ARTIFICIAL PROGRAMMABLE POLYMERASES

Robert Oppenheimer
St Catherine's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Trinity 2019

I declare that this thesis is entirely my own work, and except where otherwise stated, describes my own research.

This thesis is dedicated to
Martin and Cheryl Oppenheimer
for freedom and discipline.

ACKNOWLEDGEMENTS

Sometimes it is hard to see all the ways that others have helped. I apologise to those I have missed.

I am grateful to my supervisor Andrew Turberfield. He has an amazing capacity to cut to the heart of a problem within a mess of details, and has helped cultivate my taste for ambitious yet simple experiments. The Turberfield Group have been wonderful, in particular, Jonathan Bath, Joel Spratt and Antonio Garcia Guerra. Jon is a generous and thoughtful leader within the group - he helped me to settle in the lab and set the scope of this project. He also provided an Adobe Illustrator template I have used to represent DNA in this thesis. I had fun chats with Joel from national flags to alien ethics. I always adored his quirky pranks and helpful advice in the lab. With his broad knowledge and interests, Antonio always sparks useful conversations between lab members. I owe him a debt of coffee and cake.

As a biology student doing a chemistry project in a physics department, my co-supervisor Rachel O'Reilly (Chemistry Department, University of Birmingham) and her group have been essential. In particular Tom Wilks, Samuel Núñez-Pertíñez and Lucy Atkinson have been great collaborators. I am also grateful to St Catherine's College for financial support and delightful dinners.

The Synthetic Biology Doctoral Training Centre has given me time to learn broadly and deeply about the field. Having moved overseas, it was particularly helpful to do rotation projects in different laboratories. In particular Robert Carlisle and Antonis Papachristodoulou have been highly supportive of various extra-curricular activities: organising bioethics debates, supervising the UK's first BIOMOD team, etc. A gifted peer from my cohort, Max Jamilly, provided excellent feedback on this thesis.

Upon reflection, the experiences that developed the passion and abilities which led me to this project have been almost exclusively extra-curricular. My high-school science teacher, Kel Hardingham, designed his Forensic Science Camp to foster satisfaction in solving hard problems and a sense of responsibility for one's own learning. Nick Coleman took a chance on me when I had no experience in biochemistry but wanted to start an iGEM team at the University of Sydney. We had serious fun doing science together at a pivotal turning point in my life. Lawrence Lee was a brilliant supervisor of my Honours thesis and BIOMOD team: always as calm and clear as the eye of a storm, and expecting more of me than I expected of myself.

I have few friendships, but they have made a lasting impact. Andrew Tuckwell and I had a blast in Sydney. He taught me how work, life, mental health, energy and relationships are not competing but complementary parts of one's life. I miss Aleisa Jelbart dearly - her devotion and discipline to her art is almost inhuman. Finally, Jarrod Hore and Alex Barwick were the friends who pushed me from school to Forensic Science Camp, where I had my first taste of science.

I am outrageously fortunate to have such a loving and supportive family. My grandparents Joe Walton and Jillian Oppenheimer made it possible for me to live and study in Oxford, and even more importantly I inherited from them a joy in work and learning. Only recently have I begun to understand the impossible debt I owe my parents, Martin and Cheryl Oppenheimer.

Somehow my wife Stephanie has seen past my many blindnesses. I am forever grateful she chased me down the street and across the world. She has had incredible patience over the past four years and while proof-reading this work! Our most recent blessing is the lovely, little August, who somehow has not created pressure during my DPhil, but has instead given perspective on the important things required for life to flourish.

ABSTRACT

This thesis takes inspiration from the ribosome, an ancient nanomachine conserved throughout life for its fundamental role translating genes into proteins. In the last decade primitive replicas of the RNA ribosome have been designed and constructed using DNA as a building material, which I call 'artificial programmable polymerases'. By autonomously synthesising small molecules and aperiodic oligomers, these nanomachines have potential applications spanning directed evolution, smart materials, *in situ* drug synthesis and synthetic biology. However, numerous challenges must be resolved to make polymers of useful lengths. In this thesis I have focused on discrete design challenges relating to the construction of artificial programmable polymerases made of DNA.

DNA sequences can be designed to self-assemble into an array of 2D and 3D structures of nanoscale dimensions. By modifying DNA strands with reactive moieties attached via synthetic linkers, DNA nanostructures can colocalise reactants. In this way, a specific order of chemical reactions can be programmed by the changing the spatial proximity of reactants in a DNA nanostructure. In turn, a specific sequence of reactions during polymer synthesis may be determined by a DNA 'program'. The challenges associated with autonomous programmable polymerisation are addressed in different chapters characterising new architectures for DNA templated chemical reactions.

Chapter 1 articulates a vision for the field, introducing programmable polymerases, their role in natural systems and various approaches to build artificial mimics. This review reveals the challenges addressed in the following chapters. Chapter 2 and Chapter 3 focus on the use of enzymes to coordinate chemomechanical cycles of polymer synthesis and movement by a DNA nanomachine. Chapter 4 proposes DNA architectures where monomers react within an identical reaction environment during each step of polymer synthesis. Chapter 5 explores the protection of reactants from degradation to increase the yield of long polymers.

These advances will be useful in general for dynamic DNA nanotechnology and specifically in the construction of artificial programmable polymerases.

CONTENTS

FIGURES xiii

ABBREVIATIONS xv

1 INTRODUCTION 1

1.1 Motivation . . . 2

1.1.1 Understanding polymers and polymerases in nature . . . 2

1.1.2 Applications of artificial programmable polymerases . . . 6

1.2 Approaches . . . 9

1.2.1 Review of natural programmable polymerases . . . 9

1.2.2 Reengineering natural programmable polymerases . . . 15

1.2.3 Supramolecular chemistry . . . 17

1.2.4 DNA nanotechnology . . . 20

1.3 Aims . . . 22

1.3.1 Blueprint of an artificial programmable polymerase . . . 22

1.3.2 Challenges for the field . . . 25

1.3.3 Contributions of this thesis . . . 30

2 FEEDBACK I – ADAPTERS 33

2.1 Introduction . . . 34

2.1.1 DNA templated chemistry . . . 34

2.1.2 Coordinating DTS and DNA motors . . . 43

2.1.3 Enzyme substrates produced by DTS . . . 43

2.2 Results . . . 46

2.2.1 Flexizyme synthesis of DNA adapters . . . 46

2.2.2 DTS with 3'-ester adapters . . . 49

2.2.3 DTS produces enzyme substrates . . . 51

2.2.4 Blocking enzyme activity before DTS . . . 54

2.3 Discussion . . . 56

3 FEEDBACK II – MOTORS **59**
3.1 Introduction . . . 60
3.1.1 DNA nanomachines . . . 60
3.1.2 Design of an autonomous DTS-dependent DNA motor . . . 67
3.2 Results . . . 70
3.2.1 Assembly of structure and transfer of cargo . . . 70
3.2.2 Comparing and halting strand-displacing polymerases . . . 72
3.2.3 Polymerase-driven TMSD . . . 74
3.2.4 Nickase-triggered TMSD . . . 78
3.3 Discussion . . . 80

4 INSERTION **83**
4.1 Introduction . . . 84
4.1.1 Extension and insertion polymerisation . . . 84
4.1.2 Symmetric architectures for DTS . . . 89
4.2 Results . . . 92
4.2.1 Overhang conjugation . . . 92
4.2.2 Overhang transfer . . . 93
4.2.3 Internal modifications – abasic and nucleobasic . . . 96
4.3 Discussion . . . 100

5 PROTECTION **103**
5.1 Introduction . . . 104
5.1.1 Adapter degradation . . . 104
5.1.2 Strategies from nature . . . 104
5.2 Results . . . 106
5.2.1 Bulk environment . . . 106
5.2.2 Local environment . . . 110
5.2.3 Limiting thioester hydrolysis with abasic modifications . . . 113
5.3 Discussion . . . 117

6 CONCLUSION **121**
6.1 Reflections . . . 121
6.2 Prospects . . . 127

7 APPENDICES **129**
7.1 Methods . . . 130
7.2 Data . . . 143
7.3 Models . . . 153
7.4 Sequences . . . 162

REFERENCES **165**

FIGURES

1.1 Programmable polymerases . . . 3
1.2 The central dogma . . . 5
1.3 DNA encoded combinatorial libraries . . . 7
1.4 Transcription and translation . . . 11
1.5 Approaches: supramolecular chemistry . . . 18
1.6 Approaches: DNA nanotechnology . . . 21
1.7 Blueprint of a programmable polymerase . . . 24

2.1 DNA architectures to colocalise reactants . . . 35
2.2 DNA templated conjugation . . . 37
2.3 DNA templated translation . . . 39
2.4 DNA templated synthesis (DTS) . . . 41
2.5 Programming sequential DTS reactions . . . 42
2.6 Enzyme substrates produced by DTS . . . 45
2.7 Flexizyme synthesis of DTS adapters . . . 48
2.8 DTS with 3'-ester adapters . . . 50
2.9 LC-MS of DTS products . . . 52
2.10 Polymerase activity on DTS products . . . 53
2.11 Blocking polymerase activity with 3'-ester modifications . . . 55
2.12 DTS transition state with 3'-ester-modified adapters . . . 57

3.1 DNA nanostructures and nanomachines . . . 61
3.2 Autonomous DNA motors . . . 64
3.3 Hybridisation chain reactions (HCR) . . . 66
3.4 Design of an autonomous DNA motor with feedback . . . 69
3.5 Assembly of motor . . . 71
3.6 Comparing strand-displacing polymerases . . . 73
3.7 Halting strand-displacing polymerases . . . 74
3.8 A polymerase-driven walker . . . 76
3.9 Exploring pH, metastability, and colocalisation . . . 77
3.10 Nickase-triggered TMSD . . . 79

4.1 Extension and insertion polymerisation . . . 85
4.2 Decreasing M_{eff} during insertion polymerisation . . . 87

4.3 Overhang architectures for DTS 90
4.4 Disulfide conjugation in overhang architecture 93
4.5 Synthesis of donor adapter for overhang DTS 94
4.6 DTS in overhang architecture 96
4.7 DTS with internal abasic modifications 97
4.8 DTS with internal nucleobasic modifications 99

5.1 The effect of pH on DTS 107
5.2 The effect of temperature on DTS 108
5.3 The effect of mixed solvents on DTS 109
5.4 The effect of cations on DTS 110
5.5 DNA secondary structure and thioester hydrolysis 111
5.6 Linker length and DTS yield 112
5.7 Colocalised modifications affecting hydrolysis 113
5.8 Hydrolysis protection by abasic sites 114
5.9 Use of abasic sites during DTS 115
5.10 Testing hydrolysis protection of biotin 117
5.11 New linker designs to protect monomers from degradation 120

ABBREVIATIONS

ssDNA . . . . Single-stranded deoxyribonucleic acid.

dsDNA . . . . Double-stranded deoxyribonucleic acid.

ssRNA . . . . Single-stranded ribonucleic acid.

dsRNA . . . . Double-stranded ribonucleic acid.

mRNA . . . . Messenger ribonucleic acid, the *program* used by the ribosome during translation to determine the sequence of amino acids in proteins.

rRNA . . . . . Ribosomal ribonucleic acid, the ribonucleic acid that folds into the ribosome.

tRNA . . . . . Transfer ribonucleic acid, the *adapters* connecting each amino acid to a specific nucleotide sequence, enabling translation.

dNTP . . . . . Any deoxyribonucleotide triphosphate, e.g., dATP is adenine deoxyribonucleotide triphosphate.

rNTP . . . . . Any ribonucleotide triphosphate.

nt . . . . . Nucleotide, usually in reference to a number of single-stranded nucleotides in a row (e.g., 5nt is a region of 5 nucleotides in a nucleic acid).

bp . . . . . Basepair, usually in reference to a number of double-stranded nucleotides in a row.

PNA . . . . . Peptide nucleic acid - an artificial polymer with nucleobases connected by a polymer backbone of peptide bonds rather than ribose and phosphate bonds.

C_{eff} . . . . . Effective concentration - theoretical prediction of the number of atoms within a unit volume.

DBE . . . . . Dinitrobenzyl ester.

DIPEA . . . . N,N-Diisopropylethylamine.

DMF . . . . . Dimethylformamide.

DTS . . . . . DNA templated synthesis - the acceleration of a chemical reaction between groups colocalised by a DNA nanostructure.

FRET . . . . . Förster resonance energy transfer - the coupling between colocalised light-sensitive molecules enabling the excitation of one chromophore to trigger emission from the other.

HCR . . . . . Hybridisation chain reaction - a cascade of hybridisation events between a set of metastable nucleic acid hairpins, which can be triggered by the presence of a specific *initiator* strand.

M_{eff} . . . . . Effective molarity - empirical observation of the rate enhancement due to colocalisation of reactants.

MS or LC-MS Mass spectrometry - method for separating molecules by their mass-charge ratio. Liquid chromatography mass spectrometry adds an additional step of separation by affinity to a column.

NHS . . . . . N-Hydroxysuccinimide.

PAGE . . . . . Polyacrylamide gel electrophoresis - a method for separating molecules by their charge, size and shape.

PKS . . . . . Polyketide synthase - large enzymes containing modular protein domains that perform specific functions during polyketide synthesis. Each module hands the growing polyketide from one catalytic site to the next, so by evolving one of the modules in a PKS a different polymer can be synthesised.

PTC . . . . . Peptidyl transfer centre - the catalytic core of the ribosome responsible for peptide-bond formation.

RP-HPLC . . Reverse phase high performance liquid chromatography - a method for separating molecules by their hydrophobicity using a nonpolar stationary phase and a polar mobile phase.

UPLC . . . . . Ultra performance liquid chromatography.

RT . . . . . Room temperature (~20°C).

THF . . . . . Tetrahydrofuran.

TMSD . . . . . Toehold mediated strand displacement - the exchange of strands in a complex of nucleic acids by the spontaneous substitution of one strand with another that has a longer domain of complementary base-pairs.

1

INTRODUCTION

The ribosome and nucleic acid polymerases are programmable enzymes. They can synthesise polymers of any sequence, given a corresponding molecular template or 'program', and the corresponding set of precursor monomers. Biomimicry of programmable polymerases is being explored in a range of disciplines spanning molecular biology, DNA nanotechnology and supramolecular chemistry. This work raises questions about the minimal system required for the emergence and evolution of life, and may have applications spanning directed evolution, smart materials, *in situ* drug synthesis and synthetic biology. In this chapter I articulate a vision for the field, review general approaches to build artificial programmable polymerases, and outline challenges for future work. This thesis relies extensively on DNA templated chemistry and DNA nanomachinery, so these subjects are introduced in more detail in Chapter 2 and Chapter 3 respectively. I begin by describing how constructing artificial programmable polymerases may contribute to our understanding of biology while creating useful biotechnologies.

1.1 MOTIVATION

1.1.1 Understanding polymers and polymerases in nature

Life evolves. Evolution requires at least two tasks: storing genetic information and catalysing its replication. Genetic information must be stored in a form that is relatively stable within an organism's lifetime, yet mutable on evolutionary time scales. Erwin Schrödinger was the first to suggest that life stored genetic information in the ordered elements of "aperiodic crystals" [1]. In his words, "crystals = solids = molecules", and it is clear he was envisioning a sequence of monomers covalently bonded in a polymer that specified the plan of an organism's development. In living systems the nucleic acids DNA and RNA perform these roles. A nucleic acid is a polymer of nucleobases connected by a backbone of ribose sugars and phosphate diesters. Pairs of nucleobases have complementary arrangements of hydrogen bond acceptors and donors. Two polymers of complementary nucleotide sequence can *hybridise* by forming double-stranded duplexes. Replication is possible by separating strands and synthesising new polymers upon the isolated templates [2]. In this way information can be stored and copied.

The functions of genes and catalysts are different. A catalyst changes the rate of a reaction while not itself being consumed in the reaction [3]. Biological catalysts modify the network of reactions that replicate life, synthesising metabolites, degrading toxins, repairing structures, and so on. Biological catalysts must exhibit specificity to accelerate the reactions that increase an organism's fitness. Nature achieves this specificity by folding catalytic biopolymers into three-dimensional structures determined by their one-dimensional sequence of monomers [4]. For example, a protein is a polymer of diverse amino acid monomers connected by peptide bonds. These structures enable catalysis by binding and colocalising reactive groups, by positioning acidic and basic monomers, by recruiting charged metal ions and in general by stabilising the intermediate steps of reactions. The proteins and RNA that perform these roles are typically called *enzymes* and *ribozymes* respectively [5].

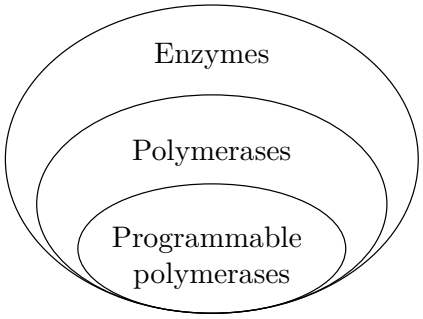


Figure 1.1: Programmable polymerases as a subset of enzymes.

Enzymes are conventionally named by adding the generic suffix '-ase' to the catalysed reaction. Though it is unconventional I use the term *polymerase* for enzymes that catalyse oligomer and polymer synthesis. Examples of this class include polyketide synthases (PKS), non-ribosomal peptide synthases, fatty acid synthases, nucleic acid polymerases and the ribosome. For some polymerases of this class to synthesise polymers with a different sequence of monomers it is necessary to create a different polymerase. For example, PKSs are large enzymes containing modular protein domains that perform specific functions during polyketide synthesis [6]. Each module hands the growing polyketide from one catalytic site to the next, so by evolving one of the modules in a PKS a different polymer can be synthesised [7]. Here, the polymerase is the program.

In contrast, a subset of polymerases are *programmable*. They can synthesise any polymer sequence given a corresponding *program* or molecular template that determines the sequence of the synthesised polymer, and an associated set of monomers. Examples include DNA-programmed RNA polymerases (e.g., T7 polymerase) or RNA-programmed protein polymerases (e.g., ribosomes). The ribosomal machinery itself does not need to evolve to synthesise a new protein: the new protein can be synthesised by providing the ribosome with a new RNA program, called a messenger RNA or mRNA. Programmability decouples the evolution of polymerases from the polymers they synthesise, permitting the sequence space of possible genes and proteins to be explored during evolution without requiring a new polymerase for each new sequence.

Programmable polymerases enable the flow of information in protein synthesis by performing the three fundamental processes of replication, transcription and translation (Figure 1.2A). One can imagine simpler systems than the three different biopolymers and over 150 genes required for protein synthesis [8], such as a single self-replicating molecule or *replicase*. Is this complexity necessary for life or is it an arbitrary convention locked in by evolution?

> *1) A biopolymer specialised to be a catalyst must have many building blocks, so that it can display a rich versatility of chemical reactivity. A biopolymer specialised to store information must have few building blocks, as a way of ensuring faithful replication.*
>
> *2) A biopolymer specialised to be a catalyst must fold easily so that it can form an active site. A biopolymer specialised to store information should not fold easily, so that it can serve as a template.*
>
> *3) A biopolymer specialised for catalysis must be able to change its physical properties rapidly with few changes in its sequence, enabling it to explore "function space" during divergent evolution. A biopolymer specialised to encode information must have physical properties largely unchanged even after substantial change in its sequence, so that the polymer remains acceptable to the enzymes required for replication.*
>
> — Benner & Hutton, 2002 [9]

As outlined in the quote above, there appears to be a trade-off between the two tasks of storing genetic information and catalysing its replication. The ideal genetic material would be composed of few, inert monomers, and have physical properties that are relatively independent of the information it carries. In contrast, the ideal catalytic material would have diverse, reactive monomers, with properties that depend on the information it carries. If this analysis is accurate, then due to these contradictory requirements it seems unlikely that either an idealised genetic or catalytic material could spontaneously evolve by itself. Each is optimised for one task and not the other. In contrast, it seems more likely that early life might evolve in a form representing a compromise between the two extremes. Indeed, this argument has been made to support the RNA world hypothesis that RNA

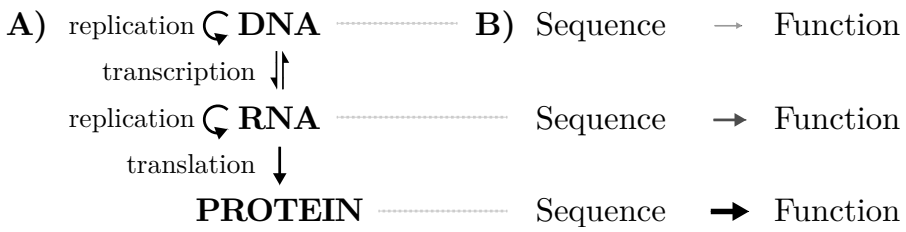


Figure 1.2: A) The central dogma of molecular biology describes the flow of information between polymers. Arrows indicate reactions catalysed by programmable polymerases, which are themselves polymers of RNA or protein. **B)** Nucleic acids and proteins differ in the extent to which sequence determines structure and function. This difference explains why nucleic acids are more ideal genes and proteins are more ideal catalysts, and offers an explanation of why the processes of transcription and translation exist.

self-replicated before the subsequent evolution of DNA and proteins [9]. RNA is more reactive than DNA due to the 2'-hydroxyl, which often plays a central role in ribozymes [10]. RNA is capable of more diverse folds than DNA due to the 2'-hydroxyl acting as both a hydrogen bond acceptor and donor, and limiting backbone flexibility by sterically inhibiting ribose puckering [11]. RNA is a simpler template for copying information than protein because base-pairing is more specific than the interactions between protein α-helices and β-sheets. Thus we might expect to find early life in a simpler, compromised form with limited catalytic potential to replicate a smaller set of sequences due to folding of the genetic polymer.

Translation is nature's solution to this trade-off. If information can be transferred between polymers, then two or more distinct types of polymer could perform the twin tasks of storing genetic information and catalysing its replication. In many living systems, information is transmitted from DNA through RNA to proteins formed of 20 distinct amino acids (in the canonical genetic code). The sequence of nucleotides in a nucleic acid are translated into protein using a triplet code, where three adjacent nucleotides in a single strand form a *codon* determining the identity of an amino acid [12]. The $4^3 = 64$ different codons typically encode 20 different amino acids: similar codons sometimes encode the same amino acid, resulting in a redundant genetic code. Translation enables a polymer with relatively few, inert monomers (RNA) to encode the sequence of a more diverse, reactive polymer (protein). As depicted in Figure 1.2A, translation and transcription permit

the flow of information from a polymer optimised as a genetic material (DNA) to a polymer optimised for catalysis (protein).

To date, all known forms of life share the same machinery for translation - a ribozyme called the *ribosome* [13]. It is an open question whether the ribosome is an optimal machine for translation, or whether it is a frozen accident that continues to be replicated because alternatives have not been explored. It is unclear if the evolution of the ribosome has created forms of life that dominate our planet's resources so successfully that they prevent the simultaneous evolution of different forms of life, or if the ribosome itself represents a symbiosis between previously separate and primitive forms of DNA-, RNA- or protein-based life.

Building artificial programmable polymerases will aid our understanding of the minimal complexity required for the open-ended evolution of complex life. Moreover, these nanomachines also represent new and powerful biotechnologies [14].

1.1.2 Applications of artificial programmable polymerases

Over the last century, synthetic chemistry has demonstrated the utility of synthetic molecules as fuels, plastics, medicines, materials, flavours, and more. The rational design of complex molecules is challenging. An alternative to rational design is the combinatorial screening of a library of possible molecules. However, as each molecule is typically prepared in a separate reaction volume before screening, the scale of synthesis rapidly becomes unwieldy.

One strategy to overcome the issue of scale is *DNA recorded synthesis*, using double-stranded DNA (dsDNA) tags covalently linked to the molecule of interest [15, 16]. During synthesis of the combinatorial library of molecules, a DNA oligomer with a unique sequence is added to each molecule during each reaction so that the identity of every molecule and chemical transformation is recorded by its own DNA tag (Figure 1.3A). Separate reaction volumes are not needed to keep track of each molecule's identity as this information can be recovered by DNA sequencing of the tags. This technique has provided large combinatorial libraries ($>10^9$ molecules) used in various commercial enterprises [16]. However, the synthesis and screening

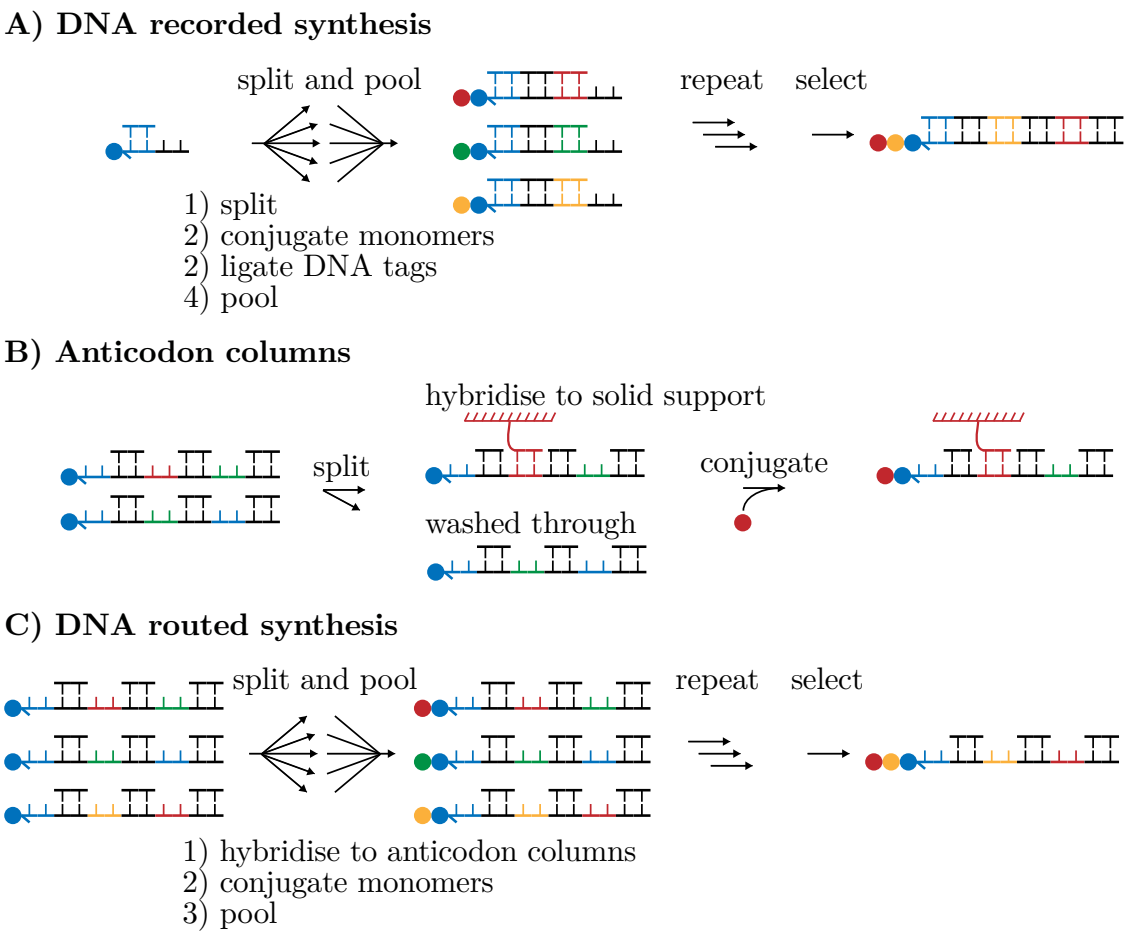


Figure 1.3: A) DNA recorded synthesis. During polymer synthesis unique DNA tags (coloured) are ligated to record the identity of monomers in a growing polymer. In this way combinatorial libraries of small molecules, oligomers and polymers can be synthesised in one-pot with DNA tags recording the sequence of chemical transformations during polymer synthesis [15]. By sequencing the library before and after selection against a target, the enrichment of a DNA tag indicates the affinity of its corresponding polymer for the target. **B) Anticodon columns.** DNA 'genes' with specific 20nt codon sequences (coloured) can be isolated from a library by hybridisation to complementary strands on a solid support called *anticodon columns*. In this way a codon sequence determines if a chemical transformation is performed on a gene. **C) DNA routed synthesis.** DNA 'genes' predetermine rather than record the steps of polymer synthesis [17–19].

of random molecules is ultimately limited as it samples a tiny fraction of possible molecules (i.e., there are more 60mer protein sequences than atoms in the universe, $\sim 10^{80}$). Directed evolution represents a compromise between the two extremes of rational design and random search: molecules that are already known to be useful are used to seed future generations of molecules.

Directed evolution has been used to identify new and beneficial DNA, RNA

and protein sequences. The utility of this approach was recognised by the 2018 Nobel Prize in Chemistry for the "directed evolution of enzymes, peptides and antibodies". But directed evolution is largely limited to biopolymers (or very closely related polymers). This is because directed evolution currently employs natural programmable polymerases for replication, transcription and translation. An alternative to enzymatic synthesis uses solid-phase chemical synthesis, as demonstrated in *DNA routed synthesis* [17–19]. Here a library of single-stranded DNA (ssDNA) *genes* are designed with subsequences called codons that are 20 nucleotides long (Figure 1.3B). Codons permit hybridisation of the genes to complementary anticodon DNA sequences covalently bound to a solid support (e.g., a resin, column or bead). A subset of genes containing a specific codon can be isolated from the pool of genes by hybridising to an anticodon column and washing away the unhybridised genes. Alternatively, libraries can be split into subsets sharing common codons by routing the library through many different anticodon columns. At this stage a different chemical transformation can be performed on each subset of the library. For example, ssDNA genes can be chemically modified by the conjugation of amino acids, and codon sequences can be used to specify the nature of the amino acid (side-chain, chirality, backbone length, etc.). In this way a library of polymers can be synthesised in parallel by many cycles of splitting, chemical synthesis and pooling. Importantly, the identity and position of a monomer in a polymer is predetermined by the presence of a *codon* in a gene and the order of chemical operations. Synthesised polymers are programmed by their genes. This permits the rounds of synthesis, selection, diversification, sequencing and re-synthesis required for directed evolution. This strategy has been used to evolve 5mer peptides [20], though in principle it could be extended to entirely non-natural polymers.

While it is inspiring to imagine evolvable chemistry and improvements to fuels, plastics, medicines, materials and more, it is important to remember that programmable polymerases are more than just chemical factories. They are autonomous nanomachines within dynamically self-organising cells. They enable *gene regulation* - the processes by which the environment inside and outside cells

affects levels of gene expression [21]. By using environmental cues to trigger or inhibit polymerisation, cells can adapt to environments by expressing an appropriate set of genes. Analogously, if an artificial programmable polymerase could be externally controlled then it could synthesise a set of polymers in response to specific cues. This might enable, for example, 1) medical systems permitting *in situ* synthesis of drugs in response to disease conditions [22], 2) chemical systems that take inspiration from self-regulating metabolic systems [23, 24] to control the relative concentrations of catalysts and reactants during production, or 3) computational systems in which molecular information could be written, stored, read, and also trained to respond to diverse cues [25, 26]. These applications rely on autonomous adaptation to the local environment and thus are impossible with the method of externally controlled solid-phase polymer synthesis described above.

Artificial programmable polymerases will form the basis of future biotechnologies by extending evolution beyond biology and enabling the synthesis of self-regulating materials. The construction of these nanomachines requires an understanding of the programmable polymerases nature has optimised over three and a half billion years of evolution. In the next section I review natural programmable polymerases. I then introduce approaches from various fields to mimic their function and enable the synthesis and evolution of non-natural polymers.

1.2 APPROACHES

1.2.1 Review of natural programmable polymerases

Transcription. Here I briefly outline the process of transcription by T7 RNA polymerase, a simple and well-characterised example of an enzyme catalysing the synthesis of ssRNA sequences programmed by dsDNA templates. T7 RNA polymerase is a single protein from the T7 bacteriophage that infects *E. coli* bacteria [27, 28]. The enzyme moves along the DNA program one nucleobase at a time, simultaneously connecting monomers to form a complementary polymer sequence (Figure 1.4A). Note that in higher organisms transcription is typically more complex.

For example, multi-protein complexes facilitate many of the steps, regulating when and where transcription occurs. Replication of nucleic acids is not described here as the general principles are similar to transcription - replication can occur via two sequential transcription reactions (as in retroviruses) or via a similar polymerisation reaction where the program and synthesised polymer are the same molecular species (as in DNA-programmed DNA replication by prokaryotes and eukaryotes).

T7 polymerase transcribes the T7 DNA genome into RNA by first binding to a specific dsDNA sequence, called a T7 promoter. Binding to this sequence causes the local denaturation of the duplex into two unwound, unhybridised strands. One of these strands acts as a template to program RNA polymerisation. Free, activated monomers (rNTPs) transiently bind to the protein and hybridise to complementary nucleotides on the exposed ssDNA template. To initiate polymerisation T7 can use a single RNA nucleotide or an RNA oligonucleotide called a *primer*. The phosphate bond in the backbone of the RNA strand is formed when the 3'-hydroxyl of the growing polymer attacks a phosphate ester of the incoming rNTP. This reaction is catalysed by two magnesium ions coordinated in position by T7 polymerase [29]. 5'→3' polymerase activity is often accompanied by the reverse reaction of 3'→5' exonuclease activity (usually catalysed by a separate protein domain), removing the nucleotides that the polymerase has just added. This would seem like a futile cycle in which a polymerase is constantly undoing the progress it has made, however, it is an example of the kinds of mechanisms polymerases can use to reduce errors, as mismatched nucleotides have slower polymerisation kinetics and a greater chance of excision by exonuclease activity [30].

Once an rNTP has been incorporated, T7 polymerase moves back and forth in an equilibrium between the current conformation (at nucleotide position i) and a conformation one nucleotide forward ($i+1$). This can be considered a Brownian ratchet [31, 32], in the sense that thermal motion drives the enzyme between the two states (i and $i+1$). In the forward state ($i+1$) there is a chance that the next rNTP may be incorporated, and if this happens then the polymerase gets trapped between the next two positions ($i+1$ and $i+2$), and so on for subsequent

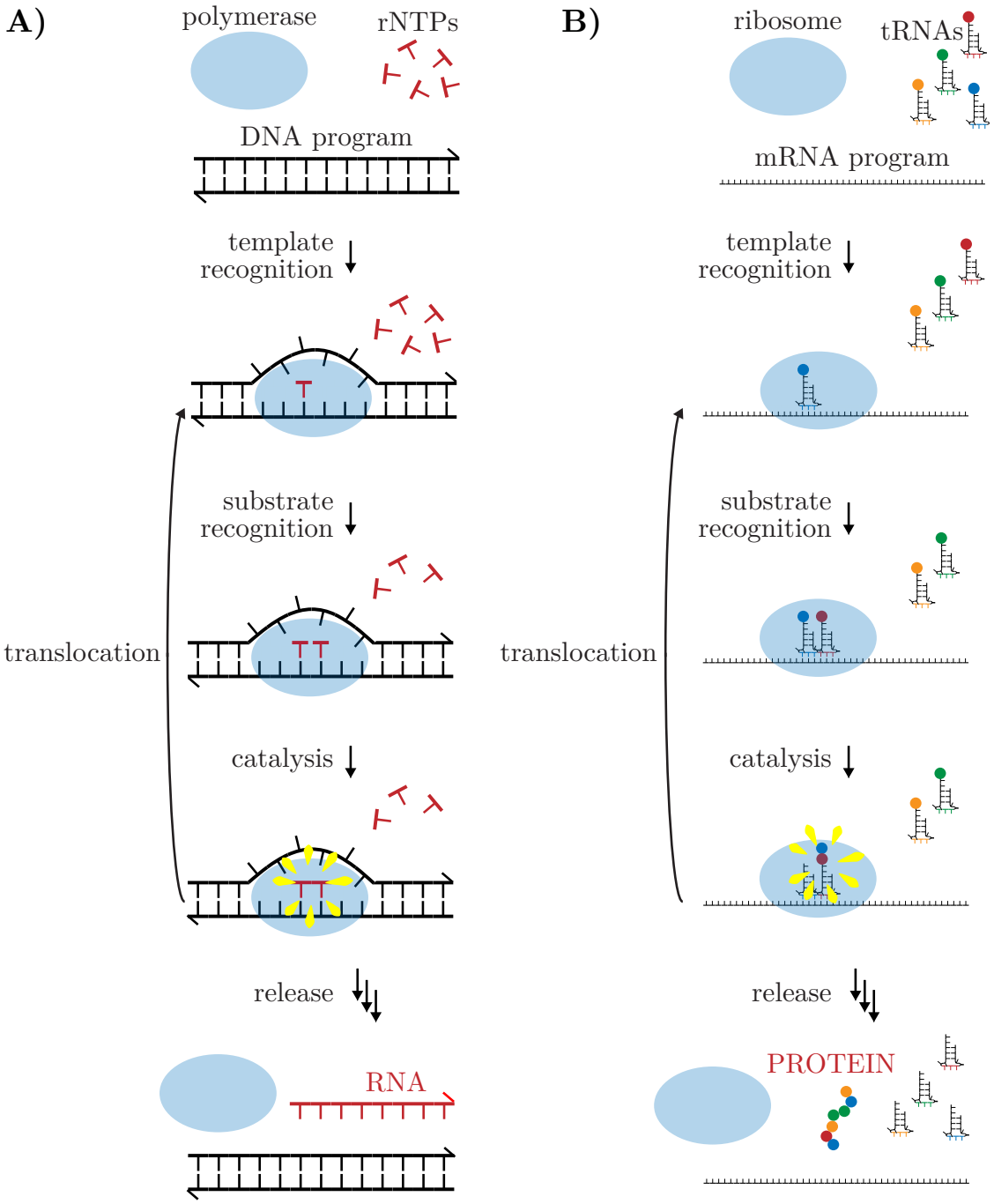


Figure 1.4: **A)** Transcription of ssRNA from a dsDNA template by T7 polymerase may be simplified into 5 key steps: 1) template recognition, 2) substrate recognition, 3) catalysis, 4) translocation and 5) release. **B)** Translation of protein from an ssRNA template by the ribosome may be similarly simplified. Note the ribosome is depicted here as a single structure, but in reality it is composed of multiple rRNA and protein subunits that self-assemble around the mRNA (see ***Translation***, page 12).

cycles of RNA polymerisation. Due to the helicity of nucleic acids, polymerases not only move along the axis of the helix, but also rotate around the helix by ~35° with each added nucleotide [33]. Polymerase activity can terminate by polymerase dissociation, by reaching the end of the program, by other proteins bound to specific nucleic acid sequences that block polymerase progression, or by the formation of specific DNA or RNA secondary structures that interfere with polymerase binding.

Finally, in prokaryotes transcription and translation can occur simultaneously. Large macromolecular complexes form when ribosomes are recruited to mRNA during transcription by DNA-programmed RNA polymerases [34]. Though I present transcription and translation here as separate processes, interactions between ribosomes and polymerases affect the rates of polymer synthesis and gene regulation in ways that are still being understood [35].

Translation. Here I briefly describe a typical prokaryotic ribosome from *E.coli*. It is important to note that eukaryotic translation is typically more complex. There may be multiple, specialised copies of ribosomes within genomes [36]; different mechanisms of recruiting ribosomes to mRNA; ribosome-mediated protein folding; simultaneous synthesis and translocation of polymers through membrane pores; and the co-existence of multiple heterogeneous ribosomes within single cells (ribosomes in mitochondria more closely resemble prokaryotic ribosomes than eukaryotic ribosomes in the endoplasmic reticulum) [37].

Over 150 genes appear to be required for the transcription of three ribosomal RNA strands (5S, 16S, and 23S rRNA) and over 30 transfer RNA strands (tRNA), for the post-transcriptional modification of rRNA and the charging of the 3' ends of tRNAs with amino acids [8]. The ribosome is composed of two major subunits, the 50S subunit containing 5S rRNA, 23S rRNA and 31 ribosomal proteins, and the 30S subunit containing 16SrRNA and 19 ribosomal proteins. The ribosome assembles around the messenger RNA (mRNA) encoding the gene to be translated, triggered by duplex formation between an mRNA *ribosome-binding site* and a complementary sequence in the 16S rRNA. From this position the ribosome moves along the mRNA

in the 3' direction until it reaches the *start codon* (AUG), where translation begins by recruiting the corresponding tRNA ($tRNA^{fMet}_{CAU}$).

During translation the ribosome is always bound to at least one tRNA though its identity changes during every step of protein synthesis. The ribosome cyclically recruits a second tRNA, catalyses a transfer reaction between them, and then discards the first tRNA. Each tRNA initially acts as an *acceptor* receiving the growing protein from the *donor* tRNA, then after a conformational change acts as a donor to the next incoming acceptor tRNA. During the transfer reaction both tRNAs are very weakly hybridised by their triplet codon to mRNA but held in place by non-covalent interactions across the body of the tRNA and rRNA. Translational fidelity is improved during recruitment of tRNAs to the ribosome by *kinetic proofreading* [38, 39]. This adds at least one energy-consuming, GTP-hydrolysing step between the tRNA-ribosome association and the incorporation of the tRNA into the heart of the ribosome. During this step mismatched tRNAs are incorporated within the ribosome slower than matched tRNAs. This means that mismatched tRNAs have a greater chance of dissociation from the ribosome before the transfer reaction occurs, thus increasing translation fidelity. Once recruited, the 3' end of each tRNA points into the heart of the ribosome - the peptidyl transfer centre (PTC) - permitting acyl transfer between donor and acceptor tRNAs.

The PTC is the ancient core of the ribosome, with an RNA sequence unchanged throughout all living species. The PTC is purely RNA, while the ribosome's exterior is covered with proteins that assist folding, structural movement, and initiation and termination of protein synthesis. The specific arrangement of rRNA in the PTC protects the linkage between tRNA and amino acid from hydrolysis by sterically blocking the 3'-ester from nucleophilic attack by water, while promoting attack by the amino acid on the acceptor tRNA [40]. Upon peptide bond formation, the 30S subunit begins fluctuating between two conformations: the small ribosomal subunit rotates ~6° relative to the large subunit and the ribosome itself translocates forwards and backwards on the mRNA. The forward position is locked into place by the protein EF-G binding the ribosome and hydrolysing GTP. This prevents further

rotation, permits the spent donor tRNA to dissociate and creates room for the next acceptor tRNA [41, 42]. The cycle of polymerisation continues until the ribosome reaches a *stop codon* in the mRNA (either UAA, UGA, UAG), where release factor proteins associate with the mRNA-ribosome complex to initiate disassembly of the ribosomal machinery and cleave the protein from the terminal tRNA.

Importantly, the ribosome itself does not participate directly in the transition state of the transfer reaction. It is an enzyme only in the sense that it brings the charged 3' ends of each tRNA into close proximity for peptide bond forming reactions - it is the 2'-hydroxyl of the acceptor tRNA that plays the crucial role. In model systems the 2'-hydroxyl of the acceptor tRNA accelerates peptide bond formation by 10^3-10^6-fold [43, 44]. Interestingly, the same 2'-hydroxyl may also accelerate hydrolysis of the tRNA's aminoacyl ester prior to recruitment of tRNAs to the ribosome. The pool of charged tRNAs is maintained by tRNA synthetases, while the protein Ef-Tu binds indiscriminately to all charged tRNAs with a binding pocket that limits hydrolysis of amino acids from tRNAs [45]. Thus translation is impossible with just the ribosome itself - an ecosystem of enzymes is necessary.

The common features of transcription and translation are the following five steps: 1) template recognition, 2) substrate recognition, 3) catalysis, 4) translocation and 5) release (Figure 1.4). The distinct features are that during transcription the nucleotide monomer recognises the template with a complementary base-pairing interface, whereas in translation these functions are divided into an adapter (tRNA) and a monomer (amino acid) that is cleaved from the adapter during protein synthesis [46]. Furthermore, the enzyme-free polymerisation of nucleic acids is possible with a single-stranded template and activated monomers [47], whereas proteins cannot spontaneously polymerise on nucleic acid templates. Finally, transcription transfers information between polymers with a monographic cipher or 1:1 exchange of letters (A→T, G→C, etc.), whereas translation is a polygraphic cipher or N:M exchange of triplet codons for amino acids (UGG→arginine, UCA→serine, etc.). Presumably these differences represent just a small subset of the possible variations that could be explored by artificial programmable polymerases.

1.2.2 Reengineering natural programmable polymerases

Here I review approaches to build programmable polymerases. Predictably, the earliest and fastest progress has been made by repurposing nature's nanomachines. In the late 1980's it was demonstrated that two natural polymerases (the Klenow fragment of DNA-programmed DNA polymerase and the T7 DNA-programmed RNA polymerase) could incorporate non-natural nucleotides with rationally designed hydrogen bonding motifs [48]. Since then, successful strategies have focused on the substrate promiscuity of polymerases, for instance, those that are known to have recently changed substrate when evolving from DNA-programmed DNA polymerases into DNA-programmed RNA polymerases [49]. By mutating the specific protein domains associated with substrate promiscuity, it has been possible to evolve polymerases capable of synthesising non-natural nucleic acids (called xeno nucleic acids or XNA, synthesised by XNA polymerases). Indeed, it has even been shown that the directed evolution of XNA aptamers is possible by iterative transcription and reverse-transcription from a library of DNA genes replicated and diversified by polymerase-chain reaction (PCR) [50]. The current XNA family includes modifications to the nucleobase, sugar and phosphate groups of nucleic acids [51]. This opens the door to an entirely orthogonal central dogma, with information replicated and transcribed from non-natural nucleic acids. However, the system is not easily amenable to rational redesign or optimisation. It is not clear why some types of XNA can be enzymatically polymerised with high yield and fidelity while others cannot. Each new type of XNA requires multiple rounds of directed evolution with unpredictable results.

Breeding ribosomes is more challenging due to the larger number of components involved and the effect of changes to the ribosome on the entire proteome and viability of cells. Significant progress was made by emulating a classic gene duplication event [52]. Using two ribosomes with distinct ribosome-binding sites *in vivo* permits one ribosome to translate the cell's set of natural proteins while the other ribosome evolves novel functions [53]. This enabled the evolution of ribosomes that can read through a 4-nucleotide codon in mRNA while incorporating an amino acid from a

tRNA with the corresponding 4-nucleotide anticodon, potentially resulting in an expanded set of $4^4 = 256$ codons [54]. However, it is impossible to harness this potentially larger protein sequence space without also changing the entire set of tRNA synthetase enzymes that charge tRNAs with amino acids [55]. This has proven challenging *in vivo* due to competition for the same tRNA substrates by natural and unnatural sets of tRNA synthetases. A general strategy has been to use a redundant codon in the natural genetic code (such as the rarest of the three stop codons) to introduce a new tRNA:tRNA synthetase pair, though this reduces host fitness [56]. As a result the most significant advances have been made with *in vitro* translation using chemical or enzymatic substitutes for tRNA synthetases. A useful tool in this area has been the directed evolution of a ribozyme that mimics tRNA synthetases, called a *flexizyme*, which has programmable tRNA specificity through complementary base-pairing between the RNA flexizyme and tRNA [57, 58]. Flexizymes have simplified the process of introducing non-canonical ribosomal substrates to tRNAs such as modified side-chains, D-stereochemistry, β- and γ-amino acids, dipeptides, α-hydroxy acids, and backbone modifications such as N-methylation and N-substitution [59].

While this access to new substrates is exciting, it comes with the length limitation that ribosomes typically stall after incorporating >2 sequential non-canonical substrates [60]. In future, this might be resolved by laboriously continuing further rounds of directed evolution in an attempt to create ribosomes capable of full-length synthesis of non-natural polymers. For example, after a decade of carefully tuning tRNA sequences and reagent concentrations, ribosomal synthesis of a peptide with <10 sequential D-amino acids was possible [61]. Interestingly, there is one case in which synthesis of full length proteins might be immediately possible without directed evolution. A *mirror ribosome* - with stereochemistry of the entire biochemistry flipped around each chiral centre - would translate *mirror mRNAs* (with L-RNA rather than the natural D-RNA) into full-length and functional *mirror proteins* (with D-amino acids rather than L-amino acids). In turn, any *mirror enzymes* produced could catalyse reactions with *mirror substrates* [62].

While this would require a currently prohibitive amount of chemical synthesis, the fundamental technologies have already been developed through the total chemical synthesis of *mirror polymerases* capable of replication, transcription, reverse-transcription and PCR of nucleic acids with flipped chirality [63–65]. Indeed, it was recently demonstrated that a 5S rRNA from a *mirror ribosome* could be transcribed from *mirror DNA* using a synthesised mirror polymerase [66]. The huge investment of resources required for this project might be justified if a simplified central dogma could become self-sustaining by *in vitro* transcription, translation and replication [67]. A *mirror cell* might be capable of autonomously reproducing without parasites or predators, which is particularly relevant for industrial biotechnology and medicine [68].

The top-down repurposing of natural polymerases will continue to make rapid progress. However, due to the complexity of biological nanotechnology and the difficulty of rational design, every new substrate and increase in polymer length necessitates additional rounds of directed evolution, rather than permitting general polymerase activity. Repurposing natural polymerases also limits the diversity of possible polymers, as they must be compatible with the aqueous solvents in which nucleic acids are soluble, foldable and active.

1.2.3 Supramolecular chemistry

There is a spectrum between top-down and bottom-up approaches to build programmable polymerases. I begin with the most extreme example of the bottom-up approach where supramolecular chemists rationally design molecules to act as nanomachine components such as switches, bearings, compartments and more [69]. This general approach to "designing and synthesising molecular machines" was recognised by the 2016 Nobel Prize in Chemistry. Various design motifs have been combined by David Leigh's group to create nanomachines where cyclic molecules called *rotaxanes* are threaded onto linear molecules called *axles*. An axle bears amino acids at regular intervals on labile ester modifications (Figure 1.5). The rotaxane is topologically trapped between a stopper at one end and the several amino

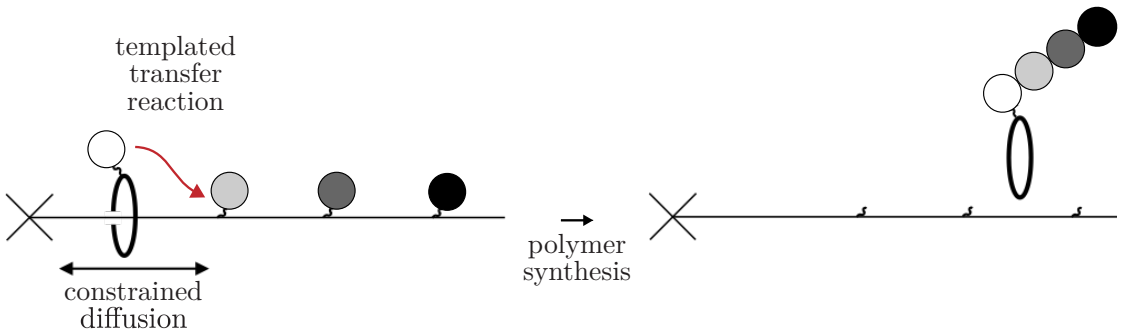


Figure 1.5: Approaches to build artificial programmable polymerases - supramolecular chemistry. A rotaxane is threaded upon an axle with a stopper (×) at one end. The rotaxane diffuses until it falls off the unstoppered end of the axle. Monomers are attached to the axle by labile ester linkers, which limit rotaxane diffusion until cleaved from the axle by templated transfer to the rotaxane [70–72].

acids, limiting the rotaxane's one-dimensional diffusion along the axle until native chemical ligation transfers each monomer to a polymer growing on the rotaxane [70]. In this way the order of monomers on the axle determines the sequence of the synthesised polymer on the rotaxane. Notably, such artificial polymerases are not limited to natural polymers nor biological solvents. They have already demonstrated synthesis of a hexapeptide with three consecutive β-amino acids in DMF [71], an improvement on recent struggles to achieve just two consecutive β-amino acids by *in vitro* translation with genetically modified ribosomes [60]. This approach has been compared with translation in the sense that it has been used to synthesise peptides. However, as there are no codons or code being translated it is perhaps more similar to transcription in the 1:1 monographic cipher transferring of information between different polymers (axle and peptide).

The length limits of this approach were recently characterised: it takes significantly longer to synthesise a 6mer peptide (~4 days) than a 3mer peptide (~1.5 days) [72], or to synthesise a β-peptide (~7 days) rather than an equivalent length α-peptide (~3 days) [71]. The synthesis time increases non-linearly with polymer length for at least three reasons: 1) the distance between reactive groups increases as the polymer grows, 2) the polymer's secondary structure may begin to interfere with synthesis, and 3) the rotaxane has a greater distance over which to diffuse on the axle to reach the next monomer. The result of long incubation times is an increased

rate of background, untemplated reactions. For example, the rotaxane-driven synthesis of a repetitive polyleucine over 4 days resulted in a distribution of lengths between 3-11 monomers with a mean of 6.5. The equivalent untemplated reaction between an unthreaded rotaxane and axle had a distribution of lengths between 1-6 monomers with a mean of 2.9 [72]. The distribution of lengths is Poisson-like, which implies that the rate of templated reactions is only ~2.2× the rate of random, untemplated reactions, corresponding to an error rate of ~0.4. By comparison, this system closely resembles a class of unprogrammable enzymes called *non-ribosomal peptide synthases* [6], which produce peptides with non-canonical amino acids and bonds, with distribution of lengths between 3-50 amino acids and a mean of 11 [73].

Unfortunately, there is no possibility of turnover in this design, as the information in the program is destroyed when amino acids are cleaved from the axle by the rotaxane. Furthermore, there is no way to read or write information onto such supramolecular machines, meaning that applications like directed evolution are impossible. However, it is likely that such schemes will be combined with other advances in polymer chemistry. Sequence-controlled polymers (polyphosphates, polyamides, polyurethanes) up to 100mers have been synthesised by solid-phase synthesis, using tandem mass spectrometry (MS/MS) to sequence the resulting polymers [74–76]. The monomers have been designed specifically to make MS/MS interpretation simple [77] and one can imagine similar techniques for high-throughput parallel sequencing by optimising monomers for nanopore sequencing. A crucial step for future development of sequence-controlled polymers will be the addition of side-chain functionality, for instance, an information-carrying polymer produced by solid-phase synthesis might program rotaxane-based artificial ribosomes [70], or enable sequence-based recognition and self-assembly analogous to protein design [78], DNA origami [79] or DNA bricks [80]. If successful, this field may eventually develop enzymatic and computational nanomachines that are functional beyond the range of biopolymers (e.g., in organic solvents, >100°C, etc.).

1.2.4 DNA nanotechnology

An alternative to rationally designing new molecules is to use natural polymers that can be designed to self-assemble into unnatural objects. DNA nanostructures are simple to design using the rules of Watson-Crick base-pairing, while the highly optimised processes for DNA synthesis, editing and sequencing permit cheap prototyping of nanostructure designs and genetic libraries. The design of a wide range of 2D- and 3D-nanostructures has become possible over three decades of research [81]. DNA sequences can be designed to encourage the self-assembly of structures with the lowest free-energy and discourage the formation of competing structures. After synthesising and mixing the component DNA strands, structures fold spontaneously or upon annealing. The process of heating and slowly cooling eliminates native secondary structure within DNA strands and permits self-assembly beginning with the most stable bonds formed at high temperature. Using these techniques it is possible to make static nanoscale objects.

DNA nanomachines are made by triggering conformational changes between multiple stable DNA nanostructures [82]. Triggers can include externally controlled changes to the environment (e.g., nucleic acids, small molecules, pH, heat, light, enzymes, etc. [83–88]). Alternatively, a system may exhibit internally programmed conformational changes through the autonomous unfolding of a reaction cascade (e.g., enzyme-driven systems [89] or hybridisation chain reactions (HCR) [90]). DNA nanomachines have included switches and compartments that open and close [22, 91], autonomous walkers that move position [92–94], primitive computers [26], and more.

David Liu's group developed the first prototype of a DNA-based artificial programmable polymerase in 2010 (Figure 1.6). They brought together two previously separate lines of work in DNA nanotechnology, DNA templated synthesis (DTS) and DNA motors. In DTS, reactive moieties are attached via synthetic linkers to DNA strands and colocalised in a DNA nanostructure [95], in analogy with the colocalisation of rotaxane and amino acids on the axle described above [70]. In this way specific transfer reactions can be programmed using the sequence of DNA adapters bearing the chemical modifications for DTS. Liu's group combined

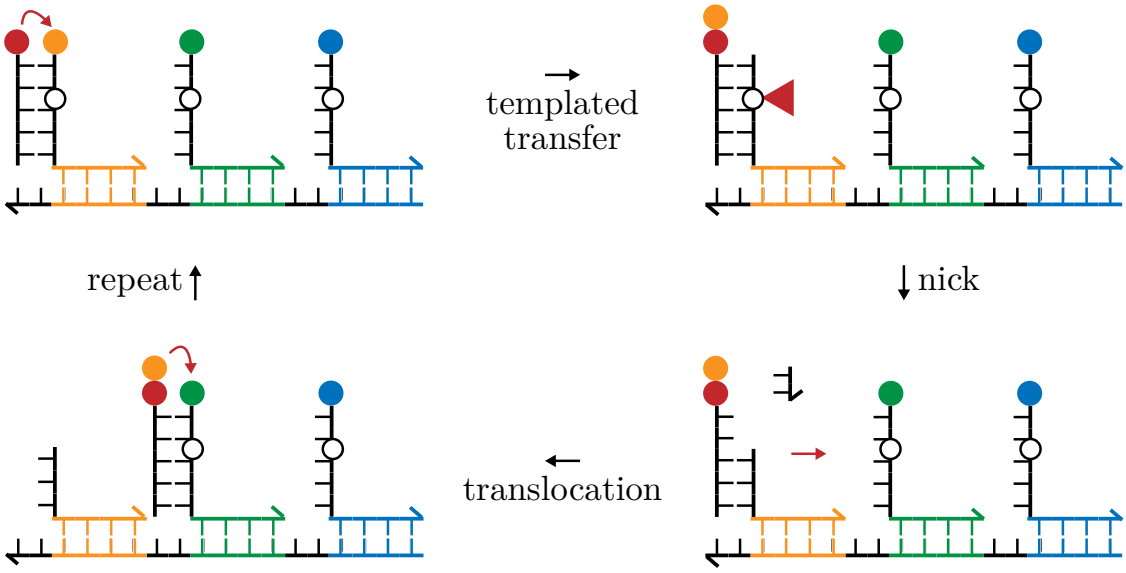


Figure 1.6: Approaches to build artificial programmable polymerases - DNA nanotechnology. A cycle of polymer synthesis can be achieved by combining DNA templated synthesis (DTS) [95] with a DNAzyme-based walker [93]. DNA strands called *legs* are covalently modified with reactive moieties (coloured circles) to form donor adapters for templated transfer reactions. A DNAzyme walks upon the track (left to right), by cleaving RNA nucleotides within legs (unfilled circles) and moving to the next leg where more base-pairs can be hybridised. The DNAzyme acts as an acceptor adapter during DTS. The DNA program or *gene* has 21nt subsequences called codons (coloured DNA) that determine the order of reactions and the sequence of the synthesised polymer [98].

DTS with a DNA walker - an RNA-cleaving DNAzyme that steps between legs on a DNA track [93]. At each position on the track Liu's DNA walker participates in sequential DTS reactions, resulting in polymer synthesis. The DNAzyme moves by cleaving an RNA nucleotide within the complementary DNA adapter. This causes the dissociation of the short ssDNA cleaved from the adapter, leaving the DNAzyme free to hybridise to the next adapter. This exchange of strands is called toehold mediated strand displacement (TMSD) [96, 97], as it is thermodynamically driven by the extra base-pairs in the toehold that becomes hybridised in the exchange. The DNAzyme translocates between adapters by TMSD. In this case a trimer peptide was autonomously synthesised by a DNAzyme according to the sequence of codons in a DNA program [98].

In summary, the most rapid progress towards artificial programmable polymerases has been made by top-down tinkering with natural polymerases, while

the greatest potential for diverse polymers lies with a bottom-up approach. DNA nanotechnology represents an interesting intersection of these two fields, as new designs of artificial programmable polymerases can be quickly conceived and tested. For this reason this thesis relies extensively on DNA templated chemistry and DNA nanomachinery. These subjects are introduced in more detail in Chapter 2 and Chapter 3 respectively. As there are common design issues that must be resolved for progress in the field, the next section attempts to articulate a general blueprint common to natural and artificial programmable polymerases.

1.3 AIMS

1.3.1 Blueprint of an artificial programmable polymerase

> *A systematic discussion of our present knowledge of protein synthesis could usefully be set out under three headings, each dealing with a flux: the flow of matter, the flow of information, and the flow of energy.*
>
> — Crick, 1958 [46]

Flow of matter. The first requirement for an artificial programmable polymerase is a polymer chemistry. This is not a programmable nor autonomous nanomachine, it is only the set of reactions that synthesise polymers from precursor monomers. The minimal set may include 1) *labile linkers*, 2) *monomers*, and 3) *active sites* (Figure 1.7A), where the labile linker and active site react to form a polymer backbone connecting monomers (Figure 1.7B). In translation the labile linker is the tRNA's 3'-ester, the monomer is the amino acid and its side chain, and the active site is the amino acid's primary amine. In transcription the labile linker is the triphosphate, the monomer is a nucleobase, and the active site is the 3'-hydroxyl. Figure 1.7C shows how a system may look in practice using N-Hydroxysuccinimide-ester (NHS-ester) labile linkers and primary amines as active sites, that interact to transfer a monomer between molecules in a peptide-bond forming reaction. Indeed, this strategy was used in peptide synthesis prior to the invention of solid-phase synthesis [99]. However, in order to produce a specific peptide rather than a mixture of random sequences,

the growing peptide was laboriously purified from and remixed with reactants at each step of synthesis. An additional feature is needed for programmability.

Flow of information. The second requirement is a way of determining the order of reactions during polymer synthesis. The three modules described above may be attached to a fourth module, called a *recognition unit*, with the capacity to position the reactive moieties in close proximity. I call this set of four modules an *adapter*, after Francis Crick's "adapter hypothesis" [46], as it abstractly describes the role of tRNAs in translation by the ribosome and the legs in Liu's peptide-synthesising DNA walker. Figure 1.7D demonstrates an example where two complementary adapters associate to colocalise 1) a polymer with an active site, and 2) a monomer attached via a labile linker. Colocalisation can determine the order of reactions by enhancing the effective concentration and rate of reaction between two specific monomers. If there are orthogonal recognition units, then multiple reactions that would otherwise interfere with each other can be carried out simultaneously in the same reaction mix [100]. The recognition unit provides a mechanism for translation, that is, a polygraphic cipher with N:M transfer of information between a program and synthesised polymer. This could take many forms (e.g., overlapping vs. non-overlapping codes, skipping or repeating regions of a specified length, etc.)

Templated synthesis has been demonstrated with coiled-coil peptides, alkanes, nucleic acids, and more [101–104], and has been used in the synthesis of <10mer peptides using DNA templated synthesis (DTS) [95, 105]. Interestingly, the choice of adapters limits the set of possible polymer chemistries (e.g., hybridised nucleic acid complexes have greatest stability in water, thus limiting the polymer chemistry to water-compatible reactions). Furthermore, the strength of the interaction between adapters determines programmability - a weak affinity between recognition elements could create a multistep cascade of indeterminate polymer length and sequence as adapters continually associate and dissociate. Programmability is possible with high affinity interactions, but in order to progress beyond a single transfer reaction it is then necessary to separate the adapters through external intervention such as

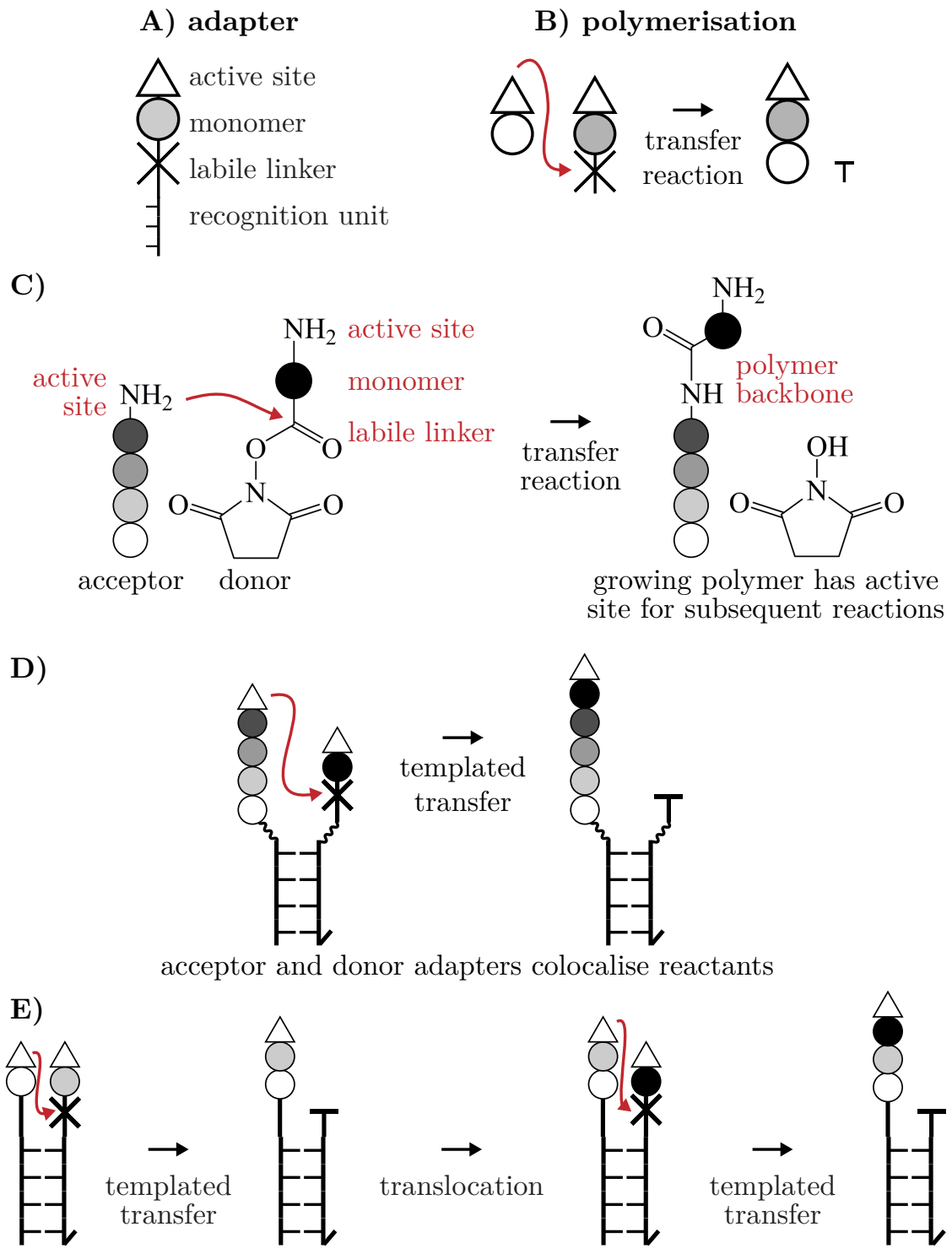


Figure 1.7: A) Adapters have three modules for reactions to synthesise polymers and a recognition unit for colocalisation. **B)** Active sites react with labile linkers to form polymer backbones. **C) Flow of matter.** An example polymer chemistry using NHS-ester labile linkers and primary amine active sites that extends a growing peptide by one monomer. **D) Flow of information.** The recognition unit enables colocalisation of adapters, increasing the effective concentration between reactants. In this way adapters can program multiple otherwise interfering reactions in reaction mix. **E) Flow of energy.** Molecular machines use fuel to drive autonomous cycles of transfer and translocation.

purification, heat, pH, or the substitution of strands by toehold mediated strand displacement (TMSD) [106–109]. An additional feature is needed for autonomy.

Flow of energy. The final requirement is molecular machinery enabling autonomous cycles of transfer reactions. Free energy must be consumed to irreversibly drive the cyclical association and dissociation of adapters, enabling programmed polymer synthesis. To avoid randomly cycling back and forth between reversible states, the machinery must bias a change of state by the consumption of free energy [31, 110]. Simultaneously, each change of state must trigger a new arrangement of adapters for polymer synthesis. In nature this corresponds to the thermal ratchet driving DNA polymerases that consumes dNTPs, or the energy consumed by the ribosome during tRNA aminoacylation, recruitment and peptide bond formation in the form of ATP and GTP. In the example of Leigh's supramolecular axle-rotaxane, the constrained diffusion and polymer-synthesising reactions represent a source of energy. In Liu's DNA walker the cleavage of the RNA backbone triggers molecular motion. If colocalisation is used to determine the order of reactions during polymer synthesis, then this implies the spatial rearrangement or relative motion of components between reactions. For this reason, I use the term *motor* for the system that converts energy into mechanical motion during polymer synthesis, though motors may take many forms beyond linear or rotational movement.

This provides a general blueprint for artificial programmable polymerases. Combining these three features is a significant challenge in itself. There are a few examples in the literature, such as a supramolecular system involving a diffusing rotaxane [70] and a DNA walker [98]. These examples indicate an array of challenges must be resolved before artificial programmable polymerases can begin to rival the length, fidelity, diversity, turnover, metastability and programmability of natural nanomachinery.

1.3.2 Challenges for the field

1. **Length.** The length of polymers synthesised by artificial programmable polymerases must be increased before useful functions are possible (folding,

binding, catalysis, etc.). A database of biologically active non-ribosomal peptides have a distribution of lengths between 3-50 amino acids and a mean of 11 [73]. From this I infer that polymers of useful lengths will be >10 monomers. Increasing polymer length might be possible by the following means:

(a) ***Stable reactants.*** Reactants for templated transfer reactions are typically reactive molecules in order to maximise yield. However, these also react with solvents or templates, resulting in the accumulation of non-reactive side products that limit the yield of long polymers. Local protection and deprotection of adapters would increase polymer yield and length. Mechanisms might be designed that are analogous to Ef-Tu protecting tRNAs for the natural ribosome.

(b) ***Catalysis.*** The yield of transfer reactions could be greatly increased by colocalising a catalyst of the transfer reaction. This would directly increase the rate of reactions but might also reduce the need for activated reactants that are prone to degradation. This is the role of the tRNA's 2'-hydroxyl in the natural ribosome [43, 44].

(c) ***Identical reaction environment.*** Templated transfer is driven by colocalisation of reactants. However, in the scheme depicted (Figure 1.7E) the distance between reactants changes and is extended by one monomer in every synthesis step. This has been shown to decrease the rate of transfer as the polymer grows longer [111], presumably due to two factors: 1) decreased effective concentration, 2) the possibility that polymer folding interferes with transfer reactions. An alternative design would take inspiration from the natural ribosome, which inserts monomers in an identical reaction environment at each synthesis step.

(d) ***Feedback.*** There is no coordination between transfer reactions and motor progression in the schemes depicted in Figures 1.6 or 1.7. If the rate of motor translocation is faster than the rate of transfer reactions,

then the resulting polymers are likely to contain deletions. If the motor is slower than the chemistry, then the motor idles at a given step while the subsequent adapters are degrading. This trade-off between length and fidelity limits the length of synthesised polymers. In contrast, the ribosome 'senses' when it has made a peptide bond, and only then releases the spent donor tRNA to accomodate the next acceptor tRNA. Feedback between the two processes would permit the motor to run as fast as possible without causing skips.

(e) ***Exponential polymerisation.*** Programmable polymerases typically extend polymers by one monomer after another in series, at best resulting in linear increase of polymer length over time. However, artificial systems might be devised in which monomers are connected in parallel (i.e. in Step 1 both reactions $A + B \rightarrow AB$ and $C + D \rightarrow CD$ occur, and in Step 2 the reaction $AB + CD \rightarrow ABCD$ occurs, and so on), resulting in exponential growth over time. In this way a 1024mer might be synthesised in 10 reaction steps rather than 1024 [14]. An alternative strategy might explore branched polymers with multiple active sites.

2. **Fidelity.** This depends on the specific application and the cost of errors, as for example protein translation has error rates ~10^{-4} while DNA replication may be as low as 10^{-10} [112]. A reasonable goal for fidelity is that >90% of polymers have the correct sequence, implying an error rate of <10^{-2} when synthesising 10mers.

   (a) ***Limit untemplated reactions.*** Transfer reactions can occur from random collisions between adapters rather than programmed colocalisation, causing errors during polymer synthesis. The rate of templated transfer reactions is typically much higher than that of untemplated reactions. However, as there are longer incubation times during multistep syntheses, untemplated reactions become more significant for longer polymers. Untemplated reactions could be limited by reducing the

concentration of all components or the total reaction time. Alternatively, colocalising a catalyst in the templated reaction environment might permit a less reactive transfer chemistry and thus limit the rate of untemplated reactions. Finally, the reaction environment might be shielded from random collisions, just as the PTC is buried within the ribosome.

(b) ***Limit out-of-order reactions.*** In addition to random, untemplated reactions, there is also the possibility of malfunctions in the execution of the program causing out-of-order reactions. In the example from Figure 1.6, this could happen if the DNAzyme stepped to the incorrect adapter during translocation, resulting in reduced fidelity. In this case, the bias towards the correct adapter is enforced only by the rigidity of the ssDNA track that serves as a program. Thus, the possibility of out-of-order reactions is determined primarily by the design of the molecular motor.

3. **Diversity.** Different monomer types are required so that different sequences can produce different folds and functions. The number of monomer types required to produce complex folds has been predicted to be >10 [113], however, it may turn out that more monomer types are required for new functions, such as catalysing exotic reactions or stabilising polymers in non-native conditions.

   (a) ***Arbitrary polymer backbones.*** Ideally, the system would be designed in a modular fashion so that the choice of polymer chemistry is independent of the molecular motor, adapters and templates performing polymer synthesis. To explore a new polymer backbone one would simply substitute a new set of monomers on adapters rather than redesigning the entire machinery.

   (b) ***Parallel translation.*** Multiple polymerases should be able to operate simultaneously on distinct programs to produce the set of corresponding polymers in one pot. This is a necessary step towards more complex

syntheses, such as a pool of polymers for directed evolution, or an ecosystem of interacting polymers mimicking the cellular milieu.

4. **Turnover.** Enzymes are not consumed in the reactions they catalyse. It should be possible for a single polymerase to synthesise multiple copies of a polymer from a single program. In future applications it would be ideal for some reactants to be recycled *in situ*, such as the activated adapters that are consumed during each polymer synthesis. This is analogous to tRNA synthetases recharging tRNAs for the ribosome.

5. **Metastability.** The free energy that drives molecular motors can come from many sources, such as interaction with external fields (e.g., heat, light, pH or electrostatic potential), polymer folding (e.g., hybridisation or rearrangement from kinetically trapped conformations), the breaking of covalent bonds in small molecule fuels (e.g., ATP), and more. Analogous small molecule fuels would be ideal for a programmable polymerase, as they are relatively stable over time and can be added in great excess of other reactants.

6. **Programmability.** The sequence of the program determines the sequence of the polymer. It should be easy, cheap and efficient to synthesise, edit and sequence programs. For example, a molecular program like DNA is ideal in this regard due to the optimisation of solid-phase synthesis, advances in enzymatic editing of DNA sequences, and a suite of technologies for high-throughput parallel sequencing.

This list identifies the central challenges for the field as it continues developing. The construction of programmable polymerases incorporating all these features is clearly beyond the scope of a single thesis, so it is worth considering where the greatest contribution may be made.

1.3.3 Contributions of this thesis

This chapter has described how the construction of artificial programmable polymerases raises questions about the minimal complexity required for open-ended evolution while creating powerful biotechnologies. I have sketched a blueprint showing how control over the flow of matter, information and energy can permit programmable polymer synthesis by autonomous nanomachines. I reviewed current approaches by others in the field and argued that the most rapid advances in design and prototyping can be made using DNA nanotechnology.

There are a few key challenges that need to be overcome before the synthesis of polymers of useful lengths will be synthesised by DNA-based artificial programmable polymerases. As a great deal of work has already been done optimising existing transfer chemistries and increasing reaction yield by using freshly synthesised reactants at each step, I have not focused on synthesising polymers in this thesis. Instead, my work explores new architectures for DNA templated synthesis (DTS) addressing the following three challenges:

1. Feedback to coordinate transfer and translocation reactions
2. Insertion of monomers in an identical reaction environment
3. Protection of adapters from degradation

Chapter 2 addresses the problem of coordinating the chemomechanical cycles of polymer synthesis and motor progression. Here I design and develop a new transfer chemistry that produces an enzyme substrate after DTS, so that DNA-acting enzymes can coordinate transfer reactions and translocation. Interestingly, rather than colocalising reactants via flexible linkers to DNA, these templated reactions occur in a tight reaction pocket resembling the DNA phosphate backbone.

Chapter 3 continues this work, but focuses on developing enzyme-driven DNA motors using the new transfer chemistry. My main contributions here are the design and demonstration of dynamic DNA nanomachinery using metastable fuel (dNTPs) and natural strand-displacing nucleic acid polymerases. This work also reveals the

issues of multistep polymer synthesis, where polymers are typically extended at an active site that changes and grows more distant during each transfer reaction.

Chapter 4 provides the background for *insertion polymerisation* that transfers monomers in an identical environment for each reaction. This is necessary for programmable polymerases to synthesise long polymers. Here I explore three new DNA architectures for DNA templated reactions for insertion, however, the failure of one DTS architecture highlights the need to protect adapters from degradation.

Chapter 5 explores adapter protection by changing the bulk and local environment of adapters. This includes the use of internal abasic sites to protect adapters from degradation, and explores the effect of DNA secondary structure, linker position and length on DTS yield.

I anticipate that some of the advances intended for artificial programmable polymerases will also be useful in general for dynamic DNA nanotechnology.

2

FEEDBACK I – ADAPTERS

To produce polymers with greater length and fidelity, programmable polymerases must coordinate the processes of catalysis and translocation so that one happens only after the other. No autonomous DNA nanomachine has achieved this to date. Here I design and develop a system using enzymes to coordinate the two processes. A new transfer chemistry is explored where DNA templated synthesis (DTS) reactions produce enzyme substrates. In this way transfer reactions could trigger enzyme recognition of DTS products, while a DNA-acting enzyme could in turn trigger translocation for the next DTS reaction. This work provides the foundation for the enzyme-driven DNA motors explored in Chapter 3. I begin by introducing the literature relevant for DNA-based artificial programmable polymerases.

2.1 INTRODUCTION

2.1.1 DNA templated chemistry

Reactions typically result from random collisions between particles. In order to simplify the design and analysis of reactions, it is typical to use few, purified substances in optimised conditions so that a specific reaction of interest tends to dominate the other possible outcomes. However, there are alternative ways to design reactions by controlling the local environment of reactants. Templated chemistry places reactants in close proximity so that their rate of reaction is increased. Colocalisation of reactants can occur when different parts of the same molecule react through an intramolecular intermediate, or alternatively, through non-covalent affinity between two molecules in a pseudo-intramolecular reaction [114]. Templated reactions have been demonstrated with a wide variety of macromolecules, for example, alkanes, peptides, nucleic acids, supramolecular motifs and more [101–104]. In these cases, if orthogonal interfaces between molecules can be designed, then each reactant can be positioned with a specific reaction partner rather than colliding with a random partner. Reactions that would normally interfere with each other can occur in the same reaction volume and the outcome of a reaction can be controlled despite the mixing of many, unpurified substances. Templates enable programmable chemistry through control of local rather than bulk reaction conditions.

DNA is a platform for programming chemical reactions. As DNA sequences can be very diverse, there is the potential for an enormous number of different DNA templated chemical reactions occurring in parallel ($<10^{14}$ in 1mL of 1 µM DNA). DNA adapters covalently modified with reactive groups can self-assemble into architectures where the reactants are held in close proximity (Figure 2.1). This increases the rate of reactions between the colocalised reactants. As such, DNA templated synthesis (DTS) enables reaction specificity - multiple otherwise incompatible or interfering reactions can occur in the same reaction volume when each reactant is colocalised with a specific reaction partner by DNA [100]. Finally, by adding or removing the DNA strands required to form the DTS complex it is

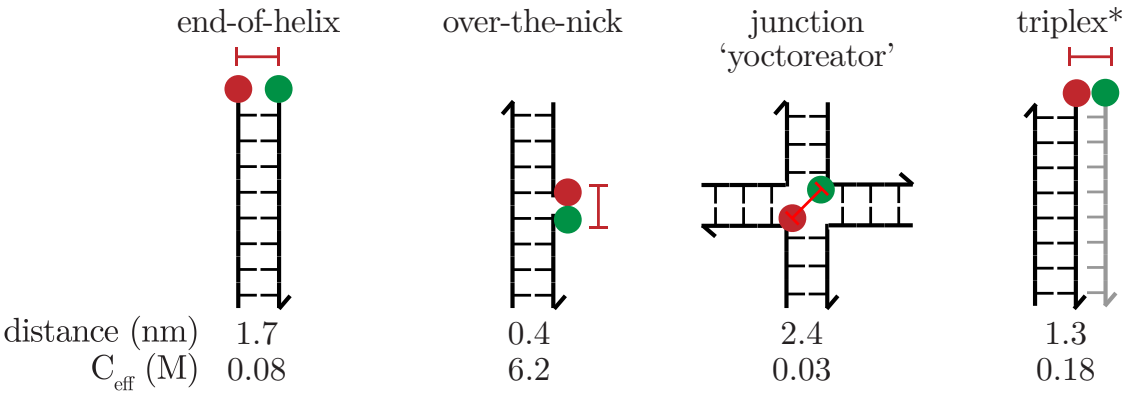


Figure 2.1: DTS architectures. Self-assembling DNA nanostructures colocalise reactive groups (coloured circles) for templated reactions. Distances between the attachments points for modifications were measured from crystal and NMR structures (PDB: 1BNA, 1WAN, 2CRX [116–118]). The effective concentration (C_{eff}) between attachment points is estimated in Appendix 7.3.1. *The triplex architecture has only been used for templated conjugation over-the-nick in triplexes [119].

possible to control when and where reactions occur [109, 115]. All these features are relevant for artificial programmable polymerases, so it is important to understand 1) how DNA templates enhance reaction kinetics, and, 2) the types of DNA templated reactions that are possible.

Two definitions have been used in attempts to quantify the effect of proximity on reaction kinetics, originating in the characterisation of ring-closing intramolecular reactions [120]: 1) the effective molarity (M_{eff}), an empirical observation of the rate enhancement caused by proximity, and, 2) the effective concentration (C_{eff}), a theoretical prediction of the number of atoms in a local reaction volume rather than the ensemble average of the bulk volume.

$$M_{\text{eff}} = \frac{k_{\text{intramolecular}}}{k_{\text{intermolecular}}}$$

M_{eff} has units of M (mol.L^{-1}), as $k_{intramolecular}$ is the rate of a pseudo-first-order reaction (s^{-1}) if the DTS complex is considered as a single macromolecule templating an intramolecular reaction, and $k_{intermolecular}$ is the rate of a second-order reaction ($M^{-1}s^{-1}$) between untemplated reactants. M_{eff} represents the theoretical increase in reactant concentration required to achieve the same observed rate with the intermolecular reaction. The reports of M_{eff} in DTS vary greatly from ~5M [121],

~25M [122], ~150M [111] to ~1000M [95]. Some of these estimates represent physically impossible concentrations (indeed $M_{eff} > 10^6$M have been reported for supramolecular catalysts [123]), but this may be explained by the additional effects of tethering and colocalisation on $k_{intramolecular}$, such as changes in reactant orientation, linker entropy and solvation. The large variation between estimates of M_{eff} during DTS may be explained by the different DNA architectures and linker chemistries between systems, and whether or not the researchers accounted for the rates of DNA hybridisation and reactant stability during DTS.

In contrast to M_{eff}, the effective concentration (C_{eff}) is a theoretical prediction of the number of atoms within a local volume (also in units M), which may be enhanced relative to the bulk concentration by non-covalent affinity or covalent bonds [120]. C_{eff} is correlated with M_{eff}, but provides physically plausible concentrations as it excludes the effect of reactant orientation, linker entropy, solvation, etc. Depending on the specific DTS architecture and linker lengths, I estimate C_{eff} ~0.01 - 10M for DTS (Appendix 7.3.1) based on measured distances between linker attachment points on DNA crystal structures. Ultimately, the rate of a templated reaction may be enhanced by a DNA architecture and high effective concentration, while the yield will also depend on the specific type of templated reaction.

Templated conjugation between DNA strands has been demonstrated with a diverse array of DNA architectures and chemical reactions. These have been reviewed extensively elsewhere [124, 125], but generally include conjugation between adjacent nucleotides using synthetic leaving groups (Figure 2.2A) or conjugation between synthetic linkers that do not resemble nucleotides (Figure 2.2B). Many DNA templated reactions have been demonstrated, with the basic limitation that reactions must be compatible with water and DNA. For example, they include nucleophilic substitution, reductive amination, amine acylation, click chemistry, Heck coupling, 1,3-nitrone cycloaddition, reversible photo-induced cycloaddition, boronic ester formation, Glaser-Eglington coupling, and more [124, 125]. This approach has been used to form libraries of small molecules that have been conjugated on the end of a DNA duplex [126–128]. Applications have also included biosensors, where a

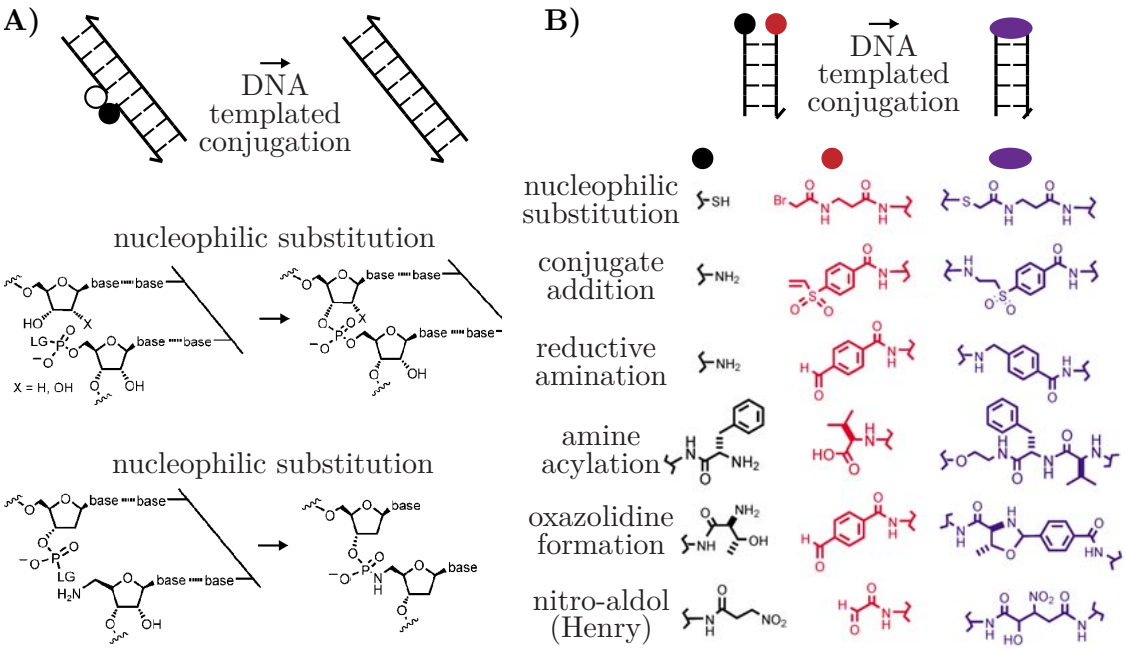


Figure 2.2: DNA templated conjugation. A) Examples of over-the-nick DNA templated conjugation between adjacent nucleotides form products that resemble a nucleic acid backbone. **B)** Examples of end-of-helix DNA templated conjugation forming unnatural nucleic acid constructs. Figure adapted with permission from [124] © 2004 John Wiley and Sons.

conjugation reaction with a fluorescent product is templated in the presence of a specific target nucleic acid [129]. By templating conjugation between linear and branched molecules, DNA templated nanowires were constructed with the potential for conductivity across a nanoscale circuit [130]. Finally, the majority of work in this area has explored how multiple rounds of templated conjugation between activated nucleotides may lead to non-enzymatic replication of information in primitive living systems [131]. This work demonstrates that DNA is a versatile platform for designing chemical reactions to template the synthesis of small molecules, oligomers and polymers.

Translation appears to require adapters. Replication and transcription involve the 1:1 copying of information between polymers. Ideally, an artificial programmable polymerase would permit N:M translation of information between a program and a synthesised polymer. In transcription, the program and synthesised polymers are of similar shape and type, whereas in translation the program and polymers may be very different (e.g., the difference between RNA and protein). Francis Crick

hypothesised that adapters overcome this difference as an intermediate with an orthogonal interface for recognises the program connected to a monomer of arbitrary shape and type [46]. The use of adapters implies both conjugation reactions that form the polymer backbone and cleavage reactions that prevent the adapter itself from becoming part of the synthesised polymer.

In practice, there are many ways to implement translation during polymer synthesis from DNA. Polymers can be translated from a DNA program using a stepwise process of 1) hybridisation of an adapter to the program, 2) DNA templated conjugation between the program and the adapter, 3) adapter cleavage by changing external conditions, 4) purification of the DNA program, and 5) repeating the process for the addition of subsequent monomers (Figure 2.3A). Additional steps can also be added to form macrocycles rather than linear polymers in the final product [126]. After optimising this process, a library of 256,000 macrocycles was translated from DNA programs and used for subsequent selection experiments [128]. Rather than the laborious serial approach, adapters may be designed so that all conjugation and cleavage reactions occur simultaneously (Figure 2.3B). For example, a set of macrocyclic peptide nucleic acid (PNA) adapters were hybridised in adjacent positions on a DNA program, permitting templated conjugation between each adjacent PNA adapter. The PNA recognition unit was subsequently cleaved to produce a polymer with a sequence programmed by the DNA template [132]. This is a remarkable effort, though there are important limitations. Firstly, macrocyclic adapters create size limits for monomers, which must be as large as the PNA anticodon sequence in the adapter. Secondly, adapters were sufficiently flexible to cyclise and self-conjugate. To prevent this from occurring each alternate adapter in the sequence was modified with either two alkyne or two azide modifications, so that intramolecular cyclisation of adapters was impossible but templated conjugation between colocalised adapters still occurred. However, this meant that two versions of each adapter had to be synthesised depending on whether the monomer was encoded at an odd- or even-numbered position in the DNA program.

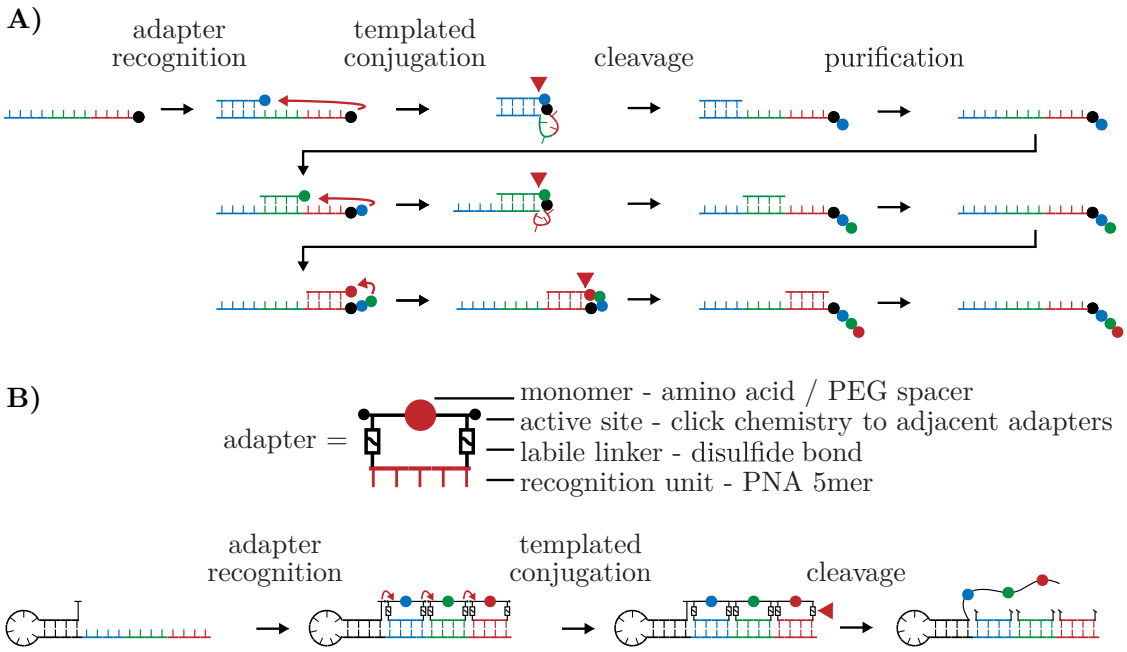


Figure 2.3: Translation of polymers with DTS. A) Sequential translation. A library of DNA programs was synthesised, and a library of macrocycles translated from them in one-pot [126]. Adapters are hybridised to the DNA program so that two-step process of templated conjugation and cleavage transfers a monomer from the adapter to the DNA program. In this case adapters were conjugated by amide coupling and cleaved via pH-sensitive linkers (bis(2-(succinimidooxycarbonyloxy)ethyl)sulfone). DNA programs were PAGE purified between successive DTS reactions. **B) Parallel translation.** A set of macrocyclic PNA adapters were synthesised, bearing a monomer (PEG spacer or amino acid), two sites for conjugation to adjacent adapters via Cu(I)-catalysed alkyne-azide 1,3-dipolar cycloaddition, two disulfide bonds that could be cleaved in reducing conditions, and a 5mer PNA sequence to recognise a DNA program [132]. In this case all adapters were hybridised simultaneously before subsequent conjugation and cleavage reactions.

In these systems two sequential conjugation and cleavage reactions were required. Single step transfer reactions are also possible and indeed may be preferable for an autonomous molecular machine as they reduce the number of operations. DTS adapters for single step transfer reactions can be prepared by chemically modifying DNA strands with the reactants for polymer synthesis. A thiol-maleimide reaction enables the addition of the labile linker NHS-ester (N-hydroxysuccinimide-ester) to disulfide-modified DNA strands, creating a donor adapter for DTS (Figure 2.4A). NHS-esters are highly reactive towards nucleophiles, which could be a primary amine active site on a colocalised DNA strand (aminolysis = DTS) or the surrounding water (hydrolysis = adapter degradation). In this case, DTS transfers

a monomer from a *donor adapter* bearing the thioester-monomer modification to an *acceptor adapter* bearing the amine, forming a peptide bond by acyl transfer. If the transferred group is bifunctional and also bears an active site − if the monomer has a primary amine − then the acceptor adapter may participate in subsequent DTS reactions. This strategy works provided the monomer is sufficiently rigid to prevent adapters self-destructing through cyclic intramolecular reactions. Alternative labile linkers have included thioesters [133] and phenol esters [S. Núñez-Pertíñez, personal communication], though they might include other conventional precursors for peptide synthesis (e.g., pentafluorophenol esters and 2,4,5-trichlorophenol esters [134]). There are alternatives to peptide bond forming reactions (Figure 2.4B). For example, a triphenyl phosphine labile linker can react with a nearby aldehyde active site (Wittig olefination = DTS) or the surrounding water (oxidation = adapter degradation) [135]. Sequential DTS reactions produce a polymer with an alkene backbone between monomers [106]. The choice of chemistry for labile linkers and active sites thus determines the nature of the polymer synthesised by artificial programmable polymerases.

DNA-based artificial programmable polymerases colocalise each adapter in a sequence of reactions programmed by a DNA template. The exchange of adapters can be externally controlled. For example, adapters of different lengths may be added to the reaction one after the other, so that each adapter displaces the last one in the cascade by toehold mediated strand displacement [109] (Figure 2.5A). This design has limited programmability because each adapter must contain information about the previous steps in the cascade. A unique set of adapters must be synthesised for each polymer sequence, rather that using a single set of adapters to synthesise any possible polymer sequence. An alternative approach designed adapters with different lengths and sequences to create different melting temperatures. In this way it was possible to crudely control the sequence of hybridisation of adapters using the bulk reaction temperature [107] (Figure 2.5B). This has the same limitations as the previous scheme, with the additional limitation of that the change in temperature

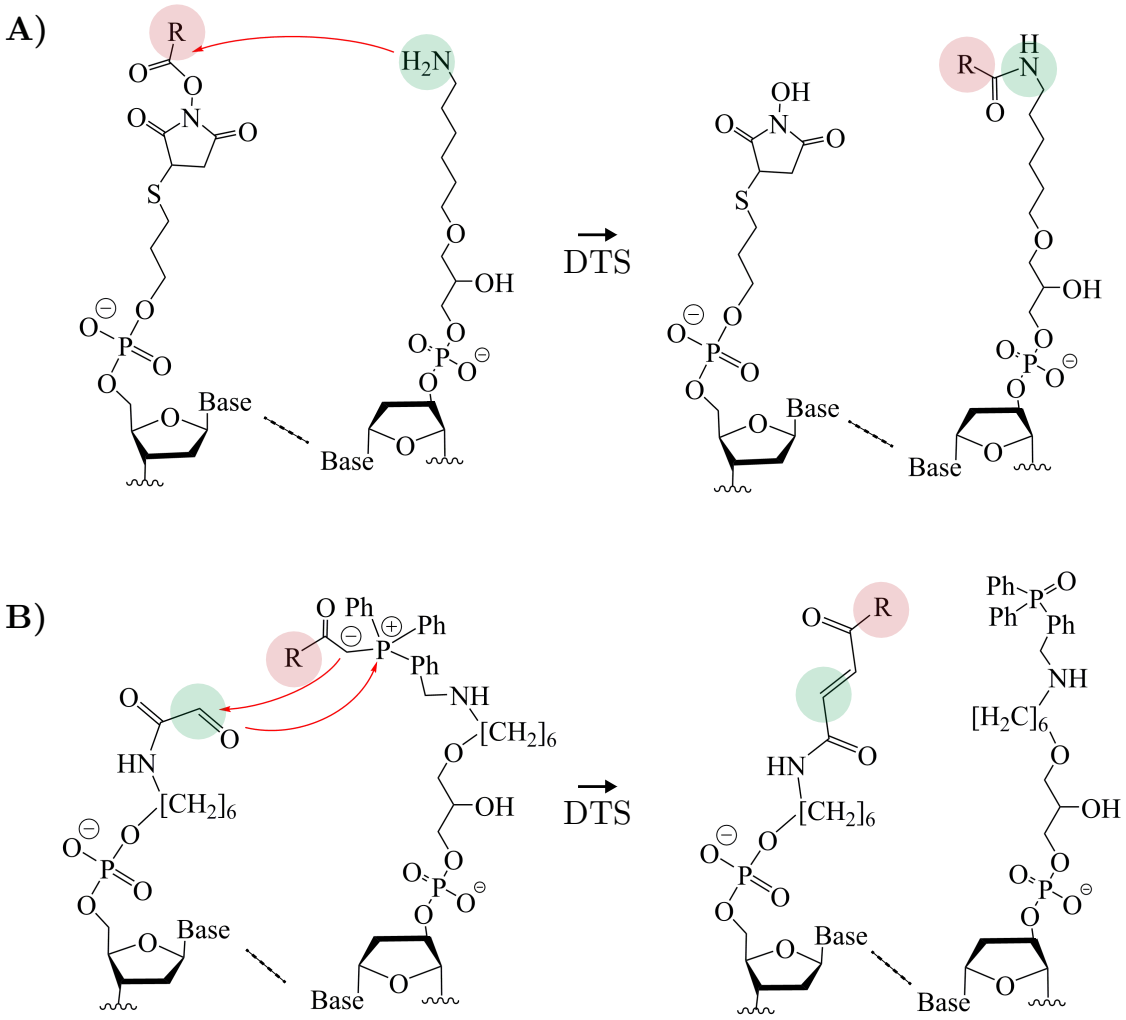


Figure 2.4: DTS transfer reactions [106]. A) DNA templated acyl transfer using NHS-ester-R donor adapters and amine acceptor adapters in an end-of-helix architecture. If the R-group contains a primary amine, then subsequent multistep reactions are possible using the same adapter. **B)** DNA templated Wittig olefination using triphenyl phosphine-R donor adapters and aldehyde acceptor adapters (Ph = phenyl). If the R-group contains an aldehyde, then subsequent multistep reactions are possible.

alters reaction yield at each step, and that polymer length is limited by the number of discrete temperature intervals.

Ideally, an artificial programmable polymerase would be an autonomous nanomachine where each reaction precedes the last without external intervention. David Liu's group combined one-step DTS transfer reactions with a DNA walker – an RNA-cleaving DNAzyme that steps between legs on a DNA track [93]. At each position on the track Liu's DNA walker can participate in a DTS reaction, resulting in polymer synthesis (Figure 2.5C). The DNAzyme moves by cleaving

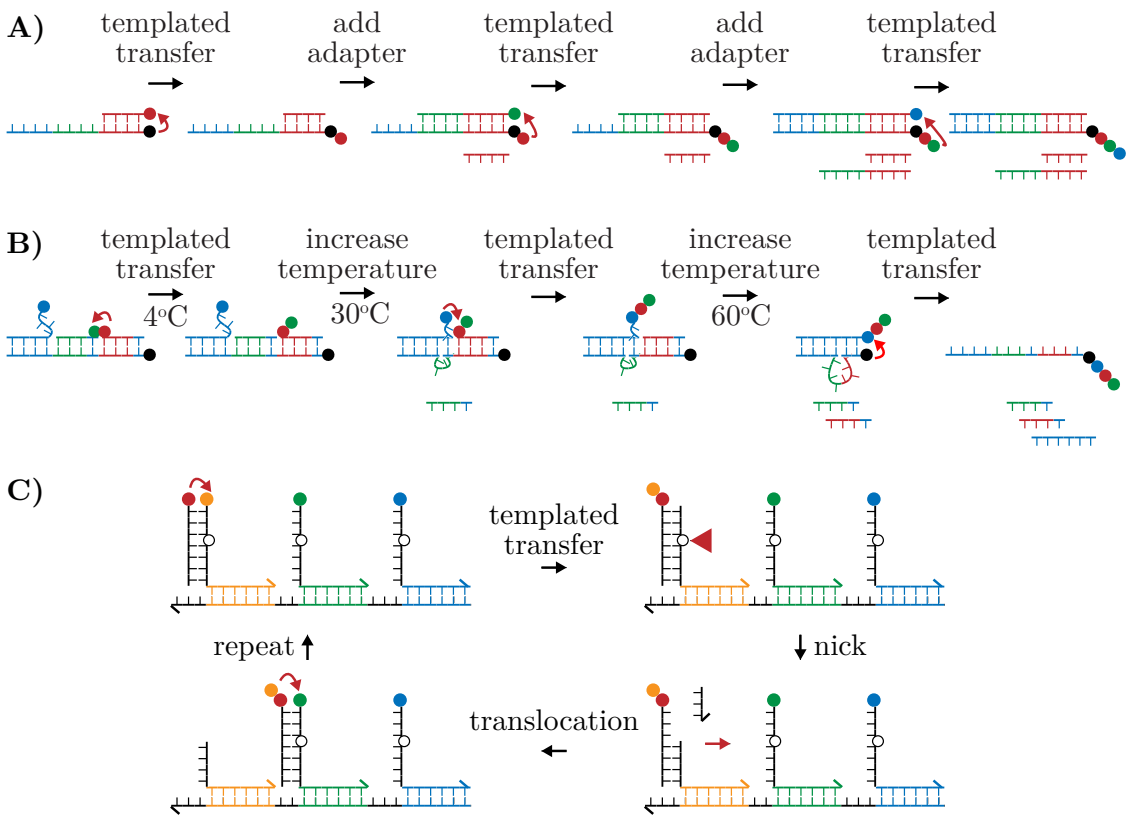


Figure 2.5: Programming sequential DTS reactions. **A)** A 6mer peptide was synthesised using acyl transfer reactions by adding longer adapters that displace shorter adapters by toehold mediated strand displacement (TMSD) [109]. **B)** A 3mer polymer was synthesised using DNA templated Wittig olefination to produce an alkene backbone [107]. Adapter lengths and sequences were designed so that upon incubation at 4°C, 30°C, and 60°C different DNA architectures would form, permitting sequential DTS reactions without adding reagents. **C)** A cycle of polymer synthesis can be achieved by combining DNA templated synthesis (DTS) with a DNAzyme-based walker [98]. Coloured circles represent monomers. Unfilled circles represent RNA nucleotides cleaved by the DNAzyme.

an RNA nucleotide within the complementary DNA adapter. This causes the dissociation of the short ssDNA cleaved from the adapter, leaving the DNAzyme free to hybridise to the next adapter. This exchange of strands is called toehold mediated strand displacement (TMSD) [96, 97], as it is thermodynamically driven by the extra base-pairs in the toehold that become hybridised in the exchange. The DNAzyme translocates between adapters by TMSD. In this case a trimer peptide was autonomously synthesised by a DNAzyme according to the sequence of codons in a DNA program [98].

This review has made it clear how DNA can be used to template chemical

reactions. This makes it possible to design nanomachines that translate information between a DNA program and a synthesised polymer. In the next sections I introduce the problem of feedback and a potential solution.

2.1.2 Coordinating DTS and DNA motors

An autonomous programmable polymerase based on DNA nanotechnology requires both DTS and a DNA motor using an energy source to drive cycles of transfer reactions. The lack of feedback between these two processes creates a contradiction between the demands for fidelity and length in polymer synthesis. If the rate of motor stepping is faster than the rate of transfer reactions, then the resulting polymers are likely to contain skips. If the motor is slower than the chemistry, then the motor idles at a given step while the subsequent adapters are degrading. This contradiction represents a fundamental design flaw of DNA-based artificial programmable polymerases [98, 136].

In contrast, nucleic acid polymerases and the ribosome appear to detect the progress of a transfer reaction, and wait until this occurs before moving into position for the next transfer reaction. Indeed, the ribosome uses the energy of peptide bond formation as part of the mechanism driving translocation [41, 42]. Introducing feedback between DTS and a DNA motor would allow an artificial programmable polymerase to run as fast as possible without causing skips and thus achieve higher yields of longer polymers. Furthermore, designing such mechanisms may be of general interest to the field of DNA nanotechnology by adding levels of complexity to current applications in molecular computation, robotics, drug delivery, and more [26, 137, 138].

2.1.3 Enzyme substrates produced by DTS

Feedback means that transfer reactions and translocation events are coordinated so that one happens after the other. There are many ways to implement feedback. In one example described previously (Figure 1.5), the one-dimensional diffusion of a supramolecular rotaxane may be blocked until a transfer reaction has occurred

[70]. Only after each transfer reaction has occurred can the rotaxane diffuse to the next transfer reaction. While DNA-based rotaxanes have been devised [139], the scale of DNA rotaxanes (10-100nm) relative to monomers (0.1-1nm) complicates this approach to blocking rotaxane progress.

An unexplored concept for implementing feedback would exploit the exquisite substrate specificity of natural enzymes. If the linker on the pre-DTS adapter blocked substrate recognition by an enzyme, while the cleaved post-DTS adapter was a viable enzyme substrate, then enzymes could be used to implement feedback by triggering translocation only after DTS. As various DNA-acting enzymes may be used to drive DNA motors, it should be simple to design a DNA motor exhibiting feedback (this is explored further in Chapter 3).

This approach requires redesigning DTS adapters to create enzyme substrates after DTS. Conventionally, DTS has used synthetic linkers that do not resemble natural enzyme substrates (e.g., Figure 2.4A-B). There are multiple ways to create enzyme substrates after DTS (Figure 2.6A-C). Monomers might be attached: 1) to the ribose 3'-hydroxyl forming a 3'-ester as an acyl transfer donor, 2) to the 5' phosphate to form a phosphate-ester, or 3) to the nucleobase itself provided that the modification sterically inhibits enzyme activity. The first and third strategies have been demonstrated in the field of DNA sequencing-by-synthesis [140–147]. In this application, fluorophores are attached by labile linkers to the 3'-hydroxyl of nucleotides in order to block polymerase activity. After the polymerase adds a nucleotide to an oligomer of unknown sequence, the polymerase cannot continue synthesis because the 3'-hydroxyl is unavailable. The identity of the added nucleotide is observed by the wavelength of fluorescence. The 3' terminator is then cleaved using an external trigger (e.g., reducing agent, transition metals, light, etc.), producing a 3'-hydroxyl for the polymerase to add the next nucleotide terminator in the sequence. This work demonstrates that the concept of combining chemistry and enzyme-based recognition to control polymerisation is plausible.

For two reasons the first scheme was chosen, attaching monomers to DNA strands through a 3'-ester labile linker (Figure 2.6A). Firstly, 3'-esters are more

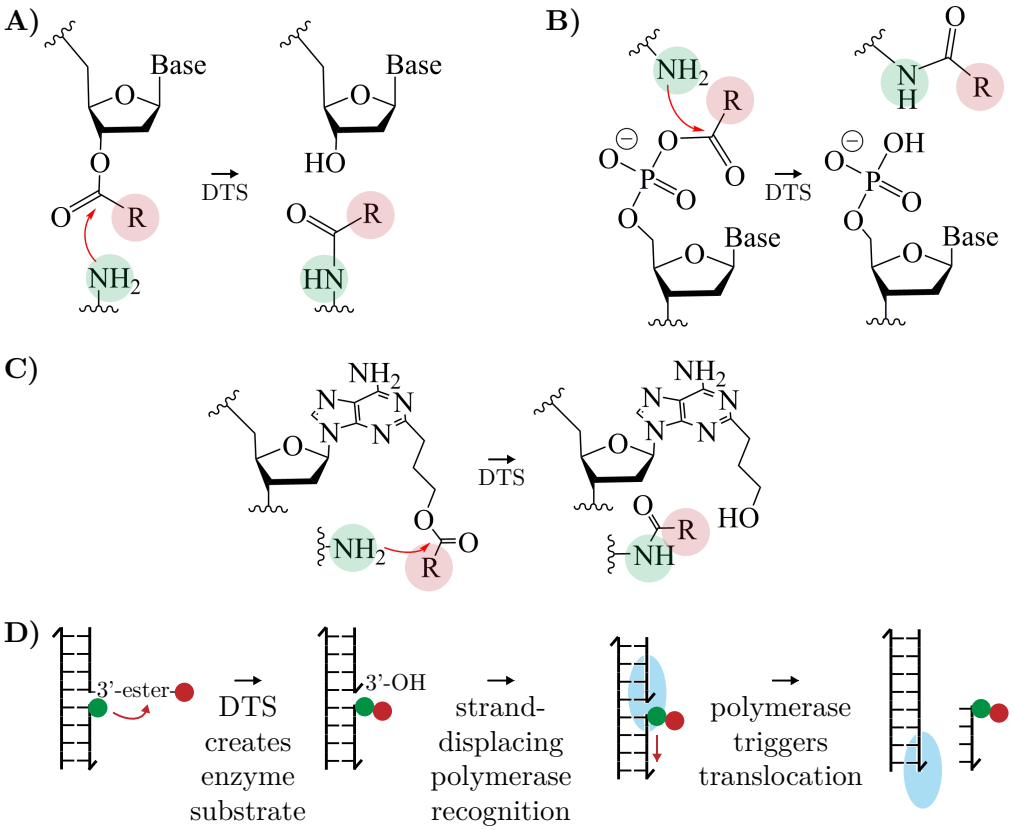


Figure 2.6: Redesigning labile linkers to produce enzyme substrates after DTS. A) Adapters with 3'-ester-modifications would inhibit polymerase or nuclease activity on the 3'-hydroxyl until after DTS occurred. This would permit a DNA-acting enzyme to recognise the 3'-hydroxyl and implement feedback by triggering translocation only after DTS. **B)** Similarly, adapters with 5'-phosphate-modified adapters would block 5' ligase or nuclease activity until after DTS. **C)** Nucleobase-modified adapters might inhibit enzyme activity until after DTS, provided that the monomer (R) is bulky enough to sterically hinder enzyme activity before DTS and the remaining linker after DTS is small enough to enable enzyme activity. **D)** An example scheme using 3'-ester adapters in an over-the-nick architecture, demonstrating how DTS could trigger the exchange of DNA strands required to drive a DNA motor.

stable to hydrolysis than 5'-phosphate-esters [148]. Secondly, in ribosomal protein synthesis the tRNA 2'-hydroxyl catalyses the reaction on the neighbouring 3'-ester, accelerating peptide bond formation by between 10^3-10^6-fold [43, 44]. There is an opportunity to explore this concept in DTS using the 2'-hydroxyl to catalyse transfer reactions. To practically implement this change in linker design it must be possible: 1) to synthesise donor adapters with the 3'-ester chemistry, and, 2) for them to exhibit DTS.

1) ***Adapter synthesis.*** Conveniently, 3'-ester modified nucleic acids can be

enzymatically synthesised using *flexizymes*. A flexizyme is a ribozyme originally evolved by Hiroaki Suga's group to aminoacylate tRNAs with non-canonical substrates in order to diversify substrates of the natural ribosome [57, 58]. Three components are mixed in a single reaction: the RNA flexizyme, the nucleic acid adapter to be modified, and a small molecule substrate (e.g., 3,5-dinitrobenzyl ester [DBE] - monomer). The flexizyme recognises the DBE handle but not the monomer, meaning that flexizymes can charge adapters with a broad set of monomers. Importantly, flexizymes are active on DNA substrates [149], so it is plausible to synthesise suitable adapters for DTS using this strategy.

2) *Adapter DTS.* In contrast, DTS with the new linkers was unprecedented. Due to the uncertainty of success of this reaction, I created a single 3'-ester donor adapters and tested various positions for the amine active site on the acceptor adapter. For example, the length of the linker tethering the nucleophilic amine to the acceptor adapter was varied. In an over-the-nick DTS architecture (Figure 2.1A), the number of nucleotides in the nick between the donor and acceptor adapters was also varied. This range of DTS architectures maximised the chance of observing DTS with the new 3'-ester transfer chemistry.

The next section aims to 1) demonstrate synthesis of the newly designed 3'-ester modified donor adapters, 2) demonstrate DTS, and 3) explore DTS-triggered enzymatic recognition. This is the key to building an artificial programmable polymerase exhibiting feedback between DTS and an enzyme-driven DNA motor.

2.2 RESULTS

2.2.1 Flexizyme synthesis of DNA adapters

This experiment aimed to produce 3'-ester modified adapters. A collaborator, Samuel Núñez-Pertíñez (O'Reilly Group, University of Birmingham), synthesised small molecule substrates for the flexizyme. The substrate DBE-biocytin was chosen (Figure 2.7A), where biocytin is a composite of biotin and lysine. It was chosen because flexizymes were originally evolved using biotin substrates, and

the biotin-streptavidin interaction can provides a useful assay of the position of the monomer before and after DTS.

The flexizyme structure is composed of two hairpins containing an interior bulge that binds DBE substrates [150]. The 3' end of the flexizyme has a sequence of variable length (4-10nt) that hybridises with the adapter, positioning the 3' end of the adapter near the DBE substrate for the acylation reaction. The longest length (10bp) of flexizyme-adapter complementarity was chosen, as this was previously shown to have the highest yield of adapter acylation [151].

I combined DBE-biocytin with commercially synthesised RNA flexizyme and DNA adapter. These components were incubated for 15 hours at 4°C in 20% DMSO (from DBE-biocytin stock in 100% DMSO) before running the samples on acid PAGE (Figure 2.7B). Acid PAGE (pH 5) was used to visualise the flexizyme product because the low pH limits ester hydrolysis. It also ensured the biocytin's amine was protonated (pKa ~9-10), so that despite the small change in mass between the adapter and adapter-3'-ester-biocytin, there is a difference in electrophoretic mobility due to charge. Furthermore, a much slower moving band was observed if streptavidin was added to the flexizyme reaction mix before acid PAGE, indicating that the specific biocytin modification was made to the DNA adapter. Before continuing, this method was used to test a range of conditions to optimise the yield of flexizyme synthesis (~40%) at pH 8 for 15 hours with 2:1:1000 ratio of flexizyme:adapter:DBE-monomer (Appendix 7.2.1).

The flexizyme reaction products were further characterised by mass spectrometry (MS, Figure 2.7C). The reaction mix contained three major peaks around the expected masses of the flexizyme, adapter and adapter-3'-ester-biocytin, with an additional mass 16Da larger than adapter-3'-ester-biocytin. This has been observed previously during biotinylation of DNA [152], and is believed to be formed by oxidation of biotin to form biotin-sulfoxide. While the dissociation rate of avidin from biotin-sulfoxide is 20-200× higher than biotin [153, 154], this still implies a stable complex with K_d ~10^{-13}M which is suitable for this application. Together this indicated successful synthesis of a donor adapter − the next step explored DTS.

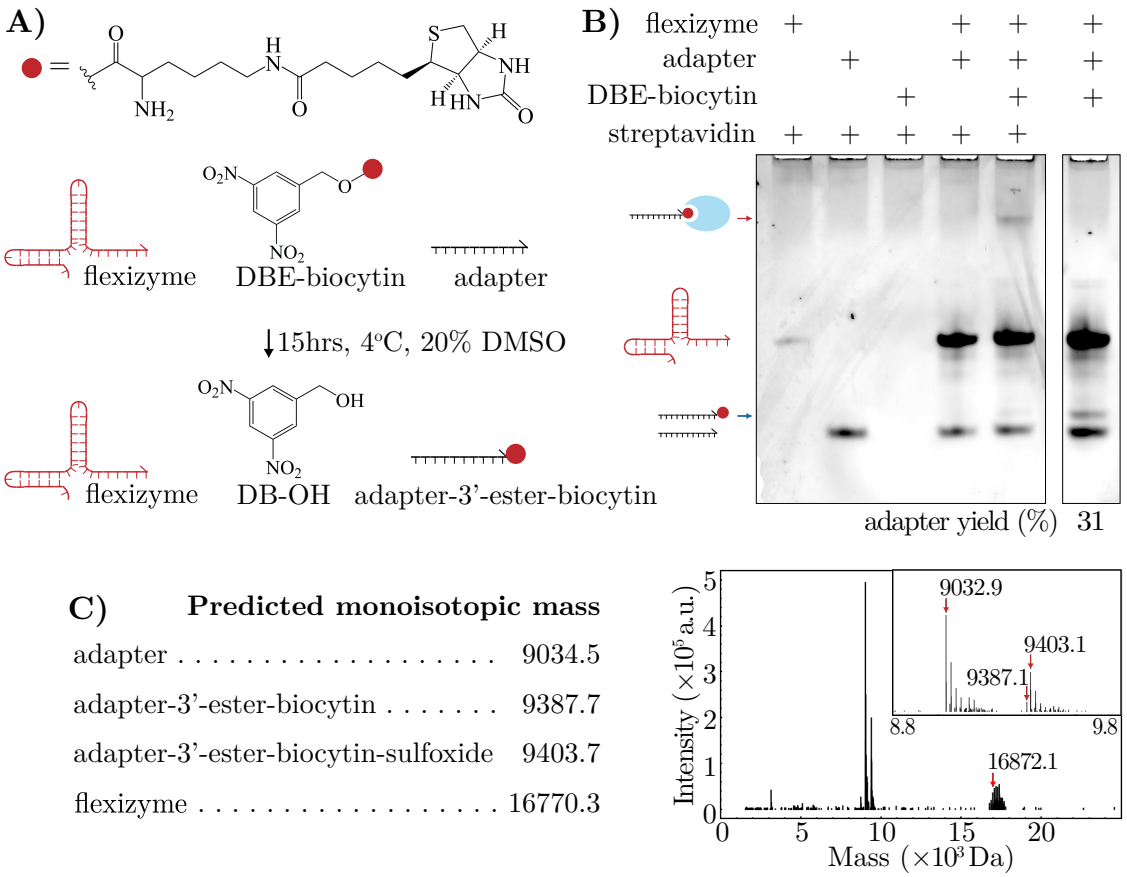


Figure 2.7: Flexizyme synthesis of DTS adapters. **A)** Scheme for flexizyme synthesis of 3'-ester-biocytin adapters. DBE-biocytin was synthesised by Samuel Núñez-Pertíñez (O'Reilly Group, University of Birmingham). Method in Appendix 7.1.2 and 7.1.3. **B)** 20% acid PAGE (pH 5) of the flexizyme reaction stained with DNA-binding dye SYBR Gold. Adapter-3'-ester-biocytin binds streptavidin, resulting in a slowly migrating band (red arrow). In the absence of streptavidin a slightly slower band is also visible (blue arrow). Gel densitometry indicates ~31% conversion to adapter-3'-ester-biocytin. Method in Appendix 7.1.5, 7.1.7 and 7.1.8. **C) MS of flexizyme reaction.** The charge-deconvoluted mass spectrum is depicted, with major peaks corresponding to expected species labelled. The inset shows the mass spectra between 8800 and 9800 Da. Comparing peak intensities indicates ~32% conversion to adapter-3'-ester-biocytin. Method in Appendix 7.1.9.

2.2.2 DTS with 3'-ester adapters

Here I aimed to test the 3'-ester adapters as donors during DNA templated synthesis (DTS) reactions. As the 3'-ester chemistry had not previously been used in DTS reactions, I explored a range of locations, orientations and lengths of reactive groups participating in DTS. It is possible to observe DTS products by PAGE, for example, if the monomers transferred between strands alter electrophoretic mobility or fluorescence. When demonstrating DTS and estimating DTS yields, PAGE is often cheaper and quicker than MS.

The flexizyme reaction mix was buffer-exchanged to remove DMSO from the flexizyme reaction. The DTS components were then added to the flexizyme reaction mix. This triggered self-assembly of an over-the-nick architecture for DTS. The 3'-ester-biocytin end of the donor adapter was colocalised with the 5'-amine modification of an acceptor adapter by hybridisation of both to a common strand called a *template* (Figure 2.8A). I used a 30bp region of complementary sequence between the template and donor adapter to ensure that the template had a greater affinity for the adapter than the 10bp complementarity with flexizyme. Samples were incubated overnight (~15 hours) at 20°C and pH 11 before hydrolysing any remaining donor adapter-3'-ester-biocytin by incubating the sample for 20 minutes at 95°C. I then added streptavidin to bind any acceptor adapter-5'-NH-biocytin produced by DTS. If DTS occurred I would expect to see a slower band on a native PAGE gel due to the biotin-streptavidin interaction with the acceptor adapter. If DTS failed to occur, this band would be absent. To test a range of positions and orientations for DTS, a variety of acceptor adapters and templates were tested against a single donor adapter with the 3'-ester-biocytin modification (Figure 2.8).

Slower migrating bands are only visible on PAGE with the shortest possible linker (C0, Figure 2.8B-C), where the amine active site is attached to the 5' carbon of the ribose. Note that a long smear of nucleic acid appears lower in the gel due to the hydrolysis of the RNA flexizyme at high pH, temperature and $[Mg^{2+}]$. This is evidence that DTS is possible with 3'-ester modified adapters but only with the shortest acceptor adapter.

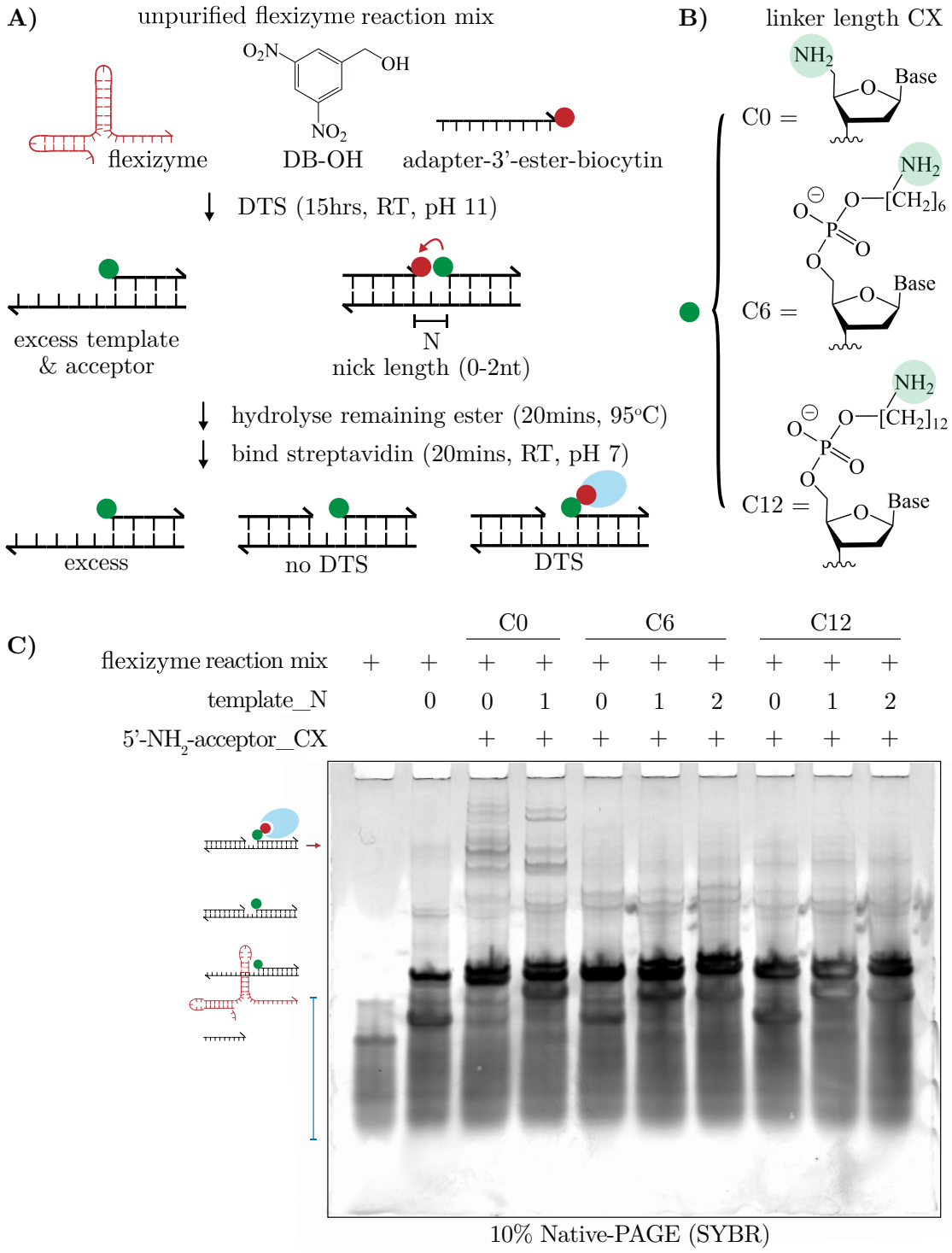


Figure 2.8: DTS with 3'-ester-modified adapters. A) The acceptor adapter and template were added to the flexizyme reaction mix. After DTS the reaction was heated to hydrolyse any remaining donor adapter, then incubated with streptavidin. Various acceptor and template combinations were tested in parallel, varying the relative position of the amine by linker length (CX) and nick length (N). Method in Appendix 7.1.10. **B)** Variations of linker length (CX) in the acceptor adapter. **C)** PAGE of DTS reaction stained with DNA-binding dye SYBR Gold. A long smear low in the gel is due to hydrolysis of the RNA flexizyme (blue band). Multiple high bands are observed with linker C0 (red arrow), indicating that some is DNA bound to streptavidin, which is only possible if DTS occurred in this sample. Method in Appendix 7.1.4, 7.1.7 and 7.1.10.

To verify whether DTS was occurring in this architecture, I performed LC-MS of a DTS sample under the supervision of a collaborator, Samuel Núñez-Pertíñez (O'Reilly Group, University of Birmingham). In this sample I tested DTS in the over-the-nick architecture with a 0nt gap between modifications, using the shortest possible linker (C0-NH_2). As before, components of the DTS reaction were added to the unpurified flexizyme reaction. Denaturing UPLC was used to separate the over-the-nick DTS complex into the component strands (using an acetonitrile mobile phase with a column temperature of 70°C). Four partially overlapping peaks were observed by absorbance at 260nm (Figure 2.9A). The deconvoluted mass of each peak corresponded to one of the expected masses of the DTS components (Figure 2.9B). As before, a mass corresponding to biocytin-sulfoxide rather than biocytin was observed (Figure 2.9C). Unfortunately, the overlapping peaks observed during LC-MS prevent quantification of DTS yield in this architecture. However, this experiment is evidence that DTS is possible with the new 3'-ester labile linkers.

2.2.3 DTS produces enzyme substrates

Using enzymes to coordinate transfer reactions and translocations relies on DTS producing a viable enzyme substrate. In this implementation, the DTS reaction must cleave the 3'-ester to produce a 3'-hydroxyl that an enzyme can recognise. To test this I performed the DTS reaction as before in the over-the-nick architecture with a 0nt gap between modifications, the shortest possible linker (C0-NH_2) transferring biocytin between strands (Figure 2.10A). As before, components of the DTS reaction were added to the unpurified flexizyme reaction. The strand-displacing Bst polymerase 3.0 from *Bacillus stearothermophilus* was added after the DTS reaction. Strand-displacing polymerases have the capacity to recognise a nick in a double helix and remove downstream nucleic acid during synthesis from the 3'-hydroxyl. Bst polymerase 3.0 is genetically engineered so that it has no nuclease or proof-reading activity. In this experiment I tested polymerase extension from the cleaved 3'-ester of the donor adapter, which would displace the acceptor adapter from the DTS complex. To the polymerase reaction I added

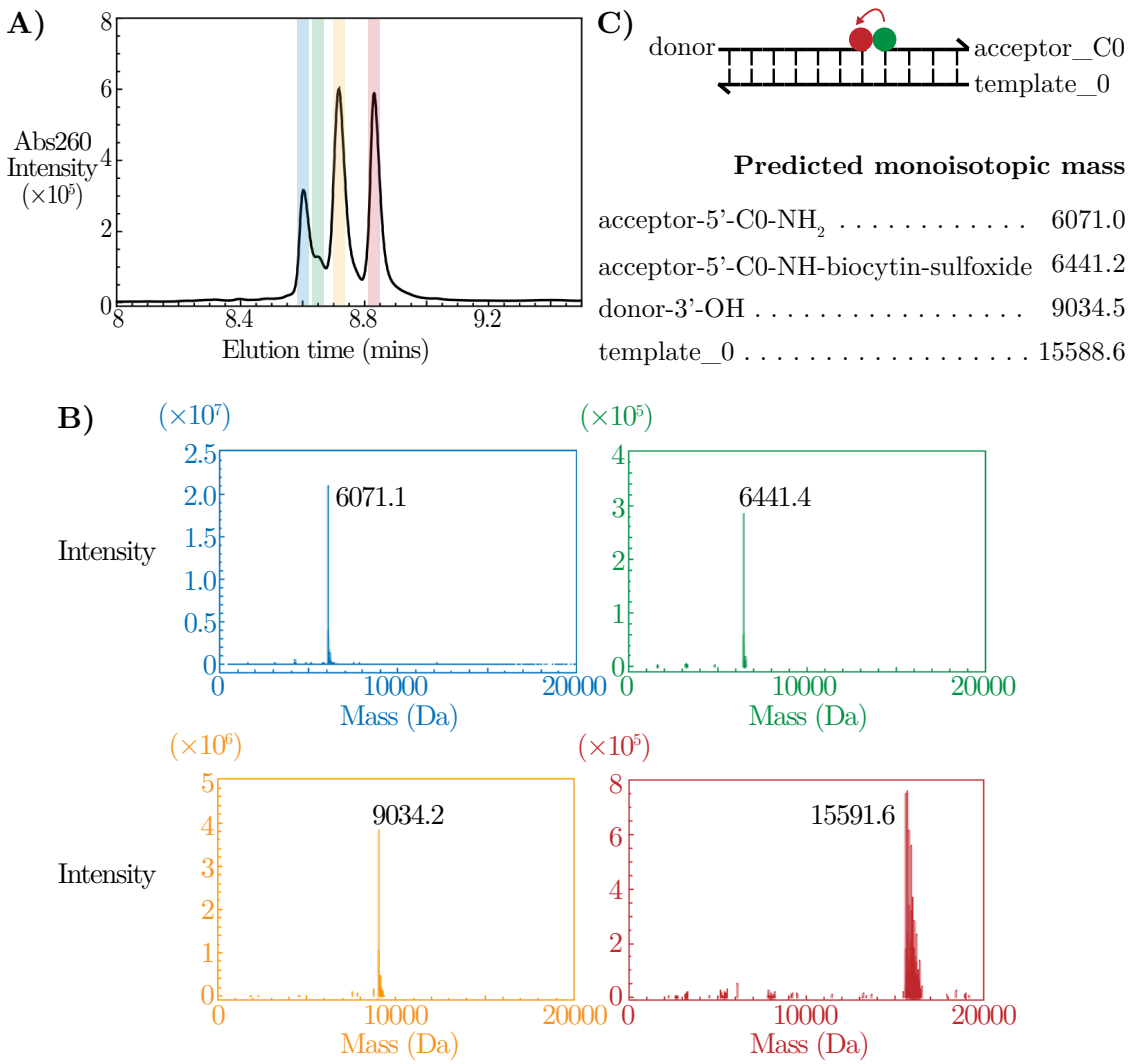


Figure 2.9: LC-MS of DTS products. A) Elution time of DTS products separated by denaturing UPLC. Method in Appendix 7.1.9. **B)** Deconvoluted mass spectra for peaks highlighted in A). **C)** DTS architecture and predicted masses for each strand and modification.

a strand that is complementary to the acceptor adapter and bears a fluorescent FAM modification, called a *detector strand*. If DTS produced a viable enzyme substrate, this would result in a dual-labelled dsDNA between the acceptor adapter and detector bearing both biotin and FAM modifications. If DTS did not occur, then only FAM-labelled dsDNA would be produced. If DTS occurred but did not produce a 3'-hydroxyl, then no polymerase extension would occur and only FAM-labelled ssDNA detector would be observed. The experiment was observed on native PAGE by incubating with streptavidin and looking for a slowly-migrating

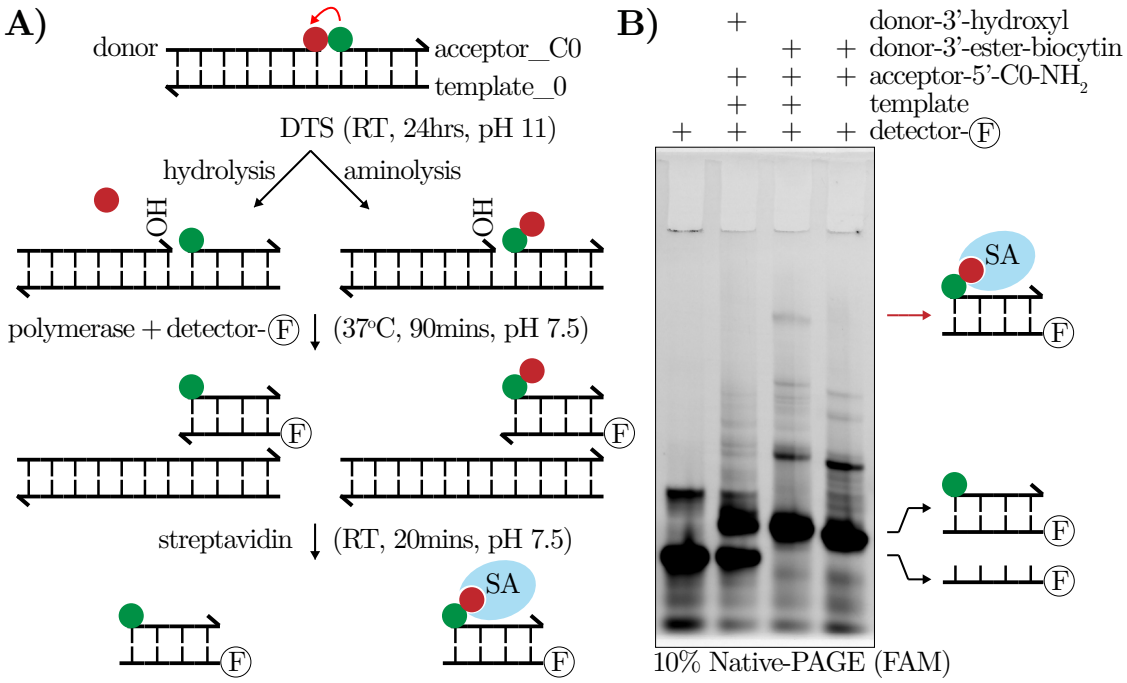


Figure 2.10: Polymerase activity on DTS products. A) During the DTS reaction, donor adapters bearing 3'-ester-biocytin (red) were either hydrolysed or aminolysed. Aminolysis transfers biocytin to the acceptor adapter (5'-C0-NH_2, green). After DTS a strand-displacing polymerase (Bst) and fluorescent detector are added. The detector is complementary to the acceptor adapter and bears a FAM modification (Ⓕ = 6-FAM). If Bst polymerase has activity on the DTS product, then a dual-labelled duplex bearing both biotin and FAM will be produced. **B)** Streptavidin-binding assay on native PAGE (FAM channel). A slowly migrating band corresponding to the dual-labelled detector and acceptor adapter duplex is observed (red arrow). Method in Appendix 7.1.4, 7.1.7, 7.1.10, 7.1.12.

fluorescent complex of streptavidin with the dual-labelled duplex.

Unfortunately, the detector had errors during solid-phase synthesis. On native PAGE this can be seen as truncations and dimers when the detector strand is run by itself (Figure 2.10B). This complicated the PAGE gel, though the overall result remained clear. If an unmodified donor adapter with a 3'-hydroxyl was used, then only a mono-labelled duplex was produced and no slowly migrating complex was observed. In contrast, if 3'-ester-biocytin modified donor adapters were used, then a dual-labelled species with slow mobility was observed. Finally, in the absence of template, no background reaction between donor and acceptor adapters was seen. This indicates that DTS can produce a 3'-hydroxyl that is a substrate for a DNA-acting enzyme.

2.2.4 Blocking enzyme activity before DTS

In order for an enzyme to coordinate transfer and translocation reactions, it must be possible for the DTS adapter to block enzyme activity until DTS. In this way the DTS reaction would trigger enzyme activity, which would in turn create conformational changes for the next DTS reaction. In this implementation, polymerase activity must be blocked by the 3'-ester-biocytin up until DTS produces the 3'-hydroxyl. I tested this by buffer-exchanging the unpurified flexizyme reaction mix to remove DMSO. I then assayed Bst polymerase activity on the flexizyme-adapter complex, as Bst should be able to use the 10bp complementary sequence between the adapter and flexizyme to synthesise a double-stranded DNA-RNA hybrid (Figure 2.11A). If the 3'-ester-biocytin blocked polymerase activity, this hybrid would not form.

I ran the samples on acid PAGE after the polymerase assay. As expected, the flexizyme showed some hydrolysis after incubating at 37°C in the presence of magnesium ions from the polymerase reaction buffer (Figure 2.11B). A positive control showed that in the absence of any 3' modification, Bst produced the expected RNA-DNA hybrid. The addition of Bst polymerase resulted in the disappearance of the band for 3'-ester-biocytin modified adapters, and the formation of the RNA-DNA hybrid.

This result may be explained by one or both of two hypotheses: 1) the polymerase can cleave 3'-esters, or, 2) a reagent in the Bst storage or reaction buffer cleaves the 3'-ester-biocytin, resulting in a 3'-hydroxyl upon which Bst acts. The first hypothesis is unlikely, as the Large Fragment of Bst polymerase 3.0 used here lacks the 3'→5' nuclease activity of the native, proof-reading Bst polymerase. The second hypothesis is highly likely, as the Bst reaction buffer contains multiple components that could cleave the 3'-ester. For example, Tris (tris(hydroxymethyl)aminomethane) with primary amines could attack the 3'-ester, magnesium ions could catalyse cleavage of the phosphate ester backbone, while a simple increase in pH could also increase ester hydrolysis. To test this hypothesis, I repeated the experiment with a range of individual components from the Bst storage and reaction buffers in the absence of

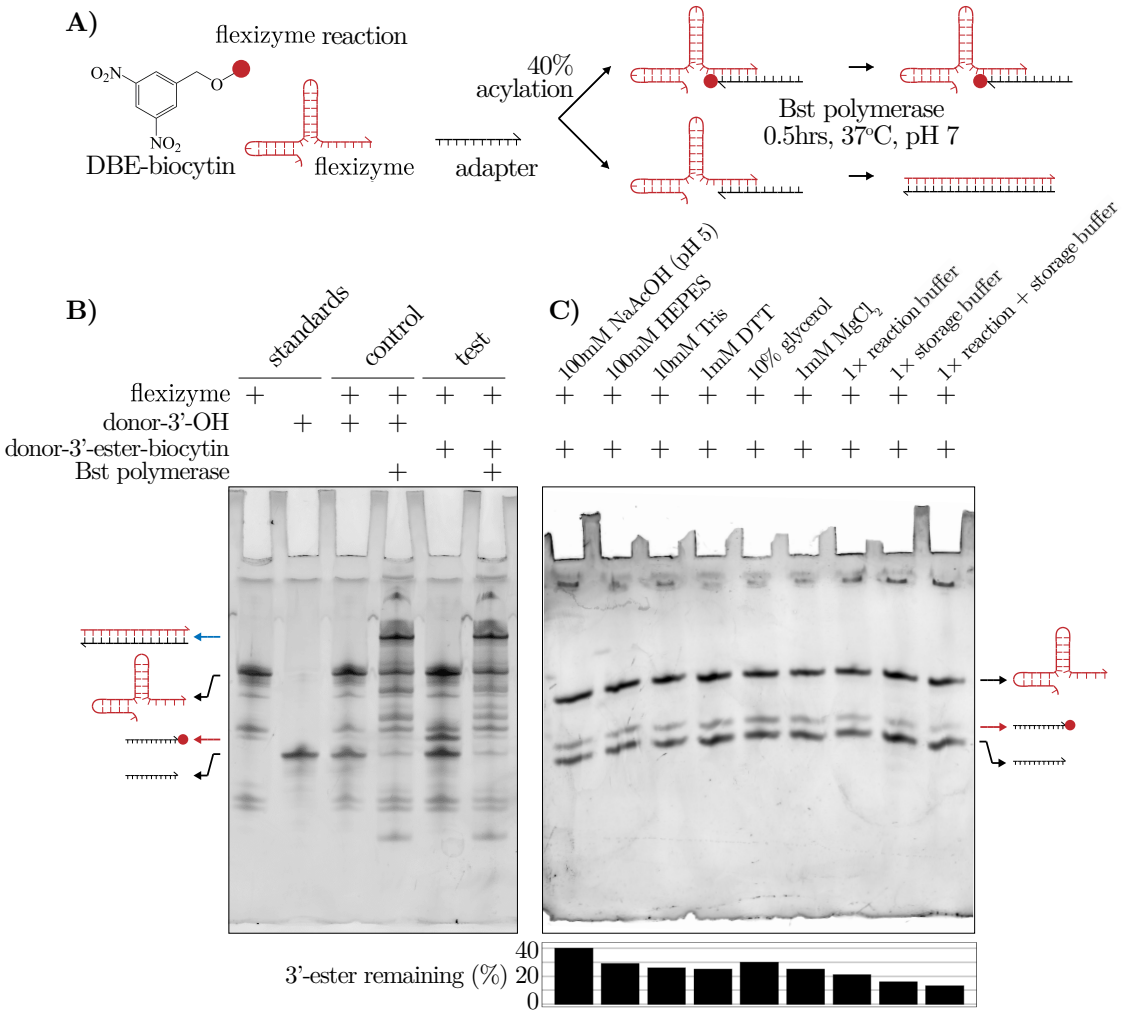


Figure 2.11: Blocking polymerase activity with 3'-ester modifications. A) The unpurified flexizyme reaction mix tested the activity of Bst strand-displacing polymerase. The 3'-ester modification blocks polymerase activity. Method in Appendix 7.1.12. **B)** 25% acid PAGE assayed polymerase extension. A control showed polymerase extension with unmodified adapter forming a larger, double-stranded RNA-DNA hybrid (blue arrow). The addition of polymerase to the reaction resulted in the disappearance of the band for 3'-ester-biocytin modified adapters (red arrow), and the formation of the RNA-DNA hybrid. Method in Appendix 7.1.5. **C)** The unpurified flexizyme reaction mix was incubated without Bst polymerase before 25% acid PAGE. The same conditions as the polymerase assay were used with various reagents from the reaction and storage buffers of Bst polymerase. Unless otherwise stated, all reactions are at pH 7. The proportion of 3'-ester adapter (red arrow) in each lane was estimated by gel densitometry. Method in Appendix 7.1.8.

Bst polymerase (Figure 2.11C). The amount of 3'-ester modified adapter remaining in each case was estimated using gel densitometry. In the absence of polymerase, these components had a cumulative effect causing hydrolysis of >70% of the 3'-ester modified adapters. In summary, this experiment indicates that this mechanism of feedback will fail when adapters degrade via side reactions with solvent.

2.3 DISCUSSION

This chapter aimed to create new designs for an artificial programmable polymerase incorporating feedback between templated transfer reactions and translocation. Here I have redesigned the adapters used in conventional DTS in such a way that DTS could trigger enzymatic activity, which in turn could be used to drive a DNA motor. The results demonstrate the synthesis of the new adapters and their capacity for DTS.

DTS was limited to a very short linker on the acceptor adapter. By comparison with the DNA double helix, one sees that the tetrahedral transition state of this acyl transfer reaction would have a similar structure to the phosphate backbone connecting adjacent nucleotides (Figure 2.12). In this sense, the redesigned architecture is a significant step away from the simple colocalisation of reactive groups on flexible tethers towards the creation of a programmable reaction pocket. With this DNA architecture and linker design, the reactive groups are in such close proximity (<0.4nm) that I estimate C_{eff} ~15M (Appendix 7.3.1). This estimate of C_{eff} is based on distances measured from crystal structures of DNA, and does not account for flexibility around the nick nor for alterations to the DNA structure caused by chemical modifications. Comparison of M_{eff} with other DTS architectures would require quantifying the rates of intra- and inter-molecular transfer reactions.

However, proximity may not be the only difference between the three acceptor adapter linkers (C0, C6, C12). There is a negatively charged phosphate group within the C6 and C12 linkers introduced during solid-phase synthesis (see Figure 2.8B). It has been shown previously that DTS linkers containing a histidine increased hydrolysis introduced during DTS reactions [133]. Perhaps the negatively charged

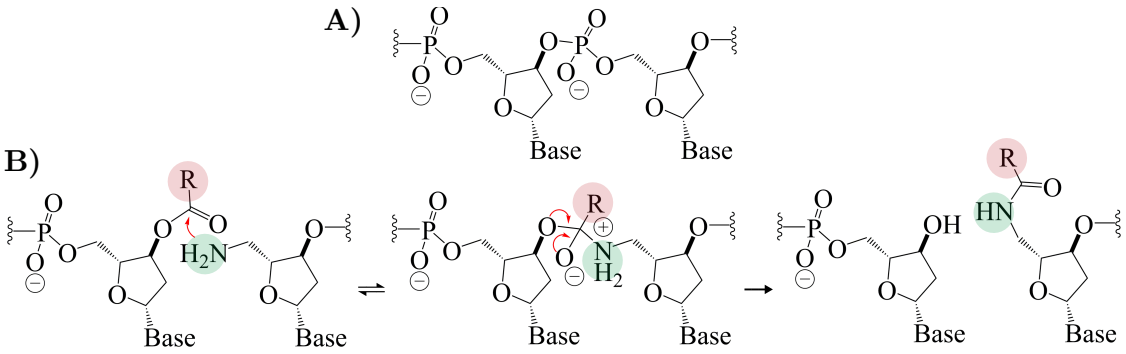


Figure 2.12: DTS transition state. A) Conventional dsDNA phosphate backbone. **B)** Transition state for acyl transfer with 3'-ester modified donor adapters and 5'-C0-NH_2 acceptor adapters in an over-the-nick DTS architecture.

ion directly repels a nucleophile, or indirectly attracts positively charged solvent (ie, $[H^+]$ or $[H_3O^+]$), creating a local environment of low pH that protonates the amine and decreases the rate of acyl transfer reactions. Such effects on local pH have been proposed to explain changes in the pH optima of enzymes conjugated to negatively and positively charged polymers [155, 156] and the fluorescence profile of pH-sensitive dyes on charged surfaces [157]. An ideal system to further explore this hypothesis would vary linker length and charge independently - this is absent from the DTS literature.

DTS with 3'-ester modifications appeared sensitive to the length of linker bearing the primary amine active site, which will complicate multistep polymer synthesis. During multistep synthesis, bifunctional monomers bear active sites for subsequent transfer reactions. The active site becomes more and more distant from the reaction environment as each monomer is added. This change in the reaction environment may make multistep DTS impossible with the 3'-ester adapters without further accommodation in future designs.

Another limitation of the current design is that adapter degradation has the same outcome as DTS. Both aminolysis and hydrolysis of 3'-esters produce a 3'-hydroxyl and thus adapter degradation would also trigger DNA motor progression. This might be resolved by using an alternative linker chemistry in which degradation by solvent has a different outcome to transfer, or by decreasing the rate of adapter degradation as explored in Chapter 5. Finally, the strategy of using enzymes

to coordinate transfer reactions and translocation limits the building material of a programmable polymerase to the substrates of natural enzymes (DNA, RNA, proteins, carbohydrates, etc.), so this approach may be inapplicable for entirely artificial supramolecular systems.

In addition to these limitations, there are also new opportunities. The successful use of unpurified flexizyme reaction mix in DTS is a step towards a scheme where adapter synthesis occurs simultaneously with DTS reactions. Flexizymes might restore degraded adapters like tRNA synthetases restoring spent tRNAs *in vivo* during protein synthesis or nucleic acid metabolism supplying cells with fresh rNTPs and dNTPs during nucleic acid synthesis. Interestingly, it is possible to use unique sequences in the 10bp duplex between flexizymes and the adapters that they charge with 3'-ester modifications, and in this way multiple adapters can be acylated simultaneously in one-pot [151]. At present, this strategy is impossible due to the differing pH requirements of DTS (pH 10-11) and flexizyme (pH 7-8). As the RNA flexizyme would hydrolyse over time at high pH, it seems more realistic that this would be achieved with changes to the DTS reaction (for example, if catalysis enabled transfer reactions at lower pH). Furthermore, in the current system donor adapters would hybridise preferentially to the DTS complex rather than the flexizyme, so an additional process is needed to achieve flexizyme turnover.

Finally, while it is interesting in itself to explore new DTS chemistries, the usefulness of this design depends on whether feedback can be implemented in an enzyme-driven DNA motor.

3

FEEDBACK II – MOTORS

Artificial programmable polymerases combine chemomechanical cycles of motion and polymer synthesis. Introducing feedback between these two processes will result in polymers with fewer skips and truncations. This chapter builds on the transfer chemistry explored in Chapter 2 by developing the autonomous DNA nanomachinery to implement feedback. In essence, DNA templated synthesis (DTS) cleaves DNA adapters to produce substrates for a strand-displacing polymerase. The polymerase in turn uses a small molecule fuel (dNTPs) to create toeholds for sequential strand translocation and transfer reactions. Here I test the design concept using fluorometry to track the movement of a fluorescently modified DNA strand, called a *cargo*, between two positions on a DNA track. Key contributions include the introduction of polymerase-driven exchange of DNA strands in a DNA nanostructure, driven by chemical fuel that remains stable for days without loss of function. I begin by introducing autonomous DNA nanomachines in more detail.

3.1 INTRODUCTION

3.1.1 DNA nanomachines

Rationally designed DNA nanostructures were first proposed by PhD candidate George Church to crystallographers including Nadrian Seeman [158], who subsequently pioneered the field by constructing DNA nanostructures using junction motifs [159, 160] (Figure 3.1A). This inspired the field of structural DNA nanotechnology [79, 161], where almost all DNA nanostructures involve variations of the junction motif. Watson-Crick base-pairing rules are used to encourage self-assembly of structures with the greatest number of base-pairs and lowest free energy, and to discourage competing structures. Moreover, complex designs can form multiple distinct structures. The equilibrium between structures can be controlled externally. The first DNA nanomachine with two structures was demonstrated in the form of tweezers that can be opened and closed by the addition of *fuel* DNA strands [91] (Figure 3.1C). This inspired the development of more complex DNA nanomachinery, included externally controlled DNA walkers that step between positions on a DNA track after the addition of DNA strands [162, 163].

An essential design motif underpinning dynamic DNA nanotechnology is toehold mediated strand displacement (TMSD) (Figure 3.1B), which is the exchange of two strands from a structure so that the one with more base-pairs is bound in the final state [164]. There are both thermodynamic and kinetic factors driving TMSD [97]. Thermodynamically, there are the additional base-pairs in the toehold domain made by the invader strand but not the incumbent strand. The invader and incumbent share a common sequence of nucleobases that are complementary to a single domain in the DNA complex. Kinetically, the toehold increases the invader's C_{eff} during competition to bind to the common domain. The invader first hybridises to the ssDNA toehold, which triggers the formation of a branch-migrating four-armed junction including two dsDNA duplexes and two ssDNA tails. Branch formation is slow relative to hybridisation and branch migration as it must wait for fraying (transient dissociation) of the end of the incumbent strand, and there is a chance

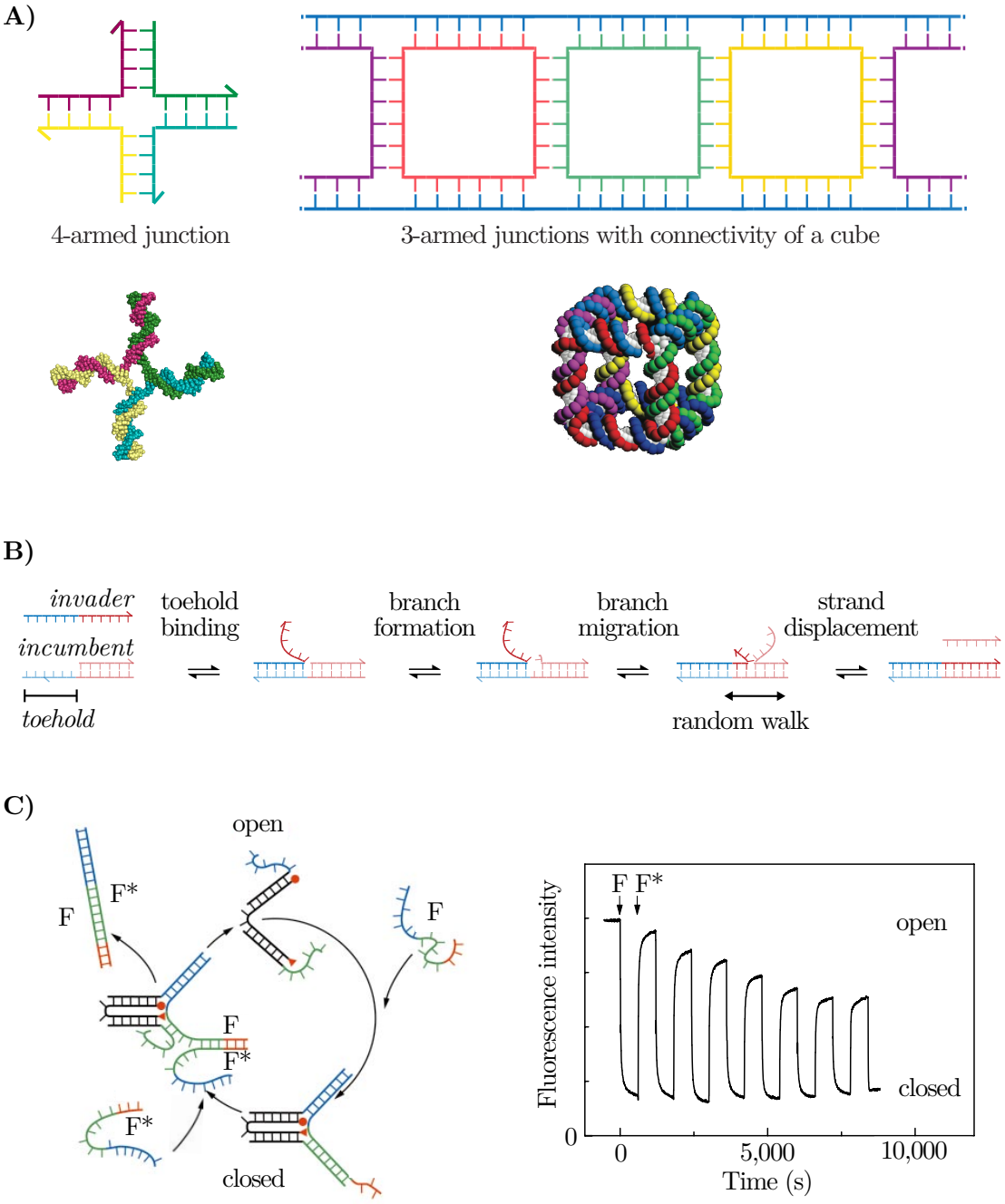


Figure 3.1: **A)** The junction motif is used to design DNA nanostructures. Holliday junction from PDB:2CRX [118] and DNA cube from [160] **B)** Toehold mediated strand displacement (TMSD) exchanges strands in a DNA complex. An *invader* strand hybridises via a toehold domain (blue) and outcompetes an *incumbent* strand for hybridisation to a common domain (red). **C)** DNA nanomachines can use TMSD to move between multiple stable structures. In the first example of DNA tweezers, a fuel strand F closes the structure and brings two fluorophores (TET and TAMRA, indicating by the triangle and circle) into close proximity for FRET, causing a decrease in fluorescence intensity. The cycle of opening and closing is fuelled by the hybridisation of F and F* through TMSD with the toehold (orange). Figure C) adapted with permission from [91] © 2011 Elsevier.

that short toeholds will dissociate before branch formation. The junction undergoes a random walk forwards and backwards along the common domain. However, if the junction ever returns to the beginning the invader can rapidly try again while if the junction reaches the end the incumbent strand is displaced and diffuses into solution. In principle an incumbent strand may return from solution and temporarily reform the branch migrating complex in a process called *blunt end strand exchange*, though blunt end exchange (~1 $M^{-1}s^{-1}$) is much slower than TMSD (~10^6 $M^{-1}s^{-1}$) so in practice the reverse process is negligible [96]. Many variations and applications of TMSD have been explored, introducing mutations and flexible linkers to alter TMSD thermodynamics and kinetics [165, 166], substituting DNA domains to trigger strand displacement with small molecules, pH, light, and more [167–170].

While this has been an exciting few decades for dynamic DNA nanotechnology, there is a large amount of work required to produce anything resembling biological nanomachines. For example, an externally controlled two-legged DNA walker requires the laborious addition and removal of strands to move at <1nm.s^{-1} [171], while the kinesin protein it is inspired by moves autonomously at >600nm.s^{-1} [172]. To move towards a DNA-based artificial programmable polymerase we require autonomous nanomachinery, which would eventually enable applications like *in situ* polymer synthesis and self-replication.

In this context *autonomous* means there is no external intervention while the machine operates. Without external intervention most of the DNA nanomachines described above would idle at a single structure or a reversible equilibrium between structures. A new energy landscape is created with each external intervention, driving the conformational change between designed structures. In contrast, an autonomous machine traverses a path between multiple local minima on a single energy landscape. Directionality can be introduced by catalysing the degradation of a thermodynamically unstable but kinetically trapped *fuel*, and coupling this source of free energy to molecular motion.

An example of chemical fuel is the making or breaking of covalent bonds. Early demonstrations of this approach in autonomous DNA walkers involved the use of

ligases and nucleases. As soon as a DNA cargo stepped to a leg on a DNA track it created a nuclease recognition site on the leg. Enzymatic cleavage of the DNA leg would in turn would trigger the next stepping event, and the creation of a new nuclease recognition site. In these cases the fuel is the cleavage of phosphate bonds (either in a DNA backbone [92], RNA backbone [93], or ATP [94]).

The first autonomous DNA motor was described in 2004 [94], and involved the careful design of sites for restriction enzyme cleavage and ligation of double-stranded DNA so that 6 nucleotides move along a track (Figure 3.2A). These 6 nucleotides can create a sticky-end overhang between two adjacent legs on the track, which is ligated by an ATP-consuming ligase enzyme. The system can idle if a sticky end in the backwards direction is hybridised, ligated, then cleaved again by the restriction enzyme. However, due to the arrangement of restriction sites only forwards steps can be made successfully, biasing the walker in the forwards direction. A simpler system was developed in 2005 [92], requiring only a single nickase enzyme that drives a cargo strand along a track (Figure 3.2B).

Finally, the nickase functionality can be incorporated within the cargo itself. A DNAzyme was evolved that hybridises to and cleaves ribose backbone linkages (RNA) within a complementary strand [173]. To create a DNA walker, Tian et al. arrayed a series of legs on a track so that a cargo DNAzyme would walk from leg to leg across the track [93]. The DNAzyme cleaved each leg into two domains − a short (<7bp) domain that dissociated and diffused into the bulk and a long (>15bp) domain that ensured the DNAzyme remained on the track (Figure 3.2C). Loss of the short domain enabled TMSD to subsequent legs on the DNA track, driving motion between legs as each is cleaved in turn. This strategy was employed in the first autonomous DNA-based artificial programmable polymerase [98] (discussed previously in the introduction, see Figure 1.6).

Alternatively, the hybridisation of DNA itself can provide the energy to drive DNA motors. Some DNA structures called *hairpins* have two self-complementary domains that hybridise to form an internal loop. It was discovered that hairpins have interesting hybridisation kinetics, where internal toeholds located within the

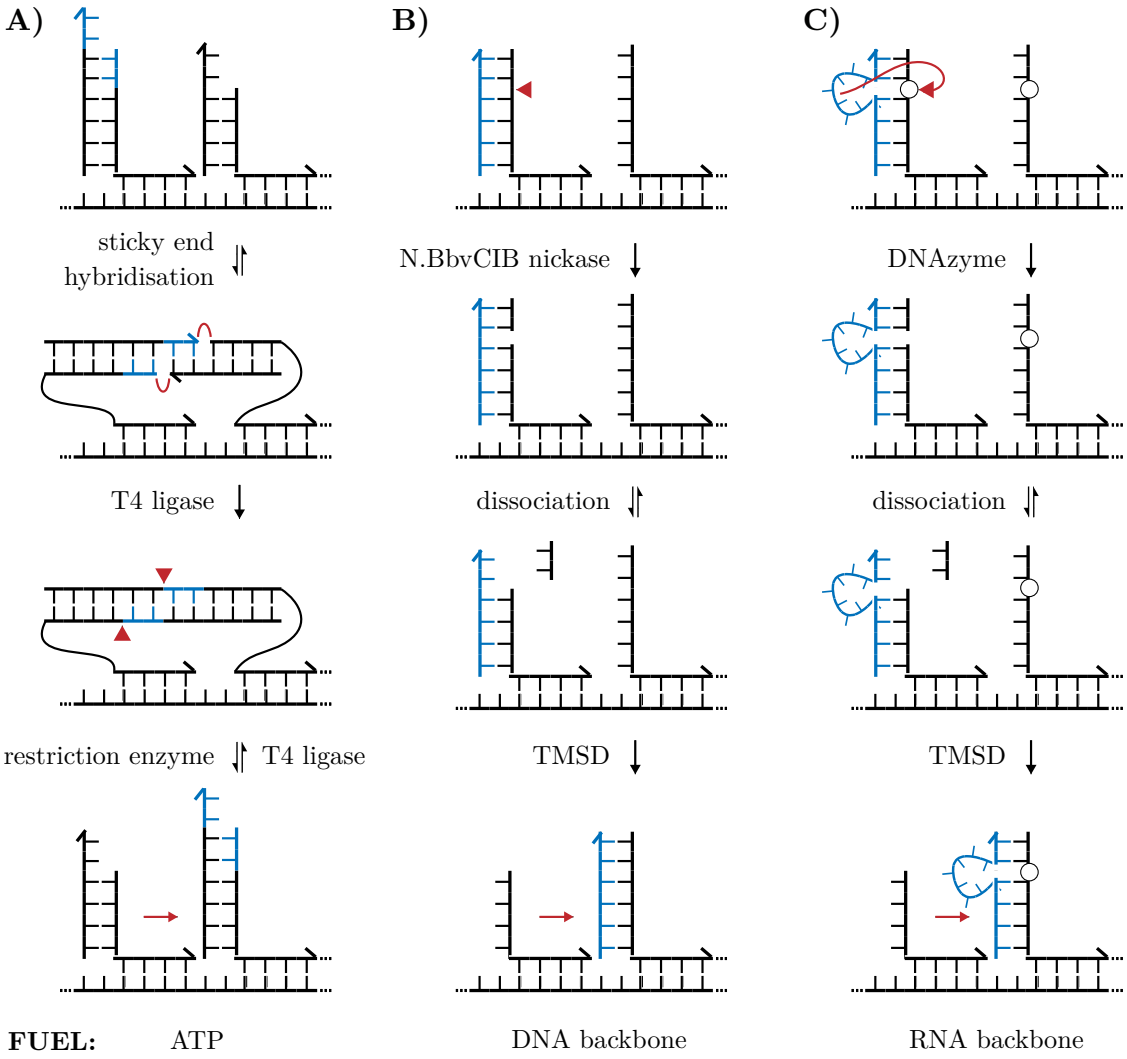


Figure 3.2: Autonomous DNA motors. A single step (left to right) of a multistep cascade is shown for three different autonomous DNA motors. On a periodic track sequential translocation steps are possible. **A)** A 6 nucleotide sticky end cargo (blue) is transferred between legs on a track [94]. T4 ligase consumes ATP when conjugating sticky ends. This forms a duplex containing a recognition site for a restriction enzyme (either PflM I and BstAP I). Cleavage of the recognition site transfers the cargo to the next leg. An idling step is possible if the sticky end reforms and is ligated and cut repeatedly. Backwards steps are impossible due to the location of the restriction enzyme recognition site. **B)** A 26 nucleotide cargo (blue) was transferred between legs on a track [92]. A nicking enzyme (N.BbvCIB) recognises a sequence formed within the cargo and leg duplex, producing a short (8nt) fragment with weak affinity for the leg. If the fragment dissociates, then the cargo will transfer to subsequent legs as more base-pairs can be made. Backwards steps are improbable, but if they occur then the cargo could idle in a random walk backward and forward between previously cut legs until reaching the uncut leg. **C)** A 38 nucleotide DNAzyme (blue) was transferred between legs on a track [93]. Cleavage of ribose backbone linkages (RNA, black circle) by the catalytic loop of the DNAzyme produces a short (7nt) fragment with weak affinity for the leg.

loop of the hairpin undergo TMSD much more slowly than external toeholds [174] (Figure 3.3A). The slow kinetics of hybridisation to hairpin loops is attributed to both geometric and topological factors [175]. The stem of the hairpin is a duplex ~1.7nm wide, which means invading strands that are >6bp (~1.8nm) require the stem to denature to fully hybridise with the internal toehold sequence. Furthermore, invading strands >11bp (the helical period of DNA) must thread through the loop of the hairpin during hybridisation. The discovery that TMSD with external toeholds rapidly opens hairpins enabled the design of a hybridisation chain reaction (HCR) [90]. Complementary, metastable hairpin structures slowly interact over days unless the presence of a truncated hairpin or *initiator* triggers the rapid opening of all hairpins using their external toeholds. Once the initiator opens a hairpin by rapid TMSD with an external toehold, the nucleobases in the loop are free to hybridise. If they hybridise to the external toehold of another hairpin, a cascade of hairpins may open one after the other in a hybridisation chain reaction.

This in turn inspired a DNA polymerisation motor where a *cargo* hybridised to the initiator moves further away from the initiator with each additional hairpin participating in the chain reaction [176]. More recently, Meng et al. [136] achieved autonomous polymer synthesis using HCR to drive a programmed sequence of DTS reactions. Meng's system features a set of DNA hairpins that interact in a predetermined sequence to opens the loop of each hairpin one after the other (Figure 3.3B). In this case, the order of hairpin interactions was programmed by a subset of *information hairpins*. Each information hairpin hybridises to a chemically modified *donor hairpin* that transfers monomers to the HCR cargo, which is a DTS acceptor adaptor. In turn, the donor hairpin recruits another information hairpin encoding the next step of in the cycle of polymer synthesis (Figure 3.3C). Upon mixing a set of hairpins, polymers were synthesised with a distribution of lengths between 1-7 monomers and a mean of ~3.5 [136].

Interestingly, the length limitations in this design have the same origin as Liu's DNA walker [98] and Leigh's supramolecular rotaxane [70]: namely, the degradation of reactive components, the changing reaction environment at each synthesis step,

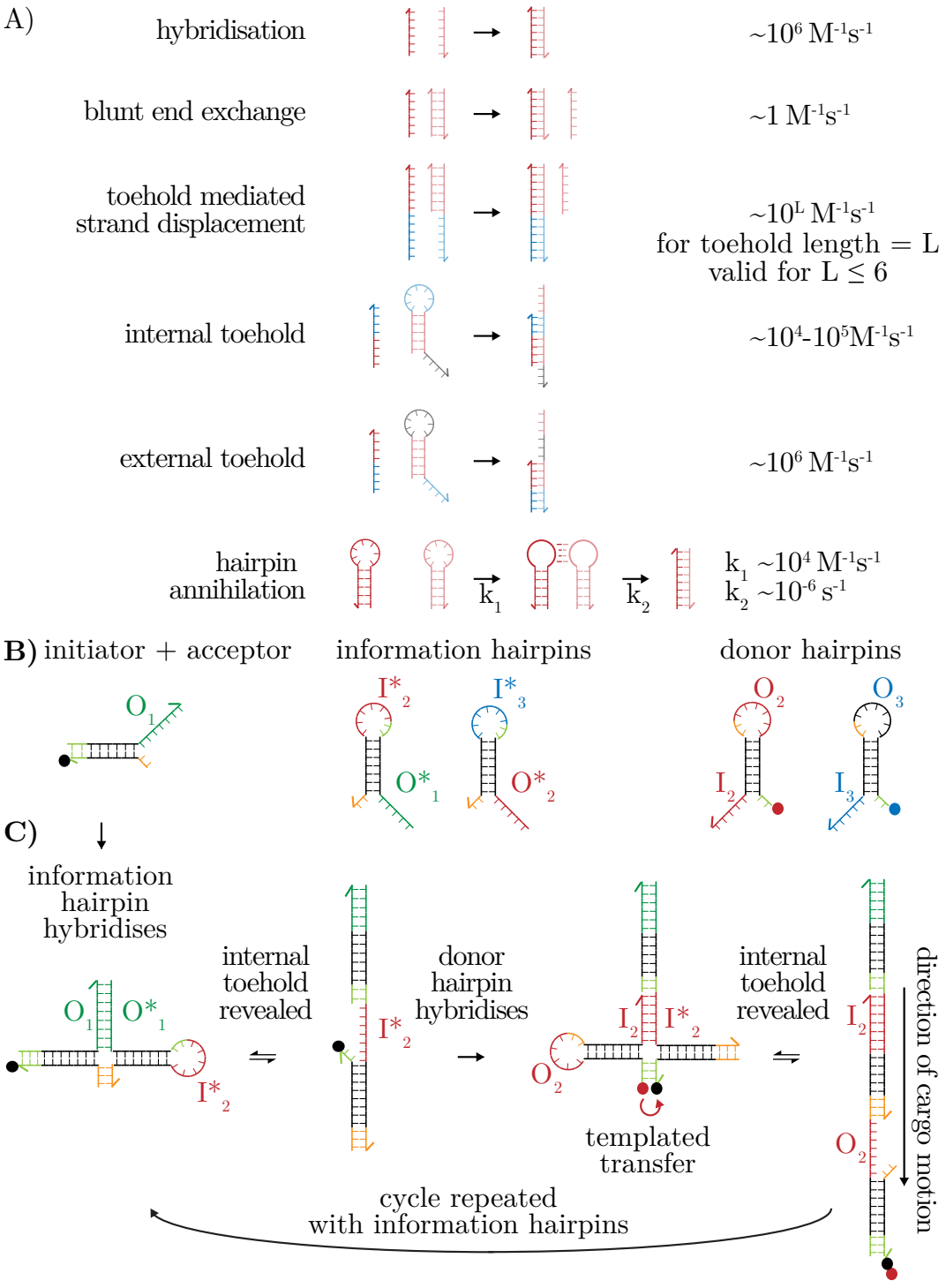


Figure 3.3: **A)** Rate constants for association of DNA strands in various architectures [175]. Kinetically trapped hairpins can be rapidly with external toeholds. This enables a hybridisation chain reaction as each opened hairpin triggers the opening of others. **B)** Components of a hybridisation chain reaction (HCR) that drives successive rounds of DNA templated synthesis (DTS) as in [136]. **C)** Hybridisation of an *information hairpin* with the initiator-acceptor complex opens the loop of the hairpin, triggering a cascade of hybridisation reactions that opens other hairpins. Every second hairpin is a *donor hairpin*, permitting successive DTS reactions adding monomers to the cargo's polymer. The sequence of DTS reactions is programmed by the information hairpins.

and so on. There are also additional issues complicating the synthesis of longer polymers and the potential applications in directed evolution. As DNA hairpins are kinetically trapped structures, they spontaneously unfold and interact given sufficient time − this becomes an issue for longer polymers with long incubation times. Furthermore, the sequence of the synthesised polymer was programmed by a pool of information hairpins rather than a single genetic polymer. In the case of a gene, the order of monomers in the polymer is determined by the order of codons in the gene, so a monomer can be encoded at a different position in the polymer by simply repeating the same codon. In contrast, as the interactions between DNA hairpins during HCR result from random collisions, each information hairpin must specify both the monomer's identity *and* its position in the polymer. Every time a monomer is repeated a new pair of unique information and donor hairpins are required. This is analogous to having a unique tRNA for every possible position of an amino acid in a protein. Synthesis is laborious and it is difficult to imagine how hairpin pairs could mutate during directed evolution. Finally, hairpins are not an ideal fuel for DNA nanomachines as truncated copies of hairpins produced during solid-phase synthesis can also initiate HCR, leading to leaking of fuel hairpins and unprogrammed chain reactions [177]. While cleavage of RNA provides a more stable fuel than DNA hairpins, RNA is also prone to hydrolysis in DTS conditions, resulting in polymers with skips or truncations as the DNA walker's legs are cleaved by water.

This section has introduced DNA nanomachines in sufficient detail to understand how an artificial programmable polymerase can be constructed using DNA as a building material. The next section explores how a DNA nanomachine could be coordinated with DNA templated chemistry so that transfer reactions trigger translocation events.

3.1.2 Design of an autonomous DTS-dependent DNA motor

In Chapter 2 I explored linker designs that produce an enzyme substrate after DTS − the 3'-hydroxyl. It is now necessary to design an autonomous DNA motor where the appearance of an enzyme substrate triggers translocation to the

next position for DTS. I envision a system where DTS occurs in an over-the-nick architecture (Figure 3.4A), where donor and acceptor adapters both hybridise to a third strand and DTS transfers a monomer across the gap between adapters. If a strand-displacing polymerase were added to this DTS reaction, then the polymerase would not have a substrate until DTS occurred. Strand-displacing polymerases have the capacity to recognise a nick in a double helix and remove downstream nucleic acid during synthesis from the 3'-hydroxyl. After DTS, the strand-displacing polymerase would act on the 3'-hydroxyl of the donor adapter produced by DTS, incorporating a set of dNTPs to synthesise a new DNA strand and displace the acceptor adapter from the DTS complex. The acceptor adapter and growing polymer would diffuse into bulk solution. This may be useful for a DNA nanomachine, however, in this particular application the loss of the acceptor adapter makes programmable polymer synthesis impossible. The polymer must remain with the DNA gene that programs its sequence. There must be a way of implementing feedback without loss of the synthesised polymer into solution.

A new design feature could be added to cause the polymerase to halt after adding only a few nucleotides rather than synthesising the entire sequence. This would prevent the acceptor adapter and the growing polymer diffusing into solution if a long sequence remained hybridised and was not displaced by the polymerase. However, before halting, the polymerase could create a short toehold in the acceptor adapter for subsequent TMSD reactions to drive an autonomous DNA motor. The polymerase might halt if it reached a nucleotide that was excluded from the reaction mix (for example, if the polymerase synthesised a sequence containing only dG, dC and dA before needing a dT and lacking a dTTP substrate). Alternatively, a dideoxynucleotide lacking a 3'-hydroxyl could be used to halt the polymerase. The DTS-triggered formation of a toehold would enable broad array of DNA nanomachines, but a simple example is given in Figure 3.4.

In this scheme, multiple legs would be positioned at intervals on a periodic DNA track. Each leg would carry a 3'-ester modified donor adapter, so that upon DTS the 3'-hydroxyl of the adapter would be revealed. In this way, DTS would

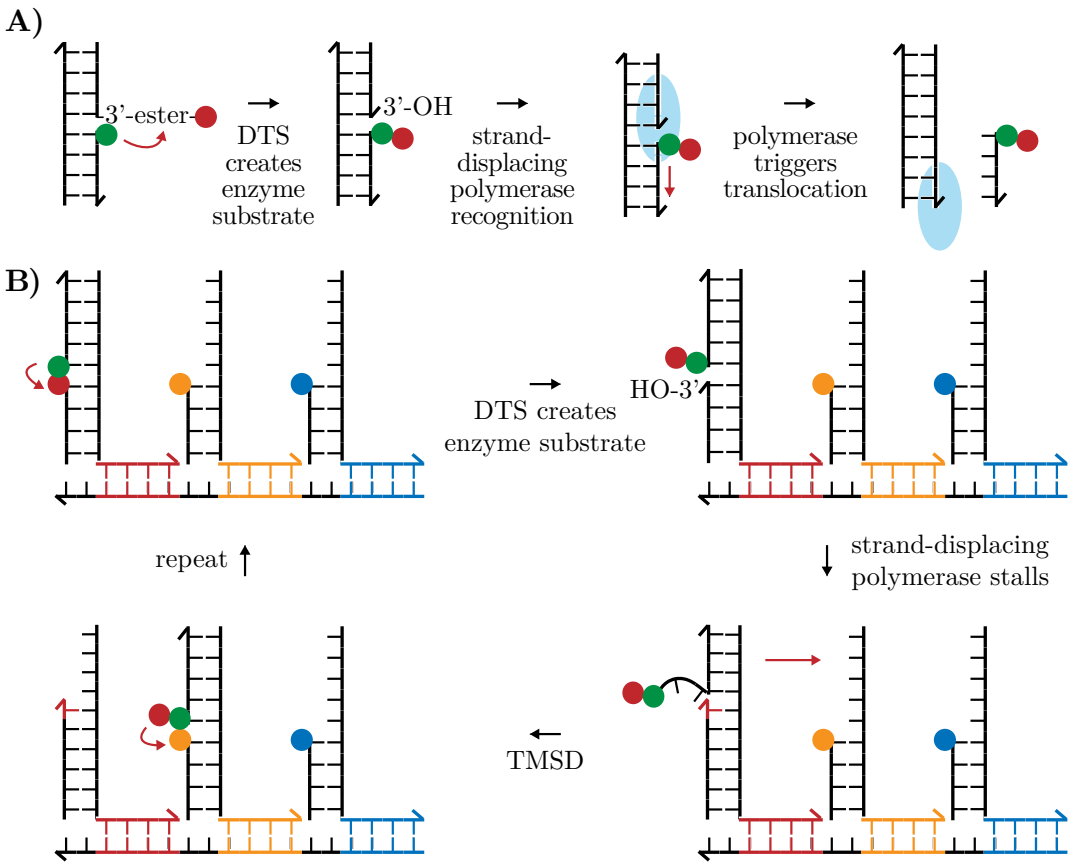


Figure 3.4: Design of an autonomous DTS-dependent DNA motor. A) DTS with a 3'-ester adapters in an over-the-nick architecture, demonstrating how DTS could trigger the exchange of DNA strands required to drive a DNA motor. **B)** A set of 3'-ester donor adapters are hybridised to the legs of a walker on a DNA program. After DTS, the strand-displacing polymerase stalls at a specific nucleotide (shown in red, either excluded from reaction mix or a dideoxynucleotide lacking 3'-hydroxyl), creating a toehold for the TMSD reactions that drive the autonomous DNA motor. The autonomous motor will run as fast as possible without causing skips, as DTS is required for the acceptor adapter to move between legs.

create a *primer* or substrate for a strand-displacing polymerase. Polymerase activity would be forced to halt at a specific position, either by excluding nucleotides to be added or using a dideoxynucleobase. This creates a toehold for TMSD that drives the cargo to the next leg for subsequent transfer and translocation reactions. The system is autonomous, yet the feedback between transfer and translocation means that it cannot skip past unreacted donor adapters.

In principle, this design has significant advantages over previous systems. Multiple dNTPs would be consumed per DTS reaction (~30 kcal/mol for a 6bp

toehold [178]), while this free energy is stored in a much more stable form than DNA hairpins or RNA backbones. Enzymatic catalysis can accelerate dNTP degradation by ~10^{10} [179], but importantly the consumption of fuel depends on DTS to produce an enzymatic substrate. However, there are also limitations to the design. Every 3'-hydroxyl that is not an intended polymerase substrate must be blocked with 3' modifications to prevent off-target polymerase activity destroying the nanostructure. Also, as in all DNA walkers, the cargo strand is complementary to every leg on the track, which means that if left for long enough then the spontaneous fraying of strands and background blunt-end transfer of cargo between legs may be possible.

In practice, it is necessary to search for a polymerase that will be active in DTS conditions, to test whether the strand-displacing polymerase can produce toeholds for TMSD, or whether the steric bulk of the polymerase on the DNA strand or off-target polymerase binding prevents TMSD. Finally, polymerase-TMSD will only be useful if it catalyses motion faster than the background exchange of strands.

3.2 RESULTS

3.2.1 Assembly of structure and transfer of cargo

I first aimed to demonstrate self-assembly of a simplified prototype of the desired motor. In this structure there were two legs (L1, L2) bearing primers resembling cleaved DTS donor adapters (P1, P2). A cargo strand (C) can hybridise to both legs L1 and L2 and move between positions (Figure 3.5A-B). As a control to confirm polymerase activity, an extended version of the primer (P1-ext) was used which transfer C from L1 to L2 by TMSD. A gene (G) was used to colocalise adjacent legs so that C made 'intramolecular' rather than 'intermolecular' steps.

Before using water-sensitive DTS adapters, I assayed the position of C using fluorophores and quenchers. The cargo (C-F) was chemically modified during synthesis with a fluorophore (3'-6-FAM, Ex:495nm, Em:520nm), while L1-Q contained a broad-spectrum quencher (5'-BHQ, A:534nm). In this way C-F fluorescence is quenched when hybridised to L1-Q. If C-F moved to L2, the fluorophore would be

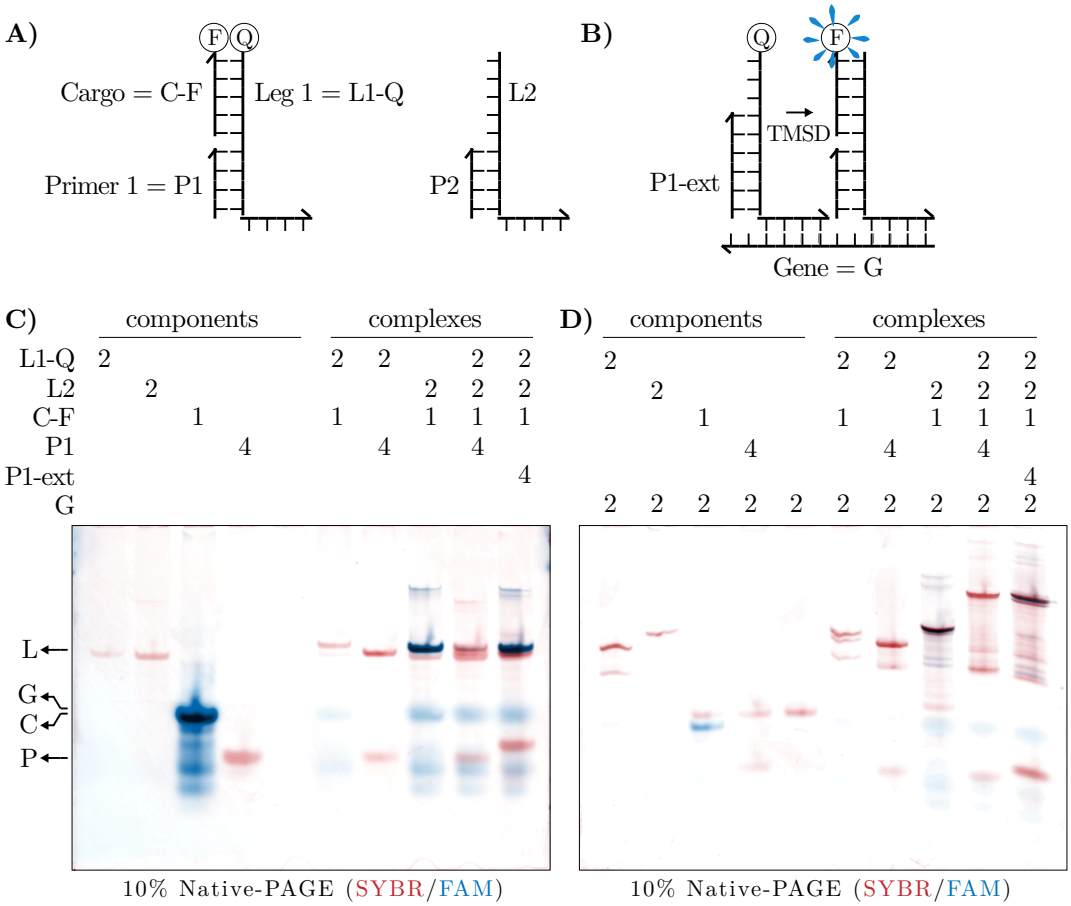


Figure 3.5: Assembly and TMSD. A) Nomenclature for components of the motor prototype. (F) = 6-FAM. (Q) = BHQ-1. **B)** The gene (G) colocalises legs. The extended primer, *P1-ext*, is a positive control for polymerase extension and drives TMSD of C-F to L2, causing increased fluorescence. **C)** PAGE demonstrating hybridisation of motor components in 10 minutes at RT. Gels were observed for FAM fluorescence (shown in blue), then stained with SYBR-gold (shown in maroon). L1-Q and C-F were added first to pre-hybridise the structure, then the remaining strands were added. [C] = 1µM, while other strands were added in the stoichiometric excess indicated. **D)** Colocalisation of motor components by G. Method in Appendix 7.1.4.

~25nm from the quencher and exceed the <10nm distance required for quenching [180]. The 3' end of all strands except P1 were blocked with modifications (either 3'-fluorophores or 3'-phosphates) that prevented off-target polymerase activity.

Self-assembly was visualised by native PAGE. When complementary strands hybridise the migration of DNA through the gel changes. Specifically, bands indicating ssDNA might disappear while more slowly migrating dsDNA appear (though secondary structure also alters migration distance). In this experiment C-F and L1-Q were pre-hybridised before subsequently adding the remaining strands.

On PAGE C-F ran slower when incubated with a 2× excess of L1-Q or L2, indicating successful formation of leg complexes (Figure 3.5C). Furthermore, fluorescence in the FAM channel was reduced when C-F was hybridised to L1-Q (the complex of C-F and L1-Q is visible when stained with DNA-binding dye SYBR Gold, but not by FAM fluorescence). The right-most two lanes of the gel compared the structure with either P1 or P1-ext, testing TMSD of cargo from L1-Q to L2. In this case there was no FAM fluorescence with P1, indicating the cargo bound to L1-Q. There was high FAM intensity with P1-ext, indicating that C-F transferred to L2. When all the samples are incubated with the gene (G) that colocalises legs, I observed similar shifts in band migration corresponding to track-leg complexes (Figure 3.5D)

In summary, this experiment confirmed both self-assembly of a two-legged motor structure and TMSD of cargo between legs. Similar gels were run for subsequent experiments with similar results. TMSD of cargo between legs acted as a positive control to test polymerase-driven TMSD.

3.2.2 Comparing and halting strand-displacing polymerases

Strand-displacing polymerases were initially tested on a hairpin architecture. Hybridisation of a primer (P) to the toehold of a hairpin (H) provided a substrate for strand-displacing polymerases to open the hairpin (Figure 3.6A-B). If the first nucleotide of the loop of H was dA, and the complementary dTTPs were excluded from the polymerisation reaction, then the polymerase might stall after opening the hairpin (Figure 3.6A). Alternatively, if all dNTPs were included, the polymerase should produce the entire complementary strand forming dsDNA (Figure 3.6B). This difference should be observable by running the products on native PAGE and comparing their lengths.

This scheme was tested with two polymerases, bacteriophage Φ29 polymerase and Bst polymerase from *Bacillus stearothermophilus*. Φ29 produced only a small amount of dsDNA in a 1000 minute polymerase reaction, whereas Bst was mostly finished in <10 minutes and began synthesising off-target DNA after 1000 minutes (Figure 3.6C-D). For this reason Bst was used for subsequent polymerase reactions.

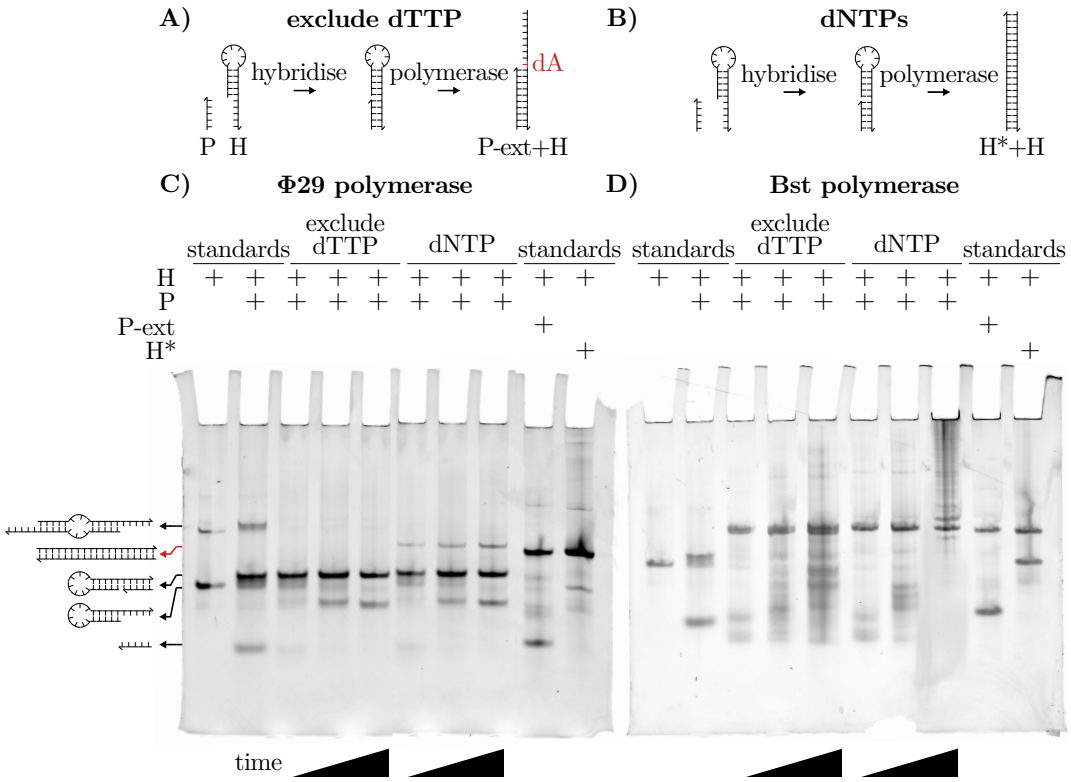


Figure 3.6: Comparing strand-displacing polymerases. **A)** A hairpin (H) hybridises with a primer (P), and P can be extended (P-ext) by a polymerase until a dT is needed. **B)** Alternatively, if all dNTPs are provided, P can be extended to form the full complement of the hairpin (H*). **C)** Φ29 polymerase was incubated with P and H for various times (10, 100, or 1000 minutes), with either dGTP, dCTP, dATP but not dTTP, or with all dNTPs. The full duplex (red arrow) is synthesised if all dNTPs are included. 10% native PAGE stained with SYBRGold. Method in Appendix 7.1.4 and 7.1.12. **D)** Bst polymerase in the same reaction. The full duplex (red arrow) is synthesised even if dTTP is excluded.

Two strategies were explored to halt Bst polymerase at a specific nucleotide to create the toehold for TMSD that would drive the DNA motor. The polymerase was initially tested on a hairpin architecture. If the first nucleotide of the loop of H was dA, and complementary dideoxynucleotides (ddTTPs) lacking a 3'-hydroxyl were included in the polymerase reaction, then the polymerase might be stalled after opening the hairpin (Figure 3.7A). Alternatively, if a run of 5 dAs was included and dTTPs were excluded, then the polymerase might also be halted (Figure 3.7B). By comparing both reactions on native PAGE (Figure 3.7C), it appeared that the first strategy of using dideoxynucleotides produced a more efficient halting and a cleaner set of DNA species.

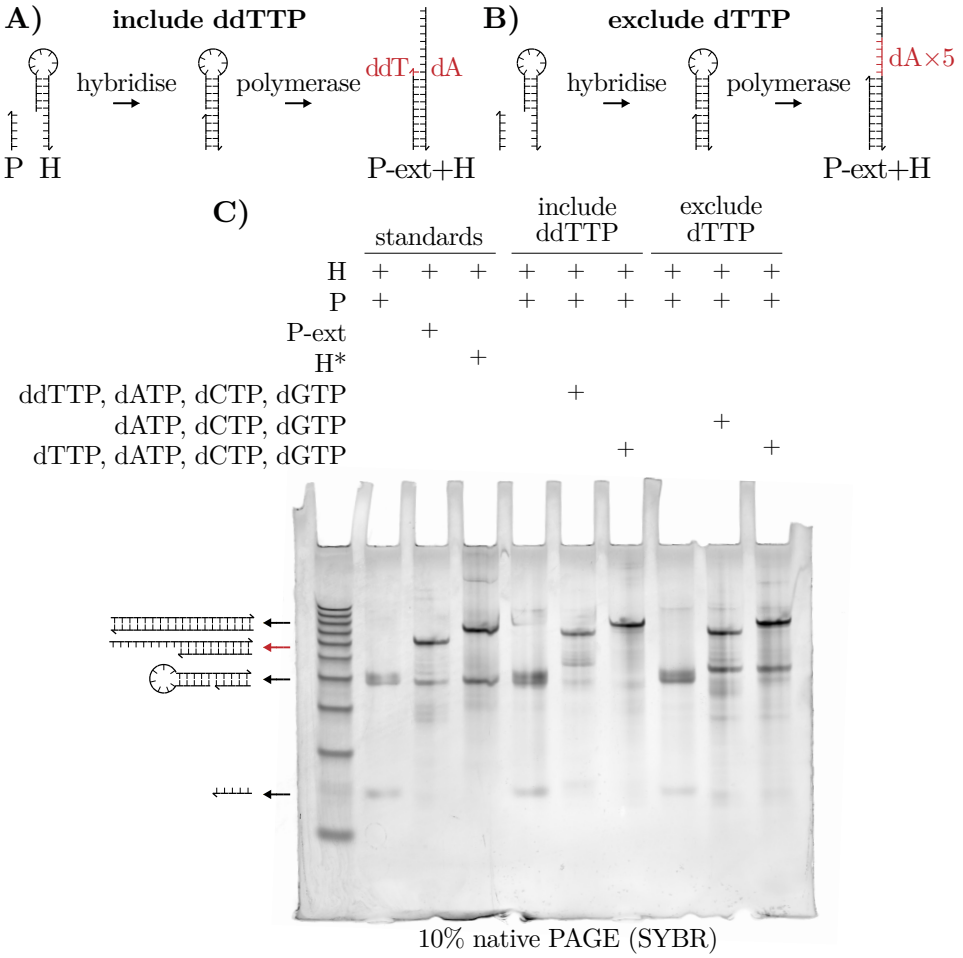


Figure 3.7: Halting a strand-displacing polymerase. A) A hairpin (H) hybridises with a primer (P), and P can be extended (P-ext) by a polymerase until a dT is needed. Halting is caused by incorporation of a ddTTP that lacks a 3'OH for further extension. **B)** Alternatively, if a run of 5 dAs is used in the loop of the hairpin, then the polymerase might be halted by excluding dTTPs from the reaction mix. **C)** Comparison of the two strategies to halt a polymerase. The left three lanes show standards where strands of the appropriate length are mixed and no polymerase reaction occurs. The middle three lanes show the strategy of using ddTTP to halt the polymerase at a specific position (red arrow), whereas the last three lanes exclude dTTP. Method in Appendix 7.1.4 and 7.1.12.

3.2.3 Polymerase-driven TMSD

Having established self-assembly, TMSD and polymerase halting, it was possible to track the position of C-F by fluorometry on a two-legged DNA walker. If the fluorescence from the cargo increased after the addition of Bst polymerase, this would indicate it had moved between legs. Importantly, a fluorescence timecourse enabled the comparison of the rate of polymerase-driven TMSD to the background

blunt end exchange of strands. After mixing the strands and incubating at 37°C, C-F had lower fluorescence when hybridised to L1-Q compared to L2 (Figure 3.8), as expected due to quenching. The signal decreased slightly over time, presumably due to fluorophore bleaching [181]. As such, in subsequent plots the signal was normalised between the two controls: 1) positive control with high fluorescence (L2 + C-F) and 2) negative control with low fluorescence (L1-Q + C-F). This normalisation more clearly communicated the position of C-F during polymerase-driven TMSD.

Upon addition of Bst strand-displacing polymerase fluorescence increased and plateaued in <15 minutes, indicating complete transfer of C-F to L2 (Figure 3.8C). If Bst or dNTPs were excluded then only blunt end transfer was possible, and no increase in fluorescence was observed until the sample was annealed, equilibrating C between L1 and L2. Upon annealing this sample I expected C would be distributed equally between L1 and L2 as there was an identical DNA sequence and length on both legs. This would be indicated by a normalised fluorescence intensity of ~0.5. In practice, I observed ~0.1, which indicated that C-F preferentially hybridised to L1-Q. I hypothesised that this was due to an affinity between fluorophore and quencher, which has been documented previously [180]. This would increase the relative stability of C-F + L1-Q and slow the apparent rates of transfer during TMSD, polymerase-driven TMSD and blunt end transfer. I tested this by reversing the order of cargo motion during blunt end transfer L2 to L1-Q, which was indeed ~10× faster than L1-Q to L2, though still much slower that polymerase-TMSD (Appendix 7.2.2). As the fluorophore-quencher interaction is not relevant for the application in DTS-triggered TMSD, it was not explored further.

To test whether the polymerase activity would be compatible with DTS, I varied pH and observed no change in the rate of polymerase-TMSD (Figure 3.9A). Furthermore, the ideal motor should remain functional after long incubations in DTS conditions: up to 5 days incubation at 37°C had no apparent effect on polymerase-TMSD (Figure 3.9B).

These experiments demonstrate that polymerase-driven TMSD is possible, however, for an autonomous programmable polymerase these complexes must

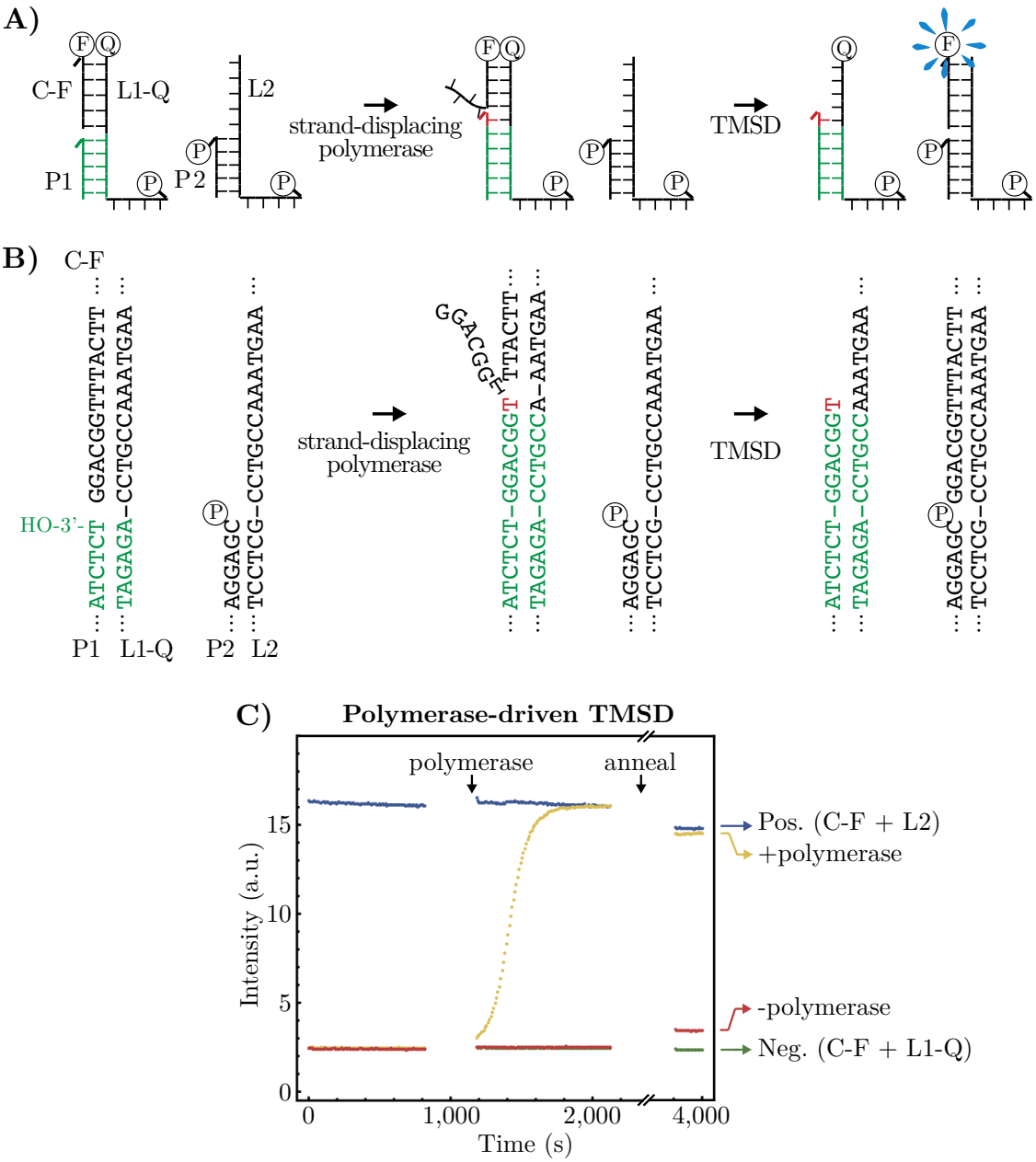


Figure 3.8: Polymerase-driven TMSD. A) Components for the reaction. C-F starts in a quenched position by pre-hybridisation with L1-Q. The addition of L2 enables a slow blunt end transfer of C-F from L1-Q, however, this transfer is accelerated by Bst strand-displacing polymerase. Bst extends from P1 and halts at a dideoxynucleotide (red), creating a toehold that transfers C-F to L2. (F) = 6-FAM. (Q) = BHQ-1. (P) = PO_4 modifications to prevent off-target polymerase activity. **B)** DNA sequences around the site of polymerase activity and TMSD shown for clarity. The dideoxythymine (red) lacking a 3'OH stalls the polymerase. **C) Timecourse of polymerase-driven TMSD.** The positive control (Pos., blue) was initiated in the end position as P1-ext is added before the reaction causing TMSD of C-F to L2. The negative control (Neg., green) remained in the start position as L2 and polymerase are excluded. The background rate of blunt end transfer was tested by excluding polymerase (-polymerase, red). Method in Appendix 7.1.11 and 7.1.12.

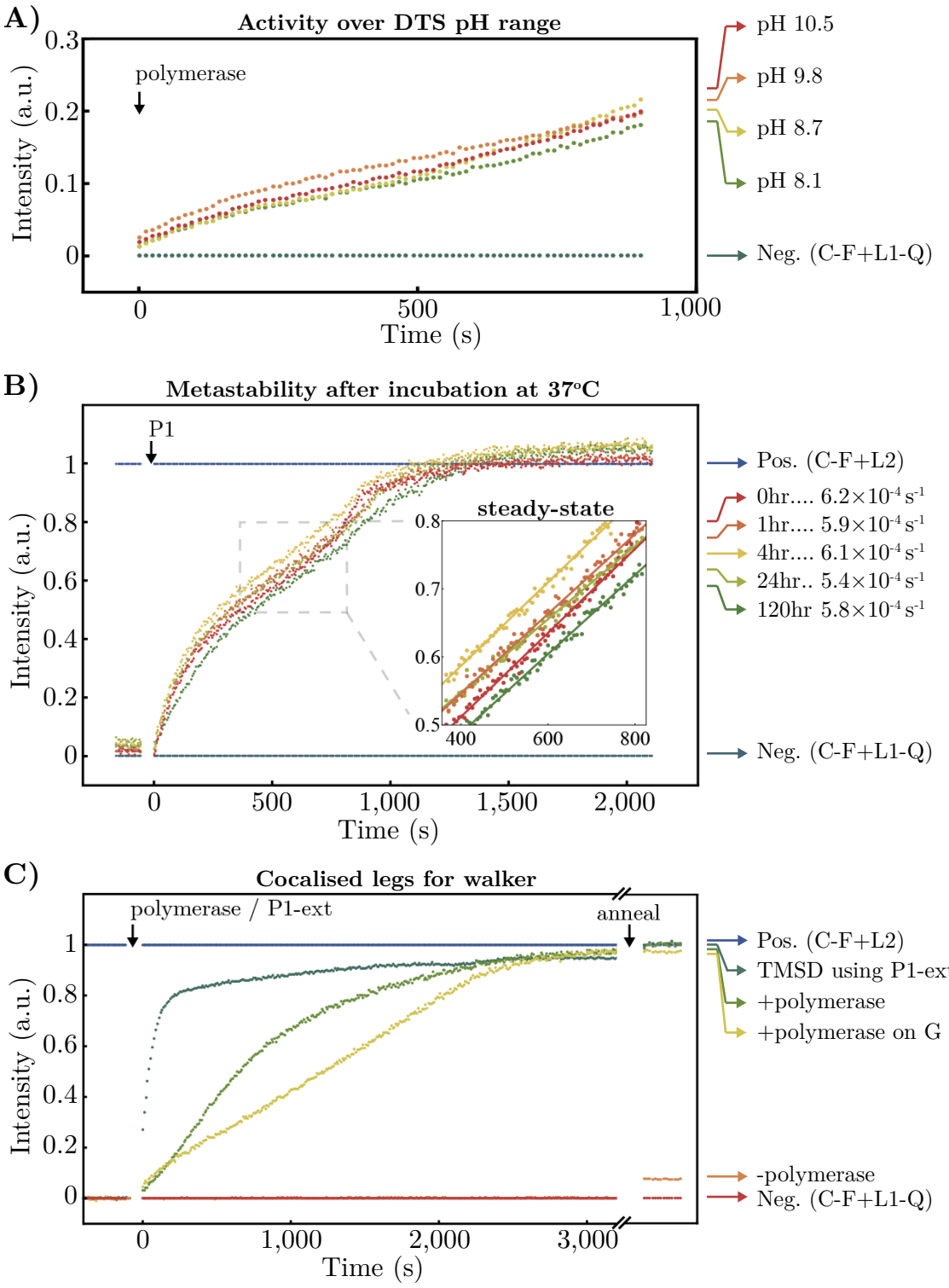


Figure 3.9: Characterising polymerase-driven TMSD for DTS. A) Exploring polymerase-driven TMSD with various pH. **B)** Metastability of polymerase-driven TMSD. Samples including polymerase and dNTPs were incubated at 37°C for up to 120 hours before triggering C-F transfer by adding the primer P1. Linear fits to the steady-state regime (inset) show a similar rate regardless of incubation time. **C)** The effect of colocalisation on polymerase-driven TMSD. Cargo C-F transferred between legs L1-Q and L2 triggered by the addition of polymerase (light green and yellow). TMSD (dark green) was faster when triggered by adding P1-ext without P1, as there is no requirement for the polymerase to create a toehold. Method in Appendix 7.1.11 and 7.1.12.

be colocalised on the gene G that programs the set of DTS reactions. Colocalising the strands on G might alter the rates of transfer through the high C_{eff} of legs, and possibly the rate of polymerase activity through steric constraints. This was tested by simply running samples side-by-side with the addition P1-ext that triggers TMSD and cargo transfer without steric constraints on polymerase activity. I also included or excluded the gene (G) that colocalises L1-Q and L2 (Figure 3.9C). C-F was transferred from L1-Q to L2 fastest by TMSD, slower by polymerase-TMSD, and slowest by colocalised polymerase-TMSD. All were faster than background transfer. Interestingly, the polymerase-driven transfer appears to have a linear phase. This makes sense as there is a ~10× molar excess of DNA over polymerase in this system, so that there is at first a rapid burst phase until all polymerases are occupied, creating a linear steady-state that is rate-limited by dissociation of polymerases from DNA after extension (as has been observed for related polymerases [182, 183]).

In summary, polymerase-driven TMSD can be used to drive cargo between legs of a DNA motor. Furthermore, this design incorporates a small molecule fuel (dNTPs) which can be used in vast excess of other motor components and is stable over long periods of time.

3.2.4 Nickase-triggered TMSD

An important feature of feedback in an artificial programmable polymerase is that a translocation reaction cannot occur before a transfer reaction. With the transfer chemistry developed in Chapter 2, this means that translocation should not occur before the 3'-ester labile linker is cleaved to produce a 3'-hydroxyl. Before working with DTS adapters which require synthesis, purification and protection from degradation, I tested a system that would emulate DTS by creating the 3'-hydroxyl that triggers polymerase-TMSD. A simple way of implementing this is through the use of a nicking enzyme, which cleaves the DNA backbone at a specific recognition site to produce a 3'-hydroxyl. I used the nickase Nt.AlwI as it operates in compatible buffer conditions with Bst polymerase and imposed the least sequence constraints.

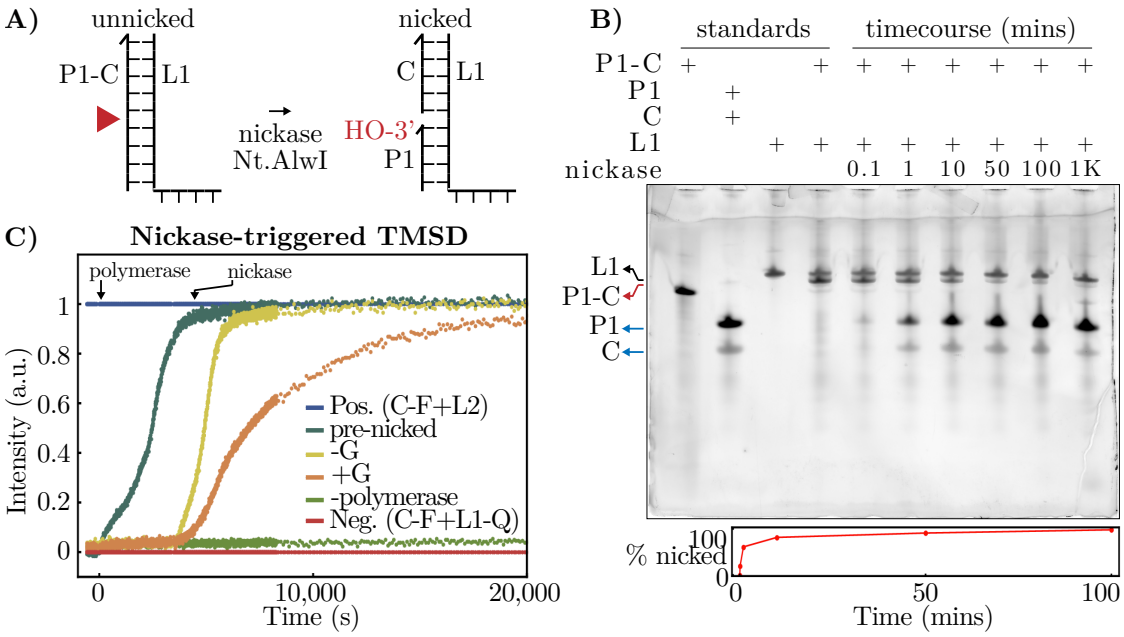


Figure 3.10: A) Simulating DTS with 3'-ester donor adapters by the nickase enzyme, Nt.AIwI, which nicks a specific sequence in P1-C. This creates a 3'-hydroxyl for the strand-displacing polymerase Bst. **B) Nickase activity.** The duplex formed of two strands, L1 and P1-C, was incubated with nickase then run on 25% denaturing PAGE and stained with SYBR Gold. Over time the P1-C band (red arrow) fades while P1 and C appear (blue arrows). The percentage nicked over time was calculated by gel densitometry. Method in Appendix 7.1.6 and 7.1.13. **C) Nickase-triggered TMSD.** As before, using a pre-nicked set of strands permits Bst polymerase to create a toehold for TMSD. However, with the P1-C strand no TMSD occurs until after the nickase is added. Method in Appendix 7.1.11 and 7.1.12.

Firstly, I designed a new strand incorporating the nickase recognition site (P1-C). If the nickase cleaved P1-C, this would produce the original strands P1 and C (Figure 3.10A). As before, P1 was a substrate for polymerase TMSD. I tested nickase activity by incubating the enzyme with P1-C and L1 dsDNA substrate. By running the sample on denaturing PAGE to dissociate dsDNA into component strands, I observed P1-C disappear over time while P1 and C appear as faster-migrating bands (Figure 3.10B). 98% of P1-C was nicked in 100 minutes.

I used the same fluorescence timecourse as before to assay cargo transfer. If the appearance of the 3'-hydroxyl triggered polymerase recognition, then no cargo transfer and fluorescence increase would be observed until after the addition of nickase. Adding the nickase by itself did not cause cargo transfer (Figure 3.10C), and neither did it affect polymerase-driven TMSD if the substrate was pre-nicked. The

addition of polymerase did not cause an increase in fluorescence until the subsequent addition of the nickase. This experiment demonstrated that the appearance of a 3'-hydroxyl is sufficient to drive a DNA motor. Combined with the previous work in Chapter 2 showing that DTS produces a 3'-hydroxyl, this provides an enzymatic mechanism for coordinating transfer and translocation reactions in an artificial programmable polymerase.

3.3 DISCUSSION

This chapter has explored DNA nanomachinery to incorporate feedback between templated transfer and translocation in DNA-based artificial programmable polymerases. In the process two interesting contributions to the field of DNA nanotechnology have been made: 1) strand-displacing polymerases can create toeholds of specific lengths for TMSD using ddNTPs, and, 2) dNTPs are a new fuel for DNA motors without appreciable decrease in activity over 5 days at 37°C. Finally, by incorporating this new motor design with previous work in Chapter 2 it will be possible to implement feedback in an artificial programmable polymerase.

I observed that polymerase-driven TMSD is slower when the strands involved are colocalised on a gene (G). At first this is surprising as one expects that the high C_{eff} of strands colocalised on G would accelerate TMSD (as observed for remote-toehold TMSD [165] or HCR on DNA origami [184]). As the new feature in this work is the strand-displacing polymerase Bst, I expect that the polymerase is somehow slowed by colocalisation. There are a few possible explanations:

1. The addition of more DNA to the reaction mix (whether colocalised or not) causes the polymerase to transiently bind and unbind off-target DNA when searching for its substrate. This would slow the rate of binding to P1 and thus polymerase-driven TMSD. I explored this hypothesis by adding unrelated DNA strands to the reaction mix, and found no effect on the rate of polymerase-driven TMSD (Appendix 7.2.3).

2. Polymerase activity might be inhibited by steric constraints in the DNA secondary structure. Bst polymerase is roughly a sphere with ~3.5nm radius (PDB ID: 1L3S [185]), and given that the distance between L1 and L2 on the prototype is >5nm this seems unlikely. Indeed, after hybridising additional DNA strands to the secondary structure, I found no effect on the rate of polymerase-driven TMSD (Appendix 7.2.3).
3. I observed TMSD was faster than polymerase activity (Figure 3.9C). It has been shown that TMSD is faster between colocalised DNA nanostructures [165, 184]. If this is the case, perhaps the cargo C transfers between legs by TMSD before the polymerase has finished synthesis, halted by incorporating a ddNTP, or released the DNA substrate. Cargo transfer means the polymerase is bound to an overhang rather than a nick during the rate-limiting step of DNA release by the polymerase [182, 183]. If there is a difference in substrate release rates then this might explain the slower kinetics in the colocalised system.

Depending on the context in which polymerase-driven TMSD is used, this effect of colocalisation on kinetics may be a desirable or undesirable feature. If explanation 3) is accurate, then this is an easily controlled feature through the stoichiometry of DNA:polymerase in the system. If [polymerase]>[DNA], then the rate-limiting step of substrate release will be irrelevant as each individual polymerase enzyme would on average drive a single cargo transfer reaction.

When this design is incorporated in multistep DNA walkers there will be additional advantages of the new polymerase-driven DNA walker over previous nuclease-driven DNA walkers. For example, in the DNAzyme-driven motor there is a chance that after a nicking event the DNAzyme steps backwards by blunt end transfer rather than forwards by TMSD. If so, then the DNAzyme begins a bounded random walk between previous legs until it eventually steps forward by TMSD. This is much less likely in the polymerase-driven design as all the previous legs are dsDNA duplexes, which should result in a more consistent and faster rate of forward stepping.

However, there are also limitations when moving to a multistep walker, such as the need for track rigidity. If a third leg (L3) were added to the prototype developed here, then there is a chance that the cargo would make out-of-order steps directly from L1 to L3. The conventionally proposed solution is to position the legs of the motor on relatively rigid DNA nanostructures, namely DNA origami [79]. Here a long ssDNA sequence is folded into a discrete object by hybridisation with many short oligomers, creating lots of positions for motor legs to be tethered at discrete intervals on a rigid nanostructure. This has been demonstrated for clocked walkers and HCR-based cascades [184, 186]. However, it is also known that these structures can have lower yields of correct assembly [187]. If a leg is missing from a track this will cause the motor to prematurely halt, causing truncations in synthesised polymers.

To maximise the length of polymers that can be synthesised, long tracks must be reliably assembled, which in turn requires detailed knowledge of the folding pathways and incorporation rates of DNA strands. Indeed it has been shown that oligomers that assemble in the centre of DNA origami tiles are incorporated with higher yield (95% vs. average incorporation yield 84% [187]) and as such this should be a design feature of multistep walkers. Finally, a unique limitation of the polymerase-driven walker is the need for every 3'-hydroxyl throughout an entire DNA origami structure to be modified to prevent strand-displacing polymerase activity displacing all strands from the nanostructure. This could be achieved cheaply by adding poly-T tails to DNA origami staples during solid-phase synthesis, creating frayed ends that are not natural substrates for polymerases.

In summary, moving to a multistep polymerase-driven DNA walker will require solving additional problems with track rigidity, assembly yield, and 3'-hydroxyl blocking. However, this only considers the aspects of the DNA motor. During multistep DTS there is also the additional challenge of the increasing distance between reactants as each monomer is added to the growing polymer. This can only be solved by creating an identical DTS reaction environment in each step, which is explored in Chapter 4.

4

INSERTION

Every polymer sequence is a unique substrate. Programmable polymerases catalyse the same reaction with diverse substrates. This substrate promiscuity is achieved by creating an identical reaction environment around the site of transfer which is sufficiently flexible to accommodate various reactants for polymer synthesis. All previous attempts to design and build autonomous artificial programmable polymerases fail in this regard, as the site of the transfer reaction changes with each monomer added during polymer synthesis. This chapter contributes new DNA nanostructures permitting identical reaction environments for DNA templated synthesis (DTS). I begin by introducing the design of DTS linkers and DNA nanostructures in greater detail to show how it is possible to create identical reaction environments during sequential templated transfer reactions.

4.1 INTRODUCTION

4.1.1 Extension and insertion polymerisation

Adapters for templated transfer have four components (Figure 4.1A): 1) the *active site*, 2) the *monomer*, 3) the *labile linker*, and, 4) the *recognition unit* that enables colocalisation of reactive groups. For example, during ribosomal protein synthesis the active site is a primary amine, the monomer is a unique amino acid side chain, the labile linker is the tRNA 3'-ester, and the recognition unit is the tRNA and its codon sequence. While the specific chemistry for templated transfer reactions may be varied greatly, there is one significant limitation − the monomer must be sufficiently rigid to limit cyclic self-cleaving from the recognition unit.

Currently, artificial programmable polymerases have limited control over the position of the active sites and labile linkers used in polymer synthesis. Unlike ribosomes, they have almost no control over reactant orientation. Without this control there are two possible outcomes of a reaction between reactive moieties on colocalised adapters. As both adapters contain active sites and labile linkers, the product could end up on either adapter depending on which active site reacts first. This makes it necessary to initiate polymer synthesis with modified adapters that lack either an active site or labile linker. This means only one of the two possible transfer reactions occur (i.e., an *acceptor adapter* with an active site but no labile linker will receive a monomer from a *donor adapter* with both an active site and labile linker). If the initiating adapter lacks a labile linker the result is *extension polymerisation* (Figure 4.1B), in which each monomer is added to the growing end of the polymer on the initiating adapter. In contrast, if the initiating adapter lacks an active site this is *insertion polymerisation* (Figure 4.1C), where each monomer is inserted at the bottom of the growing polymer, which itself moves between adapters like the baton in a relay race. The difference is profound and each strategy has its own inherent limitations.

Extension polymerisation. The consequence of extension polymerisation is a reduced rate of templated transfer with each additional monomer. This occurs for

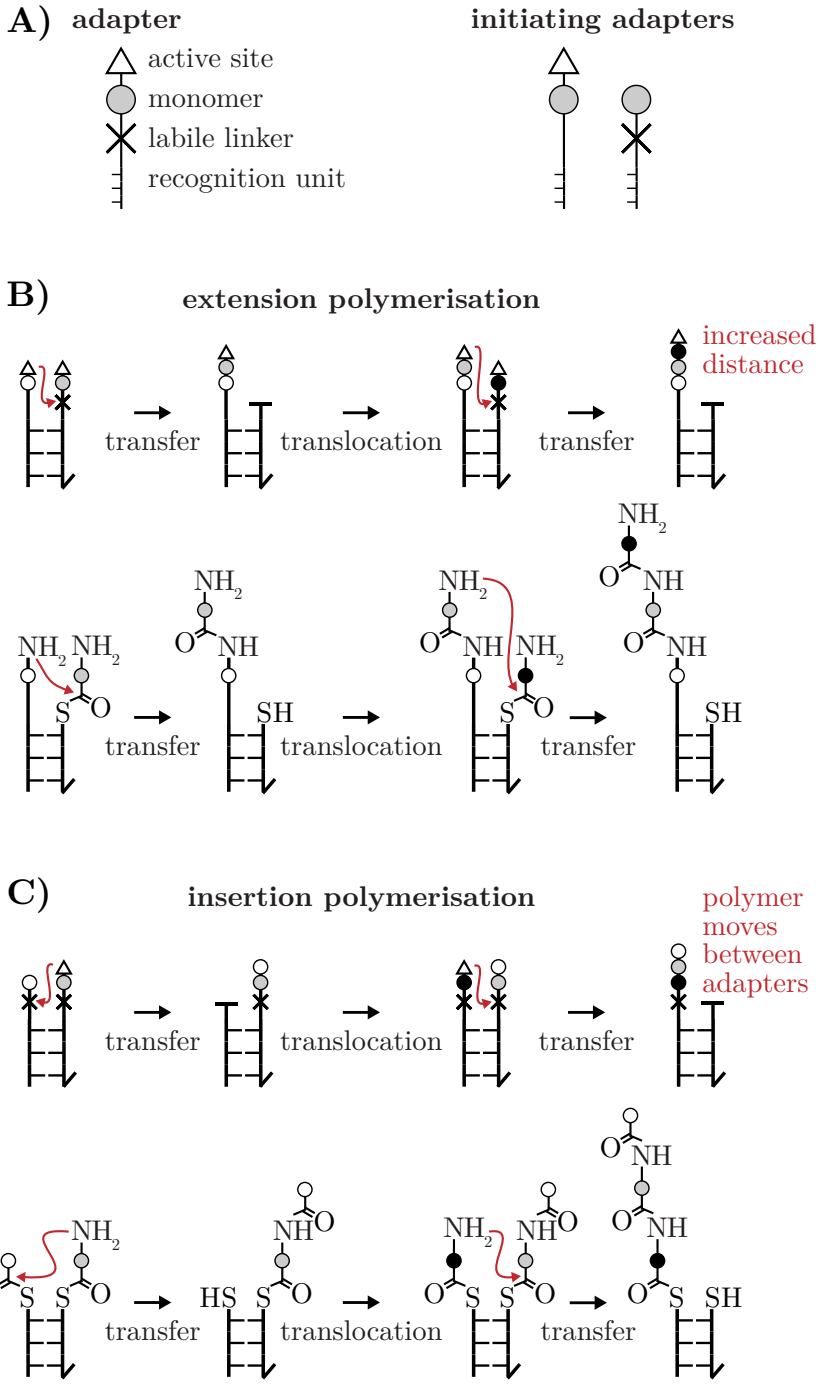


Figure 4.1: A) Nomenclature for adapters permitting polymer synthesis. **B)** If polymer synthesis is initiated with an adapter lacking a labile linker, then the same adapter will be extended and act as an acceptor during all subsequent DTS reactions. As the active site gets more distant from the labile linkers, C_{eff} decreases during polymerisation. In practice (e.g., [111, 133]), the active site could be a primary amine and the labile linker a thioester in a DNA templated reaction forming peptide bonds. **C)** If multistep synthesis is initiated with an adapter lacking an active site, then each alternate adapter will act first as an acceptor adapter and then as a donor, inserting each monomer at the bottom of the growing polymer. C_{eff} remains constant, however, the polymer is always attached to adapters via a labile linker sensitive to degradation.

two reasons: 1) the effective concentration (C_{eff}) is reduced at each step because the active site gets more distant from the labile linker, and, 2) secondary structure within the growing polymer begins to interfere with templated transfer. Eventually the polymer will grow so long that the C_{eff} is equal to or less than the bulk concentration of adapters, at which point programmable synthesis is lost as transfer reactions result from random collisions between adapters.

The reduced yield of DTS with distance has been observed in multiple systems. For example, Snyder et al. [188] explored an end-of-helix architecture with an ssDNA tail of variable length between the two reactive groups. Longer ssDNA tails resulted in lower transfer yields, though this effect depended on secondary structure within the tail. Similarly, McKee at al. [133] explored an over-the-nick architecture with a variable number of unpaired nucleotides in the ssDNA gap between adapters; they observed maximum DTS yield with a 1nt gap and reduced yield with larger distances. However, these examples measured DTS yields rather than reaction rates directly, and explored distances in DNA architectures rather than the increasing distance of the synthesised polymer itself.

Erben at al. [111] provides the only characterisation of this effect directly by comparing the rate of DTS with acceptor peptides of varying length (Figure 4.2). Crucially, they observed that the rate of DTS (k_{intra}) decreased with longer acceptor peptides, the rate of random background transfer (k_{inter}) remained constant, and the total yield of DTS remained similar (~80%). This implies a decreasing effective molarity (M_{eff}) with each additional monomer in a growing polymer (Figure 4.2B). The transfer reaction becomes slower during extension polymerisation due to the changing reaction environment as the polymer grows, and will eventually reach the point where $M_{eff} = 1$ and the rate of templated transfer is the same as background, untemplated transfer.

This effect is not unique to DNA-based artificial programmable polymerases. It is also expected in the supramolecular rotaxane-based programmable polymerase [70], though this system has not been thoroughly characterised. However, a similar trend has been observed in the study of peptides with lengths <20 amino acids,

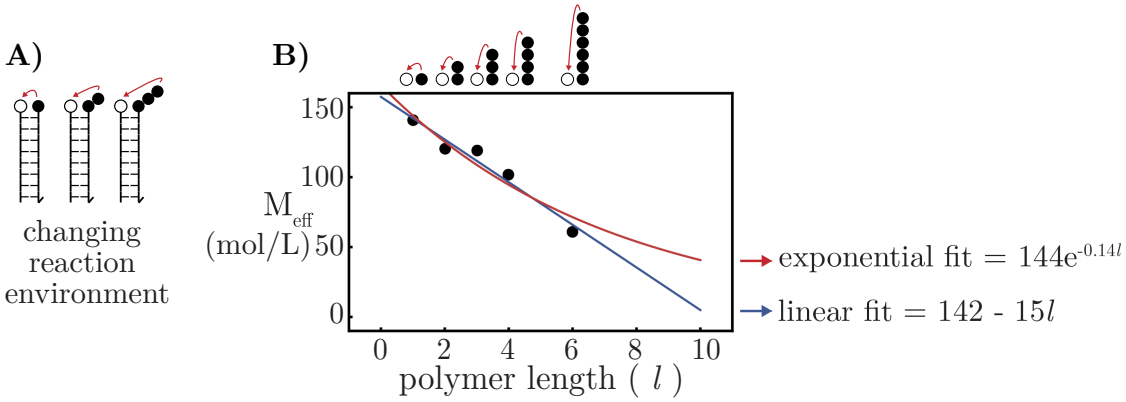


Figure 4.2: A) All published artificial programmable polymerases [70–72, 98, 136, 189], use extension polymerisation, which adds monomers to the distant end of the growing polymer. **B)** Longer polymers have lower $M_{eff} = k_{intra}/k_{inter}$ during templated transfer, measured by Erben at al. using native chemical ligation in over-the-nick PNA templated synthesis [111]. There is no effect of colocalisation on the rate of transfer when $M_{eff}= 1$. A linear fit indicates that $M_{eff} < 1$ for $l > 10$, whereas for an exponential fit this is $l > 36$. k_{intra} was measured for the templated reaction whereas k_{inter} was measured with mismatched strands, in both cases with the same chemistry transferring the quencher Dabcyl between strands. In their system, acceptor peptides have different sequences: 1 = Cys-NH_2, 2 = Ala-Cys-NH_2, 3 = Gln-Ala-Cys-NH_2, 4 = Lys-Gln-Ala-Cys-NH_2, 6 = Lys-Gln-Ala-Ile-Pro-Cys-NH_2. Data adapted with permission from [111] © 2011 Elsevier.

where intramolecular acyl migration from the thioester of internal cysteines to terminal amines proceeds through a ring-closing transition state. The rate of transfer generally decreases for longer peptides, though this also depends on peptide sequence and secondary structure [190, 191]. This is further evidence that the changing reaction environment during polymer synthesis will cause reduced yield of longer polymers. This is a key design flaw common to all attempts to build artificial programmable polymerases.

Insertion polymerisation. This is the strategy used in nature by the RNA ribosome, allowing it to synthesise proteins with thousands of monomers (e.g., titin has >30,000 amino acids [192]). It is also the strategy used by non-ribosomal peptide synthases (NRPS), unprogrammable multi-domain enzymes that catalyse the synthesis of <50mer peptides from a diverse array of non-ribosomal amino acids [6, 193]. Rather than extending on the end of the growing polymer, each monomer is inserted at the bottom in an unchanging reaction environment. Ideally, insertion polymerisation should be incorporated in an autonomous DNA-based

artificial programmable polymerase, though this requires solving two problems of symmetry and degradation.

DTS adapters are typically modified at the 5' or 3' ends during solid-phase synthesis of the DNA strands. As dsDNA is an antiparallel duplex of strands, during insertion polymerisation the polymer alternates from the 5'→3'→5' end of different adapters at each synthesis step. This means that two adapters must be synthesised for each monomer in the sequence in order to encode arbitrary polymer sequences (having only 5'-modified adapters would limit each monomer to every second position in the sequence). Furthermore, it is challenging to combine with published DNA motors. For example, if insertion polymerisation was used in the conventional DNA walker scheme (e.g., Figure 1.7C), then only a single transfer reaction could occur before the polymer is transferred from the 3'-end of a cargo to the 5'-end of a leg, where it is left stranded while the cargo continues walking. New motor designs are necessary for insertion polymerisation.

To solve the problem of symmetry the ribosome uses only 3' modified adapters. The 3'-ends of two tRNAs are held in close proximity within the peptidyl transfer centre (PTC) for each transfer reaction so that the polymer alternates from the 3'→3'→3' of tRNAs. There is two-fold symmetry of rRNA within the PTC and the tRNA's 3'-ends are translated and rotated 180° relative to each other [194, 195]. In this way symmetric adapters can act as acceptors during their first transfer reaction, then rotate to act as donors during the second transfer reaction. After the peptide bond forming transfer reaction, the donor adapter is ejected and the ribosome moves around the acceptor adapter, moving it into position to act as a donor adapter.

In addition to symmetry, there is the problem of degradation. During insertion polymerisation the polymer is always attached to adapters by a labile linker. These labile linkers either react with an active site to synthesise polymers, or degrade via side reactions with solvent or template. The polymer is almost inevitably lost from adapters due to degradation (e.g., by ester hydrolysis in acyl transfer or phosphine oxidation in Wittig olefination). While the identical reaction environment means the rate of templated transfer should remain constant during polymer synthesis,

the yield will still decrease for longer polymers due to the possibility of each labile linker degrading. Insertion synthesis has been explored during three steps of DTS with externally supplied adapters and purification between steps [196]. As expected, it produced polymers of the same length as extension polymerisation though with lower yield, presumably due to adapter degradation.

The ribosome addresses the problem of degradation with at least two mechanisms: 1) enzymatic recharging of tRNAs, and 2) protection of tRNAs from hydrolysis (both the protein Ef-Tu and the ribosome prevent water from accessing the tRNA's 3'-ester [40, 197, 198]). A change in linker chemistry limiting degradation or a mechanism to protect adapters from hydrolysis is therefore essential for artificial programmable polymerases. The current chapter explores symmetric architectures for DTS, while Chapter 5 pursues protection from degradation.

4.1.2 Symmetric architectures for DTS

Nucleic-acid templated transfer reactions were first explored by both Gerald Joyce's and Jack Szostack's groups in 1996 [199, 200]. In the first case an over-the-nick architecture was designed to transfer a peptide between modified DNA strands. In the second case RNA ribozymes were evolved with selection for the transfer of biotin between strands and an over-the-nick architecture evolved from a pool of random sequences. However, over-the-nick DTS is necessarily an asymmetric transfer between the 5' and 3' ends of DNA strands.

In 2001 David Liu's group demonstrated that an end-of-helix architecture is capable of templated transfer reactions [95], but this also has the same limitation of asymmetric adapters. In 2009 Kurt Gothelf's group were the first to use a symmetric architecture for DTS using internally modified adapters within a multi-armed Holliday junction [201]. They used a different type of transfer chemistry, where the linkers were not auto-cleaving during transfer reactions, but instead required two sequential chemical reactions (first conjugation then cleavage) driven by external changes in reaction conditions. Neither the chemistry employed nor the junction architecture are conducive to an autonomous DNA motor.

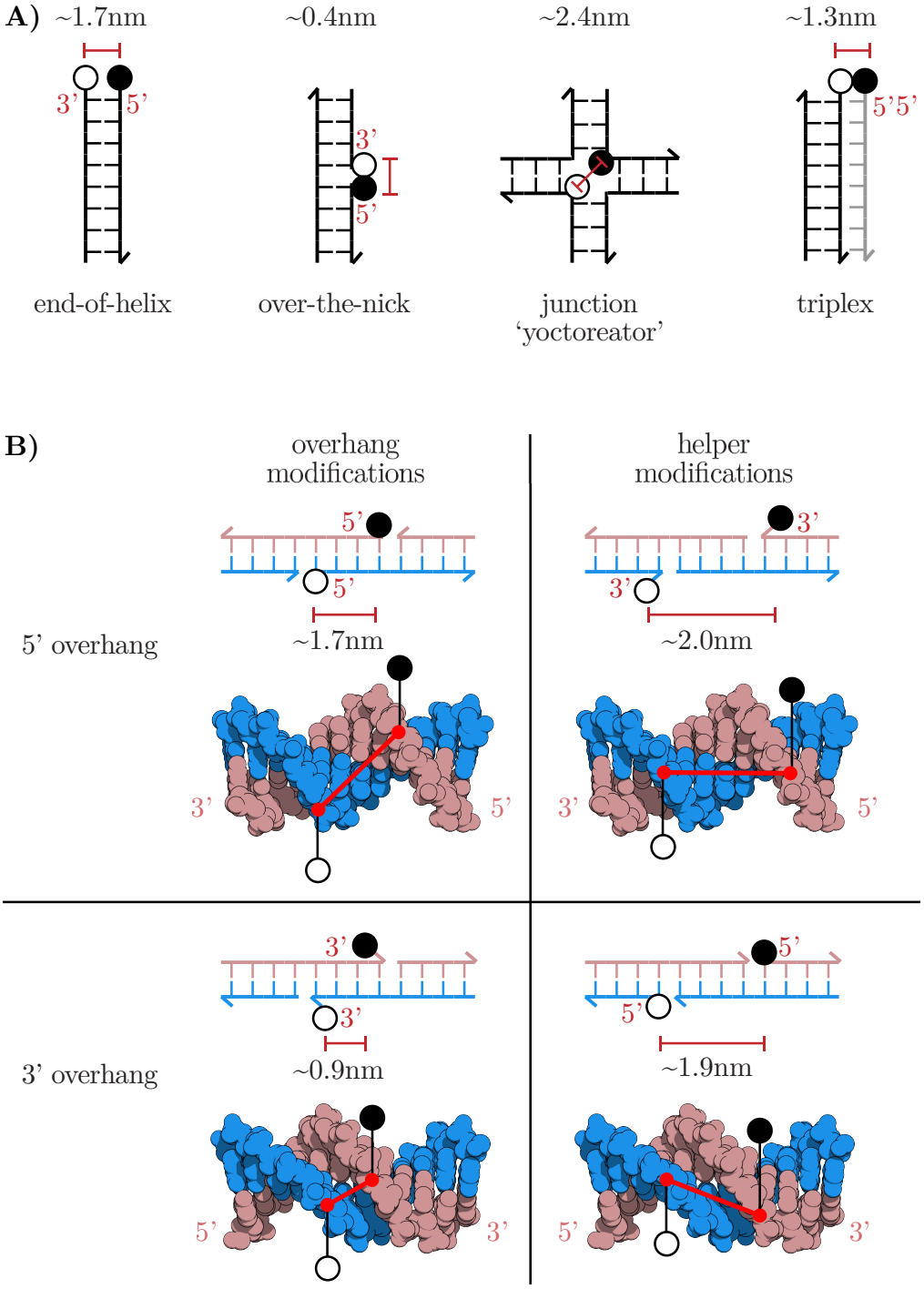


Figure 4.3: A) Comparison of published DTS architectures. The 5' to 3' asymmetry between adapters is highlighted, as are the distances between linker attachments points measured from crystal and NMR structures (PDB: 1BNA, 1WAN, 2CRX [116–118]). **B)** Four overhang architectures for DTS. A sticky end is formed by the hybridisation of four strands. Two strands form the 'overhangs', whereas the other two strands are 'helpers' that stabilise the duplex via base-stacking interactions. Upon hybridisation this could bring two reactive groups into close proximity for DTS. Interestingly, the 3' modified overhang architecture has the closest attachment points for DTS reactants, as it colocalises reactants over the minor groove. Images of PDB:1BNA [116] generated by QuteMol [202].

INSERTION

An alternative to the dsDNA duplex might be a triplex formed by non-Watson-Crick interactions between strands. Triplexes have been used in conjugation reactions over-the-nick [119], and one can imagine end-of-helix variations that colocalise two of the three strand ends with the same directionality (Figure 4.3A). While a transfer reaction in this architecture seems feasible (distance ~1.3nm, PDB:1WAN [117]), there are a few design limitations for multistep DTS [203]. Firstly, triplexes have sequence limitations, requiring runs of either purines (A/G) or pyrimidines (C/T). Secondly, some triplex motifs and sequences are pH-dependent (for example, cytosine requires protonation at pH < 6.5), which may be challenging as both DTS and degradation reactions are also pH-dependent. Finally, while strand exchange via toehold mediated strand displacement (TMSD) with triplexes has been demonstrated [85], there have been no autonomous DNA motors using triplexes upon which to develop an artificial programmable polymerase.

An alternative might be an *overhang* DTS architecture, caused by the hybridisation of 5' or 3' sticky ends. Modifications could be made to any end of the 4 different strands in this architecture. This results in a variety of architectures with symmetric adapters (Figure 4.3B), in the sense that monomers might be transferred from 3'→3'→3' end of adapters during subsequent transfer reactions. Interestingly, by inspection of DNA crystal structures, I found that two 3' modified sticky ends would not only be symmetric, but also position reactants in close proximity (~0.9nm) due to the minor groove of the DNA helix. In contrast, the attachment points of modifications on 5'-modified sticky ends with the same number of base pairs are more distant (~1.7nm) due to the position of the major groove. I decided to characterise this architecture as the effect of nucleic acid helicity has not been explored previously for DTS reactions. Furthermore, it is reasonable to imagine combining with conventional DNA nanomachines, as the first autonomous DNA motor involved the hybridisation of sticky ends [94]. Finally, the overhang architecture might also be compatible with the 3'-ester transfer chemistry explored in (Chapter 2).

4.2 RESULTS

4.2.1 Overhang conjugation

To test if reactions on overhangs are possible, I used a templated disulfide conjugation developed by De at al. [204]. Two disulfide modifications are colocalised by a DNA nanostructure and incubated with reducing agent (e.g., TCEP, tris(2-carboxyethyl)phosphine). This reduces disulfides, triggering disulfide conjugation[1]. I used denaturing PAGE to distinguish hybridised complexes that dissociate in denaturing conditions from conjugated complexes that migrated more slowly. Figure 4.4A depicts this experiment, where a complex between a helper strand and a longer adapter strand create 3' sticky ends bearing disulfide modifications. This complex can dimerise via the 4nt self-complementary palindromic sequence of the sticky ends (5'-GCGC-3'-SS), colocalising two disulfide modifications to test templated conjugation. If this formed a slower migrating band on denaturing PAGE it would indicate conjugation over the minor groove with 3' modified overhangs. If there was no slowly migrating complex, then templated conjugated is not possible on the overhangs.

Mismatched overhangs (5'-GTCT-3'-SS) did not form a slower migrating band (Figure 4.4B), whereas matched overhangs did (5'-GCGC-3'-SS). This experiment demonstrated that the overhang architecture can colocalise groups in sufficient proximity for templated conjugation reactions. However, the range of conditions for disulfide conjugation was limited. End-of-helix disulfide conjugation occurred in similar yield between 5-40°C (Appendix 7.2.5), whereas conjugation of 4bp overhangs was limited to low temperatures (Figure 4.4C). Conjugation between 0nt (blunt ends) and 2nt overhangs was not observed (Appendix 7.2.6). On 4nt overhangs templated conjugated only occurred in the presence of helper strands (Figure 4.4B),

[1]TCEP indiscriminately reduces disulfides to thiols and is expected to also break the disulfide bonds formed between DNA strands. However, TCEP is oxidised during the reaction and thus is removed over time. By incubating an end-of-helix architecture at various [TCEP] for various amounts of time, I showed that conjugation between strands requires [TCEP] > [DNA-SS], and that high concentrations of TCEP require a longer time for conjugation between strands to occur (Appendix 7.2.4), presumably as disulfides conjugate and cleave many times until all TCEP oxidises.

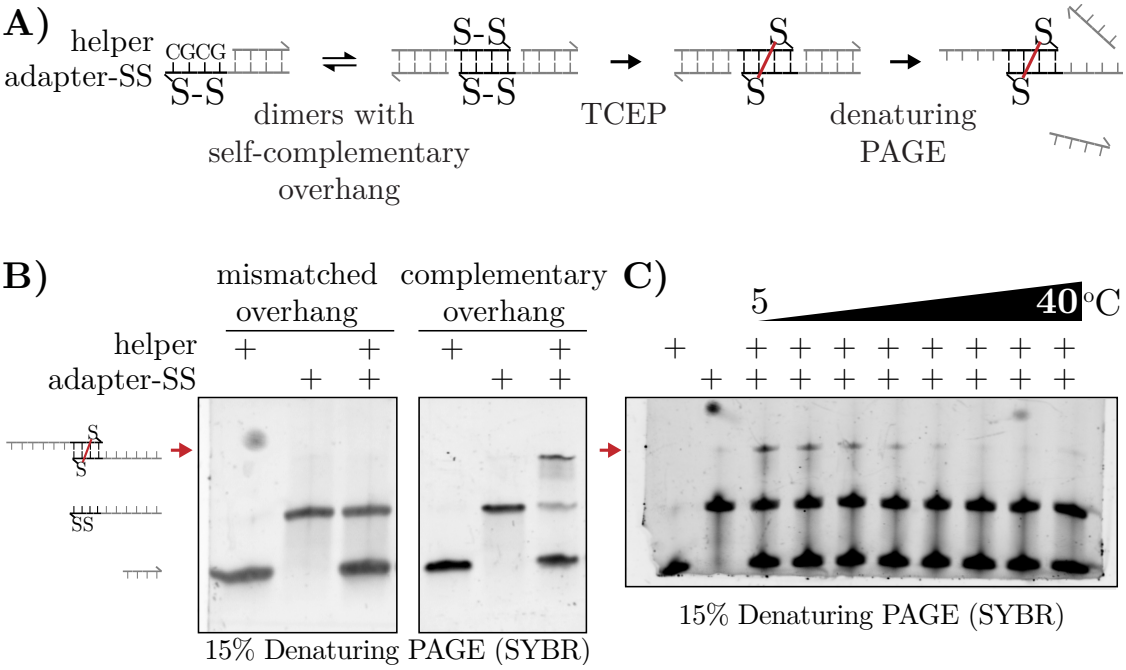


Figure 4.4: A) Scheme for disulfide conjugation between self-complementary overhangs formed by hybridisation of adapter-SS and a shorter helper (grey). **B)** A mismatched overhang (5'-GTCT-3'-SS) did not produce a slower migrating disulfide conjugated complex (red arrow), in contrast to the self-complementary palindrome (5'-GCGC-3'-SS). 1µM DNA was incubated with 500µM TCEP overnight (~15 hours) at RT. **C)** Samples were incubated with TCEP for 1hr at various temperatures between 5-40°C during the conjugation reaction. Method in Appendix 7.1.6 and 7.1.14.

which presumably stabilise overhangs through the two base-stacking interactions (predicted to increase stability of sticky ends by −6 to −9 kcal/mol [205–207]). This indicated that conjugation reactions is only possible with stable overhangs.

4.2.2 Overhang transfer

Having established that templated conjugation reactions were possible, I aimed to demonstrate templated transfer reactions. I synthesised DTS donor adapters with a thioester labile linker and a fluorescent monomer (5'-GCCG-3'-S-O-TAMRA) by reducing 3'-disulfide modifications to form 3'-thiols with TCEP and then incubating with NHS-TAMRA (Figure 4.5A). Donor adapters were purified by HPLC (Figure 4.5B) and verified by LC-MS (Appendix 7.2.7) before testing DTS. I designed adapters that transferred the fluorescent monomer TAMRA (6-carboxytetramethylrhodamine) between strands, as this provides an easy assay of

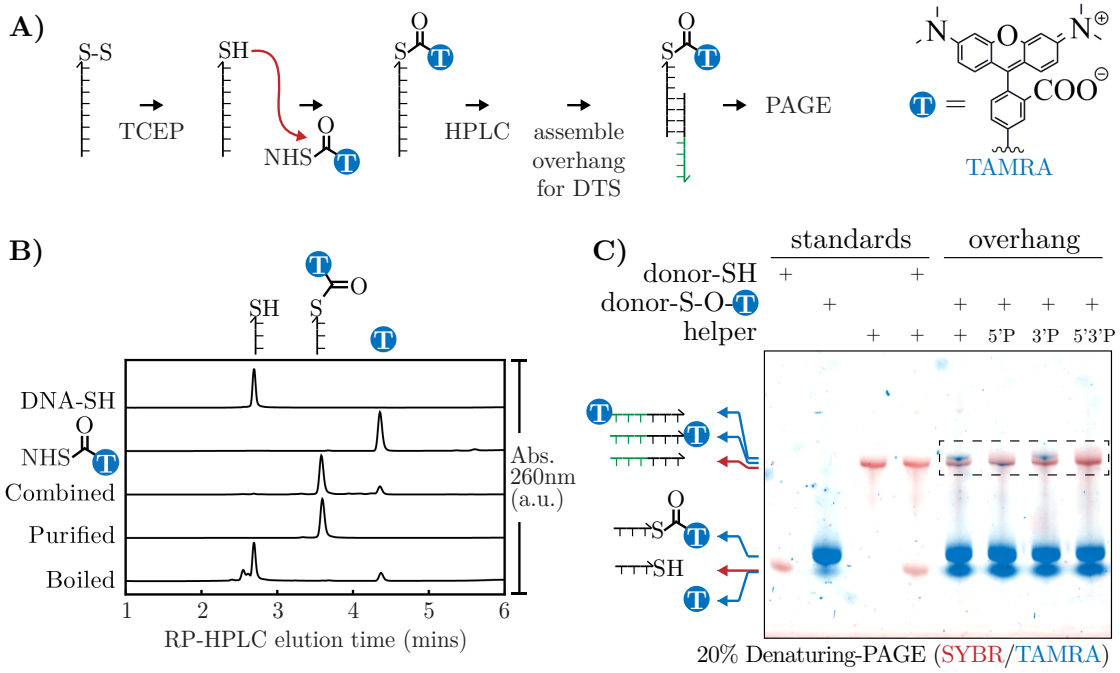


Figure 4.5: Synthesis of donor adapters for overhang DTS. A) Scheme for synthesis, purification and assembly of overhang architecture with 3' modified sticky-ends (5'-GCCG-3'-S-O-TAMRA). **B)** HPLC of reactants combined during synthesis of donor adapters. The sample was re-run on HPLC after purifying the desired product, and also after boiling at 95°C and pH 11 for 20 minutes to test thioester hydrolysis into the components (DNA and TAMRA). The product was also verified by LC-MS (Appendix 7.2.7). Method in Appendix 7.1.15. **C)** During assembly of the overhang architecture, I noticed transfer of transfer of TAMRA to the helper strand (slow, faint blue bands in the dashed black box). Adding both 5' and 3' phosphate modifications (P) prevented transfer to helpers. Method in Appendix 7.1.6.

DTS by PAGE and HPLC. Acceptor adapters for DTS were created by substituting the second 3'-disulfide modification with a 3'-amine modification (5'-CGGC-3'-NH_2). This design tests templated acyl transfer of TAMRA from a 3'-thioester to a 3'-amine when colocalised by 4bp overhangs.

When assembling the overhang architecture for DTS I noticed something unexpected. Simply hybridising the donor adapter (5'-GCCG-3'-S-O-TAMRA) with the helper strand to form the 4nt overhang caused the helpers to become fluorescent (Figure 4.5C). This indicated transfer of TAMRA from donor adapters to helpers in the absence of any nucleophilic modification of the helper strand. I speculated that transfer might occur if a native group on the helper strand acted as a nucleophile to attack the thioester (e.g., an amine on a nucleobase, or a 3'- or

5'-hydroxyl). I tested this by boiling the samples before denaturing PAGE, and the fluorescence of helpers disappeared. This indicated that a 3'- or 5'-hydroxyl on the helper was acting as a nucleophile, as this would produce a 3'- or 5'-ester which is prone to hydrolysis rather than a relatively stable peptide bond that would be formed by a nucleobasic amine. To confirm this I added phosphate modifications to the 3' and 5' ends of the helper strand, whereupon the transfer of TAMRA to helpers stopped (Figure 4.5C). To my knowledge this is the first demonstration that an unmodified DNA strand can act as an acceptor adapter for DTS. However, it is not entirely surprising as the 3'-hydroxyl is a nucleophile during nucleic acid synthesis. Subsequent experiments aimed to test DTS between two 3' modified adapters over sticky ends in the overhang architecture. To avoid the complication of unintended transfer to both helper strands and the intended acceptor adapters, I used helper strands with both 3' and 5' phosphate modifications to prevent transfer to helpers.

If DTS is possible in the overhang architecture I would expect that bands with TAMRA fluorescence migrate to different positions on denaturing PAGE, indicating that the TAMRA fluorophore has transferred between strands of different length. Indeed, this is what I observed in a positive control with the end-of-helix architecture (Figure 4.6). No DTS was observed using 4bp overhangs (Figure 4.6). DTS in the end-of-helix positive control demonstrated that templated transfer with this chemistry is possible, and the previous disulfide conjugation experiments demonstrated that the proximity on overhangs is sufficient for templated reactions. I speculated that the short 4bp overhangs only transiently hybridise, limiting the amount of time in which templated reactions can occur. It has been reported that the average hybridisation time before 6bp duplexes dissociate is ~0.1s [208]. The off-rate decreased exponentially as duplexes decreased in length from 9 to 6bp, implying that the average hybridisation time 4bp duplexes would be ~1ms. This could drive the DTS reaction towards hydrolysis rather than aminolysis, as hydrolysis of thioesters could occur constantly during the reaction whereas templated transfer by aminolysis could only occur during transient hybridisation of overhangs. I tested whether overhang hybridisation was promoted by an architecture

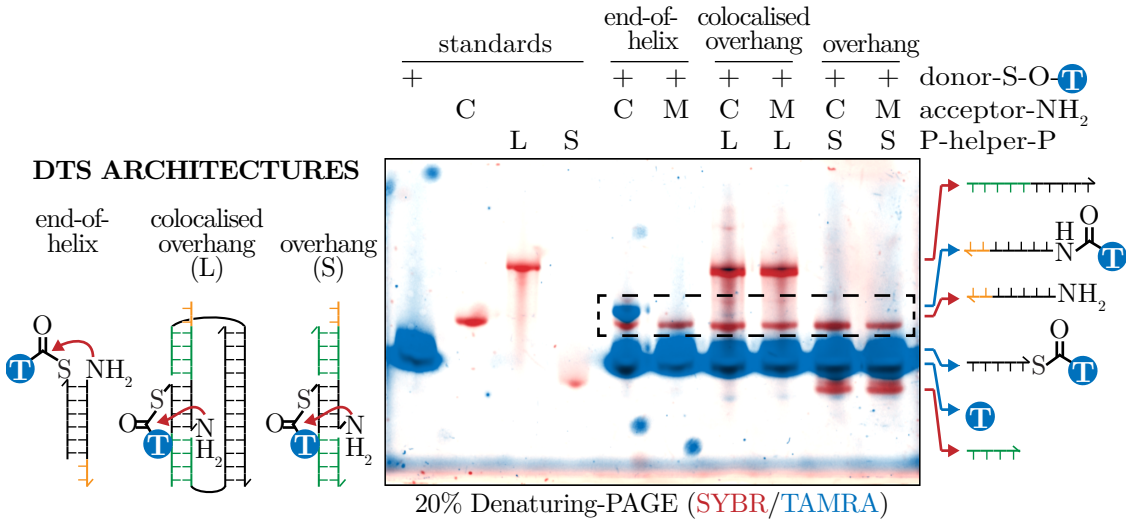


Figure 4.6: No DTS on 3' overhangs. A single donor adapter with a thioester-TAMRA labile linker was used for DTS in various DNA architectures. If DTS occurred in any architecture, a slightly slower blue band would be observed on denaturing PAGE within the dashed box, as the amine modified acceptor adaptor receives TAMRA. The end-of-helix architecture with a complementary acceptor adaptor (C) was a positive control for DTS (slow, blue band within the dashed box), whereas a mismatched acceptor adaptor (M) tested for background, untemplated transfer. Two overhang architectures were tested, either short helpers (S) creating complementary 4bp overhangs (5'-GCCG-3'-S-O-TAMRA and 5'-CGGC-3'-NH_2) or long helpers (L) designed to encourage overhang hybridisation through colocalisation with an adjacent duplex. Donor hydrolysis produced a blur of free TAMRA migrating slightly faster through the gel. Method in Appendix 7.1.6 and 7.1.10.

colocalising the two overhangs with an adjacent duplex, but this also exhibited no DTS (Figure 4.6). Without any way of limiting hydrolysis, it is impossible to observe DTS in the overhang architecture.

4.2.3 Internal modifications – abasic and nucleobasic

The hypothesis that the instability of sticky ends limits DTS can be tested by exploring another DTS architecture. Rather than using terminal modifications on sticky ends that form an unstable duplex containing two nicks in the phosphate backbone, internal modifications within a stable dsDNA duplex may be positioned with a similar proximity between the attachment points of DTS linkers. Internal modifications could be used in a symmetric architecture so that during insertion polymerisation monomers move from internal→internal→internal attachment points. This achieves the same goal as the overhang architecture explored above – symmetry

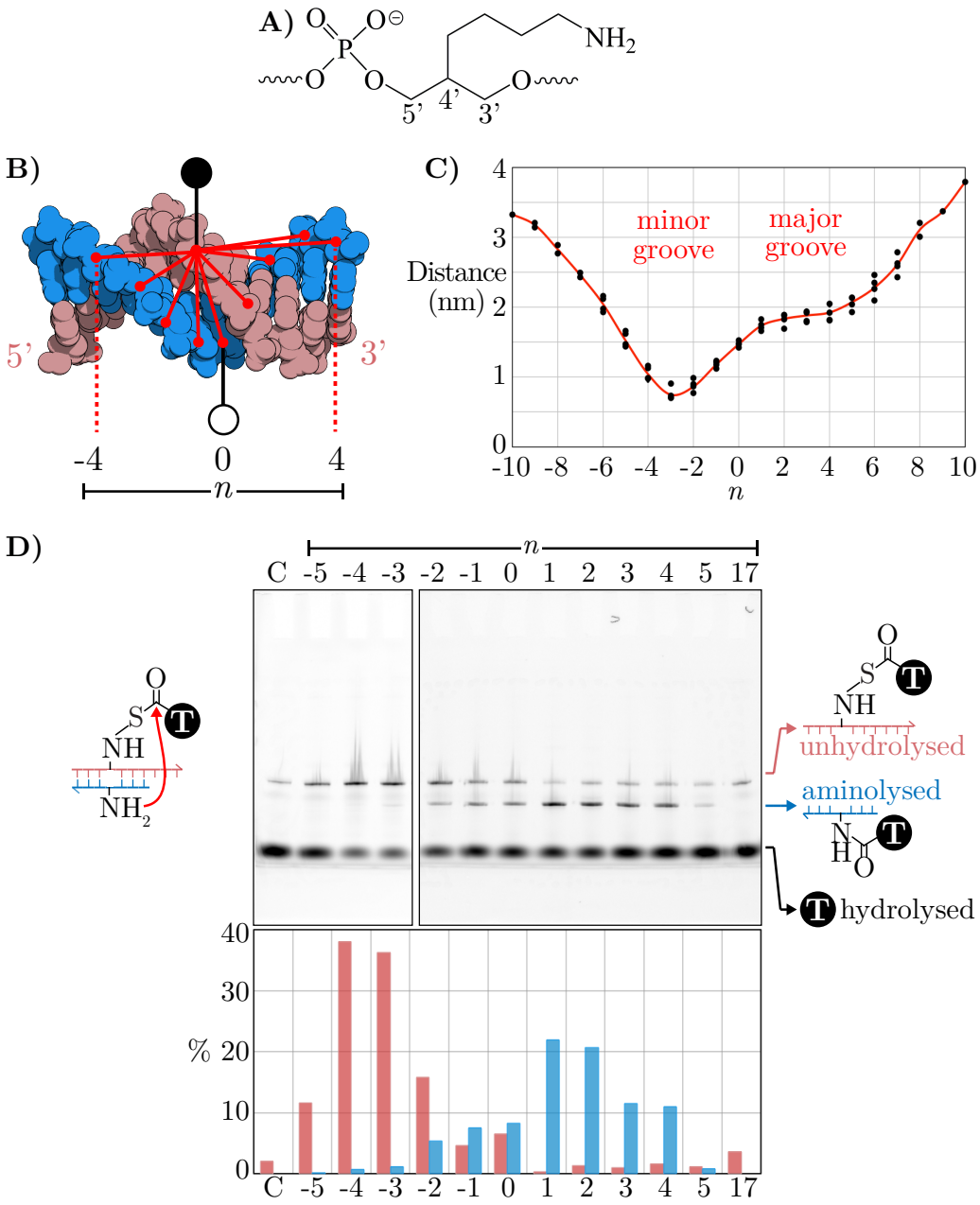


Figure 4.7: DTS with abasic modifications. A) Chemical structure of the internal abasic hexylamine modification. This modification was used as an amine active site during DTS and also as a handle to attach the labile linker thioester-TAMRA. **B)** Duplexes with internal modifications permit DTS with symmetric adapters. The position n of the active site (white circle) was varied relative to the labile linker (black circle), where n increases in the 5'→3' direction from the labile linker (black circle), so that $n < 0$ represents the minor groove and $n > 0$ the major groove. **C)** The distance between attachment points (the 4' carbon) for synthetic linkers was measured from dsDNA crystal and NMR structures (PDB: 1BNA, 1WAN, 2CRX [116–118]). Individual measurements from different structures shown as black points, the mean as a red line. **D)** 25% denaturing PAGE (TAMRA channel) separated DTS adapters to observe the location of TAMRA. The negative control (C) contains only donor adapter. The sample with 17bp between groups tests for untemplated intermolecular transfer. The proportion of TAMRA with the donor adapter (blue) and acceptor adapter (red) after DTS was estimated by gel densitometry. Method in Appendix 7.1.6, 7.1.10, and 7.1.15.

between adapters during DTS.

Two types of commercially available internal modifications were tested: 1) abasic modifications that replace the entire ribose and nucleobase with a hexylamine (Figure 4.7A), or 2) nucleobasic modifications that add a hexylamine to a thymine (Figure 4.8B). These two modifications have very different properties. While a single abasic site still forms a B-form helix with an unpaired, stacked nucleobase, there is conformational flexibility around the abasic site [209], permitting the hexylamine flexibility in position and orientation. In contrast, an internal thymine modification would create a stable base-stacked duplex where the hexylamine points out of the major groove. I explore these designs in the next two paragraphs.

I synthesised donor adapters as before, but using a preliminary step incubating the internal, abasic hexylamine with the cross-linker SPDP (N-Succinimidyl 3-(2-pyridyldithio)propionate) to create an internal, disulfide modification. The product was verified by LC-MS (Appendix 7.2.7) and donor adapters purified by HPLC as before. Due to the much greater stability of the dsDNA duplex relative to overhangs, it was possible to test a greater range of acceptor adapters by varying the position of the abasic hexylamine active site. This was used to test the prediction that DTS yield increases with linker proximity (Figure 4.7C). As before, overnight (~15 hours) DTS reactions were run on denaturing PAGE to separate duplexes and determine whether DTS had occurred. With the new abasic adapters, DTS occurred in a wide range of positions of the amine active site (Figure 4.7D). I use the convention that n increases in the 5'→3' direction from the labile linker, which means that for the position $n = 0$ the two abasic modifications are directly opposite each other. This is analogous to end-of-helix DTS, except that there is a well-formed duplex on either side of the DTS reaction environment. For an active site positioned with $n < 0$ DTS occurs over the minor groove while for $n > 0$ DTS occurs over the major groove. In contrast to my prediction based on proximity alone, DTS occurred readily over the major groove but with low yield across the minor groove, where the attachment points of modifications are physically closer. Furthermore, the fluorescence intensity from the donor adapter disappeared in most samples due to

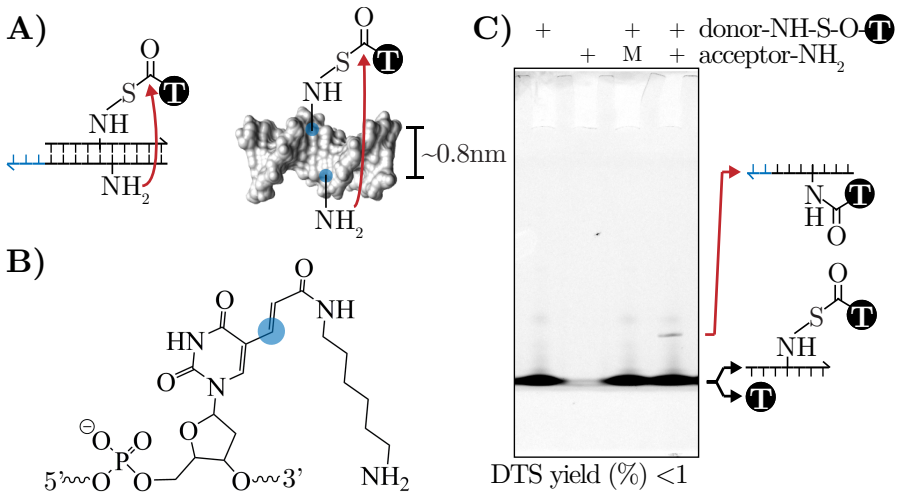


Figure 4.8: DTS with nucleobasic modifications. **A)** Duplexes with internal thymine modifications position reactants for DTS with symmetric adapters within the major groove. **B)** Chemical structure of the internal amine modification on a thymine nucleobase. The blue circle indicates the site of attachment of the hexylamine modification on a thymine within the DNA helix. This modification was used as an amine active site during DTS and also as a handle to attach the labile linker thioester-TAMRA. **C)** 20% denaturing PAGE (TAMRA channel). DTS transferred TAMRA to the longer acceptor adapter (red arrow). A mismatched acceptor adapter (M) tested for untemplated intermolecular transfer. Method in Appendix 7.1.6, 7.1.10, and 7.1.15.

either DTS or hydrolysis of the thioester, but in the minor groove with -5 $\leqslant n \leqslant$ -2 it did not. This indicated that the thioester labile linker is somehow protected from degradation in this architecture, a concept explored further in Chapter 5 for adapter protection during DTS. In summary, DTS with symmetric adapters was possible using internal abasic modifications in a stable duplex.

The second type of internal modification does not create abasic sites, but instead forms a conventional base-stacked duplex with a hexylamine modification pointing out of the major groove (Figure 4.8A-B). In principle, an adjacent pair of AT:TA base pairs would position thymine hexylamines in close proximity at ~0.8nm, while potentially promoting the interaction between by confinement in the major groove. In this case DTS between adjacent nucleobasic modifications was possible, though in very low yield (<1%) compared to other DTS architectures (e.g., ~20% with abasic modifications, Figure 4.7C).

4.3 DISCUSSION

The length of polymers produced by artificial programmable polymerases would be increased by using an identical reaction environment at each step of synthesis. This is because C_{eff} will remain constant during insertion polymerisation. Implementing this requires changes to both the linker chemistry (Figure 4.1) and the adapter design to use symmetric adapters (Figure 4.3). This is challenging due to the 5'→3' directionality of DNA, which creates asymmetry in the conventional end-of-helix and over-the-nick architectures. In this chapter I have explored new architectures for DTS between symmetric adapters. This started with a design using short, overhangs with 3' terminal modifications and proceeded towards a duplex exhibiting DTS between internal modifications.

While exploring chemical reactions on short overhangs, I demonstrated that 3'-SS-3' conjugation is achieved in decent yields (56%, Figure 4.4B). This is interesting as 3'-3' and 5'-5' ligation of DNA does not occur naturally (the closest natural analogue is 5' enzymatic capping of eukaryotic mRNA), while chemical synthesis requires extensive modification of the conventional solid-phase synthesis protocol (using both 3' and 5' phosphoramidites and a bifunctional linker). However, these kinds of 3'-3' and 5'-5' ligated strands have found applications in RT-PCR, mRNA / cDNA display, in the study of triplexes, immunomodulators and bipedal DNA walkers [210–214]. Using the overhang architecture to template 3'-3' or 5'-5' conjugation could obviously be extended to orthogonal and more stable chemistries (e.g., click chemistry [215] or reductive amination [216]) for these applications, while the sequence-specificity of overhangs could enable multiple reactions in one-pot. However, the instability of short overhangs was a key feature limiting DTS in this architecture.

Interestingly, during the characterisation of overhang DTS, I noticed that the 3'- and 5'-hydroxyl can attack a terminal thioester modification. However, it is easily rationalised considering the other instances where the 3'- and 5'-hydroxyl act as nucleophiles (e.g., during solid-phase synthesis of DNA, during enzymatic charging of tRNAs, etc.). Perhaps there is also a role played here by the TAMRA

modification itself, as it is has some regions with positive charge and it is known that similar fluorophores can stack on the ends of DNA duplexes [217]. If so this might align the 3'- and 5'-hydroxyl favourably with the thioester for transfer, which would be unique to the TAMRA modification and explain why it has not been previously reported. As flexizymes were required in Chapter 2 to synthesise 3'-ester donor adapters, and as there are currently no fluorescent flexizyme substrates, this strategy might be optimised in future for the synthesis of either 3'- or 5'-ester modifications enabling DTS to produce enzyme substrates.

DTS was demonstrated using two types of internal modifications. The two architectures with internal modifications are in a sense similar to the four-armed Holliday junction or *yoctoreactor* [201], as an ordinary dsDNA duplex could be considered a two-armed junction. One difference between the architecture here is the closer proximity of reactive groups in a duplex rather than a four-armed junction (~0.7nm vs. ~2.4nm). Another difference is the use of simpler single step transfer reactions (thioester-amine) rather than two-step conjugation then cleavage. The DTS yield with internal abasic modifications was also higher than end-of-helix DTS using the same transfer chemistry and reaction conditions (23% vs. 13% respectively).

An interesting feature of DTS on 3' overhangs was the potential compatibility with the 3'-ester transfer chemistry developed in Chapter 2, where DTS produced substrates for enzymes that coordinate transfer and translocation in an artificial programmable polymerase. However, it now appears impossible for these two design features to be combined because: 1) the 3'-ester adapters were very sensitive to linker chemistry and distance, and, 2) the 3' overhangs didn't enable DTS, presumably due to transient hybridisation and resulting hydrolysis of thioesters. While the internally modified adapters did exhibit DTS, it may be more challenging to implement feedback via enzymatic coordination as the products of DTS are not natural enzyme substrates. In principle, internal modifications might sterically hinder protein binding to the region of the modification. This could limit enzyme activity if sufficiently bulky monomers are used during DTS - as has been demonstrated

in DNA sequencing-by-synthesis [144–147]. However, limitations on the size of monomers and types of enzyme used to coordinate transfer and translocation may make it difficult or not worthwhile to implement feedback with the internally modified adapters developed here.

Implementing insertion polymerisation would enable the synthesis of longer polymers by creating an identical reaction environment during each step of synthesis. In this chapter I argued that this requires two changes to the design of artificial programmable polymerases - adapter symmetry and protection from degradation. Here I have demonstrated that a new kind of DTS adapter with internal modifications solves the problem of symmetry. Chapter 5 builds on this work by exploring mechanisms to protect adapters from degradation.

5

PROTECTION

Protection of tRNAs from hydrolysis is essential for protein synthesis by the ribosome. An analogous system will be necessary for artificial programmable polymerases to synthesise polymers of useful lengths. This might be achieved by changing either the bulk solvent or the local environment in which reactive groups exist during polymer synthesis. This chapter explores both methods for protecting adapters during DNA templated synthesis (DTS). A key contribution is the characterisation of internal abasic modifications to DNA strands which protect a nearby thioester labile linker from hydrolysis. I begin by introducing the need for protection and biomimetic strategies of protecting adapters.

5.1 INTRODUCTION

5.1.1 Adapter degradation

Polymer synthesis requires a set of reactions that form a polymer from monomers. In DNA templated synthesis (DTS) this typically requires activated reactants that are good leaving groups during substitution reactions [95, 100, 106, 109, 111, 115, 121]). For example, activated esters like NHS-esters or phenol-esters are usually easier to react with nucleophiles than thioesters or alkyl-esters, due to the increased electrophilicity of the carbonyl group, which stabilises the negative charge produced during acyl transfer [134]. However, the consequence of using reactive components is that they are also susceptible to side reactions with solvent and template that limit the yield of long polymers. This is particularly challenging in the case of autonomous one-pot syntheses. There is good evidence that preventing adapter degradation will help synthesise longer polymers with artificial programmable polymerases. For example, an externally clocked reaction adding fresh adapters at each step led to greater yield (35% of a 7mer [109]) than an autonomous system with the same transfer chemistry when adapters degraded during polymerisation (5% of a 7mer [136]). Finding ways to increase adapter stability is essential for progress in the field of artificial programmable polymerases.

5.1.2 Strategies from nature

We might take inspiration from three strategies used by natural programmable polymerases to ensure there is a constant supply of adapters for polymer synthesis: **1)** enzymatically synthesising adapters, **2)** using kinetically stable adapters and catalysts to activate them, and, **3)** using unstable adapters and creating local environments to protect them.

1) Turnover. Enzymes or catalysts analogous to tRNA synthetases might be used to continuously synthesise fresh adapters during DTS reactions. However, it is hard to imagine any natural enzymes capable of using conventional DTS linkers, which use long, flexible linkers added to synthetic DNA during or after

solid-phase synthesis. Artificial catalysts might be an alternative, but they are typically too promiscuous to promote the substrate-specific reactions required to synthesise multiple unique adapters in one pot. This strategy might be possible using modified linker chemistries that resemble natural enzyme substrates (like those explored in Chapter 2 with flexizymes). However, this will only work if there are compatible conditions for both DTS reactions and enzymatic adapter synthesis.

2) Catalysis. The dNTPs used by DNA-programmed DNA polymerases are thermodynamically unstable but kinetically stable [179], making them an ideal source of fuel. They are thermodynamically unstable as they have three negative charges in close proximity, but they are kinetically stable because they repel negatively charged nucleophiles that would cleave the phosphate esters [179, 218]. Polymerases (and other enzymes using ATP as a chemical fuel) accelerate degradation of the triphosphate and direct this free energy towards useful reactions. A collaborator, Samuel Núñez-Pertíñez (O'Reilly Group, University of Birmingham), has been working towards catalysis during DTS reactions.

3) Protection. The charged tRNAs used by the ribosome would constantly hydrolyse *in vivo* were it not for the protein Ef-Tu binding to tRNAs, limiting hydrolysis through a network of hydrogen bonds around the tRNA 3'-ester, and finally releasing tRNAs once incorporated in the ribosome-mRNA complex [219]. Analogously, protecting DTS adapters might be achieved by changing the bulk environment, through the effect of pH, temperature, mixed solvents, and so on. Alternatively, the local environment might be altered through DNA secondary structure, the local distribution of negative charges, colocalised hydrophobic groups, or perhaps by evolving a DNA or RNA mimic of Ef-Tu.

This chapter explores this final strategy for improved yield during DTS by changing the environment of labile linkers. The chemistry I use transfers monomers from thioester labile linkers to amine active sites, as this is susceptible to hydrolysis and thus changes in DTS yield can be readily observed.

5.2 RESULTS

5.2.1 Bulk environment

Here I explore the effect of changes to the bulk environment of DTS reactions on adapter degradation and DTS yield.

pH. The effect of increasing pH on DTS reactions with esters is complex - it increases the rate of ester hydrolysis through the increased concentration of nucleophilic hydroxide ions in solution, it promotes amine deprotonation making it available for DTS, while DNA duplexes destabilise at high pH due to nucleobase deprotonation. I varied pH during DTS in the end-of-helix architecture with acyl transfer between 3'-thioester-TAMRA modified donor adapters and 5'-amine acceptor adapters (Figure 5.1). The highest DTS yield (14%) was at pH 11 and 20°C. This is presumably because the double helix is destabilised at higher pH, while the amine is deprotonated at lower pH.

Temperature. The effect of increasing temperature on DTS is similarly complex - high temperatures increase reaction rates, high temperatures ionise water and increase the concentration of nucleophilic hydroxide ions, and sufficiently high temperatures denature DNA. I scanned temperature for three DNA architectures (Figure 5.2). Interestingly, it appeared that the optimal temperature depended on the DNA architecture. With a terminal thioester and internal amine the temperature with maximum DTS yield was 20% at ~25°C, whereas with the internal nucleobasic modifications it was <0.5% at <5°C.

Mixed solvents. Water is a unique solvent. Mixing solvents might affect DTS yield by decreasing the concentration of hydroxide ions causing hydrolysis, changing the reactivity of the nucleophile due to altered hydrogen bonding with solvent, and affecting the self-assembly of DNA (which in water is guided by hydrogen bonding and stabilised by hydrophobic stacking between nucleobases). Indeed, a range of organic solvents (DMF, acetonitrile, etc.) have already been characterised during DTS [220]. While in some cases organic solvents improve DTS yields between strands pre-hybridised in water and then transferred into organic solvents for DTS,

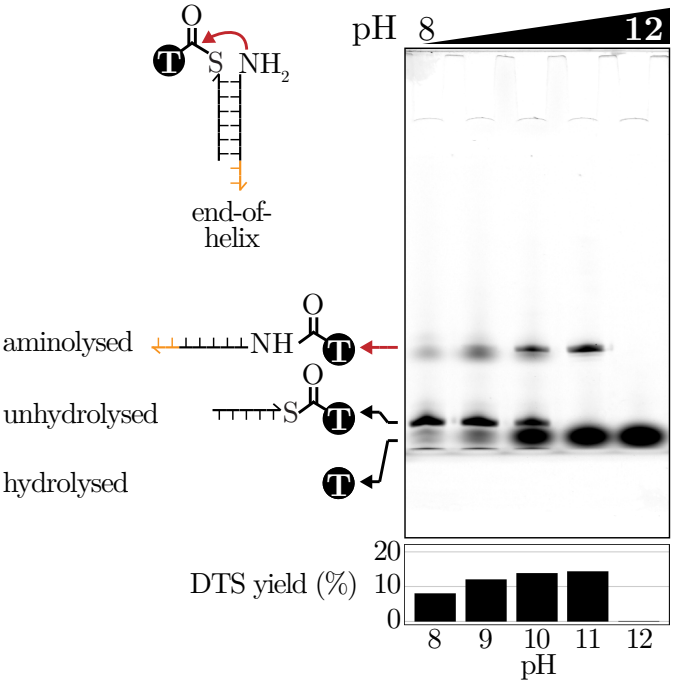


Figure 5.1: The pH was varied in an end-of-helix DTS reaction (20°C for 24 hours) transferring TAMRA (T) from the thioester labile linker on the donor adaptor to the amine active site on the acceptor adapter. DTS yield was estimated by the proportion of the intensity from the upper band (acceptor adaptor, red arrow) divided by the sum of intensity for the entire sample (acceptor + donor + hydrolysed). Method in Appendix 7.1.6, 7.1.10, and 7.1.15.

they cannot be used for multistep syntheses as they typically prevent sequential hybridisation or TMSD reactions.

I explored a range of solvents chosen for their diverse representation of polarity and hydrogen bonding capacity, and thus their capacity for DNA hybridisation. However, with each solvent I observed either increased thioester hydrolysis or no yield of DTS (Figure 5.3A). I also explored a range of polar, protic solvents which might be more likely to permit duplex formation. I used alcohols with hydroxyl groups that have a higher pKa than water so that the effective concentration of hydroxide ions might be reduced. These alcoholic solvents mostly had no effect (Figure 5.3B), though isopropanol promoted hydrolysis and ethylene glycol promoted both hydrolysis and DTS.

These results indicate that simple variations of mixed solvent exhibit reduced yield in DTS relative to water (similar to the results from McKee et al. exploring

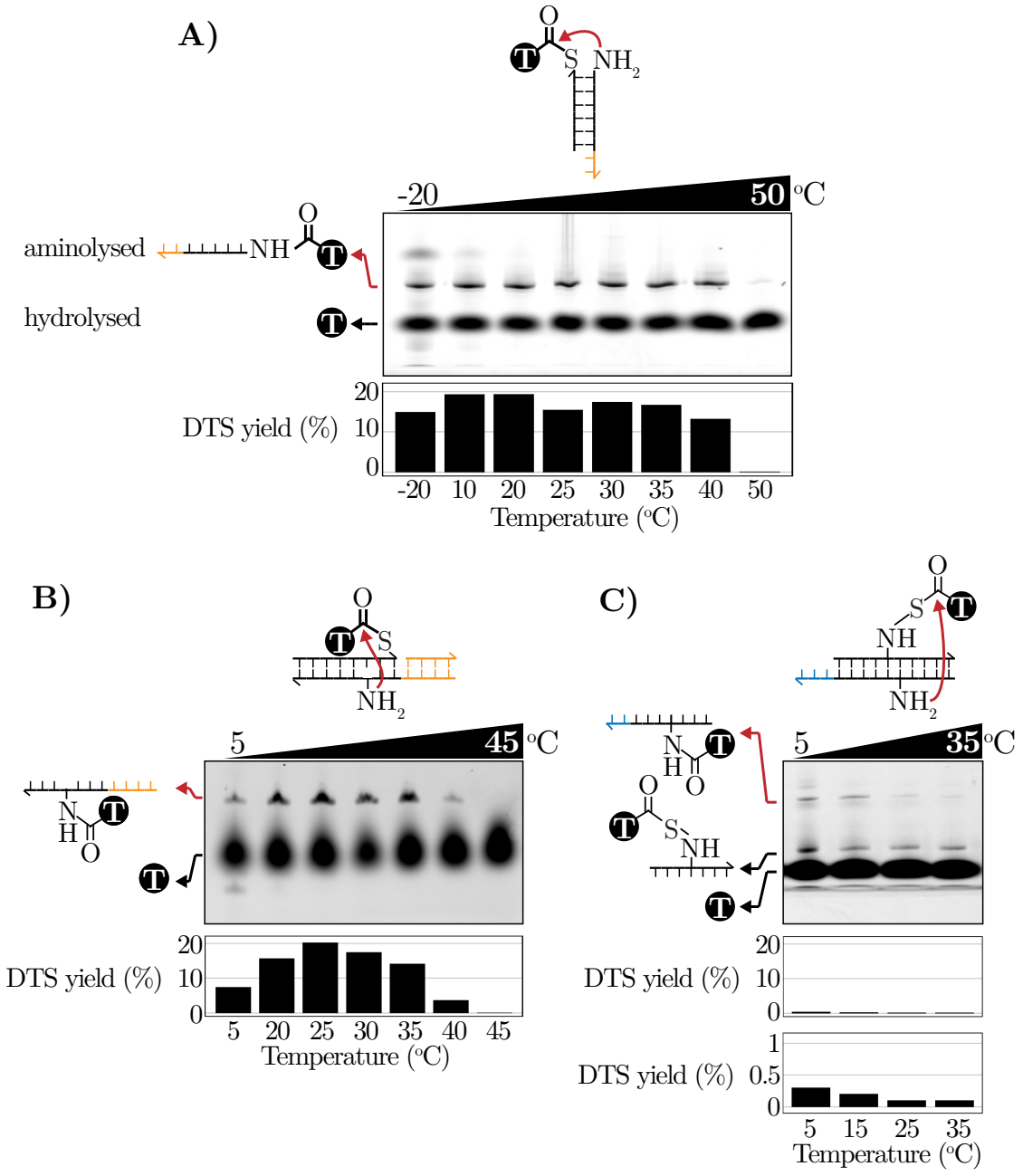


Figure 5.2: A range of temperatures were tested using three different DTS architectures (pH 11, 24 hours). All gels show 25% denaturing PAGE (TAMRA channel), with the DTS product highlighted by a red arrow. **A)** DTS in an end-of-helix architecture with a duplex of melting temperature (T_m= 66°C). **B)** DTS with a terminal thioester and internal, abasic amine modification ($n = -2$). **C)** DTS with two internal, nucleobasic modifications ($n = 1$). Method in Appendix 7.1.6, 7.1.10, and 7.1.15.

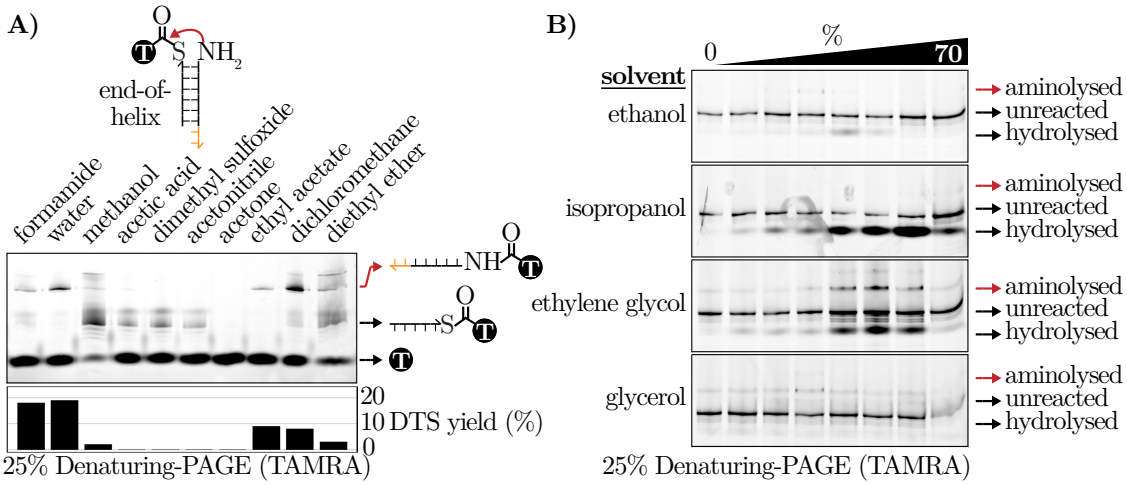


Figure 5.3: The effect of mixed solvents on DTS. All gels show 25% denaturing PAGE (TAMRA channel), with the DTS product highlighted by a red arrow. **A)** Various solvents were added at 20% v/v to an end-of-helix DTS reaction (pH 11, 24 hours, RT). **B)** The percentage of polar, protic solvents was varied to test any effect on DTS yield (pH 10, 24 hours, 4°C). Method in Appendix 7.1.6, 7.1.10, and 7.1.15.

mixed solvents and additives including DMF, THF, DIPEA and imidazole [133]). In line with this, a collaborator recently tested DTS in one of the few non-aqueous solvents in which DNA self-assembles effectively [221]. Though adapters degraded less rapidly DTS yield was also reduced [S. Núñez-Pertíñez, personal communication].

Cations. Finally, I hypothesised that the addition of charged cations might screen the negative charge of the DNA duplex, and thus decrease the local concentration of ionised water (both hydroxide and hydronium ions). I explored a range of chloride salts at different concentrations (from 1nM to 1mM of Mg^{2+}, Mn^{2+}, Al^{3+}), but observed no effect on hydrolysis or DTS (Figure 5.4). This is in contrast to a previous report by Bruick et al. [199] that spermidine (polyamine^{3+}) increased yield during over-the-nick DTS. I replicated their experiment in the same conditions but with an end-of-helix rather than their over-the-nick DTS architecture. I found that the addition of 2mM spermidine simply increased the reaction pH from 8.8 to 10.2, which may explain their observation of increased hydrolysis and DTS in the presence of spermidine.

In summary, a range of simple changes to the bulk environment of DTS reactions has been explored. DTS yield can be maximised with pH and temperature, but in

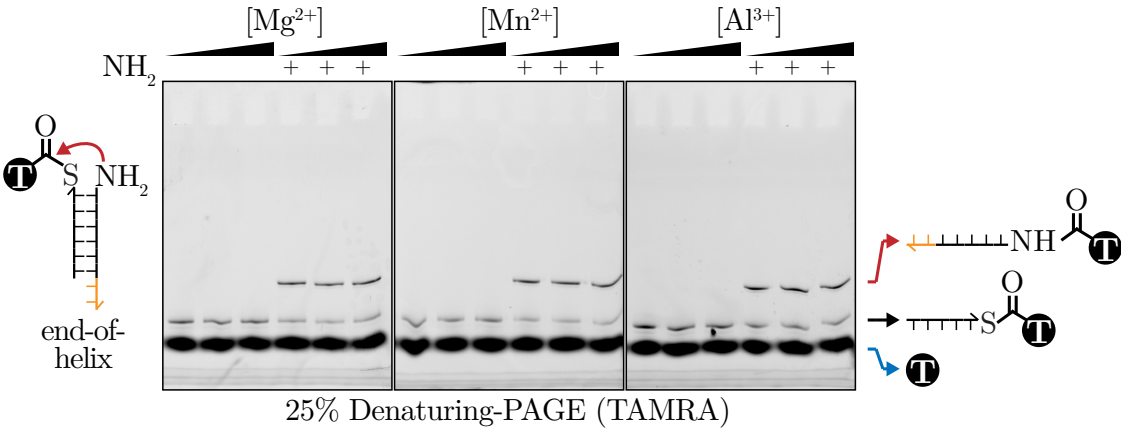


Figure 5.4: The effect of cations on DTS. Chloride salts ($MgCl_2$, $MnCl_2$, $AlCl_3$) were added to the DTS reaction to test whether cations might interact with DNA and affect thioester hydrolysis. A range of concentrations were tested (increasing from left to right 1nM, 1µM, 1mM). The presence and absence of a terminal amine modification would distinguish any effect on thioester hydrolysis (blue arrow) and DTS aminolysis (red arrow). Method in Appendix 7.1.6, 7.1.10, and 7.1.15.

these conditions both hydrolysis and aminolysis of the thioester were promoted.

5.2.2 Local environment

In this section I explore the effect of changes to the local environment of DTS reactions on adapter degradation.

DTS architectures. It has been consistently reported that DTS yield is lower in the over-the-nick architecture than the end-of-helix architecture [133, 222]. This is puzzling, as naively one would anticipate higher yields during DTS when the reactive groups are held closer together (for estimated distances in DTS architectures, see Figure 4.3). One potential explanation is that the DNA architecture affects the stability of labile linkers, perhaps due to a difference in charge density, local pH, or side reactions with the DNA template (e.g., 3'- and 5'-hydroxyl, Figure 4.5). To test this idea I measured the rate of hydrolysis of a thioester-TAMRA labile linker in a range of DNA architectures. I injected each sample into HPLC at various time points (Figure 5.5A) to separate the adapter-3'-S-O-TAMRA and the free TAMRA produced by hydrolysis, and quantify their relative abundance by TAMRA fluorescence. I observed no effect of simple changes in DNA secondary structure on the rate of hydrolysis (Figure 5.5B). As DTS yield depends on the relative rates of

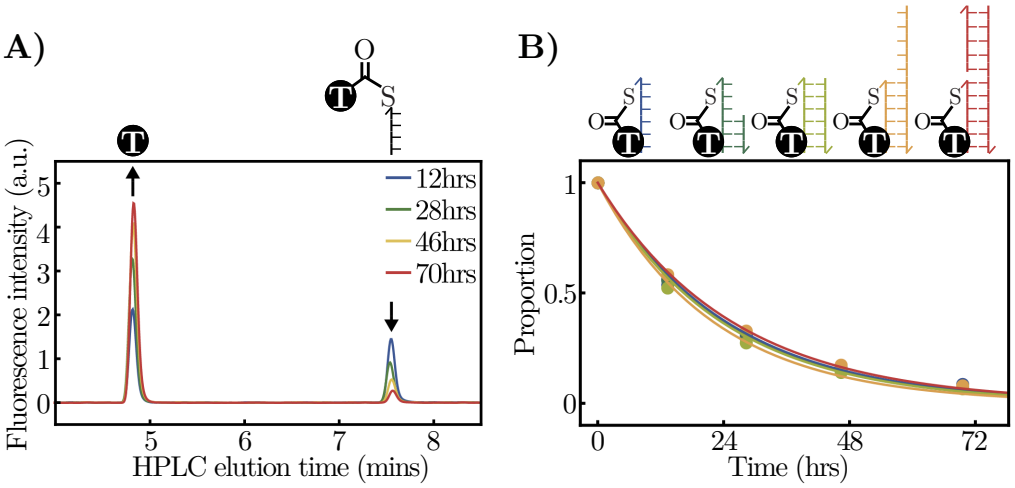


Figure 5.5: DNA secondary structure and thioester hydrolysis. A) An example HPLC trace after multiple injections of the same sample (in this case, simply the donor adapter DNA-S-CO-(T)). The fluorescence intensity of (T) was measured with Ex:530nm and Em:600nm. The peaks were identified relative to samples containing just (T) or DNA. **B)** By measuring the area under both peaks and calculating the proportion of DNA-S-CO-(T), we can compare the rate of hydrolysis in various DNA architectures. Data points are indicated with circles, whereas lines indicate fits to e^{-kt} (R^2>0.998), with k between 1.08 and $1.25\times10^{-5}s^{-1}$ for all samples. Method in Appendix 7.1.16.

transfer and degradation reactions, this result implies that the reported difference in DTS yield between end-of-helix and over-the-nick architectures must be a result of changes to the rate of aminolysis rather than the rate of hydrolysis.

Linker lengths. One explanation of the difference in DTS yield between DNA architectures is the relationship between linker length and attachment position. If the linkers are attached very close together then short linkers would be ideal whereas long linkers might be sterically constrained. In contrast, if the linkers are positioned far apart then short linkers may not reach each other to react while long linkers might have higher yields, and very long linkers would begin to have low effective concentration (a brief model is elaborated in Appendix 7.3.2). This model would explain my results in Chapter 2 where the shortest linker had highest yield in over-the-nick architecture. Similarly, in the over-the-nick architecture (linkers attached ~0.7nm apart) it has been observed that increasing linker length reduced DTS yields [111, 133]. I tested this concept in the end-of-helix architecture where linkers are attached further apart (~1.7nm). In contrast to results from the over-the-nick

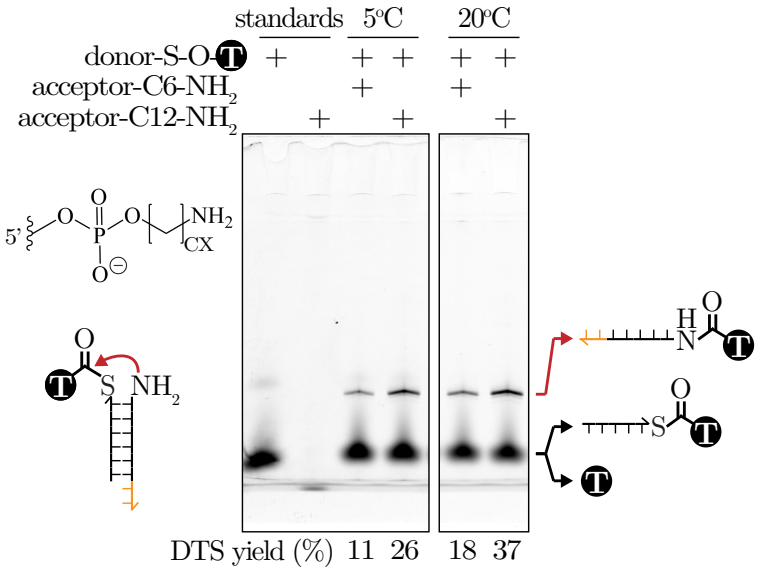


Figure 5.6: Linker length and DTS yield. Two linker lengths were compared during DTS in the end-of-helix architecture. DTS was confirmed by running samples on 25% denaturing PAGE and observing TAMRA (Ⓣ) fluorescence. DTS yield for each sample was estimated as the intensity of the upper band (acceptor-NH-CO-Ⓣ, red arrow) divided by all Ⓣ fluorescence in the sample. Method in Appendix 7.1.6, 7.1.10, and 7.1.15.

architecture [133], and my work in Chapter 2 with 3'-esters, I found that increasing linker length from C6 to C12 increased DTS yield in the end-of-helix architecture (Figure 5.6). This indicated that optimising linker length increased yield through the effect on the rate of aminolysis. The optimal linker length may be unique to each DTS architecture, as they have different distances between linker attachment points.

Colocalised groups. I speculated that colocalising a chemical group with the labile linker might affect the rate of adapter degradation. It is plausible that colocalised groups could affect degradation, as it has been shown previously that colocalising a histidine near a thioester increased hydrolysis [133]. I tested two modifications colocalised with a thioester-TAMRA: 1) a cholesterol modification that might create a local hydrophobic environment, and, 2) a short PEG polymer (15 ethylene glycol repeats) that might exclude water from the volume around the ester. These groups were colocalised in an end-of-helix architecture and the rate of hydrolysis measured via HPLC as before. I observed no effect of these modifications on the rate of hydrolysis of the thioester (Figure 5.7).

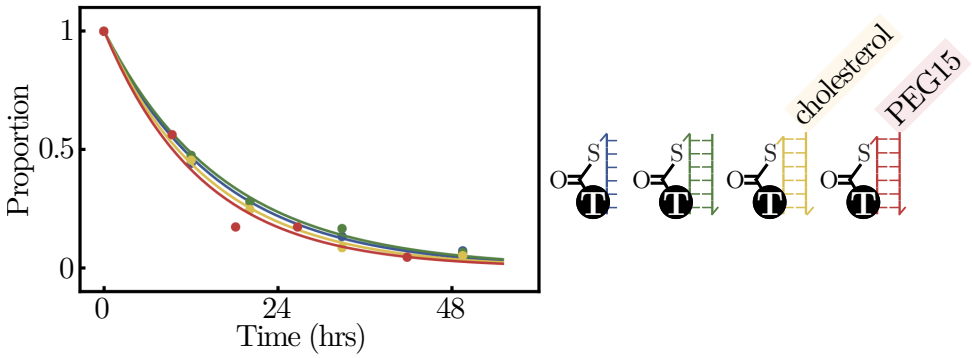


Figure 5.7: Colocalised groups and thioester hydrolysis. The rate of thioester hydrolysis was measured by HPLC as before with two DNA modifications (cholesterol and PEG15) colocalised on an end-of-helix architecture with the thioester-(T). Data points are indicated with circles, whereas lines indicate fits to e^{-kt} (R^2>0.98), with k between 1.7 and $2 \times 10^{-5}s^{-1}$ for all samples. Method in Appendix 7.1.16.

In summary, various simple changes to the local DTS environment have been explored. The rate of hydrolysis is independent of DNA secondary structure. In contrast, the rate of aminolysis appears to depend on both linker position and length, implying that DTS yield can be maximised by careful optimisation of these two parameters.

5.2.3 Limiting thioester hydrolysis with abasic modifications

Previously I explored internal modifications rather than terminal modifications to DNA strands during DTS (Chapter 4). I observed that an internal abasic amine modification had both decreased hydrolysis and aminolysis in a DTS reaction (Figure 4.7). The most obvious interaction that might explain this observation is the potential for the TAMRA fluorophore to stack into the nearby abasic site. This might position the thioester within the hydrophobic core of the DNA duplex, protecting it from hydrolysis by water. I tested this idea by measuring the rate of thioester hydrolysis with either an abasic amine, an abasic ribose modification or an unmodified strand containing only the internal thioester labile linker (Figure 5.8). The thioester was protected from hydrolysis by either abasic modification. This suggests that the internal abasic site is the necessary component and not the internal hexylamine. The rate of thioester hydrolysis is reduced by over 100-fold in the presence of either colocalised abasic site. This is significant - if extrapolated

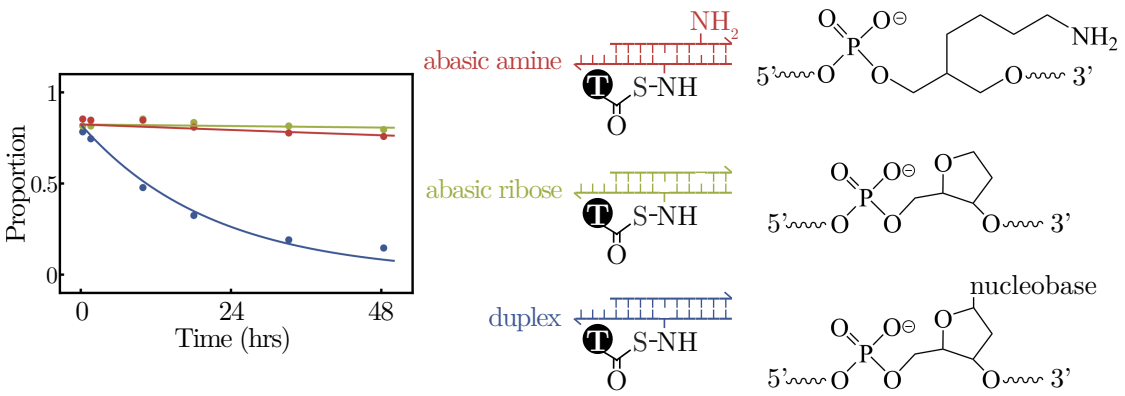


Figure 5.8: Degradation rates were measured by HPLC timecourse as before, using two abasic modifications positioned 3bp in the 5' direction from an internal thioester-(T) modification ($n = -3$, 5°C, pH 11). Data points are indicated with circles, whereas lines indicate fits to $0.82e^{-kt}$ (R^2>0.98), with k ~$1.3 \times 10^{-5}s^{-1}$ for a standard duplex (blue) and $k \leqslant 1 \times 10^{-7}s^{-1}$ with abasic modifications (red and green). Method in Appendix 7.1.16.

to the synthesis of a 10mer it is predicted to increase yield from <0.3% to ~30% (see model in Appendix 7.3.3).

To explore the utility of protection, I simulated DTS in a multistep cascade where donor adapters used later in polymer synthesis typically degrade in DTS conditions before they can be used by an artificial programmable polymerase. I incubated the donor adapter containing an internal, abasic thioester modification at pH 11 overnight (~15 hours) in either a protected or unprotected duplex, and the following day added an acceptor adapter to trigger DTS. The acceptor adapter had a toehold containing an extra 10bp with the donor adapter, triggering the exchange of strands by toehold mediated strand displacement (TMSD) to bring the thioester and amine into close proximity for DTS (Figure 5.9A). Without protection the DTS yield was 4%. In contrast, with protection DTS yield was 22%, which is the same as when the DTS adapters were freshly mixed (Figure 4.7). This demonstrated the principle that donor adapters can be protected from degradation until needed for a DTS reaction.

Previously I had observed the highest yield of DTS with the abasic amine modification positioned at $1 \leqslant n \leqslant 2$, while protection from hydrolysis occurred with $-5 \leqslant n \leqslant -2$ (Figure 4.7). It is unclear whether hydrolysis protection and aminolysis could be occurring simultaneously, in which case the high DTS yield might be

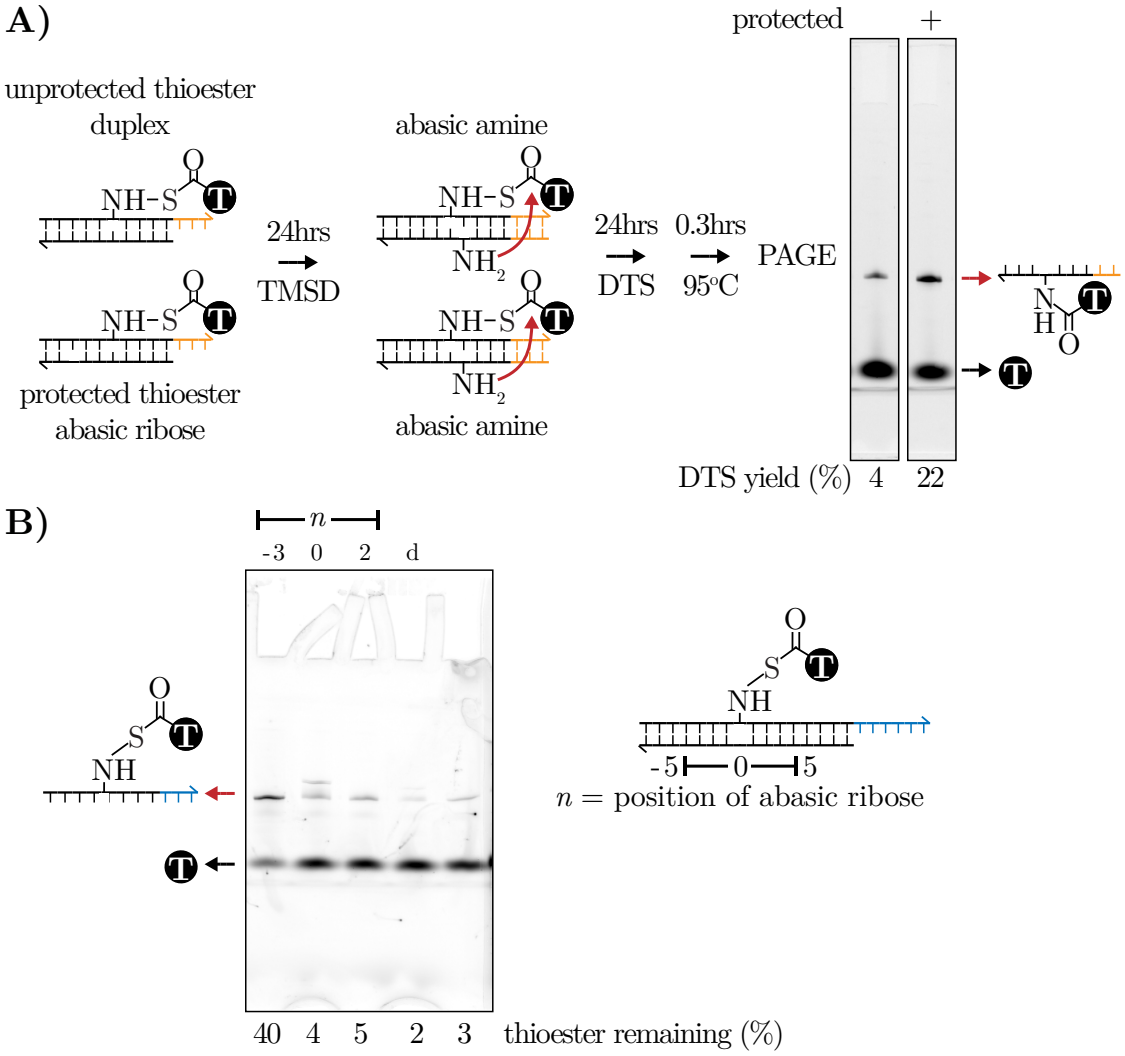


Figure 5.9: **A)** The utility of adapter protection was demonstrated by incubating donor adapters overnight (~15 hours) with either a protected abasic site or unprotected duplex (d). DTS was triggered by adding to the reaction an acceptor adapter with an internal amine modification. Strands exchanged via toehold mediated strand displacement (TMSD) to trigger DTS. The sample was boiled prior to 20% denaturing PAGE to distinguish remaining unhydrolysed donor adapter from the DTS product acceptor-NH-Ⓣ (red arrow). **B)** The position of the abasic ribose modification was varied to test if hydrolysis protection was position-dependent. The thioester donor adapter (red arrow) was hydrolysed least with the abasic ribose positioned at $n = -3$, across the minor groove.Samples were incubated for 24 hours in DTS conditions at pH 11 and 20°C (d = unprotected duplex). Method in Appendix 7.1.6, 7.1.10, and 7.1.15.

explained by protection during DTS. Therefore I varied the position of the abasic ribose (Figure 5.9B). Interestingly, protection of the thioester appears unique to specific positions of the abasic site (-5$\leqslant n \leqslant$-2). If the abasic site is opposite or in the 3' direction from the thioester-TAMRA then no protection from hydrolysis occurred.

In order to be useful during polymer synthesis it must be possible to protect many different types of monomer from degradation. If protection is caused by TAMRA stacking within the abasic site, then monomers that do not stack would not be protected. To test this idea I substituted biotin for TAMRA, which still provided a simple assay on native PAGE through the biotin-streptavidin interaction.

I synthesised donor adapters for DTS as before, substituting NHS-biotin for NHS-TAMRA during synthesis (Figure 5.10A). I confirmed that an interaction between biotin and streptavidin could be observed on native PAGE by incubating the donor adapter with thioester-biotin modification with streptavidin, either before or after boiling to hydrolyse the thioester (Figure 5.10B). In this case a slow moving band was seen, which indicated that the donor adaptor bearing a biotin modification had been synthesised correctly, that the streptavidin interaction could be used to assay the position of the biotin modification before or after DTS, and that the thioester could be hydrolysed. I next tested protection of the thioester-biotin on the donor adapter using nearby abasic sites. After incubating the donor adapter (thioester-biotin) with either a duplex as a control, an abasic ribose for protection or an abasic hexylamine for DTS, I split each sample and deliberately hydrolysed half by boiling for 20 minutes. I then added streptavidin and incubated at RT for another 20 minutes before running native PAGE. In this case if DTS was successful it would produce the acceptor adapter-NH-biotin-streptavidin. I expected to see a slow moving band on native PAGE, unaffected by boiling the sample prior to streptavidin incubation. In contrast, if the thioester-biotin was protected from hydrolysis by the abasic site then I expected a slow-moving band before boiling but not after. I saw no protection of the thioester-biotin from hydrolysis when colocalised with abasic sites (Figure 5.10C). I scanned the full range of positions for the abasic amine modification and found no hydrolysis protection at any position (Appendix 7.2.8).

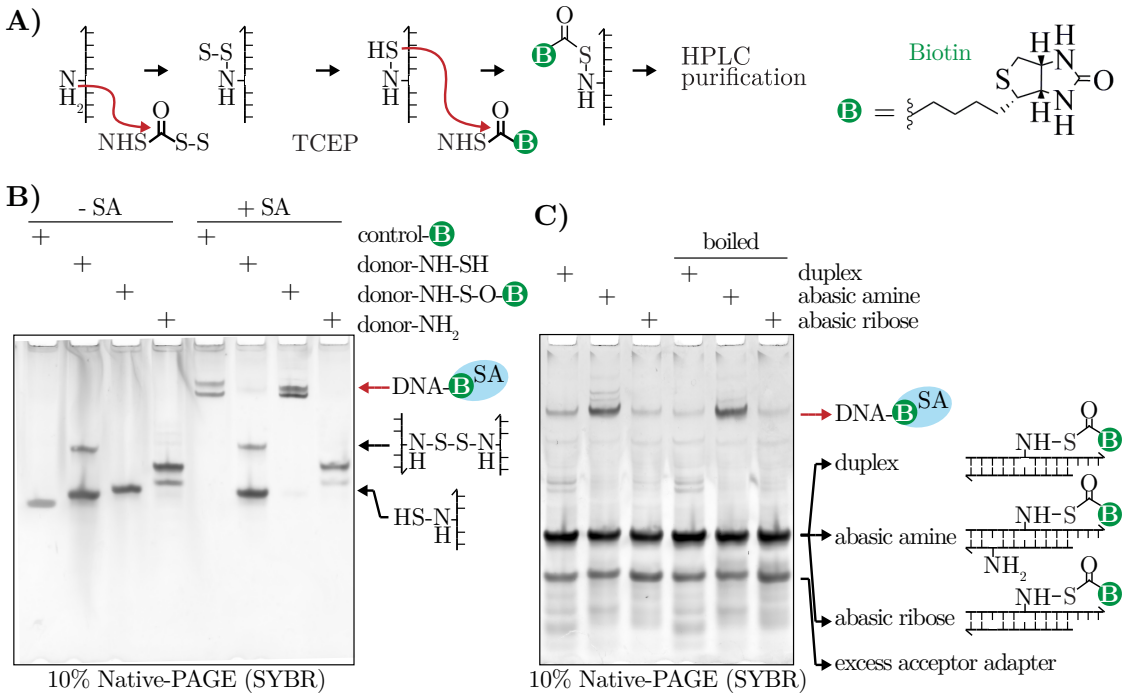


Figure 5.10: Exploring hydrolysis protection of biotin. **A)** Thioester-biotin donor adapters were built upon an internal abasic hexylamine modification. A disulfide was added using SPDP (NHS-SS), reduced to a thiol, and incubated with NHS-biotin before HPLC purification. **B)** PAGE of the synthesised donor adapter to confirm purity and streptavidin (SA) binding. A control strand had a covalent biotin modification. Slowly migrating bands are visible (red arrow), indicating formation of the DNA-biotin-straptavidin complex. Two bands appear after boiling the thioester-biotin donor adapters to produce the hydrolysed thiol, presumably due to formation of a disulfide-bonded dimer. **C)** The donor adapter with thioester-biotin modification was incubated at pH 11 in three DTS architectures. Each sample was then split into two and then half was boiled to remove any remaining donor adapter. All samples were mixed with 10× excess SA before running on PAGE. If the abasic site offered protection from hydrolysis during the incubation at pH 11, then the slow moving band (red arrow) would be stronger before boiling than after. Method in Appendix 7.1.4, and 7.1.7, and 7.1.10.

Protection of the thioester from hydrolysis appeared to be unique to the TAMRA monomer, as expected if protection is caused by TAMRA stacking in the abasic site. This means that in order to protect from degradation the various types of monomers used in polymer synthesis a new mechanism of adapter protection must be devised.

5.3 DISCUSSION

In this chapter I have summarised an array of experiments characterising DTS conditions. In analogy with the mechanisms used by the ribosome to protect

tRNA adapters from hydrolysis, I have searched for ways to protect unstable donor adapters from degradation. This is a necessary step before artificial programmable polymerases can synthesise longer polymers. In particular it is necessary to enable insertion polymerisation (Chapter 4), where each transfer reaction occurs in an identical reaction environment regardless of polymer length, but the polymer is attached to adapters via a labile linker that is prone to degradation. It is also useful to enable feedback, as the implementation I developed in Chapter 2 could fail if donor adapters modified with 3'-esters hydrolysed to produce a 3'-hydroxyl triggering motor translocation.

I have not found any convenient methods of changing the bulk environment to reduce adapter degradation. However, as reported by others [133], DTS yield can be optimised by pH, temperature and linker length. I demonstrated that DNA secondary structure does not affect the rate of hydrolysis. This implies that observed differences in DTS yield between DNA architectures are due to changes in the rate of aminolysis. Interestingly, there is an opposite effect of increasing linker length is opposite in the end-of-helix and over-the-nick architectures. Here I demonstrated that longer linkers have higher DTS yield in the end-of-helix architecture, which is in contrast to the observation of McKee et al. with the same transfer reaction in an over-the-nick architecture [133]. I have speculated that this is due to the interplay between linker length and position, which also explains the observation in Chapter 2 that DTS was only possible with very short linkers. If this is true, then the optimal linker length for DTS depends on the DTS architecture and the proximity of the linker attachment points (Appendix 7.3.2).

Changing the local environment through the addition of an abasic site near a thioester-TAMRA significantly reduced hydrolysis. There is some evidence that this is due to stacking of the fluorophore within the abasic site - substituting an abasic ribose for an abasic amine modification still enabled protection of the thioester, while substitution of TAMRA with biotin abolished protection. Indeed, fluorophores resembling TAMRA have been explored for their capacity to stack within abasic sites as a strategy in DNA sequencing [223–225]. It would be interesting future

work to determine if the protection from hydrolysis observed here is specific to the sequence around the abasic site. It would also be informative to explore the structure of the protected complex by NMR or crystallisation.

I demonstrated that a thioester-biotin was not protected from hydrolysis. Ideally, new linkers should be designed to protect arbitrary monomers from hydrolysis. For example, it would be interesting to explore linkers that invert the arrangement of labile linker, stacking motif and monomer (Figure 5.11), as this might enable protection of a diverse array of monomers and transfer chemistries. Furthermore, stacking is only one of many kinds of molecular interaction. Other interactions relying on hydrogen bonding or charge might be used to position a water-sensitive group within a hydrophobic environment.

A mechanism for limiting adapter degradation has been discovered. It is easy to imagine how the use of abasic sites for ester protection could be combined with insertion polymerisation and the symmetric DNA architectures developed in Chapter 4. Indeed, a mechanism for protection is necessary for the benefit of insertion polymerisation to be realised. However, as it relies on a long linker opposite an abasic site, it is unclear how it could be incorporated with the 3'-ester adapters developed in Chapter 2. Perhaps only the general concept of using enzymes to coordinate transfer reactions and translocation may be useful if internally-modified adapters are used because they offer protection from degradation.

This chapter has revealed an interesting interplay of the parameters affecting templated transfer. Monomer type, linker length and position affect the forward reaction of transfer, but also the potential for side-reactions degradation and protection from degradation. As suggested in Figure 5.11, further progress towards artificial programmable polymerases will require re-design of modular linkers taking advantage of the lessons learned here.

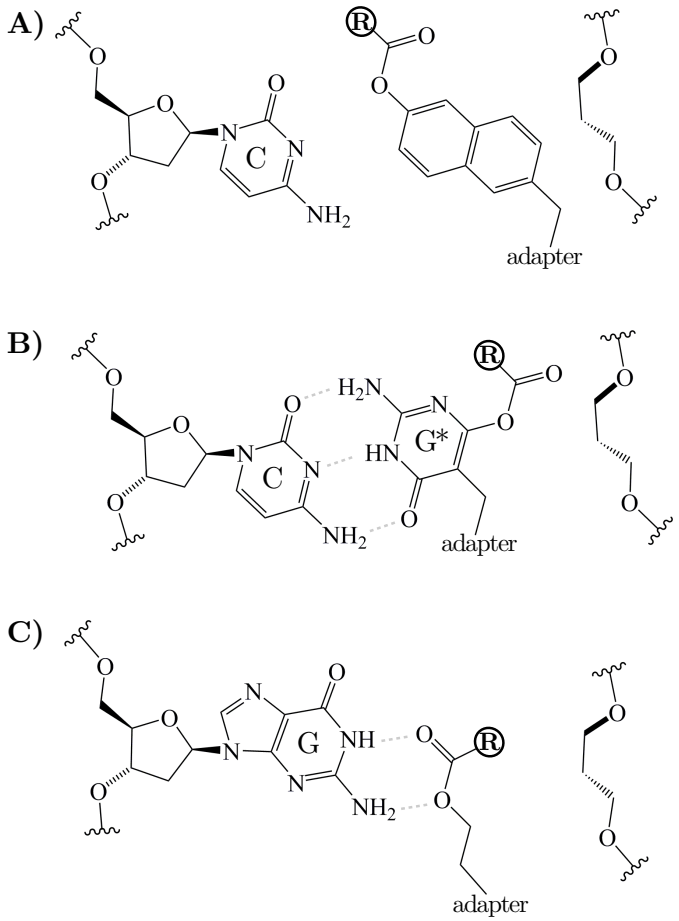


Figure 5.11: Proposed future work with protected linkers. A) A stacking motif might position a water-sensitive labile linker such as an ester within abasic sites, in order to protect a monomer (R) from hydrolysing from a donor adapter. **B)** A nucleobasic mimic (G*) might both stack and hydrogen-bond within abasic sites to precisely position the labile linker. **C)** The labile linker itself may be capable of hydrogen bonding with specific nucleobases opposite abasic sites.

6

CONCLUSION

Here I reflect on this thesis within the context of the field of artificial programmable polymerases, before discussing future research in DNA nanotechnology and beyond.

6.1 REFLECTIONS

At the heart of this thesis is the question – if it is possible to design and build almost arbitrary DNA nanostructures and primitive DNA nanomachines, is it possible to make molecular machinery that mimics the RNA ribosome? In the process of exploring this question, I offer a new definition of *polymerases* as enzymes that catalyse polymer synthesis, and *programmable polymerases* that synthesise polymer sequences according to the information in molecular templates. In Chapter 1 I outlined the requirements of artificial programmable polymerases, identified challenges for the field, and reviewed approaches spanning molecular

biology, supramolecular chemistry and DNA nanotechnology. I concluded that three challenges − feedback, insertion and protection − must be addressed to enable the synthesis of polymers of useful lengths by artificial programmable polymerases. Each chapter has suggested solutions to these challenges independently, though there have been interesting, serendipitous overlaps (e.g., an abasic architecture characterised for insertion polymerisation unexpectedly protected adapters from hydrolysis).

The challenges remain significant. Work in the field since I began my thesis has continued to exhibit these design flaws and consequently suffered length limitations during polymer synthesis by autonomous nanomachines. For example, in 2018 Bo et al. [72] made the latest iteration of the artificial programmable polymerase based on a supramolecular rotaxane sliding on an axle. Their machine synthesised a peptide catalyst with a distribution of lengths between 3-11 monomers with a mean of 6.5. However, the lack of insertion polymerisation or protection from degradation limited peptide length and caused unprogrammed background synthesis of peptides with random sequences. If these challenges can be resolved it is likely that the various applications of artificial programmable polymerases will begin to be explored.

DNA recorded synthesis is an interesting precedent in this regard. In the 1990's it was proposed that DNA sequences could encode the identity of molecules in a combinatorial library [15]. During the 2000's this was developed by academic groups and companies originating from university research [126, 201], and today large pharmaceutical companies have synthesised the largest DNA recorded combinatorial libraries [16]. It is not inconceivable that DNA programmed synthesis of small molecules, oligomers and polymers will also find similar applications once these design challenges have been resolved.

Chapter 2 introduced a new design concept for artificial programmable polymerases − coordinating the processes of chemical synthesis and molecular motion using enzymes to recognise the product of templated transfer reactions and trigger the next step of polymer synthesis. This concept was implemented using a 3'-ester transfer chemistry that produces natural substrates for DNA-acting enzymes rather than relatively longer, flexible synthetic linkers. The reliance on short linkers on

acceptor adapters for DTS is an important limitation, as the growth of polymer during multistep synthesis may make subsequent DTS reaction impossible. This may undermine the utility of the 3'-ester transfer chemistry I have developed. However, in future, methods may be developed for insertion polymerisation (as discussed in Chapter 4) that would maintain an identical reaction environment using short linkers at each step of polymer synthesis.

The 3'-ester transfer chemistry is unique in the DTS literature. The closest analogue was the very first demonstration of over-the-nick transfer of a peptide from a 5'-thioester to a 3'-amine in 1996 [199]. This also created a transition state resembling the phosphate backbone. However, the 3'-ester transfer chemistry is almost identical to the tRNA donor adapters used by the natural ribosome. The difference is the 2'-hydroxyl catalysing peptide bond formation in the ribosome [43, 44]. It would be worthwhile exploring whether the 2'-hydroxyl can catalyse DTS (or 'RNA templated synthesis' might be more appropriate), though it is reasonable to expect that this might increase hydrolysis as well as aminolysis, thus limiting transfer yield through adapter degradation. Due to the similarity with natural tRNAs it may be possible to use the Ef-Tu protein that protects tRNAs from hydrolysis to protect these artificial donor adapters. Ef-Tu has been shown to bind to highly truncated tRNAs consisting solely of a dsRNA duplex and offer some protection from ester hydrolysis [197]. Ef-Tu promiscuously binds and protects tRNAs with all canonical amino acids, while the affinity for tRNAs bearing non-canonical amino acids may be maximised by tuning tRNA sequences [198]. This means is may be possible to use Ef-Tu to protect DNA adapters from degradation during polymer synthesis by DNA-based artificial programmable polymerases. However, with this approach artificial programmable polymerases will increasingly rely on natural proteins such as Ef-Tu. The result may be increasing constraints on monomer diversity.

Finally, the most relevant future work will implement feedback in a nanomachine performing multiple steps of polymer synthesis. While it is trivial to simulate the effect of feedback on polymer synthesis (Appendix 7.3.3), in practice it will be difficult to design an experiment that isolates the effect of feedback and quantifies

any change in polymer yield or length. For example, one might naively envisage an experiment including or excluding the polymerase that recognises and acts on the product of the DTS reaction. This is insufficient, as excluding the polymerase would entirely prevent adapter transfer and multistep syntheses. Though not an ideal comparison, one experiment could compare polymer fidelity with two different types of artificial programmable polymerases with the motor running as fast as possible. In the case of the nickase-driven artificial programmable polymerase [98], this would result in a polymerase with many skips, whereas with feedback a polymerase-driven walker should have fewer skips.

Chapter 3 built on this work by exploring the use of enzymes to catalyse molecular motion and multistep synthesis. In the process, I developed a new mechanism to drive DNA nanomachinery − polymerase TMSD − with a stable small molecule fuel (dNTPs) that may have applications beyond the present focus on artificial programmable polymerases. While a diverse array of DNA-acting enzymes have been used to catalyse motion by DNA walkers, the use of polymerases is relatively rare. Relevant research by others include the use of polymerases to drive the relative rotation of concatenated DNA rings [226], the assembly of 1D tiles [227], and their use in conjunction with nucleases to create complex chemical reaction networks [89]. To my knowledge, strand-displacing polymerases have not been used to synthesise toeholds for the exchange of strands by TMSD, nor to drive DNA walkers.

To further improve the rational design of such machinery, it would be ideal to exhaustively characterise the conditions for polymerase TMSD and determine transfer kinetics in various DNA architectures − as has been done for ordinary TMSD [96]. I observed differences in the rates of polymerase TMSD when the legs of a DNA walker were colocalised (Figure 3.9C). I presented the hypothesis the track colocalising legs increased the rate of cargo transfer between legs, which changed the DNA architecture from which the polymerase has its slowest, rate-limiting step of substrate release [182, 183]. This hypothesis needs to be tested. It may also be useful to understand how sequence and DNA secondary structure affect the rate of polymerase TMSD. Finally, as I mentioned previously, the most relevant future

work would be to integrate the developments from Chapter 2 and Chapter 3 in a prototype performing multistep polymer synthesis with feedback.

Chapter 4 contrasted extension and insertion polymerisation, which either create a changing and ever-distant reaction environment or an identical reaction environment during polymer synthesis. This concept was developed by previous workers in the Turberfield Group [115]. Multistep insertion polymerisation necessitates both symmetric transfer architectures and adapter protection. I created a range of symmetric DTS architectures, initially using short sticky ends called overhangs to colocalise reactive groups and finally using internal modifications. The main contributions of this chapter included a new method for generating 3'-3' or 5'-5' conjugated DNA strands, a DTS architecture featuring internal modifications with higher transfer yield compared to the canonical end-of-helix architecture, and finally, the discovery that abasic sites can protect a thioester-TAMRA labile linker from hydrolysis.

DTS in the overhang architecture was not observed, presumably due to the instability of short sticky ends. For this reason, I moved to a DTS architecture with longer duplexes that maintained the property of symmetry using internal modifications. DTS with these adapters typically necessitates an extra synthesis step to add modifications to internal amine modifications, which in my experience increased reactant costs and reduced yields during synthesis of donor adapters. However, no attempt has been made here to optimise synthesis of adapters. It is quite likely that the use of nucleoside phosphoramidites with internal disulfide modifications and the synthesis of modifications during DNA synthesis would dramatically reduce this synthesis cost.

This is similar to previous work where conjugation reactions rather than transfer reactions were performed between internal modified DNA strands. In this system the effective molarity was measured to be M_{eff} ~25M [122]. Ideally, future work would compare the M_{eff} of various DTS architectures (end-of-helix, across-the-nick, junction, internal abasic, etc.) in order to fully understand the system and to predict polymer fidelity from the rate of off-target background transfer reactions.

In other work, the yoctoreactor was a Holliday junction that colocalised internal modifications for an externally controlled two-step transfer reaction that first coupled then cleaved strands [201]. The design developed in this thesis is simpler, with fewer components, closer proximity of reactive groups and a single-step transfer reaction, which is necessary for an autonomous artificial programmable polymerase. A clear next step for future work would be demonstrating multistep synthesis with these internally modified adapters.

Chapter 5 explored the perennial problem of adapter protection by attempting to change the bulk solvent or local environment around labile linkers. Like previous work, I found that DTS yield could be optimised with pH, temperature and linker length [111, 133, 188, 222]. Unlike previous work, I found an abasic site could reduce degradation of a thioester-TAMRA donor adapter. Importantly this mechanism of protection could be controlled by TMSD to introduce or displace the abasic site and to trigger DTS reactions. However, based on substitution experiments with a biotin monomer, this protective strategy appeared unique to the TAMRA monomer. Future work should aim for protection of a diverse range of monomers. To progress towards this goal it would be useful to substitute TAMRA with monomers that are hydrophobic, pi-stacking or charged to determine the extent to which these interactions are driving hydrolysis protection in this system. Indeed, there are examples in the DNA sequencing literature of small molecules similar to TAMRA that have been designed to stack in abasic sites [223–225]. Furthermore, it would be interesting to explore new linkers with alternative transfer chemistries designed to position cleavable linkers rather than the monomer within the hydrophobic core of the DNA duplex (e.g., Figure 5.11).

A complementary strategy to protection is the colocalisation of catalysts to increase transfer yields and decrease the reliance on highly reactive cleavable linkers. A collaborator Samuel Núñez-Pertíñez (O'Reilly Group, University of Birmingham) has been working with phenol-esters, amines and colocalised thiols to exploit native chemical ligation [228]. This work shows that phenol-esters are relatively unreactive, and thus are stable to degradation by hydrolysis and poor donor adapters for DTS

by aminolysis. The primary amine active site of an alanine cannot attack the phenol ester in detectable yield, while cysteine with a thiol can [S. Núñez-Pertíñez, personal communication]. If this strategy can be generalised to colocalised thiols rather than specific active sites like cysteine, then the use of generally less reactive transfer chemistries could help limit adapter degradation. Furthermore, when combined with other proposals in this thesis, then great improvements in length and yield of polymers should be possible by autonomous artificial programmable polymerases.

6.2 PROSPECTS

The underlying strategy in this thesis is to control reactions between components by colocalisation. Programmability by proximity has broader applications than just chemical reactions. It includes reactions between proteins [229, 230], nanoparticles [231], vesicles [232], and even cells [233]. Just as the molecular templates described here can program the set of reactions that synthesise a polymer, at larger scales an analogous sequence-controlled 1D polymer or 2D surface can program the reactions between different cell types to create synthetic tissues or multicellular communities [233, 234]. While the cost and stability of these DNA-programmed materials are still being explored and resolved [235, 236], this fine control over the structure of matter spanning nanometers to micrometers is revolutionary. DNA is currently the easiest and cheapest medium for programming molecular interactions. I anticipate that this will change as the solid-phase synthesis of self-assembling non-natural polymers becomes more accessible [237, 238], as these polymers may be more stable than DNA and as simpler design rules and cheaper building blocks are developed.

In this thesis, I have mentioned but not developed the potential for genetic regulation of artificial programmable polymerases. Once polymers of useful lengths can be synthesised, it will be interesting to create systems in which the production of polymers is contingent upon the environments that the artificial programmable polymerases encounter or create. This has been demonstrated in various other applications of DNA nanomachinery [83–88]. In the context of artificial programmable polymerases gene regulation might enable, as examples, 1) medical

systems permitting *in situ* synthesis of drugs in response to disease conditions [22], 2) chemical systems that take inspiration from self-regulating metabolic systems [23, 24] to control the relative concentrations of catalysts and reactants during production, or 3) computational systems in which molecular memories could be stored, read, written and trained to respond to diverse cues [25, 26]. These are examples in which the flow of information from the environment affects polymer synthesis by artificial programmable polymerases. There will also be the opportunity for systems in which an external flow of energy drives metabolic systems (e.g, using light to produce the chemical fuel that powers artificial programmable polymerases [239]) or embryogenetic systems (e.g., using chemical energy to create concentration gradients that drive pattern formation [240]).

The ribosome synthesises proteins with great length, yield, speed and fidelity. It outperforms highly optimised methods for solid-phase peptide synthesis. I anticipate that the ribosome is likely to always outperform solid-phase synthesis because autonomous nanomachines have the potential for error-correction during synthesis with single-molecule specificity (e.g., stalled ribosomes produce truncated proteins that are recognised and degraded [39, 241]). This has implications for all future work in synthetic biology, which is largely based on advances in solid-phase synthesis. The total chemical synthesis of genomes [242, 243] is currently feasible because synthetic DNA fragments can be reliably self-assembled into larger molecules [244]. In contrast, the total chemical synthesis of proteomes is an unreasonable challenge with current technology. However, using *in vitro* transcription and translation, programmable polymerases can be used to synthesise an arbitrary pool of proteins given the set of corresponding genes produced by total chemical synthesis. Similarly, when artificial programmable polymerases that match or outperform the ribosome are constructed, then they could synthesise mixtures of different polymer sequences in parallel by translating the corresponding genome. If this artificial proteome catalyses membrane formation, metabolism and replication, then programmable polymerases could synthesise artificial life from the molecules up.

7

APPENDICES

CONTENTS

7.1 Methods . . . 130
7.1.1 General . . . 130
7.1.2 Synthesis of DBE-biocytin . . . 130
7.1.3 Flexizyme synthesis of adapters . . . 131
7.1.4 Native PAGE . . . 132
7.1.5 Acid PAGE . . . 132
7.1.6 Denaturing PAGE . . . 132
7.1.7 Streptavidin PAGE . . . 133
7.1.8 PAGE densitometry . . . 133
7.1.9 LC-MS . . . 134
7.1.10 DTS . . . 134
7.1.11 Fluorometry . . . 137
7.1.12 Polymerase extension . . . 138
7.1.13 Nickase Nt.AlwI . . . 140
7.1.14 Templated disulfide conjugation . . . 140
7.1.15 Synthesis of thioester adapters . . . 140
7.1.16 HPLC timecourse and analysis . . . 142
7.2 Data . . . 143

7.2.1 Optimising flexizyme synthesis conditions 143
7.2.2 Reversing direction of polymerase-driven walker 144
7.2.3 Characterising the polymerase-driven walker 145
7.2.4 Time and [TCEP] during templated disulfide conjugation 146
7.2.5 Temperature during templated disulfide conjugation in end-of-helix architecture 148
7.2.6 Disulfide conjugation with 0nt and 2nt overhangs 149
7.2.7 LC-MS of thioester-TAMRA donor adapters 150
7.2.8 Position scan of DTS with thioester-biotin 151
7.3 Models 153
7.3.1 Calculating C_{eff} in DTS architectures 153
7.3.2 Linker attachment position and length 155
7.3.3 Programmable polymerase 157
7.4 Sequences 162

7.1 METHODS

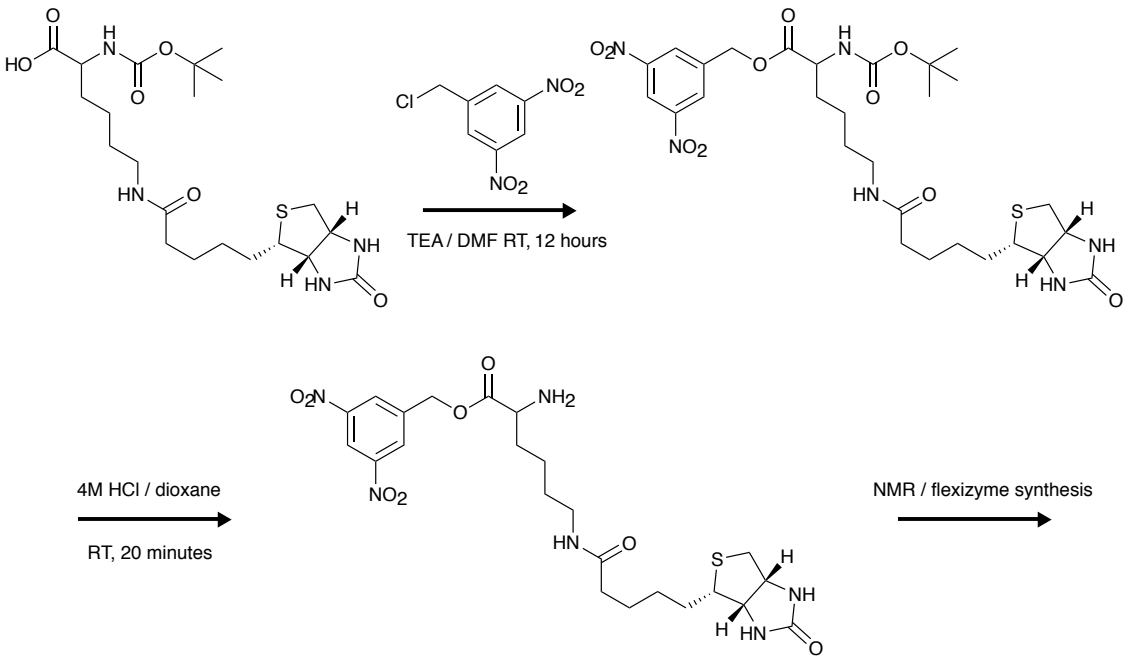
7.1.1 General

Unless otherwise stated, reagents were purchased from SIGMA. All gels were cast with a BIORAD Mini-PROTEAN Gel Kit, run with a BIORAD PowerPac Basic Power Supply and observed with an GE Amersham Typhoon FLA9500. Samples were annealed or incubated at specific temperatures using an Eppendorf Mastercycler nexus. H_2O was purified with a Milli-Q Integral Water Purification System with a 0.22 µm filter. Reactants were stored in 1.5 mL Eppendorf tubes at −20°C and reactions performed in either 1.5 mL Eppendorf tubes or 0.2 mL PCR tubes (Axygen). Centrifugation was performed with either an Eppendorf 5810R bench top ultracentrifuge or Eppendorf Minispin bench top centrifuge. TAE buffer (pH 8) was composed of 200mM Tris, 100mM acetic acid and 5mM ethylenediaminetetraacetic acid (EDTA).

7.1.2 Synthesis of DBE-biocytin

Synthesis DBE-biocytin was performed by Samuel Núñez-Pertíñez (O'Reilly Group, University of Birmingham), using a protocol adapted from [245]. 100 mg of N-Boc-biocytin (0.21 mmol, 1.2eq) and 38 mg of 3,5-dinitrobenzyl chloride (0.18mmol,

1.0eq) were dissolved in 0.1 mL DMF (peptide synthesis grade). 50 µL triethylamine (0.36mmol, 2eq) was added and the resulting solution was shaken at RT overnight (~15 hours). The solution was then diluted with 9 mL of diethyl ether and washed with 1M HCl (3×3mL), 4% NaHCO3 (3×3mL) and brine (3×3mL). The organic layer was dried over anhydrous $MgSO_4$ and concentrated under reduced pressure. The crude residue was dissolved in 2mL 4M HCl in dioxane and shaken at RT for 20 minutes. The solution was concentrated under reduced pressure and excess HCl removed by successive additions of diethyl ether and evaporation.1H/13C-NMR confirmed synthesis.



7.1.3 Flexizyme synthesis of adapters

Reactions were prepared by mixing 5 µL H_2O, 15 µL 125 µM flexizyme, and 10 µL 100 µM DNA donor adapter at RT. The sample was rapidly annealed from 95°C to 4°C over 10 minutes, before adding ice-cold 10 µL 5×flexizyme buffer (3M $MgCl_2$, 250 mM HEPES, pH 8 adjusted with KOH) and ice-cold 10 µL 100 mM DBE-monomer in DMSO. Reactions were incubated on hard-packed ice in a 4°C fridge overnight (~15 hours).

7.1.4 Native PAGE

A typical 10% 0.75 mm gel was prepared by mixing 1.8 mL 30% 29:1 acrylamide:bis-acrylamide, 2.5 mL H_2O, 1.1 mL 5×TAE, 40 µL 10% ammonium persulfate (APS), and 4 µL tetramethylethylenediamine (TEMED). Different percentage gels were prepared by varying the proportion of acrylamide and water accordingly. After removing well combs, wells were rinsed with running buffer (1×TAE) using a pipette before loading 2 µL samples typically diluted to 200 nM in loading buffer (1×TAE, 50% glycerol). Electrophoresis was typically at 150V for 20-30 minutes at RT. Gels were often observed both pre- and post-staining. Gels were stained in a mix of 2 µL $10,000\times$ SYBR Gold Nucleic acid stain (Thermofisher) and 50 mL H_2O, stirring for 5-15 minutes, before washing with H_2O and observing the gel.

7.1.5 Acid PAGE

A typical 20% 0.75 mm gel was prepared by mixing 3.7 mL 30% 29:1 acrylamide:bis-acrylamide, 1.4 mL H_2O, 0.1 mL 50×TAE, 0.2 mL 3M NaAcOH (pH 5.2 adjusted with HCl), 40 µL 10% APS, and 4 µL TEMED. Different percentage gels were prepared by varying the proportion of acrylamide and water accordingly. After removing well combs, wells were rinsed running buffer (1×TAE) using a pipette before loading 2 µL samples typically diluted to 200 nM in loading buffer (1×TAE, 50% glycerol). Electrophoresis was typically at 220V for 180 minutes at 4°C. Gels were stained in a mix of 2 µL $10,000\times$ SYBR Gold Nucleic acid stain (Thermofisher) and 50 mL H_2O, stirring for 5-15 minutes, before washing with H_2O.

7.1.6 Denaturing PAGE

Denaturing gels were made with resolving and stacking layers. A typical 20% 0.75 mm resolving gel was prepared by mixing 2.5 mL 40% 19:1 acrylamide:bis-acrylamide, 0.7 mL 2.8M Tris-HCl (pH 8.8), 2.1g urea (Alfa Aesar) and 1.25 mL formamide. Urea was dissolved by heating in a hot water bath (5 minutes at ~60°C), before adding 33 µL 10% APS and 3 µL TEMED to set the gel. Different percentage

gels were prepared by varying the proportion of acrylamide and water accordingly. A thin layer of 100% ethanol was added on top of the resolving gel while setting, and removed with a tissue before casting a stacking gel above the resolving gel. A typical 5% 0.75 mm stacking gel was prepared by mixing 0.2 mL 40% 19:1 acrylamide:bis-acrylamide, 0.16 mL H_2O, 0.17 mL 1.25M Tris-HCl (pH 6.8), 0.7g urea and 0.42 mL formamide. Urea was dissolved by heating, before adding 11 µL 10% APS and 1 µL TEMED to set the gel. After removing well combs, wells were rinsed running buffer (25 mM Tris-HCl, 200 mM, nopHadjustment glycine) using a pipette before loading 2 µL samples typically diluted to 200 nM in loading buffer (2.5 mM Tris-HCl, 5 mM glycine and 90% formamide, no pH adjustment). Samples were heated to 95°C for 2 minutes and loaded while hot, before electrophoresis at 300V for 60 minutes at RT. Gels were often observed both pre- and post-staining. Gels were stained in a mix of 2 µL 10,000× SYBR Gold Nucleic acid stain (Thermofisher) and 50 mL H_2O, stirring for 5-15 minutes, before washing with H_2O.

7.1.7 Streptavidin PAGE

Samples were mixed with a 10× molar excess of streptavidin (NEB N7021S, 1mg/ml, pH 7.4) for 15 minutes at RT, before diluting samples in Native PAGE loading buffer (1×TAE, 50% glycerol) and running native PAGE or acid PAGE as above.

7.1.8 PAGE densitometry

Gels were analysed using GE ImageQuant TL software. Lanes and bands were identified manually (with a fixed pixel width of 30), and the area underneath each peak of fluorescence intensity collected. Background subtraction was performed by subtracting from the intensity of a given band the intensity from a reference lane with no band. The yield of a species was calculated as the proportion of intensity from the desired band divided by the total from both the desired and undesired species. For example, the yield of a DTS reaction was calculated from unstained gels as:

$$Yield = \frac{Intensity_{\text{ Acceptor-NH-TAMRA}}}{Intensity_{\text{ All TAMRA in the lane}}}$$

Note that this approximation assumes that there is no change in quantum yield of TAMRA emission when attached to DNA, or of SYBRgold when staining different sequences or architecture. This means that the values given are useful for comparing relative yields between architectures but not for determining absolute DTS yields.

7.1.9 LC-MS

LC-MS with a Xevo G2-XS QTof instrument (Waters) was performed by or under supervision by Samuel Núñez-Pertíñez (O'Reilly Group, University of Birmingham). 50 µL 10 µM samples were buffer exchanged into H_2O using Micro Bio-Spin 6 columns (BIORAD) before LC-MS analysis in negative ionisation mode. An ACQUITY UPLC H-Class plus system was used for RP-LC with an ACQUITY UPLC BEH 130 Å1.7 µm C18 (2.1×75mm). Buffer A: 75 mM triethylammonium acetate (TEAA) pH 7.0 buffer in H_2O. Buffer B: 75 mM TEAA pH 7.0 buffer in MeCN. Flow rate: 0.2 mL.min^{-1}. Column temperature: 70°C. Leucine enkephalin was used as the reference for the LockSpray correction. The raw continuum data was deconvoluted to produce zero charge mass spectra using ProMass HR for MassLynx (Novatia) software. Occasionally, samples were analysed directly by MS without prior LC separation (e.g., Figure 2.7C).

7.1.10 DTS

DTS buffer was prepared with 0.25M NaCl, 0.25M TAPS (tris(hydroxymethyl)methyl aminopropanesulfonic acid), 0.25M CAPS (N-cyclohexyl-3-aminopropanesulfonic acid), 0.25M Na_2PO_4, adjusted to various pH (8-12) using HCl and NaOH.

A typical DTS reaction was prepared in a 20 µL volume by mixing: 4 µL H_2O, 2 µL 100 mM $MgCl_2$, 2 µL 20 µM acceptor-DNA-NH_2, 2 µL 10 µM donor DNA-ester-monomer, and initiating the reaction with 10 µL 0.25M DTS buffer. The reaction was incubated overnight (~15 hours). This typical reaction was varied with the addition or exclusion of different strands to create positive and negative controls for DTS, incubated at different pH and temperature, all volumes scaled up for LC-MS analysis, and so on. Specific DTS conditions are provided below:

Figure 2.8. DTS reactions were prepared in 5 µL volumes by mixing: 1 µL 100 mM $MgCl_2$, 1 µL 10 µM 5'-NH_2-acceptor-CX, 1 µL 20 µM template_N, 1 µL flexizyme reaction mix containing ~8 µM donor DNA-3'-ester-biocytin (based on 40% yield from flexizyme acylation reaction with 20uM donor adapter). The reaction was initiated with 1 µL 0.25M DTS buffer (pH 11). The reaction was incubated overnight (~15 hours) at RT. Samples were diluted with 5 µL 10uM streptavidin and incubated for 20 minutes at RT. Samples were diluted with 10 µL native PAGE loading buffer before running 3 µL samples on native PAGE (29:1, 10%, 150V, RT, 45 minutes).

Figure 2.9. DTS reactions were prepared in 80 µL volumes by mixing: 8 µL 100 mM $MgCl_2$, 16 µL 100 µM 5'-NH_2-acceptor-C0, 16 µL 100 µM template_0, 16 µL flexizyme reaction mix containing ~40 µM donor DNA-3'-ester-biocytin (based on 40% yield from flexizyme acylation reaction with 100uM donor adapter). The reaction was initiated with 24 µL 0.25M DTS buffer (pH 11). The reaction was incubated overnight (~15 hours) at RT before LC-MS.

Figure 2.10. DTS reactions were prepared in 80 µL volumes by mixing: 8 µL 100 mM $MgCl_2$, 16 µL 100 µM 5'-NH_2-acceptor-C0, 16 µL 100 µM template_0, 16 µL flexizyme reaction mix containing ~40 µM donor DNA-3'-ester-biocytin (based on 40% yield from flexizyme acylation reaction with 100uM donor adapter). The reaction was initiated with 24 µL 0.25M DTS buffer (pH 11). The reaction was incubated overnight (~15 hours) at RT before buffer exchanging sample into H_2O with 2× washes with a 3kDa Amicon centrifugal filter. Bst reactions were prepared in 50 µL volumes by mixing: 5 µL 100 mM $MgSO_4$, 5 µL 10 µM dNTPs, 5 µL 10× cutsmart buffer (NEB), 5 µL 20 µM 5'-NH_2-acceptor-C0, 5 µL 20 µM template_0, 20 µL flexizyme reaction mix containing ~2.5 µM donor DNA-3'-ester-biocytin (based on 40% yield from flexizyme acylation reaction with 10uM donor adapter). The reaction was initiated with 10 µL 8000U/ml Bst polymerase (NEB). The reaction was incubated for 90 minutes at 37°C. 5 µL 40 µM fluorescent detector was added. 2.5 µL of sample was diluted with 2.5 µL 20uM streptavidin and incubated for 20 minutes at RT. 2.5 µL of sample was diluted with 2.5 µL native PAGE loading buffer before running 2 µL samples on native PAGE (29:1, 10%, 150V, RT, 35 minutes).

Figure 4.6. DTS reactions were prepared in 10 µL volumes by mixing: 1 µL 10 µM donor-S-O-TAMRA, 1 µL 10 µM acceptor-NH_2, 1 µL 10 µM of each helper (with both 5'- and 3'-PO_4 modifications to prevent unintended transfer of TAMRA to helpers), 5 µL buffer containing 200 mM Na_2HPO_4 and 500 mM NaCl (pH 11) and 1 µL H_2O. Samples were incubated overnight (~15 hours) at RT before denaturing PAGE (29:1, 20%, 300V, RT, 90 minutes)

Figure 4.7. DTS reactions were prepared in 20 µL volumes by mixing: 2 µL 100 mM $MgCl_2$, 2 µL 20 µM acceptor-NH_2, 2 µL 10 µM donor-NH-S-O-TAMRA, 10 µL DTS buffer (pH 11), and the remaining volume H_2O. Samples were incubated overnight (~15 hours) at RT before denaturing PAGE (29:1, 25%, 250V, RT, 60 minutes)

Figure 4.8. DTS reactions were prepared as above for Figure 4.7, with only the nature of the modification site altered (a nucleobasic dT amine modification in this case rather than abasic amine).

Figure 5.1. DTS reactions were prepared in 20 µL volumes by mixing: 2 µL 10 µM acceptor-NH_2, 2 µL 10 µM donor-NH-S-O-TAMRA, 10 µL DTS buffer (various pH), and the remaining volume H_2O. Samples were incubated overnight (~15 hours) at RT before denaturing PAGE (19:1, 25%, 250V, RT, 50 minutes)

Figure 5.3A. DTS reactions were prepared in 20 µL volumes by mixing: 2 µL 20 µM acceptor-NH_2, 2 µL 10 µM donor-NH-S-O-TAMRA, 4 µL DTS buffer (pH 12), 8 µL 1 mM NaH_2PO_4 (pH 7), and 4 µL co-solvent (diethyl ether, dichloromethane, ethyl acetate, acetone, acetonitrile, dimethyl sulfoxide, acetic acid, methanol, water or formamide). Samples were incubated overnight (~15 hours) at RT before denaturing PAGE (19:1, 25%, 250V, RT, 50 minutes).

Figure 5.3B. DTS reactions were prepared in 10 µL volumes by mixing: 1 µL 10 µM acceptor-NH_2, 2 µL 10 µM donor-NH-S-O-TAMRA, 1 µL buffer containing 1M NaCl, Na_2HPO_4 (pH 11) and to generate 0-70% mixed solvents a variation of 7-0 µL H_2O with 0-7 µL co-solvent. Samples were incubated overnight (~15 hours) at 4°C before denaturing PAGE (29:1, 25%, 250V, RT, 100 minutes).

Figure 5.9A. Protection reactions were prepared in 30 µL volumes by mixing: 6 µL 0.25M DTS buffer (pH 11), 6 µL 100 mM $MgCl_2$, 6 µL 10 µM abasic ribose modified protector (n=-3) or unmodified ssDNA, 3 µL 10 µM donor-NH-S-O-TAMRA, and the remaining volume H_2O. Samples were incubated overnight at 7°C before triggering TMSD and DTS. DTS reactions were prepared by adding 6 µL 20 µM acceptor-NH_2 (n=1), with an additional 10bp complementarity with the donor adapter to outcompete the protector during TMSD. Samples were incubated overnight (~15 hours) at 7°C before denaturing PAGE (19:1, 25%, 250V, RT, 50 minutes).

Figure 5.10C. DTS reactions were prepared in 20 µL volumes by mixing: 5 µL 20 µM acceptor-NH_2 or protector-abasic site ($n = -3$), 2 µL 25 µM donor-NH-S-O-biotin, 10 µL DTS buffer (pH 11), and the remaining volume H_2O. Samples were incubated overnight (~15 hours) at RT. Samples were split and half hydrolysed by incubating at 95°C for 15 minutes. 5 µL samples were diluted in 5 µL 20 µM streptavidin and incubating at 4°C for 20 minutes, then diluted again in 5 µL native PAGE loading buffer (1×TAE, 50% glycerol) before native PAGE (29:1, 10%, 150V, 4°C, 45 minutes).

7.1.11 Fluorometry

Fluorescent samples containing DNA-FAM species were observed at 37°C in black, flat-bottomed 96-well plates with a BMG LABTECH CLARIOstar (Ex:465-20, Dichroic: 497.5, Em: 530-20, Gain: 1900, Focal height: 4mm, Orbital averaging: 3mm, Cycle time: fluorescence was recorded every 10 seconds for the first hour, and 100 seconds from then on). 200 µL samples were mixed, annealed and pre-incubated at 37°C using a thermocycler before transferring to the CLARIOstar for analysis, to minimise the temperature change when beginning observations. Samples were observed before and after interventions (e.g., the addition of polymerase, nickase, primer or annealing) to ensure samples had equilibrated. Each experiment included controls with buffer only, and negative and positive controls for the position of the fluorescent cargo (C-F).

7.1.12 Polymerase extension

A typical Bst polymerase extension reaction was prepared in a 200 µL volume with: 1× Bst reaction buffer (NEB), 5 mM $MgSO_4$, 50 µM dNTP mix (which may have substituted dTTP for the dideoxynucleotide ddTTP), 100 nM C-F (cargo DNA with FAM modification) with 2× excess of other DNA, and 40U/mL Bst 3.0 polymerase (NEB, M0374L). 10× Bst buffer stock was made with 200 mM Tris-HCl, 100 mM $(NH_4)_2SO_4$, 1500 mM KCl, 20 mM $MgSO_4$, 1% Tween 20 (pH 8.8).

Figure 2.11. The flexizyme reaction was buffer exchanged into cold 0.3mM NaAcOH (pH 5.2) by washing 2× with a 3kDa Amicon centrifugal filter. The Bst polymerase extension reaction was prepared in a 50 µL volume with: 1× cutsmart buffer (NEB), 10 mM $MgSO_4$, 1 mM dNTP mix, 5 µM flexizyme, 2.4 µM DNA adapter and 1.6 µM DNA adapter-3'-ester-biocytin (based on 40% yield from flexizyme acylation reaction with 4uM donor adapter), and 1600U/mL Bst 3.0 polymerase (NEB, M0374L). Samples were incubated at 37°C for 0.5 hours before acid PAGE (29:1, 25%, 220V, 3 hours, 4°C).

Figure 3.6. The Bst polymerase extension reaction was prepared in a 20 µL volume with: 1× cutsmart buffer (NEB), 6 mM $MgSO_4$, 0.2 mM dNTP mix, 1 µM hairpin, 5 µM primer, and 320U/mL Bst 3.0 polymerase (NEB, M0374L). Samples were incubated at 37°C for either 10 minutes, 100 minutes or 1000 minutes before heat-inactivating Bst polymerase at 80°C for 20 minutes. Samples analysed by native PAGE (29:1, 10%, 150V, 30 minutes, RT).

The Φ29 polymerase extension reaction was prepared in a 20 µL volume with: 1× Φ29 buffer (NEB), 0.2 mM dNTP mix, 1 µM hairpin, 2 µM primer, 0.2mg/mL BSA and 1000U/mL Φ29 polymerase. Samples were incubated at 30°C for either 10 minutes, 100 minutes or 1000 minutes before heat-inactivating Φ29 polymerase at 65°C for 15 minutes. Samples analysed by native PAGE (29:1, 10%, 150V, 35 minutes, RT).

Figure 3.7. The polymerase extension reaction was prepared in a 20 µL volume with: 1× cutsmart buffer (NEB), 6 mM $MgSO_4$, 0.2 mM dNTP mix (which may have substituted dTTP for the dideoxynucleotide ddTTP), 1 µM hairpin, 1 µM

primer, and 320U/mL Bst 3.0 polymerase (NEB, M0374L). Samples were incubated at 37°C for 15 minutes before heat-inactivating Bst polymerase at 80°C for 20 minutes. Samples analysed by native PAGE (29:1, 10%, 150V, 40 minutes, RT).

Figure 3.8. The polymerase extension reaction was prepared in a 200 µL volume with: 1× Bst buffer (NEB), 5 mM $MgSO_4$, 50 µM dNTP mix (which may have substituted dTTP for the dideoxynucleotide ddTTP), 100 nM C-F, 150 nM L1-Q, 150 nM L2, 200 nM P1, and 40U/mL Bst 3.0 polymerase (NEB, M0374L). Samples were incubated at 37°C for 10 minutes before the addition of Bst polymerase. After the fluorescence timecourse was complete, samples were rapidly annealed from 95°C to 4°C over 10 minutes before checking the endpoint fluorescence.

Figure 3.9A. The polymerase extension reaction was prepared as above for Figure 3.8, with the modification that stocks of the 10× Bst buffer were prepared with various pH.

Figure 3.9B. The polymerase extension reaction was prepared as above for Figure 3.8, with the modification that the DNA was diluted 10× to 10 nM C-F, 15 nM L1-Q, 15 nM L2, 20 nM P1, and 40U/mL Bst 3.0 polymerase (NEB, M0374L). Samples were incubated at 37°C for up to 5 days before the addition of P1.

Figure 3.9C. The polymerase extension reaction was prepared as above for Figure 3.8, with the modification that the DNA was diluted 10× to 10 nM C-F, 15 nM L1-Q, 15 nM L2, 20 nM P1, 20 nM G, and 40U/mL Bst 3.0 polymerase (NEB, M0374L). Samples were incubated at 37°C for 10 minutes before the addition of Bst polymerase.

Figure 3.10C. The polymerase extension reaction was prepared in a 200 µL volume with: 1× cutsmart buffer (NEB), 10 mM $MgCl_2$, 50 µM dNTP mix (which may have substituted dTTP for the dideoxynucleotide ddTTP), 10 nM P1-C-F, 20 nM L1-Q, 20 nM L2, 40 nM G, 40U/mL Bst 3.0 polymerase (NEB, M0374L) and 50U/mL Nt.AlwI nickase (NEB, R0627S). Samples were incubated at 37°C for 10 minutes before the addition of Bst polymerase and Nt.AlwI nickase .

7.1.13 Nickase Nt.AlwI

A typical nickase and combined nickase-polymerase reactions were prepared in a 200 µL volume with: 1× cutsmart buffer (NEB), 10 mM $MgCl_2$, 50 µM dNTP mix (which may have substituted dTTP for the dideoxynucleotide ddTTP), 10 nM P1-C-F (cargo DNA with FAM modification) with 2× excess of other DNA, 40U/mL Bst 3.0 polymerase (NEB, M0374L) and 50U/mL Nt.AlwI nickase (NEB, R0627S). Reagents and samples were pre-incubated at 37°C before and upon mixing. Samples were observed by PAGE and fluorometry.

Figure 3.10B. The nickase reaction extension reaction was prepared in a 20 µL volume with: 1× cutsmart buffer (NEB), 10 mM $MgSO_4$, 1 µM P1-C, 1 µM L1 and 1000U/mL Nt.AlwI nickase. The reaction was incubated at 37°C overnight with 2 µL samples taken at various timepoints (0, 0.1, 1, 10, 50, 100 and 1000 minutes) and quenched by dilution in 8 µL formamide and heating to 95°C for 10 minutes. Samples were observed by denaturing PAGE (19:1, 25%, 250V, RT, 45 minutes).

7.1.14 Templated disulfide conjugation

The templated disulfide conjugation reactions were prepared in 10 µL volumes with: 1 µM DNA, 50 µM TCEP, and 100 mM NaH_2PO_4 (pH 8). Samples were incubated at RT for 4 hours, before analysing samples by denaturing PAGE. In each experiment negative controls were run in parallel including: 1) substituting H_2O for TCEP, 2) substituting unmodified DNA strands for those containing disulfide modifications. In no case were higher order complexes formed in these samples.

7.1.15 Synthesis of thioester adapters

Reactions were prepared by mixing 10 µL 1 mM disulfide-modified DNA in H_2O with 10 µL 50 mM fresh TCEP dissolved in 200 mM Na_2HPO_4 (pH 12 adjusted with 1M NaOH). After incubating for 1 hour at RT to reduce disulfide modifications to thiols, 20 µL 50 mM 6-carboxytetramethylrhodamine-succinimidyl ester (NHS-TAMRA, Stratech) in 100% DMF was added (40 µL reaction volume, 50%

DMF, 250 µM DNA-SH, 25 mM NHS-TAMRA at 100× excess). The reaction was incubated overnight (~15 hours) at RT. Ice-cold 460 µL 0.3M NaAcOH (pH 5.2) was added to quench the reaction and limit hydrolysis.

DMF and excess TAMRA were removed by washing twice at 4°C in 0.3M NaAcOH (pH 5.2) with an Amicon-Ultra 3kDa 0.5ml centrifugal filter. The 50 µL sample was then analysed and purified by HPLC (Agilent 1200 Series). HPLC samples were stored at 5°C. HPLC buffers were A) 5% acetonitrile, 0.1M triethylammonium acetate (TEAA), and, B) 70% acetonitrile, 0.1M TEAA. Column: Waters XBridge Oligonucleotide BEH C18 column, 130Å, 2.5 µm, 4.6 × 50mm. Column temperature: 35°C. Flow rate: 1 mL/min. The buffer gradient for analysis and purification was 12% to 15% B over 10 minutes. Samples were observed by absorbance at 260nm (DNA) and 550nm (TAMRA), and fluorescence (Ex: 530nm, Em: 580nm). Fractions were collected into 10 µL 3M NaAcOH (pH 5.2) to limit hydrolysis, then washed twice at 4°C in 1mM NaH_2PO_4 (pH 7) with an Amicon-Ultra 3kDa 0.5ml centrifugal filter to remove HPLC buffers. The concentration of DNA was checked by absorbance at 260nm, while the purity was checked by HPLC.

The same protocol was followed substituting NHS-biotin for NHS-TAMRA to produce thioester-biotin modified DTS donor adapters.

When synthesising donor adapters from internal amine-modified DNA, an initial step was added to this protocol. Reactions were prepared by mixing 10 µL 1 mM internal amine-modified DNA in H_2O, 540 µL DMF, 100 µL 100 mM succinimidyl 3-(2-pyridyldithio)propionate (SPDP, Cambridge Bioscience) in DMF and 53 µL 250 mM DTS buffer (pH 12). The reaction was incubated overnight (~15 hours) at RT. To remove SPDP the reaction was ethanol precipitated and the pellet washed, and salts removed by washing twice in H_2O with an Amicon-Ultra 3kDa 0.5ml centrifugal filter. The sample was concentrated to 10 µL by centrifugal vacuum concentration (ThermoScientific DNA120 Savant), before proceeding with disulfide reduction as above.

7.1.16 HPLC timecourse and analysis

30 µL samples for an HPLC timecourse were typically composed of: 1 µM donor adapter (thioester-TAMRA modified DNA) plus a 2× molar excess of other DNA (e.g., with a colocalised amine, abasic site, cholesterol or PEG modification, etc.), 10 mM $MgCl_2$, and 125 mM DTS buffer (pH 11). Samples were stored at 5°C during the timecourse, and 2 µL volume injected at various intervals (e.g., every 400 minutes) on the HPLC (Agilent 1200). HPLC analysis buffers were A) 5% acetonitrile, 0.1M triethylammonium acetate (TEAA), and, B) 70% acetonitrile, 0.1M triethylammonium acetate (TEAA), though the HPLC column was washed with water and 100% acetonitrile between intervals. Column: Waters XBridge Oligonucleotide BEH C18 column, 130Å, 2.5 µm, 4.6 × 50mm. Column temperature: 35°C. Flow rate: 1 mL/min. The buffer gradient was 12% to 15% B over 10 minutes. Samples were observed by absorbance at 260nm (DNA) and 550nm (TAMRA), and fluorescence (Ex: 530nm, Em: 580nm). Peaks were identified using standards that contained a single species (e.g., just DNA or TAMRA).

Timecourse analysis was automated using custom code (Mathematica). Fluorescence was used for analysis as it produced a cleaner signal than absorbance. For each sample the initial fluorescence intensity was subtracted to correct for baseline drift in fluorescence intensity. Background subtraction was performed using a sample of H_2O to correct for the change in fluorescence with buffer content (12-15% buffer B) during elution of the sample from the HPLC column. The intensity under each peak was totalled using a standard window size (e.g., 30 seconds) centred around the peak. The yield of a species was calculated as the proportion of intensity from the desired peak divided by the total from both the desired and undesired species. For example, the proportion of donor adapter (DNA-S-O-TAMRA) during thioester hydrolysis was calculated as:

$$Proportion = \frac{Intensity_{\text{ Donor-S-O-TAMRA}}}{Intensity_{\text{ All TAMRA in the sample}}}$$

This proportion changed over time and was fitted to Ae^{-kt} to determine relative rate k.

7.2 DATA

7.2.1 Optimising flexizyme synthesis conditions

pH. Reactions were prepared by mixing 5 µL H_2O, 15 µL 125 µM flexizyme, and 10 µL 100 µM DNA donor adapter. The sample was rapidly annealed from 95°C to 4°C over 10 minutes, before adding ice-cold 10 µL 5×flexizyme buffer (3M $MgCl_2$, 250 mM HEPES, pH adjusted with 2M KOH) and ice-cold 100 mM DBE-leucine in DMSO. Reactions were incubated on ice overnight (~15 hours), with aliquots taken after 2 hours and 15 hours, quenched by 1/3 dilution in 0.3M NaAcOH (pH 5.2), and run on acid PAGE (29:1, 25%, 300V, 3 hours, 4°C).

Stoichiometry. Reactions were prepared as above with various concentrations of flexizyme and DBE-biocytin relative to 20 µM DNA donor adapter. Acid PAGE (29:1, 20%, 220V, 2 hours, 4°C).

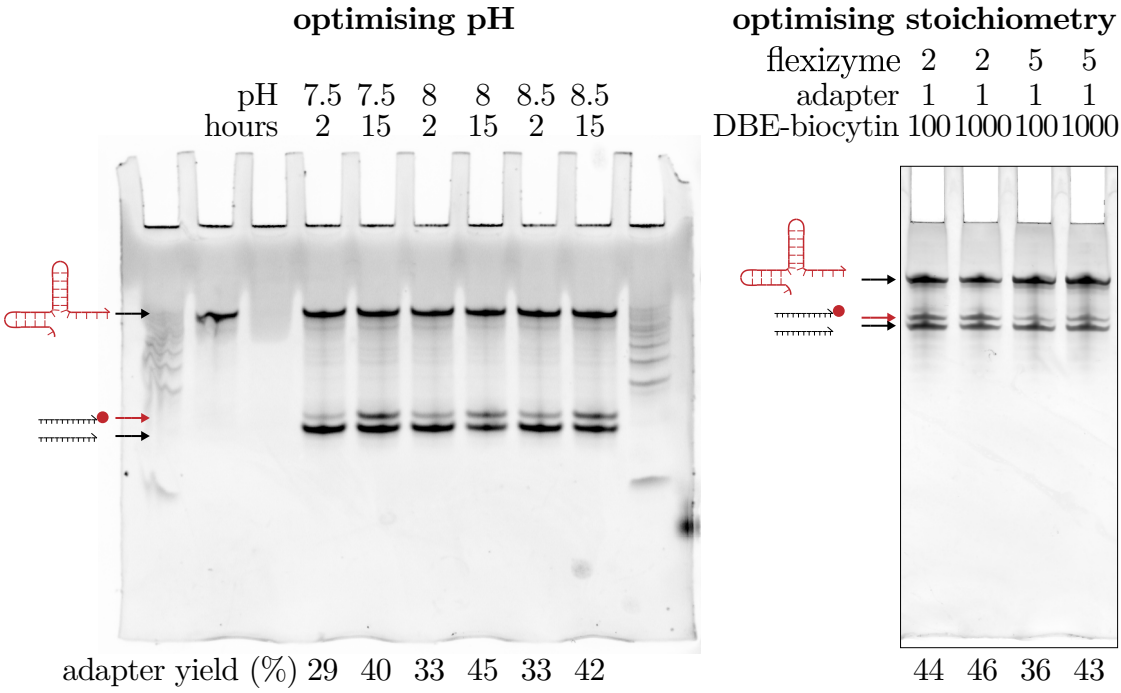


Figure 7.1: Optimising flexizyme synthesis conditions.

7.2.2 Reversing direction of polymerase-driven walker

It appeared that an affinity between fluorophore and quencher could cause the cargo (C-F) to preferentially hybridise to the first leg (L1-Q). I therefore tested the reverse step from L2 to L1-Q by blunt-end exchange (Figure 7.2), which appears to occur more readily than blunt-end transfer from L1-Q to L2. Polymerase-driven TMSD is completed in <2000 seconds, whereas blunt-end transfer takes >20,000 seconds.

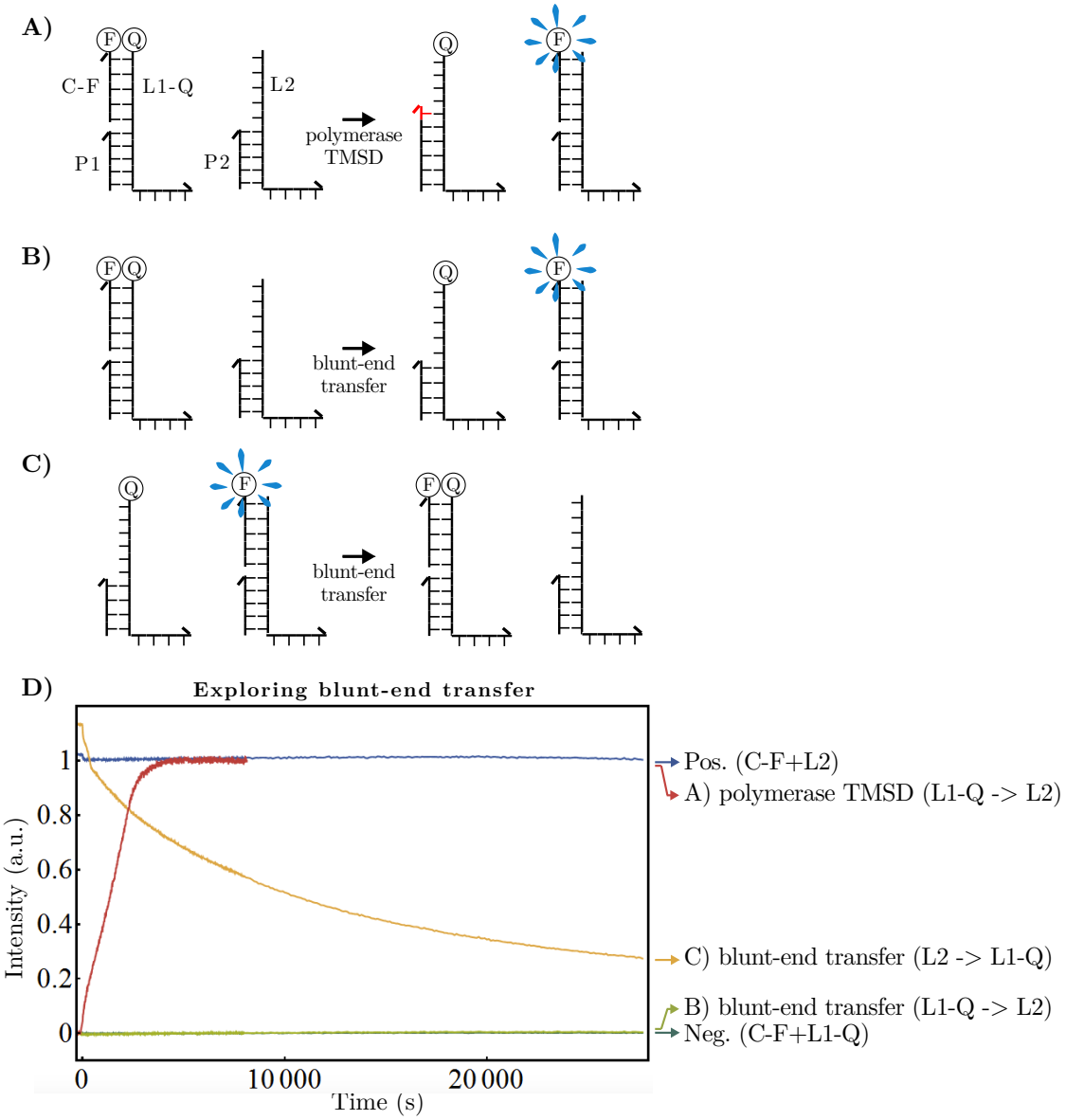


Figure 7.2: Reversing direction of polymerase-driven walker. A-C) Schemes for polymerase-TMSD and blunt-end transfer of cargo between legs. **D)** Fluorescence timecourse testing transfer in both directions between legs.

7.2.3 Characterising the polymerase-driven walker

It was demonstrated that that polymerase-driven TMSD is slower when the strands involved are colocalised on a 'gene', G (Figure 3.9C). I tested whether the addition of DNA to the reaction inhibited the polymerase through the extra 3'OH substrates (Figure 7.3A), or whether the addition of DNA around the polymerase binding site inhibited activity (Figure 7.3B).

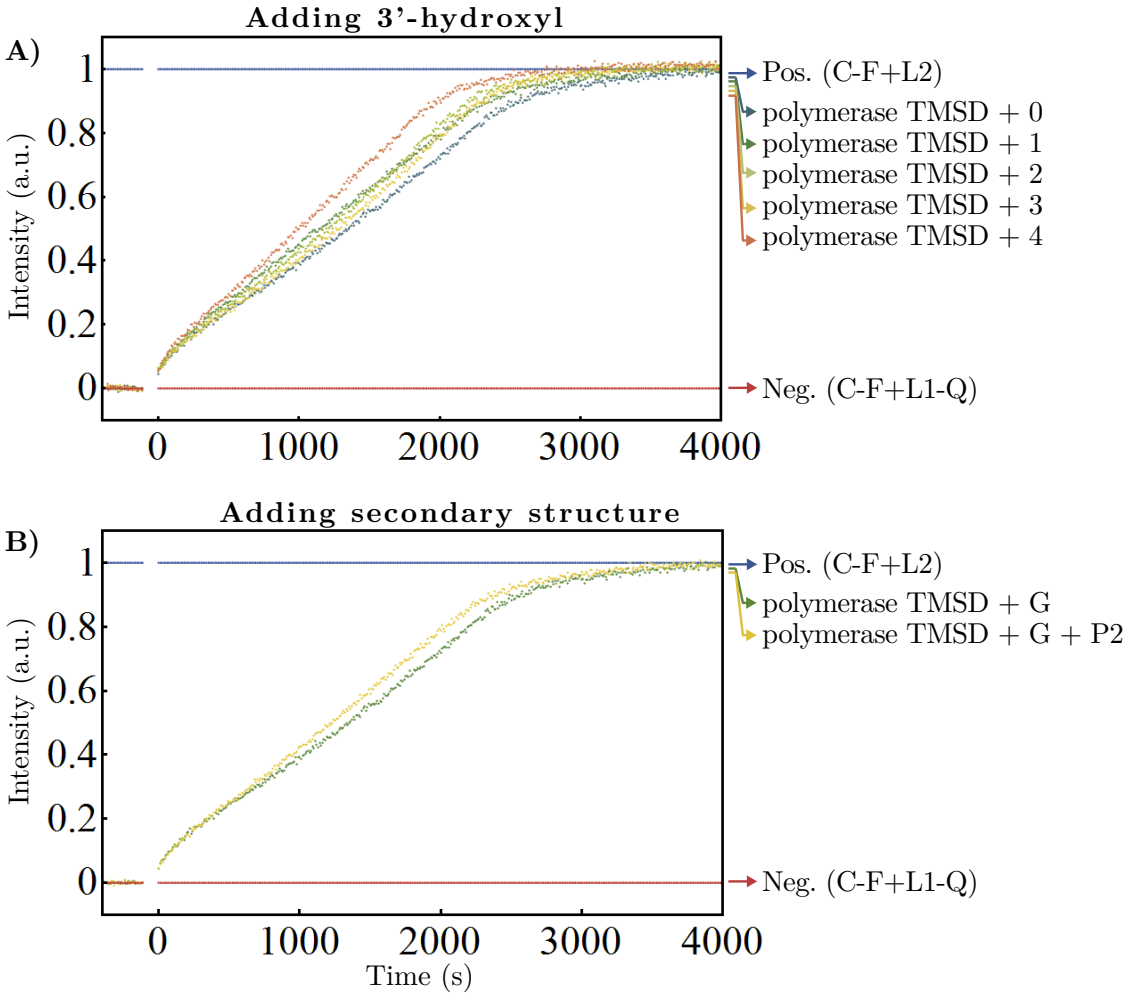


Figure 7.3: Characterising a polymerase-driven walker. A) To explore any effect of extra DNA in the reaction on polymerase activity, I tested polymerase-TMSD between uncolocalised legs with the addition of 0-4 unrelated DNA strands to the reaction mix. **B)** To explore any effect of secondary structure on polymerase activity, I tested polymerase-TMSD between colocalised legs with the addition of 1 complementary DNA strand (P2) to the reaction mix.

7.2.4 Time and [TCEP] during templated disulfide conjugation

TCEP indiscriminately reduces disulfides to thiols and should also break the disulfide bonds formed between DNA strands during templated conjugation reaction. However, TCEP is oxidised by the reaction and thus is removed over time. I characterised this reaction by incubating an end-of-helix architecture at various [TCEP] for various lengths of time (Figure 7.4). I showed that conjugation between strands requires [TCEP] > [DNA-SS], and that high concentrations of TCEP require a longer time for conjugation between strands, presumably because disulfides couple and cleave many times until all TCEP in the reaction is oxidised.

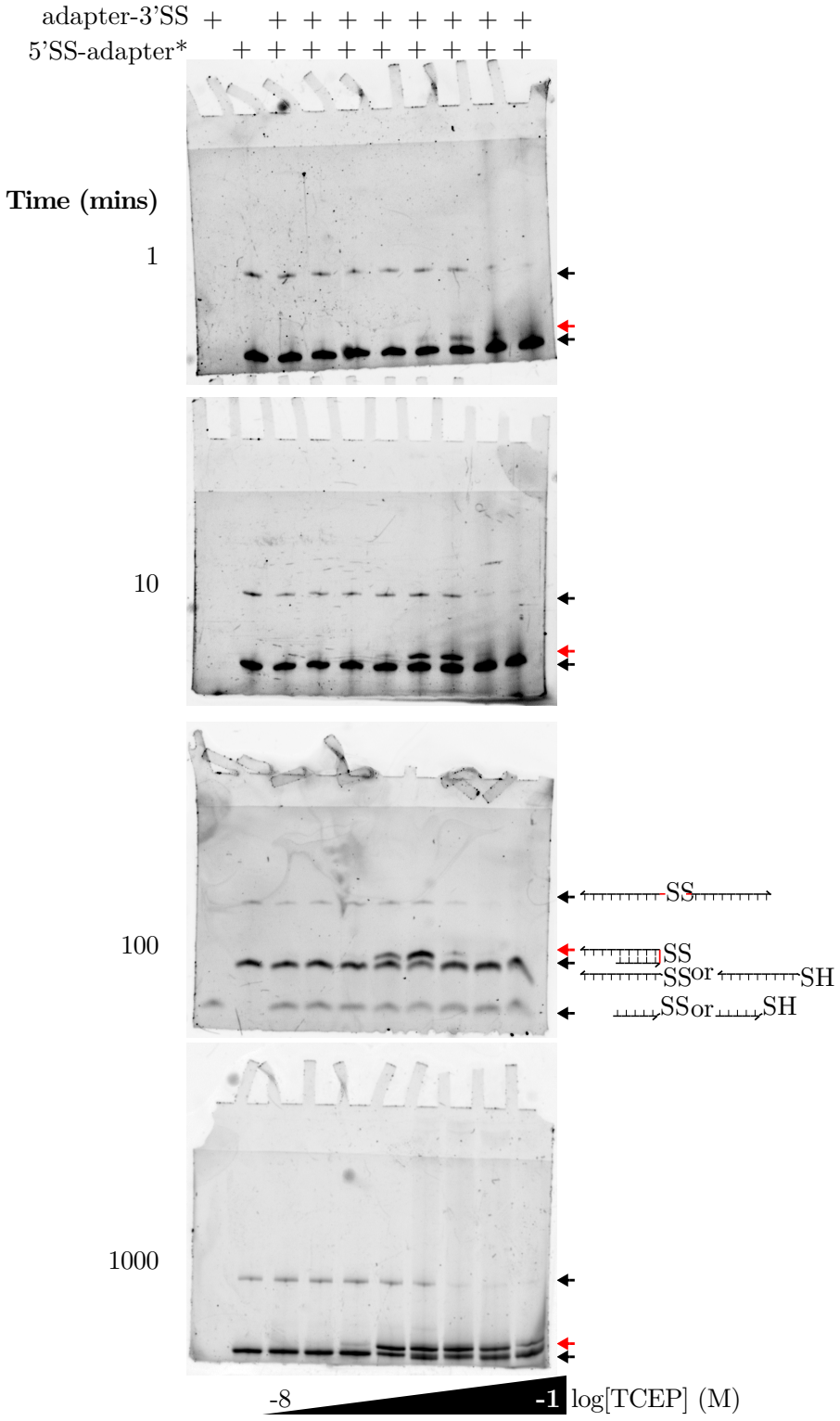


Figure 7.4: Templated conjugation in an end-of-helix architecture (1 µM disulfide-modified DNA) was explored for a range of [TCEP] (10nM to 100mM) and incubation times (1 minute to 1000 minutes). Samples run on 15% denaturing PAGE.

7.2.5 Temperature during templated disulfide conjugation in end-of-helix architecture

Templated conjugation depends on the colocalisation of reactive groups. Disulfide conjugation between reactive groups colocalised by short, 4bp overhangs was limited to low temperatures (Figure 4.4C). In contrast, conjugation occurs over a wide range of temperatures when colocalised in the end-of-helix architecture (Figure 7.5).

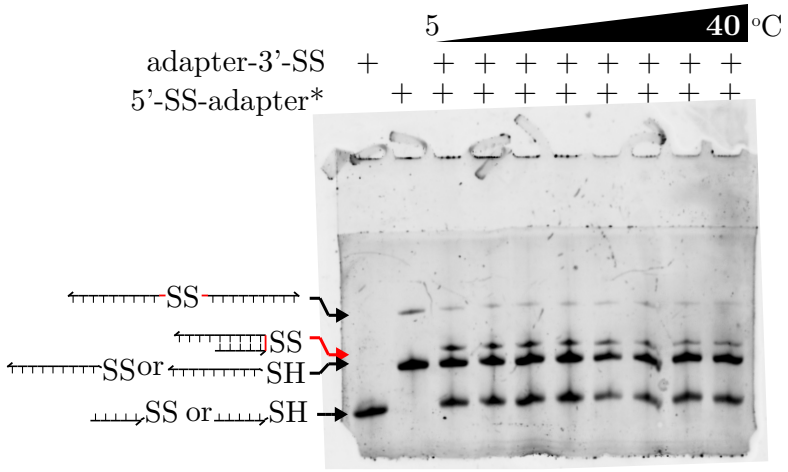


Figure 7.5: Templated conjugation in an end-of-helix architecture was explored for a range of temperatures (5°C and 40°C). Samples run on 15% denaturing PAGE.

7.2.6 Disulfide conjugation with 0nt and 2nt overhangs

Templated conjugation depends on the colocalisation of reactive groups. Disulfide conjugation between reactive groups colocalised by short overhangs was limited to overhangs >4nt (Figure 7.7).

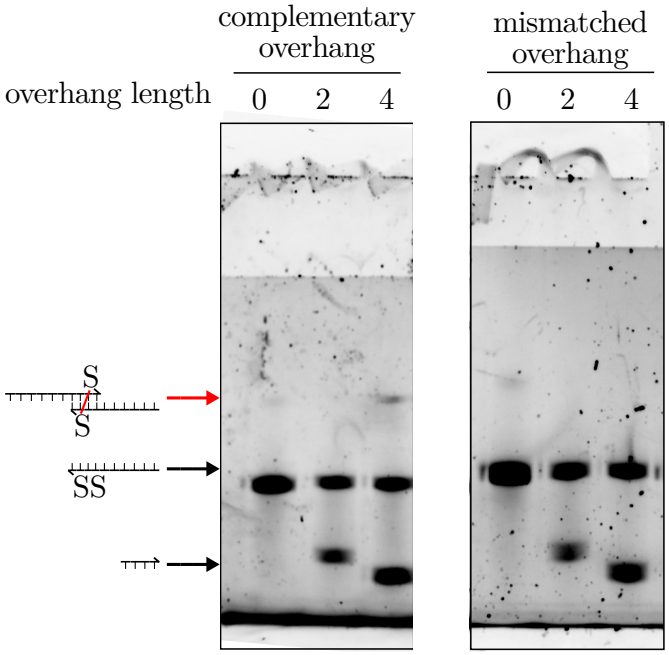


Figure 7.6: Templated conjugation in an overhang architecture was explored for a range of overhang lengths (0, 2, 4nt). Samples run on 15% denaturing PAGE.

7.2.7 LC-MS of thioester-TAMRA donor adapters

Donor adapters consisting of terminal or internal thioester-modifications were synthesised. LC-MS of donor adapters was performed by Samuel Núñez-Pertíñez (O'Reilly Group, University of Birmingham).

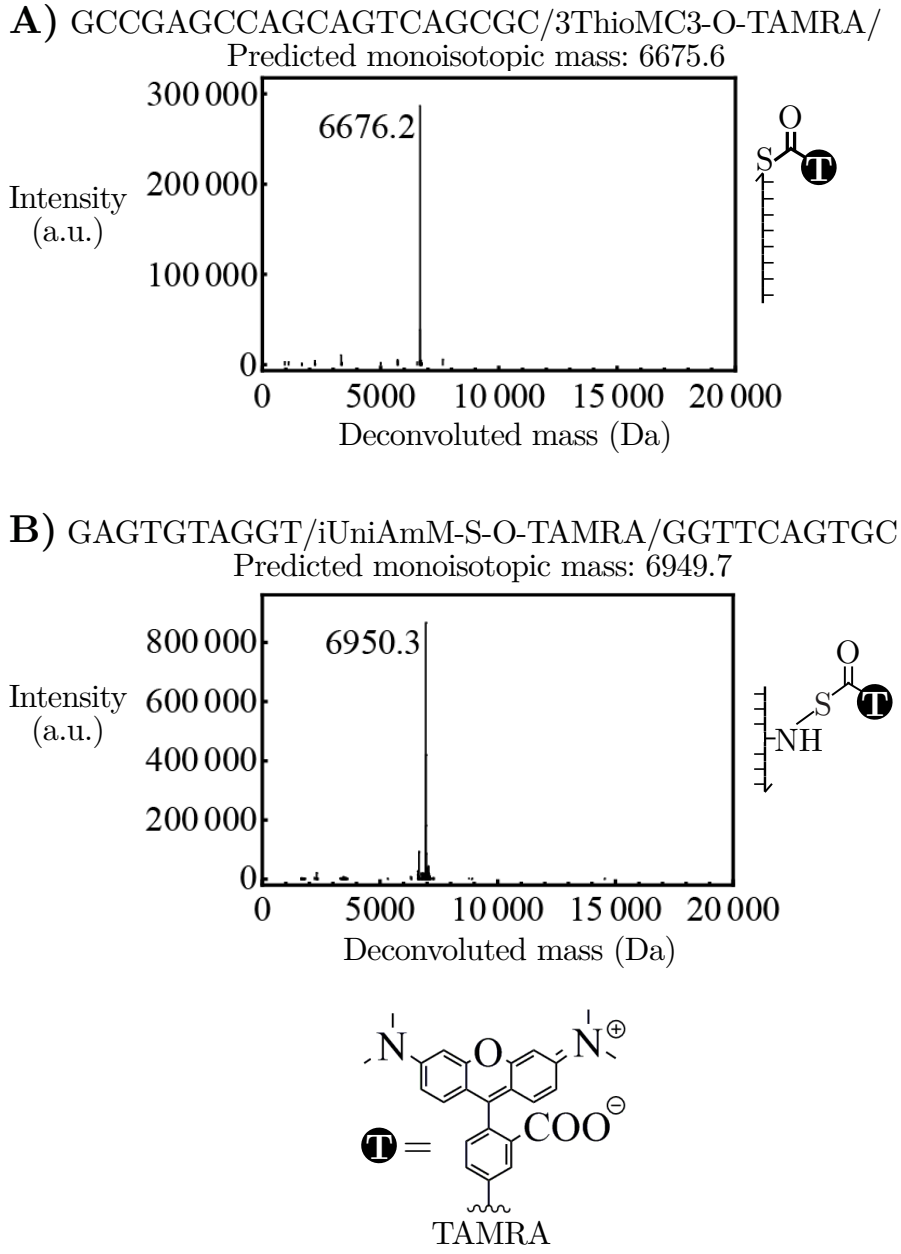


Figure 7.7: A) Deconvoluted mass spectra for a terminally modified donor adapter. **B)** Deconvoluted mass spectra for an internal, abasic modified donor adapter.

7.2.8 Position scan of DTS with thioester-biotin

DTS between an internal thioester-biotin donor and a nearby internal amine acceptor was tested (Figure 5.10), but no protection of the thioester was observed. To explore this result, various positions of the internal abasic amine modification were tested during a DTS reaction. To determine if the abasic ribose or abasic amine modifications offered any protection from hydrolysis, the DTS reaction was run on native PAGE pre- and post-hydrolysis. If protection of thioester-biotin from hydrolysis was occurring as a result of the nearby abasic site, then before hydrolysis we would expect would be more thioester-biotin remaining relative to a control (C) with an unmodified DNA sequence containing no abasic sites. After boiling these samples would appear the same. Figure 7.8 shows the results of this experiment. The control (C) lacking any abasic amine modification appeared the same as most other samples (~40% of dsDNA shifted by streptavidin), indicating no protection from hydrolysis in any position. After hydrolysis, only samples with $1 \leqslant n \leqslant 5$ showed evidence of DTS (20-40% dsDNA shifted by streptavidin).

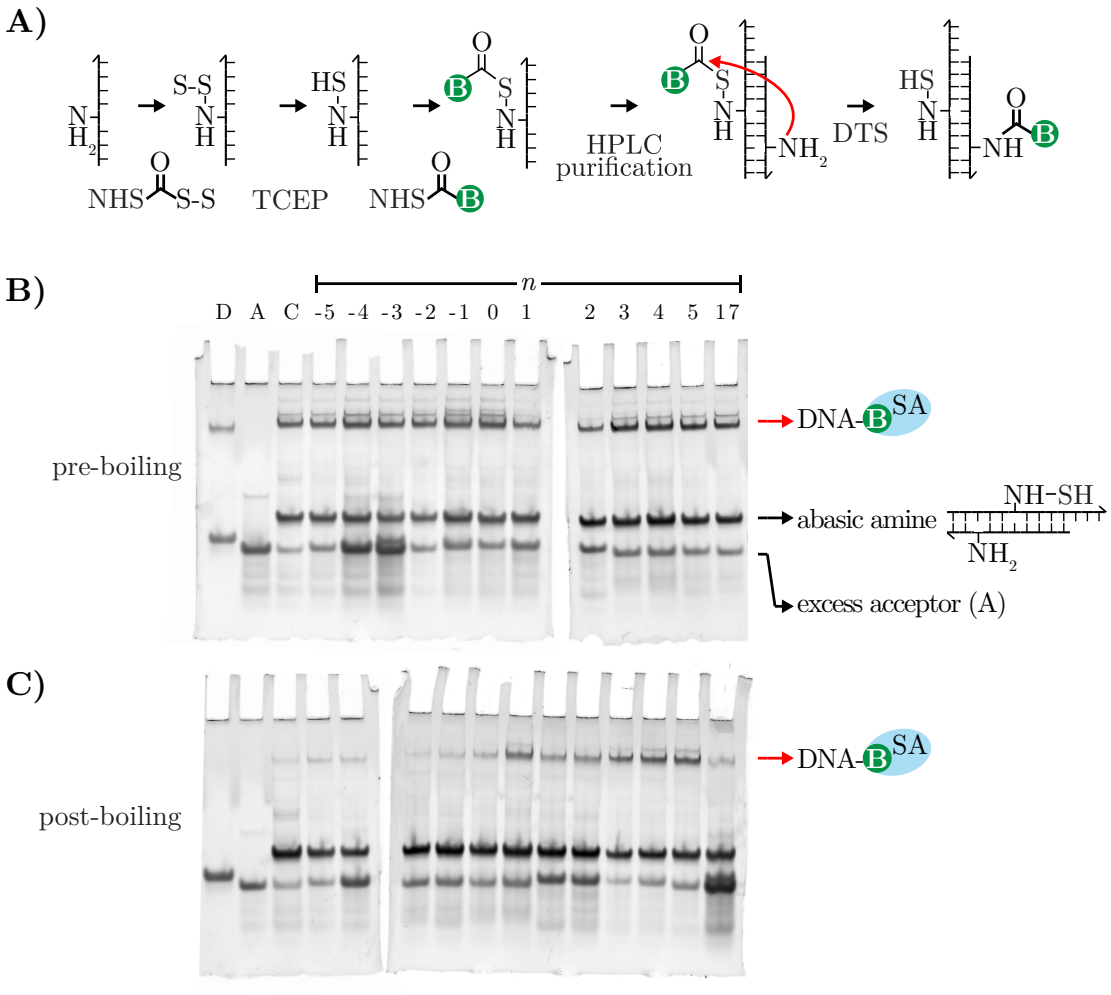


Figure 7.8: Position scan of DTS with thioester-biotin. A) Scheme for synthesis of donor adapters and DTS with an internally amine-modified acceptor. **B)** 10% Native-PAGE of the DTS reaction after incubating with streptavidin. **C)** 10% Native-PAGE of the DTS reaction after boiling and then incubating with streptavidin.

7.3 MODELS

7.3.1 Calculating C_{eff} in DTS architectures

The crudest approximation of C_{eff} considers one molecule interacting with another molecule in the 3D volume defined by the distance d between attachment points:

$$C_{\text{eff}} = \frac{3}{4\pi d^3 N_{\text{a}}}$$

Where N_{a} is Avogadro's number. This is not physically accurate as it assumes: 1) no effect of linkers or template on the interaction volume, 2) a uniform distribution of molecules within the volume, and 3) the template fixing linker attachment positions is rigid.

Using distance measurements from DNA crystal and NMR structures (PDB: 1BNA, 1WAN, 2CRX [116–118]), C_{eff} was estimated for various DTS architectures (Figure 7.9).

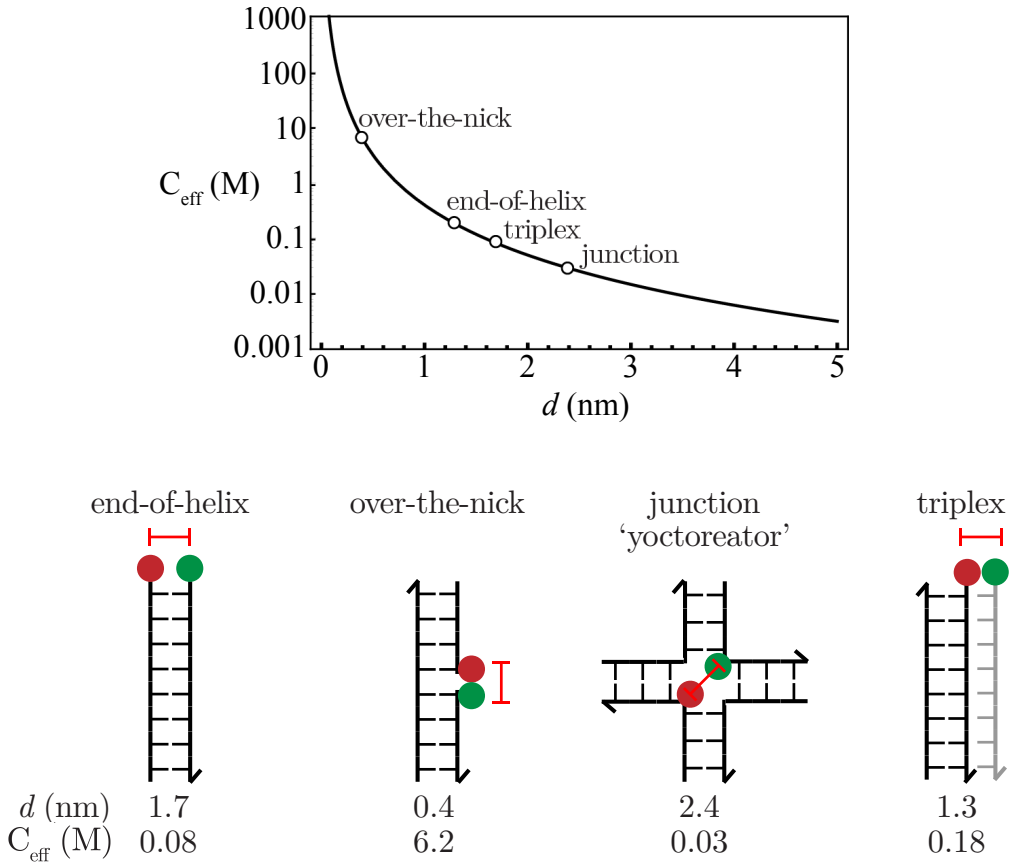


Figure 7.9: A crude model of effective concentration.

7.3.2 Linker attachment position and length

An improved model of templated transfer add details of linker attachment positions and lengths. If linkers of equal length l are attached at distance d, then there is an overlapping volume between the two spheres probes by linkers of [246]:

$$V_{\text{overlap}} = \frac{\pi}{12}((4l + d)(2l - d)^2)$$

This is valid for $l > d/2$ so that linkers are long enough to reach each other. If $l \leqslant d/2$, then $V_{\text{overlap}} = 0$. It is assumed that the proportion of time that two molecules spend together in the overlapping volume is $p = V_{\text{overlap}}/V_{\text{total}}$, where V_{total} is the volume of the sphere probed by a linker. This fraction p increases from 0 to 1 with longer linkers.

For a reaction between two components:

$$A + B \rightarrow C$$

$$\frac{dC}{dt} = k[A][B]p$$

Where the only modification to conventional second-order kinetics is multiplying by p, the proportion of time that the [A] and [B] exist in the same volume. Additionally, the concentrations [A] and [B] can be calculated directly as there is one atom per sphere encompassed by the linker $[A] = [B] = \frac{1}{V_{\text{total}}}$. Assuming that the second-order rate constant k (in atoms^{-1}.s^{-1}) is unaffected by linker position or length, then we can assume a constant $k = 1$ to calculate $\frac{dC}{dt}$ (e.g., Figure 7.10).

This model indicates that there is an optimal linker length for templated reactions. When reactants are positioned closer, it is possible for them to react with shorter linkers at faster rates. Very long linkers have slower rates due to decreased effective concentration.

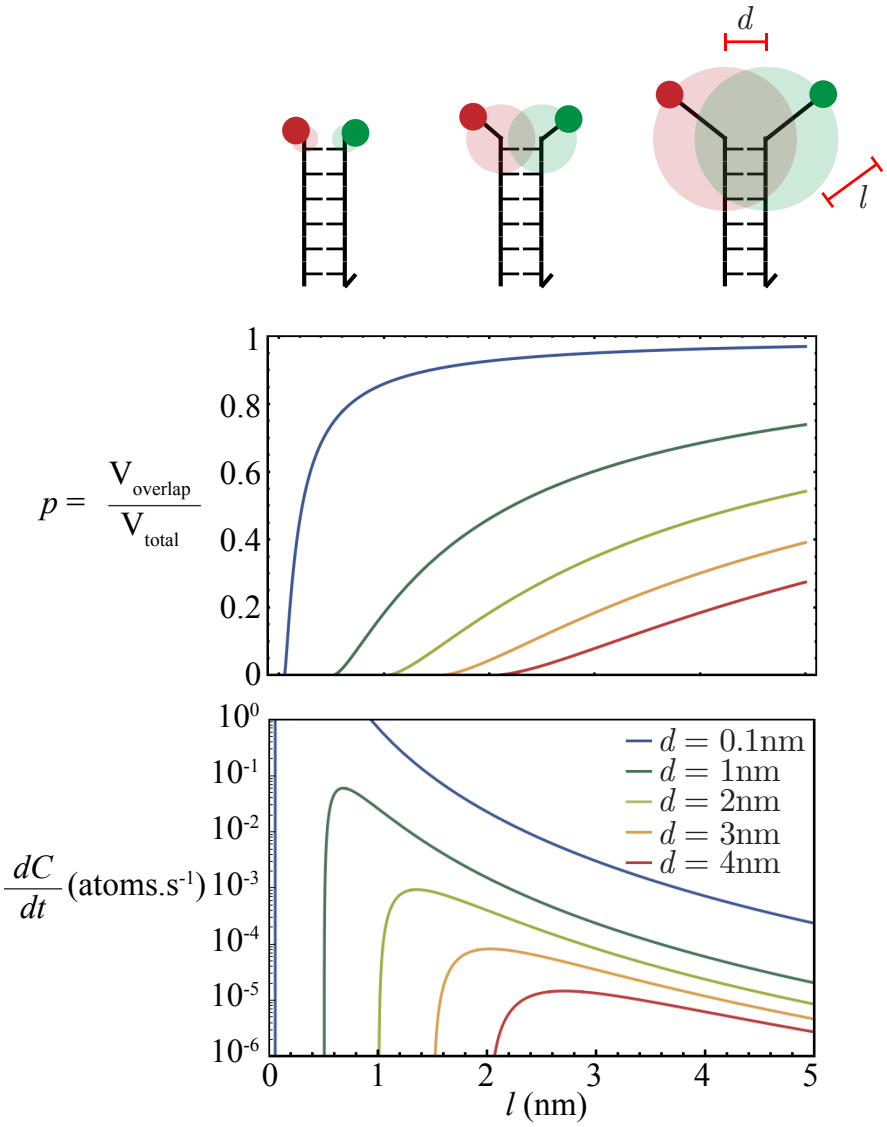


Figure 7.10: Linkers must be long enough to reach each other $l > d/2$. However, very long linkers will traverse a large volume with slower kinetics.

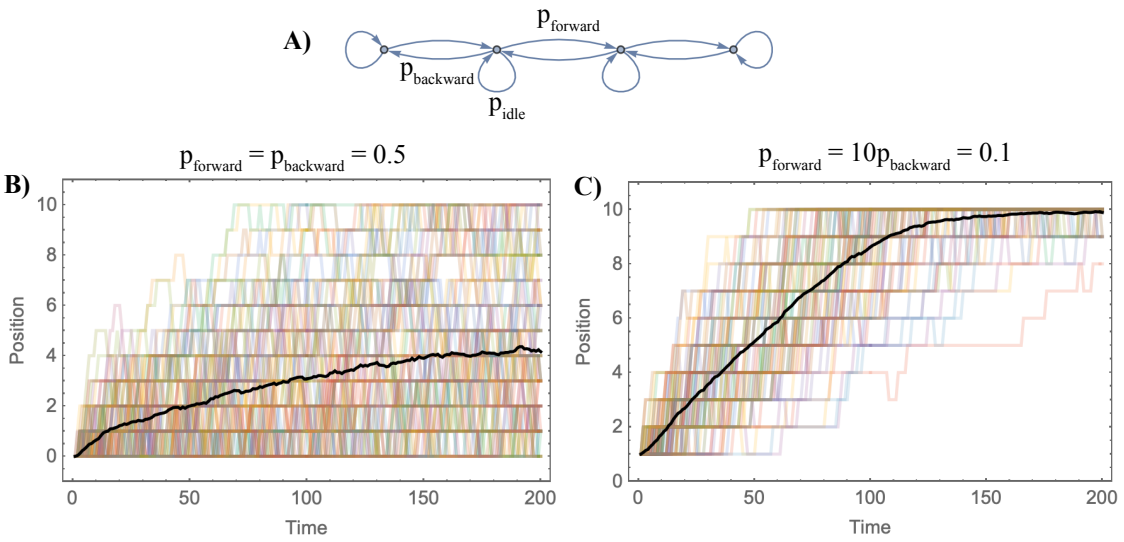


Figure 7.11: A) The motor steps between positions on a track with defined probabilities, resulting in a biased random walk to one end of the track. **B-C) Simulations of motor stepping.** The coloured trajectories show individual simulations, where the thick, black line shows the average of 200 simulations.

7.3.3 Programmable polymerase

A model of a programmable polymerase can help develop an intuition about the design modifications discussed in this thesis, such as reducing degradation or implementing feedback. In particular, it might predict where the most improvement can be made. Here I develop a Markov-chain model of a programmable polymerase.

Motion. The motor is modelled as a biased random walk between positions (or nodes of a graph Figure 7.11A). At each timepoint, t, if a random number is less than $p_{forward}$, the motor moves forward. Similarly, the motor can step backward. The motor idles if no or both transitions are made. Depending on the balance between forward and backward probabilities, the motor can be driven to one end of a track (e.g., Figure 7.11B-C). For DNA walkers driven by toehold mediated strand displacement, the probability of forward stepping would be at least 10× greater than backward stepping.

Degradation. For each adapter on a track a random number is chosen at each timepoint and if less than $p_{degrade}$, the adapter degrades irreversibly. One way of visualising this shows each adapter on a track degrading over time (Figure

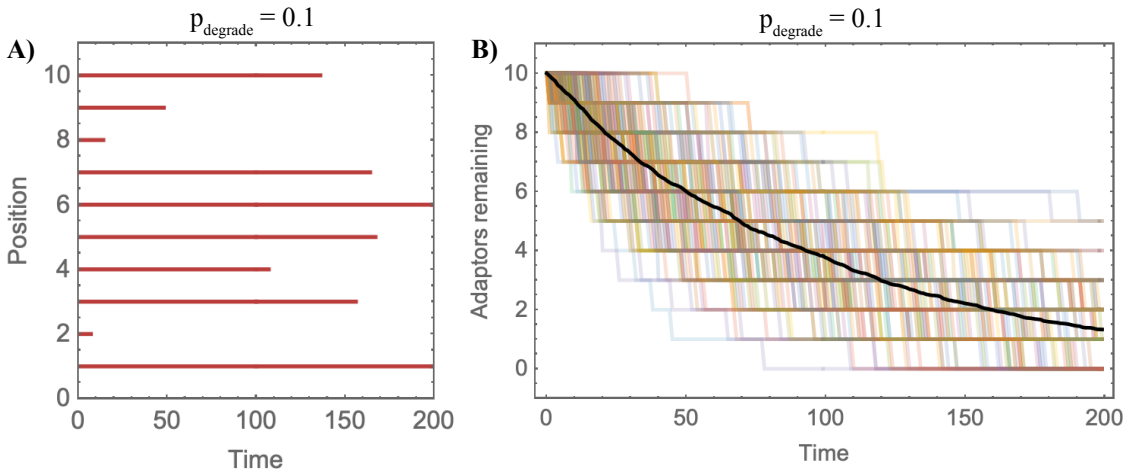


Figure 7.12: A) Simulation of a single track. The degradation of adapters at each position on a track is shown. When a red line ends, the adapter at that position has degraded. **B) Multiple simulations.** The coloured trajectories show the number of adapters remaining during individual simulations, where the thick, black line shows the average of 200 simulations.

7.12A), or alternatively, visualising the total number of adapters remaining over time (Figure 7.12B).

Transfer. At each timepoint if the motor is in the same position as an undegraded adapter, and if a random number is less than $p_{transfer}$, then there is templated transfer of a monomer from the adapter to the motor. This degrades the adapter and also adds the monomer from that position to the polymer. One way of visualising this depicts a single programmable polymerase over time (Figure 7.13A), or alternatively, visualising the total length of polymers synthesised (Figure 7.13B).

Similarly, background transfer can be added to simulate untemplated, unprogrammed interactions between adapters and polymerases. This doesn't depend on the motor position, and instead has a 50/50 probability that background transfer either degrades an adapter or adds a random monomer to the polymer, simulating the random interaction between polymerases.

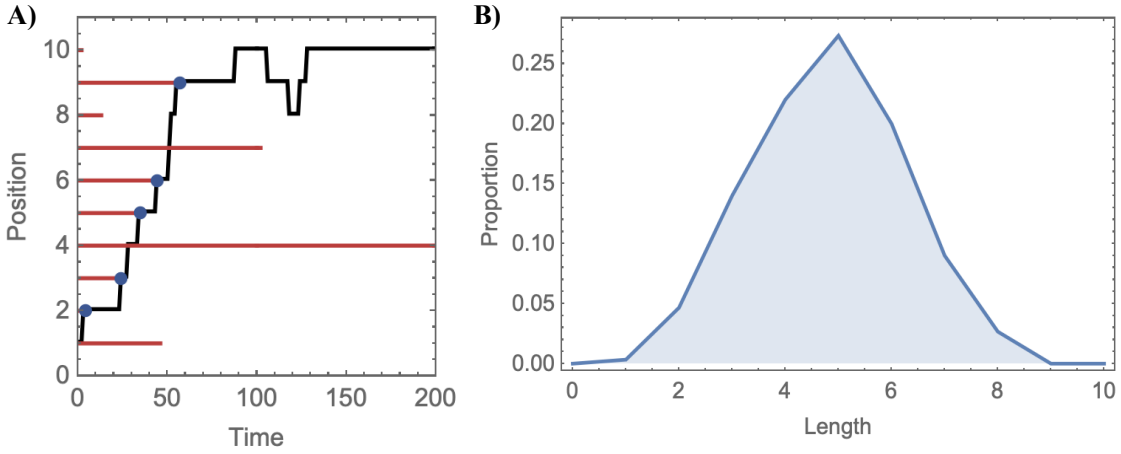


Figure 7.13: A) Simulation of a single track. When a red line ends, the adapter at that position has degraded. Alternatively, if a monomer is transferred from an adapter to the motor when colocalised, a blue circle is depicted. The position of the motor over time is shown with a black line. Sometimes the motor steps forward before transfer has occurred (e.g., positions 1, 4 and 7), whereas other times the adapter degrades before it is reached by the motor (e.g., positions 8 and 10). In this simulation, $p_{\text{forward}} = 0.1$, $p_{\text{backward}} = 0.01$, $p_{\text{degrade}} = 0.01$, $p_{\text{transfer}} = 0.2$, $p_{\text{background transfer}} = 0$. **B) Multiple simulations.** The average length of polymers synthesised with this set of parameters (200 replicates).

Parameters. It is important to set the probabilities of each transition at realistic levels. There are few well-characterised systems, though the best is Erben et al. using PNA-templated transfer in an over-the-nick architecture [111]. Unfortunately, there is no motor in this system, though from Meng et al. [136] we believe that the rate of a DNA motor can be tuned to run about 10× slower than the rate of transfer, and the rate of backward stepping is at least 10× slower than the rate of forward stepping. This was used to estimate rates for the various transitions (Figure 7.14A). From these, the relative probabilities for each transition can be calculated by choosing an appropriately small timestep, $dt = 3/10000s$, for each iteration of the simulation, and using the measured rates to infer the probability that a transition occurs within dt by: $p = 1 - e^{-kdt}$. Interestingly, the distribution of lengths produced by simulated polymerases with these parameters (Figure 7.14C) is in the same order of magnitude as that measured by Meng et al. [136], with a distribution of lengths between 1-7 monomers and a mean of ~3.5.

Design modifications. The application of the model is to determine how various design modifications to programmable polymerases are likely to affect polymer synthesis. For example, in Chapter 5 it was shown that an abasic site decreased the rate of adapter degradation by >100×. This is predicted to increase the yield of 10mer from 0.2% to 30% (Figure 7.15A).

Interestingly, adding feedback by itself does not increase the yield of longer polymers. However, feedback allows the motor to run as fast as possible without overstepping. If the rate of forward stepping is increased to the fastest rate possible ($p_{forward} = 1$) then feedback permits the synthesis of longer polymers (Figure 7.15B), increasing the yield of 10mer from 0.2% to 40%.

A)

	$pM.s^{-1}$	p	Ref.
degradation	7	0.0021	[111]
transfer	410	0.1157	[111]
background transfer	3	0.0009	[111]
forward stepping	41	0.0122	
backward stepping	4.1	0.0012	

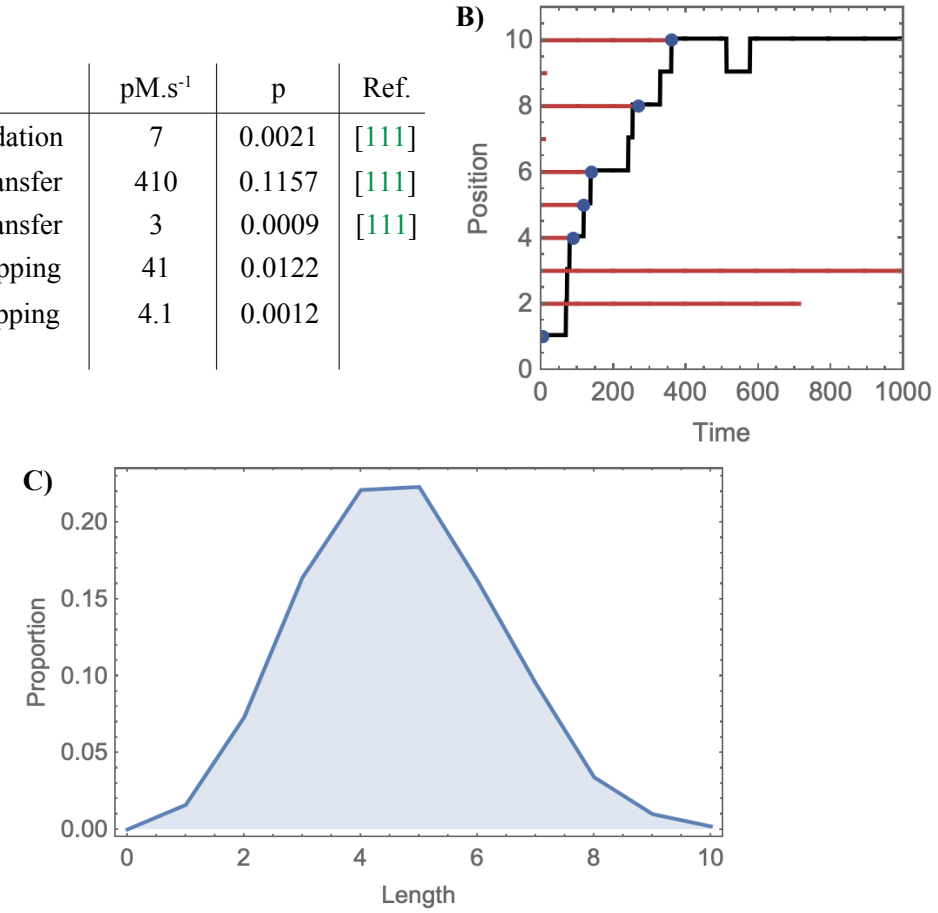


Figure 7.14: **A)** Rates and probabilities for transitions. **B)** An example simulation with the chosen probabilities. **C)** The average length of polymers synthesised with this set of parameters (1000 replicates).

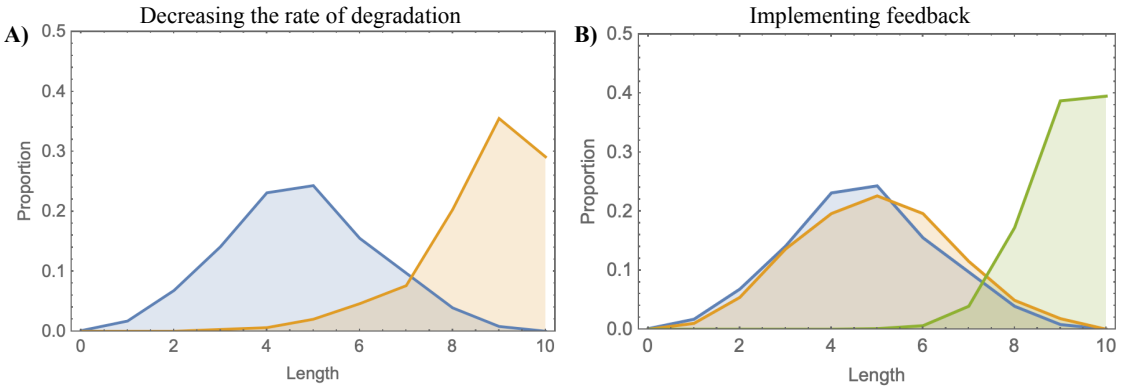


Figure 7.15: **A)** The effect of decreasing the rate of adapter degradation by 100× (orange), compared to the standard model from Figure 7.14 (blue). **B)** The effect of implementing feedback (orange), or both impementing feedback and increasing the rate of motor stepping to the fastest possible (green).

7.4 SEQUENCES

DNA and RNA were purchased from IDT. Unmodified strands were synthesised at 25nm with standard desalting. Modified strands were synthesised at 1um scale with HPLC purification. In the table below, RNA sequences are blue. RNA was synthesised with RNAse-free purification.

Name	Sequence
RNA flexizyme	GGAUCGAAAGAUUUCCGCAUCCCCGAAAGGGUACAUGGCGUUAGGUCGCAAA
donor adapter (3'ester)	CTACCTTCAATGCCTCTAACTTTGCGACCA
template0	ATAGTAAGTCCAGGCAGCAATGGTCGCAAAGTTAGAGGCATTGAAGGTAG
template1	ATAGTAAGTCCAGGCAGCAATTGGTCGCAAAGTTAGAGGCATTGAAGGTAG
template2	ATAGTAAGTCCAGGCAGCAATTTGGTCGCAAAGTTAGAGGCATTGAAGGTAG
template3	ATAGTAAGTCCAGGCAGCAATTTTGGTCGCAAAGTTAGAGGCATTGAAGGTAG
acceptor adapter (C0)	/5AmMC0/TTGCTGCCTGGACTTACTAT
acceptor adapter (C6)	/5AmMC6/TTGCTGCCTGGACTTACTAT
acceptor adapter (C12)	/5AmMC12/TTGCTGCCTGGACTTACTAT
C-F	GGACGGTTTACTTTGATGGGACTT/36-FAM/
L1-Q	/5IABkFQ/AAGTCCCATCAAAGTAAACCGTCCAGAGATCGCATTGTATTGTATTCACCAGTTAAAGTC/3Phos/
L2	AAGTCCCATCAAAGTAAACCGTCCGCTCCTACCGTAATCTATATCCCAGCCTCATCCAAA
P1	CAATACAATGCGATCTCT
P1-ext	CAATACAATGCGATCTCTGGACGGT
P2	ATAGATTACGGTAGGAGC/3Phos/
G	TTTGGATGAGGCTGGGATGACTTTAACTGGTGAATA/3Phos/
P1-C-F (nickase)	GACAAGACAGGATCACCAGGACGGTTTACTTTGATGGGACTT/36-FAM/
L1-Q (nickase)	/5IABkFQ/AAGTCCCATCAAAGTAAACCGTCCTGGTGATCCTGTCTTGTCTATTCACCAGTTAAAGTC/3Phos/
P1 (nickase)	GACAAGACAGGATCACCA
P1-ext (nickase)	GACAAGACAGGATCACCAGGACGGT
adapterA-SS (GCGC)	GCCGAGCCAGCAGTCAGCGC/3ThioMC3-D/
adapterA-SS (GCCG)	GCCGAGCCAGCAGTCAGCCG/3ThioMC3-D/
helperA (short)	TGACTGCTGGCTCGGC
helperA (long)	TGACTGCTGGCTCGGCTTTTGCTGGGACGAATCCCGTCCGTTGCTAGTTAGTACGC
P-helperA (long)	/5Phos/TGACTGCTGGCTCGGCTTTTGCTGGGACGAATCCCGTCCGTTGCTAGTTAGTACGC
helperA-P (long)	TGACTGCTGGCTCGGCTTTTGCTGGGACGAATCCCGTCCGTTGCTAGTTAGTACGC/3Phos/
P-helperA-P (long)	/5Phos/TGACTGCTGGCTCGGCTTTTGCTGGGACGAATCCCGTCCGTTGCTAGTTAGTACGC/3Phos/
adapterB-SS (mismatch)	TTTTTTTTTTGCACAGGCATGTCCTGGTCT/3ThioMC3-D/
helperB (short)	CAGGACATGCCTGTGC
helperB (long)	CAGGACATGCCTGTGCTTTTGCGTACTAACTAGCAACGGACGGGATTCGTCCCAGC
adapterA*-NH_2 (end of helix, C6)	/5AmMC6/GCGCTGACTGCTGGCTCGGCTTTTTTTTTT
adapterB-NH_2 (CGGC)	TTTTTTTTTTGCACAGGCATGTCCTGCGGC/3AmMO/
adapterB-NH_2 (mismatch)	TTTTTTTTTTGCACAGGCATGTCCTGGTCT/3AmMO/
donor-NH_2 (abasic)	GCCGAGCCAGCAGTCAGCGC/iUniAmM/GTCCTAATCTACCTG
acceptor (neg. control)	CAGGTAGATTAGGACAGCGCTGACTGCTGGCTCGGC

Name	**Sequence**
acceptor-NH_2 (abasic, -5)	CAGGTAGATTAGGACAGCGC/iUniAmM/GACTGCTGGCTCGGC
acceptor-NH_2 (abasic, -4)	CAGGTAGATTAGGACAGCG/iUniAmM/TGACTGCTGGCTCGGC
acceptor-NH_2 (abasic, -3)	CAGGTAGATTAGGACAGC/iUniAmM/CTGACTGCTGGCTCGGC
acceptor-NH_2 (abasic, -2)	CAGGTAGATTAGGACAG/iUniAmM/GCTGACTGCTGGCTCGGC
acceptor-NH_2 (abasic, -1)	CAGGTAGATTAGGACA/iUniAmM/CGCTGACTGCTGGCTCGGC
acceptor-NH_2 (abasic, 0)	CAGGTAGATTAGGAC/iUniAmM/GCGCTGACTGCTGGCTCGGC
acceptor-NH_2 (abasic, 1)	CAGGTAGATTAGGA/iUniAmM/AGCGCTGACTGCTGGCTCGGC
acceptor-NH_2 (abasic, 2)	CAGGTAGATTAGG/iUniAmM/CAGCGCTGACTGCTGGCTCGGC
acceptor-NH_2 (abasic, 3)	CAGGTAGATTAG/iUniAmM/ACAGCGCTGACTGCTGGCTCGGC
acceptor-NH_2 (abasic, 4)	CAGGTAGATTA/iUniAmM/GACAGCGCTGACTGCTGGCTCGGC
acceptor-NH_2 (abasic, 5)	CAGGTAGATT/iUniAmM/GGACAGCGCTGACTGCTGGCTCGGC
acceptor-NH_2 (abasic, -17)	CAGGTAGATTAGGACAGCGCTGACTGCTGGC/iUniAmM/CGGC
acceptor-ribose (abasic, -3)	CAGGTAGATTAGGACAGC/iSpC3/CTGACTGCTGGCTCGGC
acceptor-ribose (abasic, 0)	CAGGTAGATTAGGAC/iSpC3/GCGCTGACTGCTGGCTCGGC
acceptor-ribose (abasic, 2)	CAGGTAGATTAGG/iSpC3/CAGCGCTGACTGCTGGCTCGGC
donor-NH_2 (abasic, toehold)	GCCGAGCCAGCAGTCAGCGC/iUniAmM/GTCCTAATCTACCTGTT TTTTTTTT
acceptor-NH_2 (abasic, 1, toehold)	AAAAAAAAAACAGGTAGATTAGGA/iUniAmM/AGCGCTGACTGC TGGCTCGGC
donor-NH_2 (nucleobasic)	GAGACCTAC/iAmMC6T/ACGCACGCTC
acceptor-NH_2 (nucleobasic)	TTTTTTTTTTGAGCGTGCG/iAmMC6T/AGTAGGTCTC
adapterA*-NH_2 (end of helix, C12)	/5AmMC12/GCGCTGACTGCTGGCTCGGCTTTTTTTTTT
adapterA*-PEG	/5Sp9//iSp18/GCGCTGACTGCTGGCTCGGCTTTTTTTTTT
adapterC-SS	/5ThioMC6-D/GCGCTGACTGCTGGCTCGGCTTTTTTTTTT
adapterC*-cholesterol	AAAAAGCCGAGCCAGCAGTCAGCGC/3CholTEG/

REFERENCES

1. Schrödinger, E. What is life? *Cambridge University Press* (1943).
2. Watson, J. & Crick, F. A structure for deoxyribose nucleic acid. *Nature* **171,** 737–738 (1953).
3. Ostwald, W. *Nobel lecture − on catalysis,* https://www.nobelprize.org/prizes/chemistry/1909/ostwald/lecture/. [Accessed online 29-April-2019]. (1909).
4. Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181,** 223–230 (1973).
5. Kruger, K. *et al.* Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* **31,** 147–157 (1982).
6. Weissman, K. J. The structural biology of biosynthetic megaenzymes. *Nature Chemical Biology* **11,** 660 (2015).
7. Awakawa, T. *et al.* Reprogramming of the antimycin NRPS-PKS assembly lines inspired by gene evolution. *Nature Communications* **9,** 3534 (2018).
8. Forster, A. C. & Church, G. M. Towards synthesis of a minimal cell. *Molecular Systems Biology* **2,** 45 (2006).
9. Benner, S. A. & Hutter, D. Phosphates, DNA, and the search for nonterrean life: A second generation model for genetic molecules. *Bioorganic Chemistry* **30,** 62–80 (2002).
10. Chamberlin, S. I. *et al.* Catalysis of amide synthesis by RNA phosphodiester and hydroxyl groups. *Proceedings of the National Academy of Sciences* **99,** 14688–14693 (2002).
11. Batey, R. T. *et al.* Tertiary motifs in RNA structure and folding. *Angewandte Chemie International Edition* **38,** 2326–2343 (1999).
12. Crick, F. *et al.* General nature of the genetic code for proteins. *Nature* **192** (1961).
13. Cech, T. R. The ribosome is a ribozyme. *Science* **289,** 878–879 (2000).
14. Drexler, E. Engines of creation: the coming era of nanotechnology. *Anchor* (1986).
15. Brenner, S. & Lerner, R. A. Encoded combinatorial chemistry. *Proceedings of the National Academy of Sciences* **89,** 5381–5383 (1992).
16. Goodnow Jr, R. A. *et al.* DNA-encoded chemistry: enabling the deeper sampling of chemical space. *Nature Reviews Drug Discovery* **16,** 131 (2017).
17. Halpin, D. R. & Harbury, P. B. DNA Display I. Sequence-encoded routing of DNA populations. *PLoS Biology* **2,** e173 (2004).
18. Halpin, D. R. & Harbury, P. B. DNA Display II. Genetic manipulation of combinatorial chemistry libraries for small-molecule evolution. *PLoS Biology* **2,** e174 (2004).

19. Halpin, D. R. *et al.* DNA Display III. Solid-phase organic synthesis on unprotected DNA. *PLoS Biology* **2,** e175 (2004).

20. Krusemark, C. J. *et al.* Directed chemical evolution with an outsized genetic code. *PloS One* **11,** e0154765 (2016).

21. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* **3,** 318–356 (1961).

22. Douglas, S. M. *et al.* A logic-gated nanorobot for targeted transport of molecular payloads. *Science* **335,** 831–834 (2012).

23. Zaslaver, A. *et al.* Just-in-time transcription program in metabolic pathways. *Nature Genetics* **36,** 486 (2004).

24. Zhang, F. *et al.* Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nature Biotechnology* **30,** 354 (2012).

25. Baum, E. B. Building an associative memory vastly larger than the brain. *Science* **268,** 583–585 (1995).

26. Cherry, K. M. & Qian, L. Scaling up molecular pattern recognition with DNA-based winner-take-all neural networks. *Nature* **559,** 370 (2018).

27. Demerec, M. & Fano, U. Bacteriophage-resistant mutants in *E. coli*. *Genetics* **30,** 119 (1945).

28. Durniak, K. J. *et al.* The structure of a transcribing T7 RNA polymerase in transition from initiation to elongation. *Science* **322,** 553–557 (2008).

29. Steitz, T. A. Structural biology: a mechanism for all polymerases. *Nature* **391,** 231 (1998).

30. Johnson, K. A. The kinetic and chemical mechanism of high-fidelity DNA polymerases. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **1804,** 1041–1048 (2010).

31. Ait-Haddou, R. & Herzog, W. Brownian ratchet models of molecular motors. *Cell Biochemistry and Biophysics* **38,** 191–213 (2003).

32. Guo, Q. & Sousa, R. Translocation by T7 RNA polymerase: a sensitively poised Brownian ratchet. *Journal of Molecular Biology* **358,** 241–254 (2006).

33. Harada, Y. *et al.* Direct observation of DNA rotation during transcription by *E. coli* RNA polymerase. *Nature* **409,** 113 (2001).

34. Proshkin, S. *et al.* Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* **328,** 504–508 (2010).

35. Kohler, R. *et al.* Architecture of a transcribing-translating expressome. *Science* **356,** 194–197 (2017).

36. Leppek, K. & Barna, M. An rRNA variant to deal with stress. *Nature Microbiology* **4,** 382 (2019).

37. Genuth, N. R. & Barna, M. The discovery of ribosome heterogeneity and its implications for gene regulation and organismal life. *Molecular Cell* **71,** 364–374 (2018).

38. Hopfield, J. J. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proceedings of the National Academy of Sciences* **71,** 4135–4139 (1974).

39. Zaher, H. S. & Green, R. Fidelity at the molecular level: lessons from protein synthesis. *Cell* **136,** 746–762 (2009).

40. Schmeing, T. M. *et al.* An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA. *Nature* **438,** 520 (2005).

41. Chen, J. *et al.* Coordinated conformational and compositional dynamics drive ribosome translocation. *Nature Structural and Molecular Biology* **20,** 718 (2013).

42. Noller, H. F. *et al.* The ribosome moves: RNA mechanics and translocation. *Nature Structural and Molecular Biology* **24,** 1021 (2017).

43. Weinger, J. S. *et al.* Substrate-assisted catalysis of peptide bond formation by the ribosome. *Nature Structural and Molecular Biology* **11,** 1101 (2004).

44. Bayryamov, S. G. *et al.* Unambiguous evidence for efficient chemical catalysis of adenosine ester aminolysis by its 2'-OH. *Journal of the American Chemical Society* **129,** 5790–5791 (2007).

45. Peacock, J. R. *et al.* Amino-acid dependent stability of the acyl linkage in aminoacyl-tRNA. *RNA* **20,** 758–764 (2014).

46. Crick, F. H. C. On protein synthesis. *Symposia of the Society for Experimental Biology* **12,** 8 (1958).

47. Rosenbaum, D. M. & Liu, D. R. Efficient and sequence-specific DNA-templated polymerization of peptide nucleic acid aldehydes. *Journal of the American Chemical Society* **125,** 13924–13925 (2003).

48. Switzer, C. *et al.* Enzymatic incorporation of a new base pair into DNA and RNA. *Journal of the American Chemical Society* **111,** 8322–8323 (1989).

49. Cozens, C. *et al.* A short adaptive path from DNA to RNA polymerases. *Proceedings of the National Academy of Sciences* **109,** 8067–8072 (2012).

50. Pinheiro, V. B. *et al.* Synthetic genetic polymers capable of heredity and evolution. *Science* **336,** 341–344 (2012).

51. Pinheiro, V. B. & Holliger, P. The XNA world: progress towards replication and evolution of synthetic genetic polymers. *Current Opinion in Chemical Biology* **16,** 245–252 (2012).

52. Nei, M. Gene duplication and nucleotide substitution in evolution. *Nature* **221,** 40 (1969).

53. Rackham, O. & Chin, J. W. A network of orthogonal ribosome-mRNA pairs. *Nature Chemical Biology* **1,** 159 (2005).

54. Neumann, H. *et al.* Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **464,** 441 (2010).

55. Santoro, S. W. *et al.* An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nature Biotechnology* **20,** 1044 (2002).

56. Wang, L. *et al.* Expanding the genetic code of *E. coli*. *Science* **292,** 498 (2001).

57. Murakami, H. *et al.* A versatile tRNA aminoacylation catalyst based on RNA. *Chemistry and Biology* **10,** 655–662 (2003).

58. Goto, Y. *et al.* Flexizymes for genetic code reprogramming. *Nature Protocols* **6,** 779 (2011).

59. Rogers, J. & Suga, H. Discovering functional, non-proteinogenic amino acid containing, peptides using genetic code reprogramming. *Organic and Biomolecular Chemistry* **13,** 9353–9363 (2015).

60. Fujino, T. *et al.* Ribosomal synthesis of peptides with multiple β-amino acids. *Journal of the American Chemical Society* **138,** 1962–1969 (2016).

61. Katoh, T. *et al.* Consecutive elongation of D-amino acids in translation. *Cell Chemical Biology* **24,** 46–54 (2017).

62. Milton, R. *et al.* Total chemical synthesis of a D-enzyme: the enantiomers of HIV-1 protease show reciprocal chiral substrate specificity. *Science* **256,** 1445–1448 (1992).

63. Jiang, W. *et al.* Mirror-image polymerase chain reaction. *Cell Discovery* **3,** 17037 (2017).

64. Pech, A. *et al.* A thermostable D-polymerase for mirror-image PCR. *Nucleic Acids Research* **45,** 3997–4005 (2017).

65. Wang, Z. *et al.* A synthetic molecular system capable of mirror-image genetic replication and transcription. *Nature Chemistry* **8,** 698 (2016).

66. Wang, M. *et al.* Mirror-image gene transcription and reverse transcription. *Chem* (2019).

67. Van Nies, P. *et al.* Self-replication of DNA by its encoded proteins in liposome-based synthetic cells. *Nature Communications* **9,** 1583 (2018).

68. Church, G. M. & Regis, E. Regenesis: how synthetic biology will reinvent nature and ourselves. *Basic Books* (2014).

69. Erbas-Cakmak, S. *et al.* Artificial molecular machines. *Chemical Reviews* **115,** 10081–10206 (2015).

70. Lewandowski, B. *et al.* Sequence-specific peptide synthesis by an artificial small-molecule machine. *Science* **339,** 189–193 (2013).

71. De Bo, G. *et al.* Sequence-specific β-peptide synthesis by a rotaxane-based molecular machine. *Journal of the American Chemical Society* **139,** 10875–10879 (2017).

72. Bo, G. *et al.* An artificial molecular machine that builds an asymmetric catalyst. *Nature Nanotechnology* **13,** 381 (2018).

73. Caboche, S. *et al.* NORINE: a database of nonribosomal peptides. *Nucleic Acids Research* **36,** D326–D331 (2007).

74. Al Ouahabi, A. *et al.* Synthesis of monodisperse sequence-coded polymers with chain lengths above DP100. *ACS Macro Letters* **4,** 1077–1080 (2015).

75. Gunay, U. S. *et al.* Chemoselective synthesis of uniform sequence-coded polyurethanes and their use as molecular tags. *Chem* **1,** 114–126 (2016).

76. König, N. F. *et al.* A simple post-polymerization modification method for controlling side-chain information in digital polymers. *Angewandte Chemie International Edition* **56,** 7297–7301 (2017).

77. Charles, L. *et al.* MS/MS-assisted design of sequence-controlled synthetic polymers for improved reading of encoded information. *Journal of The American Society for Mass Spectrometry* **28,** 1149–1159 (2017).

78. Gradišar, H. *et al.* Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments. *Nature Chemical Biology* **9,** 362 (2013).

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

80. Ke, Y. *et al.* Three-dimensional structures self-assembled from DNA bricks. *Science* **338,** 1177–1183 (2012).

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

82. Bath, J. & Turberfield, A. J. DNA nanomachines. *Nature Nanotechnology* **2,** 275 (2007).

83. Zhang, D. Y. & Seelig, G. Dynamic DNA nanotechnology using strand-displacement reactions. *Nature Chemistry* **3,** 103 (2011).

84. Lu, Y. & Liu, J. Functional DNA nanotechnology: emerging applications of DNAzymes and aptamers. *Current Opinion in Biotechnology* **17,** 580–588 (2006).

85. Amodio, A. *et al.* Rational design of pH-controlled DNA strand displacement. *Journal of the American Chemical Society* **136,** 16469–16472 (2014).

86. Liu, X. *et al.* Switchable reconfiguration of nucleic acid nanostructures by stimuli-responsive DNA machines. *Accounts of Chemical Research* **47,** 1673–1680 (2014).

87. You, M. *et al.* An autonomous and controllable light-driven DNA walking device. *Angewandte Chemie International Edition* **51,** 2457–2460 (2012).

88. Green, L. N. *et al.* Autonomous dynamic control of DNA nanostructure self-assembly. *Nature Chemistry,* 1 (2019).

89. Baccouche, A. *et al.* Dynamic DNA-toolbox reaction circuits: a walkthrough. *Methods* **67,** 234–249 (2014).

90. Dirks, R. M. & Pierce, N. A. Triggered amplification by hybridization chain reaction. *Proceedings of the National Academy of Sciences* **101,** 15275–15278 (2004).

91. Yurke, B. *et al.* A DNA-fuelled molecular machine made of DNA. *Nature* **406,** 605 (2000).

92. Bath, J. *et al.* A free-running DNA motor powered by a nicking enzyme. *Angewandte Chemie* **117,** 4432–4435 (2005).

93. Tian, Y. *et al.* A DNAzyme that walks processively and autonomously along a one-dimensional track. *Angewandte Chemie* **117,** 4429–4432 (2005).

94. Yin, P. *et al.* A unidirectional DNA walker that moves autonomously along a track. *Angewandte Chemie* **116,** 5014–5019 (2004).

95. Gartner, Z. J. & Liu, D. R. The generality of DNA-templated synthesis as a basis for evolving non-natural small molecules. *Journal of the American Chemical Society* **123,** 6961–6963 (2001).

96. Zhang, D. Y. & Winfree, E. Control of DNA strand displacement kinetics using toehold exchange. *Journal of the American Chemical Society* **131,** 17303–17314 (2009).

97. Srinivas, N. *et al.* On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic Acids Research* **41,** 10641–10658 (2013).

98. He, Y. & Liu, D. R. Autonomous multistep organic synthesis in a single isothermal solution mediated by a DNA walker. *Nature Nanotechnology* **5,** 778 (2010).

99. Anderson, G. W. *et al.* The use of esters of N-hydroxysuccinimide in peptide synthesis. *Journal of the American Chemical Society* **86,** 1839–1842 (1964).

100. Calderone, C. T. *et al.* Directing otherwise incompatible reactions in a single solution by using DNA-templated organic synthesis. *Angewandte Chemie International Edition* **41,** 4104–4108 (2002).

101. Wilcoxen, K. M. *et al.* Biomimetic catalysis of intermodular aminoacyl transfer. *Journal of the American Chemical Society* **129,** 748–749 (2007).

102. Kumar, J. K. & Oliver, J. S. Proximity effects in monolayer films: Kinetic analysis of amide bond formation at the air-water interface using 1H NMR spectroscopy. *Journal of the American Chemical Society* **124,** 11307–11314 (2002).

103. Tamura, K. & Schimmel, P. Oligonucleotide-directed peptide synthesis in a ribosome-and ribozyme-free system. *Proceedings of the National Academy of Sciences* **98,** 1393–1397 (2001).

104. Ficht, S. *et al.* Single-nucleotide-specific PNA- peptide ligation on synthetic and PCR DNA templates. *Journal of the American Chemical Society* **126,** 9970–9981 (2004).

105. Milnes, P. J. *et al.* Sequence-specific synthesis of macromolecules using DNA-templated chemistry. *Chemical Communications* **48,** 5614–5616 (2012).

106. Gartner, Z. J. *et al.* Multistep small-molecule synthesis programmed by DNA templates. *Journal of the American Chemical Society* **124,** 10304–10306 (2002).

107. Snyder, T. M. & Liu, D. R. Ordered multistep synthesis in a single solution directed by DNA templates. *Angewandte Chemie* **117,** 7545–7548 (2005).

108. Chen, Y. & Mao, C. Reprogramming DNA-directed reactions on the basis of a DNA conformational change. *Journal of the American Chemical Society* **126,** 13240–13241 (2004).

109. He, Y. & Liu, D. R. A sequential strand-displacement strategy enables efficient six-step DNA-templated synthesis. *Journal of the American Chemical Society* **133,** 9972–9975 (2011).

110. Keller, D. & Bustamante, C. The mechanochemistry of molecular motors. *Biophysical Journal* **78,** 541–556 (2000).

111. Erben, A. *et al.* DNA-instructed acyl transfer reactions for the synthesis of bioactive peptides. *Bioorganic and Medicinal Chemistry Letters* **21,** 4993–4997 (2011).

112. Ganai, R. A. & Johansson, E. DNA replication—a matter of fidelity. *Molecular Cell* **62,** 745–755 (2016).

113. Fan, K. & Wang, W. What is the minimum number of letters required to fold a protein? *Journal of Molecular Biology* **328,** 921–926 (2003).

114. Mulder, A. *et al.* Multivalency in supramolecular chemistry and nanofabrication. *Organic and Biomolecular Chemistry* **2,** 3409–3424 (2004).

115. McKee, M. L. *et al.* Multistep DNA-Templated Reactions for the Synthesis of Functional Sequence Controlled Oligomers. *Angewandte Chemie International Edition* **49,** 7948–7951 (2010).

116. Drew, H. R. *et al.* Structure of a B-DNA dodecamer: conformation and dynamics. *Proceedings of the National Academy of Sciences* **78,** 2179–2183 (1981).

117. Wang, E. *et al.* Solution structure of a pyrimidine-purine-pyrimidine triplex containing the sequence-specific intercalating non-natural base D3. *Journal of Molecular Biology* **257,** 1052–1069 (1996).

118. Gopaul, D. N. *et al.* Structure of the Holliday junction intermediate in Cre-loxP site-specific recombination. *The EMBO Journal* **17,** 4175–4187 (1998).

119. Luebke, K. J. & Dervan, P. B. Nonenzymatic ligation of oligodeoxyribonucleotides on a duplex DNA template by triple-helix formation. *Journal of the American Chemical Society* **111,** 8733–8735 (1989).

120. Illuminati, G. & Mandolini, L. Ring closure reactions of bifunctional chain molecules. *Accounts of Chemical Research* **14,** 95–102 (1981).

121. Singhal, A. *et al.* Toward peptide nucleic acid (PNA) directed peptide translation using ester based aminoacyl transfer. *ACS Chemical Biology* **9,** 2612–2620 (2014).

122. Catalano, M. J. *et al.* Effective molarity in a nucleic acid-controlled reaction. *Bioorganic and Medicinal Chemistry Letters* **26,** 2627–2630 (2016).

123. Raynal, M. *et al.* Supramolecular catalysis. Part 2: artificial enzyme mimics. *Chemical Society Reviews* **43,** 1734–1787 (2014).

124. Li, X. & Liu, D. R. DNA-templated organic synthesis: nature's strategy for controlling chemical reactivity applied to synthetic molecules. *Angewandte Chemie International Edition* **43,** 4848–4870 (2004).

125. O'Reilly, R. K. *et al.* The evolution of DNA-templated synthesis as a tool for materials discovery. *Accounts of Chemical Research* **50,** 2496–2509 (2017).

126. Gartner, Z. J. *et al.* DNA-templated organic synthesis and selection of a library of macrocycles. *Science* **305,** 1601–1605 (2004).

127. Tse, B. N. *et al.* Translation of DNA into a library of 13000 synthetic small-molecule macrocycles suitable for in vitro selection. *Journal of the American Chemical Society* **130,** 15611–15626 (2008).

128. Usanov, D. L. *et al.* Second-generation DNA-templated macrocycle libraries for the discovery of bioactive small molecules. *Nature chemistry* **10,** 704 (2018).

129. Dose, C. *et al.* Reducing product inhibition in DNA-template controlled ligation reactions. *Angewandte Chemie International Edition* **45,** 5369–5373 (2006).

130. Gothelf, K. V. *et al.* Modular DNA-programmed assembly of linear and branched conjugated nanostructures. *Journal of the American Chemical Society* **126,** 1044–1046 (2004).

131. Orgel, L. E. Unnatural selection in chemical systems. *Accounts of Chemical Research* **28,** 109–118 (1995).

132. Niu, J. *et al.* Enzyme-free translation of DNA into sequence-defined synthetic polymers structurally unrelated to nucleic acids. *Nature Chemistry* **5,** 282–292 (2013).

133. McKee, M. L. *et al.* Peptidomimetic bond formation by DNA-templated acyl transfer. *Organic and Biomolecular Chemistry* **9,** 1661–1666 (2011).

134. Montalbetti, C. A. & Falque, V. Amide bond formation and peptide coupling. *Tetrahedron* **61,** 10827–10852 (2005).

135. Schnell, A. *et al.* The mechanism of hydrolysis of phosphonium ylides. *Journal of the Chemical Society, Perkin Transactions 2,* 633–636 (1976).

136. Meng, W. *et al.* An autonomous molecular assembler for programmable chemical synthesis. *Nature Chemistry* **8,** 542–548 (2016).

137. Thubagere, A. J. *et al.* A cargo-sorting DNA robot. *Science* **357,** eaan6558 (2017).

138. Chen, Y.-J. *et al.* DNA nanotechnology from the test tube to the cell. *Nature Nanotechnology* **10,** 748 (2015).

139. Ackermann, D. *et al.* A double-stranded DNA rotaxane. *Nature Nanotechnology* **5,** 436 (2010).

140. Ju, J. *et al.* Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *Proceedings of the National Academy of Sciences* **103,** 19635–19640 (2006).

141. Wu, J. *et al.* 3'-O-modified nucleotides as reversible terminators for pyrosequencing. *Proceedings of the National Academy of Sciences* **104,** 16462–16467 (2007).

142. Guo, J. *et al.* Four-color DNA sequencing with 3'-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides. *Proceedings of the National Academy of Sciences* **105,** 9145–9150 (2008).

143. Hutter, D. *et al.* Labeled nucleoside triphosphates with reversibly terminating aminoalkoxyl groups. *Nucleosides, Nucleotides and Nucleic Acids* **29,** 879–895 (2010).

144. Turcatti, G. *et al.* A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. *Nucleic Acids Research* **36,** e25–e25 (2008).

145. Litosh, V. A. *et al.* Improved nucleotide selectivity and termination of 3'-OH unblocked reversible terminators by molecular tuning of 2-nitrobenzyl alkylated HOMedU triphosphates. *Nucleic Acids Research* **39,** e39–e39 (2011).

146. Gardner, A. F. *et al.* Rapid incorporation kinetics and improved fidelity of a novel class of 3'-OH unblocked reversible terminators. *Nucleic Acids Research* **40,** 7404–7415 (2012).

147. Tan, L. *et al.* Synthesis and evaluations of an acid-cleavable, fluorescently labeled nucleotide as a reversible terminator for DNA sequencing. *Chemical Communications* **52,** 2549–2552 (2016).

148. Tamura, K. & Schimmel, P. Peptide synthesis with a template-like RNA guide and aminoacyl phosphate adaptors. *Proceedings of the National Academy of Sciences* **100,** 8666–8669 (2003).

149. Saito, H. & Suga, H. A ribozyme exclusively aminoacylates the 3'-hydroxyl group of the tRNA terminal adenosine. *Journal of the American Chemical Society* **123,** 7178–7179 (2001).

150. Xiao, H. *et al.* Structural basis of specific tRNA aminoacylation by a small in vitro selected ribozyme. *Nature* **454,** 358 (2008).

151. Ramaswamy, K. *et al.* Designer ribozymes: programming the tRNA specificity into flexizyme. *Journal of the American Chemical Society* **126,** 11454–11455 (2004).

152. Upadhya, K. *et al.* Oxidation of biotin during oligonucleotide synthesis. *Nucleosides, Nucleotides and Nucleic Acids* **24,** 919–922 (2005).

153. Garlick, R. K. & Giese, R. Dissociative binding of α-and β-sulphoxides of biotinylamidoethyl-3-(4-hydroxy-3-[125I] iodophenyl) propionamide to avidin. *Biochemical Journal* **268,** 611–613 (1990).

154. Hytönen, V. P. *et al.* Structure and characterization of a novel chicken biotin-binding protein A (BBP-A). *BMC Structural Biology* **7,** 8 (2007).

155. Murata, H. *et al.* Rational tailoring of substrate and inhibitor affinity via ATRP polymer-based protein engineering. *Biomacromolecules* **15,** 2817–2823 (2014).

156. Zhang, Y. *et al.* Increasing enzyme cascade throughput by pH-engineering the microenvironment of individual enzymes. *ACS Catalysis* **7,** 2047–2051 (2017).

157. Klitzing, R. & Moehwald, H. Proton concentration profile in ultrathin polyelectrolyte films. *Langmuir* **11,** 3554–3559 (1995).

158. Church, G. *Onward DNA Nanotechnology!,* http://arep.med.harvard.edu/gmc/pers.html. [Accessed online 11-March-2019]. (2010).

159. Kallenbach, N. R. *et al.* An immobile nucleic acid junction constructed from oligonucleotides. *Nature* **305,** 829 (1983).

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

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

162. Shin, J.-S. & Pierce, N. A. A synthetic DNA walker for molecular transport. *Journal of the American Chemical Society* **126,** 10834–10835 (2004).

163. Sherman, W. B. & Seeman, N. C. A precisely controlled DNA biped walking device. *Nano Letters* **4,** 1203–1207 (2004).

164. Yurke, B. & Mills, A. P. Using DNA to power nanostructures. *Genetic Programming and Evolvable Machines* **4,** 111–122 (2003).

165. Genot, A. J. *et al.* Remote toehold: a mechanism for flexible control of DNA hybridization kinetics. *Journal of the American Chemical Society* **133,** 2177–2182 (2011).

166. Machinek, R. R. *et al.* Programmable energy landscapes for kinetic control of DNA strand displacement. *Nature Communications* **5,** 5324 (2014).

167. Nutiu, R. & Li, Y. Structure-switching signaling aptamers. *Journal of the American Chemical Society* **125,** 4771–4778 (2003).

168. Liu, D. & Balasubramanian, S. A proton-fuelled DNA nanomachine. *Angewandte Chemie International Edition* **42,** 5734–5736 (2003).

169. Liu, H. *et al.* Light-driven conformational switch of i-motif DNA. *Angewandte Chemie International Edition* **46,** 2515–2517 (2007).

170. Wang, X. *et al.* Conformational switching of G-quadruplex DNA by photoregulation. *Angewandte Chemie* **122,** 5433–5437 (2010).

171. Pan, J. *et al.* Recent progress on DNA based walkers. *Current opinion in biotechnology* **34,** 56–64 (2015).

172. Svoboda, K. & Block, S. M. Force and velocity measured for single kinesin molecules. *Cell* **77,** 773–784 (1994).

173. Santoro, S. W. & Joyce, G. F. A general purpose RNA-cleaving DNA enzyme. *Proceedings of the national academy of sciences* **94,** 4262–4266 (1997).

174. Turberfield, A. J. *et al.* DNA fuel for free-running nanomachines. *Physical Review Letters* **90,** 118102 (2003).

175. Green, S. J. *et al.* DNA hairpins: fuel for autonomous DNA devices. *Biophysical Journal* **91,** 2966–2975 (2006).

176. Venkataraman, S. *et al.* An autonomous polymerization motor powered by DNA hybridization. *Nature Nanotechnology* **2,** 490–494 (2007).

177. Yin, P. *et al.* Programming biomolecular self-assembly pathways. *Nature* **451,** 318 (2008).

178. Olson, A. C. *et al.* The energetic difference between synthesis of correct and incorrect base pairs accounts for highly accurate DNA replication. *Journal of the American Chemical Society* **135,** 1205–1208 (2013).

179. Westheimer, F. H. Why nature chose phosphates. *Science* **235,** 1173–1178 (1987).

180. Marras, S. A. *et al.* Efficiencies of fluorescence resonance energy transfer and contact-mediated quenching in oligonucleotide probes. *Nucleic Acids Research* **30,** e122–e122 (2002).

181. Song, L. *et al.* Photobleaching kinetics of fluorescein in quantitative fluorescence microscopy. *Biophysical Journal* **68,** 2588–2600 (1995).

182. Kuchta, R. *et al.* Kinetic mechanism of DNA polymerase I (Klenow). *Biochemistry* **26,** 8410–8417 (1987).

183. Joyce, C. M. Techniques used to study the DNA polymerase reaction pathway. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **1804,** 1032–1040 (2010).

184. Chatterjee, G. *et al.* A spatially localized architecture for fast and modular DNA computing. *Nature Nanotechnology* **12,** 920 (2017).

185. Johnson, S. J. *et al.* Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations. *Proceedings of the National Academy of Sciences* **100,** 3895–3900 (2003).

186. Gu, H. *et al.* A proximity-based programmable DNA nanoscale assembly line. *Nature* **465,** 202 (2010).

187. Strauss, M. T. *et al.* Quantifying absolute addressability in DNA origami with molecular resolution. *Nature Communications* **9,** 1600 (2018).

188. Snyder, T. M. *et al.* Effects of template sequence and secondary structure on DNA-templated reactivity. *Journal of the American Chemical Society* **130,** 1392–1401 (2008).

189. De Bo, G. *et al.* Efficient assembly of threaded molecular machines for sequence-specific synthesis. *Journal of the American Chemical Society* **136,** 5811–5814 (2014).

190. Burke, H. M. *et al.* Exploring chemoselective S-to-N acyl transfer reactions in synthesis and chemical biology. *Nature Communications* **8,** 15655 (2017).

191. Panda, S. S. *et al.* Traceless chemical ligation from S-, O-, and N-acyl isopeptides. *Accounts of Chemical Research* **47,** 1076–1087 (2014).

192. Goodsell, D. *Titin: Molecule of the month* http://https://pdb101.rcsb.org/motm/185l. [Accessed online 09-May-2019]. (2015).

193. Finking, R. & Marahiel, M. A. Biosynthesis of nonribosomal peptides. *Annual Reviews of Microbiology* **58,** 453–488 (2004).

194. Gindulyte, A. *et al.* The transition state for formation of the peptide bond in the ribosome. *Proceedings of the National Academy of Sciences* **103,** 13327–13332 (2006).

195. Agmon, I. *et al.* Symmetry at the active site of the ribosome: structural and functional implications. *Biological Chemistry* **386,** 833–844 (2005).

196. McKee, M. L. *et al.* Programmable one-pot multistep organic synthesis using DNA junctions. *Journal of the American Chemical Society* **134,** 1446–1449 (2012).

197. Rudinger, J. *et al.* Minimalist aminoacylated RNAs as efficient substrates for elongation factor Tu. *Biochemistry* **33,** 5682–5688 (1994).

198. Achenbach, J. *et al.* Outwitting EF-Tu and the ribosome: translation with D-amino acids. *Nucleic Acids Research* **43,** 5687–5698 (2015).

199. Bruick, R. K. *et al.* Template-directed ligation of peptides to oligonucleotides. *Chemistry and Biology* **3,** 49–56 (1996).

200. Lohse, P. A. & Szostak, J. W. Ribozyme-catalysed amino-acid transfer reactions. *Nature* **381,** 442 (1996).

201. Hansen, M. H. *et al.* A yoctoliter-scale DNA reactor for small-molecule evolution. *Journal of the American Chemical Society* **131,** 1322–1327 (2009).

202. Tarini, M. *et al.* Ambient occlusion and edge cueing for enhancing real time molecular visualization. *IEEE Transactions on Visualization and Computer Graphics* **12,** 1237–1244 (2006).

203. Hu, Y. *et al.* Triplex DNA nanostructures: from basic properties to applications. *Angewandte Chemie International Edition* **56,** 15210–15233 (2017).

204. De Stefano, M. & Vesterager Gothelf, K. Dynamic chemistry of disulfide terminated oligonucleotides in duplexes and double-crossover tiles. *ChemBioChem* **17,** 1122–1126 (2016).

205. Kilchherr, F. *et al.* Single-molecule dissection of stacking forces in DNA. *Science* **353,** aaf5508 (2016).

206. Ng, D. T. & Sarkar, C. A. NP-Sticky: a web server for optimizing DNA ligation with non-palindromic sticky ends. *Journal of Molecular Biology* **426,** 1861–1869 (2014).

207. Zadeh, J. N. *et al.* NUPACK: analysis and design of nucleic acid systems. *Journal of Computational Chemistry* **32,** 170–173 (2011).

208. Dupuis, N. F. *et al.* Single-molecule kinetics reveal cation-promoted DNA duplex formation through ordering of single-stranded helices. *Biophysical Journal* **105,** 756–766 (2013).

209. Chen, J. *et al.* DNA oligonucleotides with A, T, G or C opposite an abasic site: structure and dynamics. *Nucleic Acids Research* **36,** 253–262 (2007).

210. Kandimalla, E. R. & Agrawal, S. 'Cyclicons' as hybridization-based fluorescent primer-probes: synthesis, properties and application in real-time PCR. *Bioorganic and Medicinal Chemistry* **8,** 1911–1916 (2000).

211. Chen, H. *et al.* Template-directed chemical ligation to obtain 3'-3' and 5'-5' phosphodiester DNA linkages. *Scientific Reports* **4,** 4595 (2014).

212. Kandimalla, E. R. & Agrawal, S. Hoogsteen DNA duplexes of 3'-3'-and 5'-5'-linked oligonucleotides and triplex formation with RNA and DNA pyrimidine single strands: experimental and molecular modeling studies. *Biochemistry* **35,** 15332–15339 (1996).

213. Yu, D. *et al.* 'Immunomers': novel 3'-3'-linked CpG oligodeoxyribonucleotides as potent immunomodulatory agents. *Nucleic Acids Research* **30,** 4460–4469 (2002).

214. Omabegho, T. *et al.* A bipedal DNA Brownian motor with coordinated legs. *Science* **324,** 67–71 (2009).

215. El-Sagheer, A. H. & Brown, T. Click chemistry with DNA. *Chemical Society Reviews* **39,** 1388–1405 (2010).

216. Li, X. *et al.* DNA-catalyzed polymerization. *Journal of the American Chemical Society* **124,** 746–747 (2002).

217. Iqbal, A. *et al.* Orientation dependence in fluorescent energy transfer between Cy3 and Cy5 terminally attached to double-stranded nucleic acids. *Proceedings of the National Academy of Sciences* **105,** 11176–11181 (2008).

218. Kamerlin, S. C. *et al.* Why nature really chose phosphate. *Quarterly Reviews of Biophysics* **46,** 1–132 (2013).

219. Peacock, J. R. *et al.* Amino acid dependent stability of the acyl linkage in aminoacyl-tRNA. *RNA* **20,** 758–764 (2014).

220. Rozenman, M. M. & Liu, D. R. DNA-templated synthesis in organic solvents. *ChemBioChem* **7,** 253–256 (2006).

221. Gallego, I. *et al.* Folding and imaging of DNA nanostructures in anhydrous and hydrated deep-eutectic solvents. *Angewandte Chemie International Edition* **54,** 6765–6769 (2015).

222. Gartner, Z. J. *et al.* Two enabling architectures for DNA-templated organic synthesis. *Angewandte Chemie International Edition* **42,** 1370–1375 (2003).

223. Yoshimoto, K. *et al.* Use of abasic site-containing DNA strands for nucleobase recognition in water. *Journal of the American Chemical Society* **125,** 8982–8983 (2003).

224. Sankaran, N. *et al.* Small-molecule binding at an abasic site of DNA: strong binding of lumiflavin for improved recognition of thymine-related single nucleotide polymorphisms. *The Journal of Physical Chemistry B* **113,** 1522–1529 (2009).

225. Wang, Y. *et al.* Recognition of DNA abasic site nanocavity by fluorophore-switched probe: Suitable for all sequence environments. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **153,** 645–650 (2016).

226. Valero, J. *et al.* A bio-hybrid DNA rotor–stator nanoengine that moves along predefined tracks. *Nature Nanotechnology* **13,** 496 (2018).

227. Garg, S. *et al.* Directed enzymatic activation of 1D DNA tiles. *ACS Nano* **9,** 1072–1079 (2015).

228. Dawson, P. E. *et al.* Synthesis of proteins by native chemical ligation. *Science* **266,** 776–779 (1994).

229. Wilner, O. I. *et al.* Enzyme cascades activated on topologically programmed DNA scaffolds. *Nature Nanotechnology* **4,** 249 (2009).

230. Spruijt, E. *et al.* DNA scaffolds support stable and uniform peptide nanopores. *Nature Nanotechnology* **13,** 739 (2018).

231. Schreiber, R. *et al.* Hierarchical assembly of metal nanoparticles, quantum dots and organic dyes using DNA origami scaffolds. *Nature Nanotechnology* **9,** 74 (2014).

232. Zhang, Z. *et al.* Placing and shaping liposomes with reconfigurable DNA nanocages. *Nature Chemistry* **9,** 653 (2017).

233. Todhunter, M. E. *et al.* Programmed synthesis of three-dimensional tissues. *Nature Methods* **12,** 975 (2015).

234. Hughes, A. J. *et al.* Engineered tissue folding by mechanical compaction of the mesenchyme. *Developmental Cell* **44,** 165–178 (2018).

235. Praetorius, F. *et al.* Biotechnological mass production of DNA origami. *Nature* **552,** 84 (2017).

236. Gerling, T. *et al.* Sequence-programmable covalent bonding of designed DNA assemblies. *Science Advances* **4,** eaau1157 (2018).

237. Stross, A. E. *et al.* Cooperative duplex formation by synthetic H-bonding oligomers. *Chemical Science* **7,** 94–101 (2016).

238. Iadevaia, G. *et al.* Backbone conformation affects duplex initiation and duplex propagation in hybridisation of synthetic H-bonding oligomers. *Organic and Biomolecular Chemistry* **16,** 4183–4190 (2018).

239. Steinberg-Yfrach, G. *et al.* Light-driven production of ATP catalysed by F0 F1-ATP synthase in an artificial photosynthetic membrane. *Nature* **392,** 479 (1998).

240. Zadorin, A. S. *et al.* Synthesis and materialization of a reaction–diffusion French flag pattern. *Nature Chemistry* **9,** 990 (2017).

241. Brandman, O. & Hegde, R. S. Ribosome-associated protein quality control. *Nature Structural and Molecular Biology* **23,** 7 (2016).

242. Cello, J. *et al.* Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. *Science* **297,** 1016–1018 (2002).

243. Gibson, D. G. *et al.* Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329,** 52–56 (2010).

244. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6,** 343 (2009).

245. Murakami, H. *et al.* A highly flexible tRNA acylation method for non-natural polypeptide synthesis. *Nature Methods* **3,** 357 (2006).

246. Weisstein, E. W. Sphere-sphere intersection. *Wolfram Research, Inc.* (2007).