at 37°C, followed by 20 minutes at 80°C to deactivate enzymes.
3. *If only one restriction enzyme used:* add 5 µL T4 PNK and 6 µL 10× T4 ligase buffer. Incubate at 37° for 120 minutes, followed by 20 minutes at 65°C.
4. Purify the sample by MWCO filtration.

Step 2: Vector Digestion/Dephosphorylation

1. In a 200 µL Eppendorf tube at room temperature, mix: 10 µL of each restriction enzyme (the same enzymes as used to digest the template, if only one restriction enzyme used also digest with HincII), 5 µL of 10× NEBuffer enzyme buffer (buffer choice determined using the NEB double digest tool) and 1 µL of 1 µM double-stranded pUC19 (also NEB).
2. Incubate for 4 hours at 37°C, followed by 20 minutes at 80°C to deactivate enzymes.
3. 5 µL Antarctic phosphatase and 6 µL 10× Ant phos buffer. Incubate at 37° for 120 minutes, followed by 20 minutes at 80°C.
4. Purify the sample by MWCO filtration (Appendix C.2.1).

Step 3: Ligation

1. Mix the insert and vector at a 6:1 molar ratio.
2. Add 17 µL of the mix to 2 µL 10× T4 ligase and 1 µL T4 Ligase.
3. Incubate for 10 minutes at 4°C, followed by 10 minutes at 25°C, then heat inactivation at 65° for 10 minutes.

Step 4: Cell transformation and template Preparation

1. Thaw 50 µL competent cells (DH5α) on ice for 10 minutes.
2. Add 5 µL of ligated DNA to the cells. Mix by flicking the tube gently.
3. Place the mixture on ice for 30 minutes.
4. Heat shock in a water bath at 42°C for 30 seconds.
5. Incubate on ice for 5 minutes.
6. Add 100 µL of room temperature SOC media to the tube.[3]

[3]Note: no 37°C recovery period is used for Ampicillin selection.

7. Prepare 4 serial dilutions of the transformed cells in SOC.
8. Plate 100 µL of each dilution onto a new Ampicillin Agar plate.
9. Incubate plates lid-side down at 37°C overnight. From this stage, template is prepared exactly according to the instructions of the QIAprep Spin Miniprep Kit.

The restriction enzymes used for each nanostructure template are given in the main thesis text, but are also provided in Table B.1 for reference.

gBlock	**Restriction Enzyme 1**	**Restriction Enzyme 2**	**NEBuffer**
Pentomino Nanostructure B, C, E	EcoRI	HindIII	2.1
Nanostructure A	-	HindIII	2.1
Origami$_{PX}$	EcoRI	-	EcoRI buffer

Table B.1: Restriction enzymes used for cloning different nanostructure templates.

B.1.3 PCR

PCR reactions are performed with plasmid or gblock templates. The following protocol is used (can be scaled up to 50 µL, for larger volumes the reaction mix is aliquotted into multiple tubes):

1. Mix reactions components on ice (Table B.2).

Component	**Volume/Amount**
Q5 High-Fidelity 2× Master Mix (NEB)	12.5 µL
Forward primer (FP), 10 µM	1.25 µL
Reverse primer (RP), 10 µM	1.25 µL
DNA template	1 ng
Milli-Q water	to 25 µL

Table B.2: PCR reaction components.

2. Transfer to a pre-heated (98°C) PCR machine and begin thermocycling (Table B.3).
3. Purify PCR product using Qiagen PCR Purification kit, eluting into a final volume of 30-50 µL.

PCR conditions for templates in this thesis are given in Table B.4.

Step	**Temperature**	**Time**
Initial Denaturation	98°C	2 minutes
	98°C	10 seconds
25 Cycles	50-72°C (T_m)[4]	30 seconds
	72°C	30 seconds/kb
Final Extension	72°C	2 minutes
Hold	10°C	

Table B.3: PCR thermocycling conditions

Product Name	**Template DNA**	**Primers**	**T_m**
Pentomino Unit 1	Pentomino gblock	FP: ACGACGGCCAGTGAATTCTAAT	67
		RP: ACGACGGCCAGTGAATTCTAAT	
Pentomino Units 1-2	Pentomino gblock	FP: ACGACGGCCAGTGAATTCTAAT	67
		RP: CTTTGTTTGATCACAGTCGCCC	
Pentomino Units 1-3	Pentomino gblock	FP: ACGACGGCCAGTGAATTCTAAT	67
		RP: GCTTGGTAAACGTGTCGCC	
Pentomino Units 1-4	Pentomino gblock	FP: ACGACGGCCAGTGAATTCTAAT	67
		RP: CTTAACACGTCGCCCATAAGC	
Unit α	Unit α gblock	FP: GCATACGTCAGAATTCTAATACGACTCACTATAGGGAGA	72
		RP: TCTGTGGTTCGTCCGGGTCA	
Unit β	Unit β gblock	FP: TTTCAGCAAGATGTTTACGTCAGAATTCTAATACG	69
		RP: AAAGAACCAGTACTTTGTAAGTCACGCATATG	
NS D	NS D gblock	FP: TGCATGATCTACGTGCGTCA	67
		RP: GAGAGACCTTGCAGCCTGTT	
NS E-apt Plasmid*	NS E plasmid	FP: TTGGGTGGGGAAAGTCCTTAAAAGAGGGCCACCACAGAAGCTGTCAGAGGTCTCACGAATC	64
		RP: TTGCGCCACGACGTTTC	

Continued on next page...

Table B.4 – continued from previous page

Staple name	Sequence			
			GTGCTCCCATCGCAGTC CAGAAGCTATCAGCCGA TAGTCGTCC	
NS D-FL	NS D gblock	FP:	CTAATACGACTCACTAT AGGGAGACCGAAGAC	70
		RP:	GCTTCGTCTGCTGAGAG ACCTTGCAGCCTGTTGAC	
NS E-FL	NS E Plasmid	FP:	CTAATACGACTCACTATA GGGAGACGAACCCA	71
		RP:	GCTTCGTCTGCTCCCAG ACCTTGCAGCGACTTG	
SDM1 Plasmid*	$\text{Origami}_{\text{PX}}$ plasmid 1[5]	FP:	CTTATTACGCGACTGCC TCCTTG	64
		RP:	TCCACAATTTTGCACGGG	
SDM2 Plasmid*	SDM1 Plasmid	FP:	ACTCCACAGTCTGGCA GGCCC	66
		RP:	GGTACATCTTCGGAT GTCGG	
$\text{Origami}_{\text{PX}}$ plasmid 2[6]*	SDM2 Plasmid		GAGCGGTACCTCCACC GACAT	67
		RP:	TCCTCTCTATTGGTTT GGACCATTAATC	
$\text{Origami}_{\text{BLKL}}-\text{s1}$	$\text{Origami}_{\text{BLKL}}$-ssl gblock	FP:	CTAATACGACTCACTAT AGGGAGAGCGTATGTAC	69
		RP:	GATAGAATGGAAGCTCT GAGGACAGATC	
$\text{Origami}_{\text{BLKL}}-\text{s2}$	$\text{Origami}_{\text{BLKL}}$-ssl gblock	FP:	CTAATACGACTCACTAT AGGGAGACGGGAGATTC AGGGTC	72
		RP:	GACTCATGGGATCAATA CGAACCAGCACAAC	
$\text{Origami}_{\text{BLKL}}$-5′	$\text{Origami}_{\text{BLKL}}$-5′ gblock	FP:	CTAATACGACTCACTAT AGGGAGACGTAGCGTAT	71
		RP:	GGGCGGAAGCTCTGAGGA	
$\text{Origami}_{\text{BLKL}}$-3′	$\text{Origami}_{\text{BLKL}}$-3′ gblock	FP:	CTAATACGACTCACTAT AGGGAGAGGCCA	71
		RP:	GACTGATGATCTCCCGCT GCC	
$\text{Origami}_{\text{BLKL}}$-N	$\text{Origami}_{\text{BLKL}}$-N gblock	FP:	CTAATACGACTCACTATA GGGAGAGAGATCATCAGT	71
		RP:	GGTATACGCTACGTCTT ACAGGTGGAAG	

Continued on next page...

[5]Plasmid from original miniprep contains three point mutations relative to designed sequence.
[6]This plasmid was linearised and used as the transcription template.

Table B.4 – continued from previous page
Staple name Sequence
*Indicated PCR reactions are performed as a step in site-directed mutagenesis (Q5 site-directed mutagenesis kit, NEB).

Table B.4: Conditions for PCR reactions for this thesis.

B.2 Transcription Reactions

B.2.1 Short RNA Transcription

Template sequences for aptamers and PX motifs were ordered as described in Appendix B.1.1. Short RNAs were transcribed using the NEB HiScribe T7 transcription kit. A master mix of equal volumes of ATP, CTP, UTP, GTP and 10×T7 buffer was prepared to reduce pipetting errors. For each transcription reaction the following protocol was used:

1. In a 500 µL Eppendorf tube, mix: 15 µL master mix, 20 µL nuclease free water (DEPC free, invitrogen), 2 µl 80 µM DNA template and 3 µL T7 polymerase mix.
2. Incubate for 16 hours at 37°C.
3. Add: 140 µL nuclease free water, 20 µL 10× DNase I buffer (New England Biolabs), 10 µL DNase I (New England Biolabs) and mix by pipetting.
4. Incubate for 30 minutes at 37°C.

B.2.2 Nanostructure Transcription

Nanostructures were transcribed using the NEB HiScribe T7 transcription kit. A master mix of equal volumes of ATP, CTP, UTP, GTP and 10×T7 buffer was prepared to reduce pipetting errors. For each transcription reaction the following protocol was used:

1. Add 10-100 ng of template DNA to a 200 µL tube, and adjust volume to 8 µL with nuclease free water (DEPC free, invitrogen).
2. Add 10 µL master mix, and 2 µL T7 polymerase mix.
3. Incubate for 1-4 hours at 37°C.
4. Add: 80 µL nuclease free water, 10 µL 10× DNase I buffer (New England Biolabs), 1 µL DNase I (New England Biolabs) and mix by pipetting.
5. Incubate for 15 minutes at 37°C.

B.3 Purification

B.3.1 Column Purification

Column purification of RNA nanostructures was performed with MEGAclear™ Kit (Ambion) as described in the user guide. All buffers (after adding ethanol to the wash solution concentrate) as well as the 100% ethanol were supplemented with 5 mM $MgCl_2$.

B.3.2 LiCl Precipitation

Purification of RNA nanostructures by LiCl precipitation was performed as follows:

1. Add 1/2 volume 7.5 M LiCl (Invitrogen).
2. Incubate at -20°C for 30 minutes.
3. Spin in a pre-chilled centrifuge at 20,000×g for 20 minutes at 4°C.
4. Wash with 70% ethanol at 20,000×g for 10 minutes at 4°C.
5. Remove excess ethanol by pipetting, and repeat wash step.
6. Remove excess ethanol by pipetting and dry sample for 10 minutes at 50°C with the lid off.
7. Resuspend to desired volume in buffer.

B.3.3 Ethanol Precipitation

Due to their shorter sequences, yield from LiCl precipitation and column purification of RNA aptamers and PX_{major} motif strands was very low. Instead, these short RNAs were purified by Ethanol precipitation. The purification protocol for a 100 µL sample of transcribed aptamer (post DNase I treatment) is as follows:

1. Add 12.5 µL sodium acetate (pH 5.5, 3 M).
2. Add 375 µL 100 % ethanol.
3. Vortex for 10 s.
4. Incubate at -80°C for 2 hours.
5. Spin in a pre-chilled centrifuge at 13,000×g for 20 minutes at 4°C.
6. Wash with 50 µL of 95% ethanol at 13,000×g for 10 minutes at 4°C.

7. Remove excess ethanol by pipetting, and dry sample for 10 minutes at 50°C with the lid off.

8. Re-suspend to desired volume in nuclease-free water.

Appendix C

Origami Methods

C.1 Annealing

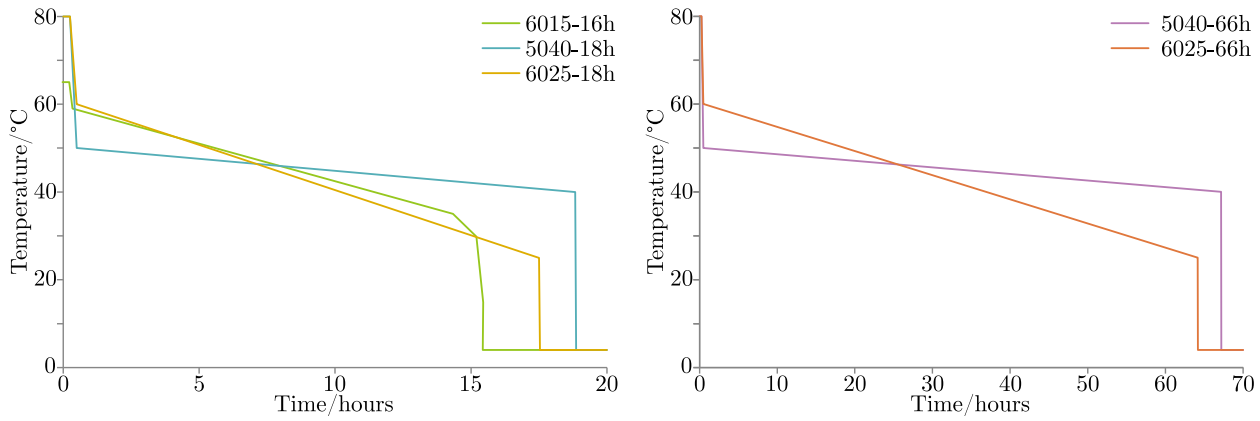


Figure C.1: Thermal annealing ramps for origami. The main protocols used are 6025-16h for honeycomb structures and 5040-18h for the square lattice signpost origami.

C.1.1 Honeycomb Lattice Origami

Honeycomb lattice origami nanostructures were folded with a 10× excess of staples at a scaffold concentration of between 10 and 40 nM in 1×TE + 12.5 mM $MgCl_2$. The thermal ramp for honeycomb lattice origami nanostructures (6015-16h) was devised by Robert Schreiber (a previous group member). The steps are:

6015-16h

1. Heating at 65°C for 15 minutes.
2. Cooling from 65°C to 59°C at 1°C every 5 minutes.
3. Cooling from 59°C to 35°C at 1°C every 35 minutes.

4. Cooling from 35°C to 30°C at 1°C every 10 minutes.
5. Cooling from 30°C to 15°C at 1°C every 5 minutes.
6. Hold at 4°C.

Total annealing time is roughly 16 hours.

C.1.2 Square Lattice Origami

The square lattice signpost origami was folded with a 10× excess of staples at a scaffold concentration of 20 nM in 1×TE + 15 mM $MgCl_2$. Four different annealing ramps were tested, and the 5040-18h program was selected. Suggestions for thermal ramps tested for the square lattice origami nanostructures were provided by William Shih. The steps are:

5040-18h

1. Heating at 80°C for 15 minutes.
2. Cooling from 80°C to 50°C at 1°C every 30 seconds.
3. Cooling from 50°C to 40°C at 1°C every 110 minutes.
4. Hold at 4°C.

Total annealing time is roughly 18 hours.

6025-18h

1. Heating at 80°C for 15 minutes.
2. Cooling from 80°C to 60°C at 1°C every 45 seconds.
3. Cooling from 60°C to 25°C at 1°C every 30 minutes.
4. Hold at 4°C.

Total annealing time is roughly 18 hours.

5040-66h

1. Heating at 80°C for 15 minutes.
2. Cooling from 80°C to 50°C at 1°C every 30 seconds.
3. Cooling from 50°C to 40°C at 1°C every 400 minutes.

4. Hold at 4°C.

Total annealing time is roughly 66 hours.

6025-66h

1. Heating at 80°C for 15 minutes.
2. Cooling from 80°C to 60°C at 1°C every 45 seconds.
3. Cooling from 60°C to 25°C at 1°C every 110 minutes.
4. Hold at 4°C.

Total annealing time is roughly 66 hours.

C.2 Purification

C.2.1 MWCO Filtration

Molecular-weight cut-off (MWCO) filtration was performed as follows:

1. Make up origami to a final volume of 500 µL in 1× origami folding buffer (unless otherwise specified).
2. Apply sample to a 100 kDa Amicon® Ultra-0.5 centrifugal filter device (Merck).
3. Spin for 15 minutes at 14,000×g.
4. Empty flow through and add 485 µL 1× origami folding buffer (unless otherwise specified).
5. Repeat steps 3, 4 and 3 again.
6. Turn device upside down in a clean tube.
7. Spin for 2 minutes at 1000×g.

C.2.2 PEG Precipitation

Prior to purification 2×PEG-precipitation buffer is prepared, containing 1×TE, 12.5 mM $MgCl_2$, 500 mM NaCl, 15% weight/volume PEG 8000.

1. Mix unpurified origami 1:1 with 2×PEG-precipitation buffer, mix thoroughly by pipetting.
2. Spin at 20,000×g for 20 minutes at 20°.
3. Remove supernatant by pipetting, and resuspend origami pellet in 1× origami buffer.
4. Repeat steps 1-3 two more times.

C.2.3 Ethanol Precipitation

Prior to purification, prepare 3 M Sodium Acetate (pH 5.2) and cool 100% ethanol in -20°C freezer.

1. Add 1/10 volume of 3 M Sodium Acetate (pH 5.2).
2. Add 2.5 volumes of cold 100% ethanol.
3. Incubate at -20° overnight.
4. Centrifuge mix at 14,000×g for 20 minutes at 4°C.
5. Remove ethanol by pipetting, and add 200 µL room temperature 70% ethanol.
6. Centrifuge the pellet again at 14,000×g for 5 minutes at 4°C to wash.
7. Remove ethanol by pipetting and leave pellet to air dry for at least 10 minutes.
8. Resuspend to desired volume/concentration in 1×TAE + 12.5 mM $MgCl_2$

C.2.4 Column Purificiation

Origami are purified with the QIAquick PCR purification kit following PEG precipitation to prepare samples for cryo-EM. Prior to purification, supplement all buffers with 5 mM $MgCl_2$.

1. Add 5 volumes of buffer PB to 1 volume of origami.
2. Spin for 1 minute at 17,900×g, and empty flow-through.
3. Add 750 µL buffer PE.
4. Spin for 1 minute at 17,900×g, and empty flow-through.
5. Spin again to remove all traces of ethanol.
6. Place column in a clean 1.5 mL Eppendorf tube, and add 30 µL buffer EB.
7. Leave for 1 minute, then spin for 1 minute at 17,900×g.

Appendix D

Microscopy

D.1 Transmission Electron Microscopy

D.1.1 Grid Preparation: Negative Stain

Carbon type-B: Formvar supported copper grids (Agar Scientific) were used for TEM. Grids are glow-discharged, and samples were applied according to the following steps:

1. Fix clean parafilm (Merck) to the lab bench with a layer of water.
2. Apply 10 μL droplets of sample to the parafilm.
3. Apply 20 μL droplets of 2% Uranyl acetate to the parafilm next to each sample droplet.
4. Place a grid is placed carbon-side down on the sample droplet for 2 minutes (hold grid with negative pressure tweezers at edge).
5. Blot sample from grid (with the grid perpendicular to the filter paper).
6. Place grid on the adjacent Uranyl acetate droplet for 10 s.
7. Blot grid a second time, and leave to dry for at least 5 minutes before storing.

D.1.2 Grid Preparation: Positive Stain

Carbon type-B: Formvar supported copper grids (Agar Scientific) were used for TEM. Grids are glow-discharged, and samples were applied according to the following steps:

1. Glow discharge grids with a plasma cleaner (20s, 15 mA).
2. Fix clean parafilm to the lab bench with a layer of water.
3. Apply 10 μL droplets of sample are to parafilm.

4. Apply 20 μL droplets of 2% Uranyl acetate to parafilm next to each sample droplet.
5. Apply 20 μL droplets of water next to each Uranyl acetate droplet.
6. Place grid is carbon-side down on the sample droplet for 2 minutes using negative pressure tweezers.
7. Blot sample off of grid with the grid perpendicular to the filter paper with section of filter paper (Whatman).
8. Place grid is on the adjacent Uranyl acetate droplet for 20 s.
9. Blot grid is a second time.
10. Place grid on the adjacent water droplet for 10 s.
11. Blot grid a third time and leave to dry for at least 5 minutes before storing in grid box.

D.1.3 Imaging

TEM imaging was performed using an FEI Tecnai T12 microscope, operated at 120 kV, with a Gatan OneView CMOS camera.

D.2 Atomic Force Microscopy

Atomic force microscopy was performed with the supervision of Natalie Haley, who also provided this protocol.[179] An Asylum Cypher ES microscope was used with AC40TS tips. 10 μL 5 nM origami in 1×TAE + 12.5 mM $MgCl_2$ was deposited on freshly cleaved mica and left for 2 minutes. After deposition, 100 μL AFM buffer[1] was added between the sample and the cantilever for imaging. Images were taken with tapping mode in fluid.

D.3 Cryo-Electron Microscopy

D.3.1 Cell Lysate Preparation

Three types of lysate have been used for figures in this thesis: mammalian cell lysate by freeze-thawing (Figure 2.13, 2.16, 2.17, 2.28), bacterial cell lysate by freeze-thawing (Figure 2.22,2.6.1) (Figure 2.6.1)and bacterial cell lysate from a 'continuous flow homogenizer' (Figure 2.3.4).

[1] 1 × TAE, pH 8.3, 12.5 mM $MgAc_2$, 4 mM $NiCl_2$.

Freeze-thawing

1. Spin down cells (200×g for mammalian cells, 4000×g for *E. coli*) for 10 minutes at 4°C.
2. Remove medium and resuspend cells in 5-10× less buffer (50 mM Tris, 150 mM NaCl).
3. Place in -80°C for 30 minutes, then thaw.
4. Repeat 3×.
5. Supernatant is used as lysate.

Continuous flow homogenizer

1. 50 mL cells are harvested by centrifugation at 4,000×g for 10 minutes at 4°C and resuspended in 10 mL cold lysis buffer (20 mM Tris pH 7,4, 150 mM NaCl in H_2O).
2. Cells are lysed in a constant flow cell disruptor (Constant Systems) at 8,000 psi without the addition of lysozyme.
3. Cell debris was removed by centrifugation at 4,000×g for 10 minutes.
4. Supernatant is used as lysate. (Membranes can also be removed from lysate with an additional 1 hour spin at 100,000 × g).

D.3.2 Grid Preparation

Quanitfoil GmbH (2/1 holey carbon support film) grids were used for cryo-EM. Grids are glow-discharged, and samples were applied according to the following steps:

1. Condense a mixture of 63% propane, 37% ethane (BOC gasses) into liquid using a metal cup cooled in liquid nitrogen.[180]
2. Pick up edge of cleaned grid with negative pressure tweezers, and clamp in plunger mounted above cryogen cup.
3. Apply 4 µL of sample to the grid with a pipetter, without touching grid surface. If required, add 0.8 µL of 10 nm Au nano-particles coated with BSA (Aurion) as fiducial markers.
4. Remove excess solution by blotting the back of the grid with Whatman No. 1 filter paper (Whatman) for 8 s.
5. Release plunger so that grid is rapidly dropped into the cryogen bath.
6. Transfer the grid using tweezers from the cryogen to liquid nitrogen, and store in grid box under liquid nitrogen until use.

D.3.3 Imaging

Cryo-EM was performed using Titan Krios (ThermoFisher Scientific)[2] or a TF30 (FEI)[3]. Both microscopes are equipped with a K2 direct electron detector (Gatan) mounted below a Quantum post-column energy filter (Gatan). Electron doses are between 10-50 $e^{-}Å^{2}$ for single images (typically 10-20 $e^{-}Å^{2}$). For tomography, images are taken at +/- 45° or +/- 60° with a 3° step size. Total dose in the tomograms is 65-200 $e^{-}Å^{2}$, evenly distributed over all tilts. The tilts series are bi-directional. Data are collected using SerialEM[181]. Tomograms were built using IMOD[182].

D.4 Fluorescence Microscopy

For fluorescent microscopy, an AxioObserver Z1 inverted microscope (Zeiss) equipped with a 63× (NA 1.4) oil objective and a monochrome AxioXam MRm camera (Zeiss) was used. A mercury arc lamp (HXP) was used for fluorescent excitation. Green (482/18 (Ex), 520/28 (Em)) and red (560/40 (Ex) and 630/75 (Em)) band pass filter cubes (AHF) were used for Dio (membrane) and Cy3 (origami) visualisation respectively. A halogen lamp was used for bright field microscopy.

[2]Used for Figures 2.22, 2.24, 2.28 and 3.10

[3]Used for Figures 3.9, 2.3.4, 2.13, 2.16 and 2.17

Appendix E

Gel Electrophoresis

E.1 Agarose Gel Electrophoresis

Agarose gels are prepared by mixing 100 ml of running buffer with agarose to give the desired weight/volume percentage agarose (0.8-2%). The mixture is heated in the microwave until the agarose is fully dissolved. The gel mixture is then cooled to $\approx 50°C$ and poured into a level gel-casting tray containing 1 or 2 combs. When completely set, the gel (in the tray) is transferred to a wide Mini-Sub cell GT electrophoresis tank (Bio-Rad) and submerged in running buffer (pre-cooled for gels run at 4°C). The comb is removed and 0.5-20 µL of samples (supplemented with final concentration 1× loading dye NEB purple) is pipetted into the wells.

For DNA origami gels in this thesis (Chapters 2 and 3), 1×TAE (ITW reagents) + 11 mM $MgCl_2$ was used as the running buffer and 2% agarose. For RNA nanostructure gels in this thesis (Chapters 4 and 5), 1×TBE (ITW reagents) + 3 mM $MgCl_2$ was used as the running buffer and 1% agarose. Both gel types were run at 60 V for 120 minutes at 4°C. Gels for analysing PCR and cloning products (not presented in thesis) used 1×TAE as the running buffer and 0.8% Agarose, and were run at 80 V for 60 minutes at room temperature.

If required, gels are post-stained with ≈1/1000 Sybr gold (Life Technologies) in 1×TAE for 40 minutes and de-stained in 1×TAE for 20 minutes.

E.2 Polyacrylamide Gel Electrophoresis

For native polyacrylamide gel electrophoresis, a 30% 29:1 Acrylamide:bis-acrylamide stock solution (Sigma) was mixed with running buffer components and water to give a final 1× buffer concentration, and the desired acrylamide concentration. 5.5 mL of mixture is prepared per plate. 1/100 volume of 10% ammonium persulphate (Sigma) and 1/1000 volume of tetramethylethylenediamine (Sigma) is added and the solution is mixed by vortexing. The mix is then poured between two glass plates with a 0.75 mm spacing, a gel comb is added and the gel is left for at least 10 minutes to polymerise. Gels are then transferred to a

Mini-PROTEAN tetra cell vertical gel electrophoresis system (Bio-Rad) and sub-merged in pre-cooled running buffer. The comb is removed and 0.5-10 mL of samples (supplemented with final concentration 1× loading dye NEB purple) is pipetted into the wells.

For aptamer gels in this thesis (Chapter 3) 0.5× TBE (ITW reagents) + 5 mM $MgCl_2$ was used as the running buffer, and 6% acrylamide was used. These gels were run at 180 V for 40 minutes at 4°C. For the PX_{major} motif gel (Figure 5.4) 1×TAE (ITW reagents) and 10% acrylamide were used. The gel was run at 200 V for 30 minutes at 4°C.

If required, gels are post-stained with ≈1/1000 Sybr gold (Life technologies) in water for 5 minutes and rinsed with water.

E.3 Visualising Gels

Gels were scanned with a PharosFX molecular imager (Bio-Rad) or an Amersham Typhoon (GE). Laser/filter combinations are shown in Table E.1.

	Bio-Rad scanner		**GE scanner**	
Fluorophore	**Laser λ/nm**	**Filter bandpass /nm**	**Laser λ/nm**	**Filter bandpass /nm**
sybr gold, FAM, GFP variants	488	530	488	525
mCherry	-	-	532	670
Alexa647	-	-	635	670

Table E.1: Lasers and filters for scanning gels.

Appendix F

Matlab Scripts

F.1 Worm-like Chain

MATLAB script for simulating worm-like chain trajectories. The output `coords` is a matrix of three columns corresponding to x, y and z coordinates of the trajectory. The output `enderror` gives the distance in nanometres between the end of the trajectory compared to a perfect (completely stiff) trajectory of the same length.

```
N = 150; % number of base pairs
r = 0.34; % helical rise in nm
P = 2400; % persistence length in nm
x = zeros(N,1); y = x; z = x;
coords = zeros(N, 6);
coords(1,6) = 1;

for i = 2:N+1
    v1 = coords(i-1,4); v2 = coords(i-1,5); v3 = coords(i-1,6);
    theta = normrnd(0, sqrt(r/P));
    phi = (rand-0.5)*2*pi;
    dzb = r*cos(theta); dxb = r*cos(phi)*sin(theta);
    dyb = r*sin(phi)*sin(theta);
    dz = -dxb*v1-dyb*v2+dzb*v3;
    dx = dxb*v3 + dzb*v1;
    dy = dyb*v3+dzb*v2;
    coords(i,1) = coords(i-1,1)+dx;
    coords(i,2) = coords(i-1,2)+dy;
    coords(i,3) = coords(i-1,3)+dz;
    v1new = coords(i,1)-coords(i-1,1);
    v2new = coords(i,2)-coords(i-1,2);
    v3new = coords(i,3)-coords(i-1,3);
    norm = sqrt(v1new^2+v2new^2+v3new^2);
```

```
    coords(i,4) = v1new/norm; coords(i,5) = v2new/norm;
    coords(i,6) = v3new/norm;
end
plot3(coords(:,1),coords(:,2),coords(:,3));
enderror = sqrt(coords(N+1,1)^2 + coords(N+1,2)^2 + (N*r-coords(N+1,3))^2);
```

F.2 Ligate

MATLAB script for ligating an RNA strand from a PDB file of multiple strands. The input PDB file must have nucleotides numbered from 1 upwards in the intended $5' \rightarrow 3'$ strand routing (e.g. using the 'edit → rename current layer' function in Swiss-PdbViewer), and all HETATM fields must be replaced with ATOM fields (e.g. find and replace 'HETATM' with 'ATOM ' in PDB text file). Operation of this script requires mat2pdb.m (Mathworks).

```
cd('path_to_pdb'); %insert path to file here
PDBStruct_in = pdbread('infile.pdb'); % insert input filename here
PDBStruct_out = PDBStruct_in;
resSeq = []
data = struct()
data.outfile = 'outfile.pdb'; % insert output filename here
for i = 1:max(size(PDBStruct_in.Model.Atom))
    resSeq(i) = PDBStruct_in.Model.Atom(i).resSeq;
end
datav = struct()
X = zeros(1,length(resSeq));
Y = zeros(1,length(resSeq));
Z = zeros(1,length(resSeq));
resNum = zeros(1,length(resSeq));
occupancy = zeros(1,length(resSeq));
resName = cell(1,length(resSeq));
atomName = cell(1,length(resSeq));
element = cell(1,length(resSeq));
AtomNo = 1:length(resSeq);
w=1;
v=1;
for j = min(resSeq):max(resSeq)
  index = resSeq == j;
  atomtmp = AtomNo(index);
  for k = min(atomtmp):max(atomtmp)
      resNum(w) = PDBStruct_in.Model.Atom(k).resSeq;
      atomNamew = PDBStruct_in.Model.Atom(k).AtomName;
      PDBStruct_out.Model.Atom(w).altLoc =
```

```
PDBStruct_in.Model.Atom(k).altLoc;
      resNamew = PDBStruct_in.Model.Atom(k).resName;
      X(w) = PDBStruct_in.Model.Atom(k).X;
      Y(w) = PDBStruct_in.Model.Atom(k).Y;
      Z(w) = PDBStruct_in.Model.Atom(k).Z;
      occupancy(w) = PDBStruct_in.Model.Atom(k).occupancy;
      elementw = ' ';
      PDBStruct_out.Model.Atom(w).charge =
PDBStruct_in.Model.Atom(k).charge;
      w = w+1;
  end
  v = v+length(atomtmp);
end
data.X = X;
data.Y = Y;
data.Z = Z;
data.atomName = atomName;
data.resName = resName;
data.resNum = resNum;
data.element = element;
data.betaFactor = ones(1, length(X)).*100;
mat2pdb(data);
```

F.3 Gel Peak Analysis

MATLAB script for visualising gels quantitatively. Firstly, plot profiles on ImageJ[131] (avoiding obvious dirt/contamination spots) and save as .csv files with the name 'lane n.csv' (n being the lane number).

```
cd('path_to_csvs'); %insert path to csv files here
bands = zeros(0,5);

for n = [1:7] % change this depending on number of lanes
   Filename = ['lane ',num2str(n),'.csv']
   xy = csvread(Filename,1,0)
   x = xy(:,1); y = -xy(:,2);
   [pks,locs,w,p] = findpeaks(y,x);
   p = p/sum(p);
   tmpbands = [ones(length(pks),1).*n,pks, locs, w, p]
   bands = [bands;tmpbands];
end
```

```
xgraph = bands(:,1);
ygraph = bands(:,3);
sizegraph = bands(:,5)./max(bands(:,5)).*1000;
refsize = (max(ygraph(xgraph==1)));
ygraph = ygraph./refsize;

scatter(xgraph,ygraph,sizegraph,'.','k')
xlabel('Lane')
ylabel('Relative Band Mobility')
```

F.4 Profile Analysis

MATLAB script for analysing side profiles of mixed lattice origami (see Figure 2.29). Firstly, plot profiles on ImageJ, and save as .csv files with the name 'Valuesn' (n being the profile number).

```
cd('path_to_csvs'); %insert path to csv files here
p = 56 % specify number of profiles
r = 0.1523 % specify nm/pixel from image
meansep = []
id = []
contrast = []
for n = 1:p
   Filename = ['Values',num2str(n),'.csv']
   xy = csvread(Filename,1,0)
   x = xy(:,1); y = smooth(xy(:,2),10);
   [pks,locs] = findpeaks(y,x);
   sep = locs(2:end) - locs(1:end-1);
   meansep = [meansep; mean(sep)]
   if(mean(sep))*r<3 % this was the determined cut-off for the two lattices
     id = [id;1] % id = 1 for square lattice
   else
     id = [id;0] % id = 0 for honeycomb lattice
   end
   [negpks,neglocs] = findpeaks(-y,x);
   contrast = [contrast; mean(pks)+mean(negpks)];
end

figure(1)
meansep = meansep*r
histogram(meansep,12)
xlabel('Mean Peak Separation/nm')
ylabel('Frequency')
```

```
figure(2)
normcontrast=contrast/max(contrast)
index1 = id == 1
index2 = id == 0
histogram(normcontrast(index1),0.2:0.1:1,'Normalization','probability')
hold on
histogram(normcontrast(index2),0.2:0.1:1,'Normalization','probability')
hold off
lgnd = legend('Square Lattice', 'Honeycomb Lattice')
xlabel('Normalised Contrast')
ylabel('Probability')
```

Bibliography

[1] Ruska H. *Über ein neues bei der bakteriophagen Lyse auftretendes Formelement.* Naturwissenschaften 29:367, 1941

[2] Ruska H., Borries B. and Ruska, E.Ruska H., Borries B. and Ruska, E. *Die Beduntung der Übermikroskopie für die Virusforschung.* Archiv für die gesamte Virusforschung 1:155, 1939

[3] Heuser J. *Whatever happened to the 'microtrabecular concept'?* Biology of the cell 94:561, 2002

[4] Brenner S. and Horne R. W. *A negative staining method for high resolution electron microscopy of viruses.* Biochimica et biophysica acta 34:103, 1959

[5] Dubochet J., Booy F. P., Freeman R., Jones A.V. and Water C. A. *Low temperature electron microscopy.* Annu. Rev. Biophys. Bioeng. 10:133, 1981

[6] Adrab M. Dubochet J. Lepault J. and McDowall A. W. *Cryo-electron microscopy of viruses.* Nature 308:32, 1984

[7] Dubochet J., Adran M., Chang J., Homo J. *Cryo-electron microscopy of vitrified specimens.* Q. Rev. Biphys. 21:129, 1988

[8] Henderson R. and Unwin P. N. *Three-dimensional model of purple membrane obtained by electron microscopy.* Nature 257:28, 1975

[9] Unwin P. N. and Henderson R. *Molecular structure determination by electron microscopy of unstained crystalline specimens.* J. Mol. Biol. 94:425, 1974

[10] Taylor K. A. and Glaeser R. M. *Electron diffraction of frozen, hydrated protein crystals.* Science 186:1036, 1974

[11] Hayways S. B. and Glaeser R. M. *Radiation damage of purple membrane at low temperature* Ultramicroscopy 4:201, 1979

[12] Bai X., McMullan G. and Scheres S. H. W. *How cryo-EM is revolutionizing structural biology.* Trends in biochemical sciences 40:49, 2015

[13] De Rosier D. J. and Klug A. *Reconstruction of three dimensional structures from electron micrographs.* Nature 217:130, 1968

[14] Cheng Y. *Single-particle cryo-EM at crystallographic resolution.* Cell 161:450, 2015

[15] Koster A. J., Grimm R., Tyoke D., Hegerl R., Stoschek A., Walz J. and Baumeister W. *Perspectives of molecular and cellular electron tomography.* Journal of structural biology 12:276, 1997

[16] Dierksen K., Typke D., Hegerl R., Koster A. J. and Baumeister W. *Towards automatic electron tomography.* Ultramicroscopy 40:71, 1992

[17] Koster A. J., H. Chen, J. W. Sedat, and D. A. Agard *Automated microscopy for electron tomography.* Ultramicroscopy 46:207, 1992

[18] Schur F. K. M., Obr M., Hagen W. J. H., Wan W., Jakobi A. J., Kirkpatrick J. M., Sachse C., Kräusslich H. and Briggs J. A. G. *An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation.* Science 353:506, 2016

[19] Beck M., Malmstrøm J. A., Lange V., Schmidt A., Deutsch E. W. and Aebersold R. *Visual proteaomics of the human pathogen Leptospira interrogans.* Nature Methods 6: 817, 2009

[20] Böhm J., Frangakis A. S., Hegerl R., Nickell S., Typke D., Baumeister W. *Toward detecting and identifying macromolecules in a cellular context: template matching applied to electron tomograms.* Proceedings of the National Academy of Sciences 92:14245, 2000

[21] Klar T. A., Jakobs S., Dyba M., Egner A. and Hell S. W. *Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission.* Proceedings of the National Academy of Sciences 97:8206, 2000

[22] Betzig E., Patterson G. H., Sougrat R., Lindwasser O. W., Olenych S., Bonifacino J. S., Davidson M. W., Lippincott-Schwartz J. and Hess H. F. *Imaging intracellular fluorescent proteins at nanometer resolution.* Science 313:1642, 2006

[23] Giepmans B. N. G., Adams S. R., Ellisman M. H. and Tsien R. Y. *The fluorescent toolbox for assessing protein location and function.* Science 312:217, 2006

[24] Briegel A., Ding H. J., Li Z., Werner J., Gitai *Location and architecture of the Caulobacter crescentus chemoreceptor array.* Molecular Microbiology 69:30, 2008

[25] van Driel L. F., Valentijn J. A., Valentijn K. M., Koning R. I. and Koster A. J. *Tools for correlative cryo-fluorescence microscopy and cryo-electron tomography applied to whole mitochondria in human endothelial cells.* European Journal of Cell Biology 88: 669, 2009

[26] Gruska M., Medalia O., Baumeister W. and Leis A. *Electron tomography of vitreous sections from cultured mammalian cells.* Journal of Structural Biology 161L 384, 2008

[27] Chang Y. W., Chen S., Tocheva E. I., Treuner-Lange A., Löbach S., Søgaard-Andersen L. and Jensen G. J. *Correlated cryogenic photoactivated localization microscopy and cryo-electron tomography.* Nature methods 11: 737, 2014

[28] Failk W. P. and Taylor G. M. *An immunocolloid method for the electron microscope.* Immunochemistry 8:1081, 1971

[29] Ackerson C. J., Powell R. D. and Hainfeld J. F. *Site-specific biomolecule labelling with gold clusters.* Methods in enzymology:481, 2010

[30] Humbel B. M., de Jong M. D., Müller W. H. and Verkleij, A. J. *Pre-embedding immunolabeling for electron microscopy: An evaluation of permeabilization methods and markers.* Microscopy research and technique, 42:43, 1998

[31] Beck, M., V. Lucić, F. Förster, W. Baumeister, and O. Medali *Snapshots of nuclear pore complexes in action captured by cryo-electron tomography.* Nature 449:611, 2007

[32] Yi Hong, Strauss J. D., Ke Z. Alonas E., Dillard R. S., Hampton C. M., Lamb K. M., Hammonds J. E., Santangelo P. J., Spearman P. W. and Wright E. R. *Native immunogold labeling of cell surface proteins and viral glycoproteins for cryo-electron microscopy and cryo-electron tomography applications.* Journal of Histochemistry and Cytochemistry 63:780, 2015

[33] Oda T. and Kikkawa M. *Novel structural labeling method using cryo-electron tomography and biotin–streptavidin system* Journal of structural biology 183:305, 2013

[34] Furey W. F., Robbins A. H., Clancy L. L., Winge D. R., Wang B. C. and Stout C. D. *Crystal structure of Cd, Zn Metallothionein* Science 231:704, 1986

[35] Nielson K. B., Atkin C. L. and Winge D. R¿ *Distinct metal-binding configurations in metallothionein* Jounral of Biological Chemistry 260:5342, 1985

[36] Mercogliano C. P., DeRosier D. J. *Concatenated metallothionein as a clonable gold label for electron microscopy.* Journal of Structural Biology 160:70, 2007

[37] Diestra E., Fontana J., Guichard P., Marco S. and Risco C. *Visualization of proteins in intact cells with a clonable tag for electron microscopy.* Journal of Structural Biology 165:157, 2009

[38] Bouchet-Marquis C., Pagratis M., Kirmse R., Hoenger A. *Metallothionein as a clonable high-density marker for cryo-electron microsocopy.* Journal of Structural Biology 177:119, 2012

[39] Nishino, Yuri, Takuo Yasunaga, and Atsuo Miyazawa. *A genetically encoded metallothionein tag enabling efficient protein detection by electron microscopy.* Journal of electron microscopy 56:92, 2007

[40] Wang Q., Mercogliano C. P., and Löwe J. *A ferritin-based label for cellular electron cryotomography.* Structure 19:147, 2011

[41] Wilkins M. H. F., Stokes A. R., Wilson H. R. *Molecular structure of nucleic acids: molecular structure of deoxypentose nucleic acids.* Nature 171:738, 1953

[42] Franklin R. E. and Gosling R. G. *Molecular configuration in sodium thymonucleate* Nature 171:740, 1953

[43] Watson J. D. and Crick F. H. C. *Molecular structure of Nucleic Acids: A structure for deoxyribose nucleic acid* Nature 171:737, 1953.

[44] Guckian K. M., Schweitzer B. A., Ren R. X. F., Sheils C. J., Tahmassebi D. C. and Kool E. T. *Factors contributing to aromatic stacking in water: evaluation in the context of DNA.* Journal of the American Chemical Society 122:2213, 2000

[45] Seeman N. C. *Nucleic acid junctions and lattices.* Journal of theoretical biology 99:237, 1982

[46] Hagerman P.J. *Flexibility of DNA.* Ann. Rev. Biophys. Chem. 17:265, 1988

[47] Smith S. B., Cui Y., Bustamante C. *Overstetching B-DNA: The elastic response of individual double-stranded and single-stranded DNA molecules.* Science 9:795, 1996

[48] Winfree E., Liu F., Wenzler L. A., Seeman N. C. *Design and self-assembly of two-dimensional DNA crystals.* Nature 394:539, 1998

[49] Yan H., Park S. H., Finkelstein G., Reif J. H. and LaBean T. H. *DNA-templated self-assembly of protein arrays and highly conductive nanowires.* Science 201:1882, 2003

[50] Rothemund P. W. K., Papadakis N., Winfree E. *Algorithmic self-assembly of DNA Sierpinski triangles.* PLoS biology 2:424, 2004

[51] Goodman R. P., Berry R. M. and Turberfield A. J. *The single-step synthesis of a DNA tetrahedron.* Chem. Commun. 12: 1372, 2004

[52] Chen J. and Seeman N. C. *Synthesis from DNA of a molecule with the connectivity of a cube.* Nature 350:621, 1991

[53] Zhang Y. and Seeman N. C. *Construction of a DNA-truncated octahedron.* J. Am. Chem. Soc. 116:1661 1993

[54] Goodman R. P., Schaap I. A. T., Tardin C.F, Erben C.M., Berry R.M., Schmidt C.F., and Turberfield A.J. *Rapid chiral assembly of rigid DNA building blocks for molecular nanofabrication.* Science 310:1661, 2005

[55] Malo J., Mitchell J. C., Vénien-Bryan C., Wille H., Sherratt D. J. and Turberfield A. J. *Engineering a 2D protein-DNA crystal* Angerwandte Chemie 117:3117, 2005

[56] He Y., Ye T. Su. M, Zhang C, Ribbe A. E., Jian W. and Mao Chengde *Hierachical self-assembly of DNA into symmetric supramolecular polyhedra.* Nature 452:198, 2008

[57] Shih W. M., Quispe J. D. and Joyce G. F. *A 1.7-kilobase single-stranded DNA that folds into a nanoscale oxtahedron* Nature 427:618, 2004

[58] Rothemund P. W. K. *Folding DNA to create nanoscale shapes and patterns.* Nature 440:297, 2006

[59] Wang J. C. *Helical repeat of DNA in solution* Proceedings of the National Academy of Sciences 76:200 1979

[60] Han D., Pal S., Nangreave J., Dens Z., Liu Y. and Yan H. *DNA origami with complex curvatures in three-dimensional space.* Science 332:342, 2011

[61] Dietz H., Douglas S. M. and Shih W.M. *Folding DNA into twisted and curved nanoscale shapes.* Science 324:725, 2009

[62] Douglas S. M., Marbestone A. H., Teerapittayanon S., Vazquez A., Church G. M. and Shih W. M. *Rapid prototyping of 3D DNA-origami shapes with caDNAno.* Nucleic acids research 37:5001, 2009

[63] Douglas S.M., Dietz H., Liedl T., Högber B., Graf F. and Shih W. M. *Self-assembly of DNA into nanoscale three-dimensional shapes.* Nature 459:414, 2009

[64] Ke Y., Douglas S. M., Liu M., Sharma J., Cheng A., Leung A., Liu Y., Shih W.M. and Yao H. *Multilayer DNA origami packed on a square lattice.* Journal of the American Chemical Society 131:15903, 2009

[65] Ke Y., Voigt N. V., Gothelf K. V. and Shih W. M. *Multilayer DNA Origami packed on Hexagonal and Hybrid Lattices.* Journal of the American Chemical Society 134:1770, 2012

[66] Tian Y., Wang T., Liu W., Xin H. L., Li H., Ke Y., Shih W. M. and Gang O. *Prescribed nanoparticle cluster architectures and low-dimensional array built using octahedral DNA origami frames.* Nature nanotechnology 10:637, 2015

[67] Gerline T., Wagenbauer K. F., Neuner A. M. and Dietz H. *Dynamic DNA devices and assemblies formed by shape-complementary, non-base pairing 3D components.* Science 347:1446, 2015

[68] Wagenbaure K. F., Sigl C. and Dietz H. *Gigadalton-scale shape-programmable DNA assemblies.* Nature 552:78, 2017

[69] Castro E.C., Kilchherr F., Kim Do-Nyun, S'hiao E. L., Wauer T., Wortmann P., Bathe M. and Dietz H. *A primer to scaffolded DNA origami.* Nature Methods 8:221, 2011

[70] Ke Y., Sharma J., Liu M., Jahn K. and Liu Y. *Scaffolded DNA origami of a DNA tetrahedron molecular container.* Nano letters 9:2445, 2009

[71] Andersen E.S., Dong, M., Nielsen M.M., Jahn, K., Subramani R., Mamdouh W., Golas M.M., Sander B., Stark H., Oliveira C.L. and Pedersen J.S. *Self-assembly of a nanoscale DNA box with a controllable lid.* Nature 459:73, 2009

[72] Han D., Pal S., Yang Y., Jiang S., Nangreave J., Liu Y. and Yan H. *DNA gridiron nanostructures based on four-arm junctions.* Science 339:1412, 2013

[73] Zhang F., Jiang S., Wi S., Li Y., Mao C., Liu Y. and Yan H. *Complex wireframe DNA origami nanostructures with multi-arm junction vertices.* Nature nanotechnology 10:779, 2015

[74] Venexiano R. Ratanalert S., Zhang K., Zhang F., Yan H., Chiu W. and Bathe M. *Designer nanoscale DNA assemblies programmed from the top down.* Science 352:1534, 2016

[75] Benson E., Mohammed A., Gardell J., Masich S., Czeiczler E., Orponen P. and Högberg B. *DNA rendering of polyhedral meshes at the nanoscale.* Nature 523:441, 2015

[76] Duncan B. K. and Miller J. H. *Mutagenic deamination of cytosine residues in DNA.* Nature 287:560, 1980

[77] Gray D. M. and Ratliff R. L. *Circular dichroism evidence for G·U and G·T base pairing and G·T base pairing in poly[r(·U)] and poly[d(G·T)].* Biopolymers: Original Research on Biomolecules 16:1331, 1977

[78] Crick F.H.C. *Codon-anticodon pairing: the wobble hypothesis.* 548, 1966

[79] Gautheret D., Konings D. and Gutell R.R. *G·U base pairing motifs in ribosomal RNA.* Rna 1:807, 1995

[80] Leontis N. B. and Westhof E. *Geometric nomenclature and classification of RNA base pairs.* Rna 7:499, 2001

[81] Berg J. M., Tymoczko J. L. and Stryer L. *Biochemistry* (5 ed.):27, W. H. Freeman and Company New York, 2002

[82] Arnott S., Huskins D. W., Dover S.D., Filler W. and Hodgson A.R. Journal of Molecular Biology 81:107, 1973

[83] Wilusz J. E., Sunwoo H. and Spector D. L. *Long noncoding RNAs: functional surprises from the RNA world.* Genes and development 23:1494, 2009

[84] Puglisi J. D., Blanchard S. C. and Green R. *Approach translation at atomic resolution.* Nature Structural Biology 7:885, 2000

[85] Staley J. P., and Guthrie C. *Mechanical devices of the spliceosome: motors, clocks, springs, and things.* Cell 92:315, 1998

[86] Evans D., Marquez S. M. and Pace N. R. *RNase P: interface of the RNA and protein worlds.* Trends in biochemical sciences 31:333, 2006

[87] Collins K. *The biogenesis and regulation of telomerase holoenzymes.* Natural Reviews Molecular Cell Biology 7:484, 2006

[88] Keenan R. J., Freymann D. M., Stroud R. M. and Walter P *The signal recognition particle.* Annual Review of Biochemistry 70:755, 2001

[89] Dohert E. A. and Doudna J. A. *Ribozyme structures and mechanisms.* Annual Review of Biochemistry 69:597, 2000

[90] Batey R. T. *Structures of regulatory elements in mRNAs.* Current opinion in structural biology 16:299, 2006

[91] Leontis N. B., and Westhof E. *Analysis of RNA motifs.* Current opinion in structural biology 13:300, 2003

[92] Hermann T, and Patel D. J. *Stitching together RNA tertiary architectures.* Journal of molecular biology 294:829, 1999

[93] Tinoco Jr I., and Bustamante C. *How RNA folds.* Journal of molecular biology 293:271, 1999

[94] Wu M. and Tinoco Jr I. *RNA folding causes secondary structure rearrangement.* Proceedings of the National Academy of Sciences 95:11555, 1998

[95] Herschlag D. *RNA chaperones and the RNA folding problem.* Journal of Biological Chemistry 270:20871, 1995

[96] Moh G., Chang A., Gianelos J. A., Belfot M., Lambowitz A. M. *The neurospora CYT-18 protein suppresses defects in the phage T4 intron by stabilizing the catalytically active structure of the intron core.* Cell 69:483, 1992

[97] Russell R. and Herschlag D. *Probing the folding landscape of the Tetrahymena ribozyme: commitment to form the naitve conformation is late in the folding pathway.* Journal of Molecular Biology 308:839, 2001

[98] Urbaneja M. A., Wu M., Casat-Finet J. R. and Karpel R. L. *HIV-1 nucleotcapside protein as a nucleic acid chaperone: spectroscopic study of its helix-destabilizing properties, structural binding specificity and annealing activity.* Journal of Molecular Biology 318:749, 2002

[99] Mayer O., Rajkowitsch L. Lorenz C., Konrat R. and Schroeder R. *RNA chaperone activity and RNA-binding properties of the E coli. protein StpA.* Nucleic Acids Research 35:1257, 2007

[100] Semrad K., Green R and Schroeder R. *RNA chaperone activity of large ribosomal subunit proteins from Escherichia coli.* Rna 10:1855, 2004

[101] Bear D. G., Ng R., Derveer D. V., Johnson N. P., Thomas G., Schleivh T. and Noller H. F. *Alteration of polynucleotide secondary structure by ribosomal protein S1.* Proceedings of the National Academy of Sciences 73:1824, 1976

[102] Coetzee T., Herschlag D. and Belfort M. *Echerichia coli proteins, including ribosomal protein S12 facilitate in vitro splicing of phage T4 introns by acting as RNA chaperones.* Genes & development 8:1575, 1994

[103] Zemora G. and Waldsich C. *RNA folding in living cells.* RNA Biology 7:634, 2010

[104] Morgan S. R. and Higgs P. G. *Evidence for kinetic effects in the folding of large RNA molecules.* Journal of chemical physics 105:7152, 1996

[105] Lewicki B. T. U., Margus T., Remme J. and Nierhaus K. H. *Coupling of rRNA transcription and ribosomal assembly in vivo: formation of active ribosomal subunits in Escherichia coli requires transcription of rRNA genes by host RNA polymerase which cannot be replaced by bacteriophage T7 RNA polymerase.* Journal of molecular biology 231:581, 1993

[106] Meyer I. M. and Miklós I. *Co-transcriptional folding is encoded within RNA genes.* BMC Molecular Biology 5:10, 2010

[107] Chworos A., Severcan I., Koyfman A. Y., Weinkam P., Oroudjev E., Hansma H. G. and Jaeger J. *Building programmable jigsaw puzzles with RNA.* Science 306:2068, 2004

[108] Severcan I., Geary C., Chworos A., Voss N., Jacovetty E. and Jaeger L. *A polyhedron made of tRNAs.* Nature chemistry 2:772, 2010

[109] Bindewald E., Hayes R., Yinglin Y. G., Kasprzak W. and Shapiro B. A. *RNAJunction: a database of RNA junctions and kissing loops for three-dimensional structural analysis and information.* Nucleic Acids Research 6:382, 2008

[110] Han D., Qi A., Myhrvold C., Wang B., Dai M., Jiang S., Bates M., Liu Y., An B., Xhang F., Yan H. and Yin P. *Single-stranded DNA and RNA origami.* Science 358:2648, 2017

[111] Afonin J. A., Bindewald E., Yaghoubian A. J., Voss N/ Jacovetty E., Sahpiro B. A. and Jaeger L. *In vitro assembly of cubic RNA-based scaffolds designed in silico.* Nature Nanotechnology 5:676, 2010

[112] Afonin K. A., Kireeva M., Grabow W. W., Kashlev M., Jaeger L. and Shapiro B. A. *Co-transcriptional assembly of chemically modified RNA nanoparticle functionalized with siRNAs.* Nano letters 12:5192, 2012

[113] Geary C. Chworos A., Verzemnieks E., Voss N. R. and Jaeger L. *Composing RNA nanostructures from a syntax of RNA structural modules.* Nano letters 17:7095, 2017

[114] Geary C., Rothemund P. W. K. and Andersen E. S. *A single-stranded architecture for cotranscriptional folding of RNA nanostructures.* Science 345:799, 2014

[115] Sparvath S. L., Geary C. W. and Andersen E. S. *Computer-aided design of RNA origami structures.* Methods and Protocols, Methods in Molecular Biology:51, 2017

[116] Geary C. W. and Andersen E. S. *Design principles for single-stranded RNA origami structures.* DNA Computing and Molecular Programming: 20th International Conference, DNA 20 Proceedings:1, 2014

[117] Elonen A. *Algorithmic design of RNA polyhedra.* Thesis, 2018

[118] Delebecque C. J., Lindner A. B., Silver P. A. and Aldaye F. A. *Organization of intracellular reactions with rationally designed RNA assemblies.* Science 222:470, 2011

[119] Sachdeva G., Garg A., Godding D., Way J. C. and Silver P. A. *In vivo co-localization of enzymes on RNA scaffolds increases metabolic production in a geometrically dependent manner.* Nucleic acids research 42:9493, 2014

[120] Li M., Zheng M., Wu S., Tian C., Liu D., Weizmann Y., Jiang W., Wang G. and Mao C. *In vivo production of RNA nanostructures via programmed folding of single-stranded RNAs.* Nature Communications 9:18, 2018

[121] Jepsen M. D. E., Sparvath S. M., Nielse T. B., Langvad A. H., Grossi G., Gothelf K. V., Andersen E. S. *Development of a genetically encodable FRET system using fluorescent RNA aptamers.* Nature Communcations 9:2196, 2018

[122] Paige, J. S., Wu, K. Y. and Jaffrey, S. R. *RNA mimics of green fluorescent protein.* Science 333:642, 2011

[123] Dolgosheina E. V., Jeng S. C., Panchapakesan S. S. S., Cojocaru R., Chen P. S., Wilson P. D., Hawkins N., Wiggins P. A., and Unrau P. J. *RNA mango aptamer-fluorophore: a bright, high affinity complex for RNA labeling and tracking.* ACS chemical biology 9:2412, 2014

[124] Pettersen E. F., Goddard T. D., Huang C. C., Couch G. S., Meng E. C., Ferrin T. E. *UCSF Chimera - a visualization system for exploratory research and analysis.* Journal of Computational Chemistry 13:1605, 2004

[125] Stahl E., Martin T. G., Praetorius F., and Dietz, H. *Facile and scalable preparation of pure and dense DNA origami solutions.* Angewandte Chemie International Edition, 53:12735, 2014

[126] Journot C. M. A. *On the interaction of DNA nanostructures with lipid bilayers.* (PhD thesis) University of OXford, 2017

[127] Liedl T., Högberg B., Tytell J., Ingber D. E., Shih W. M. *Self-assembly of 3D prestressed tensegrity structures from DNA.* Nature Nanotechnology 5:520, 2010

[128] Ditermar Porschke *Persistence length and bending dynamics of DNA from electrooptical measurements at high salt concentrations.* Biophysical Chemistry 40:169, 1991

[129] Chopra A., Swati K., and Simmel F. C. *Electrotransfection of polyamine folded DNA origami structures.* Nano letters 16:6683, 2016

[130] Agarwal N. P., Matthies M., Gür F. N., Osada K. and Schmidt T. L. *Block Copolymer Micellization as a Protection Strategy for DNA Origami* Angewandte Chemie International Edition, 56:5460, 2017

[131] Schindelin J., Arganda-Carreras I. and Frise, E. et al. *Fiji: an open-source platform for biological-image analysis.* Nature methods 9:676, 2012

[132] Stehlikova K., Kostrhunova H., Kasparkova J., and Brabec V. *DNA bending and unwinding due to the major 1, 2-GG intrastrand cross-link formed by antitumor cis-diamminedichloroplatinum (II) are flanking-base independent.* Nucleic acids research 30:2894, 2002

[133] Maaloum, M., P. Muller, and S. Harlepp. *DNA-intercalator interactions: structural and physical analysis using atomic force microscopy in solution.* Soft Matter 9:11233, 2013

[134] Alivisatos, A. P., Johnsson, K. P., Peng, X., Wilson, T. E., Loweth, C. J., Bruchez Jr, M. P., and Schultz, P. G. *Organization of 'nanocrystal molecules' using DNA.* Nature 382:609, 1996

[135] Niemeyer C. M. *Semisynthetic DNA-protein conjugates for biosensing and nanofabrication.* Angerwandte Chemie International Edition 49:1200, 2010

[136] Zimmer M. *Green fluorescent protein (GF)): Applications, structure and related photophysical behaviour.* Chem. rev. 102:759, 2001

[137] Shui B., Ozer A., Zipfel Q., Sahu N., Singh A., Lis J. T., Shi H. and Kitlikoff M. I. *RNA aptamers that functionally interact with green fluorescent protein and its derivatives.* Nucleic acids research 40:39, 2011

[138] Imburgio D., Rong M., Ma K., and McAllister W. T. *Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants.* Biochemistry 39:10419, 2000

[139] Zacharia D. A., Violin J. D., Newton A. C. and Tsien R. Y. *Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells.* Science 296:913, 2002

[140] Kremers G., Goedhart J., van Munster E. B. and Gadelle T. W. J. *Cyan and yellow super fluorescent proteins with improved brightness, protein folding and FRET Förster radius.* Biochemistry 45:6570, 2006

[141] Rana M. S., Wang X. and Banerjee A. *An improved strategy for fluorescent tagging of membrane proteins for overexpression and purification in mammalian cells.* Biochemistry 57:6741, 2018

[142] Pédelacq J., Cabantous S., Tran T., Terwilliger T. C. and Waldo G. S. *Engineering and characterization of a superfolder green fluorescent protein.* Nature Biotechnology 24:79, 2006

[143] Shu X., Shaner N. C., Yarbrough C. A., Tsien Y. R. and Remington S. J. *Novel chromophores and buried charges control color in mFruits.* Biochemistry 45:9639, 2006

[144] Fry J. Q. and Wood K. J. *Gene therapy: potential applications in clinical transplantation.* Expert Reviews in Molecular Medicine 1:1, 1999

[145] Förster F., Medalia O., Zauberman N., Baumeister W. and Fass D. *Retrovrius envelope protein complex structure in situ studied by cro-electron tomography.* PNAS 102:4729, 2005

[146] MacKrell A. J., Soong N. W., Curtis C. M. and Andersen W. F. *Identification of a subdomain in the Moloney murine leukemia virus envelope protein involved in receptor binding.*

[147] Cosset F. L., Morling F. J., Takeuchi Y., Weiss R. A., Collins M. K. and Russell S. J. *Retroviral retargeting by envelopes expressing an N-terminal binding domain.* Journal of Virology 69:6314, 1995.

[148] Gliniak B. C., Kozak S. L., Jones R. T., and Kabat D. *Disulfide bonding controls the processing of retroviral envelope glycoproteins.* Journal of Biological Chemistry 266:22991, 1991

[149] Riedel C., Vasishtan D., Siebert C. A., Whittle C., Lehmann M. J., Mothes W. and Grünewald *Native structure of a retroviral envelope protein and its conformational change upon interactions within the target cell.* Journal of Structural Biology 197:172, 2017

[150] Kelley L. A., Mezulis S., Yates C. M., Wass M. N. and Stenberg M. J. E. *The Phyre2 web portal for protein modeling, prediction and analysis.* Nature Protocols 10:845, 2015

[151] Mei Q., Wei X., Su F., Liu Y., Youngbull C., Johnson, R., Lindsay S., Yan H. and Meldrum D. *Stability of DNA origami nanoarrays in cell lysate.* Nano letters, 11:1477, 2011

[152] Schüller V. J., Heidegger S., Sandholzer N., Nickels P. C., Suhartha N. A., Endres S., Bourquin C. and Liedl T. *Cellular immunostimulation by CpG-sequence-coated DNA origami structures.* ACS Nano 5:9696, 2011

[153] Jiang Q., Song C., Nangreave J., Liu X., Lin L., Qiu D., Wang Z.G., Zou G., Liang X., Yan H. and Ding B. *DNA origami as a carrier for circumvention of drug resistance.* Journal of the American Chemical Society 134:13396, 2012

[154] Mikkila J., Eskelinen A. P., Niemela E. H., Linko V., Frilander M. J., Torma P. and Kostiainen M. A. *Virus-encapsulated DNA origami nanostructures for cellular delivery.* Nano letters 14:2196, 2014

[155] Zha, Y. X., Shaw A., Zeng X., Benson E., Nyström A. M. and Högberg B. *DNA origami delivery system for cancer therapy with tunable release properties.* ACS nano 6:8684, 2012

[156] Sheng Y., Mancino V. and Birren B. *Transformation of Escherichia coli with large DNA molecules by electroporation.* Nucleic Acids Research: 23:1990, 1995

[157] Gonzales M. F., Brooks T., Pukatzki S. U. and Provenzano D. *Rapid protocol for preparation of electrocompetent Escherichia coli and Vibrio cholerae.* Journal of Visualized Experiments 80:50684, 2013

[158] Parlea L., Puri A., Kasprzak W., Bindewald E., Zakrevsky P., Satterwhite E., Joseph K., Afonin K. A. and Shapiro B. A. *Cellular delivery of RNA nanoparticles.* ACS combinatorial science 18:537, 2016

[159] Sharei A., Zoldan J., Adamo A., Sim W. Y., Cho N., Jackson E., Mao S., Schneider S., Han M. J., Lytton-Jean A. and Basto P. A. *A vector-free microfluidic platform for intracellular delivery.* Proceedings of the National Academy of Sciences 5:2082, 2013

[160] Jackson V. A, Mehmood S., Chavent M., Roversi P., Carrasquero M., Del Toro D., Seyit-Bremer G., Ranaivoson F. M., Comoletti D., Sansom M. S. and Robinson C. V. *Super-complexes of adhesion GPCRs and neural guidance receptors* Nature communications 7:11184, 2016

[161] Connolly S. A., Jackson J. O., Jardetzky T. S., and Longnecker R. (2011). *Fusing structure and function: a structural view of the herpesvirus entry machinery.* Nature reviews Microbiology 9:369, 2011

[162] Selmer M., Dunham C. M., Murphy F. V., Weixlbaumer A., Petry S., Kelley A. C,. Weir J. R., and Ramakrishnan V. *Structure of the 70S ribosome complexed with mRNA and tRNA.* Science 313:1935, 2006

[163] Lee A. J. and Crothers D. M. *The solution structure of an RNA loop-loop complex: the Col1 inverted loop sequence* Structure 6:992, 1993.

[164] Grabow W. W., Zakrevsky P., Afonin K. A., Chworos A., Shapiro B. A. and Jaeger L. *Self-assembling RNA nanorings based on RNAI/II inverse kissing complexes.* Nano letters 11:878, 2011

[165] Guex N. and Peitsch M. C. *SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling.* Electrophoresis 18:2714, 1997

[166] Chen H., Meisburger S. P., Pabit S. A., Sutton J. L., Webb W. W., and Pollack L. *Ionic strength-dependent persistence lengths of single-stranded RNA and DNA.* Proceedings of the National Academy of Sciences 109:799, 2011

[167] Zadeh J. N., Steenberg C. D., Bois J. S., Wolfe B. R., Pierce M. B., Khan A. R., Dirks R. M. and Pierce N. A. *NUPACK: analysis and design of nucleic acid systems.* Journal of computational chemistry 32:170, 2011

[168] Zok T., Antczak M., Zurkowski M., Popenda M., Blazewicz J., Adamiak R.W. and Szachniuk M. *RNApdbee 2.0: multifunctional tool for RNA structure annotation.* Nucleic Acids Research 46:30, 2018

[169] Antczak M., Zok T., Popenda M., Lukasiak P., Adamiak R. W., Blazewicz J. and Szachniuk M. *RNApdbee – a webserver to derive secondary structures from pdb files of knotted and unknotted RNAs.* Nucleic Acids Research 42:368, 2014

[170] Ennifar E., Nikulin A., Tishchenko S., Serganov A., Nevskaya N., Garber M., Ehresmann B., Ehresmann C., Nikonov S. and Dumas P. *The crystal structure of the UUCG tetraloop.* Journal of molecular biology 204:35, 2000

[171] Zuker M. *Mfold web server for nucleic acid folding and hybridization prediction.* Nucleic Acids Research 21:3406, 2003.

[172] Lescoute A. and Westof E. *Topology of three-way junctions in folded RNAs.* Rna 12:83, 2006

[173] Ennifar E., Paillart J. C., Bodlenner A., Walter P., Weibel J. M., Aubertin A. M., Pale P., Dumas P. and Marquet R. *Targeting the dimerization initiation site of HIV-1 RNA with aminoglycosides: from crystal to cell.* Nuleic Acids Research 34:2328, 2006

[174] Bindewald E., Afonin K., Jaeger L and Shapiro B. A. *Multi-strand RNA secondary structure prediction and nanostructure design including pseudoknots.* ACS Nano 5:9542, 2011

[175] Afonin K. A., Cieply D. J. and Neocles B. Leontis *Specific RNA self-assembly with minimal paranemic motifs.* Journal of the American Chemical Society 130:93, 2008

[176] Andronescu M., Aguirre-Hernandez R., Condon A., and Hoos H. H. *RNAsoft: a suite of RNA secondary structure prediction and design software tools.* Nucleic Acids Research, 31: 3416, 2003.

[177] Bashan A., Agmon I., Zarivatch R., Schluenzen F., Harms J.M., Berisio R., Bartels H., Franceschi, F. Auerbach, T. Hansen, H.A., Kossoy E., Kessler M. and Yonath A. *Structural basis of the ribosomal machinery for Peptide bond formation, translocation, and nascent chain progression.* Molecular Cell 11:91, 2003

[178] Baker L. A., Sinnige T., Schellenberger P., de Keyzer J., Siebert C. A., Driessen A. J. M., Baldus N. and Grünewald K. *Combined ^{1}H-detected solid-state NMR spectroscopy and electron cryotomography to study membrane proteins across resolutions in native environments.* Structure 26:161, 2018

[179] Journot C. M. A. *Structures and Mechanisms for Synthetic DNA Motors.* (PhD thesis) University of Oxford, 2017

[180] Tivol W. F., Briegel A., and Jensen G. J. *An improved cryogen for plunge freezing.* Microscopy and Microanalysis 14:375, 2008

[181] Mastronarde D. N. *Automated electron microscope tomography using robust prediction of specimen movements.* Journal of structural biology 152:36, 2005

[182] Mastronarde D. N., and Held S. R. *Automated tilt series alignment and tomographic reconstruction in IMOD.* Journal of structural biology 197:102, 2017