# DNA ORIGAMI ASSEMBLY

Katherine Elizabeth Dunn

St John's College, University of Oxford


Thesis submitted for degree of Doctor of Philosophy, University of Oxford

233 pages

*Trinity Term 2014*

# DNA ORIGAMI ASSEMBLY

Katherine Elizabeth Dunn

St John's College, University of Oxford

Abstract of thesis submitted for degree of Doctor of Philosophy,
University of Oxford
*Trinity Term 2014*

This thesis describes my investigations into the principles underlying self-assembly of DNA origami nanostructures and discusses how these principles may be applied. To study the origami folding process I designed, synthesized and characterized a polymorphic tile, which could adopt various shapes. The distribution of tile shapes provided new insights into assembly.

The origami tiles I studied were based on scaffolds derived from customized plasmids, which I prepared using recombinant DNA technology. I developed a technique to monitor incorporation of individual staples in real time using fluorescence, measuring small differences in staple binding temperatures ($\sim$0.5-5°C). I examined the tiles using Atomic Force Microscopy and I found that a remarkably high proportion of polymorphic tiles folded well, which suggests that there are assembly **pathways**, arising from strong cooperation between staples. In order to analyse the tile shapes quantitatively, I developed a specialized image processing technique. For validation of the method, I generated and analysed simulated data, and the results confirmed that I could measure individual tile parameters with sub-pixel resolution.

I studied eleven variants of the polymorphic tile, and I proved that minor staple modifications can be used to change the folding pathway dramatically. The strength of cooperation between staples affects their behaviour, which is also influenced by their length and base sequences. Paired staples are particularly significant in assembly, and there are clear parallels with protein folding.

I describe in an Appendix how I applied origami assembly principles in the development of my concept for an autonomous rotary nanomotor utilizing the sequential opening of DNA hairpins (already used for linear motors). This device represents an advance over non-autonomous rotary motors and I have simulated its performance.

In this thesis I have answered important questions about DNA origami assembly, and my findings could enable the development of more sophisticated DNA nanostructures for specific purposes.

# Acknowledgements

I thank my supervisor Prof. Andrew Turberfield for his support, guidance and encouragement, and Dr Jon Bath, who first proposed the use of a dimer scaffold and assisted me with design and production of the monomer and dimer.

My experimental work was complemented by simulations performed by Frits Dannenberg of the Oxford Computer Science Department, who is co-supervised by Prof. Marta Kwiatkowska and Prof. Andrew Turberfield, and whose work has benefitted from support from Dr Tom Ouldridge. I have discussed DNA origami assembly at length with Andrew, Jon and Frits and I am grateful for the input from these conversations.

**Atomic Force Microscopy**

*Oxford:* I thank Dr Helen Carstairs for introducing me to the Oxford AFM and Dr Chandra Ramanujan, Dr Sonia Contera & Paul Kocher for providing me with help and advice on various AFM problems.

*Aarhus:* I performed initial AFM measurements during a two-week visit to the Centre for DNA Nanotechnology at Aarhus University in Denmark, and our collaborators there made me feel very welcome. I must thank Prof. Kurt Gothelf and Prof. Mingdong Dong for providing access to their lab facilities and advice on AFM, and Stinne Høst for help with organizing the visit. Anne Louise Bank Kodal, Sarah Helmig and Shuai Zhang were particularly generous with their time - I very much appreciated their assistance in the laboratory, and their tips on AFM.

**Transmission Electron Microscopy**

Although efforts to image my origami tiles using TEM generally proved to be unsuccessful, I am grateful to Dr Neil Young of Oxford Materials for training me in the use of the JEOL 2000FX system at Begbroke Science Park; I thank Dr Daniele Selmi (Oxford), who originally introduced me to TEM, plus Roslin Adamson (Oxford), Dr Thom Sharp (Oxford) and Karen Thomsen (Aarhus), who also assisted me with TEM sample preparation and imaging.

**Research group**

I thank members of the DNA group, past and present, for providing a friendly, cake-filled environment to work in, and for turning a blind eye to the ridiculous number of Eppendorf tubes I have been keeping in the lab fridge. Including both new arrivals and old hands, current group members not already mentioned above are: Dr Wenjing Meng, Dr Robert Schreiber, Robert Machinek, Aiman Entwistle, Alex Lucas, Flo Benn, Ibon Santiago Gonzalez, Mike Boemo, Antonio Garcia Guerra, Céline Journot.

I am grateful to those more senior group members, now departed for pastures new, who took the time to help me learn the ropes when I joined the lab three years ago - particularly Dr Carlos Sanchez-Cano, for teaching me how to clone plasmids. The work of our lab technician Sarah Matthews ensured a constant supply of indispensible materials such as buffers, autoclaved flasks etc.

**Funding**

Funding for my research was provided by EPSRC.

**Personal**

My life as a graduate student has been enhanced by membership of the MCR of St John's College. I am fortunate to have several close friends who deserve a special mention here: Adam Povey, Hazel Shepherd, David Lewis, & Florentia Riga (now in Athens). I very much appreciate their friendship.

Finally, I must thank my parents, for heroically backing me up all these years, and for having always been there to help me keep calm and carry on. My father is also responsible for pointing out the resemblance between the DNA motor concept I present in Appendix C and the rotary internal combustion engine (known as a Wankel engine after its creator Felix Wankel).

*Katherine Dunn*
*Trinity Term 2014*

*To my parents.*

# Contents

# Chapter 1

# Introduction

The discovery of the structure of DNA in 1953 [1, 2, 3] marked the beginning of a new era in the biological sciences, and less than 30 years later Seeman proposed that DNA could be used to build synthetic nanostructures [4]. Since then, many different DNA nanostructures have been made and characterized, but the principles underlying their self-assembly are not well-understood. One of the aims of this thesis is to address this situation by identifying factors which influence the assembly of the class of nanostructure known as DNA origami [5], enabling greater control to be exerted over the origami 'folding' process. This is of considerable significance because the number of potential applications of DNA origami is enormous, but it is only by achieving a complete understanding of origami assembly that we will be able to harness the full potential of this technology. Already, DNA origami has been proposed as the basis for new drug delivery technologies [6, 7], nanoscale workspaces for chemistry experiments [8, 9, 10], nanoscopic rulers [11], and other novel systems;

further inventions will surely follow. Furthermore, origami assembly is in some ways analogous to protein folding, a process of immense biological significance [12]. This analogy is a recurring theme in this thesis.

I will introduce the DNA molecule in this Chapter and explain its use as a building block for use in nano-engineering, before reviewing the history of the field of DNA Nanotechnology from Seeman's work in the 1980s [13] to the invention of DNA origami [5], including a description of DNA origami design, synthesis and characterization. I will discuss variations on the standard approaches, covering different origami components, unusual structural motifs and alternative synthesis protocols. To demonstrate the utility of DNA origami, I will provide an outline of the applications of the technique, together with a summary of key successor technologies and methods for producing larger scale origami structures. I will briefly consider other self-assembling structures, and draw an analogy between the hierarchical structural organization seen in proteins with the levels of design of DNA nanostructures.

The core of my thesis describes my investigations into origami folding, which I studied using a model system consisting of a structure (a polymorphic 'tile') capable of folding into a range of configurations. The observed distribution of shapes reflected the mechanisms by which the tiles self-assembled, allowing conclusions to be drawn regarding key factors in the folding process. I will describe in Chapter 2 how I developed as a reference a simpler tile with one target structure, before proceeding to explain (in Chapter 3) the method I employed to study the behaviour of

specific origami components in real time during assembly. I discuss in Chapter 4 how I synthesized the polymorphic tiles and imaged them with Atomic Force Microscopy. For these structures the number of posssible misfolds is huge, but my AFM images revealed that a surprising number of them folded well. This suggested that origami folding was not a random process and indicated the existence of assembly pathways, arising from cooperativity in the folding process, the origin of which I discuss in detail.

Chapter 5 describes the characterization of well-folded polymorphic tiles, which can be divided into categories based on tile configuration. I explain the classification system to be used and I describe the image processing strategy I developed for quantitative analysis of shapes of polymorphic tiles observed in AFM images. I demonstrate with simulated data that my technique can achieve sub-pixel resolution.

I discuss in the sixth Chapter the findings of my investigations into polymorphic tile assembly, and I show that it is possible not only to understand the folding pathway, but to control it, using comparatively small modifications. I use the methods described in Chapters 3 and 5 to examine assembly of 11 variants of the polymorphic tile and I explain the principles governing how they fold. To conclude the thesis I summarize my findings.

I present in Appendix C my concept for an autonomous rotary motor made from DNA origami. The proposed mechanism is based on well-established principles, envisaged in use in a new context. Experimental demonstration of the motor is outside the scope of this thesis, but I describe simulations I performed to verify its behaviour.

## 1.1 DNA: Deoxyribonucleic acid

The DNA molecule, depicted in Fig. 1.1a, is approximately 2 nm across [14]. In its standard 'B-form', it consists of two strands which wrap around each other to form a double helix, as first described in 1953 [1, 2, 3]. Each of the two strands is a linear sequence of sub-units called nucleotides (Fig. 1.1b).

Covalent bonds link the sugar of one nucleotide with the phosphate of the next, giving rise to the molecule's 'sugar-phosphate backbone'. Bases project from the backbone of each strand into the centre of the double helix, and pair with the bases on the opposing strand [15] (Fig. 1.1 c,d).

In DNA molecules, there are four types of base - adenine, guanine, cytosine and thymine [16]. The chemical structure of the bases (Fig. 1.1e) is such that they form hydrogen bonds with each other in a highly specific manner. Adenine (A) pairs with thymine (T), forming two hydrogen bonds, and cytosine (C) pairs with guanine (G), forming three hydrogen bonds (Fig. 1.1c). This means that the base sequence of a DNA strand dictates how well it will bind to another DNA strand; if every base on the first strand is matched by the appropriate base on the second strand the two strands are complementary and will bind to each other ('hybridize') to form a double helix (Fig. 1.1a), held together by base pairs (Fig. 1.1c,d). T and C are known as 'pyrimidines', while G and A are 'purines' [17].

Energetically, it is highly favourable for two matched strands to hybridize, but the stability of a duplex (double-helical piece of DNA) is in fact mainly assured not by

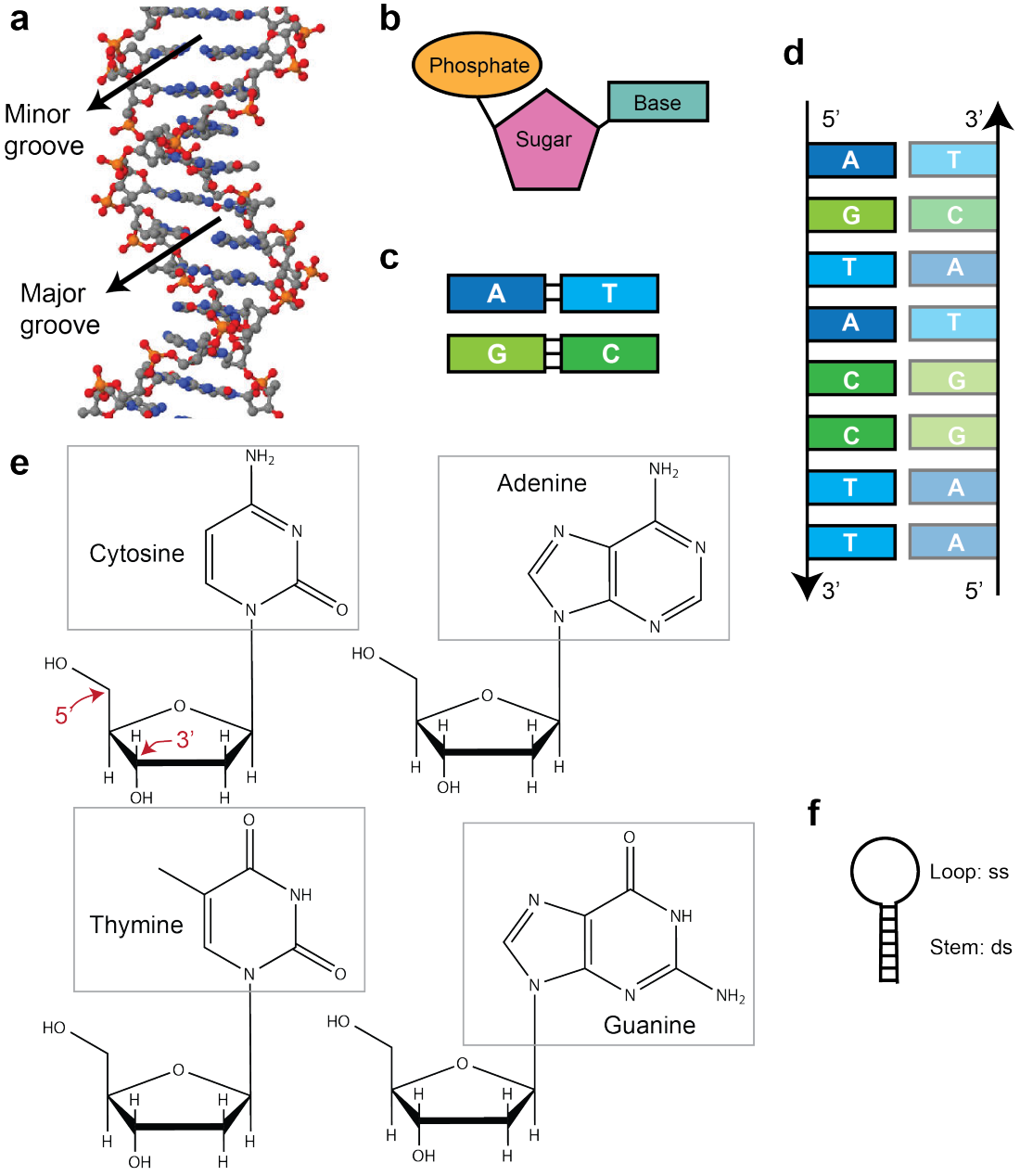


Figure 1.1: (a) DNA double helix, showing major and minor groove. Image from the RCSB PDB (www.pdb.org) of PDB ID 1BNA, Drew *et al.*, Proc. Natl. Acad. Sci.USA 78: 2179-2183 (1981). (b) Sketch of a nucleotide, showing phosphate, sugar (deoxyribose) and base. (c) Specific base-pairing: adenine pairs with thymine, while guanine pairs with cytosine. (d) Duplex formed by hybridization of complementary strands, depicted schematically. (e) Molecular structure of the four bases. Each base is shown attached to deoxyribose. The 5' and 3' carbon atoms are indicated on the diagram showing deoxyribose attached to cytosine. (f) Stem-loop structure, also known as a DNA hairpin. The loop is single-stranded (ss) and the stem is double-stranded (ds).

the base-pairing mechanism but by the interaction between consecutive base-pairs, referred to as the base-stacking interaction [14]. This results from van der Waals forces between the stacked base pairs. Additionally, the bases are hydrophobic, and when they are stacked in the centre of the helix they are not exposed to water, so this configuration maximizes the entropy of the system by minimizing the disruption to the network of hydrogen bonds in the surrounding liquid [14].

The direction of each of the two DNA strands is defined by the orientation of the deoxyribose sugar, the carbon atoms of which are numbered 1' to 5'. The sugar-phosphate bonds of the backbone involve atoms 3' and 5', and as a result the ends of the strand are labelled 3' and 5' (Fig. 1.1d,e). The two strands run anti-parallel to each other [14] and the sequence of a DNA strand is always written starting from the 5' end, unless otherwise specified. The grooves between the two strands are not the same size - the 'major groove' is larger than the 'minor groove' (Fig. 1.1a).

The stability of the molecule is enhanced by the negative charge of the phosphate backbone, which renders the phosphodiester linkages less vulnerable to hydrolysis than other esters [14]. The interaction energy for two strands with given sequences is usually computed using the nearest-neighbour model [18], which takes base-pairing and base-stacking into account.

The rigidity of a polymer such as DNA is commonly quantified by the persistence length, $\lambda$, a measure of the distance over which the molecule behaves like a straight, rigid rod [15]. It characterizes the correlation between the unit vectors $\mathbf{t}(s)$ and $\mathbf{t}(u)$

lying tangent to the polymer backbone at points $u$ and $s$ [15]:

$$\langle \mathbf{t}(s).\mathbf{t}(u) \rangle = e^{-|s-u|/\lambda} \tag{1.1}$$

For double-stranded DNA, $\lambda = 50\,\text{nm}$ [19], and at reasonably low forces the freely-jointed-chain model [15] of a polymer is a good description of its behaviour. As the force is increased, the force-extension curve starts to differ from the FJC prediction. At this stage its behaviour is like that of a continuously flexible rod which curves smoothly due to thermal fluctuations, and this is described by an inextensible worm-like chain model [19]. At higher forces the molecule behaves like a stretchable solid, and ultimately - at a force of around 65 pN - it undergoes a transition to an overstretched form which is 1.7 times longer than its B-form contour length. Single-stranded (ss) DNA is far more flexible than double-stranded DNA, and its persistence length may be less than 1 nm [20].

In addition to the standard B-form described above, DNA can take one of two other possible helical forms, which are referred to as A- and Z- DNA. A-DNA is a right-handed double helix, but the sugar ring of each nucleotide is puckered differently, with the result that the base-pairs are tilted with respect to the helical axis[1] [14]. Consequently the A-DNA helix is wide, its major groove is large and its minor groove is shallow [21]. The A-form is favoured by dehydration [14].

[1]The term 'puckering' refers to a deviation from planar geometry, where a carbon atom in the ring comes out of the plane in which the rest of the atoms in the deoxyribose ring lie. If the puckering is described as *endo* the out-of-plane atom is on the same side of the ring as the C5' atom; if it is on the opposite side, this is *exo* puckering [17].

Z-DNA [22] is a left-handed double helix, and it is less stable than B-DNA under physiological conditions. In B-DNA, nucleotides all have the same conformation, referred to as 'anti', but in Z-DNA the conformation alternates between 'syn' and 'anti' forms, which are related by a rotation around the sugar-base bond[2]. The formation of Z-DNA is favoured by alternating pyramidine and purine bases, and by high salt concentrations. Z-DNA has a greater persistence length than B-DNA [23].

When part of a DNA strand is complementary to another part of the same strand, it is possible for the strand to fold up on itself [14]; in Nature, such configurations are more common in RNA[3]. A particularly well-known nucleic acid structural motif is the stem-loop structure [14] shown in Fig. 1.1f, also known as a hairpin. This consists of a single strand of DNA which has two complementary domains separated by a sequence of several bases. The two complementary domains hybridize to form a length of B-DNA known as the stem, and the intervening sequence forms the loop. DNA hairpins can play a role in dynamic DNA nano-devices [24] (see also Appendix C) and static structures [5].

Non-duplex forms of DNA have also been observed in Nature. Of particular interest is the G-quadruplex [25], which may be formed from 1 strand or multiple separate strands. The analog of the base-pair is the G-tetrad, consisting of four guanine nucleotides in a planar configuration, stabilized by hydrogen bonding. Various quadru-

[2]In the 'syn' form the base is on the same side of the bond as the sugar, in the 'anti' form the base and sugar are on opposite sides [17]. In the 'syn' form of the nucleotide, the 'base' rectangle in Fig. 1.1b would be directly above the pentagon representing the sugar, rather than off to one side.

[3]RNA: ribonucleic acid [14]. RNA is involved in gene expression; it contains ribose rather than deoxyribose, and the base uracil replaces thymine. It is usually single-stranded, although often folded up on itself as described above.

plex topologies are possible, depending on the length of G-tracts and the length or sequence of the loops between them. The formation of G-quadruplexes requires the presence of monovalent cations such as $K^+$, and in the human genome a high proportion of the potential G-quadruplexes occur in promoter regions[4], including those of cancer genes [25]. G-quadruplexes may be involved in regulation of transcription [25].

A second quadruplex form known as the *i*-motif was discovered more recently [26]. The *i*-motif consists of a pair of intercalated poly-C duplexes. Each of the poly-C duplexes is held together by C-$CH^+$ hydrogen bonds [26]; like the G-quadruplex this structure is not formed by Watson-Crick base-pairing.

## 1.2 'Molecular Lego'

The specific base-pairing described in the preceding section underpins the whole field of DNA Nanotechnology [13]. The fundamental concept is that by designing the sequence of bases in a strand, it is possible to predetermine how it should interact with other strands, enabling the creation of a set of strands which would be expected to bind to each other in a particular arrangement to form the desired structure.

Each DNA strand acts as a building block which interacts in a highly programmable manner with others. DNA nanostructure self-assembly is an efficient bottom-up process, where each object is assembled component by component at the nanoscale, rather than being carved out of a much larger block of raw material. The latter type of pro-

[4]Promoter: region of DNA which indicates where transcription of a particular gene should start. Acts as a binding site for RNA polymerase [16].

cess ('top-down' [27]) involves considerable material wastage. An example of top-down manufacturing would be the formation of inorganic nanowires (from materials such as Si or GaAs) by chemical etching [28]; bottom-up techniques such as vapour-liquid-solid growth [29] have been developed for this kind of nanostructure but these approaches typically involve highly complex apparatus and unpleasant reagents or conditions, while DNA nanotechnology normally requires comparatively harmless reagents and benign conditions. Furthermore, DNA strands (oligonucleotides) can now be acquired from commercial suppliers at low cost[5] and techniques for working with DNA have been developed to a very advanced level.

The field of DNA Nanotechnology was effectively born in the 1980s when Seeman proposed that DNA could be used to solve a major problem in protein structure determination [4, 30]. Protein molecules fulfil many different roles in living organisms, and their highly ordered three-dimensional structures are intricately linked to their functions [12]. Obtaining the molecular structure of a protein, finding out how its constituent atoms are arranged, is therefore of paramount importance in understanding how proteins work, and what happens when they malfunction.

Protein structures are usually found from diffraction measurements made on crystals, but it is notoriously difficult to crystallize many proteins. Seeman suggested that DNA strands could be joined together to form a regular lattice (i.e. a crystal) to which proteins could be attached, creating a synthetic protein crystal [30]. Since this

[5]See e.g. www.idtdna.com. Typical price - 25 pence per base for 25 nmol of a 15-60 base long oligonucleotide.

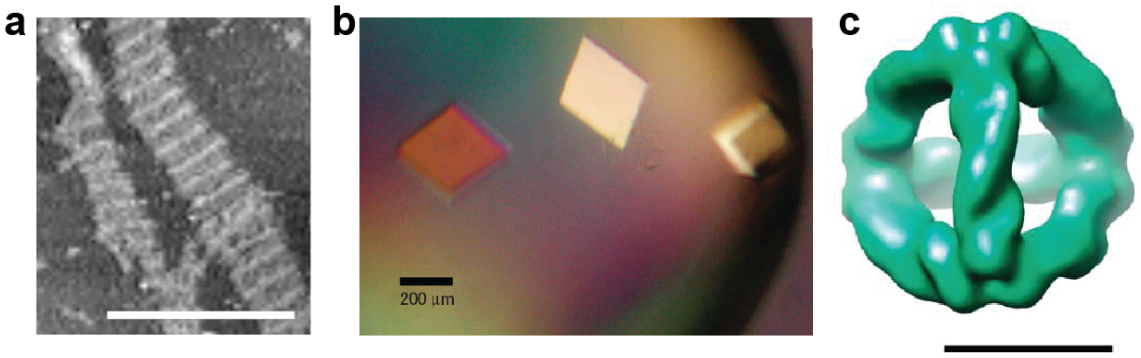


Figure 1.2: (a) Atomic Force Microscope image of a two-dimensional DNA crystal, consisting of many small tiles joined together. Reprinted by permission from Macmillan Publishers Ltd: Nature, Ref. [31], copyright 1998. Scale bar is 300 nm. (b) Optical micrograph of three-dimensional DNA crystals. Scale bar is 200 $\mu$m. Reprinted by permission from Macmillan Publishers Ltd: Nature, Ref. [32], copyright 2009. (c) Cryo-Electron Microscopy reconstruction of a DNA tetrahedron. Reprinted with permission from Ref. [34]. Copyright 2009 American Chemical Society. Scale bar is 5nm.

was first proposed, both 2D [31] and 3D [32] DNA crystals have been created (Fig. 1.2a,b), and it has been demonstrated that DNA lattices do indeed have great potential for protein structure determination [33], but the field of DNA Nanotechnology has also grown far beyond its original remit.

The first three-dimensional DNA nanostructure to be constructed was a cube, made from 10 oligonucleotides [35]. This was followed by truncated octahedra [36], octahedra [37] and tetrahedra [38, 39] (Fig. 1.2c), and it was demonstrated that such shapes could even be reconfigured after assembly by the addition of control strands, using the principle of toehold-mediated DNA strand displacement [40]. This phenomenon arises in a system of three DNA strands, where a DNA strand which is fully complementary to the target displaces a shorter strand bound previously to the target [41]. The process (Fig. 1.3a) occurs because it is energetically favourable to maximize the number of base pairs, and in the initial state the unpaired overhang

acts as a toehold to which the newcomer binds before displacing the incumbent.

DNA strand displacement [44, 45] is the foundation of dynamic DNA nanotechnology, and has been used successfully to create DNA-based molecular motors [43] and to perform computations [46]. The archetypal DNA strand displacement system is the DNA tweezers [42] shown in Fig. 1.3b. Addition of the appropriate 'fuel' strand triggers a conformational change in the tweezers, which can be reversed by addition of an antifuel for removal of the fuel. This process can be monitored experimentally by using a reporter dye (Fig. 1.3b.), the fluorescence intensity of which depends on its proximity to its FRET partner, determined by the state of the tweezers.

More recently, various researchers have created DNA motors capable of locomotion along a track [47, 48, 49, 50, 51, 52]. An example of a DNA motor is shown in Fig. 1.3c, and this motor operates with the assistance of a naturally-occurring restriction enzyme, a protein which cuts DNA strands [47]. When the motor strand is bound to a stator it forms a duplex bearing the base sequence which acts as a recognition site for the enzyme, and the enzyme then cleaves the sugar-phosphate backbone at this location. The short piece of DNA dissociates, as indicated, and the motor then steps onto the next stator by toehold-mediated strand displacement. This is referred to as a 'burnt-bridges' mechanism because the track is destroyed behind the motor, and backwards motion is impossible (see also Appendix C).

Strand displacement also underpins the field of DNA computing [53, 54], a discussion of which is outside the scope of this thesis.

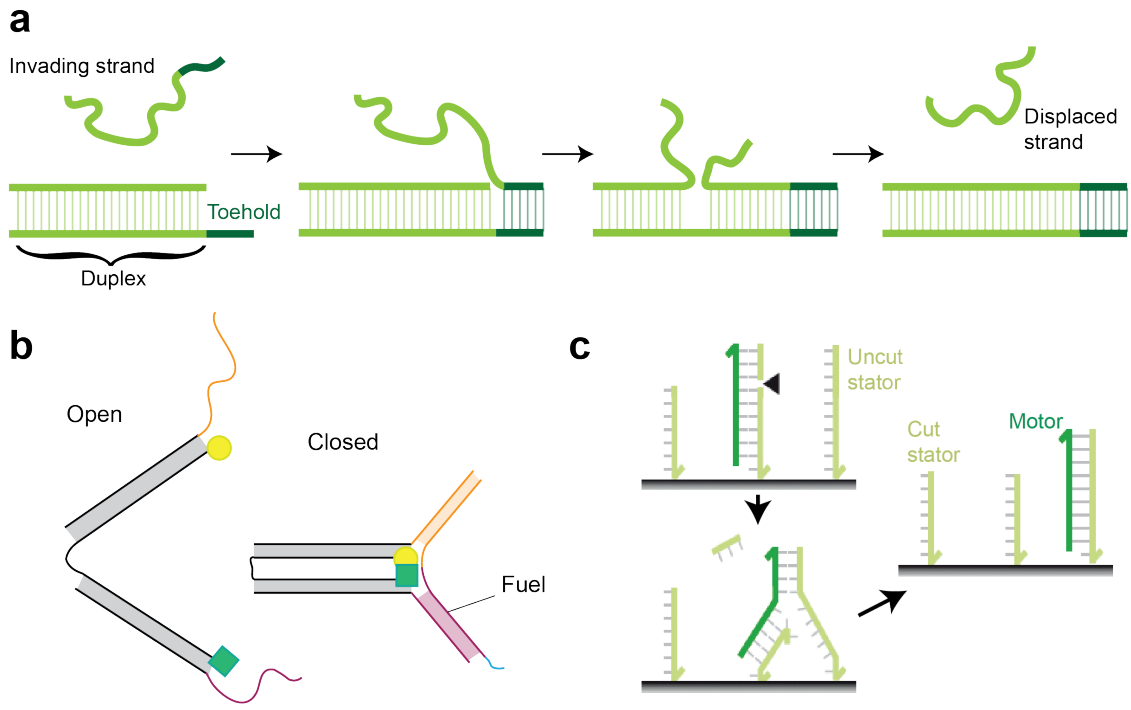


Figure 1.3: (a) Mechanism of toehold-mediated strand displacement. The invading strand binds to the toehold, and when the base pairs at the end of the adjacent duplex break as a result of thermal fluctuations, the invading strand can make new base pairs. The process is highly stochastic and each step is reversible, but the invading strand will ultimately usurp the position of the incumbent strand because the resulting duplex has a larger number of base pairs than the initial complex. (b) DNA tweezers (based on description in Ref. [42]). Addition of fuel strand closes the tweezers. Re-opening is accomplished by the addition of a strand which binds to the toehold (blue) of the fuel and strips it off. The tweezers are labelled with a pair of fluorophores which are near to each other and able to interact via FRET when the tweezers are closed, but far apart and non-interacting when they are open. (c) An example of a DNA motor - adapted by permission from Macmillan Publishers Ltd: Nature Nanotechnology, Ref. [43], copyright 2007. The motor (dark green) is initially bound to a stator (light green). A restriction enzyme nicks the duplex and the short fragment dissociates, enabling the motor to step from one site to the next by toehold-mediated strand displacement.

## 1.3 DNA Origami

### 1.3.1 Principles

This thesis focusses on the class of nanostructure known as DNA origami, invented in 2006 by Paul Rothemund [5]. The raw materials for the construction of DNA origami (shown in Fig. 1.4a) are a long single strand of DNA (the 'scaffold'), which is a few thousand nucleotides long, and a set of short strands called 'staples', each being typically 16-40 nucleotides long. The bacteriophage genome M13mp18 is commonly used as the scaffold, but an alternative may be employed - usually it is acceptable to use an arbitrary scaffold so long as the sequence is known.

Each staple strand is designed to bind specifically to two or more sections of the scaffold, pinning them together. When all staples have hybridized with the complementary scaffold domains, they force the scaffold into a particular shape, such as the iconic smiley face made by Rothemund (Fig. 1.4b). The shape adopted by the scaffold is entirely dictated by the set of staples used, and hence a completely different nanostructure may be formed using the same scaffold by changing the staple set.

The example design shown in Fig. 1.4c illustrates the typical features of a simple DNA origami object. DNA helices are linked by strand 'crossovers', formed at the points where a strand makes a connection from one helix to the next. These crossovers are usually regarded as having zero length. The 'seam' is a line which is not crossed by the scaffold except at the edges of the object, and it is usually bridged by staples.

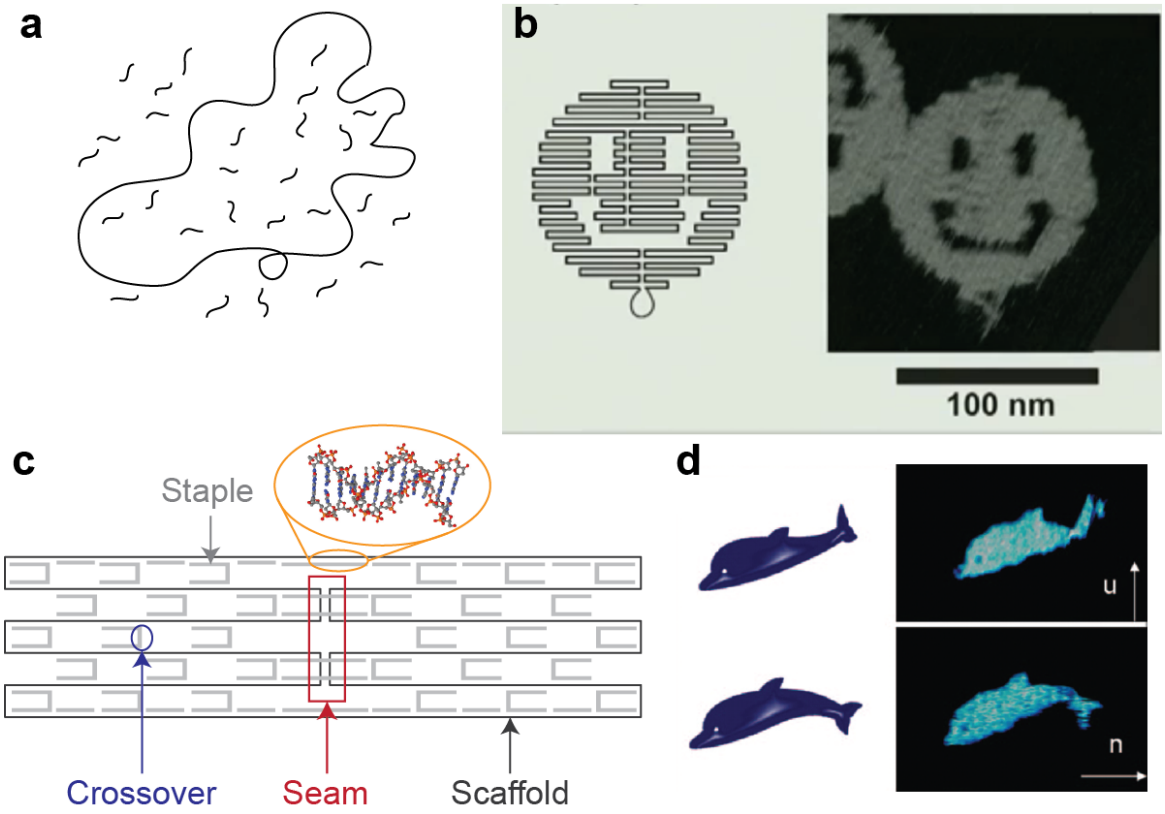


Figure 1.4: DNA origami - principle; design of two-dimensional structures. (a) The raw materials - a single-stranded loop of DNA (the scaffold) and a large number of short single-stranded pieces of DNA (the staples). (b) One of the original origami structures made by Rothemund - the smiley face [5]. Image source: `http://www.ted.com/talks/paul_rothemund_casts_a_spell_with_dna.html` The staples (not shown) pin the scaffold into the shape indicated (left), and AFM images confirm (right) that the structure folds correctly. (c) Mini origami tile, showing key features - scaffold, staples and a representative crossover (zero length) are labelled, and the seam is marked. (d) Origami dolphins with flexible tails. Reprinted with permission from Ref. [55]. Copyright 2008 American Chemical Society.

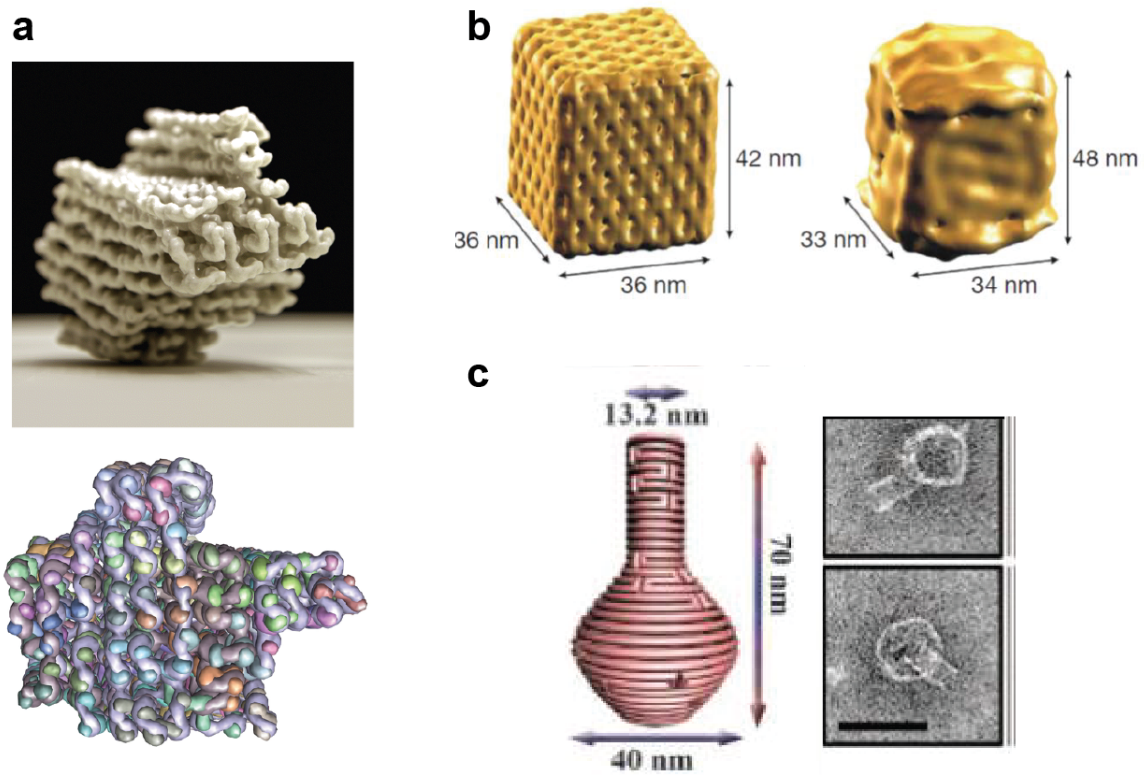


Figure 1.5: (a) A 3D DNA origami structure. Upper image - representation of the object, Copyright T.G. Martin, Cover Image, PNAS ($4^{th}$ December 2012). Lower image - molecular structure, cryo-EM reconstruction, from RCSB PDB (www.pdb.org), PDB ID 2YMH, Bai *et al.*, Proc. Natl. Acad. Sci.USA 109:20012-20017 (Ref. [58]). Individual helices are visible. (b) Origami box, image reprinted by permission from Macmillan Publishers Ltd: Nature, Ref. [59], copyright 2009. Left - design. Right - experimental data from cryo-EM. (c) Origami nanoflask. Design and TEM images. From Ref. [60]. Reprinted with permission from AAAS.

The DNA origami objects synthesized since the publication of Rothemund's original paper range from 'dolphins with flexible tails' [55] (Fig. 1.4d) to 3D structures [56, 57]. There are two main techniques for fabricating 3D DNA origami structures, both of which are represented in Fig. 1.5. One method is to to pack DNA helices together to form bundles [56, 58] (Fig. 1.5a). The alternative, exemplified by the origami cube (Fig. 1.5b [59]), is the 'flat-pack' approach, in which the object is initially designed as a two-dimensional net, which folds into a three-dimensional form.

To create a flat DNA origami tile, it is necessary to respect the natural geometry

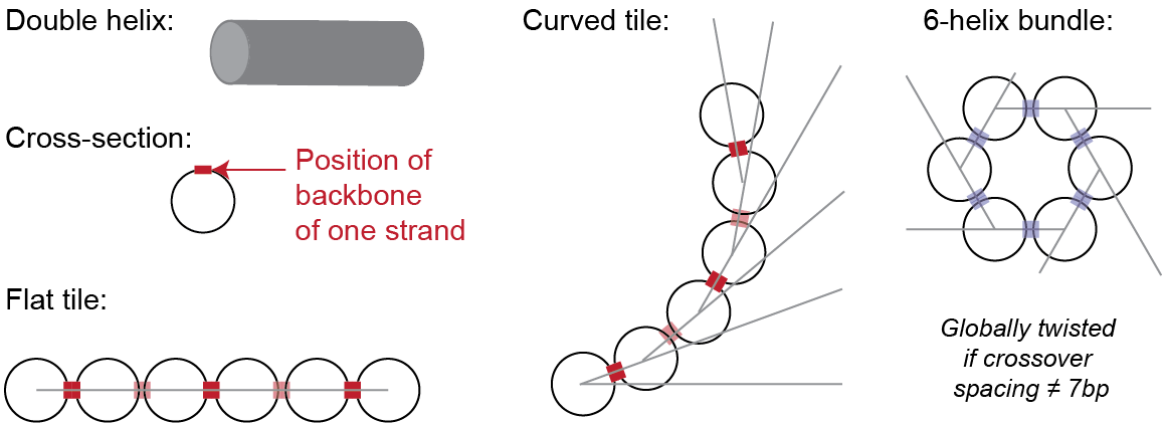


Figure 1.6: Designing flat, curved or twisted DNA origami structures. The tile is flat if the spacing of the crossovers is commensurate with the periodicity of the double helix. Otherwise, it is curved. If crossover spacing and DNA periodicity do not match but helices are locked in position by neighbours, strain or twist results.

of the DNA molecule, allowing a strand to form a crossover only when the helical twist places it in the correct plane to do so. If this rule is disobeyed, the structure will be strained, and as a result it will be bent or twisted (Fig. 1.6). While in some circumstances this effect will be highly undesirable, it may also be used to create structures with complex curvatures [61], such as nanoflasks [60] (Fig. 1.5c), a Mobius strip [62] or spheres [60].

When designing DNA origami it is frequently convenient to assume that 16 base pairs are equivalent to $1\frac{1}{2}$ turns of the helix, which is not strictly correct, and then compensate for the induced distortion by subsequent deletion of 1 base pair in every 64 to give an average of 10.5 base pairs per turn [57].

If additional base pairs are added between the crossovers, such that the number of base pairs per turn is greater than the normal value of 10.5, the assembly yield can be enhanced, but when this underwinding is increased further, yield is very low [63]. The initial improvement in yield has been attributed to increased bowing out of

helices between crossovers, resulting in reduced electrostatic repulsion, but at higher underwinding the physical stress in the system will be considerable. Correct folding of the underwound structure can be induced by addition of the intercalator ethidium bromide, at a concentration which depends on the degree of underwinding [63].

It has also been found that increasing the length of staple domains can enhance yield, by providing 'seeds' for assembly [63].

### Computational tools

The primary computational tool currently in use for design of DNA origami is caDNAno [64]. With this program, the user defines the positions of the component helices on a square or honeycomb lattice, and then designs the routing of the scaffold between these helices. The staple pattern is initially determined automatically, but is then modified by the user according to the desired design.

CaDNAno has a set of inbuilt design rules which specify the positions at which the helical twist of DNA places strands in the correct orientation to cross from one helix to another without strain. It is possible to override these rules if desired, and hairpins and loops can be added to DNA strands. CaDNAno files can be used as input to a software package called 'CanDo' [57, 65], which simulates the solution structure and flexibility of DNA origami objects. It treats a DNA helix as an elastic rod and performs a finite element analysis [66].

Coarse-grained simulations have been extremely successful in describing the properties and dynamics of DNA structures and devices [67, 45, 68], and these techniques

have begun to be applied to DNA origami. However, an origami structure contains around 100 or more different elements, which means that the assembly process is extremely difficult to simulate.

It is possible to model assembly by solving differential equations which define the average probability that a staple will be bound at a given temperature $T$, taking into account which of its neighbours were bound at the previous temperature $T - dT$ [69, 70]. However, it is perhaps preferable to model folding by explicitly tracking paths through the state space[6].

Design and analysis of small groups of interacting strands can be accomplished using the computer programs NUPACK [71] and NANEV [72]. NUPACK (the web-based 'Nucleic Acid Package') computes the interaction energies for different bases using the nearest-neighbour energy model [18], applying a salt correction to compensate for the effects of different buffers [73]. NANEV is a Windows-based computational tool used for design of nucleic acid sequences.

### 1.3.2 Synthesis

DNA origami is typically synthesized in Tris-based buffer containing 12.5 mM magnesium salt (usually either acetate or chloride), at pH 7.8-8.3 [5]. EDTA is sometimes added. For proper assembly, the concentration of staples should exceed that of the scaffold by a factor of at least 2 [5]. It is also known that origami can be folded in the absence of magnesium if the sodium concentration is comparatively high [74].

[6]This is the method used by the author's colleague, Frits Dannenberg.

The origami is folded by a simple annealing protocol. The mixture of scaffold and staples in synthesis buffer is heated to 95°C to disrupt any secondary structure which may form between parts of the scaffold, and then cooled to 25°C. For the origami used in the work described in this thesis the cooling rate was 1°C/min. However, for three-dimensional origami it was thought necessary to use a much slower rate to achieve high yield until Sobczak *et al.* demonstrated that even these structures could be folded in a matter of minutes when a tailored annealing protocol was used [75].

### 1.3.3 Purification

In this thesis the term 'purification' is generally used to refer to the removal of free staples after origami synthesis, although it may also refer to removal of unwanted chemical contaminants. Staple removal can be accomplished by a variety of techniques, but for the work described in this thesis the approach introduced by Wickham *et al.* [50] was employed. The synthesized origami was filtered through columns packed with Sephacryl S-300 size exclusion resin, which delays the passage of smaller molecules such as the staples, while allowing larger molecules such as origami structures to pass. This method is effective and fast, and the protocol is described in full in Appendix A.

Alternative methods include centrifugal filtration, but this is not always effective[7] and origami structures can be damaged if the filter cut-off size is too low or the centrifugation speed is too high [76]. Ion exchange purification [77] may also be used [76], where the sample is applied to a column in which the DNA molecules bind, to

[7]Prof. Björn Högberg, Karolinska Institutet, Sweden - personal communication.

be released by application of a buffer with a particular salt concentration.

Recently Lin *et al.* showed that rate-zonal centrifugation is an effective technique for origami purification [78]. With this method, the origami sample is applied to the top of a solution containing a density gradient of glycerol, and during centrifugation the origami and free staples travel through the solution. The distance moved by a molecule is related to its sedimentation coefficient, which depends on particle size [14]. Lin *et al.* used TEM and agarose gel analysis to confirm that the technique worked. Rate-zonal centrifugation may be valuable for larger-scale purification.

### 1.3.4 Characterization

Various experimental techniques exist for characterization of DNA origami structures. Agarose gel electrophoresis[8] may be used to verify whether DNA origami has folded correctly. This involves placing a DNA sample into a well in a slab of gel and applying an electric field along the length of the slab. DNA is negatively charged, so it will move through the gel towards the positive terminal, and the distance it migrates depends on the size of the molecule. A folded DNA origami object is generally more compact than the unfolded scaffold and therefore will tend to migrate further, but this is not always true. Even if the supposed origami and the scaffold do form distinct bands at different places in the gel, this cannot be taken as definitive proof that the origami has folded well, because it is conceivable that the origami could misfold into a compact globule-like configuration with a different gel mobility to the scaffold. This

[8]Protocol given in Appendix A.

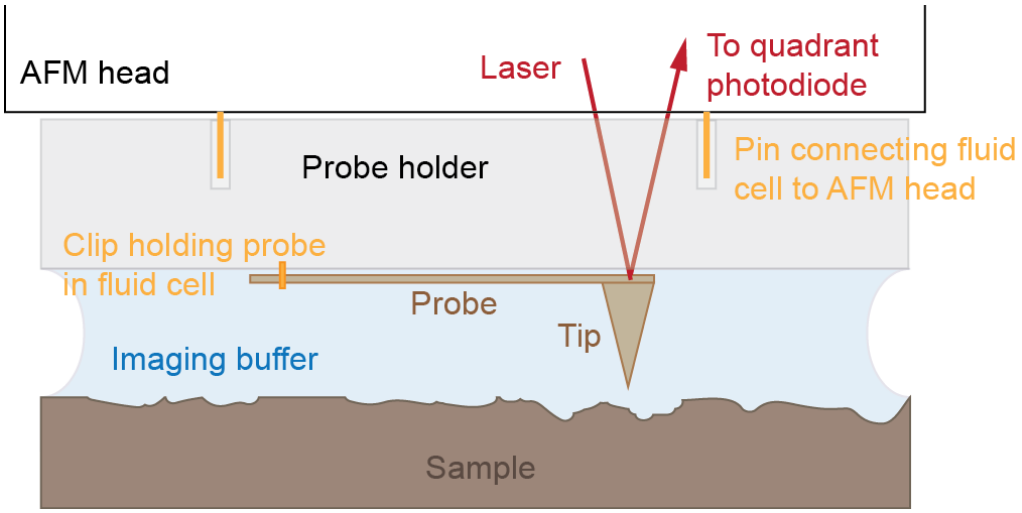


Figure 1.7: Principle of Atomic Force Microscopy in fluid (not to scale). The probe is clipped into a probe holder which is lowered into the fluid. The probe holder is connected to the AFM head, the part of the microscope containing the electronics and the laser system. In some microscopes, the imaging buffer is contained within a specially designed fluid cell, while in others the probe holder and fluid cell are combined into one unit. In the latter case the terms 'probe holder' and 'fluid cell' may be used interchangeably. As described in the text, the laser system is used to track the motion of the tip as it interacts with the surface.

means that it is ultimately essential to visualize the origami using a microscope to confirm whether it has folded into the designed structure.

At 10-100 nm in size, a typical origami object is considerably smaller than the wavelength of visible light (400-700 nm) and hence it cannot be visualized with a standard light microscope. Two-dimensional origami is usually examined using an Atomic Force Microscope (AFM), the principle of which is shown in Fig. 1.7. The AFM was invented by Binnig, Quate and Gerber in 1986 [79] and is a form of microscope in which a very sharp tip is scanned across the specimen, and its interaction with the surface is measured, building up a picture of the sample's topography, line by line.

An AFM may be operated in several different modes. Biological samples are

commonly imaged using 'tapping mode', where a piezoelectric crystal drives oscillation of the cantilever on which the tip is mounted [80, 81]. A laser beam reflects from the cantilever onto a quadrant photodiode (Fig. 1.7) and the beam moves back and forth across the photodiode as the angle of the probe varies, giving rise to a sinusoidal electrical signal. When the tip taps the surface, it experiences a force, which leads to a decrease in the amplitude of both the oscillation of the probe and the corresponding signal from the photodiode. The magnitude of the change depends on the size of the force experienced, which depends on the distance between the tip and the surface. The microscope employs a feedback system to adjust the probe height such that a constant oscillation amplitude is maintained. The topology of the surface is then determined from the feedback signal. The choice of tapping mode minimizes damage and scanning in liquid allows the sample's natural state to be preserved.

While the AFM is almost ideal for studying two-dimensional origami, it is less well-suited to three-dimensional structures because the force applied by the tip during imaging can distort them. For these types of origami object, it is usually better to use transmission electron microscopy (TEM), either negative-stain TEM or cryo-EM.

In TEM, a beam of electrons is fired at the specimen, and electrons are scattered as they interact with it. To form bright field images an objective aperture is used, which selects the direct beam of electrons (those which are not scattered far from the optical axis). In biological samples, the variation in intensity in the image is primarily due to 'mass-thickness' contrast; areas of the specimen with greater mass,

greater thickness or both scatter more electrons off-axis, such that fewer pass through the objective aperture and the corresponding region of the image is darker [82].

DNA is usually quite transparent to electrons and therefore to obtain an image it is necessary to 'stain' the sample by applying a solution containing heavy metal ions, which are attracted to the DNA and effectively block the transmission of electrons [83]. The DNA has usually been deposited on a thin carbon film which is supported on a copper grid, and the heavy metal of choice is often (non-enriched) uranium. Uranyl acetate may be used, but uranyl formate generally yields higher resolution images because it forms smaller grains when the metal ions are adsorbed to the DNA[9]. TEM is not particularly well-suited for imaging two-dimensional structures because these objects frequently roll, curl or aggregate during deposition on the grids, even after the grids have been glow-discharged - this has been observed by the author of this thesis and others[10]. It has been suggested[11] that this may be due to the low pH of the uranyl stain used (in comparison with the synthesis buffer), and that it could be corrected by use of a stain such as PTA, phosphotungstic acid (uranyl acetate stain is typically pH 4.2-4.5, and PTA is pH 5-8 [84]).

Use of these substances should give rise to negative staining, where the specimen is paler than its immediate surroundings, but positive staining can also be observed, and may be more likely when longer staining times are used.

[9]Prof. Hendrik Dietz, Technische Universität München - personal communication.

[10]Using samples provided by the author: Dr Thomas Sharp, then at Oxford, and Karen Thomsen, Aarhus University. Seen independently elsewhere: Prof. Björn Högberg, Karolinska Institutet, Sweden - personal communication.

[11]Karen Thomsen, Aarhus University, personal communication.

It has recently been shown that cryo-EM offers huge potential for imaging three dimensional origami samples with exceedingly high resolution [58]. With this method [85], the sample is flash frozen, and the structures of interest are thus embedded in a layer of ice on an EM grid. A typical TEM image of the specimen shows a large number of individual objects, with a low signal-to-noise ratio. In the image, 3D particles appear as 2D projections, at different orientations. Particles which look similar can be grouped and taking an average of all of the frames in a group forms a 'class average' which is clearer than any of the original individual images. The 3D structure can be reconstructed computationally from the 2D projections.

Origami structures can be labelled with fluorescent molecules, which can be visualized with a fluorescence microscope [11]. This allows the position of each fluorophore to be determined with high accuracy, but only the fluorescent label can be seen. Bulk fluorescence measurements, where the state of an origami object is monitored using fluorescence, can also provide valuable information, particularly in the context of the origami assembly process (see below).

Another technique which may be used to characterize DNA origami objects is small-angle X-ray scattering (SAXS) [59], a method which can measure the overall size of a nanostructure and may reveal new information about its structural properties[12].

[12]Personal communication, Dr Lawrence Lee, Victor Chang Cardiac Research Institute, Australia.

### 1.3.5 Studying origami assembly in real time

I will describe in Chapter 3 the technique I developed for studying origami assembly in real time using fluorescence, and here I review previous work on related methods. I refer to probes of the origami state either as 'non-specific', measuring the overall state, or 'specific', reflecting the behaviour of particular elements of the structure.

#### UV absorption - non-specific

A standard technique for quantifying the amount of DNA in a sample is to measure how much light it absorbs at a wavelength of 260 nm [86] (see Appendix A). The more DNA present, the more light absorbed, in accordance with Beer's Law ($A_{260} = ebc$, where $A_{260}$ is the absorbance at 260 nm, $e$ is the absorption coefficient, $b$ is the length of the light path and $c$ is the DNA concentration). The absorption coefficient $e$ is larger for single-stranded than for double-stranded DNA [87] (this effect is known as the 'hyperchromic shift'), and hence the absorption of a sample will depend on what proportion of it is double-stranded[13].

In the fully unfolded (denatured) state of an origami object, all scaffold and staples are single-stranded whereas in the folded state the origami is usually intended to be entirely double-stranded, and the proportion of dsDNA is a measure of the extent to which the origami is assembled. Consequently the hyperchromic effect can be used to probe the assembly of an origami object, by recording $A_{260}$ while the sample is annealed from high to low temperature. Arbona *et al.* used this technique to show

[13] dsDNA: $e = 0.020\,\mu\mathrm{g\,mL^{-1}\,cm^{-1}}$, ssDNA: $e = 0.026\,\mu\mathrm{g\,mL^{-1}\,cm^{-1}}$. [87].

that origami behaviour is not consistent with staples binding independently [69].

The UV-absorption method for observing origami assembly suffers from the major disadvantage that there is a considerable background signal from the excess staples which are not incorporated into an origami structure. This cannot be eliminated because a 1:1 ratio of staples to scaffold leads to a low yield of well-formed structures [5]. In addition, the background signal may be increased even further by the presence of contaminants from the scaffold purification process, such as phenol, which absorbs strongly at 270 nm [88].

The method also provides only non-specific information on the extent of origami folding i.e. the proportion of strands which have been incorporated into a double helix. It does not shed any light on the behaviour of individual staples and therefore it is of limited use for probing the details of origami folding.

**Use of fluorescent DNA binding dye - non-specific**

An alternative way to probe origami assembly non-specifically is to use a DNA-binding dye, SYBR Green, which intercalates between the base pairs and fluoresces more brightly when bound to double stranded DNA than when bound to single-stranded DNA. Investigations performed using this method [75] revealed that origami assembly occurred over a very narrow temperature range, which allowed the yield of origami assembly to be improved using an alternative annealing protocol in which cooling was slow in the vicinity of the critical formation temperature and fast otherwise.

The SYBR Green method suffers from significant disadvantages, mainly arising

from the behaviour of the dye itself. First, the intrinsic fluorescence of the dye is strongly temperature-dependent, regardless of the state of the DNA to which it is bound, and the signal tends to fall when the experiment is re-run[14]. Second, it is necessary to anneal very slowly if the dye fluorescence is to reach its equilibrium value at each temperature [75]. Third, it is possible that SYBR Green molecules may affect folding; the dye intercalates between the base pairs, and it has been shown that the presence of intercalators influences DNA nanostructure assembly [63, 89]. It has been argued that using very low concentrations of SYBR Green (approximately 1 dye molecule for every 1000 base pairs) should ensure minimum disruption to the structure [75]. Like the measurement of $A_{260}$, the SYBR-Green approach is non-specific.

**Specific probes**

The approach adopted for the work described in this thesis (Chapter 3) was to label individual staples with a fluorophore at one end and a quencher[15] at the other such that the fluorescence increased when the staple was incorporated into the origami tile, allowing the state of a specific staple to be monitored as a function of temperature.

While this investigation was in progress, another research group published a study [90] in which they labelled specific staples with fluorescent molecules but in contrast with my approach they labelled each of two staples with a single fluorophore. Each staple had three domains and when both staples were bound, the fluorophores could

[14]Based on own observations and Ref. [75].

[15]Molecule which interacts with a fluorophore and reduces the intensity of its fluorescence when they are close together.

interact via FRET[16], which was monitored. Both staples must be bound if a signal is to be seen, and hence their approach does not allow the behaviour of individual staples to be monitored. Ultimately, they found that staple stability decreased near structural defects and that measured incorporation and melting temperatures were different at different places in the origami.

#### AFM with temperature control

Origami assembly and disassembly on a surface were observed directly by Song *et al.* using a thermally controlled AFM [92]. The temperature of the sample was changed while it was in the microscope, which enabled visualization of the origami structure as it began to form, or to fall apart. During heating, the tile began to develop local defects at 55°C, and had disassembled completely at 75°C.

### 1.3.6 Variations on the theme

#### Different scaffolds

It is frequently most convenient to use the M13mp18 bacteriophage genome [5] as a scaffold because it is available commercially in single-stranded form and it is comparatively inexpensive[17]. However, where M13mp18 is inappropriate other approaches can be used, involving scaffolds of different lengths. For instance, enzymes may be

[16]Förster Resonant Energy Transfer [91] - the emission spectrum of one fluorophore (the donor) overlaps with the excitation spectrum of another (the acceptor), and energy is transferred non-radiatively from the first fluorophore to the second. The first fluorophore emits less light and the second fluorophore fluoresces. The efficiency of the FRET process depends very strongly on distance.

[17]http://www.neb.uk.com: M13mp18 Single-stranded DNA #N4040, £25 for 10 $\mu$g.

used to cut M13mp18 into sections, and a particular piece can be extracted for use as a scaffold [93]. M13mp18 is usually referred to as 'M13'.

Pound *et al.* demonstrated that PCR[18] could be used to synthesize scaffolds of various lengths (from 758 to 4808 bases) [94], while others have chosen to use plasmids[19] [56]. The latter approach was adopted for the work described in this thesis because it allows the preparation of comparatively large stocks[20] of customized scaffold by cloning using bacteria. To synthesize origami from a plasmid-derived scaffold using a standard annealing protocol it is necessary to use enzymes to make the plasmid single-stranded (see Chapter 2 and Appendix A).

It has also been demonstrated that double-stranded DNA can be used as a scaffold when the assembly buffer contains a denaturing agent. For instance, Högberg *et al.* simultaneously synthesized origami of two different designs using a double stranded scaffold with a backbone nick [95]. The scaffold was mixed with two sets of staples and it was separated into two strands by incubation at 80°C in the presence of formamide, and then exposed to a sudden temperature drop and dialysis to remove the formamide. This process resulted in folding of two distinct origami objects.

Yang *et al.* also used a double-stranded scaffold to synthesize origami in a buffer which contained formamide, but unlike Högberg *et al.* they folded two strands of their scaffold together, into a single structure [96] (Fig. 1.8a). Here, several precautions were required with the design to prevent misfolding. Firstly, it was necessary to

[18]Polymerase Chain Reaction: a technique for amplifying a target DNA sequence using a heat-stable polymerase [14].

[19]A plasmid is a closed loop of double stranded DNA found in bacteria.

[20]Hundreds of $\mu$g, or even mg.

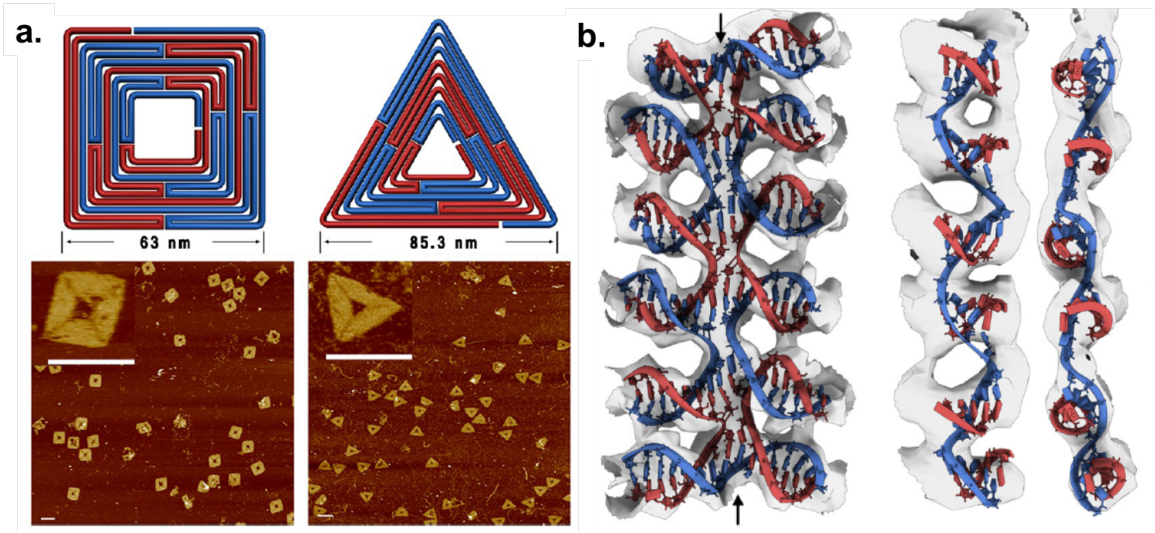


Figure 1.8: Use of double stranded scaffold and alternative structural motifs. (a) Routing a pair of complementary strands to enable origami folding; AFM image confirms success of strategy. Reprinted with permission from Ref. [96]. Copyright 2012 American Chemical Society. Scale bars 100 nm. (b) Non-standard structural motifs, incorporated into an origami object. Left side - stack of Holliday junctions. Right side - pseudohelix, viewed from two different directions. Image reprinted from Ref. [58], copyright Bai *et al.* 2012.

introduce asymmetry in the design to ensure that individual pairs of staples would not be complementary to each other. Furthermore, complementary scaffold domains were not placed too close together, and the folding paths of the two strands were designed to converge regularly to ensure they were knitted together securely.

### Novel structural motifs

The main 'motifs' involved in DNA origami design are the double helix and the crossover, and a typical structure may be thought of as an array of crossovers linked by double-helical domains. However, Bai *et al.* introduced two alternative motifs [58] - a stack of Holliday junctions [97] and a pseudohelix. The Holliday junction stack motif consisted of five such junctions in a vertical line. This motif was stabilized by the stacking of bases on either side of each junction, because the sugar-phosphate

backbone only crosses from one side of the motif to the other at the top and bottom of the stack. At first glance, this motif looks like a standard seam but it is in fact unbridged, with structural integrity maintained due to base-stacking. The pseudohelix was created by allowing a strand to twist around an axis perpendicular to the axis of the helices, crossing from one helix to the next and making a few base pairs in each.

In addition, Han *et al.* demonstrated that it is not necessary to route the scaffold in a raster pattern, and that the routing can be unidirectional (as in a spiral) [98].

### Alternative synthesis procedures

It has been demonstrated that DNA origami and similar structures can be assembled isothermally by using dialysis with formamide instead of thermal annealing, allowing the temperature of assembly to be fixed [99]. This principle was extended by Zhang *et al.* [100], who showed that complex DNA objects could be assembled without either annealing or dialysis if just the right amount of formamide was added. They pointed out that formamide has the capacity to form hydrogen bonds with the bases of DNA, competing with base-pairing, and they argued that this would tend to drive assembly forward by reducing energy barriers and increasing the penalty for mismatches.

Isothermal DNA origami assembly has been directly observed using a thermally controlled AFM [101]. The incorporation of staples was found to be ten times faster at 60°C than at room temperature, and at high temperature (60°C) the staple concentration was found to make no difference to staple binding kinetics.

Development of protocols for isothermal assembly of DNA structures could enable

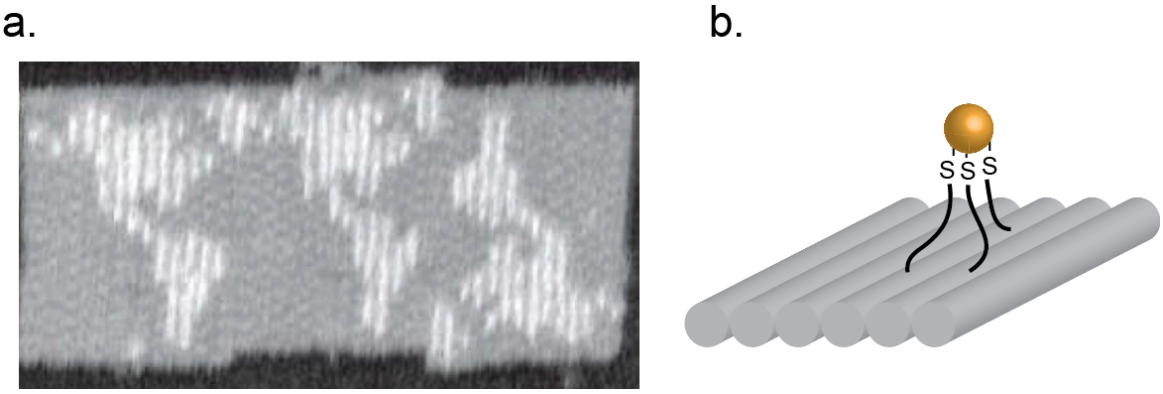


Figure 1.9: Functionalizing origami. (a) Multiple copies of map of the Americas, made from origami. Reprinted by permission from Macmillan Publishers Ltd: Nature, Ref. [5], copyright 2006. Land is represented as a raised area, created by specific placement of DNA hairpins on the surface of the tile. (b) Attaching a gold nanoparticle to an origami tile by extending and thiolating particular staples.

them to be formed within living cells, or with attachments such as proteins, which would be damaged by exposure to elevated temperatures. For such applications, it would be desirable to be able to control precisely the temperature at which assembly occurs, and in the case of structures based on single-stranded-tiles (described in section 1.3.10) this has already been achieved, by tuning the sequences, lengths and configuration of the component strands [102]. However, some of the resulting DNA objects did not appear to have a high degree of structural integrity.

### 1.3.7 Functionalizing origami

An origami tile may be thought of as an array of addressable pixels [5], where each pixel corresponds to a staple. Given that a plethora of techniques have been developed for conjugating molecules or particles to DNA, it will be seen that each staple could potentially be modified to carry a particle or functional molecule, which is then located with sub-nanometre precision on the origami surface.

For example, Rothemund created the map shown in Fig. 1.9a [5] by modifying selected staples with the addition of a hairpin loop of DNA, such that the loop projected from the origami tile, creating a raised feature. It has since become common to attach gold nanoparticles (AuNPs) to origami structures, linking the AuNPs to DNA via a thiol bond[21] (Fig. 1.9b).

Before the invention of DNA origami, gold nanoparticles had already been used as convenient DNA labels in various experiments. In 2000, gold-nanoparticle (AuNP)-labelled oligonucleotides were used for 'scanometric DNA detection', in which interactions between immobilized target strands and AuNP labelled strands were detected by means of a colour change [103]. Later, it was shown to be possible to use electrophoresis to isolate AuNPs carrying a particular number (1-5) of oligonucleotides [104]. Alternative chemical methods have also been developed to attach AuNPs to DNA [105], and assembly of an array of AuNPs has been demonstrated using a DNA lattice [106].

Initial attempts to attach gold nanoparticles to DNA origami proved unsuccessful [107] but later work demonstrated that yield could be improved by the use of bivalent linkers, where two thiol bonds connected each oligo to its AuNP [108]. In 2010, it was shown that gold nanoparticles could be assembled with a high degree of control on an origami platform when three DNA linkers were used for each AuNP [109].

Subsequently, large-area arrays of AuNPs were fabricated using lithographically confined DNA origami [110] and it was shown that AuNPs could even be encapsulated

[21]DNA oligonucleotides can be acquired commercially with a thiol (-SH) group at the end.

inside an origami cage [111]. The use of AuNPs in conjunction with DNA origami is of interest to the community because they can be used to construct plasmonic nanostructures, which can locally enhance the electric field of the incident light, resulting in an increase in the fluorescence of dye molecules, by over 100-fold [112]. AuNPs also scatter X-rays strongly, even in vivo [113], which makes them useful for X-ray measurements on DNA nanostructures.

DNA origami can also be coated with metal, for example via a seeding method, in which gold nanoparticles are attached to an origami object by standard techniques before further metallization occurs [114].

Illustrating the capacity to functionalize origami biochemically, Voigt *et al.* demonstrated that a range of reactions could be carried out on an origami tile; they extended selected staples in an origami tile and used them to create chemically different biotinylated linkers, arranged in a specific pattern [9]. These linkers were selectively cleaved if the reactions were successful, and the reaction progress was monitored by using AFM to examine the pattern of streptavidin molecules bound to the tile. Origami tiles have also been used to study the behaviour of photosensitized singlet oxygen [8].

It has also been shown that enzymatic ligation can be used for staple functionalization to enable modification of origami. For instance, single dideoxynucleotides can be added, modified with fluorophores or molecules such as biotin [10]. Furthermore, the enzyme terminal deoxynucleotidyl transferase can directly ligate DNA to nucleotide-triphosphates which have been coupled to large molecules like proteins [115].

### 1.3.8 Applications of DNA origami

Despite the fact that DNA origami was invented less than a decade ago, it has already been shown to have many potential applications. Arguably, one of the most significant is the targeted delivery of drugs using origami containers which can be opened selectively in the presence of diseased cells [7] (Fig. 1.10). The underlying principle has already been demonstrated in cell cultures, but for a similar approach to work *in vivo* it is necessary to use encapsulation or a similar technique to protect the structure from degradation by enzymes and prevent an immune reponse [116]. It has also been shown that the efficacy of the cancer drug doxorubicin in a cell culture is enhanced if the drug is adsorbed to a DNA origami bundle [6]. Furthermore, when origami structures are used to deliver doxorubicin it is possible to kill even those cells which are otherwise resistant to the drug [117].

One of the ultimate goals of nanotechnology is the creation of a nanoscale assembly line, and already the ability to attach molecules to DNA origami at precisely determined locations has enabled the creation of miniscule 'bioreactors' (Fig. 1.10b) [118]. These devices consist of origami structures with the enzymes glucose oxidase and horseradish peroxidase placed on the surface, separated by a distance of 15 nm. Hydrogen peroxide, the product of the reaction catalysed by the first enzyme, is the substrate for the second, and the close proximity of the two enzymes ensures that the local concentration of this molecule is very high indeed. Consequently, the catalytic behaviour of the enzymes is significantly enhanced in contrast to the control

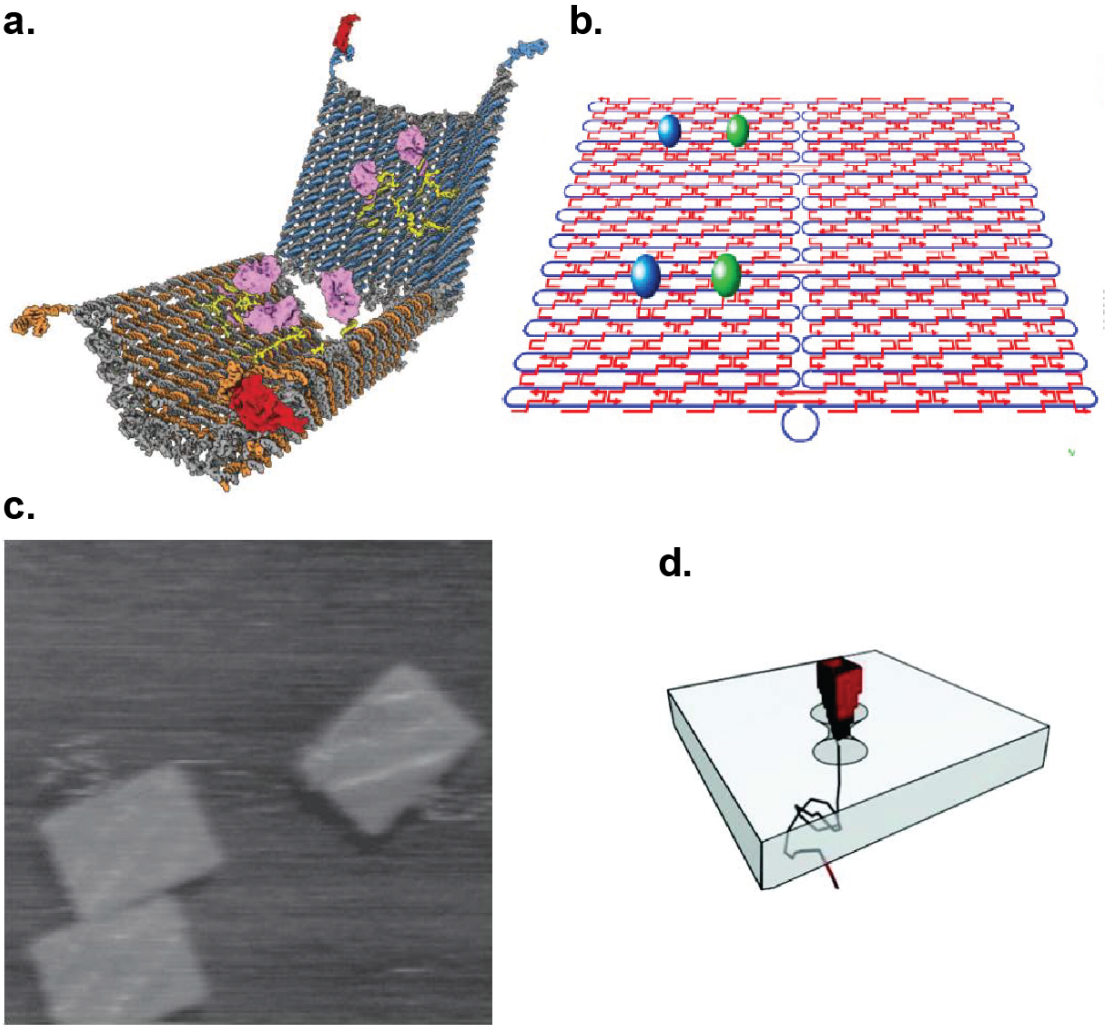


Figure 1.10: Applications of DNA origami. (a) A prototype 'nanorobot', which could have applications in targeted drug delivery. From Ref [7]. Reprinted with permission from AAAS. (b) A two-enzyme reaction cascade on an origami tile. Reprinted with permission from Ref. [118]. Copyright 2013 American Chemical Society. (c) A track for a DNA motor, constructed on an origami tile. Reprinted by permission from Macmillan Publishers Ltd: Nature Nanotechnology, Ref. [50], copyright 2011. (d) Nanopore built from DNA origami. Reprinted with permission from Ref. [119]. Copyright 2012 American Chemical Society.

experiment, in which the enzymes are free in solution.

As described earlier, origami has been used as a workspace on which to perform chemical reactions [9, 8], and directed chemical processes on origami could be facilitated by a DNA motor, such as the 'walker' reported by Wickham *et al.* [76], which operates using the mechanism shown in Fig. 1.3c and can navigate a network of tracks laid out on the origami using extended staples (Fig. 1.10c). In addition, DNA origami was employed in a study of the natural molecular motors kinesin and dynein [120].

DNA origami has been used to create a nanoscale ruler, a block of DNA with fluorophores attached in particular places, separated by a defined distance [11]. Also of interest is the use of DNA origami to fabricate nanopores - structures with a central channel through which molecules may pass [119, 121] (Fig. 1.10d). Such pores may have applications in biosensing or DNA sequencing.

### 1.3.9 Scaling up DNA origami

Individual origami tiles made using M13 or a plasmid-derived scaffold are limited to a size scale of around 100 nm, but increasing the size of origami-based structures would increase the potential of the technology. It has been shown that origami tiles can be joined together to form nanoribbons, which may be linear or twisted, by the addition of multiple linkers which bridge an intermolecular scaffold seam [122]. The monomers were folded and purified before polymerization, which took 14 days. The length of the resulting structures was in the micrometre range.

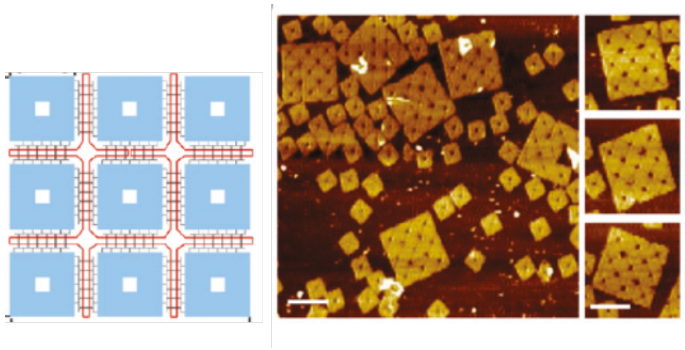

Figure 1.11: Superorigami - tiles held in place by pre-formed scaffold frame; schematic and AFM image. Array of 9 squares measures 240×240 nm. Scale bar: 200 nm. Reprinted with permission from Ref. [123]. Copyright 2011 American Chemical Society.

Similarly, corrugated (zig-zag) origami tiles of size 100×52nm, in which there was a 120-degree angle between adjacent helices, have been assembled into long ribbons using two alternative sets of 12 strands [124]. The longest observed ribbon consisted of 40 tiles, which were offset from each other, and it was also shown that the tiles could be rolled into a nanotube.

Zhao *et al.* showed that it was possible to create 'superorigami' structures by folding a PhiX174 scaffold into a pre-formed frame and using it to arrange M13-tiles into a particular layout [123] (Fig. 1.11). It has also been demonstrated that helper tiles made from comparatively short strands can be used in place of standard staples to fold M13 into a designed shape [125].

Origami tiles can also be assembled like jigsaw pieces; Rajendran *et al.* synthesized tiles which could be joined together via a mortice-and-tenon joint [126], where the joint was stabilized by staples forming inter-tile connections. Subsequently, Woo and Rothemund showed that jigsaw-piece origami tiles could be assembled by shape

complementarity alone, stabilized by base-stacking, without inter-tile stapling [127].

Similar objectives have been addressed using smaller tiles, made without a scaffold by assembly of comparatively short strands, and it has been shown that substrate-assisted-growth can lead to formation of crystals with area of order $10 \times 10^5\,\mathrm{nm}^2$ [128].

### 1.3.10 Successor technologies

Building on the success of DNA origami, ideas have begun to emerge for the next new strategy in DNA nano-engineering. For example, the technique christened 'DNA Gridiron' [129] is a means of assembling lattices from four-arm junctions which are analogous to Holliday Junctions [97]. The resulting structures can be flat, twisted or curved. Surprisingly, it has also proven to be possible to construct objects from a large number of small single-stranded bricks. It had previously been thought that this would require more precise control over stoichiometry of strands than could be achieved in practice [130], but this is not the case, although the yield of more complex structures can be low. The single-stranded tile (SST) method, illustrated in Fig. 1.12, was initially used to make two-dimensional shapes [131]. A $3 \times 7\,\mathrm{nm}$ rectangular canvas was created, and by omitting a selected sub-set of SSTs, 107 distinct two-dimensional shapes were made. The technique was quickly extended to three dimensions [132], and various shapes were fabricated.

As the field develops, it is probable that other new technologies will arise. It is also possible that DNA configurations other than the standard B-form structure

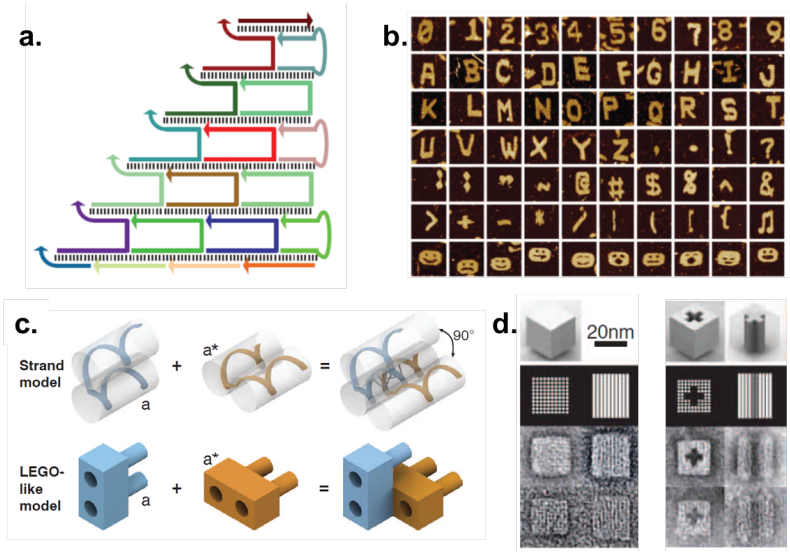


Figure 1.12: (a) The principle of assembling single-stranded tiles into a 2-dimensional structure. (b) AFM images of some of the shapes which can be made by omitting selected tiles from a single-stranded-tile canvas. (a) and (b) reprinted by permission from Macmillan Publishers Ltd: Nature, Ref. [131], copyright 2012. Each square is 150×150 nm in size. (c) and (d): Expanding the single-stranded-tile concept to three dimensions. From Ref. [132]. Reprinted with permission from AAAS. (c) Concept. (d) Two example structures, showing the design above the TEM images.

described earlier may find further use; the G-quadruplex has already been used for the construction of a switchable nanostructure [133].

## 1.4 Drawing inspiration from Nature

In living systems, most of the processes which are essential to life are carried out or facilitated by proteins, which have complex three dimensional structures, intricately linked to the functions they perform.

A protein is a long chain of amino acids linked by peptide bonds [14]. A typical protein contains 50-2000 amino acids, but the largest known protein (titin) has 27000 residues [14]. The connection between the amino acid sequence of a protein and its

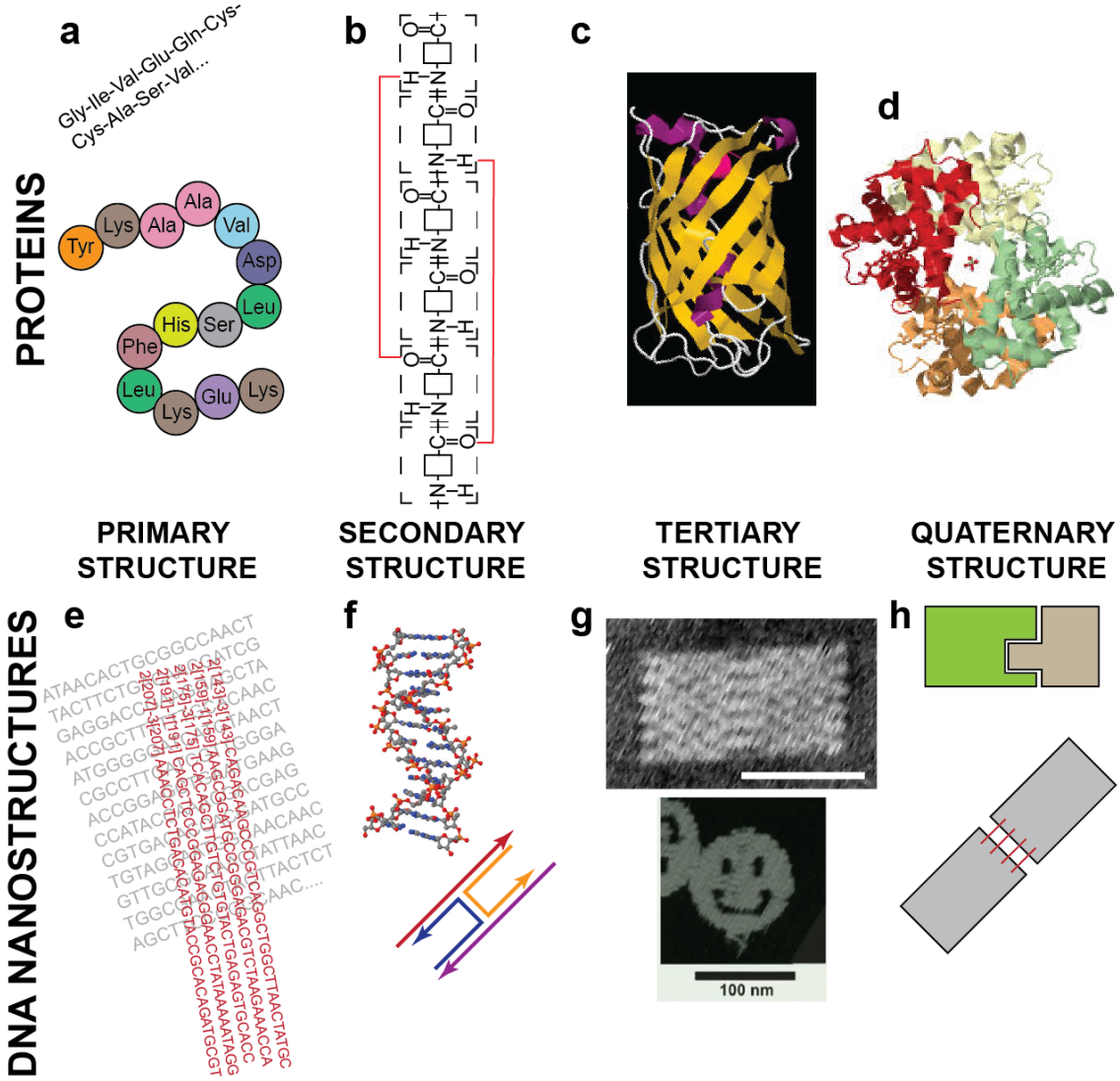


Figure 1.13: Hierarchy of structural organization in proteins (top) and DNA nanostructures (bottom). Primary structure: sequence of amino acids (a) or bases (e). (b) Example of secondary structure motif for proteins: alpha helix. Red lines indicate hydrogen bonds between oxygen atom in CO group of $i^{th}$ residue with hydrogen atom in NH group of $(i+4)^{th}$. In DNA origami structures, secondary structure motifs are double helices and crossovers - (f). (c) Example of tertiary structure in a protein - 'beta barrel', a configuration of a set of beta sheets, as seen in Green Fluorescent Protein (GFP). Image of GFP from the RCSB PDB (www.pdb.org) of PDB ID 1GFL, Yang *et al.*, Nat.Biotechnol.14: 1246-1251 (1996). Equivalent level of structure in DNA nanostructures is overall shape adopted - smiley face, rectangle etc (g). (Smiley face as already shown in Fig. 1.4, rectangular tile described later in thesis.) The quaternary structure for both proteins and DNA nanostructures is the association of individual monomers into multimers. Protein shown in (d) is the tetramer human haemoglobin - image from the RCSB PDB (www.pdb.org) of PDB ID 1HHO, Shaanan, J.Mol.Biol.171: 31-59 (1983). (h) Two mechanisms for joining origami tiles together - extended staples and base-stacking (with shape complementarity).

three-dimensional form is exceedingly complicated, and the molecular structure is hierarchical. The primary structure of a protein is the sequence of its constituent amino acids (Fig. 1.13a) and the secondary structure is the three-dimensional arrangement of the chain. Key secondary structure motifs are the alpha helix (Fig. 1.13b) and the beta sheet, both stabilized by hydrogen bonding. The tertiary structure is the arrangement of the secondary structure motifs into structures such as the beta barrel (seen in Green Fluorescent Protein, Fig. 1.13c) which is a set of beta sheets. The quaternary structure is the association of a set of several protein subunits ('monomers', or single units) to form a 'multimer' such as the haemoglobin tetramer (Fig. 1.13d).

Analogously, the primary structure of a DNA origami structure could be defined as the base sequence of its component strands (Fig. 1.13e), while the secondary structure consists of the double helix and crossovers of strands between helices (Fig. 1.13f). This is not the standard definition of secondary structure in a DNA context - normally the term 'secondary structure' refers to hairpins or similar motifs, formed when a single strand folds up on itself.

According to the scheme I discuss here, the tertiary structure is the overall shape adopted (e.g. smiley face or rectangle - Fig. 1.13g) and the quaternary structure is the association of DNA nanostructures to form multimers, examples being origami tiles joined by modified staples [122] or by base-stacking between complementary shapes [127](Fig. 1.13h).

## 1.5 Other synthetic self-assembled systems

Protein folding has been the subject of intense research for many years, the ultimate objective being a full understanding of folding principles, enabling protein structure to be predicted solely from amino acid sequence and allowing synthetic proteins to be designed from scratch. Protein folding is exceedingly complicated because there are 20 different amino acids, subject to a range of different non-specific interactions. In contrast, DNA nanostructure assembly is a more accessible problem - although not fully understood, this involves folding of strands made from only 4 types of building block, which interact via highly specific base-pairing interactions. Despite the complexity of the protein folding problem, significant advances have been made in the development of artificial protein nanostructures. For instance, a 16 nm cage has been constructed from protein subunits using fusion proteins [134]. One domain from the naturally trimeric protein bromoperoxidase was fused with another from a virus protein, which is usually dimeric. 12 fusion constructs self-assembled into a cage, the structure of which was confirmed using X-ray crystallography.

Studies of synthetic $\alpha$-helical protein fibres revealed that assembly proceeded via a nucleation and growth mechanism and that the formation process could be manipulated and controlled [135]. It has been reported that coiled-coil motifs can be selectively assembled into fibres and discrete nanostructures by modifying the length of the linker joining the two coils [136]. Furthermore, peptide units designed from scratch to form coiled-coil bundles have been used to self-assemble lattices which roll

up into cage-like structures [137], in accordance with molecular dynamics simulations.

Artificial protein assemblies have a range of possible applications, from encapsulation, which is potentially useful for raising the local concentration of biocatalysts, to engineering fibres and designing more efficient vaccines [138].

RNA has been studied as a nanoscale construction material, but less extensively than DNA[22]. RNA nanostructures may be built using similar principles to those made from DNA, but it is also possible to draw on the different structures formed by RNA in nature [139]. It has been shown that rationally designed RNA assemblies can be formed in bacteria, and used as a scaffold for enzymes catalysing hydrogen production, with the result that their activity is enhanced [140].

Non-organic materials may be self-assembled into biologically-inspired structures such as protocells [141], which may be able to respond to environmental changes [142].

## 1.6 Closing remarks

I have explained in this Chapter how DNA, the ubiquitous biomolecule, can be used as a nanoscale engineering material. I have discussed the class of nanostructure known as DNA origami, and provided a summary of key achievements in the field. Two dimensional (planar) origami shapes and three-dimensional structures can be assembled, and twist or curvature may be introduced.

DNA origami folds over a narrow temperature range (a few °C at most), somewhere

[22]Search using Thomson Reuters Web of Science on 7$^{th}$ January 2014: Topic=(RNA Nanotechnology) - 483 results *cf* Topic=(DNA Nanotechnology) - 2145 results.

between 50 and 70°C. The sharpness of the transition suggests that the process is co-operative, but to date no explanation of the mechanism has been provided. Various synthesis protocols have been used, all of which share the need to denature double-stranded sections of the scaffold and the requirement for positive ions capable of screening the charge on the DNA backbone.

These ideas have helped to advance the field of DNA origami assembly. However, important questions remain unanswered:

1. Are there pathways by which DNA origami folds?

2. If so, what factors shape the folding pathways?

3. Can we control the self-assembly process?

4. To what extent does origami assembly mimic folding of proteins?

5. How do staples co-operate during folding?

In this thesis I will set out answers to these questions, which I addressed by studying a *polymorphic* origami object. In DNA origami structures, the sequence of every piece of the scaffold is usually unique, so each staple has one and only one perfectly complementary position to bind in. However, if each scaffold domain is duplicated then every staple has a choice of binding sites and multiple staple layouts are possible. The arrangement of the staples determines the final shape of the origami object, so if a scaffold with duplicated staple binding sites is used, a range of differently shaped folded objects will result and the origami nanostructure will be *polymorphic*. The

distribution of shapes will depend on the factors influencing folding and therefore examining the shape distribution provides insight into assembly. In the Chapter which follows, I will describe how I designed, synthesized and characterized a reference structure, for comparison with the more complex polymorphic object, to be discussed separately later.

# Chapter 2

# A reference structure for investigations into DNA origami assembly

The reference structure is a 'monomer', in which every staple has a unique binding site; the staples fold the scaffold into a simple rectangle, to be referred to as a tile. In contrast, the polymorphic origami tile I describe later is a 'dimer', because it is made from a scaffold consisting of two copies of the monomer sequence, joined head-to-tail. Here I will explain the design features of the monomer tile and describe the synthesis of the scaffold. I will also present AFM images of the monomer tile, showing that it folded as expected and tends to align preferentially along certain axes when deposited on a mica surface.

## 2.1 Design

Fig. 2.1a illustrates my monomer tile design, produced using caDNAno [64]. The tile is rectangular, predicted to have a length of 75 nm and a width of 35 nm. These values are calculated by assuming that one base-pair is 0.34 nm high, the diameter of the DNA double helix is 2 nm and there is a gap of 1 nm between helices [5]. The inter-helix gap arises due to the electrostatic repulsion between helices, the backbones of which are negatively charged.

The tile has several important design features. First, across the centre of the tile there is a seam, which is bridged by pairs of staples. Second, as a simplification over the standard 3-domain staple design, all staples have two domains except at the upper and lower edges of the tile, where some staples have only one domain (Fig. 2.1a). A domain is typically 16 bases long, except where a base is deleted to compensate for unwanted twist [57]. Using domains of this length allows a highly symmetrical lattice of staple crossovers to be formed, and a 16bp duplex should be highly stable at room temperature. For reference, a 16 base-pair duplex with 50% G-C content is predicted to have a melting temperature of 54°C, while an 8bp duplex is expected to melt at 14°C and for 4bp the melting temperature is below 5°C[1].

Every U-shaped staple on the short edges of the tile has 4 unpaired thymines at each end. This inhibits inter-tile stacking, which if not suppressed would hamper efforts to identify the conformation of the dimer tiles.

[1]Calculations performed using NUPACK [71], for oligo concentrations of 20 nM, [Mg]=12.5 mM, [Na]=5mM, with 'some dangle' treatment. Sequences used: $(\texttt{ATGG})_{n=1,2,4}$ and $(\texttt{CCAT})_{n=1,2,4}$.

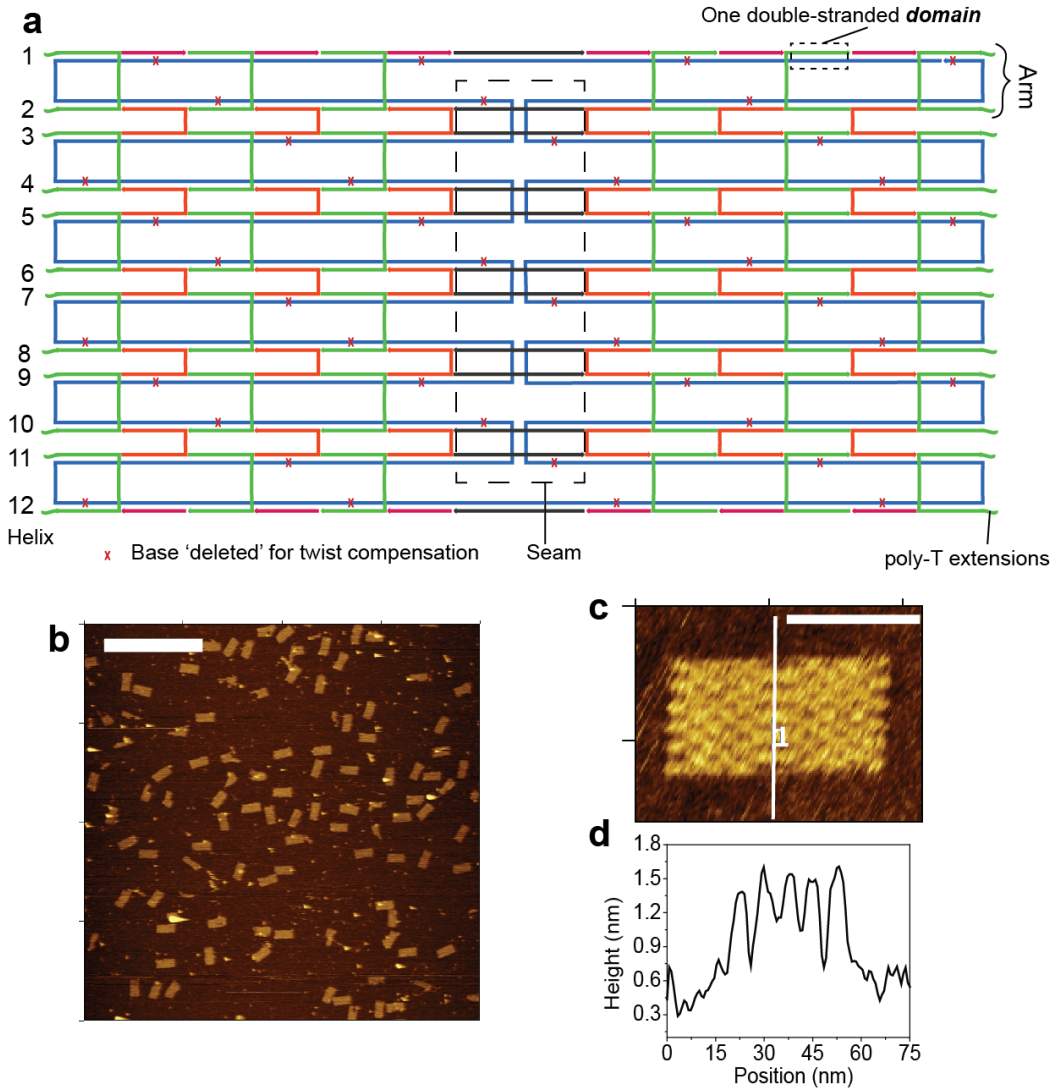


Figure 2.1: Origami monomer tile. (a) Design of tile, produced using caDNAno as described in the text. Colours are used to distinguish between staples. Poly-T extensions of staples at left and right edges inhibit inter-tile stacking. Helices are numbered 1-12. (b),(c) AFM images of monomer tiles, synthesized with 10 nM scaffold and 100 nM staples in 1×TA (pH8.3), with 12.5 mM magnesium acetate, using standard annealing protocol. Images acquired with Agilent 5500 AFM, using as imaging buffer 1×TAE (pH8) with 12.4-12.5 mM magnesium acetate (plus 4 mM $NiCl_2$ for (b)). Scale bars: 500 nm (b) and 50 nm (c). Images processed using Gwyddion (software for analysis of Scanning Probe Microscope images - available from `gwyddion.net`). (d) Height profile, along line marked (1) in (c).

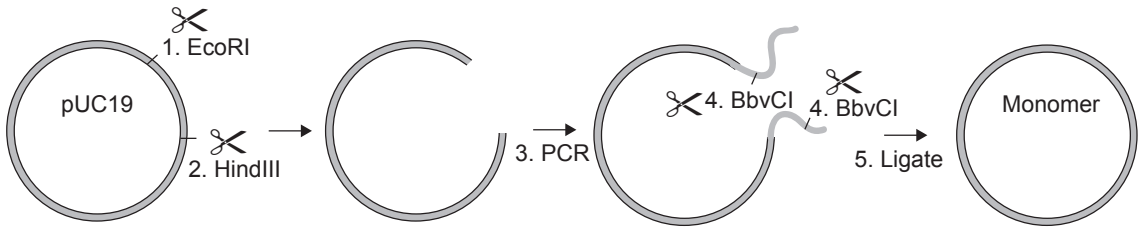


Figure 2.2: Synthesis of the monomer plasmid - schematic. The steps involved are: 1 & 2 - trimming plasmid pUC19 with restriction enzymes, 3 - PCR (including insertion of BbvCI recognition site), 4 - trimming with the restriction enzyme BbvCI and 5 - ligation of sticky ends.

The tile is highly symmetrical, without any single-stranded scaffold in the final design. Most origami objects are constructed using the bacterial genome M13mp18 [5], but this has a particular secondary structure domain which is generally left unstapled because otherwise it might interfere with folding. It is also considerably longer than required for the design shown here. The formation of a large single-stranded loop would disrupt the symmetry and potentially influence folding, and therefore I synthesized a customized plasmid as the source of the scaffold.

## 2.2 Synthesis of the monomer plasmid

I derived the plasmid from pUC19 by enzyme treatments and PCR as follows[2] (Fig. 2.2). Initially I used the restriction enzymes EcoRI and HindIII to cut open the plasmid pUC19 (successive digests, in appropriate buffers), producing a double-stranded DNA segment ready to be subjected to a PCR process. I used customized primers, such that the PCR product contained an additional length of DNA bearing the recognition site for the restriction enzyme BbvCI, enabling the product to be cut with

[2]pUC19 is a plasmid which is available commercially.

BbvCI. The final step was to join the sticky ends by incubation overnight at 16°C with the enzyme T4 DNA ligase. Transfection into *E. coli* DH5$\alpha$, followed by culture of the bacteria and extraction of the plasmid using standard methods, produced large stocks for use in experiments (see Appendix A). I had the inserted region sequenced to confirm formation of the correct product. The plasmid synthesis protocol and sequence are provided in full in Appendix A.

Fig. 2.3 shows that the monomer plasmid is of the correct size and behaves as expected when digested with restriction enzymes - BbvCI cuts the plasmid into linear fragments, while Nt.BbvCI and Nt.BspQI nick the plasmid, converting it into open circle (OC) form.

## 2.3 Origami synthesis: monomer tile

To prepare the scaffold for origami synthesis, successive enzyme digests were required [5]. I converted the plasmid into open circle (OC) form, with one strand broken and the other intact, using the nicking enzyme Nt.BspQI, and then removed the cut strand by digestion with the enzyme ExoIII, which attacks the exposed 3' end of a DNA strand (see Appendix A for detailed protocol). Fig. 2.3a illustrates the result of each stage.

The next step was the removal of the enzymes, by phenol:chloroform extraction, which separates protein and DNA as a result of their differing solubilities in phenol [87]. The extracted DNA could then be further purified if required, by means of

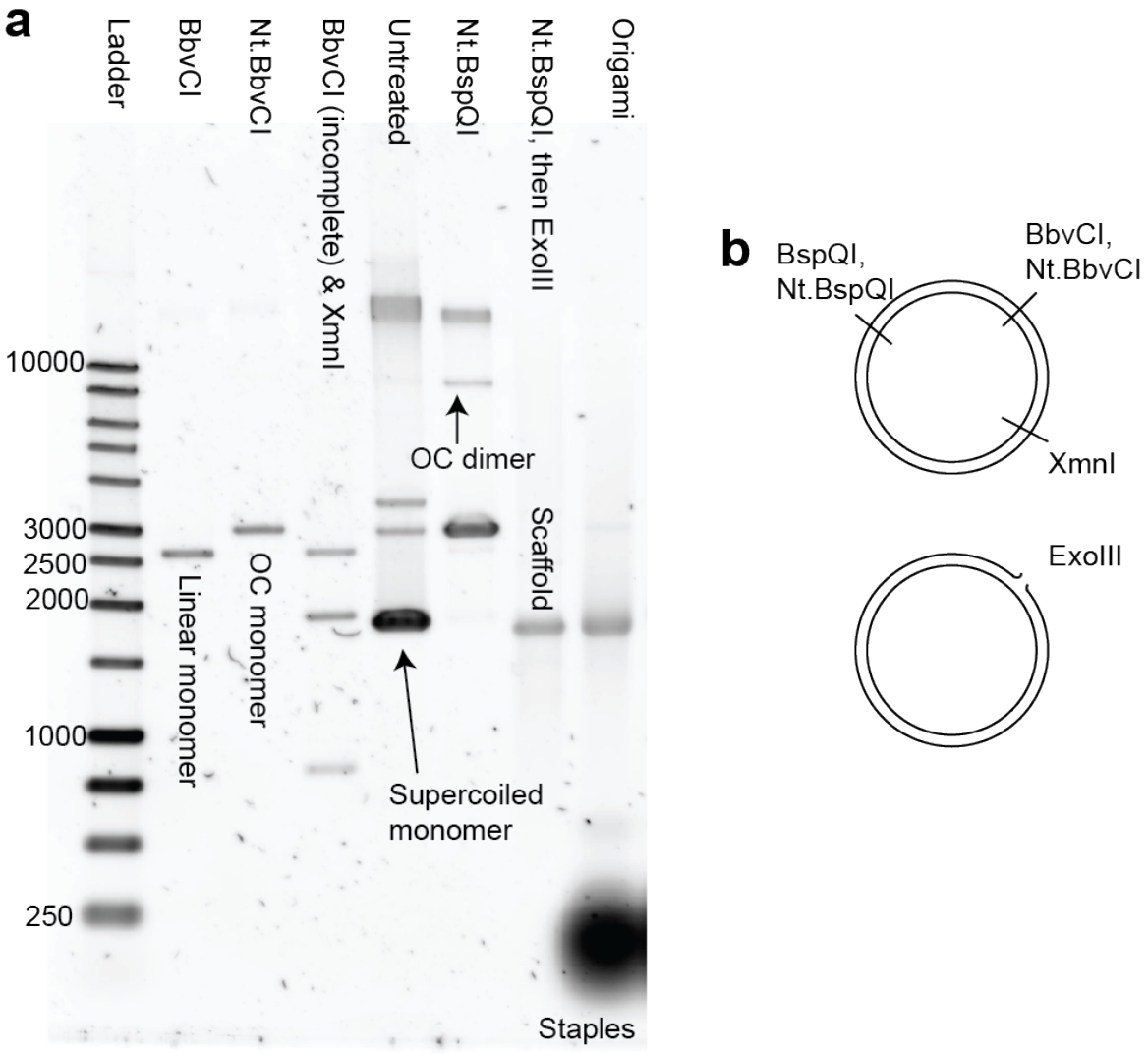


Figure 2.3: Monomer plasmid and origami - testing with agarose gel electrophoresis. (a) Agarose gel showing monomer plasmid before and after treatment with various enzymes as indicated. Gel: 1% agarose with 1×TAE buffer, run at 75 V for 2 hours and stained with SYBR Gold. Nt.BbvCI and Nt.BspQI are nicking enzymes, converting plasmid to open circle (OC) form. Apparent difference in pattern of bands following nicking with alternative enzyme results from difference in concentration of gel samples. BbvCI and XmnI convert plasmid to linear double-stranded form. ExoIII converts double-stranded nicked plasmid to single-stranded form by digestion from free 3' end. (b) Schematic illustration of positions of recognition sites for the various enzymes.

ethanol precipitation [87] (Appendix A). The origami could then be folded by annealing (Chapter 1 and Appendix A). Filtration through a size exclusion resin enabled removal of unbound staples prior to AFM imaging [50].

## 2.4 Characterization of the monomer

I studied the monomer tile using Atomic Force Microscopy (as described in Chapter 1), in tapping mode in fluid. The AFM images (Fig 2.1b,c) show that the tile folds well and is of approximately the expected size. The measured dimensions are typically slightly larger than predicted, in part due to the broadening which arises because the AFM tip is not infinitely sharp, and in part because the inter-helix gap may be slightly larger than 1 nm. The height profile (Fig 2.1d) shows that the tile is approximately 1.2 nm high, which is less than the diameter of DNA in solution because the tile is slightly flattened as a result of electrostatic interactions with the surface [5]. For the work described in this thesis, the typical AFM imaging buffer was 1×TAE (see Appendix A) with 12.5 mM magnesium acetate and 4 mM $NiCl_2$, where the nickel was used to enhance origami adhesion to the mica [143].

It has been observed previously [5] that if DNA origami was synthesized with a lower concentration of staples than scaffold, the overall yield of well-folded objects was low. Fig. 2.4 shows an AFM image of the monomer synthesized with a scaffold concentration 10 times higher than the staple concentration, and it will be seen that tiles do not fold well, in agreement with the previous study.

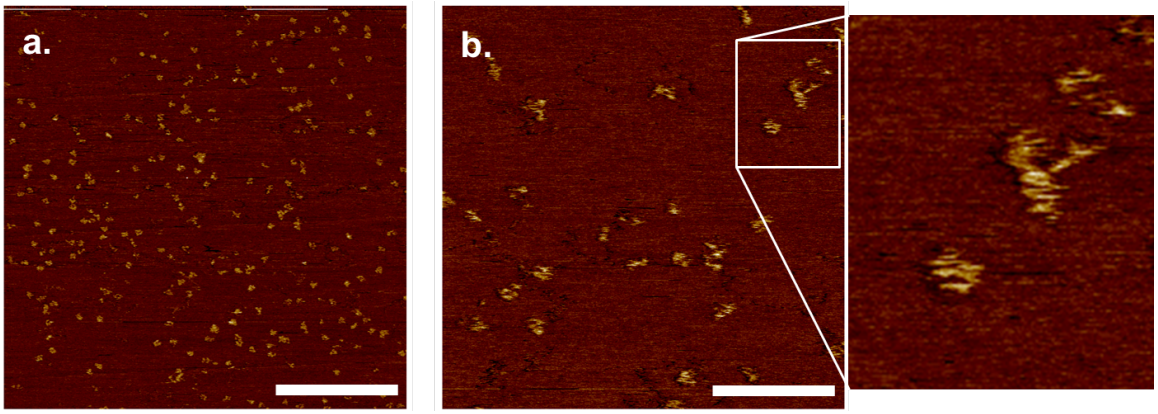


Figure 2.4: Attempt to synthesize monomer tile with excess of scaffold. Staple concentration was 1/10 the scaffold concentration. (a) 3.3 $\mu$m×3.3 $\mu$m field of view, scale bar is 1 $\mu$m. (b) 1 $\mu$m×1 $\mu$m field of view, scale bar is 330 nm.

## 2.5 Orientation of tiles on mica

Images showing large fields of view reveal that well-folded tiles appear to preferentially align along certain axes (Fig. 2.1). This effect is present in samples imaged with different microscopes and under different conditions, with different batches of mica.

A good estimate of tile orientation may be found using the MATLAB function `regionprops` on a binarized AFM image of the tiles [144]. This function computes properties of objects in a binary image, including the orientation of the ellipse which has the same second moments as each shape. Due to the symmetry of the shapes involved, the orientation of the ellipse as computed by MATLAB for an ideal rectangular tile is in fact the same as the orientation of the tile itself (Fig. 2.5a).

Calculating the orientation of the equivalent ellipse for each tile yields the distribution of alignment angles which is presented in Fig. 2.5b, with the expectation for a random arrangement. It is obvious that the experimentally observed distribution is not random, and that peaks occur at around 5° and 70°, with a separation of 65°.

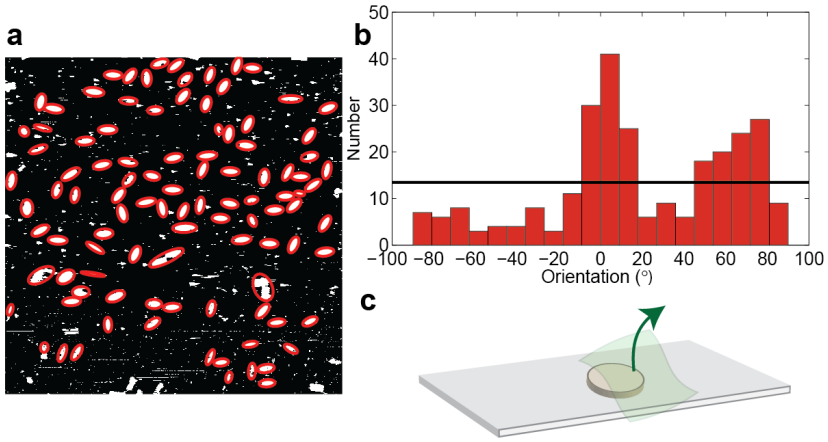


Figure 2.5: Orientation of well-folded monomer tiles on mica. (a) Binarized AFM image of tiles (made with 10-fold excess of staples over scaffold - same image as in Fig. 2.1). Each red ellipse has the same second central moment as the tile shape to which it corresponds - calculation performed by MATLAB function `regionprops`. (b) Orientation of tile-equivalent ellipses for a set of three 2 $\mu$m-scale AFM images taken of the same sample in the same imaging session. Two peaks are visible. Black line represents expectation for random distribution of angles. (c) Illustration of how mica is cleaved - tape is stuck to surface and then peeled away in one direction.

Mica has a hexagonal crystal structure, which means that there are three principal crystallographic axes, separated by 120° [145]. The most likely explanation for the distribution of tile orientation angles is that the tiles are aligning with terraces on the mica surface. The angles of these terraces would be expected to follow the crystal axes, but because the terraces are formed when the mica is cleaved with tape (Fig. 2.5b) their edges should be parallel primarily to two of the axes, not all three, as one of the crystal axes will be close to the direction in which the tape is being pulled. This would create two preferred tile alignment angles, separated by 60°, almost the same as the difference observed experimentally.

Previously, lithographically patterned surfaces have been employed to prepare

large-area arrays of aligned origami structures [110], but the finding that tiles exhibit a bias for particular orientations on the untreated mica surface suggests that simpler methods could be used to align tiles effectively.

## 2.6 Summary

I have discussed in this Chapter the monomer origami tile I designed as a reference for my studies of origami folding. The tile is rectangular and measures around 40 nm $\times$ 80 nm. AFM images show that tiles deposited on mica do not adopt random orientations, and align preferentially along axes separated by approximately 65°.

# Chapter 3

# Real time observation of origami assembly

As described earlier, a typical origami assembly protocol involves cooling scaffold and staples in solution from near-boiling to room temperature. At the cooling rate used for the work described in this thesis, the process takes 70 minutes. To understand how the staples fall into place during assembly, it is necessary to monitor them in real time. Several approaches have been used to do this [69, 75, 90, 92], as reviewed in Chapter 1. None of the techniques employed so far have allowed data to be obtained on single staples, although one method [90] allowed the behaviour of pairs of similar staples to be examined. I will explain in this Chapter the new experimental procedure I developed to probe the behaviour of individual staples. This enabled me to make measurements in real time during assembly in solution, and later in this thesis I will present results I acquired using the technique I describe here.

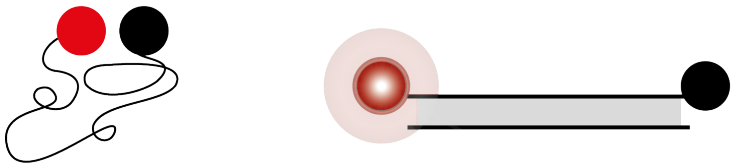

Figure 3.1: Principle of the fluorescence experiment with dual-labelled strands. When the labelled strand is free, fluorescence is effectively suppressed by the quencher. When the labelled strand hybridizes with the complementary strand, the fluorophore and quencher are far apart and fluorescence is high.

## 3.1 New approach: principle

When a staple is free, its ends are on average very close together. Upon incorporation into an origami tile, the separation of the ends increases to approximately 5 nm for a U-shaped staple and nearly 11 nm for a seam staple. Hence, if a staple is labelled with a fluorophore at the 5'-end - either Cy5 or Cy3 - and a quencher at the 3'-end, the fluorescence will increase dramatically when it binds to the scaffold. The increase will be greatest for the seam staple, for which the change in separation is largest. The same effect will be observed for such a dual-labelled strand binding to its reverse complement to form a duplex (Fig. 3.1).

Here, I measured fluorescence, $F$, as a function of temperature, $T$, using an Agilent Technologies Stratagene MX3005P RT-PCR machine using the standard origami annealing rate. The transition temperature for each experiment is defined here as the centre of a Gaussian function, fitted to the derivative $\frac{\mathrm{d}F}{\mathrm{d}T}$, where the data was normally averaged over 6 duplicates of the sample. Quoted values for the transition temperature in a sample refer to averages taken over at least 4 repeats of the same experiment. For supplementary experimental details see Appendix A.

## 3.2 Tests

Initial tests indicated that 20 nM was a good target concentration for bright fluorophores, and demonstrated that there was minimal cross-talk between the Cy3 and Cy5 channels (i.e. there was practically no signal in the Cy3 channel from a sample containing only Cy5 and vice versa). All traces had a non-zero positive offset - even when the sample was non-fluorescent (ultrapure MilliQ water). For the fluorophore-only sample, the signal exhibited near-linear dependence on fluorophore concentration (studied at 5-80 nM), over a range of temperatures from 25°C to 95°C, as expected.

Fig. 3.2 shows the normalized background-subtracted fluorescence intensity for each of two non-interacting fluorescently-labelled strands of DNA, measured with a mixture of the two strands. The intensity of the emitted fluorescence depends strongly on temperature, but the variation is smooth and there are no sharp transitions. This confirms that any peaks observed in the derivative of the measured fluorescence are not attributable to any intrinsic properties of the fluorophores themselves.

The other section of the figure illustrates the behaviour of the dual-labelled staple strands in the absence of any strand with which to hybridize. The data shows that the fluorescence is quenched very efficiently; the signal is approximately equal to that measured from a 'blank' sample consisting only of ultrapure MilliQ water (data not shown). This result demonstrates that the contribution to the signal from unincorporated staples will be low, and that any sharp transition observed in the region of interest is due to a change of state of the DNA strands.

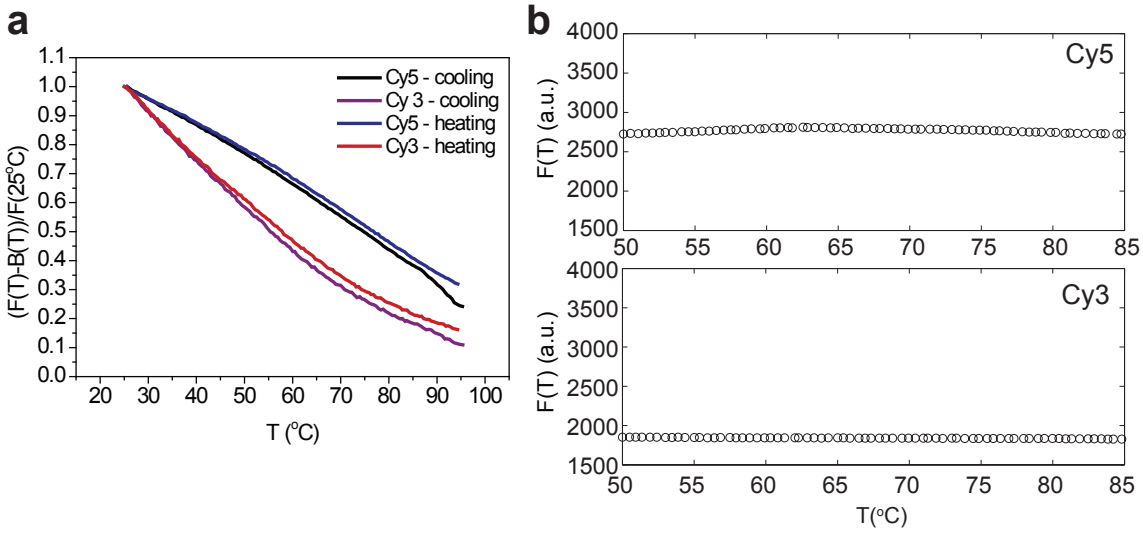


Figure 3.2: (a) Intrinsic temperature dependence of fluorescence of Cy3 and Cy5. The background signal ($B(T)$) was measured using a blank sample (ultrapure MilliQ water) and subtracted from the recorded fluorescence ($F(T)$), and the data was normalized by dividing by the background-corrected fluorescence at 25°C ($F(25°\mathrm{C})$). Sequences of strands were: /5Phos/AGGAATGTTCGAATGTGAGTTTCGA/3Cy3Sp/, /5Cy5/GGTGATAAAACGTGTAGCAAGCTGTAATCGACGGGAAGAGCATGCCCATCCACTACTATGGCG (notation - Integrated DNA Technologies). (b) Fluorescence signal during annealing, from sample containing unhybridized dual-labelled staples, in the absence of any other strands. For both (a) and (b) 6 samples were averaged, each of volume 25 $\mu$L; the buffer was the standard origami synthesis buffer (1×TA + 12.5 mM magnesium acetate) and the average rate of heating/cooling was 1°C/min. All strands were at a concentration of 20 nM and both fluorophores were present in the same solution.

| | Staple | $T_H$(°C) | NUPACK prediction (°C) |
|---|---|---|---|
| **Case 1** | U (7[48]-6[48]) | 68.7 (0.2) | 72.4 |
| | S (6[127]-6[96]) | 80.22 (0.05) | 85.4 |
| **Case 2** | U (4[175]-5[175]) | 78.2 (0.2) | 86.8 |
| | S (5[96]-5[127]) | 72.2 (0.1) | 74.9 |

Table 3.1: Hybridization temperatures for staple strands combined with reverse complements, in absence of scaffold. $T_H$ is the hybridization temperature measured experimentally (extracted by fitting Gaussian functions to dF/dT), and the number in brackets beside each is the standard error on the mean of measurements made. For NUPACK analysis, [Na]=0.05 M (minimum allowed in program; actual [Na] value was nominally zero), [Mg]=0.0125M. Concentration of each DNA strand was 200 nM.

## 3.3 Reference experiment: DNA duplex formation

For the work described in this thesis I selected four particular staples - two seam staples and two U-shaped staples - to be labelled with a quencher and a fluorophore (Cy3 for seams, Cy5 for U-shaped)[1]. As a reference experiment, I studied the hybridization of these staples with their reverse complements to form duplexes, for two different cases. Case 1 corresponds to the situation in which the seam staple duplex was predicted to be more stable than the U-staple duplex, and in Case 2, the seam staple duplex is weaker that that of its U-shaped counterpart. The results are shown in Fig. 3.3, and the hybridization temperatures are shown in Table 3.1, with the predictions made using NUPACK. The transition temperatures predicted by NUPACK are consistently a few degrees higher than those measured experimentally, but the trend is similar.

[1]Tests I conducted with the polymorphic tile revealed that swapping the two fluorophores had no effect.

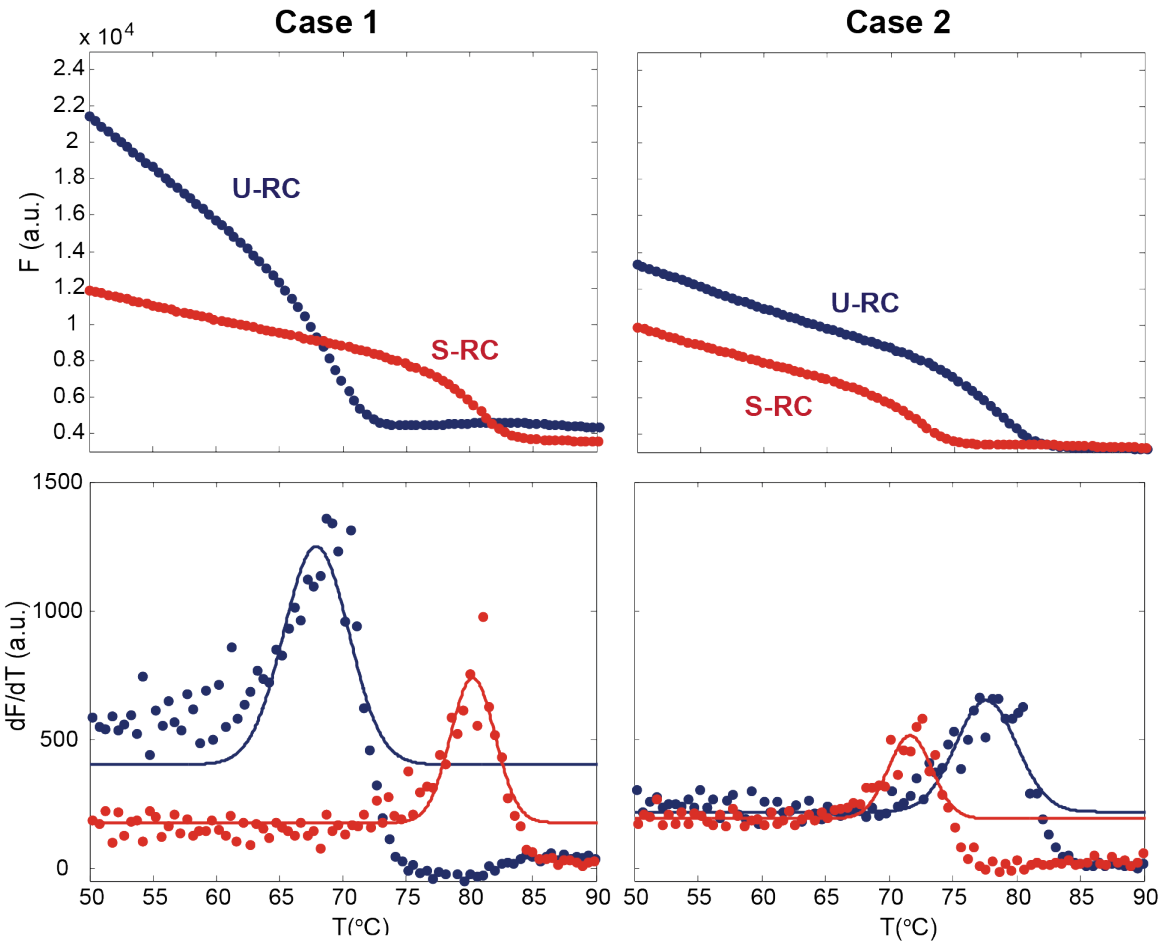


Figure 3.3: Hybridization of pairs of DNA strands to form duplexes. The dual-labelled seam and U- staples were mixed with their reverse complements in 1× standard origami synthesis buffer and annealed according to the standard protocol. The concentration of all strands was 200 nM. S-RC denotes the traces corresponding to seam-duplex formation, and U-RC to those corresponding to U-duplex formation, and the colour code in the lower figures is the same as that in the graphs above. Filled circles represent experimental datapoints, solid lines represent Gaussian fits. The results are not normalized, which gives rise to a non-zero value for dF/dT at temperatures below that of the transition, due to the intrinsic temperature dependence of the emitted fluorescence. The fluorescence, F, is scaled arbitrarily by the gain settings of the RT-PCR machine.

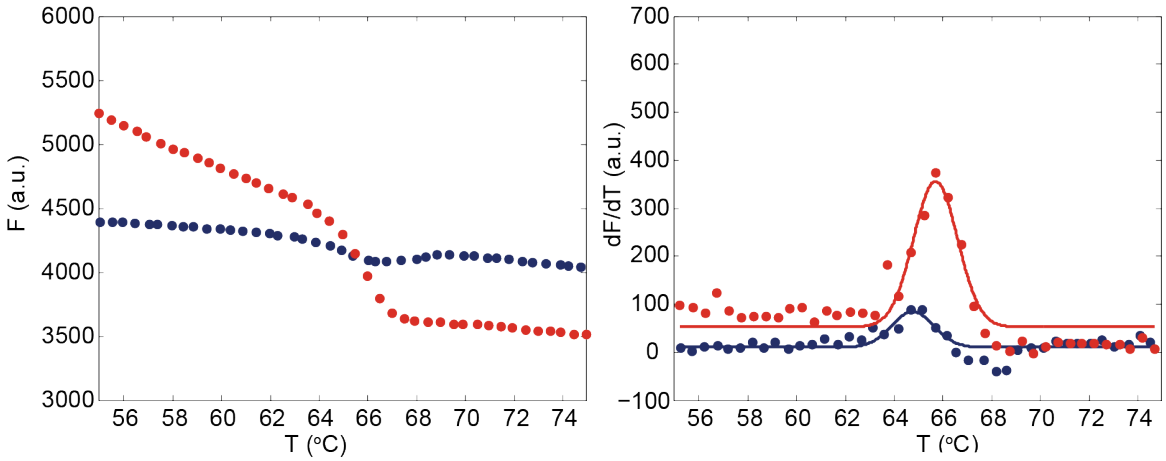


Figure 3.4: Incorporation of staples into monomer tiles. Data for a seam staple (red) and a U staple (blue) is presented. Left: fluorescence, $F$, as a function of temperature, $T$. Right: $dF/dT$. Filled circles represent experimental datapoints, solid lines in the right-hand plot represent Gaussian fits. Scaffold - 20 nM, staples - 200 nM.

## 3.4 Studying origami tiles

I monitored the incorporation of the Case 1 staples into the monomer tile, and representative results are shown in Fig. 3.4. Above the critical temperatures at which the staples are incorporated into the origami, the fluorescence is low, increasing as the transition is passed. The derivatives of the raw fluorescence data are peaked at the transition, and it will be seen that the seam staple falls into place slightly before the U-staple. Over several measurements, the difference in incorporation temperatures was 0.7°C (standard error on the mean: 0.1°C).

The fluorescence experiments I performed were designed to provide information on the sequence of events during origami assembly, and hence the cooling/heating rate used was identical to that employed during a standard origami folding protocol (1°/min). This is comparatively rapid and consequently all measurements displayed hysteresis, even those made on duplexes. Thus the hysteresis is difficult to interpret.

## 3.5 Summary

I have explained in detail the approach I developed to monitor origami assembly in real time by studying the behaviour of individual staples. I have discussed the application of the method to the monomer tile, and for reference I examined duplex formation. For the duplexes, transition temperatures are slightly lower than predicted by NUPACK. In Chapter 6, I present the results of various experiments I performed using the technique described here, and I use them to study assembly of the polymorphic tile.

# Chapter 4

# Assembly of polymorphic tiles succeeds against the odds

To synthesize the polymorphic origami tiles I used a dimer scaffold consisting of a single loop containing two copies of the monomer sequence. For the original design of my polymorphic tile, the staple set was the same as for the monomer, but with a dimer scaffold every staple domain could bind in one of two possible locations. This results in $2^{76}$ potential staple arrangements, most of which correspond to misfolded structures. Here I describe the synthesis of the dimer scaffold and present experimental results which demonstrate that a significant proportion of the polymorphic tiles fold well, despite the huge potential for misfolding. I will show that the same finding applies to other design variants; this suggests that there are pathways for folding and that origami assembly is not random. The existence of assembly pathways is attributable to cooperation between staples, and I will discuss how this arises.

## 4.1 Synthesis of the dimer scaffold

A small proportion of the plasmid stock prepared for the monomer study had dimerized, and I prepared a stock of pure dimer by extracting and cloning the dimer plasmid from the original 'monomer' stock. When the plasmid has been nicked using the restriction enzyme Nt.BbvCI it is possible to identify the different components based on mobility in an agarose gel (Appendix A). Excision of the appropriate band from the gel, followed by leaching of DNA from the gel fragment into ultrapure water, produced a dilute stock of nicked dimer plasmid. By cloning this plasmid using *E. coli* DH5$\alpha$ as for the monomer (Appendix A) I obtained a pure stock of dimer.

The product is shown in a gel in Fig. 4.1a, with the results of diagnostic enzyme reactions and comparable monomer samples. The gel shows that the untreated and open circle (nicked) dimer plasmid (lanes 6 & 7 respectively) run more slowly than the corresponding monomer samples (lanes 2 & 3 respectively), reflecting the larger size of the dimer. There is no evidence of monomer contamination in the dimer.

When treated with the enzyme BbvCI, the dimer is cut into linear monomer-sized fragments (lane 8 for dimer, lane 4 for equivalent monomer reaction), demonstrating that every dimer contains two copies of the BbvCI recognition site and that the two halves are the same size. Further proof that the dimer was a homodimer (both halves identical) was provided by a digest carried out with two enzymes, BbvCI and XmnI, where the BbvCI reaction was incomplete. The dimer and the monomer gave the same pattern of three bands, which confirmed that the two halves of the dimer have

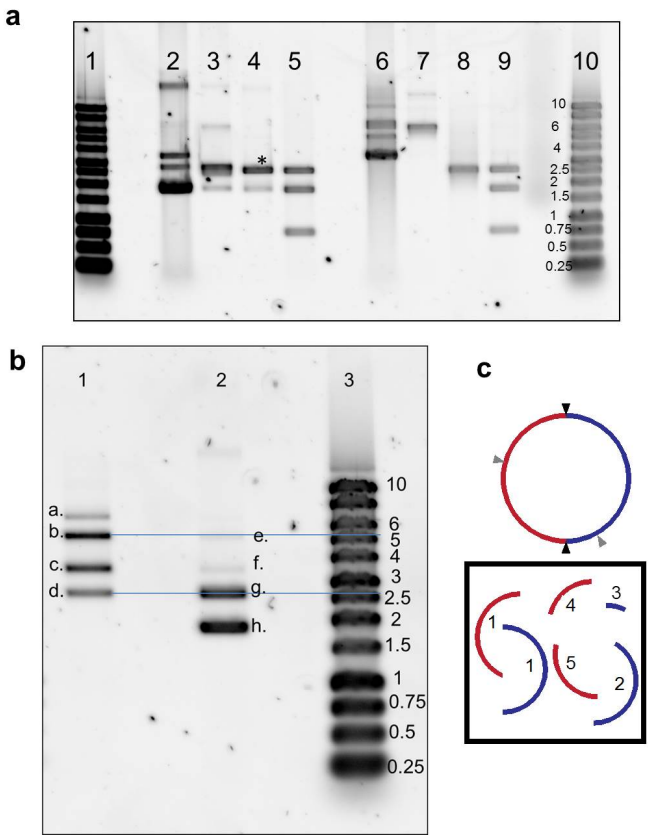


Figure 4.1: Synthesis of dimer plasmid - testing with agarose gel electrophoresis. Both gels: 0.7% agarose gel, 1×TAE buffer, stained with SYBR Gold. (a) **Monomer and dimer plasmids, as prepared and following enzyme digests**. Lane 1: ladder, labels same as Lane 10. **Lane 2-5: Monomer plasmid.** Lane 2: undigested. Lane 3: nicked with Nt.BbvCI. Lane 4: cut with BbvCI (marked band is linearized monomer); some supercoiled monomer remains undigested in lanes 3 and 4. Lane 5: double digest with XmnI and BbvCI (complete with former, incomplete with latter). **Lane 6-9: Dimer plasmid.** Lane 6: undigested. Lane 7: nicked with Nt.BbvCI. Lane 8: cut with BbvCI, yielding only linearized monomer-size fragments. Lane 9: double digest with XmnI and BbvCI (complete with former, incomplete with latter). Lane 10: Ladder, labels give size of ds fragments in kbp. (b) **Results of incomplete digest with BbvCI. Lane 1: Dimer.** a. Undigested, nicked dimer. b. Linearized dimer (plasmid cut only once) - comparison with ladder shows that it is of the correct size. c. Undigested, supercoiled dimer. d. Monomer-size linear fragments. **Lane 2: Monomer (contains some dimer component).** e. Very faint. Linearized dimer. f. Undigested supercoiled dimer. g. Linearized monomer. h. Undigested supercoiled monomer. Lane 3: Ladder (labels: size in kb). (c) The products which would be expected in Lane 9 of part (a) if the two halves of the dimer were not the same. Black triangles denote the sites at which XmnI acts, grey triangles are the recognition sites for BbvCI. If the two halves of the dimer are identical, fragment 4 is the same as fragment 3, while fragment 2 is the same as fragment 5.

the same pattern of XmnI and BbvCI recognition sites, because otherwise this digest would have yielded five bands for the dimer (Fig. 4.1c).

The formation of a small quantity of linearized dimer allowed its size to be found by comparison with a standard ladder. This involved an incomplete digest with the enzyme BbvCI, in which the DNA was severed at some but not all BbvCI recognition sites. The product of this reaction is shown on the gel in Fig. 4.1b, and the data demonstrates that linearized dimer is the correct size.

## 4.2 AFM images of polymorphic tiles

I synthesized dimer tiles using the same protocol as for the monomers, with the same set of staples. The monomer tile is rectangular, and therefore the well-folded dimer tiles have the appearance of a pair of joined rectangles (Fig. 4.2). Different arrangements are possible for the two rectangles, and so various overall shapes are observed and the tile is **polymorphic**.

Fig. 4.3 and Fig. 4.4 show large field-of-view AFM images of eleven different design variants of the polymorphic tile. The assembly of each variant will be discussed fully in Chapter 6, where full details of each design will be presented[1]. In all the AFM images, the height of most objects is between 1.2 and 1.5 nm, exactly as expected for single-layer origami structures. Most importantly, in all cases except one, many of the tiles fold well. The significance of this will be discussed in the next section.

[1]See also: the Appendices.

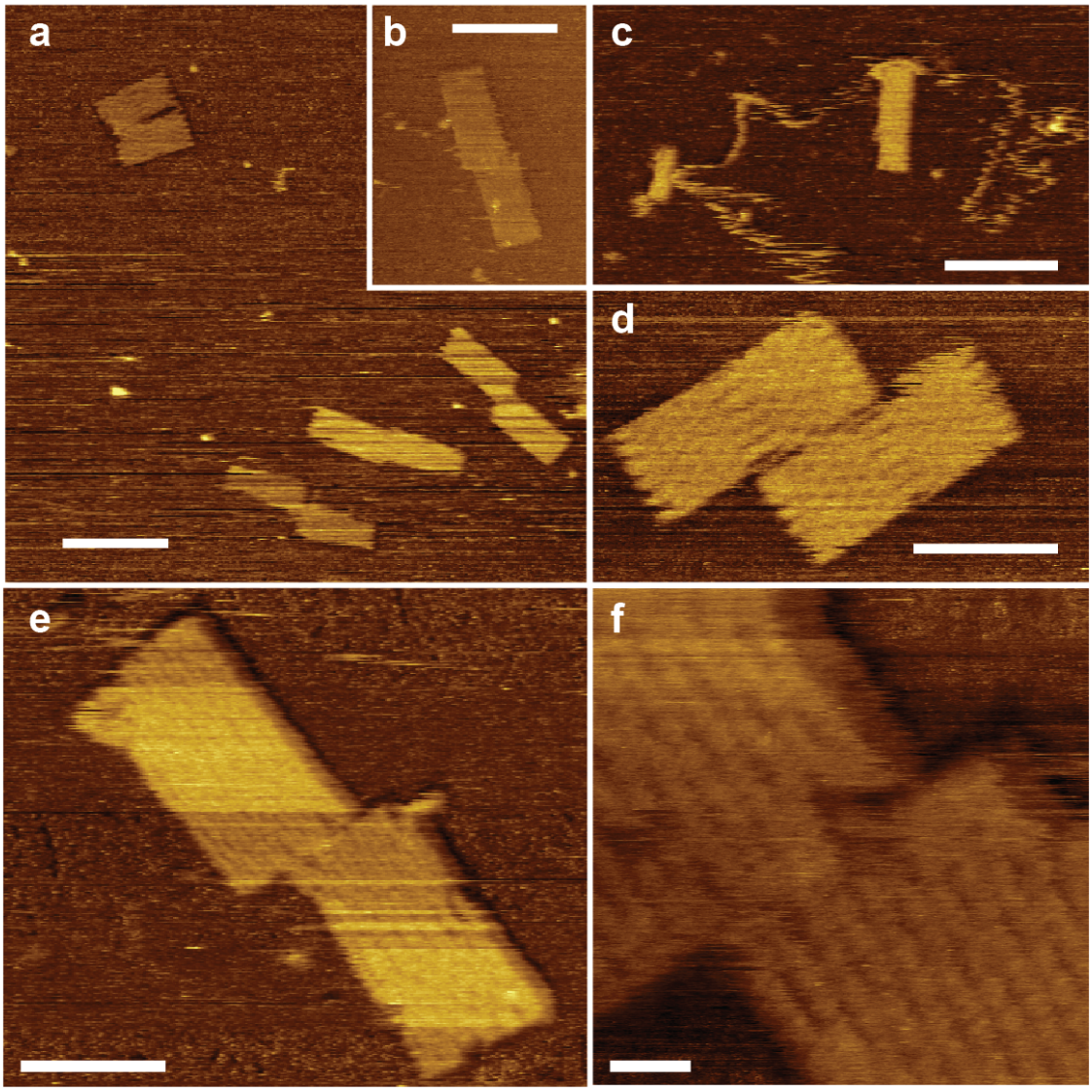


Figure 4.2: AFM images of polymorphic (dimer) tile, acquired with Agilent 5500 AFM. (a)-(c): scale bars represent 100 nm. For (d) and (e), scale bars represent 50 nm. For (f), the scale bar represents 10 nm. (c) shows a mis-folded structure, (f) is a higher magnification scan of the tile in (e). All samples were prepared in standard synthesis buffer (1×TA with 12.5 mM magnesium acetate) using standard annealing protocol, and were subjected to three rounds of purification before imaging. During synthesis, staples were present at higher concentration than scaffold in each case; staple excess is defined as staple concentration divided by scaffold concentration, and was 8.2 for the samples shown in (a,b,d,e,f) or 16.4 for (c). Imaging buffer used was: 1×TAE (pH8), 12.4 mM magnesium acetate with 4 mM $NiCl_2$ for (a) and (d); 1×TAE (pH8), 12.5 mM magnesium acetate for (b); 1×TA (pH8.3), 12.4 mM magnesium acetate with 4 mM $NiCl_2$ for (c,e,f).

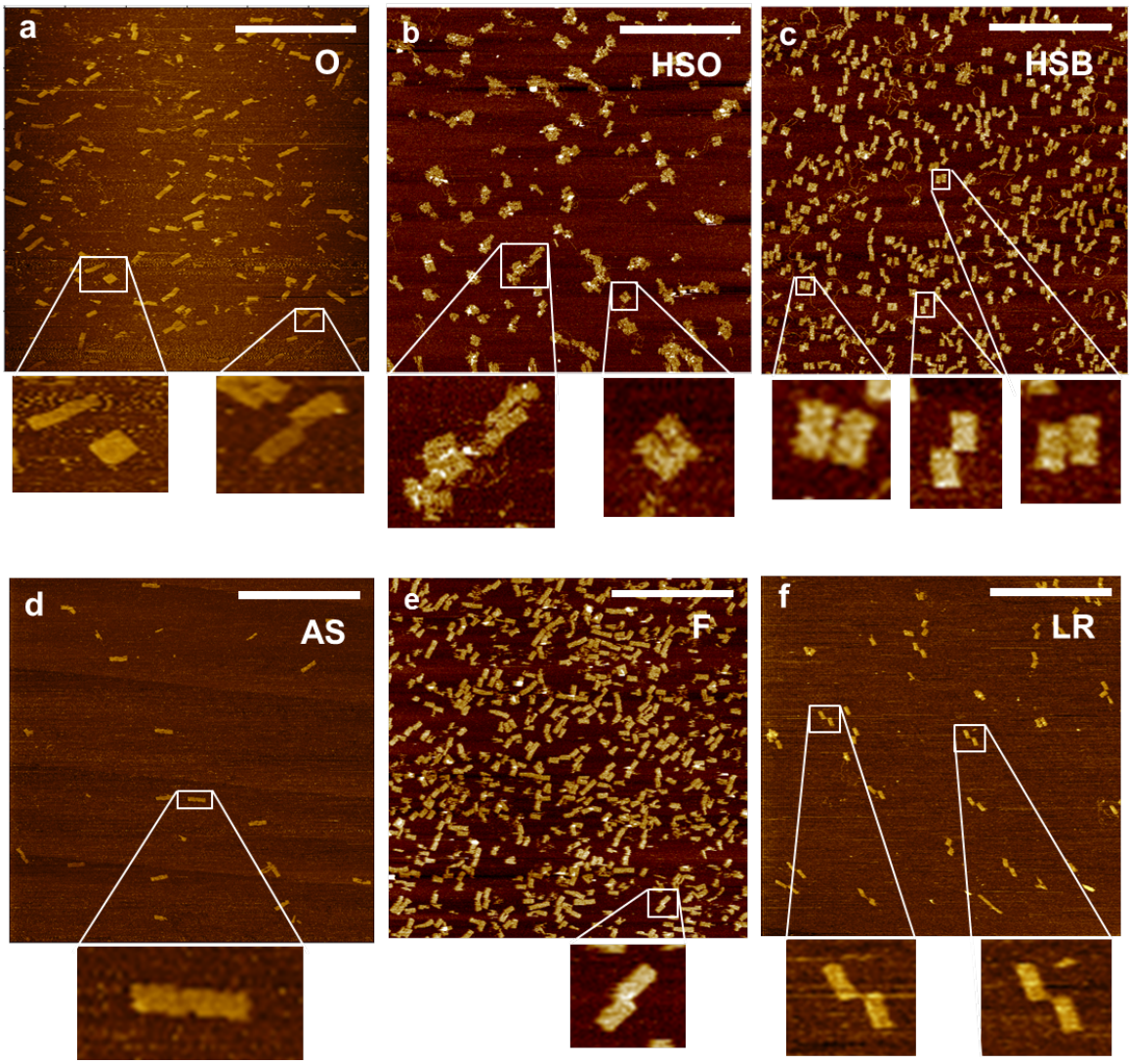


Figure 4.3: AFM images of the polymorphic tile, original design and variations on the theme (O, HSO, HSB, AS, F, LR). Tile designs are shown and discussed in full in Chapter 6; designs and specimen details are provided in the Appendices. All fields of view are $3\,\mu$m$\times 3\,\mu$m in size (512×512 pixels) and scale bars represent $1\,\mu$m. (a) Original design (O). (b) Half-seam-omitted (HSO). (c) Half-seam-broken (HSB). (d) Alternative-seam (AS). (e) Failed attempt to Force tile to fold into particular configuration (F). (f) Modification in the lower right corner (LR) of the rectangle. AFM imaging buffer: 1×TAE (pH 8-8.3) with 12.4-12.5 mM magnesium acetate and 4 mM $NiCl_2$.

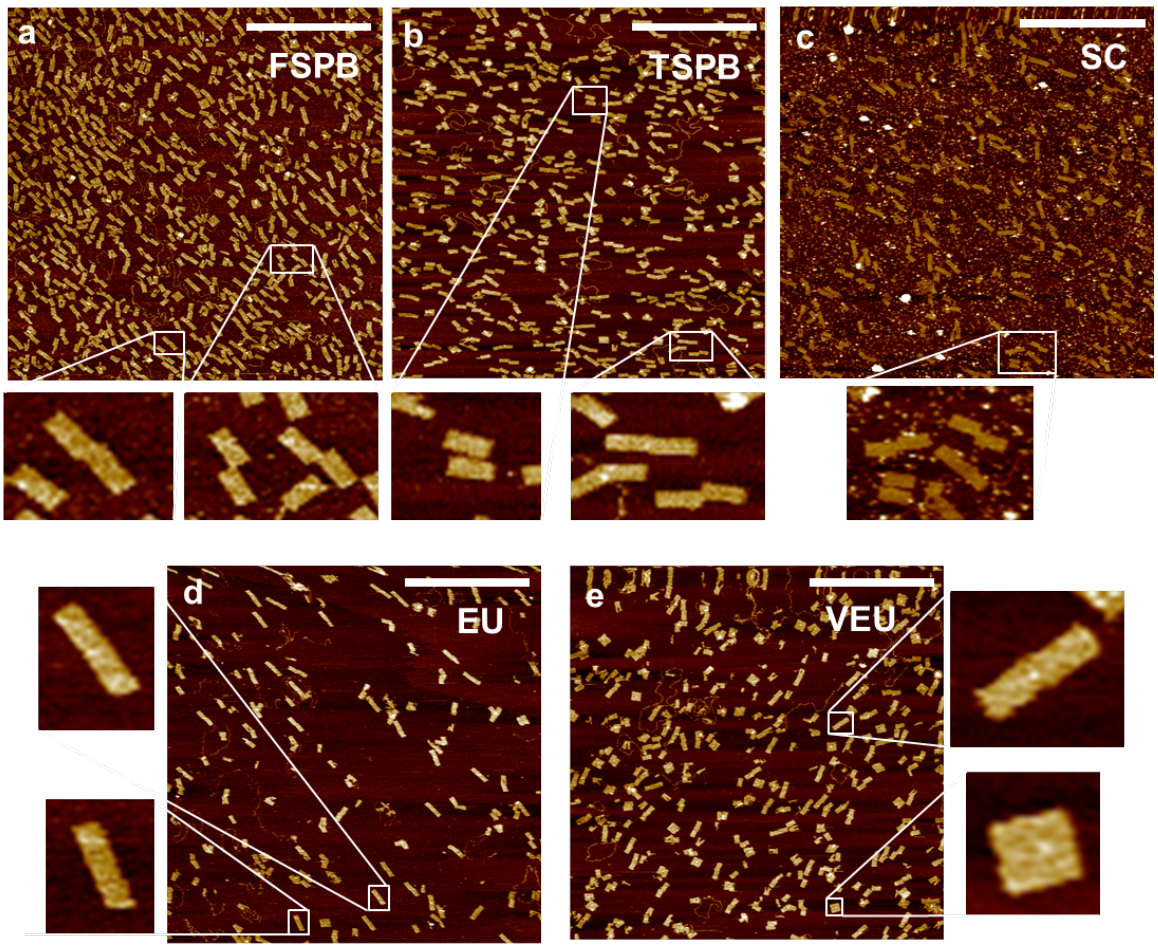


Figure 4.4: AFM images of the polymorphic tile, variations on the theme.(FSPB, TSPB, SC, EU, VEU). Tile designs are shown and discussed in full in Chapter 6; designs and specimen details are provided in the Appendices. All fields of view are $3\,\mu$m×$3\,\mu$m in size (512×512 pixels) and scale bars represent $1\,\mu$m.(a) First Seam staple Pair Broken (FSPB) (b) Third Seam staple Pair Broken (TSPB) (c) Tiles made with different staple concentrations (SC) - concentration of some staples is so elevated that many remain after purification. (d) Extended U-staples (EU) (e) Very extended U-staples (VEU). AFM imaging buffer: 1×TAE (pH 8-8.3) with 12.4-12.5 mM magnesium acetate and 4 mM $NiCl_2$.

## 4.3 Folding pathways

### 4.3.1 Polymorphic tiles fold well despite huge potential for misfolding

Fig. 4.5 shows the estimated yield of well-folded tiles for the design variants characterized quantitatively by digital image processing as described in the next Chapter. Here, the yield is defined as the percentage of observed tiles of approximately the correct size which are well-folded. It is generally around 35-60%, which is surprisingly high. This illustrates that sequence duplication within the scaffold does not preclude folding.

In the polymorphic tile each staple may connect either two sites in the same half of the dimer or sites in opposite halves. There are 76 two-domain staples, each with two choices, which means that in total there are $2^{76} \approx 7.6 \times 10^{22}$ possible staple binding patterns[2]. Most of these will not correspond to a well-folded tile, and therefore it might be expected that only a tiny proportion of tiles would fold well, because of the huge number of ways in which the folding could go wrong. It is therefore remarkable that a significant proportion of tiles do fold well. A similar puzzle was identified many years ago in protein folding, and it is known as the 'Levinthal Paradox' after the biologist who first articulated it [146].

[2]This calculation gives the number of configurations with all staples bound, under the assumption that all such arrangements are sterically feasible. In fact, many arrangements will be prohibited by geometrical constraints, but a full calculation would also include the number of states with some staples unbound or partly bound. It is impossible to calculate accurately the number of geometrically allowed states, and $2^{76}$ is a reasonable estimate.

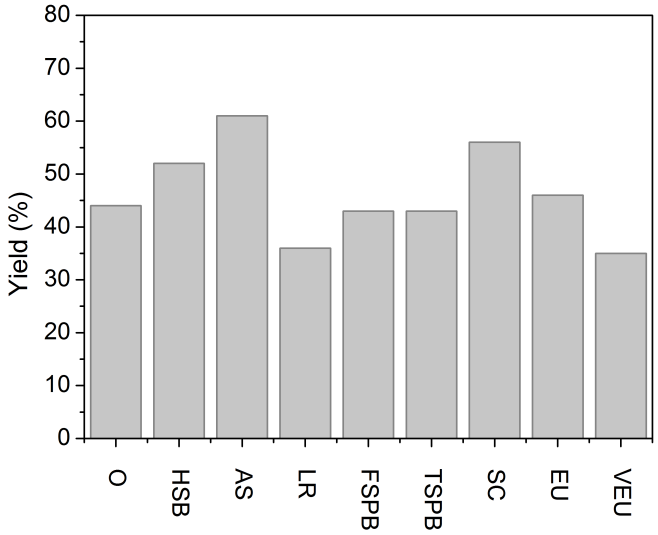


Figure 4.5: Yield of well-folded tiles for all design variants characterized quantitatively. The percentage yield is defined as the proportion of approximately the correct size which are well-folded (the calculation will be discussed in Chapter 5).

### 4.3.2 Analogy with Levinthal's Paradox

The Levinthal Paradox states that the number of possible states for a folding protein is so huge that a protein would require a time greater than the age of the Universe to fold if it explored all configurations randomly [146]. If this were the case, life as we know it would not exist, and in fact proteins fold on an exceedingly rapid timescale (some fold within milliseconds [147]). It is therefore obvious that they do not find their final structure via a random search, and the paradox was resolved with the theory that proteins fold via pathways [148] - there are intermediate semi-folded states through which a protein passes *en route* to its final structure (Fig. 4.6).

This theory has evolved since its inception, and the modern view is that proteins fold by navigating through a complex 'energy landscape' towards the minimum energy state which is the folded structure [149]. The energy landscape may be viewed as a

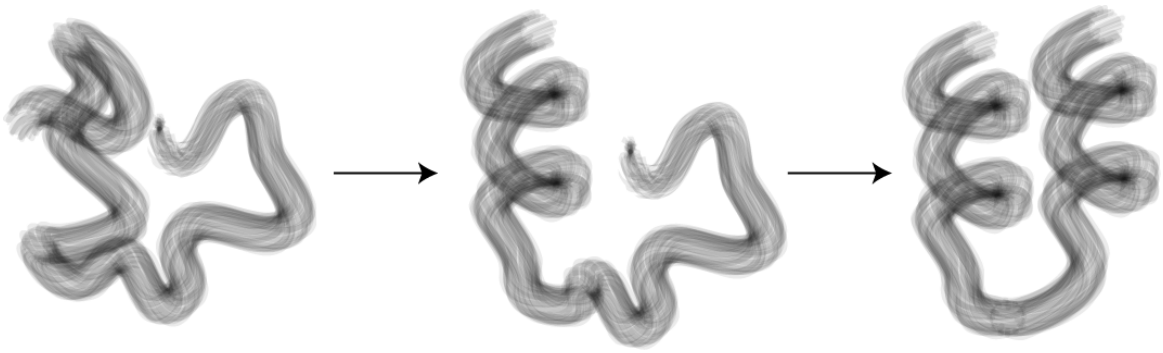

Figure 4.6: A hypothetical protein folding pathway. The unfolded chain (left) folds by successive formation of secondary structure units; here one helical domain forms and then the next folds.

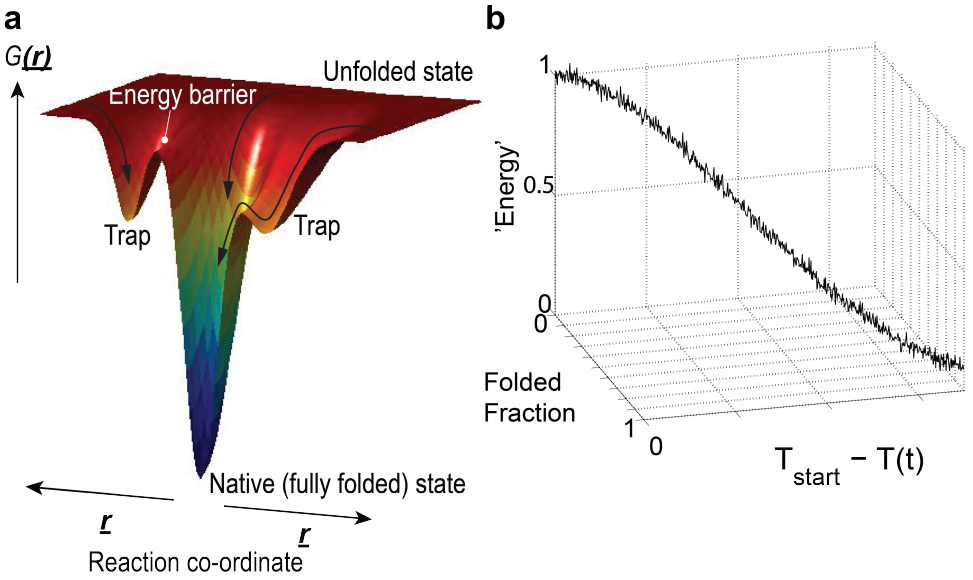


Figure 4.7: Energy landscapes and folding pathways for proteins and DNA origami. (a) Sketch of a hypothetical energy landscape for a protein, indicating the position of unfolded states, the native state and misfolded (trap) states. $G(\underline{r})$ is the energy of a state at reaction co-ordinate $\underline{r}$. An energy barrier can exist between a misfolded state and the native configuration, as shown. Arrows represent possible trajectories through the landscape towards energy minima. (b) Qualitative representation of a folding pathway for a DNA origami structure. The vertical axis represents 'free energy', and the pathway is sketched for the temperature range over which the folding transition occurs. As the solution is cooled, the folded fraction tends to 1 and the system tends towards the free energy minimum. $T(t)$ is the temperature at time $t$ and $T_{\text{start}}$ is the temperature at the start of the transition.

mathematical function $G(\underline{r})$. Here $\underline{r}$ is a 'reaction co-ordinate', usually defined only qualitatively, which represents the state of a protein in the process of folding, and the quantity $G(\underline{r})$ is the energy of that state. A sketch of an energy landscape is depicted in Fig. 4.7; in addition to the primary minimum it has secondary minima which function as traps. If the protein reaches a trap there may be insufficient energy for it to surmount the barrier which then exists between it and the folded state, in which case it will remain in the trap, a partially folded quasi-stable state. In the 'energy landscape' framework, a 'folding pathway' may be regarded as a favoured trajectory through the landscape, ending in the primary minimum. The energy landscape has a funnel-like shape which directs the protein to the native conformation [150].

Origami synthesis is usually accomplished using an annealing protocol. Assembly progress can be quantified by the number of fully bound staples, and the other reaction co-ordinate is temperature, or time. An individual scaffold follows a folding pathway, defined by the staple binding events which take place. The energy landscape is exceedingly complex, and it is more convenient to phrase arguments in the context of pathways. Fig. 4.7 shows a qualitative representation of a folding pathway for a tile which folds well. A pathway is a preferred sequence of staple binding events occurring on the way from the unfolded state to the fully folded final state. At a given point on the pathway, the partly folded structure has a certain energy, and a well-folded configuration corresponds to an energy minimum.

Thus, Levinthal's paradox may be re-phrased in the context of the polymorphic

tile. The annealing time for origami synthesis is 4200 seconds, so if the staples explored all $2^{76} \approx 7.6 \times 10^{22}$ possible configurations during folding the system would spend $5.5{\times}10^{-20}$ seconds in each configuration, which is absurdly small - two orders of magnitude smaller than the time required for a photon to cross a distance equal to the width of a double helix and therefore too short for a configuration change.

Alternatively this calculation can be performed in reverse. The time required for a change of state can be regarded as the time needed for one domain of a fully bound staple to unbind and rebind to the alternative scaffold site. This must be equal to, or longer than, the time taken for the free domain of the half bound staple to diffuse across to the new binding site. The diffusion coefficient of a half bound staple will be greater than that of a free staple, so calculating the time taken for a free staple to diffuse an equivalent distance will provide a lower limit for the time per state.

If a staple has to diffuse a distance $d$ for a change of state to occur, the time per state is approximately $t \sim d^2/D$, where D is the diffusion constant [15]. $d$ may be calculated using the freely-jointed chain model [151], according to which the root-mean-square separation of the ends of the chain is $\sqrt{N}b$, where $N$ is the number of segments in the chain and $b$ is the Kuhn length. For single-stranded DNA, the Kuhn length, $b$, is around 1.5 nm and each segment is 3 bases long [20], so if $d$ is the separation of two identical sites on opposite sides of the scaffold (i.e. sites $n$ =2646 bases apart) when no other staples are bound, $d = b\sqrt{n/3} = 45\,\text{nm}$. The staple may be approximated as a spherical object, and if it also behaves like a freely jointed chain

it has a hydrodynamic radius of $b\sqrt{(3\pi N/128)}$ [151, 152]. According to the Einstein relation, $D = k_B T/\gamma$, where $\gamma$ is the drag coefficient. For a spherical object moving through a fluid at low Reynolds number, Stokes' Law gives $\gamma = 6\pi\eta r$, where $r$ is the hydrodynamic radius of the object and $\eta$ is the viscosity of the medium [151].

Hence:

$$t \approx d^2 6\pi\eta(3\pi N/128)^{1/2} b/k_B T, \tag{4.1}$$

where $N$ is 32/3, $\eta$ is $0.5 \times 10^{-3}$ Pa s (viscosity of water at 60°C [153]) and the temperature is taken to be 60°C, the average temperature during the annealing process and the approximate staple incorporation temperature. The time to diffuse between states is therefore around 5.5 $\mu$s, which indicates that if a polymorphic tile had to sample every possible state its total folding time would exceed 13 billion years[3]. Thus, if the polymorphic tiles self-assembled randomly it would therefore be extremely remarkable to see any well-folded tiles at all.

It is evident that, like proteins, DNA origami does not fold via a random search, and that there are *folding pathways*, the existence of which is attributable to cooperation between staples.

[3]This calculation can also be performed using the time per state in the simulations performed by my colleague (Frits Dannenberg), and the conclusion is essentially the same.

## 4.4 Mechanism for cooperation between staples

### 4.4.1 Shortest-distance principle

It will be seen from Fig. 4.8 that staples in an origami tile connect domains which may be quite widely separated along the scaffold. If one staple binds, it will reduce the average separation of the binding sites of a second staple, as illustrated. The free domain of a half-bound staple will hybridize with its complement on the scaffold if the two are brought sufficiently close together by the random fluctuations of thermal motion, and this is more likely to occur if the sites are separated by a shorter length of DNA.[4] Thus, each staple which binds will help other staples to bind, and origami folding is highly cooperative.

The reasoning above suggests that staples will bind faster when their binding sites are close together on the scaffold - I call this the 'shortest distance principle'.

#### Bending DNA

When a staple binds, the scaffold must bend. Single-stranded DNA is extremely flexible, and thus the energy penalty for this will normally be negligible. However, it may not be quite so small for staples binding together two scaffold domains which are immediately adjacent to each other (e.g. the staple connecting the red domains in Fig. 4.8), and in this case, the rate of binding may be slower than expected. The

[4]This is the principle used by the author's colleague Frits Dannenberg to model origami folding computationally; the results of these simulations are generally in good agreement with the polymorphic tile experiments described in Chapter 6, but further discussion of the modelling is outside the scope of this thesis.

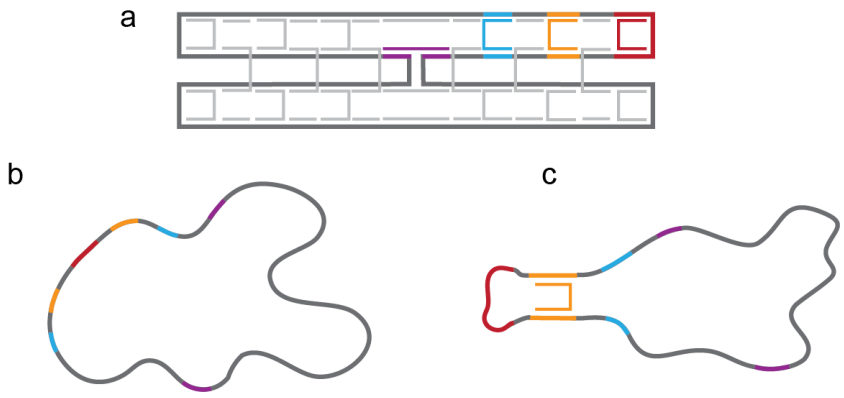


Figure 4.8: The mechanism by which staples cooperate during folding. (a) Sketch of a mini origami tile. Particular staples and their binding sites are highlighted in different colours. One of the seam staples is indicated in purple. (b) Depiction of the free scaffold. (c) A staple binding to one pair of sites (orange) brings the binding sites (blue) for another staple into close proximity by changing the arrangement of the scaffold (*cf* configuration shown in (b)).

| | |
|---|---|
| Persistence length of ssDNA | $\sim 1\,\mathrm{nm}$ [20, 154] |
| Persistence length of dsDNA | $\sim 50\,\mathrm{nm}$ ($\sim$ 150bp) [19] |
| Length of one ds staple domain | $5.44\,\mathrm{nm}$ (16bp) |

Table 4.1: Length scales.

relevant length scales are given in Table 4.1; bending of the single-stranded scaffold should not be neglected for loops shorter than the persistence length of ssDNA.

### 4.4.2 Extreme cooperativity: seam staples

As discussed above, when one staple binds it brings into close proximity the two binding sites for another. Generally there will still be some length of scaffold separating the latter pair of domains but in the original design the seam staples are paired (as in the mini-tile shown in Fig. 4.8), and when one has bound, its partner's binding sites are very close together. Thus the binding of each seam staple is greatly facilitated by the binding of its partner, and each member of the pair is expected to reinforce the other. All other staples must cooperate significantly more weakly.

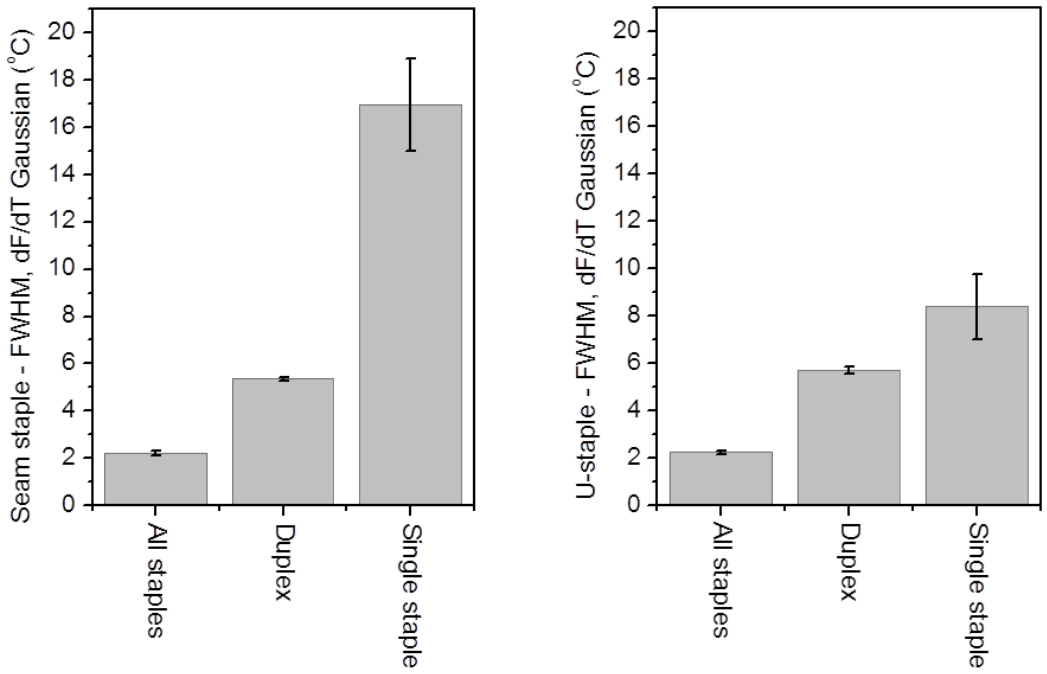


Figure 4.9: Width of transition corresponding to staple incorporation: evidence for cooperativity. Graphs show the mean FWHM of Gaussian fits to dF/dT data for the specified samples. 'All staples': Incorporation of both staples into *monomer* origami tiles (behaviour of both staples measured simultaneously). Duplex: dual-labelled staple binding to reverse complement (behaviour of both staples measured simultaneously). Single staple: single dual-labelled staple binding unassisted to *monomer* scaffold (behaviour of staples measured in different samples, each containing only the labelled staple and the scaffold). Error bars indicate the standard error on the mean. Left: seam staple, Right: U-staple (as indicated by the labels on the y axes). Labelled staples - as for 'Case 1' (see Chapter 3).

### 4.4.3 Evidence of global cooperativity in origami assembly

The width (temperature range) of a transition indicates the speed at which it occurs - the narrower a transition, the faster it is. Thus, sharper transitions are more cooperative [14, 75] - if a hybridization event happens more rapidly when helper strands are present than when they are absent, this is because the strands cooperate. Fig. 4.9 shows the measured full-width-at-half-maximum (FWHM) for three different instances of each of two staples hybridizing. Hybridization to form a duplex is non-cooperative, because only the two hybridizing strands are involved - this is the reference point for

other processes. Incorporation of staples into the origami tile is the most cooperative process - each staple is assisted by many others (data is shown for the monomer - results are very similar for the dimer). The cooperative nature of origami folding has been noted in previous studies [75], based on the temperature range over which the transition occurred, which was much narrower than would be expected if the staples did not bind cooperatively. This temperature range is the 'assembly window' [67].

For a staple which binds unassisted to the monomer scaffold, the width of the transition is increased because here the staple must bridge a long loop of scaffold. The longer the loop, the more difficult the process and the larger the width. Hence the measured width of the transition is greatest for the seam staple binding to the scaffold in the absence of other staples.

### 4.4.4 Single staple binding unassisted to scaffold

When other staples are absent the 'seam' staple and the 'U'-staple both bind to the monomer scaffold at approximately 57°C on average. Representative data is shown in Fig. 4.10, from which it will be seen that the transition is very broad and consequently data is comparatively noisy; as normal, the estimates of staple hybridization temperatures were computed by averaging over repeated experiments. In both cases, the result represents a significant downwards shift in comparison with the equivalent duplex: 8°C for the U- staple and 21°C for the seam staple. The size of the difference is related to the length of scaffold separating the two domains; when

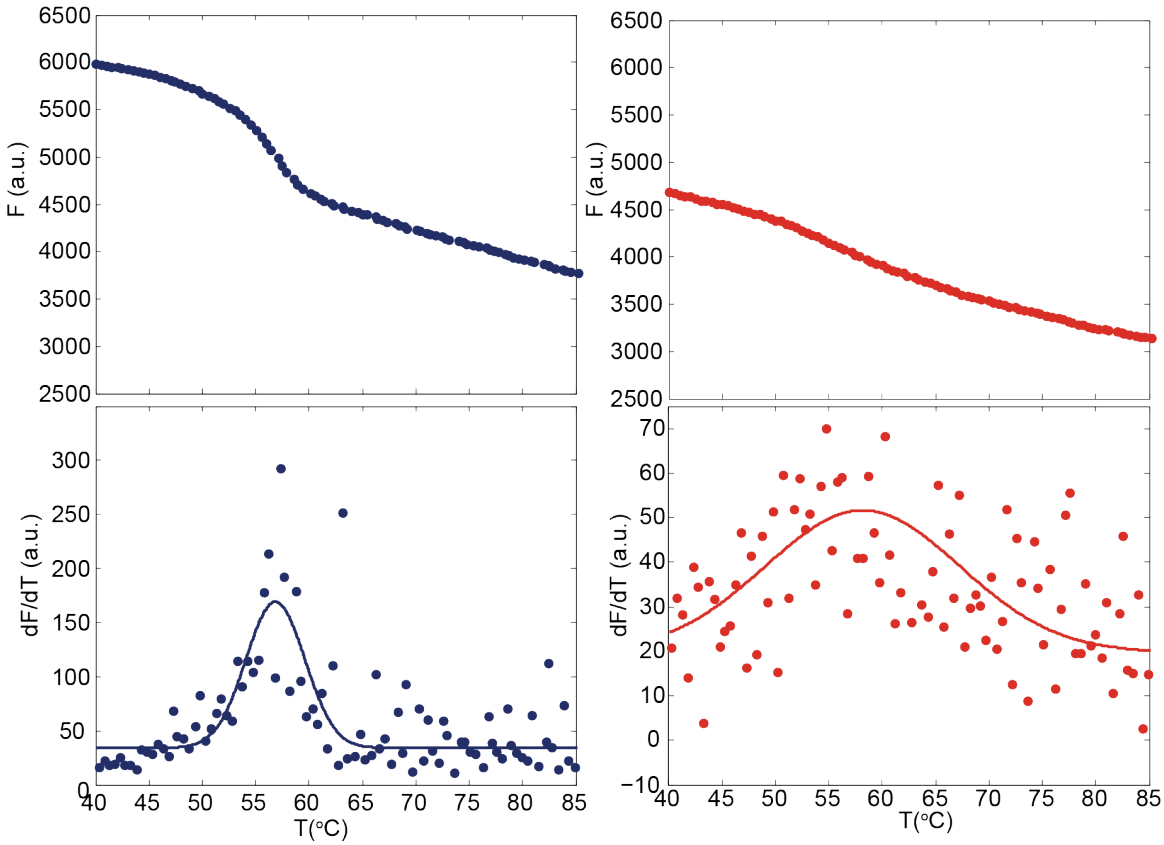


Figure 4.10: Single staples binding unassisted to the monomer scaffold (two separate samples). Upper plots: Fluorescence as a function of temperature. Lower plots: Points - derivative, lines - Gaussian fit. Graphs on the left (blue datapoints): U-staple. On the right (red datapoints): Seam staple. Note: lower pair of graphs do not have the same y-scale. Labelled staples - as for 'Case 1' (see Chapter 3).

a staple binds unassisted to the scaffold, it pinches off a loop of single-stranded DNA. In the case of the seam staple this loop is 82 domains long, but for the U-staple it is much shorter - six domains. The entropic cost of closing a loop increases with the loop length, and thus the temperature difference also increases, as does the width of the transition. This illustrates the shortest distance principle - if the separation of binding sites is increased, it is harder for the staple to bind and it binds at a lower temperature.

## 4.5 Summary

I have described in this Chapter how I synthesized the dimer scaffold and confirmed using agarose gel electrophoresis that it had formed correctly. I used the dimer scaffold to make polymorphic origami tiles, and AFM images I acquired of a range of variants of the polymorphic tiles revealed that a significant proportion of them folded well. This demonstrated the existence of folding pathways, arising due to cooperation between staples. I discussed how staples can cooperate, explaining that when one staple binds it brings into close proximity the binding sites for another, helping it to bind. The distance between the binding sites of a staple is thought to determine the speed with which it binds and the seam staples are expected to cooperate particularly strongly because they are paired.

# Chapter 5

# Quantitative characterization of polymorphic tiles

The results presented in the previous Chapter demonstrate the existence of folding pathways for the polymorphic tile, and the distribution of shapes observed for the different design variants reflects the pathways followed in each case. To shed light on the assembly processes, it was necessary to study the configurations in detail. Here, I discuss the possible well-folded shapes and explain how tiles can be classified into groups on the basis of detailed quantitative characterization.

In order to extract the required information from AFM images of the polymorphic origami tiles, I developed a custom image processing strategy which I will describe and explain fully here. Traditionally, origami objects visualized by AFM are only classified as either well-folded or mis-folded structures, and this is usually done by visual inspection [5]. A more sophisticated approach was required for analysis of

the polymorphic tile, and I therefore developed my own MATLAB procedures to extract tile parameters. My aims were to minimize subjectivity as far as possible and maximize the resolution of the measurement.

I will explain in this Chapter how my method works and show its successful application to experimental data. I will demonstrate using simulated datasets that I can achieve accurate sub-pixel resolution measurements of tile parameters.

## 5.1 Configurations of well-folded tiles

By definition, a well-folded origami dimer tile is entirely composed of double-stranded DNA, and is not associated with any loose scaffold. It does not have any visible holes or damage, it is not bent, and its outline may be fitted using the procedure described later in this Chapter. The comprehensive set of scaffold routings which give rise to stress-free folded objects is given in Appendix B, and representative examples are illustrated in Fig. 5.1.

All of the routings illustrated are planar. It may be deduced *a priori* that polymorphic tiles should not form double-layer structures as follows. During folding, branch migration will tend to maximize the length of double-helical domains, and consequently it is unlikely that any single staple domain will be bound to part of one copy of its scaffold domain and part of the other. Hence, the scaffold will cross from one rectangle to the other at the end of a staple domain. However, at the end of most of the staples the phase of the helix is such that the scaffold backbone lies in the plane

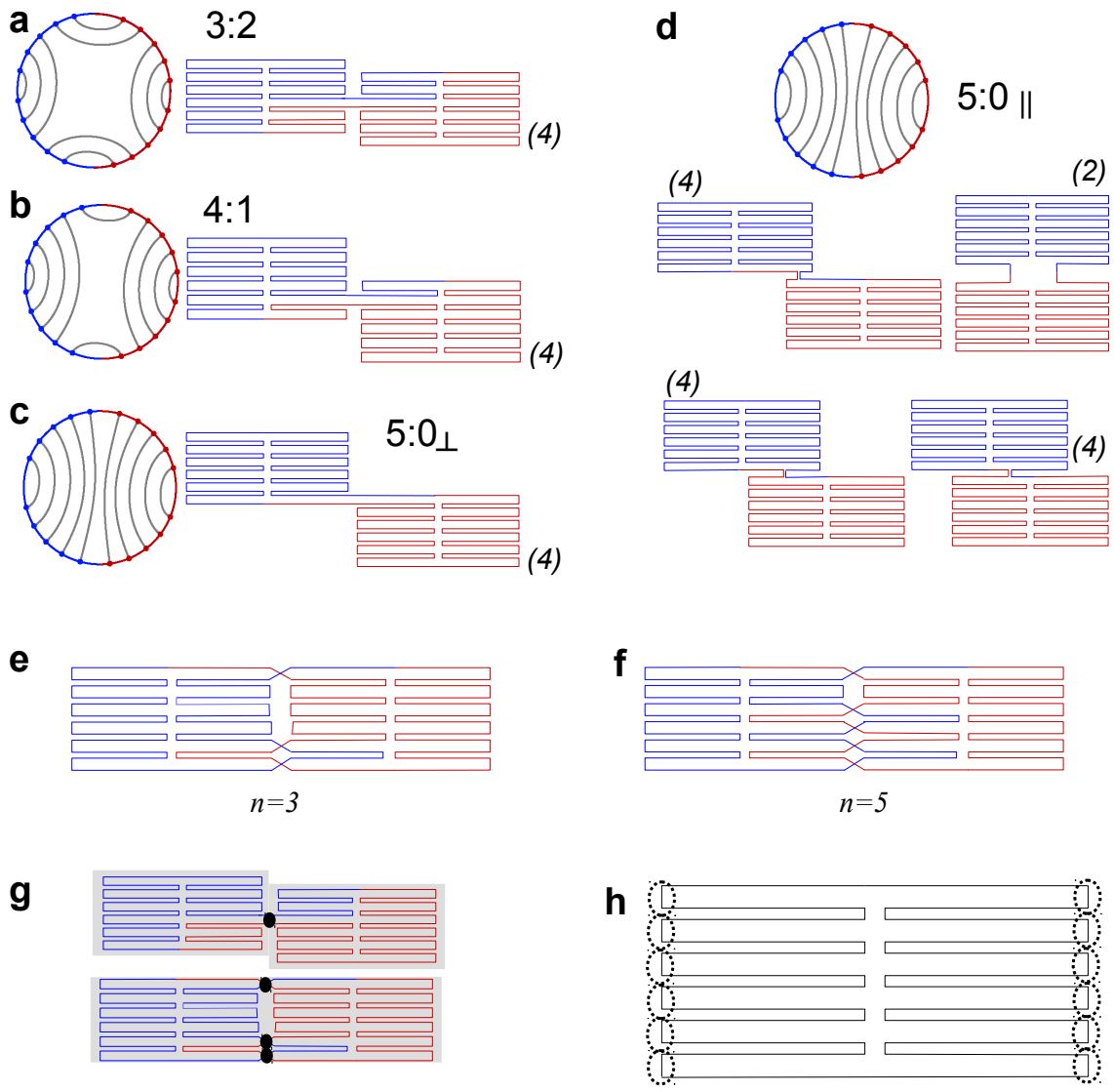


Figure 5.1: Representative well-folded configurations of polymorphic dimer tile - possible scaffold routings. Circle diagrams show connectivity of seam staple pairs - each arc represents a pair of seam staples, and the dot on the circumference represent the scaffold segments they connect. Tile configurations are labelled using the shorthand $x$:$y$, as described in the text: (a) Type 3:2, (b) Type 4:1, (c) Type $5{:}0_{\perp}$, (d) Type $5{:}0_{\parallel}$. The symbols $\perp$ and $\parallel$ indicate whether the two rectangular halves of the dimer are joined on the edge perpendicular or parallel to the helical axes, respectively. Numbers in brackets indicate the degeneracy i.e. the number of permutations of the routing which give rise to the same configuration. (e) and (f) show example scaffold routings for End-to-End (E-E) type tiles, for which $n$=3 or $n = 5$. (g) Scaffold routings depicted in (a) and (e), with 'jump points' marked by black ellipses to illustrate definition used in the text. (h) Illustration of all possible jump point locations on the short edges of the rectangle. The jump points on the long edges are not marked.

of the tile, pointing towards the next helix in the structure. Thus it is impossible for the other rectangle to form at this point, due to steric hindrance. At the edges of the rectangles there is no such restriction, and hence the other rectangle can form only at the edges of the other, where the phase of the helix is such that the two rectangles will lie in the same plane. The AFM images confirm that in general double-height tiles do not form.

A well-folded tile therefore consists of two rectangles lying side by side, joined on one edge. The number of times the scaffold crosses the join between them is $N$, and this must be even because the scaffold is a closed loop - for every occasion on which the scaffold leaves one rectangle it must return. Hence $N = 2n$, where $n$ is an integer, and it is only possible to identify continuous routings for odd values of $n$. $n$ is the number of points ('jump points') on the edge of a rectangle at which one section of the scaffold leaves and another returns (see Fig. 5.1g for definition of 'jump point'). There are 6 pairs of helices in the tile, and this means that there are up to 6 jump points per short edge (Fig. 5.1h). Hence, a given scaffold routing (for a tile consisting of two rectangles joined on the short edge) may be characterized by 1 (Fig. 5.1a-d), 3 (Fig. 5.1e) or 5 (Fig. 5.1f) jump points, which may occur at any of $p = 6$ positions. The number of such scaffold routings is given by the number of possible arrangements of jump points. The number of ways to pick $n$ jump points from $p = 6$ possibilities is:

$$N(n, p) = \frac{p!}{n!(p-n)!} \tag{5.1}$$

Hence, for n=1 there are 6 alternatives for each side of the tile, giving rise to a total of 12 scaffold routings, three of which are shown in Fig. 5.1a-c. For n=3, there are 40 routings in total (taking both sides of the tile into account), and for n=5 there are another 12 configurations. In all, there are 64 possible scaffold routings corresponding to pairs of rectangles joined on the short edge. Inspection of the design reveals that there are 7 possible jump points on each long edge of the tile and hence a total of 14 scaffold routings for tiles corresponding to pairs of rectangles on the long edge (examples are shown in Fig. 5.1d). Thus a total of 78 routings are possible[1].

The short-edge tiles with $n$=1 and the long-edge tiles can be grouped into categories based on the connectivity of the seam staples, which is represented by the circle diagrams[2] in Fig. 5.1. It is possible to group the edges in the circle diagrams which are centred on the same point, and in each diagram there are two groups of edges with $x$ members and two with $y$ members. The corresponding tile shape is therefore categorized as Type $x : y$. In the scaffold routings, if a horizontal line is drawn through the jump point, in one rectangle there are $x$ pairs of seam staples above the line and $y$ pairs below it. (In the other rectangle, there are $y$ pairs above and $x$ below.)

The circle diagrams may also be interpreted as follows. Considering the scaffold sequence to start at the centre of the tile on the top edge (or bottom edge, by symmetry), $x$ pairs of seam staples connect domains in the first half of the scaffold to other domains in the first half, while $y$ pairs connect domains in opposite halves.

[1] I performed this calculation to ensure that the set of routings I had identified by drawing was complete.

[2] Circle diagrams formatted using a modified version of an Adobe Illustrator script written by Dr Jonathan Bath.

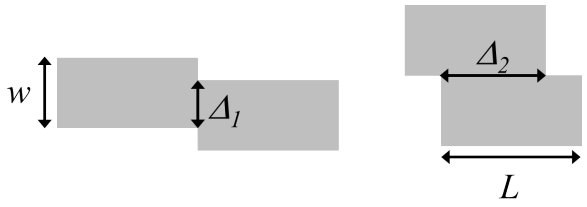


Figure 5.2: Schematic diagrams of the shape of polymorphic tiles, which look like two joined rectangles.

The three main classes of tile are Type 3:2, Type 4:1 and Type 5:0. The final class is split into sub-classes. I refer to the Type 5:0 tiles consisting of two rectangles joined on the short edge as Type $5{:}0_{\perp}$ because the short edge is perpendicular to the helical axis, and the other class is Type $5:0_{\parallel}$, where the two rectangular halves are joined on an edge parallel to the helical axis.

Tiles for which $n = 3, 5$ will all have the same shape, and I describe these as End-to-End (E-E), in which the edges of the two rectangles are flush. E-E tiles can also arise if the two halves of a Type 3:2 tile slide relative to each other. All well-folded tiles will henceforth be referred to using the E-E, 3:2, 4:1 or 5:0 labelling scheme.

## 5.2 Describing tile configurations quantitatively

The general shapes of the polymorphic tile are illustrated in Fig. 5.2, which shows that these shapes can be classified according to the value of $f_{\perp} = \frac{\Delta_1}{w}$ or $f_{\parallel} = \frac{\Delta_2}{L}$. These quantities may be expressed in terms of the width of a helix ($h$), the inter-helix gap ($g$), and the domain length ($l_d$). The tile width, $w$, is $12h + 11g$, while its length, $L$, is $14l_d$. For tile Types 3:2, 4:1 and $5{:}0_{\perp}$, values of $f_{\perp}$ are given in Table 5.1.

| Tile type | $\Delta_1$ | $f_\perp$ |
|---|---|---|
| 3:2 | $(10h + 9g)/(12h + 11g)$ | 0.83 |
| 4:1 | $(6h + 5g)/(12h + 11g)$ | 0.48 |
| 5:0$\perp$ | $(2h + g)/(12h + 11g)$ | 0.14 |

Table 5.1: Quantitative classification of type 3:2, 4:1 and 5:0$\perp$ tiles. $h$ is assumed to be 2 nm and $g$ is taken to be 1 nm. The value of $f_\perp$ is not very sensitive to the value of $g$.

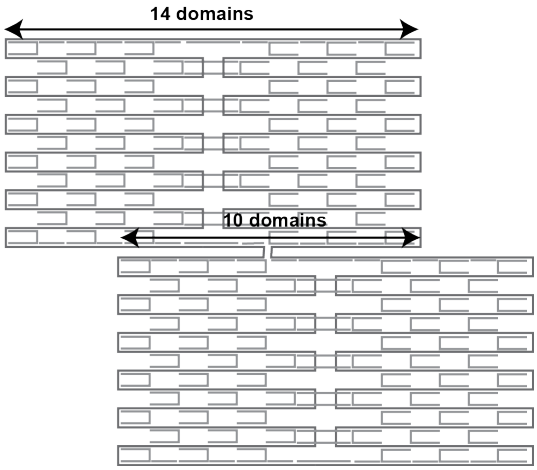


Figure 5.3: Example configuration of Type 5:0 $\parallel$ tile.

An example of a type 5:0$_\parallel$ tile is shown in Fig. 5.3. Here, the offset $\Delta_2$ is 10 domains long, and hence $f_\parallel = \Delta_2/L = 10/14 = 0.71$. The configuration shown can also exist in a form in which $f_\parallel = 1$, where double-stranded linkers extend between the tiles, as shown in Appendix B. The average value of $f_\parallel$ for these two possibilities is 0.86. For the other two 5:0$\parallel$ shapes, the possible values are 0.14 and 0.43. For classification of tiles observed experimentally, the half-way point between each pair of values forms the dividing line between tiles of different Types.

### 5.2.1 Degenerate shapes

For each value of $f_\parallel$ or $f_\perp$, there is more than one corresponding scaffold routing. For a given $f$ value, the routings are related by rotation of the circle diagrams describing

the seam staple configuration, and the full set of routings is given in Appendix B. For each shape there are generally 4 degenerate routings, giving rise to two shapes of each chirality. These are indistinguishable in the AFM images presented in this thesis because it is not possible to ascertain which way up the tile has landed on the mica.

## 5.3 Image Processing

To classify the observed tile shapes into Types, it is necessary to measure $f_{\perp}$ and $f_{\parallel}$ for each individual well-formed tile observed in AFM images. The image processing strategy I developed for characterizing the tiles quantitatively is summarized in Fig. 5.4, and I will discuss it in detail here. Before doing so, I will explain some basic concepts of image processing[3].

Every image consists of an array of *pixels*, each of which has a *brightness* or *intensity*; for an AFM image the pixel brightness is the measured height. As shown in Fig. 5.5, each pixel is surrounded by eight others, collectively referred to as the *neighbourhood* of the original pixel. It is frequently useful to plot a histogram of pixel brightness, and for an image in which each pixel is part of either the background or a bright feature, the histogram will have two peaks.

If the image is converted into a form in which all pixels have intensity values of either 0 or 1 it is said to be *binarized* or *thresholded*. Any pixels in the original image which are brighter than a certain *threshold* take a value of 1 in the final im-

[3]General image processing references: [155, 156] (textbooks).

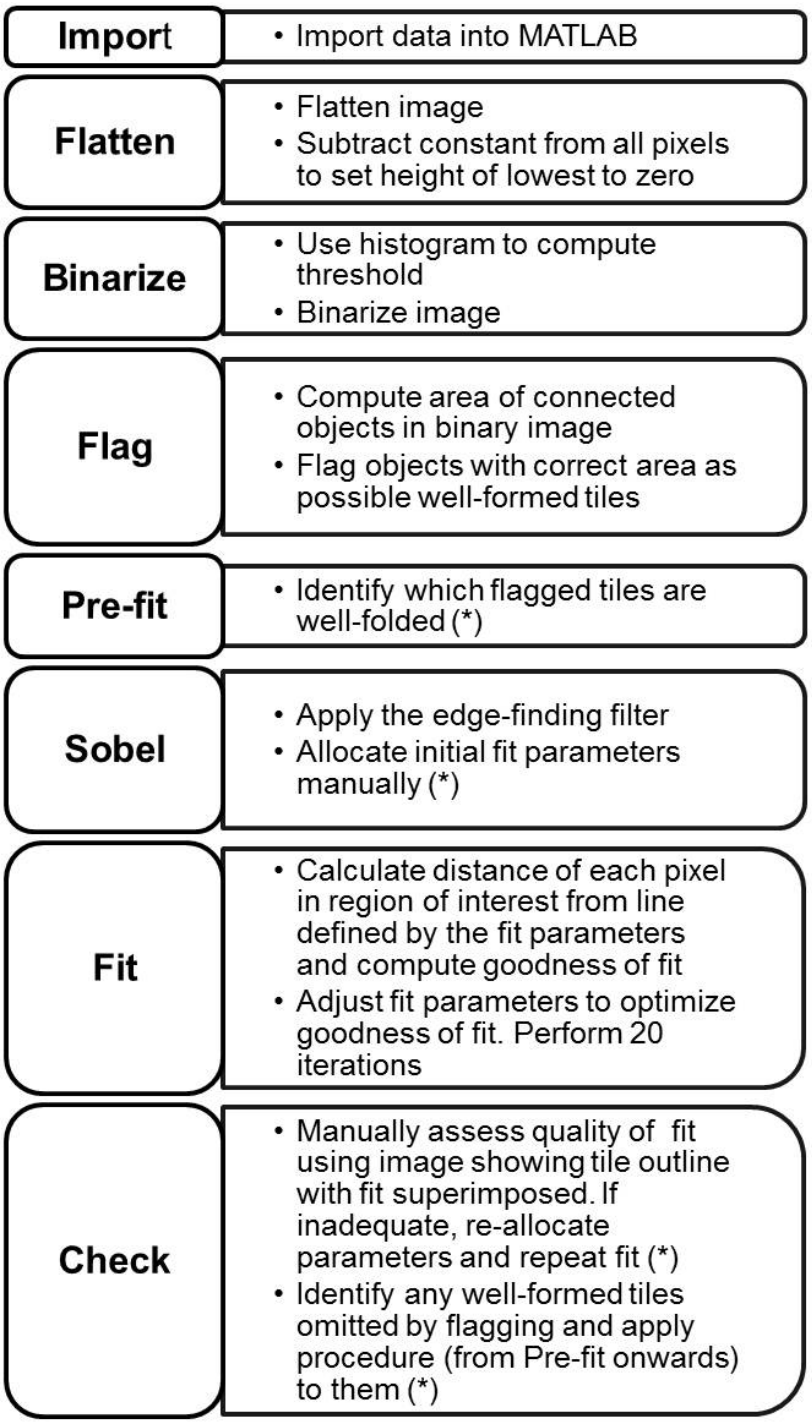


Figure 5.4: Summary of image processing procedure. All steps are automated except those marked with a *.

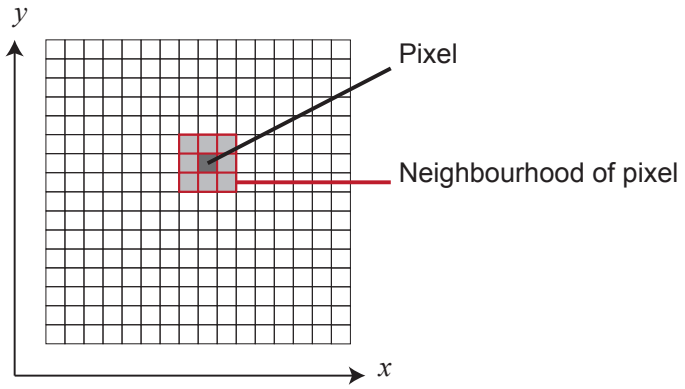


Figure 5.5: Definitions: a pixel and its neighbourhood. An image is an array of pixels, and each pixel has a neighbourhood consisting of the eight pixels which surround it.

age, whereas those which are less bright take a value of zero. The threshold may be estimated manually or computed automatically based on the distribution in the brightness histogram. The approach adopted will depend on the image to be analysed, and selected approaches will be discussed later in this Chapter.

Two neighbouring pixels in a binary image are *connected* if they both have a value of 1. A pixel (value 1) is four-connected if the four pixels which share its edges have value 1. If the pixel (value 1) is 8-connected, all pixels with which it shares an edge or a vertex also have value 1.

Operations such as edge-finding, differentiating or line-finding are applied to an image by representing them with a *kernel*, a small matrix (usually 3×3 elements) which is then *convolved* with each pixel and its neighbourhood. The convolution of a kernel $k_{ij}$ with a neighbourhood $n_{ij}$ is given by:

$$\sum_{ij} k_{ij} n_{ij} = k_{11}n_{11} + k_{12}n_{12} + k_{13}n_{13} + k_{21}n_{21} + \ldots \tag{5.2}$$

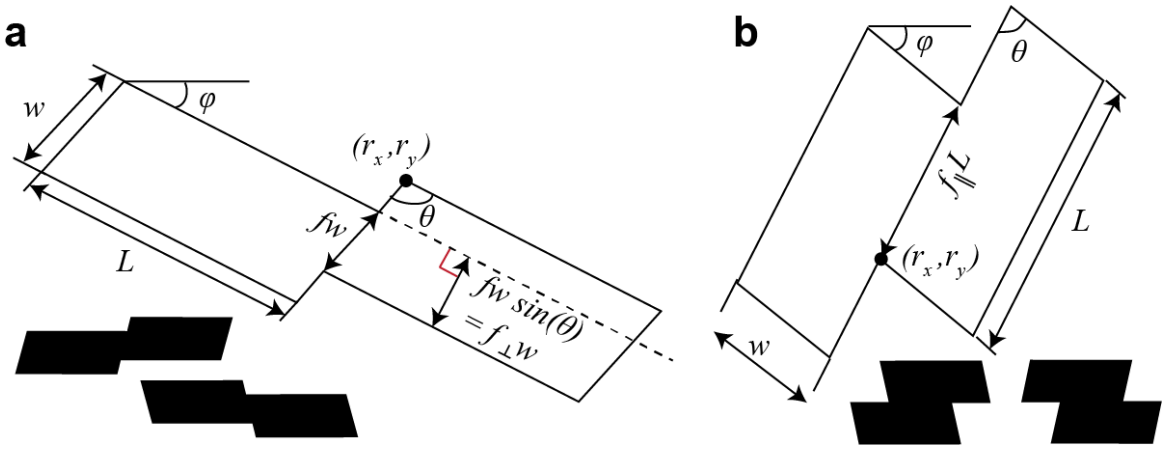


Figure 5.6: Ideal outlines of dimer tiles. (a) Shape of tiles consisting of rectangles connected on the short edge, corresponding to Types 3:2, 4:1 and $5{:}0_{\perp}$ (b) Shape of type $5{:}0_{\parallel}$ tiles, in which rectangles are joined on the long edge. Parameters of fit are shown: $L, w, \theta, \phi, f, r_x, r_y$. For perfect rectangles $\theta = 90°$, but tiles are often distorted. Block diagrams show shapes of different chirality, which is ascertained manually.

## 5.4 Processing AFM images of dimer tiles

Ideal dimer tile outlines are depicted in Fig. 5.6. The objective for the image processing routine is to extract accurate measurements of the indicated parameters, and this is achieved here by fitting the idealized outline to the observed tile shapes. The strategy adopted has three stages: (1) pre-processing, (2) identifying tiles to fit and (3) performing the fit.

### 5.4.1 Pre-processing

Each image was prepared for analysis by *flattening* - subtracting a second order polynomial from each line in the image, to remove tilt and bow [157], which arise due to the motion of the scanner head (Fig. 5.7). In addition, a constant was subtracted from all pixels in the image such that the lowest point was defined to be at zero height.

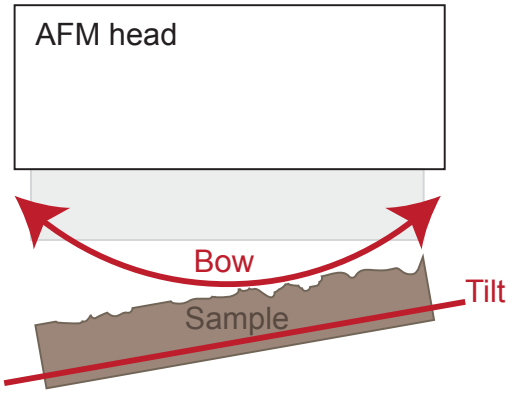


Figure 5.7: How bow and tilt arise in AFM images. The AFM head is the part of the microscope which scans the sample. Bow is introduced by the motion of the AFM head, and 'tilt' occurs when the sample is inclined with respect to the microscope.

### 5.4.2 Identifying tiles to be fitted

Before proceeding to fit tile outlines, it was necessary to make a preliminary classification of objects in the field of view as candidate well-formed tiles or unknown structures (fragments, misfolds etc). For this step, I converted the image into binary format, the threshold usually being calculated from the histogram of the pixel heights as the average of the means of the two peaks corresponding to background and tiles (Fig. 5.8a). For a distribution with two clear peaks, the threshold could have been chosen to be the lowest point between the two peaks, but this would be problematic here because the tiles are often sparsely distributed across the mica and one of the peaks is therefore very small and poorly-defined. In some cases the automatic thresholding failed because the image was particularly noisy, and in these cases I set the threshold manually (assessing data by eye). Here, I defined the threshold to be $h_b + \delta$, where $h_b$ is the height of the background peak and $\delta$ was typically around 0.3 nm.

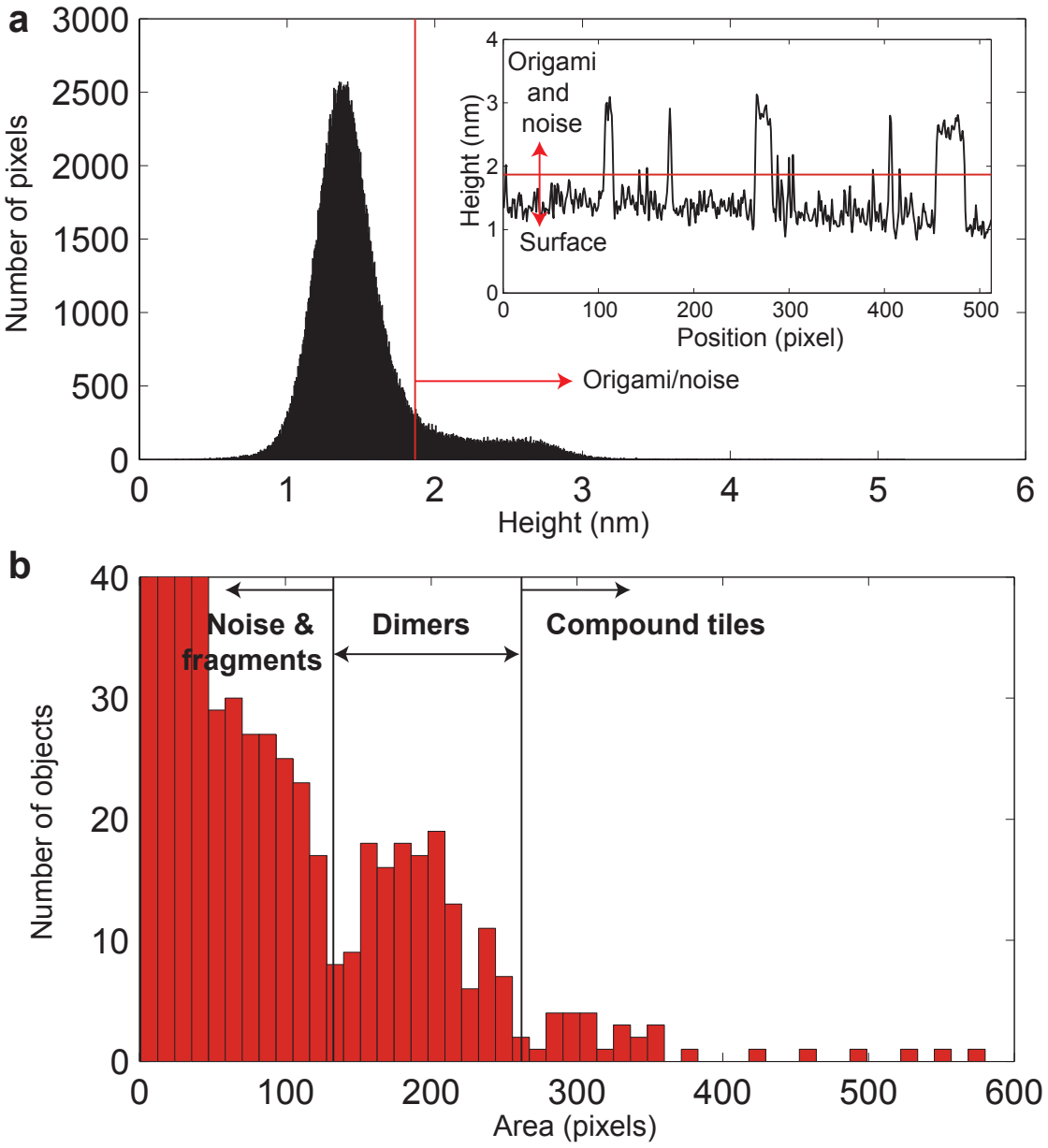


Figure 5.8: Identifying tiles to be fitted: binarizing image and measuring area of objects in field of view. (a) Distribution of pixel heights in a representative 3 $\mu$m×3 $\mu$m (512×512 pixels) AFM image of the original design of the polymorphic dimer tile. Inset shows the height profile along a single line in the image. The origami tiles are clearly distinguishable. The threshold is marked in red in both the histogram and the line profile. (b) Histogram of areas of connected objects observed in a set of three binarized 3 $\mu$m×3 $\mu$m (512×512 pixels) AFM images of the polymorphic (dimer) tile. The peak in the vicinity of 200 pixels represents well-formed dimers. The top of the low-area noise peak is cut off by the scale of the graph.

Other thresholding methods involve the use of statistical tests to divide the pixels into two categories[155], but these are more complicated. The approach adopted is simple and reasonably reliable.

My program then computed the area of each connected[4] object in the binary image. For a set of three images of dimer tiles, each image covering an area of $3 \times 3\,\mu\mathrm{m}^2$, this yielded the histogram of object areas shown in Fig. 5.8b. The peak corresponding to well-formed dimers is clearly visible. The large number of objects with an area of only a few pixels are noise spikes, whereas the few tiles with a comparatively large area are actually multiple tiles which are in contact and therefore indistinguishable.

The area of each object formed the basis for a preliminary classification. Inspection of the histogram for the original design (Fig. 5.8b) revealed well-defined minima on either side of the expected dimer tile area, and the positions of these minima were used to define maximum and minimum area limits. If the area of an object lay between these limits, the program flagged it as a possible well-folded tile. Flagged tiles were fitted as described below if no loose scaffold was visible, and it was not obviously damaged or bent. Occasionally well-formed tiles were omitted from the preliminary classification, and I identified these by visual inspection.

In the case of images displaying evidence of piezo creep[5], the affected area was masked out (for example, by ignoring the top 100 pixels in a 512×512 grid).

[4] 8-connected.

[5] The AFM scanner motion is controlled by piezoelectric elements (piezos) and when the scanner is instructed to move to a new area, the displacement is not linear [157], resulting in distortion at the start of the new image. The distortion disappears as imaging continues; piezo creep can therefore usually be eliminated by re-scanning the same area, but in the case of DNA origami this tends to damage the structures and is therefore undesirable.

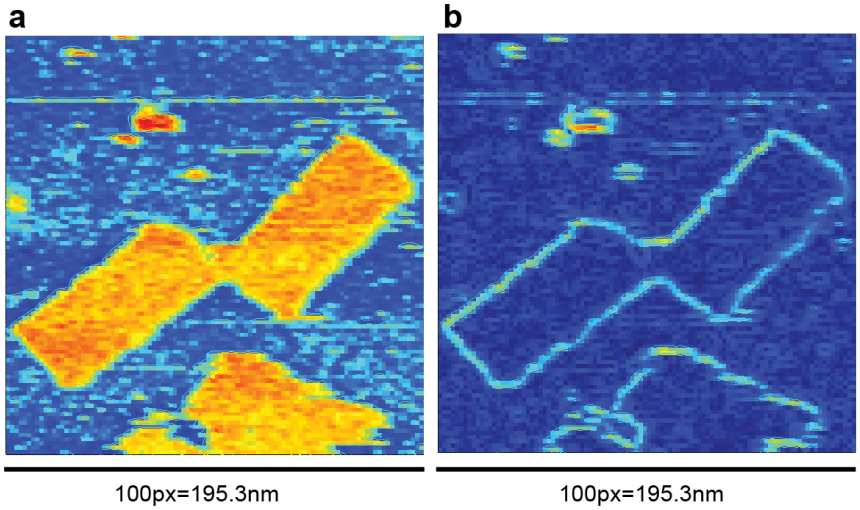


Figure 5.9: Application of the Sobel filter. (a) Tile in flattened AFM image prior to application of the filter. (b) Tile outline obtained by applying the filter. This AFM image was acquired at higher magnification than the images used for fitting.

### 5.4.3 Fitting

The observed tile outlines were extracted from the pre-processed AFM images by application of a Sobel (edge-finding [155]) filter (Fig. 5.9). This computes the horizontal and vertical derivatives ($S_x$ and $S_y$ respectively) at a particular pixel of the image by convolving the kernel representing each differential operation with the matrix representing the pixel and its neighbourhood, and then it calculates the magnitude of the vector ($S_x$,$S_y$), i.e. $\sqrt{S_x^2 + S_y^2}$. It will be seen from Fig. 5.9 that application of the Sobel Filter produces an image showing well-defined outlines of tiles.

For each tile, the fitting program adjusted the parameters describing the ideal outline until it matched the real tile shape as closely as possible. This results in more accurate data than would be obtained from simple point-to-point measurements between the corners of the tile, and in fact enables sub-pixel resolution. In this respect the technique is inspired by single-molecule localization techniques in which

a point-spread-function is fitted to the intensity pattern of a diffraction-limited spot in order to pinpoint the molecule of interest, which is located at the peak of the PSF [158, 159, 160, 161]. In the same way, by assuming that the edge of the origami is a thin straight line, it can be localized with very high precision, as confirmed later in this Chapter by validation with simulated data.

The initial values of some of the fit parameters were allocated by hand. The parameter space is extremely large, and assigning initial estimates by hand helps to ensure that the correct region of this space is searched. Other parameters were given default values, which were modified only if the fit failed. In particular, starting values for $L$ and $w$ were usually 80 nm and 40 nm respectively, and the initial internal angle $\theta$ was always 90°.

The fit proceeds by minimizing a goodness-of-fit measure, which I defined as:

$$g = \sum_{i,j} \frac{m_{ij} S_{ij} d_{ij}}{N}, \tag{5.3}$$

where $i$ and $j$ number the pixels in the region of interest, $S_{ij}$ is the brightness (height) of the $(i,j)^{th}$ pixel, $d_{ij}$ is the shortest distance between the pixel and the tile outline and $m_{ij}$ is a mask factor. The mask is defined as $m_{ij} = \exp\left(-d_{ij}^2/A^2\right)$, where an appropriate value for $A$ was found by trial and error (ultimately, I used a value of 7.2 nm, corresponding to a FWHM of 12 nm). The normalization factor is $N = \sum_{i,j} m_{ij} S_{ij}$. When $g$ is minimized, the outline defined by the fit parameters should pass close to all points on the edge of the tile. The mask reduces the effect of noise;

otherwise spikes in the 'Sobel Image' can interfere with the fitting. It also ensures that the fit is not affected by nearby tiles.

If all fit parameters remained unchanged for one iteration of the fit, the size of the change in the initial values was reduced by a factor of two for the next iteration. The fit was allowed to proceed for 20 iterations.

### 5.4.4 Checking

When the procedure had been applied to a tile, I checked the results by visual inspection. The number of fit parameters was comparatively large and consequently the fit sometimes failed. If this happened, I modified the initial values of the fit parameters and repeated the fit as necessary. Careful scrutiny of individual tiles confirmed that only well-folded tiles had been fitted and none had been omitted.

### 5.4.5 Fit results

Fig. 5.10 shows representative well-fitted tiles, and Fig. 5.11 illustrates the fitting process in operation. The results of the fit for well-folded tiles of the original design are given in Fig. 5.12, which show fit parameters extracted for tiles in a set of three $3\,\mu\text{m}\times3\,\mu\text{m}$ images. The average tile dimensions are very close to the predicted values, but it will be seen that tile dimensions ($L$ and $w$) differ between tiles. This effect is expected and has been observed before [5]. Similarly, the internal angle $\theta$ shows some variation between tiles. The data (Fig. 5.12) also shows that the dimer tiles align

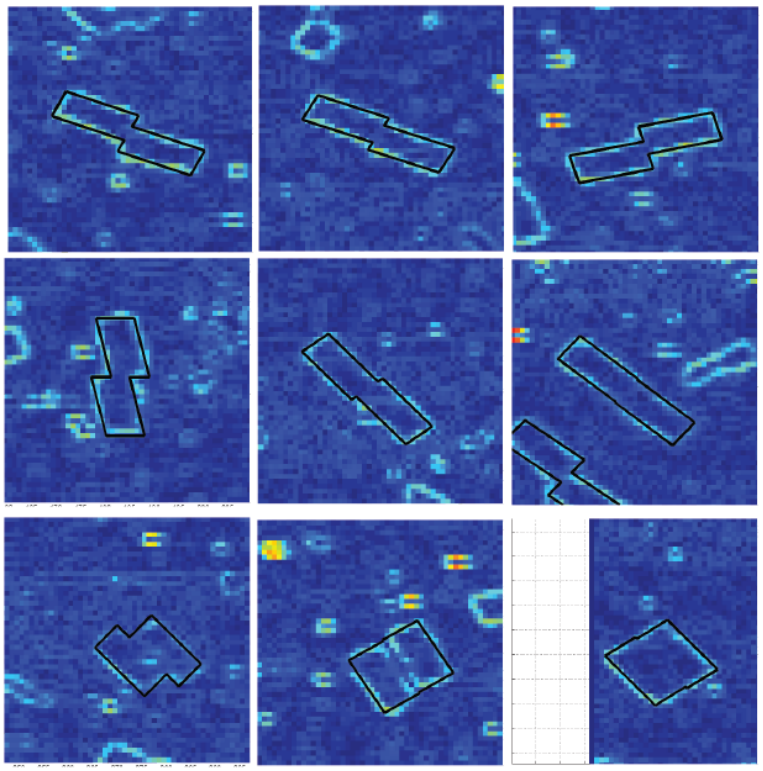

Figure 5.10: Montage: well-fitted tiles. Representative results of fitting process, for polymorphic tile of original design. Images show the results of applying a Sobel Filter to the AFM data; fits are superimposed as black lines. Each square is 50×50 pixels; each pixel measures $\sim 5.9 \times 5.9$ nm.

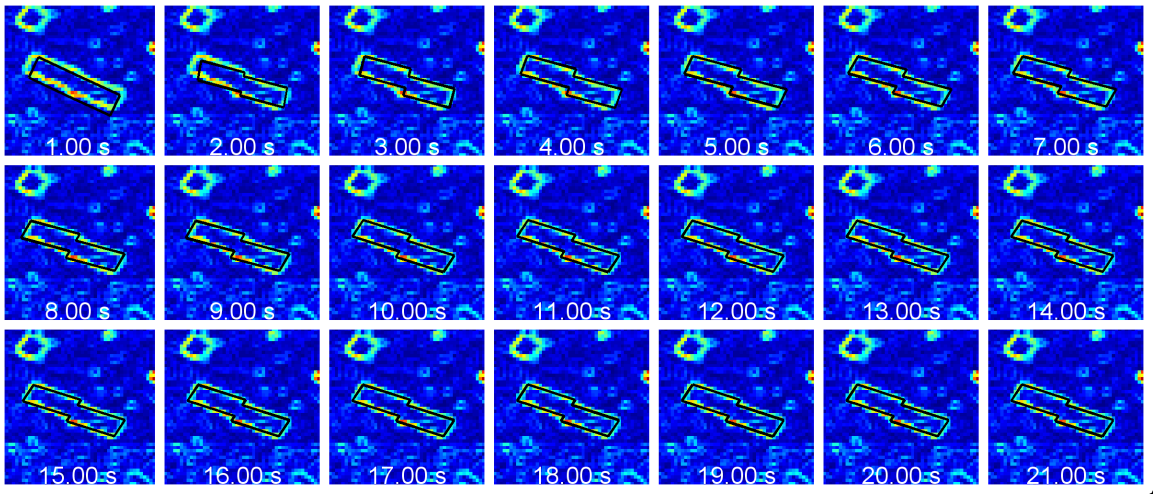


Figure 5.11: The fitting process in operation. For each iteration of the fit loop, the shape described by the fit parameters was plotted and saved as a single frame of a video (usually shown at one frame per second). The frames from a representative video are shown here, illustrating how the fit progresses from the initial parameters to the final result.

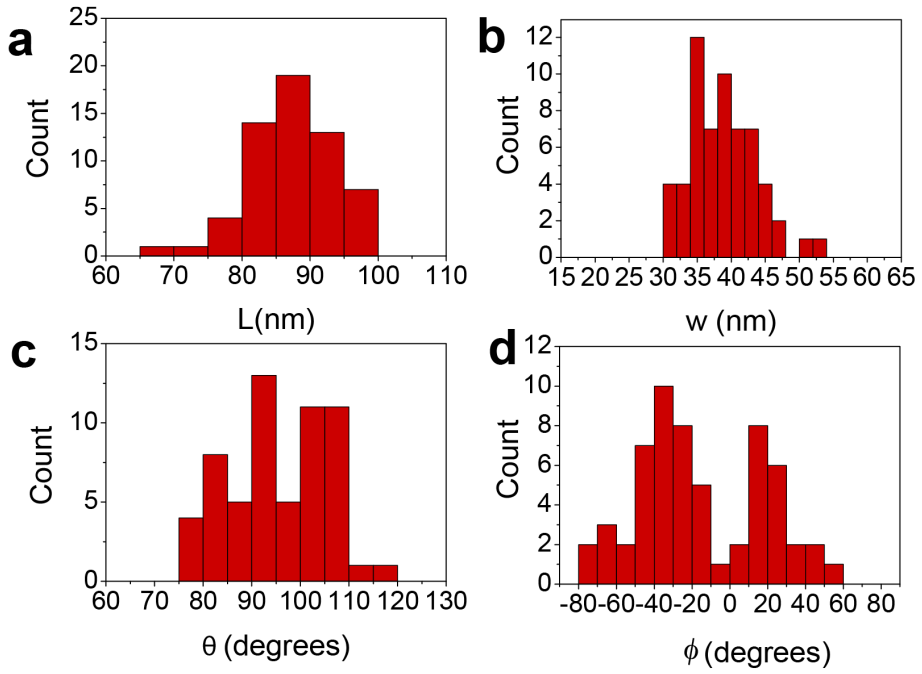


Figure 5.12: Results of fitting ideal outline to polymorphic tiles of original design - $L$ (length), $w$ (width), $\theta$ (internal angle) and $\phi$ (orientation) - for a set of three $3\,\mu$m$\times 3\,\mu$m AFM images. $L$ and $w$ are approximately distributed around the mean. The orientation histogram has two peaks because tiles align preferentially along two axes, as described earlier for the monomer tiles. The distribution of internal angles ($\theta$) is approximately centred on 90°, as expected.

preferentially along certain axes, like the monomer.

Due to the variation between objects, two tiles with the same underlying scaffold routing may have different values for $L$ and $w$. Consequently they should be classified using a dimensionless parameter which describes the overall shape. For tiles of Type $5{:}0_{\|}$, this is $f_{\|}$ (Fig. 5.6b), and for tiles of other types, it is appropriate to use $f_{\perp} = f\sin(\theta)$ (Fig. 5.6a). $f_{\perp}$ is measured along a line perpendicular to the helices in the tiles, and therefore expected $f_{\perp}$ values may be calculated (Section 5.2), whereas the value of $f$ may differ between tiles of the same routing if the two rectangular halves are skewed. Histograms of $f_{\perp}$ and $f_{\|}$ for all the variants of the polymorphic tile will be shown and discussed in the next Chapter.

### 5.4.6 Yield

I demonstrated in the previous Chapter that a significant proportion of the polymorphic tiles folded well. I defined the estimated percentage yield as $100 \times n(\text{fitted})/n(\text{candidates})$, where $n(\text{fitted})$ is the number of successfully fitted tiles, and $n(\text{candidates})$ is the number of candidate well-formed tiles, i.e. the sum of the number flagged automatically based on tile area and the number of tiles omitted by the flagging but identified by visual inspection. Very few tiles are omitted by the automated procedure - typically around 3 of every 60 tiles were left unflagged - and therefore inclusion of these objects does not make a significant difference to the estimate of yield. The calculation does not take account of objects which are the wrong size (tile fragments etc), but the quantity is well-defined and allows comparison of the results from different designs.

### 5.4.7 Validation with simulated data

#### Generating simulated data

To validate my image processing procedure I generated a simulated dataset. Here, all tiles had the same $L$, $w$, $f_{\perp}$, $\theta$ and $\phi$ values and all were of the same chirality, which means that the width of the distribution of fit parameters directly provides a measurement of the intrinsic resolution of my image processing technique. The position of a tile was determined by a random number generator, and twenty-five tiles were included in each image. The tile shapes defined by the specified parameter set were used to create a 'mask' where pixels inside a tile outline were 'on' (taking a value

of 1) and other pixels were 'off' (taking a value of zero). At this stage the image was an 8192×8192 matrix.

A Gaussian blur was applied to simulate tip convolution, and noise was represented by adding to the image an 8192×8192 pixel array of exponentially distributed random numbers, scaled to lie between 0 and 0.7× the tile height. I assumed that tiles were 1.5 nm high. Finally the image was re-sized to the conventional 512×512 pixels. I used the procedure to generate three simulated images.

These images contained a total of 58 tiles which could be fitted - very nearly the same number as the experimental data for the original design of the polymorphic tile. Six other tiles were in contact with others and the remaining 11 tiles were partially outside the field of view. Having generated the data, I fitted it using my standard procedure. Assignment of initial values for the fit parameters was not automated, and as in the case of the experimental data, I estimated the co-ordinates of a corner of the tile corner by hand. The actual value of $f_{\perp}$ was 0.5, but to test the validity of the fitting procedure, the starting point for this parameter was randomly selected to be either 0.45 or 0.55. Similarly the orientation angle $\phi$ was 'guessed' at random to be 35° or 45°. The values of $L$ and $w$ were assumed to be 90 nm and 45 nm respectively, similar but not identical to the true values. An image of representative tiles from the simulated dataset is shown in Fig. 5.13 and application of the Sobel Filter is illustrated in Fig. 5.14.

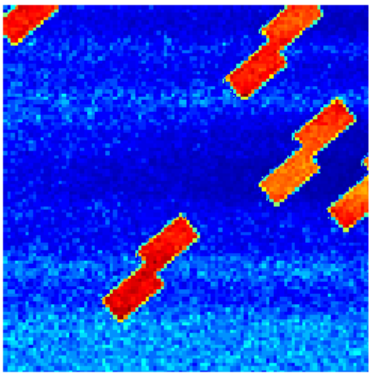

Figure 5.13: Simulated data: close-up view of tiles. Images were generated as described in the text, taking account of tip convolution and noise in the image.

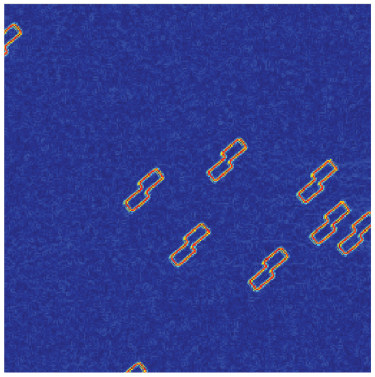

Figure 5.14: Application of Sobel Filter to simulated data - one-quarter of the whole image. Comparison with Fig. 5.9 reveals that the simulated tiles resemble those observed experimentally, as intended.

| | $\theta$ (°) | $\phi$(°) | $L$ (nm) | $w$ (nm) | $f_\perp$ |
|---|---|---|---|---|---|
| **Actual** | 90 | 40 | 93.75 | 46.875 | 0.5 |
| **Measured** | | | | | |
| Mean | 90.08 | 39.74 | 94.08 | 47.00 | 0.492 |
| Range | 10.62 | 3.28 | 3.75 | 2.34 | 0.153 |
| $\sigma$ | 1.99 | 0.68 | 0.71 | 0.48 | 0.030 |
| $3\sigma$ | 5.97 | 2.04 | 2.12 | 1.43 | 0.089 |
| $\Delta$ | 0.09% | 0.64% | 0.35% | 0.26% | 1.65% |

Table 5.2: Simulated data: results of analysis. $\sigma$ is the standard deviation of the distribution of parameters and $\Delta$ is the difference between the average measured value and the actual value. For $L$ and $w$ the standard deviation is less than 1 nm and for all parameters the error ($\Delta$) of the average measurements is below 2%.

**Analysis of simulated data: results**

My analysis of the simulated data confirmed that the precision of my image processing strategy is high. The results shown in Table 5.2 demonstrate that the average measured values of all tile parameters are within 2% of the actual values. The distributions of fitted values are sharply peaked (Fig. 5.15) and the standard deviation for the distribution of measured $L$ and $w$ values is less than 1 nanometre. This is significant because a single pixel is nearly 6 nm wide, and thus these results demonstrate that my technique measures tile parameters with sub-pixel resolution - an advance on the usual method for analysing AFM images of origami structures, which involves manual inspection and counting of objects [8].

Comparison of the analysis of the simulated images (Fig. 5.15) with experimental results (Fig. 5.12) reveals that the distribution of parameter values is considerably broader for the latter. My analysis of experimental data is therefore not limited by the resolution of the image processing procedure.

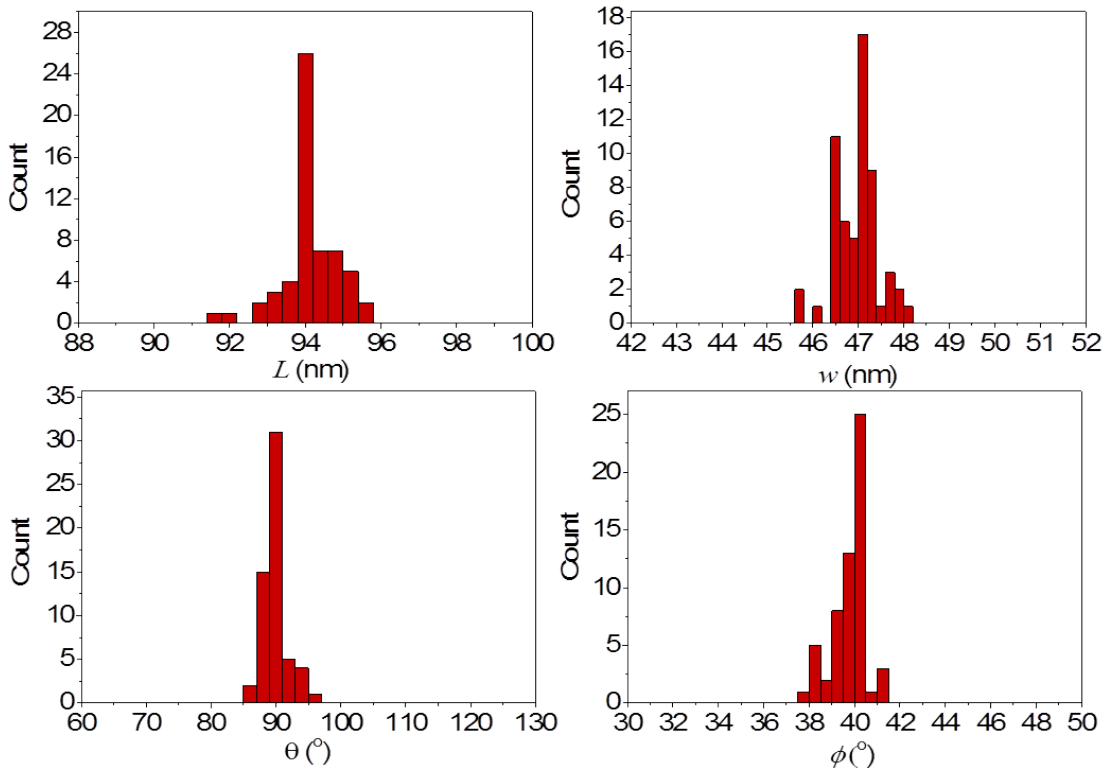


Figure 5.15: Histograms: results of fitting simulated data, for comparison with experimental results (Fig. 5.12, note different axis scales). 58 tiles were fitted. Average values of parameters are quoted in Table 5.2. The distributions are narrow, which confirms that the technique is accurate.

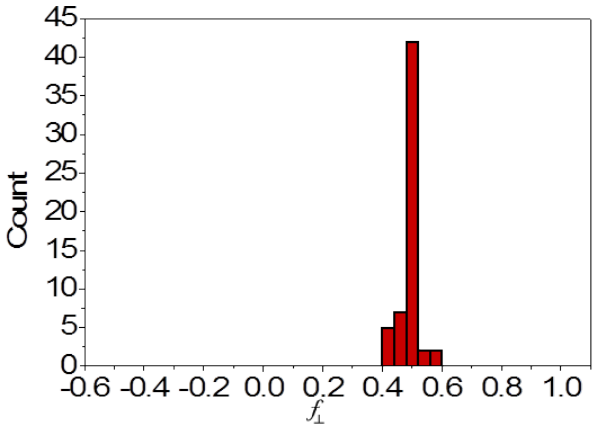


Figure 5.16: Fractional offset, $f_\perp$, for simulated data. The actual value for all tiles in the dataset was 0.5, and the distribution is sharply peaked at the correct position.

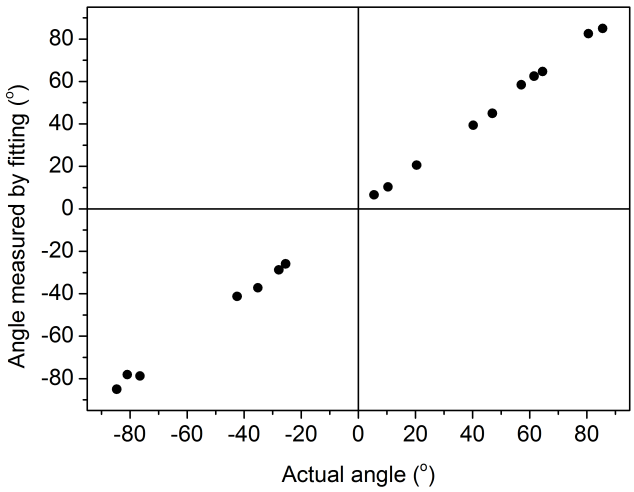


Figure 5.17: Analysis of simulated image containing 17 tiles oriented randomly. For all tiles the fitted orientation angle is almost identical to the actual value.

The parameter of most interest is the fractional offset, because it is the basis of classification of tiles into Types. The analysis of the simulated data shows that in 99.7% of cases the value for the parameter $f_\perp$ obtained from the fit is within 0.09 of the actual value. This validates the use of the image processing technique for tile classification, because it confirms that any uncertainty in the fitting process should almost never lead to misclassification of tiles (which would occur only if the measured value of $f_\perp$ differed from the expected value for the class by 0.17).

I also analysed a single simulated image consisting of 17 tiles with random orientations. All parameters apart from $\phi$ were the same as for the previous simulated dataset. Fig. 5.17 shows that in all 17 cases the value for $\phi$ extracted from the fit was almost identical to the actual value. The value of $f_\perp$ was correctly determined to be 0.50, with a standard deviation of 0.05. Similarly, $L$ and $w$ were measured to be 94.3 nm and 46.8 nm respectively, each with a standard deviation of less than 1 nm.

## 5.5 Alternative approaches

During the development of the image processing strategy I describe here I experimented with a range of different approaches. The image produced by application of the Sobel Filter clearly shows tile outlines. Initially I intended this to be an intermediate step, which would be followed by binarization and skeletonization, a process which can produce a set of outlines of width one pixel. However, this is undesirable because significant information is lost.

In early versions of my image processing procedures, I attempted to fit each tile edge with a separate line. This is overcomplicated because it involves 16 fit parameters, which should be related to each other - the method I ultimately used requires only 7. Furthermore, the process of fitting a line to one edge may be affected by the presence of the adjacent edges. These considerations provided the motivation for me to define the tile outline as shown in Fig. 5.6 rather than by using many intersecting lines. As mentioned in Chapter 2, it is possible to obtain some information on the tiles using the inbuilt MATLAB function `regionprops` but this function can only be applied to binary images, and thus may miss some details.

Many complex imaging processing algorithms exist, some created for applications such as facial recognition [162], and there is clearly scope for development of other new techniques for analysis of images of DNA nanostructures - it is possible to envisage tools which could accept as input a caDNAno file and automatically scan an AFM image for the presence of the designed shape, returning a quantitative assessment of

the nanostructures observed. Although in many respects this resembles the process of single-particle-reconstruction from cryo-EM images, here measurements would be provided for each individual object in the field of view, rather than averaged over all 'particles'. Such a powerful program would be outside the scope of this thesis.

## 5.6 Summary

I have discussed in this Chapter the configurations into which the polymorphic tile can fold, and explained how the shapes can be classified. The image processing strategy I have described yields high-precision quantitative information on the shape of the dimer tiles, which will be used in the following Chapter to draw conclusions about the folding pathways followed.

Analysis of simulated data confirmed that my image processing technique could be used to measure tile parameters with sub-pixel resolution, to within 2% of their true values. This validates its use to characterize tiles observed in AFM images.

# Chapter 6

# The folding pathways of the polymorphic tile

Having provided evidence in Chapter 4 for the existence of folding pathways for polymorphic tiles, I will now use the concepts and methods from Chapters 3 and 5 to describe in detail the self-assembly of 11 versions of the polymorphic origami tile - the original design plus 10 in which different modifications were made to the staple set[1]. I characterized AFM images of nine of them quantitatively and assessed images of the remaining two samples by visual inspection; one displayed extensive aggregation, while the modification made in the other had no noticeable effect.

Altering staples is analogous to introducing mutations to a protein, and some changes affect the choice of folding pathway dramatically, with the result that tiles made with the same homodimer scaffold but slightly different staple sets tend to fold

[1]Each design will be shown in this Chapter; full details of the specimens are given in the Appendices.

into different final configurations. I will explain the distribution of shapes observed for each variant and I will highlight analogies between assembly of the polymorphic tile and protein folding, featuring Levinthal's Paradox, the Anfinsen Experiment and studies of domain swap proteins. My results show that the folding pathway of polymorphic origami tiles can be manipulated using multiple strategies.

## 6.1 The original design

As discussed in Chapter 5, the configuration of the polymorphic tile is described quantitatively using the parameters $f_{\parallel}$ and $f_{\perp}$. It will be seen from Fig. 6.1 that

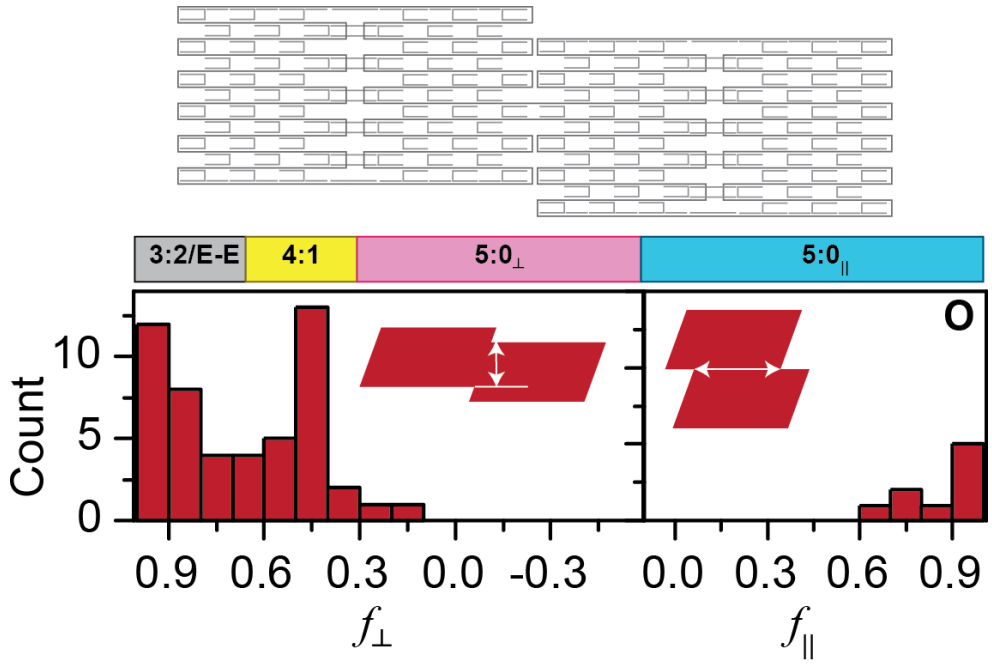


Figure 6.1: Original polymorphic tile design, depicted for one particular configuration (Type 3:2), and histogram showing distribution of shapes for tiles of original design. Shapes are described by the parameters $f_{\perp}$ and $f_{\parallel}$; lines with arrowheads on the tile shapes placed on the left and right hand graphs represent $f_{\perp}w$ and $f_{\parallel}L$ respectively, where $w$ and $L$ are the width and length of the tile. The coloured bar at the top indicates classification of shapes by Type. If $f_{\perp}$ is close to 1, the tile may be Type 3:2 or E-E.

| | **Case 1** | **Case 2** |
|---|---|---|
| Seam-duplex | 80.22(0.05) | 72.2 (0.1) |
| U-duplex | 68.7 (0.2) | 78.2 (0.2) |

Table 6.1: Measured hybridization temperature for indicated duplexes, in °C - for comparison with Table 6.2. Brackets: standard error on the mean.

| | **Case 1** | **Case 2** |
|---|---|---|
| Seam | 66.22 (0.04) | 66.39 (0.04) |
| U | 65.26 (0.06) | 68.7 (0.2) |

Table 6.2: Measured average incorporation temperatures for indicated staples, in °C: data for original design. Case 1: duplex corresponding to seam sequence forms at higher temperature. Case 2: duplex corresponding to U sequence forms at higher temperature. Numbers in brackets indicate the standard error on the mean, over repeated measurements. The two cases were monitored with different samples.

tiles of the original design fold into all Types of configuration (E-E/3:2, 4:1, 5:0)[2], but Type 5:0 tiles are less common than those of Types 3:2 and 4:1.

The distribution of shapes depends on the folding pathway followed, and to supplement my measurements of tile shapes, I examined in real time the process by which individual staples were incorporated into the origami, monitoring origami assembly using fluorescence as described in Chapter 3. I performed the experiment for two different cases - in the first case, the sequence of the selected seam staple gave rise to a higher hybridization temperature for the equivalent duplex, and in the second, the U-staple duplex formed at the more elevated temperature. Results for duplex formation were provided in Chapter 3 and they are summarized in Table 6.1.

Fig. 6.2 and Table 6.2 show the results of the fluorescence measurements of staple incorporation into the polymorphic tiles. The incorporation temperatures of the four

[2]Type E-E tiles for which $n = 3, 5$ are expected to occur infrequently because in these shapes a comparatively large number of staples must make long-range connections, and the shortest distance principle suggests this is unlikely. Tiles with a value of $f_{\perp} \approx 1$ are therefore thought to be slightly distorted Type 3:2 structures, resulting from sliding of the two rectangular halves.

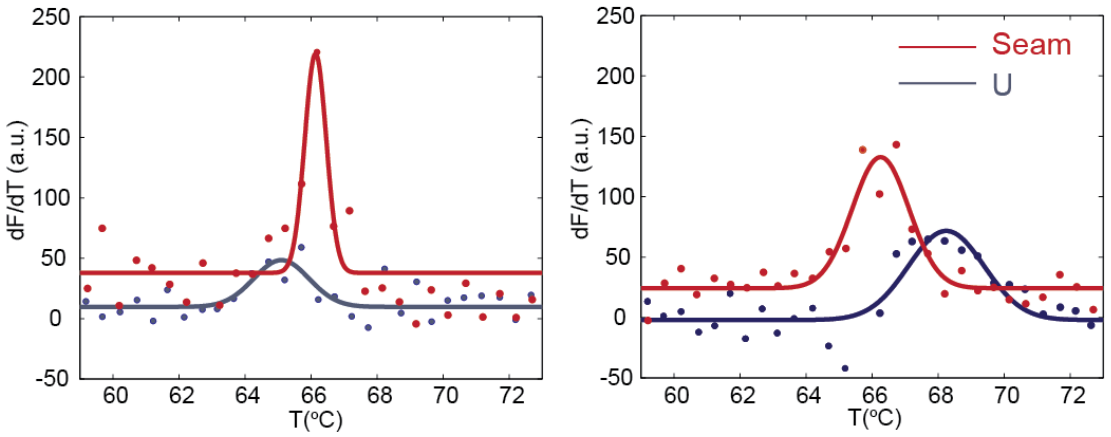


Figure 6.2: Staple incorporation - original design, polymorphic tile. Representative fluorimetric annealing curves for individual seam (red) staple and U-shaped (blue) counterpart for O tiles - left, for case where seam staple sequence is more stable and right, for case where U-staple sequence is stronger. Points represent data, while the solid line represents a Gaussian fit, the midpoint of which provides an estimate of the incorporation temperature.

staples in the origami tile differ by less than 4°C, while the hybridization temperatures of the corresponding duplexes vary by over 11°C. This is the result of cooperation between staples, discussed in Chapter 4.

The data clearly shows that the incorporation temperature of a staple is influenced by its base sequence. When the seam sequence is stronger than that of the U-staple, the seam staple falls into place first (by 0.96 ±0.05°C, Case 1), and when the sequence stability is reversed, the U-staple binds sooner (by 2.4 ±0.1°C, Case 2). Surprisingly, the two seam staples bind at about the same temperature, even though the corresponding duplexes hybridize at quite different temperatures. I attribute this to the effect of the seam staples' partners. The stronger seam staple has a weaker partner and the weaker seam staple has a stronger partner; the average of the melting temperatures predicted by NUPACK for the Case 1 seam staple duplex and its partner's duplex is the same as the average for the two corresponding Case 2 duplexes.

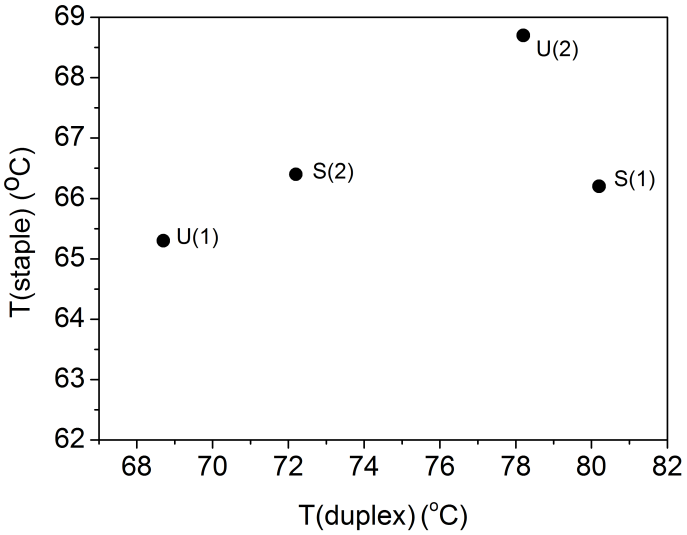


Figure 6.3: Comparing average duplex formation temperature, $T$(duplex), with average temperature ($T$(staple))) of staple incorporation into polymorphic origami tiles. Duplex measurements were made with higher concentrations of staple binding target than the origami experiments.

As discussed earlier, seam staples must cooperate strongly with their partners. Consequently the stronger seam staple binds at a lower temperature than might be expected, because it only experiences the additional stability introduced by the presence of a partner when the temperature is low enough for its partner to bind. Hence, staple sequence is significant in folding, but other factors also play a role in determining the pathway. This is illustrated by Fig. 6.3, where it will be seen that staple incorporation temperature is not entirely correlated with the formation temperature of the equivalent duplex.

Fluorescence measurements show which staple falls into place first, on average, and thus indicate the average order of staple incorporation. However, after a staple has bound to the scaffold it can unbind and rebind in the alternative configuration. The reversibility of each binding event is therefore also important. If staple A binds

| | **Earlier** | **Later** |
|---|---|---|
| **Reversible** | Some effect on pathway | Little effect on pathway |
| **Irreversible** | Significant effect on pathway | Some effect on pathway |

Table 6.3: Hierarchy of interactions.

| | **Earlier** | **Later** |
|---|---|---|
| **Reversible** | Some of the U-staples | Other U-staples |
| **Irreversible** | Some of the seam staples | Other seam staples |

Table 6.4: Nature of different families of staples.

at time $t_A$ and staple B binds at time $t_B > t_A$, but staple A is highly likely to unbind again before $t_B$, then A would be expected to have a comparatively small influence on B. On the other hand, if staple A is locked in place when B binds, it would be expected that A could potentially affect the behaviour of B.

This implies that there is a general hierarchy of interactions shaping origami folding pathways, as outlined in Table 6.3. In the case of the seam staples, the binding of one partner reinforces the other, decreasing the probability of unbinding and reconfiguring. Consequently, seam staple incorporation is less reversible than U-staple incorporation[3]. Table 6.4 shows how the suggested hierarchy of interactions in folding applies to particular families of staples in the original polymorphic tile.

This argument suggests that the seam staples are likely to dominate folding, but some of the U-staples are in a position to compete with them to some extent. These two staple families favour different final configurations, due to the 'shortest distance principle' articulated earlier. According to this principle, if the free domain of a half-bound staple could bind to either of two sites, it will be more likely to bind to the

[3]This effect is seen in simulations performed by my colleague, Frits Dannenberg, which show quantitatively that the number of times a U-staple unbinds and reconfigures is higher than the number of equivalent events for the seam staples.

nearer site. Hence, for a given set of staples, the most likely configuration is that which minimizes the length of scaffold loops formed by these staples. The circle diagrams[4] in Fig. 6.4 show the connections made by either seam staples or U-staples in tiles of Type 3:2 and $5{:}0_{\|}$. Here, each black spot represents half an arm of the tile, and these lengths of scaffold are connected by various staples. The longer the arc, the longer the loop of scaffold bridged by the corresponding staples.

The diagrams clearly show that to satisfy the shortest-distance principle, U-staples binding in the absence of seam staples should favour the $5{:}0_{\|}$ structure, whereas if the seams bound in isolation (with no U-staples present) they would favour 3:2 shapes - for the U-staples the arcs are shortest for the $5{:}0_{\|}$ tiles, and for seam staples the shortest arcs occur in the Type 3:2 tile. For the seam staples, Type 4:1 configurations are slightly less favourable than those of Type 3:2 but not as unfavourable as those of Type 5:0. Based on the hierarchy of interactions identified in Tables 6.3 and 6.4, this suggests that the original design of the polymorphic tile should tend to fold into tiles of Type 3:2, with a significant number of Type 4:1 tiles and few Type 5:0 tiles. This is exactly what is seen experimentally (Fig. 6.1).

The behaviour of the original polymorphic tile design suggests that the folding pathway and thus the final distribution of shapes will change significantly if the hierarchy of interactions is modified such that the seams bind more reversibly and at a later stage of folding. In this case it would be expected that the U-staples would dom-

[4]Circle diagrams formatted using a modified version of an Adobe Illustrator script written by Dr Jonathan Bath.

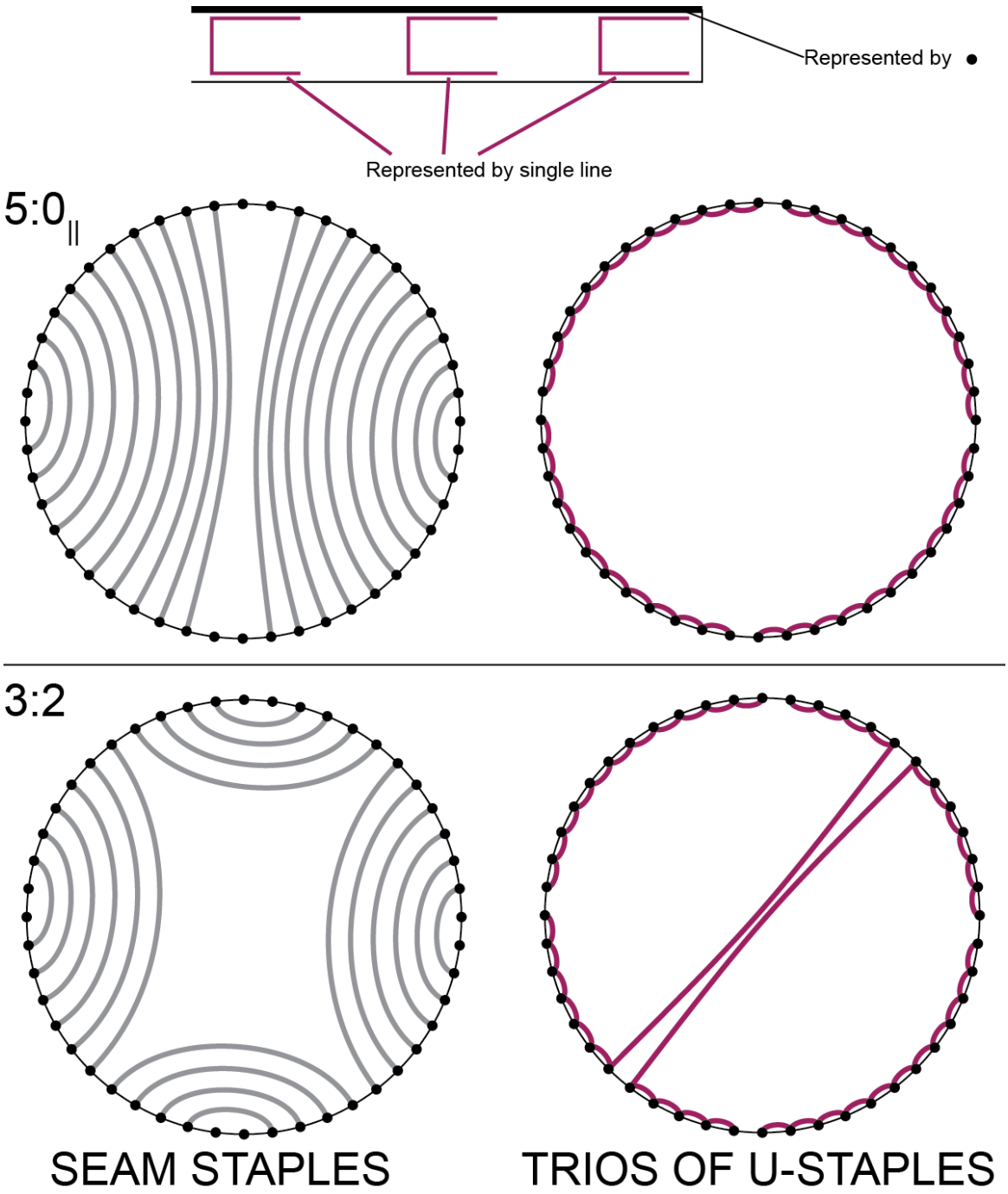


Figure 6.4: Circle diagrams for Type $5{:}0_{\|}$ and Type 3:2 configurations; arcs represent connections made by staples and black spots represent segments of the scaffold, as shown. All seam staples are shown, and one purple arc corresponds to a group of three U-staples.

inate the self-assembly process, and the shortest-distance principle dictates that the majority of the origami structures observed would fall into the Type 5:0 category. I tested this prediction by critically weakening the seam, with two different approaches.

## 6.2 Critically weakening the seam

### 6.2.1 Omitting half the seam staples; aggregation

I synthesized a variant of the polymorphic tile in which every other seam staple was omitted, and the corresponding scaffold domains left single-stranded (Fig. 6.5a). The result was that the tile folded badly, with very few 'good' tiles and extensive aggregation (Fig. 4.3b). The formation of tile aggregates was not observed in any of the other experiments I performed, so it is reasonable to attribute the clumping of partly-folded tiles to non-specific interactions between the single stranded domains of different units. To test the plausibility of this explanation, I checked the sequences of the unpaired scaffold domains for complementarity using NUPACK. There are several cases in which interactions could occur - for instance, the domain $3[112] - 3[127]$ can form a complex with $9[96] - 9[111]$ by base-pairing[5].

The aggregation observed in the half-seam-omitted polymorphic tile is reminiscent of the aggregation of proteins due to non-specific hydrophobic interactions [12], which is associated with the formation of amyloid fibrils, implicated in a range of diseases.

[5]The sequences of these two scaffold domains are: CCCGACACCCGCCAA and ACGGGGAGTCAGGCAA, and for strands at a concentration of 20 nM, NUPACK predicts that around 1/20 of the strands will form complexes at 20°, with up to 9 base pairs.

The tile aggregation phenomenon was readily suppressed by introducing strands of DNA to bind to the unpaired single domains, which is equivalent to breaking every other seam staple rather than omitting them completely (HSB sample, Fig. 6.5b and Fig. 4.3c). It is important to note that the *only* role of the new single domain staples is to suppress aggregation.

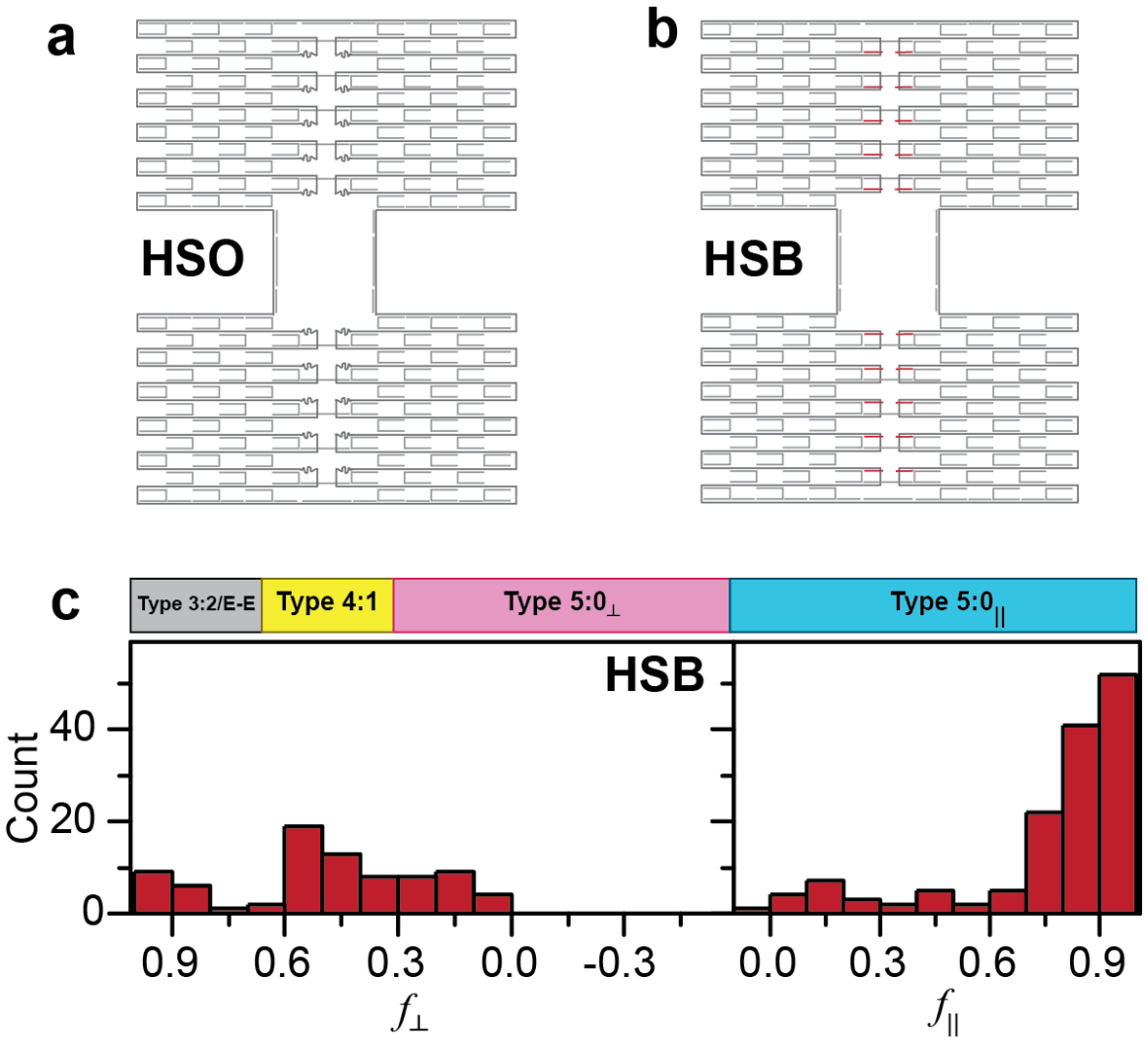


Figure 6.5: Breaking or omitting half the seam. (a) Half-seam omitted (HSO) design, depicted for tile Type expected to be most abundant. Many of the HSO tiles misfolded or aggregated. (b) Half-seam broken (HSB) design; the most common tile configuration is depicted and modifications to the original design are marked in red. (c) Distribution of tile configurations - sharp contrast with data presented for original design in Fig. 6.1.

### 6.2.2 Breaking half the seam staples

In the case of the half-seam-broken polymorphic tile, there is a strong bias in favour of one particular configuration, in comparison with the original design. This contrast is visible qualitatively in the AFM images (Fig. 4.3a,c) and quantitatively in Fig. 6.5c. Now there are far fewer tiles of Types 3:2 and 4:1, and most of the well-folded tiles are of Type 5:0. In this case the seam staple partnerships are completely non-existent - all seams are now single. This means that there is no longer such a strong cooperative effect acting on the seam staples and they now bind with less ease.

This is confirmed by fluorescence measurements (Fig. 6.6), which illustrate that representative seam staples are incorporated later than U-staples, even when the seam sequence is stronger than that of the U-staple (Table 6.5 and Fig. 6.7). The incorporation temperatures of the U staples are almost identical to the values obtained for the original design, but for each seam staple the transitions have been shifted to

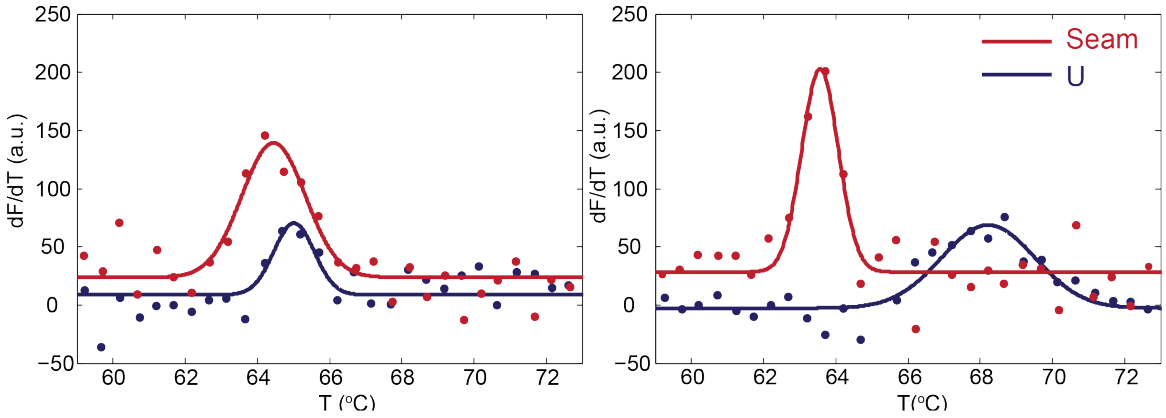


Figure 6.6: Staple incorporation - HSB tile (representative data). Fluorimetric annealing curves for individual seam (red) staple and U-shaped (blue) counterpart - left, for Case 1 (seam staple sequence stronger) and right, for Case 2 (U staple sequence stronger). Points represent data, solid line shows Gaussian fit. Average values of transition temperature obtained from fits to multiple datasets are shown in Table 6.5.

| | **Case 1** | **Case 2** |
|---|---|---|
| Seam | 64.7 (0.1) | 63.55 (0.03) |
| U | 65.2 (0.1) | 68.8 (0.2) |

Table 6.5: Measured incorporation temperatures for indicated staples, in °C: Half-seam-broken design, for comparison with Tables 6.2 and 6.1. Case 1: seam sequence stronger. Case 2: U sequence stronger. Numbers in brackets indicate the standard error on the mean. The two cases were monitored with different samples.

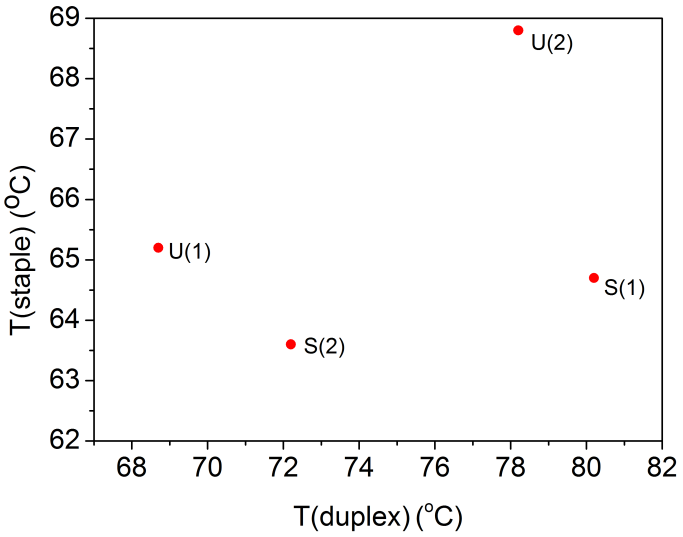


Figure 6.7: Comparing duplex formation with staple incorporation into HSB tiles. Labels: same as Fig. 6.3. Comparison with Fig. 6.3 reveals that seam staple incorporation temperature is shifted downwards significantly; U-staple incorporation occurs at the same temperature as before.

lower temperatures. The data therefore shows that in this design variant the seam staples bind later than before and have a smaller effect on folding. Hence, it may be deduced that that the U-shaped staples funnel the polymorphic tile towards the Type 5:0 shapes, while seam staples compete with them to divert it into alternative configurations.

### Analogy with Anfinsen experiment

This result is analogous to the observations made by Anfinsen and his colleagues in their classic work on protein folding, in which they demonstrated that the protein's primary sequence contains all the information required to specify its final three-dimensional form [14]. In their experiments [163] they denatured the protein bovine ribonuclease in the presence of 8 M urea and beta-mercaptoethanol, to convert it into a randomly coiled peptide with no enzymatic activity. The urea was used as a denaturing agent, while beta-mercaptoethanol reduces (i.e. breaks) disulphide bonds between cysteine residues. When both chemicals were removed, the protein slowly regained its enzymatic activity as it refolded and the sulphydryl groups became re-oxidized slowly to re-form the disulphide bridges. However, when the disulphide bridges were restored before the urea was removed, only about 1% of the functionality of the protein was regained. This is because it was only in approximately 1% of the molecules that the cysteines had been correctly linked together by random chance.

In the polymorphic origami tile, paired seam staples quasi-irreversibly form strong links between distant regions of the scaffold and they are therefore analogous to the disulphide bridges in Anfinsen's protein. When some of them bind earlier, a range of structures is possible, but when they bind later many of the U-shaped staples have already bound, constraining the seams to fall into particular places.

## 6.3 Alternative seam and lower right modifications

To investigate alternative strategies for controlling origami assembly, I prepared a modified version of the half-seam-broken tile, referred to as the 'alternative seam' ('AS') system (Fig. 6.8a). Here, while the main seam was half-broken, a set of staples on the edge of the tile was modified to create a new seam, with paired staples linking regions of the scaffold which are separated by a comparatively long distance when the origami is unfolded.

It will be seen from the data in Fig. 6.8a that the shape distribution has shifted dramatically in comparison with the half-seam-broken tile (Fig. 6.5c), and the Type 3:2 configuration is now strongly preferred (81% of the tiles are of Type 3:2/E-E). The tile is funnelled into this particular configuration because the alternative seam staples are paired and therefore should fall into place quasi-irreversibly at a comparatively early stage of folding, and because all non-target structures have an energy penalty. In this system tiles of Type 4:1 or 5:0 can only form if some staples are partially unbound or double helices are bent at extreme angles.

A second approach for controlling folding ('F') is shown in Fig. 6.8e. Here a long staple is used to increase the rigidity of a particular section of the scaffold, with the aim of forcing the polymorphic tile into the configuration depicted (Type 5:$0_{\perp}$), in which its two rectangular halves are joined at the corner by a continuous duplex. The image (Fig. 4.3e) shows a range of shapes, and there is apparently no bias in favour of the target structure. This is probably because multiple copies of the modified

staple have bound, such that connections between scaffold domains have not formed as intended and the folding pathway has been unaffected[6].

A successor to this design is presented in Fig. 6.8a, where a 63 base continuous staple has been introduced ('LR'). Longer strands of DNA hybridize at higher temperatures, and this strand is four times longer than the normal staple domain, which ensures that it will be the first staple to bind as the origami is assembled by cooling. This event creates a rigid double-helical section in the partly-folded intermediate, around which the rest of the tile must fold. Because this domain cannot bend, the target configuration is the only one which is consistent with the ultra-long staple being bound, and hence the folding pathway should be strongly biased towards the target shape, the conformational freedom of the scaffold being constrained during folding.

Fig. 6.8a ('LR') shows that this approach is extremely effective. The shape distribution is strongly peaked at the expected position for the desired structure and it is clear that the folding pathway has been directed as desired. It will be seen that a number of tiles are observed with a negative $f_{\perp}$. These shapes arise due to the lack of crossovers in the immediate vicinity of the long linker - the lower tile half can slide downwards without restriction, resulting in a negative offset.

The LR modification echoes experiments on domain swap proteins in which a key amino acid was shown to alter the tendency for domains to swap by increasing the

[6]I did not characterize this specimen quantitatively - partly because the modification did not appear to have had any effect and partly because the synthesis conditions I used for this tile were not as standard. The density of tiles on the mica was also abnormally high which would have complicated attempts to analyse this data quantitatively because many of the objects in the field of view are in contact with others.

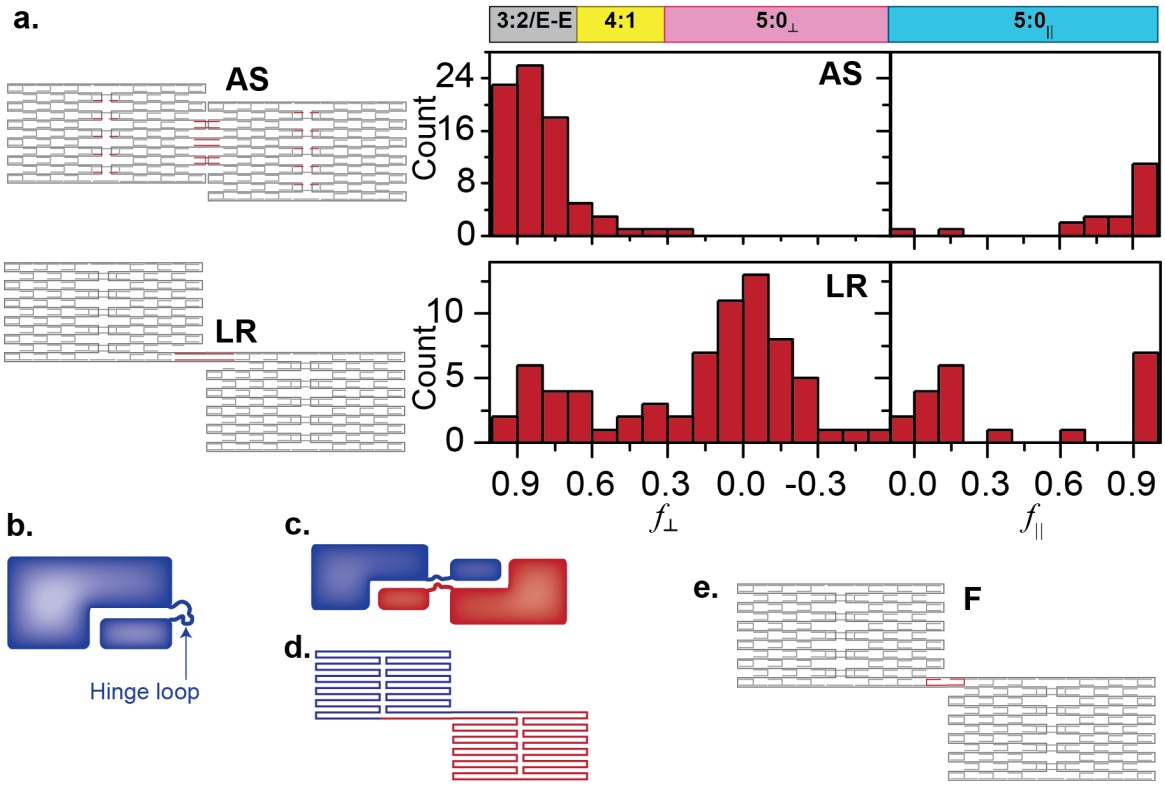


Figure 6.8: Alternative Seam (AS) and Lower Right (LR) modifications to polymorphic tile design, and predecessor of LR variant. (a) Designs of AS and LR variants, accompanied by distribution of $f_{\parallel}$ and $f_{\perp}$ for each case. Tile design is shown using the target configuration; modifications to original design are marked in red. (b) Sketch of a monomer of a domain swap protein, showing two structural domains connected by a hinge loop. Two such monomers can dimerize to form a domain-swap protein as shown in (c). (d) One of the scaffold routings for the polymorphic origami tile, for comparison. (e) Failed design 'F': predecessor of LR variant. Target configuration shown, modification to design marked in red.

rigidity of a particular section of the protein. In a domain swap protein two or more monomer units become associated via an exchange of structural components i.e. a domain from one unit takes the place of the corresponding domain in the other [164], as shown in Fig. 6.8b,c. The resulting interlocked configuration resembles a certain scaffold routing of the polymorphic origami tile (Fig. 6.8d). The point at which the scaffold crosses from one tile half to the other is a parallel of the hinge loop in the protein, the link between the swapped domain and its parent monomer [165].

It has been found that a proline residue in the hinge loop increases the tendency of the domain to swap [166, 167] . The side chain of the amino acid proline is locked to the backbone nitrogen in a cyclic arrangement [14] and consequently proline significantly restricts the conformational freedom of the peptide chain of which it is a part, increasing the chain's rigidity. This is exactly what the 63 base staple in the polymorphic tile does, and it affects the origami folding in a similar way, constraining dynamics during folding and introducing a helix-bending energy penalty for structures other than Type $5{:}0_{\perp}$.

## 6.4 Creating a single strategically located flaw

I also investigated two polymorphic tile variants with single strategically located flaws. In each of the two designs (Fig. 6.9) one pair of seam staples is broken[7]. The data

[7]I also examined a sample in which I had *omitted* the first seam. I did not observe aggregation, in contrast with the sample in which I omitted every other seam staple, but visual inspection suggested that the effect on the pathway was not significant. For better comparison with the half-seam-broken sample, I repeated the measurement with the first seam *broken* rather than *omitted*, and I did not consider the 'first seam omitted' sample further.

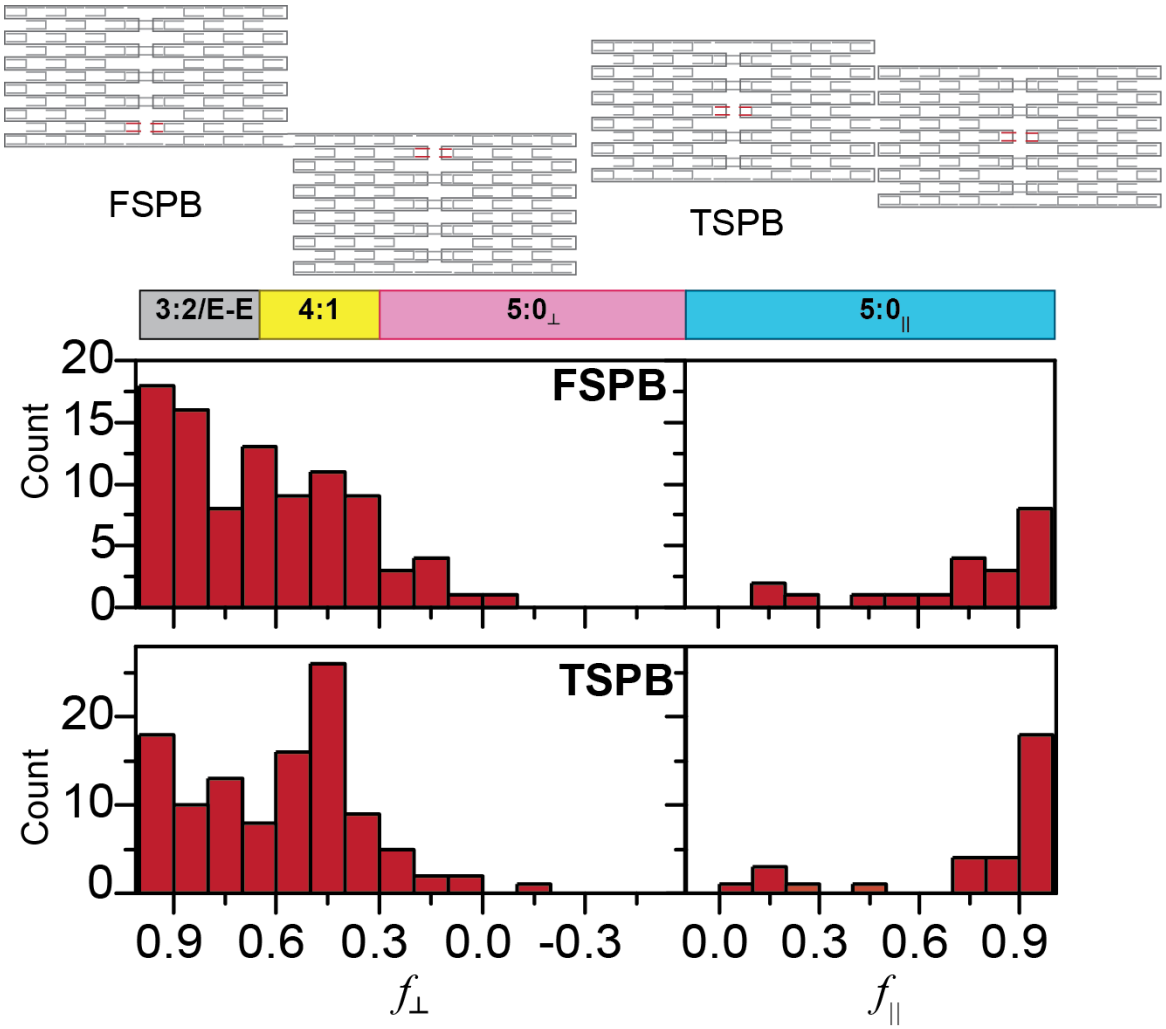


Figure 6.9: Breaking one pair of seam staples. FSPB=First Seam Pair Broken, TSPB=Third Seam Pair Broken. Distribution of $f_{||}$ and $f_{\perp}$ for these two designs. The two designs are shown - the configuration depicted in each case is that which was expected to dominate, and the broken seams are marked in red.

in Fig. 6.9 shows that breaking either the first or the third seam staple pair does not have a significant effect on the folding pathway. This reinforces the theory that staples act collectively, because breaking one pair of staples does not affect the outcome dramatically.

## 6.5 Different concentrations for selected staples

The usual rules of chemical reactions state that if the concentration of reactants is increased, the position of equilibrium will shift in favour of the products. This implies that increasing the concentration of a particular staple should increase the rate at which it binds to the scaffold. Hence, if the concentration of some staple families is changed, it might be expected that the folding pathway would change. My colleague's simulations take account of this effect, and they indicated that the setup shown in Fig. 6.10 (SC) would result in a bias towards Type 3:2. In my experiment I observed that the concentration of staples was so high that the standard purification procedure was inadequate, and a large number of staples remained, visible as small spots in the AFM image (shown in Chapter 4, Fig. 4.4c). Despite this, it was clear that the distribution

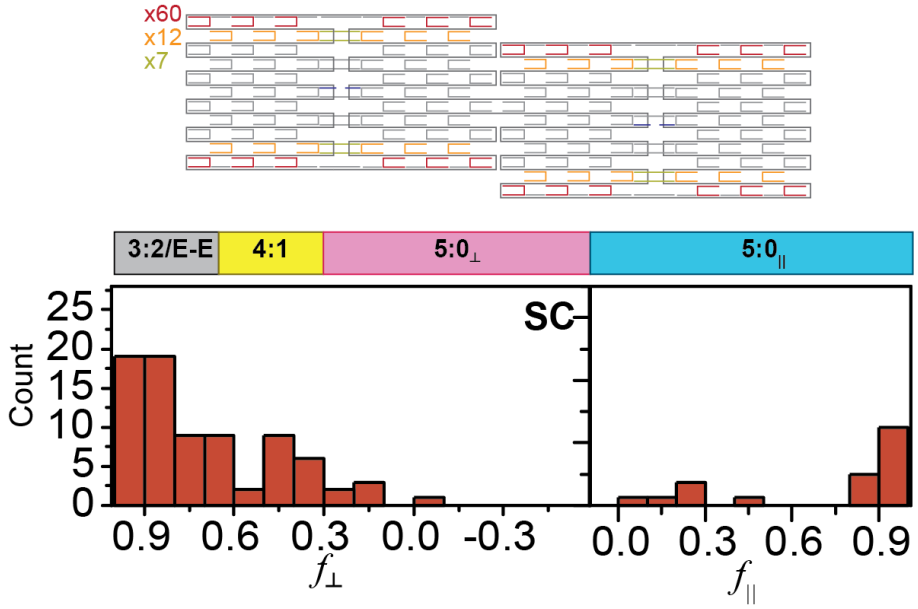


Figure 6.10: Using different concentrations for selected staple families: resulting shape distribution. Design is represented using the target configuration. Concentration of grey staples $\sim 8\times$ concentration of scaffold; yellow/orange/red staples were 7/12/60 times as concentrated as the grey staples, respectively.

of tile configurations was not significantly different from the original distribution (Fig. 6.10).

It is possible that in this sample more than two copies of some staples bound to the scaffold during folding, such that all scaffold binding sites were occupied but some staples were only bound with one domain. Here the process of strand displacement would become significant, with one staple needing to displace its competitor to bind fully. This could influence staple dynamics, and it might suppress the expected effect of increasing staple concentration.

The behaviour of this sample suggests that the exact stoichiometric ratio between scaffold and staples is irrelevant to the folding pathway, implying that pipetting errors and variation arising during staple synthesis are unlikely to change the outcome of the folding process. This agrees with the finding of Song *et al.* that staple kinetics at 60°C are not affected by an increase in staple concentration [101]. However, I confirmed earlier (Chapter 2) that if the concentration of staples is considerably lower than that of the scaffold, even the monomer tile does not fold, which suggests that staple concentration is only insignificant when the staples are present in excess.

## 6.6 Controlling the folding pathway by changing the staple pattern

I will now proceed to describe two further design modifications which were intended to force the polymorphic tile to fold preferentially into configurations of Type 3:2. In contrast to the 'Alternative Seam' design, the staple arrangements described here are compatible with formation of Type 4:1 and 5:0 tiles, and the assembly pathway is modified by a change to the order of staple incorporation rather than the introduction of a strong geometric constraint.

The 'Extended U-staple' design (Fig. 6.11) contains eight staples in which each domain is twice the normal length, while some of the staples in the 'Very Extended U-staple' version (Fig 6.11) are 80 bases long. The 'VEU' design has several other modifications, and the order in which the staples are expected to bind is indicated by the colour coding in the figure (where the design is based on considerations of staple length and pairing, without taking account of the different sequences).

When the length of the two domains of a staple is increased, it should bind more strongly to the scaffold due to the increase in the number of base pairs it can form. Consequently staples should be incorporated at a higher temperature if they have been extended. This means that their effect on the folding pathway will be increased. If the extended U-staples were to bind to the scaffold in isolation, each domain having two possible binding sites, the shortest distance principle suggests that they would

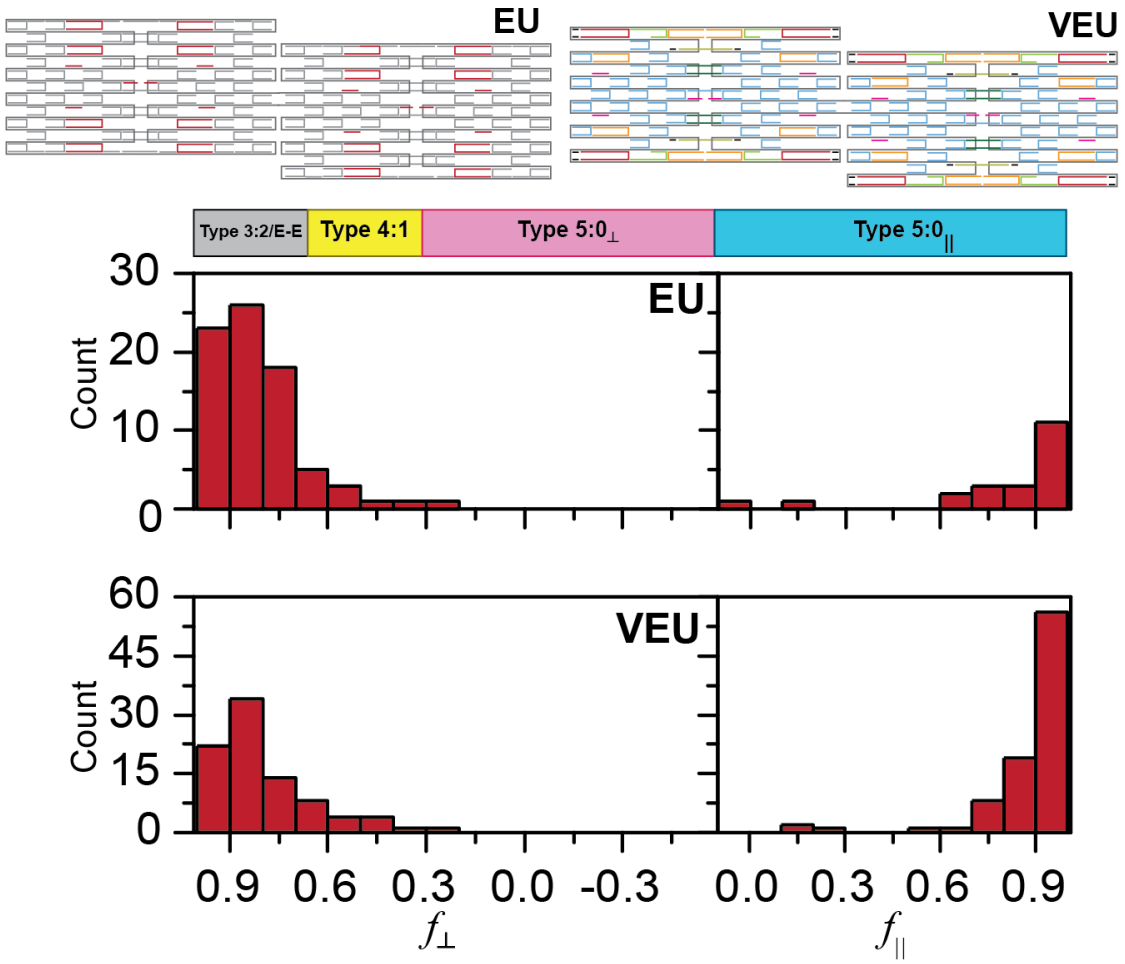


Figure 6.11: Extending selected staples: EU (Extended U-staples) and VEU (Very Extended U-staples) tile designs and distribution of tile shapes for each . Both variants were expected to give rise to a strong bias in favour of Type 3:2 tiles. Designs: for the EU variant, the modified staples are marked in red, and for the VEU design the colour coding indicates the intended sequence of incorporation (from first to last: red, orange, dark yellow, pale green, dark green, pale blue, pink, black).

be most likely to bind in the arrangement which forms the smallest loops of scaffold. This is the Type 3:2 configuration.

At the temperature of incorporation of the other staples, the long staples are unlikely to unbind and they should therefore be expected to bias tile assembly by pinning the scaffold into a Type 3:2 configuration and forcing the other staples to fall into place around them. Experimental results (Fig. 6.11) show that this works, and the number of Type 3:2 tiles is enhanced at the expense of Type 4:1.

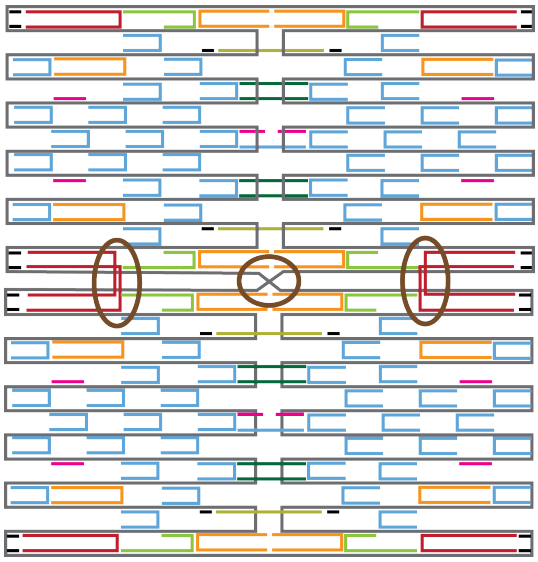

Figure 6.12: Proposed explanation for the failure of the VEU design to behave as expected - the extra-long staples are long enough to form longer range connections than predicted by the shortest distance principle. The unexpected problems are marked with brown ellipses.

The 'VEU' design was intended to generate an even stronger bias in favour of Type 3:2 tiles.[8] However, it will be seen from Fig. 6.11 that the experimentally observed number of Type 3:2 tiles in the VEU specimen is comparable to the number seen in the original design, but the number of Type 4:1 has gone down dramatically and the number of Type $5{:}0_{\|}$ is considerably enhanced.

I attribute the failure of this design to the misfolding scenario illustrated in Fig 6.12. The longest staples are of sufficient length to cross between the two halves without being significantly destabilized, and they should perhaps be regarded as polymer chains in their own right, each with the dynamics of a short freely jointed chain.

[8]Before I performed the experiments, the behaviour of these two designs was checked using my colleague's simulations; the prediction for the EU design was approximately correct, but the VEU design was predicted to induce an even stronger bias for Type 3:2, and it failed to do so.

## 6.7 Misfolds, false dimers and twisted tiles

### 6.7.1 Misfolded tiles

Fig. 6.13 shows a range of 'misfolded' or otherwise unclassifiable structures. The clarity with which misfolded tiles can be imaged varies - a very sharp tip and good imaging conditions are required if unfolded scaffold is to be seen. Thus, tiles which appear to have a fragment missing may either be broken, or attached to an unfolded loop of scaffold which is not visible in the image. Tiles of this form appear to be derived from any of the possible configurations, without significant bias[9].

A few of the tiles of the correct size which appeared to have all staples in place could not be fitted because they were bent or distorted. These tiles could not be characterized or classified by Type, but occurred comparatively infrequently. For example, for both the original design and the half-seam-broken tiles, the tiles which were eliminated from consideration on the sole basis of bending or distortion amounted to 8% or fewer of all the tiles which were examined in each case. If tiles of one type were more susceptible to bending than another, this could affect the measured shape distribution, but the fraction of tiles which are bent is sufficiently small for this potential bias to be disregarded when large differences in the morphology distribution are being discussed.

Tiles were sometimes found to be in contact with other tiles. In most cases these 'composite' objects were eliminated automatically when the size of tiles in the field

[9]Based on visual assessment of the tiles in question.

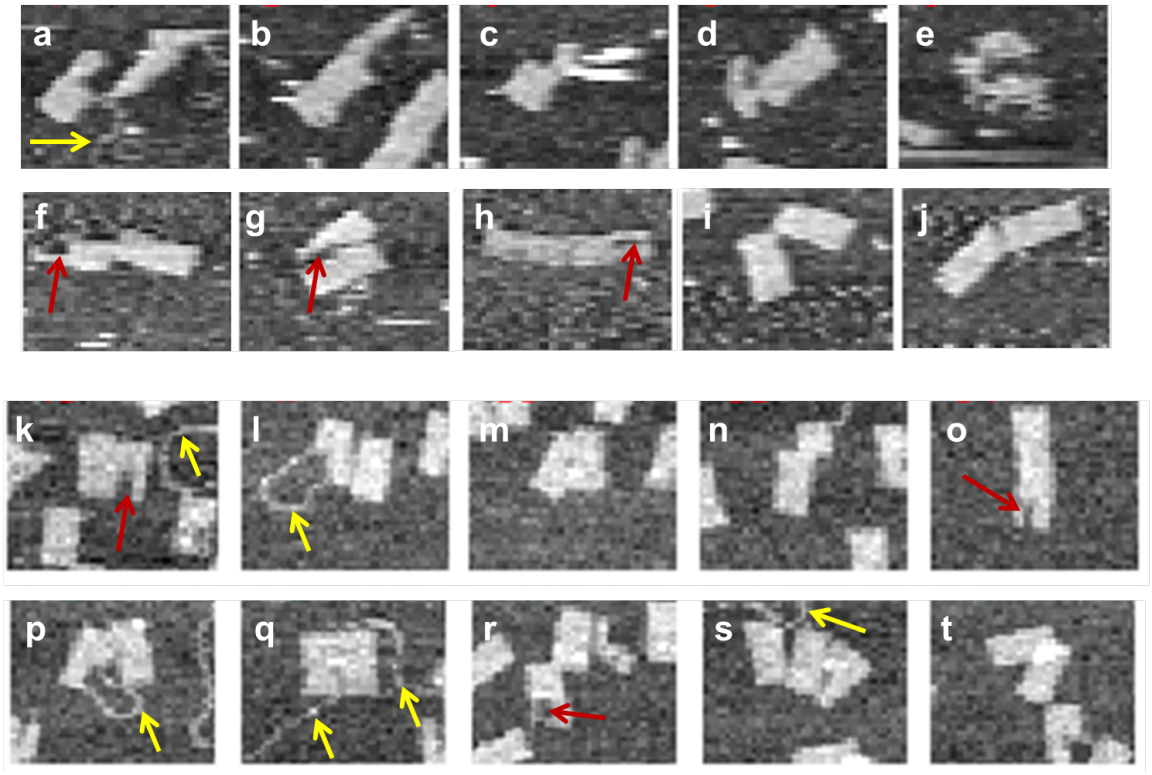


Figure 6.13: AFM images of representative misfolded and distorted tiles which were flagged as being of the correct size to be dimers. (a)-(j) taken from images of Original design tiles, (k)-(t) are from the Half Seam Broken specimen. A similar range of misfolded shapes exists in both samples. Each image is 40 pixels (234 nm) wide. Red arrows indicate structural flaws, yellow arrows mark loops of unfolded scaffold. Tiles (i), (j), (m) and (t) appear to be fully stapled but they are bent or distorted. Several others (such as (b) and (f)) appear to have a fragment missing, (e) has failed to fold into any recognisable shape and (s) consists of multiple tiles in contact with each other.

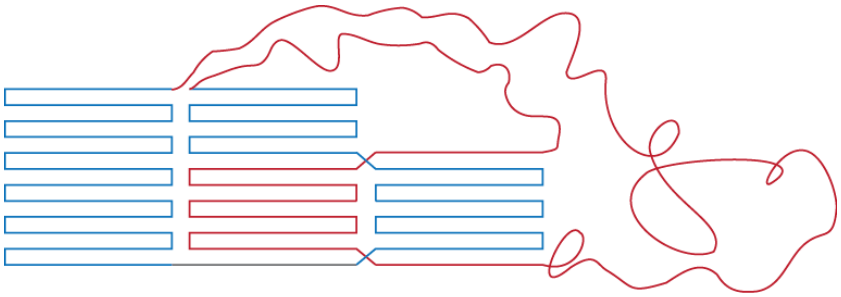

Figure 6.14: Possible mechanism for misfolding. Sections of the polymorphic origami tile fold independently in such a way that it is geometrically impossible for the tile to fold completely. The tile is depicted without staples. The straight lines correspond to folded sections and the irregular lines are the unfolded parts; in the state shown, some of the unfolded regions could still in principle fold, but others would necessarily be left single-stranded.

of view was checked, but a few composite tiles were flagged as possible well-formed dimers and these could not be fitted.

True misfolds may arise when two parts of the polymorphic tile fold separately along pathways which do not converge. A possible misfolding scenario is depicted in Fig. 6.14. The misfold depicted is similar to some of the shapes observed in the AFM images, but other misfolded tiles seem to be almost perfect, with small flaws. These faults are likely to result by chance when just a few staples fail to bind during assembly, and a small group is left out. At the other end of the spectrum, some structures form amorphous blobs. Here, only a few staples have bound.

### 6.7.2 Absence of monomer tiles

I have assumed during analysis of my data that all the structures which can be fitted are indeed dimer tiles, rather than pairs of neatly aligned monomer tiles. This assumption is justified because in most of the images the tiles are well isolated, and

gel electrophoresis confirmed that the dimer plasmid stock was not contaminated by monomer. Furthermore, the poly-T extensions to the staples at the edge of the tiles suppressed tile stacking in the monomer, and should therefore be expected to do the same here. It is also worth noting that even if there were some monomer tiles present in a specimen, and if pairs of them lined up by chance in exactly the right way to form false dimers, there is no general mechanism by which this phenomenon could cause the shape distribution to shift when the design is changed.

### 6.7.3 Twisting tiles

The stacking interaction between consecutive bases is extremely significant in determining the molecular structure of DNA (Chapter 1). It might therefore be supposed that base-stacking would have an effect on the shape distribution of polymorphic tiles, and that some tiles might twist into the end-to-end (E-E) configuration (Fig. 6.15), in order to maximize the number of stacked bases. However, as noted above, the poly-thymine 'tails' of the staples at the tile edges suppress inter-tile stacking in the

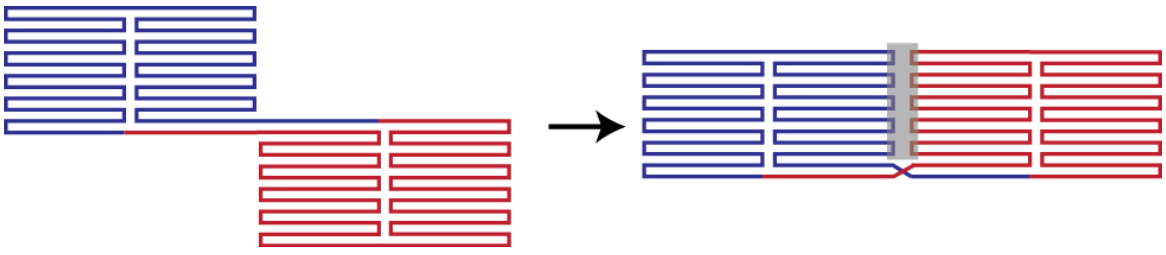

Figure 6.15: Potential change in observed shape of polymorphic tiles as a result of twisting. The tile shape on the left could be converted into that on the right by twisting, enabling base stacking interactions to occur in the shaded region. However, steric hindrance between the T-tails at the ends of the helices in this region is expected to interfere with this effect, and twisting is likely to be further suppressed due to the strain the process would place on the DNA.

monomer, and would therefore also be expected to suppress base-stacking interactions between the two rectangular halves of the dimer. Twisting a tile would also introduce strain into the structure, and this should further decrease the likelihood of forming E-E shapes by twisting. Hence, this effect should be comparatively insignificant in determining the shape distribution of the polymorphic tile.

## 6.8 Conclusions

Fig. 6.16 shows the relative abundance of the four different classes of tile configuration, for the nine different design variants which I characterized quantitatively. I have discussed two other designs in qualitative terms.

Some of the design changes had a substantial effect on the folding pathway (Table 6.6) while others did not (Table 6.7). Effective modifications involved interfering with

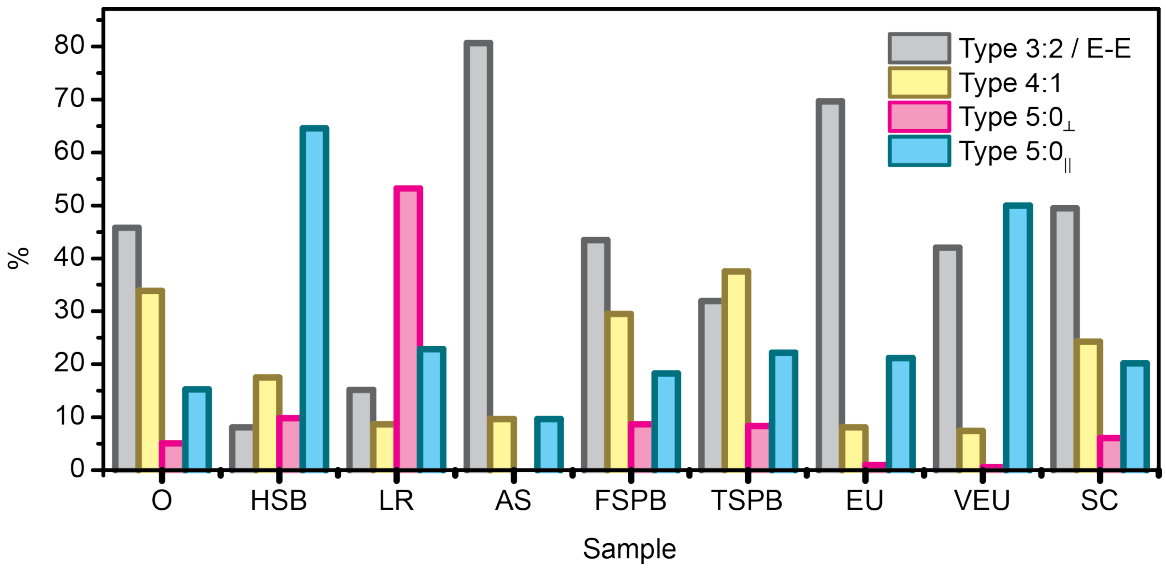


Figure 6.16: Relative abundance of different tile types: all design variants.

| Modification | Effect |
|---|---|
| Omitting half the seam staples | Aggregation |
| Breaking half the seam staples | Type 5:0 preferred |
| LR: Long 1-domain staple in lower right corner | Type $5{:}0_{\perp}$ preferred |
| Alternative seam | Type 3:2 preferred |
| Some long U-staples | Type 3:2 preferred |

Table 6.6: Summary: design modifications which significantly altered the folding pathway, and their effect.

| Modification | Apparent reason for failure |
|---|---|
| Single pair of seam staples broken | Modification too small; overcome by cooperative effects |
| Different staple concentrations | Folding pathway insensitive to staple concentration |
| Long 3-domain staple in lower right corner | Multiple copies of staple bound |
| Some very long U-staples | Staples long enough to make unintended connections |

Table 6.7: Summary: design modifications which did not significantly alter the folding pathway, and the reason.

the seam or extending selected staples. Examples of 'silent mutations' included a single broken seam, or altered concentrations of selected staples.

### 6.8.1 Closing remarks

The results described in this Chapter show that the assembly pathway of the polymorphic DNA origami tile can be controlled using a range of approaches. The outcome of the folding process for different design modifications offers insights into the principles governing origami self-assembly, and various parallels exist with protein folding.

# Chapter 7

# Conclusions

## 7.1 Summary of the thesis

The use of DNA as a nanoscale engineering material was reviewed extensively in Chapter 1, with a focus on DNA origami. I proceeded to describe (in Chapter 2) how I successfully synthesized a customized monomer plasmid, and used it as a scaffold for the fabrication of DNA origami tiles. My AFM images of monomer tiles on a mica surface revealed that tile orientation was not random, and the two preferred angles were separated by approximately 60°. This is attributable to alignment of tiles with terraces, which are expected to form when the mica is cleaved.

To shed light on the behaviour of staples during the annealing process, I developed an experimental protocol for monitoring them in real time throughout assembly, and Chapter 3 described my approach in detail. I used staples labelled with a fluorophore and a quencher; when both domains of the staple bound to the scaffold, the fluores-

cence increased. The study enabled me to examine the incorporation of individual staples. I was able to achieve high-resolution measurements, identifying the order of particular binding events even when transitions were separated by less than 1°C.

Chapter 4 covered the synthesis of the dimer plasmid and its use as an origami scaffold. When annealed with the same staple set as for the monomer, it folded into polymorphic tiles, many of which were well-formed despite the huge potential for misfolding. This suggested that there are folding pathways for origami structures.

The distribution of observed tile configurations reflects the nature of the assembly pathway, and to facilitate investigation of this I introduced in Chapter 5 a tile classification system based on the connectivity of seam staple pairs in each shape. In the same Chapter, I discussed the image-processing procedure I created and developed to extract quantitative information from AFM images. My method represents an advance over the usual approach for analysis of AFM images of DNA origami, which involves assessing structures by eye, classifying them as properly assembled or misfolded and counting them manually. I was able to demonstrate using simulated data that my technique allowed accurate quantitative characterization of complex origami shapes, and was capable of measuring tile parameters with sub-pixel resolution.

The success of my image processing strategy indicates that it might be fruitful to develop even more sophisticated image analysis programs - it is even possible to imagine 'intelligent' systems which could accept as input any images of DNA origami structures (or similar objects), and return as output a quantitative description of all

individual shapes, including an estimate of the yield of well-formed structures.

The methods I developed allowed me to conduct a detailed investigation into self-assembly of polymorphic tiles. The results were presented in Chapter 6, where I discussed 11 design variants and demonstrated that the folding pathway of these tiles could be changed dramatically by small changes to the staple set. The findings provide insights into the principles underlying origami assembly (summarized in Section 7.2).

In Appendix C, I explain my scheme for the application of DNA origami principles to the construction of a synthetic rotary motor. To the best of my knowledge, none of the rotary motors made so far using DNA have been capable of independent rotation, and all required external intervention in order to spin. In contrast, the device I have conceived would be autonomous, relying on sequential strand displacement reactions involving hairpins and hidden toeholds to accomplish rotation by stepping from one binding site to the next. Similar mechanisms were used to power the linear motors presented by Green *et al.* [48] and Omabegho *et al.* [168], which indicates that the principles underlying my new concept are sound. Further confirmation is provided by the simulations I performed, which suggested that my rotary motor would function as intended.

## 7.2 Insights into origami folding principles

It is possible to deduce from the experiments discussed in this thesis some of the key principles which govern origami folding. The observation that the polymorphic tile

folds with comparatively high yield suggests that pathways must exist for folding and that assembly is cooperative. This is underlined by the finding that staple incorporation temperatures differ less than might be expected based only on staple sequence. Furthermore, the successful synthesis of origami structures based on a dimer scaffold demonstrates that ultra-long scaffolds could be used for origami despite the presence of repeated sequences. Repetition of parts of some staple domains is probable for any scaffold longer than a few thousand bases, but the comparatively high yield of the polymorphic tile demonstrates that sequence duplication does not completely inhibit folding, and therefore this should not be a barrier to synthesis of larger origami tiles.

According to the arguments laid out in Chapter 4, if a staple has a choice of two alternative binding arrangements, the shortest distance principle will generally determine which is preferred. This is because two scaffold segments separated by a short run of bases are much more likely to be sufficiently close for a staple to bridge the gap between them. Staples cooperate because the binding of any staple can reduce the distance between the binding sites for other staples. This explains the finding of Sobczak *et al.*, who constructed an origami object consisting of two halves bridged only by scaffold links, occurring at the end of long stapled loops of scaffold [75]. They discovered that the two halves folded separately; this is consistent with the understanding of origami folding presented in this thesis, according to which there is no mechanism in such a structure for cooperation between the two halves.

My findings suggest that the folding pathway of an origami object also depends on

the order in which staples are incorporated and the reversibility of each binding event. A staple with a particularly strong sequence will tend to bind earlier in the folding process than a staple with a weak sequence, but the arrangement of neighbouring staples can affect the incorporation temperature. In particular, if a staple is paired (as in the seam) it will bind sooner than if it is not paired. Paired staples tend to reinforce each other, and the binding of a paired staple is quasi-irreversible if its partner is in place. Small modifications (involving for instance a single pair of staples) sometimes have no significant effect on folding, due to the strong cooperativity involved in the folding process.

Pairing staples to enhance their stability may prove to be a useful strategy for designing stronger origami objects. For some applications it may be desirable to ensure high structural integrity in one particular region, and this could be accomplished by extending or doubling up the staples linking together scaffold domains in that area. This trick may be of particular relevance for three-dimensional structures such as the rotary motor units shown in Appendix C.

My results also show that extending staples is an effective strategy for controlling the folding pathway. Longer staples will be more stable, binding earlier in folding, and these binding events can be significant in determining the subsequent folding pathway. The 'guiding role' of longer staples is consistent with the finding that a 14nt seed in staples can enhance the yield of well-folded structures [63]. If certain staples bind strongly at an early stage of folding, the subsequent assembly process is directed

towards the final state preferred by those staples.

In addition I found that the presence of single-stranded domains in a tile can lead to aggregation, attributable to weak inter-scaffold interactions.

### 7.2.1 Analogies with protein folding

As discussed in Chapter 1, an analogy exists between the hierarchical organization of protein structure and the levels of design of DNA nanostructures. In the case of DNA origami this is particularly strong, because the structure consists of a single scaffold strand which is folded up on itself by a set of helper strands, where the scaffold resembles a polypeptide chain. The argument I presented for the existence of folding pathways for the assembly of the polymorphic tile is identical to the reasoning used to resolve Levinthal's Paradox in protein folding. This states that the number of possible states is huge and if the system explored them all at random, the time taken to reach a final well-folded state would be so great that only misfolds would be observed.

I noted that breaking half the seam staples was analogous to the disruption of disulphide bridge formation in one of the samples studied in the Anfinsen Experiment. The paired seam staples in the origami tile make comparatively strong long-range connections, as do the cysteine-cysteine bonds, and when these connections are allowed to form first, a distribution of final structures is observed. If seam staple connections or disulphide bridges form late in folding, one structure is strongly preferred. This suggests a general principle - long-range contacts can be significant in

determining nanostructure assembly routes, but not when the connections are made late in folding.

The significance of the long-range connections made by the seam staples echoes the importance of non-local contacts in determining protein folding pathways. The proteins goat $\alpha$-lactalbumin and canine milk lysozyme apparently fold via different pathways despite having similar sequences and almost identical structures, and this has been attributed to the difference in the distribution of non-local contacts [169].

The 'Lower Right' design modification involved the use of a rigid linker in the lower right hand corner of the rectangle. The linker was formed by a long single-domain staple, designed to bind early in folding, constraining the subsequent folding pathway. The result was a strong bias in favour of the $5{:}0_{\perp}$ configuration; as discussed earlier, the mechanism for biasing assembly mirrors the effect of a proline residue in the hinge loop of a domain swap protein [166, 167].

According to the understanding of origami assembly set out in this thesis, the shortest distance principle governs the preference of a staple for one pair of binding sites over an alternative choice. The shortest distance principle has already been articulated in the context of proteins, and it is known that at an early stage of the folding process, interactions between nearby residues are usually more favourable than those between more distant units [170]. The native conformation of a polypeptide chain can be described in terms of its 'topological complexity' [171], which affects folding. The equivalent factor in an origami structure is the scaffold routing.

Furthermore, it has been shown that for some proteins it is possible to alter the folding pathway by modifying local structural elements to change their intrinsic stability [172]. As shown here, minor modifications to the staples used to fold the polymorphic tile have the same effect.

## 7.3 Conclusions

At the end of the first Chapter of this thesis, I stated that I would address a series of questions, to which I will now return, with the answers.

1. *Are there pathways by which DNA origami folds?*

   Yes. A surprisingly large number of polymorphic tiles fold well, which indicates that there must be folding pathways, and that assembly is not random.

2. *If so, what factors shape the folding pathways?*

   The key factors are the scaffold routing, the shortest-distance principle, co-operativity between staples, reversibility of staple binding and staple sequence. Paired staples co-operate strongly and can be particularly significant.

3. *Can we control the self-assembly process?*

   Yes. I presented several variations on the theme of the polymorphic tile, and demonstrated that comparatively small 'mutations' can lead to a large shift in the final distribution of shapes - which corresponds to a significant change in the folding pathway.

4. *To what extent does origami assembly mimic folding of proteins?*

   To a considerable extent. I have discussed analogies with Levinthal's Paradox, Anfinsen's Experiment and domain-swap proteins.

5. *How do staples co-operate during folding?*

   When one staple binds to the scaffold, it can reduce the distance between the binding sites of other staples, which should increase the rate at which they bind to those sites, in accordance with the shortest-distance principle.

### 7.3.1 Closing remarks

The double-helical structure of DNA has only been known for 60 years, and the first synthetic 3D DNA nanostructures were produced in the 1990s, but progress in DNA Nanotechnology has been remarkably rapid, with over 2700 papers published on 'DNA nanostructures' since 2000[1]. One of the most significant developments in the field was the invention of DNA origami in 2006, and it is a testament to the complexity of the origami folding process that only now, after several years, are we starting to comprehend in detail the principles underlying origami self-assembly. I hope that the work described in this thesis will make a significant contribution to our understanding, enabling the synthesis of even more novel structures and devices using DNA, such as the rotary motor I describe in Appendix C.

[1]Search performed using Thomson Reuters Web of Science on 10$^{th}$ May 2014. Topic=(DNA nanostructures), Timespan=2000-2014.

# Appendix A

# Materials and Methods

## A.1 Sources of materials, definitions

| **Materials** | **Supplier** |
|---|---|
| DNA strands | Integrated DNA Technologies (IDT). |
| Restriction enzymes, PCR Master Mix, T4 DNA ligase, plasmid pUC19, competent cells (DH5$\alpha$) | New England Biolabs. |
| Chemicals | Usually Sigma Aldrich. |
| Sephacryl S-300 high resolution gel filtration medium | VWR International. |
| Spin columns | Bio-Rad Laboratories. |
| AFM probes | Collaborators, Asylum Research or Bruker Ltd. |

Note - in protocols, NEBuffer refers to standard buffers supplied by New England Biolabs for use in enzyme digests. TAE denotes Tris-Acetate buffer with EDTA (40 mM Tris base, 20 mM acetate, 1mM EDTA, pH 8-8.3).

## A.2 DNA quantitation

Where necessary, the concentration of DNA in a sample was found from a measurement of light absorption, $A$, at a wavelength of 260 nm, using Beer's Law, $A = ebc$, where $e$ is the extinction coefficient for single-stranded or double-stranded DNA, as appropriate, and $b$ is the path length (here, 0.1 cm) [86]. A Cary Win UV spectrophotometer was used. The volume of sample used was 5 $\mu$L, and it was diluted to obtain a value for absorbance $A$ in the range $0.1 - 1$ (the range over which DNA can be accurately quantitated with this instrument).

## A.3 Gel electrophoresis

Generally, agarose gels were prepared using 1×TAE (pH8.3) with a concentration of 0.7-1% agarose, and run at 75-80 V for 45-60 minutes. The running buffer was 1×TAE (pH8.3). Gels were stained with SYBR Gold, in a 1×TAE buffer, for around 30 minutes. Finally gels were rinsed with ultrapure MilliQ water and imaged using a gel scanner (BioRad Pharos FX Plus Molecular Imager).

## A.4 Plasmid synthesis (monomer)

1 $\mu$g of pUC19 was incubated in a thermocycler for approximately 1 hour at 37°C with EcoRI in NEBuffer EcoRI, in a total volume of 50 $\mu$L. The enzyme was heat killed by incubation at 65°C for approximately 20 minutes and a buffer exchange was

performed using BIORAD p30 Biospin columns. 5 $\mu$L of 10× NEBuffer2 and 0.5 $\mu$L of HindIII were added and the mixture was incubated at 37°C for approximately 1 hour. The enzyme was inactivated by incubation for 20 minutes at 80°C and PCR was then performed on the cut plasmid with the following reaction (the PCR primers had previously been checked using polyacrylamide gel electrophoresis):

- 0.5 $\mu$L of a 10 $\mu$M stock of PCR primer 1, acquired from IDT[1] and resuspended in 1×TE
- 0.5 $\mu$L of a 10 $\mu$M stock of PCR primer 2, acquired from IDT[2] and resuspended in 1×TE
- 0.5 $\mu$L of 100-fold dilution (with MilliQ ultrapure water) of vector cut twice as described above
- 12.5 $\mu$L of One *Taq*® 2× Master Mix (Standard Reaction Buffer)
- 11 $\mu$L of MilliQ ultrapure water

The thermal cycle used was as follows:

- Hold at 94°C (30 seconds)
- 30 cycles: 94°C (30 seconds), 55°C (30 seconds), 68°C (3 minutes)
- Hold at 68°C (5 minutes)

[1] TGACCTAATCCTCAGCAATTCACTGGCCGTCGTTTTACAA (red: recognition site for BbvCI)

[2] ACGGACGCGCTGAGGAGCTTGGCGTAATCATGGTCATAG

Another buffer exchange was performed, and 2 $\mu$L of 10×T4 DNA ligase buffer was added, in addition to 1 $\mu$L of 1 M NaCl and 1 $\mu$L of BbvCI. The mixture was incubated at 37°C for approximately 1 hour and the enzyme was then inactivated by incubation for 20 minutes at 80°C. The cut PCR product was then ligated by incubation with 2 $\mu$L of T4 DNA ligase at 16°C for approximately 21 hours. The enzyme was later heat-inactivated by incubation for 20 minutes at 65°C.

The plasmid was sequenced by Source Life Biosciences, initially with standard M13 primers[3] and then with custom primers[4]. The scaffold sequence is given in Appendix A.13. Larger plasmid stocks were prepared as described in Appendix A.12.

The prepared plasmid was primarily in monomer form, but a small proportion had dimerized, and this was used to prepare a larger stock of pure dimer as described in Chapter 4.

## A.5 Enzyme digests

Enzyme digests were used to convert the double-stranded plasmid into a single-stranded form for use as an origami scaffold.

A volume of ultrapure MilliQ water was set up such that the total reaction volume for the nicking step was V. Next, V/10 of NEBuffer3 was added and the mixture was vortexed. (NEBuffer 3: 100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM Dithiothreitol, pH 7.9 at 25°C.)[5] Prepared plasmid was added to give a concentration

[3]Forward: TGTAAAACGACGGCCAGT, reverse: CAGGAAACAGCTATGACC [173]
[4]Forward: AGTGCACCATATGCGGTGTG; reverse: TCTCCCCGCGCGTTGGCCGA
[5]In 2013, the composition of the NEBuffers was revised - dithiothreitol was removed and BSA

at the start of the nicking step of 0.1 $\mu$g/$\mu$L, and the mixture was vortexed. V/20 of Nt.BspQI was added and the sample was mixed by pipetting[6]. The sample was incubated at 50°C for 1.5 hours, and then held at 80°C for 20 minutes to heat kill the enzyme. V/20 of ExoIII was added to the remaining solution, which was mixed by pipetting. The reaction was incubated at 37°C for 1.5 hours, then held at 80°C for 20 minutes to heat kill the enzyme. The DNA was purified by means of phenol:chloroform extraction (and ethanol precipitation where desired) - see below.

During the exonucleasing step, the total volume of the sample which is enzyme is V(reaction)/10 and this is a maximum limit; the enzyme is provided in a solution which contains glycerol and at higher concentrations glycerol can adversely affect the performance of the enzyme during the digest [174].

## A.6 Phenol/chloroform extraction

The purpose of this procedure was to isolate the DNA from a solution containing proteins and other contaminants [87]. It was carried out in a fume cupboard, with appropriate safety precautions taken. V $\mu$L of phenol was added to the sample, which was then vortexed and centrifuged for one minute at maximum speed (13.4 krpm) in tabletop centrifuge. The upper phase was then extracted and the process repeated with chloroform:isoamyl alcohol.

was added, to a concentration of 100 $\mu$g/mL. The reformulated buffer was named NEBuffer 3.1.

[6]With the pipette set to around half the total sample volume, the liquid was pipetted up and down 20 times or more.

## A.7 Ethanol precipitation of DNA

Ethanol precipitation [87] was used in some cases to purify DNA after a phenol/chloroform extraction. Ethanol was chilled at -20°C prior to use. (V/10) $\mu$L of sodium acetate (3M, pH 5.2) was added to the sample, followed by 2V $\mu$L of chilled 100% ethanol. The mixture was vortexed and placed in a freezer at -20°C for approximately 45 minutes before being spun for 10 minutes in a tabletop centrifuge at maximum speed (13.4 krpm). The supernatant was extracted and the pellet (usually barely visible) was left to dry. The DNA was then re-suspended in the desired buffer; heating briefly to 80°C in a thermocycler was found to assist, increasing final yield.

## A.8 Origami synthesis

The origami scaffold was derived from a customized plasmid as described in Appendix A.5. Origami was synthesized according to the standard protocol by annealing in a thermocycler at a rate of 1°C/min from 95°C to 25°C. The usual synthesis buffer was 1×Tris-Acetate (pH 8.3) with 12.5 mM magnesium acetate, and the staples were present in excess, at a concentration of 4-20 times the scaffold concentration. Before AFM imaging, the sample was purified as described below to remove excess staples.

## A.9 Origami purification

Origami was purified[7] using Sephacryl S-300 high resolution gel-filtration medium, which is supplied in ethanol. Before use it was prepared with an alternative buffer as follows.

25 mL of S-300 solution (in ethanol) was mixed with an equal volume of ultrapure MilliQ water and centrifuged at 1000g for 1 minute. The supernatant was decanted and the contents of the tube were topped up with MilliQ to a total volume of around 50 mL; the solution was mixed thoroughly and centrifuged again at 1000g for one minute. This step (decant/top up/mix/centrifuge) was performed three times with MilliQ, and then 5 times with buffer (NEBuffer2 analog - 50 mM NaCl, 10 mM $MgCl_2$, 1×TE). The tube was placed overnight on a shaker for mixing. S-300 was stored at 4°C both before and after the buffer change, and when smaller volumes of S-300/buffer were required the quantities given here were scaled accordingly.

For each origami sample to be purified, three spin columns (Micro Bio-spin Chromatography columns, Bio-Rad) were prepared. 500 $\mu$L of S-300/NEBuffer-2 was added to each, before centrifugation in a tabletop centrifuge at 3.8 krpm for 3 minutes. If 1.5 mL Eppendorf tubes were used to collect the flow-through, an additional 1 minute of centrifugation was required after discard of the flow-through from the first step. After disposal of the flow-through, another 500 $\mu$L of S-300/NEBuffer-2 was added to each spin column and the centrifugation step repeated. Finally the

[7]Method introduced in Supplementary Information of Ref. [50]. The description given here is based on a protocol provided by Robert Machinek, who had inherited it from the first author of Ref. [50].

origami sample was applied to a column and centrifuged at 3.8 krpm for 4 minutes; the flow-through was applied to the next column and centrifuged; this was repeated for the final column. The sample was imaged as soon as possible after purification.

As a precaution S-300/NEBuffer-2 was kept in storage for no longer than several weeks before use; S-300 in ethanol was kept indefinitely. Care was taken not to violently agitate the S-300 medium before use.

## A.10 RT-PCR machine protocols

Origami assembly and disassembly reactions were monitored using an Agilent Technologies Stratagene MX3005P RT-PCR machine, set to ramp from the initial temperature (95°C or 25°C) to the final temperature (25°C or 95°C) in steps such that the overall cooling/heating rate was 1°C/min. For each sample, duplicate measurements were made and averaged, over 3 or more repeats (normally 6). In these experiments the scaffold concentration was 20 nM and the staple concentration was 200 nM.

Data (Fluorescence, $F$, as a function of temperature, $T$) was averaged over each data collection cycle. A data collection cycle consisted of a set of datapoints acquired within a particular time frame during which the temperature was nominally constant. The number of points per cycle depends on the number of wells being used and is limited by the machine itself. All data points were recorded and smoothing was not required. The derivative $dF/dT$ was calculated and the data was fitted with a Gaussian, the centre of which was taken to be the transition temperature. Experiments

were repeated and averages taken (over all good fits) of transition temperatures or differences in transition temperatures. Data was processed with MATLAB.

At an early stage of the work described in this thesis, I attempted to study origami assembly using a fluorimeter, with a setup in which two staples were labelled with fluorophores and adjacent staples carried quenchers[8]. Ultimately, the RT-PCR machine proved to be more appropriate for fluorescence measurements on assembling origami because it has a more sophisticated temperature control system and can monitor up to 96 samples, each being tens of $\mu$L in volume. In contrast the fluorimeter can only hold a maximum of four samples, each being hundreds of $\mu$L in volume, and the apparatus is less well-suited to conducting measurements at well-defined temperatures and heating/cooling rates.

## A.11 AFM

AFM images were acquired using either an Agilent 5500 or a Veeco Multimode 8 at the Centre for DNA Nanotechnology, Aarhus University, Denmark, or a Veeco Dimension 3100 at the University of Oxford, UK. The AFM probes used were of the Olympus TR400-PSA variety (smaller cantilever) or the Bruker SNL type. A volume of a few $\mu$L of purified origami was applied to freshly cleaved mica and this specimen was imaged in tapping mode in liquid. The imaging buffer used was normally 1× Tris-

[8]The execution of these experiments - and the first of the experiments with the RT-PCR machine - pre-dated the publication of the paper by Wei *et al.* in which they described similar work using a FRET pair rather than a quencher [90].

Acetate-EDTA or 1×Tris-Acetate (pH 8-8.3), with 12.4-12.5 mM magnesium acetate and 4 mM nickel chloride.

## A.12 Plasmids - cloning and preparation

Plasmids were transfected into *E. coli* (DH5$\alpha$) competent cells[9], which had been stored at -80°C. For transfection they were defrosted in ice, and used only when clearly in a liquid state. 1 $\mu$L of the plasmid to be transformed was added and the vial gently agitated (mixing by pipetting is not advisable). The bacteria were left on ice for approximately 30 minutes, and then warmed at 40°C in a water bath for approximately 2 minutes, before being placed on ice for at least another 2 minutes. The transfected plasmid carried the ampicillin-resistance gene and hence bacteria were decanted onto an agar plate containing the antibiotic ampicillin. Typically a series of plates was prepared, having a range of bacteria concentrations. The volume of liquid (bacteria diluted with LB/ampicillin growth medium) to be spread across the plate was usually around 100 $\mu$L. Beads were used to spread the mixture across the surface evenly. Plates were incubated overnight at 37°C, left inverted to avoid the formation of condensation on the agar jelly. Growth was not allowed to proceed for longer than approximately 16 hours because after that time satellite colonies would start to form as the cells began to secrete the enzyme conferring antibiotic resistance.

A sterile loop was used to select an individual colony, for initiation of a starter

[9]Non-pathogenic strain. Cloning protocol: as instructed by Dr Carlos Sanchez-Cano.

culture in a tube containing a few mL of LB growth medium laced with ampicillin. The starter culture was normally incubated at 37°C for 6-8 hours, and then used to inoculate a larger culture, typically around 500 mL LB/ampicillin medium. Glassware and unwanted bacteria cultures were treated with the disinfectant Virkon.

Plasmids were extracted from the bacterial cultures using commercially available kits (QIAGEN plasmid purification kits or QIAFilter kits). For larger scale preparation, maximum yield was assured by using conical-bottom centrifuge tubes at lower speeds[10], which had been found to yield a more robust pellet. Yield was estimated from measurements of light absorbance at 260 nm. Where appropriate, a set of diagnostic enzyme reactions was performed with the purified plasmid, to confirm formation of the desired product. The results were analysed using agarose gel electrophoresis.

[10]These tubes are less robust than the recommended variety and they will not withstand the centrifugal force suggested in the protocol from QIAGEN.

## A.13 DNA sequences

### A.13.1 Scaffold sequence (monomer)

2646 bases.

GACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTA
ATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCA
CTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTAT
TTTTCTAAATACATTCAAATATGTATCCGCTCATGAGAC
AATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGA
AGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTC
CCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACC
CAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGT
TGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACA
GCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTT
TTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCG
CGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCG
GTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGT
ACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGA
CAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTG
ATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAG
GACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGG
ATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGA
ATGAAGCCATACCAAACGACGAGCGTGACACCACGATGC
CTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTG
GCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAG
ACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGC
GCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAAT
CTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAG
CACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTA
TCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAA
ATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGC
ATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTT
AGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCT
AGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCC
CTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCG
TAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTC
TGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGC
TACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAA
CTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGA
TACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCC
ACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCG
CTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCG
ATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGT
TACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTT
CGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCG

```
AACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCA
CGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAA
GCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTC
CAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGT
TTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCT
CGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACG
CGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTG
CTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGG
ATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTC
GCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCG
AGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCC
CCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACA
GGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAA
TTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTT
TACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTG
TGAGCGGATAACAATTTCACACAGGAAACAGCTATGACC
ATGATTACGCCAAGCTCCTCAGCAATTCACTGGCCGTCG
TTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCC
AACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCT
GGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCC
AACAGTTGCGCAGCCTGAATGGCGAATGGCGCCTGATGC
GGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACC
GCATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCG
CATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTG
ACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTT
ACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTC
AGAGGTTTTCACCGTCATCACCGAAACGCGCGA
```

### A.13.2 Staple sequences - tables

The start/end position of each staple is labelled with the notation $h[b]$, where $h$ is the helix number (1-12) and $b$ is the base number (0-223). The 'start' of the staple is its 5' terminus, the 'end' of the staple is the 3' terminus. Sequences were generated using caDNAno [64]. Staples at the edges of the tiles were extended with 4 thymines to suppress inter-tile stacking, as in previous work [5]. Section A.14 explains which staples were used in each design variant.

| Start | End | Sequence |
|---|---|---|
| 1[0] | 2[0] | TTTTTGAATACTCATACTCTGGCGACACGGAAATGTTTTT |
| 1[16] | 1[31] | TCCTTTTTCAATATT |
| 1[32] | 2[32] | ATTGAAGCATTTATCAGAAGGCAAAATGCCG |
| 1[48] | 1[63] | GGGTTATTGTCTCATG |
| 1[64] | 2[64] | AGCGGATACATATTTGTTTCACCAGCGTTTCT |
| 1[80] | 1[95] | AATGTATTTAGAAAA |
| 1[96] | 1[127] | ATAAACAAATAGGGGTTCCGCGCACATTTCCC |
| 1[128] | 1[143] | CGAAAAGTGCCACCTG |
| 1[160] | 1[175] | TTATTATCATGACATT |
| 1[192] | 1[207] | CGTATCACGAGGCCCT |
| 2[127] | 2[96] | GGCGCGTCAGCGGGTGGTGCACCCAACTGAT |
| 2[143] | 3[143] | CAGACAAGCCCGTCAGGCTGGCTTAACTATGC |
| 2[159] | 1[159] | AAGCGGATGCCGGGAGACGTCTAAGAAACCA |
| 2[175] | 3[175] | TCACAGCTTGTCTGTGTACTGAGAGTGCACC |
| 2[191] | 1[191] | CAGCTCCCGGAGACGGAACCTATAAAAATAGG |
| 2[207] | 3[207] | AAACCTCTGACACATGTACCGCACAGATGCGT |
| 2[223] | 1[223] | TTTTTCGGTGATGACGGTGATTCGTCTCGCGCGTTTTTT |
| 3[0] | 4[0] | TTTTCCGCGCCACATAGCAGCAATACGGGATAATATTTT |
| 3[16] | 2[16] | AACTTTAAAAGTGCTCCAAAAAAGGGAATAAG |
| 3[32] | 4[32] | ATCATTGGAAAACGTTTATGCGGCGACCGAGT |
| 3[48] | 2[48] | CTTCGGGGCGAAAACGGGTGAGCAAAAACAG |
| 3[64] | 4[64] | TCTCAAGGATCTTACCAGTACTCAACCAAGT |
| 3[80] | 2[80] | GCTGTTGAGATCCAGTCTTCAGCATCTTTTAC |
| 3[96] | 3[127] | TCGATGTAACCCACTCTTGGCGGGTGTCGGG |
| 4[127] | 4[96] | GTGCTGCAAGGCGATTGCCATCCGTAAGATGC |
| 4[143] | 5[143] | TGGCGAAAGGGGGATGGGTTTTCCCAGTCAC |
| 4[159] | 3[159] | TCGCTATTACGCCAGCGGCATCAGAGCAGATT |
| 4[175] | 5[175] | ATCGGTGCGGGCCTCTCGGCCAGTGAATTGCT |
| 4[191] | 3[191] | ACTGTTGGGAAGGGCGATATGCGGTGTGAAA |
| 4[207] | 5[207] | CATTCAGGCTGCGCATCATGGTCATAGCTGT |
| 4[223] | 3[223] | TTTTATCAGGCGCCATTCGCAAGGAGAAAATACCGCTTTT |
| 5[0] | 6[0] | TTTTTGCAAAAAAGCGGTTAATGATCCCCCATGTTGTTTT |
| 5[16] | 4[16] | GCTCCTTCGGTCCTCTGCTCTTGCCCGGCGT |
| 5[32] | 6[32] | CGATCGTTGTCAGAAGGCTCCGGTTCCCAAC |
| 5[48] | 4[48] | TAAGTTGGCCGCAGTGCATTCTGAGAATAGTG |
| 5[64] | 6[64] | TTATCACTCATGGTTAGTCACGCTCGTCGTTT |
| 5[80] | 4[80] | TGGCAGCACTGCATATTTTCTGTGACTGGTG |
| 5[96] | 5[127] | ATTCTCTTACTGTCATAAGTTGGGTAACGCCA |
| 6[127] | 6[96] | TTGCGCTCACTGCCCGACGTTGTTGCCATTG |
| 6[143] | 7[143] | ACTCACATTAATTGCGACCTGTCGTGCCAGCT |
| 6[159] | 5[159] | CCTAATGAGTGAGCTAGACGTTGTAAAACGA |
| 6[175] | 7[175] | GTAAAGCCTGGGGTGCAACGCGCGGGGAGAG |
| 6[191] | 5[191] | GCCGGAAGCATAAAGTGAGGAGCTTGGCGTAA |
| 6[207] | 7[207] | TCCACACAACATACGAGGCGCTCTTCCGCTTC |
| 6[223] | 5[223] | TTTTGTTATCCGCTCACAATTTCCTGTGTGAAATTTTTT |

Table A.1: Original design (1) - list continued in Table A.2.

| Start | End | Sequence |
|---|---|---|
| 7[0] | 8[0] | TTTTCCGGAAGGGCCGAGCGCAATAAACCAGCCAGTTTT |
| 7[16] | 6[16] | CAGAAGTGGTCCTGCAGATCAAGGCGAGTTAC |
| 7[32] | 8[32] | ACTTTATCCGCCTCCAAGACCCACGCTCACCG |
| 7[48] | 6[48] | TCCAGTCTATTAATTGGTATGGCTTCATTCA |
| 7[64] | 8[64] | GTTGCCGGGAAGCTAGCATCTGGCCCCAGTG |
| 7[80] | 6[80] | AGTAAGTAGTTCGCCACTACAGGCATCGTGGT |
| 7[96] | 7[127] | GTTAATAGTTTGCGCACTTTCCAGTCGGGAA |
| 8[127] | 8[96] | ACATGTGAGCAAAAGGCGTGTAGATAACTACG |
| 8[143] | 9[143] | ATAACGCAGGAAAGAAACCGTAAAAAGGCCG |
| 8[159] | 7[159] | TCCACAGAATCAGGGGGCATTAATGAATCGGC |
| 8[175] | 9[175] | GGCGGTAATACGGTTATCCATAGGCTCCGCCC |
| 8[191] | 7[191] | TATCAGCTCACTCAAAGCGGTTTGCGTATTG |
| 8[207] | 9[207] | GGCTGCGGCGAGCGGAAAAATCGACGCTCAA |
| 8[223] | 7[223] | TTTTCTGCGCTCGGTCGTTCCTCGCTCACTGACTCGTTTT |
| 9[0] | 10[0] | TTTTTAAAGTATATATGAGTAAGTTTTAAATCAATCTTTT |
| 9[16] | 8[16] | AAACTTGGTCTGACAGCTCCAGATTTATCAG |
| 9[32] | 10[32] | GTTACCAATGCTTAATTCTTCACCTAGATCC |
| 9[48] | 8[48] | CAGTGAGGCACCTATCCTGCAATGATACCGCG |
| 9[64] | 10[64] | TCAGCGATCTGTCTATAGGGATTTTGGTCATG |
| 9[80] | 8[80] | TTCGTTCATCCATAGATACGGGAGGGCTTAC |
| 9[96] | 9[127] | TTGCCTGACTCCCCGTCCAGCAAAAGGCCAGG |
| 10[127] | 10[96] | GCCTTTCTCCCTTCGGCTGACGCTCAGTGGA |
| 10[143] | 11[143] | TACCGGATACCTGTCCCTCATAGCTCACGCTG |
| 10[159] | 9[159] | TTCCGACCCTGCCGCTCGTTGCTGGCGTTTT |
| 10[175] | 11[175] | TCGTGCGCTCTCCTGGTGTAGGTCGTTCGCT |
| 10[191] | 9[191] | CCCCCTGGAAGCTCCCCCCTGACGAGCATCAC |
| 10[207] | 11[207] | AAGATACCAGGCGTTTTGCACGAACCCCCCGT |
| 10[223] | 9[223] | TTTTACCCGACAGGACTATAGTCAGAGGTGGCGAATTTT |
| 11[0] | 12[0] | TTTTGATCCGGCAAACAAACGAGTTGGTAGCTCTTTTTT |
| 11[16] | 10[16] | CACCGCTGGTAGCGGTTTTTAAATTAAAAATG |
| 11[32] | 12[32] | GGTTTTTTTGTTTGCACGCTCTGCTGAAGCCA |
| 11[48] | 10[48] | AGCAGCAGATTACGCAGATTATCAAAAAGGA |
| 11[64] | 12[64] | GCAGAAAAAAAGGATCGCTACACTAGAAGAA |
| 11[80] | 10[80] | TCAAGAAGATCCTTTGACGAAAACTCACGTTA |
| 11[96] | 11[127] | ATCTTTTCTACGGGGTGAAGCGTGGCGCTTT |
| 12[31] | 12[16] | GTTACCTTCGGAAAAA |
| 12[63] | 12[48] | CAGTATTTGGTATCTG |
| 12[95] | 12[80] | TGGTGGCCTAACTACG |
| 12[127] | 12[96] | GGTATGTAGGCGGTGCTACAGAGTTCTTGAAG |
| 12[143] | 12[128] | GGATTAGCAGAGCGA |
| 12[159] | 11[159] | GCAGCCACTGGTAACATAGGTATCTCAGTTCG |
| 12[175] | 12[160] | CTTATCGCCACTGGCA |
| 12[191] | 11[191] | ACCCGGTAAGACACGACCAAGCTGGGCTGTG |
| 12[207] | 12[192] | ATCGTCTTGAGTCCA |
| 12[223] | 11[223] | TTTTGCCTTATCCGGTAACTTCAGCCCGACCGCTGCTTTT |

Table A.2: Original design (2)

| Start | End | Sequence |
|---|---|---|
| 2[127] | 2[111] | GGCGCGTCAGCGGGTG |
| 2[110] | 2[96] | GTGCACCCAACTGAT |
| 3[96] | 3[110] | TCGATGTAACCCACTC |
| 3[111] | 3[127] | TTGGCGGGTGTCGGG |
| 4[127] | 4[111] | GTGCTGCAAGGCGATT |
| 4[110] | 4[96] | GCCATCCGTAAGATGC |
| 5[96] | 5[110] | ATTCTCTTACTGTCAT |
| 5[111] | 5[127] | AAGTTGGGTAACGCCA |
| 6[127] | 6[111] | TTGCGCTCACTGCCCG |
| 6[110] | 6[96] | ACGTTGTTGCCATTG |
| 7[96] | 7[110] | GTTAATAGTTTGCGCA |
| 7[111] | 7[127] | CTTTCCAGTCGGGAA |
| 8[127] | 8[111] | ACATGTGAGCAAAAGG |
| 8[110] | 8[96] | CGTGTAGATAACTACG |
| 9[96] | 9[110] | TTGCCTGACTCCCCGT |
| 9[111] | 9[127] | CCAGCAAAAGGCCAGG |
| 10[127] | 10[111] | GCCTTTCTCCCTTCGG |
| 10[110] | 10[96] | CTGACGCTCAGTGGA |
| 11[96] | 11[110] | ATCTTTTCTACGGGGT |
| 11[111] | 11[127] | GAAGCGTGGCGCTTT |

Table A.3: Broken seams

| Name | Sequence |
|---|---|
| AS_1 | GTCAGAGGTGGCGAAGTTATCCGCTCACAAT |
| AS_2 | TTCCTGTGTGAAATTACCCGACAGGACTATA |
| AS_3 | CTCGCTCACTGACTCGCTGCGCTCGGTCGTTC |

Table A.4: Alternative seam design - extra staples

***Case 1***

| Start | End | Sequence |
|---|---|---|
| 7[48] | 6[48] | /5Cy5/TCCAGTCTATTAATTGGTATGGCTTCATTCA/3BHQ_2/ |
| 6[127] | 6[96] | /5Cy3/TTGCGCTCACTGCCCGACGTTGTTGCCATTG/3BHQ_2/ |

***Case 2***

| Start | End | Sequence |
|---|---|---|
| 4[175] | 5[175] | /5Cy5/ATCGGTGCGGGCCTCTCGGCCAGTGAATTGCT/3BHQ_2/ |
| 5[96] | 5[127] | /5Cy3/ATTCTCTTACTGTCATAAGTTGGGTAACGCCA/3BHQ_2/ |

Table A.5: Labelled staples

| Name | Sequence |
|---|---|
| EU_1 | GGGTTATTGTCTCATGAGCGGATACATATTTGTTTCACCAGCGTTTCTGGGTGAGCAAAAACAG |
| EU_2 | CTTCGGGGCGAAAACTCTCAAGGATCTTACCAGTACTCAACCAAGTCATTCTGAGAATAGTG |
| EU_3 | CAGTGAGGCACCTATCTCAGCGATCTGTCTATAGGGATTTTGGTCATGAGATTATCAAAAAGGA |
| EU_4 | AGCAGCAGATTACGCGCAGAAAAAAAGGATCGCTACACTAGAAGAACAGTATTTGGTATCTG |
| EU_5 | TCACAGCTTGTCTGTAAGCGGATGCCGGGAGACGTCTAAGAAACCATTATTATCATGACATT |
| EU_6 | ATCGGTGCGGGCCTCTTCGCTATTACGCCAGCGGCATCAGAGCAGATTGTACTGAGAGTGCACC |
| EU_7 | TCGTGCGCTCTCCTGTTCCGACCCTGCCGCTCGTTGCTGGCGTTTTTCCATAGGCTCCGCCC |
| EU_8 | CTTATCGCCACTGGCAGCAGCCACTGGTAACATAGGTATCTCAGTTCGGTGTAGGTCGTTCGCT |
| EU_9 | TAAGTTGGCCGCAGTG |
| EU_10 | CTGCAATGATACCGCG |
| EU_11 | CGGCCAGTGAATTGCT |
| EU_12 | GGCGGTAATACGGTTA |

Table A.6: EU design - extra staples

| Name | Sequence |
|---|---|
| 8A | TTTTTGAATACT |
| 8B | GGAAATGTTTTT |
| 8C | GCGCGTTTTTT |
| 8D | TTTTTCGGTGAT |
| 8E | GCTGTTGA |
| 8F | AACTATGC |
| 8G | TCACGTTA |
| 8H | TACCGGAT |
| 8J | AGCTCTTTTTT |
| 8K | ACCGCTGCTTTT |
| 8L | TTTTGCCTTATC |
| 16A | GCTCCTTCGGTCCTC |
| 16B | TCATGGTCATAGCTGT |
| 16C | GCTCCAGATTTATCAG |
| 16D | GGCTGCGGCGAGCGG |
| 48A | GGGTTATTGTCTCATGAGCGGATACATATTTGTTTCACCAGCGTTTCT |
| 48B | AAGCGGATGCCGGGAGACGTCTAAGAAACCATTATTATCATGACATT |
| 48C | GATCCAGTTCGATGTAACCCACTCTTGGCGGGTGTCGGGGCTGGCTT |
| 48D | ACCTGTCCGCCTTTCTCCCTTCGGCTGACGCTCAGTGGAACGAAAAC |
| 48E | GCAGAAAAAAAGGATCGCTACACTAGAAGAACAGTATTTGGTATCTG |
| 48F | CTTATCGCCACTGGCAGCAGCCACTGGTAACATAGGTATCTCAGTTCG |
| 64A | GTGCACCCAACTGATCTTCAGCATCTTTTACAATGTATTTAGAAAAATAAACAAATAGGGGT |
| 64B | TCCGCGCACATTTCCCCGAAAAGTGCCACCTGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTG |
| 64C | AACTTTAAAAGTGCTCATCATTGGAAAACGTTTATGCGGCGACCGAGTTGCTCTTGCCCGGCGT |
| 64D | CATTCAGGCTGCGCAACTGTTGGGAAGGGCGATATGCGGTGTGAAATACCGCACAGATGCGT |
| 64E | AAACTTGGTCTGACAGTTACCAATGCTTAATTCTTCACCTAGATCCTTTTAAATTAAAAATG |
| 64F | AAGATACCAGGCGTTTCCCCCTGGAAGCTCCCCCCTGACGAGCATCACAAAAATCGACGCTCAA |
| 64G | TACAGAGTTCTTGAAGTGGTGGCCTAACTACGTCAAGAAGATCCTTTGATCTTTTCTACGGGGT |
| 64H | GAAGCGTGGCGCTTTCTCATAGCTCACGCTGGGATTAGCAGAGCGAGGTATGTAGGCGGTGC |
| 80A | CATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACAC |
| 80B | GACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTC |
| 80C | AAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCACGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGT |
| 80D | CGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCG |

Table A.7: VEU - list of extra staples

| ***Hairpin*** | | |
|---|---|---|
| Start | End | Sequence |
| 9[32] | 10[32] | GTTACCAATGCTTAATtcctcttttgaggaacaagttttcttgtTCTTCACCTAGATCC |
| ***F_0*** | | |
| Start | End | Sequence |
| 1[0] | 2'[0] | TGAATACTCATACTCTGGCGACACGGAAATGTTGAATACTCATACTCTGGCGACACGGAAATGT |
| ***LR_0*** | | |
| Start | End | Sequence |
| 12[192] | 11'[32] | TGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCA |

Hairpin replaced 9[32]-10[32] in HSO, HSB, FSPB and TSPB samples. Lower case letters in the Hairpin staple indicate the region which forms the hairpin itself (sequence of this region as used in Ref. [5]). In the staple F_0, the 'prime' beside the 2 indicates that this is the $2^{\text{nd}}$ helix in the *other* rectangle.

Table A.8: Other staples

## A.14 Design variants: details

The design variants of the polymorphic tile are shown diagrammatically at the end of this section. In each case the tile is depicted in the configuration expected to be observed most frequently. Details of the specimens studied are provided in Table A.9.

The staple sets used for each design variant are listed explicitly on the following pages.

| **Name** | **Label** | **[Scaffold] (nM)** | **Staple excess** | **Standard synthesis?** |
|---|---|---|---|---|
| Original | O | ∼10 | ∼8 | ✓ |
| Half Seam Omitted | HSO | ∼10 | ∼8 | ✓ |
| Half Seam Broken | HSB | ∼10 | ∼8 | ✓ |
| Alternative Seam | AS | ∼3.8 | ∼11 | ✓ |
| Failed Forcing | F | ∼12.5 | ∼8* | X |
| Lower Right modification | LR | ∼3.8 | ∼11 | ✓ |
| First Seam Pair Broken | FSPB | ∼10 | ∼8 | ✓ |
| Third Seam Pair Broken | TSPB | ∼10 | ∼8 | ✓ |
| Staple Concentration variation | SC | ∼10 | various | ✓ |
| Extended U-staples | EU | ∼10 | ∼8 | ✓ |
| Very Extended U-staples | VEU | ∼10 | ∼8 | ✓ |

Table A.9: Variants on design of polymorphic tile. All samples except O and F were imaged with a Veeco Dimension 3100 AFM in Oxford; O and F were imaged in Aarhus with an Agilent 5500 and a Veeco Multimode 8 respectively. Staple excess is defined as staple concentration divided by scaffold concentration. Standard synthesis: annealing from 95°C to 25°C at rate of 1°C/min, in buffer of 1×TA, pH 8.3, with 12.5 mM magnesium acetate, scaffold prepared by enzyme digest of plasmid followed by phenol:chloroform extraction, ethanol precipitation and resuspension in synthesis buffer (at which stage scaffold concentration was estimated by measurement of light absorbed at wavelength of 260 nm). Origami tiles were purified before imaging using procedure described in Appendix A.9, with three cycles of purification. *Variant F was prepared in 0.4×NEBuffer3 (no ethanol precipitation step) with a 40-fold excess of the extra-long forcing staple and only two cycles of purification were used.

For fluorescence experiments, dual-labelled staples (Table A.5) replaced the corresponding unlabelled staples.

### A.14.1 HSO

All even numbered seam staples omitted. Hairpin replaced 9[32]-10[32].

### A.14.2 HSB

- AFM experiment and Case 1 of fluorescence experiments: All even numbered seam staples omitted. Broken counterparts included.
- Fluorescence experiments, Case 2. Even numbered seams: intact. Odd ones: broken.
- For AFM experiment: Hairpin replaced 9[32]-10[32].
- For fluorescence experiments: no hairpin used.
- All staples not mentioned: as original design.

### A.14.3 AS

- All odd numbered seam staples omitted. Broken counterparts included.
- 6[223]-5[223], 8[223]-7[223], and 10[223]-9[223] omitted.
- Included: AS_1, AS_2, AS_3.
- All staples not mentioned: as original design.

### A.14.4 F

- 1[0]-2[0] omitted.
- Included: F_0
- All staples not mentioned: as original design.

### A.14.5 LR

- Omitted: 10[207]-11[207], 12[223]-11[223], 12[207]-12[192]
- Included: 10[207]-10[192] (half of 10[207]-11[207]), LR_0
- All staples not mentioned: as original design.

### A.14.6 FSPB, TSPB

- Omitted: 10[127]-10[96] and 11[96]-11[127] for FSPB, 6[127]-6[96] and 7[96]-7[127] for TSPB. Broken counterparts included.
- All staples not mentioned: as original design.

### A.14.7 SC

- All staples: as original design. Present at various concentrations (Fig. A.2).

### A.14.8 EU

- Omitted: 1[48]-1[63], 1[64]-2[64], 1[160]-1[175], 2[159]-1[159], 2[175]-3[175], 3[48]-2[48], 3[64]-4[64], 4[159]-3[159], 4[175]-5[175], 5[48]-4[48], 7[96]-7[127], 8[175]-9[175], 9[48]-8[48], 9[64]-10[64], 10[159]-9[159], 10[175]-11[175], 11[48]-10[48], 11[64]-12[64], 12[63]-12[48], 12[159]-11[159], 12[175]-12[160].

- Included: all staples in table marked 'EU', both halves of broken seam 7.

- All staples not mentioned: as original design.

### A.14.9 VEU

- Omitted: 1[0]-2[0], 1[16]-1[31], 1[32]-2[32], 1[48]-1[63], 1[64]-2[64], 1[80]-1[95], 1[96]-1[127], 1[128]-1[143], 1[160]-1[175], 1[192]-1[207], 2[127]-2[96], 2[143]-3[143], 2[159]-1[159], 2[191]-1[191], 2[207]-3[207], 2[223]-1[223], 3[16]-2[16], 3[32]-4[32], 3[80]-2[80], 3[96]-3[127], 4[191]-3[191], 4[207]-5[207], 5[16]-4[16], 7[96]-7[127], 8[207]-9[207], 9[16]-8[16], 9[32]-10[32], 10[127]-10[96], 10[143]-11[143], 10[191]-9[191], 10[207]-11[207], 11[0]-12[0], 11[16]-10[16], 11[32]-12[32], 11[64]-12[64], 11[80]-10[80], 11[96]-11[127], 12[31]-12[16], 12[63]-12[48], 12[95]-12[80], 12[127]-12[96], 12[143]-12[128], 12[159]-11[159], 12[175]-12[160], 12[191]-11[191], 12[207]-12[192], 12[223]-11[223]

- Included: all staples in table marked 'VEU', both halves of broken seam 7.

- All staples not mentioned: as original design.

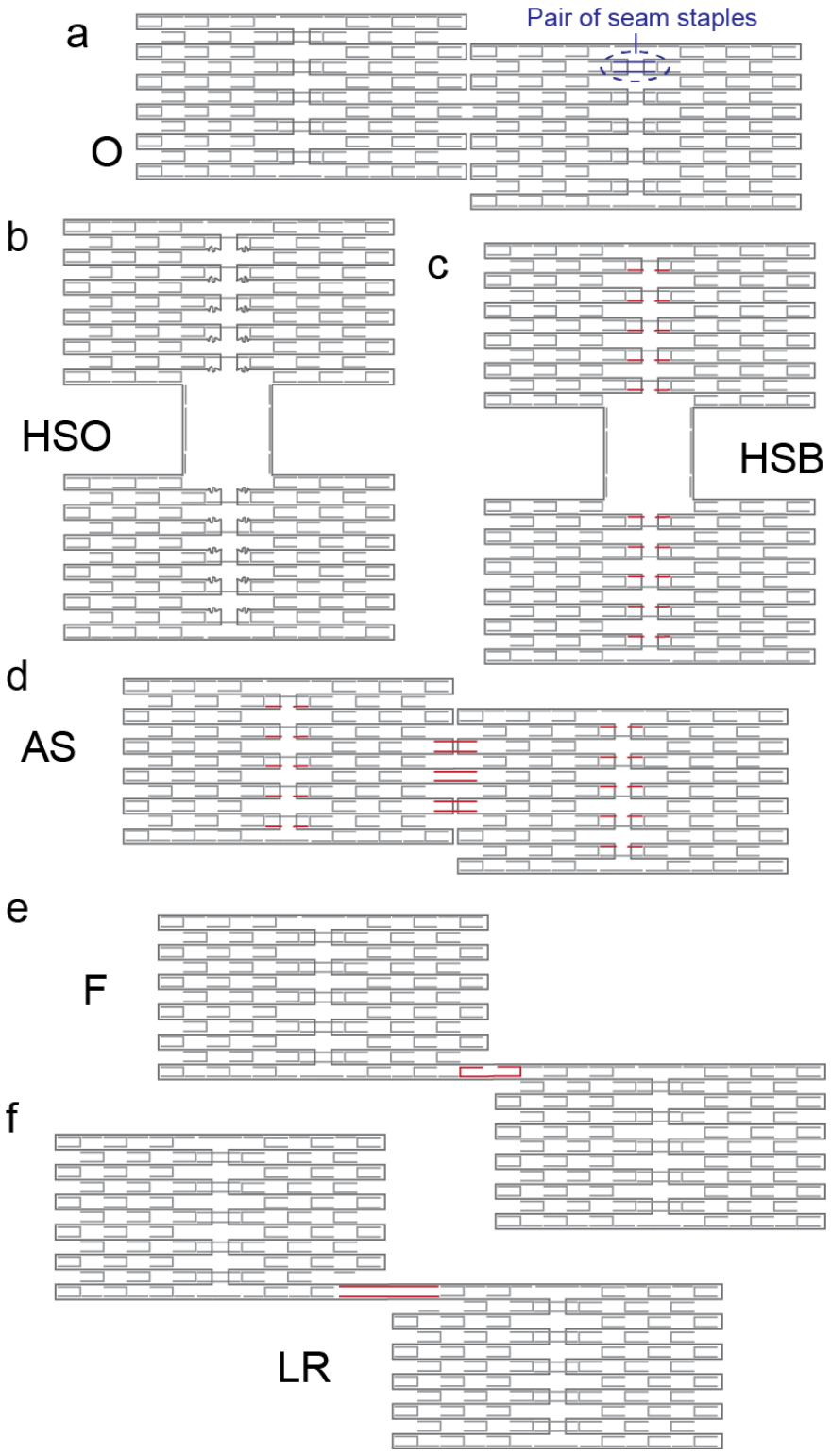


Figure A.1: Staple layouts for variants of the polymorphic tile (O, HSB, HSO, AS, F, LR), depicted for one configuration in each case. Design modifications are marked in red. (a) Original design, one pair of seam staples highlighted in blue to illustrate definition. (b)-(f) Modified designs. (b) Half Seam Omitted (HSO). (c) Half Seam Broken (HSB). (d) Alternative Seam (AS). (e) Failed attempt to Force tile to fold into particular configuration (F). (f) Modification in the Lower Right corner (LR) of the rectangle.

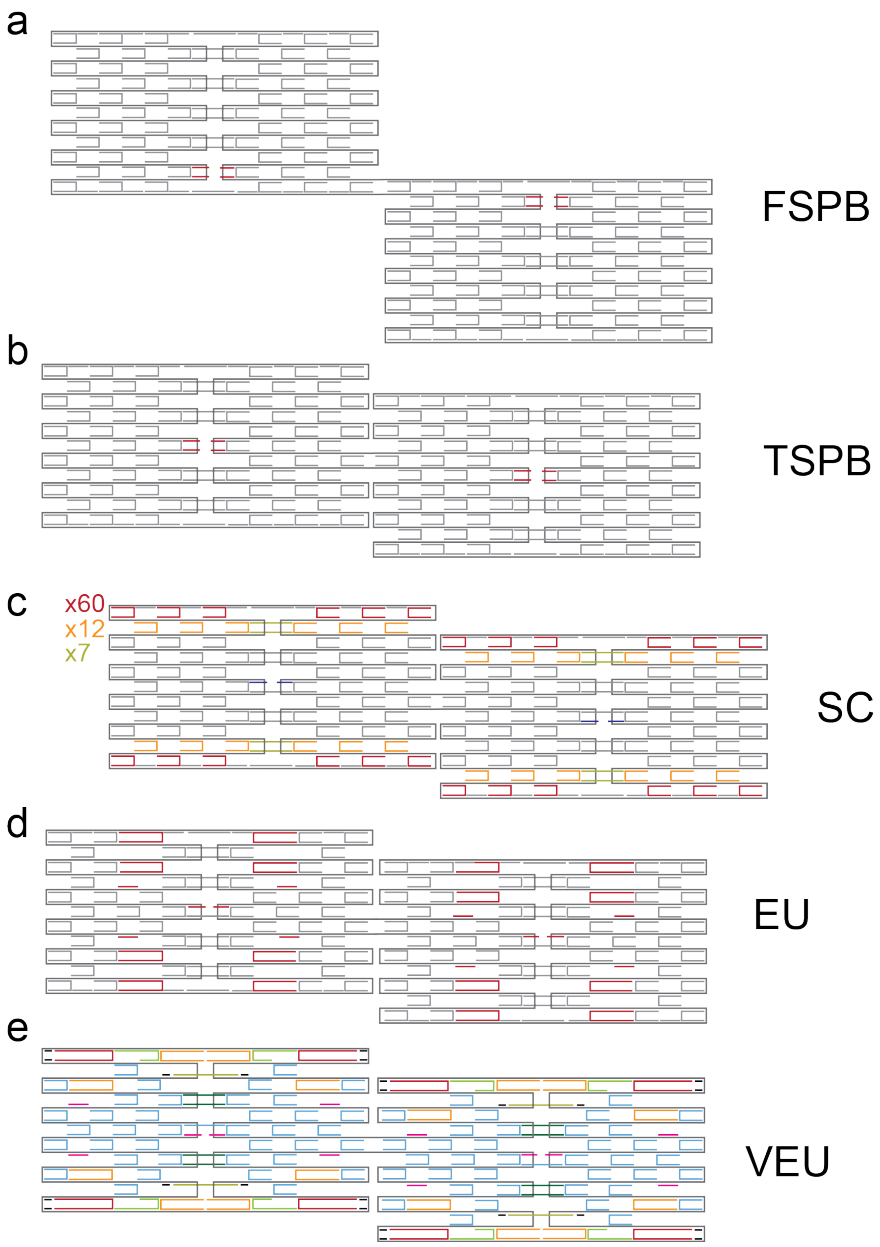


Figure A.2: Staple layouts for variants of the polymorphic tile (FSPB, TSPB, SC, EU, VEU), depicted for one configuration in each case. (a) First Seam staple Pair Broken - FSPB. (b) Third Seam staple Pair Broken - TSPB. (c) Different Staple Concentrations - SC. Concentration of grey staples $\sim 8\times$ concentration of scaffold; yellow/orange/red staples are 7/12/60 times as concentrated as the grey staples, respectively. (d) Extended U-staples (EU). (e) Very extended U-staples (VEU). Colour coding indicates intended sequence of incorporation - from first to last: red, orange, dark yellow, pale green, dark green, pale blue, pink, black. For all variants other than (c) and (e) design modifications are marked in red; for (c) and (e) colour coding illustrates intended order of incorporation.

# Appendix B

# Polymorphic tile configurations

In the polymorphic tile, the number of times the scaffold crosses between dimer halves is equal to $2n$ where $n$ is odd. For $n = 1$: the scaffold crosses over once, and then 2646 bases later it crosses back again. Figures on the next three pages show the full set of routings for $n = 1$ ($i - xxvi$), and three examples of tiles for which $n > 1$ ($xxvii - xxix$). The red section has the same sequence as the blue section. Domains with the same numerical label have the same sequence.

The tiles labelled $i$, $vi$, $xiv$ and $xix$ provide an example of shape degeneracy. These four tiles have different scaffold routings but the same overall shape. They can be split into two pairs of opposite chirality, but from the AFM images it is impossible to ascertain which face of a tile is pointing up, and the degeneracy is not lifted.

In particular instances, tiles may be converted from one shape into another without unbinding of staples. This is only possible for two groups of three tiles, which are: $(ix, x, xi)$ and $(xxii, xxiii, xxiv)$.

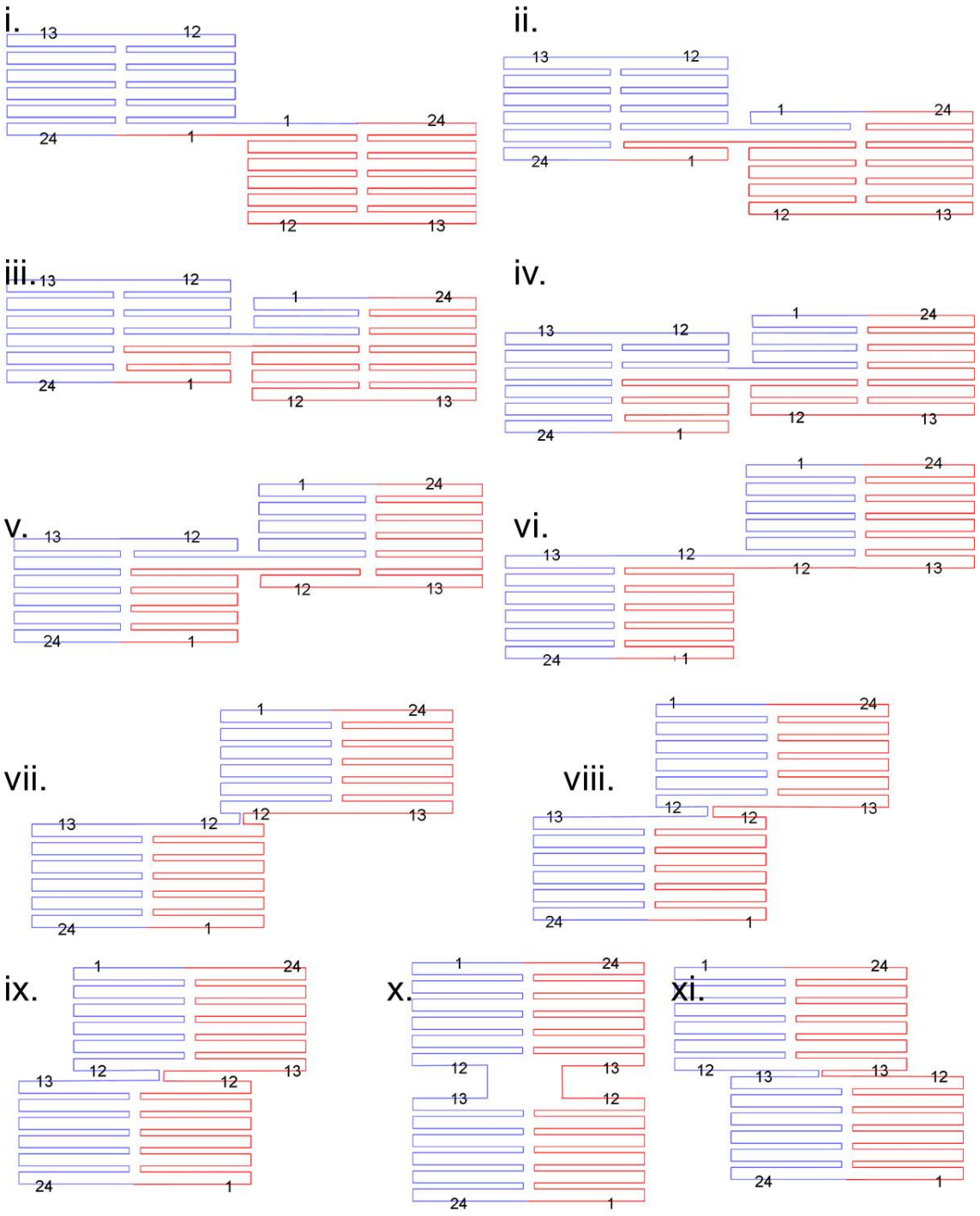


Figure B.1: Scaffold routings (1).

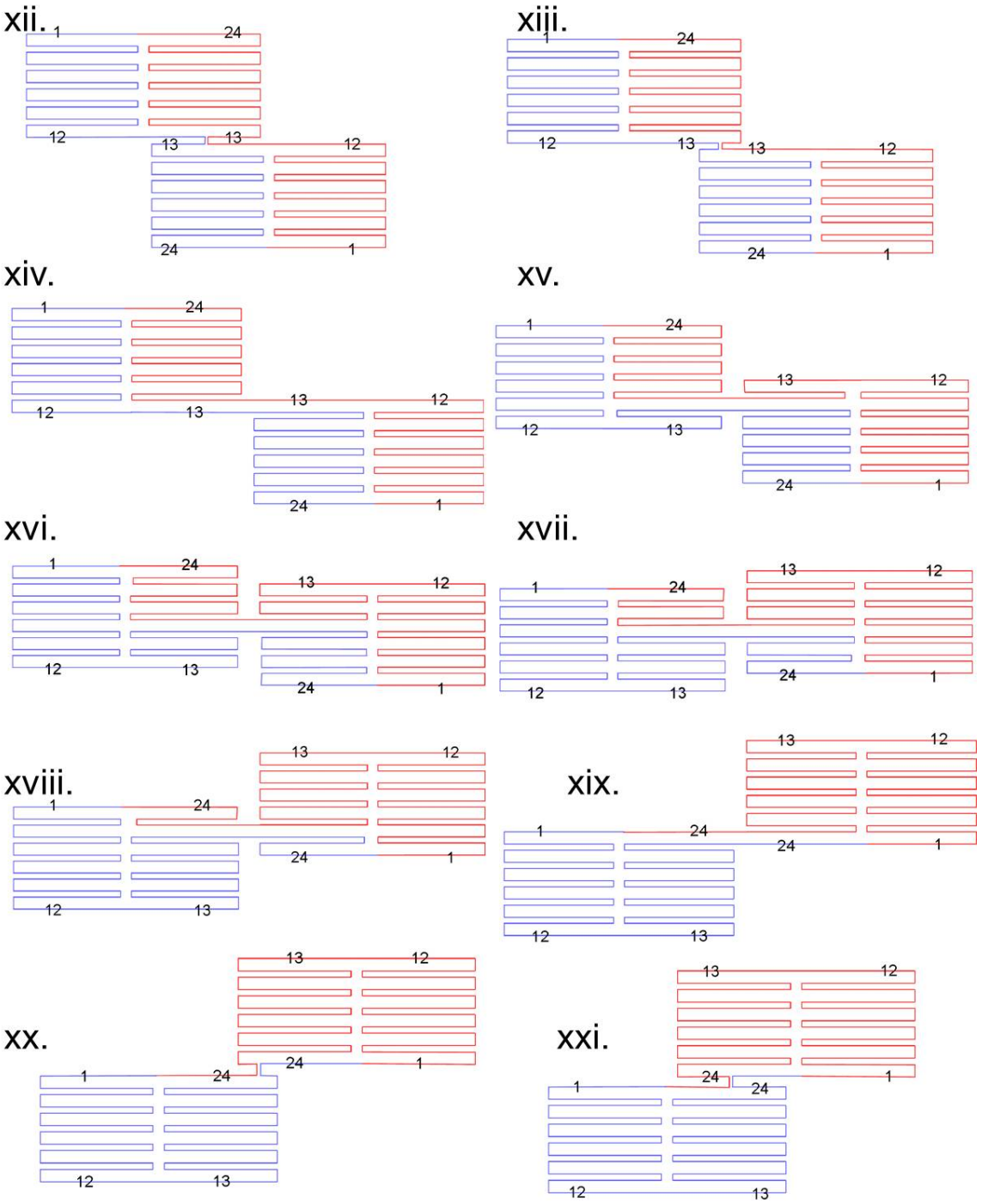


Figure B.2: Scaffold routings (2).

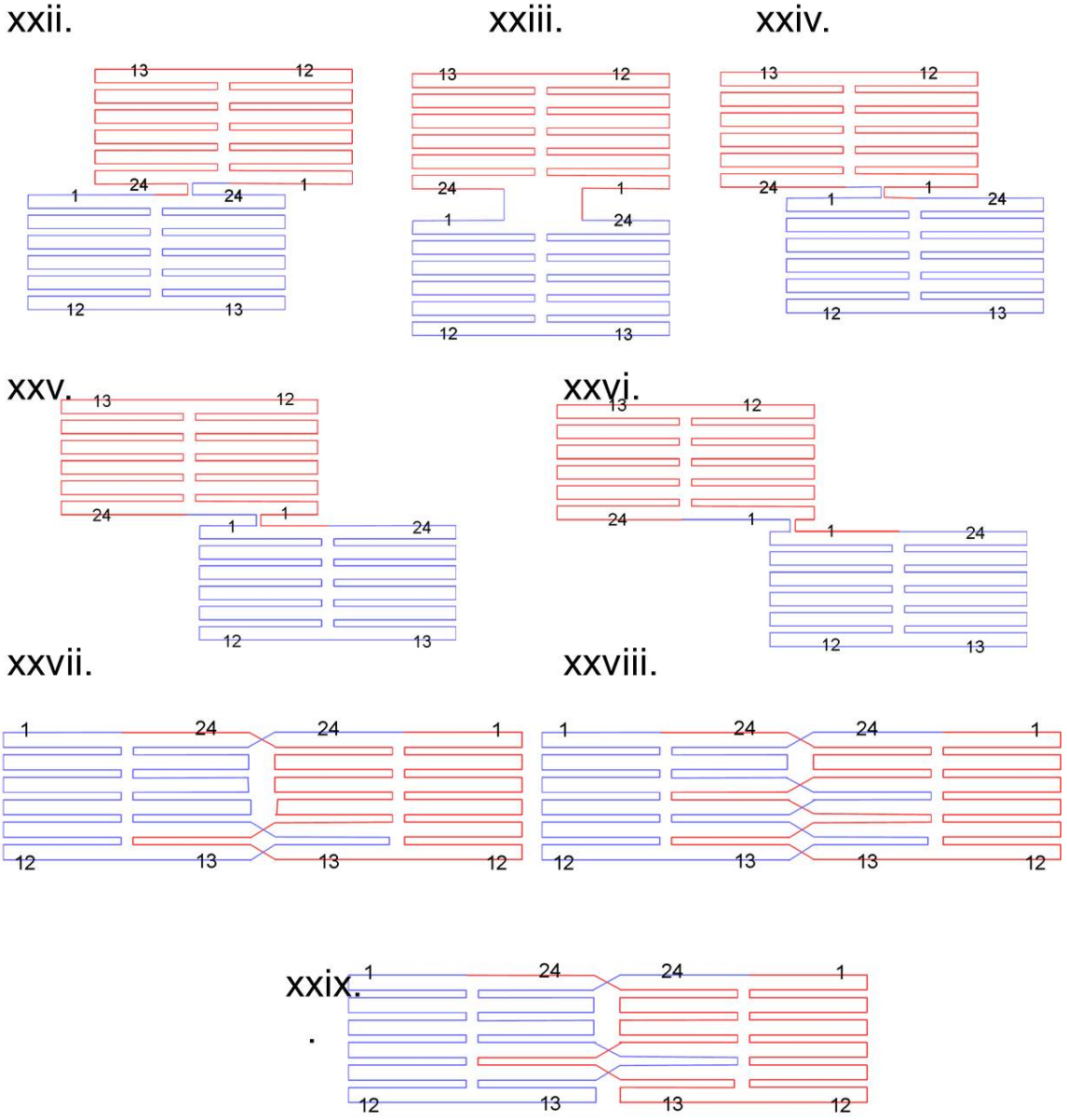


Figure B.3: Scaffold routings (3).

# Appendix C

# A synthetic autonomous rotary motor

Here I will present a proposal for the application of DNA origami assembly principles to the construction of an autonomous rotary motor, inspired by the biological molecular machine F1-ATPase [14]. The mechanism I suggest is based on the principle of sequential opening of DNA hairpins, which has already been used to drive linear DNA motors [48, 168] but has not yet been applied to rotary devices. I will explain in this Appendix how deploying DNA hairpins in this new context would enable the construction of a synthetic motor which would be able to revolve continously without external intervention, unlike previous DNA machines. As presently envisaged, it would be entirely made from and fuelled by DNA, and if realized experimentally it could be used in a range of nanostructures and devices. The experimental demonstration is outside the scope of my thesis, but I include a brief discussion of the techniques

which should be used and I describe how I have successfully simulated the operation of the motor.

## C.1 Review: the current state of the art

### C.1.1 Linear DNA motors

Linear DNA motors typically travel along a track constructed from DNA duplexes or tiles, stepping from one site to the next. Various two-footed 'walkers' have been developed, generally inspired by the naturally occurring molecular walkers myosin and kinesin. Early examples of bipedal walkers were non-autonomous, driven by manual addition of 'set' and 'unset' strands. Both 'inchworm' [175] and 'hand-over-hand' [176] mechanisms were demonstrated in 2004, and at around the same time Yin *et al.* produced an autonomous unidirectional motor consisting of a short oligonucleotide which was moved along a track by a repeating cycle of ligation and enzymatic strand cutting [177]. As described in Chapter 1, it has also been shown that a single strand of DNA (a 'motor') can be caused to step from one 'stator' to the next by using an enzyme to nick the stator-motor duplex such that a short piece of the stator dissociates, leaving a single-stranded toehold on the motor, which can then step to the next stator [47]. Such a motor can also operate on a track laid out on an origami tile [50, 76], or even navigate a network of such tracks according to 'instructions' provided [51].

The two linear motors of most relevance to the concept I describe in this Appendix are those reported by Green *et al.* [48] and Omabegho *et al.* [168], both of which are autonomous bipedal walkers fuelled by DNA hairpins. For the former, the track was a single linear strand of DNA, and the walker was initially in a configuration in which both feet were bound. A short domain at the back of the front foot competed with an identical domain at the front of the rear foot, and ultimately the domain from the rear foot was displaced, opening a loop on the rear foot and revealing a binding site for the toehold of a free fuel hairpin. As the hairpin bound, it displaced the track and caused the foot to unbind. This event revealed a binding site on the first fuel hairpin for the toehold of a second fuel hairpin, and this provided a mechanism for removing the fuel from the foot to enable the foot to rebind.

For this motor, geometric constraints introduced a bias for lifting the rear foot rather than the front foot but the raised foot was sometimes replaced in its original position rather than at the site ahead. Coarse-grained modelling has shown that it would also be possible to introduce a bias for stepping forward by placing the track under tension [68].

The motor of Omabegho *et al.* [168] also used DNA hairpins as fuel, and it walked along a track which consisted of metastable stem-loops on a double-crossover tile structure. The binding of the first fuel hairpin displaced one of the legs from the track, enabling it to step forward, activating as it did so the site to which the second fuel hairpin then bound. The other leg of the motor was then released by a

strand displacement reaction involving the second fuel hairpin, and then this leg could step forward. The motor was studied using polyacrylamide gel electrophoresis, where strands were cross-linked with psoralen to freeze the motor in each state.

A range of other DNA-based linear motors exist, including some interesting hybrid systems. For instance it was shown recently that a CdS nanocrystal could be transported along a single-walled carbon nanotube using a motor based on DNAzymes which could cleave RNA sequences attached to the nanotube [52].

### C.1.2 Rotary motors

To the best of my knowledge, none of the DNA-based rotary motors which have been reported to date have been autonomous - all have required external intervention to drive rotation. For instance, the motor developed by Yan *et al.* [178] involved a tile which underwent a conformation change involving rotation upon addition of 'fuel' strands and 'set strands'. The device was reset using a different combination of strands, and its operation was confirmed by gel electrophoresis and AFM.

More recently, a rotary motor based on a DNA catenane was reported [179]. In this device, the catenane [180] (two interlocked rings of DNA) was selectively driven between three different configurations by addition of mercury or cysteine, and changes in pH. The state of the motor was monitored using fluorescence, and it was shown that the device could be cycled from state to state.

Rajendran *et al.* showed that the transition between the right-handed B form

of DNA and the left-handed Z form could be used to drive the rotation of a motor constructed from DNA origami [181]. B-DNA is converted to Z DNA at higher salt concentrations, and they exploited this to drive rotation of a rotor attached to a DNA origami frame, which they studied using high-speed AFM.

## C.2 Challenges

Any motor, whether linear or rotary, must step from one state to the next rather than remaining in one state or fluctuating endlessly between states without making progress. In linear DNA motors, directionality is frequently assured using a 'burnt-bridges' mechanism, in which the track is destroyed behind the motor as it steps to its next binding site, ensuring that there is no going back and that the motor can only move in one direction. For a rotary motor, a 'step' from one 'state' to the next involves a rotation of part of the motor (the 'rotor'), by some fraction of a turn, and thus the motor's state can be defined by the angle through which the rotor has moved, with reference to the stationary part of the motor.

For linear motion, each step of the motor takes it to a new position, but for rotary motion the rotor must return to orientations it has occupied previously. This precludes the use of a burnt-bridges approach, in the absence of a repair mechanism. The major challenge of designing a motor which rotates is thus to establish a cycle in which the motor moves from one state to another, but is able to return to the original state at the appropriate time. For the motors described above, this problem was solved by use

of an external stimulus, which enabled them to be driven from one state to another at will. This means that these systems are not autonomous.

There are thus three inter-connected challenges inherent in the problem of how to design an autonomous rotary motor made and fuelled entirely with DNA. Firstly, the system must be cyclic. Secondly, the motor must must step from one state to the next and continue to rotate in the same direction. Thirdly, and most significantly, it must rotate on its own without intervention. Consequently all elements of the system must be in place from the start, but they must be designed such that they are not able to affect the motor dynamics until the desired point of the cycle. The first two challenges have been addressed previously by using an external stimulus to drive rotation and induce a change of state, but to the best of my knowledge the third has not previously been addressed.

## C.3 Concept

As discussed above, it has been shown that toeholds for DNA strand displacement reactions may be hidden inside hairpins, the opening of which reveals the toehold and enables it to bind to its target [48, 168]. This suggests a mechanism whereby the timing of key events in the cycle of a motor could be controlled. Fig. C.1 shows a possible design for a rotary motor constructed from DNA origami, and Fig. C.2 shows how rotation would be driven.

The motor design comprises two separate components - the rotor and the axle.

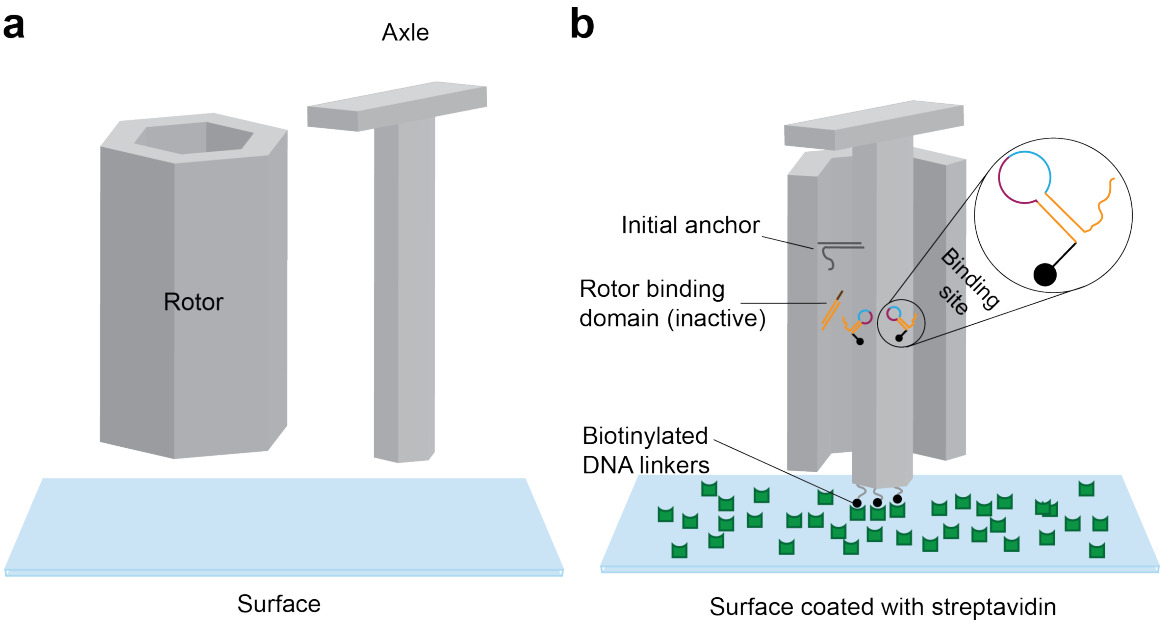


Figure C.1: Design of an autonomous rotary motor constructed from and fuelled by DNA. (a) Motor components, shown separately: the rotor, the part which rotates; the axle, which carries the binding sites for the rotor binding domain. The shape of the axle prevents the rotor from escaping. (b) Assembly of components. The rotor is shown in a cutaway view to reveal the binding sites on the axle. The motor is secured to the surface via single-stranded DNA linkers bearing a biotin molecule, which binds very tightly to the streptavidin molecules on the surface.

The rotor is the piece that moves, and it has a central hole through which the axle passes. The axle has a T-shaped cross-section to prevent the rotor from escaping when the axle is immobilized on a surface. The rotor has a single binding domain, a strand of DNA pointing towards the axle, and the axle carries three binding sites, to which the rotor's binding domain can adhere. I originally envisaged the rotor revolving inside a fixed housing, rather than rotating around an axle, but I realized that this would preclude observation of the motor using particle tracking, because there would probably be insufficient space inside the housing for a nanoparticle. Furthermore, even if the space was large, the radius of the particle's trajectory might be too small to be resolved. I overcame these potential difficulties by modifying the design - if the

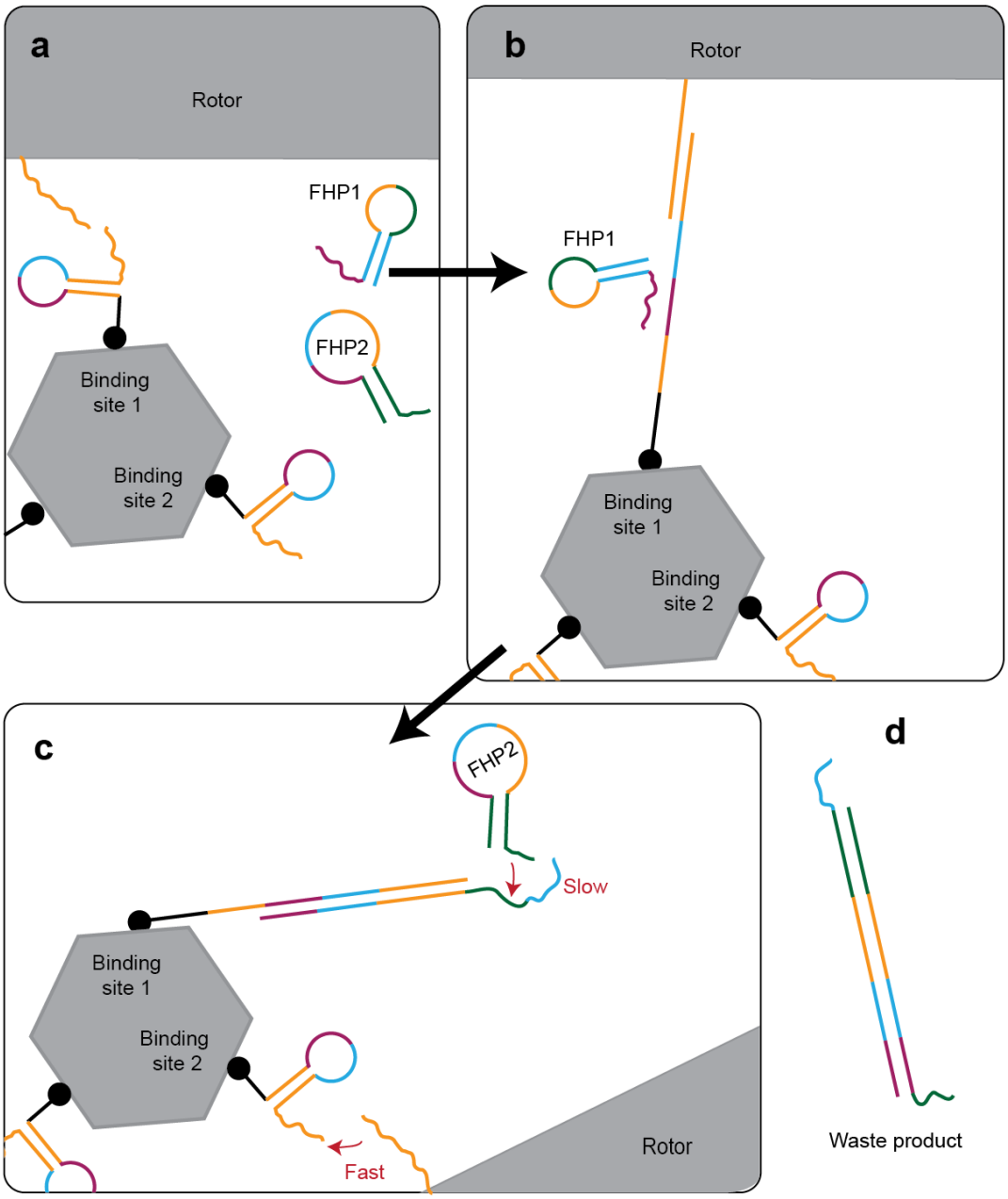


Figure C.2: Scheme for driving rotation of motor. FHP='Fuel hairpin'. (a) State of system immediately after initial anchor (not shown) released and rotor binding domain activated. (b) Rotor binding domain bound to first binding site. (c) Rotor binding domain unbound from first binding site, moving to the second. The binding of FHP2 is sufficiently slow that the rotor binding domain has moved to the third binding site before the first binding site is regenerated. (d) The waste product, a duplex containing FHP1 and FHP2.

outer piece of the motor rotates, external placement of the nanoparticle is possible.

Assembly of the motor would be accomplished as follows. The axle and rotor would be assembled separately using a standard origami synthesis protocol, but the rotor's binding domain would be blocked, so that it could not bind to the binding sites on the axle. Instead the rotor would be attached to the axle via an 'initial anchor' which would function as a brake. This is a double stranded DNA linker consisting of one strand connected to the rotor and its complement connected to the axle. The initial anchor strand on the axle is terminated with a few bases which are not complementary to the strand on the rotor, to provide a toehold to enable the initial anchor to be released.

Once the rotor and axle had been locked together and the motor immobilized on a surface via strepavidin-biotin linkers, the 'initial anchor' would be broken by introducing a strand to remove the rotor anchor strand by toehold-mediated strand displacement. In principle this anchor could be released using a DNA-cutting enzyme, but utilizing DNA strand displacement allows all components of the system to be made from DNA. Next the fuel hairpins FHP1 and FHP2 would be added, remaining inactive temporarily. The final strand to be added to the system would be a trigger strand which would unblock the rotor binding domain, enabling operation of the motor, as the rotor binding domain would then be able to hybridize with a binding site on the axle.

The binding domain of the rotor would thus open the hairpin of the binding site

on the axle, revealing a toehold for FHP1 (Fig. C.2b). This would enable binding of FHP1, the first fuel hairpin. The binding and opening of FHP1 would displace the binding domain of the rotor, which would then be free to move to the next binding site on the axle. The opening of the hairpin loop of FHP1 would reveal a toehold for FHP2. The binding and opening of FHP2 would then strip FHP1 from the binding site, but this would be a slow process, and thus the first binding site would remain inactive until after the rotor had bound to its third binding site. After this time, the first binding site would be regenerated, being restored to its original condition, ready for binding of the rotor binding domain. The slow speed of FHP2 opening would ensure that the motor would continue to rotate in the same direction and it would not return immediately to the site at which it started[1].

Binding sites on the axle would have to be sufficiently far apart that the motor cannot skip a binding site, and the concentration of FHP1 and FHP2 would need to be high to enable continuous operation of the motor. When the supply of fuel hairpins runs out, the motor will stop.

The operation of the motor is shown in Fig. C.3. Each binding site ($S$) on the axle can be in one of three possible states - rotor binding domain ($R_d$) bound, FHP1 bound (rotor unbound, site unavailable), or available. Carefully timed transitions between these states drive the rotation of the rotor. Binding of the rotor domain to

[1]According to the scheme I present here, the motor is committed to continue to rotate in the direction it initially takes, but 50% of the motors would be expected to spin anticlockwise and the other half would go clockwise. It would be possible to introduce a system for controlling directionality, using slow-release blocking strands on two of the binding sites on the axle, such that the motor always binds first to site 1, and the other sites would then be unblocked in sequence.

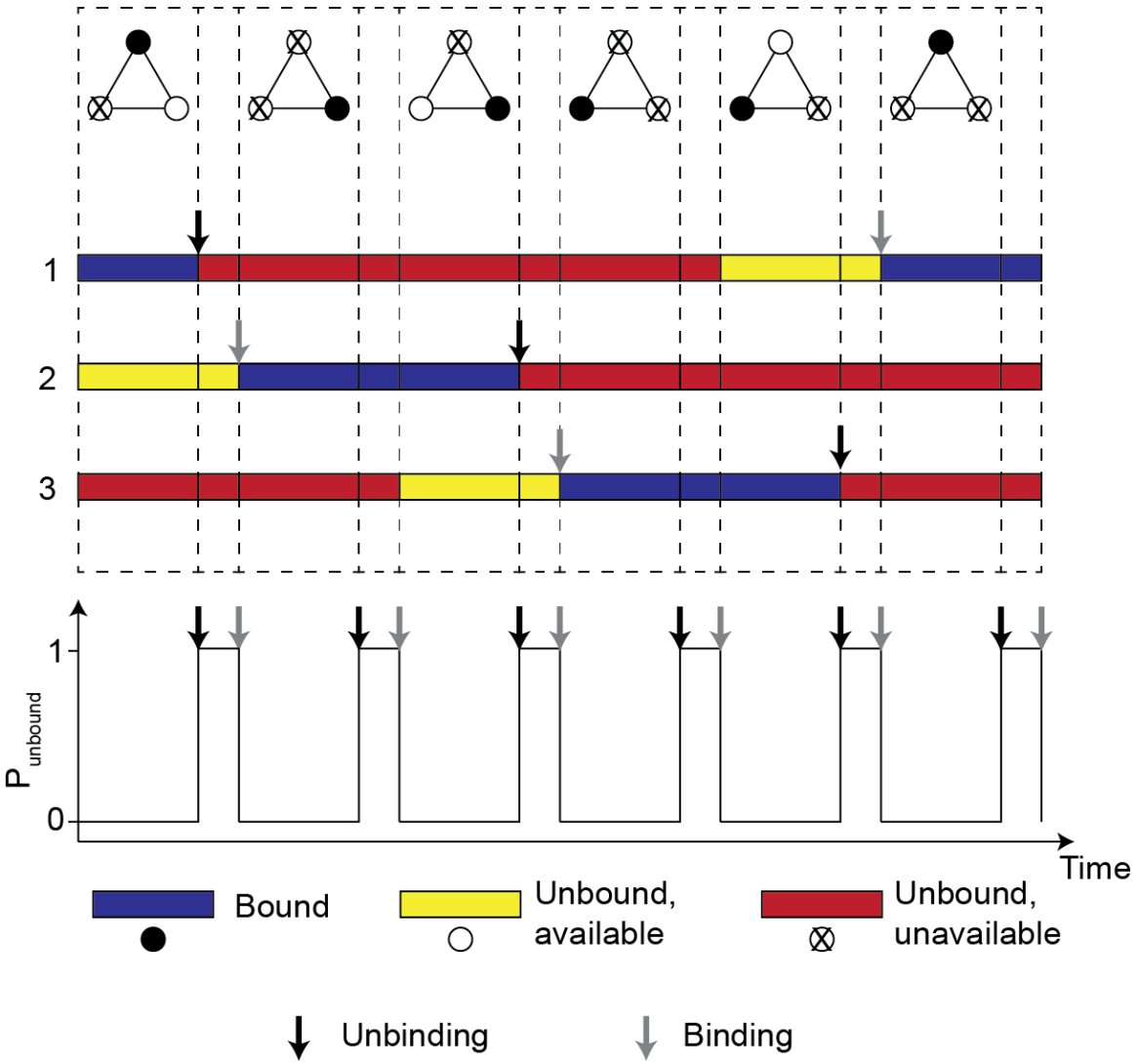


Figure C.3: Operation of rotary motor, showing states of the three binding sites on the axle as a function of time. Diagrams at top: illustrates overall state of device. Coloured bars show how the state of each site changes. The graph at the bottom represents the probability that the rotor domain will not be bound to any site. Unbinding and binding transitions are indicated. Rotor domain is unbound only for short periods. Behaviour of the motor is idealized; in reality there would be a distribution of dwell times for each configuration, as observed in the simulations (Figs C.4-C.7).

an available site is intended to be rapid (rate = $k_r$), while displacement of the rotor domain by FHP1 is slower (rate = $k_{f1}$), and the final step - which regenerates the binding site - is slowest of all (rate = $k_{f2}$). The transitions can be represented as follows:

$$R_d + S \rightarrow R_d S \tag{C.1}$$

$$R_d S + FHP_1 \rightarrow R_d + FHP_1 S \tag{C.2}$$

$$FHP_1 S + FHP_2 \rightarrow S + FHP_1 FHP_2 \tag{C.3}$$

Here, the assumption is made that the reverse reactions proceed at a negligible rate, and for the motor to operate efficiently the effect of leak reactions must be negligible. Leak reactions may occur if hairpins are opened as a result of strands binding to toeholds within the loop domain, but these reactions should be 10-100× slower than the intended reaction [24]; if this is not sufficiently slow, the efficiency of the internal toehold could be reduced by moving two of its bases into the neck of the hairpin. Kissing interactions between the loops of hairpins might also interfere with successful operation of the motor, and it might prove to be necessary to suppress these by adding short splints inside the hairpin. Other strategies could be adopted if required, and these might include the introduction of weak secondary structure within the hairpin loop or careful placement of mismatches to disrupt the formation

of unwanted complexes.

## C.4 Simulation

I simulated operation of individual rotary motors as follows. The state of the rotor binding domain was represented by a number between 1 and 3.5 - if this number was an integer, $n$, the domain was bound to axle site $n$, and for a half integer, $n+0.5$, the domain was unbound, between axle site $n$ and the next position. The state of an axle binding site was represented as 0 (no rotor domain bound, site available for binding), 1 (no rotor domain bound, site not available for binding) or 2 (rotor domain bound).

For a given angle of the motor, the transition to the next angle should occur at a characteristic rate, as suggested by the equations above. In the simulation a random number was generated for every binding site at every time step, and this was compared to the value of the function $\exp(-\text{rate} \times \text{elapsed time})$; if the random number exceeded this reference function, the state of the system changed. This process captures single-molecule heterogeneity and stochasticity, while representing the different rates at which individual transitions occur.

Results for ten motors are shown in Fig. C.4. The simulation confirms that the rotary motor can in principle operate continuously and autonomously for many cycles. The ten traces are generally similar.

In the absence of large numbers of fuel hairpins, the motor will stop after a small number of revolutions. Fig. C.5 shows this effect in operation. Although the rate

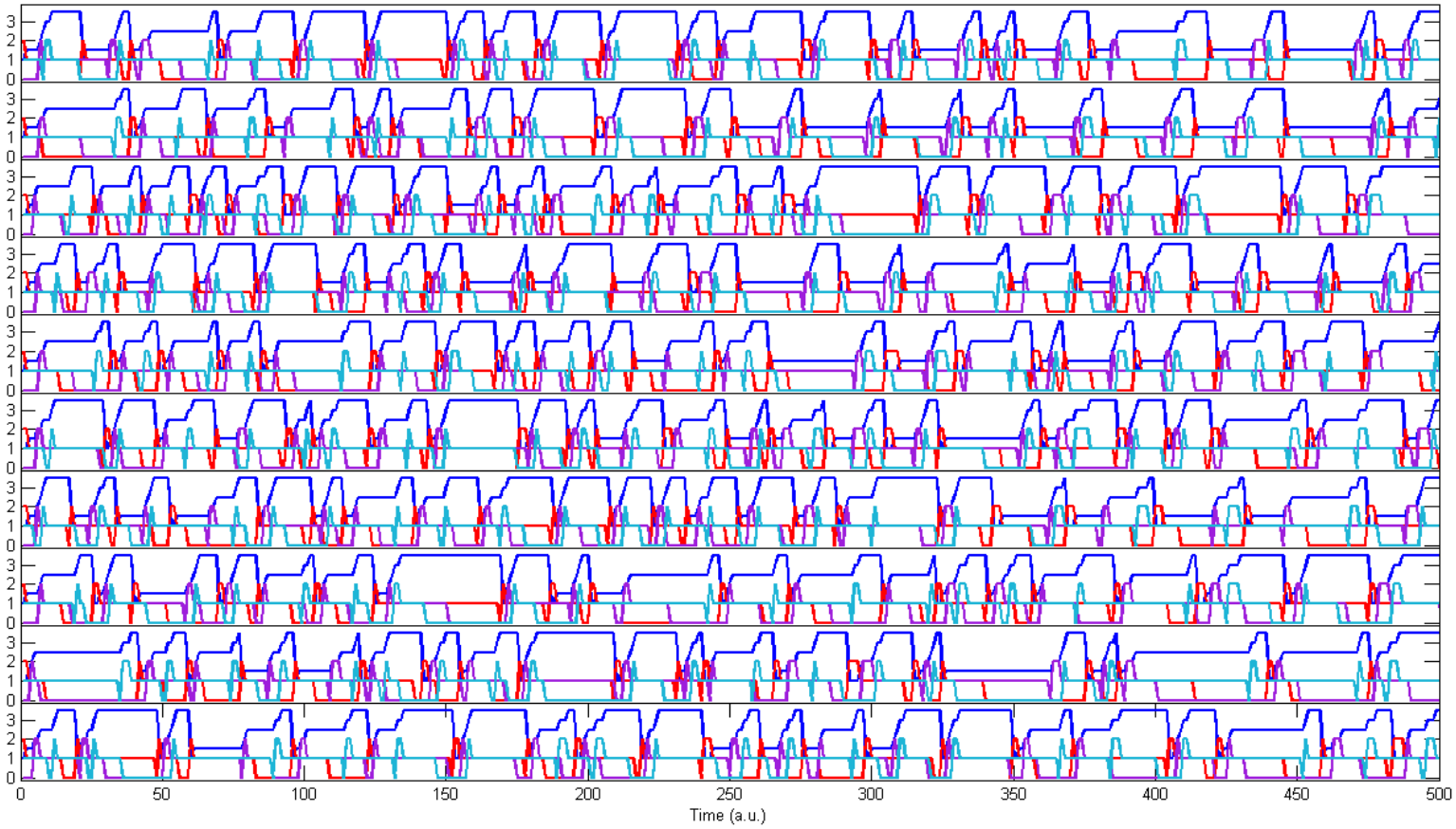


Figure C.4: Simulation: single molecule motor - normal operation. Blue line represents state of the rotor binding domain; red, lilac and turquoise represent the state of binding sites 1, 2 and 3 respectively. Definitions of the value of states are given in the text. Parameters: $k_r = 1, k_{f1} = 0.1, k_{f2} = 0.001, f = 100, N = 500, dt = 0.1$, where $k_r$ is the rate at which the binding domain of the rotor binds to an available binding site, $k_{f1}$ is the rate at which the first fuel hairpin displaces it, and $k_{f2}$ is the rate at which FHP2 strips FHP1 from the binding site. $f$ is the number of each kind of fuel hairpin for every motor, $N$ is the number of time steps, and $dt$ is the length of a time step.

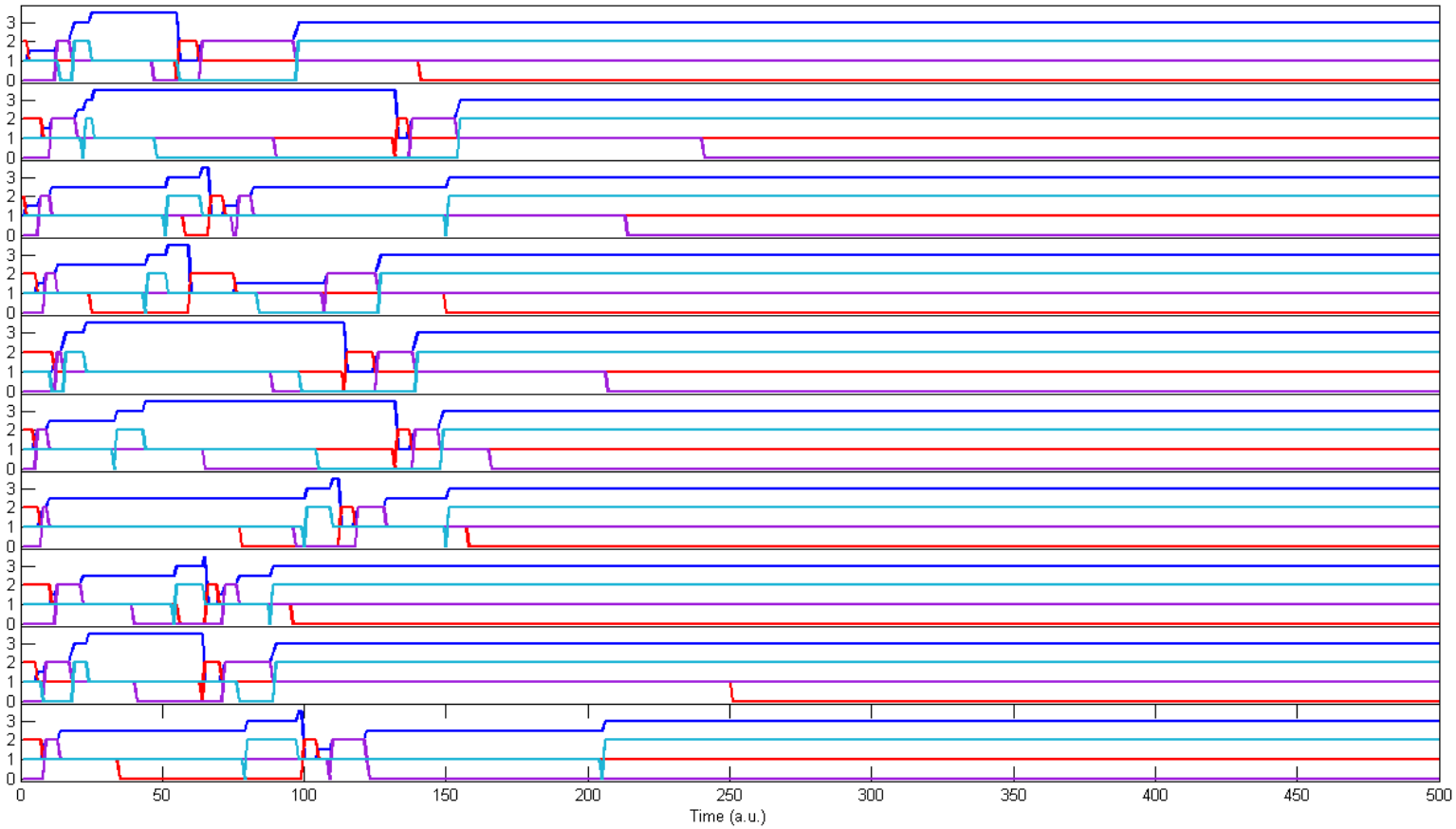


Figure C.5: Simulation: single molecule motor - limited fuel supply. Parameters: $k_r = 1, k_{f1} = 0.1, k_{f2} = 0.001, f = 5, N = 500, dt = 0.1$. Motor runs slowly and stops quickly. Colour coding as in previous figure.

constants are the same as for the simulation shown in Fig. C.4, the transitions are considerably slower because the fuel supply is limited, and the motor stops quickly.

Rotation speed depends on the values of the rate constants, which are set by the dynamics of the strand displacement reactions. These can be modified by changing parameters such as the length of the toeholds, or degree of complementarity of sequences. This effect is confirmed by the graphs in Fig. C.6. Here, all rate constants were an order of magnitude smaller than in the previous simulations. Comparison of this data with the original simulation shows that the motor now runs more slowly, and the number of rotations performed per simulation run is now 8-10 rather than around 20.

As discussed above, eventually the motor will run out of fuel and stop. This effect is demonstrated in Fig. C.7, which shows a longer timespan. It will be seen that the rate of rotation slows gradually, and eventually the motor comes to a halt. The black lines in the graphs show the normalized concentration of one of the fuel hairpins, which decays as the motor revolves. It is important to note that the motor could be refuelled at any time.

All the simulations described so far refer to the single-molecule behaviour. While many experimental techniques exist [151] which would allow single motors to be observed, it is frequently easier to make bulk measurements, and thus it would be desirable to predict the behaviour of an ensemble of motors. The simulation was performed 100,000 times, for each of the different case studies listed above, and the results were

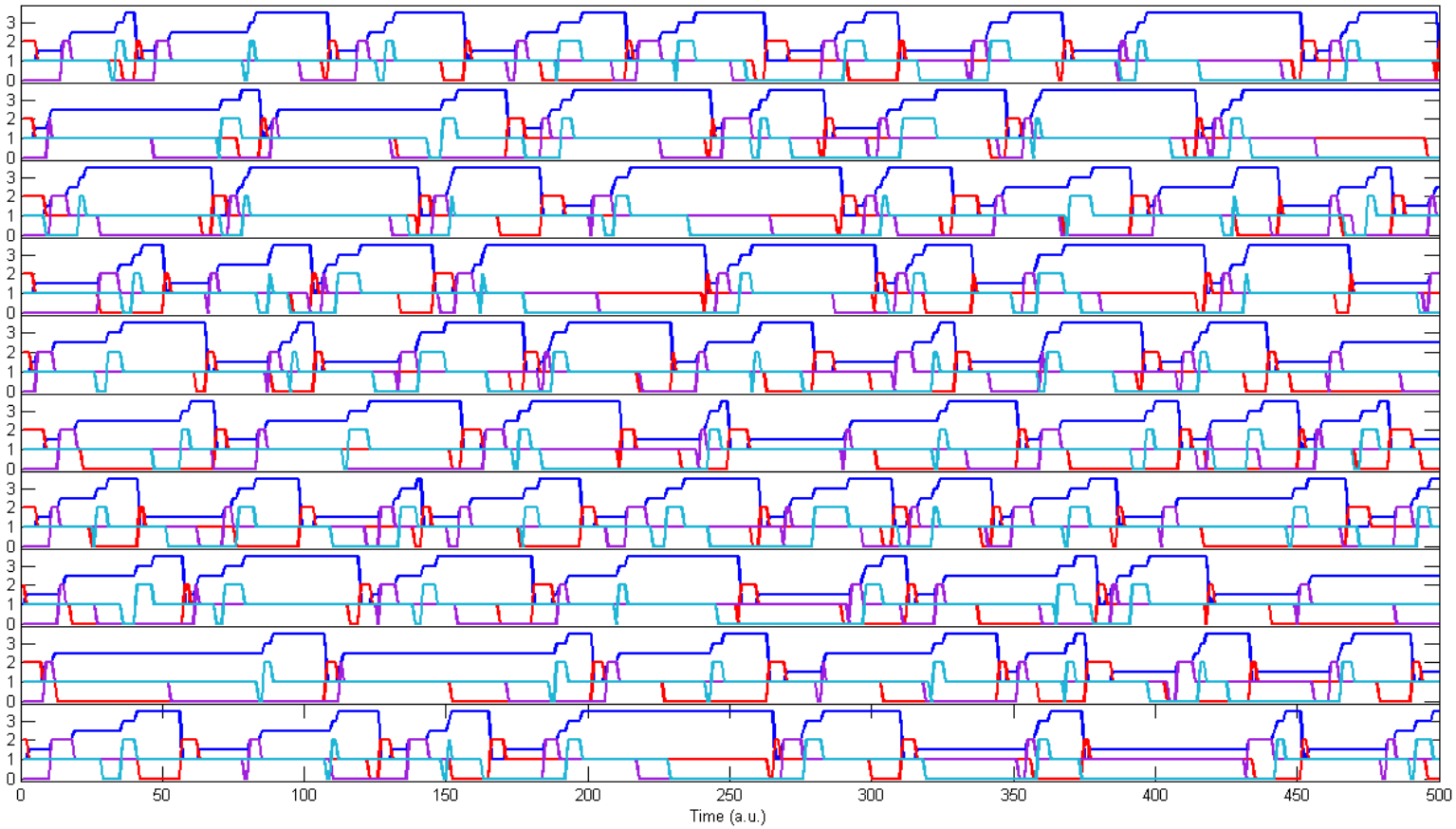


Figure C.6: Simulation: single molecule motor - modified rotation frequency. Parameters: $k_r = 0.1, k_{f1} = 0.01, k_{f2} = 10^{-4}, f = 100, N = 500, dt = 0.1$. Motor rotates more slowly when reaction rates are lower. Colour coding as in previous figures.

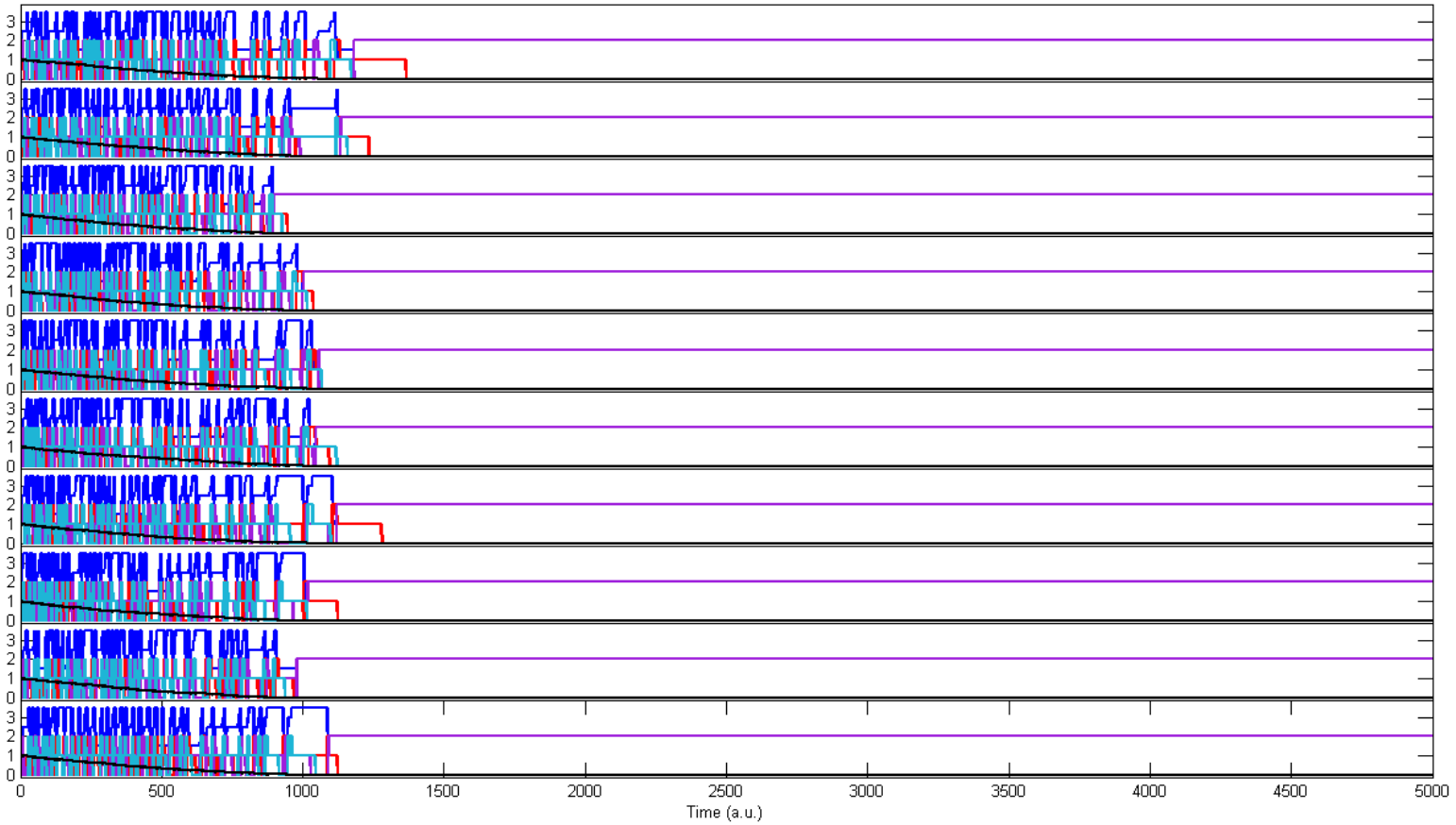


Figure C.7: Simulation: single molecule motor - behaviour on long timescales. Parameters: $k_r = 1, k_{f1} = 0.1, k_{f2} = 0.001, f = 100, N = 5000, dt = 0.1$ (all same as for Fig. C.4, except $N$). Black line indicates the normalized concentration of unused FHP1. Rotation rate slows as the fuel concentration falls. Colour coding as in previous figure.

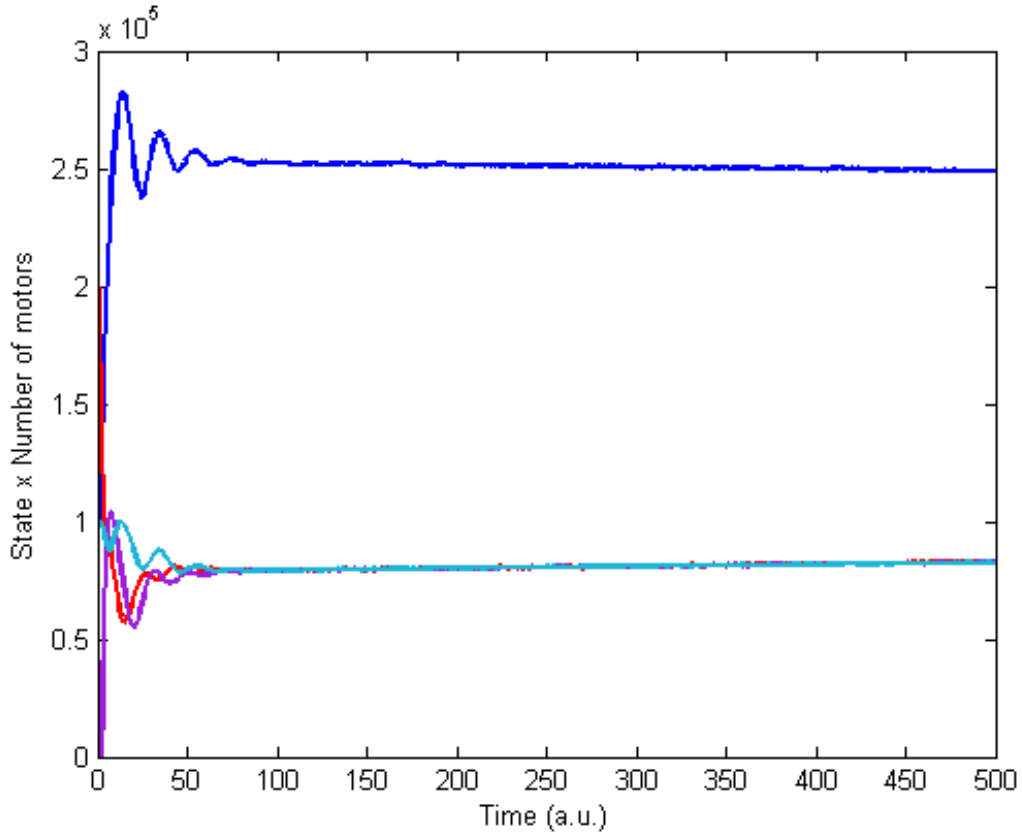


Figure C.8: Simulation: many motors - normal operation. Same parameters as for Fig. C.4, with $10^5$ motors. Colour coding as in previous figures.

consolidated to yield Figs. C.8-C.10.

The traces in Figs. C.8 and C.10 have the appearance of damped oscillations, where the amplitude of oscillation decays over time. This 'decoherence' effect arises because the transitions underlying motor operation occur stochastically. Initially all motors are in phase, but as time passes they become desynchronized. Consequently the bulk signal flatlines.

For normal operation, the oscillations corresponding to the three binding sites on the axle are related by a phase difference and all reach the same plateau. Where the fuel supply is limited, fewer oscillations are seen, and here the plateau results not from decoherence but rather from the motors stopping (Fig. C.9).

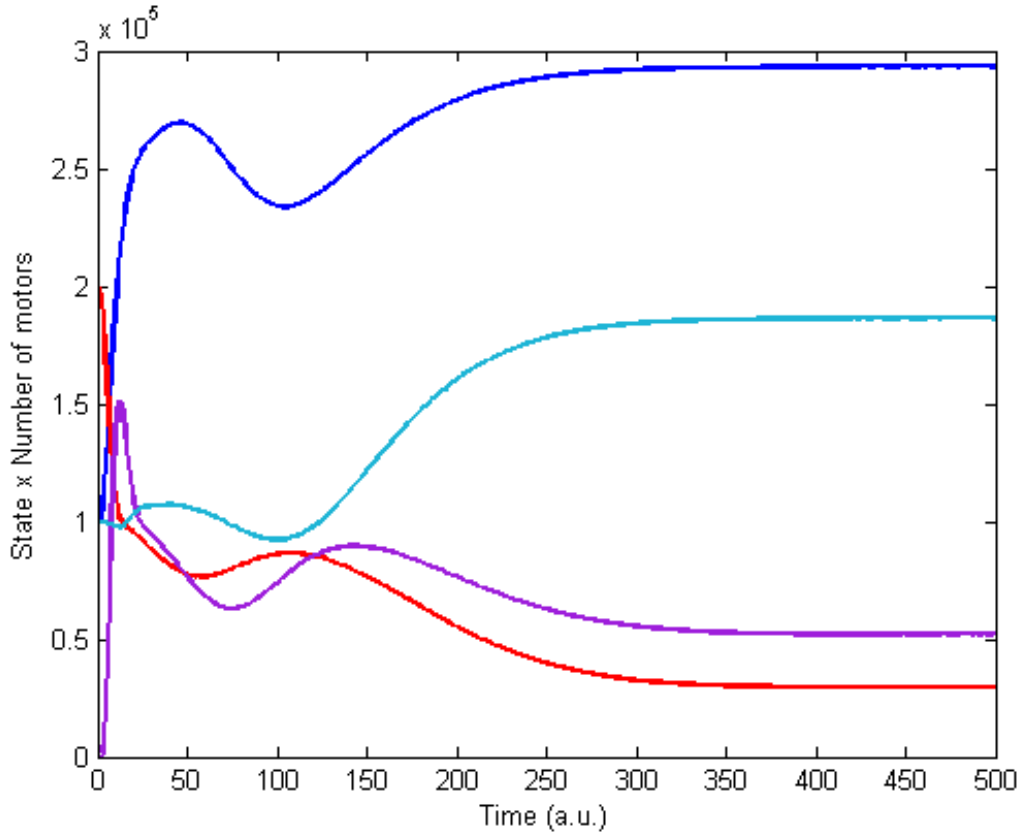


Figure C.9: Simulation: many motors - limited fuel supply. Same parameters as for Fig. C.5, with $10^5$ motors. Colour coding as in previous figures.

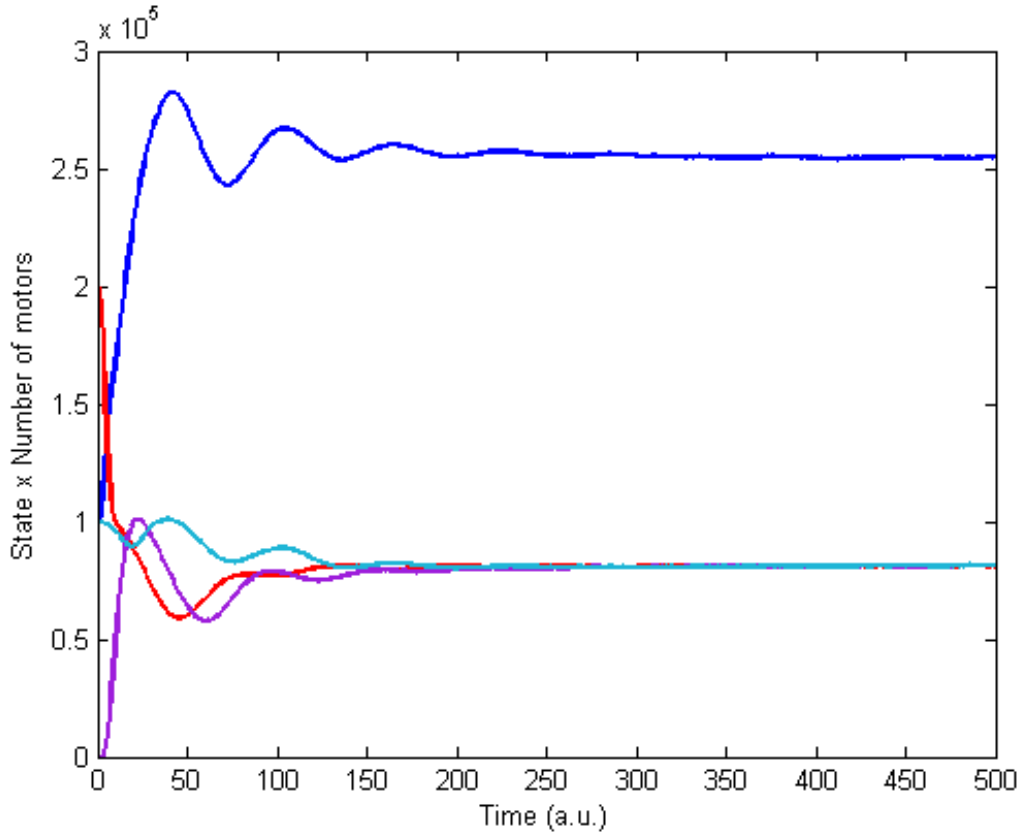


Figure C.10: Simulation: many motors - modified rotation frequency. Same parameters as for Fig. C.6, with $10^5$ motors. Colour coding as in previous figures.

The key assumptions underlying the simulation are:

- The hairpins will not open spontaneously. In other words, the hybridization energy for the hairpin necks is considerably greater than $k_{\mathrm{B}}T$. This is correct for operation at room temperature. At elevated temperatures, the motor would be prevented from rotating because the hairpins would not remain intact and control over the transitions would be lost.

- Leak reactions do not occur (as discussed above).

- Incorrectly assembled motors will not rotate at all and thus will not contribute to observed behaviour.

- There is sufficient space inside the rotor for all reactions to occur.

I also assumed that the rotor visited the axle binding sites in sequence. This equates to assuming that the rotor does not mis-step. This assumption is justified by inspection of Fig. C.4 which reveals that it is very rare for the opportunity of mis-stepping to arise (for these rate constants).

## C.5 Design

The overall geometry of my proposed motor is shown in Fig. C.11. The largest dimension of the hole in the centre of the rotor is $13H - 2r \approx 50\,\mathrm{nm}$, where $H$ is the distance between the centres of two helices diametrically opposite each other in a hexagonal unit ($\approx 4\,\mathrm{nm}$) and $r$ is the radius of a helix ($1\,\mathrm{nm}$). The cross-bar of the

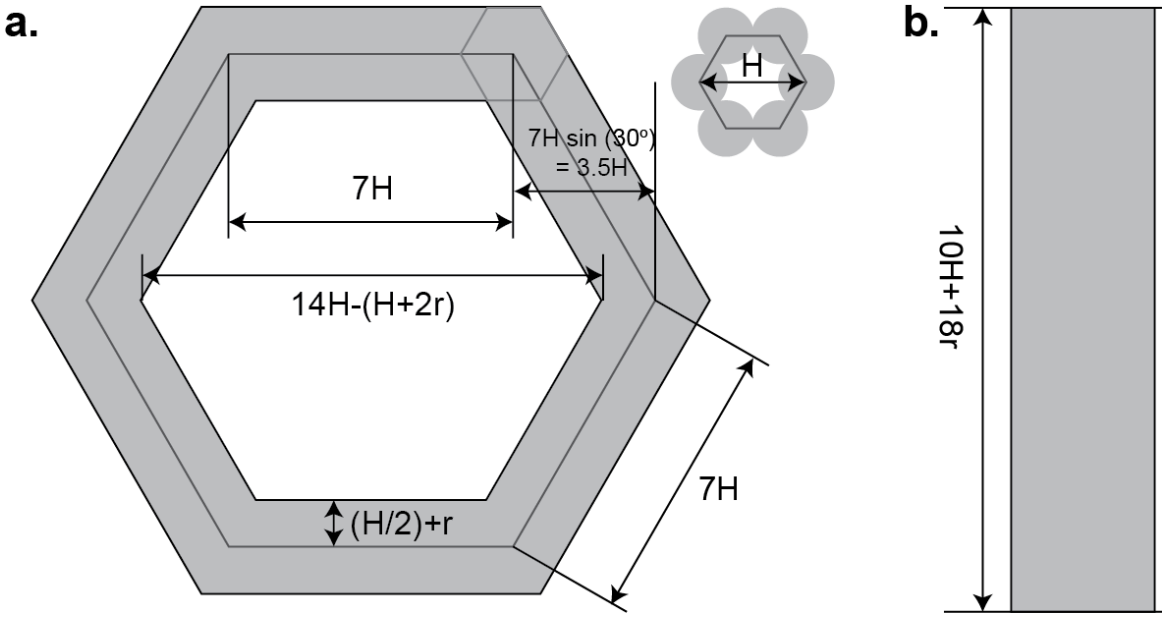


Figure C.11: Rotary motor: design of rotor and axle. Not to scale. H is approximately 4nm.

axle is designed to be $10H + 18r \approx 58\,\text{nm}$; for the rotor to escape, the two components would have to tilt at a relative angle of around 37°.

The size of the hole is determined with reference to the maximum size of complex formed by fuel hairpins with the binding sites. This will by definition be smaller than the waste product, which consists of one double-stranded 2bp domain, three double-stranded 12bp domains, one double-stranded 4bp domain and two single-stranded 12 base domains[2]. The double stranded section would be approximately 14 nm long, and this can be used as an estimate for the space required within the rotor cavity on either side of the axle pin. The axle itself measures $3H/2 + 2r \approx 8\,\text{nm}$ across, and hence the internal cavity is designed to be slightly larger than $2 \times 14.3 + 8 \approx 37\,\text{nm}$ in size.

Both parts of the motor - rotor and axle - would be based on a hexagonal lattice. Assembly of the rotor could be accomplished most effectively by connection of three

[2]See later for details of domain lengths and sequences.

identical units[3]. A caDNAno diagram of the unit design is shown in Fig. C.12, and the scaffold is routed along 54 helices, to be connected by linkers between helix 16 and helix 41. Staples on helix 16 would be extended with a sequence A, while staples on helix 41 would carry the sequence A*. Hybridization of these two domains would join together the units of the rotor. The length of scaffold needed to make a single unit is around 3200 bases, which suggests that a customized plasmid should be used, as for the tiles studied for the core work of this thesis.

A schematic diagram of the axle (again produced using caDNAno) is given in Fig. C.13, illustrating the overall shape. The required scaffold length is around 2700 bases, again requiring a custom plasmid. The pin is designed to project below the cross-bar, exceeding the height of the rotor. The staple layout of both rotor and axle would be determined with reference to the principles elucidated in Chapters 4 and 6 - in particular, staples connecting distant scaffold domains, such as those in helices 52 and 31 in the rotor, would be paired or extended to facilitate proper assembly.

### C.5.1 Sequences of strands

Fig. C.14 shows the identity of the domains explicitly, where sequence X* is complementary to X. All duplexes shown in the figure should be stable at room temperature, and thus 12 bases would be a good length for domains B, C and F (for a GC-content of 50%, a 12-base-pair duplex melts at around 40°C).

[3]For assembly as a single unit, the scaffold routing would need to be extremely complex and the yield would therefore probably be low. The rotor has six-fold symmetry, but using six separate parts would be likely to result in low yield because the six faces would not necessarily line up correctly.

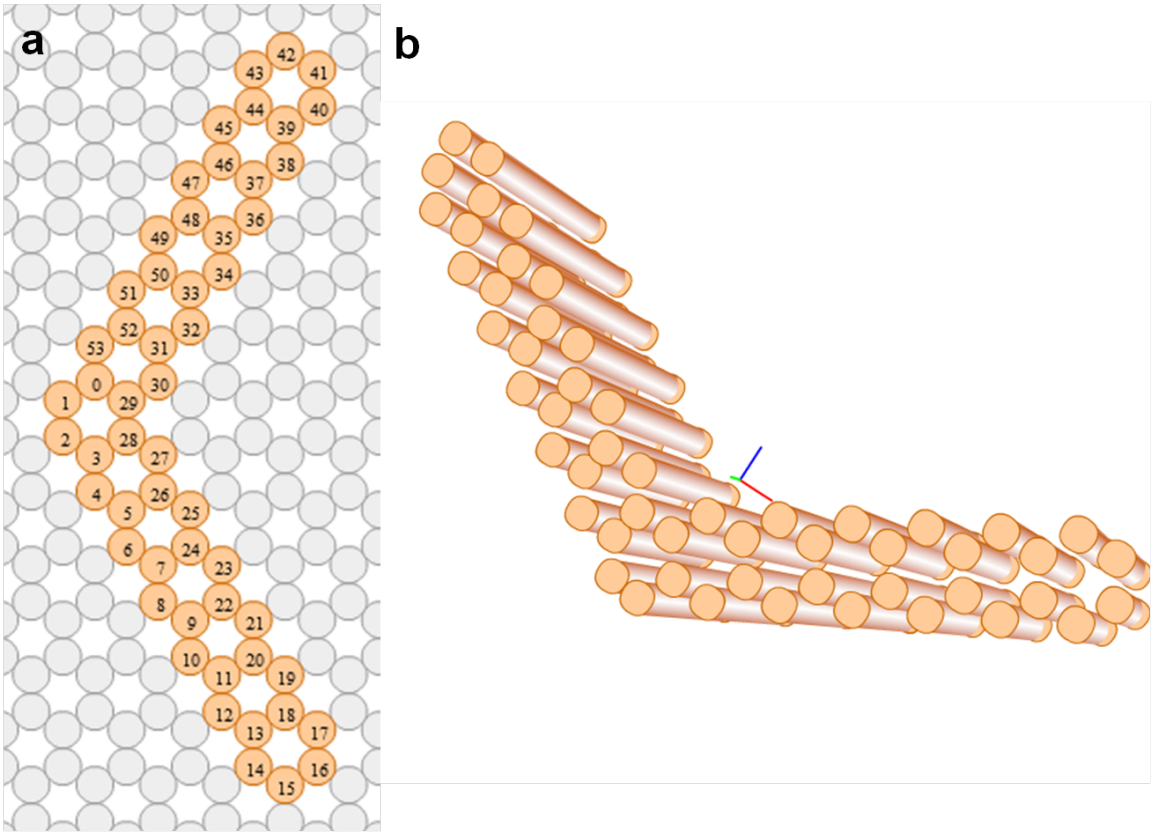


Figure C.12: Design of unit which would form 1/3 of the rotor. Linkers would join helix 16 on one unit to helix 42 on the next and the scaffold would be derived from a customized plasmid. Diagram produced using caDNAno.

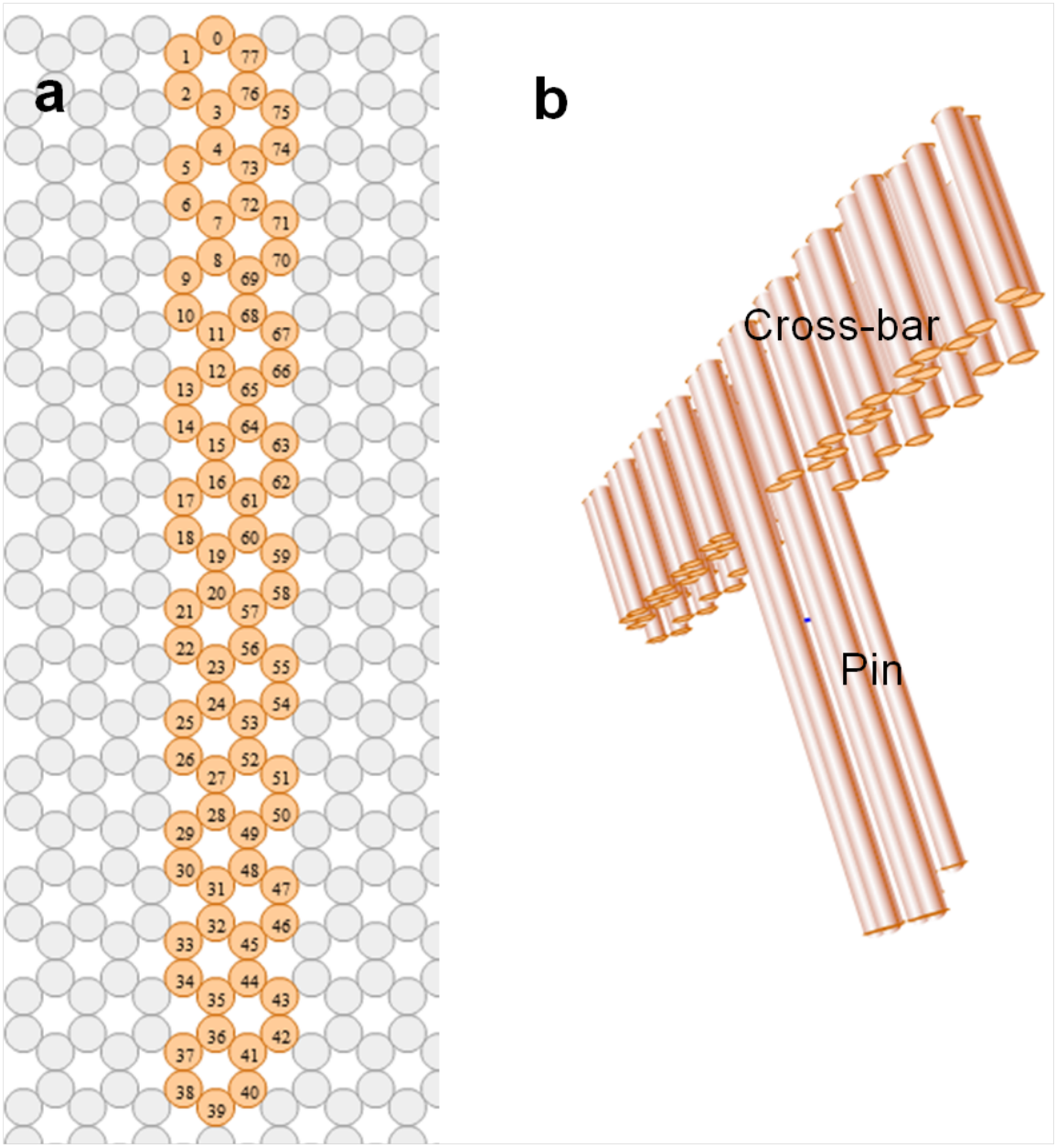


Figure C.13: Sketch of axle design, showing cross-bar and pin. Diagram produced using caDNAno.

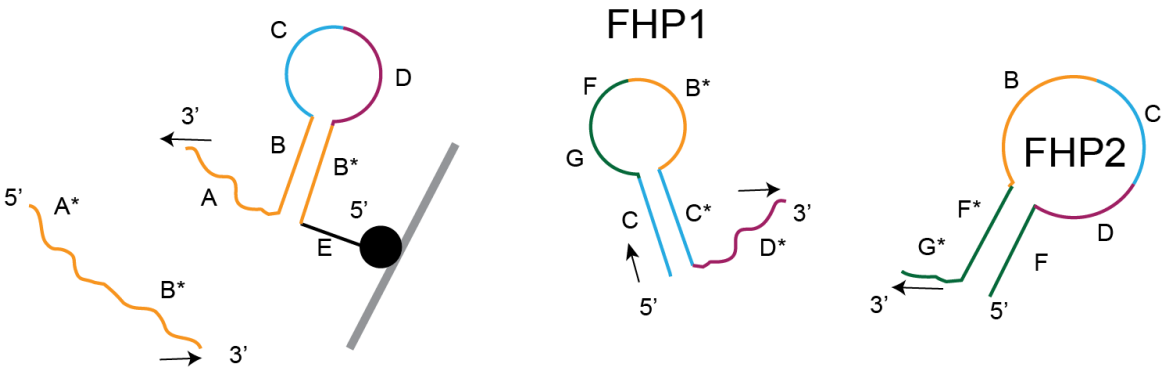


Figure C.14: Sequences for key components of rotary motor mechanism.

Domain E is simply a single-stranded linker, which should have no secondary structure. It should be both short and flexible, and thus a suitable sequence would be TTTT. The lengths of domains A, D and G determine the rates of key processes, as indicated in Table C.1. This suggests that domains A, D and G should be 5, 4 and 2 nucleotides respectively[4]. The rate of strand displacement as a function of toehold length has been studied extensively [44], and these lengths of toehold should result in a difference of one or two orders of magnitude between each rate constant and the next, while ensuring that the operation of the motor happens on an observable timescale.

| Toehold | Role | Relative rate |
|---|---|---|
| A | Rotor domain binding; opening binding site hairpin | Fast |
| D* | $FHP_1$ displacing rotor domain | Intermediate speed |
| G | $FHP_2$ displacing binding site, forming $FHP_1FHP_2$ | Slow |

Table C.1: Relative rate of processes controlled by the three toeholds A, D* and G. Simulations suggest that the fast process should be 10 times faster than the intermediate speed process, which should itself be 100 times faster than the slowest process.

The interactions between different domains are extremely important, and underpin the operation of the motor. Tables C.2 and C.3 show how strands and duplexes should

[4]If domain G was three nucleotides long, the probability of mis-stepping would increase.

| | $R_d$ | $S$ | $FHP_1$ | $FHP_2$ |
|---|---|---|---|---|
| $R_d$ | - | $\rightarrow R_dS$ | No reaction | No reaction |
| $S$ | - | - | No reaction | No reaction |
| $FHP_1$ | - | - | - | No reaction |

Table C.2: Desired interactions between strands; incubation of indicated pairs of strands at room temperature.

| | $R_dS$ | $SFHP_1$ | $FHP_1FHP_2$ |
|---|---|---|---|
| $R_d$ | - | No reaction | No reaction |
| S | - | - | No reaction |
| $FHP_1$ | $\rightarrow R_d+SFHP_1$ | - | - |
| $FHP_2$ | No reaction | $\rightarrow S+FHP_1FHP_2$ | - |

Table C.3: Desired interactions between strands; incubation at room tempertature of indicated single strands and pre-formed duplexes.

interact. Table C.2 illustrates that of all possible pairs of strands only $R_d$ and $S$ should hybridize in the absence of other components. In contrast, the interaction of individual strands with pre-formed duplexes mimics the operation of the motor and thus two of the possible processes give rise to strand displacement (Table C.3).

I obtained sequences for non-interacting domains by randomly generating sequences of the required length (or longer, for subsequent truncation), selecting those which had an appropriate GC content and checking with NUPACK whether they would interact, form any secondary structure or resemble each other. If no secondary structure features or complexes were observed, I deemed the sequences to be appropriate. Final sequences are shown in Table C.4. More elaborate approaches could also be adopted for generating orthogonal sequences, but where only a few reasonably long domains are needed random sequence generation is effective.

| Domain | Length | Sequence |
|---|---|---|
| A | 5 | AGCGT |
| B | 12 | GGGTGTGATAAG |
| C | 12 | GTCTTTGTTGGG |
| D | 4 | TTCA |
| E | 4 | TTTT |
| F | 12 | ACAGGCATTGTC |
| G | 2 | AG |

Table C.4: Sequences for domains of control strands. 30 twelve-base sequences were generated randomly and those with 50% GC content were preserved. From these, three non-interacting orthogonal linear strands were identified using NUPACK analysis (at 20° C, for a sodium concentration of 0.05 M and a magnesium concentration of 12.5 mM). 30 five-base sequences were also generated and processed similarly, lifting the constraint on GC-content and truncating appropriately, to produce sequences for domains A, D and G. It was confirmed that these three domains were orthogonal and independent of the set of three 12-base domains. For this exercise, sequences were regarded as orthogonal and independent if fewer than 2% of them interacted incorrectly.

## C.6 Possible methods for motor observation

Experimental demonstration of the operation of this motor is outside the scope of this thesis, but it is appropriate to discuss the methods which could be used.

The formation of hairpins and the interactions between individual strands could be studied by polyacrylamide gel electrophoresis (PAGE). This would indicate if the individual components of the motor system were behaving as designed. Assembly of the DNA origami units (rotor and axle) could be checked initially by agarose gel electrophoresis, but this is not necessarily conclusive and therefore visualizing the structures with a microscope would be essential. AFM could be used but the imaging process might distort the components. Hence negative stain TEM would be better, and the ideal imaging method would be cryo-EM.

Bulk measurements of motor operation in solution could be performed using either

a fluorimeter or a RT-PCR machine, although a limited number of oscillations would be observed before the motors de-synchronized, as shown by the simulation results presented earlier. To perform these experiments a fluorophore could be attached to an appropriate point on the cross-bar of the axle, with a quencher or FRET partner fluorophore placed on the rotor, such that when the motor rotated the two were brought sufficiently close to interact and produce a signal. Fluorescence could also be used for single-molecule measurements, where the fluorophore (or a quantum dot) could be tracked using super-resolution techniques. Alternatively, a larger (visible) structure could be attached to the rotor and tracked as the motor rotated. A similar technique has been used to study F1-ATPase [182].

## C.7 Possible applications

The rotary motor has several possible applications, which I list here:

- **A nanoscale winch:** The motor could be employed to twist a polymer, with one end attached to the motor and the other immobilized.

- **Mismatch detection:** The rate of rotation depends on the sequences of various strands, which means that the motor could be used to detect mismatched sequences, because it would slow down if the sequence of a key strand was not correct.

- **'Gatekeeper' for a nanopore:** If combined with techniques for constructing nanopores from DNA [119, 183], the rotary motor could act as the 'gatekeeper' for a pore. When rotation occurred, the pore would be opened periodically, but if the motor was stopped, the pore could be closed to the passage of large molecules. This would be ineffective as a means of creating an artificial ion channel because the pore would not be sealed tightly enough to prevent the passage of ions, but it might be effective for larger molecules.

- **Chemical synthesis:** The rotation of the motor could be coupled to chemical reactions, although a linear motor would probably be more versatile in this context.

- **Plasmonic antenna:** A modulated plasmonic antenna could be created by attaching a fluorophore to the rotor such that it regularly passes through a gap between a pair of gold nanoparticles, where fluorescence would be enhanced - perhaps by more than 100-fold [112].

## C.8 Context

In 1997, Boyer and Walker were awarded half of the Nobel Prize in Chemistry [184] for their work on ATP synthase, the molecular machine which synthesizes (or hydrolyzes) the energy currency molecule adenosine triphosphate (ATP) [185, 186]. $F_1F_0$ is found in the mitochondrial membrane in living cells and consists of two rotary motors back-

to-back [187]. The $F_0$ motor is driven by the proton motive force (pmf) which arises due to the hydrogen ion concentration gradient which exists across the mitochondrial membrane, and $F_1$ is driven by the hydrolysis of ATP. When the two are coupled to form $F_0F_1$ ATP synthase, $F_0$ drives $F_1$ in reverse, resulting in ATP synthesis. The rotation of $F_1$-ATPase was observed directly in 1997 [182]. The existence of this molecular machine demonstrated that a nanoscale rotary motor was not fundamentally impossible, and together with the successes of linear DNA motors, provided the inspiration for the motor described in this Appendix.

The rotary motor I have proposed in this Appendix and ATP synthase both feature three active sites. This is a result of the fact that at least three steps are necessary for any cycle involving rotary motion - (i) initialization, (ii) powerstroke, (iii) restoration. For the DNA-based rotary motor these three steps correspond to the reactions of (i) rotor domain binding, (ii) first fuel hairpin displacing rotor domain and (iii) second fuel hairpin hybridizing with first fuel hairpin. In the case of $F_1F_0$ ATP synthase the stages are (i) ADP and inorganic phosphate are loosely bound, (ii) conformational change of molecule leads to tight binding of ADP and $P_i$, resulting in conversion to ATP, (iii) ATP is released so that more substrate molecules can bind.

It is interesting to note that my rotary motor also has a macroscale equivalent, demonstrated successfully as long ago as 1956 - the rotary internal combustion engine (frequently referred to as a Wankel engine after its creator) [188, 189]. This engine comprises a three-sided rotor which revolves in a combustion chamber around an

eccentric shaft, producing three power strokes for every complete turn of the rotor. For the Wankel engine the key steps corresponding to those mentioned above are (i) compression of fuel, (ii) ignition and powerstroke, (iii) expulsion of burnt fuel (exhaust) and the induction of new fuel. (In terms of the conventional four-stroke cycle, step (ii) is two 'strokes'). In the Wankel engine, three different events are always occurring simultaneously. Not only does the operation of this machine resemble that of my rotary motor and ATP synthase, but synthetic molecular motors based on boron nanoclusters have already been referred to as 'molecular Wankel engines' [190].

# Bibliography


[1] J.D. Watson and F.H.C. Crick. Molecular structure of nucleic acids - a structure for deoxyribose nucleic acid. *Nature*, 171:737–738, 1953.

[2] M.H.F. Wilkins, A.R. Stokes, and H.R. Wilson. Molecular structure of nucleic acids: Molecular structure of deoxypentose nucleic acids. *Nature*, 171:738–740, 1953.

[3] R.E. Franklin and R.G. Gosling. Molecular configuration in sodium thymonucleate. *Nature*, 171:740–741, 1953.

[4] N. C. Seeman. Nucleic acid junctions and lattices. *Journal of Theoretical Biology*, 99:237–247, 1982.

[5] P. W. K. Rothemund. Folding DNA to create nanoscale shapes and patterns. *Nature*, 440:297–302, 2006.

[6] Y.X. Zhao, A. Shaw, X. Zeng, E. Benson, A.M. Nyström, and B. Högberg. DNA origami delivery system for cancer therapy with tunable release properties. *ACS Nano*, 6:8684–8691, 2012.

[7] S.M. Douglas, I. Bachelet, and G.M. Church. A logic-gated nanorobot for targeted transport of molecular payloads. *Science*, 335:831–834, 2012.

[8] S. Helmig, A. Rotaru, D. Arian, L. Kovbasyuk, J. Arnbjerg, P.R. Ogilby, J. Kjems, A. Mokhir, F. Besenbacher, and K.V. Gothelf. Single molecule atomic force microscopy studies of photosensitized singlet oxygen behavior on a DNA origami template. *ACS Nano*, 4:7475–7480, 2010.

[9] N.V. Voigt, T. Tørring, A. Rotaru, M.F. Jacobsen, J.B. Ravnsbæk, R. Subramani, W. Mamdouh, J. Kjems, A. Mokhir, F. Besenbacher, and K.V. Gothelf. Single-molecule chemical reactions on DNA origami. *Nature Nanotechnology*, 5:200–203, 2010.

[10] K. Jahn, T. Tørring, N.V. Voigt, R.S. Sørensen, A.L. Bank Kodal, E.S. Andersen, K.V. Gothelf, and J. Kjems. Functional patterning of DNA origami by parallel enzymatic modification. *Bioconjugate Chemistry*, 22:819–823, 2011.

[11] C. Steinhauer, R. Jungmann, T.L. Sobey, F.C. Simmel, and P. Tinnefeld. DNA origami as a nanoscopic ruler for super-resolution microscopy. *Angewandte Chemie International Edition*, 48:8870–8873, 2009.

[12] C.M. Dobson. Protein folding and misfolding. *Nature*, 426:884–890, 2003.

[13] N.C. Seeman. Nanotechnology and the double helix. *Scientific American*, 290:64–75, 2004.

[14] J.M Berg, J.L. Tymoczko, and L. Stryer. *Biochemistry, $7^{th}$ Ed.* W. H. Freeman.

[15] R. Phillips, J. Kondev, and J. Theriot. *Physical Biology of the Cell.* Garland Science.

[16] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell.* Garland Science, 4th edition, 2002.

[17] D. Voet and J. Voet. *Biochemistry.* Wiley, 2004.

[18] J. SantaLucia. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences of the United States of America*, 95:1460–1465, 1998.

[19] C. Bustamante, S.B. Smith, J. Liphardt, and D. Smith. Single-molecule studies of DNA mechanics. *Current Opinion in Structural Biology*, 10:279–285, 2000.

[20] S.B. Smith, Y. Cui, and C. Bustamante. Overstretching B-DNA: The elastic response of individual double-stranded and single-stranded DNA molecules. *Science*, 271:795–799, 1996.

[21] R.E. Dickersen, H.R. Drew, B.N. Conner, R.M. Wing, A.V. Fratini, and M.L. Kopka. The anatomy of A-, B- and Z- DNA. *Science*, 216:475–485, 1982.

[22] A. Rich, A. Nordheim, and A.H-J. Wang. The chemistry and biology of left-handed Z-DNA. *Annual Review of Biochemistry*, 53:791–846, 1984.

[23] T.J. Thomas and V.A. Bloomfield. Chain flexibility and hydrodynamics of the B and Z forms of poly(dG-dC).poly(dG-dC). *Nucleic Acids Research*, 11:1919–1930, 1983.

[24] S. J. Green, D. Lubrich, and A. J. Turberfield. DNA hairpins: Fuel for autonomous DNA devices. *Biophysical Journal*, 91:2966–2975, 2006.

[25] S. Burge, G.N. Parkinson, P. Hazel, A.K. Todd, and S. Neidle. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Research*, 34:5402–5415, 2006.

[26] M. Guéron and J-L. Leroy. The i-motif in nucleic acids. *Current Opinion in Structural Biology*, 10:326–331, 2000.

[27] L. Mazzola. Commercializing nanotechnology. *Nature Biotechnology*, 21:1137–1143, 2003.

[28] M-L. Zhang, K-Q. Peng, X. Fan, J-S. Jie, R-Q. Zhang, S-T. Lee, and N-B. Wong. Preparation of large-area uniform silicon nanowire arrays through metal-assisted chemical etching. *The Journal of Physical Chemistry C*, 112:4444–4450, 2008.

[29] Y. Wu and P. Yang. Direct observation of Vapor-Liquid-Solid nanowire growth. *Journal of the American Chemical Society*, 123:3165–3166, 2001.

[30] N.C. Seeman. DNA in a material world. *Nature*, 421:427–431, 2003.

[31] E. Winfree, F. Liu, L. A. Wenzler, and N. C. Seeman. Design and self-assembly of two-dimensional DNA crystals. *Nature*, 394:539–544, 1998.

[32] J. Zheng, J.J. Birktoft, Y. Chen, T. Wang, R. Sha, P.E. Constantinou, S.L. Ginell, C. Mao, and N.C. Seeman. From molecular to macroscopic via the rational design of a self-assembled 3D DNA crystal. *Nature*, 461:74–77, 2009.

[33] D.N. Selmi, R.J. Adamson, H. Attrill, A.D. Goddard, R.J.C. Gilbert, A. Watts, and A.J. Turberfield. DNA-templated protein arrays for single-molecule imaging. *Nano Letters*, 11:657–660, 2011.

[34] T. Kato, R.P. Goodman, C.M. Erben, A.J. Turberfield, and K. Namba. High-resolution structural analysis of a DNA nanostructure by cryoEM. *Nano Letters*, 9:2747–2750, 2009.

[35] J.H. Chen and N.C. Seeman. Synthesis from DNA of a molecule with the connectivity of a cube. *Nature*, 350:631–633, 1991.

[36] Y. Zhang and N. C. Seeman. Construction of a DNA-Truncated Octahedron. *Journal of the American Chemical Society*, 116:1661–1669, 1994.

[37] W.M. Shih, J.D. Quispe, and G.F. Joyce. A 1.7-kilobase single-stranded DNA that folds into a nanoscale octahedron. *Nature*, 427:618–621, 2004.

[38] R.P. Goodman, R.M. Berry, and A.J. Turberfield. The single-step synthesis of a DNA tetrahedron. *Chemical Communications*, pages 1372–1373, 2004.

[39] R.P. Goodman, I.A.T. Schaap, C.F. Tardin, C.M. Erben, R.M. Berry, C.F. Schmidt, and A.J. Turberfield. Rapid chiral assembly of rigid DNA building blocks for molecular nanofabrication. *Science*, 310:1661–1665, 2005.

[40] R.P. Goodman, M. Heilemann, S. Doose, C.M. Erben, A.N. Kapanidis, and A.J. Turberfield. Reconfigurable, braced, three-dimensional DNA nanostructures. *Nature Nanotechnology*, 3:93–96, 2008.

[41] D. Y. Zhang and G. Seelig. Dynamic DNA nanotechnology using strand-displacement reactions. *Nature Chemistry*, 3:103–113, 2011.

[42] B. Yurke, A.J. Turberfield, A.P. Mills, F.C. Simmel, and J.L. Neumann. A DNA-fuelled molecular machine made of DNA. *Nature*, 406:605–608, 2000.

[43] J. Bath and A.J. Turberfield. DNA nanomachines. *Nature Nanotechnology*, 2:275–284, 2007.

[44] D.Y. Zhang and E. Winfree. Control of DNA strand displacement kinetics using toehold exchange. *Journal of the American Chemical Society*, 131:17303–17314, 2009.

[45] N. Srinivas, T.E. Ouldridge, P. Šulc, J.M. Schaeffer, B. Yurke, A.A. Louis, J.P.K Doye, and E. Winfree. On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic Acids Research*, 41:10641–10658, 2013.

[46] J. H. Reif. Scaling up DNA computation. *Science*, 332:1156–1157, 2011.

[47] J. Bath, S.J. Green, and A.J. Turberfield. A free-running DNA motor powered by a nicking enzyme. *Angewandte Chemie-International Edition*, 44:4358–4361, 2005.

[48] S.J. Green, J. Bath, and A.J. Turberfield. Coordinated chemomechanical cycles: A mechanism for autonomous molecular motion. *Physical Review Letters*, 101:238101, 2008.

[49] K. Lund, A. J. Manzo, N. Dabby, N. Michelotti, A. Johnson-Buck, J. Nangreave, S. Taylor, R. J. Pei, M. N. Stojanovic, N. G. Walter, E. Winfree, and H. Yan. Molecular robots guided by prescriptive landscapes. *Nature*, 465:206–210, 2010.

[50] S.F.J. Wickham, M. Endo, Y. Katsuda, K. Hidaka, J. Bath, H. Sugiyama, and A.J. Turberfield. Direct observation of stepwise movement of a synthetic molecular transporter. *Nature Nanotechnology*, 6:166–169, 2011.

[51] S.F.J. Wickham, J. Bath, Y. Katsuda, M. Endo, K. Hidaka, H. Sugiyama, and A.J. Turberfield. A DNA-based molecular motor that can navigate a network of tracks. *Nature Nanotechnology*, 7:169–173, 2012.

[52] T-G. Cha, J. Pan, H. Chen, J. Salgado, X. Li, C. Mao, and J.H. Choi. A synthetic DNA motor that transports nanoparticles along carbon nanotubes. *Nature Nanotechnology*, 9:39–43, 2013.

[53] L. Qian, E. Winfree, and J. Bruck. Neural network computation with DNA strand displacement cascades. *Nature*, 475:368–372, 2011.

[54] L. Qian and E. Winfree. Scaling up digital circuit computation with DNA strand displacement cascades. *Science*, 332(6034):1196–1201, 2011.

[55] E.S. Andersen, M. Dong, M.M. Nielsen, K. Jahn, A. Lind-Thomsen, W. Mamdouh, K.V. Gothelf, F. Besenbacher, and J. Kjems. DNA origami design of dolphin-shaped structures with flexible tails. *ACS Nano*, 2:1213–1218, 2008.

[56] S.M. Douglas, H. Dietz, T. Liedl, B. Högberg, F. Graf, and W.M. Shih. Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature*, 459:414–418, 2009.

[57] C.E. Castro, F. Kilchherr, D.N. Kim, E.L. Shiao, T. Wauer, P. Wortmann, M. Bathe, and H. Dietz. A primer to scaffolded DNA origami. *Nature Methods*, 8:221–229, 2011.

[58] X-c. Bai, T.G. Martin, S.H.W. Scheres, and H. Dietz. Cryo-EM structure of a 3D DNA-origami object. *Proceedings of the National Academy of Sciences of the United States of America*, 109:20012–20017, 2012.

[59] E.S. Andersen, M. Dong, M.M. Nielsen, K. Jahn, R. Subramani, W. Mamdouh, M.M. Golas, B. Sander, H. Stark, C.L.P. Oliveira, J.S. Pedersen, V. Birkedal, F. Besenbacher, K.V. Gothelf, and J. Kjems. Self-assembly of a nanoscale DNA box with a controllable lid. *Nature*, 459:73–76, 2009.

[60] D. Han, S. Pal, J. Nangreave, Z. Deng, Y. Liu, and H. Yan. DNA origami with complex curvatures in three-dimensional space. *Science*, 332:342–346, 2011.

[61] H. Dietz, S.M. Douglas, and W.M. Shih. Folding DNA into twisted and curved nanoscale shapes. *Science*, 325:725–730, 2009.

[62] D. Han, S. Pal, Y. Liu, and H. Yan. Folding and cutting DNA into reconfigurable topological nanostructures. *Nat Nano*, 5:712–717, 2010.

[63] Y. Ke, G. Bellot, N.V. Voigt, E. Fradkov, and W.M. Shih. Two design strategies for enhancement of multilayer-DNA-origami folding: underwinding for specific intercalator rescue and staple-break positioning. *Chemical Science*, 3:2587–2597, 2012.

[64] S.M. Douglas, A.H. Marblestone, S. Teerapittayanon, A. Vazquez, G.M. Church, and W.M. Shih. Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nucleic Acids Research*, 37:5001–5006, 2009.

[65] D-N. Kim, F. Kilchherr, H. Dietz, and M. Bathe. Quantitative prediction of 3D solution shape and flexibility of nucleic acid nanostructures. *Nucleic Acids Research*, 40:2862–2868, 2012.

[66] B. Szabó and I. Babuška. *Finite Element Analysis*. Wiley, 1991.

[67] J.P.K Doye, T.E. Ouldridge, A.A. Louis, F. Romano, P. Šulc, C. Matek, B.E.K. Snodin, L. Rovigatti, J.S. Schreck, R.M. Harrison, and W.P.J. Smith. Coarse-graining DNA for simulations of DNA nanotechnology. *Physical Chemistry Chemical Physics*, 15:20395–20414, 2013.

[68] T.E. Ouldridge, R.L. Hoare, A.A. Louis, J.P.K Doye, J. Bath, and A.J. Turberfield. Optimizing DNA nanotechnology through coarse-grained modelling: a two-footed DNA walker. *ACS Nano*, 7:2479–2490, 2013.

[69] J. M. Arbona, J. Elezgaray, and J-P. Aimé. Modelling the folding of DNA origami. *Europhysics Letters*, 100, 2012.

[70] J. M. Arbona, J-P. Aimé, and J. Elezgaray. Cooperativity in the annealing of DNA origamis. *The Journal of Chemical Physics*, 138:015105, 2013.

[71] J.N. Zadeh, C.D. Steenburg, J.S. Bois, B.R. Wolfe, M.B. Pierce, A.R. Khan, R.M. Dirks, and N.A. Pierce. NUPACK: Analysis and design of nucleic acid systems. *Journal of Computational Chemistry*, 32:170–173, 2011.

[72] R.P. Goodman. NANEV: a program employing evolutionary methods for the design of nucleic acid nanostructures. *Biotechniques*, 38:548–550, 2005.

[73] R.T. Koehler and N. Peyret. Thermodynamic properties of DNA sequences: characteristic values for the human genome. *Bioinformatics*, 21:3333–3339, 2005.

[74] T.G. Martin and H. Dietz. Magnesium-free self-assembly of multi-layer DNA objects. *Nature Communications*, 3:1103, 2012.

[75] J.P.J. Sobczak, T.G. Martin, T. Gerling, and H. Dietz. Rapid folding of DNA into nanoscale shapes at constant temperature. *Science*, 338:1458–1461, 2012.

[76] S.F.J. Wickham. DNA origami, a substrate for the study of molecular motors. DPhil thesis, University of Oxford, 2011.

[77] P.R. Haddad and P.E. Jackson. *Ion chromatography - principles and applications*. Journal of Chromatography Library. Elsevier, 1990.

[78] C. Lin, S.D. Perrault, M. Kwak, F. Graf, and W.M. Shih. Purification of DNA-origami nanostructures by rate-zonal centrifugation. *Nucleic Acids Research*, 41:e40, 2013.

[79] G. Binnig, C.F. Quate, and C. Gerber. Atomic force microscope. *Physical Review Letters*, 56:930–933, 1986.

[80] V. Birkedal, M. Dong, M.M. Golas, B. Sander, E. S. Andersen, K. V. Gothelf, F. Besenbacher, and J. Kjems. Single molecule microscopy methods for the study of DNA origami structures. *Microscopy Research and Technique*, 74:688–698, 2011.

[81] V.J. Morris, A.R. Kirby, and A.P. Gunning. *Atomic Force Microscopy for Biologists*. Imperial College Press, 1999.

[82] D.B. Williams and C.B. Carter. *Transmission Electron Microscopy (series)*, volume I (Basics) and III (Imaging). Plenum Press, New York, 1996.

[83] J.R. Harris. *Negative Staining and Cryoelectron Microscopy*. Royal Microscopy Society Microscopy Handbooks. BIOS Scientific Publishers Ltd, 1997.

[84] Negative staining, protocol from Pathology Department, University of Oxford. `http://web.path.ox.ac.uk/~bioimaging/bitm/instructions_and_information/EM/neg_stain.pdf`.

[85] J.L.S. Milne, M.J. Borgnia, A. Bartesaghi, E.E.H. Tran, L.A. Earl, D.M. Schauder, J. Lengyel, J. Pierson, A. Patwardhan, and S. Subramaniam. Cryo-electron microscopy - a primer for the non-microscopist. *FEBS Journal*, 280:28–45, 2013.

[86] P.L.R. Bonner and A.J. Hargreaves. *Basic Bioscience Techniques: A Pocket Guide*. Wiley-Blackwell, 2011.

[87] J. Sambrook and D. W. Russell. *Molecular Cloning - A Laboratory Manual*. Cold Spring Harbor Laboratory Press, 3rd edition, 2001.

[88] G. Némethy and A. Ray. Solvent effects on the near-ultraviolet spectrum of phenol and its distribution in micellar solutions. *The Journal of Physical Chemistry*, 77:64–68, 1973.

[89] A.A. Greschner, K.E. Bujold, and H.F. Sleiman. Intercalators as molecular chaperones in DNA self-assembly. *Journal of the American Chemical Society*, 135:11283–11288, 2013.

[90] X. Wei, J. Nangreave, S. Jiang, H. Yan, and Y. Liu. Mapping the thermal behavior of DNA origami nanostructures. *Journal of the American Chemical Society*, 135:6165–6176, 2013.

[91] L Stryer. Fluorescence energy transfer as a spectroscopic ruler. *Annual Review of Biochemistry*, 47:819–846, 1978.

[92] J. Song, J-M. Arbona, Z. Zhang, L. Liu, E. Xie, J. Elezgaray, J-P. Aimé, K. V. Gothelf, F. Besenbacher, and M. Dong. Direct visualization of transient thermal response of a DNA origami. *Journal of the American Chemical Society*, 134:9844–9847, 2012.

[93] H. Said, V.J. Schüller, F.J. Eber, C. Wege, T. Liedl, and C. Richert. M1.3 - a small scaffold for DNA origami. *Nanoscale*, 5:284–290, 2013.

[94] E. Pound, J.R. Ashton, H.A. Becerril, and A.T. Woolley. Polymerase chain reaction based scaffold preparation for the production of thin, branched DNA origami nanostructures of arbitrary sizes. *Nano Letters*, 9:4302–4305, 2009.

[95] B. Högberg, T. Liedl, and W.M. Shih. Folding DNA origami from a double-stranded source of scaffold. *Journal of the American Chemical Society*, 131:9154–9155, 2009.

[96] Y. Yang, D. Han, J. Nangreave, Y. Liu, and H. Yan. DNA origami with double-stranded DNA as a unified scaffold. *ACS Nano*, 6:8209–8215, 2012.

[97] R. Holliday. Mechanism for gene conversion in fungi. *Genetical Research*, 5:282–304, 1964.

[98] D. Han, S. Jiang, A. Samanta, Y. Liu, and H. Yan. Unidirectional scaffold-strand arrangement in DNA origami. *Angewandte Chemie International Edition*, 52:9031–9034, 2013.

[99] R. Jungmann, T. Liedl, T.L. Sobey, W. Shih, and F.C. Simmel. Isothermal assembly of DNA origami structures using denaturing agents. *Journal of the American Chemical Society*, 130:10062–10063, 2008.

[100] Z. Zhang, J. Song, F. Besenbacher, M. Dong, and K.V. Gothelf. Self-assembly of DNA origami and single-stranded tile structures at room temperature. *Angewandte Chemie*, 52:9219–9223, 2013.

[101] J. Song, Z. Zhang, S. Zhang, L. Liu, Q. Li, E. Xie, K.V. Gothelf, F. Besenbacher, and M. Dong. Isothermal hybridization kinetics of DNA assembly of two-dimensional DNA origami. *Small*, 9:2954–2959, 2013.

[102] C. Myhrvold, M. Dai, P.A. Silver, and P. Yin. Isothermal self-assembly of complex DNA structures under diverse and biocompatible conditions. *Nano Letters*, 13:4242–4248, 2013.

[103] T.A. Taton, C.A. Mirkin, and R.L. Letsinger. Scanometric DNA array detection with nanoparticle probes. *Science*, 289:1757–1760, 2000.

[104] D. Zanchet, C.M. Micheel, W.J. Parak, D. Gerion, and A. P. Alivisatos. Electrophoretic isolation of discrete Au nanocrystal/DNA conjugates. *Nano Letters*, 1:32–35, 2001.

[105] J. Sharma, R. Chhabra, H. Yan, and Y. Liu. A facile in situ generation of dithiocarbamate ligands for stable gold nanoparticle-oligonucleotide conjugates. *Chemical Communications*, pages 2140–2142, 2008.

[106] J. Sharma, R. Chhabra, Y. Liu, Y. Ke, and H. Yan. DNA-templated self-assembly of two-dimensional and periodical gold nanoparticle arrays. *Angewandte Chemie International Edition*, 45:730–735, 2006.

[107] B. Högberg and H. Olin. DNA-scaffolded nanoparticle structures. *Journal of Physics: Conference Series*, 61:458, 2007.

[108] J. Sharma, R. Chhabra, C. S. Andersen, K. V. Gothelf, H. Yan, and Y. Liu. Toward reliable gold nanoparticle patterning on self-assembled DNA nanoscaffold. *Journal of the American Chemical Society*, 130:7820–7821, 2008.

[109] B. Ding, Z. Deng, H. Yan, S. Cabrini, R. N. Zuckermann, and J. Bokor. Gold nanoparticle self-similar chain structure organized by DNA origami. *Journal of the American Chemical Society*, 132:3248–3249, 2010.

[110] A.M. Hung, C.M. Micheel, L.D. Bozano, L.W. Osterbur, G.M. Wallraff, and J.N. Cha. Large-area spatially ordered arrays of gold nanoparticles directed by lithographically confined DNA origami. *Nature Nanotechnology*, 5:121–126, 2010.

[111] Z. Zhao, E. L. Jacovetty, Y. Liu, and H. Yan. Encapsulation of gold nanoparticles in a DNA origami cage. *Angewandte Chemie International Edition*, 50:2041–2044, 2011.

[112] G. P. Acuna, F. M. Möller, P. Holzmeister, S. Beater, B. Lalkens, and P. Tinnefeld. Fluorescence enhancement at docking sites of DNA-directed self-assembled nanoantennas. *Science*, 338:506–510, 2012.

[113] D. Rand, V. Ortiz, Y. Liu, Z. Derdak, J.R. Wands, M. Tatíček, and C. Rose-Petruck. Nanomaterials for X-ray imaging - gold nanoparticle enhancement of X-ray scatter imaging of hepatocellular carcinoma. *Nano Letters*, 11:2678–2683, 2011.

[114] A.C. Pearson, J. Liu, E. Pound, B. Uprety, A.T. Woolley, R.C. Davis, and J.N. Harb. DNA origami metallized site specifically to form electrically conductive nanowires. *Journal of Physical Chemistry B*, 116:10551–10560, 2012.

[115] R. S. Sørensen, A. H. Okholm, D. Schaffert, A. L. Bank Kodal, K. V. Gothelf, and J. Kjems. Enzymatic ligation of large biomolecules to DNA. *ACS Nano*, 7:8098–8104, 2013.

[116] S.D. Perrault and W.M. Shih. Virus-inspired membrane encapsulation of DNA nanostructures to achieve in vivo stability. *ACS Nano*, 8:5132–5140, 2014.

[117] Q. Jiang, C. Song, J. Nangreave, X. Liu, L. Lin, D. Qiu, Z-G. Wang, G. Zou, X. Liang, H. Yan, and B. Ding. DNA origami as a carrier for circumvention of drug resistance. *Journal of the American Chemical Society*, 134:13396–13403, 2012.

[118] Y. Fu, D. Zeng, J. Chao, Y. Jin, Z. Zhang, H. Liu, D. Li, H. Ma, Q. Huang, K. V. Gothelf, and C. Fan. Single-step rapid assembly of DNA origami nanostructures for addressable nanoscale bioreactors. *Journal of the American Chemical Society*, 135:696–702, 2012.

[119] N.A.W. Bell, C.R. Engst, M. Ablay, G. Divitini, C. Ducati, T. Liedl, and U.F. Keyser. DNA origami nanopores. *Nano Letters*, 12:512–517, 2012.

[120] N. D. Derr, B. S. Goodman, R. Jungmann, A. E. Leschziner, W. M. Shih, and S. L. Reck-Peterson. Tug-of-war in motor protein ensembles revealed with a programmable DNA origami scaffold. *Science*, 338:662–665, 2012.

[121] M. Langecker, V. Arnaut, T.G. Martin, J. List, S. Renner, M. Mayer, H. Dietz, and F.C. Simmel. Synthetic lipid membrane channels formed by designed DNA nanostructures. *Science*, 338:932–936, 2012.

[122] R. Jungmann, M. Scheible, A. Kuzyk, G. Pardatscher, C. E. Castro, and F. C. Simmel. DNA origami-based nanoribbons: assembly, length distribution, and twist. *Nanotechnology*, 22:275301, 2011.

[123] Z. Zhao, Y. Liu, and H. Yan. Organizing DNA origami tiles into larger structures using preformed scaffold frames. *Nano Letters*, 11:2997–3002, 2011.

[124] Z. Li, M. Liu, L. Wang, J. Nangreave, H. Yan, and Y. Liu. Molecular behavior of DNA origami in higher-order self-assembly. *Journal of the American Chemical Society*, 132:13545–13552, 2010.

[125] Z. Zhao, H. Yan, and Y. Liu. A route to scale up DNA origami using DNA tiles as folding staples. *Angewandte Chemie International Edition*, 49:1414–1417, 2010.

[126] A. Rajendran, M. Endo, Y. Katsuda, K. Hidaka, and H. Sugiyama. Programmed two-dimensional self-assembly of multiple DNA origami jigsaw pieces. *ACS Nano*, 5:665–671, 2011.

[127] S. Woo and P.W.K. Rothemund. Programmable molecular recognition based on the geometry of DNA nanostructures. *Nature Chemistry*, 3:620–627, 2011.

[128] J. Lee, S. Hamada, S. U. Hwang, R. Amin, J. Son, S. R. Dugasani, S. Murata, and S. H. Park. Quantitative analysis of molecular-level DNA crystal growth on a 2D surface. *Scientific Reports*, 3:2115, 2013.

[129] D. Han, S. Pal, Y. Yang, S. Jiang, J. Nangreave, Y. Liu, and H. Yan. DNA Gridiron nanostructures based on four-arm junctions. *Science*, 339:1412–1415, 2013.

[130] K. V. Gothelf. LEGO-like DNA structures. *Science*, 338:1159–1160, 2012.

[131] B. Wei, M. Dai, and P. Yin. Complex shapes self-assembled from single-stranded DNA tiles. *Nature*, 485:623–626, 2012.

[132] Y. Ke, L.L. Ong, W.M. Shih, and P. Yin. Three-dimensional structures self-assembled from DNA bricks. *Science*, 338:1177–1183, 2012.

[133] J.J. Li and W. Tan. A single DNA molecule nanomotor. *Nano Letters*, 2:315–318, 2002.

[134] Y-T. Lai, D. Cascio, and T.O. Yeates. Structure of a 16-nm cage designed by using protein oligomers. *Science*, 336:1129, 2012.

[135] E.H.C. Bromley, K.J. Channon, P.J.S. Sing, Z.N. Mahmoud, E.F. Banwell, M.F. Butler, M.P. Crump, T.R. Dafforn, M.R. Hicks, J.D. Hirst, A. Rodger, and D.N. Woolfson. Assembly pathway of a designed $\alpha$-helical protein fiber. *Biophysical Journal*, 98:1668–1676, 2010.

[136] A.L. Boyle, E.H.C. Bromley, G.J. Bartlett, R.B. Sessions, T.H. Sharp, C.L. Williams, P.M.G. Curmi, N.R. Forde, H. Linke, and D.N. Woolfson. Squaring the circle in peptide assembly: from fibers to discrete nanostructures by *de novo* design. *Journal of the American Chemical Society*, 134:15457–15467, 2012.

[137] J.M. Fletcher, R.L. Harniman, F.R.H. Barnes, A.L. Boyle, A. Collins, J. Mantell, T.H. Sharp, M. Antognozzi, P.J. Booth, N. Linden, M.J. Miles, R.B. Sessions, P. Verkade, and D.N. Woolfson. Self-assembling cages from coiled-coil peptide modules. *Science*, 340:595–599, 2013.

[138] S Howorka. Rationally engineering natural protein assemblies in nanobiotechnology. *Current Opinion in Biotechnology*, 22:485–491, 2011.

[139] P.X. Guo. The emerging field of RNA nanotechnology. *Nature Nanotechnology*, 5:833–842, 2010.

[140] C.J. Delebecque, A.B. Lindner, P. A. Silver, and F.A. Aldaye. Organization of intracellular reactions with rationally designed RNA assemblies. *Science*, 333:470–474, 2011.

[141] A.J. Dzieciol and S. Mann. Designs for life: protocell models in the laboratory. *Chemical Society Reviews*, 41:79–85, 2012.

[142] M. Li, R.L. Harbron, J.V.M. Weaver, B.P. Binks, and S. Mann. Electrostatically gated membrane permeability in inorganic protocells. *Nature Chemistry*, 5:529–536, 2013.

[143] H. G. Hansma and D. E. Laney. DNA binding to mica correlates with cationic radius: Assay by atomic force microscopy. *Biophysical Journal*, 70:1933–1939, 1996.

[144] regionprops: measure properties of image regions. `http://www.mathworks.co.uk/help/images/ref/regionprops.html?nocookie=true`.

[145] F. Ostendorf, C. Schmitz, S. Hirth, A. Kühnle, J. J. Kolodziej, and M. Reichling. How flat is an air-cleaved mica surface? *Nanotechnology*, 19:305705, 2008.

[146] C. Levinthal. How to fold graciously. In *Mössbauer Spectroscopy in Biological Systems*, volume 67. Univ. of Illinois Bulletin.

[147] J. Kubelka, J. Hofrichter, and W. A. Eaton. The protein folding 'speed limit'. *Current Opinion in Structural Biology*, 14:76–88, 2004.

[148] C. Levinthal. Are there pathways for protein folding? *Journal De Chimie Physique Et De Physico-Chimie Biologique*, 65:44, 1968.

[149] K.A. Dill and H.S. Chan. From Levinthal to pathways to funnels. *Nature Structural Biology*, 4:10–19, 1997.

[150] K.A. Dill and J.L. MacCallum. The protein-folding problem, 50 years on. *Science*, 338:1042–1046, 2012.

[151] M.C. Leake. *Single-Molecule Cellular Biophysics*. Cambridge University Press, 2013.

[152] I. Teraoka. *Polymer solutions: an introduction to physical properties*. Wiley (Google Books), 2002.

[153] Water - dynamic and kinematic viscosity. `http://www.engineeringtoolbox.com/water-dynamic-kinematic-viscosity-d_596.html`.

[154] B. Tinland, A. Pluen, J. Sturm, and G. Weill. Persistence length of single-stranded DNA. *Macromolecules*, 30:5763–5765, 1997.

[155] J.C. Russ and J.C. Russ. *Introduction to Image Processing and Analysis*. Taylor & Francis, 2008.

[156] L. O'Gorman, M. J. Sammon, and M. Seul. *Practical Algorithms for Image Analysis (Description, Examples, Programs and Projects)*. Cambridge University Press, 2008.

[157] University of Cambridge. AFM (Dissemination of IT for the Promotion of Materials Science: Teaching and Learning Packages: Atomic Force Microscopy). `http://www.doitpoms.ac.uk/tlplib/afm/index.php`, 2009.

[158] E. Betzig. Proposed method for molecular optical imaging. *Optics Letters*, 20:237–239, 1995.

[159] R.E. Thompson, D.R. Larson, and W.W. Webb. Precise nanometer localization analysis for individual fluorescent probes. *Biophysical Journal*, 82:2775–2783, 2002.

[160] A. Yildiz and P. R. Selvin. Fluorescence imaging with one nanometer accuracy: Application to molecular motors. *Accounts of Chemical Research*, 38:574–582, 2005.

[161] S. Uphoff, R. Reyes-Lamothe, F.G. de Leon, D.J. Sherratt, and A.N. Kapanidis. Single-molecule DNA repair in live bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 110:8063–8068, 2013.

[162] W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35:399–458, 2003.

[163] C.B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181:223–230, 1973.

[164] M.J. Bennett, S. Choe, and D. Eisenberg. Domain swapping: entangling alliances between proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 91:3127–3131, 1994.

[165] Y. Liu and D. Eisenberg. 3D domain swapping: as domains continue to swap. *Protein Science*, 11:1285–1299, 2002.

[166] M. Bergdoll, M.H. Remy, C. Cagnon, J.M. Masson, and P. Dumas. Proline-dependent oligomerization with arm exchange. *Structure*, 5:391–401, 1997.

[167] F. Rousseau, J.W.H. Schymkowitz, H.R. Wilkinson, and L.S. Itzhaki. Three-dimensional domain swapping in p13suc1 occurs in the unfolded state and is controlled by conserved proline residues. *Proceedings of the National Academy of Sciences of the United States of America*, 98:5596–5601, 2001.

[168] T. Omabegho, R. Sha, and N.C. Seeman. A bipedal DNA Brownian motor with coordinated legs. *Science*, 324:67–71, 2009.

[169] T. Nakamura, K. Makabe, K. Tomoyori, K. Maki, A. Mukaiyama, and K. Kuwajima. Different folding pathways taken by highly homologous proteins, goat $\alpha$-lactalbumin and canine milk lysozyme. *Journal of Molecular Biology*, 396:1361–1378, 2010.

[170] D. Baker. A surprising simplicity to protein folding. *Nature*, 405:39–42, 2000.

[171] K.W. Plaxco, K.T. Simons, and D. Baker. Contact order, transition state placement and the refolding rates of single domain proteins. *Journal of Molecular Biology*, 277:985–994, 1998.

[172] A. Zarrine-Afsar, S.M. Larson, and A.R. Davidson. The family feud: do proteins with similar structures fold via the same pathway? *Current Opinion in Structural Biology*, 15:42–49, 2005.

[173] Source Life Biosciences. M13 primers. `http://lifesciences.sourcebioscience.com/genomic-services/sanger-sequencing-service/primers-in-stock.aspx`.

[174] New England Biolabs. Optimizing restriction endonuclease reactions. `https://www.neb.com/tools-and-resources/usage-guidelines/optimizing-restriction-endonuclease-reactions`.

[175] W.B. Sherman and N.C. Seeman. A precisely controlled DNA biped walking device. *Nano Letters*, 4(7):1203–1207, 2004.

[176] J-S. Shin and N.A. Pierce. A synthetic DNA walker for molecular transport. *Journal of the American Chemical Society*, 126:10834–10835, 2004.

[177] P. Yin, H. Yan, X.G. Daniell, A.J. Turberfield, and J.H. Reif. A unidirectional DNA walker that moves autonomously along a track. *Angewandte Chemie International Edition*, 43:4906–4911, 2004.

[178] H. Yan, X. Zhang, Z. Shen, and N. C. Seeman. A robust DNA mechanical device controlled by hybridization topology. *Nature*, 415:62–65, 2002.

[179] C.H. Lu, A. Cecconello, J. Elbaz, A. Credi, and I. Willner. A three-station DNA catenane rotary motor with controlled directionality. *Nano Letters*, 13:2303–2308, 2013.

[180] E. Wasserman. The preparation of interlocking rings: a catenane. *Journal of the American Chemical Society*, 82:4433–4434, 1960.

[181] A. Rajendran, M. Endo, K. Hidaka, and H. Sugiyama. Direct and real-time observation of rotary motion of a DNA nanomechanical device. *Journal of the American Chemical Society*, 135:1117–1123, 2013.

[182] H. Noji, R. Yasuda, M. Yoshida, and K. Kinosita. Direct observation of the rotation of $F_1$-ATPase. *Nature*, 386:299–302, 1997.

[183] S. Hernández-Ainsa, N. A. W. Bell, V. V. Thacker, K. Göpfrich, K. Misiunas, M. E. Fuentes-Perez, F. Moreno-Herrero, and Ulrich F. Keyser. DNA origami nanopores for controlling DNA translocation. *ACS Nano*, 7:6024–6030, 2013.

[184] The Nobel Prize in Chemistry 1997. `http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1997/`.

[185] P.D. Boyer. ATP-synthase - past and future. *Biochimica et Biophysica Acta*, 1365:3–9, 1998.

[186] J.P. Abrahams, A.G.W. Leslie, R. Lutter, and J.E. Walker. Structure at 2.8Å resolution of $F_1$-ATPase from bovine heart mitochondria. *Nature*, 370:621–628, 1994.

[187] R.M. Berry. ATP synthesis: The world's smallest wind-up toy. *Current Biology*, 15:385–387, 2005.

[188] F. Wankel and E. Hoeppner. Rotary internal combustion engine. US Patent: US2988065A, 1961.

[189] D.E. Cole. The Wankel engine. *Scientific American*, 227:14–23, 1972.

[190] J. Zhang, A.P. Sergeeva, M. Sparta, and A.N. Alexandrova. $B_{13}^+$: A photodriven molecular Wankel engine. *Angewandte Chemie International Edition*, 51:8512–8515, 2012.