

Nanoscale patterning of complex DNA
structures with the bacterial protein
Recombinase A

Sybilla Louise Corbett

Submitted in accordance with the requirements for the degree of Doctor
of Philosophy

The University of Leeds
Centre for Doctoral Training in Molecular-Scale Engineering
School of Electronic and Electrical Engineering

February 2016

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

©2016 The University of Leeds and Sybilla Louise Corbett

For Mum and Dad

Acknowledgements

This work was made possible by the funding of the Centre for Doctoral Training in Molecular-Scale Engineering by the Engineering and Physical Sciences Research Council, for which I am very thankful.

I am deeply grateful for the help of Professors Christoph Wälti and Giles Davies, in their capacity as both CDT Directors and as my supervisors. I am also very grateful for the suggestions and assistance of the other senior members of the Bioelectronics lab, especially Dr. Rajan Sharma and Dr. Michal Szymonik. I would also like to thank the Sheffield CDT Directors, Professors Jamie Hobbs and Graham Leggett, for their support. Thanks to Katherynne Taylor for keeping the CDT ticking over.

I would also like to thank the other students on the CDT in Molecular-Scale Engineering and in the Bioelectronics lab for their friendship and aid; without them, these years would have been grey and my grasp of beam equations weak.

Special thanks to my Mum and my sister for their patience and encouragement. Finally, I thank Francis Galaway, my tame Postdoc, my rambling man, for all his love and support.

Abstract

The use of DNA as a structural material has been intensively developed since its inception in the early 1980s. The potential of DNA structures in the field of materials science is hampered by current approaches to augmentation. It is not currently possible to alter the targeting of heterogenous additional elements to structures once they have been made. The post hoc patterning of DNA architectures is therefore of great importance. The bacterial protein Recombinase A (RecA) may be able to provide this function.

This thesis will discuss the patterning of DNA structures with RecA. RecA has been shown to pattern linear dsDNA strands with high levels of efficiency. To test the potential of RecA to pattern more complex DNA, novel strategies for creating DNA topologies have been explored. This work has produced DNA strands containing regions of base pair mismatching and with terminal three-way junctions.

A method has also been developed for the creation of a 200 base product with unpaired branched junctions, using four synthetic oligomers in a scaffolded cycling ligation reaction with a heat stable ligase. A method to create longer DNA strands with three-way junctions at the termini has also been developed.

RecA patterning of a structure with internal mismatches was carried out. Mismatches proximal to the patterning area led to an increase in patterning efficiency with an increase in mismatch length. When the mismatch was separated from the patterning region a more complex relationship was observed, with intermediate-length mismatches resulting in a decrease in patterning efficiency. The introduction of a nick in the phosphate backbone proximal to the patterning region also increased patterning efficiency.

Two further DNA structures were produced on which patterning did not prove possible. The ligase chain reaction was shown to produce DNA strands that could be incorporated into a structure with central base pairing and terminal single stranded DNA regions. Attempts to create three-way junctions from these structures were not successful. A second structure was created through treatment of double stranded DNA from the polymerase chain reaction. Single strands of DNA were produced that could be annealed to produce terminal three-way junctions. Atomic force microscopy demonstrated the correct annealing of this structure. However, it did not prove possible to pattern these

structures with RecA.

Recombinant RecA production through bacterial induction produced soluble protein at a high yield. There was some evidence of DNA contamination and the purified protein showed low activity.

Contents

Dedication	iii
Acknowledgements	iv
Abstract	v
Contents	vii
List of figures	xiii
List of tables	xix
1 Introduction	1
1.1 The Importance of the Nanoscale	1
1.2 Aim	4
2 Background and Literature Review	7
2.1 Deoxyribonucleic Acid (DNA)	7
2.1.1 Protein biology	11
2.1.2 Structures in Nature	12
2.2 Creating complex DNA structures	14
2.2.1 2D structures	16
2.2.2 3D structures	22
2.2.3 Structures with switchable elements	24

2.2.4	Visualisation	28
2.2.5	Sources of DNA	29
2.3	Strategies for functionalising DNA structures	30
2.3.1	Base substitutions	30
2.3.2	Functionalising bases	32
2.3.3	DNA toeholds	33
2.3.4	DNA aptamers	36
2.3.5	Small molecules	37
2.4	Proteins in DNA nanotechnology	38
2.5	RecA	41
2.5.1	RecA function	41
2.5.2	RecA structure	43
2.5.3	Formation of a nucleoprotein filament	45
2.5.4	Filament extension	46
2.5.5	Triple strand formation	47
2.5.6	Factors affecting patterning	48
2.5.7	RecA in nanotechnology	50
2.5.8	Conclusion	53
3	Techniques, Protocols and Reagents	55
3.1	Techniques	56
3.1.1	Polymerase chain reaction	56
3.1.2	Ligase Chain reaction	57
3.1.3	DNA purification	58

3.1.4	Gel electrophoresis	60
3.1.5	Digestion	64
3.1.6	Fabricating ssDNA using T7 Exonuclease	64
3.1.7	RecA patterning	65
3.1.8	AFM	66
3.1.9	Molecular cloning	67
3.1.10	Cell lines	69
3.2	Protocols	71
3.2.1	DNA Structure Fabrication	71
3.2.2	270 base structure	75
3.2.3	Three-way junction annealing	79
3.2.4	Gel Electrophoresis	79
3.2.5	Diagnostic Digestion	83
3.2.6	AFM Visualisation	83
3.2.7	RecA Patterning	84
3.2.8	Protein production	86
3.3	Reagents	91
4	RecA patterning of 100 base structures containing base mismatches	95
4.1	Synthesis of DNA scaffolds	96
4.1.1	Design	96
4.1.2	Annealing DNA Sequences	100
4.2	Patterning proximal mismatch structures	105
4.3	Patterning 12 base mismatches	108

4.3.1	Annealing	108
4.3.2	Patterning	109
4.4	Patterning central mismatch structures	111
4.5	Patterning nicked structures	115
4.5.1	Annealing	115
4.5.2	Patterning	116
4.6	Exploring the effect of patterning oligomer complemetarity	118
4.7	Controls	120
4.8	Conclusions	121
5	RecA patterning of 200 base structures with terminal ssDNA regions	125
5.1	Fabrication	127
5.1.1	Design	127
5.1.2	Synthesis	129
5.1.3	Characterisation	131
5.1.4	Optimisation	135
5.1.5	Purification of the DNA constructs	139
5.1.6	Heat stability	143
5.2	Junction Annealing	145
5.2.1	Left (5') annealing	145
5.2.2	Right (3') annealing	146
5.2.3	Annealing oligomers to the left (5') and right (3') junctions	147
5.2.4	LCR with modified oligomers	149
5.3	Patterning	150

5.3.1	Digestion enzyme trials	150
5.3.2	Patterning	151
5.4	Conclusions	153
6	Fabrication of 270 base structures for RecA patterning	155
6.1	Synthesis	157
6.1.1	Design	157
6.1.2	Gene synthesis and DNA fabrication	158
6.1.3	T7 Exonuclease Treatment	161
6.1.4	Anneling three-way junctions	164
6.2	Atomic Force Microscopy	165
6.3	Validating structures through restriction enzyme digest	168
6.4	Patterning	171
6.5	Conclusions	175
7	Expression and purification of the bacterial protein RecA	177
7.1	Cloning	178
7.2	Expression Trials	181
7.2.1	First expression trial	181
7.2.2	Second expression trial	183
7.3	Protein Production and Purification	185
7.3.1	pET23a plasmid	186
7.3.2	pET15b plasmid	186
7.4	Refining the purification protocol	189
7.4.1	ATP wash	191

7.4.2	Addition of DTT	194
7.5	Activity assay	196
7.6	Conclusions	198
8	Discussion	201
8.1	Addressing DNA nanotechnology	201
8.2	Results	203
8.2.1	100 base structures	204
8.2.2	200 base structures	205
8.2.3	270 base structures	207
8.2.4	RecA production	209
8.2.5	Conclusions	210
8.3	Future Outlook	211
	Appendix	215
A	DNA Sequences	217
A.1	100 base sequences	217
A.2	200 base sequences	222
A.3	270 base sequences	225
B	Protein Expression	227
B.1	RecA DNA Sequence	227
B.2	RecA Protein Sequence	228
B.3	Plasmid Maps	229
	Bibliography	233

List of Figures

1.1	Diagrammatic representation of the patterning of dsDNA with RecA and a homologous ssDNA strand.	4
1.2	Schematic diagram of the structures used to assess RecA patterning of complex topologies	5
2.1	The structure of DNA; a bond diagram of the bases and hydrogen bonding and the double helix formed by two base-pairing strands of DNA	8
2.2	The transcription and translation of nuclear DNA to produce protein in the eukaryotic cell	11
2.3	Formation of RNA nanostructures through co-transcriptional folding	15
2.4	The formation of DNA double crossover motifs and the use of these motifs to form edge sharing triangles	18
2.5	Complex DNA structures in two dimensions	19
2.6	A selection of the first structures created using DNA origami	20
2.7	Complex DNA structures in three dimensions	23
2.8	DNA based molecular tweezers	25
2.9	A clam-like DNA structure encapsulating a cargo	26
2.10	Programming and imaging a DNA walker	27
2.11	DNA base pair substitute and Watson-Crick base pairing cytosine-guanine for comparison	32

2.12	Using the hybridisation of DNA to create a lattice and arrange gold nanoparticles	35
2.13	A series of diagrammatic images showing small molecule intercalation in the grooves of a DNA double helix structure.	38
2.14	Modular zinc-finger nucleases (ZFNs) can be used to direct functional groups to specific DNA sequences	41
2.15	Diagrammatic representation of RecA patterning showing monomers forming a nucleoprotein filament with ssDNA followed by binding to a homologous dsDNA strand	42
2.16	Cartoon diagram showing the secondary structural elements of the RecA monomer	44
2.17	The use of RecA for the formation of nanowires	51
3.1	Diagrammatic representation of the ligase chain reaction	59
3.2	Diagram of polyacrylamide gel electrophoresis	61
3.3	The acrylamide and agarose molecules	61
3.4	Diagrammatic representation of phosphorothioate modification of DNA.	65
3.5	The sequence of the RecA gene and its translation into a protein primary structure	86
4.1	Two ssDNA oligomers are annealed to produce the 100 base structure	97
4.2	A diagrammatic representation of the 100 base structures used in this work	98
4.3	A series of 100 base structures are synthesised and characterised using PAGE	102
4.4	AFM imaging of 100 base structures and analysis of representative images	103
4.5	Analysis of AFM images shows distribution of angles for 100 base structures with a 20 base mismatch	104

4.6	Diagrammatic representation of the 100 base proximal series of structures	106
4.7	Patterning experiment on the 100 base proximal structures	107
4.8	Diagrammatic representation of the 100 base 12 mismatch series of structures	108
4.9	Patterning analysis of a series of 12 base mismatches in the 100 base structure	110
4.10	Diagrammatic representation of the centrally radiating 100 base structures	111
4.11	PAGE and densitometry analysis of patterning experiments at different nucleoprotein filament to dsDNA ratios on the centrally radiating structures on the 100 base structure	112
4.12	Characterisation of fully base-paired 100 base structures containing a break in the phosphate backbone	116
4.13	Analysis of patterning efficiency of RecA nucleoprotein filaments on DNA structures containing mismatches or a nick	117
4.14	Analysis of patterning experiments using the reverse complement of the patterning oligomer for 100 base constructs	119
4.15	PAGE analysis of a patterning experiment carried out with heat treated RecA and of patterned and deprotected dsDNA	120
4.16	Diagrammatic representation of the release of RecA patterning-induced tension in dsDNA by non-base pairing regions	123
5.1	Diagrammatic representations of the structures produced by the ligase chain reaction and of the addition of oligomers to unpaired branched junctions for the creation of three-way junctions	126
5.2	Diagrammatic representations of the ligase chain reaction	127
5.3	PAGE analysis of oligomers for synthesis of the 200 base structure through the ligase chain reaction	129

5.4	PAGE analysis of the products of the ligase chain reaction	130
5.5	PAGE analysis of the digestion of cycling ligation reaction products and of the 140 bp scaffold	132
5.6	Identification of the cycling ligation product migration on native acrylamide through ligation experiments excluding a series of oligomers .	134
5.7	Identification of additional cycling ligation products with a single or no unpaired branched junction	135
5.8	PAGE analysis of an experiment to show the effect of additional cycles and additional ligase in the cycling ligation reaction	137
5.9	PAGE analysis of the effect of a range of hybridisation temperatures on the yield of cycling ligation product	138
5.10	PAGE analysis of the effect of higher oligomer concentration on ligase chian reaction yield of 200 base product	139
5.11	PAGE analysis of ligation products following attempted purification with the Roche High Pure PCR Purification Kit	140
5.12	Agarose gel analysis of the products of the cycling ligation reaction . . .	141
5.13	PAGE analysis of the agarose gel purified products of the cycling ligation reaction	141
5.14	Alkaline agarose gel electrophoresis of the products of the cycling ligation reaction	142
5.15	PAGE analysis of samples from the cycling ligation reaction following alkaline agarose gel electrophoresis and purification	143
5.16	PAGE analysis of the effect of magnesium concentration and temperature on the 200bp LCR product	144
5.17	PAGE analysis of the effect of a smaller range of temperatures on the 200 bp LCR product in the presence of magnesium	144

5.18	PAGE analysis of the effect of hybridising an oligomer to the unpaired branched junction of the structure with a 5' junction	146
5.19	PAGE analysis of the effect of hybridising an oligomer to the unpaired branched junction of the structure with a 3' junction	147
5.20	PAGE analysis of the effect of hybridising an oligomer to the unpaired branched junction of the structure with both 5' and 3' junctions	148
5.21	Attempted cycling ligations using oligomers with fewer nucleotides at the junction of the unpaired regions at the 3' junction point	149
5.22	PAGE analysis of XapI digestion of 623 base dsDNA	151
5.23	PAGE analysis of RecA patterning of 140 base dsDNA and the 200 base structure with unpaired branched junctions at both termini	152
6.1	Diagrammatic representation of the process for producing 270 base structures with three-way junctions at the termini	156
6.2	PCR product of the reaction to produce the central 170 bases of the 270 base structure	158
6.3	PCR products of the reactions to produce the 289 bp precursors to the 270 base structure	159
6.4	Restriction assay to assess PCR product for the 289 base dsDNA precursor to the upper strand of the 270 base structure	160
6.5	Restriction assay confirmation of correct PCR product for the 289 base dsDNA precursor to the lower strand of the 270 base structure	161
6.6	Analysis of the treatment of PCR product to create 270 base ssDNA	162
6.7	Incomplete BmtI digestion of the lower 270 base ssDNA strand	163
6.8	Annealing of ssDNA from T7 treatment and the creation of three-way junctions using additional oligomers	165
6.9	AFM images of the 270 base structures obtained using Peak Force Tapping	166

6.10	AFM images of a single representative particle of the 270 base structures obtained using Peak Force Tapping	167
6.11	PAGE analysis of the four 270 base constructs in patterning buffer and digested with XapI	169
6.12	PAGE analysis of an XapI digest of the 170 bp DNA using a series of enzyme dilutions	170
6.13	Patterning experiment at a 30:1 and 40:1 ratio on 170 bp DNA	171
6.14	Patterning 289 bp dsDNA with the patterning oligomers Adjacent and Proximal	173
6.15	Patterning 289 bp dsDNA with a series of 30 base patterning oligomers	174
6.16	Patterning 289 bp dsDNA with a series of 30 and 40 base patterning oligomers	174
7.1	Digested plasmid and inserts analysed by acrylamide gel electrophoresis	179
7.2	Agarose gel electrophoresis of plasmids with ligated inserts amplified in <i>E. coli</i>	180
7.3	First expression trial of RecA protein	182
7.4	Second expression trial of RecA protein	183
7.5	FPLC chromatogram for the purification of wild-type RecA in the pET23a plasmid, with a C terminal His-tag	187
7.6	FPLC chromatogram for the purification of wild-type RecA in the pET15b plasmid, with a N terminal His-tag	188
7.7	FPLC chromatogram for the refined protocol purification of wild-type RecA with an N terminal His-tag	190
7.8	FPLC chromatogram for the 10 mM ATP wash of the His-trap column with bound RecA	191

7.9	FPLC chromatogram of the purification of wild-type RecA with a N terminal His-tag using an updated protocol and an ATP wash	193
7.10	Surface of the RecA monomer showing cysteine accessibility	194
7.11	FPLC chromatogram for the refined protocol purification of wild-type RecA with a N terminal His-tag following the addition of DTT to buffers used in the protocol	195
7.12	Agarose gel electrophoresis showing a RecA activity assay	197
B.1	pET11a plasmid map	229
B.2	pET11a cloning/expression region	229
B.3	pET15b plasmid map	230
B.4	pET15b cloning/expression region	230
B.5	pET23a plasmid map	231
B.6	pET23a cloning/expression region	231

List of Tables

2.1	Reported properties of the <i>Escherichia coli</i> enzyme Recombinase A	49
3.1	PCR reaction mixture for the amplification of the 140bp template from the genome of the λ bacteriophage.	72
3.2	PCR cycle for the amplification of the 140 bp template from the genome of the λ bacteriophage.	72
3.3	LCR reaction mixture to create 200bp structures with ssDNA elements . .	74
3.4	LCR cycle to create 200bp structures with ssDNA elements	74
3.5	PCR reaction mixture for the amplification of the 270 base sequence from the synthetic gene.	76
3.6	PCR cycle for the amplification of the 270 base phosphorothioated products from the synthetic genes containing the template DNA sequence.	76
3.7	T7 exonuclease treatment of phosphorothioated PCR products	77
3.8	BmtI-HF digestion of phosphorothioated ssDNA	78
3.9	Native acrylamide gel composition	80
3.10	SDS acrylamide gel	82
3.11	Scan rates for capture of AFM images of 100 and 270 base structures . .	84
3.12	dsDNA patterning protocol	85
3.13	Reagents from New England Biolabs	91

3.14	Restriction enzymes	91
3.15	Reagents from Sigma-Aldrich	92
3.16	Reagents from Sigma-Aldrich	93
3.17	Additional reagents	93
4.1	Protection efficiency for structures containing base pair mismatches . . .	113
7.1	Expected sizes of the generated constructs following ligation of the insert into the plasmid	180

Chapter 1

Introduction

1.1 The Importance of the Nanoscale

The importance of the nanoscale in the development of science and technology has been acknowledged for more than 55 years [1]. In his 1959 talk 'There is plenty of room at the bottom', Richard Feynman envisaged a future where the processes of science and technology would be carried out at the nanoscale, an idea that has become increasingly tangible. This is because the development of materials and devices at the nanoscale could have a significant effect on a large number of fields.

The fabrication of progressively smaller devices has already had an enormous impact on a number of areas, for example an increase in available computing power by several orders of magnitude over the last few decades. This has been achieved through leaps forward in lithographic and deposition technology. Furthermore, other alternative technologies for the nanofabrication of materials have started to impact on the field of materials science and technology.

Having surpassed the diffraction limit of light with technologies such as dip-pen nanolithography, ever more minute and complex devices have been created, which are expected to be translated into industry over the next decade. There are many areas in which increased understanding and manipulation of materials on the nanoscale could have a huge impact. For example, the effect of nanotechnology on the field of medicine, where it has provided new ways to deliver active compounds [2], image the body [3] and design

materials [4].

One of the aims of nanotechnology is to create structures using a bottom-up approach, as opposed to the top-down approach currently used. The difference between top-down and bottom-up fabrication can be described thus: in trying to make gold nanoparticles, a top-down approach would attempt to remove pieces of gold from a larger block. A bottom-up approach to the same problem might attempt to precipitate gold from solution, or to deposit particles in a fine mist, growing the particles atom by atom. The level of control is thus greatly increased, less material is wasted, making the process more economical, and more complex, novel materials can be developed. It is thought that characterisation on the nanoscale could lead to novel properties in new and well-known materials, such as those of graphene. Identified in 2004 [5] there are a wealth of articles about the potential uses of graphene [3, 6, 7].

The next step in advanced material research is to develop adaptable scaffolds for functionalisation. Through precise control of positions and distancing it could be possible to create highly complex functionalised surfaces with novel applications. Control of enzymatic, photonic and other species could facilitate unique activities and new methods for reporting on processes at the nanoscale.

Biological fabrication processes are already carried out at the nanoscale and demonstrate the adaptability materials research desires. For this reason biological molecules have been identified as potentially useful for developing entirely new systems in engineering and materials science [8, 9]. The molecules in question have been optimised over millions of years for a role in living organisms and the objective of the materials researcher is to bend their function to serve a new purpose.

There are many examples of materials in nature that form larger structures using nanoscale components. Actin and tubulin form networks in cells, increasing in size through the addition of monomers to make structures that direct the shape and movement of the cell body, as well as facilitating the movement of substrates through the cytoplasm.

One of the materials being explored is deoxyribonucleic acid (DNA). DNA has been identified as a promising material for the development of nanoscale architectures owing to the ease with which it is possible to design and create specific sequences that occupy

a known space in the structure [10]. The prediction of 3D structures for novel proteins is inaccurate, while other polymers are too repetitive for localised ad hoc additions at the resolution of a nucleotide base.

To this end the field of DNA nanotechnology has been developed. From its early inception as a method for arranging proteins to create crystals [11], DNA structures have been imagined as a flexible scaffold for a wide range of functional groups. Through many years of research structures in 2D and 3D have been developed; more recent structures contain switchable elements that respond independently to external stimuli.

The functionalisation of DNA structures adds layers of potential complexity; the key to their widespread application is the flexibility and resolution available, made possible by the pre-defined base sequence. Methods for functionalisation largely depend on single stranded DNA (ssDNA) toeholds designed into structures; these are regions of unpaired bases that newly introduced DNA strands could bind to. Functional groups coupled to a complementary strand of DNA will be directed to bind through the specificity of the sequences. Other methods use DNA aptamers or protein-protein interactions to direct interaction, these are based on affinity for the 3D structure of the scaffold rather than base pairing. A number of potential methods for addressing DNA structures exist but are not currently developed with this aim in mind. For example, research on small molecules for targeting specific DNA sequences focuses on gene regulation [12].

Another method is to address the more stable dsDNA with an additional strand that is complementary to create a triple stranded region; this requires a mediator, like recombinase A (RecA). RecA is a protein of the bacterium *Escherichia coli* that is involved in recombination, taking part in the cellular SOS response to DNA damage. Research on the use of RecA in nanotechnology has focused on two areas. One area of inquiry has used RecA as a tool for nanowire formation [13, 14]. A second stream of experimentation has been carried out to explore the ability of extremely short RecA filaments to specifically address DNA structures [15, 16, 17, 18].

In the presence of ATP and other cofactors RecA and ssDNA will form a nucleoprotein filament, shown digramatically in Figure 1.1. Following the introduction of a double stranded DNA (dsDNA) strand containing a homologous sequence a triple stranded region

will be formed in complex with the RecA. Through the use of the non-hydrolysable ATP analogue ATP γ S it is possible to create stable complexes with RecA and three strands of DNA. RecA filaments have been demonstrated to pattern dsDNA at multiple points simultaneously and with filaments as short as 3 nm (6 RecA monomers) [15, 18].

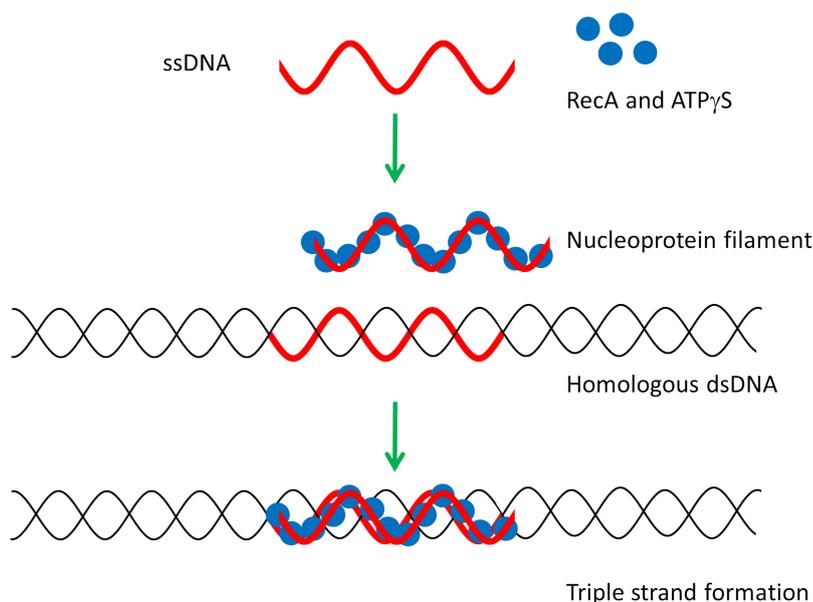


Figure 1.1: A diagrammatic representation of the patterning of dsDNA with RecA and a homologous ssDNA strand. The ssDNA and RecA form a nucleoprotein filament that searches for a region of homology in dsDNA and, in the presence of ATP γ S, forms a stable triple stranded complex.

1.2 Aim

The aim of this project is to study nanoscale patterning of complex DNA structures with the bacterial protein RecA. In the context of this thesis, patterning refers to the sequence specific placement of additional groups on DNA, for example nanoparticles or enzymes. When carried out with groups measuring between 1 and 100 nanometres in size and resolution the process is considered to be nanoscale. The potential applications of RecA are defined by its ability to interact with complex DNA structures; it is necessary to understand how the constraints of complex structures will affect patterning. Complex DNA structures containing unpaired regions and branched junctions were used to test the

potential of RecA for addressing DNA architecture at the nanoscale. The ability to direct groups to these structures in a sequence-specific manner could produce novel functional materials. Previous work has focused entirely on the use of linear dsDNA [15, 16, 17, 18].

To this end, we have produced DNA structures with different levels of complexity; a series of structures with unpaired sections at their centres, a series formed by ligation to give three-way junctions at their termini and a larger structure formed from PCR products. A schematic of these structures is shown in Figure 1.2. Additionally, we attempted to produce RecA to our own specifications, to increase the ease of visualisation and with the long term aim of manipulating the protein to increase the efficiency with which we address DNA.

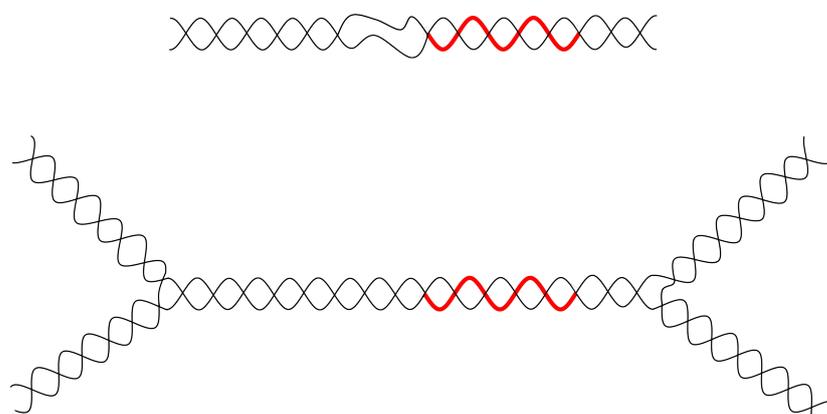


Figure 1.2: Structures with interior ssDNA regions and terminal three-way junctions were created for the purposes of RecA patterning. Schematic representations of these structures are shown. The patterning region is represented by the red part of the helix.

In this work, Chapter 2 will concern itself with some basic scientific principles that underpin the work described in this thesis and a background of the pertinent literature. In Chapter 3 the techniques and materials used will be explained in some detail. Chapter 4 discusses the effects of regions of base mismatches on the patterning of short (100 base) DNA strands, showing that these motifs can increase or decrease patterning efficiency depending on proximity to the patterning region. Chapter 5 addresses work towards understanding the effect of structures containing three-way junctions and terminal single-stranded regions on RecA patterning. A series of 200 base structures were produced through a PCR-like one-pot method for patterning experiments. Chapter 6 further

explores the effect of branched junctions on RecA patterning through the production of a 270 base structure with terminal three-way junctions that has been confirmed through AFM. Chapter 7 reviews work on the production and purification of recombinant RecA. Chapter 8 draws conclusions about this work and discusses the potential future of the field.

Chapter 2

Background and Literature Review

2.1 Deoxyribonucleic Acid (DNA)

DNA is a polymer in which the repeat units are nucleotides, shown in Figure 2.1. Nucleotides are composed of a nucleobase, a five-carbon sugar and a phosphate group. There are four possible natural nucleobases in the DNA structure; adenine, guanine, cytosine and thymine. It is the sequence of these bases in the genome that encodes the information for making the proteins that control the life of a cell. Each strand has a polarity that is commonly identified by the orientation of the carbons in the sugar; conventionally sequences are given in a 5' to 3' direction, as shown in Figure 2.1.

DNA is a biomolecule normally associated with information storage. In this context two strands hybridise to form a duplex, through a combination of hydrogen bonding (shown in Figure 2.1), pi stacking and the hydrophobic effect. Hydrogen bonding is by far the dominant force in this interaction. Hydrogen bonds will form between adenine and thymine and between cytosine and guanine to form a stable double helix structure; this is referred to as Watson-Crick base pairing. Non-Watson-Crick base pairs are possible in a transitory structure but are not tolerated in the context of a living cell. When two strands contain hydrogen-bonding bases in the correct order through their whole length and anneal in an anti-parallel alignment they are described as complementary. This exactitude is behind the 'lock and key' specificity of DNA double strand formation; through manipulation of sequences, and therefore base pairing, it is possible to direct

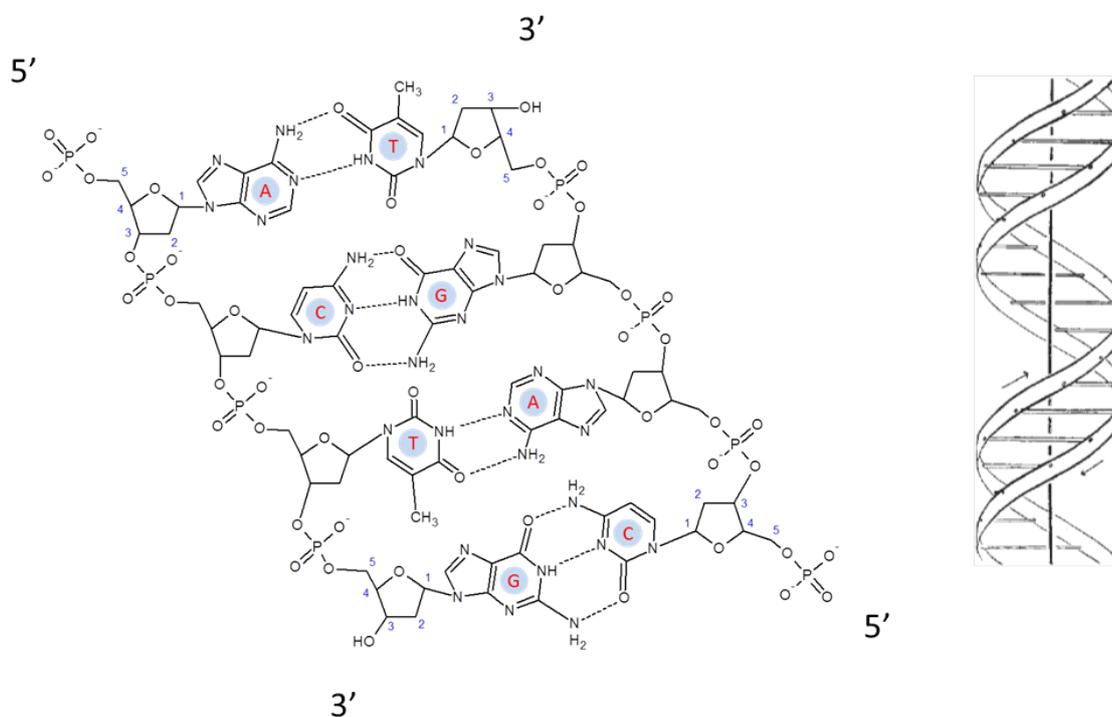


Figure 2.1: **Left:** The bonds that form a double stranded DNA structure. The covalent ester bonds between the 3' carbon of the deoxyribose and the phosphate group of the next deoxyribonucleotide can be seen. The hydrogen bonds that form between bases are shown as dotted lines. The polarity of the two strands, from 5' to 3' has been shown and the antiparallel nature of the double helix is apparent. A is adenine, T is thymine, G is guanine and C is cytosine. Figure generated using ChemSketch. **Right:** The model for the DNA double helix as drawn in Watson and Crick's original paper on the structure of DNA. The ribbons represent the phosphate and sugar groups that form the backbone of the strand and the horizontal rods represent the bases. Image taken from [19].

DNA to self-assemble into complex shapes.

A key characteristic of DNA molecules is the melting temperature (T_m), the temperature at which half of DNA remains as a duplex. DNA melting temperature can be determined empirically using UV spectrophotometry; this gives the most accurate figure for the melting temperature of a given strand of DNA.

Many methods have been developed for the calculation of DNA melting temperatures, the equations take into account a number of variables including salt concentration, the length of the strand, the number of A-T and C-G pairs and the order of the bases [20, 21, 22, 23]. It has been shown that accurate estimations of T_m can be made by taking into account the nearest neighbours of each base using data on the entropic and enthalpic contributions ascertained experimentally [24, 25, 26, 27, 28].

In vivo DNA is synthesised through the action of enzymes called DNA polymerases. These enzymes will copy an existing strand of DNA to produce its complement. A second polymerisation step produces a strand identical to the original. DNA polymerases produce double stranded DNA, though this is not always the end product of the replication reaction. Through the concurrent use of a helicase, nicking nuclease and polymerase, ssDNA viruses are able to produce their genomes as a single strand within an infected cell. The addition of excess of one primer in a PCR reaction will result in the production of ssDNA through successive iterations of melting, annealing and extension. It is also possible to synthetically create single stranded DNA through oligonucleotide synthesis. A number of methods are available that use stepped chemical reactions to sequentially add the desired bases to the terminus of a growing oligonucleotide chain. The action of DNA polymerases and repair enzymes *in vivo* produce an extremely low error rate, in the range of one error per 10^{-10} bases when proof reading is carried out by the polymerase [29]. The error rate is much higher when non-proof reading enzymes are used, for example the error rate of *Taq* polymerase is approximately 1 error in 9000 bases [30]. The error rate is much higher for synthetic oligonucleotide production and this limits the length of the DNA strands routinely produced commercially to around 150 bases.

While double stranded DNA (dsDNA) is relatively rigid, with a persistence length of $\simeq 150$ base pairs, single stranded DNA (ssDNA) is a flexible molecule that can be

bent into a specific design [31]. Creating complex DNA structures involves the use of DNA base pair specificity to make DNA anneal into structures other than simple linear duplexes. The flexibility of DNA is key as it allows a strand to be woven between different annealing partners to reinforce structures. The flexibility of DNA is both a problem and an opportunity; it allows DNA to be shaped but also dictates the need for reinforcing structural elements. The drive to create complex and stable structures with DNA has led to a fast paced and innovative field of nanoscale science.

Depending on hydration, sequence and buffer conditions, dsDNA can take a number of forms in which the strands associate into a helix. The most common form of DNA *in vivo* is B-DNA; in this state the distance between bases is 0.34 nm. The distance between base pairs is also an important aspect of DNA as a structural material; positioning with a 0.34 nm accuracy is highly desirable. Other forms that dsDNA can take include Z-DNA and A-DNA. There are many structural differences between these forms, for example while A- and B-DNA form right-handed helices, Z-DNA forms a left-handed helix [32]. The work described in this thesis is undertaken at neutral pH and physiological concentrations of salt, under these conditions the DNA used will be in the B form.

There are a wide range of naturally occurring enzymes available for manipulation of DNA sequences. DNA restriction enzymes, also called restriction endonucleases, break the sugar-phosphate backbone of the DNA on both sides of the helix to create two, shorter strands. Modified restriction enzymes have been created that will only break one of the strands; these are referred to as nicking enzymes. DNA ligases act conversely to restriction enzymes, joining two strands of DNA to create a single, long strand. DNA exonucleases cut a single base at a time from the end of a DNA strand. It is possible to prevent the action of exonucleases through base modification, for example by phosphorothioation of terminal bases. There are proteins that will methylate adenine or cytosine bases, and others will hydroxylate the methylation. The wide variety of enzymes available for DNA manipulation are one of the many advantages of using DNA for nanotechnology.

2.1.1 Protein biology

Proteins are synthesised from DNA through the action of numerous enzymes which mediate the transcription to ribonucleic acid (RNA) and translate the RNA to the primary structure of the protein, the order in which the amino acids are bonded. There are only 20 natural amino acids that are incorporated into proteins *in vivo*. The mode of protein production from nuclear DNA in eukaryotes is shown diagrammatically in Figure 2.2. From this primary structure three more levels of protein structure are classically described; the secondary structure is the formation of a strand into motifs such as alpha helices and beta sheets. The three dimensional arrangement of these structural motifs is described as the tertiary structure. The quaternary structure of a protein forms when different protein chains and accessory molecules interact.

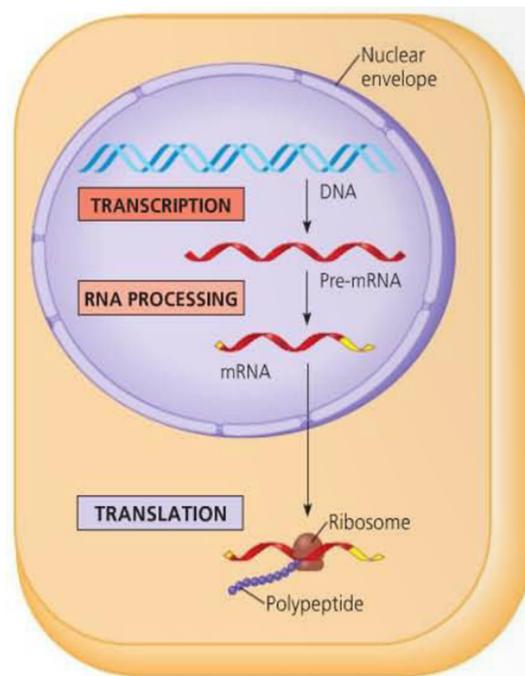


Figure 2.2: Diagrammatic representation of the relationship between nuclear DNA and protein production in a typical eukaryotic cell. The nuclear DNA is transcribed into pre-mRNA which is then processed in the nucleus to produce mRNA. When the mRNA has moved from the nuclear envelope to the cell cytoplasm, ribosomes translate the mRNA sequence into a protein molecule. The mRNA is read in triplets which correspond to a single amino acid or instruction for the ribosome (e.g. stop translation). Image taken from [33].

The structures formed by a protein are of great importance, as they dictate the functions available to it. Flexibility may be important to the binding of a molecule for catalysis; misfolding can completely remove activity. Proteins carry out a huge number of roles in the cell; due to the diversity of their folds they are able to accomplish a wide range of jobs.

In our efforts to better understand protein folding, tools have been developed for the identification of domains within proteins. By mutating individual amino acids it is possible to disrupt protein folding and identify key residues and areas of activity. Additional information can be gained through the use of developmental biology and protein knockouts. Circular dichroism gives information about the secondary structural elements of a protein. Under the correct conditions proteins can form highly ordered crystals that, when subjected to X-ray diffraction, can elucidate protein structure to very high accuracy.

The study of proteins has been greatly aided by the ability to produce proteins in large amounts in a laboratory setting. Through the use of specific bacterial strains and plasmids it is possible to induce production of proteins at much higher than native levels. Affinity tags are short sequences of amino acids or protein domains that can be added to a protein for the purposes of increasing purification yield. The use of affinity tags facilitates separation of the desired protein from contaminants. The ability to create recombinant proteins with high yields has contributed to *in vitro* studies of protein behaviour and interactions. A greater understanding of the roles a protein plays informs our knowledge of cell biochemistry and expedites their manipulation.

2.1.2 Structures in Nature

DNA is typical of large biological molecules in its fabrication through the specific bonding of a number of monomers to create a large structure. DNA in cells does not have a mechano-structural role but the potential is there as demonstrated by other natural polymers which share key features with DNA. These natural polymers form strands and matrices that provide movement and structure to cells and tissues throughout nature. Demonstrating directionality, self-assembly and a programmable tertiary structure, these

natural polymers share many of the requirements for designing nanoscale systems and can inspire the field of bionanotechnology.

An example of structural polymers can be found in the cytoskeleton of the cell, which is made up of three major fibres: tubulin microtubules, actin filaments and keratin filaments [34, 35]. Nucleating factors or organelles are necessary for polymerisation of actin and tubulin; like DNA these molecules have a polarity. The filaments and tubules grow in specific directions to fulfill many key roles in the cell. Keratin polymerisation has no polarity and filaments can spontaneously self-assemble [35]. In the cell this process is controlled to provide a more permanent network than those of actin or tubulin. Filament directed motility changes the cell location and also shapes specific regions of the cell, for example to enact phagocytosis by immune cells. The cytoskeleton provides anchorage for organelles and myosins travel through the cell along actin highways which allows life to dramatically exceed the limits imposed by diffusion. The extensions of axons are reinforced by cytoskeletal filaments. These polymers create within the cell a high density of complex, inter-dependent structures that are highly adaptable and varied, made from the same programmable building blocks.

Structures in biology also provide compartmentalisation. Virus capsids are 3D structures made of protein subunits that contain the viral genetic material for transport between infected cell and the subsequent host. The smallest virus can have a diameter of ~ 17 nm while the largest have a capsid diameter of 400 nm and package a very high density of nucleic acid. Mammalian bone is initially formed as a collagen matrix. A number of ions combine to harden the structure, forming hydroxyapatite which is stronger than collagen but flexible enough to not be brittle. Bone marrow resides within the matrix created, providing access to the blood but also compartmentalising the essential stem cells. These structures make it possible to create chemically distinct environments in close proximity and to change them independently.

These processes are tightly controlled, for example the growth of microtubules and actin filaments or the regular patterns within bone structure. Structures are built up from smaller units in the presence of a complex mixture of molecules and ions, providing robust solutions to the necessities of life.

The properties described here are all highly desirable for nanotechnology. Controlled growth and disassembly, compartmentalisation and bottom-up assembly within a closed system are all traits that could be extremely useful if applied to nanoscale networks. From this conclusion is drawn the field of bionanotechnology, in which we attempt to manipulate biological materials to fulfill our nanoscale needs.

There is work to create structural materials from proteins [36]. However, protein structure is significantly less well understood than other biological materials, i.e. nucleic acids. There is research into fabricating complex structures from RNA [37, 38, 39], which was reviewed in 2010 [40]. RNA and DNA are structurally similar but the deoxyribose sugar of DNA is less reactive to hydrolysis. While DNA is usually double stranded, RNA is single stranded and forms complex secondary structures. RNA nanotechnology uses characterised secondary structure motifs as building blocks to design larger ordered structures. Through sequence optimisation they have achieved co-transcriptional folding of RNA structures, as shown in Figure 2.3. The RNA molecule is formed by RNA polymerase through transcription of DNA and the interactions coded in the sequence cause the molecule to form tiles which can then take part in the formation of lattices [39]. While these groups have achieved some success, the applications of RNA nanotechnology are limited by the instability of RNA. A much larger field of research is the development of DNA structures. As well as being a more stable molecule, the level of control available, the tools and ability to model DNA make it an ideal candidate for creating nanoscale systems.

2.2 Creating complex DNA structures

Complex DNA structures were initially developed in the early 1980s for use in protein crystallography [41]. It was thought that lattices of DNA could be used to arrange recalcitrant proteins into ordered arrays that would allow them to be studied by X-ray diffraction. It began with the design of relatively simple junctions [31] and has increased in complexity to include structures in 3D [42] and with switchable elements [43]. Our understanding of DNA as a molecule has proved extremely important to advances in the field.

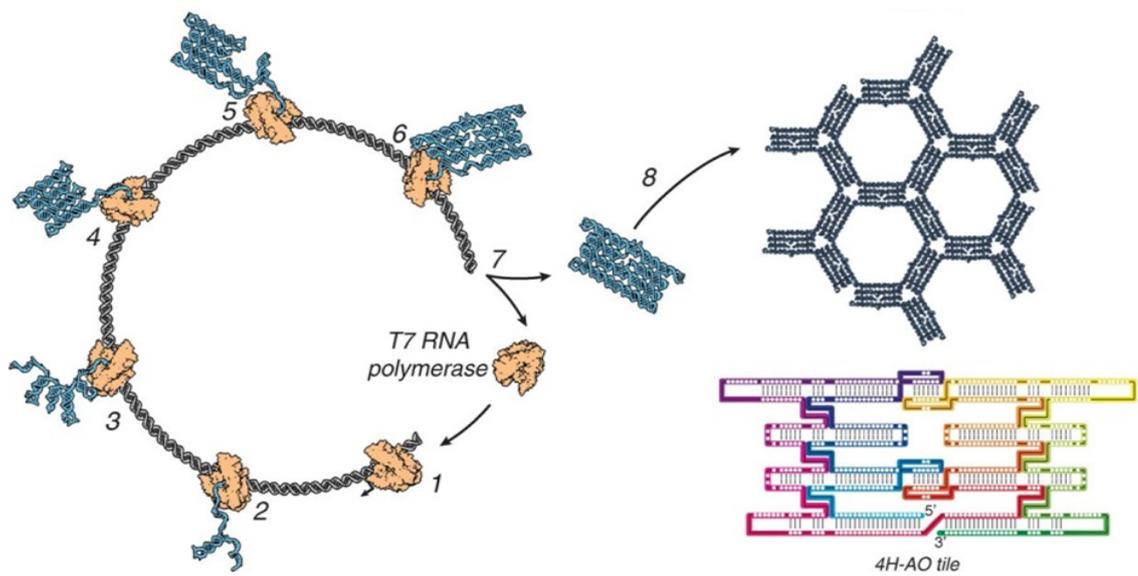


Figure 2.3: The formation of an RNA nanostructure designated the 4H-AO tile, through co-transcriptional folding. An RNA molecule is formed through the action of RNA polymerase transcribing a DNA molecule (1). As the RNA molecule is formed the base interactions direct the formation of a structure (2-6). Multiple iterations of the process are possible when the ribonucleotides are available for transcription (7). Following transcription and folding the tile is available to form lattices (8). Image taken from Geary *et al.* [39].

Central to the creation of DNA architecture is the hydrogen bonding between bases which makes the hybridisation of two strands of DNA dependent on complementary sequences. By designing the sequences of DNA strands it is possible to direct the annealing of strands to produce more complex structures. The sequences are the only programming needed for these structures to occur. Under the correct conditions, including temperature and ion concentrations, DNA will self-assemble into the programmed structures due to the specificity encoded by sequence complementarity.

DNA structures could have a number of roles as distinct entities; through functionalisation their uses could be multiplied many times [44]. For example, through the DNA scaffolded organisation of particles it is possible to tune the optical properties of plasmonic nanostructures [45]. There are a number of strategies for attaching biomolecules [46, 47] and other functional groups to DNA [48]. Important aspects of functionalisation include the minimum size of the groups that can be attached and the accuracy with which they can be positioned.

2.2.1 2D structures

Two methods have emerged for creating DNA structures in 2D. The first strategy developed uses a number of medium-length ssDNA strands to wind around each other and form unit cells which can then self-assemble into larger arrays [49]. A more recent strategy uses a very long ssDNA genome which is directed by many short staple strands in a raster motion to create a shape [50].

DNA structures were, at their inception, very simple. They were inspired by Holliday junctions, four stranded DNA junctions observed in fungi [51]. These structures are not stable *in vivo* but early work attempted to form immobile junctions of this kind [49] and to join them using short, complementary ssDNA overhangs called sticky ends [41]. The role of tension and twist in the annealing of DNA strands was not yet fully understood [49]. Structures were formed of a few strands of DNA annealing to form single helical structures.

The next major step forward in this technology was the creation of DNA double crossover motifs [52]. These structures are designed such that DNA strands interlink between

two helices in a parallel or anti-parallel fashion, reinforcing the structure and increasing stability as shown in Figure 2.4A. They were more complex to design as incorrect spacing between helix crossovers and helical mismatching could lead to increased tension in the structure that could prevent correct formation [52]. It is important to understand the effects of tension and twist when making DNA structures as crossovers between strands must be optimised to reduce the strain introduced (see Figure 2.4 A).

Further development of this idea investigated the use of crossovers as edge sharing motifs for 2D construction that does not use sticky end cohesion [53]. Crossover patterns were used by Winfree *et al.* [54] and by Ding *et al.* [55] to create 2D crystals of DNA, shown in Figure 2.5. The crossover motif has also been used in the creation of nanotubes and arrays from bundles of helices, designed with sticky ends [56].

Early work relied on native agarose and polyacrylamide gel electrophoresis to identify whether DNA structures had formed, based on the change in DNA mobility within the gel. This is no longer the case; 2D DNA structures are now most often observed through atomic force microscopy (AFM), a technique developed in the 1980s [57, 58]. Examples of 2D DNA structures can be seen in Figure 2.5. At its most advanced AFM is able to image at single molecule resolution [59]. Although DNA structures are not regularly observed at this resolution, it is possible to view individual structures and to gain data about fidelity and reproducibility of formation.

DNA origami is a field developed by Paul Rothemund, and first published in his 2006 Nature paper [50]. It involves the use of a single long strand of DNA and many short staple strands. The sequence of the long strand is known and the staple strands are designed to bend it into a set shape. The genome of the M13 bacteriophage is widely used as the source of the long strand of DNA, as it is 6.4 kb long and naturally single-stranded. Through this method it is possible to create reproducible DNA structures that assemble in a single step. Examples of these structures can be seen in Figure 2.6.

Work built on this technique included the use of DNA origami as a base for nanoarrays [60] and to create solid sides to 3D shapes [43]. Zhao *et al.* aimed to scale up origami structures through the use of origami tiles as staples in a larger array, the tiles were regular, flat, rectangular origami shapes that could be connected at their sides or corners [61]. The

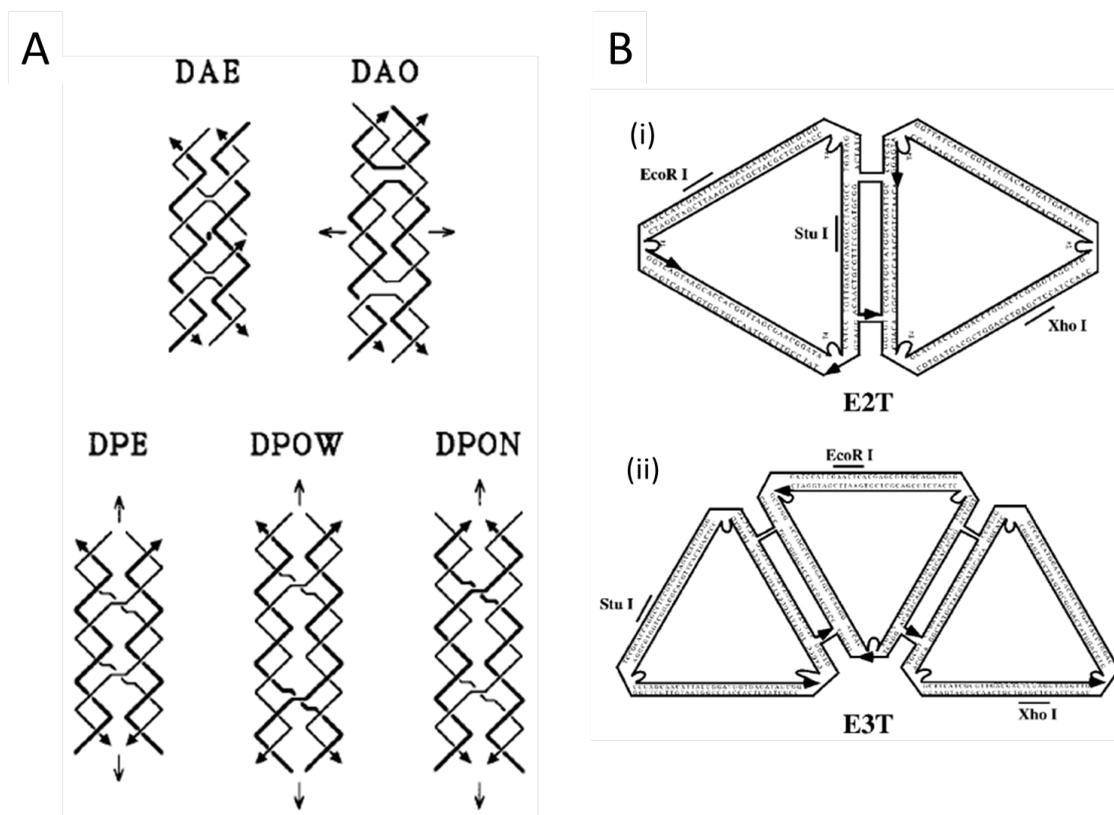


Figure 2.4: *A*: The five possible arrangements for the formation of DNA double crossovers. The letter code provides information about whether the strands are parallel (P) or antiparallel (A) and whether the number of half turns between crossovers is even (E) or odd (O). In the case of parallel helices with odd numbers of helices between turns an extra letter is added to identify whether the extra half turn corresponds to the major (W) or minor (N) groove. Image taken from Fu and Seeman [52]. *B*: The use of the DAE crossover to form edge sharing triangles. (i) Two triangles formed of four DNA molecules are held together by a DAE motif. (ii) Three triangles formed of six DNA molecules are held together by the same motif. Strand polarity is shown by arrows pointing in the 3' direction. Restriction sites are shown on both structures. Image taken from Yan and Seeman [53]

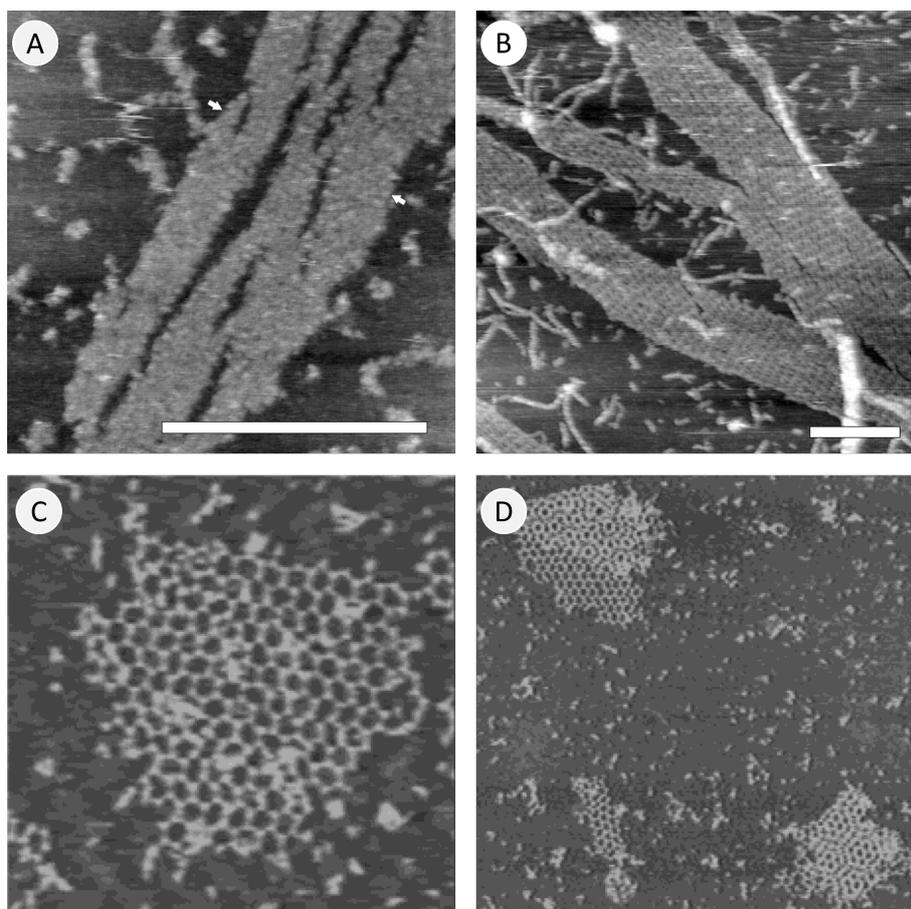


Figure 2.5: Some examples of complex DNA structures in two dimensions. Panels A and B show examples of the structures produced by Winfree *et al.*, imaged using an AFM. Images taken from [54]. Scale bars are 300nm. Panels C and D show the honeycomb pattern of the structures produced by Ding *et al.*, again imaged using an AFM. Images taken from [55]. Image C is 1.22 μm , image D is 3.13 μm .

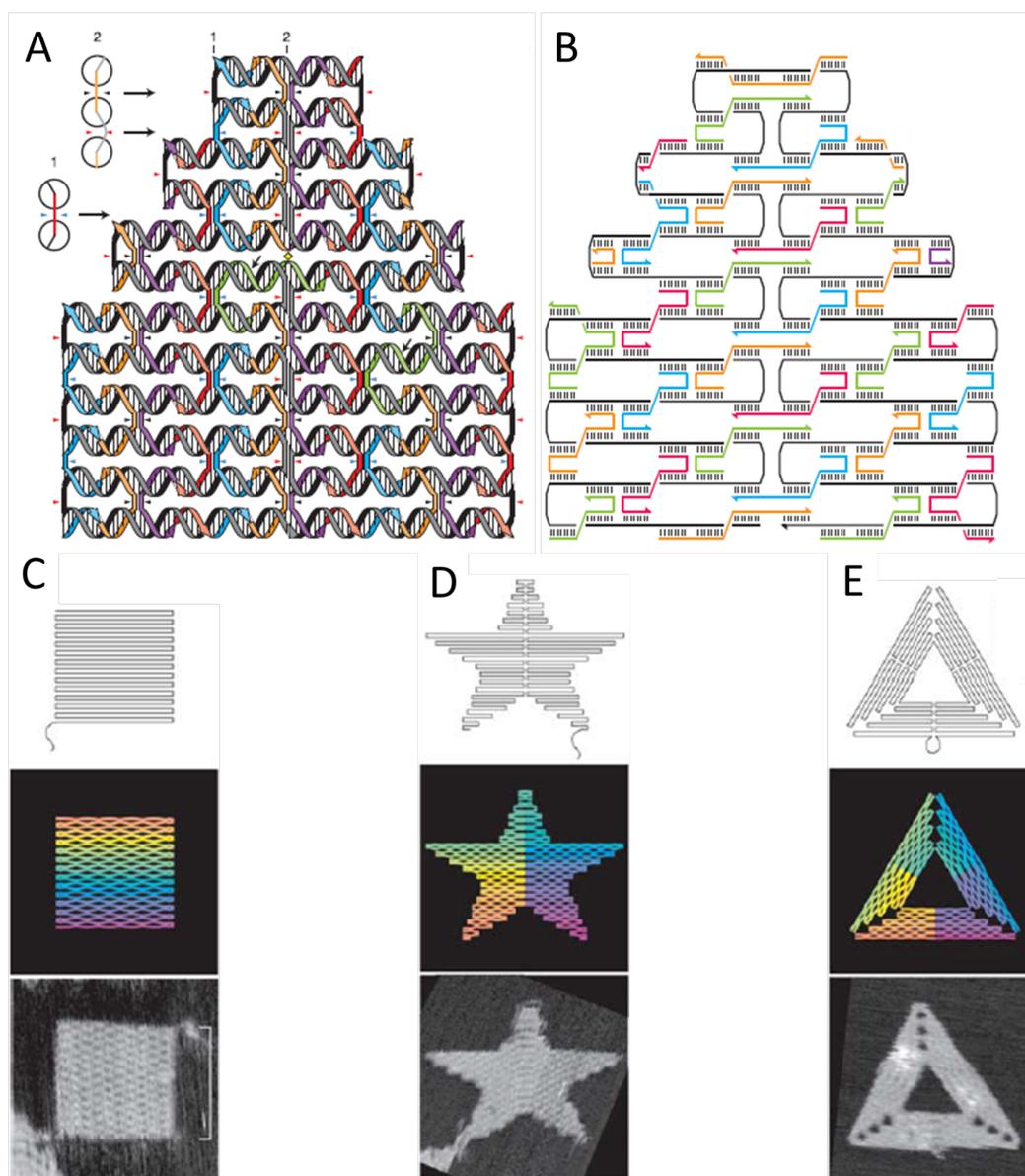


Figure 2.6: **Panel A** shows the pattern of helices that form a DNA origami structure. The staple strands are coloured and the the scaffold is shown in black. Scaffold crossovers are indicated by red arrows. The periodic crossovers where the minor groove is on the top face are indicated by black arrows and those where the minor groove is on the bottom face are indicated by blue arrows. **Panel B** shows the pattern of crossovers and base pairing for all the staple strands (coloured) and the scaffold (black). **Panels C, D and E** show the design of three DNA origami shapes using the method shown in panels A and B. Images taken from [50]. From top to bottom the images are: the folding path of the long strand of DNA; the crossover patterns of the staple strands; an AFM image of a single example (images are 165nm x 165nm).

M13 genome was folded into many rectangular tiles that were connected; the distance between the tiles increased the size of the complete structure beyond that of a traditional origami rectangle.

Endo *et al.* further developed the idea of modular DNA structures in 2D, developing the origami jigsaw which used a combination of shape and sticky ends to bind tiles end to end in a specific order [62]. Further to this work, the designs were crosslinked, which increased heat stability [63]. The heat stability of the structures was important to the formation of arrays of tiles, by allowing a higher temperature in the annealing reaction and a corresponding increase in yield. These tiles were assembled into a larger array where nine individual 'jigsaw pieces' were organised into a 3x3 self-assembled structure [64].

Additional work by the same group has developed a DNA tile that is used as a scaffold for the examination of numerous reactions. A methylation enzyme [65], a DNA repair enzyme [66], recombination [67], transcription [68], photoresponsivity of the crosslinking system discussed in Rajendran *et al.* (2011) [63, 69] and alkylation [70] have all been explored using this system. Work using this system has been recently reviewed [71].

A particular interest has been shown in G-quadruplex formation [72, 73, 74, 75]. G-quadruplexes are a form of nucleic acid secondary structure; these square planar structures can form when nucleic acid strands are rich in guanine and can include single or multiple strands. Human DNA telomeres have been shown to form G-quadruplexes *in vivo* and their presence is linked to reduced sensitivity to the enzymes DNA telomerase, polymerase and helicase [76]. Telomere length has been linked to ageing and cancer [77, 78].

In a crossover between the origami approach and the use of entirely synthetic oligomers, the work of Li *et al.* used an outline of DNA origami to nucleate and delineate the arrangement of small DNA tiles [79]. The use of DNA origami as a molecular canvas was explored by Wei *et al.*, who added or omitted staple strands to remove sections of a rectangular origami, leading to the formation of a range of precise 2D shapes [80].

2.2.2 3D structures

A wide range of complex structures can be produced in two dimensions, as demonstrated in Section 2.2.1. The logical extension of this work was to fabricate DNA architecture in three dimensions, to further take advantage of DNA as a structural material.

DNA structures in 3D have been constructed as the frame of a number of shapes [81, 82, 83, 84] or as shapes with solid sides [42, 85]. DNA frames could be used to organise the association of various functional groups, precisely controlling the distance between different elements. Structures with solid sides have the potential to sequester active substances until their release is triggered.

The development of DNA structures in three dimensions built on 2D work and shared many similar motifs. For example, the work of Shih *et al.* used the double crossover structures of Fu and Seeman [52] to create a nanoscale octahedron [81]. This work preceded the work of Rothmund [50] in using a single, long strand of DNA (1700 bases) which is then folded into a shape through the use of shorter 'staple' strands, in this case five 40 base oligonucleotides.

DNA has also been built into macroscopic structures. The use of tensegrity triangles to build DNA crystals measuring micrometres in size was demonstrated by Zheng *et al.* in 2009 [86]. Tensegrity describes a structure that relies on balanced compression or tension to maintain its shape [87]. This work was built on the initial demonstration of DNA tensegrity triangles by Liu *et al.* [88]. In a 2010 paper Liedl *et al.* used DNA origami as a starting point to build a greater range of tensegrity structures from DNA [89].

A number of modular approaches to building larger structures from DNA have had success in creating structures larger than those seen before [83, 90, 84]. The use of computer aided design and liquid handling robots greatly expands the design space that can be explored [90].

Other work has shown that it is possible to design ribbons of DNA that bend up to 180° [91]. In this work helices are stacked, connected by crossovers, to create bundles of fibres that can then be altered to introduce tension. Tension is expressed as twisting or bending of the structure, shown in Figure 2.7. This work highlights the importance of tension in

the design of DNA architecture; formation is dependent on balancing these forces.

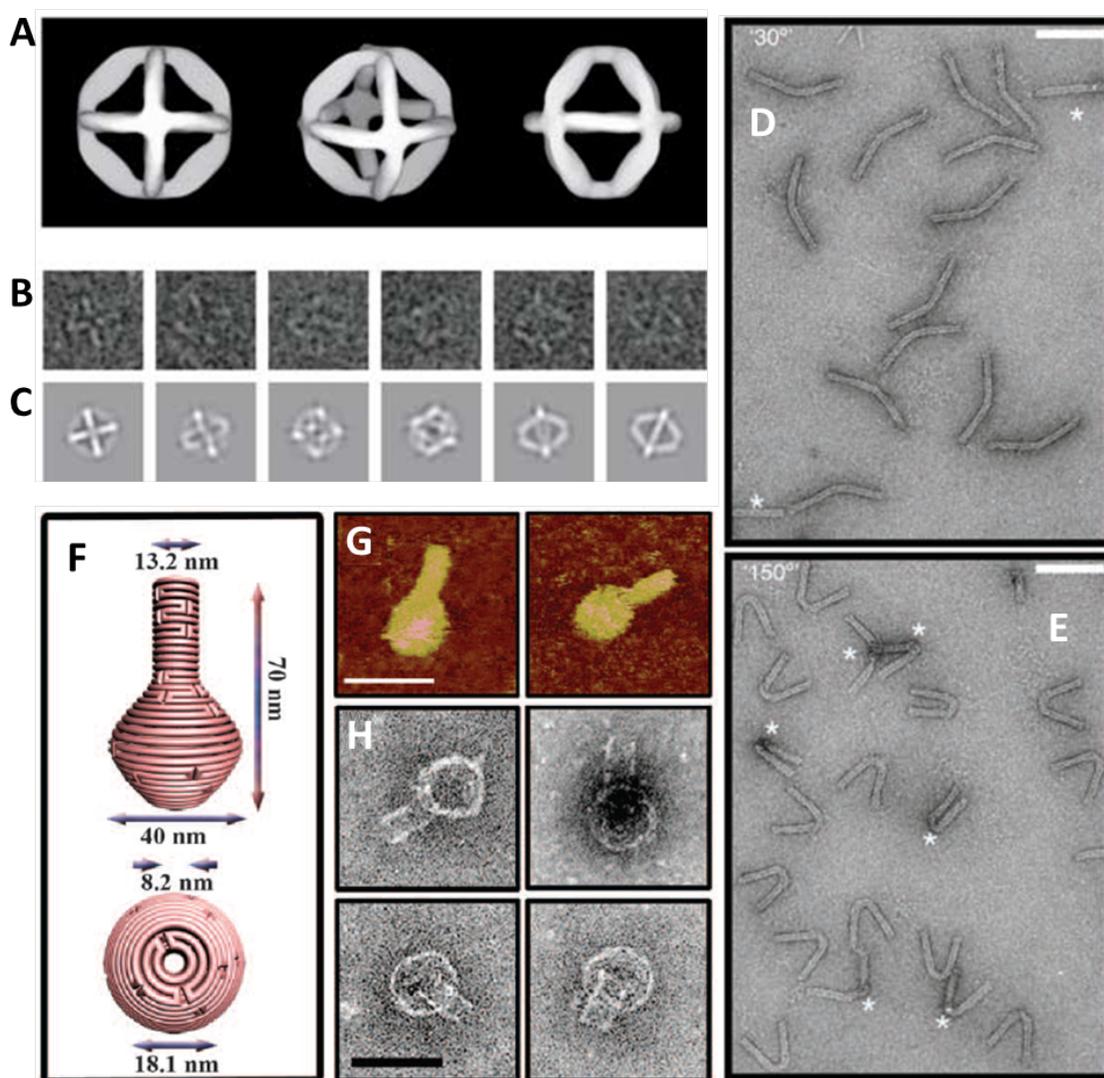


Figure 2.7: This figure illustrates some of the 3D shapes that it is possible to fabricate using DNA. A simulation of a DNA octagon is shown in A. B shows the negative stain transmission electron micrographs (TEM) taken of the structures and C shows a reconstruction of the 3D image based on averaging TEM images. Images taken from [52]. Figures D and E show TEM images of DNA bent at 30° and 150° . Images taken from [91]. Scale bars are 100nm. Panel F shows the design for a 3D nanoflask, G shows AFM images of these structures (scale bar 75nm) and H shows negative stain TEM images of the same design (scale bar 50nm). Images taken from [42].

Another group has created hollow DNA shapes with solid sides [42] using the DNA

method developed by Rothmund [50]. Manipulation of sequence length between the crossovers of staples strands allowed them to introduce curves, a method similar to the work of Dietz *et al.* [91]. Han *et al.* used these techniques to create a 'nanoflask' - a hollow structure in which the bend of the DNA changes pitch through the z axis [42]. Images of these structures can be seen in Figure 2.7. Although the DNA is able to form structures in which the conformation is very different to native, it was found that the increased strain this causes reduces the yield of correctly folded product [42].

2.2.3 Structures with switchable elements

Another interesting aspect of DNA nanotechnology is the development of structures that can react to the environment around them, using the methods described for creating 2D and 3D architectures. Shapes that are able to change confirmation take advantage of the flexibility of DNA. These structures are able to respond to external stimuli in ways that could include the sequestering and release of active compounds.

In 2009, Andersen *et al.* produced a DNA box 42 nm x 36 nm x 36 nm with a lid that could be induced to open or close through the addition of DNA oligonucleotide 'keys' [43]. They were able to observe the opening and closing of the box using Förster resonance energy transfer (FRET) experiments, where a fluorophore was attached to the lid and to the lip of the box [43]. Other groups have also used fluorescence to monitor the opening and closing of DNA structures [92]. In their work Yurke *et al.* created a set of molecular tweezers, opened and closed through the addition of DNA strands [92]. The mechanism for opening and closing, as well as the positions of the fluorescent markers, are shown in Figure 2.8.

Douglas *et al.* (2012) developed a DNA barrel that opens to reveal functional materials [85], a schematic representation is shown in Figure 2.9. They assessed the opening of the structure through negative stain TEM images of gold nanoparticles attached to the two sides.

Photoswitchable multimerisation of DNA nanostructures has been demonstrated by Yang *et al.* in 2012, who used azobenzene modified oligonucleotides to connect and disconnect origami tiles in a controlled manner using light [93].

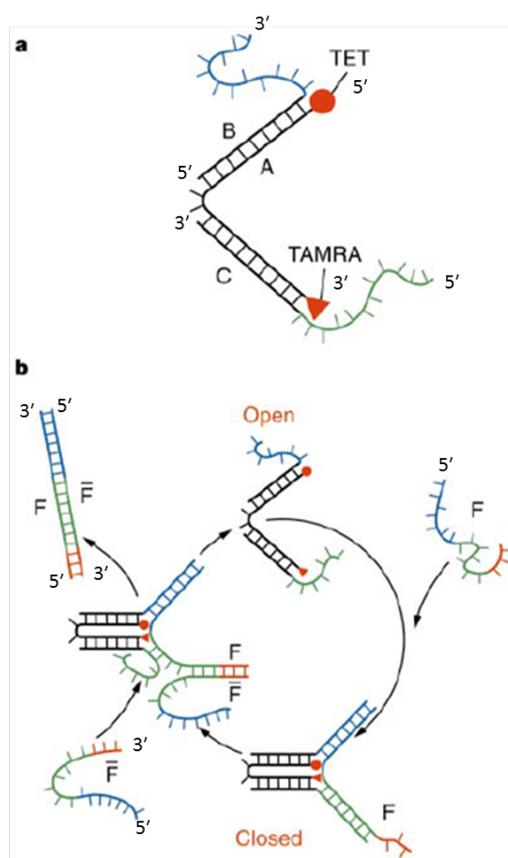


Figure 2.8: Panel A shows the positioning of the fluorescent molecules used to assess the opening and closing of the molecular tweezers. TET (5' tetrachloro-fluorescein phosphoramidite) is a fluorescent emitter, resonant energy transfer between TET and TAMRA (carboxy-tetramethylrhodamine) quenches this fluorescence when they are brought into close proximity. Panel B shows the cycle of opening and closing induced by the DNA fuel, strand colours indicate sequence complementarity. Starting at the top, the molecule is open and fluorescence can be detected. The fuel molecule (F) is added and through annealing to the blue and green sections of the tweezers it 'closes' the structure and brings the two dyes together, quenching the fluorescence. The strand \bar{F} is the anti-fuel, when added it anneals to the fuel strand by attaching to the orange foothold at the 3' end of the strand. This removes the fuel and opens the tweezers, reestablishing the fluorescent signal and producing a fully dsDNA waste strand. Images taken from [92].

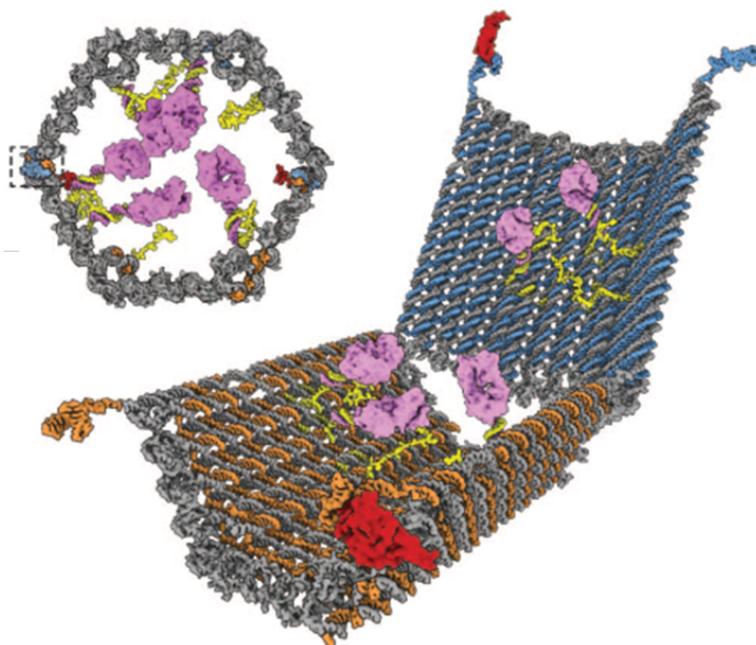


Figure 2.9: The two halves of the structure developed by Douglas *et al.* are shown here in a schematic representation, opened to reveal the magenta cargo. The image in the top left corner shows an end view of the closed structure, showing the internal attachment of the cargo. The proteins bound to the DNA aptamer are shown in red. Image taken from [85].

Recently the work of Marras *et al.* demonstrated fabrication from DNA origami of machine-like mechanisms made with the ability to move [94]. It remains impossible to machine traditional engine materials at this scale; this work creates a DNA facsimile of key designs for linear and angular motion. They believe their work could produce mechanisms akin to macromolecular engines.

DNA Walkers

A particular specialised area of switchable DNA structures is dedicated to the development of DNA walkers which will walk along a track, some will also carry a cargo. Early structures required compromises between autonomy and programmability. In 2004, Yin *et al.* produced unidirectional movement along a DNA track through the use of DNA ligase and restriction enzymes [95]. In the same year Shin *et al.* produced movement along a DNA track through the addition of a DNA fuel [96]. The work of Green *et al.*

also utilises DNA hybridisation to fuel unidirectional movement, producing papers that focus on the kinetics of hybridisation and movement [97, 98].

In 2010 Gu *et al.* developed a DNA structure that moved across a DNA tile and picked up a gold nanoparticle cargo on its journey [99]. In this case AFM was used to image the changing position of the gold nanoparticles to demonstrate the efficacy of the system [99], as shown in Figure 2.10.

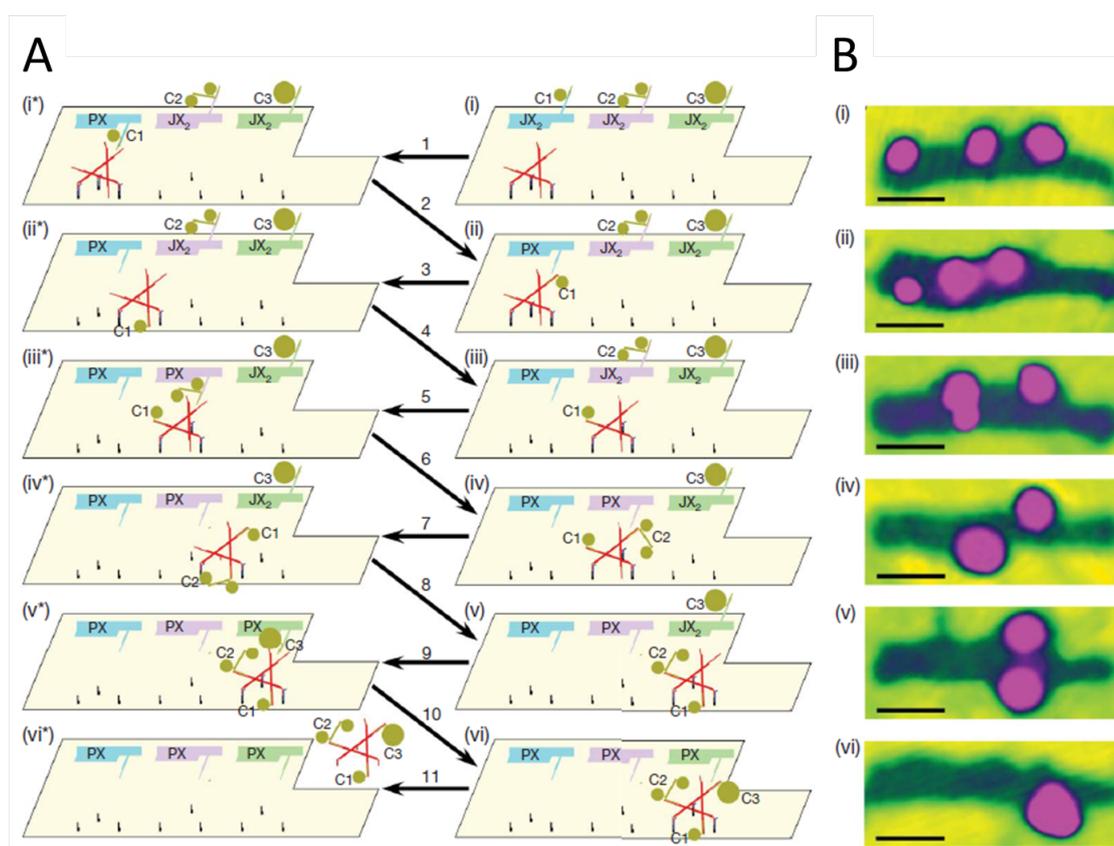


Figure 2.10: *A*: Diagrammatic representation of the action of the DNA walker as it collects gold nanoparticles. The walker is shown in red, the nanoparticles in yellow. The platform is a DNA origami tile with independently programmable machines at which gold nanoparticles can be collected (blue, purple and green). The state of the collection points can be PX (donate cargo) or JX₂ (do not donate cargo) to produce up to eight different cargos. *B*: Atomic force micrographs of the walker collecting gold nanoparticles aligned with the steps shown in diagram *A*. Image taken from [99].

Wickham *et al.* directly observed the movement of a DNA-fuelled walker across a 100

nm track on a DNA origami tile. The walker used the action of a nicking enzyme to catalyse each step [100]. This work was extended to show a network of tracks; an instruction sequence programs the path of the walker [101]. Additional DNA strand 'control elements' can block and unblock routes, allowing the route of the walker to be controlled both internally and externally [101]. The internal programming produced 76% route accuracy; the addition of strands to block control elements were found to produce a path fidelity of up to 87% [101].

In their 2010 paper Lund *et al.* used a Rothemund-style landscape [50] as the basis for the autonomous movement of a DNA-containing robot [102]. They created a structure with a streptavidin body and DNA legs which moved across a DNA tile with a path of ssDNA toeholds. This movement was rendered unidirectional through the modification of the track; the legs were DNA enzymes which cleaved the footholds as they 'stepped' [102].

A DNA walker has been used to direct chemical synthesis in an isothermal, one-pot solution. The sequence of an amine acylation reaction was programmed by the DNA track, the components of the reaction created a DNA ribosome [103]. Other work has used DNA hairpins that do not require a substrate for a similar purpose, instead using the addition of template strands to direct the order of addition. Reactive chemical groups are attached to DNA adapters which are brought into close proximity by the template strands [104, 105].

The wide range of structures discussed here provide a representative sample of the early work and developing areas of research for DNA nanotechnology. The study of structures fabricated from DNA has made great leaps since it was first envisaged and the number of citations generated each year continues to grow.

2.2.4 Visualisation

One of the additional challenges of work in three dimensions is in finding ways to image the structures made. The most common technique used has been transmission electron microscopy (TEM). In the case of Dietz *et al.* [91] and Han *et al.* [42] negative stain transmission electron microscopy was used to collect images of DNA. This technique

is possible because of the size of the structures; it is not possible to image dsDNA strands this way. Evidence shows that electrostatic forces and drying effects damage DNA structures imaged in air.

Cryo electron-microscopy (Cryo-EM) has proved its value in the imaging of DNA structures in a native-like conformation; as well as being useful in imaging larger structures [106] averaging techniques can be used to build a reconstruction of smaller structures [83]. Single molecule cryo-electron microscopy images a large number of particles randomly oriented within a solution. The images are put into classes which are averaged to show different angles of the same protein or DNA structure. A 3D structure can then be assembled from the different angles imaged. The use of fluorescence imaging has allowed small DNA structures to be observed in a native, hydrated conformation through point accumulation for imaging nanoscale topography (DNA PAINT) [84]. This method uses fluorescently labelled oligomers to bind to a DNA structure in solution; the switching fluorescence of these groups allows localisation of the structure and gives super-resolution imaging [107].

2.2.5 Sources of DNA

Sourcing DNA for complex structures is a field of research in itself. A large array of structures use synthetic oligomers [54, 92, 108, 109]. The origami technique [50] uses viral genomes as a source of long ssDNA molecules. The application of technologies based on these structures depends on the production and purification of the genome on a very large scale, as investigated by Kick *et al.* [110]. The limit to the extension of these genomes currently defines the size limit of individual DNA structures. Larger sections of DNA can be joined together, but a more effective route is through connecting many, smaller DNA structures to create a larger lattice [111].

Other approaches have also attempted to use natural sources for the production of DNA for structural work, for example rolling circle amplification [112, 113]. There have been attempts to produce DNA *in vivo* that will self assemble post- and even pre-purification [114].

2.3 Strategies for functionalising DNA structures

The wide range of DNA topologies and movements that have been developed suggest a large number of potential roles for DNA as a structural material. Functionalisation describes the process of attaching functional groups to a structure; functionalisation of DNA structures greatly increases their potential uses. Some methods for functionalisation have already been developed, including biological and synthetic routes, but there are limits to the flexibility of these approaches. There are both positive and negative aspects to each method of functionalisation. It is thought that Recombinase A might offer a more flexible option for addressing DNA structures, for reasons that will be discussed later.

The range of possibilities adds to the flexibility of DNA nanotechnology; each structure and route for addressing it introduces the opportunity for optimisation according to the chosen role. Whether to sequester active compounds or provide a surface for the creation of novel materials, positioning is key to utilising DNA nanotechnology. The distance between bases in B-DNA is 0.34 nm; any method for targeting DNA that is accurate to the distance of one base has a resolution of 0.34 nm. This is higher than the resolution of current lithographic techniques used to produce electronic structures.

2.3.1 Base substitutions

One way to attach functional groups to DNA molecules is through alteration of the chemical structures in the polymer. Modified or 'unnatural' DNA bases have been in development for a number of years, an early review can be found by Kool who discusses classes of DNA base substitutions [115]. The chemical properties of the canonical bases are limited, their fluorescence absorbance is modest, their size and structures are very similar.

A major advantage in the field of unnatural bases has been the development of a dedicated polymerase for strands including hydrophobic base analogues (HBAs) [116]. Loakes *et al.* worked on enzymes from the genus *Thermus* using an evolution strategy called compartmentalized self-replication. A DNA strand encoding the polymerase and

including HBA bases was replicated by the polymerase in an aqueous compartment. Where adaptation took place the gains were identified through increased DNA replication.

Although it is possible to include HBAs in DNA strands, the polymerisation of strands made entirely of unnatural bases is still largely inefficient. Alternative genetic polymers capable of information storage and recovery have been suggested, named xeno-nucleic acids (XNA), which contain unnatural backbone chemistries [117]. The work of Pinheiro *et al.* showed that it is possible to engineer polymerases that will transcribe DNA to XNA and reverse transcribe back to DNA. Using these molecules, a catalyst has been made that acts as an RNA endonuclease [118].

A recent review by Pinheiro and Holliger discusses the potential of XNA in nucleic acid nanotechnology [119]. They suggest uses such as providing a framework to aid structure formation which is then easily removed, or the formation of nucleic acid nanoparticles around metallic cores. They note that DNA structures are not stable *in vivo* but chemical modification could remove this limitation, or indeed add additional chemical functions [119].

A third base pair which can be amplified under standard conditions has been described [120]. The base pair d5SICS and dNaM can be incorporated and the sequence amplified through the polymerase chain reaction. The interaction between these bases is hydrophobic, instead of the hydrogen bonds that hold together A-T and G-C pairings, as shown in Figure 2.11.

Although a great deal of potential can be assigned to nanostructures from altered nucleic acids it is a field still in its infancy, developing the tools for replication and alteration that already exist for DNA structures.

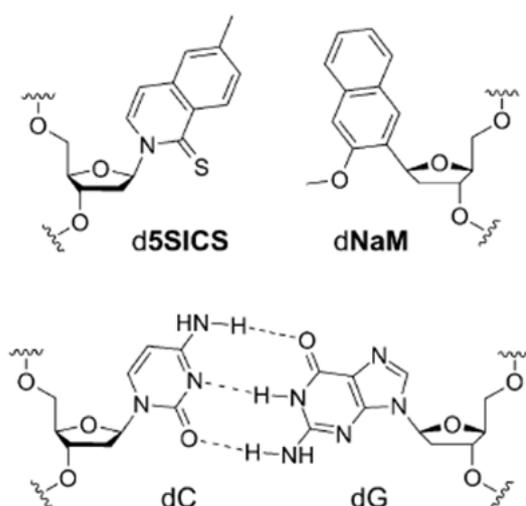


Figure 2.11: Above, the synthetic base pair d5SICS-dNaM. It has been demonstrated by Malyshev *et al.* that these molecules can be incorporated into polymers with the DNA bases cytosine, guanine, thymine and adenine [120]. It has also been shown that these short sequences can be amplified through the polymerase chain reaction. Below, the base pairing of cytosine and guanine is shown diagrammatically for comparison. Images taken from Malyshev *et al.* [120]

2.3.2 Functionalising bases

As well as synthesising novel bases, DNA can also be functionalised through the chemical alteration of the bases [48, 121]. This work requires the creation of nucleotides with additional elements that can still be polymerised, either naturally or synthetically. Alternatively, groups can be added following synthesis, potentially with much less control and can be made to sit in the interior or exterior of the DNA helix [121].

Work has been done on the cross-linking of functional groups to DNA deoxyribonucleotide triphosphates (dNTPs) which are then incorporated into a strand of DNA through the action of a polymerase. This can be done through the use of halogenated dNTPs [122]. Functional elements can be placed specifically through the use of sequences that contain the modified base (A, T, G or C) at key points. This work has been used to attach anthraquinone [123], alkylsulfanylphenyl groups [124] and amino acids [125] to bases. A polymerase was successfully used to incorporate the modified

bases into a strand of dsDNA.

Blaskjaer and Gothelf created modified DNA bases which could then be incorporated into 30 base strands of DNA through synthetic DNA polymerisation techniques [126]. Although this allows the inclusion of molecules not tolerated by DNA polymerase, it also limits the length of the resulting molecules.

In their 2011 paper Burns *et al.* discuss the use of groups placed at specific locations on a DNA nanotube, for the purpose of metal complexation; this allows for metal recognition or metalisation of the DNA structure [121]. The nanotube is designed such that the modification is present at every fifth base, so that they are present for every half turn of the helix. They used the method of Rühl and Stulz to create modified bases which could then be synthetically incorporated into strands of DNA [127].

Another group used DNA-protein interactions to bind a protein for single-molecule imaging [128]. They first created 4-way Holliday junctions with sticky ends that self-assembled into a large, 2D array. Tris-nitrilotriacetic acid (tris-NTA) was conjugated to the DNA and used to bind a number of proteins through an interaction with an affinity tag [128]. Selmi *et al.* used this technique to directly image individual proteins and a multi-protein complex [128].

A large number of molecules can be attached to DNA in this way and could potentially be removed through chemical reactions to break the covalent bonds connecting the functional group to the DNA. These methods are dependent on the yield of the chemical reactions for efficiency and on dissociation methods that do not target the bonds of the DNA polymer. There are other problems with this method; specifically, there has not yet been any attempt to incorporate the functionalised strands into more complex DNA structures. It is possible that these strands will not have the flexibility necessary to weave into crossover motifs, precluding their use in DNA architecture.

2.3.3 DNA toeholds

A common strategy for addressing DNA structures is to build in 'toeholds', single stranded elements that act as docking points through annealing with a complementary

sequence. They are easy to design and extremely versatile.

DNA annealing is the most common method for connecting DNA 'tiles' together to create larger lattices [111, 129]. This greatly extends the potential size of DNA structures beyond the limits of synthetic oligonucleotides and available genomes.

Le *et al.* used a 2D DNA array to organise gold nanoparticles, through an attached ssDNA sequence that bound to toeholds in the DNA structure [108]. There are many other examples of DNA structures being used to organise nanoparticles into regular patterns, as shown in Figure 2.12 [130, 129]. Hung *et al.* used a lithographically patterned surface to create a regular array of nanoparticles and DNA [131]. They made a DNA origami triangle with extended staple strands at the corners, complementary to sequences attached to gold nanoparticles.

DNA hybridisation as an attachment mechanism allows for the precise controlling of on/off cycles using temperature. By programming sequences for hybridisation it is possible to have a range of temperatures over which different elements would be attached to the structure. Functional elements can be directed to any accessible part of the structure and multiple different groups can be directed specifically.

When incorporating ssDNA sequences they must not be complementary to any other part of the structure, but this is true of any part of a complex DNA structure. More importantly, secondary structural elements must be considered. Temperature is an important factor in DNA annealing and can be used in the self-assembly process. DNA footholds cannot contain hairpin loops or other structures with melting temperatures similar to the dsDNA strand they will be expected to form with their complement sequence. The sequence specificity of DNA toeholds allows for large number of specifically placed groups.

Like many other methods, this strategy lacks spatial flexibility, being tied to regions of the structure where termination of the staple strand is tolerated, and requires redesign to move attachment points by even a few nanometres. In addition to this point, though the dock point on the surface is static, the length of the tether may lead to a large degree of spatial freedom.

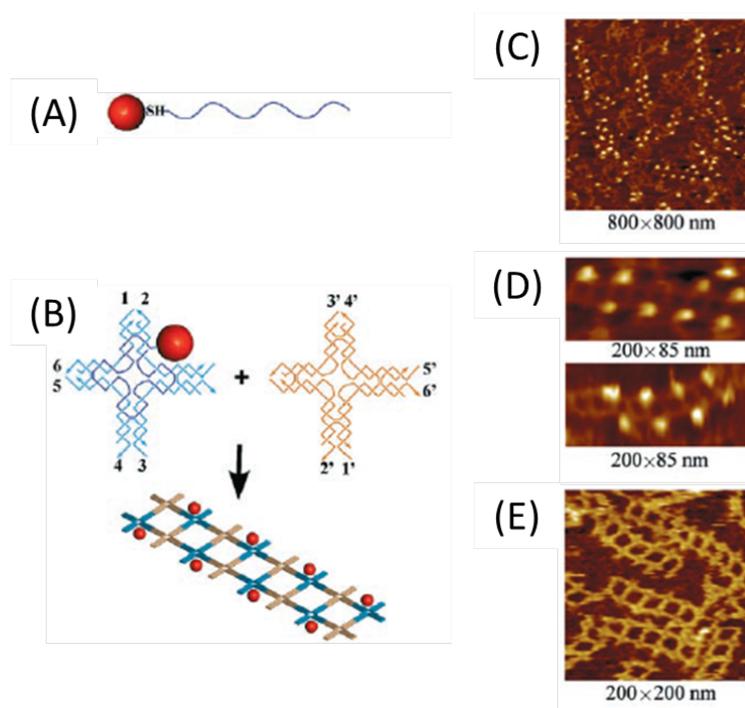


Figure 2.12: **A:** Gold particle (red) conjugated to a thiolated DNA strand (purple). **B:** Diagrammatic representation of the formation of a nanorarray of gold particles. DNA self assembles into cruciform structures (blue and yellow) with sticky ends that facilitate lattice formation. The DNA conjugated gold nanoparticles hybridise to the central strand of the blue cross. **C:** A large scan area AFM image shows a number of successfully formed structures. **D:** Two zoomed in images of individual arrays of DNA with gold nanoparticles. **E:** DNA lattice shown without gold nanoparticles. Height scales are approximately 2 nm, scan sizes are shown below the image. Images taken from [130]

2.3.4 DNA aptamers

DNA aptamers are sequences that act in the same way as an antibody. Through a combination of intermolecular forces, including charge effects and the hydrophobic effect, DNA aptamers are able to bind to specific targets such as proteins. It has been shown that DNA aptamers are able to direct the formation of protein arrays [132].

The work of Liu *et al.* used DNA aptamers as an alternative to DNA hybridisation for binding targets to specific points on a DNA structure [133]. They created a lattice of DNA tiles each of which presented a thrombin-binding DNA aptamer. The structures self-assembled to produce linear arrays of the protein [133].

The work of Douglas *et al.* incorporated DNA aptamers and complementary strand annealing to create a DNA 'nanorobot' that opened in response to the addition of a protein [85]. This work used gold nanoparticles to identify the opening of the box through TEM. The structure resembles a clam, shown in Figure 2.9, and is held closed through the annealing of two sets of partially complementary ssDNA strands on the lip of each half. When the protein is added it binds to one of these ssDNA molecules, an aptamer for the protein. In order for the structure to open both proteins must be present, making this structure equivalent to an AND gate [85]. The aptamers were designed to react to proteins on the surface of cells to deliver their payload to specific cell types.

Li *et al.* used DNA aptamers to bind single-chain antibodies, which could then be used as adapters for the binding of numerous other molecules [109]. These single-chain antibodies are the fusion of a single light and a single heavy chain, making them smaller than a conventional antibody, while retaining binding specificity. This makes them potentially useful in a number of fields where the larger size of a conventional antibody would decrease the spatial resolution of binding.

The benefits of using DNA aptamers are many: they can be built into structures, they bond specifically to chosen targets and are inexpensive to produce. DNA aptamers are easily incorporated into complex DNA structures, provided they are built into accessible facets. Some of the negative aspects of aptamers are similar to other DNA based strategies. They have to be preprogrammed into the structure, limiting their flexibility. The target specificity of aptamers could be considered a weakness in some circumstances. Most

importantly, DNA aptamers have to be discovered and optimised in a time-consuming process. This places a time constraint on the development process and could limit the usefulness of DNA aptamers for addressing DNA structures.

2.3.5 Small molecules

Synthetic chemistry allows the development of entirely new molecules with specific shapes and charges engineered for a particular role [12]. Currently these molecules are developed mostly for gene activation or inhibition for roles in disease. An example of how an oligonucleotide conjugated to a polyamide can intercalate into the major and minor groove of a DNA helix can be seen in Figure 2.13.

Gottesfeld *et al.* have produced pyrrole-imidazole polyamides that recognise sequences in the minor groove of dsDNA strands [134]. They are among a number of groups developing these structures for the inhibition or activation of gene expression [135]. This does not exclude their potential use in synthetic DNA systems: their ability to displace DNA binding proteins could indicate a high enough affinity for their target sequence to access areas of DNA with high levels of tension, as can be found in certain DNA structures [88]. Small molecules are designed for specific topologies; this could negate the negative effects of high tension on binding, as long as the level of tension was present in the optimisation steps.

An advantage of small molecules in DNA structures is ease of sequestering; it would be possible to place these groups within a DNA structure for release upon their digestion or in response to a change in conformation.

Small molecules could be an extremely versatile way of functionalising DNA structures, particularly through their ability to conjugate to a wide range of potentially therapeutic moieties. DNA structures could sequester small molecules until their degradation *in vivo*, reducing the off-target effects of cytotoxic compounds.

Work using pyrrole-imidazole polyamides in DNA nanotechnology has proved successful. These molecules are able to bind to origami tiles and, through attachment of biotin, direct the recruitment of streptavidin to the origami surface [136, 137]. Pyrrole-

imidazole polyamides have also been used to direct the formation of gold nanoparticles [138], to reinforce struts in DNA architecture [139] and modified to direct the sequence-specific alkylation of DNA in a nanoscaffold [70]. From these results it is clear that the impact of pyrrole-imidazole polyamides could be extensive.

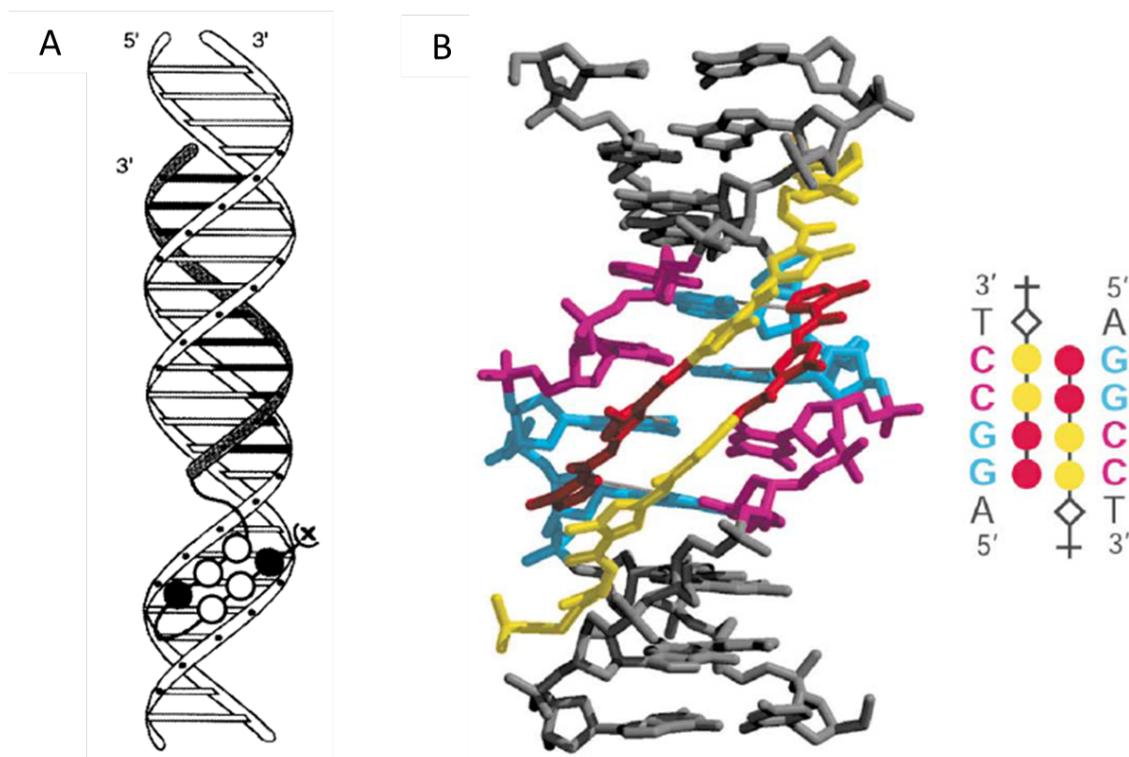


Figure 2.13: **A:** An example of an oligonucleotide conjugated to a polyamide targeting both the major and minor grooves of DNA simultaneously. **B:** Stereodiagram of the complex of polyamide and DNA shown in (A). Images taken from [12].

2.4 Proteins in DNA nanotechnology

There are a great number of proteins that interact with DNA *in vivo*. These include polymerases, which act to extend a strand of DNA by copying the sequence of another strand, and restriction enzymes, which have the ability to break the bond between bases. Restriction enzymes search for specific sequences and cut at the same point within that sequence every time.

DNA-protein interactions have already been optimised through evolution, it is thought

that proteins could potentially fill a number of roles in nanotechnology [36].

A recent review by Gradišar and Jerala [36] explores the potential of proteins for building nanostructures. This work largely explores the use of pre-existing protein motifs as building blocks, similar to the basis for work on RNA nanotechnology. The use of these motifs limits the flexibility of the structures created. Natural protein polymers contain 20 amino acids with different characteristics and chemical moieties, potentially producing a more functional molecule, but less useful as a chemically unreactive scaffold for the organisation of active groups.

Peptides including unnatural amino acids are already commercially available and could introduce a wide range of chemical functionalities to nanostructures. The length of synthetic peptides is limited due to a level of failure in the stepwise addition of amino acids that leads to truncated sequences. Peptide synthesis relies on favourable sequences, for example highly hydrophobic peptides can prove difficult to synthesise and the final product can prove difficult to solubilise. There is also work to introduce unnatural amino acids into proteins *in vivo* which has been recently reviewed [140].

The availability of biotinylated oligonucleotides and the high affinity of the interaction between biotin and streptavidin makes this a promising route for protein-DNA interactions [47]. Park *et al.* used the biotin-streptavidin interaction to create patterns on the surface of a DNA tile [111]. Their work used stepwise addition of elements to the DNA structure, starting with loop regions modified with biotin. The addition of streptavidin then created 'pixels' on the surface of the DNA array, demonstrating precise control of the patterning process. Biotin-streptavidin binding is reliable but not very flexible; streptavidin will bind to any biotin present, allowing only one species of functional element to be added to the structure.

The use of pre-existing proteins for addressing DNA nanostructures could provide a wealth of tunable interactions for functionalising scaffolds. The key to the usefulness of these interactions lies in positioning the protein on the DNA; unless a protein can provide better resolution than currently existing technologies it will never be accepted by the community.

DNA binding proteins

It has been demonstrated that restriction enzymes can be modified to reduce their activity. This has led to the development of nicking enzymes, mutated restriction enzymes that are only able to cut a single strand of dsDNA rather than both. However, the recognition site is a requirement of these enzymes and they would require significant remodelling to change the sequence to which the enzyme bound.

There are groups of proteins called zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) which are used to modify genomes in a highly sequence specific manner, as shown in Figure 2.14. TALENs in particular are extremely interesting as they have modular domains that can be programmed to target specific sequences [141]. With the nuclease activity of these proteins removed, they could potentially be used to target DNA sequences in complex structures. An example of the uses of these proteins in DNA nanotechnology can be found in the work of Nakata *et al.*, in which ZFNs acted to site-specifically pattern DNA origami tiles [142].

There are a number of other proteins described here that demonstrate binding to specific DNA sequences; their affinity for DNA may be high but they demonstrate low flexibility in directing them to different sequences. Thus Cas9 appears to be a good candidate for controllable protein-DNA interactions

Cas9 is an RNA directed nuclease that is being used for a wide range of biochemical applications [143]. The RNA sequence directs the protein to a region of complementary DNA sequence where it breaks the DNA strand and for this reason is used as a tool for genome editing [144]. Versions of Cas9 that do not cleave are used for binding DNA sequences to control expression [145].

A body of work exists that investigates the nanotechnology possibilities of a similar protein called Recombinase A (RecA). Where Cas9 is directed by RNA, RecA is directed by DNA.

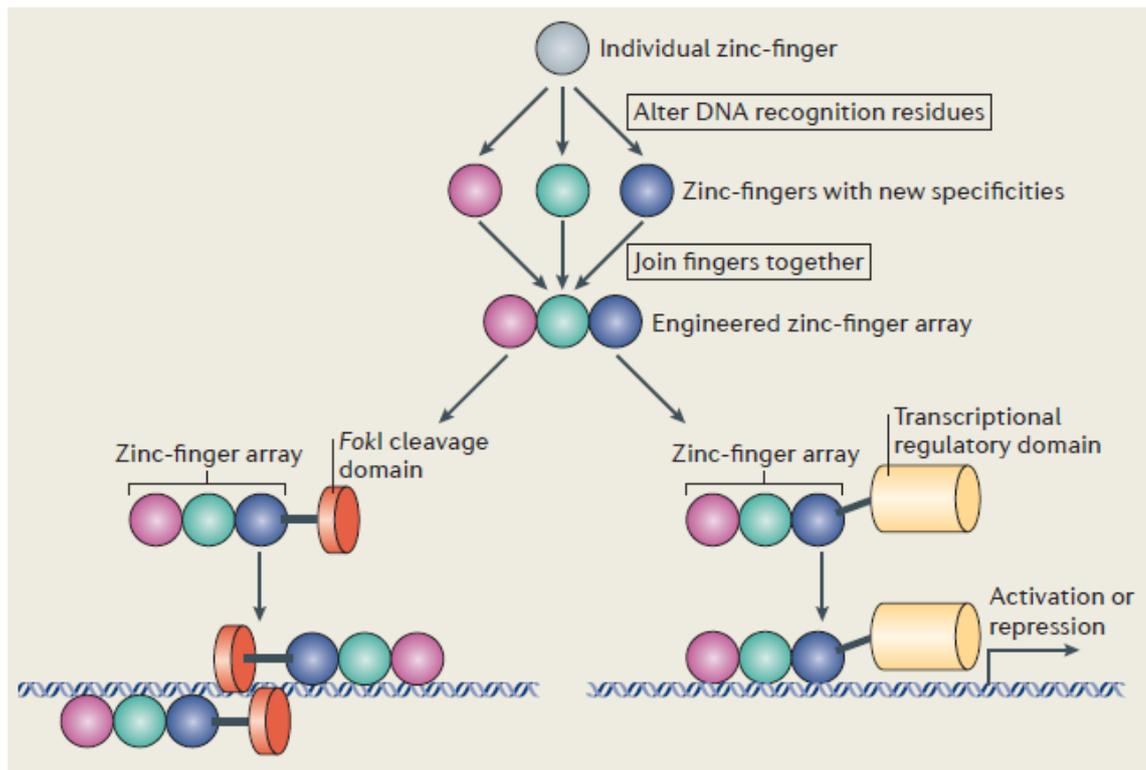


Figure 2.14: Modular zinc-finger nucleases (ZFNs) can be used to direct functional groups to specific DNA sequences. Cys2His2 zinc-fingers are directed, through a small number of amino acids, to bind \sim three base sequences of DNA. Joining a number of modules can target longer sequences, though not always with a high success rate. FokI is a non-specific restriction enzyme that is shown on the left directed to a specific sequence using a ZFN array. By mutating the enzyme to require dimerisation for activity the specificity can be further increased. The option on the right shows the fusion of a transcriptional regulatory domain. These promote or repress the action of RNA polymerase in transcribing DNA and can be used to control gene activity. Image taken from [141].

2.5 RecA

2.5.1 RecA function

Recombinase A (RecA) is an *Escherichia coli* protein; it is thought that RecA might be a valuable tool in the patterning of DNA structures. Its key role in the cell is homologous recombination, the repair of breaks in DNA [146]. It does this through identifying and

removing damaged sections of DNA, replacing them with an undamaged strand of the correct sequence. Typically the damage involves breaks in the phosphate backbone or

RecA has a number of other roles in the cell, in addition to its key role in recombination. RecA is involved in the SOS response, the cellular repair activity following an increase in DNA damage, though mediation of the cleavage of LexA and UmuD proteins [147]. Through its interaction with UmuD, RecA activates DNA polymerase V as part of SOS mutagenesis [148, 149]. This error-prone stage of the SOS response is regarded to be the last resort of cells following otherwise irreparable damage to the cellular DNA [148]. RecA has also been shown to regulate the activity of other *E. coli* polymerases, including polymerase II, III and IV [150]. The work in this thesis focusses on utilising the pathway by which RecA carries out homologous recombination, shown in Figure 2.15.

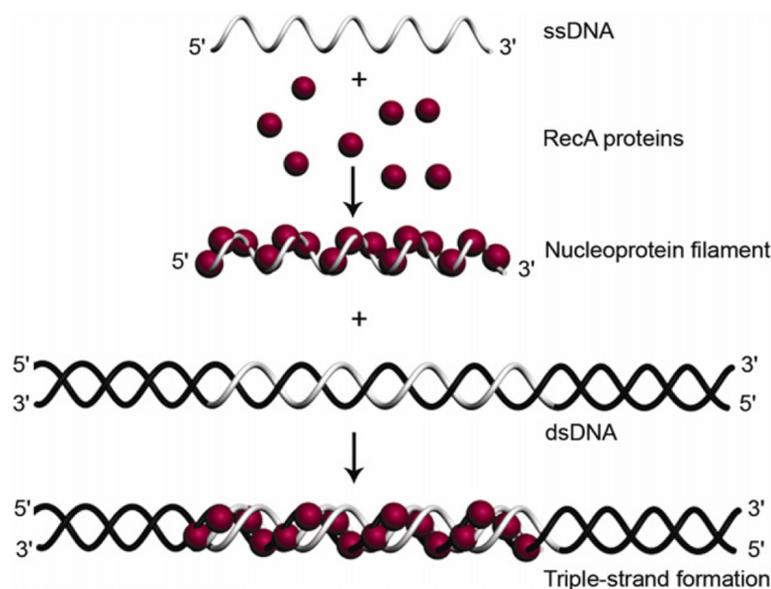


Figure 2.15: The binding of RecA to ssDNA forms a nucleoprotein filament. This complex will then bind to a homologous dsDNA strand and form a triple stranded region of DNA. Image taken from [16].

The aim is to use this system for addressing DNA in a sequence-specific manner, creating a tool by which these structures can be added to for increased functionality. In the presence of the nucleotide cofactor ATP, RecA forms a helical filament with ssDNA and searches for a region of homology within dsDNA [151, 152, 153]. When the region of homology is found a triple-stranded region is formed and RecA catalyses the exchange

of strands. It is possible through the use of a non-hydrolysable ATP analogue to maintain the triple-strand used for homologous recombination, creating a stable complex. The ATP binding site lies at the interface between monomers [152, 154, 155]. There are a number of models for the link between ATP hydrolysis and RecA mediated strand exchange [156], it is possible that the hydrolysis of ATP causes changes in the protein conformation at the monomer interface that induce the dissociation of the exchange complex [157].

2.5.2 RecA structure

RecA is a 37.8kDa protein. The structure of RecA with dsDNA has been visualised using electron microscopy [158] and AFM [159] and shows a number of different pitches for the RecA helix; it is thought these conformations may be linked to its activity in joint molecule formation. Crystal structures of RecA have also been solved [160, 161, 154]. The X-ray crystal structure of RecA is shown in Figure 2.16.

It has an acidic C terminus homologous with sequences in other ssDNA binding proteins and implicated in the modulation of DNA binding [162, 163, 164]. More than half of the terminal 25 amino acids are negatively charged or contain hydroxyl groups which would repel negatively charged DNA [163]. C-terminal truncated *E. coli* RecA protein RecA5327 has enhanced binding affinities to single- and double-stranded DNAs [165]. Benedict and Kowalczykowski found the rate of ATP hydrolysis increased from 2.47 $\mu\text{M}/\text{min}$ to 4.74 $\mu\text{M}/\text{min}$ when approximately 15% of the protein was removed at the C terminus [162]. Removal of the RecA C-terminus results in a conformational change in the RecA-DNA filament [162].

There are two loops in the RecA structures that have been associated with DNA binding at residues 151-176 and 190-227 [166, 167]. The primary site (151-176) binds the ssDNA and the secondary site (190-227) binds the dsDNA and later the ssDNA product of the exchange reaction [168, 169]. It has been suggested that a residue at position 161 prevents the binding of dsDNA but not ssDNA to the primary binding site [170]. Through the use of microfluidic channels and fluorescence microscopy, the persistence length of a RecA filament on dsDNA has been found to be 1.15 μm [171].

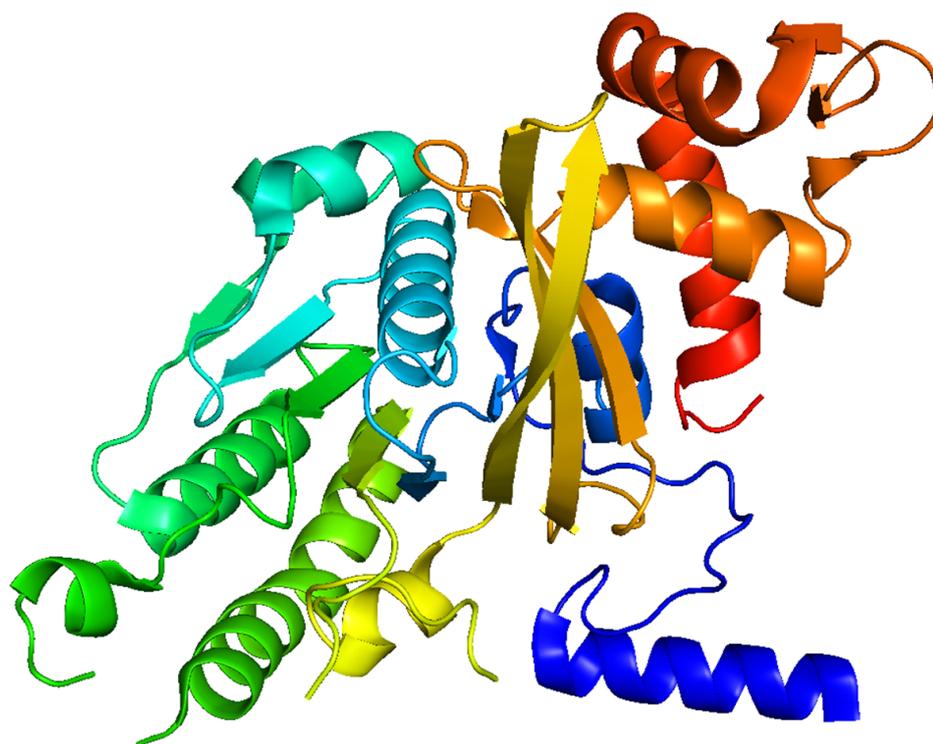


Figure 2.16: Cartoon diagram showing the secondary structural elements of RecA protein from *Escherichia coli* (Protein Data Bank file 4TWZ) Resolution is 2.8 Å. The protein is coloured from blue (N terminus) to red (C terminus). Image produced using PyMOL.

A number of RecA mutants have been characterised which show hyperrecombination through faster formation and increased stability of the nucleoprotein filament, and promotion of strand exchange, among other factors [172, 173, 174]. A hybrid *E. coli* RecA which contained 12 amino acids from *Pseudomonas aeruginosa* demonstrated enhanced ATPase activity and more dynamic DNA binding [172]. The mutation of amino acid 112 from aspartic acid to arginine, changing the position from acidic to basic, created a RecA variant that was more able to displace SSB [173]. Directed evolution of RecA protein was able to produce a large number of mutants with improved recombination functionality, the amino acid changes were targeted to the subunit-subunit interface [175].

RecA mutants have also been used to explore the mechanism of RecA mediated strand exchange [156], for example through removing the ability of RecA to hydrolyse ATP [176]. The mode of action of RecA in the induction of the SOS response has also been examined using mutants [177].

There are three phases to the recombination activity of RecA; the presynaptic phase in which the nucleoprotein filament is formed, the synaptic phase in which the filament aligns with the dsDNA strand and the postsynaptic phase in which the strands are exchanged [178]. The interaction of RecA and DNA to form nucleoprotein filaments and regions of triple stranded DNA is shown diagrammatically in Figure 2.15.

2.5.3 Formation of a nucleoprotein filament

A single RecA monomer binds to a 3 base stretch of ss or dsDNA [154]; it can bind to any sequence though some studies suggest a preference for a GT rich sequence [179]. The binding of RecA causes dsDNA unwinding to approximately 1.5X its original length [152, 180, 181, 158, 182].

As few as two RecA monomers in complex have been shown to be sufficient for nucleation, the rate of this process is dependent on RecA concentration, tensile force, nucleotide cofactor, pH, buffer and magnesium concentration [183, 184, 185, 186]. The nucleation step is slow, followed by a faster extension step [187, 188]. RecA protein has been shown to bind preferentially to longer ssDNA oligomers but the same effect is not seen for dsDNA strands [189].

ATP hydrolysis is a cooperative process, hydrolysis causes changes to the structure of RecA that induce hydrolysis in neighbouring proteins [168, 190, 191]. ATP hydrolysis is necessary for the dissociation and recycling of the RecA monomer [152, 192, 180]. The energy required for strand exchange is not taken from ATP hydrolysis [193, 194, 195], it has been proposed that entropically derived energy drives the reaction [196]. The ATPase rate of RecA was measured using coupled spectrophotometry to be 0.53 min^{-1} in the absence of DNA [179]. In the same study the rate was found to be between 9 ± 1 and $21 \pm 1 \text{ min}^{-1}$ in the presence of DNA, the rate depended on the sequence of the DNA [179].

The equilibrium dissociation constant (K_d) is used to describe the affinity between a protein and its ligand. The units of K_d are molar, describing the molarity at which half of the available binding sites are occupied. The K_d of RecA on ssDNA has been estimated using FRET experiments to be $\sim 100 \text{ nM}$ at 5' end and 8 nM at 3' end of the filament

[151]. The affinity of RecA for DNA is much higher at 3' end; filament growth occurs in the 5' to 3' direction.

There are distinct pathways *in vivo* for loading RecA onto ssDNA for recombination which take place under different circumstances and through interaction with different proteins [197, 198]. RecA is recruited to ssDNA gaps in dsDNA through the action of RecF, RecO and RecR [198], while only RecO and RecR are required when loading RecA onto a region of ssDNA not proximal to dsDNA.

The recombination activity of RecA *in vivo* is modulated by a number of proteins, including DinI and single-strand DNA-binding protein (SSB) [199]. DinI binds to RecA in cooperative manner and can substantially stabilise the RecA filament [200, 201]. RecA and SSB actively compete for binding on ssDNA [151, 202] but SSB added after RecA nucleation melts secondary structure in the ssDNA and aids filament extension [175].

2.5.4 Filament extension

While the binding rate of RecA increases with increasing RecA concentration, dissociation has been shown to be independent of this variable [151]. Many monomers bind simultaneously during nucleation, additional RecA proteins bind as single monomers [151, 180].

The secondary structure and length of the ssDNA strand can have an effect on the efficiency of the interaction with RecA, as measured by ATPase activity [168]. The change in net growth rate of filaments with respect to RecA concentration is higher in the presence of ATP γ S than in the presence of ATP [183]. The sequence of the ssDNA has been shown to have an effect on stability of the filament through changes in the rate of monomer dissociation [203]. It was found that a sequence comprised only of 5'...TGG...3' repeats showed the slowest rates of monomer dissociation of the sequences tested, 5'...CCA...3' repeats showed the highest [203].

Overall extension of the filament occurs in the 3' direction, although monomers do bind to both the 5' and 3' ends [151, 180]. Filament growth has been shown to have a rate of $44 \pm 11 \text{ nm min}^{-1}$ in the 5' to 3' direction and a rate of $27 \pm 12 \text{ nm min}^{-1}$ in the reverse

direction [183]. The dissociation rates for the 5' and 3' ends of the filaments are very similar, the directionality of filament growth is dictated by the difference in binding rates [151]. RecA monomers cannot dissociate from the middle of the filament [151]. Under physiological conditions the growth rate of RecA filaments has been shown to be 50 - 500 nm min⁻¹ [183].

2.5.5 Triple strand formation

There are multiple models describing how the RecA nucleoprotein filament searches for homology within dsDNA [204, 205]. The homology search is not well understood, particularly the scale of the search required [206, 168]. Camerini-Otero and Hsieh estimated the base-sampling frequency of the search to be $>10^3 \text{ s}^{-1}$ [207].

Recent work has shown that the nucleoprotein filament slides along dsDNA to aid its search [208]; until this study it was assumed that the search process was diffusion limited [153, 209]. The 1D filament slide was measured as $0.9 \times 10^{-3} \mu\text{m}^2/\text{s}$ using FRET measurements [208]. It is thought that the homology search includes diffusion in three dimensions, sampling between non-contiguous regions and random sliding [178].

The process by which RecA samples sequence homology between the ssDNA and dsDNA strands is not fully understood but it is believed that dsDNA is partially melted to allow Watson-Crick base pairs to form between the incoming strand and its complement [210, 211]. Evidence suggests the sampling involves three base pairs at a time [212, 209]. Rotation of the dsDNA strand around its longitudinal axis is a requirement for strand exchange [213] and mechanically stretching DNA improves the rate of the homology search [214]. A triple-stranded region is then formed in complex with the homologous region of dsDNA in a process referred to as patterning [215, 216, 192].

The fast initial searching step is followed by a slower step when formation of the triple-stranded complex takes place [217]. Joint molecule formation takes place in 3 base steps [209] followed by strand exchange, which takes place in the 5' to 3' direction [216, 218]. ATP hydrolysis is not necessary for strand exchange to take place for short DNA strands [218]. The rate of strand exchange has been calculated to be $2.25 \pm 0.04 \text{ nt/s}$ at 22 °C [182].

Enzyme activity can be described using the constant k_{cat} , which gives the rate at which the enzyme processes substrate molecules. The k_{cat} of RecA has been measured to be 30 min^{-1} when bound to ssDNA and 20 min^{-1} when bound to dsDNA [219]. These measurements indicate that the rate of enzymatic activity is decreased when bound to dsDNA, to stabilise the nucleoprotein filament during strand exchange.

It is possible to indirectly observe RecA patterning through the action of a restriction enzyme; when the recognition site of the enzyme is within the triple-stranded region the enzyme is blocked from digesting the DNA, reflecting the efficiency of triple-strand formation [217, 181, 220]. Deproteinated complexes have been shown to form very stable triple-stranded DNA regions with melting temperatures comparable to those of dsDNA sequences of the same sequence [221, 215].

It had previously been thought that only larger ssDNA oligomers were capable of directing triple strand formation [209, 181, 168, 222, 223]. It has been shown that it is possible to form a stable reaction intermediate with a ssDNA oligomer of 6 bases; larger oligomers do appear to increase the efficiency of joint-molecule formation [208, 220, 18]. The highest efficiencies are seen when the ssDNA is completely homologous to the dsDNA strand [215].

2.5.6 Factors affecting patterning

When supplied with ATP, as *in vivo*, this triple stranded region is short lived; recombination takes place and the complex disassociates. When supplied with ATP γ S, a non-hydrolysable ATP analogue, the recombination stalls and the triple stranded region becomes stable [151, 192, 180, 181, 168, 220]. The rate of ssDNA catalysed hydrolysis of ATP (20 min^{-1}) is 2000-fold higher than that of ATP γ S (0.01 min^{-1}) [190]. A number of studies have used ATP γ S in place of ATP as a method of stalling the homologous recombination activity of RecA [162, 224], allowing the DNA-RecA complex to remain in place for study. The stability of complex formation is increased at the terminus of dsDNA [221, 18]. RecA will nucleate on DNA faster in the presence of ATP γ S than ATP [180]. ADP destabilises the interaction between RecA and DNA [155].

The affinity of RecA for DNA is strongly affected by salt concentration [180], optimum

activity for wild-type RecA occurs at 10 mM magnesium ion [225].

Information on the properties of RecA can be found in Table 2.1. While these measurements provide some information about the activity of RecA as a DNA binding protein and an ATPase, these variables are heavily dependent on the concentrations of the salts, including NaCl [180] and magnesium [225]. Other factors affecting rates include DNA length, DNA sequence and the presence of additional DNA-binding proteins, particularly *E. coli* single-stranded DNA binding protein (SSB) [226, 227].

Table 2.1: Reported properties of the *Escherichia coli* enzyme Recombinase A

Property	Value range	Salts	Citations
Nucleation frequency (dsDNA)	1.6×10^{-4} to 2×10^{-5} nuclei $\text{min}^{-1} \text{bp}^{-1}$	20-80 mM NaCl	Fluorescence microscopy [180]
Nucleation occurrences (dsDNA)	ATP $1.86 \pm 0.06 \text{ min}^{-1}$ ATP γ S $10.2 \pm 0.6 \text{ min}^{-1}$	5 mM	
Assembly rate (dsDNA)	$3.6 \pm 1.1 \text{ nm s}^{-1}$	Mg(OAc) ₂	
Disassembly rate (dsDNA)	<10 pM	4 mM Mg(OAc) ₂	Fluorescence microscopy [153]
ATP hydrolysis	0.5 s^{-1} 0.18 s^{-1}	10 mM MgCl ₂	36 °C [228] 22°C
Rate of search Complex formation	2 to $4 \times 10^{-4} \text{ s}^{-1}$ 2×10^3 to 3×10^4 $\text{M}^{-1} \text{ s}^{-1}$	12.5 mM MgCl ₂	Restriction endonuclease assay [217]
Assembly ATP hydrolysis K_{cat} (ssDNA) K_{cat} (dsDNA)	1000 monomers min^{-1} $116.8 \mu\text{M min}^{-1}$ 30 min^{-1} 20 min^{-1}	17.5 mM Mg(OAc) ₂ 1.8 M NaCl	Coupled spectrophotometry 37 °C [224]
Disassembly (pH 8)	240 monomers min^{-1}	10 mM Mg(OAc) ₂ 1 mM KCl	[229]

2.5.7 RecA in nanotechnology

There have been a number of recent potential applications for RecA in the field of nanotechnology. A RecA filament has been integrated into a leaky surface acoustic wave biosensor to improve both sensitivity and detection time [230]. A graphene oxide-DNA system has been developed which uses RecA and ATP γ S as inputs for an AND logic gate [231]. RecA has also been used to increase the circuit rates of a DNA circuit up to 9-fold while also decreasing the signal to noise ratio [232].

The use of RecA for patterning DNA has been investigated by a number of groups. Keren *et al.* demonstrated the use of RecA to exclude the deposition of metal on a dsDNA structure in a sequence specific manner, allowing the production of metallic nanowires with insulating gaps, as shown in Figure 2.17 [13]. The limitation of their work was the size of the nucleoprotein filament with which they patterned the DNA; they only presented data for patterning with a filament of 500 bases, 170 nm in length.

Nishinaka *et al.* developed a protocol for metal deposition on dsDNA at the site of RecA patterning [14]. This work involved the addition of a cysteine residue to the C terminus of the protein to facilitate gold deposition.

Further work began to develop this technology with the aim of eventually patterning complex DNA structures in a sequence specific manner with nanometer resolution. In 2010 Wälti *et al.* used RecA to create triple stranded regions that were ~ 20 nm in size [15]. They were able to use AFM to image the nucleoprotein complex *in situ* and confirm that triple stranded regions were forming in a sequence specific manner [17]. This was then followed by a demonstration of the ability of RecA to pattern a single dsDNA strand at multiple points [16].

The Wälti group were able to demonstrate the positioning of the RecA nucleoprotein complex through the use of restriction enzymes; where the triple stranded region had formed over its recognition sequence, ApoI was not able to digest the strand [15, 16].

The work of Sharma *et al.* demonstrated that RecA could be used to pattern DNA in a sequence specific manner with nucleoprotein filaments formed of much shorter ssDNA sequences than those used by Keren and Nishinaka [13, 14, 15, 16]. They successfully

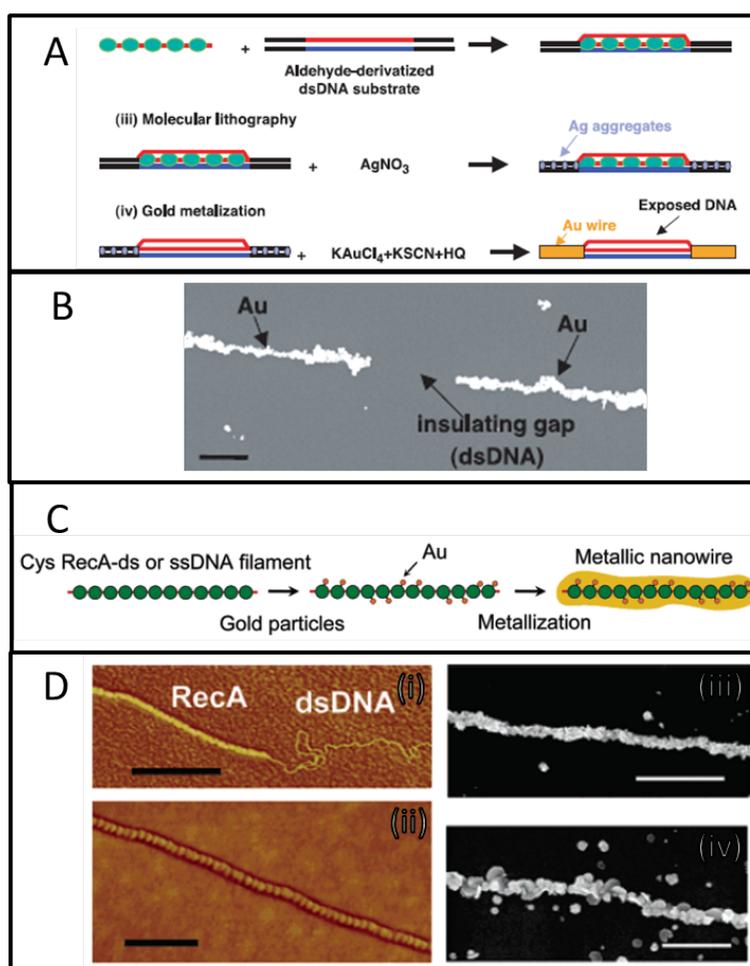


Figure 2.17: **A:** Diagrammatic representation of the method of metal deposition on DNA from Keren *et al.* [13]. The RecA-DNA filament binds to the aldehyde-derivatized dsDNA and is then incubated in an AgNO_3 solution. The aldehyde group reduces AgNO_3 and leads to the formation of Ag aggregates where the RecA is not present; these then act as nucleation points for gold deposition. **B:** Scanning electron micrograph of gold selectively deposited on DNA as described by Keren *et al.* [13]. Scale bar is 500 nm. **C:** Diagrammatic representation of the method of deposition from Nishinaka *et al.* [14]. RecA has been modified to have an additional cysteine residue on the surface at a distance from the functional site of the enzyme. Gold nanoparticles form a covalent bond with the cysteine residue and this is used for the nucleation of a gold nanowire. **D:** AFM analysis (left) of RecA polymerising on dsDNA, the proteins used are (i) wild-type RecA and (ii) cysteine modified RecA. Scanning electron micrographs (right) of metal selectively deposited on DNA as described by Nishinaka *et al.* [14]. The images show nanowires of (iii) gold and (iv) silver. Images taken from [13] and [14]. Scale bars are (iii) 1 μm and (iv) 600 nm.

patterned DNA with a nucleoprotein filament formed around a 60 base ssDNA, 20 nm in length. Their aim was to find a way to pattern preexisting DNA structures. The use of the sequence specific activity of RecA and ssDNA to pattern a dsDNA sequence could allow the positioning of conjugated species or protection from processes affecting the rest of the structure.

Further work has since been carried out which shows patterning is possible for sequences as small as 6 bases, or ~ 3 nm [18]. In order to develop RecA as a DNA patterning tool it is necessary to understand better the limitations of RecA, specifically regarding levels of tension in 3D DNA structures and how this might be affected by patterning. RecA straightens DNA molecules [159]; the reduction in flexibility produces a less energetically favourable conformation.

There is also work towards higher resolution characterisation of RecA:DNA supramolecular structures using AFM. Although many images have been taken they are not of a high enough resolution to visually confirm the effect of RecA binding on dsDNA proximal to the binding site [159, 233, 234]. This could provide information invaluable to the use of RecA as a structural patterning tool.

In order to investigate the potential of RecA as a tool for patterning complex DNA structures, it is necessary to test it on different topologies of increasing complexity. Work has focussed on the production of structures with regions of selective annealing to create non-base pairing regions and branched junctions, with the option to synthesise unpaired regions of DNA at either end or at some point in the centre. The ssDNA structures are hereafter referred to as unpaired branched junctions.

Initial work on the patterning of DNA sequences with unpaired branched junctions indicated that there was an increase in patterning efficiency associated with these structures. Work in the Wälti lab found that when there were unpaired branched junctions at both ends of the strand patterning was more efficient than when a single unpaired branched junction was present, and that the presence of a single unpaired branched junction rendered patterning more efficient than on fully base-paired DNA (Unpublished report, Faith Bateman).

The implications of these results on the patterning of DNA and more complex DNA

structures mean that further work into this area is essential. In this project new methods for making the unpaired branched junctions were attempted. By testing these structures we could draw conclusions about the effect of unpaired branched junctions on RecA patterning. This information could be used to draw further conclusions about RecA and its potential for use with DNA structures.

2.5.8 Conclusion

The applications of complex DNA structures have not been fully realised, but the potential scope is large [10]. In the last thirty years DNA structures have advanced from their inception to complex shapes with switchable elements [49, 10]. From the first, stable four-way junctions [31], structural elements were reinforced [52] and then completely re-imagined in the development of DNA origami [50]. DNA can take on a variety of complex shapes in three dimensions [42] and respond to stimuli [43, 85] in role vastly different to that it takes on *in vivo*.

Methods for addressing these structures are key to the usefulness of DNA as a material. Functionalisation of DNA structures is an area of great interest [44]. DNA nanotechnology is a multidisciplinary field, requiring the concerted work of chemists, biologists, physicists and computer scientists [44]. In return, it may be a tool for the advancement of many fields.

In keeping with the origins of DNA, there are a number of biological as well as synthetic ways to address DNA structures. The use of unnatural and functionalised bases has not yet been realised in the context of complex DNA structures. However, this is a potential avenue to highly stable, base-specific additions [48]. Single stranded regions designed into the structure as binding docks for additional elements are widely used due to ease of design and low cost. This method requires DNA to be attached to the functional group and for the temperature to be relatively stable. DNA aptamers can be designed to attach to most other biological molecules, their weakness is in the need for optimisation and their inflexibility regarding targets. Designing small molecules to interact with DNA depends on the ability to manufacture entirely new chemicals but could be extremely versatile. The use of DNA-protein interactions to address structures is very much in its infancy, but

could prove to be extremely flexible, targeting specific DNA sequences within a structure.

The loss of flexibility with an association that is built into the DNA architecture is a theme common to many of the methods discussed here. We believe that RecA can potentially address this deficiency through sequence specific interactions with complex DNA structures. In order to apply the RecA patterning system to origami and similar DNA scaffolds it is first necessary to understand the effect of these structures on the protein. By applying RecA to increasing complex topologies, beginning with simple dsDNA and proceeding in a stepwise fashion, we can understand the potential that RecA holds.

Chapter 3

Techniques, Protocols and Reagents

Recombinase A (RecA) is an *Escherichia coli* protein with potential applications in the field of DNA nanotechnology; *in vivo* RecA has a role in homologous recombination [146]. It has been used to pattern DNA to control metal deposition [13, 14]. Recent work has focussed on using short ssDNA strands in combination with RecA protein to address dsDNA in a site specific manner, allowing patterning on a much smaller scale and with multiple patterning sites on the same molecule [18, 16, 15, 17]. The techniques and methods described here are used with the aim of extending this work; to use RecA to pattern more complex DNA topologies to demonstrate its potential for patterning DNA arrays and origami.

To demonstrate the potential of RecA for DNA nanotechnology DNA was fabricated with non-base pairing regions and three-way junctions. The techniques used are introduced in the first section of this chapter. In the second section the protocols used to fabricate and investigate a series of DNA structures and RecA protein itself are outlined. In the final section the provenance of the reagents used in these experiments is identified.

The oligonucleotides needed to fabricate different DNA constructs were all designed and ordered as synthetic sequences. Where necessary, DNA was amplified using PCR and the products of the reaction then purified. The most simple construct was the result of two, annealed synthetic DNA strands and required no further purification. Longer, branched DNA was created using four synthetic oligonucleotides through a templated ligation reaction in a thermocycler, purification of which was undertaken through native

or denaturing gel electrophoresis. Another construct was synthesised by PCR from phosphorothioated DNA sequences. The restriction enzyme assay used to determine the efficiency of RecA patterning is described and DNA digestion protocols are also discussed. Gel electrophoresis was used to evaluate both the quality of fabricated constructs and the efficiency of RecA mediated patterning on these scaffolds. Atomic force microscopy (AFM) was used to visualise different DNA constructs and patterned structures.

RecA protein was expressed through *Escherichia coli* transformation, growth, induction and harvesting. The soluble cell lysate of treated bacteria was purified using fast protein liquid chromatography (FPLC). Following this, an activity assay was used to assess the quality of the protein produced.

3.1 Techniques

3.1.1 Polymerase chain reaction

The DNA polymerase enzyme was originally isolated from *Escherichia coli* [235]. In 1969 the thermophilic bacteria, *Thermus aquaticus* was identified [236]. A paper describing *T. aquaticus* DNA polymerase was published in 1976 [237]. Due to the high temperature in which these organisms operate their proteins must be unusually thermostable; *T. aquaticus* DNA polymerase has an optimum operating temperature of around 80 °C [237]. The paper introducing the polymerase chain reaction (PCR) as it is currently understood was published in 1986 [238]; the method used *E. coli* DNA polymerase which was destroyed by each melting cycle, requiring the addition of enzyme at each extension step. A patent on the use of *T. aquaticus* DNA polymerase in PCR was filed a year later by a group including scientists from the 1986 paper; due to the high heat tolerance of the enzyme it is possible to carry out multiple cycles in a closed reaction mixture.

PCR is used to amplify DNA sequences. The modern reaction uses a thermostable DNA polymerase, an enzyme that extends DNA strands by creating a bond between the existing

strand and a deoxyribonucleotide (dNTP).

DNA polymerases copy an existing DNA strand to create a complementary strand. Each new strand acts as a template for further reactions so the increase in DNA strands with number of cycles is exponential. Primers are short DNA oligomers that are complementary in sequence to the 5' ends of the upper and lower strands of the template DNA sequence that is amplified. During PCR the primers are melted and then annealed to the template DNA and are extended by the polymerase.

The reaction mixture is cycled through a melting, annealing and extension temperature, to allow templates to be reused and for the products of previous reactions to act as templates. Thermal denaturation of DNA is initially carried out through incubation at 94 - 98 °C; at this temperature all hydrogen bonds have been dissociated. The length of the primers is optimised to produce melting temperatures (T_m) within 5 °C of each other, avoiding hairpins and potential primer dimerisation. The annealing temperature is specific to each pair of primers, typically between 50 °C and the optimum temperature for the polymerase in use. The annealing temperature is typically 0-2 °C below the lowest T_m of the two primers. The extension temperature should be optimised for the enzyme used, for example *Taq* polymerase from New England Biolabs is most active at 68 °C.

The PCR reactions in this thesis were designed following reference to Molecular Cloning by Sambrook and Russel [239].

3.1.2 Ligase Chain reaction

In order to make DNA structures with ssDNA termini it was necessary to use a method other than classical PCR, which only produces dsDNA. It is possible to use PCR to create a population of DNA that is predominantly ssDNA by changing the ratio of the primers used, to bias the production of strands towards a single product. It was decided for this work instead to use the ligase chain reaction (LCR) to connect two smaller oligomers to make larger structures than were typically available synthetically. This process would double the size of the available oligomers; the size of synthetic oligomers is limited by the synthesis process. This decision was taken on the understanding that the potential yields were higher and there would be fewer undesired products to be purified away.

LCR is shown diagrammatically in Figure 3.1. The ligase chain reaction was developed as a diagnostic technique for use with, or in place of, PCR for the detection of single base changes in sequences [240]. This was possible due to the ligation only occurring if the two oligomers were directly next to each other.

LCR works in a method analogous to that of PCR. The reaction mixture is comprised of a template dsDNA strand and short ssDNA strands complementary to the template, with a heat stable ligase, for example *Taq* DNA ligase. The DNA strands which make up the 3' end of each ssDNA strand have a 5' phosphorylation for the ligase enzyme. In the first step the DNA strands are thermally denatured. The sample is then cooled to an annealing temperature at which the complementary oligomers hybridise onto the template strand. The sample is then held at a temperature at which the ligase enzyme is most active and the 3' end of one strand is ligated to the phosphorylated 5' end of the strand annealed directly next to it, as shown in Figure 3.1. The temperature is then cycled again and the ligated products of the previous reaction act as template strands.

The method allows product to be formed beyond the initial step, so that smaller amounts of scaffold can be added to the reaction. Theoretically it is possible for the results of earlier ligations to act as a template, allowing the exponential increase of products seen in PCR. Thermostable *T. aquaticus* ligase was used in order that the DNA could be heated for melting without denaturing the ligase.

3.1.3 DNA purification

Following PCR it is necessary to purify the resulting DNA to remove short strands of DNA, for example unused primers and polymerase. Purification can also remove the small proportion of the population made up of short strands of DNA that are the result of unfinished polymerisation reactions; this can be the result of exhaustion of the reagents or heat damage to the enzyme. The Roche High Pure PCR Product Purification Kit was used according to the manufacturers instructions. First, DNA is bound to a glass fiber fleece in the presence of a denaturing agent. Short strands of DNA, for example those fewer than 100 nucleotides, do not bind strongly enough to resist elution during the wash steps. The buffer used in the washing steps includes ethanol, which precipitates the DNA

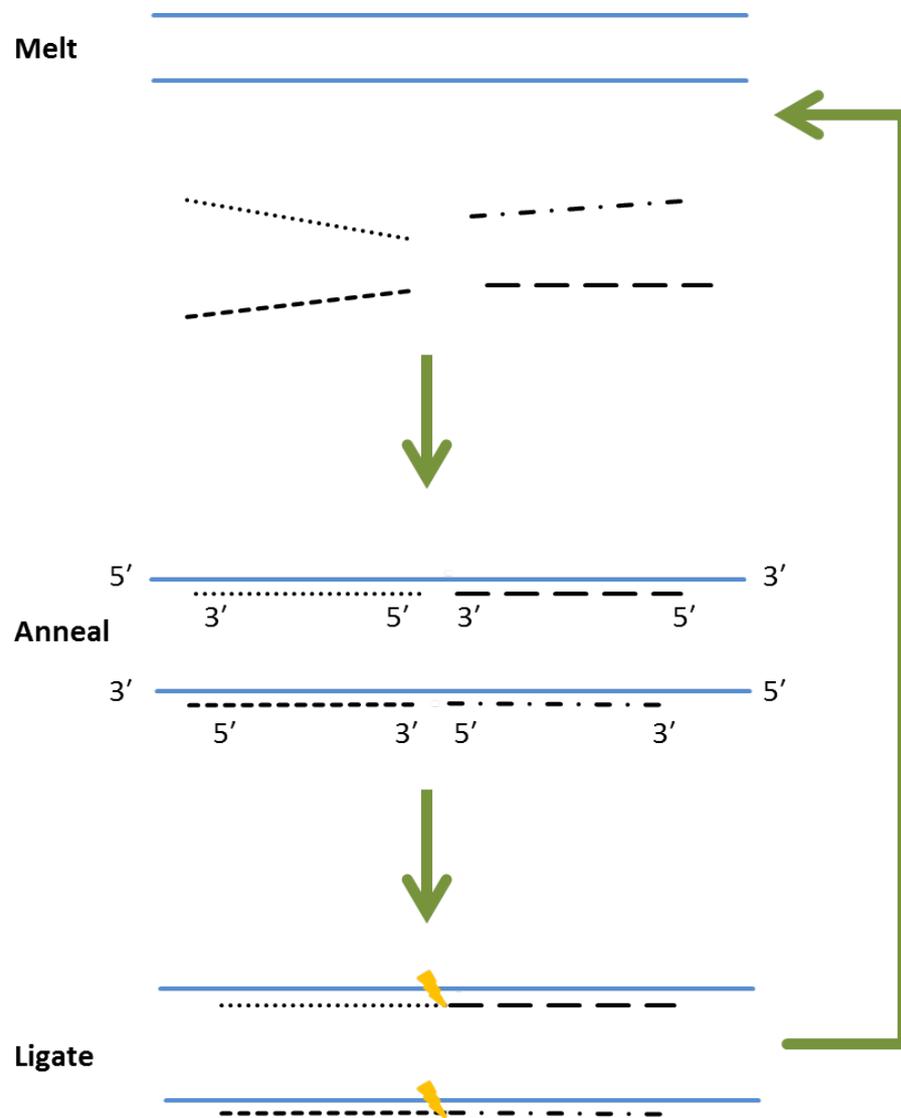


Figure 3.1: A diagrammatic representation of LCR. DNA scaffolds of the complementary strands of the desired sequence are mixed with shorter DNA strands comprising shorter sections of the same sequence. The strands are annealed and then held at a temperature at which the ligase enzyme can ligate the strands. The yellow arrow represents the action of the ligase enzyme. The entire reaction mixture is reheated to melt the DNA strands and then cooled for further annealing and ligation cycles. The successfully ligated products of earlier reactions act as scaffolds for future ligation reactions, akin to the exponential increase in product gained from PCR.

and prevents sample loss. The column is then saturated with a small volume of a low salt solution that contains no ethanol and the DNA is solubilised. Following an incubation step the solution is eluted, along with the DNA.

Purification was also carried out following T7 exonuclease treatment of partially phosphorothioated DNA, the method is described in Section 3.2.2. The action of the enzyme results in the presence of mononucleotides which purification aimed to remove; the enzyme itself cannot be heat denatured and therefore must be removed before any further treatment of the DNA. The Zymo ssDNA/RNA Clean and Concentrator works in a similar fashion to the Roche kit, but while the PCR purification kit aims to remove ssDNA primers, the Zymo kit is designed to retain ssDNA. The manufacturer states that ssDNA or RNA from 17 to 200 nucleotides can be recovered. The manufacturer gives no information about the method by which the purification works, but it can be assumed that it is analogous to the method of the PCR purification kit.

3.1.4 Gel electrophoresis

Gel electrophoresis exploits the charge and size of molecules to separate them as they run through a gel matrix, shown diagrammatically in Figure 3.2. The molecules can then be visualised and conclusions can be drawn about the population of molecules in a sample. Electrophoresis is widely used to explore populations of DNA, RNA or proteins.

There are two polymers generally used to produce the matrix for gel electrophoresis: acrylamide and agarose, shown in Figure 3.3. The polymer matrix is produced in the presence of a pH buffer and a combination of salts. The work described here uses both polyacrylamide gel electrophoresis (PAGE) and agarose gel electrophoresis.

DNA and RNA are negatively charged molecules. When an electric field is applied across a gel of polyacrylamide or agarose each nucleic acid experiences an electrostatic force and as a result the DNA or RNA is pulled through the pores in the matrix. Longer DNA or RNA molecules or those with secondary structural elements are less able to move through the matrix so they migrate more slowly, creating a series of bands of DNA species of different sizes. Samples are loaded into wells in the gel, as shown in Figure 3.2 to limit their sideways diffusion.

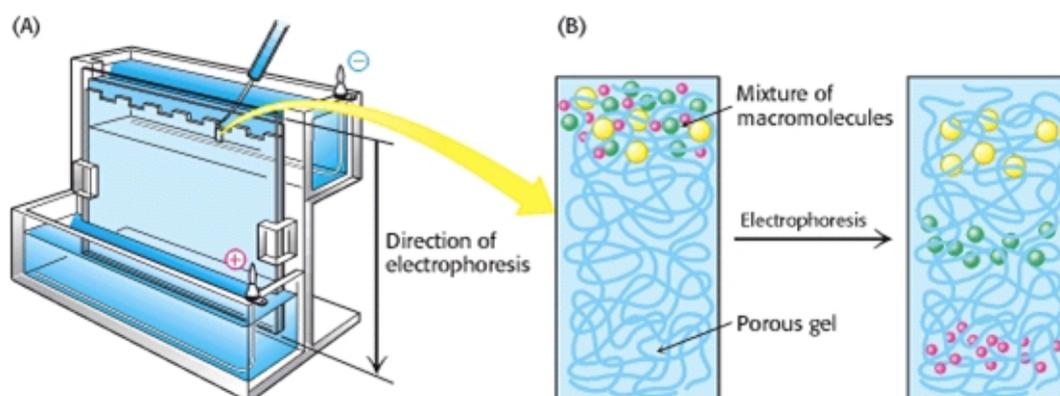


Figure 3.2: Polyacrylamide gel electrophoresis (PAGE) shown diagrammatically. (A) The setup for PAGE typically involves a gel of between 0.75 and 0.5 mm thickness, loaded at the top with a series of samples. The gel chamber is closed and a known voltage is applied at the top of the gel. (B) When the electric field is applied across the gel it causes the movement of charged molecules through the matrix. Larger molecules move more slowly through the gel. Image taken from [241].

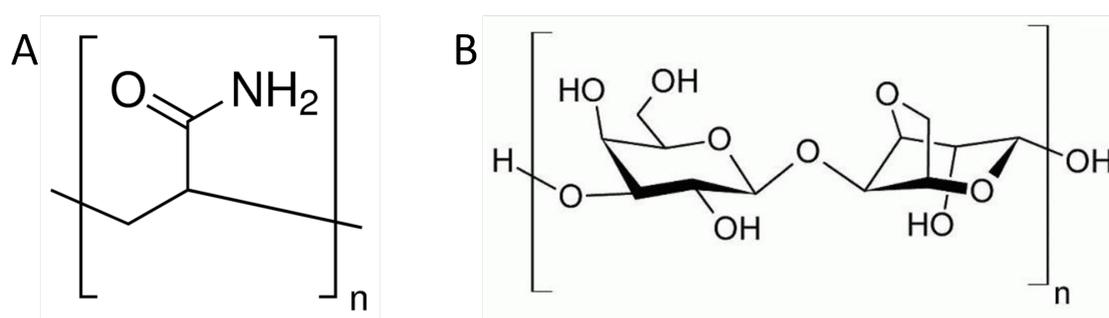


Figure 3.3: **A:** Acrylamide monomer showing the bonds that form to produce the polymer used in polyacrylamide gel electrophoresis (PAGE). Image taken from sigmaaldrich.com. **B:** Agarose monomer showing the bonds that form to produce the polymer used in agarose gel electrophoresis. Image taken from lifetechnologies.com.

As well as DNA and RNA other biological molecules, like proteins, can be separated and visualised using electrophoresis. Some of the amino acids that polymerise to form proteins can carry a charge at a range of pH values; at physiological pH glutamate and aspartate are negatively charged, lysine and arginine are positively charged. While some proteins carry an overall positive or negative charge, it is not spread evenly throughout the polymer and therefore will not react in a predictable way to an applied electric field as used in electrophoresis. For this reason proteins are usually imaged in denaturing conditions.

Native electrophoresis

In native electrophoresis samples are run in the natural fold of the DNA, RNA or protein. This method is widely used for dsDNA samples. The concentration of the polymer affects the density of the matrix, higher percentages lead to more densely packed matrix through which macromolecules move more slowly. This informs the concentration used for different sizes of molecules; as larger molecules move more slowly through the matrix, it is expected a lower percentage of polymer will be used to facilitate mobility through the gel and prevent extremely long running times. This can become important as gels tend to heat during running due to the electrical field applied.

Under native conditions dsDNA remains as a duplex, both ssDNA and RNA form hydrogen bonds with regions of base complementarity to form secondary structural elements that will affect their mobility through the gel.

Denaturing electrophoresis

There are a number of methods for imaging macromolecules under denaturing conditions. The methods used in this thesis are discussed briefly below.

SDS polyacrylamide gels Protein samples are imaged using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gels. SDS is an anionic surfactant that denatures proteins through interactions with the hydrophobic amino acids, such as

glycine. It is the charge on the SDS molecules that causes the denatured protein to move through the gel matrix. SDS is widely used in non-native protein imaging.

Urea polyacrylamide gels For denaturing polyacrylamide gels, DNA was denatured through the action of urea. The urea molecules form hydrogen bonds with the DNA bases which disrupt the base pairing between DNA strands and effectively lowers the melting point of strands. These gels are 0.5-0.75 mm thick; this concentrates samples within a small area to enhance resolution between bands, helping to distinguish between similar DNA species, and to aid visualisation of low-concentration samples. This means that only small amounts of sample can be loaded and reduces the usefulness of these gels in recovering separated DNA.

Alkaline agarose gels Denaturing agarose gels were made according to the protocol in Sambrook and Russel [242]. Alkaline gels denature DNA through a high pH; the concentration of hydrogen ions decreases, disrupting hydrogen bonding between bases. Agarose gels are much thicker than acrylamide gels, allowing a large amount of DNA to be loaded onto these gels and recovered following imaging. For this reason alkaline agarose gels were used to separate and purify DNA samples.

Gel analysis

A molecular marker, containing multiple species of known size is also run on the gel to provide a reference to the length of the nucleic acid or protein shown on the band. DNA is visualised using a fluorescent DNA intercalating dye; gels are incubated with the dye and then imaged under UV. Protein gels are imaged using Coomassie, a dye which visualises protein molecules by non-covalent interactions with basic and hydrophobic residues at acidic pH.

Densitometry is a method for analysing gels which can be used to ascertain the relative concentration of different species in a reaction, including ligation and restriction products. This involves the analysis of the total intensity in each lane and the calculation of the relative intensity of each band as a percentage of the total lane intensity. The amount

of fluorescence is proportional to the amount of DNA in the band. The background fluorescence of the gel is accounted for in this technique as bands of fluorescence are selected from the baseline and intensity is calculated without including the background.

3.1.5 Digestion

Restriction endonucleases are enzymes which cleave DNA to create multiple dsDNA strands. They are named for bacteria they were isolated in, for example EcoI is the first restriction enzyme to be identified from *E. coli*. Cleavage can result in either blunt ends, where there is no overhang, or sticky ends, where the upper and lower strand are not cut in the same place.

A recognition sequence dictates whether the enzyme will cut, where the restriction site is in relation to the recognition site depends on the type of enzyme. Recognition sites are specific to each enzyme, they can be four or more bases and can be remote from the site of restriction. Some enzymes require specific cofactors that can include ions such as Mg^{2+} or other small molecules, e.g. S-Adenosyl methionine.

Diagnostic digests were used to confirm the presence of the desired DNA products following LCR and to assess the patterning of dsDNA by RecA, described in Section 3.1.7. The activity of a unit of enzyme was calculated using activity data provided by the manufacturer on the amount of enzyme required to digest a given amount of DNA in a 50 μ l volume in 1 hour.

3.1.6 Fabricating ssDNA using T7 Exonuclease

In order to create large DNA constructs it is necessary to use PCR. The products of the PCR reaction are dsDNA, which is not useful for creating the desired topologies. In order to create ssDNA from the PCR products T7 exonuclease was used. T7 exonuclease is taken from the T7 bacteriophage, it is a DNA digestion enzyme which works by removing bases individually from the 5' terminus of the DNA strand. The exonuclease will not act on a row of 5 or more phosphorothioated bases, so these are included on the primer of the retained strand [243]. The structure of phosphorothioated bases is shown in Figure 3.4.

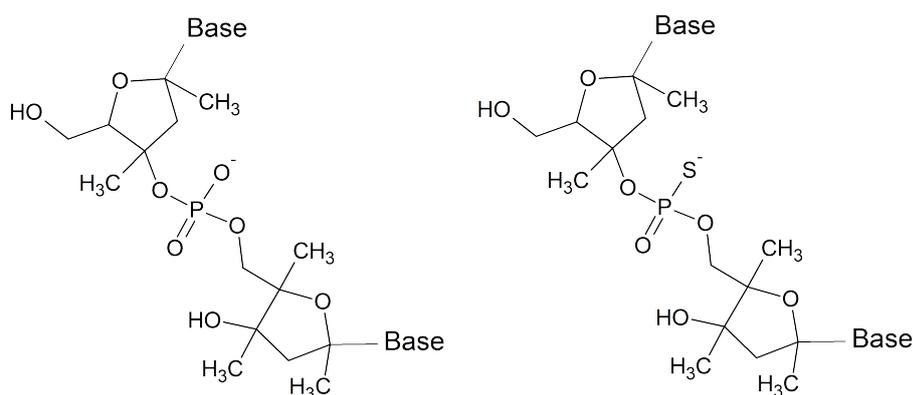


Figure 3.4: A diagrammatic representation of phosphorothioate modification of DNA. The internucleotide linker of the DNA polymer is altered from a phosphodiester group (left) to a phosphorothioate (right) through the replacement of an oxygen with a sulphur. This modification provides protection from endonuclease activity. Image taken from sigma.com.

3.1.7 RecA patterning

The patterning of dsDNA with a filament formed of ssDNA and RecA follows a number of steps that optimise the level of patterning achieved. This protocol is adapted from Wälti *et al.* [15].

A nucleoprotein filament is formed with RecA and ssDNA. An excess of a ssDNA oligomer composed of 30 thymine residues is used to react with the RecA monomers that have not bound to the complementary oligomer. This prevents the RecA from forming non-specific filaments on the dsDNA in the absence of the complementary ssDNA strand. The poly-T ssDNA has no sequence homology with the dsDNA and therefore no patterning will occur. The filaments are used to pattern the dsDNA in a sequence specific manner. A restriction enzyme is then used to test the extent of patterning. Where a RecA filament has bound to the dsDNA over a recognition site there is a reduction in the efficiency of restriction.

Proteinase K is added to digest both the restriction enzyme and the RecA bound to the dsDNA; the sample can then be run on a gel to assess the level of DNA digestion and the level of patterning that took place can be inferred from this.

3.1.8 AFM

Atomic force microscopy (AFM) is a type of scanning probe microscopy that can achieve resolution to the atomic level [59]. It involves the use of a cantilever behaving as a spring to ascertain information about a surface. Instead of directly measuring surface features, AFM measures information about forces on the tip, for example van der Waals interactions and capillary forces. This information is then used to create a topological image of the area studied.

The scanning tunneling microscope was developed in the early 1980s by Binnig, Rohrer, Gerber and Weibel at the IBM research laboratory in Zurich [57]. This was developed into the atomic force microscope by 1986 [58]. The probe is a cantilever with a sharp tip that can be only a few atoms across, most often made of silicon or silicon nitride. Piezoelectric ceramic crystals are used to raster-scan the cantilever over the surface in a highly accurate manner. A laser is reflected off the back of the cantilever onto a detector that measures the movement of the laser beam. An AFM can be used in different modes, to reflect the information required and the substrate under study.

The two most common modes of AFM imaging are contact and tapping mode. In contact mode the tip of the cantilever is firmly in contact with the surface and is rastered across it. Changes in the height of the sample cause the deflection of the laser beam to change and the height of the cantilever is changed to keep the force on the surface constant. The changes in the height of the cantilever can then be converted into an image of the changing topology of the sample. This method is appropriate for hard samples

When used in tapping mode the cantilever is oscillated at or near its resonant frequency, that is a preferential frequency at which the cantilever will oscillate with the highest amplitude. When in close proximity to the surface there are a number of tip-surface interactions that modulate the amplitude of the oscillation, such as Van der Waals and electrostatic interactions. Changes in the height of the sample cause changes in the amplitude of the oscillation of the cantilever, which can be detected using the laser deflection. This information is used to move the cantilever up and down, maintaining a constant distance from the surface, and providing the data for producing a topological image of the surface. The interaction with the surface is much less strong in tapping mode

than in contact mode, allowing for the imaging of much softer samples.

The images shown in this thesis were taken using PeakForce Tapping mode. This is similar to tapping mode, as it involves the intermittent contact of the tip with the surface. However, the cantilever is oscillating at a non-resonant frequency. As the cantilever moves closer to the surfaces the force it exerts increases. The maximum force of the interaction is set and a feedback loop allows individual peak force points to trigger retraction of the cantilever. This means that very low forces can be used while imaging giving greater detail to images of soft samples, such as biological molecules. It also gathers information about the forces between the cantilever and the surface at each point, potentially yielding large amounts of information about the surface in addition to topographical data.

AFM is used extensively for the study of DNA samples [43, 244, 60, 62, 82, 99, 42, 233, 83, 13, 108, 159, 109, 79, 88, 132, 102, 14, 111, 64, 63, 50, 130, 17, 18, 243, 234, 245, 15, 56, 246, 80, 101, 54, 53, 93, 61, 129]. In this work AFM was used to visualise DNA structures in tapping mode on a mica substrate. Mica is a naturally occurring mineral that cleaves into atomically flat layers. This makes it ideal for AFM imaging; the deposited sample is observed without obtrusive features from the surface. Both mica and DNA are negatively charged molecules, it is therefore necessary to introduce a bridging ion. Ions with positive charge can shield the charge from the mica and make it possible to deposit DNA on the surface. Divalent or trivalent ions can act as tethers for the DNA molecules.

3.1.9 Molecular cloning

Molecular cloning is the use of enzymes like restriction enzymes and ligases to isolate genes of interest and insert them into plasmids for the purposes of replication and protein expression. In the context of this body of work, a gene is a DNA sequence that codes for an RNA strand that directs the production of a single protein. A plasmid is a circular strand of dsDNA, most often found within a bacterial cell. Plasmids replicate independently of the bacterial chromosomal DNA, which can be used to increase the copy number from a single successfully ligated plasmid. Copies of plasmids can be transmitted between cells, this process is called horizontal gene transfer. Molecular biology takes advantage of the ability of bacterial cells to accept foreign DNA in this way; by inserting the gene for a

particular protein into a plasmid it is possible to induce bacterial cells to mass produce that protein in a process referred to as protein expression.

In the first step of molecular cloning, the gene is removed from the DNA surrounding it using restriction enzymes. The same enzymes are used to create complementary 'sticky ends', short strands of ssDNA, on the plasmid. The insert DNA and plasmid are mixed and the complementary sequences hold them together while the ligase acts to form a phosphodiester bond in the DNA backbone.

Three plasmids were used in the work described here; pET11a, pET15b and pET23a. Maps of these vectors and their restriction sites can be found in Appendix B.3. These plasmids were chosen to explore the most efficient system for RecA production by including a hexa-histidine tag at different ends of the protein. Histidine affinity tags are included in protein expression to aid with purification, the His-tag has an affinity for nickel that allows tagged protein to be 'pulled out' of a cell lysate. pET11a contains no affinity tag, pET15b contains an N terminal His-tag and pET23a contains a C terminal His-tag.

All three plasmids include the T7 promoter sequence for expression; this is a commonly-used DNA sequence that indicates to the cellular machinery that the DNA sequence following is to be transcribed into RNA, in this case for the purpose of protein synthesis.

The use of this promoter enables a high level of control over the expression of the desired protein; the T7 promoter requires T7 RNA polymerase for activity, it is the only polymerase that acts on this promoter. The expression of T7 RNA polymerase from the genome of the BL21 (DE3) cells is controlled by the *lac* operon, plasmids 11a and 15b contain the *lacI* repressor which acts with the *lac* operon in the cells to prevent protein expression without proper induction. IPTG is a structural lactose mimic that binds to the *lac* operon; unlike lactose it does not play a role in any of the metabolic processes of the cell. Induction with IPTG leads to the expression of high levels of the T7 RNA polymerase which bind to the T7 promoter and results in high levels of translated recombinant protein.

Control over the induction of expression is key for proteins which could be toxic to the cell, allowing cell proliferation before induction. Although RecA is an *E. coli* protein, its role in the SOS response means that large concentrations of this protein could have a toxic effect. The T7 terminator is present on all of the plasmids after the protein insertion site or

the His-tag. All three plasmids contain a number of unique restriction enzyme sites within the multicloning region to allow optimal insertion of the protein coding gene, as well as a gene coding for ampicillin resistance to allow selection of successfully transformed cells.

3.1.10 Cell lines

Plasmids containing the sequences of interest can then be transformed into cells for protein expression. The cell lines used for the work described here were One Shot MAX Efficiency DH5 α -T1R Competent Cells or BL21 Star (DE3)pLysS One Shot *Escherichia coli*. The cell types chosen were chemically competent. Competence is a term used in molecular biology to denote the ability to take up DNA from the extracellular environment. There are a number of ways to make cells chemically competent. The cells used in this work were stored in a calcium chloride solution; this aids the attachment of DNA to the membrane and is thought to induce pore formation in the membrane. When the cells are heated the pores in the membrane open and this facilitates plasmid entry.

One Shot MAX Efficiency DH5 α -T1R Competent Cells were used for increasing plasmid numbers for purification. They contain a number of genes that make them suitable for this role. The tonA genotype confers resistance to two common bacteriophages, T1 and T5. This is key in producing plasmid libraries, where phages could destroy the bacteria and hence the genetic stock. The endA1 gene results in cleaner DNA following purification steps due to a mutation in Endonuclease I, an enzyme which breaks up DNA forming structures other than duplexes. A reduction in undesired recombination is included through the recA1 gene, mutating the bacterial Recombinase A. In order to aid transformation of the cells with cloned DNA, which is unmethylated, the gene hsdR is included in this strain.

BL21 Star (DE3)pLysS cells were used for protein expression, they are chosen for their ability to produce high yields when the protein is non-toxic. They include the DE3 lysogen which encodes T7 RNA polymerase under the control of the lacUV5 promoter. This is an inducible expression system that requires the T7 promoter within the plasmid. The RNaseE gene gives high mRNA stability which increases translation of mRNA to protein. The pLysS plasmid contains the gene for T7 lysozyme which helps prevent leaky

expression, that is protein expression without the necessary induction. This cell type also contains the Cam^R plasmid, which gives resistance to the antibiotic chloramphenicol.

3.2 Protocols

3.2.1 DNA Structure Fabrication

All of the DNA sequences used in this work are shown in Appendix A. A list showing the provenance of the reagents can be found in Section 3.3.

100 base fabrication

The 100 bp structure was annealed by heating an oligomer from the 100 base series with the 100 base reverse complement to 95 °C in a thermocycler for 5 minutes and then slow cooling by 1 °C every 30 seconds to a range of temperatures between 80 °C and 60 °C. The annealing was carried out in 1X NEBuffer 4 (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM dithiothreitol, pH 7.9 at 25 °C) in a 50 μ l volume, the concentration of the two oligomers was 1 mM. Following 5 minutes at the annealing temperature the structures were immediately cooled to 4 °C. The efficiency of annealing was assessed by native PAGE.

200 base fabrication

PCR of 140 base template from Lambda DNA

The reagents for the PCR reaction to amplify the 140 base sequence from the λ genome are shown in Table 3.1. The reaction was carried out in NEB LongAmp *Taq* reaction buffer (60 mM Tris-SO₄, 20 mM (NH₄)₂SO₄, 2 mM MgSO₄, 3% glycerol, 0.06% IGEPAL CA-630, 0.05% TWEEN 20, pH 9 at 25 °C). The sequence of primer 1 and 2 for the 140 base sequence are given in Appendix A.

The PCR program used to synthesise the 140 base template DNA for the 200 base structure is shown in Table 3.2.

Following PCR the samples were purified using the Roche High Pure PCR Purification Kit. The manufacturer's protocol was followed.

Table 3.1: PCR reaction mixture for the amplification of the 140bp template from the genome of the λ bacteriophage.

Reagent	Concentration
Primer 1	800 nM
Primer 2	800 nM
dNTPs	8 mM
λ DNA template	317 pM
LongAmp <i>Taq</i> DNA polymerase	250 units
LongAmp <i>Taq</i> reaction buffer	1X
Total volume	50 μ l

Table 3.2: PCR cycle for the amplification of the 140 bp template from the genome of the λ bacteriophage.

Step	Temperature ($^{\circ}$ C)	Duration	Cycles
Heat	95	30 s	1
Melt	95	15 s	30
Annealing	52	15 s	
Extension	68	10 s	
Final Extension	68	5 minutes	1
Hold	4	As required	1

Samples were made up to a volume of 100 μl , if necessary, with 1X TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH \sim 8) and vortexed. 500 μl of Binding buffer was added and the solution vortexed again. The solution was applied to a column and centrifuged for 60 seconds at 13000 Xg. The flowthrough was discarded and 500 μl of Wash buffer was added. The column was centrifuged for 60 seconds at 13000 Xg and the flowthrough discarded. Then, 200 μl of Wash buffer was added to the column and it was centrifuged for 60 seconds at 13000 Xg. The flowthrough was discarded and the column was centrifuged again for 60 seconds at 13000 Xg. Following this step, the column was moved to a clean 1.5 ml tube and the 50 μl of the 1X TE was added. The column was incubated at room temperature for 5 minutes and then centrifuged for 60 seconds at 13000 Xg. Sample concentration was assessed using a Thermo Scientific NanoDrop 2000c spectrometer.

LCR of 200 base structure

For the purposes of the LCR reaction, *Taq* DNA ligase was used in a reaction mixture shown in Table 3.3. The reaction was carried out in *Taq* DNA ligase reaction buffer (20 mM Tris-HCl, 25 mM potassium acetate, 10 mM magnesium acetate, 1 mM NAD, 10 mM dithiothreitol, 0.1% Triton X-100, pH 7.6 at 25 °C).

The final iteration of the program is shown in Table 3.4. Optimisation of the protocol is discussed in Section 5.1.4.

Purification of 200 base structure

Following the ligase chain reaction the DNA species were separated using denaturing agarose gels.

To make the gels, agarose was melted in water and then cooled to 55 °C in a water bath. Alkaline gel buffer (500 mM sodium hydroxide, 10 mM EDTA, 10X) was added and the gel poured immediately. The gel was then run in 1X alkaline gel buffer. EDTA was added to a concentration of 10 mM to samples before adding 6X alkaline gel electrophoresis loading dye. Samples were loaded and the gels were run at 30 V for 30 minutes and then 80 V for 80 minutes. Following this, the gels were incubated for 45 minutes in gel

Table 3.3: LCR reaction mixture to create 200bp structures with ssDNA elements

Reagent	Quantity
Oligomer 1	1 μ M
Oligomer 2	1 μ M
Oligomer 3	1 μ M
Oligomer 4	1 μ M
<i>Taq</i> DNA ligase reaction buffer	1X
λ 140bp DNA scaffold	220 pM
<i>Taq</i> DNA ligase	100 units
Total volume	50 μ l

Table 3.4: LCR cycle to create 200bp structures with ssDNA elements

Step	Temperature ($^{\circ}$ C)	Duration	Cycles
Heat	98	30 s	1
Melt	98	10 s	
Annealing	72	30 s	30
Ligation	64	15 minutes	
Hold	4	As required	1

neutralising solution (1 M Tris-HCl, 1.5 M sodium chloride, pH 7.6 at 25 °C). The gels were rinsed and incubated for 30-60 minutes in TAE buffer and SYBR Gold (Invitrogen).

The excised gel slice was then suspended in 300 μ l of Binding buffer per 100 mg of gel. The slice was vortexed, and then heated to 56 °C for 10 minutes. Following the complete dissolution of the gel, 150 μ l of isopropanol per 100 mg of gel was added. The solution was thoroughly vortexed and then this solution was applied to a column of the Roche High Pure PCR Product Purification Kit. The column was centrifuged for 60 seconds at 13000 Xg and the column flow-through was discarded. After this step, the purification was continued according to the protocol for the Roche High Pure PCR Product Purification Kit, described in Section 3.2.1.

3.2.2 270 base structure

PCR of 270 bases from synthetic genes

The sequences were ordered from GeneArt as synthetic genes in the plasmid pMA-T, shown in Appendix A. A PCR reaction was undertaken using phosphorothioated primers to create 270 base dsDNA. Two separate reactions created the two strands necessary for the construct. The phosphorothioated bases were later removed through the action of a restriction enzyme, as in Tosch *et al.* [243].

The reagents for the PCR reaction to amplify the 270 base sequence from the synthetic gene is shown in Table 3.5. The reaction was carried out in NEB standard *Taq* buffer (10 mM Tris-HCl, 50 mM KCl, 1.5 mM MgCl₂, pH 8.3 at 25 °C).

The PCR program used to amplify the 270 base structure using phosphorothioated DNA primers is shown in Table 3.6.

The PCR products were purified and analysed by native PAGE as in Sections 3.2.1.

T7 Exonuclease Treatment

The PCR products from Section 3.2.2 were treated with T4 Exonuclease to create ssDNA strands. The reaction mixture shown in Table 3.7 was incubated at 37 °C for 2 hours. The

Table 3.5: PCR reaction mixture for the amplification of the 270 base sequence from the synthetic gene.

Reagent	Concentration
270 base sequence primer 1	800 nM
270 base sequence primer 2	800 nM
dNTPs	8 mM
DNA template (synthetic gene)	55 nM
<i>Taq</i> DNA polymerase	250 units
<i>Taq</i> reaction buffer	1X
Total volume	50 μ l

Table 3.6: PCR cycle for the amplification of the 270 base phosphorothioated products from the synthetic genes containing the template DNA sequence.

Step	Temperature (°C)	Duration	Cycles
Heat	95	30 s	1
Melt	95	15 s	30
Annealing	64	15 s	
Extension	68	17 s	
Final Extension	68	5 minutes	1
Hold	4	As required	1

reaction was carried out in NEBuffer 4 (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM dithiothreitol, pH 7.9 at 25 °C).

Table 3.7: T7 exonuclease treatment of phosphorothioated PCR products

Reagent	Concentration
DNA	40 - 100 nM
NEBuffer 4	1X
T7 Exonuclease	60 units
Total volume	50 μ l

Following the digestion step, a purification step was necessary to remove the mononucleotides that are a product of the reaction. It is also necessary to remove the T7 exonuclease as this enzyme cannot be heat denatured and would potentially digest the product when the phosphorothioates had been removed. The Zymo ssDNA/RNA Clean and Concentrator kit was used following the manufacturer's protocol.

To the sample add 2 volumes of the DNA/RNA Binding buffer and apply to the spin columns supplied. Centrifuge at 12,000 Xg for 60 seconds. The flow through is retained and 1 volume of ethanol is added to it. The flowthrough is then reapplied to the column for centrifugation at 12,000 Xg for 60 seconds. The flowthrough is then discarded and 400 μ l of DNA/RNA Prep buffer is applied to the column. Following centrifugation at 12,000 Xg for 60 seconds the flow through is once again discarded and 700 μ l of DNA/RNA Wash buffer is added to the column. The column is centrifuged at 12,000 Xg for 60 seconds, the flow through discarded, and a further 400 μ l of DNA/RNA Wash buffer is added to the column. When the column has been centrifuged at 12,000 Xg for 60 seconds, the flowthrough is discarded and the column is centrifuged again in the emptied collection tube for 60 seconds at 12,000 Xg. The column is moved to a clean tube and 20 μ l of water is added to elute the DNA from the column. The column is incubated for 60 seconds and then centrifuged at 10,000 Xg for 30 seconds. The concentration of the DNA can then be ascertained.

Double stranded restriction site

Following T7 exonuclease treatment the annealing oligomer for BmtI digest was added to anneal to a small section of the phosphorothioated ssDNA product. This created a dsDNA region that contained a recognition for the restriction enzyme BmtI-HF. The oligomer was added at a 2:1 ratio and the solution was heated to 95 °C for 10 minutes and then slow cooled at 1 °C/min to 4 °C.

Restriction

When the oligomer was annealed and the restriction site was available, the digestion was carried out in NEB CutSmart buffer (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 100 µg/ml BSA, pH 7.9 at 25 °C). The reaction constituents are shown in Table 3.8.

Table 3.8: BmtI-HF digestion of phosphorothioated ssDNA

Reagent	Concentration
DNA	40 - 100 nM
Cutsmart buffer	1X
BmtI-HF	20 units
Total volume	50 µl

Purification

Following the digestion step, a purification step was necessary to remove the short ssDNA and dsDNA strands that were a by-product of the reaction. The Zymo ssDNA/RNA Clean and Concentrator kit was used to separate the long ssDNA product from the shorter contaminants following the manufacturer's protocol, described above.

The concentration of samples was ascertained using a Nanodrop 2000 spectrophotometer and the DNA was analysed by native PAGE.

3.2.3 Three-way junction annealing

200 Base Structure

Three-way junctions were hybridised through heating DNA samples with the 200 base junction oligomers at a 2:1 ratio to 72 °C and then snap cooling them to 4 °C in 1X NEBuffer 4 (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM dithiothreitol, pH 7.9 at 25 °C). Experimentation established that there was no benefit to slow cooling and it significantly increased the duration of the hybridisation. The temperature and buffer composition required were identified in the experiments shown in Section 5.1.4.

270 Base Structure

A slow cooling protocol was used to anneal the two ssDNA samples and the 270 base junction oligomers for the creation of three-way junctions. The additional oligomers were added at a 2:1 ratio in a buffer containing 10 mM Tris acetate and 10 mM magnesium acetate. The samples were heated to 95 °C for 3 minutes and then cooled by 1 °C per 30 seconds until a temperature of 4 °C was reached.

3.2.4 Gel Electrophoresis

Native PAGE

Polyacrylamide gels were made to a final concentration of 15% acrylamide/bis-acrylamide (19:1) and 1X TBE (89 mM Tris-borate, 2 mM EDTA, pH 8.3), and included a 5% stacking gel. The volumes of the reactions are shown in Table 3.9. Polymerisation was catalysed by 10% APS (w/v) and TEMED, which were added into the reaction last before pouring between two glass plates. The resolving gel was cast first with a layer of isobutanol poured on top to level the interface between the two gels. When the resolving gel had set the isobutanol was thoroughly washed off and the stacking gel was added, with a comb to produce wells for the addition of DNA.

Table 3.9: Native acrylamide gel composition

Reagent	Quantity
Resolving gel	
30% acrylamide/bis-acrylamide (19:1)	7.5 ml
5X TBE	3 ml
10% APS	150 μ l
TEMED	10 μ l
H ₂ O	To 15 ml
Stacking gel	
30% acrylamide/bis-acrylamide (19:1)	1.7 ml
5X TBE	2 ml
10% APS	150 μ l
TEMED	10 μ l
H ₂ O	To 10 ml

Between 10 and 20 ng of DNA mixed with 1X Blue/Orange Loading Dye were loaded into wells. Gels were run in 1X TBE at 80 V for between 300 and 360 minutes. Following electrophoresis, gels were stained with 0.0001% SYBR Gold in 1X TBE for 20 minutes. Gels were then rinsed and imaged under UV using the Kodak EDAS 290 system and Kodak1D software, version 3.6. Densitometry analysis was carried out using ImageJ.

Denaturing PAGE

To 7.5 ml of 30% acrylamide/bisacrylamide was added 6.3g of urea and 1.5 ml of 10X TBE (890 mM Tris-borate, 20 mM EDTA, pH 8.3). The solution was heated in a water bath at 55 °C and mixed until the urea had fully dissolved. Following this the mixture was allowed to return to room temperature. 150 μ l of 10% APS and 10 μ l of TEMED were added and the gel was poured between glass plates and a lane comb was added. Once set, gels were stored at 4 °C in 1X TBE.

Denaturing gels were loaded into the gel tank with 1X TBE and pre-run at 200V for at

least 2 hours before the samples were loaded. Samples were first boiled in Gel Loading Buffer II (95% formamide, 18 mM EDTA, 0.025% SDS, 0.025% xylene cyanol, 0.025% bromophenol blue) at 100 °C for 10 minutes and then loaded onto the pre-run gel. Gels were run at 80 V for between 120 and 600 minutes and then stained with SYBR Gold in 1X TBE for 20 minutes. Gels were rinsed and imaged under UV.

Native Agarose Gel Electrophoresis

Agarose gels for DNA imaging were made with 2% agarose and run in 1X TAE (0.4 M Tris acetate, 0.01 M EDTA, pH 8.3 at 25 °C). Powdered agarose was melted in a solution of 1X TAE. Agarose gels contained 5 μ l RedSafe nucleic acid staining solution and were imaged under UV.

Denaturing Agarose Gel Electrophoresis

Ligated 200 base DNA was purified from denaturing agarose gels, made according to the protocol in Sambrook and Russel [242]. Briefly, agarose was melted in water and then cooled to 55 °C in a water bath. 10X Alkaline gel buffer (500 mM sodium hydroxide, 10 mM EDTA) was added to 1X concentration and the gel poured immediately. The gel was then run in 1X alkaline gel buffer.

EDTA was added to a concentration of 10 mM to samples before adding 6X Novex TBE-Urea loading dye (Tris base, boric acid, EDTA, acrylamide, bisacrylamide, TEMED, and APS). Samples were loaded and the gels were run at 30 V for 30 minutes and then 80 V for 80 minutes. Following this, the gels were incubated for 45 minutes in gel neutralising solution (1 M Tris-HCl, 1.5 M sodium chloride, pH 7.6 at 25 °C). The gels were rinsed and incubated for 30-60 minutes in TAE buffer (40 mM Tris-acetate, 1 mM EDTA, pH 8.3) and SYBR Gold (Invitrogen). The gels were then placed on a UV light source and the desired DNA band was excised using a scalpel.

The excised gel slice was then suspended in 300 μ l of Binding buffer (Roche High Pure PCR Product Purification Kit) per 100 mg of gel. The slice was vortexed, and then heated to 56 °C for 10 minutes. Following the complete dissolution of the gel, 150 μ l

of isopropanol per 100 mg of gel was added. The solution was thoroughly vortexed and then this solution was applied to a column of the Roche High Pure PCR Product Purification Kit. The column was centrifuged for 60 seconds at 13000 Xg and the column flow-through was discarded. After this step, the purification was continued as the protocol for the Roche High Pure PCR Product Purification Kit, described previously.

SDS polyacrylamide gel electrophoresis

The reaction mixture used is shown in Table 3.10.

The TEMED and APS were added last and then the mixture was poured between glass plates. The resolving gel was cast first with a layer of isobutanol poured on top to level the interface between the two gels. When the resolving gel had set the isobutanol was thoroughly washed off and the stacking gel was added, with a comb to produce wells for the addition of DNA. Gels were stored at 4 °C in 1X TBE.

Table 3.10: SDS acrylamide gel

Reagent	Quantity
Resolving gel	
30% acrylamide/bis-acrylamide (19:1)	5 ml
1.5 M Tris (pH 8.8)	3.8 ml
10% SDS	150 μ l
10% APS	150 μ l
TEMED	6 μ l
H ₂ O	5.9 ml
Stacking gel	
30% acrylamide/bis-acrylamide (19:1)	1.3 ml
1.5 M Tris (pH 8.8)	1 ml
10% SDS	80 μ l
10% APS	80 μ l
TEMED	8 μ l
H ₂ O	5.5 ml

Gels were run at 120 V for between 60 and 80 minutes. Following electrophoresis, gels were stained overnight in Coomassie G-250. Stained gels were placed in SDS gel destain (20% methanol (v/v), 10% acetic acid (v/v), H₂O to volume). Gels were imaged using white light.

Gel Analysis

Densitometry analysis of the gels was carried out using ImageJ software.

3.2.5 Diagnostic Digestion

The diagnostic digestion of DNA structures was carried out at 37 °C for 1 hour. In a total volume of 50 μ l, 5 pmol of the DNA was digested in 1X Buffer Tango (33 mM Tris-acetate, 10 mM magnesium acetate, 66 mM potassium acetate, 0.1 mg/mL bovine serum albumin, pH 7.9 at 37 °C) using 10 units of XapI.

3.2.6 AFM Visualisation

Imaging of the DNA structures was carried out with AFM. To prepare the sample area a cylinder of mica was glued to a magnetic stub. The mica substrate was freshly cleaved and incubated with 50 μ l 10 mM nickel chloride in 1X TE (10 mM Tris-HCl, 1 mM EDTA, pH ~8) for 2 minutes in a high humidity chamber. The 50 μ l sample containing the DNA in 1X TE was pipetted onto the surface, and allowed to incubate for 20 minutes in a high humidity chamber. The mica was transferred to the AFM and a further 100 μ l of 1X TE was added to the mica for imaging in liquid.

Samples were imaged under the AFM in liquid in peak force tapping mode using Fastscan D tips from Bruker. Scan rates are shown in Table 3.11. For all images the amplitude setpoint was 250 mV and the drive amplitude was 57.68 mV. Image analysis was undertaken with the Bruker software Nanoscope Analysis version 9.1.

Table 3.11: Scan rates for capture of AFM images of 100 and 270 base structures

100 base Structure	Scan rate
Fully base paired	2.2 Hz
20 base mismatch	2.93 Hz
270 base Structure	Scan rate
No annealed oligomers	1.46 Hz
5' annealing oligomer	3.2 Hz
3' annealing oligomer	2.34 Hz
Both annealing oligomers	2.2 Hz

3.2.7 RecA Patterning

DNA structures were patterned using the protocol shown in Table 3.12. Denatured RecA was used as a control for the digestion of bare DNA in the same conditions as those that had been patterned. RecA was held at 65 °C for 20 minutes in order to denature the enzyme, according to the NEB protocol. The heat treated protein was then used in the patterning experiment according to the protocol in Table 3.12.

Digestion was carried out in 1X Buffer Tango (33 mM Tris-acetate, 10 mM magnesium acetate, 66 mM potassium acetate, 0.1 mg/mL bovine serum albumin, pH 7.9 at 37 °C).

Table 3.12: dsDNA patterning protocol

Reagent	Final Concentration
Step 1: Filament formation	
Magnesium acetate	2 mM
Tris acetate pH	30 mM
ssDNA	10:1 to 40:1 ratio with dsDNA
RecA	Total required to completely cover the ssDNA
ATP γ S	1 mM
H ₂ O	To 20 μ l
Incubate at 37 °C for 15 minutes	
Step 2: Removal of excess RecA	
Poly-T oligomer (30 bases)	4.8 μ M
Incubate at 37 °C for 15 minutes	
Step 3: Patterning	
dsDNA	3.3% of ssDNA concentration
Tris acetate	30 mM
Magnesium acetate	10 mM
H ₂ O	To 30 μ l
Incubate at 37 °C for 90 minutes	
Step 4: Diagnostic digest	
XapI	50 units
Buffer Tango	1X
H ₂ O	To 40 μ l
Incubate at 37 °C for 5 minutes	
Step 5: Protein digestion	
Proteinase K	\geq 4 units
Incubate at 37 °C for 30 minutes	

3.2.8 Protein production

An experiment was carried out to produce RecA through recombinant gene expression. The primary aim was to develop a protocol that could be used to produce RecA mutants, for example a protein lacking the C terminus which could have a higher affinity for ss- and dsDNA [165]. It was also thought that protein production would give us greater control over the buffer conditions of RecA.

Cloning

The sequence of the wild-type protein coding region cloned into the plasmids is shown in Figure 3.5 along with the translated protein primary structure.

```

1 M A I D E N K Q K A L A A A L G Q I E K
1 ATGGCTATCGACGAAAAACAAACAGAAAGCGTTGGCGGCAGCACTGGGCCAGATTGAGAAA
21 Q F G K G S I M R L G E D R S M D V E T
61 CAATTTGGTAAAGGCTCCATCATGCGCCTGGGTGAAGACCGTTCATGGATGTGAAACC
41 I S T G S L S L D I A L G A G G L P M G
121 ATCTCTACCGGTTTCGCTTTCACCTGGATATCGCGCTTGGGGCAGGTGGTCTGCCGATGGGC
61 R I V E I Y G P E S S G K T T L T L Q V
181 CGTATCGTCGAAATCTACGGACCGGAATCTTCCGGTAAAACCACGCTGACGCTGCAGGTG
81 I A A A Q R E G K T C A F I D A E H A L
241 ATCGCCGCGAGCGCAGCGTGAAGGTAAAACCTGTGCGTTTATCGATGCTGAACACGCGCTG
101 D P I Y A R K L G V D I D N L L C S Q P
301 GACCCAATCTACGCACGTAACCTGGGCGTCGATATCGACAACCTGCTGTGCTCCCAGCCG
121 D T G E Q A L E I C D A L A R S G A V D
361 GACACCGCGAGCAGGCACTGGAAATCTGTGACGCCCTGGCGCGTTCTGGCGCAGTAGAC
141 V I V V D S V A A L T P K A E I E G E I
421 GTTATCGTCGTTGACTCCGTGGCGGCACTGACGCCGAAAGCGGAAATCGAAGGCGAAATC
161 G D S H M G L A A R M M S Q A M R K L A
481 GCGACTCTCACATGGGCCTTGGCGCACGTATGATGAGCCAGGCGATGCGTAAGCTGGCG
181 G N L K Q S N T L L I F I N Q I R M K I
541 GGTAACCTGAAGCAGTCCAACACGCTGCTGATCTTCATCAACAGATCCGATGAAAATT
201 G V M F G N P E T T T G G N A L K F Y A
601 GGTGTGATGTTCCGGTAACCCGAAACCACTACCGTGGTAACGCGCTGAAATCTACGCC
221 S V R L D I R R I G A V K E G E N V V G
661 TCTGTTCTCTCGACATCCGTCTGATCGGGCGCGTGAAAGAGGGCGAAAACGTGGTGGGT
241 S E T R V K V V K N K I A A P F K Q A E
721 AGCGAAACCCCGGTGAAAAGTGGTGAAGAACAATACTCGCTGCGCCGTTTAAACAGGCTGAA
261 F Q I L Y G E G I N F Y G E L V D L G V
781 TTCCAGATCCTCTACGGCGAAGGTATCAACTTCTACGGCGAACTGGTTGACCTGGGCGTA
281 K E K L I E K A G A W Y S Y K G E K I G
841 AAAGAGAAGCTGATCGAGAAAGCAGGCGCGTGGTACAGCTACAAAGGTGAGAAGATCGGT
301 Q G K A N A T A W L K D N P E T A K E I
901 CAGGTAAGCGAATGCGACTGCCTGGCTGAAAGATAACCCGAAACCGCGAAAGAGATC
321 E K K V R E L L L S N P N S T P D F S V
961 GAGAAGAAAGTACGTGAGTTGCTGCTGAGCAACCCGAACTCAACGCGGATTTCTCTGTA
341 D D S E G V A E T N E D F *
1021 GATGATAGCGAAGGCGTAGCAGAACTAACGAAGATTTTAA

```

Figure 3.5: The sequence of the RecA gene and its translation into a protein primary structure

The plasmids and the inserts were stored digested by a previous lab user (Rajan Sharma). Plasmid pET23a and its inserts were digested with NdeI an XhoI. Plasmid 11a and 15b

and their inserts were digested with NdeI and BamHI. Maps of these vectors and their restriction sites can be found in Appendix B.3.

Plasmids were digested in a 30 μ l reaction volume using 3 μ l of pET vector at a concentration of 1 μ g/ μ l. The restriction digest used 40 units of each restriction enzyme and contained 3 μ l of acetylated BSA at a concentration of 1 mg/ml. The restriction digest was carried out in NEBuffer 3 (100 mM sodium chloride, 50 mM Tris-HCl, 10 mM magnesium chloride, 1 mM dithiothreitol, pH 7.9 at 25 °C) at 37 °C for 4 hours.

The inserts were created using PCR, adding appropriate restriction sites at both ends of the insert. Inserts were made with and without a C-terminal cytosine codon, which was added to the genes through a modified PCR primer. The PCR products were digested with 40 units of the appropriate restriction enzymes in NEBuffer 3 at 37 °C. The digested DNA was then purified using the QIAquick PCR purification kit from Qiagen.

The ligation took place in 1X T4 DNA ligase reaction buffer (50 mM Tris-HCl, 10 mM MgCl₂, 1 mM ATP, 10 mM DTT, pH 7.5 at 25 °C) with 90 ng insert, 40 ng plasmid and 1 μ l T4 ligase. The reaction volumes were stored at 16 °C overnight. The following morning the plasmids were transformed into DH5 α *Escherichia coli* cells.

Transformation

DH5 α or BL21 Star (DE3) *Escherichia coli* cells were defrosted in ice water for 10 minutes and then split into 25 μ l aliquots. 5 μ l of the ligation solution was then added, containing 100 ng of DNA. The cells were kept on ice for 30 minutes. They were heat treated at 42 °C for 30 seconds to activate the competency before being placed back on ice for 5 minutes. Following this step, 450 μ l of SOC medium (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, 20 mM glucose) was added and the cells were placed in a shaking incubator at 250 rpm for 60 minutes at 37 °C. The LB agar plates (1% tryptone, 0.5% yeast extract, 1% NaCl, Agar) containing 0.01 % ampicillin (w/v) were warmed to 37 °C prior to spreading 50 μ l of sample onto the plate with a sterile spreader. The plates were then incubated at 37 °C overnight.

Plasmid Amplification

Colonies of DH5 α *E. coli* were picked from the plates with sterile colony pickers and grown on in 5 ml LB (1% tryptone, 0.5% yeast extract, 1% NaCl) in a 50 ml tube at 250 rpm for 60 minutes at 37 °C . Alternatively, a small sample was scraped from the top of the glycerol stocks to start 5 ml colonies. The 5 ml samples were pelleted at 8000 rpm at 25 °C for 3 minutes. The plasmids were collected by miniprep purification using the QIAprep Miniprep kit.

The 5 ml samples were pelleted at 8000 rpm at 25 °C for 3 minutes. The pellet was resuspended in 250 μ l buffer P1. The cells were then lysed with 250 μ l P2 for up to 5 minutes. Following this, 350 μ l buffer N3 was added and the solution was mixed. The cell lysate was centrifuged for 10 minutes at 13000 rpm. The supernatant was then applied to a purification column.

The column was centrifuged at 60 seconds at 13000 rpm. The flowthrough was discarded and the column was washed with 500 μ l buffer PB by centrifuging at 13000 rpm for 60 seconds. The flow-through was again discarded and the column washed with 750 μ l buffer PE by centrifuging at 13000 rpm for 60 seconds. The flow-through was discarded again and the column was centrifuged at 13000 rpm for 60 seconds in a clean collection tube to remove excess buffer. The column was moved to a new collection tube and 60 μ l of buffer EB was added and incubated for 1 minute. The DNA was then eluted by centrifuging at 13000 rpm for 60 seconds.

To check that the sequences were correct, the plasmids were sequenced by Eurofins MWG Operon.

To make glycerol stocks, 60% glycerol was first sterilised by autoclaving. The cells were then mixed with the glycerol stock in a 1:1 ratio and flash frozen in liquid nitrogen. Stocks were then stored at -80 °C.

Protein Induction

BL21 Star (DE3) *E. coli* cells were taken from glycerol stocks to grow overnight in 5 ml starter cultures of LB (1% tryptone, 0.5% yeast extract, 1% NaCl), containing 0.01%

ampicillin (w/v) and 0.0034% chloramphenicol (w/v).

In initial tests the starter culture was induced to produce protein through the addition of 0.5 mM IPTG and then grown in LB broth for 5 hours at 250 rpm at 37 °C. The cells were spun down at 8000 rpm at 25 °C for 30 minutes. The supernatant was discarded and the cell pellet was frozen at -80 °C.

The cells were defrosted and resuspended in cell resuspension buffer. The mixture was sonicated and then centrifuged at 12500 rpm for 45 minutes. The supernatant and the pellet were separated and tested by gel electrophoresis for protein expression.

For larger scale production 600 ml cultures were grown as two 300 ml cultures in LB broth in 2 litre flasks. The samples were agitated at 250 rpm at 37 °C to an optical density (OD) of 0.4-0.7. Protein production was then induced with 0.4 mM IPTG. Following 5 hours at 250 rpm at 37 °C the cultures were centrifuged at 13000 rpm for 10 minutes and then cell pellet was frozen at -80 °C.

Harvesting

BL21 Star (DE3) *E. coli* cells were resuspended in 5 ml Tris-HCl (pH 7.2) and sonicated on ice. The sample was then centrifuged at 11000 rpm for 40 minutes at 4 °C. The supernatant was collected.

Steps added in subsequent iterations include resuspension of the cell pellet in 4 ml Tris-HCl (pH 7.2) following pelleting at 4000 rpm for 20 minutes. The supernatant was filtered through a 0.2 µm membrane before FPLC.

Fast protein liquid chromatography (FPLC)

The work shown here used HisTrap HP columns from GE Healthcare. These are nickel-sepharose columns, for which the His-tag of the plasmids pET15b and pET23a have an affinity. Both the columns and the system were stored in 20% ethanol. In preparation for chromatography, the columns and system were first washed with water, then the protein elution buffer (25 mM Tris buffer pH 7.4, 100 mM NaCl, 400 mM imidazole) and finally with the column buffer (25 mM Tris buffer pH 7.4, 100 mM NaCl). Following thorough

washing, the protein-containing supernatant was introduced to the column and eluted at a low rate of flow. A gradient of protein elution buffer was run through the column; over 50 ml (10 column volumes) the gradient was increased from 0-400 mM imidazole. Fractions of 2.5 ml were collected and the fluorescence of the eluate was recorded at 260 and 280 nm wavelengths, corresponding to maximum emission peaks for DNA and protein. All chromatography took place at 4 °C.

Steps added in subsequent iterations include 20 mM imidazole in the supernatant and in the column buffer to account for non-specific binding of *E. coli* proteins. The protein elution buffer gradient was increased to 100 ml (20 column volumes). A wash step with 10 ml of 10 mM ATP was added to the protocol in an attempt to remove DNA contamination of the protein. DTT was added into the protein elution buffer to prevent aggregation.

RecA activity assay

To ascertain its activity, the formation of a stable triple-stranded complex with 3.5 kbp DNA was attempted. In a 20 μ l reaction volume, 0.088 pmol of DNA was reacted with 102.7 pmol of RecA in 30 mM Tris, 20 mM magnesium acetate and 0.2 μ M ATP γ S. Following incubation at 37 °C for 1 hour samples were run on agarose gels.

3.3 Reagents

Table 3.13: Reagents from New England Biolabs

Product Number	Reagent	Concentration
M0355S	RecA	2 mg/ml
M0273S	<i>Taq</i> DNA Polymerase	5,000 units/ml
M0323S	LongAmp <i>Taq</i> DNA Polymerase	2,500 units/ml
M0263S	T7 Exonuclease	10,000 units/ml
M0208S	<i>Taq</i> DNA Ligase	40,000 units/ml
M0202S	T4 DNA Ligase	400,000 units/ml
B7204S	Cutsmart buffer	10X
B7003S	NEBuffer 3	10X
B7004S	NEBuffer 4	10X
B0208S	<i>Taq</i> DNA Ligase Reaction Buffer	10X
B0202S	T4 DNA Ligase Reaction Buffer	10X
B9014S	Standard <i>Taq</i> Reaction Buffer	10X
B0323S	LongAmp <i>Taq</i> Reaction Buffer Pack	10X
N3231S	100 bp DNA Ladder	500 μ g/ml
N3011S	Lambda DNA	500 μ g/ml

Table 3.14: Restriction enzymes

Supplier	Number	Enzyme	Concentration	Recognition sequence
New England Biolabs	R3658S	BmtI-HF	20 units/ μ l	5' ...GCTAG [^] C... 3'
New England Biolabs	R0616S	Hpy166II	10 units/ μ l	5' ...GTN [^] NAC...3'
New England Biolabs	R0635S	BseYI	5 units/ μ l	5' ...C [^] CCAGC... 3'
Thermo Scientific	ER1381	XapI	10 units/ μ l	5' ...R [^] AATTY... 3'

Table 3.15: Reagents from Sigma-Aldrich

Product Number	Reagent	Concentration
P4850	Proteinase K	>800 units/ml
A3678	Ammonium persulphate	Powder
N6136	Nickel chloride	Powder
93337	Tris(hydroxymethyl)aminomethane acetate (TAc)	Powder
63052	Magnesium acetate (1M)	1M
T9285	Tris(hydroxymethyl)aminomethane -ethylenediaminetetraacetic acid (TE) buffer	100X
T9281	N,N,N,N-tetramethylethylenediamine (TEMED)	Liquid
A3574	30% acrylamide/bis-acrylamide	Liquid
G7893	Glycerol	Liquid
A0166	Ampicillin sodium salt	Powder
857440	Chloramphenicol	Powder
A26209	Adenosine 5-triphosphate disodium salt hydrate (ATP)	Powder
A1388	Adenosine 5-(3-thiotriphosphate) tetralithium salt (ATP γ S)	Powder
I6758	Isopropyl β -D-1-thiogalactopyranoside (IPTG)	Powder
I0125	Imidazole	Powder
L3022	Lennox L broth (LB)	Powder
L2897	LB agar	Powder
E5134	Ethylenediaminetetraacetic acid disodium salt dihydrate (EDTA)	Powder
U1250	Urea	Pellets
T3913	5X Tris-borate-EDTA (TBE) buffer	Powder
T9650	Tris-acetate-EDTA (TAE) Solution	10X
DNTP10	Deoxynucleotide set (dATP, dCTP, dGTP and dTTP)	10 mM

Table 3.16: Reagents from Sigma-Aldrich

Product Number	Reagent	Concentration
16500	Agarose	Powder
S11494	SYBR Gold Nucleic Acid Gel Stain	10,000 X
10821-015	10 bp DNA Ladder	1 $\mu\text{g}/\mu\text{l}$
LC6876	Novex TBE-Urea loading dye	2 X
C6020-03	BL21 Star (DE3)pLysS One Shot	50 μl
12297-016	Chemically Competent <i>E. coli</i> One Shot MAX Efficiency DH5 α -T1 ^R Competent <i>E. coli</i>	50 μl

Table 3.17: Additional reagents

Supplier	Product Number	Reagent	Concentration
Promega	20880.320	Blue/Orange loading dye	6X
VWR		Isopropanol	Liquid
ChemBio	A3147	RedSafe DNA dye	Liquid
PanReac AppliChem		Loading buffer	10X
Fisher Scientific		DNA III (alkaline)	
		Colloidal premixed	Liquid
		Coomassie G-250	

Oligomers were purchased from Integrated DNA Technologies and stored in 1x TE buffer.

Chapter 4

RecA patterning of 100 base structures containing base mismatches

RecA forms a nucleoprotein filament with a single strand of DNA. When a dsDNA strand is introduced a homology search takes place and RecA forms a triple-stranded region at the site of DNA homology. RecA patterning causes underwinding of the homologous DNA region and overwinding of proximal DNA regions [146]. As mentioned in Section 2.2.1 helical strain can be an important factor in DNA structures. The design of origami structures aims to reduce helical tension [50] but a certain degree of strain remains [247]. As tension plays an important role in the formation of DNA structures, for RecA to play a role in DNA nanotechnology it is important to understand the way that any deviation from linear dsDNA could affect patterning.

This chapter will present the results of experimentation with a series of 100 base constructs. These were the simplest DNA structures produced in this body of work and yielded results that cast light on how the complexity of DNA structures might affect RecA patterning. The structures were produced through annealing synthetic strands of DNA with partially complementary sequences. Regions of up to 20 bases were designed not to base pair, producing ssDNA regions within the strand. The experiments using the 100 base constructs were undertaken to explore the effect of mismatches in the DNA strand on RecA patterning efficiency.

The results in this chapter show that patterning efficiencies of higher than 95% are

possible using RecA. Furthermore, it is demonstrated that RecA can pattern structures with more complex topologies than previously shown. These results demonstrate a significant step forward in the understanding of RecA patterning in the context of complex DNA structures.

4.1 Synthesis of DNA scaffolds

4.1.1 Design

A 100 bp structure was designed to contain a non-base paired section of up to 20 bases in the central section of the two strands as the result of a region of non-complementary sequence. Synthetic oligomers were routinely and inexpensively made up to approximately 100 bases in length; this limitation defined the maximum length of the oligomers used in the fabrication of this construct.

Figure 4.1 shows a diagrammatic representation of the annealing of ssDNA oligomers to form a sequence in which the base pairing is interrupted at the centre by base pair mismatches. The DNA scaffolds were designed to incorporate an XapI restriction site and proximal sequences from the λ bacteriophage genome. The recognition sequence of XapI is R[^]AATTY where R is A or G and Y is C or T maintaining a palindromic sequence. The 30 base ssDNA oligomer used for patterning, homologous to bases 61-90 of 100, is shown in red. The XapI recognition site lies at bases 68-73, within the homologous region for the guide strand, shown as a blue box in Figure 4.1.

The sequence of the scaffold was chosen to minimise the occurrence of hairpin loops and other secondary structural elements that might interfere with the synthesis process. This was achieved using mfold, an online resource provided by the RNA Institute of the University at Albany. A 20 base randomised sequence was inserted into the Lambda sequence and the strand was analysed using mfold. The potential secondary structures of the sequence were checked to assess whether the base pairing regions of the sequence would form correctly. Randomised sequences were tested in this way until a sequence was found that would form the correct structure.

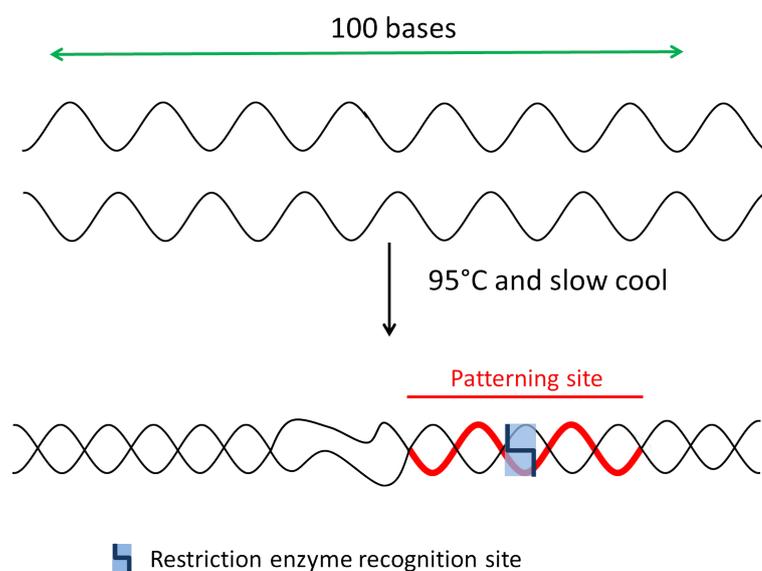


Figure 4.1: ssDNA oligomers of 100 bases are annealed to produce dsDNA sequences. Base pairing is not consistent throughout the strand; a section of the sequence has been designed to prevent interaction and introduce a non-hybridised region. A recognition site for the restriction enzyme XapI is present in the double stranded section of the sequence, represented here by a blue box. The patterning region is represented by the red part of the helix.

To make the shorter base pair mismatches the 20 base sequence was truncated and replaced with the original Lambda sequence, except where optimisation was necessary to facilitate correct base pairing. The DNA sequences used can be found in Appendix A, a diagrammatic representation of these structures is shown in Figure 4.2.

The patterning site and XapI recognition site were placed at the 3' end of the upper strand. This allowed the central region to contain the base pair mismatches, increasing the stability of the structure through longer regions of fully base paired DNA. No template was required for this structure to form, the ssDNA strands were heated to melt and cooled to anneal. Gel analysis indicated that annealing had taken place with high efficiency and no further purification was required (See Section 4.1.2). The sequences are listed in Appendix A.1

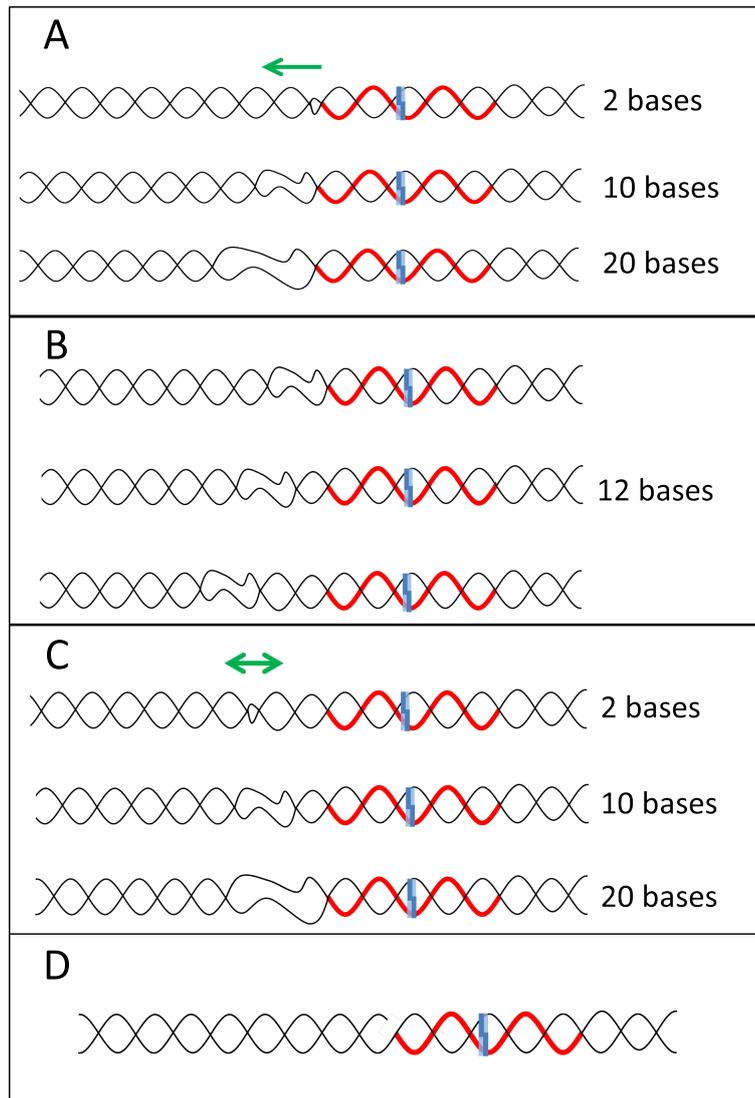


Figure 4.2: A diagrammatic representation of the 100 base structures used in this work **A:** A representation of the proximal structures, in which the mismatched region increases in size but remains at a constant distance from the patterning site. **B:** The sequence in which a 12 base mismatch was transposed in 3 base increments away from the patterning site. **C:** The central sequence in which both the distance from the patterning site and the size of the mismatch are altered. **D:** The nicked structure, in which two ssDNA strands were annealed to the lower strand creating an effective break in the phosphate backbone.

Proximal mismatches

The first set of constructs tested were sequences in which the mismatch was proximal to the patterning site, discussed in Section 4.2. A diagrammatic representation is shown in Figure 4.2 (A). This series was designed to explore the effect of an increasing number of base mismatches on patterning efficiency. The single base mismatch was at position 60, extension of the mismatch took place in the 3' to 5' direction so that the 20 base mismatch ran from base 41 to base 60. The shorter mismatches were all sequences from within the 20 base mismatch, except where this would have caused multiple stable secondary structures.

12 base mismatches

The 12 base mismatch from the first series was then moved further from the patterning site to explore the effect of proximity to the mismatch on the patterning efficiency of RecA. A diagrammatic representation is shown in Figure 4.2 (B). The 12 base mismatch was chosen as it was a convenient size for translocation on the structure and lay just within the saturation region for patterning efficiency (demonstrated in Section 4.2). The proximal 12 base mismatch began at base 49 and finished at base 60. The 'central' 12 base mismatch began at base 45 and finished at base 56 and the 'far' 12 base mismatch began at base 42 and finished at base 53. The results of these experiments showed that proximity to the patterning site was an important factor in the effect of base mismatches on patterning efficiency.

Central mismatches

Another series of sequences was designed to further explore the relationship between number of mismatches and proximity to the patterning site. The design of these sequences, hereafter referred to as the 'central' sequences, is discussed in Section 4.4; unlike the proximal sequences, the increase in size of the unpaired region is in both the 5' and 3' directions. A diagrammatic representation is shown in Figure 4.2 (C). The first base mismatch is at the centre of the sequence, at base 50. The 20 base mismatch starts at base 41 and finishes at base 60, the same as in the proximal designs.

Nicked phosphate backbone

The nicked structure was made up of three synthetic DNA strands annealed together. The 100 base reverse complement was kept the same, two strands replaced the single top strand to produce a structure with a break in the phosphate backbone. A diagrammatic representation is shown in Figure 4.2 (D). The nick was placed between bases 60 and 61 (40 and 60 base oligomers) and between bases 55 and 56 (45 and 55 base oligomers). A third structure was made with the 45 and 50 base oligomers in which a 5 base region of ssDNA was present in the reverse complement strand.

A series of experiments were done at different ratios of nucleoprotein complex to dsDNA; the proximal structures were tested at a 30:1 ratio. The ratio was limited by the volume of *RecA* that could be added and the concentration of the dsDNA necessary for PAGE. A 30:1 ratio has been demonstrated to show high patterning efficiencies while avoiding any non-specific binding [16].

4.1.2 Annealing DNA Sequences

All oligomers were received lyophilised and resuspended in 1X TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH ~8) to a concentration of 100 μ M for storage at -20 °C. Oligomers were then diluted to a working stock of 10 μ M in 1X TE buffer and run on a gel to confirm their size, shown in Figure 4.3 (A). It is important to identify bands that represent unannealed oligomer and to confirm that the structure runs as a single band.

In Figure 4.3 (A) it can be seen that all of the ssDNA strands migrate at around 150 bp. The mobility of ssDNA strands in gel electrophoresis is altered by the presence of secondary structural elements such as hairpin loops and as a result these strands appear not to have the correct length of ~ 100 bases. This effect is due to increased interaction with the gel matrix which slow the movement of the strands. A denaturing gel would have confirmed the correct size of these oligomers.

In Figure 4.3 (B) the products of annealing ssDNA oligomers containing mismatches of 0, 2, 6, 8, 10, 12, 15 and 20 bases can be seen, analysed by PAGE. The presence of a single band indicates that annealing was successful. Increasing the length of the

mismatched sequence changes the mobility of the DNA through the gel. The change in mobility indicates that there is a significant structural difference between the different constructs; single nucleotide polymorphisms in ssDNA have been shown to result in distinct mobility in polyacrylamide gels, possibly due to changes in secondary structure [248]. A diagnostic restriction digest was carried out as described in Section 3.2.5 to confirm they were the products desired. The results are shown in Figure 4.3 (C).

Only annealed DNA was digested, the mobility of the ssDNA in the gel continues to appear at ~ 150 bases when treated with XapI; this is evidence that dsDNA has been formed. The different DNA topologies interact with the gel matrix to slow the movement of the DNA through the gel, making the strands appear longer than they are. The cleavage point on the upper strand for this structure is between bases 68 and 69; when the fully base paired and 6 base mismatch sequences are digested a band appears at ~ 70 bases, confirming that the restriction site is present.

The 20 base mismatch shows a different length when it digests, appearing at ~ 600 bases. Following digestion of this construct there is only a short section of base pairing DNA on the 5' side of the digestion site following the non-base pairing region; it seems likely that the 9 bases left at the 3' end of the construct are not forming a region of dsDNA. There is evidence of increased DNA breathing in the presence of a short overhang on one strand [249], which would be present following the digestion of the construct with XapI. It seems likely that, in the absence of the 9 base pairs, the structure is forming a three-way junction and this is significantly retarding the movement of the DNA through the gel matrix. The correct annealing of the constructs was further confirmed through AFM, shown in Figure 4.4.

Image capture details are outlined in Section 3.2.6. The sequence containing the 20 base mismatch appears to be shorter than the fully base-paired DNA, indicating that the ssDNA region does not have the same rigidity as the dsDNA regions. The ssDNA can bend and form secondary structural elements. Many of the particles for the 20 base mismatched sequence show bending because of the ssDNA region, also due to the non-rigid nature of the ssDNA.

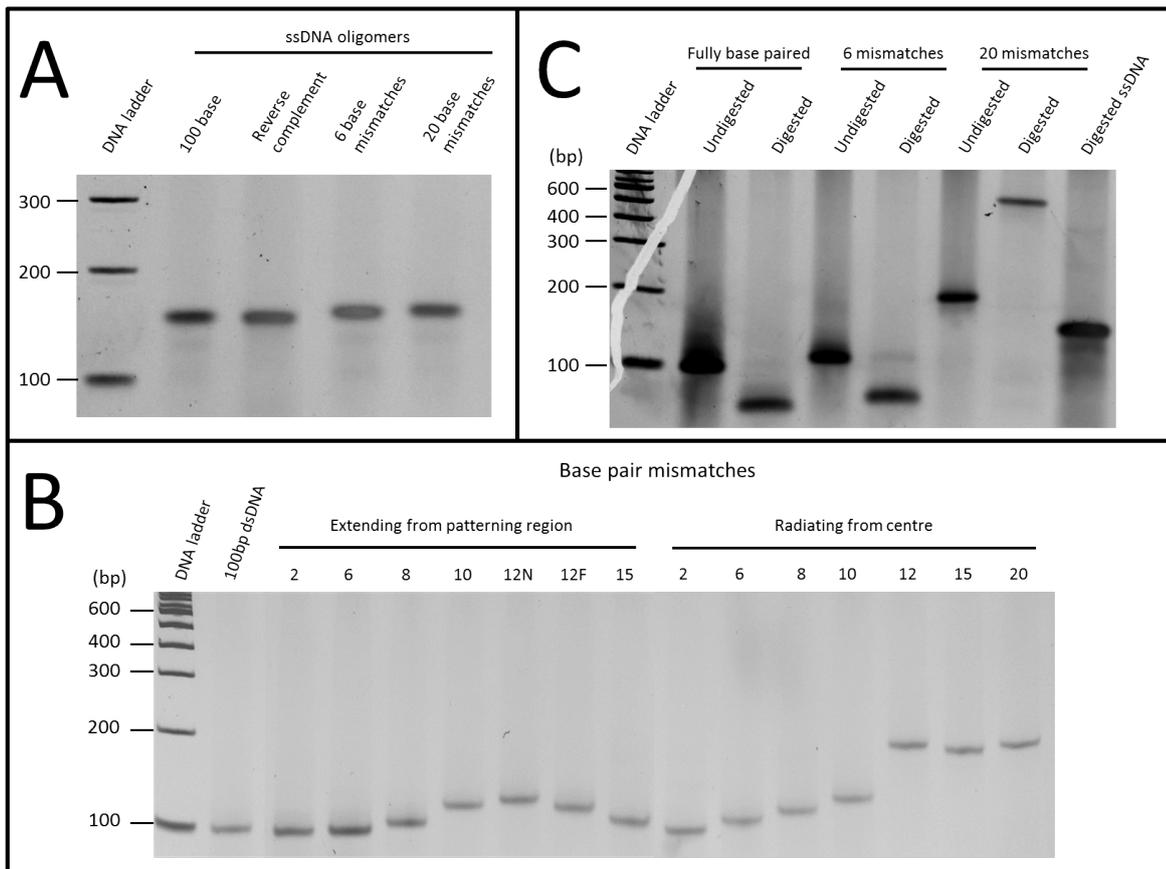


Figure 4.3: **A:** Single stranded DNA oligomers were run on a 15% acrylamide gel. The lanes are as follows: 100 base pair ladder, 100 base ssDNA, 100 base reverse complement ssDNA, 100 base ssDNA with 6 non-complementary bases, 100 base ssDNA with 20 non-complementary bases.

B: The efficiency of annealing DNA sequences is shown through PAGE analysis. A number of the structures shown diagrammatically in Figure 4.3 are visualised here. From left to right: The DNA ladder, fully base paired 100 base DNA. The proximal mismatches 1, 2, 6, 8, 10 and 12. The 12 base construct transposed away from the patterning region. The 15 base proximal mismatch. The central mismatches of 0, 2, 6, 8, 10, 12, 15 and 20 bases. This image is a composite of two gels aligned using the ladder.

C: Diagnostic digestion of annealed DNA using XapI. From left to right: 100 bp DNA ladder, undigested fully base-paired DNA, digested fully base-paired DNA, undigested DNA with a 6 base mismatch, digested DNA with a 6 base mismatch, undigested DNA with a 20 base mismatch, digested DNA with a 20 base mismatch, digested 100 base ssDNA.

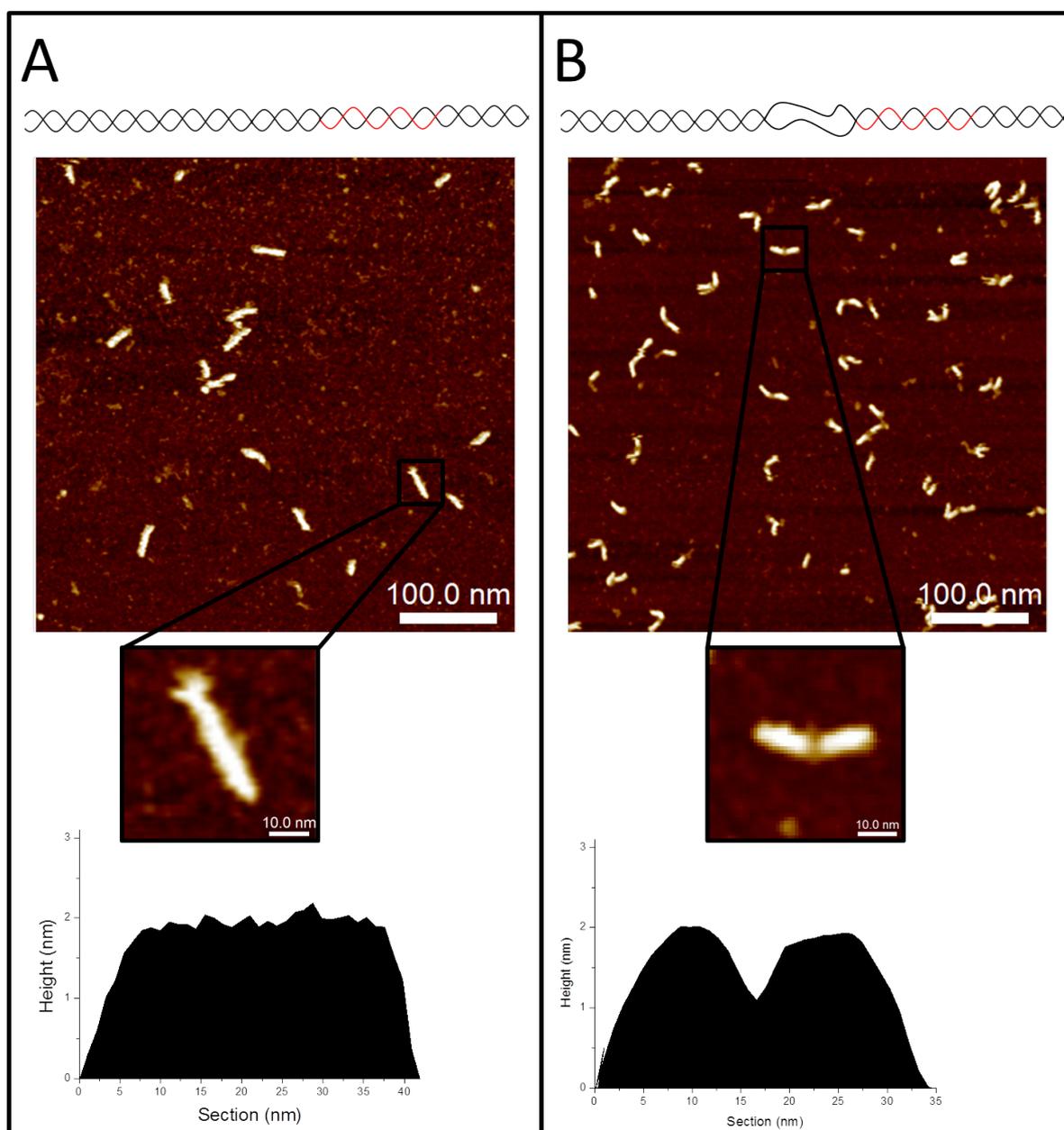


Figure 4.4: Peak force tapping images taken in liquid of DNA samples on a mica surface pre-incubated with 20 mM NiCl_2 . Diagrammatic representations of the structures are shown at the top, below are 500 nm AFM scans showing height and size scales. A single particle was used for height analysis, shown in detail below the main AFM image with the vector used for the analysis. Height and size scales are also shown for these images. A graph showing the heights of pixels across the particle is shown at the bottom. Analysis was undertaken with Nanoscope Analysis software version 9.1. **A:** Fully base paired 100 base DNA **B:** Structure containing a 20 base mismatch.

Height analysis of single particles from each image was carried out. The particles were chosen because they were straight; this was necessary for the cross-section tool of the image analysis software. Height analysis demonstrated an area of lower height at the centre of the 20 base mismatched structure likely to be the ssDNA region. This feature was not present in the fully base-paired DNA. The angle at which the particles were bent was analysed using the image in Figure 4.4, the distribution is shown in Figure 4.5.

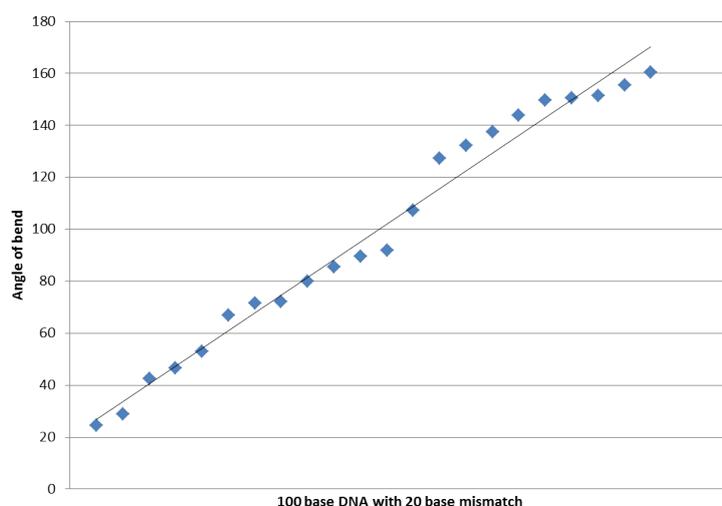


Figure 4.5: ImageJ was used to analyse an image showing 100 base structures with a 20 base mismatch. The angle between the two halves of the structure was measured and the data was plotted. Analysis showed a linear distribution of angles between 24° and 160°.

The distribution of angles is linear, showing a random alignment of these particles between 24° and 160°. It is likely that the angle is due to the orientation of the molecule at the moment it interacts with the surface and becomes immobilised due to electrostatic interactions with the bridging magnesium ions. It is possible that angles below 24° are sterically tolerated but that the image did not resolve the two halves of the molecule sufficiently for them to be analysed. When ImageJ was used to analyse particles from the same image, the average length of the particles was found to be 34 (± 2 , $n=7$) for fully base paired DNA and 33 nm (± 2 , $n=19$) when the 20 base mismatch was included. This is approximately consistent with the value of 0.34 nm per base pair which is considered to be the length of B-DNA, given that the structures are 100 bases long.

4.2 Patterning proximal mismatch structures

Patterning experiments were carried out on a series in which the patterning site was proximal to the unpaired region (proximal), shown diagrammatically in Figure 4.6. The extension of the region occurred in the 3' to 5' direction, away from the patterning region. The mismatched region remains proximal to the patterning site as the region increases in size.

A typical PAGE analysis of the patterning experiments described in Chapter 3 Section 3.1.7 is shown in Figure 4.7 (A), using the series of sequences with the mismatches proximal to the patterning region (Figure 4.6). In the patterning experiment a 30 base ssDNA oligomer was used to create a nucleoprotein filament. This was then introduced at a 30:1 ratio to the dsDNA containing a homologous sequence and incubated to allow the formation of the triple stranded complex.

Patterning was assessed through the use of a restriction enzyme (*XapI*); the recognition site for *XapI* lies within the homologous region for the ssDNA strand (Figure 4.3 A). When the patterned region was formed a digest was carried out; in the event that patterning was successful the triple stranded region occluded the enzyme recognition site and a lower level of digestion was seen. Following this step, the protein was digested using Proteinase K to improve the resolution of gel imaging by removing protein from the DNA.

Digestion was observed through PAGE analysis, the results are shown in Figure 4.7 (B). Where patterning had occurred there were two bands of intensity, the digested (lower) and undigested (upper) DNA species. The exception to this is the 20 base mismatch, in which the upper band represents the digested DNA, as demonstrated earlier in Figure 4.3 (B). By comparing the relative intensities of the upper and lower bands the efficiency of *RecA* patterning was determined [220, 18, 15]. The band or bands which are made up of the ssDNA patterning oligomer and the T30 oligomer are not present on the gel; the gel has been run for long enough that these species have run off the bottom of the gel.

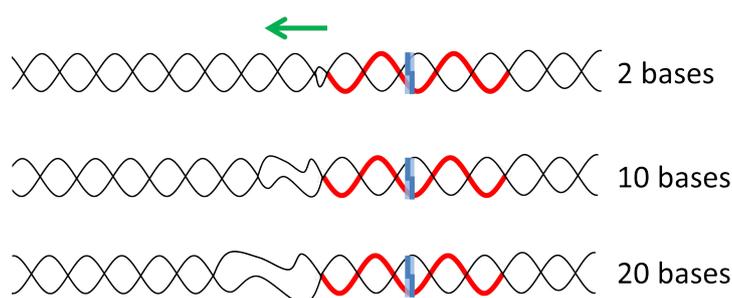


Figure 4.6: A series in which the patterning site was proximal to the unpaired region (proximal). The extension of the region occurred in the 3' to 5' direction, away from the patterning region. The patterning region is represented by the red part of the helix. The restriction site is shown as a blue box.

The densitometry data from this experiment is shown in Figure 4.7 (B). When all of the mismatched regions are at the same distance from the patterning region there is a direct increase in patterning efficiency with the increase in number of mismatched bases. This increase saturates at the 8 base mismatch when there is greater than 95% protection of the dsDNA from digestion.

An increase in patterning efficiency was observed with an increase in the size of mismatched regions. The highest patterning efficiencies can be seen for the proximal sequences with base mismatches of over 8 bases; above 95% of the dsDNA remains undigested when these mismatches are present. RecA mediated triple strand formation underwinds DNA [146] which is likely to create areas of high tension proximal to the patterned region.

The high patterning efficiency achieved with the sequence containing a longer mismatched section could be the result of the ssDNA region absorbing the increased helical tension caused by triple strand formation. It is also possible that the change in efficiency could be linked with an increase in DNA breathing around the mismatch site. DNA breathing is the transient breaking of base pairs at temperatures below the T_m [250, 251, 252, 253].

Both of these mechanisms correlate well with data showing that patterning at the terminus of a dsDNA strand increases patterning efficiency [221, 16]; it is likely the increase in

efficiency observed here arises from a combination of the relaxation of patterning-induced underwinding and increased accessibility for the nucleoprotein filament.

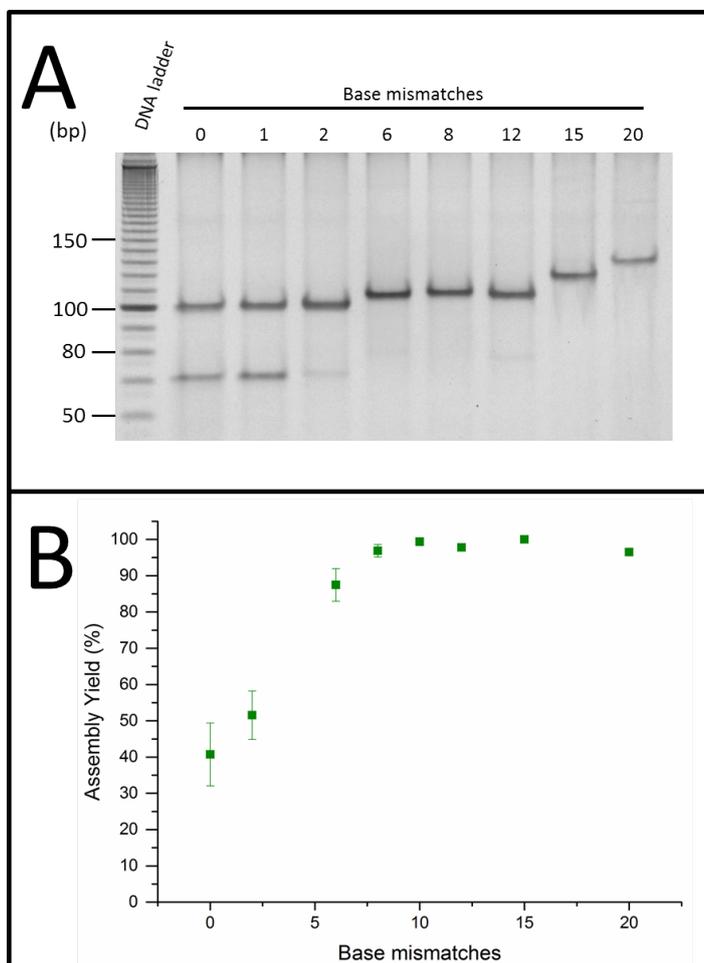


Figure 4.7: **A:** PAGE analysis of a patterning experiment using the 100 base structures containing base mismatches proximal to the patterning region shows the level of DNA digestion. A single band indicates no digestion. In lanes with two bands of intensity the lower band represents digested DNA. The number of bases that are not base paired is shown above the lanes on the gel. **B:** Densitometry analysis of gels shows the level of DNA protected from digestion by the RecA mediated triple-stranded region. As the number of base pair mismatches increases the patterning efficiency increases. Error bars show standard deviation, the sample size is 3.

This experiment identified the effect of proximal mismatched regions on patterning, further work was then necessary to elucidate whether this effect was proximity dependent.

4.3 Patterning 12 base mismatches

The 12 base proximal structure results in an extremely high level of patterning, consistently showing protection of above 95%. The 12 base mismatch was within the saturation level of the proximal experiment (Figure 4.7) and small enough to be moved some distance within the sequence without disrupting the base-pairing regions. For these reasons experiments were done on a series of 12 base mismatches that were designed to explore the effect of their proximity to the patterning region on the patterning efficiency, shown digramatically in Figure 4.8. The same sequence was transposed across the upper strand of the DNA template in a 3' to 5' direction, moving further from the patterning site. The proximal 12 base mismatch began at base 49 and finished at base 60. The 'central' 12 base mismatch began at base 45 and finished at base 56 and the 'far' 12 base mismatch began at base 42 and finished at base 53.

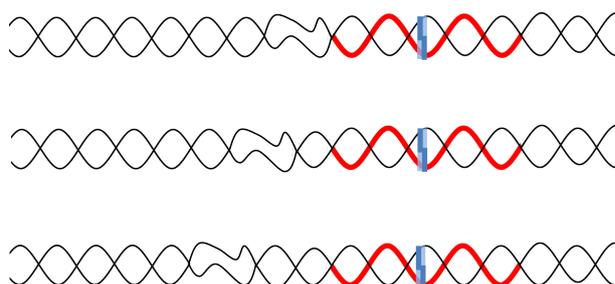


Figure 4.8: A series in which the mismatched region was maintained at 12 bases and moved in the 5' direction away from the patterning site. The patterning region is represented by the red part of the helix. The restriction site is shown as a blue box.

4.3.1 Annealing

The oligomers were annealed in the same protocol as used in the previous experiment. The result of annealing are shown in Figure 4.9 (A). Although the upper sequence of the 12 base mismatches remains the same, the sequence on the complement strand changes as the sequence is transposed. As for the structures containing the 20 base mismatch, secondary structural elements such as hairpin loops form in the non-base pairing regions of the DNA, increasing the interaction with the gel matrix which slow the movement of the

strands; this is the most likely reason for the three structures showing different mobility through the gel. A single band on the gel indicates that annealing was successful.

4.3.2 Patterning

The same 30 base ssDNA patterning oligomer was used at a 30:1 ratio in an experiment identical to that undertaken for the first set of templates. A typical patterning experiment is shown in Figure 4.9 (B). Densitometry data is shown in Figure 4.9 (C). The data show that proximity is important to the increase in patterning efficiency associated with non-base pairing regions. Transposing the patterning region by four bases causes a decrease in patterning efficiency from 96% (± 1.5) to 54% (± 15). A further move of three bases leads to patterning efficiency of 32% (± 21). The error bars for the Central and Far constructs overlap, indicating that this further distance may not lead to further changes in patterning efficiency. On the basis of this experiment it was decided to further explore the effects of size and proximity on patterning.

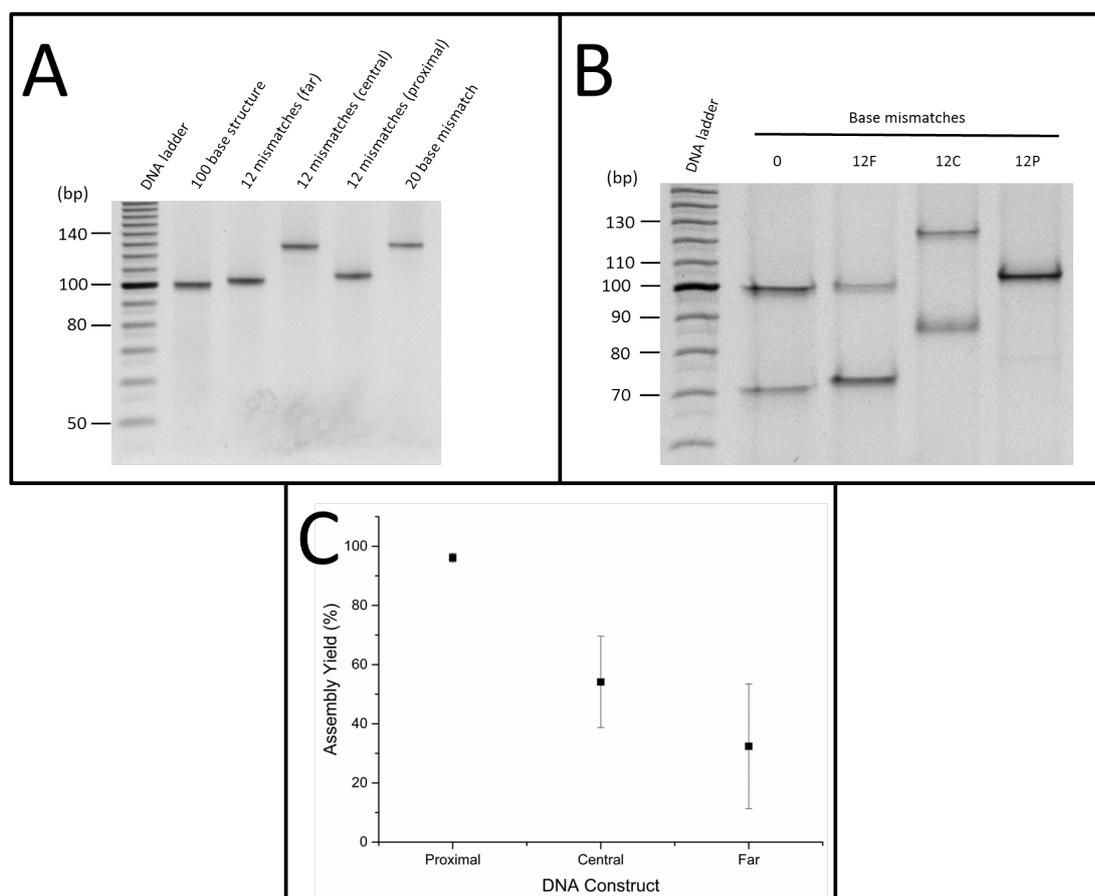


Figure 4.9: **A:** PAGE analysis of annealing of oligomers containing base mismatches. From left to right: 10bp DNA ladder, the fully base paired 100 base structure, the 12 base mismatch extending from base 42 to base 53 (far), the 12 base mismatch extending from base 45 to base 56 (central), the 12 base mismatch extending from base 49 to base 60 (proximal), the sequence containing a 20 base mismatch. **B:** PAGE analysis of a typical patterning experiment using the 100 base structures containing 12 base mismatches at base 42 to base 53 (12F), from base 45 to base 56 (12C) and base 49 to base 60 (12P) showing the level of DNA digestion. A single band indicates no digestion. In lanes with two bands of intensity the lower band represents digested DNA. **C:** Densitometry analysis of an experiment exploring the effect of proximity to the patterning region of a 12 base mismatch on the efficiency of RecA mediated protection from digestion. All mismatches are 12 bases in length but are transposed across the upper strand. Far begins at base 42, central begins at base 45 and proximal begins at base 49. Error bars show standard deviation, the sample size is 3. The results are: Proximal 96% (± 1.5), Central 54% (± 15), Far 32% (± 21)

4.4 Patterning central mismatch structures

The series of sequences shown diagrammatically in Figure 4.10 was designed to investigate the combined effect of proximity and mismatch size on patterning efficiency using a 30 base ssDNA patterning oligomer.

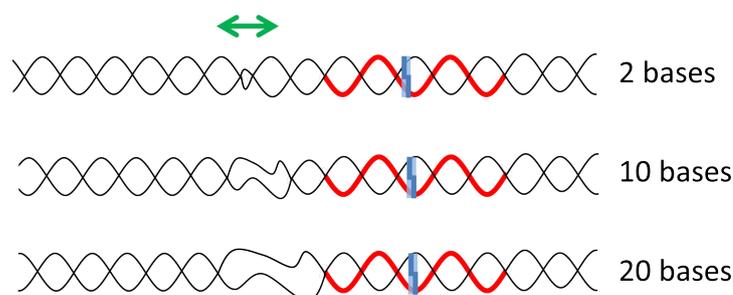


Figure 4.10: A series in which the unpaired region was at a decreasing distance from the 5' end of patterning site with the increase in the size of the unpaired region (central). The first mismatch is at the centre of the sequence and the extension of this region radiated from the centre in both directions. The patterning region is represented by the red part of the helix. The restriction site is shown as a blue box.

Experiments were carried out to explore the effect of the change in these structures on patterning efficiency. The resulting behaviour was not as expected, the experiments were repeated at different ratios of nucleoprotein filament to dsDNA to provide additional confirmation of the pattern observed. The results of an experiment investigating this effect and also the effect of changes in the nucleoprotein filament to dsDNA ratio are shown in Figure 4.11. The protection levels observed in this experiment are presented in Table 4.1.

The data shown here do not include error bars as they each represent a single experiment and the three separate experiments were carried out at different ratios of nucleoprotein filament to dsDNA. Despite the difference between the experiments, a trend can be seen in common between the three data sets. Experimental error may play a role in the differences between these results but the trend is compelling. While the series of sequences with the mismatched region extending from the patterning region (Figure 4.1 B) shows an increase in patterning efficiency with the increase in the number of mismatched bases, this is not the case for the third series of sequences, the central mismatches (Figure 4.10). Patterning

efficiency does not initially increase but, as the proximity of the mismatched region to the patterning region increases, patterning efficiency improves.

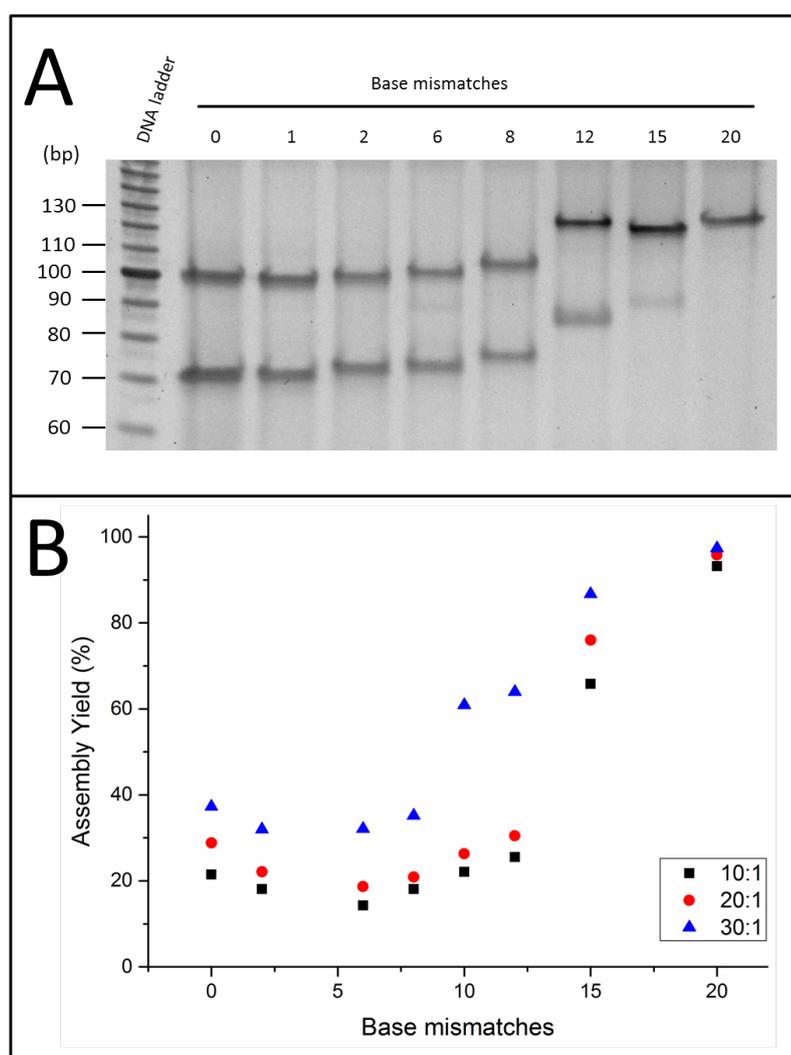


Figure 4.11: **A:** PAGE analysis of a typical patterning experiment using the 100 base structures containing 12 base mismatches at the centre of the strand. A single band indicates no digestion. In lanes with two bands of intensity the lower band represents digested DNA. Base mismatches are indicated above the lane. **B:** Densitometry analysis of PAGE experiments done on samples to show the patterning efficiency of RecA nucleoprotein filaments on the centrally radiating DNA structures containing mismatches. Experiments were done at different ratios of nucleoprotein filament to dsDNA: 10:1, 20:1 and 30:1.

Table 4.1: The level of protection from digestion in central structures afforded by RecA patterning at different nucleoprotein filament to dsDNA ratios. Protection is measured by densitometry analysis of PAGE as a percentage of the total intensity of bands in a lane.

Base mismatches	10:1 ratio	20:1 ratio	30:1 ratio
0	22	29	37
2	18	22	32
6	14	19	32
8	18	21	35
10	22	26	61
12	26	31	64
15	66	76	87
20	93	96	97

The effect of base mismatches on patterning efficiency remains consistent for all the nucleoprotein complex to dsDNA ratios explored. At a lower nucleoprotein filament ratio of 10:1 or 20:1 a decrease in patterning efficiency was observed. Previous work has shown that a patterning ratio has a strong effect on patterning efficiency [16]. The effect of an increase in the ratio of nucleoprotein complex to dsDNA is a increase in patterning efficiency up to a saturation point at which non-specific binding occurs. The increase in efficiency is probably due to an increase in the percentage occupancy of the patterning sites.

There appear to be two competing effects when patterning these DNA sequences with RecA. This results in a two phase relationship between the increase in base pair mismatches and patterning efficiency. As the number of base mismatches initially increases there is a decrease in the observed patterning efficiency; the sequence with no mismatches at a 30:1 ratio shows 37% protection; this decreases to 32% for the sequences containing a two or six base mismatch (See Table 4.1).

The effect is even more pronounced when a lower ratio of nucleoprotein complex to dsDNA is used. When a 10:1 ratio is used, the fully base-paired protection level is 22%; this drops to 18% when there is a 2 base mismatch and 14% when there is a 6 base

mismatch. When a 20:1 ratio is used the protection is 29% for fully base-paired DNA, 22% when a 2 base mismatch is present and 19% when a 6 base mismatch is present.

The 6 base mismatch consistently results in the lowest levels of protection. The 8 base mismatch in this series extends only one base further towards the patterning region, beginning at base 46 and extending to base 54. However, the recovery in protection indicates that this has a significant effect on its interaction with the patterning region. The proximally situated 8 base mismatch was sufficient to saturate the patterning, showing complete protection, as shown in Figure 4.7.

The results suggest that an increased number of base pair mismatches in some way hinders the scanning of the nucleoprotein filament across the DNA. This hindrance appears to be overcome by a higher number of mismatches and increased proximity to the patterning region; the sequence with a 10 base mismatch shows 61% protection at a 30:1 ratio and this increases to higher than 95% for the sequence with a 20 base mismatch, this increase in mismatch size is concurrent with a reduction in distance to the patterning site from 5 bases to 0 bases.

It seems possible that the mismatched region acts to increase patterning efficiency by absorbing the tension caused by DNA underwinding during triple strand formation [146]. The hindering effect of introducing a mismatch at some distance from the patterning region could be an indication that this sort of topology hinders the *RecA* in its sequence search. Given that the nucleoprotein filament slides along dsDNA to aid its search [208], it is possible that the mismatched regions interfere with the sliding and thus reduce patterning efficiency. The data indicate that the inhibitory effect is overcome when the mismatched region is proximal to the patterning region; it seems likely that the sliding activity is still hindered, but once in place the ability of the nucleoprotein filament to remain bound is increased by the tension absorbing properties of the ssDNA section.

It is also possible that the increase in patterning efficiency is the result of an increase in accessibility due to the transient breaking of the hydrogen bonds at the termini of the dsDNA region, known as 'breathing' [250, 251, 252, 253]. DNA breathing has been linked to increased DNA accessibility at the replication fork for T4 helicase [254]. It is possible that breathing also aids the formation of the *RecA*-DNA triple stranded complex.

4.5 Patterning nicked structures

Having explored the size and proximity of non-base pairing regions, further work investigated whether the effects of mismatched regions on patterning efficiency could be replicated by a nick in the DNA. The nicked structure is shown diagrammatically in Figure 4.12 (A). The lower DNA strand remains the same but the upper strand is replaced by two, shorter DNA strands with the same total number of bases. A break in the phosphate backbone would be unlikely to have the same effect in hindering the sliding of the *RecA* filament as a large region of non-base pairing DNA, but could still act to reduce the underwinding tension induced by triple-strand formation.

4.5.1 Annealing

Fully base-paired 100 base sequences containing a break in the phosphate backbone were annealed through the same protocol as the other 100 base structures. The ssDNA oligomers are shown in the last four lanes of Figure 4.12 (C). The annealed, nicked DNA structures are shown in Figure 4.12 (B).

Annealing appeared to be successful. The mobility of the 45/55 and 40/60 sequences was similar to fully base-paired DNA. The 45/50 structure, which contained a 5 base ssDNA segment, shows a different mobility, probably due to the increased flexibility of the ssDNA region or due to secondary structure formation.

To confirm that a dsDNA structure had been formed a diagnostic digest was carried out using *XapI*, shown in the first four lanes of Figure 4.12 (C). Two shorter bands were observed following the digest of approximately 30 and 70 bases; the cleavage point for *XapI* on this construct is between bases 68 and 69.

Two structures were tested, placing the nick proximal to the patterning region or at a 5 base distance. The 40/60 structure uses a 40 base and a 60 base oligomer to replace the upper strand of the 100 base structure, patterning takes place on the 40 base sequence. The 45/55 structure uses a 45 base and a 55 base oligomer to replace the upper strand of the 100 base structure, patterning takes place on the 45 base sequence.

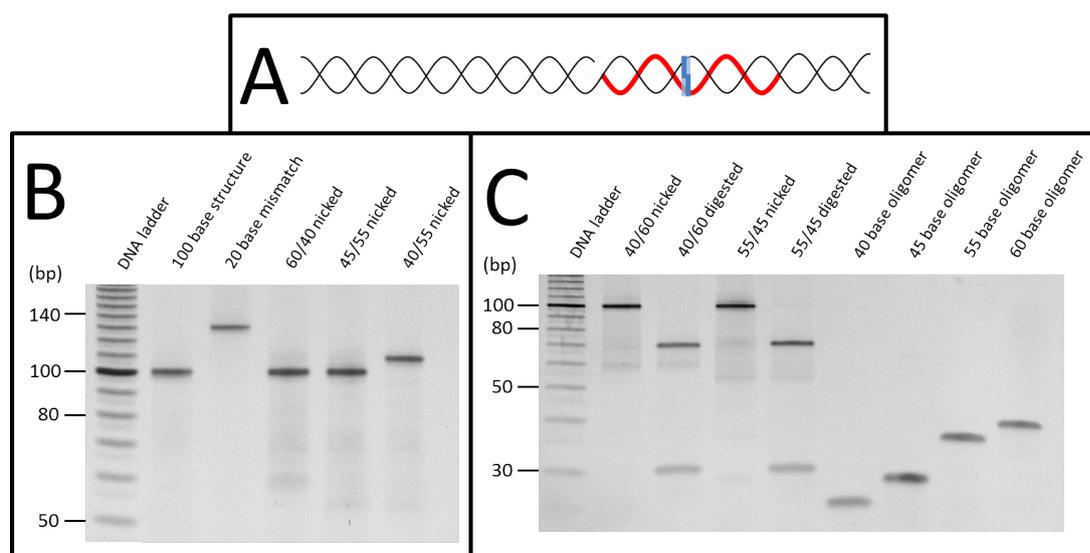


Figure 4.12: Fully base-paired 100 base structures containing a break in the phosphate backbone. **A:** Diagrammatic representation of a nicked structure. The lower DNA strand remains the same but the upper strand is replaced by two, shorter DNA strands with the same total number of bases. The patterning region is represented by the red part of the helix. The restriction site is shown as a blue box. **B:** Annealed structures containing mismatches or nicks. In the first lane is the 10bp DNA ladder. From left to right: fully base-paired 100 base dsDNA, the 100 base sequence containing a 20 base mismatch, the structure made with a 60 and a 40 base oligomer, the structure made with a 55 and a 45 base oligomer, the structure made with a 40 and a 55 base oligomer. **C:** A digestion test to confirm correct annealing of the oligomers. From left to right: 10bp DNA ladder, the structure made with a 60 and a 40 base oligomer, the 60/40 structure digested with XapI, the structure made with a 55 and a 45 base oligomer, the 55/40 structure digested with XapI, the 40 base ssDNA oligomer, the 45 base ssDNA oligomer, the 50 base ssDNA oligomer, the 60 base ssDNA oligomer.

4.5.2 Patterning

The structure in which the nick was between the first base complementary to the patterning oligomer and the base prior to it (bases 60 and 61) was tested by patterning with a 30:1 ratio of nucleoprotein filament to dsDNA. The results are shown in Figure 4.13,

along with results from the proximal and central mismatched sequences for comparison.

The nicked DNA shows an increase in efficiency when compared to the fully base paired DNA. With a DNA protection level of 85% (± 0.3), it falls between the proximal mismatches of 2 and 6 bases, which show efficiencies of 52% (± 7) and 90% (± 5) respectively.

Comparing it to the central sequences patterned with the same ratio of nucleoprotein filament to dsDNA, it shows protection equivalent to a mismatch of between 12 and 15 bases (64% and 87% respectively). The change in patterning efficiencies observed here could be attributed to tension associated with triple strand formation, which was reduced through the introduction of a break in the phosphate backbone of the strand. Easier access to the patterning site due to DNA breathing could also play a role in this effect.

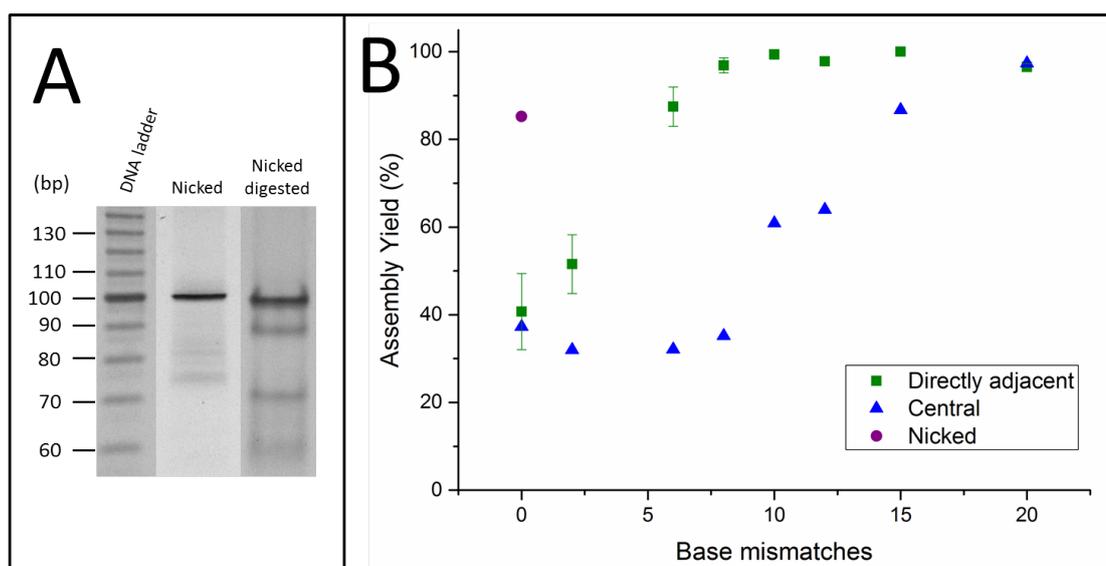


Figure 4.13: **A:** PAGE analysis showing the construct with a nick in the backbone between bases 60 and 61, next to a typical patterning experiment using this construct. This gel is a composite of two images aligned using the DNA ladders. **B:** Densitometry analysis of PAGE experiments done on samples to show the patterning efficiency of RecA nucleoprotein filaments on DNA structures containing mismatches or a nick. Proximal structures shown in green squares, central structures shown in blue triangles, the nicked structure is shown as a purple circle.

4.6 Exploring the effect of patterning oligomer complementarity

An experiment was undertaken to gain further information about the patterning oligomer used. Throughout the experiments in this section the 30 base oligomer chosen was kept constant, except in the experiments discussed in this section. The complementarity of the patterning oligomer was changed to explore the effect on patterning efficiency.

The 30 base patterning oligomer used in previous patterning experiments was used in parallel with an oligomer which was its reverse complement (RC). The 30 base oligomer was homologous to the bases 61 to 90 of the upper strand of the 100 base sequence. A series of structures containing either a mismatched region or a nick in the phosphate backbone were used in this experiment.

In the experiment shown in Figure 4.13 the nick was placed between bases 60 and 61 (40 and 60 base oligomers). This experiment utilises two further structures; one in which the nick was placed between bases 55 and 56 (45 and 55 base oligomers) and a third structure with the 45 and 50 base oligomers in which a 5 base region of ssDNA was present in the reverse complement strand. Both of these structures are shown annealed in Figure 4.12. Patterning with both the 30 base oligomer and its reverse complement is shown in Figure 4.14 (A).

The data shows very little difference in patterning efficiency between the two patterning oligomers, at both high and low efficiencies and with a range of structures. There is also very little difference in patterning efficiency between the different nicked structures used in this experiment.

The 5 base difference in the position of the nick results in almost no decrease in patterning efficiency. The 20 base mismatch of the central structures begins at base 41 and ends at base 60. The 10 base mismatch begins at base 46 and ends at base 55. The 5 base difference in this case causes a drop in patterning efficiency from 97% to 61%. This could indicate that the effect of a nick in the phosphate backbone has a longer range through the dsDNA strand, or confirm that the nick hinders the scanning of the nucleoprotein filament along the dsDNA to a much lesser extent.

If RecA were to be exploited for patterning of more complex structures, such as origami, these results might inform us about the potential behaviour of RecA in this context. In an origami structure, the short oligomers are wound around a long ssDNA genome creating a structure in which there is a high level of tension but also regular nicks in the phosphate backbone, where the short oligomers meet. It is probable that the nicks in DNA structures act to relieve some tension in these structures, as observed in the structures discussed here.

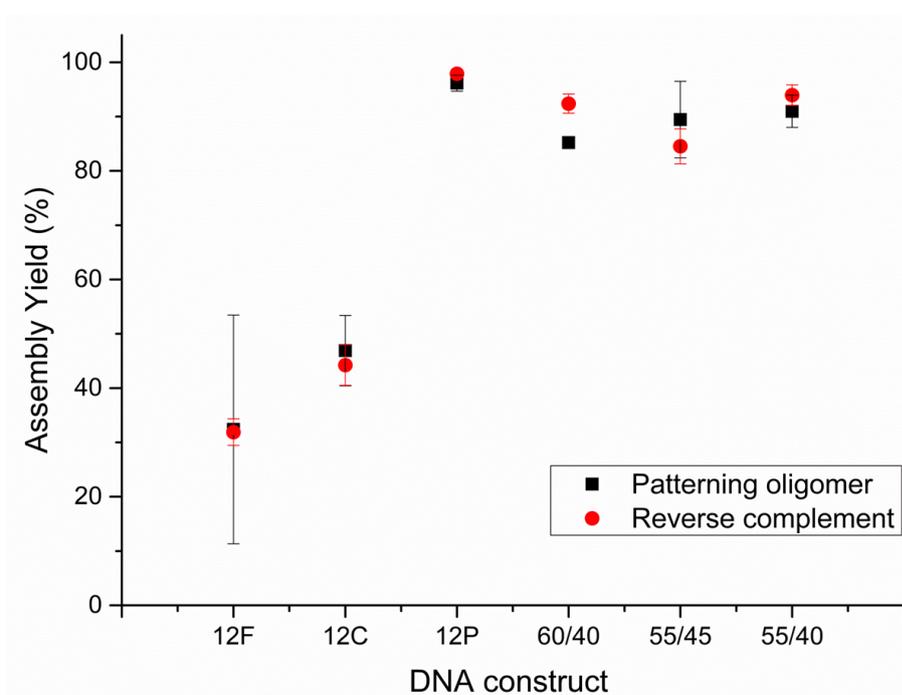


Figure 4.14: Densitometry analysis of patterning experiments using the 30 base patterning oligomer used in all previous experiments and a second patterning oligomer which was the reverse complement (RC) of this oligomer. The RC patterning oligomer was homologous to the lower strand of the 100 base sequence. The transposed 12 base mismatch and the nicked structures were patterned with both oligomers and analysed by PAGE. Experiments on the 30 base patterning oligomer are shown as black squares. Experiments with the RC oligomer are shown as red circles. Error bars show standard deviation, the sample size is 3.

4.7 Controls

To test whether the observed changes in digestion efficiency were due to RecA patterning a number of control experiments were carried out. They aimed to show that there was no inherent change in the DNA construct as a result of interaction with RecA and that in the absence of RecA digestion would occur under these reaction conditions.

Control experiments are shown in Figure 4.15. They were carried out using the central structures (Figure 4.10) and a nucleoprotein filament to dsDNA ratio of 30:1.

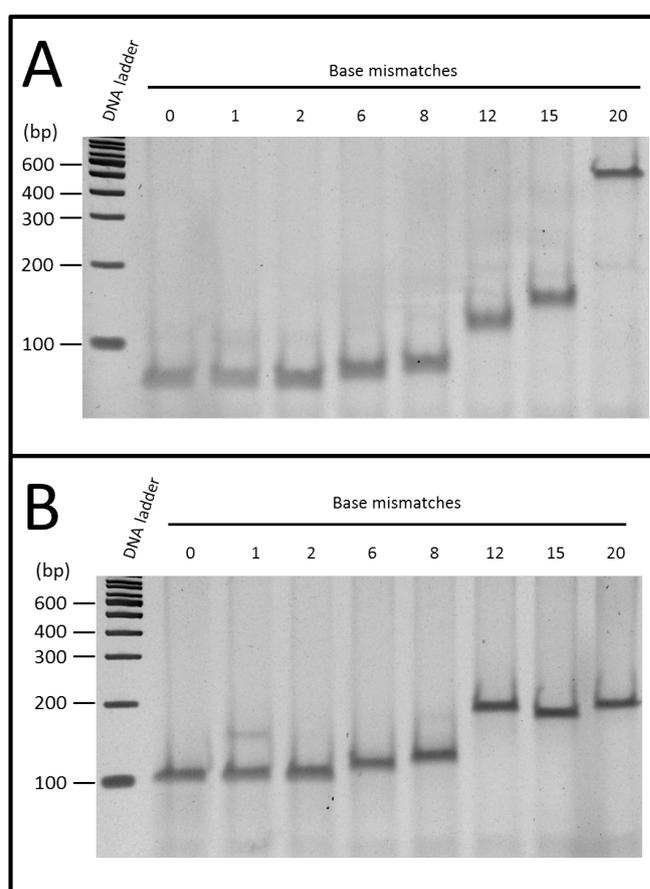


Figure 4.15: **A:** Patterning experiments carried out with heat treated RecA show complete dsDNA digestion following the same patterning protocol as previous experiments. **B:** RecA nucleoprotein filaments formed and then deprotected with Proteinase K without the intermediate digestion step show no permanent change in their structure as a result of RecA triple-strand formation.

RecA was denatured through heating to 60 °C for 20 minutes, as recommended by the

manufacturer. Following heat treatment the patterning protocol was followed and the resulting DNA species were analysed by PAGE, shown in Figure 4.15 (A). All DNA sequences show full digestion as a result of a complete lack of protection. None of the other experimental reactants provide protection from digestion without active RecA.

Figure 4.15 (B) shows undigested, deprotected DNA. The patterning protocol was carried out as in previous experiments but no restriction enzyme was added during the digestion step. As a result, the proteinase K treated DNA shows the same mobility as untreated, annealed dsDNA (Figure 4.3 C). This shows that changes in the mobility of dsDNA following patterning and digestion are not due to an intrinsic change brought about by the interaction of RecA nucleoprotein filaments or any other experimental reactant.

4.8 Conclusions

Initially a series of sequences were produced that contained a range of mismatches that extended from the patterning region. The patterning results from these structures showed that a sequence containing a longer base pair mismatch increased the patterning yield when a 30 base patterning oligomer was used.

Then a series of experiments were undertaken using a 12 base mismatch at different distances from the patterning region. The results showed that the increase in patterning efficiency was distance dependent; when the mismatched region was further from the RecA patterning site the patterning efficiency became comparable to that of the fully base paired dsDNA.

Following this, a series of sequences were produced in which both the length of the mismatch and the distance from the patterning site varied. Experiments on these structures indicated that the mismatched region, when not in close proximity to the patterning site, does not increase efficiency of RecA patterning. Mismatched regions at a distance of only a few bases from the patterning region can result in a decrease in patterning efficiency.

Finally, a nick in the phosphate backbone was introduced in the place of base pair mismatches, to show that the increase in yield was the effect of relaxation and not in some way linked to secondary structure or other single-stranded DNA properties. An

increase in patterning yield was observed when one strand of the 100 base sequence was nicked.

Control experiments showed that using the reverse complement of the patterning oligomer does not have an effect on patterning efficiency. Other control experiments included patterning with *RecA* followed by deprotection with proteinase K without the patterning step to demonstrate that patterning alone does not permanently change the DNA structure. Experiments with heat denatured *RecA* demonstrated that protection from digestion is conferred through the activity of the protein.

It has been shown that *RecA* patterning is possible on more complex DNA structures than has previously been demonstrated on linear or circular dsDNA [16, 18, 15, 17] and that different topologies have an effect on the efficiency of patterning.

Further results indicate a complex relationship between patterning efficiency and unpaired regions; experiments reveal that when not in close proximity to the patterning region mismatch bases can cause a decrease in patterning efficiency, potentially due to the *RecA* nucleoprotein filament being hindered in scanning the dsDNA.

The increase in efficiency associated with a nick in the phosphate backbone indicates that both nicks and mismatches could be increasing patterning efficiency through the release of patterning-induced tension, shown diagrammatically in Figure 4.16 or an increase in accessibility due to DNA breathing. There is no experimental basis for distinguishing between these effects in this work; a combination of these two factors is also plausible.

There is little difference in patterning efficiency between the different nicked structures. This indicates that the proximity of a nick does not have as strong an effect on patterning efficiency as the proximity of a mismatch does; when the 12 base mismatch was transposed there was a strong negative effect on patterning efficiency as the distance from the patterning region increased. This supports the hypothesis that the mismatch itself is detrimental to the interaction of the nucleoprotein filament and dsDNA strand.

The patterning efficiency of the fully base paired sequence shown here is similar to those seen for longer dsDNA [16]. Throughout these experiments it is notable that there is an increase in error with lower patterning efficiency. It seems likely that this variation is due to a greater dependence on small differences in concentrations and differences between,

for example, enzyme batches. Further work is needed to fully characterise the effect of more complex DNA topologies on RecA mediated triple-strand formation.

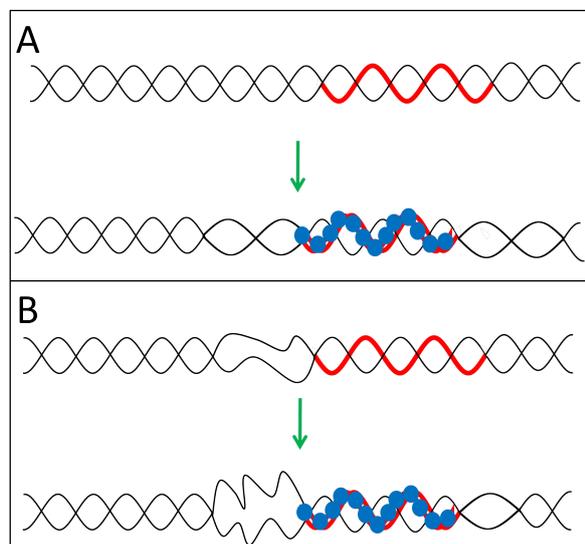


Figure 4.16: Diagrammatic representation of the release of RecA patterning-induced tension in dsDNA by non-base pairing regions. The red region represents the patterning site. Blue circles represent RecA. **A:** In patterning dsDNA the formation of the triple-stranded region induces underwinding in the proximal areas. This is entropically unfavourable and reduces the efficiency of patterning. **B:** The introduction of mismatched or nicked DNA near the patterning region provides a relaxation that reduces the proximal underwinding effects and increases the efficiency of patterning.

The data shown here indicate that including regions of non-base pairing DNA changes the topology of the strand, evidenced by both AFM imaging and the behaviour of these structures in gel electrophoresis. X-ray crystallographic data has shown that the structure of nicked DNA is not significantly different to that of unnicked strands [255]. The highest efficiencies observed in these experiments were the result of proximal mismatched regions; however, nicked DNA also showed high levels of patterning efficiencies. On the basis of the results shown here I would recommend the inclusion of nicks in the phosphate backbone of DNA strands which are to be patterned with RecA for the purposes of DNA nanotechnology.

Chapter 5

RecA patterning of 200 base structures with terminal ssDNA regions

To expand our understanding of RecA patterning it was necessary to expand the library of structures available for experimentation. A series of 200 base structures were designed that would have single stranded elements at one or more termini, shown in Figure 5.1 (Left). The annealing of additional oligomers could then be used to produce three-way junctions, shown in Figure 5.1 (Right).

The 100 base structures were limited in size by reliable and inexpensive oligomer production. A cyclizing ligation reaction was used to make the 200 base structures larger than was possible using synthetic oligomers, shown diagrammatically in Figure 5.2. Cyclizing ligation involves the use of a scaffold to align oligomers for ligation. By cycling the temperatures for melting, annealing and ligating it is possible to achieve high yields with small amounts of scaffold; the ligation products act as scaffolds in later iterations.

The 100 base sequences provided information about the effect of an internal mismatch; the 200 base structures were fabricated with and without the additional annealed oligomers to give structures that would contain three-way junctions or peripheral mismatches. The objective was to test RecA patterning of these structures, which could provide information about the effects of structural constraints. The structures could also potentially be used in the production of a larger array to create a 2D net-like structure.

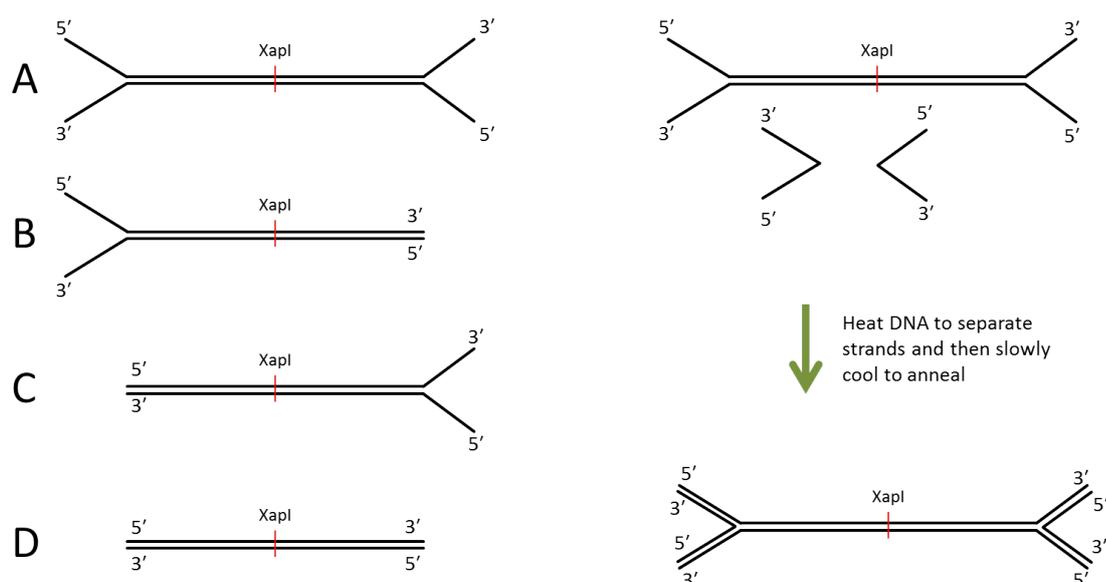


Figure 5.1: **Left:** Four structures were produced using the ligase chain reaction, the sequence of the central 140 bases was the same for all constructs. (A) Both termini have ssDNA elements, (B) the 5' end has ssDNA elements, (C) the 3' end has ssDNA elements and (D) the DNA is entirely double stranded. **Right:** A diagrammatic representation of annealing ssDNA oligomers to unpaired branched junctions to create short three-way branched junctions.

Purification of the synthesised product proved challenging but it was possible to obtain relatively pure samples. While constructs with a single terminal three-way junctions were synthesised, it did not prove possible to anneal the oligomers for both three-way junctions at the same time. Attempts were made to optimise the junction annealing oligomers and the structure itself without any success. Although computational aids are available for the design of origami structures, they were not suitable for use in the design of these structures and therefore we could not constructively change the structure further.

Due to the inability to produce double stranded three-way junctions, experiments with this structure were not pursued beyond RecA patterning of the structure containing an unpaired three way junction at both termini. Results from this experiment provided preliminary data on how transient base unpairing could effect RecA patterning.

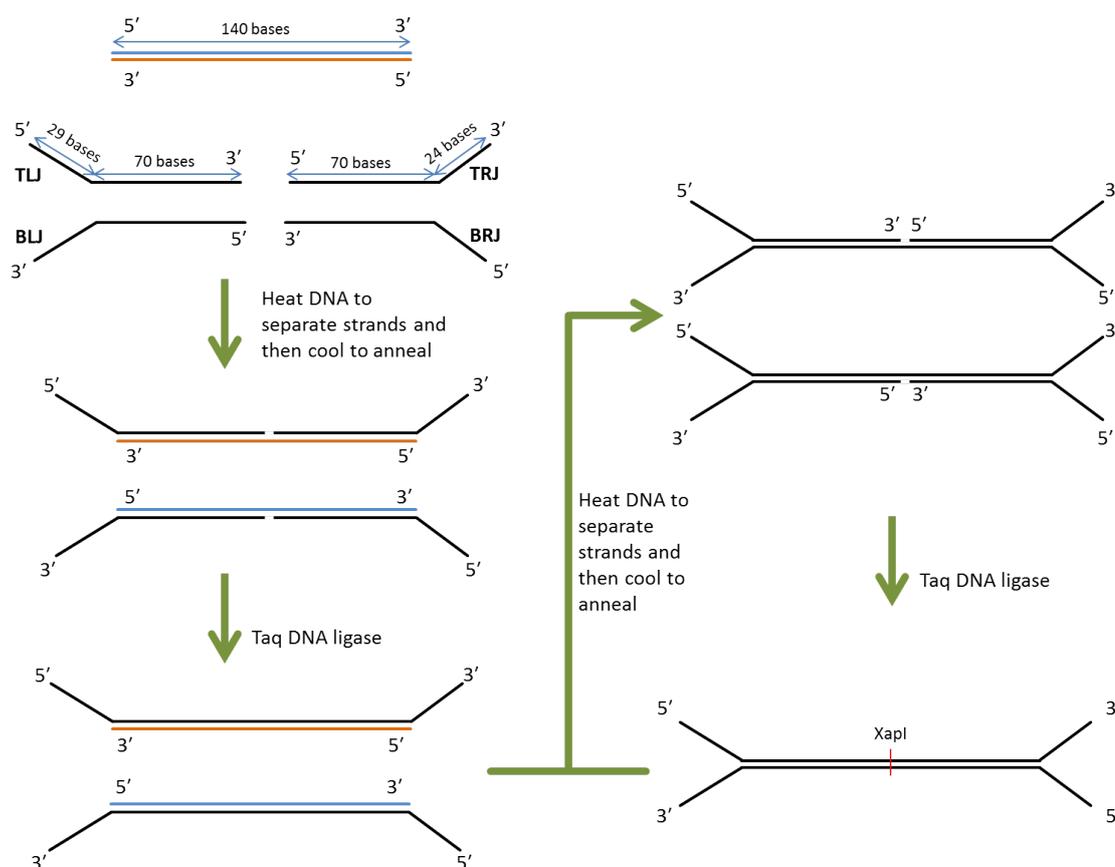


Figure 5.2: The ligase chain reaction. A 140 base dsDNA template (blue and orange) is used to bring ssDNA oligomers (black) together and a heat stable ligase creates a bond. By heating to melt the DNA and cooling to anneal the strands it is possible to repeat this process to create large amounts of product. The products of previous ligation reactions are available as templates in future iterations of the reaction. Small amounts of the original template are used to create a more pure sample of the ligation products. The oligomers are labelled as Top Left with Junction (TLJ), Top Right with Junction (TRJ), Bottom Left with Junction (BLJ) and Bottom Right with Junction (BRJ).

5.1 Fabrication

5.1.1 Design

The aim in designing the 200 base structure was to create DNA topologies that contained a central region, annealed through sequence complementarity, with the option to include

three-way junctions at both or either end, as shown in Figure 5.2. Structures were made by utilising the heat stable enzyme *Taq* ligase, which can resist denaturation at high temperatures. The structures were produced through the templated ligation of four ~ 100 base oligonucleotides. These oligomers were designated Top Left with Junction (TLJ), Top Right with Junction (TRJ), Bottom Left with Junction (BLJ) and Bottom Right with Junction (BRJ), shown labelled in Figure 5.2. Sequences can be found in Appendix A.2. The sequence of the central 100 bases of the 200 base structure is the same as the 100 base sequence, in order that the structures might be comparable.

The ligase chain reaction (LCR) is a single pot process which uses a template DNA strand to anchor two synthetic oligonucleotides proximal to each other in order to facilitate the ligation of the two oligomers into a single strand of DNA by *Taq* ligase. The polymerase chain reaction would not be appropriate for producing these structures as it is not possible to create ssDNA branches with this process. Four oligonucleotides are used to make two ssDNA strands which are complementary at their centres, allowing them to base pair and form a single dsDNA product. The oligomers Top Right with Junction (TRJ) and Bottom Left with with Junction (BLJ) make up the 3' end of the ligated strands; these strands require a 5' phosphate for the ligation to be successful. The use of heat stable *Taq* ligase allows the reaction to be cycled through different temperatures. The melting and re-annealing of the DNA strands allows multiple ligation steps and the product can act as a template for ligation in the steps that follow. The interaction between the template and the synthetic oligomers for the ligation reaction relies on the central, base pairing section of the structure.

The DNA sequences of these structures are listed in Appendix A.2. The patterning site was located from base 97 to 102 for the construct with the 5' ssDNA region, or bases 68 to 73 when this region was not present. The recognition site for the restriction enzyme XapI was placed from bases 85 to 114 for the construct with the 5' ssDNA region, or bases 56 to 85 when this region was not present.

5.1.2 Synthesis

The oligomers used for synthesis were characterised by electrophoresis on a native acrylamide gel, shown in Figure 5.3. The oligomers are 94 (TLJ and BLJ) or 99 (TRJ and BRJ) bases long. Native PAGE showed the most stable secondary structures of the ssDNA strands used in the LCR reaction.

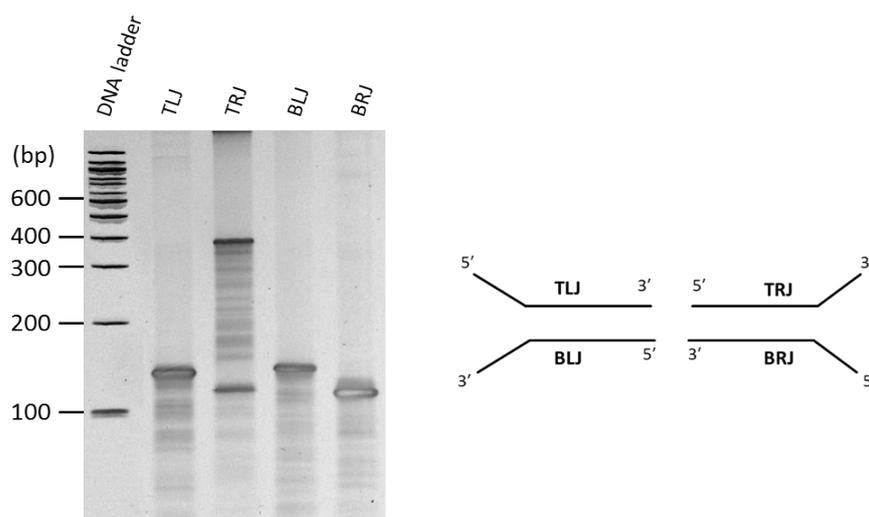


Figure 5.3: **Left:** Oligomers for LCR incorporation into the 200 base structure are shown analysed by native polyacrylamide gel electrophoresis. The samples are: DNA ladder, the 5' oligomer of the upper strands (TLJ), the 3' oligomer of the upper strand (TRJ), the 3' oligomer of the lower strand (BLJ) and the 5' oligomer of the lower strand (BRJ). **Right:** A diagrammatic representation of the oligomers in the gel, later used in the cycling ligation reaction to produce a 200 base structure.

The oligomers produced bands of intensity at ~ 140 and ~ 110 . There was also a ~ 390 base band when the top right oligomer (TRJ) which most likely represents a different stable secondary structure or multimerisation of the oligomer.

During the ligation cycle the oligomers are heated to $95\text{ }^{\circ}\text{C}$, which results in complete melting, and then cooled in the presence of an oligomer with a complementary sequence. The most energetically favourable structure is that which contains the highest proportion of paired bases; for all ssDNA oligomers used in this experiment annealing with a second ssDNA oligomer or the template is much more energetically favourable than any possible

secondary structure.

Having identified the expected band mobility of the unannealed oligomers and of the 140 base scaffold, the cycling ligation reaction was carried out as described in Section 3.2.1, the results are shown in Figure 5.4.

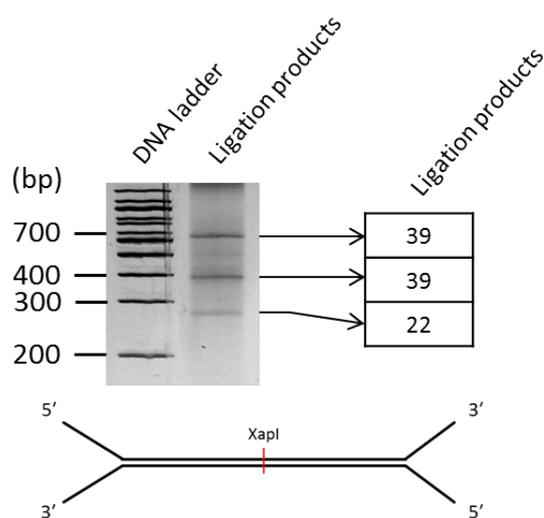


Figure 5.4: **Left:** Native PAGE analysis shows the products of the cycling ligation reaction. Three bands of intensity are shown following the ligation reaction. **Right:** Densitometry analysis shows the proportion of the total intensity in a lane that each band represents. **Below:** A diagrammatic representation of the expected product of this reaction. Given the novel nature of the structure its behaviour on the gel cannot be predicted.

The ligation reaction resulted in three bands of intensity at ~ 300 , ~ 400 and ~ 700 . None of these bands have the same mobility as the 140 base scaffold; referring to Figure 5.3 it can be seen that they do not have the same mobility as the individual oligomers. Densitometry analysis shows that the top two bands seem to represent a larger proportion of the DNA present in the sample; this could be due to the same structural elements that cause it to have a faster mobility in the gel.

It was not possible using these data to identify which band or bands represented the products of the ligation reaction. Further experiments were carried out to characterise the DNA species present following the cycling ligation reaction.

5.1.3 Characterisation

A diagnostic digest was carried out using the enzyme XapI to confirm whether any ligation events had taken place and where they appeared on the gel. The bases which make up the recognition site at the centre of the construct are split between the two oligomers; the complete sequence is only formed when the ligation reaction takes place. Any species that can be digested must be either the ligation product or the 140bp scaffold. An excess of XapI was used to digest the ligation products and the reaction was analysed by PAGE, shown in Figure 5.5.

The PAGE analysis and densitometry in Figure 5.5 show that a single band has been removed; the top band at 800-900 bp is not present following digestion of the sample. The ligation products appear at approximately the 900 base band of the ladder in this experiment and closer to 700 bases in other experiments. It is not clear why this is the case; however, the pattern of bands following digestion remained constant through all experiments. This suggests that this band represents the product of the ligation reaction and that neither of the DNA samples in the bands below contains the digestion site. These two products must be some other result of the mixture of the four oligomers. A new band showing the digestion product of the 200 base structure is not apparent.

The 140 base scaffold digests completely to produce a band at ~ 70 bases. The two products of the 140 base dsDNA are similar enough that they comigrate (See Figure 5.2), resulting in a single band of intensity.

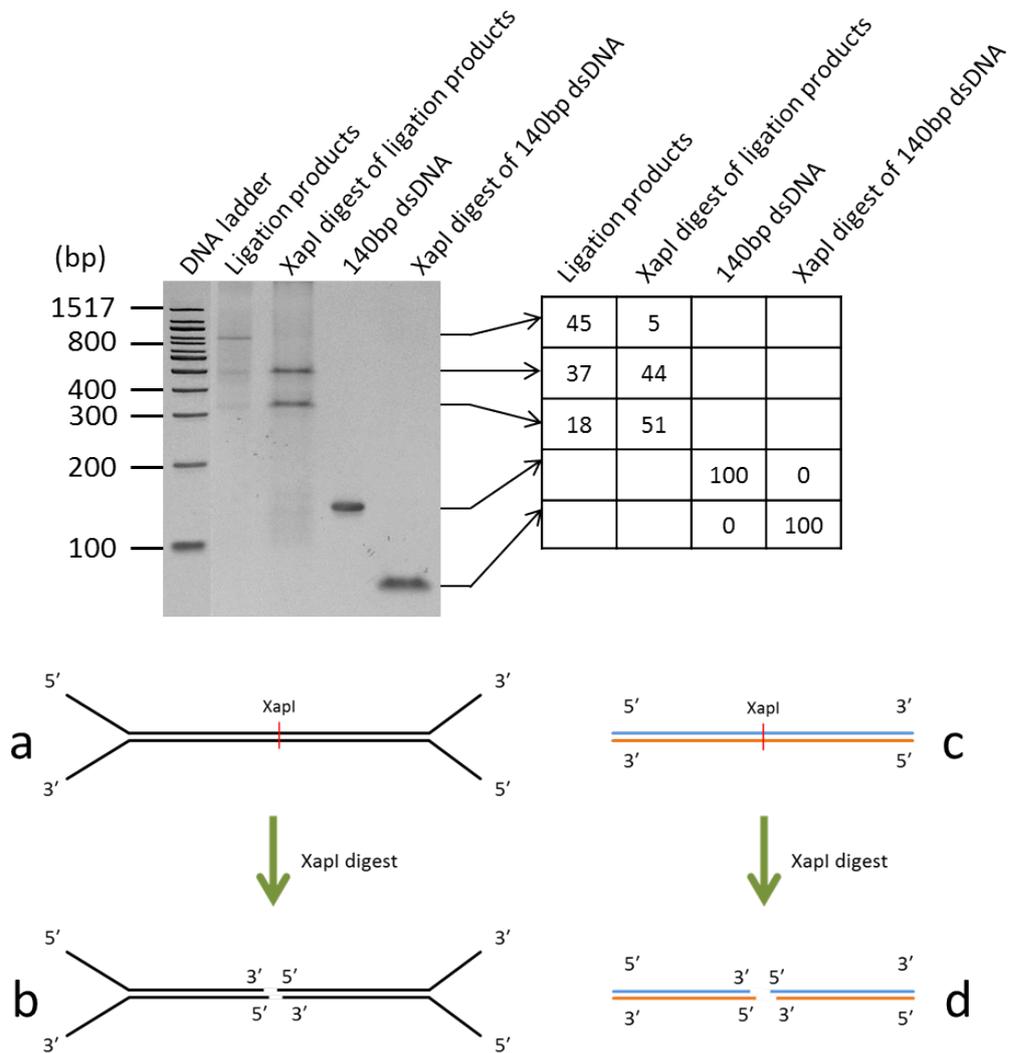


Figure 5.5: **Left:** Native polyacrylamide gel electrophoresis of XapI digestion of LCR products and of the 140 base scaffold used in the LCR reaction. From left to right: the LCR reaction products, the LCR reaction products following digestion with XapI, the 140 base scaffold, the 140 base scaffold following XapI digestion. **Right:** Densitometry analysis of DNA species showing the changes in band intensity following XapI digestion. **Below:** A diagrammatic representation of the digestion using XapI. The product of the ligation reaction (a) is digested to produce two strands (b) with an unpaired three-way junction at one end and a ~70 bp region at the other. The restriction digest of the 140 bp template (c) produces two strands of ~70 bp (d). The action of XapI leaves a 4 base overhang where the digest has taken place. The restriction site is at the centre of both DNA constructs, therefore the two strands produced by each restriction would be almost identical in size.

It was not immediately clear what the DNA species were that produced the two lower bands in the ligation products. It was thought they might be either the ssDNA products of the ligation reaction that had not annealed, or oligomer dimers. The cycling ligation reaction was carried out in the presence or absence of the different DNA species (TRJ, TLJ, BLJ, BRJ) in an attempt to identify these bands; the PAGE analysis of this experiment is shown in Figure 5.6.

When only the oligomers for the upper strand or the lower strand are added to the ligation reaction we see bands that correspond to the ssDNA oligomers; comparison with Figure 5.3 confirms that the mobility is correct for the ssDNA oligomers. The upper bands at ~400 bp or ~300 bp are only seen when there is the opportunity for the oligomer dimers to form, their size corresponds to the bands of intensity observed following XapI digestion of the ligation products. The XapI recognition site comprises three nucleotides on either side of the ligation site; this enzyme digests dsDNA to produce a four base ssDNA overhang. It seems likely that the digestion product is not apparent in Figure 5.5 because the digestion products are nearly identical to the oligomer dimers that have not been ligated.

Having identified the ligation product on the gel, further work was undertaken to ascertain if it was possible to make a series of similar structures using the cycling ligation reaction. PAGE analysis of the DNA products is shown in Figure 5.7.

The gel indicates that it is possible to produce alternative structures using the same protocol. The structures with a single terminal unpaired branched junction have higher mobility in the gel than the structure with both unpaired branched junctions. This is consistent with our previous observations about the effect of ssDNA regions and their secondary structural elements on mobility in polyacrylamide gels. The 140 bp dsDNA has a very similar mobility when produced through the cycling ligation reaction.

Having shown that it was possible to create a series of structures using a cycling ligation reaction, experiments were undertaken to increase the yield of desired product.

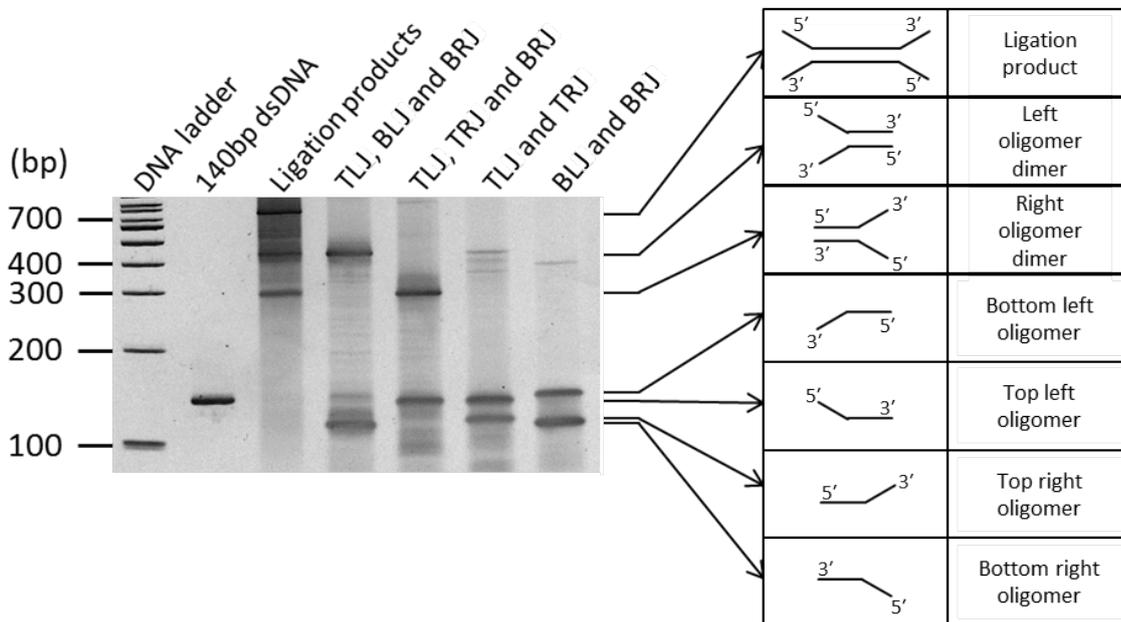


Figure 5.6: **Left:** Native PAGE analysis shows the DNA species resulting from cycling ligation experiments using a combination of oligomers for the identification of different bands. From left to right: the 140 base dsDNA template, the products of the cycling ligation reaction when all four oligomers are included, reaction products when TRJ is omitted, reaction products when BLJ is omitted, reaction products when both BLJ and BRJ are omitted, reaction products when TLJ and TRJ are omitted. Comparison with Figure 5.3 confirms that the mobility is correct for the ssDNA oligomers. **Right:** The different populations on the gel are identified diagrammatically. A ligation reaction in which only two or three oligomers are used results in oligomer dimers and ssDNA; some small proportion of the product is made when both oligomers of a strand are added. However, this product is severely limited by the concentration of scaffold added. TLJ is top left (5') oligomer with unpaired section. TRJ is top right (3') oligomer with unpaired section. BRJ is the lower left (5') oligomer with unpaired section. BLJ is bottom left (3') oligomer with unpaired section.

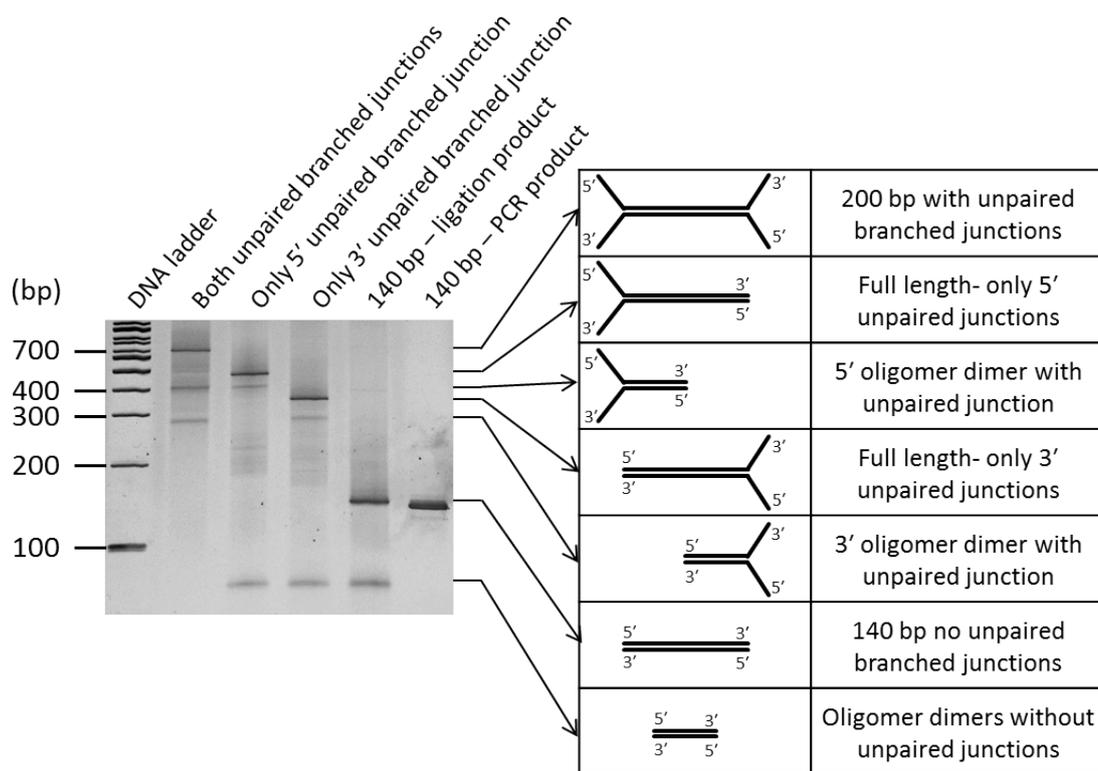


Figure 5.7: **Left:** Native PAGE analysis of cycling ligation products using a range of alternative oligomers. From left to right: structure with unpaired branched junctions at both termini, structure with unpaired branched junction at 5' terminus, structure with unpaired branched junction at 3' terminus, fully double stranded 140 base DNA formed through cycling ligation, fully double stranded 140 base DNA formed through PCR reaction. **Right:** The products of the cycling ligation reactions are identified diagrammatically.

5.1.4 Optimisation

The LCR method of construct formation yields product but it is not at a high yield. There are a number of potential reasons for this; the low yield could be in part due to insufficient 5' phosphorylation which prevents the formation of the desired product. Additionally, the formation of oligomer dimers and the hybridisation of two oligomers with a template strand are likely to occur at an approximately equal ratio, which would limit the dsDNA available for ligation in each reaction. A series of experiments were undertaken in an

attempt to optimise the yield of the reaction.

Effect of ligase activity

An experiment was carried out to ascertain if the yield of product was limited by the ligase enzyme. Calculations showed that the amount of enzyme added was sufficient to ligate all the DNA added in the reaction. The manufacturer states that one unit of *Taq* ligase enzyme is sufficient to ligate 50% of 1 μg of the BstEII fragments of Lambda DNA at 45 °C in a 50 μl volume in 15 minutes. The Lambda genome is 48.5 kbp and contains 13 BstEII sites, creating 14 fragments. The enzyme is supplied at a concentration of 40 units/ μl .

$(1 \mu\text{g Lambda} = 0.032 \text{ pmol}) \times 13 \text{ sites} = 0.416 \text{ pmol of ligation per unit of enzyme}$

$0.413 \text{ pmol} \times 40 \text{ units} = 16 \text{ pmol per } \mu\text{l of enzyme}$, but we must half this as the ligation efficiency is only 50% of the sample. The DNA for ligation is used at 100 pmol/ μl , therefore 1 μl of enzyme is equivalent to a $\sim 3\text{X}$ excess of enzyme. When 2.5 μl are used this is equivalent to a $\sim 7.5\text{X}$ excess, which is more than sufficient to carry out the ligation reaction with a high yield.

While the manufacturer advises that the ligase is stable for a minimum of 30 cycles, it is possible that the majority of the enzyme is denatured after this time. An experiment was undertaken using the structure with ssDNA regions at both termini to assess the effect of additional cycles in which extra ligase was added. The cycling ligation reaction was carried out for 30 cycles after which a further 15 (45 total) or 30 (60 total) cycles were undertaken, with or without additional ligase. The results are shown in Figure 5.8.

The yield of product in this experiment was lower than seen in the experiments already discussed; this reaction was undertaken before further optimisation of the ligation reaction had been undertaken. It can be seen that further cycles increase the yield of product from 1-2% to 10% of the total lane intensity (Figure 5.8). Additional ligase increases the yield when it is added to the reaction volume following the initial 30 cycles. Following 60 cycles the yield is almost doubled from 10% to 18% when additional ligase is added. This result conforms to the theoretically exponential increase in product, due to ligation products acting as templates in following iterations of the reaction.

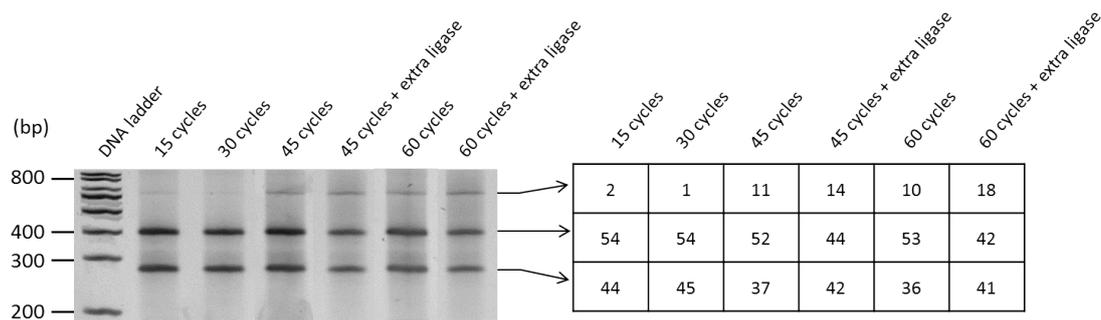


Figure 5.8: **Left:** Native PAGE analysis shows the effects of additional cycles and adding a second aliquot of *Taq* ligase to the reaction mixture. **Right:** Densitometry analysis shows the proportion of the total intensity in a lane that each band represents. The additional cycles and *Taq* ligase increase the yield of 200bp product (top band).

These results indicate that heat denaturing of the ligase is a limiting factor in later cycles and that addition of more ligase at the midpoint of the experiment is beneficial.

Varying annealing temperature

Different hybridisation temperatures were trialled to see if an increase would favour the formation of product over the formation of oligomer dimers. The results are shown in Figure 5.9.

Increasing the temperature leads to a decrease in the amount of ligation product; a change in hybridisation temperature from 72 °C to 81 °C causes a decrease in ligation product from approximately 37% of the population to only 16%. On the basis of this experiment and the melting temperatures of the oligomers involved, the hybridisation temperature was maintained at 72 °C.

Concentration of oligomers used

An experiment was carried out to ascertain whether the concentration of oligomers added to the reaction would have an effect on the yield of product, concentration varied from 10 pmol to 50 pmol. The results are shown in Figure 5.10.

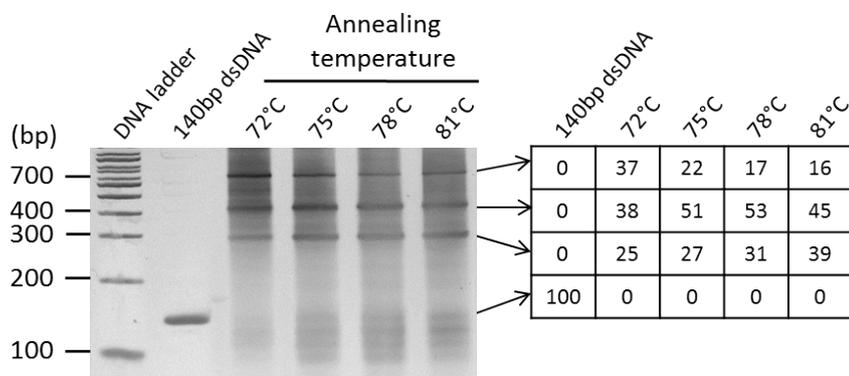


Figure 5.9: **Left:** Native PAGE analysis shows the effect on ligation efficiency of a range of temperatures for the hybridisation step of the cycling ligation reaction. **Right:** Densitometry analysis shows the proportion of the total intensity in a lane that each band represents. Increasing the hybridisation step temperature appears to decrease the ligation efficiency of the enzyme. This could be the result of approaching too closely the melting temperature of the DNA. It might also be due to the effects of increased temperature steps on the ligase enzyme.

It can be seen that the proportion of the population that is ligation product remains approximately the same independent of the concentration of oligomers. Approximately 48% of the DNA has been ligated in both case. A smaller volume of the 50 pmol sample was added to avoid overloading the gel; the absolute amount of product is much higher for the higher oligomer concentration. Therefore, it was decided to carry out further experiments adding 50 pmol of each oligomer.

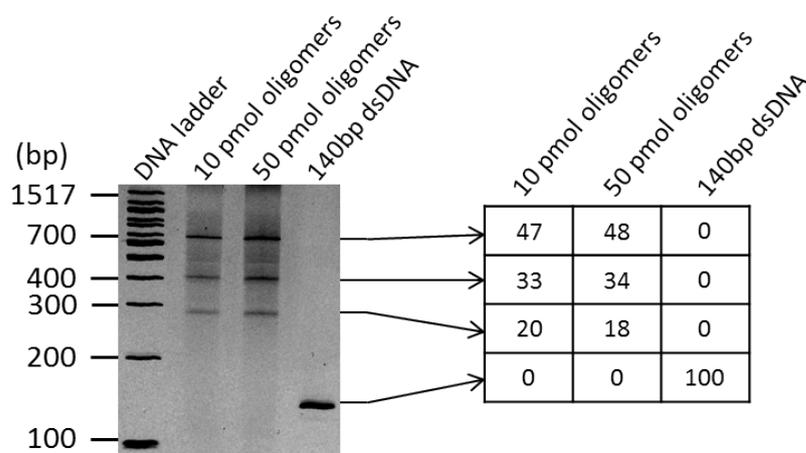


Figure 5.10: **Left:** Native PAGE analysis shows the products of cycling ligation for two concentrations of the four oligomers added to the reaction. In both cases three bands of intensity are shown following the ligation reaction. **Right:** Densitometry analysis shows the proportion of the total intensity in a lane that each band represents. It can be seen that there is not an appreciable increase in ligation efficiency associated with an increase in oligomer concentration. There is an increase in the absolute amount of product as it represents the same proportion of a more concentrated pool of DNA.

5.1.5 Purification of the DNA constructs

Following the cycling ligation reaction used to create the 200 base structure, three species of DNA were observed. In order to optimise patterning experiments, it was essential to remove the non-product DNA species. The oligomers used in this reaction were 99 or 94 bases long; the majority of unligated oligomers were present as oligomer dimers. The Roche High Pure PCR Product Purification Kit removes DNA smaller than 100bp and was used for this reason in an attempt to remove unwanted products. However, the requisite purification could not be achieved through the use of the Roche High Pure PCR Purification kit as the oligomer dimers created in the ligation reaction were purified with the ligation products, as shown in Figure 5.11.

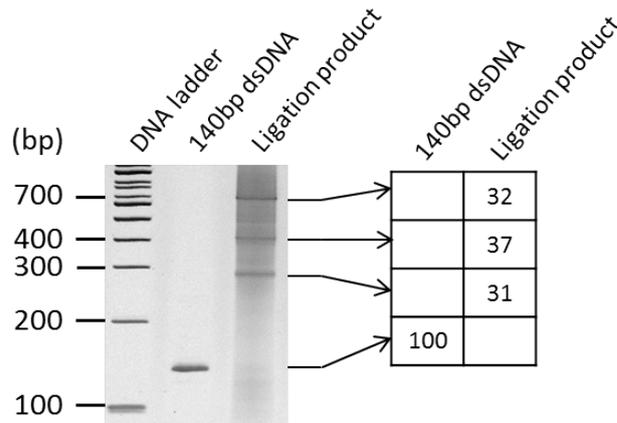


Figure 5.11: **Left:** Native PAGE analysis shows cycling ligation products following attempted purification with the Roche High Pure PCR Purification Kit. Three DNA species can be observed, as seen for unpurified ligation reactions. **Right:** Densitometry analysis of the gel shows that the oligomer dimers represent approximately two third of the DNA species in the sample.

Following the Roche kit purification it is clear that the oligomer dimers remain a significant proportion of the sample. The kit is designed to remove primers and short strands of unfinished PCR, therefore it is not unexpected that it would fail to remove the long, synthetic oligomers used in this reaction.

In a following experiment purification of the ligation products using native gel electrophoresis was investigated. In Figure 5.12 a single lane is shown in which the four bands that appear on the gel can be most easily identified, labelled A, B, C and D. The three bands with the slowest mobility (A, B and C) were extracted and gel purified to try to gain pure ligation product.

The products purified from bands A, B and C in Figure 5.12 can be seen in Figure 5.13.

A number of bands appear on the gel that are made of mixed populations and the majority of the 200 bp DNA co-migrates with the oligomer dimers. The ligation product and oligomer dimers migrate together in bands B and C of the agarose gel shown in Figure 5.12. The highest concentration of ligation product is within the mixed population found in band C, along with higher concentrations of the oligomer dimers. The upper band (A)

contains mostly ligation product, but it is at a much lower concentration.

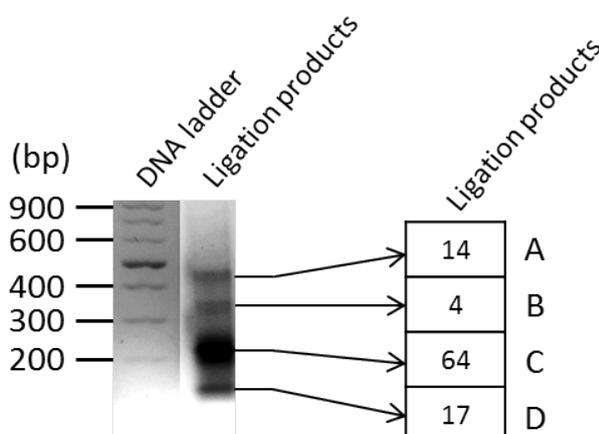


Figure 5.12: **Left:** Cycling ligation products run on a native agarose gel. **Right:** Densitometry analysis shows the proportion of the total intensity in a lane that each band represents. There are four bands visible; it was assumed that the lowest band would not represent the ligation product and the top three bands were cut out and purified using the Roche kit.

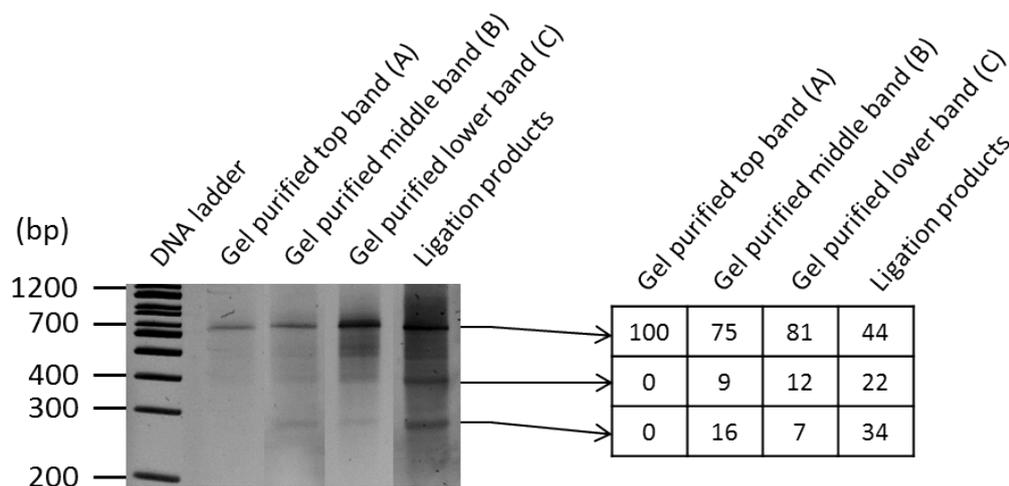


Figure 5.13: **Left:** Native PAGE analysis of the gel purified DNA from bands A, B and C of Figure 5.12. **Right:** Densitometry analysis shows the proportion of the total intensity in a lane that each band represents.

The gel purified 200 bp product from the least contaminated sample, band A, had a

concentration of approximately 1.6 ng/ μ l. The use of gel purification to isolate the 200 bp DNA is hindered by the co-migration of these DNA populations in agarose; it is not possible to gain an effective yield of product without contamination.

Denaturing agarose electrophoresis separates DNA on the basis of the number of nucleotides in the chain. Alkaline agarose gel electrophoresis denatures DNA through a high pH; the concentration of hydrogen ions decreases, disrupting hydrogen bonding between bases. An experiment was carried out to establish if this method would enable the separation of 200bp ligation product and 94 or 99 base oligomers. The ligation products would then be purified from the gel and allowed to reanneal. The 200 bp LCR products were run on a denaturing agarose gel and stained as described in Section 3.1.4. An image of the gel is shown in Figure 5.14.

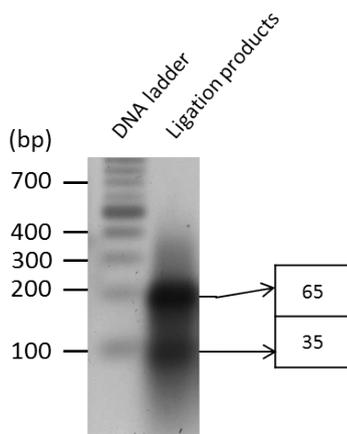


Figure 5.14: **Left:** The 200bp LCR product run on a denaturing agarose gel. **Right:** Densitometry analysis shows the proportion of the total intensity in a lane that each band represents. Two bands are partially resolved at approximately 100bp and 200bp.

Although the bands have not completely resolved, it is possible to see two distinct areas of fluorescence on the gel. The mobility of these two bands appears to be at \sim 100 and \sim 200 bases, indicating the separation of the ligated and unligated DNA strands. The top band was excised and then purified according to the protocol in Section 3.1.3. The results are shown in Figure 5.15.

It can be seen in Figure 5.15 that the purification protocol has been successful in purifying three distinct DNA species. The majority of the oligomer dimers have been removed, and

a yield of up to 17 ng/ μ l was achieved.

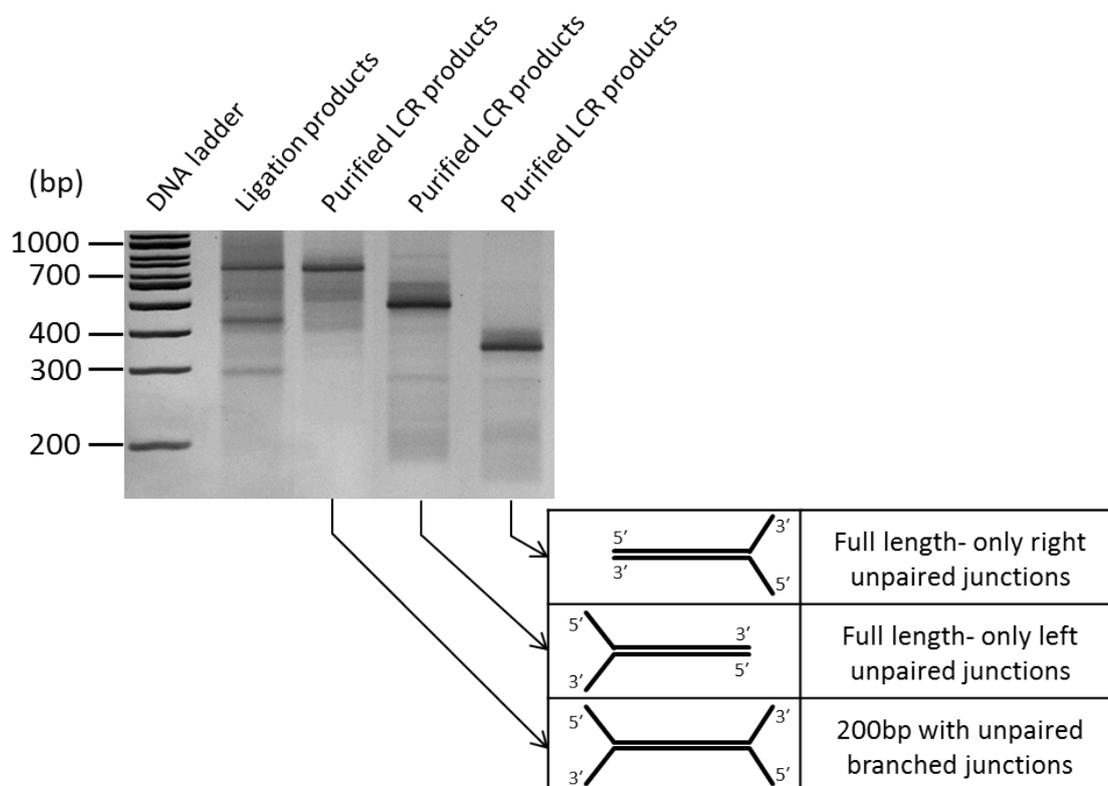


Figure 5.15: **Left:** Native PAGE analysis shows LCR products and the agarose gel purified products. **Lower Right:** The different populations on the gel are identified.

5.1.6 Heat stability

Having produced the desired structures and purified the reaction products, further tests were carried out on the DNA structures to assess the heat stability of our structures in the presence and absence of magnesium ions. These experiments were carried out to ensure that the ligation product would be stable during the hybridisation program, for the creation of three-way junctions at the termini of the structures. The results are shown in Figures 5.16 and 5.17.

Heating the construct with both ssDNA termini up to 80 °C causes some change in the structure that leads to the appearance of a secondary band on the acrylamide gel that was not present before heating. In the absence of magnesium this occurs at temperatures as low as 60 °C. The presence of magnesium in the buffer increases the temperature the

DNA can be heated to without unwanted effects to 70 °C. An experiment using a smaller ranger of temperatures in the presence of magnesium ions is shown in Figure 5.17.

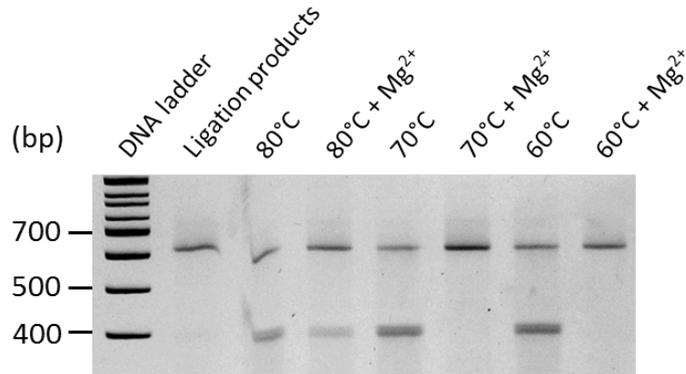


Figure 5.16: Native PAGE analysis of the effect of magnesium concentration and temperature on the 200bp LCR product. Upon heating to 80 °C - 60 °C in the absence of a magnesium-containing buffer an additional band at approximately 200bp appears. When a magnesium-containing buffer is added the DNA can be heated to 70 °C without the secondary band appearing.

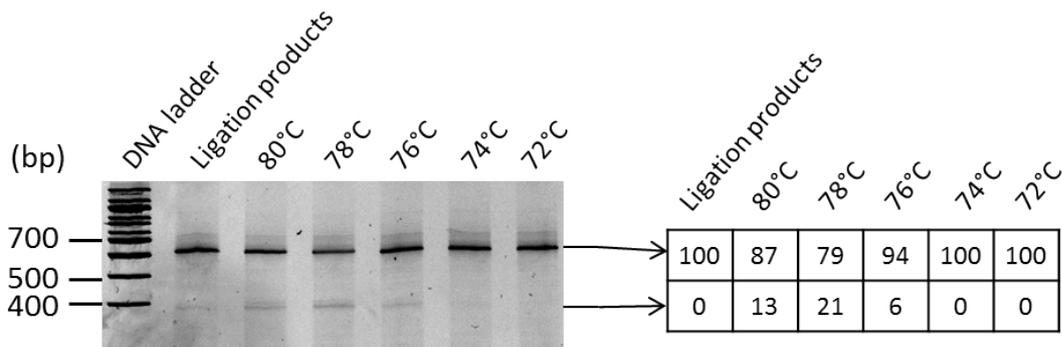


Figure 5.17: **Left:** Native PAGE analysis of the effect of a range of temperatures on the 200 bp LCR product. **Right:** Densitometry analysis shows the proportion of the total intensity in a lane that each band represents. Temperatures of 80 °C - 76 °C cause the appearance of a secondary band at approximately 300bp.

In Figure 5.17 it can be seen that temperatures above 74 °C are sufficient to cause structural changes, even in the presence of 10 mM magnesium acetate. Following this

work hybridisation experiments were carried out at 72 °C to ensure that the structures were not compromised.

5.2 Junction Annealing

The 200 base structures were designed to be fabricated with non-base paired regions at the termini, which could be made into three-way junctions. This is achieved through annealing an oligomer to the unpaired branched junctions, as shown in Figure 5.2. DNA flexibility at three-way junctions should be reduced, simulating more complex DNA structures.

The DNA structures described here, fabricated with unpaired branched junctions, were heated and then cooled in the presence of complementary ssDNA oligomers at a 2:1 ratio, which should anneal with the uncapped ends resulting in short three-way dsDNA branched junctions.

Having identified suitable buffer conditions and temperature, the hybridisation reaction was carried out. To account for the possibility of strain at the junction point in the hybridisation oligomers, a number of sequences were used in which there were a number of bases at the centre of the sequence additional to those needed for base pairing between the strands.

5.2.1 Left (5') annealing

Figure 5.18 shows the hybridisation of the oligomer to the structure with an unpaired branched junction on the left (5') side of the central region. As demonstrated in Figure 5.15, the structure with a single junction at the 5' end of the strand shows a mobility of ~500 in the gel.

Oligomers containing two, three and four additional bases were tested, as well as the sequence exactly matching the two junction strands. It appears that the highest efficiency is achieved when the hybridisation oligomer has no additional bases. In Figure 5.19 the same experiment was carried out on the structure with the unpaired branched junction on

the right (3') side of the structure. As demonstrated in Figure 5.15, the structure with a single junction at the 5' end of the strand shows a mobility of ~ 400 in the gel.

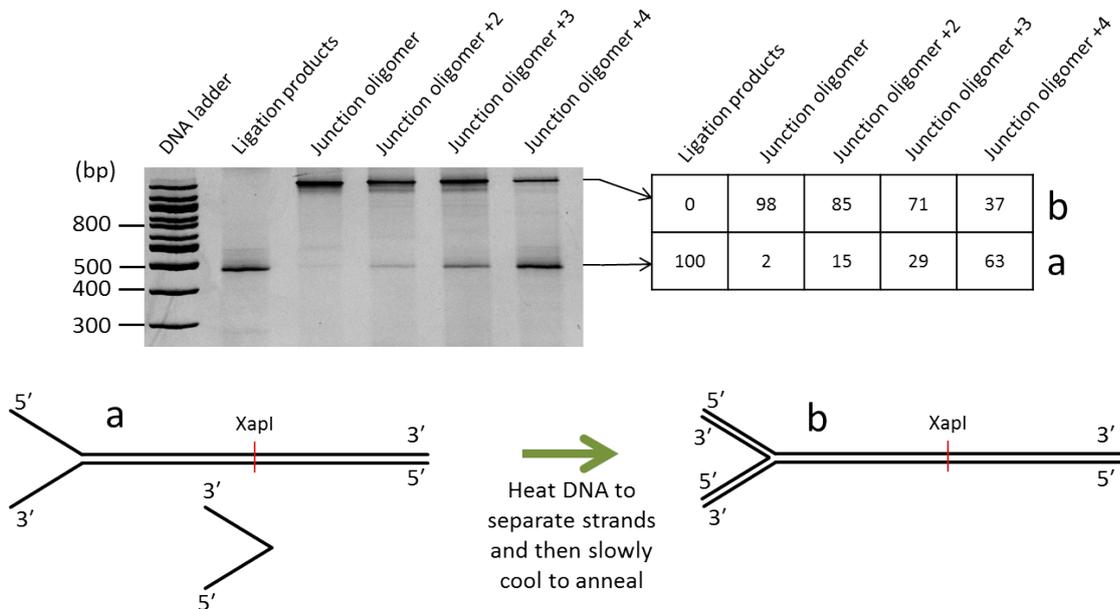


Figure 5.18: **Left:** Native PAGE analysis of the effect of hybridising an oligomer to the unpaired branched junction of the structure with a left (5') junction. **Right:** Densitometry analysis shows the proportion of the total intensity in a lane that each band represents. Labels **a** and **b** relate to species shown diagrammatically below. **Below:** A diagrammatic representation of the annealing of an additional oligomer to the ligation product (**a**) to produce a double stranded three-way junction (**b**).

5.2.2 Right (3') annealing

The results shown in Figure 5.19 indicate that, like the structure with the 5' junction, the structure with the 3' junction shows the highest efficiency when annealed to the oligomer for which the sequences of ssDNA and annealing oligomer match exactly.

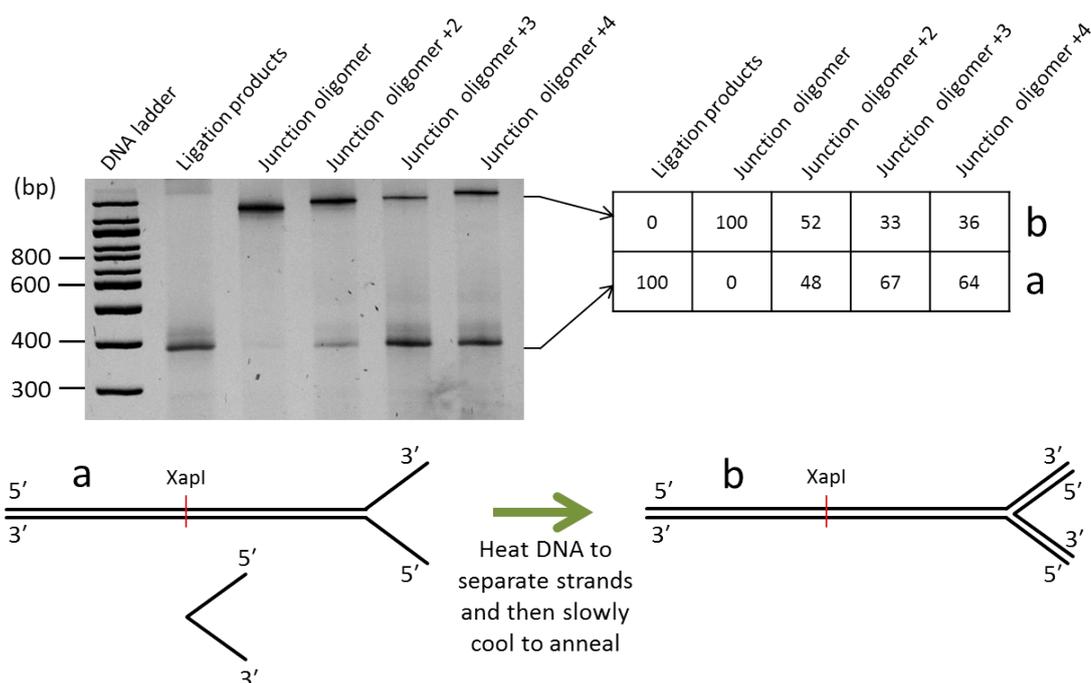


Figure 5.19: **Left:** Native PAGE analysis of the effect of hybridising an oligomer to the unpaired branched junction of the structure with a right (3') junction. **Right:** Densitometry analysis shows the proportion of the total intensity in a lane that each band represents. Labels **a** and **b** relate to species shown diagrammatically below. **Below:** A diagrammatic representation of the annealing of an additional oligomer to the ligation product (**a**) to produce a double stranded three-way junction (**b**).

5.2.3 Annealing oligomers to the left (5') and right (3') junctions

An experiment was then carried out on structures that had an unpaired branched junction at both the 5' and 3' ends. The oligomers most efficient in the single junction structures were used together, in an attempt to anneal to both ends. The results can be seen in Figure 5.20.

A single band can be seen on the gel following the binding of the left (5') hybridisation oligomer to the structure. When the right (3') oligomer binds two bands can be seen at a mobility somewhere between the 200bp unpaired structure and the left (5') annealed structure. It is possible that one of these bands is due to secondary structure formation.

However, it is not clear why this feature would form when the right (3') oligomer is annealed but not form on the structure with no annealed oligomers. When both oligomers are used two bands appear; the lower of these seems to correspond to the second band seen for individual hybridisation event when the right (3') oligomer binds.

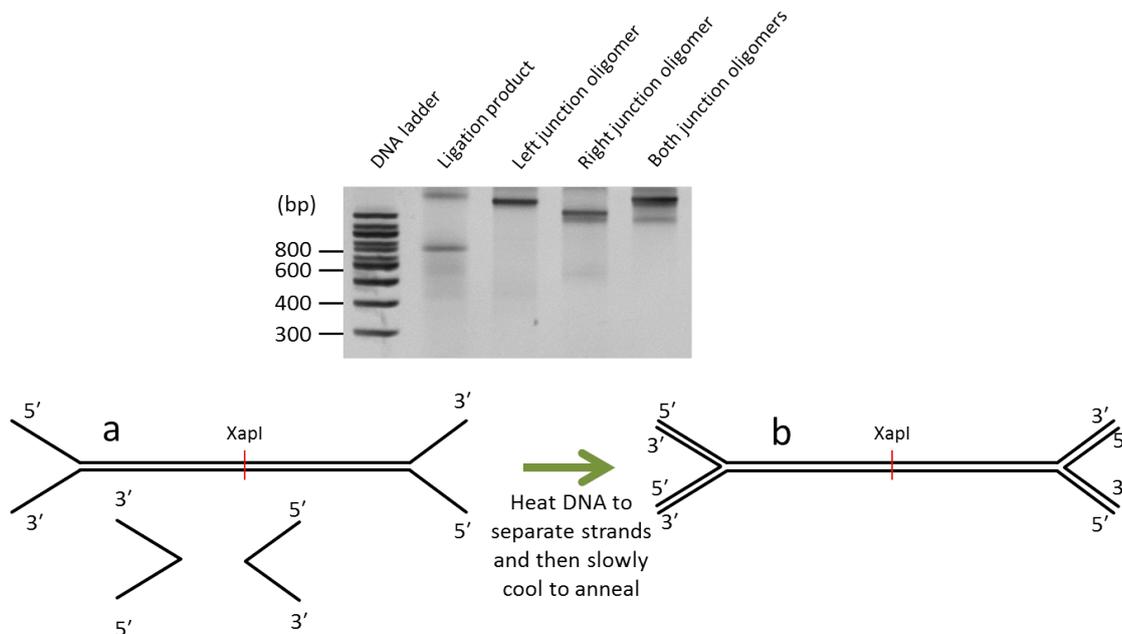


Figure 5.20: **Above:** Native PAGE analysis of the effect of hybridisation an oligomer to the unpaired branched junction of the structure with both left and right (5' and 3') junctions. Labels **a** and **b** relate to species shown diagrammatically below. Poor sample preparation has led to a secondary band in the ligation product lane. **Below:** A diagrammatic representation of the annealing of both additional oligomers to the liagtion product (**a**) to produce double stranded three-way junctions at both termini of the structure (**b**).

It appeared that the hybridisation of the right (3') oligomer was affecting the structure, possibly by creating a DNA species that presented as a second band on the gel. If this were the case it would indicate that the annealing of both oligomers at the same time was only partially successful; complete annealing should remove the second band. An attempt was made to partially redesign the LCR oligomers to see if this would increase the efficiency of junction formation. A diagnostic digest of these samples could have revealed more information about the structures observed on the gel.

5.2.4 LCR with modified oligomers

It was thought that removing bases from the oligomers making up the structure could have an effect on the efficiency of annealing the junctions; this might change the pitch of the helix as the junction oligomers approach. A cycling ligation reaction was carried out using a series of oligomers in which pairs of bases at the right (3') junction point were removed from the sequence. The products were separated by alkaline agarose gel electrophoresis. The band with the slowest mobility was excised and the DNA was purified. The resulting structures are shown in Figure 5.21.

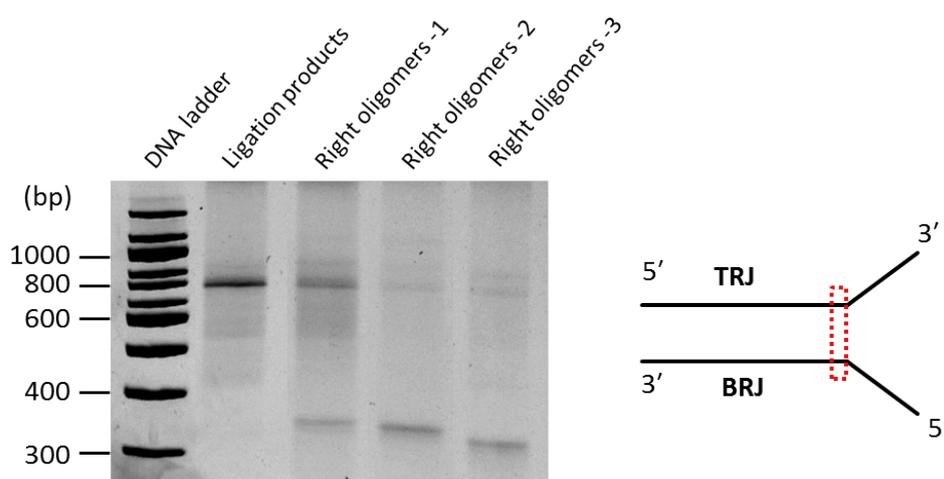


Figure 5.21: **Left:** Native PAGE analysis of an experiment to reduce the length of the ligating oligomers by one to three bases at the junction point of the 3' three-way junction. A cycling ligation reaction was carried out and the products were separated by denaturing agarose electrophoresis. The band with the slowest mobility was excised and the DNA purified. From left to right: the ligation products from the original oligomers, ligation product when a single base is removed from the TRJ and BRJ oligomers, ligation product when two bases are removed from the TRJ and BRJ oligomers and ligation product when three bases are removed from the TRJ and BRJ oligomers. **Right:** A diagrammatic representation of the position at which bases were deleted from the oligomers used in this experiment. The red box highlights the area proximal to the junction point at which bases were deleted.

In comparison to the full length structure the shortened oligomers produce little to no

product. It appears that the oligomers fail to ligate, further iterations of this experiment produced similar results.

5.3 Patterning

Patterning of the structure containing two unpaired branched junctions was carried out to test the hypothesis that unpaired branched junctions increase the efficiency of patterning for dsDNA structures. It was thought that transient breaking of the hydrogen bonds at the termini of the dsDNA region, known as 'breathing' [253], would be increased by the presence of ssDNA regions. An increase in patterning efficiency was observed with proximal mismatches and nicks in the phosphate backbone of the 100 base structures in Chapter 4. The patterning success of these structures may be linked to an increase in DNA breathing at these topologies.

5.3.1 Digestion enzyme trials

In order to use an appropriate amount of enzyme to assess patterning efficiency, trials were carried out to find the minimum level of enzyme required to digest unpatterned dsDNA completely. This enzyme concentration could then be used in the patterning experiments to give a more accurate representation of the effect of these experiments on the efficiency of patterning. The amount of enzyme required for digestion was calculated using the activity of the enzyme given by the supplier, shown below. Enzyme suppliers are listed in Section 3.3.

An experiment was carried out using XapI, a restriction enzyme from *Xylophilus ampelinus*. The digestions were carried out on 623 bp dsDNA from the Lambda genome, at 37 °C for 5 minutes.

The manufacturer states that one unit of XapI enzyme is the amount required to digest 1 μg of lambda DNA in 1 hour at 37 °C in 50 μl of recommended reaction buffer. The Lambda genome is 48.5 kbp and contains 58 XapI sites. The enzyme is supplied at a concentration of 10 units/ μl .

$(1 \mu\text{g Lambda} = 0.032 \text{ pmol}) \times 58 \text{ sites} = 1.86 \text{ pmol}$ of ligation per unit of enzyme

$1.86 \text{ pmol} \times 10 \text{ units} = 18.56 \text{ pmol}$ per μl of enzyme. Therefore a 1X digest requires $1 \mu\text{l}$ of enzyme for every 18.56 pmol of DNA containing a single restriction site.

The results of the experiment are shown in Figure 5.22. The XapI digest of the 623 base dsDNA produces a break at the centre of the strand, resulting in two ~ 300 base strands with four base ssDNA overhangs.

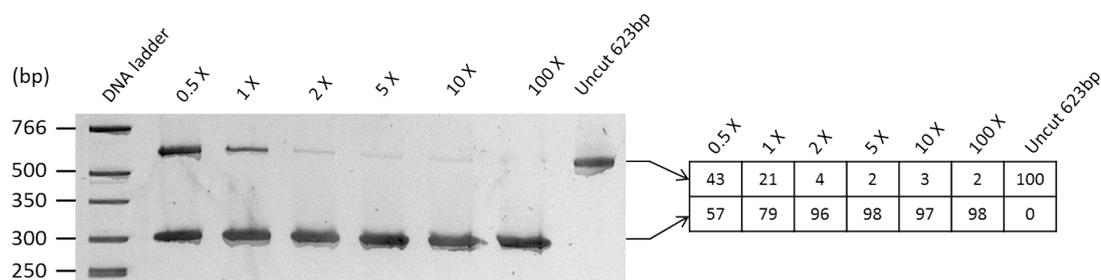


Figure 5.22: **Left:** Native PAGE analysis of the amount of XapI required to digest bare 623 bp dsDNA. **Right:** Densitometry analysis shows the proportion of the total intensity in a lane that each band represents. Increasing the enzyme concentration increases the proportion of the DNA that is digested. The XapI site of the 623 bp DNA is at the centre of the sequence so the DNA digests into two bands of approximately equal length.

The correct digestion of this DNA strand appears to have taken place, producing two strands with the same migration of ~ 300 bases.

5.3.2 Patterning

Patterning was carried out at a 30:1 ratio as described in Section 3.2.7.

Preliminary data, shown in Figure 5.23, indicates that patterning efficiency increases in the presence of unpaired branched junctions.

The 200bp DNA digests into two bands that show the same mobility as the oligomer dimers. The 140bp DNA digests into two 70bp dsDNA strands that show the same mobility, therefore only a single band of DNA can be seen following digestion. The

control experiment, using unpatterned DNA, followed the same protocol as the patterning experiment but the RecA used was heat denatured. The small amount of undigested DNA in the unpatterned samples might be due to a small population of RecA remaining active despite the heat treatment.

Patterning of the 140bp DNA decreased the level of DNA digestion from approximately 91% to 82%. This indicates a relatively low level of protection was afforded by RecA patterning. When 200bp DNA with unpaired branched junctions was patterned with RecA 58% of the DNA was digested. This is a large decrease from the 97% digested DNA for the unpatterned 200 bp sample.

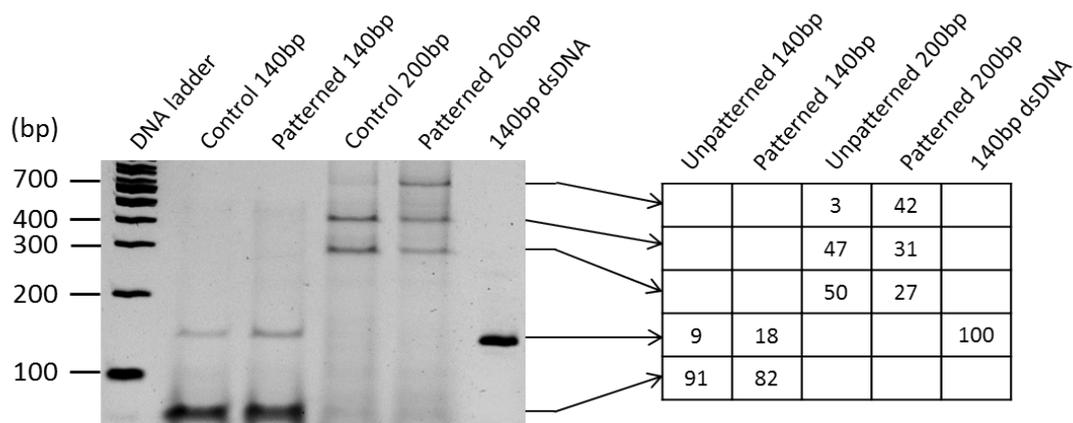


Figure 5.23: The image on the left shows native PAGE analysis of the patterning of the 140 and 200bp DNA structures using a 60 base ssDNA and RecA. In the control lane heat treated RecA was used in the patterning experiment to demonstrate the level of digestion that occurred in the absence of patterning. Densitometry analysis on the right shows the proportion of the total intensity in a lane that each band represents.

Previous results suggested that unpaired branched junctions increase patterning efficiency (unpublished report, Faith Bateman). The results shown here indicate that those findings were significant and that RecA does pattern DNA structures containing unpaired branched junctions with higher efficiency.

5.4 Conclusions

The structures discussed in this chapter were developed following interesting results from a 100 base structure containing base pair mismatches within the sequence. The structures were designed to be synthesised using a cycling ligation reaction and contain terminal base pair mismatched that could be annealed to additional oligomers for the creation of three-way junctions.

A scaffold to bring together the oligomers was produced through PCR using the Lambda genome as a template. The cycling ligation reaction was carried out and analysis identified the products of the reaction. LCR did not incorporate all the oligomers present in the reaction volume; the product of the reaction was contaminated with ~ 100 base oligomers. The yield was improved through a series of optimisation steps. A standard kit used for purification of PCR products proved insufficient to remove the ssDNA contaminants. Purification of the product was possible through the use of alkaline agarose gel electrophoresis which separated ssDNA strands on the basis of nucleotide length.

Three structures were created containing ssDNA elements at one or both ends of the central region. Attempts were made to anneal an additional oligomer to create three-way junctions. It was possible to anneal a single oligomer to structures with a single ssDNA region; the highest yield was observed when the annealing oligomer contained exactly the number bases necessary to base pair to the ssDNA strands present. This observation was repeated for the structure with two ssDNA termini. However, it did not prove possible to anneal both additional strands concurrently. Use of the oligomers containing additional bases at the junction point did not improve the yield of dsDNA with two terminal three-way junctions.

To facilitate the production of the structure with two terminal three-way junctions, experiments were undertaken in which the base-pairing region of the structure was altered. Oligomers for ligation were synthesised without a number of the base pairs at the junction point, to test the hypothesis that the annealing of a single oligomer was affecting the helical pitch of the junction point at the opposing end. These oligomers did not appear to produce ligated products that could be gel purified in the same way as the original design.

In light of the failure to produce the desired structures, a single patterning experiment was carried out to gain some preliminary data on the effect of terminal ssDNA regions. The experiment indicated that the presence of the ssDNA regions patterning efficiency was higher than for the fully base-paired 140 base dsDNA. The results showed patterning at a much lower level than was observed for the 100 base structures; in the presence of a 20 base mismatch the 100 base structures consistently showed protection of 100%.

It is difficult to draw too many conclusions from such a small dataset. The challenges in producing the structures containing dsDNA three-way junctions indicate that the base pairing tension around the junction points may be complex. This tension within a complex structure would have an effect on patterning efficiency in a way that might be difficult to predict. Unforeseen design flaws made this structure unsuitable for pursuing RecA patterning and for this reason work with different structures was pursued.

Chapter 6

Fabrication of 270 base structures for RecA patterning

Despite attempts to create a ~ 200 base structure for patterning with RecA, very little patterning data was created with these constructs. A different method was attempted for making DNA structures with non-base pairing regions, to create ssDNA strands that could then be annealed into a 270 base structure. This structure resembled the ~ 200 base structure in the presence of non-base pairing regions at the termini that could be annealed to form three-way junctions.

The 200 base structures were made through a cycling ligation reaction. To remove the need for multiple phosphorylated oligomers and ligase, a new structure was designed using a synthetic gene which acted as a PCR template. When used in conjunction with a phosphorothioated primer it was possible to create single strands of DNA with the required sequence that could be annealed to produce the desired product, shown in Figure 6.1. It was hoped that by creating the structures this way the yield and purity of the product would be increased.

This structure was large enough to be observed using atomic force microscopy (AFM), shown in Figure 6.10. The presence of three-way junctions was confirmed using peak force tapping. Following successful synthesis, patterning experiments were carried out but were not successful.

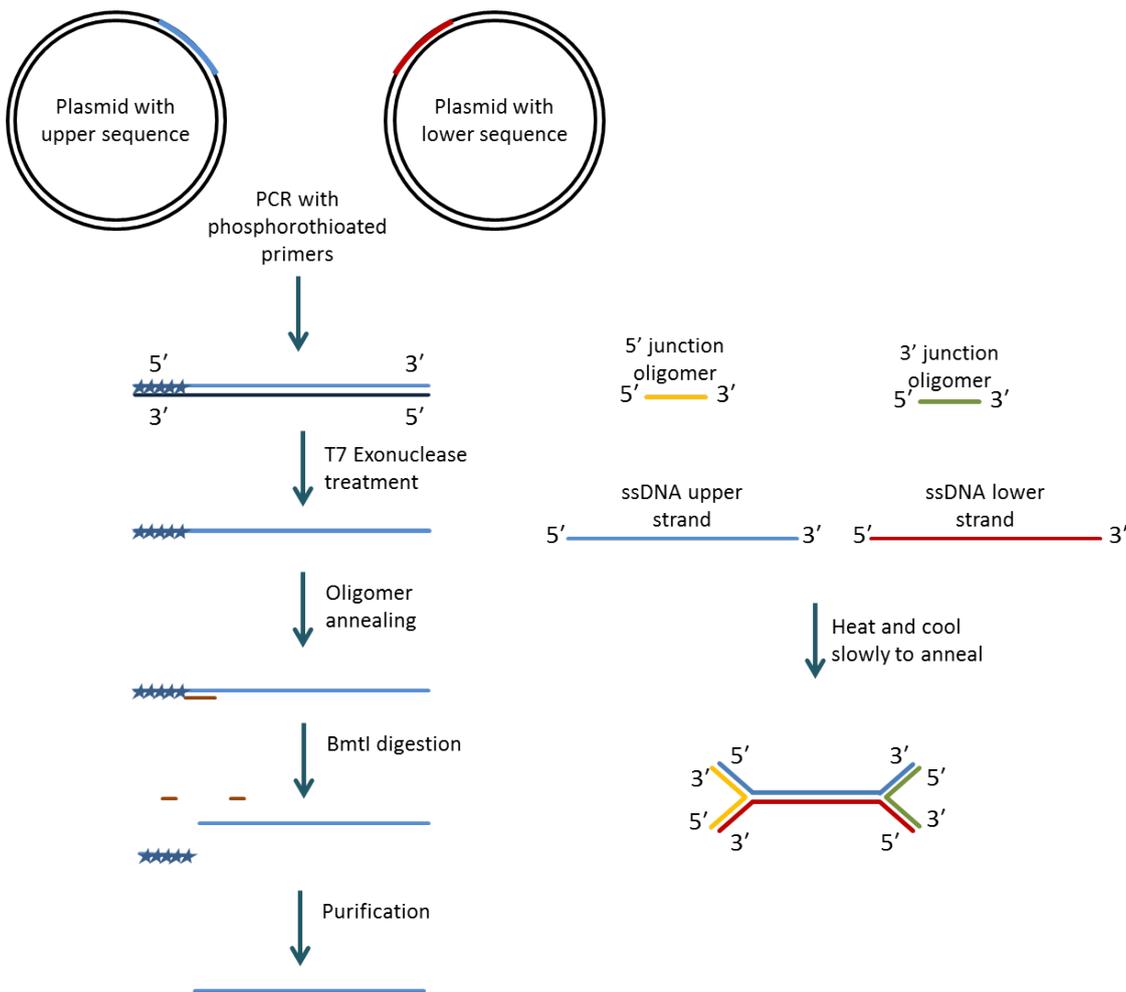


Figure 6.1: Diagrammatic representation of the process for producing 270 base structures with three-way junctions at the termini. The sequences were synthesised in plasmids which were used as the template for PCR reactions. The primer oligomer for the desired strand had 5 phosphorothioated bases at the 5' end, to protect it from the T7 exonuclease step. When the non-phosphorothioated DNA had been digested, a short oligomer was annealed to the ssDNA to produce a dsDNA restriction site. This site was used to remove the phosphorothioation from the 5' end of the strand, the enzyme and short strands of DNA were then purified away. The two ssDNA strands were annealed to produce four structures, depending on the presence of the junction-creating oligomers. The structure shown here uses both junction-creating oligomers and therefore has three-way junctions at both the 5' and 3' termini. The plasmids are shown in black, the upper strand in blue and the lower strand in red. The 5' junction oligomer is shown in yellow and the 3' junction oligomer in green. The short digestion oligomer is shown in brown.

6.1 Synthesis

6.1.1 Design

Following work on the 100 and 200 base structures, it was decided to make a larger construct that would be more easily observable through AFM. This would allow confirmation of the correctly formed structure in addition to gel electrophoresis. The sequence of the central 200 bases of the 270 base structure is the same as the 200 base sequence and 100 base sequence. In this manner all three structures had a degree of sequence homology. A series of reactions were carried out to produce ssDNA strands that could be annealed to produce the desired structures, shown in Figure 6.1. The sequences required for the upper and lower strand were ordered as synthetic genes in a pair of pMA-T plasmids. Sequencing data from the supplier confirmed the presence of the correctly synthesised sequences. From these templates a PCR reaction was carried out with phosphorothioated primers.

T7 exonuclease is a DNA digestion enzyme, it works by catalysing the removal of bases individually from the 5' terminus of the DNA strand. The exonuclease will not act on a row of 5 or more phosphorothioated bases so these were included in the primer for the desired strand of each PCR product. This allows the selective removal of one strand of the PCR product to produce the desired ssDNA strand.

The primers for the desired sequences were designed to contain five phosphorothioated bases and a linker sequence followed by a recognition sequence for the restriction enzyme BmtI. These bases were designed to occur at the 5' end of the upper strand and added 19 bases to the sequence, resulting in a 289 base PCR product.

BmtI was chosen for a number of reasons; it is not present at any other point in the plasmids used and is available as a high fidelity enzyme, reducing the chance of undesired restriction. An important factor in choosing BmtI was that the digest leaves only a single base of the recognition site on the upper (5' to 3') strand, thus only a single base would be added to the sequence we had designed.

Treatment of the double stranded product with T7 exonuclease, described by Tosch *et*

al. [243] and outlined in Section 3.2.2, yielded ssDNA oligomers of the desired length. These were then annealed with additional oligomers to produce structures with three-way junctions at their termini. The sequences are listed in Appendix A.3.

6.1.2 Gene synthesis and DNA fabrication

The desired sequences were ordered as synthetic genes within plasmids. Separate plasmids were used for the upper and lower sequences of the 270 base structure. This was necessary for the creation of the non-base pairing regions of the structure. The bases used for the phosphorothioation and the linker sequence were taken directly from the construct used by Tosch *et al.* [243]. The number and position of the phosphorothioated primers was essential to prevent T7 Exonuclease digestion of the required DNA strands. The primer for the complement strand contained no base modifications.

The PCR programs used to amplify the desired sequence from the synthetic gene are described in Section 3.2.2.

The 170 base central sequence was synthesised in a separate PCR reaction, the resulting DNA species is shown in Figure 6.2. A single species of DNA is observed with mobility lower than that of the 200 base band of the ladder, indicating that the correct product is present.

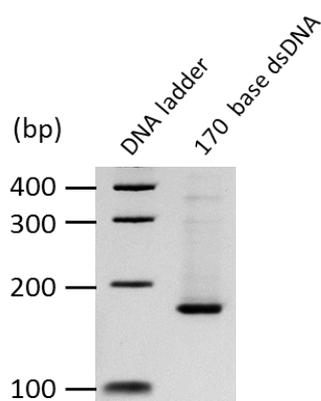


Figure 6.2: Native PAGE analysis of the purified PCR product of the reaction to produce the central 170 bases of the 270 base structure. A single species of DNA is observed with mobility indicating that the correct product is present.

The purified products from PCR reactions to create the two 289 base phosphorothioated dsDNA strands are shown in Figure 6.3. The PCR products for the upper and lower strand both show a single band of intensity can be observed around 300 base band of the DNA ladder. The dsDNA strand created should be 289 bases long; the mobility of the dsDNA strands may be slightly reduced by the phosphorothioation of five bases at the 5' terminus of on strand.

Sufficient yields of the two 289 base sequences were achieved to allow T7 Exonuclease treatment and subsequent purification and digestion.

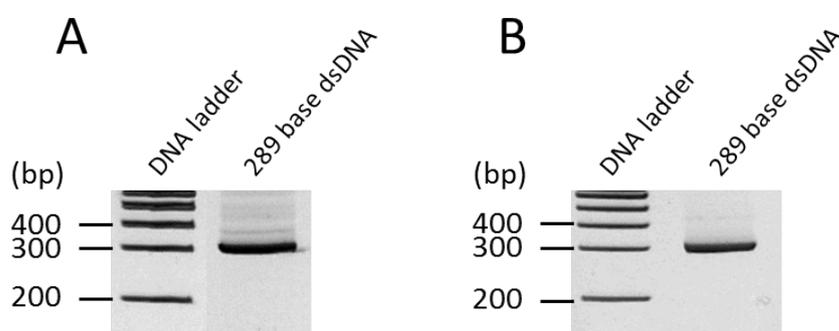


Figure 6.3: Native PAGE analysis of purified PCR products from the reactions to produce the 289 bp precursors to the 270 base structure. **A:** The 'upper' strand 289 base phosphorothioated dsDNA. **B:** The 'lower' strand 289 base phosphorothioated dsDNA.

Confirming PCR products using a restriction assay

The presence of the correct DNA structures was confirmed through a series of restriction assays using the enzymes XapI, AluI and AlwI. PAGE analysis of these experiments is shown in Figure 6.4, the numbers below each lane indicate the predicted size of the DNA species created by digestion with each enzyme.

The size of the DNA species created through restriction are correct for the position of the sites for the XapI digest. No restriction is observed for the AlwI experiment; no recognition site for this enzyme is present in the upper strand precursor. Two bands of the correct mobility are observed for the AluI digest, as well as an additional band above the ladder maximum of 330 bp. This could be the result of enzyme remaining bound to the DNA, though this seems unlikely given how well defined the band is. It is also possible

that there is a sequence very similar to the recognition sequence and we are observing some off-target digestion. It is not clear why this would lead to a restriction product with a higher mobility than the full-length product as AluI leaves a blunt end following digestion.

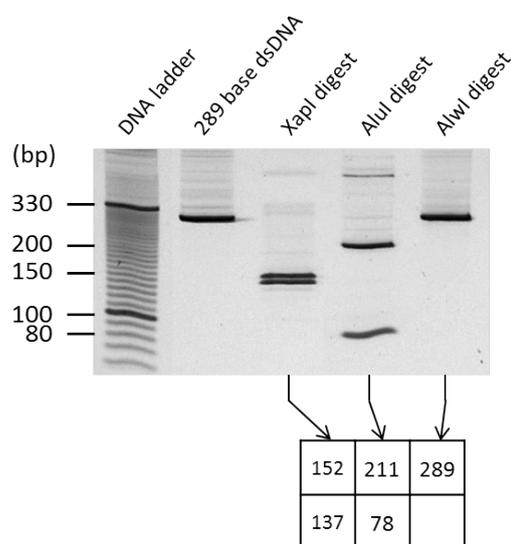


Figure 6.4: Native PAGE analysis of a restriction assay to assess PCR product for the 289 base dsDNA which is the precursor to the upper strand of the 270 base structure. The numbers below each lane indicate the predicted size of the DNA species created by digestion with each enzyme. A single recognition site is present for the enzymes XapI and AluI; no recognition site is present for AlwI.

The same assay was carried out on the 289 base precursor to the 270 base structure 'lower' strand. PAGE analysis of this experiment is shown in Figure 6.5.

The three restriction assays on this dsDNA produce the results expected for the desired 289 base sequence. Recognition sequences are present for all three enzymes and two bands with the correct mobility are seen in all three lanes. There is some undigested dsDNA present for the AluI and AlwI experiments; it is likely that an extended digestion period would remove this species.

These experiments confirm that the products of the PCR reactions on the synthetic genes are the desired dsDNA strands.

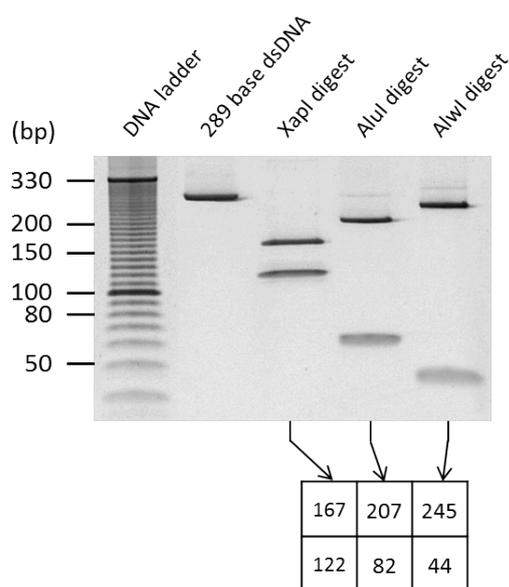


Figure 6.5: Restriction assay confirmation of correct PCR product for the 289 base dsDNA precursor to the lower strand of the 270 base structure. The numbers below each lane indicate the predicted size of the DNA species created by digestion with each enzyme

6.1.3 T7 Exonuclease Treatment

The experimental protocol for creating ssDNA from phosphorothioated dsDNA is described in Section 3.2.2. A diagram of this protocol is shown in Figure 6.6. The dsDNA PCR product was treated with T7 exonuclease, removing the 5' mononucleotides from the non-phosphorothioated strand (Step 1). It is not possible to heat inactivate T7 exonuclease so a purification step was then carried out to remove the mononucleotides and the enzyme. This removal prevented the digestion of the ssDNA product following the removal of the phosphorothioates. The ssDNA was then annealed to short, 21 base oligomers to create a short dsDNA region containing the recognition site for BmtI (Step 2). The DNA was then digested (Step 3) and purified again to remove the short DNA strands and restriction enzyme (Step 4). The ssDNA was purified into water and the concentration assessed. Figure 6.6 shows polyacrylamide gel analysis of the samples produced through the different stages of treatment to produce ssDNA strands with the desired sequence.

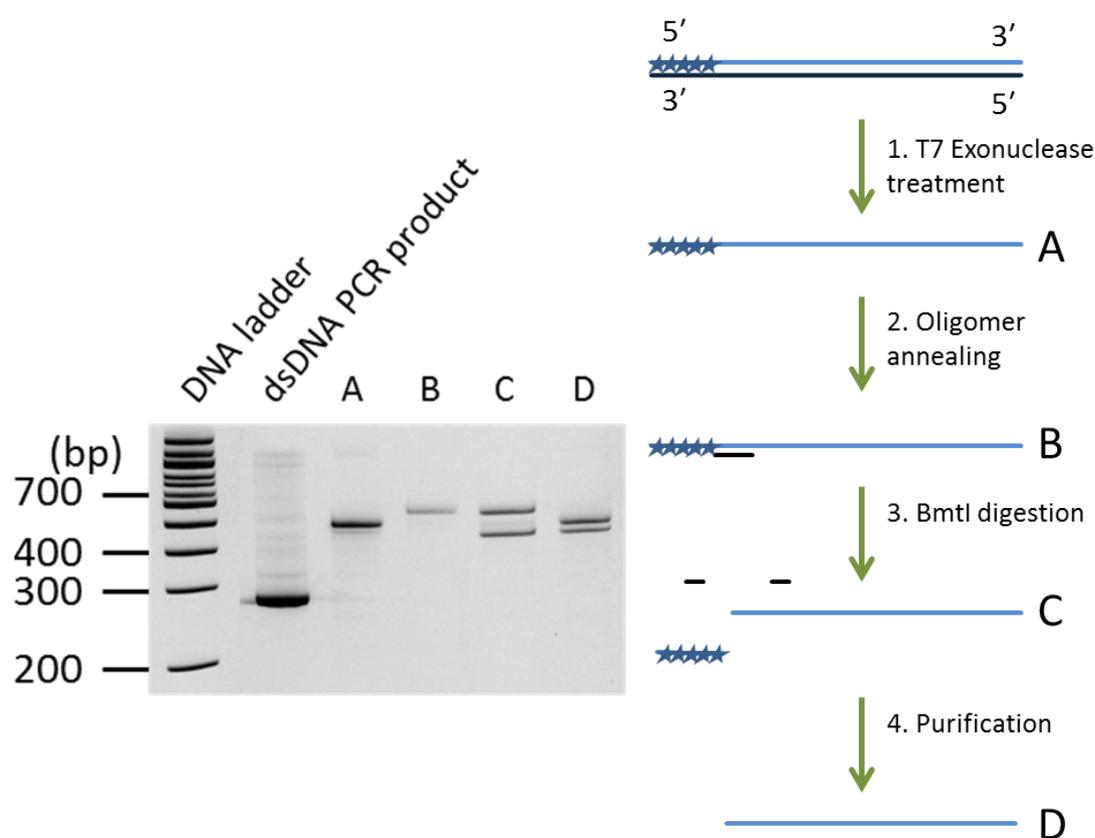


Figure 6.6: **Left:** Native PAGE analysis of the treatment of the treatment of phosphorothioated PCR product to create 270 base ssDNA. The upper strand is shown in this figure, it is also representative of the lower strand. **Right:** Flow diagram showing the treatment of phosphorothioated dsDNA to create 270 base ssDNA. Step 1: T7 exonuclease treatment produces ssDNA (gel sample A), secondary structural elements in the sequence retard the movement of the strand through the acrylamide matrix. Step 2: A 21 base oligomer is annealed to the ssDNA product of the T7 treatment to create a short dsDNA segment (gel sample B). Step 3: The DNA is incubated with the restriction enzyme BmtI. A recognition site for this enzyme lies within the dsDNA region created through oligomer annealing. Following this step two bands are visible on the gel (gel sample C); this is due to secondary structure in the ssDNA. Step 4: The reaction mixture is purified to remove the short ssDNA strands which are a by-product of the digestion, and to remove the restriction enzyme. The ssDNA is purified into water (gel sample D); the change in ionic conditions changes the mobility of the DNA through the acrylamide matrix.

A clear change in sample mobility is seen following the T7 exonuclease treatment as the sample has been digested from dsDNA to ssDNA (Step 1). A smaller change is also visible following the annealing of the 21 base oligomer used to create the short region of dsDNA containing the BmtI restriction site (Step 2). The digestion step (Step 3) removes 18 nucleotides from the 5' end of the strand and results in two strong bands of intensity, one higher and one lower than sample before digestion. There also appears to be a faint band between these two. There is a noticeable change in mobility between the sample directly following digestion (Step 3) and the sample following purification (Step 4) which could be due to the buffer change; magnesium ions are present in the restriction buffer, the DNA is then purified into water. Extremely similar results were observed at all stages for the treatment of the lower strand to produce a partially complementary ssDNA strand.

To identify the bands produced in the purified ssDNA following the BmtI digest (lane C), an experiment was carried out in which the digest step was shortened to 5 minutes instead of an hour. The resulting DNA species were analysed by PAGE, shown in Figure 6.7.

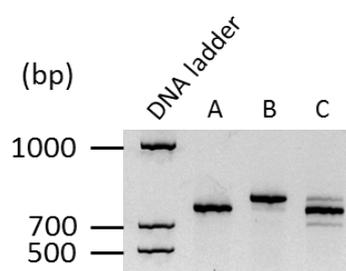


Figure 6.7: Native PAGE analysis of an experiment to identify whether the second band of the ssDNA product was undigested DNA or a second stable secondary structure. Step 1: T7 exonuclease treatment (gel sample A). Step 2: A 21 base oligomer is annealed to the ssDNA (gel sample B). Step 3: The DNA is incubated with the restriction enzyme BmtI for a reduced time period, leading to partial digestion (gel sample C).

The DNA at stages 1 and 2 were treated in the same way as those in Figure 6.6, but at stage 3 the restriction digest was prematurely halted, using Proteinase K, to show a partial digest. In comparison to the samples at stage 3 in Figure 6.6 there are three bands of intensity in stage 3 of this sample, representing the two bands of product (upper and lower bands) and the band of undigested DNA (central band). This indicates that the

double band effect of the ssDNA is not due to incomplete digestion.

The DNA species show different mobilities in Figures 6.6 and 6.7. It seems likely that there is a problem with the ladder for this gel as all of the relative mobilities of the bands remain consistent. Despite this problem we are still able to observe that the halted digestion results in a strong central band; this band can also be observed very faintly at the same position in lane C of the gel in Figure 6.6. These results indicate that the upper and lower bands of lane C are both product bands, demonstrating two stable secondary structures that lead to different gel mobility. These bands were created through the removal of 18 nucleotides from the 5' end of the strand. Analysis using Mfold demonstrated that the 270 base strand can form multiple stable secondary structures with melting temperatures of 51 °C and 50.7 °C. It is also notable that in this step the phosphorothioated bases are removed. The phosphorothioate modification has been shown to potentially alter RNA hairpins [256] and may effect the potential secondary structures of the 5' terminus. Analysing these samples using a denaturing acrylamide gel would have provided more information about the populations observed here.

6.1.4 Annealing three-way junctions

Oligomers with sequence complementarity to the unpaired junctions at the termini of the 270 base structures were annealed to the PCR purified and T7 exonuclease treated ssDNA strands. The two strands were also annealed in the absence of the additional oligomers to create structures with non-base pairing sequences at one or both termini. The two strands were also annealed in the absence of the additional oligomers to create a structure with single stranded elements at both termini. The protocol for annealing is described in Section 3.2.3. The resulting DNA samples are shown in Figure 6.8.

The two ssDNA strands were found to have the same mobility, despite the differences in sequence at both termini. Multiple bands of intensity are present when one or more termini are left as ssDNA regions, this is almost certainly due to the formation of secondary structures. The 5' and 3' capped structures show the same mobility on the gel. When both termini are annealed with additional oligomers to produce three-way junctions a single band can be seen on the gel.

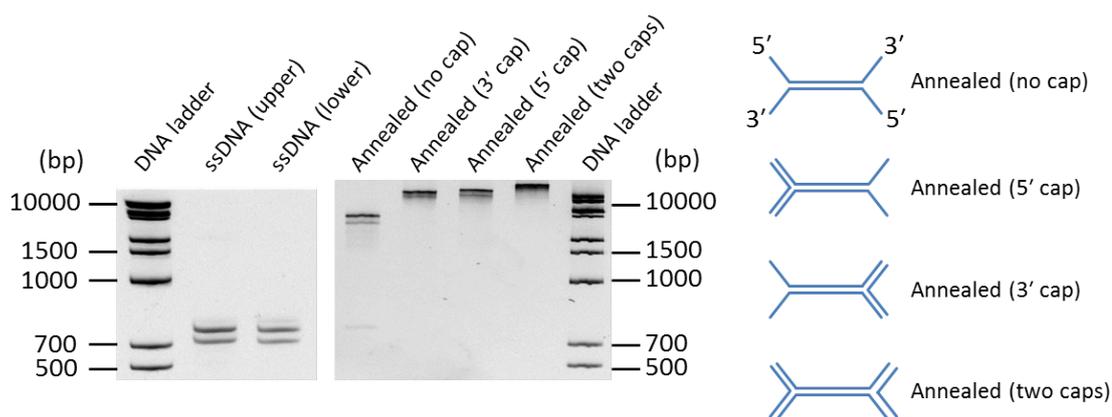


Figure 6.8: On the left, native PAGE analysis of the annealing of ssDNA strands and additional oligomers to produce a range of structures containing three-way junctions and terminal ssDNA elements. On the right, diagrammatic representations of the DNA species produced.

6.2 Atomic Force Microscopy

AFM was used to confirm the presence of the correctly annealed products shown in Figure 6.8. Sample preparation was carried out as described in Section 3.2.6. The resulting images are shown in Figures 6.9 and 6.10, alongside diagrammatic representations of the DNA species.

The structures observed on the polyacrylamide gel can be identified in these images. There is a clear change in the structure at the termini when the additional oligomers are absent or present to create terminal three-way junctions. The ssDNA elements at the termini show a lower height profile than the dsDNA when the additional oligomers have annealed, this is most obvious when a single terminus is double-stranded as in panels B and C. The ssDNA strands are less linear than the annealed dsDNA; the persistence length of ss- and dsDNA are different. The length of the ssDNA strands is not the same as the length of the fully base-paired DNA, it is probable that the ssDNA elements are forming secondary structural elements in the absence of the additional oligomers.

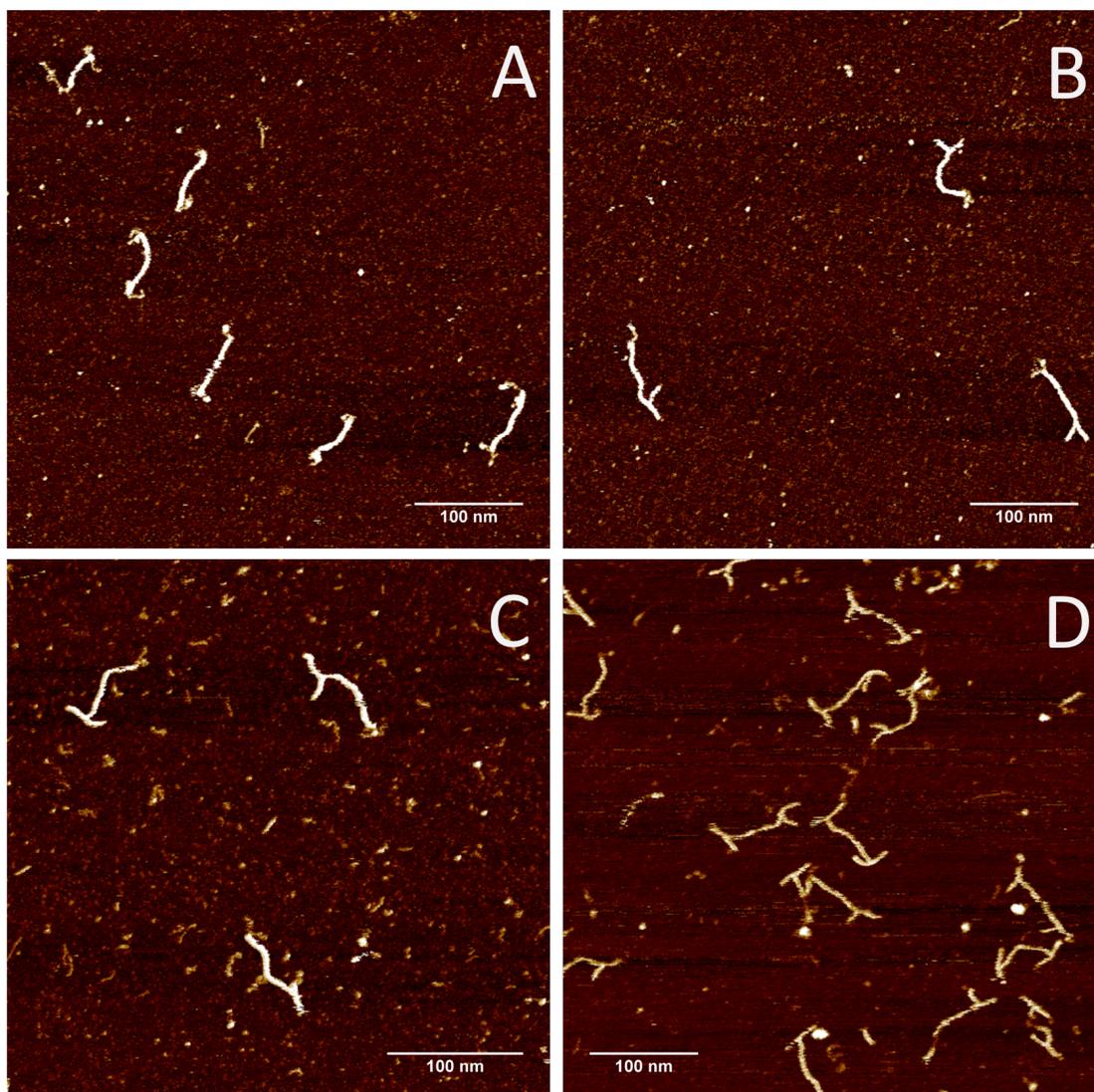


Figure 6.9: AFM images of the 270 base structures obtained using Peak Force Tapping. (A) 270 base structure with no additional oligomers, (B) 270 base structure with 5' capping oligomer, (C) 270 base structure with 3' capping oligomer, (D) 270 base structure with 5' and 3' capping oligomers.

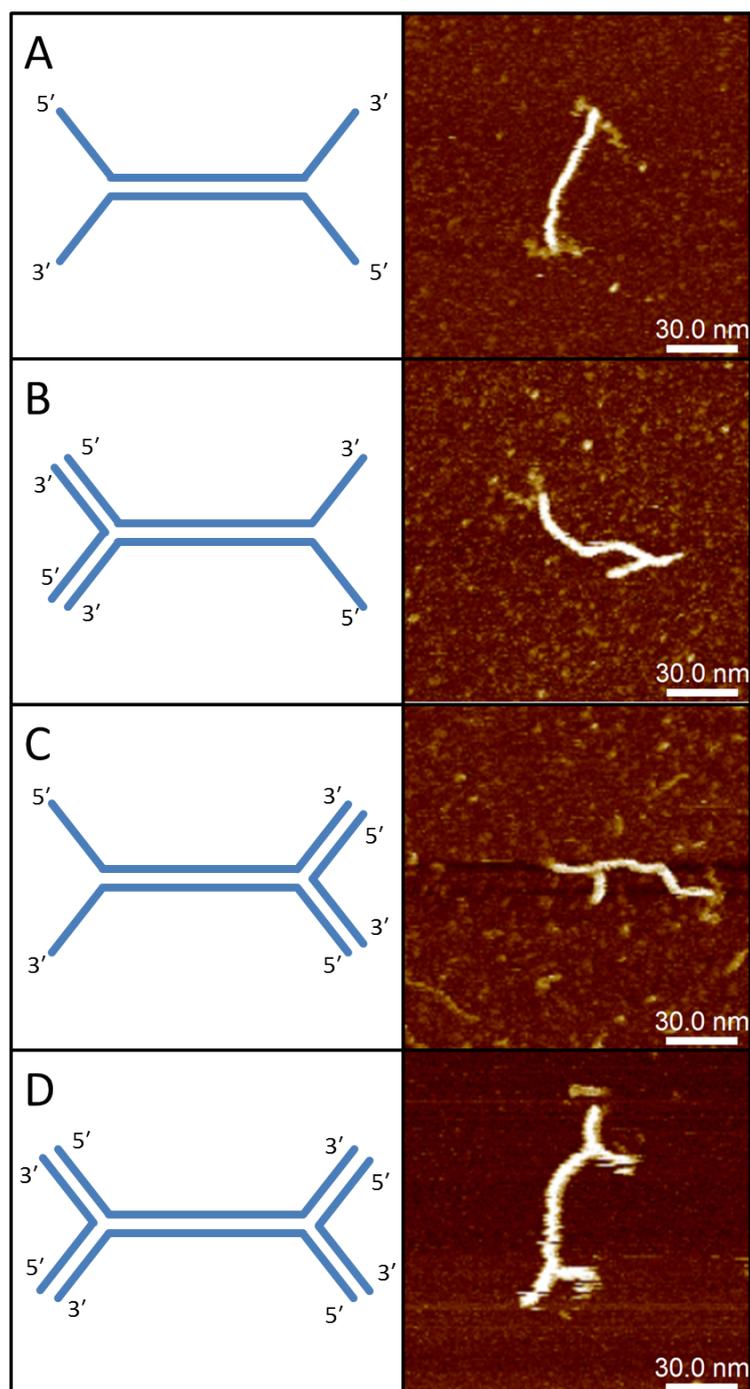


Figure 6.10: On the left, diagrammatic representations of the DNA species. On the right, AFM images of the 270 base structures obtained using Peak Force Tapping. (A) 270 base structure with no additional oligomers, (B) 270 base structure with 5' capping oligomer, (C) 270 base structure with 3' capping oligomer, (D) 270 base structure with 5' and 3' capping oligomers.

These images confirm the formation of the structures inferred from the PAGE analysis, that the additional oligomers are annealing to produce double stranded three-way junctions at the termini of the structures. Further AFM images not shown here were analysed to assess the efficiency of forming the structure with both three-way junctions using this protocol. An efficiency of $75\% \pm 4$ was found using visual assessment of the successful formation of the structures. This method of analysis is not highly rigorous due to its reliance on the orientation of the structures on the surface and on the discretion of the analyst.

6.3 Validating structures through restriction enzyme digest

To characterise the digestion products of the four constructs shown in Figure 6.10 an XapI digest was carried out in the buffer conditions of a patterning experiment. This enzyme has a recognition site from base 118 to 123 of the 270 base sequence. Concurrently all the constructs were placed in the patterning buffer without digestion, to confirm that the buffer conditions did not have a destabilising effect on the constructs. The experiment was analysed by native PAGE, shown in Figure 6.11.

A number of DNA species appear on the gel following digestion of the different constructs. A single XapI recognition site is present in all of these sequences; it seems likely that the different bands are the result of a range of secondary structures adopted following digestion. Analysis using denaturing PAGE would help reveal more information about the DNA species observed here and confirm whether the additional bands observed are the result of secondary structure. The digest is expected to produce two strands of DNA, containing either an ssDNA region or a three-way junction. The restriction products should be 118 bases and 148 bases long, both with a four base ssDNA overhang where the digest took place.

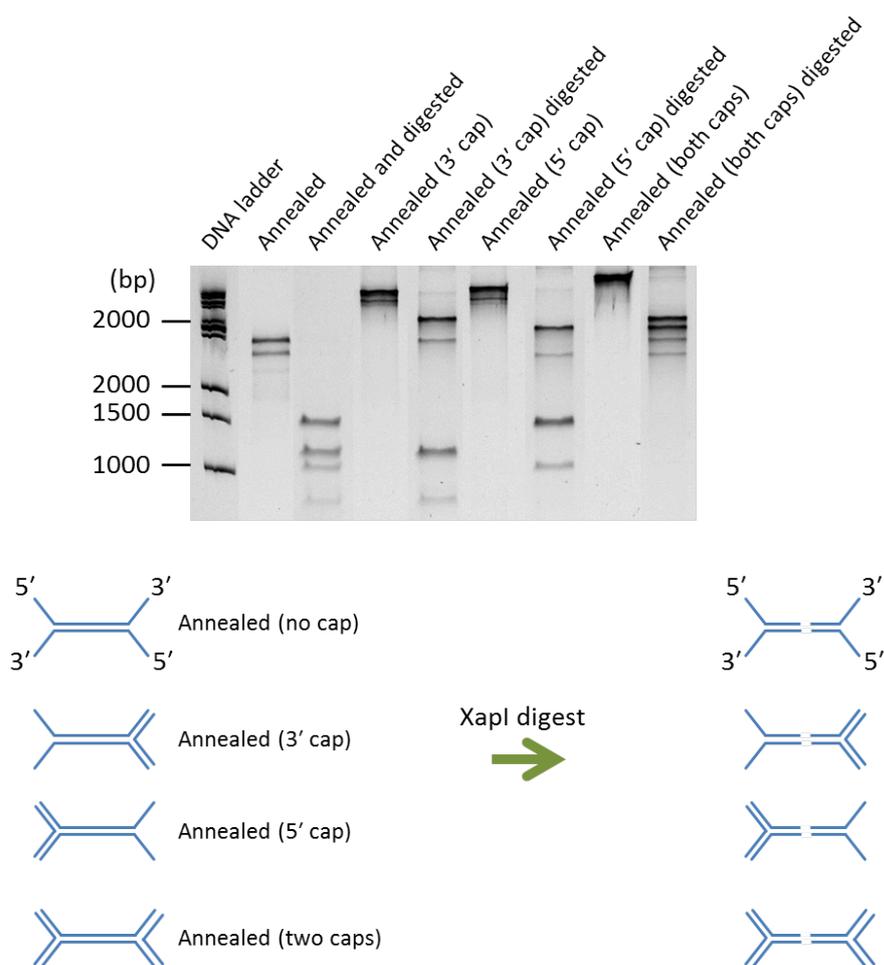


Figure 6.11: **Above:** Native PAGE analysis of the four 270 base constructs in patterning buffer and digested with XapI. Two products are produced when these four constructs are digested; four lanes appear on the gel. **Below:** A diagrammatic representation of the products expected from the XapI digestion of the 270 base constructs containing terminal single stranded or double stranded three-way junctions.

Trials were carried out to find the minimum level of enzyme required to digest unpatterned 170 bp dsDNA completely. This enzyme concentration could then be used in the patterning experiments to give a more accurate representation of the effect of our experiments on the efficiency of patterning. The digestion was carried out using a different number of enzyme units. The digestion efficiency of these enzymes is also discussed in Section 5.3.1. The results are shown in Figure 6.12.

The XapI recognition site is from 68 to base 73 of the 170 bp dsDNA; the products of

the digest should be 68 and 98 bases in length, both with a four base ssDNA overhang where the digest took place. The DNA bands observed following digestion appear to have appropriate mobility in the gel for the expected DNA products. While 10 units of XapI was sufficient to digest all the dsDNA, 5 units left a significant proportion undigested. Decreasing the number of units of enzyme decreased the proportion of undigested DNA, very little digestion occurred when 0.5 units of enzyme were used.

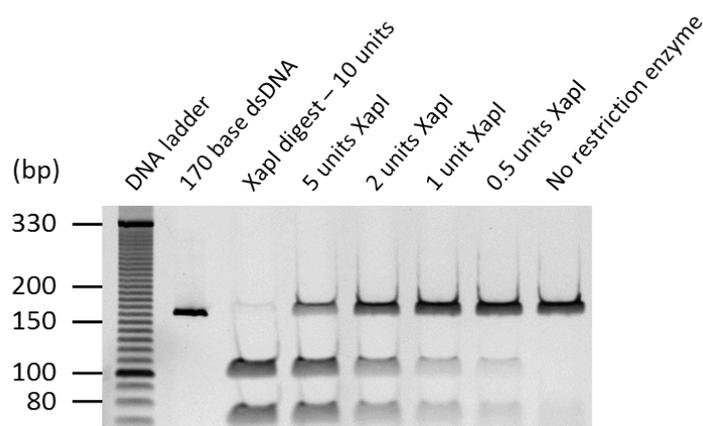


Figure 6.12: Native PAGE analysis of an XapI digest of the 170 bp DNA using a series of enzyme dilutions. From left to right: DNA ladder, undigested 170 base dsDNA, XapI digests using between 10 and 0.5 units of XapI, a control sample using the same protocol but without the restriction enzyme.

As stated in Section 5.3.1, the manufacturer states that one unit of XapI enzyme is the amount required to digest 1 μg of lambda DNA in 1 hour at 37 $^{\circ}\text{C}$ in 50 μl of recommended reaction buffer. The Lambda genome is 48.5 kbp and contains 58 XapI sites. The enzyme is supplied at a concentration of 10 units/ μl .

$(1 \mu\text{g Lambda} = 0.032 \text{ pmol}) \times 58 \text{ sites} = 1.86 \text{ pmol of digestion per unit of enzyme}$

$1.86 \text{ pmol} \times 10 \text{ units} = 18.56 \text{ pmol per } \mu\text{l of enzyme. Therefore a 1X digest requires } 1 \mu\text{l of enzyme for every } 18.56 \text{ pmol of DNA containing a single restriction site.}$

This digest was carried out with eight pmol in a 40 μl volume incubated at 37 $^{\circ}\text{C}$ for one hour. At this DNA concentration 10 units of enzyme was 2.3X the amount of enzyme needed to digest the DNA. At this concentration the enzyme was sufficient to digest all

the dsDNA. When five units of enzyme were used this represented 1.2X the amount of enzyme required; however, at this enzyme concentration, a significant proportion of the dsDNA remained undigested. This indicates that an excess of enzyme is required to fully digest dsDNA, despite the manufacturer's calculations of efficiency.

6.4 Patterning

A patterning experiment was carried out on the 170 base dsDNA using two 30 base patterning oligomers. The 170 base dsDNA is homologous to the central, fully base paired region of the 270 base structures. The results are shown in Figure 6.13.

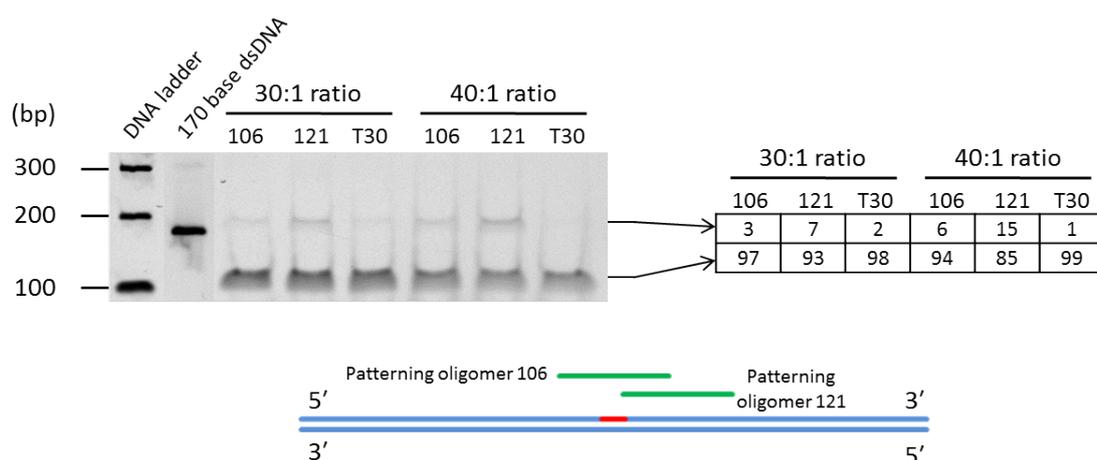


Figure 6.13: **Left:** Native PAGE analysis of a patterning experiment using three different oligomers on 170 base dsDNA. The three oligomers were used at a nucleoprotein filament to dsDNA ratio of both 30:1 and 40:1. **Right:** Densitometry analysis of the patterning experiment shown on the left. The upper band represents undigested DNA. **Below:** Diagrammatic representation of the patterning oligomers designated 106 and 121 (green) and their sequence homology with the 170 base dsDNA (blue). The recognition site for XapI is shown in red.

Densitometry analysis is shown, as well as a diagrammatic representation of the patterning oligomers designated 106 and 121. Oligomer 106 starts at base 106 of the full-length 270 base dsDNA, the centre of the patterning oligomer lies over the recognition site for XapI.

Oligomer 121 starts at base 121 of the 270 base structure; the recognition site for XapI runs from base 118 to base 123, therefore the patterning oligomer 121 only obscures half of the XapI recognition site.

The efficiency of patterning was very low for both constructs at the 30:1 nucleoprotein filament to dsDNA ratio. The same experiment was then carried out at a 40:1 ratio to investigate whether this could increase patterning efficiency.

At a 30:1 ratio the patterning oligomer starting at base 106 shows 3% of the DNA is undigested. The margin of difference between this result and the control using the T30 oligomer, which showed 2% undigested DNA, is insignificant. There is an increase in patterning efficiency seen when a higher nucleoprotein filament ratio is used, to 6% of the total lane intensity; however, the total remains low.

A higher patterning efficiency is observed for the patterning oligomer 121, this is the oligomer which only obscures half of the XapI recognition site. At a 30:1 ratio it showed undigested DNA represented 7% of the total intensity, slightly higher than the control of 2%. When used at a 40:1 ratio this increased to 15%, while the control was 1%. This indicates that total obfuscation of the recognition sequence could be unnecessary for patterning to protect a dsDNA strand, further investigation would be necessary to confirm and explore the effect of patterning oligomer positioning on digestion.

Following the low levels of success seen with the centrally patterning oligomers, a patterning experiment was carried out to test two oligomers designed to be close to the branch point of the 270 base structure at base 221. The resulting acrylamide gel can be seen in Figure 6.14.

Two oligomers were used with a 20 base overlap, within this overlap were two recognition sites for restriction enzymes. The Adjacent patterning oligomer was homologous to bases 191 to 220 and the Proximal patterning oligomer was homologous to bases 181 to 210. The Hpy166II recognition site was from base 192 to base 197 and the BseYI recognition site was from base 203 to base 208. This placed the Hpy166II site at the 5' end of the Adjacent oligomer, while the BseYI site would lie close to the centre of this oligomer. Similarly, the BseYI site was at the 3' end of the Proximal oligomer, while the Hpy166II site would lie at its centre. It was thought that this would provide information about

the effect of the alignment of the oligomer and the restriction site. An oligomer of 30 thymine bases (T30) was used as a control, containing no sequence homology to the 289 base strand.

It can be seen that there is effectively no difference in enzyme restriction when the Adjacent, Proximal and control oligomers were used in the patterning experiment. Had this experiment shown a level of RecA patterning above the control, this experiment would have been carried out using the 270 base structures containing three-way junctions. However, it was not possible to pattern dsDNA with these strands so further experimentation was carried out to identify patterning oligomers that would effectively pattern this sequence.

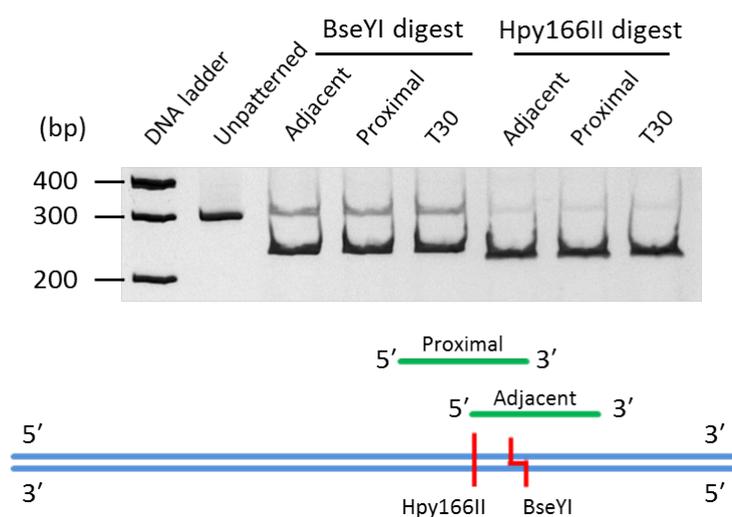


Figure 6.14: **Top:** Native PAGE analysis of a patterning experiment on 289 bp dsDNA. This scaffold was used to test the patterning oligomers on the dsDNA PCR products that are subjected to T7 treatment and annealing. The patterning regions for the oligomers adjacent and proximal were designed to be close to the branch point of the 270 base structure. Patterning was carried out with three ssDNA oligomers, the adjacent and proximal 30 base oligomers and a 30 base thymine oligomer as a control. No difference can be seen between the two patterning oligomers and the control sequence. **Below:** Diagrammatic representation of the patterning oligomers designated Adjacent and Proximal (green) and their sequence homology with the 289 base dsDNA (blue). The recognition sites for Hpy166II and BseYI are shown in red.

A range of patterning oligomers were ordered to explore the sequence space around the central XapI restriction site, from base 118-123. A series of 30 base oligomers were ordered to have homology in 3 base increments across the centre of the dsDNA region, at the 5' end of the range was an oligomer homologous to bases 94-123 and at the 3' end of the range was an oligomer homologous to bases 118-147. Two 40 base oligomers were also ordered, homologous to bases 101-147 and 115-154. Exploring the range of available oligomer sequences could bypass any region with a strong secondary structure. The results are shown in Figures 6.15 and 6.16.

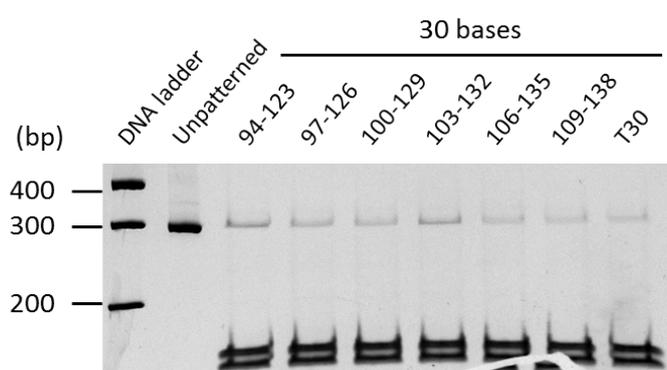


Figure 6.15: Native PAGE analysis of a patterning experiment on 289 bp dsDNA. This scaffold was used to test the patterning oligomers on the dsDNA PCR products that are subjected to T7 treatment and annealing. A range of 30 base oligomers were used to create nucleoprotein filaments for patterning dsDNA.

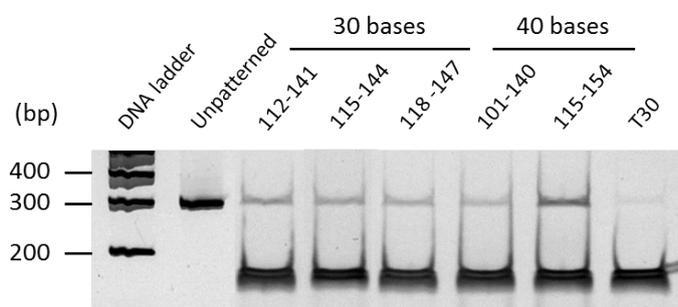


Figure 6.16: Native PAGE analysis of a patterning experiment on 289 bp dsDNA. This scaffold was used to test the patterning oligomers on the dsDNA PCR products that are subjected to T7 treatment and annealing. A range of 30 and 40 base oligomers were used to create nucleoprotein filaments for patterning dsDNA.

These experiments made it further evident that RecA patterning of the sequence was taking place at extremely low levels. For the lanes shown in Figures 6.15 the average level of undigested DNA was 6.9% with a standard deviation of 3.4% (n=9). In comparison, for the T30 control, undigested DNA made up 2.4% (Standard deviation 0.5%, n=2) of the total intensity in the lane.

A higher level of patterning appears to have taken place when one of the 40 base patterning oligomers were used. When the 40 base oligomer homologous to bases 101-147 was used, the undigested DNA made up 7% of the total lane intensity, in keeping with the average for the 30 base oligomers. When the oligomer homologous to bases 115-154 was used, undigested DNA made up 19% of the total lane intensity. Although this is a lot higher than the rest of the oligomers it remains much lower than observed in other systems.

6.5 Conclusions

The 270 base structures containing three-way junctions were produced with relative ease, in sharp contrast to the 200 base structures. The production of ssDNA with T7 polymerase, as described in Tosch *et al.* [243], was successful and produced quantities that could be used in annealing reactions. The annealing of additional oligomers to produce three-way junctions was carried out to produce structures that could be visualised through AFM. The junctions on these structures were 100 bases long in total, it is probable that the length of the strands conferred a stability that was not possible with the 60 base junctions of the 200 base structures.

For unknown reasons the patterning of these structures consistently failed or showed very low yield. A range of oligomers were tested and only moderate success was achieved with a single 40 base oligomer. It is not clear whether this was a problem with the structure or with the reagents. On multiple occasions the reagents were replaced but this did not have any discernable effect on patterning efficiencies. The protocols were the same as those used for the 100 base structures with which interesting results were achieved.

The melting temperatures of the patterning oligomers used in this work vary from 41 °C to 59.6 °C. There is some evidence that secondary structure can be deleterious to RecA

filament formation [257]. The melting temperatures of the patterning oligomers used for the 100 base structures was 47.1 °C. Given that successful patterning was carried out with these oligomers, it does not seem likely that an inability to form a filament due to secondary structure was the cause of the problems encountered in this work. This was also addressed in the final experiment of this work, in which a series of 30 base oligomers were designed to start at 3 base increments across the dsDNA. AFM analysis of these structures implies that the correct forms were created through the annealing protocol. Plasmid sequencing received from the manufacturer showed the correct sequence had been achieved. Subsequent sequencing showed a single point mutation outside the areas of interest for this study.

It has been demonstrated in this chapter that the production of DNA structures with terminal three-way junctions is achievable through the use of synthetic genes, phosphorothioated PCR primers and T7 exonuclease. Any future work on this topic should aim to understand the failure to pattern these structures. AFM analysis could be used to gain visual confirmation of whether patterning is occurring; future work should consider this to be a starting point for investigating RecA patterning of these structures. When a patterning protocol has been established important data can be collected on the effect of DNA topology on RecA patterning.

Chapter 7

Expression and purification of the bacterial protein RecA

Recombinase A (RecA) mutants lacking a C terminus have been shown to bind to DNA with a higher affinity than wild-type RecA [162, 225, 164]. RecA with this modification could increase the efficiency of patterning; it is not commercially available so it would have to be produced. The first step in the production of a truncated RecA molecule was to develop a robust protocol for purification; this was undertaken using full-length RecA which had an affinity tag to aid in chromatographic separation.

Purification of a wild-type RecA protein containing a histidine affinity tag (His-tag) was attempted, the protocols are detailed in Section 3.2.8. RecA is a 38 kilodalton (kDa) protein, the addition of the His-tag increases this to 39 kDa. A modified RecA including an additional cysteine residue was also produced in the early stages of the experiments. The cysteine residue was added to gene at the C terminus using a modified PCR primer. Cysteine residues are able to form bonds with gold surfaces and this could be used to functionalise or surface-localise structures containing these proteins. The RecA gene was cloned into three plasmids, with or without the His-tag for purification. The plasmids were transformed into *Escherichia coli* cells, DH5 α cells were used for plasmid amplification and BL21 Star (DE3) *E. coli* were used for expression of the protein. The properties of the plasmids and cells are discussed in Sections 3.1.9 and 3.1.10. A series of expression trials were undertaken to confirm the production of soluble RecA protein. On the basis of

these experiments the cysteine modified RecA was not pursued further.

The purification protocol was based on standard His-trap purification using a nickel affinity column. This requires an affinity tag of 6-8 histidine residues at one end of the protein which bind to the nickel ions on the column. Imidazole is used as a competitor to remove bound protein from the column. Two constructs were used which placed the His-tag at the N and C termini of the protein. The initial protocol produced some partially purified RecA protein, the construct containing the N terminal His-tag was taken forward to larger scale expression. Cleaner sample preparation and elution with a timed gradient of imidazole greatly improved purification but bound DNA was still a contaminant. An ATP wash was used in an attempt to decrease the DNA content and led to some improvement.

RecA protein contains three cysteine residues; the cytoplasm of *E. coli* is a reducing environment which does not allow the formation of disulphide bonds. However, these bonds can form following cell lysis and cause aggregation and misfolding. For this reason, dithiothreitol (DTT) was added to the protein buffers to reduce aggregation and appeared to have a positive effect on the yield obtained by chromatography. Purified protein showed little or no activity.

Following these results the attempt to produce RecA was halted; the benefits of success were outweighed by the potential time cost of completing the project.

7.1 Cloning

The first step was the creation of a number of plasmids with one of two RecA inserts; one was the wild-type protein and the second was a cysteine-modified RecA. Cysteine is an amino acid with a thiol group which allows it to form bonds with gold surfaces. This modification adds the sort of functionality that makes RecA patterning of DNA structures highly desirable. Although this was not relevant to the work in this thesis, the expression trials showed some interesting results and so are discussed here. Only the wild-type protein was moved forward to larger-scale expression and purification. The sequence of the *recA* gene and its translation into a protein primary structure are shown in Figure 3.5 and in Appendices RecADNASequence and RecAProteinSequence, details of the cloning

protocol can be found in Section 3.2.8. The benefits of the plasmids pET23a, pET15b and pET11a are discussed in Section 3.1.9, maps of the plasmids and their restriction sites are shown in Appendix B.3.

DNA modifying enzymes were used to ligate the *E. coli recA* gene into a plasmid for amplification and for the production of protein. The plasmids and inserts were stored digested from previous lab experiments, the preparation and ligation protocols can be found in Section 3.2.8, Cloning. The pET23a plasmid and the wild-type and cysteine modified insert are shown in Figure 7.1. The pET23a plasmid is 3.6 kilobases (kbp), the wild-type insert is 1.096 kbp and the cysteine-modified insert is 1.110 kbp. Their mobilities on the gel are in keeping with these values.

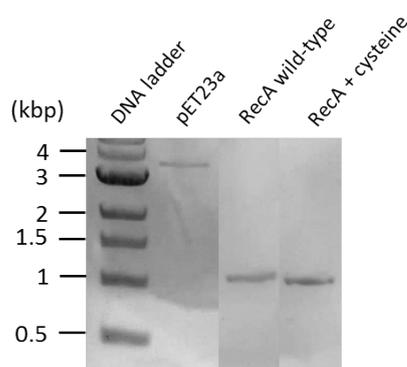


Figure 7.1: Digested plasmid and inserts analysed by acrylamide gel electrophoresis (15% 29:1 acrylamide) to confirm the size and purity of the DNA. pET23a is shown, as well as the two inserts. Some lanes have been removed for clarity, representative samples are shown.

Following the ligation plasmids were transformed into DH5 α *E. coli* cells as described in Section 3.2.8, Transformation. This cell line was used for amplification, increasing the number of plasmids. The amplified plasmids were collected by Qiagen miniprep, as per the manufacturer's recommendation. The plasmids were then sequenced by Eurofins MGW Operon and selected for transformation into a protein production cell line. While a number of ligations proved to have some sequence damage, there were enough with the correct sequence to carry out expression trials with all of the constructs desired. The amplified plasmids were also run on an agarose gel, shown in Figure 7.2. The expected

sizes of all the generated constructs can be found in Table 7.1.

Table 7.1: Expected sizes of the generated constructs following ligation of the insert into the plasmid

Plasmid + insert	Expected size (kbp)
pET11a + wild-type <i>recA</i>	6.796
pET11a + cysteine modified <i>recA</i>	6.81
pET15b + wild-type <i>recA</i>	6.796
pET15b + cysteine modified <i>recA</i>	6.81
pET23a + wild-type <i>recA</i>	4.696
pET23a + cysteine modified <i>recA</i>	4.71

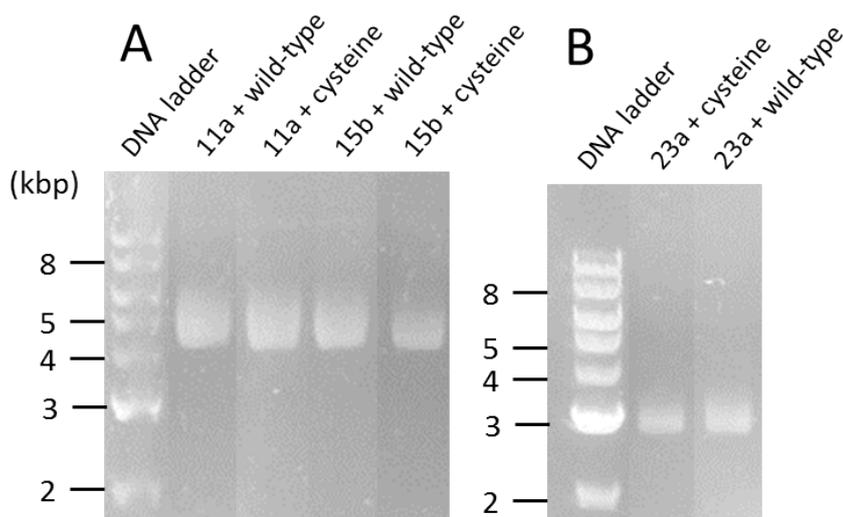


Figure 7.2: 1 % native agarose gel electrophoresis of plasmids with ligated inserts collected by miniprep from DH5 α *E. coli*. **(A)** The pET11a and 15b plasmids containing both the wild-type and cysteine-modified RecA inserts. **(B)** The pET23a vector containing both the wild-type and cysteine-modified RecA inserts.

Both pET15b and pET11a are 5.7 kbp in size, pET23a is 3.6 kbp. The pET15b and pET11a plasmids including the inserts run at between 4 and 5 kbp, the pET23a plasmids with the insert runs at approximately 3 kbp. Following amplification in *E. coli* cells the DNA should be in a supercoiled state, allowing them to migrate more quickly through

the gel than linear DNA. This is reflected in this gel, the plasmids migrate approximately 35% faster through the gel than would be expected from their length. A restriction digest of these plasmids would have provided more information about the size of the DNA and the correct insertion of the genes.

7.2 Expression Trials

Before large scale protein production was undertaken, experiments were done to confirm the constructs would produce the RecA protein under IPTG induction. The method can be found in Section 3.2.8 Induction. Plasmids containing the correct sequence were transformed into BL21 Star (DE3) *E. coli*. The bacteria were grown in 5 ml of LB broth for induction of protein production and then centrifuged for buffer removal and resuspended in a protein buffer. The cells were lysed by sonication to release the protein and centrifuged to separate the soluble and insoluble fractions. The samples were then analysed by SDS PAGE. For further information on these protocols, refer to Chapter 3 Sections 3.1.4, 3.2.4 and 3.2.8.

7.2.1 First expression trial

The results of the first expression trial are shown in Figure 7.3. Three samples are shown for each construct: the pre-induction total cell lysate (A), the post-induction soluble cell lysate (B) and the post-induction insoluble cell lysate (C). This combination of samples allows us to identify additional proteins only present following IPTG induction and to ascertain whether these proteins are soluble, indicating correct folding. Insoluble protein indicates poor folding, possibly leading to aggregation. While misfolded proteins can be refolded with the aid of denaturants and additives, it is preferable to produce soluble protein in the first instance.

Little to no soluble protein is visible in Figure 7.3. The wild-type RecA has a molecular mass of 37,973 daltons (Da), the cysteine modified protein is 38,076 Da. The addition of the His-tag leads to a construct of 38,796 Da with the wild-type RecA protein and 38,899

Da with the cysteine terminated protein. The NEB RecAf construct which is also His-tagged is 37,842 Da. Some lanes were overloaded, but there is little evidence of RecA production. What is not shown is the consistency of the cell lysate; following sonication the cell lysate was found to be extremely viscous. Centrifugation did not produce a well defined pellet, which is required for efficient separation of the soluble and insoluble cell constituents. It was decided to repeat the experiment with the addition of DNase I to the cells before they were lysed, to aid in breaking up cellular DNA.

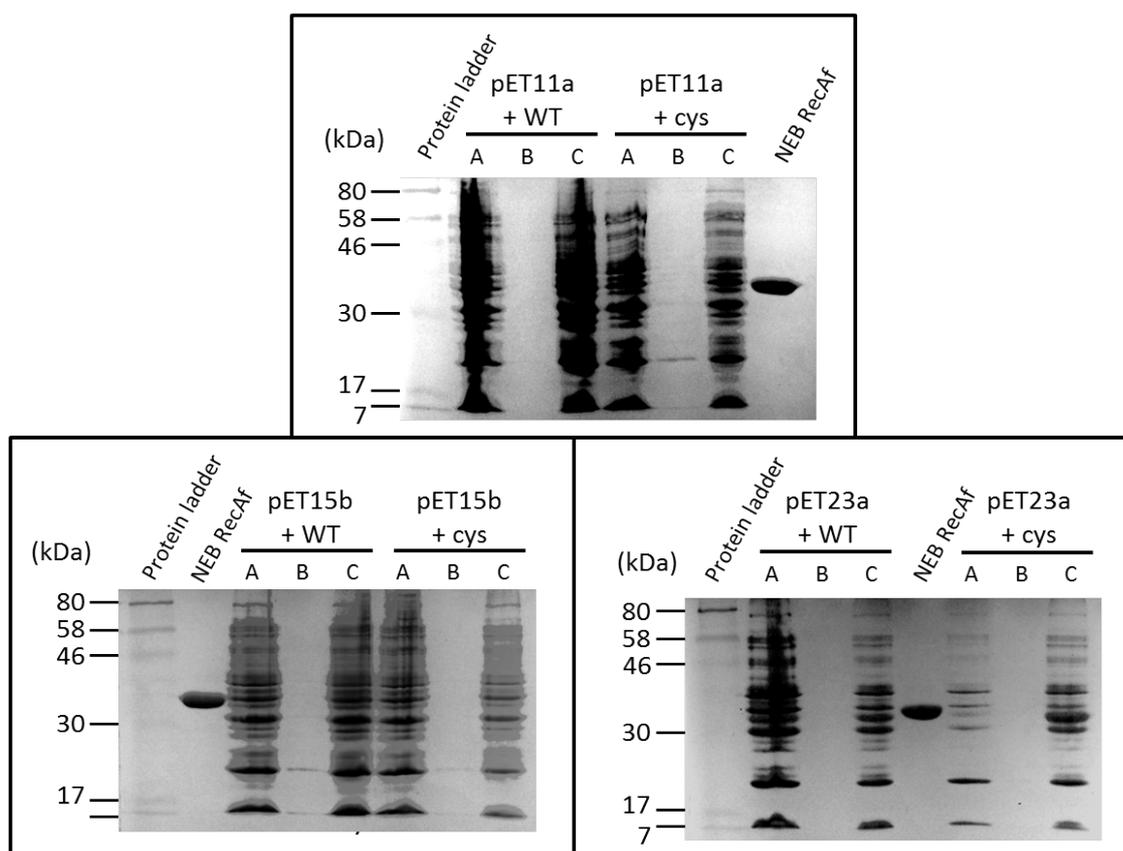


Figure 7.3: First expression trial of RecA protein. RecAf from NEB was also run on the gel to show the approximate size of the desired protein. **(A)** is the pre-induction cell lysate, an uninduced control representing the native proteins of the *Escherichia coli*. **(B)** is the post-induction soluble protein, i.e. the protein which remained in solution following centrifugation. **(C)** is the post-induction insoluble protein, the protein which was found in the pellet following centrifugation.

7.2.2 Second expression trial

The results of the second expression trial are shown in Figure 7.4.

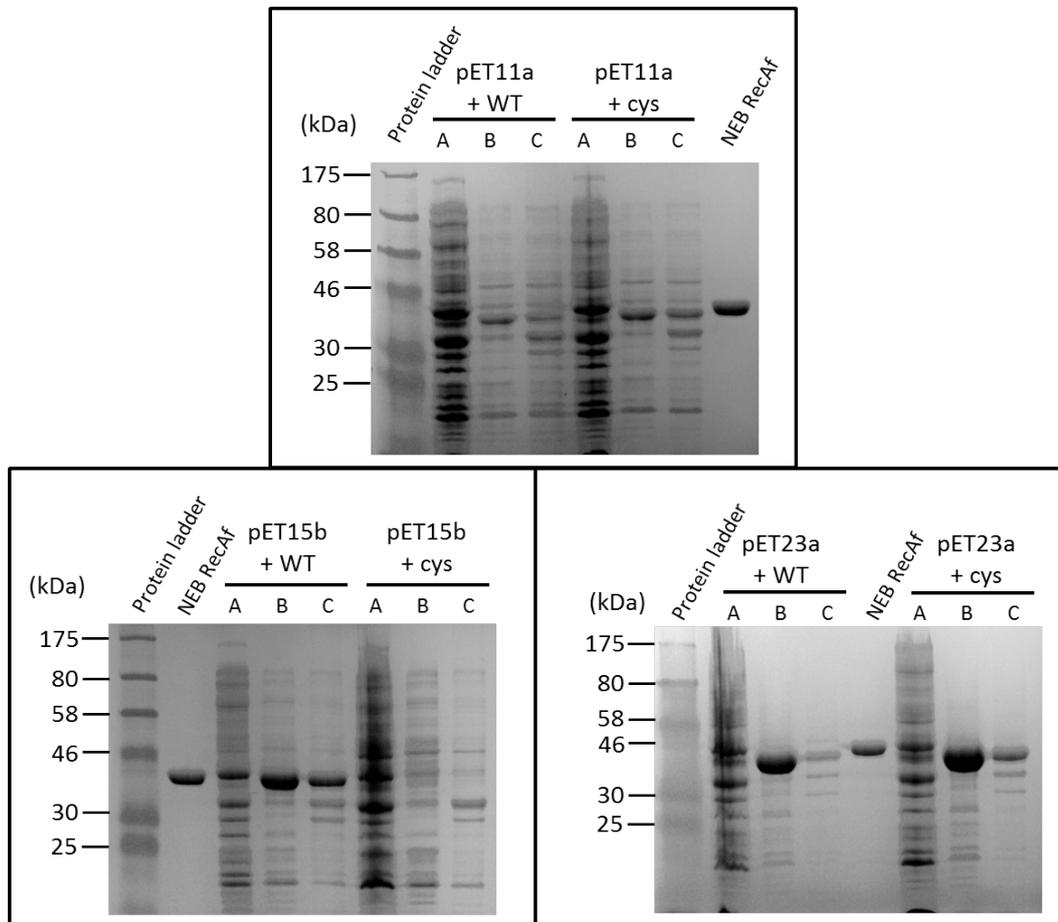


Figure 7.4: Second expression trial of RecA protein. In this instance DNase was added to the cells before they were lysed. RecAf from NEB was also run on the gel to show the approximate size of the desired protein. **(A)** is the pre-induction cell lysate, an uninduced control representing the native proteins of the *Escherichia coli*. **(B)** is the post-induction soluble protein, i.e. the protein which remained in solution following centrifugation. **(C)** is the post-induction insoluble protein, the protein which was found in the pellet following centrifugation.

The definition of the lanes is much better; individual bands of intensity can be seen and smearing has been reduced. It is clear that sonication alone was not sufficient to break up the cellular DNA and this was preventing the soluble and insoluble proteins from

being separated, as well as preventing the efficient running of proteins through the SDS acrylamide gel.

RecAf from NEB was included on the gels to give an indication of the size of the desired protein, as for the first expression trial. The protein from the pET11a plasmid appears smaller than RecAf; this is to be expected as there is no His-tag present. The mobility of RecAf in the gel is similar to that of the protein from the pET15b plasmid. The RecAf appears larger than the protein from the pET23a plasmid; it is not clear why this is the case as the correct cloning was confirmed through sequencing. In later gels the RecAf and the protein from the pET23a plasmid appear to migrate at very similar rates through the gel, indicating that the difference here may be due to an issue with the gel. The mobility of the protein in this gel does not indicate folding of the protein, SDS gels contain a denaturing agent and mobility through the gel is dictated by the length of the amino acid chain.

The wild-type and cysteine modified RecA express to some extent in pET11a (no affinity tag), shown as Constructs 1 and 2 respectively in Figure 7.4. The protein appears to be mostly soluble; expression is perhaps slightly higher when the cysteine residue is present.

The wild-type protein expresses well in pET15b (N terminal His-tag), shown as Construct 3 in Figure 7.4; though some of the protein is insoluble there is a much larger proportion in the soluble protein fraction. The cysteine modified RecA does not express at all in pET15b, shown as Construct 4.

The expression of the cysteine modified RecA was highest in the pET23a plasmid (C terminal His-tag), shown as Construct 6 in Figure 7.4. Only a small proportion of the protein is insoluble. The wild-type RecA expresses to a similar extent in pET23a as pET15b, shown as Construct 5 and Construct 3 respectively. There is less insoluble RecA when expressed in pET23a.

These results confirmed that the wild-type protein could be expressed with either an N or a C terminal His-tag for the purposes of affinity column purification. The relatively low levels of insoluble protein indicated that the protein is folding correctly; although RecA is a native *E. coli* protein and will therefore have access to the correct molecular chaperones and folding conditions within the cell, the large volume of protein produced could have had a detrimental effect on folding and therefore solubility.

The complete lack of expression of cysteine modified RecA in pET15b (Construct 4) is interesting. There is not even a high level of insoluble protein to indicate misfolding. This could be due to an error in the protocol or a problem with the plasmid used leading to a total lack of expression. No further investigations were performed, as the aim was to produce wild-type protein.

7.3 Protein Production and Purification

The expression trials showed that soluble, wild-type RecA could be produced with a N or a C terminal His-tag and that the expression levels were reasonably high. It was not immediately clear which would be a more successful protocol, so cells containing both pET15b and pET23a were grown on a larger scale for purification, in a 300 ml volume. The cysteine modified constructs were not developed further. The protocols used are described in Chapter 3 Section 3.2.8 Induction. Nickel affinity chromatography was carried out using 1 ml HisTrap HP columns, as described in Chapter 3 Section 3.2.8 FPLC.

The protocol requires a string of six or more histidine residues on a terminus of the protein; the side group of the amino acid histidine is an imidazole group which will bind to a number of metals, including nickel. The chromatography column contains immobilised metal ions and binds, free imidazole is then added which competes with the protein to release it from binding. The chromatography columns used contained agarose beads with bound nickel.

Data given during FPLC is used to assess the point at which proteins are eluted. The fluorescence absorbance of the liquid coming out of the column is measured to identify protein and DNA eluates; 260 and 280 nm wavelengths were recorded, corresponding to maximum emission peaks for DNA and protein. Additional information is collected regarding the chromatography e.g. flow rate and system pressure, as these can have an effect on the column. The samples, or fractions, are analysed by SDS PAGE to confirm the size and purity of the proteins collected.

7.3.1 pET23a plasmid

The RecA construct in the pET23a plasmid with a C terminal His-tag (Construct 5) was purified in the first experiment. The resulting FPLC data and an acrylamide gel of the proteins collected are shown in Figure 7.5.

In Figure 7.5 (A) one large peak is observed over fractions C5 to C6, with a small side peak in C6-C7. The SDS PAGE analysis in Figure 7.5 (B) shows very little protein in these samples. This could indicate that the pET23a plasmid, with a C terminal His-tag, has not bound to the column, possibly due to misfolding that leads to occlusion of the affinity tag but not a loss of solubility. It is possible that the large peak at C5-C6 was an air bubble in the lines which caused a false absorbance reading.

7.3.2 pET15b plasmid

The same protocol was carried out on the wild-type RecA protein in the pET15b plasmid (Construct 3), with an N terminal His-tag. The FPLC data and an acrylamide gel of the results are shown in Figure 7.6.

There are a number of peaks on the chromatogram in Figure 7.6 (A), starting with the introduction of the imidazole. Some of these could be contaminants, but the SDS PAGE analysis in Figure 7.6 (B) indicates that a protein of the correct size has been purified. There is little protein contamination of these samples and a large amount of protein seems to be present. Concentration was not assessed at this point.

Both 260 nm and 280 nm absorbance are high for this sample. The absorbance indicates both a high level of protein and also of bound DNA.

One of the long-term aims of the protein purification project was the production of RecA protein without the C terminus, which would have the effect of increasing the protein's affinity for DNA [162, 225, 164]. Synthesising RecA with a C terminal His-Tag at this point would ensure the need for a new purification protocol when the C terminal mutant was produced. On the basis of these factors and the results presented here, work was continued with the N terminal His-tag.

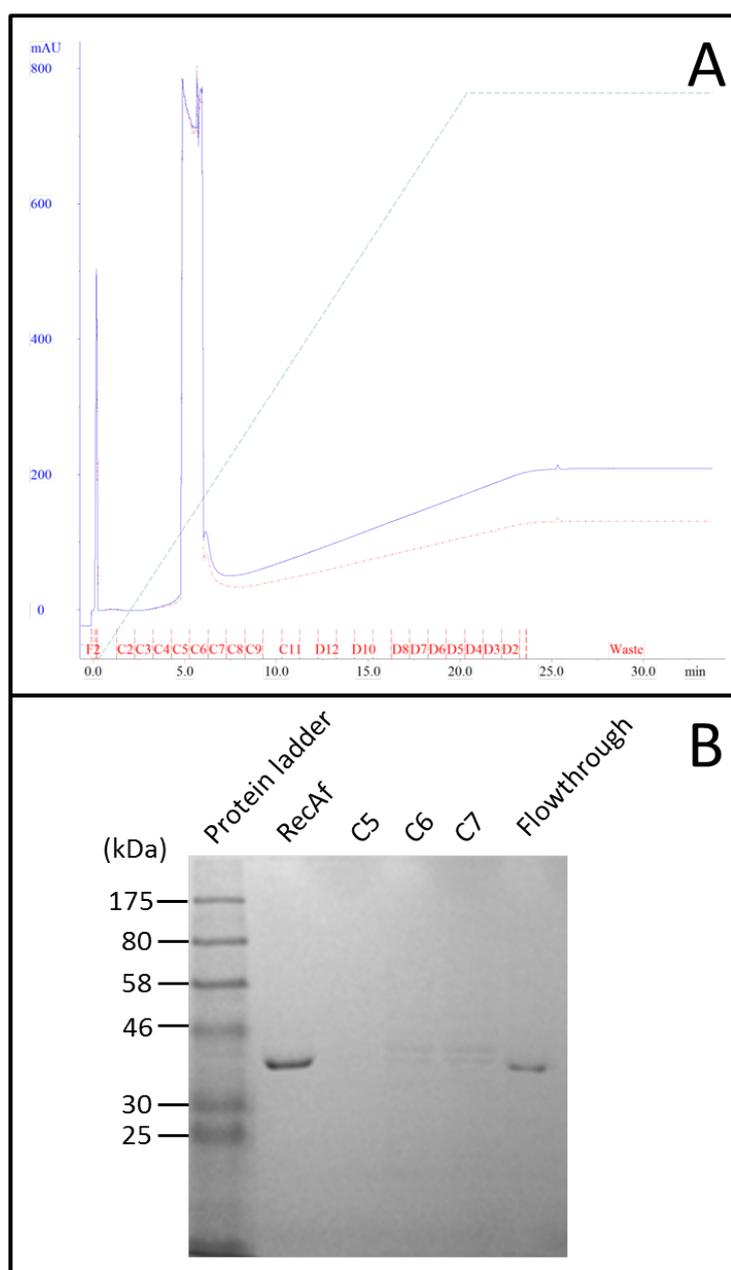


Figure 7.5: **A:** FPLC chromatogram for the purification of wild-type RecA in the pET23a plasmid, with a C terminal His-tag. 280 nm absorbance is shown in blue, 260 nm absorbance is shown in red. The gradient of the imidazole used to elute the protein is shown in green from 0 mM to 400 mM and the fractions collected are shown in red on the X axis. The Y axis is milli Absorbance Units (mAU) of the 280 nm absorbance. Time (minutes) is shown on the X axis. **B:** SDS PAGE of fractions C5 to C7 of the chromatogram above. RecAf from NEB and a protein ladder are included for size comparison. A sample retrieved as the soluble cell lysate was injected onto the column represents protein that failed to bind and is labeled 'flowthrough'.

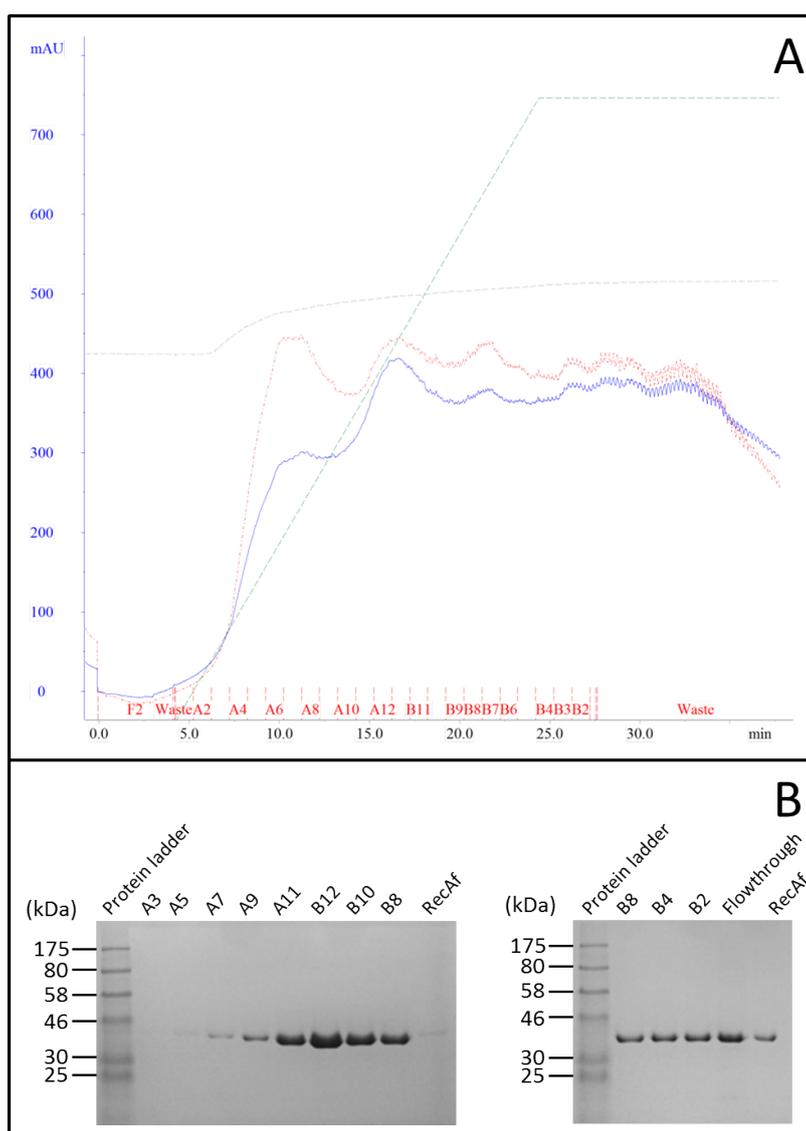


Figure 7.6: **A**: FPLC chromatogram for the purification of wild-type RecA in the pET 15b plasmid, with a N terminal His-tag. 280 nm absorbance is shown in blue, 260 nm absorbance is shown in red. The gradient of the imidazole used to elute the protein is shown in green from 0 mM to 400 mM and the fractions collected are shown in red on the x axis. The pH is shown in grey. The Y axis is milli Absorbance Units (mAU) of the 280 nm absorbance. Time (minutes) is shown on the X axis. **B**: SDS PAGE of samples from the chromatogram in panel A. The samples A3 etc. refer to the fractions eluted from the column, shown in red on the x axis of the graph in **A**. RecAf from NEB and a protein ladder are included for size comparison. A sample retrieved as the soluble cell lysate was injected onto the column represents protein that failed to bind and is labeled 'flowthrough'.

The DNA contamination observed in the experiment shown in Figure 7.6 would interfere with patterning experiments; for this reason steps were taken to refine the protein purification protocol to reduce the level of bound DNA. The scale of protein production was increased and FPLC was carried out.

7.4 Refining the purification protocol

A number of changes to the purification protocol were made between the experiments shown in Figures 7.6 and 7.7. Cells were grown and induced to produce protein in a 600 ml volume as described in Section 3.2.8, Protein Induction. The soluble protein was passed through a 0.2 μm filter before adding to the column in an attempt to remove some of the contaminants. Imidazole was added to the supernatant and to the buffer used to equilibrate the column at a concentration of 20 mM, to reduce non-specific protein binding. The length of the imidazole gradient was also increased from 10 column volumes to 20 column volumes; it was hoped this would resolve some of the peaks seen in Figure 7.6 and allow better separation of RecA from any contaminants. The second purification of the wild-type RecA construct in the pET15b plasmid is shown in Figure 7.7.

Following the changes in the protocol a single peak is observed on the chromatogram in Figure 7.7 (A). The absorbance at 260 nm was found to increase with the increase in imidazole concentration; from this it was concluded that the imidazole used also absorbed at this wavelength and it was replaced with a higher quality product from a different company. The 280 nm absorbance also increases with imidazole concentration, although to a lesser degree. This could indicate that the 260 nm peak is broad enough to increase apparent absorbance at 280 nm.

The SDS acrylamide gel in Figure 7.7 (B) shows that a large amount of quite pure protein is recovered with this protocol; no bands other than the RecA sample are observed. Protein concentration is most often gauged using absorbance; given the absorbance of the imidazole in this sample it was not possible to accurately measure this way. This protocol produces a protein free from protein contaminants but the peak in the 260 nm absorbance that appears concurrently with the protein elution indicates that the amount of

DNA bound to the RecA remains high. RecA is a DNA binding protein, so it is unsurprising that there is a large amount bound to the sample.

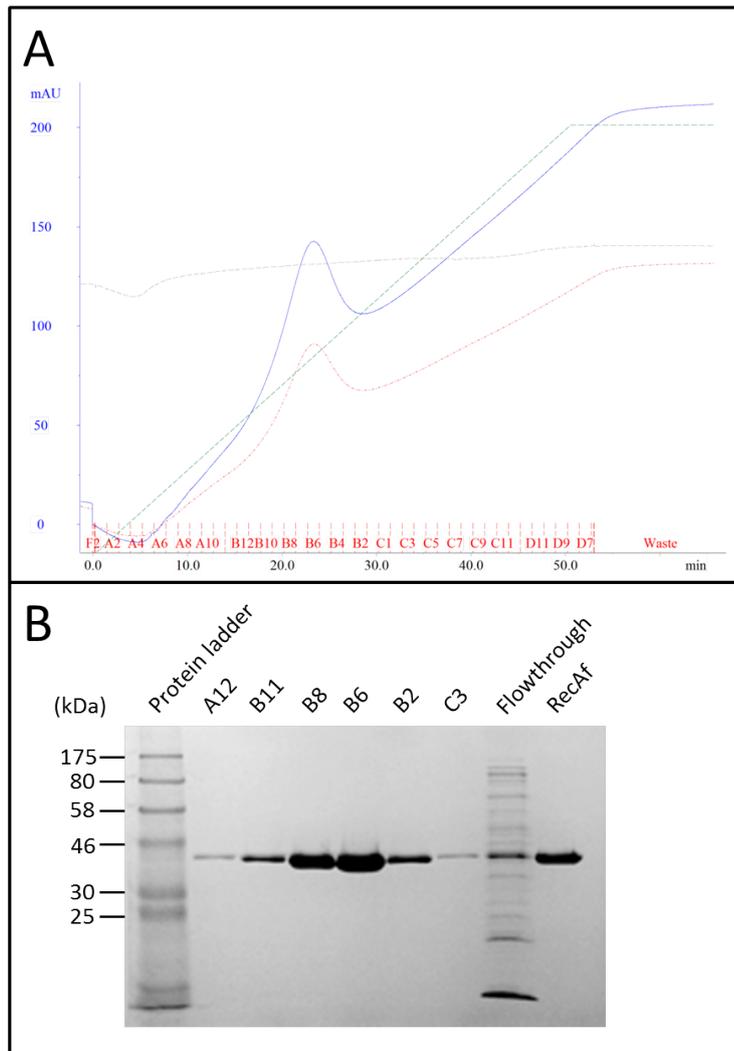


Figure 7.7: **A**: FPLC chromatogram for the refined protocol purification of wild-type RecA with an N terminal His-tag. 280 nm absorbance is shown in blue, 260 nm absorbance is shown in red. The concentration of the imidazole used to elute protein is shown in green from 0 mM to 400 mM and the fractions collected are shown in red on the x axis. The pH is shown in grey. The Y axis is milli Absorbance Units (mAU) of the 280 nm absorbance. Time (minutes) is shown on the X axis. **B**: SDS PAGE of samples from the chromatogram on the left. RecA_f from NEB and a protein ladder are included for size comparison.

7.4.1 ATP wash

Having obtained protein with a relatively high yield, experiments were carried out to reduce the amount of DNA that was bound to the RecA. DNA from *E. coli* would interfere with patterning experiments, in which the concentration of different DNA elements is carefully controlled. An ATP wash was attempted after loading the supernatant on the column and before eluting with imidazole, the chromatogram is shown in Figure 7.8. RecA uses ATP as a cofactor for binding DNA; when it is hydrolysed the RecA dissociates [192]. It was thought that introducing ATP to the column-bound RecA might induce it to release the bound DNA and help purify the sample. The wavelength at which the absorbance of the DNA was measured was changed from 260 nm to 254 nm for the ATP wash. ATP also has an absorbance peak at 260nm; however, the DNA peak is broader than that of ATP. The monitored absorbance was changed to try to visualise the DNA while reducing the contribution of ATP. The monitored absorbance was changed to try to visualise the DNA while reducing the contribution of ATP.

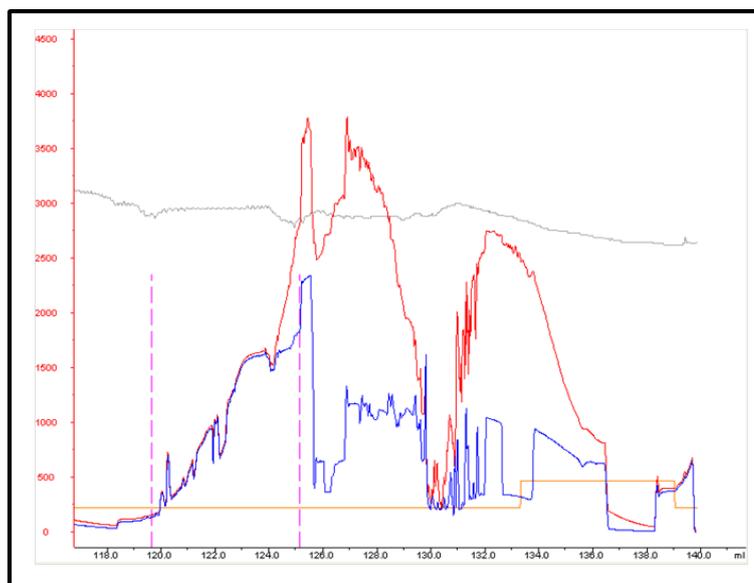


Figure 7.8: FPLC chromatogram for the 10 mM ATP wash of the His-trap column with bound RecA. 280 nm absorbance is shown in blue, 254 nm absorbance is shown in red. The flow rate is shown in orange and the injection of the ATP is shown in pink on the x axis. The pH is shown in grey. The Y axis is milli Absorbance Units (mAU) of the 260 nm absorbance. Time (minutes) is shown on the X axis.

There are a number of peaks in both 254 nm and 280 nm absorbance present during

the ATP wash. This indicates the elution of both DNA and protein. It is likely that the slopes are not smooth due to air bubbles; the ATP was not degassed before injection onto the column. The bubbles caused by the changes in pressure reflect the light giving the impression of absorbance. Following the ATP wash, the protein elution was carried out over a gradient of imidazole, shown in Figure 7.9. Higher quality imidazole was used in this protocol and as a result the absorbance does not increase with the increase in elution buffer.

The chromatogram shows a lot less protein was collected from the column, as well as less DNA. The SDS PAGE analysis in Figure 7.9 (D) shows that the initial peak does not contain discernable RecA but the following two, smaller peaks do. The greatest concentration of protein is found in the second peak, at \sim 46-56 minutes. This is at the higher end of the imidazole concentration, indicating the protein is binding with a high affinity to the column. The ATP was not buffered in this step, which may have led to some of the downstream problems with the protein.

The flowthrough is also shown on the gels in panel (D), representing protein samples that have not bound to the column. Some RecA appears to be present in this sample, as well as a number of contaminants. The flowthrough from the ATP wash in Figure 7.8 is also shown in panel (D), confirming a low concentration of protein was eluted from the column by the addition of ATP. The possibility that the protein was aggregating was considered; if this was the case it would not be bound sufficiently to the column and could be eluted during the wash step. The buffer composition of the NEB RecAf was studied and a decision was made to include DTT; there are three cysteine residues in wild-type RecA that could be forming non-native disulphide bonds. A surface rendering of the RecA protein is shown in Figure 7.10, from this image it is clear that at least one of these residues is surface accessible, making it possible that non-native dimerisation could occur in a non-reducing environment. DTT is tolerated by these columns at the required concentration (1 mM).

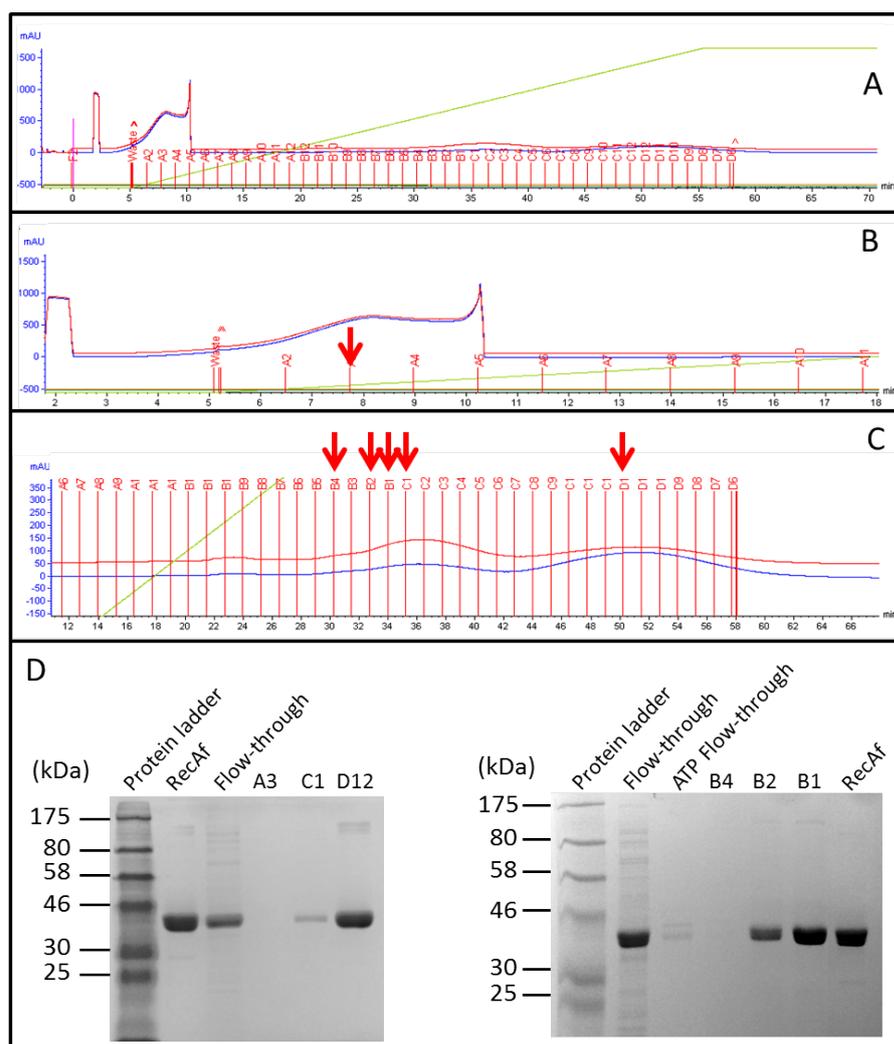


Figure 7.9: **A:** FPLC chromatogram of the purification of wild-type RecA with a N terminal His-tag using an updated protocol and an ATP wash. 280 nm absorbance is shown in blue, 260 nm absorbance is shown in red. The concentration of the imidazole used to elute protein is shown in light green from 0 mM to 400 mM and the fractions collected are shown in red on the x axis. The column pressure is shown in dark green. The Y axis is milli Absorbance Units (mAU) of the 280 nm absorbance. Time (minutes) is shown on the X axis. **B:** Closer view of minutes 2 to 18 of the chromatogram shown in panel A. The position at which fraction A3 begins is highlighted with a red arrow. **C:** Closer view of minutes 12 to 66 of the chromatogram shown in panel A. The positions at which a few of the fractions begin have been highlighted with red arrows; these fractions are shown analysed by SDS PAGE in panel D. **D:** SDS PAGE of the samples gathered from the chromatogram in panels A - C. Samples were dialysed into an imidazole-free buffer (25 mM Tris buffer pH 7.4, 100 mM NaCl) before electrophoresis.

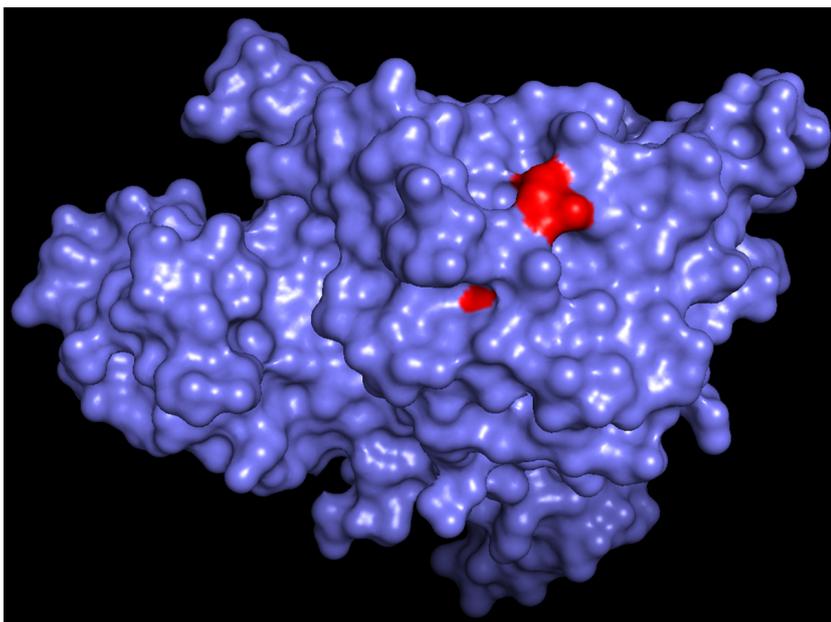


Figure 7.10: Surface rendering of the RecA protein from *Escherichia coli* (Protein Data Bank file 4TWZ) Resolution is 2.8 Å. The surface accessible residues are all shown in blue, except for the cysteine residues which are shown in red. It is clear from this image that at least one cysteine residue is surface accessible in the native fold. Image produced using PyMOL.

7.4.2 Addition of DTT

The buffer composition was updated to include 1 mM DTT and the purification protocol was repeated. The chromatogram and SDS PAGE analysis of the protein collected are shown in Figure 7.11.

The use of DTT in combination with the ATP wash and other changes in the protocol results in a much sharper peak with less DNA contamination, shown in Figure 7.11 (A). The first peak has been eliminated, indicating that it could have been aggregated protein held together by disulphide bonds. The samples from the peak of this chromatogram were pooled to produce a homogenous RecA solution for further testing. Samples were taken from the centre of the peak in 280 nm absorbance (fractions C10 to D9) and stored at 4 °C.

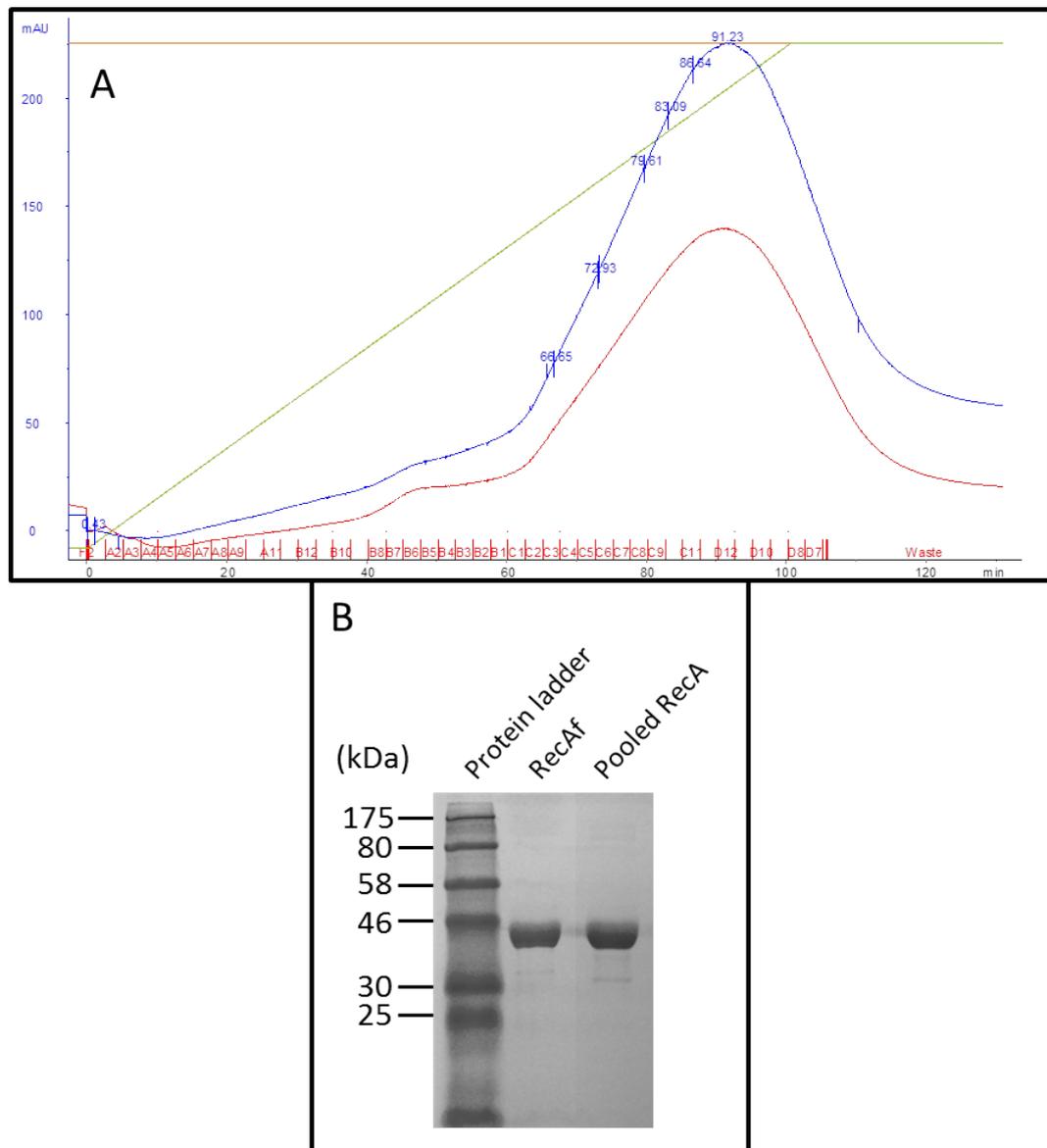


Figure 7.11: **A:** FPLC chromatogram for the refined protocol purification of wild-type RecA with a N terminal His-tag following the addition of DTT to buffers used in the protocol. 280 nm absorbance is shown in blue, 260 nm absorbance is shown in red. The concentration of the imidazole used to elute protein is shown in green from 0 mM to 400 mM and the fractions collected are shown in red on the x axis. The flow rate of liquid through the column is shown in orange. The Y axis is milli Absorbance Units (mAU) of the 280 nm absorbance. Time (minutes) is shown on the X axis. **B:** SDS PAGE of the samples collected from the FPLC shown in panel A. From left to right: NEB low molecular weight protein ladder, NEB RecAf, pooled samples from the peak intensity of the chromatography shown in panel A.

The SDS PAGE analysis of the protein is shown in Figure 7.11 (B). The number of contaminants is low; it appears to be of similar purity to the RecAf from NEB. The concentration of the pooled samples was calculated to be 1.38 mg/ml with a total volume of approximately 8 ml, making the yield of protein \sim 11 mg from 2 litres of growth medium. The binding capacity of the columns was 40 mg, so the protocol was not limited by the maximum column protein retention.

7.5 Activity assay

To gauge the success of purification the activity of the resulting protein was tested. The protocol for this assay is described in detail in Section 3.2.8. A RecA concentration of 1.38 mg/ml of protein was achieved. A DNA strand 3.5 kbp long formed nucleoprotein filaments with RecAf from NEB and with a range of ratios of the purified RecA. The ratio of RecA to dsDNA was calculated using the concentration of RecA calculated by spectrophotometry, the binding of one RecA molecule to three bases of DNA and the concentration of the 3.5 kbp DNA. The protein-DNA samples were then analysed by native agarose gel electrophoresis. The binding of RecA should result in a characteristic 'smear' on the gel, where the RecA slows the mobility of the DNA through the gel and the range of stoichiometries of bound RecA to DNA results in a continuous range of mobilities. The results are shown in Figure 7.12.

The DNA ladder used for the gel in Figure 7.12 (A) was too short and does not help identify any bands on the gel. The 3.5 kbp DNA gives a better indication of the size of the DNA bands. The DNA coated with the purified RecA shows a small amount of smearing but most of the DNA and protein appears to have remained in the wells at the top of the gel. This indicates that large aggregates have formed which are not capable of moving through the pores of the gel. The smearing expected from the purified RecA is visible for the RecAf from NEB; very little DNA remains in the well but no clear band is visible.

The experiment was repeated and the DNA was run on a second agarose gel, shown in Figure 7.12 (B). In this instance the DNA ladder used was more appropriate and confirms that the 3.5 kbp DNA is moving at an appropriate speed for its size.

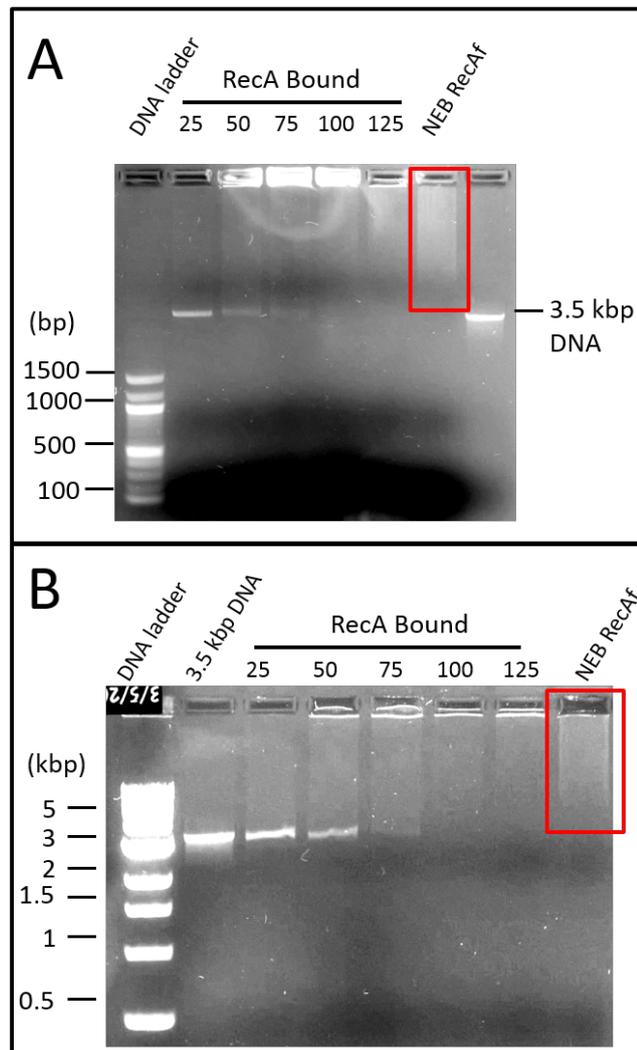


Figure 7.12: **A:** 1% agarose gel electrophoresis showing a RecA activity assay. Lanes from left to right: a 100 bp DNA ladder, 3.5 kbp DNA nucleoprotein filaments with RecA at a concentration of 25% to 125% of the protein calculated to be necessary for completely covering the DNA, 3.5 kbp DNA nucleoprotein filaments with RecA from NEB at 100% of the concentration necessary for completely covering the DNA, bare 3.5 kbp DNA. The smear on the gel created by the binding of the NEB RecA is highlighted by a red box. **B:** Agarose gel electrophoresis showing a RecA activity assay. Lanes from left to right: a 1 kbp DNA ladder, bare 3.5 kbp DNA, 3.5 kbp DNA nucleoprotein filaments with RecA at a concentration of 25% to 125% of the protein calculated to be necessary for completely covering the DNA, 3.5 kbp DNA nucleoprotein filaments with RecA from NEB at 100% of the concentration necessary for completely covering the DNA. The smear on the gel created by the binding of the NEB RecA is highlighted by a red box.

The smearing caused by NEB RecAf does not appear as prominently in this gel. It can be seen that the binding of purified RecA to the DNA continues to cause retention in the well indicating the formation of aggregates. The buffers of both NEB RecAf and the RecA produced contained 1 mM DTT, precluding the possibility that aggregation was the result of the formation of disulphide bonds between the monomers.

7.6 Conclusions

The expression and purification from protein contamination of RecA expressed with a His-tag has been successful. However, the protein obtained does not show the desired activity.

For the purposes of AFM glycerol was not included in the protocol. Glycerol is a cosolvent and stabilises proteins in solution, for this reason it is widely used in molecular biology for protein storage buffers [258, 259]. It forms a solvation layer around the protein which is extremely difficult to remove by dialysis and interferes with the probe during AFM. It was expected that the stability of the protein would be affected by the omission of glycerol, but it was hoped that it would be stable enough to use in the short period following purification. The activity assays shown in Figure 7.12 were carried out in the 48 hours following purification and the protein was not active. It could be the absence of glycerol that makes the protein too unstable for storage at 4 °C.

The lack of activity could also be independent of the glycerol, possibly indicating a low level of folding to the native state. In this event the incorrect structure would maintain solubility and access to the affinity tag but in some way affect the DNA binding activity. The use of circular dichroism and native gel analysis could have given more information about the protein folding.

There also seems to have been a large amount of DNA contamination. This could have had an effect on the activity of RecA or perhaps the formation of large aggregates in the binding experiments. If the DNA contamination reduced binding to the 3.5 kbp DNA through competition then a smear on the gel should have been visible, but at a lower molecular weight. No smear of this kind was observed.

It is possible that there were ssDNA strands present from the cell lysate and that small regions of sequence similarity with the 3.5 kbp DNA allow the formation of a sort of matrix. RecA is able to simultaneously bind ssDNA and dsDNA and, although affinity is highest for complete sequence homology, triple strand formation is possible in the presence of base pair mismatches [215]. This could explain the retention of samples in the wells, or it could simply be the effect of protein aggregation. The fact that the DNA also remains in the well with the protein indicates some level of binding.

Given the level of DNA contamination it did not seem viable to produce a mutant lacking the C terminus, which regulates DNA binding. The removal of the DNA from the wild-type protein was not successful, the C terminal deletion would increase the RecA binding affinity for DNA and increase the problem. Alternative protocols exist for the purification of RecA from *E. coli* expression [260].

This project was not carried further; it was decided that the activity of the protein was too low to be of use in AFM studies. The study of the effect of DNA topology on RecA patterning efficiency was carried out with RecAf which is widely available and increases the repeatability and potential scope of these results.

Chapter 8

Discussion

8.1 Addressing DNA nanotechnology

DNA nanotechnology began as an effort to create a matrix that would align proteins for crystallographic studies [41]. In the decades since it has shown the potential to advance a wide range of fields as a tool in medicine[3], computing [261] and materials science [48] among many other fields. DNA can self assemble into two- and three-dimensional motifs which can be developed further into structures that can move and respond to stimuli. There are a wide range of biological tools available for manipulating DNA providing flexibility and mutability.

The key to utilising DNA nanotechnology is the ability to target functional groups to them in a precise and accurate manner. Functional groups can turn DNA structures into nanoscale assembly lines or biosensors through positioning components in a controllable way on a 2D or 3D surface. The capacity to specifically place functional elements provides the ability to optimise structures for different roles. One example is the distance between fluorescent groups, which can have an exponential effect on quenching; this makes precise and accurate placement extremely important to their role, for example as a reporter.

There are a range of techniques currently used for directing functional groups to these structures, and additional routes which may be used in the future. However, there are limitations to these techniques, in particular there is rarely any low-cost flexibility in

altering targeting.

DNA toeholds are currently the most utilised method for attaching functional groups to DNA structures. They are inexpensive and are designed into the structure but there are a number of reasons why they are a flawed approach to augmenting DNA structures. Short complementary strands of DNA have relatively low melting temperatures, limiting the applications of structures utilising this method for materials in which heat stability is necessary. The use of a tether allows movement and reduces the accuracy of placement, shorter tethers provide a less stable platform for binding. Most importantly, the toeholds are designed into the DNA structure and moving them by even a single base requires redesign and new oligomers.

DNA aptamers have been used by a number of groups to attach proteins to surfaces made of DNA due to their specificity and low cost [60, 133]. While it is true that aptamers can be very specific binders, it is not possible to predict off-target binding without extensive testing, slowing the application of these structures to new environments. The development of aptamers can be a lengthy, iterative process as it relies on stochastic generation of binders. The structure of the aptamer cannot be predicted or designed for a given target.

Base substitutions and functionalisation could provide a method for introducing functionality directly into the structure of the DNA. A wide range of chemical groups are available for inclusion and could act as anchors for further chemical reactions [119]. The development of this technology is in its early stages and it is possible that the additional species could act as a barrier to inclusion in complex structures in which the DNA is under stress and proximity to other strands is high. The synthesis of non-natural DNA structures is much more expensive than the use of toeholds or aptamers and attempts at production using polymerases have not yielded strands as long as those conventionally used in origami structures. These structures would also be immutable, requiring redesign and a new synthesis to make any changes.

Small molecules are currently in development for therapeutic applications, generating *in vivo* changes in gene expression [135]. They bind to topologies in a similar fashion to aptamers and are largely targeted to the minor groove of DNA [135]. This confers the potential ability to target to even the most high-tension regions of DNA structures. Small

molecules suffer from the same development lag as aptamers and are not under research for this application. The target area is often to 4-5 bases, reducing the accuracy of binding to a large structure [135].

The work in this thesis is built on results showing it is possible to site-specifically pattern dsDNA using the bacterial protein RecA [15, 16, 17, 18]. Patterning has been shown to be possible using nanometre-scale ssDNA oligomers [18] and at multiple points on the same dsDNA strand [16]. RecA has been utilised in DNA nanotechnology for metal deposition [13, 14], biosensors [230] and for computing roles [231, 232].

All of this is possible due to the sequence specificity of RecA mediated triple-strand formation; the complex is guided by the ssDNA strand to a region of sequence homology in dsDNA, giving precise and accurate targeting. This targeting makes it flexible, allowing costly designs to be reused and small changes to targeting are easily implemented. The production of RecA is established and it is available from large manufacturers. The use of RecA could allow the recruitment of other proteins or the attachment of functional groups, placed with the resolution of a single DNA base. Through the use of non-hydrolysable ATP analogues like ATP γ S the complex can be stabilised for extended periods.

8.2 Results

The aim of the project was to investigate RecA patterning of more complex DNA structures in order to assess its potential applications in the field of DNA nanotechnology. In order to demonstrate the potential of RecA for patterning complex DNA structures, it was necessary to create a range of DNA topologies as scaffolds for triple-strand formation. The use of RecA on linear and circular dsDNA strands has been demonstrated. It was thought that changing the topology of the DNA strand would affect the efficiency of patterning and this might reveal how RecA might be utilised in more complex structures.

Three DNA constructs containing three-way junctions and internal base mismatches were created for use with RecA. The 100 base structure explored the effect of internal mismatches on patterning. The 200 and 270 base structures were designed to explore the effect of three-way junctions at the termini on the efficiency of RecA patterning.

Patterning of the 100 base structures provided interesting data and some insight into what might be necessary to utilise RecA patterning for targeting of complex DNA structures. Some data was obtained from patterning of the 200 base structure with ssDNA regions at the termini. Interestingly, the structure with both three-way junctions could not be formed while the structures with single three-way junctions formed with high efficiency. The 270 base structure could be formed with both three-way junctions and when imaged appeared to have excellent homogeneity.

Additional work aimed to produce recombinant RecA in order to explore some of the mutations available which have been shown to increase the stability of the nucleoprotein filament on dsDNA [162, 165, 172, 173, 174, 175]. A new protocol for production of RecA was implemented but patterning experiments using the protein were unsuccessful.

8.2.1 100 base structures

The results shown in this thesis demonstrate that it is possible to pattern with RecA more complex DNA structures than has been shown before. In Chapter 4, a series of 100 base structures were synthesised to contain regions of base pair mismatching, all shown diagrammatically in Figure 4.2. It was thought that the introduction of a ssDNA region in the dsDNA strand could have a positive or negative effect on the efficiency of patterning.

When the mismatch was adjacent to the patterning site, shown in Figure 4.6, an increase in the number of mismatches increased the patterning efficiency in a saturable manner. Mismatches ≥ 10 bases resulted in patterning efficiencies of $\sim 100\%$. When the mismatch was moved up to 9 bases away from the patterning site a more complex relationship was revealed, indicating that regions of ssDNA had an inhibitory effect on RecA patterning. The 12 base mismatch was transposed across the dsDNA strand to 4 and 8 bases away from the patterning site, shown in Figure 4.8 This resulted in a reduction in patterning efficiency to 47% (± 6) and 32% (± 21) respectively.

These results were further expanded through the introduction of a nick in the phosphate backbone of one strand, shown diagrammatically in Figure 4.12 (A). This topology was shown to also increase the patterning efficiency of RecA from 40% (± 9) for fully base paired dsDNA to 85% (± 0.3) when the nick was proximal to the patterning region. When

a 5 base region of ssDNA was produced proximal to the patterning site, flanked by dsDNA on either side, this increased patterning efficiency to 91% (± 3).

In light of these results, one hypothesis is that the mismatched regions and the nick act to relieve tension induced by the underwinding of the dsDNA that is in a RecA triple-strand. The reduction in tension has a stabilising effect on the triple strand and in this way is able to increase patterning. An alternative hypothesis proposes that the effect of the mismatched region is due to increased DNA breathing in the patterning area. This could have the effect of increasing the accessibility of the dsDNA to the nucleoprotein filament. It is also possible that both of these effects are acting to increase the patterning efficiency of RecA in these experiments.

If it proves to be the case that an increase in tension will have a detrimental effect on efficiency of RecA patterning, this could have an impact on the use of RecA for targeting complex DNA structures. However, if it were necessary to pattern with RecA close to the ends of staple strands, it would still present more flexibility than the use of DNA toeholds. When patterning was carried out on the nicked structures, the movement of the nick by five bases did not have a significant detrimental effect on patterning efficiency; the level of protected DNA was found to be 85% (± 0.3) when the nick was directly proximal to the patterning site and 89% (± 7) when it was moved 5 bases further away. If this were proved to be the limit to the distance of the nick from the patterning site it would give a potential movement of 10 bases that is not available when using DNA toeholds.

8.2.2 200 base structures

In Chapter 5 a cycling ligation reaction was used to create a 200 base structure containing terminal ssDNA regions. The aim was to anneal additional oligomers to these regions to create three-way junctions, as shown in Figure 5.1. These structures were developed to increase our understanding of the impact a more constrained topology might have on RecA patterning. When the additional oligomers are annealed the junction becomes more closed and this could effect the ability of RecA to form a triple-stranded complex. Many DNA structures are made through the repeated crossing over of strands, such as the origami designs seen in Figure 2.6. This creates closed regions in which tension would

be less able to release and RecA may be limited in its activity.

It did not prove possible to anneal the additional oligomers at both ends of the structure simultaneously to produce two double stranded three-way junctions. Annealing an oligomer to create either the 5' or 3' junction did prove possible with high efficiency. A number of hypotheses were generated to explain the inability to form the double junction structure. Modifications to try to account for tension and twist in the pairing of the DNA were developed but these were found experimentally to be ineffective nor did additional bases at the junction point appear to increase yield. There is not currently a commercially available program which assess structures of this kind, which could have positively identified the cause of the annealing problem. Attempts to confirm the production of these structures using AFM produced no clear images due to their small size.

The sequence of the bases in the unpaired junctions were designed to contain a series of restriction enzyme recognition sites; it is possible that this was the design error that led to the difficulty creating three-way junctions. Recognition sequences are palindromic, a sequence motif that can lead to the formation of hairpins. It was also found that reheating the structures after LCR synthesis to temperatures above 74 °C led to changes in their behaviour in the gel, indicating a change in structure. Although DNA hairpins can be melted, in combination with the limits to the temperature they could be subjected to, it is possible that this was the cause of the failure to anneal the junction oligomers.

RecA patterning was carried out on the structures with and without the ssDNA regions at the termini. Results showed that the presence of these regions increased the efficiency of patterning. While the 140 base dsDNA at the centre of the structure showed a patterning efficiency of 18%, the structure with ssDNA termini at both ends was patterning with 42% efficiency.

The patterning region for this structure was at the centre of the construct, covering the XapI site shown in Figure 5.2, at a 55 base distance from the ssDNA regions at the termini. The findings for the 100 base structures indicate that the effect of tension release may not be felt over this distance. An alternative hypothesis was created in which increased DNA breathing as a result of the single-stranded regions was the reason for the increased patterning efficiency, although this effect may also not be felt at this distance. It is possible

that this transient unpairing of the termini increases accessibility of the DNA to RecA. It has been shown that RecA has a role in the repair of stalled replication forks [198], structures which may resemble the ssDNA junctions in this work, although *in vivo* it is recruited by a series of proteins not included in our experiments.

While these results did not produce much insight into the potential uses of RecA in DNA nanotechnology, they do highlight the difficulty of producing entirely novel DNA structures. The first design rules for DNA structures published described the production of short four-way junctions [49], inspired by Holliday junctions in fungi [51]. From this inception a wide range of structures and rules have been developed and written into programs such as caDNAno [262]. This program uses DNA helices as building blocks to create 2D and 3D structures in much the same way that a house or raft might be built of logs; this prevents them from being utilised for structures with a different basic geometry, like those used in this work. A program that incorporates rules including those described here for high efficiency RecA patterning would allow easy design of addressable complex DNA structures.

8.2.3 270 base structures

A novel series of structures containing three-way junctions were successfully created using a different synthesis process and a larger sequence. Chapter 6 describes how dsDNA strands were created using PCR to include phosphorothioate modifications which could then be treated with T7 exonuclease to create ssDNA strands, as shown by Tosch *et al.* [243]. Additional oligomers were annealed to these strands to produce structures with three-way junctions, shown in Figure 6.1.

AFM images showed the efficiency of formation of the structure with both three-way junctions was 75% (± 4). The analysis was carried out manually and relied on the correct orientation of the structures and on the discretion of the analyst; it is probable that this was an underestimation of the efficiency of formation.

The ease with which these structures were created is remarkable given the difficulties producing the 200 base structures. The ssDNA regions on the 270 base structures were significantly longer than for the 200 base structures; each 'arm' was 50 bases long,

20 bases longer than those of the 200 base structure. The longer 'arms' could help stabilise a structure that is otherwise enthalpically unfavourable. Additionally, the three-way junctions were annealed in parallel with the central region of the structure, allowing for a much higher melting temperature when the junction was formed. This could have the effect of removing all secondary structure from the ssDNA 'arms' and facilitating annealing.

Experiments were carried out to pattern these structures with RecA following the protocol used successfully previously in this work and in extensive previous research [15, 16, 17, 18]. Very little success was seen in this line of experimentation. A number of approaches were taken in an attempt to identify the reason for this lack of success and it remains unclear. All reagents were discarded and purchased again, the buffers were made from the newly purchased reagents, oligomers were ordered and resuspended in newly-created buffers. The plasmids used as PCR templates were sequenced and a single base mutation was discovered in one gene at a point over 50 bases from the patterning site. The protocols followed were the same as those used for the 100 base structures and AFM images showed that the desired structure was being formed.

Patterning was tested with dsDNA with no ssDNA elements or junctions and failed to show significant levels of protection. This protocol has been used successfully on dsDNA within this project and published work. This indicated that there was a problem with the patterning reaction due to an extrinsic factor. It does not seem likely that the failure to pattern efficiently was the result of this sequence as RecA has been shown to work with a wide variety of sequences. The melting temperature of the patterning oligomer was found to be within a few degrees of the oligomer used for the 100 base structures.

A low level of patterning was seen with a 40 base patterning oligomer. In experiments with the dsDNA PCR products, the non-complementary control oligomer showed undigested DNA made up 2.4% ($\pm 0.5\%$, $n=2$) of the total DNA. When a 30 base oligomer was used 6.9% ($\pm 3.4\%$) of the DNA remained undigested, showing no significant difference from the control. Patterning of this dsDNA with two 40 base oligomers showed 7% and 19% of the total DNA remained undigested. While the 40 base oligomer showed an improvement in yield over the 30 base oligomer, this level of patterning is still very unfavourable when compared to the 40% (± 9) obtained with the 100 base dsDNA.

8.2.4 RecA production

To produce RecA to our own specifications a protocol was designed. The results are discussed in Chapter 7. The ability to produce RecA would allow us to explore a number of changes to the protein that have been found to improve binding of the nucleoprotein filament to dsDNA. This includes removing the ATPase activity of the protein and a number of mutations at the protein-protein interface that increase filament stability [162, 165, 172, 173, 174, 175]. The hydrolysis of ATP is linked to the disassembly of the strand exchange complex [176], removing this capability might further stabilise the RecA complex. The C terminus of the RecA protein has been shown to regulate its ability to bind DNA [162, 225, 164]; removing this domain could further increase the stability of the triple-stranded complex. Experiments were initially carried out with wild-type RecA in order to establish the protocol.

An affinity tag was included in the plasmid vector and fast protein liquid chromatography (FPLC) was used to capture the protein. A method for the purification of RecA was published in 1981 but this protocol is very time consuming [260]; it was thought that new technology might enable a simpler and faster protocol. Using small scale tests it was established that the affinity tag should be placed at the N terminus of the protein. Histidine is a positively charged amino acid and the C terminus of RecA has been found to be largely negatively charged [163]. It is possible that this caused misfolding which led to little soluble protein observed when the histidine chain was placed at the C terminus.

Larger scale experiments were carried out which showed that DNA contamination was a problem which was combated with an ATP wash. Despite producing protein that was sufficient to our needs, it proved to have little activity, possibly due to the difficulty of removing DNA bound into the DNA binding regions of the protein. The glycerol usually added to RecA samples was omitted so that better AFM images might be obtained. It is possible that in the absence of this stabilising agent the RecA did not maintain its correct folding. The trials were not taken further than this point.

It is possible that more extensive analysis could reveal the source of the RecA inactivity, for example an ATPase assay might show if the enzymatic function was deficient. If it did prove to be a problem with DNA contamination there are a number of potential

steps to be taken. There are well established protocols for removing DNA contamination through the use of addition chromatography columns, for example a heparin column or anion exchange chromatography. The use of benzonase nuclease, an alternative to DNaseI, is also recommended for removing DNA from samples. Protein refolding, involving treatment with denaturing chemicals followed by introduction into a refolding environment, could be used to further purify the protein. It is possible that endonuclease treatment combined with ATP could induce sufficient turnover in the RecA binding sites and remove DNA contamination.

8.2.5 Conclusions

The experiments in this work have been done almost exclusively with 30 base patterning oligomers, approximately 10 nm in length. Previous work has shown that using 30 base patterning oligomers results in patterning efficiencies similar to those achieved when using 40 base oligomers, while the use of 20 base oligomers leads to a reduction in efficiency [18]. The length of 30 bases was chosen on the basis of these results, representing a compromise between patterning efficiency and the use of smaller filaments for spatial efficiency. Higher patterning efficiencies were observed with a 40 base oligomer for the 270 base structure. Perhaps when efficiencies are lower the effect of the higher cumulative binding energy of a longer oligomer and more RecA monomers is more pronounced. In the event that tension is a factor in patterning efficiency, it seems likely that a longer oligomer would result in a larger region of patterning-induced tension.

The method by which RecA patterning efficiency is increased through the introduction of base mismatches and terminal ssDNA regions is unclear. The work from Chapter 4 could implicate tension reduction but there is no direct evidence to confirm that tension release is driving the changes in patterning efficiency observed here. It is also possible that the ssDNA region increases the accessibility of the region to RecA patterning; research has shown that patterning is more efficient towards the termini of the dsDNA strand [16]. An increase or decrease in DNA breathing may also play a role in changes in patterning efficiency. It is likely that these observations are the result of a combination of effects and not a single cause.

Including regions of ssDNA to facilitate RecA patterning of complex DNA structures could have a destabilising effect. However, a similar effect was seen when a nick is present; nicks are present throughout DNA origami structures without destabilising. If nicks are sufficient to enable high efficiency RecA patterning of complex DNA structures it would represent a large step forward in utilising RecA for DNA nanotechnology.

There is some research on DNA strands with a small number of overhangs, which could provide some insight into the structures produced in Chapters 5 and 6. It has been demonstrated that a single or pair of overhanging bases, referred to in the literature as a dangling end, in most cases stabilises dsDNA strands [263, 264, 265, 266]. The second binding loop for RecA has a strong preference for dsDNA; if dsDNA strand stabilisation is occurring as a result of longer ssDNA regions then this could be a preferred binding substrate.

While short dangling ends stabilise dsDNA strands, it is also possible that the longer overhangs used here are increasing DNA breathing. If this were the case then it could increase accessibility of the DNA to RecA; DNA breathing has been linked to increased DNA accessibility at the replication fork for T4 helicase [254]. If DNA breathing was increased at the site of ssDNA regions then it would be possible to show this through the use of hydrogen exchange experiments [267].

Although there are some potential limitations, the flexibility of RecA as a method for addressing DNA structures makes it an interesting prospect for DNA nanotechnology. The recent use of RecA in DNA circuits and biosensors shows additional potential roles for RecA and possibly other DNA binding proteins [230, 231, 232].

8.3 Future Outlook

The future of DNA nanotechnology depends on finding concrete applications for the technology. The production of complex and fascinating structures is not sufficient; without demonstrable successes funding will eventually cease, especially so if the current focus on monetisation continues. DNA computing is very much in its infancy, despite the ability of DNA strands to carry out simultaneous calculations. Medical and materials

science appear to be the most likely areas in which a breakthrough might occur, and for these applications RecA may well play a role.

For example, the inclusion of RecA in a microfluidic biosensor has a tangible positive effect on the process [230]. RecA amplifies the signal produced by pathogen detection by binding to ssDNA probes and increasing the output signal when dsDNA targets were detected. Biosensors are a large and growing field which could be in circulation in the foreseeable future.

RecA could be used for fast prototyping of arrays of functional groups to target groups to DNA nanostructures where distance plays an important role; the ability to easily move an element by 0.34 nm, the distance between two bases, without any need for redesign could significantly increase speed in the early stages of this work.

Although the limitations of DNA toeholds have been discussed at length here, this is still the most widely used method for attaching to DNA structures. Researchers seem content to accept the cost of redesign when reliability and predictability increase the speed of outcomes. This does not mean that there can be no advances in the field, only that it must be of such sufficient quality that it surpasses the currently held standard.

The further development of this work lies in demonstrating the repeatable and efficient patterning of complex DNA structures with RecA. The first aim should be the identification of the obstacle to RecA triple-strand formation on the structures containing three-way junctions. Once this has been identified, the results from the patterning experiments should reveal the potential and limitations of RecA in this field. Following work with three-way junctions, it could be possible to test the use of RecA with more complex DNA structures, including arrays and origami. Future patterning work should include the use of additional sequences to confirm that these observed effects are not sequence-dependent. Particular emphasis should be placed on the melting temperatures of the sequences used, to investigate further the effects of DNA breathing on RecA patterning. The use of optical tweezers to increase the tension of the strands as they are patterning could also aid in understanding the patterning process.

The field of DNA origami has leapt forward through the use of computer aided design, described in Douglas *et al.* (2009) [262]. Software is now freely available which allows

the user to plan 2D and 3D structures with relative ease. The programs use strict rules on the distance between allowed crossovers to reduce the tension in a structures by adhering as closely as possible to ideal helix proportions. In effect, the DNA helix becomes a cylindrical building block with a limited number of permissible connections. This is very effective for a particular class of DNA structure. Although the currently available software would not have been suitable for designing the structures discussed in this thesis, it will be very useful in further development of structures for DNA patterning and in opening the field to new research groups.

The literature identifies restriction enzyme assay as reasonable method for monitoring patterning efficiencies. It seems prudent to explore additional techniques for assessing triple-strand formation, as confirmation of our findings and to gain new data. Many studies have used Förster resonance energy transfer (FRET) and fluorescent tags to measure binding events both for ensemble measurements and the observation of single molecules [151, 183, 202]. A FRET assay could be carried out by attaching a probe to the ssDNA targeting oligomer and a site on the dsDNA in close proximity to the patterning region. This would allow us to monitor the rate of patterning and changes in patterning levels over time. The reaction could be measured in real time instead of through gel analysis. It is possible this could cast new light on the experiments discussed in this thesis; its application would undoubtedly present a new series of challenges. The use of split enzymes as reporters of binding is a field of growth and could be implemented in a similar manner.

One aim of any future work should be in increase the efficiency of patterning on any substrate. To this end, RecA mutants could be a great aid in increasing triple-strand stability. Obtaining a mutant lacking the C terminus would allow work to assess whether this might increase patterning efficiency for more complex structures, like those seen here. A range of mutants with, for example, an inability to hydrolyse ATP or an increased affinity for DNA could be utilised to increase yields in this system [162, 165, 172, 173, 174, 175].

If Richard Feynman were to look to the nanoscale with the knowledge that we have now, perhaps he would see a landscape filled with DNA architecture. With his foresight we might see the nanoscale materials, assembly lines and medical treatments of the future.

The data in this work and in published research indicates that RecA may well play a role in the breadth of activities that DNA can carry out, providing a precise and accurate route to addressing DNA structures.

"Ever tried. Ever failed. No matter. Try Again. Fail again. Fail better."

Westward Ho - Samuel Beckett

Appendices

A DNA Sequences

A.1 100 base sequences

100

5' TTT TGA TAG C TT CAA GCC AG A GTT GTC TTT TTC TAT CTA C TC TCA
TAC AA C CAA TAA ATG CTG AAA TGA A TT CTA AGC GG A GAT CGT CTA
GTG ATT TTA A 3'

100 Reverse Complement

5' AAA ACT ATC G AA GTT CGG TC T CAA CAG AAA AAG ATA GAT G AG AGT
ATG TT G GTT ATT TAC GAC TTT ACT T AA GAT TCG CC T CTA GCA GAT CAC
TAA AAT T 3'

100 (2)

5' TTT TGA TAG CTT CAA GCC AGA GTT GTC TTT TTC TAT CTA CTC TCA TAC
ATA CAA TAA ATG CTG AAA TGA ATT CTA AGC GGA GAT CGT CTA GTG ATT
TTA A 3'

100 (6)

5' TTT TGA TAG CTT CAA GCC AGA GTT GTC TTT TTC TAT CTA CTC TCA TAT
GTA GTA TAA ATG CTG AAA TGA ATT CTA AGC GGA GAT CGT CTA GTG ATT
TTA A 3'

100 (8)

5' TTT TGA TAG CTT CAA GCC AGA GTT GTC TTT TTC TAT CTA CTC TCA TCT
GTA GTG TAA ATG CTG AAA TGA ATT CTA AGC GGA GAT CGT CTA GTG ATT
TTA A 3'

100 (10)

5' TTT TGA TAG CTT CAA GCC AGA GTT GTC TTT TTC TAT CTA CTC TCA ACT
GTA GTG CAA ATG CTG AAA TGA ATT CTA AGC GGA GAT CGT CTA GTG ATT
TTA A 3'

100 (12)

5' TTT TGA TAG C TT CAA GCC AG A GTT GTC TTT TTC TAT CTA C TC TC C
CCT GT A GTG CC A ATG CTG AAA TGA A TT CTA AGC GG A GAT CGT CTA
GTG ATT TTA A 3'

100 (15)

5' TTT TGA TAG CTT CAA GCC AGA GTT GTC TTT TTC TAT CTA CTC TCC CCT
GTA GTG CCG TTG CTG AAA TGA ATT CTA AGC GGA GAT CGT CTA GTG ATT
TTA A 3'

100 (20)

5' TTT TGA TAG C TT CAA GCC AG A GTT GTC TTT TTC TAT CTA C TC TCA
TAC AA C CAA TAA ATG CTG AAA TGA A TT CTA AGC GG A GAT CGT CTA
GTG ATT TTA A 3'

100 (2)N

5' TTT TGA TAG C TT CAA GCC AG A GTT GTC TTT TTC TAT CTA C TC TCA
TAC AA C CAA TAA ACA CTG AAA TGA A TT CTA AGC GG A GAT CGT CTA
GTG ATT TTA A 3'

100 (6)N

5' TTT TGA TAG C TT CAA GCC AG A GTT GTC TTT TTC TAT CTA C TC TCA
TAC AA C CAA CGA CGT CTG AAA TGA A TT CTA AGC GG A GAT CGT CTA
GTG ATT TTA A 3'

100 (8)N

5' TTT TGA TAG C TT CAA GCC AG A GTT GTC TTT TTC TAT CTA C TC TCA
TAC AA C CCT GCT CAA CTG AAA TGA A TT CTA AGC GG A GAT CGT CTA
GTG ATT TTA A 3'

100 (10)N

5' TTT TGA TAG C TT CAA GCC AG A GTT GTC TTT TTC TAT CTA C TC TCA
TAC AA G GTT CGC GGT CTG AAA TGA A TT CTA AGC GG A GAT CGT CTA
GTG ATT TTA A 3'

100 (12)N

5' TTT TGA TAG CTT CAA GCC AGA GTT GTC TTT TTC TAT CTA CTC TCA TAC
CCG TGT AGT GCC CTG AAA TGA ATT CTA AGC GGA GAT CGT CTA GTG ATT
TTA A 3'

100 (15)N

5' TTT TGA TAG C TT CAA GCC AG A GTT GTC TTT TTC TAT CTA C TC TCA
ATG TG T AGT GCC GGT CTG AAA TGA A TT CTA AGC GG A GAT CGT CTA
GTG ATT TTA A 3'

100 (12)F

5' TTT TGA TAG CTT CAA GCC AGA GTT GTC TTT TTC TAT CTA CTG CGT ATA
TTG GCA TAA ATG CTG AAA TGA ATT CTA AGC GGA GAT CGT CTA GTG ATT
TTA A 3'

100 Nick 55

5' TTT TGA TAG CTT CAA GCC AGA GTT GTC TTT TTC TAT CTA CTC TCA TAC
AAC CAA T 3'

100 Nick 45

5' AA ATG CTG AAA TGA ATT CTA AGC GGA GAT CGT CTA GTG ATT TTA A 3'

100 Nick 60

5' TTT TGA TAG CTT CAA GCC AGA GTT GTC TTT TTC TAT CTA CTC TCA TAC
AAC CAA TAA ATG 3'

100 Nick 40

5' CTG AAA TGA ATT CTA AGC GGA GAT CGT CTA GTG ATT TTA A 3'

100 20 base patterning

5' CTG AAA TGA ATT CTA AGC GG 3'

100 20 base patterning reverse complement

5' CCG CTT AGA ATT CAT TTC AG 3'

100 30 base patterning

5' CTG AAA TGA ATT CTA AGC GGA GAT CGT CTA 3'

100 30 base patterning reverse complement

5' TAG ACG ATC TCC GCT TAG AAT TCA TTT CAG 3'

100 40 base patterning

5' CTG AAA TGA ATT CTA AGC GGA GAT CGT CTA GTG ATT TTA A 3'

100 40 base patterning reverse complement

5' TTA AAA TCA CTA GAC GAT CTC CGC TTA GAA TTC ATT TCA G 3'

A.2 200 base sequences**Full top strand (5' to 3')**

5' CGA GCT CGC CAT ATG GCT AGT CTA GAC TA TTT TGA TAG CTT CAA GCC
AGA GTT GTC TTT TTC TAT CTA CTC TCA TAC AAC CAA TAA ATG CTG AAA
TGA ATT CTA AGC GGA GAT CGC CTA GTG ATT TTA AAC TAT TGC TGG CAG
CAT TCT TGA GTC CAA TAT AAA AGT AT CCC AAG CTT GGG CGG ACT AGT
CCG 3'

Full bottom strand (5' to 3')

5' GTG CAG TTC TGT ACC GGT TTT CCT ATA CTT TTA TAT TGG ACT CAA GAA
TGC TGC CAG CAA TAG TTT AAA ATC ACT AGG CGA TCT CCG CTT AGA ATT
CAT TTC AGC ATT TAT TGG TTG TAT GAG AGT AGA TAG AAA AAG ACA ACT
CTG GCT TGA AGC TAT CAA AA ACC GGT ATC CCC GCC CGA CAA CGA CGG
CA 3'

Top left (5') oligomer

5' TTT TGA TAG CTT CAA GCC AGA GTT GTC TTT TTC TAT CTA CTC TCA TAC
AAC CAA TAA ATG CTG AAA TGA A 3'

Top right (3') oligomer

5' TTC TAA GCG GAG ATC GCC TAG TGA TTT TAA ACT ATT GCT GGC AGC
ATT CTT GAG TCC AAT ATA AAA GTA T 3'

Bottom right (5') oligomer

5' ATA CTT TTA TAT TGG ACT CAA GAA TGC TGC CAG CAA TAG TTT AAA
ATC ACT AGG CGA TCT CCG CTT AGA A 3'

Bottom left (3') oligomer

5' TTC ATT TCA GCA TTT ATT GGT TGT ATG AGA GTA GAT AGA AAA AGA
CAA CTC TGG CTT GAA GCT ATC AAA A 3'

Top left (5') oligomer with junction (TLJ)

5' CGA GCT CGC CAT ATG GCT AGT CTA GAC TAT TTT GAT AGC TTC AAG
CCA GAG TTG TCT TTT TCT ATC TAC TCT CAT ACA ACC AAT AAA TGC TGA
AAT GAA 3'

Top right (3') oligomer with junction (TRJ)

5' TTC TAA GCG GAG ATC GCC TAG TGA TTT TAA ACT ATT GCT GGC AGC
ATT CTT GAG TCC AAT ATA AAA GTA TCC CAA GCT TGG GCG GAC TAG TCC
G 3'

Bottom right (5') oligomer with junction (BRJ)

5' GTG CAG TTC TGT ACC GGT TTT CCT ATA CTT TTA TAT TGG ACT CAA GAA
TGC TGC CAG CAA TAG TTT AAA ATC ACT AGG CGA TCT CCG CTT AGA A 3'

Bottom left (3') oligomer with junction (BLJ)

5' TTC ATT TCA GCA TTT ATT GGT TGT ATG AGA GTA GAT AGA AAA AGA
CAA CTC TGG CTT GAA GCT ATC AAA AAC CGG TAT CCC CGC CCG ACA
ACG ACG GCA 3'

Central (annealed) 140 base region

5' TTT TGA TAG CTT CAA GCC AGA GTT GTC TTT TTC TAT CTA CTC TCA TAC

AAC CAA TAA ATG CTG AAA TGA ATT CTA AGC GGA GAT CGC CTA GTG ATT
TTA AAC TAT TGC TGG CAG CAT TCT TGA GTC CAA TAT AAA AGT AT 3'

Patterning oligomer

5' AAA TGC TGA AAT GAA TTC TAA GCG GAG ATC 3'

A.3 270 base sequences

Phosphorothiations are represented as asterisks. A diagrammatic representation of the structure of phosphorothiated DNA is shown in Figure 3.4.

Full top strand (5' to 3')

5' AAT GAG GCT GAT GAG TTC CAT ATT TGA AAA GTT TTC ATC ACT ACT
TAG TTT TTT GAT AGC TTC AAG CCA GAG TTG TCT TTT TCT ATC TAC TCT
CAT ACA ACC AAT AAA TGC TGA AAT GAA TTC TAA GCG GAG ATC GCC TAG
TGA TTT TAA ACT ATT GCT GGC AGC ATT CTT GAG TCC AAT ATA AAA GTA
TTG TGT ACC TTT TGC TGG GTC AGG TTG TTC TGA AAT AGG TAA TGC AAA
CTT TGG ATA CCA AAA AGC CAT AAC ACC ATG ACG 3'

Full bottom strand (5' to 3')

5' TCG CTT GTT TCA GTT TCG TTT AGT GCA TTT GAT CCT TTT ACT CCT CCT
AAA GAA CAA CCT GAC CCA GCA AAA GGT ACA CAA TAC TTT TAT ATT GGA
CTC AAG AAT GCT GCC AGC AAT AGT TTA AAA TCA CTA GGC GAT CTC CGC
TTA GAA TTC ATT TCA GCA TTT ATT GGT TGT ATG AGA GTA GAT AGA AAA
AGA CAA CTC TGG CTT GAA GCT ATC AAA ATA AAG GGA GAA TAA CGC
TGC CGG ACA TCC ACC AAT AGG TTG CAG TTT CTC 3'

Phosphorothioated upper forward (5') primer

5' A *G *T * C *A *A TAG ATT CGC TAG CAA TGA GGC TGA TGA G 3'

Upper reverse (3') primer

5' CGT CAT GGT GTT ATG GCT TTT TGG TAT CCA AAG TTT 3'

Phosphorothioated lower forward (5') primer

5' A*G*T* C*A*A TAG ATT CGC TAG CTC GCT TGT TTC AGT T 3'

Lower reverse (3') primer

5' GAG AAA CTG CAA CCT ATT GGT GGA TGT CCG 3'

270 junction oligomer (5')

5' GAG AAA CTG CAA CCT ATT GGT GGA TGT CCG GCA GCG TTA TTC TCC
CTT TAA ACT AAG TAG TGA TGA AAA CTT TTC AAA TAT GGA ACT CAT CAG
CCT CAT T 3'

270 junction oligomer (3')

5' CGT CAT GGT GTT ATG GCT TTT TGG TAT CCA AAG TTT GCA TTA CCT ATT
TCT TAG GAG GAG TAA AAG GAT CAA ATG CAC TAA ACG AAA CTG AAA
CAA GCG A 3'

Anneling oligomer for BmtI digest - Upper strand

5' TTG CTA GCG AAT CTA TTG ACT 3'

Anneling oligomer for BmtI digest - Lower strand

5' GAG CTA GCG AAT CTA TTG ACT 3'

B Protein Expression

B.1 RecA DNA Sequence

5' ATG GCT ATC GAC GAA AAC AAA CAG AAA GCG TTG GCG GCA GCA CTG
GGC CAG ATT GAG AAA CAA TTT GGT AAA GGC TCC ATC ATG CGC CTG
GGT GAA GAC CGT TCC ATG GAT GTG GAA ACC ATC TCT ACC GGT TCG CTT
TCA CTG GAT ATC GCG CTT GGG GCA GGT GGT CTG CCG ATG GGC CGT ATC
GTC GAA ATC TAC GGA CCG GAA TCT TCC GGT AAA ACC ACG CTG ACG
CTG CAG GTG ATC GCC GCA GCG CAG CGT GAA GGT AAA ACC TGT GCG
TTT ATC GAT GCT GAA CAC GCG CTG GAC CCA ATC TAC GCA CGT AAA CTG
GGC GTC GAT ATC GAC AAC CTG CTG TGC TCC CAG CCG GAC ACC GGC
GAG CAG GCA CTG GAA ATC TGT GAC GCC CTG GCG CGT TCT GGC GCA
GTA GAC GTT ATC GTC GTT GAC TCC GTG GCG GCA CTG ACG CCG AAA
GCG GAA ATC GAA GGC GAA ATC GGC GAC TCT CAC ATG GGC CTT GCG
GCA CGT ATG ATG AGC CAG GCG ATG CGT AAG CTG GCG GGT AAC CTG
AAG CAG TCC AAC ACG CTG CTG ATC TTC ATC AAC CAG ATC CGT ATG AAA
ATT GGT GTG ATG TTC GGT AAC CCG GAA ACC ACT ACC GGT GGT AAC GCG
CTG AAA TTC TAC GCC TCT GTT CGT CTC GAC ATC CGT CGT ATC GGC GCG
GTG AAA GAG GGC GAA AAC GTG GTG GGT AGC GAA ACC CGC GTG AAA
GTG GTG AAG AAC AAA ATC GCT GCG CCG TTT AAA CAG GCT GAA TTC
CAG ATC CTC TAC GGC GAA GGT ATC AAC TTC TAC GGC GAA CTG GTT GAC
CTG GGC GTA AAA GAG AAG CTG ATC GAG AAA GCA GGC GCG TGG TAC
AGC TAC AAA GGT GAG AAG ATC GGT CAG GGT AAA GCG AAT GCG ACT
GCC TGG CTG AAA GAT AAC CCG GAA ACC GCG AAA GAG ATC GAG AAG
AAA GTA CGT GAG TTG CTG CTG AGC AAC CCG AAC TCA ACG CCG GAT TTC
TCT GTA GAT GAT AGC GAA GGC GTA GCA GAA ACT AAC GAA GAT TTT TAA
3'

B.2 RecA Protein Sequence

Cysteine residues are highlighted in red.

MAIDENKQKA LAAALGQIEK QFGKGSIMRL GEDRSMDVET ISTGSLSLDI
ALGAGGLPMG RIVEIYGPES SGKTTLTLQV IAAAQREGKT CAFIDAEHAL
DPIYARKLGV DIDNLLCSQP DTGEQALEIC DALARSGAVD VIVVDSVAAL
TPKAEIEGEI GDSHMGLAAR MMSQAMRKLA GNLKQSNTLL IFINQIRMKI
GVMFGNPETT TGGNALKFYA SVRLDIRRIG AVKEGENVVG SETRVKVVKN
KIAAPFKQAE FQILYGEGIN FYGELVDLGV KEKLIEKAGA WYSYKGEKIG
QGKANATAWL KDNPETAKEI EKKVRELLLS NPNSTPDFSV DDSEGVAETN EDF

B.3 Plasmid Maps

pET11a

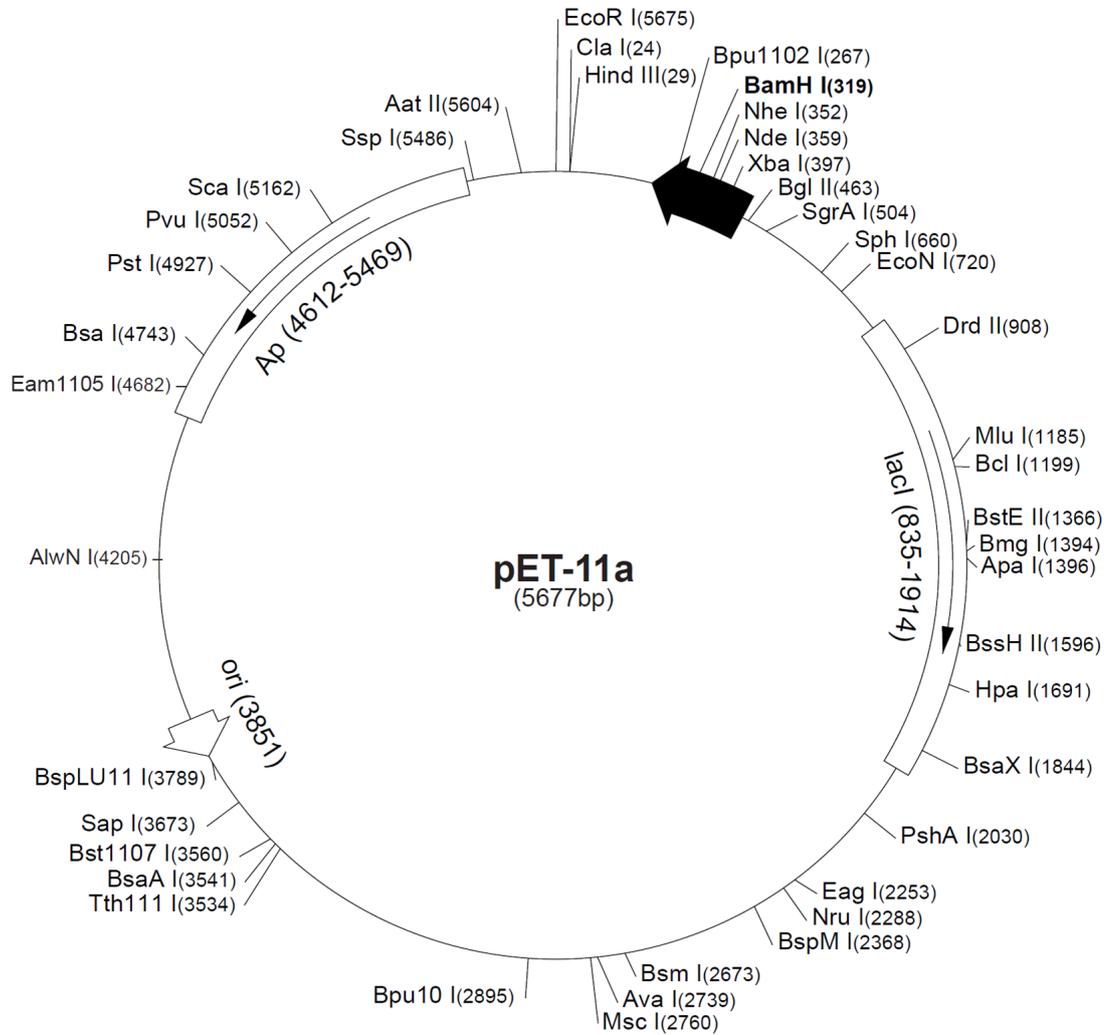


Figure B.1: pET11a plasmid map

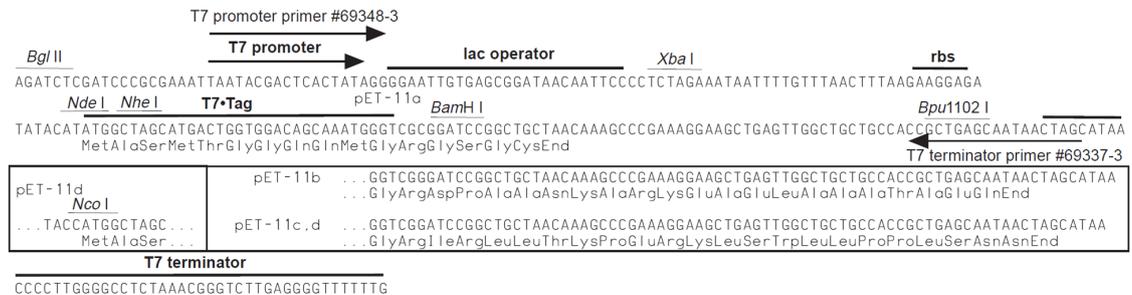


Figure B.2: pET11a cloning/expression region

pET15b

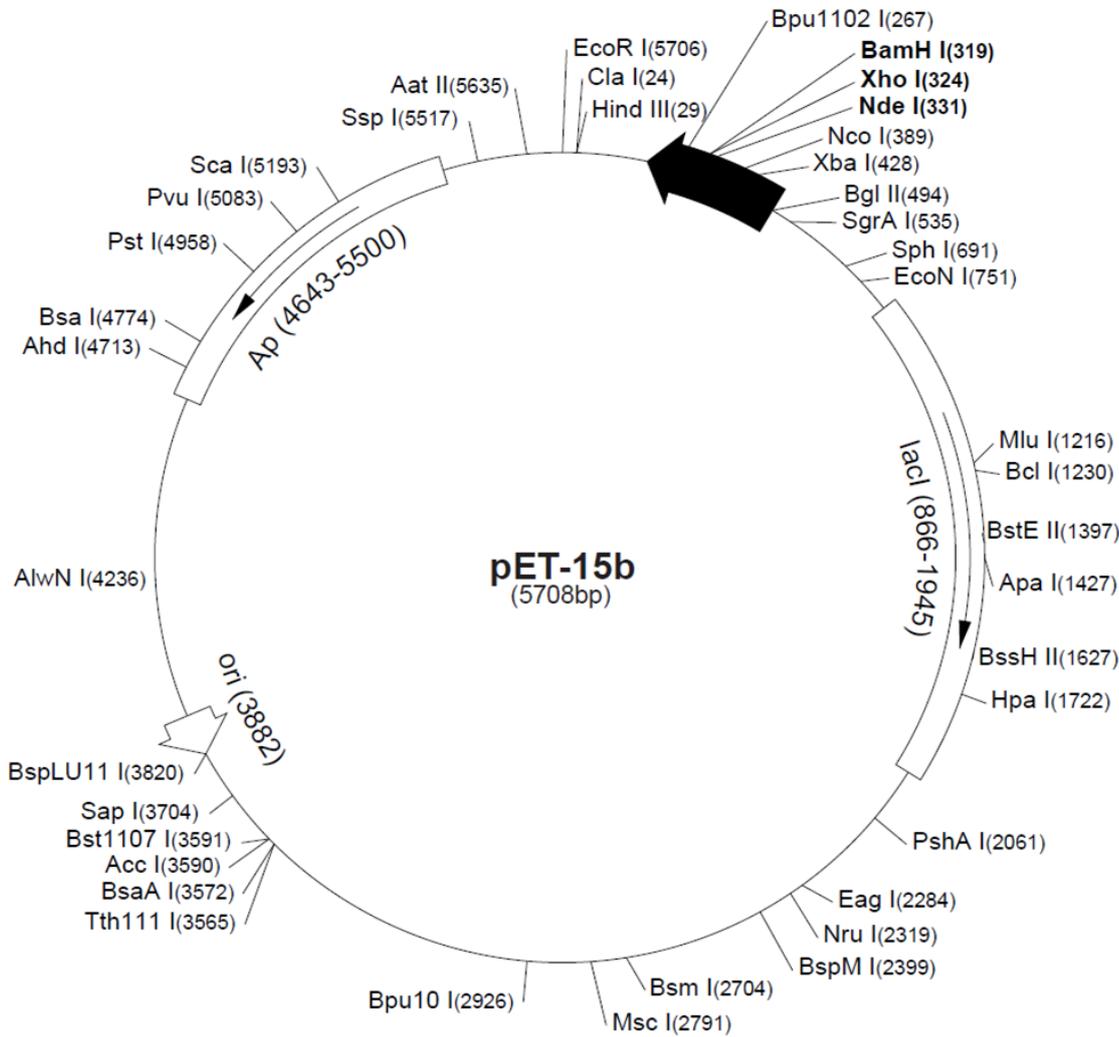


Figure B.3: pET15b plasmid map

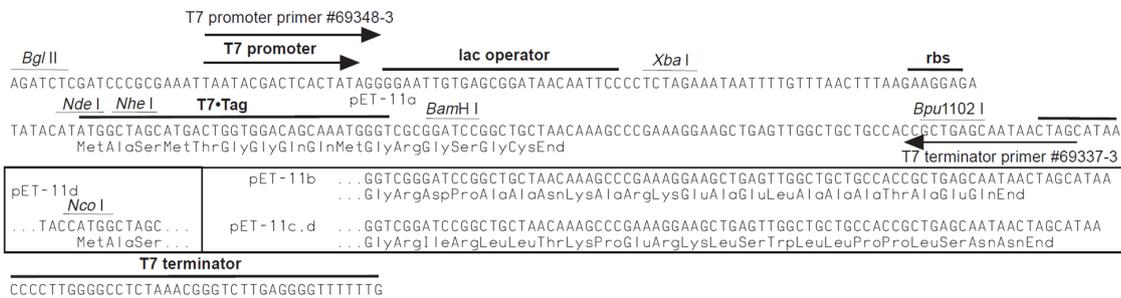


Figure B.4: pET11a cloning/expression region

Bibliography

- [1] R. Feynman, “There’s plenty of room at the bottom,” *Caltech Engineering and Science*, vol. 23, no. 5, pp. 22–36, 1960.
- [2] E. Engel, A. Michiardi, M. Navarro, D. Lacroix, and J. A. Planell, “Nanotechnology in regenerative medicine: the materials side,” *Trends in Biotechnology*, vol. 26, no. 1, pp. 39–47, 2008.
- [3] S. M. Moghimi, A. C. Hunter, and J. C. Murray, “Nanomedicine: current status and future prospects,” *The FASEB Journal*, vol. 19, no. 3, pp. 311–30, 2005.
- [4] M. J. Campolongo, J. S. Kahn, W. Cheng, D. Yang, T. Gupton-Campolongo, and D. Luo, “Adaptive DNA-based materials for switching, sensing, and logic devices,” *Journal of Materials Chemistry*, vol. 21, no. 17, p. 6113, 2011.
- [5] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov, “Electric field effect in atomically thin carbon films,” *Science*, vol. 306, no. 5696, pp. 666–9, 2004.
- [6] A. Dowling, R. Clift, N. Grobert, D. Hutton, R. Oliver, J. Pethica, N. Pidgeon, J. Porritt, J. Ryan, and Others, “Nanoscience and nanotechnologies: opportunities and uncertainties,” *London: The Royal Society & The Royal Academy of Engineering Report*, 2004.
- [7] A. K. Geim and K. S. Novoselov, “The rise of graphene,” *Nature Materials*, vol. 6, no. 3, pp. 183–91, 2007.

- [8] B. D. Reiss, C. Mao, D. J. Solis, K. S. Ryan, T. Thomson, and A. M. Belcher, "Biological routes to metal alloy ferromagnetic nanostructures," *Nano Letters*, vol. 4, no. 6, pp. 1127–1132, 2004.
- [9] J. Baumgartner and D. Faivre, "Magnetite biomineralization in bacteria," in *Progress in Molecular and Subcellular Biology* (W. Müller, P. Jeanteur, R. Rhoads, . Ugarković, and M. Reis Custódio, eds.), pp. 3–27, Springer, 2011.
- [10] N. C. Seeman, "Nanomaterials based on DNA," *Annual Review of Biochemistry*, vol. 79, pp. 65–87, 2010.
- [11] A. V. Pinheiro, D. Han, W. M. Shih, and H. Yan, "Challenges and opportunities for structural DNA nanotechnology," *Nature Nanotechnology*, vol. 6, no. 12, pp. 763–772, 2011.
- [12] P. B. Dervan, "Molecular recognition of DNA by small molecules," *Bioorganic & Medicinal Chemistry*, vol. 9, no. 9, pp. 2215–35, 2001.
- [13] K. Keren, M. Krueger, R. Gilad, G. Ben-Yoseph, U. Sivan, and E. Braun, "Sequence-specific molecular lithography on single DNA molecules," *Science*, vol. 297, no. 5578, pp. 72–5, 2002.
- [14] T. Nishinaka, A. Takano, Y. Doi, M. Hashimoto, A. Nakamura, Y. Matsushita, J. Kumaki, and E. Yashima, "Conductive metal nanowires templated by the nucleoprotein filaments, complex of DNA and RecA protein," *Journal of the American Chemical Society*, vol. 127, no. 22, pp. 8120–5, 2005.
- [15] C. Wälti, R. Sharma, and G. Davies, "RecA protein mediated nano-scale patterning of DNA scaffolds," *Journal of Nanoscience and Nanotechnology*, vol. 11, no. 12, pp. 10629–32, 2010.
- [16] R. Sharma, A. G. Davies, and C. Wälti, "Nanoscale programmable sequence-specific patterning of DNA scaffolds using RecA protein," *Nanotechnology*, vol. 23, no. 36, p. 365301, 2012.

- [17] R. Sharma, A. G. Davies, and C. Wälti, "RecA protein mediated nano-scale patterning of DNA scaffolds," *Journal of Nanoscience and Nanotechnology*, vol. 11, pp. 1–4, 2011.
- [18] R. Sharma, A. G. Davies, and C. Wälti, "Directed assembly of 3-nm-long RecA nucleoprotein filaments on double-stranded DNA with nanometer resolution," *ACS nano*, vol. 8, no. 4, pp. 3322–30, 2014.
- [19] J. D. Watson and F. H. C. Crick, "Molecular structure of nucleic acids," *Nature*, vol. 171, no. 4356, pp. 737–738, 1953.
- [20] R. B. Wallace, J. Shaffer, R. F. Murphy, J. Bonner, T. Hirose, and K. Itakura, "Hybridisation of synthetic oligodeoxyribonucleotides to $\phi\chi$ 174 DNA: the effect of single base pair mismatch," *Nucleic Acids Research*, vol. 6, no. 11, pp. 3543–3558, 1979.
- [21] P. M. Howley, M. A. Israel, M.-F. Law, and M. A. Martin, "A rapid method for detecting heterologous DNAs and mapping homology," *The Journal of Biological Chemistry*, vol. 254, no. 11, pp. 4876–4883, 1979.
- [22] A. Panjkovich and F. Melo, "Comparison of different melting temperature calculation methods for short DNA sequences," *Bioinformatics*, vol. 21, no. 6, pp. 711–722, 2005.
- [23] G. Khandelwal and J. Bhyravabhotla, "A phenomenological model for predicting melting temperatures of dna sequences," *PLoS ONE*, vol. 5, no. 8, pp. 1–9, 2010.
- [24] P. N. Borer, B. Dengler, I. Tinoco, and O. C. Uhlenbeck, "Stability of ribonucleic acid double-stranded helices," *Journal of Molecular Biology*, vol. 86, no. 4, pp. 843–853, 1974.
- [25] K. J. Breslauer, R. Frank, H. Blöcker, and L. A. Marky, "Predicting DNA duplex stability from the base sequence," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 83, no. 11, pp. 3746–50, 1986.

- [26] D. M. Gray, "Derivation of nearest-neighbor properties from data on nucleic acid oligomers. I. Simple sets of independent sequences and the influence of absent nearest neighbors," *Biopolymers*, vol. 42, no. 7, pp. 783–793, 1997.
- [27] P. Licinio and J. C. O. Guerra, "Irreducible representation for nucleotide sequence physical properties and self-consistency of nearest-neighbor dimer sets," *Biophysical journal*, vol. 92, no. 6, pp. 2000–6, 2007.
- [28] G. Weber, "Optimization method for obtaining nearest-neighbour DNA entropies and enthalpies directly from melting temperatures," *Bioinformatics*, vol. 31, no. 6, pp. 871–877, 2015.
- [29] T. A. Kunkel, "DNA replication fidelity," *Journal of Biological Chemistry*, vol. 279, no. 17, pp. 16895–8, 2004.
- [30] W. M. Barnes, "The fidelity of *Taq* polymerase catalyzing PCR is improved by an N-terminal deletion," *Gene*, vol. 112, no. 1, pp. 29–35, 1992.
- [31] N. R. Kallenbach, R.-I. Ma, and N. C. Seeman, "An immobile nucleic acid junction constructed from oligonucleotides," *Nature*, vol. 305, no. 27, pp. 829–831, 1983.
- [32] Y. Mitsui, R. Langrige, B. E. Shortle, C. R. Cantor, R. C. Grant, M. Kodama, and R. D. Wells, "Physical and enzymatic studies on poly d(IC).poly d(IC), an unusual double-helical DNA," *Nature*, vol. 228, pp. 1166–1169, 1970.
- [33] N. A. Campbell and J. B. Reece, "From gene to protein," in *Biology* (B. Wilbur, D. Gale, P. Burner, B. N. Winickoff, E. Gregg, R. Chun, and T. Amos, eds.), ch. 17, pp. 309–333, San Francisco: Benjamin Cummings, 7th ed., 2005.
- [34] N. A. Campbell and J. B. Reece, "A tour of the cell," in *Biology* (B. Wilbur, D. Gale, P. Burner, B. N. Winickoff, E. Gregg, R. Chun, and T. Amos, eds.), ch. 6, pp. 94–123, San Francisco: Benjamin Cummings, 7th ed., 2005.
- [35] A. Kölsch, R. Windoffer, T. Würflinger, T. Aach, and R. E. Leube, "The keratin-filament cycle of assembly and disassembly," *Journal of Cell Science*, vol. 123, pp. 2266–2272, 2010.

- [36] H. Gradišar and R. Jerala, “Self-assembled bionanostructures: proteins following the lead of DNA nanostructures,” *Journal of Nanobiotechnology*, vol. 12, no. 1, p. 4, 2014.
- [37] I. Severcan, C. Geary, A. Chworos, N. Voss, E. Jacovetty, and L. Jaeger, “A polyhedron made of tRNAs,” *Nature Chemistry*, vol. 2, no. 9, pp. 772–779, 2010.
- [38] W. W. Grabow, P. Zakrevsky, K. A. Afonin, A. Chworos, B. A. Shapiro, and L. Jaeger, “Self-assembling RNA nanorings based on RNAI/II inverse kissing complexes,” *Nano Letters*, vol. 11, no. 2, pp. 878–887, 2011.
- [39] C. Geary, P. W. K. Rothmund, and E. S. Andersen, “A single-stranded architecture for cotranscriptional folding of RNA nanostructures,” *Science*, vol. 345, no. 6198, pp. 799–804, 2014.
- [40] P. Guo, “The emerging field of RNA nanotechnology,” *Nature Nanotechnology*, vol. 5, no. 12, pp. 833–42, 2010.
- [41] N. C. Seeman, “Nucleic acid junctions and lattices,” *Journal of Theoretical Biology*, vol. 99, no. 2, pp. 237–247, 1982.
- [42] D. Han, S. Pal, J. Nangreave, Z. Deng, Y. Liu, and H. Yan, “DNA origami with complex curvatures in three-dimensional space,” *Science*, vol. 332, no. 6027, pp. 342–6, 2011.
- [43] E. S. Andersen, M. Dong, M. M. Nielsen, K. Jahn, R. Subramani, W. Mamdouh, M. M. Golas, B. Sander, H. Stark, C. L. P. Oliveira, J. S. Pedersen, V. Birkedal, F. Besenbacher, K. V. Gothelf, and J. Kjems, “Self-assembly of a nanoscale DNA box with a controllable lid,” *Nature*, vol. 459, no. 7243, pp. 73–6, 2009.
- [44] T. Tørring, N. V. Voigt, J. Nangreave, H. Yan, and K. V. Gothelf, “DNA origami: a quantum leap for self-assembly of complex structures,” *Chemical Society Reviews*, vol. 40, pp. 5636–5646, 2011.
- [45] A. Kuzyk, R. Schreiber, Z. Fan, G. Pardatscher, E. Roller, A. Hoge, F. C. Simmel, A. O. Govorov, and T. Liedl, “DNA-based self-assembly of chiral plasmonic

- nanostructures with tailored optical response,” *Nature*, vol. 483, pp. 311–314, 2012.
- [46] D. Bhatia, S. Sharma, and Y. Krishnan, “Synthetic, biofunctional nucleic acid-based molecular devices,” *Current Opinion in Biotechnology*, vol. 22, no. 4, pp. 475–84, 2011.
- [47] B. Saccà and C. M. Niemeyer, “Functionalization of DNA nanostructures with proteins,” *Chemical Society Reviews*, vol. 40, pp. 5910–5921, 2011.
- [48] Y. H. Roh, R. C. H. Ruiz, S. Peng, J. B. Lee, and D. Luo, “Engineering DNA-based functional materials,” *Chemical Society Reviews*, vol. 40, no. 12, pp. 5730–44, 2011.
- [49] N. C. Seeman and N. R. Kallenbach, “Design of immobile nucleic acid junctions,” *Biophysical Journal*, vol. 44, pp. 201–209, 1983.
- [50] P. W. K. Rothemund, “Folding DNA to create nanoscale shapes and patterns,” *Nature*, vol. 440, no. 7082, pp. 297–302, 2006.
- [51] R. Holliday, “A mechanism for gene conversion in fungi,” *Genetical Research*, vol. 5, no. 02, p. 282, 1964.
- [52] T. J. Fu and N. C. Seeman, “DNA double-crossover molecules,” *Biochemistry*, vol. 32, no. 13, pp. 3211–20, 1993.
- [53] H. Yan and N. C. Seeman, “Edge-sharing motifs in structural DNA nanotechnology,” *Journal of Supramolecular Chemistry*, vol. 1, no. 2001, pp. 229–237, 2001.
- [54] E. Winfree, F. Liu, L. A. Wenzler, and N. C. Seeman, “Design and self-assembly of two-dimensional DNA crystals,” *Nature*, vol. 394, no. 6693, pp. 539–44, 1998.
- [55] B. Ding, R. Sha, and N. C. Seeman, “Pseudo-hexagonal 2D DNA crystals from double crossover cohesion,” *Journal of the American Chemical Society*, vol. 126, no. 33, pp. 10230–1, 2004.

- [56] T. Wang, D. Schiffels, S. M. Cuesta, D. K. Fygenson, and N. C. Seeman, "Design and characterization of 1D nanotubes and 2D periodic arrays self-assembled from DNA multi-helix bundles," *Journal of the American Chemical Society*, vol. 134, no. 3, pp. 1606–16, 2012.
- [57] G. Binnig, H. Rohrer, C. Gerber, and E. Weibel, "Surface studies by scanning tunneling microscopy," *Physical Review Letters*, vol. 49, no. 1, pp. 57–61, 1982.
- [58] G. Binnig, F. Quate, and C. H. Gerber, "Atomic force microscope," *Physical Review Letters*, vol. 56, no. 9, pp. 930–933, 1986.
- [59] G. Binnig and H. Rohrer, "In touch with atoms," *Review of Modern Physics*, vol. 71, no. 2, pp. 324–330, 1999.
- [60] R. Chhabra, J. Sharma, Y. Ke, Y. Liu, S. Rinker, S. Lindsay, and H. Yan, "Spatially addressable multiprotein nanoarrays templated by aptamer-tagged DNA nanoarchitectures," *Journal of the American Chemical Society*, vol. 129, no. 34, pp. 10304–5, 2007.
- [61] Z. Zhao, H. Yan, and Y. Liu, "A route to scale up DNA origami using DNA tiles as folding staples," *Angewandte Chemie - International Edition*, vol. 49, no. 8, pp. 1414–7, 2010.
- [62] M. Endo, T. Sugita, Y. Katsuda, K. Hidaka, and H. Sugiyama, "Programmed-assembly system using DNA jigsaw pieces," *Chemistry - A European Journal*, vol. 16, no. 18, pp. 5362–8, 2010.
- [63] A. Rajendran, M. Endo, Y. Katsuda, K. Hidaka, and H. Sugiyama, "Photo-cross-linking-assisted thermal stability of DNA origami structures and its application for higher-temperature self-assembly," *Journal of the American Chemical Society*, vol. 133, no. 37, pp. 14488–91, 2011.
- [64] A. Rajendran, M. Endo, Y. Katsuda, K. Hidaka, and H. Sugiyama, "Programmed two-dimensional self-assembly of multiple DNA origami jigsaw pieces," *ACS nano*, vol. 5, no. 1, pp. 665–671, 2011.

- [65] M. Endo, Y. Katsuda, K. Hidaka, and H. Sugiyama, "Regulation of DNA methylation using different tensions of double strands constructed in a defined DNA nanostructure," *Journal of the American Chemical Society*, vol. 132, no. 5, pp. 1592–7, 2010.
- [66] M. Endo, Y. Katsuda, K. Hidaka, and H. Sugiyama, "A versatile DNA nanochip for direct analysis of DNA base-excision repair," *Angewandte Chemie - International Edition*, vol. 49, no. 49, pp. 9412–6, 2010.
- [67] Y. Suzuki, M. Endo, Y. Katsuda, K. Ou, K. Hidaka, and H. Sugiyama, "DNA origami based visualization system for studying site-specific recombination events," *Journal of the American Chemical Society*, vol. 136, pp. 211–218, 2013.
- [68] M. Endo, K. Tatsumi, K. Terushima, Y. Katsuda, K. Hidaka, Y. Harada, and H. Sugiyama, "Direct visualization of the movement of a single T7 RNA polymerase and transcription on a DNA nanostructure," *Angewandte Chemie - International Edition*, vol. 51, no. 35, pp. 8778–82, 2012.
- [69] M. Endo, Y. Yang, Y. Suzuki, K. Hidaka, and H. Sugiyama, "Single-molecule visualization of the hybridization and dissociation of photoresponsive oligonucleotides and their reversible switching behavior in a DNA nanostructure," *Angewandte Chemie - International Edition*, vol. 51, no. 42, pp. 10518–22, 2012.
- [70] T. Yoshidome, M. Endo, G. Kashiwazaki, K. Hidaka, T. Bando, and H. Sugiyama, "Sequence-selective single-molecule alkylation with a pyrrole-imidazole polyamide visualized in a DNA nanoscaffold," *Journal of the American Chemical Society*, vol. 134, no. 10, pp. 4654–60, 2012.
- [71] M. Endo and H. Sugiyama, "Single-molecule imaging of dynamic motions of biomolecules in DNA origami nanostructures using high-speed atomic force microscopy," *Accounts of Chemical Research*, vol. 47, pp. 1645–1653, 2014.
- [72] A. Rajendran, M. Endo, K. Hidaka, and H. Sugiyama, "Direct and single-molecule visualization of the solution-state structures of G-hairpin and G-triplex intermediates," *Angewandte Chemie - International Edition*, vol. 53, no. 16, pp. 4107–12, 2014.

- [73] A. Rajendran, M. Endo, K. Hidaka, P. L. T. Tran, J.-L. Mergny, and H. Sugiyama, "Controlling the stoichiometry and strand polarity of a tetramolecular G-quadruplex structure by using a DNA origami frame," *Nucleic Acids Research*, vol. 41, no. 18, pp. 8738–47, 2013.
- [74] A. Rajendran, M. Endo, K. Hidaka, P. L. Thao Tran, J.-L. Mergny, R. J. Gorelick, and H. Sugiyama, "HIV-1 nucleocapsid proteins as molecular chaperones for tetramolecular antiparallel G-quadruplex formation," *Journal of the American Chemical Society*, vol. 135, p. 1857518585, 2013.
- [75] Y. Sannohe, M. Endo, Y. Katsuda, K. Hidaka, and H. Sugiyama, "Visualization of dynamic conformational switching of the G-quadruplex in a DNA nanostructure," *Journal of the American Chemical Society*, vol. 132, pp. 16311–16313, 2010.
- [76] Q. Wang, J. Q. Liu, Z. Chen, K. W. Zheng, C. Y. Chen, Y. H. Hao, and Z. Tan, "G-quadruplex formation at the 3' end of telomere DNA inhibits its extension by telomerase, polymerase and unwinding by helicase," *Nucleic Acids Research*, vol. 39, no. 14, pp. 6229–6237, 2011.
- [77] N. W. Kim, M. A. Piatyszek, K. R. Prowse, C. B. Harley, M. D. West, P. L. C. Ho, G. M. Coviello, W. E. Wright, S. L. Weinrich, and J. W. Shay, "Specific association of human telomerase activity with immortal cells and cancer," *Science*, vol. 266, pp. 2011–2015, 1994.
- [78] G. B. Morin, "Telomere control of replicative lifespan," *Experimental Gerontology*, vol. 32, no. 4-5, pp. 375–382, 1997.
- [79] W. Li, Y. Yang, S. Jiang, H. Yan, and Y. Liu, "Controlled nucleation and growth of DNA tile arrays within prescribed DNA origami frames and their dynamics," *Journal of the American Chemical Society*, vol. 136, no. 10, pp. 3724–7, 2014.
- [80] B. Wei, M. Dai, and P. Yin, "Complex shapes self-assembled from single-stranded DNA tiles," *Nature*, vol. 485, no. 7400, pp. 623–6, 2012.
- [81] W. M. Shih, J. D. Quispe, and G. F. Joyce, "A 1.7-kilobase single-stranded DNA that folds into a nanoscale octahedron," *Nature*, vol. 427, pp. 260–261, 2004.

- [82] R. P. Goodman, I. A. T. Schaap, C. F. Tardin, C. M. Erben, R. M. Berry, C. F. Schmidt, and A. J. Turberfield, "Rapid chiral assembly of rigid DNA building blocks for molecular nanofabrication," *Science*, vol. 310, no. 5754, pp. 1661–5, 2005.
- [83] Y. He, T. Ye, M. Su, C. Zhang, A. Ribbe, W. Jiang, and C. Mao, "Hierarchical self-assembly of DNA into symmetric supramolecular polyhedra," *Nature*, vol. 452, pp. 198–202, 2008.
- [84] R. Iinuma, Y. Ke, R. Jungmann, T. Schlichthaerle, J. B. Woehrstein, and P. Yin, "Polyhedra self-assembled from DNA tripods and characterized with 3D DNA-PAINT," *Science*, vol. 344, no. 6179, pp. 65–9, 2014.
- [85] S. M. Douglas, I. Bachelet, and G. M. Church, "A logic-gated nanorobot for targeted transport of molecular payloads," *Science*, vol. 335, no. 6070, pp. 831–4, 2012.
- [86] J. Zheng, J. J. Birktoft, Y. Chen, T. Wang, R. Sha, P. E. Constantinou, S. L. Ginell, C. Mao, and N. C. Seeman, "From molecular to macroscopic via the rational design of a self-assembled 3D DNA crystal," *Nature*, vol. 461, no. 7260, pp. 74–7, 2009.
- [87] R. Motro and V. Raducanu, "Tensegrity systems," *International Journal of Space Structures*, vol. 18, no. 2, pp. 77–84, 2003.
- [88] D. Liu, M. Wang, Z. Deng, R. Walulu, and C. Mao, "Tensegrity: construction of rigid DNA triangles with flexible four-arm DNA junctions," *Journal of the American Chemical Society*, vol. 126, no. 8, pp. 2324–5, 2004.
- [89] T. Liedl, B. Högberg, J. Tytell, D. E. Ingber, and W. M. Shih, "Self-assembly of three-dimensional prestressed tensegrity structures from DNA," *Nature Nanotechnology*, vol. 5, no. 7, pp. 520–4, 2010.
- [90] Y. Ke, L. L. Ong, W. M. Shih, and P. Yin, "Three-dimensional structures self-assembled from DNA bricks," *Science*, vol. 338, no. 6111, pp. 1177–1183, 2012.
- [91] H. Dietz, S. M. Douglas, and W. M. Shih, "Folding DNA into twisted and curved nanoscale shapes," *Science*, vol. 325, no. 5941, pp. 725–30, 2009.

- [92] B. Yurke, A. J. Turberfield, and A. P. Mills, "A DNA-fuelled molecular machine made of DNA," *Nature*, vol. 406, pp. 605–608, 2000.
- [93] Y. Yang, M. Endo, K. Hidaka, and H. Sugiyama, "Photo-controllable DNA origami nanostructures assembling into predesigned multiorientational patterns," *Journal of the American Chemical Society*, vol. 134, no. 51, pp. 20645–53, 2012.
- [94] A. E. Marras, L. Zhou, H.-J. Su, and C. E. Castro, "Programmable motion of DNA origami mechanisms," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 2014, 2015.
- [95] P. Yin, H. Yan, X. G. Daniell, A. J. Turberfield, and J. H. Reif, "A unidirectional DNA walker that moves autonomously along a track," *Angewandte Chemie - International Edition*, vol. 43, no. 37, pp. 4906–11, 2004.
- [96] J.-S. Shin and N. A. Pierce, "A synthetic DNA walker for molecular transport," *Journal of the American Chemical Society*, vol. 126, no. 35, pp. 10834–5, 2004.
- [97] S. J. Green, D. Lubrich, and A. J. Turberfield, "DNA hairpins: Fuel for autonomous DNA devices," *Biophysical Journal*, vol. 91, no. 8, pp. 2966–2975, 2006.
- [98] S. Green, J. Bath, and A. Turberfield, "Coordinated chemomechanical cycles: A mechanism for autonomous molecular motion," *Physical Review Letters*, vol. 101, no. 23, pp. 20–23, 2008.
- [99] H. Gu, J. Chao, S.-J. Xiao, and N. C. Seeman, "A proximity-based programmable DNA nanoscale assembly line," *Nature*, vol. 465, no. 7295, pp. 202–205, 2010.
- [100] S. F. J. Wickham, M. Endo, Y. Katsuda, K. Hidaka, J. Bath, H. Sugiyama, and A. J. Turberfield, "Direct observation of stepwise movement of a synthetic molecular transporter," *Nature Nanotechnology*, vol. 6, no. 3, pp. 166–9, 2011.
- [101] S. F. J. Wickham, J. Bath, Y. Katsuda, M. Endo, K. Hidaka, H. Sugiyama, and A. J. Turberfield, "A DNA-based molecular motor that can navigate a network of tracks," *Nature Nanotechnology*, vol. 7, no. 3, pp. 169–73, 2012.

- [102] K. Lund, A. J. Manzo, N. Dabby, N. Michelotti, A. Johnson-Buck, J. Nangreave, S. Taylor, R. Pei, M. N. Stojanovic, N. G. Walter, E. Winfree, and H. Yan, "Molecular robots guided by prescriptive landscapes," *Nature*, vol. 465, no. 7295, pp. 206–10, 2010.
- [103] Y. He and D. R. Liu, "Autonomous multistep organic synthesis in a single isothermal solution mediated by a DNA walker," *Nature Nanotechnology*, vol. 5, no. 11, pp. 778–82, 2010.
- [104] M. L. McKee, P. J. Milnes, J. Bath, E. Stulz, A. J. Turberfield, and R. K. O'Reilly, "Multistep DNA-templated reactions for the synthesis of functional sequence controlled oligomers," *Angewandte Chemie - International Edition*, vol. 49, no. 43, pp. 7948–51, 2010.
- [105] M. L. McKee, P. J. Milnes, J. Bath, E. Stulz, R. K. O'Reilly, and A. J. Turberfield, "Programmable one-pot multistep organic synthesis using DNA junctions," *Journal of the American Chemical Society*, vol. 134, pp. 1446–1449, 2012.
- [106] Y. Ke, S. M. Douglas, M. Liu, J. Sharma, A. Cheng, A. Leung, Y. Liu, W. M. Shih, and H. Yan, "Multilayer DNA origami packed on a square lattice," *Journal of the American Chemical Society*, vol. 131, no. 43, pp. 15903–8, 2009.
- [107] R. Jungmann, C. Steinhauer, M. Scheible, A. Kuzyk, P. Tinnefeld, and F. C. Simmel, "Single-molecule kinetics and super-resolution microscopy by fluorescence imaging of transient binding on DNA origami," *Nano Letters*, vol. 10, no. 11, pp. 4756–4761, 2010.
- [108] J. Le, Y. Pinto, N. Seeman, K. Musier-Forsyth, T. A. Taton, and R. A. Kiehl, "DNA-templated self-assembly of metallic nanocomponent arrays on a surface," *Nano Letters*, vol. 4, no. 12, pp. 2343–2347, 2004.
- [109] H. Li, T. H. LaBean, and D. J. Kenan, "Single-chain antibodies against DNA aptamers for use as adapter molecules on DNA tile arrays in nanoscale materials organization," *Organic & Biomolecular Chemistry*, vol. 4, no. 18, pp. 3420–6, 2006.

- [110] B. Kick, F. Praetorius, H. Dietz, and D. Weuster-Botz, "Efficient production of single-stranded phage DNA as scaffolds for DNA origami," *Nano Letters*, vol. 15, no. 7, pp. 4672–4676, 2015.
- [111] S. H. Park, C. Pistol, S. J. Ahn, J. H. Reif, A. R. Lebeck, C. Dwyer, and T. H. LaBean, "Finite-size, fully addressable DNA tile lattices formed by hierarchical assembly procedures," *Angewandte Chemie*, vol. 118, no. 5, pp. 749–753, 2006.
- [112] C. Lin, M. Xie, J. J. L. Chen, Y. Liu, and H. Yan, "Rolling-circle amplification of a DNA nanojunction," *Angewandte Chemie - International Edition*, vol. 45, no. 45, pp. 7537–9, 2006.
- [113] C. Lin, X. Wang, Y. Liu, N. C. Seeman, and H. Yan, "Rolling circle enzymatic replication of a complex multi-crossover DNA nanostructure," *Journal of the American Chemical Society*, vol. 129, pp. 14475–14481, 2007.
- [114] C. Lin, S. Rinker, X. Wang, Y. Liu, N. C. Seeman, and H. Yan, "In vivo cloning of artificial DNA nanostructures," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 46, pp. 17626–31, 2008.
- [115] E. T. Kool, "Replacing the nucleobases in DNA with designer molecules," *Accounts of Chemical Research*, vol. 35, no. 11, pp. 936–943, 2002.
- [116] D. Loakes, J. Gallego, V. B. Pinheiro, E. T. Kool, and P. Holliger, "Evolving a polymerase for hydrophobic base analogues," *Journal of the American Chemical Society*, vol. 131, no. 41, pp. 14827–37, 2009.
- [117] V. B. Pinheiro, A. I. Taylor, C. Cozens, M. Abramov, M. Renders, S. Zhang, J. C. Chaput, J. Wengel, S.-Y. Peak-Chew, S. H. McLaughlin, P. Herdewijn, and P. Holliger, "Synthetic genetic polymers capable of heredity and evolution," *Science*, vol. 336, pp. 341–345, 2012.
- [118] A. I. Taylor, V. B. Pinheiro, M. J. Smola, A. S. Morgunov, S. Peak-Chew, C. Cozens, K. M. Weeks, P. Herdewijn, and P. Holliger, "Catalysts from synthetic genetic polymers," *Nature*, vol. 518, pp. 427–430, 2014.

- [119] V. B. Pinheiro and P. Holliger, "Towards XNA nanotechnology: New materials from synthetic genetic polymers," *Trends in Biotechnology*, vol. 32, no. 6, pp. 321–328, 2014.
- [120] D. A. Malyshev, K. Dhimi, H. T. Quach, T. Lavergne, P. Ordoukhanian, A. Torkamani, and F. E. Romesberg, "Efficient and sequence-independent replication of DNA containing a third base pair establishes a functional six-letter genetic alphabet," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, pp. 12005–12010, 2012.
- [121] J. R. Burns, J. Zekonyte, G. Siligardi, R. Hussain, and E. Stulz, "Directed formation of DNA nanoarrays through orthogonal self-assembly," *Molecules*, vol. 16, pp. 4912–4922, 2011.
- [122] M. Hocek, M. Vrábel, H. Cahová, J. Riedl, L. Kalachová, H. Pivonková, P. Horáková, L. Havran, and M. Fojta, "Novel base-functionalized DNA. Efficient methodology for construction and bioanalytical applications," *Nucleic Acids Symposium Series*, vol. 52, no. 1, pp. 53–4, 2008.
- [123] J. Balintová, R. Pohl, P. Horáková, P. Vidláková, L. Havran, M. Fojta, and M. Hocek, "Anthraquinone as a redox label for DNA: synthesis, enzymatic incorporation, and electrochemistry of anthraquinone-modified nucleosides, nucleotides, and DNA," *Chemistry - A European Journal*, vol. 17, no. 50, pp. 14063–73, 2011.
- [124] H. Macíčková-Cahová, R. Pohl, P. Horáková, L. Havran, J. Špaček, M. Fojta, and M. Hocek, "Alkylsulfanylphenyl derivatives of cytosine and 7-deazaadenine nucleosides, nucleotides and nucleoside triphosphates: synthesis, polymerase incorporation to DNA and electrochemical study," *Chemistry - A European Journal*, vol. 17, no. 21, pp. 5833–41, 2011.
- [125] V. Raindlová, R. Pohl, and M. Hocek, "Synthesis of aldehyde-linked nucleotides and DNA and their bioconjugations with lysine and peptides through reductive amination," *Chemistry - A European Journal*, vol. 18, no. 13, pp. 4080–7, 2012.

- [126] P. Blakskjaer and K. V. Gothelf, "Synthesis of an elongated linear oligo(phenylene ethynylene)-based building block for application in DNA-programmed assembly," *Organic & Biomolecular Chemistry*, vol. 4, no. 18, pp. 3442–7, 2006.
- [127] T. Rühl and E. Stulz, "Synthesis of new building blocks for use in supramolecular DNA architectures," *Supramolecular Chemistry*, vol. 22, no. 2, pp. 103–108, 2010.
- [128] D. N. Selmi, R. J. Adamson, H. Attrill, A. D. Goddard, R. J. C. Gilbert, A. Watts, and A. J. Turberfield, "DNA-templated protein arrays for single-molecule imaging," *Nano Letters*, vol. 11, no. 2, pp. 657–60, 2011.
- [129] J. Zheng, P. E. Constantinou, C. Micheel, A. P. Alivisatos, R. A. Kiehl, and N. C. Seeman, "Two-dimensional nanoparticle arrays show the organizational power of robust DNA motifs," *Nano Letters*, vol. 6, no. 7, pp. 1502–4, 2006.
- [130] J. Sharma, R. Chhabra, Y. Liu, Y. Ke, and H. Yan, "DNA-templated self-assembly of two-dimensional and periodical gold nanoparticle arrays," *Angewandte Chemie*, vol. 118, no. 5, pp. 744–749, 2006.
- [131] A. M. Hung, H. Noh, and J. N. Cha, "Recent advances in DNA-based directed assembly on surfaces," *Nanoscale*, vol. 2, no. 12, pp. 2530–7, 2010.
- [132] H. Liu, Y. He, A. E. Ribbe, and C. Mao, "Two-dimensional (2D) DNA crystals assembled from two DNA strands," *Biomacromolecules*, vol. 6, no. 6, pp. 2943–5, 2005.
- [133] Y. Liu, C. Lin, H. Li, and H. Yan, "Aptamer-directed self-assembly of protein arrays on a DNA nanostructure," *Angewandte Chemie - International Edition*, vol. 44, no. 28, pp. 4333–8, 2005.
- [134] J. M. Gottesfeld, C. Melander, R. K. Suto, H. Raviol, K. Luger, and P. B. Dervan, "Sequence-specific recognition of DNA in the nucleosome by pyrrole-imidazole polyamides," *Journal of Molecular Biology*, vol. 309, no. 3, pp. 615–29, 2001.
- [135] P. B. Dervan and B. S. Edelson, "Recognition of the DNA minor groove by pyrrole-imidazole polyamides," *Current Opinion in Structural Biology*, vol. 13, no. 3, pp. 284–299, 2003.

- [136] J. D. Cohen, J. P. Sadowski, and P. B. Dervan, "Addressing single molecules on DNA nanostructures," *Angewandte Chemie - International Edition*, vol. 46, pp. 7956–9, jan 2007.
- [137] J. D. Cohen, J. P. Sadowski, and P. B. Dervan, "Programming multiple protein patterns on a single DNA nanostructure," *Journal of the American Chemical Society*, vol. 130, no. 2, pp. 402–403, 2008.
- [138] Z. Krpetic, I. Singh, W. Su, L. Guerrini, K. Faulds, G. A. Burley, and D. Graham, "Directed assembly of DNA-functionalized gold nanoparticles using pyrroleimidazole polyamides," *Journal of the American Chemical Society*, vol. 134, no. 20, pp. 8356–8359, 2012.
- [139] T. L. Schmidt, C. K. Nandi, G. Rasched, P. P. Parui, B. Brutschy, M. Famulok, and A. Heckel, "Polyamide struts for DNA architectures," *Angewandte Chemie - International Edition*, vol. 46, no. 23, pp. 4382–4384, 2007.
- [140] K. Lang and J. W. Chin, "Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins," *Chemical Reviews*, vol. 114, no. 9, pp. 4764–806, 2014.
- [141] J. K. Joung and J. D. Sander, "TALENs: a widely applicable technology for targeted genome editing," *Nature Reviews. Molecular Cell Biology*, vol. 14, no. 1, pp. 49–55, 2013.
- [142] E. Nakata, F. F. Liew, C. Uwatoko, S. Kiyonaka, Y. Mori, Y. Katsuda, M. Endo, H. Sugiyama, and T. Morii, "Zinc-finger proteins for site-specific protein positioning on DNA-origami structures," *Angewandte Chemie - International Edition*, vol. 51, no. 10, pp. 2421–2424, 2012.
- [143] P. Mali, K. M. Esvelt, and G. M. Church, "Cas9 as a versatile tool for engineering biology," *Nature Methods*, vol. 10, no. 10, pp. 957–63, 2013.
- [144] B. Shen, W. Zhang, J. Zhang, J. Zhou, J. Wang, L. Chen, L. Wang, A. Hodgkins, V. Iyer, X. Huang, and W. C. Skarnes, "Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects," *Nature Methods*, vol. 11, no. 4, pp. 399–402, 2014.

- [145] L. S. Qi, M. H. Larson, L. A. Gilbert, J. A. Doudna, J. S. Weissman, A. P. Arkin, and W. A. Lim, “Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression,” *Cell*, vol. 152, no. 5, pp. 1173–83, 2013.
- [146] M. M. Cox, “Motoring along with the bacterial RecA protein,” *Nature Reviews. Molecular Cell Biology*, vol. 8, no. 2, pp. 127–38, 2007.
- [147] S. E. Burckhardt, R. Woodgate, R. H. Scheuermann, and H. Echols, “UmuD mutagenesis protein of *Escherichia coli*: overproduction, purification, and cleavage by RecA,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 85, no. 6, pp. 1811–1815, 1988.
- [148] M. Patel, Q. Jiang, R. Woodgate, M. M. Cox, and M. F. Goodman, “A new model for SOS-induced mutagenesis: how RecA protein activates DNA polymerase V,” *Critical Reviews in Biochemistry and Molecular Biology*, vol. 45, no. 3, pp. 171–184, 2010.
- [149] K. Schlacher, K. Leslie, C. Wyman, R. Woodgate, M. M. Cox, and M. F. Goodman, “DNA polymerase V and RecA protein, a minimal mutasome,” *Molecular Cell*, vol. 17, no. 4, pp. 561–572, 2005.
- [150] C. Indiani and M. Patel, “RecA acts as a switch to regulate polymerase occupancy in a moving replication fork,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 14, pp. 5410–5415, 2013.
- [151] C. Joo, S. A. McKinney, M. Nakamura, I. Rasnik, S. Myong, and T. Ha, “Real-time observation of RecA filament dynamics with single monomer resolution,” *Cell*, vol. 126, no. 3, pp. 515–27, 2006.
- [152] M. T. J. van Loenhout, T. van der Heijden, R. Kanaar, C. Wyman, and C. Dekker, “Dynamics of RecA filaments on single-stranded DNA,” *Nucleic Acids Research*, vol. 37, no. 12, pp. 4089–99, 2009.
- [153] A. L. Forget and S. C. Kowalczykowski, “Single-molecule imaging of DNA pairing by RecA reveals a three-dimensional homology search,” *Nature*, vol. 482, no. 7385, pp. 423–7, 2012.

- [154] Z. Chen, H. Yang, and N. P. Pavletich, "Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures," *Nature*, vol. 453, no. 7194, pp. 489–4, 2008.
- [155] C. E. Bell, "Structure and mechanism of *Escherichia coli* RecA ATPase," *Molecular Microbiology*, vol. 58, no. 2, pp. 358–66, 2005.
- [156] R. L. Britt, N. Haruta, S. L. Lusetti, S. Chitteni-Pattu, R. B. Inman, and M. M. Cox, "Disassembly of *Escherichia coli* RecA E38K/DeltaC17 nucleoprotein filaments is required to complete DNA strand exchange," *The Journal of Biological Chemistry*, vol. 285, pp. 3211–26, jan 2010.
- [157] A. Reymer, S. Babik, M. Takahashi, B. Norden, and T. Beke-Somfai, "ATP hydrolysis in the RecA-DNA filament promotes structural changes at the protein-DNA interface," *Biochemistry*, vol. 54, no. 30, pp. 4579–4582, 2015.
- [158] E. H. Egelman and A. Stasiak, "Structure of helical RecA-DNA complexes. Complexes formed in the presence of ATP- γ -S or ATP," *Journal of Molecular Biology*, vol. 191, no. 4, pp. 677–97, 1986.
- [159] B. S. Li and M. C. Goh, "Direct visualization of the formation and structure of RecA/dsDNA complexes," *Micron*, vol. 41, no. 3, pp. 227–31, 2010.
- [160] R. M. Story, I. T. Weber, and T. A. Steitz, "The structure of the *E. coli* RecA protein monomer and polymer," *Nature*, vol. 355, p. 318, 1992.
- [161] X. Xing and C. E. Bell, "Crystal structures of *Escherichia coli* RecA in a compressed helical filament," *Journal of Molecular Biology*, vol. 342, no. 4, pp. 1471–85, 2004.
- [162] R. C. Benedict and S. C. Kowalczykowski, "Increase of the DNA strand assimilation activity of RecA protein by removal of the C terminus and structure-function studies of the resulting protein fragment," *Journal of Biological Chemistry*, vol. 263, no. 30, pp. 15513–20, 1988.
- [163] S. L. Lusetti, E. A. Wood, C. D. Fleming, M. J. Modica, J. Korth, L. Abbott, D. W. Dwyer, A. I. Roca, R. B. Inman, and M. M. Cox, "C-terminal deletions of the

- Escherichia coli* RecA protein. Characterization of *in vivo* and *in vitro* effects,” *The Journal of Biological Chemistry*, vol. 278, no. 18, pp. 16372–80, 2003.
- [164] A. L. Eggler, S. L. Lusetti, and M. M. Cox, “The C terminus of the *Escherichia coli* RecA protein modulates the DNA binding competition with single-stranded DNA-binding protein,” *Journal of Biological Chemistry*, vol. 278, no. 18, pp. 16389–96, 2003.
- [165] S. Tateishi, T. Horii, T. Ogawa, and H. Ogawa, “C-terminal truncated *Escherichia coli* RecA protein RecA5327 has enhanced binding affinities to single- and double-stranded DNAs,” *Journal of Molecular Biology*, vol. 223, no. 1, pp. 115–29, 1992.
- [166] V. A. Malkov and R. D. Camerini-Otero, “Photocross-links between single-stranded DNA and *Escherichia coli* RecA protein map to loops L1 (amino acid residues 157-164) and L2 (amino acid residues 195-209),” *The Journal of Biological Chemistry*, vol. 270, no. 50, pp. 30230–3, 1995.
- [167] Y. Wang and K. Adzuma, “Differential proximity probing of two DNA binding sites in the *Escherichia coli* RecA protein using photo-cross-linking methods,” *Biochemistry*, vol. 35, no. 11, pp. 3563–71, 1996.
- [168] P. R. Bianco and G. M. Weinstock, “Interaction of the RecA protein of *Escherichia coli* with single-stranded oligodeoxyribonucleotides,” *Nucleic Acids Research*, vol. 24, pp. 4933–9, 1996.
- [169] A. V. Mazin and S. C. Kowalczykowski, “The function of the secondary DNA-binding site of RecA protein during DNA strand exchange,” *EMBO Journal*, vol. 17, no. 4, pp. 1161–1168, 1998.
- [170] T. Shinohara, S. Ikawa, W. Iwasaki, T. Hiraki, T. Hikima, T. Mikawa, N. Arai, N. Kamiya, and T. Shibata, “Loop L1 governs the DNA-binding specificity and order for RecA-catalyzed reactions in homologous recombination and DNA repair,” *Nucleic Acids Research*, vol. 43, no. 2, pp. 973–986, 2015.
- [171] K. Frykholm, M. Alizadehheidari, J. Fritzsche, J. Wiggenius, M. Modesti, F. Persson, and F. Westerlund, “Probing physical properties of a DNA-protein complex using nanofluidic channels,” *Small*, vol. 10, no. 5, pp. 884–887, 2014.

- [172] D. M. Baitin, I. V. Bakhlanova, D. V. Chervyakova, Y. V. Kil, V. A. Lanzov, and M. M. Cox, “Two RecA protein types that mediate different modes of hyperrecombination,” *Journal of Bacteriology*, vol. 190, no. 8, pp. 3036–45, 2008.
- [173] I. V. Bakhlanova, A. V. Dudkina, D. M. Baitin, K. L. Knight, M. M. Cox, and V. A. Lanzov, “Modulating cellular recombination potential through alterations in RecA structure and regulation,” *Molecular Microbiology*, vol. 78, no. 6, pp. 1523–1538, 2010.
- [174] J. R. Piechura, T.-L. Tseng, H.-F. Hsu, R. T. Byrne, T. A. Windgassen, S. Chitteni-Pattu, J. R. Battista, H.-W. Li, and M. M. Cox, “Biochemical characterization of RecA variants that contribute to extreme resistance to ionizing radiation,” *DNA Repair*, vol. 26, pp. 30–43, 2015.
- [175] T. Kim, S. Chitteni-Pattu, B. L. Cox, E. A. Wood, S. J. Sandler, and M. M. Cox, “Directed evolution of RecA variants with enhanced capacity for conjugational recombination,” *PLOS Genetics*, vol. 11, no. 6, p. e1005278, 2015.
- [176] R. L. Britt, S. Chitteni-Pattu, A. N. Page, and M. M. Cox, “RecA K72R filament formation defects reveal an oligomeric RecA species involved in filament extension,” *The Journal of Biological Chemistry*, vol. 286, no. 10, pp. 7830–40, 2011.
- [177] M. C. Gruenig, N. Renzette, E. Long, S. Chitteni-Pattu, R. B. Inman, M. M. Cox, and S. J. Sandler, “RecA-mediated SOS induction requires an extended filament conformation but no ATP hydrolysis,” *Molecular Microbiology*, vol. 69, no. 5, pp. 1165–1179, 2008.
- [178] J. Renkawitz, C. A. Lademann, and S. Jentsch, “Mechanisms and principles of homology search during recombination,” *Nature reviews. Molecular Cell Biology*, vol. 15, no. 6, pp. 369–83, 2014.
- [179] R. Rajan, J. W. Wisler, and C. E. Bell, “Probing the DNA sequence specificity of *Escherichia coli* RecA protein,” *Nucleic Acids Research*, vol. 34, no. 8, pp. 2463–71, 2006.

- [180] R. Galletto, I. Amitani, R. J. Baskin, and S. C. Kowalczykowski, "Direct observation of individual RecA filaments assembling on single DNA molecules," *Nature*, vol. 443, no. 7113, pp. 875–8, 2006.
- [181] M. C. Leahy and C. M. Radding, "Topography of the interaction of RecA protein with single-stranded deoxyoligonucleotides," *Journal of Biological Chemistry*, vol. 261, no. 15, pp. 6954–60, 1986.
- [182] T. van der Heijden, M. Modesti, S. Hage, R. Kanaar, C. Wyman, and C. Dekker, "Homologous recombination in real time: DNA strand exchange by RecA," *Molecular Cell*, vol. 30, no. 4, pp. 530–8, 2008.
- [183] J. C. Bell, J. L. Plank, C. C. Dombrowski, and S. C. Kowalczykowski, "Direct imaging of RecA nucleation and growth on single molecules of SSB-coated ssDNA," *Nature*, vol. 491, pp. 274–278, 2012.
- [184] S. Nayak and F. R. Bryant, "Kinetics of the ATP and dATP-mediated formation of a functionally-active RecA-ssDNA complex," *Biochemical and Biophysical Research Communications*, vol. 463, no. 4, pp. 1257–1261, 2015.
- [185] M. A. Metrick, J. E. Temple, and G. Macdonald, "The effects of buffers and pH on the thermal stability, unfolding and substrate binding of RecA," *Biophysical Chemistry*, vol. 184, pp. 29–36, 2013.
- [186] H. Fu, S. Le, K. Muniyappa, and J. Yan, "Dynamics and regulation of RecA polymerization and de-polymerization on double-stranded DNA," *PLoS ONE*, vol. 8, no. 6, p. e66712, 2013.
- [187] J. C. Register and J. Griffith, "The direction of RecA protein assembly onto single strand DNA is the same as the direction of strand assimilation during strand exchange," *Journal of Biological Chemistry*, vol. 260, no. 22, pp. 12308–12312, 1985.
- [188] B. F. Pugh and M. M. Cox, "General mechanism for RecA protein binding to duplex DNA," *Journal of molecular biology*, vol. 203, pp. 479–493, 1988.

- [189] J. D. Warfel and V. J. LiCata, “Enhanced DNA binding affinity of RecA protein from *Deinococcus radiodurans*,” *DNA Repair*, vol. 31, pp. 91–96, 2015.
- [190] B. F. Paulus and F. R. Bryant, “Time-dependent inhibition of RecA protein-catalyzed ATP hydrolysis by ATP γ S: Evidence for a rate-determining isomerization of the RecA-ssDNA complex,” *Biochemistry*, vol. 36, no. 25, pp. 7832–7838, 1997.
- [191] M. S. Vanloock, X. Yu, S. Yang, A. L. Lai, C. Low, M. J. Campbell, and E. H. Egelman, “ATP-mediated conformational changes in the RecA filament,” *Structure*, vol. 11, no. 03, pp. 187–196, 2003.
- [192] W. Rosselli and A. Stasiak, “Energetics of RecA-mediated recombination reactions. Without ATP hydrolysis RecA can mediate polar strand exchange but is unable to recycle,” *Journal of Molecular Biology*, vol. 216, no. 2, pp. 335–52, 1990.
- [193] S. C. Kowalczykowski and R. A. Krupp, “DNA-strand exchange promoted by RecA protein in the absence of ATP: implications for the mechanism of energy transduction in protein-promoted nucleic acid transactions,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 92, no. 8, pp. 3478–82, 1995.
- [194] J. P. Menetski, D. G. Bear, and S. C. Kowalczykowski, “Stable DNA heteroduplex formation catalyzed by the *Escherichia coli* RecA protein in the absence of ATP hydrolysis,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 87, no. 1, pp. 21–25, 1990.
- [195] K. P. Rice, A. L. Eggler, P. Sung, and M. M. Cox, “DNA pairing and strand exchange by the *Escherichia coli* RecA and yeast Rad51 proteins without ATP hydrolysis. On the importance of not getting stuck,” *Journal of Biological Chemistry*, vol. 276, no. 42, pp. 38570–38581, 2001.
- [196] J. Xiao, A. M. Lee, and S. F. Singleton, “Direct evaluation of a kinetic model for RecA-mediated DNA-strand exchange: The importance of nucleic acid dynamics

- and entropy during homologous genetic recombination,” *ChemBioChem*, vol. 7, no. 8, pp. 1265–1278, 2006.
- [197] A. Sakai and M. M. Cox, “RecFOR and RecOR as distinct RecA loading pathways,” *The Journal of Biological Chemistry*, vol. 284, no. 5, pp. 3264–72, 2009.
- [198] K. Morimatsu, Y. Wu, and S. C. Kowalczykowski, “RecFOR proteins target RecA protein to a DNA gap with either DNA or RNA at the 5’ terminus: Implication for repair of stalled replication forks,” *Journal of Biological Chemistry*, vol. 287, no. 42, pp. 35621–35630, 2012.
- [199] M. M. Cox, “The bacterial RecA protein: structure, function and regulation,” in *Topics in Current Genetics. Molecular Genetics of Recombination* (A. Aguilera and R. Rothstein, eds.), pp. 53–94, Heidelberg, Germany: Springer-Verlag GmbH, 2007.
- [200] V. E. Galkin, R. L. Britt, L. B. Bane, X. Yu, M. M. Cox, and E. H. Egelman, “Two modes of binding of DinI to RecA filament provide a new insight into the regulation of sos response by DinI protein,” *Journal of Molecular Biology*, vol. 408, no. 5, pp. 815–824, 2011.
- [201] S. L. Lusetti, O. N. Voloshin, R. B. Inman, R. D. Camerini-Otero, and M. M. Cox, “The DinI protein stabilizes RecA protein filaments,” *Journal of Biological Chemistry*, vol. 279, no. 29, pp. 30037–30046, 2004.
- [202] H. Fu, S. Le, H. Chen, K. Muniyappa, and J. Yan, “Force and ATP hydrolysis dependent regulation of RecA nucleoprotein filament by single-stranded DNA binding protein,” *Nucleic Acids Research*, vol. 41, no. 2, pp. 924–932, 2013.
- [203] S. H. Kim, C. Joo, T. Ha, and D. Kim, “Molecular mechanism of sequence-dependent stability of RecA filament,” *Nucleic Acids Research*, vol. 41, no. 16, pp. 7738–7744, 2013.
- [204] J. Kates-Harbeck, A. Tilloy, and M. Prentiss, “Simplified biased random walk model for RecA-protein-mediated homology recognition offers rapid and accurate

- self-assembly of long linear arrays of binding sites,” *Physical Review E*, vol. 88, no. 1, p. 012702, 2013.
- [205] D. Yang, B. Boyer, C. Prévost, C. Danilowicz, and M. Prentiss, “Integrating multi-scale data on homologous recombination into a new recognition mechanism based on simulations of the RecA-ssDNA/dsDNA structure,” *Nucleic Acids Research*, vol. 43, no. 21, pp. 10251–10263, 2015.
- [206] A. Barzel and M. Kupiec, “Finding a match: how do homologous sequences get together for recombination?,” *Nature Reviews. Genetics*, vol. 9, pp. 27–37, 2008.
- [207] R. D. Camerini-Otero and P. Hsieh, “Parallel DNA triplexes, homologous recombination and other homology-dependent DNA interactions,” *Cell*, vol. 73, pp. 217–223, 1993.
- [208] K. Ragnathan, C. Liu, and T. Ha, “RecA filament sliding on DNA facilitates homology search,” *eLife*, vol. 1, p. e00067, 2012.
- [209] K. Ragnathan, C. Joo, and T. Ha, “Real-time observation of strand exchange reaction with high spatiotemporal resolution,” *Structure*, vol. 19, no. 8, pp. 1064–73, 2011.
- [210] K. Adzuma, “Stable synapsis of homologous DNA molecules mediated by the *Escherichia coli* RecA protein involves local exchange of DNA strands,” *Genes & Development*, vol. 6, no. 9, pp. 1679–1694, 1992.
- [211] E. Folta-Stogniew, S. O’Malley, R. Gupta, K. S. Anderson, and C. M. Radding, “Exchange of DNA base pairs that coincides with recognition of homology promoted by *E. coli* RecA protein,” *Molecular Cell*, vol. 15, no. 6, pp. 965–975, 2004.
- [212] K. D. Dorfman, R. Fulconis, M. Dutreix, and J. L. Viovy, “Model of RecA-mediated homologous recognition,” *Physical Review Letters*, vol. 93, no. 26 I, pp. 3–6, 2004.

- [213] S. M. Honigberg and C. M. Radding, "The mechanics of winding and unwinding helices in recombination: Torsional stress associated with strand transfer promoted by RecA protein," *Cell*, vol. 54, no. 4, pp. 525–532, 1988.
- [214] K. Klapstein, T. Chou, and R. Bruinsma, "Physics of RecA-mediated homologous recognition," *Biophysical Journal*, vol. 87, no. 3, pp. 1466–1477, 2004.
- [215] P. Hsieh, C. S. Camerini-Otero, and R. D. Camerini-Otero, "Pairing of homologous DNA sequences by proteins: evidence for three-stranded DNA," *Genes & Development*, vol. 4, no. 11, pp. 1951–1963, 1990.
- [216] B. J. Rao, S. K. Chiu, and C. M. Radding, "Homologous recognition and triplex formation promoted by RecA protein between duplex oligonucleotides and single-stranded DNA," *Journal of Molecular Biology*, vol. 229, pp. 328–343, 1993.
- [217] J. E. Yancey-Wrona and R. D. Camerini-Otero, "The search for DNA homology does not limit stable homologous pairing promoted by RecA protein," *Current Biology*, vol. 5, no. 10, pp. 1149–58, 1995.
- [218] H.-F. Fan, M. M. Cox, and H.-W. Li, "Developing single-molecule TPM experiments for direct observation of successful RecA-mediated strand exchange reaction," *PLoS ONE*, vol. 6, no. 7, p. e21359, 2011.
- [219] S. L. Lusetti and M. M. Cox, "The bacterial RecA protein and the recombinational DNA repair of stalled replication forks," *Annual Review of Biochemistry*, vol. 71, pp. 71–100, 2002.
- [220] P. Hsieh, C. S. Camerini-Otero, and R. D. Camerini-Otero, "The synapsis event in the homologous pairing of DNAs: RecA recognizes and pairs less than one helical repeat of DNA," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, pp. 6492–6, jul 1992.
- [221] Y. Shigemori and M. Oishi, "Stable triple-stranded DNA formation and its application to the SNP detection," *DNA Research*, vol. 12, no. 6, pp. 441–9, 2005.

- [222] A. A. Volodin, H. A. Smirnova, and T. N. Bocharova, "Efficient interaction of RecA protein with fluorescent dye-labeled oligonucleotides," *FEBS Letters*, vol. 349, no. 1, pp. 65–8, 1994.
- [223] S. Cheng, B. V. Houtentii, H. B. Gamper, A. Sancarn, and J. E. Hearst, "Use of psoralen-modified oligonucleotides to trap three-stranded RecA-DNA complexes and repair of these cross-linked complexes by ABC excinuclease," *Journal of Biological Chemistry*, vol. 263, no. 29, pp. 15110–15117, 1988.
- [224] Q. Shan and M. M. Cox, "RecA protein dynamics in the interior of RecA nucleoprotein filaments," *Journal of Molecular Biology*, vol. 257, no. 4, pp. 756–74, 1996.
- [225] S. L. Lusetti, J. J. Shaw, and M. M. Cox, "Magnesium ion-dependent activation of the RecA protein involves the C terminus," *Journal of Biological Chemistry*, vol. 278, no. 18, pp. 16381–8, 2003.
- [226] S. C. Kowalczykowski, "Biochemistry of genetic recombination: Energetics and mechanism of DNA strand exchange," *Annual Review of Biophysics and Biophysical Chemistry*, vol. 20, pp. 539–75, 1991.
- [227] M. E. Bianchi and C. M. Radding, "Insertions, deletions and mismatches in heteroduplex DNA made by RecA protein," *Cell*, vol. 35, pp. 511–20, 1983.
- [228] W. A. Bedale and M. M. Cox, "Evidence for the coupling of ATP hydrolysis to the final (extension) phase of RecA protein-mediated DNA strand exchange," *Journal of Biological Chemistry*, vol. 271, no. 10, pp. 5725–5732, 1996.
- [229] J. E. Lindsley and M. M. Cox, "Dissociation pathway for RecA nucleoprotein filaments formed on linear duplex DNA," *Journal of Molecular Biology*, vol. 205, no. 4, pp. 695–711, 1989.
- [230] L. Zhang, Y. Wang, M. Chen, Y. Luo, K. Deng, D. Chen, and W. Fu, "A new system for the amplification of biological signals: RecA and complimentary single strand DNA probes on a leaky surface acoustic wave biosensor," *Biosensors and Bioelectronics*, vol. 60, pp. 259–264, 2014.

- [231] L. Tang, D. Li, and J. Li, "Formation of a graphene oxide-DNA duplex-based logic gate and sensor mediated by RecA-ssDNA nucleoprotein filaments," *Chemical Communications*, vol. 49, no. 85, pp. 9971–3, 2013.
- [232] J. N. Milligan and A. D. Ellington, "Using RecA protein to enhance kinetic rates of DNA circuits," *Chemical Communications*, vol. 51, no. 2, pp. 10–13, 2015.
- [233] H. G. Hansma, D. E. Laney, M. Bezanilla, R. L. Sinsheimer, and P. K. Hansma, "Applications for atomic force microscopy of DNA," *Biophysical Journal*, vol. 68, no. 5, pp. 1672–7, 1995.
- [234] K. Umemura, S. Ikawa, T. Nishinaka, T. Shibata, and R. Kuroda, "Imaging the RecA-DNA complex by atomic force microscopy," *Nucleic Acids Symposium Series*, no. 42, pp. 235–6, 1999.
- [235] I. R. Lehman, M. J. Bessman, E. S. Simms, and A. Kornberg, "Enzymatic synthesis of deoxyribonucleic acid," *Journal of Biological Chemistry*, vol. 233, pp. 163–170, 1958.
- [236] T. D. Brock and H. Freeze, "Thermus aquaticus gen. n. and sp. n., a non-sporulating extreme thermophile," *Journal of Bacteriology*, vol. 98, no. 1, pp. 289–297, 1969.
- [237] A. Chien, D. B. Edgar, and J. M. Trela, "Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*," *Journal of Bacteriology*, vol. 127, no. 3, pp. 1550–1557, 1976.
- [238] K. Mullis, F. Faloona, S. Scharf, R. Saiki, G. Horn, and H. Erlich, "Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction," *Cold Spring Harbour Symposium on Quantitative Biology*, vol. 51, pp. 263–273, 1986.
- [239] J. Sambrook and D. W. Russel, "*In vitro* amplification of DNA by the polymerase chain reaction," in *Molecular Cloning, A Laboratory Manual*, ch. 8, Cold Spring Harbour, New York: Cold Spring Harbour Laboratory Press, 3rd ed., 2001.
- [240] H. H. Lee, "Ligase chain reaction," *Biologicals : Journal of the International Association of Biological Standardization*, vol. 24, pp. 197–199, 1996.

- [241] J. M. Berg, J. L. Tymoczko, and L. Stryer, “The purification of proteins is an essential first step in understanding their function,” in *Biochemistry*, ch. 4, New York, USA: Freeman, W H, 5th ed., 2002.
- [242] J. Sambrook and D. W. Russel, “Gel electrophoresis of DNA and pulsed-field agarose gel electrophoresis,” in *Molecular Cloning, A Laboratory Manual*, ch. 5, Cold Spring Harbour, New York: Cold Spring Harbour Laboratory Press, 3rd ed., 2001.
- [243] P. Tosch, C. Wälti, A. P. J. Middelberg, and A. G. Davies, “Generic technique to generate large branched DNA complexes,” *Biomacromolecules*, vol. 7, no. 3, pp. 677–81, 2006.
- [244] H. Bui, C. Onodera, C. Kidwell, Y. Tan, E. Graugnard, W. Kuang, J. Lee, W. B. Knowlton, B. Yurke, and W. L. Hughes, “Programmable periodicity of quantum dot arrays with DNA origami nanotubes,” *Nano Letters*, vol. 10, no. 9, pp. 3367–72, 2010.
- [245] B. Uprety, E. P. Gates, Y. Geng, A. T. Woolley, and J. N. Harb, “Site-specific metallization of multiple metals on a single DNA origami template,” *Langmuir*, vol. 30, pp. 1134–1141, 2014.
- [246] S. M. D. Watson, M. A. Galindo, B. R. Horrocks, and A. Houlton, “Mechanism of formation of supramolecular DNA-templated polymer nanowires,” *Journal of the American Chemical Society*, vol. 136, no. 18, pp. 6649–55, 2014.
- [247] P. W. K. Rothmund, “Design of DNA origami,” in *Proceedings of the 2005 IEEE/ACM International Conference on Computer-aided design*, pp. 471–478, 2005.
- [248] Q. Liu, W. A. Scaringe, and S. S. Sommer, “Discrete mobility of single-stranded DNA in non-denaturing gel electrophoresis,” *Nucleic Acids Research*, vol. 28, no. 4, pp. 940–943, 2000.
- [249] D. Jose, K. Datta, N. P. Johnson, and P. H. von Hippel, “Spectroscopic studies of position-specific DNA ”breathing” fluctuations at replication forks and primer-

- template junctions,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 11, pp. 4231–6, 2009.
- [250] G. Altan-Bonnet, A. Libchaber, and O. Krichevsky, “Bubble dynamics in double-stranded DNA,” *Physical Review Letters*, vol. 90, no. 13, p. 138101, 2003.
- [251] M. Peyrard, S. Cuesta-López, and G. James, “Nonlinear analysis of the dynamics of DNA breathing,” *Journal of Biological Physics*, vol. 35, no. 1, pp. 73–89, 2009.
- [252] J. Fei and T. Ha, “Watching DNA breath one molecule at a time,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 43, pp. 17173–17174, 2013.
- [253] P. H. von Hippel, N. P. Johnson, and A. H. Marcus, “50 years of DNA breathing’: Reflections on old and new approaches,” *Biopolymers*, vol. 99, no. 12, pp. 923–954, 2013.
- [254] D. Jose, S. E. Weitzel, and P. H. von Hippel, “Breathing fluctuations in position-specific DNA base pairs are involved in regulating helicase movement into the replication fork,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 36, pp. 14428–14433, 2012.
- [255] J. Aymami, M. Coll, G. A. van der Marel, J. H. van Boom, A. H. Wang, and A. Rich, “Molecular structure of nicked DNA: a substrate for DNA repair enzymes,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 87, no. 7, pp. 2526–30, 1990.
- [256] J. S. Smith and E. P. Nikonowicz, “Phosphorothioate substitution can substantially alter RNA conformation,” *Biochemistry*, vol. 39, pp. 5642–5652, 2000.
- [257] R. Roy, A. G. Kozlov, T. M. Lohman, and T. Ha, “SSB protein diffusion on single-stranded DNA stimulates RecA filament formation,” *Nature*, vol. 461, pp. 1092–1097, 2009.
- [258] K. Gekko and S. N. Timasheff, “Mechanism of protein stabilization by glycerol: preferential hydration in glycerol-water mixtures,” *Biochemistry*, vol. 20, no. 16, pp. 4667–4676, 1981.

- [259] V. Vagenende and B. L. Trout, "Quantitative characterization of local protein solvation to predict solvent effects on protein structure," *Biophysical Journal*, vol. 103, no. 6, pp. 1354–1362, 2012.
- [260] M. M. Cox, K. McEntee, and I. R. Lehman, "A simple and rapid procedure for the large scale purification of the RecA protein of *Escherichia coli*," *Journal of Biological Chemistry*, vol. 256, no. 9, pp. 4676–4678, 1981.
- [261] Y. Mou, J.-Y. Yu, T. M. Wannier, C.-L. Guo, and S. L. Mayo, "Computational design of co-assembling proteinDNA nanowires," *Nature*, vol. 525, pp. 230–233, 2015.
- [262] S. M. Douglas, A. H. Marblestone, S. Teerapittayanon, A. Vazquez, G. M. Church, and W. M. Shih, "Rapid prototyping of 3D DNA-origami shapes with caDNAno," *Nucleic Acids Research*, vol. 37, no. 15, pp. 5001–6, 2009.
- [263] M. Senior, R. A. Jones, and K. J. Breslauer, "Influence of dangling thymidine residues on the stability and structure of two DNA duplexes," *Biochemistry*, vol. 27, no. 10, pp. 3879–85, 1988.
- [264] M. J. Doktycz, T. M. Paner, M. Amaratunga, and A. S. Benight, "Thermodynamic stability of the 5 dangling-ended DNA hairpins formed from sequences 5-(XY) 2GGATAC (T) 4GTATCC-3, where X, Y= A, T, G, C," *Biopolymers*, vol. 30, pp. 829–845, 1990.
- [265] J. SantaLucia and D. Hicks, "The thermodynamics of DNA structural motifs," *Annual Review of Biophysics and Biomolecular Structure*, vol. 33, pp. 415–40, 2004.
- [266] M. Kara and M. Zacharias, "Stabilization of duplex DNA and RNA by dangling ends studied by free energy simulations," *Biopolymers*, vol. 101, no. 4, pp. 418–427, 2014.
- [267] M. Gueron, M. Kochoyan, and J. Leroy, "A single mode of DNA base-pair opening drives imino proton exchange," *Nature*, vol. 328, pp. 89–92, 1987.