Construction of a Synthetic Ribosome using DNA as the Building Material

Parminder Lally

Magdalen College

University of Oxford

Thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2010

Abstract

Construction of a Synthetic Ribosome using DNA as the Building Material

Parminder Lally

Magdalen College

Thesis submitted for the degree of Doctor of Philosophy, Michaelmas 2010

This thesis forms part of an ongoing project in the DNA Group to build and operate a synthetic ribosome. We present two synthetic ribosome designs that can be combined with DNA-templated chemistry to generate libraries of functional synthetic small molecules. In Chapter 2 we use the DNA strand displacement technique to construct a mechanism that is capable of moving along a DNA track. We explore ways to control the speed and the driving force of the mechanism, and present a mathematical model of the system. We discuss the ability of the design to incorporate chemically-functionalised DNA strands. In Chapter 3 we use a 2D DNA origami tile as the basis of the synthetic ribosome mechanism. Functionalised DNA strands are arranged on the surface of the tile, and we demonstrate the ability to template reactions between the strands, and discuss the possibility of creating a library of distinct chemical products from a single origami tile.

Acknowledgements

I would like to thank the many people who have contributed in some way to my work and have supported me through the course of my D.Phil:

- Prof. Andrew Turberfield, my supervisor, for providing me with the opportunity to work in this field, scrutinising my results and demonstrating how to be a good researcher;
- Dr. Jon Bath, for patiently explaining how to carry out many of the experimental procedures I used over the last three years, as well as always being available to discuss ideas and data;
- Dr. Mireya McKee, for her guidance on the chemistry experiments;
- Shelley Wickham, for being the resident DNA origami expert, and in particular for helping me to acquire AFM images of my origami tiles;
- My collaborators, particularly Diluar Khan at Warwick University for synthesising the click adaptors, and Dr. Rachel O'Reilly for her advice on how to conduct the click chemistry;
- The rest of the DNA group, who make the office a lively place to work;
- My brother Jagjeet, for keeping my spirits high through the last weeks in the lab and for travelling to Oxford to help me to submit;
- And finally, to Oscar, for always being close at hand to fix errors in my MATLAB and LaTeX code, and for his unfaltering and continuing support.

Contents

Chapter 1

Introduction

1.1 DNA as a Novel Building Material

For over fifty years, the DNA molecule has been appreciated for its role as the keeper of genetic information, but it is only relatively recently that scientists have begun to wonder whether the molecule can be used for other, non-biological purposes. Many applications of DNA molecule analysis, such as forensic science and genetic testing, are based around the ability to read back the sequence of the molecule and understand what information is encoded within it. Whilst this is in itself an interesting and rapidly advancing research area, it is not perhaps the most imaginative and creative use of the properties of DNA and it is of no surprise, therefore, that interest has developed in using DNA outside of a purely biological setting. DNA nanotechnology was born over 20 years ago with the first demonstration of the molecule as a construction tool [Seeman, 1982], and the idea has gained impetus ever since, burgeoning into a field that attracts interest from chemists to computer scientists alike.

The pioneers of DNA nanotechnology could not have made their advances without an understanding of those properties of the molecule that could be exploited for their purposes. The use of DNA in this field is founded upon the double helix

(duplex) and the idea of complementarity. Each single strand of DNA (ssDNA) is a polymer composed of covalently bound units (nucleotides) consisting of a sugar-phosphate backbone and one of the four bases adenine, thymine, cytosine or guanine (A, T, C, G). It is the sequence of bases which encodes information; the amount of information stored within DNA is around $1\,\mathrm{bit/nm^3}$ [Condon, 2006]. The backbone of the DNA is made from alternating phosphate and sugar residues. The sugars are joined together by phosphate groups that form bonds between the 3rd and 5th carbon atoms of the adjacent sugars. This asymmetric bonding means DNA has an orientation, conventionally denoted by labelling its ends 3' and 5' (and in diagrams by an arrow pointing from 5' to 3'). When two ssDNA hybridise to form a duplex, the strands must be anti-parallel and the complementary bases must align according to the Watson–Crick base pairings of A to T and C to G [Watson and Crick, 1953]. Hydrogen bonds between the base pairs help to stabilise the duplex. The two strands are also held together partially due to hydrophobicity and, since the hydrogen bonds holding the two ssDNA together are not covalent, they can be broken and rejoined relatively simply. This means that hybridisation is reversible, a particularly useful property of DNA that is discussed further in Section 1.3. The strength of the hybridisation depends on various factors, including temperature, salt concentration in the solution [Tan and Chen, 2006], the number of basepairs formed and the strand concentration.

The melting temperature $\mathrm{T_m}$ indicates the temperature above which the duplex structure of DNA is no longer stable. The A-T base pairing creates two hydrogen bonds, and the G-C pairing involves three hydrogen bonds, thus the G-C pair is stronger. Both the percentage of G-C base pairs and the overall length of the duplex determine the strength of the association between two ssDNA. Non-Watson–Crick base pairings can occur, but the stability of the mismatched bases is lower than that of perfect-matching pairs: $GC > AT > GG > GT \approx GA > TT \approx AA > TC \geq$

$AC \geq CC$ [SantaLucia-Jr and Hicks, 2004]. A single mismatch or several within a duplex causes a bulge to be formed, which can affect the overall stability of the duplex [Zhu and Wartell, 1999], [Tanaka et al., 2004].

The persistence length of DNA is also an important factor in its use as a nanoscale building material. The persistence length is a mechanical property of polymers that describes the stiffness of the molecule. Single strands of DNA are very flexible; the persistence length is a few bases. Double-stranded DNA however, can be approximated as a rigid rod at the length scale of most DNA structures, since it has a persistence length of around 150 base pairs [Bustamante et al., 1994]. The majority of the DNA constructs presented or reviewed here use the topologically line-like DNA duplex as the cornerstone of the design, but the ability of DNA to form branched structures allows for much more complicated builds, as Seeman first proposed [Seeman, 1982]. Starting from a simple three- or four-way Holliday junction [Holliday, 1964], it has become possible to create more complicated double and triple crossover complexes, as shown in Figure 1.1 [Fu and Seeman, 1993], [Winfree et al., 1998], [LaBean et al., 2000], which have been cleverly designed to encourage tiling of a single complex into a much larger 2D array (Figure 1.2).

Regions of unpaired nucleotides at the end of a DNA strand are known as "sticky-ends", which can be exploited for a variety of purposes, including building larger structures. One of the key aims for nanotechnologists is to construct nanoscale objects using the principle of self-assembly. This is simply the idea that construction components can create larger objects via their selective affinity. The assembly takes place in solution, and the components are brought into contact through thermal motion. For example, the careful design of the sticky-end sequences of a branched molecule is what enables the self-assembly of the uniform 2D lattice shown in Figure 1.2. Seeman's revolutionary idea of manipulating the DNA molecule into branched con-

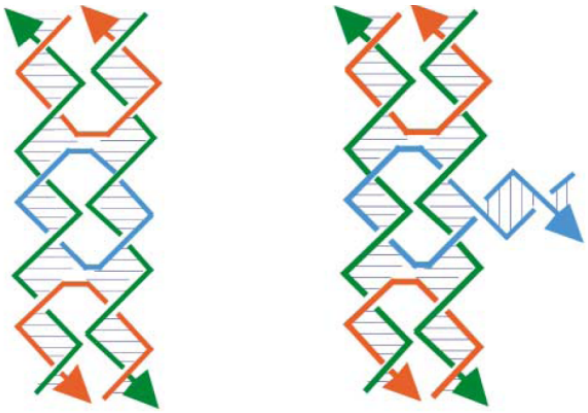


Figure 1.1: Two DNA structural motifs.
The DX molecule is made from the reciprocal exchange of individual DNA strands. A branched variant is formed using a DNA hairpin. Figure taken from [Seeman, 2003].

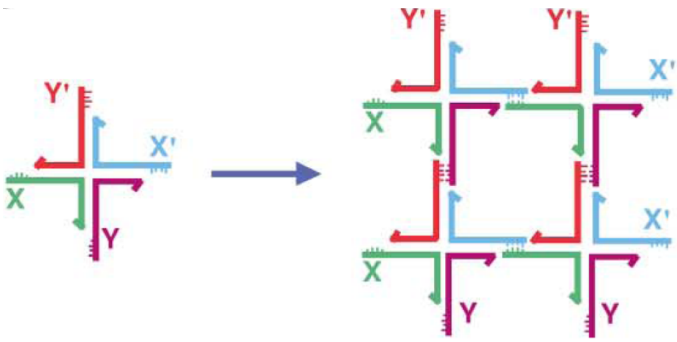


Figure 1.2: An example of a branched structure, and its assembly into larger 2D arrays.
Seeman postulated that it was possible for DNA strands to hybridise into junctioned structures (left) instead of into the more familiar linear duplexes [Seeman, 1982]. Furthermore, if such branched molecules contained regions of unpaired nucleotides known as "sticky-ends", then the formation of periodic lattices is possible. The sticky-ends X and X' are complementary to each other, as are Y and Y', such that upon hybridisation the formation of a larger structure is enabled (right). This new structure also contains the same sticky-ends. Figure taken from [Seeman, 2003].

structs paved the road to developing a wide variety of rigid 3D DNA nanostructures [Chen and Seeman, 1991], [Shih et al., 2004], [Goodman et al., 2005]. This notion enabled the building of nanoscale electronic circuits [Robinson and Seeman, 1987], [Seelig et al., 2006] and allowed the conception of DNA computation [Winfree, 1995], [Benenson, 2001], [Shin and Pierce, 2004a]. Reviews of the numerous ways in which DNA molecules have been manipulated to fashion nanoscale structures and dynamic devices are supplied by [Seeman, 2003], [Seeman, 2007], [Bath and Turberfield, 2007]. All of these examples use DNA strand complementarity to form the desired structure.

One reason behind the desire to create DNA nanostructures is to enable the very precise and controlled construction of materials with specific properties, such as the exact positioning of molecular electronic components to produce DNA nanowires [Braun et al., 1998]. Custom synthesised DNA is limited to 200 nt[1] at present by the synthesis methods used, which means much more material is required for building large-scale structures. Therein lies a problem: synthesis of such structures requires interactions between many short DNA strands, which is sensitive to the stoichiometry of the strands. The resulting yield of fully formed structures is low, and this limits the possibility of creating structures that are many orders of magnitude larger than their individual components. However, there is a newly developed alternative approach, namely that of DNA origami.

1.2 Large-Scale Nanostructures: DNA Origami

The first instance of a nanostructure built from the controlled folding of one long DNA strand is the octahedron [Shih et al., 2004], whose main component was a synthetic 1669 nt DNA template strand constructed from shorter synthetic oligonucleotides

[1] nt is used throughout this thesis as an abbreviation for "nucleotide", the structural unit of DNA strands.

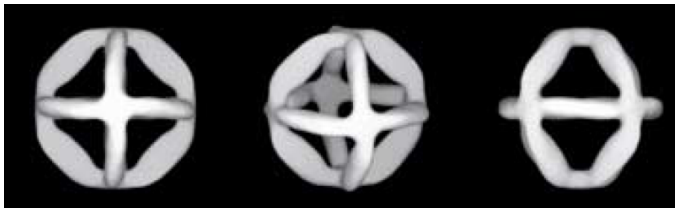


Figure 1.3: Folding a long DNA strand to form nanostructures.
A 1669 nt DNA strand was designed to fold into an octahedron when combined with five shorter DNA strands that effectively lock the structure. The octahedra were visualised using cryo-electron microscopy, and these images are the result of single-particle reconstructions. Figure taken from [Shih et al., 2004].

(using polymerase chain reaction (PCR)-based assembly). When combined with five 40 nt DNA strands, the long template was able to fold by self-interaction into the form of the octahedron, and the shorter strands locked the strand into that configuration (see Figure 1.3 for the outcome). The high yield of correctly formed octahedra suggested that creating large-scale nanostructures depends on designing the way in which long DNA template strands fold up into a desired two- or three-dimensional shape.

A "one-pot" method of folding a long ssDNA into a shape yielded one of best examples of self-assembling DNA nanostructures — DNA origami [Rothemund, 2006]. Here, instead of synthesising a long strand from shorter segments, Rothemund uses a naturally existing long ssDNA as the basis of his designs. The same template strand was used to self-assemble six different planar shapes that are approximately 100 nm in diameter, demonstrating the versatility of the process and its ability to produce large nanostructures. In contrast to Shih's octahedron, an origami structure is assembled by interactions between the template and short, synthetic "staples" — self-interactions of the template play no role.

The design of a DNA origami is a three-step process. Firstly, it is necessary to

construct a geometric model of a DNA structure that resembles the desired origami shape (Figure 1.4(a) left). The sketched shape is filled from top to bottom with an even number of parallel double helices (approximated by cylinders for ease of design), which must be an integer number of turns in length.[2] The helices are held together in this formation using a periodic array of staple strand "crossovers", which mark the positions at which staple strands running along one helix switch to running along an adjacent helix (shown by the blue points in Figure 1.4(a) left). The geometric model uses 16 bases to represent 1.5 turns of DNA (10.67 bp/turn), which is a departure from the natural length for a single turn of DNA (10.5 bp/turn). Furthermore, the DNA helix is asymmetric. The backbones of the two DNA strands that form a helix are closer together on one side of the helix than the other, such that a major groove occurs where the backbones are far apart and the minor groove occurs where they are close together. The asymmetry and geometric model combined cause the helical domains between crossovers to be slightly overtwisted. To reduce the strain caused by the twisting, it is necessary to factor into the design the orientation of the minor and major grooves of the helices. The position of the crossovers are related by a glide symmetry[3], such that the minor groove of the DNA faces alternating directions in alternating columns of periodic crossovers. As a result, the origami are, on average, flat. Without this glide symmetry to compensate for the overtwisting, the nanostructure may resemble a tube rather than a sheet. In addition, due to electrostatic repulsion between the DNA duplexes, the parallel helices are not closely packed in the origami structure. The gap between the helices depends on the spacing of the

[2]Rothemund uses a circular template strand for the origami design, as described later in this section, and an even number of double helices ensures that the start and end points of the rastered template are close together (shown in Figure 1.4(a) right). The length of each double helix row must be an integer number of turns to accomodate the natural twist of the DNA.

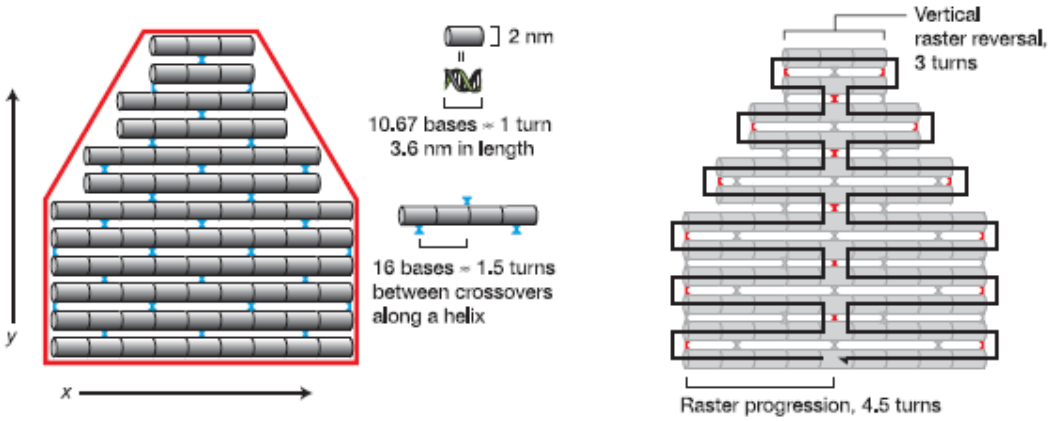
[3]Glide symmetry is the combined result of a reflection in a line and a translation along the same line. A simple example is of a reflection in the x-axis, followed by a translation by one unit along the same axis. In coordinates this is $(x, y) \mapsto (x + 1, -y)$.

crossovers, and it was found that a 1 nm gap exists when crossovers are positioned every 1.5 turns along alternating sides of a helix [Rothemund, 2006].

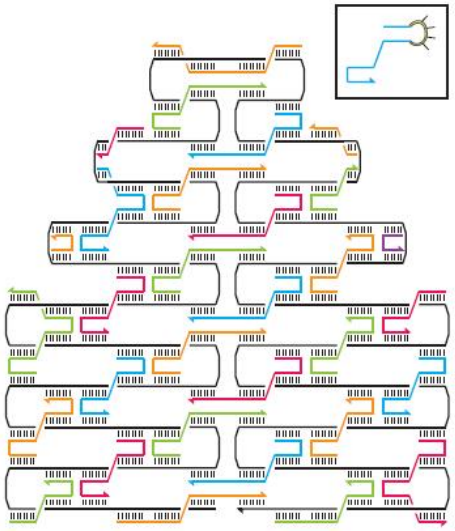
The second step is to integrate the ssDNA template into the model. A long ssDNA (scaffold) is folded back and forth in a raster pattern to fill the shape. As the scaffold moves from one helix to another, further crossovers are created. In order for the scaffold to raster from one helix to the next, the distance between successive scaffold crossovers must be an odd number of half turns, in order to accomodate the natural twist of the DNA (Figure 1.4(a) right).

The final step in the design process is to generate the "staple strand" sequences (the multiple short ssDNA which lock or "staple" the scaffold into place as shown in Figure 1.4(b)). Various software packages now exist to aid the design of an origami and the generation of staple strand sequences, with caDNAno being one of the more popular [Douglas et al., 2009b].

Rothemund chose circular genomic DNA from the virus M13mp18 as the scaffold for the DNA origami, since its natural state is single-stranded, and it has a 7249 nt sequence that can be used to create large shapes. An excess of the 200-250 staple strands required to form an origami are mixed with the scaffold, denatured at 95 °C and then cooled over 1.5 hours to 20 °C. The origami are typically visualised using atomic force microscopy (AFM) (see Figure 1.5). The yield of correctly formed origami rectangles was around 90 %, and it is this fact combined with the relative ease of production that has made DNA origami tiles so popular in recent years. The rectangular tile has been used or proposed for, among many other applications, the controlled self-assembly of much larger 2D structures [Endo et al., 2010], the biological equivalent of a silicon breadboard [Kim et al., 2010], and for the precise positioning of biological objects into arrays [Stephanopoulos et al., 2010], [Zhang et al., 2010], [Bui et al., 2010]. The



(a) Left: Step 1. An origami shape (red) is approximated by parallel duplexes that are joined by staple strand crossovers (blue). Right: Step 2. The scaffold (black) runs through every helix, and further crossover points are formed (red).



(b) Step 3. The design is completed by generating the staple strands that fix the scaffold into the desired geometry.

Figure 1.4: Design of a DNA origami.

Three steps must be completed in order to design an origami structure. Firstly, the desired shape must be filled with parallel double helices, which are joined together by periodic crossovers. Secondly, the scaffold strand is rastered to fill the shape, and in doing so, further crossovers are obtained. Thirdly, the staple strand sequences (which fix the scaffold into position) are generated. Typically, staples are 32-mers and span three helices. Figures adapted from [Rothemund, 2006].

versatility of DNA origami makes it an attractive tool for DNA nanotechnologists to either build with or build upon.

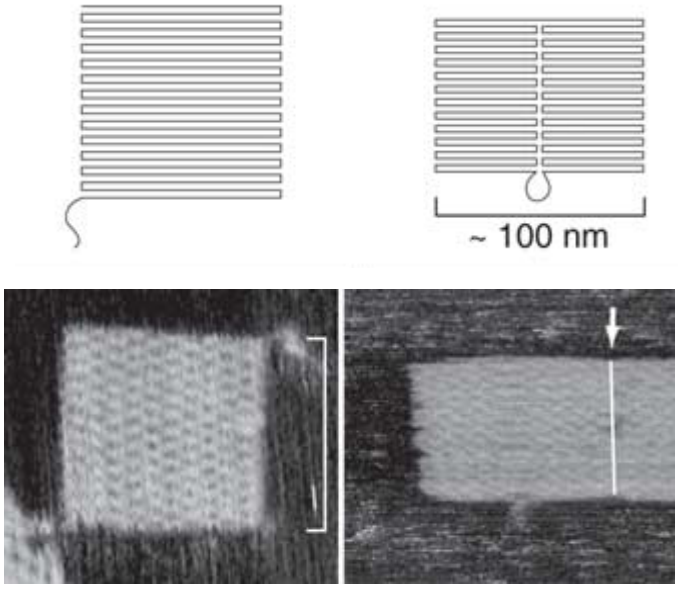


Figure 1.5: Two examples of DNA origami.
The images on the left correspond to an origami square, designed simply by rastering the scaffold strand back and forth within the shape. The images on the right are of the origami rectangle, which contains within it a central seam where the scaffold strand changes direction. The yield of successful squares was only 13 %, whereas the yield was 90 % for the rectangle, a significant difference attributed to the stabilisation of the origami from the extra crossovers caused by the central seam. Figure adapted from [Rothemund, 2006].

1.3 From Static to Dynamic DNA Machines

The ability to manipulate DNA into multi-dimensional structures has been reviewed, but these objects largely ignore a particular property of the molecule that makes it so useful to nanotechnologists. As mentioned in Section 1.1, the hydrogen bonds that form between the base pairs of a duplex are non-covalent, and so can be readily broken. Thus, the hybridisation of two strands forming a duplex is entirely reversible, and this fact has been used to develop a variety of strand displacement mechanisms

that turn static DNA constructions into dynamic machine-like creations (see Figure 1.6 for a simple example). The desire to build dynamic DNA machines stems from an interest in mimicking Nature's own biological machines for the purpose of understanding the naturally occurring systems in more detail. However, it has also sparked the development of devices that have new applications, such as DNA computation.

There are various examples of DNA machines being operated by some chemical stimulus that causes a conformational change in the DNA [Mao et al., 1999], [Fu and Seeman, 1993], [Chen et al., 2004]. However, it is those machines that are both constructed of and operated by DNA that are of special interest here. The first instance of such a device was the molecular tweezers [Yurke et al., 2000], shown in Figure 1.7. Sticky-end interactions are used to open and close the tweezers. The tweezers begin in the open state, and are formed from the hybridisation of strands A-C. Strand A is labelled with fluorescent dyes that form a donor-acceptor pair (Appendix A.1.1), and the interaction between the dyes is used to determine whether the tweezers are in the closed or open state. Strands B and C both have sticky-end regions to which the fuel strand (F) is complementary. Once the fuel is added, the sticky-end hybridisation forces the tweezers to close. In order to return to the open

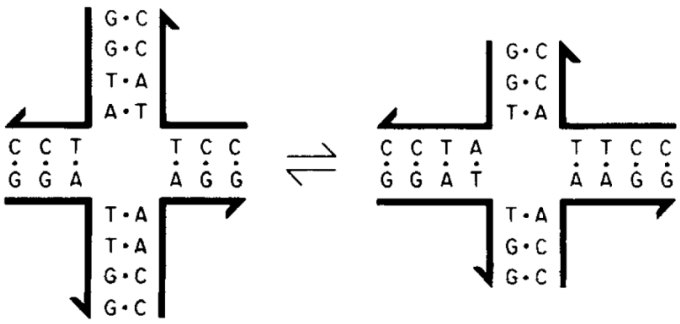


Figure 1.6: A very simple dynamic DNA structure.

A static DNA Holliday junction can be made mobile by designing the sequences such that the structure continually moves between one or other configuration. Figure taken from [Seeman, 1982].

state, the process must be reversed, and this requires a second sticky-end interaction. The fuel strand contains unpaired nucleotides to which an antifuel strand ($\overline{\text{F}}$) can hybridise and pull away strand F from the tweezers entirely. The device can cycle between open and closed states simply by the repeated addition of fuel and antifuel strands.

A consequence of the molecular tweezer work was the development of other similar machines, such as the nanoactuator [Simmel and Yurke, 2001] and a more complex three-state device [Simmel and Yurke, 2002], which demanded the manual addition of fuel and antifuel strands. The basic idea of strand displacements as a means to provide a driving force has been used to make the more mobile molecular machines known collectively as DNA walkers, which differ from the tweezers as they move relative to an external substrate. The first DNA walker was a biped device that moved along a track by sequential attachment and detachment of the "feet" from the track [Sherman and Seeman, 2004]. The motion of this motor resembles that of an inchworm, in that one foot is always behind the other. However, a biped walker which progresses by alternating which foot is in front has also been demonstrated [Shin and Pierce, 2004b]. Whilst these developments have progressed the field of DNA nanotechnology and turned the DNA constructions into machines, an autonomous machine that requires no further input once set into motion is preferred.

Autonomous DNA machines require rethinking the earlier designs which necessitated the addition of fuel strands to drive the system. A number of approaches have been developed, which often utilise enzymes to bring about the conformational change that drives the machine. One such example is the linear autonomous motor shown in Figure 1.9 and developed by [Yin et al., 2004]. The track along which the walker moves is self-assembled, and contains three anchorages to which the DNA walker can hybridise. The motor starts on the left-most anchorage and is denoted

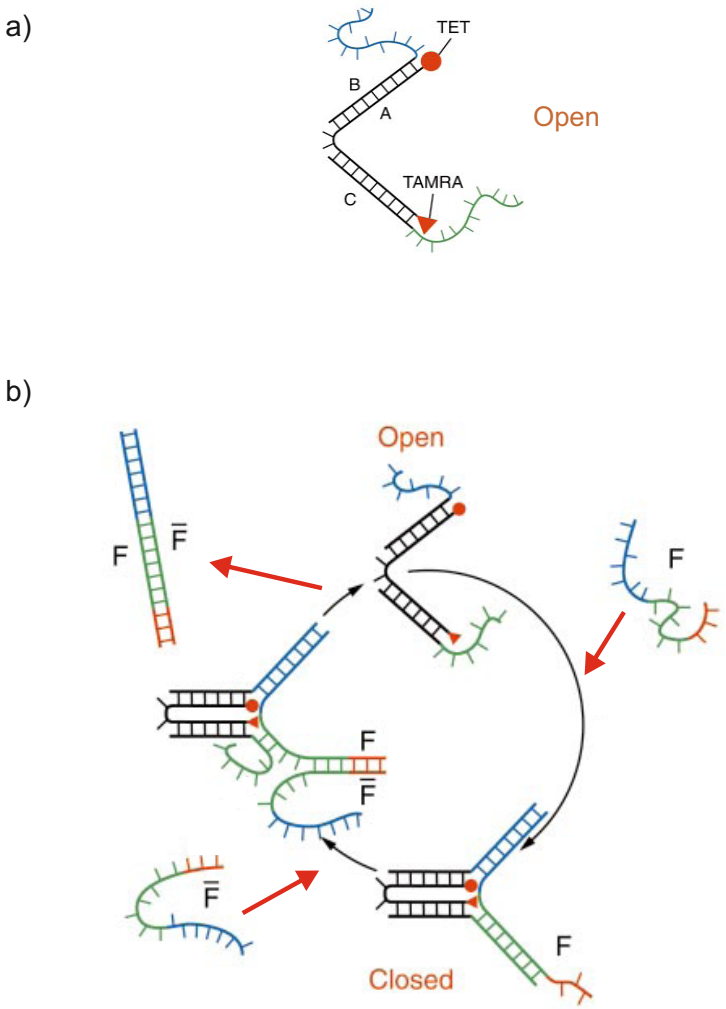


Figure 1.7: The construction and operation of molecular tweezers.

(a) The DNA tweezers are formed by the hybridisation of three strands A, B and C. Strand A is labelled with two fluorescent dyes that form a FRET pair (Appendix A.1.1), which are used to determine the state of the machine.

(b) Upon addition of a new strand F that is complementary to the sticky ends of strands B and C, the tweezers become closed. In a similar way, strand $\overline{\text{F}}$ can hybridise to the sticky end of strand F. Consequently, waste product $\text{F}\overline{\text{F}}$ is formed, and the tweezers re-open. Figure taken from [Yurke et al., 2000].

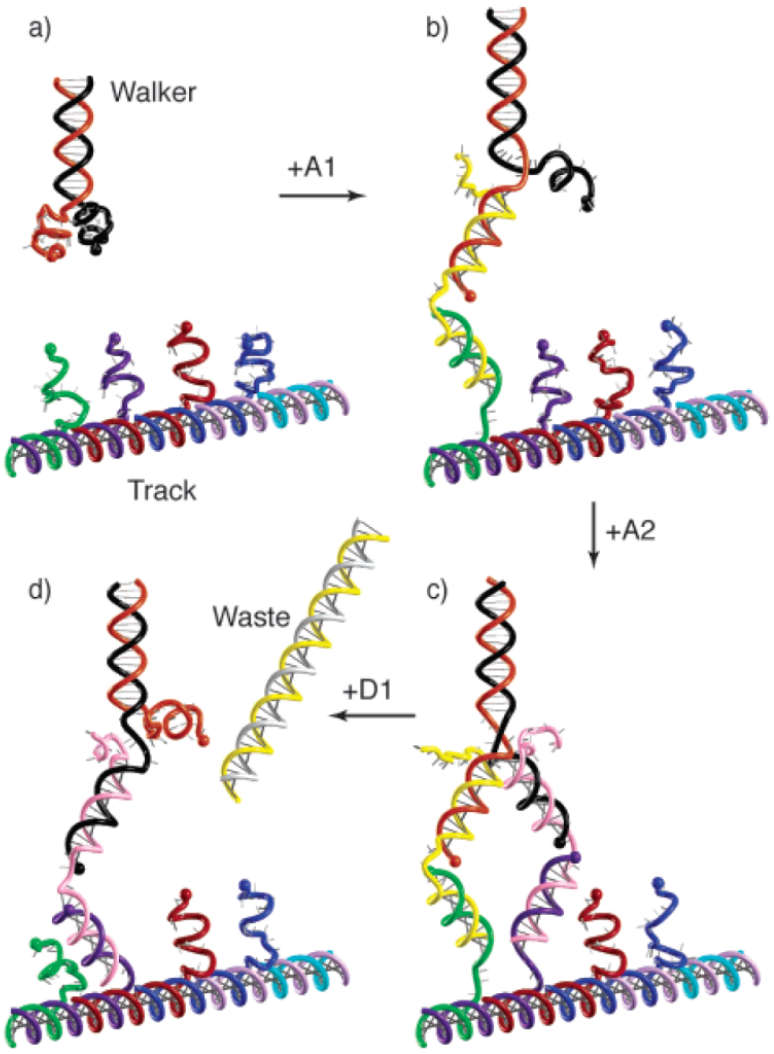


Figure 1.8: An example of a bipedal DNA walker. The walker is sequentially attached and detached from the DNA track by the addition of fuel (yellow and pink) and antifuel (grey) strands. The walker can only progress along the track if these strands are added. Figure taken from [Shin and Pierce, 2004b].

by A*. In order for it to move to the second anchorage, enzyme T4 ligase ligates together the DNA duplex formed from the sticky-end interaction of the motor with the adjacent anchorage. By repairing nicks in the new duplex, the restriction site for another enzyme (PfIM I) is formed. Now, PfIM I cleaves the new complex such that the walker progresses to the new anchorage, but more importantly, the cleavage prevents the motor from moving backwards (thus, it is dubbed the "burnt bridges" mechanism). The walker moves to the third anchorage in much the same way. All the enzymes required are present in the solution at the start of the reaction, and so once the walker is set into motion (by adding ATP [4]), the walker moves autonomously. However, there are problems with this mechanism. The yield for the first step in the reaction was 46 % (because of imprecise stoichiometry and low enzymatic efficiency), which meant the overall yield for the walker progressing from the first to the third step was low. Furthermore, in order to demonstrate that the walker can move more than three steps, much longer DNA strands are required to form the track (which calls for extra purification to guard against sequence mutations during synthesis), as is an ability to utilise the same restriction enzyme to simplify the process.

These issues have been addressed by two recent publications in which the nicking enzyme N.BbvC Ib is employed as the universal enzyme to allow the walker to progress along the track [Bath et al., 2005], [Bath et al., 2009]. As Figure 1.10 depicts, the [Bath et al., 2009] walker has two identical ssDNA feet which are able to hybridise to a single-stranded track. The track consists of two repeating domains, binding and competition, which calls for a universal enzyme to lift the feet from the track. Addition of a fuel strand, which uses exposed nucleotides to hybridise to the back foot

[4] Adenosine-5'-triphosphate (ATP) is a multifunctional nucleotide which transports chemical energy within cells for metabolism. ATP hydrolysis is the reaction by which the chemical energy stored within ATP is released. ATP hydrolysis can be combined with reactions that are thermodynamically unfavourable, as the energy released by the hydrolysis reaction can be used by the secondary reaction. Here, the energy gained from ATP hydrolysis is used to initiate the DNA walker motion.

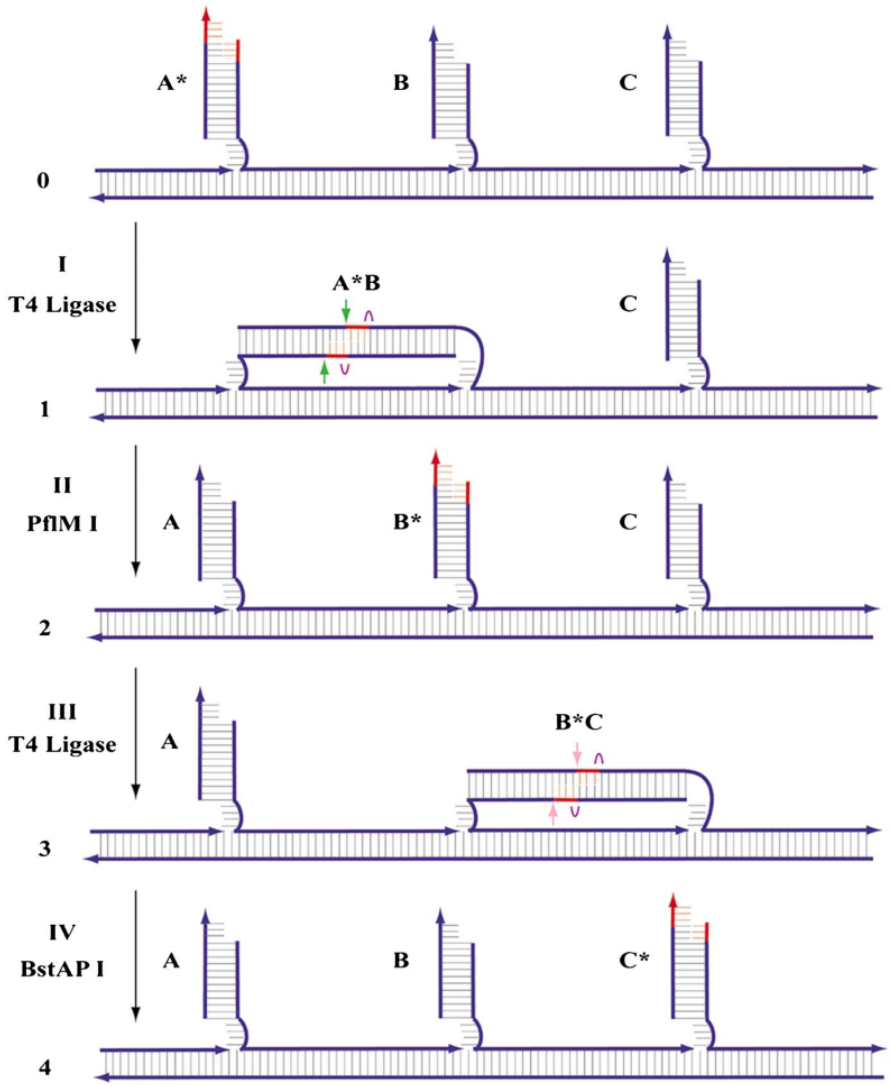


Figure 1.9: An autonomous DNA walker using a "burnt bridges" approach to encourage unidirectionality.

The DNA track is made up of three anchorages (A-C), and the walker is a 6 nt fragment which initially starts off partially hybridised to anchorage A (indicated by a $*$). The fragment is able to hybridise to the free nucleotides of the next anchorage, and enzyme T4 ligase repairs the nicks in the new duplex. A second enzyme cuts the duplex at another site along the duplex, moving the walker to the second anchorage. The nicking process also prevents the walker from moving backwards. Figure taken from [Yin et al., 2004].

of the walker, preferentially lifts the back foot off the track. In this intermediate state, with one foot off the track and one still hybridised, the nicking enzyme cuts the fuel strand in two. The DNA fragments dissociate, and the newly unpaired nucleotides can now either rehybridise to the same position on the track, or take one step forward. FRET measurements indicated that the left foot was lifted from the track at a rate of $4.5 \times 10^5\,\mathrm{M^{-1}s^{-1}}$, whilst the motion of the right foot was found to be 30-fold slower (see Appendix A.1.1 for details on FRET). This bias means the walker has the capability of taking multiple steps, as the reactivity of each foot can be controlled as the motor moves through its operating cycle.

1.3.1 DNA Sequence Design Tools

Custom-synthesised DNA sequences are available from suppliers such as Integrated DNA Technologies (IDT), and unmodified DNA strands can cost as little as 10 pence per nucleotide. The design of DNA sequences for a particular structure or device is made easier by software. Although it is possible to simply write down a sequence of nucleotides by hand, this is a naïve design process, as various important factors must be taken into consideration. For example, single strands of DNA can fold up and hybridise to themselves in hairpin-like configurations if there are self-complementary regions within the strand. This is commonly an undesirable property, as this secondary structure could affect the correct formation or operation of the designed mechanism. Tools such as NANEV, which can generate and test random sequences [Goodman, 2005] can be used in conjunction with mfold [Zuker, 2003], which predicts DNA secondary structure under user-defined reaction conditions, such as salt concentration and temperature. The Computer-Aided Nucleic Acid Design Package (CANADA) is useful for outputting large numbers of DNA sequences with specific properties and for analysing them [Feldkamp, 2010]. It is particularly useful at determining the pairwise interactions of large sets of sequences, so that subsets with little pairwise interaction can be selected (which is sometimes necessary to ensure control

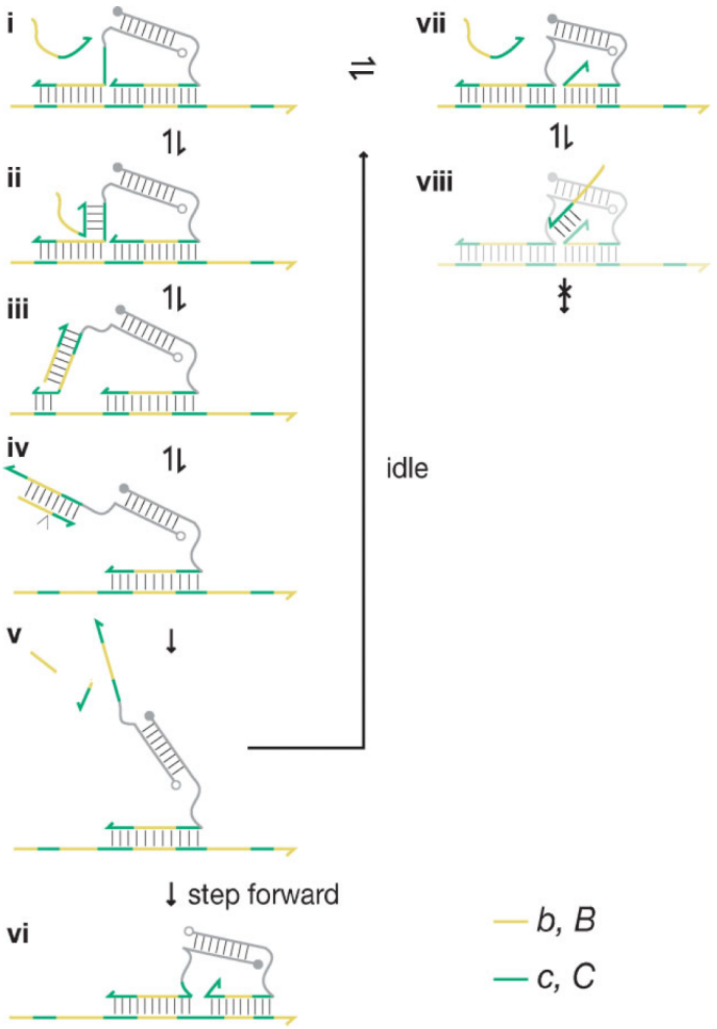


Figure 1.10: An autonomous DNA walker.

The track here is made from ssDNA, and the walker has two identical single-stranded feet which can hybridise to the track. The two feet of the walker are linked together by a dsDNA spacer (grey). The track consists of a repeating sequence of binding (yellow) and competition (green) domains. When a fuel strand is added, it binds to the single-stranded region of the left foot (ii), which causes the foot to be partially displaced from the track (iii). Only 6 nt of the foot remain hybridised to the track, which dissociate spontaneously (iv). A nicking enzyme cuts the fuel strand into two pieces (v). The foot now has the option of rehybridising to the track in either the same location as before (i/vii), or to take a step forward (vi). If the left foot takes a step forward, adding another fuel strand will result in the right foot being lifted. However, if the left foot rehybridises to the track as in step vii, then adding a fuel strand will result in a transient complex forming between the fuel and the 3' end of the right foot. This complex does not allow the walker to progress. Figure taken from [Bath et al., 2009].

over the assembly and operation of a DNA structure/machine). Finally, the two online hybridisation prediction tools Nupack [Zadeh et al., 2011], [Dirks et al., 2007] and Hyther [SantaLucia-Jr and Hicks, 2004], [SantaLucia-Jr, 1998], [Bommarito et al., 2000] have been useful for the design of the DNA mechanisms presented in the following chapters.

1.4 Towards a Synthetic Ribosome

Ribosomes are complexes of ribosomal-RNA and protein which are found in the cytoplasm of a cell, and exist to assemble proteins. The conversion of the information in DNA into proteins begins in the nucleus of cells with the synthesis of messenger-RNA (mRNA) by transcription of DNA. Several turns of the DNA duplex unwind to expose the bases of the two strands. Ribonucleotides are linked in the correct order by hydrogen bonding to their complementary bases on the DNA in the active site of the enzyme RNA polymerase. The growing RNA molecule unwinds from the DNA. Only one of the two DNA strands is transcribed into mRNA. The specific mRNA sequence forms a message that determines the order in which different amino acids are to be joined. Each "word" or codon along the mRNA chain consists of a sequence of three ribonucleotides that specify a particular amino acid [Crick et al., 1961]. As each successive codon is read by the ribosome in a process known as translation, the correct amino acids are brought into position for transfer to a growing peptide chain. When synthesis of the protein is completed, a "stop" codon signals the end and the protein is released from the ribosome.

There is growing interest in building a synthetic ribosome amongst biochemists and geneticists, but increasingly also in the DNA nanotechnology arena. For biologists the reasoning is simple: to gain a better understanding of this fundamentally important machine. However, for chemists the desire is more for the industrial benefits such

a synthetic machine could have on the discovery and synthesis of drugs. The synthetic system could be used to generate combinatorial libraries of new functional molecules more quickly than any traditional molecular evolution methods [Gartner, 2006]. Research on developing a synthetic ribosome for this purpose has tended to focus on altering the natural molecular machinery to produce new molecules. For instance, by reprogramming the codons that specify the synthesis of natural amino acids to instead code for non-natural amino acids, synthetic polypeptides have been produced [Chin et al., 2003], [Murakami et al., 2006]. However, these methods are somewhat limited, as at least some of the natural codons of a protein must remain unaltered for the synthesis to take place. Furthermore, it can be difficult to encode for "exotic" amino acids (that is, molecules that are remarkably different from the natural amino acids). In an ideal synthetic ribosome, there would be no limit on the type of polymeric molecule that could be formed.

Molecular evolution has been used to generate new biological macromolecules, and the technique has been applied to the discovery of functional synthetic small molecules and new chemical reactions [Melkko et al., 2007]. Traditional methods for evolution are being replaced by processes that make use of DNA to encourage reactions between small molecules linked to DNA strands. Section 1.5 describes methods to create new functional small molecules, through the use of DNA templates to catalyse the reaction.

1.5 DNA-Templated Chemistry

Over 40 years ago, the idea that evolutionary principles could be applied to biomolecular entities *in vitro* was presented [Mills et al., 1967], kickstarting a whole new area of research. Spiegelman and his colleagues had shown that nucleic acids capable of self-duplicating could be made to evolve outside a living cell. Selection pressures were imposed on the molecules as they replicated, such that those molecules

that mutated were preferred. The evolution was so successful that by the end of the experiment, a mere 17 % of the original genome had survived. The result directly led to more recent approaches to evolve molecules for the purpose of discovering and selecting for new functional proteins. Selection is key in molecular evolution, and synthetic molecular evolution is conceptually similar to Darwinian natural selection where functional molecules (survivors) are separated from uninteresting mutations (unfit individuals) within large libraries of different compounds (Figure 1.11). Nucleic acids are particularly suited to synthetic evolution methods because of their ability to be replicated, their high information density and their ease of manipulation [Scheuermann et al., 2006], [Rozenman et al., 2007]. DNA-templated chemistry can create polymers with strong binding affinities to targets of interest. Reviews of the recent achievements of DTS are presented in [Li and Liu, 2004], [Wrenn and Harbury, 2007], [Appella, 2010] and [Glökler et al., 2010].

1.5.1 Nucleic Acid-Templated Synthesis

The ability of nucleic acids to be functionalised, by the covalent attachment of synthetic molecules to the strands, has made it possible to induce chemical reactions upon strand hybridisation. Duplex formation encourages the attached small molecules to react, as the effective concentration of the two reacting groups is much higher than when they are unlinked [Gartner and Liu, 2001], [Calderone et al., 2002], [Gartner et al., 2002b], [Calderone and Liu, 2005]. Furthermore, undesired reactions are easily prevented by the hybridisation process, as only strands that are designed to hybridise together result in a reaction. The typical concentrations used in the reactions are low enough to prevent the reactions of chemical groups that are not brought together by hybridisation. The reaction is controlled through the sequences of the strands involved, and it is easy to imagine how large numbers of reactions can be initiated with just a few oligonucleotides by altering the order of the hybridisation process (Figure 1.12).

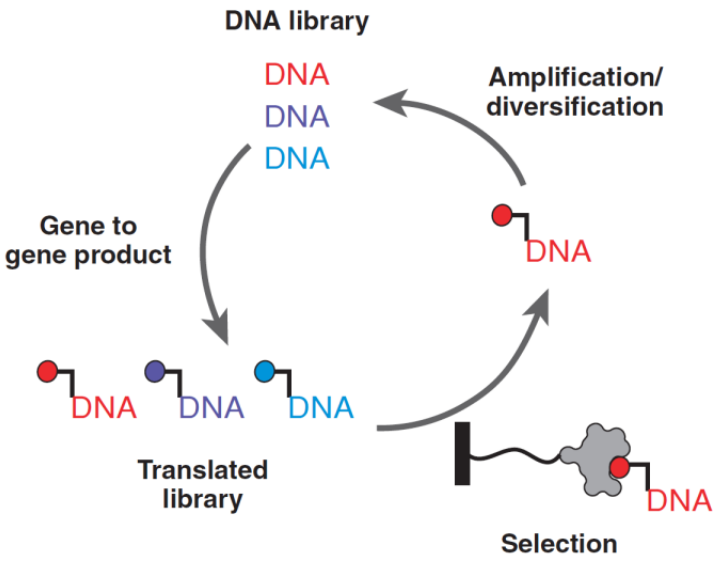


Figure 1.11: An illustration of *in vitro* evolution.

A library of nucleic acid genes (coloured DNA) is translated into corresponding gene products (coloured circles), while maintaining a physical link between gene and gene product. This coupling allows the library to be subjected to functional selection techniques. Those that are successfully pulled out from the library are then amplified and diversified. This generates the second-generation pool of strands for the next round of translation. This continues until the population converges to DNA strands encoding gene products that survive selection. Figure taken from [Wrenn and Harbury, 2007].

In the synthesis strategy presented by [Gartner et al., 2004], there is a resemblance to the natural ribosome. Here, DNA template strands with a number of distinct domains encoded within the sequences (much like mRNA codons) are used to direct the reactions. Shorter, chemically functionalised oligonucleotides are added and these effectively "read" the template sequences, their sequences determining to which strand, and where along the strand, they can hybridise (step 1 in Figure 1.12). After the reaction has finished, the short DNA strands are removed from the template. Now, a second batch of functionalised strands are added (step 2 in the diagram), which hybridise to different positions along the templates. This process continues until the entire sequence of the template strands has been read. One problem with this scheme

From: Gartner, Z. J., Tse, B. N., Grubina, R., Doyon, J. B., Snyder, T. M., and Liu, D. R. (2004). 'DNA-templated organic synthesis and selection of a library of macrocycles', Science, 305: 1601-1605. Reprinted with permission from AAAS.

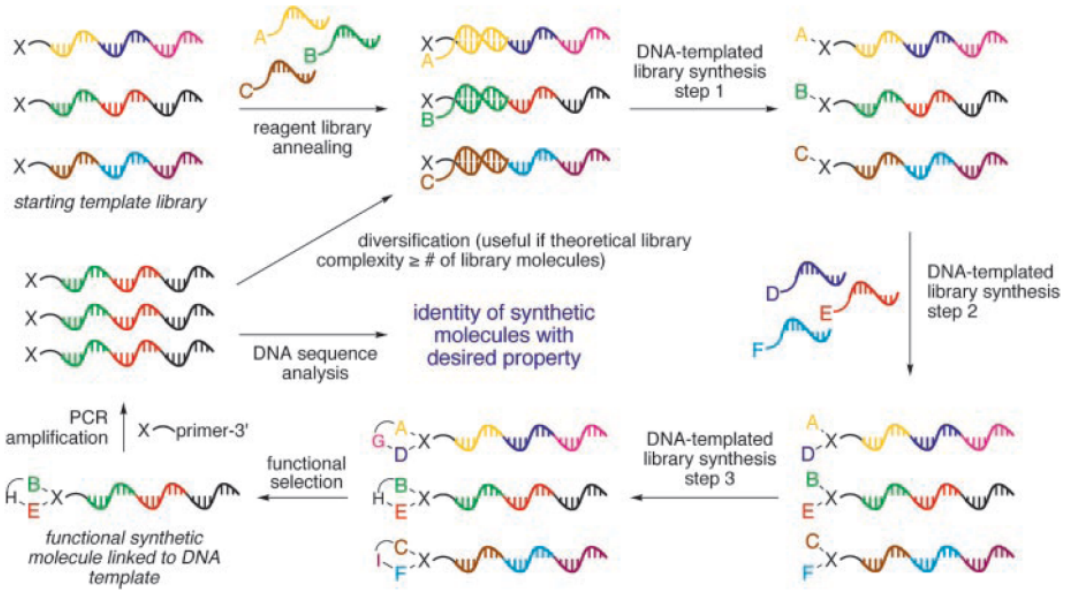


Figure 1.12: An approach to DNA-templated synthesis.

Template DNA strands are annealed with small molecule-linked DNA strands in one pot. Hybridisation enables reactions between the reactive species, where the growing chain is attached to the template DNA strand throughout. The spent strands are removed and replaced with new functionalised strands. This is repeated until the whole length of the template strand sequences have been read. Thus, a library of molecules has been created, which can now be subjected to selection and analysis. Figure taken from [Gartner et al., 2004].

is that the small molecules on the input oligonucleotides and the growing chain on the template become increasingly separated, decreasing the likelihood of a reaction. It has been found that some templated reactions are inefficient when the reacting groups are separated by a few base pairs [Gartner et al., 2002a], and a double-stranded region between reacting groups can also reduce the efficiency (necessitating a single-stranded section in the design) [Gartner and Liu, 2001]. Figures 1.13 and 1.14 show a number of different DNA architectures that have been proposed to overcome these problems. Some of the DNA-templated synthesis (DTS) mechanisms suffer from the same setback as the early DNA walkers, namely that they rely on long template DNA strands. For example, in order to produce synthetic small molecules that are made up of ten individual units, the template strand would need to have ten distinct domains, which

would need to be at least 10 nt long each in order to guarantee they were unique. Whilst 100-mer strands are commercially available, long DNA strands are likely to suffer from the formation of unwanted secondary structures that can interfere with the reactions [Snyder et al., 2008]. Therefore, an ideal DTS mechanism is one that is successful at generating combinatorial libraries but which precludes the need for long templates.

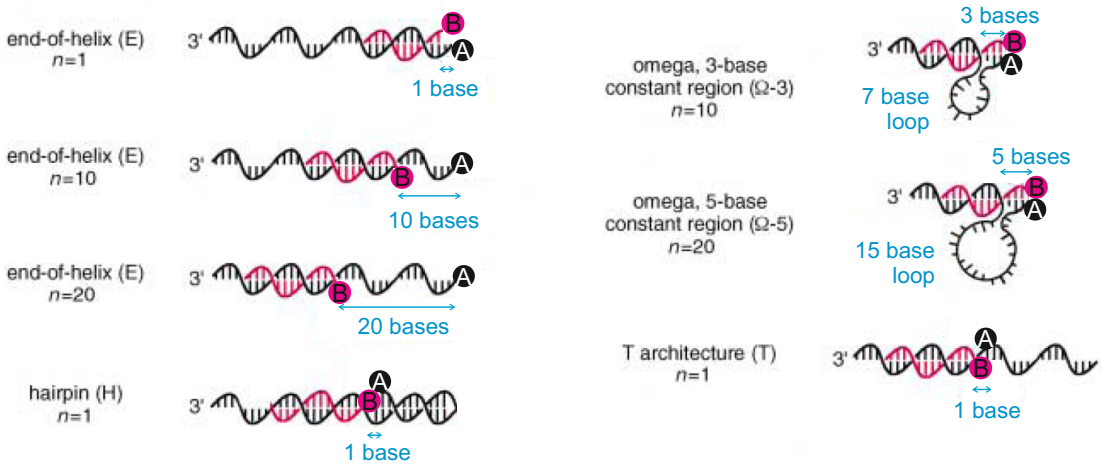


Figure 1.13: Four architectures for DNA-templated synthesis.
Template strands (black) and reagent strands (red) are functionalised at their 5' and 3' ends respectively (indicated by the circles labelled A and B). Amino-terminated DNA is covalently attached to a reacting group. In the case of the T architecture, the functional group is attached along the length of the template strand, via an amino-modified base. Once template and reagent strands hybridise, a number of unpaired bases (n) separate the reactive groups of the two strands. The success of the end-of-helix and hairpin approach to synthesise new molecules can be varied by altering the distance between the reacting groups. The omega and T architectures allow distance-dependent reactions to be efficiently mediated by DNA bases far from the reactive end of the template. This expands the types of reactions that can be encoded anywhere on a DNA template. Figure taken from [Gartner et al., 2003].

A number of different chemical reactions have been demonstrated as being especially amenable to DTS, some of which require the addition of reagents to promote chemistry between reacting groups linked to oligonucleotides, and some which are

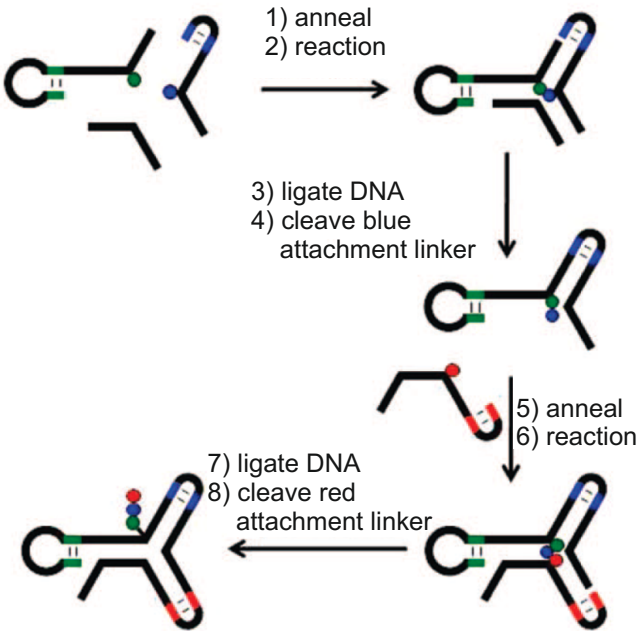


Figure 1.14: Design for the yoctoliter-scale small molecule synthesiser.

DNA three-way junctions allow for controlled synthesis of molecules. DNA strands consist of constant domains (black) and codon domains (coloured). The DNA strands are linked to chemical building blocks (coloured circles) by non-cleavable and cleavable linkers. One arm of the synthesiser is attached to its building block by a non-cleavable linker (here, this is the arm with the green codon and building block). Hybridisation between this and a second arm (blue codon) allows a reaction between the chemical building blocks to take place (1-2). The blue building block is cleaved from the arm, and the DNA is ligated together (3-4). This process is repeated for the third arm (5-8). Thus, a library of small molecules with a single covalent attachment to the DNA is created. The library can now be subjected to selection and amplification as described in Section 1.5.4. Figure taken from [Hansen et al., 2009].

reagent-free. Within this set it is possible to divide the reactions into subsets according to what happens to the reacting groups. One such subset is the group transfer reaction, which has been alluded to earlier. Here, the incoming DNA strands hybridise to a template and transfer their reactive group to the template as in [Tse et al., 2008]. Another subset of reactions involves the covalent attachment of the reacting groups to each other, thereby linking together the two oligonucleotides to which they are attached. The wide scope of possibilities are explored in [Silverman and Kool, 2006] and [Gartner et al., 2002a]. The two chemistries that are explored in the following chapters involve either a transfer reaction (Section 1.5.2) or a covalent attachment (Section 1.5.3).

1.5.2 Wittig Chemistry

One popular example of the group transfer reaction mentioned earlier is the Wittig olefination, which has particularly high yields (>90 %) in a reaction time of less than two hours [Gartner et al., 2002a], and is robust under aqueous conditions. The Wittig reaction (shown in Figure 1.15) occurs between an aldehyde and a phosphonium ylide to produce an alkene and a phosphine oxide, and has been shown to work with a variety of DNA architectures [Gartner et al., 2003], [Rozenman and Liu, 2006].

An elegant mechanism for DNA-templated Wittig chemistry has recently been presented [McKee et al., 2010]. McKee and coworkers not only produce olefin 4-mers (a chain of four covalently-bound Wittig building blocks), but have also bypassed the need for a long template strand to direct the synthesis. Figure 1.16 depicts the oligonucleotides involved in the reaction, which all consist of one of two complementary sequences (coloured black and grey), and unique toehold domains which serve to identify the strand and the building block linked to it, but are also used to facilitate strand displacement. As in most cases of DTS, the hybridisation of DNA brings the chemical building blocks into close proximity, here at the ends of the duplex. In this

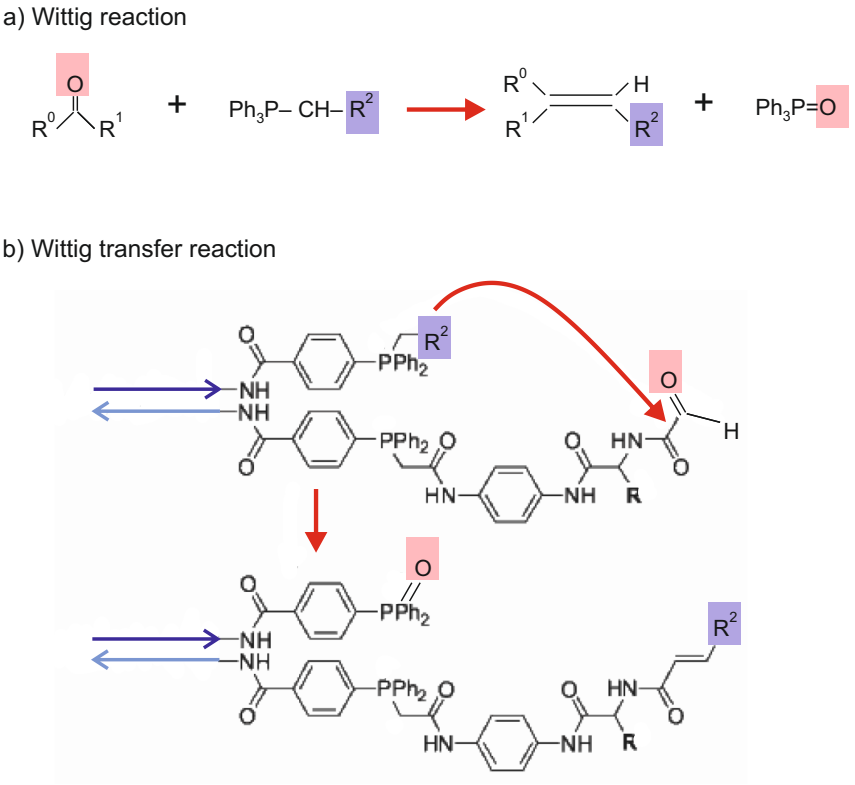


Figure 1.15: Illustration of the Wittig reaction.

(a) The Wittig reaction takes place between an aldehyde and phosphonium ylide, which react to produce an alkene and a phosphine oxide.

(b) In this case, the Wittig reaction is facilitated by the hybridisation of two DNA strands that are linked to the Wittig reactants. The dark blue strand contains only a phosphine modification, whereas the light blue strand is bifunctional and contains both a phosphine ylide and an aldehyde. This allows for the building block attached to the ylide (R^2) on the dark blue strand to be transfered to the aldehyde of the light blue strand. The unreacted ylide of the light blue strand can be used in a second Wittig reaction involving another bifunctional DNA strand. In this way, the Wittig transfer reaction can be used to build a peptide chain. Figure adapted from [McKee et al., 2010].

instance however, the Wittig reaction also includes a group transfer reaction, such that the growing chain is transferred from one DNA strand to another (in the diagram from strand A to strand B. The transfer reaction is shown in detail in Figure 1.15). The spent DNA is removed from the duplex by toehold-mediated strand displacement, and a third functionalised oligonucleotide is introduced to continue on the reaction. The end product is transferred to the final DNA strand added to the reaction, which can be separated from the solution for analysis by using a biotin-labelled removal strand. The yield of successful products was high (72 %), and the ability to create distinct products depending on the order in which the functional oligonucleotides were added to the reaction was demonstrated. The lack of a long template strand to dictate the reaction means the system is not entirely ribosome-like, and neither is it an autonomous, one-pot mechanism. However, the potential to be adapted for large library synthesis and evolution does exist by using a "split-and-pool" method for instance (described in Section 1.5.4).

1.5.3 Click Chemistry

Another reliable and robust chemistry is the copper(I)-catalysed azide-alkyne cycloaddition reaction, in which the azide and alkyne reacting groups end up covalently linked together. Click chemistry, as it is more commonly known, has been used extensively for the purposes of drug discovery perhaps because it works well with combinatorial chemistry techniques [Kolb and Sharpless, 2003]. The non-catalysed reaction requires high temperatures and is rather slow because the azides and alkynes are kinetically stable. Typically, this reaction goes to completion in 18 hours at 98 °C [Seo et al., 2003]. However, the copper-catalysed version of the reaction has been a recent but extremely useful discovery [Tornøe et al., 2002], [Rostovtsev et al., 2002]. The copper catalyst undergoes a metal insertion reaction into the alkyne group. While the reaction can be performed using copper(I), the reaction is typically performed using a mixture of copper(II) and a reducing agent (such as sodium ascorbate) to

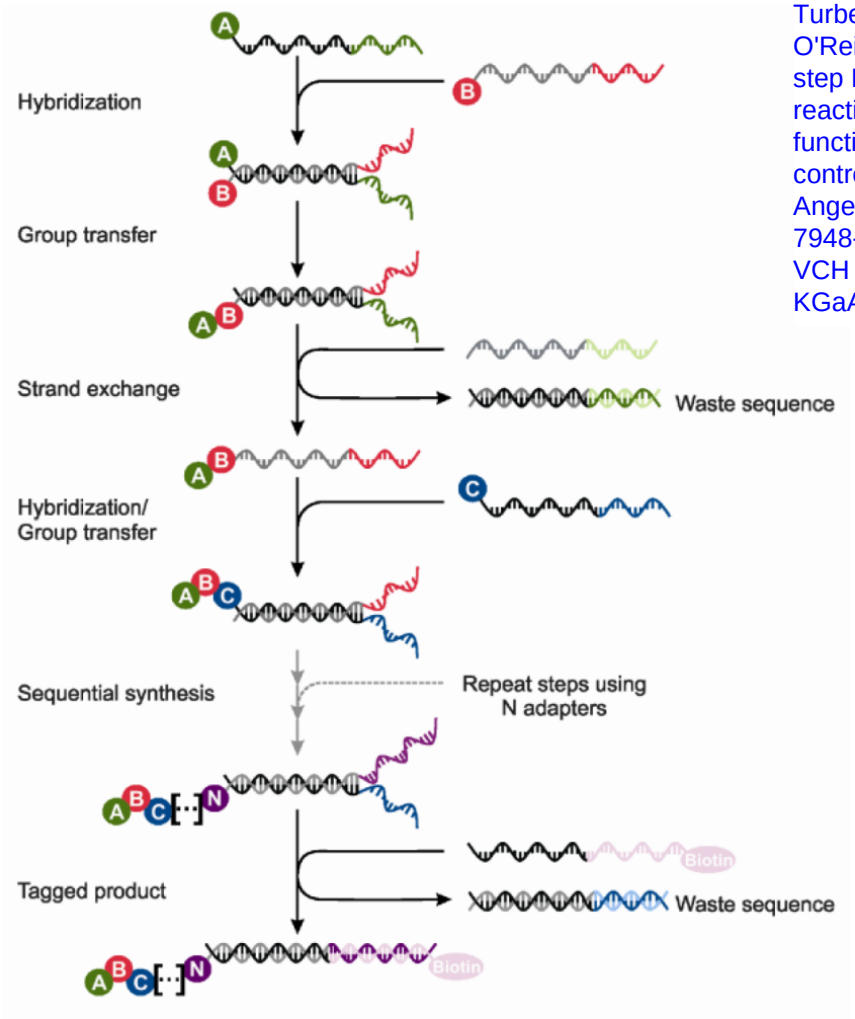


Figure 1.16: Illustration of the strand displacement mechanism for DNA-templated synthesis. The chemical reaction used here is Wittig olefination. Hybridisation of the nucleic acids bring together an aldehyde and phosphonium ylide, which react to produce an alkene and a phosphine oxide. In this case, the alkene is transferred from strand A to strand B. The transfer reaction is depicted in Figure 1.15. Toehold-mediated strand displacement is used to strip away strand A (to which the waste phosphine oxide is still attached), and by doing so reveals nucleotides on strand B for another strand (with an aldehyde functionality) to hybridise to, thereby continuing the reaction. Figure taken from [McKee et al., 2010].

produce copper(I) in situ. The copper catalyst speeds up the reaction rate significantly: under certain conditions, reaction completion is observed within 5 min [Kočalka et al., 2008]. Click chemistry is particularly popular because of its efficiency at room temperature and the fact that it can be performed in a variety of solvents, including water. The reaction has already been used in combination with DNA, specifically to cross-link two complementary DNA strands in order to produce very stable duplexes [Kumar et al., 2007], [El-Sagheer and Brown, 2010] (see Figure 1.17). Such duplexes have been shown to have much higher thermal stabilities, and in some cases the melting temperature of the duplex has been increased by up to 30 °C. However, the reaction has not yet been used explicitly for DNA-templated combinatorial chemistry.

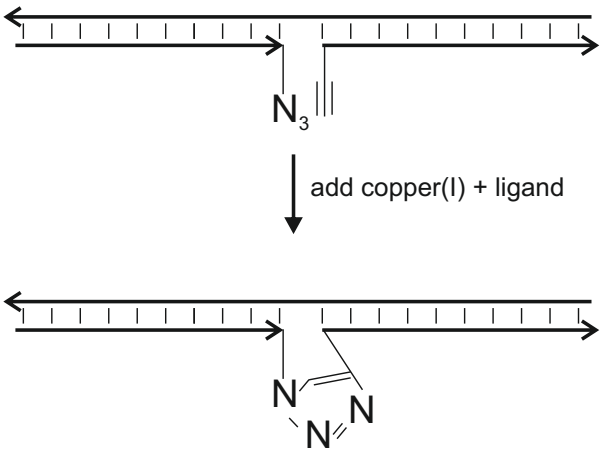


Figure 1.17: Illustration of DNA-templated click chemistry.

Two short DNA are hybridised to a template strand, one of which is functionalised with a 3' azide, while the other is functionalised is with a 5' alkyne. The click functionalities reside on neighbouring base pairs, but without the addition of the copper(I) catalyst, the cycloaddition reaction between the azide and alkyne is slow. Typically, copper(I) is produced by mixing copper(II) with the reducing agent sodium ascorbate. Copper(I) must be mixed with a copper-binding ligand (such as THPTA), in order to prevent oxidation of the copper(I) species. The catalyst and ligand are added together to the DNA and once added, the reaction can take as little as 5 min to reach completion [Kočalka et al., 2008]. Figure adapted from [Kumar et al., 2007].

Click chemistry is not without its flaws. Copper(II) is known to damage DNA through the oxidation of bases [Tkeshelashvili et al., 1991], [Liu et al., 1999], and therefore if the chemistry is to be templated by nucleic acids, the DNA must be protected against this possibility. Copper(I) is not stable in aqueous solutions and consequently, a copper-binding ligand such as the water-soluble tris-hydroxypropyltriazolylamine (THPTA), is used to stabilise the copper(I) species. The ligand does not participate directly in the reaction, but it prevents the oxidation of the copper(I) species into the copper(II) (which damages DNA), and also protects the copper ion from interactions that result in the formation of side products that may also be harmful to the DNA. The key benefits to using click chemistry for the purpose of DNA-templated synthesis are that:

1. the azides and alkynes can be attached to DNA strands without affecting the biophysical properties of the strands;

2. the azides and unactivated alkynes are almost entirely unreactive towards the functional groups normally encountered in nature, and instead they only react with each other;

3. the triazole unit (end product of the reaction) is extremely stable and non-toxic [El-Sagheer and Brown, 2010].

1.5.4 Selection and Evolution

So far, a brief review of the possible chemistries and techniques to generate large libraries of synthetic molecules has been presented, and therefore it is now necessary to discuss the application to drug discovery. Without methods to select for particular functions, the full potential of the libraries is lost. However, screening methods capable of testing each species in a library are too slow when dealing with large numbers. Therefore, an alternative is required, and one such method is that of DNA display

[Halpin and Harbury, 2004]. It requires the use of DNA to identify and select DNA-linked small molecules. Figure 1.18 shows how a combinatorial library is created by the split-and-pool method, in which reactants are repeatedly split into groups, allowed to react, and pooled to create many possible outcomes. The reactants are all attached to long ssDNA, which contain codons that specify a small-molecule synthesis. The order of the codons on the template directly correlates to the order in which the synthesis reactions take place. Thus, at the end of the reaction, it is possible to know what building blocks a particular synthetic molecule is composed of simply by reading back the DNA sequence that directed its creation. The resulting library of nonnatural peptides was subjected to selection for high-affinity to a particular antibody. (One of the template strands used in the split-and-pool reaction was specifically designed for such an affinity.) The selected DNA was amplified using PCR and used for the next round of synthesis and selection. After two rounds, twenty sequences encoded the desired property, suggesting that the split-and-pool method could be used to direct the evolution of molecules with specific attributes. A similar method using PNA (peptide nucleic acid) has also proved to be successful [Brudno et al., 2010].

1.6 Construction of a Synthetic Ribosome

The various DNA structures and devices reviewed thus far are the inspiration for the work presented in this thesis. Two different DNA architectures are described which have the potential to be used for the development of a synthetic ribosome, that is, a system capable of sequentially interpreting instructions from a tape and producing an output.

The goal to build a synthetic ribosome has two distinct elements: firstly to determine the chemistry that is most suited to producing chains of small molecules that are akin to peptides, and secondly, to design the architecture to enable the chemistry.

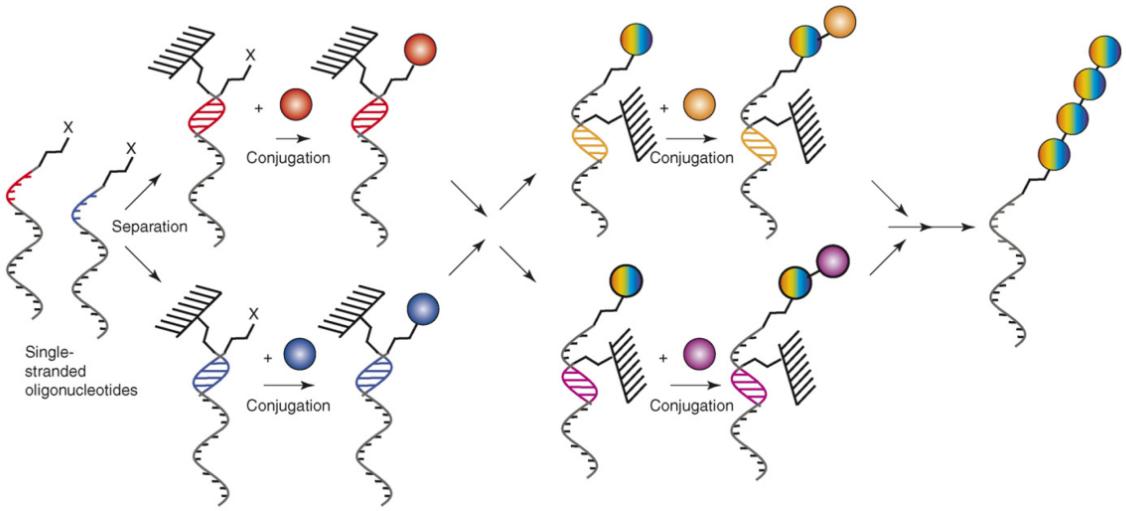


Figure 1.18: Split-and-pool synthesis of a combinatorial chemistry library.

Small molecules are synthesised at one end of 340 nt ssDNA (at the position marked by an x in the diagram). Each ssDNA template consists of multiple noncoding domains (black) and six variable codon domains (coloured). In the first round of the split-and-pool synthesis, the DNA strands are divided into codon-specific groups, based on the sequence of the first codon domain. To each group, a different chemical building block (coloured circle) is added and conjugated to the DNA. The products are pooled and separated again into codon-specific groups for the second round, where the division is based on the sequence of the second codon domain. Different chemical building blocks are added to the DNA templates in each group. Splitting the pool into n groups in each round results in n^2 distinct products after two rounds, n^3 products after three rounds and n^6 products after the entire template strand has been "read". The result after six split-and-pool rounds is a vast library of small molecules that are each attached to the DNA template responsible for its construction. The library can be subjected to selection pressures, and the surviving small molecule products can be amplified by performing PCR on the attached DNA template. Figure taken from [Melkko et al., 2007].

The majority of the work presented here is on the design and characterisation of DNA architectures to control a sequence of chemical reactions. Two DNA constructs are explored - a linear and autonomous system that is inspired by the DNA machines reviewed above and has the potential to be combined with Wittig chemistry (Chapter 2), and a one-pot system for click chemistry templated on DNA origami (Chapter 3).

As in nature, where translation involves the reading of successive codons along the mRNA chain, so too will the synthetic ribosome be capable of progressing linearly along a DNA template. While the codons in nature specify a particular amino acid, the codons in the synthetic system will be physically linked to a particular chemical building block. As the synthetic ribosome mechanism works its way along the template, chemical reactions sequentially link the building blocks together so that a molecule is formed according to the specific sequence of codons on the template. This provides the flexibility to produce different end products by rearranging the codons on the template. If the molecule produced at the end of the reaction remains attached to the template (which encodes the instructions that produced the molecule), it is possible to select useful products that have a high-affinity to a particular target and discard the remainder. The attached template can then be re-used to replicate the product. This is especially useful when combining the synthetic ribosome with the selection and evolution methods described earlier in Section 1.5.

The synthetic ribosome mechanisms presented in this thesis borrow the strand displacement method used by many DNA walkers to enable its progression along a DNA track [Bath et al., 2005] [Shin and Pierce, 2004b]. The way in which strand displacements are able to drive DNA machines is outlined in Section 1.7.

1.6.1 Drawing Conventions

It is useful to state the drawing conventions used in this thesis, to distinguish between generic and specific illustrations of DNA mechanisms. Figure 1.19 shows the different conventions employed in all diagrams created by the author. DNA strands are always drawn as lines with an arrow at one end, which indicates the 3' end of the strand. Single strands of DNA that are added to or released by a synthetic ribosome mechanism are shown as wavy lines, indicating that they are free in solution. Duplexes are depicted in two different ways. In illustrations of generic systems, the particular number of base pairs formed between two DNA strands is not shown, and the duplex is represented as two anti-parallel lines (Figure 1.19 (a)). Unpaired bases are indicated by one of the lines being shorter than the other, or by a gap between two lines that form one "side" of the DNA duplex. In specific systems, the exact number of base pairs formed between the two strands is shown by short thin lines between the two anti-parallel lines (Figure 1.19 (b)). Each of these short black lines represents 3 base pairs. Unpaired bases in a duplex are indicated by short lines that stem from one of the DNA strands. The unpaired bases can occur at the end of one of the DNA strands forming a duplex, or in the middle of a duplex, as shown in Figure 1.19 (b). The colour of these short lines refers to a specific sequence domain.

Colours are typically used to represent unique sequence domains in a DNA strand (unless otherwise stated), while black is used to represent universal/repeating domains. The track or template strands that the synthetic ribosome mechanism moves along are usually drawn in black. Chemical building blocks and fluorophores that are linked to DNA strands are shown as circles/ovals (Figure 1.19 (c)). The attachment linker is shown as a short wavey black line between the building block and the DNA. Chemical links between DNA strands, or strands which are comprised of inverted bases, are shown by small green circles joining two DNA strands together (Figure 1.19 (d)).

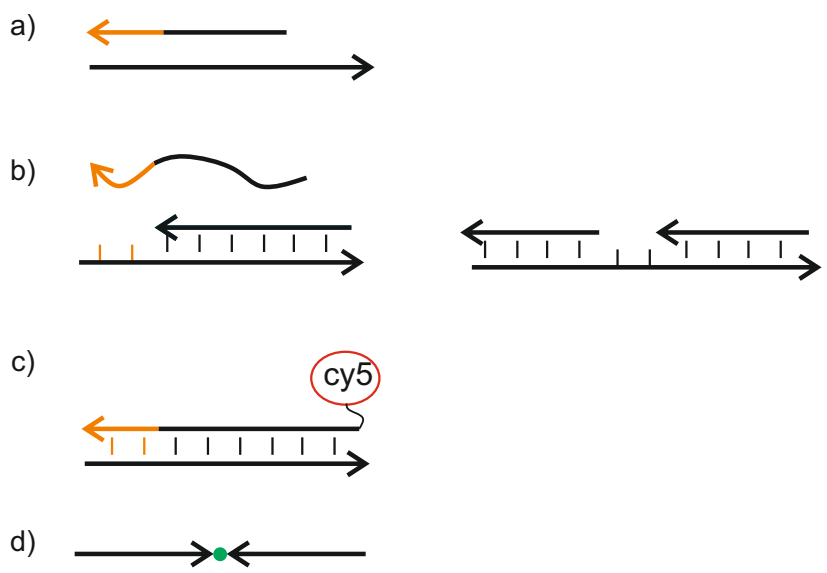


Figure 1.19: Drawing conventions (see main text for details).

1.7 Driving Dynamic DNA Machines by Toehold-Mediated Strand Displacements

Before designing a synthetic, DNA-based ribosome, it is necessary to understand the way a dynamic DNA machine operates, and in particular to examine the strand displacement mechanisms upon which many rely. The essence of these mechanisms is that an invading ssDNA is able to displace and replace one of the strands forming a duplex, thereby creating a new duplex and a "waste" strand. If the invading strand is identical to the strand it replaces, the displacement process is known as "blunt exchange" (see Figure 1.20 (a)). This displacement mechanism is slow because it is difficult for the invading strand to start the displacement reaction without the aid of a binding site on the template (complementary) strand [Reynaldo et al., 2000]. However, if the invading strand is able to hybridise to an exposed binding site on the template duplex strand (black), then the speed of the reaction is increased significantly (Figure 1.20 (b)). The free nucleotides of the binding site on the template strand are known as a "toehold" [Yurke and Mills, 2003], and it is this toehold domain

that enables the invading strand to initiate the strand displacement. The invading strand now participates in a random walk branch migration process along the template strand, competing with the incumbent strand over base pairings. This branch migration process is typically very quick ([20]µs per nucleotide), even if the branch migration occurs over a large number of base pairs [Green and Tibbetts, 1981], and it leaves the incumbent strand free in solution.

The probability of either strand "winning" and forming a duplex with the template strand is related to the relative stability of the complexes formed. In the example shown in Figure 1.20 (b), the duplex formed with the invading strand (orange-black) is the thermodynamically-favoured complex because of the extra base pairs acquired from hybridisation to the toehold. This enables it to "defeat" the incumbent strand and become completely hybridised to the template strand. However, the reaction is reversible, as thermal fluctuations may cause spontaneous dissociation of the new duplex, providing opportunity for the displaced strand to gain a toehold and rehybridise. The spontaneous dissociation rate of 10 mers is $1.1 \times 10^{-2}\,\mathrm{sec}^{-1}$ at 30.5 °C [Morrison and Stols, 1993], [Yurke et al., 2000]. With longer strands spontaneous dissociation is slow compared to the association rate of complementary strands, consistent with the equilibrium ratio between populations of hybridised and unhybridised strands.

If the toehold length is long enough to provide the system with a significant increase in energy, then a DNA machine using toeholds can be thermodynamically driven to progress to the next step along its track. Furthermore, the mechanism can be designed such that toeholds already passed along the track are converted into stable duplexes, thereby preventing the machine from reversing (similar to the "burnt bridges" mechanism depicted in Figure 1.9). This is an especially useful property for the synthetic ribosome which, like the natural ribosome, should ideally be uni-

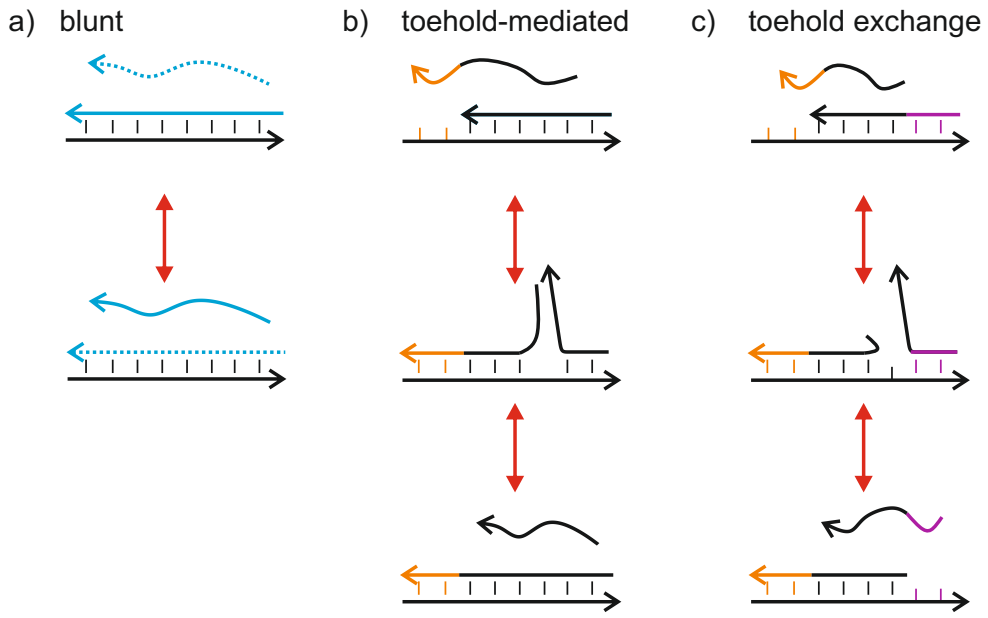


Figure 1.20: An illustration of the three forms of strand displacement.
(a) Blunt displacement. The initial duplex is formed between a template strand (black) and an incumbent strand (blue), which are fully complementary. The invading displacer strand (dashed) is also complementary to the template strand. Displacement of the incumbent strand may occur if the spontaneous melting of part of the duplex creates a toehold for the invading strand.
(b) Toehold-mediated displacement. The incumbent strand (short black arrow) is only complementary to 18 nt of the template strand, and 6 unpaired bases exist at the 5' end of the template. These form a toehold for the invading displacer strand, which is complementary to the full length of the template. Toehold-mediated displacement is significantly quicker than blunt displacement, as the invading strand can hybridise to the exposed unpaired nucleotides. The resulting gain in free energy helps to drive the displacement reaction [Yurke and Mills, 2003]. After the invading strand has hybridised to the toehold domain (orange), it begins the branch migration step in which both strands compete over the same base pairs (black). This is a random walk process in which every base pair made by the invading strand corresponds to every base pair that is broken between the original duplex.
(c) Toehold exchange. Here, both the incumbent strand and the invading strand have unique 6 nt toehold domains on the template strand. The exposed toehold allows the invading strand to hybridise and compete over the 12 nt universal domain (black) with the incumbent strand. Once displaced, the incumbent strand is able to reverse the process by using its (now) exposed toehold domain to rehybridise to the template.

directional so that it cannot reverse on itself and re-read codons.

The strand displacement reaction rate can be controlled by understanding how to alter the speeds of the toehold association/dissociation steps and of the branch migration step [Zhang and Winfree, 2009]. Typically, association of a DNA strand with its complement is fast; the rate constant was measured as $10^6\,\mathrm{M}^{-1}\mathrm{s}^{-1}$ for a 10 nt strand hybridising to its complement [Morrison and Stols, 1993], and so for typical concentrations of 10 nM, the time to half completion is about a minute. In toehold-mediated strand displacement, altering the length of the toehold can alter the rate of the reaction dramatically [Yurke and Mills, 2003]. Furthermore, a single base mismatch in one of the strands can pose a substantial barrier to branch migration [Panyutin and Hsieh, 1993].

A variation on the simple toehold-mediated strand displacement mechanism is one in which two toehold regions exist on the same strand, for two different strands to hybridise to (Figure 1.20 (c)). Only one of the toeholds is revealed or "active" at any one time, and the strand displacement mechanism alters which toehold is active. This is known as toehold exchange, and is discussed in detail in [Zhang and Winfree, 2009]. The system is illustrated in detail in Figure 1.21. Template strand S contains two toehold domains, $\overline{\gamma}$ and $\overline{\beta^m}$, to which the green invading strand and red incumbent strand are complementary, respectively. The initial state of the system is a duplex between template S and incumbent strand, with the $\overline{\gamma}$ toehold revealed. Once added to the system, the invading strand is able to quickly hybridise to the $\overline{\gamma}$ toehold, and then begins the random walk branch migration along the competition domain of template S, displacing the incumbent strand in the process. The invading strand can fully hybridise to the template, leaving the incumbent strand hybridised to its $\overline{\beta^m}$ toehold only. The incumbent strand may then spontaneously dissociate from the template altogether and, in doing so, reveal the $\overline{\beta^m}$ toehold. The free toehold has migrated

from one end of the template to the other. Clearly, the toehold exchange process is reversible, as the incumbent strand can rehybridise to its own (now free) toehold and reverse the strand exchange process.

At equilibrium there is a dynamic balance between rate of strand invasion and displacement for each strand type. Even when the initial state is more thermodynamically favourable than the final state, the exchange reaction can still occur, and can proceed quickly [Zhang and Winfree, 2009]. The invading strand can be thermodynamically favoured in the reaction, by either increasing its toehold length or strength, as this will increase the rate of association. However, if both toeholds are strengthened, then both the forward and backward reaction speeds increase. The toehold exchange process can be viewed as a three-step process consisting of invading strand hybridisation to the free toehold, branch migration over the competition domain and incumbent strand dissociation. Zhang showed that by modelling the reaction in this way, it was possible to predict the rate of the reactions. The model is discussed further in Section 2.1.2.

1.7.1 Monitoring Strand Displacement Reactions

Often, the speed at which the walker takes a step correlates to the rate at which a DNA fuel or antifuel strand added into the system interacts with the walker. The strand displacements induced by the fuel strands can be monitored by fluorescence spectroscopy to determine a change in the state of the machine. Typically, a DNA strand in the system is labelled with fluorescent dyes, and it is the interaction of the dyes with each other which is observed in fluorescence spectroscopy. The fluorescent dyes usually form a "FRET pair". FRET (Förster Resonance Energy Transfer) describes an energy transfer mechanism between two fluorophores [Förster, 1948], [Stryer and Haugland, 1967]. The energy difference between two different molecular orbitals in the fluorophore molecule falls within the range of the visible spectrum.

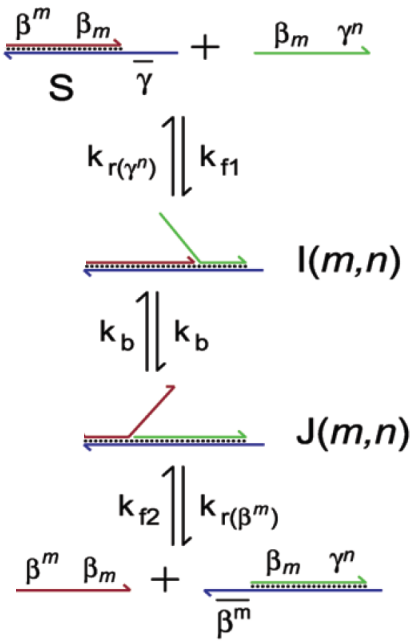


Figure 1.21: Illustration of the toehold exchange process.

See main text for the discussion of the process. Figure adapted from [Zhang and Winfree, 2009].

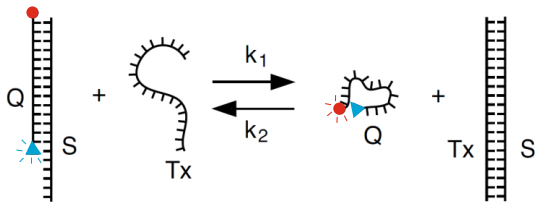


Figure 1.22: Illustration of a toehold-mediated strand displacement monitored by FRET.

The single-stranded region on strand S is the toehold, and it allows invading strand Tx to hybridise and form the more thermodynamically favourable duplex with S. Strand Q is labelled on either end with fluorophores that form a FRET pair. The blue triangle represents the donor molecule, and the red circle represents the acceptor molecule. In the initial state, the donor emission fluorescence is unquenched. Once strand Q is displaced, it exists in a random coil configuration in solution, which brings the fluorophores into close proximity. The resulting quenching of the donor fluorophore (and increased emission fluorescence of the acceptor fluorophore), allows the strand displacement reaction to be monitored by fluorescence spectroscopy. Figure adapted from [Yurke and Mills, 2003].

The fluorophore is able to absorb energy in a broad range of visible wavelengths, and in doing so, an electron is excited from its ground state to an excited state. Generally, the exciton loses energy in a cascade of transitions between excited states before recombining radiatively. Consequently, the fluorophore emits energy of a different, longer wavelength. The emission wavelength is dependent on the specific fluorophore molecule and the chemical environment of the molecule. In a FRET pair, two molecules are coupled. The bluer fluorophore is the "donor" and the other is the "acceptor". A donor fluorophore in its excited state can transfer energy by a non-radiative coupling mechanism to an acceptor fluorophore in close proximity (usually $<$10 nm apart). The result is that the donor emission fluorescence decreases (becomes quenched), while the acceptor emission fluorescence increases. The efficiency of this energy transfer is dependent on the donor-acceptor separation distance r, and specifically on the inverse sixth power of r. Therefore, FRET is sensitive to small changes in r, which makes it a useful tool to quantify molecular dynamics.

Changes in the distance between FRET pairs can be used to monitor the operation of a DNA machine. Figure 1.22 shows how a strand (Q) labelled with fluorophores at either end enables investigation into a toehold-mediated strand displacement reaction. The initial state of the system is a duplex formed between a template strand S and strand Q. In this state, the donor (blue triangle) and acceptor (red circle) molecules are far away from each other, as strand Q is in an extended state as part of a duplex. Once the displacer stand Tx is introduced to the system, strand Q is displaced. Now, the single-stranded random coil conformation of strand Q results in the dyes being much closer together on average, and therefore the donor emission fluorescence becomes quenched by the nearby acceptor molecule. There are many possible FRET pairs [Behlke et al., 2005], and so it is possible to use different pairs to monitor the different stages or steps in the operation of a DNA machine (as long as the different pairs do not interact with each other).

Chapter 2

The Cascade-Coupling Mechanism

2.1 Introduction

Toehold-mediated strand displacements are a useful mechanism to encourage a DNA walker to move along a track, and the synthetic ribosome presented in this chapter also makes use of this technique. In particular, the synthetic ribosome uses toehold exchange to drive the active part of the ribosome along a DNA template, as the migrating free toehold determines the section of the template that the ribosome will "read" next. The mechanism is similar to that used by the DNA walker discussed in Section 1.3 and Figure 1.10. The DNA bipedal walker moved along a ssDNA track with the aid of toehold-mediated strand displacements, which resulted from the addition of fuel strands to the system [Bath et al., 2009]. This walker mechanism is the inspiration behind the design of the synthetic ribosome discussed here. A simple sketch of the synthetic ribosome design, dubbed the cascade-coupling mechanism, is presented in Figure 2.1. A long ssDNA *template* forms a duplex with multiple short ssDNA *target* strands (similar to the feet of the bipedal walker, except these strands are not connected to each other). Each target strand is made up of two sections

⟨u t⟩: a "toehold" domain t and a "universal" domain u.[1] The universal domains are 18 nt in length and the sequence is the same for each target strand (universal domains are shown in black in the diagram). The toeholds are the identifiers of the target strands (much like the codons of the natural ribosome), as each toehold is a unique 6 nt sequence (shown by various colours in the figure). Each toehold sequence encodes a unique chemical building block, which is linked to the 5' end of each target strand.

At the start of the reaction (Figure 2.1 (a)), the chemical building blocks are held far apart from each other by the hybridisation of the target strands to the template. In this initial state, the building blocks are prevented from reacting by the distance separating them. The diagram also shows that there is a single-stranded region at the 3' end of the template strand. This region is called the *initial toehold*, and is necessary to begin the reaction. The cascade-coupling mechanism uses toehold-mediated strand displacements to bring consecutive chemical building blocks into close proximity. The chemical reaction that couples building blocks to each other also includes a group transfer reaction (such as that described in Section 1.5.2), which allows the chemical product to be transferred from one target strand to another. In this way, both the reaction site of the synthetic ribosome and the resulting chemical product move along the template strand.

The target strands are displaced by the addition of short ssDNA *probe* strands to the solution. Each probe strand is also made up of a universal and a toehold domain ⟨t u⟩, but in reverse order compared with the target strands ⟨u t⟩. Probe strands are able to bind to free toehold domains on the template, and use these to

[1] The notation for the domains and how they are connected is borrowed from recent work on creating a programming language for DNA computing [Phillips and Cardelli, 2009]. Single strands are divided into distinct domains, each represented by a number or in this case, the letter t or u. The notation is used here simply to indicate differences between strands.

initiate target strand displacement. While the probe and the target strand P compete over the universal domain, the 5' end of target strand P dissociates from the template and is single-stranded and flexible enough to allow it to reach the 5' end of the neighbouring target strand Q (Figure 2.1 (b)). Now, the building blocks are in close proximity and can react with each other. If the chemical reaction is fast compared to the DNA strand displacement reaction, then the chemistry can take place and the chemical product transfers to the next target strand, Q (Figure 2.1 (c)). Once the probe strand becomes completely hybridised to the template, the target strand is only hybridised to its toehold domain, which is short enough to make the duplex unstable. Thus, the spent target strand P spontaneously dissociates from the template (Figure 2.1 (d)). In doing so, a new free toehold is revealed for another probe strand, and by repeating the strand displacement process, the chemical product will move along the template by the group transfer reaction until the entire length of the template has been "read" and it becomes attached to the final target strand (Figure 2.1 (e-g)).

The strand displacement mechanism used here is entirely reversible. The toehold revealed by the displacement of a target strand can be used by a probe strand (as shown by the transition in Figure 2.1 (d-e)), but it can equally be used by the target strand to rehybridise to the template (Figure 2.1 (d-c)). The probability that this takes place depends on the relative concentrations of probe and target strands. A bigger excess of probe strands increases the probability of a probe strand hybridising to the free toehold. This reversibility is not a feature of the natural ribosome, which does not re-read codons. However, it is not too important that the strand displacement reaction is reversible, as long as the chemical reaction is not reversible. The reversibility may mean that target strands rehybridise and get displaced multiple times, but if the group transfer reaction only occurs the first time the target strand is displaced, then the chemistry is unaffected. The group transfer reaction should ensure uni-directionality of the chemical product, even if the DNA strands are capable

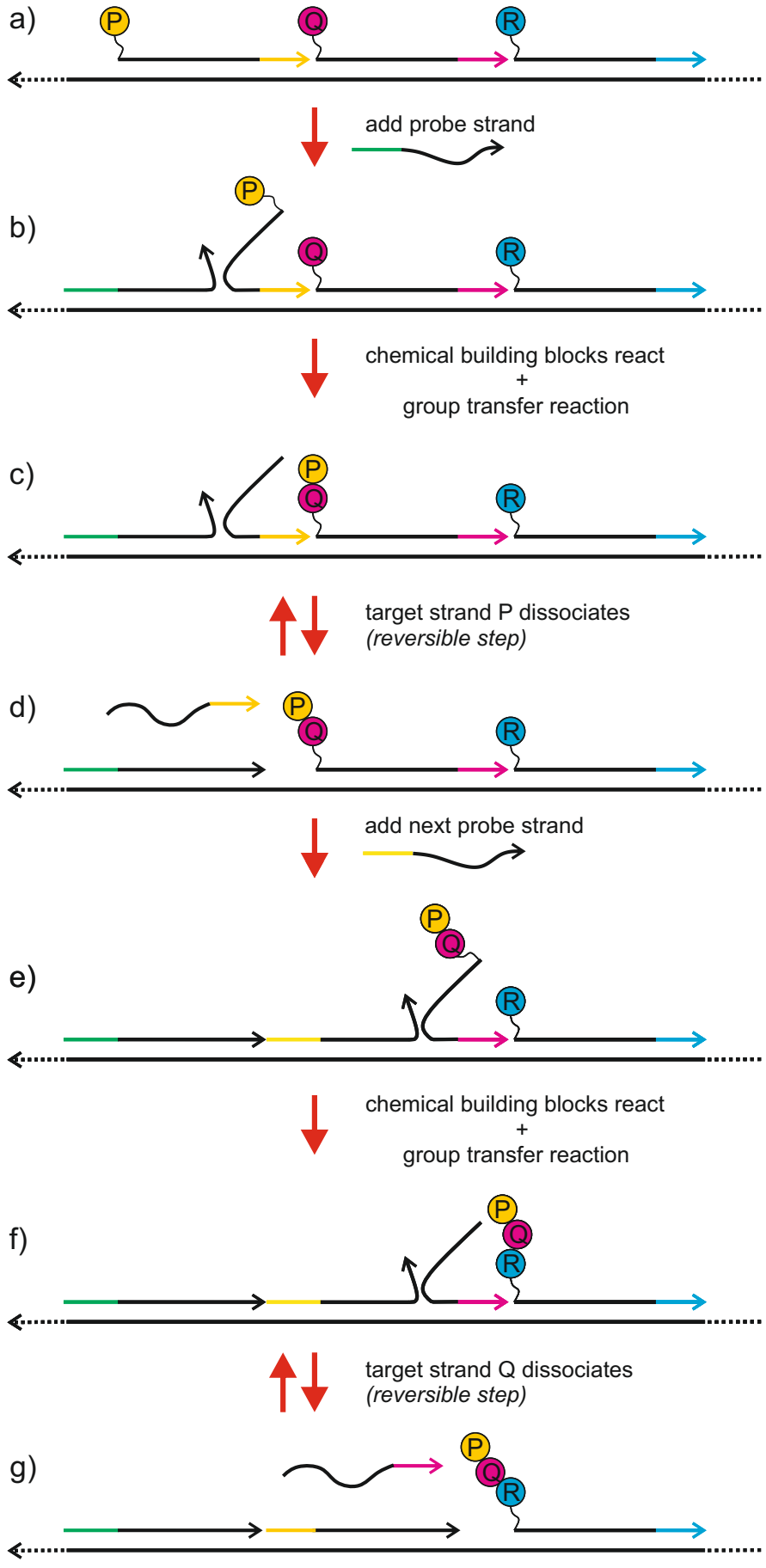

a)
P
Q
R
add probe strand
b)
chemical building blocks react
+
group transfer reaction
c)
target strand P dissociates
(reversible step)
d)
add next probe strand
e)
chemical building blocks react
+
group transfer reaction
f)
target strand Q dissociates
(reversible step)
g)

Figure 2.1: Illustration of the generic cascade-coupling mechanism.

(a) The initial state of the system is a duplex formed between a long *template* strand and several shorter *target* strands. The target strands are comprised of two domains ⟨u t⟩, the black "universal" domain which is the same for each strand, and the coloured "toehold" domain which is unique to each target strand (resembling the codons of mRNA). Chemical building blocks are attached to the 5' end of each target strand by a flexible linker, such as a C6 spacer (shown by the thin, wavey black line). The hybridisation of target strands to the template ensures that the chemical building blocks are initially held far apart, so that they are unable to react.

(b) Toehold-mediated strand displacement enables the building blocks to come into close proximity. Strand displacement is brought about by *probe* strands, which consist of the same two domains but in reverse order ⟨t u⟩. A probe strand hybridises to a free toehold, and competes with a target strand over the universal region.

(c) The distance between neighbouring building blocks decreases as a result of the strand displacement, enabling the chosen chemical reaction to occur. The chemistry includes a group transfer reaction, such as the Wittig reaction described in Section 1.5.2. Here, building block P detaches from target strand P, and the chemical product is transferred entirely to the neighbouring target strand Q.

(d) Once the first displacer strand has fully hybridised to the template, target strand P is only hybridised to the template by its toehold domain. If the toehold length is short (≤ 6 nt), the duplex is thermodynamically unstable, and the target strand spontaneously dissociates from the template entirely. In doing so, it reveals a new free toehold to which another probe strand can hybridise. This step is reversible. The spent target strand is able to rehybridise to the template using the free toehold and compete with the probe strand over the universal region. The chemistry however, is not reversible.

(e) If another probe strand is added, the synthetic ribosome mechanism can continue along the template.

(f) The growing chemical product is transferred to the neighbouring target strand at each step. It moves along the template in the same direction as the migrating free toehold.

(g) The correct sequence of strand displacements requires the probe strand with a specific toehold domain to hybridise to its complementary toehold on the template. The sequence of the toehold domains along the template determines the order in which chemical building blocks are coupled. Thus, the synthetic ribosome "reads" the template and produces a chemical output in a similar way to the natural ribosome.

of moving in both directions.

The natural ribosome synthesizes proteins from the information encoded within messenger RNA (mRNA), as described in Section 1.4. The mRNA is comprised of a sequence of codons that signal to the ribosome the types of amino acid needed to make a particular protein, as well as the order in which these amino acids should be joined together. Each codon is a sequence of three nucleotides that specifies a single amino acid. As the ribosome reads each codon, it joins the appropriate amino acid into a polypeptide chain. The synthetic ribosome described in this section works in a similar way. In the cascade-coupling mechanism, each toehold domain on the template strand can be thought of as a codon that specifies a particular chemical building block. The length of the toehold domain may not be three nucleotides, but the function of the toehold domain is essentially the same as that of the codons in the genetic code. One difference is that the toehold domains in the synthetic ribosome system do not need to only encode for amino acids. This flexibility means the cascade-coupling mechanism can be adapted for the synthesis of a chain of *any* chemical building blocks and therefore, has the potential to generate libraries of new functional molecules, as discussed in Section 1.4.

The result of both the natural and synthetic ribosomes is the same: a chain of building blocks that are attached to each other according to the sequence of the codons on the template encoding the information. In nature, transfer RNA (tRNA) sequentially delivers the amino acids to the ribosome. Each tRNA has a 3' terminal site at which amino acids can be attached, and also consists of an anticodon — three nucleotides that are complementary to a specific codon in the genetic code. The target strands are the synthetic version of tRNA in the cascade-coupling mechanism. Each target strand consists of a toehold domain that is complementary to a toehold region on the template DNA, and each can be attached to a specific chemical build-

ing block. In the synthetic system, these synthetic tRNAs are already lined-up in the order specified by the template DNA, but they are only able to join their chemical building blocks into the growing chain once the strand displacement mechanism has acted on them. Therefore, the proposed synthetic ribosome mechanism shares many of the features of the natural ribosome, whilst also having the potential to synthesize any polymeric molecule.

It is important to note the reason why the toehold domain length was chosen to be at least 6 nt. Duplexes shorter than this length are relatively unstable. Therefore, a toehold that is < 6 nt would make it difficult for a probe strand to form a stable duplex with its toehold domain to enable it to initiate the strand displacement. Equally, it could result in a very unstable duplex between the target strand and its toehold domain, causing the spontaneous dissociation rate from the toehold to quicken, and reducing the probability that the chemical group transfer reaction will occur before dissociation. However, changing the length of the toehold domain does also provide a way to control the toehold exchange reaction rate.

It is clear that there are two important parameters to control in the cascade-coupling mechanism. Firstly, there is the overall target strand displacement rate (or the rate at which the probe strand fully hybridises to the template), and secondly, there is the rate at which the partially-displaced target strand, bound only by its toehold, dissociates from the template (as shown in Figure 1.20b). Both of these rates determine the amount of time that the target strand remains hybridised to the template strand and therefore, the amount of time that the chemical building blocks are in close proximity. The chemistry is controlled by the proximity of building blocks and so, much of the work reported here concerns tuning these two rates. There are a number of methods to do so, such as altering the length of the toehold domains, and changing the concentration of the probe strands. Many of these methods involve

altering the thermodynamic advantage acquired by a strand from hybridisation to its toehold, and therefore it is necessary to understand the thermodynamics of the cascade-coupling mechanism in more detail in order to develop methods to tune the reaction to suit the chemistry.

2.1.1 Thermodynamics of the Cascade-Coupling Mechanism

Extra base pairs acquired from hybridisation to a toehold provide a DNA strand with the thermodynamic advantage that can enable it to displace and replace another strand. The cascade-coupling mechanism incorporates the toehold exchange process described in Section 1.7. A toehold is used by a probe strand to displace a target strand, and once the target spontaneously dissociates from the template, a new toehold is revealed for the next probe strand. If the two toeholds have the same strength, then this toehold exchange is thermodynamically neutral. By adding an excess of probe strands to the reaction, it is more likely that a probe strand will hybridise to the new revealed toehold than that a copy of the displaced target strand will rehybridise. Therefore, probe strand concentration drives the reaction further towards completion, as exchange of a target strand for a probe strand increases the entropy of the system.

Consider a simple one-step toehold exchange reaction in which a probe strand P reacts with the initial duplex Tt, made between a target strand T and a template strand t. The result is a new duplex Pt and a free target strand T:

$$P + Tt \underset{k_r}{\overset{k_f}{\rightleftharpoons}} Pt + T$$

The change in free energy ΔG of the strand displacement reaction can be described by the following relation:

$$\begin{aligned} \Delta G &= \Delta G^o_{Pt} + \Delta G^o_T - \Delta G^o_P - \Delta G^o_{Tt} + RT\ln(Q) \\ &= \Delta G^o_{net} + RT\ln(Q) \end{aligned} \tag{2.1}$$

where ΔG_x^o is the standard free energy of species x, R is the universal gas constant, T is the temperature of the system (usually 21 °C), and Q is the reaction quotient relative to standard conditions, such that:

$$\begin{aligned} Q &= \frac{[\text{free target strands T}][\text{final state Pt}]}{[\text{free probe strands P}][\text{initial state Tt}]} \\ &= K \text{ (at equilibrium only)} \end{aligned} \tag{2.2}$$

The reaction quotient Q is a function of the concentrations of the various species involved in the cascade reaction. Here, intermediate species are ignored for simplicity. At equilibrium, the reaction quotient equals the equilibrium constant K, which describes the concentrations of the four species specifically at equilibrium. Furthermore, $\Delta G = 0$ at equilibrium, so Equation 2.1 becomes:

$$\begin{aligned} 0 &= \Delta G_{net}^o + RT\ln(K) \\ \Delta G_{net}^o &= -RT\ln(K) \end{aligned} \tag{2.3}$$

This equation relates the net standard free energy of the reaction to the concentration at equilibrium of the four species involved in the reaction. Consider the thermodynamically neutral case, in which the toeholds are matched such that $\Delta G_{net}^o = 0$. Equation 2.3 provides useful information about the extent of the forward reaction given the initial conditions. At equilibrium, x moles of the probe strands have hybridised to the template strand, and therefore x moles of the target strands have been displaced. If probe strands are added in an α–fold excess concentration with respect to the target strands (that is, α units of probe strands for every 1 unit of target-template duplex), then:

$$\begin{aligned} K &= \frac{[\text{free target strands}][\text{final state}]}{[\text{free probe strands}][\text{initial state}]} \\ &= \frac{[x][x]}{[\alpha - x][1 - x]} \end{aligned} \tag{2.4}$$

Therefore, when $\Delta G^o_{net} = 0$, Equation 2.4 becomes:

$$\begin{aligned} \exp(0) &= K \\ 1 &= \frac{[x][x]}{[\alpha - x][1 - x]} \\ \Rightarrow x &= \frac{\alpha}{\alpha + 1} \\ &= \text{fraction completed} \end{aligned} \tag{2.5}$$

In the equimolar case ($\alpha = 1$), Equation 2.5 shows that only half of the Tt duplexes have been replaced by Pt duplexes. If the probes were added in a 10-fold excess instead ($\alpha = 10$), the higher initial probe concentration results in 90 % of the target strands being displaced at equilibrium. Therefore, changing the concentration of the reactants in the toehold exchange has an effect on the extent of the reaction at equilibrium.

In the instance where the net standard free energy change of the reaction is not zero ($\Delta G^o_{net} \neq 0$):

$$\begin{aligned} \exp(\frac{-\Delta G^o_{net}}{RT}) &= K \\ \exp(\frac{-\Delta G^o_{net}}{RT}) &= \frac{[x][x]}{[\alpha - x][1 - x]} \end{aligned}$$

Let S = $\frac{\Delta G^o_{net}}{RT}$, such that

$$\exp(-S) = \frac{[x][x]}{[\alpha - x][1 - x]}$$

Therefore, in general, the percentage completion is given by solving:

$$0 = \alpha - (\alpha + 1)x + (1 - \exp(S))x^2 \tag{2.6}$$

$$\begin{aligned} \Rightarrow x &= \frac{(\alpha + 1) \pm \sqrt{(\alpha + 1)^2 - 4\alpha(1 - \exp(S))}}{2(1 - \exp(S))} \\ &= \text{fraction completed} \end{aligned}$$

The extent of the reaction at equilibrium depends on the thermodynamic gain acquired by forming duplex Pt (Equation 2.6). The larger the net gain in free energy,

the further the reaction goes towards completion.

The equilibrium constant K relates ΔG^o_{net} to the rate constants k_f and k_r of the reaction:

$$\begin{aligned}
\text{rate of forward reaction} &= k_f[P][Tt] \\
\text{rate of reverse reaction} &= k_r[T][Pt] \\
\text{At equilibrium, } k_f[P][Tt] &= k_r[T][Pt] \\
\Rightarrow K &= \frac{k_f}{k_r} \\
&= \frac{[T][Pt]}{[P][Tt]} \\
&= \exp(\frac{-\Delta G^o_{net}}{RT}) \qquad (2.7)
\end{aligned}$$

In the case of simple toehold-mediated strand displacement (as shown by Figure 1.22), there is a strong relationship between the kinetics and the thermodynamics of the reaction, because the reaction rate can be increased simply by strengthening the toehold (either by increasing the length of the region or altering the sequence to contain more G-C base pairs) [Yurke and Mills, 2003], [Zhang et al., 2007]. Strengthening both toeholds increases both the forward and backward reaction rates. To understand the toehold exchange process so that the reaction can be tuned for the cascade-coupling chemistry, the reaction can be described as a three-step process governed by the rates at which the two strands hybridise to their toeholds and the rate of branch migration over the universal domain [Zhang and Winfree, 2009].

2.1.2 Modelling the Rate of the Toehold Exchange Reaction

It is possible to model the toehold exchange process (as illustrated by Figure 2.2), and to use the model to predict the rate of the reactions [Zhang and Winfree, 2009]. Although Zhang considers the toehold exchange reaction as a three-step process, he also demonstrates that a simplified bimolecular model can be used to determine the

forward rate constant of the probe strand hybridisation, and the reverse rate constant of the target re-hybridisation. He describes the reaction as:

$$P + Tt \underset{k_{zr}}{\overset{k_{zf}}{\rightleftharpoons}} T + Pt \tag{2.8}$$

where the bimolecular rate constants k_{zf} and k_{zr} are related to the forward and reverse second-order rate constants of the three-step process. He derives the relation between them as:

$$k_{zf} \equiv \frac{k_{TD}\ k_{HT}\ k_{BM}}{k_{PD}\,k_{TD} + k_{PD}\,k_{BM} + k_{TD}\,k_{BM}} \tag{2.9}$$

Here, k_{TD} is the rate at which a target strand dissociates from its toehold, k_{PD} is the rate at which a probe strand dissociates from its toehold, k_{BM} is the branch migration rate over the universal domain and k_{HT} is the rate at which a strand hybridises to a toehold (and is assumed to be the same for probes and targets for simplicity). A similar relation can be derived for the relation between these quantities and the reverse bimolecular rate constant k_{zr}.

The second-order rate constants are dependent on the sequences of the toehold and universal domains.

$$k_{TD} = k_{HT} \times \frac{2}{\text{universal domain length}} \exp\left(\frac{\Delta G^o(TD)}{RT}\right) \tag{2.10}$$

$$k_{PD} = k_{HT} \times \frac{2}{\text{universal domain length}} \exp\left(\frac{\Delta G^o(PD)}{RT}\right) \tag{2.11}$$

$$k_{BM} \approx 1.0\,\mathrm{s}^{-1} \tag{2.12}$$

$$k_{HT} = 10^5 - 10^6\,\mathrm{M}^{-1}\mathrm{s}^{-1} \tag{2.13}$$

Zhang parameterises the branch migration rate with a single phenomenological rate constant.[2] The rate k_{HT}, which explains how quickly the strands hybridise to their

[2]Using a single rate constant for the branch migration rate instead of modelling it as a random walk process ignores the intermediate states produced during branch migration. Zhang's choice of k_{BM} gives rise to a branch migration rate of 2.5 ms/step, which is much slower than previously reported studies on the branch migration rate of 10-100 µs/step of genomic-length DNA. However, to

toehold domains, was found to vary by an order of magnitude depending on the strength of the toeholds. The rates k_{TD} and k_{PD} depend on the free energy acquired from hybridisation of the target and probe strand to their toehold domains.[3] Therefore, this description of the toehold exchange reaction offers ways to alter the bimolecular reaction rate by altering the sequences (and therefore the energy of) the different domains in the cascade.

In the case where the free energy difference between the initial state and the final state is minimal or zero, the bimolecular rate constants k_{zf} and k_{zr} are closely balanced. Depending on the strength of the two toeholds (which for the moment are assumed to be of equal length), the forward and reverse reaction speeds can both be fast. The model is not entirely accurate, since it does not factor in the formation of intermediate states (for example, the duplex formed when the invading strand is hybridised to the template by its toehold domain only, as shown by complex I(m,n) in Figure 1.21). These intermediate states become more important when the free energy difference between the two duplexes is greater, and in particular, when the energetically favoured complex is the initial duplex. Zhang found that when the target strand toehold provides the target strand with more energy than the probe strand toehold provides the probe strand, the bimolecular forward rate constant of the toehold exchange reaction decreases with increasing target toehold strength [Zhang and Winfree, 2009]. Consequently, the probe strand either dissociates from its toehold entirely, or completes the branch migration step and fully hybridises to the template.

simplify matters and because it is difficult to select a particular rate from previous studies, Zhang's approximation is used to model the cascade here.

[3]The term 2/(universal domain length) in Equations 2.10 and 2.11 is necessary to ensure the correct relative concentrations of the branch migration intermediate states at equilibrium [Zhang and Winfree, 2009].

The experiments carried out to test the toehold exchange reaction are described in Section 2.2, and illustrated by Figure 2.2. Typically, a fluorescently-labelled target strand is displaced from the template strand by a probe strand. A reporter strand that is permanently hybridised to the template strand is also labelled with a fluorophore. The two fluorophores form a FRET pair, as described in Section 1.7.1. Two FRET pairs are used in the systems discussed here: Cy3 and Cy5, and Fam and Tamra (see Appendix A.1.1 for details). In the initial state of the system, the two fluorophores are in close proximity and so, the donor fluorescence is quenched. Once the probe strand is added to the system, the subsequent displacement of the target strand causes the distance between the fluorophores to increase. Thus, the donor fluorescence increases. It is this change in donor fluorescence which is measured in the FRET experiments discussed here, and which is conventionally reported in the graphs. The rate at which the donor fluorescence changes is proportional to the rate at which the fluorescently-labelled target strand is displaced. In a multi-step cascade, the fluorescently-labelled strand is the final target strand, so that the fluorescence measurements correspond to the rate at which the final target strand is displaced. As a result, there may be a delay between adding the probe strands to the system and observing any change in the fluorescence signals. It is possible to predict the rate at which the fluorescence signal changes in a multi-step cascade by using the bimolecular model to predict the rate constants for the individual steps of the cascade. The fluorescence signal change observed in an experiment when the fluorescent target strand is displaced is directly proportional to the concentration of the free fluorescent target. Consider a two-step cascade in which:

$$P1 + T1T2t \underset{k_{zr1}}{\overset{k_{zf1}}{\rightleftharpoons}} T1 + P1T2t$$

$$P2 + P1T2t \underset{k_{zr2}}{\overset{k_{zf2}}{\rightleftharpoons}} T2 + P1P2t$$

The initial state is the duplex T1T2t made between a template t and two target strands T1 and T2. When probe P1 is added to the system, it displaces target T1

at rate k_{zf1}. Now probe P1 and target T2 are hybridised to the template (P1T2t). The second step is initiated when probe P2 is added to the system, which displaces target T2 at a rate k_{zf2}. The final state at the end of the reaction is the duplex P1P2t. There are seven distinct species in this system (P1, T1T2t, T1, P1T2t, P2, T2, P1P2t), though the concentration of some of these species is zero at the start of the reaction (for example, there is no free T1 at the start of the reaction). This two-step reaction can be described by the following set of ordinary differential equations, which can be solved numerically to determine the change in concentration of each species over time:

$$\begin{aligned}
\frac{d(P1)}{dt} &= -k_{zf1}(P1)(T1T2t) + k_{zr1}(T1)(P1T2t) \\
\frac{d(T1T2t)}{dt} &= -k_{zf1}(P1)(T1T2t) + k_{zr1}(T1)(P1T2t) \\
\frac{d(T1)}{dt} &= +k_{zf1}(P1)(T1T2t) - k_{zr1}(T1)(P1T2t) \qquad (2.14) \\
\frac{d(P1T2t)}{dt} &= +k_{zf1}(P1)(T1T2t) - k_{zr1}(T1)(P1T2t) \\
&\quad -k_{zf2}(P2)(P1T2t) + k_{zr2}(T2)(P1P2t) \\
\frac{d(P2)}{dt} &= -k_{zf2}(P2)(P1T2t) + k_{zr2}(T2)(P1P2t) \\
\frac{d(T2)}{dt} &= +k_{zf2}(P2)(P1T2t) - k_{zr2}(T2)(P1P2t) \qquad (2.15) \\
\frac{d(P1P2t)}{dt} &= +k_{zf2}(P2)(P1T2t) - k_{zr2}(T2)(P1P2t)
\end{aligned}$$

The reversibility of the toehold exchange reaction means that even once the second target strand has been displaced, it is possible for the entire process to revert to the initial state. Thus, the extra terms for (P1T2t) describe how it is possible to create the species when the second step is reversed. If a 10-fold excess of probe strands is used, the initial conditions of the reaction are the concentrations of the seven species (in nM):

$$(P1, T1T2t, T1, P1T2t, P2, T2, P1P2t)_{t=0} = (10, 1, 0, 0, 10, 0, 0)nM$$

Here, the rate of change of the concentration of target T2 (described by Equation 2.15) corresponds to the rate of the fluorescence signal change measured in an experiment

(as it is always the final target strand which is fluorescent). This model, or a variation of it, is used to obtain the data fits presented in the following sections.

To summarise, the rate of the toehold exchange reaction and the equilibrium concentrations are linked to the change in the free energy of the system arising from the different energies of the toehold domains. The toehold exchange reaction can be broken down into three distinct steps, (toehold hybridisation, branch migration and toehold dissociation), each of which can be used to provide a way to control the overall rate of the reaction. This is important, as it allows the speed at which the target strand dissociates from the template strand to be controlled to provide the time required for the cascade-coupling chemical reactions to occur. The effect on the overall bimolecular rate constant of altering the free energies of the toehold domains in a simple one-step toehold exchange reaction is investigated in the following section.

2.2 One-Step Displacement Reaction

As discussed earlier, in order for the toehold exchange mechanism to bring successive chemical building blocks into close proximity to react, it is necessary to be able to control the bimolecular rate of the strand displacement. Strand displacement reactions of this sort type can be very quick but, if they are too fast, the chemistry will not have enough time to take place, and the yield of the final product of a multi-step cascade will be low.

2.2.1 Altering the Length of the Target Toehold

A simple experiment was designed to determine the order of magnitude of the bimolecular rate of a toehold exchange reaction. As illustrated in Figure 2.2, the initial state of the reaction is a duplex made between three strands — a template, a target strand and a reporter strand. The target toehold domain is shown in blue. The

template strand has an exposed toehold at its 3' end, which is used by the invading probe strand to initiate the displacement of the target strand. The target strand is labelled on its 3' end with a Cy3 fluorophore, and the reporter strand, which remains hybridised to the template throughout the reaction, is labelled with a Cy5 fluorophore on its 5' end. The fluorescent molecules Cy3 and Cy5 form a FRET pair (as described earlier in Section 1.7.1 and Appendix A.1.1), in which Cy3 is the donor fluorophore and Cy5 is the acceptor fluorophore.

In the initial state of the reaction, the fluorophores are located on neighbouring base pairs such that the Cy3 fluorophore is quenched by Cy5. Monitoring the emission of the donor fluorophore during the reaction enables measurement of the target strand displacement rate. At the start of the reaction, the Cy3 emission fluorescence is low: this forms the baseline of the fluorescence measurements. Once the probe strand is added to the system, it is able to hybridise to the exposed toehold domain (pink) and compete with the target strand over the universal domain. When the target strand dissociates from its toehold domain, the increased distance between the two fluorophores results in an increase in the Cy3 emission fluorescence.

In this test, the probe strand and target strand toeholds are not the same length. The probe toehold domain is 12 nt, and the target toehold length is varied to investigate the effect of altering the target toehold on the rate of the reaction. The target toehold length is altered by extending the probe strand at the 3' end, thereby extending the length of the universal domain over which both strands compete to hybridise and reducing the length of the target toehold. Five probe strands are used in this test. Probe strands A–E compete with the target strand over a universal domain of 20–16 nt respectively and, in doing so, change the target toehold length to 4–8 nt respectively. Table 2.1 shows the standard free energies of the two toeholds in each of the five cases, and the predicted bimolecular rate constants calculated using the

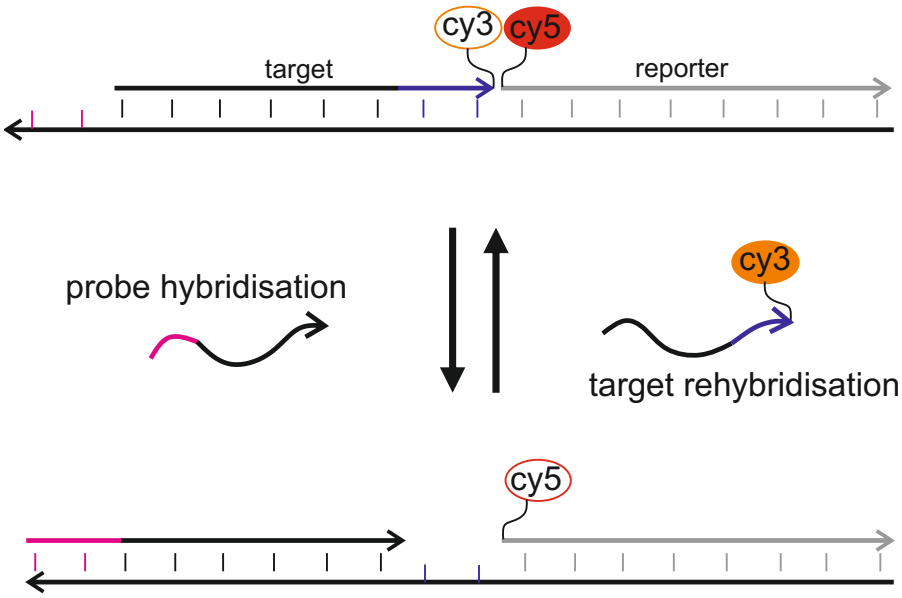


Figure 2.2: Depiction of the one-step reaction mechanism.

The initial state of the system is a duplex formed between a target strand, a reporter strand and a template strand. The target strand is labelled with a 3' Cy3 fluorophore, and the reporter strand with a 5' Cy5 fluorophore. The emission of the Cy3 fluorophore is monitored during the reaction to determine the rate of displacement of the target strand (see Appendix A.1.1) The proximity of the fluorophores in the initial state means that the Cy3 fluorescence is quenched.

Upon addition of a probe strand, which is capable of hybridising to the probe toehold, the probe and target strands compete over the universal domain of the template. The final state of the system is a duplex between the probe, reporter and template, with the target strand free in solution. As a result of the increased distance between fluorophores, Cy3 is no longer quenched: the rate at which the Cy3 emission fluorescence increases correlates to the rate of the toehold exchange reaction.

The reaction is reversible, as the displacement of the target strand reveals a toehold for it to re-hybridise to, enabling displacement of the probe strand.

model described in Section 2.1.2. The DNA strand design criteria and constraints are described below.

DNA strand design criteria

As described in Section 1.3.1, there are a number of software tools to aid the design of DNA sequences, each of which analyses a particular property of the DNA. For the DNA systems presented in this chapter, two properties were considered important.

1. Sequence specificity: This is essential to ensure that DNA strands self-assemble in the designed way. In the context of the cascade-coupling mechanism, the target and probe strands hybridise to the correct positions on the template strand because of the specificity of their toehold domains. If toeholds contain common sequences within them, there is a greater probability of the target and probe strands hybridising to incorrect toehold domains. The pairwise interactions between strands must be minimised to obtain strand specificity.

2. Secondary structure: Self-complementary regions within DNA strands can cause single strands to fold up and hybridise to themselves. This self-hybridisation may affect the correct assembly of a DNA mechanism or its operation, particularly if the secondary structure occurs in the toehold domain.

3. Toehold free energy similarity: This property is important in multi-step cascade-coupling systems. In such systems, each target toehold is also a probe toehold and therefore, it is impossible to bias one step in the cascade towards a target strand (for instance) without simultaneously and unintentionally biasing the subsequent step towards the probe strand. It is easier to design multi-step systems with unbiased toehold domains, such that each toehold in each step of the reaction provides each strand with approximately the same free energy.

For each DNA system, the selection of the toeholds was aided by the Computer-Aided Nucleic Acid Design Package (CANADA) [Feldkamp, 2010], which can not only generate sequences of a particular length, but can also calculate the strength of the pairwise interactions of the sequences with each other. The package was used to select toehold sequences which were significantly different in sequence, such that there were minimal pairwise interactions between toeholds (and between the complements of the toeholds). This is necessary to ensure that the target strands hybridise to the correct position on the template strand, which is directed by the toehold domains, and also to ensure that probe strands can only displace target strands they were specifically designed to react with. In the case of multi-step systems (Section 2.3 and 2.4), the toehold sequences were also required to have similar free energies. However, in using this method to design the system, the secondary structure of the single-stranded species (toehold plus universal domain) was largely ignored. DNA sequences were selected, after applying the sequence specificity criteria, by choosing those with the weakest secondary structures within them (using mfold to predict the secondary structure for each sequence in the set [Zuker, 2003]). It was difficult to apply the secondary structure criteria more rigourously, as it reduced the total number of available sequences. This clearly has a drawback — if free strands have significant secondary structure, the bimolecular model (Section 2.1.2) may not be applicable to this system, and the assembly and operation of the synthetic ribosome may be affected. However, the sequence specificity was considered the more important design criteria at this stage.

The standard free energies quoted for each probe and target strand in the cascade-coupling systems presented here are obtained from the online nucleic acid analysis tool Nupack [Zadeh et al., 2011], [Dirks et al., 2007]. The values of ΔG^o are the standard free energies corresponding to the hybridisation of each strand's toehold domain to its complement, at 21 °C and in a buffer containing 1.0 M NaCl. The sequences for each DNA strand are listed in Appendix A.4. In Sections 2.3 and 2.4, where

the system consists of multiple toehold domains of equal length, Nupack was used after applying the sequence specificity criteria to select those toehold sequences with similar free energies.

All the DNA strands used in the experiments presented here were purchased from Integrated DNA Technologies (IDT). DNA is shipped lyophilised, and is rehydrated either in purified water or in a salt buffer. Typically, the strands were rehydrated in NEB2 buffer (50 mM NaCl, 10 mM Tris-HCl EDTA, 10 mM $MgCl_2$, pH 8.0), as this is the buffer used in most of the toehold-exchange experiments. Unless otherwise stated, DNA strands were purchased unpurified. Purification is recommended by IDT for DNA strands which are greater than 40 nt in length and for strands that are modified. All unmodified target and probe strands were bought unpurified, and were not purified before use (by gel purification, for instance A.1.4). This means that any incorrectly synthesised strands in the samples were not removed prior to use. The likelihood of an error to be made in synthesising the DNA increases with length and therefore, since the target and probe strands are short oligomers, the strands purchased from IDT were assumed to be pure and used without purification. In addition, the assembled initial state of a reaction was not purified before starting the reaction to remove excess strands and unwanted duplexes. Instead, titration experiments were performed to obtain the minimum concentrations of the DNA strands required to assemble into the initial state duplexes. It was assumed that this would be sufficient to minimise the effect of excess strands and incorrectly assembled products on the overall strand displacement reactions.

Samples were made by annealing equimolar quantities of the target, reporter and template strands in NEB2 buffer (50 mM NaCl, 10 mM Tris-HCl EDTA, 10 mM $MgCl_2$, pH 8.0). Samples were annealed by heating the solution to 85 °C to melt all possible structures, and cooling to 20 °C over 10 min in an Eppendorf Mastercycler

Probe ⟨t u⟩	ΔG^o_{probe} **(kcal/mol)**	**Target** ⟨u t⟩	ΔG^o_{target} **(kcal/mol)**	$\mathbf{k}_{zf}$ $(\mathbf{M^{-1}s^{-1}})$	$\mathbf{k}_{zr}$ $(\mathbf{M^{-1}s^{-1}})$
A ⟨12 20⟩	-16.54	⟨20 4⟩	-7.18	5.5×10^{5}	7.48×10^{-2}
B ⟨12 19⟩	-16.54	⟨19 5⟩	-8.70	5.5×10^{5}	9.75×10^{-1}
C ⟨12 18⟩	-16.54	⟨18 6⟩	-11.33	5.5×10^{5}	8.28×10^{1}
D ⟨12 17⟩	-16.54	⟨17 7⟩	-13.93	5.43×10^{5}	6.61×10^{3}
E ⟨12 16⟩	-16.54	⟨16 8⟩	-15.72	4.4×10^{5}	1.10×10^{5}

Table 2.1: The standard free energies of the probe and target toehold domains in the one-step cascade reaction at 21 °C. The free energies correspond to the coloured regions in the DNA strands in Figure 2.2. The toehold sequence (and therefore the free energy) of all five probe strands is constant, while the toehold domain of the target strand changes. The free energy of the 8 nt target toehold domain is very close to that of the 12 nt probe strand toehold, even though it is shorter. This is primarily due to the sequence composition of the two toehold domains (as discussed in Section 1.1). The G-C content of the probe toehold is around 30 %, while the G-C content of the 8 nt target toehold is 75 % (see Appendix A.4 for the strand sequences).
The toehold free energies are used to calculate the predicted bimolecular rate constants k_{zf} and k_{zr} of the reactions as detailed in Section 2.1.2, using a toehold hybridisation rate $\mathrm{k}_{HT} = 5.5\times 10^{5}\,\mathrm{M^{-1}s^{-1}}$.

PCR machine.[4] All the fluorescence spectroscopy experiments described here were carried out using 1.5 ml quartz cuvettes to load the sample into the fluorimeter (described further in Appendix A.1.2). Sample concentrations were 6.7 nM in a 1.5 ml total volume. The experiments were conducted at 21 °C, and both the Cy3 and Cy5 emission signals were recorded. In order to obtain a "long-time" value of the fluorescence for the purpose of normalising the data, samples were annealed *in situ* at the end of the experiment and a baseline recorded.

[4]Before carrying out fluorescence spectroscopy experiments, the samples were analysed using Poly-(acrylamide) Gel Electrophoresis (PAGE) to verify that the strands had hybridised together (see Appendix A.1.3 for details on PAGE.)

Probe	ΔG^o_{probe} **(kcal/mol)**	ΔG^o_{target} **(kcal/mol)**	$\mathbf{k}_{HT}$ **($M^{-1}s^{-1}$)**	$\mathbf{k}_{zf}$ **($M^{-1}s^{-1}$)**	$\mathbf{k}_{zr}$ **($M^{-1}s^{-1}$)**
D	-16.54	-13.93	1.2×10^4	1.2×10^4	1.44×10^2
E	-16.54	-15.72	8.0×10^2	6.4×10^2	1.6×10^2

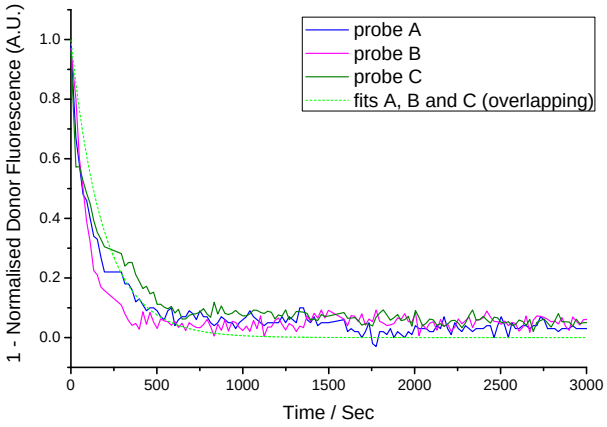
Table 2.2: The bimolecular rate constants calculated in Table 2.1 do not fit the data from the one-step reactions initiated using probes D and E (see Figure 2.3). It is possible that one value of k_{HT} (used to calculate the rate constants) to fit the data for all five experiments is not correct, as the hybridisation rate of a strand to its toehold may be disproportionately affected by the strength of base pairs. The values of k_{HT} here produce rate constants that better fit the data.

Probe strands were added in a 10-fold excess to each sample. The results are shown in Figure 2.3. The effect of adding probes A–C is very similar, however, probes D and E produce significantly slower displacement reactions. The data for probes A–C was fitted using the bimolecular model (Section 2.1.2), using the best-fit value of $k_{HT} = 5.5 \times 10^5\,M^{-1}s^{-1}$. The predicted bimolecular rate constants for each step are shown in Table 2.1.[5] The model predicts that the rate of the reactions with probes A–C are similar, as observed, because the free energy of the probe toehold is significantly larger than the free energy of the target toehold in these cases. As the difference in the free energy decreases, the reverse rate constant increases, such that in the test using probe E, the forward and reverse rate constants are of the same order of magnitude. However, the parameters in the table do not fit the experimental data obtained using probes D and E. Figure 2.3b shows that the model predicts the reaction rate to be orders of magnitude faster than experimentally observed. The manual fits shown in the graph were obtained using the parameters in Table 2.2. It is clear that the bimolecular model has some drawbacks. Firstly, the value of k_{HT}

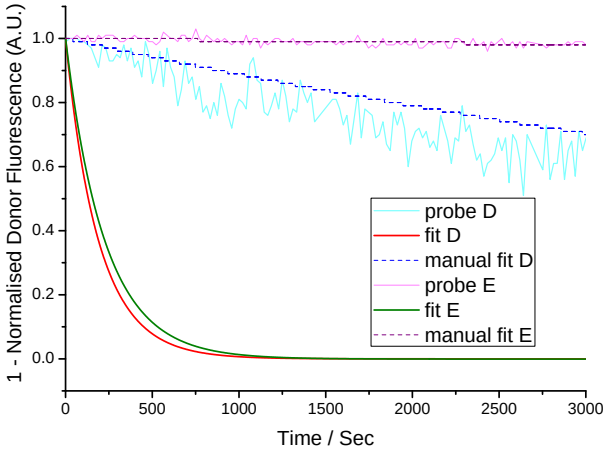
[5]The value of k_{BM} is calculated as 1.0 × (target strand length)/(universal domain length)^2 s^{-1}. This is to take into account changes in the branch migration length [Zhang and Winfree, 2009].

(the rate at which strands hybridise to their toeholds) used in the model is not easy to estimate. For instance, it is noted in [Zhang and Winfree, 2009] that strong base pairs have a disproportionate effect on the value of k_{HT}. Toeholds that were around 30 % stronger than a control set (but of the same length) were modelled using a value of k_{HT} that was twice as large. The strength of the target toeholds in cases D and E are double that of case A (though this increase in the target toehold free energy results from the increase in the length of the toeholds, and not from a change in the composition of fixed-length toeholds as in Zhang's experiments). It is possible that a different value of k_{HT} is necessary to model these reactions however, this does not explain why the same value of k_{HT} cannot be used for both cases D and E. It is likely that the model is being affected by something it does not account for, for instance, secondary structure. Zhang states that the strands used in his toehold exchange experiments were carefully designed to avoid secondary structure in the single-stranded species. If the toehold domains do form secondary structures, this could strongly reduce the strength and rate of toehold hybridisation. The strands used here were not as carefully designed (as outlined in the section on design criteria): secondary structure, weakening some toehold interactions, may go some way to explain why the reactions using probe strands D and E are so much slower than those using strands A–C.

All of the strands involved in this experiment were checked for likely secondary structure formation using mfold [Zuker, 2003] when selecting the sequences, but the secondary structure criteria was not prioritised (as discussed earlier). Therefore, it was found that the target strand and each probe strand formed secondary structures at the experimental temperature, but that these structures were localised in the universal domain. Probes D and E did not fold into more stable secondary structures than other probes. Secondary structure formation may be more important in cases D and E, where the overall change in free energy between the initial and final states



(a) The one-step reaction using probes A-C.



(b) The one-step reaction using probes D and E.

Figure 2.3: Data from the one-step toehold exchange reaction using probe strands A–E.
(a) Invading probe strands A–C cause the target to dissociate quickly, as the energy acquired from the long probe toehold drives the reaction forward. The bimolecular model (Section 2.1.2) also predicts this, since the strong bias towards the 12 nt probe toehold over the short target toehold in each case causes the overall toehold exchange reaction to be fast. The fits from the model all overlap each other.
(b) Probes D and E produce significantly slower reactions, and the bimolecular model does not predict this behaviour well. See main text for a discussion.
In each case, sample buffer was 50 mM NaCl, 10 mM Tris-HCl EDTA, 10 mM $MgCl_2$ pH 8.0. Sample concentrations were 6.7 nM in a 1.5 ml total volume, and the temperature was set to 21 °C. Cy3 (donor) fluorescence was monitored during the reaction using Ex550/Em565 nm. Probe strands were added in 10-fold excess. Data is normalised with respect to the donor fluorescence signal prior to probe strand addition, and to the baseline obtained after the sample was annealed *in situ* at the end of the experiment ("long-time" value), as described in Appendix A.1.2. The fits are normalised using the long-time value of the 100-fold excess model.

is relatively small. Since the target and probe strands have similar free energies, the rate constants of the reaction are close to being balanced, and the strands compete with each other over hybridisation for longer than in cases A–C. This means that secondary structure may be repeatedly formed and melted, which may result in the overall rate of the reaction slowing down.

The toehold exchange reaction might also be affected by the stacking interaction between the two fluorophores. It has been reported that when two dyes are located on neighbouring base pairs, the interaction between them can contribute additional stability equivalent to an extra base pair to the DNA strands they are attached to [Moreira et al., 2005]. When the difference in the free energy of the two toeholds is minimal, the extra stability imparted to the target strand by the stacking interaction may make the initial state of the system slightly more thermodynamically favourable, increasing the relative importance of the reverse reaction.

The experiments demonstrate how adjusting the length of one of the toeholds in the toehold exchange reaction (to change its strength) can provide a method to control the overall reaction rate. However, it is unclear whether the results in Figure 2.3 are primarily due to a changing toehold length or other factors, such as the rate at which secondary structures melt. To investigate this further, a new system of strands was designed with minimal secondary structure in the single-stranded species.

2.2.2 Equal Length Probe and Target Toeholds

The initial one-step experiments suggest that it is possible to alter the rate of toehold exchange by altering the strength of the target toehold. A variation of these experiments was devised in which the two toeholds are the same length (6 nt), but the probe toehold sequences are modified to change the overall free energy difference between the initial and final states. The strands were also designed carefully to min-

imise the formation of unintended secondary structures (whilst also ensuring toehold sequence specificity), as discussed earlier in the section on design criteria. Three tests were designed, and Table 2.3 shows the free energies of the probe and target toehold domains in each case. The characteristics of the three tests are:

1. the toehold sequences are different but provide the strands with approximately the same free energy (*Similar* test),

2. the probe toehold has a higher free energy than the target toehold, providing the probe strand with the thermodynamic advantage (*Downhill* scheme),

3. the target toehold has a higher free energy than the probe toehold, providing the target with the thermodynamic advantage (*Uphill* scheme).

Test	ΔG^o_{probe} **(kcal/mol)**	ΔG^o_{target} **(kcal/mol)**	$\mathbf{k}_{zf}$ $(\mathbf{M^{-1}s^{-1}})$	$\mathbf{k}_{zr}$ $(\mathbf{M^{-1}s^{-1}})$
Similar	-9.92	-10.21	1.9×10^5	3.1×10^5
Uphill	-8.48	-10.21	2.55×10^4	4.74×10^5
Downhill	-11.77	-10.21	4.67×10^5	3.35×10^4

Table 2.3: The standard free energies of the probe and target strand toehold domains in the one-step minimal secondary structure reaction at 21 °C. The reaction rates are calculated using $k_{HT} = 5 \times 10^5$ $M^{-1}s^{-1}$ and the model outlined in Section 2.1.2.

The initial state for each of the three tests was a duplex between a template strand, a reporter strand labelled with a 5' Fam fluorophore, and a target strand labelled with a 3' Tamra fluorophore (similar to the schematic shown in Figure 2.2). The two fluorophores form a FRET pair, in which Fam is the donor and Tamra the acceptor (Appendix A.1.1). In the experiments described in Section 2.2.1, the two fluorophores were located on neighbouring base pairs, and it was suggested that the stacking interaction between them could provide the target strand with extra stability

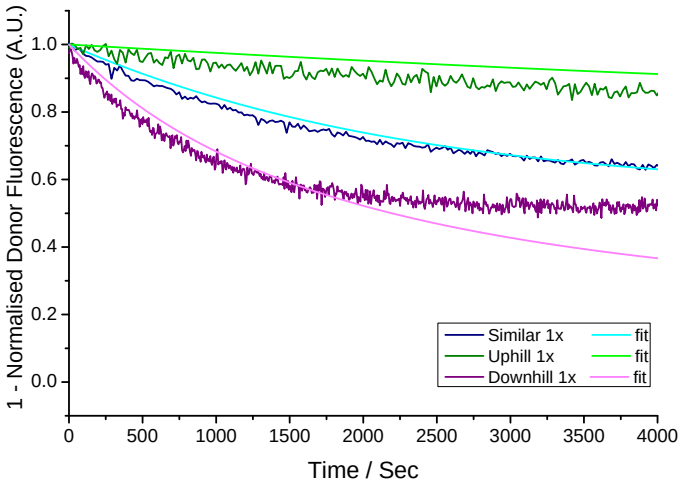
unintentionally. Since the two toehold energies in each of the three tests here have been specifically chosen, the extra stability from the fluorophore interaction could interfere with the designed experiments. To avoid this, three unpaired T nucleotides were added to the template between the reporter and target strands to minimise the fluorophore stacking interaction. The reactions were monitored using fluorescence spectroscopy, and the rate of the toehold exchange was measured by monitoring the Fam emission. Probe strands were added to samples in either a 1× excess (equimolar concentration) or 10× excess. Samples were annealed *in situ* at the end of the experiment to obtain a "long-time" value of the fluorescence. The data for each of the three tests is normalised with respect to the post-anneal value of the 10× excess sample for each test.

The results of the tests are shown in Figure 2.4. It is clear from the data that altering the probe toehold sequence has the predicted effect on the reaction rate. For instance, strengthening the probe toehold relative to the target toehold as in the *Downhill* test causes the reaction to proceed more quickly than if the probe toehold was weakened relative to the target toehold (*Uphill* test). The bimolecular rate constants used to fit the data are shown in Table 2.3, obtained using the best-fit value of $k_{HT} = 5\times10^5$ $M^{-1}s^{-1}$. The bimolecular model predicted the toehold exchange rate of the *Uphill* test well, but the fits do not match the data for the other tests as closely. A possible explanation for this is that there is an error associated with each value of the free energy used to generate the rate constants. Nupack [Zadeh et al., 2011] predicts the free energy of a DNA sequence based on experimentally-obtained thermodynamic parameters, such as the effect on stability of neighbouring base pairs and salt concentration. The error in the thermodynamic prediction of free energy is 4 % [Koehler and Peyret, 2005], and the value of ΔG^o is most accurate at temperatures near 50 °C and gets worse as the temperature deviates from this [SantaLucia-Jr and Hicks, 2004]. An error of 4 % in each of the toehold en-

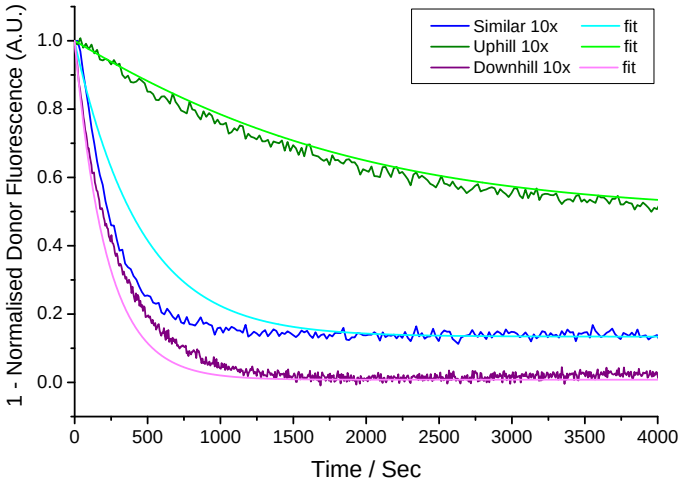
ergies could account for the difference between the experimentally observed reaction rates and those predicted by the bimolecular model.

The experiments discussed here and in Section 2.2.1 demonstrate that it is possible for the rate of the toehold exchange reaction to be altered by strengthening one of the toehold domains, either by changing the sequence or length of the domain. By altering the sequence alone of a 6 nt toehold domain, the forward rate constant of the toehold exchange reaction can be varied by a factor of 10. However, it should also be noted that the concentration of the probe strands plays an important role in how far the reaction goes towards completion. Strengthening the probe toehold domain and adding a 10-fold excess of the probe strand allowed the *Downhill* reaction to reach completion in around 30 min, but the same excess of strands only resulted in an 80 % complete *Similar* reaction. Therefore, to ensure the yield of the cascade-coupling mechanism is high, it is necessary to maximise the yield of the toehold exchange reaction by controlling the probe strand concentration.

The cascade-coupling mechanism is designed to bring together the reacting groups of one particular chemical reaction and therefore, it is desirable for each step in the cascade to proceed at approximately the same rate. This allows each chemical building block the same amount of time to react. There are two methods to achieve this. Firstly, each toehold domain in the cascade could be identical, which is a simple way to ensure the reaction rates are the same for each step. However, the cascade-coupling chemistry scheme may require the building blocks to react in a particular order. If all the toehold domains are the same, then each target strand is identical except for the building block it is attached to, and it would be difficult to assemble the system in the required configuration. The second method is to use target strands which all have the same universal domain, but differ in their toehold domains. The results of the *Similar* test show that it is possible for the toehold domains to be different in sequence but



(a) The 1× excess tests.



(b) The 10× excess tests.

Figure 2.4: The effect of altering the probe toehold sequence on the toehold exchange reaction. (a) The *Similar* test shows that when there is a small change in free energy between the initial and final state ($\Delta G^o_{net} = 0.3$ kcal/mol), strand displacement takes place rapidly. In the *Uphill* test, the probe toehold energy is lower than the target toehold energy, such that the bimolecular model predicts the forward rate constant to be slower than the reverse rate constant. The rate of the toehold exchange in this case is much slower than any of the other three cases. Similarly, when the probe toehold energy is higher than the target toehold energy (*Downhill*), the toehold exchange reaction is fast.

(b) Each reaction is driven further towards completion by adding a large excess of probe strands (Section 2.1.1).

The reactions were carried out as described in Figure 2.3, except the donor fluorescence here is Fam and was monitored using Ex490/Em520 nm.

still have similar free energies. By designing toehold domains in this way, the reaction rate for each step may be approximately the same, but the hybridisation of the target strands to the template is also controlled. These two methods are explored in the following section, which describes two-step cascade experiments.

2.3 Two-Step Reaction

It is not yet clear whether the toehold exchange process can be used to complete multiple and sequential strand displacements. To determine this, a two-step system was designed which would also help to ascertain whether the reaction is affected by using either identical or different toehold domain sequences. The initial state of the two-step system is a duplex between two target strands $\langle \mathtt{u\ t} \rangle$, a template strand and a reporter strand. The toehold domains are all 6 nt in length, including the exposed toehold on the 3' end of the template. The scheme using identical toehold domains is shown in Figure 2.5, and the scheme using different toehold domains is shown in Figure 2.6. The sequences of the universal domain and the reporter strand are the same in both systems. The one-step experiments had shown that minimising the secondary structure within the target and probe strands allowed the bimolecular model to be applied to the system (Section 2.2.2). Strands for the two-step systems were first chosen by applying the secondary structure design criteria and then selecting toeholds sequences with approximately the same free energy. As a result, the toehold domain free energies in the *Different* test are not closely matched, and there is a small bias towards the target strand at each step. This was not intended, but was a result of prioritising the secondary structure design criteria. (All DNA sequences are provided in Appendix A.4.) In order to monitor the two-step reaction, the reporter strand and the second target strand are labelled with fluorophores forming a FRET pair (Appendix A.1.1). In the *Different* scheme, the difference in the toehold domains enables the fluorescently-labelled target strand to hybridise to the correct

position on the template. However, in the *Identical* scheme, this is not possible, and so the fluorescent target strand hybridises to both positions on the template. This does not pose a problem for the fluorescence spectroscopy experiments, since the fluorophore attached to target strand 1 is too far away from the reporter strand to affect the FRET between target strand 2 and the reporter. In both schemes, a change in the emission signal of the Fam fluorophore should only be observed when the second target strand is displaced.

Scheme	**Step**	ΔG^o_{probe} **(kcal/mol)**	ΔG^o_{target} **(kcal/mol)**	$\mathbf{k}_{zf}$ $(\mathbf{M^{-1}s^{-1}})$	$\mathbf{k}_{zr}$ $(\mathbf{M^{-1}s^{-1}})$
Identical Targets	1	-10.21	-10.21	2.5×10^5	2.5×10^5
	2	-10.21	-10.21	2.5×10^5	2.5×10^5
Different Targets	1	-9.4	-9.92	1.47×10^5	3.53×10^5
	2	-9.92	-10.21	1.9×10^5	3.1×10^5

Table 2.4: The standard free energies of the probe and target strand toehold domains in the two varieties of the two-step cascade reactions at 21 °C. In the *Identical* scheme, the toehold domain sequences are identical. In the *Different* scheme, minimal secondary structure in the strands was prioritised over toehold domain free energy similarity (see main text), causing a slight unintended bias towards the target strand at each step. The bimolecular rate constants are calculated using the parameters in the table, $k_{HT} = 5 \times 10^5$ $M^{-1}s^{-1}$ as before (since the toehold size and strength matches those used in Section 2.2.2) and the model outlined in Section 2.1.2.

The strands required to form the initial states of the two schemes were mixed together in stoichiometric quantities in NEB2 buffer (50 mM NaCl, 10 mM Tris-HCl EDTA, 10 mM $MgCl_2$ pH 8.0). Samples were annealed by heating the solution to 85 °C to melt all possible structures, and cooling to 20 °C over 10 min. For the fluorescence spectroscopy experiments, sample concentrations were 6.7 nM in a 1.5 ml total volume, and the temperature of the fluorimeter was set to 21 °C. The single

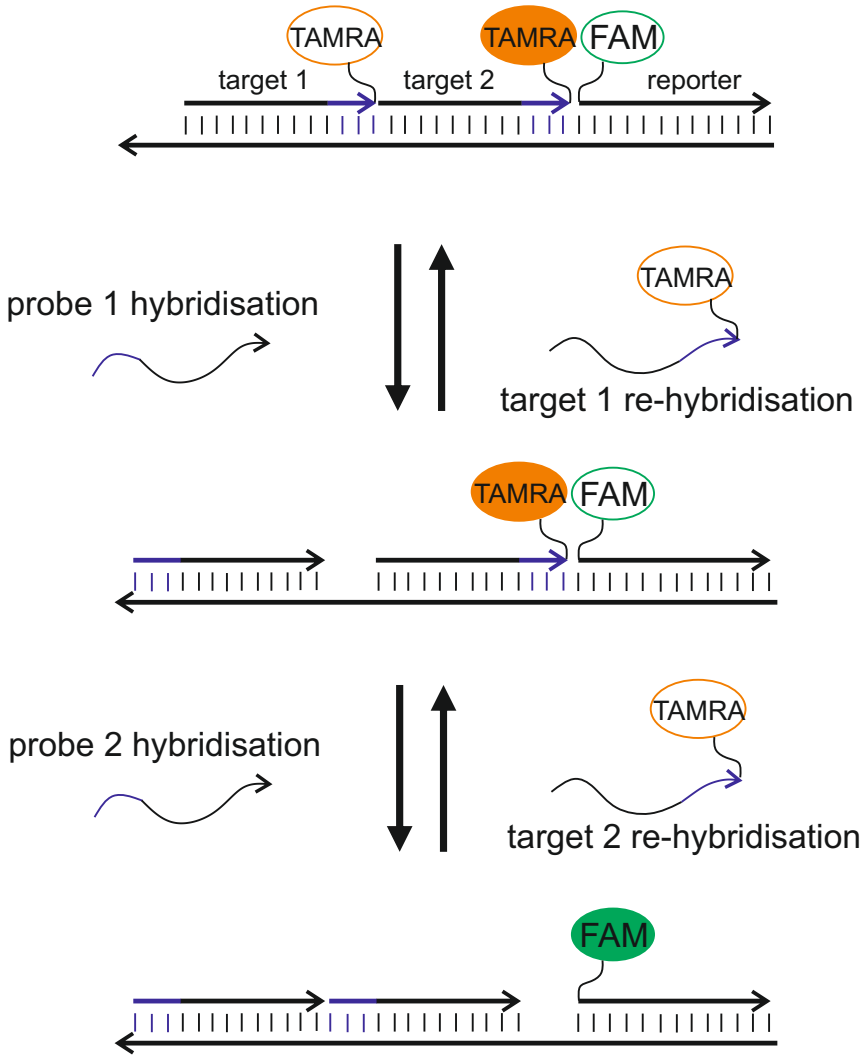


Figure 2.5: The two-step *Identical* scheme.

The toehold domains on the template strand are identical, so the same target strand hybridises to both positions along the template. The target strand is labelled with a Tamra fluorophore, which forms a FRET pair with the Fam fluorophore attached to the reporter strand. The distance between target strand 1 and the reporter strand is too large to enable FRET between the flurophores, so a significant change in the Fam fluorescence is only observed when the second target strand is displaced. Compare with Figure 2.6, in which the toehold domains are all different.

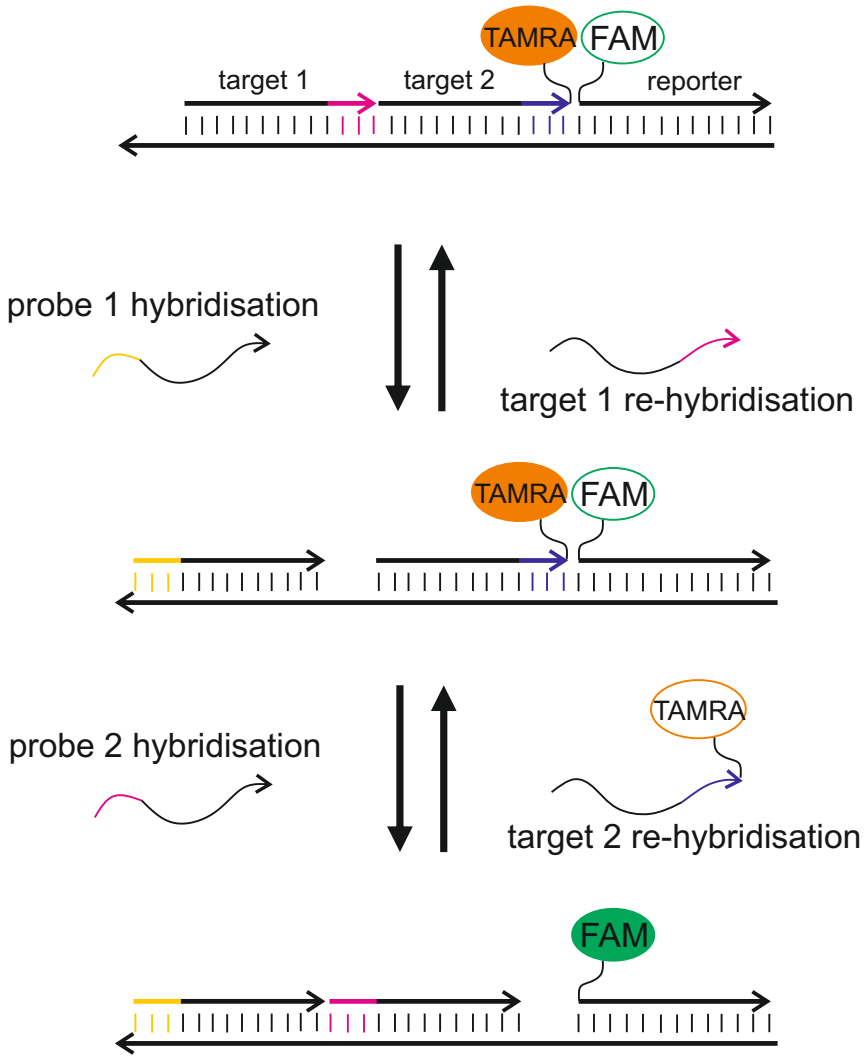


Figure 2.6: The two-step *Different* scheme.

The toehold domains on the template strand are all different, so target strands hybridise to specific positions along the template. The target strand is labelled with a Tamra fluorophore, which forms a FRET pair with the Fam fluorophore attached to the reporter strand. Displacement of the second target strand will cause a change in the donor (Fam) emission fluorescence.

probe strand was added to the *Identical* sample in either a 10- or 100-fold excess. The results are shown in Figure 2.7, and indicate that it is possible to perform two consecutive toehold exchange reactions. The reactions are quick, and are pushed further towards completion by a larger excess of the probe strand. However, the bimolecular rate constants do not fit the experimental data well, particularly in the 10-fold excess case. The system reaches the same equilibrium value as the model predicts, but the rate of the reaction does not match the prediction. The 10-fold excess reaction proceeds far quicker than the model suggests, which may partially be the result of an error in calculating the toehold free energies [Koehler and Peyret, 2005]. Another reason for the disparity between the data and the model is that the model does not take into account the fact that the toeholds are identical. This means that the number of probe strands able to complete each step is doubled, such that a $1\times$ excess of probe strands is really a $2\times$ excess. However, it may also be because one probe strand can complete both steps. It is not clear from these results whether the toehold exchange reactions are occurring sequentially as drawn in Figure 2.5. It may be possible for the probe strand to only displace the second target strand, or to simultaneously displace the first and second targets. This is only possible if the reaction is a non-toehold–mediated strand displacement, since if the probe strand has not completed the first step, there is no free toehold for it to use to complete the second step. Either way, this is an undesirable property of the synthetic ribosome, which should "read" the toehold domains (codons) in sequential order to closely match the action of the natural ribosome.

To test whether non-toehold–mediated strand displacements are possible, a universal displacer was added to the initial duplex sample in a 10-fold excess. The universal displacer is 18 nt and only complementary to the universal domain on the template. If it is able to displace the target strands, it does so without the use of a toehold. Figure 2.7 shows the change in fluorescence as a result of only adding the

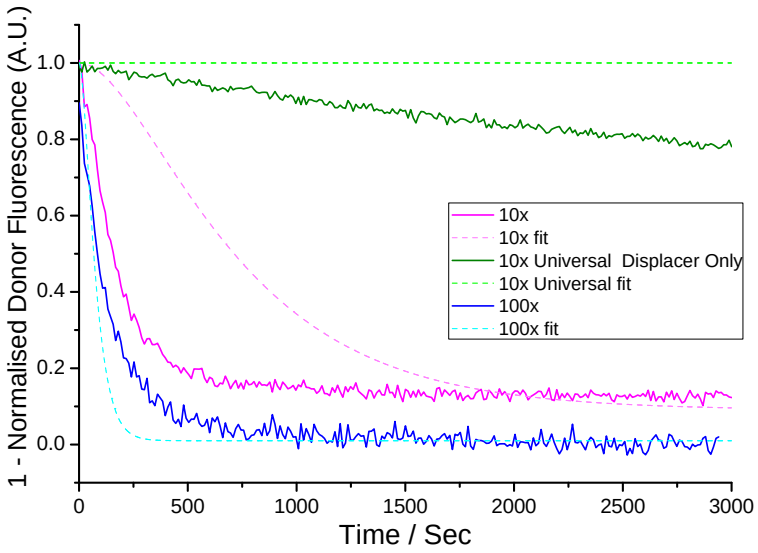


Figure 2.7: The results of the *Identical* two-step reaction.

The probe strand was added to the samples in either a 10-, or 100-fold excess. The data suggests that two consecutive toehold exchange reactions can be performed, and the reactions are high yielding. The two-step reaction was also carried out using a universal displacer, to determine whether non-toehold–mediated displacements can occur. The universal displacer was added in a 10-fold excess, and the data suggests that it is possible for the strand to displace the second target strand. However, the rate of the reaction does not match the model prediction. Without a toehold to assist the strand displacement it should be difficult for the reaction to begin, and therefore it can be assumed from the data that the universal displacer uses some unknown property of the duplex to initiate the displacement. If for instance, the target strands are frayed from the template, the universal displacer could hybridise to the few exposed toeholds and begin the displacement reaction. This could explain why the 10-fold excess reaction occurs faster than predicted by the model, as non-toehold–mediated displacements by the universal domain of the probe strand could contribute to the rate of the reaction. This mechanism may allow the two target strands to be displaced simultaneously.

The reactions were carried out as described in Figure 2.3, except the donor fluorescence here is Fam and was monitored using Ex490/Em520 nm.

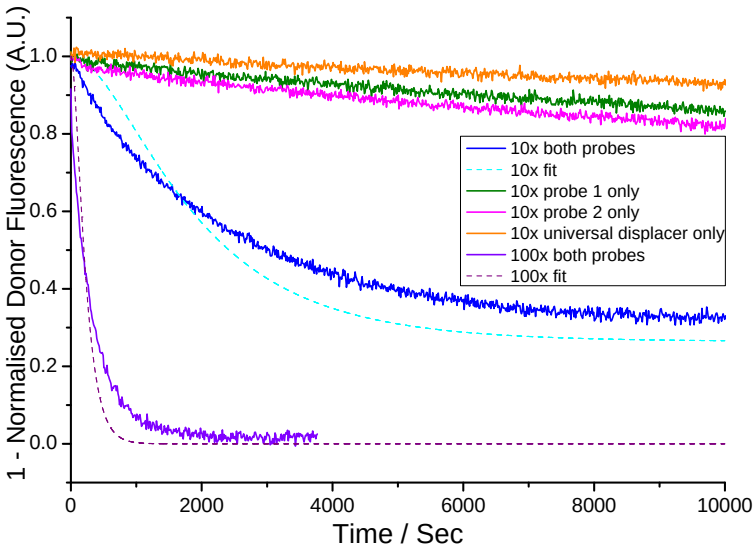


Figure 2.8: The results of the *Different* two-step reaction.

The two probe strands were added to the samples in either a 10-, or 100-fold excess. The data suggests that two consecutive toehold exchange reactions can be performed. The two-step reaction was also carried out using the probe strands individually or with a universal displacer, to determine whether non-toehold–mediated displacements can occur. The strands were added in a 10-fold excess, and the data suggests that it is possible for one of the probe strands alone or the universal displacer to displace the second target strand. However, the rate of the reaction is slow, as without a toehold to assist the strand displacement it is difficult for the reactions to begin. Thus, it can be assumed that the two-step reactions are quicker because they make use of the migrating toehold that arises if the steps are performed sequentially.

The reactions were carried out as described in Figure 2.3, except the donor fluorescence here is Fam and was monitored using Ex490/Em520 nm.

universal displacer to the sample. As expected, the universal displacer reaction rate is slow, since it is difficult for the reaction to begin without the aid of a toehold. By setting the probe toehold free energy to zero for each step, the bimolecular model (Section 2.1.2) predicts $k_{zf} \approx 10^{-2}$ $M^{-1}s^{-1}$ and $k_{zr} \approx 10^{5}$ $M^{-1}s^{-1}$, but these rate constants do not fit the data at all. The model predicts that non-toehold–mediated displacements are unlikely on the timescale of the experiment, which disagrees with the experimental evidence. This "leak" reaction seems to arise because it is possible to displace the target strands without using the toehold domain. If the base pairs of the 5' end of the target strands are not strong (more A–T pairs than G–C), it is possible that the initial duplex frays, revealing nucleotides in the universal domain of the template that can be used to initiate strand displacement. It is not easy to determine whether this is taking place, and to model the "leak" reaction. However, it is clear from these experiments that non-toehold–mediated displacements can occur, and that these "leak" reactions may contribute to the correct reaction rates, which prevents the system from being modelled accurately.

The samples for the *Different* test were prepared and the reactions carried out in the same way. The two probe strands were added in either a 10- or 100-fold excess. Three additional experiments were carried out to check that the reaction is occurring sequentially: probe 1, probe 2 and the universal displacer were added alone to reactions. The results are shown in Figure 2.8. Firstly, it is clear from comparing the data with that of the *Identical* scheme that the two-step reaction rate is much slower when different toehold domains are used than when they are identical. Table 2.4 shows that the target strands are favoured over the probe strands in both steps because the target toehold free energy is larger. As a result of a higher reverse rate constant for each step, the overall rate for the two-step reaction is slow. The yield of the reactions are also well predicted by the bimolecular model. The data suggests that in a multi-step cascade using different toehold sequences, the probe strand concentration

must be large in order to ensure the yield of the toehold exchange process is high. Finally, adding only one of the two displacers, or adding only a universal displacer, produces "leak" reactions as before. Neither of the probe strands alone should be able to displace the second fluorescent target strand. It appears that the same mechanism used by the universal displacer strand is being used in the reactions involving only one probe strand. The rate of the "leak" reaction caused by probe 2 is the fastest, presumably because it is the probe that is designed to displace target 2. It is still not clear from these experiments what enables the "leak" reactions to take place.

The results of this section indicate that the toehold exchange mechanism can be used to complete consecutive strand displacements, which is a basic requirement of the synthetic ribosome. The results of the reactions using the *Different* scheme (with three unique toehold domains), fit reasonably well to the predicted reactions generated by the bimolecular model, compared with the *Identical* scheme results which are not as easily described by the model. For this reason, and because unique toeholds allow target strands to hybridise to the template in a specific order (which may be necessary for the cascade-coupling chemistry), future toehold exchange experiments will use unique toehold domains. However, the results also show that non-toehold–mediated displacements are possible. This "leak" reaction is slow in comparison with the correct reactions, but theoretically, it should not be possible for the final target strand to be displaced without the aid of a toehold on the timescale of the experiment. This has implications for the cascade-coupling chemistry. If it is possible for a target strand to be displaced without the aid of a toehold, then non-sequential displacements can occur in the cascade. These may cause the cascade-coupling mechanism to produce truncated products, since particular target strands (and their building blocks) may have been displaced before the toehold exchange mechanism had reached them. The reason for the non-sequential displacements needs to be understood and accounted for in order for the cascade-coupling mechanism to be used in the design for the synthetic

ribosome, as one of the main properties of the ribosome is its sequential reading of the mRNA template. Therefore, the "leak" reaction was investigated further by extending the cascade to six-steps, as discussed in Section 2.4.

2.4 Six-Step Cascade

The longest custom oligonucleotide sequences that can be synthesised by Integrated DNA Technologies are 200 nt ultramers. This allowed a six-step system to be designed. The six target strands are all 24 nt in length, and consist of an 18 nt universal domain and a 6 nt unique toehold domain (which resemble the codons in mRNA). The initial state of the system is a duplex between the long template strand, the reporter strand and the six target strands. There is also a free 12 nt toehold on the 3' end of the template (Figure 2.9). The toeholds were designed to offer little or no thermodynamic advantage to the probe strands, so the reaction is driven by the high probe concentration (see Table 2.5 and Section 2.2 for the DNA strand design criteria). An exception is the initial toehold used by probe P0, though the probe strand can be shortened, if necessary, so that it hybridises to a fraction of the toehold. This toehold was designed longer than the rest so that it can be determined whether the energy acquired from hybridisation to the first toehold has a significant effect on the overall rate of the six-step reaction. All toehold sequences were first selected by choosing those sequences from a set with approximately similar free energy, and then picking out the sequences which meet the specificity design criteria. The minimal secondary structure criteria was also applied, but was not a priority. For the first attempt at designing a six-step cascade, toehold sequence specificity and free energy was considered more important than minimising secondary structure.

The natural ribosome is able to read long sequences with little error and it is hoped that the cascade-coupling synthetic ribosome mechanism can operate substan-

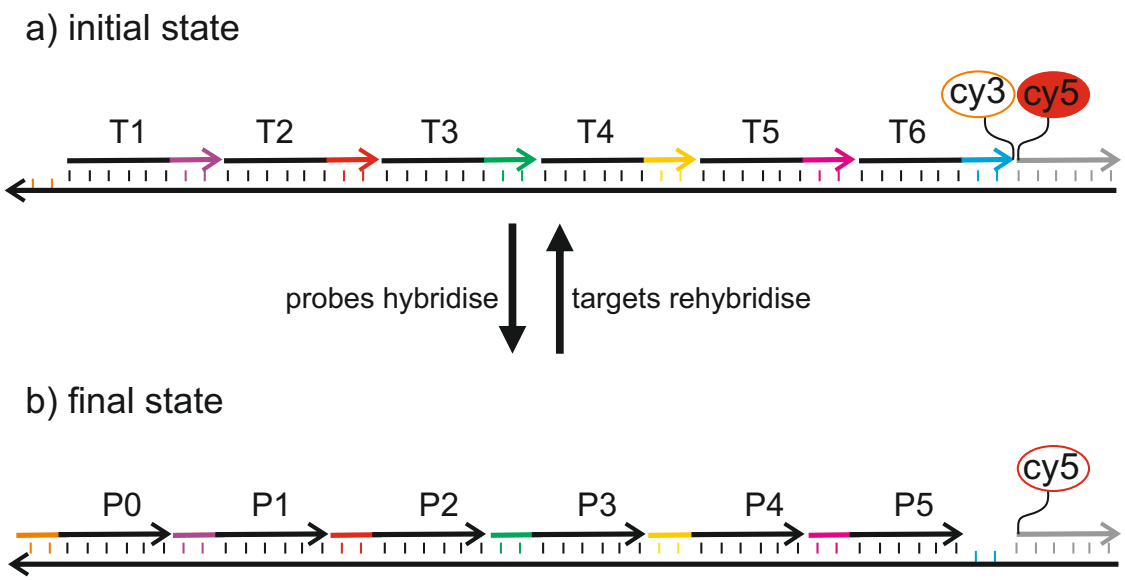


Figure 2.9: Schematic of the six-step cascade.

Each target strand consists of two domains: competition (18 nt, shown in black) and a unique toehold (6 nt). The first probe strand P0 can hybridise to some or all of the 12 nt free toehold, and once hybridised, competes with target strand T1 over the competition domain. If successful, P0 hybridises to the entire 18 nt competition domain, while leaving the target strand hybridised to its toehold only. The target strand spontaneously dissociates from the template, revealing a 6 nt toehold for the next probe strand. The migrating toehold enables the completion of the six-steps of the cascade.

The final target strand T6 is labelled with a Cy3 fluorophore, which forms a FRET pair with the Cy5 fluorophore attached to the reporter strand (shown in grey). The Cy3 (donor) emission is monitored to determine when the sixth step of the cascade reaction has been completed. Upon displacement of strand T6, the increased distance between the fluorophores results in increased Cy3 emission.

Step	ΔG^o_{probe} **(kcal/mol)**	ΔG^o_{target} **(kcal/mol)**	$\mathbf{k}_{zf}$ **($M^{-1}s^{-1}$)**	$\mathbf{k}_{zr}$ **($M^{-1}s^{-1}$)**
1	-14.16	-9.2	5.0×10^5	1.15×10^2
2	-9.2	-8.93	3.05×10^5	1.93×10^5
3	-8.93	-8.95	2.45×10^5	2.53×10^5
4	-8.95	-9.21	1.95×10^5	3.03×10^5
5	-9.21	-9.11	2.7×10^5	2.28×10^5
6	-9.11	-8.97	2.78×10^5	2.20×10^5

Table 2.5: The standard free energies of the probe and target toehold domains at each step in the six-step cascade at 21 °C. The initial toehold is 12 nt though the first probe strand need not hybridise to the full length, while the rest are 6 nt and are approximately matched in terms of free energy. In order to achieve multiple toehold domains with different (and specific) sequences, but similar free energies, the secondary structure design criteria was largely ignored (see main text). The bimolecular rate constants are generated using the parameters in the table, $k_{HT} = 5.5\times10^5$ $M^{-1}s^{-1}$ (since the toehold energies and the universal domain sequence match those of Section 2.2.1), and the model described in Section 2.1.2 and Appendix A.2.

tially error-free also. If one of the probe strands required for the reaction is not added, the cascade should be prevented from progressing past the point at which the missing probe is needed. The results of the two-step experiments in Section 2.3 showed non-toehold–mediated strand displacements were possible, but it is not clear whether these "leak" reactions will be observed in a six-step system.

Samples were made by annealing stoichiometric quantities of the six target strands, the reporter and the template strands in NEB2 buffer (50 mM NaCl, 10 mM Tris-HCl EDTA, 10 mM $MgCl_2$, pH 8.0). Samples were annealed by heating the solution to 85 °C to melt all possible structures, and cooling to 20 °C over 10 min. For the fluorescence spectroscopy experiments, sample concentrations were 6.7 nM in a 1.5 ml

total volume. The experiments were conducted at 21 °C, and both the Cy3 and Cy5 emission signals were recorded (though only the Cy3 donor emission is presented in the graphs). Reaction samples were annealed *in situ* at the end of the experiment and a fluorescence baseline recorded.

A number of tests were performed on this six-step system. All six probe strands were added to the sample, and to check whether the six-step system is prone to "leak" reactions, each of the probe strands were in turn left out of a reaction. The results are shown in Figure 2.10. It is clear from the data that the toehold exchange mechanism can be extended to multiple-step cascade reactions. This is a promising result for the cascade-coupling chemistry, as multiple building blocks can potentially be coupled together to form new functional molecules. The yield of the reaction is only 60 %, which means a larger excess of probe strands must be added to drive the reaction further towards completion. As stated earlier, since the system was designed without considering that the single-stranded species may form secondary structures, the reaction was difficult to model. The bimolecular model (described in Section 2.1.2 and by the ODEs in Appendix A.2.1) generated a reaction that was orders of magnitudes too quick, suggesting that the melting of secondary structures within strands is affecting the rate of the six-step cascade reaction. Analysis of the probe and target strands using mfold confirmed that secondary structures are formed within the single-stranded species.

Figure 2.10 also shows that "leak" reactions occur when particular probe strands are omitted from the reaction. This is an expected outcome of the reaction, given the results of the two-step experiments in Section 2.3. Each probe strand was left out of the reaction in turn, and when either probe P1 or P5 was omitted, the cascade reaction was prevented from continuing to the final step and consequently, the fluorescence signal did not change. Reactions without P1, the second probe strand,

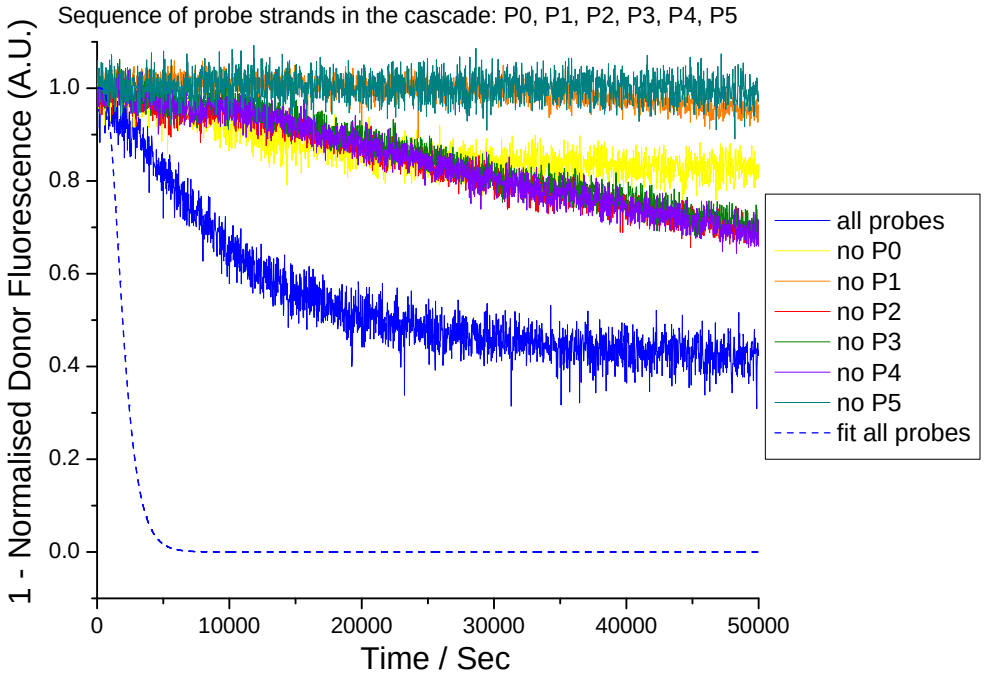


Figure 2.10: Data from the six-step cascade reaction.

Adding all six probe strands to the reaction results in a fluorescence signal change, indicating that it is possible to use toehold exchange to complete multiple, sequential strand displacements. However, the bimolecular model does not fit the data well, and appears to be at least an order of magnitude too fast. This may be explained by the fact the single-stranded species in the reaction can form relatively stable secondary structures, which can interfere with the overall reaction rate of the cascade.

"Leak" reactions are observed when probe strands are omitted from the reaction. This is expected, since the two-step experiments also produced "leaks" when a displacer was missed out. However, it is interesting that missing out either P1 or P5 does not result in a fluorescence signal change. It is not clear whether this is trelated to the strands themselves or to the position along the cascade at which the probe strand is missed out. This is investigated in Section 2.4.1.

The reactions were carried out as described in Figure 2.3.

or without the final probe strand P5, both prevent the fluorescent final target strand from being displaced. It is not immediately clear why each of these particular probe strands do not produce a "leak" reaction. One hypothesis is that "leak" reactions are in some way linked to the position along the cascade at which the missing probe occurs. This is investigated further in Section 2.4.1.

The length of the initial toehold is 12 nt, but it is not necessary for the first probe strand P0 to hybridise to the full length of this toehold. Two shorter probe strands were used in the two-step reactions to initiate the reaction. Figure 2.11 shows that the cascade reaction can be initiated by a 9 nt or 6 nt toehold. In each case, a 10-fold excess of all six probe strands was added to the reaction. It is clear from the data that a longer initial toehold helps to drive the reaction, since the rate of the reaction is not only quicker with a longer toehold, but the yield of the reaction is also greater. A long initial toehold means the first toehold exchange reaction is quick, and with a 10-fold excess of the probe strand, is also high-yielding. This has a knock-on effect on the remaining steps of the reaction.

The results so far suggest that a migrating toehold can enable multiple toehold exchange reactions to take place in the designed sequence, since in the absence of particular probe strands, the cascade reaction is prevented from continuing. Missing out a probe strand from the reaction generally results in a "leak" reaction, but the omission of either probe strand P1 or P5 does not produce a "leak". The "leak" reaction could be a position effect, such that the point at which the missing probe strand is encountered in the cascade may be the cause. The probe strands which do not produce "leaks" correspond to the second step and the final (sixth) step in the cascade, and if "leak" reactions are linked to position, then it is possible that missing out the second or the sixth probe strand will always produce the same result. This is tested by shuffling the order of the target strands along the template (Section 2.4.1).

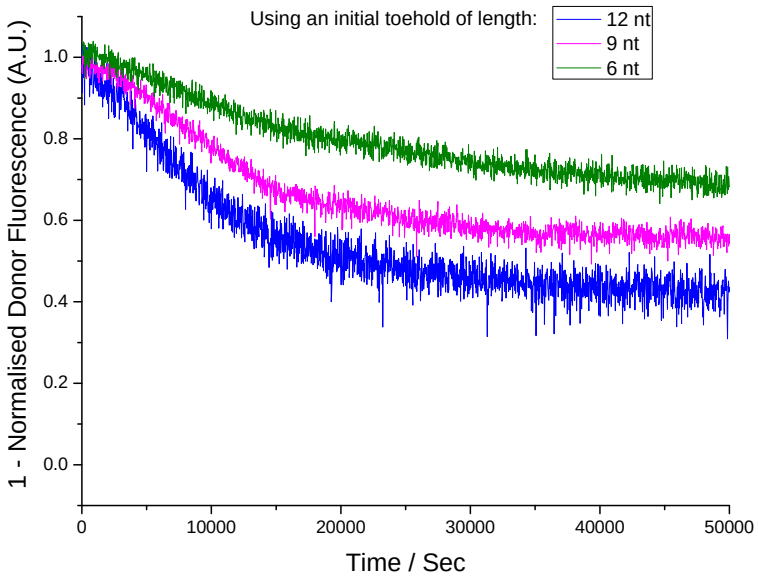


Figure 2.11: Effect of changing the length of the initial toehold on the cascade reaction.
A longer initial toehold means the first probe strand has a larger thermodynamic advantage in the toehold exchange reaction. Thus, the first step is completed quickly, and with high yield, affecting the remaining steps of the cascade.
The reactions were carried out as described in Figure 2.3.

2.4.1 The Shuffled Cascade

The shuffled cascade is constructed using the same target strands as in Figure 2.9, but a new template strand. The toehold domains on the new template strand have been shuffled, but the toehold sequences themselves have not been changed in any way. This is to determine whether the "leak" reactions are related to the position at which a missing probe is encountered. Figure 2.12 shows how the target strands are arranged on the template. In the original system, missing out the second (P1) or the sixth (P5) probe strand did not result in a "leak" reaction. If "leaks" are position-dependent, missing out the second (P6 now) or sixth (P2 now) probe strand in the shuffled cascade experiments will produce the same effect (no "leak"). However, if the strand sequences themselves are the cause, then no "leak" reaction will be observed when strand P1 is omitted from the reaction using the shuffled template. Target and probe strands are labelled according to their positions in the original cascade. The

template still has the same 12 nt toehold domain, and the final target (in this case T5 not T6) and the reporter strand are labelled with fluorophores. The toehold free energies for each step are listed in Table 2.6.

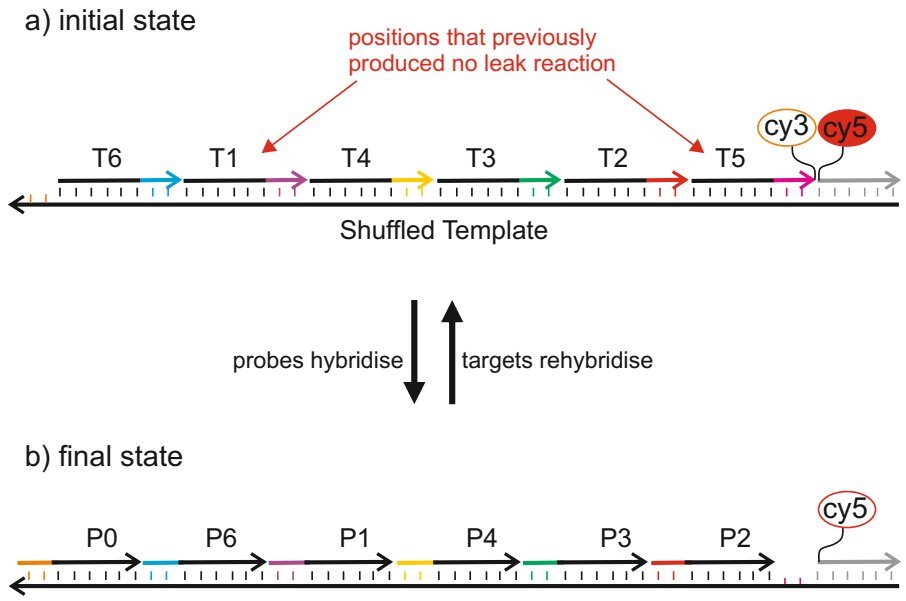


Figure 2.12: Schematic of the shuffled six-step cascade.
The toehold domains have been shuffled and the target strands now hybridise to the template at different locations (c.f. Figure 2.9). The diagram also highlights the positions along the template at which no "leak" reaction had previously been observed. In the original cascade experiments, missing out the second probe strand (P1) or missing out the final probe strand (P5) produced no "leak" reaction. If the "leak" reaction is position-dependent, then missing out probes P6 or P2 in the shuffled cascade would not result in a "leak".

Samples were prepared by annealing stoichiometric quantities of the six target strands, the reporter and the template strands in NEB2 buffer. The experiments were conducted at 21 °C. The shuffled cascade was first tested by adding a 10-fold excess of the probe strands to the reaction sample. The data, shown in Figure 2.13, indicates that the reaction rate is similar to that of the original cascade system. This is as expected, since the energies of the toehold domains at each step remain balanced. Probe strands were also omitted in turn from reactions. "Leak" reactions were ob-

Step	**Strands Involved**	ΔG^o_{probe} **(kcal/mol)**	ΔG^o_{target} **(kcal/mol)**	$\mathbf{k}_{zf}$ $(\mathbf{M^{-1}s^{-1}})$	$\mathbf{k}_{zr}$ $(\mathbf{M^{-1}s^{-1}})$
1	P0, T6	-14.16	-8.97	5.00×10^5	7.79×10^1
2	P6, T1	-8.97	-9.20	2.01×10^5	2.97×10^5
3	P1, T4	-9.20	-9.21	2.47×10^5	2.51×10^5
4	P4, T3	-9.21	-8.95	3.03×10^5	1.95×10^5
5	P3, T2	-8.95	-8.93	2.53×10^5	2.45×10^5
6	P2, T5	-8.93	-9.11	2.11×10^5	2.87×10^5

Table 2.6: The standard free energies of the probe and target toehold domains at each step in the shuffled six-step cascade at 21 °C. The initial toehold is 12 nt though the first probe strand need not hybridise to the full length, while the rest are 6 nt and are approximately matched in terms of free energy. Compare with Table 2.5.

served, but more importantly, no "leak" reaction was observed when either probe P1 or probe P2 was missed out. Figure 2.13 shows that when probe strand P2 (the final probe strand) and probe P1 (the third probe strand to be added) were individually missed out of reactions, no change in fluorescence was measured. Therefore, missing out the **final probe strand** has the same effect in each cascade system, and missing out **probe P1** also produces no "leak" reaction, independent of which step P1 is needed at.

It is possible that fluorophore stacking interactions between the fluorophores attached to the final target strand and the reporter strand provide the target strand with extra stability [Moreira et al., 2005]. This additional stability is approximately equivalent to one base pair. The *Uphill* test in Section 2.2.2 showed that even when the target toehold energy was higher than the probe toehold energy, a toehold exchange reaction leading towards the equilibrium distribution can still take place.

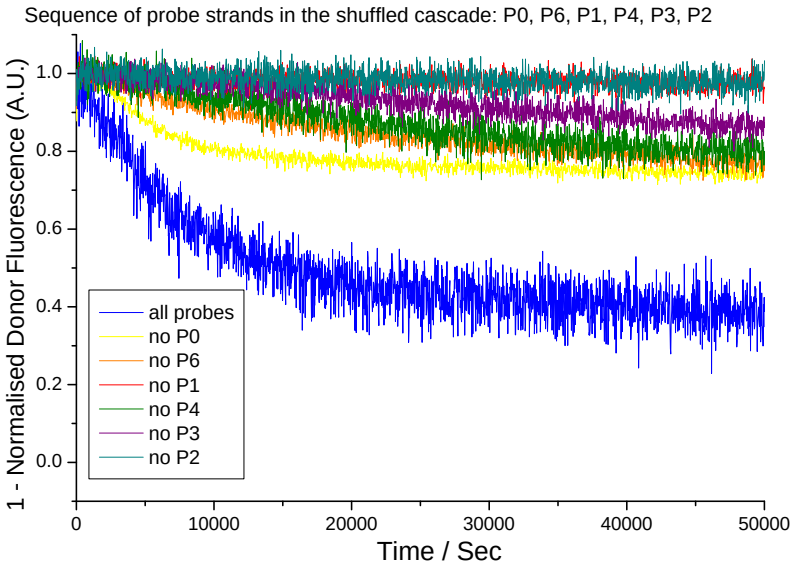


Figure 2.13: The shuffled cascade.

Adding all six probe strands to the shuffled cascade reaction produces an almost identical reaction rate to the original cascade. This is as expected, as the toehold domains are still balanced at each step, and the system is still practically thermodynamically neutral. The shuffled cascade was also tested by missing out one probe strand at a time from the reaction. The omission of most of the probe strands resulted in a "leak" reaction but as in the original cascade (c.f. Figure 2.10), there are two notable exceptions. Here, missing out either strand P1 or strand P2 (the final probe strand) prevents the cascade reaction from completing the final step. This suggests two things. Firstly, the effect of missing out probe P1 in both cascade systems is the same regardless of its position along the template, and therefore, there must be something unique to the P1 probe which prevents a "leak" in its absence from a reaction. Secondly, missing out the probe strand which displaces the final target strand in each system (P2 here and P5 in the original cascade) also prevents a "leak" reaction being observed. This implies that the correct displacer is necessary to displace the fluorescent final target strand, but it is not yet clear why this is the case. Hypotheses are described in Section 2.4.2.

The reactions were carried out as described in Figure 2.3.

Fluorophore stacking interactions alone are therefore not sufficient to explain why the final probe strand is essential for the cascade reaction.

The effect of missing out probe P1 is the same in both cascades. Shuffling the toehold domains on the template has shown that it is not the position of the P1 toehold domain (or the P1 probe) that is special, but the sequence itself. This is discussed in more detail in the following section.

2.4.2 Hypotheses for "Leak" Reaction

The "leak" reactions that have been observed in the two-step and six-step cascades are not well understood. In the two-step reactions (Section 2.3), a universal displacer was able to displace the fluorescent final target strand, which implies that non-toehold–mediated reactions can take place. A number of possible explanations for this are described below.

Toehold specificity

The toeholds were chosen from the 4096 possible 6 nt sequences by selecting those that have similar free energies and do not have strong pairwise interactions (with other sequences in the set and their complements). However, in requiring that the free energies of the toehold domains are similar, the short sequences are constrained to contain approximately the same number of G/C and A/T nucleotides. For instance, the toehold domain of target T6 is TTGCGC, while that of target T2 is AACGCG (a full list of the strands used in the cascade experiments is provided in Appendix A.4). These two toeholds have the subsequence CGC in common, which may mean the toeholds are not very specific, and strand displacements can be initiated by incorrect probe strands. Three-base-pair duplexes are unstable however, so it is unlikely that this specificity problem alone explains why "leak" reactions occur. If however, the

5' ends of the target strands are frayed, then the wrong probe strand may be able to use partial toehold binding and the exposed nucleotides of the universal domain to start the strand displacement. To test this, toehold specificity checks were conducted, in which the ability of each probe to bind to a particular toehold domain and initiate a displacement was tested. The test system resembles that of the one-step experiments (Figure 2.2). Six template strands were used in each of which the probe toehold domain was each of the six toehold sequences. Each of the six test systems was tested using each of the six probe strands in turn, to determine whether any of the probe strands can cause a strand displacement that it was not designed to carry out. The test systems were assembled in the usual way, and the experiments carried out under the same experimental conditions as the previous cascade reactions. One set of data from the experiments is presented in Figure 2.14, where the probe toehold domain is the T3/P3 toehold sequence. The results show that only probe P3 is able to participate in a toehold exchange reaction on the timescale of a one-step reaction. Each of the six tests showed that only the probe strand with a toehold domain that is fully complementary to the toehold domain on the template was able to carry out the strand displacement. Despite the similarity between some of the toehold sequences, incorrect probes did not manage to complete a step.

Targets behaving as displacers

The toehold specificity tests showed that incorrect probes are not able to carry out a one-step toehold-mediated strand displacement, but it may be possible for displaced target strands that are free in solution to displace other target strands. Figure 2.15 illustrates this hypothesis. If a target strand has been displaced but the probe strand required to continue the reaction is missing, the freed target strand is able to re-hybridise to the free 6 nt toehold. Generally, this would reverse the toehold exchange reaction, but it may be possible for the target strand to roll over on itself

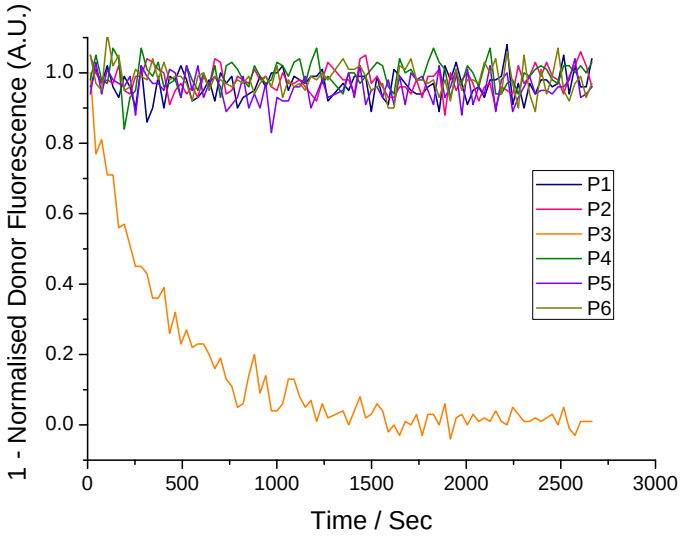


Figure 2.14: Data from one of the toehold specificity tests.

The ability of each probe to bind to each toehold to initiate a strand displacement reaction was tested using a one-step reaction similar to that shown in Figure 2.2. It was found that the the probes were unable to initiate a displacement unless the probe strand toehold domain matched the probe toehold domain on the template strand. The results shown here are from the test of each probe against the P3 template strand toehold domain. As expected, only probe P3 produced a fluorescence signal change corresponding to the strand displacement, on the timescale of a one-step experiment. These experiments were also carried out for 50000 sec: no 'incorrect' reactions were observed on this longer time scale (data not shown). Calculations of the expected yield of the toehold exchange reaction, assuming that wrong probes can hybridise to half of an incorrect toehold domain (such that $\Delta G^o_{net} > 0$) and using Equation 2.3, suggest the system is already near equilibrium (that is, the toehold exchange reaction is unlikely). Therefore, even if the toehold domains have subsequences in common, the specificity of the probe strands alone is not enough to enable "leak" reactions.

The reactions were carried out as described in Figure 2.3.

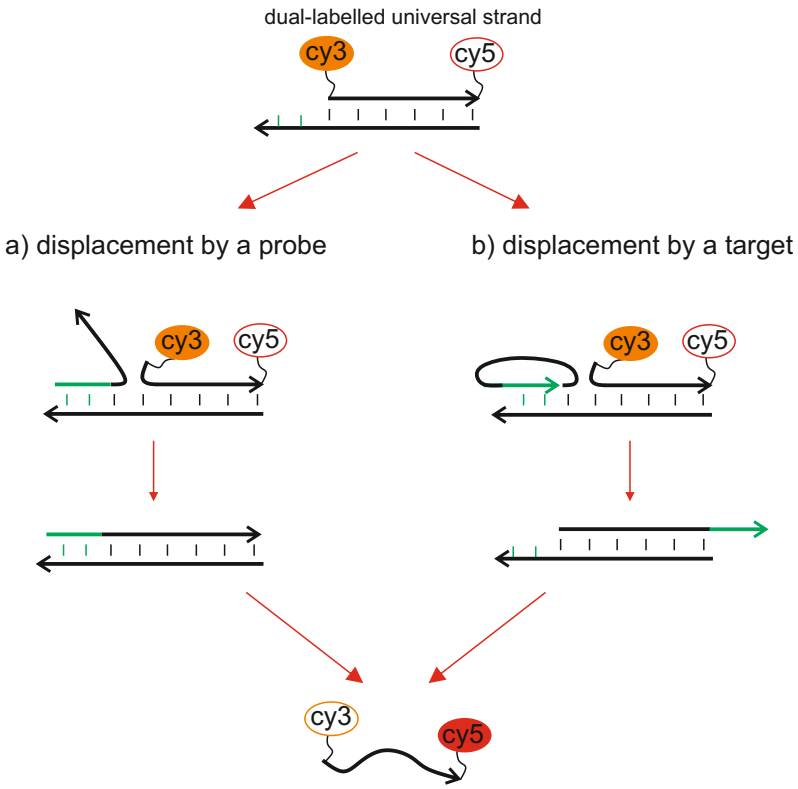


Figure 2.15: The rolling hypothesis.

It may be possible for a target strand to displace its own neighbour in the abscence of a probe strand, by rolling over on itself and hybridising to the neighbouring universal domain.

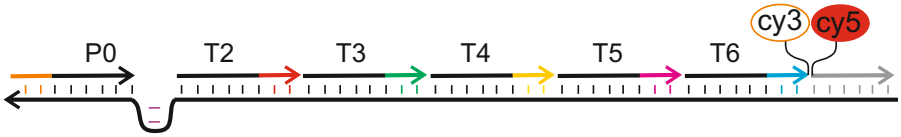


Figure 2.16: The secondary structure hypothesis.

If secondary structure exists in the toehold domains of the template strand, the correct probe strand may be the only strand able to interact with the structure and hybridise to the domain. Partial hybridisation by the wrong probes may be less energetically favourable than the secondary structure itself. This may explain why missing out probe P1 produces the same effect in both the original and shuffled cascade, and why P1 has to be added to the reactions in order for the cascade reaction to continue to the final step.

and compete with the neighbouring target strand over the universal domain. The idea was tested using the system shown in Figure 2.15. The initial state was a duplex between a dual-labelled universal strand and a template with a 3' free toehold. The system was tested by either adding a probe strand that is fully complementary to the template strand, or by a target strand which has a complementary toehold domain. The dual-labelled strand is attached to two fluorophores that form a FRET pair, and in the initial state, the universal strand is in an extended state as part of a duplex. If the universal strand is displaced, the single-stranded random coil conformation of the strand results in the fluorophores being closer together, and the donor fluorescence becomes quenched by the nearby acceptor fluorophore. The experiments (data not shown) showed that the probe strand is able to displace the universal strand, but no change in fluorescence was observed when the target strand was used as the displacer. Therefore, there is no evidence that free target strands are responsible for "leak" reactions.

Secondary structure

A third hypothesis concerns secondary structure. The strands used in the cascade experiments are known to form secondary structures, as more weight was placed on toehold specificity and free energy than on choosing strands that exhibit minimal secondary structure formation. If secondary structure exists in the toehold domain of a probe or target strand, then it also exists in the toehold domain of the template strand, which has the same sequence. When the target strands are hybridised to the template, the toehold domains are hidden. However, once a target strand has been displaced, the free toehold may be able to fold into an unintended secondary structure, as shown in Figure 2.16. The hairpin loop in the toehold domain might only be opened by the probe strand which is complementary to that domain. So far, it has been postulated that "leak" reactions occur when a probe strand uses

partial toehold hybridisation combined with the fraying of the 5' end of target strands to initiate a strand displacement. If secondary structure forms within a toehold domain of the template and the probe strand which can hybridise to the domain is not added, then the reaction may be prevented from continuing. In this case, partial hybridisation of any incorrect probe to the toehold domain may be less energetically advantageous than the secondary structure itself. This may explain why the absence of probe P1 prevents the cascade reaction from continuing to the final step in both the original and shuffled cascades. Analysis of the targets and probes using both NUPACK [Zadeh et al., 2011] and mfold [Zuker, 2003] indicate that all the strands form secondary structures at 21 °C, but none of the structures appear within the toehold domain only. It is not possible to analyse whether free toehold domains in the cascade duplex also form secondary structures, and so it is unclear whether this hypothesis is correct. One way to determine whether secondary structure plays a role in the "leak" reactions is to design a new cascade in which the secondary structure in the single-stranded species is minimal.

2.4.3 Minimal Secondary Structure Cascade

The toehold sequences for the minimal secondary structure cascade were chosen in the same way as for the previous designs, but particular care was taken to select strands which do not form secondary structures, especially structures involving their toehold domains. This design criteria was prioritised over sequence specificity and free energy similarity (see Section 2.2 for the design criteria discussion), whereas in the earlier cascade experiments, secondary structure was considered less important. As a consequence, the free energies of the toehold domains in the minimal secondary structure cascade are not as closely matched as in the earlier design (compare Table 2.7 with Table 2.5).

The six target strands, the reporter and the template strand which form the initial

Step	Strands Involved	ΔG^o_{probe} (kcal/mol)	ΔG^o_{target} (kcal/mol)	$\mathbf{k}_{zf}$ ($\mathbf{M^{-1}s^{-1}}$)	$\mathbf{k}_{zr}$ ($\mathbf{M^{-1}s^{-1}}$)
1	P0, T1	-9.39	-6.1	4.96×10^5	1.91×10^3
2	P1, T2	-6.1	-8.56	7.6×10^3	4.84×10^5
3	P2, T3	-8.56	-7.02	4.58×10^5	3.4×10^4
4	P3, T4	-7.02	-7.02	2.25×10^5	2.25×10^5
5	P4, T5	-7.02	-7.54	1.38×10^5	3.32×10^5
6	P5, T6	-7.54	-7.83	1.84×10^5	3.0×10^5

Table 2.7: The standard free energies of the probe and target toehold domains at each step of the minimal secondary structure six-step cascade at 21 °C. The initial toehold has been reduced to 6 nt. The toehold free energies are not as closely matched as in the earlier design because minimal secondary structure rather than toehold sequence was considered the design priority here. The bimolecular rate constants are generated using the parameters in the table, $k_{HT} = 5\times10^5$ $M^{-1}s^{-1}$ (since the toehold energies and the universal domain sequence match those of Section 2.2.2), and the model described in Section 2.1.2 and Appendix A.2.2.

state of the reaction were annealed together in the usual way. As in the earlier cases, the final target strand and the reporter strand were labelled with fluorophores that form a FRET pair (Fam and Tamra). The fluorescence spectroscopy experiments were conducted at 21 °C, and the donor (Fam) fluorescence recorded. The cascade system was first tested by adding all of the probe strands in either a 1-, 10- or 100-fold excess. The results of the reactions are shown in Figure 2.17. Minimising the secondary structure has had an effect on the yield of the reaction. The reaction using a 10-fold excess of the probe strands has not reached equilibrium (as the signal has not levelled off), whereas the same reaction in the original and shuffled cascades do reach equilibrium in the same time, but are only 60 % complete. As discussed in Section 2.4.2, secondary structures within any of the DNA strands in the cascade reaction could inhibit the

toehold exchange process. If free toehold domains are self-complementary and form hairpin loops (as in Figure 2.16), then these secondary structures must be opened before a probe strand can hybridise to the toehold and continue the cascade reaction. Similarly, if probe strands form secondary structures, particularly with their toehold domains, then the toehold exchange reaction is slowed down for the same reason. Since the bimolecular model assumes minimal secondary structure formation, it can be used to model this new cascade system. The rate constants were calculated from the free energies of the toeholds at each step and the results of the model are also shown in Figure 2.17. The model fits the 10-fold excess data particularly well and demonstrates that without secondary structure effects, the six-step cascade reaction rate can be accurately predicted using the relationship between the rate constants and the toehold domain free energies.

If secondary structure affects whether "leak" reactions take place, the omission of each probe strand in this minimal secondary structure cascade should produce the same result, either a "leak" or no "leak". This hypothesis was tested by performing six reactions in which each probe strand is omitted from the reaction in turn, and the results are shown in Figure 2.18. The results confirm the prediction; the omission of each probe strand results in a "leak". It can be assumed that "leak" reactions are initiated by weak interactions between an incorrect probe and an unmatched toehold. In the original cascade, secondary structure within the exposed toehold weakened binding to all probes. Where toeholds are mismatched, the strength of this effect depends on the sequences of both interacting domains. Weakening of non-specific toehold interactions, by designing intentional secondary structure in toehold domains may allow "leakage" strand displacement initiated by the exposure of certain toeholds to be effectively turned off.

Figure 2.18 also shows that the "leak" reaction rate generally becomes faster as the

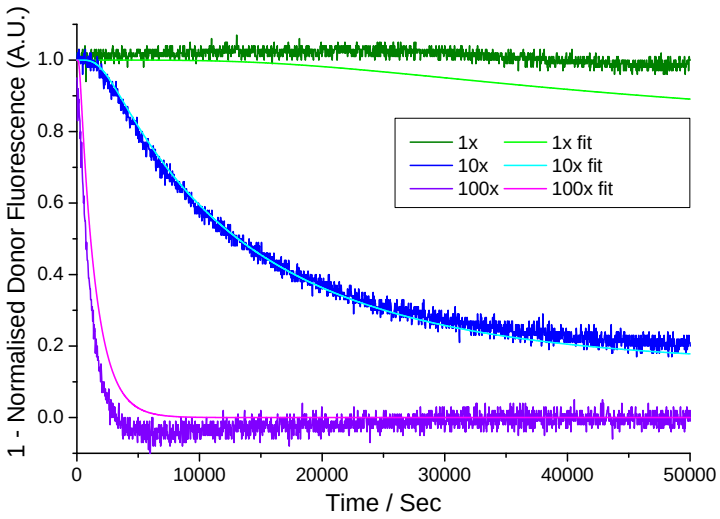


Figure 2.17: Minimal secondary structure cascade using variable probe concentration.
The six probe strands were added to the six-step cascade reaction in either a 1×, 10× or 100×. The data was fitted using the bimolecular mode, using $k_{HT} = 5{\times}10^5$ M^{-1} s^{-1}, and the toehold energies listed in Table 2.7. The model fits the data well, especially the 10-fold excess case. This indicates that by minimising the secondary structure in the design, the behaviour of the reaction is understood and can be modelled. This means that the system can be manipulated to tune the rate of the reaction to suit the cascade-coupling chemistry. This was not the case with the original cascade, as its behaviour was not predicted by the bimolecular model (Figure 2.9), and therefore tuning the system for the chemistry would be difficult.
The reactions were carried out as described in Figure 2.3, except the donor fluorescence here is Fam and was monitored using Ex490/Em520 nm.

missing probe is encountered further along the reaction. The "leak" reaction produced as a result of the penultimate probe strand (P4) being omitted is almost as fast as the reaction using all six probe strands. There is one notable exception to this observation however, as the "leak" reaction caused by missing out the final probe strand is slow and the reaction appears to be the only one to have reached equilibrium. In order to understand why the "leak" reactions are position dependent, the bimolecular model was modified to model the interactions of incorrect probes with free toeholds. In the absence of a probe strand, partially hybridised incorrect probes can initiate the strand displacement reaction. The model described in Section 2.1.2 and the equations

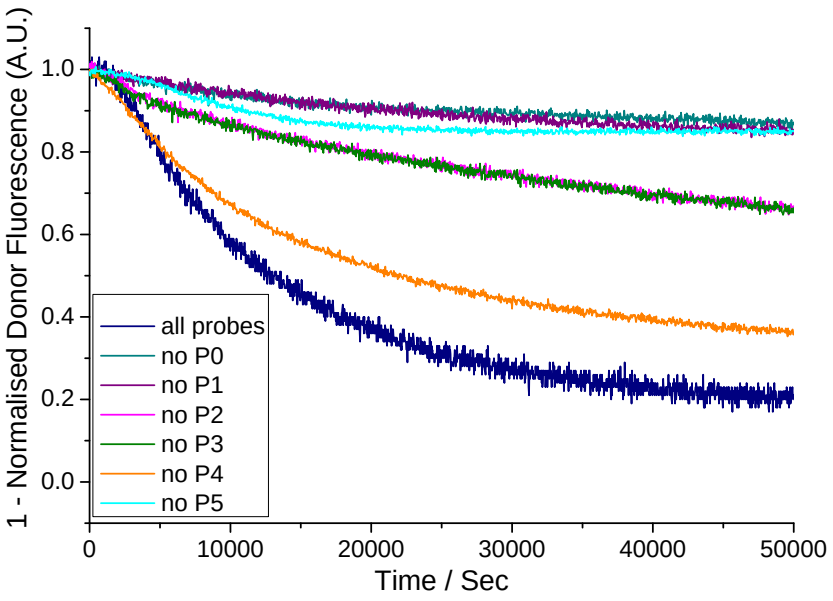


Figure 2.18: The effect of missing out a probe strand in the minimal secondary structure cascade. A "leak" reaction is observed whenever a probe strand is omitted from the cascade reaction. This is consistent with the hypothesis that secondary structure plays a role in the ability of incorrect probes to hybridise to a mismatched toehold and initiate a strand displacement. Here, secondary structure has been minimised, and as a result, incorrect probes are not prevented from performing incorrect strand displacements through weak, non-specific toehold interactions. Furthermore, it appears that the rate of the "leak" reaction is dependent on the position at which the missing probe strand is encountered along the cascade. The further along the cascade the missing probe appears, the faster the corresponding "leak" reaction. The "leak" produced by missing out the penultimate probe strand (P4) is almost as fast as the reaction using all six probes. However, missing out the final probe strand (P5) is the exception to the rule, as the resulting "leak" is slow and appears to have reached equilibrium during the experiment. This is similar to the behaviour observed in the original and shuffled cascades, where it was found that missing out the final probe strand produced no "leak" reaction. Here, a missing final probe produces the minimal "leakage".
The reactions were carried out as described in Figure 2.3, except the donor fluorescence here is Fam and was monitored using Ex490/Em520 nm.

in Appendix A.2.1 are not sufficient to model these incorrect reactions. If a wrong probe strand hybridises to a toehold, the product is different to that produced by the correct probe strand. Each toehold exchange reaction in the six-step cascade can either be completed correctly (by the correct probe) or incorrectly (by the other five probe strands), which means there are 2^6 ways of completing the six steps. Consider a two-step reaction, in which the two steps are completed by probes P1 and P2. If both steps of the reaction are completed correctly, the system is described by Equation 2.16. However, if either probe P1 or P2 are not added to the reaction, then the system is described by Equation 2.17 or 2.18 respectively.

both probe strands are added [P1 P2] (2.16)

$$P1 + T1T2t \underset{k_{zr1}}{\overset{k_{zf1}}{\rightleftharpoons}} T1 + P1T2t$$

$$P2 + P1T2t \underset{k_{zr2}}{\overset{k_{zf2}}{\rightleftharpoons}} T2 + P1P2t$$

the first probe P1 is not added [P2 P2] (2.17)

$$P2 + T1T2t \underset{kL_r}{\overset{kL_f}{\rightleftharpoons}} T1 + P2T2t$$

$$P2 + P2T2t \underset{k_{zr2}}{\overset{k_{zf2}}{\rightleftharpoons}} T2 + P2P2t$$

the second probe P2 is not added [P1 P1] (2.18)

$$P1 + T1T2t \underset{k_{zr1}}{\overset{k_{zf1}}{\rightleftharpoons}} T1 + P1T2t$$

$$P1 + P1T2t \underset{kL_r}{\overset{kL_f}{\rightleftharpoons}} T1 + P1P1t$$

The species produced by the incorrect displacement are different to that produced by the correct reaction, and therefore, these extra species need to be added to the model. In the case of this two-step system, ten equations are required to describe the reaction:

First Step: (2.19)

$$\frac{d(P1)}{dt} = -k_{zf1}(P1)(T1T2t) + k_{zr1}(T1)(P1T2t)$$

$$
\begin{aligned}
\frac{d(T1T2t)}{dt} &= -k_{zf1}(P1)(T1T2t) + k_{zr1}(T1)(P1T2t) \\
&\quad -kL_f(P2)(T1T2t) + kL_r(T1)(P2T2t) \\
\frac{d(T1)}{dt} &= +k_{zf1}(P1)(T1T2t) - k_{zr1}(T1)(P1T2t) \\
&\quad +kL_f(P1)(T1T2t) - kL_r(T1)(P1T2t) \\
\frac{d(P1T2t)}{dt} &= +k_{zf1}(P1)(T1T2t) - k_{zr1}(T1)(P1T2t) \\
&\quad -k_{zf2}(P2)(P1T2t) + k_{zr2}(T2)(P1P2t) \\
&\quad -kL_f(P2)(P1T2t) + kL_r(T2)(P1P1t) \\
\frac{d(P2T2t)}{dt} &= +kL_f(P2)(T1T2t) - kL_r(T1)(P2T2t) \\
&\quad -k_{zf2}(P2)(P2T2t) + k_{zr2}(T2)(P2P2t)
\end{aligned}
$$

Second Step: (2.20)

$$
\begin{aligned}
\frac{d(P2)}{dt} &= -k_{zf2}(P2)(P1T2t) + k_{zr2}(T2)(P1P2t) \\
&\quad -k_{zf2}(P2)(P2T2t) + k_{zr2}(T2)(P2P2t) \\
\frac{d(T2)}{dt} &= +k_{zf2}(P2)(P1T2t) - k_{zr2}(T2)(P1P2t) \\
&\quad +kL_f(P2)(P1T2t) - kL_r(T2)(P1P1t) \\
&\quad +k_{zf2}(P2)(P2T2t) - k_{zr2}(T2)(P2P2t) \\
\frac{d(P1P2t)}{dt} &= +k_{zf2}(P2)(P1T2t) - k_{zr2}(T2)(P1P2t) \\
\frac{d(P1P1t)}{dt} &= +kL_f(P2)(P1T2t) - kL_r(T2)(P1P1t) \\
\frac{d(P2P2t)}{dt} &= +k_{zf2}(P2)(P2T2t) - k_{zr2}(T2)(P2P2t)
\end{aligned}
$$

Highlighted in blue are the two species that arise from correct displacement of the target strand, whilst highlighted in red are the three species that arise from incorrect target displacements by the wrong probe strand.

To simplify the model, it is assumed that each of the five wrong probes can hybridise incorrectly with equal probability to the free toehold at each step, as there is no

experimental evidence to suggest otherwise. Furthermore, it is assumed that only one probe can be missed out per six-step reaction, to simplify the model further. This has also been experimentally observed: cascade reactions were performed by missing out two or more probe strands at a time (both consecutive and non-consecutive probes), and no change in fluorescence was observed in any of these tests (data not shown here). It is not necessary to make this simplification in the model, but it is suitable for a first attempt at modelling the "leaks".

Forty ODEs are necessary to describe a six-step system in which one incorrect displacement can take place per cascade reaction, which are listed in Appendix A.2.2. At each step, the correct displacement is highlighted in blue (for instance, the species $P_1T_{2-6}t$ is formed when the first target strand T1 has been replaced by P1, but the remaining five targets are still hybridised to template t). The species that results from an incorrect reaction contains a term P_x to indicate that one of the other probe strands have completed the step (for instance, $P_xT_{2-6}t$, in which the first target strand T1 has been incorrectly displaced by the wrong probe strands). The "leaks" can occur whether or not a probe strand has been added to the reaction, but when the correct probe is present in solution, it is more likely that it will initiate the strand displacement than the incorrect probe. The rate constants for the correct toehold exchange reactions are still those listed in Table 2.7. The "leak" rate constants kL_r and kL_f were varied to obtain the pair that best fit the "leaks" observed experimentally. The average free energy of the probe toehold domains is ΔG^o_{probe} = -7.6 kcal/mol, and the average free energy of the target toehold domains is ΔG^o_{probe} = -7.3 kcal/mol. These values were used to obtain the forward and reverse "leak" rate constants, which are assumed to be the same at each step of the cascade. These rate constants were calculated using the bimolecular model (Section 2.1.2), the average toehold free energy and a fraction of the average probe toehold energy, as incorrect probe strands cannot hybridise to the full length of the wrong toeholds. Figure 2.19 shows the "leak"

model results using "leak" rate constants corresponding to a probe toehold energy of ΔG^o_{probe} = -3.04 kcal/mol (40 % of the average probe toehold energy). The "leak" model does not exhibit the same behaviour on encountering a missing probe strand as the data shown in Figure 2.18. The speed of each "leak" reaction does not increase as the position of the missing probe moves along the cascade. The only resemblance between the data and the model is the result of missing out the final probe strand P5. Altering the rate constants by changing the probe toehold energy did not obtain better fits to the experimental data.

The results of the "leak" model indicate that the cause of the "leak" reactions are still not well understood. Certain assumptions and approximations used in the model may be incorrect. For instance, it was assumed that each probe strand can hybridise to each toehold domain and with the same strength. This is probably unrealistic, as analysis of the toehold domains and their complements using Nupack indicated that two of the probe strands could hybridise to incorrect toeholds and form duplexes with almost the same free energy as the correct probe strands, whereas other probe strands were specific and could only hybridise to domains they were designed to interact with. This may explain some differences between the model and the experimental data, as the model does not currently include this feature. However, since some toehold domains are specific, this analysis suggests that it is not possible for the wrong probe strand to bring about a strand displacement, which is contrary to the observed results. This also means that using one set of "leak" rate constants for each step of the cascade may be incorrect, but altering the rate constants for each step without any experimental evidence to support the action is not useful. The model also becomes a lot more complicated if the interaction of each probe strand with each toehold domain is considered separately rather than assuming all interactions are identical. Whilst it is interesting to be able to expand the model to account for the "leak" reactions and understand the cascade mechanism in more detail, it does not directly meet the

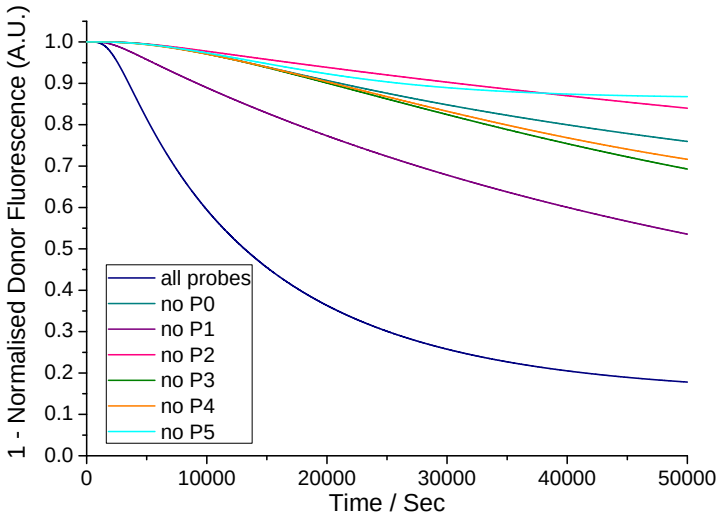


Figure 2.19: Simulation results from the "leak" cascade model.

The "leak" reactions were modelled by solving the ODEs listed in Appendix A.2.2 for each missing probe strand. The "leak" rate constants were obtained by using the average target and probe toehold free energies and the bimolecular model. The "leak" reactions are assumed to occur because incorrect probe strands are able to partially hybridise to the free toehold in the absence of the correct probe. The results of the model shown here assume that incorrect probe strands hybridise to a fraction of the (average) probe toehold and use $k_{HT} = 5{\times}10^5$ $M^{-1}s^{-1}$, $k_{zf} = 3.09{\times}10^2$ $M^{-1}s^{-1}$ and $k_{zr} = 4.43{\times}10^5$ $M^{-1}s^{-1}$. The model is not improved by changing the strength of the interaction between incorrect probes and toeholds. The model does not predict the behaviour observed in the experiments (see Figure 2.18), as the "leak" reactions shown here do not increase in speed as the missing probe is encountered further along the cascade. The only common factor between data and model is that both result in slow reactions when the final probe strand is missed out.

The fits are normalised using the long-time value of the model when all probe strands are added in a 100-fold excess.

aim of the synthetic ribosome project. For the purpose of designing a tunable DNA mechanism to suit a cascade-coupling chemistry, the focus was shifted to finding other ways to remove the "leak" reactions.

One way to reduce the probability of "leak" reactions is to purify the strands and the duplex before conducting the experiments. The decision to purchase strands unpurified from the supplier and to use them without performing any purification technique was explained in the design criteria discussion in Section 2.2.1. Since the DNA supplier recommends purifying strands which are longer than 40 nt, the short probe and target strands were purchased and used unpurified, as it was assumed that synthesis mistakes would be minimal. However, this may have been an incorrect assumption. "Leak" reactions have been observed in large-scale DNA logic circuits, and these have often been attributed to the purity of the strands used in the tests [Seelig et al., 2006]. Commonly, these undesired reactions are removed or minimised by purchasing HPLC- or PAGE-purified strands from a synthetic DNA supplier. However, another method is to assemble the initial duplex of a reaction and PAGE-purify the duplex from all the other strands in the solution. This method was used to purify the minimal secondary structure cascade (Figure 2.20), and is described further in Appendix A.1.4. The purified sample was used in a fluorescence spectroscopy experiment to determine whether a "leak" reaction was still possible. All six probe strands were added to the sample (in a 50-fold excess) to check that the cascade reaction was unaffected by the gel purification technique, and all probes except P4 were added in another reaction in a 10-fold excess. A 10-fold excess was used for the missing P4 reaction since this probe concentration had previously produced a significant "leak" (see Figure 2.18). However, a 50-fold excess was used for the reaction using all six probe strands, simply to produce a fast reaction that could quickly provide evidence of the correct operation of the cascade reaction. Ideally, a 50-fold excess experiment would also have been conducted for the missing probe reaction, as the larger probe strand

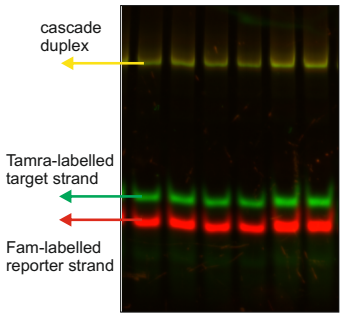


Figure 2.20: Segment of PAGE gel used to purify cascade sample.

The minimal secondary structure duplex was assembled and the sample added to 10 lanes of a 15 % non-denaturing PAGE gel (see Appendix A.1.3). The gel was run for 45 min at 240 V and 4 °C in 1× TAE buffer. The gel was scanned using the Fam and Tamra wavelengths (Em495/Ex520 nm and Em565/Ex580 nm respectively). The multi-channel image of the gel shows a slowly migrating band at the top of the gel, which corresponds to the cascade duplex as it contains both fluorophores. Excess strands appear at the bottom of the gel. The bands corresponding to the cascade duplex were cut out from the gel, and the DNA extracted using the crush and soak method described in Appendix A.1.4.

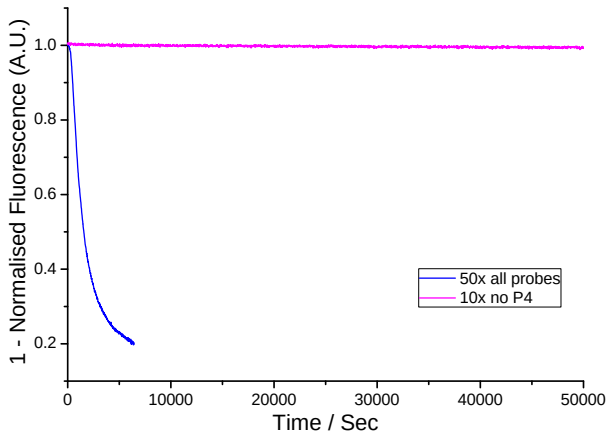


Figure 2.21: The gel-purified cascade reaction.

The minimal secondary structure duplex was assembled and PAGE purified before being used in an experiment. The purified sample was tested by adding a 50-fold excess of all six probe strands and by adding a 10-fold excess of five probe strands (see main text). The results indicate that purifying the sample in this way does not affect the operation of the cascade mechanism, but it does prevent a "leak" reaction from taking place each time a probe strand is missed out (only the result of missing out P4 is shown here as the data from each "leak" reaction obscure each other.) Compare the effect of missing out P4 here with that of Figure 2.18.

Data is normalised using the post-anneal value of the 50-fold excess reaction.

concentration would better prove the absence of "leaks" after purification. Unfortunately, time and resource constraints did not allow this to be carried out. The results are shown in Figure 2.21, and indicate that the cascade reaction is not affected by gel purification but "leak" reactions are effectively prevented. This implies that "leak" reactions are at least partially caused by incorrectly synthesised strands (which have not been removed by the DNA supplier), incorrectly assembled duplexes and free, unhybridised strands in the solution. If it is possible to PAGE purify the cascade systems that have been coupled to chemistry, then this is a simple way of removing the possibility of unintended "leak" reactions and truncated/incorrect chemical products. PAGE purification may enable the cascade-coupling mechanism to meet the design specifications of the synthetic ribosome.

2.5 Cascade-Coupling Chemistry

As discussed earlier in Section 2.1, the ultimate goal of the synthetic ribosome mechanism tested here is to facilitate the reaction of chemical building blocks and hence, to generate libraries of new functional molecules. DNA-templated synthesis (see Section 1.5.1) allows the interaction between chemical groups attached to DNA strands to be controlled via the hybridisation of the oligomers. One popular chemical reaction used in DNA-templated chemistry is the Wittig reaction, discussed in Section 1.5.2. Here, an aldehyde and a phosphonium ylide react to give an alkene and a phosphine oxide, and is widely used because of its robustness in aqueous solution and because it can be adapted to suit a variety of DNA architectures [Gartner et al., 2003], [Rozenman and Liu, 2006]. The Wittig reaction is particularly suited to the cascade mechanism because it is a group transfer reaction (see Figure 1.15). During the reaction, the Wittig product is transferred from one DNA strand to another. Furthermore, the Wittig reaction has already been applied to a DNA-templated synthesis mechanism and shown to produce a chain of four covalently-bound Wittig building

blocks [McKee et al., 2010]. The cascade-coupling mechanism was therefore adapted for the Wittig reaction to determine whether the mechanism can be used to control the reaction between chemical building blocks.

A one-step cascade-coupling system was designed to test the Wittig reaction, as shown in Figure 2.22. The Wittig reacting groups are attached to two DNA target strands, which are made of two domains, universal and toehold. In the initial state, the two targets form a duplex with a template strand, which has a single-stranded initial toehold at the 3' end. The two target strands were purchased from IDT with a 5' amine-modification to allow attachment of the Wittig building blocks. The first target strand is linked to the fluorescent phosphine ylide building block (FAM), while the second is linked to benzaldehyde building block (BAL). The synthesis of the Wittig-functionalised DNA strands is described in Appendix A.3. Once the strand displacement reaction has begun, the phosphine ylide and aldehyde groups are able to react, and the group transfer reaction transfers the Wittig product to the second target strand. This group transfer means the cascade-coupled chemical product moves in the same direction as the migrating toehold of the strand displacement reaction.

The target strands used in this test system differ from those used in the earlier experiments. Here, both target strands consist of 18 nt universal and 6 nt toehold domains, but also an additional universal 6 nt at the 5' end that is not complementary to the template $\langle \texttt{ul u t} \rangle$. This extra length `ul` is the *universal linker* domain, and is single-stranded at the start of the reaction. The linker domain was introduced for two reasons. Firstly, it provides the target strands with extra flexibility to bring their building blocks into close proximity. This means the chemical reaction could potentially start when only a few target strand nucleotides have been displaced, thereby maximising the chance that the reaction completes before the target strand dissociates from the template. The motion of the 5' end of the strand is not controllable,

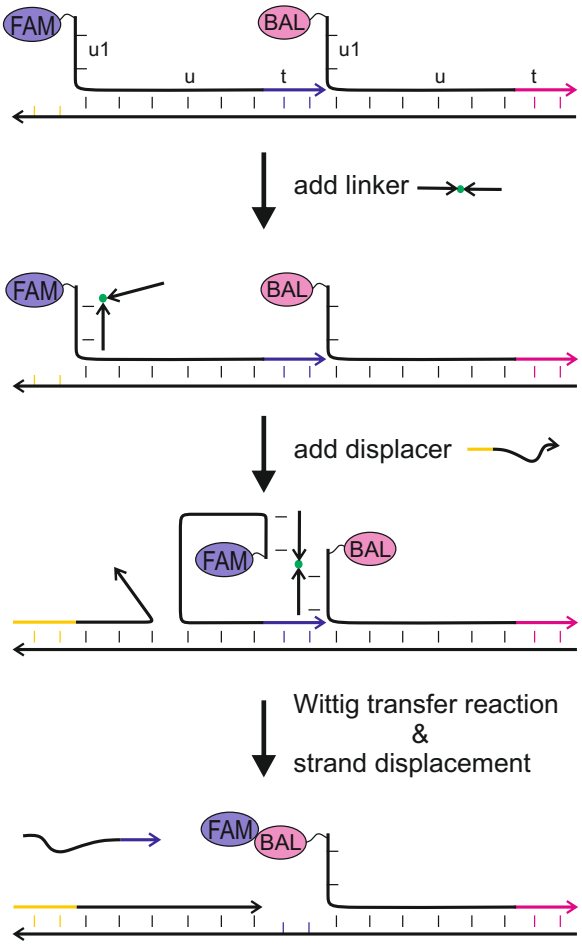


Figure 2.22: The cascade-coupling Wittig group transfer reaction.

The initial state of the system is a duplex formed between two target strands and a template strand. The target strands ⟨`ul u t`⟩ each consist of three sections: a "toehold" domain `t`, a "universal" domain `u` and a "universal linker" domain `ul`. The `ul` domains are single-stranded at the start of the reaction. Each target strand is linked to a Wittig building block on its 5' end.

Two 6 nt DNA strands are linked together at their 3' ends to form a DNA "linker" (custom synthesised by IDT). This linker is complementary to the universal linker domains of the target strands, and serves to hold the chemical building blocks in close proximity whilst the target strand is in the process of being displaced from the template. The group transfer reaction of the Wittig chemistry means the FAM group is transferred to the second target strand. The first target strand is displaced entirely, revealing a new toehold to which subsequent displacers may hybridise, and the short length of the linker DNA encourages it to disscociate from the complex also.

and so it is possible that on average it is in contact with the 5' end of the neighbouring target for only a short time. For this reason, the extra nucleotides were added to the targets, as they allow a mechanism to bring together the two target strands to be deployed. As shown in Figure 2.22, a DNA "linker" can be used to encourage the two target strands to come together. The linker is made of two short DNA strands linked to each other by their 3' ends (custom synthesised by IDT). The linker is complementary to the linker domains of the target strands, and serves to bring the building blocks into close contact while one of the target strands is in the process of being displaced from the template, as shown in the diagram. The duplex formed with the linker is unstable due to its length, which means the linker can dissociate from the two targets, and free the linker domain of the second target strand to allow the cascade-coupling reaction to continue.

Samples were made by annealing together target 1 (Adaptor 1), target 2 (Adaptor 2) and the template strand in equimolar quantities by heating the solution to 85 °C to melt all possible DNA structures, and cooling to 20 °C over 10 min. Samples were prepared in TAPS buffer (see Appendix A.3.2). The cascade-coupling mechanism was tested by adding a ten-fold excess of the displacer strand to the initial duplex, both with and without addition of the DNA linker. Figure 2.22 shows the linker hybridised to the universal linker domain of the first target strand. However, since the linker domains are identical, the linker could equally hybridise to the second target strand or to both the target strands. The latter case is undesirable, as it prevents the linker from bringing the target strands together as intended. To minimise the possibility of two linkers hybridising to one initial duplex, the concentration of the DNA linker added to the reaction was half that of the sample concentration. In the test in which both linker and displacer are added, the linker was added to the sample 10 min before the displacer was added to allow it to hybridise to the duplex before the displacement reaction begins. Reaction samples were incubated for 2 hours at 21 °C on a shaker

(as per the protocol used in [McKee et al., 2010]).

The samples were analysed by running a 15 % non-denaturing PAGE gel for 45 min at 240 V in 1 × TAE buffer at 4 °C (see Appendix A.1.3 for details on the protocols and solutions required for the non-denaturing PAGE analysis). The gel was first scanned at the Fam emission and excitation wavelengths (Ex488/Em530 nm), and then after staining with Sybr Gold nucleic acid stain, was scanned again at the Sybr Gold wavelengths (Ex488/Em530 nm). The scans are shown in Figure 2.23. Lane 5 contains the initial duplex before the strand displacement reaction takes place. Lanes 6–7 of gel A show that adding a 10-fold excess of the displacer strand results in the first target strand (A1) being displaced. However, a 10-fold excess is not sufficient to displace all the A1 strands within the 2 hour reaction time, as the Fam-labelled strand is still attached to the template (top bands in lanes 5–8). More notably, no third fluorescent complex appears in any of the reaction lanes, indicating that the strand displacement has taken place, but the Wittig group transfer reaction has not occurred. This is confirmed by the Sybr Gold stained gel (B). Highlighted in the red box is a complex that is presumed to correspond to the duplex made between the displacer strand, the second target strand and the template. This complex was not observed in gel A, and so the final product of the reaction does not contain a fluorescent group, which indicates the Wittig reaction has not taken place. This is the case whether or not the linker has been used.

There are a number of reasons why the Wittig group transfer reaction may not have taken place. Firstly, it is possible that the linker does not bring the reacting groups into close proximity as designed. The linker domains on the target strands are only 6 nt, which means the linker can hybridise to the targets but only temporarily, as a 6 base pair duplex is fairly unstable. If the linker duplex melts before the target strand is displaced from the template, or if the linker duplex lifetime is less than

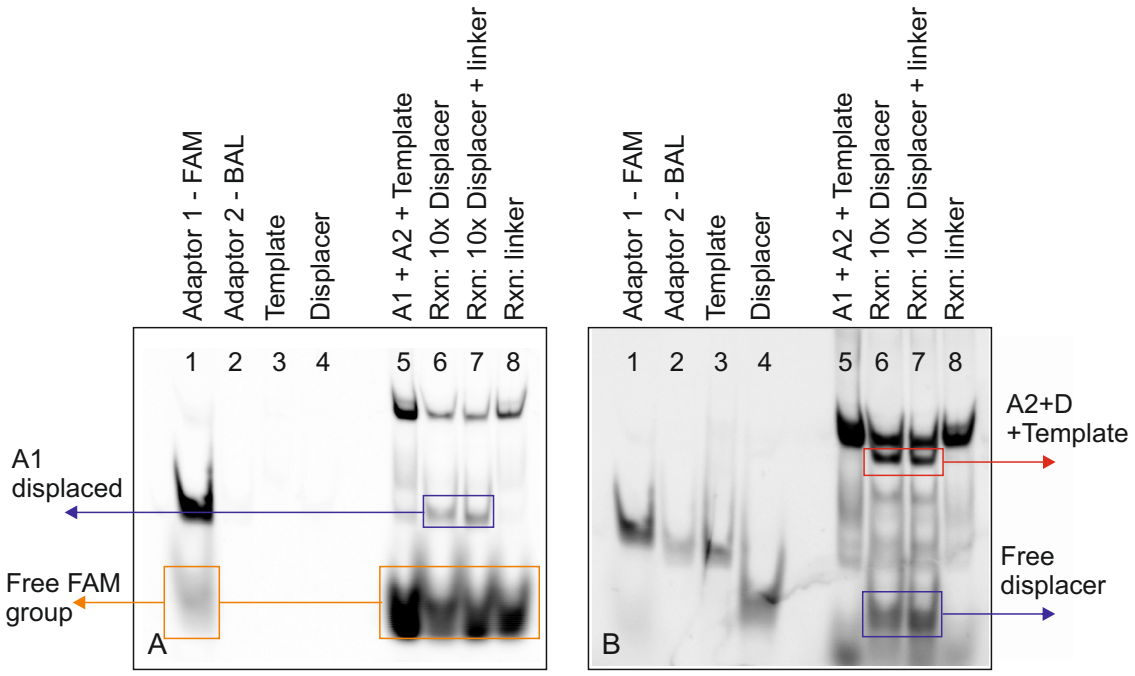


Figure 2.23: Results of the cascade-coupling Wittig reaction.

The reaction sample was formed using equimolar quantities of the Adaptor 1, Adaptor 2 and the Template (lane 5), in TAPS buffer (pH 8.5). Samples were left to react for two hours before being analysed by running a 15 % non-denaturing PAGE gel.

(A) The gel was scanned at the FAM excitation and emission wavelengths. The slowly migrating band at the top of lanes 5–8 corresponds to the initial duplex. From lanes 6–7, it is clear that adding a 10-fold excess of the displacer strand to the initial duplex results in some of the fluorescent Adaptor 1 being displaced from the template (highlighted by the blue box). Adaptor 1 is not a stable complex; it was observed that the FAM group becomes detached from the DNA over time, such that the DNA must be used soon after synthesis. The detached FAM group appears as the fast bands at the bottom of lanes 1 and 5–8 (orange box).

(B) The gel was treated with nucleic acid stain Sybr Gold, and scanned at the Sybr Gold excitation and emission wavelengths. Lanes 6–7 show that a new duplex is formed as a result of the strand displacement reaction, but that this duplex is not fluorescent. It can be assumed that whilst the strand displacement mechanism has worked, the Wittig chemistry has not taken place, as the FAM group has not been transferred to Adaptor 2.

the time required for the Wittig reaction to take place, then the linker cannot bring together the reacting groups. It is not simple to rectify this problem, since if the linker domains are made longer, the duplexes may be too stable. A degree of duplex instability is desired as it allows the linker to dissociate from the cascade complex, which in turn frees the linker domains for any subsequent steps in the cascade. The second reason may be because the mechanism to bring together the reacting groups may not be a suitable architecture for the Wittig reaction. Whilst Wittig chemistry has been shown to work in a variety of DNA-templating architectures, perhaps the linker mechanism positions the reacting groups too far apart from each other on the duplex to react. Another reason may arise from the linker hybridising to both the first and second target strand linker domains. In this case, the linker mechanism cannot operate as designed, and so the functionalised ends of the target strands are prevented rather than encouraged from reacting. This could be avoided by making the linker domain sequences of each target strand unique. However, this complicates the cascade-coupling mechanism by necessitating multiple linkers that each interact with specific pairs of target strands. It also adds an extra identifier to each target strand which complicates the sequence design of the strands. Currently, it is unclear which of these reasons resulted in the cascade-coupling Wittig reaction being unsuccessful.

2.6 Summary

The design and operation of a DNA mechanism that has the ability to sequentially couple together chemical building blocks has been investigated. The cascade-coupling mechanism uses the toehold exchange strand displacement process to walk along a DNA track that encodes the instructions for the interactions between chemical building blocks. Sequential reading of the track is a crucial requirement of this synthetic ribosome mechanism, in order to ensure the chemical building blocks interact according to the instructions on the DNA track (as in the natural ribosome). However,

the work presented here shows that whilst it is possible to use the toehold exchange process to move the cascade-coupling mechanism multiple steps along the track, it is necessary to prevent unintended and non-sequential "leak" reactions by gel-purifying the strands before use. "Leak" reactions have been observed when one of the displacer strands required for the toehold exchange has been omitted from the reaction. The natural ribosome is not able to skip over codons on the mRNA template, but the synthetic ribosome does have this ability and is capable of producing polymeric molecules that do not correspond to the instructions encoded within the DNA template. A number of hypotheses for the cause of the "leaks" have been discussed, but it is likely that the reactions are enabled by partial hybridisation of the wrong probe to a toehold domain and fraying of the 5' end of the target strand from the duplex. It may be possible to reduce the "leak" by intentionally allowing toehold domains to form secondary structures which can only be 'opened' by the probe strands designed to hybridise to the domain. However, secondary structure complicates the kinetics of a toehold exchange reaction, and to allow the reaction to be tuned to suit the cascade-coupling chemistry, it has been shown that it is better to use a minimal secondary structure system, the behaviour of which can be predicted and controlled using the bimolecular model. If the systems are gel-purified before use, the "leak" reactions disappear, and therefore the possibility of non-sequential or unintentional cascade reactions is minimal, and the system better matches the natural ribosome.

The cascade-coupling mechanism was tested by combining the strand displacement with Wittig chemistry. However, this has not been successful, and the reasons are not entirely clear. The purpose of this work is to explore DNA architectures that can be used to template chemical reactions. The potential for this design to be adapted for a chemical reaction is clear, however, a second architecture is explored as an alternative in Chapter 3.

Chapter 3

DNA Origami as a Substrate for Chemical Reactions

3.1 Introduction

In the previous chapter, a linear system to facilitate the synthesis of new functional molecules was discussed. Here, an alternative is proposed: a hexagonal arrangement of building blocks anchored to a DNA origami tile.

The principle of DNA origami is that a 2D structure can be made by folding a long genomic DNA strand into any desired shape, as discussed in Section 1.2. Since the seminal paper on this idea was published, many DNA nanotechnologists have exploited the technique for their own particular branch of research. The technique has been extended to producing more complicated geometrical shapes such as non-planar origami tiles [Li et al., 2010], Möbius bands [Han et al., 2010] and complex 3D structures [Douglas et al., 2009a], [Liedl et al., 2010]. Reviews of the versatility of DNA origami are presented in [Kuzuya and Komiyama, 2010], [Shih and Lin, 2010], [Nangreave et al., 2010]. Aside from being able to fold DNA template strands into multi-dimensional objects, the potential to use DNA origami constructs for practical

purposes has also been demonstrated. For example, a 3D DNA box made entirely from folding DNA, with a lid that can be opened and closed as desired, has been put forward as a mechanism to encapsulate and release drug molecules for the purpose of targeted drug delivery in cells [Anderson et al., 2009]. All of the examples mentioned so far exploit the ability to manipulate DNA strands into elaborate configurations, but much research is being conducted on using the basic Rothemund rectangular origami tile as a substrate. For instance, many of the DNA walkers discussed in Section 1.3 require long tracks to move along, and the origami tile can be adapted to provide such a track. In order to do this, particular staple strands can be lengthened to contain domains that do not form part of the tile structure, but which emerge out of the tile as anchorages for a DNA walker (Figure 3.1). One immediate advantage to using the origami tile as a substrate is to provide the walker with the option of moving in two dimensions, as demonstrated by two recent examples which use multiple "legs" to facilitate motion across the tile [Gu et al., 2010], [Lund et al., 2010].

The use of the rectangular origami tile is not limited to a substrate for DNA walkers to move upon. It has also been used as a template for chemical reactions [Voigt et al., 2010]. Here, staple strands were modified, this time not by increasing the length but by conjugating chemical functional groups to the ends of selected staples. To demonstrate the feasibility of using origami in this way, four staples were modified with either biotin, an alkyne, an amine or an azide (Figure 3.2 a). The biotin strand was used as a reference, while the three functional groups were selected for frequently being used in biconjugation reactions. The azide and alkyne can react together in a copper-catalysed click reaction (as described in Section 1.5.3), while the amine can react with an NHS ester to form an amide. The origami tile was formed in the usual way, and then deposited on a mica surface to enable imaging by atomic force microscopy (AFM). The chemical reactions were performed *in situ* on the surface-immobilised origami. To perform the click reaction, for instance, a biotin-linked azide

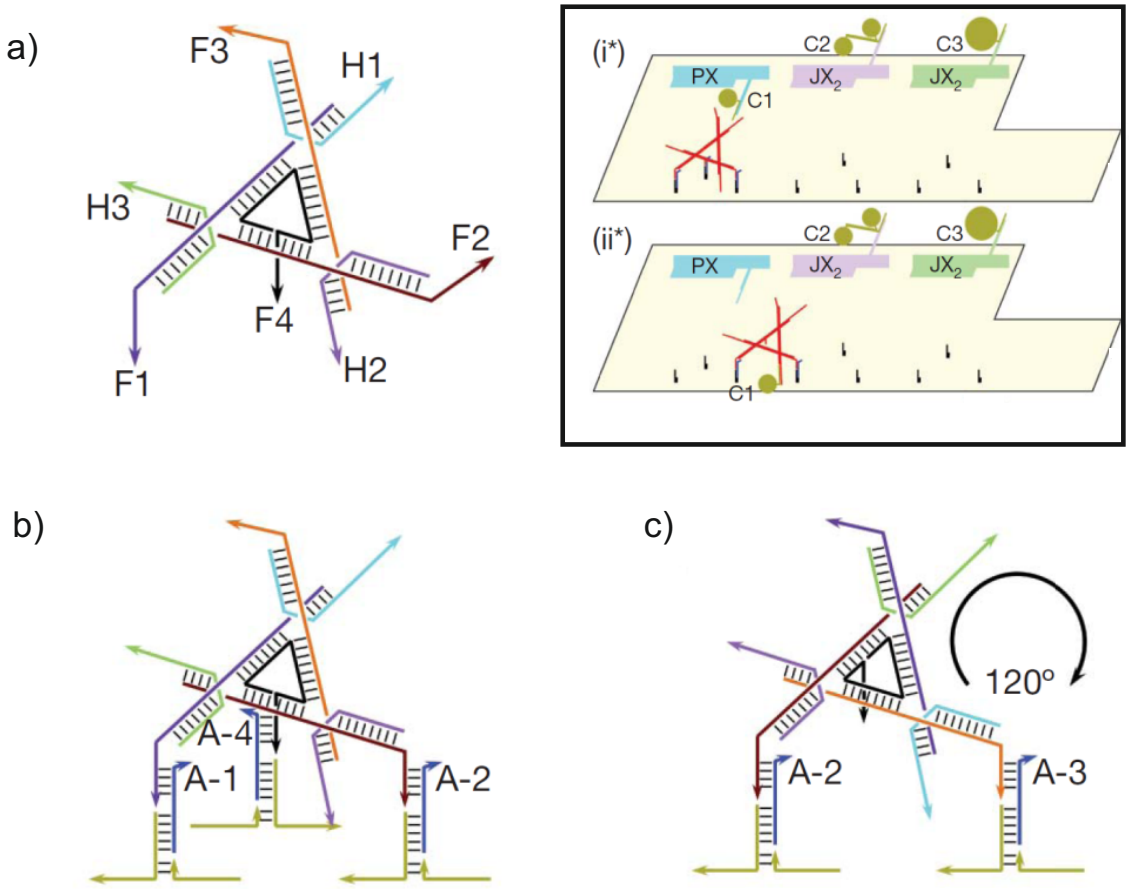


Figure 3.1: A DNA walker capable of 2D motion on an adapted DNA origami tile.

(a) The walker is a trigonal arrangement of double helices, and consists of three hands (H1-H3) and four feet (F1-F4). The feet are used to anchor the walker to the origami. The hands are free to capture cargoes that are located on the origami as the walker moves past them (top right box). Thus, not only is the walker able to move along the surface, but it is able to react with other nanoparticles in a controlled, programmable way.

(b) At any time, three of the feet are hybridised to extended origami staple strands (shown as green-brown lines) which form anchorages for the walker. Here, feet F1, F2 and F4 are hybridised to the origami. One of the hands is able to pick up a cargo (i*).

(c) Through the addition of fuel strands, a foot (F1) can be lifted off its anchorage, and the previously free foot (F3) can hybridise to an anchorage instead, thereby rotating the walker through 120°. This swivling motion means a free hand can always be positioned close to a cargo (ii*). Figure adapted from [Gu et al., 2010].

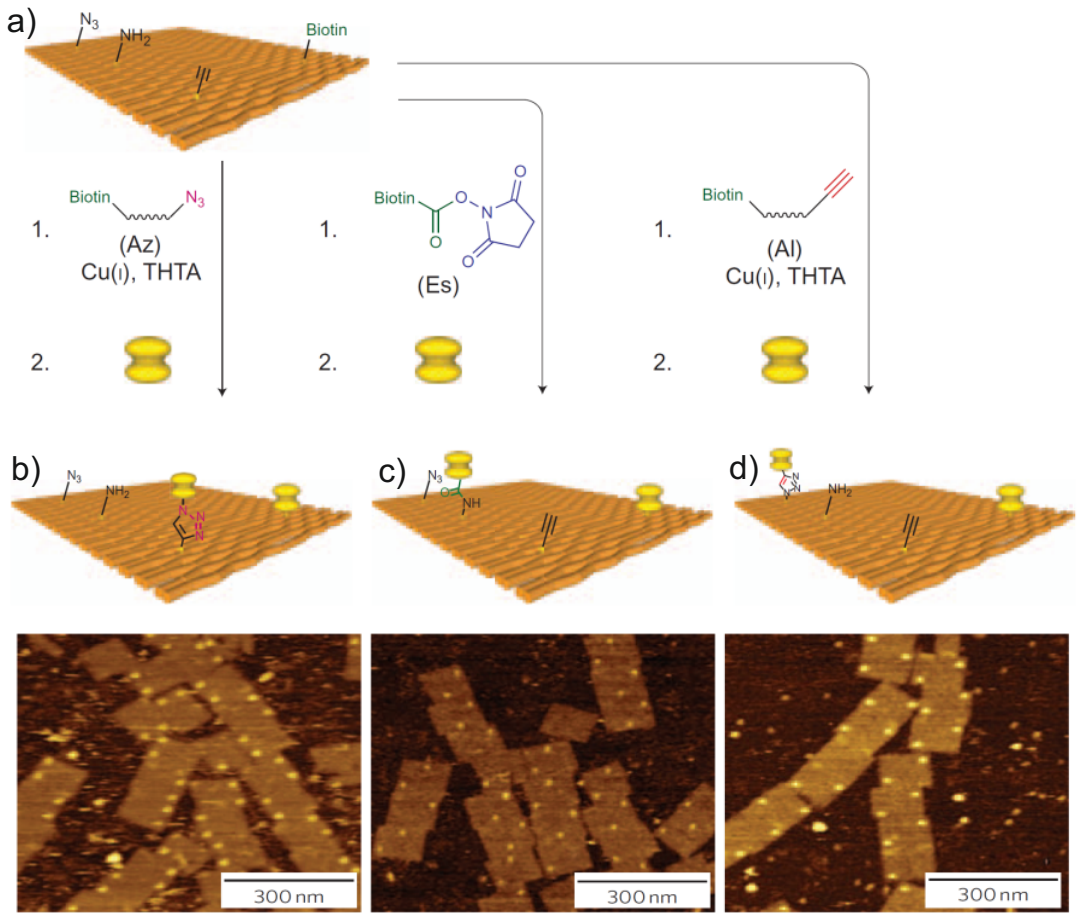


Figure 3.2: The use of DNA origami as a substrate for chemical reactions.

(a) Sketch showing the incorporation and position of azide-, alkyne- and amine-functionalised staple strands into a DNA origami tile.

(b) Addition of biotin-linked azides and the copper catalyst enables a click reaction between the alkyne-functionalised staple strand (1). In order to determine the success of the click reaction, streptavidin was added at the end of the reaction (2). The biotin-streptavidin interaction protudes from the origami surface, allowing visualisation of the covalent chemical reaction under AFM.

(c) Similarly, addition of an NHS-ester to the tile containing an amine-modified staple results in an amide product, which is visualised in the same way.

(d) The click reaction also works if the staple strand is azide-functionalised and biotin-linked alkynes are added alongside the catalyst. Figure adapted from [Voigt et al., 2010].

or biotin-linked alkyne was added to the origami in the presence of the copper catalyst. The solution was left to incubate for 20 min at room temperature. To determine if the reactions had worked, the origami samples were incubated with streptavidin for a further 5 min after each reaction. The biotin-streptavidin interaction enables the visualisation of the covalent chemical reactions under AFM, as the streptavidin protein protudes from the surface of the tile (Figure 3.2 b-d). It was found that the reactions proceed with a yield of over 80 %, and therefore it is possible to perform chemical reactions on the origami tiles, using catalysts and buffers that appear to have no destructive effect on the origami structure itself.

3.2 Origami-Templated Click Chemistry

The above examples of DNA origami being used as a substrate are the inspiration for the origami-templated chemistry architecture presented here. In a similar manner to [Voigt et al., 2010], the system described here uses an origami tile as the template for copper-catalysed click reactions. However, instead of needing to add the azides or alkynes with which the staple strands react, the functionalised staple strands are designed to interact with each other.

The design for the templating architecture is somewhat influenced by the work in the previous chapter. Functionalised strands need to be kept suitably far apart from each other on the origami to begin with, and a mechanism to bring the strands together needs to employed to bring the building blocks near each other when required. There are several ways in which the functionalised strands could be arranged to interact with each other. Firstly, the interaction could be "linear" like the cascade system. Here, the modified strands could be arranged along the diagonal of the origami tile, so that building blocks interact with each other one after the other, and the growing chemical product moves along the track. However, this is simply a way of extending

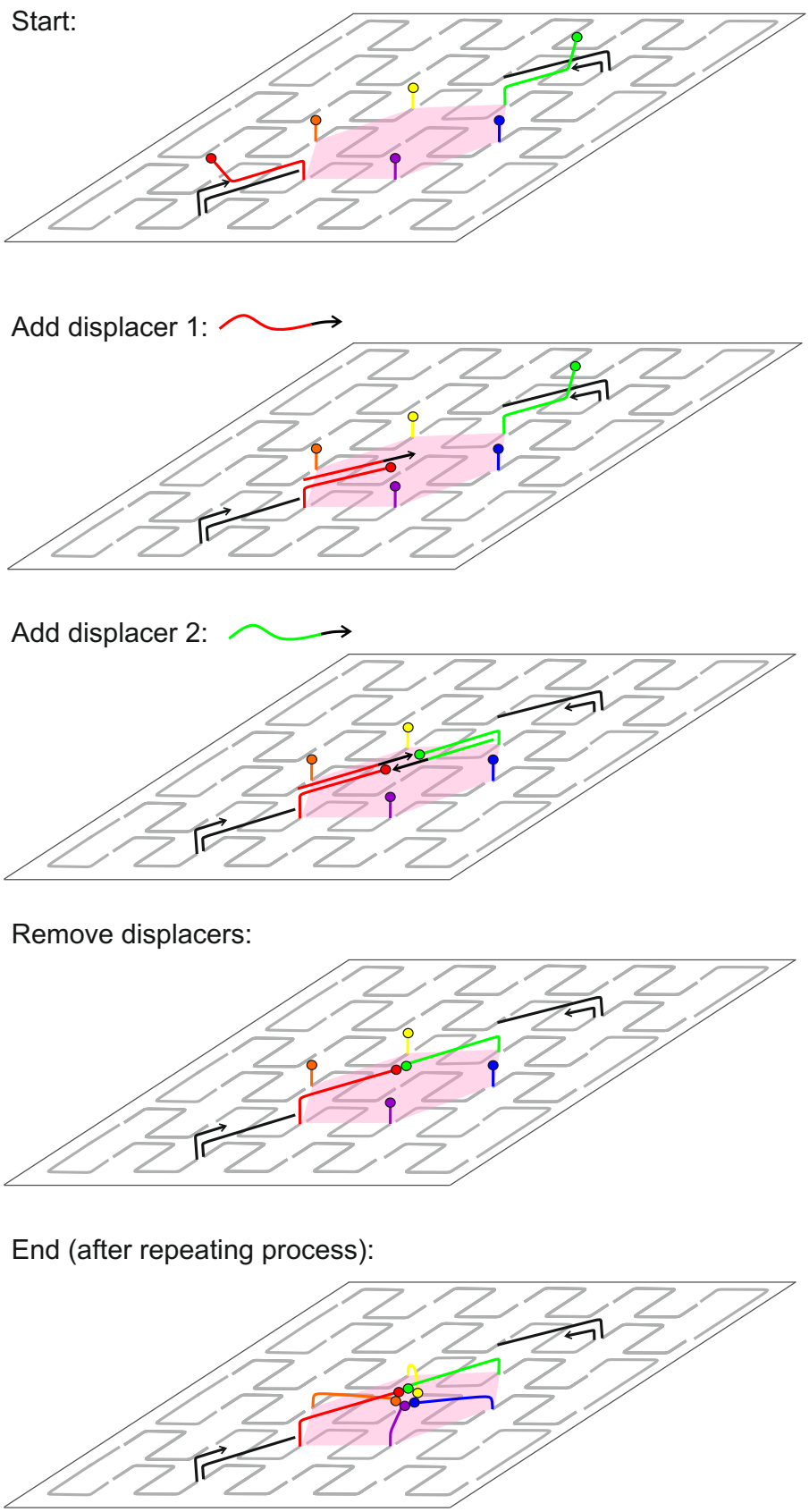

Start:
Add displacer 1:
Add displacer 2:
Remove displacers:
End (after repeating process):

Figure 3.3: Illustration of the origami "reaction centre" idea.

The six circles and coloured lines mark the positions of the functionalised staple strands on the origami tile. The circles represent chemical building blocks, which are each attached to a modified staple strand (coloured line). Each of the six modified strands is tetrafunctional — two of these are the azide and alkyne required for click chemistry, the third is unique to that particular strand, and the fourth allows the building block to be linked to a DNA strand (see Figure 3.4 for more information).

The functionalised strands are anchored outside of the hexagonal "reaction centre". The anchorage mechanism is shown for two of the functionalised staple strands. Each functionalised strand is held outside of the reaction centre with the aid of other modified staple strands (black). Typically, a portion of each functionalised strand is not anchored down and remains single-stranded. This domain forms a toehold for displacer strands to hybridise to and release the functionalised strands from their anchorages.

A strand displacement mechanism can be used to bring the strands into the reaction centre one-by-one, whereupon the click reaction can take place. The displacer strands consist of two domains; one domain is complementary to a specific functionalised strand (coloured), and the second is a sticky-end domain which is complementary to the sticky-ends of other displacer strands (black). Once functionalised strands have been released by the displacers, the sticky-ends of the displacer strands hybridise together and in doing so, bring the chemical building blocks into close proximity. The displacer strands may be removed after the chemical reaction between two functionalised strands has occurred, and new displacers to release other functionalised strands may be added to the system. By repeating this process, the six functionalised strands react in a controlled (programmed) way. The end product is a ring of the six chemical building blocks, a product which can be varied by altering the order in which the functionalised strands are released into the reaction centre.

the cascade, a mechanism which may not be entirely suited to DNA-templated chemistry. The second approach is to bring all the building blocks together in the same area on the origami. The functionalised DNA strands are arranged on the perimeter of a "reaction centre", such that it is possible for each strand to interact with any of the others. The minimum size of the reaction centre is governed by the arrangement of staple strands that form the origami, and resembles a hexagon with a modified strand at each vertex (see Figure 3.3). Each strand is anchored outside of the reaction centre at the start of the reaction, but by utilising a release mechanism one end of the strand is freed and able to diffuse in and out of the reaction centre. Potentially, the strands can be released in any order and therefore, for a six strand system loaded with six different building blocks there are a possible 720 ways of making a six-unit chemical product. This is a remarkable quantity, particularly since this templating architecture is located on a small segment of the full origami tile surface. Theoretically, an origami tile could have multiple and distinct reaction centres located on it, allowing for a large number of reactions to occur simultaneously on one tile.

In order to enable the reactions to occur in this way, each strand must be bifunctional, so that once it has undergone one reaction, it still has the capability to react with a second strand. In the case of click chemistry, the two functional groups required are an azide and an alkyne. To make use of the ability to control their sequence of reaction, the building blocks need to be distinguishable, such that each carries a unique third functionality. This is described in more detail in Section 3.2.1. It should be noted that the work presented in this chapter is a collaborative effort, the details of which are provided in Appendix B.5.1.

The cascade-coupling design presented in Section 2.1 had a clear resemblance to the natural ribosome, but the similarities between the origami-templated chemistry idea and the ribosome are less obvious. The natural ribosome is a mechanism which

reads the information encoded within an mRNA template and builds a polypeptide chain based on the sequence of instructions in the template. The template in the origami system is the long ssDNA strand which forms the basis of the origami tile. Each staple strand which hybridises to the template does so by forming base pairs with three specific sequence segments on the template (as described in Section 1.2). These sequence segments can be viewed as analogous to the codons of the natural ribosome, as each group of three segments encodes for one particular staple strand and therefore, the staple strands are the transfer RNA (tRNA), which are attached to chemical building blocks and consist of anticodon sequences. Assembly of the origami positions these synthetic tRNAs into the structure as per the instructions encoded in the template DNA strand. This is no different to the cascade-coupling system, where the functionalised target strands were already lined-up in the order specified by the template before the template sequence was 'read'. However, in the origami system, the template is not linear (and the building blocks are not lined-up one after another) and so the order of the chemical reactions is less closely linked to the template strand sequence. Instead, the order in which the chemical building blocks react is dependent on the manual addition of the displacer strands that facilitate the reactions. There may be scope for the origami idea to be developed into an autonomous process like the cascade-coupling mechanism, but at present the system is simply being tested for feasibility.

3.2.1 Tetrafunctional Click Adaptor Design

The building blocks used for the origami-templated click chemistry need to fulfill certain requirements. Firstly, these click "adaptor" molecules must carry the functional groups of the click reaction — an azide and an alkyne group. Secondly, the adaptor must be attached to a DNA strand, and thirdly, the adaptor must carry a unique functionality. Thus, the adaptor needs to have four functionalities. Rather than trying to synthesise this complicated molecule from scratch, it was decided that amino

acids could be used as the starting point for the synthesis.

Amino acids are an appropriate choice for the adaptor design for a number of reasons. The molecules contain an amine group, a carboxylic acid group and a side chain R that varies between different amino acids (see Figure 3.4). The type of reaction which this side chain can participate in is determined by its functional group, which naturally allows for a unique functionality to be incorporated into the adaptor design. The amine and carboxylic acid groups provide the amino acid with the ability to undergo most of the reactions that are associated with the functional groups, such as amide bond formation (for the amine group) or esterification (for the carboxylic acid group). Using amino acids as the basis for the building blocks is also a neat choice because the origami-templated chemical reactions will be guided by the order in which different building blocks (made from different amino acids) are released. The amino acids therefore determine what chemical product is formed, in much the same way as a protein is built in nature from a specific sequence of interactions between amino acids.

The work presented here uses bifunctional click adaptors rather than tetrafunctional ones to test the origami-templated click chemistry mechanism. These use one of the side chains of an amino acid to link the building block to a DNA strand, and another to incorporate either an azide or an alkyne. The azide-functional DNA was prepared using 2-azidoacetic acid molecules, while the alkyne-functional DNA was prepared using 4-pentynoic acid. In both cases, the free carboxylic group of the acid was activated as an N-hydroysuccinimide (NHS) ester, and this was used in an amide-coupling reaction to attach amine-modified DNA (Integrated DNA Technologies) to the small molecules. The details of how the DNA adaptor strands were prepared are provided in Appendix B.5.2.

Since each building block contains both azides and alkynes, unless chemical pro-

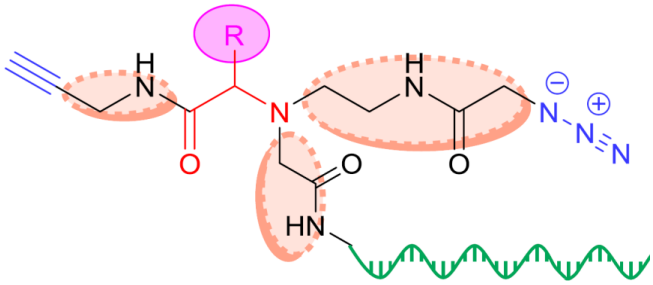


Figure 3.4: Schematic of the four functionalities of the click adaptor.
The click adaptors are required to be tetrafunctional, and are therefore complicated molecules to synthesise. To simplify the synthesis, amino acids are used as the starting material. Of the twenty natural amino acids, five have the required number of side chains to incorporate the four functional groups. Two of these functional groups are necessarily the azide and alkyne required for click chemistry. These are shown in blue on either side of the amino acid. The third functionality is used to attach the molecule to the end of a (amine-modified) DNA strand. The fourth side chain in R is used to provide the building block with a unique functionality, as this side chain varies between amino acids. Figure courtesy of D. Khan.

tecting groups are introduced into the design, the building blocks can react out of sequence and can form smaller cyclic products (as shown in Figure 3.5). To prevent this, one of the building blocks must begin with one of its functionalities, for instance the azide, protected. This "start" strand will always be released first to initiate the reaction. The "start" strand can be deprotected after the final strand has been released, so that a closed ring can be formed. (This reduces the number of six-unit outcomes from 6! to 5! (120)).

The second issue with this building block design is that the chemical product needs to be analysed at the end of the reaction by mass spectrometry or otherwise. As described so far, all the building blocks remain linked to their DNA staple strands throughout the reaction. It may be necessary to use an attachment linker that is easily broken (perhaps by using a photocleavable bond) to release all but one of the building blocks from the DNA. The product is therefore attached to one staple strand

and, by using a strand displacement mechanism or otherwise, the staple strand and the chemical product could be stripped away from the origami tile for analysis.

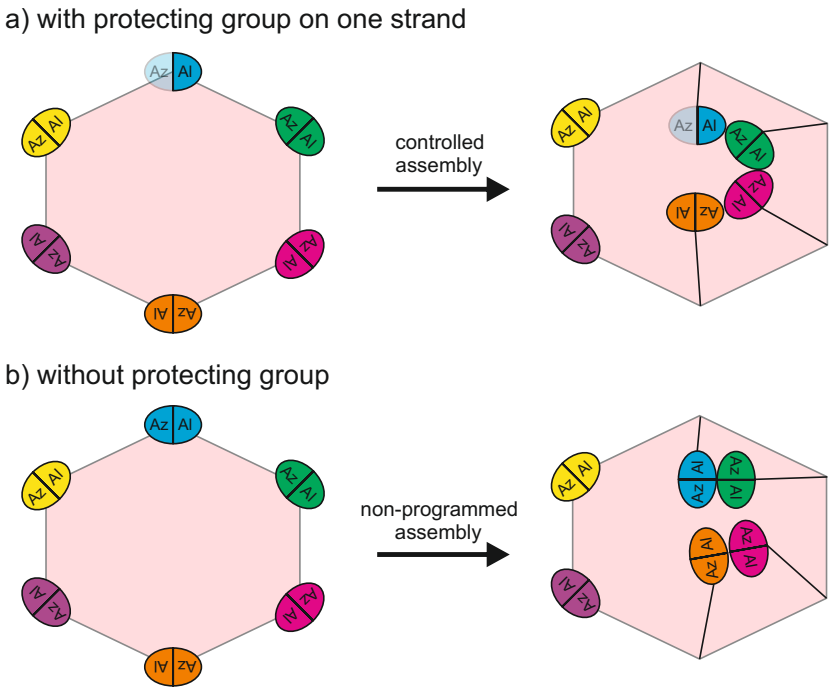


Figure 3.5: Diagram illustrating the benefit of using a protecting group on one of the functionalities of the "start" strand.
If the building blocks are released into the reaction centre without using any protecting groups, it is possible that undesired smaller chemical products are formed instead of a six-unit product (bottom right). Although the functionalised strands are released one-by-one, there is no control over how they interact with each other. This is remedied if one building block is always the first to be released (the "start" strand). By adding a protecting group to this building block (represented here by the transparent blue azide), it is possible to control the click reaction. The functionality can be deprotected at the end of the reaction to enable the formation of a six-unit ring.

3.2.2 Initial Feasibility Tests

Before carrying out click chemistry on origami, a simple experiment was designed to test the reaction in solution (off-origami). As described in Section 1.5.3, the copper-catalysed click reaction is a cycloaddition reaction which covalently links together the azide and alkyne reacting groups (Figure 3.6). The initial experiments were conducted

to determine various experimental parameters, such as the typical amount of copper catalyst required. To do this, two DNA "adaptor" strands were linked to bifunctional click building blocks (as described in the previous section, and in Appendix B.5.2). The off-origami click test mechanism is shown in Figure 3.7. Hybridisation of the two adaptor strands to a template strand brings the building blocks into close proximity. The distance between them was altered by introducing a single-stranded region on the template strand between the two adaptors. This was included to determine how sensitive the click reaction is to the distance between the building blocks. The azide adaptor strand was also fluorescently-labelled with a Fam fluorophore at the 3' end of the strand, to help visualisation of the starting material and the reaction product by PAGE analysis (see Appendix A.1.3).

Two tests were carried out, one where no catalyst was used, and one which used 50 equiv of the copper catalyst (plus the copper-binding ligand THPTA and reducing agent sodium ascorbate, as described in Section 1.5.3), relative to the DNA concentration (1 µM). The three strands were annealed together in equimolar amounts in 50 mM NaCl, which was degassed using argon for approximately 10 min prior to being used (to remove oxygen that could potentially react with the copper catalyst). Several studies on DNA-templated click chemistry describe using NaCl in their reactions, which acts only to stabilise the DNA duplex and does not participate in the reaction itself [Kumar et al., 2007], [Kočalka et al., 2008]. To anneal the strands, the solution was heated to 85 °C to melt all possible DNA structures, and then cooled to 20 °C over 10 min. In the catalysed reaction, the required copper-binding ligand THPTA and the reducing agent sodium ascorbate were added, respectively, in seven-fold excess and ten-fold excess with respect to the copper catalyst [Kočalka et al., 2008]. All solutions were degassed with argon prior to use (a detailed protocol is provided in Appendix B.5). The THPTA and sodium ascorbate were mixed together, and immediately before being mixed with the DNA, the copper was added and the solution

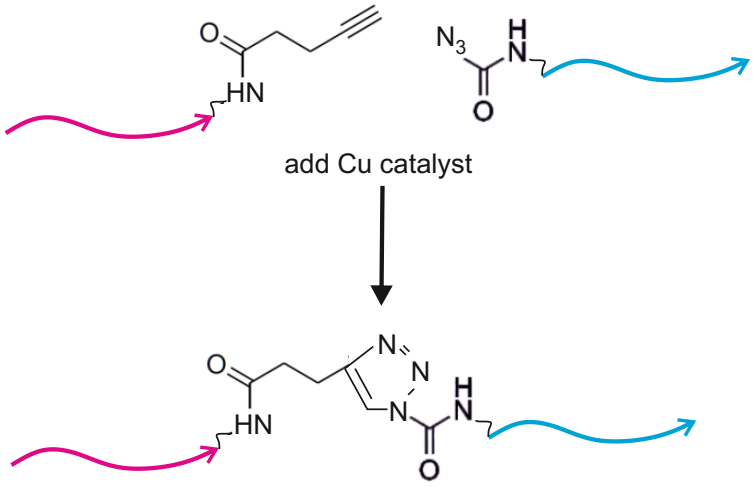


Figure 3.6: Diagram illustrating the click reaction.
The alkyne molecule attached to the magenta DNA strand and the azide molecule attached to the blue DNA strand form a covalent bond via the copper-catalysed Huisgen azide-alkyne cycloaddition reaction. In the presence of copper, reaction completion may be observed within 5 min.

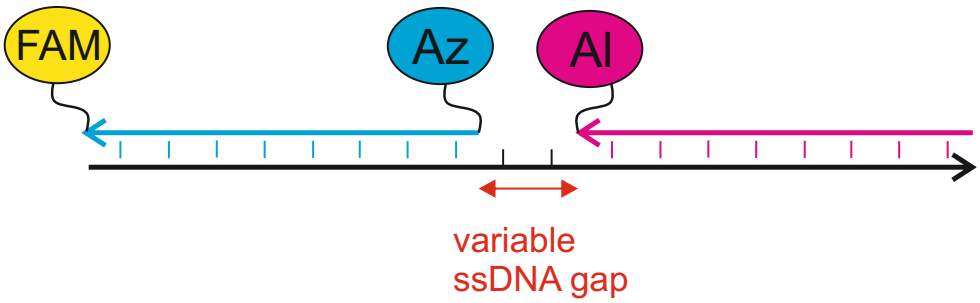


Figure 3.7: Off-origami test of the click chemistry with monofunctional building blocks.
Two DNA "adaptor" strands, each 24 nt in length, were click functionalised. The Al strand was linked to an alkyne on the 3' end of the oligomer, while the Az strand was linked to an azide on the 5' end and a fluorescent Fam molecule on the 3' end. The two adaptors form a duplex with a template strand, the length of which was varied between 48-56 nt to alter the size of the single-stranded gap between the functionalised ends of the adaptors. The gap was varied between 0–8 nt, to determine whether distance affected the click reaction. The azide adaptor was labelled with Fam to help distinguish between the click product and the template strand (which are of similar lengths) on a PAGE gel.

degassed once more. The reaction mixtures of both tests were incubated on a shaker for two hours at room temperature.

The samples were analysed by running a 15 % two-layered discontinuous super-denaturing PAGE gel for 45 min at 300 V in 1× Tris-Glycine at room temperature (see Appendix B.3 for details on the protocols and solutions required for the denaturing PAGE analysis). The gel was scanned using a BioRad Pharos FX Plus gel scanner at the Fam excitation and emission wavelengths (Ex488/Em530 nm). The gel was then stained with Sybr Gold nucleic acid stain (Invitrogen) and re-scanned using the Sybr Gold wavelengths (also Ex488/Em530 nm). The two scans are shown in Figures 3.8 and 3.9. The results confirm the need to use the copper catalyst to enable the click reaction to take place under these experimental conditions. With the catalyst, a band which migrates slowly through the gel is observed in lanes 4–12 on gel A of Figure 3.8 (highlighted in the red box), which corresponds to the 48 nt product. It is clear that a 0 nt gap does not yield as much of the product as when the gap size is increased. With a 0 nt gap the yield was merely 10 %, but with a larger gap, the yield was around 40 % (calculated directly from the gels, using the method outlined in Appendix B.4). With no catalyst, there is no slow band in any of the lanes of gel A in Figure 3.9, showing that the catalyst is required. To confirm that the catalysed reaction resulted in a 48 nt product and the uncatalysed reaction did not work, the samples that used a 1 nt gap between adaptors were analysed further by mass spectrometry (as described in Appendix B.6). The expected mass of the click product is equal to the sum of the masses of the two strands that form the product (15792.1 Da), and the results of the analysis of the catalysed reaction found a similarly sized product in the sample (see Figure 3.10). No mass corresponding to the product was found in uncatalysed sample. Thus, the initial feasibility tests have determined that it is possible to template the click reaction in this way, though the yield is low, and that the outcome can be analysed by mass spectrometry.

The click test was performed in a solution containing NaCl, however, origami structures are formed in a very specific buffer of 1× TAE, 12.5 mM $MgCl_2$. It has been found that monovalent cations such as NaCl prevent origami structures from forming correctly, and so the ratio of monovalent to divalent cations should be controlled [Dietz et al., 2009]. Since NaCl plays no role in the click reaction other than to stabilise the DNA duplexes, it is reasonable to assume the reaction will work in the origami buffer. To test this an experiment was carried out using the 0-, 1-, and 2 nt gap template strands. The same protocol was followed to perform the reaction except that the origami buffer replaced the NaCl solution. The results were analysed by running a 15 % super-denaturing PAGE gel. The scan of the Sybr Gold stained gel is shown in Figure 3.11. Lanes 4–6 contain bands which migrate at the same rate, and therefore correspond to the 48 nt click product. The template strands are also visible, and appear to run a little slower than the click product.

The yield of the click reaction is low despite a reaction time of 2 hours, which is surprising for a reaction that has been demonstrated by others to be high yielding and fast. It is possible that this partially stems from incorrect stoichiometry. Gel A in Figure 3.8 shows that there are unreacted Fam-labelled Az adaptor strands at the bottom of lanes 4–12. This may be a result of the click reaction not going to completion, but may also simply be because an excess of the Az adaptor was added through a miscalculation of the DNA concentration. To check this, the same click reaction was carried out as in the first experiment, but the Al adaptor was added in a two-fold excess with respect to the Az adaptor strand. Densitometric analysis of the results of the reaction, shown in lanes 7–9 of Figure 3.11, found the yield of the 1 nt gap reaction to be approximately 40 %, as before. Therefore, stoichiometry is not the reason behind the low yield of the reaction, and it is still not clear what the underlying cause is.

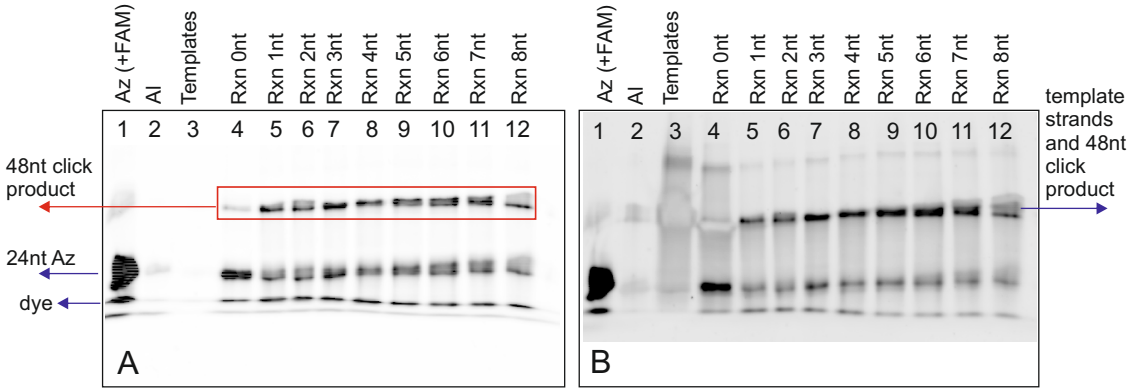


Figure 3.8: The 5'–3' click feasibility test using Cu.

(A) The gel was scanned for Fam fluorescence. The bands highlighted by the red box are fluorescent and migrate slower than the Az (FAM) strand, indicating that the reaction has taken place and produced a 48 nt product. However, it has not gone to completion as there is unreacted material in each lane.

(B) The Sybr Gold scan confirms that the click reaction has taken place, as the template strands can be observed as the bands migrating at the same rate as the click product (as they are comparable in length). The templates and product are better resolved in Figure 3.11.

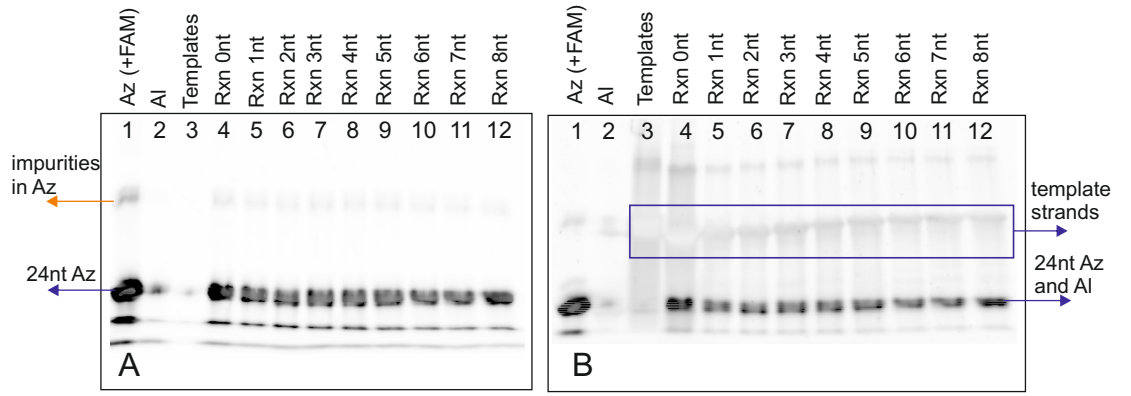
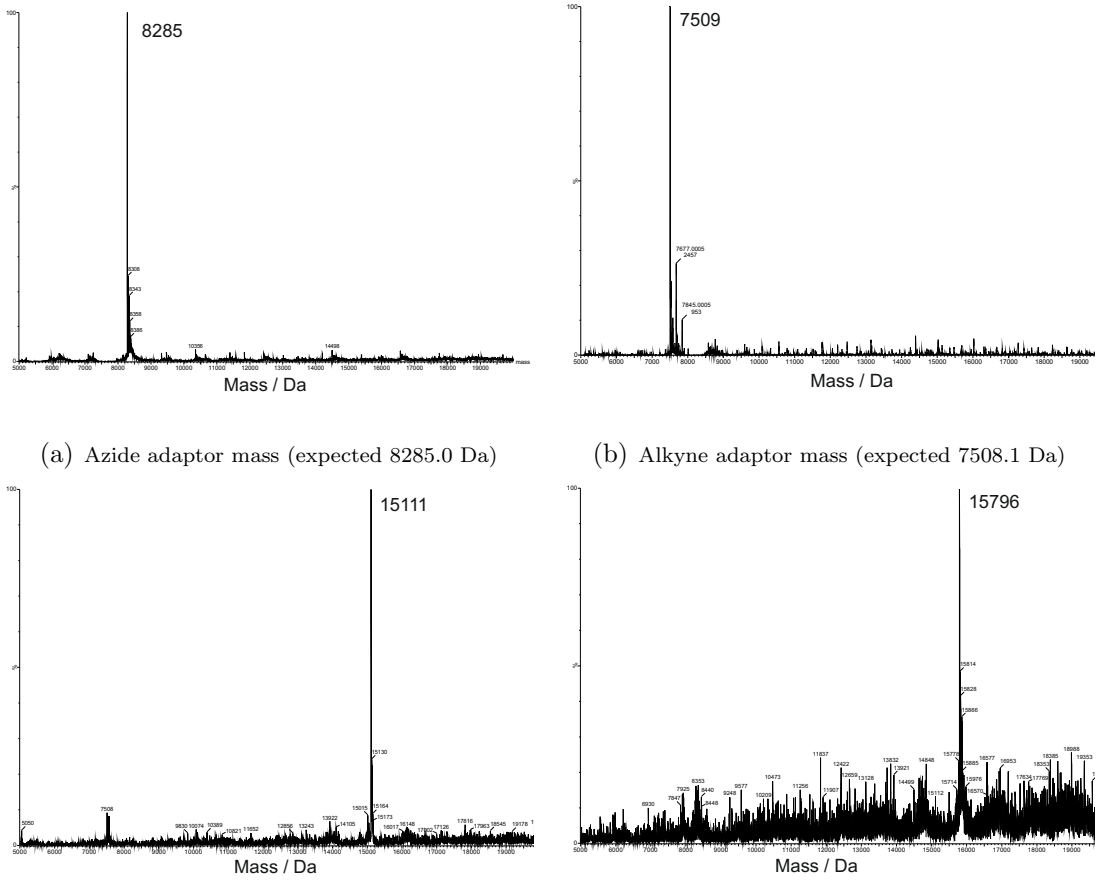


Figure 3.9: The 5'–3' click feasibility test without Cu.

(A) Unlike in Figure 3.8, there is only one fluorescent band in lanes 4–12, indicating the click reaction has not worked. The faint fluorescent band (indicated by the orange arrow) seems to correspond to impurities within the Az strand sample. These impurities can be seen in lane 1 in Figure 3.8 also.

(B) The Sybr Gold scan confirms this, as only the template strands (non-fluorescent) can be observed as the slow bands (highlighted in the blue box).



(a) Azide adaptor mass (expected 8285.0 Da)

(b) Alkyne adaptor mass (expected 7508.1 Da)

(c) 1 nt template strand mass (expected 15114.8 Da)

(d) Click product mass (expected 15792.1 Da)

Figure 3.10: Results of the mass spectrometry analysis indicate that the sample from the catalysed click reaction (using a template strand with a 1 nt gap) contained unreacted starting material, as well as a molecule with the same mass as that expected for the 48 nt click product.

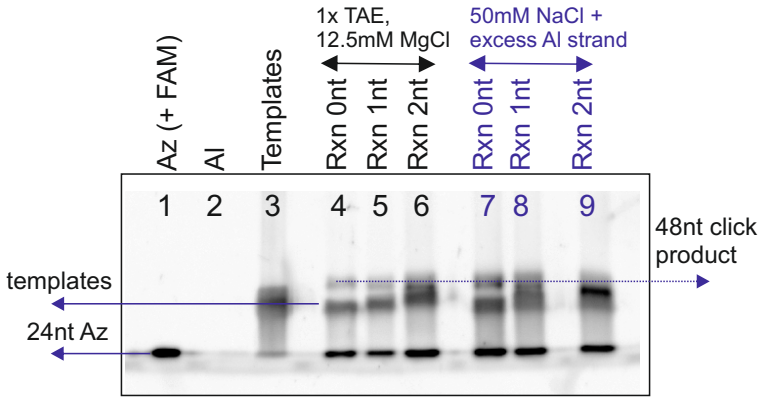


Figure 3.11: Performing the click reaction in origami buffer.

The click reaction was carried out using the origami buffer to determine whether the buffer has any effect on the outcome of the reaction. The results in lanes 4–6 show that the chemistry is unaffected by the buffer, though the yield was found to be slightly lower at approximately 30 %.

The yield of the click reaction is lower than expected for this robust chemistry. To determine whether the reaction can be pushed further, a two-fold excess of the Al strand was added to react with any unused Az adaptor. The results in lanes 7–9 do not suggest that the yield is a consequence of incorrect stoichiometry, as the yields found here match those of the earlier catalysed experiments.

3.3 5'–5' Click Chemistry

The initial feasibility experiments showed that the functionalised DNA strands could be used to perform click chemistry, although even when using a very simple DNA architecture to facilitate the reaction, the yield was less than 50 %. To test whether the click reaction could work when templated by origami, a simple origami tile was designed.

Origami Design Choices

DNA origami tiles have generally been formed using one of two template sources:

1. The full length (or a fraction of the length) of a long genomic ssDNA strand, such as M13mp18;

2. A double-stranded source, either custom synthesised or obtained by using the polymerase chain reaction (PCR)[1] to copy a specific section of a longer DNA strand.

The full origami tile (as designed by Rothemund) was not used as the substrate for this purpose. Instead of folding up the entire length of the M13mp18 virus DNA strand into an origami tile, only 960 nt were used to make a tile that was 32 nm by 30 nm (96 nt by 10 rows). A sketch of the tile is shown in Figure 3.12. (The remaining nucleotides of the scaffold strand remain unpaired but attached to the origami structure.) One reason behind using a segment of the scaffold strand was because this area is the minimum required to anchor down six adaptor strands outside a hexagonal reaction centre. The M13mp18 DNA strand is around 7000 nt in length, and only a fraction of strand was used to form an origami structure. It was assumed that this design choice would be suitable for a first attempt at performing origami-templated chemistry, and the effect of the unused segment of the template strand on the assembly of the origami and the chemical reactions was momentarily ignored.

However, it is possible to remove the unused DNA such that only the 960 nt of interest are used in the experiments. Restriction enzymes could be used to cut the ssDNA template, as each enzyme recognises a specific sequence of nucleotides and is able to cut the DNA at that site (restriction site). Several restriction enzymes were found to cut the M13mp18 DNA, but the experiment was first carried out with the full length of the template strand in order to quickly establish whether the origami-templated chemistry idea is feasible. Another method to remove the unused DNA is to use PCR to select, copy and amplify the sequence of interest. Double-stranded

[1]The polymerase chain reaction (PCR) is a technique to amplify a single copy of a DNA sequence, which can result in thousands of copies of the original sequence [Saiki et al., 1985]. Standard PCR methods are able to multiply DNA sequences that are up to 2000 nt in length, though methods to copy much longer sequences also exist [Barnes, 1994]. The technique is described further in Appendix B.9.

DNA has been used to construct origami [Högberg et al., 2009], [Pound et al., 2009], and requires following a modified annealing/assembly protocol. For a small excess of staple strand concentration, the duplex is more thermodynamically favourable than the formation of the origami structure. The electrostatic repulsion which arises from the close-packing of helices in the origami means that the synthesised duplex is more energetically favourable than the origami structure. Since the yield of origami structures created from dsDNA can be quite low, it was decided to first try the experiment with the full length of the M13mp18 strand, and then if the experiments are unsuccessful, to try using either PCR or restriction enzymes to remove unused nucleotides from the template.

All the origami experiments were carried out as bulk experiments in solution. It is possible for origami tiles in solution to become twisted, and this could affect the chemistry as the chemical building blocks may be brought into close proximity by the geometry of the tile. Furthermore, the building blocks on one tile may be able to react with the building blocks on another tile because the tiles are free in solution and may collide with each other. It was assumed that these effects would be minimal in comparison to the correct/designed reactions, and therefore, the reactions were carried out in solution. However, if the assumption is incorrect, it is possible to carry the reactions out on origami tiles that are flat and stuck down to a surface, as described in [Voigt et al., 2010]. Origami tiles can be deposited onto a mica surface and the reactions can be imaged by AFM. This does not allow fluorescence experiments to be conducted, which are useful in determining whether the strand displacements have occurred as designed. For this reason, all origami FRET experiments (to test strand displacement mechanisms) were carried out in solution.

Once folded into the tile structure, particular sections of the template strand sequence are linked to the chemical building blocks. These sections can be thought of as carrying information within their sequences about the particular chemical building blocks they are attached to. In nature, a gene exists on a segment of DNA that codes for a type of protein, and so here, the information-carrying sections of the origami template strand can be thought of as genes also. Therefore, a second reason for only using a segment of the scaffold strand is because 960 nt is approximately the right length for the PCR. By shuffling the genes on this segment of the template strand, and using PCR to amplify the result, it is possible to create many different versions of the original 960 nt sequence. Each version could encode for different chemical products, and so the origami-templated chemistry could give rise to a large number of different six-unit chemical products. When combined with the ability to generate 5! six-unit products from each tile, this method of chemical synthesis could produce large libraries of functional molecules (as discussed in Section 1.5). (However, due to the way the scaffold strand folds up to form the origami, shuffling domains on the DNA template becomes much more complicated than if the template were linear. Since the staple domains of adaptor strands each hybridise to three rows of the folded DNA template, shuffling genes means the corresponding sequences on the three rows have to be altered. Therefore, this idea has not been pursued here.)

The small origami tile design contains six DNA adaptor strands that are linked to the click building blocks on their 5' ends. The reason behind this choice was two-fold. Firstly, modifying the same end of each of the staple strands means that the extended regions of the staple strands automatically emerge from the origami on the same face of the tile. Secondly, this design means the maximum number (120) of chemical products can result from the interaction of the six strands, as the chemistry is independent of the polarity of the DNA adaptors (see Section 3.4 also). The mechanism to release the adaptors from their anchorages and to bring the functional

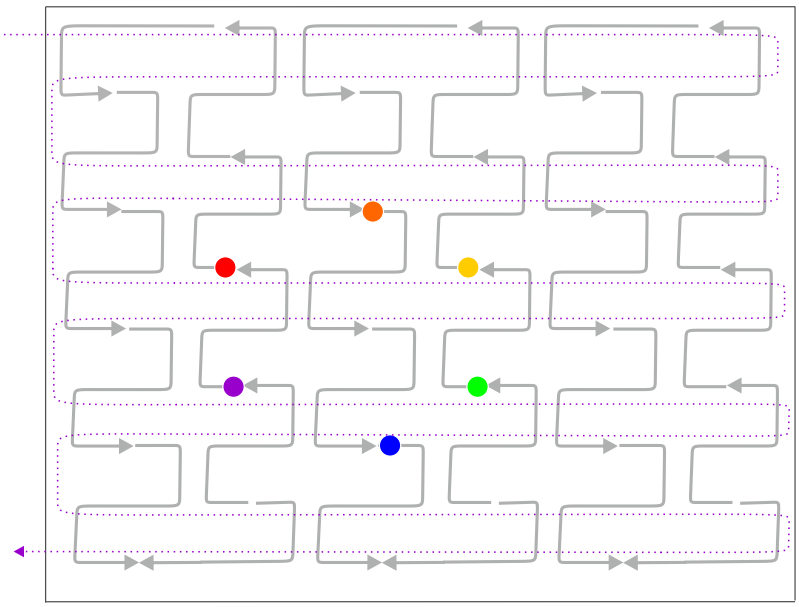


Figure 3.12: The small origami tile designed for the DNA-templated chemistry.
The tile was designed using NanoEngineer-1, a software package which enables the 3D simulation of DNA nanostructures [Sims, 2004]. Thirty staple strands (grey) are used to fold 960 nt of the M13mp18 virus strand (dotted line) into an origami tile. The unused nucleotides of the scaffold remain attached to the structure but unpaired. The position of the six staple strands linked to the chemical building blocks are marked by the coloured circles, which also highlight the perimeter of the reaction centre. Each adaptor is anchored outside of the reaction centre until required. The exact mechanism by which the strands are anchored and released is described in Section 3.3.1.

groups in close contact to enable a reaction is discussed in Section 3.3.1.

3.3.1 Strand Displacement Mechanism

As described above, six staple strands from the small origami tile were selected for 5' functionalisation with click building blocks. However, other staple strands also require modification, to help anchor the functionalised strands outside the reaction centre, and to facilitate their release when required for the chemical reaction. Figure 3.13 shows that two staple strands are involved in anchoring each of the adaptor strands outside the reaction centre. The adaptor strands consist of 18 nt reaction

domains (the length required to reach the middle of the reaction centre), which are further divided into toehold (T: 8 nt) and anchor (A: 10 nt) subdomains. The anchor subdomain is hybridised to the "fixer" staple strand (F). The toehold subdomain is unpaired until the start of the reaction. The remainder of strand F is held in place by an "extended" staple strand (E), so that it too is anchored at this position on the tile (although since the anchorage domain of strand F is not long enough when single-stranded to be able to interfere with the chemistry in the reaction centre, it may not be necessary to use the E strand at all).

To release an adaptor from its anchorage, a displacer strand is added to the system, usually in a 10-fold excess with respect to the origami concentration. The displacer is fully complementary to the reaction domain of the adaptor strand, but also contains a 3' sticky-end domain (shown in black in the diagram). The sequences of the toeholds (T) of the adaptor strands are all unique, and serve as an identifying address. Each sequence identifies the specific building block attached to the adaptor strand and therefore, each displacer strand is designed to interact with only one of the adaptor strands in the system. A displacer hybridises to the toehold of an adaptor strand to initiate a strand-displacement reaction that frees the adaptor from the fixer strand. The released adaptor is now able to move about its tether and the 5' end can diffuse into the reaction centre. In order to facilitate the chemical reaction between adaptors, consecutive displacer strands are given complementary sticky-end domains to enable them to interact. If two adaptors are in the reaction centre at the same time, the complementary sticky-end domains of the displacer strands are able to hybridise, and in doing so, bring the chemical building blocks into close proximity. As shown by Figures 3.13 and 3.3, the system is designed to bring two adaptors together, but it is easily modifiable to bring further adaptors together. By including a toehold domain on the 5' end of the displacer strands, it is possible to strip away one of the displacers from the system once the chemical reaction has completed. The displacer

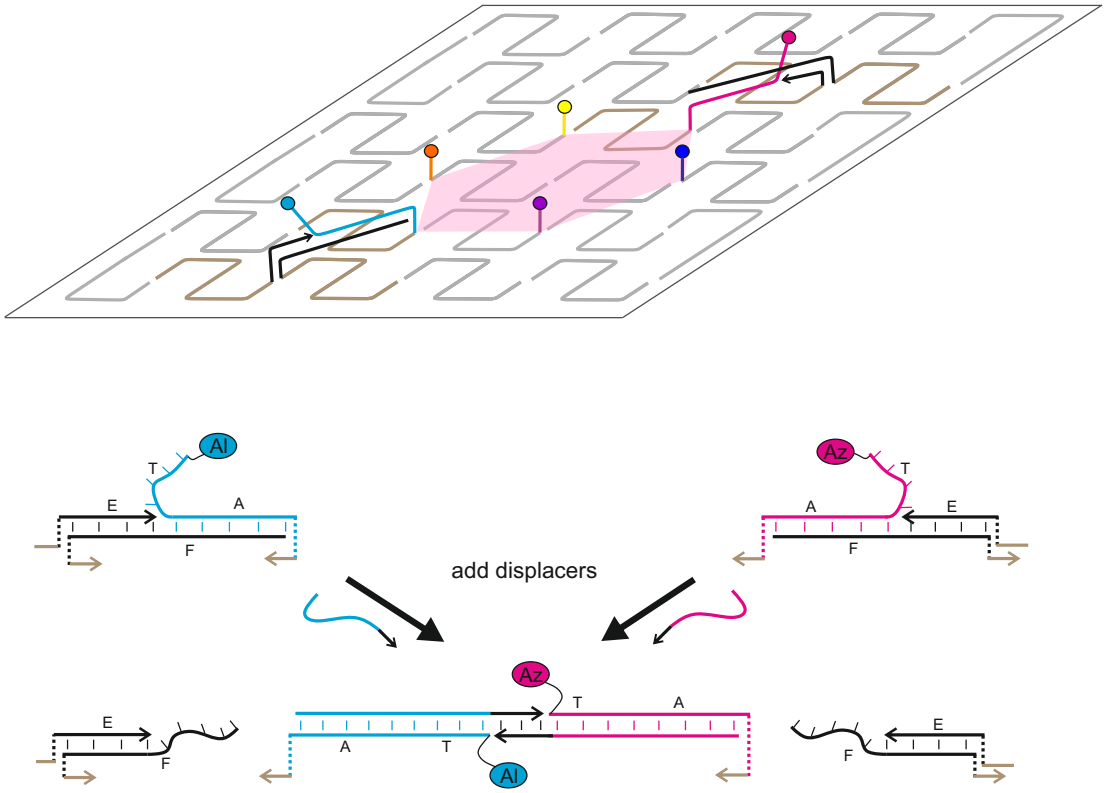


Figure 3.13: Illustration of the toehold-mediated strand displacement mechanism to bring together the click building blocks.

Note: here, each short thin line represents 2 nt, which is a break from the convention outlined in Section 1.6.1. Each adaptor strand (A) has an 18 nt domain (blue or magenta) which emerges from the origami tile (brown lines represent domains of staple strands that are hybridised to the origami template strand). Of this, 10 nt are hybridised to the fixer strand (F), while 8 nt remain unhybridised and form the toehold (T) for the strand displacement process. The chemical building blocks are linked to the 5' ends of the adaptors, by means of an amino group attached to a C6 spacer arm (shown as the black, thin and wavy lines). F strands are 18 nt in length, and the remaining nucleotides are paired with the extended staple strands (E), which serve to hold the fixer in place during the reaction.

Displacer strands are fully complementary to the 18 nt of the adaptor strands, but also contain sticky-end domains (black) which are complementary to each other. Toehold-mediated displacement releases the adaptor strand from its anchorage site, and enables it to migrate into the reaction centre. The hybridisation of the sticky-ends of the displacers bring two adaptors together, and the building blocks into close proximity.

still hybridised now has an unpaired sticky-end, which is free to interact with further adaptor-displacer duplexes.

In designing the sticky-end interaction of the displacer strands, the geometry of the DNA duplex was somewhat ignored. Though the mechanism drawn in Figure 3.13 shows the chemical building blocks being brought closer together through the sticky-end hybridisation, the diagram does not make clear that the helical twist of the DNA duplex may position the functionalised ends of the adaptor strands up to 180 ° apart (that is, at diametrically opposite points across the width of the DNA). However, it is not simple to take this property of the DNA molecule into account. The minimum angular separation between the building blocks will either be when the building blocks are linearly separated by 1 nt, in which case the sticky-end interaction would be improbable, or when they are separated by a full turn of the DNA helix, in which case the linear distance between the building blocks has increased signficantly. In the latter scenario, this linear separation may be too long a distance for the reaction to occur over, but this could be accommodated for by attaching the chemical building blocks to the DNA adaptor strands by longer and more flexible linkers (rather than the standard C6 linker). For a first test of the strand displacement mechanism on origami, these geometrical considerations were ignored. FRET experiments were conducted to determine whether the mechanism worked, with the view that adjustments could be made to the design to take the geometry into account later.

3.3.2 Testing the Release Mechanism using FRET

Prior to carrying out the click reaction on origami, the mechanism to release adaptor strands from their anchorages was tested using FRET (see Appendix A.1.1 for a description of the phenomenon). The simplest test of the system involves monitoring the interaction of two adaptor strands. The adaptors were purchased with either a 5' Tamra (strand A1) or a 5' Fam (strand A2) fluorophore. In this FRET pair, the Fam

fluorophore is the donor and the Tamra fluorophore is the acceptor. When the origami is formed with these adaptors, the large distance between the two fluorophores means that there is no FRET occuring between them, and so the donor fluorescence (Fam) should be high. However, once the adaptors are brought together by the displacer strands, FRET can take place and the donor fluorescence should decrease while the acceptor fluorescence increases.

Before carrying out the FRET experiments, it was necessary to determine whether the scaffold strand had folded into the origami structure. Origami tiles were formed using a 10× excess of the staple strands to the scaffold strand, and purified before use to remove unhybridised staples (using the protocol described in Appendix B.2). However, no origami tiles were observed by AFM. This could be a result of the large amount of unfolded, single-stranded M13mp18 that remains attached to the ∼32 nm × 30 nm origami tile. Since only around one-seventh of the scaffold strand is used to make the origami, it is possible that the remainder of the strand obscures the tile under AFM. The relatively short single-stranded region of the scaffold strand of a full-sized origami tile is clearly observable under AFM (shown in the Supplementary Information of [Rothemund, 2006]), and so it is reasonable to assume that the much larger ssDNA here can entirely obscure the small tile.

Therefore, to determine whether origami had been formed, samples were analysed by running a 0.75 % agarose gel (see Appendix B.8 for details on making the gel). Origami tiles were formed using a 10× excess of the staple strands to the scaffold strand, except in the case of the fluorescent adaptors, which were added in 3× excess. This is to reduce the amount of the unhybridised fluorophore strands in solution, which can overwhelm the fluorescence signal from the actual reaction. Origami samples are purified before use to remove excess staples, and there is more chance of removing the fluorophore strands if there are fewer of them free in solution to begin

with. Three different origami samples were made, one using 30 unmodified staple strands ("simple"), one using the six staple strands required for the fluorescence experiments ("A1 & A2"), and one in which the modified staples for the six possible adaptor strands were used ("A1-6") (DNA sequences are listed in Appendix B.1.1). The concentration of all samples was 10 nM in origami buffer, and the samples were used unpurified for the gel. The gel was loaded with the samples and run for 120 min at 70 V in a cold room (4 °C) in 1× TAE buffer. The gel was scanned at the Fam and Tamra wavelengths (Ex488/Em530 nm and Ex532/Em605 nm respectively), and then stained with Sybr Gold nucleic acid stain and scanned at the Sybr Gold wavelengths (Ex488/Em530 nm). Figure 3.14 shows the Sybr Gold scan of the gel and a multichannel image of the fluorescence scans. The gel suggests that the staple strands have hybridised to the scaffold, as there is a small shift between the single-stranded M13mp18 DNA (lanes 3 and 12), and the three origami samples (lanes 8-10). The shift is small, as only one-seventh of the scaffold strand has formed the origami, and so the overall difference in mobility between that and the ssDNA scaffold is relatively small. Furthermore, scanning the gel at the Fam and Tamra wavelengths determined that the fluorescent adaptor strands used in the "A1 & A2" sample have hybridised to the scaffold.

It is possible however, that the staple strands have hybridised to the scaffold strand but have not actually folded the scaffold into the designed tile structure. Moreover, if the origami has formed correctly, it is possible that the remainder of the scaffold could interfere with the strand displacement reaction, as highlighted in the origami design choices discussion in Section 3.3. Thus, it is beneficial to remove the unused portion of the scaffold strand, and one simple way to do this is to use the polymerase chain reaction to copy and amplify the section of the scaffold strand used for the origami, and use this shorter DNA to form the tile (see Appendix B.9 for information on how the reaction works). After performing PCR, the DNA polymerase required

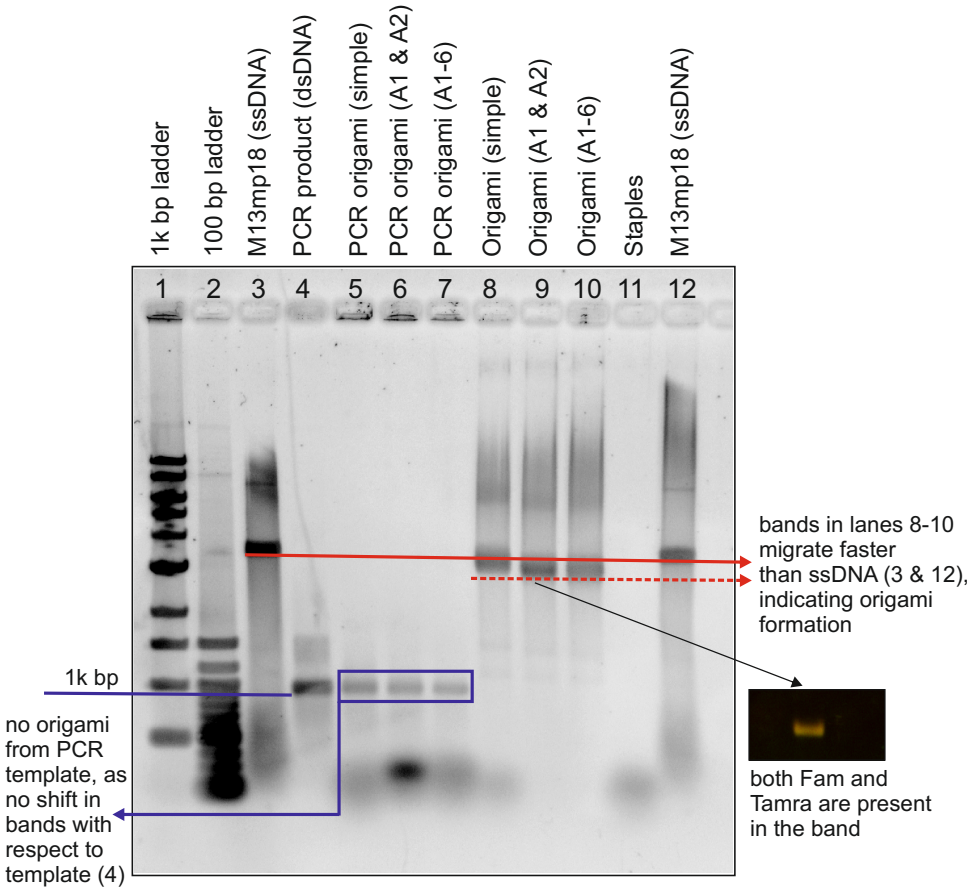


Figure 3.14: 0.75 % agarose gel to verify the formation of the 960 nt origami tile.

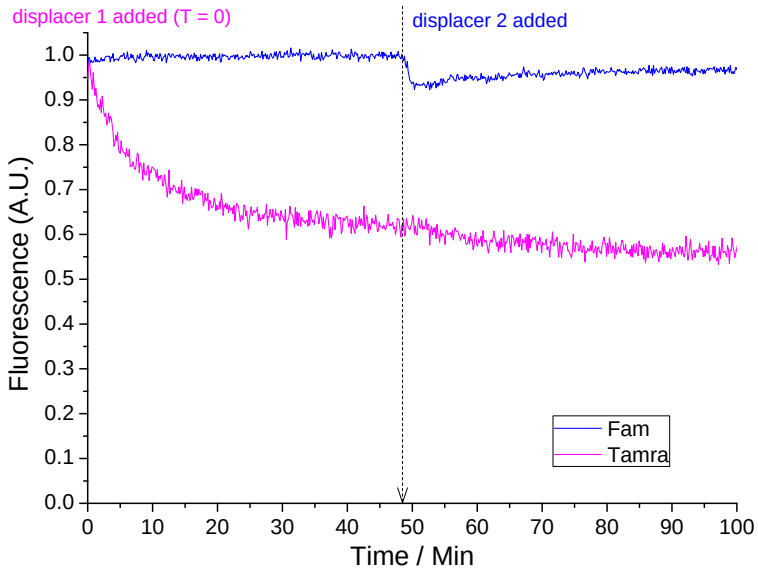
All samples were loaded in 10 nM concentrations. Two different template sources have been used to make the origami tile. Firstly, 960 nt of the M13mp18 strand were folded into an origami tile, without removing the unused nucleotides from the scaffold strand. The unfolded M13mp18 strand is in lanes 3 and 12, while the origami samples using this scaffold are in lanes 8–10. These three bands migrate slightly faster than the unfolded M13mp18 (as indicated by the dashed red arrow). Predictably, the gel shift is small, as the majority of the scaffold strand is unfolded to make the origami, and so the difference between an unfolded and a one-seventh folded scaffold is small. Fluorescence was also observed in the origami sample formed using the fluorescent adaptor strands A1 and A2 (lane 9), suggesting that the staple strands have hybridised to the scaffold.

Secondly, the 960 nt sequence of the M13mp18 was copied and amplified by PCR to produce a dsDNA template source for the origami. Lane 4 shows the ∼1000 base pair duplex DNA created by performing PCR. This dsDNA was used to form origami tiles (see main text for details), but the samples run in lanes 5–7 show that origami structures did not form using this PCR product as a scaffold (highlighted in the blue box). There is no shift in these three bands compared to the template band in lane 4, no fluorescence was observed in the bands, and all the unused staple strands are clearly visible at the bottom of lanes 5–7.

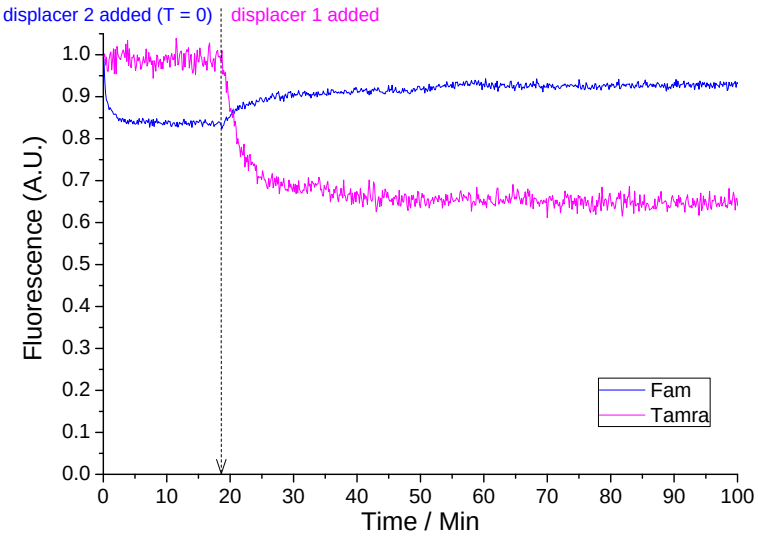
for the reaction was removed using phenol-chloroform extraction (described in Appendix B.10), and the duplex used to form origami. The staple strands were added to the synthesised duplex in 50-fold excess to encourage folding as the PCR duplex is more energetically favourable than the origami structure (fluorophore-labelled strands were still added in 3× excess), but now the samples were heated to 99 °C for 10 min to melt the DNA, rapidly cooled to 51 °C and then cooled to 21 °C at a rate of 0.1 °C per minute (as detailed in [Högberg et al., 2009]). Three samples were made in this way from dsDNA ("simple", "A1 & A2" and "A1-6"), and run on an agarose gel to determine correct formation. Lanes 5–7 of Figure 3.14 show that while a ∼960 nt duplex was produced by PCR, origami structures did not form. The unused staples are clearly visible at the bottom of the lanes, and no fluorescence was observed anywhere in lane 6 other than in the staple band. Various other techniques to pull apart the PCR duplex to form origami were attempted, such as using biotin-streptavidin extraction as described by [Pound et al., 2009], but origami structures were never observed using the synthesised DNA as the scaffold.

However, it can be inferred from Figure 3.14 that origami structures have formed using the full length of the M13mp18 strand and therefore, it is necessary to test the strand displacement mechanism (shown in Figure 3.13). Tiles were purified by spin-column purification (Appendix B.2) and used immediately in two fluorescence spectroscopy tests. Samples were made at 100 nM, though the concentration may have changed as a result of the purification process. Two small-volume cuvettes were used for the fluorescence spectroscopy experiments, containing 40 µl of the origami sample and 100 µl origami buffer. A layer of mineral oil (∼30 µl) was added to the cuvette to minimise evaporation during the experiment. The tests were carried out at 21 °C.

After obtaining a steady fluorescence signal, displacer 1 was added to the first



(a) Strand A1 (Tamra) is released first, then A2 (Fam).



(b) Strand A2 (Fam) is released first, then A1 (Tamra).

Figure 3.15: The fluorescence experiments to test the strand displacement mechanism.
(a) Upon addition of displacer 1, the Tamra-labelled adaptor A1 is released, resulting in a conformation change in the DNA and therefore a change in fluorescence. After 50 min, displacer 2 was added. This caused a further decrease in the Tamra fluorescence, and a drop (attributed to the DNA conformation change) and subsequent increase (due to the interaction between adaptors) of the Fam fluorescence.
(b) Similarly, when the displacers are added in the opposite order, an increase in the Fam fluorescence is observed, which appears to be a result of the two adaptors interacting.

cuvette and displacer 2 added to the second cuvette. Displacer 1 is used to release adaptor A1 (Tamra-labelled) from its anchorage, while displacer 2 releases adaptor A2 (Fam-labelled). Both displacers have 8 nt sticky-end domains that are complementary to each other. Figure 3.15 shows that the addition of the first displacer strand to each cuvette causes a change in the fluorescence emission of the fluorophore attached to the released adaptor in each case. The change in Tamra signal upon addition of displacer 1, and Fam signal from the addition of displacer 2, is a result of the conformational change of the adaptor DNA after hybridising to the displacer. The single-stranded toehold region of the adaptor adjacent to the fluorophore is now double-stranded, and this in turn alters the fluorescence of the attached fluorophore. In the case of strand A1, the change is large. In each sample, the fluorescence from the unreleased fluoroscent adaptor is unchanged. After the signals levelled off, the second displacer strand was added to each cuvette. The result in both cases is an increase in Fam fluorescence, which presumably results from the interaction of the two adaptor strands. In each reaction mixture, whatever the order in which the displacers are added, the Fam signal increases and the Tamra signal decreases somewhat (which is easier to see in Figure 3.15(a)). An increase in Fam fluorescence suggests that the (donor) fluorophore has become less quenched, which opposes the original hypothesis that the adaptor strand interaction would result in increased quenching of Fam by the nearby Tamra fluorophore. It is possible that the way in which the two adaptor strands are linked together results in the unexpected change in fluorescence. It appears that the fluorophores are interacting, but in a way that is not easily explained by FRET alone. Therefore, before carrying out the chemistry on origami, some further tests of the sticky-end interaction were necessary.

3.3.3 Click Off-Origami

In order to check that the sticky-end interaction of the displacer strands brings together the adaptor strands, the two strands were used in off-origami FRET exper-

iments. The 8 nt sticky-ends are long enough to enable the adaptors to come together in solution, as shown by Figure 3.16. Initially, two samples in the fluorescence spectrometer contained only one or other of the fluorescent adaptor strands. Upon obtaining steady fluorescence signals, the second adaptor was added to each cuvette and the fluorescence observed (data not shown here). No change in fluorescence was observed as arising from an interaction between the two fluorophores. The displacer strands were then added one after the other, as in the origami-templated reaction described earlier. The fluorescence signals changed in exactly the same way as when the adaptors were anchored down on the origami. This suggests that the sticky-end interaction does bring together the two adaptor strands, and that the fluorescence data obtained from the origami tests is not anomalous. The reason why in both cases the Fam fluorescence increases upon being brought into contact with the Tamra fluorophore is not clear.

The off-origami fluorescence experiments suggest that the sticky-end interaction brings the adaptor strands together as designed and therefore, the architecture was used to perform click chemistry. The 5' fluorophores of the adaptor strands were replaced with bifunctional click building blocks (as shown by Figure 3.16). Strands were annealed together in equimolar quantities in a 50 mM NaCl solution, and the usual click reaction protocol was followed (see Appendix B.5.3). The reaction was carried out for 2 hours using a 50-fold excess of the copper catalyst with respect to the DNA concentration (1 µM). The samples were analysed by running a 15 % super-denaturing PAGE gel, for 60 min at 300 V in 1× Tris-Glycine buffer, at room temperature. The gel was scanned for Fam, stained with Sybr Gold and rescanned at the Sybr Gold excitation and emission wavelengths. The results presented in Figure 3.16 clearly show that the click chemistry has not worked off-origami. No slowly migrating band corresponding to a 102 nt product was observed on the gel. This implies that although the sticky-end interaction brings together the adaptor strands, the length of

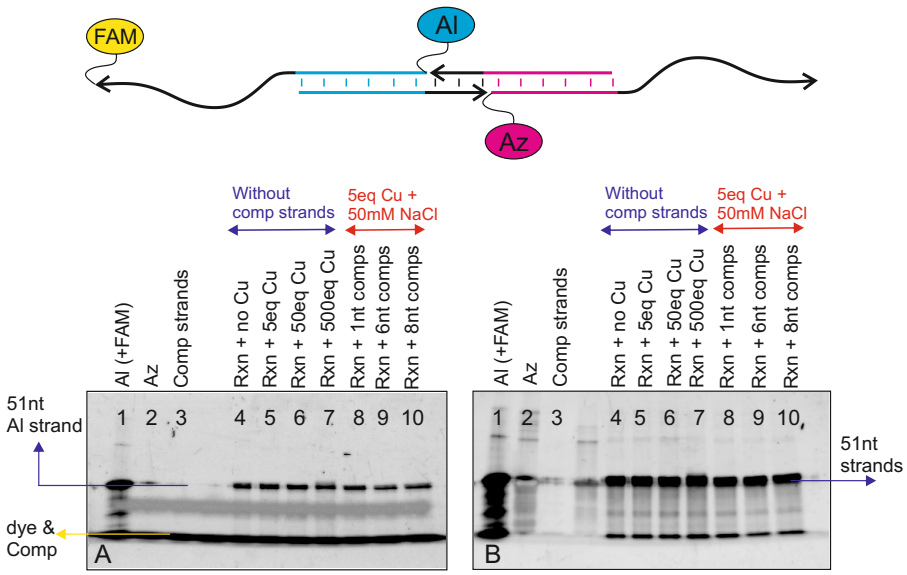


Figure 3.16: The 5'–5' click test in solution.

(A) Three different sticky-end lengths were used to bring the 51 nt adaptor strands together — 1-, 6-, and 8 nt. Here, a 6 nt sticky-end reaction is shown. It is highly improbable that a 1 base pair duplex is stable, but was used to determine if even an unstable, short duplex may be better at templating the reaction than a longer and more stable sticky-end (while also factoring in the geometrical issues discussed in Section 3.3.1). A 15 % super-denaturing PAGE gel was used to analyse the reactions. The Fam scan of the gel (Ex488/Em530 nm) shows that in lanes 4–10 only one band contains the Fam fluorophore, which migrates at the same rate as the adaptor strand in lane 1. Thus, it can be assumed the click reaction has not worked.

(B) The Sybr Gold scan confirms that the click reaction has not taken place, as there are no slowly migrating strands present in any of the lanes.

the sticky-ends and the twist of the duplex actually place the building blocks too far apart to react with each other. FRET can occur at a distance of several nanometres, but the building blocks cannot reach each other across this distance. Based on these experiments, there is no evidence that the click chemistry would work on origami using this particular mechanism to bring together the adaptor strands. Decreasing the sticky-end length, with the hope of decreasing the distance between the building blocks, would only decrease the stability of the sticky-end interaction and prevent the adaptors from interacting. Thus, a redesign is necessary.

Two options exist to remedy the problem. Firstly, it could be possible to use longer spacers to attach the chemical building blocks to the DNA adaptor strands. With longer, more flexible spacers (such as a C12 spacer arm), it may be possible for the building blocks to come into close proximity despite the distance (linear and angular) between them. A second, slightly simpler idea to implement, is to link together (chemically or otherwise) the 3' ends of the displacer strands, such that the need for sticky-ends is removed entirely. Integrated DNA Technologies is able to synthesise custom, non-catalogue DNA strands, such as oligomers containing inverted bases. A 36 nt DNA linker strand was designed to switch from ordinary bases to inverted bases midway along the length of the strand, such that half of the strand was complementary to the azide adaptor and half to the alkyne adaptor, as shown in Figure 3.17. The linker strand was used in a click reaction, with and without the aid of the copper catalyst. The two adaptors and the linker strand were annealed together in equimolar quantities in 50 mM NaCl, and the usual protocol for the reaction was followed (Appendix B.5.3). Samples were analysed via a 15 % super-denaturing PAGE gel, which was scanned at both the Fam and Sybr Gold wavelengths. The results are shown in Figure 3.17. It is clear from lane 7 of the Fam-scanned gel (A) that a longer DNA product has been formed by using the linker strand and the copper catalyst. The yield of the click reaction is again not as high as expected for click chemistry (25 % calculated directly from the gels using the method described in Appendix B.4), but the reaction has taken place. Mass spectrometry was used to confirm the creation of the 102 nt oligomer, but the results were inconclusive, possibly due to the low concentration of the product. Masses corresponding to the adaptor strands and the linker DNA were found in the sample, but a mass corresponding to the click product was not obtained. These results all confirm the need to redesign the origami anchorage architecture to bring the building blocks as close to each other as possible, and to create a system more amenable to analysis by mass spectrometry.

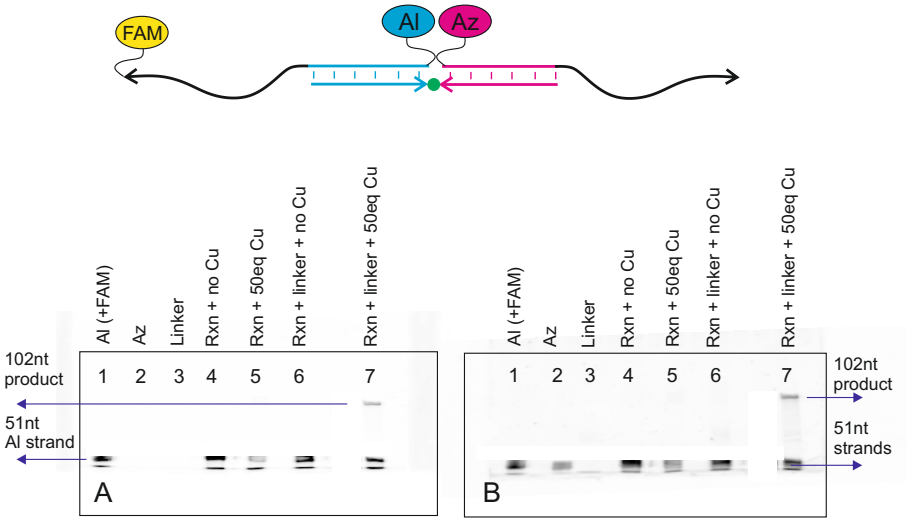


Figure 3.17: The 5'–5' click test in solution, with and without the linker DNA.
A 36 nt DNA linker was used to bring together the two adaptor strands. The linker is complementary to both adaptors since half of the strand was synthesised using inverted bases, such that the linker resembles two DNA strands joined together by their 3' ends.
(A) Reactions were carried out for 2 hours at room temperature, and samples analysed by a 15 % super-denaturing gel. The gel was scanned for Fam fluorescence, and it is clear that the reaction which utilised both linker DNA and copper catalyst have yielded a 102 nt product (lane 7).
(B) The Sybr Gold stained gel confirms that the reaction has occurred and by performing densitometric analysis, it was determined that the yield of the reaction was 25 %.

3.3.4 Conclusion

Many important conclusions can be drawn from analysing the experiments discussed so far. Firstly, whilst the idea behind forming an origami tile with only a fraction of the long M13mp18 scaffold strand is reasonable for an initial attempt at testing the origami-templated chemistry idea, in practice there are many drawbacks. Without obtaining an AFM image of the tile, there is no conclusive evidence that the origami structure formed correctly. Although the agarose gel analysis indicates that staple strands have hybridised to the scaffold, it does not confirm that they have hybridised according to the design. There are several solutions to this problem (as highlighted in the design choices discussion in Section 3.3), such as persevering with the PCR

technique to form a scaffold strand of the desired length, or using enzymes to nick the M13mp18 scaffold into shorter segments so that the unused nucleotides do not obscure the tile under AFM. However, it may simply be easier to use the full origami rectangle as the template for the chemistry. Secondly, FRET is an extremely useful tool for monitoring reactions such as strand displacements, but the results cannot necessarily be used as an indication of how the chemistry will work when the fluorophores are replaced by chemical building blocks. Whilst the fluorescence experiments suggested that the sticky-end interaction of the displacer strands brings the fluorophores into close proximity, the click experiments showed that the design was not suitable for bringing the building blocks near enough to each other to react. Therefore, a new mechanism to bring the strands together is required.

3.4 5'–3' Click Chemistry

Since the existing mechanism to bring together building blocks on the 5' end of staple strands necessitates custom synthesised DNA linker strands, a new architecture was designed which removes this complication by using building blocks attached to the 5' or 3' ends of the modified staples. Now, only strands of opposite polarity can interact in the reaction centre, and an immediate drawback of the design is that the number of possible outcomes is halved to 60. In addition to altering the building block location, the anchorage method was also changed. Previously, adaptor strands were long (over 50 nt), modified staple strands, which have to be purchased PAGE purified from IDT, adding to the cost of the oligomer. To make the system as flexible as possible, it should be simple to move the position of the adaptor strand on the origami tile, and to swap one adaptor strand with another (to allow the flavour of the building blocks to be changed, for instance). However, in the current system, these changes require new, expensive, oligomers. To make the system more flexible, the existing modified staple strands should be redesigned and split into two separate

strands: a staple strand that can be used to anchor down the adaptor strand, and an adaptor strand that is independent from the origami itself and is linked to the chemical building block. The new short adaptor strands can hybridise to lengthened staple strands that anchor the adaptors away from the reaction centre, but now the location of the adaptor on the origami can be changed with relative ease (or the adaptor can be swapped with one attached to a different building block). Furthermore, in order to analyse the chemical product that results from the origami-templated click reaction, the adaptors can be stripped away from the origami without destroying the tile itself.

The new anchorage mechanism is depicted in Figure 3.18. The diagram shows how the new adaptor strands are held away from each other by two lengthened staple strands. Adaptor A_Al has an alkyne building block linked to the 3' end of the 18 nt strand, and adaptor A_Az has an azide building block attached to its 5' end. Strand A_Az is 24 nt in length as it also contains a 6 nt toehold domain, as shown in the diagram. This toehold is used by a stripping strand to pull away the click product from the origami tile. If the click reaction works, the two adaptor strands will be chemically linked together, resulting in a 42 nt product. The stripping strand is complementary to the full length of this product, forming a duplex with it which can be purified away from the rest of the origami (in a process described later. Incidentally, if the chemistry has not worked, the stripping strand simple separates strand A_Az from the origami.) The mechanism to release the adaptors from their anchorage positions outside of the reaction centre is similar to that of the 5'–5' design. A displacer strand can hybridise to the toehold domain of an "extended" (E) staple, and consequently compete with the "fixer" (F) staple over the competition domain of the E strand. The subsequent displacement of the F strand frees the adaptor from its anchorage. The portion of the F strand which is now single-stranded forms a sticky-end domain which interacts with other F strands. In doing so, the chemical building blocks are brought into close proximity (as shown in the diagram). However, the sticky-end

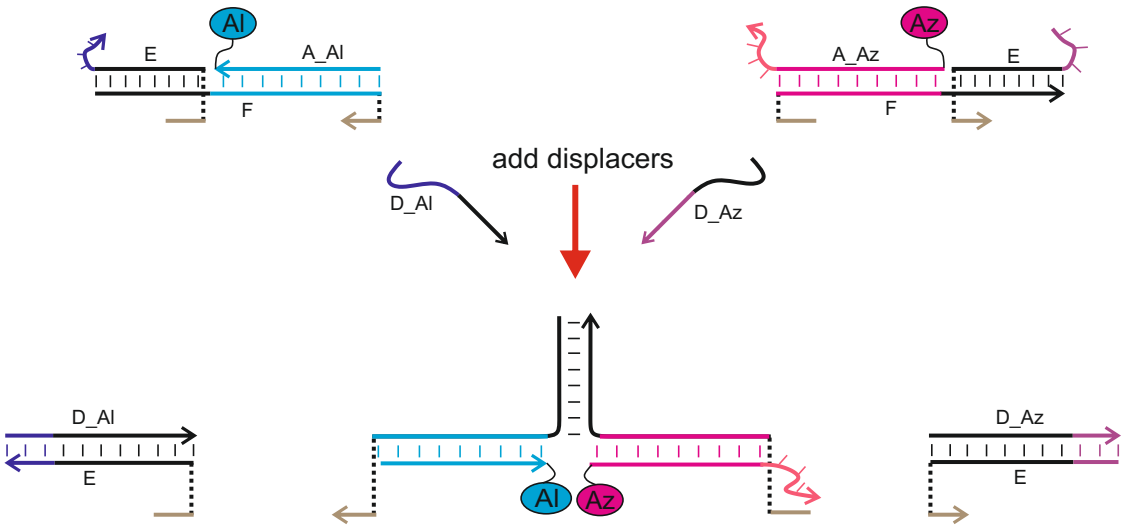


Figure 3.18: Illustration of the toehold-mediated strand displacement mechanism to bring together the 5'–3' click building blocks.

Adaptor A_Al is 18 nt and is alkyne-functionalised on its 3' end. The A_Az adaptor strand is 24 nt, and is azide-functionalised on its 5' end. In the initial state, it has a 6 nt single-stranded domain that acts as a toehold for the stripping strand which pulls away the click product from the origami. Both adaptors are hybridised to "fixer" (F) strands, which are themselves anchored down outside of the reaction domain by hybridisation to "extended" staple strands (E). The E strands have 6 nt toehold domains that are used by displacer strands (D_Al and D_Az) in the strand displacement process. As a result, the F strands are freed from their anchorages and can drift into the reaction centre. The resulting single-stranded section of the F strands forms a sticky-end domain, which can interact with other F strands. Hybridisation of the sticky-ends of the F strands brings adaptors together, and the building blocks into close proximity.

interaction here does not suffer from the geometrical problems of the previous design (discussed in Section 3.3.1). Once the reaction has taken place, the adaptor strand product is easily stripped away from the strands templating the reaction via the use of the free toehold of the azide adaptor. The mechanism is illustrated in the context of origami in Figure 3.19.

As discussed in Section 3.3.4, making an origami tile that is just large enough to hold one reaction centre is not straightforward. Therefore, to test the new click

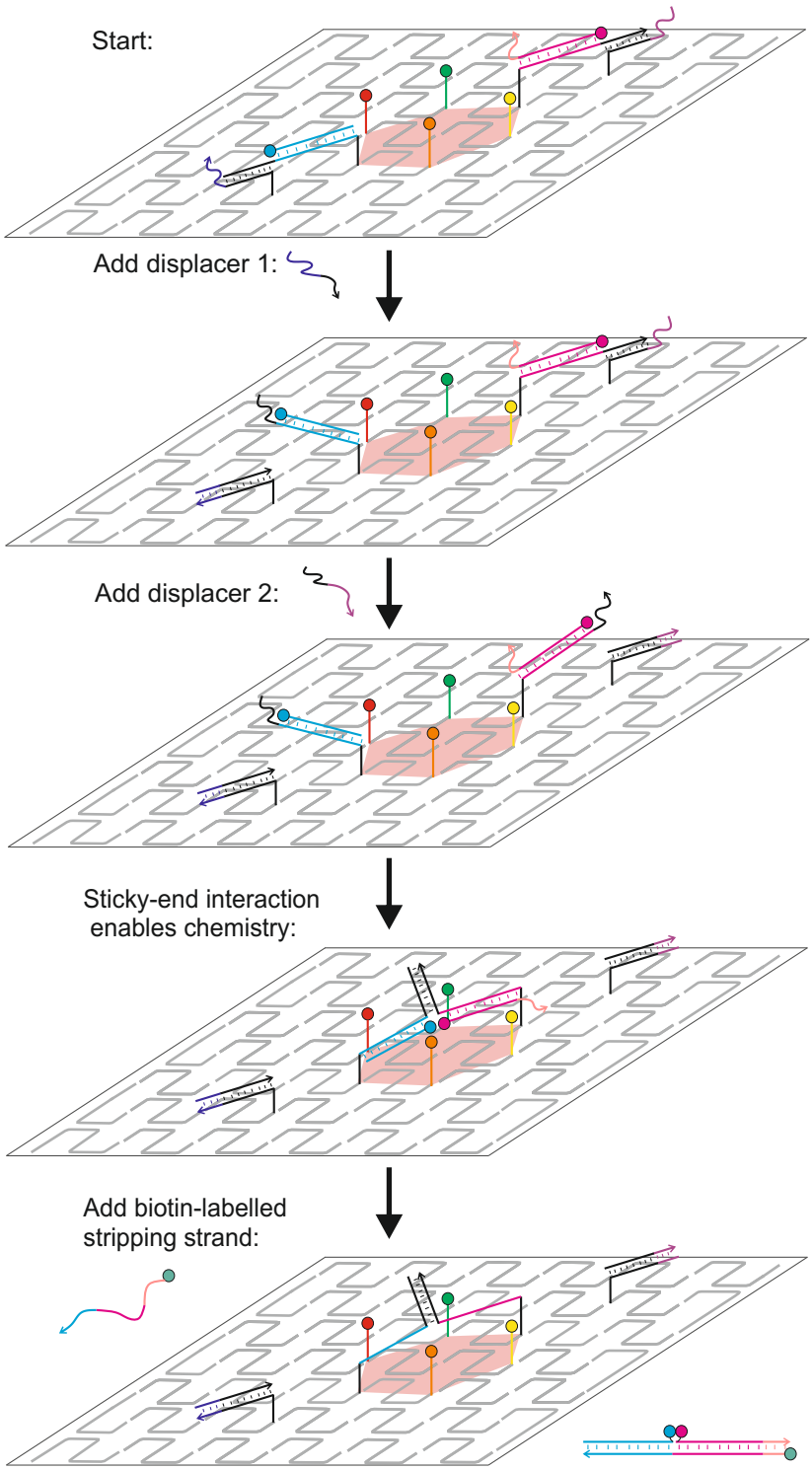


Figure 3.19: The origami-templated 5'–3' click reaction.

reaction mechanism, the full-size rectangular origami was used. However, since Rothemund published his original paper on DNA origami, it has been noted that the assumptions and models used to design the origami produce tiles that are slightly twisted. This twist may not be a concern for many of the applications of origami, but for DNA-templated chemistry it is a big problem. The anchorage system described here relies on the substrate being flat. If the substrate is twisted, the chemical building blocks may have the opportunity to interact with each other before being released simply because the geometry of the tile brings the building blocks into close contact. The modifications made to the Rothemund origami design are described in Section 3.4.2.

3.4.1 Off-Origami Test of Sticky-End Interaction

Prior to carrying out the click reaction on origami, the sticky-end interaction was tested in solution using both fluorophores and the click building blocks. For the fluorescence experiments adaptor Az is modified on the 5' end with a Fam fluorophore, and adaptor Al is linked on the 3' end to a Tamra fluorophore. The two fluorophores are brought near each other via the hybridisation of the sticky-end domains of the two complementary strands, as shown in Figure 3.20. The length of the sticky-end domains was varied between 8-14 nt to determine the length required to form a structure stable enough to bring the adaptors together. The four strands were hybridised together in equimolar quantities in origami buffer (1× TAE, 12.5 mM $MgCl_2$), and annealed by heating the solution 85 °C and cooling to 20 °C over 10 min. The samples (all at 1 µM) were analysed by running a 15 % non-denaturing PAGE gel in 1× TAE buffer for 45 min at 4 °C and 240 V. The gel was scanned using the Fam and Tamra emission and excitation wavelengths (Ex488/Em530 nm and Ex532/Em605 nm respectively). The results are shown in Figure 3.20. The Fam scan of the gel (A) shows that the concentration of adaptor Az may have been greater than 1 µM, as there is leftover DNA after hybridisation with its complement (Comp_Az) in lane 3.

Nevertheless, lanes 5-8 show that there are slowly moving bands corresponding to a four-strand product. This is confirmed by the Tamra scan (B), which reveals fluorescent bands in the same position on the gel in lanes 5-8. A multi-channel image of the gel (not shown) indicates that both fluorescent strands are present in the slowest bands in lanes 5-8. It can be inferred that the sticky-end interaction works, and that 8 nt sticky-end domains are sufficient to form a stable structure.

The previous experiments on origami-templated click highlighted the fact that FRET interactions alone may not be a useful indication that the chemical building blocks can interact successfully. Therefore, a click reaction was performed with the strands to check the architecture was amenable to the chemistry. The adaptor strands A_Az and A_Al were annealed to their complements in origami buffer. Four samples were made corresponding to the four sticky-end lengths. The copper catalyst was added to each sample in a 50 equiv excess with respect to the DNA concentration (1 µM), and the reaction was incubated at room temperature for 2 hours. The results were analysed by performing a 15 % super-denaturing PAGE gel, run for 45 min at 300 V and room temperature, in 1× Tris-Glycine buffer. (Further details of the denaturing PAGE protocol are provided in Appendix B.3). Figure 3.21 shows the gel after being stained with Sybr Gold and scanned at the Sybr Gold excitation and emission wavelengths (Ex488/Em530 nm). The single-stranded adaptor strands are visible at the bottom of the gel, although it is not possible to distinguish between them using this percentage of acrylamide in the gel. Lanes 4-7 contain the click reaction samples. The single-stranded complementary strands are clearly visible as the slowly migrating bands in lanes 4-7, and are predictably slower as the length of the strand increases. Each of the lanes contains a band which migrates at the same rate, indicating it is the desired 42 nt product of the click reaction (highlighted by the red box). Therefore, there is enough evidence to suggest that the sticky-end interaction is suitable for bringing together the click building blocks on origami.

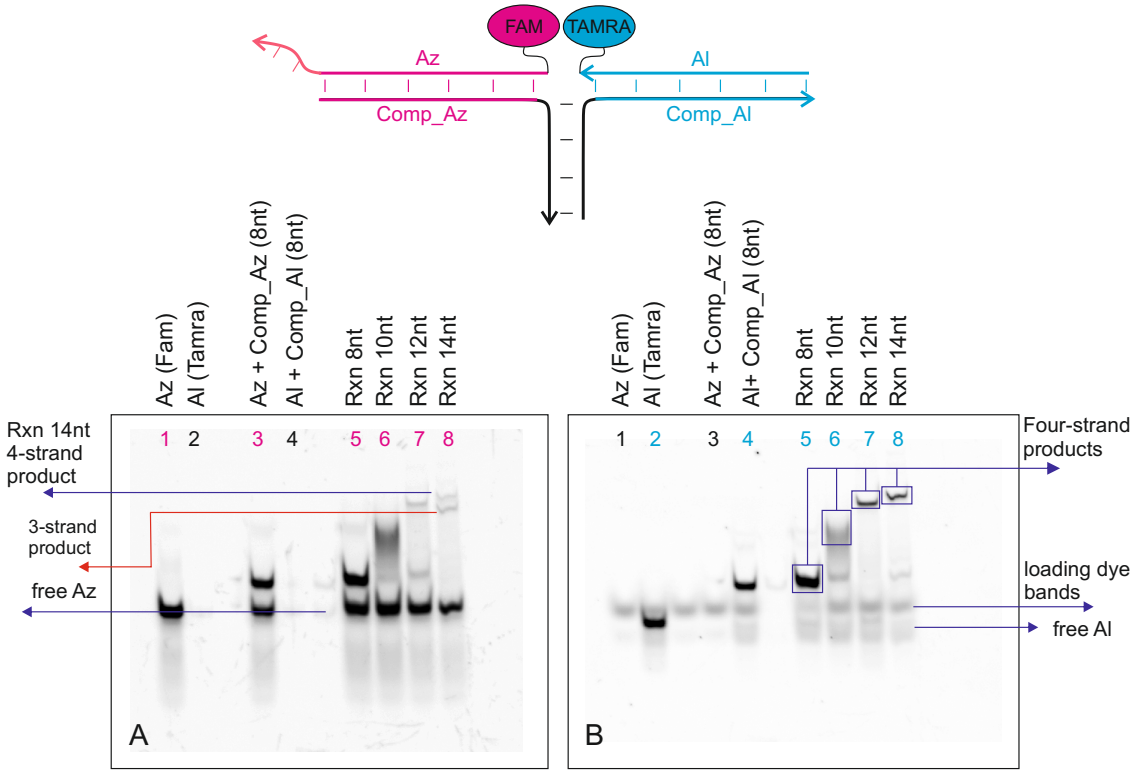


Figure 3.20: Test of the sticky-end interaction for the 5'–3' click chemistry using fluorophores. Strand Az is a 24 nt strand modified on the 5' end with a Fam fluorophore, while strand Al is a 18 nt strand modified on the 3' end with a Tamra fluorophore. The two fluorophores are brought into close proximity via the interaction of the sticky-end domains of the complementary DNA strands, Comp_Az and Comp_Al. The length of the sticky-end domains is varied between 8–14 nt, to determine the number of nucleotides necessary to produce a stable four-strand structure. Here, the diagram shows the sticky-end length as 12 nt.

The strands were annealed together in equimolar quantities and the samples analysed by a 15 % non-denaturing PAGE gel. Image A is the Fam scan of the gel, while image B is the Tamra scan. It is clear from A that the concentration of strand Az was measured incorrectly, as there is an excess of the strand in lanes 3 and 5–8. The excess also results in the two slowly moving bands at the top of lane 8. The highest band corresponds to the four-strand product (verified by the Tamra scan), while the lower of the two represents a three-strand product made without strand Al. (There is also a small excess of strand Al, as seen by the fast band in lanes 4–8). Lanes 5–8 all contain slowly moving bands that correspond to the formation of a four-strand product. The gel determines that an 8 nt sticky-end domain is sufficient to form the templating architecture.

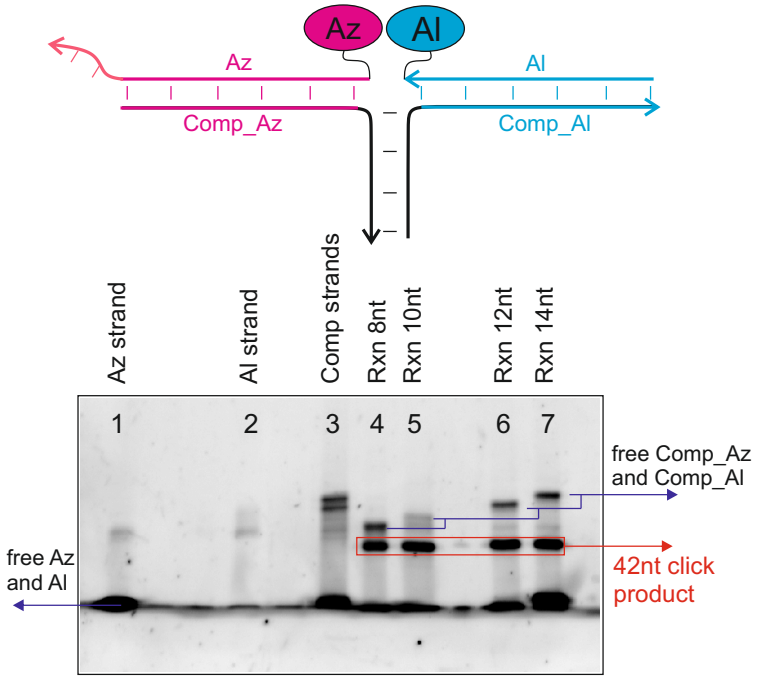


Figure 3.21: Test of the sticky-end interaction using click chemistry.

Strand Az is linked to an azide building block on the 5' end of the strand, while strand Al is linked to a 3' alkyne building block. The interaction of the sticky-end domains of the complementary strands (Comp_Az and Comp_Al) bring the building blocks into close proximity. The length of the sticky-end domains is varied between 8-14 nt.

The click reaction was performed by annealing together the four strands and adding the copper catalyst in a 50-fold excess. The samples were analysed after 2 hours by a 15 % denaturing PAGE gel, and the Sybr Gold scan after the gel had been stained is shown here. Lanes 4-7 show the single-stranded DNA strands that result from the reaction. The slowly moving bands at the top of each lane correspond to the complementary strands that templated the reaction (as shown in lane 3), which migrate at different speeds because they vary in length. The free azide and alkyne strands are at the bottom of the gel, but are indistinguishable from each other because a 15 % gel is not sufficient to separate out 18 nt and 24 nt oligomers. In each of the lanes 4-7, there exists a band that runs at the same rate on the gel, and it is likely that this corresponds to the 42 nt click product. It is unclear however, why the click product migrates faster on the gel than the shorter complementary strands.

For the purposes of origami-templated chemistry, the sticky-end domains of the "fixer" strands which bring adaptors together (as shown in Figure 3.18), were varied. Three sets of modifed staples were purchased to accommodate either 4-, 6- or 8 nt sticky-end domains. The tests described above determined that 8 nt sticky-ends were sufficient to template the chemistry. The shorter sticky-end domains are used for control experiments, as these lengths should not produce stable structures that lead to the click reaction taking place.

3.4.2 The Minimum-Twist Tile Design

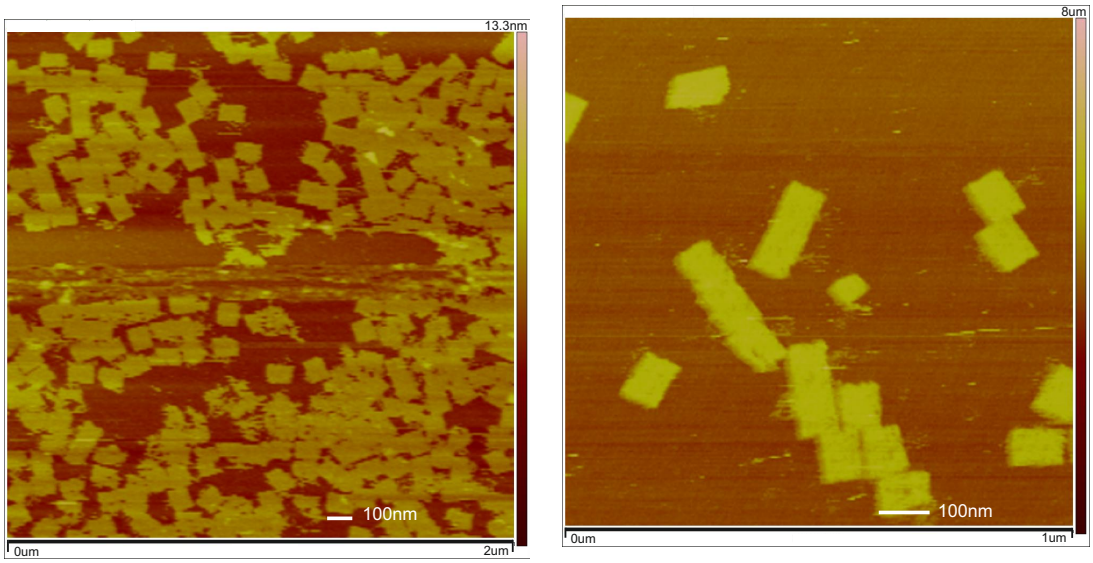
Reports claim that rectangular origami tiles designed by Rothemund do not lie flat on a mica surface when being imaged by AFM [Li et al., 2010]. This arises from under-winding in the DNA that results in the global twist of the whole structure. The minimal-twist origami tile was designed by Shelley Wickham, a PhD student in the DNA group who primarily works with DNA origami. To minimise the strain in the DNA, Wickham was influenced by the recent work on producing twisted origami [Dietz et al., 2009]. Here, the authors describe methods to purposefully alter the geometry of a tile by removing or inserting bases in each of the parallel helices that form the structure. Rothemund postulated in the Supplementary Information of [Rothemund, 2006] that the number of base pairs per turn used in his model may give rise to the twist of the tile, and Dietz determined that a minimally-twisted origami tile requires the average number of base pairs per turn to be reduced from the 10.67 bp/turn used by Rothemund to 10.42 bp/turn. In order to design the minimum-twist origami tile, Wickham used the origami design software caDNAno [Douglas et al., 2009b] to delete six bases per helix to achieve 10.42 bp/turn and generate the staple strands required to fold the M13mp18 scaffold DNA into the structure. The staple strands modified for the templated chemistry are listed in Appendix B.1.2.

3.4.3 Verification of Origami Formation

It is necessary to check that the M13mp18 ssDNA has folded into the desired 2D structure before proceeding any further. As described earlier, there are two methods to do this, AFM and agarose gel electrophoresis, both of which are used here. The origami structures were formed using a 50-fold excess of staple strands and a 5-fold excess of the adaptor strands. The standard annealing and spin column purification protocols were followed (outlined in Appendix B.2). Samples were typically made at 100 nM concentration and diluted in origami buffer to the required concentration for analysis by AFM (2-10 nM) or agarose gel (10 nM).

Firstly, AFM was used to determine whether the geometry of the origami structure was as expected (the samples were made using the 8 nt sticky-end domain staples). Origami samples were observed using the tapping mode of a Veeco 'Nanoman' Atomic Force Microscope, in 1× TAE, 20 mM $MgCl_2$ buffer. The salt concentration of this buffer is higher than that used in the origami buffer as it has been found that this helps the origami structures to stick to the mica surface. Details on the method used to obtain the images is provided in Appendix B.11. Initially a high concentration of purified origami was used, but the density of tiles on the mica surface made resolution of individual tiles difficult, as shown in Figure 3.22(a). A reduction of the concentration to 4 nM resulted in better quality images (Figure 3.22(b)). Individual structures are clearly resolved, and the Nanoman control software determined that the structures are of the expected size. The image shows that tiles aggregate end-to-end. This is a result of stacking interactions between the blunt-ended helices of the tile [Rothemund, 2006]. However, the stacking interactions only results in two or three tiles forming a chain, a significant reduction on the 5 µm chains Rothemund observed (described in the Supplementary Information of [Rothemund, 2006]). The reduction in aggregation is a consequence of the design of the origami. Wickham deployed a method described by Rothemund to lessen the stacking interactions, which involves

adding loops of four T nucleotides to staple strands on the edges of the tiles. The 4T loops replace the blunt-ends of the helices and thereby reduce the possibility of stacking. The anchored adaptor strands are located on the surface of the tile, but they were not resolved by the AFM. However, the AFM analysis of the origami clearly shows that tiles of the correct size and shape have been formed.



(a) 8 nM (b) 4 nM

Figure 3.22: AFM images of the minimum-twist origami tile.
(a) Initial imaging was carried out using origami samples at 8 nM concentration, however the density of tiles on the mica surface was too high to enable individual structures to be clearly resolved.
(b) Using a 4 nM concentration, the tiles were more easily imaged. The tiles aggregate end-to-end in two- or three-tile chains due to stacking interactions between blunt-ended helices. The unused, single-stranded portion of the M13mp18 scaffold is also just resolvable as a loop protuding from one of the longer edges of each tile.

While AFM images confirm the origami structures have the correct geometry, they do not verify whether the modified staple strands and adaptor strands have also hybridised as designed. To determine this, origami samples were analysed by agarose gels. Three samples were made, using the three sets of modified staple strands and

the fluorescently-labelled adaptor strands. The samples were used both unpurified and purified to ensure the spin-column purification method removes excess staples. Samples were run on a 0.75 % agarose gel for 180 min at 70 V at 4 °C (see Appendix B.8 for details). The gel was scanned firstly for the Fam and Tamra excitation and emission wavelengths. The gel was then stained with Sybr Gold nucleic acid stain, and scanned for the Sybr Gold excitation and emission wavelengths. The Sybr Gold scan of the gel is shown in Figure 3.23. The origami bands in lanes 4-9 run slightly slower on the gel than the ssDNA scaffold (lane 2). The fluorescence scans determined that there were fluorophores present in the bands (as shown by the yellow band obtained from a multi-channel image of the gel), implying that the adaptor strands have hybridised to the origami tiles. Furthermore, the gel was also scanned for FRET prior to staining (Ex488/Em605 nm), to determine whether the adaptor strands are hybridised to their anchorages or in the sticky-end architecture. The scan showed no FRET (image not shown), suggesting that the fluorophores were far apart and unable to interact with each other (and so, anchored outside of the reaction centre). Excess staple strands can be seen as the large bands at the bottom of the gel. Notably, the concentration of the excess staple strands is significantly reduced in the purified samples. The gel confirms that the adaptor strands and modified staples have hybridised to the origami tile as desired.

3.4.4 Origami-Templated Click

All the results discussed so far in this section suggest that not only has the origami tile formed correctly, but that the new templating architecture is also more likely to yield a successful click reaction on origami. Therefore, the click chemistry was carried out on origami. The azide and alkyne adaptors used earlier for the off-origami experiments were used to form the origami. Three tests were performed with each of the three sets of strands corresponding to the different sticky-end lengths. Either the azide displacer strand was added (D_Az), or the alkyne displacer was added (D_Al)

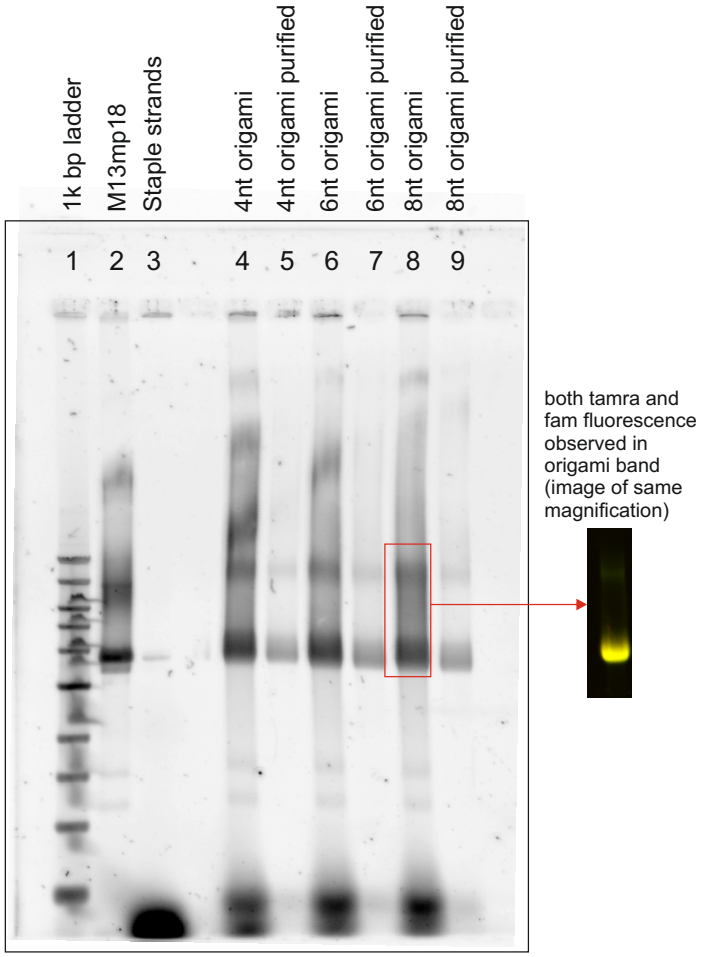


Figure 3.23: A 0.75 % agarose gel used to verify correct origami formation.

Origami samples were made using the fluorescently-modified adaptor strands, and were run on the gel both before and after spin-column purification. The gel was run for 180 min at 4 °C and 70 V in 1× TAE buffer. The gel was scanned at the Fam and Tamra excitation and emission wavelengths, then stained with Sybr Gold and rescanned at the Sybr Gold excitation and emission wavelengths. The Sybr Gold stained gel shows that in lanes 4-9 there is a large band which is presumed to be the origami tile. Scans for fluorescence revealed that these bands also contain both fluorophores (shown by the yellow-coloured band from a multi-channel view of the gel), and so the tile is assumed to have formed as designed. The excess, unused staple strands are visible at the bottom of the gel. As expected, purifying the origami significantly reduces the free staple strand concentration, and this is clearly shown in lanes 5, 7 and 9.

or both displacers were added to the origami solution. The origami concentration was 100 nM, and the copper catalyst was added in either 500 equiv excess or 50 equiv excess. Furthermore, the ratio of the copper catalyst to the reducing agent Na Asc was 1:100, and the ratio of the copper to the copper-binding ligand THPTA was 1:7. The increase in the amount of Na Asc used in these origami experiments compared to earlier click reactions carried out in solution stems from the research on origami-templated click chemistry in [Voigt et al., 2010], [Hong et al., 2009]. The authors use a 100-fold excess of Na Asc with respect to the copper concentration instead of the 10-fold excess used in [Kočalka et al., 2008], because the risk of destroying the DNA by copper oxidation is more detrimental to origami experiments than tests using a few oligomers. Thus, as a precaution against the copper oxidising and destroying the origami structures, the sodium ascorbate was used in a larger excess.

Nine different origami samples were prepared: the strands with 4-, 6-, 8 nt sticky-end domains were reacted with (i) displacer D_Az, (ii) displacer D_Al, (iii) both displacers. The displacer strands were added to each sample in a 10-fold excess immediately prior to adding the catalyst. The samples were incubated at room temperature on a shaker for 2 hours. A 60 nt biotin-labelled stripping strand was added to each sample in a 100-fold excess and incubated for a further 30 min. The stripping strand can hybridise to the 6 nt toehold on the azide adaptor strand and pull away the adaptor from the origami as depicted in Figure 3.19. In the case where both adaptor strands have been released from their anchorages and the click reaction has taken place, the stripping strand pulls the 42 nt product away from the origami. The duplex is then separated from the origami tile by using streptavidin-coated magnetic beads. Biotin has a high-affinity to bind to the streptavidin protein, and so once the biotin-labelled duplex comes into contact with the magnetic beads, the biotin-streptavidin bond enables the duplex to be removed from the origami sample. The full protocol for the extraction of the click product is provided in Appendix B.7. Once the duplex

was removed from the solution, it was necessary to separate the two strands forming the double helix, as the stripping strand is of no interest. The duplex was denatured by heating to 80 °C for 2 min and immediately applying a magnet to separate the two strands. As the biotin-labelled stripping strand is attached to the streptavidin magnetic beads, the elute after applying a magnet should largely contain the click product (although because the process is not entirely efficient, the elute also contains some biotin-labelled strands).

The elute was analysed by performing a 15 % super-denaturing PAGE gel to further separate the product from the stripping strand. In the reactions using the 4 nt and 6 nt sticky-ends, no click product was observed on the gel. Only the 60 nt biotin-labelled strand and the 24 nt azide adaptor strand were observed on the gel (data not shown). These control experiments confirm that the sticky-end domain length must be greater than 6 nt in order for the click reaction to take place. However, the click reaction was more successful using the 8 nt sticky-ends, as the stability of the 8-mer duplex allows the reaction to take place. The results are shown in Figure 3.24. The reaction is not perfect. It appears that even when only one of the displacer strands is added, the click reaction still occurs. In particular, the intensity of the band corresponding to a 42 nt product is higher when only displacer D_Al is added than when only D_Az is added. There are a number of reasons why the reaction has taken place with only one displacer present. Firstly, it is possible that the adaptor and fixer strands have not annealed to their anchorages as designed, but are already in the sticky-end configuration. However, the variation in band intensity suggests that this may not be the case, since if it were, the click reaction would now be independent of the displacer strands. Secondly, it is possible that one of the adaptors is not anchored down, such that the strand is able to diffuse into the reaction centre even when its associated displacer strand has not been added. Thirdly, and more likely, it is possible that the displacer strands are not specific enough, such that they are able to

release the wrong adaptor. From the data, it appears that the D_Al strand is the least specific of the two, as the band intensity is higher. This could be remedied in future experiments by altering the sequences of the toehold domains used by the displacer strands to release the adaptors, or by increasing the length of the toeholds. It is important to note however, that the origami-templated click reaction does appear to have worked when both displacer strands are added. The bands highlighted by the red boxes in Figure 3.24 correspond to a 42 nt product, as they migrate at the same rate as the control strand in lane 3. The gel also shows that using either a 50-fold or 500-fold excess of the catalyst does not result in the reaction going to completion. Unreacted azide strand (which is always extracted by the biotin-labelled stripping strand) is clearly observable at the bottom of lanes 4-9, in line with the band in lane 1. However, this is not surprising, as the click reactions carried out in solution did not go to completion either.

In order to confirm that the bands assumed to be the 42 nt product were actually the result of the click reaction, mass spectrometry analysis was performed. Mass spectrometry requires a minimum sample concentration of 5 µM, which is too high for an origami experiment. Origami sample concentrations are limited by the concentration of the M13mp18 scaffold strand, which is typically purchased at 420 nM (in 200 µl). A high concentration reaction was performed to enable analysis by mass spectrometry. The 50 µl of elute obtained after streptavidin magnetic bead extraction of the click product was dehydrated using a DNA120 Savant SpeedVac Concentrator, and then rehydrated in MilliQ to 30 µl, the volume required for the mass spectrometry. The sample was analysed, but no DNA was found in the sample. It is probable that the sample concentration was still too low for analysis. Therefore, it appears that the product of the click reaction has been observed on a gel, but a more conclusive method to confirm this needs to be devised.

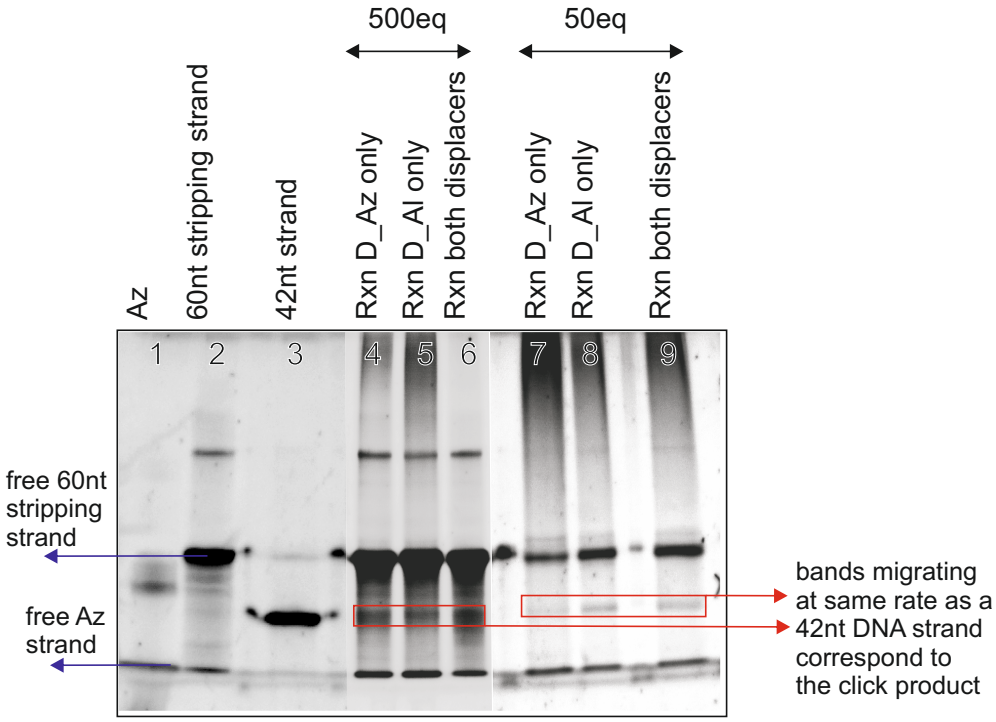


Figure 3.24: 15 % denaturing PAGE of the origami-templated click reactions.

Three tests were performed on origami: either one of the displacer strands was added to the origami to release one of the adaptors from its anchorage, or both displacers were added. Ideally, the reaction should only occur when both displacers are present. The copper-catalyst was added either in 500 equiv or 50 equiv with respect to the origami concentration (100 nM).

The image shown here is the Sybr Gold scan of the gel after it had been stained. The biotin-labelled stripping strand used to pull the click reaction product away from the origami is present in lanes 4–9. The red box highlights bands in lanes 4–9 that migrate at the same rate as the 42 nt strand in lane 3. Therefore, it can be assumed that the DNA in this band is also 42 nt and that it corresponds to the click reaction product which the stripping strand has separated from the rest of the origami strands. The click reaction appears to have taken place even when only one of the displacer strands have been added, possibly because of a lack of specificity in the displacer strand sequences.

3.5 Summary

Two templating architectures to carry out the origami-templated click chemistry have been investigated, and each has their own advantages and drawbacks. The system to bring 5' click building blocks together has the potential to create 120 possible 6-unit products from one reaction centre. However, bringing together like-polarities of DNA is difficult. Using a linker strand that is partially made from inverted bases is a way around the problem, though perhaps an expensive one, and moreover, analysis of the product is difficult. The result of the current system of bringing two opposite polarity adaptor strands together on origami is a 42 nt product, but the concentration of the product may be too low to analyse via mass spectrometry. This is indeed a problem for the origami-templated chemistry as a whole, which typically involves concentrations and volumes that are unsuitable for mass spectrometry. Furthermore, it is not yet clear how the result of six adaptors interacting will be extracted from the origami. These issues need to be resolved by chemists who are familiar with the various analysis techniques available *before* redesigning the templating architecture. Ideally, all future origami-templating architectures would use the short adaptor strands (which are all 5' functionalised with click building blocks to maximise the number of possible chemical products from one reaction centre), in combination with DNA linker strands to bring together the building blocks for the reaction, thereby making use of the advantages of the two origami architectures explored so far.

Chapter 4

Summary

The work presented here forms part of an ongoing collaboration to use DNA to build a synthetic ribosome mechanism that is capable of interpreting "instructions" in the DNA sequences to aid the synthesis of new functional chemical molecules. As outlined in Section 1.5.1, there are already several DNA-based strategies to encourage reactions between chemical building blocks to form new molecules. There are also methods to build dynamic machines that can make decisions about which direction they move in as a response to some stimulus. It is a logical progression to combine the dynamic machines, which already respond to instructions located in DNA sequences, with the DNA-templated chemical synthesis idea to form a mechanism that moves and encourages chemical reactions between molecules that it encounters whilst in motion.

There are a number of possible chemical reactions that are suited to DNA-templated synthesis, and some of these possibilities are being explored by the collaboration [McKee et al., 2010]. The bulk of the author's contribution to the project was to design different DNA architectures that could incorporate the chemical synthesis. Two designs are presented here; the first takes inspiration from the DNA walker mechanisms that use DNA strand displacement to move uni-directionally along a DNA

track, while the second moves away from this linear approach to using DNA origami as a breadboard for chemical reactions. Both designs have one thing in common: they utilise the strand displacement technique to make changes to the state of the DNA device. In the case of the cascade-coupling mechanism, the strand displacements drive the mechanism forward and in the process, enable the chemically-functionalised DNA strands hybridised to a track to interact with their neighbours. On the origami however, the strand displacement idea is used simply to encourage functionalised DNA strands to react with each other inside a reaction centre. The advantages and disadvantages of both systems were discussed in Section 2.6 and 3.5.

Designing DNA architectures for the synthetic ribosome mechanism was simpler when the particular chemical reaction it would be combined with was known from the outset. This was the case with the DNA origami architecture, which was designed with click chemistry in mind. Although the original idea to attach the click building blocks onto the 5' ends of the DNA adaptor strands was not entirely successful, changing the system to bring opposite polarities together did enable the click reaction to occur on origami. The success is partially due to the fact that click chemistry had already been used in DNA-templated reactions, and been shown to work across a range of reaction conditions. There is still much to do to prove that the architecture can produce libraries of small molecules. Much of this relies on the synthesis of a stable tetrafunctional click building block (as described in Section 3.2.1), which is a work in progress. The sticky-end interaction which brings together the functionalised DNA strands on the origami could also be improved. The results of carrying out the click reaction on origami showed that the reaction was possible, but that it could take place even when only one of the two required DNA displacer strands was added to the solution. This suggests that either there is a lack of specificity in the displacer strand sequence, enabling one displacer to react with more than one strand on the origami, or that one or both of the functionalised strands are not correctly anchored

outside of the reaction centre, or that strands from one origami tile are able to react with strands from a different origami tile (due to the reactions being carried out in solution). This latter problem could be minimised by sticking the origami tiles to a mica surface before performing the chemistry, as described in the origami design choices discussion in Section 3.3. The other two problems need to be investigated using the existing bifunctionalised DNA strands. The simplest approach would be to change the sequence of one of the displacer strands (and the appropriate domain on the strand it interacts with), to determine whether it is specificity that is the problem before trying to redesign the anchorage mechanism.

Another problem with the click reaction is the overall yield of the reaction, which has always been around 40 % in the experiments presented here. Reports on using click chemistry in DNA-templated reactions have shown the reaction to be high-yielding (nearing completion within 5 min in some cases), so it is unclear why the reaction here does not reach completion, even after several hours or leaving the samples to react overnight. The click reaction protocol includes many steps at which it is necessary to de-gas solutions, and it is possible that this has not been performed adequately. Oxygen in the solution can combine with the copper, which results in DNA damage (described in Section 1.5.3), and so it is possible that the reaction does not produce high yields because the DNA is damaged and unable to interact as designed. It may be necessary to look closely at the click reaction protocol to determine whether there are any steps at which this error can significantly impact the yield of the reaction.

The main area of focus for the cascade-coupling mechanism in the future is combining it with a particular chemical reaction. The results of the toehold exchange reactions show that the mechanism is capable of completing many steps, and that the time for each exchange reaction can be controlled. However, the entire system was

designed without knowing which chemical reaction it be used with. If the particular chemical reaction chosen is sensitive to distance between the reacting groups, then it may not suit the cascade-coupling mechanism, as the functionalised single-stranded 5' end of the target strand (after it has been partially or completely displaced from the universal domain) is flexible and drifts towards and away from the building block on the adjacent target strand. The Wittig reaction has the advantage that it involved a group transfer reaction also, to enable the growing chain of building blocks to progress along the DNA track. Any chemistry that is found to suit the way building blocks are brought into close proximity in the cascade would also need to be combined with a transfer reaction.

Whilst the possibility for the DNA origami-templated chemistry to produce libraries of small molecules is clear, it is perhaps less clear how the cascade-coupling mechanism could be used to generate many different 6-unit products. One way would be to build the initial state of the system from shorter DNA segments. This is illustrated in Figure 4.1. Here, target strands can form duplexes by hybridising to the complementary universal domain of short DNA template strands. The hybridisation is directed by the universal domains such that the target strands can combine with any of the different template 'flavours'. The 'flavour' is determined by the toehold domain of the template strand (indicated by a colour in the diagram). Thus, many varieties of these short duplexes are formed. These can then form chains of duplexes through the interactions between the toehold domains of the strands, and the result resembles the initial state of the cascade (Figure 2.9), where many target strands are hybridised to a long template. It would be necessary to ligate together the template strands so that they form one continuous strand. In this way, many different cascade duplexes could be formed where the target strands are hybridised to the template in a different order. This has the possibility of producing many different chemical products once the cascade-coupling mechanisms are set into motion. However, there

is a drawback to this idea. It is difficult to control the ligation process that is used to form the long template strand. The strands can form chains of duplexes that contain any number of target strands, and therefore, the ligation process could result in duplexes of a variety of lengths. Even if one of the short template strands and one of the target strands were designed to be unique to try and control the length of the ligated product (for instance, if a strand does not have a toehold domain for anything to hybridise to, it would naturally stop the duplex from growing in one direction), the length of the duplex before reaching this "stop" strand is not readily controllable. Whilst the cascade mechanism has been observed to work over six steps, it is not obvious that it can be extended indefinitely. However, if it were possible to control the length of the products of this duplex synthesis process, then this would be a neat way to generate libraries of small molecules, which is the ultimate aim of building this synthetic ribosome mechanism.

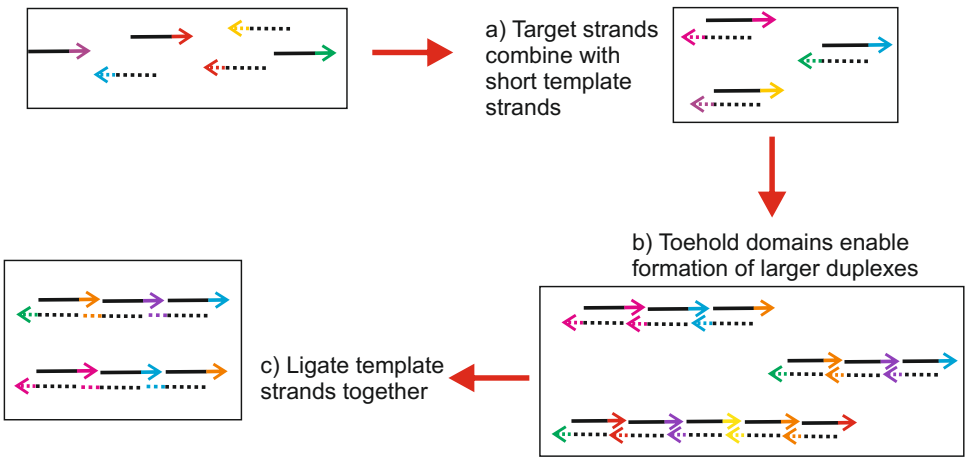


Figure 4.1: A method to produce many different cascade duplexes in a one-pot reaction. (a) Target strands are mixed with complementary short template strands (dashed lines) in solution. The strands are able to hybridise together via their universal domains (black). (b) Longer chains of strands are formed when the duplexes hybridise together by the interaction of complementary toehold domains (various colours). The result is a solution of duplexes with different combinations of target strands, however, the length of the duplexes is not easily controlled. (c) Before the duplexes can be used in a cascade-coupling reaction, the short template strands need to be ligated together in order to create one continuous template. Thus, it is possible to use relatively few target strands (and their complements) to produce many distinct duplexes that can be combined with the cascade chemistry to generate large libraries of small molecules.

Appendix A

General Experimental Protocols for the Cascade

A.1 Experimental Techniques

The two key techniques used to experiment with the cascade-coupling mechanism are detailed below. Synthetic DNA strands can be purchased with modifications on either end of the strand. For the majority of experiments conducted here, one or more strands in a particular system were purchased with fluorescent molecules attached to either their 5' or 3' end. These fluorescent molecules were particularly useful for monitoring changes in a system using FRET, or to ensure correct formation of a particular system of strands using PAGE analysis. These are described in more detail below.

A.1.1 Förster Resonance Energy Transfer

Förster Resonance Energy Transfer (FRET) describes an energy transfer mechanism between two fluorophores [Förster, 1948], [Stryer and Haugland, 1967]. Fluorophore molecules that are able to transfer energy between one another in this way form FRET pairs. In the pair, one of the fluorophores is a "donor", while the other is an

"acceptor". The energy difference between two molecular orbitals in the fluorophores falls within the range of the visible spectrum. When a donor fluorophore is excited by a particular wavelength of visible light, it emits a wavelength of a different energy. If the acceptor fluorophore is in close proximity (<10 nm), the energy is transferred to it by a nonradiative coupling mechanism to an acceptor fluorophore in close proximity. Thus, when the donor and acceptor fluorophores are near to each other, the donor emission fluorescence decreases, while the acceptor emission fluorescence increases. Similarly, as the distance between the two molecules increases, the donor emission fluorescence increases and the acceptor emission fluorescence decreases.

FRET is therefore a useful tool to quantify molecular dynamics such as protein-protein interactions, or as in this case, DNA-DNA interactions. For instance, FRET can be used to monitor the hybridisation of two ssDNA. Of the two strands, one is labelled with a donor fluorophore and the other with an acceptor fluorophore. If only the donor-labelled strand is in solution, the donor emission fluorescence is relatively high. Once the acceptor-labelled complementary strand is added to the solution, the rate at which the donor emission fluorescence decreases describes the rate at which the two strands hybridise together and bring the fluorophores into close proximity. In the experiments carried out so far, two different FRET pairs have been used.

Cy3 and Cy5: here Cy3 is the donor fluorophore Ex565/Em580 nm and Cy5 the acceptor fluorophore Ex650/Em665 nm.
FAM and TAMRA: here FAM is the donor fluorophore Ex495/Em520 nm and TAMRA the acceptor fluorophore Ex565/Em580 nm.

A.1.2 Fluorescence Spectroscopy

Fluorescence measurements were made using either one of the two available machines in the DNA Group.

JY-Horiba Fluoromax-3 spectrofluorimeter, equipped with a four-cell changer, and an external water bath to maintain the temperature of the sample compartment. A xenon arc lamp provides high intensity white light. A monochromator with adjustable slit width (set to 2 nm) is used to provide illumination of a given wavelength and dispersion. Prior to entering the sample compartment, the bean passes through a beam splitter, and a small fraction of the beam is deflected onto a photodiode that records a reference signal. Fluorescence at a given wavelength is monitored by a second monochromator that captures light at right angles to the incident beam (with this slit width set to 5 nm). The signal recorded is normalised by the reference signal. To obtain annealed baselines, the external waterbath was first set to 85 °C, with the stoppers removed from the cuvettes (heating the cuvettes with the stoppers in place could place stress on the cuvettes). Heating of the solution to this temperature took around 15 min. The temperature was maintained for 10 min and then set to the experimental temperature (typically 21 °C). The cooling time was around 45 min. The fluorescence was then recorded until a non-varying reading was obtained.

Varian Cary Eclipse fluorescence spectrophotometer with a four-cell changer, and temperature control. The machine uses an intense xenon flash lamp to provide light across the whole wavelength range. Photosensitive samples are not exposed to continuous light as the lamp flashes only to acquire a data point. The monochromator slit widths were set to 5 nm for excitation wavelengths and 10 nm for the emission wavelengths, the PMT[1] set to 600V (unless otherwise stated), and the integration

[1]Photomultiplier tubes (PMTs) are sensitive light detectors, which are often used in low intensity applications such as fluorescence spectroscopy. The detectors multiply the current produced by the incident light by several orders of magnitude. The detected current depends on the number of incoming photons and the gain of the photomultiplier. The gain of the PMT can be varied by adjusting the voltage supply across the PMT. This is used to alter the strength of the signal to accommodate both strong and weak light sources.

time to 1.0 sec. The samples were annealed *in situ* at the end of the experiment by using the temperature controls of the Cary Eclipse software. The samples were heated to 85 °C in the fluorimeter and cooled to 21 °C over 10 min. A post-anneal baseline was recorded for each sample and used to normalise the data.

Quartz cuvettes (Hellma UK Ltd) were used for all measurements. The working volume of these cuvettes was 1.5 ml. Rapid mixing of adding strands was accomplished using a 1 ml pipettor. To obtain complete mixing, a mixing time of at least 10 sec was required. To eliminate effects of evaporation, the supplied PTFE cuvette stopers were used.

A.1.3 Poly-(acrylamide) Gel Electrophoresis

PAGE is an extremely useful technique used to separate DNA, RNA or protein molecules and to observe how the molecules are interacting. The method involves loading molecules into wells within a thin, solid gel and applying an electric voltage to force the molecules to move (the "electrophoresis"). The term "gel" refers to the matrix used to contain and then separate the target molecules, and is typically a polymer whose properties (such as porosity) are chosen based on the specific weight and composition of the molecules being studied. The molecules will move through the gel at different rates, usually dependent on their mass, towards the positive anode if negatively charged and vice versa. After the electrophoresis is complete, the gel is scanned for any fluorophores or dyes attached to the molecules which will reveal the position of the molecules in the gel. Additionally, the gel can be stained to make the molecules visible, using the nucleic acid stain Sybr Gold for instance.

Each lane shows separation of the components from the original mixture as one or more distinct bands, one band per component. Bands in different lanes that end up at the same distance from the top contain molecules that passed through the gel

with the same speed, which usually means they are approximately the same size. The distance a band travels is approximately inversely proportional to the logarithm of the size of the molecule. In the case of DNA, the direction of migration from negative to positive electrodes is due to the negative charge of the phosphate backbone. Duplexes behave as rigid rods, so their migration through the gel is relative to their radius of gyration, or for non-cyclic fragments, size. Single-stranded DNA tend to fold up into molecules with complex shape and migrate through the gel in a less predictable manner.

To prepare 10 ml of a 15 % non-denaturing PAGE gel solution:

Ingredient	**Quantity**
15% acrylamide/bis (29:1)	10 ml
10 % APS	100 µl
TEMED	10 µl

To cast the gels, assemble two sets of gel plates. Pour out 10 ml of the 15 % acrylamide solution into a Falcon tube. Add 100 µl of 10 % APS (ammonium persulfate) and 10 µl of TEMED (tetramethylethylenediamine) to the Falcon tube. (TEMED catalyses the polymerisation of the acrylamide, so it is necessary to add the TEMED immediately before casting the gel.) Swirl and pour to fill the volume between the glass plates. Immediately insert clean gel combs. Once set (after around 10 min), remove the gel from the casting clamp, take out the combs, and wash out the gel wells using a syringe filled with MilliQ water.

To run the gels, place the gel into the buffer tank with 1× Tris-Acetate-EDTA (TAE) running buffer. Gels are run at 4 °C, at 240 V. Prepare DNA samples for the gel by mixing 2 parts sample to one part loading buffer (3 ml glycerol, 0.25 %

bromophenol blue, 7 ml MilliQ water). Load the gel wells with 10 µl of the samples. Typically, the gel is run for 45-60 min.

DNA was visualized using a BioRad Pharos FX Plus Molecular Imager before and after staining with SYBR Gold nucleic acid gel stain (Invitrogen).

A.1.4 PAGE Purification

In order to PAGE purify a sample, typically a 1 nM sample is added to several lanes of a 15 % PAGE gel using the protocol described in Appendix A.1.3. After running the gel for around 45 min at 240 V and 4 °C, the gel is scanned to check the sample has assembled correctly. If the sample contains fluorophore-labelled strands, then bands which are fluorescent may correspond to the assembled complex. Unhybridised strands usually migrate quickly on a gel, which allows the desired complex to be differentiated from the bands containing unhybridised fluorophore-labelled DNA.

The gel is then visualised using blue light. This illuminates bands containing the assembled duplex, and allows these bands to be cut out of the gel, thereby separating the desired product from all other DNA strands and complexes in the sample. The strip is placed in an Eppendorf 1.5 ml safe-lock tube, and is crushed and soaked in 1 ml of the buffer used in experiments (for the cascade experiments, this is NEB2 buffer). The tube is left overnight to allow the DNA duplexes to diffuse out of the gel into the buffer. The concentration of the gel-purified sample was measured before carrying out the experiments using a Cary 50 Probe UV-Vis Spectrophotometer to determine the UV absorbance at 260 nm.

A.2 MATLAB Program For Six-Step Cascade Model

The ordinary differential equations were solved using the `ode23s` function in MATLAB. This ODE solver is used because the system described by the differential equations is "stiff", containing reactions with very different time scales.

A.2.1 Simple Cascade Model

```
k = 4.0692e5; %forward 1st step
k1 = 1.47e3; %back 1st step
k2 = 5.9341e3; %forward 2nd step
k3 = 3.9757e5; %back 2nd
k4 = 3.7675e5; %forward 3rd
k5 = 2.7096e4; %back 3rd
k6 = 1.841e5; %forward 4th
k7 = 1.841e5; %back 4th
k8 = 1.1205e5; %forward 5th
k9 = 2.7252e5; %back 5th
k10 = 1.4991e5; %forward 6th
k11 = 2.461e5; %back 6th


dy(1) = - k*y(1)*y(2) + k1*y(3)*y(4);
dy(2) = - k*y(1)*y(2) + k1*y(3)*y(4);
dy(3) = + k*y(1)*y(2) - k1*y(3)*y(4) -k2*y(3)*y(5) +k3*y(6)*y(7) ;
dy(4) =  k*y(1)*y(2) - k1*y(3)*y(4);


dy(5) = -k2*y(3)*y(5) + k3*y(6)*y(7);
dy(6) = k2*y(3)*y(5) - k3*y(6)*y(7) - k4*y(6)*y(8) + k5*y(9)*y(10);
dy(7) = k2*y(3)*y(5) - k3*y(6)*y(7);


dy(8) = - k4*y(6)*y(8) + k5*y(9)*y(10);
```

```
dy(9) =  k4*y(6)*y(8) - k5*y(9)*y(10) - k6*y(9)*y(11) +
k7*y(12)*y(13);
dy(10) = k4*y(6)*y(8) - k5*y(9)*y(10);

dy(11) = - k6*y(9)*y(11) + k7*y(12)*y(13);
dy(12) = k6*y(9)*y(11) - k7*y(12)*y(13) - k8*y(12)*y(14) +
k9*y(15)*y(16);
dy(13) = k6*y(9)*y(11) - k7*y(12)*y(13);

dy(14) = - k8*y(12)*y(14) + k9*y(15)*y(16);
dy(15) = k8*y(12)*y(14) - k9*y(15)*y(16) - k10*y(15)*y(17) +
k11*y(18)*y(19);
dy(16) = k8*y(12)*y(14) - k9*y(15)*y(16);

dy(17) = - k10*y(15)*y(17) + k11*y(18)*y(19);
dy(18) = k10*y(15)*y(17) - k11*y(18)*y(19);
dy(19) = k10*y(15)*y(17) - k11*y(18)*y(19);
```

A.2.2 Leak Cascade Model

```
k = 4.96e5; %forward 1st step
k1 = 1.91e3; %back 1st step
k2 = 7.6e3; %forward 2nd step
k3 = 4.84e5; %back 2nd
k4 = 4.58e5; %forward 3rd
k5 = 3.4e4; %back 3rd
k6 = 2.25e5; %forward 4th
k7 = 2.25e5; %back 4th
k8 = 1.38e5; %forward 5th
k9 = 3.32e5; %back 5th
k10 = 1.84e5; %forward 6th
```

```
k11 = 3e5; %back 6th

kLf = 3.09e2; % forward rate for "leak" reaction

kLr = 4.43e5; %reverse rate for "leak" reaction
```

****Step 1****

$$
\begin{aligned}
d(P1) =& -k*(P1)*(T_{1-6}t)+k1*(P_1T_{2-6}t)*(T1);\ \text{[probe1]}\\
d(T_{1-6}t) =& -k*(P1)*(T_{1-6}t)+k1*(P_1T_{2-6}t)*(T1)\\
&-kLf*(T_{1-6}t)*((P2)+(P3)+(P4)+(P5)+(P6))\\
&+kLr*(P_xT_{2-6}t)*(T1);\ \text{[initial state with T1-6 bound to template t]}\\
d(P_1T_{2-6}t) =& +k*(P1)*(T_{1-6}t)-k1*(P_1T_{2-6}t)*(T1)\\
&-k2*(P_1T_{2-6}t)*(P2)+k3*(P_{1-2}T_{3-6}t)*(T2)\\
&-kLf*(P_1T_{2-6}t)*((P1)+(P3)+(P4)+(P5)+(P6))\\
&+kLr*(T2)*(P_{1,x}T_{3-6}t);\ \text{[correct displacement of T1]}\\
d(P_xT_{2-6}t) =& +kLf*(T_{1-6}t)*((P2)+(P3)+(P4)+(P5)+(P6))\\
&-kLr*(P_xT_{2-6}t)*(T1)-k2*(P_xT_{2-6}t)*(P2)\\
&+k3*(P_{x,2}T_{3-6}t)*(T2);\text{[incorrect disp of T1 by Px]}\\
d(T1) =& +k*(P1)*(T_{1-6}t)-k1*(P_1T_{2-6}t)*(T1)\\
&+kLf*(T_{1-6}t)*((P2)+(P3)+(P4)+(P5)+(P6))\\
&-kLr*(P_xT_{2-6}t)*(T1);\ \text{[free target1]}
\end{aligned}
$$

****Step 2****

$$
\begin{aligned}
d(P2) =& -k2*(P_1T_{2-6}t)*(P2)-k2*(P_xT_{2-6}t)*(P2)\\
&+k3*(P_{1-2}T_{3-6}t)*(T2)+k3*(P_{x,2}T_{3-6}t)*(T2);\text{[probe2]}\\
d(P_{1-2}T_{3-6}t) =& +k2*(P_1T_{2-6}t)*(P2)-k3*(P_{1-2}T_{3-6}t)*(T2)\\
&-k4*(P_{1-2}T_{3-6}t)*(P3)+k5*(P_{1-3}T_{4-6}t)*(T3)\\
&-kLf*(P_{1-2}T_{3-6}t)*((P1)+(P2)+(P4)+(P5)+(P6))\\
&+kLr*(P_{1-2,x}T_{4-6}t)*(T3);\text{[correct disp of T1-2]}
\end{aligned}
\tag{A.1}
$$

$$
\begin{aligned}
d(P_{1,x}T_{3-6}t) = &+ kLf * (P_1T_{2-6}t) * ((P1) + (P3) + (P4) + (P5) + (P6)) \\
&- kLr * (T2) * (P_{1,x}T_{3-6}t) - k4 * (P_{1,x}T_{3-6}t) * (P3) \\
&+ k5 * (P_{1,x,3}T_{4-6}t) * (T3); \text{[correct disp of T1, incorrect disp T2]} \\
d(P_{x,2}T_{3-6}t) = &+ k2 * (P_xT_{2-6}t) * (P2) - k3 * (P_{x,2}T_{3-6}t) * (T2) \\
&- k4 * (P_{x,2}T_{3-6}t) * (P3) + k5 * (P_{x,2-3}T_{4-6}t) * (T3); \\
&\text{[incorrect disp T1, correct disp T2]} \\
d(T2) = &+ k2 * (P_1T_{2-6}t) * (P2) + k2 * (P_xT_{2-6}t) * (P2) \\
&+ kLf * (P_1T_{2-6}t) * ((P1) + (P3) + (P4) + (P5) + (P6)) \\
&- k3 * (P_{1-2}T_{3-6}t) * (T2) \\
&- kLr * (T2) * (P_{1,x}T_{3-6}t) - k3 * (P_{x,2}T_{3-6}t) * (T2); \text{[free T2]}
\end{aligned}
$$

****Step 3****

$$
\begin{aligned}
d(P3) = &- k4 * (P_{1-2}T_{3-6}t) * (P3) + k5 * (P_{1-3}T_{4-6}t) * (T3) \\
&- k4 * (P_{1,x}T_{3-6}t) * (P3) + k5 * (P_{1,x,3}T_{4-6}t) * (T3) \\
&- k4 * (P_{x,2}T_{3-6}t) * (P3) + k5 * (P_{x,2-3}T_{4-6}t) * (T3); \text{[probe3]} \\
d(P_{1-3}T_{4-6}t) = &+ k4 * (P_{1-2}T_{3-6}t) * (P3) - k5 * (P_{1-3}T_{4-6}t) * (T3) \\
&- k6 * (P_{1-3}T_{4-6}t) * (P4) + k7 * (P_{1-4}T_{5-6}t) * (T4) \\
&- kLf * (P_{1-3}T_{4-6}t) * ((P1) + (P2) + (P3) + (P5) + (P6)) \\
&+ kLr * (P_{1-3,x}T_{5-6}t) * (T4); \text{[correct disp T1-3]}
\end{aligned}
$$

$$
\begin{aligned}
d(P_{1-2,x}T_{4-6}t) = &+ kLf * (P_{1-2}T_{3-6}t) * ((P1) + (P2) + (P4) + (P5) + (P6)) \\
&- kLr * (P_{1-2,x}T_{4-6}t) * (T3) \\
&- k6 * (P_{1-2,x}T_{4-6}t) * (P4) + k7 * (P_{1-2,x,4}T_{5-6}t) * (T4); \\
&[\text{correct disp T1-2, incorrect disp T3}] \\
d(P_{1,x,3}T_{4-6}t) = &+ k4 * (P_{1,x}T_{3-6}t) * (P3) - k5 * (P_{1,x,3}T_{4-6}t) * (T3) \\
&- k6 * (P_{1,x,3}T_{4-6}t) * (P4) + k7 * (P_{1,x,3-4}T_{5-6}t) * (T4); \\
&[\text{correct disp T1, incorrect disp T2, correct disp T3}] \\
d(P_{x,2-3}T_{4-6}t) = &+ k4 * (P_{x,2}T_{3-6}t) * (P3) - k5 * (P_{x,2-3}T_{4-6}t) * (T3) \\
&- k6 * (P_{x,2-3}T_{4-6}t) * (P4) + k7 * (P_{x,2-4}T_{5-6}t) * (T4); \\
&[\text{incorrect disp T1, correct disp T2-3}] \\
d(T3) = &+ k4 * (P_{1-2}T_{3-6}t) * (P3) - k5 * (P_{1-3}T_{4-6}t) * (T3) \\
&+ kLf * (P_{1-2}T_{3-6}t) * ((P1) + (P2) + (P4) + (P5) + (P6)) \\
&- kLr * (P_{1-2,x}T_{4-6}t) * (T3) + k4 * (P_{1,x}T_{3-6}t) * (P3) \\
&- k5 * (P_{1,x,3}T_{4-6}t) * (T3) + k4 * (P_{x,2}T_{3-6}t) * (P3) \\
&- k5 * (P_{x,2-3}T_{4-6}t) * (T3); [\text{free T3}]
\end{aligned}
$$

****Step 4****

$$
\begin{aligned}
d(P4) = &- k6 * (P_{1-3}T_{4-6}t) * (P4) + k7 * (P_{1-4}T_{5-6}t) * (T4) \\
&- k6 * (P_{1-2,x}T_{4-6}t) * (P4) + k7 * (P_{1-2,x,4}T_{5-6}t) * (T4) \\
&- k6 * (P_{1,x,3}T_{4-6}t) * (P4) + k7 * (P_{1,x,3-4}T_{5-6}t) * (T4) \\
&- k6 * (P_{x,2-3}T_{4-6}t) * (P4) + k7 * (P_{x,2-4}T_{5-6}t) * (T4); [\text{probe4}] \\
d(P_{1-4}T_{5-6}t) = &+ k6 * (P_{1-3}T_{4-6}t) * (P4) - k7 * (P_{1-4}T_{5-6}t) * (T4) \\
&- k8 * (P_{1-4}T_{5-6}t) * (P5) + k9 * (T5) * (P_{1-5}T_6t) \\
&- (kLf) * (P_{1-4}T_{5-6}t) * ((P1) + (P2) + (P3) + (P4) + (P6)) \\
&+ (kLr) * (P_{1-4,x}T_6t) * (T5); [\text{correct disp T1-4}]
\end{aligned}
$$

$$
\begin{aligned}
d(P_{1-3,x}T_{5-6}t) = &+ kLf * (P_{1-3}T_{4-6}t) * ((P1) + (P2) + (P3) + (P5) + (P6)) \\
&- kLr * (P_{1-3,x}T_{5-6}t) * (T4) - k8 * (P_{1-3,x}T_{5-6}t) * (P5) \\
&+ k9 * (P_{1-3,x,5}T_6t) * (T5); \text{[correct disp T1-3, incorrect disp T4]} \\
d(P_{1-2,x,4}T_{5-6}t) = &+ k6 * (P_{1-2,x}T_{4-6}t) * (P4) - k7 * (P_{1-2,x,4}T_{5-6}t) * (T4) \\
&- k8 * (P_{1-2,x,4}T_{5-6}t) * (P5) + k9 * (P_{1-2,x,4-5}T_6t) * (T5); \\
&\text{[correct disp T1-2, incorrect disp T3, correct disp T4]} \\
d(P_{1,x,3-4}T_{5-6}t) = &+ k6 * (P_{1,x,3}T_{4-6}t) * (P4) - k7 * (P_{1,x,3-4}T_{5-6}t) * (T4) \\
&- k8 * (P1P_xP3 - 4T_{5-6}t) * (P5) + k9 * (P_{1,x,3-5}T_6t) * (T5); \\
&\text{[correct disp T1, incorrect disp T2, correct disp T3-4]} \\
d(P_{x,2-4}T_{5-6}t) = &+ k6 * (P_{x,2-3}T_{4-6}t) * (P4) - k7 * (P_{x,2-4}T_{5-6}t) * (T4) \\
&- k8 * (P_{x,2-4}T_{5-6}t) * (P5) + k9 * (P_{x,2-5}T_6t) * (T5); \\
&\text{[incorrect disp T1, correct disp T2-4]} \\
d(T4) = &+ k6 * (P_{1-3}T_{4-6}t) * (P4) - k7 * (P_{1-4}T_{5-6}t) * (T4) \\
&+ kLf * (P_{1-3}T_{4-6}t) * ((P1) + (P2) + (P3) + (P5) + (P6)) \\
&- kLr * (P_{1-3,x}T_{5-6}t) * (T4) + k6 * (P_{1-2,x}T_{4-6}t) * (P4) \\
&- k7 * (P_{1-2,x}P4T_{5-6}t) * (T4) + k6 * (P_{1,x,3}T_{4-6}t) * (P4) \\
&- k7 * (P_{1,x,3-4}T_{5-6}t) * (T4) + k6 * (P_{x,2-3}T_{4-6}t) * (P4) \\
&- k7 * (P_{x,2-4}T_{5-6}t) * (T4) - k8 * (P_{x,2-4}T_{5-6}t) * (P5) \\
&+ k9 * (P_{x,2-5}T_6t) * (T5); \text{[free T4]}
\end{aligned}
$$

****Step 5****

$$
\begin{aligned}
d(P5) = &- k8 * (P_{1-4}T_{5-6}t) * (P5) + k9 * (T5) * (P_{1-5}T_6t) \\
&- k8 * (P_{1-3,x}T_{5-6}t) * (P5) + k9 * (P_{1-3,x,5}T_6t) * (T5) \\
&- k8 * (P_{1-2,x,4}T_{5-6}t) * (P5) + k9 * (P_{1-2,x,4-5}T_6t) * (T5) \\
&- k8 * (P_{1,x,3-4}T_{5-6}t) * (P5) + k9 * (P_{1,x,3-5}T_6t) * (T5); \text{[probe5]}
\end{aligned}
$$

$$
\begin{aligned}
d(P_{1-5}T_6t) = &+ k8 * (P_{1-4}T_{5-6}t) * (P5) - k9 * (T5) * (P_{1-5}T_6t) \\
&- k10 * (P_{1-5}T_6t) * (P6) + k11 * (P_{1-6}t) * (T6) \\
&- (kLf) * (P_{1-5}T_6t) * ((P1) + (P2) + (P3) + (P4) + (P5)) \\
&+ (kLr) * (P_{1-5,x}t) * (T6); \text{[correct disp T1-5]} \\
d(P_{1-4,x}T_6t) = &+ kLf * (P_{1-4}T_{5-6}t) * ((P1) + (P2) + (P3) + (P4) + (P6)) \\
&- kLr * (P_{1-4,x}T_6t) * (T5) - k10 * (P_{1-4,x}T_6t) * (P6) \\
&+ k11 * (P_{1-4,x,6}t) * (T6); \text{[correct disp T1-4, incorrect disp T5]} \\
d(P_{1-3,x,5}T_6t) = &+ k8 * (P_{1-3,x}T_{5-6}t) * (P5) - k9 * (P_{1-3,x,5}T_6t) * (T5) \\
&- k10 * (P_{1-3,x,5}T_6t) * (P6) + k11 * (P_{1-3,x,5-6}t) * (T6); \\
&\text{[correct disp T1-3, incorrect disp T4, correct disp T5]} \\
d(P_{1-2,x,4-5}T_6t) = &+ k8 * (P_{1-2,x,4}T_{5-6}t) * (P5) - k9 * (P_{1-2,x,4-5}T_6t) * (T5) \\
&- k10 * (P_{1-2,x,4-5}T_6t) * (P6) + k11 * (P_{1-2,x,4-6}t) * (T6); \\
&\text{[correct disp T1-2, incorrect disp T3, correct disp T4-5]} \\
d(P_{1,x,3-5}T_6t) = &+ k8 * (P_{1,x,3-4}T_{5-6}t) * (P5) - k9 * (P_{1,x,3-5}T_6t) * (T5) \\
&- k10 * (P_{1,x,3-5}T_6t) * (P6) + k11 * (P_{1,x,3-6}t) * (T6); \\
&\text{[correct disp T1, incorrect disp T2, correct disp T3-5]} \\
d(P_{x,2-5}T_6t) = &+ k8 * (P_{x,2-4}T_{5-6}t) * (P5) - k9 * (P_{x,2-5}T_6t) * (T5) \\
&- k10 * (P_{x,2-5}T_6t) * (P6) + k11 * (P_{x,2-6}t) * (T6); \\
&\text{[incorrect disp T1, correct disp T2-5]} \\
d(T5) = &+ k8 * (P_{1-4}T_{5-6}t) * (P5) - k9 * (T5) * (P_{1-5}T_6t) \\
&+ kLf * (P_{1-4}T_{5-6}t) * ((P1) + (P2) + (P3) + (P4) + (P6)) \\
&- kLr * (P_{1-4,x}T_6t) * (T5) + k8 * (P_{1-3,x}T_{5-6}t) * (P5) \\
&- k9 * (P_{1-3,x,5}T_6t) * (T5) + k8 * (P_{1-2,x,4}T_{5-6}t) * (P5) \\
&- k9 * (P_{1-2,x,4-5}T_6t) * (T5) + k8 * (P_{1,x,3-4}T_{5-6}t) * (P5) \\
&- k9 * (P_{1,x,3-5}T_6t) * (T5) + k8 * (P_{x,2-4}T_{5-6}t) * (P5) \\
&- k9 * (P_{x,2-5}T_6t) * (T5); \text{[free T5]}
\end{aligned}
$$

****Step 6****

$$
\begin{aligned}
d(P6) = &- k10 * (P_{1-5}T_6t) * (P6) + k11 * (P_{1-6}t) * (T6) \\
&- k10 * (P_{1-4,x}T_6t) * (P6) + k11 * (P_{1-4,x,6}t) * (T6) \\
&- k10 * (P_{1-3,x,5}T_6t) * (P6) + k11 * (P_{1-3,x,5-6}t) * (T6) \\
&- k10 * (P_{1-2,x,4-5}T_6t) * (P6) + k11 * (P_{1-2,x,4-6}t) * (T6) \\
&- k10 * (P_{1,x,3-5}T_6t) * (P6) + k11 * (P_{1,x,3-6}t) * (T6) \\
&- k10 * (P_{x,2-5}T_6t) * (P6) + k11 * (P_{x,2-6}t) * (T6); \text{[probe6]} \\
d(P_{1-6}t) = &+ k10 * (P_{1-5}T_6t) * (P6) - k11 * (P_{1-6}t) * (T6); \text{[correct disp T1-6]} \\
d(P_{1-5,x}t) = &+ (kLf) * (P_{1-5}T_6t) * ((P1) + (P2) + (P3) + (P4) + (P5)) \\
&- (kLr) * (P_{1-5,x}t) * (T6); \text{[correct disp T1-5, incorrect disp T6]} \\
d(P_{1-4,x,6}t) = &+ k10 * (P_{1-4,x}T_6t) * (P6) - k11 * (P_{1-4,x,6}t) * (T6); \\
&\text{[correct disp T1-4, incorrect disp T5, correct disp T6]} \\
d(P_{1-3,x,5-6}t) = &+ k10 * (P_{1-3,x,5}T_6t) * (P6) - k11 * (P_{1-3,x,5-6}t) * (T6); \\
&\text{[correct disp T1-3, incorrect disp T4, correct disp T5-6]} \\
d(P_{1-2,x,4-6}t) = &+ k10 * (P_{1-2,x,4-5}T_6t) * (P6) - k11 * (P_{1-2,x,4-6}t) * (T6); \\
&\text{[correct disp T1-2, incorrect disp T3, correct disp T4-6]} \\
d(P_{1,x,3-6}t) = &+ k10 * (P_{x,3-5}T_6t) * (P6) - k11 * (P_{1,x,3-6}t) * (T6); \\
&\text{[correct disp T1, incorrect disp T2, correct disp T3-6]} \\
d(P_{x,2-6}t) = &+ k10 * (P_{x,2-5}T_6t) * (P6) - k11 * (P_{x,2-6}t) * (T6); \\
&\text{[incorrect disp T1, correct disp T2-6]} \\
d(T6) = &+ k10 * (P_{1-5}T_6t) * (P6) - k11 * (P_{1-6}t) * (T6) \\
&+ (kLf) * (P_{1-5}T6t) * ((P1) + (P2) + (P3) + (P4) + (P5)) \\
&- (kLr) * (P_{1-5,x}t) * (T6) \\
&+ k10 * (P_{1-4,x}T_6t) * (P6) - k11 * (P_{1-4,x,6}t) * (T6) \\
&+ k10 * (P_{1-3,x,5}T_6t) * (P6) - k11 * (P_{1-3,x,5-6}t) * (T6) \\
&+ k10 * (P_{1-2,x,4-5}T_6t) * (P6) - k11 * (P_{1-2,x,4-6}t) * (T6) \\
&+ k10 * (P_{1,x,3-5}T_6t) * (P6) - k11 * (P_{1,x,3-6}t) * (T6) \\
&+ k10 * (P_{x,2-5}T_6t) * (P6) - k11 * (P_{x,2-6}t) * (T6); \text{[free T6]}
\end{aligned}
$$

A.3 Wittig Reaction Preparation

Succinimidyl 4-formylbenzoate (SFB) was purchased from Thermo Fisher Scientific. 15 % polyacrylamide gels were run at 4 °C in TAE buffer at 240 V. Denaturing gels were run as a two-layer discontinuous Tris-glycine buffer system: stacking gel (5 % acrylamide, 7 M urea, 25 %, 125 mM Tris, pH 6.8); separating gel (15 % acrylamide, 7 M urea, 25 %, 400 mM Tris, pH 8.8), running buffer (2.5 mM Tris, 0.19 M glycine). Samples were combined 1:1 with the loading buffer and heated to 95 °C for 5 min before being run on the gel at room temperature at 300 V.

Diphenylphosphine benzoic acid (DPBA) diluted in dimethyl formamide (DMF, a common solvent) (300 mM, 100 µl), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) in DMF (300 mM, 100 µl), and N-hydroxybenzotriazole (HOBt) in DMF (300 mM, 100 µl) were mixed together under an argon atmosphere for 1 hour to produce the activated ester. Amino modified oligonucleotides were diluted in phosphate buffer (PBS, 100 mM sodium phosphate, pH 7.5, 100 mM NaCl). The activated ester was added to the DNA and stirred under an argon atmosphere overnight (for approximately 18 hours). Excess small molecules were removed using an Illustra NAP-5 column (GE Healthcare), eluting with PBS buffer to give the triphenylphosphine modified oligonucleotides. These were used without any further purification.

The triphenylphosphino modified DNA strands were combined with iodo adaptors to form the DNA Wittig adaptors required for the test reaction. Excess small molecules were removed using an Illustra NAP-5 column, eluting with water. The collected DNA was concentrated in vacuo and purified by HPLC.

Synthesis of Adaptor 1 - FAM

Triphenylphosphino modified oligonucleotide A1 in PBS buffer was mixed in DMF with 5-(iodoacetamido)fluorescein to produce Adaptor 1 - FAM.

Synthesis of Adaptor 2 - BAL

A2 was diluted in PBS buffer, and 4-formlybenzoic acid N-hydroxysuccinimide ester in DMF was added and mixed for 2 hours at room temperature.

A.3.1 HPLC purification and analysis

HPLC analysis was performed using an Agilent 1200 system. Chromatography was performed on a Waters XBridge OST C18 2.5 µm particle size 4.6 × 50 mm column heated to 40 °C. Flow rate was set at 1 ml/min, with a linear gradient of two buffers.

* Buffer C is 0.1 M triethylammonium acetate, 5 % acetonitrile, pH 7.0;
* Buffer D is 0.1 M triethylammonium acetate, 70 % acetonitrile, pH 7.0.

Fractions were collected using a DNA120 Savant SpeedVac Concentrator. The concentration of the purified oligonucleotides was measured using a Cary 50 Probe UV-Vis Spectrophotometer to determine the UV absorbance at 260 nm.

A.3.2 Sample Preparation

Samples for the cascade-coupling Wittig tests were prepared by annealing the strands that form the initial duplex by heating to 85 °C and cooling to 20 °C over 10 min. Reactions were carried out in TAPS buffer: 100 mM TAPS (N-Tris(hydroxymethyl)methyl-3-aminopropanesulfonic acid), 1 M NaCl, pH 8.5. Reactions were incubated for two hours before analysis by non-denaturing PAGE.

A.4 Cascade: DNA Strands Used

The probe and target strand toehold domains are indicated by the use of upper case letters, while the universal domains are shown in lower case letters. Reporter strand domains are shown using italic lower case letters.

One-Step Reaction with Variable Target Toehold Length

Template Strand 0 5' *atctgaagctctgcgattcattaa* GCCAGCGT ccgtgatatgcctcga CTACTATTC-CTT

Reporter Strand 1 5' Cy5 + *ttaatgaatcgcagagcttcagat*

Target Strand 2 5' tcgaggcatatcacgg ACGCTGGC + Cy3

Probe Strand 3a 5' AAGGAATAGTAG tcgaggcatatcacggacgc

Probe Strand 3b 5' AAGGAATAGTAG tcgaggcatatcacggacg

Probe Strand 3c 5' AAGGAATAGTAG tcgaggcatatcacggac

Probe Strand 3d 5' AAGGAATAGTAG tcgaggcatatcacgga

Probe Strand 3e 5' AAGGAATAGTAG tcgaggcatatcacgg

One-Step Reaction with Equal Length Probe and Target Toeholds

Perfect Template Strand 5' *tagcgcctcctcatatca* ttt CGGTTG caggaagtattaagacga CGGTTG

Similar Template Strand 5' *tagcgcctcctcatatca* ttt CGGTTG caggaagtattaagacga AGCGAT

Uphill Template Strand 5' *tagcgcctcctcatatca* ttt CGGTTG caggaagtattaagacga CTATCG

Downhill Template Strand 5' *tagcgcctcctcatatca* ttt CGGTTG caggaagtattaagacga GCT-GCG

Reporter Strand 5' Fam + *tgatatgaggaggcgcta*

Target Strand 5' tcgtcttaatacttcctg CAACCG + Tamra

Perfect Probe Strand 5' CAACCG tcgtcttaatacttcctg

Similar Probe Strand 5' ATCGCT tcgtcttaatacttcctg

Uphill Probe Strand 5' CGATAG tcgtcttaatacttcctg

Downhill Probe Strand 5' CGCAGC tcgtcttaatacttcctg

Two-Step *Identical* Reaction

Template Strand 5' *tagcgcctcctcatatca* ttt CGGTTG caggaagtgttaagacga CGGTTG caggaagt-gttaagacga CGGTTG

Reporter Strand 5' Fam + *tgatatgaggaggcgcta*

Target Strands 5' tcgtcttaatacttcctg CAACCG + Tamra

Probe Strands 5' CAACCG tcgtcttaatacttcctg

Two-Step *Different* Reaction

Template Strand 5' *tagcgcctcctcatatca* ttt CGGTTG caggaagtgttaagacga AGCGAT caggaagt-

gttaagacga ATGCTG

Reporter Strand 5' Fam + *tgatatgaggaggcgcta*

Target Strand 1 5' tcgtcttaatacttcctg ATCGCT

Target Strand 2 5' tcgtcttaatacttcctg CAACCG + Tamra

Probe Strand 1 5' CAGCAT tcgtcttaatacttcctg

Probe Strand 2 5' ATCGCT tcgtcttaatacttcctg

Six-Step Cascade

Template Strand 5' *ttcgtatctgcgactcag* GCGCAA gtccgtgatatgcctcga GGGCCA gtccgtgatatgcctcga CGCTCG gtccgtgatatgcctcga GCCGTC gtccgtgatatgcctcga CGCGTT gtccgtgatatgcctcga ACGCCA gtccgtgatatgcctcga CTACTATTCCTT

Reporter Strand 5' Cy5 + *ctgagtcgcagatacgaa*

T6 5' tcgaggcatatcacggac TTGCGC

T5 5' tcgaggcatatcacggac TGGCCC

T4 5' tcgaggcatatcacggac CGAGCG

T3 5' tcgaggcatatcacggac GACGGC

T2 5' tcgaggcatatcacggac AACGCG

T1 5' tcgaggcatatcacggac TGGCGT

12 nt P0 5' AAGGAATAGTAG tcgaggcatatcacggac

9 nt P0 5' GAATAGTAG tcgaggcatatcacggac

6 nt P0 5' TAGTAG tcgaggcatatcacggac

P1 5' TGGCGT tcgaggcatatcacggac

P2 5' AACGCG tcgaggcatatcacggac

P3 5' GACGGC tcgaggcatatcacggac

P4 5' CGAGCG tcgaggcatatcacggac

P5 5' TGGCCC tcgaggcatatcacggac

Six-Step Shuffled Cascade

Shuffled Template Strand 5' *ttcgtatctgcgactcag* GGGCCA gtccgtgatatgcctcga CGCGTT gtccgtgatatgcctcga GCCGTC gtccgtgatatgcctcga CGCTCG gtccgtgatatgcctcga ACGCCA gtccgtgatatgcctcga GCGCAA gtccgtgatatgcctcga CTACTATTCCTT

Reporter Strand 5' Cy5 + *ctgagtcgcagatacgaa*

T5 5' tcgaggcatatcacggac TGGCCC

T2 5' tcgaggcatatcacggac AACGCG

T3 5' tcgaggcatatcacggac GACGGC

T4 5' tcgaggcatatcacggac CGAGCG

T1 5' tcgaggcatatcacggac TGGCGT

T6 5' tcgaggcatatcacggac TTGCGC + Cy3

12 nt P0 5' AAGGAATAGTAG tcgaggcatatcacggac

9 nt P0 5' GAATAGTAG tcgaggcatatcacggac

6 nt P0 5' TAGTAG tcgaggcatatcacggac

P6 5' TTGCGC tcgaggcatatcacggac

P1 5' TGGCGT tcgaggcatatcacggac

P4 5' CGAGCG tcgaggcatatcacggac

P3 5' GACGGC tcgaggcatatcacggac

P2 5' AACGCG tcgaggcatatcacggac

Six-Step Minimal Secondary Structure Cascade

Template 5' *tagcgcctcctcatatca* ttt CGGTTG caggaagtgttaagacga AGCGAT caggaagtgttaagacga ATGCTG caggaagtgttaagacga CAGCAT caggaagtgttaagacga CAGCGA caggaagtgttaagacga CTATCG caggaagtgttaagacga GCAGCG

Reporter 5' Fam + *tgatatgaggaggcgcta*

T6 5' tcgtcttaatacttcctg CAACCG + Tamra

T5 5' tcgtcttaatacttcctg ATCGCT

T4 5' tcgtcttaatacttcctg CAGCAT

T3 5' tcgtcttaatacttcctg ATGCTG

T2 5' tcgtcttaatacttcctg ACGCTG

T1 5' tcgtcttaatacttcctg CGATAG

P5 5' ATCGCT tcgtcttaatacttcctg

P4 5' CAGCAT tcgtcttaatacttcctg

P3 5' ATGCTG tcgtcttaatacttcctg

P2 5' ACGCTG tcgtcttaatacttcctg

P1 5' CGATAG tcgtcttaatacttcctg

P0 5’ CGCTGC tcgtcttaatacttcctg

Rolling Mechanism

Rolling Template L2R 5’ gtccgtgatcctcga CGCGTT

Rolling Template R2L 5’ CGCGTT gtccgtgatcctcga

Dual-labelled Universal Target Strand 5’ Cy3 + tcgaggcatatcacggac + Cy5

Displacer T2 5’ tcgaggcatatcacggac AACGCG

Displacer P2 5’ AACGCG tcgaggcatatcacggac

Wittig Cascade-Coupling Experiment

Template 5’ AGCGAT caggaagtattaagacga CGGTTG caggaagtattaagacga CTATCG

Adaptor 1 - FAM 5’ amine + ttccta tcgtcttaatacttcctg CAACCG

Adaptor 2 - BAL 5’ amine + ttccta tcgtcttaatacttcctg ATCGCT

Displacer 1 5’ CGATAG tcgtcttaatacttcctg

Appendix B

General Experimental Protocols for the Origami

The details of the design of the DNA origami tiles presented in Chapter 3 and the protocols for forming the origami are given here.

B.1 DNA Origami Design

The foundations for the design of the origami tiles presented in Chapter 3 are taken from [Rothemund, 2006]. Two designs were used, one which folded a small segment of the M13mp18 DNA strand, and one which used the full length of the strand.

B.1.1 The Small Tile used for 5'–5' Click

The tile was constructed using 960 nt of the M13mp18 strand, and the particular segment used was chosen as it has weak secondary structure. The tile consists of 10 rows of 96 nt each, and thirty 32 nt staple strands were required to form the structure. Two sets of staple strands were designed, one which simply folded the M13mp18 strand into a regtangular tile, and the other which included modified staples for the FRET/click experiments.

The origami rectangle is divided into a grid, with six columns of 16 nt each, and 10

rows. The name of each staple strand is a grid reference explaining where it is positioned in the rectangle with respect to the other staples. Each staple strand hybridises to three rows of the template strand. For example, "sA1" is the staple that occupies row A and column 1 (shown in red in Figure B.1); "sA2" is the next staple along the top of the rectangle (green); "sB1" is the second staple down from the top, forming the left-hand edge (blue). The following staples form the 960 nt origami.

Row	Column	Name	Sequence
A	1	sA1	acctccgg cttaggtt gggttata gtacataa
B	1	sB1	atcaatat ttttttaa tggaaaca cagtacct
C	1	sC1	tttacatc gatgaata tacagtaa aaggagcg
D	1	sD1	gaattatc acaaagaa accaccag ggttatct
E	1	sE1	aaaatatc gaaaggaa ttgaggaa taccgaac
A	2	sA2	gaattacc atgtgagt gaataacc ccttttta
B	2	sB2	taacgtca gggagaaa caataacg tttcattt
C	2	sC2	tttgcgga atcatatt cctgatta ttcaggtt
D	2	sD2	caacagta tttaggtg cactaaca attatcat
E	2	sE2	gaaccacc agcagaag ataaaaca tggcaaat
A	3	sA3	caaaatca taggtctg agagacta ttgcttct
B	3	sB3	gtaaatcg taattaca tttaacaa gattcgcc
C	3	sC3	tgattgct agaaattg cgtagatt tcagatga
D	3	sD3	tggcaatt taaaagtt tgagtaac actaatag
E	3	sE3	attagagc tcaatatc tggtcagt gaggtgag
A	4	sA4	aacaaaat tcgctatt aattaatt gaatttat
B	4	sB4	agaaataa ttgaatac caagttac atcaagaa
C	4	sC4	attaattt catcaata taatcctg cgtaaaac
D	4	sD4	accctcaa cgtcaata gataatac cgaacgtt

Row	Column	Name	Sequence
E	4	sE4	gcggtcag tattaaca ccgcctgc aatatcaa
A	5	sA5	aagacgct gagaagag tcaatagt ttccctta
B	5	sB5	gaatcctt agatgatg aaacaaac aaaatcgc
C	5	sC5	gcagaggc atcaaaat tattagca attgtttg
D	5	sD5	gattatac tattaaat cctttgcc atttgagg
E	5	sE5	atttagaa accttgct gaacctca aacagtgc
A	6	sA6	gcaaaaga gaaaacat agcgatag cttagatt
B	6	sB6	ccaaccat gaattatt catttcaa ttacctga
C	6	sC6	acaactcg ttctgaat tatggaag gaattgaa
D	6	sD6	aaagcatc gtattaga ctttacaa acaattcg
E	6	sE6	cacgctga gagccagc agcaaatg aaaaatct

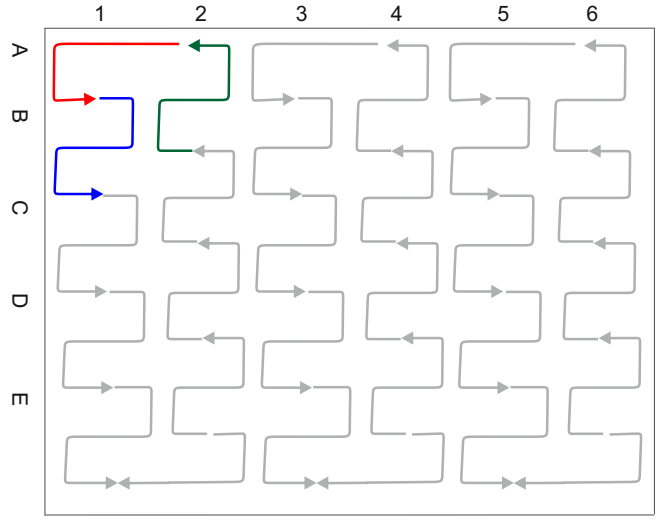


Figure B.1: Diagram illustrating the 960 nt origami tile.

The modified staple strands for the origami-templated click chemistry include adaptors (onto which either fluorophores or click adaptors are attached at the 5' end), extended strands and fixer strands. The strands are listed in order of position A-F.

Adaptors

Ad(sB2) 5' GCAAAAAT CAATACAGGG T taacgtca gggagaaa caataacg tttcattt

Ad(sC2) 5' TAMRA + GGATAAGA AACAGATGAA T tttgcgga atcatatt cctgatta ttcaggtt

Ad(sE3) 5' CCGCCGTA AACCAAGACT T attagagc tcaatatc tggtcagt gaggtgag

Ad(sC4) 5' IOWA BLACK + CAGACGAC CACTACCACG T attaattt catcaata taatcctg cgtaaaac

Ad(sB4) 5' TGAATACA CACCCAAAAG T agaaataa ttgaatac caagttac atcaagaa (T AGAATTTT)

Ad(sC3) 5' FAM + TCAACTTC AACACGCCCT T tgattgct agaaattg cgtagatt tcagatga

Extended

Ex(sB1) atcaatat ttttttaa tggaaaca cagtacct T AAAGAGGT

Ex(sD1) gaattatc acaaagaa accaccag ggttatct T GAAGGAAA

Ex(sE2) gaaccacc agcagaag ataaaaca tggcaaat T CTACGCTG

Ex(sD5) gattatac tattaaat cctttgcc atttgagg T GGAGATTG

Ex(sB5) gaatcctt agatgatg aaacaaac aaaatcgc T GGGCTAGA

Ex(sB4) agaaataa ttgaatac caagttac atcaagaa T AGAATTTT (No need to buy - buy extended version of Ad(sB4))

Fixers

F(sC1) CCCTGTATTG ACCTCTTT T tttacatc gatgaata tacagtaa aaggagcg

F(sE1) TTCATCTGTT TTTCCTTC T aaaatatc gaaaggaa ttgaggaa taccgaac

F(sD2) AGTCTTGGTT CAGCGTAG T caacagta tttaggtg cactaaca attatcat

F(sE5) CGTGGTAGTG CAATCTCC T atttagaa accttgct gaacctca aacagtgc

F(sC5) CTTTTGGGTG TCTAGCCC T gcagaggc atcaaaat tattagca attgtttg

F(sA4) AGGGCGTGTT AAAATTCT T aacaaaat tcgctatt aattaatt gaatttat

Sticky-end displacers

disp_B_stic_6 TTCATCTGTT TCTTATCC gcggcg

disp_B_stic_8 TTCATCTGTT TCTTATCC gcggcgcc

disp_F_stic_6 AGGGCGTGTT GAAGTTGA cgccgc

disp_F_stic_8 AGGGCGTGTT GAAGTTGA ggcgccgc

disp_B_pal_6 TTCATCTGTT TCTTATCC gcgcgc

disp_B_pal_8 TTCATCTGTT TCTTATCC cgcgcgcg

disp_F_pal_6 AGGGCGTGTT GAAGTTGA gcgcgc

disp_F_pal_8 AGGGCGTGTT GAAGTTGA cgcgcgcg

Strands for the click chemistry mechanism

Ad(sC2) 5' Azide + GGATAAGA AACAGATGAA T tttgcgga atcatatt cctgatta ttcaggtt

Ad(sC3) 5' Alkyne + TCAACTTC AACACGCCCT T tgattgct agaaattg cgtagatt tcagatga

compB 5' TTCATCTGTT TCTTATCC g

compF 5' AGGGCGTGTT GAAGTTGA c

B.1.2 The Minimal-Twist Origami used for 5'–3' Click

The staple strand sequences required to fold the M13mp18 ssDNA into a rectangular origami tile are:

The modified staples used in the FRET/click experiments are:

4nt sticky-ends

Position 2 (5' azide)

Fixer 2 (r3t10e) 5' TTTATCCCCGGGAGAATTAACTGAAGTTACC + T + TTC ATC TGT TTC TT AGCA + GGCACTCCGCTCTCTC 3'

Extended 2 (r3t8f) 5' AATAGT + GAGAGAGCGGAGTGCC + TGCT + T + AAAAT-ACAACCGAGGAAACGCAAAGGGAAGC 3'

Displacer 2_4nt 5' AGCA GGCACTCCGCTCTCTC TTATCA 3'

Position 5 (3' alkyne)

Fixer 5 (r5t10e) 5' ACGGTAGCAGTAGGTA + TGCT TG GGT GTG TAT TCA + T +

Row	Column	Name	Sequence
0	1	r1t0g	attccacagacagccctcatagttagcgtaa
1	1	r1t2f	ggattaggagaggctgagactcctgcataacc
2	1	r1t2e	aagtattaattagcggggttttgcctgtagc
3	1	r1t4f	gcattgacgaaccaccaccagagcggagattt
4	1	r1t4e	cgccaccaaggaggttgaggcagaaacatga
5	1	r1t6f	ttagcaagaatcaccagtagcacattgggct
6	1	r1t6e	gccagcaagccggaaacgtcaccctcagagc
7	1	r1t8f	tccttattaaaagaactggcatgatagcgag
8	1	r1t8e	gaatacccacgcagtatgttagcgaattaga
9	1	r1t10f	tacagagagtcaaaaatgaaaataggaagcaa
10	1	r1t10e	tgtttaacgaataacataaaaactaataacg
11	1	r1t12f	gcgcccaattttcatcgtaggaaaggtggca
12	1	r1t12e	cgtttttatagcaagcaaatcagcgattttt
13	1	r1t14f	ccagtaatatttaggcagaggcaaatatgat
14	1	r1t14e	aacatgtaaagagaatataaagtaaagcaagc
15	1	r1t16f	ccaatcgctatatgtaaatgctgccaatagg
16	1	r1t16e	tatataacaagacaaagaacgcgacaacgcc

A + TTTGCCAGTCAGAGGGTAATTGAGTTTTTAA 3’

Extended 5 (r7t10f) 5’ TCAGAGAGACGAGCGTCTTTCCAGGAGGTTT + T + T + AGCA + TACCTACTGCTACCGT + CTATCG 3’

Displacer 5_4nt 5’ CGATAG ACGGTAGCAGTAGGTA TGCT 3’

6nt sticky-ends

Position 2 (5’ azide)

Fixer 2 (r3t10e) 5’ TTTATCCCCGGGAGAATTAACTGAAGTTACC + T + TTC ATC TGT TTC TT AGCAAG + GGCACTCCGCTCTC 3’

Row	Column	Name	Sequence
17	1	r1t18f	tcaagaaaaaaagaagatgatgattcaggct
18	1	r1t18e	acctgagcacaaaattaattacatggttgggt
19	1	r1t20f	atggcaatatcatattcctgattcataaagt
20	1	r1t20e	gaattatctcatcaatataatcctttcaatt
21	1	r1t22f	caatcaatctgaacctcaaatattggttccg
22	1	r1t22e	tcaccttgatctggtcagttggcaaaggagcg
23	1	r1t24h	taataaaagggacattctggccactaaagca
0	3	r3t0g	tcgtcaccagtacaaactacaacgctcagtac
1	3	r3t2f	caggcggagaacctattattctggtcagacg
2	3	r3t2e	ctatttcgtaagtgccgtcgagaactgagtt
3	3	r3t4f	attggcctctcagagccaccaccaatgaaac
4	3	r3t4e	ccgccacctgatattcacaaacaacccccctgc
5	3	r3t6f	catcgatagacttgagccatttggaaacgtag
6	3	r3t6e	ccgtcaccgcagcaccgtaatcacctcagaa
7	3	r3t8f	aaaatacaaccgaggaaacgcaaagggaagc
8	3	r3t8e	agaaggaatacataaaggtggcaaattatca
9	3	r3t10f	gcattagaaatccaaataagaaaatatagaa
10	3	r3t10e	tttatccccgggagaattaactgaagttacc
11	3	r3t12f	ggcttatcgcactcatcgagaacccgacaaa
12	3	r3t12e	caagtacccggtattctaagaacgccatatta
13	3	r3t14f	aggtaaaggaatcgccatatttaagaaaact
14	3	r3t14e	ttaattgataattctgtccagactattaaac
15	3	r3t16f	ttttcaaattaacctccggcttattaacaat
16	3	r3t16e	ctacctttttatattttagttaattagtagggc
17	3	r3t18f	ttcatttggaggcgaattattcatgattgtt
18	3	r3t18e	atcgcgcaaattacctttttttaactgagaga

Row	Column	Name	Sequence
19	3	r3t20f	tggattatacaaagaaaccaccagaatcaaca
20	3	r3t20e	tttgcggaacttctgaataatgggttacaaa
21	3	r3t22f	gttgaaagcagcaaatgaaaaatacagagat
22	3	r3t22e	agagccaggaattgaggaaggttattatcat
23	3	r3t24h	agaacccttctgacctgaaagcgccacgctg
0	5	r5t0g	aataggaacccatgtaccgtaacgggttgat
1	5	r5t2f	ataagtattataaacagttaatgataaatcc
2	5	r5t2e	agtgcccgagcccggaataggtgtgcaagccc
3	5	r5t4f	tcattaaacctcagagccgccacgtagcgac
4	5	r5t4e	accgcctcgccagaatggaaagctgagtaac
5	5	r5t6f	agaatcaatattcattaaaggtgacatataa
6	5	r5t6e	acggaaatgtttgcctttagcgtccaccgga
7	5	r5t8f	aagaaacggcagatagccgaacaaacaccct
8	5	r5t8e	aaagtaacaaagacaccacggaaaaatattg
9	5	r5t10f	gaacaaagttacaaaataaacagcgaggcgt
10	5	r5t10e	tttgccagtcagagggtaattgagtttttaag
11	5	r5t12f	tttagcgacattccaagaacggggacgacaa
12	5	r5t12e	ttccttatacctcccgacttgcggagcctaa
13	5	r5t14f	taaacaacaaagccaacgctcaactcatcttc
14	5	r5t14e	accagtatatgttcagctaatgcggctgtct
15	5	r5t16f	tgacctaaatcaaaatcataggttggaaaca
16	5	r5t16e	gtgaatttatttaatggtttgaaaaattctt
17	5	r5t18f	gtacataattgctttgaataccaaaagggtta
18	5	r5t18e	tcgcctgaatcaatatatgtgagagtcaata
19	5	r5t20f	gaacctacaaaagtttgagtaacatctaaaa
20	5	r5t20e	ttaattttcatatcaaaattatttaacggat

Row	Column	Name	Sequence
21	5	r5t22f	tatctttaaccgcctgcaacagtgtaagaata
22	5	r5t22e	gtattaacggagcactaacaactgaacgtta
23	5	r5t24h	cgtggcacagacaatatttttgaggcggtca
0	7	r7t0g	caccaccctcattttcagggataatcaccgt
1	7	r7t2f	actcaggaacggggtcagtgcctgcagtctc
2	7	r7t2e	aagttttaggtttagtaccgccactcagagc
3	7	r7t4f	tgaatttaccggaaccagagccacagactgt
4	7	r7t4e	caaaatcaccgttccagtaagcgctggtaat
5	7	r7t6f	agcgcgtttgagggagggaaggttaagttta
6	7	r7t6e	caaccgatttcatcggcattttcgttcataat
7	7	r7t8f	ttttgtcatatcttaccgaagccccgctaata
8	7	r7t8e	gcaatagccaatcaatagaaaatgcgacatt
9	7	r7t10f	tcagagagacgagcgtctttccaggaggttt
10	7	r7t10e	caacgctaataacccacaagaatatgaaata
11	7	r7t12f	tgaagcctaaccaatcaataatcagaacgcg
12	7	r7t12e	catgtagataaatcaagattagttaatcttac
13	7	r7t14f	cctgtttatcatatgcgttatacataccgac
14	7	r7t14e	gtttagtatcaacaatagataagtttacgag
15	7	r7t16f	cgtgtgatttaagacgctgagaagtgaataac
16	7	r7t16e	agcttagaaaataaggcgttaaaaaaagcct
17	7	r7t18f	cttgcttccatcgggagaaacaatgcacgta
18	7	r7t18e	acctttatgtaaatcgtcgctaatagcgat
19	7	r7t20f	aaacagaattaaatcctttgcccaatagatt
20	7	r7t20e	aactcgtaataaagaaattgcgtagtaacagt
21	7	r7t22f	agagccgtataaaacagaggtgaatggctat
22	7	r7t22e	agcagaagcaatagataatacataattcgac

Row	Column	Name	Sequence
23	7	r7t24h	tagtctttaatgcgcgaactgatagaaccacc
0	-1	r-1t0g	cgatctaaagttttgtcgtctttttgcgccg
1	-1	r-1t2f	acaatgaccggtcgctgaggcttcgattata
2	-1	r-1t2e	gatatattaacaaccatcgcccaccaagagaa
3	-1	r-1t4f	ccaagcgcgcctgataaattgtgccctgacg
4	-1	r-1t4e	gtatcatcgaaacaaagtacaaccgccgcca
5	-1	r-1t6f	agaaacactttaatttcaactttacctcgttt
6	-1	r-1t6e	tgagatggcagaacgagtagtaacattacca
7	-1	r-1t8f	accagacggcaaaagaagttttgttttaatt
8	-1	r-1t8e	aggcttttacgataaaaaccaaaattaagac
9	-1	r-1t10f	cgagcttcaggtcaggattagaggctatatt
10	-1	r-1t10e	actccaacaaagcgaaccagaccgcagcctt
11	-1	r-1t12f	ttcatttgactaatagtagtagcagagaaagg
12	-1	r-1t12e	tcaattctgggcgcgagctgaaatcattacc
13	-1	r-1t14f	ccggagacgttctagctgataaattaaattt
14	-1	r-1t14e	attcaaccagtcaaatcaccatcttttcgag
15	-1	r-1t16f	ttgttaaacaaaaataattcgcgtgccggaa
16	-1	r-1t16e	aacgccattcagctcatttttttaaatgcaaat
17	-1	r-1t18f	accaggcagttgggaagggcgatctcacaat
18	-1	r-1t18e	gcgcaactaagcgccattcgccaaacaaaca
19	-1	r-1t20f	tccacacatggggtgcctaatgatgccccag
20	-1	r-1t20e	gtaaagccacatacgagccggaagatcagatg
21	-1	r-1t22f	caggcgaaaaaatcccttataaatagcggtc
22	-1	r-1t22e	aaatcggcaatcctgtttgatggcaaaccct
23	-1	r-1t24h	acgctgcgcgtaaccaccacacccacgaccag

Row	Column	Name	Sequence
0	-3	r-3t0g	tagtaaatgaattttctgtatggggtgaattt
1	-3	r-3t2f	cttaaacacgcttttgcgggatcaatacact
2	-3	r-3t2e	ttaaaggcgcttgataccgatagccagacgt
3	-3	r-3t4f	aaaacactgctccatgttacttaacaaagct
4	-3	r-3t4e	cgcgacctcatctttgacccccaggcagggag
5	-3	r-3t6f	gctcattcttatgcgattttaagcgaggcat
6	-3	r-3t6e	gaattaccagtgaataaggcttgtcgaaatc
7	-3	r-3t8f	agtaagagtaaaatgtttagactgagaggaag
8	-3	r-3t8e	ggtaatagcaacactatcataacatcattgt
9	-3	r-3t10f	cccgaaagccttttgataagaggtacatttc
10	-3	r-3t10e	taattgctacttcaaatatcgcgccagaggg
11	-3	r-3t12f	gcaaatggcatacaggcaaggcaagagtaatg
12	-3	r-3t12e	caataaattcaataacctgtttaagtacctt
13	-3	r-3t14f	tgtaggtatagctatttttgagatgtaaacg
14	-3	r-3t14e	ggagagggaagattcaaaagggtttaacatc
15	-3	r-3t16f	ttaatattccagctttcatcaaccactccag
16	-3	r-3t16e	tcctgtagttgttaaaattcgcattaatgcc
17	-3	r-3t18f	ccagctttgctattacgccagctatagctgt
18	-3	r-3t18e	gcctcttcccggcaccgcttctggtctggcct
19	-3	r-3t20f	ttcctgtgtaattgcgttgcgctagagagtt
20	-3	r-3t20e	actcacattgaaattgttatccgcggtgcgg
21	-3	r-3t22f	gcagcaagagatagggttgagtgtaaaggagc
22	-3	r-3t22e	atagcccgcggtccacgctggttgtgagcta
23	-3	r-3t24h	gggcgctagggcgctggcaagtgtcaaaaga

Row	Column	Name	Sequence
0	-5	r-5t0g	aaacaactttcaacagtttcagcaattgtat
1	-5	r-5t2f	cggtttataaagacagcatcggaaaggcacca
2	-5	r-5t2e	cagcagcgcagcttgctttcgagattttgct
3	-5	r-5t4f	acctaaaagacggtcaatcataaagaaccgg
4	-5	r-5t4e	gaggcgcacgaaagaggcaaaaggtcaccct
5	-5	r-5t6f	atattcatcagtcaggacgttggtaatgcag
6	-5	r-5t6e	cattatactacccaaatcaacgtagccggaac
7	-5	r-5t8f	atacataagcggaatcgtcataacagaagca
8	-5	r-5t8e	ccaatactcgccaaaaggaattaaactggct
9	-5	r-5t10f	aagcggatcttagagcttaattggcgaacga
10	-5	r-5t10e	gcggatggtgcatcaaaaagattagatagcgt
11	-5	r-5t12f	gtagatttagcaataaagcctcatagaaccc
12	-5	r-5t12e	caaaattaagtttgaccattagatcattttt
13	-5	r-5t14f	tcatatatcaggtcattgcctgaacaggaag
14	-5	r-5t14e	aaggctattttaaatgcaatgcctagaattag
15	-5	r-5t16f	attgtatataacaacccgtcggatgacgacga
16	-5	r-5t16e	tgagcgagagcaaatatttaaatgatctaca
17	-5	r-5t18f	cagtatcggctgcaaggcgattagggtaccg
18	-5	r-5t18e	ggggatgtgcctcaggaagatcgattaaatg
19	-5	r-5t20f	agctcgaagtcgggaaacctgtcgacagctga
20	-5	r-5t20e	gctttccattcgtaatcatggtcggcgaaag
21	-5	r-5t22f	ttgcccttaagagtccactattagaacgtgg
22	-5	r-5t22e	tttggaaccaccgcctggccctgcactgccc
23	-5	r-5t24h	cgagaaaggaagggaagaaagcgtgttccag

Row	Column	Name	Sequence
0	-7	r-7t0g	aatagaaaggaacaactaaaggactccaaaa
1	-7	r-7t2f	aaaaggcttacagaggctttgagtaaacggg
2	-7	r-7t2e	gcaacggcccaaaaggagcctttggagtgag
3	-7	r-7t4f	taaaatacaactttgaaagaggactggctgac
4	-7	r-7t4e	aactgaccgtaatgccactacgacgagggta
5	-7	r-7t6f	cttcatcataataaaacgaactaatcagttg
6	-7	r-7t6e	atctacgtagagtaatcttgacagggaaccg
7	-7	r-7t8f	agatttagcctcaaatgctttaaaaaatcag
8	-7	r-7t8e	tgaatcccgaataccacattcaacgaagaaaa
9	-7	r-7t10f	gtctttacagctcaacatgttttatcattcca
10	-7	r-7t10e	aatgctgtcctgactattatagtatattcat
11	-7	r-7t12f	tataacagcggttgtaccaaaaagcctttat
12	-7	r-7t12e	agctaaatttgattcccaattctctgaatat
13	-7	r-7t14f	ttcaacgcagagaatcgatgaacgaccccggt
14	-7	r-7t14e	agcaaacaaaggataaaaattttgagcataa
15	-7	r-7t16f	tgataatccggcggattgaccgtcatcgtaa
16	-7	r-7t16e	ggaacaaaagaaaagccccaaaagagtctgg
17	-7	r-7t18f	ccgtgcatggttttcccagtcacgcatgcctg
18	-7	r-7t18e	aacgccagctgccagtttgagggtctccgtg
19	-7	r-7t20f	caggtcgagaatcggccaacgcgggtggttt
20	-7	r-7t20e	gcattaatctctagaggatccccagttgggt
21	-7	r-7t22f	ttcttttcacgtcaaagggcgaacccgattt
22	-7	r-7t22e	ggactccaaccagtgagacgggcatgccagct
23	-7	r-7t24h	agagcttgacggggaaagccggcaagaacgt

Row	Column	Name	Sequence
0	9	r9t0i	ttttcctcagaaccgccacccccctcaga
1	9	r9t2f	accgccacttttatgatacaggagtgtatcatacat
3	9	r9t4f	ggcttttgtttttagcgtttgccatcttgtcatagc
5	9	r9t6f	ccccttatttttcgccaaagacaaaaggtcatatgg
7	9	r9t8f	tttaccagttttataagagcaagaaacatgagttaa
9	9	r9t10f	cccaatattttctacaattttatcctggctatttt
11	9	r9t12f	gcacccagtttttaatatcccatcctaatcctgaac
13	9	r9t14f	aagaaaaattttaatcataattactagataagaata
15	9	r9t16f	aacaccggtttttagaatccttgaaaacttaattaa
17	9	r9t18f	ttttccctttttgtcagatgaatatacagattttca
19	9	r9t20f	ggtttaacttttattagactttacaaacttgaggat
21	9	r9t22j	ttagaagtttttattaaaaataccgaacgccctaaaacatcgcctttt
2	-9	r-9t2i	actttttctttttttttcacgttgaaaatattgcgaataataatttttt
4	-9	r-9t4e	ggtgtacattttatgaggaagtttccatgactaaag
6	-9	r-9t6e	acattatttttttgaccaggcgcataggcagatgaac
8	-9	r-9t8e	gaaaacgattttacaggtagaaagattcacggaaca
10	-9	r-9t10e	actaaagtttttgaatgaccataaatcaacagttca
12	-9	r-9t12e	ccctgtaattttacggtgtctggaagttaatatgca
14	-9	r-9t14e	aaaactagtttttacttttgcgggagaacattatga
16	-9	r-9t16e	aggtcacgttttcatgtcaatcatatgtgtaatcgt
18	-9	r-9t18e	aaacgacgttttttggtgtagatgggcgaatgggat
20	-9	r-9t20e	ggcggtttttttgccagtgccaagcttgacgttgta
22	-9	r-9t22e	tatcacggttttgcgtattgggcgccagcggggaga
23	-9	r-9t24j	ttttaaccctaaagggagccaaaccgtc

Row	Column	Name	Sequence
		Rem1	tatttacattggcagattcaccagtcac
		Rem2	ttttgacgctcaatcgtctgaaatggat
		Rem3	aggaaaaacgctcatggaaatacctaca
		Rem4	aacaatattaccgccagccattgcaac
		Rem5	actatcggccttgctggtaatatccag
		Rem6	atcacttgcctgagtagaagaactcaa
		Rem7	agcaatacttctttgattagtaataac
		Rem8	gccgcgcttaatgcgccgctacagggc
		Rem9	gcgtactatggttgctttgacgagcac
		Rem10	gtataacgtgctttcctcgttagaatc
		Rem11	agagcgggagctaaacaggaggccgat
		Rem12	taaagggattttagacaggaacggtac
		Rem13	gccagaatcctgagaagtgtttttata
		Rem14	atcagtgaggccaccgagtaaaagagt
		Rem15	ctgtccatcacgcaaattaaccgttgt

Extended 2 (r3t8f) 5' AATAGT + GAGAGCGGAGTGCC + CTTGCT + T + AAAAT-ACAACCGAGGAAACGCAAAGGGAAGC 3'

Displacer 2_6nt 5' AGCAAG GGCACTCCGCTCTC TTATCA 3'

Position 5 (3' alkyne)

Fixer 5 (r5t10e) 5' GGTAGCAGTAGGTA + CTTGCT TG GGT GTG TAT TCA + T + A + TTTGCCAGTCAGAGGGTAATTGAGTTTTTAA 3'

Extended 5 (r7t10f) 5' TCAGAGAGACGAGCGTCTTTCCAGGAGGTTT + T + T + AGCAAG + TACCTACTGCTACC + CTATCG 3'

Displacer 5_6nt 5' CGATAG GGTAGCAGTAGGTA CTTGCT 3'

8nt sticky-ends

Position 2 (5' azide)

Fixer 2 (r3t10e) 5' TTTATCCCCGGGAGAATTAACTGAAGTTACC + T + TTC ATC TGT TTC TT AGCAAGTG + GGCACTCCGCTC 3'

Extended 2 (r3t8f) 5' AATAGT + GAGCGGAGTGCC + CACTTGCT + T + AAAAT-ACAACCGAGGAAACGCAAAGGGAAGC 3'

Displacer 2_6nt 5' AGCAAGTG GGCACTCCGCTC TTATCA 3'

Position 5 (3' alkyne)

Fixer 5 (r5t10e) 5' TAGCAGTAGGTA + CACTTGCT TG GGT GTG TAT TCA + T + A + TTTGCCAGTCAGAGGGTAATTGAGTTTTTAA 3'

Extended 5 (r7t10f) 5' TCAGAGAGACGAGCGTCTTTCCAGGAGGTTT + T + T + AGCAAGTG + TACCTACTGCTA + CTATCG 3'

Displacer 5_6nt 5' CGATAG TAGCAGTAGGTA AGCAAGTG 3'

Biotin-labelled stripping strand 5' biotin + GCTACA TTCATCTGTTTCTTATCC CTTTTGGGTGTGTATTCA CAACTCTCCTTCTCTTCC 3' (toehold + azide strand + alkyne strand + extra to make length 60 nt)

Other modified staples to ensure all functionalised strands are on the same face of the tile

r5t12e 5' TTCCTTATACCTCCCGACTTGCGGAGCCTA 3'

r7t12f 5' GAAGCCTAACCAATCAATAATCAGAACGCG 3'

r7t8f 5' TTTGTCATATCTTACCGAAGCCCCGCTAATA 3'

B.2 DNA Origami Protocol

Single-stranded, circular M13mp18 DNA was purchased from USB Europe GmbH (Staufen, Germany). Staple strands were purchased unpurified from Integrated DNA Technologies (Coralville, IA, USA), and diluted in water to 100 µM and stored at 4 °C. The desired set of staples were mixed with M13mp18 (typically using an excess of staples), in a 100 µl volume of 1x Tris-Acetate-EDTA (TAE) buffer, with 12.5 mM of $MgCl_2$, and annealed from 95 °C

to 21 °C at a rate of 1 °C per minute in 0.1 °C steps in an Eppendorf Mastercycler PCR machine.

Origami samples were generally purified before use, using Sephacryl S-300 spin columns prepared in the lab (bead resin purchased from GE Healthcare). The column matrix was chosen specifically because the beads trap particles smaller than an origami tile, such as free staple strands. To prepare the column matrix for use:

1. Pour out 25 ml of the S-300 bead resin into a Falcon tube. Add 25 ml of MilliQ water.
2. Using a centrifuge, spin the tube for 1 min at 1000 g.
3. Tip out the clear liquid, be careful not to tip away the beads.
4. Top up the tube with MilliQ water to obtain 50 ml.
5. Repeat process three times.
6. After the third cycle, tip out the clear liquid, and top up with origami buffer (1× TAE, 12.5 mM $MgCl_2$). Repeat 5× with origami buffer.
7. After the fifth repeat with buffer, top up the tube with origami buffer, and leave the tube on a slow shaker overnight.

To use the solution to purify origami samples, pipette 500 µl of the column matrix into an empty micro bio-spin column (Bio-Rad). Spin for 1 min at 1000 g, and tip out the waste. Repeat so that the column contains approximately 500 µl of the beads. Finally, spin the column for 3 min to ensure the column is dry. Add the origami sample to the prepared column (typically 50 µl), and spin for 4 min. Collect the elute, pipette into a fresh column, and spin again. Place a third column into a 1.5 ml Eppendorf safe-lock tube. Collect the elute and apply to the third column. After spinning for a third time, the sample eluted into the safe-lock tube is ready for use. Purified samples should be used immediately or within a few hours of purification, as they degrade faster than unpurified origami.

B.3 Super Denaturing Poly-(acrylamide) Gel Electrophoresis

The click reactions were typically analysed using 15 % super-denaturing PAGE gels rather than non-denaturing PAGE (Appendix A.1.3). The super-denaturing gels are composed of two different gel matrices. DNA strands quickly migrate through the thin layer of 5 % 'stacking' gel, and then are separated out by the higher percentage 'separating' gel matrix.

To prepare 40 ml of the 5 % polyacrylamide stacking gel solution:

Ingredient	**Quantity**
1M Tris pH 6.8	5.0 ml
40% acrylamide	5.0 ml
Urea	16.8 g
Deionised formamide	10 ml
H_2O	complete to 40 ml

To prepare 40 ml of the 15 % polyacrylamide separating gel solution:

Ingredient	**Quantity**
3M Tris pH 8.8	5.0 ml
40% acrylamide	15.0 ml
Urea	16.8 g
Deionised formamide	10 ml
H_2O	complete to 40 ml

To cast the gels, assemble two sets of gel plates. Pour out 10 ml of the separating gel solution into a Falcon tube. Add 100 µl of 10 % APS and 10 µl of TEMED to the Falcon

tube. Swirl and pour to fill approximately 2/3 of the volume between the glass plates. It is necessary to stop the acrylamide at the top of the gel from polymerising. To do this, immediately top up the volume between the glass plates with MilliQ water. After 10 min, the gel should have set. Remove the water by holding the clamped gel over a sink and allowing the water to flow out. Use a tissue to dry the area between the plates. Place the gel back into the casting stand. Pour 4 ml of the stacking gel solution into another Falcon tube. Mix 40 µl of APS and 4 µl of TEMED with the gel solution. Swirl and use a p1000 pipette to pour 1 ml (or the amount to fill to the top) into the glass plates. Immediately insert clean gel combs. Once set, remove the gel from the casting clamp, take out the combs, and wash out the gel wells using a syringe filled with MilliQ water. The gels can be used immediately, or kept in the fridge for three days wrapped in a plastic bag with a little water.

To run the gels, place the gel into the buffer tank with 1× Tris-Glycine running buffer (2.5 mM Tris, 0.19 M Glycine). Gels are run at room temperature, at 300 V. Heat up the loading buffer (14 M urea, 90 % deionised formamide and 10 % 1× TBE, with optional 0.05 % bromophenol blue) to 95 °C for 2-3 min. Prepare DNA samples for the gel by adding equal volumes of the heated loading buffer, and denaturing at 95 °C for 2-3 min. Load the gel wells with 10 µl of the samples. Typically, the gel is run for 45-60 min.

B.4 Analysis of Gels

DNA was visualized using a BioRad Pharos FX Plus Molecular Imager before and after staining with SYBR Gold nucleic acid gel stain (Invitrogen). The yield of synthesis was estimated from band intensities in the PAGE gels using Quantity One analysis software (BioRad). For calculation of yields, the fluorescence intensity of the band of interest was divided by the sum of intensities of all the combined bands present in the lane using densitometric analysis.

B.5 Click Chemistry Reactions

B.5.1 Project Collaborators

The synthetic ribosome project is a collaborative EPSRC-funded project between the Physics Department of the University of Oxford (Prof. Andrew Turberfield, Dr. Jonathan Bath, Dr. Mireya McKee, Parminder Lally and Richard Muscat), the Chemistry Department of the University of Warwick (Dr. Rachel O'Reilly, Dr. Phillip Milnes and Diluar Khan), the Chemistry Department of the University of Southampton (Dr. Eugen Stulz and Tom Bandy) and the Electronics Department of the University of York (Prof. Andy Tyrrell and Dr. Cristina Santini).

The work on origami-templated click chemistry was discussed by many members of the group, but the experimental work was conducted by the author and Diluar Khan. The designs of the DNA origami, the displacement mechanisms and other DNA-templated click reactions presented here are the author's contribution to the work. All click-functional DNA strands were synthesised by Diluar Khan. The experiments were conducted in Oxford by the author.

B.5.2 Synthesis of Click-Functionalised Strands

Amine-modified DNA strands were purchased from Integrated DNA Technologies.

Preparation of azide-functional DNA. Amine-modified DNA strands were coupled with 2-azidoacetic acid molecules, synthesised in the lab from sodium azide. The free carboxylic group of the acid was activated as an N-hydroysuccinimide (NHS) ester. The acid was treated with DCC (N,N'-Dicyclohexylcarbodiimide) and HOSu (N-Hydroxysuccinimide) in DCM (dichloromethane) for 4 hours, and after purification, NHS-2-azidoacetic acid was obtained with a 72 % yield. The DNA strands were diluted 100 µM with 200 mM sodium phosphate buffer (pH 7.5). The NHS-2-azidoacetic acid was added to this in DMF and DIPEA (N,N-diisopropylethylamine) in a large excess. The solution

was placed on a shaker for 16 hours at room temperature. Excess small molecules were removed using a NAP column, and the sample was purified further using reverse-phase HPLC.

Preparation of alkyne-functional DNA. Amide-modified DNA strands were coupled with 4-pentynoic acid purchased from Sigma Aldrich. The free carboxylic group of the acid was activated as an N-hydroysuccinimide (NHS) ester. The molecules were treated with TSTU (O-(N-succinimidyl)-N,N,N',N'-tetramethyluronium tetrafluoroborate) and DIPEA (N,N-diisopropylethylamine) in acetonitrile (MeCN) solvent for 2 hours. The solvent was evaporated, and the residue dissolved in DCM to produce the NHS ester of 4-pentynoic acid. The yield of this process was 74 %. Alternative synthesis methods with higher yields were also used. The DNA strands were diluted 100 µM with 200 mM sodium phosphate buffer (pH 7.5). The NHS-4-pentynoic acid was added to this in DMF and DIPEA (N,N-diisopropylethylamine) in a large excess. The solution was placed on a shaker for 16 hours at room temperature. Excess small molecules were removed using a NAP column, and the sample was further purified using reverse-phase HPLC.

B.5.3 Reaction Protocol

1. Make super-denaturing PAGE gels and 1x Tris-Gly running buffer (as described in Appendix B.3).

2. De-gas the required buffer for the experiments (i.e. 50 mM NaCl, or 1× TAE 12.5 mM $MgCl_2$) with argon for at least 10 min.

3. Anneal the required DNA strands in the de-gassed buffer, using the appropriate annealing protocol.

4. Weigh out the amount of CuSO4 (FW 249.7 g/mol), THPTA (FW 438 g/mol) and Na Asc (FW 198 g/mol) required for the experiments.

5. Dissolve Na Asc and THPTA in the required volume of MilliQ water. De-gas each for at least 10 min with argon.

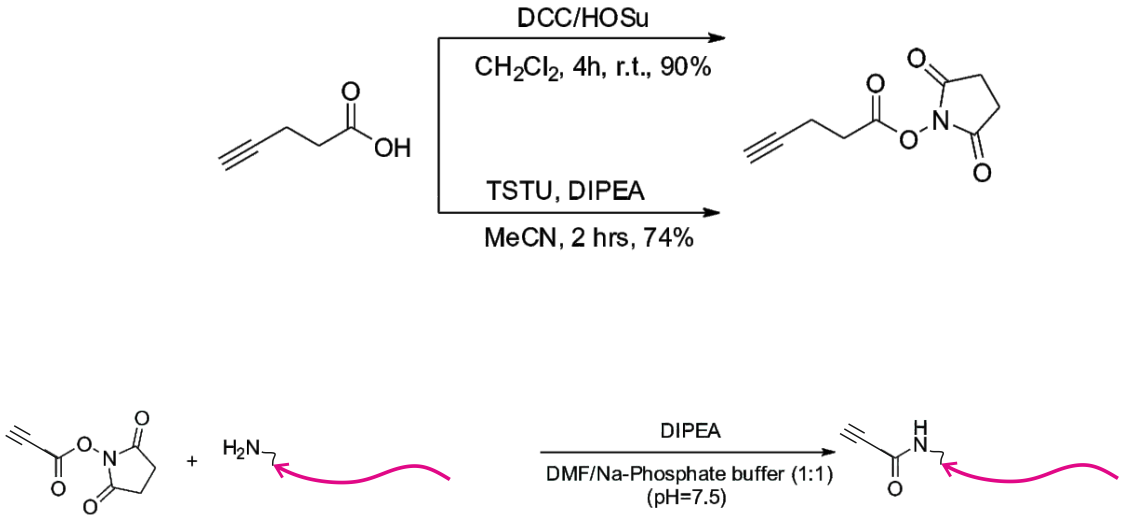


Figure B.2: Alkyne-functionalisation of DNA

After activating the carboxylic group of 4-pentynoic acid as an NHS ester, the alkyne small molecules were coupled to amine-modified oligonucleotides.

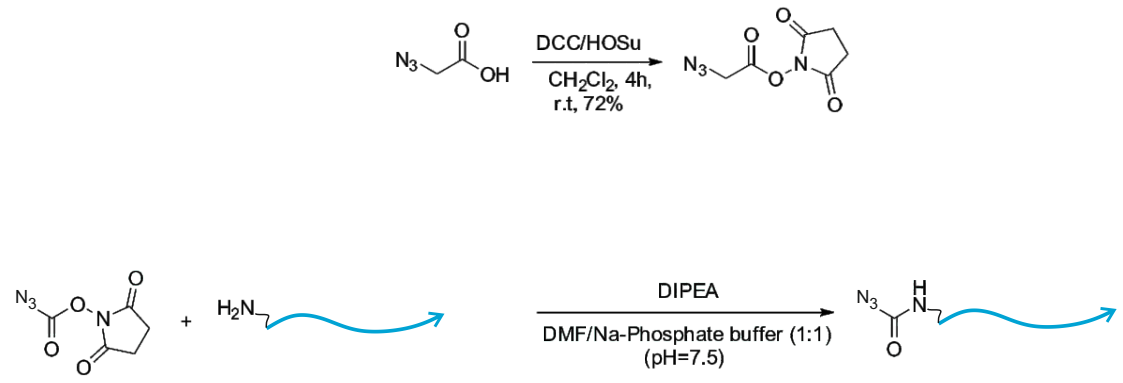


Figure B.3: Azide-functionalisation of DNA

After activating the carboxylic group of 2-azidoacetic acid as an NHS ester, the azide small molecules were coupled to amine-modified oligonucleotides.

6. Once the DNA strands have been annealed, make a stock catalyst solution in a 1.5 ml Eppendorf safe-lock tube by mixing together required volumes of Na Asc and THPTA.

7. Dissolve CuSO4 in required volume of MilliQ water, and de-gas.

8. Add appropriate volume to the catalyst stock solution, and fill the tube with argon (CuSO4 must be mixed with the THPTA copper-binding ligand prior to mixing with DNA).

9. Fill DNA sample tubes with argon. Add catalyst and fill with more argon.

10. Place all samples onto a shaker for the duration of the reaction (typically 2 hours).

11. Once the reaction has finished, prepare the samples for analysis by denaturing PAGE.

For origami-templated click chemistry, the protocol is slightly different:

1. Make super-denaturing PAGE gels and 1x Tris-Gly running buffer.

2. De-gas the 1× TAE 12.5 mM $MgCl_2$ sample buffer with argon for at least 10 min.

3. Anneal the required DNA strands in the de-gassed buffer, using the origami annealing protocol.

4. Weigh out the amount of CuSO4 (FW 249.7 g/mol), THPTA (FW 438 g/mol) and Na Asc (FW 198 g/mol) required for the experiments.

5. Dissolve Na Asc and THPTA in the required volume of MilliQ water. De-gas each for at least 10 min with argon.

6. Once the DNA strands have been annealed, the excess staple strands need to be removed (see purification protocol in Appendix B.2). Make a stock catalyst solution in a 1.5 ml Eppendorf safe-lock tube by mixing together required volumes of Na Asc and THPTA.

7. Dissolve CuSO4 in required volume of MilliQ water, and de-gas.

8. Add appropriate volume to the catalyst stock solution, and fill the tube with argon (CuSO4 must be mixed with the THPTA copper-binding ligand prior to mixing with DNA).

9. Fill purified DNA sample tubes with argon. Add catalyst and fill with more argon.

10. Place all samples onto a shaker for the duration of the reaction (typically 2 hours).

11. Once the reaction has finished, follow the biotin-streptavidin extraction protocol (Appendix B.7).

12. Run the samples on a denaturing PAGE.

B.6 Mass Spectrometry

Liquid chromatography-mass spectrometry (LC-MS) analysis of oligonucleotides was performed on an HP1050 LC coupled to a LCT Premier reflectron TOF mass spectrometer (Waters). Chromatography was carried out on a Waters XBridge OST C18 column (2.5 µm particle size, 4.6×50 mm) at a flow rate of 0.5 ml/min using a gradient of buffers A and B.

* Buffer A is 400 mM 1,1,1,3,3,3-hexafluoroisopropanol, 16.3 mM triethylamine and 5 % methanol;

* Buffer B is 400 mM 1,1,1,3,3,3-hexafluoroisopropanol, 16.3 mM triethylamine and 60 % methanol.

The eluent was directly injected into the mass spectrometer, and the data acquired in the negative-ion mode (mass range 505-3500), this data was analyzed and deconvoluted by using the manufacturer's software (MassLynx V4.1, Waters).

B.7 Protocol for Streptavidin-Coated Magnetic Bead Extraction

To extract biotin-labelled DNA strands with streptavidin-coated magnetic beads, three buffers are required:

* Wash/binding buffer (0.5 M NaCl, 20 mM Tris-HCl pH 7.5, 1 mM EDTA);
* Elution buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA);
* Low-salt buffer (0.15 M NaCl, 20 mM Tris-HCl pH 7.5, 1 mM EDTA).

The protocol for the extraction of the biotin-labelled strand from the origami sample is:

1. Pre-warm the elution buffer in a 80 °C bath.
2. Place low-salt buffer in an ice bath.
3. Aliquot a volume of the magnetic beads equivalent to a 5× excess with respect to the biotin-labelled strand, into a 1.5 ml Eppendorf safe-lock tube. Typically, this is 25 µl.
4. Add 25 µl of the binding buffer. Vortex.
5. Apply a magnet to the tube for 30 sec. Remove and discard the supernatant.
6. Repeat twice more.
7. Add the origami sample (40 µl) to the beads. Incubate for 10 min with regular agitation by hand.
8. Apply magnet, and remove and discard the supernatant.
9. Add 25 µl of the wash buffer. Apply magnet and remove and discard the supernatant.
10. Repeat.
11. Add 25 µl of the cold low-salt buffer to the beads. Vortex and apply the magnet. Remove the supernatant.

12. Add 15 µl of the pre-warmed elution buffer to the beads. Vortex and incubae at 80 °C for 2 min.

13. Apply the magnet, and transfer the supernatant to a clean tube.

14. Repeat with a fresh 15 µl of the elution buffer. Add the supernatant to the stock.

B.8 Non-denaturing Agarose Gel Electrophoresis

Agarose gels are typically used to visualise DNA origami. The nucleic acids are separated according to length by applying an electric field to move the negatively charged molecules through an agarose matrix. Fully formed origami migrate through the matrix much slower than the single-stranded M13mp18 DNA. Agarose gels are required instead of polyacrylamide gels because of the larger matrix pores that enable the large origami to migrate through it. Typically, a 0.75 % agarose solution is made in approximately 50 ml of 1x Tris-Acetate-EDTA (TAE) buffer (40 mM Tris-Acetate,1 mM EDTA disodium salt). The solution is heated in a microwave oven to dissolve the agarose, and then is cooled down to around 60 °C. The solution is then poured into the gel kit and a comb inserted at one end. After approximately 10 minutes, the gel has solidified and is ready for use. All gels were run in 1x TAE buffer in a cold room at 4 °C at a voltage between 50 V and 90 V for between 2-3 hours. Between 10-20 µl of sample of 1-10 nM is added to each well. To stain the gel, a dilution of SYBR gold (Invitrogen) is used to cover the gel, and the gel is incubated for 5 minutes.

B.9 Polymerase Chain Reaction

The polymerase chain reaction (PCR) is a technique used to amplify DNA exponentially. The procedure requires:

1. a DNA template containing the section to be amplified,

2. two primers that are complementary to the 3' ends of the sense and anti-sense strand of the DNA target,

3. Taq polymerase,

4. deoxynucleoside triphosphates (dNTPs), the building blocks from which the DNA polymerase synthesises a new DNA strand,

5. buffer solution, which provides a suitable environment for optimum activity of the DNA polymerase,

6. divalent cations, such as Mg^{2+},

7. monovalent cations, such as Na^{+} and K^{+}.

The PCR reaction is carried out in a reaction volume of 100 µl in a thin-walled reaction tube, which enabled rapid thermal equilibration. The PCR consists of a series of between 20-40 cycles of temperature changes, where each cycle consists of 3 discrete temperature steps. The reaction begins with an initialisation step, where the solution is held at 95 °C for 30 seconds, which is necessary for DNA polymerases that require heat activation. The first step heats the reaction to 95 °C for 30 seconds to melt any secondary structure in the M13mp18 DNA strand. The second step lowers the temperature to 55 °C to allow the primers to anneal to the DNA template. The annealing temperature is usually a few degress below the melting temperature of the primers used. The final step is the extension step, and takes place at the temperature at which the DNA polymerase has optimum activity. The DNA polymerase synthesises a new DNA strand which is complementary to the DNA template by adding dNTPs that are complementary to the template in the 5' to 3' direction. The extension time depends on the DNA polymerase being used and the length of the DNA sequence being amplified. The polymerase usually polymerises a thousand bases per minute. The final elongation step occurs after the final PCR cycle, to ensure the remaining single-stranded DNA is fully extended. The result of the PCR is verified by agarose gel electrophoresis.

B.10 Phenol-Chloroform Extraction

Phenol-chloroform extraction is a liquid-liquid extraction technique to purify a DNA sample from any enzymes such as polymerases or nucleases which were previously used to generate the sample. The method requires phase separation by centrifugation of a mix of an aqueous sample and a solution containing phenol. The process results in an upper aqueous phase and a lower organic phase, mainly containing phenol and the enzymes. The upper phase is removed and an equal volume of chloroform is added, and centrifugation again results in phase separation. The upper phase is the purified DNA sample, and the lower contains any phenol which was left over from the first cycle and chloroform.

B.11 Atomic Force Microscopy

Origami samples were observed using the tapping mode of a Veeco 'Nanoman' Atomic Force Microscope, in a solution of TAE/Mg^{2+} buffer. The narrow cantilever (with a tip of maximum radius 40 nm) of a Veeco NP-S AFM silicon nitride cantilever chip was used for the imaging, which has a resonance frequency of around 9kHz. A mica surface was freshly cleaved, onto which 2 µl of the origami solution was deposited for 1 min. Next, 30 µl of buffer was added, and the AFM tip, mounted in a fluid cell, was lowered into the solution. Typical drive amplitude and amplitude set-point of 100 mV and 0.3 mV respectively were used, to reduce sample damage. Images were flattened to first order using the Nanoman control software.

B.12 Other DNA strands

5'–3' click test reaction

Azide-Fam Strand: 5' azide + GGATAAGAAACAGATGAACGCAGC + FAM 3'

Alkyne Strand: 5' TCAACTTCAACACGCCCTCAGATT + alkyne 3'

Templates: 5' GCTGCGTTCATCTGTTTCTTATCC (AAGGAATA) AATCTGAGGGCGT-GTTGAAGTTGA 3'

5'–3' click in solution, using sticky-ends

Azide: 5' azide + GGA TAA GAA ACA GAT GAA TGT AGC 3' (or FAM)

Alkyne: 5' TGA ATA CAC ACC CAA AAG + alkyne 3' (or TAMRA)

Comp_Az (8nt) 5' TTC ATC TGT TTC TTA TCC AGCAAGTG 3'

Comp_Az(10nt) 5' TTC ATC TGT TTC TTA TCC AGCAAGTGAG 3'

Comp_Az(12nt) 5' TTC ATC TGT TTC TTA TCC AGCAAGTGAGAC 3'

Comp_Az(14nt) 5' TTC ATC TGT TTC TTA TCC AGCAAGTGAGACTC 3'

Comp_Al (8nt) 5' CACTTGCT CTT TTG GGT GTG TAT TCA 3'

Comp_Al (10nt) 5' CTCACTTGCT CTT TTG GGT GTG TAT TCA 3'

Comp_Al (12nt) 5' GTCTCACTTGCT CTT TTG GGT GTG TAT TCA 3'

Comp_Al (14nt) 5' GAGTCTCACTTGCT CTT TTG GGT GTG TAT TCA 3'

Bibliography

[Anderson et al., 2009] Anderson, E. S., Dong, M., Nielsen, M. M., Jahn, K., Subramani, R., Mamdouh, W., Golas, M. M., Sander, B., Stark, H., Oliveira, C. L. P., Pedersen, J. S., Birkedal, V., Besenbacher, F., Gothelf, K. V., and Kjems, J. (2009). Self-assembly of a nanoscale DNA box with a controllable lid. *Nature*, 459:73–77.

[Appella, 2010] Appella, D. H. (2010). Overcoming biology's limitations. *Nat. Chem. Bio.*, 6:87–88.

[Barnes, 1994] Barnes, W. M. (1994). PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proc. Natl. Acad. Sci. USA*, 91:2216–2220.

[Bath et al., 2009] Bath, J., Green, S. J., Allen, K. E., and Turberfield, A. J. (2009). Mechanism for a directional, processive, and reversible DNA motor. *Small*, 5:1513–1516.

[Bath et al., 2005] Bath, J., Green, S. J., and Turberfield, A. J. (2005). A free-running DNA motor powered by a nicking enzyme. *Angew. Chem. Int. Ed.*, 44:4358–4361.

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

[Behlke et al., 2005] Behlke, M. A., Huang, L., Bogh, L., Rose, S., and Devor, E. J. (2005). Fluorescence and fluorescence applications. Technical report, Integrated DNA Technologies. https://eu.idtdna.com/TechVault/TechVault.aspx.

[Benenson, 2001] Benenson, Y. (2001). Programmable and autonomous computing machine made of biomolecules. *Nature*, 414:430–434.

[Bommarito et al., 2000] Bommarito, S., Peyret, N., and SantaLucia, J. (2000). Thermodynamic parameters for DNA sequences with dangling ends. *Nuc. Acids Res.*, 28:1929–1934. DNA hybridisation prediction algorithm available at http://ozone3.chem.wayne.edu.

[Braun et al., 1998] Braun, E., Eichen, Y., Sivan, U., and Ben-Yoseph, G. (1998). DNA-templated assembly and electrode attachment of a conducting silver wire. *Nature*, 391:775–778.

[Brudno et al., 2010] Brudno, Y., Birnbaum, M. E., Kleiner, R. E., and Liu, D. R. (2010). An *in vitro* translation, selection and amplification system for peptide nucleic acids. *Nat. Chem. Bio*, 6:148–155.

[Bui et al., 2010] Bui, H., Onodera, C., Kidwell, C., Tan, Y., Graugnard, E., Kuang, W., Lee, J., Knowlton, W. B., Yurke, B., and Hughes, W. L. (2010). Programmable periodicity of quantum dot arrays with DNA origami nanotubes. *Nano Lett.*, 10:3367–3372.

[Bustamante et al., 1994] Bustamante, C., Marko, J. F., Siggia, E. D., and Smith, S. (1994). Entropic elasticity of λ-phage DNA. *Science*, 265:1599–1600.

[Calderone and Liu, 2005] Calderone, C. T. and Liu, D. R. (2005). Small-molecule diversification from iterated branching reaction pathways enabled by DNA-templated synthesis. *Angew. Chem. Int. Ed.*, 44:73837386.

[Calderone et al., 2002] Calderone, C. T., Puckett, J. W., Gartner, Z. J., and Liu, D. R. (2002). Directing otherwise incompatible reactions in a single solution by using DNA-templated organic synthesis. *Angew. Chem. Int. Ed.*, 41:4104–4108.

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

[Chen et al., 2004] Chen, Y., Lee, S., and Mao, C. (2004). A DNA nanomachine based on a duplex-triplex transition. *Angew. Chem. Int. Ed.*, 43:5335–5338.

[Chin et al., 2003] Chin, J. W., Cropp, T. A., Anderson, J. C., Mukherji, M., Zhang, Z., and Schultz, P. G. (2003). An expanded eukaryotic genetic code. *Science*, 301:964–967.

[Condon, 2006] Condon, A. (2006). Designed DNA molecules: principles and applications of molecular nanotechnology. *Nature Reviews Genetics*, 7:565–575.

[Crick et al., 1961] Crick, F. H. C., Barnett, L., Brenner, S., and Watts-Tobin, R. J. (1961). General nature of the genetic code for proteins. *Nature*, 192:1227–1232.

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

[Dirks et al., 2007] Dirks, R. M., Bois, J. S., Schaeffer, J. M., Winfree, E., and Pierce, N. A. (2007). Thermodynamic analysis of interacting nucleic acid strands. *SIAM Rev.*, 49:65–88.

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

[Douglas et al., 2009b] Douglas, S. M., Marblestone, A. H., Teerapittayanon, S., Vazquez, A., Church, G. M., and Shih, W. M. (2009b). Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nuc. Acids Res.*, 37:5001–5006. Downloadable from http://cadnano.org/.

[El-Sagheer and Brown, 2010] El-Sagheer, A. H. and Brown, T. (2010). Click chemistry with DNA. *Chem. Soc. Rev.*, 39:1388–1405.

[Endo et al., 2010] Endo, M., Sugita, T., Katsuda, Y., Hidaka, K., and Sugiyama, H. (2010). Programmed-assembly system using DNA jigsaw pieces. *Chem. Eur. J.*, 16:5362–5368.

[Feldkamp, 2010] Feldkamp, U. (2010). CANADA: designing nucleic acid sequences for nanobiotechnology applications. *J. Comput. Chem.*, 31:660–663.

[Förster, 1948] Förster, T. (1948). Zwischenmolekulare energiewanderung und fluoreszenz. *Annalen Der Physik*, 437:55–75.

[Fu and Seeman, 1993] Fu, T.-J. and Seeman, N. C. (1993). DNA double crossover structures. *Biochemistry*, 32:3211–3220.

[Gartner, 2006] Gartner, Z. J. (2006). Evolutionary approaches for the discovery of functional synthetic small molecules. *Pure Appl. Chem.*, 78:1–14.

[Gartner et al., 2003] Gartner, Z. J., Grubina, R., Calderone, C. T., and Liu, D. R. (2003). Two enabling architectures for DNA-templated organic synthesis. *Angew. Chem. Int. Ed.*, 42:1370–1375.

[Gartner et al., 2002a] Gartner, Z. J., Kanan, M. W., and Liu, D. R. (2002a). Expanding the reaction scope of DNA-templated synthesis. *Angew. Chem. Int. Ed.*, 41:1796–1800.

[Gartner et al., 2002b] Gartner, Z. J., Kanan, M. W., and Liu, D. R. (2002b). Multistep small-molecule synthesis programmed by DNA templates. *J. Am. Chem. Soc.*, 124:10304–10306.

[Gartner and Liu, 2001] Gartner, Z. J. and Liu, D. R. (2001). The generality of DNA-templated synthesis as a basis for evolving non-natural small molecules. *J. Am. Chem. Soc.*, 123:6961–6963.

[Gartner et al., 2004] Gartner, Z. J., Tse, B. N., Grubina, R., Doyon, J. B., Snyder, T. M., and Liu, D. R. (2004). DNA-templated organic synthesis and selection of a library of macrocycles. *Science*, 305:1601–1605.

[Glökler et al., 2010] Glökler, J., Schütze, T., and Konthur, Z. (2010). Automation in the high-throughput selection of random combinatorial libraries—different approaches for select applications. *Molecules*, 15:2478–2490.

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

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

[Green and Tibbetts, 1981] Green, C. and Tibbetts, C. (1981). Reassociation rate limited displacement of DNA strands by branch migration. *Nucleic Acids Res.*, 9:1905–1918.

[Gu et al., 2010] Gu, H., Chao, J., Xiao, S.-J., and Seeman, N. C. (2010). A proximity-based programmable DNA nanoscale assembly line. *Nature*, 465:202–205.

[Halpin and Harbury, 2004] Halpin, D. R. and Harbury, P. B. (2004). DNA display II. genetic manipulation of combinatorial chemistry libraries for small-molecule evolution. *PLoS Biology*, 2:1022–1030.

[Han et al., 2010] Han, D., Pal, S., Liu, Y., and Yan, H. (2010). Folding and cutting DNA into reconfigurable topological nanostructures. *Nature Nanotechnology*, 5:712–717.

[Hansen et al., 2009] Hansen, M. H., Blakskjær, P., Petersen, L. K., Hansen, T. H., Højfeldt, J. W., Gothelf, K. V., and Hansen, N. J. V. (2009). A yoctoliter-scale DNA reactor for small-molecule evolution. *J. Am. Chem. Soc.*, 131:1322–1327.

[Högberg et al., 2009] Högberg, B., Liedl, T., and Shih, W. M. (2009). Folding DNA origami from a double-stranded source of scaffold. *J. Am. Chem. Soc.*, 131:9154–9155.

[Holliday, 1964] Holliday, R. (1964). A mechanism for gene conversion in fungi. *Genet. Res.*, 5:282–304.

[Hong et al., 2009] Hong, V., Presolski, S. I., Ma, C., and Finn, M. G. (2009). Analysis and optimization of copper-catalysed azide-alkyne cycloaddition for bioconjugation. *Angew. Chem. Int. Ed.*, 48:9879–9883.

[Kim et al., 2010] Kim, K. N., Sarveswaran, K., Mark, L., and Lieberman, M. (2010). DNA origami as self-assembling circuit boards. In *Unconventional Computation Proceedings*, pages 56–68. Lecture Notes in Computer Science.

[Kočalka et al., 2008] Kočalka, P., El-Sagheer, A. H., and Brown, T. (2008). Rapid and efficient DNA strand cross-linking by click chemistry. *ChemBioChem*, 9:1280–1285.

[Koehler and Peyret, 2005] Koehler, R. T. and Peyret, N. (2005). Thermodynamic properties of DNA sequences: characteristic values for the human genome. *Bioinformatics*, 16:3333–3339.

[Kolb and Sharpless, 2003] Kolb, H. C. and Sharpless, K. B. (2003). The growing impact of click chemistry on drug discovery. *Drug Discov. Today*, 8:1128–1137.

[Kumar et al., 2007] Kumar, R., El-Sagheer, A., Tumpane, J., Lincoln, P., Wilhelmsson, L. M., and Brown, T. (2007). Template-directed oligonucleotide strand ligation, covalent intramolecular DNA circularization and catenation using click chemistry. *J. Am. Chem. Soc.*, 129:6859–6864.

[Kuzuya and Komiyama, 2010] Kuzuya, A. and Komiyama, M. (2010). DNA origami: Fold, stick and beyond. *Nanoscale*, 2:310–322.

[LaBean et al., 2000] LaBean, T., Yan, H., Kopatsch, J., Liu, F., Winfree, E., Reif, J. H., and Seeman, N. C. (2000). The construction, analysis, ligation and self-assembly of DNA triple crossover complexes. *J. Am. Chem. Soc.*, 122:1848–1860.

[Li and Liu, 2004] Li, X. and Liu, D. R. (2004). DNA-templated organic synthesis: Nature's strategy for controlling chemical reactivity applied to synthetic molecules. *Angew. Chem. Int. Ed.*, 43:4848–4870.

[Li et al., 2010] Li, Z., Liu, M., Wang, L., Nangreave, J., Yan, H., and Liu, Y. (2010). Molecular behavior of DNA origami in higher-order self-assembly. *J. Am. Chem. Soc.*, 132:13545–13552.

[Liedl et al., 2010] Liedl, T., Högberg, B., Tytell, J., Ingber, D. E., and Shih, W. M. (2010). Self-assembly of three-dimensional prestressed tensegrity structures from DNA. *Nature Nanotechnology*, 5:520–524.

[Liu et al., 1999] Liu, C., Zhou, J., Li, Q., Wang, L., Liao, Z., and Xu, H. (1999). DNA damage by copper(II) complexes: coordination-structural dependence of reactivities. *Journal of Inorganic Biochemistry*, 75:233–240.

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

[Mao et al., 1999] Mao, C., Sun, W., Shen, Z., and Seeman, N. C. (1999). A nanomechanical device based on the B-Z transition of DNA. *Nature*, 397:144–146.

[McKee et al., 2010] McKee, M. L., Milnes, P. J., Bath, J., Stulz, E., Turberfield, A. J., and O'Reilly, R. K. (2010). Multi-step DNA templated reactions for the synthesis of functional sequence controlled oligomers. *Angew. Chem. Int. Ed.*, 49:7948–7951.

[Melkko et al., 2007] Melkko, S., Dumelin, C. E., Scheuermann, J., and Neri, D. (2007). Lead discovery by DNA-encoded chemical libraries. *Drug Discovery Today*, 12:465–471.

[Mills et al., 1967] Mills, D. R., Peterson, R. L., and Spiegelman, S. (1967). An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proc. Natl. Acad. Sci. USA*, 58:217–224.

[Moreira et al., 2005] Moreira, B. G., You, Y., Behlke, M. A., and Owczarzy, R. (2005). Effects of fluorescent dyes, quenchers, and dangling ends on DNA duplex stability. *Biochem. Biophys. Res. Commun.*, 327:473484.

[Morrison and Stols, 1993] Morrison, L. E. and Stols, L. M. (1993). Sensitive fluorescence-based thermodynamics and kinetic measurements of DNA hybridisation in solution. *Biochemistry*, 32:3095–3104.

[Murakami et al., 2006] Murakami, H., Ohta, A., Ashigai, H., and Suga, H. (2006). A highly flexible tRNA acylation method for non-natural polypeptide synthesis. *Nature Methods*, 3:357–359.

[Nangreave et al., 2010] Nangreave, J., Han, D., Liu, Y., and Yan, H. (2010). DNA origami: a history and current perspective. *Current Opinion in Chemical Biology*, 14:608–615.

[Panyutin and Hsieh, 1993] Panyutin, I. G. and Hsieh, P. (1993). Formation of a single base mismatch impedes spontaneous DNA branch migration. *J. Mol. Biol.*, 230:413–424.

[Phillips and Cardelli, 2009] Phillips, A. and Cardelli, L. (2009). A programming language for composable DNA circuits. *J. R. Soc. Interface*, 6:419–436.

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

[Reynaldo et al., 2000] Reynaldo, L. P., Vologodskii, A. V., Neri, B. P., and Lyamichev, V. I. (2000). The kinetics of oligonucleotide replacements. *J. Mol. Biol.*, 297:511–520.

[Robinson and Seeman, 1987] Robinson, B. H. and Seeman, N. C. (1987). The design of a biochip: a self-assembling molecular-scale memory device. *Protein Eng.*, 1:295–300.

[Rostovtsev et al., 2002] Rostovtsev, V. V., Green, L. G., Folkin, V. V., and Sharpless, K. B. (2002). A stepwise huisgen cycloaddition process: copper(i)-catalysed regioselective ligation of azides and terminal alkynes. *Angew. Chem. Int. Ed.*, 41:2596–2599.

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

[Rozenman and Liu, 2006] Rozenman, M. M. and Liu, D. R. (2006). DNA-templated synthesis in organic solvents. *ChemBioChem*, 7:253–256.

[Rozenman et al., 2007] Rozenman, M. M., McNaughton, B. R., and Liu, D. R. (2007). Solving chemical problems through the application of evolutionary principles. *Current Opinion in Chemical Biology*, 11:259–268.

[Saiki et al., 1985] Saiki, R. K., Bugawan, T. L., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*, 230:1350–1354.

[SantaLucia-Jr, 1998] SantaLucia-Jr, J. (1998). A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA*, 95:1460–1465.

[SantaLucia-Jr and Hicks, 2004] SantaLucia-Jr, J. and Hicks, D. (2004). The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.*, 33:415–440.

[Scheuermann et al., 2006] Scheuermann, J., Dumelin, C. E., Melkko, S., and Neri, D. (2006). DNA-encoded chemical libraries. *Journal of Biotechnology*, 126:568–581.

[Seelig et al., 2006] Seelig, G., Soloveichik, D., Zhang, D. Y., and Winfree, E. (2006). Enzyme-free nucleic acid logic circuits. *Science*, 314:1585–1588.

[Seeman, 1982] Seeman, N. C. (1982). Nucleic acid junctions and lattices. *J. Theor. Biol.*, 99:237–247.

[Seeman, 2003] Seeman, N. C. (2003). At the crossroads of chemistry, biology, and materials: structural DNA nanotechnology. *Chem. Biol.*, 10:1151–1159.

[Seeman, 2007] Seeman, N. C. (2007). An overview of structural DNA nanotechnology. *Mol. Biotechnol.*, 37:246–257.

[Seo et al., 2003] Seo, T. S., Li, Z., Ruparel, H., and Ju, J. (2003). Click chemistry to construct fluorescent oligonucleotides for DNA sequencing. *J. Org. Chem.*, 68:609–612.

[Sherman and Seeman, 2004] Sherman, W. B. and Seeman, N. C. (2004). A precisely controlled DNA biped walking device. *Nano. Lett.*, 4:1203–1207.

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

[Shih and Lin, 2010] Shih, W. M. and Lin, C. (2010). Knitting complex weaves with DNA origami. *Current Opinion in Structural Biology*, 20:276–282.

[Shin and Pierce, 2004a] Shin, J.-S. and Pierce, N. A. (2004a). Rewritable memory by controllable nanopatterning of DNA. *Nano Lett.*, 4:905–909.

[Shin and Pierce, 2004b] Shin, J.-S. and Pierce, N. A. (2004b). A synthetic DNA walker for molecular transport. *J. Am. Chem. Soc.*, 126:10834–10835.

[Silverman and Kool, 2006] Silverman, A. P. and Kool, E. T. (2006). Detecting RNA and DNA with templated chemical reactions. *Chem. Rev.*, 106:3775–3789.

[Simmel and Yurke, 2001] Simmel, F. and Yurke, B. (2001). Using DNA to construct and power a nanoactuator. *Phys. Rev. E*, 63:1–5.

[Simmel and Yurke, 2002] Simmel, F. and Yurke, B. (2002). A DNA-based molecular device switchable between three distinct mechanical states. *Appl. Phys. Lett.*, 80:883–885.

[Sims, 2004] Sims, M. (2004). NanoEngineer-1. Opensource 3D multi-scale modeling and simulation program for structural DNA technology, available to download from http://nanoengineer-1.com/content/index.php.

[Snyder et al., 2008] Snyder, T. M., Tse, B. N., and Liu, D. R. (2008). Effects of template sequence and secondary structure on DNA-templated reactivity. *J. Am. Chem. Soc.*, 130:1392–1401.

[Stephanopoulos et al., 2010] Stephanopoulos, N., Liu, M. H., Tong, G. J., Li, Z., Liu, Y., Yan, H., and Francis, M. B. (2010). Immobilization and one-dimensional arrangement of virus capsids with nanoscale precision using DNA origami. *Nano. Lett.*, 10:2714–2720.

[Stryer and Haugland, 1967] Stryer, L. and Haugland, R. P. (1967). Energy transfer: a spectroscopic ruler. *Proc. Natl Acad. Sci. USA*, 58:719–726.

[Tan and Chen, 2006] Tan, Z.-J. and Chen, S.-J. (2006). Nucleic acid helix stability: Effects of salt concentration, cation valence and size, and chain length. *Biophys. J.*, 90:1175–1190.

[Tanaka et al., 2004] Tanaka, F., Kameda, A., Yamamoto, M., and Ohuchi, A. (2004). Thermodynamic parameters based on a nearest-neighbor model for DNA sequences with a single-bulge loop. *Biochemistry*, 43:7143–7150.

[Tkeshelashvili et al., 1991] Tkeshelashvili, L. K., McBride, T., Spence, K., and Loeb, L. A. (1991). Mutation spectrum of copper-induced DNA damage. *J. Biol. Chem.*, 266:6401–6406.

[Tornøe et al., 2002] Tornøe, C. W., Christensen, C., and Meldal, M. (2002). Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(i)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J. Org. Chem.*, 67:3057–3064.

[Tse et al., 2008] Tse, B. N., Snyder, T. M., Shen, Y., and Liu, D. R. (2008). Translation of DNA into a library of 13,000 synthetic small-molecule macrocycles suitable for *in vitro* selection. *J. Am. Chem. Soc.*, 130:15611–15626.

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

[Watson and Crick, 1953] Watson, J. D. and Crick, F. H. C. (1953). Molecular structure for nucleic acids. *Nature*, 171:737–738.

[Winfree, 1995] Winfree, E. (1995). On the computational power of DNA annealing and ligation. *Princeton DIMACS Technical Report*, 1:199–221.

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

[Wrenn and Harbury, 2007] Wrenn, S. J. and Harbury, P. B. (2007). Chemical evolution as a tool for molecular discovery. *Annu. Rev. Biochem.*, 76:331–349.

[Yin et al., 2004] Yin, P., Yan, H., Daniell, X. G., Turberfield, A. J., and Reif, J. H. (2004). A unidirectional DNA walker that moves autonomously along a track. *Angew. Chem. Int. Ed.*, 43:4906–4911.

[Yurke and Mills, 2003] Yurke, B. and Mills, J. A. P. (2003). Using DNA to power nanostructures. *Genetic Programming and Evolvable Machines*, 4(111-122).

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

[Zadeh et al., 2011] Zadeh, J. N., Steenberg, C. D., Bois, J. S., Woolfe, B. R., Pierce, M. B., Khan, A. R., Dirks, R. M., and Pierce, N. A. (2011). NUPACK: analysis and design of nucleic acid systems. *Journal of Computational Chemistry*, 32:170–173. NUPACK web server available to use for free at http://nupack.org.

[Zhang et al., 2007] Zhang, D. Y., Turberfield, A. J., Yurke, B., and Winfree, E. (2007). Engineering entropy-driven reactions and networks catalyzed by DNA. *Science*, 318:1121–1125.

[Zhang and Winfree, 2009] Zhang, D. Y. and Winfree, E. (2009). Control of DNA strand displacement kinetics using toehold exchange. *J. Am. Chem. Soc.*, 131:17303–17314.

[Zhang et al., 2010] Zhang, Z., Zeng, D., Ma, H., Feng, G., Hu, J., He, L., Li, C., and Fan, C. (2010). A DNA-origami chip platform for label-free SNP genotyping using toehold-mediated strand displacement. *Small*, 6:1854–1858.

[Zhu and Wartell, 1999] Zhu, J. and Wartell, R. M. (1999). The effect of base sequence on the stability of RNA and DNA single base bulges. *Biochemistry*, 38:15986–15993.

[Zuker, 2003] Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, 31:3406–3415. Available to use at http://mfold.bioinfo.rpi.edu/.