

**Concurrent DNA transcription from  
convergent and tandem promoters studied  
by atomic force microscopy**

Oliver Chammas

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

The University of  
Leeds

School of Electronic & Electrical Engineering  
School of Physics and Astronomy  
Department of Oral Biology

May 2016

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Chapter 5: Contains material published as joint publication by the candidate and the co-authors DJ Billingsley, WA Bonass and NH Thomson.

Publication: Chammas, O., Billingsley, D. J., Bonass, W. A., & Thomson, N. H. (2013). Single-stranded DNA loops as fiducial markers for exploring DNA–protein interactions in single molecule imaging. *Methods*, 60(2), 122-130.

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.

The right of Oliver Antoine Andre Chammas to be identified as Author of this work has been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

© 2016 The University of Leeds and Oliver Antoine Andre Chammas

# Acknowledgements

I would firstly like to express my sincerest gratitude and thanks to my two supervisors Neil Thomson and Bill Bonass who without their help, insight and direction this work would not have been possible. Your patience, expertise and support helped me see the light at the end of the tunnel. I would also like to thank all those who helped me along the way, mainly Ashley Stammers, Dan Burt and all those in the Molecular Nanoscale Physics group. Your conversations and friendship made my PhD immeasurably better.

I would also like thank and acknowledge the CDT for Molecular Scale Engineering for all the personal and professional support throughout my PhD. I would like to especially thank my colleagues and friends who started this journey with me, Andrew Lee, William Morton, Sybilla Corbett, Mark Sellick and Scott Bird. Without your friendship and support through the good and the bad times I would not be where or who I am now. I also want to thank Tom Milbank, you provided me with the drive to better myself, not just in the gym and your friendship gave me a chance to vent and enjoy life outside of my PhD. I would also like to thank my mother father and all my family who have been so supportive throughout my PhD.

Finally I want to express my thanks and love to my wonderful wife Lucy. You are my everything. Your encouragement and never ending patience have helped see me through to completing this work and your unquestioning love has helped keep me sane throughout.

# Abstract

With recent advances in sequencing and mapping of genomes, the occurrence of overlapping and nested transcription units is more common than previously thought in both eukaryotes and prokaryotes. The interleaved genome model means that transcriptional interference by collisions between concurrently transcribing RNA polymerases is more likely than ever before. This thesis presents a study of the outcomes of collisions between RNAPs transcribing concurrently from convergent and tandem promoters using AFM to provide a view of single populations seen after collisions.

Through the development of an improved DNA end labelling method and incorporation of an inhibitor of RNAP non-specific binding the results of collisions can be viewed with more confidence than previously possible. It was seen that collisions from both convergent and tandem promoters resulted in both RNAPs remaining bound to the template in hard contact. These collisions occurred by two main mechanisms. Either between two active elongation complexes (ECs) or between an elongation complex and an inactive complex referred to as a sitting duck (SD). EC-EC collisions were found to be the most common for convergent promoters while with tandem promoters the distinction between the two is less clear. In the case of EC-SD collisions it is shown that shunting upstream of up to 100 bp by an EC is possible.

By utilizing a linear template that is susceptible to supercoiling due to spin locking, it is shown that a region of highly positive supercoiled domain can prevent two convergently transcribing RNAPS coming into hard contact. It is also shown that topology of the DNA plays a role in the distribution of EC-EC and EC-SD collisions that occur for both promoter arrangements. This indicates that topology influences the outcomes of concurrent transcription and provides a mechanism by which RNAPs can sense one another via the DNA template.



This work is dedicated to Lucy my wife and my dog Theo. You both bring so much joy and happiness into my life.

# Contents

Acknowledgements .....	iii
Abstract .....	iv
Dedication .....	v
Contents .....	vi
List of abbreviations .....	ix
List of figures .....	xi
List of tables.....	xx
1 Introduction.....	1
1.1 Project aims .....	2
1.2 Synopsis .....	3
1.3 Chapter overview .....	5
2 Introduction to nucleic acids, DNA and transcription .....	9
2.1 History of DNA.....	9
2.2 Structure of DNA .....	10
2.2.1 Nucleic acids .....	10
2.2.2 The DNA double helix.....	12
2.2.3 RNA.....	15
2.2.4 DNA secondary structures and topology.....	15
2.2.5 Supercoiling in the cell .....	21
2.3 Central dogma of molecular biology .....	23
2.3.1 RNA polymerase structure .....	24
2.3.2 Process of transcription.....	28
2.3.3 Prokaryotic and eukaryotic gene organisation.....	39
2.3.4 Orientation and gene pairs and nested genes .....	41
2.3.5 Transcriptional interference and collisions .....	44
3 Introduction to AFM of DNA protein complexes.....	50
3.1 Atomic force microscopy for investigating transcription .....	50
3.1.1 Scanning probe microscopy.....	51
3.1.2 AFM instrumentation .....	52
3.1.3 Forces in the AFM.....	54
3.1.4 Imaging Modes .....	56
3.1.5 High speed AFM.....	57

3.2	Imaging DNA and proteins .....	60
3.2.1	Imaging substrates .....	60
3.2.2	Preparation of samples .....	63
3.3	Studying transcription by AFM .....	67
3.3.1	Practical considerations for studying biological samples by AFM .....	68
3.3.2	Previous AFM studies of transcription .....	72
4	General Methods .....	79
4.1	Introduction .....	79
4.2	Preparation of DNA constructs .....	80
4.2.1	Transcription templates .....	80
4.2.2	Polymerase chain reaction .....	82
4.2.3	Gel electrophoresis .....	84
4.2.4	Column purification .....	86
4.2.5	Measuring DNA concentration .....	87
4.3	<i>In vitro</i> transcription reactions .....	88
4.3.1	<i>In vitro</i> transcription: Open promoter complex formation .....	88
4.3.2	<i>In vitro</i> transcription: initiation of elongation .....	89
4.4	Imaging of DNA samples and <i>in vitro</i> transcription complexes .....	90
4.4.1	Sample deposition .....	91
4.4.2	Rinsing and drying .....	92
4.4.3	Analysing samples .....	92
5	High throughput labelling .....	96
5.1	Introduction .....	96
5.1.1	Labelling DNA for AFM .....	97
5.1.2	Nucleic acid based end label for AFM .....	98
5.2	Samples and preparation .....	101
5.3	PCR based labelling method design and results .....	102
5.3.1	Labelled single promoter .....	107
5.3.2	Single promoter Open promoter complexes (OPCs) .....	112
5.3.3	Elongated complexes .....	116
5.3.4	Labelling of two promoter templates .....	120
5.4	Discussion .....	124
5.4.1	Labelling .....	124
5.4.2	Effect of loop label on transcription .....	126
5.5	Conclusions .....	132
6	Convergent and tandem transcription in the presence of a non-specific binding inhibitor 134	

6.1	Introduction.....	134
6.2	Results .....	145
6.2.1	Incorporating heparin into AFM samples.....	145
6.2.2	Investigation into transcriptional collisions for <i>E.coli</i> RNA polymerase in the presence of heparan sulphate.....	154
6.2.3	Labelled convergent promoter templates.....	155
6.2.4	Transcription of tandem promoter template.....	167
6.3	Discussion .....	174
6.3.1	Incorporation of polyanionic inhibitors of RNAP non-specific binding .....	174
6.3.2	Effects of HS on OPC formation and imaging.....	176
6.3.3	Outcomes of convergent transcription in the presence of HS.....	179
6.3.4	Transcription from tandem promoters .....	182
6.4	Conclusions.....	185
7	Effects of supercoiling on transcriptional collisions .....	188
7.1	Introduction.....	188
7.1.1	Supercoiling and transcription .....	189
7.1.2	Topoisomerases.....	196
7.2	Sample preparation .....	202
7.3	Results .....	204
7.3.1	AFM analysis of wheat germ Topo IB.....	204
7.3.2	Outcomes of transcription from convergent promoters on a 2521 bp template with and without Topo IB.....	213
7.3.3	Outcomes of tandem transcription of a 2521 bp template with and without Topo IB	222
7.4	Discussion .....	229
7.4.1	Topoisomerase IB.....	229
7.4.2	Convergent transcription and the effects of Topo IB.....	231
7.4.3	Concurrent transcription from tandem promoters with and without Topo IB	236
7.5	Conclusions.....	240
8	Final conclusions and future work.....	243
8.1	Final conclusions.....	243
8.1.1	Improved imaging of DNA protein complexes .....	243
8.1.2	Transcriptional interference and the effects of transcription-coupled supercoiling.	245
8.2	Future work .....	247
9	Bibliography.....	251

# List of abbreviations

<b>3D</b>	3-Dimensional
<b>A</b>	Adenine
<b>ADP</b>	Adenine diphosphate
<b>AFM</b>	Atomic force microscopy
<b>AM</b>	Amplitude modulation
<b>AMP</b>	Adenine monophosphate
<b>ATP</b>	Adenine triphosphate
<b>Bp</b>	base pair
<b>C</b>	Cytosine
<b>C.Elegans</b>	Caenorhabditis elegans
<b>CPC</b>	Closed promoter complex
<b>DNA</b>	Deoxyribonucleic acid
<b>dNTP</b>	deoxy-nucleotide triphosphate
<b>dsDNA</b>	double stranded DNA
<b>DTT</b>	Dithiothreitol
<b><i>E.coli</i></b>	Escherichia coli
<b>EC</b>	Elongated complex
<b>EDTA</b>	Ethylenediaminetetraacetic acid
<b>FRET</b>	Förster resonance energy transfer
<b>G</b>	Guanine
<b>GlcNAc</b>	N-acetyl-D-glucosamine
<b>GlcNS</b>	N-sulphated-D-glucosamine
<b>HOPG</b>	highly oriented pyrolytic graphite
<b>HS</b>	High speed
<b>ITC</b>	Initially transcribing complex
<b>mRNA</b>	messenger RNA
<b>ncRNA</b>	Non coding RNA
<b>Nt</b>	Nucleotide

<b>NTP</b>	Nucleotide triphosphate
<b>OPC</b>	Open promoter complex
<b>PCR</b>	Polymerase chain reaction
<b>ppGpp</b>	Guanosine pentaphosphate
<b>Ppi</b>	Pyrophosphate
<b>RMS-DC</b>	Root mean squared to DC
<b>RNA</b>	Ribonucleic acid
<b>RNAP</b>	Bacterial RNA polymerase
<b>RNAP II</b>	Eukaryotic RNA polymerase II
<b>Rut</b>	Rho utilisation signal
<b>SD</b>	Sitting duck
<b>SEC</b>	Stalled elongation complex
<b>SPM</b>	Scanning probe microscopy
<b>ssDNA</b>	single stranded DNA
<b>STM</b>	Scanning tunneling microscopy
<b>T</b>	Thymine
<b>TAE</b>	Tris-Acetate EDTA
<b>Taq</b>	Thermus Aquaticus DNA polymerase
<b>TFO</b>	Triplex forming oligonucleotide
<b>TIRFm</b>	Total internal reflection fluorescence microscope
<b>Topo</b>	Topoisomerase
<b>Topo IB</b>	Wheat Germ Topoisomerase IB
<b>tRNA</b>	transfer RNA
<b>U</b>	Uracil
<b>UTP</b>	Uracil triphosphate
<b>UV</b>	Ultra violet
<b>(w/v)</b>	Weight to volume
<b><math>\sigma^{54}</math>RNAP</b>	Sigma 54 <i>E.coli</i> RNA polymerase Holoenzyme
<b><math>\sigma^{70}</math>RNAP</b>	Sigma 70 <i>E.coli</i> RNA polymerase Holoenzyme
<b><math>\sigma</math>RNAP</b>	<i>E.coli</i> RNA polymerase Holoenzyme

# List of figures

Figure 2-1: Sugar units found in nucleic acids ribose (RNA) and deoxyribose (DNA). It can be seen that deoxyribose lacks the oxygen atom the carbon 2'. .....	10
Figure 2-2: Nucleic acid bases that occur in DNA and RNA. ....	11
Figure 2-3: Nucleotides are joined together via a phosphate group linkage to form a chain. The polarity of the chain is denoted by the 5' phosphate and the 3' hydroxyl group. ....	12
Figure 2-4: Structure and base pairing of DNA. ....	13
Figure 2-5: The three main conformations of DNA.....	14
Figure 2-6: Non-B DNA structures.....	16
Figure 2-7: Examples of hairpin and cruciform structures formed by DNA. ....	17
Figure 2-8: Representation of the changes of DNA when strands are separated. ....	18
Figure 2-9: The introduction of supercoiling by cellular processes].....	19
Figure 2-10: Diagram depicting the link between changes in twist ( $T_w$ ) and writhe ( $W_r$ ). ....	21
Figure 2-11: The central dogma of molecular biology .....	24
Figure 2-12: Structure of bacterial RNAP holoenzyme .....	25
Figure 2-13: Crystal structure of bacterial RNAP with key elements highlighted, the $\sigma^{70}$ is removed but outlined in red. ....	26
Figure 2-14: Cut through cartoon of bacterial RNAP. ....	27
Figure 2-15: RNAP holoenzyme bound to a promoter DNA sequence.....	28
Figure 2-16: Mechanisms of promoter location by RNAP. ....	29
Figure 2-17: Process of isomerisation. The RNAP holoenzyme binds the promoter...	30
Figure 2-18: 3D model of the wrapping of DNA around RNAP holoenzyme.. ....	31
Figure 2-19: Cartoon of an elongating RNAP with a zoom of the active site shown....	33
Figure 2-20: Schematic of the elongation process.....	35
Figure 2-21: Depiction of a backtracked RNAP with GreB bound. ....	37
Figure 2-22: Diagrams of intrinsic and Rho dependant termination.....	38
Figure 2-23: The <i>E.coli lac</i> operon.....	39
Figure 2-24: Diagram showing the organisation of a eukaryotic gene.....	40
Figure 2-25: Diagram of different promoter arrangements.....	41
Figure 2-26: Schematic representation of nested genes.....	42

Figure 2-27: Possible mechanisms of TI as predicted by Shearwin <i>et al.</i> [2].....	45
Figure 2-28: Miller spread electron micrograph of <i>E.coli</i> rRNA 16s and 23s genes. ....	47
Figure 3-1: Diagrams of the most common AFM probe designs. ....	52
Figure 3-2: Schematic of basic sample scanning AFM setup .....	53
Figure 3-3: Graphical representation of the Lennard-Jones potential .....	55
Figure 3-4: Schematic of the stacked piezo scanner as developed by Ando.....	59
Figure 3-5: Model of muscovite mica layered structure.....	61
Figure 3-6: Counter ion correlation mediated between mica and DNA.....	62
Figure 3-7: Flow diagram of general sample preparation for AFM imaging of DNA- protein samples.....	63
Figure 3-8: Diagram depicting the two mechanisms by which DNA adsorbed to a mica surface.....	65
Figure 3-9: Diagram showing the effect of tip convolution on imaged sample size .....	69
Figure 3-10 : Diagram and AFM image of a DNA origami frame .....	71
Figure 3-11: Schematic representation of the wrapping of DNA around RNAP].....	72
Figure 3-12: AFM scans of an elongating RNAP collected by Suzuki <i>et al.</i> ....	74
Figure 3-13: Computer drawn schematic of a DNA template attached to an origami tile, as was used by Endo <i>et al.</i> ....	75
Figure 3-14: Diagram of the arrangement of the DNA template and RNA transcript in an SEC.....	75
Figure 4-1: Diagrams of the plasmids pDSU and pDSP. ....	80
Figure 4-2: Schematics of the templates used.....	81
Figure 4-3: Diagram detailing the steps involved in a PCR. ....	83
Figure 4-4: Process of column based DNA purification. ....	87
Figure 4-5: Scale bars used for all AFM height and phase images throughout the thesis. ....	93
Figure 4-6: AFM height images with the tracing lines of contour lengths measured from the loop label and from RNAPs shown (scale bars = 50nm). ....	94
Figure 5-1: Biotin labelled dNTPs can be incorporated into a DNA fragment.....	97
Figure 5-2: Process of the labelling protocol devised by Billingsley <i>et al.</i> ....	99
Figure 5-3: Comparison of the predicted structure of a labelled OPC and an AFM image of a labelled OPC. ....	100



Figure 5-4: Diagram of the 602bp DNA template used to test new labelling method. .....	101
Figure 5-5: A) Sequence of the loop primer and B) diagram of labelled template ....	102
Figure 5-6: MFOLD results for the four loop forming primers used with 20 base loops of polyA, polyC, polyG or polyT.....	103
Figure 5-7: Fragments produced from GoTaq polymerase labeling PCR.....	104
Figure 5-8: Schematic representation of the PCR labelling method .....	105
Figure 5-9: Image of 1% (w/v) agarose gel of the products from Q-BIO Taq polymerase PCR. Lanes L: 100bp ladder (Promega) ; 1: Poly(A) primer products ; 2: Poly(T) primer products; 3: Poly(C) primer products ; 4: Poly(G) primer products.....	107
Figure 5-10: Histogram plots of labelled single promoter DNA with corresponding AFM images for each loop sequence .....	108
Figure 5-11: Histogram plot of contour length measurements for poly(G) loop labelled DNA fragment (n= 385).....	110
Figure 5-12: Histograms of the loop diameter and height for all loop sequences.....	110
Figure 5-13: 3D rendering of a labelled fragment with the cross-sectional analysis shown below. ....	111
Figure 5-14: Histograms of contour length for OPCs formed on each loop labelled DNA fragment.....	113
Figure 5-15: A 3D rendering of an OPC with the loop structure imposed on the bottom of the image to provide comparison to the predicted structure .....	114
Figure 5-16: Montage of images showing different forms of complexes seen in OPC samples. ....	115
Figure 5-17: Histogram plot comparing all loop sequence OPC contour lengths with all loop sequence contour lengths of molecules with an RNAP bound to the loop end of the DNA fragment. ....	116
Figure 5-18: Montage of AFM height images showing the elongated complexes after the addition of NTPs (scale bar = 50 nm).....	117
Figure 5-19: Histogram plots of elongated complexes contour length with corresponding gaussian fitted lines (n= 257, 209, 215 and 109 for C, T, A and G respectively).....	117
Figure 5-20: Image of formaldehyde agarose gel run on elongated complexes with and without an end C-loop. ....	119

Figure 5-21: Diagram of labelled convergent and tandem promoter containing DNA templates. ....	120
Figure 5-22: Image of 1% (w/v) agarose gel showing products from labelling reaction on both tandem and convergent promoter templates .....	121
Figure 5-23: Montage of images obtained from the band of greater size seen in Lane 4, Figure 5-22. (Scale bars = 50nm).....	122
Figure 5-24: Histogram plot of labelled tandem promoter DNA. (n=100) .....	123
Figure 5-25: Histogram plot of labelled convergent promoter DNA. (n=101).....	123
Figure 5-26: Line drawings depicting possible structures that stretches of G bases can adopt, directionality is denoted by arrows. (adapted from [365]).....	125
Figure 5-27: Diagram of DNA wrapping in the formation of an OPC.....	127
Figure 5-28: Figure showing a schematic diagram of an OPC.....	129
Figure 5-29: Consequences of hairpin loops or slipped strand structures on RNAP elongation .....	130
Figure 5-30: Formation of a hybrid between the DNA and RNA transcript which can lead to stalling of elongation. ....	131
Figure 6-1: Diagrams displaying the suggested mechanisms which RNAP uses to locate its promoter .....	134
Figure 6-2: AFM height image of plasmid molecules incubated with RNAP. ....	135
Figure 6-3: AFM height image of 6136 bp linear DNA containing two promoters.....	136
Figure 6-4: AFM height images of transcription complexes taken from Crampton <i>et al.</i> .....	138
Figure 6-5: Histogram plots of the inter-RNAP contour length taken from Billingsley <i>et al</i> [293]. ....	139
Figure 6-6: Scatter plots of RNAP positions after elongation from convergent promoters collected by Billingsley <i>et al.</i> .....	140
Figure 6-7: Monosaccharide building blocks that make up heparin and HS. ....	141
Figure 6-8: Repeating disaccharide units that make up heparin and HS.....	142
Figure 6-9: Stick and space fill models of the two main conformations adopted by heparin and HS.....	143
Figure 6-10: AFM phase and height images of heparin and HS deposited on mica...	146
Figure 6-11: AFM height images of random structures formed by DNA and heparin (A) or DNA and HS (B) deposited onto mica.....	147

Figure 6-12: AFM images of DNA with heparin or HS deposited on mica where closely packed DNA can be seen) .....	148
Figure 6-13: AFM height and phase images of areas where DNA was visible but not clear.....	148
Figure 6-14: AFM height images of regions of closely packed but visible DNA seen for samples with heparin or HS. ....	149
Figure 6-15: AFM height images of OPC complexes heated to 37°C in the presence of heparin or HS.....	151
Figure 6-16: AFM height images of OPCs formed on tandem promoter DNA without any inhibitor (top) and with heparin (bottom left) or HS (bottom right).....	153
Figure 6-17: Diagram of the DNA templates with RNAPs bound at their promoters.	155
Figure 6-18: Montage of images showing OPCs formed on the convergent promoter template.....	156
Figure 6-19: Scatter plot of template position for each RNAP pair.....	157
Figure 6-20: Histogram plots of bare DNA and OPC total contour length.....	158
Figure 6-21: Histogram plot of OPC arm lengths.....	159
Figure 6-22: Montage of images showing collided complexes.....	160
Figure 6-23: Histograms comparing inter-RNAP contour length for OPC and elongated complexes. ....	161
Figure 6-24: Gaussian fit of inter-RNAP contour length of elongated complexes (n=284). ....	162
Figure 6-25: Histogram of inter-RNAP distance. A Gaussian fitted curve is shown in blue with values of the fitting shown in the inset table (n=284).....	162
Figure 6-26: Diagram depicting the effects of tip convolution of inter-RNAP separation. ....	163
Figure 6-27: Scatter plots of positions of RNAP pairs as a percentage of the total contour length of the template .....	164
Figure 6-28: Schematic representation of different classes of complexes seen for transcription from convergent promoters.....	165
Figure 6-29: Histogram plot of the movement upstream of the respective promoter travelled as a percentage of the total contour length for both long arm (blue) and short (red) arm promoters.....	166

Figure 6-30: Montage of images showing complexes with two OPCs formed on the tandem promoter template. (Scale bars = 50 nm) .....	167
Figure 6-31: Montage of AFM height images of RNAPs after elongation from two tandem promoters (Scale bars = 200 nm) .....	168
Figure 6-32: Histogram plot of total contour length of OPCs (black and red, n=149) and elongated complexes (orange and green, n=204) .....	169
Figure 6-33: Histogram plot of inter-RNAP contour length with a Gaussian distribution fitted to the main peak (n=204). .....	170
Figure 6-34: Histogram plot of inter-RNAP distance with the main peak fitted with a Gaussian curve (n=204). .....	171
Figure 6-35: Scatter plot of the positions of RNAP pairs from promoters aligned tandemly as a percentage of the total contour length .....	172
Figure 6-36: Schematic representations of the complexes seen after elongation from tandemly arranged promoters. ....	173
Figure 6-37: Comparison of electron microscopy and AFM height image of 200 µg/ml heparin absorbed onto a mica surface .....	174
Figure 6-38: Images of heparin deposited on mica collected by Zhang <i>et al</i> .....	175
Figure 6-39: Schematic representation of degree of wrapping for convergent and tandem promoter arrangements. ....	178
Figure 6-40: Scatter plot of just those RNAP pairs that displayed backtracking from the promoter which is indicated by the blue line. ....	181
Figure 6-41: Annotated scatter plot of location of RNAPs pairs as a percentage of the template .....	183
Figure 7-1: AFM height images of circularised template .....	189
Figure 7-2: Schematic diagram representing the generation of supercoils according to the twin supercoil domain theory .....	190
Figure 7-3: Experimental setup and data collected by Ma <i>et al.</i> using an optical trap setup. A) The experimental setup involved the attachment of an RNAP (green) to a surface and applying either an assisting or resistive torque to the DNA .....	192
Figure 7-4: Schematic representations of DNA rotation when considered either as a rigid rod (A) or a naturally bent semi-flexible rod .....	193
Figure 7-5: Diagram of expected supercoiling between both convergent and tandem promoter arrangements. ....	195

Figure 7-6: Crystal structure of Topoisomerase IB from Vaccinia and diagram depicting the relaxation of both positive and negative supercoils. ....	199
Figure 7-7: Montage of AFM and electron microscopy images of synapses and filaments seen for binding of Vaccinia Topo IB.....	201
Figure 7-8: AFM images of calf thymus Topo IB deposited on mica.....	202
Figure 7-9: Diagram showing both convergent and tandem promoter templates ....	203
Figure 7-10: AFM height images of wheat germ Topo IB at differing concentrations deposited on mica and imaged in air.....	205
Figure 7-11: Histogram plots of the diameter and height of Topo IB.....	206
Figure 7-12: AFM images of supercoiled plasmid and plasmid treated with Topo IB at room temperature .....	207
Figure 7-13: AFM height images of plasmid treated with 10 U of Topo IB in the presence of HS. ....	208
Figure 7-14: AFM height images of supercoiled plasmid incubated with RNAP and HS then treated with Topo IB.....	209
Figure 7-15: Column plot of number of crossovers for different sample compositions .....	210
Figure 7-16: AFM height images of Topo IB with linear tandem promoter 2521 bp DNA template with and without HS present .....	210
Figure 7-17: Histogram plots of convergent and tandem linear 2521 bp DNA template with 10 U of Topo IB.....	211
Figure 7-18: AFM height image with cross section plot shown below. The image is 918 nm by 918 nm. ....	212
Figure 7-19: Histogram plots of the contour length of the DNA to the first Topo IB molecule for both tandem and convergent promoter templates (n= 100 for both) .	213
Figure 7-20: Scatter plot of RNAP positions plotted as a percentage of total contour length. The average position is shown by the blue dashed line.....	214
Figure 7-21: AFM height images of OPCs formed on a 2521 bp template with convergent promoters. (Scale bar = 100 nm) .....	215
Figure 7-22: Histogram plots of the inter RNAP contour length after elongation from convergent promoters. ....	216
Figure 7-23: Histogram plots of inter-RNAP distance in the absence and presence of Topo IB after elongation from convergent promoters .....	217

Figure 7-24: Scatter plot of positions of RNAP pairs after elongation without Topo IB (left) and with Topo IB (right) .....	219
Figure 7-25: AFM height images of complexes imaged after the addition of NTPs on a 2521 bp template).....	220
Figure 7-26: Distribution of individual RNAPs along the template. Each count represents a RNAP at a given position.....	221
Figure 7-27: AFM height images of OPCs formed on tandem promoter 2521 bp template DNA (Scale bars = 100nm).....	222
Figure 7-28: Scatter plot of RNAP pair positions plotted as a percentage of the template for double OPCs formed on tandem promoter 2521 bp template DNA.....	223
Figure 7-29: Histogram plots of inter RNAP contour length and direct inter-RNAP distance for RNAPs elongating from tandem promoters either with (n=169) or without (n=242) Topo IBs. ....	225
Figure 7-30: Scatter plots of RNAP positions plotted as a percentage of the template. Without Topo IB (left) it can be seen that a large proportion of RNAP pairs are located at the leading promoter. With the addition of Topo IB the RNAP pairs are mainly located downstream of the leading promoter. ....	227
Figure 7-31: Distribution of individual RNAPs located along the template as a percentage of the total contour length. It can be seen that the leading RNAP (top panel) travels further before stalling with Topo IB present. This can also be seen for the trailing RNAP (bottom panel). Without Topo IB is shown in black and with Topo IB is shown in red. Dark red areas are overlaps of the two data sets.....	228
Figure 7-32: Cartoon image drawn from the crystal structure of human Topo IB bound to double stranded DNA [447]......	230
Figure 7-33: Schematic diagram showing the possible role that plectonemic structures may play in the greater, inter RNAP separation seen between the two RNAPs when Topo IB is not present to relieve supercoiling.....	232
Figure 7-34: AFM height images of stalled elongation complexes from convergent promoters with non-linear structures seen between the RNAPs.....	232
Figure 7-35: AFM height images of looped structures seen in samples of elongated complexes from convergent promoters in the presence of Topo IB (top image) .....	233
Figure 7-36: Scatter plots of Class 2 molecules from samples using the 2521 bp template without Topo IB (top left) and with Topo IB (top right)......	235

Figure 7-37: Histograms showing the inter RNAP contour length recorded for the 1144 bp template (blue n= 75) and the 2521 bp template with (green n= 65), and without Topo IB (red, n= 172)..... 237

Figure 7-38: Scatter plot of position of RNAPs on the template for samples with and without Topo IB..... 239

# List of tables

Table 2-1: Structural features of the main forms of DNA [27].	15
Table 4-1: Table of tip specifications used for imaging.	90
Table 4-2: Table of AFM setting used for imaging	91
Table 5-1: PCR conditions used for labelling with GoTaq polymerase	103
Table 5-2: PCR conditions used for the Q-BIO Taq polymerase	106
Table 5-3: Reaction mix for Q-BIO Taq polymerase PCR	106
Table 5-4: Contour lengths of DNA fragments produced with each loop primer.	109
Table 5-5: Height and diameter of the end loop feature for all loop sequences.	111
Table 5-5-6: DNA arm contour lengths for each loop sequence. The loop is at the end of the long arm of this single promoter template.	114
Table 5-7: Measurements of contour length of complexes after the addition of NTPs.	118
Table 6-1: Number of RNAPs bound to the surface as singular RNAPs or aggregates in the presence of heparin or HS.	150
Table 6-2: Percentages of complexes formed without an inhibitor and with heparin or HS. The contour lengths of molecules are also shown with single OPCs showing a decrease in contour length of approximately 25 nm, double OPC a decrease of 51 nm and the non-specifically bound DNA showing no decrease.	152
Table 6-3: Summary of measurements for convergent promoter complexes.	160
Table 6-4: Summary of measurements for tandem promoter template.	169
Table 7-1: Table of known topoisomerases showing information on type, source, structure and functionality (adapted from [27]).	198
Table 7-2: Average contour length measurements for short and long arms and inter RNAP separation for OPCs formed on a DNA template of 2521 bp with convergent promoters.	214
Table 7-3: Average contour length measurements for DNA arms and inter-RNAP contour length after the addition of NTPs with and without the addition of Topo IB for elongation from convergent promoters.	215
Table 7-4: Values of the centre of each peak from fitting the data with three Gaussian distributions. The standard error values are obtained by dividing the standard deviation of the plot by the square root of number of measurements.	218



Table 7-5: Average contour length measurements for double OPCs formed on tandem promoter 2521 bp template. ....	223
Table 7-6: Average contour length measurements of complexes elongated from tandem promoters without and with Topo IB. ....	224
Table 7-7: Values of the centre of peaks fitted to histogram plots with the standard error calculated from the standard deviation of the fitted curves. ....	226

This page is left intentionally blank

# 1 Introduction

The organisation of genomes is highly complex where it is being increasingly found that information is compressed within degenerate sequences. There are a number of genes which have transcriptional start sites nested within other genes as well as a number of genes with sections that overlap. As the technology to map genomes has developed, new genes as well as a large number of non-coding transcriptional start sites have been discovered, indicating that nested and overlapping genetic units are more common than previously believed [1]. These overlaps can occur with start sites directing transcription in the same direction (tandem) or in opposite directions, either directing transcription towards each start site (convergent) or away from each other (divergent). These arrangements allow for transcriptional machinery to interact with each other. The outcomes of these interactions has been shown to lead to alterations to the process of transcription through mechanisms known collectively as transcriptional interference (TI) [2]. One important mechanism of TI that can occur is collisions between RNA polymerase (RNAP) molecules. Investigations into collisions has often been performed using bulk biochemical methods to provide insight into the outcomes of collisions as well as the effects these events may have on gene expression. Bulk methods do not provide information about single collision events or the sub-populations of outcomes. This has led to the use of single molecule techniques such as atomic force microscopy (AFM) which can be a powerful tool to understanding sub-populations. The advantage of single molecule studies is that individual expression events can be monitored as opposed to population average events that are measured in “ensemble-type” experiments. AFM allows the study of collisions between two single RNAP molecules one complex at a time. Previous studies by Crampton *et al.* and Billingsley *et al.* were able to show that AFM is well suited to the study of these collision events [3, 4]. Both found that collisions tended to occur by two main mechanisms. Either between two elongating RNAP complexes (EC-EC collision) or between an actively elongating complex and an inactive complex, termed a sitting duck (EC-SD collisions). A number of issues were noted though when performing studies using *ex situ* AFM. Firstly, the direction travelled by each RNAP is not possible to determine after elongation had occurred. This was remedied by the use of a single stranded DNA

## Chapter 1: Introduction

labelling approach by Billingsley *et al.*, however, this method was relatively inefficient and gave low yields of labelled transcriptional templates [5]. The second issue was the inability to ensure that only specifically bound complexes were studied from a single round of transcription due to the lack of an inhibitor of non-specific binding being included in the preparation of samples.

Another interesting avenue not previously investigated is the importance that the DNA template itself plays in the occurrence of collisions. The process of transcription has been shown to alter the topological state of the DNA template. The effects that topology plays in TI and gene expression are not fully understood but it is expected that the role of topology is much greater than previously believed at all stages of transcription [6].

### 1.1 Project aims

The aims of this project were to build upon the methods used by Crampton *et al.* and Billingsley *et al.* to investigate RNAP collisions during *in vitro* concurrent transcription using *ex situ* AFM. The further development of the labelling method by Billingsley *et al.* to provide a high throughput and high yield version of this method is presented in this thesis. This was achieved through the use of a polymerase chain reaction labelling method which allows for the exponential increase in labelled template. The effects of the single stranded DNA label as well as the importance that the sequence of the label may have on transcription were also investigated.

As previously mentioned, the lack of an inhibitor of non-specific interactions between the RNAP and DNA template means that some uncertainty exists in previous studies of not just RNAP:DNA interactions but for many protein:DNA interactions investigated by AFM. To overcome this uncertainty, a method for the incorporation of such an inhibitor was required. This method was then applied to convergent and tandem, single round transcription events. The outcomes of concurrent transcription were then assessed with a higher degree of certainty than previously possible.

## *Chapter 1: Introduction*

As the development of this method allows for higher certainty in the data collection, investigations into more complex transcription templates was investigated. The role that template topology plays in these collision events was investigated through the introduction of topology resolving proteins to assess what effects this might have on collision events.

### **1.2 Synopsis**

Through the use of atomic force microscopy (AFM) this thesis details the study of the outcomes of collisions between two different promoter arrangements, tandem and convergent.

Firstly, a high throughput and efficient method for the labelling of DNA using a single stranded DNA loop was developed, which is described in Chapter 5. It was found that this loop provided a polarity marker for the DNA template. The loop structure appeared not to change as a result of altering the base composition of the loop. Upon formation of open promoter complexes, RNA polymerase (RNAP) displayed a similar affinity for binding to the loop as to its promoter region. It was seen that ~50 % of complexes analysed had a RNAP located at the loop structure. Following initiation of transcription, the RNAP travelled towards the loop where stalling resulted as a consequence of the loop, trapping RNAPs at the end of the template. The terminal loop was then used a polarity marker for experiments using templates with two promoters aligned either convergently or tandemly.

Non-specific binding can obscure those RNAPs that have undergone collisions. It is also possible that multiple rounds of transcription may occur. In order to ensure only those RNAPs that had undergone a single round of transcription were studied an inhibitor of non-specific binding and OPC formation was incorporated into samples and is presented in Chapter 6. It was found that Heparan sulphate (HS) had a number of advantages over the standardly used heparin, with heparin providing only 22 % of complexes with two open promoter complexes (OPCs) formed while HS yielded ~51 % with two OPCs formed. Upon initiation of transcription from both tandem and convergent promoters it was found that collisions led to the stalling of both RNAPs on

## *Chapter 1: Introduction*

the DNA template, with the majority coming to rest in hard contact. For convergent promoters, it was observed that collisions between two elongating complexes (ECs) were most common accounting for 61% of complexes analysed while 31 % of collisions occurred between an EC and an RNAP inactive at the promoter, known as a sitting duck (SD). EC-SD collisions appeared to occur via two distinct mechanisms. One where both RNAPs stalled upstream of both promoters indicating that an EC was able to push a SD upstream. The second displayed RNAPs located either side of the promoter. In the case of tandem transcription it was found that 59 % of collided complexes were located downstream of both promoters while 35 % had the trailing RNAP located between the promoters. Discrimination between EC-EC and EC-SD collisions is not so clear in the tandem case though as both RNAPs are travelling in the same direction. What is apparent is the inability of both RNAPs to transcribe the full template and stall at the loop.

Chapter 7 presents investigations into whether these collisions are altered by a change in template length. Transcription was initiated from a template with longer arms either side of the promoters in the presence or absence of the supercoil resolving protein Topoisomerase IB. This approach tested hypotheses about the effects molecular drag and local template topology for linear DNA templates have on concurrent transcription. It was found that for collisions occurring between convergently transcribing RNAPs stalling occurred at an increased separation of  $\sim 76$  nm as opposed to the RNAPs stalling at a distance of  $\sim 14$  nm (indicative of RNAPs coming into hard contact). This was deduced to be a consequence of transient supercoiling, as after the addition of Topo IB the majority RNAPs stalled at a distance of  $\sim 14$  nm. It was also observed that the ratio of collision types was altered by topology of the DNA with 45 % of collisions being EC-EC collisions and 43 % being EC-SD collisions in the absence of Topo IB for convergent promoters. With Topo IB present, the numbers of EC-SD collisions returned to a value similar to that seen for the shorter template (31 %), indicating that local supercoiling may play a role in promoter escape.

For transcription events from tandem promoters, the effects of topology were most apparent when considering the ratio of the types of collisions seen. Only 20 % of complexes were located downstream of both promoters and 71 % had the trailing

## *Chapter 1: Introduction*

RNAP located between the promoters in the absence of Topo IB. With the addition of Topo IB, those located downstream of both promoters accounted for 54 % of complexes analysed while only 36 % had the trailing RNAP located between the promoters, again, similar to the shorter template case in the absence of any topology resolving enzyme.

The methods devised in this work are applicable to many DNA-protein studies allowing greater certainty and accuracy when using single molecule techniques. The outcomes of convergent and tandem transcription provide evidence that RNA polymerase collision based TI can present a potent block to transcription and gene expression. The importance that topology plays in these collisions is also highlighted and provides a mechanism of sensing and feedback for concurrently transcribing RNAP.

## **1.3 Chapter overview**

### **Chapter 1**

Introduction to thesis subject and the aims of the work contained within the thesis. A synopsis of the main outcomes of the research is also provided with a brief overview of the contents of each chapter.

### **Chapter 2**

This chapter presents an overview of the important components of the transcription process. An overview of the structure of DNA and the role that it plays is discussed. The process of transcription is summarised and the abundance and current understanding of nested and overlapping genes reviewed.

### **Chapter 3**

As AFM is a core technique used throughout this thesis an overview of its basic operation as well as the interactions and forces that occur between the tip and sample is given.

The chapter also discusses the process and considerations of studying DNA protein interactions by AFM. Also discussed are previous studies of transcription using AFM and the importance that these studies have in our current understanding.

#### **Chapter 4**

This chapter introduces the main methods used throughout the thesis, in particular; Polymerase chain reaction (PCR), gel electrophoresis, DNA purification, *in vitro* transcription and AFM sample preparation and imaging. The methods for AFM image processing and data analysis are provided.

#### **Chapter 5**

The first experimental chapter presents the development of a PCR based labelling method. The efficiency and consequence of the sequence of the single stranded label may have on the labelling process and appearance under AFM is also investigated. The effects that these labels have on formation of open promoter complexes (OPCs) and on elongating complexes (ECs) are also presented.

#### **Chapter 6**

This chapter goes on to investigate and develop a method for the incorporation of the inhibitors of non-specific binding, namely, heparin or heparan sulphate (HS). The effects that these molecules have on imaging of DNA, RNAP and transcription complexes is assessed.

Once a method to incorporate HS into samples had been developed, this was applied to the investigation of concurrent transcription from both convergent and tandem promoters. The formation of OPC was investigated and the outcomes of collisions between RNAPs in the presence of HS is presented.

#### **Chapter 7**

The outcomes of concurrent transcription events are studied using a template that allows for constraint of torsional stress. The role that topology and torsional stress may play in concurrent transcription events and collisions is further interrogated by the introduction of the topology altering protein Topoisomerase IB. The outcomes of concurrent transcription from both convergent and tandem promoter arrangements in the absence and presence of Topo IB is presented. The role that topology may play in concurrent transcription events is then discussed.



**Chapter 8**

The final chapter discuss the overall conclusions that can be drawn from the work presented in previous chapters. Future studies and development of the experiments and methods presented in the thesis are also discussed.

This page is left intentionally blank

## 2 Introduction to nucleic acids, DNA and transcription

Presented in this chapter are the structure of deoxyribonucleic and ribonucleic acids (DNA and RNA) and the some of the roles these molecules have in biology. An overview of the constituent parts that make up DNA is given along with the structural properties of this biopolymer. As one of the focusses of this thesis is the study of transcription, specifically concurrent transcription of promoters arranged either in a convergent or tandem orientation an overview of the process of transcription and the relevance of such gene structures may play in biology is presented.

### 2.1 History of DNA

Gregor Mendel performed the first experiments on the transfer of specific traits while investigating plant hybridisation and progeny [7]. He was able to show that characteristics of plants were dependent on the two parent plants, leading to the concept of the heritability of specific characteristics such as color and stem length. Mendel suggested that there was a transferable unit, which was later to be coined by William Bateson as genes [7, 8]. The theory of genes was supported by Thomas-Hunt Morgan who went on to suggest that genes were organised in chromosomes through studies on fruit flies (*Drosophila*), with his student Strutevant further showing that not only were genes organised in chromosomes but occurred in a linear fashion [9, 10]. Even though the presence of genes had become an accepted theory, the material which made up these transferable elements was not elucidated until a later date.

Nucleic acids were first discovered to be present in the cell as long ago as 1868, when Friedrich Miescher purified a substance unsusceptible to protease degradation, containing high levels of phosphorus and no sulphur, indicating the material he purified was not protein. Miescher named this substance nuclein as it originated from the nuclei of cells [11]. Miescher's work paved the way for other scientists to further resolve the constituents of nuclein. Albrecht Kossel determined that five molecules, consisting of a purine or pyrimidine ring, a sugar and a phosphate group, made up what

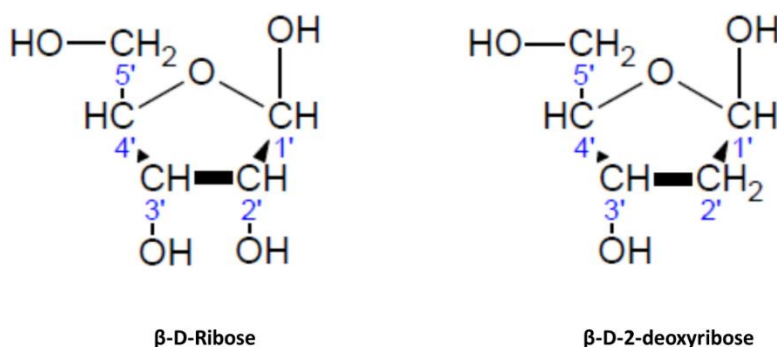
## Chapter 2: DNA and transcription

was to be eventually named as nucleic acid [12, 13]. This led to Kossel being awarded the Nobel prize for this work and his proposal that nuclein was involved in the transfer of information between cells [12]. In 1928 Griffith was able to show that the virulence of bacteria could be transferred from dead cells to live cells through the process of transformation, even though this is one of the earliest experiments to indicate that DNA was the genetic material this fact was not realised [14]. It was not until experiments by Avery *et al.* in 1944, in which it was shown that DNA alone was needed for transformation, coupled with Hershey and Chases later experiments on transduction of DNA from bacteriophage to bacteria, was it confirmed that DNA was the genetic material [15, 16]. Even though nucleic acids were known to scientists it was not until 1953 that the structure was elucidated by Watson and Crick by studying X-ray fibre diffraction patterns obtained by Rosalind Franklin, along with Chargaff's observations from 1952 that all cells had a ratio of pyrimidines to purines of close to one [17, 18]. Once the structure was available it was possible to determine how DNA was able to encode cellular information and also to replicate itself.

## 2.2 Structure of DNA

### 2.2.1 Nucleic acids

Nucleic acids consist of three main parts: a pentose sugar, a phosphate group and one of five bases. The sugar found in RNA is ribose and in DNA is 2'-deoxyribose which lacks the oxygen atom on the 2' carbon (Figure 2-1).

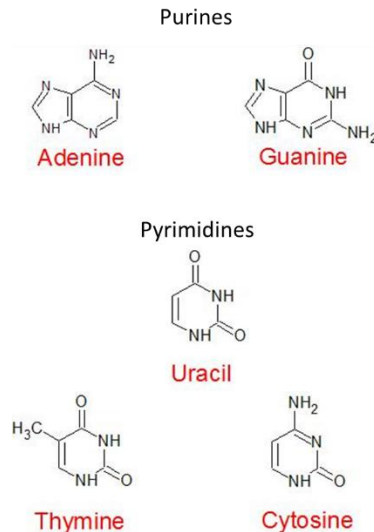


**Figure 2-1: Sugar units found in nucleic acids ribose (RNA) and deoxyribose (DNA).**

**It can be seen that deoxyribose lacks the oxygen atom the carbon 2'.**

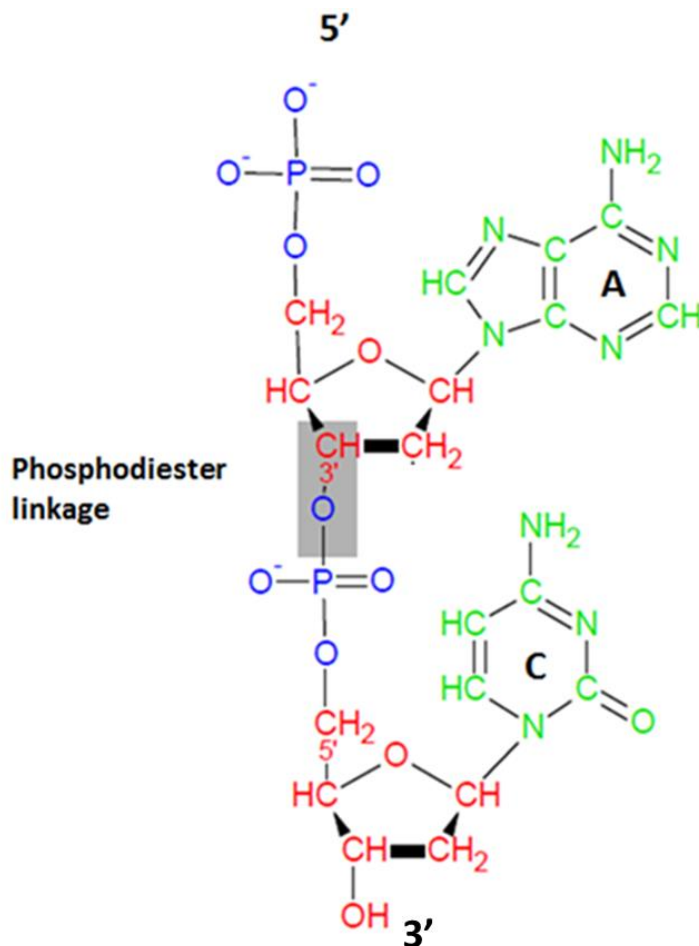
## Chapter 2: DNA and transcription

There are two types of bases found in nucleic acids, purines and pyrimidines (Figure 2-2). Nucleosides are formed when the nitrogen atom 9 of the purine bases adenine or guanine (A and G) or the nitrogen 1 of pyrimidine bases, thymine, cytosine and uracil (T, C and U) are attached to the carbon 1' of the pentose sugar.



**Figure 2-2: Nucleic acid bases that occur in DNA and RNA.**

Phosphate groups attach to the 5' carbon of the sugar to form a nucleotide. Nucleotides may contain 1, 2 or 3 phosphate groups labelled  $\alpha$ ,  $\beta$  or  $\gamma$  respectively. This gives rise to a range of nucleotides named to indicate the number of phosphate groups attached, e.g. adenosine monophosphate (AMP), adenosine diphosphate (ADP) or adenosine triphosphate (ATP). These monomers can be then joined via a phosphodiester bond between the  $\alpha$ -phosphate group on the 5' carbon and the 3' carbon of the sugar ring forming nucleic acid chains. The energy required for the formation of this bond is derived from the  $\beta$  phosphate bond of the NTP. Figure 2-3 highlights how the sequence of nucleotides has a direction resulting from the asymmetry of the terminal ends of the chains, one being a terminal phosphate attached to the 5' carbon, the 5'-end and the other having a terminal hydroxyl group attached to the 3' carbon, the 3'-end.



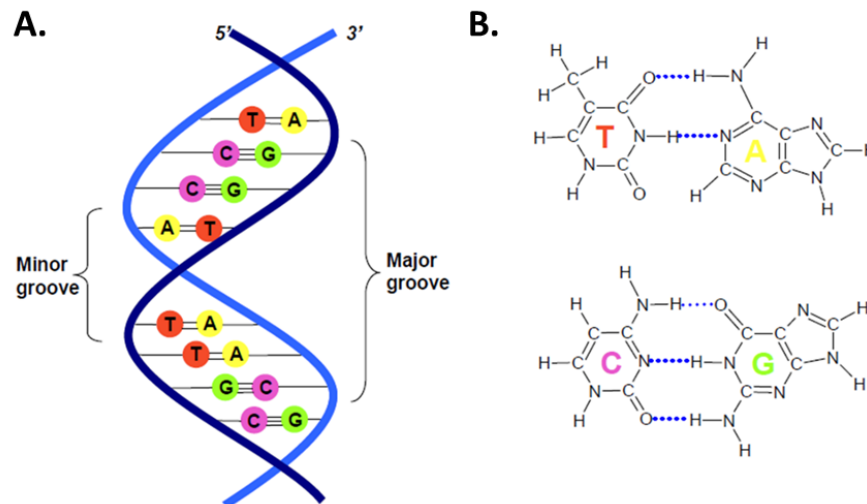
**Figure 2-3: Nucleotides are joined together via a phosphate group linkage to form a chain. The polarity of the chain is denoted by the 5' phosphate and the 3' hydroxyl group.**

### 2.2.2 The DNA double helix

In the structure proposed by Watson and Crick, the DNA adopts a double-stranded right-handed helix, with the phosphate sugar groups lying on the exterior and the bases in a perpendicular orientation in relation to the helix lying relatively flat [17]. The two strands of DNA are arranged in an antiparallel fashion with one chain running in the 5' to 3' (referred to as the sense strand) direction and the other running 3' to 5' (referred to as the anti-sense strand). The bases inside this helix bind to each other through hydrogen bonds. These hydrogen bonds are not random in nature but highly specific with A binding T with two hydrogen bonds and G binding to C with three hydrogen bonds as can be seen in Figure 2-4. This binding is referred to as Watson and Crick base pairing. Due to the extra hydrogen bond between the C and G a higher energy is

## Chapter 2: DNA and transcription

required to disrupt this bonding than A and T, meaning that sequences containing higher amounts of C-G base pairs will require more energy to disrupt the interaction and cause the two strands to denature from each other [19].

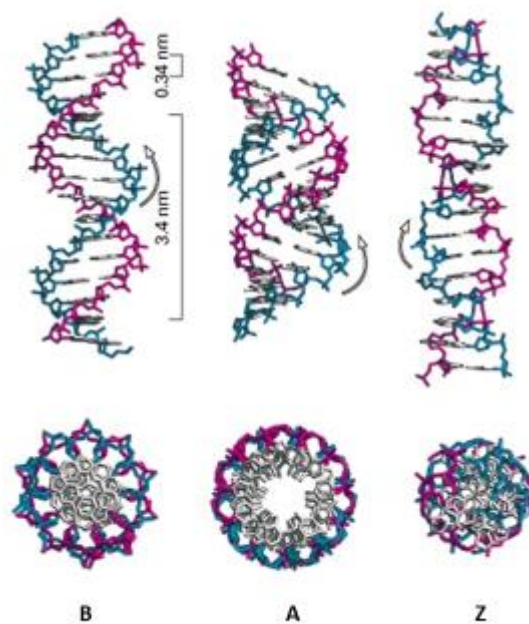


**Figure 2-4: Structure and base pairing of DNA. A) A cartoon of the grooves formed upon base pairing between two antiparallel ssDNA molecules. The DNA strands are represented as blue and light blue curved lines with their polarity denoted from the top of each strand with bases by coloured circles. B) Shown are the hydrogen bonds formed between base pairs with A and T having two hydrogen bonds and C and G three.**

As it can be seen in Figure 2-4 the helical structure contains two distinct grooves. One being wider and is therefore called the major groove and the other, minor groove being narrower. These grooves are important factors for recognition of the DNA by DNA binding proteins [20]. The organisation of the nucleotides mean that the interior of the double helix is largely hydrophobic whereas the exterior backbone with its phosphate groups is hydrophilic. This structure gives the DNA a diameter of approximately 2 nm. The double helix contains approximately 10.4 base pairs per turn of the helix with a base pair separation of 0.34 nm giving a total of 3.6 nm rise per helical turn. This structure proposed by Watson and Crick is named B-form DNA and is the most commonly occurring form of DNA *in vivo*.

## Chapter 2: DNA and transcription

Other forms of DNA exist *in vivo* as well as there being synthetic forms of DNA. The most relevant of these being A-form and Z-form DNA. DNA adopts the A-form when present in lower humidity of around 75% [21]. A-form DNA has a broader cross section of 2.6 nm with the bases being tilted off the helical axis. The major and minor groove also differ from that of B-form DNA with major groove being narrow and deep and the minor groove being wide and shallow. Due to this the grooves are often referred to as the deep and shallow grooves respectively [22]. A-form DNA is also found *in vivo* and resembles the structure of double stranded RNA and may be involved in DNA-RNA hybrids. The DNA-RNA hybrid is unable to form a B-form helix due to the steric clash of the 2'-hydroxyl group of the ribose sugar backbone. The final form of DNA was initially produced *in vitro* in the 1980's but can occur *in vivo*. This form of DNA is Z-form and is adopted when the DNA is made up of a pyrimidine purine repeat sequence at higher salt concentrations or when these sequences are under torsional stress [23]. Z-DNA adopts a left handed helix, opposite to A and B- forms [24]. The helix has a cross-sectional diameter of 1.8 nm with a helical repeat of 12 bp (in comparison to the B-forms 10.4) with a sugar phosphate backbone in a zigzag conformation [25]. A comparison of the three forms of DNA can be seen in Figure 2-5 with the structural features of each form given in Table 2-1.



**Figure 2-5: The three main conformations of DNA. B and A DNA adopt a right handed helix with B having smaller diameter and larger helical pitch. Z DNA adopts**



a left handed helix with a smaller diameter than both B and A forms and a different arrangement of the nucleotide bases as seen by the top down view [26].

DNA Form	Helix handedness	Bp/Turn	Helix Diameter (nm)	Major Groove	Minor Groove
B	Right	10.5	2.0	Wide and deep	Narrow and deep
A	Right	11.0	2.6	Narrow and deep	Wide and shallow
Z	Left	12.0	1.8	Flat	Narrow and deep


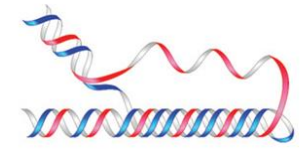
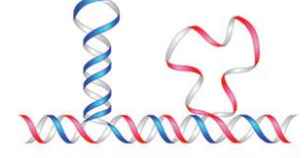
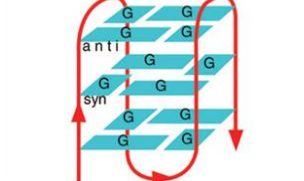
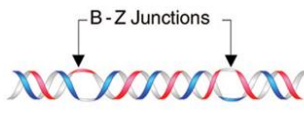
**Table 2-1: Structural features of the main forms of DNA [27].**

### 2.2.3 RNA

There are three main structural differences of RNA from DNA: the sugar being a ribose (rather than deoxy-ribose); Thymine is replaced by uracil in RNA; and typically RNA does not adopt a regular double helical structure. The presence of the hydroxyl group on the ribose sugar means that the RNA is not able to adopt the B-form helical structure of DNA. The hydroxyl group interferes with the conformations of the sugars. RNA is able to form an A-form double helix and is able to base pair, often folding upon itself to form complex secondary structures. This ability to form complex structures means that RNA is able to display some enzymatic activity as is seen in enzymes known as ribozymes such the ribosome [28].

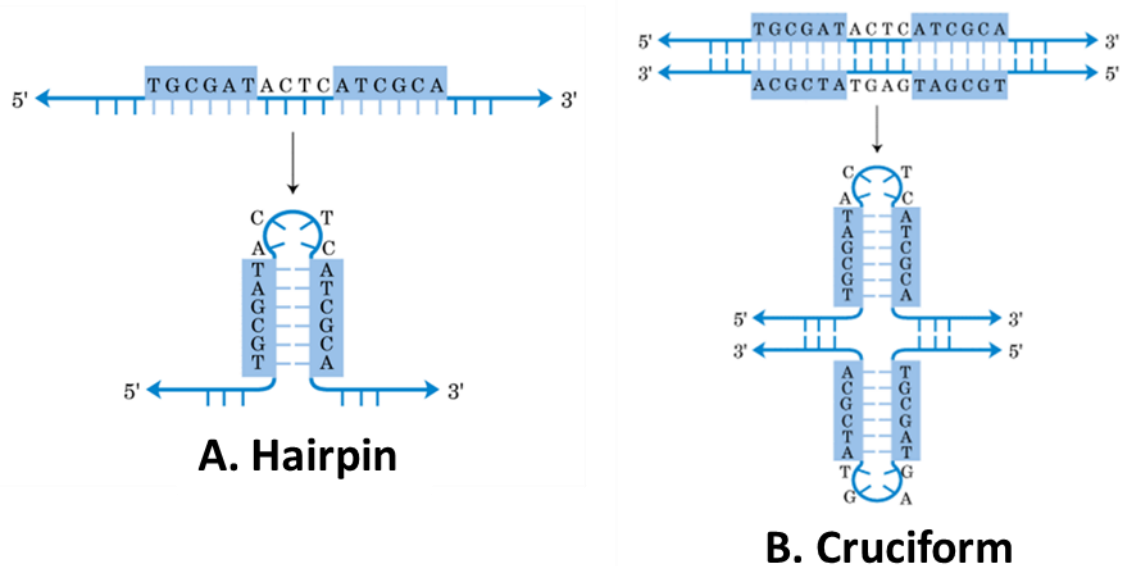
### 2.2.4 DNA secondary structures and topology

DNA is also able to adopt secondary structures. These can take a number of forms with a summary of common structures shown in Figure 2-6 [29].

Name	Conformation	General sequence requirements	Sequence
Cruciform		Inverted repeats	<pre> TCGGTACCGA AGCCATGGCT </pre>
Triplex		$(R \cdot Y)_n$ mirror repeats	<pre> AAGAGGGGAGAA TTCTCCCTCTT </pre>
Slipped (hairpin) structure		Direct repeats	<pre> TCGGTTCGGT AGCCAAGCCA </pre>
Tetraplex		Oligo (G) <sub>n</sub> tracts	<p>AG<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub> Single strand</p>
Left-handed Z-DNA		$(YR \cdot YR)_n$	<pre> CGCGTGCGTGTG GCGCACGCACAC </pre>

**Figure 2-6: Non-B DNA structures. Examples of the main types of non-B DNA structures seen *in vivo* are shown. Each example gives a schematic representation of the geometry of each structure along with the type of sequence which forms the structure and example sequences in the right hand column [30].**

Hairpin loops and cruciforms arise when an inverted repeat occurs in one strand of the double helix (hairpin) or both strands of the double helix (cruciform) leading to intrachain base pairing and extrusion of a single stranded loop from the DNA helix [31]. These structures have been seen to be able to form inverted repeats as short as 7 bp *in vivo* [32].



**Figure 2-7: Examples of hairpin and cruciform structures formed by DNA. A) Formation of a hairpin loop through intra-chain bonding at an inverted repeat highlighted in blue. B) Formation of a cruciform at an inverted repeat in both strands of a dsDNA sequence.**

The occurrence of these hairpin and cruciform forming sequences are more common in eukaryotic genomes than in prokaryotes but it has been shown that they an important role in both prokaryotic and eukaryotic DNA replication [33, 34]. Triplex structures form when the DNA contains mirrored stretches of homopurine: homopyrimidine sequence [31]. This allows for a single stranded polynucleotide sequence to bind to the major groove of the underlying DNA to form a triple stranded structure [35]. These sequences are again more abundant in eukaryotes than in prokaryotes and are often found near promoter regions of genes [36]. Slipped strand structures form when direct repeats are present and complementary strand base pairing is misaligned leading to hairpin formation or looping out of bases with this occurring during replication of DNA [34, 37, 38]. Tetraplex DNA or G-quadruplexes are formed by long stretches of G bases which are able to act as donors and acceptors in what is referred to as Hoogsteen base pairing [31]. G-quadruplexes have been found to occur at a number of different regions of eukaryotic genomes, most notably within telomeric DNA. Sequences that could form G-quadruplexes have also been found in the immunoglobulin switch region and at promoter start sites in both eukaryotes and

## Chapter 2: DNA and transcription

prokaryotes [39]. Sequences of repeating alternating purines and pyrimidines have been shown to be able to adopt the left handed zigzag helical structure of Z-DNA [31]. The crystal structure of a B to Z DNA junction was obtained in 2005 and showed extruded base pairs on each side of the double helix, with these bases believed to be more susceptible to modifications [40].

All of these structures have been recorded to occur *in vivo* and have implications in areas of DNA metabolism such as transcription, replication and recombination events. The process by which these structures alter DNA metabolism is not fully understood but in many cases these structures are involved in genome instability and diseases which have been reviewed [30, 31, 34, 38]. All these structures require some form of energy input in order to form. This energy input is often derived from changes in the topology of DNA, being most commonly formed and maintained by underwinding of the DNA helix or negative supercoiling [41].

When DNA undergoes processes of replication, transcription or any form of protein or molecule binding, the geometry and topology of the DNA is altered, especially when the two strands are separated. If we consider the DNA as two braided ropes, if these ropes are pulled apart this leads to an increase in twisting of the rope and at some point the rope will coil upon itself, as is shown in Figure 2-8.

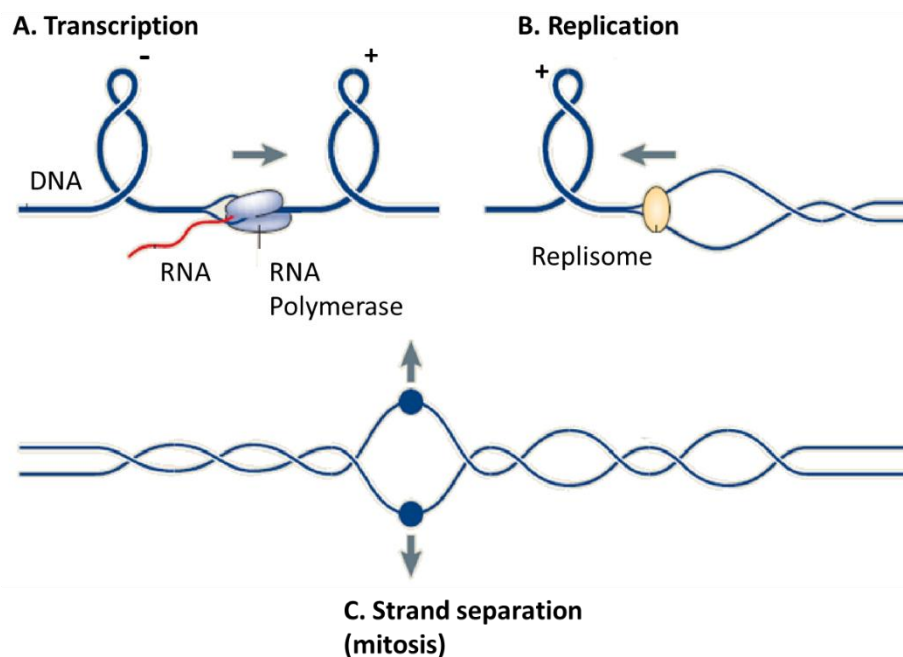


**Figure 2-8: Representation of the changes of DNA when strands are separated. As the two strands separate (left side of image) the double helical chain is under increased tension, leading to the formation of overwound structures or supercoils [42].**

These structures that form due to the rope coiling upon itself are referred to as plectonemes or toroids and are a direct result of supercoiling. Supercoiling is the

Chapter 2: DNA and transcription

higher order coiling of the DNA helix which can occur during processes where the DNA must be unwound in order to access the sequence of a single strand such as with transcription and replication, as is seen in Figure 2-9. Supercoiling also plays a role in storage of DNA in both eukaryotes and bacteria, with DNA often being supercoiled and organised into high order structures to reduce the amount of space a genome or plasmid may occupy as well as protect from degradation or help improve its transport [43, 44].



**Figure 2-9: The introduction of supercoiling by cellular processes. A) In transcription, the DNA is melted in order for the RNA polymerase to gain access to a single strand leading to over-winding in front and under-winding behind the protein. B) In replication the two strands must be separated to allow a new daughter strand to be produced, this leads to occurrence of over-winding [26].**

In Figure 2-9 structures are labelled positive and negative, indicating the type of supercoiling that is present. Positive supercoils can be simplified to an over-winding of the DNA helix and negative supercoils are an under-winding of the DNA helix. If we assume that the DNA is linear but tethered at both ends, as in nature by a protein or the chromosome itself, the DNA is in a sense closed, or if the DNA is actually a closed circle like a plasmid then this can be explained mathematically by the equation

Chapter 2: DNA and transcription

$$\Delta Lk = Lk - Lk^\circ$$

Where  $\Delta Lk$  is linking difference or change in linking number. The linking number  $Lk$  is defined as the number of times that the two strands in the helix cross over each other in a closed DNA molecule and is constrained to an integer in a closed DNA molecule. This number is unaffected by changes in the conformation of the DNA as long as both strands remain intact.  $Lk^\circ$  is the standard linking number which is the actual number of helical turns of the DNA molecule in its relaxed state and is given by the equation

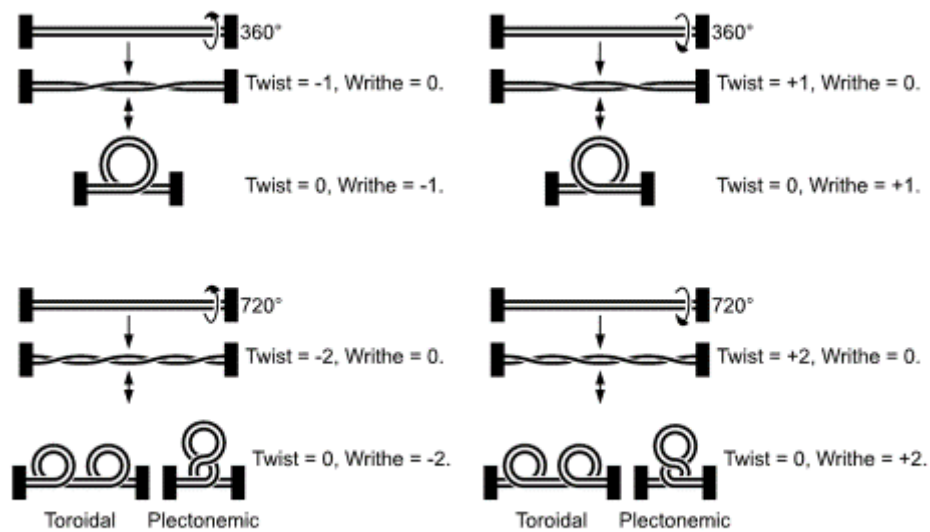
$$Lk^\circ = \frac{N}{h}$$

Where  $N$  is the number of base pairs in said molecule and  $h$  is the number of base pairs per turn of the helix which for standard DNA is 10.4-10.5.  $Lk^\circ$  therefore serves as a reference point to measure the level of supercoiling. The linking number of closed circular B-DNA is defined as positive, and so a further twisting of the helix before it is closed leads to a positive linking difference *ergo* positive supercoiling. If the helix is unwound before closure then the linking difference is negative *ergo* negative supercoiling.

The geometric consequences of supercoiling can be explained by the equation

$$Lk = Tw + Wr$$

$Tw$  is the twist and describes how the strands of the DNA coil around each other about the axis of the DNA helix.  $Wr$  is the writhe and is a measure of the coiling of the helix axis in space. Due to the fact that the linking number is invariant then changes in the twist of the DNA must be compensated for by an equal but opposite change in the writhe, and vice versa. This concept is shown by the diagram in Figure 2-10.



**Figure 2-10: Diagram depicting the link between changes in twist ( $Tw$ ) and writhe ( $Wr$ ). It can be seen that changes in twist of the DNA leads to an opposite and equal change in writhe which is accomplished by the DNA adopting different structures [45].**

Even though the linking number, twist and writhe are not easily calculated, an understanding of the structures formed due to supercoiling is important, as supercoiling can change the conformation of the DNA as well as leading to the stabilisation of structures previously discussed at the start of this chapter.

### 2.2.5 Supercoiling in the cell

Globally in both eukaryotes and prokaryotes the genomes are maintained in an underwound state, or negatively supercoiled [6, 46]. In prokaryotes this negative supercoiled state is maintained by wrapping of DNA around proteins such as HU and H-NS which constrain negative supercoils, but this only accounts for 50-60% of the supercoiling seen, whereas in eukaryotic genomes the writhing of the DNA around histones, which constrains negative supercoils, accounts for nearly all of the negative supercoiling [47, 48]. Prokaryotes contain proteins not found in eukaryotes, known as DNA gyrases which are able to introduce negative supercoils as well as topoisomerases

## Chapter 2: DNA and transcription

which are proteins capable of removing negative supercoils. These proteins are important in maintaining the unconstrained supercoiling seen bacteria [49]. Eukaryotes also have topoisomerases that are able to reduce negative supercoiling as well as positive supercoiling and these are often closely linked to DNA metabolism [50].

The genome of prokaryotes is organised into domains as shown by experiments by Sniden *et al.* who introduced single stranded nicks into the DNA, but found that the DNA maintained its negative supercoiling [51]. These domains were further defined by Postow *et al.* who was able to show that these domains were approximately 10 Kbp in size as well as being dynamic and stochastic in nature [52]. Domain structures have also been noted in eukaryotes, with domains often being much larger (10 Mbp) [53, 54]. The maintenance of these domains is suggested to be by a number of mechanisms and proteins which form a topological barrier, thereby not allowing supercoiling to spread through the whole genome. These topological barriers can be formed by insertion into the membrane of DNA bound proteins, chromosome to chromosome interactions, by nucleoid proteins bound to the DNA and by the binding and actions of proteins during processes such as transcription, replication or recombination [52, 55, 56]. Maintaining a global negative supercoiled state can help promote processes such as transcription and replication by aiding in the melting of the DNA duplex as well as in storage of a chromosome.

Supercoiling also plays an important role locally as well. Changes in local supercoiling of DNA allows for the binding of proteins or the formation of non-B DNA structures that can regulate cellular processes as well as also stabilising DNA transactions [6]. Supercoiling has been suggested as being the highest hierarchal level of gene expression control [47]. As many of these processes lead to changes in supercoiling themselves there exists a complex auto-regulatory system. For example the transcription of DNA by RNA Polymerase (RNAP) leads to the formation of a positively supercoiled domain downstream of the protein and a negative supercoiled domain upstream of the protein as presented by the twin supercoil domain theory of Lui and Wang [57]. The supercoiling linked to transcription is extensive with approximately 1 supercoil being introduced for every 10 bp transcribed [6]. This is addressed by topoisomerases in the cell in order to prevent a large build-up of torsional stress which

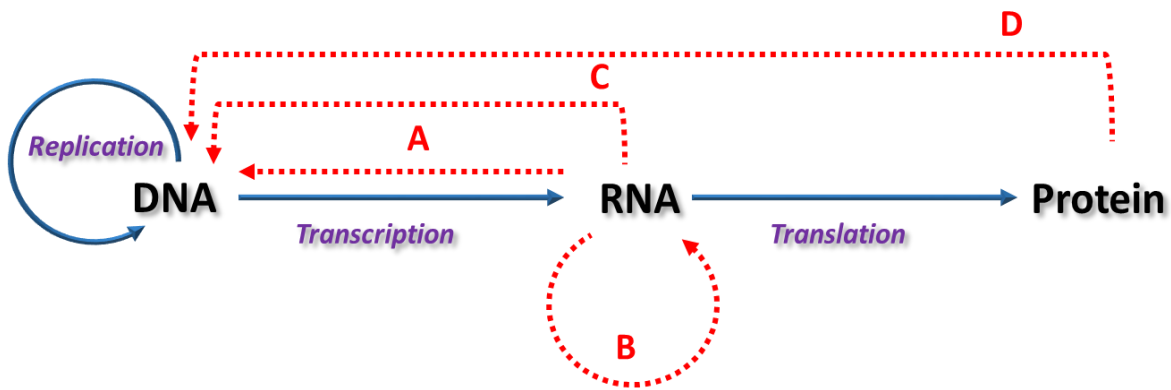


## Chapter 2: DNA and transcription

can have an inhibitory effect on transcription but evidence shows that supercoiling is not fully resolved [58-60]. The changes in supercoiling induced by transcription can lead to the formation of non-B DNA structures, promote further rounds of transcription as well as promote or destabilise binding of proteins like the nucleosome or nucleosome like structures [61, 62]. The importance of supercoiling on gene expression is confirmed by experimental work such as that carried out by Joshi *et al.* who showed that there was a reduction of ~80% of gene expression when high levels of supercoiling were present as well by Geertz *et al.* who showed that supercoiling density affected the expression profile of a cell [63, 64]. The role that supercoiling plays in cells is still being discerned but there are a number of reviews which provide insights into the possible roles that supercoiling plays in cellular processes [6, 55, 61, 65, 66].

### 2.3 Central dogma of molecular biology

*In vivo* DNA acts as a storage device for information required for the functioning of an organism. It is copied in a process called replication so that the information can be transferred to future generations. The information contained in the DNA, provides a code for production of RNA through the process of transcription, carried out by RNAPs. The RNA produced is a direct copy of the coding strand (complementary copy of the template strand) but with U in place of T. The RNA transcribed can be a non-coding RNA (ncRNA) such as transfer RNAs (tRNA) and ribosomal RNA (rRNA). Other non-coding RNAs include Small nucleolar RNAs (snoRNAs), microRNA (miRNAs), Small interfering RNAs (siRNAs), extracellular RNAs (exRNAs), piwi-interacting RNAs (piRNAs) and Small nuclear ribonucleic acids (snRNAs) [67]. Coding RNA or messenger RNA (mRNA) provides the template for the production of a protein, by the process of translation. This flow of information from DNA to RNA to protein was first proposed by Crick in what was described as the “Central Dogma” of molecular biology (Figure 2-11) [68]. As our understanding of the processes that occur in cells and organisms has increased it has been found that this flow of information is much more complex than previously thought, with metabolic processes, DNA structure, RNA and the replication of viruses indicating that DNA, RNA and proteins can play a number of different roles outside the central dogma [69, 70].

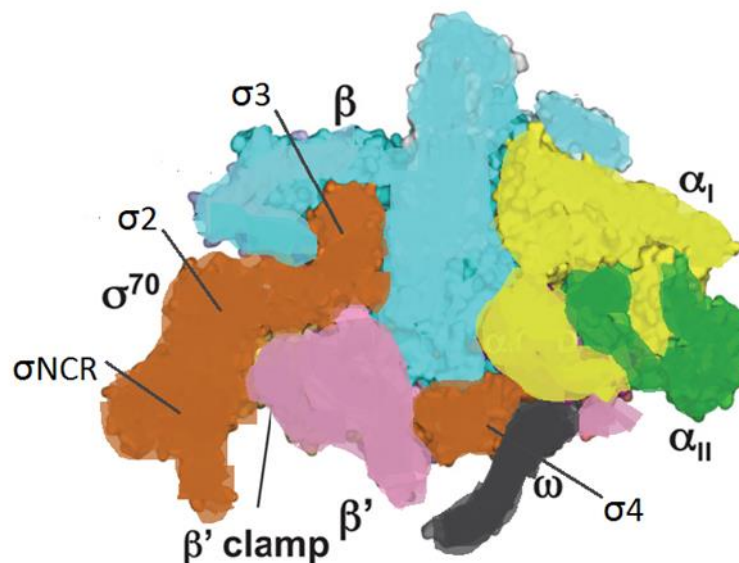


**Figure 2-11: The central dogma of molecular biology. The main pathway is shown by the blue arrows with the flow from DNA to RNA to protein. Also shown by the red dashed arrows are subsequent additions to the central dogma as our understanding has progressed. A accounts for reverse transcription seen in viruses, B is the ability of some viruses to replicate their RNA based genomes, C represents the input of RNA feedback through such mechanisms as small interfering RNAs and riboswitches and D is the feedback provided by proteins and metabolism.**

### 2.3.1 RNA polymerase structure

Transcription is a first step in the process of gene expression and is defined as the DNA templated production of RNA. In order to understand this process, knowledge of the structure of the main component, RNAP is important. The RNAP protein is found in some viruses, mitochondria, bacteria, archaea and eukaryotic cells and was first described in 1960 [71]. In the case of mitochondria and viruses such as bacteriophages, RNAP is a single subunit protein [72, 73]. In other organisms RNAP is made up of multiple subunits. Eukaryotes contain three RNAPs each with a specific role in the cell, these proteins can range from 12-17 subunits whereas archaea and bacteria only contain a single RNAP [74]. The archaeal RNAP has 14 subunits whereas the bacterial protein only has 5 [75]. The simplistic nature of the bacterial RNAP has meant that it is the most studied and best understood. The *E.coli* RNAP often serving as a simplified model of more complex RNAPs as it has been shown that all multi-subunit RNAPs consist of a conserved core region that contains the active site and major domains and show a high degree of sequence, structural and functional homology [75-79]. For this reason *E.coli* RNAP is used as a model system in this work and the process by which it performs transcription is described.

*E.coli* RNAP consists of 5 units,  $\beta$ ,  $\beta'$ ,  $\omega$  and a dimer of  $\alpha$  with a combined molecular weight of 340 kDa [80]. Bacterial RNAP requires a secondary protein in order to specifically bind to its promoter. This protein is known as sigma factor. In bacteria there are seven sigma presents [81]. These include  $\sigma^{70}$ ,  $\sigma^{19}$ ,  $\sigma^{24}$ ,  $\sigma^{28}$ ,  $\sigma^{32}$ ,  $\sigma^{38}$  and  $\sigma^{54}$ . Each acts in response to different stimuli such as stress response and heat shock. All  $\sigma$ -factors have conserved domains and activity but  $\sigma^{54}$  differs in that it is able to bind its DNA target dependant of RNAP core enzyme and also requires an activator (bacterial enhancer binding proteins) in order to initiate transcription [82]. The most abundant is  $\sigma^{70}$  which is considered the housekeeping sigma factor with the superscript number denoting its molecular weight [83].  $\sigma^{70}$  is composed of 4 domains (1-4) with each domain being broken down into regions. Domains 2-4 bind to promoter elements while domain 1 is located in the active site of the RNAP [84]. The RNAP and sigma factor combined are known as the holoenzyme ( $\sigma$ RNAP) and have a combined molecular weight of ~420 kDa.

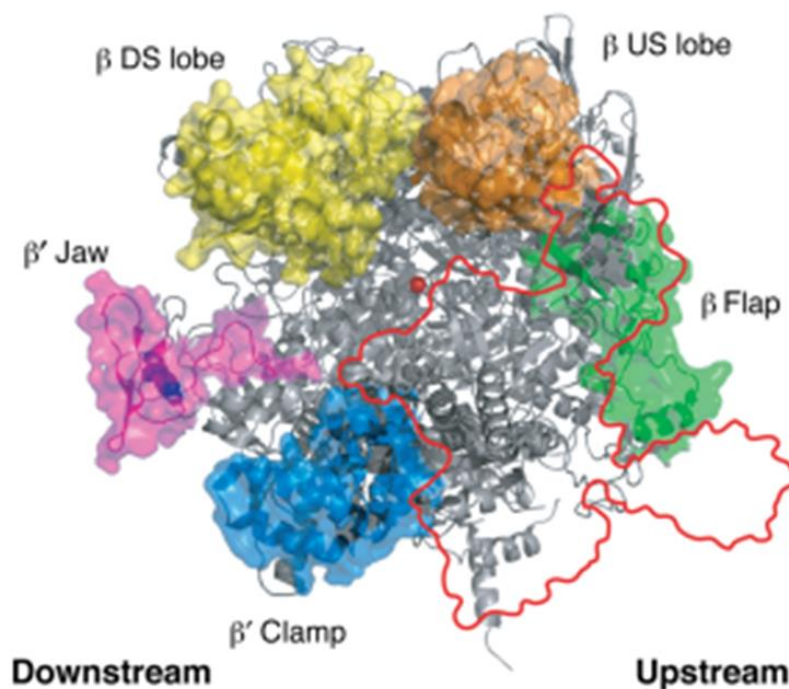


**Figure 2-12: Structure of bacterial RNAP holoenzyme. The specific subunits have been given separate colours and labels and the  $\beta'$  clamp has been labelled. Orange is  $\sigma^{70}$ , yellow is  $\alpha_i$ , green is  $\alpha_{ii}$ , cyan is  $\beta$ , pink is  $\beta'$  and grey is  $\omega$ . (adapted from [85])**

The first crystal structures reported by Zhang *et al.* were of *Thermus aquaticus* (*Taq*)  $\sigma^{70}$ RNAP which agreed well with electron crystallography data collected by Darst *et al.* with this being further confirmed in 2013 by Murakani who reported the structure of

## Chapter 2: DNA and transcription

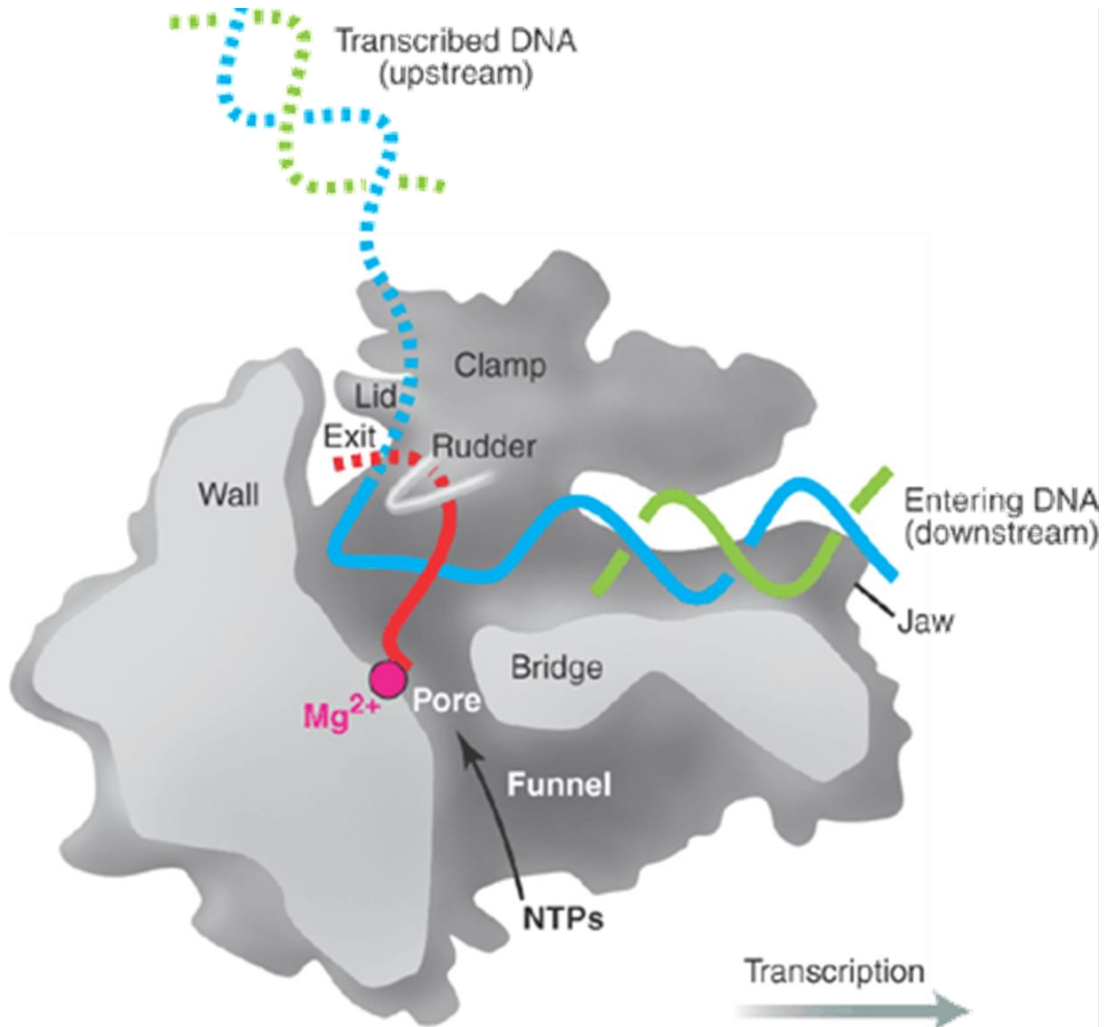
*E. coli*  $\sigma^{70}$  RNAP at a resolution of 3.6 Å using X-ray crystallography (Figure 2-13) [85-87]. The core enzyme has a number of key elements with the overall structure having the appearance of a crab claw. One pincer of the claw is formed by the  $\beta'$  unit and the other claw being from part of the  $\beta$  unit. Together they form a channel that runs the length of the protein that has a diameter of  $\sim 27$  Å [80]. There is also a secondary channel that allows for the entry of NTPs. The  $\alpha$  units are found at the  $\beta$  and  $\beta'$  interface and are involved in assembly of the  $\beta$  and  $\beta'$  units as well as interacting with elements of the DNA and activator proteins [88, 89]. The  $\omega$  unit associates with  $\beta'$  subunit and acts as a chaperone in its recruitment to the enzyme assembly [90]. The channel is surrounded by five domains: the  $\beta'$ -clamp,  $\beta'$ -jaw,  $\beta$ -flap and the up and downstream  $\beta$ -lobes (Figure 2-14) [91]. These domains can move independently and regulate the opening and closing of the main channel and loading of DNA into the active site [92, 93].



**Figure 2-13: Crystal structure of bacterial RNAP with key elements highlighted, the  $\sigma^{70}$  is removed but outlined in red. The active site  $Mg^{2+}$  is represented by the red sphere, the  $\beta$  downstream (DS) lobe is shown in yellow, the  $\beta$  upstream (US) lobe orange, the  $\beta$  flap in green, the  $\beta'$  clamp in blue and the  $\beta'$  jaw in pink (adapted from [91])**

*Chapter 2: DNA and transcription*

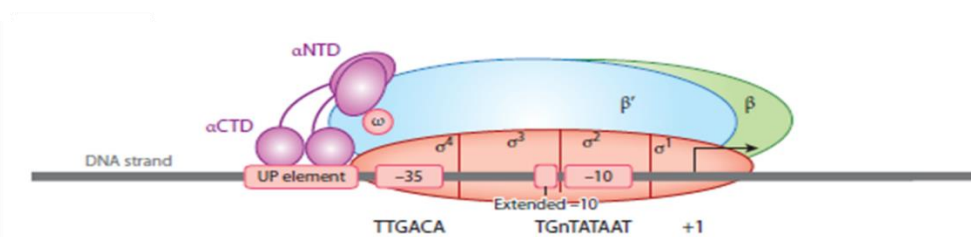
At the upstream region is the exit for the nascent RNA chain which is prevented from forming a long hybrid with the DNA by the lid. The active site of RNAP contains a rudder region which is involved in maintaining the separation between the template and coding strands. The bridge helix and the trigger loop which are involved in translocation are found opposite a  $Mg^{2+}$  ion, which is involved in the catalytic activity of the protein (Figure 2-13 and -14).



**Figure 2-14: Cut through cartoon of bacterial RNAP. The protein is shown in grey with the channel and important features labelled. The active site  $Mg^{2+}$  is shown in pink with a nascent RNA transcript attached by its 3' end (red). The template strand is shown in blue while the coding strand is shown in green. The rudder element is shown in white [94].**

### 2.3.2 Process of transcription

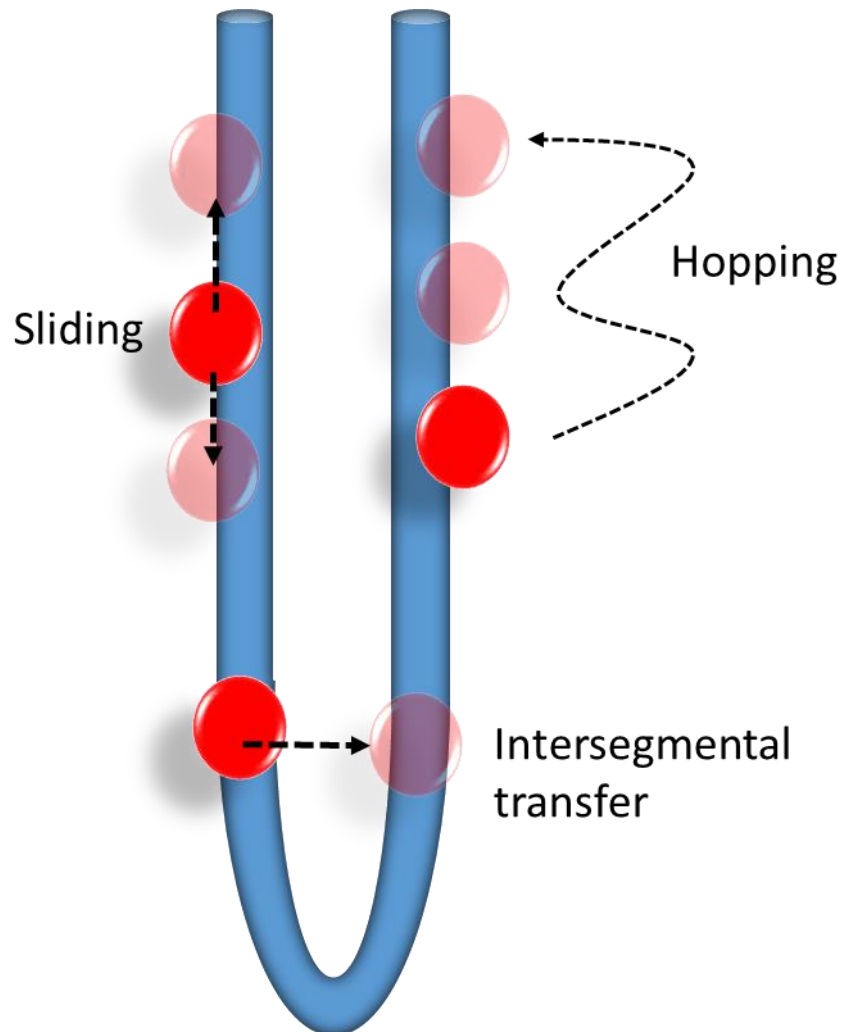
The process of transcription can be broken down into three main stages: initiation, elongation and termination. The first step in the initiation stage is the locating of a sequence known as the promoter by the  $\sigma$ RNAP. This is aided by the presence of the sigma factor which as is able to bind to specific DNA sequences. Promoters lie upstream of the transcriptional start site and are defined by the presence of certain conserved elements. In bacteria there are two main elements, the Pribnow box which has a conserved AT rich nature normally having the consensus sequence TATAAT on the template strand at a position 10 base pairs (bp) upstream of the transcription start site (-10 position). This region may have an extended AT rich sequence that can enhance promoter activity [95]. In eukaryotes this region is reflected by the TATA box found at -30 [96, 97]. The second conserved element is the -35 sequence which has consensus sequence TTGACA [98]. It is noted though that the majority of promoters only conserve three to four of the consensus nucleotides *in vivo*. These two regions are recognised and bound by the sigma factor regions, 2 for the -10 element and region 4 binding the -35 element (Figure 2- 15) and have an optimal spacing of  $17 \pm 1$  bps. It was shown by Ross *et al.* that promoters can also contain further recognition regions referred to as the UP elements, which interact with the  $\alpha$  subunits of the RNAP (Figure 2-15) [99-102].



**Figure 2-15: RNAP holoenzyme bound to a promoter DNA sequence. The  $\beta$  and  $\beta'$  subunits are shown in green and blue with the  $\sigma^{70}$  factor in red with each region designated ( $\sigma$  1, 2, 3, 4) to DNA (grey). The  $\sigma^{70}$  regions are shown to provide a visualisation of binding to the promoter elements (red boxes). The  $\alpha$  subunits are shown in purple, with the N-terminal domain (NTD) which binds the  $\beta$  and  $\beta'$  units highlighted and the C-terminal domain (CTD) shown binding to UP elements as suggested by Ross *et al.* [99].**

*Chapter 2: DNA and transcription*

For promoter binding to occur the RNAP holoenzyme must first locate the promoter. The mechanism by which this occurs is debated [103]. As the promoter is only a miniscule fraction of the total DNA, locating the promoter would take a prohibitive amount of time if it was to occur by three dimensional (3D) diffusion and so a process of facilitated diffusion has been suggested to explain the increased binding rate [104]. This has been predicted to occur by three main mechanisms: sliding, hopping and inter-segmental transfer (Figure 2-16).

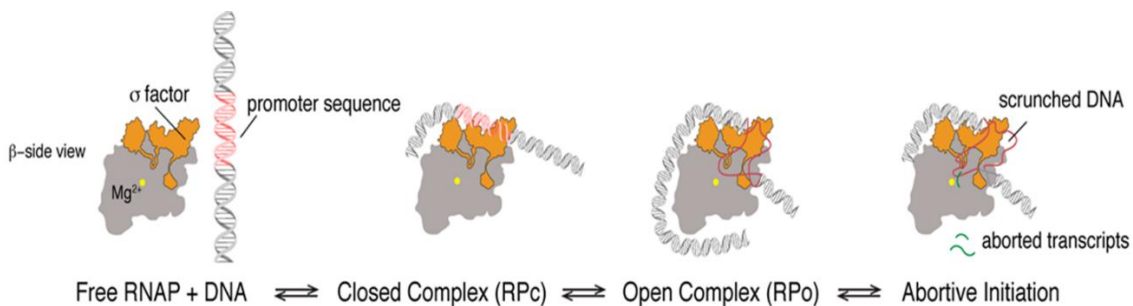


**Figure 2-16: Mechanisms of promoter location by RNAP. The RNAP is predicted to utilise three mechanisms to help locate its promoter: Sliding which involves the RNAP making non-specific contacts with the DNA and scanning; inter-segmental transfer which takes advantage of the flexible nature of DNA to allow the RNAP to transfer between two distant points through bends in the chain; hopping which involves short transient contacts of the RNAP with DNA before disassociating and rebinding at a new point.**



Sliding involves the 1-dimensional diffusion of the RNAP holoenzyme along the DNA and has been observed with AFM by Bustamante *et al.* and Endo *et al.* as well as by TIRFm by Harada *et al.* and by other single molecule methods [105-109]. There is a possibility that the  $\sigma$ RNAP is also able to transfer between positions due to looping of the DNA chain bringing further regions closer, this is known as inter-segmental transfer. The final mechanism is that of hopping, where the RNAP is able to make transient contacts with DNA until its promoter is found, it should be noted that neither inter-segmental transfer or hopping is yet to be confirmed for RNAP [110]. In recent years the observations of 1D diffusion has been regarded as more of an artefact of the buffer composition, reaction setup or concentration of protein [103]. It has been proposed that  $\sigma$ RNAP locates its promoter by 3D diffusion *in vivo* where protein concentration is high [103, 111, 112].  $\sigma$ RNAP would therefore perform small 1D diffusion events that do not dominate the promoter search.

Once the promoter has been located, the  $\sigma$ RNAP binds and forms a closed promoter complex (CPC) where the DNA is still double stranded and not yet in the active site [113]. The RNAP holoenzyme then undergoes a process called isomerisation summarised by Figure 2-17 [114].



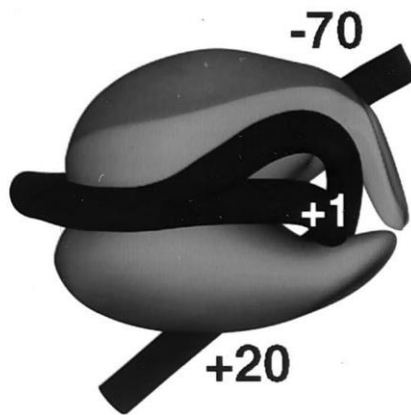
**Figure 2-17: Process of isomerisation. The RNAP holoenzyme binds the promoter. A CPC is then formed and the DNA is slightly bent, this then develops into an OPC by melting the DNA to form a transcription bubble and wrapping of the DNA around the RNAP. The RNAP then undergoes abortive transcription and scrunching of the DNA until it is able to escape it's promoter [114].**

Isomerisation involves the opening of  $\sim 13$  bp of the DNA helix around the -10 element to form a transcription bubble. The non-template strand is held by the groove formed



## Chapter 2: DNA and transcription

by the  $\beta$  subunit and the sigma factor while the template strand is threaded through the active site [115, 116]. This opening of the transcription bubble leads to the displacement of the sigma factor region 1 from the active site, as was shown by fluorescence/Förster resonance energy transfer (FRET) measurements by Mekler *et al.* [117]. The isomerisation is not complete till the complex is stabilised into an open promoter complex (OPC) by the binding of the jaw and clamp regions to the downstream DNA [118]. Upon formation of an OPC, the DNA is wrapped around the protein, leading to a bend in the DNA, this was first suggested by Dnase I footprinting assays that showed that a larger region than expected, of  $\sim 90$  bp, was protected against enzymatic degradation [119-123]. This theory was confirmed by atomic force microscopy (AFM) and magnetic tweezers measurements that were able to show that compaction of the DNA occurred upon OPC formation of  $\sim 30$  nm and that the DNA displayed a distinct bend as is shown by Figure 2-18 indicating that the DNA is wrapped through  $\sim 300^\circ$  [124-129]. This wrapping was further confirmed by FRET experiments by Sreenivasan *et al.* [130].



**Figure 2-18: 3D model of the wrapping of DNA around RNAP holoenzyme. The transcription start site is denoted as +1 and can be seen sitting at the active site. The DNA is fully wrapped around the protein and the positions of the edge of contact between the DNA and RNAP are shown [125].**

Once a stable OPC has been formed the  $\sigma$ RNAP undergoes a process of abortive initiation which involves the production of small RNA products of 2-11 nucleotides (nt) in length, but the  $\sigma$ RNAP does not escape its promoter and enter elongation. This process was believed to occur only on occasion but was shown by Hsu *et al.* to be a

## Chapter 2: DNA and transcription

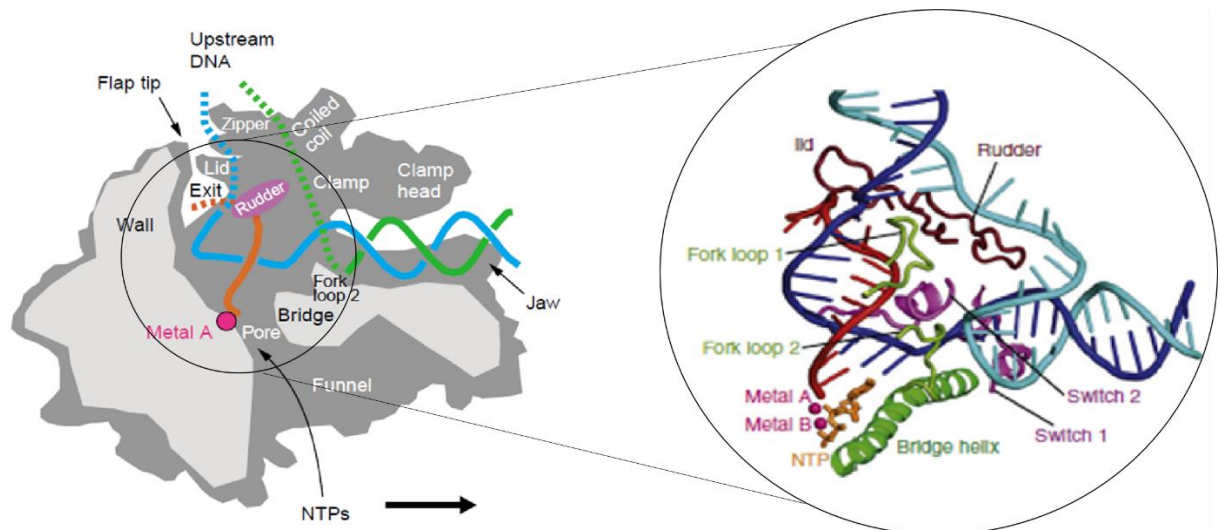
common occurrence at a number of promoters and the percentage of abortive cycles was dependent on the promoter [131]. The NTPs enter into the OPC and a small transcript is produced, while the leading edge and active site appeared to move downstream but the trailing edge maintained its contacts. This was first believed to occur either by inch-worming or transient excursions but both Kapandis and Revyankin *et al.* in 2006 were able to show by the use of FRET that the RNAP did not move but rather drew DNA into the active site, therefore expanding the size of the transcription bubble by what was termed scrunching (Figure 2-17 final panel) [132, 133]. This was further confirmed by Robb *et al.* also using FRET who showed that this scrunching mechanism could be the cause of heterogeneity seen in transcriptional starts points [116]. This scrunching leads to more DNA being held in the active site due to unwinding of the DNA helix. The unwinding and compaction lead to stress, which causes an increase in potential energy within the active site of the RNAP. This potential energy may be the driving force for RNA release and abortive initiation or for promoter escape and productive initiation [134].

The second stage of the transcription process is elongation. The RNAP escapes its promoter and translocates along the DNA unwinding the helix downstream, forming a 8-9 bp DNA-RNA hybrid at the active site [135]. It is believed that the sigma factor may be released as the RNA chain reaches 12-15 nt and the hybrid reaches 8-9bp [135]. There is some evidence that this may not be the case as Kapanadis *et al.* showed by single molecule studies that the sigma factor may be retained in up to 90% of initial elongating RNAPs and 60% of mature RNAPs. They were also able to show that this release or retention was not RNA length dependent, leaving the question of how and why the sigma factor may dissociate [136, 137]. The retention of the sigma factor has also been noted by other groups, and its retention has been implicated in pausing proximal to the promoter, especially at sites on the non-template strand that resemble the -10 element and or are AT rich [137-140]. The release of the sigma factor leads to its auto inhibition which is achieved by folding itself so it cannot bind DNA until it is bound to a new RNAP [141]. If the sigma factor is released it can then be recycled so that subsequent rounds of initiation can occur [135, 142, 143]. Within the cell RNAP is found in excess, with amounts of RNAP increasing in response to stimuli such as mitosis, stress and starvation by way of example [144]. Even though free RNAP is

## Chapter 2: DNA and transcription

considered to be in relative excess, sigma factors are found to be in competition with each other for binding to core enzyme [145]. This means that if the activity or abundance of one sigma factor increases this has a knock on effect of transcription of genes which are bound by other sigma factors [145].

The RNAP maintains the transcription bubble by using positively charged residues located in the switch 2 region and negative residues in the switch 1 region to pull apart the downstream DNA strands: re-annealing is prevented by the rudder [94]. The hybrid length is maintained by fork loop 1, and the lid region located at the RNA chain exit site prevents the nascent RNA from re-associating with the DNA (Figure 2-19) [94, 146, 147].

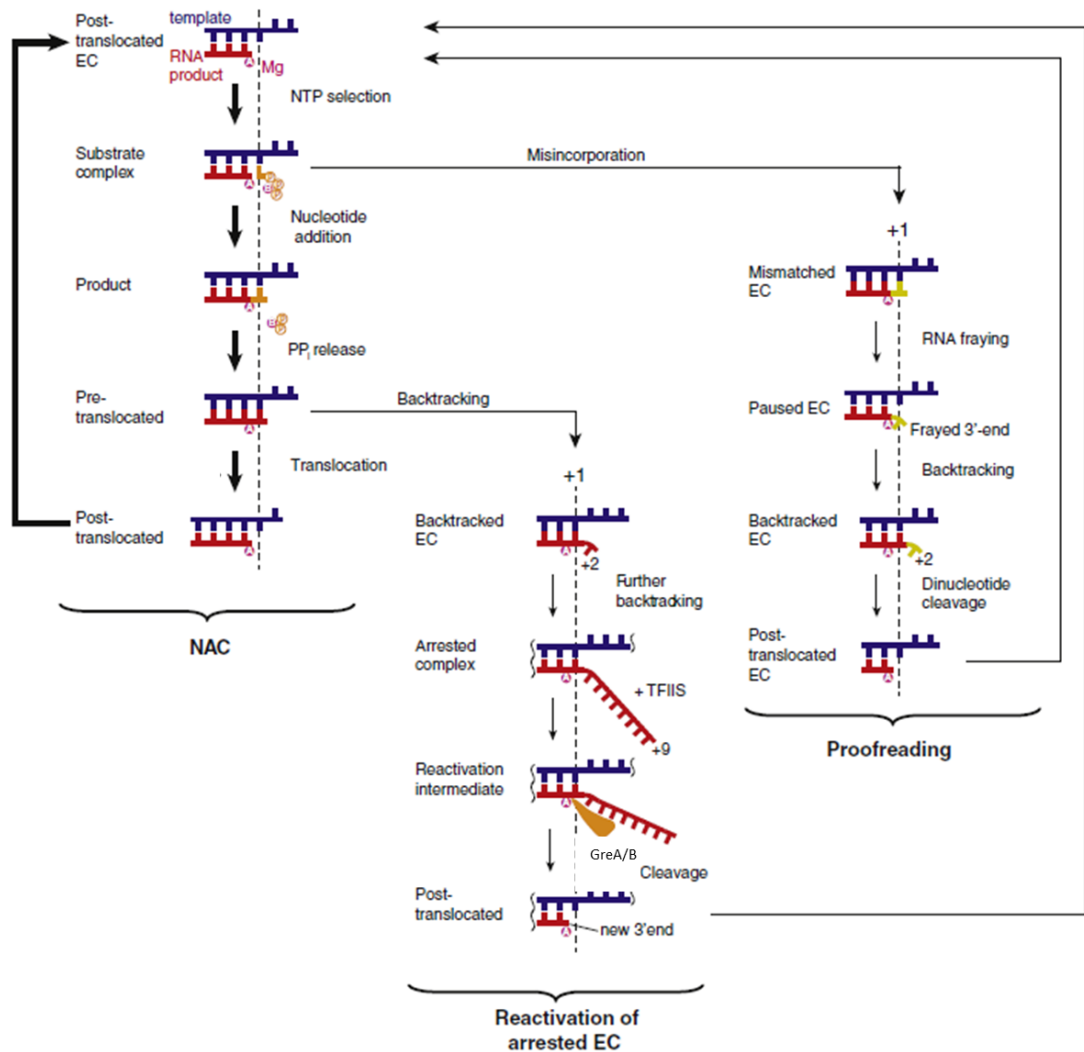


**Figure 2-19: Cartoon of an elongating RNAP with a zoom of the active site shown. Shown in the crystal structure zoom are the features of the protein involved in the elongation process. The DNA is shown in blue with the template strand in dark blue. The RNA is shown in red and the DNA-RNA hybrid can be seen between fork loops 1 and 2 (green). The incoming NTP is shown in orange with the persistent (metal A) and mobile (metal B)  $Mg^{2+}$  ions shown in pink. The switch regions can be seen in purple located at the downstream area of DNA [94].**

The formation of the nascent RNA chain is a cyclic event starting with the uptake of an NTP through the secondary channel. The NTP binds between the 3' of the RNA, the trigger loop and bridge helix of the protein. The insertion of the NTP causes the folding

## *Chapter 2: DNA and transcription*

of the trigger loop which shifts the RNAP to a closed state, if an incorrect NTP is inserted then the protein does not rearrange, biasing the equilibrium to the open state allowing the incorrect NTP to dissociate [148, 149]. Addition of the NTP to the 3' of the RNA is performed by a nucleophilic substitution reaction, involving the persistent  $Mg^{2+}$  ion located in the active site and a mobile  $Mg^{2+}$  brought in by the NTP [150]. After addition of the NTP to the nascent chain, the RNAP is held in a pre-translocation state, a pyrophosphate (PPi) is released and the bridge helix and trigger loop undergo further rearrangement and a new NTP is allowed into the secondary channel. This leads to a translocation step of one bp on the DNA. These rearrangements of the active site for eukaryotic RNAP II can be seen in a video created by Cheung *et al.* [151] elucidated by structures obtained from X-ray crystallography and FRET experiments and a summary of the cycle is shown in Figure 2-20 (left panel).



**Figure 2-20: Schematic of the elongation process with the nucleotide addition shown on the left and the process of backtracking and reactivation shown in the centre and the pausing and proofreading shown on the right. [94]**

The nature of the translocation of RNAP has been shown to most likely occur by a Brownian ratchet model which describes the translocation step as a thermally driven movement that is believed to be rectified by the binding of the next NTP [152, 153]. The addition of an NTP to the RNA chain leaves the elongating complex (EC) in a state of equilibrium between the pre and post translocation arrangements, the subsequent NTP binding shifts this equilibrium towards the post-translocation position [154]. A second theory exists known as the power-stroke model where the translocation is directly linked to the step of NTP addition and the release of the PP<sub>i</sub>, but this model has become less likely to be true with recent evidence from experimental work reviewed by Dangkulwanich which shows that the movement of RNAP at low NTP

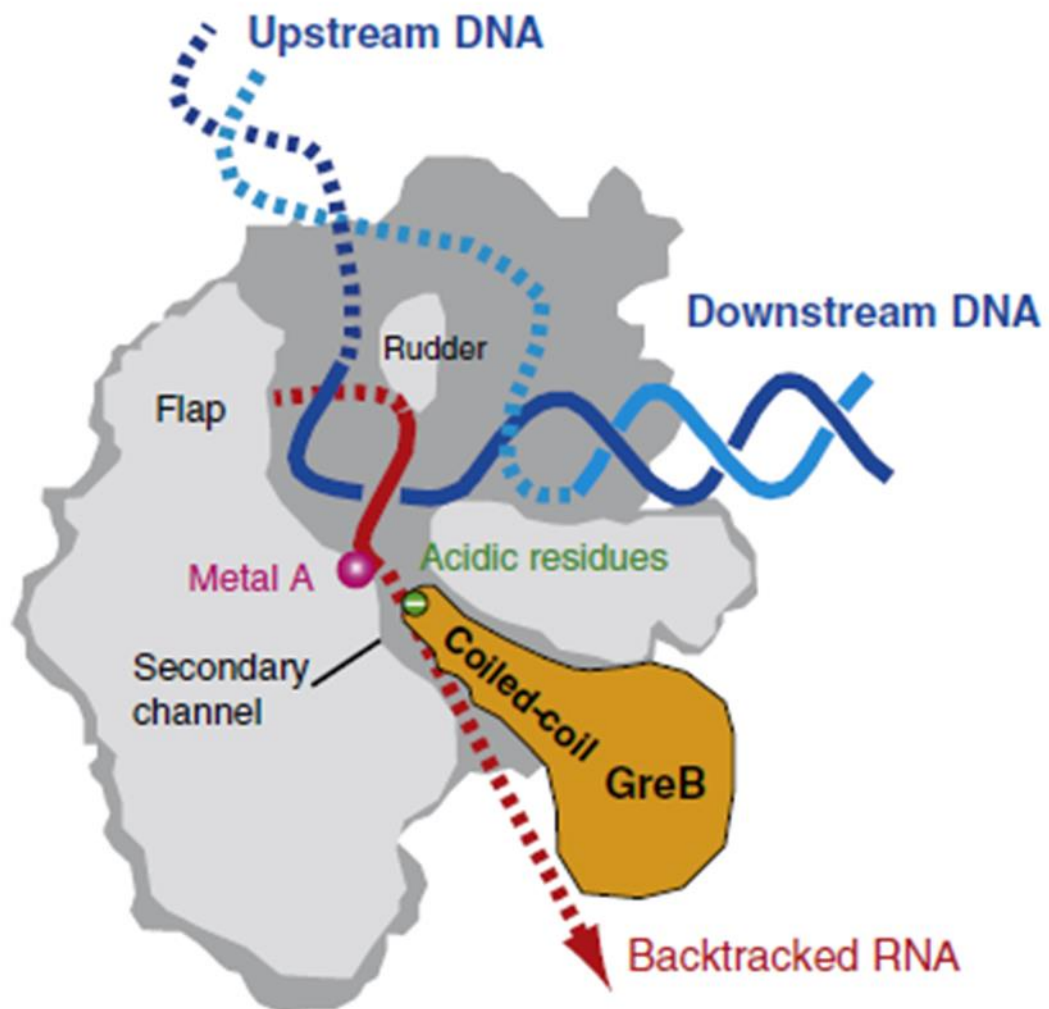
## Chapter 2: DNA and transcription

concentration is affected by force, which in a power stroke model would not hold true [152, 155-157].

As can be seen from the diagram shown in Figure 2-20 the EC also has “off-pathway” states such as: Proofreading and backtracking. Certain sequences lead to destabilisation of the DNA-RNA hybrid, this can then cause the EC to pause for a long period of time (arresting) and backtrack from the +1 site [158, 159]. The EC can also undergo short transient pauses and while translocating these can be caused by sequence, lesions in the DNA, hairpin structures and mismatching of the NTP to the template strand [160]. These pauses have been shown to be ubiquitous *in vitro* and *in vivo* are believed to play a role in regulation of the elongation process and follow a separate mechanism to that of backtracking [161-163].

Pauses that are due to a mismatch between the DNA template strand and the nascent RNA occur in a two-step process. The incorporation of the incorrect NTP into the nascent RNA leads to a lack of base pairing (fraying) at the terminal 3' of the RNA due to the unstable pairing in *E.coli* [164, 165]. In the case of RNAP II mismatch causes a slowing of incorporation of the next NTP, inducing cleavage by TFIIS. The frayed base at +1 register leads to the pausing of the RNAP in *E.coli*. The RNAP then moves back a single step inserting the frayed end into the +2 register which contains the proofreading site [166]. Once at this site, a di-nucleotide is cleaved by the endonucleolytic activity of RNAP, thereby allowing for the mismatched NTP to be removed and transcription to continue [164, 166].

In the case of backtracking it was shown that the 3' of the RNA transcript is released and extruded through a pore beneath the active site which leads to trapping of the trigger loop (Figure 2-21) [167].

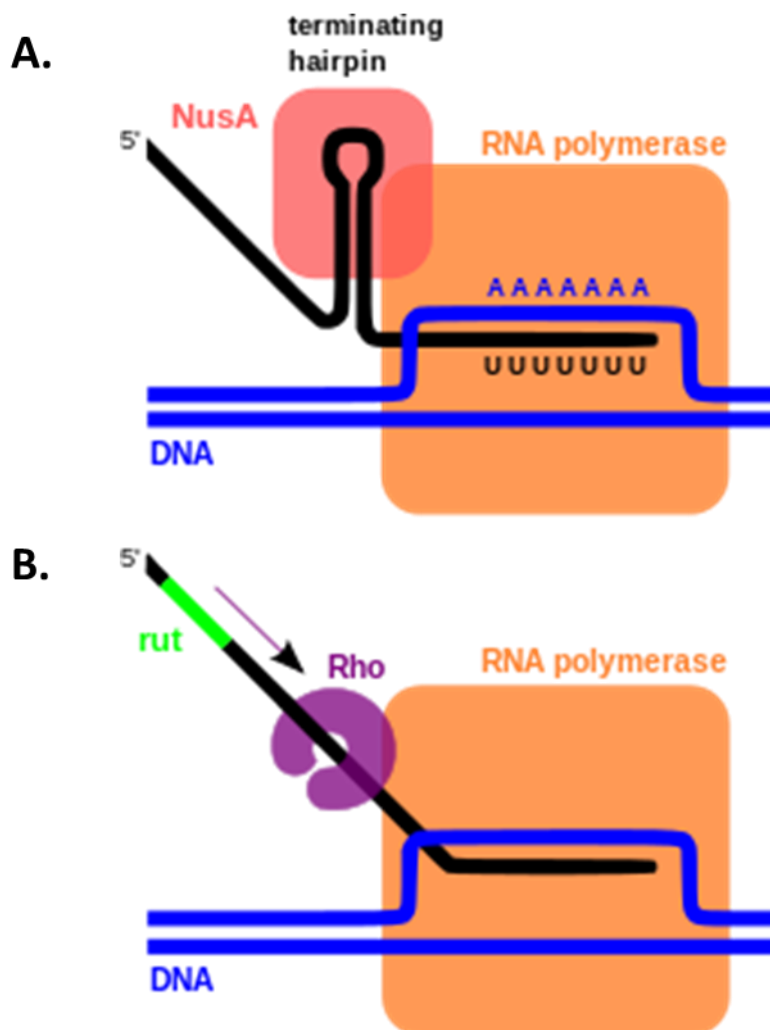


**Figure 2-21: Depiction of a backtracked RNAP with GreB bound. The RNA is extruded through the secondary channel and the RNAP is arrested until GreB is able to cleave the RNA [94].**

Backtracking can result from the same effects that cause pausing but can also be the result of events such as collisions between an EC and other proteins bound to the DNA such as the nucleosome or the Lac repressor [168]. Once in a backtracked state, the RNAP is not able to re-activate elongation until it is rescued by other protein factors, in bacteria these are GreA and GreB, TFIIIS in eukaryotes and TFS in archaea [169].

The final step of transcription is termination. In bacteria there are two forms of termination (Figure 2-23). Intrinsic termination involves the nascent RNA forming a hairpin loop which is bound by the protein NusA, caused by the presence of an inverted

repeat in the template DNA followed by a stretch of A nucleotides [170]. The long stretch of A's means that the DNA-RNA hybrid is destabilised and along with the hairpin leads to release of the RNA transcript (Figure 2-22 A)[171, 172].



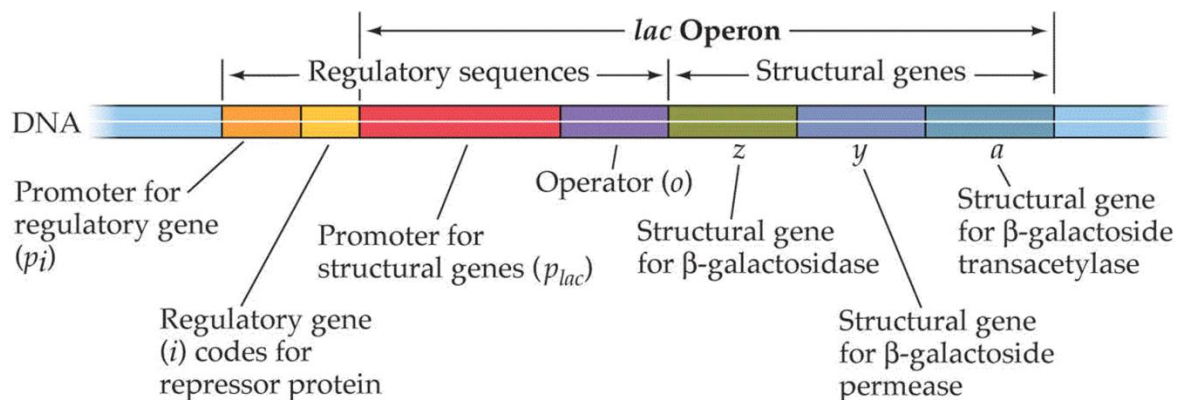
**Figure 2-22: Diagrams of intrinsic and Rho dependent termination. A) Intrinsic involves the formation of a hairpin followed by a long tract of adenosine residues in the template strand. B) Rho dependent termination involves the recruitment of the Rho protein by the rut sequence.**

The second mechanism for termination is Rho-dependent termination (Figure 2-22 B). This involves the Rho protein, which is a ring shaped ATPase motor which binds to a specific sequence known as the Rho utilisation site (rut) [173]. The Rho protein moves up the RNA transcript to the active site and disrupts the DNA-RNA hybrid while also recruiting auxiliary proteins to aid in termination, NusA and NusG, leading to the release of the transcript [174].



### 2.3.3 Prokaryotic and eukaryotic gene organisation

The majority of regions of DNA that are transcribed are known as genes. A definition of a gene is a portion or sequence of DNA that encodes for a known function or process. The production of a functional element from a gene is referred to as expression. Many genes contain certain shared elements such as promoter and terminator sequences. The organisation of genes and their regulation differs between prokaryotes and eukaryotes. Prokaryotic genes are often organised into operons which is a cluster of genes with related or complementary function controlled by a single promoter as seen in the Lac operon where the proteins required for the metabolism of lactose are maintained in one operon [175, 176]. Contained within the operon are not just structural genes but regulatory sequences such as the operator which is bound by a regulatory factor (Figure 2-23).

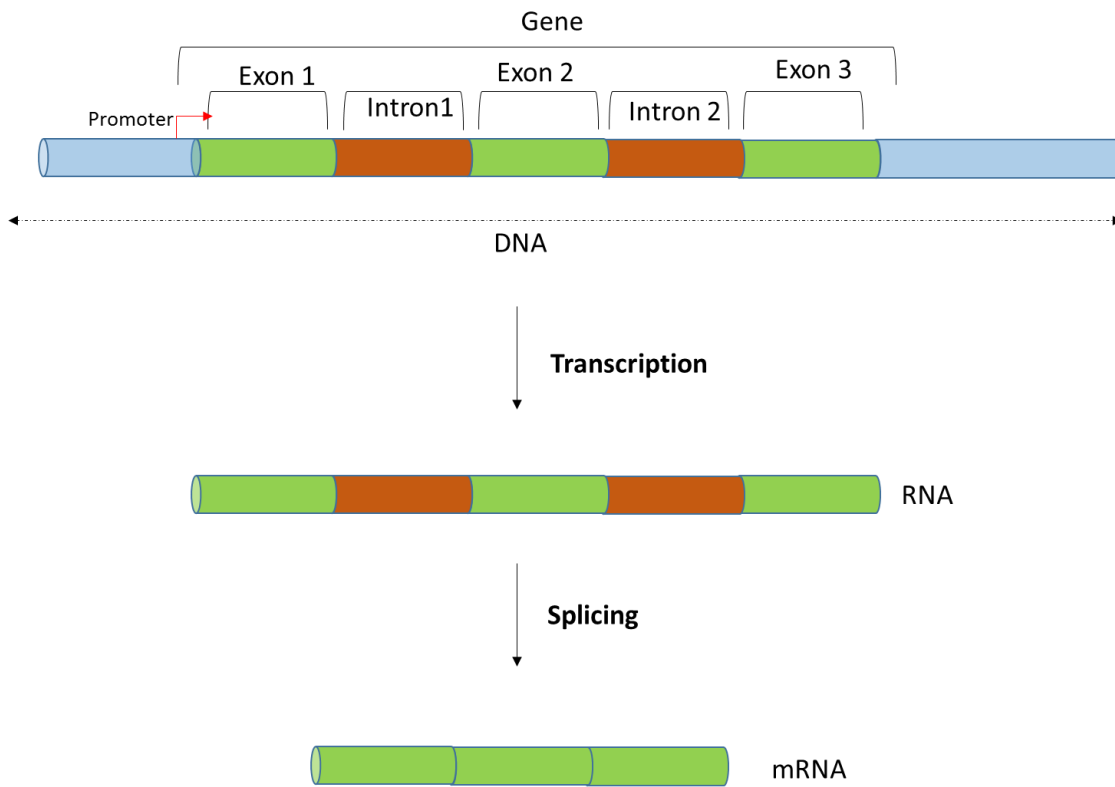


**Figure 2-23: The *E.coli lac* operon. As can be seen, an operon contains the sequence for a number of proteins under the control of a single promoter. Structural proteins and the operator are also contained within the operon.**

This structure allows for the quick adaptation to environmental factors [177]. The transcription of an operon produces a polycistronic mRNA, which encodes the amino acid sequence for each gene in the operon. As prokaryotes lack an enveloped nucleus, translation can be directly linked to transcription, with each gene being translated from a single polycistronic mRNA in a process referred to as translational coupling [178].

## Chapter 2: DNA and transcription

Eukaryotes genes generally are not organised into operons but some cases have been recorded [179]. Each gene tends to have its own promoter. Transcription occurs in the nucleus of the cell and the RNAs produced undergo a number of processes before being fully matured and exported into the cytoplasm in the case of protein coding genes. In eukaryotes genes are often formed of introns and exons (Figure 2-24).

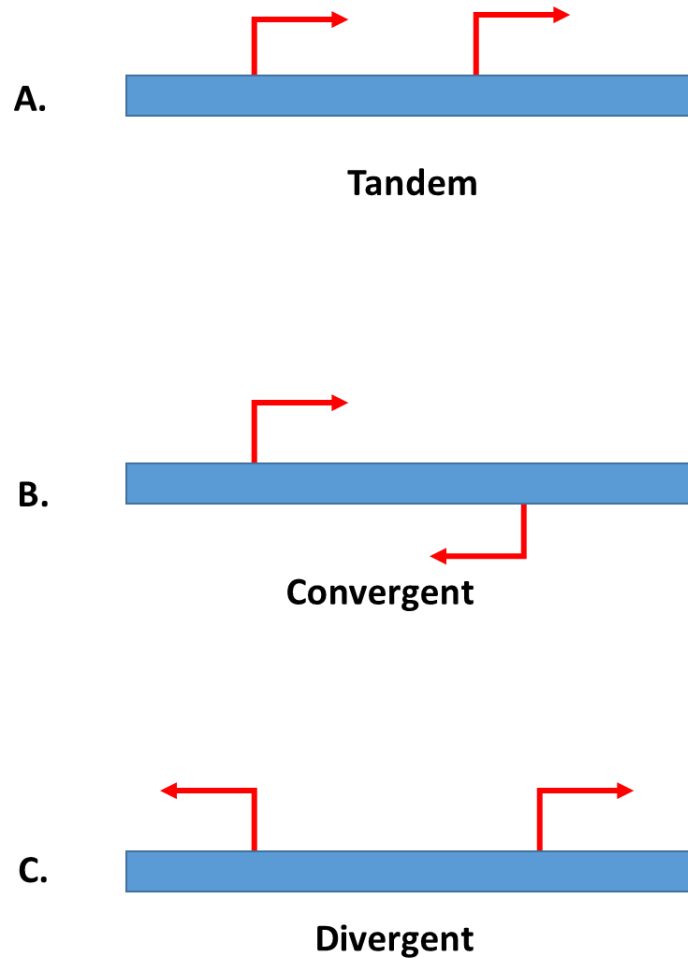


**Figure 2-24: Diagram showing the organisation of a eukaryotic gene. The DNA either side of the gene is shown in blue with the exons shown in green and introns in orange. The start of the gene is shown by the promoter (red). Both the exons and introns are transcribed and then introns are removed by splicing to provide a processed mRNA.**

Introns are non-coding sections of the gene. They are transcribed by RNAP but are subsequently removed by the process of splicing before export of the mature mRNA from the nucleus and cytoplasmic translation begins [180]. Introns originally were originally thought of as “junk” DNA, not having a use, but have since been noted as playing a number of roles in gene regulation, retro-transpositions and intron stability [181].

### 2.3.4 Orientation and gene pairs and nested genes

As promoters determine the coding and template strand during transcription they also give directionality to genes. When a second or multiple promoters are in close proximity to each other, the genes can be broken down into pairs or sets. These pairs can be seen as having three different arrangements (Figure 2-25).



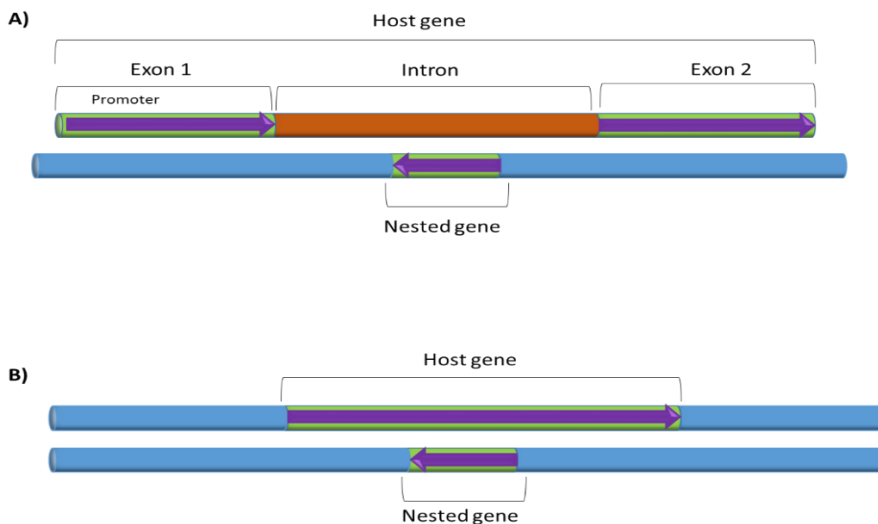
**Figure 2-25: Diagram of different promoter arrangements. A) Tandem promoters direct transcription in the same direction on the same strand of DNA. B) Convergent promoters direct transcription towards each other and are located on opposite strands of the DNA. C) Divergent promoters direct transcription away from each other and occur on the same strand of DNA.**

In many cases where genes have promoters located in close proximity to each other, genes overlap. Fukuda *et al.* analysed the genomes of 50 bacteria and found that the

## Chapter 2: DNA and transcription

number of overlapping genes increased with size of the chromosome [182]. It was found that 71.4% of overlapping genes were in a tandem orientation while convergent and divergent genes represented 14.3% each [182]. These numbers represented protein coding genes and it is expected that a much higher amount of overlapping genes are present when including ncRNAs [183]. In eukaryotes overlapping genes are also common. Veeramachaneni identified 2541 overlapping gene pairs in the human and mouse genomes and Chen and Stein identified 3971 overlapping genes with 500 bp or less between transcription start sites in *Caenorhabditis elegans* (*C.elegans*) [184, 185]. One of the most common types of overlaps seen in eukaryotes is that of nested genes.

A nested gene is a gene whose entire coding sequence lies within the boundaries of another gene. Nested genes account for up to 75 % of overlapping genes in humans and account for 4.3 % of the genome when including known ncRNAs [186, 187]. Nested genes account for 2.7 % of protein coding genes in *C.elegans* and 6.1 % in *Drosophila melanogaster* [185]. The first reported nested gene was the pupal cuticle protein which is located on the opposite strand of an intron of the adenosine 3 protein gene in *Drosophila* [188]. There are two types of nested: intronic nested genes (Figure 2-26 A) and non-intronic (exonic) nested genes (Figure 2-26 B).



**Figure 2-26: Schematic representation of nested genes. A) A gene lying within an intron of another gene known as an intronic nested gene. B) A gene nested with the coding region of another gene known as non-intronic or exonic nested gene.**

## *Chapter 2: DNA and transcription*

Intronic nested genes are the most common and only occur in eukaryotes as prokaryotic genes do not contain introns. Genes can be nested in either a convergent or tandem arrangement with Yu showing that the majority of nested genes in humans had a convergent arrangement [186]. Host genes can also carry a number of nested genes which can be arranged in both tandem and convergent orientations, as is seen in the *dunce* locus in *Drosophila* which contains six nested genes spread over two introns [189].

The occurrence of nested genes is not fully understood but it is believed that the majority occur due to transposition events such as insertions, gene duplications and the fusion of two genes [190]. The persistence of nested genes has been suggested to occur by the “sheltered island theory” presented by Chen *et al.* [185]. This theory suggests that if a gene is nested in a host gene that is essential then removal of such a gene could be fatal to a cell and so the nested gene remains intact.

The biological relevance of nested genes has also not been elucidated. In bacteria nesting of genes may help to ensure a compact genome but this would not hold true in the case of eukaryotes as space is not at such a premium. Gibson *et al.* suggest that nested genes may allow for the co-expression of related proteins but results presented by Yu *et al.* suggested this is not the case as very few nested genes share similar function with the host gene [186]. Evidence exists for the co-regulation of nested genes as it was observed by both Yu *et al.* and Chen *et al.* that nested genes showed correlated expression profiles [185, 186]. Yu *et al.* studied the expression of 45 nested genes found in humans and found that 4 tandemly arranged and 29 convergently arranged genes displayed a negative expression correlation while only 3 showed a positive correlation and 9 had no effect on each other. This was reflected in the case of *C.elegans* nested genes where Chen *et al.* noted that tandem arranged overlapping genes displayed a highly positive correlation except when the gene was nested and that convergently overlapping genes displayed a general negative correlation [185].

### 2.3.5 Transcriptional interference and collisions

A consequence of nested and overlapping genes is the possibility of transcriptional regulation via the process of transcription itself. This process is known as transcriptional interference (TI). TI is defined as the negative impact of a transcription event by a secondary transcription event. It is therefore linked to genes which have promoters in close proximity (less 600bp) as seen in some overlapping genes as well as in nested genes. The mechanisms for TI were outlined by Shearwin *et al.* (Figure 2-27) [2]:

- a) Promoter competition: This occurs when the occupation of one promoter by RNAP inhibits the occupation of a second promoter
- b) Sitting Duck (SD) interference: This can occur when an RNAP is slow to enter its elongation phase and is struck and/or dislodged by an elongating complex (EC).
- c) Occlusion: An EC from one promoter blocks the binding to a second promoter due to the EC blocking access to the promoter, therefore occluding the promoter.
- d) Collision: A collision between two ECs can lead to transcriptional arrest for one or more of the ECs.
- e) Roadblock: An RNAP bound to its promoter could be stuck by an EC and not be dislodged therefore acting as a roadblock to transcription

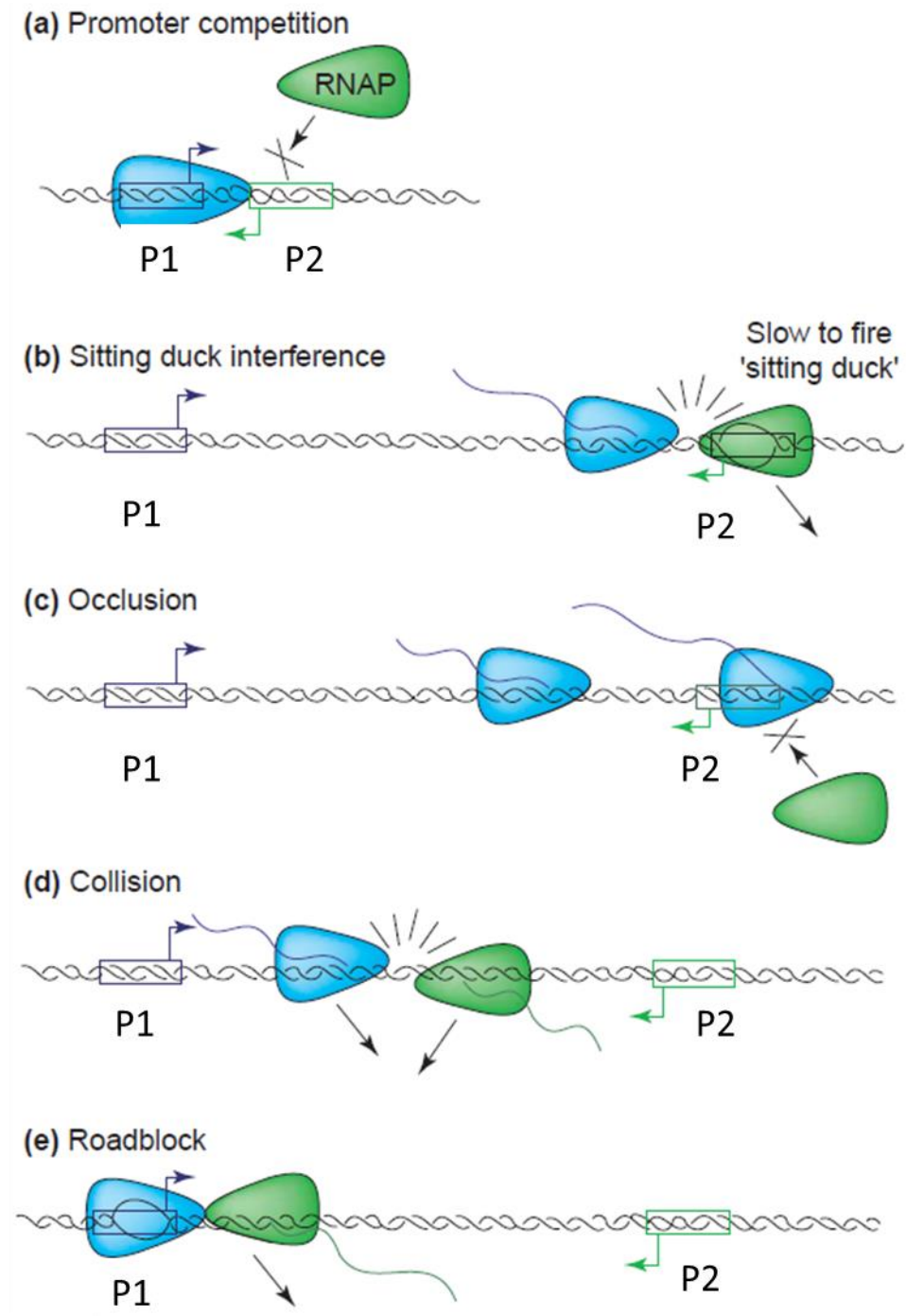


Figure 2-27: Possible mechanisms of TI as predicted by Shearwin *et al.* [2].

This thesis aims to specifically study the outcomes of collisions between two actively transcribing RNAP molecules either transcribing in a convergent or tandem orientation and so investigations into the occurrence of collisions are discussed.

## Chapter 2: DNA and transcription

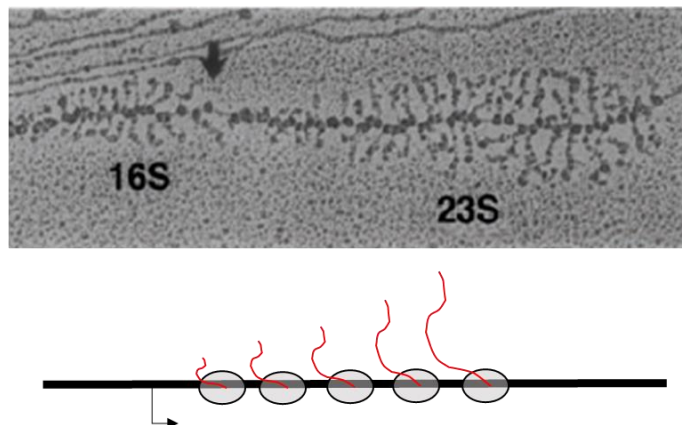
Prescott and Proudfoot were among the first to suggest that collision events are a major contributor to TI [191]. They investigated the expression of the *Gal7* and *Gal10* genes in budding yeast by RNAP II [191]. An expression vector containing two promoters arranged in a convergent orientation without termination sequences resulted in the down-regulation of both genes. As initiation events were unperturbed and no down regulation was noted from RNA interference when the genes were expressed on separate templates, it was reasoned that collision events between actively transcribing RNAPs lead to termination of transcription [191]. Callen *et al.* were able to show that in the case of a weak and strong promoter arranged in a convergent orientation resulted in the removal of OPCs from the weak promoter indicating that SD collisions were the most common form of TI and that an OPC does not act as a roadblock for another convergently transcribing RNAP [192]. This result was further supported by results of modelling by Sneppen *et al.* [193]. Sneppen *et al.* predicted that SD collisions are most likely between two promoters of differing strength while EC-EC collisions were more likely between equal promoters. It was also noted that the occurrence of EC-EC collisions increase as the distance between the two promoters increases as there is expected to be a higher probability that the inter-promoter region is occupied by two active RNAPs [193]. Investigations by Crampton *et al.* using AFM to study convergent transcription from two  $\lambda_{pr}$  promoters by *E.coli* RNAP found that EC-EC collisions and EC-SD collisions resulted in shunting (large scale backtracking) of one of the RNAPs and this was referred to as a SD collision but as transcription is arrested it could be considered to be TI via the roadblock mechanism [3]. It was also seen that both RNAPs stalled and remained on the template but were not in hard contact after collision events raising questions on the nature of stalling and arrest of transcription [194]. The stability of collided complexes was also shown by Hobson *et al.* for RNAP II [195]. After a collision between an EC and a stalled elongation complex (SEC) ubiquitination was required to remove the RNAPs from the template [195]. The results of Crampton *et al.* and Hobson *et al.* differ from those of Callen *et al.* in that they indicate that an RNAP either as an SEC or OPC can act as a transcriptional roadblock. Hobson also was able to show in cells unable to perform ubiquitination, through the use of chromatin immunoprecipitation, that for convergent genes with an inter-promoter separation of less than 400 bp collisions were common, indicated by the high levels of RNAP present on the DNA (Figure 2-28). Terminal collisions have also



## Chapter 2: DNA and transcription

been seen in bacteria as reported by Chatterjee *et al.* who studied the *prgX/prgQ* operon which is involved in the transfer of antibiotic resistance plasmids [196]. They reported that both collisions and antisense RNAs play a role in providing a bi-stable switch mechanism for controlling gene expression [196]. In the case of viral RNAPs Ma *et al.* observed that collisions may be avoided by RNAPs passing each other [197]. This is believed to be due to the temporary release of the non-template strand. This mechanism is not expected for pro- and eukaryotic RNAPs as their size would not allow passing even if release of the non-template strand was to occur.

In the case of tandemly orientated promoters, collisions are not considered to be as common. A large number of genes in all organisms undergo simultaneous transcription by RNAPs moving in tandem, as is seen by the Miller spreads of the *E. coli* and eukaryotic rRNA genes (Figure 2-28).



**Figure 2-28: Miller spread electron micrograph of *E. coli* rRNA 16s and 23s genes. Below the micrograph is a schematic representation of RNAPs transcribing in tandem [198].**

Many of these examples involve multiple transcription events from a single promoter leading to a high density of RNAPs on a single gene. In the case of transcription originating from two separate tandem promoters it is believed that TI is more likely to occur through a process of occlusion. Even so, the occurrence of collisions between RNAPs and roadblocks such as nucleoid proteins indicate that RNAPs are able to act in a synergistic manner. This has been seen for both prokaryotic and eukaryotic RNAPs. Jin *et al.* reported that two RNAP II molecules were able to act co-operatively to

## Chapter 2: DNA and transcription

overcome a nucleosome positioning element [199]. Saeki *et al.* showed that collisions between a SEC and an EC resulted in backtracking of the trailing RNAP but in the case of a paused leading RNAP II, the trailing RNAP II was able to rescue the paused RNAP and therefore acted to increase transcription rate [200]. Epstein and Nudler studied multiple transcription events from a single promoter by *E.coli* RNAP and found that a second RNAP can aid in the rescue of backtracked complexes with this rescue being further aided by addition of ECs indicating that the co-operativity of ECs is cumulative [201].

Not all cases of collisions between tandemly transcribing RNAPs results in co-operation. Kubori *et al.* reported the formation of stalled inactive (moribund) complexes when collisions occurred between two tandem transcribing RNAPs, reasoning that the leading RNAP was able to exert a negative effect on the trailing RNAP [202]. This data is in concurrence with findings by Ponnambalam and Busby who reported the reduction in expression of a downstream promoter as well as production of a truncated transcript from an upstream promoter when transcribed tandemly [203]. This interference only occurred when the two promoters were located at a distance of 86 bp apart and was reasoned to be due to collisions between the two RNAPs leading to premature termination of transcription.

In many of the studies presented here collision events are monitored between SECs rather than between two actively transcribing RNAPs. Many studies also utilise promoters that have differing strengths meaning that collisions are not the most likely outcome. Work carried out by Crampton *et al.* addresses some of these issues but the experiments performed lacked a competitive inhibitor in order to rule out non-specific interactions as well as mainly investigating SEC-EC collisions [194]. This thesis aims to investigate the outcomes of concurrent transcription from both convergent and tandem gene arrangements using simplified gene models as presented in Chapter 4 for *E.coli* RNAP. The outcomes of these transcription events is studied using AFM, which is discussed in Chapter 3 to provide a single molecule view of the occurrence, type and outcomes of collisions and TI that may occur during concurrent transcription events.

This page is left intentionally blank

# 3 Introduction to AFM of DNA protein complexes

## 3.1 Atomic force microscopy for investigating transcription

Atomic force microscopy (AFM) is a technique that is highly versatile and suited for the study of biological samples. Many techniques do not allow the direct visualisation of molecules as AFM does. Since the development of the first AFMs one of the most commonly visualised biological molecules has been DNA. The versatility of AFM has meant that not just the structure of DNA alone but also the interactions of DNA-binding proteins and other molecules has been investigated as is shown by the wealth of literature in these areas [204-211]. AFM can allow for the imaging of samples in ambient conditions, allowing a snapshot of a process to be visualised in 3 dimensions (3D), but can also be used in liquid to provide information on the dynamics of biological processes. Moreover, with the development of high speed AFM techniques, there is now the possibility to visualise dynamic processes not just with high spatial resolution but with high temporal resolution as well [208, 212].

The study of DNA-protein interactions by AFM is aided by the relatively non-destructive sample preparations. Unlike electron microscopy and other single molecule techniques such as fluorescence based systems, AFM does not require the use of stains or dyes which can alter the native state and function of DNA and associated proteins [213]. Sample preparation for DNA and DNA-protein samples tend to be similar in many studies meaning that a number of different systems can be studied easily [214-216]. The focus of this thesis is the study of transcription and the outcomes of concurrent transcription from multiple promoters. Transcription has three main stages as already discussed and so AFM allows the study of each of these stages. Transcription is an active process but it is possible to study the process at different stages, such as OPC formation or post elongation, using AFM in ambient conditions in an *ex situ* manner. Studies in liquid can provide insight into the dynamics of this process as a whole but using AFM in ambient conditions it is possible

### *Chapter 3: AFM of DNA and proteins*

to gain insight into the spatial arrangement of RNAP molecules and the DNA at certain time points due to the ability of AFM to distinguish molecules through its high resolution in the z axis and in the x/y axis. This means that molecules of different size can be distinguished from each other. The single molecule nature of AFM means that it is possible to study sub-populations of molecules that would normally be masked in studies using bulk biochemical methods.

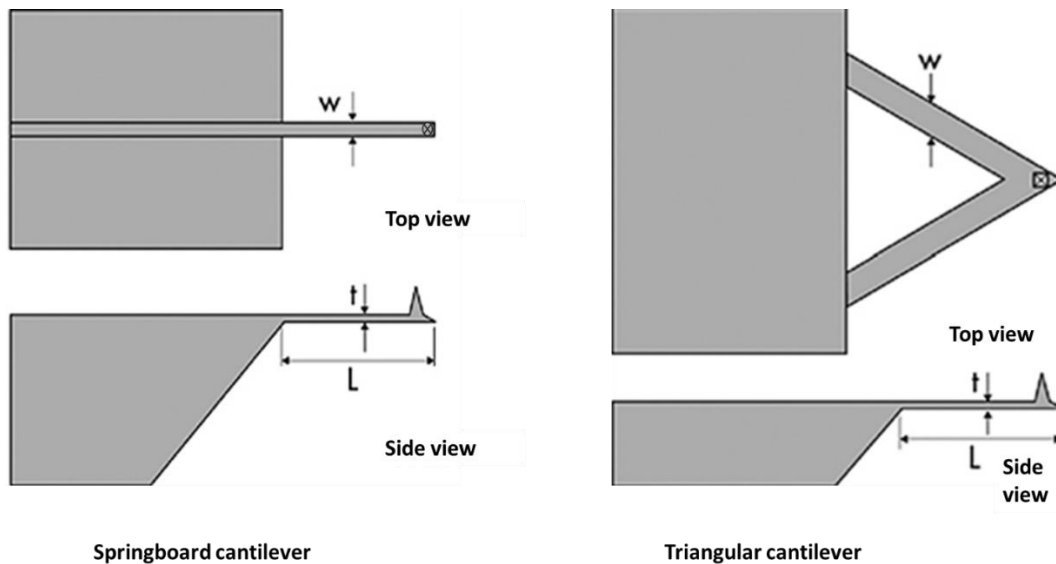
#### **3.1.1 Scanning probe microscopy**

The AFM is a type of instrument known as a scanning probe microscope (SPM). These instruments utilise a probe which is mechanically moved with respect to a sample. The interactions between the probe and the surface are recorded to allow for a 3D representation of the surface to be constructed. The nature of the probe and the interactions measured can mean that different properties as well as topography of the surface can be mapped. The original SPM was the scanning tunneling microscope (STM) developed by Binnig *et al.* in the 1980s [217]. The STM detects the tunneling current between a sharp probe and the sample surface. Due to the exponential dependence of the tunneling current on the tip-sample separation the device has high sensitivity to changes in the surface topography. The main drawback of the STM is the need for the sample to be electrically conductive. Even though atomic scale images of DNA were obtained by Driscoll *et al.* on highly orientated pyrolytic graphite (HOPG) it has been suggested that the use of HOPG as a surface substrate can lead to misinterpretation of the DNA structure, due to HOPGs periodic structure, meaning that reliable imaging of DNA or biological samples is not viable [218, 219]. This requirement for conductive samples was one of the key drivers that led to the invention of the AFM [220]. AFM utilises a flexible cantilever that is deflected by surface features and events due to ubiquitous inter-molecular forces. The deflection is measured to provide a map of the surface and its features, meaning any sample surface can be examined. AFM has been shown to be highly sensitive with sub-molecular imaging being possible and can be operated in either ambient conditions or in a liquid environment [221-225]. When operated in liquid it is possible to study biological processes in a physiologically relevant environment providing information on dynamics of DNA-protein interactions whereas

studying samples in air bound to a surface provides more information on the positioning of elements, such as binding proteins on the DNA [226].

### 3.1.2 AFM instrumentation

The key component of the AFM is the probe, which consists of a cantilever with an integrated tip that interacts with the sample through short and long range forces that are localized to the tip. Design of cantilevers vary but they are usually micro-fabricated from silicon or silicon nitride. The most common designs are triangular (often used for contact mode) and the “diving board” shape (mainly used in dynamic modes) (Figure 3-1). Cantilevers can have a range of spring constants typically from  $0.01$  to  $100 \text{ Nm}^{-1}$  [227]. The tip has a radius of curvature on the order of nanometers.

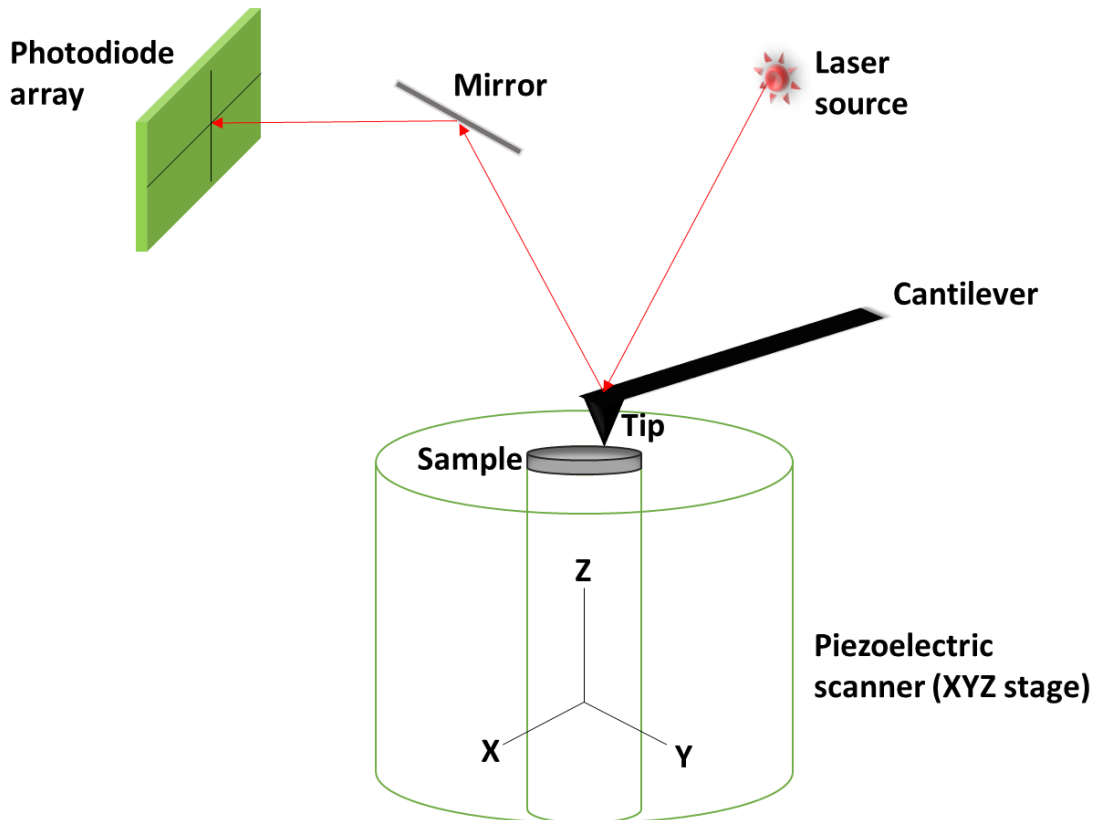


**Figure 3-1: Diagrams of the most common AFM probe designs. On the left is a springboard design, often used for probes used in dynamic AFM modes. On the right is the triangular design often used for contact mode imaging. On both the dimensions referred to are highlighted, W is width, t is thickness and L is length.**

**The position of the tip in the top view is denoted by the circled cross.**

The deflection or changes in the behavior of the cantilever are caused by forces arising from the tip-surface interaction and so a sensitive detection method of changes in the cantilever is needed. Original AFMs utilised an additional tip on the topside of the cantilever to detect tunneling current changes upon deflection but it was found that

this tunneling tip could exert forces on the cantilever distorting measurements [220]. The most commonly used method for detection currently is the optical lever method developed by Meyer *et al.* [228]. This involves the focusing of a laser onto the backside of the cantilever which is then reflected to an adjustable mirror which in turn reflects the laser onto a quadrant photodiode array (see Figure 3-2 which provides a schematic representation of typical AFM setup).



**Figure 3-2: Schematic of basic sample scanning AFM setup. The scanner has the outer tube used for x and y translation and an inner tube for z movement. This arrangement is not true for all AFMs but many utilise a system based on similar concepts. The sample is placed on the scanner and the tip is held on or near the sample surface. The laser reflects off the cantilever backside and is reflected onto the mirror. The mirror is then adjusted to centre the laser on the photodiode array.**

The adjustable mirror is used to position the laser onto the centre of the photodiode array. This means that changes in the cantilever lead to changes in the position of the laser on the photodiode array therefore providing a measurement of the deflection of the cantilever. The amplified deflection of the laser can provide a sensitivity of up to 0.01 nm [229, 230]. As the forces that act upon the tip are short range, the AFM must

be able to accurately maintain a small tip to sample distance (control in the Z dimension). There is also a requirement for accurate movement in the x and y directions in order to provide precise movement when moving and scanning over the surface. The majority of AFMs achieve this precision by using piezo ceramic materials [231]. These materials contract or expand when a voltage is applied to them [232]. Commonly used is a sample scanner in which the scanner has a piezoelectric tube divided into four areas, as voltage is applied to each area X and Y movement is achieved. Z movement is controlled by a second piezo tube. There is a number of different AFM designs with piezo elements being organised differently, but many draw on the same concepts to provide functionality. Some AFMs use a tip scanner which involves the movement of the tip rather than the sample.

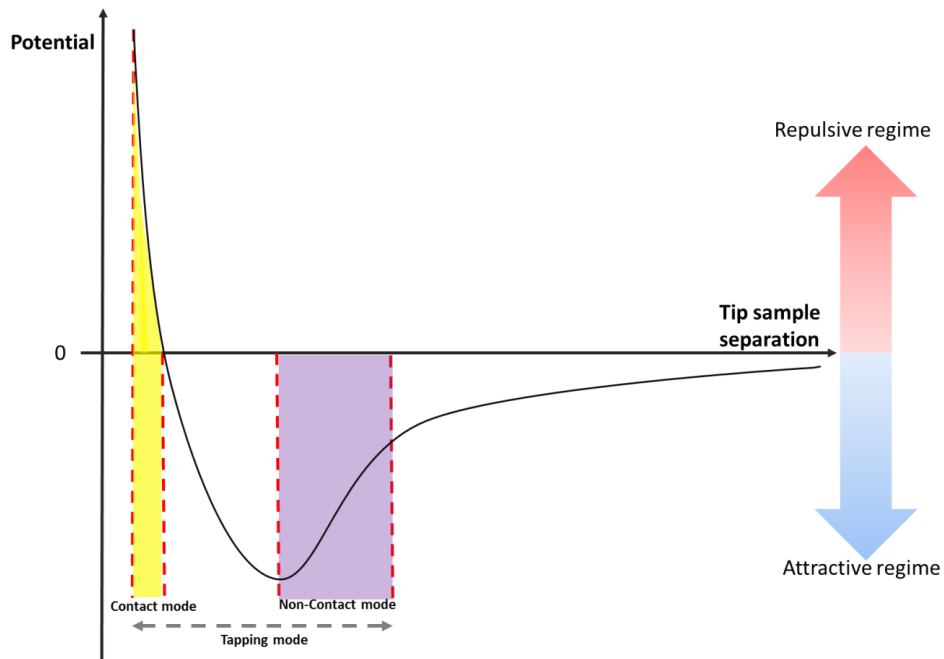
### **3.1.3 Forces in the AFM**

Image contrast in AFM is provided by forces that occur between the tip and sample. There are multiple forces that act on the tip leading to deflection of the cantilever (these forces are reviewed in detail by Israelachvili [233]). The relationship between force and deflection can be summarised by Hooke's law:

$$F = -kx$$

where  $F$  is force,  $k$  is spring constant and  $x$  is displacement. The origin of the interaction forces can be modelled most simply by the "Lennard-Jones" potential. The model provides the relationship of interactions between to atoms or molecules when brought into close proximity (Figure 3-3).





**Figure 3-3: Graphical representation of the Lennard-Jones potential. As the separation between the tip and sample decreases it undergoes different forces. As the separation decreases there is a slow change in potential due to attractive forces such as Van der Waals forces acting on the probe. As the tip moves closer there is steep rise in potential as the two molecules or atoms repel each other according to the Pauli Exclusion Principle. Highlighted on the graph are the separations at which the tip is operating in the attractive regime (blue arrow) and repulsive regime (red arrow). Also highlighted is the separations where the tip is situated on the graph when operated in different modes. Non-contact in purple and contact shown in yellow. When operated in tapping mode the tip moves in and out of both these regimes in one oscillation cycle (adapted from [234]).**

As the separation between the tip and sample changes so does the potential energy. At low separations there is sharp increase in potential as the atoms or molecules repel each other due to the Pauli exclusion principle that states that two identical fermions and therefore electrons cannot occupy the same space [235]. At larger separations there is more gradual decrease in the potential which is due to Van der Waals forces dominating the interaction. Van der Waals forces arise from the correlation of neighboring atoms dipole moments. Van der Waals forces are relatively weak, always attractive and act over distances from 0.2 to 10 nm.

### *Chapter 3: AFM of DNA and proteins*

There is also adhesive forces acting between the tip and sample surface when imaging, which are dominated by capillary force in ambient conditions [236]. When imaging in ambient conditions a water layer can form on the substrate surface and on the probe. If the substrate surface and probe are hydrophilic it is possible that a capillary neck can form between the tip and surface [237-239]. These capillary forces can range from 10 nN to 100 nN and are dependent on tip radius, tip geometry, humidity and temperature [223, 240, 241]. The effect of these capillary forces is not fully understood especially when the AFM is operated in a dynamic mode. There has been number of attempts at modelling the effects these forces have on tip sample interactions and image contrast [242-244].

#### **3.1.4 Imaging Modes**

Originally AFMs were only able to operate in one mode, contact mode. This is still used today but is generally not suitable for biological samples. In contact mode the tip is held in constant contact with the surface and can be operated in two ways: constant height and constant force. Constant height mode holds the tip at a constant average Z position above the surface with no feedback systems. The cantilever is deflected by surface features providing a topographical map of the surface. In constant height mode rough samples can lead to high forces but provides a good vertical sensitivity. In order to avoid these high forces a feedback loop can be used to maintain the tip deflection or tip-sample force (constant force mode) by altering the height of the tip. The vertical position of the tip provides a representation of the surface topography. In contact mode there are large shear forces present due to the tracking of the tip on the surface. These shear forces mean that soft samples, such as biological samples will be dislodged or damaged by the tip. Early images of DNA were collected by contact mode imaging but it was found that in ambient conditions capillary forces accentuated the effects of shear forces making imaging difficult [245, 246].

This inability to image soft samples led to the development of tapping mode [247]. Tapping mode involves oscillating the cantilever at or near its resonance frequency through the use a piezo element on the cantilever holder or by an oscillating magnetic field with a magnetised cantilever. The tip is oscillated with an amplitude large enough

to overcome adhesion forces (normally between 1-25 nm) and intermittently strikes the surface. Upon striking the surface, energy is dissipated and the amplitude of the oscillation decreases. The feedback loop works to ensure that the imaging amplitude remains at a set value by altering the height of the tip. The amount the amplitude must be modulated can be used to provide a topographic map of the surface. This mode of operation can therefore be referred to as amplitude modulation (AM) AFM. This can be an important distinction from using the term tapping and non-contact imaging as the in AM AFM the amplitude setpoint can be set to image in different imaging regimes, attractive or repulsive, and so the tip does not necessarily have to make any contact with the surface (Figure 3-3) meaning that AM AFM covers both intermittent contact imaging (tapping mode) and non-contact imaging. Imaging in different regimes can provide different advantages and disadvantages and depends on the sample being imaged [248]. It has also been shown that imaging in the repulsive regime with low amplitude of oscillation can help overcome capillary forces, meaning that the tip is more stable above the surface and provide better resolution images [221, 249-252].

Due to the intermittent nature of the tip sample contact, the shear forces are reduced to sub nN values [250]. The low shear forces and minimal sample deformation by the tip mean that AM AFM is good for imaging biological and soft samples. As the contrast is provided by the energy dissipation at the end of each oscillation cycle it is possible to map material properties such as adhesion and friction of some samples as well as topography [253, 254]. This can be achieved by comparing the phase of the cantilever to the phase of the drive frequency. A softer sample will lead to a greater damping of the oscillation leading to greater phase lag than would be seen for a harder sample [255, 256]. The phase and topography image can be collected simultaneously. In some cases the phase image can allow for higher scan rates and lower forces to be used, as well as the contrast helping to highlight features that may not be visible in the topography [257].

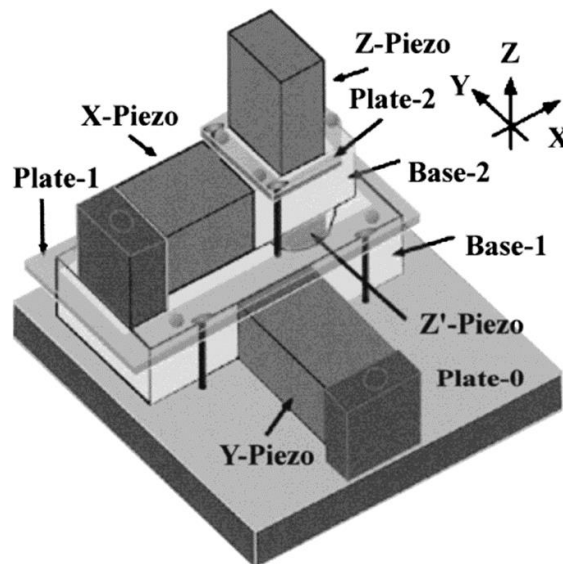
### **3.1.5 High speed AFM**

One of the major issues with AFM is the speed at which images can be collected (reviewed in detail by Ando in [208]). The speed at which an AFM can operate is

### Chapter 3: AFM of DNA and proteins

governed by a number of limitations. In the early 2000's a number of new AFM designs were reported with scanning speeds equivalent to video rate, in both contact and tapping modes [212, 258-262]. In the case of tapping mode one of the major limitations was the probe. As the probe has to oscillate through one cycle to collect each pixel, the time this cycle takes limits the speed a pixel can be obtained. To increase this cycle speed the resonance frequency of the cantilever had to be increased, but this is difficult due to the increase in spring constant needed to achieve high resonance frequency of the cantilever. This issue was overcome with the development of much smaller cantilevers, that had resonant frequencies in the range of MHz rather than KHz [208]. The development of small cantilevers went hand in hand with the development of a more precise laser sources to ensure that the laser spot on the cantilever was not too diffuse as well as the development of a RMS-DC converter system that was capable of keeping up with the signal and information from these new cantilevers. This was achieved by the development of an RMS-DC converter that was capable of utilising half a wave to converter signal rather than the standard 5-6 waves [208, 212]. The laser spot issue was remedied by the use of an objective lens based system rather than the simple mirror system seen in many conventional AFMs [208].

Another factor that limits the speed of AFM is the movement of the scanner. The scanner must be able to move in accordance with the speed of the tip oscillation and be able to provide feedback at higher speeds, as well as not suffer from interference from the oscillation of the cantilever. Ando *et al.* designed a scanner that had stacked piezo elements with the X and Y piezo in one plane and the Z piezo in another allowing for fast and smooth movement in X and Y while minimizing vibrations in the Z direction (Figure 3-4) [208].



**Figure 3-4: Schematic of the stacked piezo scanner as developed by Ando. The piezo elements are stacked on top of each other rather than using the conventional tube arrangement in standard AFMs [212].**

Presented here is one example of a high speed AFM developed by Ando *et al*, it is noted though that a number of machines were developed around the same time by other groups as well and now both the Ando machine as well as others, such as the Bruker Fast Scan and Asylum Cypher are commercially available.

In the case of contact mode scanning, higher speeds were obtained through the use of a micro resonator as the sample stage with a passive mechanical feedback loop to maintain the average force over a timescale of one frame [261]. This allowed for frame rates of around 70 frames/second to be achieved. This was further developed to incorporate a tuning fork which the sample could be placed on and so vibrated at higher speeds relative to the tip, thus providing a scan rate of around 1000 frame/second [260].

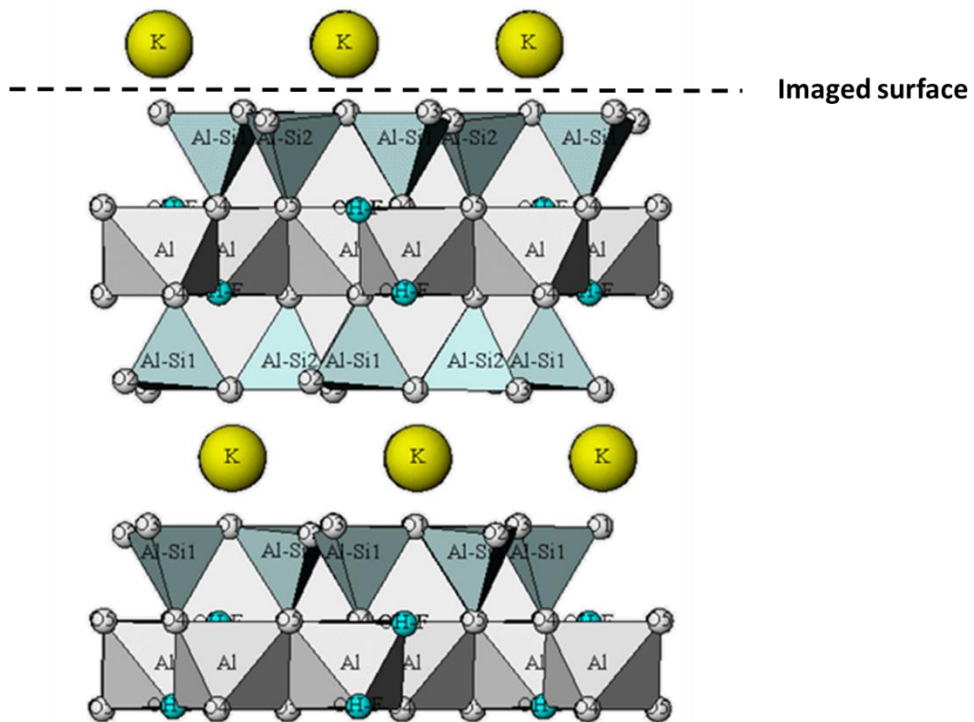
The development of high speed AFMs has meant that it is possible to study dynamics of biological processes in real time rather than using time lapse methods as well aiding in the collection of large data sets over shorter time periods in ambient and liquid studies.

## 3.2 Imaging DNA and proteins

### 3.2.1 Imaging substrates

As AFM is a surface scanning technique, samples must be adsorbed onto a surface to be imaged. The substrate used must conform to a number of criteria. The sensitive nature of AFM means that the substrate must have a low surface roughness to ensure that features of the sample are not occluded from the tip and so that samples can be easily distinguished from the surface. It is also necessary that the substrate has a strong enough interaction with the sample so that the sample is readily adsorbed onto the surface from solution and once adsorbed is held strongly enough that it will be detected by the tip and not dislodged by the motion of the tip.

The most commonly used substrate in the case of DNA and proteins is mica [226, 229, 263]. Mica is the collective name of a number of silicate based minerals. Commonly used in AFM is muscovite (ruby) mica. The chemical structure of muscovite mica is  $\text{KAl}_2(\text{AlSi}_3\text{O}_{10})(\text{OH})_2$ . Muscovite mica has a layered structure which is easily cleaved with adhesive tape, to leave a near atomically flat and clean surface, making it an ideal substrate for AFM studies (Figure 3-5)

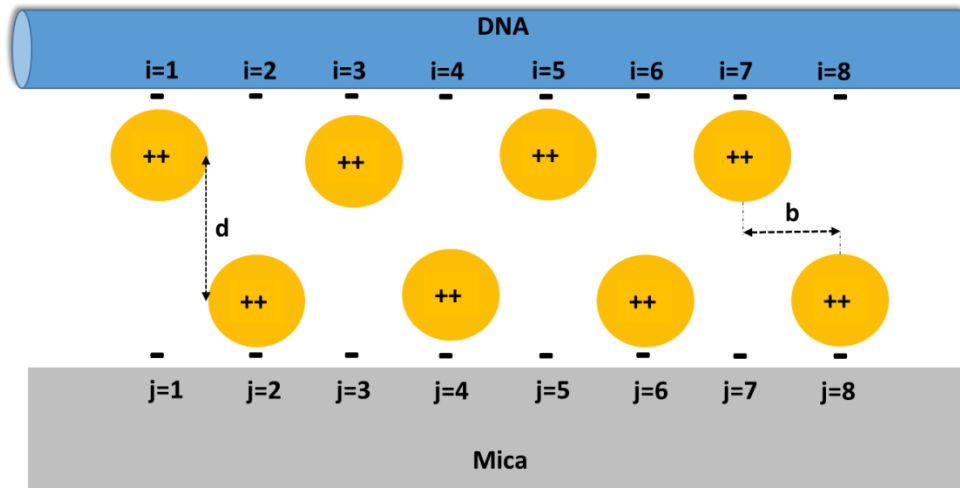


**Figure 3-5: Model of muscovite mica layered structure. The arrangement of the silicate layers can be seen with the interchelating  $K^+$  ions holding two layers together. The top surface has a dashed line to indicate the surface imaged, the  $K^+$  ions are free to dissociate from the surface and form a counter ion cloud when mica is submerged in solution.**

In the case of DNA there is an issue due to the presence of potassium ions ( $K^+$ ) which ionically bond together the mica layers. When freshly cleaved mica is immersed in solution the  $K^+$  ions dissociate from the lattice and form a counter ion cloud above the surface leaving the mica surface with a net negative charge [264]. Due to the negative charge of the DNA phosphate backbone there is no longer an attraction between the DNA and mica meaning that very few DNA molecules are adsorbed onto the surface.

This effect can be overcome by introducing cations into the system, either in the buffer the sample is deposited in or by pre-treating the mica, where, in the case of DNA, divalent cations are preferred [265-267]. These cations replace the  $K^+$  ions on the mica surface leaving it positively charged and form a diffuse layer of positive ions above the surface, neutralising the negative charge of the DNA phosphate backbone, thereby

allowing the DNA to adsorb to the mica [268]. Pastre *et al.* modelled the binding of DNA to the mica surface and proposed a model by which the counter ions bind the DNA in staggered configuration, allowing the cations to be shared between the DNA and the mica (Figure 3-6).



**Figure 3-6: Counter ion correlation mediated between mica and DNA. Divalent counter ions are shown in yellow and can be seen to bind on alternating sites on the DNA and mica, therefore adopting a staggered arrangement. The vertical spacing between two cations is shown as  $d$  and the spacing between binding sites of the cations is shown by  $b$ .**

Pastre *et al.* were also able to show that lateral diffusion of the DNA on the surface is inhibited by frictional forces arising from the electrostatic interactions between the two surfaces [269]. For the DNA to move laterally it requires the divalent cations to jump from one position to the next, this jumping has a large energy barrier and so is not likely to occur [268-270].

Different cations have been shown to allow for varying degrees of binding DNA. The most effective are the divalent transition metal cations, especially  $\text{Ni}^{2+}$ ,  $\text{Co}^{2+}$  and  $\text{Zn}^{2+}$  [265, 271].  $\text{Mg}^{2+}$  is also able to bind DNA to mica but due to the weak nature of its binding is only useable in ambient imaging conditions. As shown by Hansma and Laney, there is a critical ionic radius for binding of DNA to the mica of  $0.82\text{\AA}$  [265]. These transition metals have ionic radii below this value and so are able to fill the spaces



above the mica's hydroxyl groups, binding DNA tightly to the surface. In the case of  $Mg^{2+}$  the ionic radius is small enough but it has a much lower enthalpy of hydration in comparison to the three transition metals. This low hydration enthalpy means that the  $Mg^{2+}$  is not able to form as strong bonds with the DNA or mica [265]. It is also noted that transition metals are able to bind the major and minor grooves of DNA allowing for tighter binding [272-274] whereas  $Mg^{2+}$  binds the backbone through non-specific electrostatic interactions which are much weaker. The strong binding cations can result in changes to DNA conformation whereas  $Mg^{2+}$  does not have as drastic effect. In general for investigations in liquid  $Ni^{2+}$  is the preferred ion for binding DNA and proteins as it provides a tight and secure binding. In ambient conditions  $Mg^{2+}$  is the most common.

### 3.2.2 Preparation of samples

The methods of preparation for DNA and DNA-protein complexes for analysis by AFM tend to follow a very similar process, with a few differences between preparation for ambient and liquid imaging. A general flow diagram for sample preparation is provided in Figure 3-7.

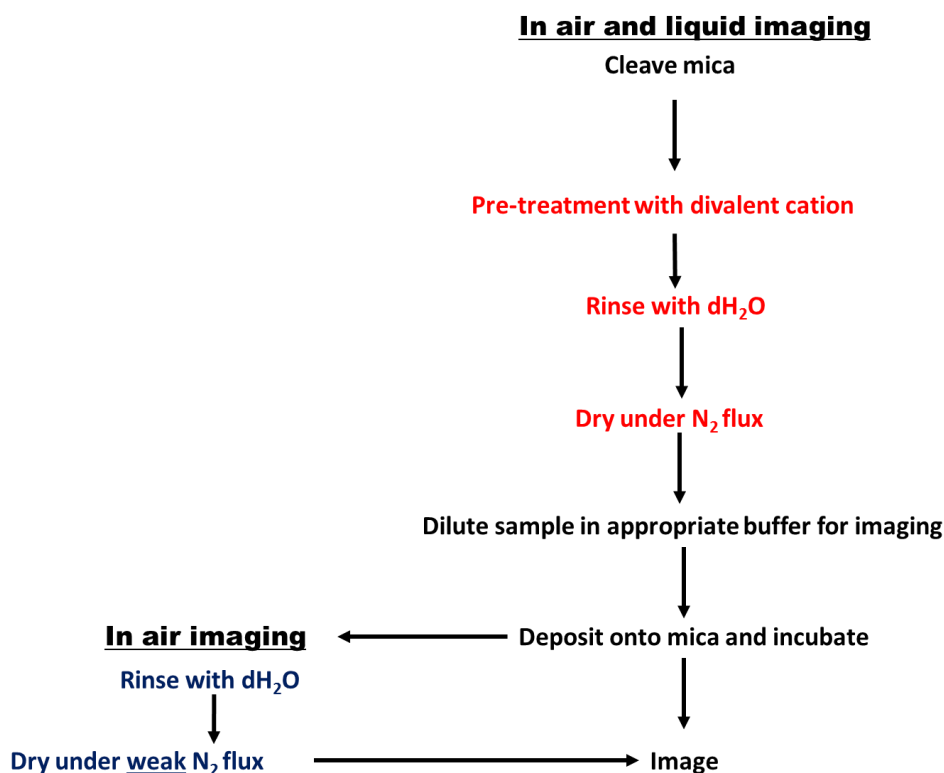


Figure 3-7: Flow diagram of general sample preparation for AFM imaging of DNA-protein samples. The steps shown in red are not required and are optional steps

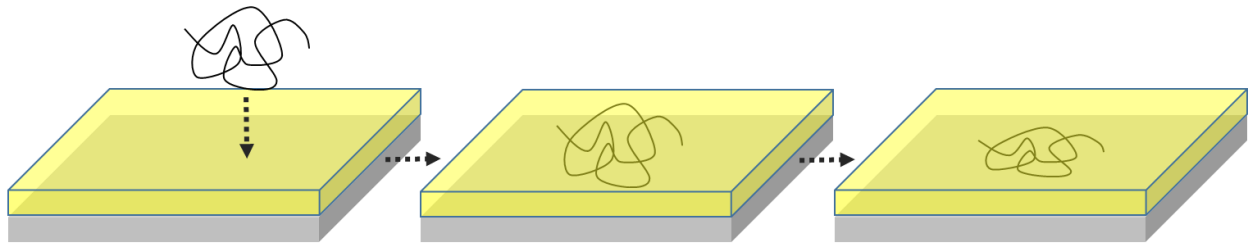
**depending on the sample being imaged. Those steps shown in blue are used solely for imaging in ambient conditions.**

The majority of proteins do not greatly affect the deposition kinetics of the DNA as long as they are at relatively low concentrations (<50nM) [226]. A large number of proteins readily adsorb to mica in a much broader range of buffers than DNA and as long as the protein in question is pure (>80%) and free from such additives as BSA, large amounts of detergent or highly charged molecules, then there is often no issue [215]. Deposition of DNA onto the surface has been shown to occur in two stages. Firstly, the DNA must transport to the surface, secondly the DNA must adsorb to the surface. Lang and Coates were able to show by electron microscopy that diffusion dominated the process of the DNA leaving solution in dilute solutions to reach the surface [275]. It was proposed that the number of molecules at a given time bound to the surface was given by the equation;

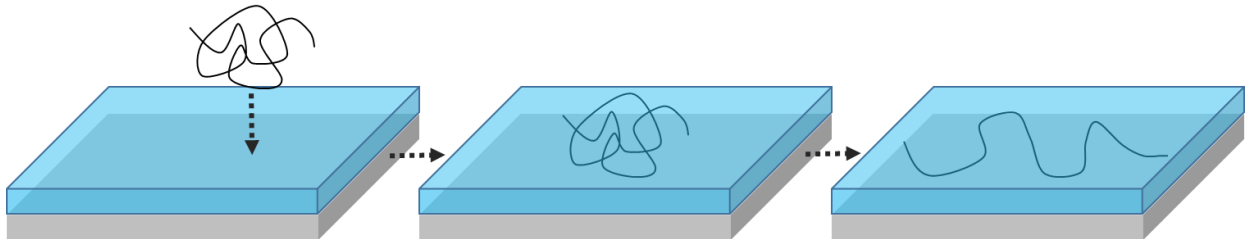
$$\frac{n_F(t)}{n_0} = \sqrt{\frac{4D}{\pi}} \sqrt{t}$$

Where  $n_F(t)$  is the number of molecules on the surface at time  $t$ ,  $n_0$  is the total number of molecules in solution at  $t$  zero and  $D$  is the diffusion constant. This was confirmed by Rivetti, who used AFM to show that for small aliquots of DNA on mica, incubated for up to 30 minutes also obeyed the same behavior [276]. One important consideration when using AFM is whether the surface population is representative of the population in solution. Rivetti investigated the conformations adopted by DNA when deposited onto a mica surface and observed two main modes of binding for DNA, surface equilibrated and kinetically trapped (Figure 3-8) [277]. When the DNA is kinetically trapped it is held by strong short range interactions. Once a segment of the DNA is bound to the surface it is effectively pinned at these points leading to the collapse of the rest of the DNA chain, providing a 2D projection of the DNA's 3D structure.

**A) Kinetic trapping**



**B) Surface equilibrated**



**Figure 3-8: Diagram depicting the two mechanisms by which DNA adsorbed to a mica surface. A) shows the process of kinetic trapping, where the DNA lies on the surface in a conformation that is a 2D projection of its 3D shape in solution. B) In the case of surface equilibrated binding the DNA has some lateral movement allowing it to lie on the surface in its lowest energy conformation.**

Kinetic trapping is most often seen when the surface has been pre-treated with transition metal divalent cations or they are present in deposition buffer, such as  $\text{Ni}^{2+}$ , as these provide the strong bonding interaction needed for the DNA to adopt this conformation [278]. In the case of surface equilibration, the DNA is able to move laterally on the surface to adopt its minimum energy conformation. Surface equilibration occurs through weak long range interactions and is often adopted by DNA when  $\text{Mg}^{2+}$  is used in the deposition buffer [279]. It is possible to determine the mode of binding by accessing the mean squared end-to-end distance ( $\langle R^2 \rangle$ ) which when the DNA is in a trapped conformation is one third of that measured in an equilibrated state [276, 280]. The different binding conformations can provide different advantages depending on the system being studied. Kinetic trapping can provide an observation of the solution conformations of molecules, whereas equilibrated enables a clearer

view of the molecule, allowing for detailed analysis of contour lengths and bends in the DNA to be accessed.

Recent studies have shown that it is possible for both binding conformations to occur and that this can be fine-tuned. Billingsley *et al.* proposed that a mica surface pre-treated with low levels of  $\text{Ni}^{2+}$  has a patchy structure [281]. When short fragments of DNA bind the surface they can show binding heterogeneity whether they fall in a  $\text{Ni}^{2+}$  patch or not [281]. Lee *et al.* used this finding to show that it is possible to alter the strength of binding of DNA in liquid by increasing or decreasing the concentration of  $\text{Ni}^{2+}$  to pre-treat the mica surface [282].

Once the DNA has adsorbed to the surface, the sample is rinsed and dried for imaging in ambient conditions. Typically this would not be the case if imaging in a liquid environment. The incubation time is decided by weighing the number of molecules absorbed to a surface against the formation of salt crystals which can form if slow evaporation of the buffer occurs [215, 226, 229]. The sample is then rinsed to remove any unbound molecules. Drying is then performed under a weak flux of gas such as nitrogen to remove any bulk liquid from the surface and in some cases desiccated. In some cases desiccation of the sample can also lead to the formation of salt crystals or the condensation of DNA and is not always performed [211, 215, 216, 226].

In ambient conditions samples are imaged and contour lengths, bend angles and spatial arrangements of DNA and proteins are analysable. There are a number of DNA tracing software available and each has a number of advantages and disadvantages associated with it, but hand traced measurements tend to be more suited for analysing DNA-protein complexes as algorithms can have issues with height and direction changes in the DNA when bound by proteins [283-287]. Both Sanchez-Savilla and Rivetti performed in depth studies of the accuracy of DNA contour length measurements giving a value of 0.29-0.33 nm per base pair for DNA fragments ranging from 500-4300 bp in length [288, 289]. The range of measurements was believed to be due to transitions of the DNA from B to A form in certain deposition conditions, but are in good agreement with X-ray crystallography data that gives a base pair rise of 0.33-0.44 nm for B-form DNA. It is also possible to analyse the volumes of objects such as

proteins to provide insight into structure of multimeric proteins and binding events [290].

Working in liquid does not offer the chance to perform in depth analysis of contour lengths and structural relationship of proteins but provides real-time information on binding and movement of proteins and DNA in physiologically relevant buffers. Studies can be performed using a time lapse approach, altering reaction conditions to allow for imaging. With the development of high speed AFM technology, time lapse experiments are becoming less common. It should be noted though that association and dissociation rates can be altered by surface and tip interactions [291, 292].

### **3.3 Studying transcription by AFM**

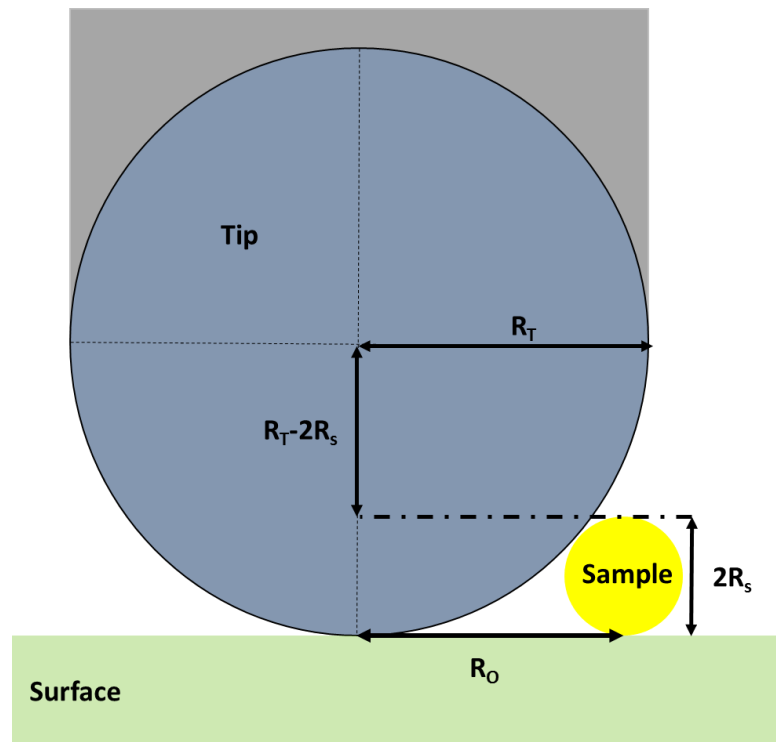
The single molecule nature of AFM enables the study of sub-populations that are often obscured by bulk methods, as previously mentioned. As each molecule is analysed separately to then provide a distribution of data, it is often beneficial to divide molecules in sub-categories or classes [215]. This can be done by looking at each molecule and seeing whether it meets certain conditions in order to categorise. For example, in the case of transcription the three stages can be visualised separately in ambient conditions. The formation of OPCs can then be accessed by measuring the spatial organisation of RNAPs, the wrapping of the DNA and the contraction of the DNA [3, 126, 293]. Those that do not meet all these criteria can be classed as being bound in a non-specific manner. OPCs can then be induced to form either elongated complexes (ECs) or stalled elongated complexes (SECs) by introducing NTPs into the reaction mix before deposition. In the case of SECs again the position of RNAPs on the template can be analysed and complexes that display more RNAPs bound than expected can be discounted or separately classed [124, 294], this can also be applied to those complexes where the DNA contains a specific termination site. This classification of molecules does mean that some information may be lost but at the same time insight into specific outcomes can be garnered. This method of classing complexes allowed Billingsley and Crampton to specifically study the outcomes of transcriptional collisions from convergent promoters, specifically focusing on those complexes that had two RNAPs bound after elongation [4, 5, 194]. In liquid such

processes like the promoter search undertaken by RNAP can be studied, with information on the mechanism utilised and rates of association and dissociation being obtained [295, 296]. The process of active elongation can also be studied allowing for translocation events to be studied for a single RNAP [124, 297, 298].

### 3.3.1 Practical considerations for studying biological samples by AFM

Both *ex situ* (ambient) and *in situ* (liquid) imaging have a number of practicalities that have to be considered. Even though preparation of transcription reactions is relatively simple and often similar in many studies, the process can be inconsistent at times [215, 216, 226, 299]. Due to this the introduction of new species into a reaction mix are often avoided, especially for molecules that can form aggregates or have high charge density as these can alter the binding behavior of complexes. This has led to many AFM studies of transcription not including non-specific binding inhibitors such as heparin, heparan sulphate (HS) (discussed in detail in Chapter 6) or bovine serum albumin (BSA), meaning that many complexes seen must be discounted. This factor is a part of the motivation behind studies presented in this thesis.

Another issue that is common in AFM studies of DNA and proteins is that of tip convolution, i.e. probe broadening effect. This is an effect that occurs when features have a size smaller than the radius of curvature of the AFM tip. The effect leads to an increase in the size of features seen, due to the cantilever being deflected upwards before the lowest point of the tip reaches the sample (Figure 3-9).



**Figure 3-9: Diagram showing the effect of tip convolution on imaged sample size. The tip and sample are modelled as spheres. The diameter ( $2R_s$ ) of the sample and the radius of the tip ( $R_T$ ) mean that the tip senses the sample before it reaches the apex of the tip. This leads to the tip being deflected and so gives an image that is broader than the actual size of the sample.**

This amount of broadening can be given by the equation;

$$R_0 = 2\sqrt{R_T R_s}$$

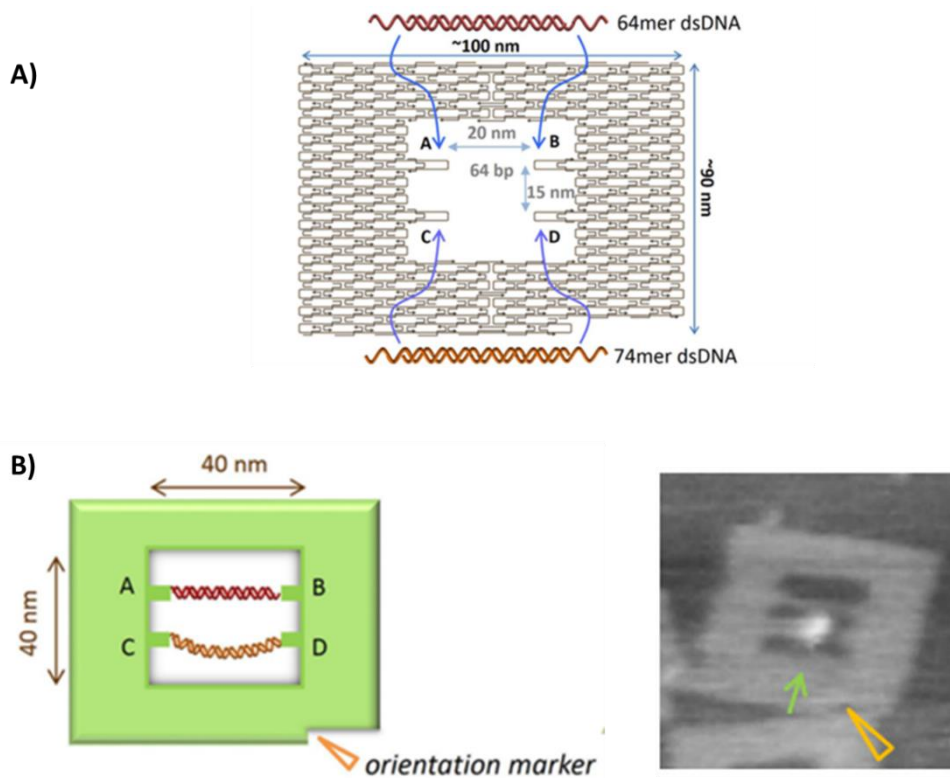
Where  $R_0$  is the observed radius,  $R_T$  is the tip radius and  $R_s$  is the radius of the sample. This effect means that the study of multiple proteins can be difficult if in close proximity or close in molecular weights as they may not be distinguishable. To overcome such issues, the combining of AFM with fluorescence based microscopy has been developed. Examples include the combination of total internal reflection microscopy (TIRFm), forester resonance energy transfer (FRET) and confocal microscopy, allowing two proteins to be distinguished as well as giving better time resolution [300]. An example of which is shown by Sanchez *et al.* who investigated Rad54 and Rad51 interactions on DNA [301, 302]. Another tip induced issue is that of sample deformation, even though forces are low it has been reported that DNA and proteins on the surface are not their expected heights with DNA often having a height of 0.2-0.6 nm even though its

### Chapter 3: AFM of DNA and proteins

diameter is 2nm [303, 304]. These effects can hinder volumetric measurements, but Fuentes-Perez *et al.* were able to demonstrate a method of using dsDNA as fiducial marker for calibrating volume measurements [290]. Methods for correcting for these anomalies using models of the tip geometry and interactions with the water layer have also been used [305, 306].

When imaging *in situ* there is a need to ensure that the DNA is mobile enough for the RNAP or protein to associate and bind tight enough to the support surface to allow imaging. Some groups have achieved this by using liquid cells that allow for the exchange of buffers, using a transcription buffer to allow the DNA to be only partly bound enabling binding of RNAP, then exchanging for an imaging buffer to secure complexes to the surface [279]. In recent years there has been the development of novel approaches to this issue. DNA origami developed by Rothemund [307] has found a number of uses, one of which is its use as a platform for DNA templates and proteins in AFM studies [308]. DNA origami is a method for forming 3D structures by utilising a long ssDNA folded and held in a specific shape by smaller single stranded “staple” strands [307]. It can be used to form either a platform where DNA can be attached, meaning the DNA is held in place while not being tightly bound to the support surface as was shown by Endo *et al.* when investigating T7 RNAP [105]. DNA origami can also be used to make DNA “frames” (Figure 3-10) where a void is left in the centre allowing for a DNA fragment to be bound across this void. This method was used by Yamamoto *et al.* to investigate the binding of the proteins Sox 2 and Pax 6 [309]





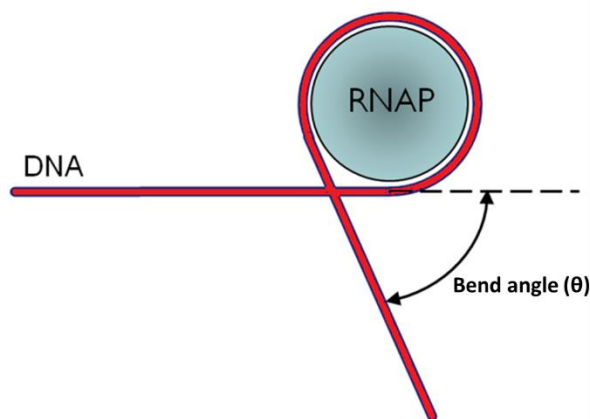
**Figure 3-10 : Diagram and AFM image of a DNA origami frame. A) Shown is the computer design of the origami frame with the void empty. The DNA template strands to bind to the frame are shown in red and orange. These attach via single stranded regions at the end of the DNA. B) A schematic diagram of the frame with the template DNAs bound in the void is shown, alongside an AFM image of the frame with a protein bound. A notch is left at one corner of the DNA origami (highlighted by orange triangle) so the orientation of the tile can be discerned when imaging ( adapted from [309])**

The use of frames and tiles is still a relatively new approach but provides opportunity when combined with high speed AFM for an increase in events of protein binding and translocation analysable in real time. Thomson *et al.* also utilised a method that involved the functionalisation of a gold surface with alkane thiols [297]. Some of these alkane thiols were able to bind to poly histidine-tagged RNAP molecules, allowing for the protein to be orientated on the surface. It was shown using a single stranded DNA template that the RNAP was active on the surface but the RNA transcript could not be imaged directly under liquid. This method may be an option to utilise with high speed AFM for imaging of immobilised RNAP during transcription of dsDNA templates.

### 3.3.2 Previous AFM studies of transcription

Some of the first studies of transcription by AFM visualised the promoter search mechanism. Guthold and Bustamante used *in situ* imaging to observe the rates of  $\sigma$ RNAP movement when undergoing a promoter search [106, 295]. Guthold used a promoter-less DNA template and time-lapse imaging to observe the  $\sigma$ RNAP sliding on the template [295, 296]. This sliding displayed a diffusion distance proportional to the square root of time [296]. This indicated that  $\sigma$ RNAP undergoes a 1D diffusion on the DNA. Bustamante also used time lapse imaging and was able to observe the  $\sigma$ RNAP undergoing what was believed to be hopping and intersegmental transfer as well as sliding [106]. This behavior of the RNAP was later confirmed by Suzuki *et al.* who used high speed AFM to visualise these events in real time also seeing a combination of sliding, hopping and intersegmental transfer [310].

The formation of OPCs by *ex situ* AFM has been investigated for the *E.coli*  $\sigma^{70}$ RNAP as well as for the  $\sigma^{54}$ RNAP. In the case of the  $\sigma^{70}$  RNAP both Rees *et al.* and Rivetti *et al.* noted that complexes often adopted a bent shape [124, 126]. The bend angle of the DNA template (Figure 3-11) was measured and it was found that in the case of the  $\lambda_{pL}$  promoter the DNA had a mean bend angle of  $54^\circ$  while for the  $\lambda_{pR}$  promoter the bend angle was between  $55^\circ$  and  $88^\circ$  [124, 126]



**Figure 3-11: Schematic representation of the wrapping of DNA around RNAP. The DNA is shown in red, wrapped by a full turn around the RNAP (blue). The total length of wrapping corresponded to approximately 90 bp (30nm). The angle measured is shown as the bend angle ( $\theta$ ). Adapted from [4].**

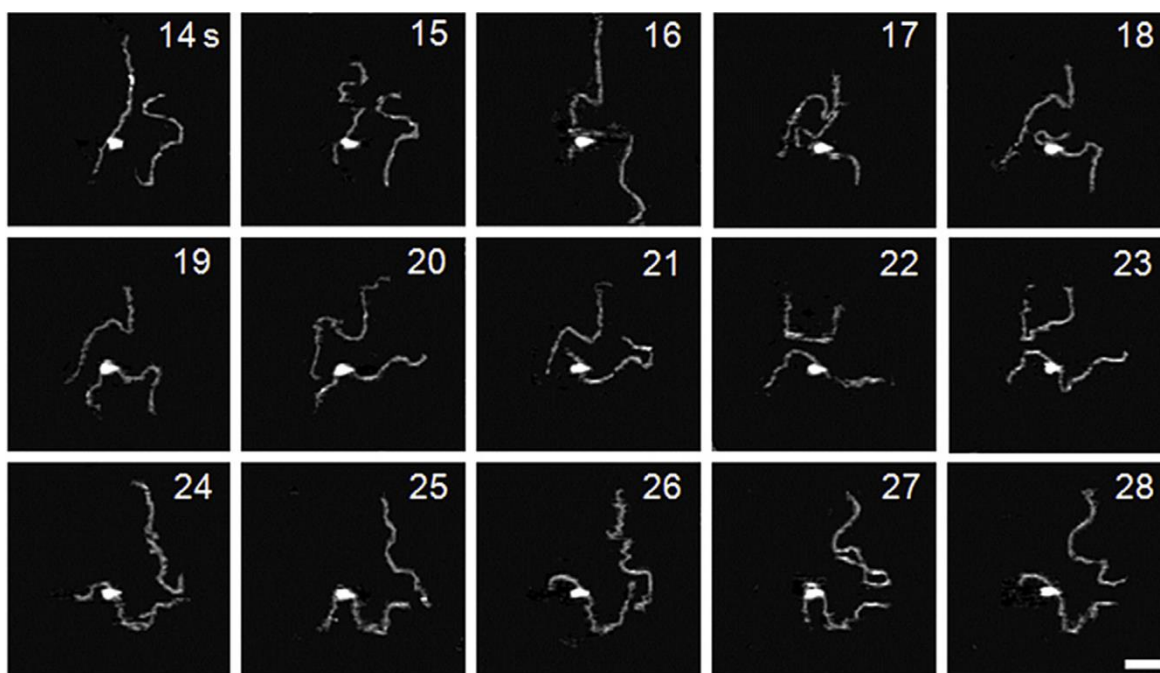
Rivetti *et al.* also showed that the contour length of the DNA decreased by  $\sim 30$  nm which is equivalent to 90 bp, combined with the bending of the DNA it was concluded that the DNA wraps the  $\sigma^{70}$ RNAP with this outcome being supported by DNA footprinting data and FRET based studies [311, 312]. Studies by Cellai *et al.* were able to show that this wrapping was mediated by the  $\alpha$ -subunits carboxyl terminal domain: with sequential removal of this domain the degree of wrapping decreased [129]. This idea was previously suggested by Mangiarotti *et al.* when investigating the binding and wrapping of two different promoters in close proximity indicating that closely spaced promoters can be linked and that the  $\alpha$ -subunits contact with the UP-elements of a promoter play a role in forming fully wrapped complexes [127]. Doniselli *et al.* showed that wrapping is also affected by the stringent response modulator of *E. coli* guanosine tetraphosphate (ppGpp) [313]. It was shown that the ppGpp was able to allosterically prevent conformational changes that lead to stabilisation of the OPC. *Ex situ* imaging of a DNA template containing multiple *fis* promoters by Gerganova *et al.* revealed that the binding of one  $\sigma$ RNAP at its promoter site can encourage or hinder the binding of subsequent RNAPs, believed to be due to changes in topology of the template upon OPC formation [314].

For  $\sigma^{54}$ RNAP the formation of OPCs was shown to go through an intermediate stage, where the RNAP binds its promoter and loops the DNA to form contacts with an auxiliary factor NtrC, allowing for the activator to also interact with the promoter element [315, 316]. Bending of the DNA was also observed with two distinct bend angles being seen for CPCs and OPCs. A bend angle of  $\sim 49^\circ$  was recorded for CPCs and  $\sim 114^\circ$  for OPCs [316].

Elongation has been visualised by time lapse imaging *in situ* by Kasas *et al.* and Guthold *et al.*, using low NTP concentrations to limit the rate of elongation [296, 298]. Kasas deposited SECs onto a mica surface before adding NTPs to the reaction buffer. The arms of the DNA were seen to increase and decrease respectively either side of the RNAP in the direction determined by the promoter, indicating that the DNA was being threaded through the RNAP. A rate of elongation of  $0.5 - 2 \text{ nt s}^{-1}$  was recorded. RNA transcripts could not be visualised but activity was confirmed by using ssDNA template

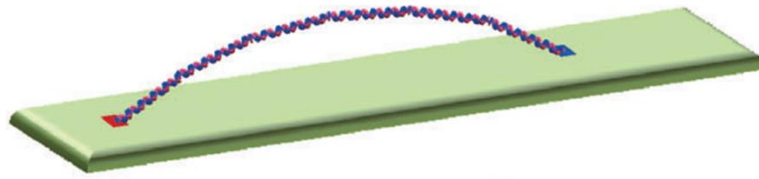
to induce rolling circle transcription at the surface in liquid, before drying the sample and visualising the RNA transcripts. Guthold *et al.* performed similar experiments and recorded a rate of  $1.5 \pm 0.8 \text{ nt s}^{-1}$  [296].

The use of high speed AFM allowed Suzuki *et al.* to visualise all three stages of the transcription cycle for *E.coli* RNAP [310]. In these experiments it was found that the RNAP had an elongation rate of up to  $15 \text{ nt s}^{-1}$  which is relatively similar to rates seen in some biochemical and biophysical experiments [317]. The DNA was again seen to be pulled through the RNAP (Figure 3-12).



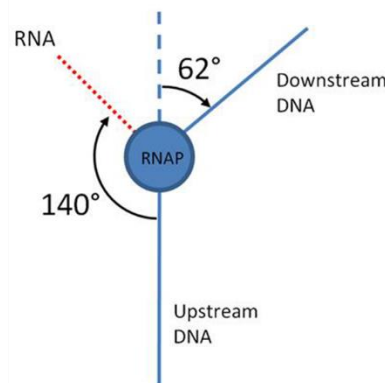
**Figure 3-12: AFM scans of an elongating RNAP collected by Suzuki *et al.* using high speed AFM. The time of each frame is given at the top of each image. The RNAP can be seen to feed the DNA through in a unidirectional fashion after the addition of NTPs, indicating that elongation is occurring [310].**

Endo *et al.* were able to use an origami tile to attach the template to (Figure 3-13) and then visualised the stages of the transcription for T7 RNAP [105]. In this experiment, the elongation rate was not measurable due to the DNA making transient contacts with mica, but RNA products were occasionally seen during *in situ* imaging and also made more visible by the use of a biotinylated UTP which was then labelled with streptavidin, meaning that activity for the specific dsDNA template could be directly confirmed.



**Figure 3-13: Computer drawn schematic of a DNA template attached to an origami tile, as was used by Endo *et al.* The tile is shown in green with tethering points for the template DNA (blue and red helix) shown by the blue and red squares [105].**

*Ex situ* imaging of ECs and SECs has also provided information on structural elements of the process. Rees *et al.* saw that SECs formed by the omission of specific NTPs, had a bend angle larger than that seen for OPCs and this was concluded to be due to loss of contact of the DNA with RNAP and reduced wrapping [124]. This was later confirmed by Rivetti *et al.* who noted that the compaction of the DNA in a SEC was reduced to  $\sim 22\text{nm}$  and by Billingsley *et al.* who noted a reduction of  $\sim 13\text{ nm}$  for SECs [4, 294]. Both Rivetti *et al.* and Billingsley *et al.* were also able to visualise the RNA transcript. In both cases it was seen that the RNA exited the RNAP at angle of  $140^\circ$  from the DNA on the opposing side to the template (Figure 3-14).



**Figure 3-14: Diagram of the arrangement of the DNA template and RNA transcript in an SEC. The bend angle of the DNA can be seen to have an average value of  $62^\circ$  with the RNA exiting from the RNAP at an angle of  $140^\circ$  from the upstream arm of the DNA.**

Billingsley *et al.* also investigated transcription by *E.coli* RNAP from two identical promoters [4]. This allowed the outcomes of simultaneous transcription from two

### Chapter 3: AFM of DNA and proteins

promoters to be studied. It was found for convergent promoters that the RNAPs stalled in close proximity and remained on the template, which was in agreement with observations by Crampton *et al.* and biochemical studies of RNAPII by Hobson *et al.* [3, 4, 318]. Through the use of nucleotide base labelling system, it was possible to show that there were two main outcomes of convergent transcription [5]. Collisions which resulted in one of the RNAPs being pushed back by the other RNAP. This is thought to be the consequence of a collision where one RNAP has failed to commence elongation, alternatively collisions occurred between two active RNAPs resulted in stalled complexes in the region between the two promoters. In the case of tandemly aligned promoters, a label was not used but it was seen that the RNAPs did not reach the end of the template DNA, but stalled in close proximity downstream of both promoters, indicating that the two RNAPs are able to allosterically regulate each other [4].

Multiple RNAPs and RNAP interactions with other proteins have also been investigated by AFM. Ebenstein *et al.* used combined AFM and fluorescence microscopy to distinguish the binding of T7 and *E.coli* RNAPs labelled with quantum dots, to the T7 genome [319]. This provided a new outlook for further developing combined techniques for investigations into transcription by AFM. Horn *et al.* explored RNAP II interactions with nucleosomal proteins [320]. They were able to show through the analysis of height and volume, that when the RNAP transcribed into the nucleosomal proteins the DNA was looped to allow for passing of the RNAP and that the nucleosomal proteins underwent disassembly and rearrangement. This work further confirmed experiments performed by Bintu *et al.* who also investigated these events by AFM [321].

Investigations by AFM into termination of transcription have not been common. In the studies by Suzuki *et al.* and Endo *et al.*, the RNAPs were just reported as dissociating from the templates. Limanskaya and Limanskii imaged T7 RNAPs interacting with termination sites at a reduced reaction temperature of 31°C [322]. This reduced temperature was shown to diminish the disassociation rate of the RNAPs from the template. A number of RNAPs were seen at the terminator sites indicating that multiple rounds of transcription had occurred. It was also seen that some complexes displayed RNAPs at the terminator, at the promoter and bound between the two. This

### *Chapter 3: AFM of DNA and proteins*

shows that it is possible for RNAPs to initiate transcription as soon as the promoter has been cleared by the previous RNAP, explaining the high rates of transcription seen *in vivo*. Kotlajick *et al.* observed Rho-dependent termination and the role H-NS plays by imaging RNAP interactions with H-NS filaments formed on the DNA template [323]. They were able to show that bridged H-NS filaments were able to cause pausing and backtracking of the RNAP. This was seen to occur due to the bridged filaments surrounding the RNAP causing extended pauses which allow for Rho protein action. They suggested that the results promote a theory that transcription driven supercoiling, when constrained by the H-NS protein leads to pausing.

As it can be seen by previous studies, AFM provides a high resolution technique that is able to distinguish between DNA and RNAP. The ability to recognize different time points within the transcription process mean that AFM is well suited to the study of concurrent transcription. By recording contour lengths of the DNA and position of RNAPs to a high accuracy, it is possible to recognize those RNAPs that have formed OPCs as well as study the final resting positions after elongation. This means that it is possible to observe the outcomes of collisions and stalling in concurrent transcription systems. In air the stages of transcription are studied by averaging over a large set of molecules providing both a view of subsets of molecules during OPC formation and after elongation. When operated in liquid either by time lapse methods or at high speed, it is possible to gain real time information on the kinetics and interactions between RNAPs. Other single molecule techniques do not allow for accurate mapping of RNAP position on the DNA for small sub-populations of molecules.

This page is left intentionally blank



## 4 General Methods

### 4.1 Introduction

AFM is a direct imaging technique that can investigate the nature of the interactions of the DNA template and RNAP at different stages throughout the transcription process providing a high degree of accuracy on the spatial arrangement of DNA and proteins [126, 294]. The single molecule nature of AFM provides a different outlook to these systems compared with traditional biochemical methods that usually look at outcomes of events on a bulk scale [213]. AFM allows study of individual outcomes of transcription events and to use each outcome to provide a distribution of results. The AFM probe is sensitive to all features on a surface, because it detects force, and therefore it is important that samples be free of any contaminants that can lead to misinterpretation of data. It is also important that biomolecular samples are securely attached to the support surface due to forces and interactions between the tip and sample that could lead to alterations of complexes. The surface binding must be carried out in a manner that is representative of the complexes formed in solution to provide an accurate interpretation of data [227, 283]. Upon binding to the surface, complexes provide a 2D representation of the 3D structure in solution. This means that in order to study the position of proteins bound to a DNA template, the complexes must be in a conformation that allows for visualisation of the DNA ends and the contour of the DNA backbone. This means that surface bound complexes may not be bound as they occur in solution, but in a manner that allows for interpretation of the 3D organisation of complexes on a 2D surface.

This chapter offers a general overview of the generic techniques and methods used for the preparation of samples that were subsequently imaged using the AFM and begins with the molecular biology techniques used throughout to produce and prepare DNA samples. These samples were subsequently used for *in vitro* transcription reactions. The methods for these are described along with the process for sample deposition, AFM imaging and data analysis are presented. Subsequent chapters will present methods specific to the experiments in those chapters in more detail.

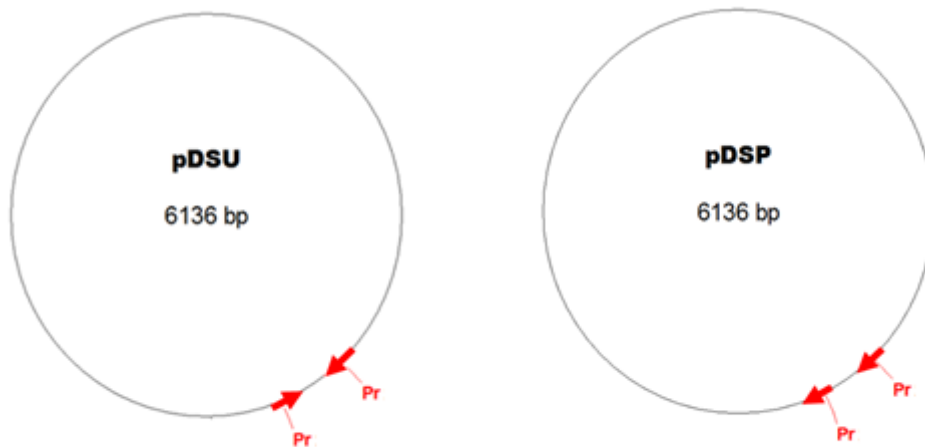
## 4.2 Preparation of DNA constructs

### 4.2.1 Transcription templates

DNA constructs used throughout this thesis were generated from either of two plasmids containing two transcriptional promoter elements. These plasmids were pDSU, which contained two convergent  $\lambda_{pr}$  promoters and pDSP which contained two  $\lambda_{pr}$  promoters in a tandem arrangement (Figure 4-1). The  $\lambda_{pr}$  promoter is a promoter from bacteriophage  $\lambda$  and is able to direct  $\sigma^{70}$  mediated transcription by *E.coli* RNAP and has the sequence:

5' ACCTCTGGCGGTGATAATGGTTGCATGTACTAAGGAGGTTG 3'

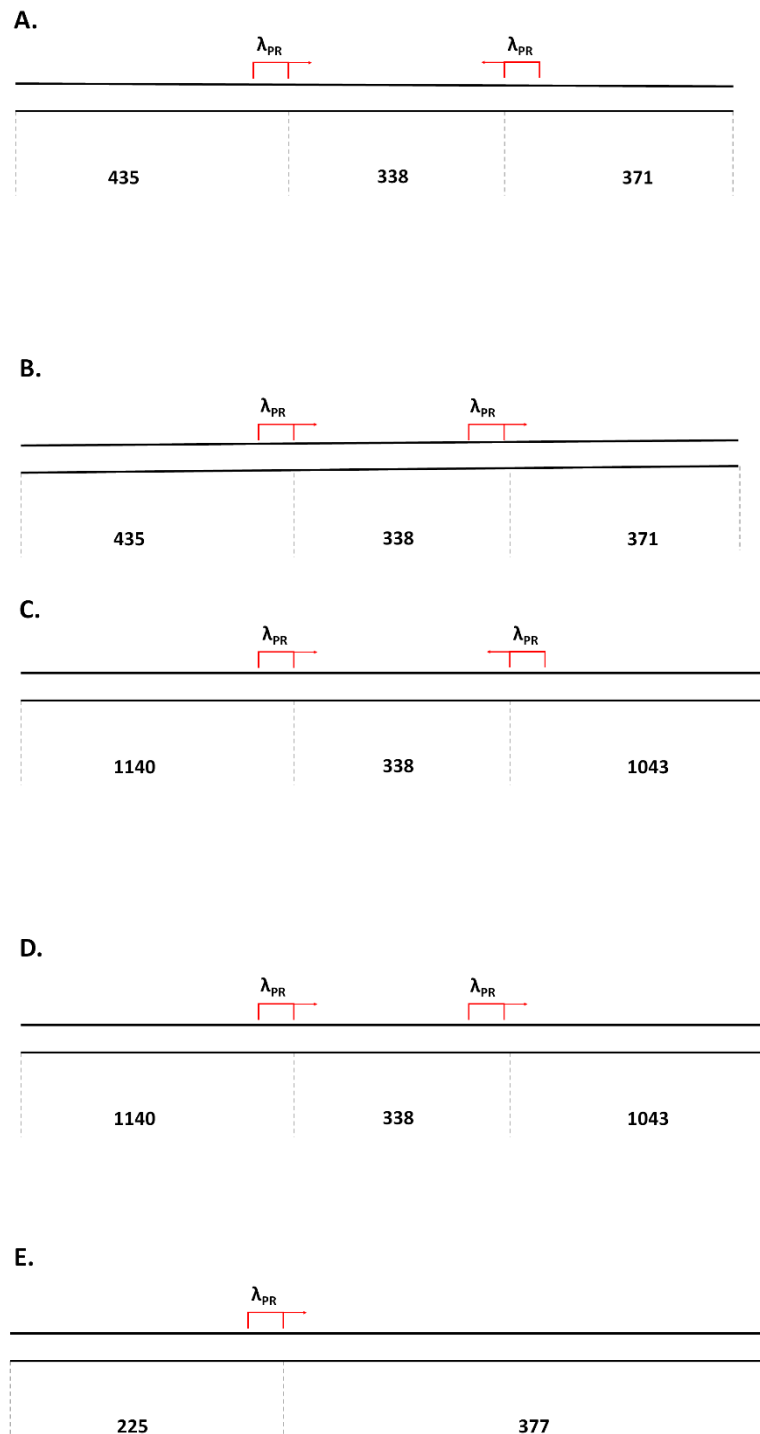
The transcriptional start point is marked by the nucleotide shown in red [324, 325].



**Figure 4-1: Diagrams of the plasmids pDSU and pDSP. Both plasmids are shown with the number of base pairs that make up each plasmid and the positions of the  $\lambda_{pr}$  promoter position and direction of transcription are indicated by the red arrows**

Five different templates were produced by polymerase chain reaction (PCR). These consisted of a template 1144 bp in size and a template 2521 bp produced from both pDSU and pDSP to provide templates with convergent and tandem promoter arrangements. A template of 602 bp was also produced from pDSP containing only a single promoter. Each of these templates has different numbers of base pairs either side of the promoter or promoters while having the same number of base pairs, 338 bp, between the two promoters when applicable.

Chapter 4: General methods



**Figure 4-2: Schematics of the templates used. A and B are the 1144 bp templates with convergent and tandem promoters. C and D are the 2521 bp templates with convergent and tandem promoters. E is the 602 bp template derived from pDSP with a single promoter. Each template has the promoters indicated by the red boxes with arrows to indicate the direction of transcription. The lengths of each section are given from the transcription start point in bps.**

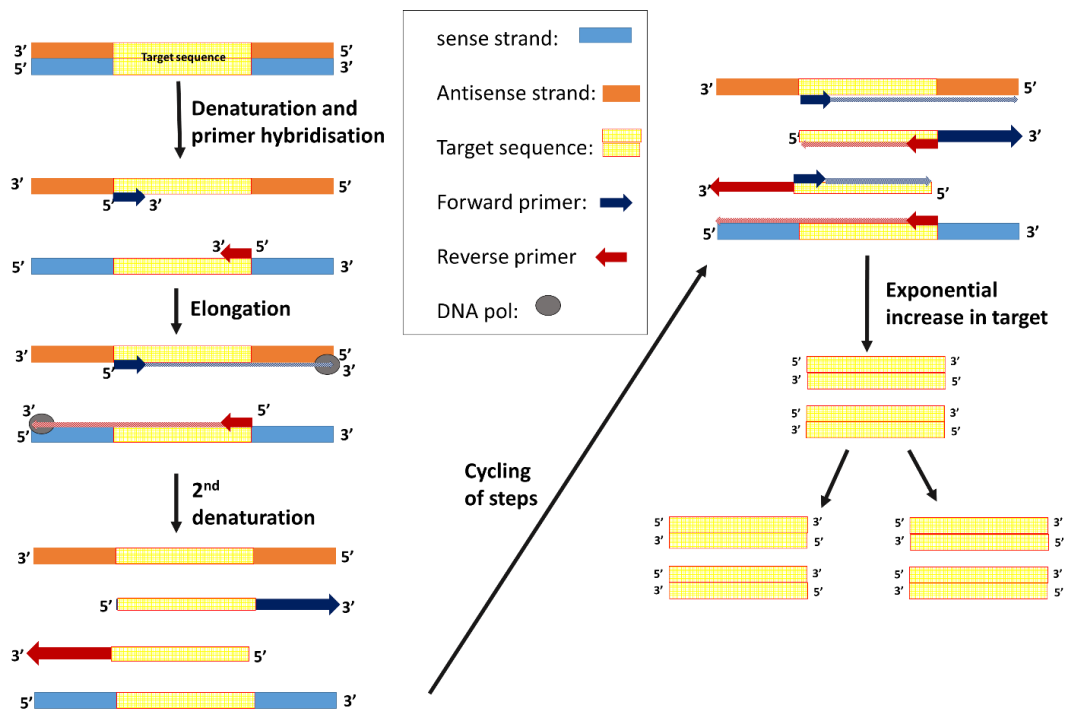
#### *Chapter 4: General methods*

The asymmetric lengths of the arms allows for the templates orientation to be determined under AFM analysis when open promoter complexes (OPCs) have been formed. The arms of each template are measured from the transcriptional start site of each promoter.

#### **4.2.2 Polymerase chain reaction**

A number of different DNA templates were used in this thesis. These templates of DNA were produced using a method known as the Polymerase Chain Reaction (PCR). PCR is common microbiology technique being used in a number of fields such as forensics to phenotyping [326]. The discoverer of the method, Kary Mullis in the 1983 was awarded the Nobel Prize due to its importance and usefulness [327, 328]. The process is able to take a single double stranded DNA (dsDNA) molecule and produce a huge number of copies of the whole molecule or a region of interest in a test tube using a relatively simple procedure by mimicking DNA replication that occurs in nature [329]. The major development of PCR since its discovery is the use of thermostable *Taq* DNA polymerases by Saiki in 1988 [330, 331]. Found in bacteria populating hot thermal springs, the *Taq* polymerase is able to withstand higher temperatures making the PCR process more efficient and simpler as it could be run in a single reaction vessel [330]. Initially the process involved the use *E.coli* DNA polymerase which denatured during the heating steps and therefore needed replacing after each cycle [332, 333].

Chapter 4: General methods



**Figure 4-3: Diagram detailing the steps involved in a PCR. The dsDNA is shown as two strands in orange and blue, with the target sequence shown in hatched yellow.**

**The parent strand is denatured and the forward and reverse primers anneal (hybridise) to their complementary sequences. Once annealed DNA pol produces a complementary strand of DNA in the 5'-3' direction. The steps of denaturation, annealing and elongation are then cycled to allow for exponential increase in the target DNA.**

A PCR reaction using a heat stable *Taq* polymerase begins by mixing a small amount of dsDNA to act as a parent molecule (shown by orange and blue boxes in Figure 4-3), *Taq* polymerase, the four deoxynucleotide triphosphate (dNTPs) monomers and two short oligonucleotide sequences referred to as primers, in a stabilising buffer containing a divalent ion (usually  $Mg^{2+}$ ). The reaction mix is initially heated to a high temperature to denature the parent DNA into two separate strands (sense and anti-sense) as is shown in the first step of Figure 4-3. There also maybe a prior high temperature heating step if using a hot-start *Taq* polymerase which is conjugated to an antibody to prevent activity until heated [334]. The denaturation into two single stands allows for the primers to hybridise to their complementary sequences. One primer hybridises to the 3' end of the target sequence on the sense strand (shown by the blue arrow in Figure 4-3) acting as the reverse-primer and the other binds to the 3' of the anti-sense

#### Chapter 4: General methods

strand acting as a forward primer (red arrow in Figure 4-3). The sequence between the two primers is the target sequence to be amplified. The annealing of the primers is driven by cooling the reaction mix to a temperature 3-5°C below the melting temperature of the primers. The temperature has to be carefully selected to allow for stable but specific binding of the target and is dependent on the primers and their sequence. Once the primers have hybridised, the temperature is raised to allow the DNA polymerase to produce a complementary strand of DNA in a 5' to 3' direction using the primer to direct the production as is shown in the third step of Figure 4-3 by the hatched arrows [330]. For *Taq* polymerase the optimum temperature is 75-80°C but a temperature of 72°C is used for standard *Taq* polymerase, but this is dependent on the specific *Taq* being used as these can vary between manufactures [332, 335, 336]. The three steps of denaturation, hybridisation and elongation are cycled between 20-40 cycles to allow exponential amplification of the target DNA. A final extended elongation step is used to ensure completion of all complementary strand synthesis.

This process enables the amplification of DNA targets of interest to a high concentration that can be used in subsequent protocols and reactions. Throughout this thesis DNA templates were produced using GoTaq Hot Start Polymerase (Promega, Madison, WI) in 50µL reactions as per the manufacturer's instructions unless stated otherwise. The use of alternative PCR reagents and protocols is given in more detail in subsequent chapters.

### 4.2.3 Gel electrophoresis

Gel electrophoresis is a routine technique that can provide a rapid method for checking the outcome of PCRs and other enzymatic reactions such as restriction digests and ligations. The method provides a rapid sizing of all DNA fragments within a reaction mix.

Gel electrophoresis utilises a gel substance such as agarose or polyacrylamide to form a matrix [337]. This matrix acts as a molecular sieve to separate out molecules of different size. Smaller molecules move quicker through the gel than larger molecules meaning they can be separated from each other by size [338]. When placed in an

#### *Chapter 4: General methods*

electric field the negative charge of the phosphate back bone means the DNA is attracted to the positive electrode and migrates through the gel [337, 338]. The gel is submerged in a buffer that can keep the pH stable as well as provide an ionic field when a current is applied. Fragments can be visualised after separation by casting the gel with an interchelator, post-staining the gel or loading samples with an interchelating agent. Interchelators are able to stack between the base pairs of the DNA double helix. The selected interchelators show fluorescence under UV-B light [339]. This enables the visualisation of the DNA within the gel as defined bands which can then subsequently be imaged using a camera. Run alongside the test samples is set of DNA fragments of known size, known as a DNA ladder[340]. These fragments allow for reference markers for determining the size of samples [340].

Gels used were formed by dissolving 1 g of agarose in 100 ml Tris-acetate EDTA (TAE) buffer to give final concentration of 1% (w/v) agarose. 50 ml of the still warm gel was poured into a BioRad mini (BioRad, Hercules CA) gel tank gel tray (5 cm x 5 cm) and if used, 1 $\mu$ L of ethidium bromide at a concentration of 10mg/ml was added and mixed into the gel. The gels were left to set before being submerged in TAE buffer up to 1-2mm over the top of the gel. DNA samples were mixed with either blue/orange gel loading dye (Promega, Madison WI) or GelRed loading dye (Biotium, Hayward, CA) using 1 part loading dye and 5 parts sample. These loading buffers contained glycerol to ensure samples settled in the wells. The loading buffers also contained dye to allow tracking of the migration of samples. The tank was then sealed and a constant voltage of 80 V was applied to the gel. Once the loading dye had migrated through 70% of the gel, it was removed from the tank and visualised using an InGenius gel documentation system (Sygene, Cambridge UK).

In order to study RNA transcripts produced formaldehyde agarose gels were used. Formaldehyde was chosen due to its denaturing effects on single stranded RNA, which is able to readily form secondary structures which may lead to it not running uniformly through the gel. To form a 1.5% gel, 1.5% (w/v) agarose was dissolved into formaldehyde agarose gel buffer (for 10x buffer, 200mM 3-[N-morpholino] propanesulfonic acid, 50mM sodium acetate, 10mM EDTA, pH 7) heated and cooled and then 37% (v/v) formaldehyde was added along with 0.01mg of ethidium bromide

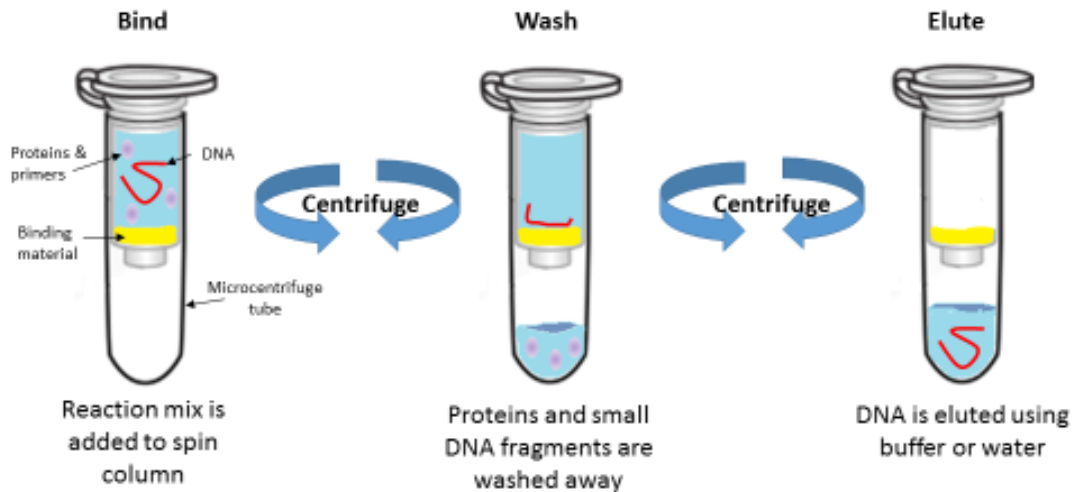
#### *Chapter 4: General methods*

or GelRed (Biotium, Hayward CA). The gel was then left to set in a BioRad mini gel casting tray (5 cm x 5 cm). Once set, the gel was placed into a tank and submerged in formaldehyde gel running buffer (for 1 liter of 1x buffer, 100ml 10x Formaldehyde agarose gel buffer, 20ml 37% (v/v) formaldehyde, 880ml RNase-free water). Samples to be analysed were then mixed with RNA sample loading buffer (Sigma Aldrich, St Louis MO) and heated to 65°C and chilled on ice before being loaded onto the gel. The gel was then run at a constant voltage of 50V until the marker dye within the loading buffer had migrated through 70% of the gel. The gel was then visualised using the InGenius gel documentation system (Syngene, Cambridge UK).

#### **4.2.4 Column purification**

For analysis by AFM, DNA samples must be of high purity. Due to the sensitive nature of AFM, impurities from any reactions, enzymatic or PCR, must be removed in order that they do not obscure objects of interest. The DNA must be suspended in a solution that does not have any salts or other compounds that may affect adsorption of the DNA and protein complexes to the AFM support surface, as well as not affecting function of the protein in subsequent reactions. DNA throughout was purified using one of two purification systems depending on the source material. For PCR and enzymatic reactions DNA species of a single size were purified using Qiagen's QIAquick PCR purification kit (Qiagen Valencia CA). This system removes proteins, nucleotides and DNA shorter than 100bp [341]. For fragments of differing size which were separated on an agarose gel, the Qiagen QIAquick Gel extraction kit was used. This involves an additional step to solubilise and remove agarose from the sample after excision of DNA band from a gel. Both systems utilise a spin column set up as shown in Figure 4-4.





**Figure 4-4: Process of column based DNA purification. The three main steps used in the process are shown. Firstly the DNA is bound to the membrane within the column. The sample is then washed to remove any proteins, nucleotide monomers and oligonucleotides below 100bp. The DNA is then suspended in a new buffer before being eluted from the column.**

Both systems utilise a silica based membrane to bind the DNA in high salt concentration buffers containing for example guanidine hydrochloride and isopropanol. Once the DNA is bound, the membrane is washed and centrifuged in order to remove impurities. The DNA is then eluted from the membrane by adding a low salt buffer (10 mM Tris pH 8.5) or deionised water. The column is spun again to collect the eluted DNA. This method removes unwanted proteins but also leads to the DNA being suspended in a solution that is suitable for AFM analysis and *in vitro* transcription reactions.

#### **4.2.5 Measuring DNA concentration**

In order to use the DNA in further reactions the concentration of the DNA in solution is needed. It is possible to use the intensity of the bands on a gel to determine concentration relative to DNA standards but it is more efficient and accurate to use spectrophotometry. The solution containing the DNA has UV-light of a wavelength of 260 nm passed through it. DNA absorbs light in the UV region at 260nm due to the aromatic nature of the bases [342]. By first using a reference solution that is the same as that which the DNA is suspended in, the change in absorbance from the reference

## Chapter 4: General methods

to the DNA can be used to determine the concentration using the Beer-Lambert law [343]. The concentration in DNA is given in ng/ $\mu$ L and so is independent of the sequence and length of the DNA being analysed.

It is also possible to measure the purity of a sample, as proteins absorb UV light around 280nm, by comparing the ratio of absorbance at  $A_{260}$  to  $A_{280}$  protein contamination is determined. The  $A_{260}/A_{280}$  ratio should be approximately 1.8 for a pure dsDNA sample.

DNA samples in this thesis had their concentration determined using a NanoDrop 2000C (Thermoscientific, Waltham MA). This is a spectrophotometer specifically designed for small sample volumes. The NanoDrop was calibrated using the same buffer as the DNA was suspended in, this was either dH<sub>2</sub>O or Qiagen's elution buffer. One or two microliters of sample was then placed in the light path and measured. Samples with a concentration lower than 5ng/ $\mu$ L or a  $A_{260}/A_{280}$  ratio lower than 1.8 were disposed of as this was below a useful concentration or considered contaminated with protein.

### 4.3 *In vitro* transcription reactions

Transcription reactions were performed in two steps. The reaction was performed in this manner in order to provide confirmation that the polymerase was able to find its promoter and specifically bind before undergoing elongation. The protocol used is based upon previous experiments used in the group and in the literature [126, 293, 296].

#### 4.3.1 *In vitro* transcription: Open promoter complex formation

The formation of open promoter complexes (OPCs) is the process of the  $\sigma$ RNAP locating the promoter sequence, melting the DNA helix and wrapping the DNA around the active site of the protein. The  $\sigma$ RNAP requires an increase of temperature to 37°C to efficiently melt the promoter. The process by which the  $\sigma$ RNAP locates its promoter is not fully understood but is believed to occur through non-specific interactions with the DNA backbone as in previously mentioned in Chapter 1. The sample preparation was

#### *Chapter 4: General methods*

designed to provide time for the RNAP to locate its promoter and melt the DNA in order to form specific interactions.

The given DNA template was mixed with *E.coli*  $\sigma^{70}$ RNAP (Epicentre, Madison MI) in 10 $\mu$ L transcription buffer. The transcription buffer consisted of 20mM Tris-HCl (pH 7.9), 50mM KCl, 5mM MgCl<sub>2</sub>, 1mM DTT. The buffer used differed from that provided by the manufacturer, having no Triton X-100®. This was carried out in order to prevent any interference with the AFM analysis but still be able the protein to perform its function. DTT is added to help prevent aggregation of RNAP molecules as well maintain protein stability by acting as reducing agent. It acts as a reducing agent and reduces any disulphide bond that may occur between cysteine residue side-chains located on the protein surface.

The amount of DNA used in the reaction was 200 fmol of DNA chains for all templates. The amount of RNAP used was based on the number of promoters the DNA fragment contained. A 1:1 ratio of promoter to RNAP was selected as this provided the highest yield of OPCs while limiting the levels of non-specific interactions of RNAP with a given DNA template. The activity of RNAP was assumed to be 100 %. RNAP was stored at -80°C in small aliquots to ensure repeat freeze thaw cycles did not occur.

Once the samples had been mixed by gently pipetting up and down they were incubated at 37°C for 20 minutes in order to allow OPCs to form. Once incubated, samples of the 1144 bp and 602 bp template were diluted by a factor of 10 in imaging buffer. Samples containing the 2521 bp template were diluted by a factor of 20. The difference between the two samples is to ensure that well dispersed molecules are seen upon the surface when imaging.

#### **4.3.2 *In vitro* transcription: initiation of elongation**

In order to initiate transcriptional elongation, all four NTPs were added to a sample of OPCs to a final concentration of 100  $\mu$ M each. The sample was then incubated at room temperature (between 20-25 °C) for 15 minutes. The NTPs were added in excess to ensure that they were not limiting the elongation reaction. The lower temperature

helps prevent the formation of new OPCs during the elongation step, due to the fact that there will be some RNAPs that are free in solution and not yet located their promoters. This also helps to reduce the occurrence of multiple rounds of transcription from occurring. This would lead to a skewing of results collected. Again once incubated samples were diluted as with OPC samples.

#### 4.4 Imaging of DNA samples and *in vitro* transcription complexes

Detailed in this section is the process which samples undergo in order to view them under the AFM. This work is based on previous experiments [5, 215, 344]. The samples were imaged using either a Multimode 8 AFM (Bruker, Billerica MA) or a FastScan Bio AFM (Bruker, Billerica MA). Both machines were operated in tapping mode in air at room temperature (22-25°C). The tips used were TESPAs for the Multimode (Bruker, Billerica, MA) and Fastscan A (Bruker, Camarillo CA) for the FastScan bio the specification of both tips is given in Table 4-1.

	TESPA	Fastscan A
<b>Material</b>	(n) doped silicon	Silicon nitride
<b>Cantilever geometry</b>	Rectangular	Triangular
<b>Thickness (nominal)</b>	4 $\mu\text{m}$	0.58 $\mu\text{m}$
<b>Back side coating</b>	Reflective aluminum	Reflective aluminum
<b>Tip radius (nominal)</b>	8 nm	12 nm

**Table 4-1: Table of tip specifications used for imaging.**

The cantilever auto tune function was used to locate the resonance frequency for each probe and the frequency of oscillation was offset by a maximum of 5% below the resonance. The offset was used to ensure that the cantilever was oscillating close to the resonance when imaging as changes in the resonance frequency can occur when the tip comes close to the surface. The amplitude of oscillation was initially set to 500 mV but this was changed throughout imaging to provide the best image. The set point

#### Chapter 4: General methods

was initially selected automatically but was altered along with the integral and proportional gains to provide the best image quality. The images on the Multimode were collected with a setting of 512 samples per line at a scan rate of 2.9 Hz. For the Fastscan Bio the majority of images were between 512-1024 samples per line at a scan rates ranging from 11-22.4 Hz but this was changed depending on the sample tip interactions and image quality on the day. The standard ranges for operation of both AFM is given in Table 4-2.

	<b>Multimode</b>	<b>Fastscan</b>
<b>Setpoint</b>	200-1000 mV	500-1500 mV
<b>Integral gain</b>	0.2	1.0
<b>Proportional gain</b>	0.5	5.0
<b>Scan rate</b>	2.0-3.2 Hz	11.0-22.4 Hz
<b>Samples per line</b>	512	512-1024
<b>Amplitude of oscillation</b>	5-10 nm	2-20 nm

**Table 4-2: Table of AFM setting used for imaging**

All images were collected using a Z-range of 5 nm unless otherwise stated.

#### 4.4.1 Sample deposition

Once transcription complexes had been formed, 1 $\mu$ L of the sample was diluted into 9 $\mu$ L of imaging buffer (Tris-HCL (4mM, pH 7.5) and 4mM MgCl<sub>2</sub>) for samples containing 1144 bp and 602 bp DNA or 0.5 $\mu$ L was diluted into 9.5 $\mu$ L for samples containing 2521 bp DNA. For bare DNA, 1 $\mu$ L of DNA at a concentration of 2.5ng/ $\mu$ L was diluted in 9 $\mu$ L of imaging buffer. The imaging buffer contains MgCl<sub>2</sub> in order to enable 2D equilibration of the DNA on mica surface [277]. The full 10 $\mu$ L of sample in imaging buffer was deposited onto freshly cleaved muscovite mica (Agar scientific, Essex UK) that had been cleaved using sticky tape. Samples were incubated on the surface for 5 mins to ensure that complexes adsorbed to the mica surface.

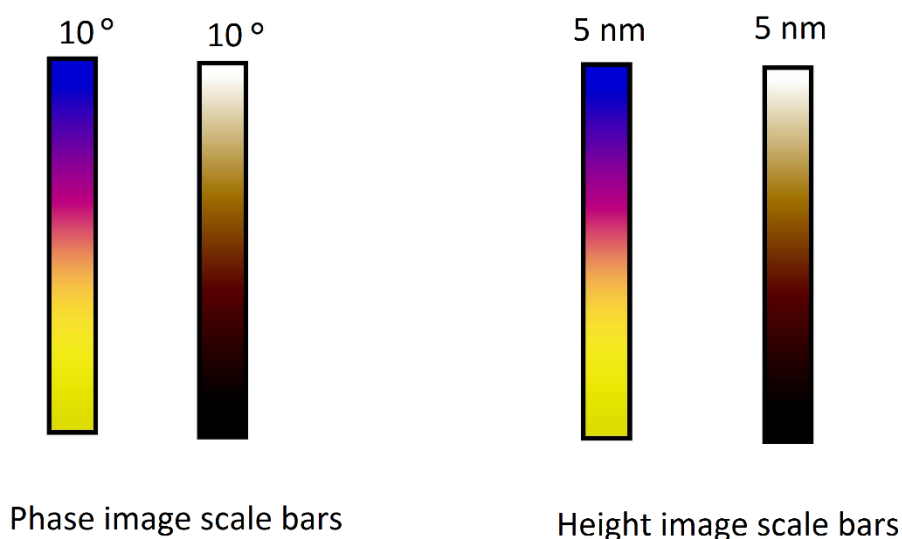
#### **4.4.2 Rinsing and drying**

Samples were rinsed using ultra-pure water with a resistivity of 18.2 M $\Omega$ .cm (at 25°C), after incubation. The rinsing step helps to remove any DNA and protein molecules that may not be tightly bound to the surface, thereby helping to reduce the likelihood of any of the molecules binding to the AFM tip and affecting image quality. Rinsing also removes any excess buffer which is important as the salts in the buffer can form crystals on the surface upon drying and lead to issues when imaging as well as having a detrimental effect on the DNA. A volume of 5-8ml was used to rinse the samples. The mica disc was held with tweezers freshly cleaned with ethanol and tilted to prevent pooling of water on the surface. The water was exuded at a flow of approximately 0.5-1 ml per second. The flow rate was fast enough to prevent pooling of water on the surface but not so fast as to remove any molecules loosely bound.

Once a sample was rinsed, it was dried in steady stream of nitrogen at 1 bar of pressure until all liquid was removed from the surface. Higher pressures were found to damage the sample but there is a need for a relatively quick drying process in order to discourage the formation of salt crystals. Once the sample was dried it was stored in a standard petri dish and imaged within 24 hours.

#### **4.4.3 Analysing samples**

Once AFM images had been collected, the raw files were exported to the AFM manufacturer's software Nanoscope Analysis 1.4 or 1.5 (Bruker, Billerica MA). The files were then flattened in the 0<sup>th</sup> order to centre the data in the digital Z range and subsequently flattened in the 3<sup>rd</sup> order to remove tilt and bow. Flattening utilises a best fit polynomial fit for each line of data to centre data (0<sup>th</sup> order) or remove tilt and bow (3<sup>rd</sup> order). After flattening, the height scale of the data was set to real time and the Z-range for all images was set to 5 nm. For phase images the data scale used for all images was 10°. Representative scale bars are shown in Figure 4-5.

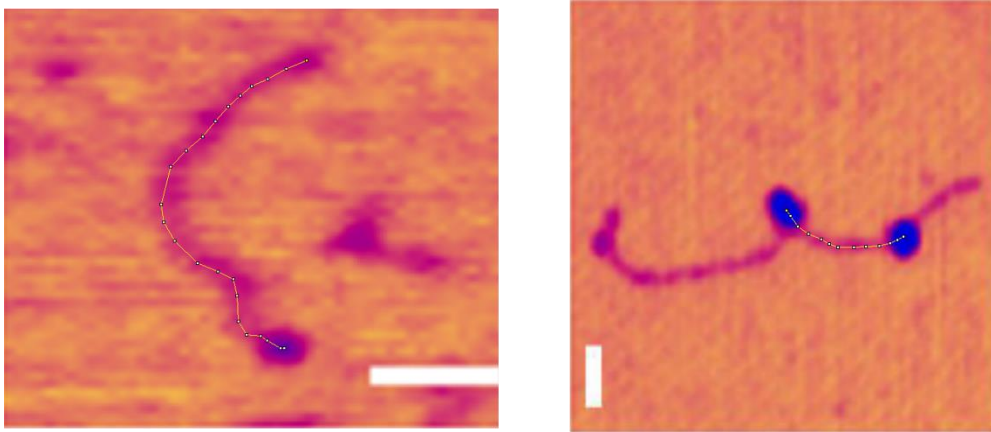


**Figure 4-5: Scale bars used for all AFM height and phase images throughout the thesis.**

Height measurements, inter-RNAP distances and cross sectional analysis was performed using the software's cross section analysis tool. Height measurements were taken from the highest point of an object of interest to a point that was at 0 nm on the mica. Cross sectional analysis was performed by selecting the width at half the height of the object.

In order to perform length measurements files were exported as JPEG or BMP files which maintained the pixel ratio of the original image before being loaded into the analysis suite ImageJ [345]. As the scan size of an image is known and the number of pixel per line that make up an image are known it is possible to set a scale in order to provide accurate measurements of length scales.

DNA contour length measurements were performed by tracing a line along the DNA back bone. When RNAP was bound to the DNA the contour length was measured to the centre of the RNAP molecule. Samples which had a loop label attached had the length measurements made from the centre of the loop as is shown in Figure 4-6.



**Figure 4-6: AFM height images with the tracing lines of contour lengths measured from the loop label and from RNAPs shown (scale bars = 50nm).**

Lengths were recorded and plotted in histograms using OriginPro in order to provide insight into the distribution of lengths seen sample sets of 100 molecules or greater were collected. This number was selected in order to provide data sets that could be plotted in histograms plots, as small data sets were not seen as representative of the populations believed to exist. Histograms had bin number selected using the square root choice (square root of number of data points). Errors for length measurements are given as the standard error of the mean value throughout the thesis.

Fitting of peak values was performed using the OriginPro fitting Gaussian fitting function with the Y value set zero where possible. For single peaks the equation used was:

$$y = y_0 + \frac{A}{w\sqrt{\pi}/2} e^{-2\frac{(x-x_c)^2}{w^2}}$$

The parameters used were area (A), offset ( $y_0$ ), center ( $x_c$ ) and width (w). Derived parameters were the full width half maxima, standard deviation and height of curve.

For multiple peak fittings the same equation was used and peaks were manually selected by designating the approximate center of each peak using OriginPro peak designator.



This page is left intentionally blank

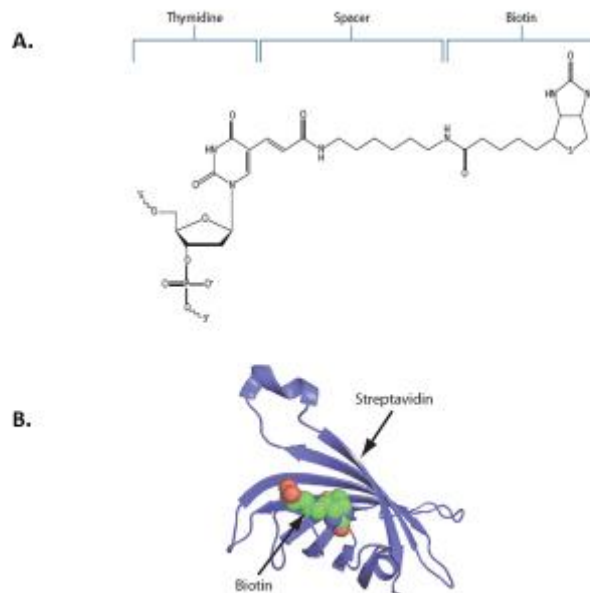
# 5 High throughput labelling

## 5.1 Introduction

Throughout this work a linear DNA fragment was utilised as a transcription template. Upon formation of OPCs it is possible to determine that the RNAPs are at their promoters using contour length measurements of the DNA backbone. The polarity of the DNA template can also be determined from these measurements if it is designed to appear asymmetric when visualised by AFM. Even so, after the addition of NTPs into the system, elongation occurs and the RNAPs leave their promoter sites and can be distributed across the whole length of the template. This draws into question which polymerase originated from which promoter and therefore raises doubt about the position of RNAPs in relation to their starting point. In the case of convergent and tandem transcription *in vitro* it is assumed that the RNAPs are unable to pass each other based on previous studies [3, 193, 318]. With this in mind, a fiducial marker or label that can provide information on the polarity of the DNA template is required. This label must be specifically attached to a known position of the DNA in order to be informative and would be best situated at the end in order to not influence the process of elongation. This label is also required to be compatible with AFM sample preparation and analysis. The requirements for an end label led to the development of a method by Billingsley *et al.* to incorporate a nucleic acid based single stranded loop to one end of the template [5]. This method was successful in discerning the polarity of the DNA template by AFM and is discussed in detail in 5.1. This chapter discusses the reasoning behind such a label and the method utilised by Billingsley *et al.* Then presented is the development of a labelling method based on this initial protocol, going from a single step PCR process to a multistep PCR based process. Also discussed is the use of the four different homopolynucleotide sequences for the loop region of the label, in order to compare the effect that the sequence may have on loop appearance and on transcription reactions.

### 5.1.1 Labelling DNA for AFM

The end labelling of dsDNA is not a new concept and there are a number of labels and methods available. The most common of these methods is to use a biotin linked dNTP as is seen in Figure 5-1a [346-349].



**Figure 5-1: Biotin labelled dNTPs can be incorporated into a DNA fragment. Thymidine labelled with biotin, a spacer is incorporated in order to allow binding to streptavidin. B) Biotin is shown by the green and red dots, bound into a pocket of streptavidin, shown by blue ribbons, with a  $K_d$  in the order of  $4 \times 10^{-14}$  M, yielding very tight binding image was edited from reference [350].**

This in turn can be linked to a streptavidin protein as shown in Figure 5-1b which binds extremely tightly to the biotin moiety [347]. This method has been further developed to incorporate other proteins, such as ferritin as well as other non-biological molecules such as colloidal gold and certain dye molecules [319, 351-353]. The use of biotin-streptavidin based methods have been utilised for AFM studies, but they were deemed to be unsuitable for use in our transcription system. This is mainly due to the fact that they may affect the adsorption of the DNA-protein complexes to the mica surface due to the importance of local chemistry on the interaction [194]. The effect that such molecules may have on the function of the RNAP and its interaction with the DNA is also not fully understood and so may introduce unwanted interactions in the model

### Chapter 5: High throughput labelling

transcription system. The use of proteins may lead to unwanted protein-protein interactions which are normally disrupted by increased salt concentrations. High salt concentrations, however, decrease the adsorption of complexes to the mica surface and leave deposits in air-dried samples that can obscure the biomolecules, such that it may not be possible to prevent such interactions in our system. The appearance of a protein or bulky marker under AFM may also be hard to distinguish from the RNAP depending on its molecular weight [354].

Due to all these factors, a nucleic acid based label was decided upon. This would mean that there would be very little change to the local chemistry when the complexes were adsorbed onto the mica, no issue with increasing protein-protein interactions and a minimisation of detrimental effects on the function of the RNAP. A prevalently used method for the incorporation of nucleic acid based labels is triplex forming oligonucleotides (TFOs). These are short stretches of ssDNA that are designed to insert into the dsDNA at specific regions. These TFOs can be conjugated to other molecules or designed to contain secondary structures such as hairpins that stand out from the DNA backbone [355-357]. TFOs have two major drawbacks for use in RNAP studies. Firstly, they are often used to label internal sites along the DNA and secondly they rely on topological differences in the DNA, such as the supercoiling state of the DNA to form [355]. This means that they are not ideal for labelling the ends of the DNA and also that they may not be viable for linear DNA templates [358].

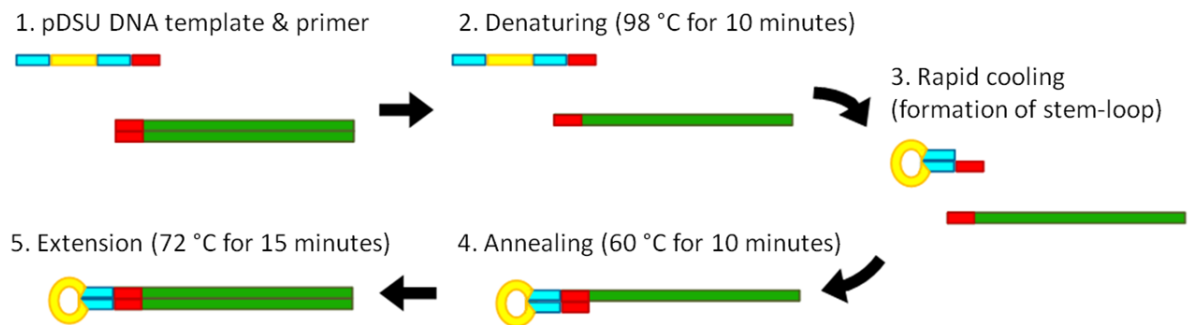
As a consequence, Billingsley *et al.* devised a new method to incorporate a single stranded hairpin loop to the end of a linear DNA template. This hairpin molecule was easily distinguishable from the DNA backbone as well as from the RNAPs.

#### 5.1.2 Nucleic acid based end label for AFM

The method of Billingsley *et al.* was based on a PCR reaction, however, the protocol was not that of a typical PCR with a large number cycled steps, but was referred to as a “single step” PCR. The method utilised a small ssDNA hairpin loop that was designed to contain a double stranded “neck” region to seal the loop and small extended single stranded tail to allow for annealing to the template. The single stranded loop was

## Chapter 5: High throughput labelling

made up of 20 adenosines. Prior investigations by Billingsley *et al.* as well as others showed that these small DNA secondary structures had an appearance that was distinct from a typical dsDNA backbone under AFM [4]. The “neck” region held the hairpin loop in a closed conformation. The tail also acted as a primer for the DNA polymerase during the extension stage of the reaction which is summarised in Figure 5-2.



**Figure 5-2: Process of the labelling protocol devised by Billingsley *et al.*** The reaction starts by melting the original DNA template in the presence of the hairpin forming oligonucleotide. The reaction is then cooled rapidly to encourage the formation of the hairpin loop. The reaction is then heated back to 60°C to allow annealing of the tail region (primer) to its complementary sequence within the original DNA template. DNA polymerase was then added and the reaction heated to 72°C in order to allow the DNA polymerase to produce a complementary strand, leaving a double stranded fragment with the hairpin loop attached to the end.

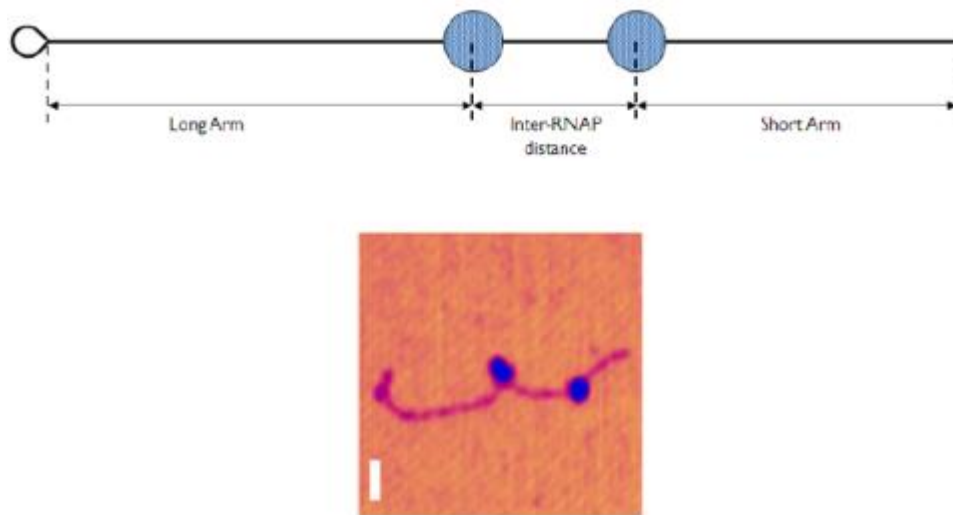
The steps of the reaction were similar to that of a PCR. The two strands of the target DNA were first melted at 98°C in a PCR buffer with just the hairpin loop structure in high excess and no second primer. The sample was then rapidly cooled, in order to drive the annealing of the neck region to form the loop. As with a PCR, the reaction was then held at 60°C to allow the single stranded tail to anneal to its complementary sequence within the target DNA. This newly formed double stranded region acts as the primer for the DNA Polymerase which when heated up to 72°C, is then able to fill in the second strand of the DNA target, leaving a double stranded molecule with the single stranded loop attached to the end of the target DNA.

This method provided a number of advantages. Firstly, its simplistic nature meant that it was relatively easy to perform. Secondly, the single stranded primer region could

### Chapter 5: High throughput labelling

have its sequence altered to attach to any DNA target. The specific base pairing of nucleotides means that the loop structure can be attached to a specified end with some certainty.

It was noted that the loop adopted three distinct structures when analysed by AFM, but the majority of structures appeared triangular or globular with a height greater than that of the DNA backbone and a diameter of ~20 nm. Upon the addition of RNAP it was seen that OPCs formed at the expected promoter sites and showed that the loop structure was attached to the correct DNA arm. The loop feature was distinguishable from the DNA backbone and the RNAP molecules. A comparison of the predicted structure and an AFM image of a labelled complex can be seen in Figure 5-3.



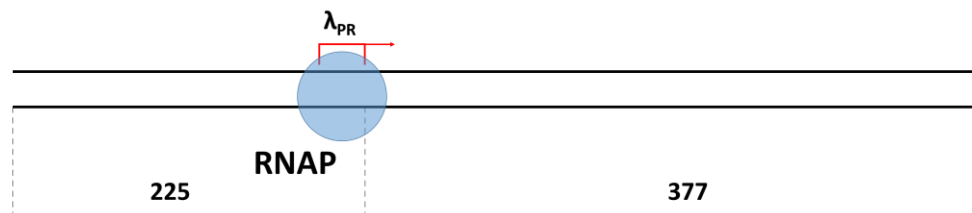
**Figure 5-3: Comparison of the predicted structure of a labelled OPC and an AFM image of a labelled OPC. As it can be seen, the loop is attached to the desired long arm and has an appearance that differs from the DNA backbone and the RNAPs (Scale bar is 50nm).**

This method was deemed successful in labelling the target DNA at the specified end with a feature that was easily recognisable from the DNA backbone and the RNAP. There was also no noticeable effects on the adsorption of complexes to the mica support, the formation of OPCs or the elongation process [5]. Even so, there were a number of issues that needed to be addressed. Mainly, the method only provided a labelling efficiency of 48% making collection of data slow. The process also left a low yield of DNA after purification of the labelled products due to lack of amplification of

the DNA template. Due to these facts it was decided that a new method that was high-throughput and more efficient would be essential in order to allow for the collection of data on labelled templates in a timely fashion.

## 5.2 Samples and preparation

The template used to develop the high-throughput labelling method was a template derived from the pDSP plasmid containing only the leading promoter (Figure 5-4). This template was decided upon as the transcribed sequence is the same as that transcribed by the leading RNAP during concurrent tandem transcription reactions. This means that the template can act as a control for elongation from tandem promoters.



**Figure 5-4: Diagram of the 602bp DNA template used to test new labelling method. The promoter region is shown with red with the RNAP overlaid in blue. The lengths of each DNA arm are shown in base pairs from the transcription start position.**

The 602 bp template was formed by PCR amplification using GoTaq Polymerase (Promega, Madison WI) as per the manufacturer's instructions by using pDSP as a template and the forward primer: 5' GCTACCAGGGAAGAACGGGAAGG 3' and reverse primer: 5' AAGCTTGATGCCTGCAGGTC 3'. The 1144 bp template was formed using the primers 5' AGGTGAGAACATCCCTGC 3' (forward) and 5' GCATGCCTGCAGGTC 3' (reverse). The template size was checked using a 1% (w/v) agarose gel electrophoresis and was purified using the QiaQuick PCR purification kit (Qiagen, Valencia CA). The labelled samples were then used for *in vitro* transcription reactions as detailed in Chapter 4. Once it was confirmed as successful, the labelling was applied to the two promoter templates derived from the plasmids pDSU (convergent promoters) and pDSP (tandem promoters).

## Chapter 5: High throughput labelling

Samples were prepared for AFM as previously described and deposited onto freshly cleaved muscovite mica before being imaged in air on a Multimode 8 AFM in standard tapping mode using TESPA V2 probes (Bruker, Camarillo CA).

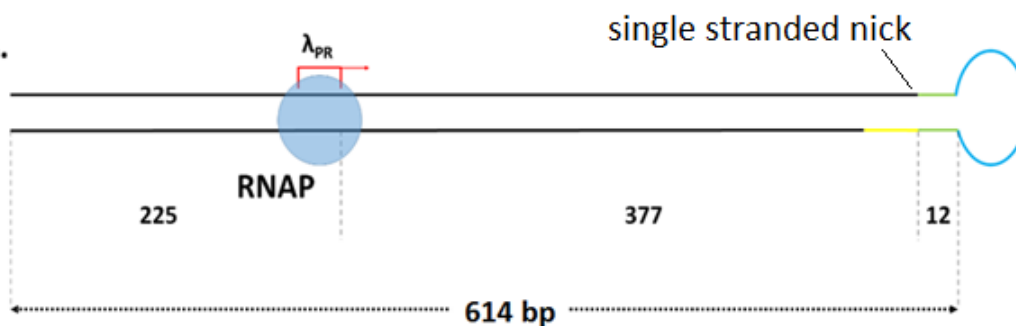
### 5.3 PCR based labelling method design and results

It was decided that the development of PCR based labelling protocol would be most beneficial, as it would provide exponential amplification of the labelled DNA target. In order to incorporate the label, the reverse primer in the reaction contained the loop structure and was to attach downstream from the promoter meaning that upon addition of NTPs the RNAP elongates towards the loop in order to investigate the outcomes of elongation towards the loop. The sequence used for the loop and the proposed structure is shown in Figure 5-5.

A.

***GGCCCTGGAGGG***-----**CCCTCCAGGGCC****GCATGCCTGCAGGTC**

B.



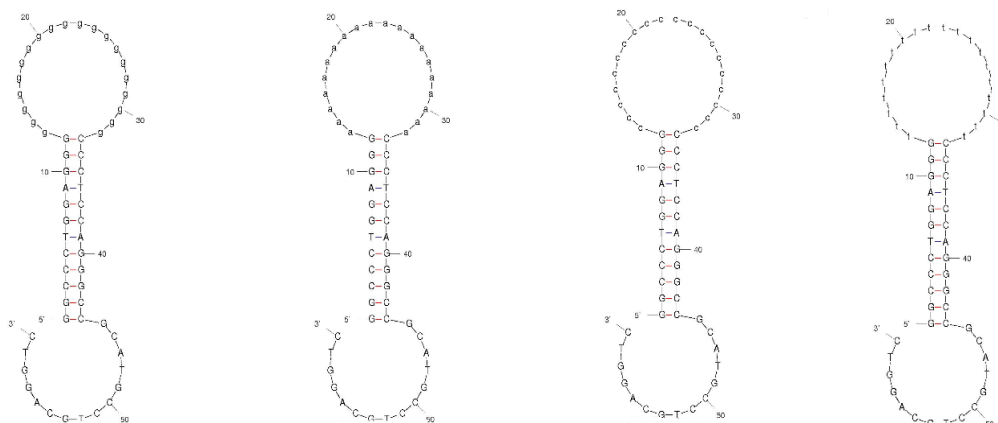
**Figure 5-5: A) Sequence of the loop primer and B) diagram of labelled template** The base pairs that form the loop primer: the neck region (Green) sequence is shown in italics, the 20 base homopolynucleotide that makes the loop (Cyan) are shown by dashes and the primer tail (yellow) sequence is shown in bold.

As can be seen from the diagram in Figure 5-5, the neck region is formed via an inverted repeat and anneals to the template through the primer region. The bases that make up the loop are not included as they were all the same, either poly-A, C, G or T in order to investigate whether the different bases had any effect on the structure and



## Chapter 5: High throughput labelling

appearance of the loop, as well as the effect this change might have on production of the loop. The folding of the loop was checked using the MFold online folding tool [359], an image of the folded structures are shown in Figure 5-6.



**Figure 5-6: MFOLD results for the four loop forming primers used with 20 base loops of polyA, polyC, polyG or polyT. The polarity of the DNA strand is shown with the bases that form the loop in lower case.**

Initial PCR attempts were performed using a standard PCR reaction setup with GoTaq Polymerase (Promega, Madison WI). The reaction conditions are given in Table 5-1.

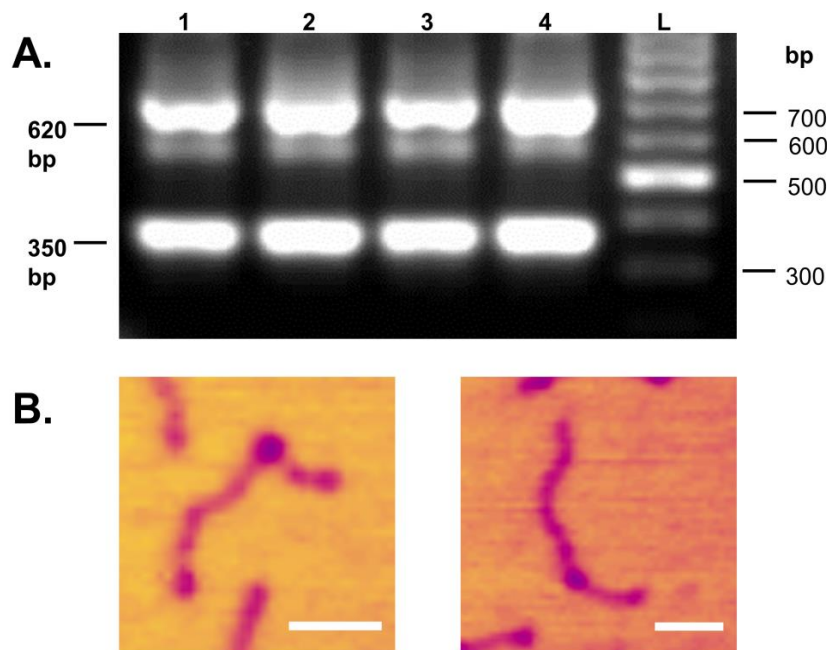
Stage	Temperature (°C)	Time (Minutes) / No. of Cycles
Initial Denaturation	96	5/1
Denaturation	94	1/30
Annealing	60	1/30
Elongating	72	1min15/30
Final Elongation	72	10/1

**Table 5-1: PCR conditions used for labelling with GoTaq polymerase**

With the GoTaq reaction mix and the stated conditions it was seen that a number of DNA species with differing sizes were produced. The amplified band at approximately 602 bp for the A loop was excised from the gel and purified as detailed in Chapter 4 and imaged by AFM. The agarose gel and an example of these molecules can be seen in Figure 5-7. It was observed by AFM that a number of templates had a globular structure located along the backbone of the DNA (B). This was believed to be a

Chapter 5: High throughput labelling

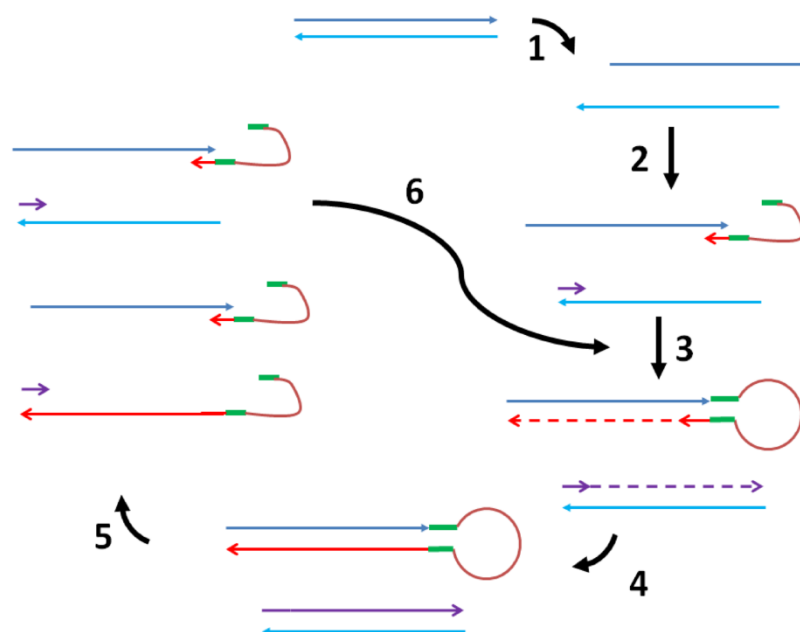
secondary structure incorporated into the DNA backbone due to its globular appearance and height greater than the DNA backbone. The other bands observed in the agarose gel were not excised and imaged.



**Figure 5-7: Fragments produced from GoTaq polymerase labeling PCR. A) Image of a 1% (w/v) agarose gel of products from PCR. Lanes: 1- Poly(A) primer products ; 2 - Poly(T) primer products; 3 - Poly(C) primer products ; 4 - Poly(G) primer products ; L: 100bp ladder (Promega). Three bands can be seen for each product, with high levels of DNA at approximately 620bp and 350bp. B) AFM height images of products purified from the band seen at 620bp for the poly A loop. The presence of a globular region with greater height than the DNA backbone can be seen (Scale bars = 50nm)**

It was reasoned that this occurrence was due to unwanted structures forming during the PCR reaction therefore leading to the loop structure forming in the incorrect place. It was also noted that the enzyme GoTaq polymerase contained 5'-3' exonuclease activity. This means that the enzyme is not only able to produce DNA in 3'-5' direction but degrade DNA in 5'-3' direction, which is common for standard commercially available *Taq* polymerases and is involved in correction of any errors during the PCR reaction. This would mean the loop containing primer may be degraded during the

extension step [360]. This is highlighted by the flow diagram of the PCR process shown in Figure 5-8.



**Figure 5-8: Schematic representation of the PCR labelling method (arrows represent 3' to 5' direction). Blue: forward strand; light blue: reverse strand; green: inverted repeats; brown: loop; red: reverse primer region; purple: forward primer. 1) Initial denaturation; 2) Primer annealing; 3) annealing of stem loop and extension of primers 4) products formed from first extension; 5) second cycle of denaturation and primer annealing; 6) second cycle of primer extension. Steps 3 to 6 are repeated for 30 cycles leading to exponential propagation of a labelled double stranded template.**

It can be seen that at step 5 shown in Figure 5-8 the *Taq* polymerase might degrade part of the inverted repeat of the loop primer. It was therefore decided to use a *Taq* polymerase that did not contain the 5'-3' exonuclease activity: QBIO-*Taq* Polymerase (MP Biomedical). This specific *Taq* polymerase has a truncated N-terminal meaning that the enzyme is unable to degrade the loop primer. It was also thought that the loop may not be able to form with high efficiency, due to the lack of a sudden cooling step as was applied in the method of Billingsley *et al.* The introduction of a sudden cooling step may deteriorate the action of the *Taq* polymerase. In order to ensure the loop remained in a loop structure, the extension temperature was reduced to 68°C. This was believed to prevent fluctuations of the annealing of the neck region or the

Chapter 5: High throughput labelling

formation of alternative structures that the loop primer might adopt. In order to counteract this drop in temperature the extension time was increased as can be seen in the reaction conditions given in Table 5-2.

Stage	Temperature (°C)	Time (Minutes) /No. of Cycles
Initial Denaturation	98	5/1
Denaturation	94	1/30
Annealing	63	1/30
Elongating	68	2/30
Final Elongation	68	10/1

**Table 5-2: PCR conditions used for the Q-BIO Taq polymerase**

The QBIO-Taq required a different reaction mix as given in Table 5-3.

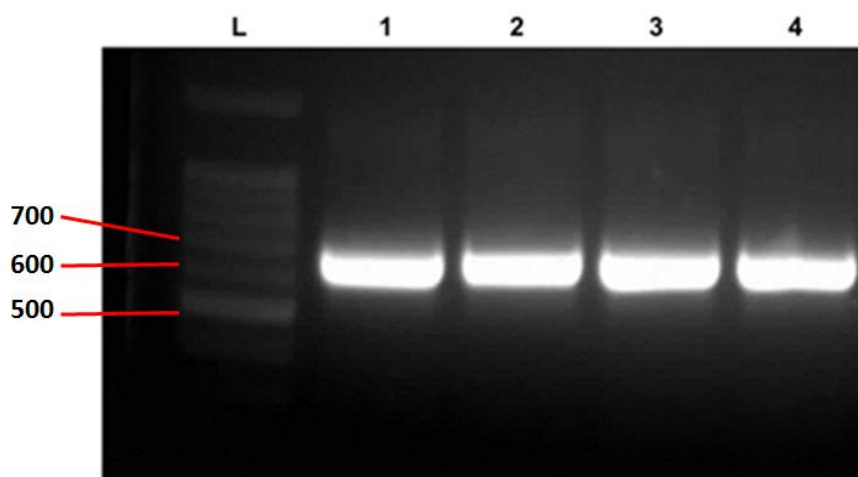
Reagent	Volume (1x)	Final Concentration
10x Qbio-Taq Buffer + MgCl <sub>2</sub>	5µL	1x
10mM dNTPs	1µL	200µM
Qbio-Taq (MP biochemical)	0.4µL	2U
Forward Primer	2µL	0.4µM
Reverse Primer ( <i>Loop primer</i> )	2µL	0.4µM
Template	(concentration dependent)	30-100ng
H <sub>2</sub> O	To give final volume of 50µL	-

**Table 5-3: Reaction mix for Q-BIO Taq polymerase PCR**

This reaction was decided upon having first performed a number of tests. Firstly the MgCl<sub>2</sub> concentration was altered to provide the highest amount of amplification and a range of annealing temperatures were tested for the primers in order to give the highest amount of amplification without the production of unwanted DNA species. This reaction mix and conditions were then applied to the different loop sequence primers.

### 5.3.1 Labelled single promoter

The four different loop sequences were amplified by PCR using the conditions given previously and the products were analysed by running on a 1% (w/v) agarose gel electrophoresis.



**Figure 5-9: Image of 1% (w/v) agarose gel of the products from Q-BIO Taq polymerase PCR. Lanes L: 100bp ladder (Promega) ; 1: Poly(A) primer products ; 2: Poly(T) primer products; 3: Poly(C) primer products ; 4: Poly(G) primer products.**

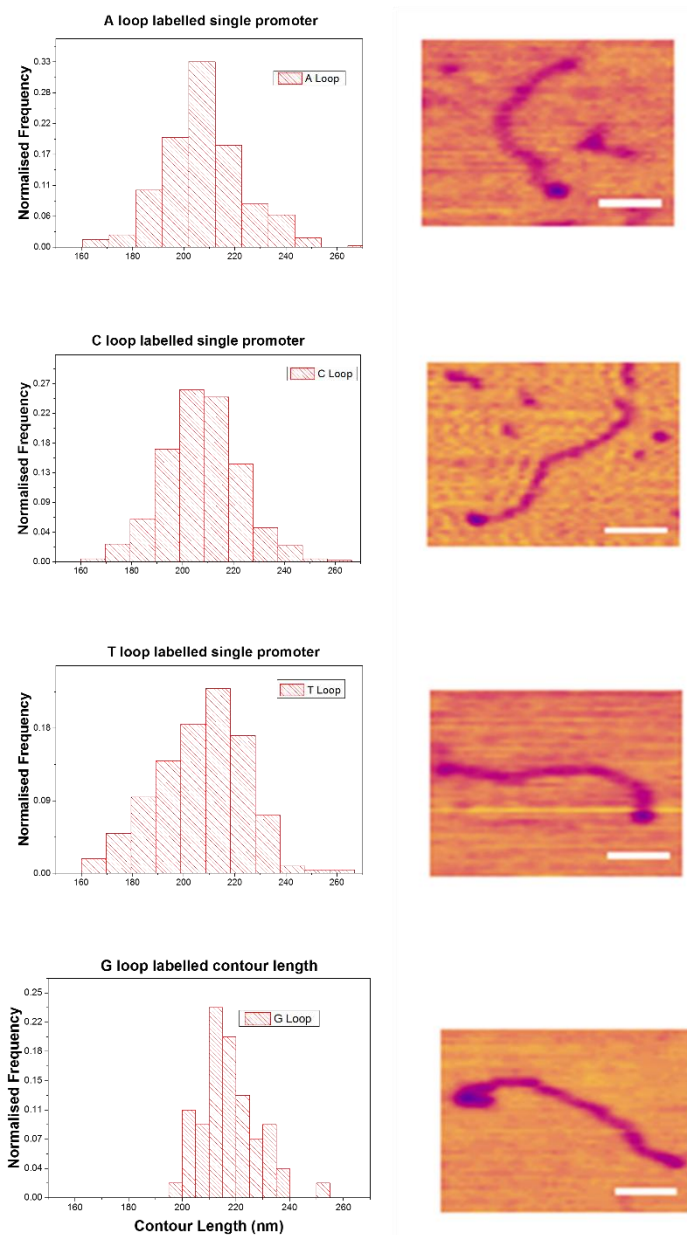
All four of the loop primers produced a single defined band at the expected size of approximately 600 bp, indicating that the PCR reaction was successful (Figure 5-9).

These products were purified using the QiaQuick PCR purification Kit (Qiagen, Valencia CA) and the concentration of each product was tested using a Nanodrop 2000c spectrophotometer (Thermo scientific MA, USA). All four templates had a concentration greater than the amount of template added to the PCR mix confirming that the PCR reaction had been successful in amplifying the target DNA.

These products were then diluted into imaging buffer and analysed by AFM. For each loop sequence the contour length of the backbone was measured from the center of the globular feature seen at the end of the DNA fragment. Histogram plots for the

Chapter 5: High throughput labelling

contour lengths of all loop sequences are shown with an AFM height image of each (Figure 5-10).



**Figure 5-10: Histogram plots of labelled single promoter DNA with corresponding AFM images for each loop sequence (n= 363, 449, 217 and 55 for A, C, T, and G respectively. Scale bars: 50nm)**

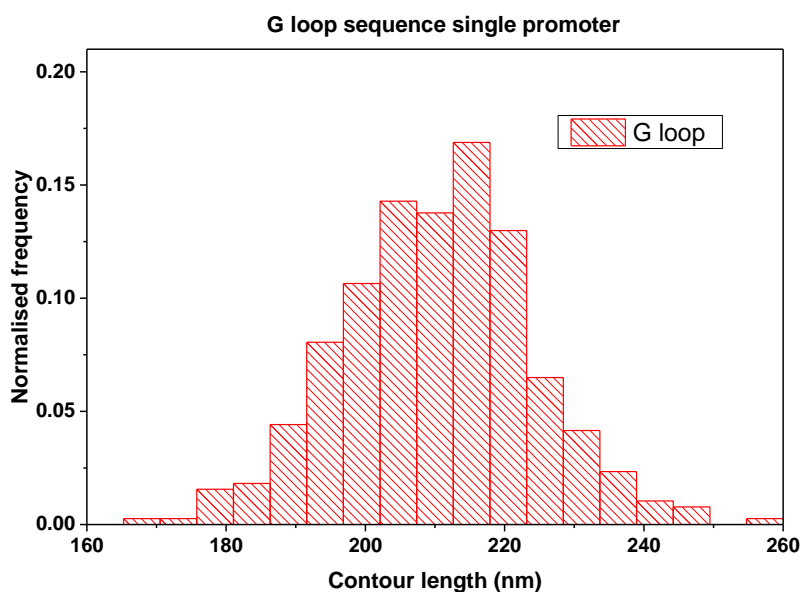
The mean contour length for the loop labelled single promoter templates are given in Table 5-4.

Loop sequence	n	Contour length (nm)
A	363	207.7 ± 0.8
C	449	207.5 ± 0.7
T	217	207.1 ± 1.2
G	55	217.1 ± 2.4

**Table 5-4: Contour lengths of DNA fragments produced with each loop primer.**

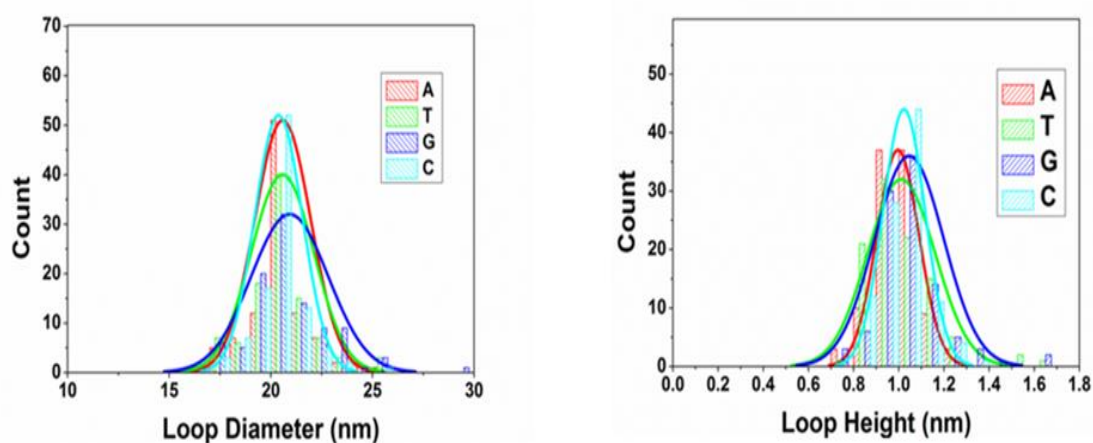
The DNA also shows an increase from the unlabeled template which had a contour length of  $194.6 \pm 1.3$  nm. The poly(G) loop labelled template displayed a contour length approximately 10 nm higher than that of the other loop sequences. When accessing the efficiency of the labelling by counting the number of molecules with an end feature present, the poly(G) loop sequence only had a feature on 10% of molecules, whereas for the poly(A, C and T) loops there was an appearance of the feature for over 70% of molecules. With this fact in mind it was reasoned that the poly(G) loop was not forming in a manner similar to the other loop sequences. This was believed to be due to an erroneous effect during the PCR amplification.

In order to investigate this further, the PCR conditions were altered for the G-loop sequence. The extension temperature was dropped to 66°C and the time was increased to 4 minutes for the cycled extensions and increased to 20 minutes for the final extension. With these changes, the poly(G) loop labelled fragment showed a mean contour length of  $210.2 \text{ nm} \pm 0.89$  (Figure 5-11) and a feature occurrence of 73%, consistent with the other homo-polynucleotide 20 base loop labels.



**Figure 5-11: Histogram plot of contour length measurements for poly(G) loop labelled DNA fragment (n= 385)**

In order to further characterise the labelling reaction, the dimensions of the globular feature itself for each loop sequence was measured. The height profile of the feature and its diameter was measured, the histogram in Figure 5-12 shows all of the loop heights and diameters plotted on the same axis for comparison between each loop sequence with the values shown below in Table 5-5.



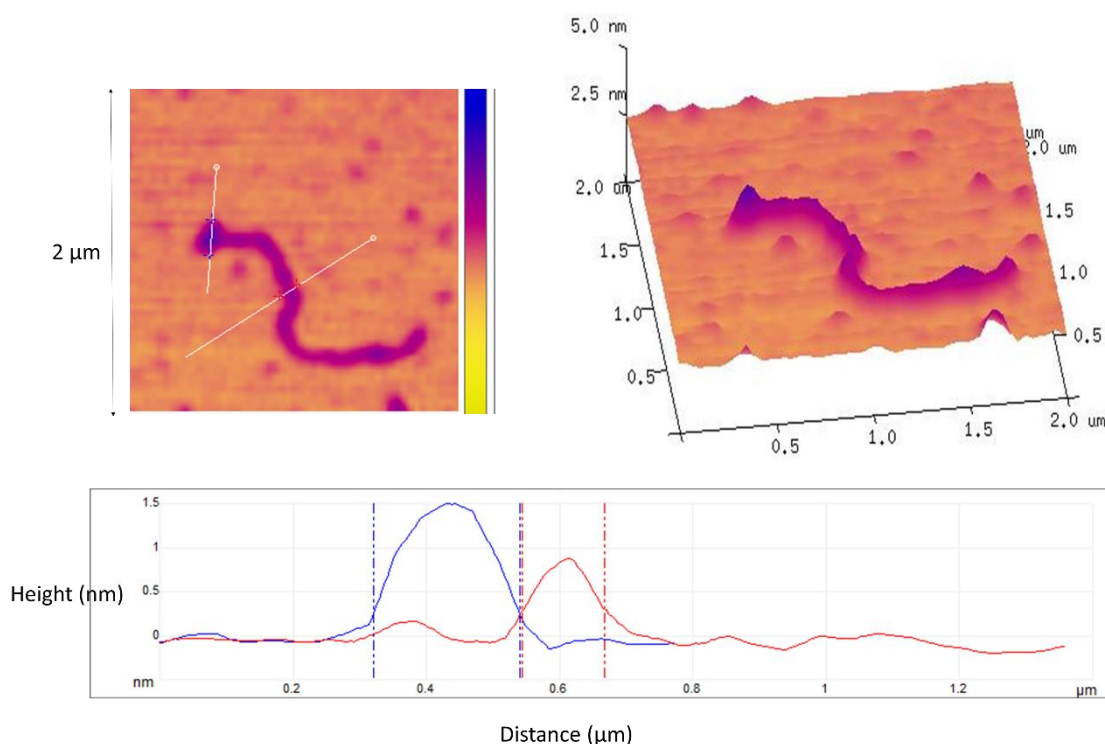
**Figure 5-12: Histograms of the loop diameter and height for all loop sequences. Red is poly(A), green poly(T), blue poly(G) and cyan poly(C).**



Loop sequence	N	Height (nm)	Diameter (nm)
A	100	0.99 ± 0.01	20.57 ± 0.14
T	100	1.00 ± 0.01	20.48 ± 0.14
G	100	1.04 ± 0.02	20.74 ± 0.27
C	100	1.02 ± 0.01	20.38 ± 0.12

**Table 5-5: Height and diameter of the end loop feature for all loop sequences.**

Both the height and diameter for all the loop sequences are similar. The height of the loop was noted as being greater than that of the DNA backbone which across all samples taken had an average height of  $0.42 \pm 0.08$  nm ( $n = 120$ ). This difference in height is highlighted by the 3D rendering of an AFM image on a labelled fragment in Figure 5-13.

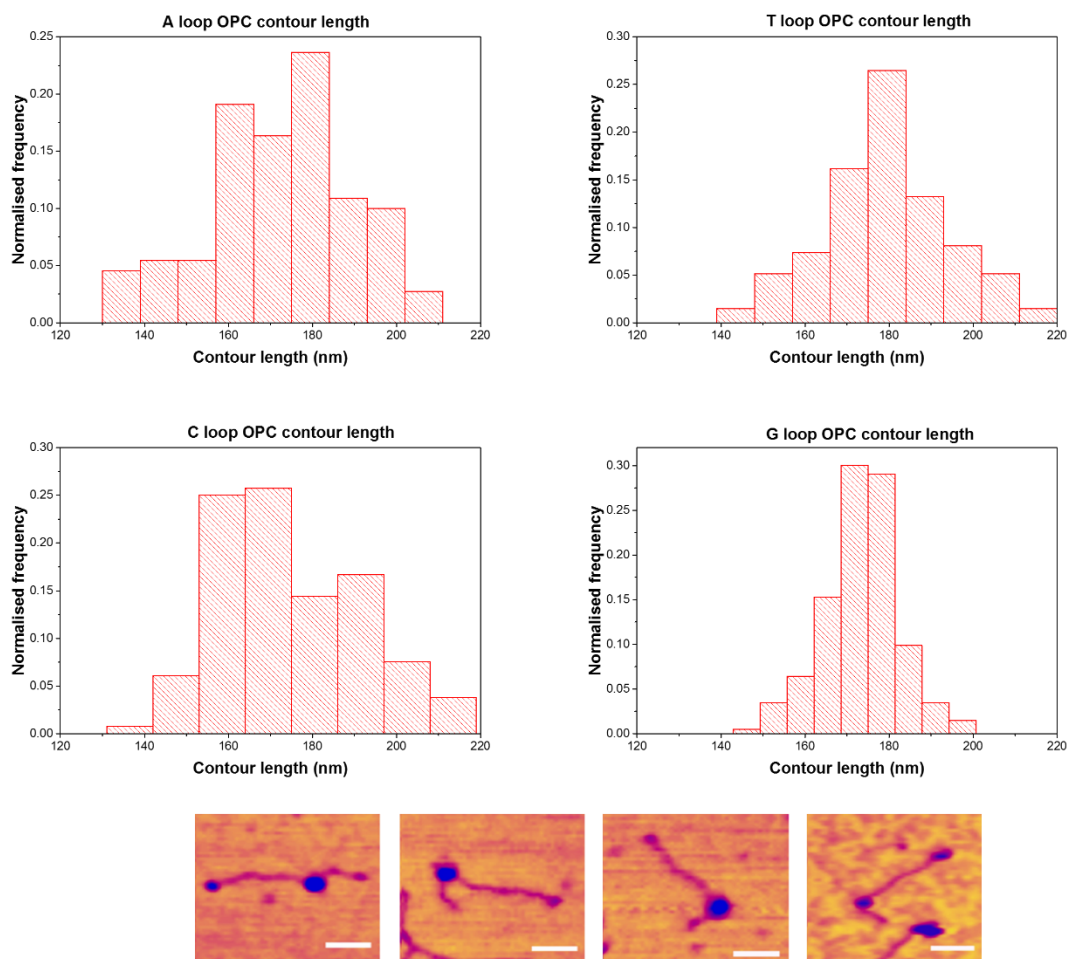


**Figure 5-13: 3D rendering of a labelled fragment with the cross-sectional analysis shown below. On the cross-section plot the blue is the loop and red is the DNA backbone. The height difference between the loop feature and the DNA backbone can be clearly seen.**

The diameter of the loop is in keeping with contour length measurements, with the addition of the loop adding approximately 13 nm to the length of the DNA from the centre of the loop which agrees well with a diameter of approximately 20 nm. This value is greater than expected as the inter phosphate spacing is expected to be between 5.9-7 Å for ssDNA [361]. This would give a circumference approximately 130-140 Å and a diameter of 4.1-4.5 nm. The observed diameter may be higher due to tip convolution or alterations to the inter phosphate distance upon surface deposition. The additional 3 nm observed when measuring contour length to the centre of the loop is most likely due to the addition of the neck region to the contour length. The width of the backbone of the DNA was measured for 30 molecules in each sample and gave an average  $10.9 \pm 0.3$  nm.

### 5.3.2 Single promoter Open promoter complexes (OPCs)

The purpose of adding a loop to the fragment of DNA was to provide a marker of polarity when visualising transcription complexes by AFM. Therefore the labelled DNA fragments, with the differing loop sequences were used for *in vitro* transcription reactions to test their influence on RNAP activity. The position of the loop was determined by measuring the contour lengths either side of a bound RNAP in an OPC. Only the poly(A) loop was used in the original one step method, but the new method was used to create the four different loop sequences, not only to quantify any differences in the loops appearance but also to determine if the sequence of the loop affected the interaction of the RNAP with the DNA template. The labelled DNA was incubated to form open promoter complexes (OPC) as described in Chapter 4 and then imaged by AFM. Examples of the complexes can be seen in Figure 5-14 along with histograms of the full contour lengths of the complexes for each loop sequence.



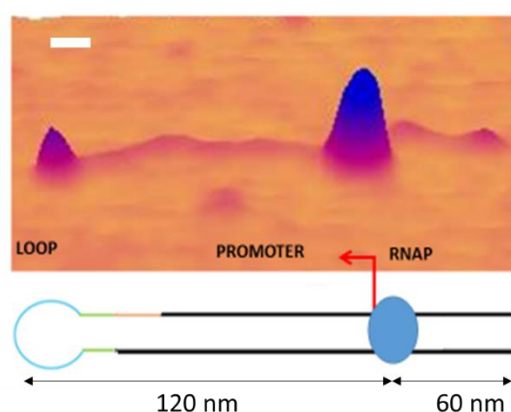
**Figure 5-14: Histograms of contour length for OPCs formed on each loop labelled DNA fragment AFM images of OPCs from each loop sequence (A, T, C, G). For histograms n= 110, 116, 132, 203 for A, T, C and G loop sequences (Scale bars = 50nm)**

The mean contour lengths of the OPCs were 174.2 nm ( $\pm 1.8$ ), 176.1 nm ( $\pm 1.4$ ), 174.4 nm ( $\pm 1.4$ ) and 174.31 nm ( $\pm 1.6$ ) for Poly (A, T, C and G) loops, respectively. The OPCs were easily identified by the distinct bend angle of the DNA observed. The bend angle was not measured. The shortening of the contour length due to wrapping as reported by previous studies, was used to determine whether OPCs had been formed [3, 126]. The loop labelled DNA arm is easily distinguishable not just by the end feature, but also by its greater contour length (Table 5-6).

Loop sequence	N	Short arm (nm)	Long arm (nm)
A	110	54.1 ± 3.5	119.5 ± 1.4
T	116	54.5 ± 4.8	120.2 ± 1.1
G	203	54.7 ± 3.8	119.3 ± 0.9
C	132	53.2 ± 3.2	120.4 ± 1.4

**Table 5-5-6: DNA arm contour lengths for each loop sequence. The loop is at the end of the long arm of this single promoter template.**

For all the poly-nucleotide loop templates, the short arms are similar in contour length. The degree of error in these results is a reflection of the fact that due to its short length this arm was hard to measure with as high a degree of accuracy as it was sometimes partially obscured by the RNAP molecule and tip convolution. The long arm measurements for the poly(A, T, C and G) loops all agree well with each other indicating that the sequence of the loop has no significant effect on the contour length or appearances of OPCs. The main reason for adding the label to the DNA was to distinguish the polarity of the DNA after transcription elongation. It is therefore a requirement that the loop not just be distinguishable from the DNA backbone but also from the RNAPs molecules bound to the surface.

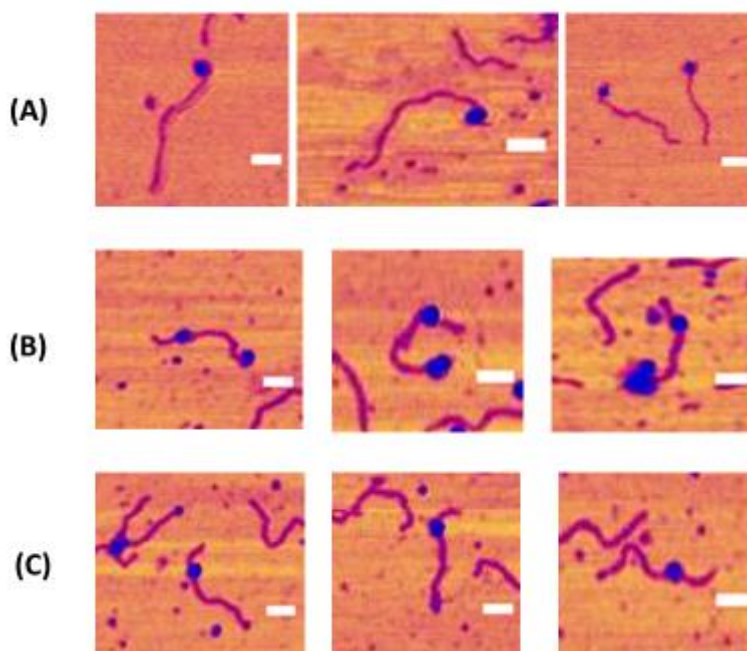


**Figure 5-15: A 3D rendering of an OPC with the loop structure imposed on the bottom of the image to provide comparison to the predicted structure. The predicted lengths of the arms are provided (Scale bar = 20 nm)**

### Chapter 5: High throughput labelling

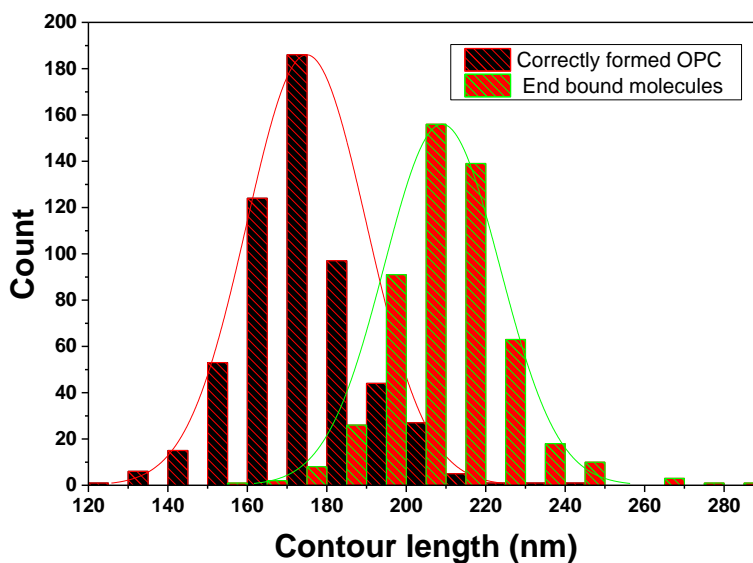
As can be seen by the 3D rendering in Figure 5-15, the loop has a lower height profile and diameter than the RNAP. This difference is further supported by the average height and diameter of RNAPs in the OPC samples. A total of 120 RNAPs (30 from each loop sequence) in an OPC confirmation had the diameter and height measured. The average diameter was  $33.7 \pm 0.46$  nm and the height was  $4.6 \pm 0.16$  nm. This is in agreement with height and diameter observed by Wyman *et al.* and Billingsley *et al.* who recorded approximately 35 nm for the diameter and 3-5 nm for the height of RNAPs in OPCs [4, 362]. This difference from the diameter and height of the loop made the two structures easily identifiable.

Upon analysis of the images there were two other types of complexes observed. Those that showed a RNAP at the promoter only made up approximately 47% (Figure 5-16 C). For all loop sequences, a species with an RNAP attached to the end of the loop arm accounted for 48 % of molecules seen (Figure 5-16 A) and a small percentage (5 %) had a RNAP at the promoter site and attached to the loop arm (Figure 5-16 B).



**Figure 5-16: Montage of images showing different forms of complexes seen in OPC samples. A) Images of samples with an RNAP bound to the loop end of the template, therefore occluding the loop from the AFM tip. B) Images where an RNAP was bound at the promoter as well as having an RNAP bound to the loop. C) Images of correctly formed OPCs with a visible loop and an RNAP bound at the expected location of the promoter. (Scale bars: 50nm)**

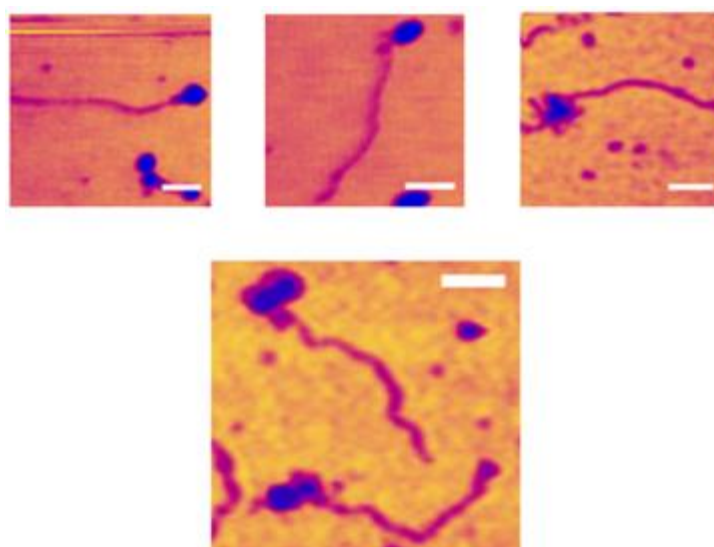
Molecules that had only one RNAP bound to the loop end of the template did not show the shortening of the contour length as is seen for those with the RNAP at the promoter and had an average contour length of  $208.8 \pm 0.98$  nm. This longer contour length indicates that the RNAP has not wrapped and therefore compacted the DNA as is seen upon formation of an OPC. This was seen for all loop sequences and a comparison of all the contour lengths for correctly formed OPCs and contour length of the end bound molecules is shown in Figure 5-17.



**Figure 5-17: Histogram plot comparing all loop sequence OPC contour lengths with all loop sequence contour lengths of molecules with an RNAP bound to the loop end of the DNA fragment. The difference in length indicates that the RNAP has not wrapped the DNA when bound solely to the loop (n =345 for OPCs and 335 for end bound) .**

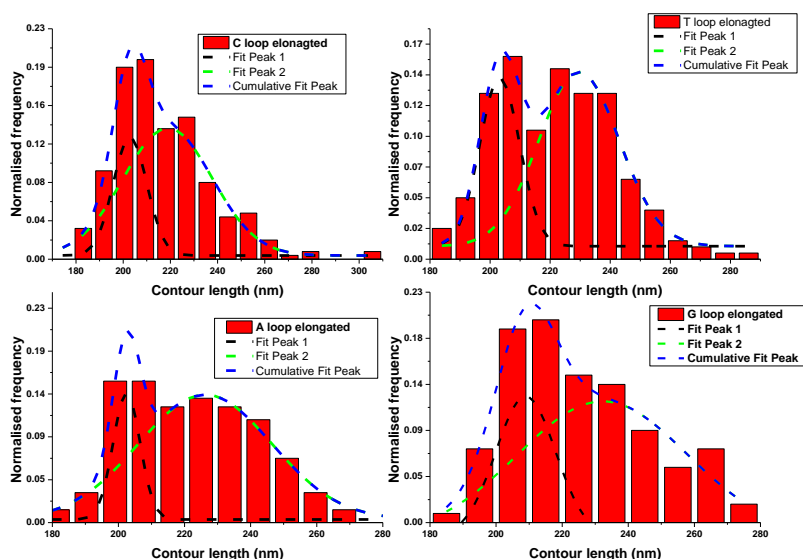
### 5.3.3 Elongated complexes

Once it was shown that it was possible to form OPCs on the labelled template, NTPs were added to the OPC reaction mixes and incubated at room temperature in order to allow the RNAP to initiate elongation, in the direction of the loop. Upon analysis of the AFM images the majority of molecules had RNAPs at the end of the template as can be seen in Figure 5-18.



**Figure 5-18: Montage of AFM height images showing the elongated complexes after the addition of NTPs (scale bar = 50 nm).**

Upon analysis of contour lengths of the molecules it was noticed that there was a greater spread of contour lengths than expected. The data collected for elongated complexes was fitted with two Gaussian peaks in order to determine whether the data represented two distinct populations. These fittings are shown in Figure 5-19.



**Figure 5-19: Histogram plots of elongated complexes contour length with corresponding gaussian fitted lines (n= 257, 209, 215 and 109 for C, T, A and G respectively).**

The centres of each fitted peak along with the average for all measurements is given in Table 5-7.

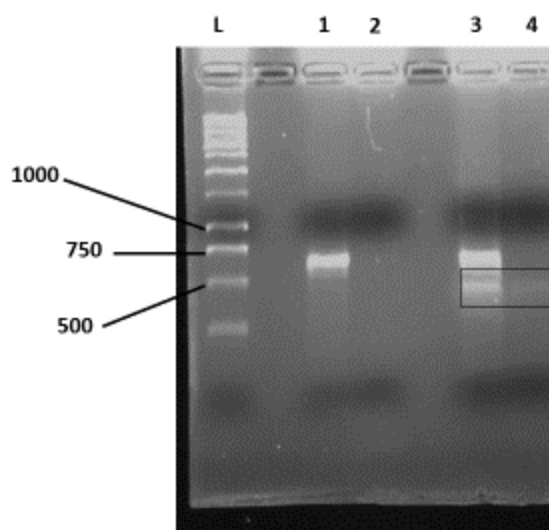
Loop sequence	n	Peak 1 (nm)	Peak 2 (nm)	Average Value (nm)
A	215	202.5 ± 0.2	229.6 ± 1.4	221.5 ± 0.8
T	209	203.2 ± 0.5	229.6 ± 0.9	221.4 ± 0.7
G	109	208.9 ± 0.9	232.2 ± 2.4	224.9 ± 1.2
C	257	202.9 ± 0.4	218.9 ± 1.2	217.1 ± 0.8

**Table 5-7: Measurements of contour length of complexes after the addition of NTPs.**

Even though the data has been fitted with two peaks there is most likely a combination of four different complexes present after elongation: 1) those complexes which have an RNAP bound to the loop as seen in OPC samples; 2) those complexes where a RNAP has elongated into the loop and has stalled at a position so that the loop is occluded; 3) complexes that have had an RNAP elongate to the loop and subsequently a RNAP has non-specifically bound to the loop, RNAP or RNA transcript; 4) complexes where an RNAP was bound to both the loop and promoter before elongation and the elongating RNAP has collided with the loop bound RNAP. The high number of complexes with RNAPs bound to one end indicate that the loop is able to cause stalling and prevent dissociation of an elongating RNAP as well as displaying affinity for  $\sigma^{70}$  RNAP.

The effect of the loop on elongation was further investigated by analysing the RNA transcript using a 1.5 % v/v formaldehyde agarose gel using the method described in Chapter 4. The DNA template with and without the poly(C) loop were run on the gel after completion of an *in vitro* transcription reaction using 10 times the standard concentrations of DNA and RNAP to ensure that large enough levels of RNA were produced. The transcription reaction was incubated at 37°C for 15 minutes to encourage multiple rounds of transcription. The samples were split into two and one half of each sample was treated with DNase I to remove the DNA template (Figure 5-20).





**Figure 5-20: Image of formaldehyde agarose gel run on elongated complexes with and without an end C-loop. Lanes L: 1kbp DNA ladder (Promega), 1: Loop Labelled transcription complexes, 2: Loop Labelled transcription complexes treated with Dnase I, 3: Unlabelled transcription complexes, 4: Unlabelled transcription complexes treated with Dnase I.**

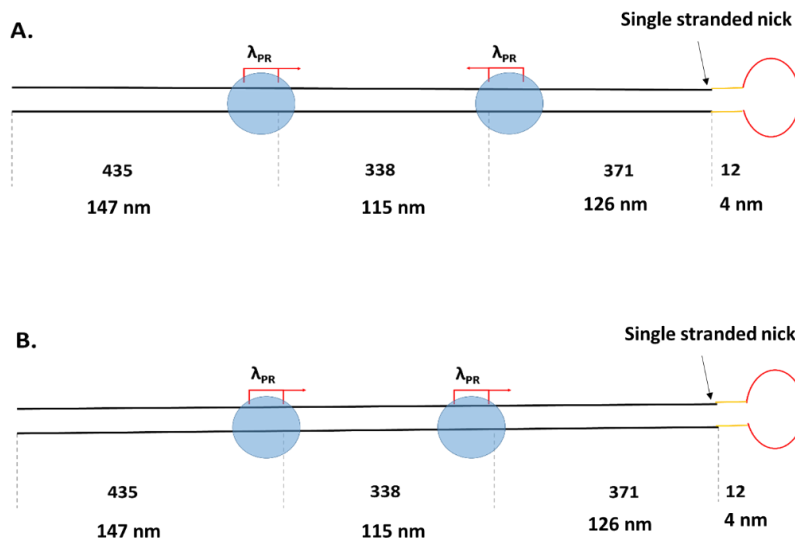
It can be seen that the C-loop labelled sample shows a band at the expected size for the transcription template before DNase I treatment but after DNase I treatment there are no distinguishable bands seen, indicating that there is not a detectable level of RNA produced. For the unlabeled sample there is a band seen at the expected length for the DNA with a second band below this, which is also present after DNase I treatment. This indicates that the band observed is RNA, and this result suggests that without the loop a higher level of RNA is produced. The simplest interpretation of this result in conjunction with the AFM data, is that the end loop label captures transcribing RNAP and prevents dissociation from the template. This would lead to a lower number of active  $\sigma^{70}$ RNAP able to undergo elongation, therefore reducing the amount of RNA produced. Whereas multiple rounds of transcription are possible on the unlabeled template. This confirms the unexpected advantage of the end loop ssDNA label allowing capture and retention of the transcribing RNAP for subsequent *ex-situ* analysis by AFM.

The hypothesis that the loop prevents dissociation of RNAP from the template by performing footprinting experiments to map the position of RNAP on the DNA after elongation.

### 5.3.4 Labelling of two promoter templates

The labelling reaction was successful for all polynucleotide sequence loops for the single promoter fragment with no differences in yield or behavior of RNAP. Due to this similarity, the selection of a loop sequence for labelling the 1144 bp two promoter DNA fragments was somewhat arbitrary, however, the loop sequence chosen was that of poly(C) as it was thought that long stretches of A or T bases could resemble the AT rich promoter elements which may encourage non-specific binding and that the G-loop required different PCR conditions to produce.

To label the 1144 bp two promoter DNA templates, the primer sequence was adjusted and therefore the annealing step of the reaction was changed to 60°C after testing a range of annealing temperatures. Diagrams of the convergent and tandem 1144 bp two promoter DNA templates can be seen in Figure 5-21.

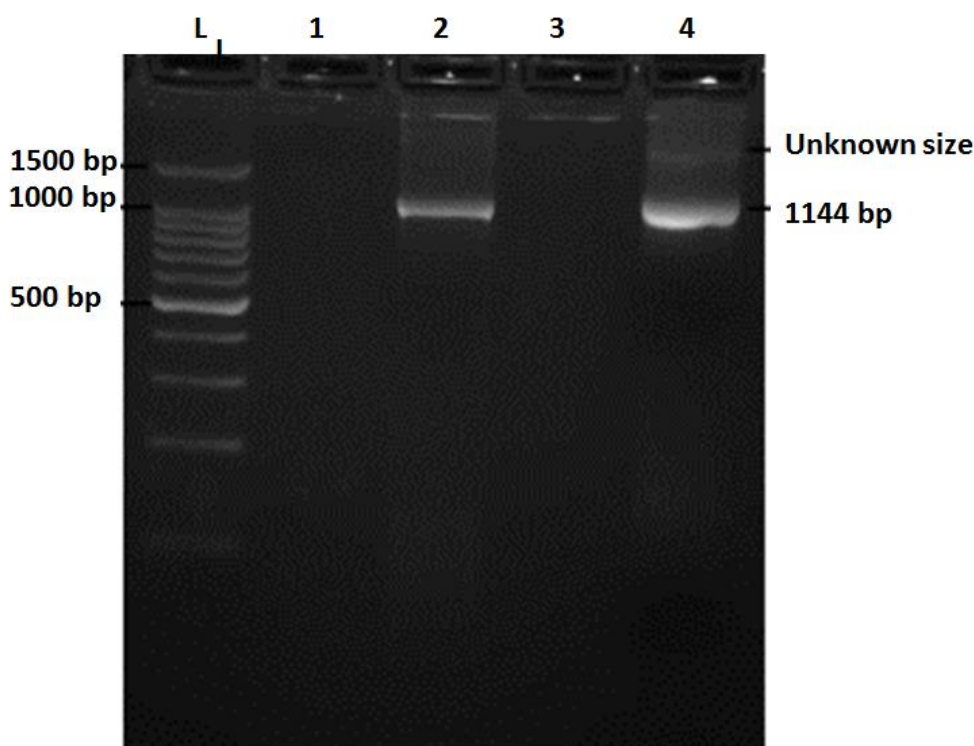


**Figure 5-21: Diagram of labelled convergent and tandem promoter containing DNA templates. In both panels the lengths of each section are shown in base pairs and nanometres, promoters are shown in red and the RNAP binding positions are overlaid in blue. The ssDNA loop label is shown in red and the neck is shown in orange.**

### Chapter 5: High throughput labelling

For the tandem promoter template, the loop was attached to the short arm, such that both RNAPs transcribed towards the loop. This design was finalised from the outcome on the single promoter template, which demonstrated that the loop will capture transcribing RNAP and prevent dissociation from the template. This would enable us to study tandem transcription using an *ex-situ* AFM approach with greater certainty that only a single round of transcription had occurred.

The products of the PCR labelling reaction were analysed firstly by 1% (w/v) agarose gel electrophoresis (Figure 5-22).



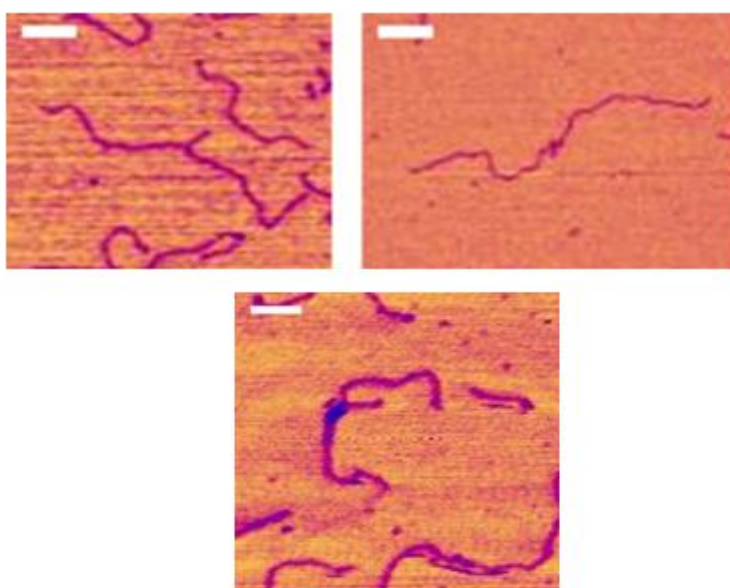
**Figure 5-22: Image of 1% (w/v) agarose gel showing products from labelling reaction on both tandem and convergent promoter templates. Lanes L: 100bp DNA ladder (Promega), 1: Sample containing tandem template at concentration added to labelling PCR, 2: tandem promoter fragment labelling products, 3: Sample containing convergent template at concentration added to labelling PCR, 4: convergent template labelling products. The unexpected band believed to be due to dimers is highlighted.**

As can be seen from the gel, the PCR reaction showed amplification of the original template DNA. There is however the presence of band of unknown size which indicates

### Chapter 5: High throughput labelling

a larger product being formed. This was only seen on occasion but was investigated further. In order to study this band as well as those that were amplified, the bands were excised from the gel and purified separately in order to image each with the AFM. Purified samples were prepared and deposited onto mica as previously described.

The band of higher molecular weight upon imaging had a contour length of 791.05 nm ( $\pm 3.2$ ) with the occurrence of a higher point approximately located in the centre of the template, as is shown by the image in Figure 5-23.



**Figure 5-23: Montage of images obtained from the band of greater size seen in Lane 4, Figure 5-22. (Scale bars = 50nm)**

The size and appearance of these fragments indicate that they are most likely dimers formed by the annealing of the two loop regions to each other giving a secondary structure at the centre of each fragment. This might occur due to the annealing of two loop primers to one another during the PCR reaction and so may be unavoidable. The issue raised by their presence is minimal as it was only a rare occurrence and it is possible to remove these fragments by using a gel extraction protocol to purify the templates. Interestingly this was not seen for single promoter templates labelled with the C-loop.

## Chapter 5: High throughput labelling

The correct fragments purified from the 1% (w/v) agarose gel for both tandem and convergent promoter templates had an average contour length of 392.10 nm ( $\pm 1.72$ ) and 390.56 nm ( $\pm 2.19$ ) respectively when imaged by AFM. The distributions of measurements are shown in Figures 4-24 and 4-25.

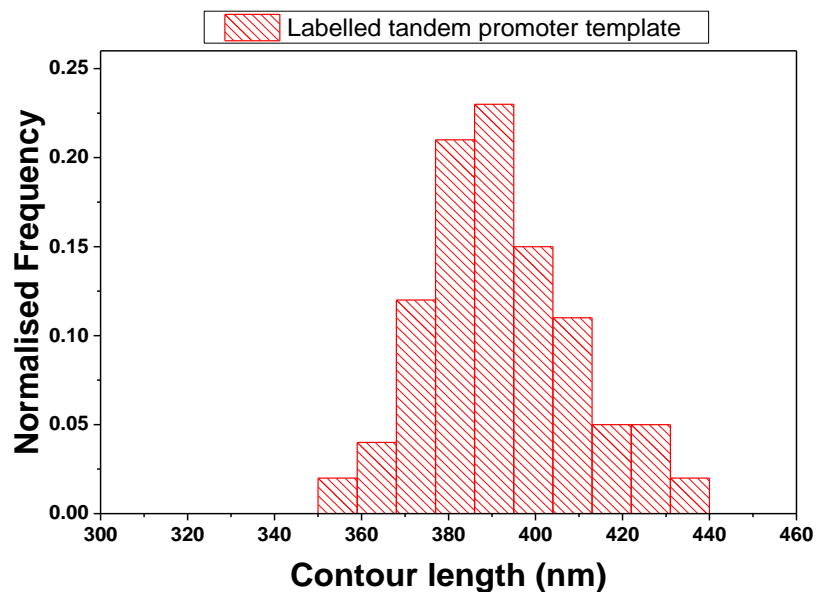


Figure 5-24: Histogram plot of labelled tandem promoter DNA. (n=100)

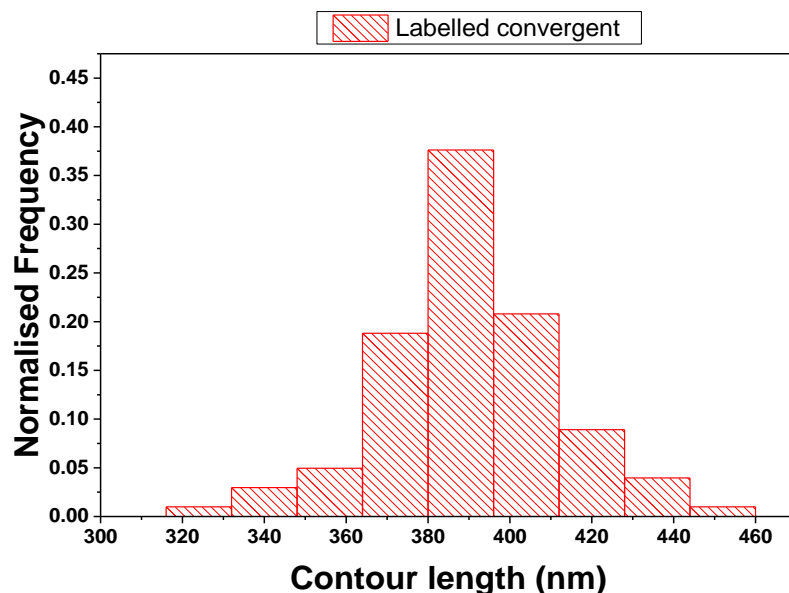


Figure 5-25: Histogram plot of labelled convergent promoter DNA. (n=101)

The percentage of molecules with a globular structure attached to the end was >70% for both tandem and convergent promoter fragments. The feature had an average height of 1.02 nm ( $\pm 0.12$ ) and a diameter of 18.56 nm ( $\pm 2.1$ ) for both fragments. This

similarity in size and appearance to that of the labelled single promoter template indicated that the PCR successfully amplified the 1144 bp two promoter DNA fragments with the ssDNA loop attached to the end of the template for both convergent and tandem promoter configurations.

## 5.4 Discussion

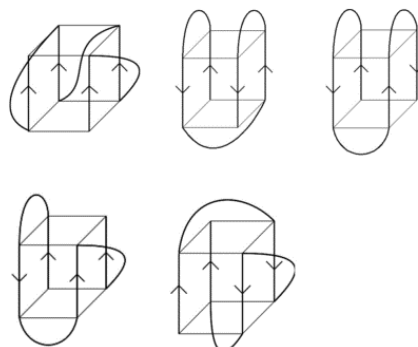
### 5.4.1 Labelling

The PCR based labelling reaction can be considered successful by the presence of a single band when products were analysed by gel electrophoresis, that when analysed by AFM had a globular structure attached to one end of the DNA. The use of a 5'-3' exonuclease deficient *Taq* polymerase meant that the PCR was able to amplify the parent strand of DNA without degrading the loop region of the primer. This result is novel in the fact that PCR primers are specifically designed to contain no secondary structures as this might interfere with the annealing and amplification stages of the reaction [363]. The use of a reduced DNA extension temperature meant that the loop was more likely to remain intact during the PCR, therefore preventing the *Taq* polymerase from reading through the loop region of the primer and producing the complementary sequence. The simplicity of PCR means that it is possible to perform this labelling reaction with any DNA fragment as long as time is taken to design primers and optimise reaction conditions to the template and prevent unwanted formation of secondary structures.

The method used showed a labelling efficiency of >70% which shows an improvement on the previous single round method used by Billingsley *et al.* which achieved a labelling efficiency of 48% [5]. There is also a marked increase in the concentration of labelled DNA produced leading to a higher yield after purification since this full PCR method amplifies the template. The use of different homo-polynucleotide loop sequences was able to show that for the poly(A, C, and T) there was no effect on the formation of the loop and amplification, confirmed by the similarity in yield, efficiency and appearance of loops in AFM imaging. For the poly(G) loop, a lower extension temperature was required in order to produce an appropriate globular structure. The original annealing temperature of 68°C only produced a labelling efficiency of 10%. The

### Chapter 5: High throughput labelling

reason for this lower level of loop formation was not investigated in depth but it has been shown that poly(G) stretches of DNA can form a number of structures that differ from the standard B-form DNA. The most common of these structures are G-quadruplexes, which are often seen in telomeric DNA [39]. Examples of possible structures that could be formed are shown Figure 5-26 [364, 365].



**Figure 5-26: Line drawings depicting possible structures that stretches of G bases can adopt, directionality is denoted by arrows. (adapted from [365])**

It was not possible to discriminate particular G-loop structures under the AFM but if these structures formed during the PCR, this could lead to amplification of unwanted regions of the loop, leading to a loop not forming correctly and giving an increased contour length and lack of end feature. A drop in extension temperature resolved this issue indicating that the lack of the end feature displayed temperature dependence. AFM imaging of G-quadruplexes has shown that they can form different structures such as spurs, blobs or no visible structure at all, but this was not investigated in this study [366].

The lack of a globular feature on ~30% of molecules analysed by AFM could be due to a number of reasons. It has been seen through a number of studies that hairpin loop structures have a relatively unstable nature. Hairpins have been shown to form and break over time scale of milliseconds. These fluctuations have been seen to happen for hairpins that are surface immobilised as well as free in solution [367-369]. The closed state has been shown to be stabilised by higher NaCl concentrations. Even though these fluctuations have mainly been tested with neck regions containing fewer base pairs than in our loop primers, the low NaCl concentration used could mean that

## Chapter 5: High throughput labelling

some of the fragments may have loops in an open structure when absorbed to the mica surface as well as during the PCR reaction. Upon binding to the mica the flexibility of the single stranded loop structure may mean that it is not tightly secured to the surface or may lay in an orientation that is not readily visible to the AFM. This is supported by studies performed by Ohta *et al.* on palindromic promoter elements from *Staphylococcus aureus HSP70* gene [370]. Using AFM analysis Ohta *et al.* were able to show that the stem loop formed by the promoter was visualised as a globular feature but only on 10% of templates analysed, indicating the inherent instability of these structures. Investigations into the stability of DNA hairpin loops suggest that loops are most stable when they contain 4 to 5 nucleotides and stability is lost for loops larger than this [371]. It is also a possibility that when adsorbed to the mica surface the orientation of the loop is such that it cannot be elucidated by the AFM tip.

The size and appearance of the loop is also consistent with previous AFM studies of hairpin loops. Billingsley *et al.* used a loop of 20 nt to end label a DNA template and reported a diameter of  $20.5 \pm 0.5$  nm and a height of  $1.05 \pm 0.05$  nm which is consistent with the average height and diameter for all loop sequences measured in this study ( $1.05 \pm 0.02$  nm and  $20.56 \pm 0.14$  nm respectively) [5]. The structures in Ohta *et al.*'s study were composed of an 11 nt loop which displayed a diameter of  $7 \pm 2$  nm and a height of  $2.2 \pm 0.2$  nm [370]. This observed diameter for a loop of half the number of nucleotides supports the theory that the feature observed is a single stranded loop. The height of the loop in Ohta *et al.*'s studied is approximately double the height, but this is expected as they believed that two stem loops were lying on top of each other, which in turn would give a greater height [370]. A height greater than the backbone of the DNA for hairpin loops was also noted by Duzdevich *et al.* when studying CAG repeats related to Huntington's disease [372]. Upon AFM analysis of samples regions of greater height were noted extruding from the DNA backbone and were confirmed as small hairpin loops by Mung bean digestion assay.

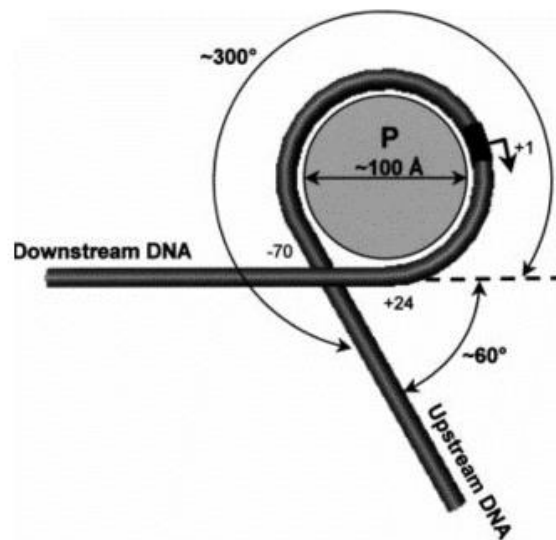
### 5.4.2 Effect of loop label on transcription

The labelling of the DNA was done to provide a polarity marker for analysis of transcription complexes under AFM. Upon formation of OPCs there were two distinct



### Chapter 5: High throughput labelling

classes of molecules. Those that had an RNAP located along the DNA backbone and those with an RNAP located at one end of the template. Those that had an RNAP located along the backbone of the DNA showed a contour length reduced by approximately 33 nm for all loop sequences. This reduction in contour length is expected to occur when the RNAP binds to its promoter and wraps the DNA as can be seen in Figure 5-27 taken from Rivetti *et al.* [126]. This wrapping of the DNA has been shown to involve approximately 90 bp of the DNA and Rivetti *et al.* reported a reduction of 32 nm for the formation of an OPC at a  $\lambda_{pr}$  promoter [126, 294]. This reduction in contour length along with the position of the RNAP determined by arm length measurements indicate that these are OPCs.

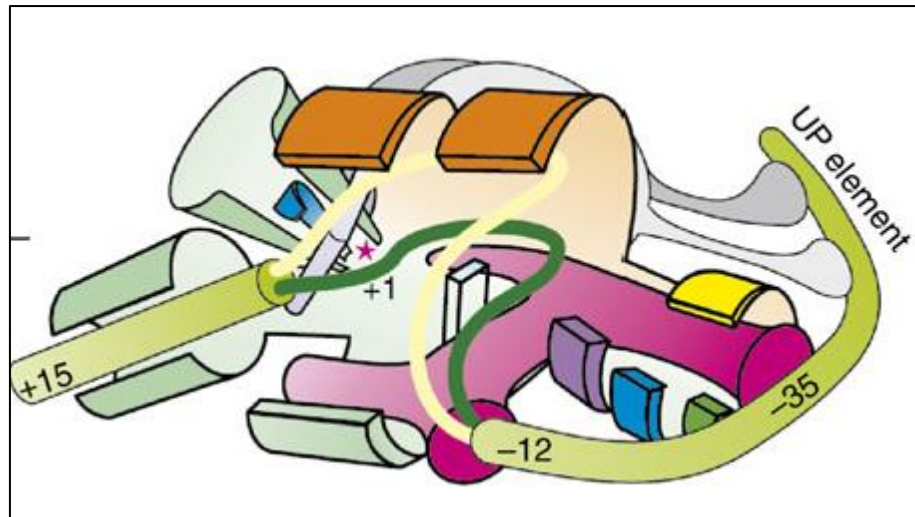


**Figure 5-27: Diagram of DNA wrapping in the formation of an OPC. The DNA wrapped around the protein, which is depicted as a sphere, the promoter element and the upstream and downstream elements involved in the wrapping are shown. The angle by which the DNA is seen after wrapping is also noted on the diagram [126].**

This wrapping also leads to a distinct bend in the DNA at the binding site which was also seen in the complexes analysed. The presence and position of the loop on the longer arm confirm that the loop label has attached to the desired arm.

The second class of molecules with an RNAP bound to the end of the template did not display this reduction in contour length. As no end feature was visible on the opposing

end of the fragment it is reasoned that the RNAP was bound to the loop label. This interaction is most likely due to a non-specific binding of the RNAP to the single stranded region of the loop. This binding is not completely unexpected as end binding of RNAP to DNA has been noted before but it is common practice to discount molecules that show end binding [294, 373]. Due to the large population of complexes that displayed this end binding (50% of complexes measured) for all of the loop sequences, it was considered significant.  $\sigma$ RNAP and core enzyme have both been shown to have affinity for ssDNA as well as displaying some affinity for binding to hairpin loops [374, 375]. This affinity for ssDNA could be due to the increased flexibility of the DNA which is a common property of DNA-protein binding sites as well as these regions resembling melted DNA [110, 376]. Experiments investigating the binding of *E.coli* RNAP to templates containing single stranded bubbles have shown that binding can occur regardless of sequence, but is more likely for the core enzyme alone rather than the holoenzyme [377]. Investigations by De Haseth *et al.* and Huang *et al.* were able to show that binding of RNAP had a higher affinity for ssDNA than for promoter-less dsDNA [375, 378]. Studies have reported through competition assays that this binding occurred in a non-specific manner but was more stable than  $\sigma$ RNAP bound non-specifically to dsDNA [375, 376, 378]. When looking at the structure of an OPC as shown in Figure 5-28 it can be seen that a ssDNA loop would be able to fit into the binding site of the protein. These interactions are believed to be mediated through electrostatic interactions with the negative ssDNA which can mimic polyanionic inhibitors of RNAP such as heparin. It is also possible that the holoenzyme may interact through base interactions mediated by the sigma factor which contains a single stranded binding region in region 2.3 of the sigma factor which is accessible when bound by the RNAP core enzyme [378, 379]



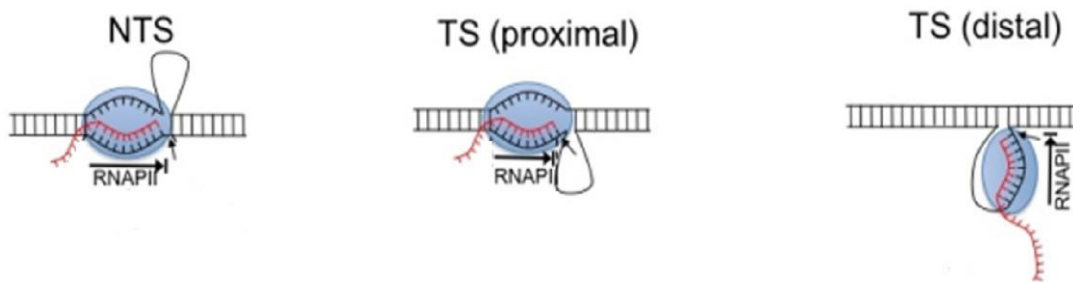
**Figure 5-28: Figure showing a schematic diagram of an OPC. The sigma factor is magenta, dsDNA is in green, the template strand is in dark green and the non-template strand is in cream. The binding region that binds to the non-template strand is in orange. Adapted from [380].**

This binding may also be further assisted by the ssDNA binding sites within the  $\sigma$ -factor [379]. Callaci *et al* performed experiments on the binding of holoenzyme to ssDNA. It was observed that holoenzyme bound to non-specific DNA had a dissociation constant of 11  $\mu\text{M}$ , to randomized non-template strands with a dissociation constant of 17.6  $\mu\text{M}$  and a non-template strand with a dissociation constant of 0.5  $\mu\text{M}$  [379]. This is in comparison to the observed dissociation constant of 10.6  $\mu\text{M}$  for holoenzyme binding to template strand [379]. When considering the results after the addition of NTPs it is evident that the RNAP is not able to dissociate from the loop structure.

After the addition of NTPs, the spread of contour lengths measured was greater than expected. As previously mentioned this is most likely due to the presence of a number of species within the population. As a high number of templates had at least one RNAP bound to the end of the template and no loop visible it is assumed that elongation has occurred into the loop and the RNAP is unable to dissociate from the template after transcription. This is further confirmed by the RNA analysis performed on a C-loop labelled template. The reduction in the level of RNA produced indicates that there is undetectable levels of run off transcription and re-initiation. The stalling at the loop structure could be due to a number of reasons. Zhou and Doetsch were able to show

Chapter 5: High throughput labelling

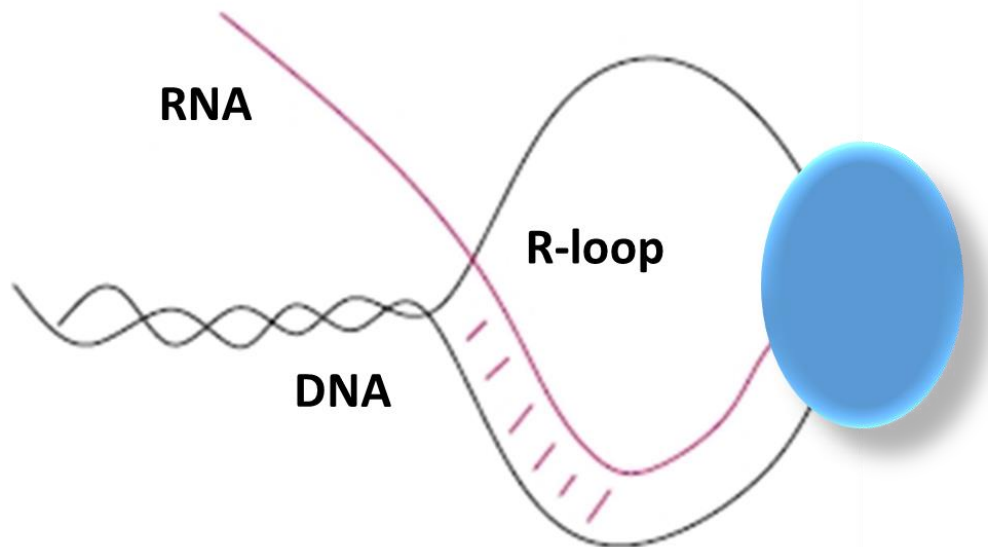
that when the transcription template contained a single stranded nick that *E.coli* RNAP was only able to produce truncated transcripts, indicating that the RNAP is unable to elongate past the nick [381]. The labelling reaction did not involve a subsequent ligation so a single stranded nick would be left in the template on the template strand of the DNA, which may cause the stalling at the loop. The loop itself may also lead to the stalling. In the case of T7 RNAP stalling at cruciform and hairpin structures was not noted, but stalling was seen for the eukaryotic RNAP II [382]. The stalling seen was reasoned to lead to one of the three situations shown in Figure 5-29.



**Figure 5-29: Consequences of hairpin loops or slipped strand structures on RNAP elongation. Each case represents a different scenario, the left panel shows a loop on the non-template strand (NTS) and the middle and right panel show consequences of a loop on the template stand (TS). Adapted from [383].**

The AFM data collected indicates that if the loop is causing the stalling then the template strand distal mechanism is most likely. This is due to the fact that for stalled elongation complexes (SECs) Rivetti *et al.* showed that DNA templates displayed a shortening of 22 nm believed to be due to some wrapping being maintained in elongation complexes [294]. This shortening of the template is not recorded in the data which indicates that the RNAP has lost some or all of its wrapping. If the RNAP has stalled “on” the loop then this shortening would not be expected to be as pronounced. This may also account for the formation of complexes that have a greater contour length than that seen for the bare DNA as the diameter of the RNAP is larger than the loop and could therefore lead to a measurement of a longer contour length. In the case of two RNAPs located at the loop after elongation, it is also possible that co-operative action of a second RNAP could lead to shunting of the loop bound RNAP leading to extension of the loop structure, giving a larger contour length. The final

mechanism by which stalling may occur is the formation of an overly stable RNA:DNA hybrid between the homo-polynucleotide DNA and the RNA transcript. This structure would be similar to those referred to as R-loops (Figure 5-30) [384].



**Figure 5-30: Formation of a hybrid between the DNA and RNA transcript which can lead to stalling of elongation. These structures are found at points and times when there is opening of the DNA duplex, and so can be expected to be more likely if transcription of the loop itself was to occur.**

R-loops have been shown to lead to arrest and stalling of RNAPs *in vitro* but are normally found at regions of high negative supercoiling that allow for invasion of the RNA transcript into the DNA template [384, 385]. In the case of a single strand loop, negative supercoiling would not be required as the RNA would be able to easily access the DNA and base pair to form an RNA:DNA hybrid.

The success of the labelling and investigation into the effects of different loop sequences helped inform the protocol for labelling the two promoter templates. The choice of a poly(C) loop was intended to reduce any similarity to promoter sites, which are often AT rich. The reaction was shown again to be successful producing a high yield of labelled DNA. The production of dimers was unexpected as this was not seen with the single promoter fragment but was not considered an issue as these dimers were rare and easily removed by gel purification. The template length in base pairs before

addition of the loop is 1144 bp which gives an expected contour length of 377.5 nm assuming a base pair rise of 0.33 nm/bp. With the loop attached the contour length is 13.8 nm greater than this length. The neck region of the loop adds 12 bp to the template which would be approximately 3.9 nm. If this length is subtracted from the radius of the loop it gives a radius of approximately 10 nm as was recorded for loops attached to the 602 bp DNA template and by Billingsley *et al.* [5].

## **5.5 Conclusions**

The aim of this work was to design a method for incorporating a nucleotide based end label onto DNA fragments that was high throughput and produced high yields. The increase in yield over previous methods as well the increased labelling efficiency shows that this has been achieved. The method has been shown that it can be applied to different templates using loops of all four nucleotides. The structure of the loop does not show any change with the nucleotide used to form it. The only difference noted between sequences was for the poly-G loop which required a lower extension temperature during the PCR to produce a readily visible loop under AFM analysis.

The label itself was easily distinguishable from the DNA backbone as well as from RNAP upon formation of OPCs, meaning that the polarity of the DNA can be easily recognised under AFM analysis. The RNAP showed a high affinity for binding to the loop structure which was unexpected and is believed to be a non-specific interaction due to a lack of wrapping of the DNA. The loop proved to be an insurmountable blockage to elongation by RNAP leading to stalling of the RNAP at the loop. This is not unexpected but the exact mechanism by which this occurs could not be fully determined.

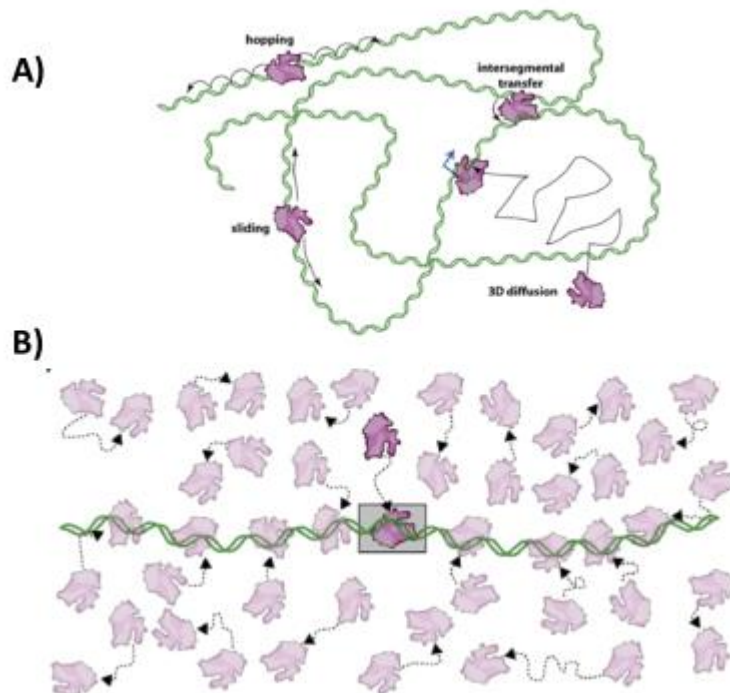
The high levels of non-specific binding to the loop are not unexpected as previous AFM studies of RNAP have all shown that when imaging, a large number of complexes have RNAPs bound outside the promoter in a non-specific manner. In order to address this issue a method for incorporating non-specific inhibitor is presented in Chapter 5 with data for the transcription of labelled templates possessing either tandem or convergent promoters.

This page is left intentionally blank

# 6 Convergent and tandem transcription in the presence of a non-specific binding inhibitor

## 6.1 Introduction

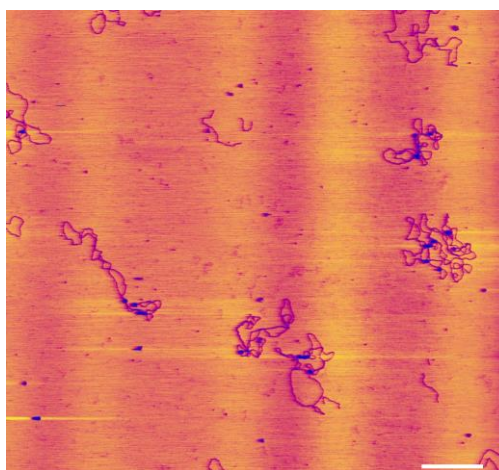
RNA polymerase holoenzyme ( $\sigma$ RNAP) must first locate its promoter to form a closed complex and melt the DNA helix before being able to form a stable open promoter complex (OPC). The mechanism by which RNAP locates its promoter has been shown to be through three main mechanisms: hopping; sliding; and inter-segmental transfer (Figure 6-1). There is also some evidence that under certain conditions it may undergo 3-dimensional diffusion which is also shown in Figure 6-1 [112].



**Figure 6-1: Diagrams displaying the suggested mechanisms which RNAP uses to locate its promoter, RNAP is shown in purple and dsDNA in green. A) Shown are the four possible mechanisms of promoter location hopping, sliding, inter-segmental transfer and 3D diffusion. B) Diagram depicting how concentration of the protein relative to its target can drive 3D diffusion [112].**

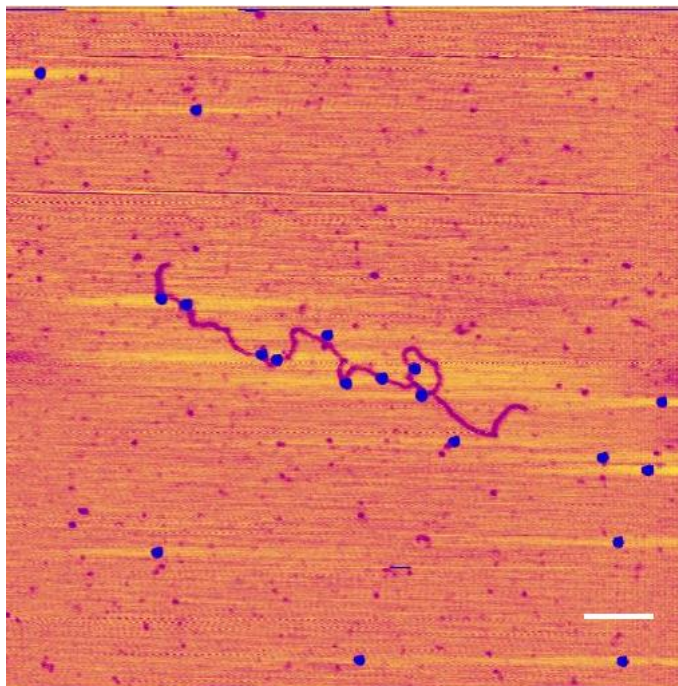


These mechanisms have been modelled and observed in a number of studies [292, 386-388]. In all these scenarios, there is a need for the RNAP to make contact with the DNA. This contact means that the RNAP is able to use facilitated diffusion to locate its promoter and this view is widely supported even though recent studies suggests this may not be the case [112, 386, 388-390]. No matter which process leads to  $\sigma$ RNAP locating its promoter there is need to contact the DNA in a non-specific manner at some point. The nature of this contact between the  $\sigma$ RNAP, RNAP and DNA has not been fully elucidated but is believed to be mediated by electrostatic interactions and influenced by ionic concentration [112, 391]. Even though these interactions are non-specific, they have been shown to have lifetimes of up to 3.3 seconds in bulk aqueous liquid and 600 seconds when complexes are deposited onto a surface such as mica, and have dissociation rates ranging from  $3.0-8.4 \text{ s}^{-1}$  [106].  $\sigma$ RNAP and RNAP have also been shown to have some affinity for blunt ends of DNA and from the labelling study presented in Chapter 5 the RNAP clearly displays affinity for ssDNA. AFM provides only a topographic image of the sample surface and so it is not always possible to distinguish the binding conformation of the protein. This can be highlighted by looking at images of plasmid DNA containing two  $\lambda_{pr}$  promoters, incubated with RNAP holoenzyme as shown in Figure 6-2.



**Figure 6-2: AFM height image of plasmid molecules incubated with RNAP. It can be seen that the plasmids are highly decorated by RNAP molecules even though the plasmid contains only two promoters. Multiple protein-protein interactions makes further analysis difficult. (Scale bar = 500 nm)**

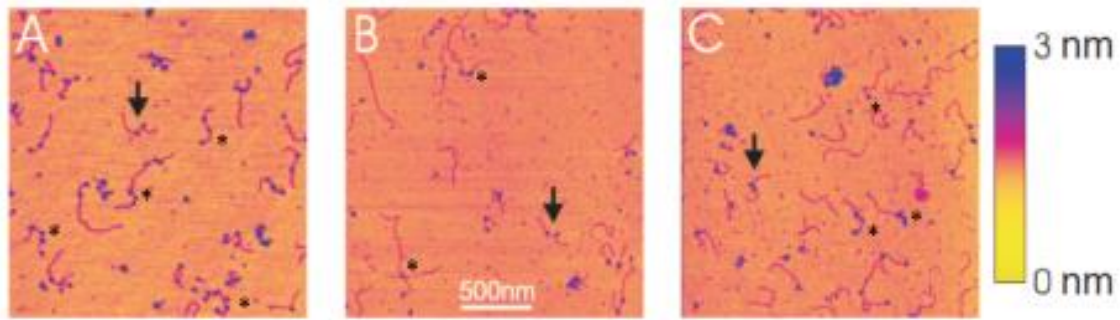
As it can be seen the plasmid DNA has a high number of RNAPs associated with it, making analysis difficult, if not impossible. The presence of RNAPs on the DNA may not only be due to non-specific binding but also due to deposition of RNAP in close proximity to or on top of the DNA which may be further accentuated by protein-protein interactions. In a sample of OPCs it is possible to separate out non-specific binding from specific binding by analysing the bend in the DNA, change in contour length and template lengths but if the DNA has more than the expected number of RNAPs bound this can lead to analysis being unfeasible, due to the DNA being obscured from the tip. An example of a linear DNA fragment of 6136 bp with two promoters is shown in Figure 6-3, where it can be seen that those  $\sigma$ RNAPs which might be promoter bound cannot be distinguished from those that are non-specifically bound. The use of contour length measurements and bend angle measurements can help indicate those that lie at a promoter, but in such a scenario uncertainty is high.



**Figure 6-3: AFM height image of 6136 bp linear DNA containing two promoters. As it can be seen there are a number of RNAPs bound along the DNA template, even though bend angles and length can be measured these additional RNAP add uncertainty and errors to measurements (Scale bar = 200nm)**

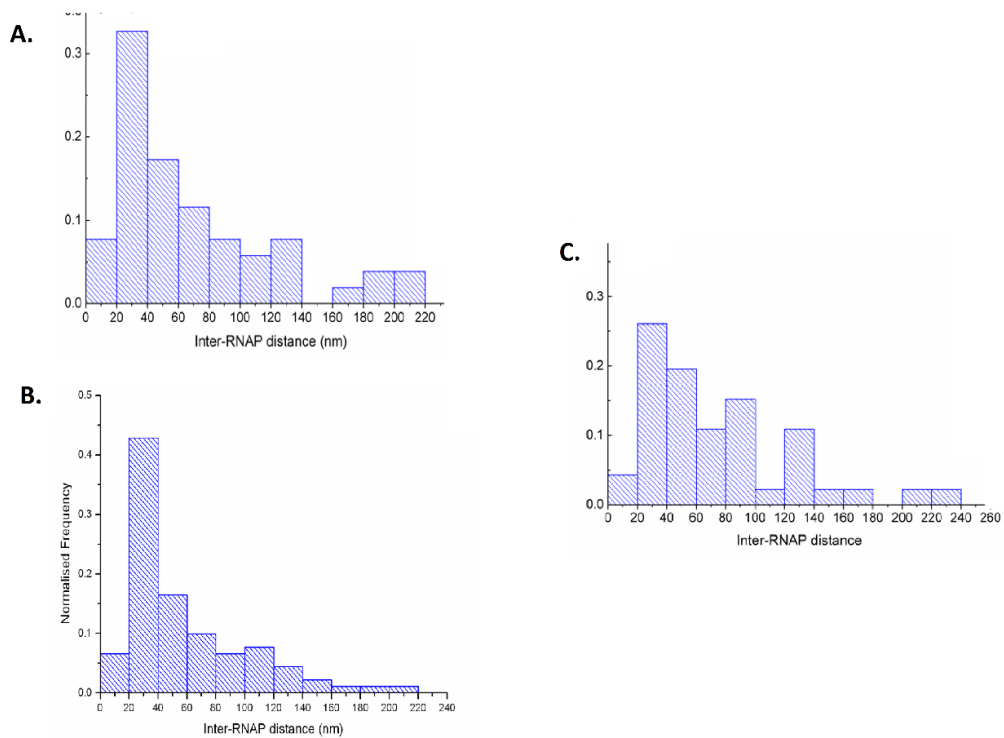
With shorter DNA fragments this non-specific binding may not be as prevalent and so molecules that have more than the expected number of RNAPs associated with the DNA can be discounted, but this means that the number of complexes analysed is considerably less and analysis is overly time consuming. After the addition of NTPs, elongation occurs and the non-specific binding has a more profound effect on analysis. Once the RNAP has undergone elongation, DNA compaction and bend angles are less pronounced meaning that non-specific bound RNAPs are harder to discern. There is also the possibility that after elongation, if a RNAP reaches the end of the template and “runs off”, the RNAP may be recycled and start a new search for its promoter. This theory is supported by evidence that the  $\sigma$ -factor can be recycled and lead to the formation of new holoenzyme that is capable of undergoing a new promoter search and subsequent rounds of transcription, as well as by recent studies indicating that the  $\sigma$ -factor may remain bound for certain RNAPs [135, 392]. It is also possible that those  $\sigma$ RNAPs that are still free in solution may subsequently bind to a promoter leading to multiple initiation and elongation events as well as there being evidence that RNAP core enzyme is able to initiate transcription from blunt ends as well as from single stranded nicks.

Previous studies on similar DNA fragments as studied in this thesis, carried out by Crampton and Billingsley did not use an inhibitor to prevent non-specific binding and re-initiation in AFM samples as this was the standard approach [194, 293]. In images taken from [3] shown in Figure 6-4 it can be seen that non-specific binding was prevalent, limiting the number of analysable complexes.



**Figure 6-4: AFM height images of transcription complexes taken from Crampton *et al.* The images show that analysis is possible but high levels of non-specifically bound RNAP are seen with the black arrows indicating complexes with the desired number of RNAPs bound and the asterisks indicates those where an  $\sigma$ RNAP is bound to a blunt end of the template [3].**

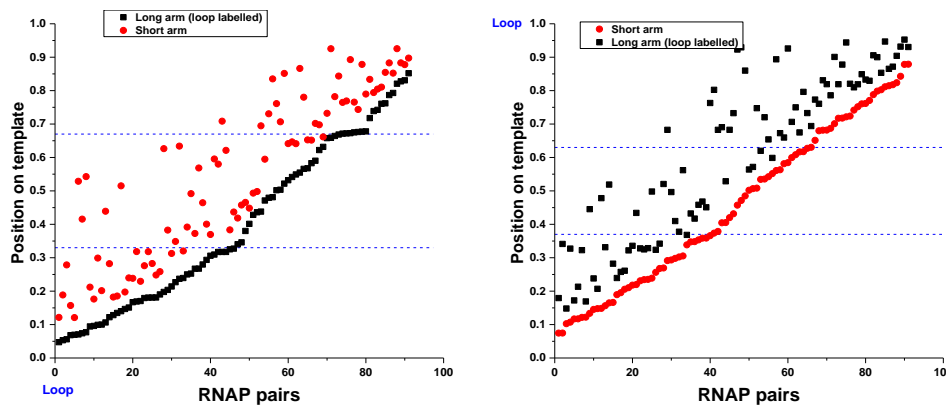
For both studies performed it was noted that RNAPs could not pass each other as was further shown by Billingsley *et al.* through end-labelling of the DNA. The data indicated that there was a high occurrence of long range backtracking or shunting seen for up to 70% of complexes [4]. The distribution of the separation between RNAPs after elongation for unlabelled and labelled convergent promoter DNA templates and tandem promoter templates are shown in Figure 6-5.



**Figure 6-5: Histogram plots of the inter-RNAP contour length taken from *Billingsley et al* [293]. It can be seen that for all cases the histograms display a range of the inter-RNAP contour lengths with a peak at approximately 40 nm. A) Inter-RNAP contour length for convergent promoter DNA after elongation (n = 62). B) Inter-RNAP contour length for tandem promoter DNA after elongation (n = 65). C) Inter-RNAP contour length for labelled convergent promoter DNA after elongation (n = 91).**

As it can be seen from the graphs shown in Figure 6-5 the distance between RNAPs has a negatively skewed distribution with a range of distances present. The distribution of distances and locations of the RNAPs on the template can be explained by backtracking events but without any method of distinguishing from RNAPs that are bound non-specifically or currently undergoing a new transcription cycle to those that have undergone elongation, this conclusion cannot be confirmed with a high degree of accuracy. The large range of distances could be explained by non-specific binding due to its random nature. The distribution of RNAPs pairs on the template also displayed a random nature and this can be seen when the positions of RNAP pairs were plotted as a percentage of the total template contour length (Figure 6-6). The contour lengths of

the designated arm (short or long) to the nearest RNAP were organised in ascending order and the position of the second RNAP was plotted in relation to the first RNAP. This therefore provides a plot displaying the distance of each RNAP from each arm and the distance between each RNAP pair.



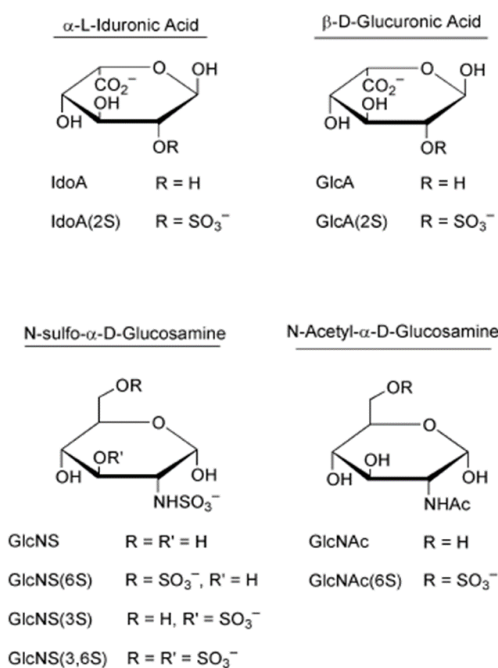
**Figure 6-6: Scatter plots of RNAP positions after elongation from convergent promoters collected by Billingsley *et al.* The RNAP originating from the long arm promoter (loop labelled arm) is shown in black and the RNAP originating from the short arm promoter in red. The left plot shows the RNAPs originating from the long arm promoter plotted in ascending order from the loop to the short arm while the plot on the right shows RNAP originating from the short arm promoter plotted in ascending order from the short arm to the long arm [4]. The distribution of RNAP pairs on the template can be seen to have large number of RNAPs located upstream of the promoters, which was believed to have occurred due to shunting of an inactive RNAP by an active RNAP.**

The plots show that a large number of RNAP pairs are positioned upstream of the promoters as stated by *Billingsley et al.* but the distribution of RNAP pairs over the template is also quite random in nature. This uncertainty in the validity of measurements requires a further examination of collision events in the presence of an inhibitor of non-specific interactions.

The occurrence of crowded DNA fragments due to overlaying of protein and DNA can be overcome by ensuring that samples are sufficiently diluted before deposition, allowing for protein and DNA-protein complexes to be evenly distributed over the

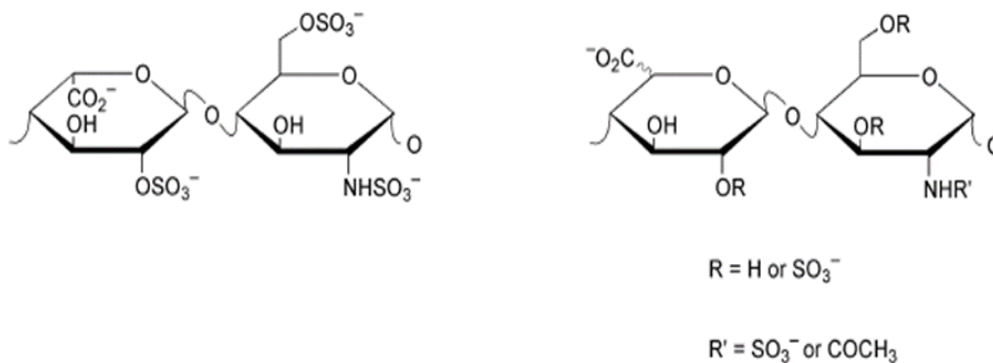
surface. The problem of non-specific binding and re-initiation is more complex. There are number of methods to reduce and prevent non-specific binding of proteins, but due to the sensitive nature of AFM and surface deposition kinetics these methods are not all viable for producing samples suitable for AFM analysis. One of the simplest methods to reduce non-specific interactions is to increase salt concentration or ionic strength of the buffers used which leads to decrease in the net electrostatic potential of the DNA. It is noted that the rate of promoter binding decreases with increased salt concentration [393]. This change would mean that for *in vitro* transcription reactions to be analysed by AFM there would be a low number of OPC for analysis. The effect of monovalent salts at high concentration can also alter the binding of the DNA and DNA-protein complexes to the mica surface and so may not be feasible for use in AFM [103, 394-396]. This would also not solve the issue of multiple rounds of transcription occurring. The most common method used in biochemical assays is the addition heparin to samples after the formation of OPCs.

Heparin is a polyanionic polysaccharide of the glycosaminoglycan family which includes the closely related macromolecule heparan sulphate (HS). Both heparin and HS are linear polysaccharides made up of the same monosaccharides shown in Figure 6-7.



**Figure 6-7: Monosaccharide building blocks that make up heparin and HS.**

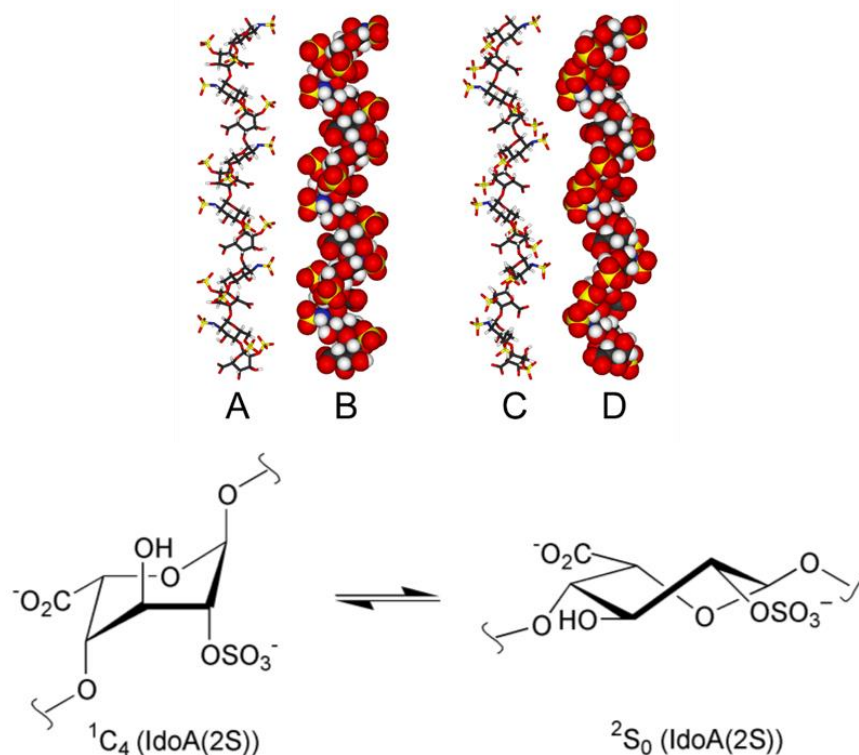
The overall structure of heparin and HS can vary due to their non-templated production [397, 398]. Both heparin and HS are produced in the same manner and are made up of two repeating disaccharide units (Figure 6-8).



**Figure 6-8: Repeating disaccharide units that make up heparin and HS.**

Heparin is produced by mast cells and has a molecular weight range of 60-100 kDa but when purified for biochemical uses has a size distribution of 12-15 kDa as purified from porcine intestine. HS is produced by all cell types and is produced attached to a protein core to form a proteoglycan. Free chains of HS are rarely found *in vivo* but can be purified from bovine kidney cells free of the attached protein. The molecular weight of HS has a similar range to that of heparin, but the average molecular weight of purified chains is slightly higher at approximately 20 kDa and is less well characterised than heparin. The major biochemical difference between HS and heparin is the number of GlcN-sulphate groups that occur. Porcine mucosa heparin chains have approximately 88.6 sulphate groups per 100 disaccharide units with a ratio of GlcNS to GlcNAc of around 4 to 1 [398]. HS purified from bovine kidney has 30.8 N-sulphates per 100 and ratio of a ratio of GlcNS to GlcNAc of around 1 to 1 [398]. Both molecules have homologous structures and often are considered to display the same properties, and used as models of each other when necessary. Heparin and HS chains adopt one of two right handed helical structures with a 1.63-1.73 nm tetra saccharide sequence shown in Figure 6-9.





**Figure 6-9: Stick and space fill models of the two main conformations adopted by heparin and HS. A and B are the conformation adopted when all the 2-O-sulfo- $\alpha$ -L-iduronic acid groups are in a  ${}^2S_0$  conformations. C and D show the structure adopted when the 2-O-sulfo- $\alpha$ -L-iduronic acid groups are in a  ${}^1C_4$  arrangement. Below is shown the iduronic acid in the two different conformations. Image adapted from reference [399].**

As is shown in Figure 6-9 heparin and HS can adopt different conformations. These refer to the pyranose ring of the iduronic acid [398]. The iduronic acid pyranose ring exhibits conformational flexibility and this is one reason why both heparin and HS are able to bind such a vast number of proteins [398]. Heparin and HS have a number of functions *in vivo* and heparin is a common pharmaceutical agent. Heparin is predominately used as an anticoagulant, but due to its varying structure can perform a wide range of functions [397]. The use of heparin for *in vitro* transcription assays is due to the fact that its structure is similar to that of DNA as well its polyanionic nature. Both DNA and heparin/HS have a negatively charged backbone and adopt a helical structure. Heparin/HS have a residue rise of 0.4nm in comparison to 0.34nm for DNA. Both are able to mimic DNA and bind to RNAP via its DNA binding domain located in the active site [400]. The similar structure of HS means that it is also able to bind to

DNA binding proteins in the same manner but is not commonly used due to its less well characterised chemical composition.. The binding of heparin/HS occludes the DNA from the active site therefore preventing the formation of OPCs. If an RNAP has already formed an OPC, then heparin/HS are unable to bind as they cannot gain access to the binding site [400]. This means that heparin/HS are able to bind free RNAP molecules that have not formed OPCs, non-specifically bound  $\sigma$ RNAP, and RNAPs that have undergone elongation and are recycled for subsequent rounds of transcription [142, 400, 401]. This sequestering of free and non-specifically bound RNAPs has meant that heparin is used as a competitor in *in vitro* transcription reactions to ensure that only RNAPs that have formed OPCs, and therefore those that are actively transcribing, are being studied. The prevalent use of heparin as a competitor in biochemical assays suggests that heparin would be a good choice for AFM experiments, solving non-specific binding events and preventing re-initiation and multiple rounds of transcription. However, addition of heparin into AFM samples has not been performed to date due to the potential risk of heparin interfering with the binding of DNA to the surface, with researchers preferring to perform bulk biochemical assays only with heparin present [127]. Heparin is believed to inhibit the binding of DNA to the surface which is a reasonable assumption due to its highly anionic structure. The use of HS instead of heparin has not been investigated to date but it is expected to exhibit similar effects, due to its similar structure and chemistry. HS has a lower level of sulphonation as well a different chain length and so may interact with the mica surface in a different manner. As both are capable of inhibiting non-specific binding of RNAP both were considered for incorporation into samples.

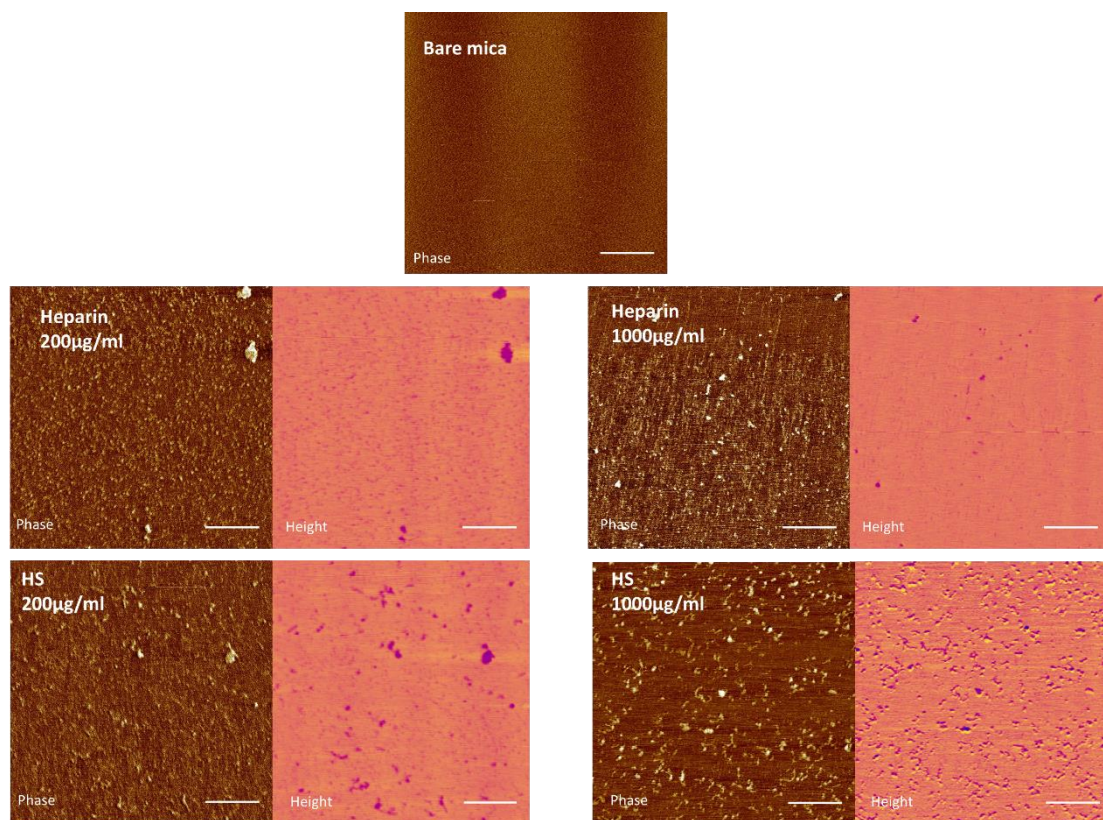
Some studies of transcription have utilised heparin attached to Sepharose beads [319]. Even though this method would remove any non-specifically bound RNAPs after formation of OPCs, it would not prevent re-initiation events due to the lack of a competitor in the reaction during and post transcription initiation.

## 6.2 Results

### 6.2.1 Incorporating heparin into AFM samples

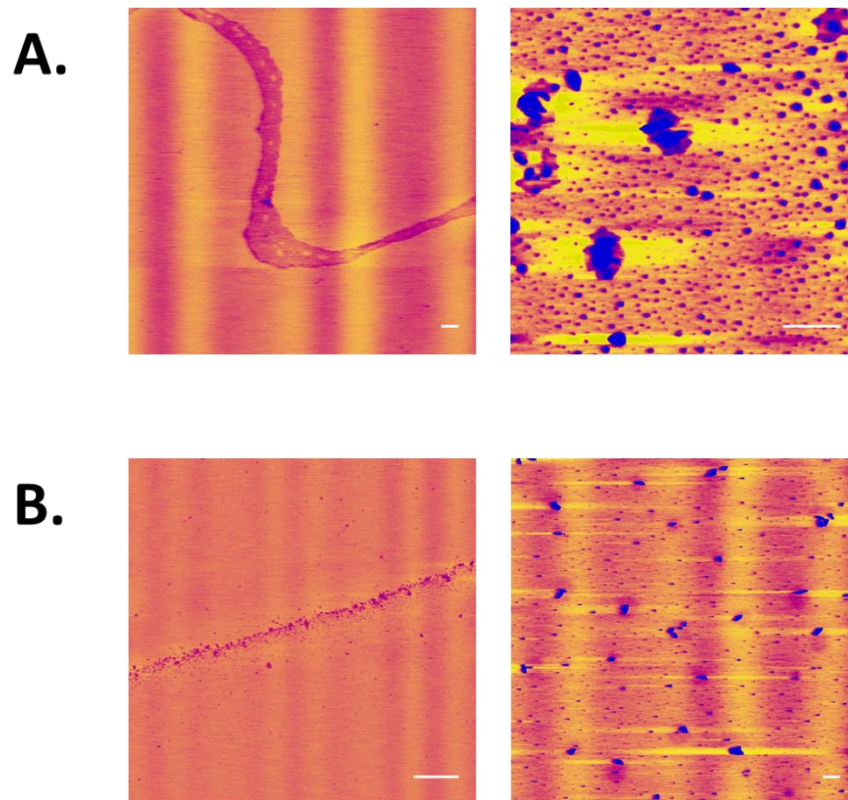
Heparin or HS have not been widely used for AFM samples due to a reasonable hypothesis that the highly negative charge of polysaccharide chains will compete with or prevent binding of the DNA to the mica surface. Rivetti *et al.* utilised heparin in samples for investigations into the wrapping of DNA by RNAP upon formation of OPCs even though in later papers heparin was considered to not be compatible with AFM analysis [126, 127]. The other potential disadvantage of using heparin or HS free in the sample is that RNAP not bound to DNA will remain in the solution even when inhibited and so may still overlay the DNA or contact the DNA.

Rivetti *et al.* used 200 µg/ml heparin in OPC samples containing 600 fmol of RNAP holoenzyme in transcription buffer to ensure heparin was in molar excess [294]. This protocol was the basis of the protocol used here. The appearance of heparin and HS on mica was first investigated as a control. Heparin or HS was added to 10µL of transcription buffer at final concentration of 200µg/ml or 1000µg/ml before being diluted 1 in 10 in imaging buffer. This was then deposited onto mica and incubated for 3 minutes, rinsed with dH<sub>2</sub>O and dried with nitrogen before imaging. Both heparin and HS were found bound to the surface with heparin having a greater coverage of the surface, while HS was seen to be dispersed over the surface (Figure 6-10).



**Figure 6-10: AFM phase and height images of heparin and HS deposited on mica. It can be seen that heparin is not visible in the height image when at high concentration but can be seen in the phase image. The appearance of heparin and HS differ with HS having a more dispersed appearance (Scale bar = 200 nm).**

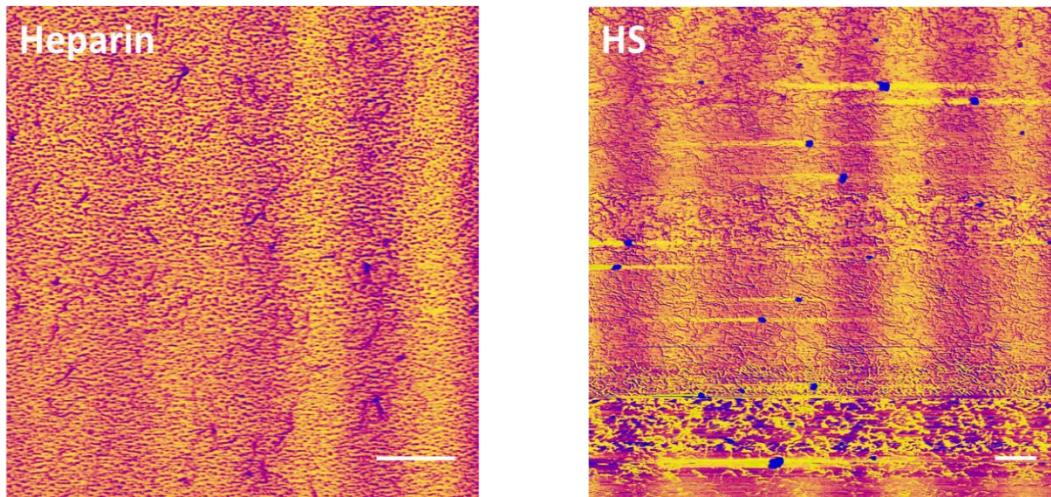
The binding of heparin and HS to the mica is not surprising as with DNA, the  $Mg^{+}$  ions most likely help to bridge between negative charges [270]. The difference in appearance between the two molecules may be due to the levels of charge or the differing chain lengths and flexibility. Next the influence of heparin and HS on the binding of DNA to mica was investigated. 200 fmol of DNA was added to 10 $\mu$ L of transcription buffer with heparin or HS at final concentration of either 200 $\mu$ g/ml or 1000 $\mu$ g/ml. These were then prepared for imaging as in previous samples. Upon imaging it was found that the surface with both inhibitors displayed a number of different morphologies. Some of these appeared random in nature with no DNA visible.



**Figure 6-11: AFM height images of random structures formed by DNA and heparin (A) or DNA and HS (B) deposited onto mica. The images on the left were taken at a concentration of 200µg/ ml while the images on the right had 1000µg/ml of heparin or HS (Scale bars = 500 nm).**

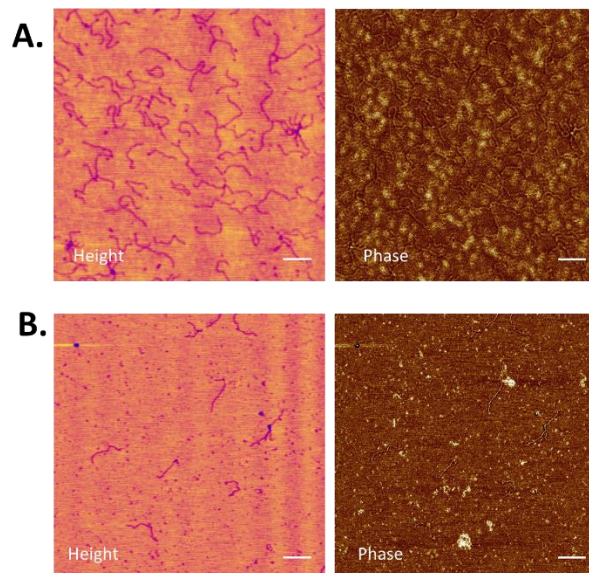
These areas were randomly dispersed over the surface of the mica, and may be formed due to the DNA not equilibrating onto the surface correctly and then upon drying, becoming condensed. Areas containing regions of closely packed DNA were seen for both HS and heparin at 1000µg/ml with what was assumed to be a network of DNA and heparin or HS (Figure 6-12). The DNA in the HS samples was much more visible but these structures were not securely attached to the surface and were easily detached by the motion of the tip.





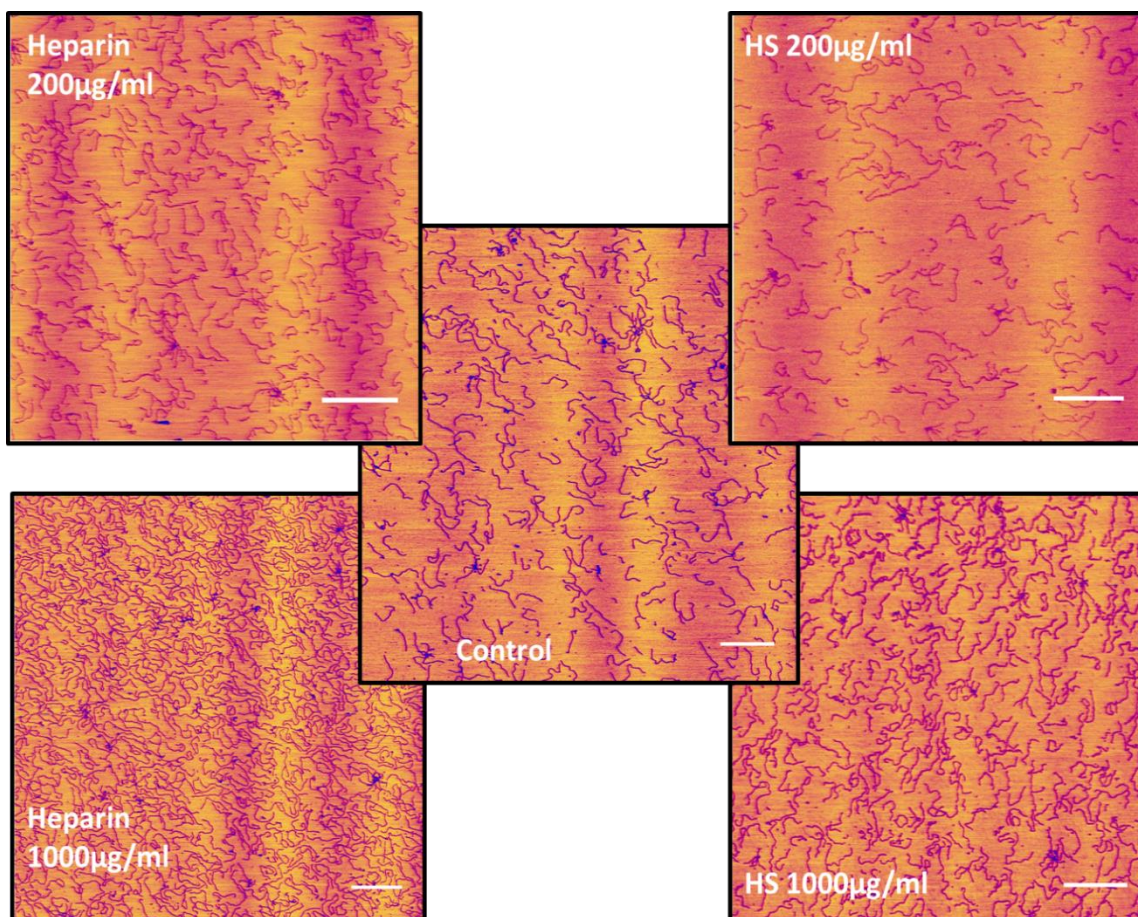
**Figure 6-12: AFM images of DNA with heparin or HS deposited on mica where closely packed DNA can be seen. The quality of the images is low due to the difficulty in imaging the loosely attached structures (Scale bars = 500 nm)**

This morphology could be caused by the DNA lying on top of areas of adsorbed HS. The DNA can be seen in the height and phase images but are not easily analysed due to the high surface roughness. At low concentrations of heparin and HS some areas of the mica did have visible DNA bound to the surface (Figure 6-13). These areas were more common and clearer with HS added as compared to heparin.



**Figure 6-13: AFM height and phase images of areas where DNA was visible but not clear. It can be seen from the phase images (right) that for both heparin (A) and HS (B) containing samples, the DNA has what appears to be the inhibitors bound to the mica surface around or underneath the DNA. (Scale bars = 200 nm)**

The final regions seen were those where DNA was clearly visible and appeared similar to control samples of DNA (Figure 6-14). These areas did show a higher density of DNA than the control sample and this increased with HS or heparin concentration. The amount of DNA seen in these regions was greater for heparin as compared to HS.



**Figure 6-14: AFM height images of regions of closely packed but visible DNA seen for samples with heparin or HS. A control sample containing the same amount of DNA but no inhibitor molecule is shown in the centre (Scale bars = 500 nm).**

Next, the effect of heparin on adsorption of RNAP holoenzyme alone was tested. Samples were prepared containing 400 fmol of RNAP holoenzyme and 200µg/ml of heparin or HS in 10µL transcription buffer. These were diluted 1 in 10 in imaging buffer before being deposited and dried for imaging. It was seen that with the addition of both heparin and HS the number of small aggregates of RNAP decreased. Molecules of RNAP were counted using the particle detect feature of the Nanoscope software. This features detect particles by using the pixel height. An upper limit of 35 nm for

diameter was used to determine aggregates. The software was used to detect any features above a height of 2 nm. After detection molecules that had not been detected by the software were manually added. Aggregates incorrectly detected were discounted manually. The number of aggregates and singular RNAPs for each sample are shown in Table 6-1. The number of RNAPs in four 2  $\mu\text{m}^2$  images was counted for each sample.

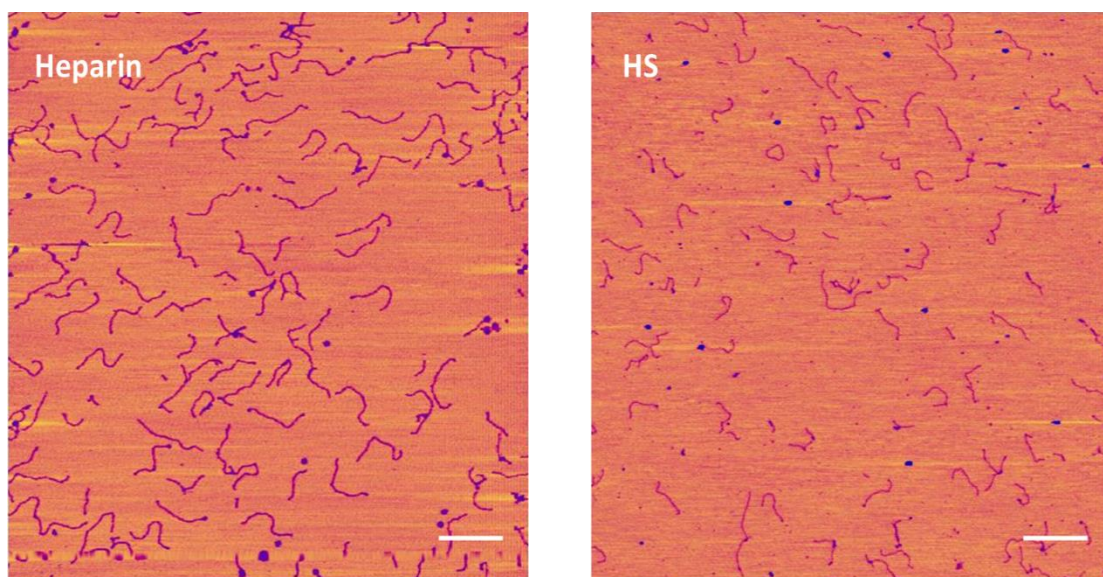
Sample	n (number of images)	Single RNAP molecules	Aggregated RNAPs
Control	5	159 $\pm$ 20	42 $\pm$ 4
Heparin	5	516 $\pm$ 26	26 $\pm$ 4
HS	5	557 $\pm$ 53	26 $\pm$ 4

**Table 6-1: Number of RNAPs bound to the surface as singular RNAPs or aggregates in the presence of heparin or HS.**

It can be seen that the presence of heparin or HS leads to fewer aggregates being found on the surface as well as a higher number of RNAPs visible overall. The binding of RNAP to the surface is not inhibited by either heparin or HS, and the data indicates that its binding may be increased by the presence of the polyanionic molecules.

As RNAP was seen to be able to bind the mica in the presence of heparin and HS the effects on DNA-RNAP complexes was investigated. In order to provide comparison to the previous experiments performed by Rivetti *et al.* and Crampton *et al.* without heparin or HS present, a similar two promoter template without a label was used to form OPCs in the presence of heparin or HS. OPCs were formed as detailed in Chapter 3 using 1144 bp DNA templates containing two tandem promoters. After incubation at 37°C for 15 minutes heparin or HS was added to a final concentration of 200  $\mu\text{g}/\text{ml}$ . The samples were then incubated at room temperature or 37°C. Complexes were imaged by AFM and analysed. Very few DNA-RNAP complexes were observed in samples incubated to 37°C (Figure 6-15).





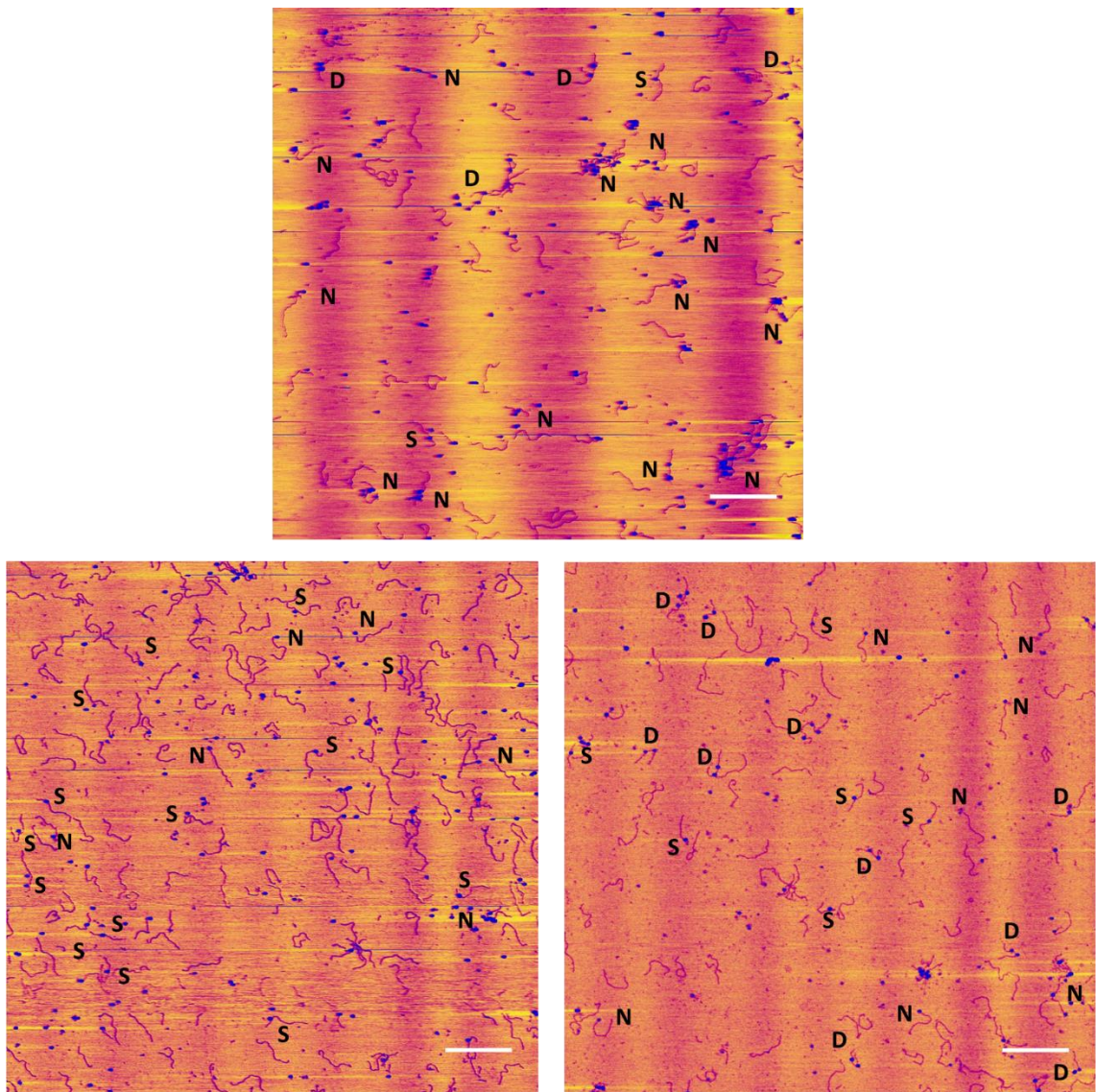
**Figure 6-15: AFM height images of OPC complexes heated to 37°C in the presence of heparin or HS. There are very few RNAP molecules visible on the surface and very few DNA molecules that have RNAP bound (Scale bars = 500 nm).**

With samples incubated at room temperature this was not the case: much higher amounts of OPCs were observed in these samples. The percentage of molecules with a single OPC formed, two OPCs formed and non-specifically bound RNAPs bound were determined by measuring the full contour length and counted (see Table 6-2).

Sample	Single OPC (nm)	Two OPCs (nm)	Non-specifically bound (nm)
-heparin / - HS (n=236)	30 % (356.4 ± 1.5)	17 % (331.4 ± 1.8)	53 % (382.9 ± 1.3)
+ heparin (n=200)	60 % (350.2 ± 1.4)	22 % (328.5 ± 1.8)	17 % (379.4 ± 6.5)
+ HS (n= 233)	33 % (358.4 ± 1.0)	51 % (328.3 ± 0.9)	14 % (379.5 ± 2.9)
Bare DNA (n=102)	380.16 ± 1.2		

**Table 6-2: Percentages of complexes formed without an inhibitor and with heparin or HS. The contour lengths of molecules are also shown with single OPCs showing a decrease in contour length of approximately 25 nm, double OPC a decrease of 51 nm and the non-specifically bound DNA showing no decrease.**

The addition of both heparin and HS reduces the level of non-specific binding observed. It was also noted that when heparin or HS was present, non-specific binding was mainly limited to the ends of the template, whereas without either inhibitor a number of aggregated complexes were seen. This can be seen in the AFM images shown in Figure 6-16.



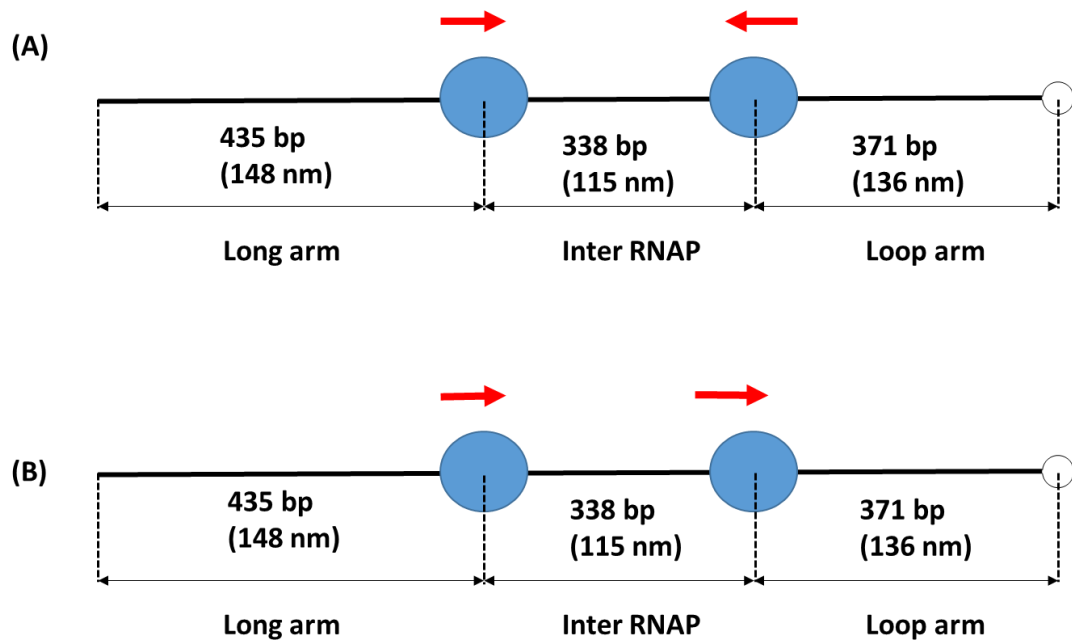
**Figure 6-16: AFM height images of OPCs formed on tandem promoter DNA without any inhibitor (top) and with heparin (bottom left) or HS (bottom right). The images have been annotated with an S to signify those with single OPCs, a D to signify those with two OPCs and an N to signify those that have non-specifically bound  $\sigma$ RNAP (Scale bars = 500 nm)**

It can be seen also that with the addition of heparin or HS that there is an increase in bare DNA in comparison to samples without either heparin or HS. This is expected as both RNAP binders reducing non-specific interaction of RNAP with DNA. As the HS sample contained the greatest amount of two OPC complexes it was decided that HS was preferable to heparin for the study of concurrent transcription events.

The protocol for HS challenged samples involved the formation of OPCs as previously described in Chapter 4 and then straight after removal from incubation at 37°C, HS from bovine kidney (Sigma, Saint Louis, MO) with an average molecular weight of 20 kDa was added to a final concentration of 200µg/ml and the sample was incubated at room temperature for 15 mins. Afterwards it was diluted by a factor of 10 in imaging buffer before being deposited, rinsed and dried as described previously, or had transcription initiated as detailed in Chapter 4.

### **6.2.2 Investigation into transcriptional collisions for *E.coli* RNA polymerase in the presence of heparan sulphate**

With the development of a method to incorporate heparin into *in vitro* transcription reactions that was compatible with AFM, this method was applied to samples with template DNA harboring two promoters. The two arrangements investigated were convergent and tandem promoters. These give rise to the possibility of transcriptional interference (TI) caused by the interaction between two RNAPs transcribing the same template. By taking samples at the different time points in *in vitro* transcription reactions, OPC and post elongation, and depositing them onto mica, it is possible to use AFM to track the outcomes of transcription events from these promoter arrangements. This is achieved by using the high spatial resolution and high signal-to-noise in the AFM to map the positions of RNAP molecules on the template. It is possible to use the AFM to provide snapshots of transcription events by measuring the change in position of RNAPs for OPCs to elongating complexes (ECs). Contour length measurements along the DNA backbone are taken from the images, to give RNAP positions on the DNA template and RNAP-RNAP (inter-RNAP) contour length separation as well as inter-rnap distances from cross sectional analysis of the 3-dimensional topographic data. The lengths of the different parts of the template can then be compared between OPCs and ECs. The templates and measurements possible are schematically highlighted in Figure 6-17.



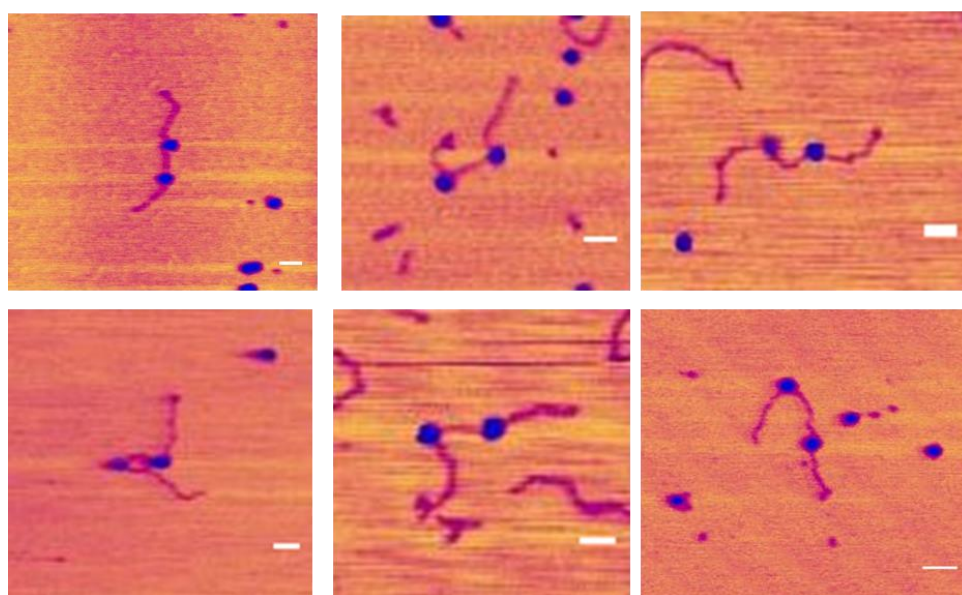
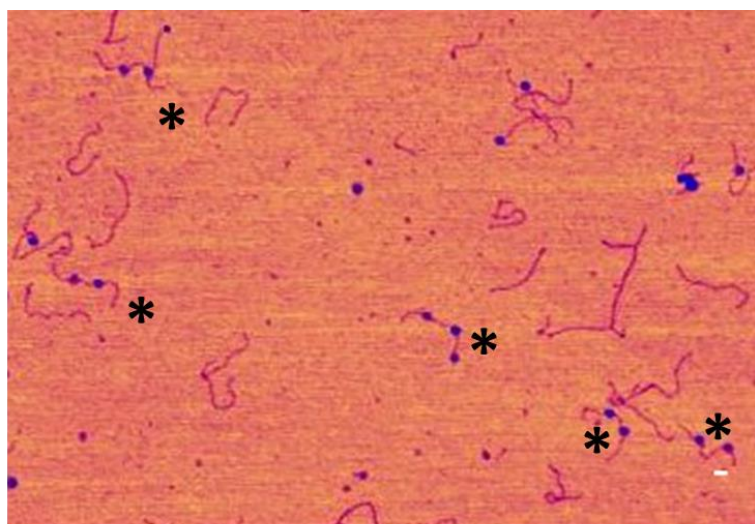
**Figure 6-17: Diagram of the DNA templates with RNAPs bound at their promoters. A) Convergent promoters are shown with the RNAPs bound to the promoter site. B) Tandem promoter arrangement. (DNA= black, RNAP= blue and promoter = red)**

Previous investigations into TI using AFM by Billingsley *et al.* and Crampton *et al.* using similar templates without the same labels or added heparin, as well as studies utilising other techniques such as footprinting experiments by Hobson *et al.*, have shown that *in vitro* RNAPs are not able to pass each other in either promoter arrangement and remain stable on the DNA template [4, 194, 201, 202, 318]. The use of the fiducial marker allows the starting point of each RNAP to be determined in ECs as RNAPs have not been observed to pass. This along with the removal of non-specifically bound RNAPs by incorporating heparin means that outcomes of TI when two RNAPs are acting on the same template can be investigated with more confidence and in more detail.

### 6.2.3 Labelled convergent promoter templates

Firstly OPCs were formed on a convergent promoter template using *E.coli* RNAP holoenzyme before being challenged with HS as previously described. These complexes were imaged using AFM (Figure 6-18) and those that had two RNAPs bound to the template were analysed.

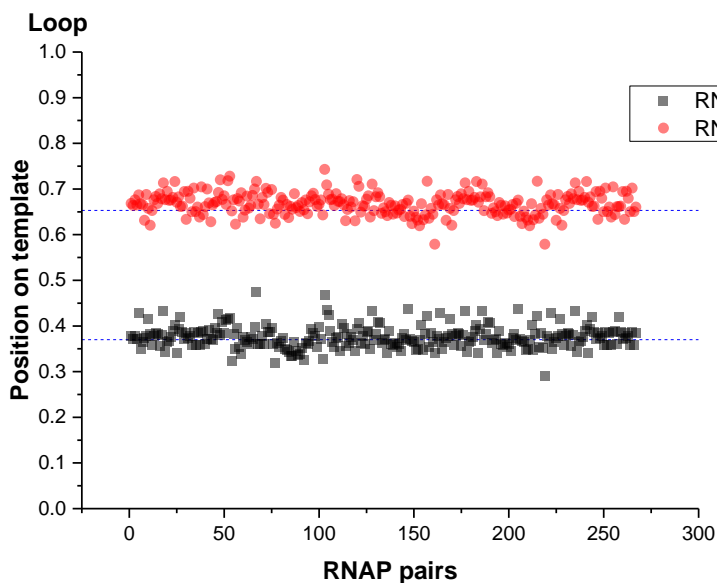




**Figure 6-18: Montage of images showing OPCs formed on the convergent promoter template. The top images show a wide scan with OPCs marked with a black star, the images below are high resolution views of OPCs. (Scale bars = 50 nm)**

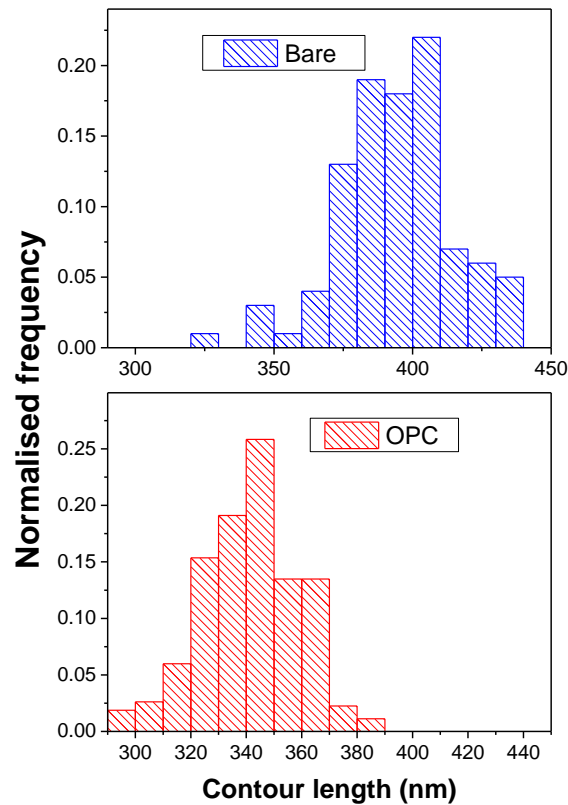
Measurements were made of the arm lengths and inter-RNAP separation in order to determine whether the RNAPs were located at the promoter sites. The contour length measurements were plotted as percentages of the total contour length as shown in Figure 6-19 on the y-axis. Along the x-axis is the count of each measurement (RNAP Pair). Each count position denotes two RNAP molecules one shown in red and the corresponding RNAP bund to the same template in black. When arranged like this it can be seen that for the vast majority of complexes analysed, the RNAPs are bound at the expected promoter sites which are located at approximately 0.32 for the short arm

(loop labelled arm) promoter and 0.37 for the long arm promoter from the template ends, indicated by the blue dashed lines. The average position of the RNAPs at the loop arm promoter was  $0.67 \pm 0.02$  (0.33 from the end of the template) and the average position of the RNAP located at the long arm promoter was  $0.37 \pm 0.02$ .



**Figure 6-19: Scatter plot of template position for each RNAP pair. The expected position of the promoter sites are denoted by the blue dashed lines. It can be seen that the majority of RNAPs have bound in the region of the promoter, indicated by the tight grouping of points at the promoter position.**

In order to further confirm that these complexes were in OPCs, the total contour length was compared with that of the bare DNA template. A shortening of  $52.9 \pm 3.2$  nm in the contour length was recorded, which is expected due to the wrapping of the DNA around the RNAP upon formation of two OPCs. This shortening falls in the range expected, with one RNAP being shown to reduce the contour length by  $27.5 \pm 4.0$  nm [126]. Figure 6-20 shows histograms comparing the change in contour length.



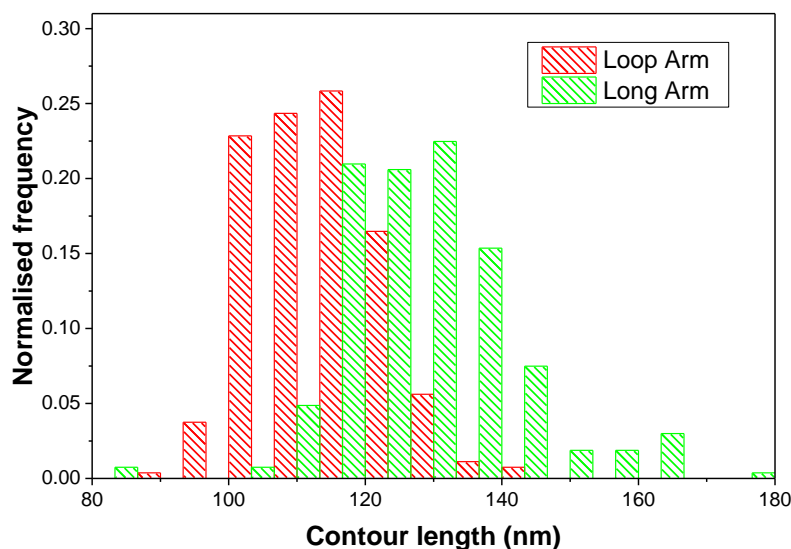
**Figure 6-20: Histogram plots of bare DNA and OPC total contour length. It can be seen that there is definite shift seen in the position of the peak value upon formation of OPCs (n=100 for bare and n=270 for OPCs)**

The wrapping of the DNA upon formation of OPCs also leads to a distinct bend angle in the DNA which can be easily distinguished by eye, allowing for selection of complexes that had formed two OPCs [126]. To access the effects of adding HS, complexes were designated specific classes: double OPC complexes; single OPC complexes; randomly bound RNAP or greater than two RNAPs bound. The percentages of each were as follows: 56% for double OPCs; 27% single OPCs; 17% with randomly bound or more than 2 bound RNAPs. The number of DNA molecules with no RNAP bound was not counted due to their high number. In comparison to studies by *Crampton et al.* and *Rivetti et al.* approximately 20% of DNA molecules seen were bare, the number observed here outnumbered the RNAP bound DNA [3, 126]. This increase in bare DNA when considered with the low levels of RNAP bound at sites outside of the promoters suggests that HS is having the desired effect of decreasing the number of non-



specifically bound complexes. It has been seen that for studies utilising a template with two  $\lambda_{pr}$  promoters arranged in tandem, that binding to one promoter is prevalent, *Rivetti et al.* noted 50% of complexes had only a single promoter site occupied [126].

The arm lengths for the complexes recorded as having two OPCs is shown in Figure 6-21.



**Figure 6-21: Histogram plot of OPC arm lengths. The distribution of lengths for both arms highlights the difference between the initial lengths of the two arms. (n=270)**

The arms are different lengths and the shorter arm is marked by the single stranded loop, meaning that determination of RNAP positioning is easily discerned.

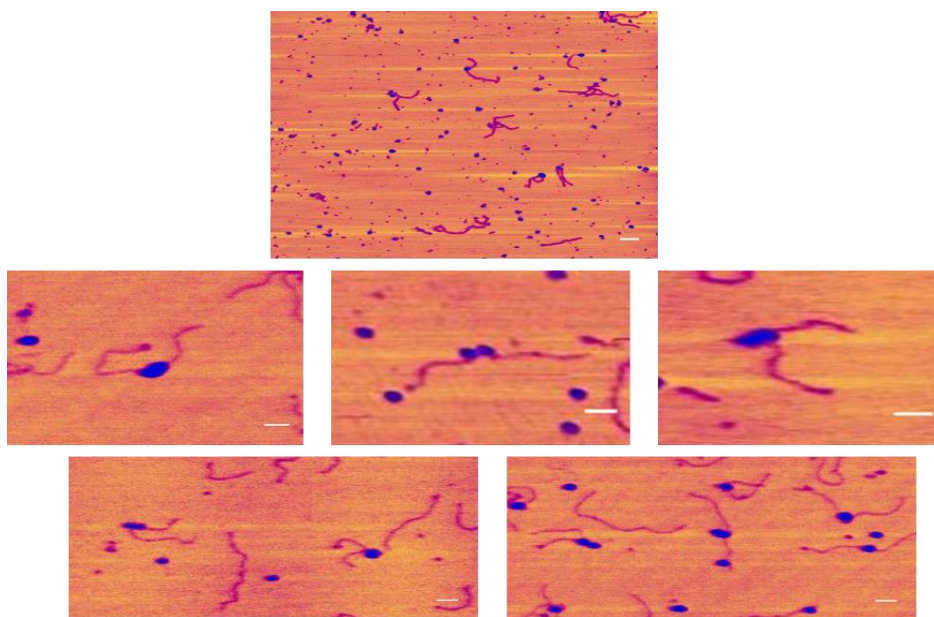
Once it was confirmed that OPCs were formed on the DNA, elongation was initiated as described in Chapter 3, by the addition of all four NTPs. Samples were then imaged by AFM and templates with two RNAPs still attached were analysed. It was observed that the full contour length of the DNA increased from that measured in OPCs by  $32.3 \pm 3.1$  nm. In comparison to the length of bare DNA, this is a decrease in contour length of  $20.6 \pm 3.0$  nm. This increase is believed to occur due to reduced wrapping of the DNA around the RNAP once elongation has been initiated and RNAP has lost contact with the promoter.

The arm lengths measurements showed an increase in the length whereas the average inter-RNAP separation measurements showed a decrease. This is as expected as both RNAPs are moving towards the centre of the template. The contour lengths of each measurement for OPCs and elongated complexes are given in Table 6-3.

Sample	n	Long arm (nm)	Inter-RNAP contour length (nm)	Loop arm (nm)	Total contour length (nm)
OPC	270	127.9 ± 0.8	100.1 ± 0.6	113.4 ± 0.6	341.5 ± 1.0
Elongated	284	178.3 ± 3.2	35.9 ± 1.8	153.6 ± 3.1	369.8 ± 2.1

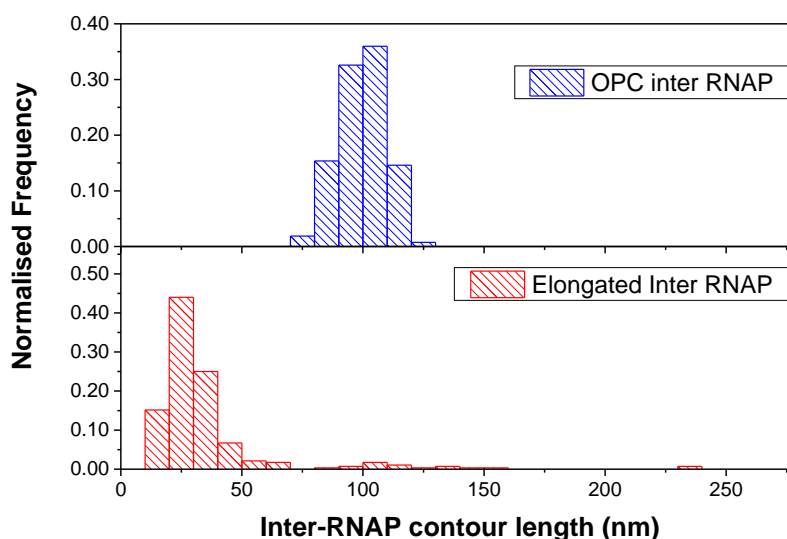
**Table 6-3: Summary of measurements for convergent promoter complexes.**

The majority of complexes analysed (96%) had two RNAPs in close proximity to each other (Figure 6-22). These complexes were identified as collided complexes (CC).



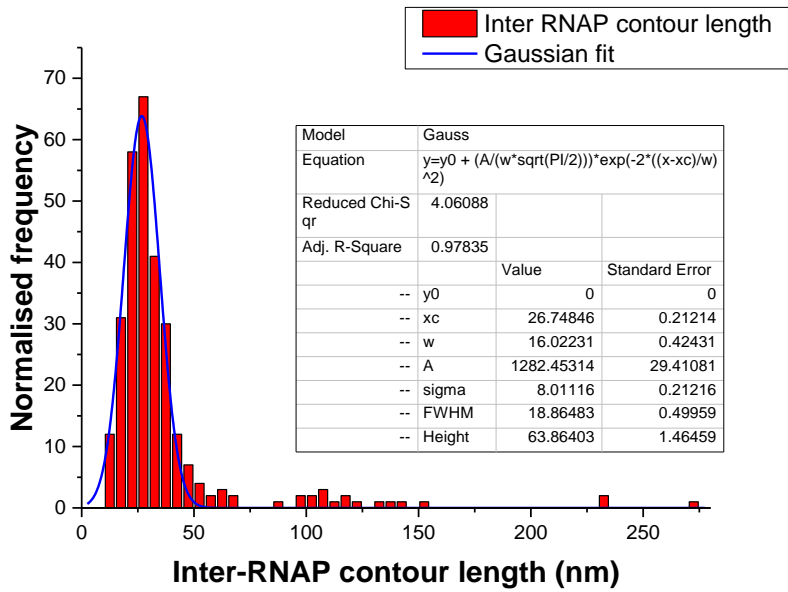
**Figure 6-22: Montage of images showing collided complexes. The images show that the RNAPs are in close proximity to each other while remaining bound to the template, indicating these are the outcomes of elongation. (Scale bars for top= 100 nm, centre and bottom images = 50 nm).**

In order to investigate the nature of collision events further, the inter-RNAP contour length between complexes that had two RNAPs located on the template was measured. Figure 6-23 shows a plot of the inter-RNAP contour length from OPCs and elongated complexes.



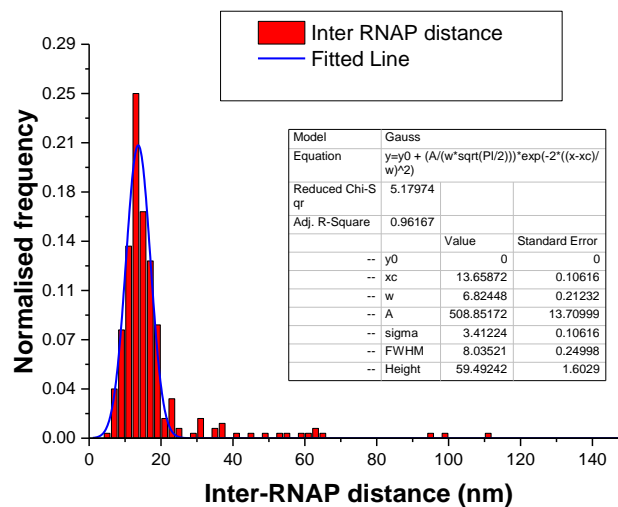
**Figure 6-23: Histograms comparing inter-RNAP contour length for OPC and elongated complexes. It can be seen that there is decrease in the average value after the addition of NTPs, indicating that RNAPs have undergone elongation and collided (n for OPCs = 270 and n for elongated complexes = 284).**

There is an obvious decrease in the separation between the two RNAPs after addition of NTPs and elongation has occurred, indicating that the majority of RNAPs have left their promoters and travelled in the expected directions, stalling on the template with an average separation of  $35.9 \pm 1.8$  nm. A Gaussian fit of the histogram gave a value of  $26.8 \pm 0.2$  nm (Figure 6-24).



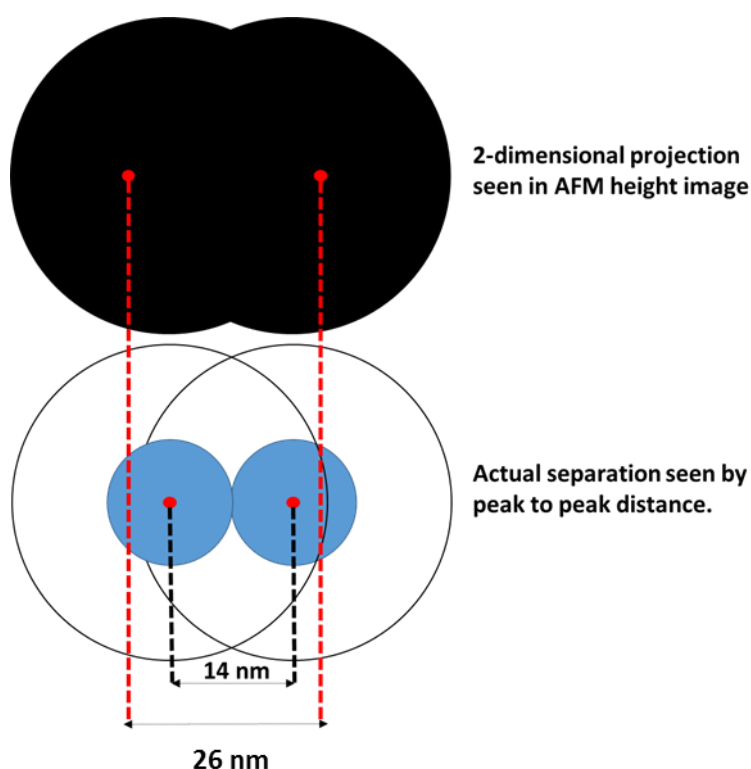
**Figure 6-24: Gaussian fit of inter-RNAP contour length of elongated complexes (n=284).**

The outlying points in the histogram of inter-RNAP contour length are most likely due to complexes where both RNAPs have failed to escape the promoter or due to low levels of RNAPs overlaying the DNA template therefore appearing as if bound. In order to provide greater detail into the separation of RNAPs that were in close proximity the straight line distance between the highest point of each RNAP was measured (intermolecular distance) and plotted as a histogram (Figure 6-25)



**Figure 6-25: Histogram of inter-RNAP distance. A Gaussian fitted curve is shown in blue with values of the fitting shown in the inset table (n=284).**

As the intermolecular distance between the RNAPs does not take into account the path of any DNA between the two RNAPs those that had DNA visible are expected to display a shorter distance than the inter-RNAP contour length. In the case of RNAPs that are in close proximity and appear convoluted it provides a more accurate measurement of the distance between the two RNAPs. When in close proximity, tip convolution can lead to uncertainty when measuring the centre of both RNAPs as is shown by the diagram in Figure 6-26.

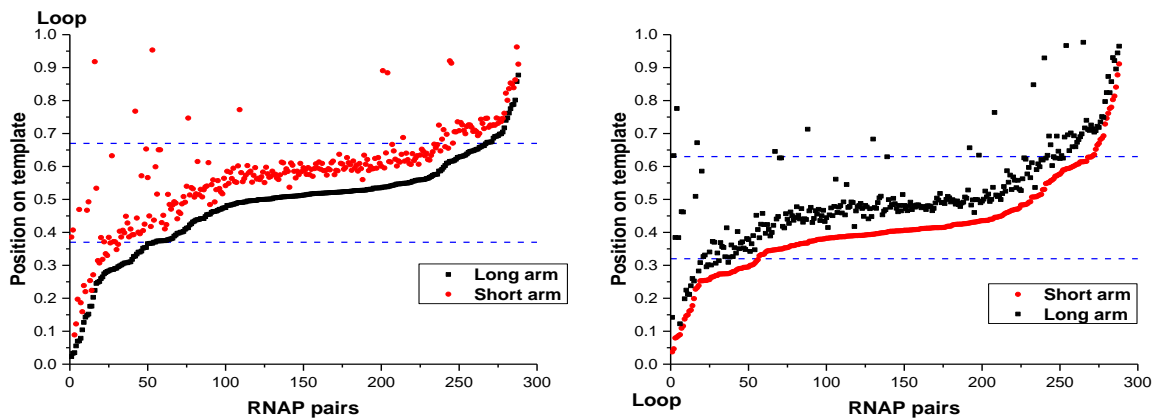


**Figure 6-26: Diagram depicting the effects of tip convolution of inter-RNAP separation. The appearance of two RNAPs located in close proximity in two – dimensions from an AFM height image is shown in black. The centre of the feature is shown not to be the true centre of both RNAPs as is shown by the schematic representation below (image drawn to scale assuming a RNAP diameter of 32 nm in the AFM height image and an actual diameter of 14 nm for an RNAP molecule).**

The average inter-RNAP distance was  $16.6 \pm 0.8$  nm which is lower than measurements taken of the contour length from the 2-dimensional images. Fitting the plot to Gaussian

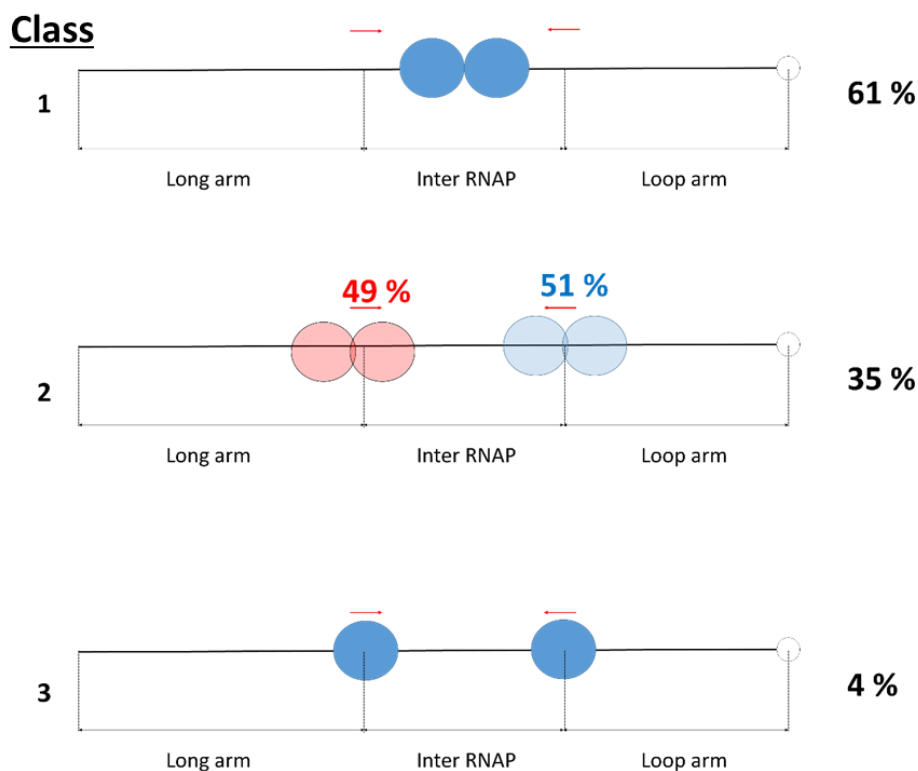
distribution gave a value of  $13.7 \pm 0.1$  nm. This distance is more in keeping with that expected for two RNAP molecules in hard contact.

By plotting the position of each RNAP as a percentage of the total template length the distribution of complexes on the template can be seen (Figure 6-27).



**Figure 6-27: Scatter plots of positions of RNAP pairs as a percentage of the total contour length of the template. On the left the RNAPs originating from the long arm are plotted in ascending order towards the short (loop) arm. On the right the RNAPs from the short arm are plotted in ascending order away from the short (loop) arm.**

By looking at the distribution of RNAPs the molecules can be further broken down into classes: 1) those pairs that have both RNAPs located within the inter-promoter region with both RNAPs in close contact; 2) those that have a small inter-RNAP separation and have one or both RNAPs located outside the inter-promoter region; 3) those with both RNAPs separated by a distance equal to or greater than that observed in OPCs. Schematic representations of these classes are shown in Figure 6-28 with the respective percentage of each class.

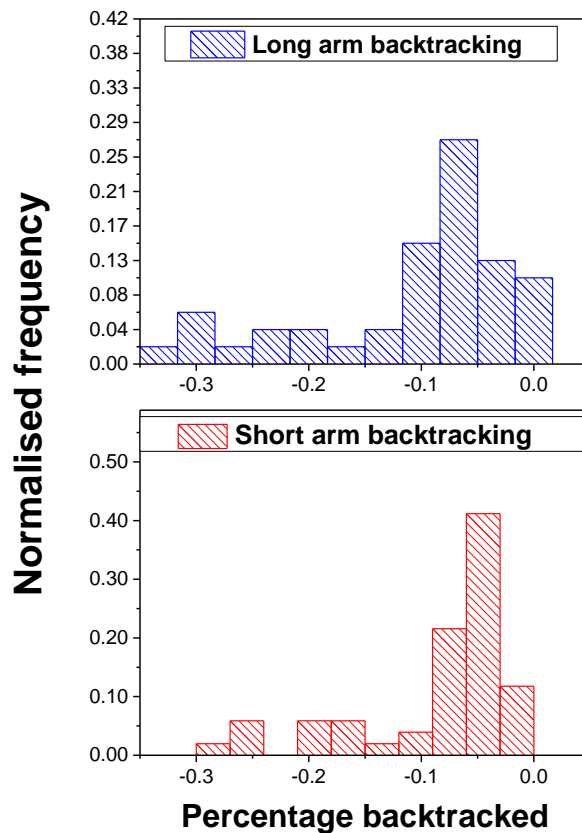


**Figure 6-28: Schematic representation of different classes of complexes seen for transcription from convergent promoters.**

Class 1 molecules had a contour length separation of  $30.6 \pm 0.9$  nm. By analysing arm lengths it was found that the RNAP originating from the short (loop) arm had travelled an average of  $30.9 \pm 1.3$  nm or 100 bp and the RNAP originating from the long arm had travelled an average of  $49.9 \pm 1.4$  nm or 151 bp. The position of these complexes in the centre of the template indicate that these are the outcome of two actively transcribing RNAPs colliding with each other and stalling.

Class 2 complexes had a contour length separation of  $29.3 \pm 1.6$  nm. The location of the RNAP pairs were relatively evenly distributed between the short and long arms with 49 % being located at the long arm promoter and 51 % at the short arm promoter. Arm length analyses showed that RNAPs located upstream of the long arm promoter had moved an average of  $37.9 \pm 5.0$  nm. For RNAPs initiating from the short arm promoter the RNAP from that promoter had moved  $25.8 \pm 3.4$  nm upstream. These distances are equivalent to  $80 \pm 10$  bp for the short arm RNAP and  $110 \pm 14$  bp for the long arm RNAP. If the percentage of the template travelled upstream from the promoters is plotted as

a histogram (Figure 6-29) it can be seen that the majority of RNAPs only travel a value between 0.050 to 0.064 upstream of the promoter.



**Figure 6-29: Histogram plot of the movement upstream of the respective promoter travelled as a percentage of the total contour length for both long arm (blue) and short (red) arm promoters.**

This would equate to a backwards movement of 17-22 nm or 50 to 60 bp. The location of these complexes around the promoter region as well as upstream of the promoter leads to the interpretation that they are the outcomes of a collision between one RNAP that has elongated through the inter-promoter region and collided with the other RNAP while it is located at its promoter. It is noted though that the distances measured are at the limit of resolution for the AFM and some contour length may not be equated for due to tip convolution effects.

Class 3 complexes displayed a separation that was equal to or greater than that seen in OPC samples. The rare occurrence of such complexes (4 %) is most likely because

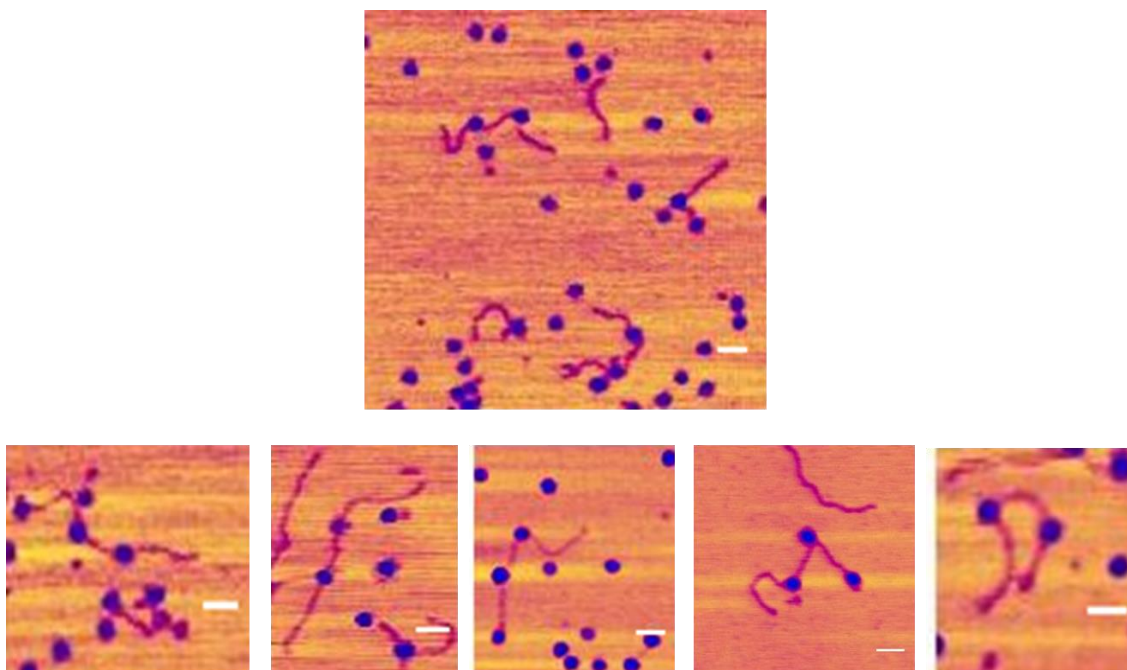


they are artefacts of either non-specific binding events or are complexes in which both RNAPs have failed to escape the promoter and initiate elongation.

#### 6.2.4 Transcription of tandem promoter template

The above experiments were repeated for a DNA template containing two tandemly arranged promoters with the direction of elongation being towards the loop.

For the tandem promoter template, OPCs were set up as described in 5.2.3. These complexes appeared similar to those seen for the convergent promoter template, examples of which are shown in Figure 6-30.

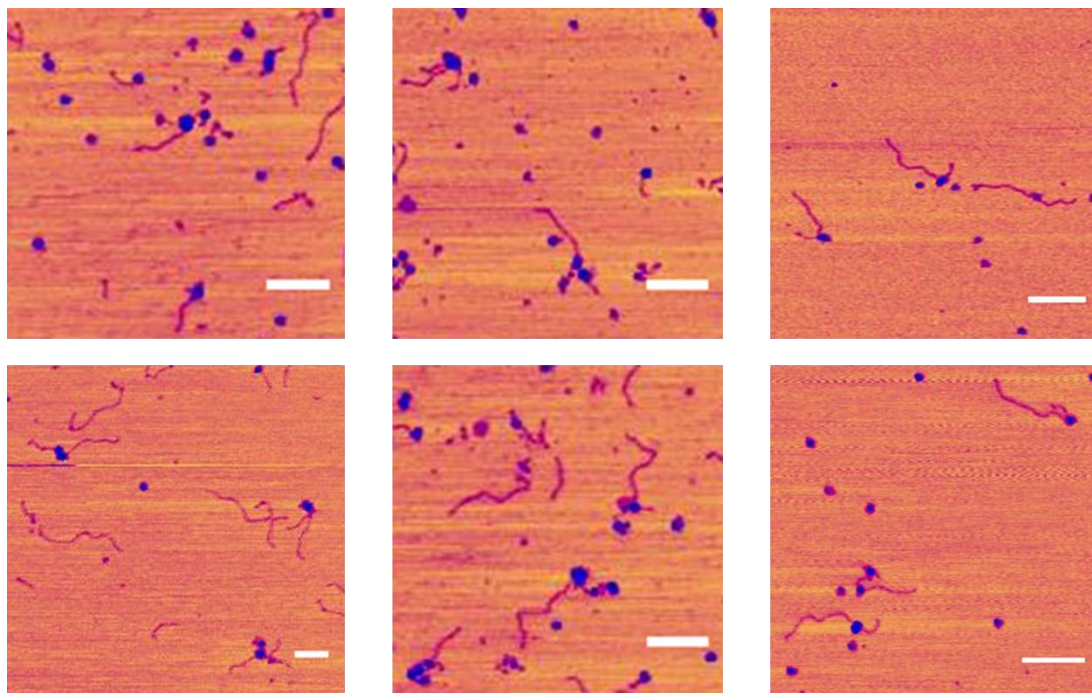


**Figure 6-30: Montage of images showing complexes with two OPCs formed on the tandem promoter template. (Scale bars = 50 nm)**

Populations of complexes in specific arrangements were counted and are as follows: 57% had double OPCs; 31% had a single RNAP OPC; 12% displayed random or more than two RNAPS bound to the template. These values agree well with what was seen for the convergent template. Again the amount of free DNA seen in the images was in excess and not counted.

A decrease of  $57.9 \pm 3.9$  nm in the DNA contour length was seen for those complexes that had two RNAPs bound which is expected due to the wrapping of the DNA upon formation of an OPC [126].

Once it was observed that OPCs could be formed on the tandem template, all four NTPs were added to the reaction mix in order to initiate elongation. Post elongation samples were imaged to determine the outcomes of tandem transcription in the presence of HS. It was found that for the majority of complexes there was a decrease in the average length of the arm downstream of both promoters and an increase in arm upstream of both promoters as compared to templates with OPCs (Figure 6-31).



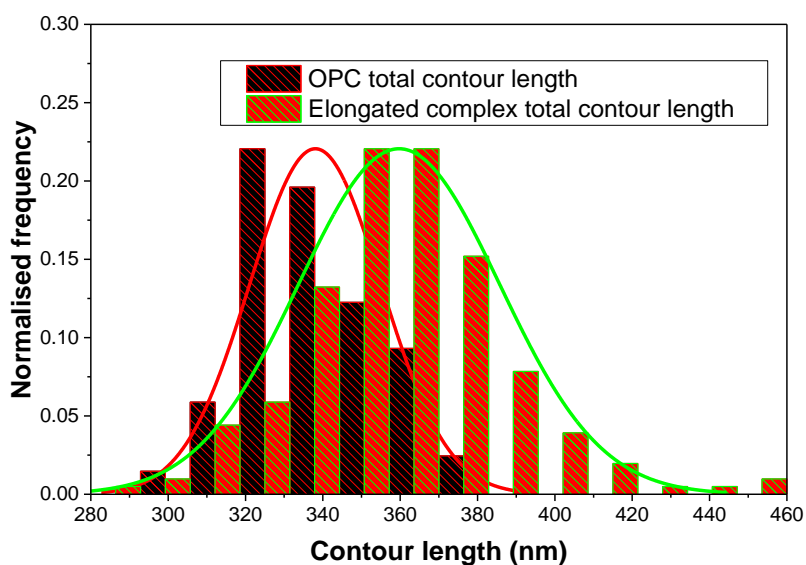
**Figure 6-31: Montage of AFM height images of RNAPs after elongation from two tandem promoters (Scale bars = 200 nm)**

There was also a decrease in the average inter-RNAP contour length. This is consistent with the majority of the RNAPs escaping from their promoters and elongating towards the loop end, but the RNAPs come closer together than the promoter spacing. These measurements are given in Table 6-4.

Sample	n	Long arm (nm)	Inter RNAP (nm)	Loop arm (nm)	Total contour (nm)
OPC	149	130.0 ± 0.7	93.7 ± 0.9	114.4 ± 0.8	338.1 ± 1.4
Elongated	206	237.3 ± 4.3	66.0 ± 3.5	57.8 ± 2.7	361.1 ± 1.75

**Table 6-4: Summary of measurements for tandem promoter template.**

As can also be seen from Table 6-4 that there was an increase observed in the overall contour length of complexes measured when compared to templates with OPCs formed, of  $23.1 \pm 3.2$  nm, which is expected due to the reduction in wrapping seen upon escape from the promoter (Figure 6-32). This equates to a decrease in contour length from the bare template of  $34.8 \pm 3.5$  nm.

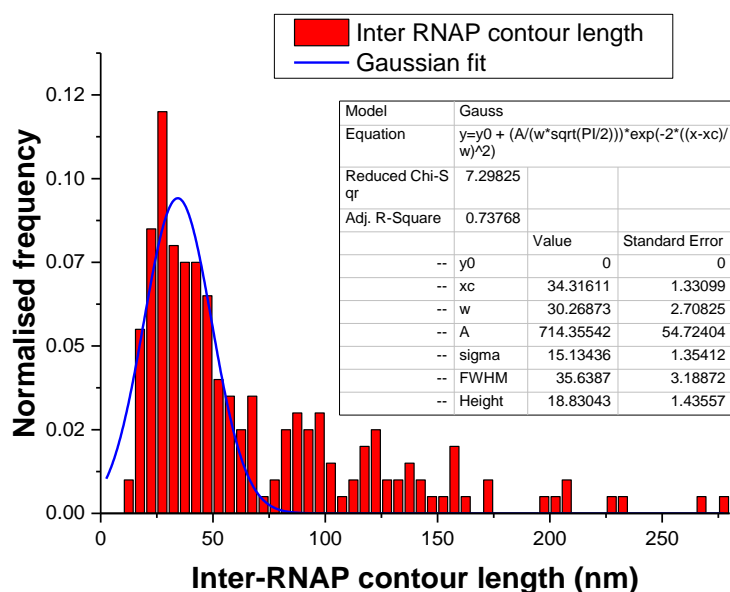


**Figure 6-32: Histogram plot of total contour length of OPCs (black and red, n=149) and elongated complexes (orange and green, n=204). The plots have been overlaid with a normal distribution to allow better visualisation of their differences.**

The decrease in the inter-RNAP contour length seen along with the average decrease in the loop arm indicate that the majority of RNAPs have escaped from their promoters and transcribed towards the end of the template. The average length of the loop arm indicates that the majority of complexes have not reached the loop, but have stalled

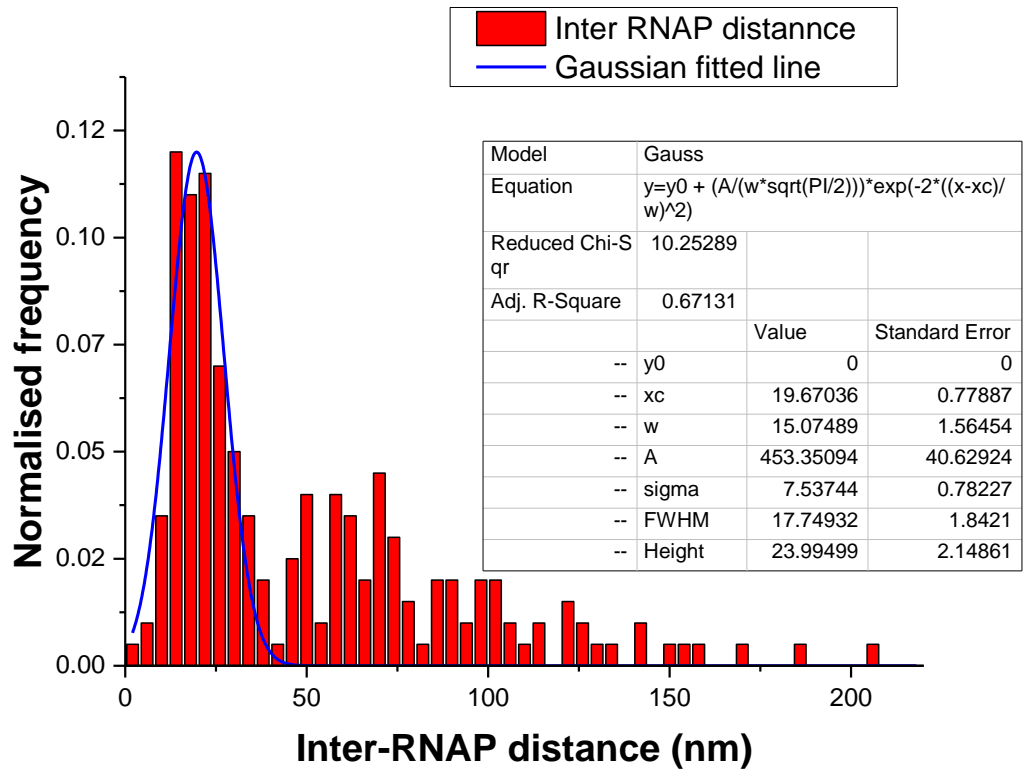
outside of the inter-promoter region. If both RNAPs were found to be at the loop, the contour length would be expected to be equivalent to that seen for the single promoter template, where no compaction of DNA after elongation was seen.

The average contour length of inter-RNAP contour length was  $65.9 \pm 3.5$  nm and inter-RNAP distance gave a separation of  $49.3 \pm 2.7$  nm.



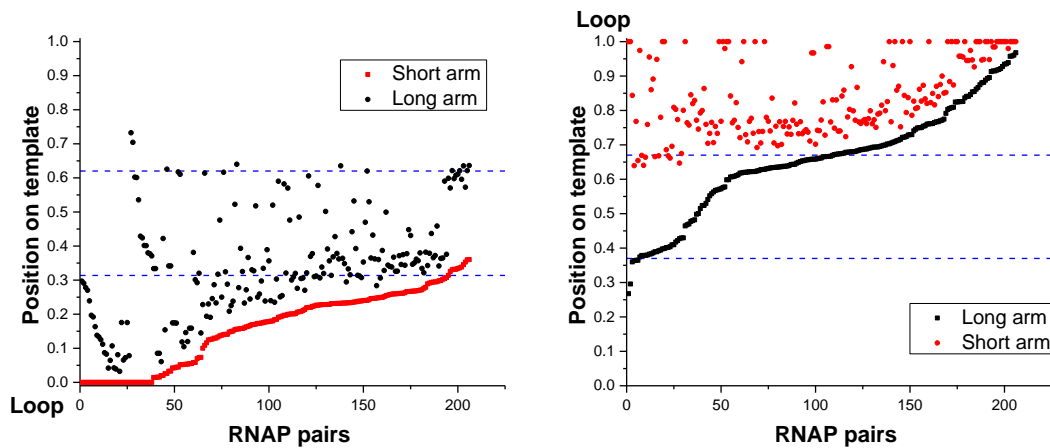
**Figure 6-33: Histogram plot of inter-RNAP contour length with a Gaussian distribution fitted to the main peak (n=204).**

The histogram plot of the inter-RNAP contour length (Figure 6-33) was fitted with a single Gaussian peak which had a value of  $34.3 \pm 1.3$  nm. This peak is most likely represents those RNAPs which have undergone a collision due their close proximity. A second peak can be seen at approximately 90-100 nm which is most likely those where both RNAPs have failed to initiate elongation and a third broader peak is seen at around 120 nm. This third peak may be due to the RNAPs originating from the short arm promoter escaping the promoter and entering elongation while the RNAPs at the long arm promoter fail to initiate. In the case of the inter-RNAP distance a Gaussian fit of the data gave a value of  $19.7 \pm 0.8$  nm (Figure 6-34).



**Figure 6-34: Histogram plot of inter-RNAP distance with the main peak fitted with a Gaussian curve (n=204).**

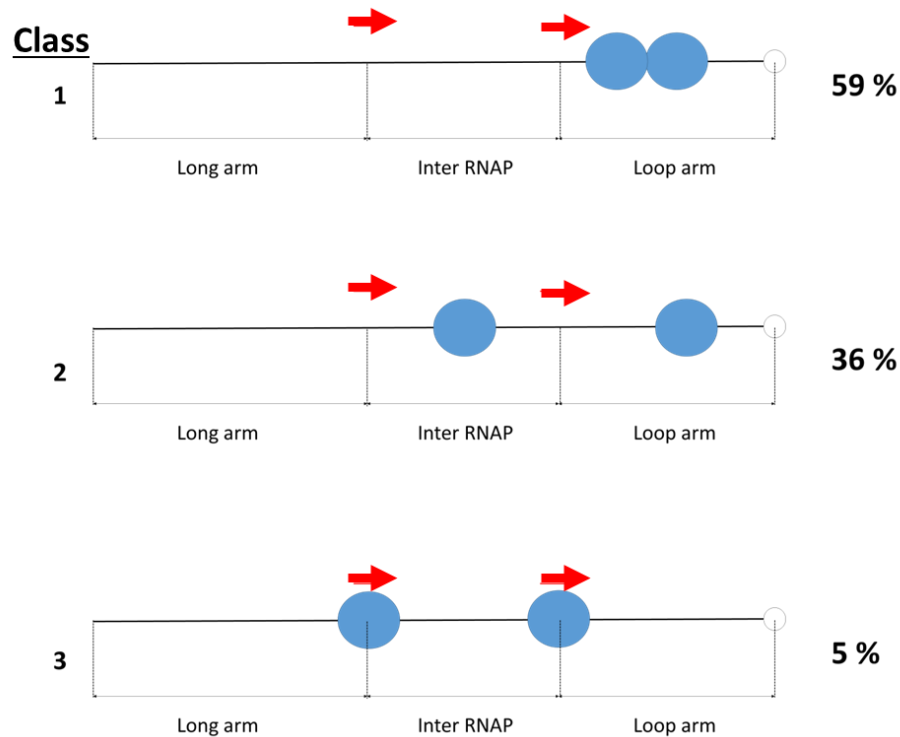
If the positions of the RNAPs is plotted as a percentage of the template it is possible to separate out a number of different classes from the data as can be seen from Figure 6-35.



**Figure 6-35: Scatter plot of the positions of RNAP pairs from promoters aligned tandemly as a percentage of the total contour length. On the left the RNAPs originating from the short arm are plotted in ascending order. On the right the RNAPs originating from the long arm promoter are plotted in ascending order.**

On average the leading RNAP travelled  $175 \pm 9$  bp from its promoter while the trailing RNAP travelled  $314 \pm 12$  bp. By viewing the scatter plot it can be seen that only a smaller number of RNAPs reach the end of the template and are trapped by the loop. The majority of RNAPs are located around the short arm promoter with both RNAPs just outside the inter-promoter region.

It is possible to discern three broad classes of complexes from the scatter plot but due to the RNAPs travelling in the same direction class designation is not as clear as in the case of convergent promoters: 1) those where both the leading (short arm promoter) RNAP and trailing (long arm promoter) RNAP are located downstream of the inter-promoter region; 2) those where the trailing RNAP is located between the promoters and the leading RNAP is downstream of the two promoters; 3) those that have both RNAPs located at or between the promoters. An example of each class is shown in Figure 6-36.



**Figure 6-36: Schematic representations of the complexes seen after elongation from tandemly arranged promoters.**

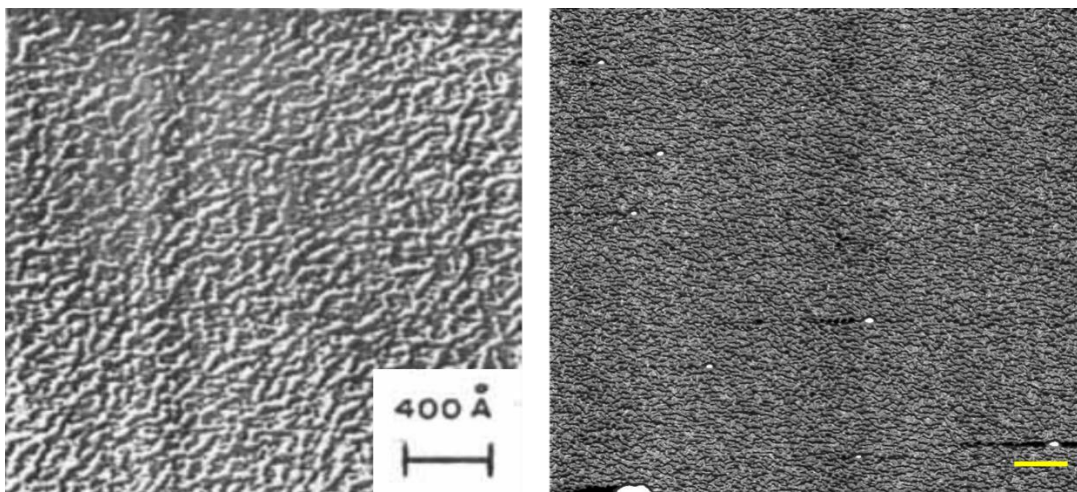
Class 1 complexes appear where a collision has occurred between the RNAPs as the separation seen is similar to that in the case of collisions between convergently transcribing RNAPs. It is not possible to determine whether these are the outcomes of an EC-EC collision or EC-SD collision as both RNAPs are moving in the same direction. If the leading RNAP pauses or arrests then a collision would occur. This is also true for those leading RNAPs which reach the loop and stall. Class 2 complexes with a smaller RNAP separation are most likely those that have undergone an EC-SD collision due to the failure of the leading RNAP to escape its promoter. Those that have a greater separation could be the consequence of only the leading RNAP escaping its promoter leading to a greater separation. It is also possible that the escape from the leading promoter happens first and the trailing RNAP then pauses or arrests preventing a collision from occurring. The final group of complexes are most likely those where both RNAPs have failed to escape their promoters, which is indicated by both RNAPs being located near to their respective promoters.



## 6.3 Discussion

### 6.3.1 Incorporation of polyanionic inhibitors of RNAP non-specific binding

Heparin alone at higher concentration was observed to form a layer over the mica surface that has areas of aggregation which are loosely attached. This ability of heparin to form a layer on mica was noted by Stoner *et al.* who showed it was possible to passivate a mica surface for investigations of implants using electron microscopy to image the surface [402]. The transmission electron micrographs show a rough layer covering the mica similar to what is observed by AFM (Figure 6-37) [402].

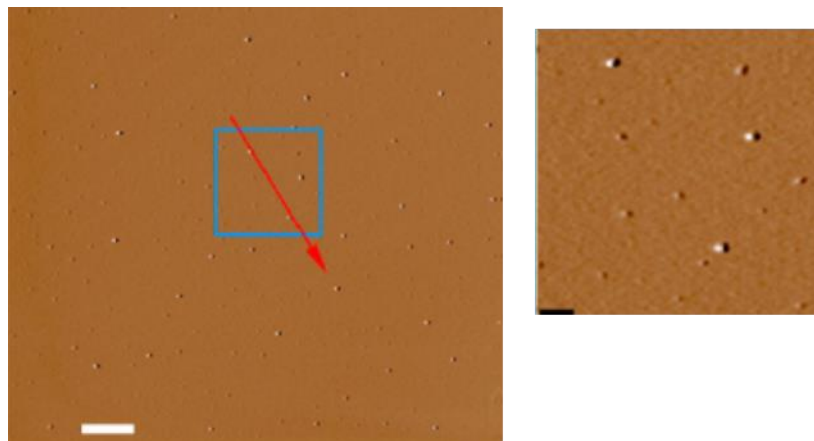


**Figure 6-37: Comparison of electron microscopy and AFM height image of 200 µg/ml heparin absorbed onto a mica surface. It can be seen from the images that heparin forms a rough homogeneous layer on the mica (Scale bar for AFM image is 100 nm)**

HS was found on the mica surface in a more dispersed state rather than forming a layer. This may be due to fact that HS has domains of negative charge whereas heparin is negatively charged along the whole chain of the molecule [398]. Upon cleavage of the mica, potassium ions dissociate from the surface leaving a net negative charge at the surface. The heparin may be binding to the surface through counter ion correlation as proposed by Pastre *et al.* in a manner similar to DNA due to the presence of  $Mg^{2+}$  in the buffer [270, 402, 403]. In the case of HS binding would be limited to the domains of higher sulphonation. This could mean that chains either side of the HS fold giving the appearance of small globular features on the surface. It also may be that the chains of



HS can associate with each other through interactions mediated by the divalent  $Mg^{2+}$  forming a mesh like layer upon the surface. At lower concentrations both heparin and HS appeared dispersed on the surface and had an appearance similar to that seen by Zhang *et al.* when investigating adeno virus interactions (Figure 6- 38).



**Figure 6-38: Images of heparin deposited on mica collected by Zhang *et al.* (Scale bars = 1 $\mu$ m white and 250 nm black) [404].**

In the presence of DNA, both heparin and HS lead to areas of differing morphology. At higher concentrations both heparin and HS lead to areas of aggregates, presumably formed by DNA with heparin or HS. Areas were also seen that appeared to have DNA bound on top of the heparin or HS. This is supported by the instability of these regions, which were easily dislodged by the AFM tip. The final areas seen did not appear to have any heparin or HS, but only high density of DNA molecules. These areas may form due to separation of the DNA and heparin or HS on the surface. It is believed that the high density areas of DNA observed, were caused by the highly negative heparin excluding DNA from where heparin was bound on the mica. This therefore caused the DNA to be repelled to where heparin was not bound causing it to be contained in regions of high density.

RNAP in the presence of heparin or HS appeared to readily bind to the mica. This may be due to the binding strength of the protein to the mica preventing heparin or HS from forming a layer on mica, with appearance of the samples being similar to other studies which co-adsorbed proteins and heparin to mica [404-407]. The increase in RNAPs seen on the surface may be due to the reduction in the number of RNAPs found in

aggregates. Experiments by Houska *et al.* on the layering of protein/polyelectrolyte on a surface showed that heparin could interact with the outer surfaces of a protein with a net negative charge due to local charge clusters [408]. This would lead to the repulsion of RNAPs with heparin or HS bound, preventing aggregation. If this was the case then the RNAPs would repel each other more than when heparin or HS is not present. This effect of heparin and HS as an anti-aggregation agent for proteins has been noted for some proteins [409], but due to the diverse structure and sequence a complete understanding of this effect has not been elucidated [410]. It is also possible that the binding of heparin or HS to the surface of the RNAP may allow for more  $Mg^{2+}$  mediated binding to occur leading to more RNAPs remaining tightly bound to the surface after rinsing and drying.

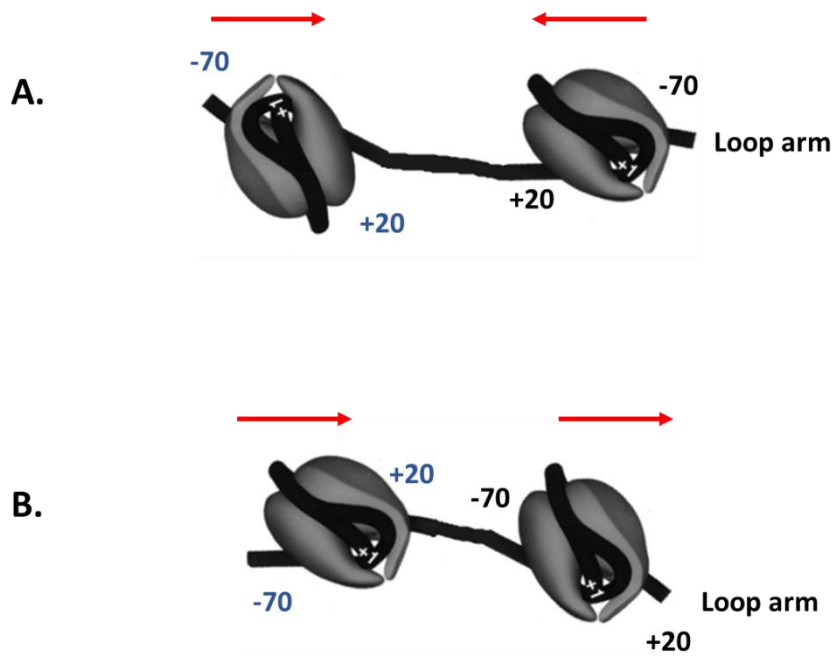
### **6.3.2 Effects of HS on OPC formation and imaging.**

For deposition of OPCs incubated with heparin or HS there was no noticeable effect on adsorption of complexes to mica, as was also reported by Rivetti *et al.* when investigating wrapping of DNA by RNAP, where it was found that RNAP pre-incubated with heparin did not prevent binding of DNA-RNAP complexes to the surface [126]. An increase of bare DNA was noted along with a decrease in RNAPs bound outside of the promoter regions, indicating that both heparin and HS are preventing non-specific binding. In the case of the two promoter template it was found that heparin and HS had differing effects on the number of each type of complexes seen. Both led to a decrease in the amount of non-specifically bound  $\sigma^{70}$ RNAP as well as an increase in bare DNA present in the samples. Studies on two promoter templates by Rivetti *et al.* gave 20% bare DNA molecules on the surface, but here with heparin and HS bare DNA is the dominant species seen on the surface, indicating that less non-specific binding is occurring [126]. Studies by Crampton *et al.* reported that 50 % of complexes they analysed for a two promoter template displayed non-specifically bound  $\sigma^{70}$ RNAP, with heparin or HS this number decreased to 17 % and 14 %, respectively, indicating that they both reduce non-specific interactions of  $\sigma^{70}$ RNAP with the DNA. With the introduction of heparin though there was a greater number of single OPC complexes in comparison to when HS was present. Studies by Rivetti *et al.* and Crampton *et al.* on two promoter templates reported single OPCs to account for 50 % and 30 % of

complexes formed, respectively [3, 126]. This increased amount is unexpected as neither heparin nor HS are expected to remove OPCs. It is however reported in the literature that  $\sigma^{70}$ RNAP complexes may be susceptible of OPCs to removal by heparin as was shown for the T7 phage promoter A1 by Pfeiffer *et al.* They also reported the susceptibility of OPCs to attack by heparin was dependent on the structure of the promoter and the promoter-RNAP complex [401]. Schlax *et al.* were able to provide evidence that for the lacUV5 promoter the steps between free  $\sigma^{70}$ RNAP binding and the formation of an initially transcribing complex (ITC) were reversible and susceptible to removal by heparin [411]. It is possible that this is the case for the  $\lambda_{pr}$  promoter, as OPCs formed in the absence of NTPs may dissociate as they are not capable of forming ITCs. It is also possible that with the formation of two OPCs, one of the OPCs may form a less stable complex due to the influence of the wrapping from the second OPC. This could also account for the lack of DNA bound by heparin when the sample is incubated at 37°C as the increased temperature would make the OPCs less stable, allowing the removal of OPCs from the template. The fact that this increase is not seen with HS may be due to the lower levels of sulphonation of HS. This would give HS a lower net negative charge and so it is possible that the interaction of HS with  $\sigma^{70}$ RNAP is not as strong as for heparin. OPCs formed on the 1144 bp two promoter template still show low levels of non-specifically bound RNAP. Rivetti *et al.* noted that when pre-incubated with heparin there were still some interactions seen between the  $\sigma^{70}$ RNAP and DNA when  $\sigma^{70}$ RNAP was in excess [126]. This fact indicates that some low levels of non-specific binding may be unavoidable. This binding was mainly limited to the ends of the DNA template as was observed by Rivetti *et al.* and therefore would have minimal effect on sample analysis. The relative amount of complexes seen for the labelled two promoter templates are similar to that of the unlabelled template, indicating that the addition of the loop does not alter the effects of HS. This also confirms that the binding to the loop structure seen for single promoter template is most likely due to non-specific interactions.

The change of contour length measured for the labelled two promoter templates reflects values reported from foot-printing studies and AFM studies where wrapping of the DNA was shown to involve approximately 90 bp [116, 121, 324]. The difference in contour length seen for the inter-RNAP contour length between the convergent and

tandem promoters is reported by Billingsley *et al.* and can be expected due to the promoter arrangements for each template which is highlighted in Figure 6-39 showing the approximate amounts of wrapping expected up and downstream of the promoters [125].



**Figure 6-39: Schematic representation of degree of wrapping for convergent and tandem promoter arrangements. A) It can be seen that for convergent promoters there is 20bp downstream of each promoter involved in wrapping the DNA, whereas for the tandem arrangements (B) there is 70bp upstream of the leading RNAP involved in wrapping and 20bp downstream of the trailing RNAP, giving a difference from convergent promoters of 50bp.**

By looking at Figure 6-39, it can be seen that a larger number of base pairs (bp) from the inter-RNAP space will be involved in wrapping around the RNAP for the tandem template than the convergent, which is reflected in the measurements. The extent of this difference is lower than would be expected but may be due to the orientation of bends in the DNA not being truly represented when the complexes are deposited onto a mica surface and dried for imaging.

### 6.3.3 Outcomes of convergent transcription in the presence of HS

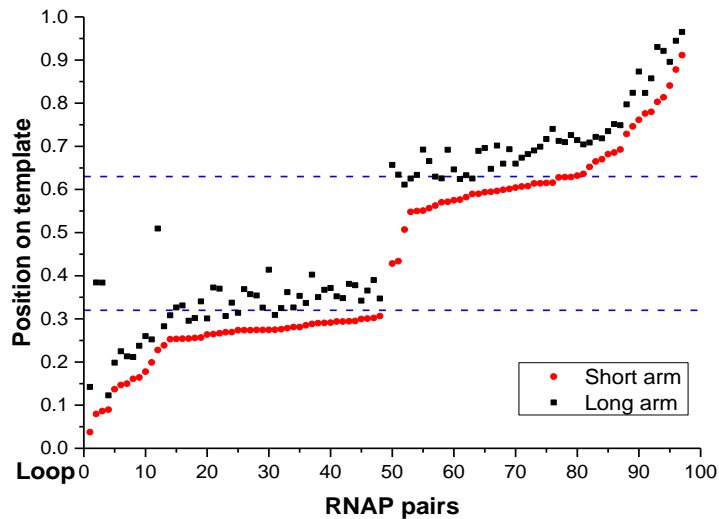
After addition of NTPs the majority of RNAP pairs are seen in close proximity to each other. As this decrease in distance was only seen once NTPs had been added and not in OPCs samples it is reasoned that these are the outcomes of transcriptional collisions. This is supported by studies by Hobson *et al.* on eukaryotic RNAPs and Crampton *et al.* on *E.coli* RNAPs who both noted that RNAPs come into close contact upon collisions and remain stably bound to the template [3, 318]. The stability of collided complexes is further indicated by their resistance to HS which should remove any non-specifically bound RNAPs, as well as by the difference in total contour length in comparison to bare DNA. Rivetti *et al.* were able to show that a stalled elongation complex (SECs) caused a decrease in contour length of 22nm but the data agrees better with that of footprinting studies which show that the RNAP interacts with approximately 32 bp which would be equivalent to 10-15 nm [294, 311, 412].

The majority of complexes (61 %) were designated Class 1. These were believed to be the outcome of EC-EC collisions as both RNAPs had travelled downstream of their promoters. These complexes displayed an inter-RNAP distance which indicated that they have stalled in hard contact, this agrees well with studies carried out by Hobson *et al.* who found that eukaryotic RNAP II molecules stalled with a separation of 35 bp between active sites of both RNAPs which is similar to the inter-RNAP distance seen between collided complexes [318].

Experiments by both Billingsley *et al.* and Crampton *et al.* using AFM to investigate collisions between *E.coli* RNAPs found that the majority of complexes were found outside the inter-promoter region (Class 2) [3, 4]. These studies were performed in the absence of heparin or HS. The change in the distribution of classes as compared with their studies indicates that some of those RNAPs believed to have moved upstream of their promoter were the consequence of non-specific binding or initiation events from the blunt ends or nicks rather than collisions [113]. For both eukaryotic and prokaryotic RNAP collisions truncated transcripts were observed [191, 318]. This supports the observation of the high percentage of complexes seen between the promoters. This outcome also agrees with modelling by Sneppen *et al.* who suggested that head on EC-

EC collisions were the most likely form of TI to occur when promoters were separated by a relatively large number of base pairs and both promoters were of similar strength [193]. This is due to the probability of both RNAPs escaping the promoters is equal. The model system used here negates the effects of promoter occlusion by performing OPCs and analysing only a single round of transcription, thus EC-EC collisions are expected to dominate. It is noted that on average the RNAP originating from the long arm travelled 50 bp further than those originating from the short arm promoter. This is unexpected but could be due to the presence of a -10 element like sequence located on the non-template strand 28 bp downstream of the transcription start point. It has been seen that such elements can cause prolonged pausing of the holoenzyme [413]. Experiments by Harden *et al.* indicated that ECs can retain  $\sigma^{70}$  and when an EC passes over such a pause site that transcription rates decrease [414]. This means that it is possible that the RNAP elongating from the short arm promoter may pause and therefore not travel as far along the template as those travelling from the long arm promoter, causing a slight difference in distance travelled.

Collisions between what Crampton *et al.* termed an EC and sitting ducks (EC-SD collisions) still occur, making up 35 % of complexes analysed (designated as Class 2 molecules). Class 2 molecules had either one or both RNAPs downstream of the promoters. These collisions are believed to occur between an EC and an OPC. It has been seen for a number of promoters that two forms of initial transcribing complexes (ITCs) can form, productive and unproductive [415]. Vo *et al.* showed evidence that the type of complex formed occurred at the stage of OPC formation and that productive ITCs were quick to escape the promoter, whereas non-productive ITCs remained at the promoter indefinitely. The formation of unproductive ITCs was found to occur for 28 % of complexes at a consensus promoter by Margeat *et al.* [143]. This study used a single promoter but if the probability of promoter escape is assumed to be the same for the two promoters in the system used here a probability that only one RNAP enters elongation 40 % of the time. Class 2 make up 35 % of complexes analysed indicating that these may be collisions between inactive ITCs and an EC. If Class 2 molecules are plotted separately as a scatter plot it can be seen that the majority of complexes are situated near to the promoter (Figure 6-40).



**Figure 6-40: Scatter plot of just those RNAP pairs that displayed backtracking from the promoter which is indicated by the blue line. The promoter position corresponds to the average value of the position of  $\sigma^{70}$  RNAP in OPCs.**

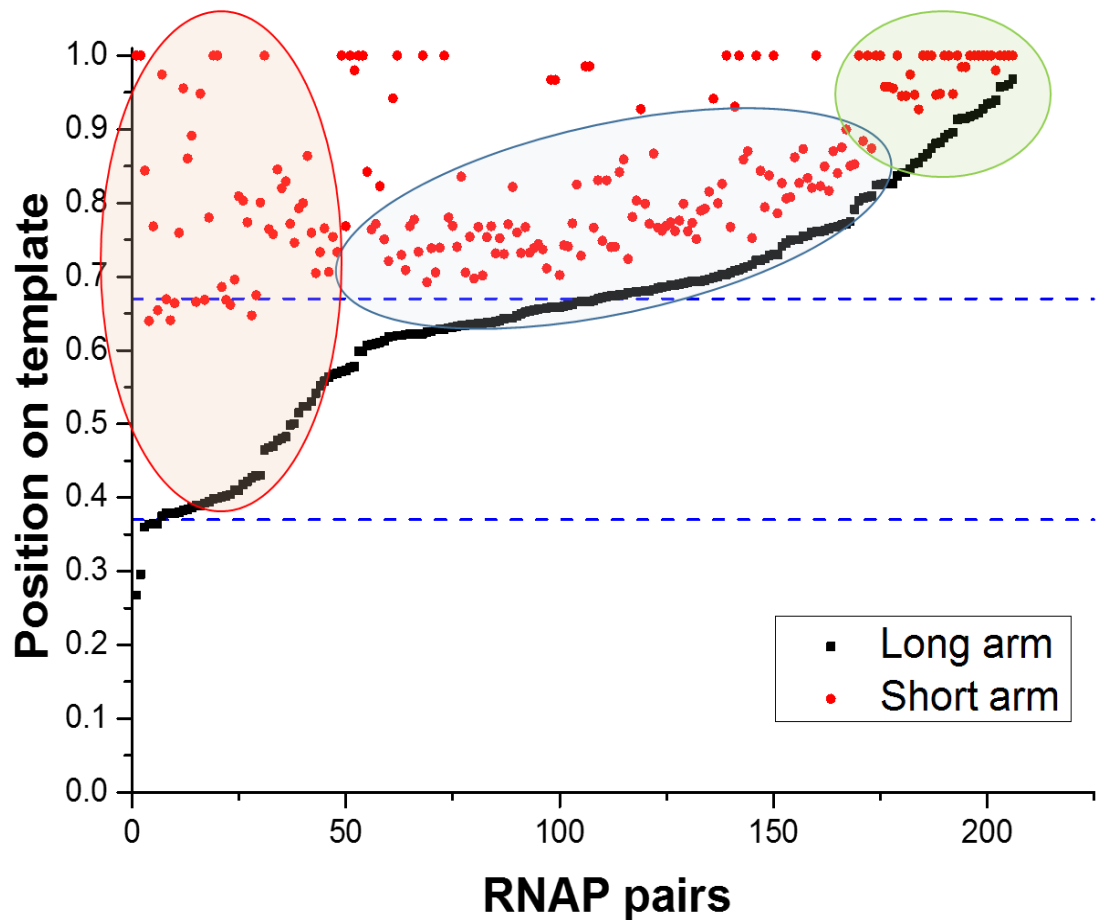
It was also noted by Kuburi *et al.* and Sen *et al.* that these unproductive ITCs can take two main forms of arrested complex [416, 417]. One form is a moribund complex, which is stuck in the abortive cycle but retains transcriptional activity, having a structure similar to that of an active OPC. These moribund complexes can convert to the second type of arrested complex over time which is referred to as a dead end complex which displayed no elongating activity [417]. It is possible that those RNAP pairs that are backtracked by a large amount upstream of the promoter may be the consequence of a collision between a dead end complex and an EC, resulting in a further distance being travelled upstream due to the dead end complex having already undergone some backtracking. Those where only one RNAP is located downstream of its promoter may be the outcome of a collision between a moribund complex which is still located at the promoter undergoing abortive initiation. Upon collision the moribund complex may lose any wrapping of the DNA that may be maintained in this state and then form a dead end complex and then backtrack. This could account for the slightly greater separation between the two RNAPs seen for those located either side of a promoter. Callen *et al.* provided data that suggested OPCs formed at the  $\lambda$  phage promoter  $P_L$  do not act as a roadblock and are removed from the template [192].

The promoters in this experiment were of differing strength from the  $P_L$  promoter and the data indicate that in the case of the  $\lambda_{pr}$  promoter an OPC can act as a roadblock.

#### **6.3.4 Transcription from tandem promoters**

In the case of tandem transcription it is expected that the majority of RNAPs would be located at the loop end of the template if no interference was occurring between the two RNAPs. This is not the case as is shown by the scatter plot and arm length measurements. In the majority of cases both RNAPs are located downstream of the long arm promoter with a separation similar to that seen for head on collisions during convergent transcription. It is reasonable to assume from the separation of the two RNAPs that these are outcomes of collisions but Class 1 complexes may be a combination of EC-SD and EC-EC collisions. Collisions between an unproductive ITC at its promoter are also contained with Class 2 complexes with a lower inter-RNAP contour length. This can be observed by viewing the annotated scatter plot shown in Figure 6-41.





**Figure 6-41: Annotated scatter plot of location of RNAPs pairs as a percentage of the template. Specific areas of interest have been highlighted and are discussed in the text. The red areas highlights molecules that have failed to initiate elongation. The pale blue oval highlights those RNAPs that are believed to have undergone a collision. The pale green oval highlights those RNAPs that have reached the ssDNA loop label.**

Highlighted by the pale blue oval are those complexes that are believed to have undergone a collision and stalled. A collision between an EC-SD could lead to shunting of the SD as in the case of convergent transcription. Studies on collisions between two tandemly transcribing RNAPs have been performed often by stalling a leading EC at a stall site then allowing a trailing RNAP to elongate from the same promoter. In such studies the leading RNAP enters a backtracked state and it has been seen by Epshtein *et al.* that a trailing RNAP can help recover the activity of backtracked complexes [201]. The data here however agrees better with studies by Wang *et al.* who were able to

show that upon collision both RNAPs remained on the template at the stall site [418]. This stalling due to collisions between co-transcribing RNAPs was also suggested by Bentin *et al.* when investigating effects of long nascent RNAs on transcription [419]. It was reasoned that aggregates of RNAPs could be caused by collisions between two active ECs leading to the arrest of both ECs in close proximity [419]. Unproductive ITCs would only account for some of the collisions observed and only those in the vicinity of the leading promoter. Transcription by RNAP has been shown to involve short ubiquitous pauses that are separate from pauses leading to backtracking [162]. These short pauses or the asynchronous escape from each promoter could lead to collisions occurring between two active EC-ECs. It has also been shown that such pauses are unaffected by force and so would not be overcome by collision induced cooperativity [420]. Saeki *et al.* reported that collisions between an eukaryotic RNAP at a pause site and an active EC lead to a higher number of RNAPs escaping their paused state, but approximately 40 % did not escape the pause site indicating that not all collisions lead to re-activation in a system utilising a single promoter [200]. The separation between the two promoters may also play a key role in how the two RNAPs interact. Ponnambalam and Busby found that TI between two tandem promoters only occurred when they were spaced by a set number of base pairs [203]. They noted that the level of transcription from the leading promoter dramatically decreased and the trailing RNAP produced a truncated transcript only when the promoters were separated by 86 bp indicating that a collision between the two RNAPs lead to inactivation of both [203]. This sensitivity to the spacing between promoters may be indicative of the global shape or geometry of the DNA playing a role in the outcome of collisions and TI in general. Cooperation between RNAPs may still be occurring in this system as it can be seen that a fraction of RNAP pairs do reach the end of the template which are highlighted by the pale green oval on Figure 6-41. If RNAPs do exhibit cooperative behavior it is also possible that a second RNAP may be able to push both RNAPs past the single stranded loop as the bypass of roadblocks through RNAP cooperation has been noted in some studies [199, 421]. If this was the case then such molecules would be discounted from analysis.

The final region of the Figure 6-41 highlighted by the pale red oval contains Class 3 complexes which have both RNAPs still located at the promoter, indicating that both

RNAPs have failed to fire. These complexes most likely have not undergone a collision due to the high separation between the leading RNAP and the trailing RNAP. Those where the leading RNAP reaches the loop are most likely due to the failure of the trailing RNAP to fire. The random distribution of both leading and trailing RNAPs over the length of the template in this region may be explained by the previously mentioned occurrence of intrinsic pausing and backtracking of complexes, which can lead to some of these entering an arrested state [160]. It is also important to mention those where the leading RNAP has reached the end of the template and been trapped by the loop but have the trailing RNAP situated near the leading promoter. These complexes may form due to the pausing of the trailing RNAP at the leading promoter. Palmer *et al.* noted that at strong promoters prolonged pausing of an actively transcribing RNAP can occur which in turn can lead to occlusion based TI and may lead in some cases to the RNAP entering a backtracked state [422].

## 6.4 Conclusions

It has been shown that it is possible to incorporate an inhibitor of non-specific binding into *in vitro* transcription reactions for analysis by AFM. The use of HS or heparin in samples containing protein is shown to be a useful method for reducing non-specific interactions of RNAP with the DNA. Deposition of DNA in the presence of the polyanionic molecules can lead to the formation of different regions of varying structure and DNA concentration, indicating that a protein or other molecule is needed to ensure samples adsorb to the surface in an analysable manner. The use of this method is not limited to RNAP alone but is important for AFM studies of any proteins that are inhibited from non-specific binding by heparin or HS. It is noted that in the absence of RNAP, heparin and HS are able to prevent binding of DNA to a mica surface. This is most likely due to the highly polyanionic nature of these macromolecules.

The reduction of non-specific interactions means that the outcomes of transcription from both tandem and convergent promoters can be studied with a higher degree of accuracy than achieved previously using *ex-situ* AFM. It is shown that in the case of convergent promoters, RNAPs irreversibly stall when they have undergone EC-EC or EC-SD collisions. Those complexes that undergo an EC-SD collision have been shown

to occur through two different mechanisms involving the different forms of ITCs that can form at a promoter. Our results suggest that one RNAP can shunt another RNAP when a collision occurs between and EC and SD. The cause of stalling in the case of convergent transcription is most likely due to steric hindrance between the two RNAPs and may be due to a disengagement of the template DNA from the active site, but AFM is unable to elucidate the exact cause for the loss of activity. This loss of activity and stability of RNAPs that have undergone a collision indicate that occurrence of such events *in vivo* would lead to altered expression levels of both transcribed genes. This would mean for convergently arranged nested genes that concurrent transcription is not possible and the nesting of genes in such a way may allow for regulation through TI. It is most likely that *in vivo* mechanisms would exist in order to clear such transcriptional blockages as was seen in the case of eukaryotic RNAP II, where complexes were cleared by ubiquitinylation [318].

In the case of transcription from tandem promoters it was observed that the majority of RNAPs do not reach the end of the template but stall outside the promoter region, again in close proximity. This stalling is likely a consequence of different factors and it is not possible to determine which of these is dominant. It is however obvious that addition of second RNAP leads to changes in the behavior of both RNAPS, suggesting a combination of cooperative and/or competitive behavior. It maybe that *in vivo* other factors ensure that tandemly transcribing RNAPs are able to act in a cooperative manner in order to allow transcription to occur concurrently and at a high rate. It is also possible that the spacing of tandem promoters plays a role in how the RNAPs interact with each other. It is not possible elucidate the cause of collisions using *ex situ* AFM as only the aftermath of collisions is analysed. The data does provide insight into how two RNAPs may interact with each other when transcribing in tandem.

This page is left intentionally blank

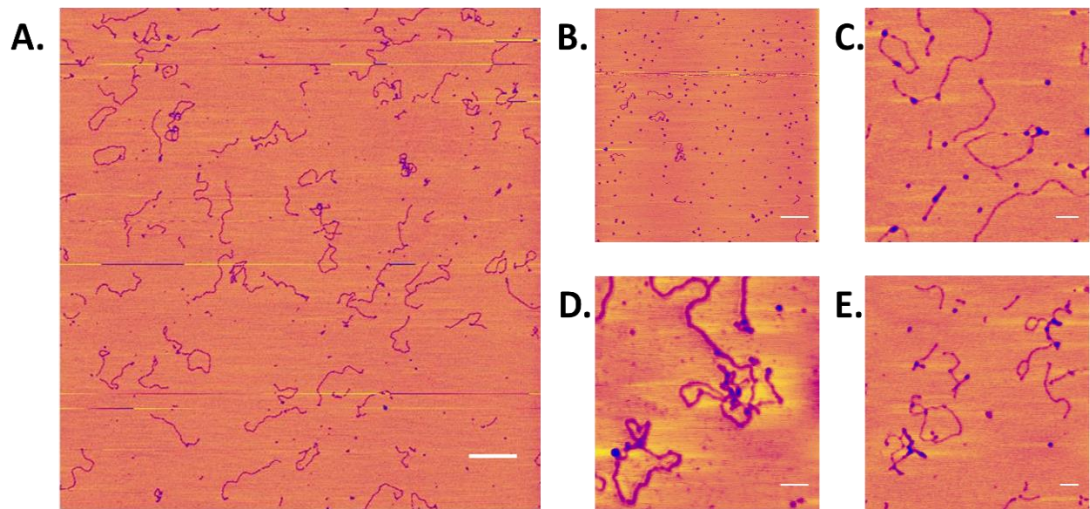
# 7 Effects of supercoiling on transcriptional collisions

## 7.1 Introduction

The data presented in Chapter 5 provides insight into the occurrence of collisions between co-transcribing RNAPs. Information on the spatial arrangement of RNAPs after a collision indicate that RNAPs arrest in close if not hard contact and remain stably bound to the template even in the presence of an inhibitor of non-specific binding, heparin sulphate (HS). One question that arises is what factors govern and affect these collision events. DNA topology has been shown to play an important role both at a global and local level in transcription and gene expression. With RNAP having to open the DNA duplex during OPC formation and elongation, transcription is associated with changes in DNA topology and supercoiling. This change in topology was first proposed and modelled by Liu and Wang using the twin supercoiled domain theory [57]. This theory states that DNA downstream of an elongating RNAP will be positively supercoiled and DNA upstream will be negatively supercoiled when the RNAPs or template are not free to rotate [57]. The DNA templates used for studies of RNAP collisions in Chapter 5 are considered topologically open as they are linear and not anchored, allowing for the rapid dissipation of any topological alterations, such as supercoiling, that may arise from the elongation process. In order to investigate the effect that topological changes arising from elongation may have on collisions, topologically constrained templates as well as a method for detecting topology is needed.

The incorporation of HS into samples allows a more accurate and precise study of larger templates which would otherwise be difficult if not impossible to analyse due to non-specific binding. With this in mind a new template consisting of longer arm lengths was synthesised and initially circularised to provide a topologically closed DNA template. Upon AFM analysis the number of circular templates was found to be very low. It was also noted that upon formation of OPCs an even lower

number of templates remained as closed circles. It was also seen that binding of the RNAP before the addition of NTPs lead to the formation of DNA crossovers and unanalysable structures (Figure 7-1).



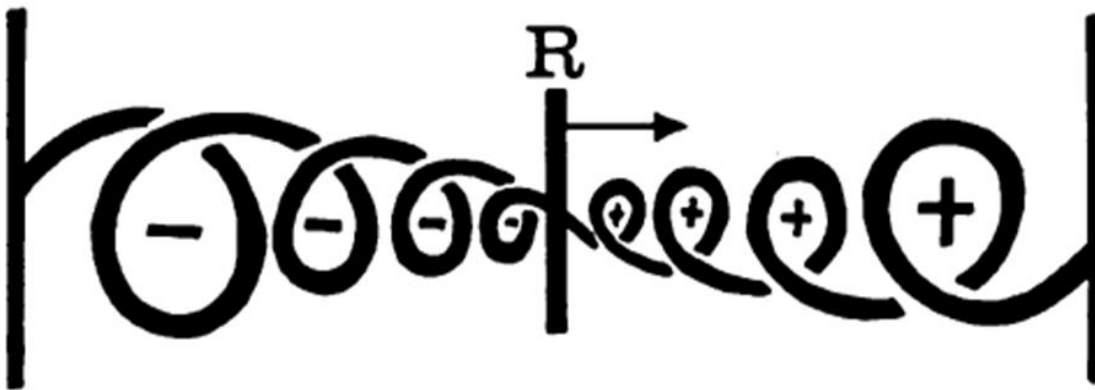
**Figure 7-1: AFM height images of circularised template. A) Bare circularised template. A few closed circles can be seen but the majority of DNA is linear (Scale bar = 1µm). B, C, D, E) AFM height images showing structures formed upon addition of RNAP to form OPCs. It can be seen that crossovers and convoluted structures were common (Scale bars: B = 1µm; C = 100 nm; D = 100 nm; E = 100nm)**

These structures meant that a circular template would not be useful for statistical analysis. Theoretical and experimental evidence (which is discussed in section 6.1.1 in this chapter) suggested however that with an increased template length transient supercoiling may still occur in linear templates. As this supercoiling is only transient and is expected to dissipate relatively quickly it was not necessarily expected to be visualised by *ex situ* AFM. In order to test whether any changes seen during collisions were due to topological changes occurring at the elongation stage, topoisomerase IB (Topo IB) was added when elongation was initiated. Presented in this chapter are the outcomes of transcriptional collisions on linear templates 2521 bp in length, in the absence and presence of Topo IB.

### 7.1.1 Supercoiling and transcription

The link between transcription elongation and supercoiling was first explained and modelled by Liu and Wang (1987) in light of a number of experiments that produced

previously unexplainable results [57]. Liu and Wang proposed that as the RNAP tracks the helical DNA duplex opening up the DNA downstream and resealing the DNA upstream creating the transcription bubble, either the RNAP or the DNA must rotate. It was postulated that as the transcription complex, comprising the RNAP and nascent RNA as well as other associated proteins, tracked along the DNA it would be easier to rotate the DNA [57]. This rotation leads to a wave of positive supercoiling downstream of the transcription complex and negative supercoiling upstream (Figure 7-2).



**Figure 7-2: Schematic diagram representing the generation of supercoils according to the twin supercoil domain theory. R is the RNAP molecule with the arrow indicating its direction of transcription. Downstream of the RNAP positive supercoiling occurs while upstream negative supercoiling occurs [57].**

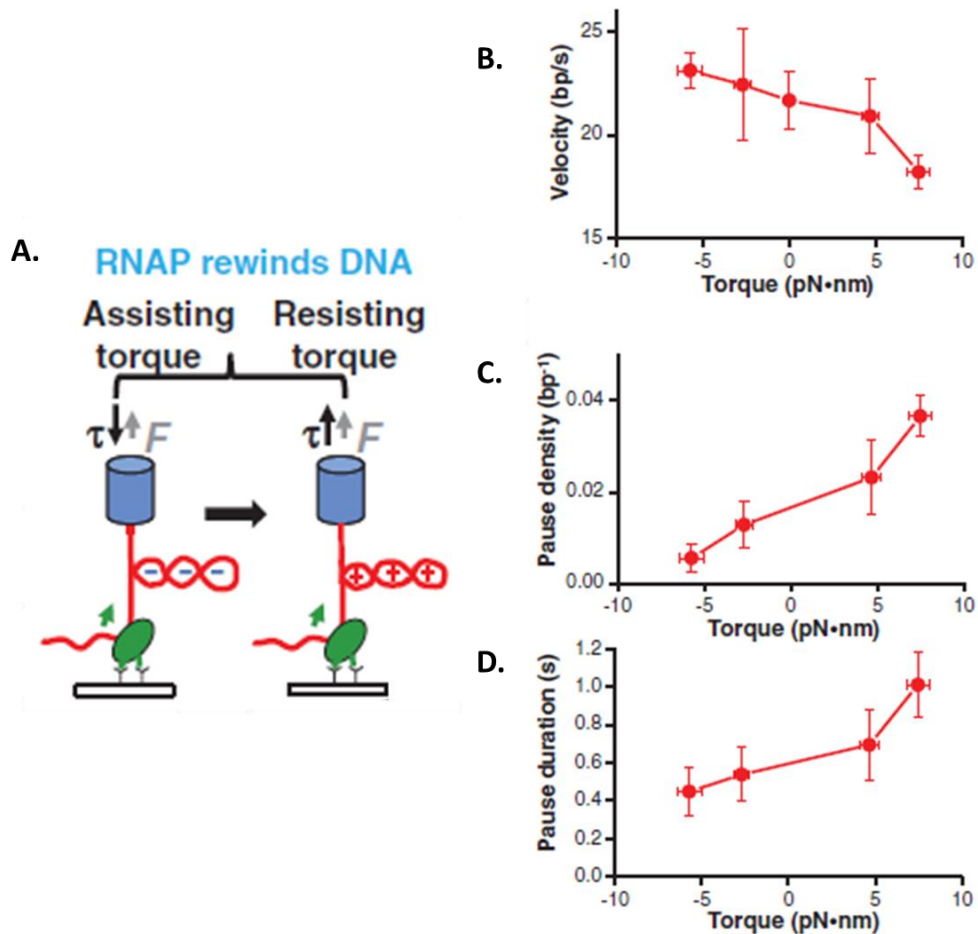
This hypothesis appeared to fit well with results obtained in subsequent experiments by a number of groups [423-425]. Both Giaver *et al.* and Wu *et al.* provided biochemical evidence that supercoiling by bacterial and eukaryotic RNAPs occurred during the elongation stage of transcription [423, 424]. Wu *et al.* were able to show that the levels of supercoiling induced by transcription elongation was affected by the RNA transcript length and rate of transcription [424]. Further confirmation of the validity of the twin supercoiled domain theory came from Tsao *et al.* who were able to show through biochemical methods that the level of supercoiling and therefore the torsional stress applied to the DNA by RNAP was larger than first suggested [425]. The effect of transcription-generated supercoiling has been noted at positions as far as 2 Kbp upstream of a transcriptional start site indicating that transcription may have the ability to influence a number of processes including transcription itself at relatively



### *Chapter 7: Supercoiling and collisions*

distant sites [6, 426]. More recently, single molecule experiments using optical and magnetic tweezers have provided a deeper understanding of forces exerted on the template DNA during elongation. Ma *et al.* used an angular optical trap to provide a value of the torsional force that RNAP was able to exert [427, 428]. It was found that *E.coli* RNAP was able to exert a torque of  $11 \pm 4$  piconewton-nanometre before stalling. This is consistent with an earlier study by Yin *et al.* who measured a stall force of  $14 \pm 4$  pN [317, 427]. This level of force is noted as being large enough to melt arbitrary sequences of DNA, form non-B DNA structures and also form plectonemic or toroidal structures [429]. The ability of RNAP to alter the structure of DNA suggests the importance of local supercoiling on the global gene expression of a cell [426, 430].

The experiments by Ma *et al.* also indicated that stalling of elongation can occur due to both positive and negative supercoiling [427]. Results also showed that after removal of torsion 50 % of the RNAPs regained activity [427]. It was also found that with an assisting torque applied, transcription rate increased and pausing decreased, with the opposite of this being true for a resisting torque as is seen in the graphs shown in Figure 7-3 [427].



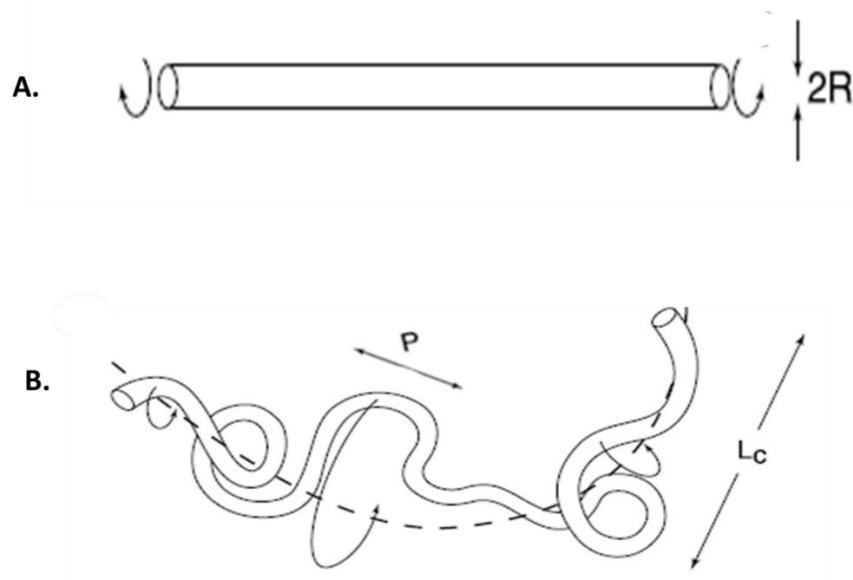
**Figure 7-3: Experimental setup and data collected by Ma *et al.* using an optical trap setup. A) The experimental setup involved the attachment of an RNAP (green) to a surface and applying either an assisting or resistive torque to the DNA. B) Plot showing the change in velocity seen when torque is increased. C) Plot showing the increase in pause density as higher resisting torque is applied to the RNAP. D) Plot showing the relationship between pause duration and torque. As resisting torque increases pause duration also increases [428].**

Transcriptional stalling due to supercoiling or torsional stress is a consequence that had been noted before, but not at the single molecule level. Joshi *et al.* showed that when high levels of positive supercoiling were induced in a cell, 80 % of genes in the yeast *Saccharomyces cerevisiae* displayed down regulation [63]. The regulation of transcription due to excess levels of both positive and negative supercoiling has been subject of a number of reviews [6, 60, 61, 429]. In many of the *in vivo* experiments the accumulation of supercoiling was induced by the removal of topoisomerases. Topoisomerases are proteins that can introduce and remove supercoiling from the

Chapter 7: Supercoiling and collisions

DNA template and so play an important role in regulating the levels of supercoiling in the cell. A more detailed overview of topoisomerases is given in section 6.2.

In a system where the DNA, whether linear or a closed circle, is free to rotate supercoiling would not be expected to occur. This assumption is due to the fact that the free rotation of the DNA would not be hindered enough by its own viscous drag to allow for supercoiling. This has been suggested to not be entirely true. Many theories and models of DNA rotation use the speedometer-cable model proposed by Levinthal and Crane [431]. This theory assumes that DNA when rotated would be like a rigid rod and so would display a low viscous drag (Figure 7-4 A) [431].



**Figure 7-4: Schematic representations of DNA rotation when considered either as a rigid rod (A) or a naturally bent semi-flexible rod. A) The Levinthal and Crane model assumes DNA to be rigid and therefore its viscous drag when cranked or rotated is limited by the diameter of the rod. B) Representation of DNA as a naturally bent semi-flexible rod. When cranked the viscous drag experienced by the rod is related to the persistence length ( $P$ ) and the crossover scale  $L_c$ . On length scales shorter than  $L_c$  the DNA rotates as a rigid rod and over  $L_c$  the DNA will flex and therefore rotate via hybrid motion [432].**

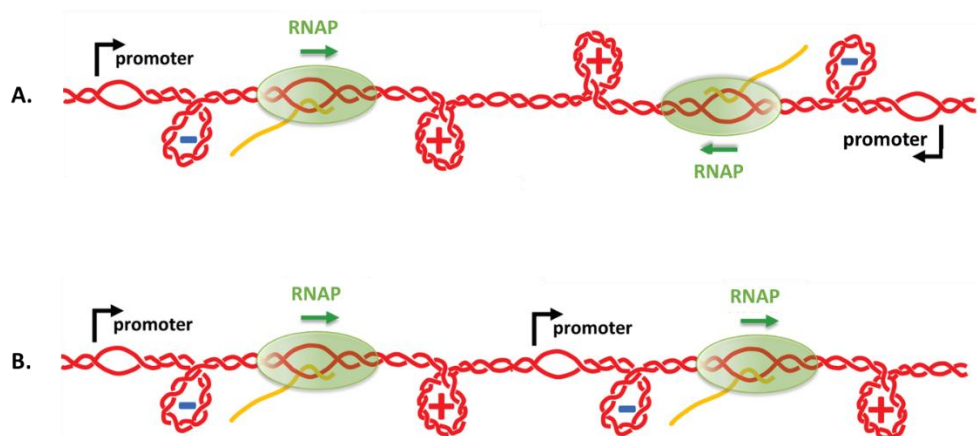
## Chapter 7: Supercoiling and collisions

This model was considered as “naïve” by Nelson who highlighted that DNA cannot be regarded as a rigid rod, as it is known that DNA with a length greater than its persistence length would have intrinsic bends and curvature (Figure 7-4 B) [432]. Nelson proposed the hybrid-motion theory which takes into consideration these intrinsic bends and curves found in DNA. The theory states that if the DNA has a length greater than its persistence length ( $P$  in Figure 7-4 which is approximately 52 nm for DNA) then only sections of the DNA below a crossover length ( $L_c$ ) will rotate as a rigid rod, while the rest of the DNA will not [432]. Nelson calculated a value for  $L_c$  of approximately 1.4 Kbp or 450 nm for DNA being transcribed by T7 RNAP and slightly longer for slower rate RNAPs [432]. The difference in  $L_c$  due to the speed of an RNAP and its molecular weight is due to the crossover length being effected by the speed that the flexible rod is cranked. Therefore a faster moving RNAP or protein will rotate the DNA faster, leading to a lower  $L_c$  value. The theory suggests that when a double-stranded DNA fragment greater than approximately 1 Kbp is cranked, as in the process of transcription, the DNA will become “spin locked” and so the rate of supercoiling diffusion greatly diminishes. As  $L_c$  is influenced by bends in the DNA, molecular weight and transcription rate, the introduction of multiple transcribing RNAPs, which introduce significant bends into the DNA, would decrease the length  $L_c$ . Support for this model is found in experiments by Tsao *et al.* as well as by experiments by Dröge and Nordheim who studied levels of transcriptional supercoiling of a 3 Kbp plasmid [433]. They noted a change in the DNA from B-form to Z-form as a consequence of supercoiling [433]. Further experiments by Dröge where the site specific recombination acted as a reporter of supercoiling indicated that supercoiling also occurred on linear templates of 7.5 Kbp [434]. A similar study by Kouzine *et al.* using divergent promoters and a recombination-reporting system showed that a highly negative supercoiled domain was able to form and was sustained between the two promoters on a linear template of 3 Kbp [435]. The anchoring of the DNA or RNAP was also shown to not be required for supercoiling to occur in experiments by Drolet *et al.* as well as by Lilley *et al.* who studied the effect of anchoring of the DNA templates *in vivo* [436, 437]. Further results which support the hybrid-motion and spin locking theory can be seen from work carried out by Dunaway *et al.* [438]. These experiments involved the injection of bacterial RNAP and linear DNA fragments of 3 – 4 Kbp into *Xenopus* oocytes which are unable to transcribe linear DNA. The results indicated that

## Chapter 7: Supercoiling and collisions

local and transient domains of supercoiling could occur in open DNA templates therefore allowing for the transcription of the linear DNA [438]. This theory has come under some scrutiny in the light of results obtained by DNA unzipping experiments carried out by Thomen *et al.* who originally calculated a value for the rotational drag of the DNA to be 38,000 times smaller than that predicted by the hybrid model. This was later rebuked and accepted as a calculation error with the actual value being 4 times smaller than that predicted by the hybrid rod theory and 10 times larger than the rigid rod theory [439, 440]. Theoretical modelling also provides support for the hybrid-rod model many of which have been reviewed by Nelson [432, 441].

When considering the occurrence of supercoiling in the case of convergent promoters it is expected that a region of hyper-positive supercoiling would occur downstream of both promoters (Figure 7-5 A). This was observed in experiments by Wu *et al.* who investigated the occurrence of transcriptional supercoiling from different promoter arrangements, finding that promoter orientation plays a role in the levels and handedness of supercoiling observed [424]. In the case of a head-on collision, either between two ECs or an EC and SD, the effect of local supercoiling is not known. Indirect evidence exists from studies of convergent transcription by RNAP II from eukaryotes that elongation is impaired due to supercoiling [442]. The formation of separate domains of supercoiling through the binding of RNAPs and spin locking is expected, due to the wrapping during OPC formation but also due to bends introduced by ECs as well as the increased molecular weight and dimensions of the DNA-RNAP transcription complex.



**Figure 7-5: Diagram of expected supercoiling between both convergent and tandem promoter arrangements. A) With convergent promoters a region of hyper-positive**

**supercoiling is expected to occur between two active ECs. B) In the case of tandem promoters, it is expected that two ECs would produce opposite supercoiling, meaning that opposite supercoils should be able to merge and obliterate each other (adapted from [429]).**

In the case of tandemly arranged promoters if both RNAPs are elongating at the same time it is expected that the opposite regions of supercoiling would obliterate each other as is seen in Figure 7-5 B. The merging of oppositely orientated supercoils was shown to not be hindered by intrinsic bends in the DNA through investigations by Stupina and Wang [443]. They did find however that a second transcription unit impeded the merging of oppositely orientated supercoils, suggesting that this may be due to a stable bend introduced by the RNAP as well as greatly increased frictional barrier to free rotation of the DNA [443]. If the transition into elongation is asynchronous then there is the possibility that a region of supercoiling may form between the two RNAPs, this has been suggested as the cause of transcriptional bursting seen in highly transcribed genes, with supercoiling becoming too great for elongation to continue until it has been relieved by topoisomerases or other factors [444, 445].

It is the outcome of collisions on a DNA template of a length that would allow for spin locking to occur that is investigated here. Using the high spatial resolution of the AFM and the incorporation of HS into samples to prevent non-specific interactions from having an adverse effect on sample analysis, the outcomes of both convergent and tandem collisions can be analysed. As the supercoiling is expected to only be transient and relatively low, imaging of structural transitions may not be likely, as imaging supercoiled DNA poses a number of issues during analysis as noted by Lyubchenko [206]. In order to overcome this issue an indirect method to determine whether supercoiling is occurring is needed. This was achieved through the use of a topoisomerase which is discussed in the subsequent section.

### **7.1.2 Topoisomerases**

There are a number of topoisomerases known, the first being discovered by James Wang in 1971 in *E.coli* [446]. All topoisomerases share the ability to relax negatively

*Chapter 7: Supercoiling and collisions*

supercoiled DNA, but not all share the same mechanism. Some topoisomerases are also able to decatenate DNA rings as well as introduce supercoiling into DNA. A summary of the known topoisomerases along with their features is given in Table 7-1.

Chapter 7: Supercoiling and collisions

Enzyme	Type	Source	Size (kDa)	Notes
Bacterial topoisomerase I ( $\omega$ protein)	IA	Bacteria ( <i>E.coli</i> )	97 (monomer)	Cannot relax positive supercoils
Eukaryotic topoisomerase I	IB	Eukaryotes	91 (monomer)	Relaxes both positive and negative supercoils
Vaccinia topoisomerase I	virus IB	Vaccinia virus	37 (monomer)	Relaxes both negative and positive supercoils. Stimulated by ATP
Topoisomerase III	IA	Bacteria	73 (monomer)	High decatenating activity
Reverse gyrase	IA	Thermophilic Archaea	143 (monomer)	ATP-dependant introduction of positive supercoils
DNA gyrase	IIA	Bacteria	97 and 90 (A <sub>2</sub> B <sub>2</sub> )	ATP-dependant introduction of negative supercoils
T4 topoisomerase	IIA	Bacteriophage T4	58, 51, 18. (2 of each subunit)	Can relax but not supercoil DNA (ATP dependant)
Eukaryotic topoisomerase II	IIA	Eukaryotes	174 (homodimer)	Can relax but not supercoil DNA (ATP dependant)
Topoisomerase IV	IIA	Bacteria	84 and 70 (C <sub>2</sub> E <sub>2</sub> )	Can relax but not supercoil DNA. High decatenating activity (ATP dependant)
Topoisomerase VI	IB	Archaea	45 and 60 (A <sub>2</sub> B <sub>2</sub> )	Can relax but not supercoil DNA (ATP dependant)

**Table 7-1: Table of known topoisomerases showing information on type, source, structure and functionality (adapted from [27]).**

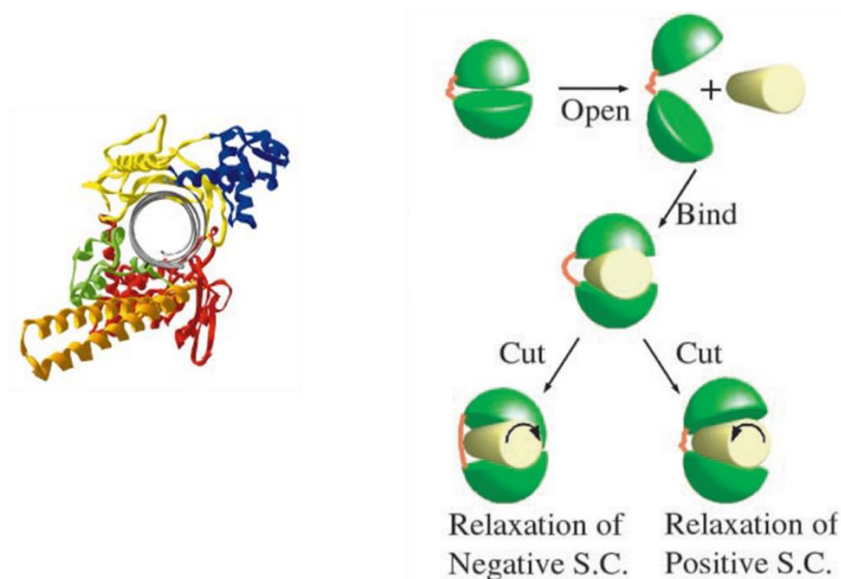
As shown in Table 7-1, topoisomerases are broken down into two types (I and II) and these are then subdivided (A and B). Type I topoisomerases perform reactions in which



## Chapter 7: Supercoiling and collisions

only one strand of the DNA duplex is broken whereas type II break both strands of the DNA helix [27]. For the purpose of this thesis only type IB topoisomerases are discussed but a review of the structure, function and mechanism of all topoisomerases can be found elsewhere [59].

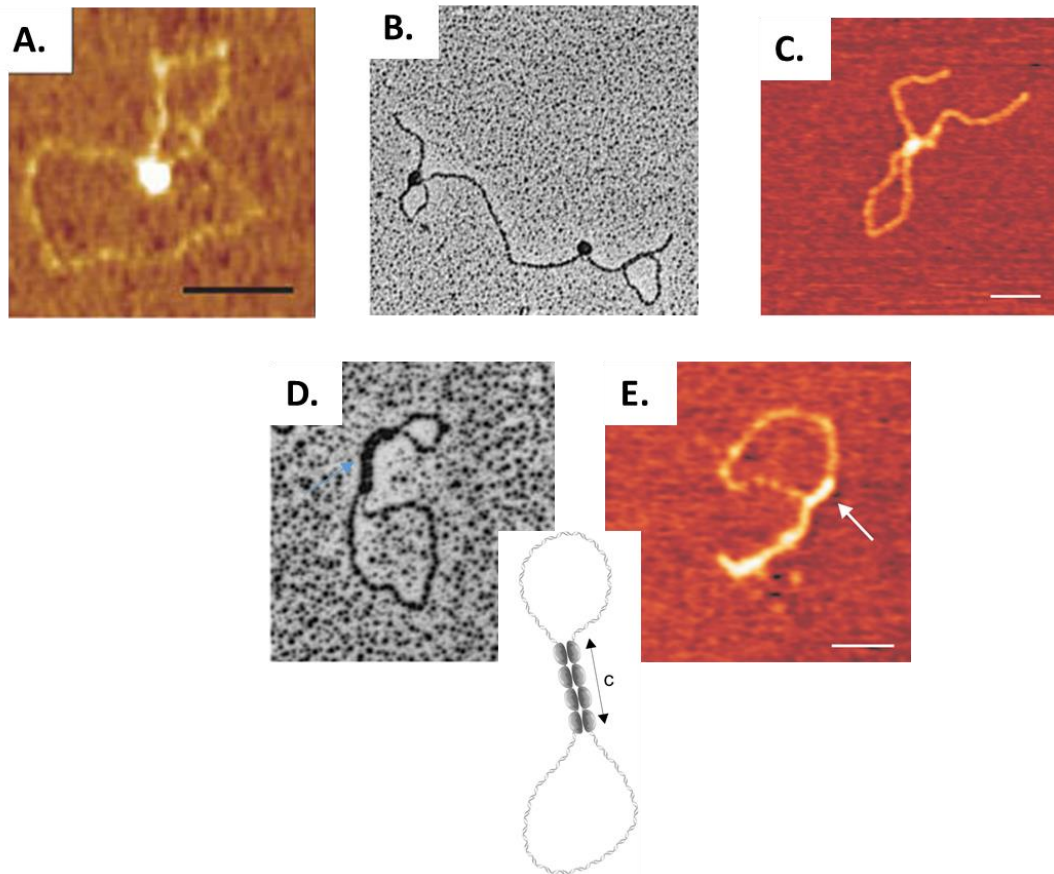
For the purpose of addressing transcriptionally derived supercoiling between convergent and tandem promoters a topoisomerase that is able to relax both positive and negative supercoils was desirable, so a type IB enzyme was selected. There are three main classes of type IB topoisomerases and of these *Vaccinia* virus and eukaryotic Topo IB are commercially available. Both topoisomerases are believed to work by the same mechanism which occurs in five steps: 1) non covalent binding; 2) generation of a single stranded break; 3) rotation of the DNA; 4) re-ligation of the DNA; 5) dissociation. Both of these topoisomerases have a clamp like structure which has two lobes joined by hinge (Figure 7-6).



**Figure 7-6: Crystal structure of Topoisomerase IB from *Vaccinia* and diagram depicting the relaxation of both positive and negative supercoils. On the left Topoisomerase IB in complex with DNA is shown. On the right are the steps involved in binding and relaxing both negative and positive supercoils. The enzyme binds and then allows rotation of the DNA within the active site. Different conformational changes are associated with different handedness of supercoiling [59, 447].**

As is seen in Figure 7-6, Topo IB closes around the duplex DNA and cuts a single strand before allowing the DNA to rotate within the active site. The enzyme binds the DNA and cuts it through a nucleophilic attack by hydroxyl group of the tyrosine 723 on the scissile phosphodiester bond which breaks a single strand of the DNA [59]. A covalent intermediate in which the active site tyrosine attaches to the 3' phosphate end of the cleaved strand is then formed [59, 448]. The DNA is free to rotate with the direction of rotation depending on the handedness of supercoiling present [447]. The DNA downstream of the binding site has been shown to rotate while the enzyme is in a closed state, from crystal structures [449]. It was shown by single molecule experiments carried out by Koster *et al.* that this rotation does not occur freely but is a controlled mechanism governed by the levels of torque and friction [450]. The rotation was also seen to occur in steps and the number of supercoils removed from the DNA was dependent on the torque of the supercoiling, with between 5-15 supercoils removed in each cycle [450]. The process does require additional factors such as  $Mg^{2+}$  or ATP, even though ATP can stimulate the activity of *Vaccinia* Topo IB.

One main difference between *Vaccinia* and eukaryotic Topo IB is the binding specificity. The determined consensus binding sequence for eukaryotic Topo IB is: 5' (A/T) (G/C) (A/T) T 3'. This sequence in itself is quite non-descript and a consensus sequence has been shown to not be required for binding and cleavage to occur [59]. It has been noted though that *Vaccinia* Topo IB has a higher sequence specificity than eukaryotic Topo IB having a consensus sequence of: 5' (T/C) C C T T 3' [59, 451]. Studies have revealed that DNA topology has a greater role in encouraging binding of Topo IB than sequence, with eukaryotic Topo IB having been shown to have at least a 60 fold higher affinity for supercoiled DNA over linear DNA [452, 453]. An interesting observation seen for both types of Topo IB is that binding occurs at DNA synapses or crossovers. This has been observed by AFM for *vaccinia* Topo IB as well as by electron microscopy as is shown in Figure 7-7 [454-456].

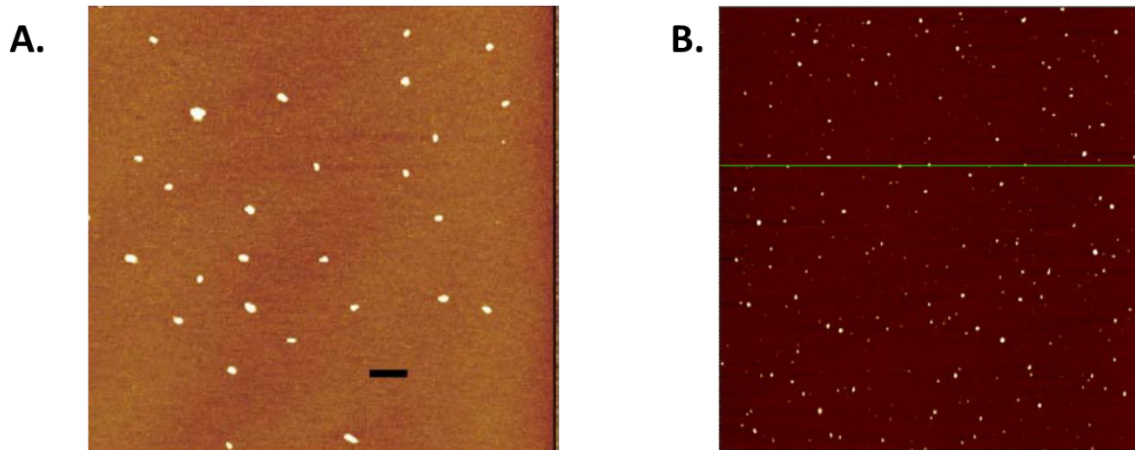


**Figure 7-7: Montage of AFM and electron microscopy images of synapses and filaments seen for binding of Vaccinia Topo IB. A, C and E show AFM images collected by Argaman *et al.* (A) and Moreno-Herrero *et al.* (C and E). B and D show electron micrographs collected by Shuman *et al.* Inset between D and E is a drawing showing the predicted structure of filamentous synapses formed by Vaccinia Topo IB [454-456].**

AFM and electron microscopy studies also found that it is possible for cooperativity to occur for *Vaccinia* Topo IB when binding, leading to the formation of filamentous structures as is shown in Figure 7-7 D and E [455, 456]. The lower sequence specificity of eukaryotic Topo IB means that it can be used with the templates in this work which do not contain specific cleavage sites for *Vaccinia* Topo IB. Another advantage of eukaryotic Topo IB is that it has been shown to be required for transcription elongation and so its role in relaxing transcription-complex supercoiling is confirmed [457, 458].

Previous AFM studies of eukaryotic Topo IB indicate that it is readily adsorbed onto the mica surface. In studies by Argaman *et al.* calf thymus Topo IB, at a final concentration

of  $0.4 \text{ U}/\mu\text{l}$ , was deposited onto mica and imaged in air [454]. The Topo IB was seen to have a height of  $2.8 \pm 0.4 \text{ nm}$  and a diameter of  $27.6 \pm 3.1 \text{ nm}$  [454]. It can be seen from Figure 7-8 A that this provided well dispersed Topo IB molecules on the surface.



**Figure 7-8: AFM images of calf thymus Topo IB deposited on mica and imaged in air. A) Calf thymus Topo IB imaged in air at a concentration of  $0.4 \text{ U}/\mu\text{l}$  [454] (Scale bar = 500 nm). B) AFM height image collected by Liu *et al.* of human Topo IB in air on mica at a concentration of  $0.3 \text{ U}/\mu\text{l}$  [459] (Scale bar =  $3 \mu\text{m}$ ). It can be seen from both images that the adsorption of the Topo IB to the surface can differ between samples.**

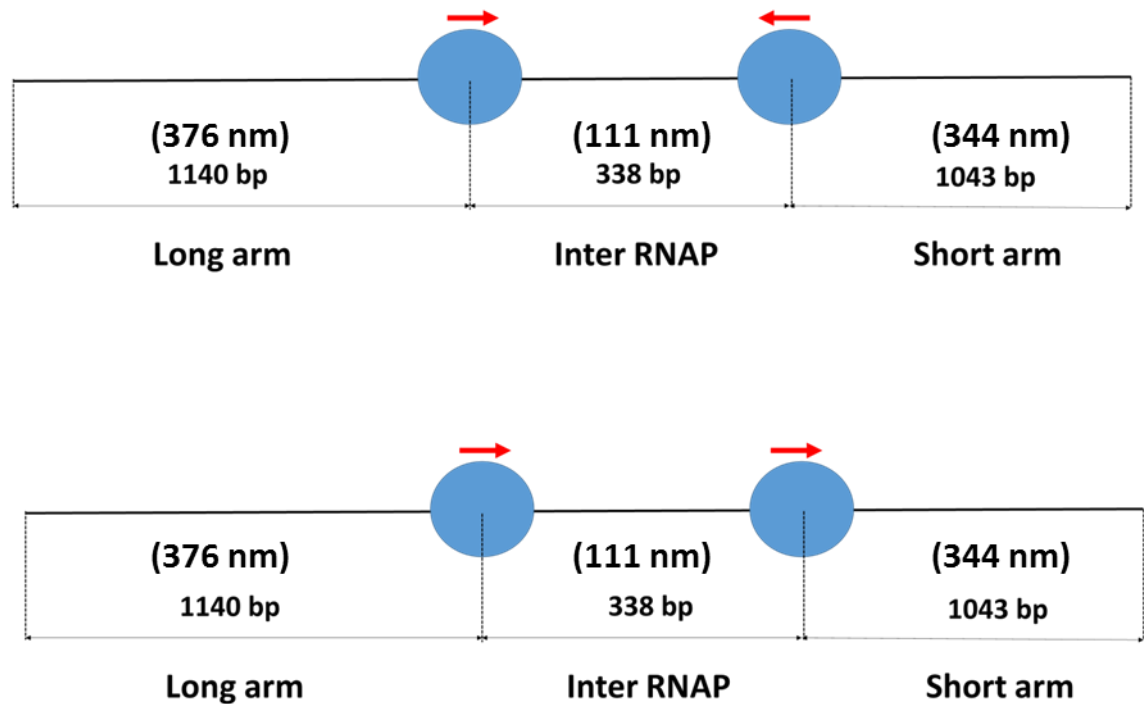
Studies by Liu *et al.* found that in air human Topo IB had a diameter of  $20.3 \pm 3.0 \text{ nm}$  and a height of  $2.6 \pm 0.3 \text{ nm}$  [459]. A final concentration of  $0.3 \text{ U}/\mu\text{L}$  was deposited onto mica but unlike that seen by Argaman *et al.* the amount of adsorbed Topo IB was much higher (Figure 7-8 B). Both these studies provided precedent for the incorporation of eukaryotic Topo IB into AFM samples.

## 7.2 Sample preparation

The DNA template was produced by a standard GoTaq Polymerase PCR reaction using the primers  $5' \text{ ATCTTCAACTGAAGCTTTAGAGCG } 3'$  (forward) and  $5' \text{ GTGTGAAATACCGCACAGATG } 3'$  (reverse) for both convergent and tandem promoter templates. The products had their size checked by 1% (w/v) agarose gel electrophoresis before being purified using the QiaQuick purification column system.

Chapter 7: Supercoiling and collisions

The DNA template used is shown in Figure 7-9 with the promoter locations and arm lengths denoted.



**Figure 7-9: Diagram showing both convergent and tandem promoter templates. RNAPs are represented by blue circles, DNA by the black line and promoters by the red arrows. The size of each region of the template is shown in base pairs and is the same in both cases.**

OPCs were formed by mixing 200 fmol of DNA with 400 fmol of  $\sigma^{70}$ RNAP in transcription buffer before incubation at 37°C for 15 minutes. HS was then added to a final concentration of 200  $\mu\text{g}/\mu\text{L}$ . In order to initiate transcription all four NTPs were added to a final concentration of 100  $\mu\text{M}$  and incubated for 20 minutes at room temperature. Samples were then diluted 1 in 20 in imaging buffer before being deposited onto freshly cleaved mica. Samples were then incubated for 5 minutes before being rinsed with  $\text{dH}_2\text{O}$  and dried under a weak flux of nitrogen.

Wheat germ topoisomerase IB (Sigma, St. Louis, MS) was stored in single-use 5  $\mu\text{L}$  aliquots at -80 °C in the manufacturer's buffer, to avoid multiple freeze thaw cycles. The Topo IB had a concentration of 8 U/ $\mu\text{L}$  with one unit defined as the amount required to relax 1  $\mu\text{g}$  of the plasmid pGEM97 in 30 minutes at 37 °C. Samples containing just topoisomerase were made by diluting the Topo IB in transcription buffer before then

## *Chapter 7: Supercoiling and collisions*

further diluting 1 in 10 in imaging buffer before deposition. The units of Topo IB quoted in the text refer to the final concentration that was deposited onto the mica surface.

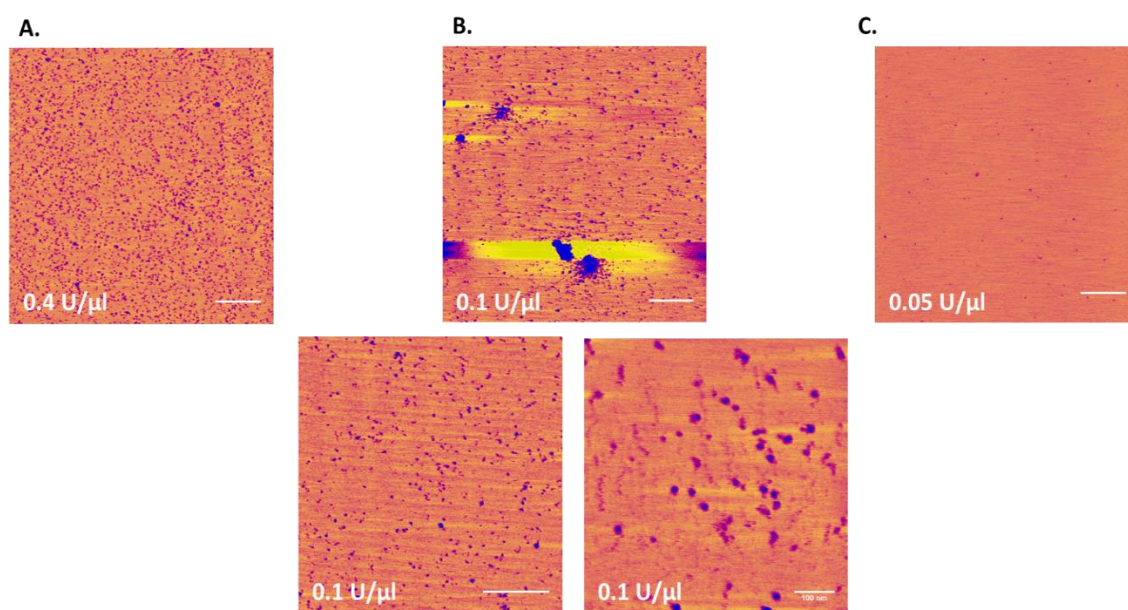
For samples containing plasmid and topoisomerase, 200 fmol of plasmid was mixed with 10 units of Topo IB in transcription buffer then either incubated at room temperature or 37 °C for 20 minutes. When required HS was added to a final concentration of 200 µg/µL before the addition of Topo IB. Samples containing plasmid and RNAP were formed using 200 fmol of plasmid and 400 fmol of RNAP before being incubated at 37 °C for 20 minutes. After incubation, HS was added to a final concentration of 200 µg/ml and incubated for 10 minutes at room temperature. In samples containing Topo IB, 10 units of the enzyme were added and samples were incubated at room temperature for 20 minutes. As 200 fmol of plasmid is too high to distinguish single separated plasmid molecules, samples were diluted 1 in 30 in imaging buffer.

*In vitro* transcription reactions containing Topo IB had OPCs formed in the same way as reactions without Topo IB, except that the Topo IB was added along with the NTPs, prior to incubation at room temperature for 20 min.

## **7.3 Results**

### **7.3.1 AFM analysis of wheat germ Topo IB**

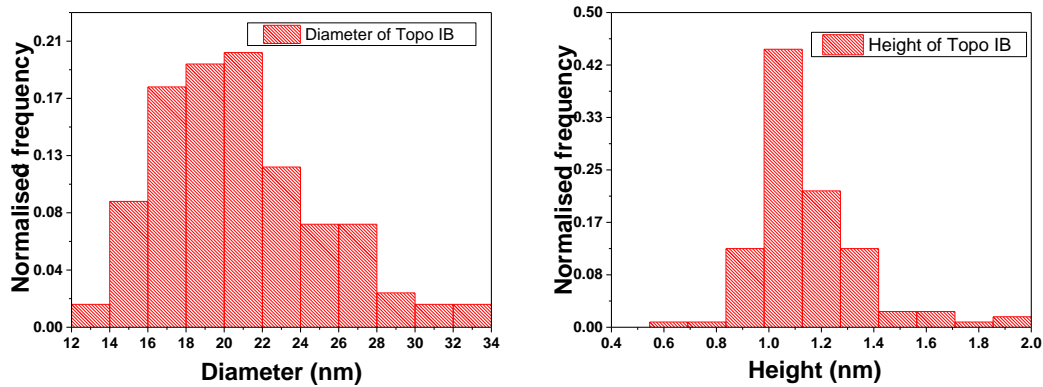
In order to determine how Topo IB interacted with and appeared on the mica surface when imaged by AFM, samples of just Topo IB were prepared and analysed. A range of concentrations were tested (Figure 7-10).



**Figure 7-10: AFM height images of wheat germ Topo IB at differing concentrations deposited on mica and imaged in air. A, B and C correspond to concentrations given in in each image (Scale bars = 1 $\mu$ m). The bottom panel of B shows a zoomed image of Topo IB molecules (Scale bar = 100 nm)**

As it can be seen in the scans shown in Figure 7-10 the Topo IB readily binds to the mica surface. Some areas of aggregation were observed but the majority of Topo IB was dispersed over the surface. At a concentration of 0.4 U/ $\mu$ l surface coverage was considered to be too great to be used in samples containing DNA and RNAP. This was also noted at a concentration of 0.1 U/ $\mu$ l and so it was decided that a final concentration to be deposited would need to be lower. At a concentration of 0.05 U/ $\mu$ L single Topo IB was visible with very few aggregates observed on the surface. This final concentration would require the addition of 10 units of enzyme to a sample of DNA or OPCs before dilution with imaging buffer. The height and diameter of individual topoisomerases were measured; histogram plots of these measurements are shown in Figure 7-11.



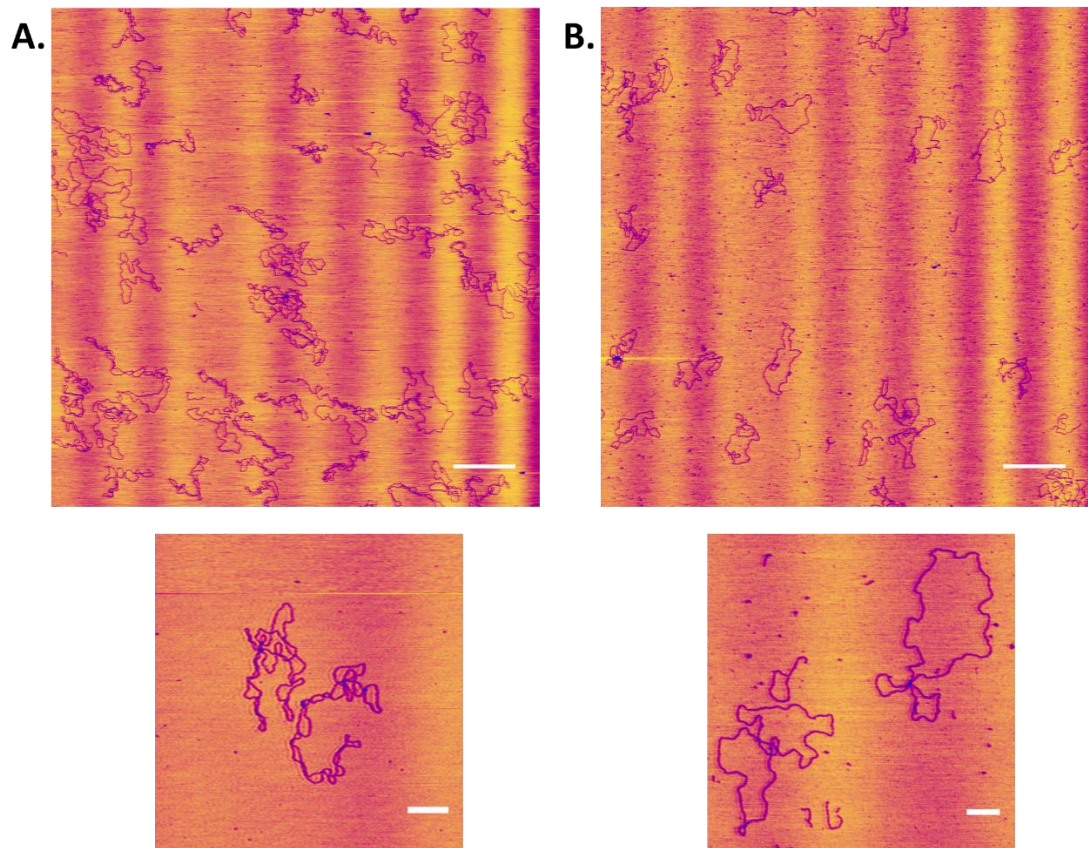


**Figure 7-11: Histogram plots of the diameter and height of Topo IB. The diameter has two major peaks visible, at approximately 17 and 21 nm. These may correspond to different orientations of the protein on the surface. The height shows a single peak at approximately 1-1.2 nm. (n = 120 for both graphs)**

The diameter had an average value of  $20.7 \pm 4.1$  nm but the histogram has a large spread. This may be due to the Topo IB being adsorbed in different orientation on surface. There is also the possibility that the TopoIB has been damaged when adsorbed to the mica. The height data shows a single peak with an average value of  $1.15 \pm 0.2$  nm. The height and diameter of Topo IB is greater than that of DNA but less than that of RNAP, leading to the assumption that it should be distinguishable from both in AFM images.

In order to test the activity of 10 U of the Topo IB in the transcription buffer as well as at room temperature, samples containing supercoiled pDSU plasmid were treated with Topo IB before being deposited and imaged.

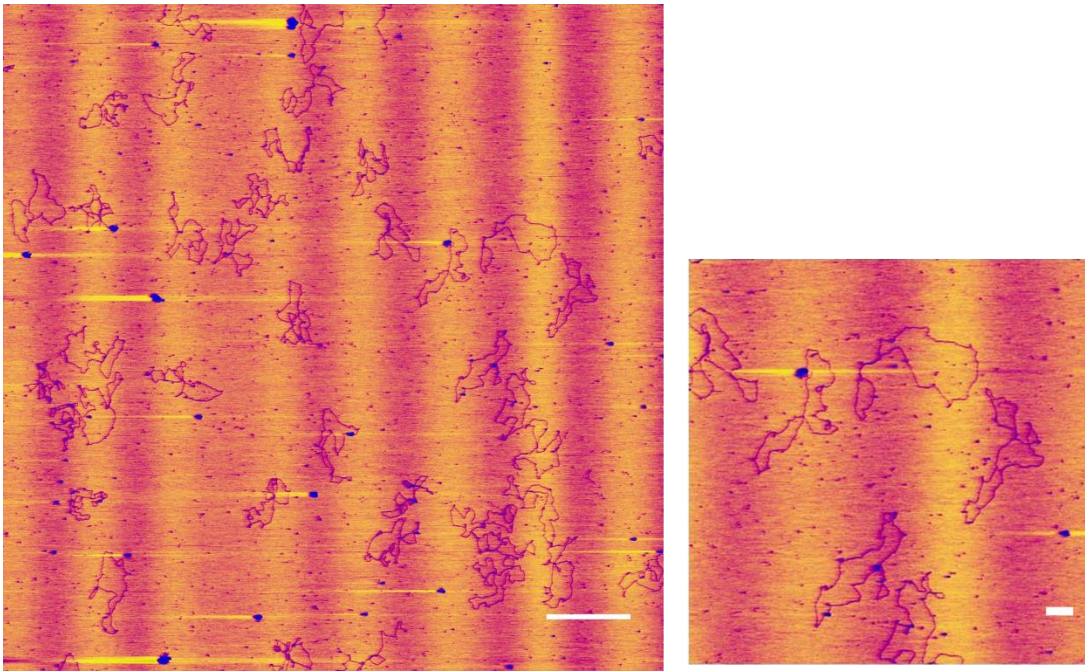




**Figure 7-12: AFM images of supercoiled plasmid and plasmid treated with Topo IB at room temperature. A) Plasmid molecules can be seen to be compacted and demonstrate a large number of crossovers as expected for supercoiled plasmid DNA. The bottom panel shows a zoomed image of two typical plasmids. B) After the addition of Topo IB, the structure of the plasmids can be seen to be more open with fewer crossovers, indicating that there is now a lower level of supercoiling. (Scale bars = 1 $\mu$ m for top panel and 200 nm for bottom panel)**

Plasmid conformation was assessed by counting the number of crossovers of the DNA chain. This method allows for the determination of how supercoiled a structure is as accounts for the writhe and twist of the molecule. The higher the number of crossover points, or nodes, the greater the level of supercoiling, as observed by Jiang *et al* [460, 461]. It was found that the plasmid samples had on average  $14.1 \pm 0.4$  crossovers. With the addition of Topo IB this number decreased to  $3.8 \pm 0.4$  crossovers. As the number of crossovers has been shown to be related to plasmid conformation this decrease indicates that the plasmid is less supercoiled with the addition of Topo IB. This result indicates that Topo IB is active in the transcription buffer. In order to

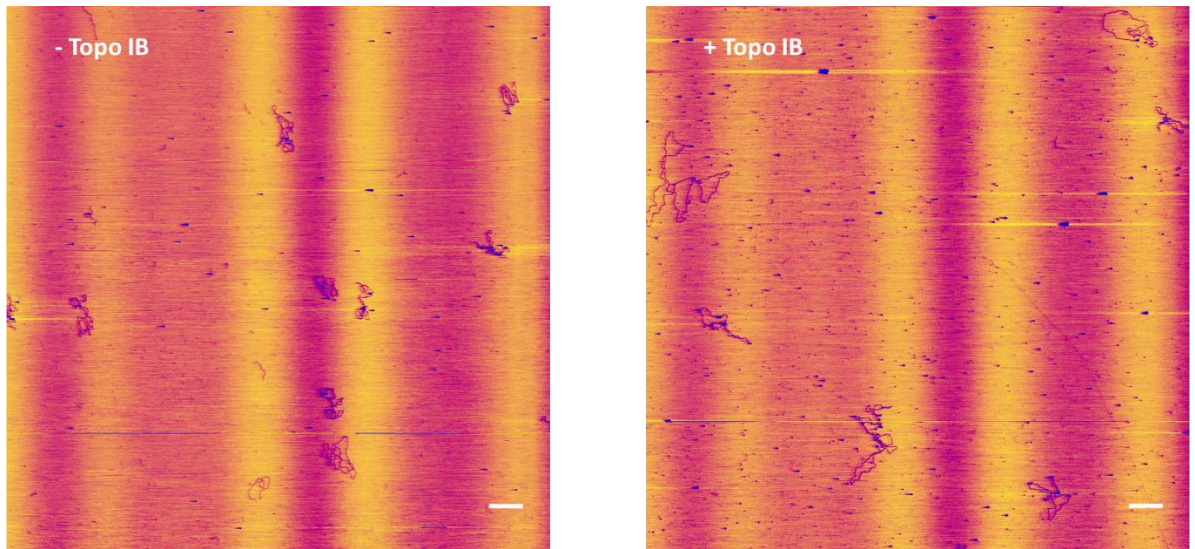
investigate the effect that HS had on activity of Topo IB samples were prepared and imaged that contained HS, Topo IB and plasmid (Figure 7-13).



**Figure 7-13: AFM height images of plasmid treated with 10 U of Topo IB in the presence of HS. It can be seen that the structure of the plasmids changes to a relaxed state compared with the absence of Topo IB. The zoomed image shows relaxed plasmids with Topo IB bound at crossover points. (Scale bar for large image is 1 $\mu$ m and for the zoomed image is 100 nm).**

With the addition of HS there is a decrease in the number of relaxed plasmids seen when compared to samples lacking HS, with the number of crossovers decreasing to  $7.1 \pm 0.4$  in the presences of HS. There is still a significant change in the plasmid conformation indicating that Topo IB is still able to relax supercoils. The lower decrease in crossover number is most likely due to Topo IB being inhibited by the HS.

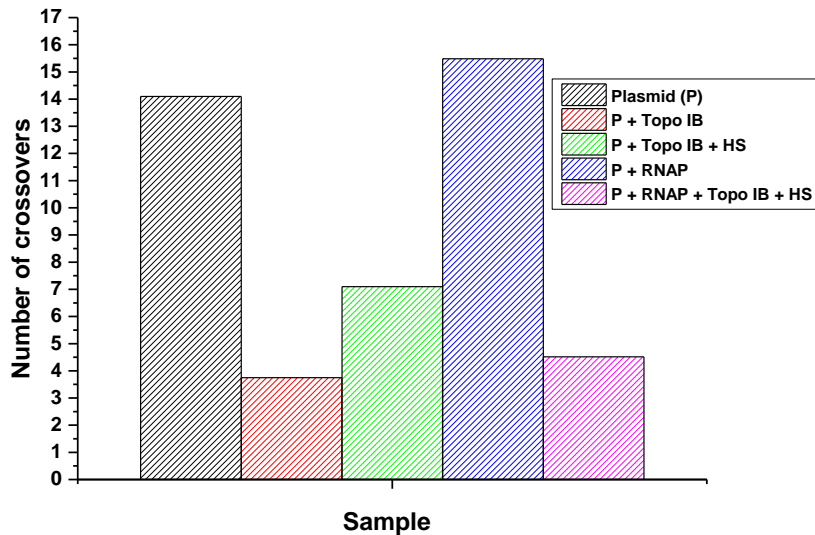
In order to test the effect of the presence of RNAP on the ability of Topo IB to relax supercoiled plasmid, samples of plasmid incubated with RNAP and HS in the absence or presence of Topo IB (Figure 7-14). In this case, one might expect the RNAP to form OPCs on the plasmid, but no NTPs are present to initiate transcription.



**Figure 7-14: AFM height images of supercoiled plasmid incubated with RNAP and HS then treated with Topo IB. The left panel shows compacted highly supercoiled plasmids in the absence of Topo IB while in the presence of Topo IB the plasmid appears in a more open conformation (Scale bar = 500 nm).**

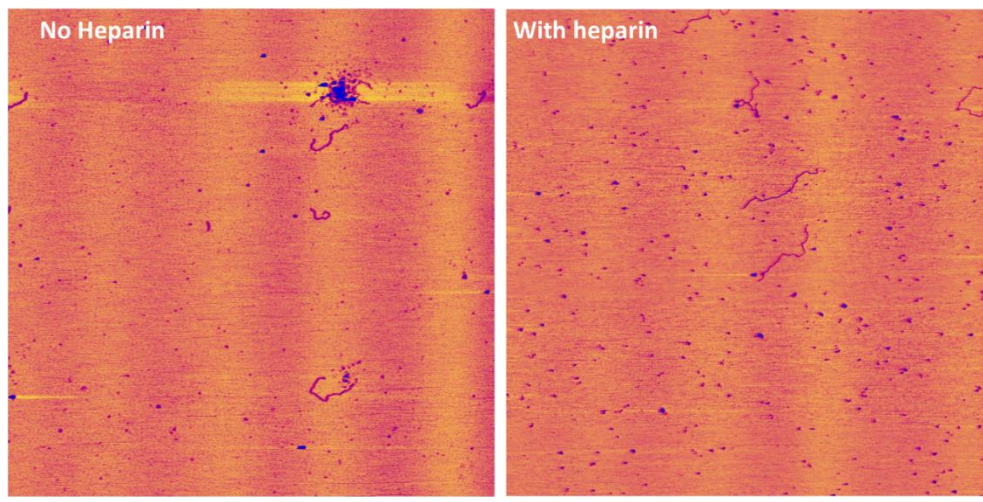
With RNAP present the number crossovers increases a small amount from the bare plasmid, to value of  $15.5 \pm 0.4$ . This increase is most likely due to the RNAP causing bending of the DNA and therefore leading to a higher number of crossovers occurring. With the addition of Topo IB and HS the number of crossovers decreases to  $4.5 \pm 0.4$ . The decrease in the number of crossovers is greater than that observed for plasmid, Topo IB and HS. This difference most likely is caused by some of the HS being sequestered due to binding to the RNAP, therefore decreasing the HS free to bind Topo IB. The difference in crossover number are show by the column plot shown in Figure 7-15.





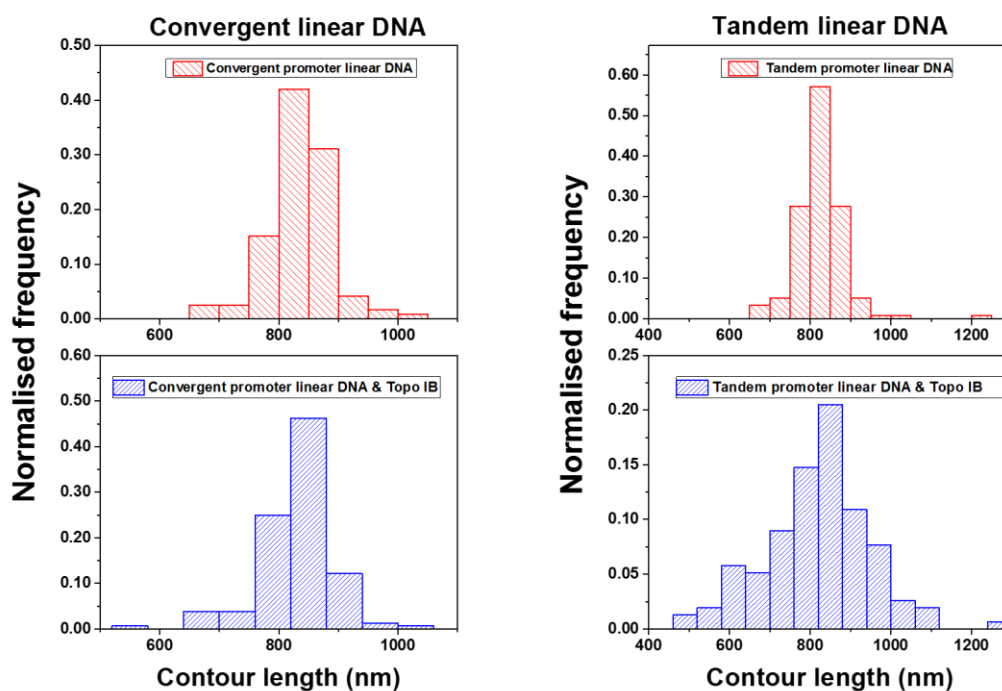
**Figure 7-15: Column plot of number of crossovers for different sample compositions. As it can be seen Topo IB shows activity in the buffer by decreasing crossover number. This is also true when HS is present and when RNAP is present in the sample. (n=120 for all samples).**

In order to test the effect of Topo IB on linear DNA, samples were imaged which contained the linear 2521 bp tandem and convergent promoter DNA and Topo IB with and without HS (Figure 7-16)



**Figure 7-16: AFM height images of Topo IB with linear tandem promoter 2521 bp DNA template with and without HS present. Occasional binding of Topo IB to linear DNA was seen but no change in contour length or DNA shape was recorded. (Scale bar = 1µm)**

It was observed that 30 % of DNA molecules imaged in the absence or presence of HS had Topo IB bound. In order to test whether the Topo IB affected the DNA upon binding, the contour length of the DNA was measured and compared to bare DNA template (Figure 7-17).

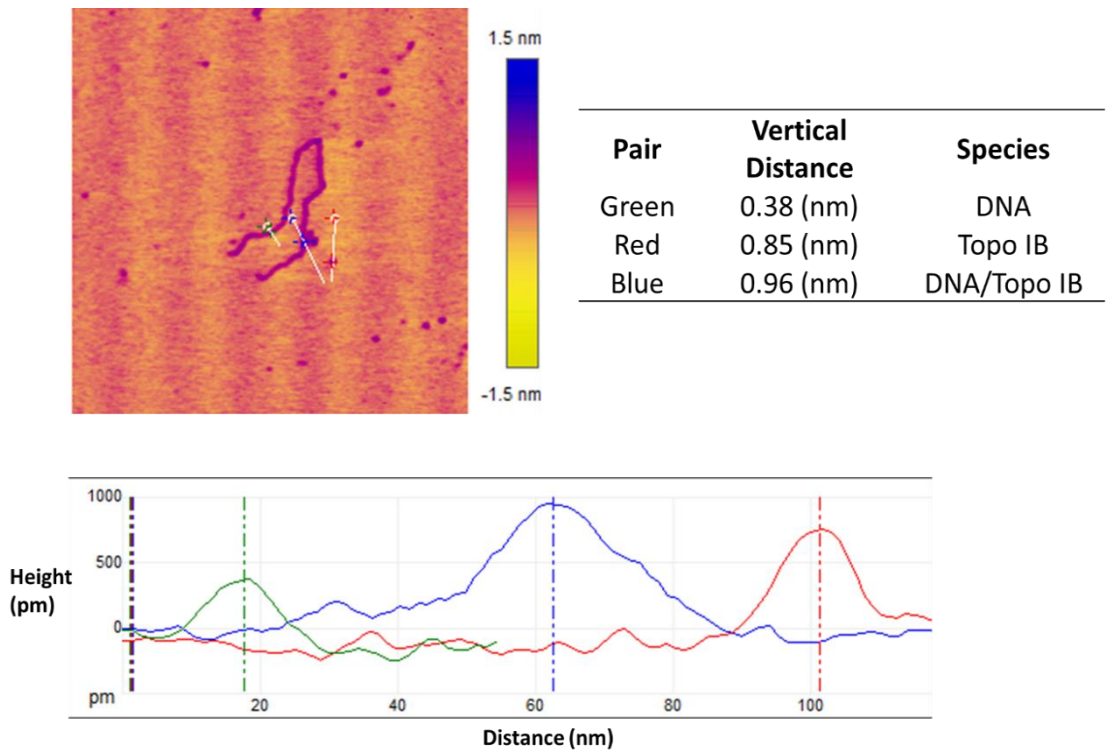


**Figure 7-17: Histogram plots of convergent and tandem linear 2521 bp DNA template with 10 U of Topo IB. For both DNA templates the presence of Topo IB does not alter the contour length indicating that the Topo IB does not wrap the DNA (n= 119 and 146 for convergent without and with Topo IB and n=153 and n=128 for tandem without and with Topo IB).**

The contour length of the convergent promoter 2521 bp template was  $834.5 \pm 4.7$  nm which gives a base pair rise of 0.33 nm which is within the usual experimental range. When bound by Topo IB a small change in the average to  $832.3 \pm 5.3$  nm was seen. For the tandem promoter template, the DNA had a contour length of  $828.1 \pm 4.9$  nm and when bound by Topo IB a contour length of  $821.0 \pm 11.3$  nm. There is a small decrease seen in the contour length which can be considered insignificant but may be due to bends in the DNA backbone being obscured by binding of Topo IB. Examples of DNA

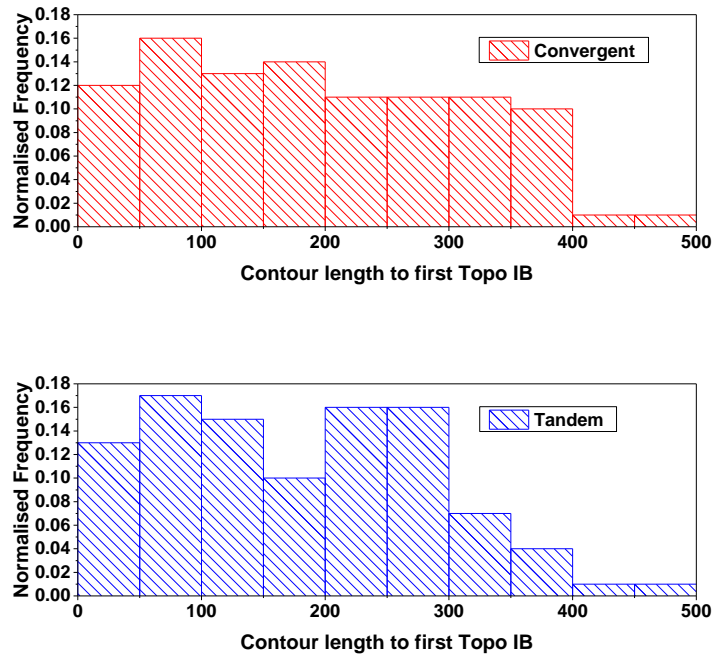
Chapter 7: Supercoiling and collisions

bound by Topo IB are shown in Figure 7-18, the cross section analysis shows the difference in height between, free DNA and Topo IB bound DNA.



**Figure 7-18: AFM height image with cross section plot shown below. The image is 918 nm by 918 nm. The table provides the difference in vertical height of the markers placed at the highest point of each cross section line. It can be seen that DNA bound by Topo IB shows a slightly increased height from Topo IB alone and over double the height of bare DNA.**

The low level of binding to linear DNA is not surprising as Topo IB does not have particular high specificity for topologically open DNA. The DNA bound Topo IB had a slightly greater height than that seen for free Topo IB and on average a height 3.35 times higher than that of the DNA. The binding of Topo IB did not show any preference for a location on the DNA and was considered to be occurring in a non-specific manner. In order to access the binding specificity the contour length to the first Topo IB bound to the DNA, from the shortest arm was measured. The histogram plot shown in Figure 7-19 shows a random distribution of contour lengths.



**Figure 7-19: Histogram plots of the contour length of the DNA to the first Topo IB molecule for both tandem and convergent promoter templates (n= 100 for both)**

The binding of the Topo IB to linear DNA had no apparent effect on the DNA appearance or length. This meant any changes seen during the *in vitro* transcription reactions to the contour length of the DNA and its shape could be assumed to be due to binding of RNAP or changes in the topological state of the DNA. The activity of Topo IB in the *in vitro* transcription buffer and reaction conditions was confirmed by its ability to relax supercoiled plasmid molecules. As this was confirmed, the effects of Topo IB on convergent and tandem transcription on a 2521 bp DNA template was then investigated.

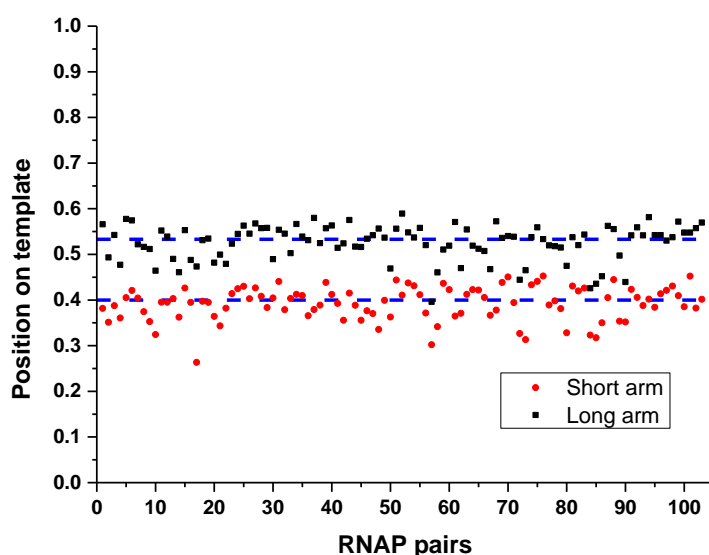
### **7.3.2 Outcomes of transcription from convergent promoters on a 2521 bp template with and without Topo IB**

In order to investigate transcription from convergent promoters the formation of OPCs was first studied. The arm lengths and inter-RNAP contour length was measured. As the DNA template did not possess an end label the arms were only discernible by length. The measurements are summarised in Table 7-2.

n	Short arm (nm)	Inter arm contour length(nm)	RNAP Long arm (nm)	Total length (nm)
103	$310.1 \pm 3.8$	$101.8 \pm 5.5$	$362.7 \pm 5.6$	$774.6 \pm 10.9$

**Table 7-2: Average contour length measurements for short and long arms and inter RNAP separation for OPCs formed on a DNA template of 2521 bp with convergent promoters.**

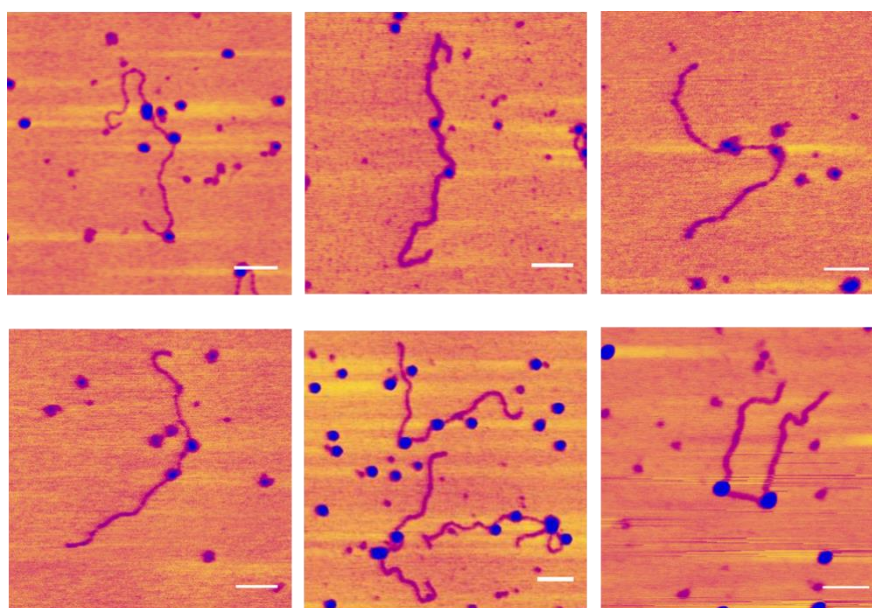
The template had a decrease in contour length of  $59.9 \pm 6$  nm. This reduction is expected due to the wrapping of the DNA upon formation of two OPCs [126, 130]. If the position of the two promoters is expressed as percentage of the template on bare DNA their position would be 0.45 of the total length for the long arm and 0.41 for the short arm from the ends of the template. The average measurements give a value of  $0.40 \pm 0.04$  and  $0.46 \pm 0.04$  indicating that the RNAPs are located at the promoters. This is visualised by the scatter plot shown in Figure 7-19.



**Figure 7-20: Scatter plot of RNAP positions plotted as a percentage of total contour length. The average position is shown by the blue dashed line. This is believed to correspond to the position of the promoters on the DNA.**



Examples of OPCs can be seen in Figure 7-21.



**Figure 7-21: AFM height images of OPCs formed on a 2521 bp template with convergent promoters. (Scale bar = 100 nm)**

Once the formation of OPCs was confirmed, elongation was initiated in the absence and presence of Topo IB. Images were collected and the position of the RNAPs was analysed. The average measurements are shown in Table 7-3.

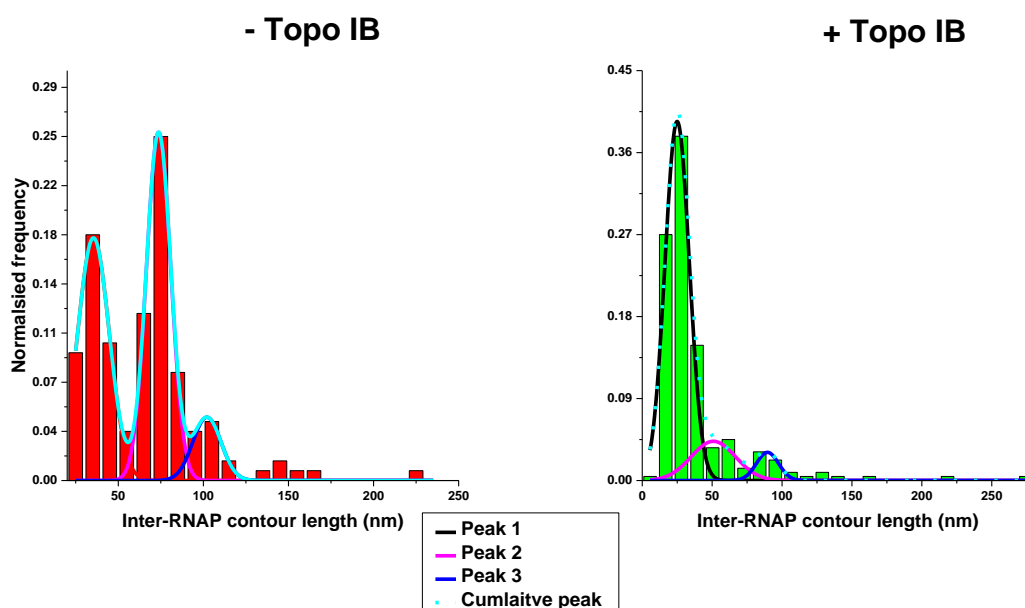
Convergent	n	Short arm (nm)	Inter-RNAP contour length (nm)	Long arm (nm)	Total (nm)
-Topo	139	354.7 ± 7.4	64.4 ± 2.7	384.1 ± 7.4	803.2 ± 7.0
+Topo	220	357.9 ± 4.4	37.7 ± 2.1	416.5 ± 3.8	812.1 ± 1.9

**Table 7-3: Average contour length measurements for DNA arms and inter-RNAP contour length after the addition of NTPs with and without the addition of Topo IB for elongation from convergent promoters.**

## Chapter 7: Supercoiling and collisions

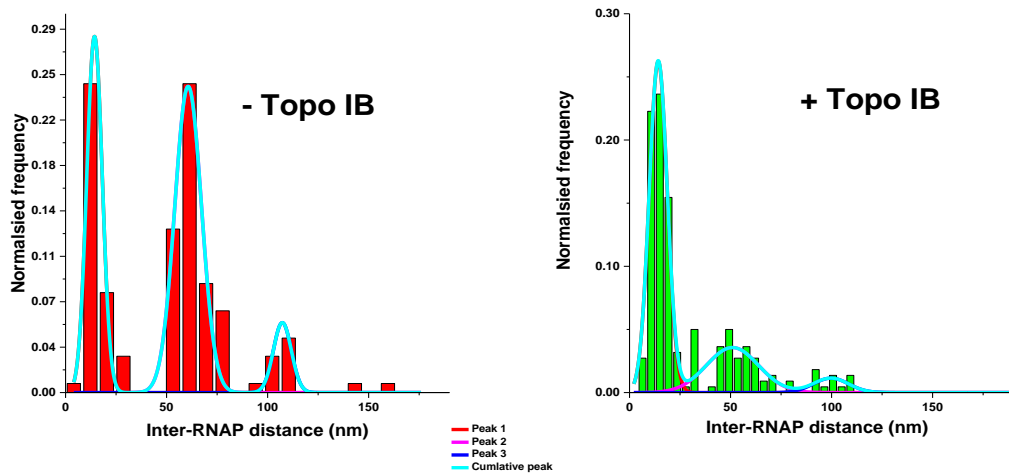
As it can be seen in both cases the inter-RNAP contour length has decreased as compared with the OPCs and the arm lengths have increased from that seen for OPC samples. This along with the increase in total contour length from OPC samples indicates that the wrapping of the DNA has decreased and elongation has commenced. The slightly shorter total contour length from that of the bare template ( $31.1 \pm 5.9$  nm and  $20.2 \pm 3.6$  nm, without and with Topo IB, respectively) also indicates that these complexes are actively bound to the template as two ECs are expected to reduce contour length by between 20-30 nm due to maintaining a small amount of wrapping during elongation [294]. The most striking difference between the two samples is the inter-RNAP contour length.

In samples without Topo IB it was found that the average inter-RNAP contour length was much larger than when Topo IB is present. Three distinct peaks were observed in the histogram (Figure 7-22).



**Figure 7-22: Histogram plots of the inter RNAP contour length after elongation from convergent promoters either without (red)  $n=139$  or with Topo IB (green)  $n=220$ . Plotted on the graph are three Gaussian distributions.**

These peaks are also seen when measuring the inter-RNAP distance as shown Figure 7-23.



**Figure 7-23: Histogram plots of inter-RNAP distance in the absence and presence of Topo IB after elongation from convergent promoters. The plots have been fitted with three peaks using a Gaussian function (n=139 for –Topo IB and 220 for +Topo IB).**

The peaks were fitted with Gaussian curves and the values of the fitted peak centres are shown in Table 7-4.

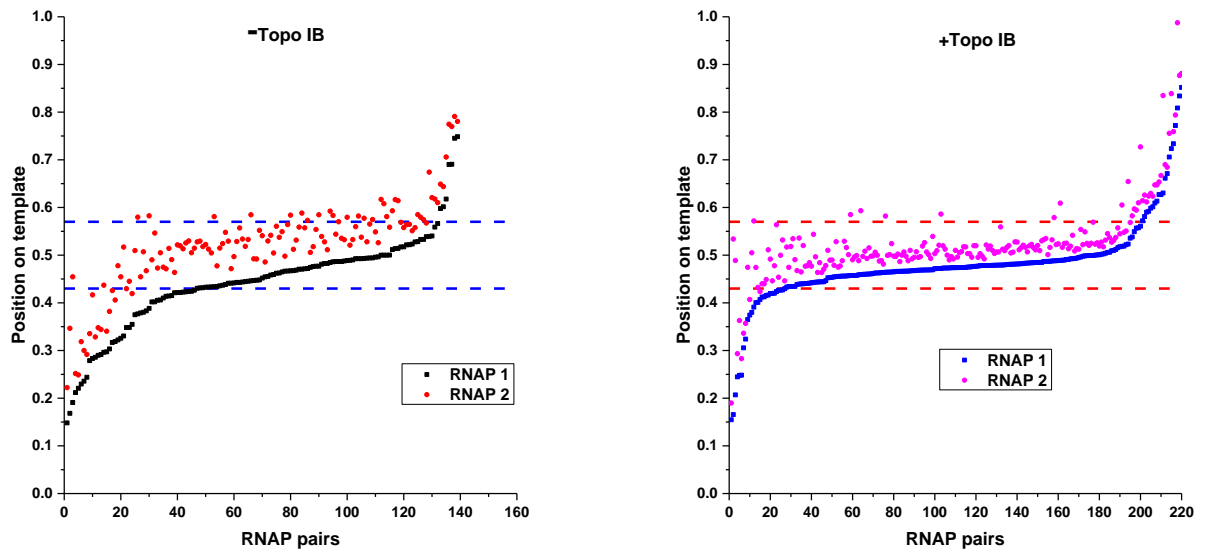
	Peak	Inter-RNAP contour length (nm)	Standard Error	Inter-RNAP distance (nm)	Standard Error
- Topo (n=139)	<b>1</b>	35.4	0.8	14.3	0.3
	<b>2</b>	73.7	0.6	60.7	0.6
	<b>3</b>	102.1	0.7	107.3	0.4
+ Topo (n=220)	<b>1</b>	24.9	0.3	14.2	0.3
	<b>2</b>	50.9	0.9	50.9	0.9
	<b>3</b>	89.6	0.7	99.7	0.6

**Table 7-4: Values of the centre of each peak from fitting the data with three Gaussian distributions. The standard error values are obtained by dividing the standard deviation of the plot by the square root of number of measurements.**

The peak with lowest  $X_c$  value, peak 1 in the absence of Topo appears to correspond with those RNAPs that have collided and stalled in hard contact indicated by the similar distance seen with the 1144 bp template. Peak 2 are RNAPs that have stalled at a greater distance.

Peak 2 is greatly reduced in the presence of Topo IB in comparison to peak 1. In the presence of Topo IB the height of peak 1 and number of RNAP pairs represented in this peak increases indicating that more RNAPs are able to collide and stall in hard contact.

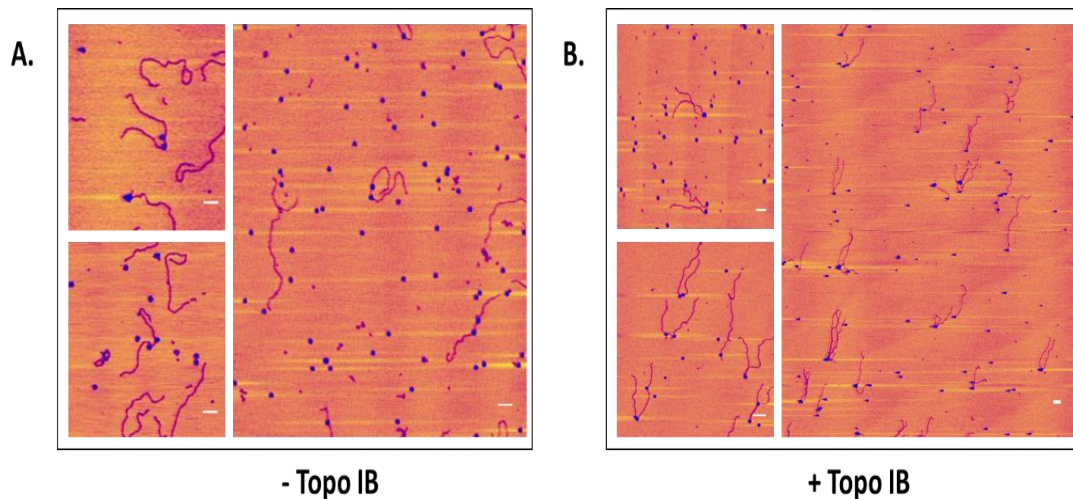
When RNAP pairs are plotted as a percentage of the template this difference between the two samples can be seen (Figure 7-24).



**Figure 7-24: Scatter plot of positions of RNAP pairs after elongation without Topo IB (left) and with Topo IB (right). It can be seen that without Topo IB RNAPs stall with a greater distance between them, whereas with Topo IB present the RNAPs stall in hard contact.**

As can be seen from the scatter plots shown in Figure 7-24, the RNAP pairs can be split into classes as with the 1144 bp template: Class 1 where both RNAPs are located between the promoters; Class 2 where one or both RNAPs are located upstream of a promoter; Class 3 where the RNAPs had a separation equal to or greater than that seen for OPCs. In the absence of Topo IB, Class 1 made up 45 % of complexes, 43 % of complexes were Class 2 and 12 % Class 3. In the presence of Topo IB, the distribution of classes was different with Class 1, 2 and 3 accounting for 62 %, 31 % and 7 %, respectively.

Example images of complexes from both samples are shown in Figure 7-25.



**Figure 7-25: AFM height images of complexes imaged after the addition of NTPs on a 2521 bp template. A) In the absence of Topo IB RNAPs were found to stall with a separation greater than would be expected if in hard contact. B) With the addition of Topo IB, RNAPs can be seen to stall at a distance which indicates that they have collided in hard contact. (Scale bars = 100 nm)**

For Class 1 complexes in the absence of Topo IB both RNAPs travelled on average  $30.4 \pm 2.7$  nm which is equivalent to approximately 92 bp. With Topo IB present the average distance travelled by both RNAPs was  $42.7 \pm 2.3$  nm or 129 bp. The inter-RNAP contour length of this class was  $49.9 \pm 2.7$  nm and  $29.1 \pm 1.1$  nm in the absence and presence of Topo IB, respectively. In the presence of Topo IB, a higher number of RNAPs stall in hard contact whereas without Topo IB 49 % of the RNAPs stall at a distance equivalent to approximately 178 bp apart. These complexes are taken to be the outcomes of EC-EC collision events as both RNAPs have escaped their respective promoters.

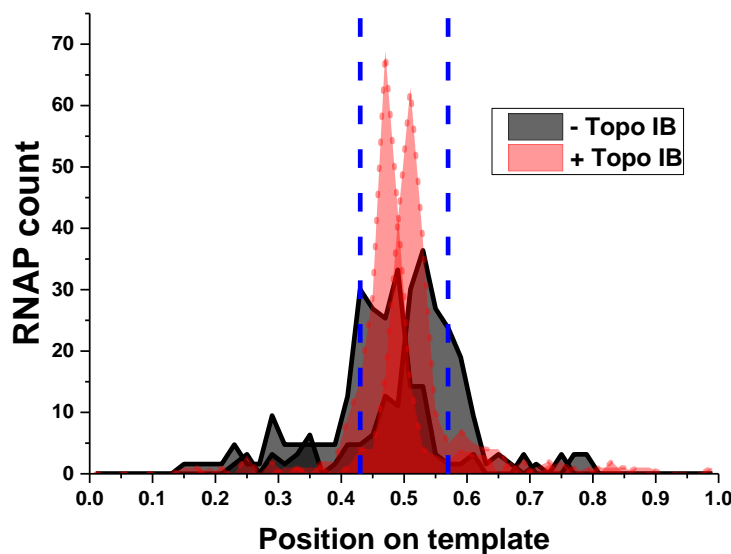
The data for Class 2 molecules suggest that the majority of RNAPs had moved upstream of the promoter by  $0.02 \pm 0.005$  and  $0.024 \pm 0.006$  of the total template length in the absence and presence of Topo IB. This equates to a backtracking of approximately 50 bp in both samples. Those that had both RNAPs upstream of a promoter had a separation of  $58.8 \pm 2.4$  nm in the absence of Topo IB and  $26.09 \pm 2.9$  in the presence of Topo IB. These complexes, as with the 1144 bp template, most likely account for EC-SD collisions which lead to shunting (backtracking) of one or both of the RNAPs upstream of the promoter. An increase in the occurrence of EC-SD collisions is noted

## Chapter 7: Supercoiling and collisions

in the absence of Topo IB but the distances backtracked remains approximately the same, while the average separation between the two RNAPs is reduced with the addition of Topo IB.

Class 3 complexes were much more common than seen previously when using the 1144 bp template, in the absence of Topo IB, accounting for 13 % of complexes analysed. These complexes include those where both RNAPs have failed to escape their promoters as well as any non-specific binding events. This increase could be due to higher levels of non-specific binding with a longer template, however, this is unlikely as, with Topo IB present, Class 3 complexes only accounted for 7 %.

The distribution of Class 1 and 2 complexes can be visualised by plotting the count of individual RNAPs at each position on the template (Figure 7-26).

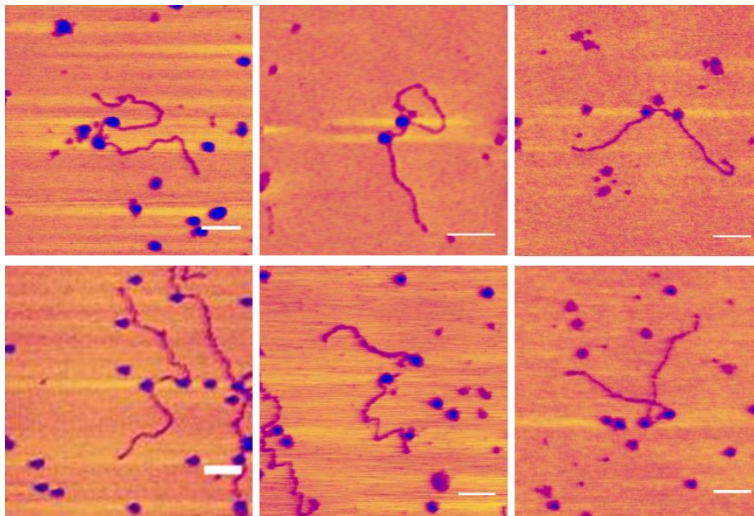


**Figure 7-26: Distribution of individual RNAPs along the template. Each count represents a RNAP at a given position. Those elongated in the absence of Topo IB are shown in black. The count of RNAPs at each point is overlaid with each other from each arm. Those elongated in the presence of Topo IB are shown in red.**

The higher number of RNAPs located at the promoters can be seen in the absence of Topo IB as well as the greater number of RNAPs located between the promoters with the addition of Topo IB.

### 7.3.3 Outcomes of tandem transcription of a 2521 bp template with and without Topo IB

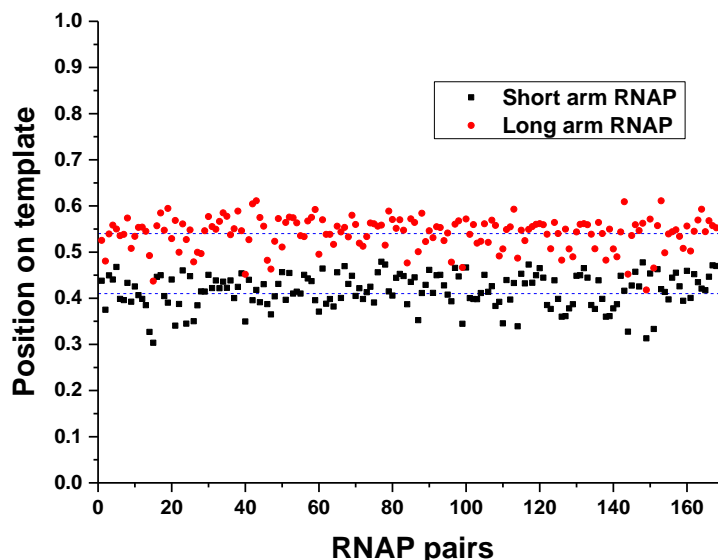
Having investigated the outcomes of convergent transcription, and the effects of Topo IB, on the 2521 bp template, the outcomes of transcription from tandemly arranged promoters were studied. Firstly OPCs were formed on the 2521 bp template and imaged (Figure 7-27).



**Figure 7-27: AFM height images of OPCs formed on tandem promoter 2521 bp template DNA (Scale bars = 100nm).**

The average RNAP position as a percentage of the template for each RNAP was  $0.41 \pm 0.04$  and  $0.46 \pm 0.04$  which can be seen by the scatter plot shown in Figure 7-28.





**Figure 7-28: Scatter plot of RNAP pair positions plotted as a percentage of the template for double OPCs formed on tandem promoter 2521 bp template DNA. The average position is shown by the blue dashed line.**

The arm lengths and inter-RNAP contour length are shown in Table 7-5. The contour length measurements and positioning of the RNAPs indicate these are OPCs.

n	Short arm (nm)	Inter-RNAP contour length (nm)	Long arm (nm)	Total (nm)
168	314.9 ± 3.5	94.4 ± 2.0	348.6 ± 4.4	757.9 ± 7.5

**Table 7-5: Average contour length measurements for double OPCs formed on tandem promoter 2521 bp template.**

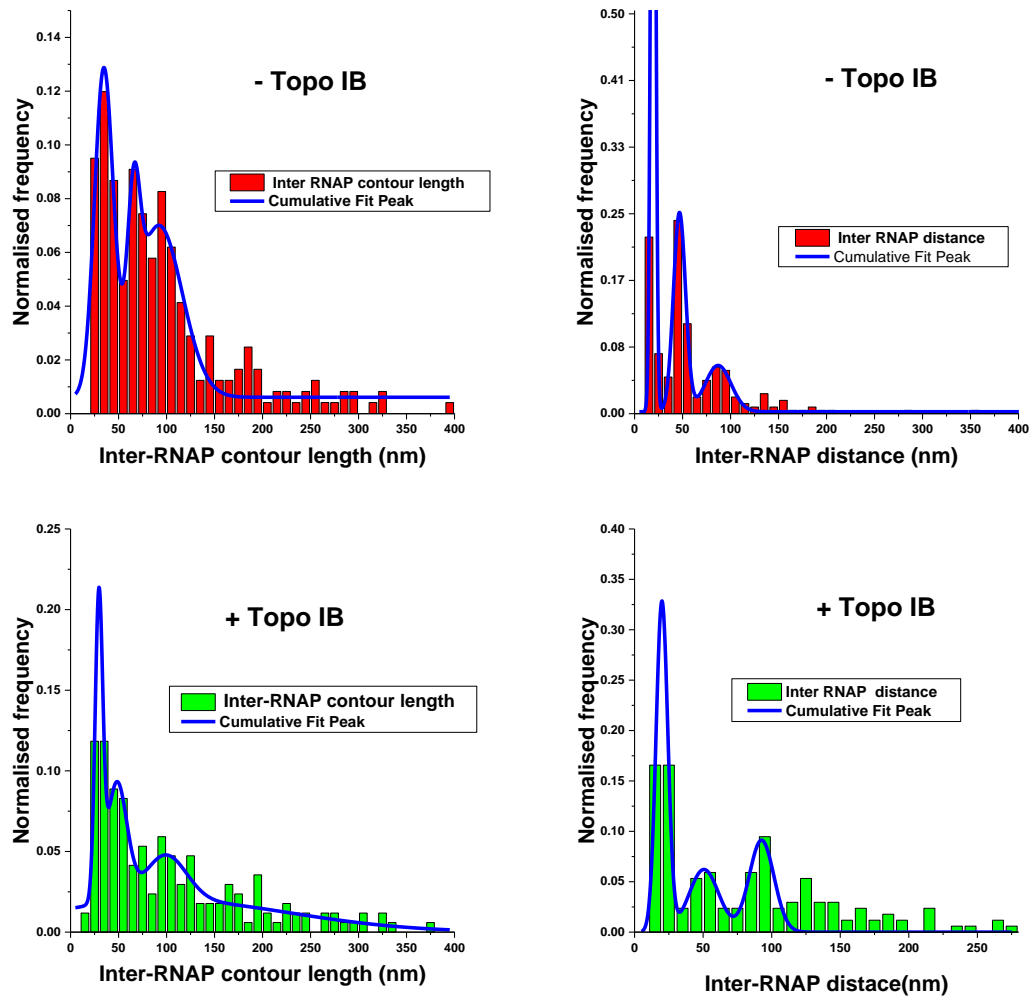
The template shows a contour length decrease of  $70 \pm 7$  nm which is confirmation that OPCs have been formed. After formation of OPCs, elongation was initiated with and without Topo IB. The average contour length measurements are shown in Table 7-6.

Tandem Promoter template	n	Short arm (nm)	Inter-RNAP contour length (nm)	Long arm (nm)	Inter-RNAP distance (nm)	Total (nm)
-Topo	242	256.9 ± 4.2	96.5 ± 4.5	429.9 ± 3.9	58.9 ± 3.5	783.4 ± 1.4
+Topo	169	197.9 ± 7.0	103.9 ± 6.2	483.6 ± 6.9	79.4 ± 4.9	784.8 ± 2.0

**Table 7-6: Average contour length measurements of complexes elongated from tandem promoters without and with Topo IB.**

After the addition of NTPs, the short arm can be seen to decrease while the long arm increases in contour length. This along with the increase in the total contour length from OPC samples indicates that elongation has occurred. It can be seen that upon addition of Topo IB the average distance moved by both RNAPs increases but there is no significant change in the contour length separation between RNAPs but a slightly greater inter-RNAP distance is measured in the presence of Topo IB. In the absence of Topo IB the leading RNAP travelled  $64.2 \pm 4.2$  nm ( $194 \pm 12$  bp) and the trailing RNAP travelled  $77.35 \pm 3.8$  nm ( $234 \pm 11$  bp). When Topo IB was present the leading RNAP travelled on average  $124.5 \pm 7.3$  nm ( $377 \pm 21$  bp) while the trailing RNAP travelled  $130.4 \pm 6.7$  nm ( $395 \pm 19$  bp).

Histogram plots of the inter-RNAP contour length and distance provided a more detailed view of the distribution of measurements (Figure 7-29).



**Figure 7-29: Histogram plots of inter RNAP contour length and inter-RNAP distance for RNAPs elongating from tandem promoters either with (n=169) or without (n=242) Topo IB. The histograms have been fitted with a Gaussian fitting function indicating three main peaks.**

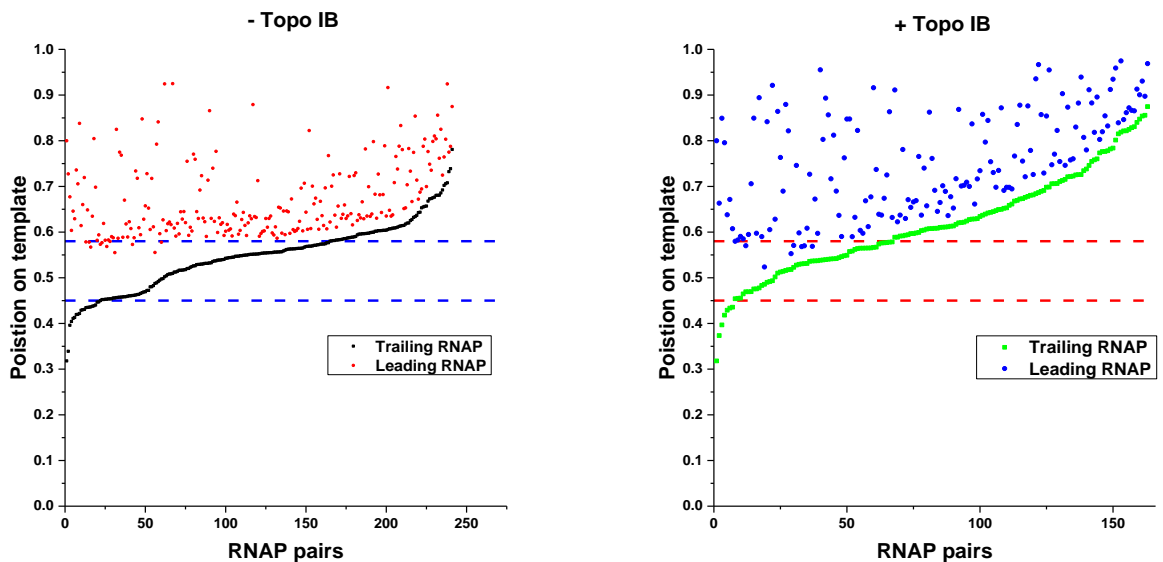
Inter RNAP contour length and inter-RNAP distances were fitted with three peaks using a Gaussian function, the locations of the centre of each peak are shown in Table 7-7.

	Peak	Inter-RNAP contour length (nm)		Inter-RNAP distance (nm)	
		Xc	Standard Error	Xc	Standard Error
<b>- Topo</b>	<b>1</b>	34.6	0.6	19.1	0.2
	<b>2</b>	65.9	0.4	46.7	0.4
	<b>3</b>	91.9	1.6	87.2	0.9
<b>+ Topo</b>	<b>1</b>	29.5	0.3	19.9	0.33
	<b>2</b>	48.2	0.8	50.6	0.83
	<b>3</b>	99.2	9.9	92.6	0.69

**Table 7-7: Values of the centre of peaks fitted to histogram plots with the standard error calculated from the standard deviation of the fitted curves.**

The peak at the lowest value most likely corresponds to those RNAPs that have come into hard contact. The second peak corresponds to RNAP pairs that have stalled with a greater separation. The third peak is at a value equivalent to the separation between RNAPs seen in OPCs, indicating that this corresponds to RNAPs that have failed to initiate elongation. In the absence of Topo IB, the first peak accounts for approximately the same number of complexes as those in the second peak. With the addition of Topo IB the second peak decreases while the first peak's height increases relative to the second peak indicating that in the presence of Topo IB a higher proportion of RNAPs come into hard contact.

The Gaussian fitted peaks indicate that after the addition of Topo IB a higher proportion of stalled complexes are closer to each other. When the positions of the RNAPs is plotted as a scatter plot details of the distribution of RNAP pairs over the template can be seen (Figure 7-30).

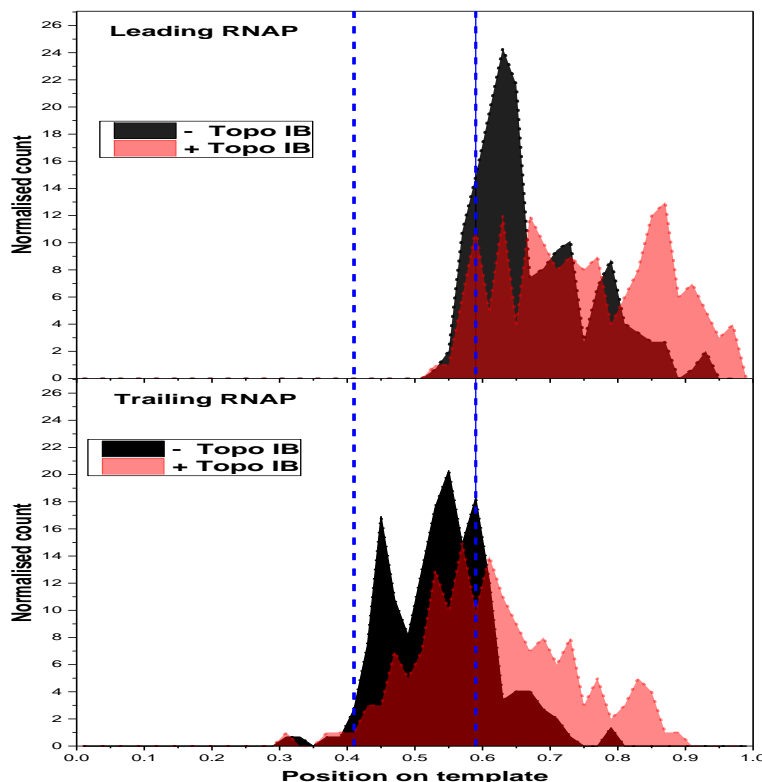


**Figure 7-30: Scatter plots of RNAP positions plotted as a percentage of the template. Without Topo IB (left) it can be seen that a large proportion of RNAP pairs are located at the leading promoter. With the addition of Topo IB the RNAP pairs are mainly located downstream of the leading promoter.**

As with the 1144 bp template the complexes can be broken down into classes. Class 1 consists of those with both RNAPs downstream of both promoters. This class accounted for 20 % of complexes analysed in the absence of Topo IB but in the presence of Topo IB this was the most common class making up 54 % of complexes analysed. This class most likely occurs when both RNAPs have initiated elongation and transcribed away from the promoters before stalling. Some of these molecules could also be the outcomes of EC-SD collision where an elongating trailing RNAP has led to shunting of a slow to escape or promoter bound leading RNAP. The second class of complexes were those that had the trailing RNAP located between the promoters and the leading RNAP located downstream of both promoters. This was the most common class in the absence of Topo IB (71 %), however, this class of complex only accounted for 36 % in the presence of Topo IB. These most likely represent a combination of pairs where the leading RNAP has escaped its promoter while the trailing molecule has failed to initiate, and are represented by those RNAP pairs on the left of the scatter plots shown in Figure 7-30. Also included in this class of complexes are those where the leading RNAP may have failed to initiate transcription. These will have undergone an

Chapter 7: Supercoiling and collisions

EC-SD collision, but the trailing RNAP has not travelled downstream of the leading promoter as well as some that may have both RNAPs still resident at their respective promoters. The final class of complexes are those with both RNAPs at or within the inter promoter region. Some of these may be a result of non-specific binding events but are most likely to be complexes where both RNAPs have failed to escape their promoters and remain as OPCs. Additionally a few may represent a small amount of EC-SD collisions. These accounted for 9 % and 10 % of complexes in the absence and presence of Topo IB, respectively. When RNAP position is plotted against count of individual RNAPs at each position on the template the differences between the distribution of each RNAP on the template can be seen between samples without and with Topo IB (Figure 7-31).



**Figure 7-31: Distribution of individual RNAPs located along the template as a percentage of the total contour length. It can be seen that the leading RNAP (top panel) travels further before stalling with Topo IB present. This can also be seen for the trailing RNAP (bottom panel). Without Topo IB is shown in black and with Topo IB is shown in red. Dark red areas are overlaps of the two data sets.**

It can be seen in the case of the leading RNAP that addition of Topo IB leads to a more even spread of RNAPs over the short arm, with peaks at, and just downstream of the promoter. These peaks in the region of the promoter are seen in both plots, but are much larger when the Topo IB is not added. The trailing RNAP in both sample sets has a peak close to the trailing promoter, most likely this represents those RNAPs that have failed to escape. There are also two peaks seen in both samples, one slightly upstream of the leading promoter and one slightly downstream, or at the leading promoter. This is most likely due to pausing at the promoter as was noted by Palmer *et al.* as well as the occurrence of EC-SD collisions leading to stalling [422].

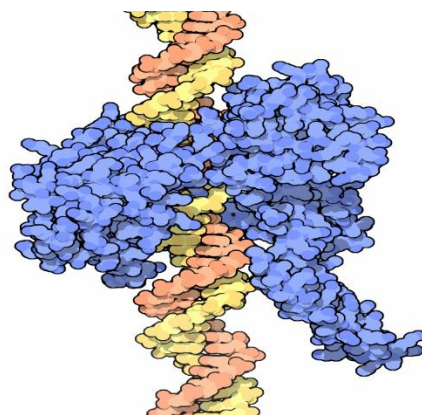
## 7.4 Discussion

### 7.4.1 Topoisomerase IB

The data shows that Topo IB can be readily adsorbed to a mica surface as was seen in previous studies by Arganam *et al.* and Liu *et al.* [454]. The diameter measurements show that there are two peaks at approximately 17 and 22 nm. This is similar to results obtained by Liu *et al.* for human Topo IB imaged in air in which had a width of  $21.5 \pm 3.4$  nm and length of  $19.17 \pm 2.5$  nm was measured [459]. This may be due to the protein's asymmetric dimensions. The crystal structures of the homologous human Topo IB gives approximate dimensions of  $70 \text{ \AA} \times 60 \text{ \AA} \times 60 \text{ \AA}$  and Argaman *et al.* give an expected diameter of 7.6 nm for Calf thymus Topo IB [448, 454]. Broadening due to the AFM tip will also increase the observed diameter [305]. It is also suggested that upon adsorption to the mica surface it is possible that the protein's structure is altered due to the high surface free energy of the mica leading to greater adsorption of hydrophilic amino acids on the protein surface which may in turn alter the conformation, or structure of the protein [459, 462]. This reasoning is also given as an explanation for the reduced height of the protein seen by Liu *et al.* who recorded an average height of  $2.6 \pm 0.3$  nm [459]. The reaction buffer and experimental conditions used in this work differ from that of Liu *et al.* and may account for the discrepancy in average height which was measured as  $1.15 \pm 0.2$  nm. This latter measurement is in agreement with that seen in liquid by Subramani *et al.* who reported a height of  $1.4 \pm 0.3$  nm [463].

The Topo IB activity in the buffer used in this study was confirmed by changes in the topological appearance of plasmids following the addition of Topo IB. The activity of the Topo IB was seen to be affected by the presence of HS in the reaction, however, relaxation of supercoils was still observed. This may be due to the HS binding to the active site of the Topo IB as might occur with RNAP due to the polyanionic nature of HS and its similarity to DNA. The inhibition of Topo IB by HS has been reported previously, with 300  $\mu\text{g/ml}$  giving a 50 % inhibition of human Topo IB [464, 465], and therefore some loss of activity might be expected. The higher activity of the Topo IB seen in the presence of RNAP is most likely due to RNAP competing for HS and therefore HS being bound by RNAP meaning that the effective concentration of HS is decreased.

The binding of Topo IB in the presence of linear DNA template occurs at a low level as expected and does not seem to show any specificity as is shown by the position of Topo IB on the DNA template. Vaccinia Topo IB has been shown to bind to linear, and open circular DNA, at high DNA:Topo IB ratios, forming loops in the DNA [455]. Moreno-Herrero *et al.* noted that the contour length of the DNA was not altered by the binding of Vaccinia Topo IB and Argaman *et al.* when investigating calf thymus Topo IB did not note any change of DNA contour length with Topo IB binding [454, 455]. This can be expected as changes in the DNA contour length are believed to only occur when wrapping of the DNA around the protein occurs, for example with RNAP [126, 294]. As can be seen from Figure 7-32, Topo IB binds in a clamp like manner and does not require wrapping.

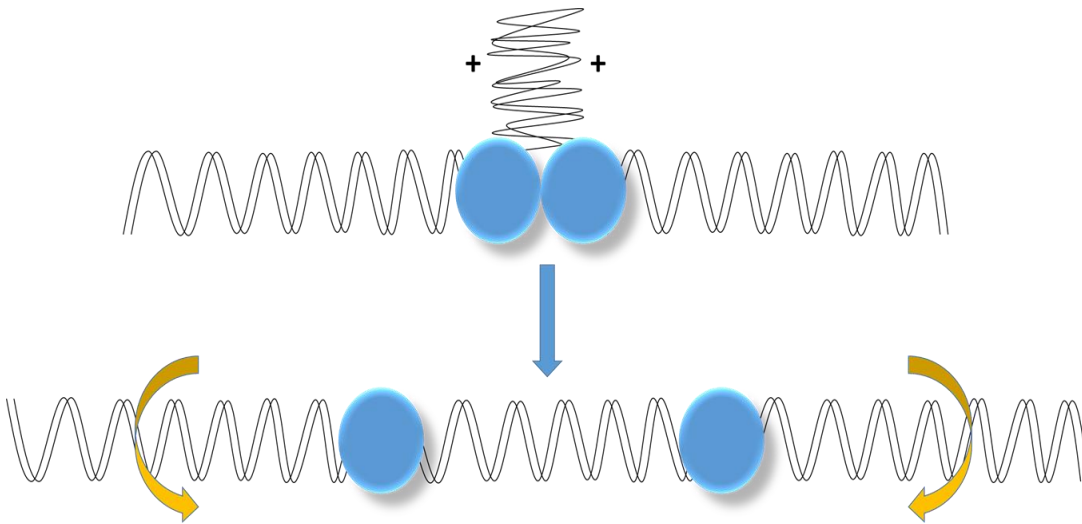


**Figure 7-32: Cartoon image drawn from the crystal structure of human Topo IB bound to double stranded DNA [448].**



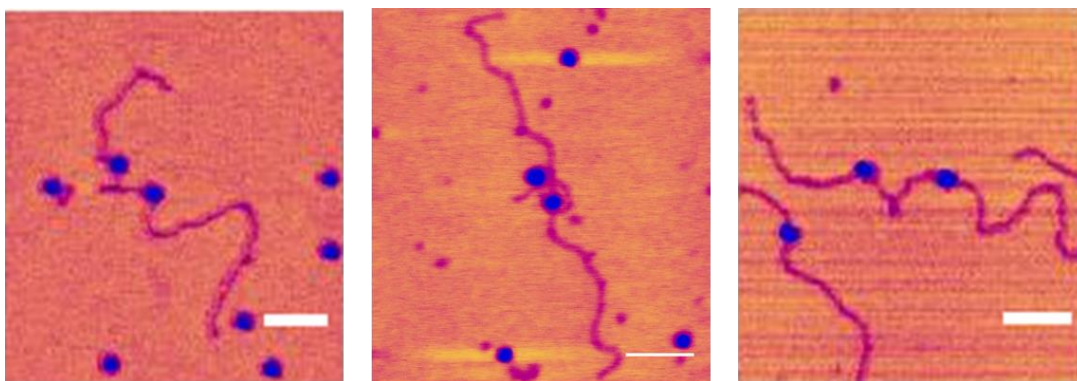
### 7.4.2 Convergent transcription and the effects of Topo IB

The outcomes of convergent transcription on the 1144 bp and 2521 bp templates differed. The most notable difference is in the distribution of inter-RNAP contour length. The difference between the average separation of RNAPs on the 2521 bp and 1144 bp templates was  $28.5 \pm 2.3$  nm. This value decreased upon the addition of Topo IB and is interpreted as a difference in levels of supercoiling occurring in these templates with two RNAPs bound. Kouzine *et al.* were able to show that it was possible to have an isolated domain of negative supercoiling between two divergent promoters on linear template of similar size [435]. The occurrence of supercoiling is expected to be short lived but its dissipation would be hindered by not just natural bends in the DNA but also by the presence of larger bends induced by the two ECs as was noted by Stupina *et al.* when investigating the merging of oppositely supercoiled domains *in vivo* [294, 443]. The introduction of bends as well as the increased hydrodynamic radius and molecular weight when two RNAPs are bound would also encourage supercoiling. This is not unexpected when considering that Nelson's hybrid-motion theory estimates that a length of DNA over 400 nm would be spin locked [432]. The template used here has a contour length of over double this length and so supercoiling may occur. Upon the addition of Topo IB, the separation decreases which further supports the role that supercoiling is playing in convergent transcription. Direct evidence for the stalling of RNAPs due to the introduction of both positive and negative supercoiling has been reported by Ma *et al.* and Yin *et al.* [317, 427]. In both experiments it was noted that 36 % of RNAPs studied by Yin *et al.* and 50% of RNAPs studied by Ma *et al.* were able to regain activity if supercoiling was removed in a timely fashion [317, 427]. The presence of a peak in the inter-RNAP contour length in the absence of Topo IB, which accounts for 43 % of complexes analysed, supports this because if the supercoiling is dissipated over time by the global rotation of the complexes, some RNAPs should regain activity and come into hard contact. The increased inter-RNAP contour length may arise not just from the force exerted by supercoiling of the template but if plectonemic structures occur then this may bring the proteins into contact as detailed in Figure 7-33. After the supercoiling structures have dissipated the RNAPs may rest a greater distance apart.



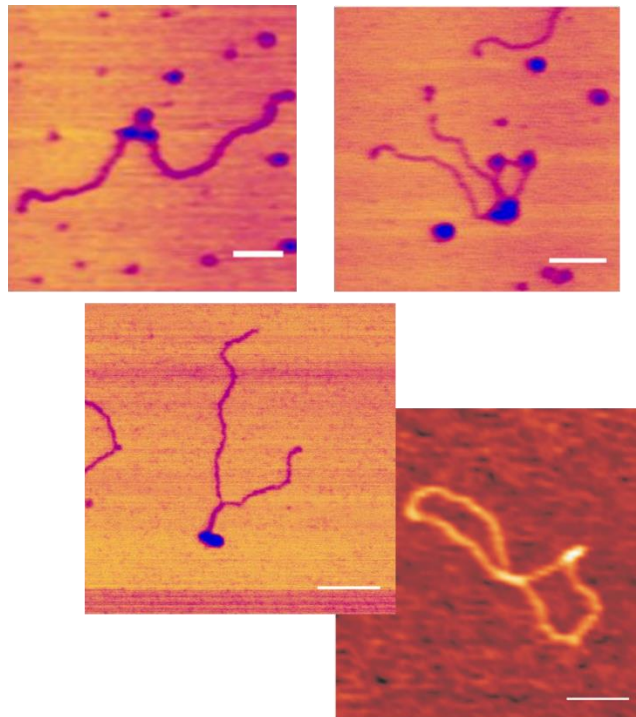
**Figure 7-33: Schematic diagram showing the possible role that plectonemic structures may play in the greater, inter RNAP separation seen between the two RNAPs when Topo IB is not present to relieve supercoiling. The positive supercoiling ahead of each RNAP (blue) leads to a hyper-positive supercoiled region between the two RNAPs. This may form a plectonemic structure which may bring the RNAPs into contact leading to stalling. The dissipation of the supercoiling due to the rotation of the template (yellow arrows) leads to the RNAPs being stalled at a greater distance apart than when no supercoiling occurs.**

On rare occasions non-linear DNA structures were noted between two RNAPs. As these structures were random in nature, in-depth analysis is not possible but examples are shown in Figure 7-34.



**Figure 7-34: AFM height images of stalled elongation complexes from convergent promoters with non-linear structures seen between the RNAPs. These structures appear different to the backbone DNA (Scale bars = 100 nm).**

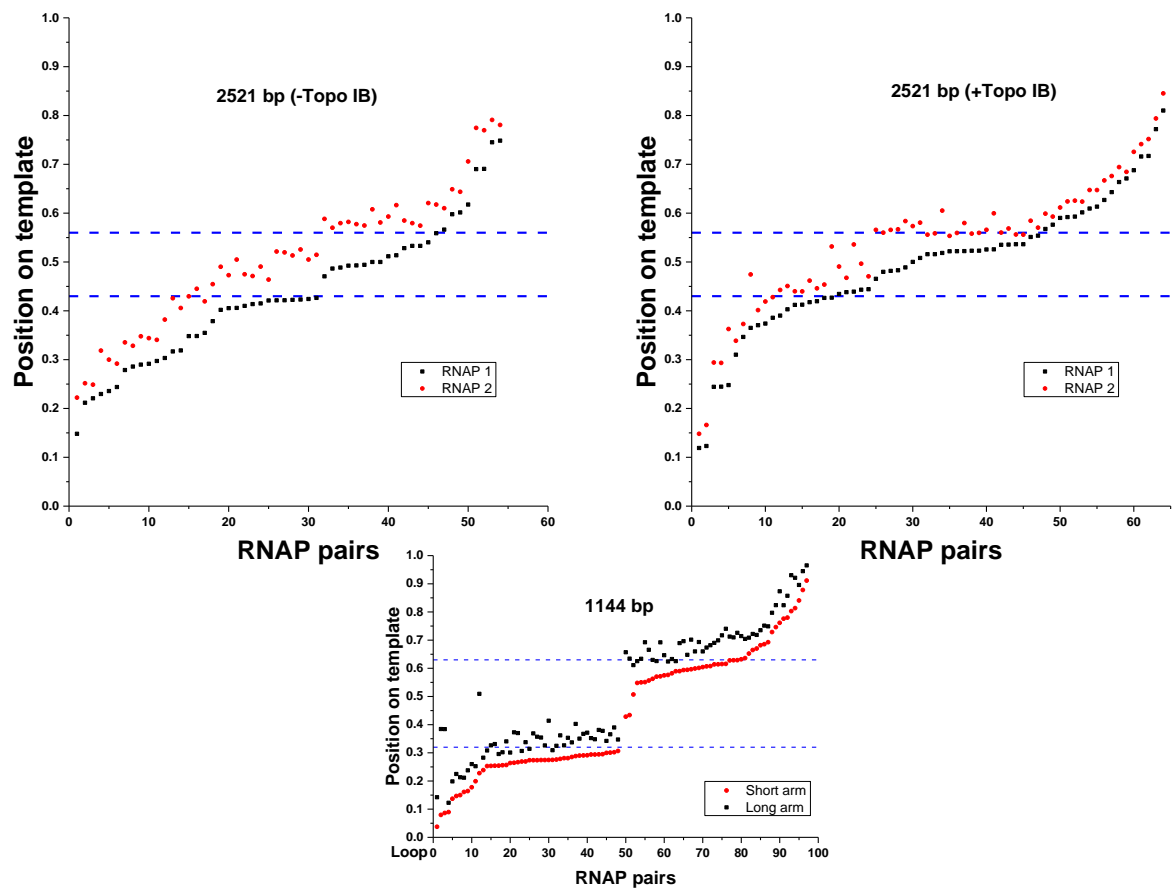
Upon addition of Topo IB these structures were not seen but a number of structures not previously recorded were noted. These structures appeared similar to the synapses formed by Vaccinia Topo IB (Figure 7-35) as well as the occurrence of structures that appeared to be looped DNA bound by multiple proteins.



**Figure 7-35: AFM height images of looped structures seen in samples of elongated complexes from convergent promoters in the presence of Topo IB (top image). Below an image is shown of a structure that appears similar to synapses formed by viral Topo IB by Moreno-Herrero *et al.* [455]. (Scale bars = 100 nm).**

Again these structures were not analysed due to their disorganized nature but it is possible that these are outcomes of Topo IB remaining bound to the template. The involvement of supercoiling between convergent promoters has been suggested when studying RNAP II by Rubio *et al.* [442]. Using a plasmid based system it was found that when Topo I, Topo II or both topoisomerases were inhibited a reduction in full length transcripts was recorded. This indicated that supercoiling can play a role in transcriptional interference during elongation stage [442].

The second change noted with the 2521 bp template is the distribution of the different classes analysed. There is an increased number of Class 2 complexes which are believed to be EC-SD collisions while there is a decrease in Class 1 complexes (EC-EC collisions). In the presence of Topo IB, the distribution of classes returns to values similar to that observed with the 1144 bp convergent template. RNAP scrunches DNA during initiation before formation of an ITC, it has been suggested that this scrunching of the DNA may be negatively affected by the presence of positive supercoils in the template as this would resist the unwinding of the DNA required for this process [60, 132, 133]. In the case of convergent promoters, if the initiation of elongation occurs asynchronously, positive supercoils downstream of an EC may have an inhibitory effect on the promoter escape of a RNAP still in the abortive initiation stage. This would lead to a higher number of EC-SD collisions occurring. The removal of supercoiling by Topo IB would lead to a distribution similar to that observed when using the 1144 bp which is the case. This result indicates that supercoiling derived from transcription is able to influence the promoter escape mechanism and may influence the formation of a productive ITC. The separation observed between the RNAP pairs that have undergone EC-SD collisions is greater on the 2521 bp template in the absence of Topo IB and decreases with the addition of Topo IB. This indicates that supercoiling may also play a role in the shunting or movement upstream of an inactive complex. It is also possible that supercoiling may lead to the stalling of the EC and disrupt the wrapping of an SD. This could cause the final resting positions of RNAPs to be further apart. The scatter plot has a similar appearance to that observed for the 1144 bp template with those RNAPs located near to a promoter having a greater separation than those that have been shunted upstream of the promoters. This is highlighted by the scatter plot shown in Figure 7-36 which shows only those RNAP pairs that are Class 2 molecules on the 2521 bp template with and without Topo IB and on the 1144 bp template.



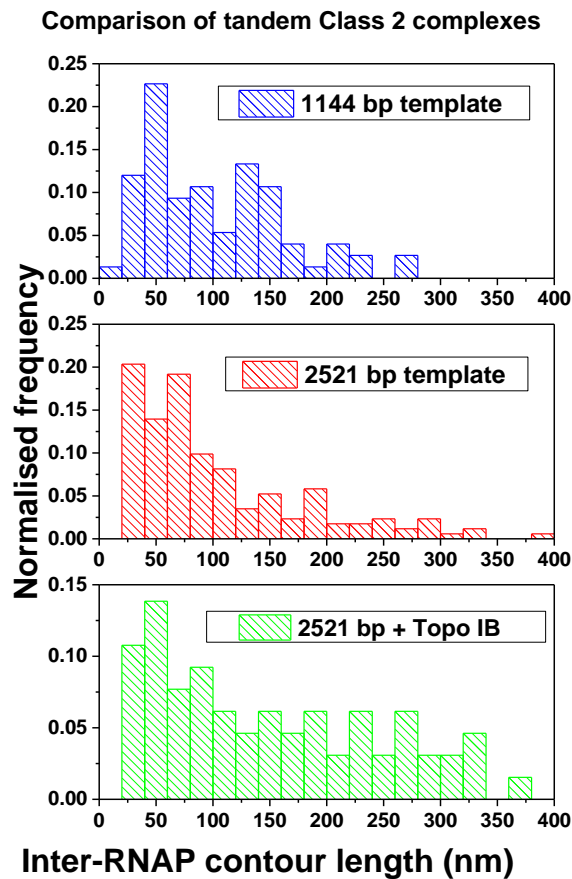
**Figure 7-36: Scatter plots of Class 2 molecules from samples using the 2521 bp template without Topo IB (top left) and with Topo IB (top right). Also shown is the scatter plot for Class 2 molecules taken from samples using the labelled 1144 bp convergent promoter template.**

The similarity observed between those RNAPs believed to have undergone EC-SD collisions may be due to these collisions occurring different mechanism that do not involve supercoiling. It is only the distance between the RNAPs after a collision that differs.

### **7.4.3 Concurrent transcription from tandem promoters with and without Topo IB**

In the absence of Topo IB a distinct difference in the classes as identified can be seen with an increased template length. With a template of 1144 bp the majority of complexes analysed have both RNAPs located outside of the inter promoter region, (Class 1), indicating that collisions are occurring between two RNAPs that have successfully escaped their promoters and stalled downstream of both promoters (59 %). With an increased template length this class of molecules drops to 20 % of complexes seen. Upon the introduction of Topo IB into the reaction mixture the percentage of Class 1 complexes is found to be at a level similar to that seen on the 1144 bp template, 54 %. This change in the number of Class 1 complexes is accompanied by a change in Class 2 complexes.

In the absence of Topo IB, Class 2 complexes represent the majority of those analysed (71%). These complexes are believed to be those that represent an EC-SD collisions between an active EC originating from the trailing promoter and an OPC or inactive RNAP located at the leading promoter. The fact that the addition of Topo IB leads to a decrease in this class indicates that supercoiling may be a factor in the escape of the leading RNAP from its promoter or may influence an ITC from the leading promoter. The separation seen between the RNAPs in these classes appears to be similar between the 2521 bp with and without Topo IB and those seen on the 1144 bp template (Figure 7-37)



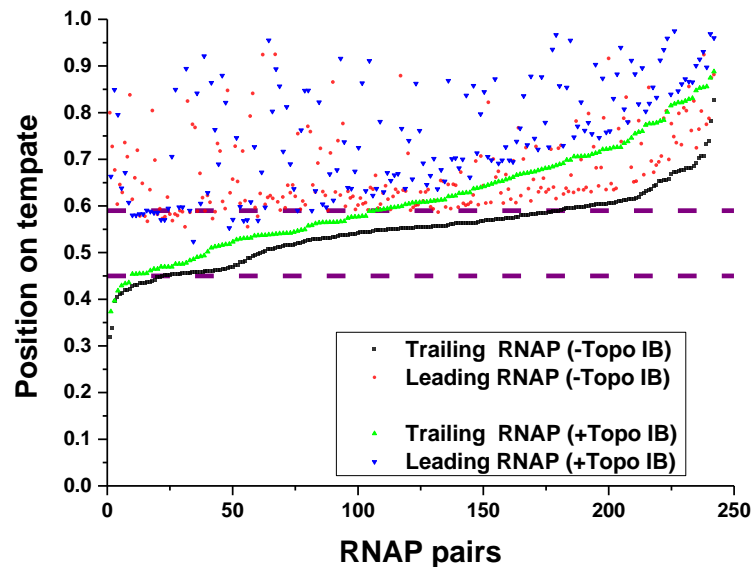
**Figure 7-37: Histograms showing the inter RNAP contour length recorded for the 1144 bp template (blue n= 75) and the 2521 bp template with (green n= 65), and without Topo IB (red, n= 172). It can be seen that the separation seen between RNAPs is similar for all three data sets.**

This similarity in the separation may be due to collisions occurring through a similar mechanism. If the leading complex is slow to escape its promoter due to it being an unproductive ITC or moribund complex an EC-SD collision may occur. If this is the case then negative supercoiling behind the leading RNAP would not be able to obliterate the positive supercoiling ahead of the trailing RNAP. This positive supercoiling may have an influence on the ability of the leading RNAP to escape its promoter, giving rise to an increase in the number of EC-SD collisions that occur. The separation seen between the two RNAPs may be a consequence of the loss of wrapping of an OPC after a collision or may be due to some shunting of the leading RNAP by the trailing RNAP or both. The stability of OPCs and the formation of moribund complexes appears to be modulated by the presence of supercoiling as shown by the reduction in the number of these

complexes upon addition of Topo IB. The influence of positive supercoiling upstream of an OPC has not been investigated previously as many studies utilise tandem promoter templates with differing promoter strengths in experiments involving multiple rounds of transcription. Such studies have been able to show that not only does transcriptional interference occur through promoter occlusion but the generation of positive supercoils from a strong promoter upstream leads to suppression of the downstream promoter due to the reduced rate of CPC to OPC transition [6, 422, 466, 467]. This system *in vivo* may provide a mechanism to avoid EC-SD collisions as is seen in this *in vitro* system. Even though no previous evidence exists for what the effects of upstream transcription might have on promoter escape, it has been suggested that promoter escape and formation of productive ITCs may be influenced by supercoiling. Both Hsu and Susa suggest that supercoiling may be an important factor in the transition from abortive to active transcription [468, 469]. It was found by Vo *et al* that blockage of promoter escape could occur at the nucleotide addition stage or at the elongation step. [415]. It was determined through promoter element mutation studies that the ability to enter productive elongation may involve interactions between the upstream promoter elements, the sigma factor and RNAP molecule. The generation of positive supercoils may lead to re-arrangement of the DNA held in the active site. Sen *et al.* commented that moribund complexes stuck in the abortive cycle had a reduced number of open base pairs at the -10 position [417]. Positive supercoiling may be able to stabilise this structure leading to more OPCs forming moribund complexes. It may not be the supercoiling itself that alters the transition from abortive initiation to productive elongation but rather changes in the geometry of the DNA. Bends in the DNA may lead to an overly stable interaction between the  $\sigma^{70}$ RNAP and the promoter, therefore preventing escape from the promoter. The repression of promoter clearance due to bends introduced into the DNA was noted by McAllister *et al.* for the *Bal129* gene of the phage *SP82* [470]. The effects of bending on promoter activity is varied but repression at each stage of the initiation process has been observed [471]. If the bending induced by the formation of an upstream OPC as well as the bends introduced by an upstream EC inhibit or slow down the clearance of the downstream promoter then an increase in EC-SD collisions would be expected. As collisions lead to transcriptional stalling this could lead to traffic jams occurring on genes possessing tandem promoters, which could have a detrimental effect on a cells expression profile.



The addition of Topo IB in the system allows for a larger number of RNAP pairs to elongate in conjunction with each other. Stalling still occurs as with the 1144 bp template but the distribution of RNAPs over the template highlight the change in the data seen upon addition of Topo IB (Figure 7-38).



**Figure 7-38: Scatter plot of position of RNAPs on the template for samples with and without Topo IB. It can be seen that with Topo IB present a much greater number of RNAP pairs are located downstream of both promoters in comparison to when Topo IB is absent.**

If Topo IB is able to relieve torsional stress that may occur between an EC and OPC or SD and this torsional stress, or changes in topology induced by this torsional stress, does inhibit promoter escape, then it would be expected that a higher number of RNAP pairs would be able to escape both promoters when Topo IB is present. The data shows evidence that supercoiling does not play a role in the stalling of two ECs as the separation seen between RNAPs is similar to that of the 1144 bp template. This may be expected; because if both RNAPs are actively transcribing then ablation of supercoiling would be expected with tandem promoters. In the presence of Topo IB it was also noted that the amount of bare DNA increased by 20% with the addition of Topo IB, indicating that Topo IB may play a role in allowing synchronized transcription to occur.

## 7.5 Conclusions

With an increased template length the patterns of stalling of RNAPs from both convergent and tandem promoters changes. In the case of transcription from convergent promoters a number of RNAP pairs stall at a distance that indicates that they are not in hard contact. The decrease in the number of complexes with this greater separation observed upon the addition of Topo IB indicates that supercoiling between the two RNAPs is involved in their stalling. A change in the types of collisions observed also indicates that supercoiling may be an important factor influencing the process of transcription initiation. The outcome of collisions remains the same in that both RNAPs remain stably bound to the template. It is also seen from the data that collisions between an EC and an SD still lead to a roadblock to transcription and that the shunting of the SD occurs to the same extent. The change observed in the number of EC-SD collisions in this system may indicate that *in vivo* dynamic supercoiling may play a role in preventing EC-EC collisions occurring as often therefore providing a higher chance of a complete transcript to be synthesized.

Studies by Kouzine *et al.* observed supercoiling on DNA templates of similar size when studying multiple rounds of transcription[472]. This study only investigated single rounds of transcription. This study provides evidence that even with a single round of transcription, supercoiling is transient, but can still play a role in collisions between two ECs, as well as influencing initiation from convergent promoters [432]. In regard to the spin locking hypothesis of Nelson the data indicates that it is possible for supercoiling to occur on a linear template. It is noted though that this supercoiling dissipates over time. This could be considered to be due to the lacking of a cranking mechanism when transcription has stalled, therefore allowing supercoiling to dissipate. Thus allowing some RNAPs to continue elongation while others remain stalled at a distance greater than hard contact. The influence of supercoiling on interactions between RNAPs provides a new insight into the roles that supercoiling may have in the initiation of convergent genes with supercoiling providing a feedback system to help prevent collisions occurring in the first place.

### *Chapter 7: Supercoiling and collisions*

With transcription from tandem promoters the role that supercoiling plays is not as clear cut. The apparent increase in the number of collisions occurring at and around the leading promoter again indicate that dynamic and transient supercoiling may play a role in regulating simultaneous transcription and be important in avoiding transcriptional arrest of highly transcribed genes. The stalling of two active ECs is unexpected as pausing and backtracking have been shown to be reduced through collisions between tandemly transcribing RNAPs. This stalling process may have implications on the phasing of the two elongating RNAPs or may be due to other factors that are not apparent in this study. The increase in the number of stalled RNAP pairs at the leading promoter and the change in this population with the addition of Topo IB provide evidence that the transition of an OPC into an active elongation complex is hindered by supercoiling directly or via the global topology of the DNA complex.

This page is left intentionally blank

## 8 Final conclusions and future work

### 8.1 Final conclusions

The study of concurrent transcription by AFM has been presented in this work. Through the development of new protocols to increase accuracy and precision of data it has been possible to investigate the occurrence of TI that occurs for both tandem and convergent promoter arrangements. It has also been possible to investigate the role that supercoiling may have in collisions.

#### 8.1.1 Improved imaging of DNA protein complexes

AFM provides a powerful technique that allows the study of single molecules or complexes of DNA and protein, which may be overlooked when using bulk biochemical techniques.

Even so AFM still has a number of shortcomings, some of which have been addressed in this work. The addition of ssDNA loop to the end of the DNA was shown by Billingsley *et al* to be distinguishable from the DNA backbone and RNAP [5]. Thus showing the ability to use end bound ssDNA loops as labels of polarity for the DNA. The method devised by Billingsley *et al* had low yield of labelled DNA. This method was improved by the incorporation of ssDNA loop label into a PCR reaction. This method was shown to produce high yields of labelled DNA due to the exponential increase in DNA template by the process of PCR. It was observed that the PCR reaction had specific requirements so that a labeled template could be produced.

These included the use of a 5'-3' exonuclease deficient DNA polymerase so that the primer incorporating the loop was not degraded. It was also observed that the temperature of the extension stage of the PCR was important as the structure of the loop was require to be maintained throughout the reaction. Meaning that the extension temperature had to be decreased to ensure that the loop was integrated into the backbone of the DNA.

### *Chapter 8: Final conclusions*

This method provided a highly efficient one pot reaction that allowed for production of a high yield of labeled DNA. This therefore allowed for faster collection of data and more detailed information to be obtained from the data due to the fact that the loop provides a marker of the polarity of the DNA template.

The loop structure was shown to have some effects on transcription. This was observed at the stages of promoter location and elongation. The loop structure was observed to prevent dissociation of the RNAP during the promoter searching process. Once elongation had occurred, those RNAPs that transcribed into the loop structure were observed to be bound to the loop structure. The cause of the binding the loop structure during promoter search was not determined but may be due to the single stranded nature of the loop. The stalling of elongation was also not investigated in detail as this was not seen as a disadvantageous effect.

The binding of the RNAP to the loop highlighted another issue known to AFM when studying DNA protein interactions. This is the occurrence of non-specific binding. In many studies the inhibition of non-specific binding has not been possible. Bulk biochemical methods often utilised polyanionic molecules that are able to competitively bind to the DNA binding proteins. In the case of RNAP heparin is often used, as it is able to bind non-specifically bound RNAPs but not those that have formed OPC [401].

The use of heparin in AFM samples has been seen as not viable due to the highly negative charge of heparin being able to alter or prevent the binding of DNA to the mica surface [127]. Investigations presented in Chapter 6 show that both the polyanionic molecules heparin and heparin sulphate (HS) do alter the binding of DNA to the surface. This is not the case though when included in samples that contain proteins. The presence of proteins in the sample prevents the heparin and HS from forming over the mica surface.

It was found though that when investigating transcription from two promoters, heparin did not allow for the regular formation of two OPCs. This effect may be due to the instability of OPCs due to wrapping of the DNA upon formation of an OPC. It was found

though that the less negative structurally similar HS did not prevent the formation of two OPCs and so was utilised in investigations of concurrent transcription. The incorporation of HS providing more certainty in data and also improving the speed at which data can be collected, by reducing the number of complexes that need to be discounted due to large numbers of RNAPs being bound.

### **8.1.2 Transcriptional interference and the effects of transcription-coupled supercoiling.**

With inclusions of HS in samples it was observed, for convergent promoters, that the majority of collisions occurred between the two promoters (EC-EC collisions). It was also observed that the majority of these collisions resulted in the RNAPs stalling in hard contact. The stalled RNAPs remained bound to the DNA even in the presence of HS indicating that they would induce TI and therefore seriously interfere with gene expression.

The second most common class of collisions observed was thought to be a consequence of an EC colliding with a RNAP still bound to its promoter or in an inactive state. These collisions appeared to occur by two distinct mechanisms. This was reasoned due to the distance between the two RNAPs as well as the position on the template. Some collisions resulted in large scale movement of both RNAPs upstream of the promoter. With the RNAPs stalling in hard contact. Other collisions resulted in RNAPs being bound near to the promoter with the RNAPs being in close proximity rather than hard contact. These differences were thought to show that the stability of an RNAP on the DNA template differs at different stages of the initiation cycle.

In order to investigate whether the occurrence of these collisions was affected by the topology of the DNA template, experiments using larger templates in the presence or absence of Topo IB were performed. A change in the average distance observed between RNAPs that had undergone EC-EC collisions was seen after initiation of elongation on the 2552 bp templates, with no more than 50 % of collisions resulting in hard contact. A change in the populations of collision types was also recorded. The

## Chapter 8: Final conclusions

number of EC-SD collisions observed increased. In the presence of Topo IB the number of EC-EC collisions was similar to that seen on the 1145 bp template. The separation between the collided RNAPs indicated that a greater number came into hard contact. As the presence Topo IB altered the outcomes of collisions it was reasoned that supercoiling of the template played a role. The data indicates that supercoiling affects the transition of an OPC into an elongating complex. The increased EC-EC RNAP separation indicates that positive supercoiling may build up between two actively elongating RNAPs and cause them stall. This may occur without hard contact occurring or by a mechanism of supercoil driven backtracking.

When considered in the context of nested genes the results indicate that TI would lead to effects on gene expression, due to the inability of RNAPs to pass. It is most likely that *in vivo* mechanisms would exist prevent collisions and to clear stalled collided complexes. One mechanism by which convergent genes may be regulated is through the local supercoiling introduced by transcription itself. The increase in EC-SD collisions is an indication that supercoiling may provide at least one mechanism by which RNAPs may be able to communicate to each other in order to avoid collisions. An increase in positive supercoiling has been observed *in vivo*, to reduced formation of OPC [127]. It may be possible that the ability of these complexes to be shunted back by an EC would allow for a stronger promoter to efficiently drive transcription through the cumulative action of multiple ECs.

In the case of tandem promoter arrangements the discerning of EC-EC an EC-SD collision is more difficult. Data collected in the presence of heparin when two promoters are present indicate that a second RNAP leads to changes in initiation and elongation from a downstream promoter. Stalling of both RNAPs is recorded. It may be expected that an upstream RNAP may be able to act cooperatively for a downstream RNAP to overcome pauses and stalling, but the results indicate that this is not the case in this system. This could be due to the fact that both RNAPs are active and not in an intentionally stalled or backtracked state upon collision [201, 421]. The spacing of promoter may play a role with how RNAPs originating from separate promoters interact, similar to what was observed by Ponnambalam *et al.* [203]. Collisions show similar outcomes as convergent, where RNAPs are unable pass each other and that



## *Chapter 8: Final conclusions*

collisions result in RNAPs stalling in hard contact. With a longer template an increase in collisions around the downstream promoter was observed, which were reasoned to be EC-SD collisions, where the downstream RNAP had failed to initiate elongation. This increase in EC-SD collisions was not observed when Topo IB was present. This indicates that topology and transcription driven supercoiling are factors that are able to influence concurrent transcription.

The observation of changes with the introduction of Topo IB was reasoned as being an indirect indicator of DNA topology changes occurring due to transcription. The selection of a longer template was made in light of the spin locking theory suggested by Nelson [432]. The results obtained indicate that supercoiling is occurring on the linear template even though it is topologically open. Nelson theory suggest that the DNA is able to become spin locked when cranked at a certain speed and the DNA is over a certain length [432]. The results here agree with this in that supercoiling can occur on the 2552 bp template. When RNAPs stall there is loss of the cranking mechanism, allowing for supercoiling to dissipate. This dissipation is most likely the cause for the high number of EC-EC collisions observed with the 2552 bp template that resulted in hard contact stalling. Ma *et al* observed that when positive supercoiling is applied up or downstream of an RNAP stalling occurs, and for 50 % of stalled RNAPs, elongation resumed upon removal of supercoiling.

## **8.2 Future work**

The two methods presented in this body of work, high-throughput labelling and the incorporation of HS into AFM samples, are both novel. The use of these methods in other AFM studies of DNA protein interactions will help to provide detail in a number of studies as well as helping provide more precise and confident data. Further investigation into the binding of HS to the mica surface would allow for a better understanding of the interactions that occur between DNA, proteins and mica, allowing the further development of methods to incorporate HS and other similar molecules in AFM samples in a controlled and defined manner.

## Chapter 8: Final conclusions

In order to further investigate the occurrence of collisions between RNAPs originating from both tandem and convergent promoters a number of experiments are desirable to further confirm the importance of topology on these interactions and the outcomes of different classes of collisions. The use of linear templates with single stranded nicks in the DNA between promoters would provide a relevant control. The nick in the DNA would allow the dissipation of torsional stress and so it would be expected that the distance at which RNAPs stall would be similar to that seen when using the 1145 bp template, or when in the presence of Topo IB. For convergent promoter arrangement the use of topologically closed DNA templates would provide a more direct read of the topology of DNA. As elongation occurs the template would increase in its writhe and or twist, thereby adopting a different conformation when imaged by AFM. This conformation and the extent of any supercoiling could be assessed by counting the number of visible crossovers, or nodes as well as accessing the topology of the DNA using gel electrophoresis [460, 461]. This could be further extended to the use of template containing a cruciform structure. Cruciform structures have been observed to be responsive to topology, and by careful design can be made highly sensitive as observed by Yang *et al* and Shlyakhntenko *et al*. [473, 474]. By placing the cruciform outside the inter-promoter region it would be expected that negative supercoiling would cause cruciform intrusion, whereas when placed between the two promoters cruciform extrusion would be expected.

The data collected in this study could be further supported by the use of bulk biochemical techniques. The method of reconstructing elongation complexes, assured by Hobson *et al*. provides a combination of Dnase I mapping of the active sites of RNAPs, providing highly accurate positioning of RNAPs on a template [200, 318]. This method of purifying those EC that have undergone a collision, combined with radiolabelled sequencing of the transcripts would provide information on the activity of those RNAPs that have undergone EC-SD collisions. The use of FRET may also be viable in helping distinguish such collisions. FRET could be used to monitor the structure of the active site of the RNAP and determine what stage the of the initiation process an RNAP is at. The sequential formation of OPCs on the template using an inactive and active RNAPs would also allow in depth study of just EC-SD collisions.

### *Chapter 8: Final conclusions*

The use of combined optical /AFM techniques such as combined AFM and TIRFm would allow the locating of specific proteins on the DNA template. Distinction between RNAPs, or the distinction between RNAP and the Topo IB, would provide more detailed information on how these proteins may interact throughout transcription.

For tandem promoter arrangements further studies using templates possessing more promoters as well as different spaced promoters would provide detail on whether cooperativity can be observed as well as the importance of the phasing of the RNAPs has in the stalling process. The spacing of promoters in convergent templates would also provide more information on the effects of topology. It would be expected that an increased inter-promoter distance would allow for more supercoiling to occur, therefore causing RNAPs to stall at a greater distance. The size of templates could also be studied by increasing the length of the template from 1145 bp to greater than 2552 bp. This would allow for a value of  $L_c$  from the spin locking theory of Nelson to be determined [432]. The larger template may also prevent the dissipation of supercoils once transcription has stalled, therefore allowing supercoiled structures to be readily imaged by AFM.

One final set of future experiments would be the study of collisions *in situ*. This could be achieved using high speed AFM setups in a liquid environment. This would provide real-time observations of collisions as they occur, giving better insight into the process of shunting and backtracking in both tandem and convergent promoters. The use of DNA origami tiles as platforms for the attachment of templates as used by Endo *et al* or the use of DNA origami cassettes would provide a method to locate template DNA on the surface. The ability to fine tune the DNA and mica interactions as shown by Lee *et al*. would also enable better control over the system, as previous experiments found that the DNA may intermittently contact the mica surface, therefore effecting the elongation process. Even though *in situ* imaging provides a real-time visualisation of collision events, current methods are slow and not fully suited to studying complex reactions such as concurrent transcription and so may need further development before being truly viable.

This page is left intentionally blank

## 9 Bibliography

### Bibliography

1. Mellor, J., R. Woloszczuk, and F.S. Howe, *The Interleaved Genome*. Trends in Genetics, 2016. **32**(1): p. 57-71.
2. Shearwin, K.E., B.P. Callen, and J.B. Egan, *Transcriptional interference—a crash course*. TRENDS in Genetics, 2005. **21**(6): p. 339-345.
3. Crampton, N., et al., *Collision events between RNA polymerases in convergent transcription studied by atomic force microscopy*. Nucleic acids research, 2006. **34**(19): p. 5416-5425.
4. Billingsley, D.J., *Convergent transcription and nested gene models studied by AFM*. 2012, The University of Leeds.
5. Billingsley, D.J., et al., *Single-stranded loops as end-label polarity markers for double-stranded linear DNA templates in atomic force microscopy*. Nucleic acids research, 2012. **40**(13): p. e99-e99.
6. Baranello, L., et al., *The importance of being supercoiled: How DNA mechanics regulate dynamic processes*. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 2012. **1819**(7): p. 632-638.
7. Mendel, G., *Versuche über Pflanzenhybriden*. Verhandlungen des naturforschenden Vereines in Brunn 4: 3, 1866. **44**.
8. Bateson, W. and G. Mendel, *Mendel's Principles of Heredity: A Defence, with a Translation of Mendel's Original Papers on Hybridisation*. 1902: the University Press.
9. Strutevant, A., *The linear arrangement of six sex-linked factors in *Drosophila* as shown by their mode of association*. Molecular and General Genetics MGG, 1913. **10**(1): p. 293-294.
10. Morgan, T.H., *The theory of the gene*. American Naturalist, 1917: p. 513-544.
11. Miescher, F., *Die Histochemischen und Physiologischen Arbeiten von Friedrich Miescher—Aus dem Wissenschaftlichen Briefwechsel von F. Miescher*. 1869.
12. Dahm, R., *Discovering DNA: Friedrich Miescher and the early years of nucleic acid research*. Human genetics, 2008. **122**(6): p. 565-581.
13. Jones, M.E., *Albrecht Kossel, a biographical sketch*. The Yale journal of biology and medicine, 1953. **26**(1): p. 80.

14. Griffith, F., *The significance of pneumococcal types*. Journal of Hygiene, 1928. **27**(02): p. 113-159.
15. Avery, O.T., C.M. MacLeod, and M. McCarty, *Studies on the chemical nature of the substance inducing transformation of pneumococcal types induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III*. The Journal of experimental medicine, 1944. **79**(2): p. 137-158.
16. Hershey, A.D. and M. Chase, *Independent functions of viral protein and nucleic acid in growth of bacteriophage*. The Journal of general physiology, 1952. **36**(1): p. 39-56.
17. Watson, J.D. and F.H. Crick, *Molecular structure of nucleic acids*. Nature, 1953. **171**(4356): p. 737-738.
18. Chargaff, E., S. Zamenhof, and C. Green, *Human desoxypentose nucleic acid: Composition of human desoxypentose nucleic acid*. 1950.
19. Myers, R.M., et al., *Modification of the melting properties of duplex DNA by attachment of a GC-rich DNA sequence as determined by denaturing gradient gel electrophoresis*. Nucleic acids research, 1985. **13**(9): p. 3111-3129.
20. Privalov, P.L., et al., *What drives proteins into the major or minor grooves of DNA?* Journal of molecular biology, 2007. **365**(1): p. 1-9.
21. Saenger, W., W.N. Hunter, and O. Kennard, *DNA conformation is determined by economics in the hydration of phosphate groups*. 1986.
22. Neidle, S., *Oxford handbook of nucleic acid structure*. 1999: Oxford University Press on Demand.
23. Pohl, F.M. and T.M. Jovin, *Salt-induced co-operative conformational change of a synthetic DNA: equilibrium and kinetic studies with poly (dG-dC)*. Journal of molecular biology, 1972. **67**(3): p. 375-396.
24. Drew, H., et al., *High-salt d (CpGpCpG), a left-handed Z' DNA double helix*. 1980.
25. Wang, A., et al., *Molecular structure of a left-handed double helical DNA fragment at atomic resolution*. Nature, 1979. **282**(5740): p. 680-686.
26. Lodish, H., *Molecular cell biology*. 2008: Macmillan.
27. Bates, A.D. and A. Maxwell, *DNA topology*. 2005: Oxford university press.
28. Cech, T.R., *RNA as an enzyme*. Scientific American, 1986. **255**(5): p. 64-75.

29. Bacolla, A. and R.D. Wells, *Non-B DNA conformations, genomic rearrangements, and human disease*. Journal of Biological Chemistry, 2004. **279**(46): p. 47411-47414.
30. Wells, R.D., *Non-B DNA conformations, mutagenesis and disease*. Trends in Biochemical Sciences, 2007. **32**(6): p. 271-278.
31. Zhao, J., et al., *Non-B DNA structure-induced genetic instability and evolution*. Cellular and molecular life sciences, 2010. **67**(1): p. 43-62.
32. Nag, D.K. and T.D. Petes, *Seven-base-pair inverted repeats in DNA form stable hairpins in vivo in Saccharomyces cerevisiae*. Genetics, 1991. **129**(3): p. 669-673.
33. Schroth, G.P. and P.S. Ho, *Occurrence of potential cruciform and H-DNA forming sequences in genomic DNA*. Nucleic acids research, 1995. **23**(11): p. 1977-1983.
34. Wang, G. and K.M. Vasquez, *Non-B DNA structure-induced genetic instability*. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 2006. **598**(1-2): p. 103-119.
35. Htun, H. and J.E. Dahlberg, *Single strands, triple strands, and kinks in H-DNA*. Science, 1988. **241**(4874): p. 1791.
36. Jain, A., G. Wang, and K.M. Vasquez, *DNA triple helices: biological consequences and therapeutic potential*. Biochimie, 2008. **90**(8): p. 1117-1130.
37. Sinden, R.R., M.J. Pytlos-Sinden, and V.N. Potaman, *Slipped strand DNA structures*. Frontiers in bioscience: a journal and virtual library, 2006. **12**: p. 4788-4799.
38. Wang, G. and K.M. Vasquez, *Impact of alternative DNA structures on DNA damage, DNA repair, and genetic instability*. DNA repair, 2014. **19**: p. 143-151.
39. Wong, H.M., L. Payet, and J.L. Huppert, *Function and targeting of G-quadruplexes*. Curr Opin Mol Ther, 2009. **11**(2): p. 146-55.
40. Ha, S.C., et al., *Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases*. Nature, 2005. **437**(7062): p. 1183-1186.
41. Zhao, F., J. Xu, and S. Liu, *Atomic force microscopy visualization of the DNA network and molecular morphological transition on a mica surface*. Thin Solid Films, 2008. **516**(21): p. 7555-7559.
42. Batholomew, B. *DNA Topology*. 2007; Lecture on DNA topology]. Available from: <http://www.siumed.edu/~bbartholomew/lectures/Supercoiling%2007.pdf>.
43. Dorman, C.J., *DNA supercoiling and environmental regulation of gene expression in pathogenic bacteria*. Infection and immunity, 1991. **59**(3): p. 745.

44. Cozzarelli, N.R. and J.C. Wang, *DNA topology and its biological effects*. 1990: Cold Spring Harbor Laboratory Press.
45. Mirkin, S.M., *DNA topology: fundamentals*. eLS, 2001.
46. Travers, A. and G. Muskhelishvili, *A common topology for bacterial and eukaryotic transcription initiation?* EMBO reports, 2007. **8**(2): p. 147-151.
47. Hatfield, G.W. and C.J. Benham, *DNA topology-mediated control of global gene expression in Escherichia coli*. Annual review of genetics, 2002. **36**(1): p. 175-203.
48. Worcel, A., S. Strogatz, and D. Riley, *Structure of chromatin and the linking number of DNA*. Proceedings of the National Academy of Sciences, 1981. **78**(3): p. 1461-1465.
49. Zechiedrich, E.L., et al., *Roles of topoisomerases in maintaining steady-state DNA supercoiling in Escherichia coli*. Journal of Biological Chemistry, 2000. **275**(11): p. 8103-8113.
50. Richmond, T.J. and C.A. Davey, *The structure of DNA in the nucleosome core*. Nature, 2003. **423**(6936): p. 145-150.
51. Sinden, R.R. and D.E. Pettijohn, *Chromosomes in living Escherichia coli cells are segregated into domains of supercoiling*. Proceedings of the National Academy of Sciences, 1981. **78**(1): p. 224-228.
52. Postow, L., et al., *Topological domain structure of the Escherichia coli chromosome*. Genes & Development, 2004. **18**(14): p. 1766-1779.
53. Dixon, J.R., et al., *Topological domains in mammalian genomes identified by analysis of chromatin interactions*. Nature, 2012. **485**(7398): p. 376-380.
54. Nora, E.P., et al., *Spatial partitioning of the regulatory landscape of the X-inactivation centre*. Nature, 2012. **485**(7398): p. 381-385.
55. Gilbert, N. and J. Allan, *Supercoiling in dna and chromatin*. Current opinion in genetics & development, 2014. **25**: p. 15-21.
56. Deng, S., R.A. Stein, and N.P. Higgins, *Transcription-induced barriers to supercoil diffusion in the Salmonella typhimurium chromosome*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(10): p. 3398-3403.
57. Liu, L.F. and J.C. Wang, *Supercoiling of the DNA template during transcription*. Proceedings of the National Academy of Sciences, 1987. **84**(20): p. 7024-7027.
58. Koster, D.A., *Topoisomerase and the unwinding of stressed DNA*. 2007: TU Delft, Delft University of Technology.
59. Champoux, J.J., *DNA topoisomerases: structure, function, and mechanism*. Annual review of biochemistry, 2001. **70**(1): p. 369-413.



60. Roca, J., *Transcriptional inhibition by DNA torsional stress*. *Transcription*, 2011. **2**(2): p. 82-85.
61. Teves, S.S. and S. Henikoff, *Transcription-generated torsional stress destabilizes nucleosomes*. *Nature structural & molecular biology*, 2014. **21**(1): p. 88-94.
62. Sheinin, M.Y., et al., *Torque modulates nucleosome stability and facilitates H2A/H2B dimer loss*. *Nature communications*, 2013. **4**.
63. Joshi, R.S., B. Piña, and J. Roca, *Positional dependence of transcriptional inhibition by DNA torsional stress in yeast chromosomes*. *The EMBO journal*, 2010. **29**(4): p. 740-748.
64. Geertz, M., et al., *Structural coupling between RNA polymerase composition and DNA supercoiling in coordinating transcription: A global role for the omega subunit?* *MBio*, 2011. **2**(4).
65. Kouzine, F., et al., *Transcription-dependent dynamic supercoiling is a short-range genomic force*. *Nature structural & molecular biology*, 2013. **20**(3): p. 396-403.
66. Lavelle, C., *Pack, unpack, bend, twist, pull, push: the physical side of gene expression*. *Current opinion in genetics & development*, 2014. **25**: p. 74-84.
67. Mattick, J.S. and I.V. Makunin, *Non-coding RNA*. *Human molecular genetics*, 2006. **15**(suppl 1): p. R17-R29.
68. Crick, F., *Central dogma of molecular biology*. *Nature*, 1970. **227**(5258): p. 561-563.
69. de Lorenzo, V., *From the selfish gene to selfish metabolism: revisiting the central dogma*. *BioEssays*, 2014. **36**(3): p. 226-235.
70. Shapiro, J.A., *Revisiting the central dogma in the 21st century*. *Annals of the New York Academy of Sciences*, 2009. **1178**(1): p. 6-28.
71. Hurwitz, J., *The discovery of RNA polymerase*. *Journal of Biological Chemistry*, 2005. **280**(52): p. 42477-42485.
72. Ringel, R., et al., *Structure of human mitochondrial RNA polymerase*. *Nature*, 2011. **478**(7368): p. 269-273.
73. Cermakian, N., et al., *On the evolution of the single-subunit RNA polymerases*. *Journal of molecular evolution*, 1997. **45**(6): p. 671-681.
74. Cramer, P., et al., *Structure of eukaryotic RNA polymerases*. *Annu. Rev. Biophys.*, 2008. **37**: p. 337-352.
75. Cramer, P., *Multisubunit RNA polymerases*. *Current opinion in structural biology*, 2002. **12**(1): p. 89-97.
76. Sweetser, D., M. Nonet, and R.A. Young, *Prokaryotic and eukaryotic RNA polymerases have homologous core subunits*. *Proceedings of the National Academy of Sciences*, 1987. **84**(5): p. 1192-1196.

77. Archambault, J. and J.D. Friesen, *Genetics of eukaryotic RNA polymerases I, II, and III*. Microbiological reviews, 1993. **57**(3): p. 703-724.
78. Biggs, J., L.L. Searles, and A.L. Greenleaf, *Structure of the eukaryotic transcription apparatus: features of the gene for the largest subunit of Drosophila RNA polymerase II*. Cell, 1985. **42**(2): p. 611-621.
79. Minakhin, L., et al., *Bacterial RNA polymerase subunit  $\omega$  and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly*. Proceedings of the National Academy of Sciences, 2001. **98**(3): p. 892-897.
80. Darst, S.A., *Bacterial RNA polymerase*. Current opinion in structural biology, 2001. **11**(2): p. 155-162.
81. Pérez-Rueda, E. and J. Collado-Vides, *The repertoire of DNA-binding transcriptional regulators in Escherichia coli K-12*. Nucleic acids research, 2000. **28**(8): p. 1838-1847.
82. Gross, C.A., M. Lonetto, and R. Losick, *6 Bacterial Sigma Factors*. Cold Spring Harbor Monograph Archive, 1992. **22**: p. 129-176.
83. Gruber, T.M. and C.A. Gross, *Multiple sigma subunits and the partitioning of bacterial transcription space*. Annual Reviews in Microbiology, 2003. **57**(1): p. 441-466.
84. Murakami, K.S. and S.A. Darst, *Bacterial RNA polymerases: the whole story*. Current opinion in structural biology, 2003. **13**(1): p. 31-39.
85. Murakami, K.S., *X-ray crystal structure of Escherichia coli RNA polymerase  $\sigma$ 70 holoenzyme*. Journal of Biological Chemistry, 2013. **288**(13): p. 9126-9134.
86. Darst, S.A., E.W. Kubalek, and R.D. Kornberg, *Three-dimensional structure of Escherichia coli RNA polymerase holoenzyme determined by electron crystallography*. 1989.
87. Zhang, G., et al., *Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 Å resolution*. Cell, 1999. **98**(6): p. 811-824.
88. Haugen, S.P., W. Ross, and R.L. Gourse, *Advances in bacterial promoter recognition and its control by factors that do not bind DNA*. Nature Reviews Microbiology, 2008. **6**(7): p. 507-519.
89. Gourse, R.L., W. Ross, and T. Gaal, *UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition*. Molecular microbiology, 2000. **37**(4): p. 687-695.

90. Mathew, R. and D. Chatterji, *The evolving story of the omega subunit of bacterial RNA polymerase*. Trends in microbiology, 2006. **14**(10): p. 450-455.
91. Ghosh, T., D. Bose, and X. Zhang, *Mechanisms for activating bacterial RNA polymerase*. FEMS microbiology reviews, 2010. **34**(5): p. 611-627.
92. Murakami, K.S., et al., *Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex*. Science, 2002. **296**(5571): p. 1285-1290.
93. Wigneshweraraj, S.R., et al., *Interplay between the  $\beta'$  Clamp and the  $\beta'$  Jaw Domains during DNA Opening by the Bacterial RNA Polymerase at  $\sigma$  54-dependent Promoters*. Journal of molecular biology, 2006. **359**(5): p. 1182-1195.
94. Martinez-Rucobo, F.W. and P. Cramer, *Structural basis of transcription elongation*. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 2013. **1829**(1): p. 9-19.
95. Artem'ev, I., G. Vasil'ev, and A. Gurevich, [*Statistical study of the nucleotide sequence of a prokaryotic promoter. Structural elements determining the efficacy of the transcription initiation stages*]. Bioorganicheskaja khimiia, 1983. **9**(11): p. 1544-1557.
96. Pribnow, D., *Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter*. Proceedings of the National Academy of Sciences, 1975. **72**(3): p. 784-788.
97. Smale, S.T. and J.T. Kadonaga, *The RNA polymerase II core promoter*. Annual review of biochemistry, 2003. **72**(1): p. 449-479.
98. Malhotra, A., E. Severinova, and S.A. Darst, *Crystal structure of a  $\sigma$  70 subunit fragment from E. coli RNA polymerase*. Cell, 1996. **87**(1): p. 127-136.
99. Ross, W., et al., *A Third Recognition Element in Bacterial Promoters: DNA Binding by the Subunit of RNA Polymerase*. SCIENCE-NEW YORK THEN WASHINGTON-, 1993. **262**: p. 1407-1407.
100. Murakami, K.S., S. Masuda, and S.A. Darst, *Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution*. Science, 2002. **296**(5571): p. 1280-1284.
101. Vassylyev, D.G., et al., *Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution*. Nature, 2002. **417**(6890): p. 712-719.
102. Rhodius, V.A., V.K. Mutalik, and C.A. Gross, *Predicting the strength of UP-elements and full-length E. coli  $\sigma$ E promoters*. Nucleic acids research, 2012. **40**(7): p. 2907-2924.

103. Halford, S., *An end to 40 years of mistakes in DNA-protein association kinetics?* Biochemical Society Transactions, 2009. **37**(2): p. 343.
104. von Hippel, P.H. and O. Berg, *Facilitated target location in biological systems.* Journal of Biological Chemistry, 1989. **264**(2): p. 675-678.
105. Endo, M., et al., *Direct Visualization of the Movement of a Single T7 RNA Polymerase and Transcription on a DNA Nanostructure.* Angewandte Chemie, 2012. **124**(35): p. 8908-8912.
106. Bustamante, C., et al., *Facilitated Target Location on DNA by Individual Escherichia coli RNA Polymerase Molecules Observed with the Scanning Force Microscope Operating in Liquid.* Journal of Biological Chemistry, 1999. **274**(24): p. 16665-16668.
107. Kim, J.H. and R.G. Larson, *Single-molecule analysis of 1D diffusion and transcription elongation of T7 RNA polymerase along individual stretched DNA molecules.* Nucleic acids research, 2007. **35**(11): p. 3848-3858.
108. Kabata, H., et al., *Visualization of single molecules of RNA polymerase sliding along DNA.* Science, 1993. **262**(5139): p. 1561-1563.
109. Harada, Y., et al., *Single-molecule imaging of RNA polymerase-DNA interactions in real time.* Biophysical Journal, 1999. **76**(2): p. 709-715.
110. Halford, S.E. and J.F. Marko, *How do site-specific DNA-binding proteins find their targets?* Nucleic Acids Research, 2004. **32**(10): p. 3040-3052.
111. Friedman, L.J., J.P. Mumm, and J. Gelles, *RNA polymerase approaches its promoter without long-range sliding along DNA.* Proceedings of the National Academy of Sciences, 2013. **110**(24): p. 9740-9745.
112. Redding, S. and E.C. Greene, *How do proteins locate specific targets in DNA?* Chemical physics letters, 2013. **570**: p. 1-11.
113. Saecker, R.M. and M.T. Record, *Mechanism of bacterial transcription initiation: RNA polymerase-promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis.* Journal of molecular biology, 2011. **412**(5): p. 754-771.
114. Opalka, N., et al., *Complete structural model of Escherichia coli RNA polymerase from a hybrid approach.* PLoS biology, 2010. **8**(9): p. 2283.
115. Naryshkin, N., et al., *Structural organization of the RNA polymerase-promoter open complex.* Cell, 2000. **101**(6): p. 601-611.

116. Robb, N.C., et al., *The transcription bubble of the RNA polymerase–promoter open complex exhibits conformational heterogeneity and millisecond-scale dynamics: implications for transcription start-site selection*. *Journal of molecular biology*, 2013. **425**(5): p. 875-885.
117. Mekler, V., et al., *Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex*. *Cell*, 2002. **108**(5): p. 599-614.
118. Drennan, A., et al., *Key roles of the downstream mobile jaw of Escherichia coli RNA polymerase in transcription initiation*. *Biochemistry*, 2012. **51**(47): p. 9447-9459.
119. Carpousis, A.J. and J.D. Gralla, *Interaction of RNA polymerase with lacUV5 promoter DNA during mRNA initiation and elongation: footprinting, methylation, and rifampicin-sensitivity changes accompanying transcription initiation*. *Journal of molecular biology*, 1985. **183**(2): p. 165-177.
120. Spassky, A., K. Kirkegaard, and H. Buc, *Changes in the DNA structure of the lac UV5 promoter during formation of an open complex with Escherichia coli RNA polymerase*. *Biochemistry*, 1985. **24**(11): p. 2723-2731.
121. Schickor, P., et al., *Topography of intermediates in transcription initiation of E. coli*. *The EMBO journal*, 1990. **9**(7): p. 2215.
122. Mecsas, J., D.W. Cowing, and C.A. Gross, *Development of RNA polymerase-promoter contacts during open complex formation*. *Journal of molecular biology*, 1991. **220**(3): p. 585-597.
123. Cowing, D.W., et al., *Intermediates in the formation of the open complex by RNA polymerase holoenzyme containing the sigma factor  $\sigma$  32 at the groE promoter*. *Journal of molecular biology*, 1989. **210**(3): p. 521-530.
124. Rees, W.A., et al., *Evidence of DNA bending in transcription complexes imaged by scanning force microscopy*. *Science*, 1993. **260**(5114): p. 1646-1649.
125. Coulombe, B. and Z.F. Burton, *DNA bending and wrapping around RNA polymerase: a “revolutionary” model describing transcriptional mechanisms*. *Microbiology and Molecular Biology Reviews*, 1999. **63**(2): p. 457-478.
126. Rivetti, C., M. Guthold, and C. Bustamante, *Wrapping of DNA around the E. coli RNA polymerase open promoter complex*. *The EMBO journal*, 1999. **18**(16): p. 4464-4475.
127. Mangiarotti, L., et al., *Sequence-Dependent Upstream DNA–RNA Polymerase Interactions in the Open Complex with  $\lambda$ P R and  $\lambda$ P RM Promoters and Implications for the Mechanism of Promoter*

- Interference*. Journal of molecular biology, 2009. **385**(3): p. 748-760.
128. Revyakin, A., R.H. Ebright, and T.R. Strick, *Promoter unwinding and promoter clearance by RNA polymerase: detection by single-molecule DNA nanomanipulation*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(14): p. 4776-4780.
129. Cellai, S., et al., *Upstream promoter sequences and  $\alpha$ CTD mediate stable DNA wrapping within the RNA polymerase–promoter open complex*. EMBO reports, 2007. **8**(3): p. 271-278.
130. Sreenivasan, R., et al., *FRET Characterization of DNA Wrapping in Closed and Open Escherichia coli RNA Polymerase- $\lambda$ PR Promoter Complexes*. 2016.
131. Hsu, L.M., et al., *In vitro studies of transcript initiation by Escherichia coli RNA polymerase. 1. RNA chain initiation, abortive initiation, and promoter escape at three bacteriophage promoters*. Biochemistry, 2003. **42**(13): p. 3777-3786.
132. Revyakin, A., et al., *Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching*. Science, 2006. **314**(5802): p. 1139-1143.
133. Kapanidis, A.N., et al., *Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism*. Science, 2006. **314**(5802): p. 1144-1147.
134. Tang, G.-Q., et al., *Real-time observation of the transition from transcription initiation to elongation of the RNA polymerase*. Proceedings of the National Academy of Sciences, 2009. **106**(52): p. 22175-22180.
135. Mooney, R.A., S.A. Darst, and R. Landick, *Sigma and RNA polymerase: an on-again, off-again relationship?* Molecular cell, 2005. **20**(3): p. 335-345.
136. Kapanidis, A.N., et al., *Retention of transcription initiation factor  $\sigma$  70 in transcription elongation: single-molecule analysis*. Molecular cell, 2005. **20**(3): p. 347-356.
137. Bar-Nahum, G. and E. Nudler, *Isolation and characterization of  $\sigma$  70-retaining transcription elongation complexes from Escherichia coli*. Cell, 2001. **106**(4): p. 443-451.
138. Mukhopadhyay, J., et al., *Translocation of  $\sigma$ 70 with RNA Polymerase during Transcription: Fluorescence Resonance Energy Transfer Assay for Movement Relative to DNA*. Cell, 2001. **106**(4): p. 453-463.

139. Hatoum, A. and J. Roberts, *Prevalence of RNA polymerase stalling at Escherichia coli promoters after open complex formation*. Molecular microbiology, 2008. **68**(1): p. 17-28.
140. Nickels, B.E., et al., *The  $\sigma 70$  subunit of RNA polymerase mediates a promoter-proximal pause at the lac promoter*. Nature structural & molecular biology, 2004. **11**(6): p. 544-550.
141. Schwartz, E.C., et al., *A full-length group 1 bacterial sigma factor adopts a compact structure incompatible with DNA binding*. Chemistry & biology, 2008. **15**(10): p. 1091-1103.
142. Dieci, G., B. Fermi, and M.C. Bosio, *Investigating transcription reinitiation through in vitro approaches*. Transcription, 2014. **5**(1): p. e27704.
143. Margeat, E., et al., *Direct observation of abortive initiation and promoter escape within single immobilized transcription complexes*. Biophysical journal, 2006. **90**(4): p. 1419-1431.
144. Klumpp, S. and T. Hwa, *Growth-rate-dependent partitioning of RNA polymerases in bacteria*. Proceedings of the National Academy of Sciences, 2008. **105**(51): p. 20245-20250.
145. Mauri, M. and S. Klumpp, *A model for sigma factor competition in bacterial cells*. PLoS Comput Biol, 2014. **10**(10): p. e1003845.
146. Naji, S., et al., *Structure–function analysis of the RNA polymerase cleft loops elucidates initial transcription, DNA unwinding and RNA displacement*. Nucleic acids research, 2008. **36**(2): p. 676-687.
147. Vassylyev, D.G., et al., *Structural basis for transcription elongation by bacterial RNA polymerase*. Nature, 2007. **448**(7150): p. 157-162.
148. Vassylyev, D.G., et al., *Structural basis for substrate loading in bacterial RNA polymerase*. Nature, 2007. **448**(7150): p. 163-168.
149. Kireeva, M.L., et al., *Transient reversal of RNA polymerase II active site closing controls fidelity of transcription elongation*. Molecular cell, 2008. **30**(5): p. 557-566.
150. Abbondanzieri, E.A., et al., *Direct observation of base-pair stepping by RNA polymerase*. Nature, 2005. **438**(7067): p. 460-465.
151. Cheung, A.C. and P. Cramer, *A movie of RNA polymerase II transcription*. Cell, 2012. **149**(7): p. 1431-1437.
152. Dangkulwanich, M., et al., *Molecular mechanisms of transcription through single-molecule experiments*. Chemical reviews, 2014. **114**(6): p. 3203-3223.
153. Kolomeisky, A.B. and M.E. Fisher, *Molecular motors: a theorist's perspective*. Annu. Rev. Phys. Chem., 2007. **58**: p. 675-695.
154. Kireeva, M., M. Kashlev, and Z.F. Burton, *Translocation by multi-subunit RNA polymerases*. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms, 2010. **1799**(5–6): p. 389-401.

155. Guajardo, R. and R. Sousa, *A model for the mechanism of polymerase translocation*. *Journal of molecular biology*, 1997. **265**(1): p. 8-19.
156. Yin, Y.W. and T.A. Steitz, *The structural mechanism of translocation and helicase activity in T7 RNA polymerase*. *Cell*, 2004. **116**(3): p. 393-404.
157. Bar-Nahum, G., et al., *A ratchet mechanism of transcription elongation and its control*. *Cell*, 2005. **120**(2): p. 183-193.
158. Nudler, E., et al., *The RNA–DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase*. *Cell*, 1997. **89**(1): p. 33-41.
159. Komissarova, N. and M. Kashlev, *Transcriptional arrest: Escherichia coli RNA polymerase translocates backward, leaving the 3' end of the RNA intact and extruded*. *Proceedings of the National Academy of Sciences*, 1997. **94**(5): p. 1755-1760.
160. Weixlbaumer, A., et al., *Structural basis of transcriptional pausing in bacteria*. *Cell*, 2013. **152**(3): p. 431-441.
161. Artsimovitch, I. and R. Landick, *Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals*. *Proceedings of the National Academy of Sciences*, 2000. **97**(13): p. 7090-7095.
162. Neuman, K.C., et al., *Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking*. *Cell*, 2003. **115**(4): p. 437-447.
163. Lis, J. *Promoter-associated pausing in promoter architecture and postinitiation transcriptional regulation*. in *Cold Spring Harbor symposia on quantitative biology*. 1998. Cold Spring Harbor Laboratory Press.
164. Sydow, J.F., et al., *Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA*. *Molecular cell*, 2009. **34**(6): p. 710-721.
165. Touloukhonov, I., et al., *A central role of the RNA polymerase trigger loop in active-site rearrangement during transcriptional pausing*. *Molecular cell*, 2007. **27**(3): p. 406-419.
166. Zenkin, N., Y. Yuzenkova, and K. Severinov, *Transcript-assisted transcriptional proofreading*. *Science*, 2006. **313**(5786): p. 518-520.
167. Cheung, A.C. and P. Cramer, *Structural basis of RNA polymerase II backtracking, arrest and reactivation*. *Nature*, 2011. **471**(7337): p. 249-253.
168. Nudler, E., *RNA polymerase backtracking in gene regulation and genome instability*. *Cell*, 2012. **149**(7): p. 1438-1445.



169. Fish, R.N. and C.M. Kane, *Promoting elongation with transcript cleavage stimulatory factors*. *Biochimica et biophysica acta*, 2002. **1577**(2): p. 287-307.
170. Gusarov, I. and E. Nudler, *The mechanism of intrinsic transcription termination*. *Molecular cell*, 1999. **3**(4): p. 495-504.
171. Henkin, T.M., *Control of transcription termination in prokaryotes*. *Annual review of genetics*, 1996. **30**(1): p. 35-57.
172. Porrua, O. and D. Libri, *Transcription termination and the control of the transcriptome: why, where and how to stop*. *Nature Reviews Molecular Cell Biology*, 2015.
173. Brennan, C.A., A.J. Dombroski, and T. Platt, *Transcription termination factor rho is an RNA-DNA helicase*. *Cell*, 1987. **48**(6): p. 945-952.
174. Burns, C.M., L.V. Richardson, and J.P. Richardson, *Combinatorial effects of NusA and NusG on transcription elongation and Rho-dependent termination in Escherichia coli*. *Journal of molecular biology*, 1998. **278**(2): p. 307-316.
175. Rocha, E.P., *The organization of the bacterial genome*. *Annual review of genetics*, 2008. **42**: p. 211-233.
176. Lewis, M., *Allostery and the lac Operon*. *Journal of Molecular Biology*, 2013. **425**(13): p. 2309-2316.
177. Yin, Y., et al., *Genomic arrangement of bacterial operons is constrained by biological pathways encoded in the genome*. *Proceedings of the National Academy of Sciences*, 2010. **107**(14): p. 6310-6315.
178. Gowrishankar, J. and R. Harinarayanan, *Why is transcription coupled to translation in bacteria?* *Molecular microbiology*, 2004. **54**(3): p. 598-603.
179. Pi, H., L.-W. Lee, and S.J. Lo, *New insights into polycistronic transcripts in eukaryotes*. *Chang Gung medical journal*, 2009. **32**(5): p. 494-498.
180. Lodish, H., et al., *Processing of eukaryotic mRNA*. 2000.
181. Chorev, M. and L. Carmel, *The function of introns*. *Front genet*, 2012. **3**: p. 55.
182. Fukuda, Y., Y. Nakayama, and M. Tomita, *On dynamics of overlapping genes in bacterial genomes*. *Gene*, 2003. **323**.
183. Georg, J. and W.R. Hess, *cis-antisense RNA, another level of gene regulation in bacteria*. *Microbiology and Molecular Biology Reviews*, 2011. **75**(2): p. 286-300.
184. Veeramachaneni, V., et al., *Mammalian overlapping genes: the comparative perspective*. *Genome research*, 2004. **14**(2): p. 280-286.

185. Chen, N. and L.D. Stein, *Conservation and functional significance of gene topology in the genome of Caenorhabditis elegans*. Genome research, 2006. **16**(5): p. 606-617.
186. Yu, P., D. Ma, and M. Xu, *Nested genes in the human genome*. Genomics, 2005. **86**(4): p. 414-422.
187. Makalowska, I., C.-F. Lin, and W. Makalowski, *Overlapping genes in vertebrate genomes*. Comput Biol Chem, 2005. **29**.
188. Henikoff, S. and M.K. Eghtedarzadeh, *Conserved arrangement of nested genes at the Drosophila Gart locus*. Genetics, 1987. **117**(4): p. 711-725.
189. Gibson, C.W., et al., *Nested genes: biological implications and use of AFM for analysis*. Gene, 2005. **350**(1): p. 15-23.
190. Assis, R., et al., *Nested genes and increasing organizational complexity of metazoan genomes*. Trends in genetics, 2008. **24**(10): p. 475-478.
191. Prescott, E.M. and N.J. Proudfoot, *Transcriptional collision between convergent genes in budding yeast*. Proceedings of the National Academy of Sciences, 2002. **99**(13): p. 8796-8801.
192. Callen, B.P., K.E. Shearwin, and J.B. Egan, *Transcriptional interference between convergent promoters caused by elongation over the promoter*. Molecular cell, 2004. **14**(5): p. 647-656.
193. Sneppen, K., et al., *A Mathematical Model for Transcriptional Interference by RNA Polymerase Traffic in Escherichia coli*. Journal of Molecular Biology, 2005. **346**(2): p. 399-409.
194. Crampton, N., *Convergent transcription studied by atomic force microscopy*. 2005.
195. Hobson, D.J., *RNA Polymerase II Collision & Its Role in Transcript Elongation*. 2012.
196. Chatterjee, A., et al., *Convergent transcription in the butyrolactone regulon in Streptomyces coelicolor confers a bistable genetic switch for antibiotic biosynthesis*. PLoS One, 2011. **6**(7): p. e21974.
197. Ma, N. and W.T. McAllister, *In a head-on collision, two RNA polymerases approaching one another on the same DNA may pass by one another*. Journal of molecular biology, 2009. **391**(5): p. 808-812.
198. Gotta, S.L., O. Miller, and S.L. French, *rRNA transcription rate in Escherichia coli*. Journal of bacteriology, 1991. **173**(20): p. 6647-6649.
199. Jin, J., et al., *Synergistic action of RNA polymerases in overcoming the nucleosomal barrier*. Nat Struct Mol Biol, 2010. **17**(6): p. 745-752.

200. Saeki, H. and J.Q. Svejstrup, *Stability, flexibility, and dynamic interactions of colliding RNA polymerase II elongation complexes*. *Molecular cell*, 2009. **35**(2): p. 191-205.
201. Epshtein, V. and E. Nudler, *Cooperation between RNA polymerase molecules in transcription elongation*. *Science*, 2003. **300**(5620): p. 801-805.
202. Kubori, T. and N. Shimamoto, *Physical interference between Escherichia coli RNA polymerase molecules transcribing in tandem enhances abortive synthesis and misincorporation*. *Nucleic acids research*, 1997. **25**(13): p. 2640-2647.
203. Ponnambalam, S. and S. Busby, *RNA polymerase molecules initiating transcription at tandem promoters can collide and cause premature transcription termination*. *FEBS letters*, 1987. **212**(1): p. 21-27.
204. Ido, S., et al., *Beyond the helix pitch: direct visualization of native DNA in aqueous solution*. *ACS nano*, 2013. **7**(2): p. 1817-1822.
205. Leung, C., et al., *Atomic force microscopy with nanoscale cantilevers resolves different structural conformations of the DNA double helix*. *Nano letters*, 2012. **12**(7): p. 3846-3850.
206. Lyubchenko, Y.L., A.A. Gall, and L.S. Shlyakhtenko, *Visualization of DNA and Protein–DNA Complexes with Atomic Force Microscopy*, in *Electron Microscopy*. 2014, Springer. p. 367-384.
207. Rajendran, A., M. Endo, and H. Sugiyama, *State-of-the-art high-speed atomic force microscopy for investigation of single-molecular dynamics of proteins*. *Chemical reviews*, 2013. **114**(2): p. 1493-1520.
208. Ando, T., et al., *A high-speed atomic force microscope for studying biological macromolecules in action*. *Japanese Journal of Applied Physics*, 2002. **41**(7S): p. 4851.
209. Fotiadis, D., et al., *Imaging and manipulation of biological structures with the AFM*. *Micron*, 2002. **33**(4): p. 385-397.
210. Kasas, S., et al., *Biological applications of the AFM: from single molecules to organs*. *International journal of imaging systems and technology*, 1997. **8**(2): p. 151-161.
211. Moreno-Herrero, F., et al., *Atomic force microscopy contact, tapping, and jumping modes for imaging biological samples in liquids*. *Physical Review E*, 2004. **69**(3): p. 031915.
212. Ando, T., et al., *A high-speed atomic force microscope for studying biological macromolecules*. *Proceedings of the National Academy of Sciences*, 2001. **98**(22): p. 12468-12472.
213. Liu, S. and Y. Wang, *Application of AFM in microbiology: a review*. *Scanning*, 2010. **32**(2): p. 61-73.

214. Lyubchenko, Y.L., L.S. Shlyakhtenko, and T. Ando, *Imaging of nucleic acids with atomic force microscopy*. *Methods*, 2011. **54**(2): p. 274-283.
215. Ristic, D., H. Sanchez, and C. Wyman, *Sample preparation for SFM imaging of DNA, proteins, and DNA–protein complexes*, in *Single molecule analysis*. 2011, Springer. p. 213-231.
216. Buechner, C.N. and I. Tessmer, *DNA substrate preparation for atomic force microscopy studies of protein–DNA interactions*. *Journal of Molecular Recognition*, 2013. **26**(12): p. 605-617.
217. Binnig, G. and H. Rohrer, *Scanning tunneling microscopy*. *Surface science*, 1983. **126**(1): p. 236-244.
218. Driscoll, R.J., M.G. Youngquist, and J.D. Baldeschwieler, *Atomic-scale imaging of DNA using scanning tunnelling microscopy*. *Nature*, 1990. **346**(6281): p. 294-296.
219. Clemmer, C.R. and T.P. Beebe, *Graphite: a mimic for DNA and other biomolecules in scanning tunneling microscope studies*. *Science*, 1991. **251**(4994): p. 640-642.
220. Binnig, G., C.F. Quate, and C. Gerber, *Atomic force microscope*. *Physical review letters*, 1986. **56**(9): p. 930-933.
221. San Paulo, A. and R. Garcia, *High-resolution imaging of antibodies by tapping-mode atomic force microscopy: attractive and repulsive tip-sample interaction regimes*. *Biophysical Journal*, 2000. **78**(3): p. 1599-1605.
222. Czajkowsky, D.M. and Z. Shao, *Submolecular resolution of single macromolecules with atomic force microscopy*. *FEBS letters*, 1998. **430**(1): p. 51-54.
223. Gan, Y., *Atomic and subnanometer resolution in ambient conditions by atomic force microscopy*. *Surface Science Reports*, 2009. **64**(3): p. 99-121.
224. Putman, C.A., et al., *Tapping mode atomic force microscopy in liquid*. *Applied physics letters*, 1994. **64**(18): p. 2454-2456.
225. Hansma, P., et al., *Tapping mode atomic force microscopy in liquids*. *Applied Physics Letters*, 1994. **64**(13): p. 1738-1740.
226. Yang, Y., H. Wang, and D.A. Erie, *Quantitative characterization of biomolecular assemblies and interactions using atomic force microscopy*. *Methods*, 2003. **29**(2): p. 175-187.
227. Alessandrini, A. and P. Facci, *AFM: a versatile tool in biophysics*. *Measurement science and technology*, 2005. **16**(6): p. R65.
228. Meyer, G. and N.M. Amer, *Novel optical approach to atomic force microscopy*. *Applied physics letters*, 1988. **53**(12): p. 1045-1047.
229. Bustamante, C. and C. Rivetti, *Visualizing protein-nucleic acid interactions on a large scale with the scanning force microscope*.

- Annual review of biophysics and biomolecular structure, 1996. **25**(1): p. 395-429.
230. Rugar, D. and P. Hansma, *Atomic force microscopy*. Physics today, 1990. **43**(10): p. 23-30.
231. Binnig, G. and D. Smith, *Single-tube three-dimensional scanner for scanning tunneling microscopy*. Review of Scientific Instruments, 1986. **57**(8): p. 1688-1689.
232. Pohl, D.W., *Dynamic piezoelectric translation devices*. Review of Scientific Instruments, 1987. **58**(1): p. 54-57.
233. Israelachvili, J.N., *Intermolecular and surface forces: revised third edition*. 2011: Academic press.
234. Cross, S.J., *Combining magnetic tweezers and single-molecule fluorescence microscopy to probe transcription-coupled DNA supercoiling*. 2013, University of York.
235. Pauli, W., *On the connexion between the completion of electron groups in an atom with the complex structure of spectra*. Zeitschrift für Physik, 1925. **31**: p. 765.
236. Blackman, G., C. Mate, and M. Philpott, *Interaction forces of a sharp tungsten tip with molecular films on silicon surfaces*. Physical review letters, 1990. **65**(18): p. 2270.
237. Sirghi, L., et al., *Atomic force microscopy study of the hydrophilicity of TiO<sub>2</sub> thin films obtained by radio frequency magnetron sputtering and plasma enhanced chemical vapor depositions*. Langmuir, 2001. **17**(26): p. 8199-8203.
238. Xu, L., et al., *Wetting and capillary phenomena of water on mica*. The Journal of Physical Chemistry B, 1998. **102**(3): p. 540-548.
239. Colchero, J., et al., *Observation of liquid neck formation with scanning force microscopy techniques*. Langmuir, 1998. **14**(9): p. 2230-2234.
240. Hu, J., D. Ogletree, and M. Salmeron, *The structure of molecularly thin films of water on mica in humid environments*. Surface science, 1995. **344**(3): p. 221-236.
241. Yang, L., Y.-s. Tu, and H.-l. Tan, *Influence of atomic force microscope (AFM) probe shape on adhesion force measured in humidity environment*. Applied Mathematics and Mechanics, 2014: p. 1-8.
242. Stifter, T., O. Marti, and B. Bhushan, *Theoretical investigation of the distance dependence of capillary and van der Waals forces in scanning force microscopy*. Physical Review B, 2000. **62**(20): p. 13667.

243. Pakarinen, O., et al., *High-resolution scanning force microscopy of gold nanoclusters on the KBr (001) surface*. Physical Review B, 2006. **73**(23): p. 235428.
244. Zitzler, L., S. Herminghaus, and F. Mugele, *Capillary forces in tapping mode atomic force microscopy*. Physical Review B, 2002. **66**(15): p. 155436.
245. Hansma, H., et al., *Progress in sequencing deoxyribonucleic acid with an atomic force microscope*. Journal of Vacuum Science & Technology B, 1991. **9**(2): p. 1282-1284.
246. Lindsay, S., et al., *STM and AFM images of nucleosome DNA under water*. Journal of Biomolecular Structure and Dynamics, 1989. **7**(2): p. 279-287.
247. Zhong, Q., et al., *Fractured polymer/silica fiber surface studied by tapping mode atomic force microscopy*. Surface Science Letters, 1993. **290**(1-2): p. L688-L692.
248. Anselmetti, D., et al., *Attractive-mode imaging of biological materials with dynamic force microscopy*. Nanotechnology, 1994. **5**: p. 87.
249. Santos, S., et al., *A method to provide rapid in situ determination of tip radius in dynamic atomic force microscopy*. Review of Scientific Instruments, 2012. **83**(4): p. 043707.
250. Round, A.N. and M.J. Miles, *Exploring the consequences of attractive and repulsive interaction regimes in tapping mode atomic force microscopy of DNA*. Nanotechnology, 2004. **15**(4): p. S176.
251. Santos, S., et al., *Bi-stability of amplitude modulation AFM in air: deterministic and stochastic outcomes for imaging biomolecular systems*. Nanotechnology, 2010. **21**(22): p. 225710.
252. Santos, S., et al., *Stability, resolution, and ultra-low wear amplitude modulation atomic force microscopy of DNA: Small amplitude small set-point imaging*. Applied Physics Letters, 2013. **103**(6): p. 063702.
253. Cleveland, J., et al., *Energy dissipation in tapping-mode atomic force microscopy*. Applied Physics Letters, 1998. **72**(20): p. 2613-2615.
254. Tamayo, J. and R. García, *Relationship between phase shift and energy dissipation in tapping-mode scanning force microscopy*. Applied Physics Letters, 1998. **73**(20): p. 2926-2928.
255. Schmitz, I., et al., *Phase imaging as an extension to tapping mode AFM for the identification of material properties on humidity-sensitive surfaces*. Applied surface science, 1997. **115**(2): p. 190-198.

256. Magonov, S., V. Elings, and M.-H. Whangbo, *Phase imaging and stiffness in tapping-mode atomic force microscopy*. Surface Science, 1997. **375**(2): p. L385-L391.
257. Kim, J.M., et al., *AFM phase lag mapping for protein–DNA oligonucleotide complexes*. Analytica chimica acta, 2004. **525**(2): p. 151-157.
258. Viani, M.B., et al., *Probing protein–protein interactions in real time*. Nature Structural & Molecular Biology, 2000. **7**(8): p. 644-647.
259. Fantner, G.E., et al., *Data acquisition system for high speed atomic force microscopy*. Review of Scientific Instruments, 2005. **76**(2): p. 026118.
260. Picco, L., et al., *Breaking the speed limit with atomic force microscopy*. Nanotechnology, 2007. **18**(4): p. 044030.
261. Humphris, A., M. Miles, and J. Hobbs, *A mechanical microscope: high-speed atomic force microscopy*. Applied physics letters, 2005. **86**(3): p. 34106-34106.
262. Humphris, A.D., J.K. Hobbs, and M.J. Miles, *Ultrahigh-speed scanning near-field optical microscopy capable of over 100 frames per second*. Applied Physics Letters, 2003. **83**(1): p. 6-8.
263. Benedetti, F., et al., *Effects of physiological self-crowding of DNA on shape and biological properties of DNA molecules with various levels of supercoiling*. Nucleic acids research, 2015: p. gkv055.
264. Crampton, N., et al., *Studying silane mobility on hydrated mica using ambient AFM*. Ultramicroscopy, 2006. **106**(8): p. 765-770.
265. Hansma, H.G. and D.E. Laney, *DNA binding to mica correlates with cationic radius: assay by atomic force microscopy*. Biophysical journal, 1996. **70**(4): p. 1933.
266. Thomson, N.H., et al., *Protein tracking and detection of protein motion using atomic force microscopy*. Biophysical journal, 1996. **70**(5): p. 2421-2431.
267. Vesenska, J., et al., *Substrate preparation for reliable imaging of DNA molecules with the scanning force microscope*. Ultramicroscopy, 1992. **42**: p. 1243-1249.
268. Pastré, D., et al., *Anionic polyelectrolyte adsorption on mica mediated by multivalent cations: a solution to DNA imaging by atomic force microscopy under high ionic strengths*. Langmuir, 2006. **22**(15): p. 6651-6660.
269. Pastré, D., et al., *Study of the DNA/ethidium bromide interactions on mica surface by atomic force microscope: influence of the surface friction*. Biopolymers, 2005. **77**(1): p. 53-62.

270. Pastré, D., et al., *Adsorption of DNA to mica mediated by divalent counterions: a theoretical and experimental study*. Biophysical journal, 2003. **85**(4): p. 2507-2518.
271. Pashley, R., *Hydration forces between mica surfaces in aqueous electrolyte solutions*. Journal of Colloid and Interface Science, 1981. **80**(1): p. 153-162.
272. Arjmand, F. and M. Aziz, *Synthesis and characterization of dinuclear macrocyclic cobalt (II), copper (II) and zinc (II) complexes derived from 2, 2', 2''-S, S [bis (bis-N, N-2-thiobenzimidazolyloxalato-1, 2-ethane)]*: DNA binding and cleavage studies. European journal of medicinal chemistry, 2009. **44**(2): p. 834-844.
273. Cheng, H., et al., *Polynucleotide adsorption to negatively charged surfaces in divalent salt solutions*. Biophysical journal, 2006. **90**(4): p. 1164-1174.
274. Abrescia, N.G., T. Huynh-Dinh, and J.A. Subirana, *Nickel-guanine interactions in DNA: crystal structure of nickel-d [CGTGTACACG] 2*. JBIC Journal of Biological Inorganic Chemistry, 2002. **7**(1-2): p. 195-199.
275. Lang, D. and P. Coates, *Diffusion coefficient of DNA in solution at "Zero" concentration as measured by electron microscopy*. Journal of molecular biology, 1968. **36**(1): p. 137-151.
276. Rivetti, C., C. Walker, and C. Bustamante, *Polymer chain statistics and conformational analysis of DNA molecules with bends or sections of different flexibility*. Journal of molecular biology, 1998. **280**(1): p. 41-59.
277. Rivetti, C., M. Guthold, and C. Bustamante, *Scanning Force Microscopy of DNA Deposited onto Mica: Equilibration versus Kinetic Trapping Studied by Statistical Polymer Chain Analysis*. Journal of molecular biology, 1996. **264**(5): p. 919-932.
278. Bhushan, B. and S. Kawata, *Applied scanning probe methods VI*. 2007: Springer.
279. Thomson, N., et al., *Reversible binding of DNA to mica for AFM imaging*. Langmuir, 1996. **12**(24): p. 5905-5908.
280. Bustamante, C., et al., *Estimating the persistence length of a worm-like chain molecule from force-extension measurements*. Science, 1994. **265**: p. 1599-1600.
281. Billingsley, D.J., et al., *Patchiness of ion-exchanged mica revealed by DNA binding dynamics at short length scales*. Nanotechnology, 2014. **25**(2): p. 025704.
282. Lee, A.J., et al., *Tuning the translational freedom of DNA for high speed AFM*. Nano Research, 2014: p. 1-11.



283. Sundstrom, A., et al., *Image analysis and length estimation of biomolecules using AFM*. Information Technology in Biomedicine, IEEE Transactions on, 2012. **16**(6): p. 1200-1207.
284. Rivetti, C., *A simple and optimized length estimator for digitized DNA contours*. Cytometry Part A, 2009. **75**(10): p. 854-861.
285. Mikhaylov, A., S.K. Sekatskii, and G. Dietler, *DNA trace: a comprehensive software for polymer image processing*. Journal of Advanced Microscopy Research, 2013. **8**(4): p. 241-245.
286. Spisz, M.T., et al., *Automated sizing of DNA fragments in atomic force microscope images*. Medical and Biological Engineering and Computing, 1998. **36**(6): p. 667-672.
287. Pastushenko, V.P., et al., *Reconstruction of DNA shape from AFM data*. Single Molecules, 2002. **3**(2-3): p. 111-117.
288. Rivetti, C. and S. Codeluppi, *Accurate length determination of DNA molecules visualized by atomic force microscopy: evidence for a partial B-to A-form transition on mica*. Ultramicroscopy, 2001. **87**(1): p. 55-66.
289. Sanchez-Sevilla, A., et al., *Accuracy of AFM measurements of the contour length of DNA fragments adsorbed on mica in air and in aqueous buffer*. Ultramicroscopy, 2002. **92**(3): p. 151-158.
290. Fuentes-Perez, M.E., M.S. Dillingham, and F. Moreno-Herrero, *AFM volumetric methods for the characterization of proteins and nucleic acids*. Methods, 2013. **60**(2): p. 113-121.
291. Herbert, K.M., W.J. Greenleaf, and S.M. Block, *Single-molecule studies of RNA polymerase: motoring along*. Annual review of biochemistry, 2008. **77**: p. 149.
292. Singer, P. and C.-W. Wu, *Promoter search by Escherichia coli RNA polymerase on a circular DNA template*. Journal of Biological Chemistry, 1987. **262**(29): p. 14178-14189.
293. Billingsley, D.J., et al., *Single-molecule studies of DNA transcription using atomic force microscopy*. Physical biology, 2012. **9**(2): p. 021001.
294. Rivetti, C., et al., *Visualizing RNA extrusion and DNA wrapping in transcription elongation complexes of bacterial and eukaryotic RNA polymerases*. Journal of molecular biology, 2003. **326**(5): p. 1413-1426.
295. Guthold, M., et al., *Following the assembly of RNA polymerase-DNA complexes in aqueous solutions with the scanning force microscope*. Proceedings of the National Academy of Sciences, 1994. **91**(26): p. 12927-12931.

296. Guthold, M., et al., *Direct observation of one-dimensional diffusion and transcription by Escherichia coli RNA polymerase*. Biophysical journal, 1999. **77**(4): p. 2284-2294.
297. Thomson, N.H., et al., *Oriented, active Escherichia coli RNA polymerase: an atomic force microscope study*. Biophysical journal, 1999. **76**(2): p. 1024-1033.
298. Kasas, S., et al., *Escherichia coli RNA polymerase activity observed using atomic force microscopy*. Biochemistry, 1997. **36**(3): p. 461-468.
299. Costa, L.T., et al., *Chemical treatment of mica for atomic force microscopy can affect biological sample conformation*. Biophysical chemistry, 2004. **109**(1): p. 63-71.
300. Kellermayer, M.S., *Combined atomic force microscopy and fluorescence microscopy*. Atomic Force Microscopy in Biomedical Research: Methods and Protocols, 2011: p. 439-456.
301. Fukuda, S., et al., *High-speed atomic force microscope combined with single-molecule fluorescence microscope*. Review of Scientific Instruments, 2013. **84**(7): p. 073706.
302. Sanchez, H., et al., *Combined optical and topographic imaging reveals different arrangements of human RAD54 with presynaptic and postsynaptic RAD51–DNA filaments*. Proceedings of the National Academy of Sciences, 2013. **110**(28): p. 11385-11390.
303. Santos, S., et al., *The intrinsic resolution limit in the atomic force microscope: implications for heights of nano-scale features*. PLoS One, 2011. **6**(8): p. e23821.
304. Van Noort, S.J.T., et al., *Height anomalies in tapping mode atomic force microscopy in air caused by adhesion*. Ultramicroscopy, 1997. **69**(2): p. 117-127.
305. Winzer, A., et al., *Correcting for AFM tip induced topography convolutions in protein–DNA samples*. Ultramicroscopy, 2012. **121**: p. 8-15.
306. Fuentes-Perez, M., et al., *Using DNA as a fiducial marker to study SMC complex interactions with the atomic force microscope*. Biophysical journal, 2012. **102**(4): p. 839-848.
307. Rothmund, P.W., *Folding DNA to create nanoscale shapes and patterns*. Nature, 2006. **440**(7082): p. 297-302.
308. Bald, I. and A. Keller, *Molecular Processes Studied at a Single-Molecule Level Using DNA Origami Nanostructures and Atomic Force Microscopy*. Molecules, 2014. **19**(9): p. 13803-13823.
309. Yamamoto, S., et al., *Single molecule visualization and characterization of Sox2–Pax6 complex formation on a regulatory*

- DNA element using a DNA origami frame*. Nano letters, 2014. **14**(5): p. 2286-2292.
310. Suzuki, Y., et al., *Fast microscopical dissection of action scenes played by *Escherichia coli* RNA polymerase*. FEBS letters, 2012.
311. Metzger, W., P. Schickor, and H. Heumann, *A cinematographic view of Escherichia coli RNA polymerase translocation*. The EMBO journal, 1989. **8**(9): p. 2745.
312. Roy, R., S. Hohng, and T. Ha, *A practical guide to single-molecule FRET*. Nature methods, 2008. **5**(6): p. 507-516.
313. Doniselli, N., et al., *New insights into the regulatory mechanisms of ppGpp and DksA on Escherichia coli RNA polymerase–promoter complex*. Nucleic acids research, 2015. **43**(10): p. 5249-5262.
314. Gerganova, V., et al., *Upstream Binding of Idling RNA Polymerase Modulates Transcription Initiation from a Nearby Promoter*. Journal of Biological Chemistry, 2015. **290**(13): p. 8095-8109.
315. Rippe, K., et al., *Transcriptional activation via DNA-looping: visualization of intermediates in the activation pathway of E. coli RNA polymerase·  $\sigma$  54 holoenzyme by scanning force microscopy*. Journal of molecular biology, 1997. **270**(2): p. 125-138.
316. Schulz, A., et al., *Scanning force microscopy of Escherichia coli RNA polymerase·  $\sigma$  54 holoenzyme complexes with DNA in buffer and in air*. Journal of molecular biology, 1998. **283**(4): p. 821-836.
317. Yin, H., et al., *Transcription against an applied force*. Science, 1995. **270**(5242): p. 1653-1657.
318. Hobson, D.J., et al., *RNA polymerase II collision interrupts convergent transcription*. Molecular cell, 2012. **48**(3): p. 365-374.
319. Ebenstein, Y., et al., *Combining atomic force and fluorescence microscopy for analysis of quantum-dot labeled protein–DNA complexes*. Journal of Molecular Recognition, 2009. **22**(5): p. 397-402.
320. Horn, A.E., J.A. Goodrich, and J.F. Kugel, *Single molecule studies of RNA polymerase II transcription in vitro*. Transcription, 2014. **5**(1): p. e27608.
321. Bintu, L., et al., *The elongation rate of RNA polymerase determines the fate of transcribed nucleosomes*. Nature structural & molecular biology, 2011. **18**(12): p. 1394-1399.
322. Limanskaya, O. and A. Limanskii, *Study of elongation complexes for T7 RNA polymerase*. Biophysics, 2012. **57**(4): p. 428-441.
323. Kotlajich, M.V., et al., *Bridged filaments of histone-like nucleoid structuring protein pause RNA polymerase and aid termination in bacteria*. Elife, 2015. **4**: p. e04970.

324. Craig, M.L., W.-C. Suh, and M.T. Record Jr, *HO. bul. and DNase I Probing of E. sigma. 70 RNA Polymerase-. lambda. PR Promoter Open Complexes: Mg<sup>2+</sup> Binding and Its Structural Consequences at the Transcription Start Site*. *Biochemistry*, 1995. **34**(48): p. 15624-15632.
325. Walz, A. and V. Pirrotta, *Sequence of the PR promoter of phage lambda*. *Nature*, 1975. **254**(5496): p. 118-121.
326. Rolfs, A., *PCR: clinical diagnostics and research*. 1992: Springer.
327. Mullis, K.B., *The unusual origin of the polymerase chain reaction*. *Scientific American*, 1990. **262**(4): p. 56-61.
328. Mullis, K.B., et al., *One of the first Polymerase Chain Reaction (PCR) patents*. 1987, Google Patents.
329. Hadidi, A. and T. Candresse, *Polymerase chain reaction*. *Viroids*, 2003: p. 115-122.
330. Saiki, R.K., et al., *Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase*. *Science*, 1988. **239**(4839): p. 487-491.
331. Bartlett, J.M. and D. Stirling, *A short history of the polymerase chain reaction*, in *PCR protocols*. 2003, Springer. p. 3-6.
332. Chien, A., D.B. Edgar, and J.M. Trela, *Deoxyribonucleic acid polymerase from the extreme thermophile Thermus aquaticus*. *Journal of bacteriology*, 1976. **127**(3): p. 1550-1557.
333. Eckert, K.A. and T.A. Kunkel, *High fidelity DNA synthesis by the Thermus aquaticus DNA polymerase*. *Nucleic Acids Research*, 1990. **18**(13): p. 3739-3744.
334. Sharkey, D.J., et al., *Antibodies as thermolabile switches: high temperature triggering for the polymerase chain reaction*. *Nature Biotechnology*, 1994. **12**(5): p. 506-509.
335. Viljoen, G.J., L.H. Nel, and J.R. Crowther, *Molecular diagnostic PCR handbook*. 2005: Springer Science & Business Media.
336. Lawyer, F.C., et al., *High-level expression, purification, and enzymatic characterization of full-length Thermus aquaticus DNA polymerase and a truncated form deficient in 5' to 3' exonuclease activity*. *Genome research*, 1993. **2**(4): p. 275-287.
337. Sambrook, J. and D.W. Russell, *Agarose gel electrophoresis*. Cold Spring Harb. Protoc, 2006.
338. Johansson, B., *Agarose gel electrophoresis*. *Scandinavian Journal of Clinical & Laboratory Investigation*, 1972. **29**(S124): p. 7-19.
339. Grundemann, D. and E. Schomig, *Protection of DNA during preparative agarose gel electrophoresis against damage induced by ultraviolet light*. *BioTechniques*, 1996. **21**(5): p. 898-903.

340. Carlson, D.P., P.C. Watkins, and L. Klevan, *Size markers for electrophoretic analysis of DNA*. 1994, Google Patents.
341. Handbook, Q.P.P., *Qiagen Inc*. Valencia, CA, 1998.
342. Hinegardner, R.T., *An improved fluorometric assay for DNA*. Analytical biochemistry, 1971. **39**(1): p. 197-201.
343. Swinehart, D., *The beer-lambert law*. Journal of chemical education, 1962. **39**(7): p. 333.
344. Crampton, N., et al., *Imaging RNA polymerase–amelogenin gene complexes with single molecule resolution using atomic force microscopy*. European journal of oral sciences, 2006. **114**(s1): p. 133-138.
345. Abràmoff, M.D., P.J. Magalhães, and S.J. Ram, *Image processing with ImageJ*. Biophotonics international, 2004. **11**(7): p. 36-42.
346. Reed, J., et al., *Identifying individual DNA species in a complex mixture by precisely measuring the spacing between nicking restriction enzymes with atomic force microscope*. Journal of The Royal Society Interface, 2012. **9**(74): p. 2341-2350.
347. Gebeyehu, G., et al., *Novel biotinylated nucleotide-analogs for labeling and colorimetric detection of DNA*. Nucleic acids research, 1987. **15**(11): p. 4513-4534.
348. Niemeyer, C.M., et al., *Self-assembly of DNA-streptavidin nanostructures and their use as reagents in immuno-PCR*. Nucleic Acids Research, 1999. **27**(23): p. 4553-4561.
349. Neish, C.S., et al., *Direct visualization of ligand-protein interactions using atomic force microscopy*. British journal of pharmacology, 2002. **135**(8): p. 1943-1950.
350. Weber, P.C., et al., *Structural origins of high-affinity biotin binding to streptavidin*. Science, 1989. **243**(4887): p. 85-88.
351. Muzard, G., B. Theveny, and B. Revet, *Electron microscopy mapping of pBR322 DNA curvature. Comparison with theoretical models*. The EMBO journal, 1990. **9**(4): p. 1289.
352. Shaiu, W.-L., et al., *Atomic force microscopy of oriented linear DNA molecules labeled with 5nm gold spheres*. Nucleic acids research, 1993. **21**(1): p. 99-103.
353. Marilley, M., A. Sanchez-Sevilla, and J. Rocca-Serra, *Fine mapping of inherent flexibility variation along DNA molecules. Validation by atomic force microscopy (AFM) in buffer*. Molecular Genetics and Genomics, 2005. **274**(6): p. 658-670.
354. Buzio, R., et al., *Label-free, atomic force microscopy-based mapping of DNA intrinsic curvature for the nanoscale comparative analysis of bent duplexes*. Nucleic acids research, 2012. **40**(11): p. e84-e84.

355. Zohar, H. and S.J. Muller, *Labeling DNA for single-molecule experiments: methods of labeling internal specific sequences on double-stranded DNA*. *Nanoscale*, 2011. **3**(8): p. 3027-3039.
356. Escudé, C., et al., *Multiple topological labeling for imaging single plasmids*. *Analytical biochemistry*, 2007. **362**(1): p. 55-62.
357. Géron-Landre, B., T. Roulon, and C. Escudé, *Stem-loop oligonucleotides as tools for labelling double-stranded DNA*. *FEBS Journal*, 2005. **272**(20): p. 5343-5352.
358. Maxwell, A., N.P. Burton, and N. O'Hagan, *High-throughput assays for DNA gyrase and other topoisomerases*. *Nucleic acids research*, 2006. **34**(15): p. e104-e104.
359. Zuker, M., *Mfold web server for nucleic acid folding and hybridization prediction*. *Nucleic acids research*, 2003. **31**(13): p. 3406-3415.
360. Holland, P.M., et al., *Detection of specific polymerase chain reaction product by utilizing the 5'----3'exonuclease activity of *Thermus aquaticus* DNA polymerase*. *Proceedings of the National Academy of Sciences*, 1991. **88**(16): p. 7276-7280.
361. Guy, A.T., T.J. Piggot, and S. Khalid, *Single-stranded DNA within nanopores: conformational dynamics and implications for sequencing; a molecular dynamics simulation study*. *Biophysical journal*, 2012. **103**(5): p. 1028-1036.
362. Wyman, C., et al., *Unusual oligomerization required for activity of NtrC, a bacterial enhancer-binding protein*. *Science*, 1997. **275**(5306): p. 1658-1661.
363. Vallone, P.M. and J.M. Butler, *AutoDimer: a screening tool for primer-dimer and hairpin structures*. *Biotechniques*, 2004. **37**(2): p. 226-231.
364. Panyutin, I., et al., *G-DNA: a twice-folded DNA structure adopted by single-stranded oligo (dG) and its implications for telomeres*. *Proceedings of the National Academy of Sciences*, 1990. **87**(3): p. 867-870.
365. Burge, S., et al., *Quadruplex DNA: sequence, topology and structure*. *Nucleic acids research*, 2006. **34**(19): p. 5402-5415.
366. Neaves, K.J., et al., *Direct visualization of G-quadruplexes in DNA using atomic force microscopy*. *Nucleic acids research*, 2009. **37**(18): p. 6269-6275.
367. Tsukanov, R., et al., *Detailed study of DNA hairpin dynamics using single-molecule fluorescence assisted by DNA origami*. *The Journal of Physical Chemistry B*, 2013. **117**(40): p. 11932-11942.
368. Nayak, R.K. and A. Van Orden, *Counterion and Polythymidine Loop-Length-Dependent Folding and Thermodynamic Stability of DNA*

- Hairpins Reveal the Unusual Counterion-Dependent Stability of Tetraloop Hairpins*. *The Journal of Physical Chemistry B*, 2013. **117**(45): p. 13956-13966.
369. Grunwell, J.R., et al., *Monitoring the conformational fluctuations of DNA hairpins using single-pair fluorescence resonance energy transfer*. *Journal of the American Chemical Society*, 2001. **123**(18): p. 4295-4303.
370. Ohta, T., et al., *Atomic Force Microscopy Proposes a Novel Model for Stem-Loop Structure That Binds a Heat Shock Protein in the *Staphylococcus aureus* HSP70 Operon*. *Biochemical and biophysical research communications*, 1996. **226**(3): p. 730-734.
371. Varani, G., *Exceptionally stable nucleic acid hairpins*. *Annual review of biophysics and biomolecular structure*, 1995. **24**(1): p. 379-404.
372. Duzdevich, D., et al., *Unusual structures are present in DNA fragments containing super-long Huntingtin CAG repeats*. *PLoS one*, 2011. **6**(2): p. e17119-e17119.
373. Melancon, P., R.R. Burgess, and M.T. Record Jr, *Direct evidence for the preferential binding of Escherichia coli RNA polymerase holoenzyme to the ends of deoxyribonucleic acid restriction fragments*. *Biochemistry*, 1983. **22**(22): p. 5169-5176.
374. Bikard, D., et al., *Folded DNA in action: hairpin formation and biological functions in prokaryotes*. *Microbiology and Molecular Biology Reviews*, 2010. **74**(4): p. 570-588.
375. DeHaseth, P.L., et al., *Nonspecific interactions of Escherichia coli RNA polymerase with native and denatured DNA: differences in the binding behavior of core and holoenzyme*. *Biochemistry*, 1978. **17**(9): p. 1612-1622.
376. Helmann, J.D. and P.L. deHaseth, *Protein-nucleic acid interactions during open complex formation investigated by systematic alteration of the protein and DNA binding partners*. *Biochemistry*, 1999. **38**(19): p. 5959-5967.
377. Aiyar, S.E. and J.D. Helmann, *A mismatch bubble in double-stranded DNA suffices to direct precise transcription initiation by Escherichia coli RNA polymerase*. *Journal of Biological Chemistry*, 1994. **269**(18): p. 13179-13184.
378. Huang, X., F.J.L. de Saro, and J.D. Helmann,  *$\sigma$  factor mutations affecting the sequence-selective interaction of RNA polymerase with -10 region single-stranded DNA*. *Nucleic acids research*, 1997. **25**(13): p. 2603-2609.
379. Callaci, S. and T. Heyduk, *Conformation and DNA binding properties of a single-stranded DNA binding region of  $\sigma$ 70 subunit from*

- Escherichia coli* RNA polymerase are modulated by an interaction with the core enzyme. *Biochemistry*, 1998. **37**(10): p. 3312-3320.
380. Borukhov, S. and E. Nudler, *RNA polymerase: the vehicle of transcription*. *Trends in Microbiology*, 2008. **16**(3): p. 126-134.
381. Zhou, W. and P.W. Doetsch, *Effects of abasic sites and DNA single-strand breaks on prokaryotic RNA polymerases*. *Proceedings of the National Academy of Sciences*, 1993. **90**(14): p. 6601-6605.
382. Salinas-Rios, V., B.P. Belotserkovskii, and P.C. Hanawalt, *DNA slip-outs cause RNA polymerase II arrest in vitro: potential implications for genetic instability*. *Nucleic acids research*, 2011. **39**(17): p. 7444-7454.
383. Belotserkovskii, B.P., S.M. Mirkin, and P.C. Hanawalt, *DNA sequences that interfere with transcription: implications for genome function and stability*. *Chemical reviews*, 2013. **113**(11): p. 8620-8637.
384. Gowrishankar, J., J.K. Leela, and K. Anupama, *R-loops in bacterial transcription: their causes and consequences*. *Transcription*, 2013. **4**(4): p. 153-157.
385. Tous, C. and A. Aguilera, *Impairment of transcription elongation by R-loops in vitro*. *Biochemical and biophysical research communications*, 2007. **360**(2): p. 428-432.
386. Zupancic, M.L. and M.T. Record, *RNA polymerase-promoter interactions: the comings and goings of RNA polymerase*. *Journal of Bacteriology*, 1998. **180**(12): p. 3019-3025.
387. Suzuki, Y., M. Endo, and H. Sugiyama, *Studying RNAP-promoter interactions using atomic force microscopy*. *Methods*, 2015.
388. Wang, F., et al., *The promoter-search mechanism of Escherichia coli RNA polymerase is dominated by three-dimensional diffusion*. *Nature structural & molecular biology*, 2013. **20**(2): p. 174-181.
389. Tafvizi, A., L.A. Mirny, and A.M. van Oijen, *Dancing on DNA: kinetic aspects of search processes on DNA*. *Chemphyschem*, 2011. **12**(8): p. 1481-1489.
390. Mirny, L., et al., *How a protein searches for its site on DNA: the mechanism of facilitated diffusion*. *Journal of Physics A: Mathematical and Theoretical*, 2009. **42**(43): p. 434013.
391. Ucci, J.W. and J.L. Cole, *Global analysis of non-specific protein-nucleic interactions by sedimentation equilibrium*. *Biophysical chemistry*, 2004. **108**(1): p. 127-140.
392. Wade, J.T. and K. Struhl, *Association of RNA polymerase with transcribed regions in Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, 2004. **101**(51): p. 17777-17782.



393. Roe, J.H. and M.T. Record Jr, *Regulation of the kinetics of the interaction of Escherichia coli RNA polymerase with the. lambda. PR promoter by salt concentration*. Biochemistry, 1985. **24**(18): p. 4721-4726.
394. Wang, Y., R.H. Austin, and E.C. Cox, *Single molecule measurements of repressor protein 1D diffusion on DNA*. Physical Review Letters, 2006. **97**(4): p. 048302.
395. Ellis, J., et al., *Direct atomic force microscopy observations of monovalent ion induced binding of DNA to mica*. Journal of microscopy, 2004. **215**(3): p. 297-301.
396. Mantelli, S., et al., *Conformational analysis and estimation of the persistence length of DNA using atomic force microscopy in solution*. Soft Matter, 2011. **7**(7): p. 3412-3416.
397. Shriver, Z., et al., *Heparin and heparan sulfate: analyzing structure and microheterogeneity*, in *Heparin-A Century of Progress*. 2012, Springer. p. 159-176.
398. Rabenstein, D.L., *Heparin and heparan sulfate: structure and function*. Natural product reports, 2002. **19**(3): p. 312-331.
399. Mulloy, B., et al., *N.m.r. and molecular-modelling studies of the solution conformation of heparin*. Biochemical Journal, 1993. **293**(Pt 3): p. 849-858.
400. Walter, G., et al., *Initiation of DNA-Dependent RNA Synthesis and the Effect of Heparin on RNA Polymerase*. European Journal of Biochemistry, 1967. **3**(2): p. 194-201.
401. Pfeffer, S.R., S. Stahl, and M. Chamberlin, *Binding of Escherichia coli RNA polymerase to T7 DNA. Displacement of holoenzyme from promoter complexes by heparin*. Journal of Biological Chemistry, 1977. **252**(15): p. 5403-5407.
402. Stoner, G.E., *Protein Interactions with Materials as Related to Thromboresistance*. Biomaterials, medical devices, and artificial organs, 1973. **1**(1): p. 155-162.
403. Stoner, G.E., S. Srinivasan, and E. Gileadi, *Adsorption inhibition as a mechanism for the antithrombogenic activity of some drugs. I. Competitive absorption of fibrinogen and heparin on mica*. The Journal of physical chemistry, 1971. **75**(14): p. 2107-2111.
404. Zhang, F., et al., *Characterization of interactions between heparin/glycosaminoglycan and adeno-associated virus*. Biochemistry, 2013. **52**(36): p. 6275-6285.
405. Chen, C.H. and H.G. Hansma, *Basement membrane macromolecules: insights from atomic force microscopy*. Journal of structural biology, 2000. **131**(1): p. 44-55.

406. Mitsi, M., et al., *Heparin-mediated conformational changes in fibronectin expose vascular endothelial growth factor binding sites*. *Biochemistry*, 2006. **45**(34): p. 10319-10328.
407. Khurana, R., et al., *A general model for amyloid fibril assembly based on morphological studies using atomic force microscopy*. *Biophysical journal*, 2003. **85**(2): p. 1135-1144.
408. Houska, M., E. Brynda, and K. Bohatá, *The effect of polyelectrolyte chain length on layer-by-layer protein/polyelectrolyte assembly—an experimental study*. *Journal of colloid and interface science*, 2004. **273**(1): p. 140-147.
409. Hopf, C. and W. Hoch, *Heparin Inhibits Acetylcholine Receptor Aggregation at Two Distinct Steps in the Agrin-induced Pathway*. *European Journal of Neuroscience*, 1997. **9**(6): p. 1170-1177.
410. Xu, Y., et al., *Effect of heparin on protein aggregation: Inhibition versus promotion*. *Biomacromolecules*, 2012. **13**(5): p. 1642-1651.
411. Schlax, P.J., M.W. Capp, and T.M. Record Jr, *Inhibition of Transcription Initiation by the lac Repressor*. *Journal of molecular biology*, 1995. **245**(4): p. 331-350.
412. Milan, S., L. D'Ar, and M.J. Chamberlin, *Structural analysis of ternary complexes of Escherichia coli RNA polymerase: ribonuclease footprinting of the nascent RNA in complexes*. *Biochemistry*, 1999. **38**(1): p. 218-225.
413. Brodolin, K., et al., *The  $\sigma 70$  subunit of RNA polymerase induces lacUV5 promoter-proximal pausing of transcription*. *Nature structural & molecular biology*, 2004. **11**(6): p. 551-557.
414. Harden, T.T., et al., *Bacterial RNA polymerase can retain  $\sigma 70$  throughout transcription*. *Proceedings of the National Academy of Sciences*, 2016. **113**(3): p. 602-607.
415. Vo, N.V., et al., *In vitro studies of transcript initiation by Escherichia coli RNA polymerase. 2. Formation and characterization of two distinct classes of initial transcribing complexes*. *Biochemistry*, 2003. **42**(13): p. 3787-3797.
416. Kubori, T. and N. Shimamoto, *A Branched Pathway in the Early Stage of Transcription by Escherichia coli RNA Polymerase*. *Journal of molecular biology*, 1996. **256**(3): p. 449-457.
417. Sen, R., H. Nagai, and N. Shimamoto, *Polymerase arrest at the  $\lambda P_R$  promoter during transcription initiation*. *Journal of Biological Chemistry*, 2000. **275**(15): p. 10899-10904.
418. Wang, F. and C. Reiss, *The collision of cotranscribing E. coli RNA polymerases studied in vitro*. *Biochemistry and molecular biology international*, 1993. **30**(5): p. 969-981.

Chapter 9: Bibliography

419. Bentin, T., et al., *Transcription arrest caused by long nascent RNA chains*. Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression, 2005. **1727**(2): p. 97-105.
420. Dalal, R.V., et al., *Pulling on the nascent RNA during transcription does not alter kinetics of elongation or ubiquitous pausing*. Molecular cell, 2006. **23**(2): p. 231-239.
421. Epshtein, V., et al., *Transcription through the roadblocks: the role of RNA polymerase cooperation*. The EMBO journal, 2003. **22**(18): p. 4719-4727.
422. Palmer, A.C., et al., *Potent transcriptional interference by pausing of RNA polymerases over a downstream promoter*. Molecular cell, 2009. **34**(5): p. 545-555.
423. Giaever, G.N. and J.C. Wang, *Supercoiling of intracellular DNA can occur in eukaryotic cells*. Cell, 1988. **55**(5): p. 849-856.
424. Wu, H.-Y., et al., *Transcription generates positively and negatively supercoiled domains in the template*. Cell, 1988. **53**(3): p. 433-440.
425. Tsao, Y.-P., H.-Y. Wu, and L.F. Liu, *Transcription-driven supercoiling of DNA: direct biochemical evidence from in vitro studies*. Cell, 1989. **56**(1): p. 111-118.
426. Wang, Z. and P. Dröge, *Long-range effects in a supercoiled DNA domain generated by transcription in vitro*. Journal of molecular biology, 1997. **271**(4): p. 499-510.
427. Ma, J., L. Bai, and M.D. Wang, *Transcription under torsion*. Science, 2013. **340**(6140): p. 1580-1583.
428. Ma, J. and M.D. Wang, *RNA polymerase is a powerful torsional motor*. Cell Cycle, 2014. **13**(3): p. 337-338.
429. Ma, J. and M. Wang, *Interplay between DNA supercoiling and transcription elongation*. Transcription, 2014. **5**(3): p. e28636.
430. Dröge, P., *Protein tracking-induced supercoiling of DNA: A tool to regulate DNA transactions in vivo?* Bioessays, 1994. **16**(2): p. 91-99.
431. Levinthal, C. and H. Crane, *On the unwinding of DNA*. Proceedings of the National Academy of Sciences, 1956. **42**(7): p. 436-438.
432. Nelson, P., *Transport of torsional stress in DNA*. Proceedings of the National Academy of Sciences, 1999. **96**(25): p. 14342-14347.
433. Dröge, P. and A. Nordheim, *Transcription-induced conformational change in a topologically closed DNA domain*. Nucleic acids research, 1991. **19**(11): p. 2941-2946.
434. Dröge, P., *Transcription-driven site-specific DNA recombination in vitro*. Proceedings of the National Academy of Sciences, 1993. **90**(7): p. 2759-2763.

435. Kouzine, F., et al., *The dynamic response of upstream DNA to transcription-generated torsional stress*. Nature structural & molecular biology, 2004. **11**(11): p. 1092-1100.
436. Drolet, M., X. Bi, and L.F. Liu, *Hypernegative supercoiling of the DNA template during transcription elongation in vitro*. Journal of Biological Chemistry, 1994. **269**(3): p. 2068-2074.
437. Lilley, D.M., D. Chen, and R.P. Bowater, *DNA supercoiling and transcription: topological coupling of promoters*. Quarterly reviews of biophysics, 1996. **29**(03): p. 203-225.
438. Dunaway, M. and E.A. Ostrander, *Local domains of supercoiling activate a eukaryotic promoter in vivo*. Nature, 1993. **361**(6414): p. 746-748.
439. Thomen, P., U. Bockelmann, and F. Heslot, *Rotational drag on DNA: a single molecule experiment*. Physical review letters, 2002. **88**(24): p. 248102.
440. Nelson, P., *Comment on "Rotational Drag on DNA: A Single Molecule Experiment"*. Physical review letters, 2004. **92**(15): p. 159801.
441. Nelson, P.C., *Spare the (elastic) rod*. Science, 2012. **337**(6098): p. 1045-1046.
442. García-Rubio, M.L. and A. Aguilera, *Topological constraints impair RNA polymerase II transcription and causes instability of plasmid-borne convergent genes*. Nucleic acids research, 2011: p. gkr840.
443. Stupina, V.A. and J.C. Wang, *DNA axial rotation and the merge of oppositely supercoiled DNA domains in Escherichia coli: Effects of DNA bends*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(23): p. 8608-8613.
444. Levens, D. and D.R. Larson, *A new twist on transcriptional bursting*. Cell, 2014. **158**(2): p. 241-242.
445. Chong, S., et al., *Mechanism of transcriptional bursting in bacteria*. Cell, 2014. **158**(2): p. 314-326.
446. Wang, J.C., *Interaction between DNA and an Escherichia coli protein  $\omega$* . Journal of molecular biology, 1971. **55**(3): p. 523-IN16.
447. Sari, L. and I. Andricioaei, *Rotation of DNA around intact strand in human topoisomerase I implies distinct mechanisms for positive and negative supercoil relaxation*. Nucleic acids research, 2005. **33**(20): p. 6621-6634.
448. Redinbo, M.R., et al., *Crystal structures of human topoisomerase I in covalent and noncovalent complexes with DNA*. Science, 1998. **279**(5356): p. 1504-1513.

449. Carey, J.F., et al., *DNA relaxation by human topoisomerase I occurs in the closed clamp conformation of the protein*. Proceedings of the National Academy of Sciences, 2003. **100**(10): p. 5640-5645.
450. Koster, D.A., et al., *Friction and torque govern the relaxation of DNA supercoils by eukaryotic topoisomerase IB*. Nature, 2005. **434**(7033): p. 671-674.
451. Krogh, B.O. and S. Shuman, *Catalytic mechanism of DNA topoisomerase IB*. Molecular cell, 2000. **5**(6): p. 1035-1041.
452. Madden, K., L. Stewart, and J. Champoux, *Preferential binding of human topoisomerase I to superhelical DNA*. The EMBO journal, 1995. **14**(21): p. 5399.
453. Yang, Z., et al., *Recognition of forcible curvature in circular DNA by human topoisomerase I*. Chemical Communications, 2011. **47**(40): p. 11309-11311.
454. Argaman, M., et al., *Revealing the mode of action of DNA topoisomerase I and its inhibitors by atomic force microscopy*. Biochemical and biophysical research communications, 2003. **301**(3): p. 789-797.
455. Moreno-Herrero, F., et al., *Atomic force microscopy shows that vaccinia topoisomerase IB generates filaments on DNA in a cooperative fashion*. Nucleic acids research, 2005. **33**(18): p. 5945-5953.
456. Shuman, S., D.G. Bear, and J. Sekiguchi, *Intramolecular synapsis of duplex DNA by vaccinia topoisomerase*. The EMBO journal, 1997. **16**(21): p. 6584-6589.
457. Collins, I., A. Weber, and D. Levens, *Transcriptional consequences of topoisomerase inhibition*. Molecular and cellular biology, 2001. **21**(24): p. 8437-8451.
458. Mondal, N., et al., *Elongation by RNA polymerase II on chromatin templates requires topoisomerase activity*. Nucleic acids research, 2003. **31**(17): p. 5016-5024.
459. Liu, Z., et al., *Imaging and studying human topoisomerase I on mica surfaces in air and in liquid by atomic force microscopy*. Scanning, 2009. **31**(4): p. 160.
460. Jiang, Y., et al., *Separating DNA with different topologies by atomic force microscopy in comparison with gel electrophoresis*. The Journal of Physical Chemistry B, 2010. **114**(37): p. 12162-12165.
461. Jiang, Y., et al., *Detecting ultraviolet damage in single DNA molecules by atomic force microscopy*. Biophysical journal, 2007. **93**(5): p. 1758-1767.

462. Kim, D.T., H.W. Blanch, and C.J. Radke, *Direct imaging of lysozyme adsorption onto mica by atomic force microscopy*. Langmuir, 2002. **18**(15): p. 5841-5850.
463. Subramani, R., et al., *A novel secondary DNA binding site in human topoisomerase I unravelled by using a 2D DNA origami platform*. ACS nano, 2010. **4**(10): p. 5969-5977.
464. Ishii, K., et al., *Inhibition of topoisomerase I by heparin*. Biochemical and biophysical research communications, 1982. **104**(2): p. 541-547.
465. Ishii, K., et al., *Mechanism of inhibition of mammalian DNA topoisomerase I by heparin*. Biochemical Journal, 1987. **241**(1): p. 111-119.
466. Adhya, S. and M. Gottesman, *Promoter occlusion: transcription through a promoter may inhibit its activity*. Cell, 1982. **29**(3): p. 939-944.
467. Kouzine, F., D. Levens, and L. Baranello, *DNA topology and transcription*. Nucleus, 2014. **5**(3): p. 195-202.
468. Hsu, L.M., *Promoter clearance and escape in prokaryotes*. Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression, 2002. **1577**(2): p. 191-207.
469. Susa, M., T. Kubori, and N. Shimamoto, *A pathway branching in transcription initiation in Escherichia coli*. Molecular microbiology, 2006. **59**(6): p. 1807-1817.
470. McAllister, C.F. and E.C. Achberger, *Effect of polyadenine-containing curved DNA on promoter utilization in Bacillus subtilis*. Journal of Biological Chemistry, 1988. **263**(24): p. 11743-11749.
471. Pérez-Martín, J., F. Rojo, and V. De Lorenzo, *Promoters responsive to DNA bending: a common theme in prokaryotic gene expression*. Microbiological reviews, 1994. **58**(2): p. 268-290.
472. Kouzine, F., et al., *The functional response of upstream DNA to dynamic supercoiling in vivo*. Nature structural & molecular biology, 2008. **15**(2): p. 146-154.
473. Yang, X., et al., *Torsional control of double-stranded DNA branch migration*. Biopolymers, 1998. **45**(1): p. 69-83.
474. Shlyakhtenko, L.S., et al., *Structure and dynamics of supercoil-stabilized DNA cruciforms*. Journal of molecular biology, 1998. **280**(1): p. 61-72.